

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Análise intersticial de complexos pMHC para predição de afinidade de ligação

Martela Vaz de Freitas

Porto Alegre

2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Análise intersticial de complexos pMHC para predição de afinidade de ligação

Martiel Vaz de Freitas

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do título de Mestre em Genética e Biologia Molecular.

Linha de pesquisa: Genômica, Bioinformática e Sistemas Biológicos.

Orientador: Dr. Gustavo Fioravanti Vieira

Porto Alegre

2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Análise intersticial de complexos pMHC para predição de afinidade de ligação

Martiel Vaz de Freitas

BANCA EXAMINADORA

Dr. Maurício Garcia de Souza Costa

Dr. Maurício Menegatti Rigo

Dra. Ursula da Silveira Matte

Dr. Márcio Dorn (Suplente)

Dr. José Artur Bogo Chies (Suplente)

Porto Alegre, 2017

Este trabalho foi desenvolvido no Núcleo de Bioinformática do Laboratório de Imunogenética do Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul.

Aos que me ajudaram a chegar aqui.

AGRADECIMENTOS

Inicialmente, agradeço ao PPGBM. Agradeço aos meus professores, indistintamente. Ao Elmo, por todos os cafés.

Agradeço imensamente ao meu orientador, Dr. Gustavo Fioravanti Vieira, pela abertura de um novo caminho, pela amizade e sobretudo pela confiança. Essa foi essencial em todos os momentos.

Agradeço ao NBLI. Maurício e Meg, por me escolherem para integrar o laboratório ainda em 2013; Dinler, pela boa vontade em me ambientar na Imuno; Matheus, por entrar comigo e partilhar dos “*Sócrates Cafés*”; Renata, porque ficaria feio não agradecer também.

Agradeço e me desculpo com o Prof. Dr. José Artur Bogo Chies, por todas as vezes em que alguém o chamava para responder dúvidas que eram minhas.

Agradeço ao Marcus e ao Bragatte, por serem as pessoas que chamavam o Prof. José Artur. Verdadeiros companheiros em pesquisa, sempre alegres e bem-dispostos, pacientes e extremamente calorosos comigo.

Agradeço aos amigos da Imunogenética: Francis, Jacqueline, Joel e Valéria: alegria é imprescindível. Aos amigos de outros ambientes do pós e da graduação, da FURG, da Unisinos e da vida! Não cito nomes para que eu não esqueça ninguém.

Agradeço a Sra. Cleuza e ao Sr. Jackson Medeiros pelo suporte incondicional.

E não há como não agradecer ao Dr. Jackson Medeiros. Tu sabes da importância que tens na minha história. Sabes o tamanho do carinho, do amor e do zelo que guardo em mim. Obrigada pelas críticas, sugestões e apoio. Agradeço os livros emprestados (voluntariamente ou não), os cafés e as piadas sem graça. Obrigada por todos esses anos de convivência e por me ensinar a permanecer alerta e em movimento. Não importa como, sigo contigo.

Agradeço a minha gata Momo, por me acordar TODOS os dias, as 6h da manhã, impreterivelmente. Incontáveis vezes me salvou de atrasos.

Agradeço a minha família, principalmente. Obrigada por acreditarem que eu ainda chego em algum lugar, em algum momento. Amo vocês. Profundamente.

Em tempo, agradeço à relatora e aos membros da banca que avaliaram este trabalho.

“Estudar custa caro.
Aproveitem enquanto o Governo paga e estudem. “

Mario Rocha Retamoso, Prof. Dr.
Em alguma aula de Cálculo II, 2008

RESUMO

O MHC é uma molécula apresentadora que possui um papel importante tanto na apresentação de peptídeos próprios como não-próprios, para que todo o sistema imune possa diferenciar a produção normal de proteínas de uma célula, da produção de proteínas de uma célula infectada por um patógeno. Existem duas vias de apresentação de antígenos: a via de Classe I, que apresenta peptídeos endógenos e a via de Classe II que apresenta preferencialmente peptídeos exógenos. Este trabalho atém-se à via de Classe I, pelo fato de que a própria estrutura deste MHC facilita o processo de modelagem. Isso se deve às características de sua fenda, que possui um espaço de apresentação fechado nas extremidades, permitindo apenas a acomodação de peptídeos menores (entre 8 e 12 aminoácidos). Aqui, buscou-se analisar estruturalmente a interação entre o peptídeo e o MHC, para inferir se um peptídeo é um ligante ou não. A capacidade de ligação de um peptídeo ao MHC é um ponto crucial para a imunogenicidade do mesmo. Até então a maioria das ferramentas que pretendem realizar a predição deste fenômeno são baseadas apenas na análise da sequência de aminoácidos.

Palavras-chave: Complexo pMHC, ligação, peptídeo, volume.

ABSTRACT

MHC is a presenting molecule that plays an important role both in presentation of cleaved proteins from our own cells and non-specific peptides. That is one of mechanisms how immune system can differentiate normal production of proteins from a cell infected by a pathogen. There are two antigens presenting pathways: Class I pathway, which shows endogenous peptides and Class II pathway, which preferably exhibits exogenous peptides. We set the Class I pathway, due to the fact that the structure of this MHC itself facilitates the modeling process. MHC's cleft has a closed presentation space at the ends, allowing only the accommodation of small peptides (between 8 and 12 amino acids). We aim to analyze structurally the interaction between the peptide and the MHC to infer if a peptide is a ligand or not, since the ability of a peptide to bind MHC is a crucial point for the immunogenicity of them. Until now, most part of tools that intend to carry out the prediction of this phenomenon are based on amino acid sequence analysis.

Keywords: Binding, peptide, pMHC Complex, Volume.

LISTA DE ILUSTRAÇÕES

Figura 1: Mapa gênico humano da região HLA.

Figura 2: Diagrama esquemático do MHC de classe I.

Figura 3: Diagrama esquemático do MHC de classe II.

Figura 4: Via endógena da apresentação de antígenos.

Figura 5: Interação entre TCR e complexo pMHC.

Figura 6: Esquema visual da técnica D1-EM-D2.

Figura 7: Representação gráfica de uma rede neural.

Figura 8: Estrutura em texto plano de um arquivo PDB.

Figura 9: Volume intersticial para a cadeia completa do epitopo complexado ao Cristal de MHC-I.

Figura 10: Volume intersticial para cada resíduo do epitopo complexado ao Cristal de MHC-I.

Figura 11: Parâmetros da janela de extração de volumes via SURFNET.

Figura 12: Motivos do MHC humano HLA-A*02:01

LISTA DE GRÁFICOS

Gráfico 1: Plot dos dados modelados para cadeia C do complexo.

Gráfico 2: Volume intersticial para os dados de volume de 190 sequências consideradas ligantes.

Gráfico 3: Volume intersticial para os dados de volume de 190 sequências consideradas não ligantes.

Gráfico 4: Volume intersticial para a posição 2 da cadeia C.

Gráfico 5: Volume intersticial para a posição 7 da cadeia C.

Gráfico 6: Volume intersticial para a posição 9 da cadeia C.

Gráfico 7: Curva ROC – Melhor rede treinada.

Gráfico 8: Curva ROC – Segunda melhor rede.

LISTA DE TABELAS

Tabela 1: Dados da extração de volumes de cristais

Tabela 2: Estatística primária para alvos positivos

Tabela 3: Estatística primária para alvos negativos

Tabela 4: Resultado do teste de Wilcoxon-Mann-Whitney

Tabela 5: Resultado dos testes de Equações de Estimativas Generalizadas

Tabela 6: Valores de entrada para ligantes.

Tabela 7: Valores de entrada para não ligantes.

Tabela suplementar 1: Alvos positivos iniciais pós-docagem por D1-EM-D2

Tabela suplementar 2: Alvos negativos iniciais pós-docagem por D1-EM-D2

Tabela suplementar 3: Dados de extração de valores de volume de pMHC "Positive".

Tabela suplementar 4: Dados de extração de valores de volume de pMHC
"Negative".

Tabela suplementar 5: Matriz de correlações.

Tabela suplementar 6: Aplicação de GEE.

LISTA DE ABREVIATURAS

ANN _ Redes Neurais Artificiais (do inglês, *Artificial Neural Network*).

BCR _ Receptor de célula B (do inglês, *B Cell Receptor*).

CD4 _ Grupamento de Diferenciação 4 (do inglês, *Cluster of Differentiation 4*).

CD8 _ Grupamento de Diferenciação 8 (do inglês, *Cluster of Differentiation 8*).

CDR _ Regiões Determinantes de Complementaridade (do inglês, *Complementarity Determining Region*).

CTL _ Linfócito T Citotóxico (do inglês, *Cytotoxic T Lymphocyte*).

D1-EM-D2 _ Docking 1-Energy Minimization-Docking 2.

EM _ Minimização de Energia (do inglês, *Energy Minimization*).

HLA _ Antígeno Leucocitário Humano (do inglês, *Human Leukocyte Antigen*)

MHC _ Complexo Principal de Histocompatibilidade (do inglês, *Major Histocompatibility*)

MHC-I _ Complexo Principal de Histocompatibilidade de Classe I (do inglês, *Major Histocompatibility Complex Class I*)

MHC-II _ Complexo Principal de Histocompatibilidade de Classe II (do inglês, *Major Histocompatibility Complex Class II*)

PDB _ *Protein Data Bank*. A sigla “pdb” poderá ser utilizada para se referir ao formato dos arquivos do PDB.

PLC _ Complexo de Carregamento do Peptídeo (do inglês, *Peptide-Loading Complex*)

pMHC _ Complexo peptídeo:MHC.

RE _ Retículo Endoplasmático

TAP _ Transportador Associado ao Processamento de Antígenos (do inglês, *Transporter associated with Antigen Processing*).

TCR _ Receptor de Linfócitos T (do inglês, *T Cell Receptor*).

SUMÁRIO

Resumo	8
Abstract	9
Lista de ilustrações	10
Lista de Gráficos	11
Lista de tabelas	12
Lista de abreviações	13
1 Introdução	15
1.1 Uma visão geral do sistema imune	15
1.1.1 O mecanismo de reconhecimento de antígenos	16
1.1.2 A via de processamento de antígenos	19
1.1.3 Reconhecimento pelos linfócitos T	20
1.2 Bioinformática e sistema imune	22
1.2.1 Bancos de dados e preditores de sequências	23
1.2.2 Ancoramento molecular e conformação de epitopo	24
1.2.3 Aplicações de Redes Neurais Artificiais	26
2 Objetivos	28
3 Materiais e Métodos	29
3.1 Parametrização dos volumes das regiões intersticiais a partir de dados de estruturas	29
3.2 Obtenção de dados e estruturas modeladas.....	32
3.3 Heurística inicial e avaliação dos dados modelados	33
3.4 Treinamento de redes neurais artificiais	34
4 Resultados	35
4.1 Obtenção e análise volumétrica dos cristais	35
4.2 Seleção dos grupos ligantes e não ligantes	36
4.3 Análise completa dos volumes intersticiais nos conjuntos finais de ligantes e não ligantes	38
4.4 Valores de entrada e treinamento de redes neurais artificiais	42
5 Discussão	44
Referências	50
Apêndices	55
Anexos	65

1 INTRODUÇÃO

1.1 UMA VISÃO GERAL DO SISTEMA IMUNE

Nosso ambiente é composto por uma variedade de agentes infecciosos que podem vir a causar algum tipo de patologia e, até mesmo, levar seu hospedeiro à morte. Como uma forma de defesa, organismos desenvolveram diferentes estratégias de defesa, culminando com o sistema imune propriamente dito a partir de vertebrados mandibulados (Kindt et al, 2008). Esse sistema é composto de uma infinidade de células e moléculas acessórias com as mais variadas funções que agem, basicamente, através de duas atividades intimamente relacionadas: reconhecimento e resposta (Abbas et al, 2012; Roitt et al, 2014).

O reconhecimento diz respeito à capacidade do sistema em encontrar invasores, através de mecanismos que são capazes de perceber diferenças sutis entre aquilo que compõe o organismo e aquilo que é estranho a ele. Além do reconhecimento de marcas de infecção, o que culmina com a estimulação da resposta contra esse agente invasor, o sistema imune tem a capacidade de guardar informações sobre estes sinais de perigo. Esse processo é o que chamamos de memória imunológica, o princípio base da vacinação (Kindt et al, 2008).

A resposta trata da neutralização do invasor. O sistema imune pode ser, então, didaticamente categorizado em imunidade inata e imunidade adaptativa. A imunidade inata é formada especialmente por um grupo importante de leucócitos, as células fagocitárias, que se ligam ao microrganismo, os englobam e digerem com alta efetividade, realizando a eliminação dos patógenos poucas horas após seu contato com as suas células (Roitt et al, 2014). A imunidade inata também é composta por barreiras físicas e químicas; sendo exemplo de cada um desses, respectivamente, a pele e as alterações de pH.

A imunidade adaptativa, que constitui a segunda linha de defesa na eliminação dos patógenos que escaparam do sistema inato, pode ser dividida em imunidade humoral, cuja resposta é orquestrada por moléculas proteicas conhecidas como anticorpos; e imunidade celular, cuja resposta é realizada por linfócitos T CD4+ (auxiliares) e linfócitos T CD8+ (citotóxicos). Cabe salientar que os linfócitos T CD4+ também realizam a sinalização, com consequente geração de um ambiente inflamatório, com efeito sobre as células da resposta humoral. O sistema adaptativo

desenvolve resposta contra um agente infeccioso e adapta-se para reconhecer, eliminar e gerar células de memória contra este organismo, onde em uma infecção futura, irá ocorrer o *clearance* de forma mais efetiva e agilizada. Os sistemas inato e adaptativo agem em cooperação, proporcionando uma defesa mais eficaz contra infecções (Abbas et al, 2012; Roitt et al, 2014).

1.1.1 O MECANISMO DE RECONHECIMENTO DE ANTÍGENOS

O reconhecimento de partículas infecciosas no sistema imune adaptativo ocorre, principalmente, pela interação destas com receptores contidos nas células de defesa: receptor de célula B (BCR) nos linfócitos produtores de anticorpos e o receptor de célula T (TCR) nos linfócitos auxiliares e citotóxicos (Zoete et al, 2013; Halle et al, 2017). Neste trabalho o enfoque se dará sobre a resposta celular desencadeada pelos linfócitos T citotóxicos.

Os TCRs interagem com uma molécula proteica denominada Complexo Principal de Histocompatibilidade (MHC). Em humanos, o MHC é traduzido a partir de uma região gênica altamente polimórfica situada no braço curto do cromossomo 6 (posição 6p21.3), e pertence a região da família gênica do antígeno leucocitário humano (HLA), como pode ser visto na Figura 1. (Mungall et al., 2003; Kindt et al., 2008)

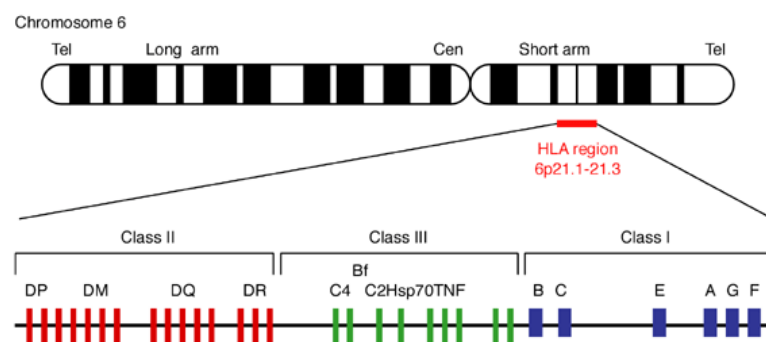


FIGURA 1: Mapa gênico humano da região HLA. A Região do MHC, encontrada no braço curto do cromossomo 6, subdividida em três *loci* principais: o MHC de classe I, o MHC de classe II e o MHC de classe III. (Modificado de Mehra & Kaur, 2003. 4p.)

Há duas classes principais de MHC: MHC de classe I (MHC-I) e MHC de classe II (MHC-II). O MHC-I é composto de uma cadeia α (contendo os domínios α 1, 2 e 3), com tamanho de 45 quilodáltons, codificada por genes polimórficos dentro das regiões A, B e C do HLA. É composto, também, de uma molécula de β_2 -microglobulina (β_2m),

proteína altamente conservada, transcrita de outro cromossomo. A cadeia α e a β_2m se associam de forma não covalente e essa associação é fundamental para garantir a expressão da molécula na superfície celular. (Kindt et al, 2008; Wiczorek et al, 2017)

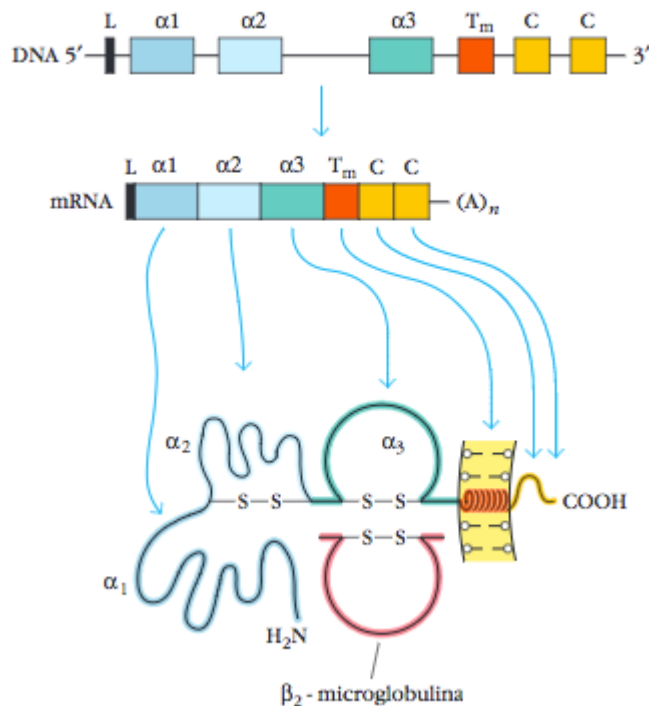


FIGURA 2: Diagrama esquemático do MHC de classe I. Correspondência entre os éxons e os domínios no produto gênico. A β_2 -microglobulina é proveniente de um cromossomo diferente. T_m e C dizem respeito aos domínios transmembrana e citoplasmático, respectivamente. (Adaptado de Kindt et al., 2008)

Na superfície celular, a estrutura apresenta para os linfócitos T CD8+ (citotóxicos) pequenas sequências proteicas (peptídeos) possuindo, em média, entre 8 a 10 aminoácidos, oriundas de proteínas maiores do meio intracelular (Neefjes et al, 2011; Wiczorek et al, 2017). Se esse peptídeo for reconhecido pelo sistema imunológico como sendo pertencente a uma proteína não-própria, uma resposta citotóxica será desencadeada.

Já o MHC-II apresenta, para linfócitos T CD4+, peptídeos maiores (até 30 aminoácidos) oriundos do meio extracelular, gerando uma resposta auxiliar. O produto da sinalização TCR:MHC-II:peptídeo age recrutando mais moléculas e células de defesa para o local da infecção, auxiliando no processo de eliminação do patógeno (Abbas et al, 2012). O MHC-I está presente em todas as células nucleadas (Neefjes et al, 2011), enquanto o MHC-II é restrito a um conjunto específico de células

apresentadoras de antígenos (APC). São exemplos de APCs os macrófagos, as células dendríticas e, também, os linfócitos B (LeibundGut-Landmann et al, 2004).

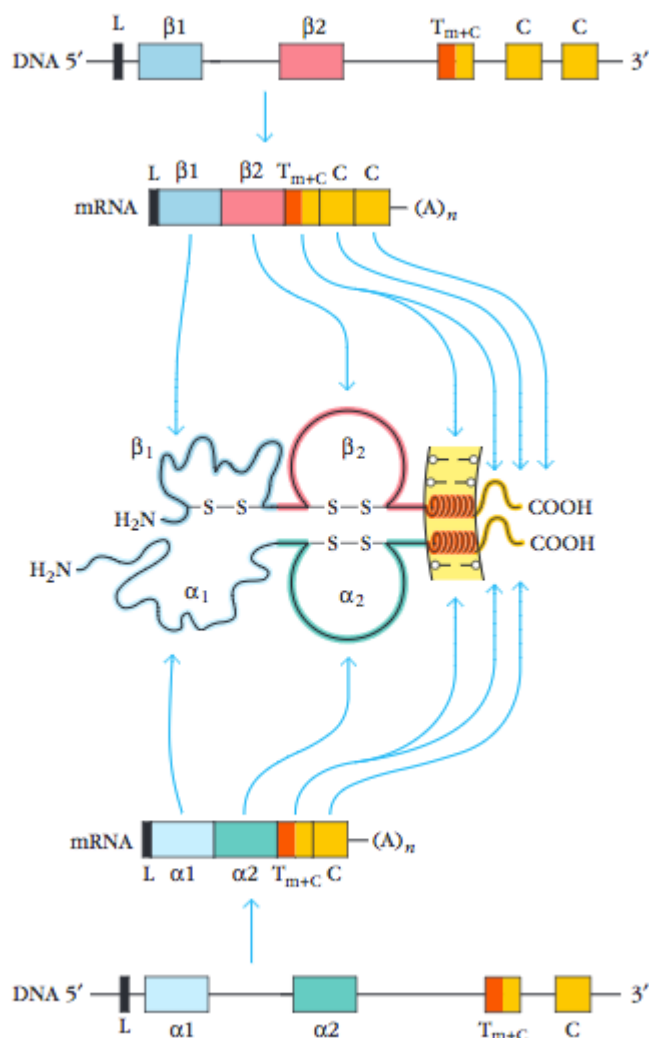


FIGURA 3: Diagrama esquemático do MHC de classe II. Correspondência entre os éxons e os domínios nos produtos gênicos das cadeias α e β . T_m e C dizem respeito dos domínios transmembrana e citoplasmático, respectivamente. (Adaptado de Kindt, et al., 2008)

Seguindo o foco deste trabalho, será dada maior ênfase a interação peptídeo-MHC, com atenção especial na via endógena. Por esse motivo, faz-se necessário compreender o processo como um todo, desde a invasão do patógeno ao ambiente intracelular, até a apresentação de peptídeos oriundos de suas proteínas na superfície da célula. A partir da compreensão destes processos, é que será apresentado o enfoque dado neste trabalho que investiga as características moleculares e estruturais envolvidas no desfecho entre um peptídeo que apresenta uma estabilização da ligação peptídeo-MHC-I (ligante) e peptídeos não aptos a se manter estáveis nestes

complexos.

1.1.2 A VIA DE PROCESSAMENTO DE ANTÍGENOS

Uma das melhores maneiras de perceber como o sistema imune trabalha sobre a vigilância de nossas células em busca de infecções, é analisar a rota de apresentação de antígenos por parte dos MHC's. A via endógena de apresentação de peptídeos apresenta fragmentos de proteínas celulares na superfície da célula através das moléculas de MHC de classe I. Caso ocorra uma infecção celular por vírus, por exemplo, peptídeos virais podem ser gerados e apresentados, a partir do momento em que suas proteínas passam a ser transcritas pela maquinaria celular. É dessa forma que o sistema imune discrimina e elimina uma célula infectada (Swan & Smith, 2007).

Dentre as moléculas envolvidas na via de processamento de antígenos, se destacam o proteossomo e o transportador associado ao processamento de antígenos (TAP). As proteínas que são produzidas pela célula e que já realizaram a sua função, ou que apresentam algum grau de desnaturação, recebem marcação por ubiquitinação e são encaminhadas ao proteossomo para degradação (Lecter et al, 2006).

O proteossomo é um complexo multiproteico formado por um núcleo central (20S) altamente conservado e por subunidades regulatórias (19S ou 11S), formando o proteossomo 26S. A subunidade 19S tem alta afinidade pela ubiquitina. No interior do proteossomo é que ocorre a digestão da proteína, liberando pequenos fragmentos da proteína original. Estes fragmentos proteicos são degradados por proteases citosólicas até que restem apenas aminoácidos ou, no caso de peptídeos, estes são translocados pela TAP para o Retículo Endoplasmático Rugoso (RE) (Murata, et al, 2009; Kish-Trier & Hill, 2013; van de Weijer et al, 2015).

Os peptídeos apresentados pela molécula de MHC-I apresentam a região C-terminal gerada na clivagem pelo proteossomo. Já a extremidade N-terminal não é gerada pelo proteossomo e os peptídeos transportados pela TAP passam ainda por um processo de desbastamento adicional nessa região. Esse processo é realizado em parte por proteases para que possuam o tamanho correto antes de serem carregadas na estrutura parcialmente enovelada do MHC-I. (Falk & Rotzschke, 2002; Neefjes et al., 2011)

O encaixe e dobramento do peptídeo na fenda do MHC é realizado por um complexo de carregamento do peptídeo (PLC) que, além das duas cadeias do MHC-I recém-sintetizado, envolve chaperonas e moléculas acessórias, destacando-se as proteínas Bap31, ERp57, tapasina, além da própria TAP (Praveen et al., 2010). A estrutura parcialmente enovelada do MHC-I interage, então, com a TAP e recebe o peptídeo direcionando-o para o interior do RE. Essa interação estabiliza o MHC-I que agora pode ser transportado até a superfície celular, através do aparato do Complexo de Golgi. (Peaper & Cresswell, 2008; Anders et al, 2011; Neefjes et al, 2011)

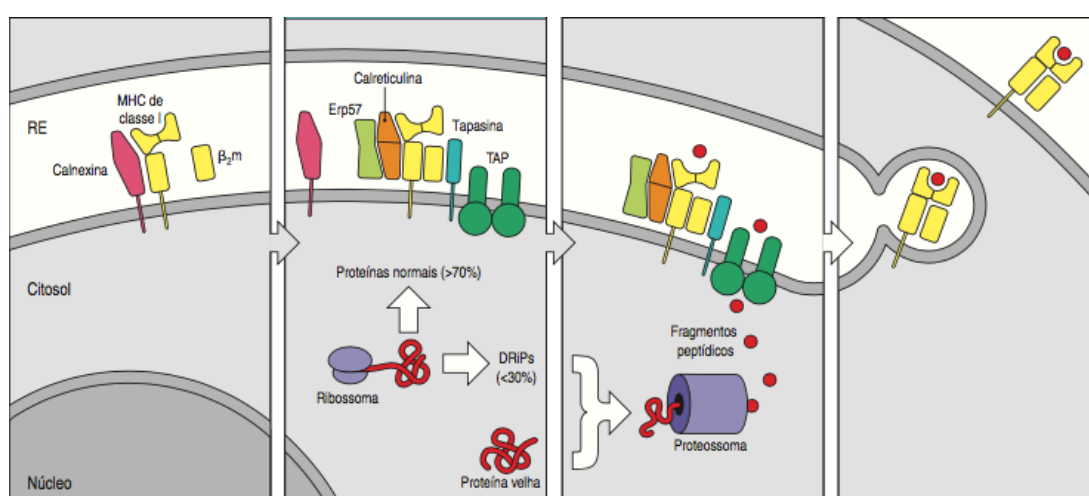


FIGURA 4: Via endógena da apresentação de antígenos. Os fragmentos de proteínas próprias ou virais são clivados pelo proteossomo e ficam disponíveis no citosol. Esses fragmentos são, então, encaminhados para o retículo endoplasmático, onde são associados ao MHC-I. Após são direcionados o Complexo de Golgi que se encarrega da apresentação do complexo na superfície celular. (Adaptado de *Murphy et al, 2010. p. 188*)

Resumidamente, o proteossomo cliva proteínas endógenas ubiquitinadas em partículas menores, que serão desbastadas até que adquiram o tamanho aproximado de nove aminoácidos, tamanho preferencial para ligação dentro da fenda do MHC-I (Paul et al, 2013). Partículas associadas ao processamento de antígenos (TAP), levam o peptídeo ao interior do RE. Nesse momento, chaperonas ajudam o complexo pMHC em seu enovelamento e associação com a molécula β_2m .

Após todo o processo de montagem do complexo pMHC, e de seu encaminhamento a superfície da célula, a molécula encontra-se pronta para interagir com o TCR.

1.1.3 RECONHECIMENTO PELOS LINFÓCITOS T

Os linfócitos T CD8⁺ sofrem um processo de seleção no timo, mediante competição por estímulos. Linfócitos que não forem capazes de interagir com moléculas de MHCs são eliminados. Linfócitos auto-reativos, sofrem seleção negativa. Restando apenas as células T que possuem afinidade moderada ou baixa por peptídeos próprios (Antunes, 2011). É dessa maneira que os TCR's se tornam capazes de interagir efetivamente e reconhecer alvos não próprios, no contexto do MHC-I (Sohn et al, 2007). Tal reconhecimento que, em linhas gerais, leva a eliminação da célula alvo.

O receptor de uma célula T é um composto heterodímero formado de uma cadeia α e uma cadeia β , ou de uma cadeia γ e uma cadeia δ . Em cada uma dessas cadeias há três regiões hipervariáveis chamadas de Regiões Determinantes de Complementariedade (CDR) que realizam o contato com resíduos do MHC e do peptídeo. (Brehm et al., 2004; Parham, 2005).

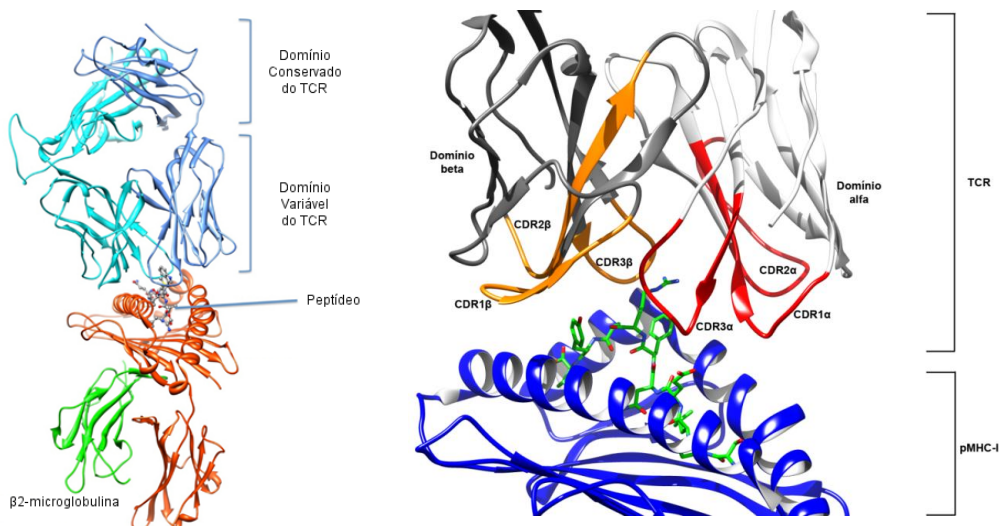


FIGURA 5: Interação entre TCR e complexo pMHC. Ressaltadas as regiões de alta variabilidade do TCR e sua interação com a região da fenda do MHC. (Adaptado de Rigo, 2015a. p.18)

Uma série de moléculas acessórias auxilia no processo de estabilização da interação o que influencia diretamente na correta "leitura" do complexo pMHC-I, pelo TCR, para que o linfócito T receba o estímulo e desencadeie uma resposta de maneira correta. A sinapse imunológica corresponde a uma região de agrupamento das tríades formadas pelos complexos pMHCs/TCRs a qual mantém próximas a célula

apresentadora e o linfócito envolvido na resposta (Chen et al., 2009; Thauland & Parker, 2010; Saito et al., 2010).

Após esta breve introdução sobre a resposta imune celular, podemos pensar nas diversas formas de abordagem deste processo, contribuindo dessa maneira para uma melhor compreensão das regras que regem este sistema e que possam levar a eficácia na resolução de processos patológicos, como as infecções virais. Entretanto, um ponto que deve ser considerado é o elevado custo para a realização de ensaios *in vitro/in vivo* neste tipo de estudo, considerando a magnitude das possíveis variáveis envolvidas nas respostas efectoras do sistema imune. Dessa maneira é que estudos computacionais se mostram úteis. No próximo item, após uma breve introdução sobre o início da imunoinformática, são enumeradas ferramentas para a predição dos processos da via do MHC, em especial a via endógena, de modo individual ou combinado, sendo que algumas delas foram aplicadas neste trabalho.

1.2 BIOINFORMATICA E SISTEMA IMUNE

Com a criação do *International Immunogenetics Information System*¹ (IMGT) em 1989, surgiu uma nova interface entre a imunologia e a bioinformática, denominada imunoinformática. Essa área surgiu da necessidade de entendimento acerca do funcionamento do sistema imune, aliada à dificuldade de experimentação *in vitro* e *in vivo*, fazendo com que os esforços se concentrem em gerar ferramentas computacionais que auxiliem na compreensão desses processos de forma mais rápida e com menor custo. (Lefranc et al, 1999; Lefranc et al, 2015)

A imunoinformática trabalha diversas áreas dentro do espectro imunológico, como os estudos de interação entre patógenos e hospedeiros, predição de alvos de células B e T, predição de alvos para vacinação, identificação de genes, criação e manutenção de bancos de dados entre outros. A partir daí, se desenvolveu a área da vacinologia reversa (Sette & Rappuoli, 2010; Tomar & De, 2010).

A metodologia tradicional de produção de vacinas trata da aplicação de técnicas bioquímicas e sorológicas para identificação de patógenos. Todavia, nem sempre é trivial que esses patógenos sejam cultivados *in vitro*. Assim, a vacinologia reversa traz, *in silico*, a possibilidade de identificação dos antígenos candidatos que

¹ Acessível em: <<http://www.imgt.org/>>

podem estar envolvidos na causa, ou no reconhecimento (como *targets*), de uma determinada doença (Rappuoli, 2000; Davies & Flower, 2007; Donati & Rappuoli, 2013). Não são poucas as ferramentas disponíveis com essa finalidade, principalmente para a via do MHC-I: preditores de sequências de peptídeos gerados pelo proteossomo, passando pelas vias TAP, até a docagem de um peptídeo e avaliação de sua afinidade com o receptor. Nas seções a seguir são apresentadas algumas das ferramentas de predição existentes para o entendimento da ligação do peptídeo ao MHC-I.

1.2.1 BANCOS DE DADOS E PREDITORES DE SEQUÊNCIAS

A disponibilidade crescente de informações faz com que novas técnicas e bancos de dados imunológicos surjam ou sejam ampliados de modo a comportá-las. Esses dados podem então ser utilizados para treinar métodos de aprendizagem de máquina com o objetivo de prever cada um destes processos. Dentre os vários programas e bancos de dados existentes para esses fins, podemos citar:

- NetChop 3.1: um servidor online baseado em redes neurais e dados experimentais com a finalidade de prever a clivagem proteica realizada pelo proteossomo, a partir da sequência linear de uma proteína. (Nielsen et al., 2005)
- TAPPred: uma ferramenta com objetivo de prever a afinidade de peptídeos gerados pelo proteossomo à proteína TAP. (Bhasin & Raghava, 2004)
- NetMHC e BIMAS: ambos direcionados para a predição de afinidade de ligação entre epitopo e MHC-I. (Lundegaard et al., 2008; Parker et al., 1994)
- *Immune Epitope Database* (IEDB) permite a análise conjunta de todos os passos da via de processamento de antígenos endógenos citados anteriormente. Todavia, mais do que uma plataforma para ferramentas de imunoinformática, o IEDB é um dos principais bancos de dados experimentais relacionados a epitopos testados para imunogenicidade de células T e B. (Fleri et al, 2017)

- CrossTope: um repositório de estruturas tridimensionais de complexos pMHC de classe I, obtidos por modelagem *in silico*. (Sinigaglia et al, 2013)

Excetuando-se a plataforma de imunoinformática CrossTope, que contém e utiliza informação estrutural para estudar e desenvolver ferramentas com enfoque imunológico, quase a totalidade dos preditores são baseados em análise de sequências. Apesar de extremamente úteis, deixam de lado algumas informações importantes a respeito da estrutura de uma molécula e sua influência sobre a ocorrência ou não de determinado evento. No caso da via de processamento de antígenos, a partir do momento em que um peptídeo é clivado e ligado a fenda do MHC, este passa a fazer parte de um complexo maior, que é composto pela combinação atômica do peptídeo original com a do alelo do MHC no qual ele está ligado. Para que estas características possam ser consideradas é necessário utilizar outros métodos computacionais de modelagem molecular, os quais podem simular a conformação e comportamento destes complexos, tais como: ancoramento e dinâmica molecular.

1.2.2 ANCORAMENTO MOLECULAR E CONFORMAÇÃO DE EPITOPOS

A associação de peptídeos às moléculas de MHC é uma interação com lenta taxa de associação e de dissociação, além de saturável. A baixa taxa de dissociação entre o peptídeo e as moléculas do MHC permite que os complexos pMHC persistam complexados por tempo suficiente na superfície das células apresentadoras de antígenos, assegurando interações produtivas com as células T antígeno-específicas. (Murphy, et al. 2010)

O encaixe do peptídeo na fenda do MHC, que se dá entre as duas hélices, é baseado na formação das ligações de hidrogênio entre as cadeias laterais do MHC-I e o *backbone* ou cadeia principal do peptídeo e, especificamente, no correto encaixe das cadeias laterais do peptídeo nos *pockets* do MHC-I. Os *pockets* são regiões onde a ligação do MHC-I com os resíduos do peptídeo funcionam como âncoras do peptídeo dentro da fenda. No caso do MHC-I esses âncoras seriam os de posição P2 ou P5/6 e P Ω (terminal), dependendo do alelo analisado. No caso do MHC-II, seriam as posições P1, P4, P6, e P9. (Wieczorek, 2017)

A natureza da interação do peptídeo com o MHC-I vai depender da sua geometria, distribuição de cargas e a hidrofobicidade da fenda (Antes; Siu; Lengauer, 2006). Sendo assim, a maior parte das técnicas de docagem molecular se valem dessas informações. Com relação à geometria do peptídeo, sabe-se que a fenda do MHC-I pode conferir a ele um formato específico, dependendo do alelo (Antunes et al, 2010).

Levando em consideração as metodologias desenvolvidas para *docking* estrutural para construção de complexos pMHC, é possível destacar as ferramentas DynaPred (Antes; Siu; Lengauer, 2006), e DockTope (Rigo et al, 2015), para MHC-I, EpiDock (Atanasova et al, 2013) para MHC-II e pDOCK (Khan & Ranganathan, 2010), para ambas as moléculas de MHC. Cada uma dessas ferramentas solicita como dado de entrada as sequências dos epitopos de interesse e retorna um arquivo que contém as coordenadas tridimensionais do complexo gerado.

A ferramenta DockTope, em especial, faz uso da técnica D1-EM-D2; uma técnica para docagem de peptídeos em fendas de MHC de classe I, sendo necessária apenas a sequência FASTA do epitopo para gerar o modelo. O algoritmo assume então um padrão conformacional para o epitopo, de acordo com o alelo de interesse para a modelagem. Esses padrões foram obtidos a partir da análise de cristais dos alelos de interesse (Antunes et al, 2010; Antunes, 2014; Rigo, 2015b)

A partir disso e utilizando o pacote Autodock Vina, uma primeira docagem (D1) é realizada. Após um passo de minimização de energia (EM) para acomodação das cadeias laterais das estruturas de epitopo e do MHC-I, é realizada uma segunda docagem (D2) para que haja o refinamento final da estrutura. O arquivo final, em formato PDB, contém as coordenadas tridimensionais do complexo de interesse. Um esquema visual do processo pode ser visto na Figura 6, a seguir:

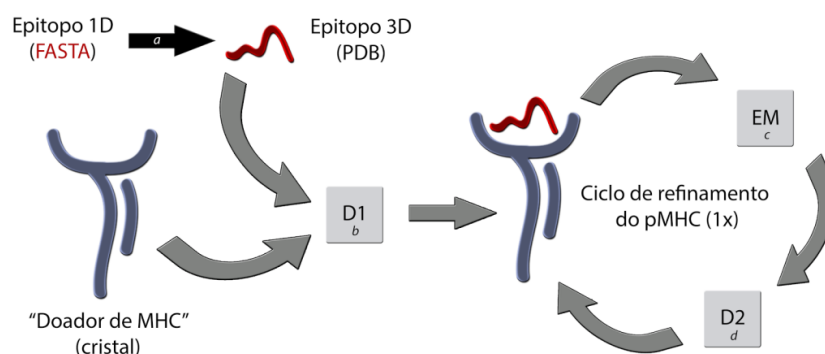


Figura 6: Esquema visual da técnica D1-EM-D2. A estratégia de docagem de peptídeos em MHC-I consiste em 4 etapas. Como entrada para a criação dos modelos é solicitada apenas a sequência fasta

do peptídeo (a), que receberá uma conformação pré-determinada. O peptídeo sofre uma primeira docagem (b) seguido de um ciclo de minimização de energia (c). É então retirado da fenda do MHC e recolocado em uma segunda docagem, para que as cadeias laterais se ajustem (d). (Adaptado de Antunes, 2014)

Essa abordagem permite gerar complexos pMHC a partir de sequências de interesse, permitindo uma análise estrutural, valendo-se de um processo rápido, confiável e de baixo custo.

1.2.3 APLICAÇÃO DE REDES NEURAIS ARTIFICIAIS

As abordagens apresentadas anteriormente são ferramentas poderosas que podem ser aplicadas à imunoinformática. Entretanto, o número de dados gerados a partir delas é de difícil manuseio para extração de informações, mesmo utilizando análises robustas de matemática e estatística. Uma alternativa interessante nestes casos é o uso de métodos de aprendizagem de máquina, dentre eles as redes neurais artificiais (ANN). A ideia base da aplicação de redes neurais artificiais, vem de uma área conhecida como neurocomputação. Seu desenvolvimento inicial aconteceu depois que McCulloch e Pitts, um neuroanatomista e um matemático, simularam o comportamento de um neurônio utilizando resistores e amplificadores (McCulloch & Pitts, 1943).

Computacionalmente, ANNs são modelos capazes de reconhecer padrões e, assim, aprender o comportamento de um conjunto de dados observados. (Ripley, 1995). Uma ANN é composta de neurônios de processamento, cada um deles com dois ou mais receptores de estímulos (entradas). Estes são os encarregados de perceber um sinal. Possui ainda um corpo de processamento, ou camada interna, capaz de se reprogramar dependendo dos dados de entrada. Por fim, cada neurônio apresenta uma saída 1 ou 0 (sim ou não) dependendo do que é processado. A principal característica de um neurônio é que ele é capaz de realizar apenas um processamento. Isso, porque, apesar de poder possuir várias entradas, cada entrada só pode perceber um tipo de estímulo. Essa estrutura permite a rede habilitar ou desabilitar conexões de forma dinâmica (Hebb, 1949).

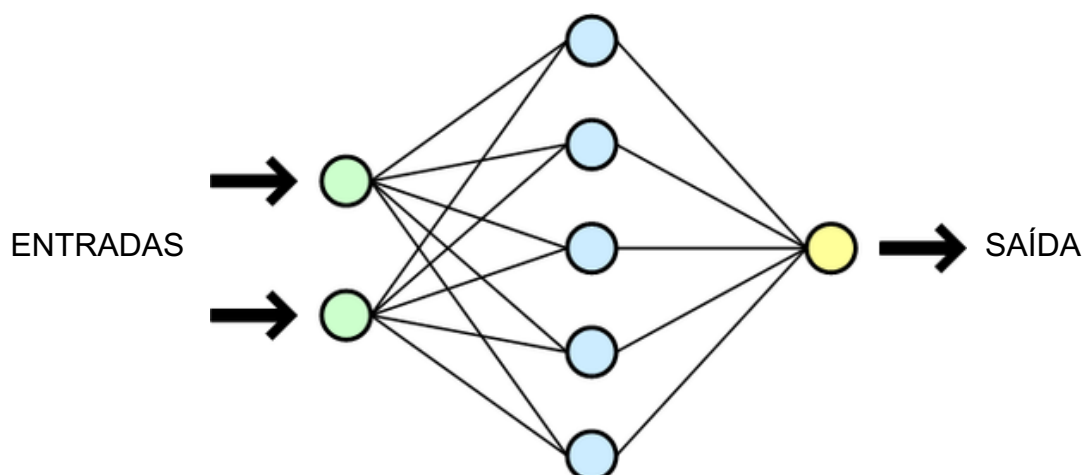


FIGURA 7: Representação gráfica de uma rede neural. Uma rede neural artificial é composta de grupos de nodos, interconectados. Aqui, cada nodo representa um neurônio artificial e cada conexão entre o output de um neurônio e o input de outro é representado pelas setas. No caso deste trabalho as entradas são os valores de volume para sequências escolhidas como ligantes e não ligantes.

Em imunoinformática, as ANNs têm se mostrado úteis, especialmente no desenvolvimento de preditores que utilizam unicamente sequências de aminoácidos. Muitos métodos de avaliação de ligação de proteínas e epitopos utilizam o treinamento de redes neurais, ou métodos de aprendizagem de máquina, como um todo.

A apresentação inicial dos conceitos, ainda que de forma breve, é importante para dar suporte mínimo à fundamentação da metodologia apresentada a seguir. No contexto do sistema imune se mostra importante entender os processos de ligação de um peptídeo ao MHC, para que o desenvolvimento da resposta à uma infecção seja plenamente compreendido. Há algum tempo procura-se entender melhor essas questões levando em consideração abordagens estruturais. Assim, dada a complexidade da via de apresentação de antígenos em si, cada avanço realizado se mostra importante.

É nesse sentido que este trabalho aborda a análise das áreas de interação entre o peptídeo e o MHC. A intenção é compreender como essa interação ocorre e se há algum tipo de diferenciação entre características moleculares dos ligantes, levando em consideração as estruturas envolvidas.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Inferir se existe uma diferença no espaço intersticial dos complexos pMHC, entre grupos de ligantes e não ligantes.

2.2 OBJETIVOS ESPECÍFICOS

Averiguar se esta diferença reside no volume total da cadeia peptídica principal ou se está restrita a alguma posição específica da cadeia.

Utilizar os melhores discriminadores para treinar um método de aprendizagem de máquina capaz de prever bons e maus ligantes.

3 MATERIAIS E MÉTODOS

3.1 PARAMETRIZAÇÃO DOS VOLUMES DAS REGIÕES INTERSTICIAIS A PARTIR DE DADOS DE ESTRUTURAS CRISTALOGRÁFICAS

O passo inicial, após o desenho da pesquisa, foi a avaliação das fendas de cristais de MHC's do alelo HLA-A*02:01. Foram selecionados cristais de alta resolução para o alelo especificado. As estruturas foram obtidas no *Protein Data Bank*² (RCSB PDB) tendo sido filtradas por:

- MHC de classe I;
- HLA-A*0201.

Como critério adicional foram escolhidos apenas alvos com ligantes complexados ao MHC, bem como ausência de TCR na estrutura, para que não houvesse qualquer influência externa na interação entre o peptídeo e o MHC.

Foram recuperados arquivos em formato PDF para o caso de informações biológicas, FASTA para informações a respeito de sequências e, em formato PDB, para as coordenadas tridimensionais (x,y,z) das estruturas em questão. Cada um dos domínios alfa, beta e peptídeo, são designados por uma cadeia A, B e C, respectivamente. Como o enfoque deste trabalho é a análise dos volumes resultantes da interação entre a fenda do MHC e o peptídeo, utilizamos as cadeias A e C para a realização dos cálculos. Cada uma dessas estruturas foi avaliada utilizando o software *UCSF Chimera*³, um aplicativo para visualização e análise de estruturas tridimensionais.

ATOM	2244	CD	GLU	A	275	9.943	-33.682	-4.986	1.00	29.58	C
ATOM	2245	OE1	GLU	A	275	10.919	-34.282	-5.526	1.00	30.51	O
ATOM	2246	OE2	GLU	A	275	10.088	-32.875	-4.018	1.00	30.54	O
ATOM	2247	OXT	GLU	A	275	5.042	-34.393	-5.945	1.00	25.72	O
TER	2248		GLU	A	275						
ATOM	2249	N	MET	B	0	0.020	16.618	-18.694	1.00	25.89	N
ATOM	2250	CA	MET	B	0	0.296	16.126	-17.311	1.00	25.74	C
ATOM	2251	C	MET	B	0	0.483	14.582	-17.294	1.00	24.43	C

Figura 8: Estrutura em texto plano de um arquivo PDB. Os quadros azul e verde ressaltam os resíduos e as cadeias aos quais pertencem cada átomo, respectivamente. No quadro em vermelho podemos visualizar as coordenadas tridimensionais (x, y, z) da estrutura.

² O *Protein Data Bank* está disponível para acesso no endereço: <<http://www.rcsb.org>>

³ UCSF Chimera está disponível para download e documentado em: <<https://www.cgl.ucsf.edu/chimera/>>

O software UCSF Chimera é capaz de interpretar o arquivo PDB, gerando a visualização tridimensional da estrutura modelada. O aplicativo contempla uma diversidade de ferramentas que são disponibilizadas para análise das estruturas de nosso interesse (Pettersen, 2004).

Inicialmente, foi realizada a busca por *gaps* (espaço resultante entre dois conjuntos de átomos) e o cálculo do volume desses *gaps* para cada resíduo (aminoácido), como pode ser visto nas figuras 9 e 10 a seguir.

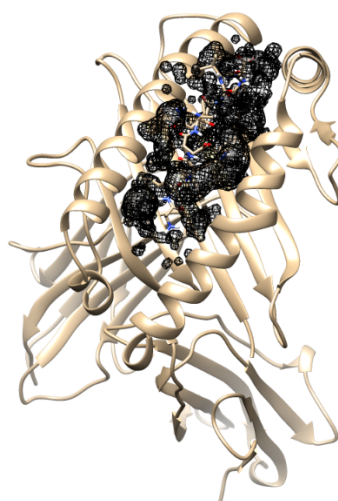


Figura 9: Volume intersticial para a cadeia completa do epitopo complexado ao Cristal de MHC-I. O volume foi calculado tendo como referência a cadeia α e computando os valores de volumes para o epitopo completo.

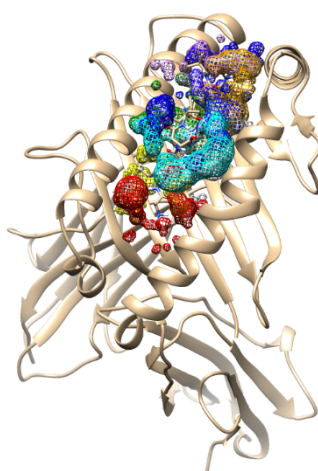


Figura 10: Volume intersticial para cada resíduo do epitopo complexado ao Cristal de MHC-I. Cada volume foi calculado tendo como referência a cadeia α e computando os valores de volumes individualmente para cada aminoácido.

Os módulos que permitiram o cálculo volumétrico nos gaps das estruturas cristalográficas foram o *Surfnet* e *Measure Volume and Area*. O primeiro trata-se de um programa escrito inicialmente em linguagem Fortran, adaptado em Python, que através de uma sonda de tamanho variável (0.1 – 1.0 angstroms) busca por espaços disponíveis entre as moléculas envolvidas na interação, utilizando as informações das coordenadas tridimensionais fornecidas pelo arquivo PDB (Laskowski, 1995).

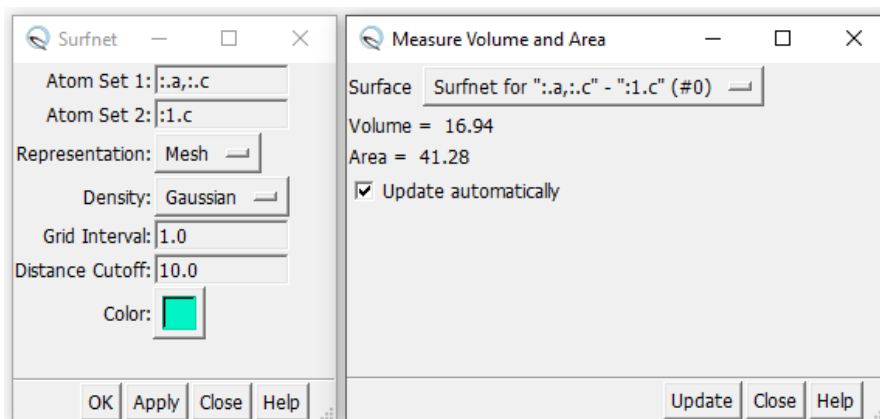


Figura 11: Parâmetros da janela de extração de volumes via SURFNET. Cada aminoácido do epitopo era referenciado por um valor de 1-9, da cadeia C. Para computar o valor para a cadeia C completa, basta remover qualquer referência aos resíduos específicos. Os valores calculados pelo Surfnet são exibidos pelo módulo *Measure Volume and Area*.

Considerando que, se um peptídeo está complexado à fenda do cristal ele é um ligante, para formar um conjunto de não-ligantes como grupo negativo de comparação, pensou-se em mutar esses peptídeos de acordo com a informação do banco de dados *Imune Epitope Data Base*⁴ (IEDB) de que determinados aminoácidos inviabilizam a ligação (Figura 11). Todavia, o espectro possível de mutações era elevado e nada garantiria que esses peptídeos manualmente mutados seriam realmente não-ligantes. Foi definido, então, que os grupos de ligantes e não ligantes seriam recuperados de dados de experimentação *in vitro* depositados no IEDB.

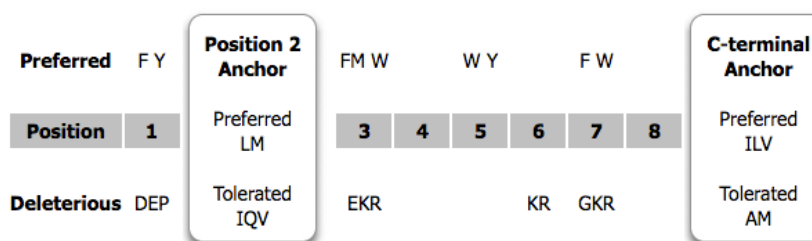


Figura 12: Motivos do MHC humano HLA-A*02:01. Aminoácidos preferidos, tolerados e deletérios nas nove posições para este respectivo alelo. Acessível em: <http://www.iedb.org/MHCalleleId/143>

⁴ O IEDB está disponível para consulta em: <<http://iedb.org>>

3.2 OBTENÇÃO DOS DADOS E ESTRUTURAS MODELADAS

O IEDB é um banco de dados de epitopos imunogênicos, mantido pelo instituto *La Jolla for Allergy and Immunology*, com foco na disseminação de informação para facilitar a produção de novas ferramentas para pesquisa, diagnóstico, e produção de vacinas. Nele se encontram depositados dados relacionados a anticorpos, epitopos de célula T para humanos, primatas e outras espécies animais, bem como dados de ligação ao MHC e imunogenicidade de epitopos, compreendendo diversas fontes antigênicas. (Vita et al, 2014)

Para a obtenção dos grupos ligantes e não ligantes foram utilizados os seguintes critérios de busca:

- MHC-I;
- Alelo HLA-A*0201;
- High-positive (para ligantes);
- Negative (para os não-ligantes).

Os parâmetros *high-positive* e *negative*, se referem ao tipo de resposta imunogênica que o epitopo alvo atingiu em ensaios para essa finalidade. Sabe-se que nem todo epitopo negativo para imunogenicidade será necessariamente um mau ligante. Dessa maneira, todos os peptídeos inicialmente selecionados foram avaliados quanto ao seu IC50 descrito no banco. Essa medida se trata de uma media da inibição de um composto (50% de concentração inibitória máxima). Para casos de ensaios de ligação competitiva, o IC50 é a medição sumária mais comum da curva de dose-resposta. Podendo, então, ser interpretado como um ensaio de ligação, especialmente para os casos estudados neste trabalho (Neubig, 2003).

Os peptídeos selecionados nos grupos de ligantes e não ligantes foram então modelados com a ferramenta DockTope, contendo a estrutura do MHC de interesse com as sequências selecionadas. As estruturas geradas foram visualmente inspecionadas para averiguação da correta sequência modelada.

3.3 HEURÍSTICA INICIAL E AVALIAÇÃO DOS DADOS MODELADOS

Os gráficos gerados e as análises estatísticas iniciais foram produzidos através do software MATLAB⁵ (*Matrix Laboratory*), um software de alta performance voltado para o cálculo numérico, que é capaz de analisar, processar e construir matrizes e gráficos sem necessidade de redimensionamento das matrizes dos dados.

Depois das análises iniciais, foi possível notar que os dados relacionados aos âncoras dos peptídeos poderiam prover informações mais precisas sobre o comportamento esperado para cada caso, ligantes e não ligantes. Assim sendo, foram avaliadas, mais uma vez, as tabelas de dados obtidas através do IEDB, dessa vez, aplicando critérios mais rigorosos, de modo a escolher amostras bastante distintas para os dois grupos estudados, sendo eles:

- Apenas dados publicados;
- Apenas 9(mer);
- Testados para IC50;
- Removidas as sequências redundantes.

Como não seria possível modelar todos os alvos, selecionamos os alvos com os menores valores de IC50 (menores que 0,5nM) para ligantes e os com maiores valores para não-ligantes (valores iguais ou superiores a 70.000nM). De acordo com a literatura, são considerados ligantes, alvos com índice de IC50 de até 500nM (Paul et al, 2013). As análises estatísticas desses dados foram feitas no software R⁶.

Todos as sequências modeladas tiveram seus complexos submetidos ao *Prodigy*⁷ (Xue et al, 2016), uma ferramenta online para cálculo do potencial de dissociação entre as estruturas. O cálculo de dissociação pode ser entendido como a afinidade da interação entre duas proteínas, sendo descrita através da constante de dissociação de equilíbrio K_d , ou em termos da termodinâmica, pela energia livre de Gibbs, ΔG .

$$\Delta G = RT \ln(k_d)$$

onde: R é a constante equivalente a constante de Boltzmann, T é a temperatura do ambiente em K, e k_d é a constante de desequilíbrio. Esses dados foram usados

⁵ Licença disponibilizada pela Universidade Federal do Rio Grande do Sul.

⁶ O software está disponível para download no endereço: <<https://cran.r-project.org/>>

⁷ O PRODIGY pode ser acessado no endereço: <milou.science.uu.nl/services/PRODIGY/>

posteriormente para conferência individual de sequências com comportamento considerado aberrante.

3.4 TREINAMENTO DE REDES NEURAS ARTIFICIAIS

Após todas as análises anteriormente descritas, foi construído o algoritmo para o treinamento de redes neurais que fossem capazes de classificar, com base nos dados obtidos, se um peptídeo seria um ligante ou um não ligante. O treinamento das redes neurais artificiais se deu com os últimos conjuntos de dados extraídos, levando em consideração apenas as posições estatisticamente significantes nas separações dos grupos.

Foram selecionados, de forma randômica, 70 alvos positivos para o conjunto de treino e os valores referentes as outras 15 sequências para o conjunto de validação. O mesmo para o conjunto de dados negativos. Todos os dados foram normalizados para valores entre zero e um para que não houvesse uma flutuação muito grande nos decimais tratados.

4 RESULTADOS

4.1. OBTENÇÃO E ANÁLISE VOLUMÉTRICA DOS CRISTAIS

Foram obtidos os arquivos PDB de oito cristais diferentes referenciados pelos códigos: 1DUZ, 1HHG, 2X4O, 2X4R, 2X4S, 3MRE, 3MRH e 4K7F⁸. Todas as estruturas apresentaram valores de resolução inferiores a 2 angstroms. Esse subconjunto com um pequeno número de cristais é decorrente do preenchimento dos diferentes critérios de exclusão definidos na seção materiais e métodos.

Os volumes intersticiais formados entre a cadeia peptídica e a fenda do MHC foram calculados para todos os cristais (cadeia principal do peptídeo e por aminoácidos) e podem ser encontrados na Tabela 1.

Tabela 1: Dados da extração de volumes de cristais para resíduos individuais do epitopo da estrutura cristalográfica e do volume total do epitopo.

Resíduos											
IEDB	Sequência	1	2	3	4	5	6	7	8	9	Cadeia C
1DUZ	LLFGYPVYV	118.7	174.8	258.5	474.2	762.4	641.3	177.8	452.1	48.4	486.1
1HHG	TLTSCNTSV	179.4	222.4	401.3	385.9	522.5	724.7	302.9	293.8	30.56	744
2X4O	KLTPLCVTL	163.8	152.2	236.1	509.9	708.8	750.5	208.1	334.6	49.52	709.5
2X4R	NLVPMVATV	175.2	199.7	311.5	400	817.3	461.6	226.7	171.1	27.85	560.1
2X4S	AMDSNTLEL	92.37	117.3	307	647.3	526	733.8	225.5	314.4	28.49	628
3MRE	GLCTLVAML	74.44	68.5	271.8	529.1	454.5	744.3	326.3	436.7	35.34	183.6
3MRH	CISGVCWTV	119.1	100.1	286.1	592	672.9	637.6	516	203	5.381	591.5
4K7F	VCWGELMNL	113.5	120.2	296.9	600.8	867.8	686	416.1	409.2	60.7	445.2

FONTE: Dados de pesquisa.

⁸ Referência para os cristais: 1DUZ (Khan, A.R. et al. 2000); 1HHG (Madden, D.R. et al., 1993); 2X4O, 2X4R, 2X4S (Celie, P.H.N. et al., 2009); 3MRE, 3MRH (Machillot, P. et. al, 2014); 4K7F (Zhang, Y. et. al, 2013).

4.2. SELEÇÃO INICIAL DOS GRUPOS LIGANTES E NÃO LIGANTES

Após a filtragem definida na seção 3.2, foram obtidas duas tabelas: uma para *high-positive* com um total de 4.414 alvos e outra, para *negatives*, com 3.850 alvos. Essas tabelas foram então ordenadas para evidenciar as sequências com maiores e menores valores de IC50 em cada grupo. A partir disso, foram gerados dois conjuntos compostos dos primeiros vinte e cinco alvos de cada tabela apresentando os valores mais extremos de IC50.

Das cinquenta sequências selecionadas para modelagem, duas não foram passíveis de serem geradas pelo processo (uma de cada grupo), resultando num conjunto de 48 modelos pMHC. Os resultados dos volumes calculados para todas estas estruturas, separados pelos diferentes grupos, podem ser vistos na Tabela Suplementar 1 e Tabela Suplementar 2.

Após a extração dos dados, foram calculadas as médias e os desvios padrão para cada resíduo que compunha o peptídeo e o mesmo para a cadeia completa (Tabelas 2 e 3). A intenção foi compreender o comportamento específico para cada posição e averiguar se havia algum indício sobre esse comportamento que pudesse ser inferido apenas analisando esses valores. Posteriormente aos cálculos foram plotados os dados como é possível observar no Gráfico 1, que apresenta um panorama representativo do comportamento dos volumes calculados para o epitopo (cadeia C) contra o MHC.

Tabela 2: Estatística primária para alvos positivos.

	Resíduo									Cadeia C
	1	2	3	4	5	6	7	8	9	
Media	144.5	161.9	454.1	666.2	872.9	661.5	401.6	342.0	38.5	548.0
Desvio	22.4	7.9	68.1	63.2	84.4	21.7	19.7	23.7	8.4	65.5
Mínimo	109.3	143.9	345.8	567.4	765.5	626.8	366.3	290.8	28.35	377.7
Máximo	204.3	182.9	574	829	1068	694.6	440.8	382.2	58.03	643.9
Mediana	143.2	161.4	441.0	683.3	875.8	664.2	399.9	341.1	35.4	548.9

FONTE: Dados de pesquisa.

Tabela 3: Estatística primária para alvos negativos.

Resíduo										
	1	2	3	4	5	6	7	8	9	Cadeia C
Media	198.1	171.6	479.2	659.7	838.2	669.6	442.2	367.8	46.2	608.3
Desvio	92.5	11.2	58.3	73.0	63.2	32.4	79.0	92.4	16.6	67.0
Minimo	121.5	154.1	386.6	563.5	723.4	598.6	364.6	289.7	22.9	476.5
Maximo	460.3	200.3	603.6	799.3	963.7	723.1	767.2	693.4	91.6	737.6
Mediana	166.9	169.1	460.7	671.4	833.2	667.2	425.9	340.7	41.3	603.4

FONTE: Dados de pesquisa.

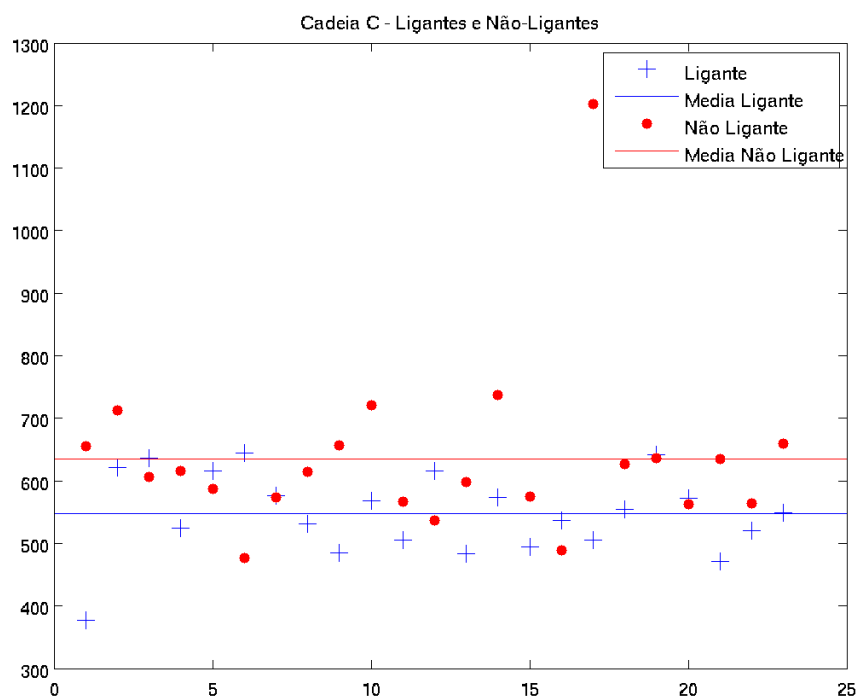


Gráfico 1: Plot dos dados modelados para cadeia C do complexo pMHC. Plot dos dados de volume (eixo vertical) para a cadeia completa do peptídeo. Os alvos (eixo horizontal) tidos como ligantes (positivos para teste de IC50) podem ser vistos em azul, e os não ligantes (negativos para teste de IC50) em vermelho.

4.3. ANÁLISE COMPLETA DOS VOLUMES INTERSTICIAIS NOS CONJUNTOS FINAIS DE LIGANTES E NÃO LIGANTES

O conjunto mais estrigente de filtros da seção 2.3 diminuiu a tabela inicial de 4416 para 756 alvos a serem considerados ligantes. Os mesmos filtros foram aplicados aos de resposta negativa, diminuindo a tabela de 3852 para 2267 alvos. Os limiares $<0.5\text{nM}$ e >70.000 foram os critérios finais de seleção. Restaram, então, 95 alvos positivos para ligação a serem modelados da mesma forma que na seção anterior. Por questões de tamanho amostral, recuperamos os 95 alvos da tabela de negativos apresentando os piores valores de ligação para compor o grupo de alvos não-ligantes.

Como na Figura 13 da seção 4.2, foram gerados os *plots* com os volumes da cadeia C para os 190 peptídeos selecionados como ligantes e não ligantes (Gráficos 2 e 3). Utilizando os testes de Wilcoxon-Mann-Whitney e Equações de Estimativas Generalizadas encontramos significância tanto para a cadeia C como para as posições 2, 7 e 9 dos peptídeos como bons discriminadores de ligação (Gráficos 4, 5, e 6). Imagens com os valores ordenados para estas posições são apresentadas na Figura 15.

TABELA 4: Resultados do teste de Wilcoxon-Mann-Whitney por resíduo e cadeia C.

Posição	W	p	p corrigido	h1
1	3957	0.1427	0.1431	true location shift is not equal to 0
2	1618.5	2.25E-14	2.27E-14	true location shift is not equal to 0
3	3665	0.02534	0.02543	true location shift is not equal to 0
4	5359	0.02552	0.02561	true location shift is not equal to 0
5	4971	0.2264	0.2269	true location shift is not equal to 0
6	3784.5	0.05476	0.05492	true location shift is not equal to 0
7	3345	0.002067	0.002076	true location shift is not equal to 0
8	4403	0.7726	0.7737	true location shift is not equal to 0
9	3155.5	0.0003431	0.0003448	true location shift is not equal to 0
C	2992	6.03E-05	6.06E-05	true location shift is not equal to 0

FONTE: Dados de pesquisa.

TABELA 5: Resultado dos testes de Equações de Estimações Generalizadas por resíduo e cadeia C.

Posição	Média ± DP	Média ± DP	Diferença	GEE
	NÃO LIGANTE (n=95)	LIGANTE (n=95)		P (ajustado por sequencial Sidák)
1	86,17 ± 31,09	78,53 ± 25,06	7,64	0,393
2	42,66 ± 14,02	29,77 ± 11,97	12,90	<0,001
3	121,43 ± 41,00	108,28 ± 36,74	13,15	0,177
4	205,21 ± 69,51	188,62 ± 62,09	16,59	0,477
5	221,86 ± 74,72	231,71 ± 77,03	-9,84	0,900
6	164,92 ± 45,16	150,43 ± 43,74	14,50	0,195
7	166,74 ± 42,57	147,93 ± 41,99	18,81	0,032
8	61,36 ± 35,22	60,99 ± 36,30	0,37	0,943
9	31,13 ± 21,93	23,21 ± 10,88	7,92	0,025

FONTE: Dados de pesquisa.

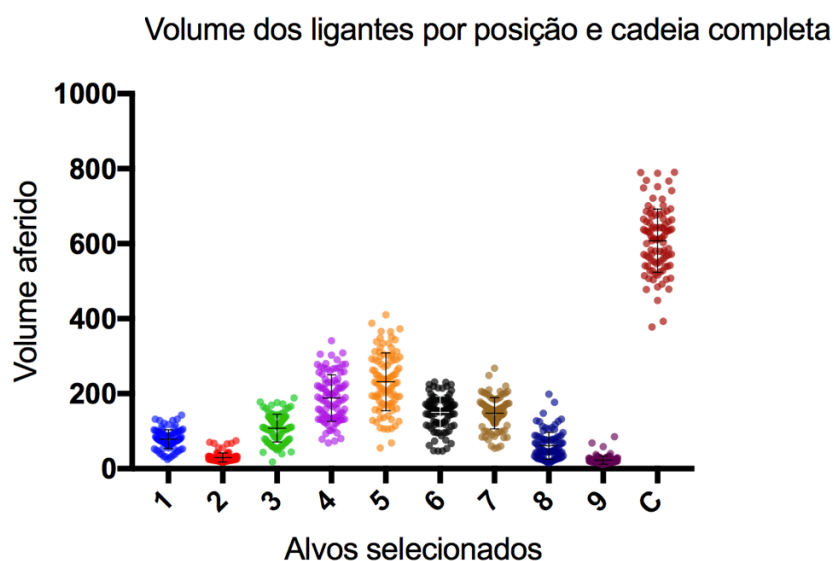


Gráfico 2: Volume intersticial para os dados de volume de 190 sequências consideradas ligantes. Cada volume foi calculado tendo como referência a cadeia *alpha* e computando os valores de volumes para toda a cadeia C (epitopo) do complexo pMHC. IC considerado: 95%.

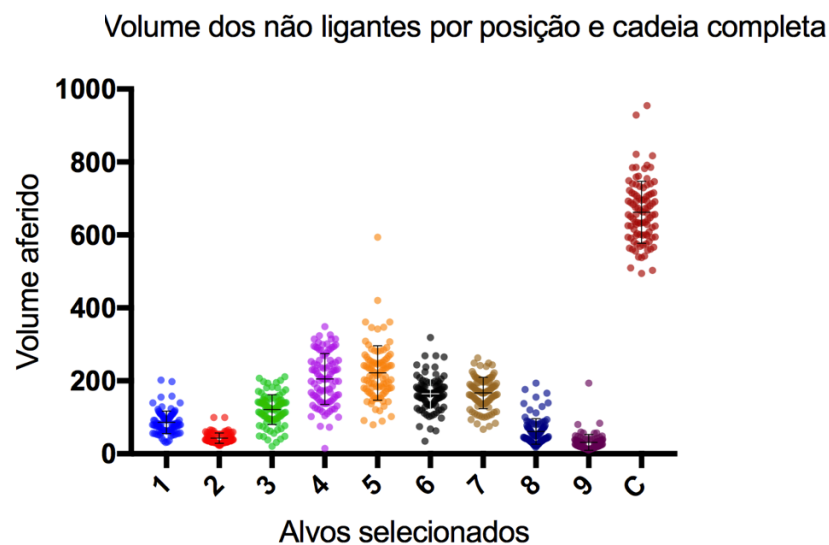


Gráfico 3: Volume intersticial para os dados de volume de 190 sequências consideradas não ligantes . Cada volume foi calculado tendo como referência a cadeia *alpha* e computando os valores de volumes para todo a cadeia C (epitopo) do complexo pMHC bem como para cada posição específica de cada aminoácido. IC considerado: 95%.

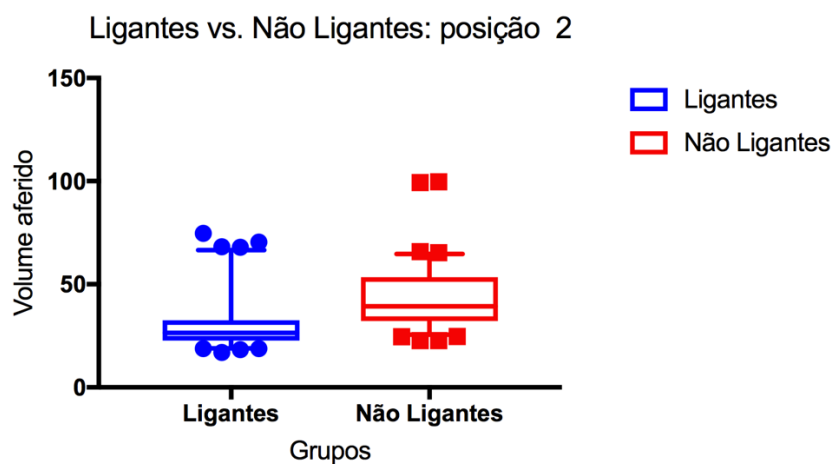


Gráfico 4: Volume intersticial posição 2 da cadeia C . Para os dados de volume de 190 sequências consideradas cada volume foi calculado tendo como referência a cadeia *alpha* e computando os valores de volumes para o aminoácido na posição 2 da cadeia C (epitopo) do complexo pMHC, com indicação de média e desvio para cada grupo, para valor de α igual a 0,05.

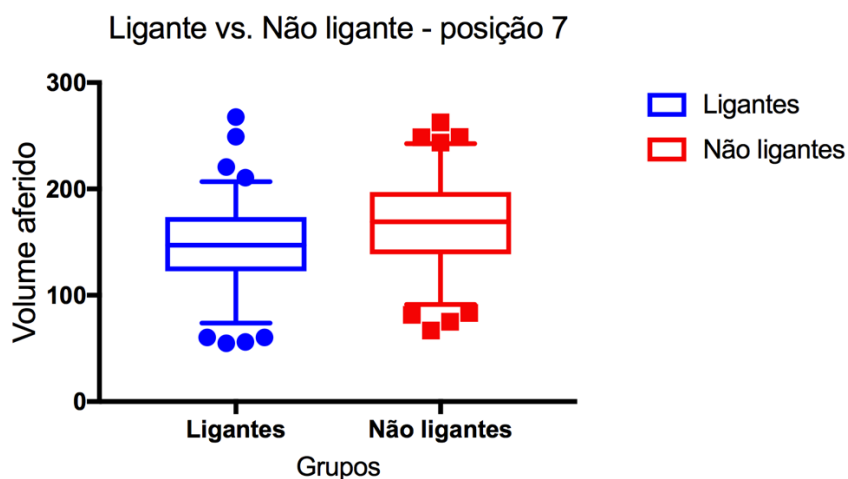


Gráfico 5: Volume intersticial posição 7 da cadeia C. Para os dados de volume de 190 sequências consideradas cada volume foi calculado tendo como referência a cadeia *alpha* e computando os valores de volumes para o aminoácido na posição 7 da cadeia C (epitopo) do complexo pMHC, com indicação de média e desvio para cada grupo, para valor de α igual a 0,05.

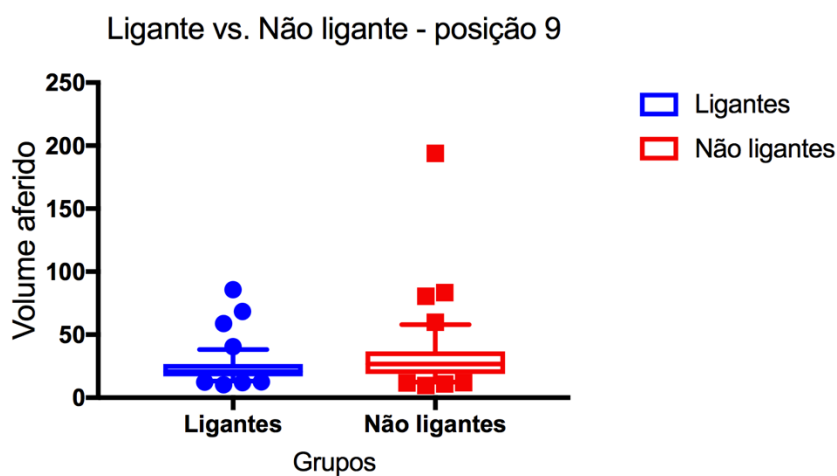


Gráfico 6: Volume intersticial posição 9 da cadeia C. Para os dados de volume de 190 sequências consideradas cada volume foi calculado tendo como referência a cadeia *alpha* e computando os valores de volumes para o epitopo na posição 9 da cadeia C (epitopo) do complexo pMHC, com indicação de média e desvio para cada grupo, para valor de α igual a 0,05.

4.4. VALORES DE ENTRADA E TREINAMENTO DE REDES NEURAIAS ARTIFICIAIS

Tanto os valores dos volumes intersticiais da cadeia C como os valores dos volumes das posições 2, 7 e 9 contra a fenda do MHC foram utilizados para o treinamento dos métodos de aprendizagem de máquina. As redes geradas foram então testadas para gerar curvas ROC. As curvas ROC auxiliam a visualização das taxas de acerto das redes treinadas, de modo que, quanto maior a área abaixo da curva, maior o acerto da rede. Os gráficos 7 e 8 apresentam as curvas ROC para as melhores redes geradas. Os valores AUC (*area under the curve*) encontrados demonstram uma alta taxa de acerto no conjunto de validação.

Tabela 6: Valores de entrada para ligantes.

Exemplo dos dados para treinamento das redes neurais.

Volumes - Ligantes - Valores de Entrada				
Alvos	aa2	aa7	aa9	Cadeia c
1	66.16	138.3	23.96	567.9
2	19.63	145.8	13.82	636.4
3	22.47	155.3	20.26	567.2
(...)				(...)
94	26.4	83.48	68.52	549.6
95	36.04	220.7	58.91	665.2

Tabela 7: Valores de entrada para não ligantes.

Exemplo dos dados para treinamento das redes neurais.

Volumes - Não Ligantes - Valores de Entrada				
Alvos	aa2	aa7	aa9	Cadeia c
1	32.09	166.4	49.13	571.7
2	30.37	196.7	14.06	628.6
3	30.58	106.5	15.92	494.8
(...)				(...)
94	33.28	248.8	36.85	791.5
95	30.04	197.1	26.47	656.4

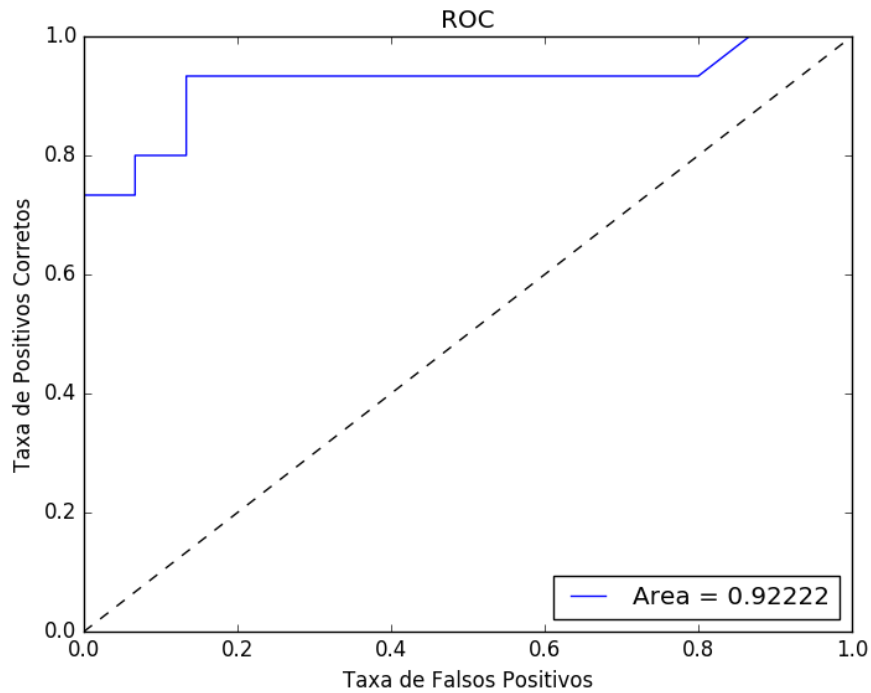


Gráfico 7: Curva ROC – Melhor rede treinada. Área abaixo da curva indica a qualidade da rede em termos de acertos para o conjunto de treinamento da rede.

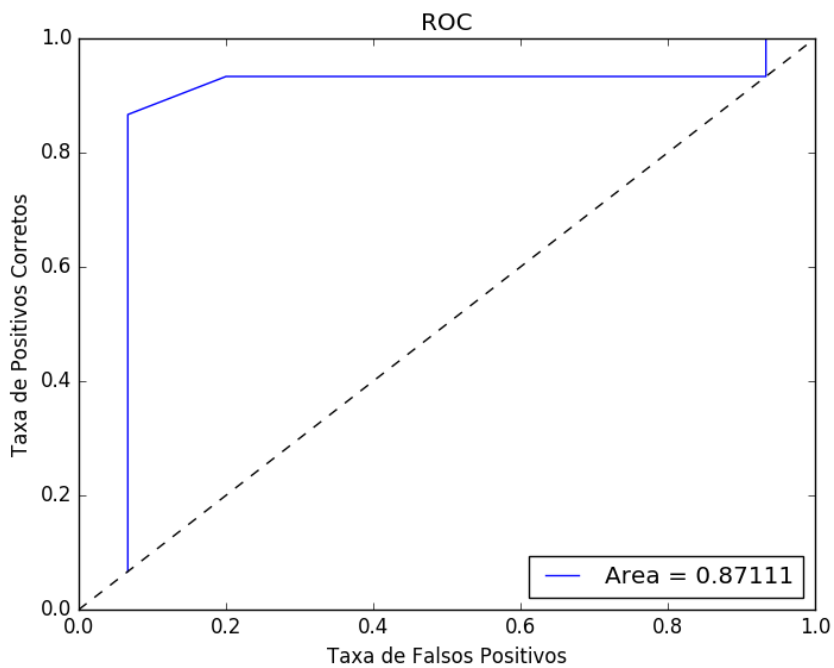


Gráfico 8: Curva ROC – Segunda melhor rede. Área abaixo da curva indica a qualidade da rede em termos de acertos para o conjunto de treinamento da rede.

5 DISCUSSÃO

Este trabalho teve como principal objetivo caracterizar a diferença entre alvos ligantes e não ligantes de MHC-I, através da avaliação volumétrica do interstício da fenda de complexos pMHC-I. Além disso, também foi proposta uma metodologia baseada em treinamento de redes neurais que fosse capaz de fazer esse discernimento entre ligantes e não ligantes após a docagem do peptídeo na fenda do MHC-I de interesse. Isso tudo, dentro do contexto da imunoinformática de buscar soluções simplificadas para problemas complexos, e com o objetivo de colaborar no preenchimento das lacunas existentes na avaliação estrutural da via endógena da apresentação de antígenos.

É possível perceber que a ligação de peptídeos provenientes de proteínas endógenas ao MHC-I é um passo importante durante a resposta contra patógenos. Assim, baseados em sequências e no contexto da apresentação de antígenos da via endógena, há programas direcionados para avaliar a afinidade do epitopo pelo MHC-I. Dois deles, citados anteriormente, o NetMHC e o BIMAS são largamente utilizados (Lundegaard et al., 2008; Parker et al., 1994). Apesar de computacionalmente eficientes, ao processar apenas a sequência, e manter de lado a informação estrutural do complexo pMHC, perdemos algumas informações importantes que acredita-se que seriam melhor acessadas através da estrutura terciária da proteína. (Ishikawa, 2017; Antes et al., 2006)

No âmbito do sistema imune, utilizar dados estruturais pode permitir uma melhor avaliação da afinidade do epitopo pelo MHC através de dados de área acessível ao solvente (ASA) e do volume ocupado na fenda do MHC-I. Esse tipo de abordagem permite, por exemplo, o desenho de peptídeos análogos para melhorar a ligação do epitopo a um MHC (Hoppes et al., 2014; Bordner & Abagyan, 2006).

Um dos maiores impasses no uso de estruturas tridimensionais, como um todo e ao longo desse trabalho, se deu por sua pouca disponibilidade. Se faz necessário o emprego de técnicas que gerem os modelos a serem estudados, uma vez que a existência de cristais de boa resolução é pequena. A necessidade do uso de técnicas como Cristalografia, Ressonância Magnética Nuclear e Difração de Raios-X, exigem tempo e mão de obra qualificada, o que acaba elevando os custos desses processos.

Em uma consulta rápida ao banco de dados PDB, de onde foram obtidas as estruturas cristalográficas que deram início a esse trabalho, é possível notar que há cerca de 130 mil estruturas depositadas. Todavia, ao restringir a busca às estruturas e alelos de interesse desse trabalho o número cai drasticamente. Nesse sentido, uma abordagem que faça uso da modelagem molecular se fez bastante útil e viável. Entretanto, haja visto que há cerca de 6000 alelos de MHC-I, que podem codificar alótipos com diferenças de apenas um aminoácido, mas que interferem na ligação de cada peptídeo, a técnica utilizada precisa ser bem desenhada e generalizações feitas a partir destes resultados devem ser discutidos com cuidado.

As três principais técnicas utilizadas atualmente para a docagem de peptídeos em MHC são: o MHCsim; primeiro servidor disponível online para construção de complexos dessa natureza; pDOCK e outro protocolo de ancoragem baseado na aplicação de Monte Carlo (Rigo, 2015a; Todman et al., 2008; Khan & Ranganathan, 2010; Bordner & Abagyan, 2006).

A técnica escolhida, D1-EM-D2, disponível online sob o nome Docktope, apresenta uma vantagem essencial sobre as outras três técnicas supracitadas. Ela não se baseia em similaridade de sequências, como é o caso do MHCsim, mas em dados estruturais recuperados de estruturas cristalografadas. Sua validação é feita por *cross-docking* e não *re-docking*, como é o caso do pDOCK (Rigo, 2015a).

Mesmo assim, em cada grupo de alvos escolhidos do banco de dados IEDB, ao menos três epitopos não foram corretamente docados na fenda do MHC-I e, por esse motivo, foram excluídos das análises. Pode-se pensar que a não correta modelagem do epitopo possa ter se dado por conta de, originalmente, o padrão conformacional adotado pelo epitopo na fenda no MHC-I diferir do padrão utilizado pela técnica durante a criação do modelo (padrão estrutural tipicamente encontrado para o alelo HLA-A*02:01). Sendo assim, é necessário um aprofundamento nessas questões conformacionais que sejam capazes de explorar sub padrões que possam existir para um mesmo alelo de MHC.

Assim, foram gerados complexos confiáveis quanto a sua estrutura para a análise pretendida. Nesse sentido, podemos dizer também que uma das

razões para os bons resultados apresentados, está na forma em como as sequências para as estruturas geradas foram escolhidas.

Uma das principais dificuldades em se trabalhar com bancos de dados com informação recuperada de forma automática e/ou alimentado diretamente pelos autores são as informações de retorno. Foi necessário retomar os artigos onde as sequências foram publicadas para garantir que os alvos haviam mesmo sido testados em ensaios de IC50, bem com garantir que o alelo de MHC estava correto.

A acuidade ao obter as sequências e as várias etapas de refinamento na recuperação dos dados, trouxe a confiança necessária para a construção de complexos pMHC-I. Esses complexos refletem a diversidade dentro do ambiente intracelular, ainda que se tenha tomado apenas os dois extremos em relação aos ensaios de ligação. É importante salientar que este foi um estudo exploratório e, ainda que essa seja uma limitação da avaliação, foi a forma encontrada para garantir a diferença máxima entre os dados apresentados.

Comentando especificamente dos passos de execução do projeto, a avaliação inicial de cristais de HLA-A*02:01 se mostrou muito útil para compreensão das ferramentas que seriam utilizadas, especialmente o UCSF Chimera e os módulos de extração e cálculo de volumes, *Surfnet* e *Measure Volume and Area*. No momento da extração de valores dos volumes existentes nas fendas dos cristais de MHC, deparou-se com a necessidade de definir o tamanho da sonda a ser utilizada ao longo da análise para recuperação dos dados de volume. Esse foi um momento crucial da análise, uma vez que uma sonda muito pequena nos retornava muitos valores a respeito do volume acima do peptídeo. Enquanto uma sonda de tamanho maior nos retornava mais informação a respeito da parte interna - entre o peptídeo e a base da fenda do MHC. Essa análise inicial possibilitou a escolha da sonda que melhor discriminava as variações de volume bem como os comportamentos individuais dos aminoácidos ao longo da cadeia do peptídeo

Posteriormente, para verificar se existiam valores volumétricos intersticiais diferentes entre ligantes e não ligantes de MHC-I, 190 sequências alvo foram obtidas de bancos de dados (95 ligantes e 95 não ligantes), as quais foram modeladas dentro do alelo HLA-A*02:01 e extraídos os valores de volumes de áreas disponíveis para interação entre o peptídeo e a fenda deste

MHC-I de todas as estruturas. Cada uma dessas sequências foi avaliada resíduo a resíduo e, também, levando em conta a cadeia completa do peptídeo na fenda do MHC. As tabelas completas com os dados podem ser encontradas na seção de Apêndices deste trabalho.

O Gráfico 1, traz os valores de volume disponíveis na fenda do MHC, levando em consideração o peptídeo completo. É possível notar uma clara tendência dos valores tomados como ligantes de permanecerem abaixo do valor médio obtido com os dados para não ligantes. Da mesma forma, é possível observar o contrário para os não ligantes, os quais tendem a permanecer acima da média dos ligantes.

Através desse mesmo gráfico é possível notar que alguns dos valores para ligantes ficam muito acima das médias dos não ligantes, caracterizando um comportamento inesperado inicialmente. Na tentativa de explicar esses dados, alguns deles foram tomados aleatoriamente para realização de dinâmica molecular. O que foi possível perceber em um desses complexos foi o desprendimento do epitopo em relação a fenda ao longo da dinâmica molecular, de modo que maiores investigações seriam necessárias.

O contrário também foi analisado, peptídeos tidos como não ligantes que possuem comportamento de ligantes, ou seja, espaços muito pequenos entre o peptídeo e o MHC-I. Nesses casos, os *frames* extraídos das dinâmicas não mostraram qualquer alteração significativa ao longo do tempo de simulação. A possível explicação, além de que esse realmente é um comportamento natural, é de que esses peptídeos são sim positivos, apenas não tendo sido realizado um ensaio onde eles seriam caracterizados dessa maneira. Uma análise mais completa por dinâmica molecular (com cálculo dos volumes em diferentes intervalos de tempo das simulações) requer mais tempo e pretendemos explorar essa possibilidade em um estudo posterior.

Uma forma mais direta para investigar estes ligantes com valores acima da média e não ligantes com baixos volumes intersticiais, foi a condução de uma análise individual para cada resíduo da cadeia peptídica dos ligantes (alguma posição específica poderia estar explicando estes valores aberrantes em alguns peptídeos). Visualmente, foi possível notar a mesma tendência para com os resíduos tidos como âncoras na ligação do peptídeo com o MHC-I, referentes aos aminoácidos das posições 2 e 9.

Partiu-se então para uma análise estatística dos dados gerais. Apesar de não ter sido possível distinguir se havia diferenças entre as médias nos grupos de dados estudados, foi possível encontrar uma diferença estatística através do teste não paramétrico de Wilcoxon-Mann-Whitney e perceber que o grupo de não ligantes apresenta valores (mediana) de volume mais altos que os ligantes. Isso corrobora a ideia inicial do trabalho de que haveria um volume maior no espaço formado entre a fenda e o peptídeo entre não ligantes. Os resultados podem ser encontrados na Tabela 4.

Na intenção de confirmar os achados, uma série de outros testes foram realizados, os quais podem ser encontrados na seção Apêndices desse trabalho. Como conclusão geral dos testes podemos dizer que: quanto ao volume, a diferença entre ligantes e não-ligantes depende da posição do aminoácido, sendo que, não ligantes têm volumes maiores nos resíduos das posições 2 (GEE, $P < 0,001$), 7 (GEE, $P=0,032$) e 9 (GEE, $P=0,025$). O mesmo ocorre com relação ao volume disponível levando o peptídeo completo em consideração (média: $662,12 \times 607,86$; teste t: $P < 0,001$). Quanto às demais posições de aminoácidos, não há diferença entre ligantes e não ligantes.

Buscando verificar se essas posições que apresentaram significância estatística poderiam ser utilizadas como variantes preditoras na separação dos ligantes, partiu-se para o treinamento das redes neurais. Visou-se a observação do fato destas serem capazes de acertar a classificação dos alvos ligantes e não ligantes. As melhores redes apresentaram valores de AUC de 0,8711 e 0,922; valores realmente altos para o método aplicado. É interessante observar que a metodologia utilizada apresentou resultados ainda melhores que os desempenhados pelo NetMHC3.0 (Lundegaard et al, 2008), por exemplo.

No trabalho mencionado, o treinamento da rede foi feito com sequências de epitopos com IC_{50} inferior a 500nM (como critério de positividade de ligação), onde 76% deles foram corretamente preditos. A média da correlação de Pearson nesse caso, foi de 0.71 e a área abaixo da curva ROC para esse limiar de IC_{50} , foi de 0.86. Obviamente, a quantidade de dados de entrada interfere nesses resultados. O conjunto utilizado nesse trabalho é muitas vezes menor que o *dataset* utilizado pelo NetMHC3.0.

Levando em consideração o fato de que as redes treinadas para esse trabalho tiveram como dados de entrada apenas os volumes considerados

significativos (sem nenhuma informação dos resíduos de aminoácidos que compunham as sequências - exemplo dos dados de entrada na Tabela Suplementar 5), obteve-se um resultado animador e indicativo de que há sentido em tentar explicar a ligação dos peptídeos levando em consideração o espaço da fenda do MHC pós ancoramento.

Sabe-se que ainda se faz necessário modelar e testar dados em maior volume, utilizando as mesmas sequências que compunham o grupo de treinamento do NetMHC3.0. Da mesma forma, devem ser modelados e avaliados os volumes em sequências de peptídeos tidos como intermediários em testes de ligação. Todavia, as análises à que foram submetidos os dados nos mostram que realmente há uma clara distinção volumétrica no espaço intersticial formado entre a fenda do MHC e a cadeia peptídica entre os grupos de ligantes e não ligantes analisados. Especialmente para as posições 2, 7 e 9.

Além disso, utilizando apenas estes valores diferenciais fomos capazes de realizar um método de aprendizagem de máquina capaz de classificar, com índices altamente satisfatórios, ligantes e não ligantes do grupo de validação. É importante salientar que essa classificação é dependente de métodos de obtenção de estruturas confiáveis em larga escala, o que ainda não é disponível para a maioria dos alelos de MHC. Este projeto demonstra a importância da informação estrutural, até então praticamente negligenciada, no discernimento da ligação entre peptídeo e MHC. O direcionamento de estudos nesta área abre caminho para o desenvolvimento de ferramentas de predição inovadoras abordando essa informação estrutural de uma maneira rápida, eficiente e confiável.

REFERÊNCIAS

- Abbas AK, Lichtman AH, Pillai S (2012) *Imunologia Celular e Molecular*. 7 edition. ELSEVIER, Rio de Janeiro, 576 pp.
- Anders AK, Call MJ, Schulze MS, Fowler KD, Schubert DA, Seth NP (2011) HLA-DM captures partially empty HLA-DR molecules for catalyzed removal of peptide. *Nat Immunol* 12:54–61
- Antes I, Siu SWI, Lengauer T (2006) DynaPred: A structure based method for prediction of MHC class I binding peptide sequences and conformations. *Bioinformatics* 22. 14: 15-24.
- Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, Chies JAB and Vieira GF (2011) Structural in silico analysis of cross-genotype-reactivity among naturally occurring HCV NS3-1073-variants in the context of HLA-A*02:01 allele. *Molecular Immunology* 48:1461-1467.
- Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, Chen W, Levin AM, Connolly JM, Zhu C (2011) T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* 35:681-693.
- Bhasin M and Raghava GP (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 13:596-607.
- Bordner AJ, Abagyan R (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins: Structure, Function and Bioinformatics* 63(3):512-526
- Brehm MA, Selin LK and Welsh RM (2004) CD8 T cell responses to viral infections in sequence. *Cell Microbiol* 6:411-421
- Celie PHN, Toebes M, Rodenko B, Ovaas H, Perrakis A, Schumacher TNM (2009) Uv-Induced Ligand Exchange in Mhc Class I Protein Crystals. *J.Am.Chem.Soc.* 131: 12298-12304
- Chen M, Bouvier M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. *EMBO J* (2007) 26:1681–90
- Davies MN, Flower DR (2007) Static energy analysis of MHC class I and class II peptide-binding affinity. *Methods in Molecular Biology*. 409:309-320
- Donati C and Rappuoli R (2013) Reverse vaccinology in the 21st century: improvements over the original design. *Ann N Y Acad Sci* 1285:115-132.
- Expert Reviews in Molecular Medicine (2003) Cambridge University Press: Cambridge Clinical School
- Falk K, Rötzschke O (2002) The final cut: how ERAP1 trims MHC ligands to size.

Nature Immunology 3(12):1121-2

Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B and Sette A (2017) The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Front. Immunol.* 8:278.

Guan P, Doytchinova IA, Zygouri C, Flower DR(2003) MHCpred: bringing a quantitative dimension to the online prediction of MHC binding. *Applied Bioinformatics* 2(1):63-6.

Halle S, Halle O, Förster R (2017) Mechanism and dynamics of T-cell-mediated cytotoxicity in vivo. *Trends in Immunology* 38(6):432-443

Hebb D (1949) *The Organization of Behavior*. New York: John Wiley & Sons.

Hoppes R, Oostvogels R, Luimstra JJ, et al (2014) Altered peptide ligands revisited: vaccine design through chemically modified HLA-A2-restricted T cell epitopes. *Journal in Immunology* 193(10):4803-4813

Ishikawa T (2017) Prediction of peptide binding to a major histocompatibility complex class I molecule based on docking simulation. *Journal of Computer-Aided Molecular Design* 30:875-887

Khan AR, Baker BM, Ghosh P, Biddison WE, Wiley DC (2000) The structure and stability of an HLA-A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. *Journal of Immunology* 164: 6398-6405

Khan JM, Ranganathan S (2010) pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Research* 6 Suppl 1:S2.

Kindt TJ, Goldsby RA and BA O (2008) *Kuby Immunology*. 6th edition, 704 pp.

Kish-Trier E, Hill CP (2013) Structural Biology of the Proteasome. *Annual Reviews of Biophysics* 42:1.1-1.21

Laskowski, RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics* 5:323-30

Lecker SH, Goldberg AL, Mitch WE (2006) Protein degradation by ubiquitin-proteasome pathway in normal and disease states. *Journal of American Society of Nephrology* 17, 1807-1819

Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Barbie V, Chaume D (1999) IMGT, the international ImMunoGeneTics database *Nucleic Acids Res* 27 (1): 209-212.

Lefranc MP, Giudicelli V, Dourox P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse A (2015) IMGT,

the international ImMunoGeneTics information system 25 years on. *Nucleic Acids Res* 33 (D1): D413-D422.

- LeibundGut-Landmann S, Waldburger JM, Krawczyk M, Otten LA, Suter T, Fontana A, Acha-Orbea H, Reith W (2004) Specificity and expression of CIITA, the master regulator of MHC class II genes. *Eur J. Immunol.* 2004. 34: 1513–1525
- Lundegaard C, Lamberth K, Harndahl M, Buys S, Lund O, Nielsen M (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acid Research*: W509-W512
- Madden DR, Garboczi DN, Wiley DC (1993) The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell(Cambridge,Mass)* 75: 693-708
- McCulloch, Warren; Walter Pitts (1943) A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics.* 5 (4): 115–133.
- Mehra NK, Kaur G (2003) MHC-based vaccination approaches: progress and perspectives. *Expert Reviews in Molecular Medicine*, Vol. 5:1-17.
- Mungall AJ, Palmer SA, Sims SK, et al (2003) The DNA sequence and analysis of human chromosome 6. *Nature* 425(6960):805-11
- Murata S, Yashiroda H, Tanaka K (2009) Molecular mechanisms of proteasome assembly *Nature Reviews Molecular Cell Biology* 10, 104-115
- Murphy K, Travers P, Walport M (2010) *Immunobiologia de Janeway.* 7 edition. Artmed, Porto Alegre. 899 pp.
- Neefjes J, Jongstra ML, Paul P, Bakke O (2011) Towards a system understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology* 11:823-836
- Nielsen M, Lundegaard C, Lund O and Kesmir C (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57:33-41.
- Neubig RR, Spedding M, Kenakin T, Christopoulos A (2003) XXXVIII Update on terms and symbols in quantitative pharmacology. *Pharmacological Reviews* 55:597–606.
- Praveen PV, Yaneva R, Kalbacher H, Springer S. Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange. *European Journal of Immunology* (2010) 40:214–24
- Peaper DR, Cresswell P (2008) Regulation of MHC class I assembly and peptide

- binding. *Annual Review of Cell and Developmental Biology* 24:343-68
- Parham P (2005) MHC class I molecules and KIRs in human history, health and survival. *Nature Reviews in Immunology*. 5(3):201-14
- Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Journal Immunology* 152(1):163-75.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*:1605-1612.
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA and Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213-219.
- Rappuoli R (2000) Reverse vaccinology. *Current opinion in microbiology* 3:445-450.
- Reiser JB, Legoux F, Gras S, Trudel E, Chouquet A, Leger A, Le Gorrec M, Machillot P, Bonneville M, Saulquin X, Housset D (2014) Analysis of Relationships between Peptide/MHC Structural Features and Naive T Cell Frequency in Humans. *J.Immunol.* 193: 5816-5826
- Rigo MM, Antunes D, Vieira G and Chies J (2009) MHC: Peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein. *Lecture Notes in Computer Science* 5676:160-163.
- Rigo MM, (2015a) Modelagem automatizada e ensaios de estabilidade In Silico sobre complexos peptídeo:MHC-I. Lume – UFRGS. Disponível em: <<http://hdl.handle.net/10183/128956>>
- Rigo MM, Antunes DA, Freitas MV, Mendes MFA, Meira L, Sinigaglia M, Vieira (2015b) Docktope: A web-based tool for automated pMHC-I modelling. *Nature - Scientific Reports: Lecture Notes in Computer Science*. 5: online published.
- Saito T, Yokosuka T and Hashimoto-Tane A (2010) Dynamic regulation of T cell activation and co-stimulation through TCR-microclusters. *FEBS letters* 584:4865-4871.
- Sette A, Rappuoli R (2010) Reverse vaccinology: developing vaccines in the era of genomics. *Immunity* 33(4):530-41
- Sinigaglia M, Antunes DA, Rigo MM, Chies JA and Vieira GF (2013) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database (Oxford)* 2013:bat002.
- Sohn SJ, Thompson J and Winoto A (2007) Apoptosis during negative selection

of autoreactive thymocytes. *Curr Opin Immunol* 19:510-515.

Swann JB, Smyth MJ (2007) Immune surveillance of tumors. *The Journal of Clinical Investigation* 117(5):1137-1146

Thauland TJ, Parker DC (2010) Diversity in immunological synapse structure. *Immunology*. 131(4):466-72

Tomar N, De RK (2010) Immunoinformatics: an integrated scenario. *Immunology*131(2):153-68

Todman SJ, Halling-Brown MD, Davies MN, Flower DR, Kayikci M, Moss DS
Toward the atomistic simulation of T cell epitopes automated construction of MHC: peptide structures for free energy calculations. *Journal of Molecular Graphics and Modelling* 26(6):957-61

Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. (2014) The immune epitope database (IEDB) 3.0. *Nucleic Acids Research(Database issue)*:D405-12

Vangone A, Bonvin AM (2015) Contacts-based prediction of binding affinity in protein-protein complexes. *Elife* 4:e07454

van de Weijer ML, Luteijn RD and Wiertz EJ (2015) Viral immune evasion: Lessons in MHC class I antigen presentation. *Semin Immunol*.

Wieczorek M, Abualrous ET, Sticht J, Alvaro-Benito M, Stolzenberg S, Noé F, Freund C (2017) Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Frontiers in immunology*. 8: 292. 16p.

Xue L, Rodrigues J, Kastiris P, Bonvin AMJJ., Vangone A (2016) PRODIGY: a web-server for predicting the binding affinity in protein-protein complexes. *Bioinformatics* 32(23): 3676-3678

Zhang Y, Ren Y, Wu Y, Zhao B, Qiu L, Li X, Xu D, Liu J, Gao GF, Meng SD (2013) The L60V variation in HBV core protein elicits new epitope-specific cytotoxic T lymphocytes and enhances viral replication. *J.Virol*.

Zoete V, Irving M, Ferber m, Cuendet MA, Michielin O (2013) Structure-based rational design of T cell receptors. *Frontiers in Immunology* 2013:4(268)

APÊNDICES

Tabela Suplementar 1: Alvos positivos iniciais pós-docagem por D1-EM-D2

Resíduo										
Sequências	1	2	3	4	5	6	7	8	9	Epitopo
FILGIIITV	143.2	162.7	406.3	583.7	800.8	668.7	366.3	339.4	31.2	377.7
FLAHLQWFA	121.6	160.7	441	625.2	900.1	638.8	392.2	316	50.2	621.8
FLIDLAFI	125.7	158.3	438.7	679.5	938.3	690.6	399.9	301.9	31	636.4
FLLLSIPI	128.5	162.1	418.8	693.8	780.7	690.1	370.6	341	29.25	524.9
FLNISWFYI	123.1	182.9	518.6	730.1	784.2	650.6	404.2	341.1	32.92	615.2
FLSNGHVTI	132.8	161.4	574	708	765.5	630.8	386.8	382.2	28.35	643.9
FLSRLVLYA	139.1	160.4	553	829	794.1	664.2	394.8	378.5	45.73	575.9
FLVPFVFL	133.9	171.9	491.8	582.3	937.9	640.3	385	343	49.64	531.2
GLLDRLYDL	109.3	155.4	414.5	689.9	797.1	671.2	428.1	367.7	45.13	484.6
GLLQFIVFL	115.3	164.8	384.2	724.3	875.8	682.9	391.5	330.1	36.63	567.9
LLMMTLPSI	161.3	157.7	349	702.5	913.4	626.8	375.2	290.8	29.76	504.8
LLNLLWPL	162.3	155	519.6	702.8	887.6	688.8	440.8	350.8	40.03	616.2
LMSFTILCL	162.2	159.7	518.4	596.1	875.5	664.4	402.1	352.9	46.49	483
LMSTLLIYL	170.1	143.9	565.4	665.8	809.9	693.7	389.8	376.8	34.93	573.6
MMLVPLITV	128.8	150.5	345.8	683.3	792	632.7	407.4	334.7	37.24	494.7
NLLLWPLYV	204.3	161.2	400.5	704.9	890.2	638.6	403.9	339.1	31.62	537
QLLDVKLAL	158.8	159.3	458.9	697	845.2	645.3	417.6	342.9	58.03	505.3
TLLGLILFV	180.7	164.1	358	594	970.9	655.3	425	341.8	35.38	554.5
WLIGDFDFV	148	162	409.3	567.4	1068	694.6	390.2	359.3	29.19	642.1
YLYALYSPL	134.6	165.7	475.8	632.1	933.7	655.4	427.8	368.1	44.95	572.3
YMMGIEYGL	149.4	161.4	430.4	578.7	888.5	665.2	429.1	318.3	34	471.3
YTYSGLFCV	146.3	175.3	462.8	696.5	782.9	650.9	412.1	317.5	35.22	520.9

Tabela Suplementar 2: Alvos iniciais negativos pós-docagem por D1-EM-D2

Resíduo										
Sequências	1	2	3	4	5	6	7	8	9	Epitopo
AARNIVRRA	145.1	175.3	518.4	728.2	889	670.6	767.2	693.4	63.95	654.8
AAVSADPLA	138.5	178.9	412	667.2	804.7	704.2	403.7	328.1	62.4	712.4
AVKGVGTMV	140.1	154.1	536	587.5	843.3	655.1	432.4	334.7	41.41	605.7
CVYCKQQLL	202.4	166.2	474.5	715.6	828.9	665.5	457.9	340.6	66.93	616.2
ELQTTIHDII	155.2	154.1	574.5	689.2	800.7	642.7	438.3	341.7	32.52	586.6
FISGIQYLA	121.5	163.7	523.6	574.4	844.1	668.9	402.3	314.4	52.73	476.5
IYAGSLSAL	165.6	168.4	446.5	564.2	786.3	655.9	467.1	309	41.12	573.6
KISEYRHYC	254.5	168	558.8	731.6	963.7	649.2	438.4	352.4	55.04	614
LCTELQTTI	159.8	189.5	386.6	799.3	837.4	682.3	409.8	340.2	32.69	657.3
LLDDEAGPL	161.6	160.1	450.1	693.8	901.8	716.8	418.4	356.9	37.64	721
LQTTIHDII	168.1	169.8	436	653.2	811.8	624.2	414.4	337	26.8	567.3
PLEEELPRL	127	175	485.1	631.5	871.1	672.2	377.6	553.9	47.32	536.9
QAQTRHPAT	274.2	200.3	560	675.5	744.5	598.6	364.6	289.7	50.38	598.6
RAAVSADPL	460.3	185.8	447.1	677.5	874.1	716.1	449.6	344.4	40.82	737.6
RAFTEEGAI	458	178.4	472.9	678.9	777.8	672	474.5	314.2	38.49	575.1
RVRFFFPSL	161.5	177.8	603.6	764.1	948.7	643	400.8	340.8	37.23	489.3
SAAFEDLRL	194.2	179	428.9	607.1	723.4	721.8	406.8	366.2	37.72	601.1
SILGIGTVL	199.2	164.1	420.6	581.2	828.3	640.9	444.9	337	68.75	636
TEVGQDQYV	183.2	173.8	442.7	563.5	871.7	687.4	419.7	344.7	22.93	563
TLHEYMLDL	180.3	163	444	770.6	922.8	660.2	432	349	41.51	634.5
TRHPATATV	179.1	165.3	453.8	576	801.3	661.5	411.4	324.9	26.13	564.6
YLEPAIAKY	129.1	165.4	467.6	584.2	764.1	723.1	497.6	479.1	91.63	659.9

Tabela Suplementar 3: Dados de extração de valores de volume de pMHC “Positive”. Ordenada por valor de IC50 e contendo valores de ΔG e K_d .

Extração de Volumes											IEDB	Prodigy	
Alvos	aa1	aa2	aa3	aa4	aa5	aa6	aa7	aa8	aa9	Cadeia c	IC50	deltaG	Kd
GLLQFIVFL	40.44	66.16	52.91	239.4	211.2	154.6	138.3	26.9	23.96	567.9	0.0075	-10	4.90E-08
FLIDLAFLL	69.83	19.63	131.2	257.4	289.6	158	145.8	68.34	13.82	636.4	0.041	-10	4.70E-08
TLLGLILFV	80.09	22.47	66.22	122.1	298.5	154.9	155.3	54.52	20.26	567.2	0.045	-11	8.10E-09
WLIGDFDFV	92.1	21.93	125	131.6	365.2	163.4	150.2	85.55	15.5	691.7	0.06	-9.4	1.20E-07
QLLDVKLAL	73.02	32.13	59.81	144.8	162	185.2	144.1	33.84	33.84	508.2	0.06	-11.4	4.20E-09
YVMTMILFL	69.83	16.91	130.4	235.8	251.3	116.7	121.3	45.05	31.36	587	0.065	10.2	3.50E-08
LLNLLLWPL	75.16	26.55	135.9	215.2	248.9	146.2	143.2	29.59	20.76	616.2	0.073	-10	4.60E-08
FLSRLVLYA	36.61	25.17	92.9	214.5	272.3	146.3	130.3	125.7	32.98	575.9	0.075	-10.3	2.60E-08
FLAHLQWFA	100.8	23.25	103.8	228.4	322.9	188.8	206.1	118.1	31.34	621.8	0.1	-9.6	8.70E-08
LLMMTLPSI	73.68	27.17	43.37	174.2	164.5	88.16	86.1	40.99	13.43	504.8	0.11	-8.9	3.20E-07
FLLTSIPI	75.97	20.59	139.1	199.8	197.8	173.5	156	26.05	14.79	524.9	0.11	-9.3	1.50E-07
FILGIIITV	49.57	22.02	39.75	68.67	168.2	76.57	99.31	37.35	14.33	377.7	0.11	-9.5	1.20E-07
FLSNGHVTI	99.45	24.93	147.1	272	202.3	224.2	144	46.11	15.2	643.9	0.11	-8.2	9.60E-07
NLLLWPLYV	103.4	20.73	104.9	120.4	250.2	180.3	170.1	54.17	14.11	555.8	0.12	-11.5	3.80E-09
YMMGIEYGL	64.21	20.12	135.6	73.76	157.2	166.4	164.7	36.71	20.18	478.4	0.12	-9.7	7.20E-08
YTYSGLFCV	74.01	39.55	133.3	194.9	173.2	162.1	162	22.1	25.23	546.9	0.12	-11.3	5.50E-09
LMSTLLIYL	86.62	24.24	111.8	227.9	242.6	168.5	135.9	84.31	24.05	613.6	0.13	-10.5	2.10E-08
YLYALYSPL	93.06	18.98	118.8	158.9	155.7	146.4	131.6	104.3	31.51	612.2	0.13	-9.5	1.10E-07
GLLDRLYDL	24.84	68.24	56.16	199.1	130.9	115.3	96.37	23.64	29.12	484.6	0.13	-10.1	3.80E-08
LMSFTILCL	78.38	28.86	109.9	124.5	196.3	159.5	120.9	19.72	28.28	492.3	0.14	-10.1	3.90E-08
RLVDFFPDI	47.09	28.5	117.7	219.3	322.3	190.2	140.9	66.92	21.02	719.1	0.14	-10.3	2.90E-08
ALWIPDLFM	59.96	32.56	112.4	183.5	123.8	95.62	145.1	42.88	19.19	570.7	0.14	-10.2	3.20E-08
TLYDFDYI	97.09	22.9	97.79	148.4	228.4	97.14	95.45	131.7	15.33	675.9	0.14	-11.7	2.70E-09
ALWEIQQVV	67.91	31.41	146.2	159.8	106.1	123.3	150.5	52.27	40.51	560.5	0.15	-9.2	1.70E-07
MLNFSIYSV	46.29	27.5	145.7	266.2	215.8	164.4	167.5	24.75	12.59	582	0.15	-9.5	1.10E-07
MLMEVFPQL	142.7	25.16	62.97	128.7	106.8	114.1	99.73	79.31	30.16	516.4	0.15	-8.7	3.80E-07
VLMIKALEL	73.84	24.56	113	230.7	239	141.9	112.2	21.85	22.52	613.1	0.15	-9.1	2.00E-07
FLFFATSDV	75.56	21.56	138.7	137.1	153.6	175.3	160.8	84.91	23.77	539.3	0.16	-9.1	2.00E-07
RLAVYIDKV	84.24	29.09	116	269.5	334.4	153.6	152.4	34.67	17.1	681.6	0.16	-10	4.80E-08
MLFTKFFYL	128.8	23.84	107.1	182.1	191.7	155.2	135.6	109	26.7	638	0.17	-10.3	2.60E-08
ILQYDLWNV	69.67	25.88	143.9	172.8	208.7	208.7	155.1	44.27	19.5	572.6	0.17	-9.2	1.80E-07
YLYALYIFL	77.21	21.15	149.1	137.5	197.8	143.5	164.4	106	24.43	579.9	0.18	-10.5	2.10E-08

SIMAFILGI	98.65	40.21	67.1	123.4	215.3	113.1	147	53.31	21.5	613.6	0.18	-11.1	6.70E-09
FMDGKQACV	52.61	23.94	103.4	134.8	227.4	154.6	84.99	101.3	20.61	506.9	0.18	-9.2	1.90E-07
VLALYSPPL	65.82	22.82	143	278.3	366.3	149.4	137.6	73.91	22.25	741.6	0.19	-8.9	2.90E-07
WLYGYNFII	109.5	19.96	188.4	93.8	344	180.6	141	63.64	16.84	651.4	0.19	-9.8	7.00E-08
TIYSHLLLV	89.65	31.74	117.6	95.55	206.9	154.3	205.9	19.24	17.67	541	0.2	-11.3	5.60E-09
LLLWPLYVL	81.13	23.78	69.03	207.9	225	169.5	173.7	13.74	23.24	552.1	0.21	-10.1	3.80E-08
YLSIYGFGV	105.2	21.81	178.2	268.6	410.7	230.3	203.8	28.32	11.98	752.3	0.21	-9.9	5.60E-08
GLLRVISGV	31.8	42.1	109.8	266.1	249.6	192.8	198.8	47.85	19.8	701.4	0.21	-9.9	5.30E-08
FVDEFYAYL	116.2	35.62	69.76	130.7	268.1	182.6	90.22	177	29.72	686.3	0.22	-10.1	3.70E-08
MMISAGFSL	52.5	27.32	153.1	256.7	291.5	231.4	190.1	15.45	14.79	672.6	0.22	-9.2	1.80E-07
LLLEWLAEV	106.6	30.36	79.31	187.3	334.6	122.8	100.8	46.67	18.9	640.8	0.22	-9.4	1.30E-07
TLAPFNFLV	88.61	26.6	105.2	149.8	338.6	199.4	210.5	45.3	20.38	666.1	0.22	-10.4	2.20E-08
RLFSYNFTT	72.51	36.92	169.8	136.7	311.7	185.7	206	26.16	27.82	655.1	0.23	-10.2	3.40E-08
FLAIKLYGV	75.21	26.43	58.64	165.5	113.6	106.7	101.9	56.63	13.62	448.8	0.23	-9.9	5.30E-08
SLLYLILFL	117.6	27.28	56.77	179.4	305.2	160.5	150.7	87.54	18.06	659.1	0.23	-11.8	2.10E-09
FLAFVVFL	38.74	18.86	119.8	302.5	297.5	159.5	161.1	43.68	22.23	658.9	0.24	-10	4.30E-08
VLYNGVNYL	83.72	27.87	158.8	201.5	175.7	158.7	199.2	75.08	23.01	638.7	0.25	-9.8	6.70E-08
YLMPYSVYI	86.12	22.91	139.2	221.4	278.9	125.7	97.71	88.52	14.1	634.7	0.26	-10.1	4.20E-08
SMWALVISV	97.44	34.33	77.32	78.92	55.08	47.82	83.68	23.26	13.79	392.9	0.27	-10.9	1.10E-08
YLLFNHFSV	95.73	18.28	124.8	116.8	179.5	190.7	170.8	18.14	20.41	527.8	0.27	-10.1	4.20E-08
VLLTYFCFV	90.18	24.01	81.19	308.7	351.2	172.3	172.8	68.55	16.04	787.5	0.28	-10.4	2.30E-08
LLSKNTFYL	68.26	32.66	118.5	214.5	302.8	174.1	187.9	77.62	23.68	663.9	0.28	-10	4.60E-08
KMIYDLNAV	55.29	43.47	103.6	277.9	262.4	177.8	201.8	19.91	18.16	682.3	0.29	-10.4	2.30E-08
FIADIGIV	75.65	19.85	138.9	280	305.1	215.3	157.8	38.8	21.54	677.8	0.3	-9.7	7.70E-08
LLYDANYFV	101.5	34.37	175.9	141.8	180.9	180.5	200.5	67.61	15.33	607.8	0.31	-10.1	3.70E-08
FAAELTIGV	53.73	56.41	104.6	260.5	261	143.4	134.2	31.48	17.49	620	0.32	-9.5	1.00E-07
LMIFISSFL	99.78	25.64	129.4	102.3	237.6	172.7	164.4	77.9	23.47	600.5	0.32	-9.8	6.80E-08
FLITGVFDI	69.77	19.68	114.4	258.7	214.1	142.2	142.9	39.86	14.68	582.9	0.33	-9.5	1.20E-07
YLSSWTPVV	78.75	21.04	114.9	290.7	373.3	98.2	56.2	76.89	20.08	686.9	0.33	-9.7	8.10E-08
FLLPSLATV	67.65	21.72	69.35	193.5	194.9	174.8	142.5	75.69	29.72	541.7	0.34	-8.6	5.20E-07
YLMCLSPML	96.17	20.96	135.3	214.8	243.6	131.2	121.7	89.05	33.67	642.5	0.34	-9.5	1.00E-07
KVYDKLFPV	32.07	41.86	59.17	132	192.6	47.13	60.32	41.14	17.17	527.5	0.34	-10.5	1.90E-08
SMSSYDFST	96.21	32.74	123.1	238.4	388	100.8	122.6	41.23	29.12	766.9	0.34	-10.7	1.40E-08
KVYDKLFPV	32.07	41.86	59.17	132	192.6	47.13	60.32	41.14	17.17	527.5	0.34	-10.5	1.90E-08
YLCFLAFLL	50.95	19.03	161.2	305	348.8	224.9	184.4	49.38	25.2	768.4	0.35	-10.4	2.50E-08
KIDYYIPYV	102.9	32.28	110.1	183.2	286.4	145.8	145.4	112.8	24.03	693	0.35	-11	8.00E-09
MLNGIMYRL	79.02	24.18	17.94	80.44	173.9	53.47	54.97	198.6	23.55	514.9	0.35	-9.7	8.20E-08

YLLCCNYKL	74.15	22.71	136.5	251.9	311.9	219.9	267.5	128.3	18.49	790	0.36	-9.3	1.50E-07
MLMFIFTGI	37.81	30.2	70.13	220.5	231.3	170.7	172.5	44.91	22.52	632.9	0.38	-9	2.40E-07
YILCNMALL	93.77	22.65	44.24	159.9	198.2	180.4	134.4	70.37	29.68	563.4	0.38	-9.8	6.70E-08
VMCGGSLYV	65.84	25.56	172.6	168.6	311.3	182.5	177.5	70.26	16.9	629.8	0.39	-9.2	1.80E-07
AQFLYLYAL	72.77	55.12	53.8	158.7	136.2	67.29	77.32	46.57	37.58	555.7	0.39	-11.3	5.50E-09
MIFISSFLL	121.5	33.21	110.9	130.3	108.4	95.47	115.3	52.27	30.74	539.1	0.39	-9.4	1.30E-07
AAAWYLWEV	102.1	70.41	74.51	229.7	105.6	61.82	143.7	35.74	17.31	571.2	0.39	-9.8	6.80E-08
VLWAHGFEL	67.27	25.13	163	108.2	238.6	186.8	174.9	24.41	22.26	580.5	0.4	-9.4	1.30E-07
KTMVAITYEL	64.97	40.93	129.6	148.8	245.6	158.7	186.1	35.17	22.67	602.5	0.4	-11.2	6.30E-09
GMANTTFHV	28.74	67.88	107.2	340.9	263	152	159.2	35.26	10.28	702.3	0.41	-9	2.30E-07
MLLALVALV	45.64	27.16	96.98	101.7	228.1	146.2	195.5	41.9	19.47	518.6	0.42	-9.7	8.30E-08
VLAWLYAAV	72.77	26.22	90.52	142.5	275.7	219.2	174.8	79.47	20.78	642.6	0.43	-10	4.80E-08
KLGNLLLLL	132.2	31.01	84.4	189.4	302.3	211.9	159.5	25.76	15.97	633.1	0.43	-10.6	1.60E-08
YLDFLLLLL	91.39	22.75	80.85	131.1	191.5	125.7	134.3	31.24	23.17	478	0.43	-10.9	1.10E-08
MLDDFSAGA	128.8	27.25	126.1	156.8	286.5	214.8	184.3	67.56	31.12	721.4	0.43	-9.1	2.20E-07
GLFQLIFFL	54.07	74.72	85.44	267.5	68.72	74.14	82.83	56.63	28.11	571.8	0.44	-10.3	2.80E-08
ALNHLVLSL	73.13	32.88	166.4	204	229.6	119.5	163.8	27.14	18.2	638.5	0.45	-9.4	1.30E-07
KLFTIAMWL	132	24.92	165.7	134.5	290.8	198.6	189.4	147.6	24.92	748.4	0.46	-10.8	1.20E-08
ALIARCWYL	62.03	31.45	76.73	225	160.5	123.7	142.9	126	18.94	610.1	0.46	-10	4.60E-08
LLSYVVVVV	86.23	23.59	127.7	214.9	292.5	120.4	131.1	96.52	12.35	644.8	0.46	-10.3	3.00E-08
FLHTTFIDV	78.01	19.83	104.5	210.9	141.6	158.5	127.2	27.55	16.41	503.8	0.46	-9.7	7.70E-08
TLMNVITLV	79.6	29.32	83.78	191.3	239.8	150.7	157.2	48.26	15.91	633.8	0.47	-10.4	2.30E-08
NLLDSYFVV	80.92	27.06	109.5	279	134.9	116.4	249.1	78.36	85.75	789.6	0.48	-10.6	1.50E-08
FTLIDIWFL	85.14	18.76	102.2	134.1	170.6	166	200.9	69.34	30.1	539.2	0.49	-11.3	4.90E-09
TLIDIWFLA	70.49	26.4	49.46	171.2	126.4	96.49	83.48	84.28	68.52	549.6	0.49	-10.5	1.90E-08
KLFVYYYNL	112.3	36.04	79.9	198.4	128.9	116.8	220.7	109.4	58.91	665.2	0.49	-10.1	3.80E-08

Tabela Suplementar 4: Dados de extração de valores de volume de pMHC “Negative”. Ordenada por valor de IC50 e contendo valores de ΔG e K_d .

Negatives (95 alvos)													
Extração de Volumes											IEDB	Prodigy	
Alvos	aa1	aa2	aa3	aa4	aa5	aa6	aa7	aa8	aa9	Cadeia c	IC50	deltaG	Kd
TEVGQDQYV	84.5	30.37	96.8	110.6	257.4	131.7	196.7	111.8	14.06	628.6	250000	-10.9	9.70E-09
TRHPATATV	81.38	32.09	143.3	115.7	193.6	172	166.4	49.13	49.13	571.7	250000	-10.8	1.10E-08
RVRFFFPSL	57.22	30.58	144.2	160	161.7	112.6	106.5	36.57	15.92	494.8	200000	-9.7	7.70E-08
SILGIGTVL	97.36	40.69	153.9	73.03	287	215.4	208	39.43	39.32	646.7	100000	-10.3	2.60E-08
YLEPAIAKY	81.01	22.71	126.4	155.4	164.3	195.1	212.5	120.3	83.41	691.5	100000	-9.6	8.80E-08
AAEQRRSTI	56.51	61.21	40.3	229.4	177.6	153.2	127.3	40.8	13.35	620.8	70000	-8.8	3.70E-07
AARLKRSAT	71.19	64.58	133.6	314.6	142.5	157.4	66.92	43.61	29.01	740.2	70000	-9.6	9.00E-08
AARNIVRRA	59.87	65.82	56.54	252.6	102.4	34.76	81.74	155.2	47.81	673.4	70000	-9.7	7.40E-08
AAVSADPLA	57.53	60.55	134.9	228.6	232.5	124.6	166.8	50.23	42.08	712.4	70000	-9	2.50E-07
AVKGVGTMV	51.59	43.81	201.2	101.7	285.5	183.3	173.7	49.88	15.6	631.9	70000	-9.5	1.20E-07
AVVKSDNKL	55.65	45.65	147.6	314.2	193.1	135.3	209.1	71.57	40.71	785.8	70000	-8.5	5.90E-07
CIPRSKML	61.17	36.58	207.1	231.8	298.8	127.7	248.7	58.3	35.37	761	70000	-8.9	2.80E-07
CIRNASKFV	75.4	34.7	123.3	285.1	91.37	74.46	101.7	53.09	15.96	594.7	70000	-10	5.00E-08
CTEETKRNI	85.21	40.14	20.12	212.6	79.58	98.14	104.6	88.04	18.9	537.9	70000	-10.4	2.20E-08
DAKRNSKSL	117.4	56.81	157.2	141.7	159.1	158	154.9	48.42	14.45	598.8	70000	-9.1	2.00E-07
DICSKHMDA	91.69	30.96	129.2	161.6	242.1	152.6	130.5	100.3	80.45	712.5	70000	-10.4	2.50E-08
DINESMSQM	94.58	32.18	102.6	299.9	202.5	162.3	162	63.58	41.09	754.4	70000	-9.6	9.20E-08
DISVNASKT	72.11	31.26	31.26	205.3	361.3	183.2	205.1	82.39	35.5	733.2	70000	-9.3	1.40E-07
DITFLRPVL	79.59	29.87	77.02	284.3	186.4	127.7	114.1	32.97	18.37	615.8	70000	-10.7	1.50E-08
DLNRMPTDM	60.14	27.34	166.8	162.2	182.4	158.4	153.7	60.72	44.46	593.3	70000	-9.8	6.30E-08
DTCGASINI	64.1	47.11	193.5	122.3	271.2	189	180.8	26.7	11	600.9	70000	-10.3	3.00E-08
DTEFINKFL	109.8	25.82	175.2	144.4	213.3	175.2	203.4	47.18	16.14	629.1	70000	-11.5	3.90E-09
DTLVKSGLT	93.32	36.37	96.96	104.6	252	199.2	199.6	89.98	30.29	650	70000	-10.8	1.20E-08
DVEKEKFVA	55.65	45.65	147.6	314.2	193.1	135.3	209.1	71.57	40.71	785.8	70000	-10.4	2.20E-08
DVGCLLTD	81.13	31.32	109.8	247.2	306.7	186.6	169	72.41	29.57	729.4	70000	-8.6	4.90E-07
DVHPGEPVV	64.2	39.98	126.5	121	143.5	163.3	150.5	38.02	22.91	539.9	70000	-9.8	6.00E-08
DVRTLLGLI	83	42.97	107	192.1	152.5	107.5	110.7	72	22.33	598.6	70000	-10.6	1.70E-08
EAFETQSGA	139.3	51.11	94.85	122.7	171.2	166.5	163.2	44.97	56.26	625.3	70000	-9.5	9.90E-08
EIARIENEM	93.48	33.28	95.74	235	360.9	237.9	248.8	45.43	36.85	791.5	70000	-10	5.00E-08
EIDVSEVKT	123.8	32.49	109.3	176.8	177.5	162	138.6	77.76	29.97	644.9	70000	-9.9	5.60E-08
EIEKVEKYL	97.8	31.4	66.95	167.9	184.4	154.2	158.2	137.5	45.16	692.9	70000	-10.7	1.40E-08
ELVMDKNHA	103	24.6	48.9	174.6	238.9	183.9	187.4	50.68	57.19	667.2	70000	-9.6	9.70E-08

EMKINRQIL	95.97	30.04	143.3	138.3	121.3	265.7	197.1	81.05	26.47	656.4	70000	-10	4.80E-08
ETGLSASDV	76.96	53.38	142.2	301.1	346.9	243.4	195	67.78	12.01	784.6	70000	-10.4	2.40E-08
ETIQKDINI	102.1	49.15	104.4	141.5	191.9	133	157.8	31.93	20.03	574.8	70000	-10.7	1.40E-08
ETLNEYKQL	80.18	29.12	101.9	230.6	142.8	111.2	242.4	72.68	49.35	693.3	70000	-10.8	1.20E-08
ETVKMGAFM	102.7	54.08	120.6	168.2	291	199.1	173	88.33	35.16	748.6	70000	-10.8	1.20E-08
EVKTLSSYI	122.3	34.61	210.9	158.7	265.8	177.1	216.3	93.97	12.4	722.3	70000	-9.8	7.00E-08
EVRIPVDLV	110.3	34.99	113.5	138.2	137	62.56	115.2	52.18	9.663	509.2	70000	-10.7	1.40E-08
FTELENKKV	75.94	35.76	139	234.8	171.6	173.6	134.2	41.81	19.73	604.1	70000	-9.9	5.60E-08
FTSYKRFVT	35.61	43.71	109.2	348.5	234.5	193.8	193.8	193.8	193.8	688.6	70000	-11.1	7.30E-09
GAAVTLNRI	32.34	99.3	126.9	246.5	245.5	133.8	160.9	166.1	17.49	737.3	70000	-8.5	6.00E-07
GAGVLDKDL	31.47	99.67	149	270.4	282.6	205.5	262.6	74.09	26.81	707.3	70000	-8.9	2.80E-07
ITETIPIGM	53.49	47.26	182.5	172.9	249.3	152.4	186.8	67.66	40.02	674.9	70000	-10.6	1.80E-08
ITNILGGVL	74.92	35.73	168.5	203.9	342.2	269	239.8	86.62	29.81	781.7	70000	-10.6	1.60E-08
IVKNIREGT	106.7	42.37	47.78	256.4	89.44	104.9	75.13	62.39	33.04	593.9	70000	-10.2	3.40E-08
IVSHLRAST	51.91	39.85	62.22	240.6	201	175.4	93.34	29.68	24.94	608.6	70000	-9.3	1.50E-07
KLKHRDGFT	92.17	35.79	123.2	185.7	180.4	143.5	105.5	83.72	22.73	561.2	70000	-10.8	1.30E-08
KVDDTFYIV	108.6	28.13	94.58	122.5	283.3	198.8	243.7	139.2	23.06	705.7	70000	-10.4	2.50E-08
LAARLKRSA	76.52	54.56	65.19	268.6	240.5	152.3	148.9	26.29	32.68	663.3	70000	-8.6	5.30E-07
LLKNMKQCT	119	30.07	74.52	244.1	223	141.3	139.7	42.2	33.81	692.7	70000	-9.8	6.80E-08
LTHGADPNA	74.32	35.81	145.1	75.33	179.4	137.1	143.8	57.64	40.26	558.3	70000	-9.4	1.20E-07
LTSNCTRRT	91.04	39.38	112.1	256.6	183.3	102.4	221.8	41.33	25.84	633.4	70000	-10.2	3.10E-08
LVDKEDTDI	73.28	37.89	118.5	100.2	233	185.2	168.7	82.06	14.18	601	70000	-8.9	2.90E-07
LVSKHWELT	64.86	38.9	95.35	219.3	254.2	182.7	136.9	53.58	59.95	656.6	70000	-9.8	7.00E-08
MAAEQRRST	114.5	53.96	37.92	208.7	117.9	136.2	108.3	22.6	25.56	568.7	70000	-8.7	3.90E-07
NAKIAARIL	108.3	59.12	147	291.8	145.7	150.3	120.9	43.52	20.28	697.4	70000	-10.9	9.60E-09
NAQQFANVI	106.7	56.75	171.1	323.6	223.4	157.2	193.8	42.1	19.79	816.7	70000	-10.5	1.90E-08
NAVTTNLL	92.57	62.25	103.2	218.6	267.5	177.3	194.2	31.2	22.13	715.4	70000	-10.3	2.90E-08
NIIKNKCSI	79.64	41.63	47.47	276	152.3	106.2	100.5	26.94	19.53	563.8	70000	-10.7	1.50E-08
NITHNITT	128.4	42.76	139.4	286.9	234	179.2	143.7	33.51	34.51	695.8	70000	-10	5.00E-08
NMKQCTNDI	115.8	31.55	152.6	225.4	217.5	120.1	209.4	79.43	18.4	738.9	70000	-10.4	2.20E-08
NTCKPTILA	71.2	38.36	114.2	229.5	251.5	146.5	145.7	34.64	57.41	635.1	70000	-12.6	5.40E-10
NTKSDIDVI	92.84	38.99	161.4	141.7	247.6	169.1	171.3	20.45	11.85	581.8	70000	-12	1.70E-09
PANINDKQI	55.22	55.18	195.6	236.8	210.9	148.9	174.9	36.11	14.73	641.2	70000	-8.7	4.50E-07
PMIIGEPII	66.52	32.49	114.4	294.3	152	152.5	135.9	44.61	14.36	684.1	70000	-9.6	9.00E-08
PTDMLKLFT	73.43	37.51	89.2	298.7	225.5	174.5	171.5	78.07	33.04	715	70000	-11.4	4.30E-09
PTIEDDKIV	51.99	39.79	139.2	255.9	202.9	119.8	174	42.64	20.74	654.6	70000	-10.8	1.20E-08
PTILATLNT	48.37	45.37	136.8	250.1	196.5	119.6	159.5	41.96	40.6	655.7	70000	-9.5	1.10E-07

PTLFGRGVI	62.37	22.74	112.6	174.2	214.4	318.9	201.8	87.43	18.8	679.4	70000	-10.8	1.10E-08
PVVKDKIKL	64.33	39.81	124.4	283.8	242.1	208.5	158.9	58.58	29.73	700.1	70000	-9.2	1.60E-07
QAELTSNCT	139.2	52.69	181.6	202.6	257.9	163.7	207.8	54.63	24.73	746.1	70000	-9.8	6.80E-08
RAAVSADPL	197.9	59.25	134.6	200.2	308.9	218.1	178.8	37.64	26.48	740.8	70000	-9.4	1.20E-07
RAFTEEGAI	155.5	57.14	68.73	14.69	152.4	171.1	183.9	69	21.26	581	70000	-9.9	5.70E-08
RASTTENAA	202.3	57.4	138	167.9	249.4	194.3	170	23.22	36.21	675.5	70000	-11.1	6.80E-09
RVCWLHECT	35.92	40.05	86.24	115.9	165.6	155.1	155.4	41.8	38.51	502.6	70000	-11.1	6.80E-09
SAAFEDLRL	115.8	65.28	98.37	139.1	208.9	164.4	175.9	175.9	21.54	633.6	70000	-10.7	1.50E-08
SAPQQLCTM	112.4	58.45	119.1	290.4	280.8	175	225.3	52.09	29.52	821.5	70000	-9.6	9.90E-08
SASDMQKFT	94.36	62.98	133.3	305.6	229.4	190.6	148.2	48.34	34.42	758.7	70000	-10.7	1.40E-08
TIEDDKIVT	97.18	29.51	92.84	208.8	182.2	167.3	144.7	22.18	38.04	591.1	70000	-11.1	7.60E-09
TIQKDINIT	81.7	32.32	139.4	154.4	205.4	119.7	197	48.56	25.06	599.5	70000	-10.8	1.20E-08
TQFNFNIGHT	80.96	39.09	84.28	133	248.7	192.4	193.3	90.3	32.2	686.2	70000	-11.5	3.60E-09
TTGEWPLII	76.24	53.91	156.5	295.4	594	213.7	202.6	29.31	14.5	954.3	70000	-12.2	1.20E-09
VIGLTHCT	62.58	35.14	119.3	293	346.4	200.9	184.2	24.62	36.53	709.7	70000	-8.9	3.10E-07
VINRVSENT	79.68	32.16	140.3	198.6	242.6	166.8	183	38.43	32.85	624.3	70000	-8.6	5.10E-07
VIRNEVNDT	79.68	32.16	140.3	198.6	242.6	166.8	183	38.43	32.85	624.3	70000	-9.5	1.10E-07
VQRTRCKYV	83.46	34.08	95.64	180.8	130.3	68.24	83.43	129.2	14.16	566.1	70000	-11.6	3.30E-09
VTKRDESSI	105.3	58.54	138.7	126.3	167	185.1	161.6	24.69	12.45	560	70000	-10.1	4.00E-08
VTVPTNDHI	63.05	31.3	102.6	149.3	241.7	165.3	166.3	30.47	19.67	555.6	70000	-9.3	1.50E-07
VVYHDDNT	82.51	39.36	131.7	131.5	192.1	171.5	190.1	34.65	33.4	575.2	70000	-8.7	4.40E-07
WLRAKRKPA	108.4	24.72	151.1	226.9	181.1	167	101.4	35.78	38.19	634.6	70000	-9.8	7.10E-08
WTGNYFTDT	157.8	58.79	124.6	326	420.5	224.8	137.9	60.5	21.33	928.9	70000	-10.1	4.20E-08
WVKKGGHVT	42.09	40.4	197.5	205.6	272.8	238	205.7	18.01	34.37	659.9	70000	-9.7	7.80E-08
YNFSLGAAV	96.02	35.87	181.6	198.8	261.5	268.5	218.8	42.01	16.47	673.2	70000	-10.3	2.90E-08
YTAGNKVDV	51.95	34.67	76.89	149.8	236.9	172.3	114.8	69.13	24.8	542.1	70000	-10.2	3.30E-08

Os próximos dados se referem aos testes estatísticos realizados para confirmação dos primeiros resultados encontrados com o Teste de Wilcoxon-Mann-Withney. O objetivo era comparar peptídeos ligantes ($IC_{50} < 0,5$ nM) e não ligantes ($IC_{50} > 70.000$ nM) quanto ao tamanho do *gap* na molécula, sendo este medido pelo volume e partindo do princípio que quanto menor o volume, melhor a ligação.

Tabela suplementar 5: Matriz de correlações

	aa1	aa2	aa3	aa4	aa5	aa6	aa7	aa8	aa9
Pearson	1	-.036	.030	-.202**	.036	.131	.063	.075	-.045
aa1 Sig. (2-tailed)		.619	.677	.005	.625	.072	.389	.304	.538
N									
Pearson	-.036	1	-.008	.244**	-.098	-.081	.032	-.062	.033
aa2 Sig. (2-tailed)	.619		.911	.001	.180	.267	.665	.395	.650
N									
Pearson	.030	-.008	1	.075	.313**	.394**	.462**	-.144*	-.079
aa3 Sig. (2-tailed)	.677	.911		.302	.000	.000	.000	.048	.276
N									
Pearson	-.202**	.244**	.075	1	.228**	.063	.052	-.088	.133
aa4 Sig. (2-tailed)	.005	.001	.302		.002	.390	.480	.227	.067
N									
Pearson	.036	-.098	.313**	.228**	1	.518**	.373**	.008	-.113
aa5 Sig. (2-tailed)	.625	.180	.000	.002		.000	.000	.915	.122
N									
Pearson	.131	-.081	.394**	.063	.518**	1	.596**	-.061	.027
aa6 Sig. (2-tailed)	.072	.267	.000	.390	.000		.000	.405	.710
N									
Pearson	.063	.032	.462**	.052	.373**	.596**	1	.000	.137
aa7 Sig. (2-tailed)	.389	.665	.000	.480	.000	.000		.996	.059
N									
Pearson	.075	-.062	-.144*	-.088	.008	-.061	.000	1	.300**
aa8 Sig. (2-tailed)	.304	.395	.048	.227	.915	.405	.996		.000
N									
Pearson	-.045	.033	-.079	.133	-.113	.027	.137	.300**	1
aa9 Sig. (2-tailed)	.538	.650	.276	.067	.122	.710	.059	.000	
N	190	190	190	190	190	190	190	190	190

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Tabela suplementar 6: Aplicação de GEE

Prob distribution	Link function	Working correlation matrix structure	Quasi likelihood under independence model criterion (QIC)	Corrected quasi likelihood under independence model criterion (QICC)	P grupos	P AA	P interação grupo* AA
normal	identity	independent	3 369 852,2	3 369 852,2	0,002	<0,001	0,030
normal	identity	AR(1)	3 369 852,2	3 369 852,2	0,002	<0,001	0,030
normal	identity	Exchangeable ^a	3 369 852,2	3 369 852,2	0,002	<0,001	0,030
normal	identity	unstructured	3 369 852,2	3 369 852,2	0,002	<0,001	0,030

^aO mesmo que compound symmetry

A publicação abaixo informada consta dos anexos dessa dissertação, pois tem uma relação direta com a metodologia de modelagem aqui aplicada. E, também, sendo coautora, por ter tido um grande envolvimento no seu desenvolvimento.

SCIENTIFIC REPORTS

OPEN

DockTope: a Web-based tool for automated pMHC-I modelling

Maurício Menegatti Rigo^{1,*}, Dinler Amaral Antunes^{1,2,*}, Martiela Vaz de Freitas¹, Marcus Fabiano de Almeida Mendes², Lindolfo Meira³, Marialva Sinigaglia² & Gustavo Fioravanti Vieira¹

Received: 28 June 2015

Accepted: 18 November 2015

Published: 17 December 2015

The immune system is constantly challenged, being required to protect the organism against a wide variety of infectious pathogens and, at the same time, to avoid autoimmune disorders. One of the most important molecules involved in these events is the Major Histocompatibility Complex class I (MHC-I), responsible for binding and presenting small peptides from the intracellular environment to CD8⁺ T cells. The study of peptide:MHC-I (pMHC-I) molecules at a structural level is crucial to understand the molecular mechanisms underlying immunologic responses. Unfortunately, there are few pMHC-I structures in the Protein Data Bank (PDB) (especially considering the total number of complexes that could be formed combining different peptides), and pMHC-I modelling tools are scarce. Here, we present DockTope, a free and reliable web-based tool for pMHC-I modelling, based on crystal structures from the PDB. DockTope is fully automated and allows any researcher to construct a pMHC-I complex in an efficient way. We have reproduced a dataset of 135 non-redundant pMHC-I structures from the PDB (C α RMSD below 1 Å). Modelling of pMHC-I complexes is remarkably important, contributing to the knowledge of important events such as cross-reactivity, autoimmunity, cancer therapy, transplantation and rational vaccine design.