

Ministério da Saúde  
Fundação Oswaldo Cruz  
Instituto René Rachou  
Programa de Pós-graduação em Ciências da Saúde

ANÁLISE DE ASPECTOS ESTRUTURAIS EM IMUNOINFORMÁTICA UTILIZANDO  
CANDIDATOS VACINAIS CONTRA LEISHMANIOSE QUE FORAM SELECIONADOS  
USANDO VACINOLOGIA REVERSA

por

João Paulo Linhares Velloso

Belo Horizonte  
2018

**DISSERTAÇÃO**

**MCS – IRR J. P. L. VELLOSO**

**2017**

João Paulo Linhares Velloso

ANÁLISE DE ASPECTOS ESTRUTURAIS EM IMUNOINFORMÁTICA UTILIZANDO  
CANDIDATOS VACINAIS CONTRA LEISHMANIOSE QUE FORAM SELECIONADOS  
USANDO VACINOLOGIA REVERSA

Dissertação apresentada ao Programa de Pós-graduação em Ciências da Saúde, do Instituto René Rachou, como requisito parcial para a obtenção do título de mestre em Ciências da Saúde.

Orientador: Dr. Jeronimo Conceição Ruiz  
Coorientadora: Dra. Daniela de Melo Resende

Belo Horizonte - Minas Gerais  
2018

Catálogo-na-fonte  
Rede de Bibliotecas da FIOCRUZ  
Biblioteca do IRR  
CRB/6 1975

V443a Velloso, João Paulo Linhares.  
2018

Análise de aspectos estruturais em imunoinformática utilizando candidatos vacinais contra Leishmaniose que foram selecionados usando vacinologia reversa/João Paulo Linhares Velloso – Belo Horizonte, 2018

XV, 135 f.: il.; 210 x 297mm.

Bibliografia: f. 97-109

Dissertação (mestrado) – Dissertação do título de Mestre em Ciências pelo Programa de Pós - Graduação em Ciências da Saúde do Instituto René Rachou. Área de concentração: Biologia Celular e Molecular, Genética e Bioinformática.

1. *Leishmania braziliensis*/imunologia 2. *Leishmania infantum*/imunologia 3. Palavra-chave I. Título. II. Ruiz, Jeronimo Conceição (Orientação). III. Resende, Daniela de Melo (Coorientação)

CDD – 22. ed. – 616.936

## **Ata de defesa**

## **AGRADECIMENTOS**

Aos meus pais, Edilene e Roberto, por sempre apoiarem minhas decisões, sempre estarem dispostos a me ouvir e acreditarem em mim.

Ao meu irmão, Marco, minha vó (segunda mãe) Any e vó Zélia por estarem dispostos a aceitar minha ausência em momentos de confraternização e sempre apoiarem os meus passos.

Aos orientadores, Jeronimo e Daniela, por todo incentivo, paciência, dedicação na realização deste trabalho e acreditarem no meu potencial.

À Amanda, pela parceria, por todo apoio, carinho, paciência, incentivo e por me ajudar inúmeras vezes na escrita deste trabalho.

Ao Dr. Pedro pela supervisão em meu estágio no IOC e pela contribuição inestimável para a realização deste trabalho.

Ao Dr. Ernesto por me aceitar no IOC durante parte da realização deste trabalho.

À Dra. Carolina pela revisão incansável do meu texto final e pelas contribuições que foram essenciais.

Ao Dr. Alexandre Reis e ao Dr. Rory por terem compartilhado as informações acerca dos antígenos vacinais que foram os objetos de estudo deste trabalho.

Ao Paul por todos os ensinamentos em programação e pela grande amizade.

À Leilane e à Grace por serem colaboradoras no desenvolvimento deste trabalho e me ajudarem a dar os primeiros passos durante o mestrado.

Ao Luiz pelas discussões científicas e ajuda na revisão do texto.

Ao Frederico, pelas dicas de programação e sempre estar disposto a escutar e dar conselhos.

Ao Jáder, pelo suporte computacional, essencial para execução de todas as etapas deste projeto.

Ao Henrique, pelas discussões científicas e momentos de descontração.

Aos outros membros do laboratório, pela convivência harmoniosa e pela amizade.

Aos amigos bioinformatas do anexo, pela amizade, várias ajudas e pelas confraternizações.

Aos meus amigos do Grupo de Biofísica Computacional e Modelagem Molecular, Rafael, Ronald, Vanessa, Deborah, Alexandra, Artur, Valdemir e Gisele que me receberam tão bem no Rio, forneceram preciosas dicas e tornaram minha estadia lá tão agradável.

Ao comitê estudantil, por contribuir para minha formação científica e como pessoa.

Aos docentes da Fiocruz-MG, em especial Laila, pelos vários ensinamentos.

A todos os servidores da Fiocruz-MG que me deram todo suporte necessário para o desenvolvimento deste trabalho, em especial Andrea, Seu Carlos, Shirley, Nilson e Júlio César, por me receberem sempre com um grande sorriso.

A todas as agências de fomento em especial CAPES e Fiocruz-MG, pelo auxílio financeiro recebido. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001.

Aos componentes da banca, por terem aceitado o convite de avaliar este trabalho.

## RESUMO

As leishmanioses são um conjunto de doenças causadas por parasitos do gênero *Leishmania*. A leishmaniose cutânea no Brasil é causada principalmente pela espécie *Leishmania braziliensis* enquanto a leishmaniose visceral é causada pela espécie *L. infantum*. Segundo a Organização Mundial da Saúde, existem cerca de 12 milhões de pessoas infectadas e estima-se que, anualmente, são reportados dois milhões de novos casos (1,5 milhões de casos de leishmaniose cutânea e 500 mil de leishmaniose visceral). Apesar da gravidade da doença, até o momento não existem vacinas para uso humano comercialmente disponíveis, o que poderia ser uma medida efetiva de controle. Neste contexto, abordagens computacionais são utilizadas em estratégias de vacinologia reversa para identificar novos antígenos e permitem a seleção de potenciais candidatos vacinais em um curto tempo e a um custo reduzido. No entanto, a maioria dos preditores de epítomos atualmente utilizada nesse tipo de abordagem não consideram informações estruturais proteicas. Considerando a importância da estrutura nas interações imunológicas, torna-se necessário avaliar os antígenos vacinais utilizando programas que considerem informações estruturais. Neste projeto, analisamos características estruturais de seis proteínas de *L. braziliensis* e seis proteínas de *L. infantum* selecionadas através da vacinologia reversa. Adicionalmente, também foi analisada a afinidade de ligação de epítomos oriundos das proteínas de *L. infantum* anteriormente mencionadas com o Complexo de Histocompatibilidade Principal (MHC) de classes I e II através de atracamento (docking) molecular. Inicialmente, a predição estrutural das proteínas foi feita através de modelagem computacional. Os modelos estruturais foram posteriormente validados e foram empregados na identificação de epítomos conformacionais de células B. Para essas proteínas foi também realizada a predição de regiões desordenadas e a investigação da correlação de desordem estrutural proteica com regiões imunogênicas de células B. Por último usamos análises de simulação de atracamento molecular para quantificar a afinidade de ligação entre epítomos e diferentes moléculas de MHC. Foi possível a modelagem estrutural de seis proteínas, nas quais realizamos a predição de epítomos

conformacionais de células B usando quatro preditores. Em seguida realizamos o consenso da predição desses quatro preditores. Obtivemos alguns possíveis epítopos com características similares às descritas na literatura. Muitas regiões de desordem preditas para as proteínas analisadas coincidiram com regiões de epítopos, o que demonstra uma possível correlação entre estas duas características. Através do atracamento molecular foi possível priorizar os epítopos estudados e adicionar mais um grau de confiança ao resultado obtido através da vacinologia reversa.

**Palavras-chave:** imunoinformática estrutural, vacina, *Leishmania braziliensis*, *Leishmania infantum*, modelagem de proteínas, atracamento molecular, desordem estrutural proteica.



## ABSTRACT

Leishmaniasis is a wide-spectrum of diseases caused by parasites from *Leishmania* genus. In Brazil cutaneous leishmaniasis is mainly caused by *Leishmania braziliensis*, while the main cause of visceral leishmaniasis is *L. infantum*. According to the World Health Organization, two million new cases (1.5 million cases of cutaneous leishmaniasis and 500,000 of visceral leishmaniasis) are estimated to occur annually, with about 12 million people currently infected. Despite the severe problem, there is no human vaccine commercially available, and this could be an effective control measure. In this context, computational approaches have been used in reverse vaccinology strategies in order to discover novel antigens. These approaches allow selection of potential vaccine candidates in a short period of time, including proteins that are not expressed at high levels. Nevertheless, the majority of epitope predictors used in reverse vaccinology does not take into account structural information. Considering the importance of protein structure in immunological interactions and in vaccine development, it is necessary to evaluate vaccine antigen candidates using structural bioinformatic tools. In this project, we analysed structural features of six *L. braziliensis* and six *L. infantum* proteins selected via reverse vaccinology, and also their respective epitopes binding affinity with Major Histocompatibility Complex (MHC) using molecular docking analysis. Initially, protein structures were predicted, and models for these proteins were generated. Afterwards, validated models were used in B cell discontinuous epitope prediction. Disordered regions of these proteins were also predicted using amino acid sequence and the correlation of their occurrence with immunogenic regions were further investigated. By using docking analyses, we quantified the binding affinity between epitopes and different MHC molecules. We could build structural models for six proteins, which we used for B cell discontinuous epitope prediction. After the predictions we obtained the consensus predictions. Through these analyses we found possible B cell epitopes with corresponding B cell epitopes characteristics described in literature. Additionally, many protein regions predicted as disordered co-occurred with B cell epitopes, demonstrating a possible correlation between these two characteristics. The molecular docking analysis added another layer

of data strength to the previous results and allowed prioritizing the study of best epitopes.

**Key words:** structural immunoinformatics, vaccine, *Leishmania braziliensis*, *Leishmania infantum*, protein modeling, molecular docking, structural protein disorder.

## LISTA DE QUADROS

|  |    |
|--|----|
| Quadro 1: Características das proteínas analisadas de <i>L. braziliensis</i> .....   | 36 |
| Quadro 2: Características das proteínas analisadas de <i>L. infantum</i> .....   | 37 |
| Quadro 3: Peptídeos preditos como sendo epítopos lineares pela vacinologia reversa e que foram utilizados nas análises de atracamento molecular..... | 50 |
| Quadro 4: Regiões desconsideradas nas predições devido à divergência estrutural com os moldes.....   | 68 |
| Quadro 5: Trechos de sequências preditos como desordenados pelos programas DisEMBL, GlobPipe, IUPred, e VSL2B.....                                   | 74 |

## LISTA DE FIGURAS

|   |     |
|---|-----|
| Figura 1: Tipos de epítomos reconhecidos por anticorpos: conformacionais e lineares.  | 20  |
| Figura 2: Esquema da modelagem comparativa.....   | 25  |
| Figura 3: Atracamento molecular. ....   | 28  |
| Figura 4: Ângulos diedrais de uma ligação peptídica.. ....  | 42  |
| Figura 5: Gráfico de Ramachandran do modelo da proteína LbrM22_V2.0110.. ....   | 43  |
| Figura 6: Gráfico com os escores DOPE da cadeia A da estrutura 2HJL. ....   | 44  |
| Figura 7: Esquema do delineamento experimental da etapa de Avaliação dos Modelos..  | 45  |
| Figura 8: Esquema do <i>pipeline</i> de predição de desordem proteica. ....   | 48  |
| Figura 9: Esquema do delineamento experimental das simulações de atracamento molecular.....   | 52  |
| Figura 10: Gráficos DOPE para os três modelos selecionados para as proteínas de <i>Leishmania braziliensis</i> .....  | 65  |
| Figura 11: Gráficos DOPE para os três modelos selecionados para as proteínas de <i>Leishmania infantum</i> .. ....  | 66  |
| Figura 12: Alinhamento estrutural entre modelos e moldes.....   | 67  |
| Figura 13: Alinhamento entre a proteína LinJ.28.1850 e seu respectivo molde.....  | 69  |
| Figura 14: Regiões preditas como imunogênicas nos modelos das proteínas <i>Leishmania braziliensis</i> .....  | 70  |
| Figura 15: Regiões preditas como imunogênicas nos modelos das proteínas de <i>L. infantum</i> . ....  | 71  |
| Figura 16: Estrutura das proteínas de <i>L. braziliensis</i> destacando os resíduos classificados como epítomos conformacionais de células B consenso entre os preditores. .... | 72  |
| Figura 17: Estrutura das proteínas de <i>L. infantum</i> destacando os resíduos classificados como epítomos conformacionais de células B consenso entre os preditores. ....     | 74  |
| Figura 18: Localização das regiões preditas como desordenadas nos modelos estruturais das proteínas .....   | 75  |
| Figura 19: Trechos proteicos preditos como desordenados que coincidem com predições consenso de epítomos conformacionais de célula B.....                                       | 76  |
| Figura 20: Possibilidades de conformações geradas pelo atracamento molecular .....  | 77  |
| Figura 21: Resultado das simulações de atracamento molecular entre os peptídeos avaliados, peptídeos originais e moléculas de MHC I.....  | 80  |
| Figura 22: Resultado das simulações de atracamento molecular entre os peptídeos avaliados, peptídeos originais e moléculas de MHC II.....                                       | 81  |
| Figura 23: Gráfico de Ramachandran das proteínas modeladas de <i>L. braziliensis</i> .....  | 126 |
| Figura 24: Gráfico de Ramachandran das proteínas modeladas de <i>L. infantum</i> .....  | 127 |

## LISTA DE TABELAS

|  |     |
|--|-----|
| Tabela 1: Moldes proteicos identificados para <i>L. braziliensis</i> .....   | 55  |
| Tabela 2: Moldes proteicos identificados para <i>L. infantum</i> . ....  | 56  |
| Tabela 3: Características dos possíveis moldes para cada uma das proteínas modeladas.<br>.....   | 57  |
| Tabela 4: Melhores moldes identificados para cada proteína de acordo com o servidor<br>I-TASSER. ....  | 60  |
| Tabela 5: Melhores moldes identificados para cada proteína de acordo com o servidor<br>RaptorX.....  | 61  |
| Tabela 6: Valores de pontuações dos modelos obtidos pelas três abordagens após a<br>submissão dos mesmos à minimização energética.....   | 62  |
| Tabela 7: Ranqueamento dos peptídeos analisados em função do número de alelos para<br>os quais as simulações apresentaram I <sub>sc</sub> menor em comparação à simulação MHC I-<br>peptídeo cocrystalizado. ....  | 82  |
| Tabela 8: Ranqueamento dos peptídeos analisados em função do número de alelos para<br>os quais as simulações apresentaram I <sub>sc</sub> menor em comparação à simulação MHC II-<br>peptídeo cocrystalizado. .... | 82  |
| Tabela 9: Resultados da análise de similaridade para <i>L. braziliensis</i> .....  | 120 |
| Tabela 10: Resultados da análise de similaridade para <i>L. infantum</i> .....   | 123 |
| Tabela 11: Posição dos resíduos nas proteínas de estudo identificados como parte de<br>epítomos de células B conformacionais pelos preditores Bepro, Discotope e EPCES....   | 128 |
| Tabela 12: Coordenadas dos resíduos nas proteínas de estudo identificados como parte<br>de epítomos de células B pelo Ellipro. ....  | 130 |
| Tabela 13: Valores de I <sub>sc</sub> para atracamentos entre peptideos e MHC de classe I .....  | 133 |
| Tabela 14: Valores de I <sub>sc</sub> para atracamentos entre peptideos e MHC de classe II .....   | 134 |

## LISTA DE ABREVIATURAS E SIGLAS

**BLAST** - Basic Local Alignment Search Tool

**DOPE** - Discrete Optimized Protein energy

**IFN** - Interferon

**Ig** - Imunoglobulinas

**LOMETS** - Local Meta-Threading-Server

**MHC** - Complexo Principal de Histocompatibilidade

**NCBI** - National Center for Biotechnology Information

**PDB** - Protein Data Bank

**RMSD** - Desvio da raiz do valor quadrático médio

## SUMÁRIO

|       |   |    |
|-------|---|----|
| 1     | INTRODUÇÃO.....   | 17 |
| 1.1   | Vacinologia.....  | 17 |
| 1.2   | Vacinologia estrutural.....   | 18 |
| 1.3   | Resposta imune humoral.....   | 19 |
| 1.4   | Regiões desordenadas e imunologia.....  | 20 |
| 1.5   | Resposta imune celular e processamento de antígenos.....  | 21 |
| 1.6   | Determinação Estrutural.....  | 22 |
| 1.6.1 | Métodos baseados em molde.....  | 23 |
| 1.6.2 | Otimização da geometria.....  | 26 |
| 1.6.3 | Avaliação dos Modelos.....  | 26 |
| 1.7   | Métodos de Atracamento Molecular ( <i>Docking</i> ).....  | 27 |
| 1.8   | Leishmanioses.....  | 29 |
| 1.8.1 | Imunidade na Leishmaniose.....  | 29 |
| 1.8.2 | Vacinas existentes para Leishmaniose.....   | 30 |
| 2     | JUSTIFICATIVA.....  | 32 |
| 3     | OBJETIVO GERAL.....   | 34 |
| 3.1   | Objetivos específicos - Parte 1.....  | 34 |
| 3.2   | Objetivos específicos - Parte 2.....  | 34 |
| 4     | MÉTODOS.....  | 35 |
| 4.1   | Parte 1- Predição estrutural das proteínas, predição de epítomos de células B conformacionais e correlação com desordem.....  | 35 |
| 4.1.1 | Origem dos dados utilizados na primeira parte deste trabalho.....   | 35 |
| 4.1.2 | Predição estrutural das proteínas.....  | 39 |
| 4.1.3 | Predição de epítomos conformacionais de células B.....  | 46 |
| 4.1.4 | Predição de regiões desordenadas.....   | 47 |
| 4.2   | Parte 2 - Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular..... | 50 |
| 4.2.1 | Origem dos dados utilizados na segunda parte deste trabalho.....  | 50 |
| 4.2.2 | Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular.....           | 52 |
| 5     | RESULTADOS.....   | 55 |
| 5.1   | Parte 1 - Predição de epítomos de células B conformacionais e correlação com desordem.....                                    | 55 |

|       |  |                                      |
|-------|--|--------------------------------------|
| 5.1.1 | Predição estrutural das proteínas .....  | 55                                   |
| 5.1.2 | Predição de epítomos conformacionais de células B .....  | 67                                   |
| 5.1.3 | Predição de regiões desordenadas.....  | 74                                   |
| 5.2   | Parte 2- Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular..... | 77                                   |
| 5.2.1 | Atracamento molecular aplicado à análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II.....      | 77                                   |
| 6     | DISCUSSÃO.....   | 84                                   |
| 6.1   | Parte 1- Predição estrutural, predição de epítomos de células B conformacionais e correlação com desordem .....              | 84                                   |
| 6.1.1 | Modelagens .....   | 84                                   |
| 6.1.2 | Predição de epítomos conformacionais de células B .....  | 86                                   |
| 6.1.3 | Predição de regiões desordenadas.....  | 90                                   |
| 6.2   | Parte 2 – Análise estrutural da afinidade de ligação de peptídeos com MHC de classe I e II usando atracamento molecular..... | 91                                   |
| 7     | CONCLUSÕES.....  | 94                                   |
| 8     | PERSPECTIVAS .....   | 95                                   |
|       | REFERÊNCIAS .....  | <b>Erro! Indicador não definido.</b> |
|       | Apêndice A - Seleção dos moldes.....   | 110                                  |
|       | Apêndice B - Modelagem por Homologia usando Modeller .....   | 112                                  |
|       | Apêndice C - Comandos utilizados nos preditores de epítomos de Cél. B conformacionais ...                                    | 115                                  |
|       | Apêndice D - Execução do <i>pipeline</i> de predição de desordem proteica .....  | 116                                  |
|       | Apêndice E – Script em python para a troca dos peptídeos cocristalizados pelos peptídeos das proteínas em estudo .....       | 117                                  |
|       | Apêndice F - Comandos utilizados no atracamento molecular .....  | 119                                  |
|       | Apêndice G - Resultados detalhados da análise de similaridade para <i>L. braziliensis</i> .....                              | 120                                  |
|       | Apêndice H - Resultados detalhados da análise de similaridade para <i>L. infantum</i> .....                                  | 123                                  |
|       | Apêndice I – Gráficos de Ramachandran para os melhores modelos de cada uma das abordagens.....                               | 126                                  |
|       | Apêndice J - Predições do Bepro, Discotope e EPCES e Ellipro .....   | 128                                  |
|       | Apêndice K – Tabela com os valores de $I_{sc}$ para atracamentos entre peptídeos e MHC de classe I .....                     | 133                                  |
|       | Apêndice L – Tabela com os valores de $I_{sc}$ para atracamentos entre peptídeos e MHC de classe II.....                     | 134                                  |



# 1 INTRODUÇÃO

## 1.1 Vacinologia

O desenvolvimento de técnicas de sequenciamento de DNA que iniciaram em 1970 permitiu a determinação de genomas em poucas semanas e atualmente milhares de organismos possuem seu genoma depositado em bancos de dados de domínio público, incluindo muitos patógenos. A disponibilidade desses dados permitiu o surgimento da vacinologia reversa, técnica que visa descobrir antígenos vacinais através da predição tanto das proteínas codificadas pelos genes de determinado organismo, quanto dos possíveis epítomos existentes nas mesmas, usando programas computacionais. Existem atualmente vários programas capazes de prever características importantes para geração de uma resposta imune efetiva como por exemplo: localização subcelular, tipo de processamento de antígenos ou ativação de células B e T. A busca por antígenos deve sempre considerar as características de cada organismo e o conjunto de preditores usado deve ser avaliado para a otimização da predição.

O refinamento da vacinologia reversa permitiu o desenvolvimento de vacinas mais seguras, com menos efeitos adversos, reduzindo componentes não protetores e capazes de causar alergias ou doenças autoimunes. No ano 2000, Pizza e colaboradores usaram o genoma de uma cepa virulenta de *Neisseria meningitidis* para identificar candidatos vacinais (PIZZA et al., 2000). Um total de 350 antígenos selecionados foram expressos em *Escherichia coli*, permitindo assim, a identificação de proteínas expressas em superfície celular, conservadas em várias outras cepas e capazes de induzir respostas de anticorpos. Esse trabalho apoiou o desenvolvimento de uma vacina que em 2013 teve sua comercialização liberada pela Comissão Europeia (KUHDARI et al., 2016).

No entanto, a vacinologia reversa ainda não é capaz de identificar antígenos não proteicos, como polissacarídeos e glicolipídios. Além disso, a maioria dos preditores não considera aspectos estruturais das proteínas, como os epítomos conformacionais e interações entre porções conformacionalmente próximas, que são essenciais na resposta imune.

## 1.2 Vacinologia estrutural

Nas últimas décadas, os avanços em técnicas experimentais de determinação estrutural de biomoléculas, como cristalografia por raios-X e espectrometria por ressonância magnética nuclear permitiu a elucidação estrutural de várias proteínas de interesse biológico. Atualmente no Protein Data Bank (PDB) (BERMAN et al., 2000), banco de dados de domínio público de estruturas proteicas e ácidos nucleicos, estão depositadas mais de 130.000 estruturas. Além disso, existem vários métodos computacionais capazes de prever a estrutura de uma proteína, como a modelagem *ab initio* e as modelagens baseadas em molde (DORMITZER; ULMER; RAPPUOLI, 2008).

Considerando que as interações entre alguns componentes do sistema imune e os patógenos são dependentes de aspectos estruturais, a disponibilidade desse tipo de informação apoiou uma nova área de estudo, a imunoinformática estrutural. Os dados estruturais de alta resolução revelam superfícies eletrostáticas e arquiteturas das regiões de uma determinada proteína analisada que podem ser reconhecidas por anticorpos e células do sistema imune, sendo assim extremamente úteis durante a produção das vacinas, uma vez que podem indicar quais porções do epítipo devem ficar intactas e quais podem ser alteradas (DORMITZER; ULMER; RAPPUOLI, 2008). Adicionalmente, a informação estrutural permite o desenvolvimento de vacinas termodinamicamente mais estáveis, homogêneas, eficientemente produzidas, tornando a imunização mais prática e acessível (DORMITZER; ULMER; RAPPUOLI, 2008).

Apesar desse grande avanço na elucidação estrutural de proteínas citado anteriormente, muitas proteínas de tripanossomatídeos ainda não possuem estruturas resolvidas pelos métodos experimentais. Podemos citar como exemplo as espécies: *Leishmania (Viannia) braziliensis* em que a busca em bancos de dados de estruturas proteicas retornou nove proteínas; e *Leishmania (Leishmania) infantum*, em que o mesmo tipo de busca retornou 18 proteínas.

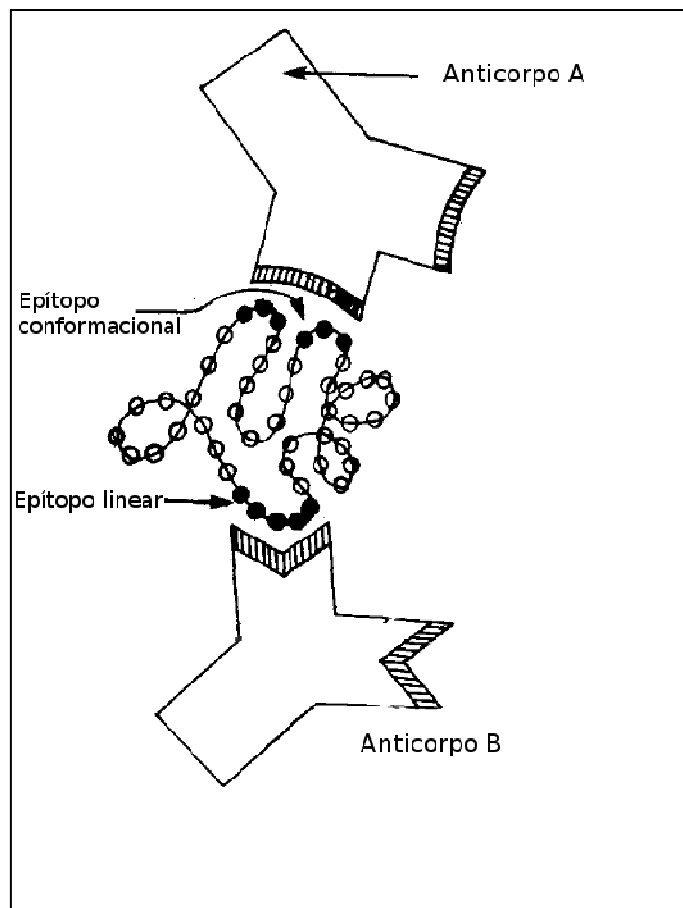
Portanto, para estudo envolvendo estrutura de proteínas em vacinologia estrutural muitas vezes é necessária a utilização de métodos experimentais ou computacionais. Como exemplo pode-se citar a modelagem computacional baseada em molde, a qual permite a construção de um modelo para uma proteína a partir da

sequência de seus resíduos utilizando informações estruturais de proteínas homólogas elucidadas experimentalmente.

### **1.3 Resposta imune humoral**

A resposta imune adaptativa, que pode ser dividida em celular ou humoral, é a maior responsável por eliminar antígenos estranhos do corpo humano. Dentre as células envolvidas nesse tipo de resposta imune, as células B são importantes pela sua habilidade ampla de neutralização e opsonização de microorganismos através de anticorpos, que são secretados pelos plasmócitos, células diferenciadas a partir da ativação dos linfócitos B. Os antígenos, por sua vez, são reconhecidos pelos anticorpos através de trechos específicos denominados epítopos. O contato entre antígeno e anticorpo ocorre a partir de complementaridade estrutural e afinidade entre resíduos do epítipo do antígeno e do parátipo (o lugar onde se unem o antígeno e o anticorpo) do anticorpo, que faz parte da região variável responsável pelo reconhecimento.

Existem dois tipos de epítopos reconhecidos por anticorpos: lineares e conformacionais (Figura 1). Os epítopos lineares são definidos como uma sequência contínua de resíduos, enquanto os conformacionais são definidos como resíduos que não são contíguos, mas que são posicionados próximos uns dos outros no dobramento do epítipo (VAN REGENMORTEL, 1986).



**Figura 1: Tipos de epítopos reconhecidos por anticorpos: conformacionais e lineares.** O Anticorpo A está interagindo com um epítipo conformacional enquanto o anticorpo B está interagindo com um epítipo linear. Resíduos participantes das interações estão indicados em preto (modificada de VAN REGENMORTEL, 1986).

#### 1.4 Regiões desordenadas e imunologia

Apesar de alguns estudos de três décadas atrás (BARLOW et al., 1988; LYNCH; RISEMAN; BRETSCHER, 1987) demonstrarem a existência de proteínas e domínios de proteínas capazes de realizar suas funções biológicas sem uma conformação estrutural definida e estável, somente em 1999 foi observada a generalidade desse fenômeno (WRIGHT; DYSON, 1999).

O aumento de estudos relacionados à identificação e à caracterização funcional de desordem estrutural apoiou o desenvolvimento de programas computacionais para a predição dessa característica. Muitos desses programas seguem a premissa de que uma vez que a sequência de aminoácidos determina a estrutura adotada por uma proteína, a sequência também pode determinar a falta de uma estrutura definida (DUNKER; OBRADOVIC, 2001). Desse modo, a predição de regiões desordenadas pode ser feita utilizando-se como referência a sequência de resíduos. Podemos citar

como exemplo os programas DisEMBL (LINDING et al., 2003), GlobPipe (LINDING, 2003), IUPred (DOSZTÁNYI et al., 2005), e VSL2B (PENG et al., 2006).

Com base em predições computacionais, foi também recentemente demonstrado que patógenos eucariotos, como tripanossomatídeos (RUY et al., 2014), *Plasmodium* e *Toxoplasma* (MOHAN et al., 2008), possuem uma média de 30% de proteínas com regiões desordenadas nos seus respectivos proteomas e muitas dessas proteínas estão envolvidas em funções cruciais, como regulação gênica, transdução de sinal e reconhecimento de biomoléculas (DYSON; WRIGHT, 2005; WRIGHT; DYSON, 2015). Além disso, Dunker e colaboradores demonstraram que essas proteínas com desordem são frequentemente *hubs* em redes de interação proteína-proteína (DUNKER et al., 2005).

A correlação de desordem estrutural proteica com a resposta imune é ainda muito controversa. Alguns estudos apontam que essas regiões elicitam resposta imune fraca (DUNKER et al., 2002) e que a flexibilidade poderia em alguns casos servir como meio de evasão ao sistema imune. No entanto, há também relatos de que essas regiões são imunodominantes (STAHL et al., 1984).

### **1.5 Resposta imune celular e processamento de antígenos**

A resposta imune celular é caracterizada, principalmente, pela ativação das células T CD8<sup>+</sup> e T CD4<sup>+</sup> a partir da apresentação de antígenos através da interação entre seus receptores e o complexo principal de histocompatibilidade (MHC) associado a um peptídeo. Nesse contexto, as moléculas de MHC de classes I e II possuem funções semelhantes: estão envolvidas na apresentação de antígenos para células T CD8<sup>+</sup> e T CD4<sup>+</sup>, respectivamente, e possuem alto índice de polimorfismos e em termos gerais apresentam estruturas semelhantes. No entanto, diferem na origem dos antígenos, sendo que para a classe I a origem é intracelular e para classe II é extracelular (NEEFJES et al., 2011).

As moléculas de MHC de classe I são expressas em todas as células nucleadas e apresentam fragmentos proteicos na superfície celular. Esses fragmentos proteicos, denominados antígenos, são derivados da degradação de proteínas por proteassomas nucleares e citosólicos e transportados por transportadores de antígeno peptídico até o retículo endoplasmático. Nesse local os antígenos são

inseridos em moléculas do MHC de classe I e transportados via complexo de Golgi para a membrana plasmática (NEEFJES et al., 2011).

As moléculas de MHC de classe II são expressas por células apresentadoras de antígenos profissionais, como as células dendríticas, linfócitos B e macrófagos. As cadeias desse complexo são montadas no retículo endoplasmático sem o antígeno, mas com um peptídeo estabilizador denominado peptídeo classe II associado. Antígenos derivados da degradação endossômica de proteína exógena substituem esses peptídeos associados e o MHC de classe II é transportado para a membrana plasmática (NEEFJES et al., 2011). Quando linfócitos T CD4+ reconhem MHC de classe II contendo antígeno em linfócitos B, eles enviam sinais ativadores para esses linfócitos B. Os sinais ativadores são citocinas interleucinas IL-6, IL-5, IL-4 e ligante de CD40 que induzem a transformação de linfócitos B em plasmócitos, células produtoras de anticorpos (MURPHY et al., 2008).

A afinidade do MHC pelo peptídeo é essencial para imunogenicidade e já foi descrita como crítica na seleção de novos antígenos vacinais. De acordo com Sette e colaboradores (1994), peptídeos com alta afinidade por MHC de classe II foram altamente correlacionados com indução de resposta imune enquanto epítomos com interações por MHC menos ávidas foram associados em menor proporção (60% dos casos) com imunogenicidade (SETTE et al., 1994). No caso de MHC de classe I, essa associação também já foi demonstrada (CHEN et al., 1994).

## **1.6 Determinação Estrutural**

A determinação estrutural de uma proteína e seu padrão de enovelamento provêm informações importantes, como a provável função biológica, e podem inclusive auxiliar a busca de novas vacinas. O enovelamento de uma proteína é capaz de proporcionar a interação de resíduos sequencialmente distantes em uma mesma região espacial da proteína. Através dessas interações e de determinadas condições físico-químicas ocorre a formação de diferentes padrões de conformação das proteínas.

A determinação estrutural pode ser realizada de forma experimental e através de métodos teóricos. Métodos teóricos incluem métodos *ab initio* e baseados em molde. Os métodos *ab initio* buscam construir o modelo estrutural das proteínas baseando-se em princípios físicos e não usam informações de estruturas elucidadas

que possuem similaridade com a sequência a ser modelada. Tipicamente, a modelagem *ab initio* conduz a uma busca conformacional que tem como objetivo a identificação de conformações energeticamente favoráveis. Esse processo gera possíveis conformações para uma determinada sequência de resíduos e a partir dessas conformações são derivados modelos estruturais para a proteína em estudo (RIGDEN, 2009). Já os métodos baseados em molde dependem de estruturas de proteínas que já foram determinadas experimentalmente (SALI; BLUNDELL, 1993). Tendo como ponto de partida a sequência de resíduos de uma determinada proteína, esses métodos utilizam as informações estruturais já existentes de proteínas relacionadas e têm como objetivo prever modelos estruturais com grande acurácia.

Essas estratégias computacionais permitem a geração rápida de modelos proteicos que podem ser usados no desenvolvimento de fármacos; análises de função proteica, interações e comportamento antigênico (através de preditores de epítomos de células B); e no desenho racional de proteínas com maior estabilidade ou com novas funções. Adicionalmente, são as únicas opções quando as técnicas experimentais se mostram ineficientes, como é o caso quando as proteínas são muito grandes para serem cristalizadas (GU; BOURNE, 2009).

### **1.6.1 Métodos baseados em molde**

Os métodos baseados em molde usam como ponto de partida a sequência de resíduos de uma determinada proteína, e utilizando as informações estruturais já existentes de proteínas relacionadas, são capazes de prever modelos estruturais com grande acurácia. Podemos citar duas abordagens bastante usadas a modelagem comparativa e a modelagem por *Threading*.

Esses métodos se baseiam na observação de que a estrutura terciária de uma proteína é mais conservada do que a sua sequência de aminoácidos. Desse modo, mesmo proteínas que são divergentes quando comparamos suas sequências de aminoácidos, podem compartilhar propriedades estruturais. Nesse contexto, é possível através dos métodos computacionais de modelagem baseados em molde, utilizarmos a estrutura terciária de uma proteína com estrutura já elucidada experimentalmente como molde para a modelagem estrutural de uma proteína de interesse (KACZANOWSKI; ZIELENKIEWICZ, 2010).

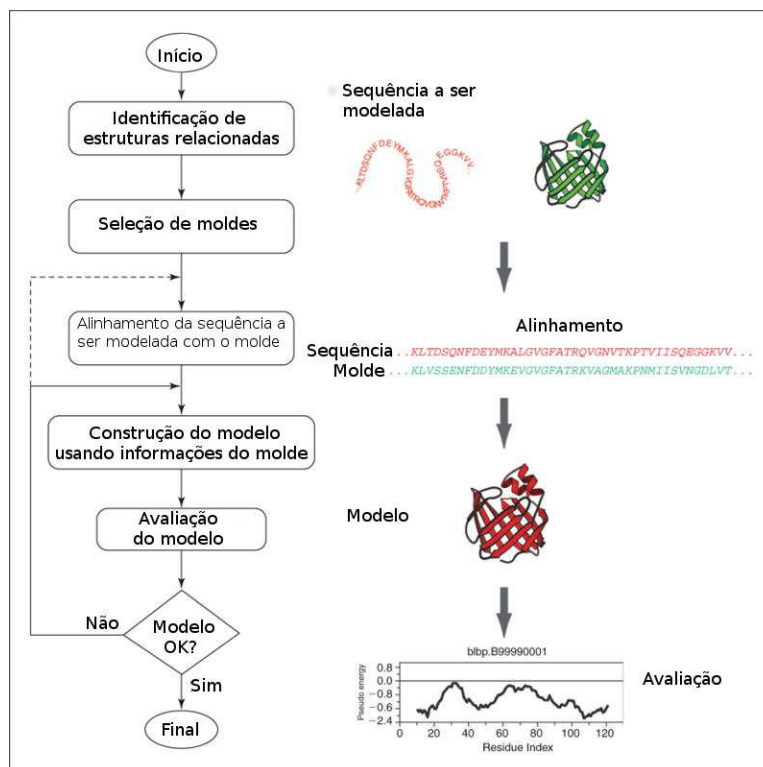
### 1.6.1.1 Modelagem comparativa por meio de Satisfação de Restrição Espaciais

Como dito anteriormente, a modelagem comparativa é baseada na observação de que a estrutura terciária de uma proteína tende a ser mais conservada do que a sua sequência de aminoácidos. A modelagem comparativa é possível uma vez que pequenas mudanças na sequência de resíduos de uma determinada proteína, geralmente, acarretam pequenas mudanças na conformação estrutural dessa proteína (MARTÍ-RENOM et al., 2000). Essa abordagem consiste em quatro passos sequenciais (Figura 2). O primeiro passo é a identificação de proteínas com estrutura elucidada que estão relacionadas com a proteína em estudo. Nesse passo pode ser usado um programa de busca por similaridade de sequências como o programa BLAST (Basic Local Alignment Search Tool) (ALTSCHUL et al., 1990). O programa compara a sequência da proteína a ser modelada com o banco de dados contendo proteínas com estrutura definida, o PDB (BERMAN et al., 2000). O segundo passo consiste no alinhamento das sequências similares com a proteína a ser modelada e a definição de quais proteínas serão usadas na modelagem, considerando o alinhamento e a similaridade entre as proteínas. As proteínas selecionadas para serem usadas na modelagem são chamadas moldes.

O terceiro passo é a construção do modelo da proteína com base no alinhamento com as proteínas moldes. Nesse passo são geradas as estruturas: da cadeia principal, considerando as coordenadas do nitrogênio proveniente do grupo amina do aminoácido, do carbono alfa e da carbonila da proteína molde; dos *loops*; e das cadeias laterais considerando isomerismos conformacionais. Essa etapa busca satisfazer os impedimentos estéricos espaciais obtidos através do alinhamento da proteína em estudo com os modelos (GU; BOURNE, 2009). Os impedimentos estéricos são expressos como funções de densidade de probabilidade e estão diretamente relacionados às características específicas da região atômica associada àquele impedimento específico. Por exemplo, a probabilidade de uma determinada distância entre dois carbonos alfa de uma estrutura proteica desconhecida é uma função Gaussiana com média igual à distância equivalente na estrutura usada como molde (ŠALI; OVERINGTON, 1994). O último passo é a avaliação dos modelos



obtidos. Caso necessário, o passo de alinhamento ou construção é repetido até a obtenção de um modelo sem regiões estericamente instáveis (ŠALI et al., 1995).



**Figura 2: Esquema da modelagem comparativa.** Quatro passos da modelagem comparativa: 1) Seleção de moldes; 2) Alinhamento de seqüências entre molde e proteína alvo; 3) Construção do modelo da proteína alvo; 4) Avaliação do modelo (modificado de WEBB; SALI, 2016).

Um dos programas mais utilizados atualmente para a modelagem comparativa é o MODELLER (WEBB; SALI, 2016). Especificamente, a estratégia empregada pelo programa deriva distâncias e ângulos diedrais associados aos impedimentos estéricos na seqüência a ser modelada com base no alinhamento dela com a estrutura proteica molde. Em seguida, os impedimentos estéricos e os requerimentos energéticos condizentes com uma estereoquímica apropriada são combinados em uma função objetiva. Em uma terceira etapa, os modelos são obtidos por otimização da função objetiva no espaço cartesiano. Essa otimização é realizada através do uso do método *variable target function*, que emprega gradientes conjugados e dinâmica molecular com anelamento simulado. Esse procedimento otimiza as posições de todos os átomos ao mesmo tempo (ŠALI et al., 1995).

### 1.6.1.2 Modelagem por *Threading*

A predição de estruturas de proteínas utilizando a abordagem *threading* se inicia com a computação de vários modelos com base em estruturas depositadas nos bancos de dados para sequência de proteína a ser modelada. Em seguida esses modelos são avaliados para se determinar a compatibilidade entre o que foi gerado e estruturas moldes. (GIBAS; JAMBECK, 2001).

Assim como na abordagem de modelagem comparativa, se a identidade entre a sequência a ser modelada e o molde for abaixo de 30%, o enovelamento predito pode ser completamente errado (GIBAS; JAMBECK, 2001). Podemos citar dois servidores Web que utilizam essa abordagem, o I-TASSER (ROY; KUCUKURAL; ZHANG, 2010) e o RaptorX (KÄLLBERG et al., 2012).

### 1.6.2 Otimização da geometria

Os modelos estruturais preditos têm muitas vezes rearranjos incorretos de *loops* e posicionamento incorreto de elementos da estrutura secundária (KRIEGER et al., 2009). Sendo assim precisam passar por refinamento estrutural. O refinamento serve para ajustar e corrigir conformações estruturais desfavoráveis, incluindo interações atômicas e comprimento de ligações químicas. Esse processo de otimização estrutural geométrico da proteína envolve o estabelecimento de uma condição estrutural que representa o mínimo de energia livre possível. Podemos citar como exemplo o servidor *Web* YASARA que emprega um modelo matemático próprio para a realização de minimização de energia (KRIEGER et al., 2009). Também é possível usar programas como GROMACS (BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995) e Amber (SALOMON-FERRER; CASE; WALKER, 2013).

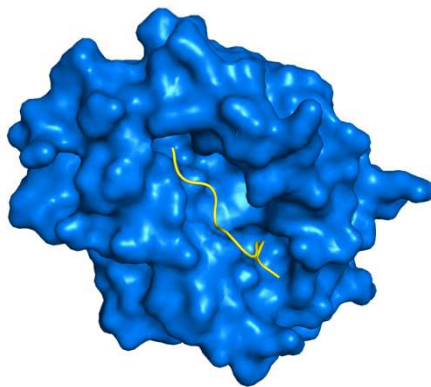
### 1.6.3 Avaliação dos Modelos

Após a modelagem, é importante a avaliação do modelo para detecção de possíveis erros. O número de erros está relacionado à porcentagem de identidade entre a proteína e o molde e ao número de erros da estrutura molde. Se a identidade entre toda extensão da proteína a ser modelada e a proteína molde é maior que 90%, a acurácia do modelo pode ser comparada à de estruturas determinadas por cristalografia (CHOTHIA; LESK, 1986; SIPPL, 1993). Se a identidade está entre

50% e 90%, o erro relacionado à média da distância entre os átomos das coordenadas dos modelos pode ser da ordem de 1,5 Å, com consideráveis erros locais. Se a identidade é menor que 25%, os modelos gerados tendem a conter erros muito grandes (GU; BOURNE, 2009). A minimização de energia e a avaliação dos modelos são usadas conjuntamente para otimizar a geometria de uma molécula.

### **1.7 Métodos de Atracamento Molecular (*Docking*)**

O método de atracamento molecular (*docking*) é amplamente utilizado na compreensão da interação entre biomoléculas. Uma das principais aplicações do método consiste na predição de interação favorável ou não de ligantes que ainda não foram testados experimentalmente. Como exemplo, podemos citar a interação entre peptídeos e moléculas de MHC que está representada na Figura 3. Nessa abordagem é possível obter orientações otimizadas energeticamente dos peptídeos em relação às moléculas de MHC e pontuações acerca da estabilidade dos posicionamentos espaciais dos mesmos, considerando aspectos estruturais e eletroquímicos, como efeitos estéricos de outras moléculas e interações do tipo ligação de hidrogênio, dipolo-dipolo ou dispersão. Podemos citar dois trabalhos recentes que utilizaram essa abordagem para a seleção de antígenos vacinais de *Leishmania* o trabalho de Silva e colaboradores em 2016 e o de Dikhit e colaboradores em 2017 nos quais, além de atracamento molecular, foram utilizadas simulações de dinâmica molecular. Essa abordagem computacional permite a redução do tempo e custo envolvidos na descoberta de novos antígenos.



**Figura 3: Atracamento molecular.** Peptídeo (em amarelo) na fenda de ligação da molécula de MHC de classe I, código do PDB 2HJL (em azul). (GILLESPIE et al., 2006)

A interação entre proteínas baseia-se na complementaridade de aspectos físico-químicos e estruturais das macromoléculas que estão interagindo. Os aspectos físico-químicos se relacionam com as interações intermoleculares existentes no complexo e conferem o grau de afinidade e de especificidade do ligante pela macromolécula. As interações em questão se tratam de ligações de hidrogênio, interações hidrofóbicas, interações do tipo dipolo-induzido, interações eletrostáticas e ligações covalentes. Os aspectos estruturais estão relacionados aos arranjos espaciais moleculares que compreendem variações na orientação, posicionamento espacial e rotações de ligações químicas entre as moléculas (VERLI, 2014).

Para predição do modo e a afinidade de ligação, usando atracamento molecular, são necessários métodos de busca e funções de avaliação. Os métodos de busca são utilizados para predizer as possíveis orientações do ligante interagindo ou não com o alvo proteico. Para isso ele considera complementaridade geométrica, flexibilidade das moléculas e interações interatômicas (DIAS; DE AZEVEDO JR., 2008). Como resultado desses métodos de busca são obtidas várias orientações preditas e para o ranqueamento das mesmas em relação à afinidade e energia livre de ligação são empregadas as funções de avaliação. Estas funções podem ser divididas em três classes: funções baseadas em campos de força, funções empíricas e funções baseadas em conhecimento. As funções baseadas em campos de força fundamentam-se em estimar a energia livre entre moléculas através da soma de interações intermoleculares. As funções empíricas se baseiam em comparar aspectos estruturais simulados com dados experimentais. As funções baseadas em conhecimento utilizam dados estatísticos de contatos intermoleculares

identificados em estruturas depositadas em bancos de dados para avaliação da conformação simulada (DIAS; DE AZEVEDO JR., 2008).

## **1.8 Leishmanioses**

As leishmanioses são um conjunto de doenças negligenciadas com grande número de casos por todo o mundo, sendo causadas por parasitos do gênero *Leishmania*. Em humanos podem gerar lesões cutâneas, no caso da leishmaniose cutânea, e manifestações viscerais no caso da leishmaniose visceral. No Brasil, o maior número de casos de leishmaniose cutânea é causado pela espécie *Leishmania braziliensis* e no caso da manifestação visceral pela espécie *Leishmania infantum*. Nos dois casos os vetores são os flebotomíneos, que possuem ampla distribuição no Brasil, sendo a espécie *Lutzomyia longipalpis* uma das mais comuns.

Atualmente existem poucos fármacos disponíveis para o tratamento das leishmanioses. O Sistema Único de Saúde preconiza o Antimoniato de meglumina para o tratamento da leishmaniose tegumentar e os medicamentos antimoniato de N-metil glucamina, a anfotericina B lipossomal e o desoxicolato de anfotericina B para o tratamento de leishmaniose visceral. Nesse sentido o desenvolvimento de uma vacina efetiva e segura é altamente desejável e é considerada como uma das melhores formas de controle da doença (MODABBER, 1995).

### **1.8.1 Imunidade na Leishmaniose**

A resposta imune associada ao controle das leishmanioses e à eliminação do parasito está principalmente associada às células T CD4+ e T CD8+, produtoras de interferon gama (IFN- $\gamma$ ) e ativação de macrófagos (DA SILVA SANTOS; BRODSKYN, 2014; OGHUMU et al., 2010). Ela tem início quando células dendríticas especializadas em apresentação de antígenos ativam as células T *naïve* através da apresentação de antígenos por proteínas do MHC e da co-estimulação realizada por meio da secreção de citocinas (BENNETT et al., 2001; XIN; LI; SOONG, 2007).

A ativação de células T CD8+ é resultante da interação destas com proteínas do MHC de classe I ligadas a peptídeos do patógeno. Já a ativação de células T CD4+ resulta da interação entre elas e proteínas do MHC de classe II ligadas a peptídeos do patógeno (E SILVA et al., 2016).

### 1.8.2 Vacinas existentes para Leishmaniose

Inúmeros candidatos vacinais têm sido testados tanto para humanos quanto para cães. Em 1940, Pessoa e Pestana desenvolveram uma vacina constituída por 18 cepas dermatrópicas de *Leishmania*, oriundas de seis regiões do estado de São Paulo e realizaram o primeiro ensaio clínico vacinal realizado em humanos para leishmaniose, obtendo-se uma eficácia protetora de 80%. Apesar dos resultados promissores, essa preparação foi abandonada por apresentar problemas relacionados à sua padronização (PESSOA e PESTANA, 1940; PESSOA, 1941a; 1941b). Em 1996, Mayrink e colaboradores avaliaram, em cães, antígenos de *L. (Viannia) braziliensis* associados ao adjuvante BCG (*Bacillus Calmette-Guérin*). Apesar dessa preparação ter se mostrado eficaz em estudos de fases I e II, não se demonstrou eficaz em ensaios de fase III.

No ano de 2011 foi licenciada na Europa a CaniLeish<sup>®</sup> (LiESP/QA-21), uma vacina para cães composta de proteínas excretadas e secretadas de *Leishmania infantum* purificadas e adjuvante QA-21, que é uma fração altamente purificada da saponina. Os estudos com essa vacina indicaram que após a imunização, os cães desenvolveram forte imunidade mediada por células do tipo Th1, resposta indicada como protetora no caso de leishmanioses e resposta humoral baseada na produção de IgG2 (MORENO et al., 2012).

Estudos com duas vacinas recentemente desenvolvidas a LBSap (composta por antígenos de *L. braziliensis* associado ao adjuvante saponina) e a LBSapSal (mesma composição da vacina LBSap acrescida de extrato de glândula salivar de flebotomíneos) indicaram alto potencial imunogênico protetor em cães (GIUNCHETTI et al., 2007, 2008).

No Brasil, atualmente, se encontra disponível para a comercialização em clínicas veterinárias particulares, a LeishTec<sup>®</sup>. A formulação dessa vacina foi analisada em um estudo de fase III duplo cego randomizado e apresentou 71% de eficácia ao serem avaliados os resultados parasitológicos isoladamente. Sobrepondo o xenodiagnóstico e o exame parasitológico, a eficácia de proteção foi de 58% (REGINA-SILVA et al., 2016).

A vacina Leishmune<sup>®</sup> desenvolvida sob supervisão da Dra. Clarissa Palatinik, composta pelo antígeno fucose-manose ligante (FML) de *L. donovani* em

combinação com o adjuvante saponina (Leishmune<sup>®</sup>, *Fort Dodge Saúde Animal Ltda*), demonstrou resultados promissores em ensaios vacinais de fase III que apontaram de 92% a 95% de proteção vacinal contra a leishmaniose visceral canina no grupo imunizado (76% de eficácia vacinal) (DA SILVA et al., 2000; BORJA-CABRERA et al., 2002). Apesar dos resultados promissores, recentemente, o registro da vacina foi suspenso pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), através de uma nota técnica n° 038/2014/DFIP/DAS, por não atender os requisitos dos estudos de fase III segundo a instrução normativa 31/2007 (MINISTÉRIO DA AGRICULTURA, 2014).

## 2 JUSTIFICATIVA

O desenvolvimento de vacinas é, reconhecidamente, um processo longo, complexo e de alto custo. Para tanto, é necessário o conhecimento da doença, de dados epidemiológicos, genômicos e de antígenos capazes de prevenir ou tratar a doença.

Dentro das abordagens mais promissoras podemos mencionar a vacinologia reversa. Essa abordagem representa um grande avanço no desenvolvimento de novas vacinas, uma vez que vem contribuindo com a geração de modelos preditivos eficazes para a identificação de novos antígenos vacinais e gerando, como desdobramento, a redução dos custos e do tempo de desenvolvimento da pesquisa, uma vez que direciona assertivamente a experimentação.

Um dos aspectos que ainda merece atenção, e que está vinculado ao desenvolvimento metodológico da estratégia, diz respeito ao fato da maioria das ferramentas preditivas não considerar as informações estruturais dos epítomos. Essas informações são fundamentais em diferentes aspectos das interações imunológicas, como, por exemplo, no que se refere a epítomos conformacionais de células B. A informação estrutural é também essencial para o processamento e o atracamento de peptídeos com as moléculas de MHC de classes I e II, que estão envolvidas no direcionamento da resposta imune. Além disso, alguns estudos apontam que a falta de estrutura (regiões desordenadas da proteína) pode estar relacionada com epítomos imunoestimuladores.

No Brasil, no ano de 2015 o número de casos de leishmaniose visceral era de 3.289 e de leishmaniose tegumentar era de 19.395, de acordo com o Ministério da Saúde. Não existem vacinas para uso humano comercialmente disponíveis e as estratégias atuais de controle estão centradas na gestão do reservatório canino, aplicação de inseticidas, diagnóstico e tratamento. Essas medidas, muitas vezes realizadas de forma isolada, não se demonstram muito efetivas.

Considerando esse cenário e ainda que os fármacos preconizados no tratamento das leishmanioses são altamente tóxicos, o presente trabalho visa contribuir para o desenvolvimento de uma vacina através da avaliação estrutural de proteínas que foram previamente selecionadas através de vacinologia reversa. Para isso o



trabalho foi dividido em duas partes: a primeira consistiu na análise de epítomos de células B conformacionais e possível correlação com desordem estrutural; e uma segunda parte que consistiu na análise de epítomos de linfócitos T, através de atracamento molecular.

### 3 OBJETIVO GERAL

Avaliar aspectos estruturais utilizando candidatos vacinais selecionados através de vacinologia reversa contra *Leishmania braziliensis* e *L. infantum*.

#### 3.1 Objetivos específicos - Parte 1

Construir modelos estruturais, através de modelagem computacional de proteínas de *L. braziliensis* e *L. infantum* previamente selecionadas como boas candidatas vacinais pela vacinologia reversa e, em seguida, avaliar os modelos construídos utilizando programas computacionais específicos

Predizer epítomos de células B conformacionais usando as estruturas construídas e validadas das proteínas de *L. braziliensis* e *L. infantum*

Predizer regiões de desordem estrutural com base na sequência de resíduos das proteínas de *L. braziliensis* e *L. infantum* selecionadas pela vacinologia reversa como boas candidatas vacinais e verificar possível correlação com imunogenicidade.

#### 3.2 Objetivos específicos - Parte 2

Analisar a afinidade de ligação de peptídeos oriundos das proteínas de *L. infantum* selecionadas como boas candidatas vacinais pela vacinologia reversa com MHC I e MHC II de humano através de atracamento molecular.

## 4 MÉTODOS

### 4.1 Parte 1- Predição estrutural das proteínas, predição de epítomos de células B conformacionais e correlação com desordem

#### 4.1.1 Origem dos dados utilizados na primeira parte deste trabalho

Brito e colaboradores, em 2014, utilizaram a vacinologia reversa para seleção de alvos vacinais para *L. infantum* e *L. braziliensis*. O método usado inicialmente envolveu o *download* dos genomas das espécies de *Leishmania* do TritypDB, versão 3.2 para *L. infantum* e 2.0 para *L. braziliensis*. Em seguida, os proteomas preditos, a partir desses genomas, foram usados nas análises. As predições foram feitas utilizando-se os seguintes programas: para MHC-I, NetCTL (LARSEN et al., 2007; NIELSEN et al., 2005; PETERS et al., 2003) e NetMHC (BUUS et al., 2003; NIELSEN et al., 2003, 2004); para MHC-II, NetMHCII (NIELSEN; LUND, 2009; NIELSEN; LUNDEGAARD; LUND, 2007); para células B, BepiPred (LARSEN; LUND; NIELSEN, 2006), AAP12 e BCPred12 (CHEN et al., 2007; EL-MANZALAWY; DOBBS; HONAVAR, 2008a, 2008b). Foram analisados 12 alelos MHC-I humanos e sete alelos MHC-I de camundongos, e no contexto de MHC-II, foram analisados 14 alelos humanos e três de camundongos. Para a determinação da localização subcelular das proteínas, foram utilizados os programas: WoLF PSORT (HORTON et al., 2007) para a localização de proteínas em vários compartimentos subcelulares; TargetP (EMANUELSSON et al., 2000) para proteínas de secreção e mitocondriais; e Sigcleave (VON HEIJNE, 1986) para peptídeos sinais para a excreção.

Um banco de dados relacional usando MySQL foi idealizado e construído para integrar os resultados das predições e definir os possíveis alvos vacinais. Seis proteínas foram selecionadas para *L. braziliensis* e seis para *L. infantum*. Como informações complementares nos quadros 1 e 2, referente às 12 proteínas, na terceira coluna está representada as predições realizadas para essas proteínas acerca da família proteica a qual fazem parte, processo biológico, função molecular e componente celular com os quais estão envolvidas.

Quadro 1: Características das proteínas analisadas de *L. braziliensis*.

| Identificador  | Descrições Preditas | Predições realizadas através do Interproscan*   |
|----------------|---------------------|---|
| LbrM11_V2.0090 | Processo Biológico  | <a href="#">GO:0006364</a> processamento de rRNA<br><a href="#">GO:0042254</a> biogênese de ribossomos  |
|                | Função Molecular    | <a href="#">GO:0005515</a> ligação de proteínas   |
|                | Componente celular  | Nenhum predito  |
|                | Família Proteica    | <a href="#">WD repeat BOP1/Erb1 (IPR028598)</a>   |
| LbrM28_V2.1900 | Processo Biológico  | <a href="#">GO:0042176</a> regulação de processo catabólico de proteínas  |
|                | Função Molecular    | <a href="#">GO:0005488</a> ligação<br><a href="#">GO:0030234</a> atividade reguladora de enzimas  |
|                | Componente celular  | <a href="#">GO:0000502</a> complexo do proteassoma  |
|                | Família Proteica    | <a href="#">26S Proteasome non-ATPase regulatory subunit 1 (IPR035266)</a>  |
| LbrM28_V2.3070 | Processo Biológico  | <a href="#">GO:0005975</a> processo metabólico de carboidratos<br><a href="#">GO:0006108</a> processo metabólico de malato<br><a href="#">GO:0019752</a> processo metabólico de ácido carboxílico<br><a href="#">GO:0055114</a> processo de óxido-redução                                     |
|                | Função Molecular    | <a href="#">GO:0003824</a> atividade catalítica<br><a href="#">GO:0016491</a> atividade de óxido-redutase<br><a href="#">GO:0016615</a> atividade malato desidrogenase<br><a href="#">GO:0016616</a> atividade de óxido-redutase, atividade no grupo CH-OH de doadores, NAD ou NADP aceptores |
|                | Componente celular  | Nenhum predito  |
|                | Família Proteica    | <a href="#">Malate dehydrogenase, type 2 (IPR010945)</a>  |
| LbrM13_V2.1480 | Processo Biológico  | <a href="#">GO:0055114</a> processo de óxido-redução  |

|                |                    |   |
|----------------|--------------------|---|
|                | Função Molecular   | <a href="#">GO:0004506</a> atividade de esqualeno mono-oxygenase<br><a href="#">GO:0050660</a> ligação de flavina adenina dinucleotídeo<br><a href="#">GO:0071949</a> Ligação de FAD  |
|                | Componente celular | <a href="#">GO:0016021</a> componente integral de membrana  |
|                | Família Proteica   | Nenhuma predita   |
| LbrM22_V2.0110 | Processo Biológico | <a href="#">GO:0006164</a> processo bio-sintético de nucleotídeo de purina<br><a href="#">GO:0006177</a> Processo de bio-sintético de GMP   |
|                | Função Molecular   | <a href="#">GO:0003922</a> Atividade de GMP sintase (hidrólise de glutamina)<br><a href="#">GO:0005524</a> Ligação de ATP<br><a href="#">GO:0016462</a> atividade de pirofosfatase  |
|                | Componente celular | Nenhum predito  |
|                | Família Proteica   | <a href="#">Queuosine biosynthesis protein QueC</a> (IPR018317)   |
| LbrM05_V2.0730 | Processo Biológico | <a href="#">GO:0008033</a> processamento de tRNA  |
|                | Função Molecular   | <a href="#">GO:0003824</a> atividade catalítica<br><a href="#">GO:0010181</a> Ligação de FMN<br><a href="#">GO:0051536</a> ligação de cluster de ferro-enxofre<br><a href="#">GO:0051539</a> 4 iron, 4 sulfur cluster binding |
|                | Componente celular | Nenhum predito  |
|                | Família Proteica   | <a href="#">S-adenosyl-L-methionine-dependent tRNA 4-demethylwyosine synthase</a> (IPR034556)   |

Fonte: Elaborado pelo autor.

\*Predições realizadas através do servidor Web <https://www.ebi.ac.uk/interpro/>

Quadro 2: Características das proteínas analisadas de *L. infantum*.

| Identificador | Descrições Preditas | Predições realizadas através do Interproscan*                                  |
|---------------|---------------------|--|
| LinJ.32.0970  | Processo Biológico  | <a href="#">GO:0015991</a> Transporte de prótons acoplado com hidrólise de ATP |

|              |                    |   |
|--------------|--------------------|---|
|              | Função Molecular   | <a href="#">GO:0015078</a> Atividade de transporte transmembrana de íon hidrogênio  |
|              | Componente celular | <a href="#">GO:0000220</a> Transportador de próton vacuolar tipo V ATPase, domínio V0<br><a href="#">GO:0033179</a> Transportador de próton tipo V ATPase, domínio V0   |
|              | Família Proteica   | <a href="#">ATPase, V0 complex, subunit 116kDa, eukaryotic</a> (IPR026028)  |
| LinJ.19.0350 | Processo Biológico | <a href="#">GO:0009107</a> Processo biosintético de lipoato   |
|              | Função Molecular   | <a href="#">GO:0003824</a> atividade catalítica<br><a href="#">GO:0016992</a> Atividade de lipoato sintase<br><a href="#">GO:0051536</a> ligação de cluster de ferro-enxofre<br><a href="#">GO:0051539</a> 4 iron, 4 sulfur cluster binding |
|              | Componente celular | Nenhum predito  |
|              | Família Proteica   | <a href="#">Lipoyl synthase</a> (IPR003698)   |
| LinJ.18.1500 | Processo Biológico | <a href="#">GO:0120029</a> transporte de próton através da membrana plasmática  |
|              | Função Molecular   | <a href="#">GO:0000166</a> ligação de nucleotídeo<br><a href="#">GO:0008553</a> atividade de ATPase transportadora de prótons   |
|              | Componente celular | Componente integral de membrana   |
|              | Família Proteica   | <a href="#">P-type ATPase, subfamily IIIA</a> (IPR006534)   |
| LinJ.36.5700 | Processo Biológico | <a href="#">GO:0055114</a> processo de óxido-redução  |
|              | Função Molecular   | <a href="#">GO:0016491</a> atividade de óxido-redutase  |
|              | Componente celular | Nenhum predito  |
|              | Família Proteica   | Nenhum predito  |
| LinJ.36.2160 | Processo Biológico | Nenhum predito  |

|              |                    |  |
|--------------|--------------------|--|
|              | Função Molecular   | <a href="#">GO:0016758</a> atividade de transferase, transferência de grupos hexosyl                             |
|              | Componente celular | <a href="#">GO:0005783</a> retículo endoplasmático<br><a href="#">GO:0016021</a> componente integral de membrana |
|              | Família Proteica   | <a href="#">Glycosyltransferase, ALG3</a> (IPR007873)  |
| LinJ.28.1850 | Processo Biológico | <a href="#">GO:0042176</a> regulação de processo catabólico de proteínas   |
|              | Função Molecular   | <a href="#">GO:0005488</a> ligação<br><a href="#">GO:0030234</a> atividade reguladora de enzimas                 |
|              | Componente celular | <a href="#">GO:0000502</a> complexo do proteassoma   |
|              | Família Proteica   | <a href="#">26S Proteasome non-ATPase regulatory subunit 1</a> (IPR035266)                                       |

Fonte: Elaborado pelo autor.

\*Predições realizadas através do servidor Web <https://www.ebi.ac.uk/interpro/>

As 12 proteínas selecionadas foram utilizadas nesse trabalho. Nesse sentido a sequências destas proteínas foram obtidas através do banco de dados *TritypDB* (<http://tritypdb.org/tritypdb/>) respeitando as versões utilizadas na etapa de predição realizada por Brito e colaboradores, versão 3.2 para *L. infantum* e 2.0 para *L. braziliensis*.

#### 4.1.2 Predição estrutural das proteínas

##### 4.1.2.1 Seleção das proteínas a serem modeladas

Para a seleção das proteínas passíveis de serem modeladas, primeiramente foi feita uma busca para levantamento dos possíveis moldes disponíveis. Para isso, um banco de dados local foi criado usando as sequências de aminoácidos das proteínas com estruturas depositadas no PDB ([ftp://ftp.wwpdb.org/pub/pdb/derived\\_data/pdb\\_seqres.txt.gz](ftp://ftp.wwpdb.org/pub/pdb/derived_data/pdb_seqres.txt.gz)), (data de acesso 03/12/2017) composto de 42.820 estruturas. As sequências de aminoácidos foram formatadas, usando o programa *readseq* (GILBERT, 2003) e então indexadas, no banco utilizando o programa *formatdb* (ALTSCHUL et al., 1990). Após a criação do banco de dados, foram realizadas buscas por similaridade de sequências localmente, usando o programa Protein-Protein BLAST 2.2.31+ (ALTSCHUL et al., 1990), entre as sequências de aminoácidos das proteínas a serem modeladas e as

de todas as proteínas com estrutura depositada no PDB e contidas no banco de dados local.

Foram modeladas somente as proteínas que tiveram *hits* com no mínimo 35% (WEBB; SALI, 2016) de identidade e 70% de cobertura. Tal critério de corte foi estabelecido como o mínimo necessário para se obter um modelo estrutural confiável para as proteínas analisadas. Para cada uma dessas proteínas foi selecionado um molde, dando prioridade para aquelas que apresentaram maior identidade e cobertura na busca de similaridade citada anteriormente, e estruturas depositadas com melhor resolução em Å (angstroms). Os procedimentos realizados nesta etapa estão descritos no apêndice A.

#### **4.1.2.2 Construção dos modelos estruturais**

A construção dos modelos estruturais das proteínas foi realizada empregando-se métodos de modelagem baseados em molde utilizando *software* local e servidores Web. A construção dos modelos foi realizada localmente utilizando o *software* Modeller (WEBB; SALI, 2016) e nos servidores Web I-TASSER (ROY; KUCUKURAL; ZHANG, 2010; YANG et al., 2015; ZHANG, 2008) e RaptorX (KÄLLBERG et al., 2012). Em seguida, o melhor modelo para cada proteína foi escolhido com base nos melhores resultados obtidos na etapa de avaliação.

##### **4.1.2.2.1 Modelagem comparativa usando Modeller**

O programa Modeller, versão 9.19, foi utilizado para a modelagem comparativa realizada localmente. Esse programa de modelagem comparativa emprega o método de satisfação de restrição espacial. O modelo estrutural é construído a partir do alinhamento entre a sequência de aminoácidos da proteína a ser modelada e uma proteína homóloga (que compartilha ancestralidade com a proteína a ser modelada) com estrutura conhecida (molde). Escolhemos a opção do programa que utiliza apenas um molde para cada proteína de interesse. Os moldes foram selecionados com base na busca inicial realizada no item 4.1.2.1. Os critérios de seleção estiveram centrados na maior porcentagem de cobertura, maior porcentagem de identidade e melhor resolução (quanto menor o valor em angstroms, melhor a resolução).



Para a utilização do Modeller inicialmente foi necessário converter manualmente o arquivo do tipo fasta, contendo as sequências de aminoácidos, para o arquivo do tipo PIR com extensão “.ali”, reconhecido pelo Modeller.

Em seguida um alinhamento par a par entre o molde e a proteína alvo foi realizado utilizando o arquivo “align2d.py” disponível no tutorial oficial do Modeller. De posse do alinhamento, quatro modelos estruturais da proteína alvo foram construídos utilizando o arquivo “model-single.py” também disponível no tutorial oficial do Modeller (<https://salilab.org/modeller/tutorial/>). Os procedimentos realizados nesta etapa estão descritos no apêndice B.

#### **4.1.2.2 Modelagens usando servidores Web**

As sequências das proteínas modeladas também foram submetidas a dois servidores, o I-TASSER e o RaptorX.

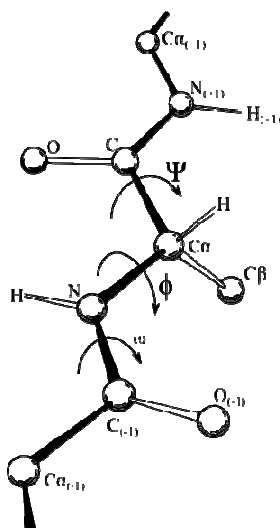
O servidor I-TASSER (Iterative Threading Assembly Refinement) combina os métodos *threading*, *ab initio* e refinamento estrutural para construção de modelos. A predição é constituída basicamente por três etapas: 1) identificação de moldes com conformações similares para a sequência de resíduos através de uma abordagem chamada LOMETS; 2) agrupamento dos fragmentos contínuos retirados dos moldes identificados no primeiro passo em modelos completos através de uma abordagem de troca baseada em simulações de Monte Carlo; as regiões da proteína alvo que não se alinharem a nenhum molde são construídas utilizando modelagem *ab initio*; 3) e por último é realizado o refinamento através de remontagem de conformações por fragmentos utilizando as estruturas modelo da etapa anterior e a seleção de configurações de menor energia.

O servidor de predição de estrutura RaptorX pode ser aplicado à predição de sequências de proteína que não possuem proteínas homólogas próximas com estruturas depositadas no PDB. Inicialmente é realizado o alinhamento entre a sequência submetida e os possíveis moldes usando um algoritmo de alinhamento de molde único. Em seguida, é predita a qualidade do alinhamento e feito o ranqueamento de moldes seguindo ordem decrescente de qualidade de alinhamento. Se for possível, a modelagem é feita utilizando-se múltiplos moldes através da abordagem *threading*. Caso contrário, um modelo estrutural é construído através do alinhamento par a par com um molde apenas (PENG; XU, 2011).

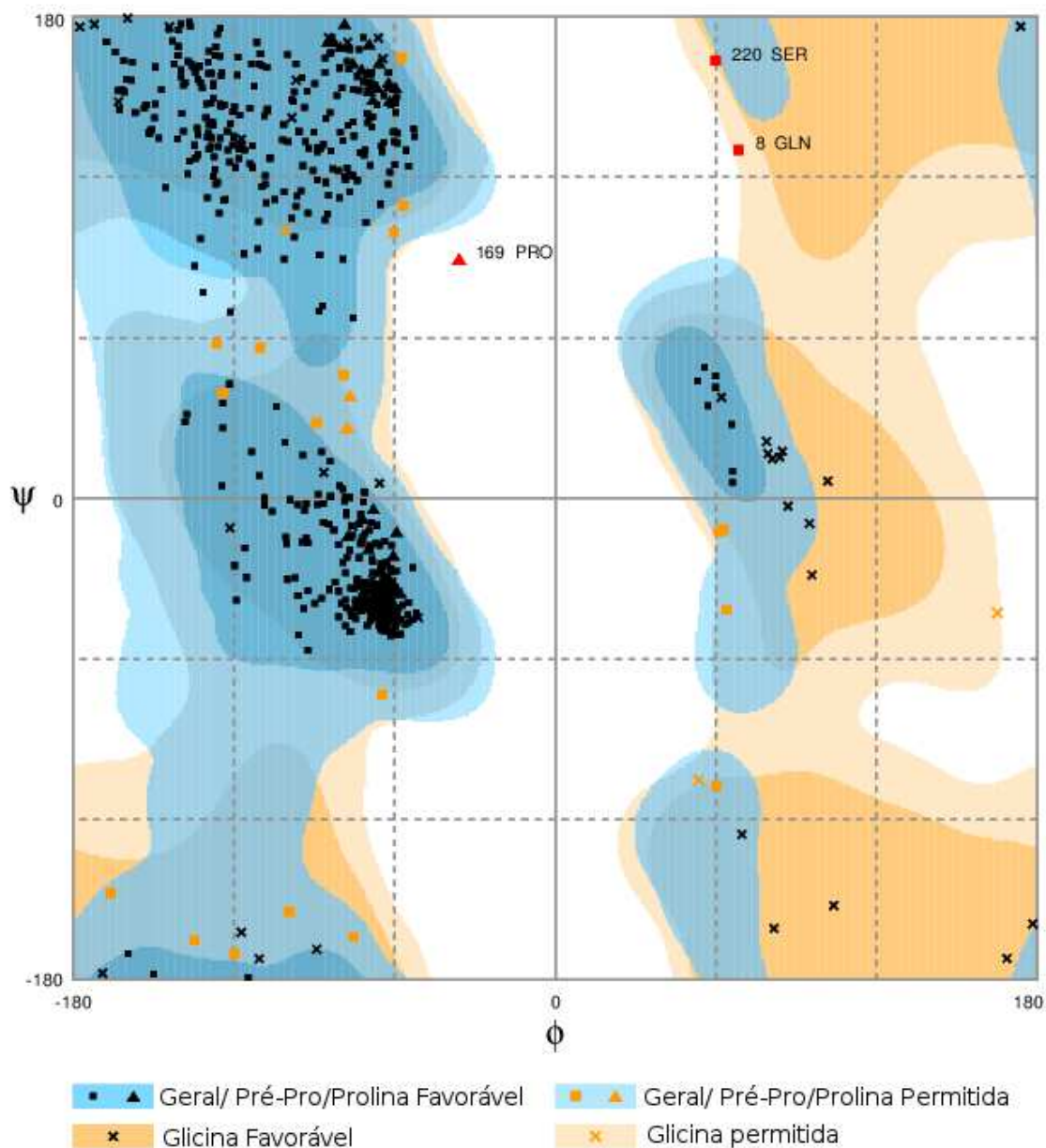
#### 4.1.2.3 Avaliação das estruturas previstas

Os modelos foram avaliados de três formas: a) por meio de gráficos de Ramachandran (RAMACHANDRAN; SASISEKHARAN, 1968) gerados pelo servidor Rampage (LOVELL et al., 2003) (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>); b) através de gráficos gerados a partir do escore DOPE (SHEN; SALI, 2006) e c) RMSD.

Nos gráficos de Ramachandran é possível visualizar todas as combinações de ângulos diedrais  $\Psi$  (psi),  $\Phi$  (phi) ilustrados na Figura 4. Esses ângulos determinam a conformação da cadeia principal da proteína. O ângulo  $\Phi$  define a rotação em torno da ligação carbono  $\alpha$  - nitrogênio do resíduo, e o  $\Psi$  define a rotação em torno da ligação carbono  $\alpha$  - carbono da carbonila. Em gráficos de Ramachandran de estruturas proteicas favoráveis, os pontos, os quais representam cada um dos resíduos da proteína, tendem a se acumular em determinadas regiões favoráveis e não ocorrem em regiões desfavoráveis devido ao impedimento estérico das cadeias laterais dos aminoácidos. Na Figura 5 temos um exemplo desse tipo de gráfico. Nele, destacam-se três valores em porcentagem, sendo o primeiro referente ao número de resíduos em regiões favorecidas, o segundo referente ao número de resíduos em regiões permitidas e o terceiro referente ao número de resíduos em regiões não permitidas.



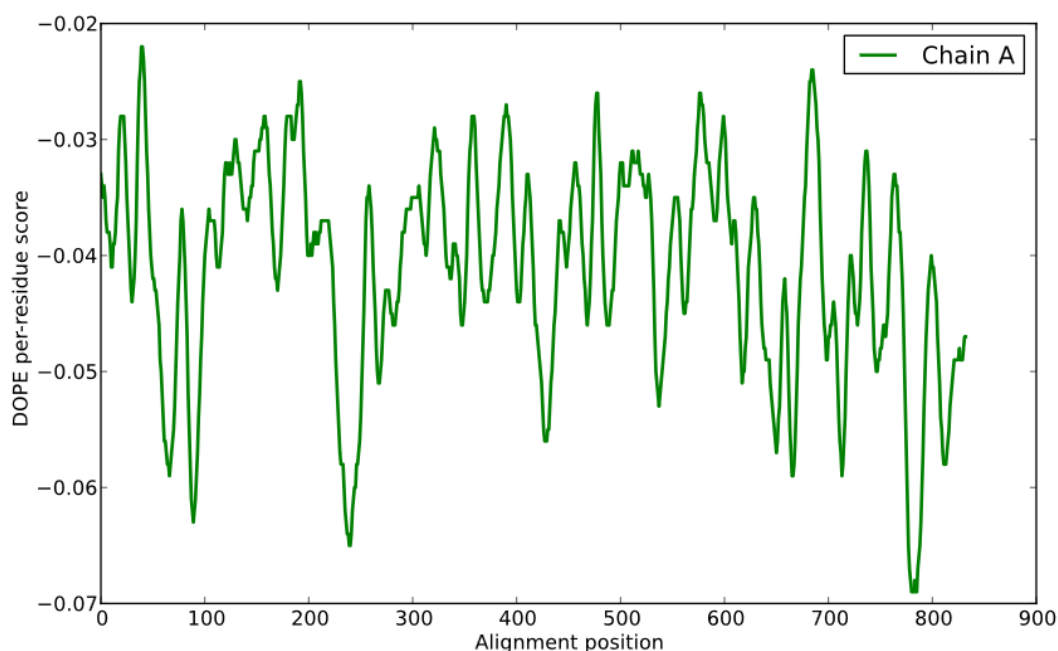
**Figura 4:** Ângulos diedrais de uma ligação peptídica. Os ângulos diedrais chamados  $\Psi$  (psi) e  $\Phi$  (phi) definem a rotação em torno da ligação Carbono  $\alpha$  – Carbono da carbonila e Carbono  $\alpha$  - Nitrogênio do mesmo resíduo, respectivamente (Fonte: Dcrjsr, vetorizado por Adam Rędzikowski, wikipedia).



Número de resíduos em região favorável (~98.0% expected) : 628 (96.0%)  
 Número de resíduos em região permitida (~2.0% expected) : 23 (3.5%)  
 Número de resíduos em região não permitida : 3 (0.5%)

**Figura 5: Gráfico de Ramachandran do modelo da proteína LbrM22\_V2.0110.** No eixo x temos representados os ângulos  $\phi$ , no eixo y os ângulos  $\psi$ . As áreas em azul representam combinação de ângulos favoráveis, as regiões em amarelo representam combinação de ângulos permitidos, as regiões em branco representam combinação de ângulos desfavoráveis. Cada uma das formas geométricas representam um resíduo da proteína. . Fonte: Elaborado pelo autor usando o servidor Rampage.

Os modelos que apresentaram melhores gráficos de Ramachandran foram avaliados em relação ao valor de Energia Proteica Otimizada Discreta (DOPE, do inglês, Discrete Optimized Protein Energy), um parâmetro quantitativo que indica o potencial energético de modelos gerados. Esse potencial estatístico foi desenvolvido por Shen e Sali em 2006 e baseia-se em 1472 estruturas cristalográficas. Para geração dos escores e plotagem dos gráficos foi utilizado o programa Modeller (WEBB; SALI, 2016).



**Figura 6:** Gráfico com os escores DOPE da cadeia A da estrutura 2HJL (GILLESPIE et al., 2006). No eixo x temos a posição dos resíduos. No eixo y estão representados os escores DOPE, quanto menor o escore, menor a energia potencial da região indicada e maior a estabilidade energética.

Na Figura 6: Gráfico com os escores DOPE da cadeia A da estrutura 2HJL (GILLESPIE et al., 2006). temos um exemplo de um gráfico plotado com base nos escores DOPE da estrutura da proteína 2HJL (GILLESPIE et al., 2006).

Adicionalmente, utilizando o Pymol, os modelos que apresentaram melhores gráficos de Ramachandran também foram avaliados através do RMSD dos mesmos relativos a seus respectivos moldes. RMSD do inglês *Root Mean Square deviation*, é o desvio da raiz do valor quadrático médio da distância entre átomos de duas estruturas. Um valor de RMSD inferior a 2,0 Å obtido através da sobreposição entre

modelo obtido através de modelagem baseada em molde e seu respectivo molde é indicativo de boa qualidade (VERLI, 2014).

O delineamento experimental da etapa de avaliação dos modelos está ilustrado na Figura 7.

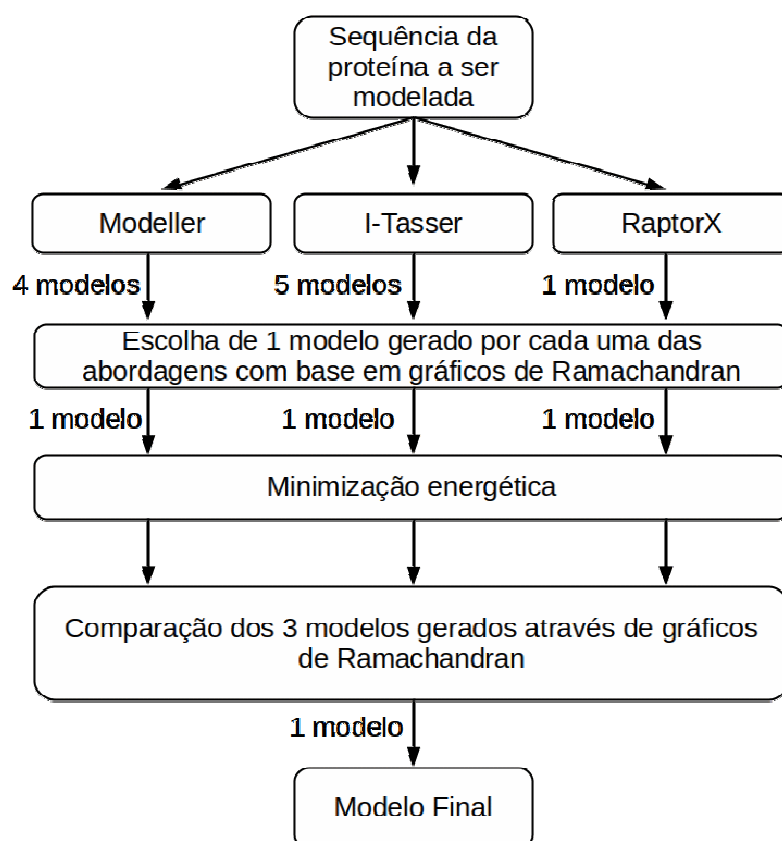


Figura 7: Esquema do delineamento experimental da etapa de Avaliação dos Modelos.

#### 4.1.2.4 Minimização energética

No intuito de refinar os modelos obtidos com as modelagens baseadas em molde foi utilizado o servidor de minimização de energia YASARA (<http://www.yasara.org/minimizationserver.htm>) (KRIEGER et al., 2009), o qual consiste em utilizar um modelo matemático para reduzir a energia potencial do modelo construído. O modelo matemático desse servidor combina o campo de força AMBER (CORNELL et al., 1996), com potenciais de torção multidimensionais baseados em conhecimento (KEEDY et al., 2009) e um consistente conjunto de parâmetros de campos que maximizam a acurácia do modelo.

O melhor modelo obtido por cada uma das abordagens foi selecionado com base nos melhores valores do gráfico de Ramachandran gerado pelo servidor Rampage e submetido ao refinamento estrutural. Após o refinamento, o servidor retorna um arquivo com as coordenadas dos resíduos proteicos refinadas e as estruturas proteicas resultantes foram submetidas à avaliação pelos programas citados no item 4.1.2.3. Após essa avaliação usando o Programa Pymol (SCHRÖDINGER, LLC, 2015), os modelos foram alinhados estruturalmente a seus respectivos moldes e as regiões de incongruências foram desconsideradas nas análises seguintes.

### **4.1.3 Predição de epítomos conformacionais de células B**

A predição dos epítomos conformacionais de células B nas proteínas de *Leishmania* foi feita com base nas melhores estruturas obtidas na etapa de modelagem. Os epítomos preditos serão representados pelos peptídeos de *Leishmania* no decorrer deste trabalho. Os programas utilizados nesta etapa foram: EPCES (LIANG et al., 2007), Bepro (antes denominado PEPITO) (SWEREDOSKI; BALDI, 2008), DiscoTope (KRINGELUM et al., 2012) e Ellipro (PONOMARENKO et al., 2008). O DiscoTope e o Ellipro estão disponíveis para execução local, o Bepro e o EPCES são servidores de Web.

#### **4.1.3.1 EPCES**

O EPCES (<http://sysbio.unl.edu/EPCES>) é um servidor Web que realiza predição de epítomos antigênicos na superfície proteica fazendo um consenso de pontuações. Este preditor utiliza uma função de energia empírica, desenvolvida para atracamento molecular entre proteínas, adicionada de um estado de referência simples capaz de remover dependências irreais de afinidade de ligação em conformações de atracamento em regiões da interface acessíveis a solvente, porém encobertas. A pontuação mínima considerada para este preditor neste trabalho foi de 70,00.

#### **4.1.3.2 BEpro**

BEpro (<http://pepito.proteomics.ics.uci.edu/>) é um preditor de epítomos conformacionais de células B que utiliza uma combinação de pontuações de propensão de aminoácidos (pontuação utilizada para predição de estruturas secundárias a partir de cálculos da probabilidade do resíduo se encontrar na

interface pela probabilidade de estar na superfície) e valores de meia esfera de exposição (medida de exposição da proteína ao solvente, mensura o quão um aminoácido está enterrado na estrutura) em múltiplas distâncias para as predições de epítomos. A pontuação mínima considerada para este preditor neste trabalho foi de 1,30.

#### **4.1.3.3 DiscoTope**

O programa DiscoTope 1.1 (<http://www.cbs.dtu.dk/services/DiscoTope>) também utiliza uma combinação de cálculos de acessibilidade de superfície (estimada em termos de números de contatos) e estatísticas de propensão de aminoácidos baseada em epítomos já validados. A pontuação mínima considerada para este preditor neste trabalho foi de -7,7. O procedimento utilizado encontra-se descrito no apêndice C.

#### **4.1.3.4 Ellipro**

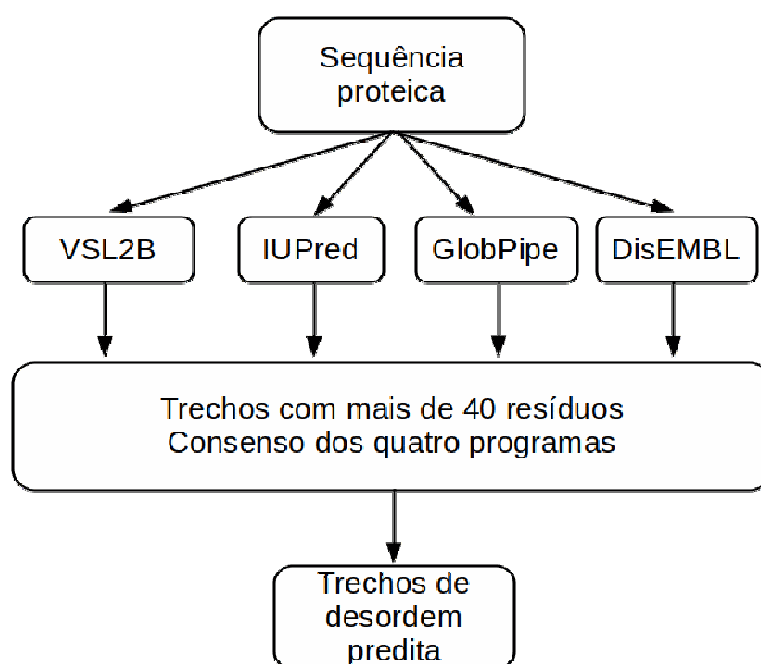
O programa Ellipro (<http://tools.iedb.org/ellipro>) implementa uma versão modificada do método de Thornton (originalmente uma abordagem para reconhecimento de conformação proteica) e um algoritmo de clusterização. O método desenvolvido por Thornton se baseia em localizar peptídeos antigênicos a partir de coordenadas de carbonos-alfa de uma proteína, baseando-se em protruções de áreas globulares da proteína (THORNTON et al., 1986). A pontuação final da predição é definida como *Protusion Index* - índice de protusão, baseado na aproximação da estrutura 3D por um número de elipsoides, ou seja, na massa do resíduo que fica fora do maior elipsoide possível. Dessa forma resíduos com altas pontuações estão associados a uma alta acessibilidade ao solvente. Após a definição dos epítomos, estes são agrupados com base na distância entre os centros de massa dos resíduos, sendo que altos valores de clusterização se associam a maiores epítomos conformacionais. A pontuação mínima considerada para este preditor neste trabalho foi de 0,5. O procedimento utilizado encontra-se descrito no Apêndice C

#### **4.1.4 Predição de regiões desordenadas**

As sequências de aminoácidos das proteínas estudadas foram utilizadas como arquivo de entrada para o *pipeline* desenvolvido por Patrícia Ruy e colaboradores

publicado em 2014 e validado para predição de desordem estrutural em proteínas provenientes de tripanossomatídeos.

O *pipeline* emprega a utilização de quatro programas de predição e integra informações de anotação funcional e estrutural, predição de localização subcelular e propriedades físico-químicas (Figura 8). No centro do *pipeline* há um banco de dados relacional que descreve o conhecimento de desordem proteica de modo consistente (RUY et al., 2014).



**Figura 8: Esquema do *pipeline* de predição de desordem proteica.** A sequência de resíduos da proteína a ser analisada é submetida a quatro preditores de desordem. Os trechos preditos com mais de 40 resíduos são usados para a realização do consenso dos quatro programas.

Os quatro preditores utilizados foram: DisEMBL (LINDING et al., 2003), GlobPipe (LINDING, 2003), IUPred (DOSZTÁNYI et al., 2005), e VSL2B (PENG et al., 2006). As etapas realizadas estão descritas no apêndice D.





## 4.2 Parte 2 - Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular

### 4.2.1 Origem dos dados utilizados na segunda parte deste trabalho

Brito e colaboradores, em 2014 (BRITO; RESENDE; REIS, 2014), utilizaram a vacinologia reversa para seleção de alvos vacinais para *L. infantum*, conforme descrito no item 4.1.1 dos resultados da parte 1 deste trabalho. Seis proteínas foram selecionadas e os epítomos preditos dessas proteínas como imunogênicos, através da vacinologia reversa foram utilizados nesta segunda parte do trabalho (Quadro 3), que consistiu de análises de atracamento molecular na avaliação das interações entre esses epítomos e moléculas de MHC de classes I e II.

Quadro 3: Peptídeos preditos como sendo epítomos lineares pela vacinologia reversa (Brito et al. 2014) e que foram utilizados nas análises de atracamento molecular.

| Identificador da Proteína | Sequência do epítomo | Resíduo inicial do epítomo | Classe de MHC* |
|---------------------------|----------------------|----------------------------|----------------|
| LinJ.18.1500              | LVVSIPIAL            | 303                        | MHC I          |
|                           | YLLARFYET            | 283                        |                |
|                           | KPQRRQSVL            | 29                         |                |
|                           | QMVYNQDEI            | 453                        |                |
|                           | RGYTCAMTG            | 596                        |                |
|                           | GGHFFFYVPPSPILF      | 810                        | MHC II         |
|                           | FFFYVPPSPILFCGA      | 813                        |                |
| LinJ.32.0970              | YLRLWALSL            | 798                        | MHC I          |
|                           | FMLLGMPYI            | 652                        |                |
|                           | SGYTWAPIS            | 489                        |                |
|                           | EYVLSSVSNTASYLR      | 786                        | MHC II         |
|                           | TSFVVLCSPTMIVR       | 235                        |                |
|                           | QKGTTYPTTPNGLPS      | 499                        |                |
|                           | KGTTYPTTPNGLPSV      | 500                        |                |
| LinJ.36.2160              | LAYDHRAYL            | 215                        | MHC I          |

|              |                 |     |        |
|--------------|-----------------|-----|--------|
|              | PARDNEPKL       | 423 | MHC II |
|              | TLPFWRVAW       | 190 |        |
|              | TVATWVLLWRTRWAA | 262 |        |
|              | IRQGFESFPPTPKTS | 377 |        |
|              | GFESFPPTPKTSMM  | 37  |        |
| LinJ.36.5700 | RLCPRGHSL       | 119 | MHC I  |
|              | QSGHNSGCL       | 97  |        |
|              | IGPGSALAL       | 353 |        |
|              | PRVTVPATMDKKMLL | 454 | MHC II |
|              | VATAREIRQKYPRKR | 69  |        |
| LinJ.28.1850 | FALKRLSSL       | 25  | MHC I  |
|              | YWYWYPYLL       | 791 |        |
|              | SVIHNATVV       | 379 |        |
|              | RGGSRMVAF       | 247 |        |
|              | FYVANVLLQDITFRR | 189 | MHC II |
|              | GAPVAAAVPESPEQK | 324 |        |
| LinJ.19.0350 | YLQPSRTRL       | 332 | MHC I  |
|              | KMLTKSSIM       | 294 |        |
|              | SAVTLGQYL       | 325 |        |
|              | RPPPLDPEE       | 173 |        |
|              | EMGVDIYVMTMVDRD | 191 | MHC I  |
|              | LVEQLAVTPLSVYAH | 244 |        |
|              | LKLVEQLAVTPLSVY | 242 |        |

Fonte: Elaborado pelo autor.

\*Classe de MHC predita na etapa de vacinologia reversa como ligante ao peptídeo

#### 4.2.2 Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular

Os passos seguidos nesta etapa foram baseados no trabalho de Silva e colaboradores publicado em 2016 (SILVA et al., 2016) e se baseiam em atracamento molecular para a busca de potenciais peptídeos imunogênicos capazes de se ligarem a diferentes moléculas de MHC de classe I e II de diferentes populações humanas. O delineamento experimental desta etapa está representado na Figura 9.

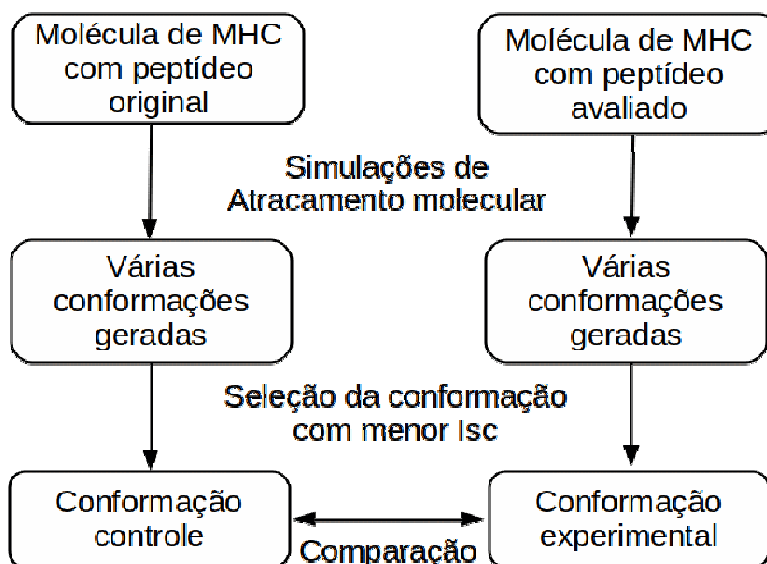


Figura 9: Esquema do delineamento experimental das simulações de atracamento molecular. Isc: Interface score.

Inicialmente foi realizado o download das estruturas dos alelos de MHC I (20 estruturas do PDB, identificadores: 2HJL (GILLESPIE et al., 2006), 3C9N (RODER et al., 2008), 3HCV (BELTRAMI et al., 2009), 3KPP (MACDONALD et al., 2009), 3L3D (THEODOSSIS; ELY; ROSSJOHN, 2010), 3RL1 (ZHANG et al., 2012), 3VCL (PETERSEN; ROSSJOHN, 2012), 3VFS (LIU; ROSSJOHN; GRAS, 2012), 4F7M (LIU et al., 2012), 4G8G (GRAS; WILMANN; ROSSJOHN, 2013), 4HWZ (NIU et al., 2013), 4JQX (THEODOSSIS et al., 2013), 4MJ5 (LIU et al., 2014), 4MJJ (RIZKALLAH et al., 2014), 4NQV (ROSSJOHN; GRAS, 2013), 4O2C (SUN et al., 2014), 4QRR (GRAS; CHABROL; ROSSJOHN, 2014), 4QRU (GRAS; TWIST; ROSSJOHN, 2015), 4WU5 (SHIMIZU et al., 2015) e 4XXC (HIBBERT; ROSSJOHN;

GRAS, 2015)) e de MHC II (12 estruturas do PDB, identificadores: 1A6A (GHOSH et al., 1998), 1BX2 (SMITH et al., 1998), 1H15 (LANG et al., 2002), 1S9V (KIM et al., 2004), 1UVQ (SIEBOLD et al., 2004), 1YMM (HAHN et al., 2005), 2NNA (HENDERSON et al., 2007), 2Q6W (PARRY, 2007), 3C5J (DAI; KAPPLER, 2008), 3LQZ (DAI, 2010), 3PL6 (SETHI; WUCHERPFENNIG, 2010) e 3WEX (KUSANO et al., 2014)).

As estruturas foram preparadas removendo-se as moléculas de água, as cadeias não participantes das interações entre o peptídeo e o MHC e os carboidratos. Após as etapas descritas acima, as estruturas cocrystalizadas com seus respectivos peptídeos, (aqui denominado peptídeo cocrystalizado), foram submetidas ao processo de atracamento molecular e utilizadas como controle para a etapa descrita a seguir.

Usando o programa Pymol (SCHRÖDINGER, LLC, 2015) a cadeia de peptídeos cocrystalizada em cada estrutura foi modificada para conter o mesmo número de resíduos, ou seja, nove resíduos para os peptídeos com afinidade para MHC I e 15 para MHC II. Resíduos de alanina foram adicionados quando necessário para alongar a cadeia peptídica, pois o peptídeo cocrystalizado era menor que o número de resíduos estipulado inicialmente (9 resíduos para moléculas de MHC de classe I e 15 resíduos para moléculas de MHC de classe II). Para isso foi utilizada a opção “Builder” do Pymol. Utilizou-se também o programa UCSF-Chimera (PETTERSEN et al., 2004) para renumerar os resíduos das cadeias e renomear as cadeias quando necessário. Em seguida a cadeia de peptídeos editada foi trocada pelos peptídeos das proteínas em estudo, equivalentes aos epítomos preditos. Para esse fim foi desenvolvido um *script* em linguagem Python que utiliza a função *mutagenesis* do programa Pymol (SCHRÖDINGER, LLC, 2015), descrito no apêndice E.

Os complexos obtidos foram então submetidos a uma etapa de pré-empacotamento utilizando o *script* *flexpep\_prepack* do pacote Rosetta (LEAVER-FAY et al., 2011; RAVEH; LONDON; SCHUELER-FURMAN, 2010). Após essa etapa, foram realizadas as simulações de atracamento utilizando o *script* *FlexPepDocking*, também do pacote Rosetta. Os comandos estão descritos no Apêndice F.

Para quantificar a afinidade entre os receptores de MHC e os peptídeos em estudo foi utilizada uma pontuação do protocolo *FlexPepDock* do pacote Rosetta,

denominada Isc (Interface score), que consiste na soma das contribuições energéticas da interface de resíduos entre os receptores e os epítomos em estudo. O Isc foi a pontuação escolhida por Silva e colaboradores em 2016 citado anteriormente. De acordo com esse trabalho um menor Isc se correlaciona com maior afinidade de ligação. De acordo com Holland e colaboradores (2013) uma maior afinidade entre peptídeos e MHC estaria correlacionada, principalmente, com maior número de ligações de hidrogênio estabelecidas entre esses dois componentes (HOLLAND; COLE; GODKIN, 2013).

## 5 RESULTADOS

### 5.1 Parte 1 - Predição de epítomos de células B conformacionais e correlação com desordem

#### 5.1.1 Predição estrutural das proteínas

##### 5.1.1.1 Seleção de proteínas a serem modeladas

O primeiro passo para as modelagens foi a utilização das sequências das proteínas de interesse em buscas por proteínas similares com estruturas depositadas em bancos de dados públicos. Os resultados desta análise para as 12 proteínas estudadas são sumarizados nas Tabela 1 e Tabela 2.

Seguindo os critérios antes estabelecidos (identidade acima de 35% e com mais de 70% de cobertura) somente foram selecionadas para modelagem as proteínas identificadas pelos códigos LbrM28\_V2.1900, LbrM28\_V2.3070 e LbrM22\_V2.0110 de *L. braziliensis* e LinJ.19.0350, LinJ.18.1500 e LinJ.28.1850 de *L. infantum*.

Um maior detalhamento desta etapa pode ser encontrado nos apêndices G e H.

Tabela 1: Moldes proteicos identificados para *L. braziliensis*

| <b>LbrM11_V2.0090</b>    |                         |                        |
|--------------------------|-------------------------|------------------------|
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 4u7a cadeia A            | 32,44                   | 93                     |
| 5em2 cadeia A            | 30,92                   | 55                     |
| 5cxc cadeia B            | 30,27                   | 53                     |
| 5cxb cadeia B            | 30,27                   | 53                     |
| <b>LbrM28_V2.1900***</b> |                         |                        |
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 5vhs cadeia U            | 36,45                   | 98                     |
| 5vhi cadeia U            | 36,45                   | 98                     |
| 5vhh cadeia U            | 36,45                   | 98                     |
| 5vgz cadeia U            | 36,45                   | 98                     |
| 5t0j cadeia U            | 36,45                   | 98                     |
| <b>LbrM28_V2.3070***</b> |                         |                        |
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |

|                          |                         |                        |
|--------------------------|-------------------------|------------------------|
| 4i1i cadeia B            | 85,49                   | 100                    |
| 4h7p cadeia A            | 85,49                   | 100                    |
| 2cvq cadeia B            | 50,47                   | 99                     |
| <b>LbrM13_V2.1480</b>    |                         |                        |
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 5dbj cadeia A            | 38,67                   | 13                     |
| <b>LbrM22_V2.0110***</b> |                         |                        |
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 2vx0 cadeia A            | 47,79                   | 98                     |
| 2ywc cadeia B            | 36,79                   | 96                     |
| <b>LbrM05_V2.0730</b>    |                         |                        |
| <b>Molde*</b>            | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 2yx0 cadeia A            | 34,67                   | 35                     |
| 2z2u cadeia A            | 30,00                   | 38                     |

Fonte: Elaborado pelo autor

\* A descrição corresponde ao identificador da estrutura proteica do PDB e sua respectiva cadeia

\*\*Comparação entre o possível molde e as proteínas de *L. braziliensis* a serem modeladas

\*\*\*Proteínas que foram selecionadas para modelagem segundo o critério de possuírem moldes com identidade maior ou igual a 35% e cobertura acima de 70%

Tabela 2: Moldes proteicos identificados para *L. infantum*.

|                        |                         |                        |
|------------------------|-------------------------|------------------------|
| <b>LinJ.32.0970</b>    |                         |                        |
| <b>Molde*</b>          | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 5voz cadeia b          | 30,50                   | 98                     |
| 5voy cadeia b          | 30,50                   | 98                     |
| 5vox cadeia b          | 30,50                   | 98                     |
| 3j9v cadeia b          | 30,50                   | 98                     |
| 3j9u cadeia b          | 30,50                   | 98                     |
| <b>LinJ.19.0350***</b> |                         |                        |
| <b>Molde*</b>          | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 4u0p cadeia B          | 41,75                   | 71                     |
| 4u0o cadeia B          | 41,75                   | 71                     |
| 5exk cadeia K          | 43,14                   | 72                     |
| <b>LinJ.18.1500***</b> |                         |                        |
| <b>Molde*</b>          | <b>Identidade (%)**</b> | <b>Cobertura (%)**</b> |
| 5ksd cadeia A          | 36,74                   | 85                     |
| 1mhs cadeia A          | 33,41                   | 84                     |
| 4xe5 cadeia A          | 26,88                   | 69                     |



| LinJ.36.5700    |                  |                 |
|-----------------|------------------|-----------------|
| Molde*          | Identidade (%)** | Cobertura (%)** |
| 3dme cadeia A   | 28,04            | 61              |
| 4x9n cadeia A   | 31,45            | 30              |
| 4x9m cadeia A   | 31,45            | 30              |
| LinJ.28.1850*** |                  |                 |
| Molde*          | Identidade (%)** | Cobertura (%)** |
| 5vhs cadeia U   | 38,01            | 86              |
| 5vhi cadeia U   | 38,01            | 86              |
| 5vhh cadeia U   | 38,01            | 86              |
| 5vgz cadeia U   | 38,01            | 86              |
| 5t0j cadeia U   | 38,01            | 86              |

Fonte: Elaborado pelo autor.

\* A descrição corresponde ao identificador da estrutura proteica do PDB e sua respectiva cadeia

\*\*Comparação entre o possível molde e as proteínas de *L. infantum* a serem modeladas

\*\*\*Proteínas que foram selecionadas para modelagem segundo o critério de possuírem moldes com identidade maior ou igual a 35% e cobertura acima de 70%

De acordo com a busca realizada utilizando o programa BLASTp nenhum potencial molde foi encontrado para a proteína LinJ.36.2160 de *L. infantum*.

### 5.1.1.2 Modelagem comparativa usando Modeller

No processo de modelagem por homologia utilizamos a opção do programa Modeller que utiliza apenas um molde para cada proteína de interesse. Foram selecionados os moldes com maior porcentagem de cobertura, maior porcentagem de identidade e melhor resolução da estrutura depositada no PDB (quanto menor o valor em Å, melhor a resolução) dentre os apresentados nas Tabela 1 e Tabela 2. Os moldes selecionados para cada uma das proteínas estão destacados em cinza na Tabela 3.

**Tabela 3: Características dos possíveis moldes para cada uma das proteínas modeladas.** Os moldes destacados em cinza foram os escolhidos para a modelagem usando o programa Modeller.

| LbrM28_V2.1900 |               |                     |                              |                |               |
|----------------|---------------|---------------------|------------------------------|----------------|---------------|
| Molde          | Resolução (Å) | Organismo           | Anotação Funcional (Classes) | Identidade (%) | Cobertura (%) |
| 5vhs_U         | 8,8           | <i>Homo sapiens</i> | Hidrolase                    | 36,45          | 98            |

|        |     |                     |           |       |    |
|--------|-----|---------------------|-----------|-------|----|
| 5vhi_U | 6,8 | <i>Homo sapiens</i> | Hidrolase | 36,45 | 98 |
| 5vhh_U | 6,1 | <i>Homo sapiens</i> | Hidrolase | 36,45 | 98 |
| 5vgz_U | 4,5 | <i>Homo sapiens</i> | Hidrolase | 36,45 | 98 |
| 5t0j_U | 8,0 | <i>Homo sapiens</i> | Hidrolase | 36,45 | 98 |

#### LbrM28\_V2.3070

| Molde  | Resolução | Organismo                   | Anotação Funcional (Classes) | Identidade (%) | Cobertura (%) |
|--------|-----------|-----------------------------|------------------------------|----------------|---------------|
| 4i1i_B | 1,5       | <i>Leishmania major</i>     | Óxido-redutase               | 85,49          | 100           |
| 4h7p_A | 1,3       | <i>Leishmania major</i>     | Óxido-redutase               | 85,49          | 100           |
| 2cvq_B | 2,08      | <i>Thermus thermophilus</i> | Óxido-redutase               | 50,47          | 99            |

#### LbrM22\_V2.0110

| Molde  | Resolução | Organismo                   | Anotação Funcional (Classes) | Identidade (%) | Cobertura (%) |
|--------|-----------|-----------------------------|------------------------------|----------------|---------------|
| 2vx0_A | 2,5       | <i>Homo sapiens</i>         | Ligase                       | 47,79          | 98            |
| 2ywc_B | 2,2       | <i>Thermus thermophilus</i> | Ligase                       | 36,79          | 96            |

#### LinJ.19.0350

| Molde  | Resolução | Organismo                            | Anotação Funcional (Classes) | Identidade (%) | Cobertura (%) |
|--------|-----------|--------------------------------------|------------------------------|----------------|---------------|
| 4u0p_B | 1,62      | <i>Thermosynechococcus elongatus</i> | Transferase                  | 41,75          | 71            |
| 4u0o_B | 1,62      | <i>Thermosynechococcus elongatus</i> | Transferase                  | 41,75          | 71            |
| 5exk_K | 1,86      | <i>Mycobacterium tuberculosis</i>    | Transferase                  | 43,14          | 72            |

#### LinJ.18.1500

| Molde  | Resolução | Organismo                   | Anotação Funcional (Classes) | Identidade (%) | Cobertura (%) |
|--------|-----------|-----------------------------|------------------------------|----------------|---------------|
| 5ksd_A | 3,5       | <i>Arabidopsis thaliana</i> | Transporte de proteína       | 36,74          | 85            |

#### LinJ.28.1850

| Molde | Resolução | Organismo | Anotação Funcional | Identidade (%) | Cobertura (%) |
|-------|-----------|-----------|--------------------|----------------|---------------|
|-------|-----------|-----------|--------------------|----------------|---------------|

| (Classes) |     |                     |           |       |    |
|-----------|-----|---------------------|-----------|-------|----|
| 5vhs_U    | 8,8 | <i>Homo sapiens</i> | Hidrolase | 38,01 | 86 |
| 5vhi_U    | 6,8 | <i>Homo sapiens</i> | Hidrolase | 38,01 | 86 |
| 5vhh_U    | 6,1 | <i>Homo sapiens</i> | Hidrolase | 38,01 | 86 |
| 5vgz_U    | 4,5 | <i>Homo sapiens</i> | Hidrolase | 38,01 | 86 |
| 5t0j_U    | 4,5 | <i>Homo sapiens</i> | Hidrolase | 38,01 | 86 |

Fonte: Elaborado pelo autor.

O melhor molde está relacionado à proteína LbrM28\_V2.3070 (85,49% de identidade e 100% de cobertura) e o pior à proteína LinJ.18.1500 (36,74% de identidade e 85% de cobertura).

Em seguida um alinhamento par a par entre o melhor molde e a proteína alvo foi realizado utilizando o arquivo “align2d.py” disponível no tutorial oficial do Modeller. Em seguida, de posse do alinhamento, quatro modelos estruturais para proteína alvo foram construídos.

As informações obtidas através do PDB acerca das características biológicas das proteínas moldes selecionadas apresentaram consoância com as informações obtidas através do servidor Interpro das proteínas a serem modeladas e apresentadas nos quadros 1 e 2, indicando funções biológicas relacionadas entre proteínas a serem modeladas e seus respectivos modelos.

### 5.1.1.3 Modelagem usando o servidor Web I-TASSER

O servidor Web I-TASSER seleciona moldes através da metodologia denominada LOMETS (LOcal MEta-Threading-Server), um meta-servidor que combina a seleção de moldes realizada por nove programas de “threading”: (PPA-I (WU; ZHANG, 2007), SP3 (ZHOU; ZHOU, 2005), PPA-II (WU; ZHANG, 2007), SPARKS (ZHOU; ZHOU, 2004), PROSPECT (XU; XU, 2000), FUGUE (SHI; BLUNDELL; MIZUGUCHI, 2001), HHSEARCH (SÖDING, 2005), PAINT (WU; ZHANG, 2007) e SAM-T02 (KARPLUS et al., 2003)). Os melhores moldes identificados para cada proteína, segundo o I-TASSER, estão sumarizados na Tabela 4.

**Tabela 4: Melhores moldes identificados para cada proteína de acordo com o servidor I-TASSER.**

| <b>Proteína a ser modelada</b> | <b>Molde selecionado</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
|--------------------------------|--------------------------|-----------------------|----------------------|
| LbrM28_V2.1900                 | 5l4k - cadeia N          | 35                    | 88                   |
| LbrM28_V2.3070                 | 4h7p - cadeia A          | 84                    | 97                   |
| LbrM22_V2.0110                 | 2vx0 - cadeia B          | 49                    | 96                   |
| LinJ.19.0350                   | 5exi - cadeia A          | 29                    | 67                   |
| LinJ.18.1500                   | 2zxe - cadeia A          | 26                    | 87                   |
| LinJ.28.1850                   | 5l4k - cadeia N          | 36                    | 88                   |

Fonte: Elaborado pelo autor.

O melhor molde identificado está relacionado à proteína LbrM28\_V2.3070 (84.0% de identidade e 97.0% de cobertura) e o pior à proteína LinJ.19.0350 (29.00% de identidade e 67.00% de cobertura).

#### 5.1.1.4 Modelagem usando o servidor Web RaptorX

O servidor Web RaptorX seleciona moldes a partir de uma busca inicial por moldes únicos. Em seguida é montada uma matriz de probabilidade com base nos resultados dessa primeira busca. Essa matriz é então utilizada para alinhamentos múltiplos entre a sequência a ser modelada e os possíveis moldes (mais de um, se possível), ao final os moldes são selecionados.

Os melhores moldes identificados para cada proteína, segundo o RaptorX, estão sumarizados na Tabela 5.

Tabela 5: Melhores moldes identificados para cada proteína de acordo com o servidor RaptorX.

| Proteína a ser modelada | Molde selecionado | Identidade (%) | Cobertura (%)* |
|-------------------------|-------------------|----------------|----------------|
| LbrM28_V2.1900          | 4ady - cadeia A   | 31             | 99             |
| LbrM28_V2.3070          | 4h7p - cadeia A   | 85             | 100            |
| LbrM22_V2.0110          | 2vxo - cadeia A   | 48             | 98             |
| LinJ.19.0350            | 5exi - cadeia A   | 43             | 77             |
| LinJ.18.1500            | 5ksd - cadeia A   | 37             | 86             |
| LinJ.28.1850            | 4ady - cadeia A   | 31             | 99             |

Fonte: Elaborado pelo autor.

\* A porcentagem relacionada a cobertura foi obtida através do programa BLASTp e não através do servidor Web RaptorX

O melhor molde identificado está relacionado à proteína LbrM28\_V2.3070 (85.0% de identidade e 100% de cobertura) e o pior à proteína LinJ.18.1500 (37.00% de identidade e 86.00% de cobertura).

#### 5.1.1.5 Avaliação e minimização energética das estruturas preditas

Dependendo da abordagem utilizada (Modeller, I-TASSER e Raptor-X), o par da proteína de interesse-molde pode gerar um ou mais modelos se faz necessária uma

análise e avaliação desses modelos. Dessa forma, esta avaliação foi feita através de gráficos de Ramachandran e teve como objetivo final o estabelecimento de um único modelo para cada uma das três abordagens, para cada proteína de interesse. Os modelos escolhidos foram submetidos à minimização de energia potencial e posteriormente a uma segunda etapa analítica envolvendo gráficos de Ramachandran. Os gráficos de Ramachandran gerados para cada uma das estruturas estão apresentados no apêndice I. Os resultados da avaliação estão apresentados na Tabela 6 e os melhores modelos, destacados em cinza, foram selecionados com base em menores porcentagens de resíduos em regiões energeticamente desfavoráveis (regiões não permitidas) e maiores porcentagens de resíduos em regiões favoráveis.

Em relação ao RMSD, com exceção do modelo da proteína LinJ.28.1850, todos apresentaram um valor abaixo de 2,0 Å, o que é indicativo de compatibilidade entre o modelo gerado e seu molde (VERLI, 2014). No caso do modelo da proteína LinJ.28.1850 o RMSD foi ligeiramente superior, e isso pode ter acontecido devido à menor porcentagem de identidade entre a sequência dessa proteína e a sequência do seu respectivo molde (31%).

**Tabela 6: Valores de pontuações dos modelos obtidos pelas três abordagens após a submissão dos mesmos à minimização energética.** Modelos selecionados para etapa posterior estão destacados em cinza.

| <b>Abordagem</b>      | <b>Regiões favorecidas (%)</b> | <b>Regiões permitidas (%)</b> | <b>Regiões não permitidas (%)</b> | <b>RMSD (Å)</b> |
|-----------------------|--------------------------------|-------------------------------|-----------------------------------|-----------------|
| <b>LbrM22_V2.0110</b> |                                |                               |                                   |                 |
| I-TASSER              | 93.7                           | 5.5                           | 0.8                               | –               |
| Modeller              | 92.5                           | 6.0                           | 1.5                               | –               |
| RaptorX               | 96.0                           | 3.5                           | 0.5                               | 0.480           |
| <b>LbrM28_V2.1900</b> |                                |                               |                                   |                 |
| I-TASSER              | 91.4                           | 7.1                           | 1.5                               | 1.871           |

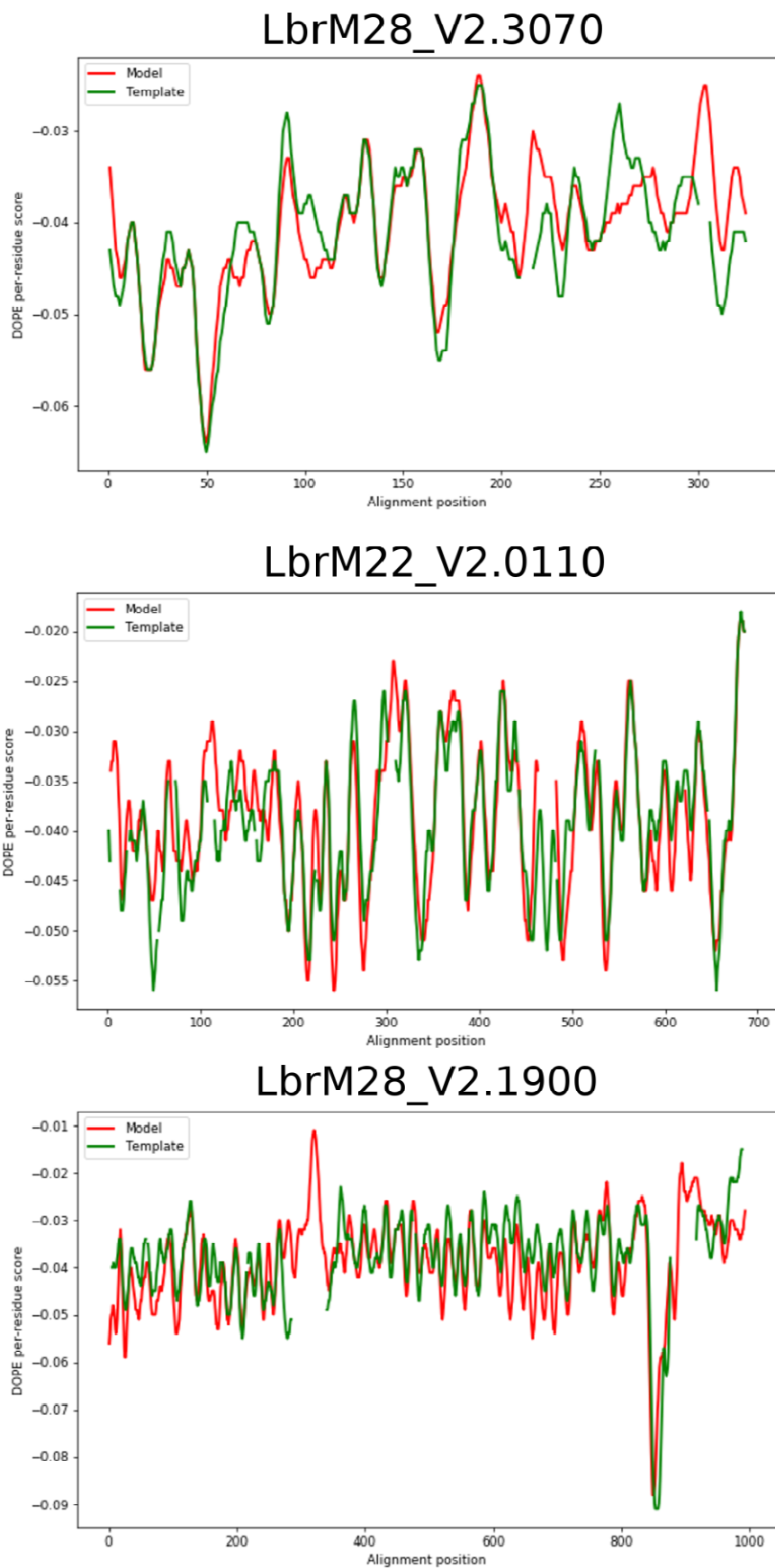
|                       |      |      |     |       |
|-----------------------|------|------|-----|-------|
| Modeller              | 93.3 | 5.0  | 1.6 | –     |
| RaptorX               | 92.6 | 5.4  | 1.9 | –     |
| <b>LbrM28_V2.3070</b> |      |      |     |       |
| I-TASSER              | 96.6 | 3.1  | 0.3 | –     |
| Modeller              | 98.8 | 0.9  | 0.3 | 0.359 |
| RaptorX               | 98.1 | 1.2  | 0.6 | –     |
| <b>Linj.19.0350</b>   |      |      |     |       |
| I-TASSER              | 87.1 | 10.7 | 2.2 | –     |
| Modeller              | 92.7 | 6.3  | 1.0 | –     |
| RaptorX               | 95.4 | 3.2  | 1.5 | 0.683 |
| <b>LinJ.18.1500</b>   |      |      |     |       |
| I-TASSER              | 89.6 | 7.5  | 2.9 | –     |
| Modeller              | 92.0 | 6.1  | 2.0 | –     |
| RaptorX               | 94.1 | 4.9  | 0.9 | 0.564 |
| <b>LinJ.28.1850</b>   |      |      |     |       |
| I-TASSER              | 90.3 | 7.6  | 2.1 | –     |
| Modeller              | 94.9 | 4.2  | 0.8 | –     |
| RaptorX               | 95.0 | 3.4  | 1.5 | 2.071 |

Fonte: Elaborado pelo autor.

Nas Figura 10 e Figura 11 estão representados os gráficos referentes aos escores DOPE para os seis modelos escolhidos. Os modelos escolhidos para as três proteínas de *L. braziliensis* e para as proteínas LinJ.18.1500 e LinJ.28.1850 de *L. infantum* apresentaram gráficos com os escores DOPE bem semelhantes a seus respectivos moldes. Essas semelhanças se destacam principalmente nas regiões

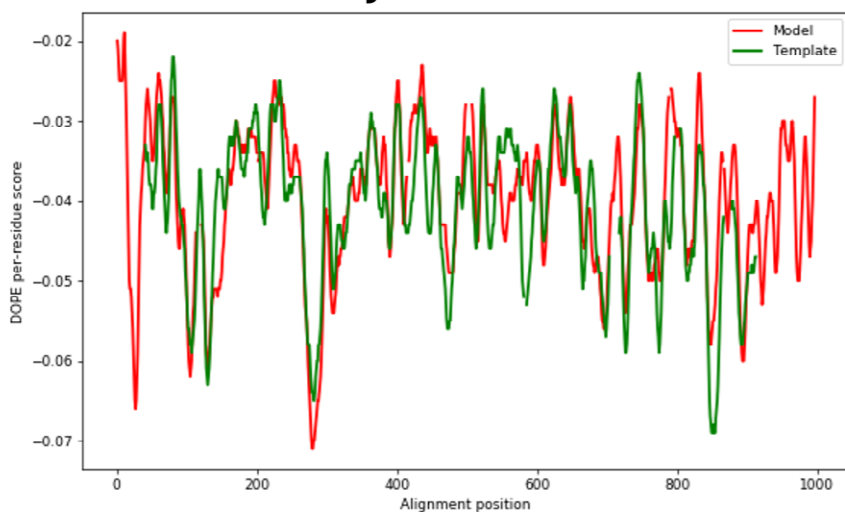
que apresentam baixo escore DOPE, indicativo de estabilidade energética, ou alto escore DOPE, indicativo de menor estabilidade energética, tanto no modelo quanto no molde utilizado. Isso pode ser observado na posição 50 do alinhamento entre o molde e o modelo da proteína LbrM28\_V2.3070 na qual destaca-se um escore DOPE inferior a -0,06 para ambos e próximo à posição 190 para essa mesma proteína na qual destaca-se um escore DOPE acima de -0,03 para o molde e o modelo. No caso do modelo feito para a proteína LinJ.19.0350 os gráficos com os escores DOPE do modelo e do molde se mostraram divergentes em alguns pontos. Essas divergências podem ser justificadas pelas menores cobertura e identidade apresentadas no alinhamento entre as sequências do molde e da proteína em questão.



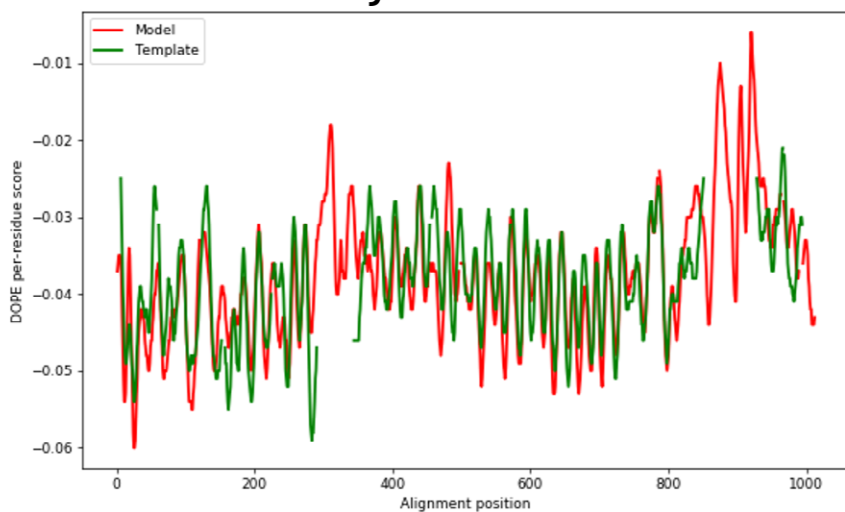


**Figura 10:** Gráficos DOPE para os três modelos selecionados para as proteínas de *Leishmania braziliensis*. Em verde estão representados os escores para os moldes utilizados e em vermelho estão representados os escores para os modelos gerados.

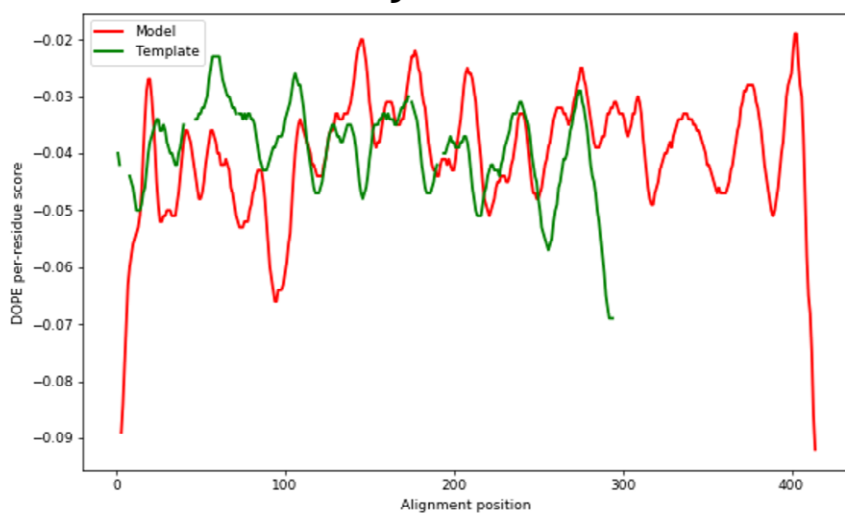
## LinJ.18.1500



## LinJ.28.1850



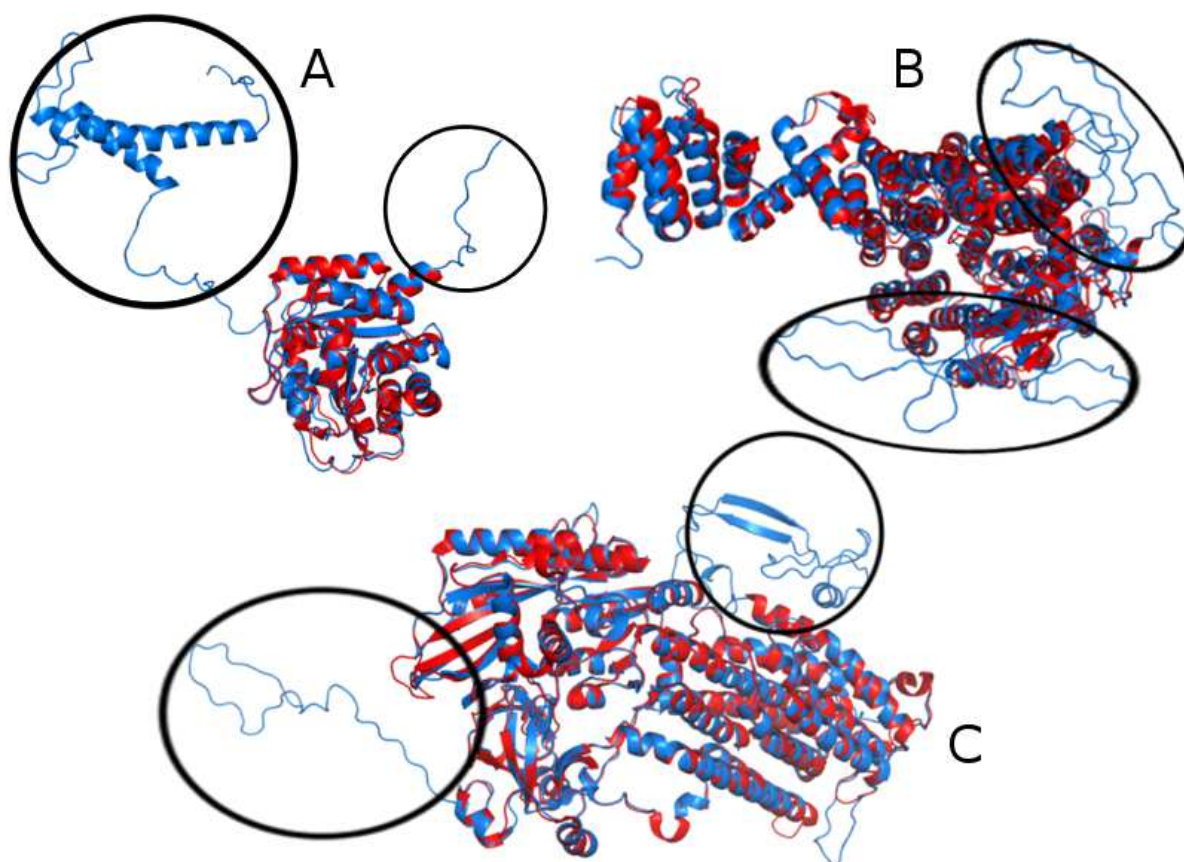
## LinJ.19.0350



**Figura 11: Gráficos DOPE para os três modelos selecionados para as proteínas de *Leishmania infantum*.** Em verde estão representados os escores para os moldes utilizados e em vermelho estão representados os escores para os modelos gerados.

### 5.1.2 Predição de epítomos conformacionais de células B

Os modelos selecionados e validados foram alinhados estruturalmente a seus respectivos moldes e inspecionados visualmente através do programa Pymol com o objetivo de identificar as sobreposições das estruturas alinhadas assim como incongruências. Em três modelos (Figura 12) foram observadas regiões que apresentaram grande divergência estrutural com o molde utilizado (destacadas por um círculo na figura em questão). Essas regiões foram removidas para evitar predições equivocadas. Os modelos foram então utilizados na predição de epítomos conformacionais de células B.



**Figura 12:** Alinhamento estrutural entre modelos (em azul) e moldes (em vermelho). A) corresponde à proteína LinJ.19.0350, B) à proteína LinJ.28.18.50, C) à proteína LinJ.18.1500

O

Quadro 4 sumariza as regiões, incluindo posições de resíduos, que não contribuíram para o alinhamento estrutural. Nos casos das proteínas Linj.19.0350 e Linj.18.1500 essas regiões foram especificamente aquelas não cobertas pelo molde.

Coincidentemente, essas duas proteínas foram aquelas que apresentaram menor cobertura, 77% e 85%, respectivamente.

**Quadro 4: Regiões desconsideradas nas predições devido à divergência estrutural com os moldes.**

| Proteína     | Regiões desconsideradas            |
|--------------|------------------------------------|
| LinJ.19.0350 | 1-109, 394-412                     |
| LinJ.18.1500 | 1-41, 890-974                      |
| LinJ.28.1850 | 283-331, 842-866, 874-919, 979-991 |

Fonte: Elaborado pelo autor.

Em relação à proteína LinJ.28.1850, na qual também identificamos visualmente um trecho que não contribuiu para o alinhamento estrutural (Figura 12C), é possível notar que, apesar da alta cobertura, a menor identidade gerou trechos de sequência que não se alinharam ao molde (Figura 13), gerando um modelo com divergência estrutural em relação ao molde. Os modelos das proteínas de *Leishmania braziliensis* não apresentaram regiões com grande divergência estrutural em relação ao molde e nenhuma região foi desconsiderada.

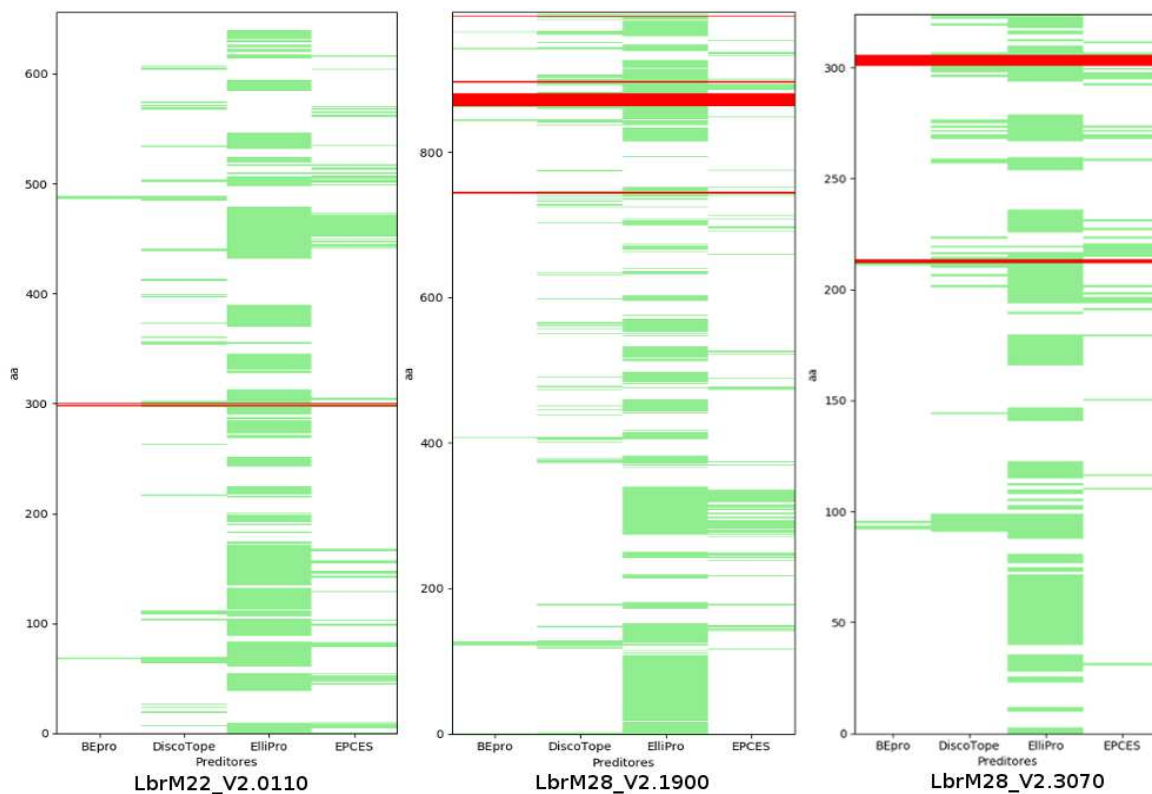
|              |  |     |
|--------------|--|-----|
| LinJ.28.1850 | MVQGLSSAKAVLALLSEEENLVVLFALKRLSSLMDTFWHEVSAKLPLIEELAASEKLADE | 60  |
| 4adyA        | -----TTAAPLLALLRENQDSVKTYALESINNVDQLWSEISNELPDIEALYDDDTFS--  | 60  |
| LinJ.28.1850 | TRRLASLVASQVYFHLGDYSNSVKHALAAGTAFDATTSLFTDTILSRCIDTYVAYQETP  | 120 |
| 4adyA        | DREMAALIASKVYYNLGEYESAVKYALAAKDRFDIDEKSQFVETIVSKSIEMYVQEASKQ | 120 |
| LinJ.28.1850 | -----ESERAELPPKLEELFVSLTKSWMENETMADLKEMVGFTVRAFRLDFLEKVLRQ   | 180 |
| 4adyA        | YTKDEQFYTKDIIDPKLTSIFERMIEKCLKAS----ELKLALGIALEGYRLDIIESALKS | 180 |
| LinJ.28.1850 | ALS--KTHSAEILNFTFYVANVLLQDITFRRKVLRLADLYTDGLSTIDYYSLAHCLLFL  | 240 |
| 4adyA        | KLDQSTSENVKIINYLLTLAITTVTNSKFRSSILRKSFDLMMN-PNCDYLTLNKVVVNL  | 240 |
| LinJ.28.1850 | GDVEATSNLICGLWRGGSRMVAFQLAFDLFEYGNQEYLSGVVAHLKQLGVAEQALAVQP  | 300 |
| 4adyA        | NDAGLALQLFKKLEENDEGLSAQIAFDLVSSASQQLLEILVTELTAAQGY-----      | 300 |
| LinJ.28.1850 | APSTSDAAGAAAAPSAPATSLSPPTTAGDASGAPVAAAVPESPEQKLLSVLSGQVTTNLN | 360 |
| 4adyA        | -----DPALLNILSGLPTCDYY                                       | 360 |
| LinJ.28.1850 | VKYLVSRCVADVYVLMHIKKMTDPNRSVIHNATVVANAFMYSGTMDGFLRDNMKWLGA   | 420 |
| 4adyA        | NTFLLNKNIDIGLLNKSKSSLDGKFSLFHTAVSVANGFMHAGTTDNSFIKANLPWLGA   | 420 |
| LinJ.28.1850 | QHWAKFTAVASTGAIHRGHTTEAMRVLEQYLPKGPSVPTLPYQEAGALYALGLIYSPLGA | 480 |
| 4adyA        | QNWAKFTATASLGVVHKGNLLEGKKVMAPYLPGS--RASSRFIKGGSYGLGLIYAGFGR  | 480 |
| LinJ.28.1850 | TRDRRTIQYLEENLQK--FSTNVQMVHGASLGIGLTAMGLQDEGLYDALFTCVTGMDAVA | 540 |

|              |  |      |
|--------------|--|------|
| 4adyA        | DTTDYLNKNIIVENSGTSGDEDDVDVLLHGASLGIGLAAMGSANIEVYEALKEVLYNDSATS | 540  |
| LinJ.28.1850 | AEGASVGVGMLMLGSGNDIVLQSLKNVAYEENQKEKQVIRGVCMAMALINLGREDEALLA   | 600  |
| 4adyA        | GEAAALGMGLCMLGTGKPEAIHDMFTYSQET-QHGNITRGLAVGLALINYGRQELADDLI   | 600  |
| LinJ.28.1850 | EELLESGBPWWRLGGCFVLGLAYAGTENAKTIEKLLSVTVKMSDDVRRTAVTMVGFLLTF   | 660  |
| 4adyA        | TKMLASDESLLRYGGAFITIALAYAGTGNNSAVKRLLHVAVSDSNDDVRRAAVIALGFVLL  | 660  |
| LinJ.28.1850 | KDPNLCCLDLIRVLVDSYSPHVRVYGVAMALAVSAAGTGNAAVIDVLWDMLEDIVDYVRQGA | 720  |
| 4adyA        | RDYTTVPRIVQLLSKSHNAHVRCGTAFALGIACAGKGLQSAIDVLDPLTKDPVDFVRQAA   | 720  |
| LinJ.28.1850 | AMGLAMVMVQLTEKENPRVKDFRLLLEKKIEDRNEGRCSKFGYVLASGLLDAGGRNCTFA   | 780  |
| 4adyA        | MIALSMILIQQTEKLNQVADINKNFLSVITNKHQEGFLAKFGACVAQGIMNAGGRNVTIQ   | 780  |
| LinJ.28.1850 | LHKQ--RHRLDKAVVGVFMFLQYWYWPYLLMITLAMQPTCIIGLNESE-ELPEYTFKSN    | 840  |
| 4adyA        | LENADTGTLDTKSVGLVMFVSQFWYWFPLAHFLSLSFTPTTVIGIRGSDQAIKPFQMNCY   | 840  |
| LinJ.28.1850 | APPSTYAVPKSVLQEKRVKTSEVQAVVLSSTRKEEELRQRRHQGPGSPGLAAGASATGAP   | 900  |
| 4adyA        | AKEDAFSYPRM-----   | 900  |
| LinJ.28.1850 | GGDGGAAEAGSKRGGDKDGGGESGEAAEQEPAFELHNPARTAHQFSVISHDQVDRYVP     | 960  |
| 4adyA        | -----KYSSKPKYVDNMTRILPQQSRYISFIKDDRFVP                         | 960  |
| LinJ.28.1850 | LKP---KPMGICLLKDTKPEMGAEVLVAP-----VVLSDRDEAPVPEPFSYP           | 1012 |
| 4adyA        | VRKFKG-NNGVVVLRDREPKEPVALIETVRQMKD-----                        | 1012 |

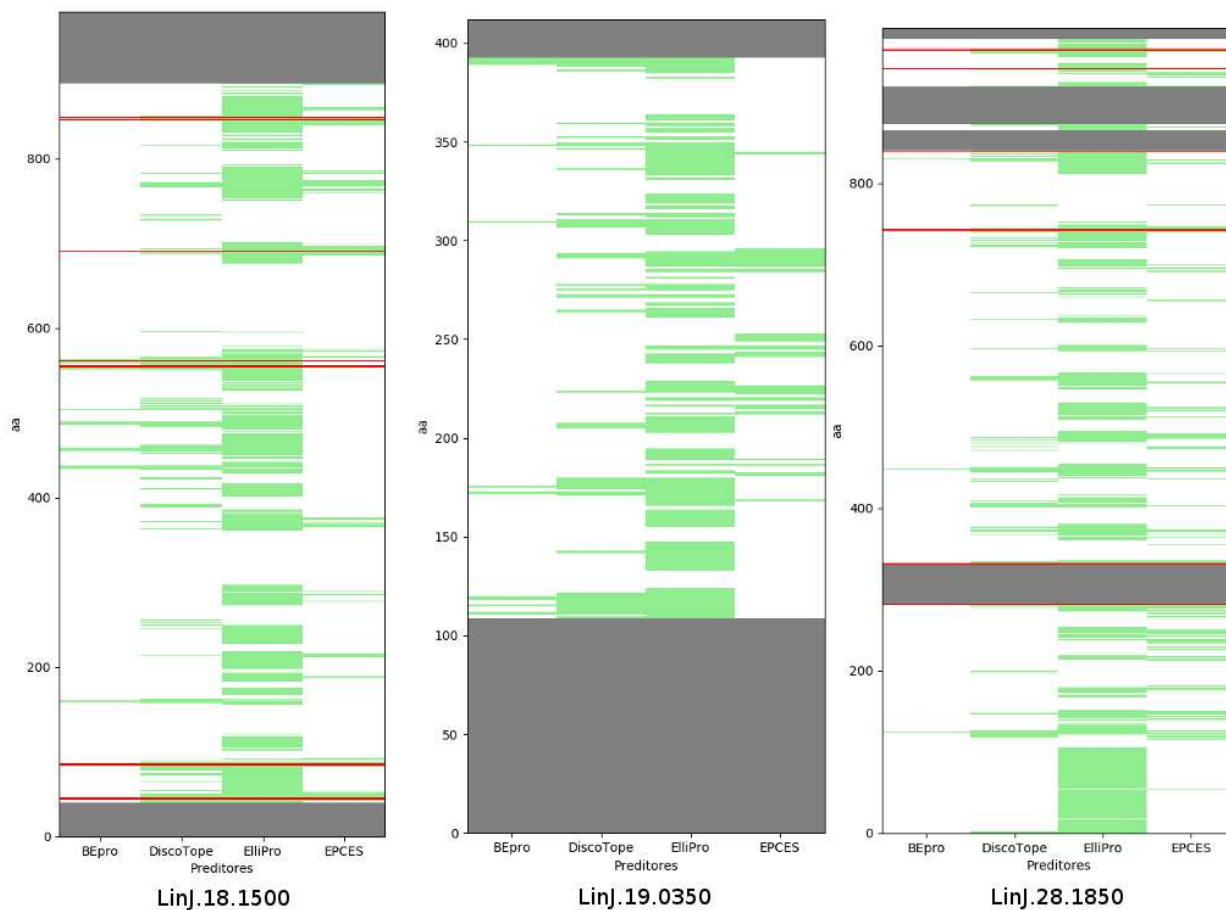
**Figura 13: Alinhamento entre a proteína LinJ.28.1850 e seu respectivo molde.** As regiões destacadas em cinza representam regiões que não se alinharam.

Após a curadoria e edição dos alinhamentos estruturais, as predições de epítopos conformacionais de células B foram feitas usando os preditores BÉpro, DiscoTope, Ellipro e EPCES usando como base os modelos estruturais proteicos construídos para as proteínas em estudo.

Os resultados dos preditores foram utilizados na elaboração de mapas de distribuição de resíduos de epítopos preditos (Figura 14 e Figura 15) que facilitaram a visualização de regiões consenso preditas como epítopos conformacionais de células B nas proteínas analisadas. Os resultados referentes às predições estão descritos no apêndice J

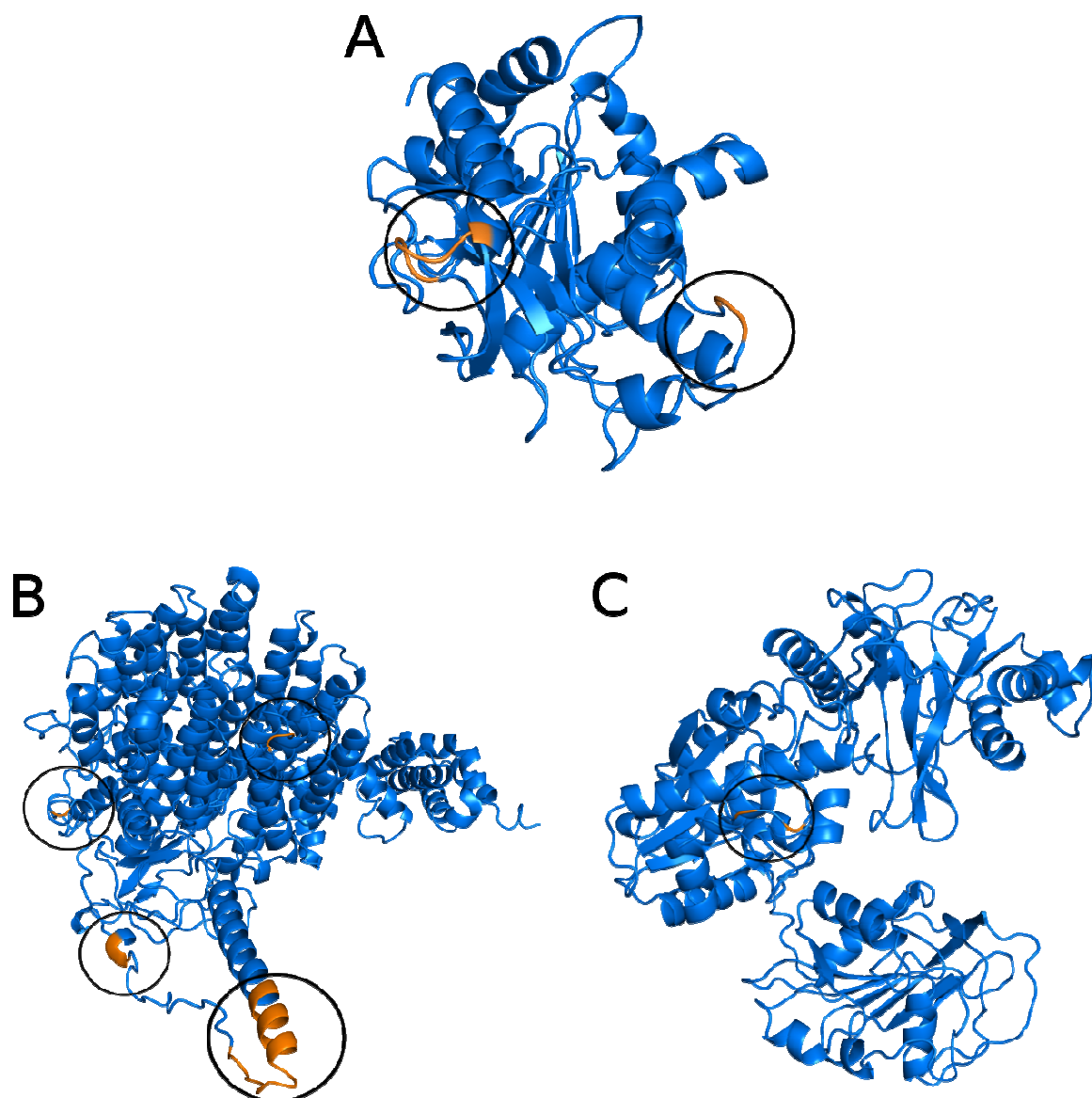


**Figura 14: Regiões preditas como imunogênicas nos modelos das proteínas *Leishmania braziliensis*.** O eixo y representa as posições de resíduos, o eixo x representa os preditores de epítopos conformacionais de células B. As regiões verdes representam as predições de imunogenicidade e as regiões vermelhas representam o consenso dos quatro preditores.



**Figura 15: Regiões preditas como imunogênicas nos modelos das proteínas de *L. infantum*.** O eixo y representa as posições de resíduos, o eixo x representa os preditores de epítomos conformacionais de células B. As regiões verdes representam as predições de imunogenicidade, as regiões vermelhas representam o consenso dos quatro preditores e as regiões em cinza representam as partes das proteínas que foram desconsideradas nas predições.

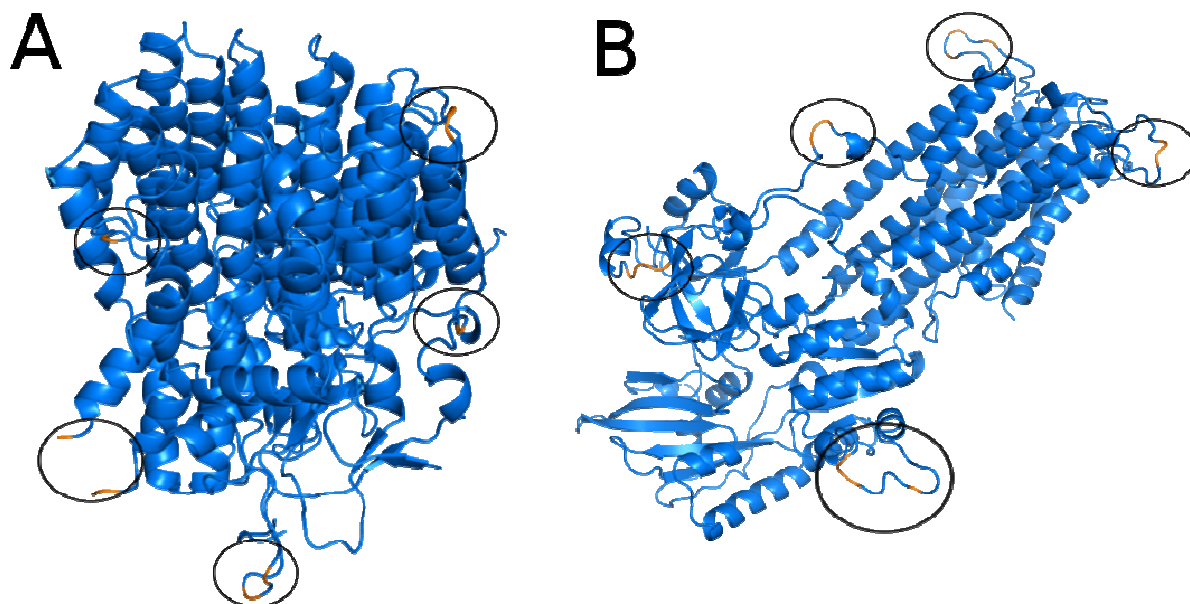
Nas Figura 16 e Figura 17 é possível observar a localização dos resíduos identificados como epítomos conformacionais de células B, consenso pelos preditores, nos modelos estruturais construídos para cada proteína (destacadas por um círculo).



**Figura 16:** Estrutura das proteínas de *L. braziliensis* destacando em laranja os resíduos identificados como epítomos conformacionais de células B consenso entre os preditores. A) Proteína LbrM28\_V2.3070, o consenso está nos resíduos S213, F214, D302, P303, P304, Q305, Q306; B) Proteína LbrM28\_V2.1900, o consenso está nos resíduos R745-D746, 864-881 (EELRQRRHQGGPGLPA), E897-A898, E988; C) Proteína LbrM22\_V2.0110, o consenso está nos resíduos K2999 e G301.







**Figura 17:** Estrutura das proteínas de *L. infantum* destacando em laranja os resíduos identificados como epítomos conformacionais de células B consenso entre os preditores. A) Proteína LinJ.28.1850, o consenso está nos resíduos G282, P332, E333, P447, R743, N744, L840, E964, G966; B) Proteína LinJ.18.1500, o consenso está nos resíduos D45, E46, R47, P85, S86, W87, K555, D556, A557, E562, S691, V692, D847, V849, L850.

Com relação à predição de epítomos da proteína LinJ.19.0350 não ocorreu concordância entre todos os preditores.

### 5.1.3 Predição de regiões desordenadas

Abordagens preditivas visando à identificação de trechos desordenados foram empregadas para as seis proteínas modeladas. Quatro proteínas tiveram regiões preditas como desordenadas e estão sumarizadas no Quadro 5.

**Quadro 5:** Trechos de sequências preditos como desordenados em consenso pelos programas DisEMBL, GlobPipe, IUPred e VSL2B.

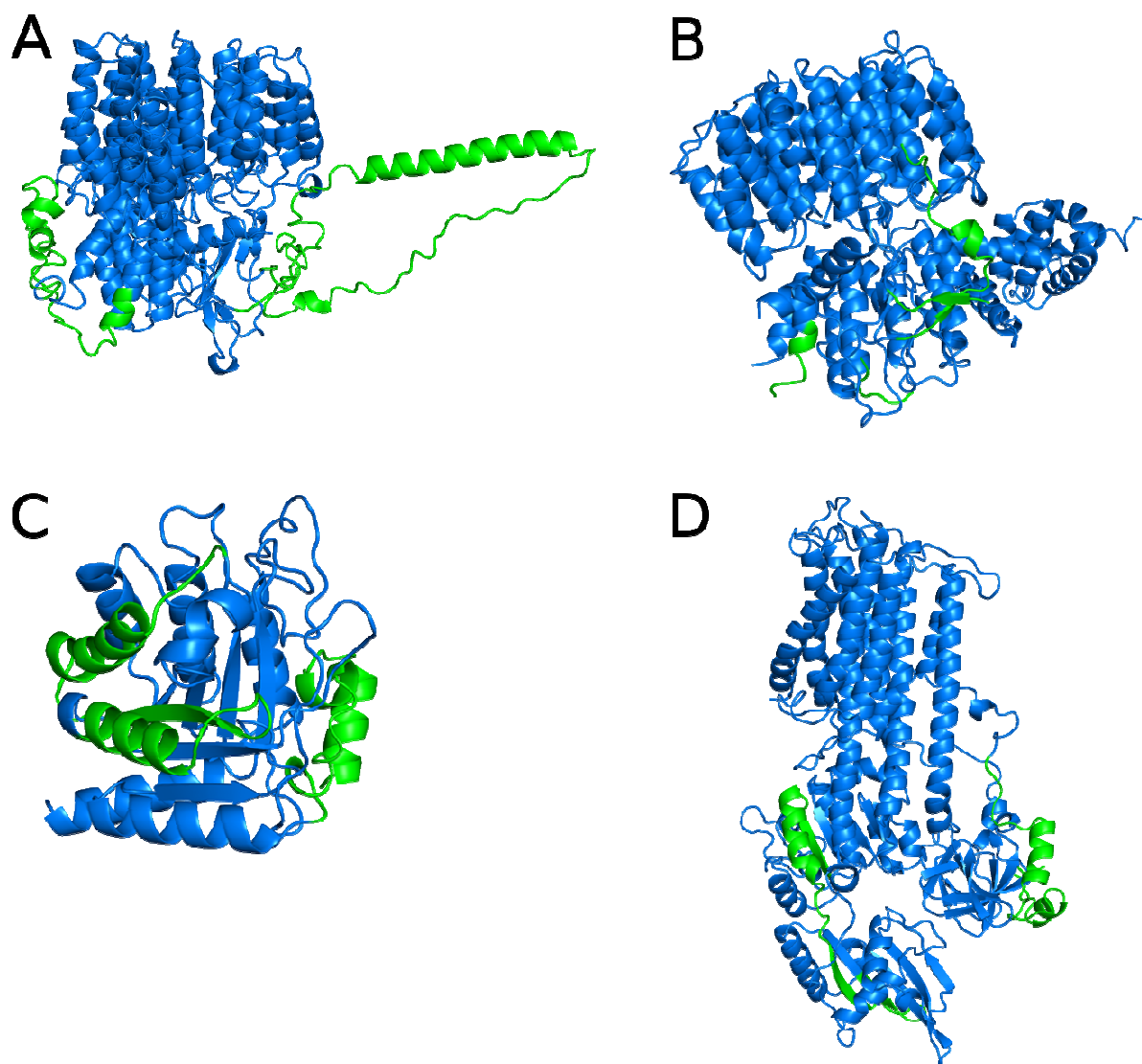
| Proteína original | Trechos desordenados <sup>*,**</sup> |
|-------------------|--------------------------------------|
| LinJ.18.1500      | 1..82, 482..525, 904..964            |
| LinJ.19.0350      | 1..44, 52..141, 273..320             |
| LinJ.28.1850      | 287..339 822..925                    |
| LbrM28_V2.1900    | 288..341 823..927                    |

Fonte: Elaborado pelo autor.

\*As posições dizem respeito aos resíduos nas proteínas de estudo.

\*\*Coordenadas agrupadas por .. indicam trechos consecutivos da proteína

Na Figura 18 é possível observar os trechos desordenados nos modelos estruturais gerados na etapa de modelagem.



**Figura 18: Localização das regiões preditas como desordenadas (em verde) nos modelos estruturais das proteínas.** Proteína de *L. braziliensis* (A) LbrM28\_V2.1900 e proteínas de *L. infantum* (B) LinJ.28.1850, (C) LinJ.19.0350 e (D) LinJ.18.1500

Podemos citar alguns trechos proteicos preditos como desordenados que coincidem com predições consenso de epítomos conformacionais de célula B, representados na Figura 19 (destacados por um círculo). Entre eles, os resíduos 864-881 (EELRQRRHQGGPGLPA), E897-A898, E988 da proteína LbrM28\_V2.1900, os resíduos D45, E46, R47 da proteína LinJ.18.1500 e os resíduos G282, P332, E333, L840, da proteína LinJ.28.1850.

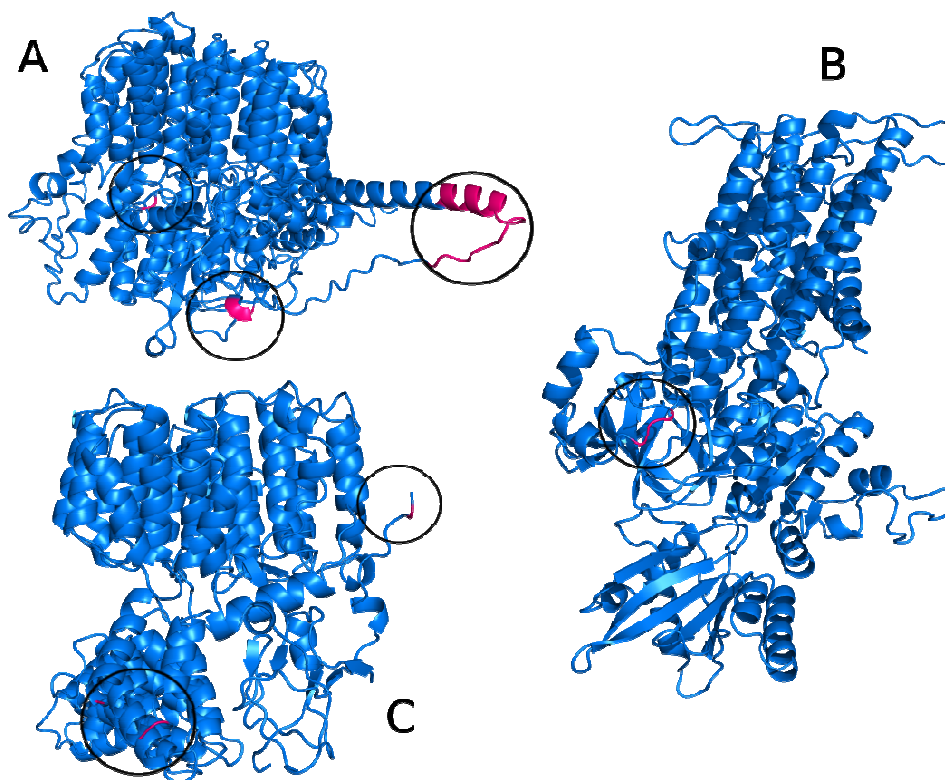


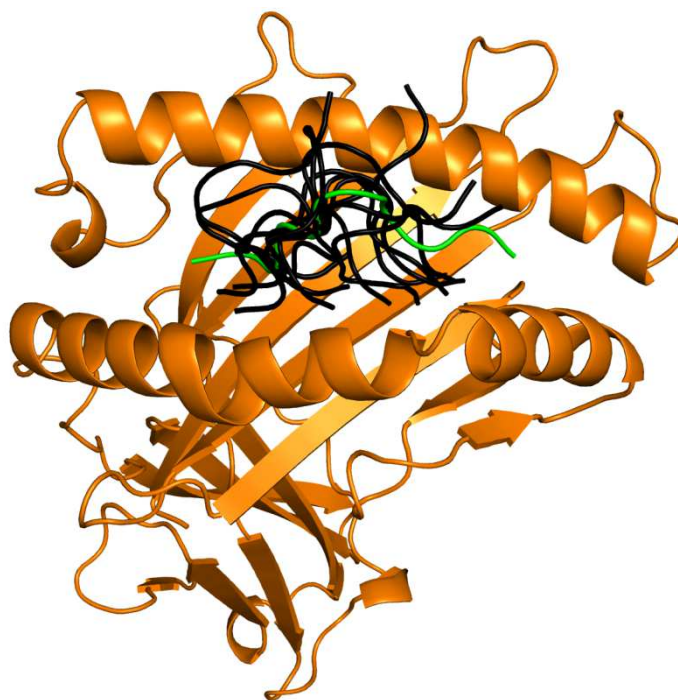
Figura 19: Trechos proteicos preditos como desordenados que coincidem com predições consenso de epítomos conformacionais de célula B, em rosa. Proteína de *L. braziliensis* (A) LbrM28\_V2.1900 e proteínas de *L. infantum* (B) LinJ.18.1500. (C) LinJ.28.1850.

## 5.2 Parte 2- Análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II usando atracamento molecular

### 5.2.1 Atracamento molecular aplicado à análise estrutural da afinidade de ligação de peptídeos com MHC de classes I e II

O atracamento molecular foi realizado com o pacote de programas Rosetta. Particularmente, utilizamos o programa denominado *Flex Pep Docking*, direcionado especificamente para estudos de atracamento molecular de peptídeos. Este programa foi implementado para cada um dos peptídeos ligantes com cada uma das moléculas de MHC de classe I e II selecionadas para este trabalho.

Após as simulações de atracamento molecular foram obtidas várias conformações preditas para cada uma das possibilidades de combinações entre peptídeos e moléculas de MHC (Figura 20).

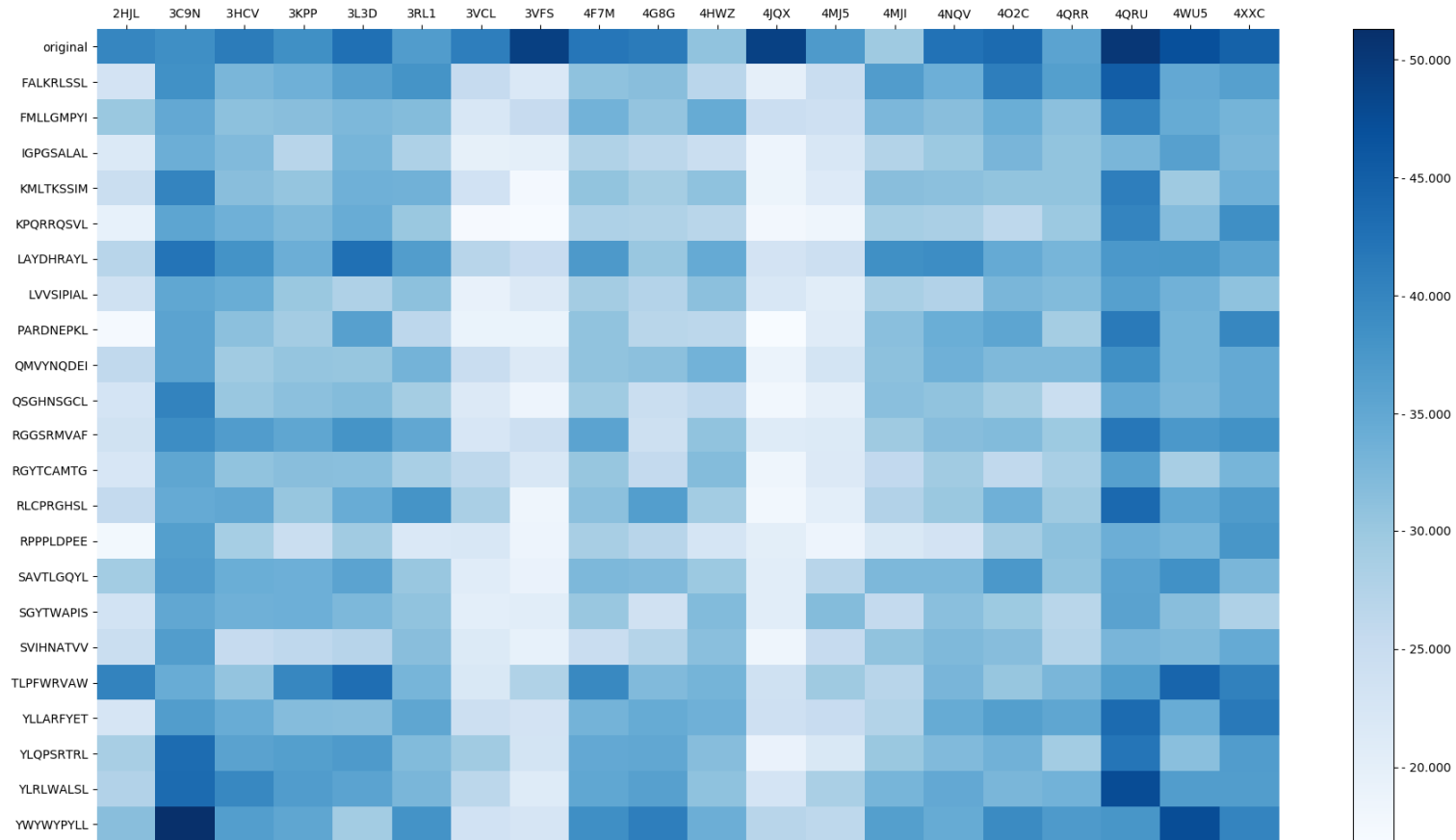


**Figura 20: Possibilidades de conformações geradas pelo atracamento molecular.** Em laranja, a molécula de MHC I de humano 2HJL (GILLESPIE et al., 2006); em verde, a conformação do peptídeo cocrystalizado; em preto, 10 possibilidades de conformações para o peptídeo avaliado (FALKRLSSL) segundo as simulações de atracamento.

Para cada uma das possíveis conformações são geradas pontuações denominadas *Isc*. A conformação com o melhor *Isc* foi selecionada (resultados completos

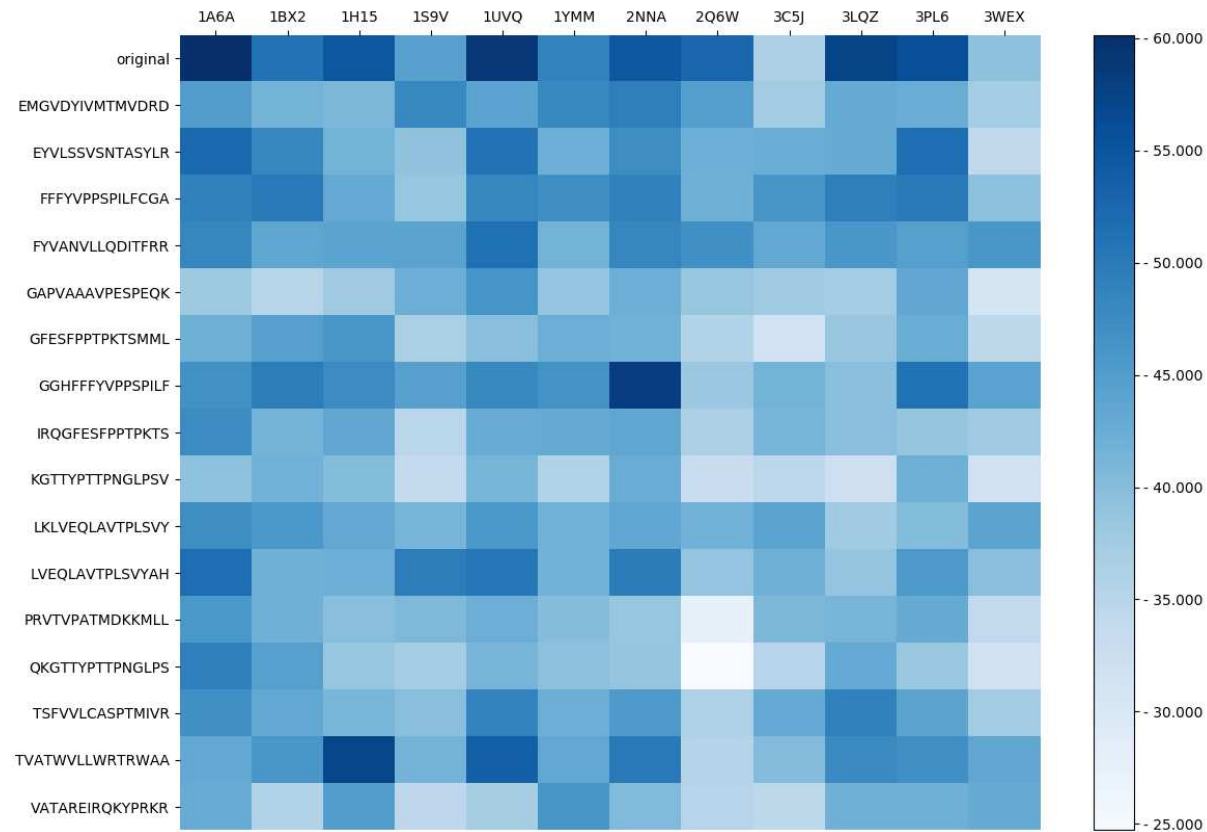
detalhados nos apêndices K e L) e utilizada para construção dos mapas de interação para as duas classes de MHC (Figura 21 e Figura 22), quanto mais clara a cor, maior o valor de  $I_{sc}$  e mais fraca a estabilidade da conformação gerada pela simulação de atracamento molecular. Na primeira linha desses mapas de interação estão os resultados para os atracamentos realizados entre as moléculas de MHC e os peptídeos que foram cocristalizados com as mesmas (originais). Nas linhas seguintes estão representados os resultados das simulações de atracamento molecular realizadas entre as moléculas de MHC e as sequências de peptídeos em estudo.





**Figura 21:** Resultado das simulações de atracamento molecular entre os peptídeos avaliados, peptídeos originais e moléculas de MHC I. Quanto mais intensa a cor menor o valor de  $I_{sc}$ , o que é indicativo de maior possibilidade de uma apresentação bem-sucedida do peptídeo para o sistema imune.





**Figura 22: Resultado das simulações de atracamento molecular entre os peptídeos avaliados, peptídeos originais e moléculas de MHC II.** Quanto mais intensa a cor menor o valor de  $I_{sc}$ , o que é indicativo de maior possibilidade de uma apresentação bem sucedida do peptídeo para o sistema imune.

As Tabela 7 e Tabela 8 mostram um ranqueamento dos peptídeos analisados em função do número de alelos em que cada peptídeo alcançou um Isc menor, indicativo de maior afinidade, do que o do peptídeo original. Dentro dessa perspectiva, os peptídeos que apresentaram maior afinidade com um maior número de alelos de MHC foram: o peptídeo YWYWYPYLL (seis simulações), para simulações com as estruturas de MHC I, e o peptídeo GGHFFFYVPPSPILF (quatro simulações) para simulações com as estruturas de MHC II. Peptídeos com um valor de Isc menor quando comparados ao peptídeo original para mais de um alelo tendem a possuir uma maior capacidade de interação com variações estruturais (resultante das variações alélicas) das moléculas de MHC, e por isso podem ser mais abrangentes quando utilizados para indução de resposta imune na população humana.

**Tabela 7: Ranqueamento dos peptídeos analisados em função do número de alelos para os quais as simulações apresentaram Isc menor em comparação à simulação MHC I- peptídeo cocristalizado.**

| Peptídeo<br>(posição no<br>ranqueamento) | Número de<br>alelos | Peptídeo      | Número<br>de alelos | Peptídeo      | Número<br>de alelos |
|--|---------------------|---------------|---------------------|---------------|---------------------|
| YWYWYPYLL (1)                            | 6/21                | QSGHNSGCL (4) | 2/21                | YLLARFYET (5) | 1/21                |
| TLPFWRVAW (2)                            | 4/21                | QMVYNQDEI (4) | 2/21                | IGPGSALAL (6) | 0/21                |
| KMLTKSSIM (3)                            | 3/21                | RGGSRMVAF (4) | 2/21                | RPPPLDPEE (6) | 0/21                |
| LAYDHRAYL (3)                            | 3/21                | SGYTWAPIS (5) | 1/21                |               |                     |
| FALKRLSSL (3)                            | 3/21                | LVVSIPIAL (5) | 1/21                |               |                     |
| YLRLWALSL (3)                            | 3/21                | SAVTLGQYL (5) | 1/21                |               |                     |
| YLQPSRTRL (3)                            | 3/21                | RGYTCAMTG (5) | 1/21                |               |                     |
| FMLLGMPYI (4)                            | 2/21                | RLCPRGHSL (5) | 1/21                |               |                     |
| SVIHNATVV (4)                            | 2/21                | PARDNEPKL (5) | 1/21                |               |                     |

Fonte: Elaborado pelo autor.

**Tabela 8: Ranqueamento dos peptídeos analisados em função do número de alelos para os quais as simulações apresentaram Isc menor em comparação à simulação MHC II- peptídeo cocristalizado.**

| Peptídeo<br>(posição no<br>ranqueamento) | Número de alelos | Peptídeo             | Número<br>de alelos |
|--|------------------|----------------------|---------------------|
| GGHFFFYVPPSPILF (1)                      | 4/12             | TSFVVLICASPTMIVR (6) | 1/12                |

|                     |      |                     |      |
|---------------------|------|---------------------|------|
| LVEQLAVTPLSVYAH (2) | 3/12 | IRQGFESFPPTPKTS (6) | 1/12 |
| TVATWVLLWRTRWAA (2) | 3/12 | FFFYVPPSPILFCGA (6) | 1/12 |
| EMGVYIVMTMVDRD (4)  | 2/12 | VATAREIRQKYPRKR (6) | 1/12 |
| LKLVEQLAVTPLSVY (5) | 2/12 | QKGTTPPTPNGLPS (7)  | 0/12 |
| FYVANVLLQDITFRR (5) | 2/12 | KGTTTPPTPNGLPSV (7) | 0/12 |
| GAPVAAAVPESPEQK (6) | 1/12 | GFESFPPTPKTSMML (7) | 0/12 |
| PRVTVPATMDKKMLL (6) | 1/12 |                     |      |
| EYVLSSVSNTASYLR (6) | 1/12 |                     |      |

---

Fonte: Elaborado pelo autor.

## 6 DISCUSSÃO

### 6.1 Parte 1- Predição estrutural, predição de epítomos de células B conformacionais e correlação com desordem

#### 6.1.1 Modelagens

A determinação de estruturas proteicas a partir de moldes tem sido usada com sucesso em estudos de modelagem de complexos de ligação entre biomoléculas (XU et al., 1996), predição de sítios de ligação de proteínas (MATSUMOTO et al., 1995), modelagem molecular de organelas (SPAHN et al., 2001) e anotação funcional de genes (GOPAL et al., 2001). Um ponto crítico conhecido da técnica está relacionado à necessidade da existência de estruturas similares depositadas em bancos de dados estruturais. Este foi um fator que viabilizou a modelagem de apenas seis moléculas das 12 inicialmente propostas.

Utilizamos três diferentes programas de modelagem vinculados aos algoritmos I-TASSER, Modeller e RaptorX, e em cada abordagem foi utilizada uma forma distinta de busca de moldes. É importante ressaltar que dentro dessas estratégias encontramos consensos e não consensos na seleção dos moldes. Especificamente para a proteína LbrM28\_V2.3070 houve consenso entre as três abordagens e um mesmo molde foi selecionado como mais adequado para as modelagens estruturais computacionais. Em relação à proteína LbrM28\_V2.3070 houve consenso na seleção do melhor molde somente entre o I-TASSER e o Modeller. Para as proteínas LbrM22\_V2.0110 e LinJ.19.0350 houve consenso na seleção do molde entre o Raptor-X e I-TASSER. Já para a proteína LinJ.28.1850 o mesmo molde foi selecionado pelo Modeller e RaptorX. Para a proteína LbrM28\_V2.1900 não houve consenso na seleção do molde entre as abordagens.

Essas diferenças na seleção de moldes estão relacionadas à diferença na forma de busca de moldes empregadas em cada uma dessas abordagens. Enquanto a busca por moldes utilizada na modelagem pelo Modeller foi baseada em comparações de sequências proteicas par a par (BLAST), as duas outras abordagens foram baseadas em métodos threading, que envolvem comparação de sequências e estruturas de proteínas. O método threading é muito útil quando existem poucos

moldes disponíveis com alta similaridade com a sequência de proteína que se deseja modelar (FISER, 2010). Esse foi o método que demonstrou melhores resultados, uma vez que em cinco das seis proteínas modeladas, ele foi o método utilizado capaz de gerar modelos com menor proporção de resíduos em regiões não permitidas de acordo com análises realizadas utilizando gráficos de Ramachandran. Em relação às buscas empregadas, aquela utilizada através do servidor Web I-TASSER utiliza um meta-servidor (LOMETS) que emprega nove diferentes métodos Threading, e os moldes finais são selecionados a partir do consenso desses nove métodos (WU; ZHANG, 2007). O método de busca de moldes do RaptorX se baseia inicialmente na identificação de moldes únicos, para determinar os melhores; a partir dessa busca inicial é montada uma matriz de probabilidade que é utilizada para alinhamentos múltiplos entre a sequência a ser modelada e os possíveis moldes. Ao final os moldes são selecionados (PENG; XU, 2011). No caso das modelagens locais usando o Modeller selecionamos um único molde a partir de uma busca por similaridade usando o pacote BLAST.

A utilização de múltiplos moldes pode ser vantajosa em alguns sentidos: possibilidade de aumentar a cobertura do alinhamento entre a proteína a ser modelada e os moldes; e utilização de moldes que se complementam (uma dada proteína pode ser mais similar a um molde em uma região e a outro molde em outra região) (PENG; XU, 2011). Essas duas vantagens podem justificar o fato de a maioria das proteínas analisadas neste trabalho terem os melhores modelos construídos a partir de mais de um molde.

Em relação à qualidade, todos os modelos apresentaram uma porcentagem baixa de resíduos em regiões consideradas estruturalmente desfavoráveis nos gráficos de Ramachandran. E com exceção da proteína LinJ.19.0350, os modelos das proteínas apresentaram gráficos similares de escore DOPE quando comparados com os gráficos de escore DOPE dos molde, o que é indicativo de um bom modelo para se fazer predições de epítomos conformacionais de células B. Finalmente, considerando a qualidade dos modelos gerados, o seu potencial emprego na identificação de motivos de interação proteína/sistema imune e os baixos custos associados, avaliamos que o emprego da técnica é bastante promissor em análises de imunoinformática visando ao direcionamento de uma experimentação mais assertiva.

### 6.1.2 Predição de epítomos conformacionais de células B

Nos últimos anos ocorreu um crescente fortalecimento do emprego da imunoinformática no contexto de desenvolvimento de vacinas e ferramentas de diagnóstico. Essa estratégia computacional de identificação de potenciais moléculas indutoras de resposta imune permite a busca por imunofármacos, novos antígenos vacinais, em um curto período de tempo com custos reduzidos (THOMAS; DILBAROVA; RAPPUOLI, 2016, BAMBINI; RAPPUOLI, 2009), como foi o caso da vacina contra *Meningococcus B* que foi licenciada na Europa em 2013 e nos Estados Unidos da América em janeiro de 2015 (GIULIANI et al., 2006).

Nos últimos anos, além da análise de dados de sequências de aminoácidos, vem sendo adotada na imunoinformática a utilização de informações estruturais (DORMITZER; ULMER; RAPPUOLI, 2008), o que é imprescindível para responder questões acerca das interações entre sistema imune e antígenos. Nesse contexto, o presente trabalho emprega de maneira integrativa as mais recentes estratégias de modelagem estrutural visando à identificação e a avaliação de epítomos de *L. infantum* e *L. braziliensis* a partir de antígenos previamente selecionados através de vacinologia reversa.

A primeira etapa deste trabalho consistiu na modelagem estrutural de proteínas de *Leishmania* para posterior predição de epítomos de células B conformacionais. Os epítomos de células B podem ser classificados em lineares e conformacionais. Os lineares são compostos de uma sequência contínua de resíduos, enquanto os conformacionais são constituídos de resíduos com descontinuidade sequencial, mas com vizinhança espacial (ABBAS; LICHTMAN; PILLAI, 2015). Um estudo anterior demonstrou que a maioria dos epítomos de células B são conformacionais (HASTE ANDERSEN; NIELSEN; LUND, 2006), o que subsidia uma atenção especial em métodos capazes de identificar esse tipo de epítomo, que foi um dos objetos de estudo deste trabalho.

Até o momento, o desempenho de métodos para mapeamento *in silico* de epítomos de células B é considerado moderado. Podemos levantar dois pontos como os mais significantes para explicar tal fato, são eles: a) atualmente, nem mesmo as abordagens experimentais de referência (métodos baseados em cristalografia, métodos baseados em espectrometria de massa, ressonância magnética nuclear e

ressonância plasmônica de superfície) são capazes de mapear todos os potenciais epítomos em um determinado antígeno, e assim seria difícil esperar tal desempenho das abordagens computacionais; e b) a produção de anticorpos é elicitada contra um antígeno frente a um dado contexto biológico multifatorial (presença de moléculas coestimulatórias, microambiente) e não contra o monômero antigênico isolado avaliado nas abordagens *in silico*. A desconsideração desses aspectos leva à predição de falsos positivos e consequentemente a baixos valores de desempenho (KRINGELUM et al., 2012). No sentido de diminuir o viés relacionado ao item b, o resultado consenso de quatro preditores foi aplicado neste trabalho.

Em relação ao papel da resposta humoral em leishmaniose, em 2006, Reis e colaboradores, estudando leishmaniose visceral, avaliaram cães infectados com a *L. (L.) chagasi* e relataram uma correlação positiva entre parasitismo tecidual e níveis de IgG, IgG2, IgM e IgA. De forma diferente para IgG1 nesse mesmo trabalho foi relatada uma correlação negativa entre essa subclasse de imunoglobulina e parasitismos e sinais clínicos (REIS et al., 2006). Outros trabalhos em cães também relataram essa associação (ALMEIDA et al., 2005). Em outro trabalho foi demonstrado que a subclasse de imunoglobulina IgG2 estaria associada à resistência contra leishmaniose canina, enquanto IgG1 estaria relacionada à suscetibilidade (DE OLIVEIRA MENDES et al., 2003). Além disso, de acordo com Day (2007) a resposta humoral não seria polarizada, uma vez que essas duas subclasses de imunoglobulinas se elevariam com o passar do tempo e essa observação também seria verdadeira para casos humanos de leishmaniose visceral (DAY, 2007).

Para leishmaniose cutânea, Ulrich e colaboradores (1995) reportaram altos níveis de imunoglobulinas da subclasse IgG4 em manifestação do tipo cutânea difusa (o tipo associado ao pior prognóstico) comparando as subclasses de IgG nas respostas em diferentes infecções caracterizadas por anergia de resposta celular (leishmaniose visceral e hanseníase) (ULRICH et al., 1995). Em relação às subclasses IgG1, IgG2 e IgG3, Rodriguez e Ulrich (1996) indicaram uma possível correlação entre detecção dessas subclasses e imunidade mediada por célula (RODRIGUEZ; CENTENO; ULRICH, 1996).

Considerando os aspectos citados acerca da real contribuição da resposta humoral nas leishmanioses, os resultados de epítomos conformacionais de células B deste

trabalho poderão contribuir para a elucidação dessa questão, uma vez que se estes epítomos forem confirmados como imunoprotetores em análises experimentais, o subtipo de imunoglobulinas associados ou não à sua proteção poderão ser determinados e avaliados.

Entre os preditores de epítomos de células B conformacionais houve ausência de consenso entre os resultados para as proteínas analisadas, o que pode ser explicado pela diferença nas abordagens empregadas por cada algoritmo. O programa Ellipro faz predições a partir de identificações de protusões na estrutura e foi desenvolvido sem treinamento de máquinas. Ele foi o preditor que gerou maior número de predições para todos os modelos avaliados. Esse maior número pode estar associado à não utilização de escala de propensão de aminoácidos. O BePro combina pontuações de propensão de aminoácidos e identificação de superfície de acessibilidade do solvente ao resíduo e foi o preditor que gerou menos predições. O Discotope, de forma similar ao BePro, combina pontuações de propensão de aminoácidos e superfície de acessibilidade do solvente ao resíduo e apresentou maior consenso de predição com o preditor Bepro. O EPCES utiliza uma abordagem baseada em atracamento molecular e gerou predições com poucos resíduos consenso com os outros preditores. Como os preditores utilizam abordagens diferentes de predição, os resíduos preditos por todos eles possuem maiores chances de induzir resposta imune protetora conforme algumas publicações (LIANG et al., 2009; MUKONYORA, 2015).

As características físico-químicas dos epítomos de células B foram analisadas em diversos trabalhos. Em 1981, Hopp e Woods relataram uma possível correlação entre aminoácidos hidrofílicos carregados e epítomos. Desde então vários parâmetros físico-químicos foram propostos para assistirem na localização de epítomos de células B. Em relação ao número de resíduos de aminoácidos os estudos sugerem que os epítomos de células B possuem de 10 a 20 resíduos (KRINGELUM et al., 2013). Nesse contexto, baseado nas predições consenso deste trabalho, apenas para a proteína LbrM28\_V2.1900 (Figura 16B) observamos uma região com 18 resíduos contíguos. Para as outras predições consenso com até cinco resíduos contíguos foram identificados (LbrM28\_V2.3070). Tal fato poderia ser explicado pela existência de epítomos conformacionais, nos quais os resíduos não precisam necessariamente estar contíguos para interação com o anticorpo. Assim



sendo, para algumas das proteínas estudadas foram observadas regiões consenso com vizinhança espacial, como por exemplo, nas proteínas, LbrM22\_V2.0110, LinJ.28.1850, LinJ.18.1500, (Figura 16C, Figura 17A e Figura 17B). É importante ainda citar que para o preditor BEpro pouquíssimas regiões foram preditas para todas as proteínas e, como usamos o consenso de quatro preditores, regiões capazes de estimular resposta de células B *in vivo* podem ter sido descartadas. Sendo assim, os resíduos preditos neste trabalho como capazes de estimular resposta de células B devem ser vistos como indicadores de uma possível interação com anticorpo na região estudada. Estudos posteriores visando à avaliação experimental desses resíduos devem, portanto, incluir a análise do domínio estrutural no qual os resíduos se inserem.

De acordo com artigos publicados em 2008 por Rubinstein e colaboradores e Ofran e colaboradores, os epítomos de células B estão presentes em sua maioria em regiões de *loops* e em menor frequência nas conformações secundárias (alfa-hélices e folhas-beta) (OFRAN; SCHLESSINGER; ROST, 2008; RUBINSTEIN et al., 2008). Neste trabalho a maior parte das regiões do consenso representam *loops* (Figura 16 e Figura 17), sendo que apenas para as proteínas LbrM28\_V2.3070 e LbrM28\_V2.1900 foram preditos epítomos sobre alfa-hélices (Figura 16A e Figura 16B, respectivamente). Considerando ainda o aspecto de acessibilidade estrutural, todas regiões consenso preditas estão localizadas em porções expostas das proteínas, a princípio acessíveis para o anticorpo.

É importante considerar que preditores de células B como um todo ainda apresentam valores de sensibilidade e especificidade baixos (BACKERT; KOHLBACHER, 2015). Além disso, esses resultados foram obtidos a partir de modelos estruturais utilizando abordagens *in silico* que, apesar da boa qualidade, podem conter imprecisões, principalmente para as proteínas LinJ.19.0350, LinJ.18.1500 e LinJ.28.18.50 (Figura 12) que apresentaram menor identidade e cobertura em relação aos seus respectivos moldes. Por esse motivo todos os resultados advindos desses modelos devem ser considerados com cautela, mas são de qualquer forma um indicador valioso na busca de epítomos em proteínas, principalmente quando se deseja avaliar antígenos em larga escala.

### 6.1.3 Predição de regiões desordenadas

As estratégias de seleção de candidatos vacinais através da vacinologia reversa permitiram a busca de antígenos em diversos patógenos (CHIANG et al., 2015; MAIONE, 2005; MONTIGIANI et al., 2002; NAZ et al., 2015; TALUKDAR et al., 2014; WIZEMANN et al., 2001). Em 2012 foi licenciada na Europa com base nesse método a vacina contra meningococos B (SERRUTO et al., 2012). Apesar de todo o sucesso, ainda há espaço para aprimoramento da técnica. Nesse intuito é fundamental a busca de novas características relacionadas à resposta imune e ainda, como já foi citada, a utilização de características estruturais na seleção de novos antígenos. A predição de regiões desordenadas neste trabalho seguiu esse propósito.

O trabalho pioneiro de Westhof e colaboradores (1984) descreveu que a flexibilidade da cadeia de resíduos de determinada proteína pode ser um indicativo de antigenicidade (WESTHOF et al., 1984). Nesse mesmo ano Stahl e colaboradores também correlacionaram imunodominância com regiões desordenadas estruturais (STAHL et al., 1984). Os resultados gerados pelo presente estudo corroboram esses achados, uma vez que seis regiões de um total de 23 regiões proteicas preditas como imunogênicas, neste trabalho, coincidem com regiões preditas como desordenadas estruturalmente. Essas observações podem estar correlacionadas a uma maior exposição da região imunogênica proporcionado pela desordem estrutural proteica. Sendo assim, a ocorrência dos dois fenômenos em uma mesma região pode resultar em maiores chances de indução de resposta imunológica in vivo. Corroborando essa hipótese, podemos citar alguns antígenos que apresentam epítomos com desordem estrutural proteica, como o antígeno AMA-1 de *Plasmodium vivax* que foi validado experimentalmente e se localiza em uma região predita como desordenada (BUENO et al., 2011). Existe também a proteína de superfície do merozoíto 2 (ADDA et al., 2012), e a proteína de HIV do tipo 1, wt-Tat (FOUCAULT et al., 2009). A correlação entre desordem estrutural proteica e resposta imune é ainda muito pouco elucidada, se comprovada, através de avaliações experimentais, pode ser uma grande aliada na busca computacional de antígenos vacinais.

Muitas regiões preditas como desordenadas coincidem com regiões sem estrutura secundária definida nos modelos das proteínas estudadas, como é possível observar na Figura 18. Contudo existiram situações onde regiões desordenadas foram preditas coincidindo com a estrutura secundária predita para os modelos validados, exemplos disso podem ser observados na alfa-hélice da proteína LbrM28\_V2.1900 (Figura 18A) e na folha-beta da proteína LinJ.19.0350 (Figura 18). Esse achado pode estar relacionado a erros nas predições de desordem estrutural proteica, erros na modelagem estrutural das proteínas, ou ainda, relacionados à existência nesses locais de inconsistência de regiões denominadas elementos de reconhecimento molecular. Esses elementos passam de um estado estrutural desorganizado para um estado estrutural organizado, após interagirem com algum ligante (OLDFIELD et al., 2005). Considerando esses elementos de reconhecimento molecular, hipotetizamos que alguns moldes utilizados na etapa de modelagem (Threading e Comparativa) pode ter sido cristalizado com algum ligante (MOHAN et al., 2006; VAN DER LEE et al., 2014) que pode, por sua vez, ter induzido o estabelecimento de uma conformação espacial na estrutura proteica. Essa conformação espacial induzida pelo ligante pode ter sido então transferida para o modelo e assim culminado na inconsistência observada.

## **6.2 Parte 2 – Análise estrutural da afinidade de ligação de peptídeos com MHC de classe I e II usando atracamento molecular**

O processamento e a apresentação de antígenos via MHC são fundamentais na resposta imune a várias doenças incluindo as leishmanioses, nas quais a resposta celular possui papel central na resistência às doenças. Os linfócitos T CD8+ são estimulados através de reconhecimento de antígenos ligados a MHC I presentes em todas as células nucleadas, atuando como células citotóxicas e induzindo morte em células infectadas por parasitos, além de participarem na modulação da resposta imune. Os linfócitos T CD4+ são estimulados por antígenos ligados a MHCII presentes em células apresentadoras de antígenos e são essenciais na resposta contra leishmaniose. Depois de estimulados, os linfócitos T CD4+ podem se diferenciar em dois subtipos principais conhecidos como linfócitos Th1 ou Th2. O linfócito Th1 é o subtipo que está associado ao controle da leishmaniose, pois é capaz de direcionar a resposta imune contra parasitos intracelulares, como é o caso das amastigotas de leishmania (CASTELLANO et al., 2009). Já o subtipo de

linfócitos Th2 direciona a resposta imune contra parasitos extracelulares. Considerando a importância dessas respostas, a predição de epítomos de células T é imprescindível para a descoberta de novos antígenos e desenvolvimento de novas vacinas contra *Leishmania* (DARRAH et al., 2007).

Uma das grandes revoluções da atualidade são as abordagens científicas capazes de produzir resultados em larga escala. No entanto a alta produção de dados muitas vezes não é acompanhada de novos conhecimentos devido a impossibilidades técnicas e científicas. O mesmo paradoxo acontece na imunoinformática. Preditores de epítomos lineares, quando aplicados em escala proteômica geram um grande número de potenciais epítomos e é impraticável testar todos experimentalmente. Nessa segunda parte do trabalho visamos ao estabelecimento de um processo para ranquear os epítomos lineares preditos em um trabalho anterior (Brito e colaboradores, 2011) para experimentação. O protocolo utilizado envolveu o atracamento molecular, através do qual é possível analisar interações entre peptídeos e moléculas de MHC. Essa metodologia permite triagem e ranqueamento em grande escala de epítomos em um curto período.

Os epítomos utilizados neste trabalho foram selecionados com base em predições realizadas em sequências de aminoácidos de proteínas de *Leishmania*. Esses preditores de epítomos lineares, em sua maioria, se baseiam em aprendizado de máquina, método que possui grande dependência de bons conjuntos de dados para treinamento. Até o momento nenhum dos preditores disponível foi treinado usando dados experimentais de *Leishmania* (E SILVA et al., 2016), desse modo, esses preditores devem ser utilizados com cautela em tripanossomatídeos. Além disso, apesar dos preditores de epítomos capazes de predizer afinidade com MHC I gerarem predições com altas taxas de acerto para muitos organismos avaliados, os preditores de epítomos capazes de predizer afinidade com MHC II ainda não são tão eficientes quanto os de MHC I. Essa menor taxa de acerto é correlacionada com o tamanho dos peptídeos de MHC II que são maiores (BACKERT; KOHLBACHER, 2015). Sendo assim, a análise de atracamento molecular realizada neste trabalho visa a suplantiar esses obstáculos e gerar mais confiabilidade aos resultados da predição linear de epítomos realizada anteriormente.

A resposta imunológica induzida por linfócitos T CD8+ pode contribuir para a proteção contra leishmaniose ou induzir mecanismos patológicos. Esse tipo de linfócito induz resposta protetora quando produz IFN- $\gamma$ , crítico para o

desenvolvimento de linfócitos T CD4<sup>+</sup> Th1, mas promove patologia quando induz resposta citolítica (NOVAIS; SCOTT, 2015). Considerando esses aspectos, neste trabalho fizemos a busca por epítomos usando tanto estruturas de MHC I quanto MHC II. A combinação de epítomos capazes de se ligarem a MHC I e epítomos capazes de se ligarem a MHC II, em uma posterior avaliação em testes vacinais, poderão induzir uma resposta mais balanceada e capaz de prevenir a progressão da leishmaniose. Além disso, as análises, através de atracamento molecular, foram capazes de indicar peptídeos que tendem a apresentar boa afinidade de ligação com mais de uma molécula de MHC (Tabela 7 e Tabela 8) e por isso tendem a possuir maior potencial de induzir resposta imune protetora de forma mais abrangente na população, premissa essencial para o desenvolvimento de uma vacina.

Considerando os ranqueamentos gerados (Tabela 7 e Tabela 8) podemos supor que os peptídeos de *Leishmania infantum* que apresentaram melhores resultados com múltiplas moléculas de MHC nos atracamentos moleculares, como por exemplo, o peptídeo YWYWYPYLL para MHC I e o peptídeo GGHFFFYVPPSPILF para MHC II, irão preferencialmente ser apresentados ao sistema imune e terão maiores chances de elicitarem resposta imune. Desse modo, estudos posteriores envolvendo experimentação *in vitro* e *in vivo* deverão priorizar esses peptídeos, melhor ranqueados.

A metodologia de priorização de candidatos vacinais, usada neste trabalho, através de atracamento molecular, mostrou-se viável, rápida e de baixa complexidade, possibilitando o ranqueamento de 22 epítomos em 22 estruturas de MHC I e 16 epítomos em 12 estruturas de MHC II. Além disso, essas análises foram capazes de identificar epítomos com bom potencial de estimularem o sistema imune humano a desenvolver uma resposta protetora.

## 7 CONCLUSÕES

As principais conclusões desta dissertação podem ser condensadas nos seguintes tópicos:

- Para seis das 12 proteínas avaliadas foi possível a realização da modelagem estrutural utilizando moldes depositados nos bancos de dados de domínio público.
- Entre as abordagens usadas, o servidor RaptorX foi o que produziu um melhor modelo para mais proteínas, quatro de um total de seis proteínas modeladas.
- Sequências com menor cobertura e identidade com seus respectivos moldes apresentaram resultados relativos à modelagem estrutural menos satisfatórios do que aquelas com maior cobertura e identidade.
- Através do consenso das predições de epítomos conformacionais de células B utilizando quatro preditores foram obtidas regiões de possíveis epítomos para cinco das seis proteínas modeladas.
- Muitas das possíveis regiões indicadas como imunoestimuladoras seguem algumas das características físico-químicas de epítomos de células B descritas na literatura.
- Seis regiões preditas como estruturalmente desordenadas coincidiram com regiões preditas como capazes de estimular resposta humoral, correlação que vem sendo relatada por alguns pesquisadores. No entanto, mais estudos são necessários para consolidar essa informação.
- Simulações de atracamento molecular para predições de interação entre peptídeos e moléculas de MHC de classes I e II realizadas com programas do pacote Rosetta foram capazes de realizar um ranqueamento em dados obtidos a partir de predições de epítomos de células T CD8<sup>+</sup> e T CD4<sup>+</sup>, realizadas anteriormente. Além disso, essa abordagem foi capaz de adicionar mais uma camada de confiança aos dados e de selecionar epítomos com

potencial de induzir de forma mais abrangente uma resposta imune protetora nas populações humanas.

## **8 PERSPECTIVAS**

Uma das principais perspectivas deste estudo é realização da validação experimental e construção de uma quimera proteica utilizando os epítomos considerados mais promissores.

Pretendemos também realizar o desenvolvimento de um fluxo de análise e definição de aspectos estruturais importantes para uma vacina contra leishmaniose utilizando os métodos propostos por este trabalho para a avaliação de outros candidatos vacinais em estudo.

Verificar se as proteínas selecionadas são membros de famílias proteicas e analisar se os epítomos são conservados entre os membros dessas famílias proteicas.

Adicionalmente, além da abordagem utilizando atracamento molecular, pretendemos agregar ao fluxo de análise simulações de dinâmica molecular para refinamento das predições e geração de resultados mais confiáveis.





## REFERÊNCIAS

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Cellular and molecular immunology**. Philadelphia, PA: Elsevier Saunders, 2015.
- ADDA, C. G. et al. Antigenic Characterization of an Intrinsically Unstructured Protein, Plasmodium falciparum Merozoite Surface Protein 2. **Infection and Immunity**, v. 80, n. 12, p. 4177–4185, dez. 2012.
- ALMEIDA, M. A. O. et al. Antileishmanial antibody profile in dogs naturally infected with Leishmania chagasi. **Veterinary Immunology and Immunopathology**, v. 106, n. 1–2, p. 151–158, jun. 2005.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990a.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990b.
- BACKERT, L.; KOHLBACHER, O. Immunoinformatics and epitope prediction in the age of genomic medicine. **Genome Medicine**, v. 7, n. 1, dez. 2015a.
- BACKERT, L.; KOHLBACHER, O. Immunoinformatics and epitope prediction in the age of genomic medicine. **Genome Medicine**, v. 7, n. 1, dez. 2015b.
- BAMBINI, S.; RAPPUOLI, R. The use of genomics in microbial vaccine development. **Drug Discovery Today**, v. 14, n. 5–6, p. 252–260, mar. 2009.
- BARLOW, P. N. et al. Synthesis and some properties of constrained short-chain phosphatidylcholine analogues: (+)- and (-)-(1,3/2)-1-O-(phosphocholine)2,3-O-dihexanoylcyclopentane-1,2,3-triol. **Chemistry and Physics of Lipids**, v. 46, n. 3, p. 157–164, mar. 1988.
- BELTRAMI, A. et al. Crystal structure of HLA-B\*2709 complexed with the double citrullinated vasoactive intestinal peptide type 1 receptor (VIPR) peptide (residues 400-408). 9 jun. 2009.
- BENNETT, C. L. et al. Silent infection of bone marrow-derived dendritic cells by Leishmania mexicana amastigotes. **European Journal of Immunology**, v. 31, n. 3, p. 876–883, mar. 2001.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1–3, p. 43–56, set. 1995.
- BERMAN, H. M. et al. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 1 jan. 2000.

BRITO, R. C. F. DE; RESENDE, D. DE M.; REIS, A. B. **Identificação de novos antígenos candidatos vacinais contra Leishmaniose Visceral Canina no genoma de *L. infantum* utilizando a Bioinformática como ferramenta**, 2014.

BUENO, L. L. et al. Identification of a Highly Antigenic Linear B Cell Epitope within *Plasmodium vivax* Apical Membrane Antigen 1 (AMA-1). **PLoS ONE**, v. 6, n. 6, p. e21289, 21 jun. 2011.

BUUS, S. et al. Sensitive quantitative predictions of peptide-MHC binding by a “Query by Committee” artificial neural network approach. **Tissue Antigens**, v. 62, n. 5, p. 378–384, nov. 2003.

CASTELLANO, L. R. et al. Th1/Th2 immune responses are associated with active cutaneous leishmaniasis and clinical cure is associated with strong interferon- $\gamma$  production. **Human Immunology**, v. 70, n. 6, p. 383–390, jun. 2009.

CHEN, J. et al. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. **Amino Acids**, v. 33, n. 3, p. 423–428, set. 2007.

CHEN, W. et al. Determinant selection of major histocompatibility complex class I-restricted antigenic peptides is explained by class I-peptide affinity and is strongly influenced by nondominant anchor residues. **The Journal of Experimental Medicine**, v. 180, n. 4, p. 1471–1483, 1 out. 1994.

CHIANG, M.-H. et al. Identification of novel vaccine candidates against *Acinetobacter baumannii* using reverse vaccinology. **Human Vaccines & Immunotherapeutics**, v. 11, n. 4, p. 1065–1073, 3 abr. 2015.

CHOTHIA, C.; LESK, A. M. The relation between the divergence of sequence and structure in proteins. **The EMBO Journal**, v. 5, n. 4, p. 823–826, abr. 1986.

CORNELL, W. D. et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197. **Journal of the American Chemical Society**, v. 118, n. 9, p. 2309–2309, jan. 1996.

DA SILVA SANTOS, C.; BRODSKYN, C. I. The Role of CD4 and CD8 $\alpha$  T Cells in Human Cutaneous Leishmaniasis. **Frontiers in Public Health**, v. 2, 29 set. 2014.

DAI, S. Crystal Structure of HLA-DP2. 14 abr. 2010.

DAI, S.; KAPPLER, J. W. Crystal structure of HLA DR52c. 5 ago. 2008.

DAY, M. J. Immunoglobulin G subclass distribution in canine leishmaniosis: A review and analysis of pitfalls in interpretation. **Veterinary Parasitology**, v. 147, n. 1–2, p. 2–8, jun. 2007.

DE OLIVEIRA MENDES, C. et al. IgG1/IgG2 antibody dichotomy in sera of vaccinated or naturally infected dogs with visceral leishmaniosis. **Vaccine**, v. 21, n. 19–20, p. 2589–2597, jun. 2003.

DIAS, R.; DE AZEVEDO JR., W. Molecular Docking Algorithms. **Current Drug Targets**, v. 9, n. 12, p. 1040–1047, 1 dez. 2008.

DIKHIT, M. R. et al. Identification of Potential MHC Class-II-Restricted Epitopes Derived from *Leishmania donovani* Antigens by Reverse Vaccinology and Evaluation of Their CD4+ T-Cell Responsiveness against Visceral Leishmaniasis. **Frontiers in Immunology**, v. 8, 14 dez. 2017.

DORMITZER, P. R.; ULMER, J. B.; RAPPUOLI, R. Structure-based antigen design: a strategy for next generation vaccines. **Trends in Biotechnology**, v. 26, n. 12, p. 659–667, dez. 2008a.

DORMITZER, P. R.; ULMER, J. B.; RAPPUOLI, R. Structure-based antigen design: a strategy for next generation vaccines. **Trends in Biotechnology**, v. 26, n. 12, p. 659–667, 1 dez. 2008b.

DOSZTÁNYI, Z. et al. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. **Journal of Molecular Biology**, v. 347, n. 4, p. 827–839, abr. 2005.

DUMONTEIL, E. Vaccine development against *Trypanosoma cruzi* and *Leishmania* species in the post-genomic era. **Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases**, v. 9, n. 6, p. 1075–1082, dez. 2009.

DUNKER, A. K. et al. Intrinsic disorder and protein function. **Biochemistry**, v. 41, n. 21, p. 6573–6582, 28 maio 2002.

DUNKER, A. K. et al. Flexible nets. The roles of intrinsic disorder in protein interaction networks. **The FEBS journal**, v. 272, n. 20, p. 5129–5148, out. 2005.

DUNKER, A. K.; OBRADOVIC, Z. The protein trinity—linking function and disorder. **Nature Biotechnology**, v. 19, n. 9, p. 805–806, set. 2001.

DYSON, H. J.; WRIGHT, P. E. Intrinsically unstructured proteins and their functions. **Nature Reviews. Molecular Cell Biology**, v. 6, n. 3, p. 197–208, mar. 2005.

E SILVA, R. DE F. et al. Combination of In Silico Methods in the Search for Potential CD4+ and CD8+ T Cell Epitopes in the Proteome of *Leishmania braziliensis*. **Frontiers in Immunology**, v. 7, 29 ago. 2016.

EL-MANZALAWY, Y.; DOBBS, D.; HONAVAR, V. Predicting flexible length linear B-cell epitopes. **Computational Systems Bioinformatics. Computational Systems Bioinformatics Conference**, v. 7, p. 121–132, 2008a.

EL-MANZALAWY, Y.; DOBBS, D.; HONAVAR, V. Predicting linear B-cell epitopes using string kernels. **Journal of molecular recognition: JMR**, v. 21, n. 4, p. 243–255, ago. 2008b.

EMANUELSSON, O. et al. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. **Journal of Molecular Biology**, v. 300, n. 4, p. 1005–1016, 21 jul. 2000.

FISER, A. Template-Based Protein Structure Modeling. **Methods in molecular biology (Clifton, N.J.)**, v. 673, p. 73–94, 2010.

FOUCAULT, M. et al. UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. **Proteins: Structure, Function, and Bioinformatics**, p. NA-NA, 2009.

GHOSH, P. et al. THE STRUCTURE OF AN INTERMEDIATE IN CLASS II MHC MATURATION: CLIP BOUND TO HLA-DR3. 27 maio 1998.

GIBAS, C.; JAMBECK, P. **Developing bioinformatics computer skills**. 1st ed ed. Beijing : Cambridge: O'Reilly, 2001.

GILBERT, D. Sequence File Format Conversion with Command-Line Readseq. In: BAXEVANIS, A. D. et al. (Eds.). . **Current Protocols in Bioinformatics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2003.

GILLESPIE, G. M. A. et al. Strong TCR Conservation and Altered T Cell Cross-Reactivity Characterize a B\*57-Restricted Immune Response in HIV-1 Infection. **The Journal of Immunology**, v. 177, n. 6, p. 3893–3902, 15 set. 2006a.

GILLESPIE, G. M. A. et al. Strong TCR conservation and altered T cell cross-reactivity characterize a B\*57-restricted immune response in HIV-1 infection. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 177, n. 6, p. 3893–3902, 15 set. 2006b.

GILLESPIE, G. M. A. et al. Crystal Structure of HLA-B5703 and HIV-1 peptide. 3 out. 2006c.

GIULIANI, M. M. et al. A universal vaccine for serogroup B meningococcus. **Proceedings of the National Academy of Sciences**, v. 103, n. 29, p. 10834–10839, 18 jul. 2006.

GIUNCHETTI, R. C. et al. Immunogenicity of a killed Leishmania vaccine with saponin adjuvant in dogs. **Vaccine**, v. 25, n. 44, p. 7674–7686, 1 nov. 2007.

GIUNCHETTI, R. C. et al. A killed Leishmania vaccine with sand fly saliva extract and saponin adjuvant displays immunogenicity in dogs. **Vaccine**, v. 26, n. 5, p. 623–638, 30 jan. 2008.

GOPAL, S. et al. Homology-based annotation yields 1,042 new candidate genes in the *Drosophila melanogaster* genome. **Nature Genetics**, v. 27, n. 3, p. 337–340, mar. 2001.

GRAS, S.; CHABROL, E.; ROSSJOHN, J. Crystal Structure of HLA B\*3501-IPS in complex with a Delta-Beta TCR, clone 12 TCR. 10 dez. 2014.

GRAS, S.; TWIST, K.-A.; ROSSJOHN, J. Crystal Structure of HLA B\*0801 in complex with ELR\_MYM, ELRRKMMYM. 4 fev. 2015.

GRAS, S.; WILMANN, P. G.; ROSSJOHN, J. Crystal Structure of C12C TCR-HA B2705-KK10. 20 mar. 2013.

GU, J.; BOURNE, P. E. (EDS.). **Structural bioinformatics**. 2nd ed ed. Hoboken, N.J: Wiley-Blackwell, 2009.

HAHN, M. et al. TCR/HLA-DR2b/MBP-peptide complex. 3 maio 2005.

HASTE ANDERSEN, P.; NIELSEN, M.; LUND, O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. **Protein Science**, v. 15, n. 11, p. 2558–2567, nov. 2006.

HENDERSON, K. N. et al. Structure of the MHC class II molecule HLA-DQ8 bound with a deamidated gluten peptide. 4 set. 2007.

HIBBERT, K. M.; ROSSJOHN, J.; GRAS, S. HLA-B\*1801 in complex with a self-peptide, DELEIKAY. 8 abr. 2015.

HOLLAND, C. J.; COLE, D. K.; GODKIN, A. Re-Directing CD4+ T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core is Not Enough. **Frontiers in Immunology**, v. 4, 2013.

HOPP, T. P.; WOODS, K. R. Prediction of protein antigenic determinants from amino acid sequences. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 6, p. 3824–3828, jun. 1981.

HORTON, P. et al. WoLF PSORT: protein localization predictor. **Nucleic Acids Research**, v. 35, n. Web Server issue, p. W585–W587, jul. 2007.

HORVÁTH, I. et al. Tyrosine aminotransferase induction in normal and tumor-bearing chickens. **International Journal of Cancer**, v. 16, n. 6, p. 897–904, 15 dez. 1975.

KACZANOWSKI, S.; ZIELENKIEWICZ, P. Why similar protein sequences encode similar three-dimensional structures? **Theoretical Chemistry Accounts**, v. 125, n. 3–6, p. 643–650, mar. 2010.

KÄLLBERG, M. et al. Template-based protein structure modeling using the RaptorX web server. **Nature Protocols**, v. 7, n. 8, p. 1511–1522, 19 jul. 2012.

KARPLUS, K. et al. Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. **Proteins**, v. 53 Suppl 6, p. 491–496, 2003.

KEEDY, D. A. et al. The other 90% of the protein: assessment beyond the Calphas for CASP8 template-based and high-accuracy models. **Proteins**, v. 77 Suppl 9, p. 29–49, 2009.

KIM, C.-Y. et al. Crystal structure of HLA-DQ2 complexed with deamidated gliadin peptide. 2 mar. 2004.

KRIEGER, E. et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. **Proteins: Structure, Function, and Bioinformatics**, v. 77, n. S9, p. 114–122, 2009.

KRINGELUM, J. V. et al. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. **PLoS Computational Biology**, v. 8, n. 12, p. e1002829, 27 dez. 2012.

KRINGELUM, J. V. et al. Structural analysis of B-cell epitopes in antibody:protein complexes. **Molecular Immunology**, v. 53, n. 1–2, p. 24–34, jan. 2013.

KUHDARI, P. et al. Meningococcal B vaccination: real-world experience and future perspectives. **Pathogens and Global Health**, v. 110, n. 4–5, p. 148–156, 3 jul. 2016.

KUSANO, S. et al. Crystal structure of HLA-DP5 in complex with Cry j 1-derived peptide (residues 214-222). 23 jul. 2014.

LANG, H. et al. X-ray crystal structure of HLA-DRA1\*0101/DRB5\*0101 complexed with a peptide from Epstein Barr Virus DNA polymerase. 3 out. 2002.

LARSEN, J. E. P.; LUND, O.; NIELSEN, M. Improved method for predicting linear B-cell epitopes. **Immunome Research**, v. 2, p. 2, 24 abr. 2006.

LARSEN, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. **BMC bioinformatics**, v. 8, p. 424, 31 out. 2007.

LEAVER-FAY, A. et al. Rosetta3. In: **Methods in Enzymology**. [s.l.] Elsevier, 2011. v. 487p. 545–574.

LIANG, S. et al. A simple reference state makes a significant improvement in near-native selections from structurally refined docking decoys. **Proteins: Structure, Function, and Bioinformatics**, v. 69, n. 2, p. 244–253, 1 nov. 2007.

LIANG, S. et al. Prediction of antigenic epitopes on protein surfaces by consensus scoring. **BMC Bioinformatics**, v. 10, n. 1, p. 302, 2009.

LINDING, R. GlobPlot: exploring protein sequences for globularity and disorder. **Nucleic Acids Research**, v. 31, n. 13, p. 3701–3708, 1 jul. 2003.

LINDING, R. et al. Protein Disorder Prediction. **Structure**, v. 11, n. 11, p. 1453–1459, nov. 2003.

LIU, J. et al. Crystal Structure of HLA-A\*2402 Complexed with a Newly Identified Peptide from 2009 H1N1 PA (649-658). 10 out. 2012.

LIU, J. et al. Crystal Structure of HLA-A\*1101 in complex with H1-22, an influenza A(H1N1) virus epitope. 8 out. 2014.

LIU, Y. C.; ROSSJOHN, J.; GRAS, S. crystal structure of HLA B\*3508LPEP-P5Ala , peptide mutant P5-ala. 22 fev. 2012.

LOVELL, S. C. et al. Structure validation by C $\alpha$  geometry:  $\phi, \psi$  and C $\beta$  deviation. **Proteins: Structure, Function, and Bioinformatics**, v. 50, n. 3, p. 437–450, 8 jan. 2003.

LYNCH, W. P.; RISEMAN, V. M.; BRETSCHER, A. Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. **The Journal of Biological Chemistry**, v. 262, n. 15, p. 7429–7437, 25 maio 1987.

MACDONALD, W. A. et al. Crystal Structure of HLA B\*4405 in complex with EEYLQAFTY a self peptide from the ABCD3 protein. 22 dez. 2009.

MAIONE, D. Identification of a Universal Group B Streptococcus Vaccine by Multiple Genome Screen. **Science**, v. 309, n. 5731, p. 148–150, 1 jul. 2005.

MARTÍ-RENOM, M. A. et al. Comparative Protein Structure Modeling of Genes and Genomes. **Annual Review of Biophysics and Biomolecular Structure**, v. 29, n. 1, p. 291–325, jun. 2000.

MATSUMOTO, R. et al. Packaging of Proteases and Proteoglycans in the Granules of Mast Cells and Other Hematopoietic Cells A CLUSTER OF HISTIDINES ON MOUSE MAST CELL PROTEASE 7 REGULATES ITS BINDING TO HEPARIN SERGLYCIN PROTEOGLYCANS. **Journal of Biological Chemistry**, v. 270, n. 33, p. 19524–19531, 18 ago. 1995.

MODABBER, F. Vaccines against leishmaniasis. **Annals of Tropical Medicine and Parasitology**, v. 89 Suppl 1, p. 83–88, dez. 1995.

MOHAN, A. et al. Analysis of Molecular Recognition Features (MoRFs). **Journal of Molecular Biology**, v. 362, n. 5, p. 1043–1059, out. 2006.

MOHAN, A. et al. Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. **Molecular bioSystems**, v. 4, n. 4, p. 328–340, abr. 2008.

MONTIGIANI, S. et al. Genomic approach for analysis of surface proteins in Chlamydia pneumoniae. **Infection and Immunity**, v. 70, n. 1, p. 368–379, jan. 2002.

MORENO, J. et al. Use of a LiESP/QA-21 Vaccine (CaniLeish) Stimulates an Appropriate Th1-Dominated Cell-Mediated Immune Response in Dogs. **PLoS Neglected Tropical Diseases**, v. 6, n. 6, 19 jun. 2012.

MUKONYORA, M. A Review of Important Discontinuous B-Cell Epitope Prediction Tools. **Journal of Clinical & Cellular Immunology**, v. 6, n. 5, 9 out. 2015.

MURPHY, K. et al. **Janeway's immunobiology**. 7th ed ed. New York: Garland Science, 2008.

NAZ, A. et al. Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: A reverse vaccinology based approach. **Infection, Genetics and Evolution**, v. 32, p. 280–291, jun. 2015.

NEEFJES, J. et al. Towards a systems understanding of MHC class I and MHC class II antigen presentation. **Nature Reviews Immunology**, 11 nov. 2011.

NIELSEN, M. et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. **Protein Science: A Publication of the Protein Society**, v. 12, n. 5, p. 1007–1017, maio 2003.

NIELSEN, M. et al. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. **Bioinformatics (Oxford, England)**, v. 20, n. 9, p. 1388–1397, 12 jun. 2004.

NIELSEN, M. et al. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. **Immunogenetics**, v. 57, n. 1–2, p. 33–41, abr. 2005.

NIELSEN, M.; LUND, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. **BMC bioinformatics**, v. 10, p. 296, 18 set. 2009.

NIELSEN, M.; LUNDEGAARD, C.; LUND, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. **BMC bioinformatics**, v. 8, p. 238, 4 jul. 2007.

NIU, L. et al. Structure of HLA-A68 complexed with an HIV derived peptide. 16 out. 2013.

NOVAIS, F. O.; SCOTT, P. CD8+ T cells in cutaneous leishmaniasis: the good, the bad, and the ugly. **Seminars in Immunopathology**, v. 37, n. 3, p. 251–259, maio 2015.

OFRAN, Y.; SCHLESSINGER, A.; ROST, B. Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 181, n. 9, p. 6230–6235, 1 nov. 2008.

OGHUMU, S. et al. Role of chemokines in regulation of immunity against leishmaniasis. **Experimental Parasitology**, v. 126, n. 3, p. 389–396, nov. 2010.

OLDFIELD, C. J. et al. Coupled Folding and Binding with  $\alpha$ -Helix-Forming Molecular Recognition Elements<sup>†</sup>. **Biochemistry**, v. 44, n. 37, p. 12454–12470, set. 2005.

PARRY, C. S. The structure of HLA-DRA, DRB3\*0101 (DR52a) with bound platelet integrin peptide associated with fetal and neonatal alloimmune thrombocytopenia. 24 jul. 2007.

PENG, J.; XU, J. Raptorx: Exploiting structure information for protein alignment by statistical inference. **Proteins: Structure, Function, and Bioinformatics**, v. 79, n. S10, p. 161–171, 2011a.

PENG, J.; XU, J. A multiple-template approach to protein threading. **Proteins: Structure, Function, and Bioinformatics**, v. 79, n. 6, p. 1930–1939, jun. 2011b.

PENG, J.; XU, J. A multiple-template approach to protein threading. **Proteins: Structure, Function, and Bioinformatics**, v. 79, n. 6, p. 1930–1939, jun. 2011c.



PENG, K. et al. Length-dependent prediction of protein intrinsic disorder. **BMC Bioinformatics**, v. 7, p. 208–208, 2006.

PETERS, B. et al. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 171, n. 4, p. 1741–1749, 15 ago. 2003.

PETERSEN, J.; ROSSJOHN, J. Crystal Structure of HLA-B7 with the HCMV pp65 peptide RIPHERNGFTVL. 21 nov. 2012.

PETTERSEN, E. F. et al. UCSF Chimera?A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, out. 2004.

PIZZA, M. et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. **Science (New York, N.Y.)**, v. 287, n. 5459, p. 1816–1820, 10 mar. 2000.

PONOMARENKO, J. et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. **BMC Bioinformatics**, v. 9, p. 514, 2 dez. 2008.

RAMACHANDRAN, G. N.; SASISEKHARAN, V. Conformation of Polypeptides and Proteins\*\*The literature survey for this review was completed in September 1967, with the journals which were then available in Madras and the preprinta which the authors had received.††By the authors' request, the publishers have left certain matters of usage and spelling in the form in which they wrote them. In: ANFINSEN, C. B. et al. (Eds.). . **Advances in Protein Chemistry**. [s.l.] Academic Press, 1968. v. 23p. 283–437.

RAVEH, B.; LONDON, N.; SCHUELER-FURMAN, O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. **Proteins: Structure, Function, and Bioinformatics**, p. NA-NA, 2010.

REIS, A. B. et al. Isotype patterns of immunoglobulins: Hallmarks for clinical status and tissue parasite density in brazilian dogs naturally infected by Leishmania (Leishmania) chagasi. **Veterinary Immunology and Immunopathology**, v. 112, n. 3–4, p. 102–116, ago. 2006.

RIGDEN, D. J. (ED.). **From protein structure to function with bioinformatics**. Dordrecht: Springer, 2009.

RIZKALLAH, P. J. et al. T cell response to a HIV reverse transcriptase epitope presented by the protective allele HLA-B\*51:01. 28 maio 2014.

RODER, G. A. et al. Crystal Structure of a SARS Corona Virus Derived Peptide Bound to the Human Major Histocompatibility Complex Class I molecule HLA-B\*1501. 26 fev. 2008.

RODRIGUEZ, V.; CENTENO, M.; ULRICH, M. The IgG isotypes of specific antibodies in patients with American cutaneous leishmaniasis; relationship to the cell-mediated immune response. **Parasite Immunology**, v. 18, n. 7, p. 341–345, jul. 1996.

ROSSJOHN, J.; GRAS, S. Crystal Structure of HLA A\*0101 in complex with NP44, an 9-mer influenza epitope. 25 dez. 2013.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725–738, abr. 2010.

RUBINSTEIN, N. D. et al. Computational characterization of B-cell epitopes. **Molecular Immunology**, v. 45, n. 12, p. 3477–3489, jul. 2008.

RUY, P. et al. Intrinsically disordered proteins (IDPs) in trypanosomatids. **BMC Genomics**, v. 15, n. 1, p. 1100, 2014.

ŠALI, A. et al. Evaluation of comparative protein modeling by MODELLER. **Proteins: Structure, Function, and Genetics**, v. 23, n. 3, p. 318–326, nov. 1995.

SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of Molecular Biology**, v. 234, n. 3, p. 779–815, 5 dez. 1993.

ŠALI, A.; OVERINGTON, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. **Protein Science**, v. 3, n. 9, p. 1582–1596, set. 1994.

SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the Amber biomolecular simulation package: Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 3, n. 2, p. 198–210, mar. 2013.

SCHRÖDINGER, LLC. **The PyMOL Molecular Graphics System, Version 1.8.** nov. 2015.

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

SETHI, D. K.; WUCHERPFENNIG, K. W. Structure of Autoimmune TCR Hy.1B11 in complex with HLA-DQ1 and MBP 85-99. 22 dez. 2010.

SETTE, A. et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. **The Journal of Immunology**, v. 153, n. 12, p. 5586, 15 dez. 1994.

SHEN, M.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein Science**, v. 15, n. 11, p. 2507–2524, nov. 2006.

SHI, J.; BLUNDELL, T. L.; MIZUGUCHI, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. **Journal of Molecular Biology**, v. 310, n. 1, p. 243–257, 29 jun. 2001.

SHIMIZU, A. et al. HLA-A24 in complex with HIV-1 Nef134-8(wt). 6 maio 2015.

SIEBOLD, C. et al. Crystal structure of HLA-DQ0602 in complex with a hypocretin peptide. 5 fev. 2004.

SIPPL, M. J. Recognition of errors in three-dimensional structures of proteins. **Proteins**, v. 17, n. 4, p. 355–362, dez. 1993.

SMITH, K. J. et al. CRYSTAL STRUCTURE OF HLA-DR2 (DRA\*0101,DRB1\*1501) COMPLEXED WITH A PEPTIDE FROM HUMAN MYELIN BASIC PROTEIN. [s.d.].

SÖDING, J. Protein homology detection by HMM-HMM comparison. **Bioinformatics (Oxford, England)**, v. 21, n. 7, p. 951–960, 1 abr. 2005.

SPAHN, C. M. T. et al. Structure of the 80S Ribosome from *Saccharomyces cerevisiae*—tRNA-Ribosome and Subunit-Subunit Interactions. **Cell**, v. 107, n. 3, p. 373–386, 2 nov. 2001.

STAHL, H. D. et al. Differential antibody screening of cloned *Plasmodium falciparum* sequences expressed in *Escherichia coli*: procedure for isolation of defined antigens and analysis of human antisera. **Proceedings of the National Academy of Sciences of the United States of America**, v. 81, n. 8, p. 2456–2460, abr. 1984.

SUN, M. et al. An Nt-acetylated peptide complexed with HLA-B\*3901. 23 jul. 2014.

SWEREDOSKI, M. J.; BALDI, P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. **Bioinformatics (Oxford, England)**, v. 24, n. 12, p. 1459–1460, 15 jun. 2008.

TALUKDAR, S. et al. Identification of Potential Vaccine Candidates Against *Streptococcus pneumoniae* by Reverse Vaccinology Approach. **Applied Biochemistry and Biotechnology**, v. 172, n. 6, p. 3026–3041, mar. 2014.

THEODOSSIS, A. et al. HLA-B\*44:03 in complex with Epstein-Barr virus BZLF1-derived peptide (residues 169-180). 26 jun. 2013.

THEODOSSIS, A.; ELY, L. K.; ROSSJOHN, J. Crystal structure of HLA-B\*4402 in complex with the F3A mutant of a self-peptide derived from DPA\*0201. 16 mar. 2010.

THOMAS, S.; DILBAROVA, R.; RAPPUOLI, R. Future Challenges for Vaccinologists. In: THOMAS, S. (Ed.). . **Vaccine Design**. New York, NY: Springer New York, 2016. v. 1403p. 41–55.

THORNTON, J. M. et al. Location of “continuous” antigenic determinants in the protruding regions of proteins. **The EMBO Journal**, v. 5, n. 2, p. 409–413, fev. 1986.

ULRICH, M. et al. Differing antibody IgG isotypes in the polar forms of leprosy and cutaneous leishmaniasis characterized by antigen-specific T cell anergy. **Clinical and Experimental Immunology**, v. 100, n. 1, p. 54–58, abr. 1995.

VAN DER LEE, R. et al. Classification of Intrinsically Disordered Regions and Proteins. **Chemical Reviews**, v. 114, n. 13, p. 6589–6631, 9 jul. 2014.

VAN REGENMORTEL, M. H. V. Which structural features determine protein antigenicity? **Trends in Biochemical Sciences**, v. 11, n. 1, p. 36–39, jan. 1986.

VERLI, H. **Bioinformática : da biologia à flexibilidade molecular**. [s.l.] Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.

VON HEIJNE, G. A new method for predicting signal sequence cleavage sites. **Nucleic Acids Research**, v. 14, n. 11, p. 4683–4690, 11 jun. 1986.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]**, v. 54, p. 5.6.1-5.6.37, 20 jun. 2016.

WESTHOF, E. et al. Correlation between segmental mobility and the location of antigenic determinants in proteins. **Nature**, v. 311, n. 5982, p. 123–126, 13 set. 1984.

WIZEMANN, T. M. et al. Use of a Whole Genome Approach To Identify Vaccine Molecules Affording Protection against Streptococcus pneumoniae Infection. **Infection and Immunity**, v. 69, n. 3, p. 1593–1598, 1 mar. 2001.

WRIGHT, P. E.; DYSON, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. **Journal of Molecular Biology**, v. 293, n. 2, p. 321–331, out. 1999.

WRIGHT, P. E.; DYSON, H. J. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. **Nature reviews. Molecular cell biology**, v. 16, n. 1, p. 18–29, jan. 2015.

WU, S.; ZHANG, Y. LOMETS: A local meta-threading-server for protein structure prediction. **Nucleic Acids Research**, v. 35, n. 10, p. 3375–3382, maio 2007a.

WU, S.; ZHANG, Y. LOMETS: A local meta-threading-server for protein structure prediction. **Nucleic Acids Research**, v. 35, n. 10, p. 3375–3382, maio 2007b.

XIN, L.; LI, Y.; SOONG, L. Role of Interleukin-1 in Activating the CD11chigh CD45RB Dendritic Cell Subset and Priming Leishmania amazonensis- Specific CD4+ T Cells In Vitro and In Vivo. **Infection and Immunity**, v. 75, n. 10, p. 5018–5026, 1 out. 2007.

XU, L. Z. et al. Ligand specificity of brain lipid-binding protein. **The Journal of Biological Chemistry**, v. 271, n. 40, p. 24711–24719, 4 out. 1996.

XU, Y.; XU, D. Protein threading using PROSPECT: design and evaluation. **Proteins**, v. 40, n. 3, p. 343–354, 15 ago. 2000.

YAN, R. et al. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. **Scientific Reports**, v. 3, p. 2619, 2013.

YANG, J. et al. The I-TASSER Suite: protein structure and function prediction. **Nature Methods**, v. 12, n. 1, p. 7, jan. 2015.

ZHANG, S. et al. HIV RT derived peptide complexed to HLA-A\*0301. 29 fev. 2012.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, v. 9, n. 1, p. 40, 2008.

ZHOU, H.; ZHOU, Y. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. **Proteins**, v. 55, n. 4, p. 1005–1013, 1 jun. 2004.

ZHOU, H.; ZHOU, Y. Fold Recognition by Combining Sequence Profiles Derived From Evolution and From Depth-Dependent Structural Alignment of Fragments. **Proteins**, v. 58, n. 2, p. 321–328, 1 fev. 2005.

## Apêndice A - Seleção dos moldes

Comando para formatação das sequências usando o programa *readseq*:

```
java -jar readseq.jar -inform fasta -f fasta -o pdb_seqres_03-12-17.fasta  
pdb_seqres_03-12-17.txt
```

-inform = formato de arquivo de entrada

-f= formato do arquivo de saída

-o= Arquivo de saída

Essas sequências foram então indexadas utilizando o programa *makeblastdb*. A linha de comando usada foi a seguinte:

```
makeblastdb -in pdb_seqres_03-12-17.txt -input_type fasta -dbtype prot
```

-in = arquivo de entrada

-input\_type = tipo de arquivo de entrada

-dbtype = Informação biológica utilizada

O pacote BLAST foi então executado utilizando as sequências indexadas do PDB e as proteínas de *Leishmania* selecionadas para esse trabalho, a seguinte linha de comando foi usada:

```
blastp -query seq_leishmania_structural.faa -db ./blastdb/pdb_seqres_03-12-17.txt  
-out seq_leishmania_structuralxpdb_blastp-e5-des5-alig5-outfmt7.txt -evalue  
0.00001 -num_descriptions 5 -num_alignments 5 -outfmt "7 qacc sacc evalue  
qstart qend sstart send pident qcovs"
```

-query= sequência em formato FASTA a ser analisada

-db= banco criado com o makeblastdb  
-e-value= e-value máximo considerado  
-out= nome para o arquivo de saída  
-num\_descriptions= número de descrições  
-num\_alignments= número de alinhamentos  
-outfmt = formato do arquivo de saída

## Apêndice B - Modelagem por Homologia usando Modeller

Para a utilização do Modeller inicialmente foi necessário converter manualmente o arquivo fasta contendo as sequências de aminoácidos para o arquivo, do tipo PIR, capaz de ser reconhecido pelo Modeller, de extensão “.ali”.

### Exemplo de arquivo fasta

```
>LinJ.18.1500 | organism=Leishmania_infantum | product=P-type
H -ATPase, putative | location=LinJ.18:656653-659577(+) |
length=974

MSSKKYELDAAAFEDKPESHSDAEMTPQKPQRRQSVLSKAVSEHDERATGPATDLLPPSKGL
TTEEAEEELLKKYGRNELPEKKTPSWLIYVRGLWGPMPAALWIAIIIEFALENWPDGAILFAI
QIANATIGWYETIKAGDAVAALKNSLKPTATVYRDSKWQQIDAAVLVPGDLVKLASGSAVPA
DCSINEGVIDVDEAALTGESLPVTMGPEHMPKMGSNVVRGEVEGTVQYTGSLTFFGKTAALL
QSVESDLGNIHVILRRVMLALCAISFILCMCCFIYLLARFYETFRHALQFAVVVLVVSIPIA
LEIVVTTTLAVGSKHLSKHKIIVTKLSAIEMSGVNMMLCSDKTGTLTLNKMEIQEQCFTFEE
GNDLKSTLVLAALAAKWREPPRDALDTMVLGAADLDECDNYQQLNFVFPDPTTKRТАATLVD
RRSGEKFDVTKGAPHVILQMVYNQDEINDEVVDIIDSLAARGVRCLSVAKTDQQGRWHMAGI
LTFDPPRPDPTKDTIRRSKEYGVDVKMITGDHLLIAKEMCRMLDLDPNILTADKLPQIKDAN
DLPEDLGEKYGDMMLSVGGFAQVFPEHKFMIVETLRQRGYTCAMTGDGVNDAPALKRADVGI
AVHGATDAARAAADMVLTEPGLSVVVEAMLVSREVFQRMLSFLTYRISATLQLVCFFFIACF
SLTPKAYGSVDPNFQFFHLPVLMFMLITLLNDGCLMTIGYDHVIPSERPQKWNLPVVVFSAS
ILAAVACGSSMLLLWIGLEGYSSQYYENSWFHRLGLAQLPQGKLVТMМYLKISISDFLTLFS
SRTGGHFFFYVPPSPILFCGAIISLLVSTMAASFVHKSRPDNVLTEGLAWGQTNAEKLLPLW
VWIYCIVWWFVQDVVKVLAHICMDAVDLFGCVSDASGSGPIKPYSDDMKVNGFEPVKKPAEK
STEKALNLSVSSASHKALEGLREDTHSPIEEARPVNVYVSCDKK
```

### Exemplo de arquivo PIR”

```
>P1;LinJ.18.1500
sequence:LinJ.18.1500:::::0.00: 0.00
```



```

MSSKKYELDAAAFEDKPESHSDAEMTPQKPQRRQSVLSKAVSEHDERATGPATDLLPPSKGL
TTEEAEELLKKYGRNELPEKKTPSWLIYVRGLWGPMPAALWIAIIIEFALENWPDGAILFAI
QIANATIGWYETIKAGDAVAALKNSLKPTATVYRDSKWQQIDA AVLVPGLVKLASGS AVPA
DCSINEGVIDVDEAALTGESLPVTMGPEHMPKMGSNVVRGEVEGTVQYTGSLTFFGKTAALL
QSVESDLGNIHVILRRVMLALCAISFILCMCCFIYLLARFYETFRHALQFAVVVLVVSIPIA
LEIVVTTTLAVGSKHLSKHKIIVTKLSAIEMSGVNMMLCSDKTGTLTLNKMEIQEQCFTFEE
GNDLKSTLVLAALAAKWREPPRDALDTMVLGAADLDECDNYQQLNFVFPDPTTKRRTAATLVD
RRSGEKFDVTKGAPHVILQMVYNQDEINDEVVDIIDL AARGVRCLSVAKTDQQGRWHMAGI
LTFLDPPRPDTKD TIRRSKEYGVDVKMITGDHLLIAKEMCRMLDLDPNILTADKLPQIKDAN
DLPEDLGEKYGDMMLSVGGFAQVFPEHKFMIVETLRQRGYTCAMTGDGVNDAPALKRADVGI
AVHGATDAARAAADMVLTEPGLSVVVEAMLVSREVFQRMLSFLTYRISATLQLVCFFFIACF
SLTPKAYGSVDPNFQFFHLPVLMFMLITLLNDGCLMTIGYDHVIPSERPQKWNLPVVVFSAS
ILAAVACGSSMLLLWIGLEGYSSQYYENSWFHRLGLAQLPQGKLV TMMYLKISISDFLTLFS
SRTGGHFFFYVPPSPILFCGAIISLLVSTMAASF WHKSRPDNVLTEGLAWGQTNAEKLLPLW
VWIYCIVWWFVQDVVKVLAHICMDAVDLFGCVSDASGSGPIKPYSDDMKVNGFEPVKKPAEK
STEKALNLSVSSASHKALEGLREDTHSPIEEARPVNVVYVSCDKK*

```

Em seguida o alinhamento entre o molde e a proteína alvo foi realizado utilizando o arquivo “align2d.py” disponível no tutorial oficial do Modeller.

```

from modeller import *

env = environ()
aln = alignment(env)
mdl = model(env, file='lbdm', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='lbdmA', atom_files='lbdm.pdb')
aln.append(file='TvLDH.ali', align_codes='TvLDH')
aln.align2d()
aln.write(file='TvLDH-lbdmA.ali', alignment_format='PIR')
aln.write(file='TvLDH-lbdmA.pap', alignment_format='PAP')

```

De posse do alinhamento, os modelos estruturais da proteína alvo foram calculados através da auto-modelagem do Modeller, utilizando o arquivo model-single.py também disponível no tutorial oficial do Modeller.

```

from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

```

```
env = environ()
a = automodel(env, alnfile='TvLDH-lbdaA.ali',
              knowns='lbdaA', sequence='TvLDH',
              assess_methods=(assess.DOPE,
                              #soap_protein_od.Scorer(),
                              assess.GA341))
a.starting_model = 1
a.ending_model = 5
a.make()
```

## Apêndice C - Comandos utilizados nos preditores de epítomos de Células B conformacionais

### Discotope

```
./discotope -f arquivo_de_entrada -chain A > arquivo_de_saida  
  
-f= arquivo de entrada  
-chain= cadeia a ser analisada  
> = direcionador para arquivo de saída
```

Observação: Em alguns casos foi necessário atribuir o identificador da cadeia usando o *chimera* para utilização do Discotope, uma vez que era exigida a identificação da cadeia na linha de comando acima.

### Ellipro

```
java -jar ElliPro.jar -f arquivo_de_entrada -o arquivo_de_saida  
-f= arquivo de entrada  
-o= arquivo de saída
```

Observação: para esse programa é necessário ter o JAVA instalado

## Apêndice D - Execução do *pipeline* de predição de desordem proteica

As linhas de comando abaixo foram utilizadas. A primeira para a criação de um banco de dados relacional e a segunda para execução do *pipeline* e inserção dos dados gerados no banco inicialmente criado:

```
create_db_iup-pipeline-optimus.pl -d nome_do_banco -h banco -u user
```

-d: nome do banco de dados

-h: servidor onde está hospedado o banco de dados

-u: usuário com permissão de alteração do banco de dados

```
iup_pipeline_V7-optimus.pl -h banco -d idp_epítopo_11_2017 -u user
-f especie -fdb /storage3/BANCOS/nr/nr -o lin_jpcm5 -ls 100 -met F
-func T -a T -fa F -bl T -in entrada.fasta -ld 40 -nt 65 -mts 50 -
ev 0.000001 2
```

-h: servidor de banco de dados

-d: banco de dados

-u: usuário com privilégio de edição do banco de dados

-f: nome base para nomenclatura de arquivos  
que serão gerados

-fdb: caminho para o banco de dados não redundante

-o: nome base do nome dos arquivos resultantes

-ls: tamanho mínimo das sequências que serão consideradas na  
análise

-met: verifica se as proteínas iniciam com Metionina

-func: classifica as proteínas analisadas em preditas e hipotéticas

-a: verifica a existência de erros de anotação

-fa: anotação funcional

-bl: realizar busca por similaridade de sequências

-in: proteoma a ser analisado

-ld: tamanho mínimo a considerar para cada disordem

-nt: número de processadores para a busca por similaridade de  
sequências

-mts: número de alinhamentos e descrições nos resultados a busca  
por similaridade de sequências

-ev: e-value mínimo empregado na a busca por similaridade de  
sequências

## Apêndice E – Script em python para a troca dos peptídeos cocrystalizados pelos peptídeos das proteínas em estudo

Linha para a execução:

```
pymol -qc ./mutagenesis_peptide.py <arquivo de entrada> <nome do
arquivo de entrada> <cadeia da proteína a ser trocada> <sequência de
aminoácidos a ser utilizada na troca>
```

```
#!/usr/bin/env python2.7

from pymol import cmd #importar biblioteca cmd
import sys #importar biblioteca sys

aa_dict = {'V':'VAL', 'I':'ILE', 'L':'LEU', 'E':'GLU', 'Q':'GLN',
'D':'ASP', 'N':'ASN', 'H':'HIS', 'W':'TRP', 'F':'PHE', 'Y':'TYR',
'R':'ARG', 'K':'LYS', 'S':'SER', 'T':'THR', 'M':'MET', 'A':'ALA',
'G':'GLY', 'P':'PRO', 'C':'CYS'} #biblioteca para conversão de
identificacao de uma letra de residuo pelo indentificador de 3
letras

x = 1 #iniciar contador

while x <= 9: #substituir somente 9 aminoacidos (caso de MHCI, MHC
II mudar para 15)

y = x + 5 #considerar a partir do quinto parametro da linha de
comando o residuo a ser usado na troca

pdb = (sys.argv[3]) #argumento indicando qual pdb usar
name = (sys.argv[4]) #argumento indicando identificador do pdb
selection = (str(sys.argv[5])+("/")+str(x)+("/")) #argumento
indicando quais aminoacidos a serem subsituídos
mutant = (aa_dict[sys.argv[y]]) #argumento indicando quais
aminoacidos serao utilizados na substituicao

cmd.wizard("mutagenesis") #carregar a ferramenta mutagenesis do
pymol

cmd.load(pdb) #carregar pdb

cmd.refresh_wizard()

cmd.get_wizard().do_select(selection)# selecionar aminoacidos a
```

```
serem substituidos

cmd.get_wizard().set_mode(mutant) #realizar troca

cmd.get_wizard().apply() #aplicar mudanca

cmd.set_wizard()
x+=1 #incrementar contador em 1

peptide =
((sys.argv[6])+(sys.argv[7])+(sys.argv[8])+(sys.argv[9])+(sys.argv[1
0])+(sys.argv[11])+(sys.argv[12])+(sys.argv[13])+(sys.argv[14]))
#serie de aminoacidos a serem usadas no nome do ouput

cmd.save("%s_%s.pdb" %(name, peptide), name, 0, 'pdb') #salvar pdb
criado

print (peptide)
cmd.quit() #fechar função mutagenesis
```

## Apêndice F - Comandos utilizados no atracamento molecular

Comando utilizado no script flexpep\_prepack:

```
./FlexPepDocking.static.linuxgccrelease -s receptor_e_ligante -
flexpep_prepack -ex1 -ex2aro

-s = estrutura do ligante e receptor (arquivo de entrada)
-flexpep_prepack = Modo de preempacotamento. Otimização de cadeias
laterais, sem realização de atracamento molecular
-ex1/-ex1aro -ex2/-ex2aro = Adição de rotâmetros de cadeias
laterais.
```

Comando utilizado no script FlexPepDocking:

```
./FlexPepDocking.static.linuxgccrelease -s
receptor_e_ligante_previamente_empacotados -out:file:silent
decoys.silent -out:file:silent_struct_type binary -pep_refine -ex1
-ex2aro -use_input_sc -nstruct 100 -lowres_preoptimize

-s = estrutura do ligante e receptor previamente empacotada
(arquivo de entrada)
-out:file:silent= formato de saída do arquivo de saída
-out:file:silent_struct_type= Formato a ser lido do arquivo de
saída
-pep_refine= Modo de Refinamento
-ex1/-ex1aro -ex2/-ex2aro = Adição de rotâmetros de cadeias
laterais.
-use_input_sc = Incluir conformações dos rotâmetros da estrutura de
entrada durante o re-empacotamento de cadeias laterais
-nstruct = Número de conformações geradas na simulação
-lowres_preoptimize = Realizar uma rodada preliminar do modo de
otimização de centroides
```

**Apêndice G - Resultados detalhados da análise de similaridade para *L. braziliensis***

**Tabela 9: Resultados da análise de similaridade para *L. braziliensis***

| <b>LbrM11_V2.0090</b> |                |                           |                          |                         |                       |                       |                      |
|-----------------------|----------------|---------------------------|--------------------------|-------------------------|-----------------------|-----------------------|----------------------|
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 4u7a_A                | 8,00E-117      | 55                        | 742                      | 149                     | 825                   | 32,44                 | 93                   |
| 5em2_C                | 9,00E-64       | 336                       | 741                      | 9                       | 387                   | 30,92                 | 55                   |
| 5em2_A                | 9,00E-64       | 336                       | 741                      | 9                       | 387                   | 30,92                 | 55                   |
| 5cxc_B                | 2,00E-58       | 347                       | 741                      | 1                       | 368                   | 30,27                 | 53                   |
| 5cxb_B                | 2,00E-58       | 347                       | 741                      | 1                       | 368                   | 30,27                 | 53                   |
| <b>LbrM28_V2.1900</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 5vhs_U                | 0              | 4                         | 976                      | 1                       | 928                   | 36,45                 | 98                   |
| 5vhi_U                | 0              | 4                         | 976                      | 1                       | 928                   | 36,45                 | 98                   |
| 5vhh_U                | 0              | 4                         | 976                      | 1                       | 928                   | 36,45                 | 98                   |
| 5vgz_U                | 0              | 4                         | 976                      | 1                       | 928                   | 36,45                 | 98                   |
| 5t0j_U                | 0              | 4                         | 976                      | 1                       | 928                   | 36,45                 | 98                   |
| <b>LbrM28_V2.3070</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 4i1i_B                | 0              | 1                         | 324                      | 22                      | 345                   | 85,49                 | 100                  |
| 4i1i_A                | 0              | 1                         | 324                      | 22                      | 345                   | 85,49                 | 100                  |
| 4h7p_B                | 0              | 1                         | 324                      | 22                      | 345                   | 85,49                 | 100                  |
| 4h7p_A                | 0              | 1                         | 324                      | 22                      | 345                   | 85,49                 | 100                  |



| 2cvq_B                | 9,00E-98       | 2                         | 322                      | 3                       | 321                   | 50,47                 | 99                   |
|-----------------------|----------------|---------------------------|--------------------------|-------------------------|-----------------------|-----------------------|----------------------|
| <b>LbrM13_V2.1480</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 5dbj_D                | 5,00E-06       | 29                        | 103                      | 2                       | 74                    | 38,67                 | 13                   |
| 5dbj_C                | 5,00E-06       | 29                        | 103                      | 2                       | 74                    | 38,67                 | 13                   |
| 5dbj_B                | 5,00E-06       | 29                        | 103                      | 2                       | 74                    | 38,67                 | 13                   |
| 5dbj_A                | 5,00E-06       | 29                        | 103                      | 2                       | 74                    | 38,67                 | 13                   |
| 5dbj_E                | 5,00E-06       | 29                        | 103                      | 2                       | 74                    | 38,67                 | 13                   |
| <b>LbrM22_V2.0110</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 2vxo_B                | 0              | 12                        | 656                      | 32                      | 697                   | 47,79                 | 98                   |
| 2vxo_A                | 0              | 12                        | 656                      | 32                      | 697                   | 47,79                 | 98                   |
| 2ywc_D                | 6,00E-82       | 12                        | 531                      | 2                       | 490                   | 36,79                 | 96                   |
| 2ywc_D                | 2,00E-08       | 549                       | 656                      | 394                     | 503                   | 33,91                 | 96                   |
| 2ywc_C                | 6,00E-82       | 12                        | 531                      | 2                       | 490                   | 36,79                 | 96                   |
| 2ywc_C                | 2,00E-08       | 549                       | 656                      | 394                     | 503                   | 33,91                 | 96                   |
| 2ywc_B                | 6,00E-82       | 12                        | 531                      | 2                       | 490                   | 36,79                 | 96                   |
| 2ywc_B                | 2,00E-08       | 549                       | 656                      | 394                     | 503                   | 33,91                 | 96                   |
| <b>LbrM05_V2.0730</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>          | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 2yx0_A                | 4,00E-52       | 427                       | 721                      | 21                      | 310                   | 34,67                 | 35                   |

|        |          |     |     |   |     |       |    |
|--------|----------|-----|-----|---|-----|-------|----|
| 2z2u_A | 1,00E-39 | 430 | 753 | 2 | 306 | 30,00 | 38 |
|--------|----------|-----|-----|---|-----|-------|----|

---

**Apêndice H - Resultados detalhados da análise de similaridade para *L. infantum***

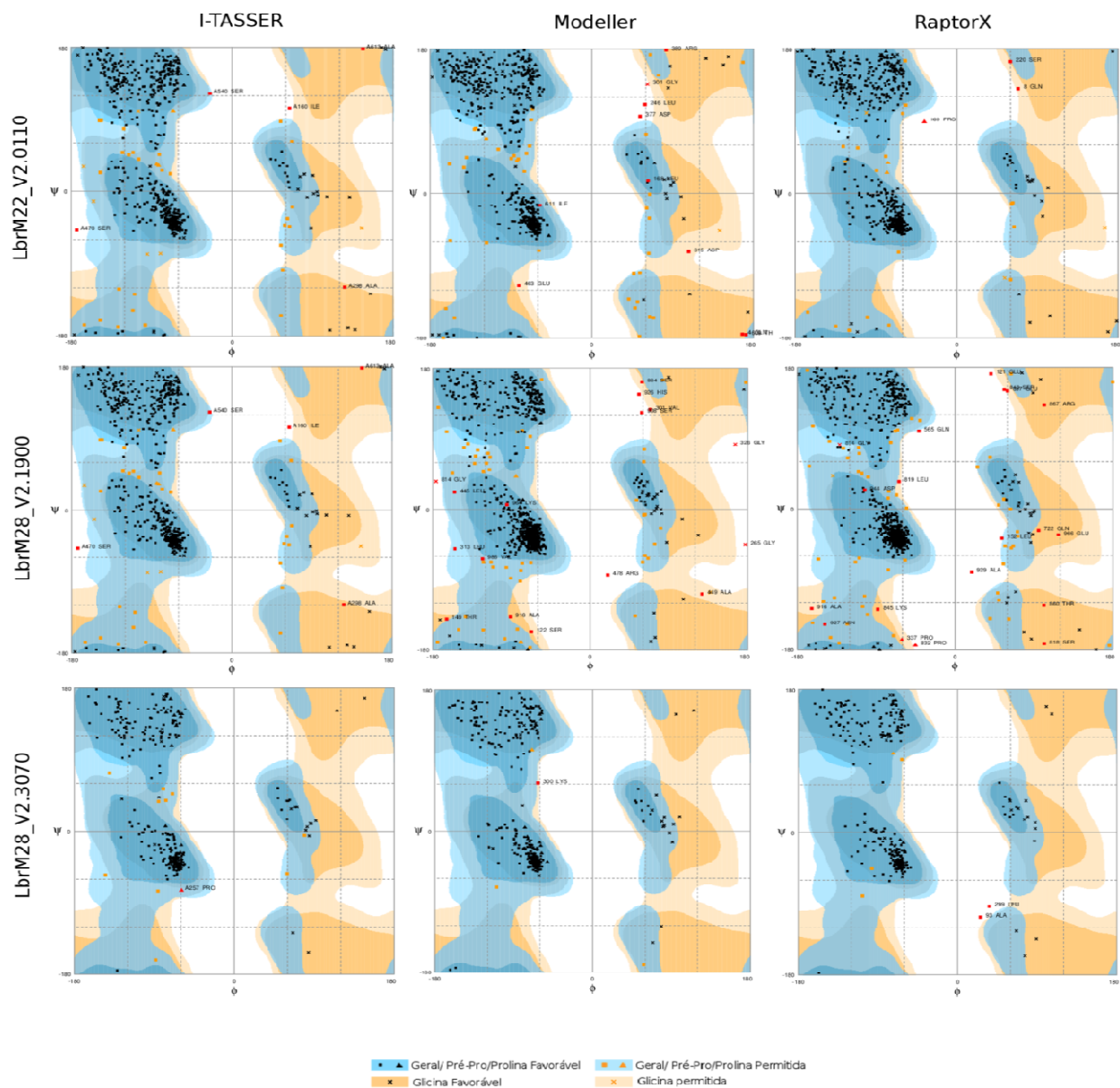
**Tabela 10: Resultados da análise de similaridade para *L. infantum***

| LinJ.32.0970 |           |                    |                   |                  |                |                |               |
|--------------|-----------|--------------------|-------------------|------------------|----------------|----------------|---------------|
| Molde        | e-value   | Início da proteína | Final da proteína | Início do Moldes | Final do Molde | Identidade (%) | Cobertura (%) |
| 5voz_b       | 4,00E-123 | 6                  | 884               | 3                | 820            | 30,50          | 98            |
| 5voy_b       | 4,00E-123 | 6                  | 884               | 3                | 820            | 30,50          | 98            |
| 5vox_b       | 4,00E-123 | 6                  | 884               | 3                | 820            | 30,50          | 98            |
| 3j9v_b       | 4,00E-123 | 6                  | 884               | 3                | 820            | 30,50          | 98            |
| 3j9u_b       | 4,00E-123 | 6                  | 884               | 3                | 820            | 30,50          | 98            |
| LinJ.19.0350 |           |                    |                   |                  |                |                |               |
| Molde        | e-value   | Início da proteína | Final da proteína | Início do Moldes | Final do Molde | Identidade (%) | Cobertura (%) |
| 4u0p_B       | 4,00E-68  | 90                 | 380               | 4                | 288            | 41,75          | 71            |
| 4u0o_B       | 4,00E-68  | 90                 | 380               | 4                | 288            | 41,75          | 71            |
| 5exk_K       | 3,00E-66  | 85                 | 382               | 41               | 319            | 43,14          | 72            |
| 5exk_I       | 3,00E-66  | 85                 | 382               | 41               | 319            | 43,14          | 72            |

| 5exk_G              | 3,00E-66       | 85                        | 382                      | 41                      | 319                   | 43,14                 | 72                   |
|---------------------|----------------|---------------------------|--------------------------|-------------------------|-----------------------|-----------------------|----------------------|
| <b>LinJ.18.1500</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>        | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 5ksd_B              | 1,00E-150      | 60                        | 884                      | 20                      | 827                   | 36,74                 | 85                   |
| 5ksd_A              | 1,00E-150      | 60                        | 884                      | 20                      | 827                   | 36,74                 | 85                   |
| 1mhs_B              | 2,00E-137      | 61                        | 883                      | 87                      | 882                   | 33,41                 | 84                   |
| 1mhs_A              | 2,00E-137      | 61                        | 883                      | 87                      | 882                   | 33,41                 | 84                   |
| 4xe5_A              | 3,00E-63       | 20                        | 692                      | 18                      | 803                   | 26,88                 | 69                   |
| <b>LinJ.36.5700</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>        | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 3dme_B              | 1,00E-21       | 58                        | 376                      | 6                       | 294                   | 28,04                 | 61                   |
| 3dme_A              | 1,00E-21       | 58                        | 376                      | 6                       | 294                   | 28,04                 | 61                   |
| 4x9n_A              | 2,00E-10       | 58                        | 211                      | 39                      | 194                   | 31,45                 | 30                   |
| 4x9m_A              | 2,00E-10       | 58                        | 211                      | 39                      | 194                   | 31,45                 | 30                   |
| <b>LinJ.36.2160</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>        | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| -                   | -              | -                         | -                        | -                       | -                     | -                     | -                    |
| <b>LinJ.28.1850</b> |                |                           |                          |                         |                       |                       |                      |
| <b>Molde</b>        | <b>e-value</b> | <b>Início da proteína</b> | <b>Final da proteína</b> | <b>Início do Moldes</b> | <b>Final do Molde</b> | <b>Identidade (%)</b> | <b>Cobertura (%)</b> |
| 5vhs_U              | 0              | 5                         | 858                      | 2                       | 836                   | 38,01                 | 86                   |
| 5vhi_U              | 0              | 5                         | 858                      | 2                       | 836                   | 38,01                 | 86                   |
| 5vhh_U              | 0              | 5                         | 858                      | 2                       | 836                   | 38,01                 | 86                   |
| 5vgz_U              | 0              | 5                         | 858                      | 2                       | 836                   | 38,01                 | 86                   |

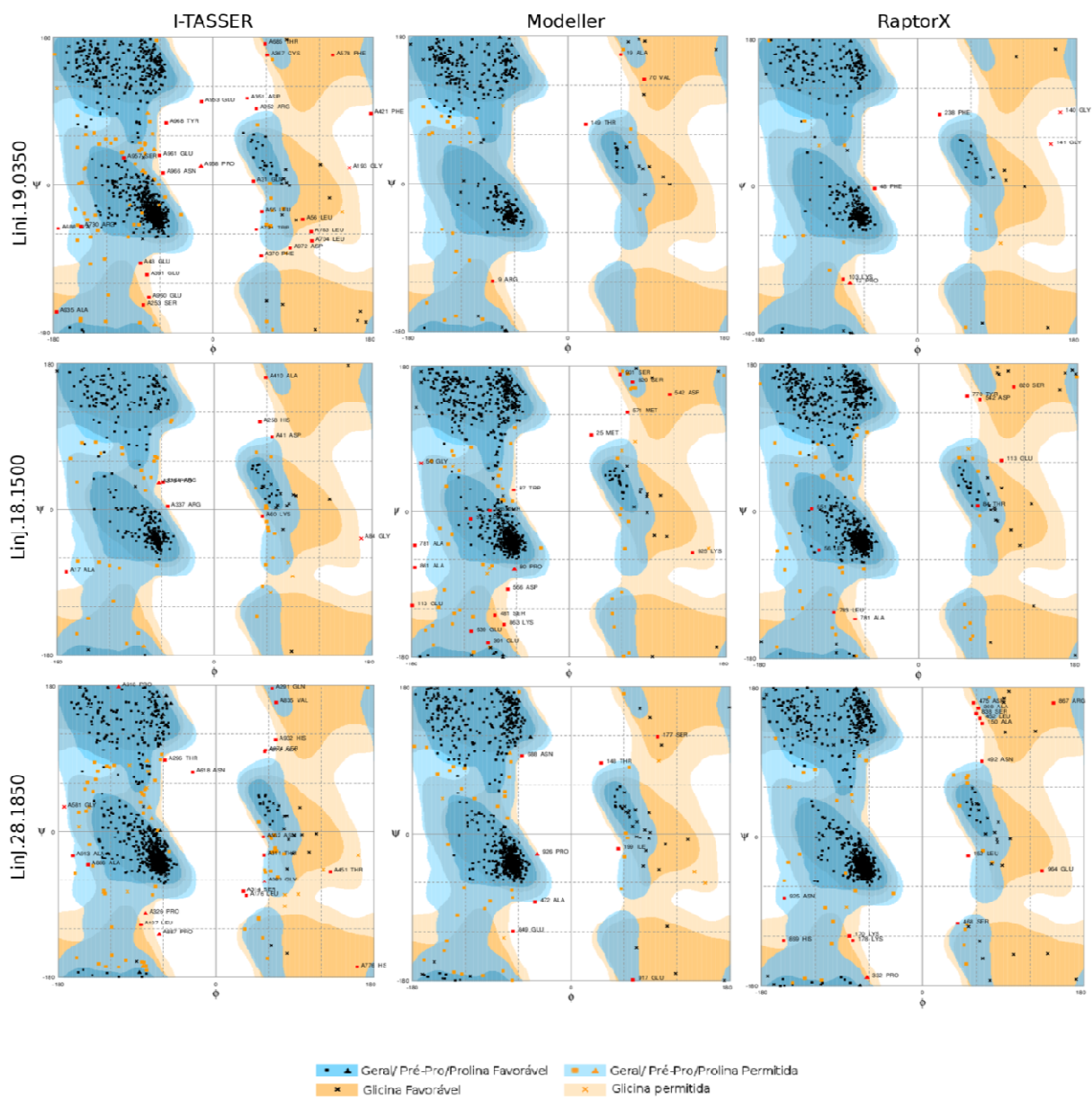
|        |   |   |     |   |     |       |    |
|--------|---|---|-----|---|-----|-------|----|
| 5t0j_U | 0 | 5 | 858 | 2 | 836 | 38,01 | 86 |
|--------|---|---|-----|---|-----|-------|----|

---



## Apêndice I – Gráficos de Ramachandran para os melhores modelos de cada uma das abordagens

**Figura 23: Gráfico de Ramachandran das proteínas modeladas de *L. braziliensis*.** No eixo x temos representado os ângulos  $\phi$ , no eixo y os ângulos  $\psi$ . As áreas em azul representam combinação de ângulos favoráveis, as regiões em amarelo representam combinação de ângulos permitidos, as regiões em branco representam combinação de ângulos desfavoráveis. Cada uma das formas geométricas representam um resíduo da proteína.



**Figura 24: Gráfico de Ramachandran das proteínas modeladas de *L. infantum*.** No eixo x temos representado os ângulos  $\phi$ , no eixo y os ângulos  $\psi$ . As áreas em azul representam combinação de ângulos favoráveis, as regiões em amarelo representam combinação de ângulos permitidos, as regiões em branco representam combinação de ângulos desfavoráveis. Cada uma das formas geométricas representam um resíduo da proteína.

## Apêndice J - Predições do BEpro, Discotope e EPCES e Ellipro

**Tabela 11: Posição dos resíduos nas proteínas de estudo identificados como parte de epítomos de células B conformacionais pelos preditores BEpro, Discotope e EPCES.**

|                | BEpro*  | Discotope*  | EPCES*  |
|----------------|---|---|---|
| LbrM22_V2.0110 | 69,299,301,487-489  | 8,20,24,27,65-70,104-105,109-112,217-218,264,298-303,355-357,361,374,398,400,413-414,440-441,486-489,503-504,535,569-570,572,575,605,606,608  | 1,6-9,11,46,49-53,55,80-83,99-100,104,130,143-144,146-148,156-158,167-168,298-301,304-306,442,444-446,448-449,451,453-472,474,500,502-504,506-508,510-511,514-515,518,536,562-563,566,569,571,605,617   |
| LbrM28_V2.1900 | 1,123-126,128,409,745-746,844-845,863-881,897-898,943-944,966,987-988 | 1-2,119-120,122-129,148-149,178-179,374-377,379-380,402,405-409,435,439,447,452,475,478-479,492,551,558,562-563,565,567,599,622,633,635,669,704,726,728-729,733,736,740,743-747,775-776,839,842-846,861-884,893-902,904-907,943-945,952,963-968,983,986-989,991,993 | 118,142,144-146,148-150,178-179,218,239-240,243-244,246-250,272,275-276,278-281,283-284,286-295,298-299,303-305,309-310,312-316,320-336,371,375,475-478,489,491,523,526-527,661,692,696,698-699,709,713,741,743-746,752,776,826,850,864-881,887-898,901,934,936-938,954,988 |
| LbrM28_V2.3070 | 93-   | 92, 93, 94, 95, 96, 97, 98, 99,   | 32, 111, 117, 151,  |



94,96,212,213,  
214, 302-306

145, 202, 207, 211, 212, 213,  
214, 215, 217, 220, 224, 258,  
259, 269, 270, 272, 274, 276,  
277, 297, 299, 300, 301, 302,  
303, 304, 305, 306, 307, 320,  
323, 324

180, 192, 195, 196,  
197, 199, 202, 213,  
214, 216, 217, 218,  
219, 220, 221, 224,  
228, 232, 259, 269,  
270, 272, 274, 293,  
296, 297, 298, 300,  
302, 303, 304, 305,  
306, 307, 312

LinJ.18.1500

45, 46, 47, 81,  
83, 85, 86, 87,  
160, 161, 435,  
436, 437, 438,  
439, 456, 458,  
459, 460, 487,  
488, 489, 490,  
505, 553, 554,  
555, 556, 557,  
559, 561, 562,  
563, 564, 565,  
691, 692, 847,  
849, 850

42, 43, 44, 45, 46, 47, 48, 49,  
50, 51, 55, 66, 73, 74, 75, 80,  
81, 82, 83, 84, 85, 86, 87, 88,  
89, 159, 160, 161, 162, 163,  
215, 247, 250, 251, 253, 256,  
364, 373, 390, 391, 392, 393,  
411, 412, 423, 424, 425, 434,  
435, 436, 437, 438, 439, 453,  
455, 456, 457, 458, 459, 460,  
461, 462, 463, 485, 486, 487,  
488, 489, 490, 505, 506, 508,  
509, 512, 513, 515, 516, 518,  
552, 553, 554, 555, 556, 557,  
558, 559, 560, 561, 562, 563,  
564, 565, 566, 567, 597, 689,  
690, 691, 692, 695, 729, 730,  
733, 735, 768, 769, 770, 771,  
772, 773, 784, 816, 845, 847,  
848, 849, 850, 851

12, 13, 24, 36, 37,  
38, 39, 40, 44, 45,  
46, 47, 48, 49, 50,  
51, 52, 54, 84, 85,  
86, 87, 88, 89, 92,  
93, 189, 191, 213,  
214, 215, 216, 279,  
286, 287, 290, 367,  
368, 369, 370, 372,  
375, 376, 377, 555,  
556, 557, 562, 567,  
568, 574, 576, 687,  
688, 689, 690, 691,  
692, 693, 694, 695,  
696, 697, 698, 760,  
763, 764, 765, 766,  
769, 770, 771, 772,  
773, 774, 775, 783,  
784, 785, 787, 841,  
842, 843, 844, 845,  
846, 847, 848, 849,  
850, 859, 860, 861,  
862, 888, 889, 892,  
893, 894, 895, 896,  
897, 898, 899, 900,  
901, 902, 903, 904,  
905, 906, 907, 908,  
911, 912, 913, 914,  
915, 916, 917, 918,  
919, 920, 921, 923,  
926, 927, 928, 941,  
943, 945, 947, 950,  
953, 954, 955, 956,  
957, 958, 959, 960,  
961, 967, 971, 972,  
973, 974

|              |   |   |  |
|--------------|---|---|--|
| LinJ.19.0350 | 112,116,119-<br>120,173,176,310<br>,337,349,390-<br>393                 | 110,112-122,143,172-<br>173,175-180,206-<br>208,224,265,272-<br>273,276,278,292-294,308-<br>311,314,337,347,349-<br>350,353,360,387,389-393   | 169,182-<br>183,187,190,213-<br>214,216-217,220-<br>221,223-227,242-<br>244,246-247,250-<br>253,269,285-<br>286,288-296,345  |
| LinJ.28.1850 | 125,282,332-<br>333,447,449,743<br>-<br>744,829,831,840<br>,942,964,966 | 1-3,119-127,147-<br>149,198,200,279,281,282,332<br>-335,369,372-373,375,377-<br>378,403-<br>407,410,434,437,445-<br>451,472,477,481,484-<br>485,488,559-<br>563,597,620,633,667,723-<br>725,727,730-731,734,741-<br>745,773-774,793,828-<br>832,834,839-841,872-<br>873,920-921,941-942,961-<br>966 | 56,116,118-121,123-<br>127,141-142,144-<br>145,147-151,178-<br>180-183,210,214-<br>215,217-<br>219,227,230-<br>231,235-237,239-<br>240,243-244,246-<br>251,266,268,271-<br>272,275-276,278-<br>282,332-<br>336,357,366,369,372<br>-375,404,437,446-<br>448,450-451,474-<br>477,484,487-<br>492,514,520-<br>522,524-525,555-<br>557,567,595,597,655<br>,658,691,693-<br>694,696-<br>697,700,739,741-<br>747,774,825,827,829<br>,840-<br>841,870,920,932,934<br>-937,942-943,963-<br>966 |

\*A tabela apresenta estas coordenadas separadas por vírgula e quando um trecho contínuo é identificado este aparece separado por um traço.

**Tabela 12: Coordenadas dos resíduos nas proteínas de estudo identificados como parte de epítomos de células B pelo Ellipro.**

| <b>Ellipro*</b> |   |
|-----------------|---|
| LbrM22_V2.0110  | Primeiro: 433-480, 499-507,510-511,533-547, 586-595, 630-631, 633-640<br>Segundo: 1-10, 40-55, 62-84, 90-105, 108-112, 114-133, 136-172, 174-175, 183-184, 191-191, 194-199, 201-201<br>Terceiro: 216-216, 219-225, 244-252, 270-272, 274-285, 287-288, 291-313, 329-330, 332-346, 356-356, 371-390 |

---

|                |  |
|----------------|--|
|                | Quarto: 518-518, 521-525<br>Quinto : 615-618, 621-623, 626-627   |
| LbrM28_V2.1900 | Primeiro: 1-17 19-109 112-112 115-115 123-124 126-152<br>Segundo: 243-251 275-340 726-726 816-834 839-844 846-915 917-927<br>940-948 961-981 983-989 991-991<br>Terceiro: 367-367 371-371 373-383 406-416 419-419 442-442 445-460<br>477-477 482-483 485-498 514-516 518-533 549-549 551-551 554-571<br>577-577 597-604 634-637 640-640 664-664 667-672 674-674 701-707<br>709-709 736-737 740-752 795-795<br>Quarto: 174-181 215-220  |
| LbrM28_V2.3070 | Primeiro: 1-3 11-12 24-26 29-36 41-72 74-75<br>Segundo: 89-99 102-103 167-180 190-190 195-217 220-220 255-260<br>268-279 295-310 313-313 316-317 319-324<br>Terceiro:227-236<br>Quarto: 78-81 106-106 109-110 113-113 116-123 142-147  |
| LinJ.18.1500   | Primeiro: 103-104 106-119 122-122 274-275 277-288 290-298 678-702<br>752-753 755-791 794-794 825-825 828-829 832-875<br>Segundo: 42-81 157-161 163-163 169-176 184-194 199-219 229-250<br>Terceiro: 362-382 384-387 402-417 430-435 437-443 447-449 451-477<br>479-479 482-498 500-502 504-506 508-509 512-512 527-531 533-534<br>536-537 539-571 573-576 580-580 596-596<br>Quarto: 878-879 882-882 885-886 889-889<br>Quinto: 82-87<br>Sexto: 88-89 92-92<br>Sétimo: 811-811 813-821 824-824 |
| LinJ.19.0350   | Primeiro: 309-309 313-313 356-357 359-360<br>Segundo: 167-167 334-339<br>Terceiro: 183-183 217-217 220-221 224-225<br>Quarto: 110-124 134-148 187-187 190-195 226-229<br>Quinto: 156-164 168-180 184-184 204-211 213-213 262-266 268-269<br>272-273 276-277 304-308 310-311 314-314 332-332 340-350 352-353<br>Sexto: 3-324 362-364 383-383 386-393<br>Setimo:239-243 246-247 278-278 282-282 285-286 288-295 317-318<br>321-322   |
| LinJ.28.1850   | Primeiro: 1-17 19-55 57-106 122-122 124-134 136-137 140-141 143-<br>152<br>Segundo: 239-240 242-245 247-254 275-282 332-336 362-366 368-381<br>403-406 408-414 417-417 439-439 441-456 481-481 483-496 510-510<br>512-514 516-531 547-550 552-568 595-602 630-635 638-638 661-661<br>664-666 668-670 672-672 693-693 696-697 699-707 721-728 730-750<br>753-753 812-841 867-873 920-925 936-936 938-949 956-972 974-978<br>Terceiro: 169-171 174-180 215-220                                   |

---

\*A tabela apresenta estas coordenadas separadas por vírgula e quando um trecho contínuo é identificado este aparece separado por um traço.



## Apêndice K – Tabela com os valores de Isc para atracamentos entre peptídeos e MHC de classe I

**Tabela 13: Valores de Isc para atracamentos entre peptídeos e MHC de classe I**

|           | 2HJL   | 3C9N   | 3HCV   | 3KPP   | 3L3D   | 3RL1   | 3VCL   | 3VFS   | 4F7M   | 4G8G   | 4HWZ   | 4JQX   | 4MJ5   | 4MJI   | 4NQV   | 4O2C   | 4QRR   | 4QRU   | 4WU5   | 4XXC   |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| original  | -39889 | -38704 | -41331 | -38551 | -42768 | -36861 | -40968 | -49166 | -41947 | -41464 | -30901 | -49112 | -37142 | -29704 | -42483 | -43444 | -35759 | -50224 | -47051 | -44698 |
| FALKRLSSL | -23274 | -38465 | -32848 | -33757 | -36213 | -38043 | -25455 | -21904 | -31024 | -31581 | -26762 | -20050 | -25035 | -36795 | -33835 | -40935 | -36412 | -45368 | -34912 | -36290 |
| FMLLGMPYI | -30039 | -34970 | -31213 | -31529 | -32579 | -31957 | -22340 | -25414 | -33414 | -30661 | -34546 | -24579 | -23970 | -32587 | -31559 | -34113 | -31362 | -40210 | -34528 | -33211 |
| IGPGSALAL | -21712 | -33979 | -32204 | -26983 | -33047 | -28115 | -19926 | -20151 | -27746 | -26630 | -24750 | -18616 | -22341 | -27480 | -29907 | -32945 | -30814 | -32793 | -36146 | -32758 |
| KMLTKSSIM | -24729 | -40131 | -31588 | -30546 | -33674 | -33532 | -23537 | -17530 | -30715 | -28985 | -31021 | -18716 | -21351 | -31594 | -31251 | -30767 | -30802 | -41057 | -29724 | -33750 |
| KPQRQSVL  | -19652 | -35481 | -33704 | -32447 | -34423 | -30122 | -17243 | -17055 | -28022 | -27845 | -26873 | -17834 | -18432 | -28934 | -28539 | -26303 | -29990 | -40206 | -31812 | -38711 |
| LAYDHRAYL | -26954 | -42162 | -38235 | -33946 | -42726 | -36651 | -26988 | -25176 | -37268 | -30283 | -34625 | -23193 | -24473 | -38612 | -39033 | -34651 | -33006 | -37382 | -37548 | -35609 |
| LVVSIPIAL | -23878 | -35262 | -34219 | -30039 | -27965 | -31148 | -19467 | -21654 | -29117 | -27363 | -31336 | -22222 | -20600 | -28545 | -27682 | -32790 | -32135 | -36218 | -33577 | -31007 |
| PARDNEPKL | -17461 | -35789 | -31294 | -29087 | -36096 | -26449 | -19093 | -19046 | -30964 | -26924 | -26655 | -16830 | -21032 | -31477 | -34163 | -35490 | -29050 | -41589 | -33173 | -39760 |
| QMVYNQDEI | -26020 | -35745 | -29456 | -30499 | -30326 | -33280 | -24968 | -21550 | -30892 | -31356 | -33461 | -18976 | -23044 | -31228 | -33665 | -32445 | -32375 | -38584 | -33235 | -34820 |
| QSGHNSGCL | -22988 | -40312 | -30213 | -31362 | -32000 | -29038 | -21634 | -18626 | -29510 | -24695 | -26125 | -18037 | -19952 | -31486 | -30918 | -29031 | -24732 | -34826 | -32725 | -34790 |
| RGGSRMVAF | -23677 | -38974 | -36807 | -35347 | -37980 | -35086 | -22369 | -24476 | -35687 | -24220 | -30840 | -21157 | -21509 | -29726 | -31690 | -32175 | -29861 | -41857 | -37396 | -38334 |
| RGYTCAMTG | -22456 | -35367 | -30849 | -31537 | -31488 | -28605 | -26309 | -22227 | -30313 | -25714 | -32003 | -18521 | -21670 | -25860 | -29443 | -26059 | -28547 | -36297 | -28808 | -33031 |
| RLCPRGHSL | -25710 | -34618 | -35039 | -30416 | -34459 | -38039 | -28485 | -18399 | -31290 | -36516 | -29280 | -17958 | -20322 | -27725 | -30116 | -33705 | -29730 | -43749 | -35212 | -36990 |
| RPPPLDPEE | -17752 | -36408 | -28843 | -24695 | -29388 | -21833 | -22343 | -18628 | -28713 | -27043 | -22883 | -20444 | -18630 | -21950 | -23172 | -29093 | -31231 | -34006 | -33051 | -37789 |
| SAVTLGQYL | -29343 | -36771 | -34073 | -33923 | -35789 | -30264 | -20769 | -19190 | -32551 | -32239 | -29923 | -20702 | -27019 | -32460 | -32500 | -37338 | -30926 | -35809 | -38422 | -32783 |
| SGYTWAPIS | -23342 | -35082 | -33756 | -33887 | -32489 | -30863 | -20277 | -20646 | -30243 | -23781 | -32067 | -20696 | -31956 | -25594 | -31455 | -29859 | -26898 | -35942 | -31521 | -27947 |
| SVIHNATVV | -24666 | -36711 | -25578 | -26127 | -27112 | -31534 | -21521 | -19595 | -25007 | -27239 | -31428 | -18375 | -25470 | -30962 | -32388 | -31721 | -27280 | -32942 | -32337 | -34603 |
| TLPFWRVAW | -40378 | -34447 | -30599 | -39734 | -43082 | -33075 | -22045 | -27781 | -39576 | -32267 | -33248 | -23933 | -29702 | -27062 | -32940 | -30420 | -32841 | -36381 | -44255 | -40636 |
| YLLARFYET | -22640 | -36695 | -34413 | -31871 | -31671 | -35319 | -24467 | -23273 | -33356 | -34520 | -33689 | -23988 | -25176 | -27497 | -34571 | -36399 | -35322 | -43464 | -34440 | -41661 |
| YLQPSRTRL | -28718 | -43347 | -35942 | -36428 | -37072 | -32117 | -29472 | -23087 | -34995 | -35089 | -31695 | -19156 | -22077 | -30145 | -32236 | -33656 | -29263 | -42027 | -31471 | -36765 |
| YLRLWALSL | -27668 | -43363 | -39614 | -36855 | -35703 | -32806 | -26622 | -21182 | -35206 | -36126 | -31066 | -23012 | -28436 | -33075 | -34996 | -32824 | -33455 | -47557 | -36702 | -36710 |
| YWYWYPYLL | -31544 | -51296 | -36566 | -35351 | -29305 | -38235 | -23522 | -22683 | -38709 | -41018 | -33970 | -27001 | -26266 | -36364 | -34503 | -39206 | -37078 | -37772 | -47430 | -40239 |

## Apêndice L – Tabela com os valores de Isc para atracamentos entre peptídeos e MHC de classe II

Tabela 14: Valores de Isc para atracamentos entre peptídeos e MHC de classe II

|                  | 1A6A   | 1BX2   | 1H15   | 1S9V   | 1UVQ   | 1YMM   | 2NNA   | 2Q6W   | 3C5J   | 3LQZ   | 3PL6   | 3WEX   |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| original         | -60157 | -50889 | -54836 | -44419 | -59002 | -48842 | -54701 | -52741 | -36422 | -57418 | -56123 | -39508 |
| EMGVDIYIVMTMVDRD | -45152 | -41740 | -40871 | -48187 | -44236 | -48086 | -49397 | -44926 | -37580 | -42952 | -42513 | -37209 |
| EYVLSSVSNTASYLR  | -52296 | -48380 | -41642 | -39228 | -50893 | -42300 | -47322 | -42179 | -42460 | -42966 | -51484 | -34198 |
| FFFYVPPSPILFCGA  | -49148 | -50130 | -43014 | -38589 | -48276 | -47201 | -49004 | -42101 | -46251 | -49451 | -49929 | -39480 |
| FYVANVLLQDITFRR  | -48428 | -43723 | -43897 | -44011 | -51385 | -41701 | -48491 | -46919 | -43334 | -46030 | -44682 | -46054 |
| GAPVAAAVPESPEQK  | -38038 | -35182 | -37994 | -42224 | -46443 | -38970 | -42308 | -38584 | -37624 | -37232 | -43506 | -30739 |
| GFESFPPTPKTSMML  | -42143 | -44585 | -46080 | -36606 | -39947 | -42307 | -41855 | -35753 | -31396 | -38521 | -42628 | -34508 |
| GGHFFFYVPPSPILF  | -46790 | -49770 | -47623 | -44472 | -48056 | -46689 | -58107 | -38205 | -41735 | -39673 | -51097 | -44159 |
| IRQGFESFPPTPKTS  | -47642 | -41535 | -43421 | -34951 | -42849 | -42950 | -43570 | -36487 | -41343 | -39944 | -38776 | -37858 |
| KGTTYPTTPNGLPSV  | -39373 | -41770 | -40306 | -33774 | -41465 | -35662 | -42588 | -32916 | -34664 | -32343 | -42047 | -31929 |
| LKLVEQLAVTPLSVY  | -47360 | -45879 | -43407 | -41216 | -45805 | -41821 | -43645 | -41775 | -44051 | -37762 | -40275 | -43974 |
| LVEQLAVTPLSVYAH  | -51704 | -42047 | -42240 | -49699 | -50372 | -41774 | -49818 | -38948 | -42146 | -38735 | -45668 | -39844 |
| PRVTVPATMDKKMLL  | -45825 | -41932 | -39852 | -40740 | -42391 | -40064 | -38637 | -27417 | -40904 | -41202 | -42966 | -33873 |
| QKGTTYPTTPNGLPS  | -49337 | -44659 | -38656 | -37296 | -41267 | -39631 | -38798 | -24724 | -35349 | -43120 | -38358 | -31803 |
| TSFVVLCASTMIVR   | -46956 | -43401 | -41178 | -39891 | -48875 | -42426 | -45583 | -36194 | -43123 | -48946 | -43914 | -37529 |
| TVATWVLLWRTRWAA  | -43184 | -45950 | -57141 | -41473 | -53616 | -43407 | -50100 | -35443 | -40115 | -47766 | -47092 | -43542 |
| VATAREIRQKYPRKR  | -42825 | -35796 | -45041 | -34352 | -36917 | -46249 | -40505 | -35266 | -34477 | -41906 | -41936 | -42877 |



