

Universidade Federal do Rio grande do Sul

Dissertação de Mestrado

Construção de redes de reatividade cruzada a partir da inferência de relações de similaridade da área acessível ao solvente e da distribuição eletrostática da região de interação com o receptor de células T entre complexos peptídeo:HLA:0201

Marcus Fabiano de Almeida Mendes

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

Orientador: Prof. Dr. Gustavo Fioravanti Vieira

Porto Alegre

Abril de 2014

Apoio Financeiro

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

FAPERGS – Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul

Bill & Melinda Foundation (Grand Challenges Exploration – Round 2)

AGRADECIMENTOS

Quero agradecer a todos que me ajudaram nessa longa jornada que foi minha vida acadêmica. Primeiramente a minha mãe Sandra, meu irmão Gustavo e minha vó Celi, por sempre estarem do meu lado, mesmo estes não querendo que eu estivesse tão longe de casa.

Depois agradeço ao meu Orientador Gustavo e toda a equipe do laboratório NBLI, principalmente a Meg, o Dinler e o Mauricio pelo apoio tanto no lado acadêmico como no lado pessoal. Este mestrado é mais de vocês que meu!

Agradeço minha família, por sempre me acolherem tão bem quando volto para casa e sempre acreditarem em mim. Em especial ao tio Didi pelos churrascos e cervejadas, a tia Eliane pelos mimos e a todos os outros tios, tias, ao meu primo Edinho e as primas, e outros parentes que me fortalecem e me incentivam a crescer cada vez mais.

Aos meus amigos de Campo Grande, em especial o Aruã, o Heverson, o Carlos, o Kenji e o Tiaguinho, pelos tererés e a amizade de vocês. Um agradecimento também aos amigos de Dourados, que foram de especial ajuda principalmente no começo da minha jornada acadêmica.

Aos meus amigos de Porto Alegre, que não deixam eu ir embora! Em especial ao Artur, Marcelo, Liana, Ivi, Paula, Gabriel, Rafael, Rodrigo, Lucio, Matzen e tantos outros que são grandes pessoas e fazem eu me sentir em casa mesmo estando tão longe de Campo Grande. A Aida, o Gilson e a Luana por terem sido fundamentais nessa minha ida a Porto Alegre.

Ao pessoal do laboratório de Imuno do centro de pesquisa biomédica da PUCRS, que me incentivaram a seguir nesta área e abriu as portas para o universo de imunologia, em especial a Cristina Bonorino, Zanin, Borges, Taiane, Lisboa e a Deise. Ao pessoal do laboratório de Imunogenética da UFRGS, em especial ao Zeca e a Francis. E também ao Elmo, por sempre estar disponível para questões de mestrado, congressos e afins.

E por ultimo, mas não menos importante, o pessoal da Housing, o qual tive grande contato, troca de ideias e uma amizade muito grande. Apesar de interagir com quase todos, mando um agradecimento especial para o Eder, Orlando, Diego, Gabrieli, Evelyn, Reginaldo, Eduardo, Dani, Esperanza, Thays, Leticia, Jenifer, Dario, Flávio, Gustavo, Jimena, Fernando, Giovani e tantos outros que sempre me animaram e me enriqueceram com ideias, alegrias e amizades.

E por fim, a todas as outras pessoas que eu não citei, porém participaram dessa minha jornada, um muito obrigado! Essa é para vocês!

*“The right man in the wrong place can make all the
difference in the world.”*

G-Man

SUMÁRIO

| | |
|--|-----------|
| Abreviaturas | 6 |
| Resumo..... | 7 |
| Abstract | 8 |
| <i>Capítulo I.....</i> | <i>9</i> |
| <i>Introdução e Objetivos.....</i> | <i>9</i> |
| Introdução Geral | 10 |
| Introdução histórica | 10 |
| Resposta Imune Inata e Adaptativa | 11 |
| Rota de Processamento e Apresentação de Antígeno..... | 13 |
| MHC-I E TCR: protagonistas da resposta citotóxica | 14 |
| Reatividade Cruzada..... | 15 |
| Vírus da Dengue | 16 |
| Vírus da Hepatite C..... | 17 |
| Imunoinformática aplicada ao desenvolvimento de vacinas..... | 18 |
| Objetivos | 21 |
| Objetivo Geral: | 21 |
| Objetivos Específicos:..... | 21 |
| <i>Capítulo II.....</i> | <i>22</i> |
| Improved structural method for T-cell cross-reactivity prediction..... | 23 |
| <i>Capítulo III.....</i> | <i>43</i> |
| Discussão Geral | 44 |
| Figuras | 50 |
| Referencias Bibliográficas | 53 |

Abreviaturas

- ASA – Superfície Acessível ao Solvente (do inglês *accessible surface area*).
- 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*).
- ER – Retículo Endoplasmático (do inglês *Endoplasmic Reticulum*).
- ERAP – Aminopeptidase do Retículo Endoplasmático (do inglês *Endoplasmic Reticulum Aminopeptidase*).
- HCA – Análise de Agrupamento Hierarquico (do inglês *Hierarchical Clustering Analysis*).
- HCV – Vírus da Hepatite C (do inglês *Hepatitis C Virus*).
- HLA – Antígeno Leucocitário Humano (do inglês *Human Leucocyte Antigen*).
- MHC – Complexo Principal de Histocompatibilidade (do inglês *Major Histocompatibility Complex*).
- NK – Células Assassinas Naturais (do inglês *Natural Killer*).
- ORF – Fase Aberta de Leitura (do inglês *Open Reading Frame*).
- PDB – *Protein Data Bank*
- PLC – Complexo de Carregamento do Peptídeo (do inglês *Peptide Loading Complex*).
- p:MHC – Complexo MHC com o peptídeo ancorado em sua fenda.
- TAP – Transportador Associado ao Processamento de Antígeno (do inglês *Transporter Associated with Antigen Processing*).
- TCR – Receptor de Linfócitos T (do inglês *T cell Receptor*).
- Treg – Células T Reguladoras (do inglês *Regulatory T cells*).

Resumo

O sistema imunológico continua, ao menos parcialmente, sendo uma caixa preta. Com grande quantidade de pesquisa na área, estamos aos poucos decifrando os eventos que se procedem dentro desta caixa, como por exemplo, a reatividade cruzada. Este fenômeno ocorre essencialmente quando um receptor de célula T (TCR) interage com diferentes moléculas de MHC carregadas com peptídeo (p:MHC). A reatividade cruzada é um evento importante, com consequências sobre o desenvolvimento de vacinas e imunoterapias. Entretanto, a complexidade deste sistema tem atrasado o progresso nesta área, e o desenvolvimento de um método confiável de predição de reatividade cruzada se mantém um desafio para os imunologistas. Neste trabalho, nós utilizamos características estruturais dos complexos p:MHC de regiões de contato com o TCR e um diferente método de análise estatística para melhorar a técnica de predição de reatividade cruzada previamente publicada pelo nosso grupo. Assim, o nosso grupo de análise inclui um grupo de 28 epitopos de hepatite C, usado para testar a antiga abordagem, mais 8 epitopos de 4 subtipos de dengue testados *in vitro* em um trabalho previamente publicado. Utilizando a análise de clusterização hierárquica (HCA) com cálculo de bootstraps com os dados extraídos do potencial eletrostático e a superfície acessível ao solvente (ASA) de regiões selecionadas do p:MHC, foi criada uma rede com complexos de p:MHC, os quais agrupam corretamente complexos que possuem reatividade cruzada em um ramo e os que não possuem reatividade cruzada em outros ramos. A inclusão dos valores de ASA adiciona uma informação importante em relação à topografia do p:MHC, representando um grande ganho em relação a nossa antiga técnica. Uma representação gráfica da rede foi feita com a saída do HCA, para uma melhor visualização da rede. Esta técnica tem aplicação direta sobre nova estratégia de imunização. Por enquanto, pode ser utilizado para selecionar alvos que ativam ou não reatividade cruzada entre diferentes sorotipos de dengue e HCV.

Abstract

The immune system is still, in part, a black box. Advanced studies are starting helping us to better understand some events inside this black box, such as cross-reactivity. According to this phenomenon, a T cell receptor (TCR) can interact with many different peptide-loaded Major Histocompatibility Complexes (pMHCs) and a pMHC can interact with manifold TCRs. Cross-reactivity is an important issue, with consequences over vaccine development and immunotherapy. However, the complexity of this system has delayed progress in the field, and the development of a reliable method for T-cell cross-reactivity prediction remains a major goal for several immunologists. In the present work, we suggest the use of pMHC structural features as data input for a multivariate statistical analysis, in order to build an improved cross-reactivity prediction method. Our approach was an upgrade of your published techniques, including a subset with 28 HCV epitopes used to test your old implement plus 8 epitopes from the four Dengue Virus serotypes, which were previously published in an *in vitro* study. Using a Hierarchical Clustering Analysis (HCA) with data extracted from charges distribution and Accessible Surface Area (ASA) of selected residues over each pMHC structure, we succeed to create a network which groups pMHC complexes that have cross-reactive response. Complexes that were previously described as non-cross-reactive fall in independent branches of the HCA. The inclusion of ASA values added important information on pMHC topography, representing a great improvement in relation to our previously described approach. Moreover, the use of *Gephi* software to plot the output in a network simplified the interpretation of results. This technic has direct applications for vaccine design. For instance, could be used to select targets that do not trigger cross-reactive responses among dengue and HCV virus serotypes.



Capitulo 1

Introdução e Objetivos



Introdução Geral

Introdução histórica

A ciência passa por revoluções constantemente nas suas mais diferentes áreas. Tivemos na física a revolução da mecânica clássica desenvolvida por Isaac Newton e a teoria da relatividade geral de Albert Einstein que serviu de pilar para a física moderna. Já na química podemos citar Niels Bohr com seu modelo atômico e Ernest Rutherford com sua contribuição para a física nuclear. Na biologia, temos Charles Darwin com o livro "A Origem das Espécies" que revolucionou toda a forma de visualizar a variabilidade dinâmica dos organismos. A história da imunologia é um pouco mais recente, embora existam registros de tentativas muito antigas de proteger os indivíduos da ação nefasta dos patógenos. Temos os primórdios da imunização praticada pelos chineses através da inalação de cascas de feridas de indivíduos infectados pela varíola (variolação). Somente em 1796 que Edward Jenner, através de um experimento ousado envolvendo uma criança que foi imunizada com o material da pústula de ordenhadeiras contaminadas com um vírus similar ao da varíola (cowpox), conseguiu desenvolver a primeira vacina similar as que conhecemos hoje, ganhando assim o título de "pai da vacinação".

Porém o sistema imunológico permanece, ao menos parcialmente, sendo uma caixa preta para nós. Sabemos o "*input*" e o "*output*" clínico das patologias, mas o processo que culmina com a eliminação ou não da doença ainda é um grande mistério. Muitos dos métodos que utilizamos hoje em dia para a vacinação são basicamente os mesmos que Edward Jenner usou para comprovar sua metodologia, utilizando vírus atenuado, com um porém de que hoje em dia há um melhor entendimento do processo de imunização. Novas abordagens então se fazem necessárias, tanto para um melhor entendimento da imunização, como também para o desenvolvimento de vacinas mais eficazes e de ampla proteção. A importância deste tipo de conhecimento

é evidenciada pelo fato de que 2 dos 10 últimos Nobéis de Fisiologia e Medicina tenham sido destinados a estudos envolvendo o sistema imune.

Buscando contribuir para o entendimento dos mecanismos de reconhecimento imunogênico dessa caixa preta, venho por meio desta dissertação de mestrado apresentar resultados obtidos durante os dois últimos anos de pesquisas no laboratório de imunoinformática da UFRGS, possuindo este trabalho um enfoque no reconhecimento de peptídeos apresentados pela molécula de MHC de classe I e seu impacto na reatividade cruzada entre alvos peptídicos virais, utilizando como modelo epitopos de sorotipos do vírus HCV e da dengue. O desenvolvimento deste estudo foi totalmente realizado, utilizando-se ferramentas *in silico*, sobre dados experimentais (*in vitro* ou *in vivo*) previamente publicados.

Resposta Imune Inata e Adaptativa

Possuímos diversas formas de defesas contra patógenos. Para fins didáticos, separam-se estes tipos em duas categorias: resposta imune inata e resposta imune adaptativa. Podem-se citar como pertencentes à resposta imune inata as barreiras físicas, como a pele e as mucosas; químicas, como mudanças de pH e celulares, as quais abrangem fagócitos e células *Natural Killers*, por exemplo. Como característica deste tipo de resposta evidencia-se sua maior rapidez no mecanismo de ação, apesar de sua menor especificidade e sua incapacidade em gerar uma memória imunológica, sendo observados apenas alguns casos de memória imunológica em *NK-like* de insetos. A resposta imune adaptativa também apresenta uma subdivisão: resposta humoral e resposta celular. Ambas são respostas de desenvolvimento mais lento quando comparadas a inata, porém mais específicas e com posterior geração de células de memória. Esta é uma característica evolutivamente mais recente nos organismos, compartilhada entre os membros da superclasse Gnathostomata (Cooper and Alder (2006), ou seja, presente em mais de 99% de todos os vertebrados atuais.

A resposta humoral é definida primariamente pela produção de imunoglobulinas (anticorpos) pelos linfócitos B ((Mauri & Bosma, 2012). Esta é a rota mais estudada e utilizada para a concepção e produção de vacinas. Apesar de serem normalmente avaliadas separadamente, ambas as vias de respostas estão intimamente interligadas, sendo que raramente ocorre a ativação de apenas uma das vias no combate aos patógenos ((Shlomchik *et al.*, 2001). Apesar de estarmos consciente de que ambas as respostas são cruciais e intrincadas, o enfoque do nosso trabalho é sobre a resposta celular, mais especificamente sobre a resposta celular citotóxica, sendo que esta será melhor discutida nos tópicos a seguir.

A resposta celular citotóxica é caracterizada pela estimulação de linfócitos T CD8+ através da interação do receptor de células T (TCR) com a superfície formada pelo peptídeo ancorado ao complexo principal de histocompatibilidade (MHC) das células apresentadoras ((Rangarajan & Mariuzza, 2014). Esta estimulação também é dependente da sinalização de outras moléculas produzidas pela interação de outro conjunto de células, como a resultante da interação das células apresentadoras de antígenos profissionais e os linfócitos T CD4+ (linfócitos T auxiliares).

A molécula de MHC é a principal proteína apresentadora de antígenos, apresentando um alto grau de variabilidade genética entre os diferentes alelos presentes nos indivíduos. Os genes das diferentes classes de MHC estão contidos dentro do super-*locus* do MHC, uma região altamente polimórfica presente no cromossomo 6 de humanos (Mungall *et al.*, 2003). Existem as classes de MHC I, II e III. O MHC de classe I, presente em todas as células nucleadas, tem a função de apresentar partículas proteicas, chamadas de epitópos, oriundas do meio intracelular e que apresentam tamanho entre 8 a 12 aminoácidos (figura 1.A). Esta restrição de tamanho (normalmente 9 aminoácidos) deve-se a natureza restritiva da fenda deste tipo de molécula de MHC (fechada nas extremidades). As moléculas de MHC de classe I apresentam os epitópos para as células T CD8+, as quais podem gerar uma resposta citotóxica quando o epitopo que está sendo apresentado pertence a uma proteína não própria (Li *et al.*, 2013). Já o MHC de classe II, presente somente nas células apresentadoras de antígenos profissionais (APCs), apresenta epitópos oriundos do

meio extracelular, maiores do que os contidos no MHC-I (até 30 aminoácidos), em função da abertura das extremidades da fenda (Figura 1.B). O MHC de classe II apresenta o epitopo para o linfócito CD4+, o qual, quando estimulado, gera uma resposta T auxiliar, sinalizando para que outras células de defesa possam "auxiliar" na eliminação do patógeno (Li *et al.*, 2013). O MHC de classe III está envolvido na codificação de citocinas sinalizadoras, chaperonas e proteínas do sistema complemento (Qin & Mamotte, 2008).

Nas seções a seguir, será dado um enfoque maior na resposta citotóxica das células T CD8+ e sua interação com o MHC de classe I, sendo descrito abaixo a rota de processamento intracelular e apresentação de antígenos (epitopos de células T).

Rota de Processamento e Apresentação de Antígeno

Toda infecção viral e alguns tipos de infecções bacterianas, apresentam uma forma intracelular predominante (Li, 2011). Com isso, sistemas de defesa que agem na região extracelular como os anticorpos neutralizantes, são incapazes de debelar essa fase infecciosa. Para conseguir combater estes agentes infecciosos, é necessário um sistema de defesa intracelular especializado. A resposta citotóxica celular é capaz de localizar a ocorrência de uma infecção subjacente à membrana plasmática e eliminar a célula infectada. Para isto, várias etapas de preparação e apresentação dos antígenos endógenos precisam ocorrer.

A primeira etapa da via de processamento é a adição de uma molécula de ubiquitina nas proteínas intracelulares que são marcadas para degradação, processo conhecido como ubiquitinação. Este passo é necessário para o encaminhamento da proteína para o proteossomo (Frankes & Kleinschmidtfl, 1994).

O proteossomo 26S é formado por um núcleo central de 20S, que tem como função a reação proteolítica, e duas subunidades periféricas de 19S, as quais reconhecem a proteína ubiquitinada parcialmente desnaturada, clivando-a em

fragmentos regulares a partir da extremidade c-terminal da proteína (Wang & Maldonado, 2006). Estes pequenos fragmentos resultantes, ou peptídeos, do processo proteolítico deixam o complexo proteossômico e são capturados pelo transportador associado ao processamento de antígenos (TAP), formado por duas subunidades (TAP1 e TAP2) ancoradas no retículo endoplasmático (RE). Nestes dois primeiros passos da via de processamento de antígenos já ocorre uma pré-seleção dos peptídeos que serão apresentados, visto que tanto a proteólise quanto a translocação por TAP ocorre em fragmentos com combinações preferenciais de aminoácidos (motivos).

Dentro do RE, os peptídeos ainda sofrem algumas modificações, como o desbaste realizado pela ação de enzimas como as aminopeptidases, do retículo endoplasmático (ERAPs, do inglês **ER** aminopeptidases) (Endert *et al.*, 2011; Peaper & Cresswell, 2008). Após estas modificações finais de ajuste de tamanho, o peptídeo é posicionado na fenda do MHC de classe I. Este encaixe se faz com o auxílio de chaperonas específicas e outras moléculas acessórias (Hammer *et al.*, 2007). Para uma melhor visualização de todos os processos descritos consulte a figura 2.

MHC-I E TCR: protagonistas da resposta citotóxica

O MHC é formado por duas cadeias polipeptídicas, sendo uma cadeia alfa, definida como cadeia pesada e uma cadeia beta 2-microglobulina (figura 1.C), sendo esta responsável pela estabilização do complexo (Peaper & Cresswell, 2008). A cadeia alfa é subdividida em alfa 1, alfa 2 e alfa 3, sendo que alfa 1 e 2 formam uma espécie de fenda, local onde são ancorados os peptídeos. Estes peptídeos ou epitopos são ancorados em posições específicas, definidas como “pockets”, os quais, por sua natureza (tamanho, propriedades químicas, etc.) permitem ou limitam o encaixe de resíduos preferenciais. Após o acoplamento, esse complexo formado entre o peptídeo e o MHC (p:MHC) é exportado pelo complexo de golgi para ser apresentado na superfície extracelular, ficando acessível para a interação com o receptor de células T (TCR) de Linfócitos T CD8+ (Li *et al.*, 2013).

O TCR é um heterodímero da superfamília das imunoglobulinas formado por uma cadeia α e uma cadeia β , ou por uma cadeia γ e uma cadeia δ . Estas cadeias são compostas por dois domínios extracelulares, o variável e o constante. A região constante fica mais próxima da membrana da célula T e a região variável, a qual é composta por três sítios hipervariáveis (CDRs), sendo responsável por realizar o contato com o epítipo e com a cadeia alfa do MHC (Bhati *et al.*, 2014). A molécula CD8, também presente na superfície da célula T, auxilia na estabilização da interação TCR/p:MHC. Se a célula T reconhecer um epítipo como não próprio, esta irá desenvolver uma resposta citotóxica contra a célula que está realizando a apresentação, fazendo com que esta sofra apoptose (Diamond *et al.*, 2000).

Reatividade Cruzada

A reatividade cruzada é um fenômeno onde um determinado TCR pode reconhecer e gerar resposta contra mais de um epítipo distinto (oriundo de patógenos diferentes, por exemplo). Desta maneira, uma infecção por um patógeno seria capaz de gerar linfócitos T de memória capazes de produzir uma resposta contra um patógeno não relacionado posteriormente (Clute *et al.*, 2005).

Um exemplo deste fenômeno foi observado no experimento citado anteriormente, onde Edward Jenner percebeu que mulheres ordenhadeiras não desenvolviam a varíola. Na verdade, estas mulheres eram infectadas pela varíola bovina a qual causava sintomas mais brandos do que a varíola humana. O pesquisador observou que as mulheres que haviam tido um contato com o vírus da varíola bovina, sem saber na época que se tratava de um vírus, apresentavam uma forma mais leve da doença quando entravam em contato com a varíola humana, o que pode ser explicado pelo fenômeno da reatividade cruzada. A partir destas observações Edward Jenner utilizou o vírus da varíola bovina atenuado para criar a primeira vacina, de grande importância na erradicação da varíola.

Os primeiros trabalhos que estudaram o fenômeno da reatividade cruzada eram focados primariamente na busca por similaridades entre as sequências primárias de aminoácidos dos epítomos que estimulavam a reatividade cruzada. Entretanto, sabe-se que epítomos que compartilham menos de 50% de identidade entre seus resíduos também são capazes de gerar uma resposta cruzada (Clute *et al.*, 2005). Este fato salienta a importância da avaliação de outras características além da sequência linear do epítomo que contribuem para o desencadeamento da resposta imune, como por exemplo, as suas propriedades bioquímicas e estruturais que o compõe.

No âmbito de desenvolvimento de vacinas, é de grande interesse produzir uma vacina que confira proteção contra mais de um tipo de doença, ou pelo menos, vacinas que protejam contra um maior número de linhagens ou genótipos virais em patógenos altamente variáveis como o Vírus da Hepatite C (HCV) ou o Vírus da Imunodeficiência Humana (HIV). Para isto, é necessário encontrar os padrões de similaridade que definem a ocorrência da reatividade cruzada. É importante salientar que nem sempre este fenômeno é benéfico. Podemos citar, por exemplo, doenças autoimunes que são induzidas por mimetismo molecular, onde um antígeno externo faz com que células T gerem respostas cruzadas contra antígenos próprios (reatividade cruzada com antígenos próprios) (Regner & Article, 2001). Outra situação deste tipo é a que acontece com a dengue, onde a forma mais severa da doença (a dengue hemorrágica) é causada pela reinfeção por um subtipo diferente, originando uma resposta cruzada bem mais intensa que a primeira.

Porém, o organismo tem vantagem em possuir essas células T promíscuas, pois este tem um espaço limitado para o armazenamento de células de memória. Seria fisicamente impossível ter um TCR para cada epítomo que os linfócitos são passíveis de encontrar durante o curso de vida de um indivíduo. Essa promiscuidade, entretanto, não diminui a especificidade da resposta gerada por uma infecção (Regner & Article, 2001).

Vírus da Dengue

O vírus da dengue é um arbovírus pertencente à família Flaviviridae, gênero *Flavivirus*. É formado por uma única fita simples de RNA, a qual codifica três proteínas estruturais (C, M e E) e sete não estruturais (NS1, NS2a, NS2b, NS3, NS4a, NS4b e NS5) e possui quatro subtipos imunológicos: D1V, D2V, D3V e D4V (Duan *et al.*, 2012). A infecção por qualquer um dos subtipos confere proteção permanente para o mesmo sorotipo, porém quando infectado por outro subtipo, como citado anteriormente, ocorre a dengue hemorrágica.

A doença ocasionada pelo vírus pode ser assintomática ou possuir sintomas brandos a médios, entretanto podem surgir casos mais graves, levando inclusive ao óbito do indivíduo. A dengue é mais frequente em locais de clima tropical, ocorrendo geralmente em países em desenvolvimento ou de terceiro mundo, com uma taxa de infecção em torno de 100 milhões de casos por ano (Bhatt *et al.*, 2013).

A forma de contágio é através da picada do mosquito do gênero *Aedes*, sendo o mais comum a espécie *Aedes aegypti* (Bhatt *et al.*, 2013). A única tentativa de erradicação da doença é através do combate ao vetor, pois não existem vacinas eficazes disponíveis até agora.

De acordo com a literatura, a resposta citotóxica através das células T CD8+ é de fundamental importância no processo de combate à infecção pelo vírus da dengue (Duan *et al.*, 2012). Estas produzem interferon gama, o qual serve de sinalização para que ocorra o “clearance” da infecção (Diamond *et al.*, 2000; Duan *et al.*, 2012). Posto isso, nosso foco é gerar conhecimentos que possam ser aplicados na criação de vacinas baseadas em epitopos apresentados no contexto do MHC de classe I. O maior problema que deve ser evitado na criação da vacina da dengue é a ocorrência da reatividade cruzada entre seus subtipos, sendo desejável buscar alvos que confirmem proteção para cada subtipo e que não tenham reatividade (similaridades que possam desencadear resposta) entre si.

Vírus da Hepatite C

O vírus da hepatite C é relativamente pequeno, envelopado, possuindo uma fita simples de RNA. Pertence a família Flaviviridae. Seu RNA possui uma única ORF (fase aberta de leitura, do inglês *open reading frame*) contendo mais de 9600 nucleotídeos (KATO, 2000). Possui sete diferentes genótipos e, pela alta taxa de mutação destes vírus, estes genótipos possuem ainda subdivisões (Nakano *et al.*, 2012). O subtipo 1a e 1b respondem por mais de 60% dos casos de Hepatite C em todo o mundo.

A doença causada por este vírus atinge seres humanos e outros tipos de primatas. A infecção localiza-se principalmente no fígado. Em geral ela é assintomática, mas pode vir a se tornar uma infecção crônica e levar a fibrose ou até mesmo a uma cirrose hepática. O vírus em geral permanece no fígado, sendo que o tratamento convencional é realizado com interferon peguilado e ribavarina (Hiroishi *et al.*, 2010). Não existe ainda vacina aprovadas disponível para este vírus (Halliday *et al.*, 2011).

Mesmo considerando todas as limitações impostas por um vírus com tamanha variabilidade e diferentes estratégias de escape (Yu & Chiang, 2010) é muito importante que se pense no desenvolvimento racional de vacinas anti-HCV que cubram o maior número de genótipos possíveis. Pelo menos para aqueles genótipos que ocorrem dentro da região da população imunizada. Dessa maneira, estratégias como a busca de padrões compartilhados entre peptídeos de HCV contidos em complexos p:MHC pode permitir que sejam geradas estratégias vacinais que gerem resposta contra uma grande parte dos genótipos predominantes e que funcionem no contexto de alelos de MHC altamente prevalentes.

Imunoinformática aplicada ao desenvolvimento de vacinas

Graças aos avanços nas áreas da tecnologia, processamento e no avanço dos ambientes de simulação, as abordagens computacionais surgem como ferramentas poderosas em estudos biológicos. Assim, a imunoinformática (área da bioinformática que compreende preditores e abordagens *in silico* relacionadas à imunologia) torna-se

uma opção cada vez mais confiável e viável financeiramente em estudos envolvendo novas abordagens para o combate e tratamento de doenças infecciosas.

Utilizando conceitos da vacinologia reversa, é possível escolher alvos imunogênicos *in silico* para vacinas, sendo que este processo possui um baixo custo de operação e uma eficiência crescente em função da disponibilização de novas informações. Grandes bancos de dados estão disponíveis na rede mundial de computadores, os quais se tornam grandes aliados para o desenvolvimento de novos métodos para predição de novos alvos.

Os bancos estruturais de proteínas, por exemplo, disponibilizam cristais de pMHCs (código 1HLA no *Protein Data Bank* por ex.) e TCRs (código 2V2W no *Protein Data Bank* por ex.) que podem ser utilizados para analisar os elementos que interagem na formação de um complexo p:MHC, e que resíduos deste complexo são importantes na interação com o receptor de célula T. Porém, o custo de uma cristalização e sua taxa de sucesso é uma grande barreira para a geração de cristais para o número de combinações de epitopos e alelos de MHC desejável para uma compreensão mais completa da estimulação da imunogenicidade. Então, alternativamente, temos abordagens computacionais estruturais para suplantar este tipo de limitação técnica, como o a ancoragem molecular. Esta ferramenta é utilizada na exploração de sítios e modos de ligação entre ligantes e seus receptores. No nosso trabalho, o objetivo final é realizar a ancoragem do epitopo no MHC de classe I. A ferramenta calcula as afinidades entre diferentes modos de ligação do epitopo e o receptor (MHC), de onde pode ser escolhida a melhor posição de ancoramento. Após esse passo, o complexo gerado sofre uma rodada de minimização de energia para acomodação das cadeias laterais das estruturas formadoras do complexo, as quais são novamente separadas e reancoradas, gerando assim uma estrutura 3D (Antunes *et al.*, 2010a).

Após o processo de modelagem dos complexos, pode-se utilizar uma abordagem complementar que forneça informações adicionais sobre as estruturas geradas. Uma delas é o cálculo de potencial eletrostático. Com ele podemos inferir a carga residual em cada uma das diferentes regiões da superfície da molécula. Esta

análise é de fundamental importância em diversos estudos, como nos desenhos de ligantes (fármacos), por exemplo. No nosso trabalho é um dos parâmetros utilizados para identificar padrões que estimulam a reatividade cruzada entre epitopos distintos (Antunes *et al.*, 2011; Li *et al.*, 2012b). Para o cálculo de carga residual em moléculas, utilizou-se a equação de Poisson-Boltzmann. Também objetivando encontrar os "fingerprints" imunogênicos, nós utilizamos os dados de superfície de acessibilidade ao solvente (SAS) dos resíduos que normalmente contactam o receptor de célula T. Com essas duas variáveis, topográfica e físico-química, procura-se a separação entre epitopos com reatividade cruzada e os que não induzem este fenômeno entre si.

Para realizar o agrupamento dos dois tipos de epitopos, utilizamos o método estatístico de clusterização hierárquica (HCA). Este método é capaz de agrupar em ramos os alvos que compartilham similaridade nas características de entradas, através de medidas de dissimilaridade que são obtidas pelas distâncias observadas entre cada par observado. A saída é apresentada através de um dendograma, mostrando as ramificações formadas.

Considerando o acima exposto, apresentamos os objetivos esperados para este trabalho.

Objetivos

Objetivo Geral:

- O objetivo deste trabalho é criar um preditor de redes de reatividade cruzada baseado em parâmetros estruturais e de potencial eletrostático, utilizando como controle um conjunto de epitopos publicados na literatura com reatividade cruzada pré-definida. Adicionalmente pretendemos gerar um modo de visualização destas redes de similaridade, para que futuramente esta ferramenta possa ser hospedada no sítio do “*Crosstope Data Bank* (www.crosstope.com.br)”.

Objetivos Específicos:

- Fazer um mapeamento dos pontos de contatos entre o TCR e o p:MHC (alelo HLA-A:0201) utilizando cristais disponíveis em bancos de dados para estabelecer os locais da superfície da molécula que serão comparados.
- Calcular os dados de ASA para os resíduos pré-definidos na análise anterior de todos os epitopos incluídos no estudo;
- Calcular o potencial eletrostático dos pontos pré-selecionados da superfície de todos os complexos p:HLA-A:0201 incluídos no estudo;
- Combinar os dados de ASA e distribuição de cargas para utilizar como entrada na análise de HCA para posterior agrupamento dos epitopos com características similares;
- Observar se os agrupamentos corroboram os dados experimentais de reatividade cruzada extraídos da literatura para os epitopos incluídos na análise;
- Utilizar uma ferramenta visualizadora de redes proteicas e gênicas para representação das redes de reatividade cruzada como protótipo de nossa ferramenta final;



Capitulo II

*Improved structural method for T-cell cross-reactivity
prediction*



Improved structural method for T-cell cross-reactivity prediction

Marcus FA Mendes^{12*}, Dinler A Antunes^{12*}, Maurício M Rigo¹², Marialva Sinigaglia¹², Gustavo F Vieira^{12§}

¹NBLI – Núcleo de Bioinformática do Laboratório de Imunogenética. Departamento de Genética, Universidade Federal do Rio Grande do Sul. Av. Bento Gonçalves 9500, Building 43323, room 225.

²Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Universidade Federal do Rio Grande do Sul (UFRGS), Rio Grande do Sul, Porto Alegre, Brazil.

*These authors contributed equally to this work

§Corresponding author

Email addresses:

MFAM: cla_atm_milo@hotmail.com

DAA: dinler@gmail.com

MMR: mauriciomr1985@gmail.com

MS: msinigaglia@gmail.com

GFV: gusfioravanti@yahoo.com.br

Abstract

Background

Cytotoxic T-Lymphocytes (CTLs) are the key players of adaptive cellular immunity, being able to identify and eliminate infected cells through the interaction with peptide-loaded Major Histocompatibility Complexes class I (pMHC-I). Despite of the high specificity of this interaction, a given lymphocyte is actually able to recognize more than just one pMHC-I complex, a phenomena referred as cross-reactivity. In the present work, we describe the use of pMHC-I structural features as input for multivariate statistical methods, in order to perform standardized structure-based predictions of cross-reactivity among viral epitopes.

Results

Our improved approach was able to successfully identify cross-reactive targets among 28 naturally occurring HCV variants and among 8 epitopes from the four Dengue Virus serotypes. In both cases, our results were supported by multiscale bootstrap resampling and by data from previously published *in vitro* experiments.

Conclusions

The combined use of data from charges and Accessible Surface Area (ASA) of selected residues over the pMHC-I surface provided a powerful way of assessing the structural features involved in triggering cross-reactive responses. Moreover, the use of an R package (pvclust) for assessing the uncertainty in the hierarchical cluster analysis provided an statistical support for interpretation of results. Taken together, these methods can be applied for vaccine design, both for the selection of candidates capable of inducing immunity against different targets, and to identify epitopes that could trigger undesired immunological responses.

Keywords

Cross-reactivity, pMHC-I, HCA, ASA, pvclust, vaccine development.

Background

Cellular immunity is one of the two main branches of the adaptive immunologic response, focused on specific functions of the Cytotoxic T-Lymphocytes (CTLs). Although both cellular and humoral immunity are desired for an ideal and longstanding immunization, CTL response plays a central role in which regards to antiviral immunity (Brehm *et al.*, 2004). After infecting a host cell, the virus will use the host molecular machinery to replicate its genome and produce new virions. In addition to all the mechanisms that allow virus escape from circulating neutralizing antibodies, during its intracellular replication cycle the virus is virtually hidden from the action of humoral immunity. However, some viral proteins will unavoidably be marked to enter the endogenous antigen presentation pathway. Through this route, virus-derived peptides will be presented at the cell-surface in the context of Major Histocompatibility Complex (MHC) class I molecules, forming stable peptide:MHC-I (pMHC-I) complexes. Each CTL produced by the host has one specific T-Cell Receptor (TCR), which is able to recognize pMHC-I complexes presenting nonself peptides. Therefore, through the interaction between pMHC-I complexes and TCRs, CTLs are able to identify and eliminate infected cells.

The TCR-pMHC-I interaction is highly specific, which allows the development of memory T-cells that will be once again triggered in future challenges with the same target. However, a given lymphocyte is able to recognize more than just one pMHC-I complex. This capacity of a CTL to recognize non-related peptides derived from the same virus, or even peptides from heterologous viruses, was defined as cross-reactivity (Vieira & Chies, 2005). As expected, cross-reactivity has direct implications over vaccine development, autoimmunity and heterologous immunity, a process by which the immunization with one pathogen confers protection against another (Cornberg *et al.*, 2010; Selin *et al.*, 1994; Welsh & Fujinami, 2007; Welsh & Selin, 2002). Understanding of the molecular features driving these cross-reactivities became a major goal for several immunologists, but the system's complexity has delayed progress in the field. Wedemeyer *et al.* 2001 (Wedemeyer *et al.*, 2001) has proposed that cross-recognition of two heterologous epitopes could be triggered by the high amino acid sequence similarity between them. Similarity in terms of biochemical properties was also proposed as being the key for cross-recognition (Vieira & Chies, 2005), and was even applied with some success to predict cross-reactivity (Frankild *et al.*, 2008; Moise *et al.*, 2013a).

However, structural studies have shown that even epitopes with low sequence and biochemical similarity might present quite identical pMHC-I surfaces (Antunes *et al.*, 2011; Sandalova *et al.*, 2005), defending that this structural similarity should account for the cross-stimulation of a given T-cell population.

Structural analysis of pMHC-I complexes can provide a level of information much closer to that presented *in vivo* to interaction with the TCR. On the other hand, structural approaches are frequently limited by the number of pMHC-I structures already produced by experimental methods, such as X-Ray crystallography and NMR (Nuclear Magnetic Resonance). Our group has used structural bioinformatics tools to build *in silico* models of pMHC-I complexes that were not yet determined by experimental methods. This approach, referred as *D1-EM-D2* (*Docking 1 - Energy Minimization - Docking 2*), was previously validated through the successful reproduction of several crystal structures (Antunes *et al.*, 2010b; Sinigaglia *et al.*, 2013a) and has been used to provide novel complexes for the CrossTope Data Bank for Cross-Reactivity Assessment (Sinigaglia *et al.*, 2013a). Our group has also combined this approach with the use of multivariate statistical methods in order to make structural-based cross-reactivity predictions (Antunes *et al.*, 2011). In this context, we present here an improved and standardized structural-based method for T-cell cross-reactivity prediction of HLA-A*02:01-restricted epitopes. The predictive capacity of our method was enhanced by the inclusion of new features, and our results with the analysis of viral epitopes (from Hepatitis C Virus and Dengue Virus) support its use as an important auxiliary tool for vaccine development.

Results and Discussion

Identification of conserved contacts among TCR-pMHC-I crystal structures

The human HLA-A*02:01 is largely studied for being the most frequent MHC allele in human populations (<http://www.allelefreqencies.net/>) (Fernandez-Vina *et al.*, 1992). For this reason, the protein encoded by this specific allele (called allotype) also presents the larger number of crystal structures available at the Protein Data Bank (PDB). Aiming to identify the residues involved in the recognition of this allotype by different TCRs, we performed an extensive search for all available crystal structures of TCR-pMHC-I complexes, restricted to

HLA-A*02:01. This search returned 29 complexes ([Additional file 1](#)), presenting 15 different TCRs and 18 different epitopes. Despite this variability, 5 epitope positions and 4 MHC residues were consistently indicated as involved with TCR interactions, being presented in more than 85% of these complexes. Several residues over the pMHC-I surface might participate of the interaction with the TCR, influencing the specific level of T-cell stimulation that will be triggered by each pMHC-I. However, we here postulate that changes in these nine conserved contacts might have greater impact over the T-cell recognition, therefore preventing cross-reactivities.

An image-based strategy for pMHC-I clustering

In a previous study, our group used images of the electrostatic potential distribution over the pMHC-I surface to predict the cross-reactivity pattern among 28 naturally occurring HCV variants, in the context of HLA-A*02:01 (Antunes *et al.*, 2011). In that study, one specific region over the pMHC-I surface was defined, based on the observation of the main spots of variation among the 28 complexes analysed. From this specific region, were extracted values of Mean and Standard deviation for each RGB (Red, Green, Blue) component, which were used as input for a Principal Component Analysis (PCA). Based only on the extracted information from the pMHC-I structures, the PCA was able to predict the same clusters of cross-reactivity observed *in vitro* (Antunes *et al.*, 2011; Fyttili *et al.*, 2008). Despite of the success of this approach, the same parameters could not be applied to other subsets, once different regions of the pMHC-I surface might have greater influence over the TCR recognition. In the same study, seven variable regions over the pMHC-surface were defined and used to perform a structure-based virtual screening of putative cross-reactive targets among 55 non-related pMHC-I complexes (Antunes *et al.*, 2011). This experiment predicted a cross-reactivity between two unrelated epitopes (one from HCV, other from EBV), which shared no amino acids in their sequences, and this prediction was later confirmed with *in vitro* experiments (Zhang S, personal communication).

This image-based clustering of pMHC-I complexes is an innovative approach that has been shown to be a fast and efficient way to predict cross-reactivity using structural information, being able to identify cross-reactive targets even between epitopes which shared no amino acids in sequence. It might be argued that a better structure-based clustering would use the actual charge information for each atom, as well as more accurate topographic descriptors. However, this “full-structure-based” comparison of pMHC-I complexes remains an

evasive and expensive computational challenge. In order to run a structure-based virtual screening such as that previously described by our group, it is important to compare the pMHC-I complexes presenting epitopes in different conformations (the equivalent atom might be in a completely different position), with different sequences (and therefore different number of atoms in each complex), and considering the charge distribution over the “TCR-interacting surface” as a whole (a “sum” of peptide and MHC), instead of considering only the specific charges of epitope atoms.

Definition of key areas for TCR-pMHC-I contact

In the present work, we aimed to improve our structure-based prediction method and to provide a defined set of “gates” that could indicate the key interactions involved in cross-reactive responses, which could be applied to any subset of epitopes restricted to HLA-A*02:01. Considering the nine key positions identified in crystal structures and described above, we defined a group of seven regions over the pMHC-I surface (**Figure 1**). These regions, or “gates”, were defined considering the specific contribution of each one of these residues to the combined surface of the pMHC-I. Three regions were defined covering the epitope surface. The contribution of epitope positions p4 and p5 were collected by two independent gates (G1 and G2). In the case of positions p6, p7 and p8, only one gate was defined, centered over p7 (G3). This was decided considering that given the conformation of the epitope backbone inside the cleft of HLA-A*02:01, p7 is much more exposed to the contact with the TCR, while p6 and p8 have lower contribution to the pMHC-I surface. Other four gates were defined over selected MHC-I residues (G4, G5, G6 and G7). These seven key regions are placed within the area previously described by other groups as the “TCR footprint” for this allotype (Gras *et al.*, 2012; Gras *et al.*, 2009; Rudolph *et al.*, 2006) and, therefore, will be primarily responsible for triggering cross-reactive responses.

Inclusion of ASA values

Other limitation of our previously described approach was its entire dependence on the electrostatic potential information (Antunes *et al.*, 2011). As discussed by the authors, there are experimental evidences suggesting that charge similarity is more important than subtle topography differences between the cross-reactive complexes (Jorgensen *et al.*, 1992; Kessels *et al.*, 2004). However, pMHC-I complexes are 3D structures and topography variation certainly has some influence over the TCR recognition. The Accessible Surface Area (ASA) of a residue can provide a quantitative measure of how exposed or buried its side chain is, which

will have impact over the pMHC-I topography. ASA values of the epitope residues, for instance, were previously related to immunogenicity (Meijers *et al.*, 2005) and were also able to identify non-cross-reactive complexes (Antunes *et al.*, 2010b). Since direct 3D topography comparison remains an expensive computational challenge and ASA values can be easily calculated, we decided to include ASA values together with electrostatic potential information in order to improve our prediction method. When analyzing a subset with different epitopes in the context of the same allotype, the ligand variation will affect not only the charges and the ASA values of the epitope itself, but also of the MHC residues that directly interact with the peptide. For that reason, in addition to the ASA values for the nine epitope residues, we also included in our approach ASA values from 28 selected MHC-I residues ([Additional file 2](#)).

Use of multiscale bootstrap resampling

As previously described, our prediction method was based on the use of pMHC-I structural features as input for multivariate statistical methods (Antunes *et al.*, 2011). Originally, only information on electrostatic potential was used to define the clusters of putative cross-reactive complexes, being now combined with additional information on ASA values. Another remarkable improvement of our approach relates to the use of an R package (*pvclust*) for assessing the uncertainty of the Hierarchical Cluster Analysis (HCA) (Suzuki & Shimodaira, 2006). This package provides both Bootstrap Probability (BP) and Approximately Unbiased (AU) *p*-value, which is computed by multiscale bootstrap resampling and has been shown to be less biased than other methods in typical cases of phylogenetic tree selection (Shimodaira, 2002). This improvement adds a statistical validation to the dendrogram, enriching the discussion on the results, and avoiding unsubstantiated conclusions.

Method validation with a previously studied subset

To confirm its usability, our approach must be also able to identify the cross-reactive complexes in the previously described HCV subset (Antunes *et al.*, 2011; Fyttili *et al.*, 2008) ([Additional file 3](#)). These 28 variants, covering all six HCV genotypes, were tested *in vitro* against the same T-cell population, which was obtained from an individual vaccinated with the wild-type epitope HCV-NS3₁₀₇₃ (CINGVCWTV). The level of IFN-gamma production stimulated against a highly cross-reactive variant from genotype 1 (G1-01: CVNGVCWTV) was defined as a reference of high response, which was used to classify the other variants into high, intermediate or low cross-reactive complexes ([Additional file 3](#)).

An HCA based in our improved approach was able to divide the complexes into three main clusters (**Additional file 4**). A threshold was defined with the *pvrct* function to highlight these groups ($\alpha=0.95$), which are corroborated by AU *p*-values with low standard error (**Additional file 5**). The variant G3-18 (from Genotype 3) fell in a completely independent branch, forming a cluster on their one. This result is in agreement with our previous analysis and with the experimental data, since the G3-18 was the only one among the 28 complexes that presented no detectable response *in vitro* (Fyttili *et al.*, 2008). All the high cross-reactive complexes fell in the same main cluster. Of note, in the *in vitro* assay, the complexes with the higher levels within the cross-reactive complexes were G1-02, G1-07, G5-22, G6-25 and G6-27 (Antunes *et al.*, 2011; Fyttili *et al.*, 2008). In our HCA, four of these complexes fell in the same sub-cluster of the reference variant G1-01, although the AU/BP values does not strongly support the existence of this independent subgroup ($AU=77$). Most of the remaining high responders fell in a close related sub-cluster. The high responder variant G6-26 and the intermediate responder G1-05 fell in separate branches ($AU=99$), but still within the main cluster of the cross-reactive complexes ($AU=97$). It's also important to observe, that the original analysis of these complexes presented the intermediate responder G1-05 as the closest related complex to the reference complex G1-01 (Antunes *et al.*, 2011). The authors discussed this unexpected result defending that despite of the surface similarity other issues might account for the lower response presented by G1-05, such as binding affinity of the epitope to the MHC and complex stability. Our improved approach was able to identify neglected structural differences between G1-01 and G1-05, and correctly place G1-05 outside of the sub-clusters of high responders.

All low cross-reactive complexes fell in an independent main cluster ($AU=97$). The low responders from genotype 1, G1-03 and G1-04, fell correctly into this main low responders cluster, as well as the intermediate responders G1-06 and G1-08. The complex G1-06 was also placed within the low responders in the original analysis. Of note, a trend to the separation of the variants according to their genotypes is also observed, since we have a sub-cluster only with G3 complexes ($AU=77$) and a sub-cluster with majority of G2 complexes ($AU=99$). Our HCA results also provide other suggestions, such as that G1-08 is closer related to G2-11 and G3-20 ($AU=96$) than to G1-06. However, we have no experimental background to support this level of speculation. Observe that the *in vitro* assay performed with these 28 HCV variants was to verify the cross-reactivity against the wild-type HCV-NS3₁₀₇₃. Cross-reactivity also depends on

the T-cell population involved, so in order to evaluate the cross-reactivity against G1-08, an assay with a G1-08-specific T-cell population would be needed.

Cross-reactivity prediction among dengue virus serotypes

Dengue virus (DV) represents a major challenge for vaccine development (Halstead, 2013). Despite effective immunization against one serotype is easy to achieve, and protective T-cell response is observed, challenge of an immunized individual with an heterologous serotype often leads to severe symptoms, such as dengue haemorrhagic fever and dengue shock syndrome (DHF/DSS). In this context, cross-reactive T-cells are believed to mediate the immunopathogenesis of DHF/DSS during secondary heterologous challenge (Duan *et al.*, 2012). Therefore, the identification of non-cross-reactive immunogenic targets, specific for each DV serotype, is one way to develop a combined tetravalent vaccine. In a recent publication, Zhi-Liang Duan and colleagues identified HLA-A*02:01-restricted peptides from the four DV serotypes, and examined their immunogenicity and cross-reactivity (Duan *et al.*, 2012). From their data, we extract the epitope sequence of two groups of targets, being one identified as (i) cross-reactive variants and other as (ii) non-cross-reactive variants (**Additional file 6**).

We performed a new prediction with the combined data from both subsets (HCV and DV), totalling 36 pMHC-I complexes. The HCV and DV variants fell in independent main clusters, and the same threshold ($\alpha=0,95$) was able to identify cross-reactive and non-cross-reactive complexes within these groups (**Figure 2**). A plot of the standard error is provided in **Additional file 7**. All four NS4b variants fell in the same cluster ($AU=100$). This was expected, since cross-reactive response was indeed observed for these four variants. The same level of clustering was not observed for the NS4a variants ($AU=83$), a group that did not present cross-reactivity in the study of Zhi-Liang Duan and colleagues. Together with the AU/BP values, the dendrogram y-axis (Height) can also provide information on the dissimilarity among the complexes. A higher distance among the edges of the NS4a cluster can be observed when compared with the distance among the edges of the other main clusters.

The variants D1V-NS4a₁₄₀ and D4V-NS4a₁₄₀ fell in independent branches, while the other two (D2V-NS4a₁₄₀ and D3V-NS4a₁₄₀) fell in the same cluster ($AU=95$). Our HCA, therefore, indicates a possible cross-reactivity between D2V-NS4a₁₄₀ and D3V-NS4a₁₄₀, which could be understood as a false positive result. However, is important to highlight that cross-reactivity is also dependent on the specific T-cell population involved, and normally produces responses

with lower intensity when compared to the challenge with the cognate peptide. Of note, the D2V-NS4a₁₄₀ presented really low levels of response even upon challenge with the cognate epitope (**Additional file 6**) (Duan *et al.*, 2012). Despite of a possible structural similarity, a cross-reactive response would be probably undetectable with this T cell population. Moreover, our approach relies exclusively on structural features of the pMHC-I surface, such as charges distribution and ASA values, being capable of identifying the closer related complexes. However, other features such as MHC binding and pMHC-I stability might also interfere with the T cell stimulation process, preventing the occurrence of cross-reactive responses.

Finally, the combined HCA (HCV and DV) was able to reproduce the same results observed in the independent HCV analysis. This combined approach corroborates the consistency of our method, even with a greater number of complexes, suggesting its possible use in a larger scale as a virtual screening method. In this sense, we also explored an alternative way to present our HCA results. Instead of a dendrogram, this data can be used as input for relational networks, which can provide more intuitive information about the cross-reactive-networks studied (**Additional file 8**).

Applicability to vaccine development

Although in this case our analysis of the HCV subset was performed only as a proof of concept, this approach could be applied for real HCV vaccine development. Several immunogenic targets were identified and successful immunization can be achieved, but HCV diversity remains a major challenge. The identification of targets capable of triggering cross-genotype responses could drive the efforts to develop a new generation of vaccines, protective against all genotypes.

On the other hand, cross-reactivity is an issue to be avoided in a DV vaccine development, since it is involved in the immunopathogenesis of DHF/DSS. Once again, our improved structural based prediction could be applied as a virtual screening method to identify possible cross-reactivities that are yet unknown, and must be tested before the use of predicted targets in an anti-DV vaccine.

Traditional methods of vaccine development provided some successful results, but have been unable to overcome some of the major challenges for global health, such as HIV and HCV. In that context, a new generation of rationalized vaccines are starting to be planed, and bioinformatics tools are playing a major role in this process (Donati & Rappuoli; Dormitzer *et*

al.). Combined *in silico* approaches can save time and money, identifying the candidates more likely to stimulate the desired immune response, which can then be tested with *in vitro* and *in vivo* experiments to confirm its safety and efficacy for the use in a new vaccine.

Conclusions

The CD8+ T-cell cross-reactivity is a complex phenomenon triggered by the structural similarity between two different pMHC-I complexes that are recognized by the same TCR. Despite the enormous variability of TCRs and epitopes involved in these interactions, there are few conserved contacts that are shared by all TCR-pMHC-I crystal structures available, providing a map of the most important regions over the pMHC-I surface. Moreover, cross-reactivity between two pMHC-I complexes can be predicted based on the electrostatic potential over these selected regions. Our innovative approach showed that use of ASA values can improve this prediction, adding valuable information on the topography of these complexes. Finally, the use of a R package to assess the uncertainty of the hierarchical clustering provided a statistical validation of the results. Taken together, these findings provide an improved structural method for cross-reactivity prediction, with direct application over vaccine development.

Methods

Identification of conserved contacts between TCRs and pMHCs

An extensive search for all available crystal structures of TCR-pMHC-I complexes restricted to HLA-A*02:01 was performed in Protein Data Bank (Berman *et al.*, 2000) and IMGT/3Dstructure-DB (Kaas *et al.*, 2004). Curated and calculated contacts between TCR and pMHC, for each complex, were obtained from IEDB-3D (Ponomarenko *et al.*, 2011). Although available at IMGT/3Dstructure-DB, complexes 1QSE, 3QDG, 3QEQ, 3QDM, 3QDJ and 3UTS were not yet included in IEDB-3D and, therefore, were not included in our analysis. Information on included complexes is provided in [Additional file 1](#).

Construction of pMHC-I complexes

All our structural analysis were performed with pMHC-I complexes obtained through the previously described *D1-EM-D2* approach (Antunes *et al.*, 2010b). Briefly, only the FASTA

sequence of the epitopes was recovered from the reference studies (Duan *et al.*, 2012; Fytli *et al.*, 2008) and used as input to produce 3D structures of these epitopes, with PyMOL scripts. A “donor” structure of an empty HLA-A*02:01 was obtained by removing the epitope from a reference PDB structure (PDB code 2V2W). The new pMHC-I structure, harboring the epitope of interest in the context of HLA-A*02:01, was then obtained by a combined sequence of Molecular Docking and Energy Minimization steps. These steps were performed with Autodock Vina (Trott *et al.*, 2010) and GROMACS 4.5.1 package (Pronk *et al.*, 2013), respectively. The accuracy and reliability of this *D1-EM-D2* approach was tested in previous studies (Antunes *et al.*, 2010b; Sinigaglia *et al.*, 2013a).

Calculations over the pMHC-I complexes

Electrostatic potential of each pMHC-I structure was calculated with Delphi (Li *et al.*, 2012a), with custom parameters (e.g.: *indi=1.0*, *exdi=80.0*, *prbrad=1.4*, *salt=0.2*). Accessible Surface Area (ASA) of the selected residues from each pMHC-I complex was calculated with NACCESS V2.1.1 using default parameters (<http://www.bioinf.manchester.ac.uk/naccess/>).

Images acquisition and data extraction

Images of the electrostatic potential distribution over the “TCR-interacting surface” of each pMHC-I were obtained with UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) (Pettersen *et al.*, 2004). The “Electrostatic surfacing coloring” option of Chimera is used to import and visualize the electrostatic potential calculated with Delphi, using a range from -3 kiloteslas to +3 kiloteslas. Selected regions over these images were defined, and color histograms (RGB) of these areas were obtained with ImageJ 1.43u software (National Institute of Health, USA, <http://rsb.info.nih.gov/ij/>). In total, 42 values were obtained from the seven histograms of each image, such as color mean and standard deviation for each RGB component. Figures included in the article were edited with Adobe Photoshop CS2 v.9.0. program (Adobe, San Jose, CA).

Clustering Analysis

Hierarchical cluster analysis was performed with R package Pvcust (Suzuki & Shimodaira, 2006), assisted by the RStudio IDE v0.97 (<http://www.rstudio.com/>). The “average” linkage method was used with “correlation” distance, and the number of bootstrap replications was set to 10000. Results were plotted as dendrograms with Bootstrap

Probabilities (BP) and Approximately Unbiased (AU) p -values. Main clusters were identified with *pirect* ($\alpha=0,95$) and standard errors for AU p -values were obtained with *seplot*. Relation networks were plotted with the open-source platform Gephi (<https://gephi.org>).

List of Abbreviations used

ASA - Accessible Surface Area.

AU - Approximately Unbiased p -value.

BP – Bootstrap Probability value.

CTLs - Cytotoxic T-Lymphocytes.

DHF/DSS - Dengue Haemorrhagic Fever and Dengue Shock Syndrome.

DV – Dengue Virus.

HCA - Hierarchical Cluster Analysis.

HCV – Hepatitis C Virus.

HIV - Human Immunodeficiency Virus

MHC - Major Histocompatibility Complex.

PCA - Principal Component Analysis.

PDB - Protein Data Bank.

pMHC-I - peptide:MHC-I complexes.

RGB – Red, Green, Blue.

TCR - T-Cell Receptor.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

MFAM performed most of the experiments and reviewed the manuscript. DAA participated in all steps of the project and drafted the manuscript. MMR collaborated with the improvement of the algorithm used to build the structures, discussed the data and reviewed the manuscript. MS suggested and supervised the analysis of the TCR-pMHC crystal structures. GFV conceived of the study,

participated in its design and coordination and have given final approval of the version to be published. All authors read and approved the final manuscript.

Acknowledgements

We thank the students Jader Peres da Silva, Ártur Krumberg Schüller and Marina Roberta Scheid for the collaboration in some steps of this work.

References

1. Brehm MA, Selin LK, Welsh RM: **CD8 T cell responses to viral infections in sequence.** *Cellular microbiology* 2004, **6**:411-421.
2. Vieira GF, Chies JAB: **Immunodominant viral peptides as determinants of cross-reactivity in the immune system--Can we develop wide spectrum viral vaccines?** *Medical hypotheses* 2005, **65**:873-879.
3. Welsh RM, Selin LK: **No one is naive: the significance of heterologous T-cell immunity.** *Nat Rev Immunol* 2002, **2**:417-426.
4. Welsh RM, Fujinami RS: **Pathogenic epitopes, heterologous immunity and vaccine design.** *Nature reviews Microbiology* 2007, **5**:555-563.
5. Cornberg M, Clute SC, Watkin LB, Saccoccio FM, Kim S-k, Naumov YN, Brehm MA, Aslan N, Welsh RM, Selin LK: **CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections.** *Journal of immunology (Baltimore, Md : 1950)* 2010, **184**:2825-2838.
6. Selin LK, Nahill SR, Welsh RM: **Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses.** *The Journal of experimental medicine* 1994, **179**:1933-1943.
7. Wedemeyer H, Mizukoshi E, Davis AR, Bennink JR, Rehmann B: **Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells.** *Journal of virology* 2001, **75**:11392-11400.
8. Frankild S, de Boer RJ, Lund O, Nielsen M, Kesmir C: **Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire.** *PLoS ONE* 2008, **3**:0.
9. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, Verberkmoes NC, Sztein MB, Losikoff PT, Martin WD *et al*: **The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix.** *Hum Vaccin Immunother* 2013, **9**(7).
10. Sandalova T, Michaelsson J, Harris RA, Odeberg J, Schneider G, Karre K, Achour A, Michaëlsson J, Kärre K: **A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity.** *The Journal of biological chemistry* 2005, **280**:27069-27075.

11. Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, Chies JAB, Vieira GF: **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* 2011, **48**:1461-1467.
12. Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M, Chies JAB: **Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment.** *PloS one* 2010, **5**:e10353.
13. Sinigaglia M, Antunes DA, Rigo MM, Chies JA, Vieira GF: **CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes.** *Database (Oxford)* 2013, **2013**:bat002.
14. Fernandez-Vina MA, Falco M, Sun Y, Stastny P: **DNA typing for HLA class I alleles: I. Subsets of HLA-A2 and of -A28.** *Hum Immunol* 1992, **33**(3):163-173.
15. Fytili P, Dalekos GN, Schlaphoff V, Suneetha PV, Sarrazin C, Zauner W, Zachou K, Berg T, Manns MP, Klade CS *et al*: **Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073.** *Vaccine* 2008, **26**:3818-3826.
16. Rudolph MG, Stanfield RL, Wilson Ia: **How TCRs bind MHCs, peptides, and coreceptors.** *Annual review of immunology* 2006, **24**:419-466.
17. Gras S, Burrows SR, Turner SJ, Sewell AK, McCluskey J, Rossjohn J: **A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned.** *Immunol Rev* 2012, **250**(1):61-81.
18. Gras S, Saulquin X, Reiser J-B, Debeaupuis E, Echasserieau K, Kissenpfennig A, Legoux F, Chouquet A, Le Gorrec M, Machillot P *et al*: **Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope.** *Journal of immunology (Baltimore, Md : 1950)* 2009, **183**:430-437.
19. Jorgensen JL, Esser U, Fazekas de St Groth B, Reay PA, Davis MM: **Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics.** *Nature* 1992, **355**:224-230.
20. Kessels HWHG, de Visser KE, Tirion FH, Coccoris M, Kruisbeek AM, Schumacher TNM: **The impact of self-tolerance on the polyclonal CD8+ T cell repertoire.** *Journal of immunology (Baltimore, Md : 1950)* 2004, **172**:2324-2331.
21. Meijers R, Lai C-CC, Yang Y, Liu J-HH, Zhong W, Wang J-HH, Reinherz EL: **Crystal structures of murine MHC Class I H-2 D(b) and K(b) molecules in complex with CTL epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance.** *J Mol Biol* 2005, **345**:1099-1110.
22. Suzuki R, Shimodaira H: **Pvclust: an R package for assessing the uncertainty in hierarchical clustering.** *Bioinformatics* 2006, **22**(12):1540-1542.
23. Shimodaira H: **An approximately unbiased test of phylogenetic tree selection.** *Syst Biol* 2002, **51**(3):492-508.
24. Halstead SB: **Identifying protective dengue vaccines: Guide to mastering an empirical process.** *Vaccine* 2013.

25. Duan ZL, Li Q, Wang ZB, Xia KD, Guo JL, Liu WQ, Wen JS: **HLA-A*0201-restricted CD8+ T-cell epitopes identified in dengue viruses.** *Virology* 2012, **9**:259.
26. Donati C, Rappuoli R: **Reverse vaccinology in the 21st century: improvements over the original design.** *Ann N Y Acad Sci*, **1285**:115-132.
27. Dormitzer PR, Grandi G, Rappuoli R: **Structural vaccinology starts to deliver.** *Nat Rev Microbiol*, **10**(12):807-813.
28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242.
29. Kaas Q, Ruiz M, Lefranc M-P: **IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data.** *Nucleic acids research* 2004, **32**:D208-210.
30. Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, Bourne PE: **IEDB-3D: structural data within the immune epitope database.** *Nucleic acids research* 2011, **39**:D1164-1170.
31. Trott O, Olson AJ, News S: **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.** *J Comput Chem* 2010, **31**:455-461.
32. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D *et al*: **GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.** *Bioinformatics* 2013, **29**(7):845-854.
33. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M, Alexov E: **DelPhi: a comprehensive suite for DelPhi software and associated resources.** *BMC Biophys* 2012, **5**:9.
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera-a visualization system for exploratory research and analysis.** *J Comput Chem* 2004, **25**:1605-1612.

Figures

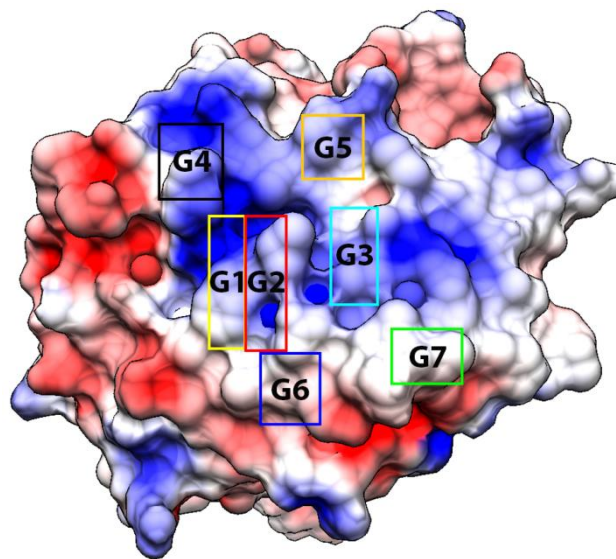


Figure 1 – Seven Gates defined to obtain colour histograms.

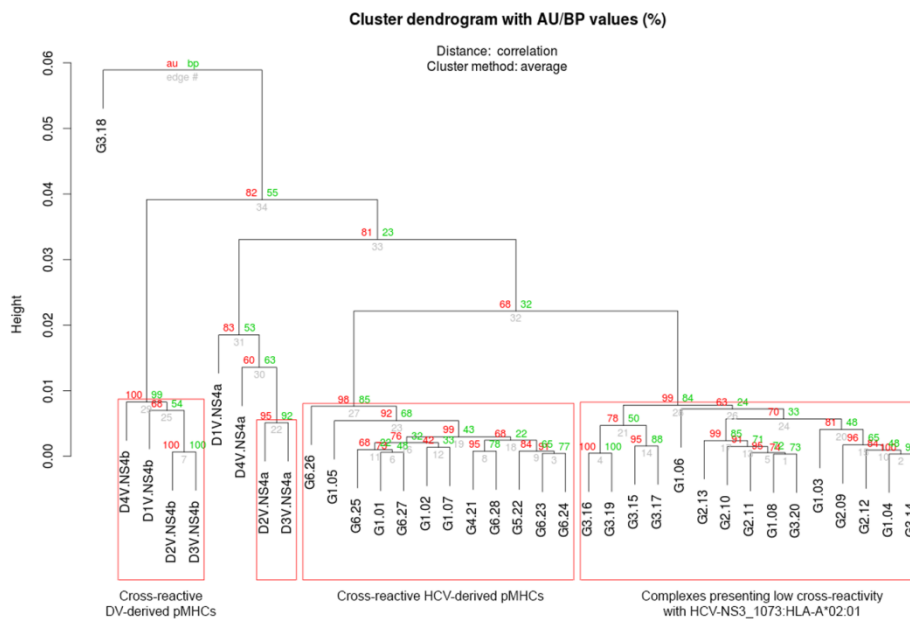


Figure 2 - Structure-based hierarchical clustering of pMHC-I complexes.

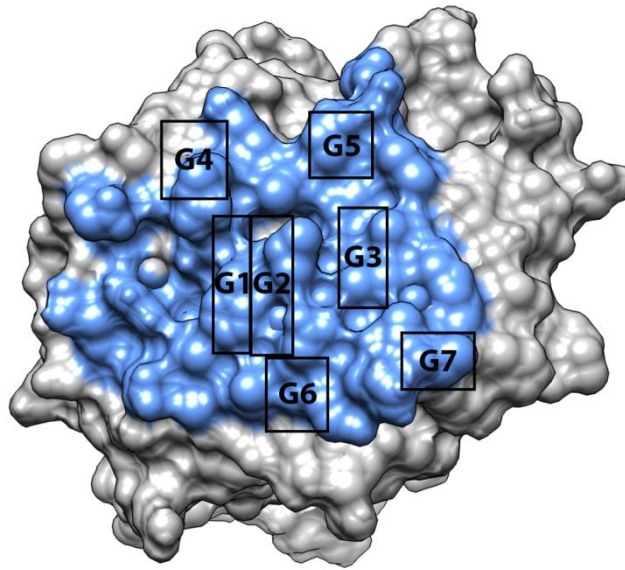


Figure S1 – Selected residues for ASA assessment.

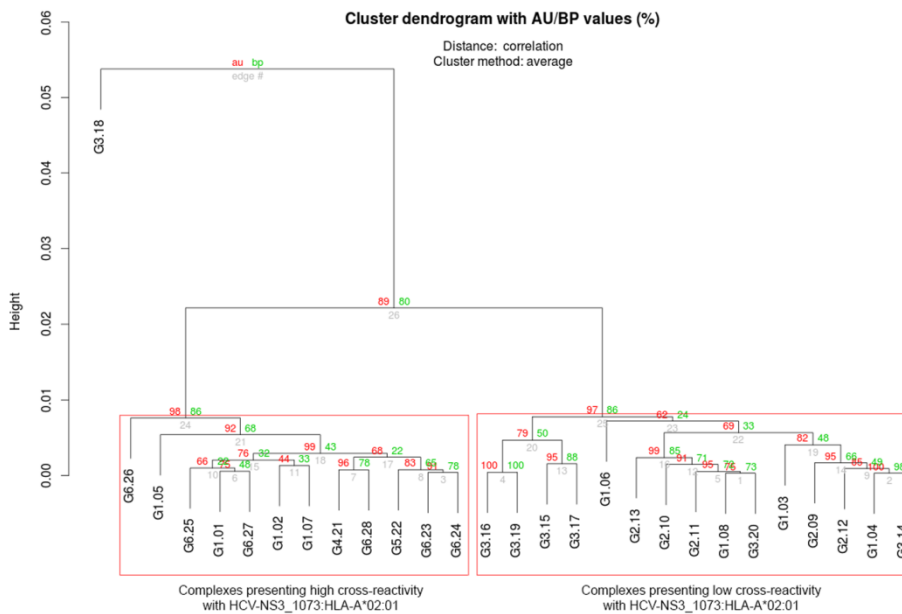


Figure S2 – HCA of 28 HCV natural occurring variants

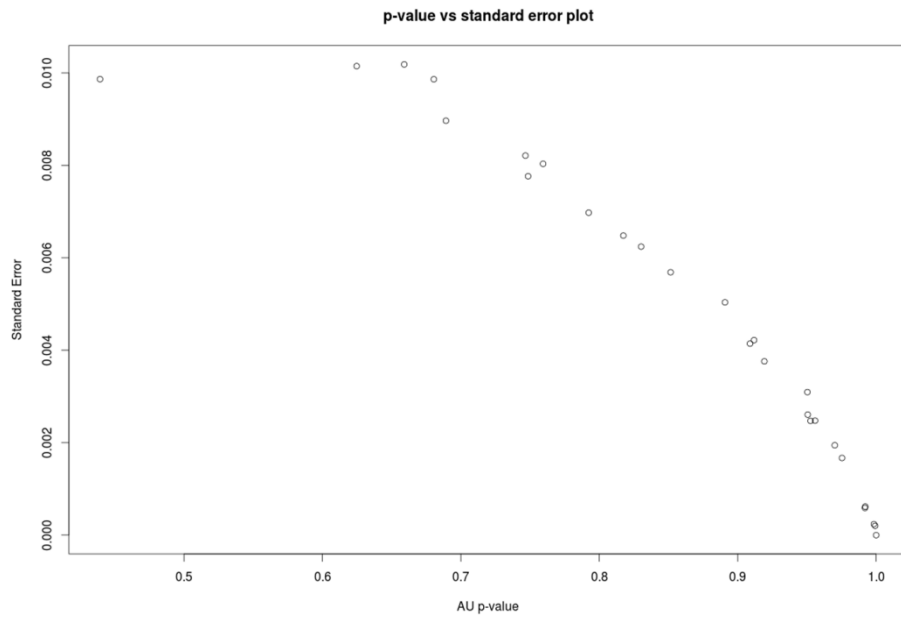


Figure S3 – Standard Error plot from the clustering of HCV variants.

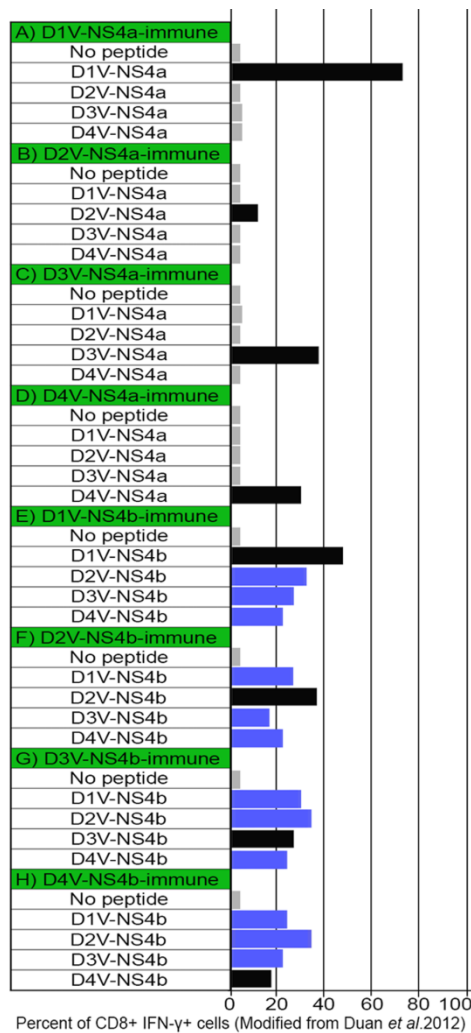


Figure S4 – Cross-reactivities among DV serotypes.

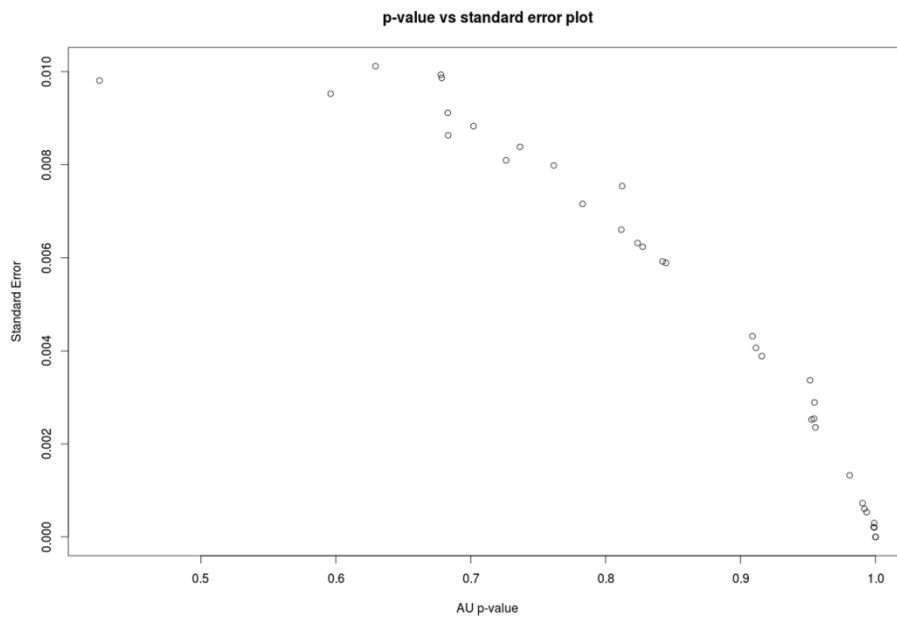


Figure S5 – Standard Error plot from the clustering of 36 pMHC-I complexes.

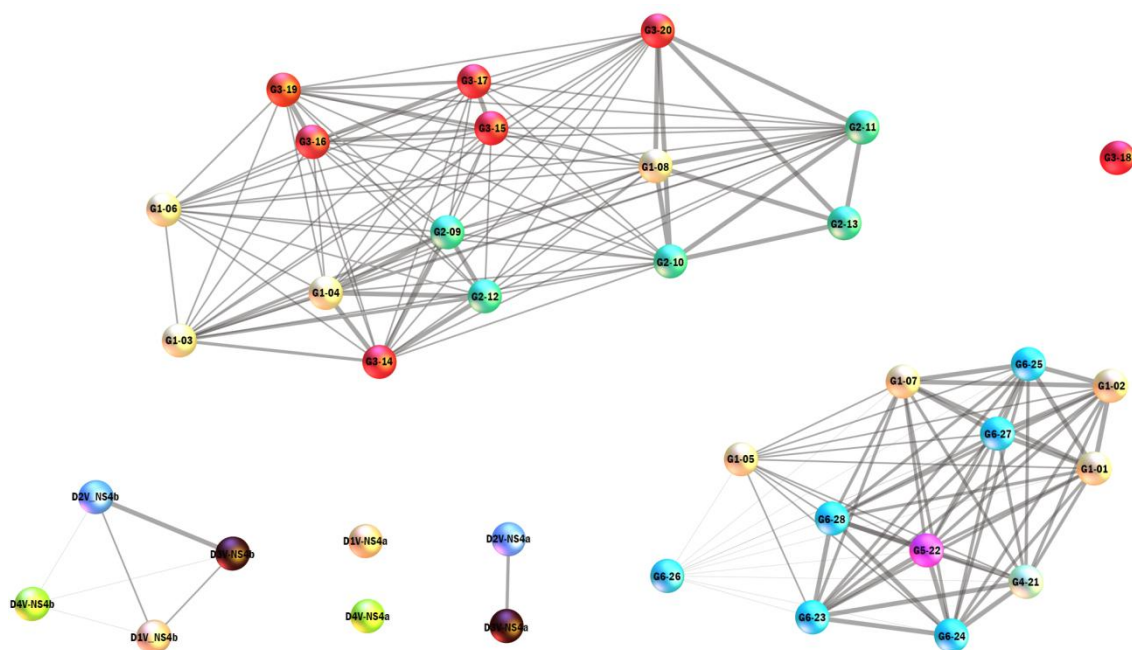


Figure S6 – Relational network of 36 pMHC-I complexes.

Capítulo III

Discussão geral

Discussão Geral

Infecções causadas pelos vírus da hepatite C e da dengue não possuem vacinas eficazes até hoje e são alvos de diversas pesquisas em busca de uma forma profilática funcional e adequada. Nosso trabalho se desenvolveu em torno da otimização de predições de redes de reatividade cruzada *in silico* entre epitopos de subtipos destes vírus, todos apresentados no contexto do alelo de MHC humano HLA-A:0201. Para isto, utilizamos um *pipeline* com diversos softwares envolvidos, que incluem desde a montagem dos complexos, extrações dos padrões de potencial eletrostático e ASA, análises estatísticas e a ferramenta para representação das redes. Detalhes deste *pipeline* são mostrados na figura 3.

Conforme já havia sido observado em trabalhos anteriores do nosso grupo (Antunes *et al.*, 2011), é possível, utilizando valores extraídos do campo eletrostático das superfícies de interação com o TCR de complexos p:MHC, agrupar epitopos que possuem reatividade cruzada pré estabelecida (Fyttili *et al.*, 2008). O primeiro diferencial neste trabalho foi a padronização das regiões de contato com o TCR dos complexos p:MHC, definidas a partir da análise de 28 cristais extraídos do PDB contendo complexos ternários (TCR/p:MHC), as quais foram selecionadas para a extração dos valores RGB (quantidade de Red, Green e Blue de cada região). Foram selecionados 9 regiões, sendo 4 da superfície do epitopo e 3 do MHC, onde constatou-se interações de resíduos do TCR com estes resíduos do complexo p:MHC em pelo menos 85% dos cristais analisados. No trabalho anterior, estas regiões foram selecionadas arbitrariamente, tendo como critério de seleção os pontos que aparentemente variavam mais entre as superfícies dos 28 complexos p:MHC de HCV. Além disso, nós utilizamos pela primeira vez valores numéricos que refletem a topografia dos resíduos que contactam o TCR, utilizando os valores calculados de ASA para cada resíduo incluído na análise. O método estatístico de agrupamento foi o mesmo utilizado previamente, a clusterização hierárquica com a utilização de uma validação estatística adicional contendo os valores de *bootstrap* para cada

agrupamento formado. Dessa maneira, obteve-se uma robustez maior para cada uma das relações observadas.

A distribuição do campo eletrostático na superfície de interação dos complexos p:MHC vem se mostrando um dos principais elementos que devem contribuir para o reconhecimento pelo Linfócito T (Flower, 2007). Como já foi citada anteriormente, essa variável foi informativa para ter um acerto considerável no agrupamento de epitopos que apresentavam reatividade cruzada separando-os daqueles que não eram imunogênicos. Além disso, utilizando essa abordagem nós fomos capazes de prever, inclusive, reatividade cruzada para um epitopo sem evidências prévias *in vitro*, o qual foi posteriormente testado e confirmado em experimentos *in vitro* posteriores, confirmando a validade da nossa técnica (Corneberg, M.). Uma implementação neste trabalho foi que o cálculo do potencial eletrostático, diferentemente do trabalho anterior que aplicava a ferramenta grasp, foi realizado com o delphi, que embora use a mesma equação como base para o cálculo (Poisson-Boltzmann) (Li *et al.*, 2012b), possui uma forma de implementação melhor para scripts, podendo ser executado por linhas de comandos, o que facilita a automatização do processo.

A topologia foi utilizada como um dado adicional para as nossas análises, servindo de complemento aos dados da distribuição eletrostática do p:MHC. As cadeias laterais dos resíduos do epitopo podem apresentar diversos tamanhos e conformações dentro da fenda, o que definirá se elas estarão mais ou menos exposta ao solvente e, conseqüentemente, se poderão interagir em uma proporção maior ou menor com o TCR. Esta exposição dos resíduos ao solvente é calculada facilmente pela ferramenta *naccess*, o qual gera como saída uma tabela com todos os valores da porcentagem do aminoácido que está exposto. Foram extraídos os valores de ASA dos 9 resíduos que compõe o epitopo, mais 28 resíduos do MHC formadores da interface que contacta o TCR.

Para testar a validade desta nova técnica, em relação à técnica anterior, foi utilizado o mesmo conjunto contendo os 28 subtipos de HCV, publicados previamente (Fyttili *et al.*, 2008). Aplicando nossa nova seleção das regiões e a inclusão dos dados de ASA dos resíduos de contato, conseguimos reproduzir um resultado melhor que o da

técnica anterior. Um ganho observável com essa nova abordagem é que, em alguns casos, epitopos que apresentavam resposta intermediária agrupavam junto com aqueles que apresentavam reatividade cruzada com o epitopo CVNGVCWTV (G1_01 - o qual tinha resposta 100% equivalente ao epitopo selvagem CINGVCWTV contido na vacina IC41). Podemos citar, por exemplo, o caso do complexo G1_08 (intermediário), o qual aparece na técnica antiga no ramo dos que possuem reatividade cruzada (ao lado do G1_01) e na nossa técnica este se agrupa em um ramo com outros epitopos com resposta intermediária (G2_10, G2_11, G2_13 e G3_20), o que evidencia um ganho de resolução com a adição do ASA e das novas regiões definidas, conforme pode ser observado na figura 4.

Para análise comparativa dos agrupamentos, utilizamos primeiramente o programa estatístico SPSS, a qual utiliza um algoritmo de agrupamento hierárquico (HCA), gerando como saída um dendograma mostrando os grupos formados. Os parâmetros utilizados do HCA para a clusterização foram o método “*between-groups linkage*” e para calcular a distância utilizou-se o “*squared Euclidean distance*”, sendo estes parâmetros os mesmos utilizados no trabalho anterior (Antunes *et al.*, 2011). No trabalho prévio a seleção do limiar foi feita por observação visual, definida pela altura do ramo na qual a separação dos grupos está relacionada com os grupos formados no experimento da imunização, a qual se demonstrou apropriada agrupando corretamente a maioria dos 28 complexos testados pela técnica anterior, sendo mantido este limiar (figura 4). Porém neste trabalho foi utilizada uma validação estatística adicional, através da técnica de cálculo de *bootstrap*, a qual foi obtida através de um pacote em R para clusterização hierárquica com valores de p (*pvclust*). Estes valores de p calculados conferem uma robustez estatística ao agrupamento formado, onde valores de AU (“Approximately Unbiased” bootstrap –valores em vermelho no dendograma) superiores a 95% são fortemente suportados pelos dados. O *pvclust* utiliza para a clusterização o método de “*Average*” e para calcular a distância utiliza o método de “*correlation*”, conforme explicado no capítulo 2 desta dissertação.

Também foi feito o mesmo tipo de análise em outra ferramenta disponível na internet, o PIPSA (Blomberg *et al.*, 1999). Esta ferramenta provê um serviço de comparação de propriedades eletrostáticas de proteínas, permitindo a classificação

destas de acordo com suas cargas. Enviamos os 28 PDB's do grupo do HCV e o resultado foi um agrupamento que não condiz com o que é observado nos estudos anteriores, conforme é mostrado na figura 3. Ele agrupou, por exemplo, o G1_01, o qual estimula 100% de resposta nos linfócitos do indivíduo imunizado com o epitopo selvagem, junto com G3_14 e G3_16, que estão entre os intermediários, e que possuem uma baixa reatividade cruzada com o epitopo selvagem, conforme mostrado na figura 4. Um dos motivos pelo qual esta ferramenta não agrupou corretamente refere-se ao fato que é calculado o potencial eletrostático da molécula inteira, ao contrario da nossa técnica que só leva em consideração determinados pontos com os quais o TCR interage com o pMHC. Isto provavelmente está interferindo na análise, criando um fator de confusão e, conseqüentemente, ramificações incorretas. Para efeitos comparativos, a figura 4 mostra a saída do pipsa, da técnica antiga e da nova técnica.

Com a observação de que nossa nova abordagem apresentava ganhos reais na capacidade de agrupamentos, especialmente dos epitopos com valores intermediários, nós a aplicamos para a análise de um conjunto adicional de alvos virais, composto por quatro subtipos de vírus da dengue. Estes conjuntos de dados foram extraídos de um artigo que apresenta os dados de resposta celular *in vitro* destes alvos e a ocorrência de reatividade cruzada entre esses subtipos. Os epitopos são derivados de duas proteínas, sendo aqui dividido em Ns4a e Ns4b, os quais são apresentados pelo alelo de MHC HLA-A:0201 (o mesmo do outro conjunto com os 28 epitopos de HCV). Os dados *in vitro* revelam a ocorrência de uma reatividade cruzada entre os epitopos provenientes da proteína Ns4b (entre todos os 4 subtipos), o que não ocorre entre os epitopos da proteína Ns4a.

A nossa técnica conseguiu reproduzir os mesmos resultados para os epitopos derivados da proteína Ns4b, onde os 4 subtipos do vírus da Dengue ficaram agrupados no mesmo ramo (Figura 2 do Capítulo 2). Entretanto, para os epitopos da proteína Ns4a, houve um agrupamento inesperado entre o subtipo D2V-Ns4a e D3V-Ns4a, pois no artigo não foi constatada reatividade cruzada da proteína a nos subtipos. Este falso positivo pode ser discutido, pois no artigo usa-se apenas um pequeno *pool* de linfócitos T colhidos de um único indivíduo. É possível que outros indivíduos, ou até

mesmo outros linfócitos do mesmo indivíduo, tenham capacidade de gerar respostas cruzadas contra esses 2 alvos. Outro ponto que não pode ser descartado é que ainda precisamos fazer uma parametrização com outros conjuntos de epitopos para obtenção de um limiar que reflita com melhor precisão a formação de agrupamentos que estimulem resposta cruzada.

Para uma melhor visualização dos agrupamentos, utilizamos o programa Gephi para montar uma rede entre todos os epitopos analisados, conforme demonstrado na figura suplementar 6, situada no capítulo 2 desta dissertação. As espessuras das arestas mostram o quão próximos estão cada complexo do nó que está ligado. Para calcular esta proximidade, foi utilizado a distâncias dos ramos gerados pelo HCA, sendo que para cada ramo foi aferido um valor. Com isto, complexos que estão mais próximos em determinados ramos possuem valores maiores, possuindo este par uma espessura maior na aresta do que complexos em ramos mais distantes. Para exemplificar, podemos citar o D2V_NS4b, o qual está mais próximo ao D3V_NS4b em relação ao D4V_NS4b, possuindo assim uma aresta de maior intensidade que o conecta ao D3V, comparado com a aresta que o liga ao D4V.

Os avanços computacionais aliados ao desenvolvimento de novas ferramentas de simulação e predição cada vez mais precisas, alavancam cada vez mais as pesquisas nesta área da imunoinformática. Os baixos custos e o reduzido período de tempo despendido com uma análise *in silico* faz com que mais imunologistas lancem mão inicialmente deste tipo de abordagem para depois recorrerem a experimentos *in vitro* e *in vivo*. Nossa meta primordial é a padronização desta técnica de predição de agrupamentos entre alvos que possuem reatividade cruzada para outros alelos tornando-a disponível no sitio do Crosstope Structural Data Bank (Sinigaglia *et al.*, 2013b) para que outros pesquisadores possam realizar suas próprias predições de maneira “customizada”, inclusive com a implementação completa de uma rede gráfica para a visualização dos grupos formados.

Nossa abordagem de predição de reatividade cruzada utilizando um enfoque estrutural é única, sendo que outros trabalhos focam-se na análise e comparação de sequências de epitopos (Moise *et al.*, 2013b).

Apesar de muitos avanços e das nossas tentativas iniciais de abordar um problema antigo através de um método inovador, ainda sabemos pouco sobre a infinidade de elementos que regem a interação do complexo ternário TCR/p:MHC. Somente a partir do continuado desenvolvimento de mais pesquisas básicas nessa área que poderemos decifrar esse encantador mundo da imunogenicidade. A utilização desses conhecimentos a nosso favor permitirá o desenvolvimento de novas vacinas e imunoterapias, fazendo com que doenças como a hepatite C, dengue e AIDS apareçam apenas nos livros de histórias.

Figuras

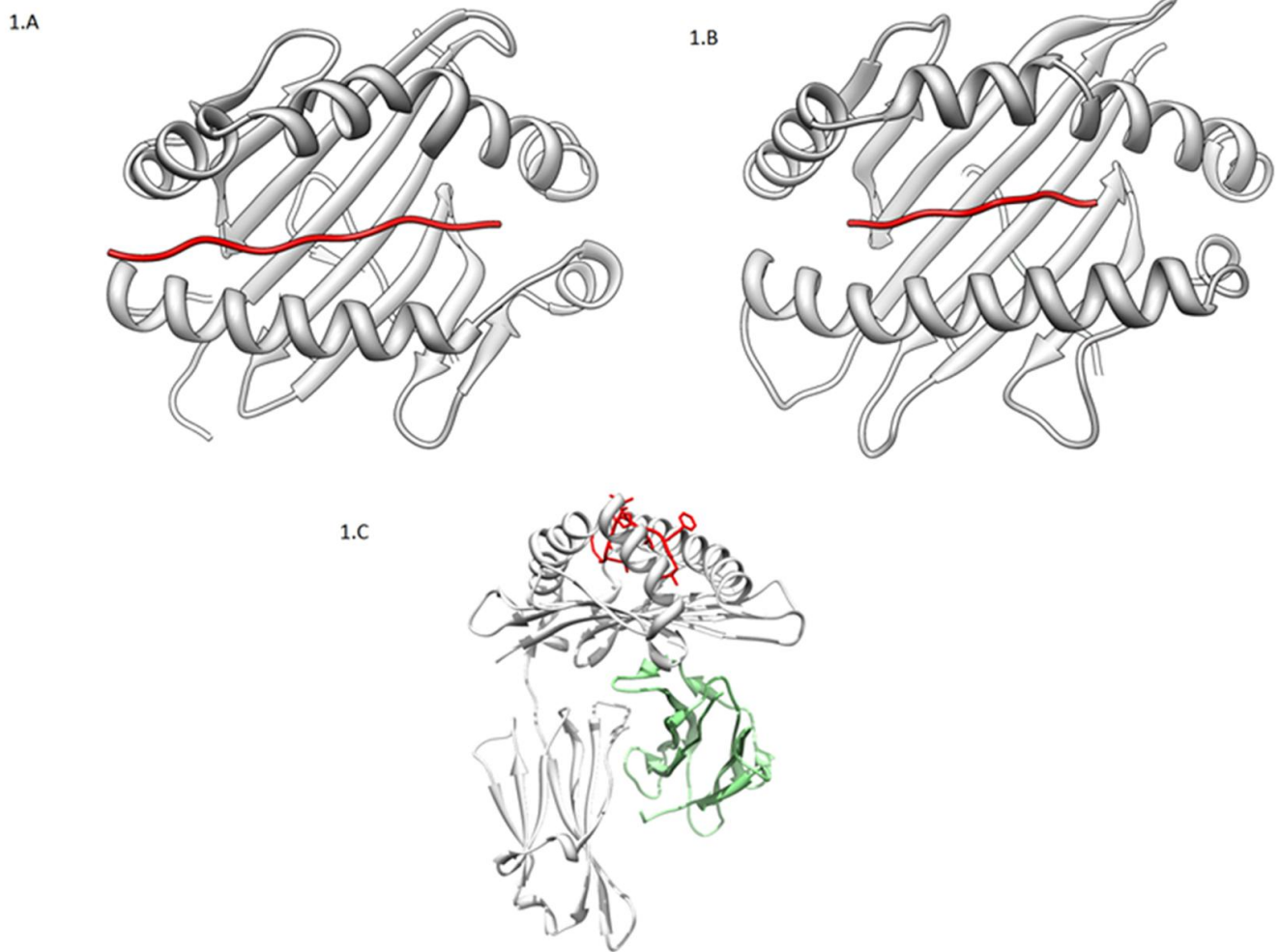


Figura 1: 1.A) Figura mostrando uma fenda de MHC-II com o epitopo ancorado. 1.B) Figura mostrando a fenda do MHC-I, com o epitopo ancorado. 1.C) Figura mostrando um complexo p:MHC, sendo as estruturas alfa na cor cinza, beta 2-microglobulina na cor verde e o epitopo na cor vermelha.

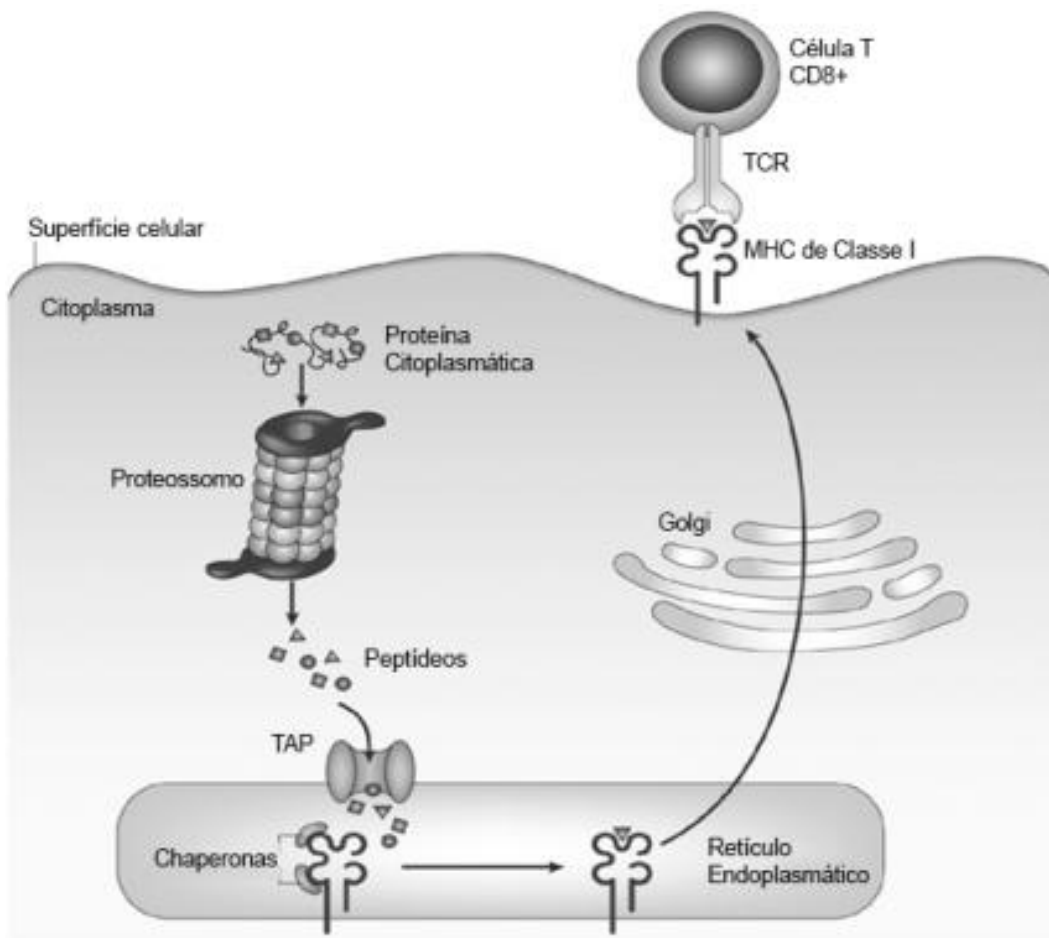


Figura 2: Rota de apresentação de peptídeos endógenos. Modificado de Yewdell e cols., 2003.

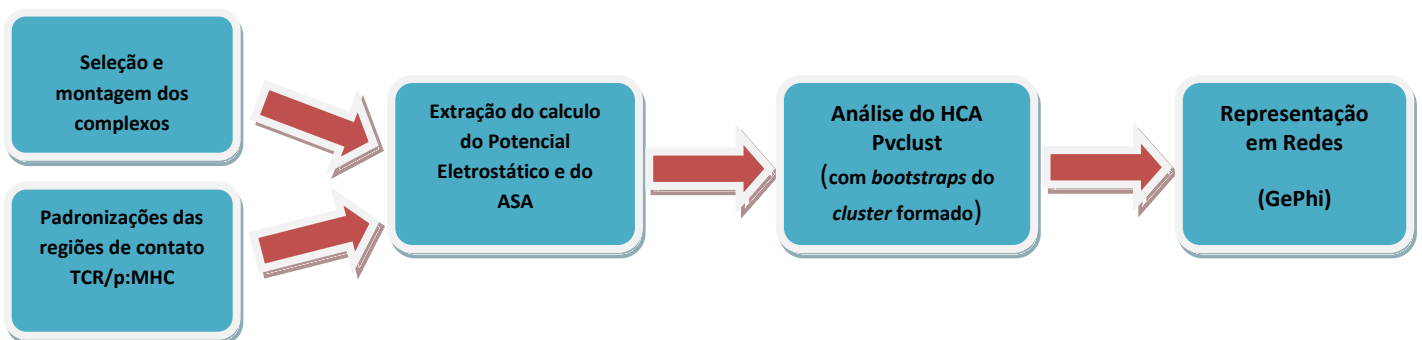


Figura 3: Fluxograma mostrando as etapas de todo o trabalho executado.

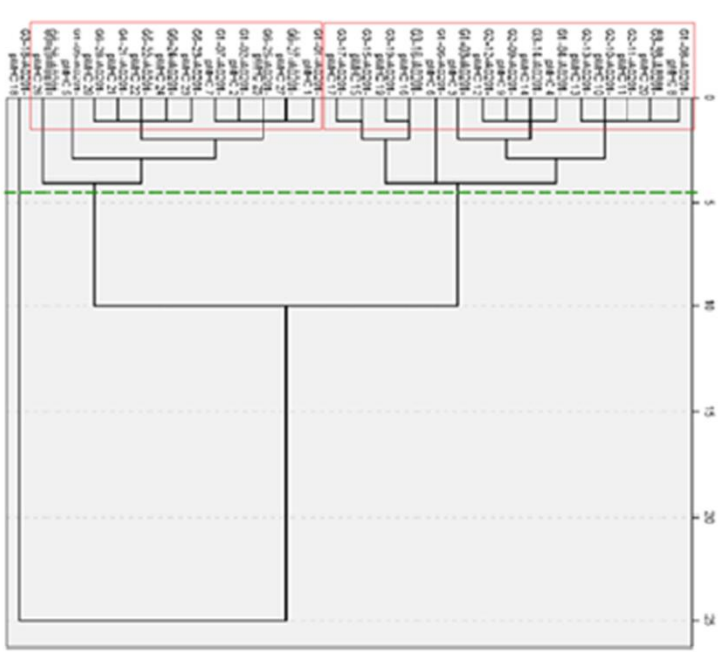
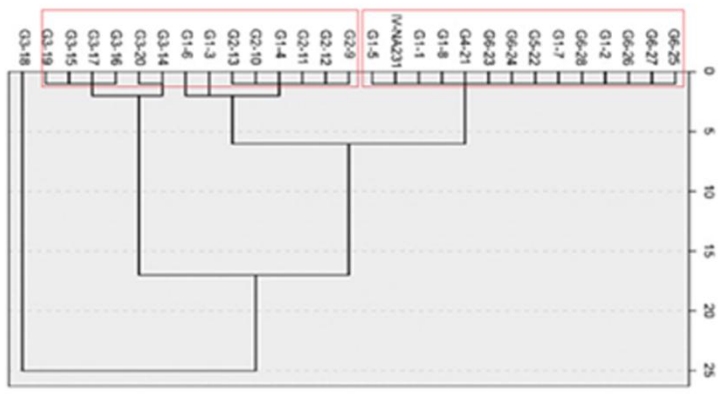
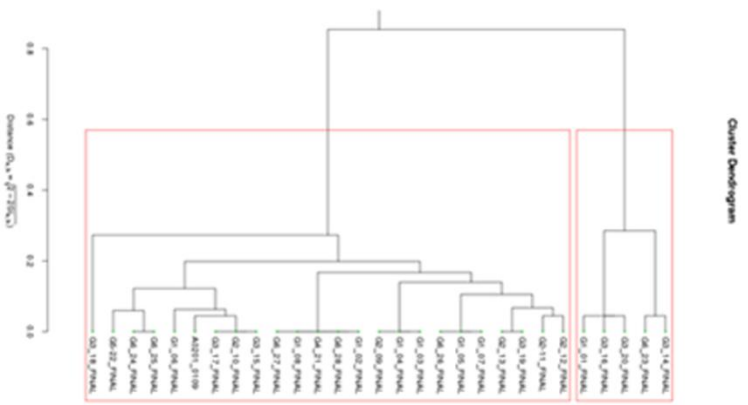
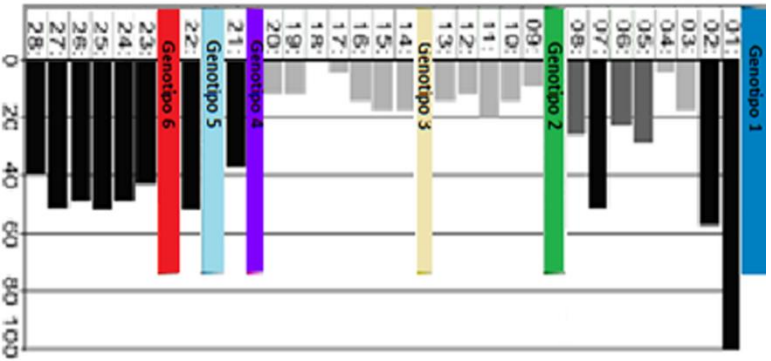


Figura 4: Figura mostrando, na primeira caixa, o índice de reatividade contra o genótipo selvagem, medido pela produção de interferon gama (modificado de Fylli et al. 2008). Os 3 dendogramas seguintes, mostrado da esquerda pra direita são os seguintes resultados: Agrupamento do PPSA, Agrupamento da técnica antiga e Agrupamento da técnica nova. Observa-se uma melhor ramificação da técnica nova em comparação com as outras técnicas, tendo como parâmetro o resultado do artigo da Fylli et. al. 2008

Referencias Bibliográficas

- 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.
- Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M and Chies JA (2010a) Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment. *PLoS One* 5:e10353.
- Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M and Chies JAB (2010b) Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment. *PloS one* 5:e10353.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235-242.
- Bhati M, Cole DK, McCluskey J, Sewell AK and Rossjohn J (2014) The versatility of the $\alpha\beta$ T-cell antigen receptor. *Protein Science* 23:260-272.
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O *et al.* (2013) The global distribution and burden of dengue. *Nature* 496:504-507.
- Blomberg N, Gabdoulline RR, Nilges M and Wade RC (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins* 37:379-387.
- Brehm MA, Selin LK and Welsh RM (2004) CD8 T cell responses to viral infections in sequence. *Cellular microbiology* 6:411-421.
- Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, Luzuriaga K, Welsh RM and Selin LK (2005) Cross-reactive influenza virus – specific CD8 + T cells contribute to lymphoproliferation in Epstein-Barr virus – associated infectious mononucleosis. 115.
- Cooper M and Alder M (2006) The evolution of adaptive immune systems. *Cell* 124:815-822.
- Cornberg M, Clute SC, Watkin LB, Saccoccio FM, Kim S-k, Naumov YN, Brehm MA, Aslan N, Welsh RM and Selin LK (2010) CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections. *Journal of immunology (Baltimore, Md. : 1950)* 184:2825-2838.
- Diamond MS, Roberts TG, Edgil D, Lu B, Ernst J, Harris E, Roberts TGUY and Harris EVA (2000) Modulation of Dengue Virus Infection in Human Cells by Alpha , Beta , and Gamma Interferons Modulation of Dengue Virus Infection in Human Cells by Alpha , Beta , and Gamma Interferons.
- Donati C and Rappuoli R Reverse vaccinology in the 21st century: improvements over the original design. *Ann N Y Acad Sci* 1285:115-132.
- Dormitzer PR, Grandi G and Rappuoli R Structural vaccinology starts to deliver. *Nat Rev Microbiol* 10:807-813.
- Duan ZL, Li Q, Wang ZB, Xia KD, Guo JL, Liu WQ and Wen JS (2012) HLA-A*0201-restricted CD8+ T-cell epitopes identified in dengue viruses. *Virology* 9:259.
- Ender PV, Irap ÁEÁ and li ÁTPP (2011) Post-proteasomal and proteasome-independent generation of MHC class I ligands.1553-1567.
- Fernandez-Vina MA, Falco M, Sun Y and Stastny P (1992) DNA typing for HLA class I alleles: I. Subsets of HLA-A2 and of -A28. *Hum Immunol* 33:163-173.
- Flower DR (2007) *Immunoinformatics: Predicting Immunogenicity in Silico*. Humana Press

- Frakes WW and Kleinschmidtfl JA (1994) Distinct 19 S and 20 S Subcomplexes of the 26 S Proteasome and Their Distribution in the Nucleus and the Cytoplasm *. 269:7709-7718.
- Frankild S, de Boer RJ, Lund O, Nielsen M and Kesmir C (2008) Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. PLoS ONE 3:0.
- Fytili P, Dalekos GN, Schlaphoff V, Suneetha PV, Sarrazin C, Zauner W, Zachou K, Berg T, Manns MP, Klade CS *et al.* (2008) Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. Vaccine 26:3818-3826.
- Gras S, Burrows SR, Turner SJ, Sewell AK, McCluskey J and Rossjohn J (2012) A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned. Immunol Rev 250:61-81.
- Gras S, Saulquin X, Reiser J-B, Debeauvais E, Echasserieu K, Kissenpfennig A, Legoux F, Chouquet A, Le Gorrec M, Machillot P *et al.* (2009) Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. Journal of immunology (Baltimore, Md. : 1950) 183:430-437.
- Halliday J, Klenerman P and Barnes E (2011) Vaccination for hepatitis C virus: closing in on an evasive target. Expert Rev Vaccines 10:659-672.
- Halstead SB (2013) Identifying protective dengue vaccines: Guide to mastering an empirical process. Vaccine.
- Hammer GE, Kanaseki T and Shastri N (2007) The Final Touches Make Perfect the Peptide-MHC Class I Repertoire. Immunity 26:397-406.
- Hiroishi K, Eguchi J, Ishii S, Hiraide a, Sakaki M, Doi H, Omori R and Imawari M (2010) Immune response of cytotoxic T lymphocytes and possibility of vaccine development for hepatitis C virus infection. J Biomed Biotechnol 2010:263810.
- Jorgensen JL, Esser U, Fazekas de St Groth B, Reay PA and Davis MM (1992) Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. Nature 355:224-230.
- Kaas Q, Ruiz M and Lefranc M-P (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic acids research 32:D208-210.
- KATO N (2000) Genome of Human Hepatitis C Virus (HCV):
Gene Organization, Sequence Diversity, and Variation. MICROBIAL & COMPARATIVE GENOMICS 5:22.
- Kessels HWHG, de Visser KE, Tirion FH, Coccoris M, Kruisbeek AM and Schumacher TNM (2004) The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. Journal of immunology (Baltimore, Md. : 1950) 172:2324-2331.
- Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M and Alexov E (2012a) DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys 5:9.
- Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M and Alexov E (2012b) DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC biophysics 5:9.
- Li RYWK (2011) Host Factors in the Replication of Positive-Strand RNA Viruses. Chang Gung Med J 35:13.
- Li Y, Yin Y and Mariuzza Ra (2013) Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. Frontiers in immunology 4:206.
- Mauri C and Bosma A (2012) Immune regulatory function of B cells. Annual review of immunology 30:221-241.
- Meijers R, Lai C-CC, Yang Y, Liu J-HH, Zhong W, Wang J-HH and Reinherz EL (2005) Crystal structures of murine MHC Class I H-2 D(b) and K(b) molecules in complex with CTL

- epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance. *J Mol Biol* 345:1099-1110.
- Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, Verberkmoes NC, Sztejn MB, Losikoff PT, Martin WD *et al.* (2013a) The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. *Hum Vaccin Immunother* 9.
- Moise L, Gutierrez AH, Bailey-kellogg C, Terry F, Leng Q, Hady KMA, Verberkmoes NC, Sztejn MB, Losikoff PT, Martin WD *et al.* (2013b) Examining the host-microbe interface with JanusMatrix The two-faced T cell epitope.1577-1586.
- Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE *et al.* (2003) The DNA sequence and analysis of human chromosome 6.805-812.
- Nakano T, Lau GM, Lau GM, Sugiyama M and Mizokami M (2012) An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver Int* 32:339-345.
- Peaper DR and Cresswell P (2008) Regulation of MHC class I assembly and peptide binding. *Annual review of cell and developmental biology* 24:343-368.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612.
- Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A and Bourne PE (2011) IEDB-3D: structural data within the immune epitope database. *Nucleic acids research* 39:D1164-1170.
- Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845-854.
- Qin J and Mamotte C (2008) A map of the class III region of the sheep major histocompatibility complex. 9:1-9.
- Rangarajan S and Mariuzza Ra (2014) T cell receptor bias for MHC: co-evolution or co-receptors? *Cellular and molecular life sciences : CMLS*.
- Regner M and Article R (2001) Cross-reactivity in T-cell antigen recognition. *Immunology and cell biology* 79:91-100.
- Rudolph MG, Stanfield RL and Wilson Ia (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annual review of immunology* 24:419-466.
- Sandalova T, Michaelsson J, Harris RA, Odeberg J, Schneider G, Karre K, Achour A, Michaëlsson J and Kärre K (2005) A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity. *The Journal of biological chemistry* 280:27069-27075.
- Selin LK, Nahill SR and Welsh RM (1994) Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. *The Journal of experimental medicine* 179:1933-1943.
- Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492-508.
- Shlomchik MJ, Craft JE and Mamula MJ (2001) From T to B and back again: positive feedback in systemic autoimmune disease. *Nat Rev Immunol* 1:147-153.
- Sinigaglia M, Antunes DA, Rigo MM, Chies JA and Vieira GF (2013a) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database (Oxford)* 2013:bat002.
- Sinigaglia M, Antunes DA, Rigo MM, Chies JAB and Vieira GF (2013b) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database : the journal of biological databases and curation* 2013:bat002.
- Suzuki R and Shimodaira H (2006) Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22:1540-1542.

- Trott O, Olson AJ and News S (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455-461.
- Vieira GF and Chies JAB (2005) Immunodominant viral peptides as determinants of cross-reactivity in the immune system--Can we develop wide spectrum viral vaccines? *Medical hypotheses* 65:873-879.
- Wang J and Maldonado Ma (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. *Cellular & molecular immunology* 3:255-261.
- Wedemeyer H, Mizukoshi E, Davis AR, Bennink JR and Rehermann B (2001) Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells. *Journal of virology* 75:11392-11400.
- Welsh RM and Fujinami RS (2007) Pathogenic epitopes, heterologous immunity and vaccine design. *Nature reviews. Microbiology* 5:555-563.
- Welsh RM and Selin LK (2002) No one is naive: the significance of heterologous T-cell immunity. *Nat Rev Immunol* 2:417-426.
- Yu CI and Chiang B-L (2010) A new insight into hepatitis C vaccine development. *Journal of biomedicine & biotechnology* 2010:548280.