

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

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 conseqüente 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)

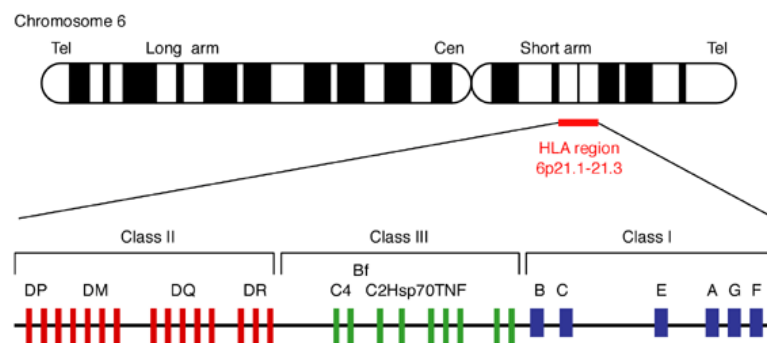


FIGURA 1: Mapa gênico humano da região HLA. A Região do MHC, encontrada no braço curto do cromossomo 6, subdividida em três *loci* principais: o MHC de classe I, o MHC de classe II e o MHC de classe III. (Modificado de *Mehra & Kaur, 2003. 4p.*)

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; Wieczorek et al, 2017)

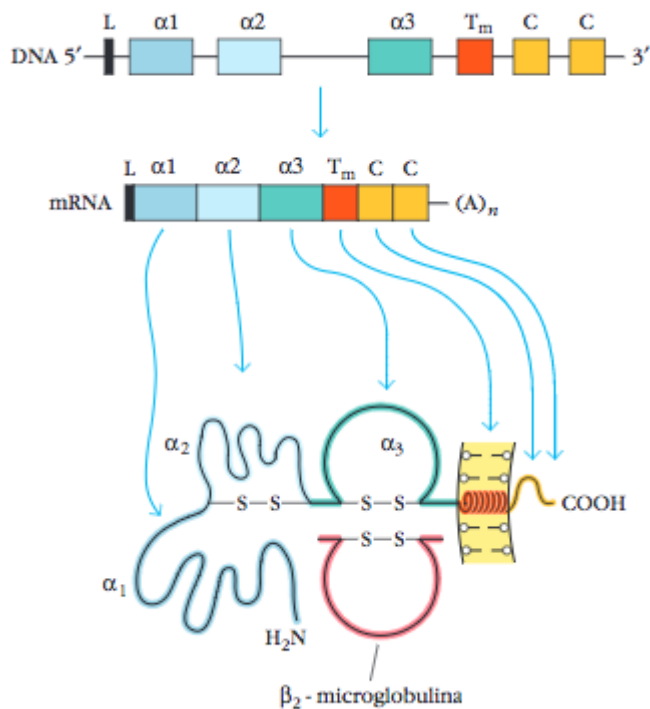


FIGURA 2: Diagrama esquemático do MHC de classe I. Correspondência entre os éxons e os domínios no produto gênico. A β_2 -microglobulina é proveniente de um cromossomo diferente. T_m e C dizem respeito aos domínios transmembrana e citoplasmático, respectivamente. (Adaptado de Kindt et al., 2008)

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; Wieczorek 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).

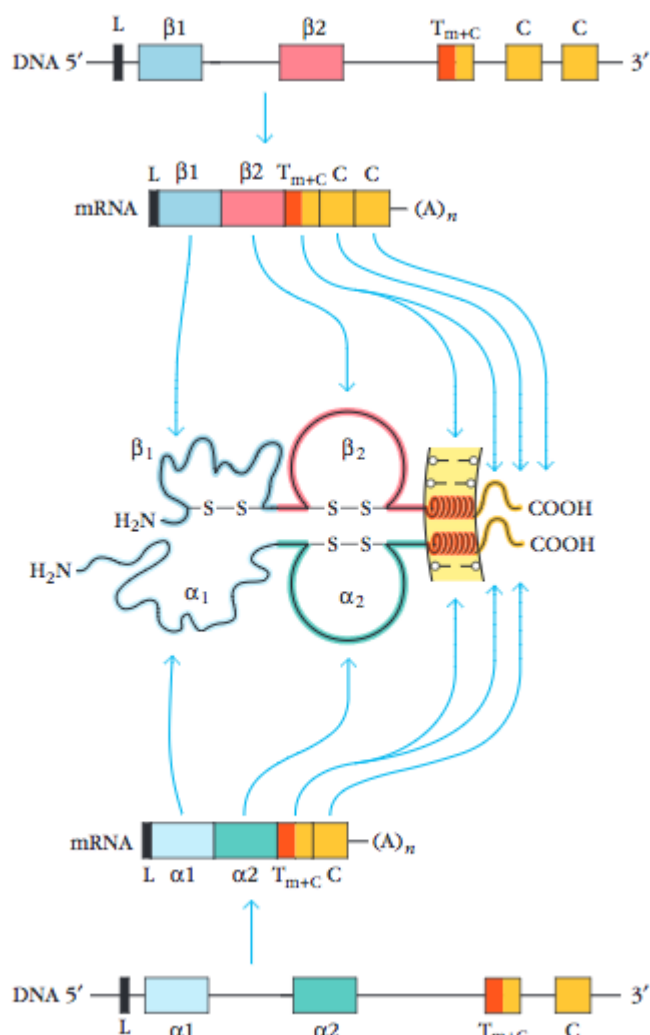


FIGURA 3: Diagrama esquemático do MHC de classe II. Correspondência entre os éxons e os domínios nos produtos gênicos das cadeias α e β . T_m e C dizem respeito dos domínios transmembrana e citoplasmático, respectivamente. (Adaptado de *Kindt, et al., 2008*)

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)

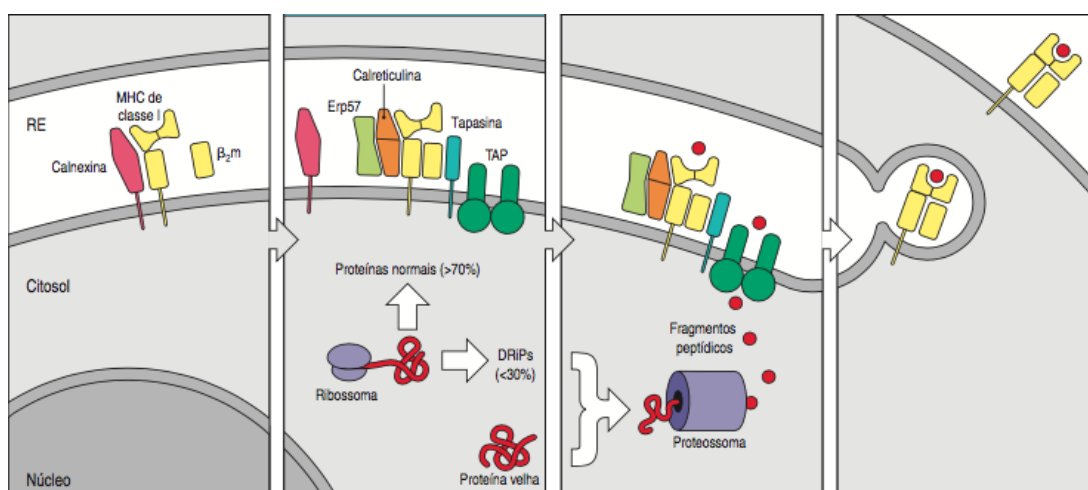


FIGURA 4: Via endógena da apresentação de antígenos. Os fragmentos de proteínas próprias ou virais são clivados pelo proteossomo e ficam disponíveis no citosol. Esses fragmentos são, então, encaminhados para o retículo endoplasmático, onde são associados ao MHC-I. Após são direcionados o Complexo de 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).

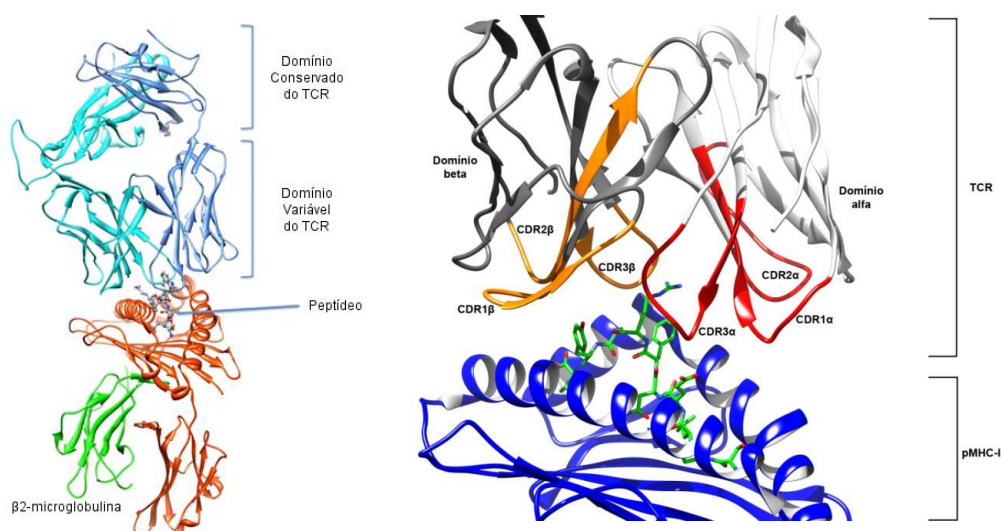


FIGURA 5: Interação entre TCR e complexo pMHC. Ressaltadas as regiões de alta variabilidade do TCR e sua interação com a região da fenda do MHC. (Adaptado de Rigo, 2015a. p.18)

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 efetoras 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

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

reversa traz, *in silico*, a possibilidade de identificação dos antígenos candidatos que 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:

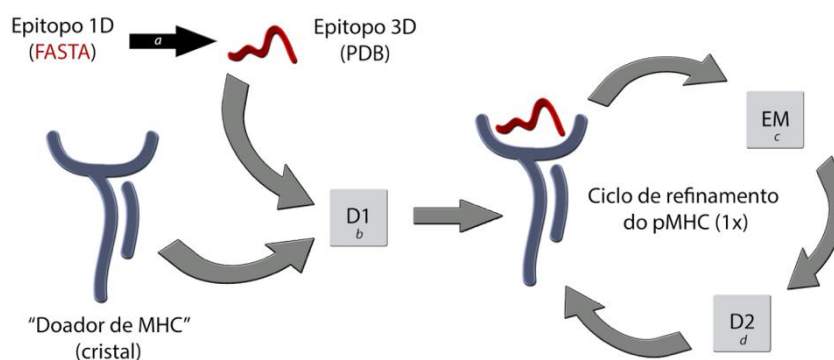


Figura 6: Esquema visual da técnica D1-EM-D2. A estratégia de docagem de peptídeos em MHC-I consiste em 4 etapas. Como entrada para a criação dos modelos é solicitada apenas a sequência

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).

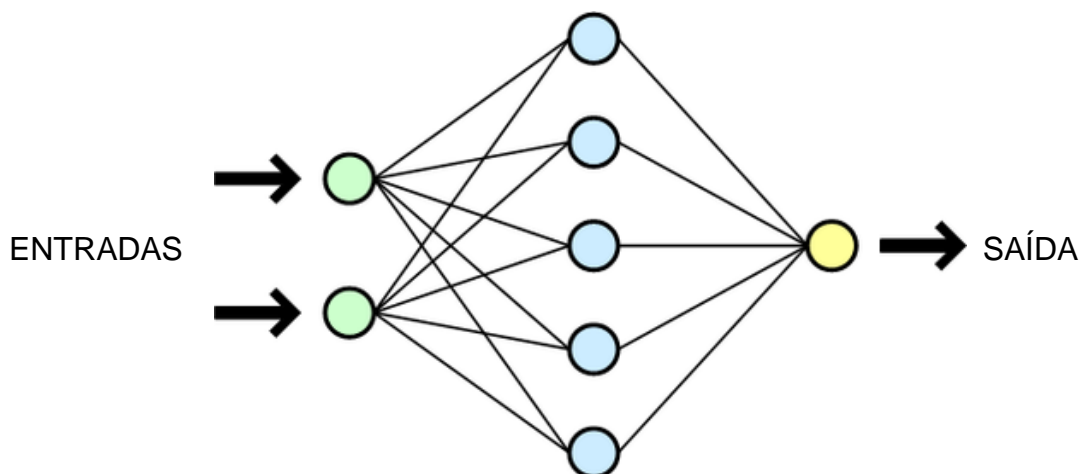


FIGURA 7: Representação gráfica de uma rede neural. Uma rede neural artificial é composta de grupos de nodos, interconectados. Aqui, cada nodo representa um neurônio artificial e cada conexão entre o output de um neurônio e o input de outro é representado pelas setas. No caso deste trabalho as entradas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

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 Docktopé, 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 X 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 IC50 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 IC50, 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, Lemaltre 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, Carilion E, Duvergey H, Houles A, Paysan-Lafosse A (2015) IMGT, the international ImMunoGeneTics information system 25 years on. *Nucleic*

Acids Res 43 (D1): D413-D422.

- LeibundGut-Landmann S, Waldburguer 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) *Imunobiologia de Janeway.* 7 edition. Artmed, Porto Alegre. 899 pp.
- Neefjes J, Jongsma MLM, 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 in 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 peptic 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 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

^a O mesmo que compound symmetry

ANEXOS

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.

www.nature.com/scientificreports

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 Sinigaglia¹ & 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.