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 expetativas. 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**, **Derlin** 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 Derlin, 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, Liisa 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. Conseguiste 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 pvclust 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*

Paraskevi Fytili. *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 llevaré 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

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 ingles *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) _ Virus 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).

Figura 1. A corrida da Rainha Vermelha. Ilustração de Sir John Tenniel para o livro *Through the Looking-Glass* (1871), de Lewis Carrol. 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 reposta 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).

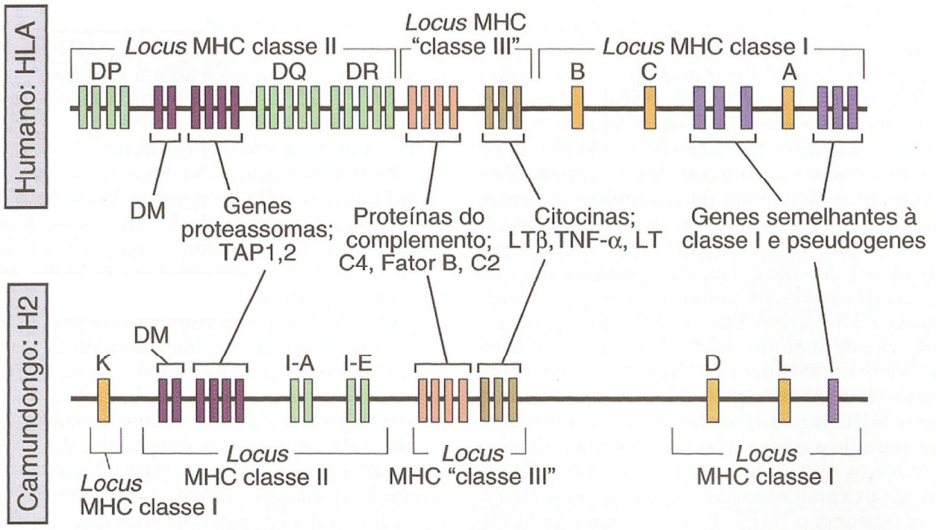


Figura 2. Mapas esquemáticos dos *loci* do MHC humano e murino. Os genes semelhantes à classe I no *locus* do MHC de classe I humano, se referem aos chamados HLA não clássicos, como o HLA-G, os quais não estão diretamente envolvidos com a apresentação de peptídeos endógenos, embora desempenhem outros 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 (β2-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 epitopos) 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).

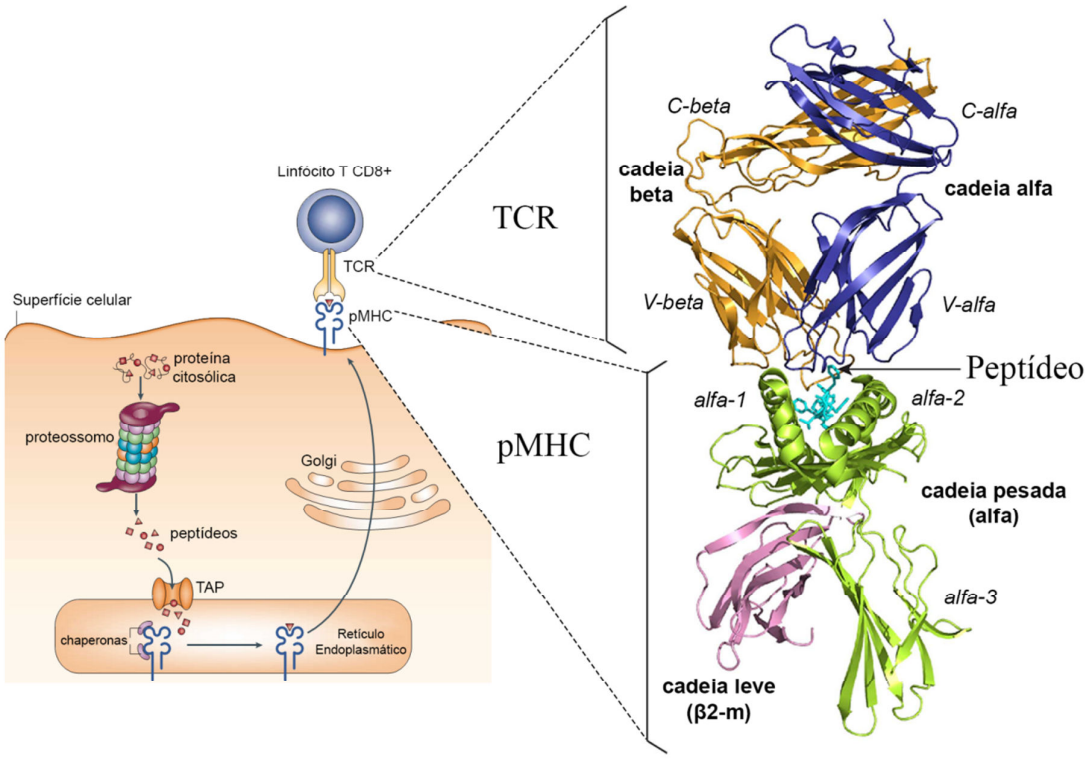


Figura 3. Estrutura e função do complexo pMHC. A esquerda, representação esquemática da via de apresentação de peptídeos endógenos. Uma amostra das proteínas citosólicas é digerida pelo proteossomo e os peptídeos derivados são ultimamente apresentados na superfície celular, complexados a molécula de 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 compõe 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).

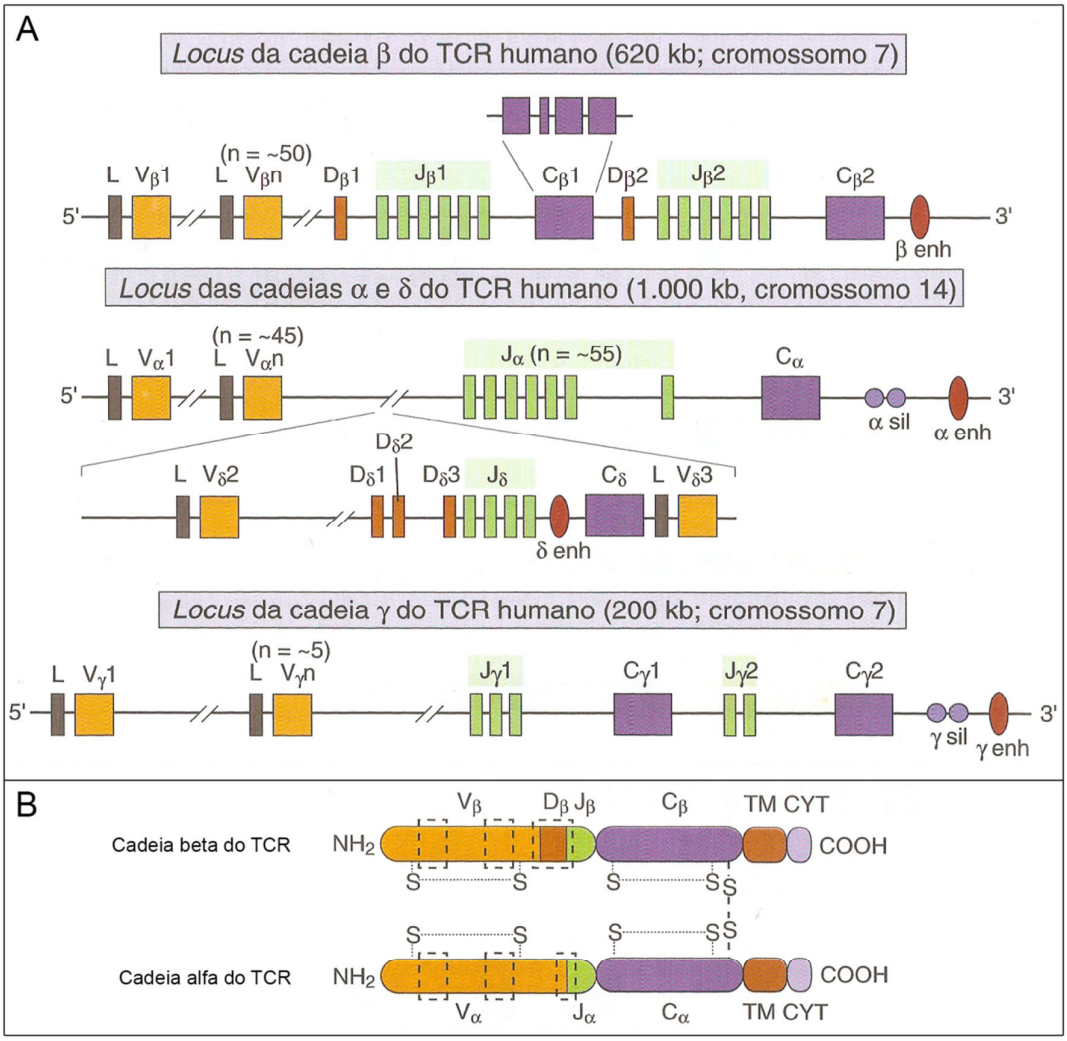


Figura 4. Organização do TCR. A. Organização dos *loci* das cadeias α, β, γ e δ do TCR humano, na linhagem germinativa. Éxons e íntrons não estão representados em escala e pseudo-genes não são indicados. Os genes constantes (C) são representados como blocos simples, mas são na verdade compostos por vários éxons (ex. Cβ1). 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 avidez/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 avidez/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 (Birnbaum *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).

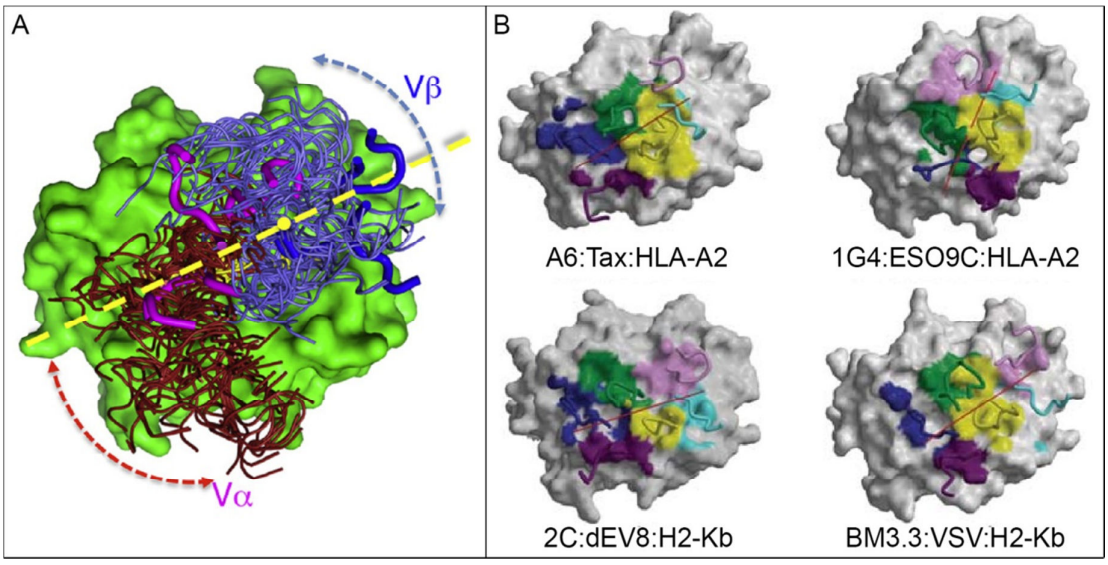


Figura 5. Orientação da interação TCRpMHC. A. Espectro de orientações de ancoramento para 40 TCRs agonistas, alinhados sobre a "face" de contato de um pMHC. As alças de um TCR não agonista (p3A1) são representadas com linhas mais grossas em azul e roxo, com a linha tracejada em amarelo indicando sua 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 contatados pelas alças do TCR são representados em cinza na superfície do pMHC. Demais resíduos são coloridos de acordo com a alça do TCR contatada, seguindo o esquema: CDR1α, azul escuro; CDR2α, magenta; CDR3α, verde; CDR1β, ciano; CDR2β, rosa; CDR3β, 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). A despeito 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.

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

ShiHong Zhang[1#]; Rakesh K. Bakshi[1,2#]; Paraskevi Fytili[1]; Dinler A Antunes[3]; Gustavo F. Vieira[3]; Roland Jacobs[4]; Christoph S. Klade[5]; Michael P. Manns[1]; Anke Kraft[1]; Heiner Wedemeyer[1]; Verena Schlaphoff[1*] and Markus Cornberg[1*]

1 Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; 2 Hannover Biomedical Research School (HBRS); 3 NBLI - Núcleo de Bioinformática do Laboratório de Imunogenética, Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil; 4 Department of Clinical Immunology and Rheumatology, Hannover Medical School, Hannover, Germany; 5 Intercell AG, Vienna, Austria; # and * authors contributed equally to this work

Corresponding author:
PD Dr. Markus Cornberg
Department of Gastroenterology, Hepatology and Endocrinology
Hannover Medical School, Hannover, Germany
Telephone: 001495115326821
Fax: 001495115326820
E-mail: cornberg.markus@mh-hannover.de

Running title: HCV-specific CD8+ T-cells in seronegatives

Abstract

T-cells specific for pathogens have been detected in unexposed individuals, but little is known about their characteristics or origin. The existence of such T-cells might have consequences for the outcome of infection and response to vaccination. In this study, we systematically investigated how frequently HCV-specific CD8+ T-cells can be found in HCV-seronegative individuals (HCV-SN), which properties these cells have and whether they might impact on response to HCV vaccination. We found that CD8+ T-cells recognizing the immunodominant epitope HCV NS3-1073 were surprisingly abundant in HCV-SN irrespective of risk factors for HCV exposure. These cells were partially memory-phenotype cells and displayed variable private and in some individuals skewed T-cell receptor repertoires. *In vitro* cross-recognition of peptides from unrelated viruses suggests that the HCV-specific T-cells found in HCV-SN might originate from previous heterologous infections. Of note, in healthy individuals vaccinated with an HCV peptide vaccine, the *in vitro* response to HCV before vaccination correlated with a more vigorous and earlier response towards the vaccine. Thus, we suggest that virus-specific CD8+ T-cells can be present in unexposed individuals due to cross-reactivity and might influence responses to vaccines.

Author summary

Immune responses to vaccination show a wide individual variability. Also, responses in humans towards infections like HCV are heterogeneous regarding severity of disease and natural outcome. The reasons for this are not yet fully understood. Here, we report that HCV-specific CD8+ T-cells can readily be found *ex vivo* and expanded *in vitro* in healthy HCV seronegative individuals. These cells were partially memory-phenotype cells and displayed characteristics of selective clonal expansion. We further found cross-recognition of EBV-, Influenza- and HIV-derived peptides by these HCV-specific CD8+ T-cells, which demonstrate a possible implication of cross-reactivity in the generation of the HCV-specific response in seronegative individuals. Of interest, we could show that in healthy individually who received in a phase I study a HCV peptide vaccine T cell responses towards HCV seem to be influenced by the pre-existence of HCV-specific CD8+ T-cells. In those vaccinated individuals responding *in vitro* to HCV already before the first vaccination, the CD8+ T cell response towards the vaccine occurred earlier and with a higher magnitude. Overall, our data show that virus-specific T-cells exist in unexposed individuals and that these cells might have a clinical impact influencing immune responses towards vaccination.

Introduction

The natural courses of infections as well as response to vaccines can be quite variable. For example, hepatitis C virus (HCV) infection show a strong heterogeneity ranging from asymptomatic to symptomatic acute phases with spontaneous clearance or viral persistence [1]. Also the response to vaccination likewise is different between individuals, where the magnitude and the timing of appearance of immune responses might differ substantially and some individuals even are non-responder to the particular vaccine [2]. In addition, vaccines may induce non-specific immune effects [3]. The reasons for this variability are not known in detail, factors like dosage, route of infection or vaccination as well as pathogen and host genetics are considered to play a role.

Another element influencing the natural course of infection or vaccination might be a pre-existence of virus-specific cells, which become re-activated and participate in the immune response thereby changing the course of disease. Virus-specific T-cells have repeatedly been documented in individuals who have not encountered these pathogens before. Recently, human immunodeficiency virus (HIV) or herpes simplex virus (HSV) specific CD4+ T-cells were described to be present in uninfected persons and these cells displayed a phenotype and expression of genes associated with memory cells [4]. In other cases, T-cells specific for a pathogen were documented unintended in unexposed persons, as for example in the case of severe acute respiratory syndrome virus (SARS) or HIV specific CD4+ T-cells [5,6]. Similarly, T-cells specific for the hepatitis C virus (HCV) were shown to be present in persons not exposed to and not infected with HCV [7-10]. However, little is known about the characteristics, phenotype, functionality or the possible origin of these T-cells as they have not yet been systematically investigated. Also, possible clinical impacts of such cells in case of infection or vaccination remain largely unknown. In the case of HCV specific T-cells in unexposed individuals different explanations were discussed including activation of naïve precursor cells [7], exposure to HCV without seroconversion [8] or cross-reactivity with other epitopes [4,11]. The aim of this study was to analyze the frequencies of HCV specific CD8+ T-cells in HCV-seronegative (HCV-SN) individuals with different risk factors for HCV exposure and to systematically characterize these cells in regard to their phenotype, functionality, T-cell receptor repertoire and clonality. For these analyses we used the HCV model epitope NS3-1073 which is frequently targeted in HCV-infected patients [12]. This peptide is also included in a HCV peptide vaccine candidate and was applied in clinical trials to healthy individuals [13,14].

We here show that CD8+ T-cells specific for HCV are astonishingly frequent in HCV-SN and that they partially are memory-phenotype cells. Cross-recognition of peptides derived from Epstein-Barr-Virus (EBV), Influenza A Virus (IAV) or HIV suggests a possible origin of these cells due to cross-reactivity. Analyses in healthy individuals vaccinated with a HCV peptide vaccine suggest influence on vaccine responses.

Results

HCV-specific CD8+ T-cells are present in seronegative individuals

The study cohort for the analysis of HCV-specific CD8+ T-cells in HCV-seronegative (HCV-SN) individuals comprised altogether 164 individuals. To answer the question to what extent HCV-specific CD8+ T-cells exist in HCV-SN, healthy blood donors first were screened in an unbiased approach for the presence of HCV NS3-1073 specific CD8+ T-cells by MHC-I multimer staining directly *ex vivo*. In 35 of the 126 individuals analyzed *ex vivo* (27.8%) HCV-specific CD8+ T-cells were detectable with a frequency of 0.05% to 0.8% of CD8+ T-cells (median 0.021%, Figure 1a and 1b). Further, another sensitive approach was used by testing cells *ex vivo* for NS3-1073 specific IFNγ production by EliSpot. Four of 18 HCV-SNs (22%) showed peptide-specific responses above the set cut-off (9 SFU/300,000 cells). In two of the samples the IFNγ secretion was even higher with >27 SFU/300,000 cells (Figure 1b). Characteristics of all HCV-SN included in this study are given in supplementary table S1.

HCV-specific CD8+ T-cells in seronegative individuals partially display a memory phenotype

In order to gain more insight into the properties of the NS3-1073 specific CD8+ T-cells that can be found among healthy blood donors, we aimed to phenotypically analyze the memory status of these cells. As the *ex vivo* frequency was too low to allow this analysis, the method of enrichment of antigen-specific cells using MHC Class I multimers by magnetic bead isolation was used. This approach was performed with samples from donors already showing NS3-1073 specific CD8+ T-cells *ex vivo*. HCV NS3-1073 specific CD8+ T-cells could be enriched in 16 out of 21 selected blood donors. Parts of the NS3-1073 specific CD8+ T-cells were identified as memory-phenotype T-cells as they were negative for CCR7 and CD45RA (Figure 1c and Supplementary Figure S1) with frequencies varying between 1.9% and 89.5%. However, as the identification of memory phenotype by CCR7 and CD45RA can be imprecise as also antigen-experienced T-cells can be positive for these two markers thus appearing to be naïve cells (e.g. as seen by CD95 data [15]). Therefore we additionally stained the memory-associated markers CD11a in some of the individuals analysed. Here, the NS3-1073 specific CD8+ T-cells showed a high expression of this marker (Figure 1c, lower panel) and identified them as potentially being previously exposed to antigen. Calculating the precursor frequency of NS3-1073 specific CD8+ T-cells according to Alanio et al. [16]

revealed a precursor frequency ranging from 5.7e-05 to 3.66e-06 thus being comparable to previously published data. Thus, HCV specific CD8+ T-cells can be found already *ex vivo* in HCV-SN individuals and these cells partially display a memory-phenotype.

HCV-specific CD8+ T-cells from seronegative individuals can be expanded *in vitro*

The initial cohort of healthy blood donors recruited from the local blood donation centre represents a well characterized cohort of individuals, who were stringently tested for diseases including infection or exposure to HCV by HCV-RNA and anti-HCV screening. Nevertheless, a possible contact to HCV explaining the presence of HCV-specific CD8+ T-cells cannot be excluded. Thus, for the following analyses we decided to extend our cohort of HCV-seronegatives by including individuals with different risk factors for HCV exposure such as risk-free individuals (RF, n= 23), healthy blood donors (HBD, n= 49), health care workers (HCW, n= 9), sexual partners of HCV-infected persons (SP, n= 17) and healthy intravenous drug users (HDU, n=4). All individuals were selected in an unbiased fashion only by being HLA-A2-positive and without applying any other selection criteria. Characteristics of the cohort used for the *in vitro* assays are listed in Supplementary Table S1.

We established T cell lines by stimulating CD8+ T-cells or total PBMCs *in vitro* from 91 HCV-SN using HLA-A2 transgenic TAP transporter deficient cells (T2) pulsed with the HCV NS3-1073 peptide for 3 weeks. Corresponding to our *ex vivo* data, we were able to expand HCV NS3-1073 specific CD8+ T-cells in about one third of the individuals tested (36/102 individuals, 35.3%). The magnitude of responses as well the NS3-1073 multimer staining pattern varied widely between individuals (Figure 2a), possibly indicating diverse T cell repertoires. However, no obvious differences in the frequencies, response rates of seronegative individuals or the clonality of the NS3-1073 specific response could be observed between cell lines established with isolated CD8+ T-cells or total PBMCs (average frequencies of NS3-1073 positive CD8+ T-cells 2.3% vs. 1.8% and response rates 18/49 (36.7%) and 18/53 (33.9%) individuals, respectively). The frequencies of NS3-1073 specific CD8+ T-cells after three week *in vitro* stimulation were 2.0% of total CD8+ T-cells in average and even reached 48.8% and 88.8% in two individuals. According to the percentage of NS3-1073 specific CD8+ T-cells all samples tested were classified as non- or weak-responders (<0.5% multimer+ CD8+ T-cells), intermediate responders (0.5%-3% multimer+ CD8+ T-cells) or strong responders (>3% multimer+ CD8+ T-cells). Altogether 29 individuals were assigned as intermediate responders (28.4%), while 7 individuals showed a strong response (6.9%; see supplementary table S1). Interestingly, response rates did not vary

between groups with different risk factors for HCV exposure (RF: 8/23, 34.8%; HBD: 16/49, 32.6% and potentially exposed (PE) individuals (SP, HCW and HDU: 12/30, 40.0%; see Figure 2b). Even more, the two individuals among the strongest responders were risk-free individuals (RF1 and RF 25). Similarly, the median percentages of NS3-1073 specific CD8+ T-cells after three weeks of culture were comparable between those groups (RF: 0.3%; HBD: 0.19% and PE: 0.35%).

The presence of NS3-1073 specific CD8+ T-cells in HCV-SN individuals and the fact that they partially were memory cells suggests that these HCV responses were generated by cross-reactive cells originating from previous heterologous infections. To further strengthen this assumption, we had the opportunity to analyze cells from cord blood samples for the presence of HCV-specific T-cells. Memory-type T-cells are present in cord blood only in low frequency and thus, no significant numbers of memory T-cells cross-reactive to NS3-1073 should be present. We established NS3-1073 specific T-cell lines from 5 HLA-A2 positive cord blood samples. In none of the samples analyzed we were able to expand NS3-1073 specific CD8+ T-cells during our three week *in vitro* stimulation (data not shown).

Importantly, three risk-free individuals were studied shortly after an acute EBV infection. All three individuals responded to NS3-1073 *in vitro*, though in varying magnitudes (Figure 2c). For two of these individuals follow-up samples were collected over a period of several years after the infection. Both individuals continually showed expansion of NS3-1073 specific cells *in vitro*. However, frequencies of multimer+ cells declined over the years (Figure 2d) and also the T-cell repertoire changed significantly (see acEBV1, Supplementary Table S2). Thus, CD8+ T-cell responses against HCV are abundant in seronegative individuals and are present irrespective of possible previous exposure to HCV. Further, NS3-1073 specific T-cell responses in seronegatives can show great fluctuations over time.

HCV-specific CD8+ T-cells from HCV-seronegatives produce cytokines in response to peptide stimulation

To investigate whether the HCV NS3-1073 specific CD8+ T-cells are able to exert effector functions, we also analyzed their response to peptide stimulation by measuring the production of the cytokines IFNγ, TNF and MIP-1β and determined their cellular cytotoxicity by analyzing degranulation via surface CD107 expression. The majority of the *in vitro* cell lines analyzed (11/16) were capable of responding with IFNγ production upon re-challenge with the NS3-1073 peptide (Figure 3a). However, the strength and the quality of the response varied considerably between individuals. In most cases, only a fraction of the NS3-1073

specific cells produced IFNγ in response to peptide (ratio between multimer+ CD8+ T-cells and IFNγ+ CD8+ T-cells). Further, only in some samples we could detect multiple effector functions upon re-stimulation with NS3-1073, in six cases triple responses (IFNγ, TNF and MIP-1β) were seen and three samples showed quadruple responses (IFNγ, TNF, MIP-1β and CD107a).

Likewise, the avidity of the IFNγ response towards different peptide concentrations varied between the NS3-1073 specific T-cell lines. While in some cases the avidity was intermediate with EC50 values for IFNγ responses below 1µg/ml peptide, others only were able to respond to the higher dosages with the calculated EC50 lying above 1µg/ml (Figure 3b). Hence, NS3-1073 specific T-cells from HCV-SN are functional and can respond with cytokine production.

The TCR repertoires of HCV specific CD8+ T-cells in HCV-seronegatives show private specificities

The origin of the memory phenotype HCV NS3-1073 specific CD8+ T-cell in HCV-SN is still unclear. Previous publications have shown that T-cells can be cross-reactive thereby altering the response to other peptides and differentially shaping the T-cell receptor (TCR) repertoire [17]. Thus, we next aimed to investigate the TCR repertoire of the NS3-1073 specific CD8+ T-cell response after 3 weeks *in vitro* stimulation. Therefore, multimer+ NS3-1073 specific CD8+ T-cells were isolated by FACS sorting from *in vitro* cell lines of 14 HCV-SN (4 RF, 4 HBD, 1 SP, 2 HCW and 3 acEBV) and TCR α and β chains were sequenced by a combination of an unbiased anchored reverse transcription method followed by an unrestricted TCR-specific PCR.

The clonal composition of the TCR β chain repertoires showed to be rather different between individuals. Each individual showed a unique clonal dominance pattern and usage of individual clones and Vβ and Jβ families. While in some cell lines a broad and polyclonal repertoire was visible, we found a highly skewed oligoclonal TCR usage with one clone amounting to more than 75% of the response suggesting a memory like response (e.g. HCW4, HBD21, Figure 4a). This divergent repertoire among individuals is a phenomenon known as private specificity [18]. Yet, we could also find features of public specificity in these *in vitro* responses as seen by a common usage of certain Vβ gene families for the NS3-1073 specific response. Here, Vβ6 and Vβ4 were the most dominant family (Figure 4b) and often the clone dominating used Vβ6. The clonality of TCR alpha chains was generally more broad and polyclonal as compared to the TCR beta chain usage. We observed a prevailing usage of the

Vα4 gene family with more than one third of all clones using this particular variant (Figure 4c). Again, no significant differences in the TCR repertoires could be found between cell lines established with isolated CD8+ T-cells or total PBMCs. Taken together, these data show that HCV specific CD8+ T-cells from HCV-SN show private specificities with often restricted TCR repertoires *in vitro,* but also public motifs of TCR usage can be found.

Common TCR motifs determine the functionality of HCV NS3-1073 specific CD8+ T-cells

We next asked whether a common motif in the composition of the hypervariable CDR3β was visible, as it can be found e.g. in the case of the Influenza A specific T-cell response (Vβ17-IRSS-like motif, [19]). However, no ubiquitous motif of the CDR3 could be found. However, some amino acid motifs within the CDR3β region were detectable, which could repeatedly be found among several different though not all individuals. Here, clones bearing the amino acid pattern ExAG-, xxGAP-, PxTGG- and clones carrying multiple glycines could be identified (see Table 1). The sequences of all T-cell clones identified in *in vitro* cell lines are summarized in the Supplementary Table S2.

Grouping all samples analyzed in regard to the functionality of these cells we asked whether differences in the CDR3 could be seen there. As in most cases the frequencies of IFNγ+ CD8+ T-cells was lower than the percentages of NS3-1073 multimer+ CD8+ T-cells seen *in vitro*, we stratified the quality of the responses according to the correlation between frequencies of multimer+ CD8+ T-cells and IFNγ+ CD8+ T-cells. Samples, where more than one fourth of NS3-1073 specific CD8+ T-cells responded with IFNγ production were classified as IFNγ intermediate/high responder (n= 7). Interestingly, we could find a correlation between the Vβ family usage and the IFNγ production capacity. The frequency and dominance of Vβ4, Vβ6 and Vβ24 usage was associated with a strong IFNγ production as seen in the cytokine assay for most individuals (Figure 4d). As no common CDR3 motifs could be found within these two groups, the main factor might be the composition of the other hypervariable regions CDR1 and CDR2. Interestingly, Vβ4, Vβ6 and Vβ24 share certain similarities in the amino acid composition of the CDR1 using a QxMxH-NxMY motif, while those of the other Vβ families associated with poor IFNγ response (e.g. Vβ9, Vβ29 or Vβ12) show clearly different amino acid sequences (see Supplementary Table S3). Thus, functionality of HCV-specific CD8+ T-cells from HCV-SN seems to be at least partially dependent on the composition of the TCR and Vβ gene usage.

HCV NS3-1073 specific CD8+ T-cells from HCV-seronegatives are cross-reactive to EBV, Influenza and HIV-derived peptides

The narrowed clonal dominances of the HCV NS3-1073 specific TCR repertoires in HCV-SN and the strong response in patients after acute EBV infection suggest that cross-reactive T-cells might play a role in the generation of the response. We thus aimed to further analyze whether a cross-recognition of other potentially cross-reactive peptides could be found by testing HCV NS3-1073 specific T-cell lines generated *in vitro* in a cytokine and degranulation assay.

As the identification of such peptide candidates by mere comparison of amino acid sequences is not accurate enough and testing of numerous candidates would be inefficient, different HLA-A2 restricted peptides were selected by an *in silico* structure-based approach (Figure 5a). Nine 3D structures of pMHC complexes presenting selected peptides (see Supplementary Table S4) were recovered from CrossTope Data Bank [20]. Electrostatic potential distribution over the peptide:MHC surface were used as input to perform a structure-based Hierarchical Cluster Analysis (HCA). In this analysis, the previously suggested cross-reactive peptide candidate Influenza A Virus (IAV) NA-231 [11], Epstein-Barr-virus (EBV) LMP2-329 and human immunodeficiency virus (HIV) GAG-77 [21] were included. The well-known cross-reactive variant of HCV (GT1b) was used as a positive control. Other four EBV-derived epitopes were also taken into account in order to search for new possible cross-reactive candidates.

Our results indicated only the LMP2-329 and the Gag-77 epitopes as possible cross-reactive targets for HCV NS3-1073, in agreement with an HCA previously published by Antunes et al. 2011 [21]. Interestingly, despite the similarity in amino acid sequence, the NA-231 was less similar to NS3-1073 as also were the other four EBV epitopes. Stimulation of NS3-1073 specific cell lines with the EBV LMP2-329 peptide showed that indeed HCV-specific T-cells were partially able to respond with IFNγ production when re-stimulated with the LMP2 peptide (see Figure 5b upper panel). To exclude the potential influence of LMP2 specific memory T-cells surviving the *in vitro* culture, cell lines cultured without any peptide for three weeks were included into the analyses and re-stimulated likewise with the respective peptides during the cytokine assay. Responses to peptide candidates were considered as cross-reactive responses if the IFNγ production by NS3-1073 specific *in vitro* cell lines in response to the peptide candidate was higher as compared to the medium control cell lines likewise re-stimulated with peptides. Here, in 5 out of 22 cases NS3-1073 specific cells were able to recognize also the LMP2-329 peptide (Figure 5b, lower panel and Figure 5c). However, only

a fraction of the NS3-1073 specific cells were able to respond to the LMP2 peptide, the production of IFNγ usually was less than towards the original NS3-1073 peptide. In two cases we were able to detect also a marginal cross-reactivity to the IAV NA-231 peptide and cytokine responses were usually weak. Only in few individuals a weak *in vitro* cross-reactivity could be observed for another peptide candidate derived from the HIV-Gag protein (e.g. RF22, Figure 5c).

Taken together, these data indicate that HCV NS3-1073 specific CD8+ T-cells from HCV-SN are able to cross-recognize multiple other peptides and thus might possibly be generated due to cross-reactivity towards these tested peptide candidates or towards other unknown epitopes derived from heterologous pathogens.

Cross-reactive T-cell responses influence responses to vaccination

The presence of memory T-cells that are potentially cross-reactive to other pathogenic epitopes might have an important impact on other infections or vaccines. We had the unique opportunity to analyze the NS3-1073 specific CD8+ T-cell responses in six healthy HCV-seronegative individuals who received an experimental HCV peptide vaccine which also includes the NS3-1073 epitope. The magnitude of NS3-1073 specific responses and the time of their emergence were analyzed, considering that T-cells cross-reactive to NS3-1073 might alter the vaccine response. The *ex vivo* frequency of HCV NS3-1073 specific CD8+ T-cells was determined and cells were expanded *in vitro*.

Three individuals analyzed already showed a growth of NS3-1073 specific CD8+ T-cells *in vitro* before first administration of the vaccine (see individuals RF22, RF24 and RF25, Figure 6a). Here, the vaccine response and proliferation of cells *in vitro* occurred earlier already being detectable and strong at visit 08 and with a higher frequency as compared to two other individuals where no *in vitro* proliferation of NS3-1073 specific before vaccination could be observed (see individuals RF23, RF26 and RF27, Figure 6a). However, as these are *in vitro* responses which might be affected by the cell culture, we also wanted to know whether the actual vaccine response differed between these individuals. Thus, the *ex vivo* frequencies of NS3-1073 specific CD8+ T-cells and their dynamics in the vaccinated individuals was analyzed and individuals were grouped according to whether a NS3-1073 responsiveness was visible *in vitro* already before vaccination (visit 00, Figure 6b). Similarly, the *in vivo* response to the vaccine also showed clear differences between these individuals, as the increase of NS3-1073 specific CD8+ T-cells seen by *ex vivo* staining was stronger for those subjects who responded to the NS3-1073 peptide *in vitro* at the time point before first

vaccination. Those three individuals that were *in vitro* non-responder to NS3-1073 before vaccination did also mount NS3-1073 specific CD8+ T-cells, but the *ex vivo* frequencies did not show significant increases during the early time points and also responses occurring later during vaccination did not reach the same magnitudes in comparison.

For those three individuals being *in vitro* responder to NS3-1073 before vaccination we also investigated the functionality of the cell lines. In two of the individuals a moderate cross-reactivity to the EBV LMP2-329 peptide could be observed (compare Figure 5c, RF22 and RF25). In case of RF22 also a low cross-reactivity towards the other peptide candidate derived from IAV NA-231 was visible. Overall, these data suggest that pre-existing HCV-specific CD8+ T-cells may influence the strength and time of appearance of vaccine responses.

Discussion

The natural course of viral infections is often surprisingly diverse and varies between individuals. One possible reason might be that pre-existing T-cells specific for these viruses are involved in the immune response thereby altering the outcome of infection. Here we demonstrate that HCV-specific CD8+ T-cells are present in unexposed individuals and that these T-cells partially display a memory phenotype. We observed a high variability of HCV-specific T-cell repertoires in unexposed individuals and a partial cross-reactivity towards unrelated peptides. In addition and importantly, we show that HCV-specific CD8+ T-cell responses in HCV seronegatives may impact the response to vaccination with a peptide-based vaccine.

The presence of HCV-specific T-cells in both uninfected and in unexposed HCV-seronegative individuals has been described previously [8,9,22-24]. Different reasons for the existence of these cells are under debate including low-level exposure to HCV without seroconversion [25], loss of HCV-antibodies after infection [26], presence of naïve precursor T-cells [27] and T-cell cross-reactivity. Of note, our data show that a large fraction of the NS3-1073 specific CD8+ T-cells displayed a memory-like phenotype in some individuals as judged by the expression of CD45RA and CCR7. On the other hand, in other individuals the frequency of naïve cells was prevailing. However, expression of CCR7 and CD45RA might not always indicate bona-fide naïve T-cells [15]. Thus, we included the additional memory marker CD11a and found high CD11a expression in some individuals with a predominant CCR7+ CD45RA+ T-cell population. CD11a is highly expressed on memory T-cells and low or absent on naïve T-cells [28]. Thus, we suggest that a large portion of HCV NS3-1073 specific cells display a phenotype of previously activated memory T-cells. Supporting these findings, we could expand HCV NS3-1073-specific CD8+ T-cells *in vitro* in about one third of HCV seronegative individuals. The presence of memory cells *ex vivo* and the variability of the *in vitro* response argues against the exclusive involvement of naïve precursor cells as suggested previously [27]. Moreover, we could not find a significant impact of the risk factors for HCV exposure on the frequency or magnitude of *in vitro* HCV-specific CD8+ T-cell responses. Importantly, those individuals showing the highest response were primarily risk-free individuals and three individuals shortly after acute EBV infection. This finding might argue against an impact of low-level exposure and also discourages the idea of a potential loss of anti-HCV antibodies after infection, especially as the risk-free individual with the strongest response is aged below 30 years and a loss of HCV-antibodies was shown to occur after decades only [26]. Further and importantly, we had the opportunity to test cord blood cells

from newborns for their *in vitro* responsiveness to the NS3-1073 peptide. Here, we were not able to see any expansion of cells using our cell culture method, a finding that also suggest that naïve precursor cells are not the main reason for the presence of HCV-specific CD8+ T-cells in HCV-SN.

The *in vitro* proliferation of HCV-specific memory CD8+ T-cells was further supported by the T cell receptor repertoire (TCR) data. High purity cell sorting of NS3-1073 specific CD8+ T-cells after *in vitro* expansion and subsequent sequencing of the TCR β and α chains revealed a frequent strong skewing and clonal focussing of the TCR repertoire with a single T-cell clone clearly dominating the response. This finding is intriguing, as a normal naïve T-cell response usually comprises multiple clones with a fractal distribution and no clear shift of clonality or a single clone dominating [29]. Arguably, this might be an effect of the *in vitro* stimulation of cells for three weeks. However, a previous report shows that no skewing of the TCR repertoire occurs through *in vitro* stimulation of T-cells under optimized conditions [30]. However, it needs to be considered that the NS3-1073 peptide is not the cognate peptide for these responses and thus, the optimal concentration cannot be determined. Thus, a limitation of our studies is that the NS3-1073 peptide concentration used here for establishing T-cell lines was originally determined using T-cells isolated from HCV-infected patients, which might not represent the optimal concentration in HCV-SN. The optimal condition for cross-reactive responses in seronegatives is not known and difficult to explore. However, cell lines derived from several HCV-SN in our cohort also showed polyclonal TCR repertoires after our *in vitro* culture.

The restricted TCR clonality of CD8+ T-cells in response to NS3-1073 *in vitro* and the high level of private specificity of the TCR suggest that only few T-cells are selectively expanding. This lead us to investigate whether the mechanism of cross-reactivity might be the rationale for the existence of HCV-specific CD8+ T-cells in HCV-SNs. Opposing to the initial dogma that one T-cell has only one specificity, several reports of the last years clearly indicate that the TCR-peptide recognition is highly promiscuous and flexible and that one T-cell can recognize multiple peptides sharing certain biochemical properties in amino acid composition and thus are cross-reactive [31]. Recent research using a monoclonal T-cell expressing the 1E6 TCR isolated from a patient with autoimmune type 1 diabetes showed that this single TCR can potentially react to more than one million different peptides [32]. The fact of T-cells being cross-reactive to a number of different peptides [33] forms the basis for the concept of heterologous immunity. Memory T-cells being present due to previously encountered infections become reactivated upon a secondary infection with an unrelated

heterologous pathogen as some memory T-cells are cross-reactive to epitopes encoded by the second pathogen. The identification of potentially cross-reactive peptide candidates is difficult, because comparison of the mere amino acid sequence is not sufficient and cross-reactivity can exist despite little sequence similarities. We therefore employed an *in silico* approach using bioinformatic tools to model the three-dimensional structures and charge distribution of the peptide: HLA-A0201 complexes. This approach has previously been used to successfully predict cross-reactivity among various genotype variants of the HCV NS3-1073 peptide [19], and has also suggested structural similarity between the HCV NS3-1073 variant CVNGVCWTV and two unrelated targets. The first is derived from the Epstein-Barr-Virus (EBV) LMP2 protein (LMP2-329), and the second is derived from the human immunodeficiency virus (HIV)-Gag protein (Gag-77). Intriguingly, the EBV-LMP2 epitope has no similarities in amino acids sequence with the here studied HCV NS3-1073 epitope (CINGVCWTV), and share only 33% identity if biochemical properties are considered. The real similarity between them only arises in a higher level of complexity, when presented in the context of HLA-A*02:01. This is possible because not all amino acids will be exposed and because the charges and the properties of the MHC cleft will interact with the properties of the epitope providing a new combined surface that, in this case, presents a similar structural pattern for both complexes. The Influenza-A virus (IAV) NA-231 epitope, which has been previously described as cross-reactive to HCV NS3-1073 [11,35] presented much smaller structural similarity to the HCV-derived complex despite a sequence similarity of 66%. Testing these peptide candidates in a cytokine assay revealed that indeed in some but not all cases NS3-1073 specific CD8+ T-cells indeed were able to cross-react towards the LMP2 peptide. This is surprising, as even epitopes with no sequence similarity can trigger cross-reactive responses, thus highlighting the prospective potential of structure-based cross-reactivity prediction approaches. However, only a fraction of the NS3-1073 specific cells appeared to be able to respond with IFNγ production towards the LMP2 epitope. Additionally, cross-recognition of IAV NA-231 and HIV Gag-77 was observed, though with lesser extents. The finding that IAV NA-231 with a high amino acid sequence similarity presents low cross-reactivity to NS3-1073 match to the predictions seen in the HCA and might also explain, why other reports described a limited cross-recognition of the NA-231 peptide by NS3-1073 specific CD8+ T-cells in HCV patients [34].

Overall our observations suggest that cross-reactivity seems to be quite a frequent phenomenon. Likewise, a recent report by Su and colleagues [4] showed the presence of CD4+ T-cells specific for pathogen-derived antigens in unexposed individuals. Similar to our

observations, cross-reactivity of HIV-1 and IAV specific CD4+ T-cells towards other epitopes from unrelated pathogens was described. The hypothesis of induction of HCV-specific T-cells in seronegative individuals through cross-reactivity could be further supported by analyses of the TCR repertoires. The skewed and oligoclonal repertoires suggest a selective expansion of single T-cell clones caused by cross-reactivity has been shown before in mice [17]. The CDR3β regions of NS3-1073 specific CD8+ T-cells showed that the repertoire is rather heterogeneous with no single clone or amino acid motif dominating. This is not unexpected, as the TCR usage is known to be unique to every individual and multiple cross-reactive epitopes might exist [17]. Interestingly, among different individuals we found several clones which had multiple glycines in the CDR3, such TCRs are assumed to be more flexible in the binding to the peptide:MHC complex and exert high a high degree of cross-reactivity [35]. However, public CDR3 amino acid motifs and dominant usage of certain Vβ and Vα gene families indicate preferred TCR structures for targeting a specific peptide.

Overall, cross-reactivity in humans is not a new phenomenon, it has been described beforehand between two epitopes derived from EBV and IAV [36] and has also been documented for HCV and IAV [11]. Likewise, our data suggest that a cross-reactive relation exists also between HCV and EBV as suggested by the response in three acute EBV patients. Here, we found a uniform *in vitro* responsiveness to NS3-1073 which declined during the years after acute EBV infection with a changing of the TCR repertoire. These data suggests that cross-reactive responses are very dynamic and are underlying many fluctuations and that the responses in seronegative individuals observed at a given time point might just be snap-shots underestimating the real extent of cross-reactivity. Individual differences in the memory T-cell pool and exposure to varying pathogens might explain the vast heterogeneity in the natural course of diseases as it is also seen for HCV. Cross-reactivity of T-cells seems to be a common phenomenon and is discussed to be even a necessary phenomenon due to the vastness of pathogenic epitopes and the comparably limited T-cell pool present in humans [37].

Considering the concept of Heterologous Immunity, we finally aimed to directly investigate a possible clinical impact of pre-existing HCV-specific CD8+ T-cells in unexposed individuals. We therefore used samples from healthy individuals which have been vaccinated with a HCV peptide vaccine candidate, a formulation which also comprises the NS3-1073 epitope [13]. Here, we could observe strong individual differences in the magnitude of *in vitro* expansion of NS3-1073 specific CD8+ T-cells in six individuals analyzed longitudinally before, during and after vaccination. This is at first not surprising and

might be attributed to the method of vaccination. However, we could see clear differences in the *in vivo* response to the vaccine when we grouped individuals according to the fact whether they showed an *in vitro* response to NS3-1073 already before the first vaccine dosage. Those individuals being *in vitro* responder before vaccination generally had an earlier and stronger responsiveness to the vaccine as seen by *ex vivo* staining for NS3-1073 specific CD8+ T-cells as compared to individuals with no *in vitro* response. This finding indicates that pre-existing HCV-specific CD8+ T-cells in unexposed are able to influence an immune response towards a vaccine. How this might impact on the natural course of an HCV infection is unclear. Cross-reactive responses might have different effects on infections in humans ranging from protection or enhanced immunopathology [38], effects which have also been documented for mice [39].

The specific TCR repertoire of the cross-reactive response may be of importance. In this context interesting, we observed an apparent impact of the Vβ family usage on the functionality of NS3-1073 specific CD8+ T-cells. Usage of certain Vβ families (Vβ4 and Vβ6) was associated with the ability to respond with IFNγ production towards the peptide. This shows the importance of the CDR1 and CDR2 regions, here we were able to find a common amino acid motif between the positively associated Vβ families, which was clearly different among those Vβ families correlating with a low or absent IFNγ response. A previous infection might generate T-cells cross-reactive to HCV NS3-1073 using either beneficial or non-beneficial Vβ families. Thus, these memory cells could easily become re-activated and consequently influence the recruitment of naïve T-cells but might impact the cytokine response. In two cases of risk-free individuals who showed an *in vitro* response to NS3-1073 before vaccination we had the opportunity to analyze the TCR repertoire these cells. Interestingly, these cells were not able to produce IFNγ and in accordance to our previous observations these responses dominantly used Vβ families that were negatively associated with IFNγ response (Vβ11 and Vβ29, data not shown). Thus, under certain conditions immunogenic epitopes may not be beneficial in vaccines and might therefore be considered pathogenic as discussed previously [40].

In conclusion, concepts such as pre-existing memory responses, i.e. due to cross-reactivity, as well as the individual TCR repertoire are important for a number of implications both with regard to immunobiology and practical applications in vaccine design and development.

Materials & Methods

Ethics statement

This study was conducted in accordance with the guidelines of the Declaration of Helsinki. The study was approved by the local ethics committee of Hannover Medical School. All individuals gave written informed consent.

Patient cohort

Heparinized blood samples were collected from 164 HCV-seronegative individuals (HCV-SNs) included in the study. Samples were negatively tested for presence of HCV-RNA and anti-HCV antibodies. HCV-SNs were grouped according to their risk of HCV exposure into risk-free (RF; n=23), healthy blood donors (HBD; n=106) and potentially exposed individuals (PE; n=32). The selection of individuals was done according to the expression of HLA-A2 and no further criteria were taken into account. Additionally, a special pre-selected cohort of individuals with acute EBV infection (acEBV; n=3) was included. HBDs were recruited from Hannover Medical School blood donation centre. All blood donors were stringently tested for diseases, including infection or exposure to HCV by HCV-RNA and anti-HCV screening. PEs included professional health care workers (HCW; n=11), sexual partners of HCV-infected individuals (SP; n=17) and healthy intravenous drug users (HDU; n=4). All individuals were tested HLA-A2 positive from whole blood by flow cytometry using anti-HLA-A2 antibodies (BioLegend, San Diego, CA, USA). Characteristics of all individuals included in this study are summarized in supplementary table S1.

Isolation of peripheral blood mononuclear cells

Peripheral blood mononuclear cells (PBMCs) were isolated using the standard Ficoll Hypaque Density Centrifugation method (BioColl Separating Solution, Biochrom AG, Berlin, Germany). Cells were either used directly *ex vivo* for experiments or cryopreserved in freezing medium containing 10% DMSO (Sigma-Aldrich, St. Louis, MO, USA) in liquid nitrogen.

Detection of HCV-RNA in PBMCs

Extraction of viral RNA from cyro-preserved PBMCs was performed using the Qiagen miniRNA kit (QIAGen, Hilden, Germany) according to manufacturer's protocol. Synthesis of cDNA was performed using the Superscript III System (Invitrogen, Carlsbad, CA, USA)

according to manufacturer's protocol. HCV-specific primers from HCV core region were designed according to the HCV sequence database (http://hcv.lanl.gov). Detection of HCV RNA was performed by qualitative nested PCR using the Roche-COBAS® AmpliPrep/COBAS® TaqMan® HCV Test (Roche, Mannheim; limit of detection of 15 HCV RNA IU/ml).

Monoclonal antibodies, synthetic peptides and MHC class I multimeric complexes

The following fluorochrome-labeled mouse anti-human monoclonal antibodies were used for cell surface staining: anti-CD8-APC-H7, anti-CD45RA-FITC, anti-CD107a-PE-Cy5, anti-IFNγ-FITC, anti-TNF-APC, anti-MIP-1β-PE (BD Pharmingen, BD Biosciences, La Jolla, CA, USA) and anti-CCR7-PerCP/Cy5.5 (Biolegend, San Diego, CA, USA). Dump channel for exclusion of cell populations was done using anti-CD14, anti-CD19 and anti-CD56 (BD Pharmingen). Synthetic HLA-A*0201 restricted peptides originating from HCV NS3-1073-1081 (CINGVCWTV), human Epstein-Barr Virus (EBV) BMLF1-280-288 (GLCTLVAML), EBV LMP2-329-337 (LLWTLVVLL), Influenza A Virus (IAV) NA-231-239 (CVNGSCFTV) and HIV Gag-77-85 (SLYNTVATL) were used, Tyrosinase 368–376 (YMDGTMSQV) served as a negative control. All peptides were purchased from ProImmune Ltd. (Oxford, UK) with a purity >98% and were dissolved in sterile endotoxin-free DMSO (Sigma-Aldrich). Final concentrations of peptides used depended on individual titration experiments and was 1µg/ml EBV BMLF1-280 and 10µg/ml for EBV LMP2-329, Influenza-A NA-231 and HCV NS3-1073. PE-labelled HLA-A*0201 restricted MHC class I multimeric complexes (multimers) of HCV NS3-1073 were purchased from Beckman Coulter (Fullerton, CA, USA) and Immudex (Copenhagen, Denmark). Peptide sequence of the multimer was identical to the synthetic peptide used.

Cell surface and multimer staining

For cell surface staining $0.3x10^6$ PBMCs were washed with FACS Buffer (PBS + 2% FCS) and stained with the respective antibodies for 10-15 minutes at 4°C. After washing thrice cells were resuspended and analyzed by flow cytometry (BD FACS Calibur or BD FACS CantoII, BD Biosciences). For detection of antigen-specific CD8+ T-cells, samples were stained with multimer for 20 minutes at room temperature followed by cell surface staining as described above. All samples were analyzed within 30 minutes by flow cytometry. Analysis of flow cytometry data was done using FlowJo Software (TreeStar Inc., Ashland, OR, USA).

In vitro long-term T-cell lines

For establishing long-term antigen-specific T-cell lines, CD8+ T-cells were isolated from fresh PBMCs using magnetic beads isolation Kit (Miltenyi Biotech, Bergisch Gladbach, Germany) according to manufacturer's protocol. $1x10^6$/well isolated CD8+ T-cells were stimulated with $0.2x10^6$/well previously irradiated (40 Gy) HLA-A*0201-transgenic TAP transporter deficient T2 cells loaded with peptides. Cryopreserved PBMCs were thawed and plated at $4x10^6$ cells/well together with $0.2x10^6$/well T2 cells and peptides. Every 7 days cells were harvested and again plated at $1x10^6$ cells/ml with T2 cells and peptides. Cells were kept in CTL medium (AIM-V (Invitrogen, Carlsbad, CA, USA) containing 10% human AB serum (Cambrex, East Rutherford, NJ, USA), 100U/ml Penicillin, 0.1mg/ml Streptomycin, 1% MEM non-essential amino acids, 1% sodium pyruvate and 2mM L-Glutamine (PAA Laboratories GmbH, Parching, Austria) and 5U/ml rhIL-2 (Invitrogen)) and the medium was changed every 3-4 days. Growth and functionality of antigen-specific CD8+ T-cells was analyzed after 21 days of culture, cell lines with frequencies of NS3-1073+ CD8+ T-cells higher than 0.5% and distinct multimer staining pattern were considered as responder. Characteristics of all samples and cell culture method used for *in vitro* analyses are given in Supplementary Table S1.

Multimer associated magnetic bead enrichment

For the analysis of the memory phenotype of HCV-multimer+ CD8+ T-cells were magnetically enriched as described by Alanio et al [16]. Donors were selected previously according to the detectability of NS3-1073 specific CD8+ T-cells *ex vivo*. PBMCs were stained with HCV NS3-1073 multimer for 30 minutes at room temperature. After washing and incubation with anti-PE microbeads (Miltenyi Biotech) cells were isolated according to manufacturer's instructions. Multimer-enriched cells were stained for cell surface expression of CCR7 and CD45RA and partially for CD11a, and analyzed by flow cytometry. Gating of memory cell populations was done according to the staining of bulk CD8+ T-cells.

Analysis of functionality of antigen-specific T-cells

For the detection of *ex vivo* NS3-1073 specific responses $1x10^5$ PBMCs were stimulated using respective peptides and analyzed using an IFNγ ELISpot Assay as described previously [41]. For the analysis of T-cell lines a cytokine and degranulation assay was performed. $0.3x10^6$ cells per well were stimulated with peptides for six hours in the presence

of anti-CD107a. Brefeldin A (2µg/ml; Sigma-Aldrich) was added during the last five hours. Cells were then washed with FACS Buffer and stained for CD8, CD107a and exclusion markers. Cell fixation and permeabilization was performed by using BD CytoFix/CytoPerm kit (BD Biosciences) according to manufacturer's protocol. Intracellular staining of IFNγ, TNF and MIP-1β was analyzed by flow cytometry. Cell function was analyzed by comparing NS3-1073 stimulated T-cell lines with non-stimulated medium control cell lines after re-stimulation with the same peptide. Responses of NS3-1073 stimulated cell lines were considered as significant if they were more than 2 folds higher compared to medium control cell lines from the same donor re-stimulated with the same peptide.

Avidity assays

The analysis of avidities of CD8+ T-cells for the respective epitope was performed by intracellular cytokine staining as described above after re-stimulating T-cells lines with serial dilutions of the NS3-1073 peptide. For the judgement of the avidity the EC50 was calculated, samples were considered to have a high avidity to the NS3-1073 peptide when the EC50 was lower than 1µg/ml peptide.

***In silico* structure-based cross-reactivity prediction**

Nine 3D structures of pMHC complexes presenting selected peptides (see Supplementary Table S4) were recovered from CrossTope Data Bank [20]. Electrostatic potential distribution over the peptide:MHC surface was computed with GRASP2 program and images of the TCR-interacting surfaces were used as input to perform a structure-based Hierarchical Cluster Analysis (HCA). The HCA was performed with SPSS software (PASW Statistics 18, IBM, Chicago, IL, USA), using values extracted from seven selected regions, as previously described [21].

Cell sorting and RNA extraction

After 21 days *in vitro* stimulation cells were stained with the respective multimer, CD8 and exclusion markers. Multimer+ CD8+ T-cells were then isolated by high-resolution flow cytometric sorting at the Central Sorting Facility of Hannover Medical School using a BD FACS Aria (BD Biosciences). Purity of sorted multimer+ CD8+ T-cells was higher than 95%. mRNA was then isolated using the Oligotex Direct mRNA kit (QIAGEN, Hilden, Germany) according to manufacturer's protocol. Isolated mRNA was stored at -80°C until further use.

Analysis of T-cell receptor using template-switch anchored RT-PCR and sequencing

T-cell receptor analysis was performed by using an adapted SMART-RACE method as described by MF Quigley et al. [42] using the SMARTer RACE cDNA Amplification Kit (Clontech Laboratories Inc.). Touchdown PCR was performed with Advantage 2 PCR Kit (Clontech) with custom-synthesized primers specific for the constant region of the TCR α and β chain under the following conditions: 1 cycle 95°C for 30 seconds, 5 cycles at 95°C for 5 seconds and 72°C for 2 minutes, 5 cycles at 95°C for 5 seconds, 70°C for 10 seconds and 72°C for 2 minutes, 30 cycles at 95°C for 5 seconds, 68°C for 10 seconds and 72°C for 2 minutes, 1 cycle of 72°C for 10 minutes.

Gel-purified SMART-RACE products were cloned into the pCR4®-TOPO® vector using the TOPO TA Cloning Kit (Invitrogen) according to manufacturer's instructions. For propagation, the plasmids were transformed into TOP10® chemically competent *E. coli* (Invitrogen) and spread on a LB agar plate. After incubation at 37°C over night single bacterial colonies were selected, for each sample 50 clones were sequenced and analyzed. Alignment and identification of TCR gene family usage and of CDR3α and CDR3β regions was done according to IMGT standard sequences (http://www.imgt.org) using the Sequencher software (GeneCodes, Ann Arbor, MI, USA). Clones with identical nucleic acid sequences were regarded as one clone. Frequency of individual clonotype distribution within the sequenced clones was analyzed and a ratio of clonality was calculated by dividing the number of unique clonotypes identified by the total number of sequences analyzed. All TCR sequencing data obtained are summarized in Supplementary Table S3.

Acknowledgements

The Authors want to thank Liisa K. Selin (Department of Pathology, University of Massachusetts Medical School, Worcester, Massachusetts, USA) for helpful discussion of data.

References

1. Bengsch B, Thimme R, Blum HE (2009) Role of host genetic factors in the outcome of hepatitis C virus infection. Viruses 1: 104-125.
2. Kimman TG, Vandebriel RJ, Hoebee B (2007) Genetic variation in the response to vaccination. Community Genet 10: 201-217.
3. Benn CS, Netea MG, Selin LK, Aaby P (2013) A small jab - a big effect: nonspecific immunomodulation by vaccines. Trends Immunol.
4. Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM (2013) Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. Immunity 38: 373-383.
5. Yang J, James E, Roti M, Huston L, Gebe JA, et al. (2009) Searching immunodominant epitopes prior to epidemic: HLA class II-restricted SARS-CoV spike protein epitopes in unexposed individuals. Int Immunol 21: 63-71.
6. Ritchie AJ, Campion SL, Kopycinski J, Moodie Z, Wang ZM, et al. (2011) Differences in HIV-specific T cell responses between HIV-exposed and -unexposed HIV-seronegative individuals. J Virol 85: 3507-3516.
7. Cerny A, McHutchison JG, Pasquinelli C, Brown ME, Brothers MA, et al. (1995) Cytotoxic T lymphocyte response to hepatitis C virus-derived peptides containing the HLA A2.1 binding motif. J Clin Invest 95: 521-530.
8. Koziel MJ, Wong DK, Dudley D, Houghton M, Walker BD (1997) Hepatitis C virus-specific cytolytic T lymphocyte and T helper cell responses in seronegative persons. J Infect Dis 176: 859-866.
9. Scognamiglio P, Accapezzato D, Casciaro MA, Cacciani A, Artini M, et al. (1999) Presence of effector CD8+ T cells in hepatitis C virus-exposed healthy seronegative donors. J Immunol 162: 6681-6689.
10. Thurairajah PH, Hegazy D, Chokshi S, Shaw S, Demaine A, et al. (2008) Hepatitis C virus (HCV)--specific T cell responses in injection drug users with apparent resistance to HCV infection. J Infect Dis 198: 1749-1755.
11. Wedemeyer H, Mizukoshi E, Davis AR, Bennink JR, Rehermann B (2001) Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells. J Virol 75: 11392-11400.
12. Koziel MJ, Dudley D, Afdhal N, Grakoui A, Rice CM, et al. (1995) HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. J Clin Invest 96: 2311-2321.
13. Firbas C, Boehm T, Buerger V, Schuller E, Sabarth N, et al. (2010) Immunogenicity and safety of different injection routes and schedules of IC41, a Hepatitis C virus (HCV) peptide vaccine. Vaccine 28: 2397-2407.
14. Firbas C, Jilma B, Tauber E, Buerger V, Jelovcan S, et al. (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.
15. Gattinoni L, Lugli E, Ji Y, Pos Z, Paulos CM, et al. (2011) A human memory T cell subset with stem cell-like properties. Nat Med 17: 1290-1297.
16. Alanio C, Lemaitre F, Law HK, Hasan M, Albert ML (2010) Enumeration of human antigen-specific naive CD8+ T cells reveals conserved precursor frequencies. Blood 115: 3718-3725.
17. Cornberg M, Chen AT, Wilkinson LA, Brehm MA, Kim SK, et al. (2006) Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. J Clin Invest 116: 1443-1456.
18. Welsh RM (2006) Private specificities of heterologous immunity. Curr Opin Immunol 18: 331-337.

19. Lehner PJ, Wang EC, Moss PA, Williams S, Platt K, et al. (1995) Human HLA-A0201-restricted cytotoxic T lymphocyte recognition of influenza A is dominated by T cells bearing the V beta 17 gene segment. J Exp Med 181: 79-91.
20. Sinigaglia M, Antunes DA, Rigo MM, Chies JA, Vieira GF (2013) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. Database (Oxford) 2013: bat002.
21. Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, et al. (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.
22. Semmo N, Barnes E, Taylor C, Kurtz J, Harcourt G, et al. (2005) T-cell responses and previous exposure to hepatitis C virus in indeterminate blood donors. Lancet 365: 327-329.
23. Zeremski M, Shu MA, Brown Q, Wu Y, Des Jarlais DC, et al. (2009) Hepatitis C virus-specific T-cell immune responses in seronegative injection drug users. J Viral Hepat 16: 10-20.
24. Choi YS, Lee JE, Nam SJ, Park JT, Kim HS, et al. (2013) Two Distinct Functional Patterns of Hepatitis C Virus (HCV)-Specific T Cell Responses in Seronegative, Aviremic Patients. PLoS One 8: e62319.
25. Meyer MF, Lehmann M, Cornberg M, Wiegand J, Manns MP, et al. (2007) Clearance of low levels of HCV viremia in the absence of a strong adaptive immune response. Virol J 4: 58.
26. Takaki A, Wiese M, Maertens G, Depla E, Seifert U, et al. (2000) Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C. Nat Med 6: 578-582.
27. Schmidt J, Neumann-Haefelin C, Altay T, Gostick E, Price DA, et al. (2011) Immunodominance of HLA-A2-restricted hepatitis C virus-specific CD8+ T cell responses is linked to naive-precursor frequency. J Virol 85: 5232-5236.
28. Hamann D, Baars PA, Rep MH, Hooibrink B, Kerkhof-Garde SR, et al. (1997) Phenotypic and functional separation of memory and effector human CD8+ T cells. J Exp Med 186: 1407-1418.
29. Naumov YN, Naumova EN, Hogan KT, Selin LK, Gorski J (2003) A fractal clonotype distribution in the CD8+ memory T cell repertoire could optimize potential for immune responses. J Immunol 170: 3994-4001.
30. Naumov YN, Naumova EN, Clute SC, Watkin LB, Kota K, et al. (2006) Complex T cell memory repertoires participate in recall responses at extremes of antigenic load. J Immunol 177: 2006-2014.
31. Welsh RM, Selin LK (2002) No one is naive: the significance of heterologous T-cell immunity. Nat Rev Immunol 2: 417-426.
32. Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, et al. (2012) A single autoimmune T cell receptor recognizes more than a million different peptides. J Biol Chem 287: 1168-1177.
33. Cornberg M, Clute SC, Watkin LB, Saccoccio FM, Kim SK, et al. (2010) CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections. J Immunol 184: 2825-2838.
34. Kasprowicz V, Ward SM, Turner A, Grammatikos A, Nolan BE, 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.
35. Naumov YN, Naumova EN, Yassai MB, Kota K, Welsh RM, et al. (2008) Multiple glycines in TCR alpha-chains determine clonally diverse nature of human T cell memory to influenza A virus. J Immunol 181: 7407-7419.

36. Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, et al. (2005) Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. J Clin Invest 115: 3602-3612.
37. Sewell AK (2012) Why must T cells be cross-reactive? Nat Rev Immunol 12: 669-677.
38. Urbani S, Amadei B, Fisicaro P, Pilli M, Missale G, et al. (2005) Heterologous T cell immunity in severe hepatitis C virus infection. J Exp Med 201: 675-680.
39. Chen AT, Cornberg M, Gras S, Guillonneau C, Rossjohn J, et al. (2012) Loss of anti-viral immunity by infection with a virus encoding a cross-reactive pathogenic epitope. PLoS Pathog 8: e1002633.
40. Welsh RM, Fujinami RS (2007) Pathogenic epitopes, heterologous immunity and vaccine design. Nat Rev Microbiol 5: 555-563.
41. Wiegand J, Cornberg M, Aslan N, Schlaphoff V, Sarrazin C, et al. (2007) Fate and function of hepatitis-C-virus-specific T-cells during peginterferon-alpha 2b therapy for acute hepatitis C. Antiviral Therapy 12: 303-316.
42. Quigley MF, Almeida JR, Price DA, Douek DC (2011) Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. Curr Protoc Immunol Chapter 10: Unit10 33.

Figure Legends

Figure 1

***Ex vivo* characterization of HCV NS3-1073 specific CD8+ T-cells in HCV-SN.** (**A**) PBMCs from healthy HCV-seronegative individuals were stained *ex vivo* for the detection of HCV NS3-1073 specific CD8+ T-cells using MHC-1 multimers. FACS plots of six selected HVC-SN are shown, gates include NS3-1073 specific CD8+ T-cells gated on T-cells after exclusion of dump channel. Numbers indicated display the frequencies of mulitmer-positive CD8+ cells referring to CD8+ T-cells. The gating strategy employed for all analyzes is depicted (upper panel) with selection of lymphocytes, exclusion of CD14+CD19+CD56+ cells (dump channel) and gating on CD8+ T-cells. (**B**) Summary of *ex vivo* frequencies of HCV NS3-1073 specific CD8+ T-cells detected by multimer staining in all 125 HBDs analysed (left plot). Frequencies given were calculated by the frequencies of NS3-1073 multimer-positive CD8+ T-cells gated on CD8+ T-cells and subtracting the background of multimer staining (NS3-1073 multimer-positive CD8+ T-cells). Samples were considered as being ex vivo positive for NS3-1073+ CD8+ T-cells if the frequency was at least 0.05% of total CD8+ T-cells. Mean frequencies are indicated by a solid horizontal line. Samples considered as positive for NS3-1073 specific CD8+ T-cells *ex vivo* are depicted by open circles. HCV NS3-1073 specific responses were further detected by *ex vivo* IFNγ ELISpot in 18 HBD (right plot). Mean of responses is shown by a solid horizontal line, the horizontal dashed line indicates threshold of 9 SFU/300,000 cells. (**C**) Enrichment of HCV NS3-1073 specific CD8+ T-cells *ex vivo* from HCV-SNs using magnetic bead isolation of multimer+ CD8+ T-cells for analysis of memory phenotypes. The left panel shows representative plots of multimer staining before and after enrichment in two individuals, where the enrichment was successful (HBD33, left plots) or did not result in enumeration of cells (HBD106, right plots). The memory phenotype of NS3-1073 specific CD8+ T-cells was analyzed by co-staining for CD45RA and CCR7 and by staining for CD11a. FACS plots of four representative individuals are shown, numbers indicated indicate the frequencies of NS3-1073+ CD8+ T-cells and frequencies of memory T-cells gated on multimer-positive CD8+ T-cells. The scatter plot (right side) summarizes the frequencies of memory-like T-cells (CCR7+CD45RA- and CCR7-CD45RA+ and CCR7-CD45RA- cells) of the NS3-1073 specific CD8+ T cell populations of all HCV-SN included (n=16).

Figure 2

Detection of HCV NS3-1073 specific CD8+ T-cells in HCV-SN *in vitro*. (**A**) NS3-1073 multimer staining of T-cell lines after *in vitro* stimulation using the NS3-1073 peptide. Representative FACS plots of high (left), intermediate (middle) and low/non-responders are shown (right). Numbers represent frequencies of multimer-positive cells gated on CD8+ T-cells after subtracting values from medium control cell lines. Plots show cell populations gated on total T-cells after exclusion of the dump channel (**B**) Scatter plot summarizing *in vitro* frequencies of NS3-1073 specific CD8+ T-cells in all HCV-SN analysed (n=102). Samples are grouped according to risk factors for HCV exposure (RF = risk free; HBD = healthy blood donor; PE = potentially exposed). Horizontal lines indicate median frequencies of responses. **(C**) *In vitro* expansion of NS3-1073 specific CD8+ T-cells in individuals with acute EBV infection reveals a high responsiveness towards the peptide. FACS plots of multimer+ CD8+ T-cells after three weeks *in vitro* stimulation with NS3-1073 are shown. Plots show cell populations gated on total T-cells after exclusion of the dump channel (**D**) Longitudinal analysis over nine years after acute EBV infection in 2004 in one individual shows a continuous *in vitro* response towards NS3-1073. FACS plots of multimer staining

from three selected time points are shown. Clonalities of the TCR repertoire of NS3-1073 specific CD8+ T-cells for each time point are depicted above as bar graphs.

Figure 3

Functional characterization of *in vitro* HCV NS3-1073 specific CD8+ T-cells from HCV-SN. (**A**) NS3-1073 specific T-cell lines generated *in vitro* respond with production of multiple cytokines and degranulation against the NS3-1073 peptide. Representative FACS plots of IFNγ+, MIP-1β+, TNF+ and CD107+ CD8+ positive T-cells from four individuals are shown. Plots show cell populations gated on total T-cells after exclusion of the dump channel, numbers indicated refer to frequencies of positive cells gated on CD8+ T cell populations (**B**) HCV NS3-1073 specific cells display different avidities towards the NS3-1073 peptide after expansion *in vitro* as measured by serial peptide dilution in a cytokine assay. Representative graphs of six individuals with intermediate/high (upper panel) and low avidity (lower panel) are shown. IFNγ responses of CD8+ T-cells are expressed as stimulation indices referring to medium controls. The calculated EC50 was used to judge on the avidity on NS3-1073 specific CD8+ T-cells *in vitro*, the individual EC50 values are indicated.

Figure 4

TCR repertoires of HCV NS3-1073 specific CD8+ T-cells. (**A**) Clonal compositions of NS3-1073 specific CD8+ T-cell responses of six selected HCV-SN are shown. Bar graphs depict numbers of unique clones found within the NS3-1073 specific CD8+ T-cell populations of *in vitro* T-cell lines. The individual clone ratio (numbers of T-cell clones identified / total number of sequences obtained) is indicated for each individual. (**B**) NS3-1073 specific CD8+ T-cell responses show public pattern with a preferential Vβ family usage. TCR repertoires of representative individuals are shown as pie charts, frequencies of different Vβ family usages are summarized for each individual, respectively. Summary of Vβ family usages of all individuals analyzed are given in a scatter plot revealing a preferential usage of Vβ4 and Vβ6. Standard deviation is indicated by whiskers. (**C**) TCR Vα repertoires of NS3-1073 specific CD8+ T-cell responses show a preferential Vα family usage. TCR repertoires of six representative individuals are shown as pie charts, frequencies of different Vα families usages are summarized for each individual, respectively. Summary of Vα family usages of all individuals analyzed are given in a scatter plot revealing a preferential usage of Vα4 and Vα38. Standard deviation is indicated by whiskers. (**D**) The ability of IFNγ production in response to NS3-1073 correlates with Vβ family usage. Samples showing high or low IFNγ responses towards the NS3-1073 peptide were grouped. Comparison of Vβ family usages between both groups displays a preferential usage of Vβ4, Vβ6 and Vβ24 in samples with a high response. Whisker plots show the minimum and maximum of percentages, the median is indicated by a horizontal line.

Figure 5

HCV NS3-1073 specific CD8+ T-cells recognize different cross-reactive peptides *in vitro*. (**A**) *In silico* structure-based prediction revealed peptide candidates potentially cross-reactive to NS3-1073. Hierarchical Cluster Analysis (HCA) was performed based on charge distribution over the TCR-interacting surface of peptide:MHC complexes. The green line indicates the expected threshold for cross-reactivity. Epitope information is presented on the left, amino acid exchanges in relation NS3-1073 indicated in red. Only the complexes in the upper branch of the dendrogram were selected for testing (blue line). Three selected peptide:MHC surfaces are depicted below. Alpha-1 and alpha-2 domains of the MHC are

indicated, delimiting the region occupied by the peptide. Charged areas over the peptide:MHCs surface were computed using GRASP2 and represented as red (negative) and blue (positive charges) spots ranging from −5 to +5 kT. Differences in topography and electrostatic distribution on the NA-231:HLA-A2 complex are indicated by arrows. (**B**) *In vitro* stimulation of NS3-1073 specific CD8+ T-cell lines from HCV-SN with different cross-reactive peptide candidates show a recognition of peptides derived from EBV LMP2-329 and IAV NA-231. FACS plots of IFNγ responses from three representative individuals are depicted, plots show cells gated on total T-cells after exclusion of the dump channel, the numbers indicated refer to frequencies of IFNγ+ CD8+ T-cells gated on CD8+ T-cells. Peptides used for *in vitro* cell culture and for re-stimulation during the cytokine assay are indicated. (**C**) Individual IFNγ responses of NS3-1073 specific CD8+ T-cell lines show cross-recognition of different peptide candidates by NS3-0173 specific T-cells. Bar graphs of IFNγ responses from four selected HCV-SN are shown representative. White bars represent frequencies of IFNγ+ CD8+ T-cells of medium control and black bars of NS3-1073 specific T-cell lines towards peptides indicated and gated on total CD8+ T-cells. Dashed horizontal red lines indicate background IFNγ production after re-stimulation of NS3-1073 specific T-cell lines with Tyrosinase negative control peptide. n. a. = not analyzed

Figure 6

T-cell responses to HCV peptide vaccination is influenced by pre-existing NS3-1073 specific CD8+ T-cells (**A**) *In vitro* expansion of NS3-1073 specific CD8+ T-cells in all six healthy vaccinated individuals analysed before (visit 00), during (visit 08 and 12) and after HCV peptide vaccination (visit 20) differs between individuals with some individuals responding to NS3-1073 before vaccination. The individuals were grouped into ex vivo responder (R) and non-responder (NR). A response to NS3-1073 before vaccination corresponds to the strength of the *in vitro* expansion of NS3-1073 specific CD8+ T-cells during early time point after vaccination (visit 08, indicated by dotted box). FACS plots of NS3-1073 multimer staining after 3 weeks *in vitro* stimulation from four individuals are shown, plots show cells gated on total T-cells after exclusion of dump channel. Numbers indicated refer to frequencies of NS3-1073 specific CD8+ T-cells gated on total CD8+ T-cells. n.a.= not analyzed (**B**) *In vivo* responsiveness to HCV peptide vaccination was stronger and occurred earlier in those individuals being *in vitro* NS3-1073 responder before vaccination. Frequencies of NS3-1073 specific CD8+ T-cells directly *ex vivo*, during (visit 08-12) and after vaccination (visit 16 – 20) are shown gated on total CD8+ T-cell populations. Horizontal dashed lines indicate the set thresholds as doubling of the baseline frequencies of specific cells before vaccination.

Supplementary Table Legends

Supplementary Figure S1: Memory phenotype of HCV NS3.1073 specific CD8+ T-cells after enrichment
FACS plots of staining after enrichment of NS3-1073 specific CD8+ T-cells of all 16 HCV-SN individuals where enrichment was successful are displayed (upper panels). Numbers indicate the frequencies of multimer+ CD8+ T-cells gated on total CD8+ T-cells. The respective lower panels show FACS plots for the corresponding memory phenotyping using CCR7 and CD45RA (lower panels). Cells were sub-gated on NS3-1073+ CD8+ T cells.

Supplementary Table S1: Cohort characteristics of HCV-SN individuals included
Characteristics of all 164 HCV-SN included in this study. Given are the sample codes, age and gender of individuals with their risk group affiliation for HCV exposure, the frequencies of NS-173+ CD8+ T cells ex vivo (background (% of multimer+ CD8- cells) subtracted, the frequency of HCV NS3-1073 multimer+ CD8+ T-cells after three weeks *in vitro* stimulation (values of medium line controls are subtracted) and the method used for establishing NS3-1073 specific T-cell lines. Cell lines regarded as *ex vivo* positive and as *in vitro* responder to NS3-1073 are highlighted in grey. UK = unknown, m = male, f = female, RF = risk free, HBD = healthy blood donor, HCW = health care worker, SP = HCV sexual partner, HDU = healthy drug user, acEBV = acute EBV

Supplementary Table S2: Summary and sequences of T cell clones identified for all HCV-SN.
Amino acid and nucleotide sequences of individual TCR β chain and TCR α chain clones are given in addition to the Vβ/Vα and Jβ/Jα usages and the respective frequencies and length of the CDR3 region. The number of sequences and number of individual clones found for each sample is indicated for each individual. The clonotype ratio is calculated by dividing the total number of sequences obtained by the number of individual clones identified. A clonotype ratio of >3 indicates a skewing of the clonal composition.

Supplementary Table S3: Comparison of CDR1 and CDR2 amino acid sequences
Amino acid sequences of the CDR1 and CDR2 regions are compared for different TCR Vβ families. The families Vβ4, Vβ6 and Vβ24 are associated with a strong IFNγ production in response to the NS3-1073 peptide and show a common amino acid motif of the CDR1 and CDR2 regions (indicated by green highlights). The TCR Vβ families Vβ9, Vβ29 and Vβ12, however, which are associated with a weaker IFNγ response, display different amino acid usages in these regions (indicated by red highlights).

Supplementary Table S4: Characteristics of peptide candidates for *in silico* structure-based approach
Different HLA-A*0201 restricted peptides were recovered from CrossTope Data Bank and used for *in silico* structure-based approaches to identify peptides potentially cross-reactive to the HCV NS3-1073 peptide. The table indicates the CrossTope ID and the epitope ID identifying the epitope in the Immune Epitope Database (IEDB). The structure type used for *in silico* modelling approaches, virus and proteins from which the peptides originate as well as the position, length and amino acid sequence of the epitopes are given.

Tables

Table 1: CDR3 regions show common amino acid motifs in HCV-SN

	Vb family	CDR3 motif	Jb family
ExAG-motif	24	**EAG**STTEA	1-1
	4	**E**V**AG**GNEQ	2-1
	4	**E**V**A**S**G**TPGEL	2-2
xxGAP-motif	27	ATW**GAP**YEQ	2-7
	4	EP**GAP**EKL	1-4
	28	ITR**GAP**YEQ	2-7
	7	LRPE**GAP**NEKL	1-4
	27	MGQ**GAP**YEQ	2-7
	28	MTS**GAP**YEQ	2-7
	28	SSA**GAP**YEQ	2-7
	28	TS**GAP**YEQ	2-7
PxTG-motif	4	**P**AW**TGG**NQPQ	1-5
	4	**P**E**TG**A**G**GTEA	1-1
	4	**P**G**TG**A**G**GTEA	1-1
	5	PL**PTG**S**G**NTEA	1-1
	5	**PTG**VPANYGY	1-2
	24	**PT**H**G**T**G**IYD	1-2
	6	**PT**P**G**QLNEKL	1-4
multiple Gs	27	A**GG**SYNEQ	2-1
	20	A**G**T**G**T**GG**YEQ	2-7
	24	DGD**G**A**G**LPQ	1-5
	20	D**G**DS**GG**SYEQ	2-7
	12	E**GG**PHEQ	2-1
	12	FA**G**Q**GG**TYEQ	2-7
	5	F**GGGG**TEA	1-1
	12	F**GG**Y	1-2
	6	**G**TP**G**Q**GG**EKL	1-4
	9	**G**Y**GG**M**G**TDTQ	2-3
	9	LL**G**S**GG**NYGY	1-2
	4	PAWT**GG**NQPQ	1-5
	10	S**GG**ITEA	1-1
	24	SL**GG**DYYNEQ	2-1
	12	VA**GGG**MQPQ	1-5
	20	V**GG**P**GG**EL	2-2
	7	VR**GGG**ETEA	1-1

Sequencing and analysis of TCRs of NS3-1073 specific CD8+ T-cells in HCV-SN show usage of distinct amino acid sequences within the CDR3β region among different individuals. Clones using ExAG-, xxGAP-, PxTGG-motifs or containing multiple glycins can be repeatedly found among HCV-SN. Usages of Vβ and Jβ families as well as the respective samples analyzed are additionally indicated.

Figure 1

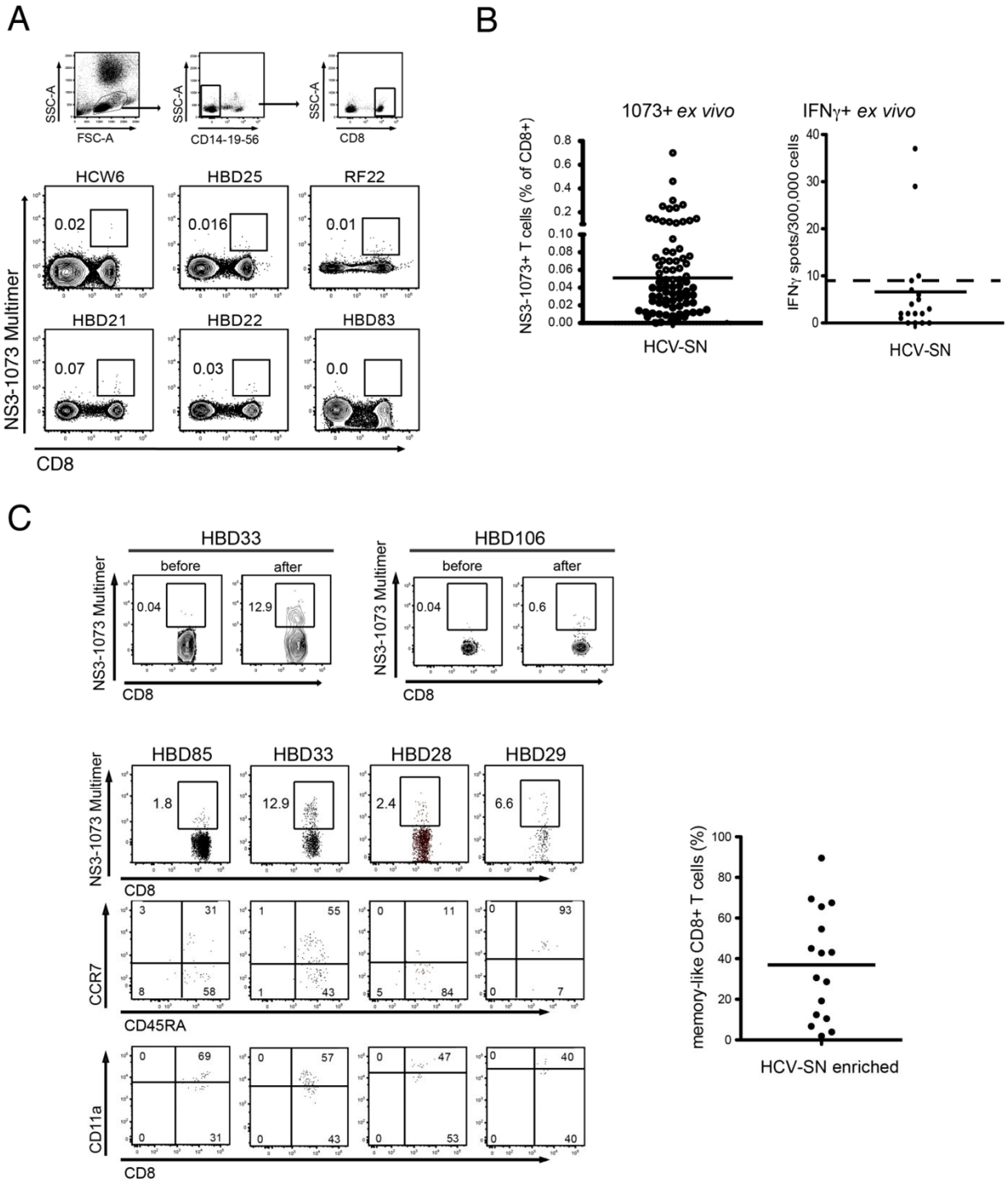

A
SSC-A
FSC-A
CD14-19-56
CD8
HCW6
HBD25
RF22
0.02
0.016
0.01
HBD21
HBD22
HBD83
0.07
0.03
0.0
NS3-1073 Multimer
CD8
B
1073+ ex vivo
IFNγ+ ex vivo
NS3-1073+ T cells (% of CD8+)
IFNγ spots/300,000 cells
HCV-SN
HCV-SN
C
HBD33
HBD106
before
after
0.04
12.9
0.04
0.6
NS3-1073 Multimer
CD8
HBD85
HBD33
HBD28
HBD29
1.8
12.9
2.4
6.6
CCR7
CD45RA
CD11a
memory-like CD8+ T cells (%)
HCV-SN enriched

Figure 2

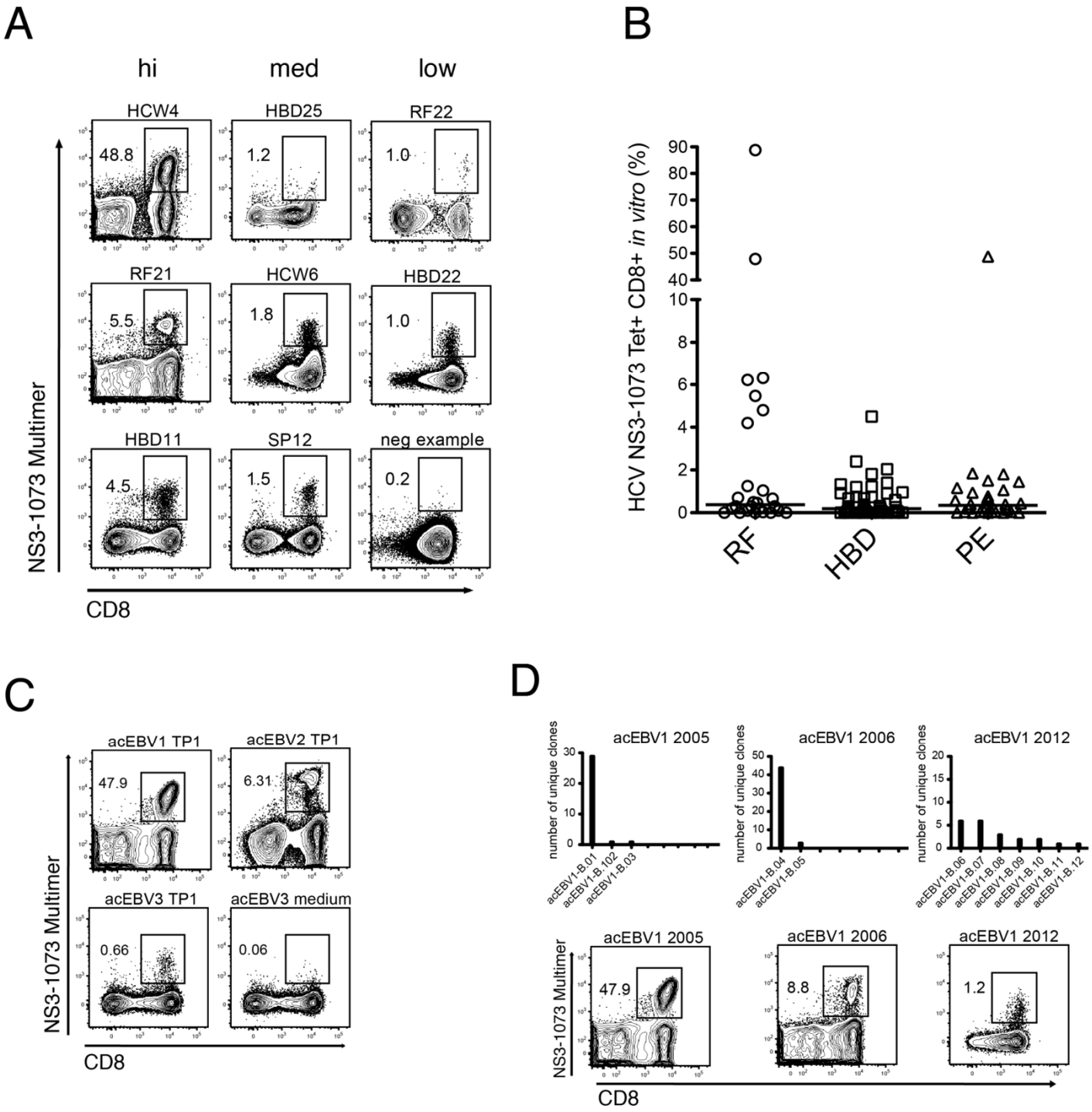

A
hi
med
low
HCW4
HBD25
RF22
48.8
1.2
1.0
RF21
HCW6
HBD22
5.5
1.8
1.0
HBD11
SP12
neg example
4.5
1.5
0.2
NS3-1073 Multimer
CD8
B
HCV NS3-1073 Tet+ CD8+ in vitro (%)
RF
HBD
PE
C
acEBV1 TP1
acEBV2 TP1
47.9
6.31
acEBV3 TP1
acEBV3 medium
0.66
0.06
NS3-1073 Multimer
CD8
D
acEBV1 2005
acEBV1 2006
acEBV1 2012
number of unique clones
acEBV1-B.01
acEBV1-B.102
acEBV1-B.03
acEBV1-B.04
acEBV1-B.05
acEBV1-B.06
acEBV1-B.07
acEBV1-B.08
acEBV1-B.09
acEBV1-B.10
acEBV1-B.11
acEBV1-B.12
47.9
8.8
1.2
NS3-1073 Multimer
CD8

Figure 3

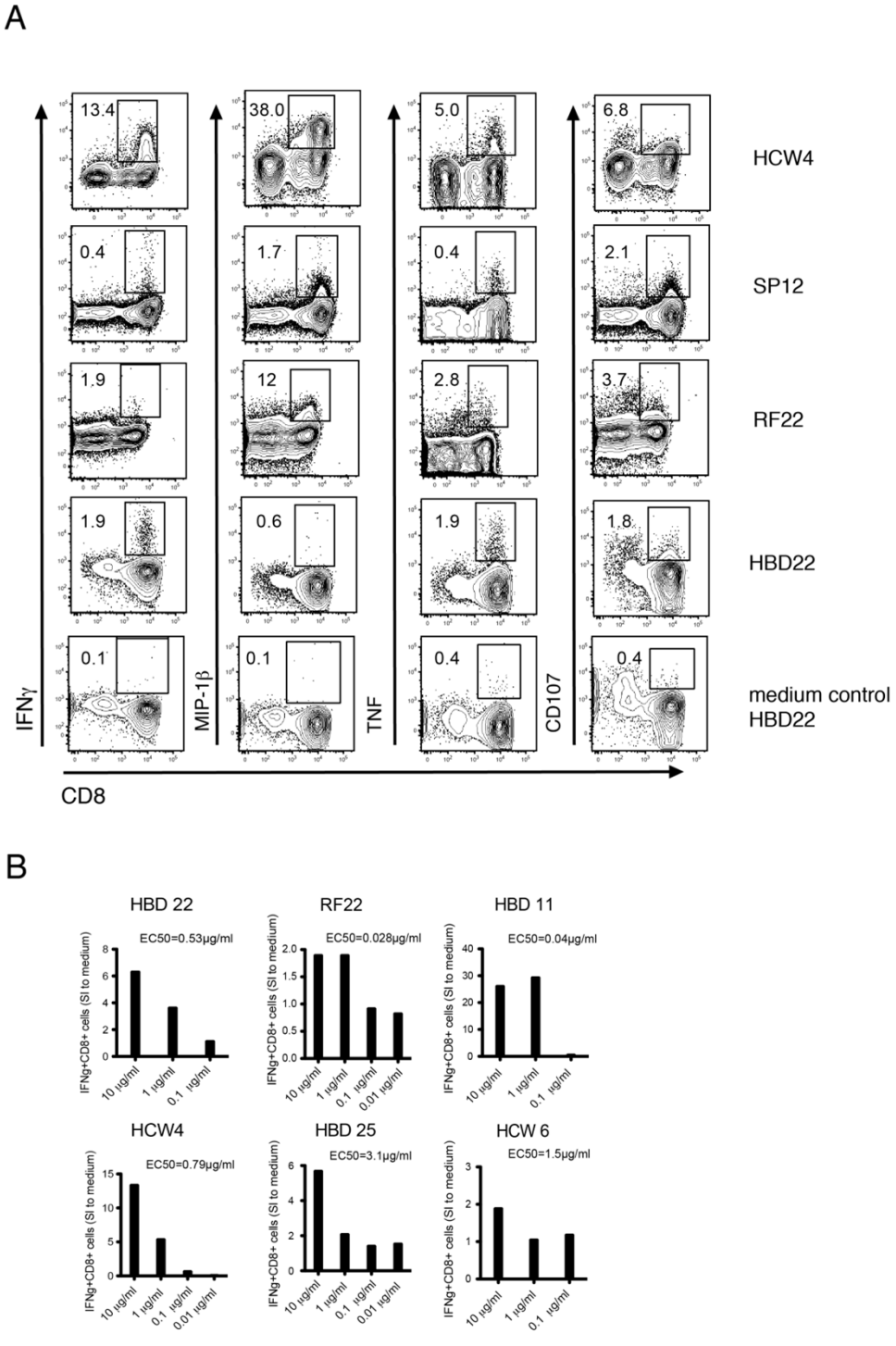

A
13.4
38.0
5.0
6.8
HCW4
0.4
1.7
0.4
2.1
SP12
1.9
12
2.8
3.7
RF22
1.9
0.6
1.9
1.8
HBD22
0.1
0.1
0.4
0.4
medium control
HBD22
IFNγ
MIP-1β
TNF
CD107
CD8
B
HBD 22
EC50=0.53µg/ml
RF22
EC50=0.028µg/ml
HBD 11
EC50=0.04µg/ml
HCW4
EC50=0.79µg/ml
HBD 25
EC50=3.1µg/ml
HCW 6
EC50=1.5µg/ml
IFNg+CD8+ cells (SI to medium)
10 µg/ml
1 µg/ml
0.1 µg/ml
0.01 µg/ml

Figure 4

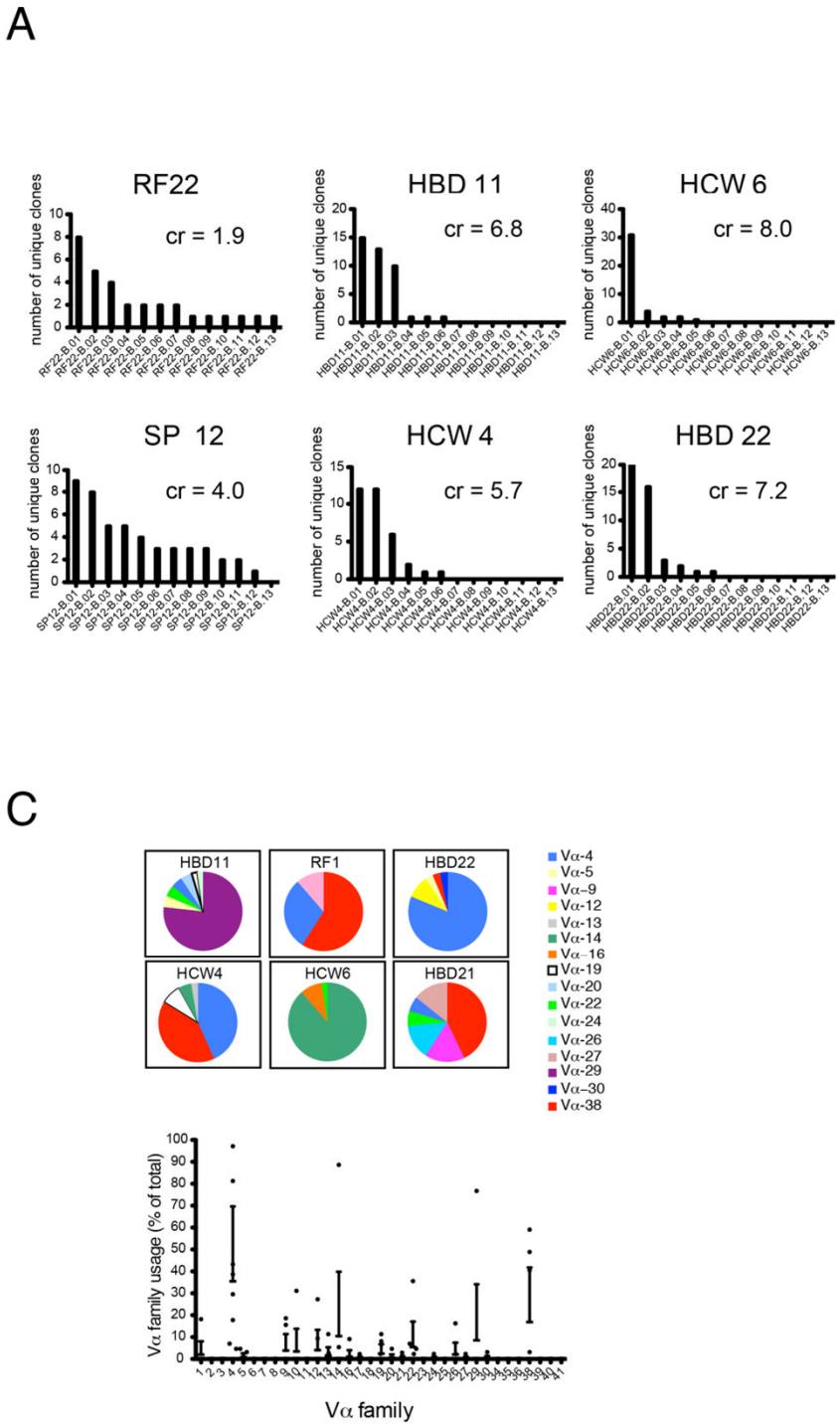
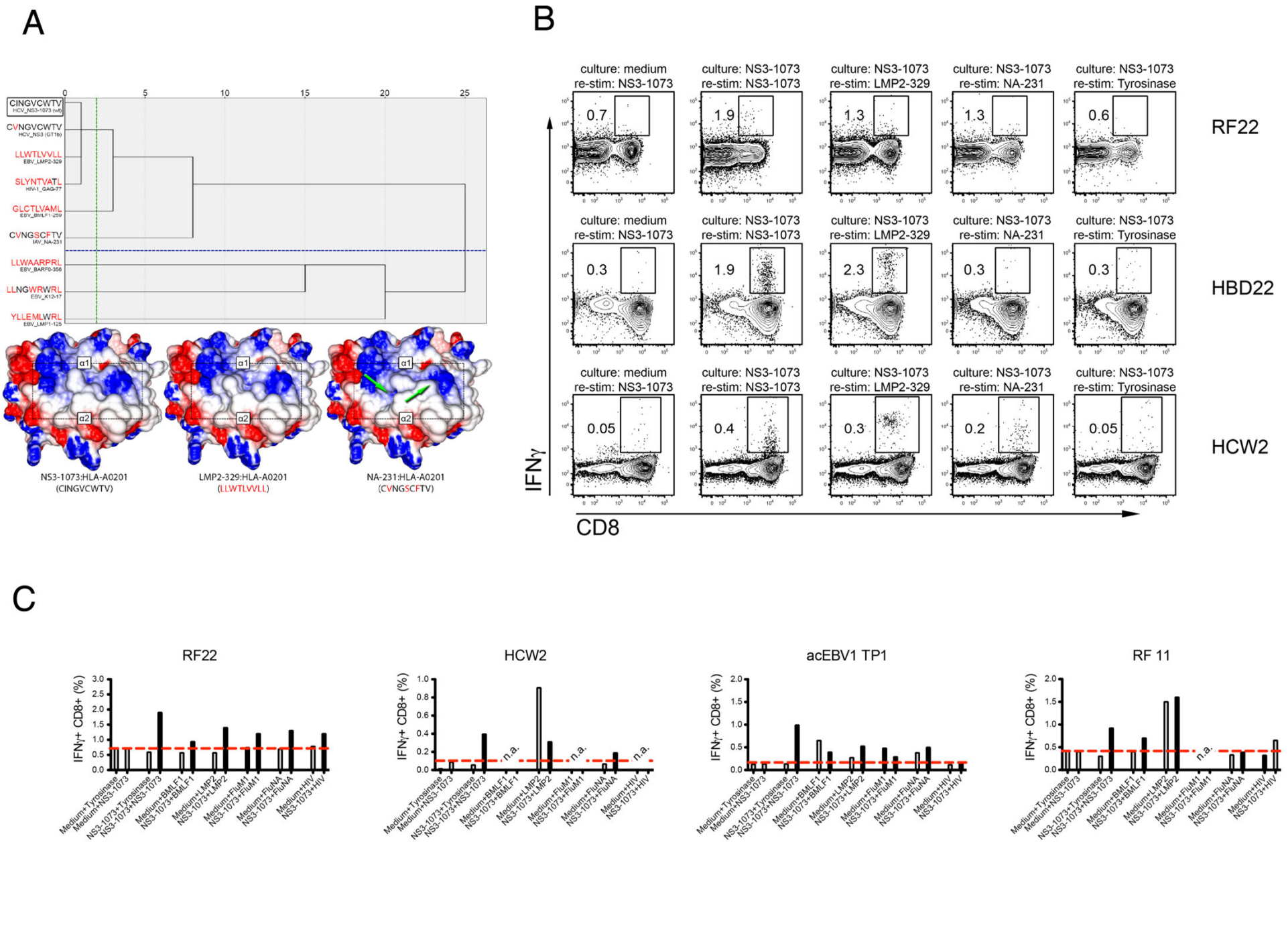
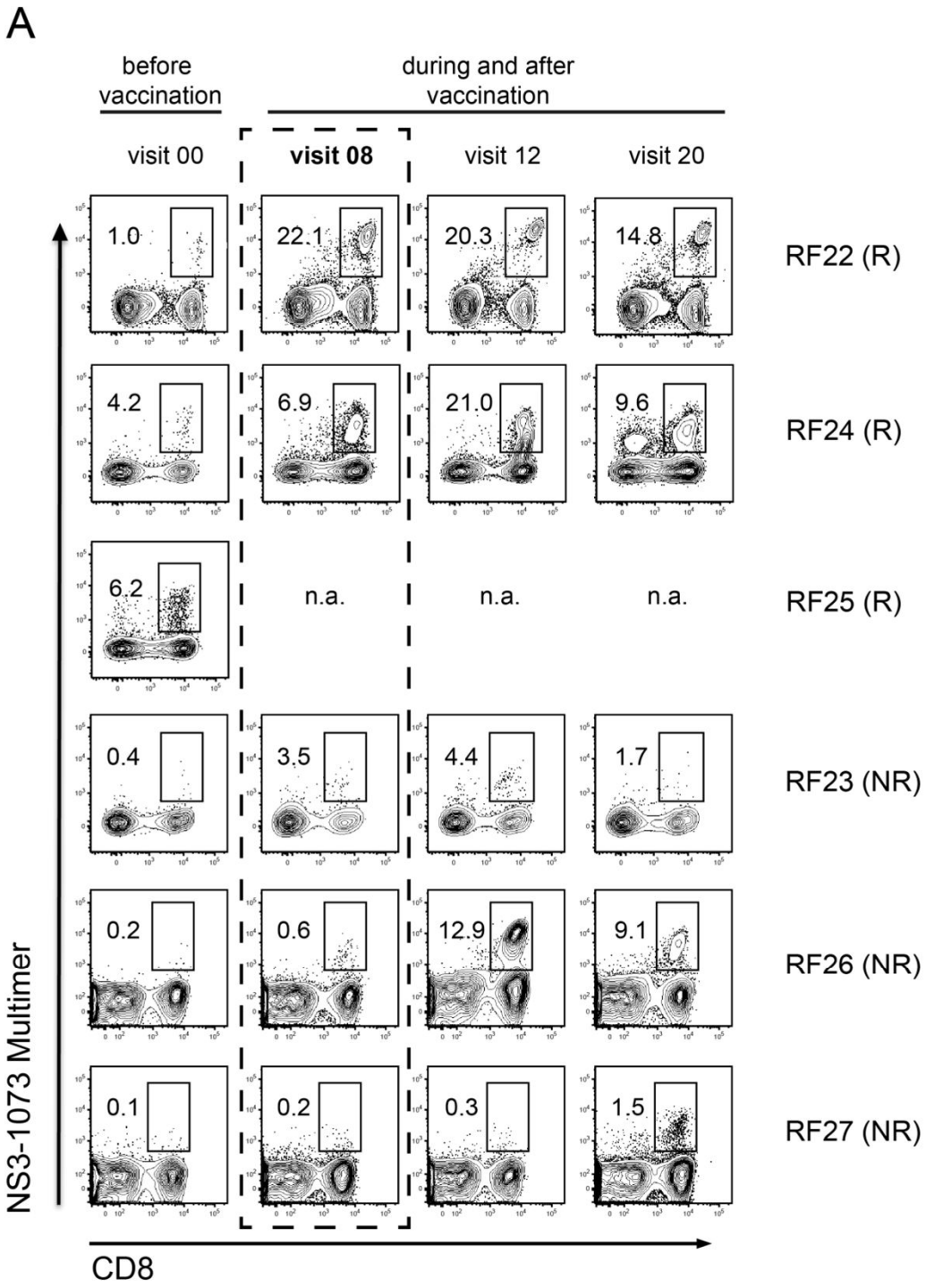

A
RF22
cr = 1.9
HBD 11
cr = 6.8
HCW 6
cr = 8.0
SP 12
cr = 4.0
HCW 4
cr = 5.7
HBD 22
cr = 7.2
number of unique clones
B
HBD11
HCW6
HCW4
HBD22
SP12
RF1
RF20
HBD21
RF22
Vβ-2
Vβ–4
Vβ-5
Vβ-6
Vβ-9
Vβ-10
Vβ-11
Vβ-12
Vβ-13
Vβ–14
Vβ-19
Vβ-20
Vβ-24
Vβ-27
Vβ-28
Vβ-29
Vβ family usage (% of total)
Vβ family
C
HBD11
RF1
HBD22
HCW4
HCW6
HBD21
Vα-4
Vα-5
Vα–9
Vα-12
Vα-13
Vα-14
Vα–16
Vα-19
Vα-20
Vα-22
Vα-24
Vα-26
Vα-27
Vα-29
Vα–30
Vα-38
Vα family usage (% of total)
Vα family
D
high IFNγ
low IFNγ
Vβ family usage (% of total)
Vβ4+Vβ6+Vβ24
Vβ9+Vβ29
other Vβ

Figure 5


A
B
C
CINGVCWTV
CVNGVCWTV
LLWTLVVLL
SLYNTVATL
GLCTLVAML
CVNGSCFTV
LLWAARPRL
LLNGWRWRL
YLLEMLWRL
NS3-1073:HLA-A0201
(CINGVCWTV)
LMP2-329:HLA-A0201
(LLWTLVVLL)
NA-231:HLA-A0201
(CVNGSCFTV)
culture: medium
re-stim: NS3-1073
culture: NS3-1073
re-stim: NS3-1073
culture: NS3-1073
re-stim: LMP2-329
culture: NS3-1073
re-stim: NA-231
culture: NS3-1073
re-stim: Tyrosinase
RF22
HBD22
HCW2
IFNγ
CD8
RF22
HCW2
acEBV1 TP1
RF 11
IFNγ+ CD8+ (%)
n.a.


A
before vaccination
during and after vaccination
visit 00
visit 08
visit 12
visit 20
1.0
22.1
20.3
14.8
RF22 (R)
4.2
6.9
21.0
9.6
RF24 (R)
6.2
n.a.
n.a.
n.a.
RF25 (R)
0.4
3.5
4.4
1.7
RF23 (NR)
0.2
0.6
12.9
9.1
RF26 (NR)
0.1
0.2
0.3
1.5
RF27 (NR)
NS3-1073 Multimer
CD8

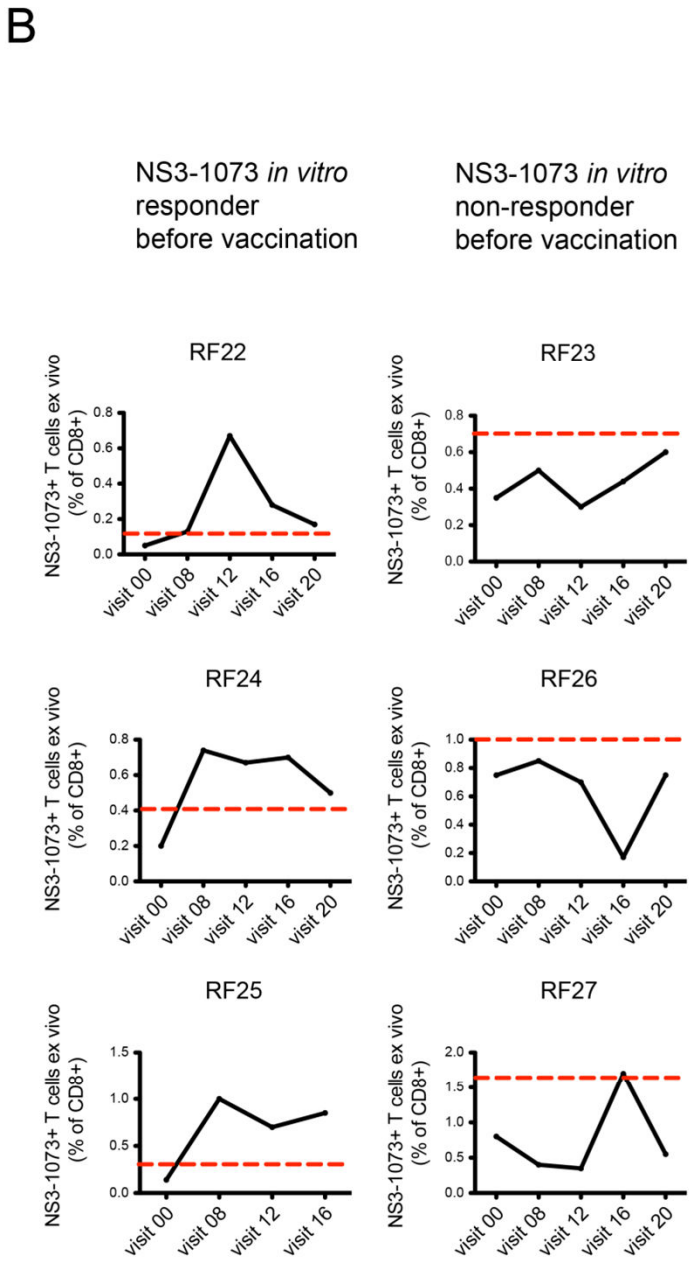
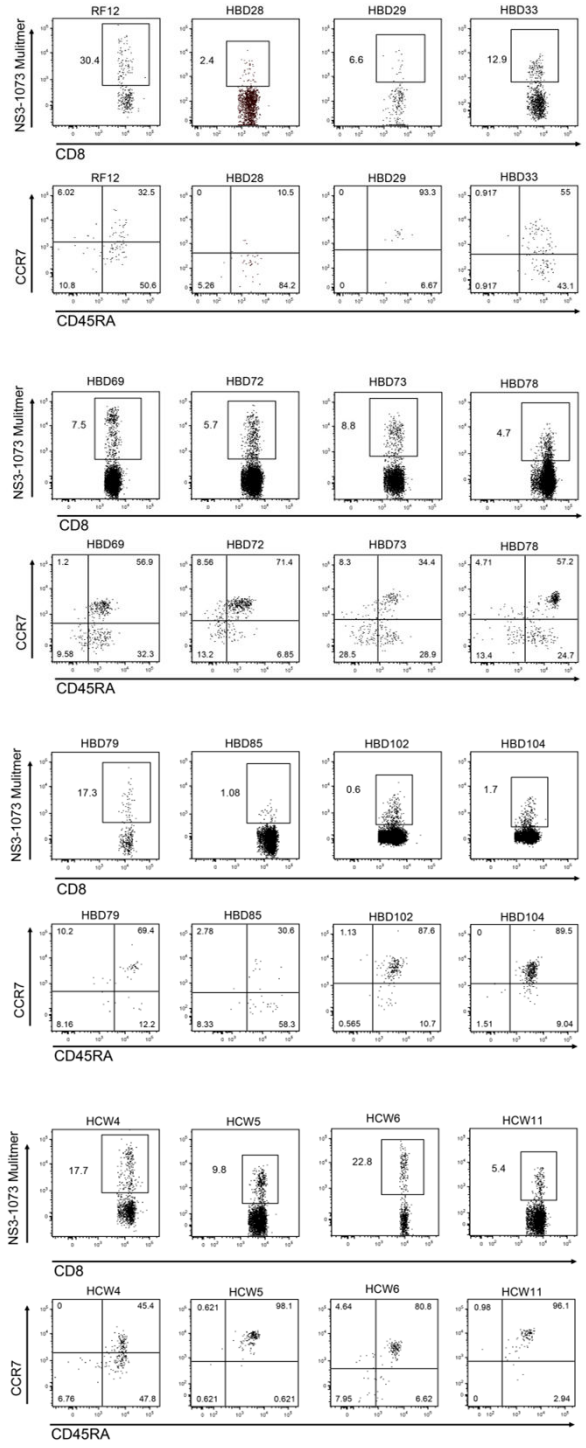

B
NS3-1073 in vitro responder before vaccination
NS3-1073 in vitro non-responder before vaccination
RF22
RF23
RF24
RF26
RF25
RF27
NS3-1073+ T cells ex vivo (% of CD8+)
visit 00
visit 08
visit 12
visit 16
visit 20

Figure 6



Supplementary Figure S1

Supplementary Table 1: Summary of all HCV seronegative individuals enrolled for the study.

code	sex	age	risk group	% NS3-1073+CD8+ *ex vivo* (background substracted)	% NS3-1073+CD8+ *in vitro*	cell culture method
RF01	f	30	risk-free	*n.a.*	88.80	CD8
RF02	m	32	risk-free	0.046	1.25	PBMC
RF03	m	31	risk-free	*n.a.*	0.00	PBMC
RF04	f	32	risk-free	*n.a.*	0.03	PBMC
RF05	f	30	risk-free	0.083	0.20	PBMC
RF06	m	30	risk-free	*n.a.*	0.10	CD3+DC
RF07	m	35	risk-free	*n.a.*	0.00	CD3+DC
RF08	m	31	risk-free	*n.a.*	0.30	CD3+DC
RF09	m	34	risk-free	*n.a.*	0.00	CD8
RF10	f	25	risk-free	0.049	0.30	PBMC
RF11	f	32	risk-free	*n.a.*	0.70	PBMC
RF12	f	31	risk-free	0.040	0.00	PBMC
RF17	m	64	risk-free	*n.a.*	0.30	CD8
RF18	f	38	risk-free	*n.a.*	0.00	CD8
RF19	uk	18-50	risk-free vacc.	0.073	0.46	PBMC
RF20	uk	18-50	risk-free vacc.	0.125	4.80	PBMC
RF21	uk	18-50	risk-free vacc.	0.034	5.47	PBMC
RF22	uk	18-50	risk-free vacc.	0.044	1.04	PBMC
RF23	uk	18-50	risk-free vacc.	0.35	0.49	PBMC
RF24	uk	18-50	risk-free vacc.	0.20	4.20	PBMC
RF25	uk	18-50	risk-free vacc.	0.14	6.22	PBMC
RF26	uk	18-50	risk-free vacc.	0.75	0.20	PBMC
RF27	uk	18-50	risk-free vacc.	0.8	0.10	PBMC
HBD01	uk	uk	healthy blood donor	0.000	0.10	PBMC
HBD02	uk	uk	healthy blood donor	0.000	0.30	PBMC
HBD03	uk	uk	healthy blood donor	0.000	0.02	PBMC
HBD04	uk	uk	healthy blood donor	0.005	0.00	PBMC
HBD05	uk	uk	healthy blood donor	0.000	0.00	PBMC
HBD06	uk	uk	healthy blood donor	0.237	0.02	CD8
HBD07	uk	uk	healthy blood donor	0.110	0.03	CD8
HBD08	uk	uk	healthy blood donor	0.300	0.03	CD8
HBD09	uk	uk	healthy blood donor	0.121	0.00	CD8
HBD10	uk	uk	healthy blood donor	*n.a.*	0.00	PBMC
HBD11	uk	uk	healthy blood donor	0.049	4.50	PBMC
HBD12	uk	uk	healthy blood donor	*n.a.*	1.40	PBMC
HBD13	uk	uk	healthy blood donor	0.071	0.32	PBMC
HBD14	uk	uk	healthy blood donor	0.147	0.26	PBMC
HBD15	uk	uk	healthy blood donor	0.039	0.00	PBMC
HBD16	uk	uk	healthy blood donor	0.141	0.01	CD8
HBD17	uk	uk	healthy blood donor	0.251	*n.a.*	*n.a.*
HBD18	uk	uk	healthy blood donor	0.000	0.67	PBMC
HBD19	uk	uk	healthy blood donor	0.018	0.28	CD8
HBD20	uk	uk	healthy blood donor	0.008	0.38	CD8
HBD21	uk	uk	healthy blood donor	0.076	0.19	CD8
HBD22	uk	uk	healthy blood donor	0.032	0.95	CD8

HBD23	uk	uk	healthy blood donor	0.030	0.60	CD8
HBD24	uk	uk	healthy blood donor	0.023	2.40	CD8
HBD25	uk	uk	healthy blood donor	0.047	1.20	CD8
HBD26	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD27	uk	uk	healthy blood donor	0.062	*n.a.*	*n.a.*
HBD28	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD29	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD30	uk	uk	healthy blood donor	0.011	*n.a.*	*n.a.*
HBD31	uk	uk	healthy blood donor	0.025	*n.a.*	*n.a.*
HBD32	uk	uk	healthy blood donor	0.028	*n.a.*	*n.a.*
HBD33	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD34	uk	uk	healthy blood donor	0.119	*n.a.*	*n.a.*
HBD35	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD36	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD37	uk	uk	healthy blood donor	*n.a.*	0.21	CD8
HBD38	uk	uk	healthy blood donor	0.042	0.00	CD8
HBD39	uk	uk	healthy blood donor	0.231	0.00	CD8
HBD40	uk	uk	healthy blood donor	0.000	0.00	CD8
HBD41	uk	uk	healthy blood donor	0.065	0.10	CD8
HBD42	uk	uk	healthy blood donor	0.000	0.08	CD8
HBD43	uk	uk	healthy blood donor	*n.a.*	0.74	CD8
HBD44	uk	uk	healthy blood donor	*n.a.*	0.10	CD8
HBD45	uk	uk	healthy blood donor	0.000	1.33	CD8
HBD46	uk	uk	healthy blood donor	0.084	*n.a.*	*n.a.*
HBD47	uk	uk	healthy blood donor	0.012	*n.a.*	*n.a.*
HBD48	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD49	uk	uk	healthy blood donor	0.036	*n.a.*	*n.a.*
HBD50	uk	uk	healthy blood donor	0.033	*n.a.*	*n.a.*
HBD51	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD52	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD53	uk	uk	healthy blood donor	0.049	*n.a.*	*n.a.*
HBD54	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD55	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD56	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD57	uk	uk	healthy blood donor	0.014	*n.a.*	*n.a.*
HBD58	uk	uk	healthy blood donor	0.461	*n.a.*	*n.a.*
HBD59	uk	uk	healthy blood donor	0.150	*n.a.*	*n.a.*
HBD60	uk	uk	healthy blood donor	0.263	*n.a.*	*n.a.*
HBD61	uk	uk	healthy blood donor	0.026	*n.a.*	*n.a.*
HBD62	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD63	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD64	uk	uk	healthy blood donor	0.021	*n.a.*	*n.a.*
HBD65	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD66	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD67	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD68	uk	uk	healthy blood donor	0.007	*n.a.*	*n.a.*
HBD69	uk	uk	healthy blood donor	0.035	*n.a.*	*n.a.*
HBD70	uk	uk	healthy blood donor	0.041	*n.a.*	*n.a.*
HBD71	uk	uk	healthy blood donor	0.081	*n.a.*	*n.a.*
HBD72	uk	uk	healthy blood donor	*n.a.*	*n.a.*	*n.a.*
HBD73	uk	uk	healthy blood donor	0.030	*n.a.*	*n.a.*
HBD74	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD75	uk	uk	healthy blood donor	0.070	2.04	CD8

HBD76	uk	uk	healthy blood donor	0.060	*n.a.*	*n.a.*
HBD77	uk	uk	healthy blood donor	0.042	1.23	CD8
HBD87	uk	uk	healthy blood donor	0.002	0.00	CD8
HBD79	uk	uk	healthy blood donor	0.000	0.00	CD8
HBD80	uk	uk	healthy blood donor	0.000	0.00	CD8
HBD81	uk	uk	healthy blood donor	0.053	*n.a.*	*n.a.*
HBD82	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD83	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD84	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD85	uk	uk	healthy blood donor	0.000	0.11	CD8
HBD87	uk	uk	healthy blood donor	0.000	0.07	CD8
HBD88	uk	uk	healthy blood donor	0.000	0.11	CD8
HBD89	uk	uk	healthy blood donor	0.002	*n.a.*	*n.a.*
HBD90	uk	uk	healthy blood donor	0.000	0.27	CD8
HBD91	uk	uk	healthy blood donor	0.000	0.04	CD8
HBD92	uk	uk	healthy blood donor	0.028	*n.a.*	*n.a.*
HBD93	uk	uk	healthy blood donor	0.002	0.22	CD8
HBD94	uk	uk	healthy blood donor	0.021	1.81	CD8
HBD95	uk	uk	healthy blood donor	0.021	*n.a.*	*n.a.*
HBD96	uk	uk	healthy blood donor	0.001	0.68	CD8
HBD97	uk	uk	healthy blood donor	0.012	*n.a.*	*n.a.*
HBD98	uk	uk	healthy blood donor	0.012	*n.a.*	*n.a.*
HBD99	uk	uk	healthy blood donor	0.010	*n.a.*	*n.a.*
HBD100	uk	uk	healthy blood donor	0.074	*n.a.*	*n.a.*
HBD101	uk	uk	healthy blood donor	0.000	*n.a.*	*n.a.*
HBD102	uk	uk	healthy blood donor	0.056	0.73	CD8
HBD103	uk	uk	healthy blood donor	0.040	0.90	CD8
HBD104	uk	uk	healthy blood donor	0.060	1.75	CD8
HBD105	uk	uk	healthy blood donor	0.001	*n.a.*	*n.a.*
HBD106	uk	uk	healthy blood donor	0.023	*n.a.*	*n.a.*
HBD107	uk	uk	healthy blood donor	0.010	*n.a.*	*n.a.*
HCW01	m	41	health care worker	*n.a.*	0.57	CD8
HCW02	m	33	health care worker	*n.a.*	1.16	PBMC
HCW03	f	39	health care worker	0.026	0.04	PBMC
HCW04	m	32	health care worker	0.019	48.78	PBMC
HCW05	f	30	health care worker	0.040	1.06	CD8
HCW06	m	37	health care worker	0.019	1.84	CD8
HCW07	f	30	health care worker	*n.a.*	0.14	CD8
HCW08	f	35	health care worker	*n.a.*	0.68	PBMC
HCW09	m	31	health care worker	0.009	*n.a.*	*n.a.*
HCW10	f	42	health care worker	0.010	*n.a.*	*n.a.*
HCW11	m	44	health care worker	0.045	1.80	CD8
SP01	m	36	HCV sexual partner	*n.a.*	0.07	PBMC
SP02	f	65	HCV sexual partner	0.000	0.94	PBMC
SP03	f	55	HCV sexual partner	0.044	0.44	PBMC
SP04	m	uk	HCV sexual partner	0.000	0.77	PBMC
SP05	f	50	HCV sexual partner	*n.a.*	0.20	PBMC
SP06	m	71	HCV sexual partner	*n.a.*	0.00	PBMC
SP07	m	46	HCV sexual partner	0.000	0.05	PBMC
SP08	f	uk	HCV sexual partner	*n.a.*	0.01	PBMC
SP09	f	uk	HCV sexual partner	*n.a.*	1.45	PBMC

SP10	m	40	HCV sexual partner	*n.a.*	0.39	PBMC
SP11	f	29	HCV sexual partner	*n.a.*	0.03	PBMC
SP12	m	55	HCV sexual partner	*n.a.*	1.51	PBMC
SP13	m	uk	HCV sexual partner	*n.a.*	0.04	PBMC
SP14	f	53	HCV sexual partner	*n.a.*	0.20	PBMC
SP15	f	33	HCV sexual partner	*n.a.*	0.60	PBMC
SP16	m	65	HCV sexual partner	*n.a.*	0.04	PBMC
SP17	f	40	HCV sexual partner	*n.a.*	0.00	PBMC
HDU1	m	uk	healthy drug user	*n.a.*	0.30	PBMC
HDU2	m	uk	healthy drug user	*n.a.*	0.00	PBMC
HDU3	m	uk	healthy drug user	*n.a.*	0.40	PBMC
HDU4	m	uk	healthy drug user	*n.a.*	0.00	PBMC
acEBV1	m	25	acute EBV	0.070	47.90	PBMC
acEBV2	m	26	acute EBV	0.007	6.31	PBMC
acEBV3	m	22	acute EBV	*n.a.*	0.66	CD8

Supplementary Table S2: Summary and sequences of T cell Vβ and Vα chains and clones identified for all HCV-SN analyzed. Clonotype ratio = number of sequences obtained / number of individual clones.

TRBV sequences						
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
RF1-B.01	12-4	1-2	25	CASS-FGGY-TFGSG	4	TGTGCCAGCAGTTT-TGGG-GGCTAC
RF1-B.02	24	2-1	6	CA-TGTSGTYNEQ-FFGPG	10	TGTGCCAC-GGGGACTAGCGGGACT-TACAATGAGCAG
RF1-B.03	6-1	2-7	4	CAS-NVLGESIYEQ-YFGPG	10	TGTGCCAGCA-ATGTCTTAGGGGAATCAAT-CTACGAGCAG
RF1-B.04	4-1	1-1	2	CASSQ-DYNTEA-FFGQG	6	TGTGCCAGCAGCCAAGA-TTAC-AACACTGAAGCT
RF1-B.05	28	1-2	2	CASSL-SHTGGLDGY-TFGSG	9	TGTGCCAGCAGTTTAT-CCCACACAGGGGGCTTGG-ATGGCTAC
RF1-B.06	29	2-1	2	CSV-VPPGRGDNEQ-FFGPG	10	TGTAGCGTTG-TCCCCCCCGGGAGAGGAG-ACAATGAGCAG
RF1-B.07	14	2-7	1	CASSQ-DFDEQ-YFGPG	5	TGTGCCAGCAGCCAAGA-CTTCG-ACGAGCAG
no. of seqs			**42**			
no. individual clones			**7**			
clonotype ratio			**6.00**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
RF20-B.10	29-1	1.2	14	CSA-PGVGY-TFGSGTRLTVVED	5	TGCAGCGCC-CCAGGGGTCGGCTACACCTTCGGT TCGGGGACCAGGTTAACCGTTGTA GAGGAC
RF20-B.09	2-1	2.4	9	CASS-GSGKNIQ-YFGAGTRLSVLED	7	TGTGCCAGTTCG-GGGTCCGGGAAAAACATTCAGTAC TTCGGCGCCGGGACCCGGCTCTCA GTGCTGGAG
RF20-B.08	4-2/4-3	1.1	6	CASS-PGTGAGGTEA-FFGQGTRLTVVE	10	TGCGCCAGCAGC-CCAGGGACAGGCGCCGGGGGAAC TGAAGCTTTCTTTGGACAAGGCACC AGACTCAC
RF20-B.07	27-1	2.7	2	CASS-ATWGAPYEQ-YFGPGTRLTVTED	9	TGTGCCAGCAGT-GCAACATGGGGGGCCCCCTACGAG CAGTACTTCGGGCCGGGCACCAGG CTCACGG
RF20-B.01	27-1	2.1	1	CAS-CVAGGFNEQ-FFGPGTRLTVLED	9	TGTGCCAGCTGC-GTAGCGGGGGGCTTCAATGAGCAG TTCTTCGGGCCAGGGACACGGCTC ACCGTG
RF20-B.02	6-5	2.3	1	CASS-LTAGTSGGPSTD TQ-YFGPGTRLTVLED	14	TGTGCCAGCAGT-CTTACAGCCGGGACTAGCGGGGGG CCTAGCACAGATACGCAGTATTTTG GC
RF20-B.03	4-2/4-3	1.1	1	CASS-PETGAGGTEA-FFGQGTRLTVVE	10	TGCGCCAGCAGC-CCAGAGACAGGCGCCGGGGGAAC TGAAGCTTTCTTTGGACAAGGCACC AGACTCA

RF20-B.04	4-2/4-3	2.2	1	CASSQ-EPAASTGEL-FFGEGSRLTVLED	9	TGCGCCAGCAGC-CAAGAACCGGCAGCTAGCACCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGAC
RF20-B.05	27-1	2.2	1	CASS-SYARTGGRL-FFGEGSRLTVLED	9	TGTGCCAGCAGT-TCCTACGCCCGGACAGGGGGTAGGCTGTTTTTTGGAGAAGGCTCTAGGCTGACCGTA
RF20-B.06	20-1	2.2	1	CSAS-VGGPGGEL-FFGEGSRLTVLED	8	TGCAGTGCT-AGCGTGGGGGGGCCGGGCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGACCGTACT
no. of seqs			**37**			
no. individual clones			**10**			
clonotype ratio			**3.7**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
RF21-B.09	11-2	1.4	18	CASS-PYATNEKL-FFGSGTQLSVLED	8	TGTGCCAGCAGC-CCTTATGCAACTAATGAAAAACTGTTTTTTGGCAGTGGAACCCAGCTCTCTGTCTTGGAGGAC
RF21-B.07	28-1	2.7	3	CASS-ITRGAPYEQ-YFGPGTRLTVTED	8	TGTGCCAGCAGC-ATAACTAGAGGGGCTCCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACA
RF21-B.08	28-1	2.3	3	CASS-SFQGALDTQ-YFGPGTRLTVLED	9	TGTGCCAGCAGT-TCATTCCAGGGGGCTTTAGATACGCAGTACTTTGGCCCAGGCACCCGGCTGACAGTGCTCG
RF21-B.05	4-2/4-3	1.1	2	CASSQ-DGTEA-FFGQGTRLTVVED	5	TGCGCCAGCAGC-CAAGATGGCACTGAAGCTTTCTTTGGACAAGGCACCAGACTCACAGTTGTAGAGGAC
RF21-B.06	28-1	2.6	2	CASS-VVVGTVYSGANVL-TFGAGSRLTVLED	12	TGTGCCAGCAGT-GTAGTTGTGGGGACAGTCTACTCTGGGGCCAACGTCCTGACTTTCGGGGCCGGCAGCAGGC
RF21-B.01	12-3	2.5	1	CAS-RQLGETQ-YFGPGTRLLVLED	7	TGTGCCAGCAGA-CAACTGGGGGAGACCCAGTACTTCGGGCCAGGCACGCGGCTCCTGGTGCTCGAGGAC
RF21-B.02	2-1/2-2	1.6	1	CASS-PGQVFSYNSPL-HFGNGTRLTVTE	11	TGTGCCAGCAGC-CCGGGACAGGTCTTCTCCTATAATTCACCCCTCCACTTTGGGAACGGGACCAGGCTCAC
RF21-B.03	4-1	1.1	1	CASS-PTLNTEA-FFGQGTRLTVVED	7	TGCGCCAGCAGC-CCGACACTAAACACTGAAGCTTTCTTTGGACAAGGCACCAGACTCACAGTTGTAGAGGAC
RF21-B.04	9-1	2.1	1	CASSV-AHNEQ-FFGPGTRLTVLED	5	TGTGCCAGCAGC-GTAGCCCACAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTAGAGGAC
no. of seqs			**32**			
no. individual clones			**10**			
clonotype ratio			**3.2**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**

RF22-B.04	28-1	2.7	1	CASR-SSAGAPYEQ-YFGPGTRLTVTED	9	TGTGCCAGCAGATCTTCGGCAGGGGCTCCTTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.05	3-1/3-2	2.7	1	CASS-HEVAAAYEQ-YFGPGTRLTVTE	9	TGTGCCAGCAGCCATGAGGTAGCGGCAGCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.09	4-1	2.7	1	CASSQ-EEGSEQ-YFGPGTRLTVTE	6	TGCGCCAGCAGCCAAGAAGAAGGGTCCGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.10	24-1	2.7	1	CATS-AATGADEQ-YFGPGTRLTVTED	8	TGTGCCACCAGTGCTGCGACAGGGGCAGACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.11	20-1	2.7	1	CSA-ETSGNGYEQ-YFGPGTRLTVTED	9	TGCAGTGCTGAGACTAGCGGGAACGGCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.12	29-1	2.7	1	CSV-IAGRGDAQ-YFGPGTRLTVLED	8	TGCAGCGTGATAGCGGGACGAGGGGATGCGCAGTATTTTGGCCCAGGCACCCGGCTGACAGTGCTCGAGGAC
RF22-B.14	27-1	2.7	2	CASS-PLGSSYEQ-YFGPGTRLTVTED	9	TGTGCCAGCAGTCCCCTCGGGAGCTCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGACCTG
RF22-B.17	28-1	2.7	4	CASS-MTSGAPYEQ-YFGPGTRLTVTED	9	TGTGCCAGCAGTATGACTAGCGGAGCTCCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.18	4-2/4-3	2.7	5	CASSQ-LTSAPYEQ-YFGPGTRLTVTE	9	TGTGCCAGCAGCCAACTGACTAGCGCGCCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTCACGGTCACAGAGGAC
RF22-B.19	28-1	2.7	8	CASSL-TSGAPYEQ-YFGPGAGLTVTE	8	TGTGCCAGCAGTTTAACTAGCGGGGCACCCTACGAGCAGTATTTCGGGCCGGGCGCCGGGCTCACGGTCACAGAGGAC
RF22-B.02	28-1	2.5	1	CAS-QGTDQETQ-YFGPGTRLLVLEDLK	8	TGTGCCAGCCAGGGGACAGACCAAGAGACCCAGTACTTCGGGCCAGGCACGCGGCTCCTGGTGCTCGAGGACCTGAAA
RF22-B.15	29-1	2.5	2	CSVE-VDRVGETQ-YFGPGTRLLVLED	8	TGCAGCGTTGAAGTGGACCGGGTAGGAGAGACCCAGTACTTCGGGCCAGGCACGCGGCTCCTGGTGCTCGAGGAC
RF22-B.08	14-1	2.3	1	CASSQ-DRDWITDTQ-YFGPGTRLTVLED	9	TGTGCCAGCAGCCAGGACAGGGATTGGATCACAGATACGCAGTATTTTGGCCCAGGCACCCGGCTGACAGTGCTCGAGGAC
RF22-B.16	30	2.2	2	CAW-RVQATGEL-FFGEGSRLTVLED	8	:TGTGCCTGGAGA:GTACAGGCCACCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGACCGTACTGGAGGAC
RF22-B.03	12-2	2.1	1	CASRL-EGGPHEQ-FFGPGTRLTVLED	7	TGTGCAAGTCGCTTAGAAGGAGGGCCGCATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTAGAGGAC
RF22-B.06	27-1	2.1	1	CASSL-AGGSYNEQ-FFGPGTRLTVLED	8	TGTGCCAGCAGTTTAGCGGGAGGCTCCTACAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTAGAGGAC
RF22-B.01	10-3	1.2	1	CAISE-SVEGVAAGY-TFGSGTRLTVVE	9	TGTGCCATCAGTGAGTCCGTGGAGGGGGTAGCGGCTGGCTACACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAG

RF22-B.13	11-2	1.2	2	CASSL-SGFYGY-TFGSGTRLTVVED	6	TGTGCCAGCAGCTTGTCGGGGTTCTATGGCTACACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAGGAC
RF22-B.07	28-1	1.1	1	CASSL-GGNTEA-FFGQGTRLTVVEDLN	6	TGTGCCAGCAGTTTAGGAGGGAACACTGAAGCTTTCTTTGGACAAGGCACCAGACTCACAGTTGTAGAGGACCTGAAC
no. of seqs			**37**			
no. individual clones			**19**			
clonotype ratio			**1.94736842**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
HBD11-B.01	6-5	2-1	15	CASSY-GREQ-FFGPG	4	TGTGCCAGCAGTTAC-GGAAGG-GAGCAG
HBD11-B.02	4	2-3	13	CASS-QVPGDTQ-YFGPG	6	TGTGCCAGCAGCCA-GGTTCCGGG-AGATACGCAG
HBD11-B.03	6-5	1-2	10	CASS-GPYGY-TFGSG	5	TGTGCCAGCAGT-GGTCCC-TATGGCTAC
HBD11-B.04	4	2-1	1	CASSQ-ELGENEQ-FFGPG	7	TGTGCCAGCAGCCAA-GAATTAGGGGAG-AATGAGCAG
HBD11-B.05	20	1-2	1	CSAR-WTVNYGY-TFGSG	7	TGTAGTGCTAGA-TGGACAGTG-AACTATGGCTAC
HBD11-B.06	28	2-7	1	CASS-QLYEQ-YFGPG	5	TGTGCCAGCAG-CCAGCTA-TACGAGCAG
no. of seqs			**41**			
no. individual clones			**6**			
clonotype ratio			**6.83**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
HBD12-B.01	4-3	2-3	5	CASSQ-DHPRGGTDTQ-YFGPG	10	TGTGCCAGCAGCCAAGA-TCACCCCCGGGGGG-GCACAGATACGCAG
HBD12-B.02	4-1	2-3	4	CASSQ-AQGQGVADTQ-YFGPG	10	TGTGCCAGCAGCCAAG-CCCAGGGACAGGGCGTGG-CAGATACGCAG
HBD12-B.03	20	2-7	3	CS-AGTGTGGYEQ-YFGPG	10	TGTAGTGC-AGGCACCGGGACAGGCGGT-TACGAGCAG
HBD12-B.04	9	1-5	3	CASS-ESGQVIEPQ-HFGDG	10	TGTGCCAGCAGCG-AGTCAGGACAGGTGATAG-AGCCCCAG
HBD12-B.05	12-4	1-5	2	CASSL-VAGGGMQPQ-HFGDG	9	TGTGCCAGCAGTTTAG-TAGCCGGAGGGGGGATG-CAGCCCCAG
HBD12-B.06	27	2-6	2	CAS-STTGTSGANVL-TFGAG	11	TGTGCCAGCAG-CACTACAGGGAC-CTCTGGGGCCAACGTCCTG
HBD12-B.07	6-6	1-6	2	CAS-GDLSSYNSPL-HFGNG	10	TGTGCCAGC-GGTGATCTCAG-CTCCTATAATTCACCCCTC
HBD12-B.08	9	2-2	2	CAS-NDRGLSTGEL-FFGEG	10	TGTGCCAGCA-ATGACAGGGGGCTTAG-CACCGGGGAGCTG
HBD12-B.09	4-3	2-7	1	CASSQ-VAQGWYEQ-YFGPG	8	TGTGCCAGCAGCCAAG-TGGCACAGGGCTGG-TACGAGCAG
HBD12-B.10	5-1	2-7	1	CASSL-EGQASSYEQ-YFGPG	9	TGTGCCAGCAGCTTGG-AGGGACAGGCGAG-CTCCTACGAGCAG

HBD12-B.11	7-2	1-1	1	CASS-RGSGQGPTEA-FFGQG	10	TGTGCCAGCAGC-CGAGGAAGCGGACAGGGGCC-CACTGAAGCT
HBD12-B.12	7-6	1-4	1	CASS-LRPEGAPNEKL-FFGSG	11	TGTGCCAGCAGC-CTCCGGCCCGAAGGGGCCC-CTAATGAAAAACTG
HBD12-B.13	24	1-1	1	CATS-EAGSTTEA-FFGQG	8	TGTGCCACCAGTGA-GGCCGGGTCCAC-CACTGAAGCT
HBD12-B.14	24	1-2	1	CATSD-PWTDVNYGY-TFGSG	9	TGTGCCACCAGTGAT-CCATGGACAGACGT-TAACTATGGCTAC
HBD12-B.15	24	1-5	1	CATS-DGDGAGLPQ-HFGDG	9	TGTGCCACCAGTGA-CGGCGACGGGGCGGGGTT-GCCCCAG
no. of seqs			**30**			
no. individual clones			**15**			
clonotype ratio			**2.00**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
HBD21-B.01	9	1-2	23	CASS-LLGSGGNYGY-TFGSG	10	TGTGCCAGCAGCCTCCTGGGATCGGGTGGAAACTATGGCTAC
HBD21-B.02	6-2/6-3	2-7	2	CAS-RTSNYEQ-YFGPG	7	TGTGCCAGCAGGACATCTAACTACGAGCAG
HBD21-B.03	4-2	1-1	1	CASS-HGGNTEA-FFGQG	7	TGTGCCAGCAGCCACGGGGGAAACACTGAAGCT
HBD21-B.04	2	2-7	1	CASSE-VGQGIYEQ-YFGPG	8	TGTGCCAGCAGTGAAGTGGGACAGGGGATCTACGAGCAG
no. of seqs			**27**			
no. individual clones			**4**			
clonotype ratio			**6.75**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
HBD22-B.01	4-1	1-2	20	CASSQ-DQQSYYGY-TFGSG	8	TGTGCCAGCAGCCAAGA-CCAACAGAGTTAC-TATGGCTAC
HBD22-B.02	12-3/12-4	2-7	16	CASS-FGTSGDEQ-YFGPG	8	TGTGCCAGCAGTTT-TGGGACTAGCGGGGAC-GAGCAG
HBD22-B.03	12-3/12-4	2-1	3	CASS-LGQYNEQ-FFGPG	7	TGTGCCAGCAGT-CTTGGACAG-TACAATGAGCAG
HBD22-B.04	5-1	1-2	2	CASS-PTGVPANYGY-TFGSG	10	TGTGCCAGCAGC-CCGACAGGGGTGCCAGCT-AACTATGGCTAC
HBD22-B.05	5-6	2-3	1	CAS-TSLPDTQ-YFGPG	7	TGTGCCAGC-ACGTCTCTCCCA-GATACGCAG
HBD22-B.06	19	2-7	1	CA-TRPSATYYEQ-YFGPG	10	TGTGCCA-CCCGACCAAGCGCAACCTAC-TACGAGCAG
no. of seqs			**43**			
no. individual clones			**6**			
clonotype ratio			**7.17**			

Clone ID	Vb	Jb	frequency	CDR3aa sequence	CDR3 aa length	Nucleotide sequence
HCW4-B.01	6-1	1-2	12	CASSE-LEGHYGY-TFGSG	7	TGTGCCAGCAGTGAA-TTGGAGGGCC-ACTATGGCTAC-ACC
HCW4-B.02	6-6	2-7	12	CASSY-TTGTDSYEQ-YFGPG	9	TGTGCCAGCAGTTAC-ACCACCGGGACTGAT-TCCTACGAGCAG-TAC
HCW4-B.03	9	2-1	6	CASSV-AGTYNEQ-FFGPG	7	TGTGCCAGCAGCGTAG-CGGGCA-CCTACAATGAGCAG-TTC
HCW4-B.04	6-1	1-6	2	CASSE-MAPILTNNSPL-HFGNG	11	TGTGCCAGCAGTGAA-ATGGCCCCAATTTTAACGA-ATAATTCACCCCTC-CAC
HCW4-B.05	4-2	2-1	1	CASSQ-EVAGGNEQ-FFGPG	8	TGTGCCAGCAGCCAAGA-GGTAGCGGGAGGT-AATGAGCAG-TTC
HCW4-B.06	4-2	2-7	1	CASSQ-ASGDTYEQ-YFGPG	8	TGTGCCAGCAGCCAAG-CTTCGGGAGACA-CCTACGAGCAG-TAC
no. of seqs			**34**			
no. individual clones			**6**			
clonotype ratio			**5.67**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
HCW6-B.01	6-1	1-2	31	CASSE-LSDSPYGY-TFGSG	8	TGTGCCAGCAGTGAA-CTTTCGGACAGCCC-CTATGGCTAC
HCW6-B.02	5-1	1-1	4	CASS-FGGGGTEA-FFGQG	8	TGTGCCAGCAGCTT-TGGAGGGGGAGG-CACTGAAGCT
HCW6-B.03	6-1	1-1	2	CASS-DLTGQGYGEA-FFGQG	10	TGTGCCAGCAGTGA-TCTTACGGGACAGGGTTACGG-TGAAGCT
HCW6-B.04	20-1	2-7	2	CSA-SVSSGVPEQ-YFGPG	9	TGTAGTGCTAG-CGTGTCTAGCGGGGTCCC-CGAGCAG
HCW6-B.05	10-3	2-7	1	CA-SWDSDEQ-YFGPG	7	TGTGCC-TCCTGGGATTCAGACG-AGCAG
no. of seqs			**40**			
no. individual clones			**5**			
clonotype ratio			**8.00**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
SP12-B.01	29	1-2	9	CSVE-EQQGNYGY-TFGSG	8	TGTAGCGTTGAAGA-GCAACAGGGG-AACTATGGCTAC
SP12-B.02	9	2-1	8	CASS-VPLGDYNEQ-YFGPG	9	TGTGCCAGCAGCGT-GCCCCTCGGGGA-CTACAATGAGCAG
SP12-B.03	13	2-7	5	CASSL-GSGPWEQ-YFGPG	7	TGTGCCAGCAGCTTAGG-ATCAGGCCCTTGG-GAGCAG
SP12-B.04	4-2	1-5	5	CASS-PAWTGGNQPQ-HFGDG	10	TGTGCCAGCAGCC-CAGCCTGGACAGGGG-GCAATCAGCCCCAG
SP12-B.05	4-3	2-7	4	CASSQ-VGIAAPYEQ-YFGPG	9	TGTGCCAGCAGCCAAG-TTGGTATAGCGGCTC-CCTACGAGCAG
SP12-B.06	2	1-2	3	CA-RQTELYGY-TFGSG	8	TGTGCCAG-ACAGACAGAGCT-CTATGGCTAC

SP12-B.07	24	2-1	3	CATSD-SLGGDYYNEQ-YFGPG	10	TGTGCCACCAGTGATT-CCCTCGGGGGGGACTAT-TACAATGAGCAG
SP12-B.08	6-1	1-5	3	CAS-REEPSGNQPQ-HFGDG	10	TGTGCCAGC-CGCGAAGAACCGTCGG-GCAATCAGCCCCAG
SP12-B.09	6-1	2-7	3	CASSE-SISYEQ-YFGPG	6	TGTGCCAGCAGTGAA-TCAAT-CTCCTACGAGCAG
SP12-B.10	20	2-7	2	CSAR-DGDSGGSYEQ-YFGPG	10	TGTAGTGCTAGAGA-CGGGGATAGCGGGGG-CTCCTACGAGCAG
SP12-B.11	5-6	2-1	2	CASS-GLKTTSSYNEQ-FFGPG	11	TGTGCCAGCAGC-GGACTCAAGACAACAAG-CTCCTACAATGAGCAG
SP12-B.12	11-3	2-3	1	CAS-RRINRAGSTDTQ-YFGPG	12	TGTGCCAGCAG-ACGGATTAACAGGGCGGGG-AGCACAGATACGCAG
no. of seqs			**48**			
no. individual clones			**12**			
clonotype ratio			**4.00**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
acEBV1-B.01	27	2-7	1	CAS-KIGQGAPYEQ-YFGPGTRLTVTED	10	CCTGAAGGGTACAAAGTCTCTCGAAAAGAGAAGAGGAATTTCCCCCTGATCCTGGAGTCGCCCAGCCCCAACCAGACCTCTCTGTACTTC
acEBV1-B.02	4-2/4-3	2-7	1	CASS-PGQGAPYEQ-YFGPGTRLTVTED	9	CCTGAATGCCCCAACAGCTCTCACTTATTCCTTCACCTACACACCCTGCAGCCAGAAGACTCGGCCCTGTATCT
acEBV1-B.03	28	2-7	29	CASSL-STGAPYEQ-YFGPGTRLTVTED	8	AAAGAAAAAGGAGATATTCCTGAGGGGTACAGTGTCTCTAGAGAGAAGAAGGAGCGCTTCTCCCTGATTCTGGAGTCCGCCAGCACCAACCAGACATCTATGTACCTC
no. of seqs			**31**			
no. individual clones			**3**			
clonotype ratio			**10.33**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
acEBV2-B.01	24	2-2	18	CATSD-WEGREAGELFF-GEGSRLTVLED	11	TGTGCCACCAGTGAT-TGGGAGGGCCGAGAGGCCGGGGAGCTGTTTTTTGGAGAAGG
acEBV2-B.02	2	2-1	5	CASS-GTGTYNEQ-FFGPGTRLTVLED	8	TGTGCCAGCAGT-GGGACAGGTACCTACAATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCG
acEBV2-B.03	4-2/4-3	2-2	4	CASSQ-EVASGTPGEL-FFGEGSRLTVLED	10	TGTGCCAGCAGCCAA-GAAGTGGCTAGCGGGACCCCCGGGGAGCTGTTTTTTGGAGAAGGCTCTA
acEBV2-B.04	24	1-2	3	CATSD-PTHGTGIYD-YTFGSGTRLTVVE	9	TGTGCCACCAGTGAT-CCGACCCACGGGACAGGGATATACGACTACACCTTCGGTTCGGG
acEBV2-B.05	10-3	1-1	1	CAISE-CSGLEA-FFGQGTRLTVVE	6	TGTGCCATCAGTGAG-TGTTCGGGGCTTGAAGCTTTCTTTGGACAAGGCACCAGACTCACAGTTGTAGA
acEBV2-B.06	28	1-1	1	CAS-GGYEGHTEA-FFGQGTRLTVVEDL	9	TGTGCCAGC-GGGGGTTATGAGGGACACACTGAAGCTTTCTTTGGACAAGGCACCAGACTCACAGTTG

acEBV2-B.07	6-6	1-4	1	CAS-GTPGQGGEKL-FFGSGTQLSVLED	10	TGTGCCAGC-GGCACACCCGGACAGGGAGGTGAAAAACTGTTTTTTGGCAGTGGAACCCAGCTCTCT
acEBV2-B.08	27	2-3	1	CASSL-SPSEAITDT-QYFGPGTRLTVLE	10	TGTGCCAGCAGTTTA-TCTCCTAGCGAGGCGATCACAGATACGCAGTATTTTGGCCCAGGCACCCGGC
acEBV2-B.09	7-8	1-1	1	CASSL-VRGGGETEA-FFGQGTRLTVVE	9	TGTGCCAGCAGCTTA-GTGCGAGGAGGGGGGGAAACTGAAGCTTTCTTTGGACAAGGCACCAGACTC
acEBV2-B.10	28	2-2	1	CASS-MTGSGEL-FFGEGSRLTVLED	7	TGTGCCAGCAGT-ATGACAGGGTCCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGACCGTAC
acEBV2-B.11	?	1-4	1	CAT-NEKL-FFGSGTQLSVLED	4	TGTGCAACT-AATGAAAAACTGTTTTTTGGCAGTGGAACCCAGCTCTCTGTCTTGGAGGAC
no. of seqs			**37**			
no. individual clones			**11**			
clonotype ratio			**3.36**			
Clone ID	**Vb**	**Jb**	**frequency**	**CDR3aa sequence**	**CDR3 aa length**	**Nucleotide sequence**
acEBV3-B.01	29-1	1-1	18	CSV-GDRQGYTEA-FFGQGTRLTVVE	9	TGCAGCGTC-GGTGATCGGCAGGGTTACACTGAAGCTTTCTTTGGACAAGGCACCAGACTCA
acEBV3-B.02	28	2-7	8	CASSL-AGQAYEQ-YFGPGTRLTVT	7	TGTGCCAGCAGTTTA-GCGGGACAGGCCTACGAGCAGTACTTCGGGCCGGGCACCAGGCTC
acEBV3-B.03	6-5	2-2	3	CASS-PAGPGEL-FFGEGSRLTVL	7	TGTGCCAGCAGC-CCTGCGGGGCCCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGA
acEBV3-B.04	6-6	1-4	2	CASS-PTPGQLNEKL-FFGSGTQLSV	10	TGTGCCAGCAGT-CCCACCCCGGGACAGCTTAATGAAAAACTGTTTTTTGGCAGTGGAA
acEBV3-B.05	6-5	2-2	2	CA-TSGAGTGEL-FFGEGSRLTVLED	9	TGTGCC-ACCTCAGGAGCGGGAACCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTG
acEBV3-B.06	24-1	2-2	2	CATS-KGVDTGEL-FFGEGSRLTVLE	8	TGTGCCACCAGC-AAGGGGGTTGACACCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGC
acEBV3-B.07	29-1	2-2	2	CSV-LGSGEL-FFGEGSRLTVLED	5	TTCAGCGTT-CTGGGATCCGGGGAGCTGTTTTTTGGAGAAGGCTCTAGGCTGACCGTACT
acEBV3-B.08	9	2-3	1	CASS-GYGGMGTDTQ-YFGPGTRLTVLED	10	TGTGCCAGCAGC-GGATACGGGGGGATGGGCACAGATACGCAGTATTTTGGCCCAG
acEBV3-B.09	12-3/12-4	2-7	1	CASSL-DSGTGLYEQ-YFGPGTRLTVTED	9	TGTGCCAGCAGTTTA-GACTCCGGGACAGGACTCTACGAGCAGTACTTCGGGCCGG
acEBV3-B.10	7-9	1-1	1	CASSL-IPGMGNTEA-FFGQGTRLTVVE	9	TGTGCCAGCAGCTTA-ATCCCAGGGATGGGGAACACTGAAGCTTTCTTTGGACAAGG
acEBV3-B.11	5-1	1-1	1	CASS-PLPTGSGNTEA-FFGQGTRLTVVE	11	TGCGCCAGCAGC-CCTCTACCGACAGGGTCGGGGAACACTGAAGCTTTCTTTGGACC
acEBV3-B.12	4-1	2-7	1	CASS-PNTDRSLQ-YFGPGTRLTVTED	8	TGCGCCAGCAGC-CCTAATACAGACAGGTCGTTACAGTACTTCGGGCCGGGCACCA

no. of seqs	**42**			
no. individual clones	**12**			
clonotype ratio	**3.50**			

TRAV sequences					
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
RF1-A.01	38-2/DV8	52	26	CA-LL-NAGGTSYGKLT	GCTTTACTAAATGCTGGTGGTACTAGCTATGGAAAGCTGACATTTGGACAAGGGACCATCTTGACTGTCCATCCA
RF1-A.02	4	52	5	CLVGD-T-NAGGTSYGKLT	CTCGTGGGTGACACTAATGCTGGTGGTACTAGCTATGGAAAGCTGACATTTGGACAAGGGACCATCTTGACTGTCCATCCA
RF1-A.03	4	43	5	CLV-VDL-NNDMR	CTCGTGGTCGATCTTAACAATGACATGCGCTTTGGAGCAGGGACCAGACTGACAGTAAAACCA
RF1-A.04	13-1	4	5	CAAS-TPASA-GGYNKLI	GCAGCAAGTACCCCCGCTTCGGCTGGTGGCTACAATAAGCTGATTTTTGGAGCAGGGACCAGGCTGGCTGTACACCCA
RF1-A.05	4	22	2	CLVGD--GSARQLT	CTCGTGGGTGATGGTTCTGCAAGGCAACTGACCTTTGGATCTGGGACACAATTGACTGTTTTACCT
RF1-A.06	4	34	1	C-HV-TDKLI	CACGTCACCGACAAGCTCATCTTTGGGACTGGGACCAGATTACAAGTCTTTCCA
no. of seqs			**44**		
no. individual clones			**6**		
clonotype ratio			**7.3**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
HBD11-A.01	29/DV5	40	33	CAAS-VP-TSGTYKYI	GCAGCAAGCGTACCTACCTCAGGAACCTACAAATACATCTTTGGAACAGGCACCAGGCTGAAGGTTTTAGCA
HBD11-A.02	5	42	2	CAE-G-GGSQGNLI	GCAGAGGGTGGAGGAAGCCAAGGAAATCTCATCTTTGGAAAAGGCACTAAACTCTCTGTTAAACCA
HBD11-A.03	22	35	2	CA-EEGL-GFGNVLHC/GSGTQVIVLP	GCTGAGGAGGGGTTAGGCTTTGGGAATGTGCTGCATTGCGGGTCCGGCACTCAAGTGATTGTTTTACCA
HBD11-A.04	4	23	1	CLVG-E-NQGGKLI	CTCGTGGGCGAGAACCAGGGAGGAAAGCTTATCTTCGGACAGGGAACGGAGTTATCTGTGAAACCC
HBD11-A.05	4	37	1	CLVG-VSQ-GNTGKLI	CTCGTGGGTGTGTCCCAGGGCAACACAGGCAAACTAATCTTTGGGCAAGGGACAACTTTACAAGTAAAACCA
HBD11-A.06	20	39	1	CA-GPF-NNAGNMLT	GCTGGGCCCTTTAATAATGCAGGCAACATGCTCACCTTTGGAGGGGGAACAAGGTTAATGGTCAAACCC

HBD11-A.07	20	30	1	CA-A-DDKII	GCTGCCGATGACAAGATCATCTTTGGAAAAGGGACACGACTTCATATTCTCCCCA
HBD11-A.08	19	28	1	CALSE-A-GAGSYQLT	GCTCTGAGTGAGGCTGGGGCTGGGAGTTACCAACTCACTTTCGGGAAGGGGACCAAACTCTCGGTCATACCA
HBD11-A.09	24	57	1	CAF-RNL-TQGGSEKLV	GCCTTTAGGAACTTAACTCAGGGCGGATCTGAAAAGCTGGTCTTTGGAAAGGGAATGAAACTGACAGTAAACCCA
no. of seqs			**43**		
no. individual clones			**9**		
clonotype ratio			**4.77**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
HBD21-A.01	38-2	45	20	CAYRS--GGGADGLT	GCTTATAGGAGCGGGGGAGGTGCTGACGGACTCACCTTTGGCAAAGGGACTCATCTAATCATCCAGCCC
HBD21-A.02	9-2	12	8	CAL-EG-DSSYKLI	GCCCTCGAGGGGGATAGCAGCTATAAATTGATCTTCGGGAGTGGGACCAGACTGCTGGTCAGGCCT
HBD21-A.03	26-2	52	4	CILRD--NAGGTSYGKLT	ATCCTGAGAGATAATGCTGGTGGTACTAGCTATGGAAAGCTGACATTTGGACAAGGGACCATCTTGACTGTCCATCCA
HBD21-A.04	22	17	3	CA-PFPP-AGNKLT	GCTCCCTTCCCCCCTGCAGGCAACAAGCTAACTTTTGGAGGAGGAACCAGGGTGCTAGTTAAACCA
HBD21-A.05	26-2	40	3	CILR-SS-TSGTYKYI	ATCCTGAGATCTTCTACCTCAGGAACCTACAAATACATCTTTGGAACAGGCACCAGGCTGAAGGTTTTAGCA
HBD21-A.06	4	4	2	CLVG-E-GGYNKLI	CTCGTGGGTGAAGGAGGCTACAATAAGCTGATTTTTGGAGCAGGGACCAGGCTGGCTGTACACCCA
HBD21-A.07	4	17	1	CLVG-AGH-KAAGNKLT	CTCGTGGGTGCCGGTCACAAAGCTGCAGGCAACAAGCTAACTTTTGGAGGAGGAACCAGGGTGCTAGTTAAACCA
HBD21-A.08	38-1	28	1	CAFM-F-SGAGSYQLT	GCTTTCATGTTCTCTGGGGCTGGGAGTTACCAACTCACTTTCGGGAAGGGGACCAAACTCTCGGTCATACCA
HBD21-A.09	27	23	1	CAG-MAGG-QGGKLI	GCAGGGATGGCCGGGGGGCAGGGAGGAAAGCTTATCTTCGGACAGGGAACGGAGTTATCTGTGAAACCC
no. of seqs			**43**		
no. individual clones			**9**		
clonotype ratio			**4.77**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
HBD22-A.01	4	37	15	CLVG-I-SGNTGKLI	CTCGTGGGTATCTCTGGCAACACAGGCAAACTAATCTTTGGGCAAGGGACAACTTTACAAGTAAAACCA
HBD22-A.02	4	38	4	CLV-DL-NAGNNRKLI	CTCGTGGATCTCAATGCTGGCAACAACCGTAAGCTGATTTGGGGATTGGGAACAAGCCTGGCAGTAAATCCGAA
HBD22-A.03	4	44	3	CLVGD-P-TGTASKLT	CTCGTGGGTGACCCTACCGGCACTGCCAGTAAACTCACCTTTGGGACTGGAACAAGACTTCAGGTCACGCTC
HBD22-A.04	4	33	2	CLVGD-SPT-DSNYQLI	CTCGTGGGTGACAGCCCCACGGATAGCAACTATCAGTTAATCTGGGGCGCTGGGACCAAGCTAATTATAAAGCCA
HBD22-A.05	12-3	18	2	CAMS-P-DRGSTLGRLY	GCAATGAGCCCCGACAGAGGCTCAACCCTGGGGAGGCTATACTTTGGAAGAGGAACTCAGTTGACTGTCTGGCCT

HBD22-A.06	5	6	1	CAE-IGRA/FGRGTSLIVHP	GCAGAGATCGGTCGGGCATTTGGAAGAGGAACCAGCCTTATTGTTCATCCG
HBD22-A.07	12-2	3	1	CA-AP-YSSASKII	GCCGCCCCGTACAGCAGTGCTTCCAAGATAATCTTTGGATCAGGGACCAGACTCAGCATCCGGCCA
HBD22-A.08	38-2	31	1	CAYR-RS-NARLM	GCTTATAGAAGGAGCAATGCCAGACTCATGTTTGGAGATGGAACTCAGCTGGTGGTGAAGCCC
HBD22-A.09	4	43	1	CLVG-G-NDMR	CTCGTGGGTGGAAATGACATGCGCTTTGGAGCAGGGACCAGACTGACAGTAAAACCA
HBD22-A.10	4	37	1	CLVG-VP-SGNTGKLI	CTCGTGGGTGTCCCCTCTGGCAACACAGGCAAACTAATCTTTGGGCAAGGGACAACTTTACAAGTAAAACCAGA
HBD22-A.11	30	26	1	CGT-V-DNYGQNFV	GGCACTGTCGATAACTATGGTCAGAATTTTGTCTTTGGTCCCGGAACCAGATTGTCCGTGCTGCCC
no. of seqs			**32**		
no. individual clones			**11**		
clonotype ratio			**2.91**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
HCW4-A.01	4	28	9	CRP-GAGSYQLT	CGCCCGGGGGCTGGGAGTTACCAACTCACT
HCW4-A.02	38-2	45	8	CAY-SD-SGGGADGLT	GCTTATAG-TGATTCAGGAGGAGGTGCTGACGGACTCACC
HCW4-A.03	38-2	42	7	CA-P-YGGSQGNLI	GC-CCCCTATGGAGGAAGCCAAGGAAATCTCATC
HCW4-A.04	4	3	3	CLVG-S-GYSSASKII	CTCGTGGG-GTCGGGGTACAGCAGTGCTTCCAAGATAATC
HCW4-A.05	19	40	2	CALSE-GS-SGTYKYIA	GCTCTGAGTGAGGGGAGCTCAGGAACCTACAAATACATC
HCW4-A.06	14	11	2	CAMR-AAR-NSGYSTLT	GCAATGAGAGCCGCCCGGAATTCAGGATACAGCACCCTCACC
HCW4-A.07	4	3	1	CLVG-ER-SSASKII	CTCGTGGGTGA-AAGGAGCAGTGCTTCCAAGATAATC
HCW4-A.08	4	50	1	CLVGD-NL-KTSYDKVI	CTCGTGGGTGACA-ACCTAAAAACCTCCTACGACAAGGTGATA
HCW4-A.09	4	30	1	CLVGD-E-NRDDKII	CTCGTGGGTGAC-GAGAACAGAGATGACAAGATCATC
HCW4-A.10	4	13	1	CLVG-GSA-GGYQKVT	CTCGTGGGTG-GGTCCGCCGGGGGTTACCAGAAAGTTAC
HCW4-A.11	19	21	1	CAL-GE-NFNKFY	GCTCTGGGGGAAAACTTCAACAAATTTTAC
HCW4-A.12	13-1	6	1	CAA-TLE-SGGSYIPT	GCAGCAACTCTTGAATCAGGAGGAAGCTACATACCTACA
no. of seqs			**37**		
no. individual clones			**12**		
clonotype ratio			**3.1**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
HCW6-A.01	14/DV4	13	39	CAMRE-V-SGGYQKVT	GCAATGAGAGAGGTCTCTGGGGGTTACCAGAAAGTTACCTTTGGAACTGGAACAAAGCTCCAAGTCATCCCA

HCW6-A.02	16	4	4	CAL-FM-FSGGYNKLI	GCTCTCTTCATGTTTTCTGGTGGCTACAATAAGCTGATTTTTGGAGCAGGGACCAGGCTGGCTGTACACCCA
HCW6-A.03	22	35	1	CA-EEGL-GFGNVLHC/GSGTQVIVLP	GCTGAGGAGGGGTTAGGCTTTGGGAATGTGCTGCATTGCGGGTCCGGCACTCAAGTGATTGTTTTACCA
no. of seqs			**44**		
no. individual clones			**3**		
clonotype ratio			**14.66**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
acEBV-A.01	4	53	20	CLVGD-EGR-SGGSNYKLTFGKGTLLTVNPN	TGCCTCGTGGGTGACGAGGGAGGAGTGGAGGTAGCAACTATAAACTGACATTTGGAAAAGGAACTCTCTTAACCGT
acEBV-A.02	4	31	11	CLVG-AD-NARLMFGDGTQLVVKPNIQNPDP	TGCCTCGTGGGTGCCGACAATGCCAGACTCATGTTTGGAGATGGAACTCAGCTGGTGGTGAAGCCCAATATCCAGAA
acEBV-A.03	4	30	3	CLV--RDDKIIFGKGTRLHILPNIQNPDPAV	TGCCTCGTAAGAGATGACAAGATCATCTTTGGAAAAGGGACACGACTTCATATTCTCCCCAATATCCAGAACCCTGA
no. of seqs			**34**		
no. individual clones			**3**		
clonotype ratio			**11.3**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
acEBV2-A.01	22	35-orf	16	CAVE-G-GFGNVLHCGSGTQVIVLPHIQNPDP	TGTGCTGTGGAGGGGGGCTTTGGGAATGTGCTGCATTGCGGGTCCGGCACTCAAGTGATTGTTTTACCACATATCCAGAACCCT
acEBV2-A.02	10	4	14	CVV-RGPP-SGGYNKLIFGAGTRLAVHPYIQNPDP	TGTGTGGTGAGGGGCCCTCCTAGTGGTGGCTACAATAAGCTGATTTTTGGAGCAGGGACCAGGCTGGC
acEBV2-A.03	9?	36	7	CAL-GDLNP-GANNLFFGTGTRLTVIPYIQN	TGTGCTCTTGGGGACCTTAACCCCGGGGCAAACAACCTCTTCTTTGGGACTGGAACGAGACTCACCGTTATTCCCTATATCCAGAACCC
acEBV2-A.04	4	41	7	CLV-DPP-NSGYALNFGKGTSLLVTPHIQNPDP	TGCCTCGTGGACCCACCAAATTCCGGGTATGCACTCAACTTCGGCAAAGGCACCTCGCTGTTGGTCACACCCCATATCCAGA
acEBV2-A.05	4	5	1	CLVG-GRLE-DTGRRALTFGSGTRLQVQPNIQNPDP	TGCCTCGTGGGTGGACGCCTGGAAGACACGGGCAGGAGAGCACTTACTTTTGGGAGTGGAACAAGACTCCAAGTGCAA
no. of seqs			**45**		
no. individual clones			**5**		
clonotype ratio			**9.0**		
Clone ID	**Va**	**Ja**	**frequency**	**CDR3aa sequence**	**Nucleotide sequence**
acEBV3-A.01	12-3	18	10	CAMS-P-DRGSTLGRLYFGRGTQLTVWPDIQNP	TACCTCTGTGCAATGAGCCCCGACAGAGGCTCAACCCTGGGGCGGCTATACTTTGGAAGAGGAACTCAGTTGACTGTCTGGCCTGATATCCAGAACCC
acEBV3-A.02	4	39	8	CLVG-GP-NNAGNMLTFGGGTRLMVKPHIQNPDPAVYQL	TACTACTGCCTCGTGGGTGGCCCAAATAATGCAGGCAACATGCTCACCTTTGGAGGGGGAACAAGGTTAATGGTCAAACCCCATATGTACCAGCTG

acEBV3-A.03	1-2	7	6	C-GD-NRLAFGKGNQVVVIPNIQNPDP	TACCTCTGTGGGGACAACAGACTCGCTTTTGGGAAGGGGAACCAAGTGGTGGTCATACCAAATATCCAGAACCCTGACCCT
acEBV3-A.04	19	29	5	CALS-DP-SGNTPLVFGKGTRLSVIANIQNPDP	TACTTCTGTGCTCTGAGTGACCCTTCAGGAAACACACCTCTTGTCTTTGGAAAGGGCACAAGACTTTCTGTGATTGCAAATATCCAGAACCCTGACCC
acEBV3-A.05	4	47	3	CLVGD-GL-YGNKLVFGAGTILRVKSYIQNPDPAVYQL	TACTACTGCCTCGTGGGTGACGGTTTATATGGAAACAAACTGGTCTTTGGCGCAGGAACCATTCTGAGAGTCAAGTCCTATATCCAGAACCCTGACCCTGCCGTGTACCAGCTG
acEBV3-A.06	12-2	50	2	CA-GTEE-TSYDKVIFGPGTSLSVIPNIQNPDPAVYQL	TACCTCTGTGCCGGGACGGAGGAAACCTCCTACGACAAGGTGATATTTGGGCCAGGGACAAGCTTATCAGTCATTCCAAATATCCAGAACCCTGACCCTGCCGTGTACCAGCT
acEBV3-A.09	17	42	1	CATD-EG-GGSQGNLIFGKGTKLSVKPNIQNPDP	TACTACTGCCTCGTGGGTGACCATGAACAGAGATGACAAGATCATCTTTGGAAAAGGGACACGACTTCATATTCTCCCCAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTG
acEBV3-A.10	1-2	12	1	CAV--MDSSYKLIFGSGTRLLVRPDIQNPDP	TACTTCTGCGCTCCTCTCCGGGGGATAACCAGGGAGGAAAGCTTATCTTCGGACAGGGAACGGAGTTATCTGTGAAACCCAATATCCAGAACCCTG
acEBV3-A.11	26-1	31	1	CIV-PY-NARLMFGDGTQLVVKPNIQNPDP	TACTTCTGTGCTACGGACGAAGGGGGAGGAAGCCAAGGAAATCTCATCTTTGGAAAAGGCACTAAACTCTCTGTTAAACCAAATATCCAGAACCCTGACCCT
acEBV3-A.12	4	15	1	CLV-D-QAGTALIFGKGTTLSVSSNIQNPDP	TACCTCTGTGCTGTGATGGATAGCAGCTATAAATTGATCTTCGGGAGTGGGACCAGACTGCTGGTCAGGCCTGATATCCAGAACCCTGACCCT
acEBV3-A.13	4	8	1	CLVGD-NQS-TGFQKLVFGTGTRLLVSPNIQNPDPAVYQL	TACTATTGCATCGTCCCTTACAATGCCAGACTCATGTTTGGAGATGGAACTCAGCTGGTGGTGAAGCCCAATATCCAGAACCCTGACCCT
acEBV3-A.14	4	9	1	CLVG-VPGGA-GGFKTIFGAGTRLFVKANIQNPDPAVYQL	TACTACTGCCTCGTCGACCAGGCAGGAACTGCTCTGATCTTTGGGAAGGGAACCACCTTATCAGTGAGTTCCAATATCCAGAACCCTGACCCT
acEBV3-A.15	4	23	1	C-PL-YNQGGKLIFGQGTELSVKPNIQNPDPAVYQL	TACTACTGCCTCGTGGGTGACAATCAAAGCACAGGCTTTCAGAAACTTGTATTTGGAACTGGCACCCGACTTCTGGTCAGTCCAAATATCCAGAACCCTGACCCTGCCGTGTACCAGCTG
no. of seqs			**41**		
no. individual clones			**15**		
clonotype ratio			**3.15**		

Supplementary Table S3: Comparison of CDR1 and CDR2 amino acid sequences

	CDR1											CDR2									
TRBV-4	Q	H	M	G	H	-	N	A	M	Y		F	V	Y	S	x	-	E	E	x	x
TRBV-6	Q	D	M	N	H	-	N	Y	M	Y		Y	Y	S	V	x	-	A	G	x	T
TRBV-24	Q	T	M/K	G	H	-	D	R	M	Y		Y	Y	S	F	D	-	V	K	D	I
TRBV-9	P	R	S	G	D	-	L	S	Y	Y		I	Q	Y	Y	N	-	G	E	E	R
TRBV-12	P	I	S	G	H	-	N	x	L	F		I	Y	F	x	N	-	x	x	P	x
TRBV-29	V	D	S	Q	V	-	T	M	M	F		T	A	N	Q	G	-	S	E	A	T

CrossTope ID	Structure Source	Structure Type	Source Virus	Source Protein	Epitope Position	Sequence	Epitope ID (IEDB)
A0201_0068	CrossTope	Model (D1-EM-D2)	Hepatitis C virus (WT)	NS3	1073-1081	CINGVCWTV	6435
A0201_0031	CrossTope	Model (D1-EM-D2)	Hepatitis C virus (GT1b)	NS3	1073-1081	CVNGVCWTV	7292
A0201_0109	CrossTope	Model (D1-EM-D2)	Influenza A Virus	NA	231-239	CVNGSCFTV	7291
A0201_0095	PDB	Crystal (2V2W)	Human Immunodeficiency Virus 1	GAG	77-85	SLYNTVATL	59613
A0201_0073	CrossTope	Model (D1-EM-D2)	Epstein-Barr Virus/Human Herpesvirus 4	LMP2	329-337	LLWTLVVLL	37960
A0201_0016	CrossTope	Model (D1-EM-D2)	Epstein-Barr Virus/Human Herpesvirus 4	BMLF1	259-267	GLCTLVAML	20788
A0201_0072	CrossTope	Model (D1-EM-D2)	Epstein-Barr Virus/Human Herpesvirus 4	Putative BARF0 protein	356-364	LLWAARPRL	37938
A0201_0080	CrossTope	Model (D1-EM-D2)	Epstein-Barr Virus/Human Herpesvirus 4	LMP1	125-133	YLLEMLWRL	74774
A0201_0110	CrossTope	Model (D1-EM-D2)	Epstein-Barr Virus/Human Herpesvirus 4	K12	17-25	LLNGWRWRL	37607

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.

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

Dinler A Antunes[1*], Maurício M Rigo[1], Martiela V Freitas[1], Marcus FA Mendes[1], Marialva Sinigaglia[1], Liisa K Selin[2], Markus Cornberg[3], Gustavo F Vieira[1§]

[1] NBLI – Núcleo de Bioinformática do Laboratório de Imunogenética, Department of Genetics, UFRGS, RS/Brazil.

[2] Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA.

[3] Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany.

[§] Corresponding author.

Email addresses:

DAA: dinler@gmail.com

MMR: mauriciomr1985@gmail.com

MVF: martielafreitas@gmail.com

MFAM: cla_atm_milo@hotmail.com

MS: msinigaglia@gmail.com

LKS: liisa.selin@umassmed.edu

MC: cornberg.markus@mh-hannover.de

GFV: gusfioravanti@yahoo.com.br

Abstract

Memory T cells specifically recognize peptide-loaded Major Histocompatibility Complexes (pMHC). This specificity, however, is tempered by cross-reactivity, a phenomenon with direct consequences over heterologous immunity between viruses. Here we summarize into three intuitive cross-reactivity networks data previously tested through careful experimentation with virus-derived epitopes in context of both human and murine MHC allotypes. Furthermore, using crystal structures and modeled pMHC complexes we perform an innovative structure-based hierarchical clustering of cross-reactive targets. Our predictions are well fitted to experimental data, even in cases involving nonrelated epitopes from heterologous viruses, with less than 40% of sequence similarity. Besides of its use as a tool for prospecting unknown cross-reactive targets, this approach can also provide insights into complex features involved in heterologous immunity. Our data is in agreement with recent studies suggesting the importance of shared structural features to trigger cross-reactive T cell responses. The impressive complexity of this system, involving MHC polymorphism, viral sequence variability and T cell somatic recombination, represents a major immunology puzzle, which will not be easily solved or predicted. However, we here suggest that pMHC structural similarity can be used as an index of the probability for cross-reactive T cell responses, which would have several applications from basic immunology research to vaccine design.

Introduction

Antiviral immunity is one of the most important and challenging tasks performed by adaptive immune system in jawed vertebrates [1,2]. Viruses are mandatory intracellular parasites which uses the host molecular machinery to replicate [3,4]. They usually have fast replication cycles and poor accuracy control over genome replication, which among other features are responsible for fast evolution and sequence diversity [5]. Through co-evolution, jawed vertebrates developed a complex and highly variable system to identify and eliminate these parasites [6,7,8], as well as to prevent future infections [9]. The endogenous peptide presenting pathway, found in most vertebrates' nucleated cells, allows a "quality control" system, by sampling cytoplasmic proteins which will be digested into small peptides (or epitopes) and presented at the cell surface in the context of Major Histocompatibility Complex (MHC) molecules [6]. Through this system, virus-infected cells will be "targeted" for cellular immunity, by presenting virus-derived epitopes at cell surface. The "MHC region", which encodes most proteins involved in this pathway, is the most polymorphic and dense region of human genome [6], highlighting the influence of virus diversity (scape mutations) over human cellular immunity.

In order to eliminate infected cells, however, these peptide:MHC complexes (pMHC) must be specifically recognized by Cytotoxic T Lymphocytes (CTLs), which raises a fundamental question: How to produce and store a pool of memory CTLs able to specifically recognize most of these hugely variable pMHC complexes? The answer involves a combination of two important features of cellular immunity, (i) somatic recombination of antigen receptor genes and (ii) cross-reactivity. The first, allows a potential combinatorial diversity of the T cell receptor (TCR) which exceeds 10^{20} [10]. The second, allows to optimize the repertoire of T cells to recognize most of possible targets, despite the limited number of CTLs possible to exist in a given individual, in a given time ($\approx 10^{11}$ in humans) [10,11].

T cells are polyspecific, in the sense that they can specifically interact with more than one pMHC complex [12], and can be cross-reactive, meaning that can be activated by two or more heterologous targets [11]. This cross-reactivity can even mediate a heterologous immunity, when a contact with one pathogen generates a

partial immunity against a second (heterologous) pathogen [13]. Heterologous immunity can be protective and desired for wide spectrum vaccine development [14,15], but can also mediate impaired cellular response, chronic infection and immunopathology [15,16,17,18]. The random nature of TCR specificity generation (through somatic recombination) entails that each individual has a unique set of TCRs (private specificity)[13]. In addition to that, given the referred size limit of the CTL repertoire and the constant challenges with a diverse variety of pathogens, our immunological memory is ever changing. In time, cross-reactive cells will represent an important part of our memory repertoire and our immunity against each new challenge will be directly influenced by our immunological history [11,13,19,20].

Recent studies are corroborating the idea that cross-reactivity in CTL recognition is rather the rule than the exception [17,20,21], and that structural features involved in specific TCRpMHC interactions are the main features driving cross-reactive responses against heterologous targets [21,22,23]. In previous works, our group described an *in silico* approach to predict the 3D structures of pMHC complexes that were not yet determined by experimental methods [24,25], and also described a structure-based analysis to predict cross-reactivity among nonrelated virus-derived epitopes in the context of the human MHC allotype HLA-A*02:01 [26]. Some of these predictions, performed with hierarchical clustering analysis of modeled pMHCs, were later confirmed by *in vitro* experiments (Zhang *et al.* personal communication), highlighting the prospective potential of this innovative method. Here, we further explore these structural analyses over a subset of previously tested cross-reactive targets, providing new insights on the molecular features driving cross-reactivity and heterologous immunity.

Results/Discussion

Mapping cross-reactivity networks

In 2010, Cornberg and colleagues [20] described cross-reactivity networks (CRNs) involving virus-derived epitopes, both within human and murine CD8+ memory T cell pools. For instance, they demonstrate that one Vaccinia-derived epitope (VV-$A11_{198}$) was able to activate three different LCMV-specific memory populations (LCMV-GP_{34}, LCMV-GP_{118} and LCMV-NP_{205}), and that the pattern of cross-reactivity was partially determined by changes in private specificities of memory repertoires. Cross-reactivity is far more common than initially predicted [21], and graphical representations of CRNs have been used in other works [13,27,28,29].

In addition to experiments that supported the CRNs described by Cornberg *at al.* 2010 [20], new experimental and structural data on these cross-reactive targets has been recently made available [17,23,26,30], helping to explain and expand these networks. Based on these previous publications (Table S1), we developed a new representation scheme to summarize the data on three cross-reactivity networks (Figure 1). Inspired by representations of enzymatic reactions, this scheme allows to intuitively indicate the directionality and reciprocity of cross-reactive reactions, as well as to represent tested noncross-reactive targets.

Reviewing these publications, we observed that some targets were referred with different abbreviations, especially in which regards to epitope position. For instance, the same EBV-derived epitope "GLCTLVAML" was previously referred as $BMLF1_{259}$, $BMLF1_{280}$ and $BMLF1_{300}$. Performing a careful verification of each sequence at Uniprot [31] we were able to determine the correct epitope position and recommended protein name for all targets, providing a standardized reference for future studies in the field (Table S1).

Structure-based clustering of crystal structures

Out of the 25 virus-derived epitopes included in the H2-K^b-restricted network, only 6 had its structure determined by experimental methods (Figure 1A). Inspired on a previously described structure-based approach [26], we performed an innovative hierarchical clustering analysis (HCA) of these 6 crystallographic structures (Figure S1A). Of note, 3TID is referred as the crystal structure of LCMV-GP_{34}:H2-K^b complex

[23], despite presenting an amino acid exchange at P8 (LCMV-GP_{34}-C8M). According to the authors who described the structure, this exchange has no significant impact on TCRpMHC interactions and this C8M variant was used in previous studies as an "equivalent" to the wild-type sequence. Here, sequence divergence between LCMV-GP_{34} and LCMV-GP_{34}-C8M is indicated in Figure 1A, but 3TID was considered as being the crystal structure of LCMV-GP_{34} for all structure-based analysis performed.

Supported by multiscale bootstrap resampling with *pvclust* R package [32], the HCA reproduced experimental data. The cross-reactive targets VV-$A11_{198}$ (3TIE) and LCMV-GP_{34} fall in the same cluster, and the same is observed for the highly cross-reactive targets LCMV-NP_{205} (3P4M) and PV-NP_{205} (3P4N). All these four targets are closer to one another than with the noncross-reactive target OVA_{258} (1VAC). Finally, the most divergent structure in this analysis was 3P4O, which contains the noncross-reactive scape variant LCMV-NP_{205}-V3A [19,30].

Expanding the K^b-restricted network with pMHC modeling

In their study with the H2-K^b-restricted CRN [20], Cornberg and colleagues discussed that observed patterns of cross-reactivity presented a within-individual variation which was driven by private specificities of each memory T cell repertoire, and immunological history. For instance, VV-$A11_{198}$-specific T cells expanded *in vitro* from a polyclonal pool of CTLs harbored from LCMV-immune mice presented cross-reactivity with LCMV-GP_{34}, LCMV-GP_{118}, LCMV-NP_{205} and PV-NP_{205} [20]. However, cross-reactivity against other VV-derived epitope, VV-$E7_{130}$ was not observed for these cells. Of note, if the donor had no previous contact with VV epitopes, than these expanded VV-$A11_{198}$-specific T cells should be actually cross-reactive cells, primarily expanded *in vivo* by recognizing some LCMV target. On the other hand, if the same experiment was performed expanding VV-$A11_{198}$-specific T cells from VV-immune mice, cross-reactivity with VV-$E7_{130}$ and LCMV-GP_{34} was observed, but no cross-reactivity was observed with LCMV-GP_{118}, LCMV-NP_{205} and PV-NP_{205}. These results suggest the use of a different T cell population, with a different specificity [20]. They also suggest a greater structural similarity between VV-$A11_{198}$ and LCMV-GP_{34}, since this cross-reactivity was observed both for LCMV-immune and VV-immune background. Structural similarity between these targets was later discussed by Shen *et*

al. 2013 [23], which solved the crystal structures of VV-A11$_{198}$ (3TIE) and LCMV-GP$_{34}$-C8M (3TID).

In previous studies, our group described and tested a docking-based protocol for modeling pMHC complexes [24,25,26]. Here, this approach was used to expand the HCA of H2-K^{b}-restricted targets, modeling these other complexes previously tested by Cornberg *et al.* 2010 [20] (Figure 1). We also included in this analysis two unrelated epitopes, VV-C4$_{125}$ and LCMV-GAG$_{70}$, as putative noncross-reactive controls (Table S1). Our expanded HCA corroborates the idea of greater structural similarity between VV-A11$_{198}$ and LCMV-GP$_{34}$, since both complexes fall in the same cluster, with the *edge* presenting the lowest *Height* and the highest AU/BP values (Figure 2). Epitopes LCMV-GP$_{118}$ and VV-E7$_{130}$, which are cross-reactive with VV-A11$_{198}$, fall in the next branch, followed by a cluster with the other cross-reactive targets LCMV-NP$_{205}$ and PV-NP$_{205.}$ All these cross-reactive targets were grouped into a bigger cluster (*edge* 5), separated from all the noncross-reactive targets. This is an impressive result. As discussed by the authors [20], these cross-reactivities could not be easily predicted based on epitope sequence similarity, since all these epitopes present less than 50% of amino acids identity. For instance, identity between VV-A11$_{198}$ and LCMV-GP$_{34}$ is of only 37.5%, the same percentage shared between VV-A11$_{198}$ and the noncross-reactive target OVA$_{258}$. Therefore, this cluster of cross-reactive targets (*edge* 5) involving nonrelated epitopes of three different viruses, could be predicted by an *in silico* analysis of these pMHC structures.

Cross-reactivity was observed among these targets [20,23] and they present structural similarities, being clustered together in our structure-based HCA. However, there was no experimental evidence of one T cell population able to recognize all these six epitopes [20]. As already discussed, cross-reactivity patterns would depend on the specific T cell population tested. Trying the summarize the information, we could say that CTLs from LCMV-immune mice were more cross-reactive, and we could represent them with a higher threshold, including all members of *edge* 5 in our HCA (Figure 2). This threshold would correctly predict most of the observed cross-reactivities, with the exception of VV-E7$_{130}$ (which was not recognized). On the other hand, CTLs from VV-immune mice could be represented with a lower threshold (*edge* 4), since in this context VV-A11$_{198}$-specific T cells do not recognized both NP$_{205}$ epitopes. The

exception in this case, would be LCMV-GP_{118}. These exceptions cannot be predicted considering the information provided by pMHC structures, since they are driven by TCR variability and private specificities. In spite of that, our data suggests a correlation between structural similarity and the probability to find cross-reactivity among targets. Therefore, we could postulate that although cross-reactivity between LCMV-GP_{118} and VV-$E7_{130}$ was not observed using VV-$A11_{198}$-specific or VV-$E7_{130}$-specific T cells [20], this cross-reactivity should be observed using some other T cell population, maybe with LCMV-GP_{118}-specific T cells (Figure S3).

We also further extended our HCA by including all the remaining complexes depicted in our H2-K^b-restricted network (Figure 1). Unfortunately, in its current state, our clustering approach was not able to perform an accurate prediction of close related targets, such as some alanine exchanged epitopes (data not shown). For instance, in the HCA with all the K^b-restricted network, our main noncross-reactive controls fall correctly in separated branches (*e.g.* OVA_{258} and LCMV-GAG_{70}) and some of our main cross-reactive controls fall together in a cluster (*e.g.* PV-NP_{205} and LCMV-NP_{205}). However, a clear separation was not observed for several noncross-reactive mutated epitopes, such as LCMV-GP_{34}-T7A-C8M and OVA_{258}-AA. Although presenting a great potential to prospect nonrelated cross-reactive targets, even without any sequence similarity [26](Zhang *et al*, personal communication), our approach needs to be refined in order to have a higher "resolution" for clustering close related epitopes.

Peculiarities in other cross-reactivity networks

Out of the 9 virus-derived epitopes included in the HLA-A*02:01-restricted network (Figure 1B), only 4 had available crystal structures. We modeled the remaining complexes and performed an HCA with *pvclust* (Figure S2). As expected, the cross-reactive targets EBV-$BMLF1_{300}$, IAV-$M1_{58}$, HCV-$NS3_{1073}$, HIV-GAG_{77} and EBV-$LMP2_{329}$ were clustered together (*edge* 5). These last two structures were actually the most similar pair of structures inside this cluster, in agreement with previous HCAs performed by our group [26].

The two noncross-reactive variants of HCV-$NS3_{1073}$ derived from HCV genotype 3, previously referred as *G3-14* and *G3-18* [26,33], fall in separate branches. Despite of being the outermost branch at the main cluster (*edge* 6), the small distance between *G3-14* and the cross-reactive targets suggest that cross-reactivity with this HCV-derived

scape variant might be observed depending on the T cell population tested. Interestingly, the complex presenting EBV-BRLF1$_{109}$ falls in the same branch as *G3-18*, which is far from its cross-reactive targets IAV-M1$_{58}$ and EBV-BMLF1$_{300}$. This HCA result was due to a negatively charged spot in the surface of EBV-BRLF1$_{109}$:H2-K^{b} complex, which was not seen in its cross-reactive counterparts (Figure 3A). If we remove from our analysis this negatively charged spot, EBV-BRLF1$_{109}$ is clustered with IAV-M1$_{58}$ and EBV-BMLF1$_{300}$ (data not shown). But these cross-reactivities involving EBV-BRLF1$_{109}$ have some peculiarities. For instance, it is not observed for most T cell populations and normally respects a given directionality, from EBV-BMLF1$_{300}$ (or IAV-M1$_{58}$) to EBV-BRLF1$_{109}$ [20]. EBV-BRLF1$_{109}$-specific T cells recovered from EBV-immune individuals and expanded *in vitro* in the presence of the cognate epitope present higher affinity/avidity in TCRpMHC interaction, and are not cross-reactive with EBV-BMLF1$_{300}$ or IAV-M1$_{58}$. On the other hand, EBV-BMLF1$_{300}$-specific T cells expanded *in vitro* in the presence of the cognate epitope might also recognize EBV-BRLF1$_{109}$ [20]. Further expansion of this population with this heterologous epitope, will produce (cross-reactive) EBV-BRLF1$_{109}$-specific T cells with lower affinity/avidity in TCRpMHC interaction (data not shown).

A similar situation is observed in the H2-D^{b}-restricted network (Figure 1C). Some cross-reactivity was observed between IAV-PA$_{224}$ and LCMV-GP$_{276}$ [17], and no cross-reactivity was observed with LCMV-NP$_{366}$. Our structural analysis agrees with that result, with LCMV-NP$_{366}$ (1HOC) falling in a completely separated branch (Figure S4). However, IAV-PA$_{224}$ (1WBY) presents a positively charged spot which differs from LCMV-GP$_{276}$ (1JPF), thus preventing their clustering (Figure 3). This positively charged spot over IAV-PA$_{224}$ surface is given by an arginine residue at P7. We included in our HCA the structure of a mutated epitope IAV-PA$_{224}$-R7A (1YN7), which falls in the same cluster as LCMV-GP$_{276}$ (1JPF). The authors who described these structures discussed that IAV-PA$_{224}$ has a “stronger flavor” [34,35], stimulating a diverse repertoire of TCRs in a process driven by private specificities. On the other hand, the “vanilla” (or featureless) peptide IAV-PA$_{224}$-R7A selected a more limited repertoire of TCRs in each mouse, with common TCR usage among mice (public TCRs).

TCRs interact with pMHCs in a more or less “canonical” binding mode [21,36,37,38], but it was shown that a given TCR can preferentially use distinct

residues to contact different complexes [39] or even modify its CDR loops to accommodate different peptides [40]. Considering these issues, it could be argued that immunization with a "featured" epitope (such as EBV-BRLF1$_{109}$ or IAV-PA$_{224}$) will trigger a highly polyclonal T cell response, with a broad spectrum of TCR specificities. Some of these are less specific to the homologous target, and more cross-reactive with other epitopes, probably "focusing" the interaction in surface regions which are shared among these targets (Figure 3). On the other hand, some of these cells present higher affinity/avidity with this homologous epitope, by "focusing" the interaction in unique features of its surface. Therefore, cross-reactivity between a "featured" and a "featureless" epitope will depend on which of these possible T cell populations are being tested.

Since these higher affinity/avidity cells will expand preferentially, they will dominate the pool of responding T cells in a homologous challenge with the "featured" epitope. However, in a heterologous challenge with a "featureless" epitope, the pool of responding cells will be dominated by cross-reactive T cells which are able to recognize both targets [17,19,20]. This helps to explain the referred "peculiarities" in the cross-reactivities of "featured" epitopes, such as the response directionality [41]. It is easier to find cross-reactivity against EBV-BRLF1$_{109}$ after consecutive rounds of expansion with the "featureless" EBV-BMLF1$_{300}$, but is far more difficult to find cross-reactivity performing the inverse experiment.

Conclusion

Structural similarity among pMHC complexes, in terms of topography and electrostatic potential over the TCR-interacting surface, is one of the main features driving the probability of cross-reactive T cell responses. Cross-reactivity is highly likely to be observed between two structurally identical complexes, for most T cell populations which recognize one of the complexes, and in both directions. On the other hand, is highly unlikely to find any T cell population capable of recognizing two completely different pMHC complexes. In most cases, however, two complexes will share some features and diverge about others. In this situation, cross-reactivity can be estimated by the level of structural similarity, but its occurrence, intensity and

directionality will be driven by the specific T cell population stimulated with the first target and selectively expanded after heterologous challenge.

Our innovative structure-based approach for predicting cross-reactive targets seems to be a promising tool for vaccine development, especially in which regards to prospecting nonrelated virus-derived epitopes. However, its resolution to analyze close related epitopes needs to be improved. Moreover, cross-reactivities among “featured” and “featureless” epitopes represent an especially difficult challenge for structure similarity based approaches, and additional tools/strategies must be developed to address this issue. Nevertheless, despite all variability involved in this highly complex system, structure-based analysis is an important tool capable of providing insightful ideas to help testing and explaining how T cells recognize their targets.

Methods

Experimental data on CRNs

Cross-reactivity networks (CRNs) depicted in Figure 1 were compiled from previously published experiments. Most data was made available by Cornberg and colleagues, who first represented these CRNs [20]. They also described a scape variant of LCMV-NP_{205} with a V3A substitution [19], suggested its sequence similarity with epitopes from Old World Arenaviruses (MOPV-NP_{205} and LASV-LNP_{209}) and finally solved its 3D crystal structure in the context of H2-K^b [30]. This study with murine cross-reactivities was further explored by Shen *et al*. 2013 [23]. The murine H2-D^b-restricted network was depicted with data from Wlodarczyk *et al.* 2013 [17].

Cornberg *et al.* 2010 [20] also described a human HLA-A*02:01-restricted network. We expanded this network including a cross-reactive target prospected through structural *in silico* analysis [26] and already confirmed experimentally (Zang S, personal communication) and two noncross-reactive targets described by Fytili *el al.* 2008 [33]. These tested noncross-reactive targets were included both in human and murine CRNs in order to provide further experimental information to test our structure based cross-reactivity prediction method.

A careful verification of epitopes' information was performed to determine the correct protein name and epitope position for each target, providing an updated reference for future studies (Table S1). Curated information from Uniprot [31] was used as the main reference, and GenBank (NCBI) was also consulted. References to Immune Epitope Database (IEDB) and Protein Data Bank (PDB) are also provided, if available.

Crystal structures

Crystal structures were obtained from Protein Data Bank (PDB) [42]. In each case, duplicated structures, water molecules and heteroatoms were removed from the "pdb" coordinates file using Pymol Viewer [43] and the remaining pMHC structure was submitted to a short energy minimization with Gromacs 4.5.1 package [44].

Modeled structures

Peptide:MHC complexes without experimentally-determined structures available at PDB were predicted using an automated version of the previously

described *D1-EM-D2* approach [24]. A text file with the epitope sequence in the FASTA format was provided, and the MHC allotype was selected, triggering a pipeline which returns a 3D structure of the pMHC complex in the pdb format. Briefly, the initial ligand structure was obtained by mutating an epitope structure obtained in the context of the same allotype ("Epitope_pattern"), step performed with Pymol scripts. A reference crystal structure of the MHC allotype of interest (without its ligand) was used as a receptor ("MHC_donor") for a molecular docking with the new ligand (flexible side chains) using Autodock Vina 1.1.2 [45]. The resultant pMHC structure was then refined through a full atom energy minimization step with Gromacs 4.5.1 package [44] and a new docking search was performed keeping flexible only the epitope side chains. This automated approach for pMHC structure prediction was largely validated against crystal structures and is being currently prepared for publication as a web-server (Rigo MM, personal communication).

Electrostatic potential calculation and image analysis

Electrostatic potential over the TCR-interacting surface of pMHCs (for both crystals and models) were calculated using Delphi [46], through the molecular viewer software GRASP2 [47]. Automated scripts were used to prepare the structures for this analysis, allowing all pMHC to be observed in the same fixed position. Images of the TCR-interacting surfaces were saved and imported to ImageJ 1.46r software (National Institute of Health, USA, http://rsb.info.nih1.43u.gov/ij). An ImageJ plugin was developed by our team to import RGB values from predetermined regions over the pMHC surface. This regions were selected based on spots of variation over the "TCR interacting surface" of pMHCs, as previously described [26]. Values were exported as "csv" tables for further analysis.

Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) was performed with Pvclust [32], an R package for assessing the uncertainty in hierarchical clustering. The option average was selected as linkage method and option correlation was used as distance method. The number of bootstrap replications was set to 10000. Results were plotted as dendrograms with Bootstrap Probabilities (BP) and Approximately Unbiased (AU) *p*-values. BP values are calculated by normal bootstrap resampling and AU values are

computed through multiscale bootstrap resampling, which is referred as a better approximation to unbiased *p*-value [32]. Standard errors for AU *p*-values were obtained with *seplot*, presenting values lower than 0.01 for all HCAs performed.

Acknowledgments

We thank the *Centro Nacional de Supercomputação* (CESUP-RS) for allowing access to its computational resources. We also thank Dr. Zu Ting Shen and Dr. Lawrence Stern, from the University of Massachusetts Medical School (Worcester, MA/USA), for sharing the crystal structures of VV-$A11_{198}$:H2-K^b (3TIE) and LCMV-GP_{34}-V8M:H2-K^b (3TID) before publication.

References

1. Welsh RM, Selin LK, Szomolanyi-Tsuda E (2004) Immunological memory to viral infections. Annu Rev Immunol 22: 711-743.
2. Ressing ME, Luteijn RD, Horst D, Wiertz EJ (2013) Viral interference with antigen presentation: trapping TAP. Mol Immunol 55: 139-142.
3. Salazar MI, Del Angel RM, Lanz-Mendoza H, Ludert JE, Pando-Robles V (2014) The role of cell proteins in dengue virus infection. J Proteomics: [Epub ahead of print].
4. Schmid M, Speiseder T, Dobner T, Gonzalez RA (2014) DNA virus replication compartments. J Virol 88: 1404-1420.
5. Lauring AS, Frydman J, Andino R (2013) The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol 11: 327-336.
6. Vandiedonck C, Knight JC (2009) The human Major Histocompatibility Complex as a paradigm in genomics research. Brief Funct Genomic Proteomic 8: 379-394.
7. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, et al. (2010) Antagonistic coevolution accelerates molecular evolution. Nature 464: 275-278.
8. Kubinak JL, Ruff JS, Hyzer CW, Slev PR, 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.
9. Welsh RM, Selin LK, Szomolanyi-Tsuda E (2004) Immunological memory to viral infections. Annu Rev Immunol 22: 711-743.
10. Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. Front Immunol 4: 485.
11. Welsh RM, Selin LK (2002) No one is naive: the significance of heterologous T-cell immunity. Nat Rev Immunol 2: 417-426.
12. Wucherpfennig KW, Allen PM, Celada F, Cohen IR, De Boer R, et al. (2007) Polyspecificity of T cell and B cell receptor recognition. Semin Immunol 19: 216-224.
13. Welsh RM, Che JW, Brehm Ma, Selin LK (2010) Heterologous immunity between viruses. Immunol Rev 235: 244-266.
14. Vieira GF, 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.
15. Welsh RM, Fujinami RS (2007) Pathogenic epitopes, heterologous immunity and vaccine design. Nat Rev Microbiol 5: 555-563.
16. Selin LK, Cornberg M, Brehm MA, Kim SK, Calcagno C, et al. (2004) CD8 memory T cells: cross-reactivity and heterologous immunity. Semin Immunol 16: 335-347.
17. Wlodarczyk MF, Kraft AR, Chen HD, Kenney LL, 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.
18. Cornberg M, Kenney LL, Chen AT, Waggoner SN, Kim SK, et al. (2013) Clonal exhaustion as a mechanism to protect against severe immunopathology and death from an overwhelming CD8 T cell response. Front Immunol 4: 475.
19. Cornberg M, Chen AT, Wilkinson LA, Brehm MA, Kim SK, et al. (2006) Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. J Clin Invest 116: 1443-1456.
20. Cornberg M, Clute SC, Watkin LB, Saccoccio FM, Kim S-k, et al. (2010) CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections. J Immunol 184: 2825-2838.
21. Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, et al. (2014) Deconstructing the peptide-MHC specificity of T cell recognition. Cell 157: 1073-1087.
22. Yin Y, Li Y, Mariuzza RA (2012) Structural basis for self-recognition by autoimmune T-cell receptors. Immunol Rev 250: 32-48.
23. Shen ZT, Nguyen TT, Daniels KA, Welsh RM, 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.
24. Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M, et al. (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.

25. Sinigaglia M, Antunes DA, Rigo MM, Chies JA, Vieira GF (2013) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. Database (Oxford) 2013: bat002.
26. Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, et al. (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.
27. Selin LK, Wlodarczyk MF, Kraft AR, Nie S, Kenney LL, et al. (2011) Heterologous immunity: immunopathology, autoimmunity and protection during viral infections. Autoimmunity 44: 328-347.
28. Petrova GV, Naumova EN, Gorski J (2011) The polyclonal CD8 T cell response to influenza M158-66 generates a fully connected network of cross-reactive clonotypes to structurally related peptides: a paradigm for memory repertoire coverage of novel epitopes or escape mutants. J Immunol 186: 6390-6397.
29. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, et al. (2013) The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. Hum Vaccin Immunother 9: 1577-1586.
30. Chen AT, Cornberg M, Gras S, Guillonneau C, Rossjohn J, et al. (2012) Loss of anti-viral immunity by infection with a virus encoding a cross-reactive pathogenic epitope. PLoS Pathog 8: e1002633.
31. UniProt-Consortium (2014) Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 42: D191-198.
32. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22: 1540-1542.
33. Fytili P, Dalekos GN, Schlaphoff V, Suneetha PV, Sarrazin C, et al. (2008) Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. Vaccine 26: 3818-3826.
34. Turner SJ, Kedzierska K, Komodromou H, La Gruta NL, Dunstone MA, 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.
35. Turner SJ, Doherty PC, McCluskey J, Rossjohn J (2006) Structural determinants of T-cell receptor bias in immunity. Nat Rev Immunol 6: 883-894.
36. Garcia KC, Adams JJ, Feng D, Ely LK (2009) The molecular basis of TCR germline bias for MHC is surprisingly simple. Nat Immunol 10: 143-147.
37. Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, et al. (2011) T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. Immunity 35: 681-693.
38. Gras S, Burrows SR, Turner SJ, Sewell AK, McCluskey J, et al. (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.
39. Santori FR, Holmberg K, Ostrov D, Gascoigne NR, Vukmanovic S (2004) Distinct footprints of TCR engagement with highly homologous ligands. J Immunol 172: 7466-7475.
40. Mazza C, Auphan-Anezin N, Gregoire C, Guimezanes A, Kellenberger C, et al. (2007) How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? EMBO J 26: 1972-1983.
41. Kasprowicz V, Ward SM, Turner A, Grammatikos A, Nolan BE, 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.
42. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.
43. DeLano WL, Bromberg S (2004) PyMOL User's Guide. San Francisco: DeLano Scientific LLC
44. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29: 845-854.
45. Trott O, Olson AJ, 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.
46. Li L, Li C, Sarkar S, Zhang J, Witham S, et al. (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys 5: 9.
47. Petrey D, Honig B (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. Methods Enzymol 374: 492-509.

Figures

Figure 1. Cross-reactivity networks (CRNs) compiled from previous publications. Arrows indicate the directionality of reactions observed experimentally, with colors indicating stronger (black) or weaker (grey) responses. Segmented connectors indicate noncross-reactive targets. Each ellipse indicates one epitope in the context of (A) murine H2-K^b, (B) human HLA-A*02:01 or (C) murine H2-D^b MHC allotypes. Each ellipse contains the epitope sequence, abbreviation and PDB code (if available). Ellipses colors indicate the source of the information on cross-reactivity. Most data was compiled from Cornberg *et al.* 2010 (orange and red ellipses) and expanded with data from Shen *et al.* 2013 (cyan) and Fytili *et al.* 2008 (yellow). Grey ellipses indicate data from Wlodarczyk *et al.* 2013 and green ellipses indicate targets included based on sequence/structural analysis (see Methods). The symbol # was used to indicate reactions suggested by sequence/structural analysis, which were not yet tested *in vitro*/*in vivo*. Purple areas indicate complexes with greater structural similarity according to a hierarchical clustering analysis.

Figure 2. Extended H2-K^b-restricted HCA. Structure-based hierarchical cluster analysis (HCA) performed with *pvclust*. Each putative cluster is represented by a specific *edge* (grey numbers), in order of increasing *Heights* (y axis). Cluster confidence is measured with two *p*-values, Bootstrap Probabilities (BP) and Approximately Unbiased (AU). Lines highlighted in purple indicate structures with greater structural similarity (as represented in Figure 1). Lines highlighted in blue and pink indicate putative cross-reactivity thresholds for different memory T cells (see Results/Discussion). Each target is colored according to Figure 1. Epitope abbreviation and sequence are provided, with red amino acids indicating changes in relation to VV-$A11_{198}$. (*) Crystal structure 3TID was used to represent LCMV-GP_{34}, despite presenting a C8M exchange, as indicated by its sequence (see Results/Discussion).

Figure 3. TCR-interacting surfaces of selected pMHC complexes. A. Four HLA-A*02:01-restricted complexes. B. Two H2-K^b-restricted pMHCs. Regions with positive (blue) and negative (red) charges are represented with a scale from -5 kT to +5 kT. Abbreviation of the specific "peptide:MHC" depicted is provided below each complex. Some "unique" features in terms of topography or electrostatic potential are indicated with green arrows.

Supplementary Figures

Figure S1. Crystal-based H2-K^b-restricted HCA. Structure-based hierarchical cluster analysis (HCA) performed with *pvclust*. Each putative cluster is represented by a specific *edge* (grey numbers), in order of increasing *Heights* (y axis). Cluster confidence is measured with two *p*-values, Bootstrap Probabilities (BP) and Approximately Unbiased (AU). Lines highlighted in purple indicate structures with greater structural similarity (as represented in Figure 1). Epitope abbreviation and the respective PDB code for each crystal structure (in blue) are provided. (*) Crystal structure 3TID was used to represent LCMV-GP_{34}, despite presenting a C8M exchange (see Results/Discussion).

Figure S2. Extended HLA-A*02:01-restricted HCA. Structure-based hierarchical cluster analysis (HCA) performed with *pvclust*. Each putative cluster is represented by a specific *edge* (grey numbers), in order of increasing *Heights* (y axis). Cluster confidence is measured with two *p*-values, Bootstrap Probabilities (BP) and Approximately Unbiased (AU). Abbreviation of crystal structures includes their PDB code (in blue), while the termination "Mod" indicates modeled structures. Lines highlighted in purple indicate structures with greater structural similarity (as represented in Figure 1).

Figure S3. TCR-interacting surfaces of predicted cross-reactive targets. Regions with positive (blue) and negative (red) charges are represented with a scale from -5 kT to +5 kT. Abbreviation of the specific "peptide:MHC" depicted is provided below each complex, as well as epitope sequence. Amino acids depicted in red (VV-$E7_{130}$) indicate exchanges in relation to LCMV-GP_{118}. Great structural similarity is observed between these complexes, both in terms of topography and electrostatic potential over the TCR-interacting surface.

Figure S4. Crystal-based H2-D^b-restricted HCA. Structure-based hierarchical cluster analysis (HCA) performed with *pvclust*. Each putative cluster is represented by a specific *edge* (grey numbers), in order of increasing *Heights* (y axis). Cluster confidence is measured with two *p*-values, Bootstrap Probabilities (BP) and Approximately Unbiased (AU). Epitope abbreviation and the respective PDB code for each crystal structure (in blue) are provided. Lines highlighted in purple indicate structures with greater structural similarity (as represented in Figure 1).

Figure 1.

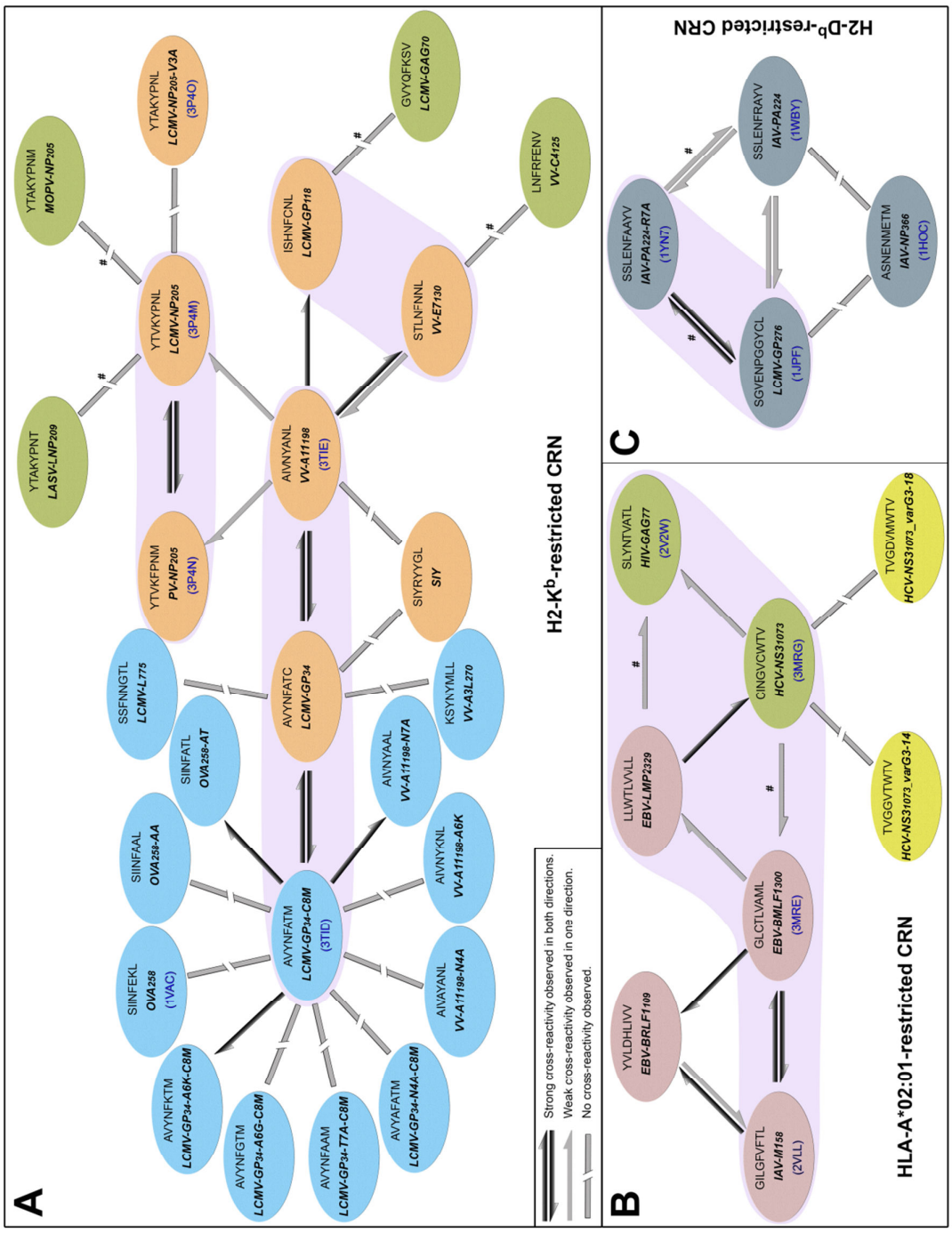

A
H2-K^b-restricted CRN
YTAKYPNT LASV-LNP209
YTAKYPNM MOPV-NP205
YTVKYPNL LCMV-NP205 (3P4M)
YTAKYPNL LCMV-NP205-V3A (3P4O)
YTVKFPNM PV-NP205 (3P4N)
AIVNYANL VV-A11198 (3TIE)
ISHNFCNL LCMV-GP118
GVYQFKSV LCMV-GAG70
STLNFNNL VV-E7130
LNFRFENV VV-C4125
SIYRYYGL SIY
SSFNNGTL LCMV-L775
SIINFAAL OVA258-AA
SIINFEKL OVA258 (1VAC)
SIINFATL OVA258-AT
AVYNFATC LCMV-GP34
AVYNFATM LCMV-GP34-C8M (3TID)
AVYNFKTM LCMV-GP34-A6K-C8M
AVYNFGTM LCMV-GP34-A6G-C8M
AVYNFAAM LCMV-GP34-T7A-C8M
AVYAFATM LCMV-GP34-N4A-C8M
AIVNYAAL VV-A11198-N7A
AIVNYKNL VV-A11198-A6K
AIVAYANL VV-A11198-N4A
KSYNYMLL VV-A3L270
Strong cross-reactivity observed in both directions.
Weak cross-reactivity observed in one direction.
No cross-reactivity observed.
B
HLA-A*02:01-restricted CRN
YVLDHLIVV EBV-BRLF1109
LLWTLVVLL EBV-LMP2329
SLYNTVATL HIV-GAG77 (2V2W)
GILGFVFTL IAV-M158 (2VLL)
GLCTLVAML EBV-BMLF1300 (3MRE)
CINGVCWTV HCV-NS31073 (3MRG)
TVGGVTWTV HCV-NS31073_varG3-14
TVGDVMWTV HCV-NS31073_varG3-18
C
H2-D^b-restricted CRN
SSLENFAAYV IAV-PA224-R7A (1YNT)
SSLENFRAYV IAV-PA224 (1WBY)
SGVENPGGYCL LCMV-GP276 (1JPF)
ASNENMETM IAV-NP366 (1HQC)

Figure 2.

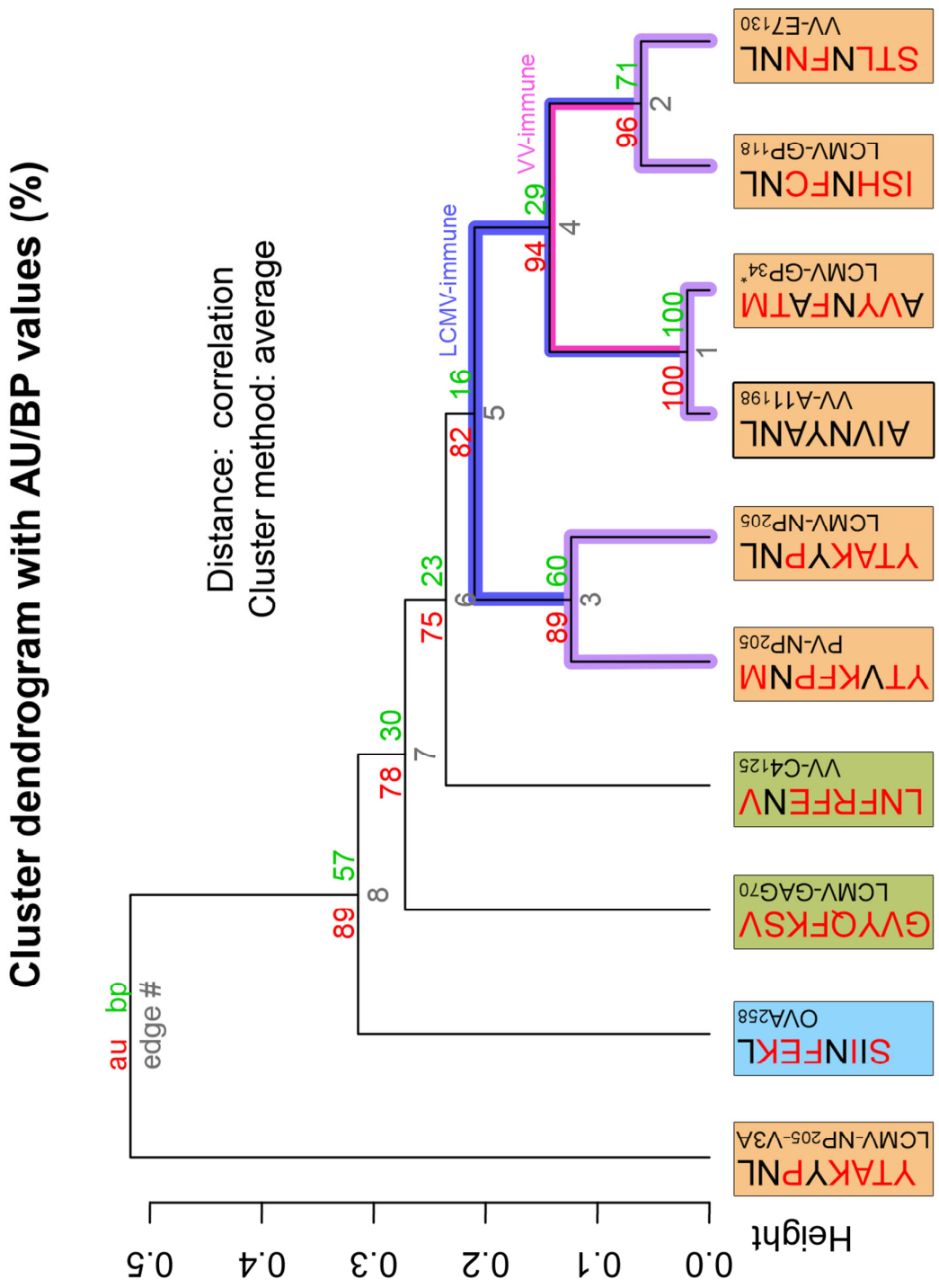

Cluster dendrogram with AU/BP values (%)
Distance: correlation
Cluster method: average
au bp
edge #
Height
0.0
0.1
0.2
0.3
0.4
0.5
YTAKYPNL
LCMV-NP205-V3A
SIINFEKL
OVA258
GVYQFKSV
LCMV-GAG70
LNFRFENV
VV-C4125
YTVKFPNM
PV-NP205
YTAKYPNL
LCMV-NP205
AIVNYANL
VV-A11198
AVYNFATM
LCMV-GP34*
ISHNFCNL
LCMV-GP118
STLNFNNL
VV-E7130
LCMV-immune
VV-immune

Figure 3.

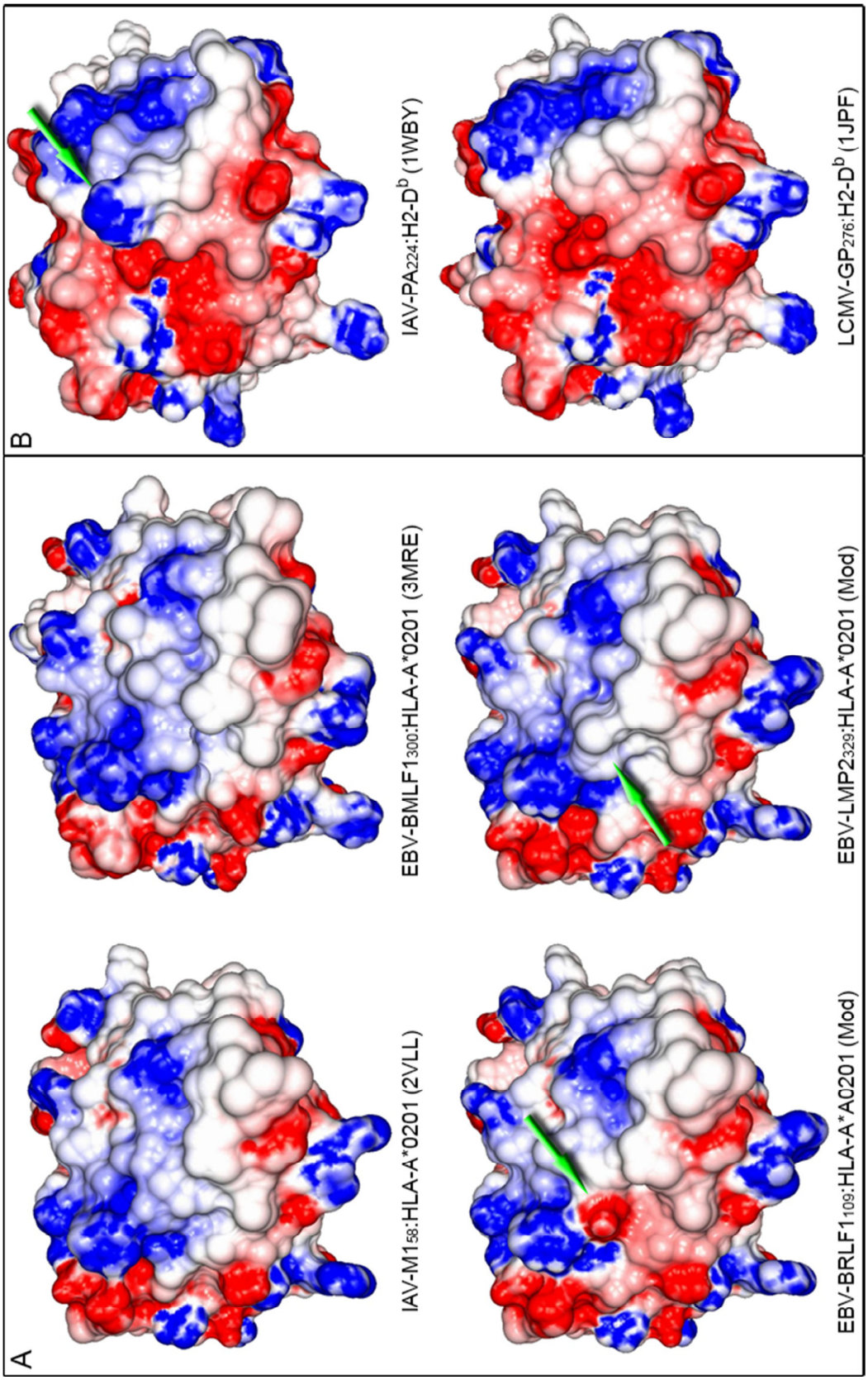

A
IAV-M1$_{58}$:HLA-A*0201 (2VLL)
EBV-BMLF1$_{300}$:HLA-A*0201 (3MRE)
EBV-BRLF1$_{109}$:HLA-A*A0201 (Mod)
EBV-LMP2$_{329}$:HLA-A*0201 (Mod)
B
IAV-PA$_{224}$:H2-D^{b} (1WBY)
LCMV-GP$_{276}$:H2-D^{b} (1JPF)

Figure S1.

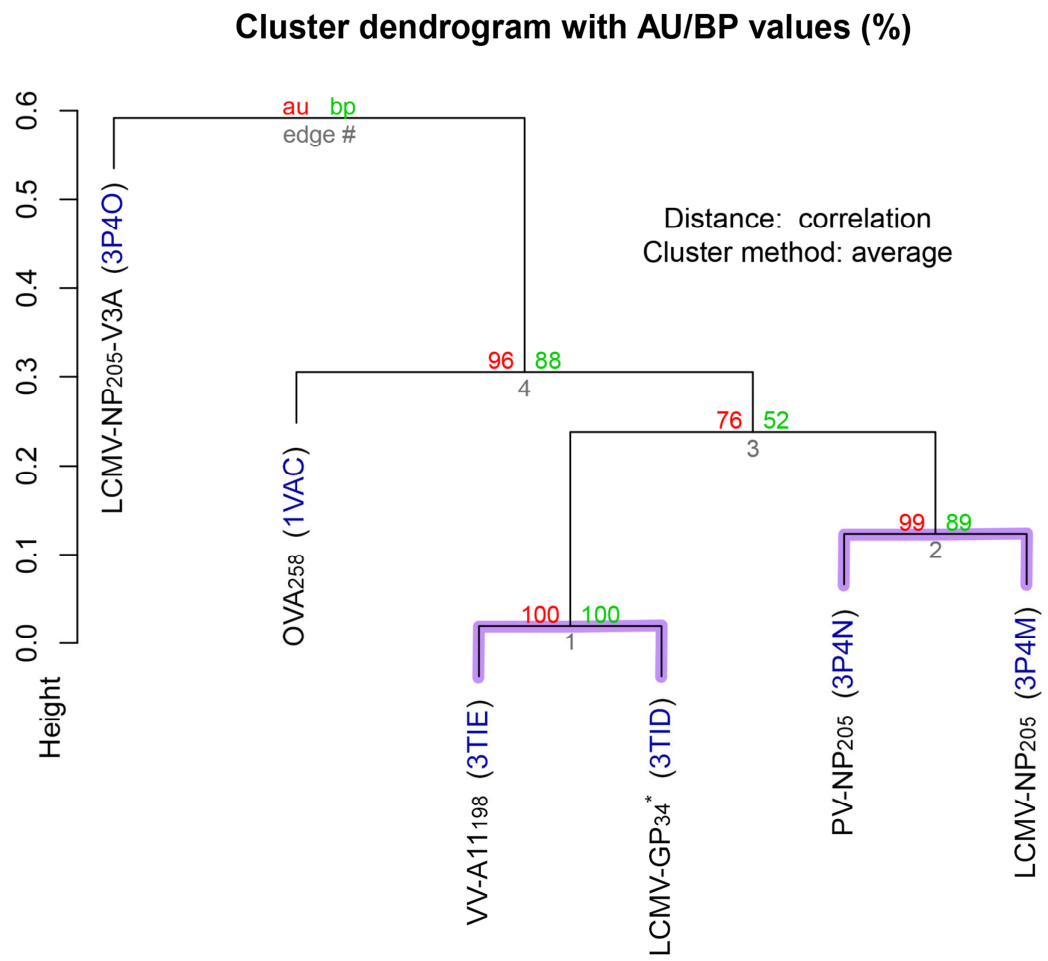

Cluster dendrogram with AU/BP values (%)
au bp
edge #
Distance: correlation
Cluster method: average
Height
0.0
0.1
0.2
0.3
0.4
0.5
0.6
96 88
4
76 52
3
99 89
2
100 100
1
LCMV-NP205-V3A (3P4O)
OVA258 (1VAC)
VV-A11198 (3TIE)
LCMV-GP34* (3TID)
PV-NP205 (3P4N)
LCMV-NP205 (3P4M)

Figure S2.

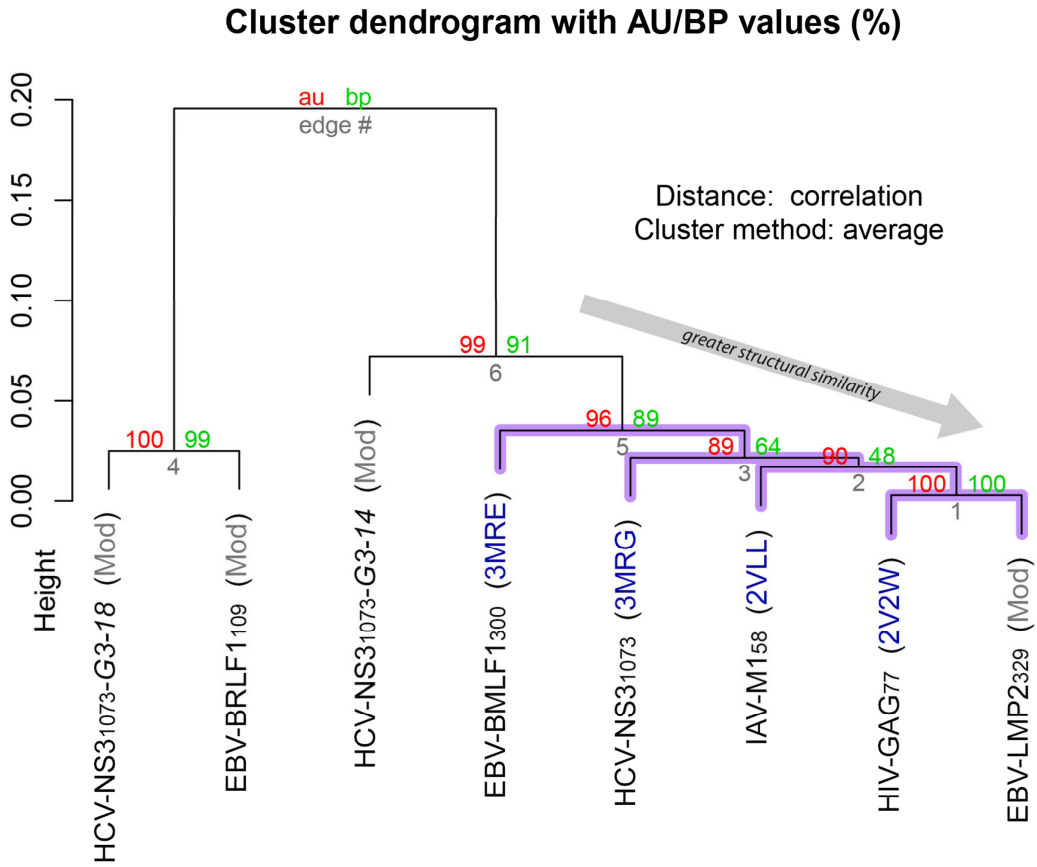

Cluster dendrogram with AU/BP values (%)
au bp
edge #
Distance: correlation
Cluster method: average
greater structural similarity
Height
0.00
0.05
0.10
0.15
0.20
100 99
4
99 91
6
96 89
5
89 64
3
90 48
2
100 100
1
HCV-NS31073-G3-18 (Mod)
EBV-BRLF1109 (Mod)
HCV-NS31073-G3-14 (Mod)
EBV-BMLF1300 (3MRE)
HCV-NS31073 (3MRG)
IAV-M158 (2VLL)
HIV-GAG77 (2V2W)
EBV-LMP2329 (Mod)

Figure S3.

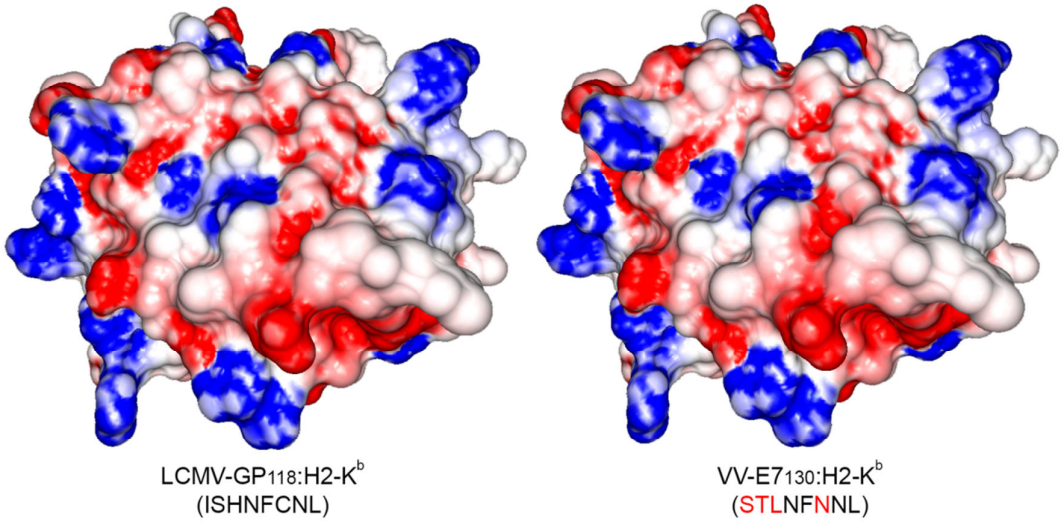


Figure S4.

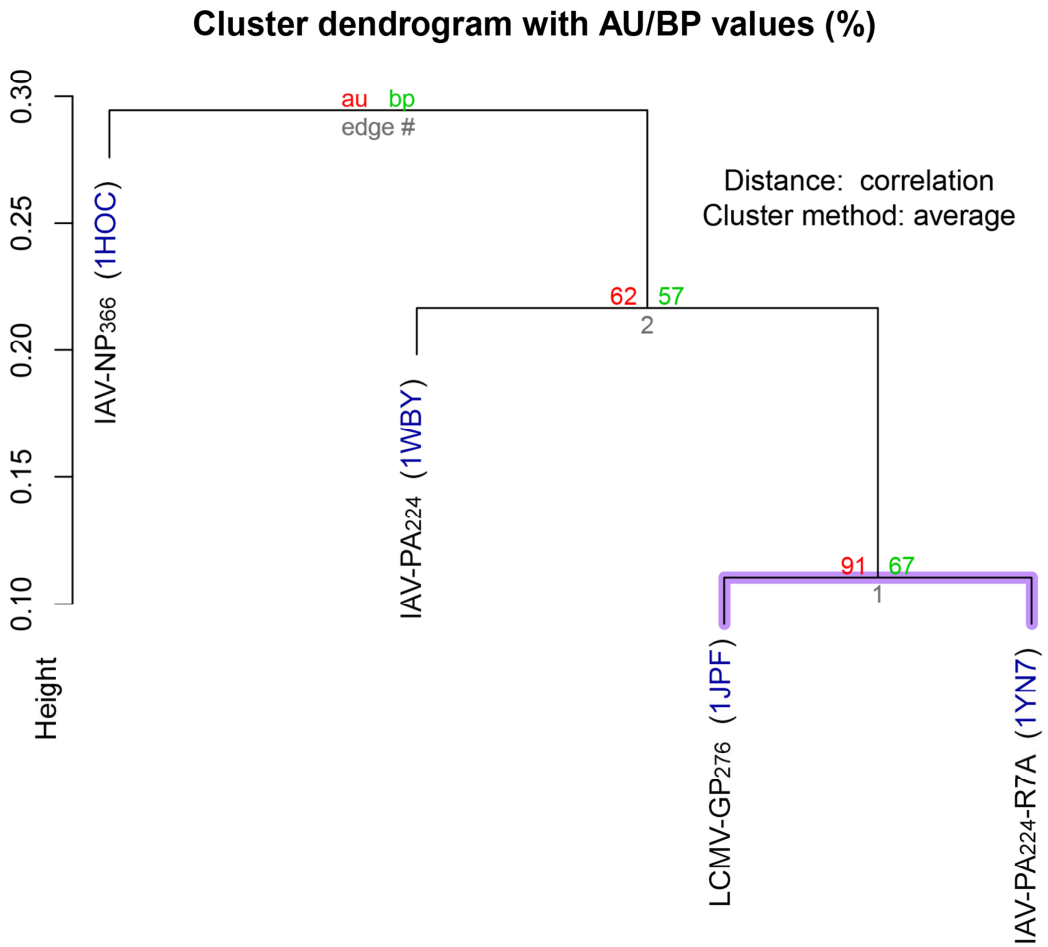


Table S1. Epitope additional Information.

Abbreviation	Aliases	MHC Allele	Virus	Protein	Protein ID (Uniprot/NCBI)	Epitope Position	Sequence	Epitope ID (IEDB)	Complex ID (PDB)	Reference (PubMed ID)
EBV-$BMLF1_{300}$	$BMLF1_{259}$, $BMLF1_{280}$, $BMLF1_{300}$	HLA-A*02:01	Epstein-Barr vírus/Human herpesvirus 4	mRNA export factor ICP27 homolog	Uniprot: Q04360	300-308	GLCTLVAML	20788	3MRE	20164414
EBV-$BRLF1_{109}$	$BRLF1_{190}$	HLA-A*02:01	Epstein-Barr vírus/Human herpesvirus 4	Transcription activator BRLF1	Uniprot: P03209	109-117	YVLDHLIVV	76333	-	20164414
EBV-$LMP2_{329}$	-	HLA-A*02:01	Epstein-Barr vírus/Human herpesvirus 4	Latent membrane protein 2	Uniprot: P13285	329-337	LLWTLVVLL	37960	-	20164414
HCV-$NS3_{1073}$	-	HLA-A*02:01	Hepatitis C virus	Genome polyprotein (Serine protease NS3)	Uniprot: P26664	1073-1081	CINGVCWTV	6435	3MRG	18582999
HCV-$NS3_{1073}$_varG3-14	-	HLA-A*02:01	Hepatitis C virus (genotype 3)	Genome polyprotein (Serine protease NS3)	GenBank: AAC03058.1	1079-1087	TVGGVTWTV	95938	-	18582999
HCV-$NS3_{1073}$_varG3-18	-	HLA-A*02:01	Hepatitis C virus (genotype 3)	Genome polyprotein (Serine protease NS3)	-	-	TVGDVMWTV	95935	-	18582999
HIV-GAG_{77}	-	HLA-A*02:01	Human immunodeficiency virus 1	gag polyprotein	Uniprot: P05889	77-85	SLYNTVATL	59613	2V2W	14527342
IAV-$M1_{58}$	-	HLA-A*02:01	Influenza A virus	Matrix protein 1	Uniprot: P35937	58-66	GILGFVFTL	20354	2VLL	14527342/ 20164414
LASV-LNP_{209}	-	H2-K^b	Lassa virus	Nucleoprotein	Uniprot: P13699	209-216	YTAKYPNT	-	-	16614754
LCMV-GAG_{70}	-	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	70-77	GVYQFKSV	23229	-	-

Table S1. (contin.)

Abbreviation	Aliases	MHC Allele	Virus	Protein	Protein ID (Uniprot/NCBI)	Epitope Position	Sequence	Epitope ID (IEDB)	Complex ID (PDB)	Reference (PubMed ID)
LCMV-GP_{118}	-	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	118-125	ISHNFCNL	28528	-	20164414
LCMV-GP_{34}	-	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYNFATC	5625	-	20164414
LCMV-GP_{34}-A6G-C8M	gp34 (P6G)	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYNFGTM	194909	-	24127554
LCMV-GP_{34}-A6K-C8M	gp34 (P6K)	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYNFKTM	194910	-	24127554
LCMV-GP_{34}-C9M	-	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYNFATM	5628	3TID	24127554
LCMV-GP_{34}-N4A-C8M	gp34 (P4A)	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYAFATM	194907	-	24127554
LCMV-GP_{34}-T7A-C8M	-	H2-K^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	34-41	AVYNFAAM	194908	-	24127554
LCMV-L_{775}	-	H2-K^b	Lymphocytic choriomeningitis virus	RNA-directed RNA polymerase L	Uniprot: P14240	775-782	SSFNNGTL	60998	-	24127554
LCMV-NP_{205}	-	H2-K^b	Lymphocytic choriomeningitis virus	Nucleoprotein	Uniprot: P09992	205-212	YTVKYPNL	76205	3P4M	20164414/ 12055626
LCMV-NP_{205}-V3A	-	H2-K^b	Lymphocytic choriomeningitis virus	Nucleoprotein	Uniprot: P09992	205-212	YTAKYPNL	75945	3P4O	16614754/ 22536152

Table S1. (contin.)

Abbreviation	Aliases	MHC Allele	Virus	Protein	Protein ID (Uniprot/NCBI)	Epitope Position	Sequence	Epitope ID (IEDB)	Complex ID (PDB)	Reference (PubMed ID)
MOPV-NP_{205}	-	H2-K^b	Mopeia virus	-	-	-	YTAKYPNM	-	-	16614754
OVA_{258}	OVA_{257}	H2-K^b	Gallus gallus	ovalbumin	Uniprot: P01012	258-265	SIINFEKL	58560	3P9L	9469429
OVA_{257}-E7A-K8A	-	H2-K^b	-	-	-	-	SIINFAAL	-	-	24127554
OVA_{257}-E7A-K8T	-	H2-K^b	-	-	-	-	SIINFATL	-	-	24127554
PV-NP_{205}	-	H2-K^b	Pichinde arenavirus	Nucleoprotein	Uniprot: P03541	205-212	YTVKFPNM	76204	3P4N	20164414/ 12055626
SIY	-	H2-K^b	-	-	-	synthetic construct	SIYRYYGL	58773	-	22233579
VV-$A11_{198}$	$A11R_{198}$	H2-K^b	Vaccinia virus	Protein A11	Uniprot: P20988	198-205	AIVNYANL	2124	3TIE	20164414/ 24127554
VV-$A11_{198}$-A6K	$A11R_{198}$ (P6K)	H2-K^b	Vaccinia virus	Protein A11	Uniprot: P20988	198-205	AIVNYKNL	194905	-	24127554
VV-$A11_{198}$-N4A	$A11R_{198}$ (P4A)	H2-K^b	Vaccinia virus	Protein A11	Uniprot: P20988	198-205	AIVAYANL	194902	-	24127554
VV-$A11_{198}$-N7A	$A11R_{198}$ (P7A)	H2-K^b	Vaccinia virus	Protein A11	Uniprot: P20988	198-205	AIVNYAAL	194903	-	24127554
VV-$A3L_{270}$	-	H2-K^b	Vaccinia virus	Major core protein 4b	Uniprot: P20643	270-277	KSYNYMLL	33586	-	24127554

Table S1. (contin.)

Abbreviation	Aliases	MHC Allele	Virus	Protein	Protein ID (Uniprot/NCBI)	Epitope Position	Sequence	Epitope ID (IEDB)	Complex ID (PDB)	Reference (PubMed ID)
VV-$C4_{125}$	-	H2-K^b	Vaccinia virus	Protein C4	Uniprot: P17370	125-132	LNFRFENV	-	-	-
VV-$E7_{130}$	$E7R_{130}$	H2-K^b	Vaccinia virus	Protein E7	Uniprot: P68446	130-137	STLNFNNL	61731	-	20164414
LCMV-GP_{276}	-	H2-D^b	Lymphocytic choriomeningitis virus	Pre-glycoprotein polyprotein GP complex	Uniprot: P09991	276-286	SGVENPGGYCL	58282	1JPF	23408839
IAV-PA_{224}	-	H2-D^b	Influenza A virus	Polymerase acidic protein	Uniprot: Q809J3	224-233	SSLENFRAYV	61151	1WBY	23408839
IAV-PA_{224}-R7A	-	H2-D^b	Influenza A virus	Polymerase acidic protein	-	-	SSLENFAAYV	-	1YN7	15735650
IAV-NP_{366}	-	H2-D^b	Influenza A virus	Nucleoprotein	Uniprot: B4URE0	366-374	ASNENMETM	4602	1HOC	23408839

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 incialmente 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; Plewczynski *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.

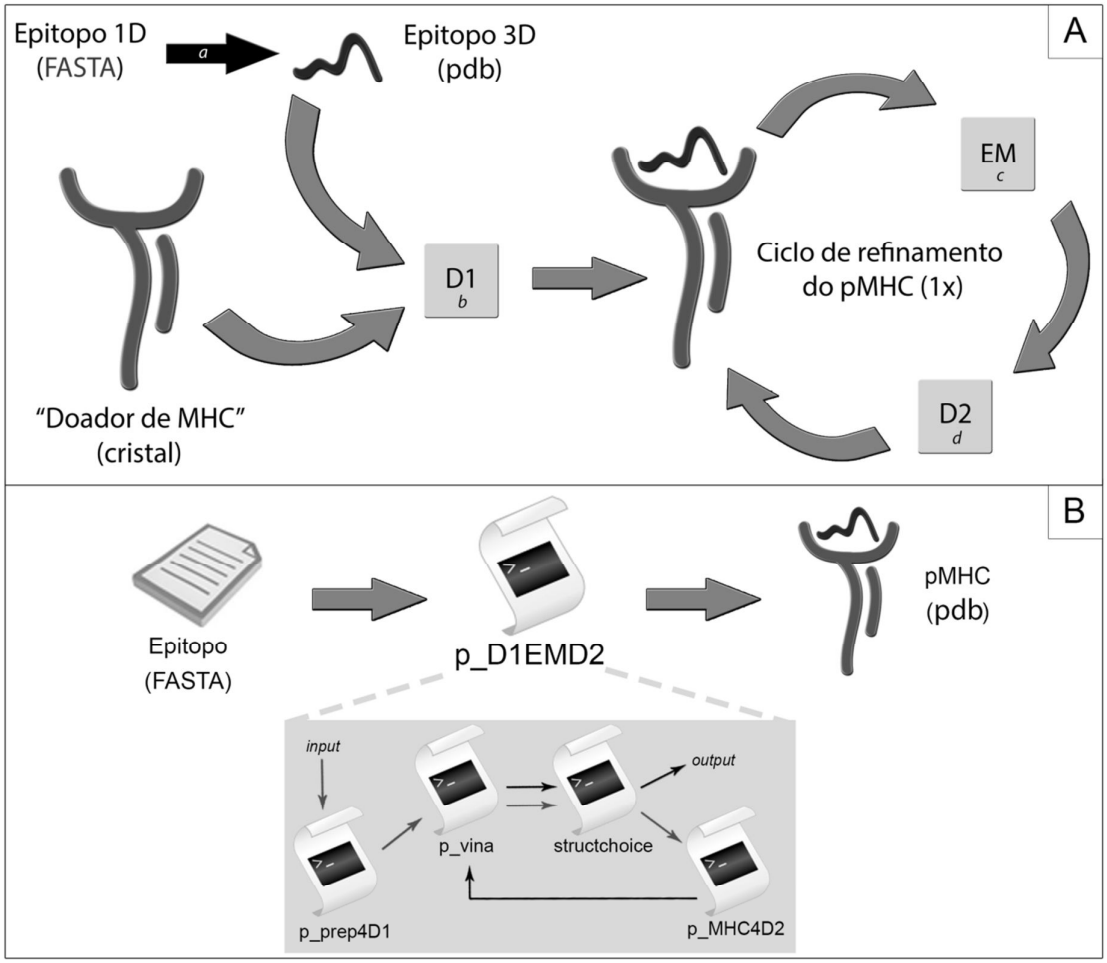


Figura 6. Abordagem *D1-EM-D2.* A. A estretégia consiste em 4 etapas principais, nas quais são utilizados os programas PyMOL (*a*), Autodock Vina (*b,d*) e o pacote de programas Gromacs (*c*). Um cristal de referência para o alotipo de interesse, previamente preparado, é utilizado como "Doador de MHC". 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 envolvento 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 sequencia 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 sequencias 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://mgltools.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 epitopo 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 epitopos (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 epitopo 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 (± 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 (cα)		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; cα, 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 usa 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).

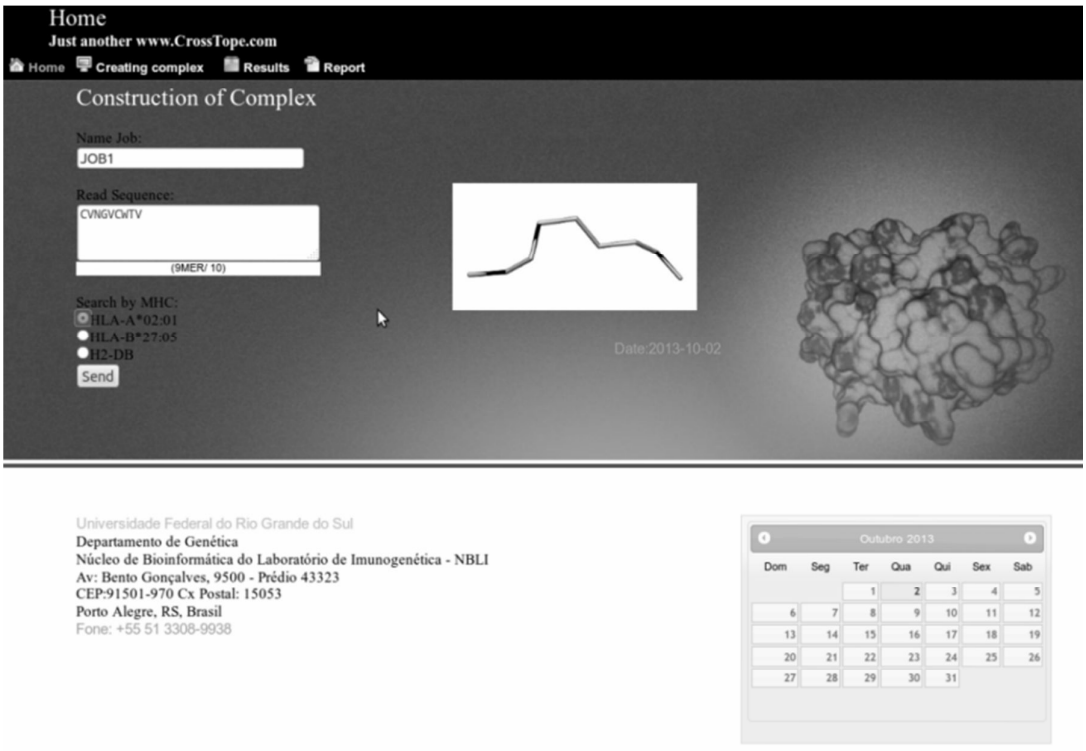


Figura 7. Interface *web* para a construção automática de complexos pMHC. Após informar a sequência e selecionar o MHC, o usuário clica em enviar e aguarda o resultado do processamento, o qual será realizado em um *cluster* de alto desempenho. Esta interface está em fase de testes e ainda não foi disponibilizada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

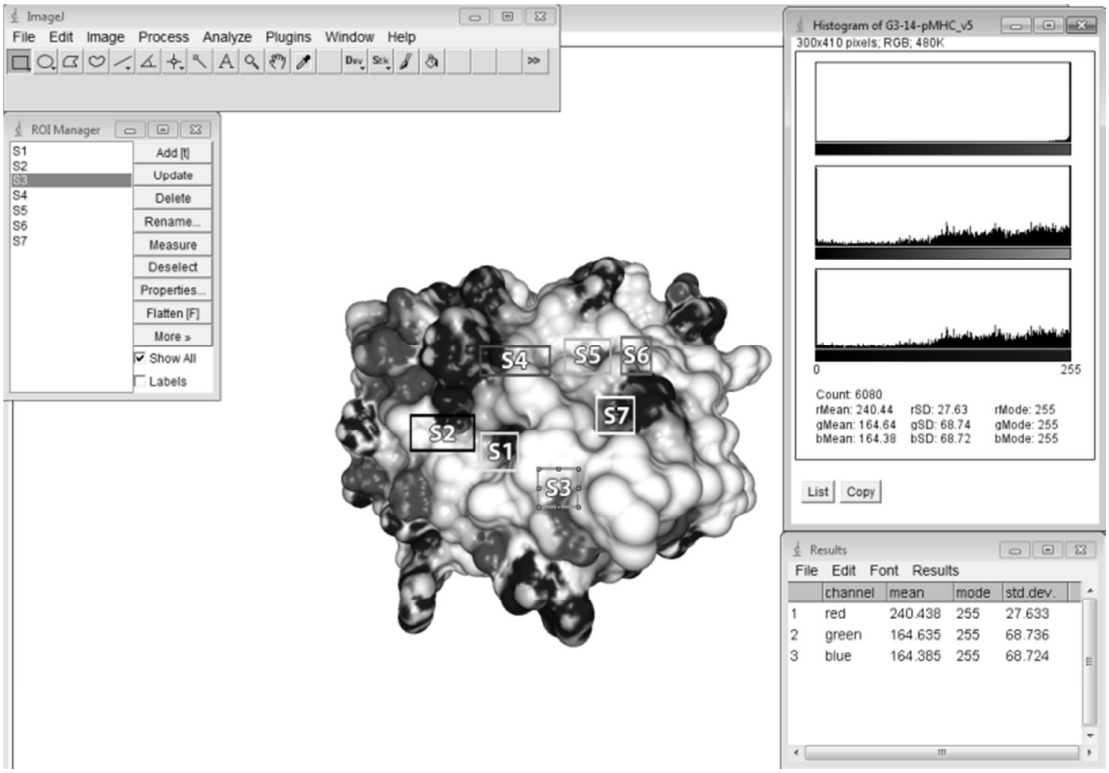


Figura 8. Utilização de um *plugin* para importação de valores RGB no programa ImageJ. Ao centro é possível observar a superfície de um complexo pMHC, sobre a qual estão identificadas as sete regiões de interesse previamente determinadas. A posição destas regiões foi salva, permitindo a seleção independente 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 $NS3_{1073}$ (C**V**/**I**NGVCWTV) (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-$NS3_{1073}$ 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-NA_{231} (CVNGSCFTV), que compartilhava 77% (7/9) de sua sequência linear de aminoácidos com o HCV-$NS3_{1073}$, 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-NA_{231} (CVNGSCFTV) o *primer* para esta reatividade cruzada. Corroborando a hipótese dos autores, 44% (4/9) dos controles HCV- que respondiam para HCV-$NS3_{1073}$ também reconheceram o alvo IAV-NA_{231} 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-NA_{231} 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-NA_{231} sugere o possível envolvimento de reatividade cruzada com outros alvos ainda desconhecidos.

Posteriormente, um trabalho publicado por Kasprowicz e colaboradores sugeriu que esta reatividade cruzada entre *Influenza* e HCV era relativamente fraca e apresentava uma direcionalidade preferencial no sentido HCV → IAV (Kasprowicz *et al.*, 2008). Eles encontraram poucas evidências de resposta contra IAV-NA_{231} em controles saudáveis (HCV-), mas descreveram a geração de células específicas para IAV-NA_{231} 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-$NS3_{1073}$, cobrindo os seis genótipos de HCV (Fytili *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-$NS3_{1073}$ (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-NA_{231}. Como pode ser observado na Figura 9, o PCA posicionou este

possível alvo de reatividade cruzada exatamente ao lado do epitopo selvagem de HCV-$NS3_{1073}$ (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-$NS3_{1073}$, 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).

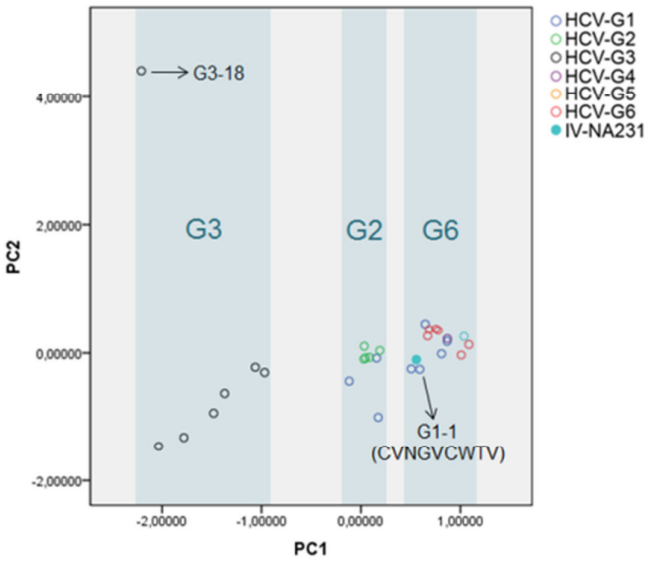


Figura 9. Análise de componentes principais. O único alvo sem resposta detectável *in vitro* (G3-18) ficou isolado, enquanto o alvo selvagem (G1-1) agrupou com alvos que apresentaram maior produção de IFN-gamma. De acordo com o PC1, foi possível observar a distribuição de faixas que incluíam todos os 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-$NS3_{1073}$ (Antunes *et al.*, 2011). No artigo que apresentou os dados, foi salientada a similaridade estrutural com os alvos EBV-$LMP2_{329}$ e HIV-GAG_{77},

cuja reatividade cruzada com HCV-$NS3_{1073}$ foi posteriormente confirmada (capítulo III). Mas o HCA de 2011 também incluía no mesmo *cluster* os alvos IAV-$M1_{58}$, CMV-$pp65_{495}$ e EBV-$GP85_{420}$. Os dois primeiros foram recentemente testados *in vitro* frente a diferentes populações de célula T específicas para HCV-$NS3_{1073}$ (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-NA_{231} também foi incluído neste HCA com 55 complexos. No entanto, contrariando o resultado do PCA, ele não foi agrupado com o HCV-$NS3_{1073}$. 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-NA_{231} e HCV-$NS3_{1073}$ (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).

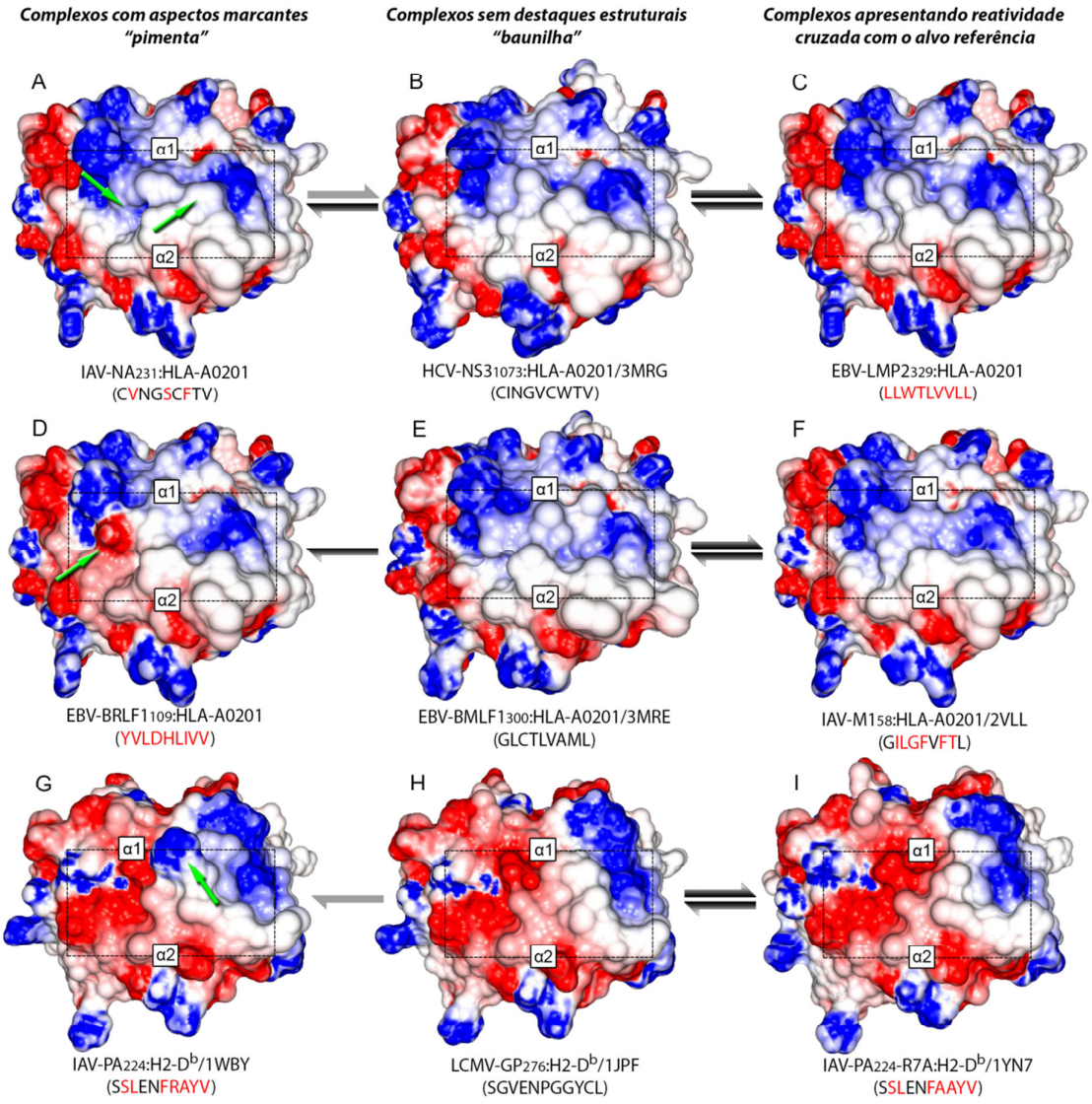


Figura 10. Comparação entre superfícies de complexos envolvidos em reatividade cruzada. Imagens da superfície de contato com o TCR, para cada pMHC, foram obtidas com o programa GRASP2. Os domínios "alfa 1" e "alfa 2" da cadeia pesada do MHC são indicados para cada complexo, bem como a região ocupada 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-NA_{231}:HLA-A*0201 e HCV-$NS3_{1073}$: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-$BRLF1_{109}$ e EBV-$BMLF1_{300}$, 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-$BMLF1_{300}$ → EBV-$BRLF1_{109}$ (Selin LK, comunicação pessoal). Nas nossas análises, um HCA incluindo alvos que apresentavam reatividade cruzada com EBV-$BMLF1_{300}$ excluiu o EBV-$BRLF1_{109}$ 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-PA_{224}:H2-D^b e IAV-NP_{366}: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-PA_{224}: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-PA_{224}-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-PA_{224} e LCMV-GP_{276}, 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-PA_{224}-R7A e LCMV-GP_{276}, do que entre este último e o alvo selvagem IAV-PA_{224} ou o controle negativo IAV-NP_{366} (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-$BMLF1_{300}$ ou o próprio HCV-$NS3_{1073}$, 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 reposta 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).

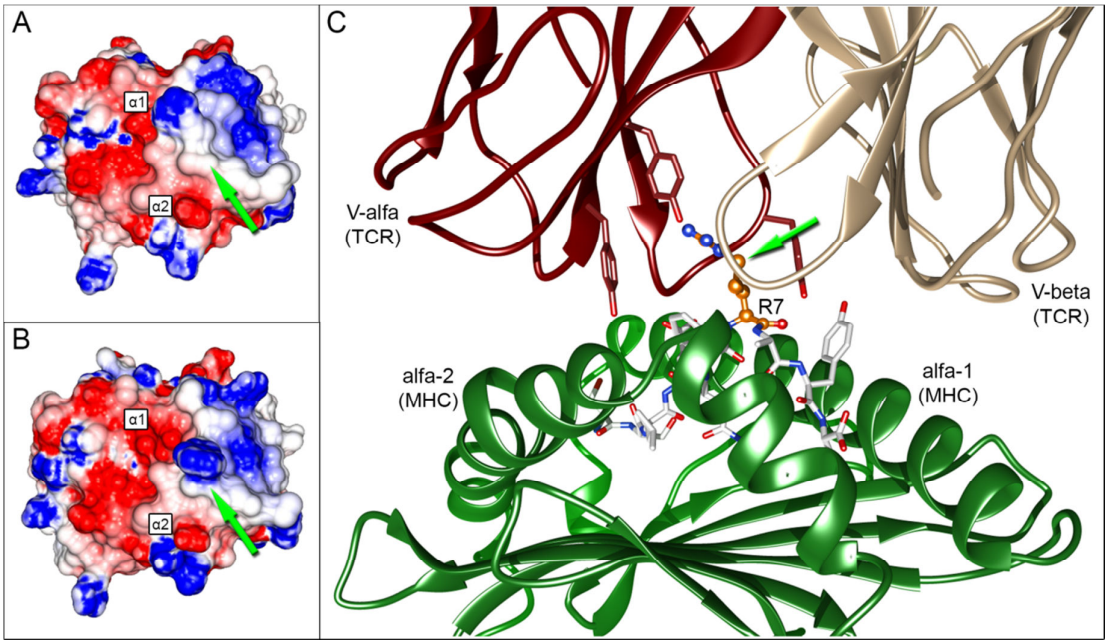


Figura 11. Interação específica entre o TCR e uma Arginina do epitopo. A. Superfície do complexo IAV-PA_{224}:H2-D^b calculada com o programa GRASP2 utilizando-se o cristal obtido na ausência do TCR (1WBY). B. Superfície do mesmo complexo, utilizando-se o cristal obtido na presença do TCR (3PQY). C. 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-BRLF1$_{109}$, 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-BMLF1$_{300}$, 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-BRLF1$_{109}$), mas também consegue reconhecer com certa afinidade/avidez o alvo heterólogo. Nesta rodada de estimulação frente ao EBV-BMLF1$_{300}$, 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-BMLF1$_{300}$. Um novo desafio com EBV-BRLF1$_{109}$ 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-$NS3_{1073}$ 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-NA_{231} e HCV-$NS3_{1073}$ não seriam agrupados na análise com todas as regiões, mas seriam agrupados em várias das análises com um número menor de regiões.

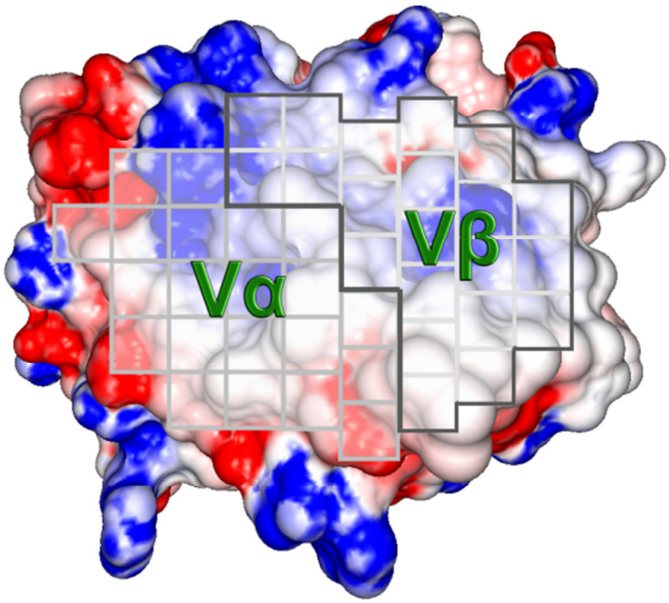


Figura 12. Nova proposta de seleção de regiões para agrupamento de complexos pMHC. Ao total, 42 regiões de tamanho uniforme foram definidas sobre a imagem da superfície do pMHC, cobrindo as principais áreas de interação com as cadeias Vα e Vβ do TCR. O arquivo delimitando estas regiões (*RoiSet*) foi 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. 5ª edição. 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. Trabalha de Conclusão de Curso em Biomedicina (UFRGS).

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, Wucherpfennig 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.
Fytili P, Dalekos GN, Schlaphoff V, Suneetha PV, Sarrazin C, Zauner W, Zachou K, Berg T, Manns MP, Klade CS *et al.* (2008) Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. Vaccine 26:3818-3826.
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.
Kasprowicz 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 Giralt 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, Traunecker 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, Sztein 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, Shirts MR, Smith JC, Kasson PM, van der Spoel D *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845-854.

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, Michaëlsson J and Kärre K (2005) A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity. J Biol Chem 280:27069-27075.

Schatz DG, Oettinger MA and Schlissel 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 Rigo[12], Marialva Sinigaglia[12], Gustavo F Vieira[12§]

[1]NBLI – 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.

[2]Programa 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.

Introduction

Cellular immunity is one of the two main branches of the adaptive immunologic response, focused on specific functions of the Cytotoxic T-Lymphocytes (CTLs). Although both cellular and humoral immunity are desired for an ideal and longstanding immunization, CTL response plays a central role in which regards to antiviral immunity [1]. After infecting a host cell, the virus will use the host molecular machinery to replicate its genome and produce new virions. In addition to all the mechanisms that allow virus scape from circulating neutralizing antibodies, during its intracellular replication cycle the virus is virtually hidden from the action of humoral immunity. However, some viral proteins will unavoidably be marked to enter the endogenous antigen presentation pathway. Through this route, virus-derived peptides will be presented at the cell-surface in the context of Major Histocompatibility Complex (MHC) class I molecules, forming stable peptide:MHC-I (pMHC-I) complexes. Each CTL produced by the host has one specific T-Cell Receptor (TCR), which is able to recognize pMHC-I complexes presenting nonself peptides. Therefore, through the interaction between pMHC-I complexes and TCRs, CTLs are able to identify and eliminate infected cells.

The TCR-pMHC-I interaction is highly specific, which allows the development of memory T-cells that will be once again triggered in future challenges with the same target. However, a given lymphocyte is able to recognize more than just one pMHC-I complex. This capacity of a CTL to recognize non-related peptides derived from the same virus, or even peptides from heterologous viruses, was defined as cross-reactivity [2]. As expected, cross-reactivity has direct implications over vaccine development, autoimmunity and heterologous immunity, a process by which the immunization with one pathogen confers protection against another [3-6]. Understanding of the molecular features driving these cross-reactivities became a major goal for several immunologists, but the system's complexity has delayed progress in the field. Wedemeyer *et al.* 2001 [7] has proposed that cross-recognition of two heterologous epitopes could be triggered by the high amino acid sequence similarity between them. Similarity in terms of biochemical properties was also proposed as being the key for cross-recognition [2], and was even applied with some success to predict cross-reactivity [8,9]. However, structural studies have shown that even epitopes with low sequence and biochemical

similarity might present quite identical pMHC-I surfaces [10,11], defending that this structural similarity should account for the cross-stimulation of a given T-cell population.

Structural analysis of pMHC-I complexes can provide a level of information much closer to that presented *in vivo* to interaction with the TCR. On the other hand, structural approaches are frequently limited by the number of pMHC-I structures already produced by experimental methods, such as X-Ray crystallography and NMR (Nuclear Magnetic Resonance). Our group has used structural bioinformatics tools to build *in silico* models of pMHC-I complexes that were not yet determined by experimental methods. This approach, referred as *D1-EM-D2* (*Docking 1 - Energy Minimization - Docking 2*), was previously validated through the successful reproduction of several crystal structures [12,13] and has been used to provide novel complexes for the CrossTope Data Bank for Cross-Reactivity Assessment [13]. Our group has also combined this approach with the use of multivariate statistical methods in order to make structural-based cross-reactivity predictions [11]. In this context, we present here an improved and standardized structural-based method for T-cell cross-reactivity prediction of HLA-A*02:01-restricted epitopes. The predictive capacity of our method was enhanced by the inclusion of new features, and our results with the analysis of viral epitopes (from Hepatitis C Virus and Dengue Virus) support its use as an important auxiliary tool for vaccine development.

Results and Discussion

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

The human HLA-A*02:01 is largely studied for being the most frequent MHC allele in human populations (http://www.allelefrequencies.net/) [14]. For this reason, the protein encoded by this specific allele (called allotype) also presents the larger number of crystal structures available at the Protein Data Bank (PDB). Aiming to identify the residues involved in the recognition of this allotype by different TCRs, we performed an extensive search for all available crystal structures of TCR-pMHC-I complexes, restricted to HLA-A*02:01. This search returned 29 complexes (Table S1), presenting 15 different TCRs and 18 different epitopes. Despite this variability, 5 epitope positions

and 4 MHC residues were consistently indicated as involved with TCR interactions, being presented in more than 85% of these complexes. Several residues over the pMHC-I surface might participate of the interaction with the TCR, influencing the specific level of T-cell stimulation that will be triggered by each pMHC-I. However, we here postulate that changes in these nine conserved contacts might have greater impact over the T-cell recognition, therefore preventing cross-reactivities.

An image-based strategy for pMHC-I clustering

In a previous study, our group used images of the electrostatic potential distribution over the pMHC-I surface to predict the cross-reactivity pattern among 28 naturally occurring HCV variants, in the context of HLA-A*02:01 [11]. In that study, one specific region over the pMHC-I surface was defined, based on the observation of the main spots of variation among the 28 complexes analyzed. From this specific region, were extracted values of Mean and Standard deviation for each RGB (Red, Green and Blue) component, which were used as input for a Principal Component Analysis (PCA). Based only on the extracted information from the pMHC-I structures, the PCA was able to predict the same clusters of cross-reactivity observed *in vitro* [11,15]. Despite of the success of this approach, the same parameters could not be applied to other subsets, once different regions of the pMHC-I surface might have greater influence over the TCR recognition. In the same study, seven variable regions over the pMHC-surface were defined and used to perform a structure-based virtual screening of putative cross-reactive targets among 55 non-related pMHC-I complexes [11]. This experiment predicted a cross-reactivity between two unrelated epitopes (one from HCV, other from EBV), which shared no amino acids in their sequences, and this prediction was later confirmed with *in vitro* experiments (Zhang S, personal communication).

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

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

Definition of key areas for TCR-pMHC-I contact

In the present work, we aimed to improve our structure-based prediction method and to provide a defined set of "gates" that could indicate the key interactions involved in cross-reactive responses, which could be applied to any subset of epitopes restricted to HLA-A*02:01. Considering the nine key positions identified in crystal structures and described above, we defined a group of seven regions over the pMHC-I surface (Figure 1). These regions, or "gates", were defined considering the specific contribution of each one of these residues to the combined surface of the pMHC-I. Three regions were defined covering the epitope surface. The contribution of epitope positions p4 and p5 were collected by two independent gates (G1 and G2). In the case of positions p6, p7 and p8, only one gate was defined, centered over p7 (G3). This was decided considering that given the conformation of the epitope backbone inside the cleft of HLA-A*02:01, p7 is much more exposed to the contact with the TCR, while p6 and p8 have lower contribution to the pMHC-I surface. Other four gates were defined over selected MHC-I residues (G4, G5, G6 and G7). These seven key regions are placed within the area previously described by other groups as the "TCR footprint" for this allotype [16-18] and, therefore, will be primarily responsible for triggering cross-reactive responses.

Inclusion of ASA values

Other limitation of our previously described approach was its entire dependence on the electrostatic potential information [11]. As discussed by the authors, there are experimental evidences suggesting that charge similarity is more important than subtle topography differences between the cross-reactive complexes [19,20]. However, pMHC-I complexes are 3D structures and topography variation certainly has some influence over the TCR recognition. The Accessible Surface Area (ASA) of a residue

can provide a quantitative measure of how exposed or buried its side chain is, which will have impact over the pMHC-I topography. ASA values of the epitope residues, for instance, were previously related to immunogenicity [21] and were also able to identify non-cross-reactive complexes [12]. Since direct 3D topography comparison remains an expensive computational challenge and ASA values can be easily calculated, we decided to include ASA values together with electrostatic potential information in order to improve our prediction method. When analyzing a subset with different epitopes in the context of the same allotype, the ligand variation will affect not only the charges and the ASA values of the epitope itself, but also of the MHC residues that direct interact with the peptide. For that reason, in addition to the ASA values for the nine epitope residues, we also included in our approach ASA values from 28 selected MHC-I residues (Figure S1).

Use of multiscale bootstrap resampling

As previously described, our prediction method was based on the use of pMHC-I structural features as input for multivariate statistical methods [11]. Originally, only information on electrostatic potential was used to define the clusters of putative cross-reactive complexes, being now combined with additional information on ASA values. Another remarkable improvement of our approach relates to the use of an R package (*pvclust*) for assessing the uncertainty of the Hierarchical Cluster Analysis (HCA) [22]. ´This package provides both Bootstrap Probability (BP) and Approximately Unbiased (AU) *p*-value, which is computed by multiscale bootstrap resampling and has been shown to be less biased than other methods in typical cases of phylogenetic tree selection [23]. This improvement adds a statistical validation to the dendrogram, enriching the discussion on the results, and avoiding unsubstantiated conclusions.

Method validation with a previously studied subset

To confirm its usability, our approach must be also able to identify the cross-reactive complexes in the previously described HCV subset [11,15] (Table S2). These 28 variants, covering all six HCV genotypes, were tested *in vitro* against the same T-cell population, which was obtained from an individual vaccinated with the wild-type epitope HCV-$NS3_{1073}$ (CINGVCWTV). The level of IFN-gamma production stimulated against a highly cross-reactive variant from genotype 1 (G1-01: CVNGVCWTV) was

defined as a reference of high response, which was used to classify the other variants into high, intermediate or low cross-reactive complexes (Table S2).

An HCA based in our improved approach was able to divide the complexes into three main clusters (Figure S2). A threshold was defined with the *pvrect* function to highlight these groups (*alpha=0.95*), which are corroborated by AU *p*-values with low standard error (Figure S3). The variant G3-18 (from Genotype 3) fell in a completely independent branch, forming a cluster on their one. This result is in agreement with our previous analysis and with the experimental data, since the G3-18 was the only one among the 28 complexes that presented no detectable response *in vitro* [15]. All the high cross-reactive complexes fell in the same main cluster. Of note, in the *in vitro* assay, the complexes with the higher levels within the cross-reactive complexes were G1-02, G1-07, G5-22, G6-25 and G6-27 [11,15]. In our HCA, four of these complexes fell in the same sub-cluster of the reference variant G1-01, although the AU/BP values does not strongly support the existence of this independent subgroup (*AU=77*). Most of the remaining high responders fell in a close related sub-cluster. The high responder variant G6-26 and the intermediate responder G1-05 fell in separate branches (*AU=99*), but still within the main cluster of the cross-reactive complexes (*AU=97*). It's also important to observe, that the original analysis of these complexes presented the intermediate responder G1-05 as the closest related complex to the reference complex G1-01 [11]. The authors discussed this unexpected result defending that despite of the surface similarity other issues might account for the lower response presented by G1-05, such as binding affinity of the epitope to the MHC and complex stability. Our improved approach was able to identify neglected structural differences between G1-01 and G1-05, and correctly place G1-05 outside of the sub-clusters of high responders.

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

speculation. Observe that the *in vitro* assay performed with these 28 HCV variants was to verify the cross-reactivity against the wild-type HCV-NS3$_{1073}$. Cross-reactivity also depends on the T-cell population involved, so in order to evaluate the cross-reactivity against G1-08, an assay with a G1-08-specific T-cell population would be needed.

Cross-reactivity prediction among dengue virus serotypes

Dengue virus (DV) represents a major challenge for vaccine development [24]. Despite effective immunization against one serotype is easy to achieve, and protective T-cell response is observed, challenge of an immunized individual with an heterologous serotype often leads to severe symptoms, such as dengue hemorrhagic fever and dengue shock syndrome (DHF/DSS). In this context, cross-reactive T-cells are believed to mediate the immunopathogenesis of DHF/DSS during secondary heterologous challenge [25]. Therefore, the identification of non-cross-reactive immunogenic targets, specific for each DV serotype, is one way to develop a combined tetravalent vaccine. In a recent publication, Zhi-Liang Duan and colleagues identified HLA-A*02:01-restricted peptides from the four DV serotypes, and examined their immunogenicity and cross-reactivity [25]. From their data, we extract the epitope sequence of two groups of targets, being one identified as (i) cross-reactive variants and other as (ii) non-cross-reactive variants (Figure S4).

We performed a new prediction with the combined data from both subsets (HCV and DV), totaling 36 pMHC-I complexes. The HCV and DV variants fell in independent main clusters, and the same threshold (*alpha=0,95*) was able to identify cross-reactive and non-cross-reactive complexes within these groups (Figure 2). A plot of the standard error is provided in Figure S5. All four NS4b variants fell in the same cluster (*AU=100*). This was expected, since cross-reactive response was indeed observed for these four variants. The same level of clustering was not observed for the NS4a variants (*AU=83*), a group that did not present cross-reactivity in the study of Zhi-Liang Duan and colleagues. Together with the AU/BP values, the dendrogram y-axis (Height) can also provide information on the dissimilarity among the complexes. A higher distance among the edges of the NS4a cluster can be observed when compared with the distance among the edges of the other main clusters.

The variants D1V-NS4a$_{140}$ and D4V-NS4a$_{140}$ fell in independent branches, while the other two (D2V-NS4a$_{140}$ and D3V-NS4a$_{140}$) fell in the same cluster (*AU=95*). Our

HCA, therefore, indicates a possible cross-reactivity between D2V-NS4a$_{140}$ and D3V-NS4a$_{140}$, which could be understood as a false positive result. However, is important to highlight that cross-reactivity is also dependent on the specific T-cell population involved, and normally produces responses with lower intensity when compared to the challenge with the cognate peptide. Of note, the D2V-NS4a$_{140}$ presented really low levels of response even upon challenge with the cognate epitope (Figure S4) [25]. Despite of a possible structural similarity, a cross-reactive response would be probably undetectable with this T cell population. Moreover, our approach relies exclusively on structural features of the pMHC-I surface, such as charges distribution and ASA values, being capable of identifying the closer related complexes. However, other features such as MHC binding and pMHC-I stability might also interfere with the T cell stimulation process, preventing the occurrence of cross-reactive responses.

Finally, the combined HCA (HCV and DV) was able to reproduce the same results observed in the independent HCV analysis. This combined approach corroborates the consistency of our method, even with a greater number of complexes, suggesting its possible use in a larger scale as a virtual screening method. In this sense, we also explored an alternative way to present our HCA results. Instead of a dendrogram, this data can be used as input for relational networks, which can provide more intuitive information about the cross-reactive-networks studied (Figure S6).

Applicability to vaccine development

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

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

Traditional methods of vaccine development provided some successful results, but have been unable to overcome some of the major challenges for global health, such as HIV and HCV. In that context, a new generation of rationalized vaccines is starting to be planed, and bioinformatics tools are playing a major role in this process [26,27]. Combined *in silico* approaches can save time and money, identifying the candidates more likely to stimulate the desired immune response, which can then be tested with *in vitro* and *in vivo* experiments to confirm its safety and efficacy for the use in a new vaccine.

Conclusions

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

Materials and Methods

Identification of conserved contacts between TCRs and pMHCs

An extensive search for all available crystal structures of TCR-pMHC-I complexes restricted to HLA-A*02:01 was performed in Protein Data Bank [28] and IMGT/3Dstructure-DB [29]. Curated and calculated contacts between TCR and pMHC, for each complex, were obtained from IEDB-3D [30]. Although available at IMGT/3Dstructure-DB, complexes 1QSE, 3QDG, 3QEQ, 3QDM, 3QDJ and 3UTS were not yet included in IEDB-3D and, therefore, were not included in our analysis. Information on included complexes is provided in Table S1.

Construction of pMHC-I complexes

All our structural analysis were performed with pMHC-I complexes obtained through the previously described *D1-EM-D2* approach [12]. Briefly, only the FASTA sequence of the epitopes was recovered from the reference studies [15,25] and used as input to produce 3D structures of these epitopes, with PyMOL scripts. A "donor" structure of an empty HLA-A*02:01 was obtained by removing the epitope from a reference PDB structure (PDB code 2V2W). The new pMHC-I structure, harboring the epitope of interest in the context of HLA-A*02:01, was then obtained by a combined sequence of Molecular Docking and Energy Minimization steps. These steps were performed with Autodock Vina [31] and GROMACS 4.5.1 package [32], respectively. The accuracy and reliability of this *D1-EM-D2* approach was tested in previous studies [12,13].

Calculations over the pMHC-I complexes

Electrostatic potential of each pMHC-I structure was calculated with Delphi [33], with custom parameters (e.g.: *indi=1.0*, *exdi=80.0*, *prbrad=1.4*, *salt=0.2*). Accessible Surface Area (ASA) of the selected residues from each pMHC-I complex was calculated with NACCESS V2.1.1 using default parameters (http://www.bioinf.manchester.ac.uk/naccess/).

Images acquisition and data extraction

Images of the electrostatic potential distribution over the "TCR-interacting surface" of each pMHC-I were obtained with UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) [34]. The "Electrostatic surfacing coloring" option of Chimera is used to import and visualize the electrostatic potential calculated with Delphi, using a range from -3 kiloteslas to +3 kiloteslas. Selected regions over these images were defined, and color histograms (RGB) of these areas were obtained with ImageJ 1.43u software (National Institute of Health, USA, http://rsb.info.nih.gov/ij). In total, 42 values were obtained from the seven histograms of each image, such as color mean and standard deviation for each RGB component. Figures included in the article were edited with Adobe Photoshop CS2 v.9.0. program (Adobe, San Jose, CA).

Clustering Analysis

Hierarchical cluster analysis was performed with R package Pvclust [22], assisted by the RStudio IDE v0.97 (http://www.rstudio.com/). The "average" linkage method was used with "correlation" distance, and the number of bootstrap replications was set to 10000. Results were plotted as dendrograms with Bootstrap Probabilities (BP) and Approximately Unbiased (AU) *p*-values. Main clusters were identified with *pvrect* (*alpha=0,95*) and standard errors for AU *p*-values were obtained with *seplot*. Relation networks were plotted with the open-source platform Gephi (https://gephi.org).

Acknowledgements

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

References

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

peptide ligands: implications for molecular mimicry in autoreactivity. The Journal of biological chemistry 280: 27069-27075.
11. Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, et al. (2011) Structural in silico analysis of cross-genotype-reactivity among naturally occurring HCV NS3-1073-variants in the context of HLA-A*02:01 allele. Molecular immunology 48: 1461-1467.
12. Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M, et al. (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.
13. Sinigaglia M, Antunes DA, Rigo MM, Chies JA, Vieira GF (2013) CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. Database (Oxford) 2013: bat002.
14. Fernandez-Vina MA, Falco M, Sun Y, Stastny P (1992) DNA typing for HLA class I alleles: I. Subsets of HLA-A2 and of -A28. Hum Immunol 33: 163-173.
15. Fytili P, Dalekos GN, Schlaphoff V, Suneetha PV, Sarrazin C, et al. (2008) Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. Vaccine 26: 3818-3826.
16. Rudolph MG, Stanfield RL, Wilson Ia (2006) How TCRs bind MHCs, peptides, and coreceptors. Annual review of immunology 24: 419-466.
17. Gras S, Burrows SR, Turner SJ, Sewell AK, McCluskey J, et al. (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.
18. Gras S, Saulquin X, Reiser J-B, Debeaupuis E, Echasserieau K, et al. (2009) Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. Journal of immunology (Baltimore, Md : 1950) 183: 430-437.
19. Jorgensen JL, Esser U, Fazekas de St Groth B, Reay PA, Davis MM (1992) Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. Nature 355: 224-230.
20. Kessels HWHG, de Visser KE, Tirion FH, Coccoris M, Kruisbeek AM, et al. (2004) The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. Journal of immunology (Baltimore, Md : 1950) 172: 2324-2331.
21. Meijers R, Lai C-CC, Yang Y, Liu J-HH, Zhong W, et al. (2005) Crystal structures of murine MHC Class I H-2 D(b) and K(b) molecules in complex with CTL epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance. J Mol Biol 345: 1099-1110.
22. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22: 1540-1542.
23. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol 51: 492-508.
24. Halstead SB (2013) Identifying protective dengue vaccines: Guide to mastering an empirical process. Vaccine.
25. Duan ZL, Li Q, Wang ZB, Xia KD, Guo JL, et al. (2012) HLA-A*0201-restricted CD8+ T-cell epitopes identified in dengue viruses. Virol J 9: 259.
26. Donati C, Rappuoli R Reverse vaccinology in the 21st century: improvements over the original design. Ann N Y Acad Sci 1285: 115-132.
27. Dormitzer PR, Grandi G, Rappuoli R Structural vaccinology starts to deliver. Nat Rev Microbiol 10: 807-813.

28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.
29. Kaas Q, Ruiz M, Lefranc M-P (2004) IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic acids research 32: D208-210.
30. Ponomarenko J, Papangelopoulos N, Zajonc DM, Peters B, Sette A, et al. (2011) IEDB-3D: structural data within the immune epitope database. Nucleic acids research 39: D1164-1170.
31. Trott O, Olson AJ, 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.
32. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29: 845-854.
33. Li L, Li C, Sarkar S, Zhang J, Witham S, et al. (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys 5: 9.
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

Figures

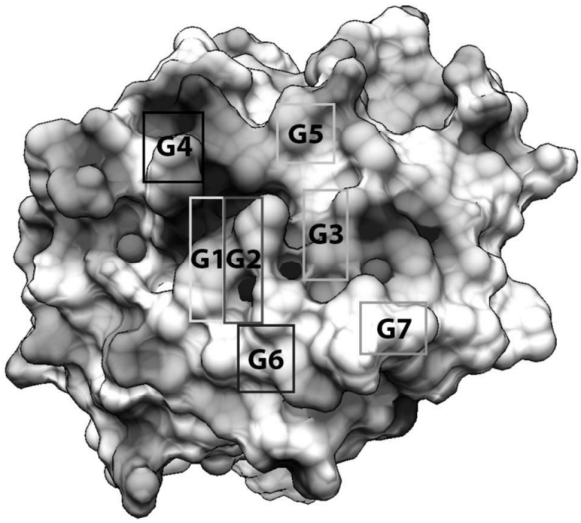


Figure 1 – Seven Gates defined to obtain color histograms.

Top view of a pMHC-I complex presenting a Dengue-derived epitope in the cleft of HLA-A*02:01, obtained with UCSF Chimera package [31]. Electrostatic potential over the surface was computed with Delphi program and represented as red (negative charges) and blue (positive charges) spots, with a range from -3 kiloteslas to +3 kiloteslas. The seven gates (G1 to G7) relate to conserved contacts with different TCRs, as observed in the crystal structures available, and were selected for the RGB analysis with ImageJ.

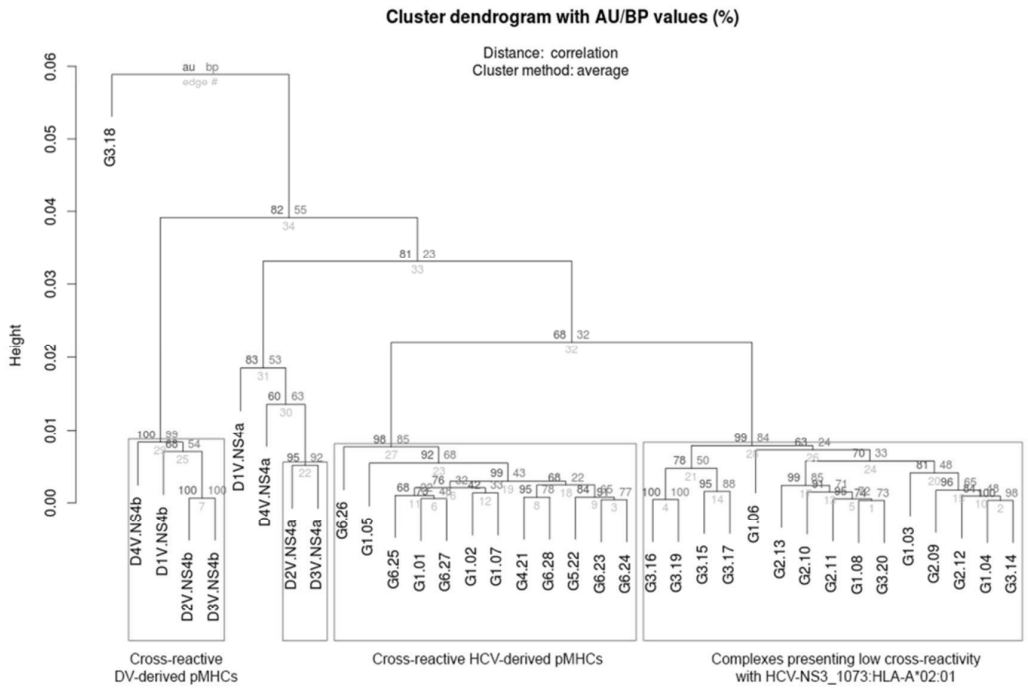


Figure 2 - Structure-based hierarchical clustering of pMHC-I complexes.

Dendrogram of 36 pMHC-I complexes representing the Hierarchical Cluster Analysis performed with Pvclust R package. The input data was Accessible Surface Area values and color histograms (RGB) for each pMHC-I, which provided information on topography and charges distribution over the surface. Red boxes indicate the main clusters identified (*alpha=0,95*). Cross-reactive and non-cross-reactive complexes of both subsets (HCV and DV) fell in independent clusters. AU, Approximately Unbiased; BP, Bootstrap Probability.

Supporting Files

Table S1 – Conserved contacts between TCR and pMHC-I in available crystal structures.

Table with the twenty-nine non-redundant HLA-A*02:01 restricted crystal structures of TCR-pMHC-I complexes available. Information on PDB code, MHC, TCR, epitope ID and interactions with TCR are provided. Curated and calculated contacts between TCR and pMHC-I, for each complex, were obtained from IEDB (www.iedb.org).

Table S2 – Information on the 36 pMHC complexes analyzed.

Table containing complete information on the 36 pMHC-I complexes analyzed. Details on epitope sequence, source protein and ID codes for databanks hosting complementary information are provided. In each subset (HCV and DV), wild-type epitope sequence is depicted with all letters in black, while red letters indicate mutated amino acids in relation to respective wild-type.

Table S1

PDB	Assay ID (IEDB)	MHC	TCR	Epitope	Epitope ID (IEDB)	Interaction Epitope/TCR (from IEDB)	Interaction MHC/TCR (from IEDB)
1AO7	1584448	HLA-A*02:01	A6	HTLV1-Tax_11-19	37257	C: L1, L2, G4, Y5, P6, V7, Y8	A: E58, R65, K66, K68, A69, Q72, A149, A150, H151, Q155, A158, Y159, T163, E166, W167, R170
1BD2	1478873	HLA-A*02:01	B7	HTLV1-Tax_11-19	37257	C: L1, G4, Y5, P6, V7, Y8	A: Y59, G62, R65, A69, Q72, A150, Q155, A158, G162, T163, W167
1LP9	1617229	HLA-A*02:01	AHIII 12.2	Self peptide P1049	2999	C: L2, W3, G4, F5, F6, P7, V8	A: R65, K66, K68, A69, Q72, K146, W147, A149, A150, H151, V152, E154, Q155, A158, Y159, G162, T163, E166, W167
1OGA	1005001	HLA-A*02:01	JM22 (Vβ17Vα10.2)	IAV-M1_58-66	20354	C: G4, F5, V6, T8	A: R65, K68, A69, Q72, T73, R75, V76, A149, A150, H151, V152, E154, Q155
1QRN	1496444	HLA-A*02:01	A6	HTLV1-Tax_11-19 (P6A)	37253	C: L2, G4, Y5, A6, V7, Y8	A: R65, K66, K68, A69, Q72, A150, H151, E154, Q155, A158, Y159, T163, E166, W167, R170
1QSF	1496445	HLA-A*02:01	A6	HTLV1-Tax_11-19 (Y8A)	37255	C: L1, L2, G4, Y5, V7	A: E58, R65, K66, K68, A69, Q72, A150, Q155, A158, Y159, T163, E166, W167, R170
2BNQ	1617378	HLA-A*02:01	1G4	NY-ESO-1_157-165 (C9V)	59283	C: M4, W5, I6, T7, Q8	A: R65, K66, K68, A69, Q72, T73, A150, H151, E154, Q155
2BNR	1617377	HLA-A*02:01	1G4	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: R65, K66, K68, A69, Q72, T73, A150, H151, E154, Q155
2F53	1618325	HLA-A*02:01	1G4 (C49c50)	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: G62, R65, K66, K68, A69, S71, Q72, T73, R75, V76, A150, H151, E154, Q155
2F54	1618324	HLA-A*02:01	1G4 (AV-wt)	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: R65, K66, K68, A69, Q72, T73, K146, A150, H151, E154, Q155
2GJ6	1511838	HLA-A*02:01	A6	HTLV1-Tax_11-19 (Y5K)	130542	C: L1, L2, F3, G4, K5, P6, V7, Y8	A: R65, K66, K68, A69, Q72, T73, K146, A149, A150, H151, Q155, A158, Y159, T163, E166, W167, R170
2J8U	1617381	HLA-A*02:01 (K66A)	AHIII 12.2	Self peptide P1049	2999	C: G4, F5, F6, P7, V8	A: G62, R65, A66, K68, A69, Q72, K146, W147, A150, H151, V152, E154, Q155, A158, G162, T163, E166
2JCC	1617380	HLA-A*02:01 (W167A)	AHIII 12.2	Self peptide P1049	2999	C: L2, W3, G4, F5, F6, P7, V8	A: R65, K66, K68, A69, Q72, K146, W147, A150, H151, V152, E154, Q155, A158, Y159, G162, T163, E166
2P5E	1511834	HLA-A*02:01	1G4 (c58c61)	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: R65, K66, K68, A69, Q72, T73, R75, V76, A150, H151, E154, Q155, T163
2P5W	1511835	HLA-A*02:01	1G4 (c58c62)	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: E19, G62, R65, K66, K68, A69, S71, Q72, T73, R75, V76, A149, A150, H151, E154, Q155, T163
2PYE	1511833	HLA-A*02:01	1G4 (c5c1)	NY-ESO-1_157-165	59278	C: M4, W5, I6, T7, Q8	A: R65, K66, K68, A69, Q72, T73, V76, A150, H151, Q155
2UWE	1617379	HLA-A*02:01 (T163A)	AHIII 12.2	Self peptide P1049	2999	C: L2, W3, G4, F5, F6, P7, V8	A: R65, K66, K68, A69, Q72, K146, W147, A149, A150, H151, V152, E154, Q155, A158, Y159, G162, A163, E166, W167
2VLJ	1505355	HLA-A*02:01	JM22 (Vβ17:Vα10.2)	IAV-M1_58-66	20354	C: G4, F5, V6, T8	A: A69, Q72, V76, A149, A150, H151, V152, E154, Q155
2VLK	1511521	HLA-A*02:01	JM22 (Vβ17:Vα10.2)	IAV-M1_58-66	20354	C: G4, F5, V6, T8	A: A69, Q72, V76, A149, A150, H151, V152, E154, Q155
2VLR	1511522	HLA-A*02:01	JM22 (Vβ17:Vα10.2)	IAV-M1_58-66	20354	C: G4, F5, V6, T8	A: A69, Q72, T73, R75, V76, A149, A150, H151, V152, E154, Q155
3D39	1975805	HLA-A*02:01	A6	HTLV1-Tax_11-19 (Y5F*)	186691	C: L1, L2, G4, (PFF)5, P6, V7, Y8	A: R65, K66, K68, A69, Q72, T73, A149, A150, H151, E154, Q155, A158, Y159, T163, E166, W167, R170
3D3V	1975807	HLA-A*02:01	A6	HTLV1-Tax_11-19 (Y5F**)	186691	C: L1, L2, F3, G4, (F2F)5, P6, V7, Y8	A: E58, R65, K66, K68, A69, Q72, T73, A150, H151, E154, Q155, Y159, T163, E166, W167, R170
3GSN	1714843	HLA-A*02:01	RA14	HCMV-pp65_485-493	44920	P: N1, P4, M5, V6, A7, T8	H: G62, R65, K66, A69, Q72, R75, V76, K146, A149, A150, Q155, A158
3H9S	1851293	HLA-A*02:01	A6	*S.c*.-TEL1_549-557	42094	C: M1, L2, W3, G4, Y5, L6, Q7, Y8	A: R65, K66, K68, A69, Q72, H151, E154, Q155, A158, T163, E166, W167, R170
3HG1	1883845	HLA-A*02:01	MEL5	MART-1 _26-35	12941	C: E1, L2, A3, G4, I5, G6, I7, L8, T9	A: G62, R65, K66, K68, Q72, T73, R75, V76, H151, E154, Q155, A158, Y159, T163, E166, W167
3O4L	1831437	HLA-A*02:01	AS01	EBV-BMLF1_259-267	20788	C: C3, T4, L5, V6, A7, M8	A: G62, R65, K66, A69, Q72, R75, K146, A150, V152, E154, Q155, A158, Y159, T163, E166, W167
3PWP	1850160	HLA-A*02:01	A6	HuD_87-95	36357	C: L1, G2, G4, F5, V6, N7, Y8	A: E58, R65, K66, K68, A69, Q72, T73, K146, A150, E154, Q155, A158, Y159, T163, E166, W167, R170
3QFJ	1976480	HLA-A*02:01	A6	HTLV1-Tax_11-19 (Y5F)	184429	C: L1, L2, F3, G4, F5, P6, V7, Y8	A: R65, K66, K68, A69, Q72, T73, K146, A150, H151, E154, Q155, A158, Y159, T163, E166, W167, R170
3UTT	1930414	HLA-A*02:01	1E6	Self peptide P01308	103041	C: L2, G4, P5, D6, P7, A8	A: G62, R65, K66, Q72, V76, A150, H151, V152, Q155

* Position Y5 of the natural sequence is substituted with a 4-fluoro-phenylalanine residue

** Position Y5 of the natural sequence is substituted with a 3,4-difluoro-phenylalanine residue

Table S1 (contin.)

Interaction TCR/Epitope (from IEDB)	Interaction TCR/MHC (from IEDB)
D: Q30, S31, T93, D99, S100; E: E30, R95, L98, G100, G101, P103	D: K1, D26, R27, Q30, Y50, N52, K68, T98, D99, W101, G102; E: L98, G101, R102, P103
D: M28, D30, Y31, M93, G95, A96; E: Y96, P97, G98, Y104	D: S27, M28, S50, S51, I52, E94, Q102, K103; E: Y48, I54, G100, Y104
E: F93, A97, S98, S100, F101, S102; F: Y31, W97, S99	E: Y28, S29, F31, F50, T51, F93, S98, S99, F101; F: D30, Y48, Y50, V51, E56, W97, V98, Y100
D: S95, Q96, G97; E: D32, Q52, I53, R98, S99, S100	D: S31, V51, G94; E: I53, V54, N55, D56, Q58, R98, S100, Y101
D: Q30, S31, T93, D99, S100; E: E30, R95, L98, A99, G100, G101, P103	D: R27, Q30, Y50, N52, K68, T98, D99, W101, G102; E: G100, G101, R102, P103
D: G28, Q30, S31, T93, D99, S100; E: G100, P103	D: S25, D26, R27, G28, Q30, S31, Y50, S51, N52, K68, T98, D99, W101, G102; E: L98, G100, G101, P103
D: Y31, R93, P94, T95, S96, G97, G98, Y100; E: N27, E29, Y94, V95, G96, N97	D: Y31, Q51, S52, S53, Q54, G97, G98, S99, Y100; E: E29, Y47, V49, G50, A51, I53, D55, T70, V95, N97
D: Y31, R93, P94, T95, S96, G97, G98, S99, Y100; E: N27, E29, Y94, V95, G96, N97	D: Y31, Q51, S53, Q54, T95, G98, S99, Y100; E: E29, Y47, V49, G50, A51, I53, D55, T70, V95, N97
D: Y30, R92, P93, T94, S95, G97, S98, Y99; E: N26, E28, Y93, V94, G95, N96	D: Y30, F51, W52, G96, G97, Y99; E: N26, E28, Y46, V48, S49, V50, M52, D54, T69, V94, N96
D: Y30, R92, P93, T94, S95, G96, G97, S98, Y99; E: N26, E28, Y93, V94, G95, N96	D: Y30, Q50, S51, S52, Q53, G97, S98, Y99; E: E28, Y46, V48, G49, A50, I52, D54, T69, V94, N96
D: Q30, T98, D99, S100; E: E30, G97, L98, A99, P103	D: R27, Q30, Y50, S51, N52, K68, T98, D99, W101, G102; E: L98, G100, G101, R102, P103
E: F93, S99, S100, F101, S102; F: Y31, W97, S99	E: Y28, S29, F31, F50, T51, F93, L96, A97, S98, S100; F: Y48, Y50, V51, E56, W97, V98, Y100
E: F93, A97, S98, S100, F101, S102; F: Y31, W97, S99	E: Y28, S29, F31, F50, F93, S98, S99, F101; F: Y48, Y50, V51, E56, W97, V98, Y100
D: Y31, R93, P94, L95, L96, D97, G98, T99, Y100; E: N26, E28, Y93, L94, G95, N96	D: Y31, W53, L96, G98, T99, Y100; E: N26, E28, Y46, V48, A49, I50, T52, D54, I69, L94, N96
D: Y31, R93, P94, L95, L96, D97, G98, T99, Y100; E: N26, E28, Y93, L94, G95, N96	D: Y31, W53, Q54, L96, G98, T99, Y100; E: E28, Y46, V48, S49, V50, M52, D54, I69, L94, N96
D: Y31, R93, P94, L95, L96, D97, G98, T99, Y100; E: N26, E28, Y93, L94, G95, N96	D: Y31, Q51, S53, G98, T99, Y100; E: N26, E28, Y46, V48, G49, A50, T52, D54, I69, L94, N96
E: F93, A97, S98, S100, F101, S102; F: Y31, W97	E: Y28, S29, F31, F50, T51, F93, S98, S99, F101; F: Y48, Y50, V51, E56, W97, Y100
D: S95, Q96, G97; E: D32, Q52, I53, R98, S99, S100	D: S31, V51, A93, G94; E: I53, N55, D56, R98, S100, Y101
D: S95, Q96, G97; E: D32, Q52, I53, S99	D: S31, V51, G94; E: I53, V54, N55, D56, R98, S100, Y101
D: S95, Q96, G97; E: D32, Q52, I53, S100	D: S31, V51, A93, G94; E: I53, V54, N55, D56, R98, S100, Y101
D: G28, Q30, S31, T93, T98, D99, S100; E: E30, L98, A99, G100, G101, P103	D: R27, G28, Q30, Y50, K68, T98, D99, W101, G102; E: L98, G100, G101, R102, P103
D: G28, Q30, S31, D99, S100; E: E30, L98, A99, G100	D: D26, R27, G28, Q30, Y50, K68, T98, D99, W101, G102; E: L98, G100, G101, R102, P103
A: N29, F30, Y31, N93, G95, N96; B: E30, T97, G98, G99	A: N29, Y31, T51, L52, T94, N96; B: E30, Y48, V50, I54, D56, V96, I100, Y101
D: G28, Q30, D99, S100; E: E30, L98, A99, G100, G101	D: R27, Q30, Y50, S51, N52, K55, K68, T98, D99, W101, G102; E: L98, G101, R102, P103
D: G29, Q31, S32, Y51, N92; E: T96, L98, G99	D: R28, G29, Q31, Y51, A94, K96; E: N30, V51, Q55, E59, T96, G97, G99, T100
D: Y32, D93, N95, A96; E: R98, G100, T101, G102, N103	D: S29, T31, Y32, Y49, F51, S52, K69, N95, A96, R97; E: N52, E53, Y59, E60, T101, N103
D: G28, Q30, S31, D99, S100; E: E30, L98, A99, P103	D: D26, R27, G28, Q30, Y50, K68, T98, D99, W101, G102; E: L98, A99, G101, R102, P103
D: G28, Q30, S31, T98, D99, S100; E: E30, L98, A99, G100, P103	D: R27, G28, Q30, Y50, S51, N52, K68, T98, D99, W101, G102; E: L98, A99, G100, G101, R102, P103
D: R92, D94, S95, S96, Y97; E: Y31, W97	D: Y32, D94, S95, S96; E: N50, N51, V53, I55, D56, W97, E98, A101, K102

Table S2

Code	Sequence	Immune response	MHC Allele	Virus	Protein	Epitope Position	Epitope ID (IEDB)	Complex ID (CrossTope)
G1-01	CVNGVCWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	7299	A0201_0031
G1-02	CTNGVCWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95297	A0201_0032
G1-03	CVSGACWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95299	A0201_0033
G1-04	CISGVCWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95281	A0201_0034
G1-05	CINGACWTV	Intermediate*	HLA-A*02:01	HCV	NS3	1073-1081	6430	A0201_0035
G1-06	CVNGACMTV	Intermediate*	HLA-A*02:01	HCV	NS3	1073-1081	95298	A0201_0036
G1-07	CINGVCWSV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95279	A0201_0037
G1-08	CINGVCWSI	Intermediate*	HLA-A*02:01	HCV	NS3	1073-1081	95278	A0201_0038
G2-09	CISGVLWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95282	A0201_0039
G2-10	TISGVLWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95910	A0201_0040
G2-11	SISGVLWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95855	A0201_0041
G2-12	SIAGVLWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95851	A0201_0042
G2-13	TISGILWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95909	A0201_0043
G3-14	TVGGVTWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95938	A0201_0044
G3-15	SVGGVMWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95889	A0201_0045
G3-16	TIGGVMWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95907	A0201_0046
G3-17	AIGGVMWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95228	A0201_0047
G3-18	TVGDVMWTV	No response*	HLA-A*02:01	HCV	NS3	1073-1081	95935	A0201_0048
G3-19	TVGGVMWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95937	A0201_0049
G3-20	TVGGVIWTV	Low Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95936	A0201_0050
G4-21	AVNGVMWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95265	A0201_0051
G5-22	CINGVMWTL	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95280	A0201_0052
G6-23	SINGVMWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95854	A0201_0053
G6-24	AINGVMWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	2033	A0201_0054
G6-25	TVNGVMWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95890	A0201_0055
G6-26	AVNGVLWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95264	A0201_0056
G6-27	TINGVLWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95908	A0201_0057
G6-28	TVNGVLWTV	High Responder*	HLA-A*02:01	HCV	NS3	1073-1081	95940	A0201_0058
D1V-NS4a	GLLFMILTV	Non-cross-reactive**	HLA-A*02:01	DV	NS4a	140-148	179798	-
D2V-NS4a	AILTVVAAT	Non-cross-reactive**	HLA-A*02:01	DV	NS4a	140-148	179761	-
D3V-NS4a	GILTLAAIV	Non-cross-reactive**	HLA-A*02:01	DV	NS4a	140-148	179796	-
D4V-NS4a	TILTIIGLI	Non-cross-reactive**	HLA-A*02:01	DV	NS4a	140-148	179915	-
D1V-NS4b	LLMRTTWAL	Cross-reactive**	HLA-A*02:01	DV	NS4b	183-191	179847	-
D2V-NS4b	LMMRTTWAL	Cross-reactive**	HLA-A*02:01	DV	NS4b	182-190	150389	-
D3V-NS4b	LLMRTSWAL	Cross-reactive**	HLA-A*02:01	DV	NS4b	182-190	179845	-
D4V-NS4b	LLMRTTWAF	Cross-reactive**	HLA-A*02:01	DV	NS4b	179-187	179846	-

* Considering the level of IFN-gamma production when presented to a T-cell population specific to the HCV-$NS3_{1073}$ wild-type epitope (CINGVCWTV), as tested by Fytili *et al.*2008.

** Considering the frequency of $CD8^+$/IFN-gamma$^+$ T-cells after stimulation with the cognate or an heterologous peptide (*see* Figure X), as tested by Duan *et al.* 2012.

Letters depicted in red indicate the variant amino acid. HCV, Hepatitis C Virus; DV, Dengue Virus.

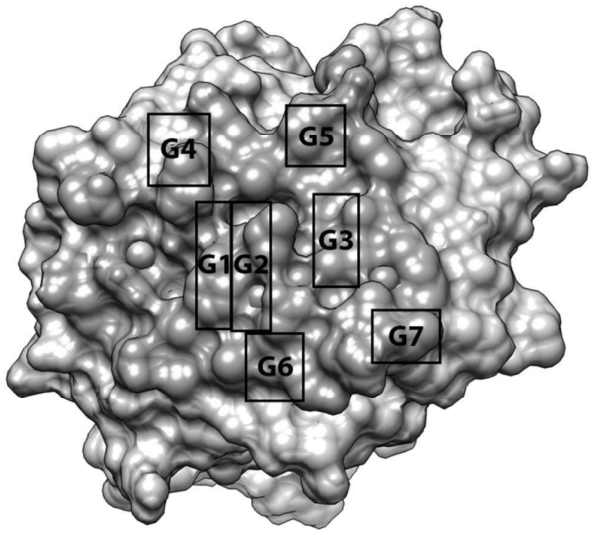


Figure S1 – Selected residues for ASA assessment.
Top view of a pMHC-I complex presenting a Dengue-derived epitope in the cleft of HLA-A*02:01, obtained with UCSF Chimera package [31]. Complex surface is depicted in grey while surface of all residues selected for ASA assessment are depicted in blue. Black rectangles indicate the seven gates (from G1 to G7) used in the RGB analysis.

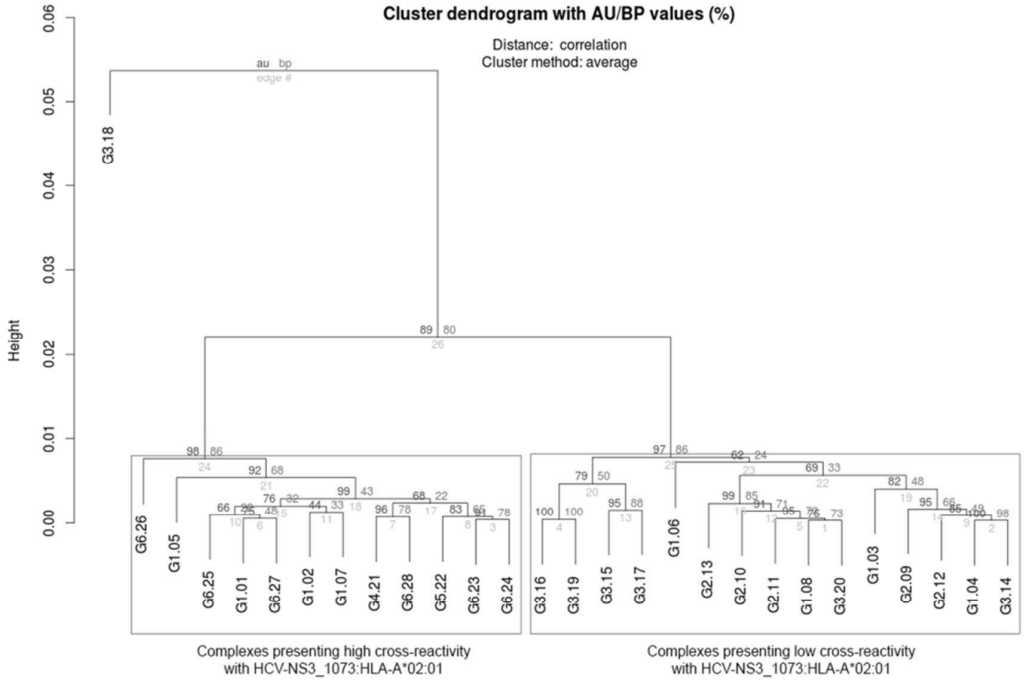


Figure S2 – HCA of 28 HCV natural occurring variants

Dendrogram representing the Hierarchical Cluster Analysis (HCA) of 28 pMHC-I complexes loaded with HCV-derived epitopes covering all six HCV genotypes (from G1 to G6). The input data was Accessible Surface Area values and color histograms (RGB) for each pMHC-I, which provided information on topography and charges distribution over the surface. Red boxes indicate the main clusters identified (*alpha=0,95*). High cross-reactive and low cross-reactive complexes fell in independent main clusters. The only complex that presented no response *in vitro*, G3-18, fell alone in an independent branch. Information on the specific response presented by each complex in cross-reactivity tests (*in vitro*) is provided in Additional file 3. G, HCV Genotype; AU, Approximately Unbiased; BP, Bootstrap Probability.

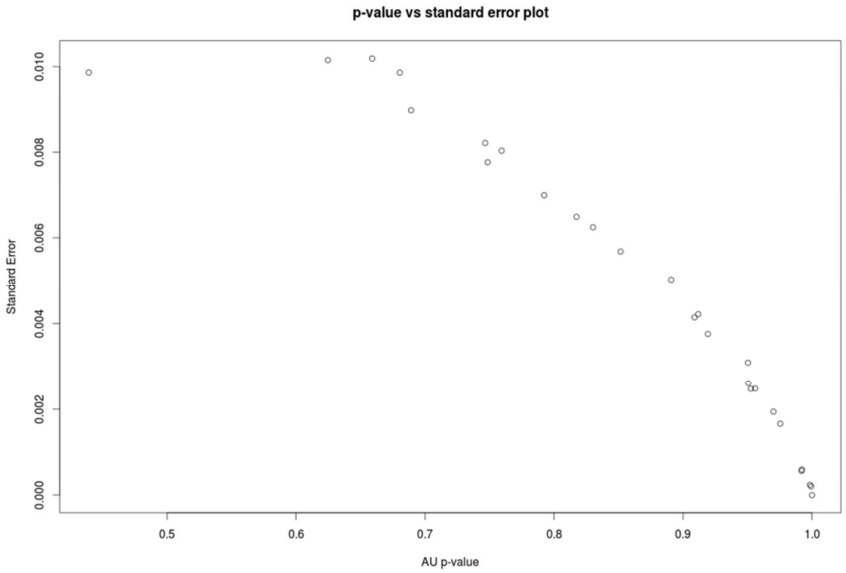


Figure S3 – Standard Error plot from the clustering of HCV variants.
Simple scatter plot of the Standard Error (SE) of each AU *p*-value calculated in the Hierarchical Cluster Analysis of 28 pMHC-I complexes presenting HCV-derived epitopes (Additional file 4). Observe that all AU *p*-values presented very low values of SE. Plot generated with *seplot* function assisted by RStudio IDE.

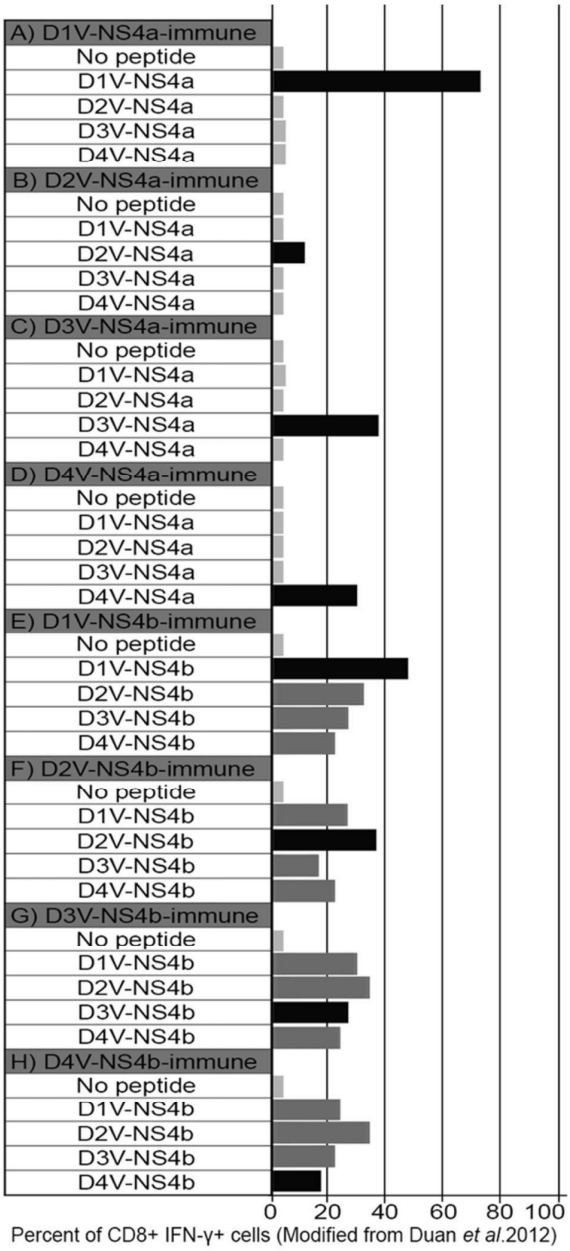


Figure S4 – Cross-reactivities among DV serotypes.

Schematic representation of cross-reactivities among Dengue Virus (DV) serotypes as previously tested by Duan *et al.* 2012. Splenocytes were isolated from peptide-immunized mice and were stimulated in vitro with cognate or heterologous peptide. In each case, response against cognate peptide is depicted in black, cross-reactive responses are depicted in blue, and negative responses are depicted in grey. (A) Responses observed with splenocytes from D1V-NS4a-immune mice. (B) Responses observed with splenocytes from D2V-NS4a-immune mice. (C, D) Responses observed with splenocytes from D3V-NS4a-immune mice and D4V-NS4a-immune mice, respectively. (E, F, G, H) Responses observed with splenocytes from mice immunized with D1V-NS4b, D2V-NS4b, D3V-NS4b and D4V-NS4b, respectively.

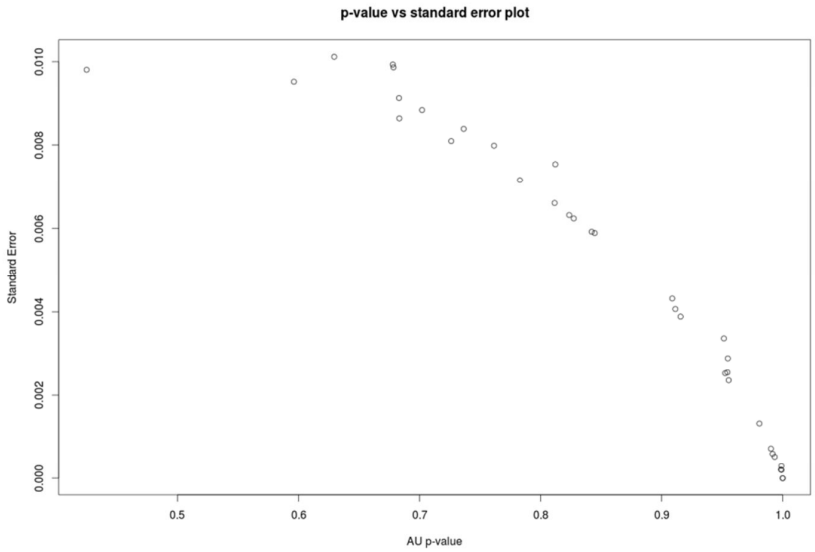


Figure S5 – Standard Error plot from the clustering of 36 pMHC-I complexes. Simple scatter plot of the Standard Error (SE) of each AU *p*-value calculated in the Hierarchical Cluster Analysis of 36 pMHC-I complexes (Figure 2). Observe that all AU *p*-values presented very low values of SE. Plot generated with *seplot* function assisted by RStudio IDE.

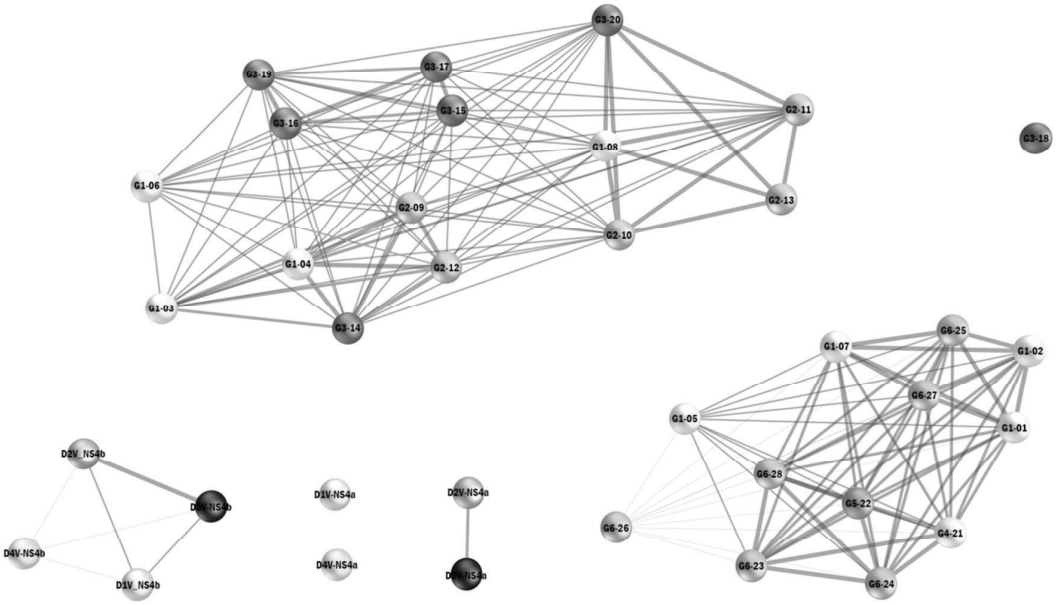


Figure S6 – Relational network of 36 pMHC-I complexes.
Relational network generated with Gephi program, based on the dendrogram of 36 pMHC-I complexes (Figure 2). Each sphere represents a given pMHC-I and different colors indicate different HCV genotypes or DV serotypes. For instance, red spheres indicate pMHC-I complexes loaded with HCV genotype 3 epitopes. Lines (*edges*) indicate cross-reactivity between the connected complexes (*nodes*), complexes without connections are considered non-cross-reactive. The strength of each line indicates the similarity between the connected complexes, being a structure-based indicative of the strength of the cross-reactivity between them. The distribution of the clusters is merely representative, and distance between *nodes* in the picture has no meaning.


NÚCLEO DE BIOINFORMÁTICA DO LABORATÓRIO DE IMUNOGENÉTICA UFRGS