

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

TESE DE DOUTORADO

**Similaridade Estrutural de Complexos peptídeo:MHC como
um Indicador para a Ocorrência de Reatividade Cruzada**

DINLER AMARAL ANTUNES

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular).

Orientador: Prof. Dr. Gustavo Fioravanti Vieira

Coorientadora: Prof.^a Dr.^a Marialva Sinigaglia

PORTO ALEGRE

AGOSTO DE 2014

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

Apoio financeiro

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

Bill & Melinda Gates Foundation (*Grand Challenges Explorations - Round 2*)

*“Dedico este trabalho aos meus pais, Aristeu e Ana Cristina,
pois tudo que sou e tudo que estou conquistando
é apenas um reflexo do seu exemplo de conduta,
de trabalho e de perseverança.”*

AGRADECIMENTOS

Parafraseando Sir. Isaac Newton (1643-1727), “se vi mais longe foi por estar sobre os ombros de gigantes”. Assim, dedico aqui algumas palavras em agradecimento àqueles que de forma direta ou indireta colaboraram com a minha formação acadêmica, a qual se vê materializada nesta Tese de Doutorado.

*Em primeiro lugar agradeço a **Laura Stertz**, minha noiva, minha amiga, minha companheira. Foste uma constante em minha vida, desde o início da faculdade. Acompanhaste cada dia desta minha trajetória acadêmica, me incentivando nas derrotas e me contendo nas conquistas. Celebraste comigo, choraste comigo, fizeste planos e modificaste tua vida para que o desenvolvimento de nossas carreiras não fosse incompatível com o nosso relacionamento. Agora, foste a responsável por uma revolução em nossas vidas. Um novo capítulo se abre, cheio possibilidades e expectativas. Obrigado por tudo e que nesta nova “temporada” continues sendo a personagem central em minha vida, minha principal referência, meu amor.*

*Agradeço a **meus pais** pelo suporte incondicional. Por acreditarem em mim e por estarem sempre ao meu lado, apesar da distância (geográfica). Vocês foram meus maiores professores, sempre me ensinando com seu exemplo de vida. Esta conquista, também é de vocês.*

***Jonier**, assim como dizes que eu fui uma influência à tua formação científica, saibas que foste a mais forte influência para a minha visão cética do mundo. Tua inteligência, tua cultura e tua visão de mundo me impuseram um padrão de autocrítica, que me obrigou a trazer para o cotidiano a visão científica das coisas, a embasar com referências mesmo as mais coloquiais conversas nas manhãs de sábado e a desfazer a maioria dos paradoxos filosóficos que se estabeleceram em meus primeiros anos de faculdade.*

*Agradeço aos **meus queridos primos** pelo carinho e amizade. Em especial, agradeço a **Larissa**, que além de prima e afilhada tornou-se uma grande companheira. Parceira de corridas nas manhãs de sábado, madrinha e guardiã do **Pacato** e VÍTIMA de infinitos e-mails intermináveis.*

*Agradeço aos meus padrinhos, **Éder, Dina, Derlín e Preta**, que sempre esbanjaram amor e orgulho deste afilhado e que foram muito importantes em diversos momentos da minha formação. Em particular, agradeço ao Tio Derlín, que neste ano foi meu colega de apartamento e que tem sido muito presente em minha vida.*

*Muito obrigado a todos os meus demais tios e familiares, que sempre torceram muito por mim. Em especial agradeço aos meus avós, **Valmi, Wolmy e Vanda**, que foram ao mesmo tempo meus orientadores e meus fãs. Não tenho palavras para expressar meu agradecimento por tudo o que vocês fizeram por mim. Em particular, agradeço a Vandinha, que colou o logo do NBLI na parede da sala de estudos e sempre alardeou aos quatro cantos do Alegrete os feitos do neto Biomédico. Tenho certeza que, assim como o Vô **Gaspar**, ela está muito feliz com a trajetória que estou percorrendo.*

*Agradeço também a família da Laura, em especial ao Sr. **Eldo**, Dona **Therez, Suzi, Roberto, Eduardo e Carina**. Vocês me adotaram como parte da sua família e eu agradeço por todo o carinho, pelos diversos presentes e por toda a atenção que vocês me deram.*

*Agradeço ao **Maurício**, pelos vários anos de amizade e de parceria, neste e em outros projetos. És coautor desta Tese e coautor da minha formação acadêmica. Tuas críticas, sugestões e ideias foram fundamentais para aprimorar o meu trabalho. Entre muitas outras coisas, foste para mim um exemplo de organização e planejamento, de quem “importei” diversas ideias para a confecção de planilhas de controle da taxa de bancada, controle financeiro, projetos e relatórios de pedido de auxílio, desenho dos experimentos, etc. Muito obrigado por tudo! Torço muito pelo teu sucesso e pela tua realização pessoal!*

*Ao **Gustavo**, meu orientador, agradeço por ter me recebido de braços abertos ainda na iniciação científica e por ter me dado a liberdade de incluir (progressivamente) análises estruturais nos projetos do grupo. Foi uma colaboração de sucesso, que rendeu mais de uma dezena de publicações e diversos feitos inenarráveis! Quem poderia imaginar nossos feitos, ao nos olhar em 2006 escondidos no depósito que hoje abriga a cozinha da Imunogenética? Talvez só o **Zéca** tenha conseguido ver mais do que dois malucos da fronteira oeste sem nenhuma base computacional iniciando um projeto de “bioinformática” (seja lá o que isso signifique). Mesmo depois da vinda do **Maurício** e da **Meg**, provavelmente nem nós acreditaríamos que nos próximos anos faríamos contato*

com Morten Nielsen, Rinno Rappuoli, Heiner Wedemeyer, Darren Flower, Alessandro Sette, Líisa Selin, Raymond Welsh e diversos outros pesquisadores incríveis, tanto do Brasil quanto do exterior. Quis o acaso que eu fosse defender a minha Tese de Doutorado ao final do mês de Agosto, mesmo mês em que iniciei minha Iniciação Científica, concluindo assim um ciclo de 8 anos de parceria.

Também não poderia deixar de agradecer a **Meg**, minha amiga, co-orientadora e protetora. Nestes anos, acrescentaste muito ao nosso grupo e aos projetos em que estivemos envolvidos. Consequiste ser gentil e amigável, mas ao mesmo tempo ser muito séria e profissional. Sempre atenta aos detalhes do trabalho, sempre sugerindo formas de melhorá-lo. Muito Obrigado!

Ao **Marcus**, agradeço pela grande parceria, dentro e fora do lab. Pelos convites para tomar Tereré, tomar Chopp e para assistir aos jogos da Copa. Teu trabalho na implementação do pvcIust foi fundamental para este projeto, agregando valor as nossas análises. Obrigado também pela rápida execução das inúmeras tarefas solicitadas em e-mails gigantes, tanto com relação aos HCAs quanto às dinâmicas de IDUA. Não vou entrar no debate sobre tua cidade de origem, para não gerar polêmica, mas fico feliz que tenhas escolhido Porto Alegre para iniciar tua carreira. Desejo sorte e Sucesso!

Agradeço também ao meu referencial como imunologista e como pesquisador, o **Prof. Dr. José Artur Bogo Chies** (viu Zéca, mantive a formalidade!). Muito obrigado por estares sempre disposto a responder minhas dúvidas, fossem científicas, filosóficas ou de qualquer outra natureza. Também agradeço pelas aulas de Paddle, esporte que sem dúvida já é meu favorito. Por fim, agradeço por teres aceito o convite para ser relator deste projeto. Teus comentários serão sempre bem vindos!

Saliento ainda meu agradecimento a todos os demais membros e ex-membros do NBLI, em especial a **Cassiana, Jader, Marina, Caio, Martiela, Bragatte, Matheus e Renata**, por suas participações em etapas deste projeto. Em particular, agradeço a imprescindível colaboração da **Martiela**, que vem materializando alguns de nossos projetos mais antigos.

Neste quesito, cabe um agradecimento especial a todos os meus colaboradores estrangeiros, em especial a **Markus Cornberg, Verena Schlaphoff, Shihong Zhang** e

Paraskeví Fytíli. Muito obrigado pelo carinho com que me receberam em Hannover e pela seriedade com que avaliaram os nossos resultados. Sem dúvida este projeto de doutorado não seria possível sem a participação direta do Dr. Markus Cornberg e sua equipe.

Agradeço também a todos os colegas do laboratório de Imunogenética e aos colegas do PPGBM. Em especial, agradeço a **Francis**, que teve contribuições citadas neste projeto e cuja amizade se estende muito além dos nossos vínculos acadêmicos.

Aproveito para agradecer a todos os meus amigos, em especial ao sempre presente **DINGREMARROG**, ao **Leandro** e a **Raquel**, e ao **Leonardo** “Alegrete”. É muito bom perceber que a nossa amizade permanece, a despeito das mudanças, do tempo e da distância. Neste sentido agradeço também aos **ex-colegas da Biomedicina** e a todos “**Los Malucos**”, pela frequente distração no WhatsApp. Em particular, agradeço ao hermano **José Vargas**, una gran amistad que me llevará conmigo a todas las partes.

Obrigado a **Porto Alegre** e a **UFRGS**, duas “entidades” que estarão pra sempre casadas em minha memória. Foi pela UFRGS que eu vim para Porto Alegre, mas a cidade me mostrou que a Universidade era apenas um de seus vários atrativos. Fui muito feliz nestes quase 10 anos de POA/UFRGS, que terminaram de moldar minha personalidade e consolidaram as bases da minha carreira.

Por fim, agradeço a **todos os meus professores** (tanto da graduação quanto da pós-graduação), aos **funcionários da Universidade** e ao **PPGBM**. Em especial agradeço ao **Elmo Cardoso** (Coord. Administrativo-PPGBM), por todo o suporte fornecido ao longo da minha pós-graduação.

Muito Obrigado a Todos!!!

“Success is the ability to go from one failure to another with no loss of enthusiasm.”

Sir Winston Churchill (1874 - 1965)

Sumário

Abreviaturas	3
Resumo	4
<i>Abstract</i>	5
Capítulo I - Introdução e Objetivos.....	6
Introdução.....	7
O paradoxo da rainha vermelha e a resposta imunológica celular	7
A região do MHC e a família gênica do HLA.....	9
Recombinação somática e seleção tímica	12
Características e consequências da interação TCRpMHC.....	15
Genética, Imunologia e Bioinformática	17
Objetivos.....	20
Capítulo II - <i>Structural Immunoinformatics and Vaccine Development</i>.....	21
Capítulo III - <i>Abundance and privacy of CD8+ HCV-specific T-cells in seronegatives:</i>	
<i>implication for vaccine response</i>	56
Capítulo IV - <i>Peptide:MHC structural similarity as a probability for cross-reactive T cell</i>	
<i>responses</i>	116

Capítulo V - Automação de processos em Imunoinformática	146
Automatização da abordagem <i>D1-EM-D2</i>	148
Revalidação da metodologia de predição de pMHCs	154
Instalação em equipamentos de alto desempenho	155
Disponibilização de uma ferramenta baseada em <i>web</i>	156
Preparação de estruturas para o cálculo do potencial eletrostático.....	157
Desenvolvimento de um <i>plugin</i> para importação dos valores RGB	158
Automatização do HCA com <i>bootstrap</i>	160
Capítulo VI - Discussão Geral	161
Referências Complementares	175
Anexo I.....	179
<i>Improved structural method for T-cell cross-reactivity prediction</i>	179

Abreviaturas

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

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

CRN _ Rede de Reatividade Cruzada (do inglês *Cross-Reactive Network*)

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

D1-EM-D2_ *Docking 1-Energy Minimization-Docking 2*.

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

EBV _ Vírus Epstein-Barr (do inglês *Epstein-Barr Virus*).

HCV _ Vírus da Hepatite C (do inglês *Hepatitis C Virus*).

HLA _ Antígeno Leucocitário Humano (do inglês *Human Leucocyte Antigen*).

IAV _ Vírus Influenza A (do inglês *Influenza A Virus*)

IFN _ Interferon.

LCMV _ Vírus da Coriomeningite Linfocítica (do inglês *Lymphocytic Choriomeningitis Virus*)

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

NMR _ Ressonância Magnética Nuclear (do inglês *Nuclear Magnetic Resonance*).

PCA _ Análise de Componentes Principais (do inglês *Principal Component Analysis*).

PDB _ *Protein Data Bank*. A sigla "pdb" é utilizada para se referir ao formato dos arquivos do PDB.

pMHC _ Complexo peptídeo:MHC.

RGB _ Sistema que utiliza três cores para compor uma imagem (do inglês *Red, Green, Blue*).

RMSD _ Desvio Quadrático Médio (do inglês *Root Mean Square Deviation*).

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

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

TCRpMHC _ Complexo TCR:peptídeo:MHC.

VV (ou VACV) _ Vírus da Vaccinia, agente utilizado na vacina que erradicou a varíola (*smallpox*) em humanos.

Resumo

A coevolução parasita-hospedeiro pode ser apontada como uma das principais responsáveis pela grande diversificação de genes envolvidos na resposta imunológica. A chamada “região do MHC” (na sigla em inglês para *Major Histocompatibility Complex*), localizada no braço curto do cromossomo 6 humano, é a região mais polimórfica e densa do nosso genoma. Os três genes mais polimórficos deste *locus* codificam a cadeia pesada de um complexo referido como MHC de classe I, responsável pela apresentação (na superfície celular) de peptídeos provenientes da degradação de proteínas intracelulares. Este mecanismo é central na resposta antiviral, permitindo que células infectadas sejam identificadas e eliminadas pelos Linfócitos T Citotóxicos. Apesar de estruturalmente similares, cada molécula de MHC apresenta maior afinidade por peptídeos com determinadas características bioquímicas. Assim, quanto maior a variabilidade de MHCs em uma dada população, menor o risco de que todos os indivíduos sejam incapazes de apresentar pelo menos alguns alvos derivados de um determinado vírus. Por outro lado, a resposta imunológica celular e a geração de memória contra este alvo apresentado pelo MHC, depende do reconhecimento específico deste complexo peptídeo:MHC (pMHC) por uma dada população de linfócitos. Neste trabalho empregamos ferramentas de bioinformática para realizar a análise estrutural de complexos pMHC, identificando propriedades envolvidas na estimulação da resposta imunológica celular. Nossos resultados *in silico*, corroborados por experimentos *in vitro* e *ex vivo*, sugerem que a similaridade estrutural de complexos pMHC (em termos de topografia e potencial eletrostático) desempenha um papel central na reatividade cruzada de linfócitos T, com implicações sobre imunidade heteróloga, imunopatologia e desenvolvimento de vacinas.

Abstract

Host-pathogen coevolution can be implicated as one of the main features driving the great diversity of genes involved with immunological response. The so-called “MHC region” (*Major Histocompatibility Complex*), located at the short arm of human chromosome 6, is the most polymorphic and dense region of our genome. The three most polymorphic genes in this *locus* encode the heavy chain of a complex referred as MHC class I, which is responsible for presentation (at cell surface) of peptides derived from the digestion of cytosolic proteins. This mechanism plays a key role in antiviral immune response, allowing infected cells to be identified and eliminated by Cytotoxic T Lymphocytes. Although structurally similar, each MHC molecule presents higher affinity for peptides with certain biochemical properties. Therefore, the greater the variability of MHCs in a given population, the smaller the risk that all individuals are unable to present at least some targets derived from a given virus. On the other hand, cellular immune response and memory generation against the target presented by the MHC, depends on specific recognition of this peptide:MHC (pMHC) complex by a given T cell population. In this work, we use bioinformatics tools to perform structural analysis of pMHC complexes, identifying features involved in triggering cellular immune responses. Our *in silico* results, corroborated by *in vitro* and *ex vivo* experiments suggest that structural similarity among pMHC complexes (topography and electrostatic potential) plays a central role in cross-reactivity of cytotoxic T cells, with implications over heterologous immunity, immunopathology and vaccine development.

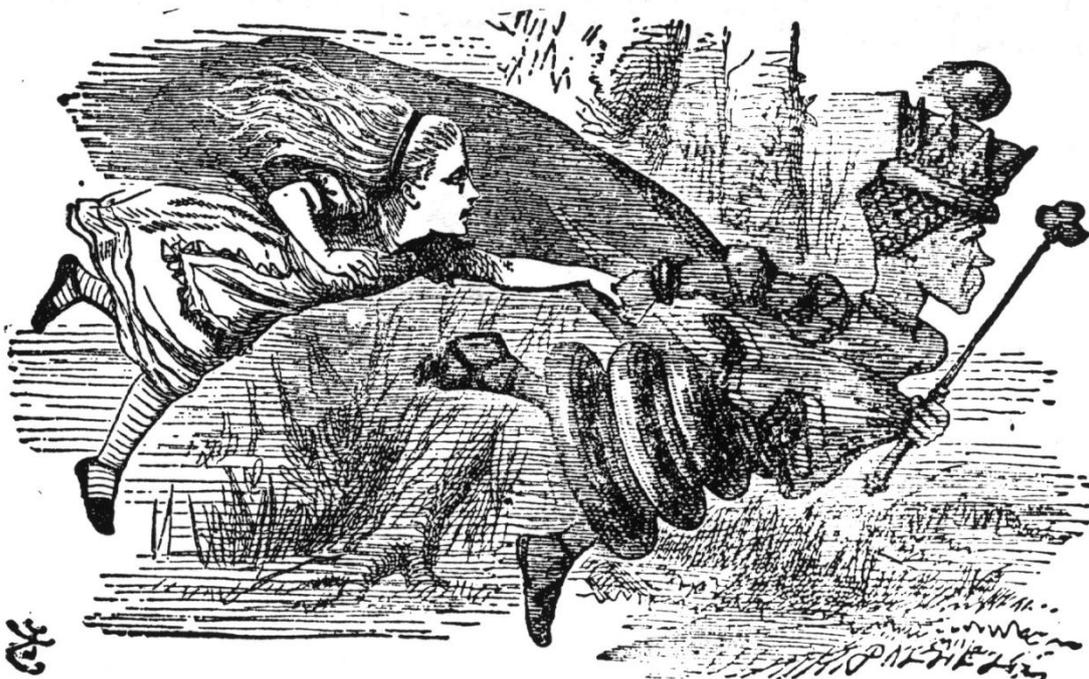
Capítulo I

Introdução e Objetivos

Introdução

O paradoxo da rainha vermelha e a resposta imunológica celular

Em seu famoso livro *Through the Looking-Glass* de 1871 (adaptado para o português como “*Alice no país do espelho*”) o autor Lewis Carroll descreve uma situação incomum na qual é preciso correr o máximo possível para se permanecer no mesmo local (Figura 1). Esta curiosa condição imposta à personagem Alice durante sua visita a um mundo imaginário, acabou servindo como uma perfeita analogia para uma questão central no estudo da biologia, a coevolução entre espécies. A chamada Hipótese da Rainha Vermelha foi inicialmente proposta por Leigh Van Valen para explicar as taxas de extinção no registro paleontológico, sendo posteriormente empregada para descrever a influência da relação parasita-hospedeiro sobre as taxas de evolução molecular (Paterson *et al.*, 2010).



"it takes all the running you can do, to keep in the same place"

Figura 1. A corrida da Rainha Vermelha. Ilustração de Sir John Tenniel para o livro *Through the Looking-Glass* (1871), de Lewis Carroll. Na cena, a Rainha Vermelha informa a Alice que é preciso correr o máximo possível, para se permanecer no mesmo local. Esta frase acabou servindo de analogia para coevolução, em especial entre parasitas e hospedeiros, sendo referenciada como a “Hipótese da Rainha Vermelha”.

Os vírus são parasitas intracelulares obrigatórios, que utilizam a maquinaria molecular do hospedeiro para realizar sua replicação (Salazar *et al.*, 2014; Schmid *et al.*, 2014). Eles normalmente apresentam ciclos rápidos e um controle reduzido da fidelidade durante a replicação genômica (se comparados a eucariotos), o que colabora para a rápida diversidade de sequências, especialmente em vírus com genoma de RNA (Lauring *et al.*, 2013). No entanto, se por um lado nos patógenos são selecionadas estratégias para maximizar sua dispersão na população hospedeira, por outro lado, no organismo hospedeiro são selecionadas estratégias para controlar e eliminar estes patógenos. Graças à reprodução sexuada, a duplicação gênica e a seleção natural, mecanismos de identificação e controle de patógenos foram gradativamente desenvolvidos a partir de mecanismos mais simples de reconhecimento célula-célula (Barclay, 1999). Neste processo, uma série de células e moléculas passou a atuar em conjunto, caracterizando um sistema imunológico. A chamada imunidade natural (ou inata) integra as barreiras epiteliais, as células fagocitárias e um conjunto de moléculas capazes de reconhecer padrões típicos de patógenos, como receptores semelhantes à Toll e proteínas do Sistema Complemento (Huang & Wells, 2014; Nakamoto & Kanai, 2014).

Adicionalmente, a “corrida armamentista” contra os patógenos levou ao desenvolvimento de um novo conjunto de ferramentas, altamente variável, complexo e especializado (Kubinak *et al.*, 2012; Vandiedonck & Knight, 2009a). Além de permitir a identificação de virtualmente qualquer alvo, a despeito de padrões “típicos” de patógenos, este novo “arsenal” também permite a prevenção contra infecções futuras (Welsh *et al.*, 2004). A chamada imunidade adquirida possui duas “frentes de combate”, (i) a resposta humoral e a (ii) resposta celular. A primeira envolve principalmente os linfócitos B e a produção/secreção de anticorpos (imunoglobulinas), enquanto a segunda é mediada pelos linfócitos T. Ambas são fundamentais para a imunidade contra patógenos, mas neste trabalho iremos dar enfoque aos mecanismos envolvidos na resposta celular. Cabe ainda salientar que apesar da frequente representação do sistema imune como um “exército”, sempre pronto a “atacar o inimigo”, os mecanismos da imunidade natural e adquirida desempenham um papel constante na tolerância e

modulação da microbiota (vigilância imunológica), a qual é fundamental para a saúde do hospedeiro (Geuking *et al.*, 2014).

A região do MHC e a família gênica do HLA

Uma das regiões genômicas mais diretamente afetadas por esta coevolução com os vírus foi a chamada região do complexo principal de histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*) (Kubinak *et al.*, 2012; Vandiedonck & Knight, 2009b). Em humanos, ela se encontra no braço curto do cromossomo 6 (6p21.3) e constitui a região mais polimórfica e mais densa do genoma, a qual alcança em alguns trechos a média de 8,5 genes por 100 Kb (Vandiedonck & Knight, 2009a; Xie *et al.*, 2003). Esta região possui envolvimento direto na resposta imunológica, sobretudo na suscetibilidade às doenças infecciosas e autoimunes (Fellay *et al.*, 2007; Vandiedonck & Knight, 2009a).

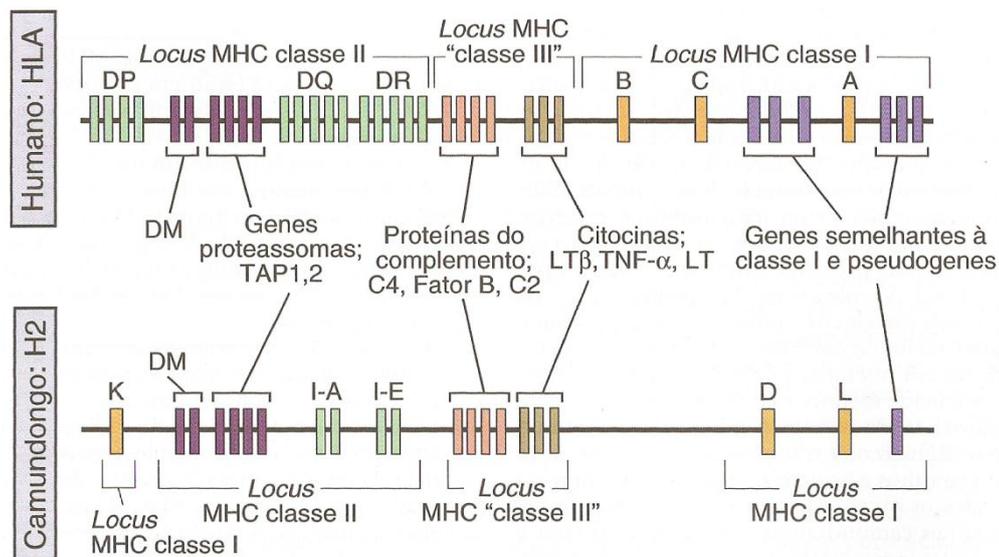


Figura 2. Mapas esquemáticos dos *loci* do MHC humano e murino. Os genes semelhantes à classe I no *locus* do MHC de classe I humano, se referem aos chamados HLA não clássicos, como o HLA-G, os quais não estão diretamente envolvidos com a apresentação de peptídeos endógenos, embora desempenhem outros papéis imunomodulatórios. Modificado de Imunologia Celular e Molecular, 5ª edição (Abbas & Lichtman, 2005).

A região do MHC pode ser subdividida em três *loci* (Figura 2). Um deles, o *locus* do MHC de classe III, codifica citocinas e proteínas do Sistema Complemento, enquanto os outros dois codificam proteínas envolvidas na apresentação de peptídeos (Horton *et al.*, 2004). As moléculas codificadas pelos *loci* DP, DQ e DR, no *locus* do MHC de classe II,

estão envolvidas na via de apresentação de peptídeos exógenos, a qual é fundamental para o desencadeamento da resposta imunológica humoral. Neste trabalho, enfocaremos as moléculas codificadas pelo *locus* do MHC de classe I, o mais polimórfico dentro da região do MHC. Destacam-se neste *locus* os genes HLA-A, HLA-B e HLA-C, os quais codificam a cadeia pesada do complexo responsável pela apresentação de peptídeos endógenos na superfície das células (Kelley et al., 2005). Em humanos, estas moléculas são referidas como antígeno leucocitário humano ou simplesmente HLA (do inglês *Human Leukocyte Antigen*), recebendo outras denominações dependendo da espécie. Em camundongos, por exemplo, são referidas como Antígeno H2 (H2-K, H2-D e H2-L). De maneira genérica, podemos nos referir a todas estas proteínas apresentadoras como moléculas de MHC (de classe I ou de classe II).

Atualmente, são descritos 8.976 alelos distintos para os HLAs de classe I em humanos, segundo dados de Julho de 2014 (hla.alleles.org). O mais polimórfico é o HLA-B (3.590), seguido por HLA-A (2.884) e HLA-C (2.375). Os demais são dados pela soma dos chamados HLA de classe I não clássicos (Figura 2), os quais são muito menos polimórficos. Esta expressiva variabilidade de genes codificando proteínas apresentadoras de antígeno confere à espécie uma enorme vantagem em termos populacionais, maximizando a capacidade de reconhecimento de agentes infecciosos por pelo menos uma parcela dos indivíduos (Vandiedonck & Knight, 2009a). A distribuição destes alelos, no entanto, não é homogênea. O alelo HLA-A*02:01, por exemplo, é referido como sendo o mais frequente na população humana (<http://www.allelefrequencies.net/>). A herança dos alelos de MHC de classe I clássicos é realizada em haplótipos (Zuniga *et al.*, 2013), de modo que cada indivíduo pode apresentar até seis alelos distintos (três herdados do pai e três herdados na mãe). Conforme mencionado anteriormente, cada um destes alelos codifica uma variante da cadeia pesada (cadeia alfa) do complexo responsável pela apresentação de peptídeos endógenos, sendo estas variantes proteicas referidas como alotipos (Bordner & Abagyan, 2006). Os alotipos de MHC são estáveis na superfície celular apenas na forma de um heterotrímero, formado em conjunto com uma cadeia constante (β 2-microglobulina) e um peptídeo de origem intracelular (Figura 3). Este heterotrímero será doravante referido como complexo peptídeo:MHC, ou pMHC.

Em linhas gerais, esta via permite que peptídeos derivados de proteínas citosólicas (também referidos como epítopos) sejam apresentados na superfície celular para reconhecimento pelos linfócitos T citotóxicos (CTLs, do inglês *Cytotoxic T Lymphocyte*). Tendo em vista a capacidade dos CTLs de reconhecer peptídeos não próprios apresentados pelos MHCs do organismo, fruto do processo de seleção de linfócitos no timo, esta via desempenha papel central no controle de infecções virais constituindo a base da resposta imunológica celular (Yewdell *et al.*, 2003).

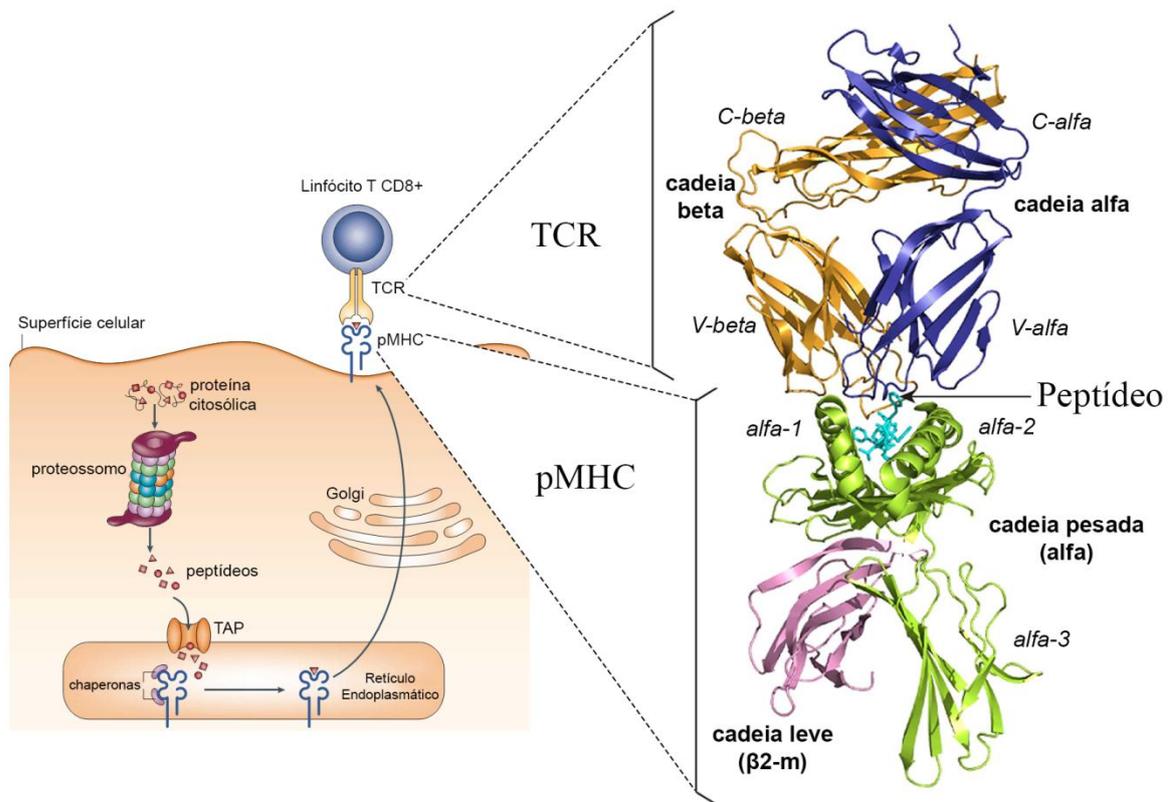


Figura 3. Estrutura e função do complexo pMHC. A esquerda, representação esquemática da via de apresentação de peptídeos endógenos. Uma amostra das proteínas citosólicas é digerida pelo proteossomo e os peptídeos derivados são ultimamente apresentados na superfície celular, complexados a molécula de MHC de classe I. Este complexo (pMHC) será reconhecido por linfócitos T citotóxicos (CD8+), os efetores da resposta imunológica celular. Modificado de Yewdell *et al.*, 2003. A direita, representação em *cartoon* da estrutura molecular de um complexo TCRpMHC (em destaque). Cada cor representa uma cadeia proteica distinta. Os nomes das cadeias (em negrito) e de seus domínios (em itálico) são indicados. Modificado de Lefranc MP, 2014. TAP, transportador associado à apresentação de antígenos; TCR, receptor de células T; CD8, glicoproteína que interage especificamente com o MHC de classe I, auxiliando a interação com o TCR; MHC, complexo principal de histocompatibilidade; pMHC, complexo peptídeo:MHC.

Recombinação somática e seleção tímica

A via de apresentação de peptídeos endógenos permite a identificação de células infectadas por vírus (ou outros parasitas intracelulares), uma vez que peptídeos derivados de proteínas do patógeno serão expostos na superfície celular, no contexto de um dos alotipos de MHC do hospedeiro. A eficácia deste sistema, no entanto, depende da capacidade de reconhecimento de pMHCs apresentando alvos não próprios. Esta tarefa é executada pelos linfócitos T CD8+ através da interação mediada pelo receptor de linfócitos T (TCR, do inglês *T Cell Receptor*). Mais do que isso, conforme discutido anteriormente, uma das características da resposta imunológica adaptativa é o reconhecimento específico de alvos patogênicos, capaz de conferir imunidade protetora contra futuras infecções pelo mesmo patógeno (memória imunológica). Considerando o grande polimorfismo dos MHCs e a expressiva variabilidade dos patógenos, que proporcionam em conjunto um número incalculável de complexos pMHC únicos, parece improvável a existência de um mecanismo capaz de gerar uma variabilidade de TCRs que se aproxime desta variabilidade de alvos. A constatação empírica de que os organismos hospedeiros são de fato capazes de gerar respostas específicas contra uma ampla gama de patógenos e o desconhecimento dos mecanismos geradores desta variabilidade de TCRs (e imunoglobulinas), intrigou os imunologistas por décadas. A resposta para este enigma envolvia um surpreendente mecanismo de recombinação somática, o qual foi revelado ao longo das décadas de 80 e 90 (Fanning *et al.*, 1998; Kurosawa *et al.*, 1981; Maki *et al.*, 1981; Schatz *et al.*, 1992).

A estrutura do TCR é um heterodímero formado pelas cadeias alfa e beta (Figura 3) ou, alternativamente, gama e delta. Cada uma destas cadeias apresenta um domínio C (*constant*) e um domínio V (*variable*) (Lefranc, 2014). No domínio variável, existem ainda três sítios hipervariáveis, os quais são referidos como regiões determinantes de complementariedade (CDRs, do inglês *Complementarity Determining Region*). Na estrutura tridimensional, estas CDRs compõem as alças que realizam o contato direto com resíduos do peptídeo e da molécula de MHC (Brehm *et al.*, 2004). Diferentemente do domínio C, o domínio V é codificado por uma combinação de diferentes segmentos gênicos. A porção N-terminal do domínio V contém informação de um segmento J

(*joining*) e de um segmento D (*diversity*), este último estando presente apenas na cadeia beta (Figura 4).

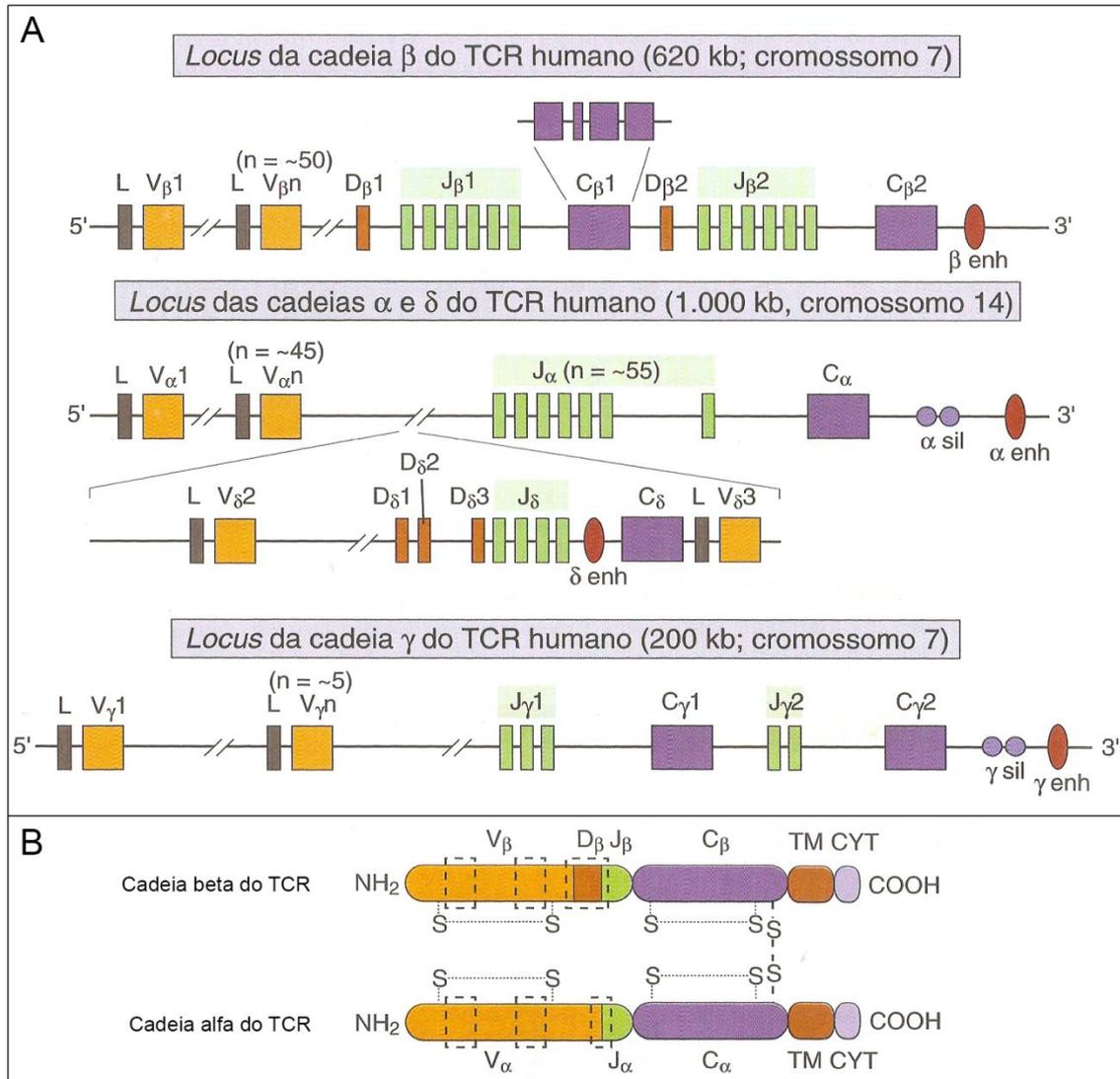


Figura 4. Organização do TCR. A. Organização dos *loci* das cadeias α , β , γ e δ do TCR humano, na linhagem germinativa. Éxons e íntrons não estão representados em escala e pseudo-genes não são indicados. Os genes constantes (C) são representados como blocos simples, mas são na verdade compostos por vários éxons (ex. $C\beta 1$). Os segmentos gênicos ilustrados são L, *leader*; V, *variable*; D, *diversity*; J, *joining*; C, *constant*, enh, *enhancer*; sil, *silencer*. B. Domínios das proteínas α e β do TCR. A localização das cadeias internas e das pontes dissulfeto (S-S) é representada de maneira aproximada. As áreas pontilhadas (retângulos) indicam as regiões hipervariáveis (CDRs). As cadeias α e β , assim como os domínios TM (*transmembrane*) e CYT (*cytoplasmic*), são codificados por diferentes éxons. Modificado de *Imunologia Celular e Molecular*, 5ª edição (Abbas & Lichtman, 2005).

Os *loci* que codificam as cadeias do TCR apresentam uma estruturação característica, marcada pela repetição sequencial de diversos segmentos que codificam para os domínios V, D e J (Figura 4). Durante o processo de maturação na medula, cada linfócito T sofrerá um rearranjo único destes segmentos somáticos, gerando um TCR com uma combinação “V-D-J” específica (Schatz *et al.*, 1992). Adicionalmente, existe um mecanismo de edição de nucleotídeos nas junções destes segmentos V-D-J (adição e remoção ao acaso), aumentando ainda mais a variabilidade do TCR produzido (Fanning *et al.*, 1998). Em conjunto, estes mecanismos fornecem ao TCR uma diversidade potencial que supera 10^{20} combinações (Zarnitsyna *et al.*, 2013).

Tendo em vista a natureza aleatória dos mecanismos de edição, que e em dois terços dos casos poderá resultar em troca da fase de leitura do DNA, muitos dos rearranjos gerados não serão bem sucedidos na produção de uma cadeia de TCR (Elhanati *et al.*, 2014). Além disso, diferentemente das Imunoglobulinas, os TCRs reconhecem apenas antígenos apresentados no contexto de moléculas de MHC. Assim, um novo linfócito T só será útil para o organismo caso seu TCR seja capaz de interagir com peptídeos apresentados pelos alotipos de MHC do indivíduo. Para assegurar esta especificidade, linfócitos T passam por um rigoroso processo de seleção no timo, onde células epiteliais (tímicas) realizam a apresentação de uma ampla gama de peptídeos próprios do contexto dos MHCs do indivíduo. Através da competição por estímulos, linfócitos não responsivos, apresentando TCRs defectivos ou incapazes de interagir com os MHCs próprios, são negligenciados e morrem. Linfócitos altamente autoreativos são negativamente selecionados, uma vez que estão apresentando alta avides/afinidade por peptídeos próprios. Deste modo, sobrevive a seleção tímica uma população de linfócitos cujo TCR consegue interagir com os MHCs do indivíduo, mas possui apenas moderada ou baixa afinidade por peptídeos próprios (Sohn *et al.*, 2007). Ao serem liberadas na circulação, estas células serão expostas a diversos antígenos próprios e não-próprios (sempre no contexto do MHC), sendo ativadas apenas frente a um pMHC para o qual elas apresentem alta avides/afinidade (em tese, apresentando um peptídeo não-próprio).

Características e consequências da interação TCRpMHC

Diversas outras moléculas auxiliam na formação do complexo TCRpMHC, estabilizando e prolongando a interação a ponto de permitir a estimulação do linfócito T (Chen et al., 2009; Rudolph et al., 2006). Uma das mais importantes é a glicoproteína CD8 (do inglês *cluster of differentiation 8*), que se liga ao domínio alfa-3 da molécula de MHC de classe I (Chen *et al.*, 2009). Cabe ainda salientar que a estimulação do linfócito não é desencadeada pela interação de um único TCR com um único pMHC. Durante a interação entre o linfócito e a célula apresentadora, ocorre um rearranjo nas membranas plasmáticas de ambas as células a fim de concentrar um grande número de complexos TCRpMHC e uma diversidade de correceptores e outras moléculas de adesão, em uma área restrita de superfície celular. Esta intensa interação entre as células foi referida como sinapse imunológica (Saito *et al.*, 2010; Thauland & Parker, 2010).

A interação TCRpMHC apresenta uma fina regulação, podendo desencadear diferentes níveis de estimulação do linfócito (van der Merwe & Dushek, 2010). Conforme discutido anteriormente, cada linfócito T produzido na medula possui um TCR específico. Um dos desfechos possíveis da interação TCRpMHC é a chamada expansão clonal, na qual a célula mãe dará origem a uma população de células filhas que compartilharão o mesmo rearranjo V-D-J da célula original. Este mecanismo assegura que, uma vez identificado um alvo não-próprio, sejam geradas várias células (clones) capazes de percorrer o organismo eliminando células infectadas. Mais do que isso, ele permite que uma parcela destes clones seja diferenciada em células de memória, as quais permanecerão em nichos específicos para serem futuramente ativadas em caso de um novo contato com o mesmo patógeno (Seder *et al.*, 2008).

Durante a imunização de um indivíduo, seja por infecção ou vacinação, diversos linfócitos serão simultaneamente ativados. Cada um deles poderá responder a um epitopo distinto, ou alguns deles poderão responder de forma diferencial a um mesmo epitopo. Estes diferentes linfócitos poderão expandir clonalmente, gerando uma população heterogênea de células. Assim, a resposta primária a uma imunização é policlonal (Cornberg *et al.*, 2006). Além disso, um mesmo imunógeno irá estimular conjuntos diferentes de clones em cada indivíduo (diferentes recombinações V-D-J), um

fenômeno referido como repertório privado (*private specificity*). A existência dos repertórios privados tem papel central na heterogeneidade de resposta entre indivíduos. Entretanto, já foram descritos alguns rearranjos V-D-J que parecem ser mais frequentes na população (ainda que não apresentem sequências idênticas), sendo chamados de TCRs públicos (Elhanati *et al.*, 2014).

A especificidade do contato TCRpMHC é uma característica fundamental, que permite a geração de memória imunológica. No entanto, considerando o processo de geração dos linfócitos, podemos observar que a mesma célula que reconheceu fracamente um complexo pMHC apresentando um peptídeo próprio (no timo), reconheceu fortemente outro complexo pMHC (não relacionado), apresentando um peptídeo não-próprio. Esta situação caracteriza uma propriedade intrínseca dos linfócitos T, referida como poli-especificidade (Wucherpfennig *et al.*, 2007). O reconhecimento TCRpMHC não é degenerado, mas um mesmo TCR pode interagir (especificamente) com diferentes complexos pMHC. Em alguns casos, complexos pMHC apresentando alvos heterólogos poderão desencadear de maneira similar a ativação de uma mesma população de linfócitos, um evento referido como reatividade cruzada (Welsh & Selin, 2002).

A reatividade cruzada, de certo modo, se contrapõe a especificidade da interação TCRpMHC, o que pode acarretar diversas consequências para o organismo. Por um lado, ela maximiza a capacidade de resposta de um dado conjunto de linfócitos, permitindo que um mesmo clone responda contra diferentes alvos (Welsh *et al.*, 2010). Por outro lado, ela pode acarretar respostas contra alvos próprios (autoimunidade) e mediar respostas ineficientes ou crônicas, que não eliminam o alvo e acabam lesando o organismo (imunopatologia) (Welsh & Fujinami, 2007; Wlodarczyk *et al.*, 2013).

Estudos recentes sugerem que semelhanças estruturais entre complexos pMHC apresentando alvos heterólogos sejam um fator decisivo em eventos de reatividade cruzada (Birnbbaum *et al.*, 2014; Shen *et al.*, 2013). A interação TCRpMHC obedece uma geometria relativamente restrita, na qual as cadeias V-alfa e V-beta do TCR irão normalmente interagir com regiões determinadas da superfície do pMHC (Adams *et al.*, 2011; Garcia *et al.*, 2009; Gras *et al.*, 2012). A “face” do pMHC que interage com o TCR é

delimitada pelos domínios alfa-1 e alfa-2, formando uma superfície única em conjunto com o epitopo (Figura 5). Os padrões típicos de interação entre diferentes TCRs e pMHCs podem ser mapeados sobre a “face” do pMHC, sendo referidos como “impressões digitais dos TCRs” (*TCR footprints*) (Rudolph *et al.*, 2006).

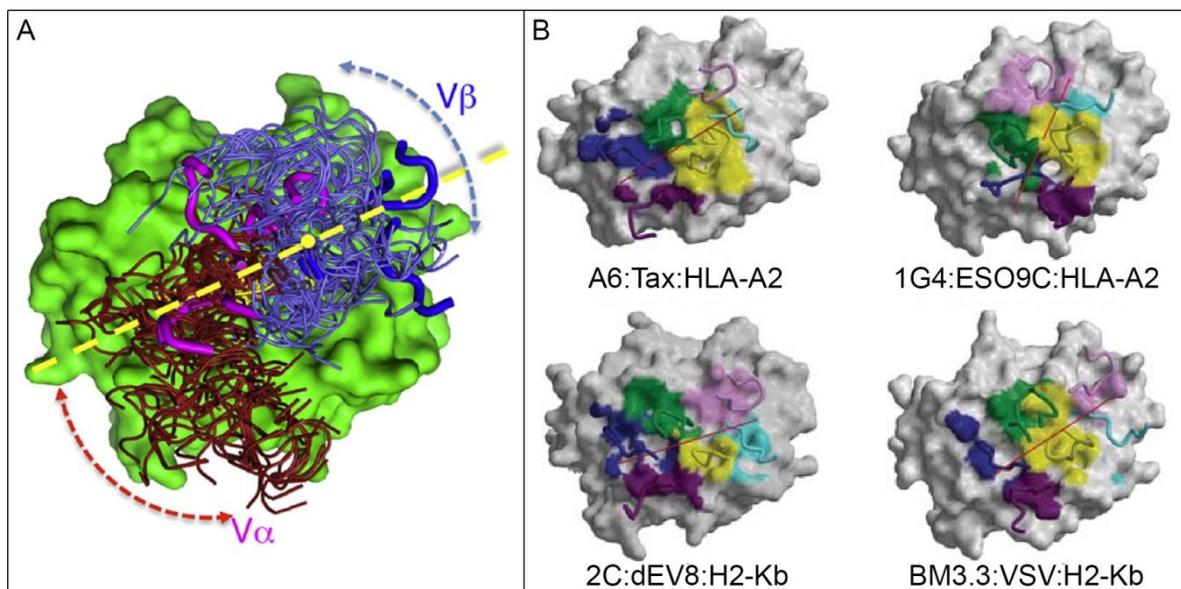


Figura 5. Orientação da interação TCRpMHC. A. Espectro de orientações de ancoramento para 40 TCRs agonistas, alinhados sobre a “face” de contato de um pMHC. As alças de um TCR não agonista (p3A1) são representadas com linhas mais grossas em azul e roxo, com a linha tracejada em amarelo indicando sua orientação. Modificado de Adams *et al.* 2011. B. *Footprints* de diferentes TCRs com epitopos restritos a HLA-A2 e H2-K^b. A identificação das proteínas envolvidas é fornecida abaixo de cada imagem, seguindo o padrão TCR:peptídeo:MHC. Resíduos não contactados pelas alças do TCR são representados em cinza na superfície do pMHC. Demais resíduos são coloridos de acordo com a alça do TCR contactada, seguindo o esquema: CDR1 α , azul escuro; CDR2 α , magenta; CDR3 α , verde; CDR1 β , ciano; CDR2 β , rosa; CDR3 β , amarelo. A linha vermelha indica a orientação do ancoramento do TCR. Modificado de Rudolph *et al.*, 2006.

Genética, Imunologia e Bioinformática

A interdisciplinaridade é fundamental para a resolução de problemas complexos. Por exemplo, a descoberta dos mecanismos envolvidos na geração de variabilidade dos TCRs e imunoglobulinas representa uma das maiores conquistas da imunologia, a qual só foi possível graças ao desenvolvimento da genética e de ferramentas de biologia molecular. Reciprocamente, a motivação imunológica levou pesquisadores a desvendarem mecanismos genéticos até então desconhecidos. Esta interface entre

genética e imunologia tem sido extremamente frutífera e próspera, sendo reconhecida como uma disciplina independente, a imunogenética.

O desenvolvimento do Projeto Genoma Humano nos anos 90 reuniu pesquisadores de diversas áreas e financiamento de múltiplas fontes, visando um objetivo comum. Esta iniciativa grandiosa acabou acelerando o desenvolvimento da ciência, com impactos diretos em diversas áreas (Lander, 2011). Uma das necessidades básicas do projeto dizia respeito ao desenvolvimento de ferramentas de informática, para armazenar, processar e compartilhar uma quantidade de dados biológicos sem precedentes. Desenvolveu-se assim uma área batizada como bioinformática (Ikekawa & Ikekawa, 2001). Surgiram diversos bancos de dados e ferramentas específicas para se trabalhar com sequências biológicas, marcando o início do período das “ômicas” (genômica, proteômica, transcriptômica, etc). Com o tempo, o interesse no estudo da estrutura e função das proteínas ganhou força, amparado por avanços no campo da cristalografia de raios X e na rápida evolução da capacidade de processamento de dados. Desenvolveu-se assim a bioinformática estrutural, com uma série de ferramentas e desafios próprios (Sliwoski *et al.*, 2013).

A imunologia, no entanto, parece insistir em nos surpreender com sua complexidade e a peculiaridade de seus mecanismos. Os bancos de dados desenvolvidos para armazenar e organizar a informação de genes e proteínas envolvidas em outros sistemas, aparentemente não forneciam uma estrutura compatível com as definições e características da imunogenética. Desta necessidade de desenvolver ferramentas de bioinformática voltadas ao armazenamento e processamento de informações da imunogenética, surge a imunoinformática (Korber *et al.*, 2006; Lefranc, 2014; Tomar & De, 2010).

Sendo um tema central para a imunologia, a vacinologia sempre esteve presente na pesquisa em imunoinformática. O termo "vacinologia reversa" foi cunhado em 2000 pelo italiano Rino Rappuoli (Rappuoli, 2000). Ele se refere a um processo de desenvolvimento de vacinas que se inicia pela análise de sequências genômicas *in silico*, seguido pela identificação de alvos e posterior expressão de proteínas recombinantes para testes *in vivo*. Esta estratégia continua sendo discutida e aplicada, acumulando

diversos resultados interessantes na imunização contra vírus e bactérias (Donati & Rappuoli, 2013; Vivona *et al.*, 2008). Recentemente, começa a popularizar-se o termo "vacinologia computacional", que também tem sido discutida como uma alternativa na pesquisa em câncer (Pappalardo *et al.*, 2013). Apesar do termo utilizado, o fato é que o uso de ferramentas computacionais para o desenvolvimento de vacinas é uma realidade que vem ganhando destaque.

Estudos baseados em cristalografia de raios X nos forneceram pistas imprescindíveis sobre mecanismos imunológicos complexos, como a reatividade cruzada e o impacto de mutações sobre a diversidade do repertório de células T (Sandalova *et al.*, 2005; Shen *et al.*, 2013; Turner *et al.*, 2005). No entanto, os elevados custos, o tempo necessário e a dificuldade para se obter um cristal de complexos proteicos com baixa afinidade (como alguns complexos TCRpMHC), ainda tem limitado a popularização do uso desta ferramenta. Alternativamente, a bioinformática estrutural fornece muitas ferramentas rápidas e de baixo custo, que podem ser empregadas para o estudo das interações peptídeo:MHC e TCR:peptídeo:MHC. Tais ferramentas podem nos ajudar a analisar estruturas previamente cristalografadas ou até mesmo modelar estruturas ainda não determinadas, fornecendo assim um valioso subsídio para a formulação de hipóteses, as quais podem ser posteriormente testadas *in vitro* ou *in vivo*.

Objetivos

O objetivo do presente trabalho foi identificar características estruturais dos complexos peptídeo:MHC que possam estar envolvidas na estimulação da reatividade cruzada de linfócitos T, utilizando ferramentas de bioinformática estrutural e desenvolvendo novas estratégias que permitam realizar predições acerca destes eventos.

Capítulo II

Structural Immunoinformatics and Vaccine Development

(Capítulo publicado no livro “*Bioinformatics Research: New Developments*”)

Neste capítulo aprofundaremos a discussão sobre a possível aplicação de ferramentas de bioinformática para a pesquisa em imunologia, com especial interesse no desenvolvimento de vacinas. Apresentaremos alguns dos principais bancos de dados voltados a pesquisa em imunoinformática estrutural, com enfoque na interação entre TCR e pMHC.

A análise estrutural de complexos pMHC revela aspectos moleculares importantes no desfecho de fenômenos imunológicos complexos, mas os métodos experimentais para obtenção destas estruturas apresentam limitações quanto ao uso em larga escala. Serão apresentadas algumas alternativas para a modelagem e predição estrutural de complexos pMHC, utilizando ferramentas de bioinformática. Por exemplo, a grande conservação estrutural da molécula de MHC permite a modelagem por homologia de alotipos cuja estrutura 3D ainda não foi determinada. Adicionalmente, a conformação de um dado complexo pMHC, apresentando um epitopo de interesse, pode ser predita com o uso de ferramentas como o ancoramento molecular (*docking*).

Finalmente, revisaremos alguns resultados de nosso grupo referentes a predição de reatividade cruzada entre complexos pMHC. Uma abordagem inovadora, desenvolvida pela nossa equipe, utiliza o potencial eletrostático dos complexos pMHC para agrupar alvos estruturalmente semelhantes, através da aplicação de métodos estatísticos multivariados.

Capítulo III

Abundance and privacy of CD8+ HCV-specific T-cells in seronegatives: implication for vaccine response

(Artigo completo submetido a revista PLoS Pathogens)

Buscando alvos que apresentassem similaridade estrutural com o epitopo imunodominante de HCV NS3-1073, nosso grupo sugeriu a possível reatividade cruzada deste peptídeo com um epitopo do vírus Epstein-Barr (EBV). Neste capítulo apresentaremos os resultados de uma colaboração estabelecida entre o Núcleo de Bioinformática do Laboratório de Imunogenética (NBLI) e a equipe coordenada pelo Prof. Dr. Markus Cornberg, da *Medizinische Hochschule Hannover* (MHH, na sigla em alemão para Escola de Medicina de Hannover). A equipe do MHH detectou a presença de células T específicas para este mesmo alvo de HCV (NS3-1073) em 32,6% dos 46 indivíduos analisados, sendo estes doadores de sangue saudáveis, HCV-, sem histórico de exposição ao HCV. Sugere-se que estas células tenham sido expandidas pela exposição prévia destes indivíduos a algum alvo heterólogo (outro vírus), respondendo *in vitro* contra o epitopo de HCV por um mecanismo de reatividade cruzada de células T.

Para confirmar nossos resultados prévios e prospectar outros alvos que pudessem ser testados *in vitro*, foi realizada uma nova análise hierárquica de agrupamentos baseada nas estruturas de 9 epitopos virais, no contexto do alotipo de MHC humano HLA-A*02:01. Esta análise corroborou os resultados prévios, apontando os epitopos LMP2-329 (EBV) e GAG-77 (HIV) como alvos estruturalmente similares ao epitopo de HCV. Outro epitopo de EBV e um epitopo de *Influenza* ficaram em ramos próximos, enquanto três candidatos derivados de EBV não apresentaram similaridade estrutural.

Testes *in vitro* confirmaram a reatividade cruzada entre os alvos preditos e indicaram que esta resposta heteróloga (prévia) altera o desfecho da vacinação contra HCV (com a vacina experimental IC41). Os resultados salientam que a identidade estrutural entre complexos pMHC parece ter maior impacto no desencadeamento da reatividade cruzada do que a similaridade de sequência entre os epitopos testados.

Capítulo IV

Peptide:MHC structural similarity as a probability for cross-reactive T cell responses

(Manuscrito em preparação)

Levantamentos anteriores sugeriram a existência de verdadeiras redes de reatividade cruzada (CRNs) entre epitopos virais. Neste capítulo, apresentaremos uma visão integrada sobre estas redes, tanto em humanos, quanto em murinos. O possível envolvimento de características estruturais de complexos pMHC no desencadeamento destes eventos de reatividade cruzada atraiu a atenção de cristalógrafos, de modo que a estrutura de alguns destes complexos já foi determinada experimentalmente.

Tendo em vista nossos resultados prévios, a análise hierárquica de agrupamentos (HCA) baseada em estrutura foi aplicada sobre estes cristais, fornecendo dendrogramas que se correlacionam com os dados experimentais. Um novo algoritmo foi utilizado para realizar o HCA, fornecendo uma estimativa da confiabilidade dos agrupamentos obtidos (com valores de *bootstrap* associados).

Mais do que uma nova validação acerca do potencial prospectivo desta abordagem de agrupamentos, nossos resultados salientam uma relação entre o grau de similaridade estrutural entre complexos pMHC e a probabilidade de se observar eventos de reatividade cruzada utilizando-se diferentes populações de linfócitos T. A discussão integrada da análise estrutural de complexos restritos aos alotipos HLA-A*02:01, H2-D^b e H2-K^b nos sugere que características estruturais dos complexos pMHC podem ser responsáveis por determinar diversas características da resposta celular, como a clonalidade e a direcionalidade de eventos de reatividade cruzada.

Capítulo V

Automatização de processos em Imunoinformática

Interessados em padronizar e otimizar os processos envolvidos na predição e análise estrutural de complexos pMHC, nós desenvolvemos uma série de códigos (*scripts*) utilizando diferentes linguagens de programação. Neste capítulo iremos descrever as etapas envolvidas na execução da abordagem *D1-EM-D2* e como foi possível automatizar estes processos. A automatização leva a um ganho de desempenho e previne diversos erros que poderiam ocorrer em função de descuidos por parte do usuário. Ela também permite aplicar a técnica para um conjunto maior de alvos, algo extremamente custoso utilizando-se o procedimento manual.

Diferentes programas são utilizados neste fluxograma (*pipeline*) para a predição estrutural de complexos pMHC. O papel dos *scripts* é fazer a ligação entre as etapas, convertendo arquivos e transmitindo parâmetros necessários para a execução dos programas. A automatização também permite um maior controle de qualidade sobre o processo (rastreabilidade).

Finalmente, a automatização abre o caminho para a possível disponibilização desta abordagem como uma ferramenta *online*, permitindo sua utilização por usuários que não estão familiarizados com a instalação e a execução dessas ferramentas no Linux. Discutiremos avanços feitos pela nossa equipe visando oferecer este serviço para livre utilização pela comunidade científica. Também serão apresentados alguns avanços no sentido de aprimorar a nossa metodologia de agrupamento estrutural de complexos pMHC.

Automatização da abordagem D1-EM-D2

Conforme apresentado no capítulo II, a variabilidade dos complexos pMHC supera em muito nossa capacidade de resolver estruturas por métodos experimentais, como cristalografia de raios X e ressonância magnética nuclear. Além dos custos e do tempo necessário para se resolver uma estrutura proteica, novas variantes virais surgem a cada dia, gerando novos epitopos. Assim, a “modelagem” ou predição conformacional de complexos pMHC é um objetivo importante e atual dentro da imunoinformática estrutural (Bordner, 2013; Dhanik *et al.*, 2013; Khan & Ranganathan, 2010).

Nos trabalhos aqui apresentados (capítulos II, III e IV), utilizamos uma estratégia própria para a predição estrutural de complexos pMHC (“*Docking 1 – Energy Minimization – Docking 2*” ou simplesmente *D1-EM-D2*). Esta abordagem foi inicialmente desenvolvida durante meu trabalho de conclusão de curso em Biomedicina (Antunes, 2008), baseada no uso de ancoramento molecular (*docking*) e minimização de energia (Figura 6A). A maioria dos programas de *docking* trabalha bem com até 10 ligações flexíveis, “limite” acima do qual o custo computacional aumenta e a precisão dos resultados diminui consideravelmente (Dhanik *et al.*, 2013; Plewczyński *et al.*, 2011). No entanto, um epitopo típico apresentado por MHC de classe I possui cerca de 9 aminoácidos, podendo apresentar longas cadeias laterais e totalizando mais de 40 ligações flexíveis. Conforme descrito no capítulo II, o nosso grupo identificou padrões conformacionais compartilhados por epitopos no contexto de um mesmo alotipo de MHC. O uso destes padrões permitia manter rígida a cadeia principal do peptídeo durante o *docking*, empregando o programa apenas para resolver a conformação das cadeias laterais do ligante. Esta foi a premissa para o desenvolvimento de uma nova abordagem de predição de complexos pMHC (Antunes *et al.*, 2010).

A primeira etapa deste processo envolve gerar um arquivo pdb com as coordenadas do epitopo, baseado simplesmente em sua sequência linear de aminoácidos (formato FASTA). Outro arquivo pdb, com a estrutura de um epitopo apresentado pelo mesmo MHC, é utilizado como molde nesta etapa. O objetivo é obter uma estrutura 3D do ligante, que possa ser utilizada na etapa de ancoramento molecular. Paralelamente, uma estrutura do MHC de interesse deve ser preparada para servir de receptor. Em cada

simulação de ancoramento molecular, uma conformação inicial aleatória é atribuída ao ligante, dentro de uma região de interesse (*GRID box*) definida pelo usuário (em nosso caso, incluindo a fenda do MHC). A conformação inicial do ligante será refinada por etapas iterativas de mudança conformacional e medidas de afinidade de interação (utilizando um algoritmo genético Lamarckiano). Assim, em duas simulações realizadas com o mesmo par “ligante/receptor”, caminhos distintos serão percorridos pelo algoritmo, podendo resultar em conformações finais distintas (diferentes mínimos locais de energia). Para se garantir que todos os mínimos energéticos sejam amostrados, nosso protocolo repete a etapa de ancoramento molecular 20 vezes com o programa Autodock Vina (Trott *et al.*, 2010), gerando uma população final de 1000 conformações distintas. A seguir, é preciso escolher o melhor resultado, com base na energia de ligação (*binding energy*) e na frequência dos mínimos amostrados.

Uma vez obtido o novo complexo pMHC, com o epitopo de interesse, realiza-se uma etapa de refinamento através da minimização de energia. Sobretudo, isso permite ajustar as cadeias laterais do receptor, que por limitações computacionais também foram mantidas rígidas durante o *docking*. Após esta etapa, uma nova rodada de ancoramento molecular é realizada, permitindo que o programa explore o sítio refinado e encontre resultados ainda melhores (Figura 6A).

Todas as etapas descritas acima eram realizadas manualmente, desde a leitura do arquivo FASTA e geração de uma estrutura inicial para o ligante, até a escolha do melhor resultado gerado pelo *docking*. Muitas etapas também envolvem a conversão de arquivos, e a etapa de minimização de energia envolve a execução de pelo menos seis programas distintos. Apesar das dificuldades técnicas, esta abordagem (manual) foi validada através da reprodução de 46 estruturas cristalográficas (RMSD médio de 1,754 Å para todos os átomos do epitopo) (Antunes *et al.*, 2010) e utilizada para a predição dos 55 complexos posteriormente incluídos no HCA descrito no capítulo II (Antunes *et al.*, 2011).

Tendo em vista nosso interesse em utilizar esta estratégia para a triagem virtual de complexos pMHC e o objetivo de disponibilizar as estruturas preditas através do banco de dados CrossTope (Sinigaglia *et al.*, 2013), a automatização e a padronização das etapas envolvidas na *D1-EM-D2* foram metas paralelas durante a execução do presente projeto.

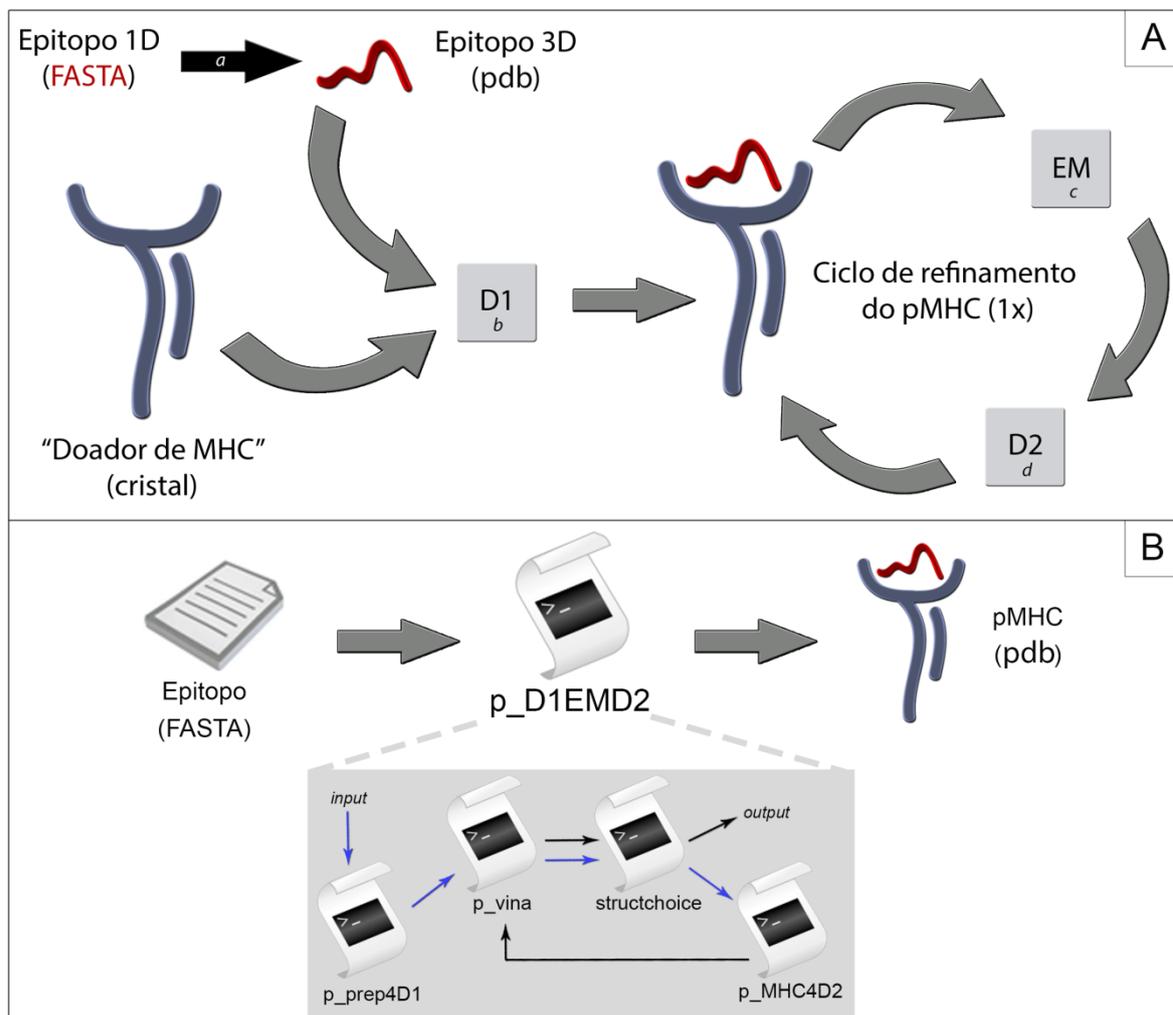


Figura 6. Abordagem *D1-EM-D2*. A. A estratégia consiste em 4 etapas principais, nas quais são utilizados os programas PyMOL (*a*), Autodock Vina (*b,d*) e o pacote de programas Gromacs (*c*). Um cristal de referência para o alotipo de interesse, previamente preparado, é utilizado como “Doador de MHC”. Um epitopo cristalografado no contexto deste MHC é utilizado como “padrão do peptídeo” (molde), para gerar uma estrutura 3D do epitopo de interesse. Após a obtenção de um complexo pMHC, gerado por *docking* (D1), realiza-se um ciclo de refinamento envolvendo uma etapa de minimização de energia (EM) e uma segunda rodada de ancoramento molecular (D2). Modificado de Sigaglia *et al.*, 2013. B. A *D1-EM-D2* foi completamente automatizada através do uso de *scripts*. Em um computador previamente configurado, basta fornecer um arquivo de texto com a sequência do epitopo no formato FASTA e executar o comando “*p_D1EMD2*”, para obter-se uma estrutura 3D (formato pdb) do pMHC de interesse. Este comando desencadeia a execução de *scripts* secundários (quadro cinza) que realizam as etapas sequenciais da abordagem. Figura em preto e branco na versão impressa.

Nos últimos três anos, uma série de código (*scripts*) foi desenvolvida para automatizar etapas deste processo (utilizando a distribuição Ubuntu do Linux). Uma das primeiras e mais importantes automatizações foi referente a escolha do melhor resultado

do *docking*. Além de agilizar o processo, a automatização elimina possíveis vieses advindos da interpretação do usuário. As etapas envolvidas na minimização de energia também foram rapidamente automatizadas com o uso de *scripts*, bem como as etapas de conversão de arquivos. Neste sentido cabe salientar o uso de alguns *scripts* em *python* desenvolvidos e distribuídos pelo *Molecular Graphics Laboratory* do *The Scripps Research Institute* (<http://mglttools.scripps.edu/>). Aos poucos, as funções destes vários *scripts* independentes foram sendo agregadas em fluxogramas maiores (*pipelines*), até alcançarmos a completa automatização da abordagem *D1-EM-D2* (Figura 6B).

Atualmente é possível obter-se a estrutura 3D de um dado complexo pMHC fornecendo-se apenas a sequência do epitopo (formato FASTA) e executando-se apenas um *script* principal (*p_D1EMD2*). Este comando irá orquestrar a execução serial de outros 4 *scripts* secundários, os quais por sua vez executarão de forma automatizada todos os demais *scripts* e programas envolvidos no processo. Em conjunto, este *pipeline* integra 9 *scripts* em *shell*, 13 *scripts* em *python*, 7 executáveis em *c++* e 2 executáveis em *python*. Evidentemente, este procedimento precisa ser executado em um computador no qual todos os *scripts* e programas tenham sido previamente instalados, o que também pode ser realizado de forma automatizada utilizando-se pacotes desenvolvidos pela nossa equipe (para instalação na distribuição Ubuntu do sistema operacional Linux).

As principais funções dos 4 *scripts* secundários serão apresentadas abaixo:

- **p_prep4D1**: Executa a *pipeline* de preparação para o *docking* 1 (D1). Ele recebe como parâmetros um arquivo com a sequência do epitopo (*.fasta) e o nome do alelo de interesse. Atualmente nossa estratégia permite a modelagem de 4 alelos de MHC, sendo dois humanos (HLA-A*02:01 e HLA-B*27:05) e dois murinos (H2-D^b e H2-K^b). A identificação do MHC de interesse permite importar para o diretório de trabalho o “Doador de MHC” e seu correspondente arquivo de configuração para o *docking* (com as coordenadas e dimensões do *GRID*). Também permite selecionar o padrão de cadeia principal que será utilizado como molde pelo PyMOL para a geração da estrutura 3D inicial do peptídeo. Após a geração do pdb do ligante, uma etapa de minimização de energia é realizada para acomodar

possíveis conflitos entre a conformação “padrão” adotada para a cadeia principal e a distribuição de cadeias laterais específicas da sequência de interesse. Após esta etapa, realizada com o pacote Gromacs 4.5.1 (Pronk *et al.*, 2013), o arquivo pdb do epitopo minimizado é convertido em um arquivo PDBQT (formato de entrada para o *docking*).

- **p_vina:** Este *script* consiste em um laço que permite a execução sequencial de 20 simulações de ancoramento molecular, realizadas pelo Autodock Vina (Trott *et al.*, 2010), utilizando sempre os mesmos arquivos de entrada (receptor e ligante). O mesmo *script* é utilizado tanto no D1 quanto no D2, mas os arquivos de entrada não serão os mesmos. As saídas são numeradas sequencialmente, facilitando a análise dos resultados. Tendo em vista que o ancoramento molecular representa a etapa com maior custo computacional na *D1-EM-D2*, o *script* também realiza o monitoramento das atividades em tempo real. Em caso de erro na primeira rodada do laço, o *script* é interrompido e o problema é reportado. Caso contrário, o *script* continua reportando os resultados ao final de cada *docking* (lista das energias de ligação obtidas e localização dos arquivos de saída).
- **structchoice:** Este *script* permite escolher o melhor resultado dentre as múltiplas conformações produzidas pelo *docking*, gerando um arquivo pdb com as coordenadas do complexo pMHC. A nomenclatura do complexo gerado varia dependendo se os resultados analisados se referem ao D1 ou ao D2, o que deve ser informado por parâmetro juntamente com a identificação do MHC utilizado. Cada simulação de ancoramento molecular gera um arquivo PDBQT contendo as múltiplas conformações obtidas (até o máximo teórico de 50 estruturas). O *structchoice* separa estas diferentes conformações em arquivos PDBQT independentes, utilizando o executável *vina_split*, que é distribuído em conjunto com o Autodock Vina (Trott *et al.*, 2010). O melhor resultado de cada uma das 20 simulações realizadas é convertido de volta para o formato pdb. Através do uso de comandos em

linguagem *Perl* (*sed*, *grep* e *awk*), os arquivos com os logs destes 20 melhores resultados são analisados, e os alvos são ordenados de acordo com suas energias de ligação ao MHC. Alvos com energia superior a média calculada são excluídos (quanto mais alta a energia de ligação em kcal/mol, mais fraca a interação). As estruturas dos demais são importadas para uma rodada de cálculos de RMSD (*Root Mean Square Deviation*) com o programa *g_confrms* do pacote Gromacs 4.5.1 (Pronk *et al.*, 2013). Deste modo, será escolhido como melhor resultado aquela conformação que apresenta o menor RMSD em relação às demais conformações (estrutura “média”, mais frequente), dentre aquelas que apresentaram as menores energias observadas nas 20 rodadas. O *pdb* do ligante escolhido, juntamente com o *pdb* do “doador de MHC”, são combinados para gerar o *pdb* do novo pMHC. Finalmente, uma etapa de verificação da posição do epítipo é realizada com o *script RMSCheck*, o qual sobrepõe a estrutura do pMHC modelado com a estrutura de um pMHC referência, calculando o RMSD entre os epítopos (para carbono alfa). Esta etapa, realizada tanto após o D1 quanto após o D2, assegura que o resultado obtido pelo *docking* esteja de acordo com os dados cristalográficos, no que se refere orientação e a localização do epítipo na fenda do MHC. Caso o valor de RMSD obtido nesta verificação supere um ponto de corte previamente definido, todo o *pipeline* do “p_D1EMD2” será interrompido e o erro será reportado.

- **p_MHC4D2:** O pMHC resultante do D1, após escolha e verificação pelo *structchoice*, deverá passar por uma minimização de energia e posteriormente ser preparado para o D2. Inicialmente o arquivo (*.pdb) do complexo pMHC é convertido para o formato do pacote Gromacs (*.gro) com o programa *pdb2gmx*. Este programa também gera arquivos de topologia da molécula. Uma caixa tridimensional é gerada ao entorno da molécula alvo, com programa *editconf*. Moléculas de água são adicionadas até completar o volume da caixa com programa *genbox*. Os arquivos de coordenadas (*.gro) e de topologia (*.top e *.itp) que descrevem este

sistema são combinados com um arquivo de parâmetros da simulação (*.mdp) através do programa *grompp*, gerando um arquivo único de entrada para a minimização de energia (*.tpr). A minimização de energia é realizada pelo programa *mdrun* e as coordenadas finais do pMHC são escritas novamente no formato pdb com o programa *trjconv*. Para permitir a reutilização do arquivo de configuração do Vina (com as coordenadas do *GRID* utilizado no D1), este pMHC é ajustado às coordenadas do “Doador de MHC”. Isso é obtido através do alinhamento estrutural com o programa PyMOL. Finalmente, peptídeo e MHC são separados em arquivos independentes e convertidos ao formato PDBQT (para utilização no D2).

Posteriormente, visando a predição de complexos pMHC em larga escala, foi desenvolvida uma versão modificada do *script* principal, batizada de “p_mD1EMD2”. Ao invés de receber como parâmetro um arquivo no formato FASTA (que permitiria a modelagem de um pMHC), este *script* lista todos os arquivos FASTA disponíveis em determinado diretório. Alternativamente, também pode ser fornecida uma tabela (CSV) contendo a lista dos epitopos de interesse (no formato “Nome, Sequência”). Primeiro, o *script* irá gerar arquivos FASTA correspondentes a cada um dos epitopos listados. Depois, realizará a *D1-EM-D2* para cada um destes alvos, organizando os resultados em subdiretórios. Um relatório dos processos é gerado em tempo real, atualizando uma lista que informa se houve sucesso (ou não) na modelagem de cada um dos complexos solicitados.

Revalidação da metodologia de predição de pMHCs

Esta completa automatização da abordagem *D1-EM-D2* também permitiu uma nova validação através da reprodução de cristais, utilizando um conjunto maior de complexos. Todos os cristais de complexos pMHC disponíveis no *Protein Data Bank* foram reproduzidos, excluindo-se complexos redundantes e estruturas em que havia interação com TCR, anticorpos ou outras moléculas (visto que estas interações externas podem interferir na conformação do epitopo). Nos casos de complexos redundantes, o RMSD do complexo modelado foi calculado em relação ao cristal de melhor resolução. Ao todo, 130

estruturas de pMHC foram reproduzidas apresentando um RMSD médio de 1,95 Å (\pm 0,63) para todos os átomos do peptídeo (Tabela 1). Estes dados confirmam a confiabilidade da técnica, uma vez que são usualmente consideradas reproduções válidas aquelas com um desvio do ligante igual ou inferior a 2,2 Å (Madurga *et al.*, 2005; Trott *et al.*, 2010).

Tabela 1. Reprodução de 130 cristais utilizando a abordagem *D1-EM-D2*.

Alotipo	C.Seq.	Nº de pMHCs	RMSD (α)		RMSD (total)	
			Média	DP	Média	DP
HLA-A*02:01	9	68	0,926	0,437	1,875	0,971
HLA-B*27:05	9	10	1,027	0,503	2,239	1,322
H2-D ^b	9	33	0,671	0,326	1,901	0,912
H2-K ^b	8	19	1,132	0,355	2,076	0,401
TOTAL		130	0,899	0,435	1,945	0,628

RMSD, *Root Mean Square Deviation*, C.Seq., comprimento da sequência do peptídeo; DP, desvio padrão; α , carbono alfa; total, todos os átomos do ligante.

Instalação em equipamentos de alto desempenho

Uma das vantagens da utilização do Autodock Vina para o cálculo de ancoramento molecular é sua eficiência. Graças a algumas implementações, como a paralelização de processos, este programa apresenta alto desempenho sem perder a acurácia (Chang *et al.*, 2010; Trott *et al.*, 2010). Ainda assim, em função do grande número de ligações flexíveis de muitos peptídeos e da necessidade de se repetir o cálculo diversas vezes para garantir uma amostragem representativa, as etapas de *docking* da nossa abordagem apresentam um elevado custo computacional. O tempo necessário para a predição de uma estrutura, a partir da sequência do peptídeo, é normalmente superior a 5 horas. Esta estimativa considera o uso de uma máquina com processador *quad-core* de alto desempenho (ex.: i7-920 ou superior).

Tendo conhecimento acerca da existência do Centro Nacional de Supercomputação (CESUP-UFRGS), que por sua vez integra o Sistema Nacional de Processamento de Alto Desempenho (SINAPAD), nosso grupo iniciou os trâmites para instalação dos *scripts* e programas necessários para a *D1-EM-D2* em um *cluster* de alto

desempenho. Apesar do *cluster* “Newton” (*Sun Fire*) do CESUP também utilizar o sistema operacional Linux, ele opera com a distribuição *Red Hat*. Além disso, o sistema possui um gerenciador de filas que distribui os mais de 122 núcleos de processamento entre as tarefas solicitadas pelos usuários (*jobs*). Assim, uma série de ajustes precisou ser realizada, tanto pela nossa equipe, quanto pela equipe do CESUP, para que a versão automatizada da abordagem *D1-EM-D2* pudesse ser executada no cluster “Newton”. Esta etapa de implementação foi recentemente concluída, diminuindo o tempo de execução da nossa *pipeline* e permitindo a submissão de vários *jobs* em paralelo.

Disponibilização de uma ferramenta baseada em *web*

Desde a publicação do banco de dados CrossTope, que disponibiliza estruturas modeladas pela técnica *D1-EM-D2*, nossa equipe tem interesse em também oferecer uma versão *online* deste método de predição. Em paralelo às etapas de instalação dos *scripts* no CESUP, nossa equipe também desenvolveu uma interface *web* que poderá ser utilizada para gerenciar clientes e disparar *jobs* (Figura 7). Após a conclusão das etapas de desenvolvimento da ferramenta e de instalação e verificação dos *scripts*/programas no CESUP, estamos atualmente trabalhando na comunicação entre estes dois sistemas.

Além de possuir um banco de clientes que permite o cadastro de usuários e o monitoramento dos *jobs* submetidos (para cada cliente), a ferramenta conta com uma série de controles para evitar a submissão de dados incorretos. Por exemplo, do campo onde o usuário informa a sequência do peptídeo de interesse, são permitidas apenas letras que representam aminoácidos. Além disso, a sequência deve conter entre 8 e 10 resíduos. À medida que a sequência é informada, são oferecidas as opções de MHC compatíveis com aquele tamanho de sequência. Por exemplo, atualmente a modelagem de complexos com H2-K^b só é possível para epitopos com 8 resíduos (mais frequente para este alelo), de modo que uma sequência com 9 resíduos não apresenta este alotipo de MHC como uma das alternativas (Figura 7).

A interface *web* já está instalada em um servidor dentro do CESUP, comunicando-se diretamente com o cluster “Newton” e permitindo a submissão de *jobs*. No entanto, ela ainda não foi aberta ao público. Os dados referentes à automatização da abordagem *D1-EM-D2*, sua revalidação em maior escala e a disponibilização desta ferramenta *web*

para utilização gratuita pela comunidade científica, estão sendo preparados para publicação na forma de artigo científico (Rigo MM, comunicação pessoal).

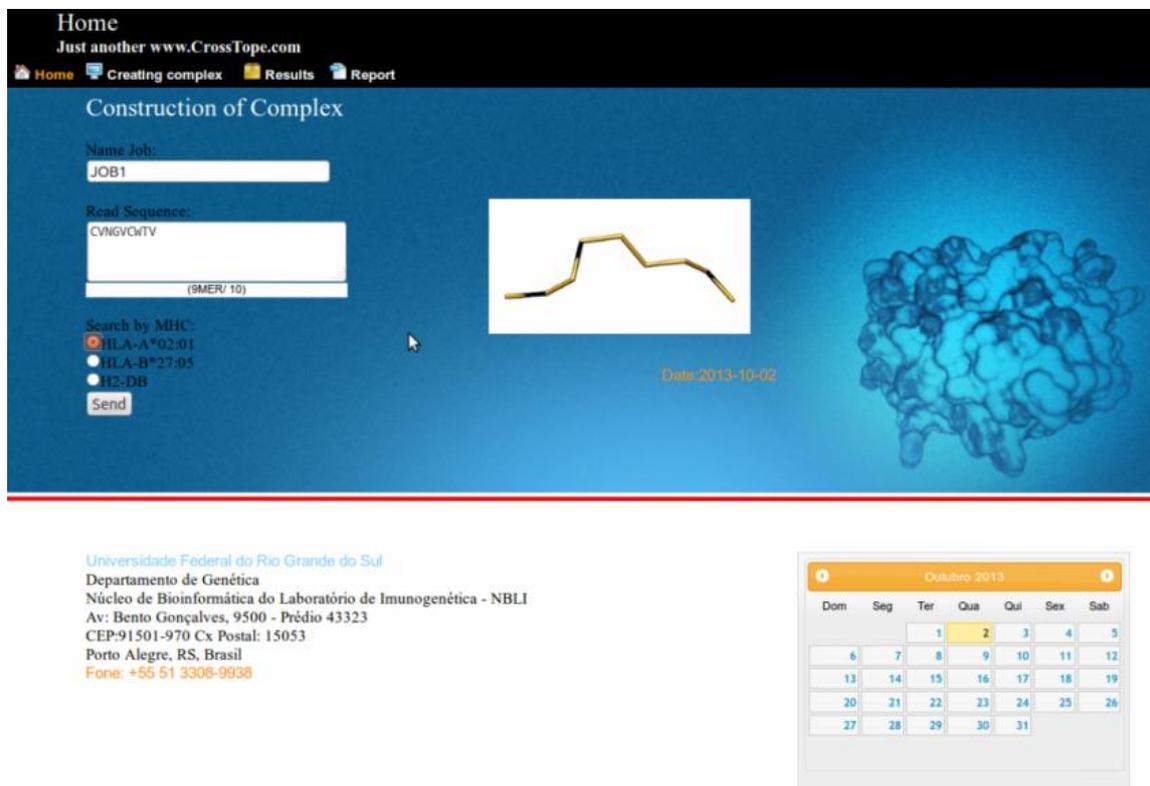


Figura 7. Interface *web* para a construção automática de complexos pMHC. Após informar a sequência e selecionar o MHC, o usuário clica em enviar e aguarda o resultado do processamento, o qual será realizado em um *cluster* de alto desempenho. Esta interface está em fase de testes e ainda não foi disponibilizada para o público. Figura em preto e branco na versão impressa.

Preparação de estruturas para o cálculo do potencial eletrostático

Apesar da predição estrutural de complexos pMHC apresentar diversas aplicações (como análises de ligação ao MHC, estudos de dinâmica molecular, etc), o enfoque do nosso grupo sempre esteve voltado ao estudo da reatividade cruzada de linfócitos T citotóxicos (Rigo *et al.*, 2009; Vieira & Chies, 2005). Embasados por estudos prévios apontando para a importância do potencial eletrostático dos pMHCs no reconhecimento pelo TCR (Kessels *et al.*, 2004; Sandalova *et al.*, 2005), nós passamos a utilizar o programa GRASP2 (Petrey & Honig, 2003) para comparar a superfície de interação (com TCR) de diferentes complexos pMHC (vide capítulo II).

Infelizmente, o programa GRASP2 é disponibilizado apenas para o sistema operacional Windows, além de apresentar algumas outras limitações. Um dos aspectos primordiais para que pudéssemos padronizar uma análise de complexos baseada nas imagens da superfície dos pMHCs era a necessidade de visualizar estas moléculas sempre na mesma orientação. Além disso, frequentemente desejamos calcular o potencial para diversas estruturas pMHC, o que pode gerar conflitos no GRASP2. Para padronizar ao máximo esta etapa, nossa equipe desenvolveu um *script* de preparação para o cálculo do potencial eletrostático (*prep2grasp*). Entre outras funções, este *script* lista todos os complexos pMHC disponíveis no diretório de interesse (*.pdb) e realiza o alinhamento estrutural destas moléculas utilizando uma estrutura de referência (para assegurar a orientação desejada). Ele também edita os nomes de cadeias de todas as moléculas listadas, para evitar que duas cadeias de pMHCs distintos sejam identificadas pela mesma letra, evitando conflitos no GRASP2. Após esta preparação, o arquivo gerado pode ser importado para o GRASP2 onde será calculada a superfície molecular e o potencial eletrostático (etapa ainda não automatizada). Alternativamente, o *script* permite realizar o cálculo do potencial eletrostático diretamente pelo programa Delphi (Li *et al.*, 2012) e calcular a área acessível ao solvente (ASA) de resíduos selecionados (Anexo I), utilizando o programa NACCESS (<http://www.bioinf.manchester.ac.uk/naccess/>).

Desenvolvimento de um *plugin* para importação dos valores RGB

Desde sua implementação em 2011 (Antunes *et al.*, 2011), a nossa metodologia de análise hierárquica de agrupamentos (HCA) baseada em estrutura utiliza imagens geradas pelo programa GRASP2. O programa ImageJ (<http://imagej.nih.gov/ij/>) foi utilizado para definir regiões de interesse (*roiset*), das quais são importados os valores da distribuição RGB (*red, green, blue*). Estas regiões foram definidas através da identificação de áreas de maior variabilidade do potencial eletrostático entre 55 pMHCs não relacionados (Antunes *et al.*, 2011). Os valores utilizados para a análise estatística são a média e o desvio padrão da intensidade de cada componente RGB, calculados sobre todos os *pixels* de cada área selecionada. Estes valores são fornecidos pelo programa ImageJ, mas apresentados na forma de imagem junto ao histograma de cores. Inicialmente, os valores eram importados de forma manual e salvos em uma planilha, para posterior análise estatística. Além da

maior suscetibilidade a erro, este procedimento trabalhoso limitava a análise frequente de grandes conjuntos de complexos.

Recentemente, a nossa equipe desenvolveu um *plugin* para o programa ImageJ que permite realizar a exportação dos valores de interesse diretamente para uma planilha em formato XLS (Figura 8). Uma nova versão do *plugin* está em desenvolvimento, visando permitir a importação simultânea de 42 regiões (capítulo VI) e a exportação dos valores tabelados no formato de entrada para o *pvclust* (programa utilizado para o HCA).

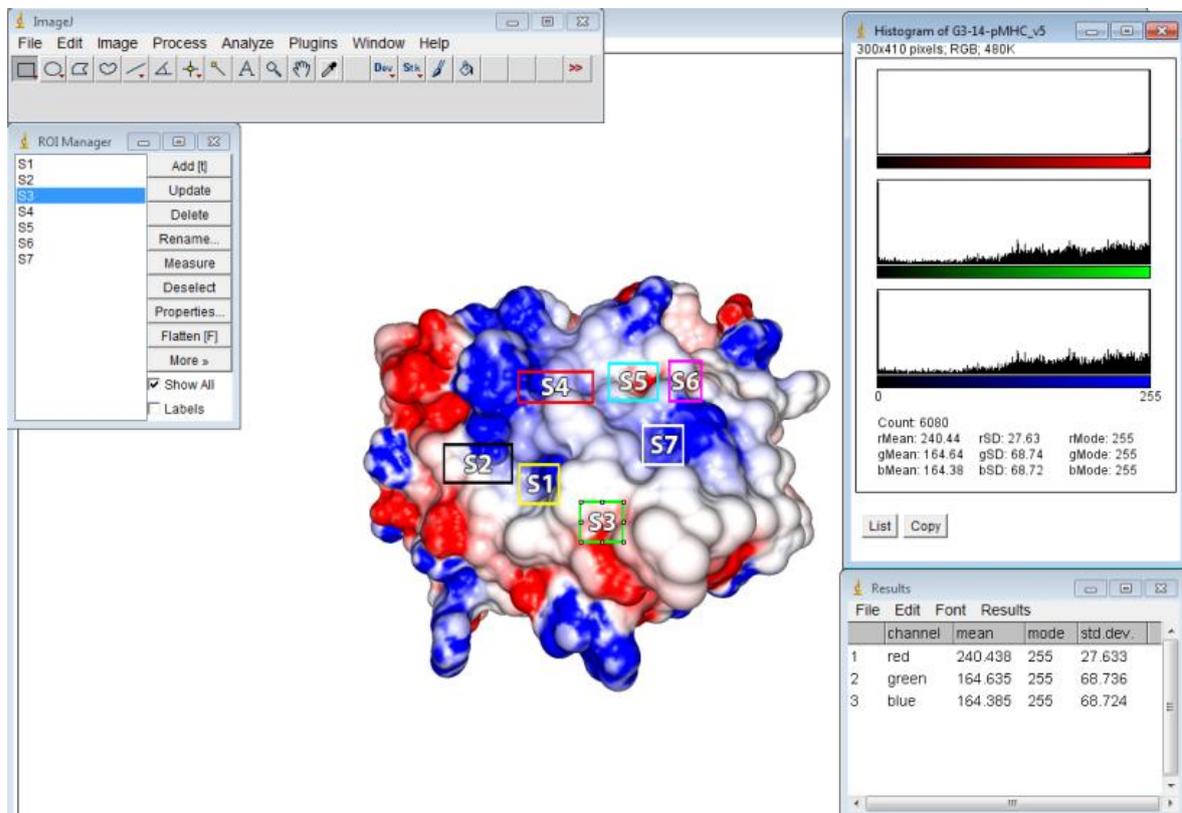


Figura 8. Utilização de um *plugin* para importação de valores RGB no programa ImageJ. Ao centro é possível observar a superfície de um complexo pMHC, sobre a qual estão identificadas as sete regiões de interesse previamente determinadas. A posição destas regiões foi salva, permitindo a seleção independente de cada uma delas através do *ROI Manager*. Após a seleção de uma região e execução do *plugin*, são apresentados os histogramas de cores de cada componente RGB e uma planilha com os valores de média, moda e desvio padrão para cada componente. As 7 seleções apresentadas na imagem foram definidas com base nas regiões de variabilidade entre 55 complexos não relacionados, sendo que a região S1 corresponde a seleção utilizada no primeiro PCA com as 28 variantes de HCV (capítulos II e VI). Figura em preto e branco na versão impressa.

Automatização do HCA com *bootstrap*

Conforme apresentado nos capítulos II e III, o nosso grupo discutia os resultados do HCA baseado em estrutura através da análise de dendrogramas gerados pelo programa SPSS (PASW Statistics 18, IBM, Chicago IL. USA). Além de ser implementado no sistema operacional Windows, o SPSS não fornecia no dendrograma uma validação estatística ou ponto de corte referente à confiabilidade dos agrupamentos gerados. Preocupados com esta questão, nós passamos a utilizar o pacote *pvclust* para o cálculo do HCA. Esta ferramenta desenvolvida em R calcula os agrupamentos, fornecendo ainda uma validação estatística (*p-value*) sobre a confiabilidade dos ramos. Dois valores são fornecidos, sendo um *bootstrap* padrão e um *bootstrap* refinado (*multiscale bootstrap resampling*).

Esta nova abordagem foi testada sobre conjuntos previamente estudados, apresentando bons resultados (Anexo I). Atualmente já foram desenvolvidos *scripts* para automatizar a conversão de tabelas para o formato de entrada do *pvclust*, bem como a execução do HCA com parâmetros pré-definidos e a exportação dos resultados para arquivos em formato PDF.

Capítulo VI

Discussão Geral

Discussão

A reatividade cruzada de linfócitos T permite que alvos heterólogos, no contexto do MHC, desencadeiem a ativação de uma mesma população de células CD8+ (Welsh & Selin, 2002). Este fenômeno, por sua vez, é consequência de uma propriedade intrínseca dos linfócitos T, a poli-especificidade (Wucherpfennig *et al.*, 2007). Em conjunto, poli-especificidade e reatividade cruzada apresentam diversas implicações sobre a resposta imunológica celular e em particular sobre a imunidade heteróloga contra os vírus.

O vírus da Hepatite C (HCV, do inglês *Hepatitis C Virus*) representa um sério problema global de saúde pública, afetando cerca de 3% de toda a população humana (Walker, 2010; Zeisel *et al.*, 2009). A maior parte das infecções, cerca de 70% dos casos, resulta em persistência do vírus no organismo do hospedeiro, sendo a principal causa de doença crônica do fígado, cirrose hepática e carcinoma hepatocelular. Os demais indivíduos, menos de 30% dos casos, resolvem espontaneamente a infecção, normalmente adquirindo uma imunidade protetora contra futuras exposições ao patógeno.

A resposta imune celular, sobretudo quando desencadeada de forma intensa nas fases iniciais da infecção, parece desempenhar um papel fundamental no controle e erradicação do vírus. Alguns alvos imunodominantes são frequentemente observados em pacientes HCV+, dentre os quais se destaca o epitopo NS3₁₀₇₃ (CV/INGVCWTV) (Hiroishi *et al.*, 2010). No entanto, mesmo uma limitada variação em um destes alvos pode levar a ação defectiva de CTLs HCV-específicos, o que levaria à persistência viral e à infecção crônica (Wedemeyer *et al.*, 2002).

Em um famoso estudo realizado em 2001, Wedemeyer e colaboradores conseguiram expandir células T específicas para este alvo (CVNGVCWTV) a partir do sangue de 55% (11/20) dos pacientes HCV+. Curiosamente, eles conseguiram expandir células com a mesma especificidade a partir de 60% (9/15) dos controles saudáveis, constituídos por doadores de sangue sem histórico de infecção por HCV (Wedemeyer *et al.*, 2001). No entanto, um segundo epitopo da mesma proteína foi reconhecido apenas por linfócitos dos pacientes HCV+, não sendo reconhecido por nenhum dos controles. Os

pesquisadores demonstraram ainda que os linfócitos expandidos na presença do peptídeo de HCV apresentavam um fenótipo de memória, defendendo a hipótese de que estas células haviam sido previamente estimuladas *in vivo* na presença de algum alvo heterólogo, respondendo *in vitro* contra o HCV-NS3₁₀₇₃ por um mecanismo de reatividade cruzada.

Através de uma busca por identidade de sequência realizada com ferramentas do *GenBank* (NCBI), os pesquisadores identificaram três possíveis alvos de reatividade cruzada. Destacou-se nesta análise o epitopo de *Influenza* IAV-NA₂₃₁ (CVNGSCFTV), que compartilhava 77% (7/9) de sua sequência linear de aminoácidos com o HCV-NS3₁₀₇₃, incluindo os dois resíduos de ancoragem ao alotipo de MHC humano HLA-A*02:01. Além da similaridade de sequência, a origem do epitopo era coerente com a hipótese dos pesquisadores, uma vez que infecções por *Influenza* são frequentes em humanos. Assim, a infecção prévia por IAV poderia ser a explicação para a origem das células específicas para HCV observadas em controles HCV-, sendo o epitopo IAV-NA₂₃₁ (CVNGSCFTV) o *primer* para esta reatividade cruzada. Corroborando a hipótese dos autores, 44% (4/9) dos controles HCV- que respondiam para HCV-NS3₁₀₇₃ também reconheceram o alvo IAV-NA₂₃₁ em um ensaio *ex vivo* de produção de IFN-gamma (*Elispot*). Os autores realizaram vários outros experimentos para demonstrar que a resposta celular específica contra o alvo IAV-NA₂₃₁ também era gerada durante um processo normal de infecção por *Influenza* (utilizando camundongos transgênicos HLA-A2+) e que existia reatividade cruzada entre estes dois alvos. Eles discutem ainda que apesar destas evidências, o fato de nem todos os controles terem reconhecido o alvo IAV-NA₂₃₁ sugere o possível envolvimento de reatividade cruzada com outros alvos ainda desconhecidos.

Posteriormente, um trabalho publicado por Kasproicz e colaboradores sugeriu que esta reatividade cruzada entre *Influenza* e HCV era relativamente fraca e apresentava uma direcionalidade preferencial no sentido HCV → IAV (Kasproicz *et al.*, 2008). Eles encontraram poucas evidências de resposta contra IAV-NA₂₃₁ em controles saudáveis (HCV-), mas descreveram a geração de células específicas para IAV-NA₂₃₁ após o encontro com HCV.

Conforme descrito no capítulo II, o grupo coordenado pelos professores Markus Cornberg e Heiner Wedemeyer publicou um estudo avaliando a reatividade cruzada entre 28 variantes naturais do epitopo HCV-NS3₁₀₇₃, cobrindo os seis genótipos de HCV (Fyttili *et al.*, 2008). Entre outros experimentos, eles imunizaram um indivíduo saudável com a vacina experimental IC41 (Firbas *et al.*, 2006), que continha o epitopo imunodominante HCV-NS3₁₀₇₃ (CINGVCWTV). Após a coleta de linfócitos e expansão *in vitro* de uma população específica para este alvo, eles testaram a produção de IFN-gamma frente as 28 variantes naturais (no contexto do HLA-A*02:01). Neste estudo foi observada a reatividade cruzada entre o tipo selvagem e variantes dos genótipos 4, 5 e 6 de HCV. Por outro lado, variantes do genótipo 1 (G1) apresentaram resposta heterogênea e não foi observada resposta significativa contra alvos dos genótipos 3 e 4 (capítulo II, Figura 10).

Com base nestes resultados, o nosso grupo utilizou pela primeira vez métodos estatísticos multivariados para realizar o agrupamento de complexos pMHC de acordo com sua similaridade estrutural. Após a modelagem dos 28 epitopos no contexto do HLA-A*02:01, nós realizamos o cálculo do potencial eletrostático na superfície dos complexos e observamos que diferenças de carga em uma região específica poderiam explicar a variação na produção de IFN-gamma frente a uma mesma população de linfócitos. Nós então extraímos valores (RGB) desta região da imagem (obtida da superfície do complexo) e utilizamos como entrada para uma análise de componentes principais (PCA).

O PCA realizado com dados de uma região da superfície dos pMHCs conseguiu agrupar corretamente os complexos (Figura 9). O único complexo que não apresentou nenhuma resposta detectável *in vitro*, G3-18, ficou completamente separado dos demais. Além disso, foi possível verificar a concentração de todas as variantes do genótipo 3 e de todas as variantes do genótipo 2 em faixas bem determinadas, de acordo com a distribuição do PC1 (eixo x). Os complexos com fraca resposta no genótipo 1 também caíram na faixa do genótipo 2, enquanto todos os complexos com elevada produção de IFN-gamma ficaram agrupados em uma faixa dominada pelos complexos do genótipo 6. Tendo conhecimento da possível reatividade cruzada com um alvo de *Influenza* (Wedemeyer *et al.*, 2001), nós também incluímos neste PCA um complexo apresentando o peptídeo de IAV-NA₂₃₁. Como pode ser observado na Figura 9, o PCA posicionou este

possível alvo de reatividade cruzada exatamente ao lado do epitopo selvagem de HCV-NS3₁₀₇₃ (G1-01).

O passo seguinte, também abordado no capítulo II, consistiu em uma triagem virtual de 55 complexos apresentando epitopos de vírus não relacionados. Estes complexos haviam sido previamente modelados para a inclusão no banco de dados CrossTope. Embora a análise de apenas uma região na superfície dos complexos pMHC houvesse sido suficiente para “classificar” as 28 variantes de HCV-NS3₁₀₇₃, nós observamos que a variação estrutural (topografia e cargas) entre estes 55 complexos envolvia outras regiões na superfície. Assim, outras seis regiões foram delimitadas e incluídas no nosso procedimento de extração dos valores RGB (capítulo V, Figura 8).

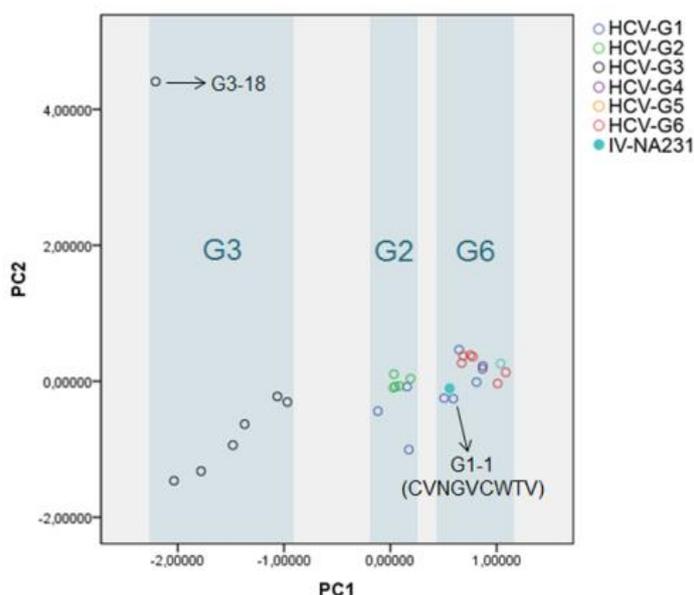


Figura 9. Análise de componentes principais. O único alvo sem resposta detectável *in vitro* (G3-18) ficou isolado, enquanto o alvo selvagem (G1-1) agrupou com alvos que apresentaram maior produção de IFN-gamma. De acordo com o PC1, foi possível observar a distribuição de faixas que incluíam todos os alvos dos genótipos 2 e 3 (não respondedores), agrupando todos os alvos com reatividade cruzada na faixa dominada por alvos do genótipo 6. Modificado de Antunes e colaboradores (Antunes *et al.*, 2011).

Os valores extraídos destas sete regiões foram utilizados como entrada para uma análise hierárquica de agrupamentos (HCA), a qual indicou outros possíveis alvos de reatividade cruzada para o HCV-NS3₁₀₇₃ (Antunes *et al.*, 2011). No artigo que apresentou os dados, foi salientada a similaridade estrutural com os alvos EBV-LMP2₃₂₉ e HIV-GAG₇₇,

cuja reatividade cruzada com HCV-NS3₁₀₇₃ foi posteriormente confirmada (capítulo III). Mas o HCA de 2011 também incluía no mesmo *cluster* os alvos IAV-M1₅₈, CMV-pp65₄₉₅ e EBV-GP85₄₂₀. Os dois primeiros foram recentemente testados *in vitro* frente a diferentes populações de célula T específicas para HCV-NS3₁₀₇₃ (coletadas de diferentes pacientes HCV+), tendo sido reconhecidos em uma parcela dos casos (Zhang S, comunicação pessoal). Em conjunto, estes dados demonstram o sucesso da nossa metodologia de modelagem de complexos pMHC (*D1-EM-D2*) e da triagem virtual baseada nestas estruturas, a despeito da variabilidade do sistema e das simplificações realizadas (como o uso de imagens dos complexos).

Cabe salientar, que o complexo apresentando o epitopo IAV-NA₂₃₁ também foi incluído neste HCA com 55 complexos. No entanto, contrariando o resultado do PCA, ele não foi agrupado com o HCV-NS3₁₀₇₃. Mais do que a diferença entre as estatísticas utilizadas, este resultado contraditório reflete a inclusão de sete regiões no HCA. Uma delas (S7) recuperava valores especificamente de um ponto que apresentava divergência (topográfica) entre os alvos IAV-NA₂₃₁ e HCV-NS3₁₀₇₃ (Figura 10). Excluindo-se esta região do HCA, os dois alvos eram agrupados (dados não apresentados). Esta aparente discrepância entre os resultados do PCA e do HCA vem ao encontro da discussão sobre a inconsistência dos resultados de reatividade cruzada envolvendo estes dois alvos. O próprio Wedemeyer, que descreveu esta reatividade cruzada em um estudo realizado nos Estados Unidos, estava tendo dificuldade em replicar seus dados após seu retorno à Alemanha (Wedemeyer H, comunicação pessoal). Os epitopos e o MHC eram os mesmos testados anteriormente, o que mudava era a origem dos linfócitos utilizados para os ensaios (de pacientes HCV+ ou doadores saudáveis, obtidos nas instituições locais). A despeito de possíveis problemas técnicos com a padronização dos experimentos, estes resultados sugerem que esta reatividade cruzada não é muito frequente, depende da população de linfócitos utilizada no estudo e apresenta ainda uma direcionalidade preferencial (Kasprowicz *et al.*, 2008).

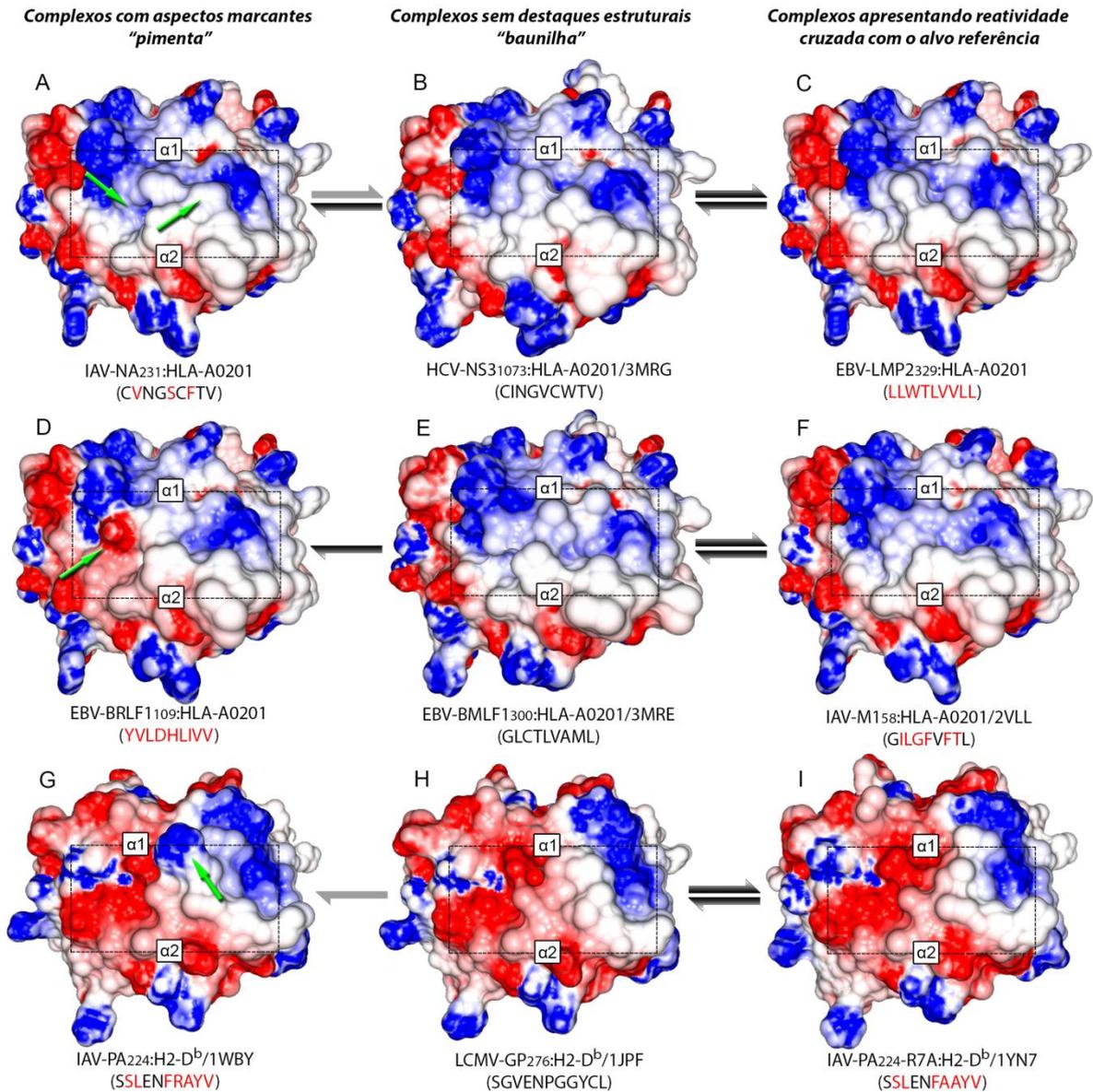


Figura 10. Comparação entre superfícies de complexos envolvidos em reatividade cruzada. Imagens da superfície de contato com o TCR, para cada pMHC, foram obtidas com o programa GRASP2. Os domínios "alfa 1" e "alfa 2" da cadeia pesada do MHC são indicados para cada complexo, bem como a região ocupada pelo peptídeo na fenda (quadro delimitado em preto). Cores indicam a variação do potencial eletrostático em uma escala que varia de -5 kT (vermelho) a +5 kT (azul). A identificação de cada peptídeo:MHC é fornecida abaixo de cada complexo, juntamente com o código PDB da estrutura. Complexos que não possuíam estrutura disponível no PDB foram modelados. A sequência dos epitopos também é indicada em cada complexo, com aminoácidos em vermelho representando alterações em relação ao epitopo referência (coluna do meio). Setas verdes indicam os aspectos marcantes dos epitopos com reatividade cruzada limitada (coluna da esquerda, "sabor pimenta"). Epitopos da coluna central não apresentam características marcantes ("sabor baunilha"). Setas pretas e cinzas indicam a intensidade e a direcionalidade preferencial da reatividade cruzada observada *in vitro* (capítulo IV, Figura 1).

Considerando nosso sucesso na prospecção de alvos de reatividade cruzada baseada em estrutura, independentemente da similaridade de sequência, e tendo em vista as recentes publicações corroborando a ideia de que a similaridade estrutural entre complexos pMHC é um dos principais fatores envolvidos no desencadeamento de eventos de reatividade cruzada (Birnbaum *et al.*, 2014; Shen *et al.*, 2013), concluímos que as diferenças estruturais observadas entre os complexos IAV-NA₂₃₁:HLA-A*0201 e HCV-NS3₁₀₇₃:HLA-A*0201 (Figura 10) são as prováveis responsáveis pelas limitações quanto a reatividade cruzada observada *in vitro* e *ex vivo*.

Uma situação equivalente foi descrita no capítulo IV, envolvendo os epitopos EBV-BRLF1₁₀₉ e EBV-BMLF1₃₀₀, também restritos ao contexto do HLA-A*02:01. Embora reatividade cruzada entre estes alvos tenha sido descrita anteriormente (Cornberg *et al.*, 2010), dados posteriores indicam uma baixa frequência de reatividade cruzada e uma direcionalidade preferencial no sentido EBV-BMLF1₃₀₀ → EBV-BRLF1₁₀₉ (Selin LK, comunicação pessoal). Nas nossas análises, um HCA incluindo alvos que apresentavam reatividade cruzada com EBV-BMLF1₃₀₀ excluiu o EBV-BRLF1₁₀₉ como um alvo estruturalmente relacionado (capítulo IV, Figura S2). Novamente, esta exclusão era ocasionada pela divergência estrutural (de potencial eletrostático) em uma das regiões aferidas (Figura 10). Por sua vez, esta diferença estrutural poderia também ser a responsável pelas limitações quanto à reatividade cruzada observada entre estes alvos.

Adicionalmente, esta diferença estrutural entre complexos pMHC pode ter consequências sobre o perfil da resposta celular desencadeada por diferentes alvos, sobretudo no que tange a clonalidade e a seleção de células com maior propensão a reatividade cruzada. Conforme discutido no capítulo IV, estudos realizados com os complexos IAV-PA₂₂₄:H2-D^b e IAV-NP₃₆₆:H2-D^b indicaram que a presença de uma característica estrutural “marcante” na superfície do pMHC estimula um conjunto diverso de linfócitos T (Turner *et al.*, 2005), com destaque para a dominância de um repertório privado de TCRs (*private specificities*). No caso do IAV-PA₂₂₄:H2-D^b o “sabor marcante” era proporcionado pela presença de uma arginina em P7, um aminoácido com cadeia longa, carregado positivamente, ocupando uma posição no epitopo em que se projeta para fora do MHC. Ao gerar um mutante com a perda desta arginina (IAV-PA₂₂₄-R7A), os autores

observaram uma alteração no padrão de clonalidade, caracterizado pela estimulação de um conjunto mais restrito de linfócitos T e dominado pelo uso de TCRs públicos (compartilhados entre diferentes animais). Estas estruturas também foram incluídas em nossas análises em função da recente publicação de uma fraca reatividade cruzada entre IAV-PA₂₂₄ e LCMV-GP₂₇₆, com implicações sobre a imunopatologia associada à infecção por *Influenza* (Wlodarczyk *et al.*, 2013). Um HCA realizado com estes complexos indicou maior similaridade estrutural entre os alvos IAV-PA₂₂₄-R7A e LCMV-GP₂₇₆, do que entre este último e o alvo selvagem IAV-PA₂₂₄ ou o controle negativo IAV-NP₃₆₆ (capítulo IV, Figura S3).

Em conjunto, estes dados sugerem que eventos de reatividade cruzada podem ocorrer entre complexos que apresentam divergências pontuais na superfície de interação com o TCR, desde que ainda assim compartilhem uma área (maior) de similaridade estrutural. No entanto, esta divergência estrutural apresenta um impacto sobre a frequência e as características dos eventos de reatividade cruzada observados entre estes alvos (Figura 10). A presença de uma característica estrutural marcante (“sabor pimenta”), por exemplo, acaba selecionando clones que interagem com grande afinidade e de maneira específica para esta característica (Figura 11). Assim, a ausência desta característica em um alvo heterólogo dificulta o reconhecimento por esta população de linfócitos, prevenindo eventos de reatividade cruzada. Por outro lado, a seleção frente a um alvo sem uma característica muito marcante, como o EBV-BMLF1₃₀₀ ou o próprio HCV-NS3₁₀₇₃, favorece a seleção de células com maior potencial para o reconhecimento heterólogo. Estas células mais “promíscuas” reconhecem regiões compartilhadas por um conjunto maior de alvos, podendo incluir complexos com divergências pontuais, revelando assim amplas redes de reatividade cruzada (capítulo IV, Figura 1).

Conforme discutido nos capítulos III e IV, o estreitamento da clonalidade é uma característica marcante em eventos de imunidade heteróloga mediada por reatividade cruzada (Cornberg *et al.*, 2006; Welsh & Selin, 2002). Cabe aqui uma diferenciação entre o estreitamento de repertório discutido por Turner em 2005 e aquele discutido por Cornberg em 2006, uma vez que se referem a momentos distintos da reposta celular.

Turner discutiu que um alvo com “sabor marcante” estimula uma resposta policlonal, cujos TCRs dos clones dominantes divergem bastante de um animal para o outro. Alvos menos marcantes também estimulam uma resposta policlonal, mas existe menor variabilidade de TCRs tanto comparando os diferentes clones de um mesmo indivíduo, quanto comparando os clones dominantes entre indivíduos diferentes. Desafios homólogos, com qualquer um dos alvos discutidos, devem manter a mesma (poli)clonalidade observada no desafio primário, tendendo a manter também o mesmo perfil de dominância (Cornberg *et al.*, 2006). Por outro lado, Cornberg e colaboradores demonstraram que em desafios heterólogos não existe manutenção da resposta original. Ocorre estreitamento da resposta (oligoclonal), sendo esta dominada por clones distintos, que poderiam estar muito abaixo na hierarquia de dominância observada na primeira resposta (Cornberg *et al.*, 2006).

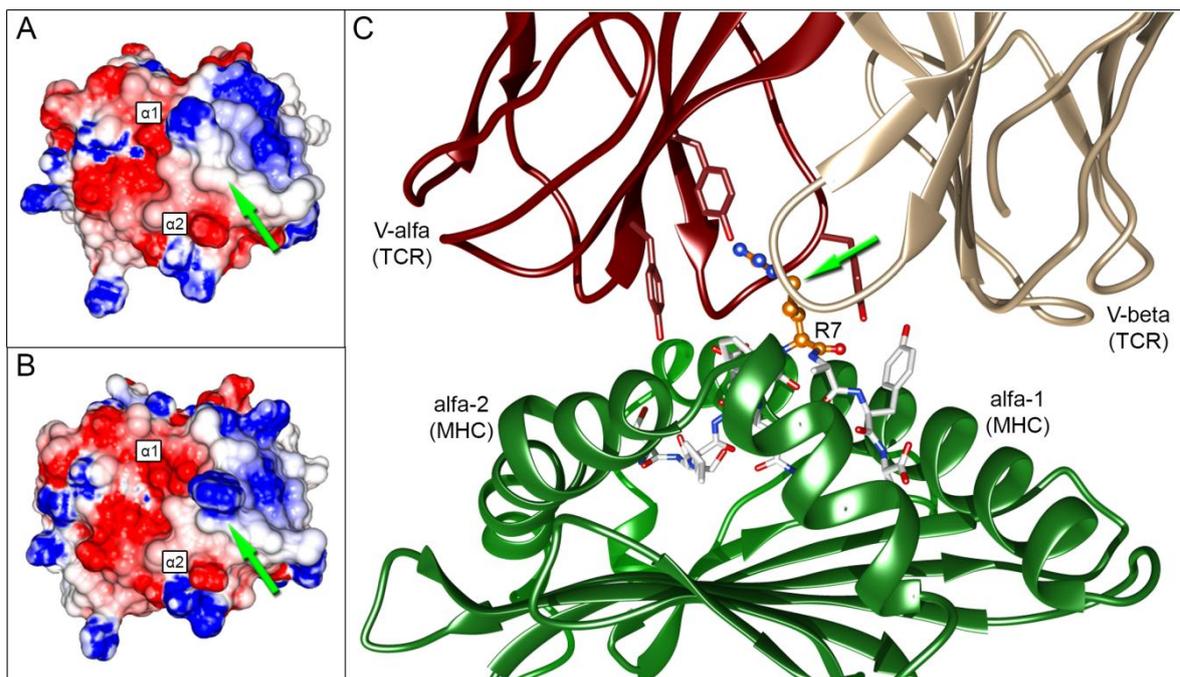


Figura 11. Interação específica entre o TCR e uma Arginina do epitopo. A. Superfície do complexo IAV-PA₂₂₄:H2-D^b calculada com o programa GRASP2 utilizando-se o cristal obtido na ausência do TCR (1WBY). B. Superfície do mesmo complexo, utilizando-se o cristal obtido na presença do TCR (3PQY). C. Representação em *cartoon* do cristal 3PQY indicando resíduos de Tirosina do TCR que interagem diretamente com a Arginina P7 do epitopo (R7), formando uma cavidade negativamente carregada (inversamente complementar). A cadeia lateral do resíduo R7 é apresentada em *ball & stick*. Os domínios do TCR e do MHC são identificados com legendas e a localização do resíduo R7 é indicada pelas setas verdes.

Combinando estes momentos distintos de estreitamento da resposta, podemos compreender a relação entre as diferenças estruturais de complexos pMHC e a direcionalidade da reatividade cruzada. A imunização com EBV-BRLF1₁₀₉, por exemplo, estimula uma ampla variedade de linfócitos do hospedeiro, muitos dos quais apresentam recombinações V-D-J que são únicas daquele indivíduo. Por ter uma característica marcante (carga negativa na superfície de contato com V-alfa), a resposta primária será dominada por clones com alta especificidade por esta característica. Em um novo encontro com o mesmo alvo (desafio homólogo), esta população dominante continua sendo a melhor “alternativa” para reconhecer o alvo, mantendo sua dominância. Em um desafio com EBV-BMLF1₃₀₀, no entanto, a ausência da característica marcante faz com que o clone dominante (original) perca a disputa para um dos outros clones viáveis, o qual conseguia reconhecer com menor afinidade/avidez o alvo primário (EBV-BRLF1₁₀₉), mas também consegue reconhecer com certa afinidade/avidez o alvo heterólogo. Nesta rodada de estimulação frente ao EBV-BMLF1₃₀₀, todos os clones que “dependiam” da interação específica com aquela característica marcante serão perdidos, restando apenas aqueles que conseguem reconhecer ambos os alvos (células propensas à reatividade cruzada). Existe, obviamente, a possibilidade de não existir no conjunto amostrado nenhuma célula capaz de responder para ambos os alvos. Caso haja, estas células poderão ser expandidas na presença de EBV-BMLF1₃₀₀. Um novo desafio com EBV-BRLF1₁₀₉ provavelmente indicará uma resposta mais fraca do que aquela observada na primeira estimulação homóloga. Além disso, neste processo de estreitamento da resposta e troca de dominância, induzido pelo desafio heterólogo, existe uma grande chance de um clone com uma recombinação específica daquele indivíduo se tornar o clone dominante. Isso aumenta a heterogeneidade da resposta entre indivíduos.

São inúmeras as variáveis envolvidas neste sistema, o que dificulta sua compreensão, a testagem de hipóteses (reprodução de dados) e a aplicação destes conhecimentos teóricos no desenvolvimento de produtos ou serviços. A reatividade cruzada não pode ser predita com precisão absoluta através da análise de complexos pMHC. Por mais refinadas que estas análises se tornem, existirá sempre o componente variável e dinâmico da manutenção das populações de linfócitos em cada indivíduo. É o

linfócito T CD8+, com sua específica recombinação V-D-J, que dará a resposta definitiva em cada caso, sobre a ocorrência ou não de uma reatividade cruzada. No entanto, nossos dados sugerem que cuidadosas análises estruturais de complexos pMHC podem nos fornecer estimativas confiáveis a cerca da probabilidade de ocorrência destes eventos, uma informação relevante que pode ter diversas aplicações práticas.

Conforme discutido no capítulo III, imunidade prévia ao alvo vacinal HCV-NS3₁₀₇₃ influenciou significativamente o perfil da resposta frente a imunização com a vacina IC41. No entanto, ainda precisa ser determinado se a antecipação na resposta (observada *in vitro*) e sua manutenção após vacinação (observada *ex vivo*) se refletem em imunidade protetora frente ao desafio com o vírus. Conforme discutido anteriormente, uma imunidade parcial conferida por células de memória que apresentam reatividade cruzada pode ter efeito patogênico, sendo mediadoras de imunopatologias associadas a infecções virais (Cornberg *et al.*, 2013; Welsh & Fujinami, 2007; Wlodarczyk *et al.*, 2013). Do mesmo modo, a imunidade parcial conferida pelo reconhecimento de alvos de reatividade cruzada em infecções mais frequentes, como EBV, Influenza, Herpes Simplex, HPV, entre outros, pode ser uma das explicações para a variabilidade nos desfechos observados frente à infecção por HCV e para a prevalência de infecções crônicas (levando a imunopatologias hepáticas).

Seja para projetar vacinas com maior abrangência e eficácia ou para detectar possíveis desfechos patológicos, uma estimativa de reatividade cruzada seria de extremo interesse. Durante muitos anos, a análise de similaridade de sequências foi a única ferramenta disponível para a identificação de alvos de reatividade cruzada. Mais tarde, o compartilhamento de propriedades bioquímicas dos resíduos dos epitopos foi sugerido como uma das explicações para eventos de reatividade cruzada envolvendo alvos com menor similaridade de sequência (Vieira & Chies, 2005), conceito que chegou a ser empregado para a predição de reatividade cruzada (Frankild *et al.*, 2008).

Recentemente, a equipe coordenada pela pesquisadora Anne S. De Groot publicou o JanusMatrix, uma ferramenta da empresa *EpiVax, Inc.* para a predição de reatividade cruzada em larga escala (Moise *et al.*, 2013). Baseada na análise de cristais de complexos TCRpMHC, foram identificados resíduos do epitopo (posições) que normalmente

interagem com o TCR e resíduos que normalmente servem de âncora para o MHC. Assim, a ferramenta divide o epitopo de interesse em duas “faces”: a face que contata o MHC e a face que contata o TCR. Na prática isso significa mapear quais resíduos estão em cada face e utilizar este “padrão” como entrada para uma busca por identidade de sequência em larga escala. O algoritmo permite ainda certa variabilidade dos resíduos que compõe a face do MHC (aumentando sua sensibilidade), desde que não prejudiquem a ligação com o mesmo. A ferramenta foi inicialmente utilizada para mapear a ocorrência de potenciais alvos de reatividade cruzada em sequências proteicas obtidas do genoma humano, assim como do microbioma e do genoma de vírus e bactérias patogênicas. Os autores salientam as possíveis aplicações da ferramenta, mas ponderam que esta análise em larga escala esta voltada a aspectos populacionais, não sendo precisa no que se refere a respostas de indivíduos ou o contexto de MHCs específicos. Salientam ainda que ela pode estar sujeita a uma série de vieses, como erros de anotação nas sequências fornecidas pelos bancos pesquisados.

Apesar de o JanusMatrix considerar alguma flexibilidade nos padrões utilizados e permitir uma análise em larga escala, sua base continua sendo a comparação por similaridade de sequências. Conforme discutido nos capítulos III e IV, nossos dados salientam a grande divergência de sequência entre alvos que compartilham características estruturais e apresentam confirmada reatividade cruzada *in vitro*, considerando o contexto de um MHC específico. A nossa técnica possui um grande potencial de prospecção, mas ainda apresenta limitações no que se refere à separação de alvos com grande similaridade estrutural (capítulo IV). Uma das possíveis explicações é o fato de ainda estarmos utilizando um número limitado de regiões, as quais não contém toda a informação presente na superfície de interação com o TCR.

No momento, estamos desenvolvendo um novo *plugin* para o software ImageJ, o qual permitirá a importação automatizada de um número muito maior de regiões (Figura 12). Associado ao uso do *pvclust*, que permite estimar a confiabilidade dos agrupamentos no HCA, nós acreditamos que a ferramenta poderá apresentar maior especificidade na identificação de variações estruturais entre os complexos. Adicionalmente, outros descritores dos complexos pMHC, como a acessibilidade de resíduos de contato, podem

ser também incorporados a análise (Anexo I). Por outro lado, isso diminui a sensibilidade da técnica, sobretudo no que se refere aos casos de reatividade cruzada entre complexos com divergências pontuais. Visando a identificação destes candidatos, nosso grupo estuda também a possibilidade de gerar múltiplos HCAs alternativos para um mesmo conjunto de complexos, alternando as regiões importadas para análise. Desta forma, pares como IAV-NA₂₃₁ e HCV-NS3₁₀₇₃ não seriam agrupados na análise com todas as regiões, mas seriam agrupados em várias das análises com um número menor de regiões.

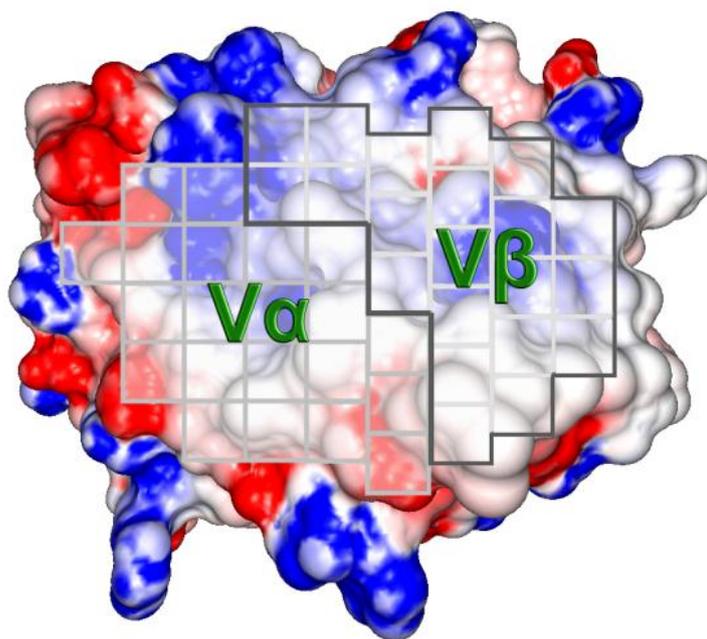


Figura 12. Nova proposta de seleção de regiões para agrupamento de complexos pMHC. Ao total, 42 regiões de tamanho uniforme foram definidas sobre a imagem da superfície do pMHC, cobrindo as principais áreas de interação com as cadeias V α e V β do TCR. O arquivo delimitando estas regiões (*RoiSet*) foi salvo para uso no software ImageJ e um *plugin* está sendo desenvolvido para realizar a importação automática dos valores RGB destas regiões.

Apesar das limitações técnicas e da gigantesca complexidade deste sistema, a reatividade cruzada é uma temática atual na imunologia, atraindo a atenção e o investimento de diversos setores. Neste contexto, o estudo de características estruturais de complexos pMHC pode fornecer estimativas úteis para o planejamento de vacinas e a prevenção de imunopatologias, bem como contribuir para a compreensão dos mecanismos moleculares envolvidos na ocorrência de fenômenos imunológicos complexos.

Referências Complementares (capítulos I, V e VI):

- Abbas AK and Lichtman AH (2005) *Imunologia Celular e Molecular*. 5 edition. ELSEVIER, Rio de Janeiro, 576 pp.
- Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, Chen W, Levin AM, Connolly JM, Zhu C *et al.* (2011) T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity* 35:681-693.
- Antunes DA, 2008 *Utilização de Ferramentas de Bioinformática para a Análise do Potencial de Reatividade Cruzada entre Epitopos Virais*, pp.
- 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. *Mol Immunol* 48:1461-1467.
- Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M and Chies JAB (2010) 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.
- Barclay AN (1999) Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition. *Proc Natl Acad Sci U S A* 96:14672-14674.
- Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, Ozkan E, Davis MM, Wucherpennig KW and Garcia KC (2014) Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* 157:1073-1087.
- Bordner AJ (2013) Structure-based prediction of Major Histocompatibility Complex (MHC) epitopes. *Methods Mol Biol* 1061:323-343.
- Bordner AJ and Abagyan R (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* 63:512-526.
- Brehm MA, Selin LK and Welsh RM (2004) CD8 T cell responses to viral infections in sequence. *Cell Microbiol* 6:411-421.
- Chang MW, Ayeni C, Breuer S and Torbett BE (2010) Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS One* 5:e11955.
- Chen Y, Shi Y, Cheng H, An Y-Q and Gao GF (2009) Structural immunology and crystallography help immunologists see the immune system in action: how T and NK cells touch their ligands. *IUBMB life* 61:579-590.
- Cornberg M, Chen AT, Wilkinson LA, Brehm MA, Kim SK, Calcagno C, Ghersi D, Puzone R, Celada F, Welsh RM *et al.* (2006) Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. *J Clin Invest* 116:1443-1456.
- 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. *J Immunol* 184:2825-2838.
- Cornberg M, Kenney LL, Chen AT, Waggoner SN, Kim SK, Dienes HP, Welsh RM and Selin LK (2013) Clonal exhaustion as a mechanism to protect against severe immunopathology and death from an overwhelming CD8 T cell response. *Front Immunol* 4:475.
- Dhanik A, McMurray JS and Kavraki LE (2013) DINC: A new AutoDock-based protocol for docking large ligands. *BMC Struct Biol* 13:S11.
- Donati C and Rappuoli R (2013) Reverse vaccinology in the 21st century: improvements over the original design. *Ann N Y Acad Sci* 1285:115-132.
- Elhanati Y, Murugan A, Callan CG, Jr., Mora T and Walczak AM (2014) Quantifying selection in immune receptor repertoires. *Proc Natl Acad Sci U S A* 111:9875-9880.
- Fanning L, Bertrand FE, Steinberg C and Wu GE (1998) Molecular mechanisms involved in receptor editing at the Ig heavy chain locus. *Int Immunol* 10:241-246.
- Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A *et al.* (2007) A whole-genome association study of major determinants for host control of HIV-1. *Science* 317:944-947.
- Firbas C, Jilma B, Tauber E, Buerger V, Jelovcan S, Lingnau K, Buschle M, Frisch J and Klade CS (2006) Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a

- randomized, placebo controlled trial for dose optimization in 128 healthy subjects. *Vaccine* 24:4343-4353.
- 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:e1831.
- Fyttili 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.
- Garcia KC, Adams JJ, Feng D and Ely LK (2009) The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat Immunol* 10:143-147.
- Geuking MB, Koller Y, Rupp S and McCoy KD (2014) The interplay between the gut microbiota and the immune system. *Gut Microbes* 5:[Epub ahead of print].
- 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.
- 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. *Journal of biomedicine & biotechnology* 2010:263810.
- Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC, Wright MW *et al.* (2004) Gene map of the extended human MHC. *Nat Rev Genet* 5:889-899.
- Huang E and Wells CA (2014) The Ground State of Innate Immune Responsiveness Is Determined at the Interface of Genetic, Epigenetic, and Environmental Influences. *J Immunol* 193:13-19.
- Ikekawa A and Ikekawa S (2001) Fruits of human genome project and private venture, and their impact on life science. *Yakugaku Zasshi* 121:845-873.
- Kaspruwicz V, Ward SM, Turner A, Grammatikos A, Nolan BE, Lewis-Ximenez L, Sharp C, Woodruff J, Fleming VM, Sims S *et al.* (2008) Defining the directionality and quality of influenza virus-specific CD8+ T cell cross-reactivity in individuals infected with hepatitis C virus. *J Clin Invest* 118:1143-1153.
- Kelley J, Walter L and Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56:683-695.
- 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. *J Immunol* 172:2324-2331.
- Khan JM and Ranganathan S (2010) pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Res* 6 Suppl 1:S2.
- Korber B, LaBute M and Yusim K (2006) Immunoinformatics comes of age. *PLoS Comput Biol* 2:e71.
- Kubinak JL, Ruff JS, Hyzer CW, Slev PR and Potts WK (2012) Experimental viral evolution to specific host MHC genotypes reveals fitness and virulence trade-offs in alternative MHC types. *Proc Natl Acad Sci U S A* 109:3422-3427.
- Kurosawa Y, von Boehmer H, Haas W, Sakano H, Trauneker A and Tonegawa S (1981) Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature* 290:565-570.
- Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* 470:187-197.
- Lauring AS, Frydman J and Andino R (2013) The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol* 11:327-336.
- Lefranc MP (2014) Immunoglobulin and T Cell Receptor Genes: IMGT((R)) and the Birth and Rise of Immunoinformatics. *Front Immunol* 5:22.
- Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M and Alexov E (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys* 5:9.
- Madurga S, Belda I, Llorà X and Giralte E (2005) Design of enhanced agonists through the use of a new virtual screening method: application to peptides that bind class I major histocompatibility complex (MHC) molecules. *Protein science : a publication of the Protein Society* 14:2069-2079.
- Maki R, Roeder W, Trauneker A, Sidman C, Wabl M, Raschke W and Tonegawa S (1981) The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes. *Cell* 24:353-365.

- Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, Verberkmoes NC, Sztejn MB, Losikoff PT, Martin WD *et al.* (2013) The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. *Hum Vaccin Immunother* 9:1577-1586.
- Nakamoto N and Kanai T (2014) Role of toll-like receptors in immune activation and tolerance in the liver. *Front Immunol* 5:221.
- Pappalardo F, Chiacchio F and Motta S (2013) Cancer vaccines: state of the art of the computational modeling approaches. *Biomed Res Int* 2013:106407.
- Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B *et al.* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275-278.
- Petrey D and Honig B (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol* 374:492-509.
- Plewczynski D, Lazniewski M, Augustyniak R and Ginalski K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J Comput Chem* 32:742-755.
- Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shiels 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.
- Rappuoli R (2000) Reverse vaccinology. *Current opinion in microbiology* 3:445-450.
- Rigo M, Antunes D, Vieira G and Chies J (2009) MHC: Peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein. *Lecture Notes in Computer Science* 5676:160-163.
- Rudolph MG, Stanfield RL and Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annual review of immunology* 24:419-466.
- Saito T, Yokosuka T and Hashimoto-Tane A (2010) Dynamic regulation of T cell activation and co-stimulation through TCR-microclusters. *FEBS letters* 584:4865-4871.
- Salazar MI, Del Angel RM, Lanz-Mendoza H, Ludert JE and Pando-Robles V (2014) The role of cell proteins in dengue virus infection. *J Proteomics*:[Epub ahead of print].
- Sandalova T, Michaelsson J, Harris RA, Odeberg J, Schneider G, Karre K, Achour A, Michaelsson J and Karre K (2005) A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity. *J Biol Chem* 280:27069-27075.
- Schatz DG, Oettinger MA and Schlessel MS (1992) V(D)J recombination: molecular biology and regulation. *Annu Rev Immunol* 10:359-383.
- Schmid M, Speiseder T, Dobner T and Gonzalez RA (2014) DNA virus replication compartments. *J Virol* 88:1404-1420.
- Seder RA, Darrah PA and Roederer M (2008) T-cell quality in memory and protection: implications for vaccine design. *Nat Rev Immunol* 8:247-258.
- Shen ZT, Nguyen TT, Daniels KA, Welsh RM and Stern LJ (2013) Disparate epitopes mediating protective heterologous immunity to unrelated viruses share peptide-MHC structural features recognized by cross-reactive T cells. *J Immunol* 191:5139-5152.
- Sinigaglia M, Antunes DA, Rigo MM, Chies JA and Vieira GF (2013) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database (Oxford)* 2013:bat002.
- Sliwoski G, Kothiwale S, Meiler J and Lowe EW, Jr. (2013) Computational methods in drug discovery. *Pharmacol Rev* 66:334-395.
- Sohn SJ, Thompson J and Winoto A (2007) Apoptosis during negative selection of autoreactive thymocytes. *Curr Opin Immunol* 19:510-515.
- Thauland TJ and Parker DC (2010) Diversity in immunological synapse structure. *Immunology* 131:466-472.
- Tomar N and De RK (2010) Immunoinformatics: an integrated scenario. *Immunology*:153-168.
- 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.
- Turner SJ, Kedzierska K, Komodromou H, La Gruta NL, Dunstone MA, Webb AI, Webby R, Walden H, Xie W, McCluskey J *et al.* (2005) Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus-specific CD8+ T cell populations. *Nat Immunol* 6:382-389.
- van der Merwe PA and Dushek O (2010) Mechanisms for T cell receptor triggering. *Nature reviews. Immunology* 11:47-55.

- Vandiedonck C and Knight JC (2009a) The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief Funct Genomic Proteomic* 8:379-394.
- Vandiedonck C and Knight JC (2009b) The human Major Histocompatibility Complex as a paradigm in genomics research. *Briefings in functional genomics & proteomics* 8:379-394.
- 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? *Med Hypotheses* 65:873-879.
- Vivona S, Gardy JL, Ramachandran S, Brinkman FSL, Raghava GPS, Flower DR and Filippini F (2008) Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends in biotechnology* 26:190-200.
- Walker CM (2010) Adaptive immunity to the hepatitis C virus. *Advances in virus research* 78:43-86.
- Wedemeyer H, He X-S, Nascimbeni M, Davis AR, Greenberg HB, Hoofnagle JH, Liang TJ, Alter H and Rehermann B (2002) Impaired Effector Function of Hepatitis C Virus-Specific CD8+ T Cells in Chronic Hepatitis C Virus Infection. *J. Immunol.* 169:3447-3458.
- 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. *J Virol* 75:11392-11400.
- Welsh RM, Che JW, Brehm Ma and Selin LK (2010) Heterologous immunity between viruses. *Immunol Rev* 235:244-266.
- Welsh RM and Fujinami RS (2007) Pathogenic epitopes, heterologous immunity and vaccine design. *Nat Rev Microbiol* 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.
- Welsh RM, Selin LK and Szomolanyi-Tsuda E (2004) Immunological memory to viral infections. *Annu Rev Immunol* 22:711-743.
- Wlodarczyk MF, Kraft AR, Chen HD, Kenney LL and Selin LK (2013) Anti-IFN-gamma and peptide-tolerization therapies inhibit acute lung injury induced by cross-reactive influenza A-specific memory T cells. *J Immunol* 190:2736-2746.
- Wucherpfennig KW, Allen PM, Celada F, Cohen IR, De Boer R, Garcia KC, Goldstein B, Greenspan R, Hafler D, Hodgkin P *et al.* (2007) Polyspecificity of T cell and B cell receptor recognition. *Semin Immunol* 19:216-224.
- Xie T, Rowen L, Aguado B, Ahearn ME, Madan A, Qin S, Campbell RD and Hood L (2003) Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. *Genome Res* 13:2621-2636.
- Yewdell JW, Reits E and Neefjes J (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nat Rev Immunol* 3:952-961.
- Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN and Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front Immunol* 4:485.
- Zeisel MB, Fafi-Kremer S, Robinet E, Habersetzer F, Baumert TF and Stoll-Keller F (2009) Adaptive Immunity to Hepatitis C Virus. *Viruses* 1:276-297.
- Zuniga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, Acuna-Alonzo V, Yunis M, Granados-Montiel J, Cruz-Lagunas A *et al.* (2013) HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations. *PLoS One* 8:e74442.

Anexo I

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

(Artigo completo submetido a revista PLoS One)

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

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 phenomenon 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. 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. 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 a 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.

