UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

DISSERTAÇÃO DE MESTRADO

ESTUDO *IN SILICO* DA REATIVIDADE CRUZADA ENTRE EPITOPOS DE HANTAVÍRUS

MAURÍCIO MENEGATTI RIGO

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

Orientador: Prof. Dr. Gustavo Fioravanti Vieira

Co-orientador: Prof. Dr. José Artur Bogo Chies

PORTO ALEGRE

MARÇO DE 2011

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

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

Bill & Melinda Gates Foundation

AGRADECIMENTOS

Agradeço à minha família, especialmente ao meu pai e a minha mãe, por atuarem como órgãos linfóides primários, provendo meu sustento e ajudando no meu desenvolvimento.

Agradeço a minha namorada Mariana, meu TCR privado e que não apresenta reatividade cruzada com nenhum outro complexo pMHC.

Agradeço à família da minha namorada, por atuarem como órgãos linfóides secundários, me auxiliando em diversos momentos.

Agradeço aos meus amigos, por tolerarem minhas citocinas citotóxicas e agirem como verdadeiros linfócitos T auxiliares.

Agradeço a meus orientadores, por atuarem como um super antígeno, me estimulando freqüentemente para que eu desse o meu melhor.

Agradeço aos meus colegas de laboratório, pelo convivo diário no contexto da imunologia.

Agradeço aos meus professores da pós, pela seleção positiva e negativa, contribuindo para que eu não fosse mais uma célula academicamente anérgica.

Agradeço ao Elmo e à Ellen, moléculas acessórias importantíssimas para manutenção do meu mestrado.

Agradeço aos meus psicoterapeutas, pela imunomodulação do meu ser.

Agradeço ao apoio financeiro, pelo “booster” mensal.

Agradeço a Deus, por me manter responsivo, quando todo o resto falhou.

“*Try to learn something about everything and everything about something!*”

(Thomas Henry Huxley)

SUMÁRIO

LISTA DE ABREVIATURAS

ADT	*AutoDock Tools*
APC	Célula Apresentadora de Antígenos (do inglês *Antigen Presenting Cell*)
BE	Energia de Ligação (do inglês *Binding Energy*)
CD4	Grupamento de Diferenciação 4 (do inglês *Cluster of Differentiation 4*)
CD8	Grupamento de Diferenciação 8 (do inglês *Cluster of Differentiation 8*)
CDR	Regiões Determinantes de Complementaridade (do inglês *Complementarity Determining Regions*)
CTL	Linfócito T Citotóxico (do inglês *Cytotoxic T Lymphocyte*)
DC	Células Dendríticas (do inglês *Dendritic Cells*)
DNA	Ácido desoxirribonucleico (do inglês *Desoxirribonucleic acid*)
EBV	Vírus do Epstein-Barr (do inglês *Epstein-Barr Virus*)
ER	Retículo Endoplasmático (do inglês *Endoplasmic reticulum*)
ERAP	Amino Peptidase do Retículo Endoplasmático (do inglês *Endoplasmic Reticulum Amino Peptidase*)
Gc	Glicoproteína C-terminal (do inglês *Glycoprotein C-terminal*)
Gn	Glicoproteína N-terminal (do inglês *Glycoprotein N-terminal*)
HCPS	Síndrome Cardiopulmonar causada por Hantavírus (do inglês *Hantavirus Cardiopulmonary Syndrome*)
HFRS	Síndrome Renal com Febre Hemorrágica (do inglês *Hemorrhagic Fever with Renal Syndrome*)
HIV	Vírus da Imunodeficiência Human (do inglês *Human Immunodeficiency virus*)
HLA	Antígeno Leucocitário Humano (do inglês *Human Leucocitary Antigen*)
HTNV	Vírus Hantaan (do inglês *Hantaan virus*)
IEDB	*Immune Epitope Data Base*
LCMV	Coriomeningite linfocítica (do inglês *Lymphocytic choriomeningitis*)
LMP	Proteína Latente da Membrana (do inglês *Latent Membrane Protein*)
MCMV	Citomegalovirus Murino (do inglês *Murine Cytomegalovirus virus*)
MHC	Complexo Principal de Histocompatibilidade (do inglês *Major Histocompatibility Complex*)
MS	Esclerose Múltipla (do inglês *Multiple Sclerosis*)
PDB	*Protein Data Bank*
PHV	Vírus *Prospect Hill* (do inglês *Prospect Hill virus*)
pMHC	Peptídeo ligado ao MHC (do inglês *petide:MHC*)
PUUV	Vírus Puumala (do inglês *Puumala virus*)
PV	Vírus Pichinde (do inglês *Pichinde virus*)
RMSD	Desvio Quadrático Médio (do inglês *Root Mean Square Deviation*)
RMSF	Flutuação Quadrática Média (do inglês *Root Mean Square Fluctuation*)
RNA	Ácido ribonucleico (do inglês *Ribonucleic acid*)
SNV	Vírus Sin Nombre (do inglês *Sin Nombre virus*)
SPDBv	*Swiss-PDB viewer*
SVM	Máquina de Suporte Vetorial (do inglês *Support Vector Machine*)
TAP	Transportador associado ao Processamento Antigênico (do inglês *Transporter associated with Antigen Processing*)
TCR	Receptor de Célula T (do inglês *T Cell Receptor*)
VV	Vírus Vaccinia (do inglês *Vaccinia virus*)

RESUMO

Diariamente, o organismo humano é desafiado através da invasão de patógenos. No caso da invasão viral, o principal meio de defesa ocorre através do Receptor de Linfócito T (TCR), o qual interage com o peptídeo viral (epitopo) que é apresentado pelo Complexo Principal de Histocompatibilidade de classe I (MHC-I). Dada a imensidão de epitopos possíveis e a limitação do organismo em possuir um único TCR para reconhecer cada antígeno, o sistema imunológico desenvolveu o mecanismo de reatividade cruzada, através do qual um mesmo TCR pode reconhecer diferentes epitopos virais. Nesse trabalho nos propomos a estudar a seqüência da proteína de nucleocapsídeo (proteína N) de diferentes espécies do gênero *Hantavírus* através de uma abordagem *in silico*, utilizando ferramentas de imunoinformática. No primeiro trabalho (capítulo dois) apresentamos uma revisão sobre o uso de ferramentas de imunoinformática na resolução de problemas relacionados à imunogenicidade. No terceiro capítulo, analisamos a seqüência de aminoácidos da proteína N de 34 espécies do gênero *Hantavírus*. Nossos resultados mostraram que epitopos murinos localizam-se preferencialmente em áreas conservadas dentro do alinhamento de sequencias da proteína N. Duas regiões da proteína N – $N_{94\text{-}101}$ e $N_{180\text{-}188}$ - de três espécies virais – Sin Nombre (SNV), Puumala (PUUV) e Hantaan (HTNV) - possuem aminoácidos com grande similaridade físico-química, mas geram respostas imunológicas diferenciadas. Portanto, no quarto capítulo, nosso objetivo foi construir os complexos pMHC referentes às regiões $N_{94\text{-}101}$ e $N_{180\text{-}188}$ da proteína N das espécies SNV, PUUV e HTNV através de uma abordagem *in silico* desenvolvida pelo nosso grupo e analisar as diferenças estruturais e moleculares que possam estar envolvidas na indução da resposta imunológica diferenciada observada *in vitro*. Nossos resultados mostram uma diferença no padrão de cargas entre os complexos pMHC de SNV/PUUV e HTNV referente à região $N_{94\text{-}101}$ e uma diferença na topologia referente à região $N_{180\text{-}188}$. Nosso trabalho foi o primeiro a analisar os mecanismos de reatividade cruzada entre epitopos da proteína N de espécies virais do gênero *Hantavírus* a partir da construção de complexos pMHC partindo da estrutura primária dos epitopos. Ainda, ao final da dissertação, é apresentado um trabalho que visa, através da docagem e dinâmica molecular, diferenciar bons e maus ligantes de MHC-I.

ABSTRACT

The human organism is constantly challenged against several pathogens. In the viral context, the main defense occurs through T Lymphocyte Receptor (TCR) which interacts with the viral peptide (epitope) presented by the Major Histocompatibility Complex of class I (MHC-I). Since there is an infinitude of epitopes and the organism could not be able to accommodate one single TCR for each epitope, the immune system developed the cross-reactivity response, where the same TCR could recognize several viral epitopes. In this work, we studied the nucleocapside (N) protein sequence from several species of the *Hantavirus* genus through an *in silico* approach, using immunoinformatics tools. In the first work (second chapter) we presented a review about the use of immunoinformatics tools to solve issues concerning immunogenicity. In the third chapter, we analyzed the N protein sequence from 34 hantaviruses species. Our results have shown that murine epitopes are preferentially localized over conserved areas of the N protein alignment. Two N protein regions – $N_{94\text{-}101}$ and $N_{180\text{-}188}$ – from three different species – Sin Nombre (SNV), Puumala (PUUV) and Hantaan (HTNV) – have remarkable physical-chemical similarity, however, generated different immunological responses. Based on that, in the fourth chapter we aimed to construct pMHC complexes from $N_{94\text{-}101}$ and $N_{180\text{-}188}$ regions from SNV, PUUV and HTNV, through an *in silico* approach developed by our research group, to analyze structural and molecular differences that would be involved in the differential immunological response induction observed *in vitro*. Our results point to a difference of the charges distribution on the pMHC complexes from SNV/PUUV and HTNV in the $N_{94\text{-}101}$ region and topology differences in the $N_{180\text{-}188}$ region. Our work was the first to analyze the cross-reactivity mechanisms among epitopes of the N protein from hantaviruses species through *in silico* construction of pMHC complexes from linear epitope sequence. Also, at the end, it will be presented a work that uses molecular docking and molecular dynamics, attempting to differentiate good and bad MHC-I ligands.

Capítulo I

Introdução e Objetivos

1. INTRODUÇÃO

1.1. O Sistema Imunológico

1.1.1. Visão geral

Todos os organismos vivos possuem mecanismos de defesa, seja através de estruturas morfológicas diferenciadas, liberação de substâncias tóxicas ou através da ação específica de células e moléculas especializadas. O estudo dessas defesas é inerente ao campo da imunologia. O conceito de imunologia surge com as observações de Elie Metchnikoff, em 1882. Após inserir um acúleo de roseira em uma larva transparente, o pesquisador observou, no decorrer de 24 horas, um acúmulo de células cercando o objeto perfurante. A partir destas observações, Elie Metchnikoff conclui que a larva possui um mecanismo de defesa para proteger-se contra eventuais danos, estendendo essa conclusão a todos os seres vivos (Tan and Dee, 2009).

No estudo do sistema imunológico, podemos dividi-lo, para fins didáticos, em imunidade inata e imunidade adaptativa. Do ponto de vista evolutivo, a imunidade inata foi a primeira a surgir. A imunidade adaptativa surge apenas a partir dos peixes cartilaginosos, na classe dos Chondrichthyes (Kindt *et al.*, 2008).

A imunidade inata é composta basicamente por barreiras físicas, como a pele; químicas, como as mudanças no pH em diferentes compartimentos do corpo; e, biológicas, compreendendo células específicas para defesa do organismo, tais como fagócitos e células "Natural Killer". A resposta imunológica inata é caracterizada por ser rápida, eficaz, pouco específica e sem geração de memória imunológica. A imunidade inata age em compasso com a imunidade adaptativa, a qual é caracterizada por ser uma resposta de longa duração, bastante específica e com o diferencial de que há geração de memória imunológica. A imunidade adaptativa, por sua vez, pode ser dividida em imunidade humoral – cuja resposta é orquestrada por moléculas protéicas denominadas anticorpos - e imunidade celular – cuja resposta é realizada por linfócitos T $CD4^+$ (auxiliares) e linfócitos T $CD8^+$ (citotóxicos).

A capacidade de reconhecimento dos antígenos pelos linfócitos T depende da interação do receptor de células T (TCR, do inglês "T Cell Receptor") com uma molécula

protéica denominada Complexo Principal de Histocompatibilidade (MHC, do inglês "Major Histocompatibility Complex"), traduzida a partir de uma região gênica altamente polimórfica situada no braço curto do cromossomo 6 (posição 6p21.3) (Mungall *et al.*, 2003). Há duas classes principais de MHC: MHC de classe I (MHC-I) e MHC de classe II (MHC-II). O MHC-I é responsável pela apresentação de pequenas seqüências protéicas (epitopos) de 8 a 12 aminoácidos (Rammensee *et al.*, 1999), oriundas do meio intracelular, para linfócitos T $CD8^+$. Se esse epitopo for reconhecido pelo sistema imunológico como sendo pertencente a uma proteína não-própria, uma resposta citotóxica será desencadeada. Já o MHC-II apresenta peptídeos maiores (até 30 aminoácidos), oriundos do meio extracelular, para linfócitos T $CD4^+$, gerando uma resposta auxiliar, ou seja, recrutando mais moléculas e células de defesa para o local da infecção, auxiliando no processo de eliminação da doença. O MHC-I está presente em todas as células nucleadas, enquanto o MHC-II é restrito a um conjunto específico de Células Apresentadoras de Antígenos (APCs, do inglês "Antigen Presenting Cells"). São exemplos de APCs os macrófagos, as células dendríticas e os linfócitos B.

Seguindo o foco do nosso trabalho, será dada atenção especial à resposta imunológica adquirida desencadeada por linfócitos T $CD8^+$ mediante interação com MHC-I. Para tanto, é necessário entender o processo como um todo, desde a invasão do patógeno no ambiente intracelular, até a apresentação de peptídeos oriundos de suas proteínas na superfície celular.

1.1.2. Rota de Apresentação de Antígenos Endógenos

Ao invadir o ambiente intracelular, o patógeno fica protegido contra o ataque de anticorpos (resposta humoral), os quais possuem ação neutralizante apenas no ambiente extracelular. Sendo assim, um segundo mecanismo da resposta imunológica entra em ação: a resposta celular mediada por linfócitos T $CD8^+$. Mas um longo caminho deve ser percorrido antes que os peptídeos oriundos do patógeno sejam apresentados na superfície celular pelas moléculas de MHC-I.

O primeiro passo envolve a marcação das proteínas virais através do processo de ubiquitinação. A ubiquitina é uma proteína formada por aproximadamente 76 aminoácidos

e altamente conservada entre eucariotos (Nandi *et al.*, 2006). O processo de ubiquitinação é realizado através da ação de uma cascata enzimática, mediada pelas enzimas E1, E2 e E3, às quais adicionam várias ubiquitinas na proteína alvo (poliubiquitinação) (Ciechanover, 1994).

O segundo passo envolve a degradação das proteínas poliubiquitinadas por um complexo multienzimático denominado proteassomo. O proteassomo é formado por um complexo denominado 26S, o qual é subdividido em complexo 20S – central e responsável pela atividade proteolítica – e duas partículas periféricas regulatórias denominadas 19S, as quais controlam o reconhecimento das proteínas poliubiquitinadas, o desdobramento protéico dependente de ATP e a abertura do canal do complexo 20S, permitindo a entrada da proteína no sítio catalítico (Wang and Maldonado, 2006). É importante ressaltar que a extremidade C-terminal do epitopo, o qual será apresentado pelo MHC-I, é gerada durante o corte do proteassomo. Entretanto, o mesmo não ocorre para a extremidade N-terminal, a qual sofrerá um desbastamento por aminopeptidases específicas presentes no citoplasma e no retículo endoplasmático, antes da ligação ao MHC-I (Falk and Rotzschke, 2002).

Apenas uma pequena parte dos peptídeos gerados através da clivagem pelo proteassomo será transportada para o lúmen do retículo endoplasmático rugoso. Esse transporte é mediado por uma complexa maquinaria protéica associada à apresentação de antígenos (TAP, do inglês "Transporter associated with Antigen Processing"). A TAP é uma proteína homodimérica transmembrana do retículo endoplasmático rugoso. Os peptídeos transportados pela TAP possuem tamanhos bastante variáveis, entre 8 e 40 aminoácidos (Schmitt and Tampe, 2000). Quanto a sua especificidade, TAP parece favorecer peptídeos cuja extremidade C-terminal possua aminoácidos hidrofóbicos ou básicos (Kindt *et al.*, 2008).

O último passo antes da apresentação do peptídeo na superfície celular envolve a interação do epitopo com o MHC-I. O MHC-I é formado por duas cadeias polipeptídicas conectadas uma a outra de forma não-covalente, sendo uma cadeia pesada (cadeia alfa) e uma cadeia invariável denominada beta 2-microglobulina. O segmento N terminal da cadeia alfa (alfa 1 e alfa 2) interage formando uma fenda, a qual é composta por uma plataforma de aproximadamente oito beta-folhas e um par de alfa-hélices paralelas, lateralmente dispostas (Figura 1). Esse espaço é suficiente para que o MHC-I consiga

acomodar peptídeos de 8 a 12 aminoácidos de extensão (Hammer *et al.*, 2007). Ainda, é importante ressaltar que um mesmo MHC-I pode se ligar a diferentes peptídeos, com diferentes especificidades, e vice-versa. Os peptídeos são ancorados a "pockets" dentro da fenda do MHC-I em determinadas posições, denominadas posições de ancoragem. Cada alelo de MHC-I possui seu próprio conjunto de "pockets", os quais possuem afinidade por diferentes resíduos. Uma vez que o peptídeo esteja ligado à fenda de MHC-I, esse complexo se torna estável o suficiente para que seja transportado para a superfície celular, cumprindo o seu papel na apresentação de antígenos intracelulares. Cabe salientar que proteínas próprias (isto é, pertencentes ao oragnismo humano) também passam pela via de processamento antigênico.

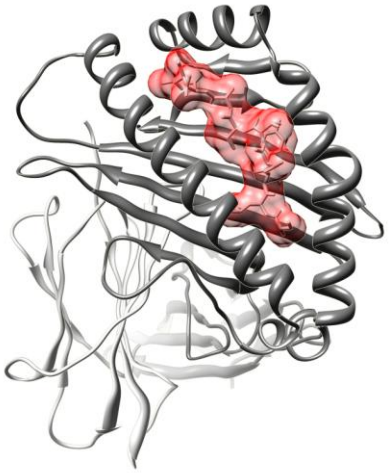

Figura 1. Representação do MHC-I (cinza) interagindo com um peptídeo (epitopo). Em vermelho, a superfície energética gerada pelo epitopo na fenda do MHC.

Uma vez na superfície celular, o complexo peptídeo:MHC-I (pMHC) passará pelo processo de reconhecimento mediado pelo TCR. O TCR possui uma estrutura heterodimérica composta por uma cadeia alfa e uma cadeia beta ou, alternativamente, uma cadeia gama e uma cadeia delta. Cada uma dessas cadeias possui três sítios hipervariáveis, também conhecidos como Regiões Determinantes de Complementaridade (CDRs, do

inglês “Complementariety Determinant Regions”), que protraem como alças do TCR e realizam contato direto com os aminoácidos centrais do peptídeo ligado ao MHC-I (Bjorkman, 1997). Se nesse contexto, o peptídeo for reconhecido como algo não-próprio, ocorrerá a liberação de substâncias tóxicas do linfócito T, levando à morte por apoptose da célula infectada.

Em teoria, há uma infinidade de peptídeos que podem ser apresentados pelo MHC-I. Dessa forma, não podemos imaginar um sistema onde exista apenas um TCR capaz de reconhecer cada complexo pMHC possível. Sendo assim, é de se esperar que o sistema imunológico tenha desenvolvido uma maneira de sobrepor essa carência numérica. Atualmente, sabe-se que um mesmo TCR possui a habilidade de reconhecer diferentes peptídeos ligados ao mesmo MHC-I. Na verdade, estima-se que um mesmo TCR tem o potencial de interagir com aproximadamente 10^6 complexos pMHC diferentes (Mason, 1998). Essa promiscuidade na interação TCR:pMHC é denominada reatividade cruzada, e será abordada no próximo tópico.

1.1.3. Reatividade Cruzada

Um mecanismo de reconhecida importância no contexto evolutivo do sistema imunológico é denominado reatividade cruzada, o qual consiste na capacidade de um mesmo TCR reconhecer peptídeos não relacionados derivados de um mesmo vírus, ou oriundos de vírus heterólogos (Snyder *et al.*, 2004; Williams *et al.*, 2002). Por isso, a reatividade cruzada é também conhecida como imunidade heteróloga (Welsh and Selin, 2002). Esse fenômeno é observado principalmente envolvendo linfócitos T citotóxicos, embora também ocorra com linfócitos T auxiliares (Selin *et al.*, 1994), anticorpos e, com menor especificidade, macrófagos (Welsh *et al.*, 2010).

Do ponto de vista clínico, o decorrer da infecção em um indivíduo infectado por um determinado patógeno pode ser muito menos severo se o mesmo possuir imunidade heteróloga contra o agente infeccioso (Selin *et al.*, 1998). A maior pandemia do século XXI, causada pela cepa recombinante do vírus Influenza A(H1N1)-2009, serve como exemplo de reatividade cruzada contra epitopos virais. Estudos mostram que indivíduos imunizados antes da década de 50 contra uma cepa do vírus Influenza apresentaram

imunidade heteróloga contra a cepa pandêmica A(H1N1)-2009. A maior parte dos indivíduos acima de 60 anos que entraram em contato com a cepa A(H1N1)-2009 não apresentou sintomatologia, visto que os mesmos já possuíam células e moléculas de defesa contra o agente infeccioso (Skountzou *et al.*, 2010). Além disso, observou-se que muitos dos peptídeos derivados da cepa A(H1N1)-2009 são representativos da catastrófica gripe espanhola de 1918 (Gras *et al.*, 2010).

Um trabalho importante sobre reatividade cruzada de Selin e cols (1998) demonstrou que a memória imunológica do indivíduo pode ser moldada através da infecção por diferentes agentes virais não relacionados. Foram analisados quatro vírus heterólogos – vírus da coriomeningite linfocítica (LCMV), citomegalovírus murino (MCMV), vírus pichinde (PV) e vírus da vaccinia (VV). A reatividade cruzada foi observada principalmente entre os vírus LCMV e VV e entre LCMV e PV, sendo que a intensidade da resposta variou de acordo com a ordem de imunização (Selin *et al.*, 1998).

A ocorrência de reatividade cruzada tanto ao nível do reconhecimento de células T (TCR), quanto na ligação do peptídeo ao MHC dá idéia de um sistema imunológico dinâmico, que apesar de apresentar degeneração no reconhecimento de antígenos, possui evidente eficácia na eliminação de infecções virais. Para averiguar e analisar os dados oriundos de experimentos *in vivo*/*in vitro*, não apenas de reatividade cruzada, mas referentes à imunologia como um todo, um novo ramo da bioinformática surge: a imunoinformática.

1.2. Imunoinformática

1.2.1. Visão geral

O número de experimentos *in vivo*/*in vitro* tem aumentado vertiginosamente em diversas áreas da ciência, sendo que na imunologia isso não é diferente, trazendo à tona algumas questões: como lidar com esse enorme volume de dados? O que podemos inferir a partir deles? Como organizar e realizar novos experimentos a partir dos dados gerados? Para auxiliar na resolução dessas questões, surge a imunoinformática, como um novo ramo da bioinformática, voltada para a aceleração da pesquisa na área de imunologia. Aos

experimentos realizados com base na imunoinformática, assim como na bioinformática, damos o nome de experimentos *in silico*.

Historicamente, a imunoinformática começou no início do século XX a partir da modelagem epidemiológica da malária (Ross, 1916). Naquele tempo, a ênfase era dada no uso da matemática para guiar o estudo da transmissão da doença. Desde então, o campo da imunoinformática expandiu para cobrir outros aspectos inerentes ao processo imunológico como um todo. Com o avanço da tecnologia computacional, aliado ao maior número de dados experimentais publicados, surgiram novos programas na área da imunoinformática para realizar estudos *in silico*. Um novo termo foi criado, a imunômica, o qual combina a imunologia tradicional com a ciência da computação, matemática, física, química, bioquímica, genômica e proteômica, visando analisar e entender a função do sistema imune em larga escala (Tong and Ren, 2009).

Como este trabalho se foca na resposta imunológica por linfócitos T $CD8^+$ frente à apresentação de peptídeos por MHC de classe I, daremos ênfase aos programas de imunoinformática que se relacionam com essa área do conhecimento.

1.2.2. Programas Preditores de Epitopos de MHC de classe I

Dentre os principais programas que simulam etapas da rota de apresentação de antígenos podemos citar os seguintes: Netchop (Nielsen *et al.*, 2005), TAPPred (Bhasin and Raghava, 2004), SYFPEITHI (Rammensee *et al.*, 1999), NetMHC (Lundegaard *et al.*, 2008), IEDB (Peters *et al.*, 2005) e EpiJen (Doytchinova *et al.*, 2006). Todos esses programas utilizam algoritmos avançados que foram desenvolvidos com base nos dados oriundos de experimentos *in vitro* e/ou *in vivo*.

O Netchop é um programa que prediz os cortes realizados pelo proteassomo na extremidade C-terminal das proteínas. O método utilizado pelo programa se baseia em uma combinação de várias redes neurais, cada uma treinada sobre um banco de dados montado a partir de dados experimentais *in vitro*. Após submeter a sequência de uma proteína no formato FASTA (formato de texto onde cada aminoácido da seqüência é representado por uma letra), o programa retorna um valor, variando de 0 (valor mínimo) a 1 (valor máximo),

o qual representa a probabilidade de que o proteassomo clive a proteína naquela posição específica.

Como foi mencionado na seção anterior, assim que a proteína viral passa pelo proteassomo ela é clivada em pequenos peptídeos (variando de 8 a 30 aminoácidos). O próximo passo, portanto, consiste no transporte de parte desses peptídeos para dentro do retículo endoplasmático rugoso. Esse transporte é realizado pela proteína TAP, a qual possui atividade diferenciada dependendo do peptídeo que se liga a ela. Baseado no fato de que determinados peptídeos serão favorecidos em prol de outros, foi desenvolvido o programa TAPPred. Este programa foi construído sobre uma matriz quantitativa baseada em SVM (do inglês "Support Vector Machine") e é utilizado para predizer a probabilidade de que um determinado peptídeo seja transportado, ou não, para dentro do retículo endoplasmático. Diferente do Netchop, o TAPPred retorna um valor que o próprio programa interpreta, categorizando os peptídeos como tendo probabilidade baixa, média ou alta de serem transportados. O TAPPred leva em consideração não apenas o tamanho dos aminoácidos, mas também o grupo físico-químico ao qual eles pertencem.

Uma vez que os peptídeos foram transportados para dentro do retículo endoplasmático, apenas alguns terão afinidade suficiente para se ligar ao MHC. A fim de predizer o grau de afinidade de peptídeos pelo MHC, foram criados diversos programas, sendo que o SYFPEITHI e NetMHC são os que se destacam. O SYFPEITHI possui um amplo banco de dados oriundo de dados experimentais descritos na literatura sobre ligação de peptídeos aos seus respectivos MHCs. Baseado na posição dos aminoácidos âncora e de aminoácidos auxiliares, ele computa um valor para afinidade de ligação do peptídeo, para diversos alelos de MHC. Já o NetMHC, diferentemente do SYFPEITHI, prediz um valor de ligação dos peptídeos ao MHC através de algoritmos baseados em redes neurais artificiais e matrizes ponderadas.

Obviamente, assim como na ciência, a imunoinformática não poderia deixar de ser um campo interdisciplinar. Há ferramentas utilizadas para resolver questões referentes a dados genômicos, por exemplo, que podem ser bem empregadas no dia-a-dia de um imunoinformata. É o caso do alinhamento múltiplo de seqüências, ferramenta muito utilizada para evidenciar padrões de conservação protéica que não poderiam ser distinguidos pela simples análise textual.

Todas essas abordagens são válidas e importantes para solucionar os desafios impostos pelo conhecimento imunológico-científico acumulado. Ainda, esforços estão sendo empregados para análises estruturais do complexo pMHC apresentado na superfície celular, já que a análise da topologia e distribuição de cargas, por exemplo, é de grande importância para inferir a imunogenicidade de um determinado peptídeo (Kessels *et al.*, 2004).

1.2.3. Análise Estrutural de complexos pMHC

Uma abordagem que tem sido empregada para análise estrutural de complexos pMHC e a sua interação com linfócitos T é a cristalografia de proteínas (Chen *et al.*, 2009). Esta técnica fornece um entendimento em nível molecular das interações que estão ocorrendo tanto entre o peptídeo e o MHC, bem como entre o pMHC e o TCR. Entretanto, a cristalografia de proteínas é uma técnica cara, morosa e complexa, sendo que nem sempre é possível obter os complexos desejados devido à instabilidade exacerbada em determinadas regiões da proteína (principalmente regiões de alça).

Uma alternativa tem sido a utilização de programas que realizam docagem (Trott and Olson, 2010) e dinâmica molecular (Van Der Spoel *et al.*, 2005) de complexos pMHC (Rognan *et al.*, 1994; Sieker *et al.*, 2007; Zacharias and Springer, 2004). O primeiro tem a capacidade de procurar por sítios de interação entre ligantes (no nosso caso, o peptídeo) e seus respectivos receptores (como o MHC, por exemplo) baseado num índice de afinidade. Já a dinâmica molecular, como o próprio nome sugere, fornece dados sobre a dinâmica da proteína e seu ligante em um determinado ambiente (vácuo, água, óleo, entre outros) ao longo do tempo. A partir dessas técnicas, é possível fazer uma análise da estrutura no seu estado tridimensional, apontando interações e padrões conformacionais que não poderiam ser obtidos pela simples análise linear de seqüência.

Devido ao baixo número de complexos pMHC disponíveis no Banco de Dados de Proteínas (PDB, do inglês “Protein Data Bank”), principal banco de dados de estruturas cristalografadas, nosso grupo desenvolveu uma técnica baseada em docagem e dinâmica molecular para construir complexos pMHC a partir da seqüência FASTA de peptídeos (Antunes *et al.*, 2010). Esta técnica foi validada a partir da reprodução de todas as estruturas depositadas no PDB para três diferentes alelos de MHC (46 estruturas

englobando os alelos HLA-A*0201, H2-K^b e H2-D^b)[1]. Como nosso estudo está focado na resposta imunológica, utilizamos esta técnica para reproduzir complexos tridimensionais de moléculas de MHC apresentando peptídeos oriundos de patógenos virais. No caso deste trabalho, trabalhamos com a família viral *Bunyaviridae*, mais especificamente com o gênero *Hantavírus*.

1.3. Gênero *Hantavírus*

1.3.1. Visão geral

O gênero *Hantavírus* pertence à família Bunyaviridae. Essa família compreende mais de 300 espécies virais, distribuídas entre de cinco gêneros (Karabatsos, 1985). Enquanto a grande maioria das espécies virais pertencentes a essa família infectam, particularmente, membros da classe Arthropoda, os hantavírus possuem a peculiaridade de infectar membros da ordem Rodentia, família Muridae (subfamília Murinae) e família Cricetidae (subfamílias Arvicolinae, Sigmodontinae e Neotominae) (Lee *et al.*, 1978; Mir, 2010) e, ocasionalmente, humanos. Em humanos, os hantavírus podem desencadear a Síndrome Renal com Febre Hemorrágica (HFRS, do inglês "Hemorrhagic Fever with Renal Syndrome") ou a Síndrome Cardiopulmonar (HCPS, do inglês "Hantavirus Cardiopulmonary Syndrome"), com taxas de mortalidade bastante elevadas (Guilfoyle and Macnab, 2008). A HFRS e a HCPS possuem uma distribuição geográfica característica. A HFRS ocorre preponderantemente na Europa, África e Ásia, sendo que os casos relatados na África são bastante raros, não tendo nenhum hantavírus isolado nessa região até o momento. Já a HCPS ocorre mais comumente nas Américas (Nichol *et al.*, 1993; Schmaljohn *et al.*, 1987).

Os hantavírus, isolados pela primeira vez em 1978, são vírus envelopados cujo tamanho varia de 80 a 140 nanômetros (Lee *et al.*, 1978). Seu genoma consiste em três segmentos de RNA de polaridade negativa: L, M e S (Schmaljohn, 1996). O segmento L codifica uma RNA polimerase dependente de RNA; o segmento M codifica duas glicoproteínas de membrana (Gn e Gc); e, o segmento S, codifica a proteína de nucleocapsídeo, também conhecida como proteína N (Jonsson and Schmaljohn, 2001),

[1] Foram utilizados apenas complexos pMHC cujo epitopo apresentava uma padrão alelo-específico, como apresentado no trabalho de Antunes e cols (2010).

além de uma proteína não estrutural, denominada NsS (Figura 2). A proteína N, por ser o foco do nosso trabalho, será abordada com maiores detalhes.

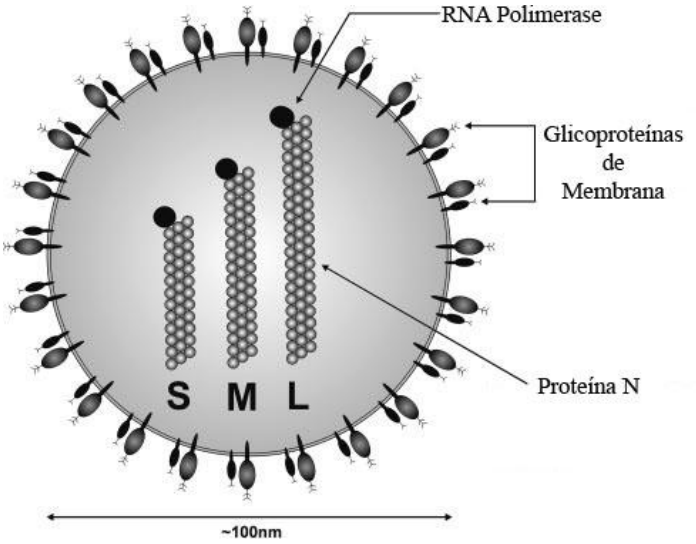

Figura 2. Figura esquemática do vírion do gênero *Hantavírus*. Os segmentos S, M e L codificam a proteína N, as glicoproteínas de membrana (Gn e Gc) e a RNA Polimerase, respectivamente. Modificado de Muranyi e cols. 2005.

1.3.2. Proteína N

Os segmentos de RNA viral são revestidos pela proteína N, formando os segmentos individuais denominados L, M e S (Figueiredo *et al.*, 2008). A proteína N é sintetizada logo no início da infecção, se mantendo abundante durante todo curso da doença (Elliott, 2000). Sua importância para o vírus se reflete no fato de que ela age protegendo o RNA da degradação por nucleases e, em elevadas quantidades, ativa a replicação genômica do vírus (Mir and Panganiban, 2006). Além disso, a proteína N interage diretamente com a extremidade citoplasmática da glicoproteína de superfície Gn, iniciando a montagem viral (Hepojoki *et al.*, 2010). Também já foi observado que a proteína N de alguns hantavírus interage com algumas proteínas da célula hospedeira procurando impedir mecanismos antivirais da mesma (Khaiboullina *et al.*, 2005; Li *et al.*, 2002; Maeda *et al.*, 2003).

Por ser um componente viral interno, a proteína N sofre uma menor pressão seletiva do sistema imunológico, se comparada com as glicoproteínas de membrana (Tischler *et al.*, 2008), o que pode explicar a sua elevada conservação estrutural (Plyusnin *et al.*, 1996).

Além disso, já foram observados vários eventos de reatividade cruzada entre epitopos oriundos da proteína N de diferentes espécies de hantavírus, como será abordado adiante (Asada *et al.*, 1988; Asada *et al.*, 1989; Ennis *et al.*, 1997; Maeda *et al.*, 2004; Van Epps *et al.*, 1999).

1.3.3. Evidências de Reatividade Cruzada em Hantavírus

Em um trabalho pioneiro, Asada e cols (1988) infectaram um grupo de camundongos com a cepa B-1 do hantavírus Seoul e um segundo grupo com a cepa 76-118 do hantavírus Hantaan. Observou-se que os esplenócitos do primeiro grupo, quando desafiados *in vitro* com a cepa 76-118 de Hantaan, foram capazes de se diferenciar, cumprindo sua função efetora. O mesmo ocorreu quando o caminho inverso foi realizado, ou seja, desafiando esplenócitos do segundo grupo com a cepa B-1 de Seoul. No segundo caso, entretanto, foi observada uma resposta mais fraca (Asada *et al.*, 1988). Esse trabalho demonstrou de forma inédita, que a reatividade cruzada, por imunização heteróloga de diferentes cepas virais, pode ocorrer em nível celular nos hantavírus.

Em um segundo trabalho realizado pelo mesmo grupo (Asada *et al.*, 1989), o tema da reatividade cruzada foi novamente abordado, mas dessa vez utilizando outras cepas de hantavírus. Esplenócitos de camundongos BALB/c, imunizados com o hantavírus Prospect Hill (PHV) ou Puumala (PUUV), foram desafiados *in vitro* contra células apresentadoras de antígenos pulsadas com a cepa viral 76-118 de Hantaan. Observou-se geração de resposta imunológica cruzada entre a cepa de Hantaan e as cepas de PHV e PUUV. Ainda, realizou-se a transferência de soro contendo anticorpos e linfócitos T dos camundongos imunizados com PHV e PUUV para dois grupos de camundongos "nude", os quais foram desafiados no dia posterior com a inoculação da cepa 76-118 de Hantaan. Observou-se a ocorrência da resposta celular cruzada contra Hantaan em ambos os grupos de camundongos. Entretanto, a resposta humoral cruzada só se mostrou eficaz contra o grupo imunizado com PUUV. Este estudo não apenas realça a importância da reatividade imunológica cruzada contra infecção por hantavírus, como aponta o mecanismo de resposta celular cruzada como mecanismo importante para a sua defesa.

No estudo de Van Epps e cols (1999) foram realizados estudos de reatividade cruzada utilizando o soro de três pacientes infectados pelo hantavírus Hantaan. Observou-se que em dois de três pacientes os linfócitos T foram capazes de se diferenciar em sua função efetora ao reconhecer peptídeos da proteína N. Linfócitos dos pacientes infectados foram desafiados com os peptídeos reconhecidos durante a infecção de diferentes espécies de hantavírus. Ao fim, foi constatado que a proteína N das espécies de hantavírus Hantaan, Puumala e Seoul é capaz de estimular reatividade cruzada entre diferentes linhagens celulares de linfócitos T $CD8^+$.

Um trabalho mais atual e bastante elucidativo no campo da reatividade imunológica cruzada entre diferentes espécies de hantavírus foi realizado por Maeda e cols (2004). Um dos focos do trabalho foi identificar diferentes epitopos oriundos da proteína N de três espécies de hantavírus - Sin Nombre (SNV), Hantaan (HTNV) e Puumala (PUUV) – e averiguar se esses epitopos eram capazes de gerar uma reposta imunológica cruzada. Duas regiões da proteína N chamam a atenção – região 91-105 e 175-189. Ambas as regiões de SNV e PUUV apresentam epitopos, com forte geração de resposta imunológica, reconhecidos pelo mesmo "pool" de linfócitos T $CD8^+$ (Tabela 1). Entretanto, nas mesmas regiões da proteína N, a espécie de HTNV não gerou nenhuma resposta imunológica, apesar de apresentar elevada similaridade de seqüência com a região de SNV e PUUV.

Tabela 1. Epitopos identificados por Maeda e cols (2004)

Região da Proteína N	Vírus	Epitopo	Geração de resposta
$\mathbf{N_{91\text{-}105}}$	SNV	^{94}SSLRYGNV^{101}	SIM
	PUUV	^{94}SSLRYGNV^{101}	SIM
	HTNV	^{94}S<u>M</u>L<u>S</u>YGNV^{101}	NÃO
$\mathbf{N_{175\text{-}189}}$	SNV	^{180}SMPTAQSTM^{188}	SIM
	PUUV	^{180}SMPTAQSTM^{188}	SIM
	HTNV	^{180}S<u>L</u>P<u>N</u>AQS<u>S</u>M^{188}	NÃO

Tendo como enfoque principal o gênero *Hantavírus* e o uso da imunoinformática para auxiliar na elucidação de questões de reatividade imunológica cruzada, traçamos os objetivos deste trabalho como segue.

1.4. Objetivos

1.4.1. Objetivos Gerais

Analisar os elementos que influenciam a imunogenicidade em diferentes regiões da proteína N do gênero *Hantavírus* tanto no âmbito da análise de estrutura primária de seqüências bem como em relação à estrutura terciária da proteína.

1.4.2. Objetivos Específicos

1) Procurar por epitopos, descritos na literatura e reconhecidamente imunogênicos, específicos da proteína N e sobrepô-los ao alinhamento das diferentes espécies do gênero *Hantavírus*, buscando averiguar relações com áreas de conservação da proteína.

2) Construir e analisar complexos pMHC referentes às regiões da proteína N de hantavírus (N_{94-101} e $N_{180-188}$) para as seqüências das espécies Puumala (PUUV), Sin Nombre (SNV) e Hantaan (HTNV).

3) Procurar identificar padrões moleculares específicos que possam ser os responsáveis pela resposta imunológica diferenciada entre os complexos pMHC.

4) Computar a energia total do sistema e analisar a distribuição de cargas na superfície dos complexos pMHC.

5) Analisar o padrão de topologia entre os complexos pMHC gerados.

Capítulo II

Molecular Aspects Involved in the Immunogenicity against Viral Epitopes - an Immunoinformatic Perspective

2. MOLECULAR ASPECTS INVOLVED IN THE IMMUNOGENICITY AGAINST VIRAL EPITOPES - AN IMMUNOINFORMATIC PERSPECTIVE

CAPÍTULO DO LIVRO INTITULADO *IMMUNOGENICITY* (ISBN: 978-1-61761-591-7), DA SÉRIE *IMMUNOLOGY AND IMMUNE SYSTEM DISORDERS*, PUBLICADO PELA EDITORA *NOVA PUBLISHERS*
(EDITOR: Christian J. Villanueva)

Chapter 1

Molecular Aspects Involved in the Immunogenicity against Viral Epitopes: An Immunoinformatic Perspective

Maurício Menegatti Rigo[1], Dinler Amaral Antunes[1], Marialva Sinigaglia[1], Cassiana Chassot Fülber[2], José Artur Bogo Chies[1] and Gustavo Fioravanti Vieira[1]

[1] Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil Bento Gonçalves Avenue, 9500 - Build 43323 - Room 225 - ZIP Code: 91501-970.

[2] Informatic Institute, Federal University of Rio Grande do Sul, Porto Alegre, Brazil Bento Gonçalves Avenue, 9500 - Build 43323 - Room 225 - ZIP Code: 91501-970.

Abstract

Since the beginning of studies about immune response stimulation against viral infections, many pieces have already been added to complete this puzzle. Once uncovered, it will provide the key elements to the development of more efficient vaccines against a wide diversity of infectious diseases. A thorough understanding of this system involves not only the analysis of epitopes that are presented to the immune system but also the elucidation of the steps required for the generation and selection of immunogenic epitopes, such as the proteolysis mediated by proteasomes, transport associated with antigen processing (carried out by TAP protein) and binding of epitopes to the Major Histocompatibility Complex (MHC). Besides the steps in the antigen intracellular processing pathway, other aspects involved in the selection of immunogenic epitopes, such as its intrinsic avidity by the MHC cleft (on the extracellular surface) and the epitope constitution , are necessary to stimulate the T cell receptor (TCR). Certainly, the epitopes linear sequences are important; however, an epitope is much more than a simple combination of amino acids, which present a distribution of charges and topological features that actively interact with the other elements that form the MHC cleft. These elements will be essential for the induction (or not) of an immune response. In this chapter, we will discuss several aspects involved in the recognition of a viral infection. What are the typical elements of a viral infection, and how does the immune system recognize them? How could different amino acid sequences stimulate the same TCR generating cross-reactive responses? In which way can a virus infection trigger an autoimmune disease? Allied to all these immunological approaches, we intend to present how bioinformatics—or more specifically immunoinformatics—tools can be used to elucidate these processes. Thus, several tools, such as docking and molecular dynamics, peptide:MHC binding and proteasome cleavage predictors will be presented. We will also present and discuss the main molecular immunological databases, including SYFPEITHI, IEDB and IMGT. In summary, we will discuss how population, immunological, and molecular aspects are essential in the comprehension of vaccine design and autoimmunity.

1. Introduction

A close interaction characterizes the relationship between humans and the environment. This interaction is much more intimate than the simple contact with the environment and its compounds, since it applies to the body composition itself, i.e. in addition to our own cells, our body contains and interacts with several other organisms. Data obtained from the intestinal microbiota, for example, suggest that each human being carries, only at this compartment, a set of as much as 400 different bacterial species. This intimate contact between microorganisms and our bodies constantly shapes the immune system and drives our decisions leading to tolerance or to immune responses. Certainly, humans are aware that these interactions and attempts to interfere in the balance that controls immune responses are not a novelty.

An elucidative example of attempts to modulate the immune response is given by the history of smallpox. The Chinese, for example, long before the 1700s used "variolation". They would deliberately infect individuals with smallpox through the inhalation of a preparation composed of dried smallpox scabs. This procedure led to the development of a mild disease (in most cases) and, contrary to persons not subject to variolation, among which the mortality of the naturally contracted disease reached 30%, only 1% to 2% of the variolated individuals died if they later contracted the disease. However, it was only after Edward Jenner's experiments, published in 1798, that the procedures of human interference in the immune system, through vaccination, had become effective.

Interestingly, the work of Jenner not only marks the beginning of a period where modulation of the immune system becomes a reality (despite the almost complete ignorance of the processes that control this modulation at the time, and even recently) but it is also characterized by the induction of an immune response through cross-reactivity. Actually, Jenner managed to induce resistance to smallpox by inoculating cowpox in the subjects. The fact that Jenner used a related pathogen to induce reactivity against another pathogen remained virtually forgotten for a long time. Nevertheless, this situation, much more than an academic curiosity, is extremely important and the effectiveness of such a cross-reactive vaccine gives us important insights into the functioning of the immune system. First, the induction of such a protective immune response via cross-reactivity shows that different pathogens share features, and that these shared features can be good targets for the immune system, even acting as immunodominant viral determinants. Second, the existence of shared characteristics among different pathogens also suggests that they are not able to evade the immune response by altering these epitopes (or such targets would be negatively selected through the pathogen evolution), indicating the existence of some phenomena that restrict pathogen evolution. From these two characteristics we can consider strategies aimed at the development of wide spectrum vaccines (see Vieira and Chies, 2005). Another important feature of the immune system is that, though often described as a lock and key system, antigen recognition by antibodies is rather promiscuous, with several antigens being recognized by a unique antibody and conversely several antibodies recognizing the same antigen (although different antibodies may have different affinities to the same antigen). This same reasoning applies to the recognition of peptide:MHC complexes by T cell receptors (TCR).

In the present chapter we will discuss how these characteristics, namely cross-reactivity and promiscuous antigen recognition, affect the development of an immune response, specifically in the context of a viral infection. Additionally, we intend to discuss the different steps in the antigen processing pathway which allows an epitope to be properly presented to T cells by a MHC molecule on the cell surface. We will also present how immunoinformatics tools could be used to elucidate these processes. Finally, we will discuss how the comprehension of the cross-reactivity and of the promiscuous antigen recognition phenomena in the immune system can help in the development of wide spectrum viral vaccines.

2. Endogenous Antigen Processing Pathway

Through the antigen processing pathway, self and non-self antigens are presented for immune system cells, which will or will not trigger an immune response. There are independent ways for the presentation of exogenous and endogenous antigens, each one causing a different profile of immune response. In this section we will focus on endogenous antigens presentation pathway, which occurs in all nucleated cells of the vertebrates, and cellular immune response.

Five different aspects will be highlighted.

2.1. Ubiquitin-Proteasome System (UPS)

Ubiquitin is a protein formed by 76 amino acids with very high conservation among eukaryotes [1]. The selective binding of ubiquitin to a determined protein is the first sign of this protein direction to degradation. This mechanism is extremely important in the removal of proteins, denatured or even improperly translated, as well as in cyclins levels and transcription factors regulation [2]. The ubiquitination process is performed by an enzymatic cascade, mediated by enzymes E1 (ubiquitin-activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin-protein ligase). First, an ubiquitin molecule is activated by binding to E1 enzyme, involving ATP consumption. The second step involves transferring the recently bound ubiquitin to E1 enzyme to one of E2 forms. Subsequently, occurs the addition of ubiquitin to the target protein, and this reaction is catalyzed by one of E3 forms. This process is repeated until a polyubiquitin chain is formed and connected to the target protein, which will be posteriorly recognized by the proteasome [3]. Since ubiquitins can be removed with the assistance of *de-ubiquitylating enzymes* (such as A20), this process is reversible [4].

Degradation of proteins marked with ubiquitin is performed by a multienzymatic complex, called proteasome. The proteasome is formed by a complex named 26S. This is subdivided in another complex, 20S, central, responsible for proteasome proteolitic activity, and two periferic regulatory particles 19S, which control ubiquitined proteins recognition, the ATP-dependent unfolding and the opening of the channel in the 20S proteasome that allows entry into the proteolytic chamber [5]. Proteasome's cleavage will generate accurately the carboxy-terminal (C-term) extremity for epitopes presented by MHC-I (discussed in subsection 1.3), having a notable agreement among proteasome, TAP (discussed in subsection 1.2) and MHC in terms of C-term affinity by an amino acid (or an physical-chemical group). This agreement is likely to be resultant of the co-evolution among fundamental steps in the same pathway [6]. The final N-term extremity, however, is not generated by the proteasome, and the peptides transported by TAP will suffer an amino-terminal trimming – by Endoplasmic Reticulum (ER) proteases such as ERAP (ER Amino Peptidase) – before being finally loaded in the MHC-I nascent chain [7].

The process of protein degradation by the proteasome occurs constitutively, generating peptides arising from self proteins, as well as proteins arising from pathogens, in case the cell is infected [8]. In viral infection context, the continuous presence of pathogen induces, by a refined event that involves regulatory mechanisms not only of the own cell, but also of other immune system's cells, the production of interferon gamma cytokine which induces the association of the PA28 regulatory particle to the 19S regulatory subunit [9]. Besides, LMP2, LMP7 and LMP10 expression, encoded by genes in MHC region, is also induced [10]. These proteins association, as well as PA28 subunit, leads to a variation in proteasome's catalytic specificity, advantaging the generation of appropriate and smaller peptides, which will bind to MHC-I posteriorly [11].

Peptides generated by the proteasome will have different destinations, some of them being degraded in cytosol by aminopeptidases or translocated to the ER via TAP. The peptides transportation process from cytosol to ER involves a complex protein machinery, which will be discussed in next section.

2.2. TAP Transport

The TAP (Transporter associated with Antigen Processing) is composed of an endoplasmic reticulum transmembrane heterodimer of homologous TAP1 and TAP2 subunits, and each subunit posses a domain directed to ER interior and an ATP binding domain that projects itself to the cytosol. Both TAP1 and TAP2 belong to a family of transmembrane proteins that bind to ATP (ABC transporters), found in a variety of cells, including bacterias [12]. Their function is to transport peptides generated in the cytosol into ER lumen. Peptides transported by TAP can vary from 8 to 40 amino acids, which is why commonly this protein is referenced as promiscuous among the members of ABC transporters family [13]. TAP is also capable of recognizing a great variety of substrates, which leads to the quest of how a protein can deal with a wide range of ligands without losing its affinity, specificity or efficiency [14]. Considering specificity, TAP seems to favor peptides with hidrofobic or basic carboxyterminal aminoacids [10], anchors preferred by MHC, as will be seen in next item.

In each TAP C-term extremity, the NBD (Nucleotide Binding Domain) is present, which drives translocation process. In NBD's C-term extremity, there are three motifs (Walker A, signature and Walker B), responsible for ATP binding and hydrolysis [15]. These motifs are highly conserved among a great variety of organisms. How the two NBDs domains communicate with each other and at which stage substrate transport takes place remains unclear.

2.3. MHC Binding

The MHC was discovered as an extense *locus* with highly polymorphic genes, which determines the results of organ transplant. Despite the mistaken name, inheritance of the context on which it was discovered, the physiological function of this molecule has no relation to histocompatibility, being responsible for peptides presentation in cell surface. MHC locus still harbors other genes involved in the generation of this protein segments, highlighting proteasome and TAP coding sequences. MHC genes can still be subdivided in at least two main groups – MHC Class I and MHC Class II – which encode proteins structurally different, and yet homologous. MHC Class II molecules are involved in the presentation of exogenous peptides to recognition by T CD4+ lymphocytes. They are present in a restrict group of cells (e.g. macrophages, dendritic cell and B lymphocytes) known as Antigen Presenting Cells (APCs). On the other hand, MHC Class I molecules are present in every nucleated cell, including in APCs.

MHC Class I is formed by two polypeptide chains connecting in a non-covalent way, being one heavy chain (α chain) and an invariable non-coded subunit in the same *locus*, called β_2-microglobulin. The amino-terminal segments (N-term) of the α chain (α1 e α2) interact forming a platform composed by eight β-strands on which two parallel α-helix rely. This ensemble forms MHC-I cleft, a space with sufficient size to harbor peptides from eight to eleven amino acids. MHC-I molecule fully composed is an heterodimer composed by two chains (α e β_2-microglobulin) and an endogenous peptide[1] binding to them, and the stable expression of this complex in cell surface require the presence of these three components [16].

The epitope's fit in MHC cleft is performed by a Peptide Loading Complex (PLC) which, besides the recently synthesized MHC two chains, involves an ER chaperone series and accessory molecules, highlighting calnexin protein, Bap31, ERp57, tapasin and also TAP. It is known that Bap31 is involved in peptide:MHC complexes aggregation in exit sites to ER exportation [17]. The rising MHC-I chain (α) is directed to the ER, where it is recruited by chaperones Grp78 and calnexin. The binding of β_2-microglobulin induces the substitution of calnexin by an ortholog chaperon, the calreticulin (CRT). In this step, ERp57 and tapasin are also recruited, forming the PLC. The ERp57 – a thiol oxireductase - is known for catalyzing the formation of disulfide bonds, being, along with the CRT, involved in the correct folding and in the maturation of a series of ER glicoproteins, as well as of the own MHC-I. The

long time of interaction of ERp57 with the MHC and with the other PLC integrants indicates a probable function of this chaperon in cleft formation and in peptide binding [18]. Tapasin is a protein responsible for collocation between PLC and TAP. This protein binds covalently to ERp57 and posses different known functions, over MHC and also over TAP. It is an assumption that this heterodimer, ERp57-tapasin, is responsible for the peptide editing process after dissociation between TAP and the recently built peptide:MHC complex [19].

2.4. TCR Interaction

After the complete assembly of peptide:MHC complex, it is routed to Golgi Complex and then to cell surface, where it will be recognized by Cytotoxic T Lymphocytes. This recognition occurs by the interaction with the T Cell Receptor (TCR), which is a process established by accessory molecule CD8. TCR is a heterodimeric structure, comprehended by an α chain and a β chain or, alternatively, a γ and a δ chain. Inside each of these chains there are three hyper variable sites, known as complementariety determinant regions (CDRs), which protract as TCR coils and perform direct contact with peptide and MHC molecule sites [20]. T Lymphocytes suffers a rigid selection process in thymus, where, by stimulus competition, non-responsive lymphocytes are neglected and dye, while highly-reactive lymphocytes are negatively selected, remaining potentially a population of lymphocytes whose TCR posses only low affinity for self peptides [21].

Through the endogenous peptides presentation route, which involves from the proteasome to the TCR presentation, self peptides are being constantly presented to CD8+ lymphocytes by all nucleated cells. According to the thymic selection process, peptide:MHC complexes presenting self peptides in the cell periphery are not recognized by TCR. This constant surveillance is of vital importance in the prevention and in the response against cellular pathologies, specially viral infection and oncogenesis [16].

Being intracellular mandatory parasites, viruses infect its target cell and induce the performance of necessary processes to its replication. Independently of the strategy, which varies according to the virus and its type of genome, viral proteins will be produced in large scale and liberated in the infected cell's cytoplasm. Inevitably, some molecules of these proteins will suffer the ubiquitination process, being then routed to MHC-I's via. Viral epitopes in MHC's context, usually distinct from self peptide:MHC complex, will be recognized by TCRs, leading to T CD8+ lymphocyte activation, liberation of its cytotoxic content and the consequent lyses of the infected cell.

The liberation of cellular cytotoxity is not, however, the only possible consequence after the contact between peptide:MHC complex and TCR. The intensity of this interaction can vary among a wide range of possibilities, triggering different behaviors in the lymphocyte, among which highlights the activation, clonal expansion and the anergy process [22, 23]. Despite, a consequence of this contact can be the generation of T lymphocytes memory, central mechanism in the adaptative immune response and with vital importance to the development of prophylactic vaccines [24].

2.5. Cross Presentation

Immunology is usually presented in a static way, with a series of subdivions that, apparently, do not correlate: humoral and cellular response, MHC-I and MHC-II pathway, Th1 and Th2 profile, etc. These subdivisions are very useful from didactic point of view, but they are not a faithful picture of the dynamic immune system that has been presented by most recent researches. Everything seems to be more complex than the first thought. For example, though the main source of peptides presented by MHC-I is the pathway that involves UPS and PLC, there are alternative pathways that should not be neglected. There are other cytosolic peptidases that are also able to generate peptides, which can,

eventually, be transported by TAP and routed to MHC. Besides, loading mechanisms occurring outside the ER, in process independent from TAP, have been suggested.

One of the most interesting examples of this interaction is the so called cross presentation, phenomenon by which exogenous antigens are presented in cell surfaces in MHC Class I context. According to mentioned before, exogenous peptides are usually presented in MHC Class II context, arising from endosomal degradation of internalized proteins. Thus, MHC-II pathway is only present in cells that have phagocytic capacity and, in the same way, the cross presentation is restricted to this same set of molecules.

Although it has been already described in macrophages, B lymphocytes and liver sinusoidal endothelial cells [25, 26], the cross presentation seems to occur primarily in dendritic cells. These cells have the unique capacity to stimulate T naïve cells and can also internalize and degrade non-immune infected cells or fragments derived from cells and subsequently connect peptide fragments to MHC Class I. After phagocyting the antigens, the DC migrates to lymphonodes where it can perform the peptide presentation to a T CD8 cell. The mechanisms involved in this atypical presentation have not been completely elucidated, but it is supposed that in most cases the process requires antigenic proteins exportation to the cytosol, where they are routed to degradation by the proteasome. The generated peptides are translocated by TAP to cross presentation-loading compartment *lumen*, to MHC-I binding. This compartment may be either the ER or a mix phagosome-ER compartment. MHC class I egress from the loading compartment to cell surface remains to be analyzed [25].

The physiological outcome of cross presentation of cell-associated antigens by MHC class I may be either tolerization ("cross tolerization") or priming ("cross priming") of CD8+ T cells. In the context of viral immunity, cross presentation may be especially relevant when virus infections are confined to non-hematopoietic tissues [27].

3. Immunogenicity: Heaven or Hell in T-Lymphocyte's Judgment

After passing through the different phases of antigen processing era, the resulting epitopes – the chosen ones – will be offered in cell surfaces over MHC-I and will wait for the cytotoxic T lymphocyte's judgment, which might condemn the death of the hosting cell. Mythological similarities aside, the bases that rule cytotoxic response mechanisms against viral infection are basically these. Every instant, CTLs are supervising cell surface, alert for any minimal sign of infection.

Certainly, the bottlenecks (proteolysis, TAP transport, MHC loading) where shaped by evolution so that, in an infection scenario, among the epitopes presented, will be the ones which have characteristics that indicate to the immune system that the cell is infected. Events like immunodominance and the formation of immune proteasome are only some of the evidences that peptide generation is not a random phenomenon.

It is possible to suggest the following scenario so for a certain viral epitope to become immunogenic: it needs to be generated, carried and recognized by the immune system as a potentially harmful signal. Another key issue for the epitope to be recognized by T lymphocyte is that it must be there! Though this affirmation may seem obvious, it has many implications. To perform a calm migration to cell surface and to maintain itself stable in the extracellular surface, the epitope requires specific physicochemical conditions that keep it accommodated with high affinity to its MHC receptor. These requirements refer mainly to anchor residues.

Anchor residues, as the name indicates, keep the epitope anchored to MHC cleft pockets, providing stability to the peptide:MHC complex. Since each MHC allele can differ on amino acids that compose the pockets, there is a preference over some residues according to each allele or group of related alleles.

Thus, a same peptide can have high affinity with an MHC allele and, on the other hand, be a bad ligand for another allele and vice-versa.

Therefore, different individuals have different abilities when responding to some infections, because, depending on its MHC allele constitution, it will be capable of presenting a determined pool of epitopes. This might allow some individuals to respond to some infections in a more efficient way [28]. On the other hand, some viral variants are capable of escaping from immune system recognition, by avoiding its presentation by MHC alleles [29].

The third and crucial phase is the recognition by the T cell receptor, here referred as the T lymphocyte eye. This structure has the ability to discriminate self from non-self peptides. In part, this capacity is provided by thymic education, in which only the TCRs that are able to weakly interact with self peptide:MHC complexes are selected (positive selection), while the highly self-reactive ones suffer apoptosis (negative selection).

The traditional approach, which claims that there is one T cell receptor for each peptide, gave rise to a more dynamic vision of this interaction, on which the receptor has a more promiscuous behavior that allows it to contact with thousands of different ligands [30]. However, its capacity to trigger a cytotoxic immune response against the correct target is highly specific.

In the next three sub-sections we will approach issues intimately related to the result of the interaction between the TCR and the peptide:MHC complex: the recognition of antigens, the cross reactivity and the molecular mimicry.

3.1. Antigens Recognition

The major issue that permeates perhaps almost the whole immunology history is how the immune system differentiates self from non-self peptides. In cellular response against viral infections, this task is assigned to the T cell receptor. Though the TCR sees the MHC molecule, it only has the peptide sequence (which usually has from 8 to 10 amino acids) as a reference to take this decision. Also, the context is extremely important in differential recognition, and we must not neglect the role of all accessory molecules and cytokines that participate in the immune response assembling. The logic of this process is like when we see a suspicious individual in a dark street. Even though the context leads us to taking some attitude more defensive or negligent, the fact that the individual has a suspicious phenotype is determinant. How, thus, can we detect which peptides are suspicious? In the last two decades, the focus to find this answer was in sequence analyzes. Both the search for conserved motifs and the prevalence of a specific amino acid on each epitope position are still being investigated [31]. However, restricting the complex interaction between T cell receptors and peptide:MHC complexes to simply analyzing sequences is a naïve simplification of the system. The surface contacted by the TCR (cleft and epitope) is more than a simple sequence. It has the signs, both topological and the combination of charges, which allows the TCR to establish a more or less stable binding, that determines if the required cytotoxic mechanisms will be executed or not. Obviously we understand that this kind of approach (sequence analyzes) comes from the low number of three-dimensional peptide: MHC structures available in PDB [http://www.pdb.org], which allows large scale comparison. Nowadays, reliable techniques of peptide: MHC complexes construction are available [32, 33]. The only limiting point is that, for the best accuracy in the complex construction, a crystal structure of the MHC allele of interest is needed.

When analyzing structures instead of sequences, epitopes that looked distantly related have very similar molecular surfaces [34], while variants of a same peptide may have great variations in the generated complexes [32]. Lots of this observation might explain the next two presented phenomena.

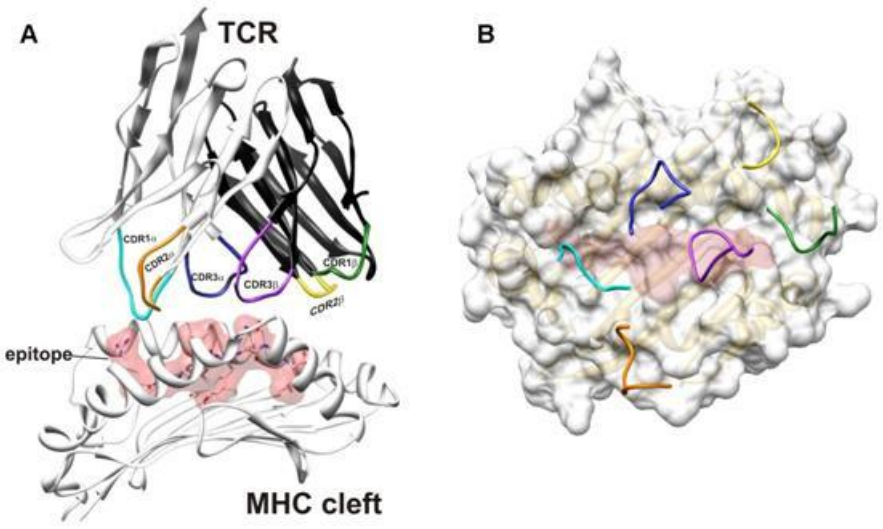

Figure 1. (A) Side view of the interaction between a peptide:MHC complex and the T Cell Receptor. Peptide is presented inside the MHC cleft, with surface depicted in red. Both TCR chains are represented (alpha in grey and beta in black) with CDR loops enhanced. (B) Top view of a peptide:MHC complex. TCR-interacting surface depicted in grey and the peptide corresponding area enhanced in light red. Positions of CDR loops are also presented, in the same colors as in (A)

3.2. Cross Reactivity

In cellular response to viral infections, cross reactivity refers to the capacity of a single T cell receptor to recognize distinct epitopes [35]. This degenerated recognition urges from the necessity of response plasticity, elapsed from the high number of theoretical peptides that can be generated. As observed previously, the gain in the recognition capacity to different targets does not result in specificity loss [36]. These mechanisms can also allow many of epitopes variants, elapsed from viral mutations, to be recognized by the same TCR. However, this is not a rule. In the same way that epitopes with low similarity can stimulate a cross response, epitopes with only one alanine modification can abort an immune response [37]. The explanation to these discrepancies may be because, once more, we are talking about sequences, symbols we attribute to structures that have size, polarity and, in some cases, charges. Thus, the surface analyzes of epitopes we considered different might reveal similar structures, as well as the variation of a single residue can change an epitope backbone or even the charge distribution in the TCR contact area.

The complete comprehension of cross reactivity mechanisms between viral epitopes will make room for the development of wide spectrum viral vaccines.

3.3. Molecular Mimicry

In cellular response is very difficult to dissociate the concept of cross reactivity from molecular mimicry. As a matter of fact, it all depends on the point of view. In cross reactivity, the immune system takes advantage of TCR plasticity to optimize its responses range, being able to response against different pathogens epitopes from one single T cell receptor. In molecular mimicry's case, mutations that occur in various viral proteins make some epitopes structurally similar to self proteins. Thus, or the immune system ignores this epitope, which gives a gain for the pathogen to proliferate; or some T cells become not only responsive to variant epitopes, but also become reactive to peptides arising from self proteins. The nefarious result of this second option is an event unknown until recently: some viral

infections might be involved in autoimmune diseases etiology [38]. Among the more studied associations, is the Epstein-Barr Virus (EBV) viral infection and the development of Multiple Sclerosis (MS) [39-41]. A mechanism that really proves how these autoimmune disturbance is induced from viral infection is still unknown, but it is possible to imagine the following scenario. During a viral infection, specially chronic ones, the presence of viral antigens is notorious. If we consider the issue approached in the beginning of this chapter that some protein regions are more likely to contain epitopes that will be constantly presented to the immune system (immunodominat epitopes) and that these epitopes, in the context of some MHC alleles, may form complexes similar to self peptide:MHC complexes, it is possible that the constant stimulation with some peptides in an inflammatory context that favors the recognition supports the violation of tolerance of some T cytotoxic cells. If these cells are in a compartment that contains self peptides similar to mimetic viral epitopes, or that are exposed to these antigens, the trigger for the autoimmunity to be established may be initiated. Thus, when working with immunoinformatics, an issue that must be remembered when we search for vaccine targets is that we must avoid viral targets that have sequences similar to human proteins, because there are evidences that vaccines can trigger autoimmune diseases [42]. A complicating point in this case is that even epitopes apparently different can be structurally similar, and it would be necessary to perform comparative structural analyzes even between sequences without conspicuous similarity. Since we have quoted immunoinformatics approach, in the next sections we will explain specifically what it is, which are its applications and which are the main tools used in this prominent area of vaccinology.

4. Softwares That Simulate the Antigens Presentation Pathway

There are a series of softwares that are used to simulate different steps of intracellular antigens presentation. Next, we present some softwares that are commonly used and that have *in vitro* validation.

4.1. EpiJen

EpiJen is a software that simulates the endogenous peptides presentation route, from the cleavage by proteasome until the binding with MHC-I. The advantage of EpiJen is that it performs this chain reaction, simulating from the proteasome cut to the MHC binding. On the other hand, the disadvantage is that in specific points of the route it does not generate such a better result as other softwares that perform only this step. Besides, this software does not use static algorithms, which are less advanced than systems that use neural networks. After the job is submitted, the peptide (9 amino acids) rank is presented. Two values follow the sequences: IC50 (nM), which corresponds to half of this peptide concentration (in nM) necessary for the MHC binding to occur, in other words, the smaller, the better; and also logIC50(M), which is a increscent scale of the affinity value, in other words, the bigger, the better. This last has a smaller variation between the first and the last data, which is better for graphic plotting and analyzes [http://www.jenner.ac.uk/EpiJen/] [43].

4.2. Netchop 3.0

NetChop 3.0 [http://www.cbs.dtu.dk/services/NetChop/] [44] is indicated by literature as the proteolytic simulation software, presenting superior performance in comparison to other simulators. This software uses neural networks trained to identify and generate, in one protein (FASTA), the same cuts generated by human proteasome. Although it was trained to work with human data, considering the

great conservation of proteasome 20S subunit through evolution, NetChop prediction is assumed to be correctly extrapolated to other mammals.

This server presents two different networks to generate the output data. “C-term 3.0” uses a neural network trained in a 1260 published MHC-I ligands database, and, thus, is appropriated to simulate immuneproteasome and to indicate the carboxy-terminal extremity of epitopes presented to cytotoxic T lymphocytes. “20S 3.0” uses a network trained in published *in vitro* data, and it is indicated to simulate the proteasome.

The standard output has five columns, from left to right: amino acid position in FASTA sequence input; amino acid name (one-letter code); cleavage presence (“S”) or absence (“.”) ; cleavage score in that point (probability); and sequence identifier (first line of fasta file, containing the “>” character).

4.3. PAProC I

This software presents the same function as NetChop (more similar to 20S 3.0), simulating cuts by proteasome, but it uses a static algorithm. This algorithm calculates the global affinity of each amino acid pair on the proteasome catalytic site (in subunit 20S), estimated by *in vitro* digestions. In case this pair of amino acids presents great affinity by catalytic amino acids of 20S subunit, the software indicates a possible cut spot. [http://www.paproc2.de/paproc1/paproc1.html].

4.4. TAPPred

This server performs the prediction of nonamers capable of binding to TAP by a protein amino acids sequence [http://www.imtech.res.in/raghava/tappred/]. The sequence can be in FASTA format or a simple letter sequence, and the calculation considers the amino acids and their properties (Cascade SVM).

The software can show all nonamers analyzed or only the ones with a determined affinity (high, intermediate or low). The output brings two different visualizations: in the first one, the query sequence is presented and the colors indicate nonamers with high affinity (green marks the amino-terminal extremity of the nonamer, while purple marks the other amino acids); the second one is a ranking of all analyzed nonamers, sort by its affinity to TAP.

5. Molecular Docking

The peptide:MHC complexes recognition by TCR is a key step in the construction of cellular immune responses, which, in turn, are involved in a number of fundamental mechanisms such as surveillance against tumors, protection against intracellular parasites and development of autoimmunities. Thus, the study of structural details involved in this interaction is of extreme interest, and may lead us to a broader understanding of immunological phenomena, such as cross reactivity and immunodominance. However, the great diversity of this complexes and the still limited number of X-ray structures restricts structural studies. MHC molecules are extremely polymorphic, with thousands of alleles already described and with most variability located in the peptide binding cleft. Besides that, a given allele can bind a wide range of different peptides, varying both in amino acids composition as in length (from 8 to 11 residues). Each combination of peptide sequence and MHC allele will produce a unique complex that will interact with a given TCR in a specific manner. Since determining the crystallographic structure of all possible peptide:MHC complexes is not a reasonable task, an accurate

prediction of these structures became a critical goal for computational biology [33, 45]. Among the tools that are already used to address these issues, molecular docking stands out.

Molecular docking, or just docking, is a method to predict the preferred orientation of a ligand within a receptor (macromolecule), or even to search for possible binding pockets in a given macromolecule surface [46]. This approach is largely used for ligand screening and drug design [47, 48]. There are many softwares, both free and paid, developed to perform this task, and some of the most commonly used are AutoDock, GOLD, DOCK, FlexX, ICM, QXP, GLIDE, and FRED [49-56]. In all cases, to perform the docking two input structures are needed, for the ligand and the receptor.

The receptor structure, in this case an MHC-I structure, could be both from the Protein Data Bank - in case this allele has its structure determined - or acquired by homology modeling, using another allele's PDB structure as a template [33].

Most of molecular docking studies usually use ligands with up to 10 rotatable bonds into rigid or partially flexible targets, since ligands with a larger amount of rotatable bonds are much more difficult to dock, even using a rigid receptor [57-59]. A typical 9-mer peptide, however, has about 35 to 44 rotatable bonds and this number can be even higher according to amino acid sequence of the epitope. Besides, if we are using a unique "MHC-donor" structure, to receive different peptides that can be presented by this allele (procedure referenced as cross-docking), we must consider that the side-chains in the cleft of the donor structure are not in the best conformation to receive all these different ligands and, in order to have a successful docking, we need to adjust this side chains to each different ligand.

There are several works applying docking to build peptide:MHC complexes, as well summarized in the article by Bordner and Abagyan [33]. In order to deal with all the variability and degrees of freedom, most of described methods divide the structure prediction into separate parts, e.g. the peptide termini and central portion or peptide main chain and side chains. A combined approach of molecular docking and energy minimization was also described (*D1-EM-D2*) [32]. In this case, the peptide backbone is fitted to a known allele-specific pattern, and then maintained rigid during the docking. After the first docking, a short energy minimization is performed in the complex, allowing the relaxation of MHC side chains, which improves the interaction with the peptide. This approach proved to be useful for cross-docking of different peptides in the same MHC structure, but its application is limited to alleles with at least some structures experimentally determined. Alternatively to docking approaches, the MHCSim server [60] brings a fast and easy way to build peptide:MHC complexes. The server uses your input (peptide sequence and MHC allotype) as a query to search for similar structures at PDB. The most similar structure is then mutated to generate your complex of interest. Besides the limitation for human MHC alleles and peptide length (9-mers), sequence similarity may not necessarily be responsible for structural similarity, especially in what regards to the conformation of peptide inside the MHC-I cleft.

After the peptide:MHC complexes construction, these structures can be used to perform several analysis, such as Accessible Surface Area (ASA), hydrogen bonds pattern, topology and electrostatic potential comparison. Structural studies of these complexes may also be used to binding affinity assessment. Although we have sequence-based methods that can accurately predict peptide:MHC binding affinity, their use is limited to a small number of MHC allotypes with sufficient quantities of experimental data. Once we have reliable methods for modeling MHC alleles and accurate methods to predict peptide:MHC complexes structure, physics-based scoring functions can be used to provide general binding affinities, applicable to different peptide lengths and MHC allotypes [33].

6. Immunology Usage of Molecular Dynamics

Molecular Dynamics consists of a technique widely used in many work fields. In biological sciences this methodology is becoming more useful each day, due to the possibility to observe, analyze and study details so far obscure. Through the application of advanced mathematical function and

algorithms, it is possible to study structural molecular fluctuations over time, based on Newton's movement laws. In this chapter, we will focus in the study of peptide-MHC interaction using a molecular dynamics technique.

From the theoretical point of view, molecular dynamics was conceived in the 50's decade. At the end of de 50's, researchers performed a molecular dynamic using a *hard-sphere model*, in which the atoms interacted only through perfect collisions [61]. The next progress occurred in 1964, when Rahman simulated a system of 864 particles interacting with a Lennard-Jones potential in order to study the molecular dynamics in liquid argon at 94.4ºK. That was the first time that a realistic potential was truly used [62]. In the 70 decade, with the dissemination of computers usage, a rising in publications using molecular dynamics technique occurred. In 1974, Stillinger and Rahman performed a molecular dynamics in a realistic system using 216 water molecules as solvent [63]. The first molecular dynamics simulation of a protein was performed in 1977, using bovine pancreatic tripsin inhibitor (BPTI), in a total of 9.2ps considering a system of approximately 500 atoms [64]. Nowadays, molecular dynamics panorama is different. With the enhanced evolution of computational power, combined with molecular dynamics software updates, it is possible to simulate systems with over 10^4 atoms for up to 1 microsecond [65]. A quick search about molecular dynamics in PubMed database results in an enormous amount of articles quoting simulation of solvated proteins, complexes protein-DNA, as well as systems involving lipids to solve thermodynamics ligand binding issues and little proteins folding.

In immunology field, molecular dynamics can be used for a series of studies, both for analyzing the interaction between antigens and immune system molecules and also for wider studies involving immunogenicity. Many works are arising in the context of analyzing details of peptide:MHC binding.

Rognan et al. (1994) describes a simulation using a total of six ligands arising from heat shock protein of *E. coli* or *C. trachomatis*, with different binding affinities (experimentally proved) by human MHC-I HLA-B27, in order to identify good and bad ligands through molecular dynamics. In this work, researchers were able to, through a 150ps simulation, analyze the RMSD of each ligand, the solvent access surface, the RMSF of each residue and the pattern of hydrogen bonds formation between each ligand and HLA-B27. Results were useful to evaluate the abnormality and the differences of chemical interactions in MCH cleft, providing clues about the binding pattern of different ligands to HLA-B27 [66].

Another study, also with focus directed to MHC molecule and using molecular dynamics was performed by Zacharias et al. (2004). With a very superior simulation time (approximately 25ns), researchers were able to evaluate the conformational flexibility of MHC-I HLA-A*0201 molecule in the presence of ligand (GVYDGREHTV - a peptide derived from a tumor-specific antigen) [67]. Besides the numerical data, interesting information was withdrawn through visualization of the molecule in different points of the simulation. A posterior study, performed by the same group, used molecular dynamics to evaluate two MHC-I alleles, one tapasin-dependent (HLA-B*4402) and the other non-dependent (HLA-B*4405). Both alleles differ in only one residue, in position 116. Both molecules were studied in the presence and also in the absence of ligands. The data arising from molecular dynamics showed that, in peptide absence, HLA-B*4402 molecule adopts a more open structure in the anchoring region of the peptide's N-terminal extremity (pocket F). On the other hand, HLA-B*4405 allele, in the same region, presents a more closed conformation when compared to its binding state. These are very relevant data to understand peptide binding mechanism to MHC-I molecules that depend or not of tapasin [68].

Simulations of bigger systems and involving more molecules require a higher computational power. However, currently, this is not a limiting factor anymore, such as highlighted in Wan et al. [69]. In this work, researchers simulated by molecular dynamics, during 10ns, the smallest system necessary to enable an immunological synapse, containing a TCR (receptor responsible for peptide:MHC recognition), a peptide:MHC complex and an accessory CD4 glycoprotein molecule. Besides, they also included in the system part of the lipid bi-layer membrane which anchors these three molecules.

Other immune system focus can also be approached, besides peptide:MHC systems. Su et al. [70] used molecular dynamics to study the binding mechanism of lysozyme and the Camelid VHH HL6 antibody, paying special attention to hydrogen bonds formation pattern between the two complexes. Another work, Perryman et al., 2004 [71], studied two important mutations in HIV-I proteases, V82F/I84F, which are responsible for decreasing inhibitors affinity of proteases used currently in infected people treatment. Through this study it was possible to observe important variations that occur in mutated proteases in comparison to natural proteases and to understand better inhibitor's binding mechanisms. This has an important clinical effect, since knowing inhibitor's binding mechanism and mutations in HIV-I proteins of infected patients allows a previous reassignment so that patients with a certain mutation do not receive a determined inhibitor which, certainly, will not have clinical efficiency, and might even lead to generation of multi-resistant strains.

To run molecular dynamics, specific softwares are required. There are innumerous softwares available, so we quote the main. AMBER (Assisted Model Building with Energy Refinement) refers to a suite of programs which are used in order to perform molecular dynamics. The AMBER software suite is divided in two parts: AmberTools and Amber11. There are approximately 50 programs that, together, work reasonably well. Among the main programs we can cite *sander* (Simulated annealing with NMR-derived energy restraints), which allows NMR refinement based on NOE-derived distance restraints, torsion angle restraints, and penalty functions based on chemical shifts and NOESY volumes and can also be used to integrate thermodynamics, replica-exchange and potential mean force (PMF) calculations; *pmed*, a modified version of *sander* optimized for periodic, PME simulations and GB simulations, which works better in parallel processing machines; *ptraj*, used to analyze molecular dynamics trajectories; among others. This product is sold under a license agreement, and its price varies considering the client and his objective [http://ambermd.org/].

CHARMM (Chemistry at Harvard Macromolecular Mechanics) is a package of programs to perform molecular dynamics and also the name given to a set of force fields commonly used. It is a versatile software, and it can run in a parallel environment involving multiple molecules. Its development was focused for usage primarily in the study of biological systems' molecules [http://www.charmm.org/html/info/intro.html].

A software developed not only for biological systems, but also directed to computational chemistry is GROMACS [http://www.gromacs.org/]. It has a very varied group of equations used to calculate bonded and non-bonded interactions. One of its differentials is the usage of a preprocessor (GROMPP), which intends to optimize input files. It gathers information about topology, coordinates and velocities in a single file, which is used to perform the molecular dynamics. This file can be parallelized among various processing cores, allowing the job optimization. Another differential are the innumerous utilitarian that can be used to analyze the final molecular dynamics trajectory.

Besides molecular dynamics specific softwares, there are also specific hardwares to perform these simulations. In New York is situated Anton, a specialized, massively parallel supercomputer designed to execute molecular dynamics simulations. Also, in Japan, there is another supercomputer, named MDGRAPE: a special purpose system built for molecular dynamics simulations, especially protein structure prediction.

7. Immunoinformatics Databases

Nowadays, there are several immunological databases available for the scientific community, which emerged with the intent of organizing and storing the huge amount of information collected through decades of immunological experimental analysis.

In November 2009 Journal Nucleics Acids Research edition, approximately 30 databases were quoted in immunological databases section [http:// www. oxfordjournals. org/ nar/ database/cat/14]. Since there are a lot of immunological databases available, we choose those that contain data related to

immune epitopes. IEDB, SYFPHEITI and IMGT databases are the most quoted immunological databases, with over 100 quotes per year [72].

7.1. Immune Epitope Database (IEDB)

IEDB [http://www.iedb.org] provides information about B cell and T cell for different organisms, including human. Also, it holds manually curated data for epitopes related to allergens, infectious diseases, and even autoimmunity [73]. In section *Browse by 3D structure*, in IEDB, it is possible to access structural data referent to B cell, T cell and also MHC binding. Let's say you are interested in evaluating every peptide:MHC complex structure of HLA-A*0201 allele. Selecting the option *HLA-A*0201*, the program retrieves a table containing the following information: MHC binding ID (IEDB identifier), reference, epitope, MHC restriction, assay description and quantitative measures. Besides the possibility of visualizing the tridimensional structure of each complex, relevant information considering epitope residues interacting with MHC, as well as MHC's residues interacting with epitope are available. In addition to its database, IEDB has a series of tools to predict and analyze immune epitopes, which goes from the epitope generation by proteasome, through TAP transport, to MHC Class I and Class II binding.

7.2. SYFPEITHI

SYFPEITHI [http://www.syfpeithi.de/] is one of the most commonly used epitope databases. This database hosts more than 7000 peptide sequences known both as MHC Class I and Class II binders, collected from different scientific articles. One characteristic of this server is that it is restricted to hold naturally processed epitopes from humans and other species, in other words, it excludes information about synthetic peptides. For users interested in cancer research, this database provides access to a list of manually curated epitopes derived from cancer, grouped by MHC allele. Besides the epitope sequence, information such as peptide source and the referenced paper are available. In addition to the database, SYFPEITHI also holds tools for natural MHC binders (including processing, transport and MHC binding) [74].

7.3. International Immunogenetics Information System (IMGT)

Another important database for autoimmune disorders and infectious diseases is the IMGT [http://www.imgt.org/], which hosts information about immunoglobulins (IGs), TCRs, MHCs as well as related immune systems proteins (RPI). IMGT is composed by various databases, including three sequence databases (IMGT/LIGM-DB, IMGT/MHC-DB and IMGT/PRIMER-DB**)** one genome database **(**IMGT/GENE-DB), one 3D structure database (IMGT/3Dstructure-DB**)** and one monoclonal antibodies database (IMGT/mAb-DB) [75]. Among these, IMGT/3Dstructure-DB should be highlighted, since it holds 3D structures of IG, TCR, MHC and RPI available in PDB. In this database, the searching is efficient and allows filters; for example, you can separate peptide:MHC structures deposited in PDB that are complexed with the TCR from the ones that are not. Important information such as chain details, contact analysis between peptide and MHC, as well as structures visualization can be accessed. IMGT's website also holds several tools for IG, TCR and MHC sequence analysis; however, it does not hold any tool related to epitope prediction and analysis. In section *IMGT Web* resources, it is worth checking the link "*The IMGT Immunoinformatics page*", where you can find information and links about different immunological databases and tools.

7.4. HIV Molecular Immunology Database

When focusing in HIV, HIV Molecular Immunology Database [http://www.hiv.lanl.gov/content/immunology/], which holds highly curated information on HIV derived epitopes, stands out [76]. According to his interest, the user can choose among three different research sections: (i) CTL (CD8+); (ii) T helper (CD4+); and (iii) antibody. In these sections, it is possible to find information that go from response against HIV natural infections, through results of vaccine studies in animal and human models, to cross-reactivity and escape mutation.

7.5. AntiJen

AntiJen [http://www.darrenflower.info/antijen/antijenhomepage.htm] is not only a T cell and B cell database; its difference comparing to other databases is that it includes experimental data about immunological molecular interactions, including thermodynamic and kinetic measures [77]. This database has many ways to perform a search, including MHC ligand, MHC kinetics, T cell epitope, TCR-MHC, TAP ligand, protein interactions and diffusion coefficient.

7.6. EPIMHC

EPIMHC is a database hosted in http://immunax.dfci.harvard.edu/epimhc/ website, focusing specifically in epitopes ligands to MHC [78]. In this database, the search can be performed by peptide sequence or size, by MHC allele, among others. Besides, it is also possible to search by peptide binding level (“high”, “medium” or “low”) and immunogenicity level (“high”, “medium” or “low”). The result is a list of epitopes containing information about the MHC allele, the epitope sequence, the source protein name, the epitope, the epitope immunogenicity level, the peptide source organism and finally the sequence’s size. Additionally, along with each epitope there is a check box so that the user can select the epitopes for download; after the selection, the result can be saved in a table format (containing all information previously referenced) or in FASTA format (containing only peptides sequence).

7.7. MHCBN

MHCBN [http://www.imtech.res.in/raghava/mhcbn/], on the other hand, is a curated database that holds over 1600 antigens and 450 MHC alleles, contemplating information about MHC peptides binder and non-binders, as well as TAP binders and non-binders [79]. In section *Peptide Search* it is possible to search submitting peptide sequence of interest against MHCBN peptide binders database; also, it is possible to restrain the search defining parameters such as MHC class, host organism, binding strength and T-cell activity. Another interesting section is *Dataset creation*, where it is possible to create an specific dataset to a determined MHC allele, for different peptide binding affinity (“all”, “all binders”, “high”, “moderate”, “low”, “non-binders”) and different peptide T-cell activity (“all T cell epitope”, “high”, “moderate”, “low”, “non T cell epitope”). Recently, autoimmune diseases’ information where uploaded to this database. You can perform the search: (i) by disease’s name, resulting in a list of MHC alleles related to the disease; or (ii) by MHC allele, resulting in a list of diseases associated to this allele.

Regarding the scope of each database described above, we can observe some overlap of information between them. Although some data is repeated in different databases, each database also has exclusive information. Thus, it is advisable to perform the same search in more than one database, in order to avoid the gaps that each database has. Another relevant issue that must be considered when using a database is to verify whether the website is being updated or not, and when was the last update.

8. Conclusion

In this chapter we passed through many topics that rule cellular immunogenicity against viral infections and discussed crucial matters that must be considered when choosing vaccine targets. Besides, we presented tools and databases that represent the state of art in the recent field of immunoinformatics. An increasing complexity level arose along with the progress of the chapter, demonstrating that from the beginnings of immunization attempts, centuries ago, until the most modern computational techniques that can help in the development of more rational vaccines, many elements have been unraveled. Though we have not approached the experimental part, we are conscious of its incontestable importance in this process. What we wanted to present is that the number of targets that can be experimentally tested can be drastically diminished with the help of immune informatics. This represents gain of time, money and security. We evidence, for example, that the passage of a protein through the antigens processing pathway excludes the majority of theoretical epitopes that can be generated from a determined virus. We also demonstrated that even these generated epitopes need to have structural requirements and specific physicochemical characteristics to establish a stable binding with the T cell receptor and to stimulate a full cytotoxic response. Besides, we must consider the caution we have to take not to choose targets that can bring risks to the patients, such as the development of autoimmune diseases elicited by viral epitopes similar to self proteins. All these factors can be simulated and accessed through different bioinformatic tools that we presented in the last section along with databases that are extremely useful as a starting point to this kind of research. Using information contained in these databases, we can obtain information about sequences, the organism we are studying and also about the immunological profile of populations that we intend to immunize. This is of major importance when we think that population groups are heterogeneous and probably a single vaccine that would work on a group with people that have a certain pool of MHC alleles would also present different results in a population with another allele pool. Finally, our objective was to demonstrate that the number of ingredients in the recipe for a vaccine construction is great and meticulous and certainly lots of them still a secret.

9. References

1. Nandi, D; Tahili, P; Kumar, A; Chandu, D. The ubiquitinproteasome system. *J Biosci*., 2006, (31), 137-55.
2. Goldberg, AL. Protein degradation and protection against misfolded or damaged proteins. *Nature*, 2003, Dec 18, 426(6968), 895-9.
3. Ciechanover, A. The ubiquitin-proteasome proteolytic pathway. *Cell*, 1994, Oct 7, 79(1), 13-21.
4. Liu, YC; Penninger, J; Karin, M. Immunity by ubiquitylation: a reversible process of modification. *Nat Rev Immunol*, 2005, Dec, 5(12), 941-52.
5. Wang, J; Maldonado, MA. The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. *Cell Mol Immunol*, 2006 Aug, 3(4), 255-61.
6. Paulsson, KM. Evolutionary and functional perspectives of the major histocompatibility complex class I antigen-processing machinery. *Cell Mol Life Sci*., 2004, Oct, 61(19-20), 2446-60.
7. Falk, K; Rotzschke, O. The final cut: how ERAP1 trims MHC ligands to size. *Nat Immunol*, 2002, Dec, 3(12), 1121-2.
8. Adams, J. The proteasome: structure, function, and role in the cell. *Cancer Treat Rev*., 2003, May, 29 Suppl 1, 3-9.

9. Fabunmi, RP; Wigley, WC; Thomas, PJ; DeMartino, GN. Interferon gamma regulates accumulation of the proteasome activator PA28 and immunoproteasomes at nuclear PML bodies. *J Cell Sci*., 2001, Jan, 114(Pt 1), 29-36.
10. Kindt, TJ; Goldsby, RA; Osborne, BA. Kuby *Immunology*. Palgrave/W.H. Freeman, 2006.
11. Foss, GS; Larsen, F; Solheim, J; Prydz, H. Constitutive and interferon-gamma-induced expression of the human proteasome subunit multicatalytic endopeptidase complex-like 1. *Biochim Biophys Acta*, 1998, Mar 12, 1402(1), 17-28.
12. Procko, E; Gaudet, R. Antigen processing and presentation: TAPping into ABC transporters. *Curr Opin Immunol*, 2009, Feb, 21(1), 84-91.
13. Schmitt, L; Tampe, R. Affinity, specificity, diversity: a challenge for the ABC transporter TAP in cellular immunity. *Chembiochem*, 2000, Jul 3, 1(1), 16-35.
14. Schmitt, L; Tampe, R. Structure and mechanism of ABC transporters. *Curr Opin Struct Biol*., 2002, Dec, 12(6), 754-60.
15. Reits, EA; Griekspoor, AC; Neefjes, J. How does TAP pump peptides? insights from DNA repair and traffic ATPases. *Immunol Today*, 2000, Dec, 21(12), 598-600.
16. Hammer, GE; Kanaseki, T; Shastri, N. The final touches make perfect the peptide-MHC class I repertoire. *Immunity*, 2007, Apr, 26(4), 397-406.
17. Zhang, Y; Williams, DB. Assembly of MHC class I molecules within the endoplasmic reticulum. *Immunol Res*., 2006, 35(1-2), 151-62.
18. Purcell, AW; Elliott, T. Molecular machinations of the MHC-I peptide loading complex. *Curr Opin Immunol*, 2008, Feb, 20(1), 75-81.
19. Wearsch, PA; Cresswell, P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat Immunol*, 2007, Aug, 8(8), 873-81.
20. Brehm, MA; Selin, LK; Welsh, RM. CD8 T cell responses to viral infections in sequence. *Cell Microbiol*, 2004, May, 6(5), 411-21.
21. Sohn, SJ; Thompson, J; Winoto, A. Apoptosis during negative selection of autoreactive thymocytes. *Curr Opin Immunol*, 2007, Oct, 19(5), 510-5.
22. Nel, AE. T-cell activation through the antigen receptor. Part 1: signaling components, signaling pathways, and signal integration at the T-cell antigen receptor synapse. *J Allergy Clin Immunol*, 2002, May, 109(5), 758-70.
23. Nel, AE; Slaughter, N. T-cell activation through the antigen receptor. Part 2: role of signaling cascades in T-cell differentiation, anergy, immune senescence, and development of immunotherapy. *J Allergy Clin Immunol*, 2002, Jun, 109(6), 901-15.
24. Seder, RA; Darrah, PA; Roederer, M. T-cell quality in memory and protection: implications for vaccine design. *Nat Rev Immunol*, 2008, Apr, 8(4), 247-58.
25. Guermonprez, P; Amigorena, S. Pathways for antigen cross presentation. *Springer Semin Immunopathol*, 2005, Jan, 26(3), 257-71.
26. Burgdorf, S; Scholz, C; Kautz, A; Tampe, R; Kurts, C. Spatial and mechanistic separation of cross-presentation and endogenous antigen presentation. *Nat Immunol*, 2008, May, 9(5), 558-66.
27. Heath, WR; Carbone, FR. Cross-presentation in viral immunity and self-tolerance. *Nat Rev Immunol*, 2001, Nov, 1(2), 126-34.
28. Kosmrlj, A; Read, EL; Qi, Y; Allen, TM; Altfeld, M; Deeks, SG; et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature*, May 20, 465(7296), 350-4.
29. Kaslow, RA; McNicholl, J; Hill, AVS. *Genetic Susceptibility to Infectious Diseases*. US: Oxford University Press, 2008.
30. Welsh, RM; Selin, LK. No one is naive: the significance of heterologous T-cell immunity. *Nat Rev Immunol*, 2002, Jun, 2(6), 417-26.

31. Frankild, S; de Boer, RJ; Lund, O; Nielsen, M; Kesmir, C. Amino acid similarity accounts for T cell cross-reactivity and for "holes" in the T cell repertoire. *PLoS One*, 2008, 3(3), e1831.
32. Antunes, DA; Vieira, GF; Rigo, MM; Cibulski, SP; Sinigaglia, M; Chies, JA. Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment. *PLoS One*, 2010, 5(4), e10353.
33. Bordner, AJ; Abagyan, R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins*, 2006, May 15, 63(3), 512-26.
34. Sandalova, T; Michaelsson, J; Harris, RA; Odeberg, J; Schneider, G; Karre, K; et al. A structural basis for CD8+ T cell-dependent recognition of non-homologous peptide ligands: implications for molecular mimicry in autoreactivity. *J Biol Chem*., 2005, Jul 22, 280(29), 27069-75.
35. Regner, M. Cross-reactivity in T-cell antigen recognition. *Immunol Cell Biol*., 2001, Apr, 79(2), 91-100.
36. Frank, SA. Immunology and Evolution of Infectious Disease. *Princeton*, New Jersey.: Princeton University Press, 2002.
37. Fytili, P; Dalekos, GN; Schlaphoff, V; Suneetha, PV; Sarrazin, C; Zauner, W; et al. Cross-genotype-reactivity of the immunodominant HCV CD8 T-cell epitope NS3-1073. *Vaccine*, 2008, Jul 23, 26(31), 3818-26.
38. Munz, C; Lunemann, JD; Getts, MT; Miller, SD. Antiviral immune responses: triggers of or triggered by autoimmunity? *Nat Rev Immunol*, 2009 Apr, 9(4), 246-58.
39. Salvetti, M; Giovannoni, G; Aloisi, F. Epstein-Barr virus and multiple sclerosis. *Curr Opin Neurol*, 2009, Jun, 22(3), 201-6.
40. Pohl, D. Epstein-Barr virus and multiple sclerosis. *J Neurol Sci*., 2009, Nov 15, 286(1-2), 62-4.
41. Bagert, BA. Epstein-Barr virus in multiple sclerosis. *Curr Neurol Neurosci Rep*., 2009, Sep, 9(5), 405-10.
42. Salemi, S; D'Amelio, R. Could autoimmunity be induced by vaccination? *Int Rev Immunol*, Jun, 29(3), 247-69.
43. Doytchinova, IA; Guan, P; Flower, DR. EpiJen: a server for multistep T cell epitope prediction. *BMC Bioinformatics*, 2006, 7, 131.
44. Nielsen, M; Lundegaard, C; Lund, O; Kesmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, 2005 Apr, 57(1-2), 33-41.
45. Sieker, F; May, A; Zacharias, M. Predicting affinity and specificity of antigenic peptide binding to major histocompatibility class I molecules. *Curr Protein Pept Sci*., 2009, Jun, 10(3), 286-96.
46. Dhaliwal, B; Chen, YW. Computational resources for protein modelling and drug discovery applications. *Infect Disord Drug Targets*, 2009, Nov, 9(5), 557-62.
47. Kolb, P; Ferreira, RS; Irwin, JJ; Shoichet, BK. Docking and chemoinformatic screens for new ligands and targets. *Curr Opin Biotechnol*, 2009, Aug, 20(4), 429-36.
48. Fukunishi, Y. Structure-based drug screening and ligand-based drug screening with machine learning. *Comb Chem High Throughput Screen*, 2009, May, 12(4), 397-408.
49. Morris, GM; Goodsell, DS; RS. H; Huey, R; Hart, WE; Belew, RK; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Computational Chemistry*, 1998, 19(14), 1639-62.
50. Jones, G; Willett, P; Glen, RC; Leach, AR; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*., 1997, Apr 4, 267(3), 727-48.
51. Ewing, TJ; Makino, S; Skillman, AG; Kuntz, ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*., 2001, May, 15(5), 411-28.
52. Rarey, M; Kramer, B; Lengauer, T; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*., 1996, Aug 23, 261(3), 470-89.

53. Abagyan, RA; Totrov, MM; Kuznetsov, DA. ICM: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation. *J Comp Chem*., 1994, 15, 488-506.
54. McMartin, C; Bohacek, RS. QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des*., 1997, Jul, 11(4), 333-44.
55. Friesner, RA; Banks, JL; Murphy, RB; Halgren, TA; Klicic, JJ; Mainz, DT; et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*., 2004, Mar 25, 47(7), 1739-49.
56. McGann, MR; Almond, HR; Nicholls, A; Grant, JA; Brown, FK. Gaussian docking functions. *Biopolymers*, 2003, Jan, 68(1), 76-90.
57. Cecchini, M; Kolb, P; Majeux, N; Caflisch, A. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J Comput Chem*., 2004, Feb, 25(3), 412-22.
58. Kang, L; Li, H; Jiang, H; Wang, X. An improved adaptive genetic algorithm for protein-ligand docking. *J Comput Aided Mol Des*., 2009, Jan, 23(1), 1-12.
59. Fuhrmann, J; Rurainski, A; Lenhof, HP; Neumann, D. A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J Comput Chem*., Jan 15.
60. Todman, SJ; Halling-Brown, MD; Davies, MN; Flower, DR; Kayikci, M; Moss, DS. Toward the atomistic simulation of T cell epitopes automated construction of MHC: peptide structures for free energy calculations. *J Mol Graph Model*, 2008, Feb, 26(6), 957-61.
61. Alder, BJ; Wainwright, TE. Phase Transition for a Hard Sphere System. *J Chem Phys*., 1957, 27.
62. Rahman, A. Correlations in the Motion of Atoms in Liquid Argon. *Phys Rev A*, 1964, 136, 405-11.
63. Stillinger, FH; Rahman, AJ. Propagation of sound in water. A molecular-dynamics study. *Chem Phys*., 1974, 60.
64. McCammon, JA; Gelin, BR; Karplus, M. Dynamics of folded proteins. *Nature*, 1977, Jun 16, 267(5612), 585-90.
65. Duan, Y; Kollman, PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 1998, Oct 23, 282(5389), 740-4.
66. Rognan, D; Scapozza, L; Folkers, G; Daser, A. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry*, 1994, Sep 27, 33(38), 11476-85.
67. Zacharias, M; Springer, S. Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study. *Biophys J*, 2004 Oct, 87(4), 2203-14.
68. Sieker, F; Springer, S; Zacharias, M. Comparative molecular dynamics analysis of tapasin-dependent and -independent MHC class I alleles. *Protein Sci*., 2007, Feb, 16(2), 299-308.
69. Wan, S; Flower, DR; Coveney, PV. Toward an atomistic understanding of the immune synapse: large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Mol Immunol*., 2008, Mar, 45(5), 1221-30.
70. Su, ZY; Wang, YT. A Molecular Dynamics Simulation of the Human Lysozyme - Camelid VHH HL6 Antibody System. *Int J Mol Sci*., 2009, Apr, 10(4), 1719-27.
71. Perryman, AL; Lin, JH; McCammon, JA. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci*., 2004, Apr, 13(4), 1108-23.
72. Salimi, N; Fleri, W; Peters, B; Sette, A. Design and utilization of epitope-based databases and predictive tools. *Immunogenetics*, Apr, 62(4), 185-96.
73. Vita, R; Zarebski, L; Greenbaum, JA; Emami, H; Hoof, I; Salimi, N; et al. The immune epitope database 2.0. *Nucleic Acids Res*., Jan, 38(Database issue), D854-62.

74. Rammensee, H; Bachmann, J; Emmerich, NP; Bachor, OA; Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 1999, Nov, 50(3-4), 213-9.
75. Lefranc, MP; Giudicelli, V; Kaas, Q; Duprat, E; Jabado-Michaloud, J; Scaviner, D; et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*., 2005, Jan 1, 33(Database issue), D593-7.
76. Korber, BTM; Brander, C; Haynes, BF; Koup, R; Moore, JP; Walker, BD; et al. HIV molecular immunology 2006/2007. Los Alamos, New Mexico, *Los Alamos National Laboratory*, 2007.
77. Toseland, CP; Clayton, DJ; McSparron, H; Hemsley, SL; Blythe, MJ; Paine, K; et al. AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res*., 2005, Oct 6, 1(1), 4.
78. Reche, PA; Zhang, H; Glutting, JP; Reinherz, EL. EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics*, 2005, May 1, 21(9), 2140-1.
79. Lata, S; Bhasin, M; Raghava, GP. MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res Notes*., 2009, 2, 61.

Capítulo III

Immunogenic Epitopes of Hantaviruses' N Protein are Restricted to Conserved Regions

3. IMMUNOGENIC EPITOPES OF HANTAVIRUSES' N PROTEIN ARE RESTRICTED TO CONSERVED REGIONS

ARTIGO COMPLETO SUBMETIDO PARA A REVISTA *FRONTIERS IN BIOSCIENCE*

Immunogenic epitopes of Hantaviruses' N protein are restricted to conserved regions

Mauricio Menegatti Rigo[1], Dinler Amaral Antunes[1], Samuel Paulo Cibulski[1], Marialva Sinigaglia[1], Jose Artur Bogo Chies[1], Gustavo Fioravanti Vieira[1]

[1]Nucleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500 - Prédio 43323, Porto Alegre, RS, Brasil

TABLE OF CONTENTS

Corresponding author:

Gustavo Fioravanti Vieira
Department of Genetics, UFRGS
Av. Bento Gonçalves, 9500 – P.O.Box 15053 – Zip Code 91501-970
Porto Alegre, RS, Brazil.
Phone: +55 51 3308 9938
Fax: +55 51 3308 7311
E-mail: gusfioravanti@yahoo.com.br

1. ABSTRACT

The *Bunyaviridae* virus family is composed by five genera, of which the *Hantavirus* genus is one of the most important representatives. Occasionally, these viruses can be transmitted to humans, giving rise to severe diseases that present high mortality rates. We analysed the amino acid sequences of the nucleocapsid (N) proteins of 34 different hantaviruses to investigate the potential mechanisms involved in immunogenicity against hantaviruses. Immunogenic epitopes described in the literature through experimental analyses for Sin Nombre (SNV), Puumala (PUUV), and Hantaan (HTNV) viruses' species were retrieved. We identified and characterized the regions believed to be responsible for the induction of immune response in hosts. We found that N protein epitopes described in the literature for PUUV, SNV and HTNV viruses are all located in highly conserved regions of the protein. The high conservation of these regions suggests that a cross-reactive immune response among different hantaviruses can be induced.

2. INTRODUCTION

The *Bunyaviridae* family includes more than 300 viral species distributed as five genera (1), and most of these viruses are transmitted by arthropods. Members of the *Hantavirus* genus, first isolated in 1978, are enveloped viruses with 80-140 nm virions and are the only non-arbovirus group, i.e. not transmitted by arthropods, in the *Bunyaviridae* family (2). Hantaviruses genome consists of three negative strand RNA segments: L (large), M (medium), and S (small) (3). The L segment encodes an RNA-dependent RNA polymerase; the M segment encodes two glycoproteins (Gn and Gc) inserted in the viral envelope membrane; the S segment encodes a nucleocapsid protein (~50 kDa) also referred to as N protein (4). Because the N protein is an internal virion component, it suffers less selective pressure from immune responses than the envelope glycoproteins (5) what could explain the high structural conservation of hantavirus N protein (6). Also, the N protein is highly abundant during the course of hantavirus infection, (7) and cross-reactive immune responses have already been observed in different hantaviruses species (8-12).

In humans, hantaviruses can cause hemorrhagic fever with renal syndrome (HFRS) and hantavirus pulmonary syndrome (HPS), which can reach case fatality rates even higher than 50% (13). HFRS is more prevalent in Asia and Europe, while HPS occurs mainly in the Americas (14, 15). Although efforts have been made in the last years (16), no effective and safe vaccines have been developed until now (17).

In order to understand hantavirus immunogenicity we searched the literature for epitopes and N protein sequences from different hantaviruses species. Immunogenic epitopes from three different hantaviruses species – Sin Nombre virus (SNV), Puumala virus (PUUV) and Hantaan virus (HTNV) – all of which described through experimental analyses, such as ELISPOT and chromium-release assay, were retrieved from Immune Epitope Data Bank (IEDB) and analysed. To investigate if these regions are specific to the *Hantavirus* genus, we created a motif of each region and searched a non-redundant protein databank for similar proteins. Also, we analysed the entire N protein through sliding window method and compared to proteins derived from rodents. This would reveal similarities between self (rodent) and non-self (the N protein) proteins and would allow to examine whether the immune system chooses regions in the N protein that are different from the host, consequently minimizing the risks of autoimmune episodes. Finally, taking into account the immunological differences between humans and rodents, we submitted the N protein to the human antigen processing pathway, in order to suggest the use of such epitopes in a human vaccine.

3. MATERIALS AND METHODS

3.1. Search for hantaviruses epitopes described in the literature

A search in the literature for immunogenic epitopes derived from hantaviruses nucleocapsid protein was performed. Only epitopes that have been proved to elicit a predominantly CD8+ T cell and strong murine immune response *in vivo* or *in vitro* assays were selected and included in our analysis. A total of nine epitopes from three different hantaviruses species – SNV, PUUV and HTNV – were retrieved (Table 1). Also, several human hantaviruses epitopes from different species were retrieved from Immune Epitope Data Bank (18) and colocalized in the alignment.

3.2. N protein search and alignment

All selected epitopes (8 to 10 amino acids in size) from murines (Table 1) were submitted to a *BLASTp* (19) in order to identify similar sequences. This approach allowed us to retrieve the N protein from several hantaviruses species. We used the default parameters of the software, except "*Word size*" (value = 2) and "*Max target sequences*" (value = 250), which were used to optimize the search. The results were filtered considering the original organism where the sequence was described. At this point of the analysis only non-redundant and complete N protein sequences were considered as valid results. Finally, 34 completed sequences from different species of the *Hantavirus* genus were retrieved and aligned using the Muscle software (20). The visualization and identification of conserved regions were performed using the GeneDoc software (http://www.psc.edu/biomed/genedoc). The 34 different species and their respective NCBI accession numbers are: Sin nombre (Accession No.: AAG03030.1), Puumala (Accession No.: CAB65379.1), Hantaan (Accession No.: AAW02944.1), Seewis (Accession No.: ABR24795.1), Cao Bang (Accession No.: ABR29825.1), Soochong (Accession No.: AAT78474.1), Amur (Accession No.: BAD04845.1), Saarema (Accession No.: CAE83602.1), Seoul (Accession No.: AAK96243.1), Thailand (Accession No.: CAL37107.1), Muju (Accession No.: AAZ67072.1), Topografov (Accession No.: CAB42097.1), Vladivostok (Accession No.: BAA25145.1), Tula (Accession No.: CAA11466.1), Malacky (Accession No.: CAA92340.1), Neembucu (Accession No.: ABC70873.1), Bermejo (Accession No.: AAL82648.1), Lechiguanas (Accession No.: AAL82649.1), Andes (Accession No.: AAK11225.1), Oran (Accession No.: AAL82650.1), Araucaria (Accession No.: AAW57482.1), Pergamino (Accession No.: AAL82652.1), Maciel (Accession No.: AAL82651.1), Alto Paraquay (Accession No.: ABC70872.1), Rio Mamore (Accession No.: AAC58450.1), Laguna Negra (Accession No.: AAB87601.1), Maporal (Accession No.: AAP92140.1), Choclo (Accession No.: ABB90557.1), Catacamas (Accession No.: ABB83548.1), Bayou (Accession No.: AAA61691.1), Muleshoe (Accession No.: AAD00082.1), Cano Delgadito (Accession No.: AAB71815.1), Limestone Canyon (Accession No.: AAK49899.1) and Thottapalayam (Accession No.: AAS19458.1).

3.3. Sliding Window Method

As already described (21), we used the sliding window method to compare the N protein from hantaviruses with self rodent proteins. The N protein sequence of SNV was used as reference. A sliding 10-residue window was set to generate a BLAST score against host sequences (*Rodentia* databank) to each N protein decamer. We assumed arbitrary values of up to 70%, 71-80%, and 81-100% for low, intermediate, and high similarity, respectively.

3.4. Pattern Search

To further analyze the corresponding regions of the described epitopes – four regions - we generated consensus motifs, based on the alignment of the N protein of the 34 hantaviruses members (Table 1). A search was performed with each motif as the entry sequence. In this search, all non-redundant proteins

available in the Swiss-prot data bank were investigated through the PatternSearch tool from the Max Planck Institute (http://toolkit.tuebingen.mpg.de/pat-search).

3.5. Epitope generation in humans

To investigate if the N protein epitopes described in the literature could be generated during an immune response in humans we performed an *in silico* approach simulating the steps of the antigenic processing pathway. We used tools for each specific steps to this pathway: proteasome cleavage - Netchop 3.0 (22); TAP transporting – TapPred (23) and MHC binding - NetMHC 3.2 (24). Based on the prevalence of the MHC alleles in different human populations we chose the following HLAs in NetMHC analysis: HLA-A*0201, HLA-A*0301, HLA-A*1101, and HLA-B*07.

3.6. Human Autoimmune Propensity Investigation

We performed a comparison between N protein epitopes and human proteins in order to investigate if these sequences could elicit a cross immune response against related self-proteins in humans. Therefore, a *BLASTp* against *Homo sapiens* protein data bank (NCBI) was performed. The *BLASTp* parameters were maintained as described above. Human proteins presenting similarity with the N protein epitopes higher than 75% were downloaded, and underwent the same approach as described in section 3.5.

4. RESULTS AND DISCUSSION

In the present work we analysed the amino acid sequence of the N protein from different hantaviruses species in order to investigate the potential mechanisms involved in hantaviruses immunogenicity. Immunogenic hantaviruses epitopes, described in the literature, were used as the starting point of our work, since they were defined by *in vitro* or/and *in vivo* experimental approaches, and therefore represent the immunogenic regions of the N protein that are more relevant in the induction of the immune response against hantaviruses in murines. In this context, our approach based on bioinformatics tools is greatly strengthened by the usage of experimental data.

Immunogenic peptides already described in murines were identified from three hantaviruses species: PUUV, SNV, and HTNV (Table 1) (12, 25). When these epitopes were submitted to a BLASTp search against the whole NCBI protein databank, we observed that 210 out of the first 250 output sequences presented high similarity degree with N protein sequences from *Hantavirus* genus members. This indicates that these sequences are characteristic of this viral group. This analysis allowed the identification of N protein sequences from 34 different members of the *Hantavirus* genus.

Considering that tables of alignments are able to highlight amino acid conservation patterns with high confidence, the 34 N protein amino acid sequences were aligned to provide an overview of N protein from various species of the Hantavirus genus. We found that murine epitopes are concentrated in more conserved regions (Figure 1). Contrastingly, human epitopes presented a more disperse localization along the hantavirus N protein. This is the first work to correlate hantaviruses conserved regions with immunogenic epitopes. This use of more disperse targets by humans can be involved in the fact that human immune responses against hantaviruses are not as efficient as murine immune responses. In this way, the identification of possible epitopes recognized by humans into the conserved regions of the N protein could

bypass the problems of such an inefficient response and would potentially be an interesting strategy in vaccine development.

An immune response directed against a region that presents low variability among different viruses can potentially allow the occurrence of cross-reactivity. In fact, Maeda *et al* (2004) already showed the existence of cross-reactivity directed against different regions of the N protein (N_{94-101}, $N_{180-188}$ and $N_{218-226}$) between the SNV and PUUV viruses (12). We believe that alterations in these regions affect negatively the viral infectivity or maintenance in the host and, therefore, although under a selective pressure as targets to immune responses, these regions do not show high variability among the different viruses species. In fact, it is known that the N_{1-125} region, which encompasses the $PUUV_{94-101}$ and SNV_{94-101} epitopes, is involved in protein oligomerization and that the region encompassing the $PUUV_{180-188}$ and $SNV_{180-188}$ epitopes is a RNA binding region (26).

The sliding window analysis evidenced that decamers superposing the epitope regions presented low similarity levels (up to 70%) with protein sequences from rodents, pointing towards a long-term coevolution between rodents immune system and hantaviruses. The SNV_{94-103} decamer, which contains the epitope SNV_{94-101}, presented less than 60% of similarity with any of the sequences from Rodentia. All the decamers that superposes the region of the epitope SNV_{94-101} [SNV_{85-94}; SNV_{86-95}; …; $SNV_{101-110}$] also showed low similarities with rodent proteins, reaching a maximum value of 70%. Similar results were obtained to the $SNV_{180-189}$, $SNV_{218-227}$, and $SNV_{328-335}$ regions.

To investigate if the immunogenic epitopes from N protein described in table 1 are specific to hantaviruses species we created a consensus motif of each region based on the alignment of the N protein of the 34 hantaviruses members. Each consensus motif was used as an input sequence against all sequences in the GenBank through the PatternSearch tool (27). These results are summarized in Table 2. The outputs from the PatternSearch tool suggest that the consensus motifs are characteristic (or even exclusive) of the Hantavirus group. The only exception was the consensus motif for the region 221-228. It is important to point out that this region is less conserved than the other N protein regions, as can be observed both in Figure 1 and Table 1, and contains proteins from different sources, ranging from organisms like bacteria and fungus to proteins from *Homo sapiens* (less than 1%).

Considering the development of a vaccine to be used in humans, we assessed the possibility that the N protein epitopes described in the literature could be generated during an immune response in humans through the intracellular antigen processing pathway. We submitted the original protein sequences from SNV, PUUV and HTNV to immunoinformatic tools (28, 29). First, we analysed the complete N protein sequences with Netchop 3.0 (22), which simulates the immunoproteasome cleavage. Subsequently we submitted the resulting epitopes to TapPred (23), in order to evaluate the peptide binding to the transporter associated with antigen processing (TAP). Finally, we evaluated the epitope binding affinity for the high prevalent human alleles HLA-A*0201, HLA-A*0301, HLA-A*1101, and HLA-B*07, using the NetMHC 3.2 (24).

The analysis of the *in silico* proteasomal cleavage (Netchop) and TAP affinity (TapPred) showed that, although the correct requirements for proteasomal cleavage were present in some murine epitopes of the N protein, only the epitope SVIGFLALA ($N_{221-229}$) had all the predicted requirements to be generated and presented for a T lymphocyte in humans. All results are presented in Table 3. This suggests that immunogenic epitopes for the murine immune system will not necessarily be generated in humans. Such differences can be explained by the different interactions of the murine and the human immune systems to hantaviruses. Considering that rodents are the natural hosts for hantaviruses, we can envisage the superposition of two scenarios, where the murine immune system efficiently selects good targets to minimize the morbidity associated to the infection and simultaneously selects strains that present low lethality rates, therefore favoring hantaviruses maintenance (30).

After the selection of immunogenic epitopes (SVIGFLALA - $N_{221-229}$) that can be generated and presented by the human immune system, we performed a second step – the search for similar human sequences that can trigger cross-reactive events and the propensity to autoimmune disorders. When we compared the hantavirus immunogenic epitope (SVIGFLALA - $N_{221-229}$) with all human proteins available at GenBank databank, only one protein (RTN1) presented a region (SVVAYLALA - $RTN1_{548-556}$) with similarity to the foreign epitope higher than 80%. The RTN1 protein was submitted to imunoinformatics tools, which simulate the complete antigen processing pathway, in order to evaluate if the human sequence SVVAYLALA 548-556 will be generated and presented by the whole process. Although the sequence SVVAYLALA has a high affinity by TAP, the protein RTN1 has no predict cleavage site flanking the region, which render more difficult the generation of this epitope. Besides, the sequence only can be presented by the HLA-A*0201 allele, as a weak ligand.

In conclusion, we postulate that the murine immune system is preferentially directed against hantaviruses epitopes from conserved regions and that do not present a significant identity to self proteins. This phenomenon minimizes the risk of autoimmune responses. This idea is supported by our results from the comparison between the N protein and proteins from the *Rodentia* family (the natural hosts of hantaviruses) obtained through the sliding window method. Also, although hantaviruses use mechanisms to evade immune responses (as any other pathogen), the potential to change some protein regions is not unlimited, since this alteration can negatively affect the viral infectivity or maintenance in the host. Therefore, although under a selective pressure as targets to immune responses, some proteins do not show high variability across different viruses and can therefore be considered potential candidates in vaccine development strategies. The development of reliable and effective vaccines is quite important, and to achieve such objective, we should understand the mechanisms underlying the selection of viral targets by the immune system. The high conservation of the immunogenic regions of hantaviruses, evidenced in our study, suggests that a cross-reactive immune response among different hantaviruses can be induced (31).

Also, the lack of similarity among the tested epitopes and humans points to these as potential targets for vaccine development, with reduced risk of adverse effects due to autoimmune responses.

6. ACKNOWLEDGMENT

This work was funded by a grant from the Bill & Melinda Gates foundation through the Grand Challenges Exploration Initiative. Supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil. No competing financial interests exist.

7. REFERENCES

1. N. Karabatsos: International catalogue of arboviruses including certain other viruses of vertebrates. In: *American Society of Tropical Medicine and Hygiene*. San Antonio (1985)

2. H. W. Lee, P. W. Lee and K. M. Johnson: Isolation of the etiologic agent of Korean Hemorrhagic fever. *J Infect Dis*, 137(3), 298-308 (1978)

3. C. S. Schmaljohn: Molecular biology of hantaviruses. In: *The Bunyaviridae*. Ed R. Elliott. Plenum Press, New York (1996)

4. C. B. Jonsson and C. S. Schmaljohn: Replication of hantaviruses. *Curr Top Microbiol Immunol*, 256, 15-32 (2001)

5. N. D. Tischler, M. Rosemblatt and P. D. Valenzuela: Characterization of cross-reactive and serotype-specific epitopes on the nucleocapsid proteins of hantaviruses. *Virus Res*, 135(1), 1-9 (2008)

6. A. Plyusnin, O. Vapalahti and A. Vaheri: Hantaviruses: genome structure, expression and evolution. *J Gen Virol*, 77 (Pt 11), 2677-87 (1996)

7. R. M. Elliott, M. Bouloy, C. H. Calisher, R. Goldbach, J. T. Moyer, S. T. Nichol, R. Pettersson, A. Plyusnin, and C. S. Schmaljohn: Bunyaviridae. In: *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*. Ed C. M. F. M. H. V. van Regenmortel, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, R. B. Wickner. Academic Press, San Diego (2000)

8. F. A. Ennis, J. Cruz, C. F. Spiropoulou, D. Waite, C. J. Peters, S. T. Nichol, H. Kariwa and F. T. Koster: Hantavirus pulmonary syndrome: CD8+ and CD4+ cytotoxic T lymphocytes to epitopes on Sin Nombre virus nucleocapsid protein isolated during acute illness. *Virology*, 238(2), 380-90 (1997)

9. H. Asada, M. Tamura, K. Kondo, Y. Dohi and K. Yamanishi: Cell-mediated immunity to virus causing haemorrhagic fever with renal syndrome: generation of cytotoxic T lymphocytes. *J Gen Virol*, 69 (Pt 9), 2179-88 (1988)

10. H. Asada, K. Balachandra, M. Tamura, K. Kondo and K. Yamanishi: Cross-reactive immunity among different serotypes of virus causing haemorrhagic fever with renal syndrome. *J Gen Virol*, 70 (Pt 4), 819-25 (1989)

11. H. L. Van Epps, C. S. Schmaljohn and F. A. Ennis: Human memory cytotoxic T-lymphocyte (CTL) responses to Hantaan virus infection: identification of virus-specific and cross-reactive CD8(+) CTL epitopes on nucleocapsid protein. *J Virol*, 73(7), 5301-8 (1999)

12. K. Maeda, K. West, T. Toyosaki-Maeda, A. L. Rothman, F. A. Ennis and M. Terajima: Identification and analysis for cross-reactivity among hantaviruses of H-2b-restricted cytotoxic T-lymphocyte epitopes in Sin Nombre virus nucleocapsid protein. *J Gen Virol*, 85(Pt 7), 1909-19 (2004)

13. J. F. Guilfoyle and A. J. Macnab: Hantavirus cardiopulmonary syndrome: implications for transport management and care. *Air Med J*, 27(6), 299-302 (2008) doi:S1067-991X(08)00183-1

14. S. T. Nichol, C. F. Spiropoulou, S. Morzunov, P. E. Rollin, T. G. Ksiazek, H. Feldmann, A. Sanchez, J. Childs, S. Zaki and C. J. Peters: Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science*, 262(5135), 914-7 (1993)

15. N. Lopez, P. Padula, C. Rossi, M. E. Lazaro and M. T. Franze-Fernandez: Genetic identification of a new hantavirus causing severe pulmonary syndrome in Argentina. *Virology*, 220(1), 223-6 (1996)

16. J. Li, K. N. Li, J. Gao, J. H. Cui, Y. F. Liu and S. J. Yang: Heat shock protein 70 fused to or complexed with hantavirus nucleocapsid protein significantly enhances specific humoral and cellular immune responses in C57BL/6 mice. *Vaccine*, 26(25), 3175-87 (2008)

17. M. Zeier, M. Handermann, U. Bahr, B. Rensch, S. Muller, R. Kehm, W. Muranyi and G. Darai: New ecological aspects of hantavirus infection: a change of a paradigm and a challenge of prevention--a review. *Virus Genes*, 30(2), 157-80 (2005)

18. B. Peters, J. Sidney, P. Bourne, H. H. Bui, S. Buus, G. Doh, W. Fleri, M. Kronenberg, R. Kubo, O. Lund, D. Nemazee, J. V. Ponomarenko, M. Sathiamurthy, S. Schoenberger, S. Stewart, P. Surko, S. Way, S. Wilson and A. Sette: The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol*, 3(3), e91 (2005)

19. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman: Basic local alignment search tool. *J Mol Biol*, 215(3), 403-10 (1990)

20. R. C. Edgar: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5), 1792-7 (2004)

21. L. Berglund, J. Andrade, J. Odeberg and M. Uhlen: The epitope space of the human proteome. *Protein Sci*, 17(4), 606-13 (2008)

22. M. Nielsen, C. Lundegaard, O. Lund and C. Kesmir: The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, 57(1-2), 33-41 (2005)

23. M. Bhasin and G. P. Raghava: Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci*, 13(3), 596-607 (2004)

24. C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund and M. Nielsen: NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res*, 36(Web Server issue), W509-12 (2008)

25. J. M. Park, S. Y. Cho, Y. K. Hwang, S. H. Um, W. J. Kim, H. S. Cheong and S. M. Byun: Identification of H-2K(b)-restricted T-cell epitopes within the nucleocapsid protein of Hantaan virus and establishment of cytotoxic T-cell clones. *J Med Virol*, 60(2), 189-99 (2000)

26. P. Kaukinen, A. Vaheri and A. Plyusnin: Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties. *Arch Virol*, 150(9), 1693-713 (2005)

27. A. Biegert, C. Mayer, M. Remmert, J. Soding and A. N. Lupas: The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res*, 34(Web Server issue), W335-9 (2006)

28. B. Korber, M. LaBute and K. Yusim: Immunoinformatics comes of age. *PLoS Comput Biol*, 2(6), e71 (2006)

29. J. C. Tong and E. C. Ren: Immunoinformatics: Current trends and future directions. *Drug Discov Today*, 14(13-14), 684-9 (2009)

30. J. D. Easterbrook and S. L. Klein: Immunological mechanisms mediating hantavirus persistence in rodent reservoirs. *PLoS Pathog*, 4(11), e1000172 (2008)

31. M. M. Rigo, D. A. Antunes, G. F. Vieira and J. A. B. Chies: MHC: Peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein. *Lecture Notes in Computer Science*, 5676, 160-163 (2009)

Figure legends

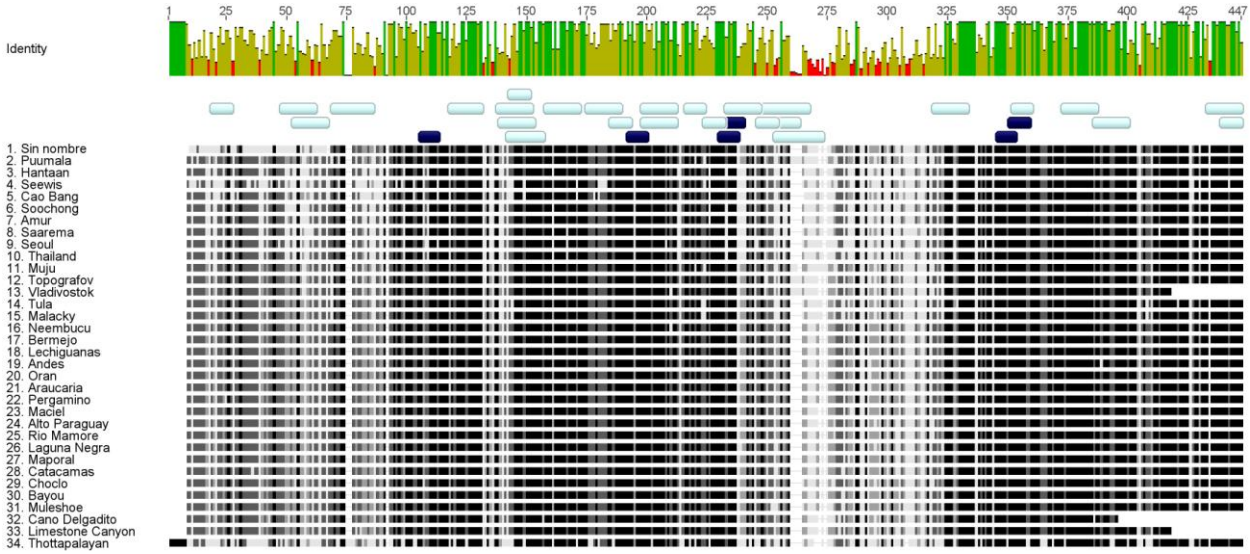

Figure 1. Multiple alignment of sequences, evidencing the conserved regions pattern. Alignment of N protein amino acid sequences from 34 different hantaviruses species. Each box, above the alignment, represents the screened hantaviruses epitopes in murines (dark blue boxes) or in humans (soft blue boxes).

Tables and table legends

Table 1. Immunogenic epitopes for each N protein region.

Specie	Sequence[1]	Region[1]	Reference
SNV	^{94}SSLRYGNV^{101}	94 – 101	Maeda et al., 2004
PUUV	^{94}SSLRYGNV^{101}		
SNV	^{180}SMPTAQSTM^{188}	180 -188	Maeda et al., 2004
PUUV	^{180}SMPTAQSTM^{188}		
SNV	^{218}PVMGVIGFS^{226}	218 – 226	Maeda et al., 2004
PUUV	^{218}PVMGVIGFS^{226}		
SNV	^{332}AILQDMRNT^{340}	332 – 340	Maeda et al., 2004
HTNV	^{221}SVIGFLAL^{228}	221 – 228	Park et al., 2000
HTNV	^{328}LGAFFSIL^{335}	328 - 335	Park et al., 2000

Abbreviations: SNV: Sin Nombre virus; HTNV: Hantaan virus; PUUV: Puumala virus.

[1] The numbers refer to the first and the last amino acid of the sequence.

Table 2. Output sequences containing the consensus motif for each N protein region.

Motif	Region	Output sequences containing the motif
		All sequences on GenBank / only hantaviruses sequences
S [SMAT] [LP] [RS] Y G N [VTI]	94 - 101	391/389 (99.5%)
S [ML] P [TN] [AS] Q S [TS] [MI]	180 - 188	490/489 (99.8%)
P V M [GS] V [IV] G F [SLPMAGNQ]	218 - 226	557/553 (99.3%)
[AS] [IVL] [LM] Q D [MI] R [NG] [TG]	332 - 340	337/335 (99.4%)
[SG] V [IV] G F [SLPMAGNQ] [FAVSH] [FLI]	221 - 228	917/543 (59.2%)
L G A F [FL] [SA] [IVL] [LM]	328 - 335	435/369 (84.8%)

Table 3. TapPred and NetChop predictions for TAP binding and C-terminal precise cutting, respectively

Species	**Sequence**	**NetChop (C-terminal Precise Cutting)**	**TapPred (Predicted Affinity)**	**NetMHC affinity(nM)[1]**			
				HLA-A*0201	**HLA-A*0301**	**HLA-A*1101**	**HLA-B*0702**
SNV	^{91}KSSLRYGNV^{99}	0.114010	High	15089	17381	19792	22930
SNV	^{92}SSLRYGNVL^{100}	0.925992	Intermidate	18668	22460	21336	10917
SNV	^{178}SMPTAQSTM^{186}	0.977767	Intermidate	4221	20782	24776	18394
SNV	^{216}PVMGVIGFS^{224}	0.036510	Low or undetectable	20904	19477	24749	24774
SNV	^{330}AILQDMRNT^{338}	0.024778	High	17525	22468	29295	25590
PUUV	^{93}RSSLRYGNV^{101}	0.111766	High	17364	17202	19128	20101
PUUV	^{94}SSLRYGNVL^{102}	0.869308	Intermidate	18668	22460	21336	10917
PUUV	^{180}SMPTAQSTM^{188}	0.976805	Intermidate	4221	20782	24776	18394
PUUV	^{218}PVMGVIGFS^{226}	0.034059	Low or undetectable	20904	19477	24749	24774
HTNV	^{220}MSVIGFLAL^{228}	0.964342	Intermidate	8573	18622	17295	4695
HTNV	^{221}SVIGFLALA^{229}	0.707558	High	113 (Weak Binding)	16908	4641	19615
HTNV	^{327}ELGAFFSIL^{335}	0.870844	Low or undetectable	6608	20665	25331	22379
HTNV	^{328}LGAFFSILQ^{336}	0.033648	Intermidate	22791	19690	22075	27706
Homo sapeins (RNT1)	^{548}SVVAYLALA^{556}	0.396658	High	123 (Weak Binding)	16343	9098	23252

[1] Affinity as IC50 value in nM. The value is inversely proportional to the binding affinity.

Abbreviations

BLAST: Basic Local Alignment Search Tool, ELISPOT: Enzyme-linked Immunosorbent Spot, HCPA: Hantavirus Pulmonary Syndrome, HLA-A: Human Leukocyte Antigen, HFRS: Hemorrhagic Fever with Renal Syndrome, HTNV: Hantaan virus, IEDB: Immune Epitope Database, L: Large, M: Medium, MHC: Major Histocompatibility Complex, N: Nucleocapsid, PUUV: Puumala virus, S: Small, SNV: Sin Nombre virus, RNA: Ribonucleic Acid, TAP: Transporter Associated with Antigen Processing.

Key Words: *Hantavirus* genus, immunogenicity, *Bunyaviridae* family, nucleocapsid protein, immunoinformatics.

Send correspondence to: Dr. Gustavo Fioravanti Vieira, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500 – Prédio 43323, P.O. BOX 15053, Porto Alegre, Brazil. Tel: +55-51-3308-9938, Fax: +55-51-3308-7311, E-mail: gusfioravanti@yahoo.com.br

Running title: Immunogenic epitopes of Hantaviruses' N protein

Capítulo IV

MHC:Peptide Analysis: Implications On The Immunogenicity Of Hantaviruses' N Protein

4. MHC:PEPTIDE ANALYSIS: IMPLICATIONS ON THE IMMUNOGENICITY OF HANTAVIRUSES' N PROTEIN

ARTIGO COMPLETO PUBLICADO NA REVISTA *LECTURE NOTES IN COMPUTER SCIENCE*

MHC:peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein.

Maurício Menegatti Rigo[1], Dinler Amaral Antunes[1], Gustavo Fioravanti Vieira[1], José Artur Bogo Chies[1],

[1] Núcleo de Bioinformática, Department of Genetics, UFRGS, Porto Alegre, Brazil
{mauriciomr1985, dinler}@gmail.com, gusfioravanti@yahoo.com.br, jabchies@terra.com.br

Abstract. Hantaviruses, members of the *Bunyaviridae* family, are enveloped negative-stranded RNA viruses with tripartite genomes – S, M and L. The S genome codes for a nucleocapsid (N) protein, which is quite conserved among different species of the hantavirus genus and possess a recognized immunogenicity. In this work we analyzed the sequence of two regions in this protein (N_{94-101} and $N_{180-188}$), which presents T cell epitopes for two species of hantaviruses – Sin Nombre and Puumala. Interestingly, the same region has no described epitopes for Hantaan virus, despite its similarity. A study using a bioinformatic approach for the construction of MHC:peptide complexes was performed to detect any variation on the cleft region that could explain such degrees of immunogenicity. Our results shown topological and charges differences among the constructed complexes.

Keywords: MHC-I, epitopes, Hantavirus, molecular docking

1 Introduction

As members of the *Bunyaviridae* family, hantaviruses have a tripartite ssRNA(−) genome coding for a RNA-dependent RNA polymerase, two glycoproteins which are inserted into the viral envelope membrane, and the N protein associated with the viral genome [1]. The hantavirus nucleocapsid (N) protein fulfills several key roles in virus replication and assembly [1]. Also, it presents cross-reactivity among different members of the hantavirus genus [2]. Considering the recognized immunogenicity of this protein and that the final objective of our work is the development of a vaccine, two aspects should be here considered: the analysis of antigenic processing pathway and the presentation of epitopes to T cell on the MHC class I context. In a previous work (personal communication) we observed a concordance between conserved regions in the N protein and epitopes of hantaviruses described in literature [2]. Maeda *et al*. described epitopes for Puumala virus (PUU) and Sin Nombre virus (SNV) at region N_{94-101} and $N_{180-188}$, however epitopes for Hantaan virus (HTN) in the same region were not observed, although the high similarity between these sequences (75% and 67%, respectively). Additionally, *in silico* simulations showed that epitopes of HTN in these regions can be generated by the antigen processing pathway in the same way as PUU and SNV. This data suggests a strong influence from some amino acids of the epitopes on the TCR recognition and on immune response induction. Therefore, the construction of a MHC:peptide complex for topological and charge analysis of these regions is quite important.

In the present work we analyzed two specific regions of the N protein (94-101/180-188) from three different species of hantaviruses – PUU, SNV and HTN. This analysis was made on modeled MHC:peptide complexes, mainly on the TCR surface contact, searching for charges and topological differences that could explain the different levels of immunogenicity observed among the epitopes from different hantaviruses.

2 Material & Methods

The epitope peptide sequences related to N_{94-101} and $N_{180-188}$ region from PUU, SNV and HTN were obtained from literature and written in the FASTA format. The crystal structure of murine MHC alleles, H2-K^b (1G7P) and H2-D^b (1WBY), were used to modeling the complexes of interest. Each peptide sequence was fitted on the parameters of the specific MHC allele epitope pattern using the SPDBV program [3]. Since we have the

amino acids sequences within the parameters of three-dimensional shape, energy minimization was performed with the GROMACS 3.3.3 program [4] to stabilize the molecule, simulating an aqueous solution. The next step was a construction of the MHC:peptide complex using AutoDock 4.0 program [5]. After the construction of the first complex, a second energy minimization was performed aiming to a better interaction of the side chains of the MHC with the epitope. The MHC was separated from its epitope and a second docking was carried out for the construction of the final MHC:peptide complex. The surface of the resulting complex was visualized with the GRASP2 program [6] (Figure 1), where the electrostatic charge distribution and the shape were analyzed.

3 Results

All FASTA sequences were perfect fitted with PDB allele-specific peptides. Information about the positions of the anchor residues for H2-K^b and H2-D^b was obtained from SYFPEITHI [7], a MHC-ligand databank. Epitopes described in literature for SNV and PUU at the studied regions have the same sequence, therefore they were analyzed as a unique epitope. After two rounds of molecular docking and energy minimization we found good values of binding energy (BE) for each region (Table 1). At the topological and charge levels, categorical differences were found in MHC:peptide complex for $N_{94\text{-}101}$ region, on the discordant residues. These differences could be important in TCR recognition. The $N_{180\text{-}188}$ region showed almost none differences at charge distribution, but a topological discrepancy was verified.

Table 1: Best binding energy values for each region of Hantavirus species.

Hantavirus specie	Protein/Region	Peptide Sequence*	MHC allele	Best BE-1st docking (Kcal/mol)	Best BE - 2nd docking (Kcal/mol)
Hantaan	$N_{94\text{-}101}$	SMLS<u>Y</u>GN<u>V</u>	H2-K^b	-6,11	-6,48
Sin Nombre & Puumala	$N_{94\text{-}101}$	SSLR<u>Y</u>GN<u>V</u>	H2-K^b	-4,98	-7,49
Hantaan	$N_{180\text{-}188}$	SLPN<u>A</u>QSS<u>M</u>	H2-D^b	-11,07	-14,73
Sin Nombre & Puumala	$N_{180\text{-}188}$	SMPT<u>A</u>QST<u>M</u>	H2-D^b	-12,01	-14,55

*Anchor amino acids are underlined.

4 Discussion

The described epitopes for $N_{94\text{-}101}$ region are presented by H2-K^b allele, while $N_{180\text{-}188}$ are presented by H2-D^b allele [2]. There are a difference of only two amino acids at $N_{94\text{-}101}$ region between SNV/PUU and HTN; at $N_{180\text{-}188}$ region, the difference is 3 amino acids. These substitutions seems to affect the recognition by the immune system, since there are no epitopes described for hantavirus in that region [2].

A first approach analyzed the binding energy value of molecular docking with peptides from these two regions of SNV/PUU and HTN. The computational program used was AutoDock. Values for binding energy were all negative. There are no data in literature about the best value for binding energy of peptide and MHC. However, this value is directly proportional to the entropy of the system. Thus, we admitted that the lowest value should be the best value. Accordingly to Binding energy data, all peptides have a good potential for attachment to the MHC. The conservation of anchor amino acids among the sequences for each MHC restriction could explain these good binding values.

It is known that changes in the distribution of charges interfere with the TCR recognition [8]. The best MHC:peptide models were visualized and analyzed at GRASP program, which provides a molecular surface with charge distribution, through the adjust of electric potential (Figure 1). The negative potentials are represented in red, while the positive potentials in blue. We could observe a topological and, especially, a

charge difference between the MHC:peptide complexes of $N_{94\text{-}101}$ region, mainly on the fourth (Ser/Arg) residue. The $N_{180\text{-}188}$ region showed only topological changes, and the charges seems to remain equally distributed in both complexes. Our results showed that more than just charge differences, topological differences could explain the abrogation of an immune response.

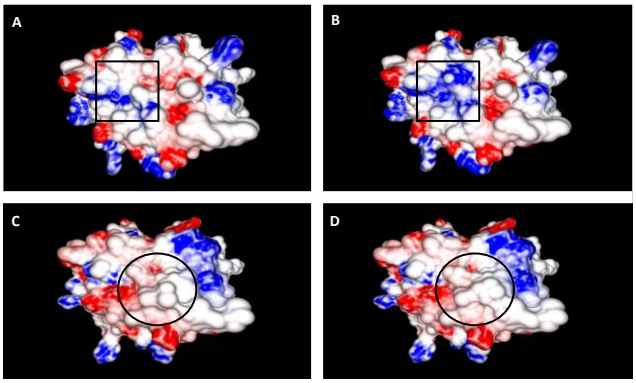

Fig. 1: Charge distribution of MHC:peptide complexes (Top-view) visualized at GRASP program. Peptide sequences from HTN (A) and SNV/PUU (B) on H2-Kb allele in N94-101 region; and from HTN (C) and SNV/PUU (D) on H2-Db allele in N180-188 region. Black squares show charge and topological differences between the first two complexes. Black circles show only topological difference on the remaining complexes.

References

1. Tischler, N.D., Rosemblatt, M., Valenzuela, P.D.T.: Characterization of Cross-reactive and Serotype-specific Epitopes on the Nucleocapsid Proteins of Hantaviruses. Virus Res. 135, 1--9 (2008)
2. Maeda, K., West, K., Toyosaki-Maeda, T., Rothman, A.L.: Identification and Analysis for Cross-reactivity Among Hantaviruses of H-2^b-restricted Cytotoxic T-lymphocyte Epitopes in Sin Nombre Virus Nucleocapsid Protein. J. Gen Virol. 85, 1909--1919 (2004)
3. Guex, N., Peitsch, M.C.: SWISS-MODEL and the Swiss-PdbViewer: an Environment for Ccomparative Protein Modeling. Electrophoresis 18, 2714-2723 (1997)
4. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.: GROMACS: Fast, Flexible, and Free. J. Comput. Chem. 26, 1701-1718 (2005)
5. Morris, G.M., Goodsell, D.S., Halliday R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J.: Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. J. Comput. Chem. 19, 1639--1662 (1998)
6. Petrey, D., Honig, B.: GRASP2: Visualization, Surface Properties, and Electrosttics of Macromolecular Structures and Sequences. Meth. Enzymol. 374, 492--509 (2003)
7. Rammensee, H., Bachmann, J., Emmerich, N.N., Bachor, O.A., Stevanovic, S.: SYFPEITHI: Database for MHC Ligands and Peptide motifs. Immunogenetics. 50, 213--219 (1999)
8. Kessels, H.W., de Visser, K.E., Tirion, F.H., Coccoris, M., Kruisbeek, A.M., Schumacher, T.N.: The Impact of Self-tolerance on the Polyclonal CD8+ T-cell Repertoire. J. Immunol. 172, 2324--2331 (2004)

Capítulo V

Discussão

5. DISCUSSÃO

Este trabalho se propôs a estudar os mecanismos envolvidos na apresentação cruzada de epitopos de hantavírus, presentes na proteína de nucleocapsídeo, bem como discutir aspectos imunogênicos gerais, tendo como base o uso de ferramentas de imunoinformática.

Nesse contexto, o segundo capítulo desta dissertação foi organizado como uma revisão da apresentação endógena de peptídeos virais e das principais ferramentas de imunoinformática disponíveis até o momento e que auxiliam, de uma forma ou outra, a entender e analisar processos imunológicos oriundos de dados *in vitro/in vivo*. A imunoinformática também pode ser utilizada para traçar o direcionamento de um projeto de pesquisa, economizando tempo, dinheiro e suplementos laboratoriais, antes de iniciar os experimentos em bancada. Deste modo, este novo ramo da ciência surge como uma ferramenta do imunologista moderno, o qual deve dividir o seu tempo entre a administração e manutenção do laboratório e a realização, da forma mais adequada possível, dos projetos de pesquisa.

A partir dessa contextualização da imunogenicidade, direcionamos o foco do estudo para o gênero *Hantavírus*. Sabe-se que a infecção por hantavírus em humanos pode ser letal, principalmente em decorrência da resposta celular exacerbada, mediada por linfócitos T $CD8^+$ (Khaiboullina and St Jeor, 2002). O mesmo não acontece em roedores reservatórios, onde embora o sistema imunológico não consiga eliminar totalmente o vírus, este se mantém em níveis não letais para o hospedeiro na maioria das vezes, fruto de uma co-evolução parasita-hospedeiro de muitos anos (Easterbrook and Klein, 2008b).

Tendo este cenário como plano do fundo, e sabendo da importância da proteína N na resposta imunológica associada aos hantavírus, tanto em roedores como em humanos, o terceiro capítulo desta dissertação discorre sobre a análise da seqüência da proteína N de 34 espécies do gênero *Hantavírus*. Seqüências completas e não redundantes foram obtidas na literatura, a partir do banco de dados do NCBI (do inglês "National Center for Biotechnology Information"). Ainda, epitopos murinos e humanos presentes na proteína N, reconhecidamente imunogênicos, foram obtidos a partir do IEDB (do inglês, "Immune

Epitope Data Bank"), banco de dados de epitopos de células T e células B (Vita *et al.*, 2010).

Foram encontrados, em murinos, 9 epitopos oriundos da proteína N, pertencentes às espécies SNV, PUUV e HTNV de hantavírus. Por outro lado, foram encontrados muito mais epitopos reconhecidos por linfócitos T citotóxicos humanos, de diferentes espécies de hantavírus. O alinhamento das seqüências da proteína N nos permitiu observar áreas de conservação de aminoácidos. Ao co-localizar os epitopos murinos sobre o alinhamento, sua disposição sobre áreas de elevada conservação ficou evidente, ao contrário dos epitopos gerados e reconhecidos pelo sistema imunológico humano, os quais apresentam um padrão de distribuição dispersa ao longo da proteína N, se situando tanto sobre regiões conservadas quanto regiões não conservadas.

A elevada atividade pró-inflamatória decorrente da infecção por hantavírus em humanos é necessária para que ocorra a eliminação do agente viral, mas isso ocorre à custa de um elevado dano tecidual (Easterbrook and Klein, 2008b). Já foi observado que em roedores reservatórios, entretanto, a resposta imunológica contra hantavírus em células pulmonares pode ser suprimida, através da redução de moléculas envolvidas na defesa antiviral, tais como IFN-beta e IFN-gamma, e de moléculas pró-inflamatórias, diminuindo o dano tecidual naquela região (Easterbrook *et al.*, 2007; Hannah *et al.*, 2008). Porém, a resposta na periferia continua ocorrendo de forma sítio-específica, pois já foi demonstrado que a produção de moléculas pró-inflamatórias e de fatores antivirais no baço é constante na fase aguda da infecção pela espécie Seoul, do gênero *Hantavírus* (Easterbrook and Klein, 2008a; Klein *et al.*, 2001). Isso parece ser um dos fatores contribuintes para que roedores reservatórios não sejam acometidos pela doença da mesma forma como ocorre em humanos. Além disso, estudos com roedores infectados demonstraram que há um aumento significativo de células T regulatórias, as quais estão diretamente envolvidas no controle e regulação da infecção, impedindo a manifestação da doença (Easterbrook *et al.*, 2007). Nossa hipótese, portanto, é a de que a inflamação exacerbada não ocorreria se o sistema imunológico humano conseguisse suprimir a resposta contra hantavírus de forma sítio-específica e, ao mesmo tempo, direcionar a resposta por linfócitos T $CD8^+$ contra alvos específicos, assim como parece estar ocorrendo entre roedores reservatórios.

A utilização de epitopos de hantavírus que estimulam a resposta em murinos para incitar uma resposta celular direcionada parece ser uma abordagem atraente. Não apenas pelo direcionamento da resposta em si, mas também pelo fato de que tais epitopos se encontram em regiões conservadas. Além disso, já foi demonstrado que tais regiões são importantes para a manutenção do vírus no hospedeiro, diminuindo assim a probabilidade de que organismos que tenham sofrido mutações nessas regiões permaneçam viáveis e consigam se multiplicar (Kaukinen *et al.*, 2005). Isso aumenta a chance de que através da utilização de um ou mais epitopos específicos de uma espécie de hantavírus, possamos gerar reatividade cruzada contra epitopos de outras espécies de hantavírus.

Entretanto, para que a vacina seja eficaz a longo prazo, precisamos ter certeza de que tais epitopos serão gerados no contexto imunológico humano, ou seja, a proteína N precisa passar por toda a via intracelular antigênica e os epitopos devem ser capazes de se ligar ao MHC humano e serem apresentados para linfócitos T $CD8^+$. Ao mesmo tempo, deve-se ter o cuidado de escolher alvos virais que não tenham semelhança com proteínas do hospedeiro, o que poderia resultar em uma resposta autoimune originada por reatividade cruzada.

Tendo esses dois conceitos em mente, autoimunidade e geração de epitopos descritos em murinos no contexto do sistema imunológico humano, submetemos primeiramente as seqüências de proteína N encontradas (apenas de SNV, PUUV e HTNV, já que os epitopos encontrados são oriundos dessas espécies) às ferramentas de imunoinformática que simulam o corte proteassomal, o endereçamento para o retículo endoplasmático mediado por TAP e o grau de afinidade de ligação ao MHC. A partir das nossas análises *in silico*, dos 9 epitopos murinos, apenas um (SVIGFLALA – $N_{221-229}$) tem a capacidade de ser gerado através da via de processamento intracelular de peptídeos endógenos. Isso aponta para o fato de que, embora as famílias *Muridae* e *Hominidae* se encontrem filogeneticamente próximas, as interações imunológicas não são as mesmas. Na verdade, elas diferem sutilmente, pois foram moldadas com base na história de infecções pelas quais as duas famílias passaram ao longo do processo evolutivo (Mestas and Hughes, 2004).

O epitopo SVIGFLALA foi submetido a uma busca por similaridade com proteínas humanas, através do BLAST (do inglês “Basic Local Alignment Search Tool”). A maior

similaridade encontrada foi com a seqüência SVVAYLALA da proteína RTN-1, a qual é encontrada na membrana do retículo endoplasmático de células do sistema nervoso (Melino *et al.*, 2009). Entretanto, ao contrário do epitopo SVIGFLALA, a seqüência SVVAYLALA de RTN-1 não possui a capacidade de ser gerada pelo processamento intracelular endógeno para ser apresentada para linfócitos T $CD8^+$, tomando-se por base os estudos realizados *in silico*. Esse trabalho, portanto, fornece uma nova alternativa para o estudo da prevenção de patologias associadas ao gênero *Hantavírus*, fornecendo um novo alvo peptídico que pode ser utilizado em estudos *in vitro*/*in vivo*.

A partir do alinhamento das seqüências da proteína N, partimos para uma análise mais minuciosa das áreas conservadas. Duas regiões em particular, das espécies SNV, PUUV e HTNV, chamaram a atenção: as regiões compreendendo as seqüências que vão do aminoácido 94 ao aminoácido 101 (N_{94-101}) e do aminoácido 180 ao aminoácido 188 ($N_{180-188}$). Isso porque segundo Maeda e cols. (2004) o peptídeo oriundo da proteína N de HTNV, em ambas as regiões, não desencadeia uma resposta imunológica celular, apesar da elevada similaridade peptídica com os peptídeos de SNV e PUUV, os quais geram resposta celular por linfócitos T $CD8^+$ (Maeda *et al.*, 2004). Aqui vale ressaltar que a sequência de SNV e PUUV, para ambas as regiões, é a mesma. Sendo assim, no terceiro capítulo desta dissertação nos propomos a estudar, através de uma abordagem *in silico*, o mecanismo de resposta imunológica contra complexos pMHC apresentando as seqüências das regiões N_{94-101} e $N_{180-188}$ oriundas de HTNV, SNV e PUUV. Nosso objetivo principal, portanto, foi estudar os padrões moleculares na interface pMHC:TCR que possam estar envolvidos na inibição de uma resposta imunológica cruzada entre os peptídeos de SNV/PUUV e HTNV.

Como foi mencionado anteriormente, estima-se que um único TCR seja capaz de reconhecer mais de 10^6 combinações diferentes de pMHCs (Mason, 1998). Isso, por si só, já indica que peptídeos diferentes não precisam ter necessariamente elevada similaridade de seqüência para apresentarem reatividade cruzada entre si. Portanto, outras características moleculares do complexo pMHC devem ser levadas em consideração.

Sabe-se que o padrão topológico na interface de interação entre o pMHC e o TCR, por exemplo, bem como a distribuição de cargas decorrente da presença de aminoácidos com diferentes grupos químicos, podem influenciar na reposta imunológica cruzada (Kessels *et al.*, 2004). Também é sabido que o compartilhamento de características físico-

químicas entre diferentes epitopos também é importante, tanto quanto (ou mais que) a similaridade na estrutura primária de aminoácidos. Na verdade, já foi demonstrado através de experimentos *in vitro* que epitopos com baixa similaridade na estrutura primária, mas elevada similaridade físico-química e estrutural, são capazes de gerar uma resposta imunológica a partir do mesmo "pool" de células T $CD8^+$ (Sandalova *et al.*, 2005). Além disso, já foi demonstrado que existem padrões moleculares compartilhados entre epitopos, independente da sua estrutura primária, quando apresentados pelo mesmo alelo de MHC (Antunes *et al.*, 2010). Esses dados apontam para a importância da análise estrutural dos complexos pMHC.

Embora a análise estrutural tenha se mostrado cada vez mais útil para desvendar os mistérios por trás das interações imunológicas moleculares, são poucas as estruturas tridimensionais de complexos pMHC cristalografadas disponíveis no PDB. Em partes pela dificuldade da técnica, em partes pelo preço excessivamente elevado para sua obtenção. Para se ter uma idéia, enquanto o banco de dados do PubMed conta com aproximadamente 130 milhões de seqüências, o banco de proteínas tridimensionais do PDB conta com apenas cerca de 70 mil estruturas, até o momento. No caso do gênero hantavírus, os números são ainda menores. Até o momento, foram resolvidas menos de cinco estruturas tridimensionais (apenas pequenos trechos) referentes a proteínas deste gênero viral (Boudko *et al.*, 2007; Estrada *et al.*, 2009; Wang *et al.*, 2008).

Para contornar o problema do baixo número de estruturas tridimensionais, foram desenvolvidos programas de modelagem molecular que conseguem, até certo grau de acurácia, resolver a estrutura terciária de proteínas a partir da sua estrutura primária. Na área da imunoinformática, nosso grupo desenvolveu uma nova técnica (Antunes *et al.*, 2010; Rigo *et al.*, 2009), a qual permite não só a construção de complexos pMHC sem que haja referência cristalográfica do mesmo, bem como a possibilidade de analisar-se em nível molecular a interface de interação entre pMHC e TCR.

O método de docagem molecular, empregado para construção do complexo pMHC, nos permite inferir a afinidade entre a molécula ligante e o receptor, no nosso caso, o epitopo e o MHC, respectivamente. Essa medida é referida como Energia de Ligação (BE, do inglês "Binding Energy") a qual é inversamente proporcional à afinidade de ligação, ou seja, quanto menor o valor de BE, maior a afinidade existente entre o receptor e o ligante.

Como uma das principais características que influenciam a imunogenicidade se refere à estabilidade do complexo pMHC, a qual depende, entre outros fatores, da afinidade entre o peptídeo e o MHC, utilizamos o valor de BE para nossa análise. Os valores encontrados para HTNV e SNV/PUUV (BE = -6,48 e BE = -7,49 para a região $N_{94\text{-}101}$; e, BE = -14,73 e BE = -14,55 para a região $N_{180\text{-}188}$) indicam alta afinidade entre os epitopos e seus respectivos MHCs, para ambas as regiões. Isso era esperado, já que esses peptídeos possuem os mesmos aminoácidos âncoras (os quais são ideais para ligação com o seu respectivo MHC) (Rammensee *et al.*, 1999), contribuindo para que a afinidade entre o peptídeo e o MHC se mantenha elevada. Além disso, os peptídeos comparados entre si, em cada região da proteína N, possuem similaridade estrutural elevada, a qual é decorrente da elevada similaridade na sua estrutura primária.

Observa-se, ainda, que os valores de BE para os peptídeos da região 180-188 são melhores, do ponto de vista da afinidade com o MHC, do que os da região 94-101. As interações entre o receptor e o ligante são calculadas baseadas em um algoritmo genético (Forrest, 1993), o qual se baseia na teoria da seleção natural para procurar e encontrar boas conformações do ligante, no sítio de interação do receptor. Aqui, não apenas dados quantitativos, mas dados qualitativos serão levados em consideração. Como as duas regiões estão sendo apresentadas por diferentes alelos de MHC murino (H2-Kb e H2-Db), isso explica a diferença no valor de BE observado entre as duas regiões. Por serem duas regiões distintas, esses valores não devem ser comparados entre si. A comparação restringe-se às seqüências dentro de cada região.

Uma vez terminada a construção dos complexos, aliando a dinâmica molecular às técnicas de docagem molecular, o potencial eletrostático, bem como a topologia gerada a partir da distribuição de aminoácidos, na superfície de interação de cada pMHC foi calculado a partir do programa GRASP2 (Petrey and Honig, 2003). A partir deste ponto foram observadas duas diferenças importantes. A primeira com relação à distribuição de cargas entre o complexo pMHC de SNV/PUUV referente à região 94-101; a segunda com relação à topologia, entre o complexo pMHC de SNV/PUUV e HTNV referente aos peptídeos da região 180-188.

A primeira diferença constatada entre os complexos pMHC de SNV/PUUV e HTNV, na região 94-101 da proteína N, era esperada, já que houve troca de uma serina

(aminoácido polar não-carregado) por uma arginina (aminoácido polar com carga positiva) na posição 4 do epitopo. Por estar numa posição central em relação ao complexo pMHC, esse aminoácido está envolvido na interação direta com o TCR (Vieira and Chies, 2005; Welsh and Selin, 2002). Além disso, já foi observado que a complementaridade de cargas entre o TCR e o pMHC é um fator importante na geração da resposta imunológica (Chang *et al.*, 1997).

Nossa hipótese, portanto, é a de que o aumento de cargas negativas no sítio de interação do pMHC de HTNV com o TCR, na região $N_{94\text{-}101}$, é a causa maior para que não haja resposta imunológica cruzada entre os epitopos de SNV/PUUV e HTNV. Embora a mudança no padrão de cargas fosse algo sabido desde o início, não sabíamos se essa carga estaria exposta na superfície de interação com o TCR antes de construir o complexo pMHC. Sendo assim, a construção do complexo nos forneceu uma visão mais abrangente do panorama eletrostático do pMHC, no contexto da apresentação antigênica. Ainda, embora a diferença nas cargas seja pequena, é suficiente para que haja uma interação diferenciada com o TCR, já que a constante de dissociação do TCR pelo complexo pMHC – na ordem de 10^{-4}M a 10^{-5}M - é muito baixa (Davis and Chien, 1993). Assim, pequenas mudanças podem levar à diminuição da interação entre o pMHC e o TCR.

Tal análise estrutural é mais elucidativa do que a simples análise linear de seqüências. Quanto à diferença na topologia dos dois complexos pMHC referentes à região 94-101, não foram observadas mudanças qualitativamente significativas.

Já a segunda diferença foi na topologia dos complexos pMHC referentes à região 180-188 da proteína N. É sabido que a topologia não é tão importante para o reconhecimento pelo TCR quanto o perfil eletrostático do complexo pMHC (Jorgensen *et al.*, 1992). Entretanto, já existem estudos demonstrando que dois epitopos, estruturalmente idênticos, com o mesmo padrão de cargas, que apresentam reatividade cruzada a partir do mesmo clone de TCR, apresentam conformações diferentes na fenda do MHC quando não estão interagindo com o TCR. A reatividade cruzada, nesse caso, pode ser explicada pela mobilidade diferencial do epitopo na fenda após a aproximação e interação do TCR (Borbulevych *et al.*, 2009). Portanto, uma diferença na topologia levando a uma interação imunológica diferencial.

Ainda, a partir das análises realizadas, podemos inferir que o fato de não haver uma resposta imunológica cruzada entre os epitopos de HTNV e de SNV/PUUV da região 180-188 da proteína N se deve a outro fator que não a interação com o TCR ou a topologia propriamente dita, mas à fraca afinidade do epitopo pelo MHC e/ou à instabilidade do complexo pMHC na superfície celular. Portanto, uma abordagem *in silico* utilizando dinâmica molecular de complexos pMHC seria de grande valia para uma análise mais profunda sobre a reatividade cruzada. Pensando nisso, nosso grupo vem desenvolvendo e padronizando uma ferramenta onde, a partir da dinâmica molecular de complexos pMHC em ambiente aquoso, será possível diferenciar complexos estáveis de não estáveis e, a partir daí, analisar o comportamento da estrutura ao longo do tempo na presença ou ausência de um TCR (Anexo 1).

Levando-se em consideração os achados deste trabalho e aceitando a hipótese de que não há reatividade cruzada entre os epitopos estudados devido à diferença no padrão de cargas ou na topologia dos complexos pMHC, podemos discutir outros fatores que estão inibindo a resposta imunológica contra os epitopos de HTNV. Nesse âmbito, um ponto que deve ser levado em consideração diz respeito à imunodominância de epitopos. Este termo é utilizado para descrever o modo como o sistema imunológico se direciona no reconhecimento de apenas um pequeno número de determinantes ou epitopos (Assarsson *et al.*, 2007). Ou seja, mesmo havendo um elevado número de epitopos sendo gerados, apenas alguns serão capazes de gerar resposta imunológica devido a uma hierarquia pré-estabelecida, a qual é influenciada pelo contexto infeccioso como um todo (Yewdell and Bennink, 1999). Pensando nisso, podemos imaginar um cenário onde os epitopos de HTNV não estão sendo apresentados por não estarem em um patamar elevado na hierarquia de imunodominância do sistema imunológico (pela influência de fatores como o processamento intracelular, por exemplo).

As análises se baseiam, até o momento, em dados qualitativos, visto que há uma dificuldade em encontrar programas que quantifiquem diferenças topológicas e de cargas entre complexos pMHC. Nesse momento, estamos começando a desenvolver uma ferramenta que compute diferentes valores referentes aos dados apresentados (área do epitopo acessível ao solvente, análise de componentes variáveis entre epitopos imunogênicos e não-imunogênicos, estabilidade do epitopo na fenda do MHC ao longo do

tempo, entre outros) e que consiga agrupar valores de imunogenicidade a epitopos candidatos.

É claro que para que uma resposta imunológica efetiva seja desencadeada, uma série de fatores são necessários, desde a quantidade de complexos pMHC presentes na superfície celular até a presença de moléculas acessórias e de adesão que não possuem interação direta com o pMHC (Assarsson *et al.*, 2007). Embora não tenhamos analisado todos esses fatores, nosso trabalho foi o primeiro a analisar os mecanismos de reatividade cruzada entre epitopos da proteína N de espécies virais do gênero Hantavírus em nível molecular a partir de uma metodologia relativamente simples e rápida. Acessamos, dessa forma, informações valorosas que contribuem para a descoberta de fatores relevantes na imunogenicidade e que, até o momento, estavam sendo negligenciadas por técnicas mais tradicionais.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M and Chies JA (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.

Asada H, Tamura M, Kondo K, Dohi Y and Yamanishi K (1988) Cell-mediated immunity to virus causing haemorrhagic fever with renal syndrome: generation of cytotoxic T lymphocytes. J Gen Virol 69 (Pt 9)**:**2179-2188.

Asada H, Balachandra K, Tamura M, Kondo K and Yamanishi K (1989) Cross-reactive immunity among different serotypes of virus causing haemorrhagic fever with renal syndrome. J Gen Virol 70 (Pt 4)**:**819-825.

Assarsson E, Sidney J, Oseroff C, Pasquetto V, Bui HH, Frahm N, Brander C, Peters B, Grey H and Sette A (2007) A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. J Immunol 178**:**7890-7901.

Bhasin M and Raghava GP (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. Protein Sci 13**:**596-607.

Bjorkman PJ (1997) MHC restriction in three dimensions: a view of T cell receptor/ligand interactions. Cell 89**:**167-170.

Borbulevych OY, Piepenbrink KH, Gloor BE, Scott DR, Sommese RF, Cole DK, Sewell AK and Baker BM (2009) T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. Immunity 31**:**885-896.

Boudko SP, Kuhn RJ and Rossmann MG (2007) The coiled-coil domain structure of the Sin Nombre virus nucleocapsid protein. J Mol Biol 366**:**1538-1544.

Chang HC, Smolyar A, Spoerl R, Witte T, Yao Y, Goyarts EC, Nathenson SG and Reinherz EL (1997) Topology of T cell receptor-peptide/class I MHC interaction defined by charge reversal complementation and functional analysis. J Mol Biol 271**:**278-293.

Chen Y, Shi Y, Cheng H, An YQ 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.

Ciechanover A (1994) The ubiquitin-proteasome proteolytic pathway. Cell 79**:**13-21.

Davis MM and Chien Y (1993) Topology and affinity of T-cell receptor mediated recognition of peptide-MHC complexes. Curr Opin Immunol 5**:**45-49.

Doytchinova IA, Guan P and Flower DR (2006) EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics 7**:**131.

Easterbrook JD, Zink MC and Klein SL (2007) Regulatory T cells enhance persistence of the zoonotic pathogen Seoul virus in its reservoir host. Proc Natl Acad Sci U S A 104**:**15502-15507.

Easterbrook JD and Klein SL (2008a) Seoul virus enhances regulatory and reduces proinflammatory responses in male Norway rats. J Med Virol 80**:**1308-1318.

Easterbrook JD and Klein SL (2008b) Immunological mechanisms mediating hantavirus persistence in rodent reservoirs. PLoS Pathog 4**:**e1000172.

Elliott RM, M. Bouloy, C. H. Calisher, R. Goldbach, J. T. Moyer, S. T. Nichol, R. Pettersson, A. Plyusnin, and C. S. Schmaljohn (2000) Bunyaviridae. In: M. H. V.

van Regenmortel CMF, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, R. B. Wickner (eds) Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, pp 599-621.

Ennis FA, Cruz J, Spiropoulou CF, Waite D, Peters CJ, Nichol ST, Kariwa H and Koster FT (1997) Hantavirus pulmonary syndrome: CD8+ and CD4+ cytotoxic T lymphocytes to epitopes on Sin Nombre virus nucleocapsid protein isolated during acute illness. Virology 238**:**380-390.

Estrada DF, Boudreaux DM, Zhong D, St Jeor SC and De Guzman RN (2009) The Hantavirus Glycoprotein G1 Tail Contains Dual CCHC-type Classical Zinc Fingers. J Biol Chem 284**:**8654-8660.

Falk K and Rotzschke O (2002) The final cut: how ERAP1 trims MHC ligands to size. Nat Immunol 3**:**1121-1122.

Figueiredo LT, Moreli ML, Borges AA, Figueiredo GG, Souza RL and Aquino VH (2008) Expression of a hantavirus N protein and its efficacy as antigen in immune assays. Braz J Med Biol Res 41**:**596-599.

Forrest S (1993) Genetic algorithms: principles of natural selection applied to computation. Science 261**:**872-878.

Gras S, Kedzierski L, Valkenburg SA, Laurie K, Liu YC, Denholm JT, Richards MJ, Rimmelzwaan GF, Kelso A, Doherty PC *et al.* (2010) Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. Proc Natl Acad Sci U S A 107**:**12599-12604.

Guilfoyle JF and Macnab AJ (2008) Hantavirus cardiopulmonary syndrome: implications for transport management and care. Air Med J 27**:**299-302.

Hammer GE, Kanaseki T and Shastri N (2007) The final touches make perfect the peptide-MHC class I repertoire. Immunity 26**:**397-406.

Hannah MF, Bajic VB and Klein SL (2008) Sex differences in the recognition of and innate antiviral responses to Seoul virus in Norway rats. Brain Behav Immun 22**:**503-516.

Hepojoki J, Strandin T, Wang H, Vapalahti O, Vaheri A and Lankinen H (2010) Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein. J Gen Virol 91**:**2341-2350.

Jonsson CB and Schmaljohn CS (2001) Replication of hantaviruses. Curr Top Microbiol Immunol 256**:**15-32.

Jorgensen JL, Esser U, Fazekas de St Groth B, Reay PA and Davis MM (1992) Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. Nature 355**:**224-230.

Karabatsos N, 1985 International catalogue of arboviruses including certain other viruses of vertebrates, pp. in *American Society of Tropical Medicine and Hygiene*, San Antonio.

Kaukinen P, Vaheri A and Plyusnin A (2005) Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties. Arch Virol 150**:**1693-1713.

Kessels HW, de Visser KE, Tirion FH, Coccoris M, Kruisbeek AM and Schumacher TN (2004) The impact of self-tolerance on the polyclonal CD8+ T cell repertoire. J Immunol 172**:**2324-2331.

Khaiboullina SF and St Jeor SC (2002) Hantavirus immunology. Viral Immunol 15**:**609-625.

Khaiboullina SF, Rizvanov AA, Deyde VM and St Jeor SC (2005) Andes virus stimulates interferon-inducible MxA protein expression in endothelial cells. J Med Virol 75**:**267-275.

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

Klein SL, Bird BH and Glass GE (2001) Sex differences in immune responses and viral shedding following Seoul virus infection in Norway rats. Am J Trop Med Hyg 65**:**57-63.

Lee HW, Lee PW and Johnson KM (1978) Isolation of the etiologic agent of Korean Hemorrhagic fever. J Infect Dis 137**:**298-308.

Li XD, Makela TP, Guo D, Soliymani R, Koistinen V, Vapalahti O, Vaheri A and Lankinen H (2002) Hantavirus nucleocapsid protein interacts with the Fas-mediated apoptosis enhancer Daxx. J Gen Virol 83**:**759-766.

Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O and Nielsen M (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 36**:**W509-512.

Maeda A, Lee BH, Yoshimatsu K, Saijo M, Kurane I, Arikawa J and Morikawa S (2003) The intracellular association of the nucleocapsid protein (NP) of hantaan virus (HTNV) with small ubiquitin-like modifier-1 (SUMO-1) conjugating enzyme 9 (Ubc9). Virology 305**:**288-297.

Maeda K, West K, Toyosaki-Maeda T, Rothman AL, Ennis FA and Terajima M (2004) Identification and analysis for cross-reactivity among hantaviruses of H-2b-restricted cytotoxic T-lymphocyte epitopes in Sin Nombre virus nucleocapsid protein. J Gen Virol 85**:**1909-1919.

Mason D (1998) A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunol Today 19**:**395-404.

Melino S, Nepravishta R, Bellomaria A, Di Marco S and Paci M (2009) Nucleic acid binding of the RTN1-C C-terminal region: toward the functional role of a reticulon protein. Biochemistry 48**:**242-253.

Mestas J and Hughes CC (2004) Of mice and not men: differences between mouse and human immunology. J Immunol 172**:**2731-2738.

Mir MA and Panganiban AT (2006) The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. RNA 12:272-282.

Mir MA (2010) Hantaviruses. Clin Lab Med 30:67-91.

Muranyi W, Bahr U, Zeier M and van der Woude FJ (2005) Hantavirus Infection. J Am Soc Nephrol 16**:**3669-3679.

Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE *et al.* (2003) The DNA sequence and analysis of human chromosome 6. Nature 425**:**805-811.

Nandi D, Tahiliani P, Kumar A and Chandu D (2006) The ubiquitin-proteasome system. J Biosci 31**:**137-155.

Nichol ST, Spiropoulou CF, Morzunov S, Rollin PE, Ksiazek TG, Feldmann H, Sanchez A, Childs J, Zaki S and Peters CJ (1993) Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science 262:914-917.

Nielsen M, Lundegaard C, Lund O and Kesmir C (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics 57**:**33-41.

Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O *et al.* (2005) The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol 3**:**e91.

Petrey D and Honig B (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. Methods Enzymol 374**:**492-509.

Plyusnin A, Vapalahti O and Vaheri A (1996) Hantaviruses: genome structure, expression and evolution. J Gen Virol 77 (Pt 11)**:**2677-2687.

Rammensee H, Bachmann J, Emmerich NP, Bachor OA and Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50**:**213-219.

Rigo MM, Antunes DA, Vieira GF and Chies JAB (2009) MHC: Peptide Analysis: Implications on the Immunogenicity of Hantaviruses' N protein. Lecture Notes in Computer Science 5676**:**160-163.

Rognan D, Scapozza L, Folkers G and Daser A (1994) Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. Biochemistry 33**:**11476-11485.

Ross R (1916) An application of the theory of probabilities to the study of a priori pathometry. Part I. Proc. R. Soc. Lond. Ser. A 92**:**204-230.

Sandalova T, Michaelsson J, Harris RA, Odeberg J, Schneider G, Karre K and Achour A (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.

Schmaljohn CS, Schmaljohn AL and Dalrymple JM (1987) Hantaan virus M RNA: coding strategy, nucleotide sequence, and gene order. Virology 157**:**31-39.

Schmaljohn CS (1996) Molecular biology of hantaviruses. In: Elliott R (eds) The Bunyaviridae. Plenum Press, New York, pp 63-90.

Schmitt L and Tampe R (2000) Affinity, specificity, diversity: a challenge for the ABC transporter TAP in cellular immunity. Chembiochem 1**:**16-35.

Selin LK, Nahill SR and Welsh RM (1994) Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. J Exp Med 179**:**1933-1943.

Selin LK, Varga SM, Wong IC and Welsh RM (1998) Protective heterologous antiviral immunity and enhanced immunopathogenesis mediated by memory T cell populations. J Exp Med 188**:**1705-1715.

Sieker F, Springer S and Zacharias M (2007) Comparative molecular dynamics analysis of tapasin-dependent and -independent MHC class I alleles. Protein Sci 16**:**299-308.

Skountzou I, Koutsonanos DG, Kim JH, Powers R, Satyabhama L, Masseoud F, Weldon WC, Martin Mdel P, Mittler RS, Compans R *et al.* (2010) Immunity to pre-1950 H1N1 influenza viruses confers cross-protection against the pandemic swine-origin 2009 A (H1N1) influenza virus. J Immunol 185**:**1642-1649.

Snyder JT, Belyakov IM, Dzutsev A, Lemonnier F and Berzofsky JA (2004) Protection against lethal vaccinia virus challenge in HLA-A2 transgenic mice by immunization with a single CD8+ T-cell peptide epitope of vaccinia and variola viruses. J Virol 78**:**7052-7060.

Tan SY and Dee MK (2009) Elie Metchnikoff (1845-1916): discoverer of phagocytosis. Singapore Med J 50**:**456-457.

Tischler ND, Rosemblatt M and Valenzuela PD (2008) Characterization of cross-reactive and serotype-specific epitopes on the nucleocapsid proteins of hantaviruses. Virus Res 135**:**1-9.

Tong JC and Ren EC (2009) Immunoinformatics: Current trends and future directions. Drug Discov Today 14**:**684-689.

Trott O and Olson AJ (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.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26**:**1701-1718.

Van Epps HL, Schmaljohn CS and Ennis FA (1999) Human memory cytotoxic T-lymphocyte (CTL) responses to Hantaan virus infection: identification of virus-specific and cross-reactive CD8(+) CTL epitopes on nucleocapsid protein. J Virol 73**:**5301-5308.

Vieira GF and Chies JA (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.

Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A and Peters B (2010) The immune epitope database 2.0. Nucleic Acids Res 38**:**D854-862.

Wang J and Maldonado MA (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. Cell Mol Immunol 3**:**255-261.

Wang Y, Boudreaux DM, Estrada DF, Egan CW, St Jeor SC and De Guzman RN (2008) NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. J Biol Chem 283**:**28297-28304.

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, Che JW, Brehm MA and Selin LK (2010) Heterologous immunity between viruses. Immunol Rev 235**:**244-266.

Williams OM, Hart KW, Wang EC and Gelder CM (2002) Analysis of CD4(+) T-cell responses to human papillomavirus (HPV) type 11 L1 in healthy adults reveals a high degree of responsiveness and cross-reactivity with other HPV types. J Virol 76**:**7418-7429.

Yewdell JW and Bennink JR (1999) Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. Annu Rev Immunol 17**:**51-88.

Zacharias M and Springer S (2004) Conformational flexibility of the MHC class I alpha1-alpha2 domain in peptide bound and free states: a molecular dynamics simulation study. Biophys J 87**:**2203-2214.

7. ANEXOS

INFERENCE OF GOOD AND BAD HLA LIGANDS THROUGH MOLECULAR DYNAMICS APPROACH

ARTIGO EM PREPARAÇÃO PARA A REVISTA "*BMC Genomics"*

Inference of Good and Bad HLA ligands Through Molecular Dynamics Approach

Maurício M. Rigo, Dinler A. Antunes, Cassiana C. Fülber, José A.B. Chies, Marialva Sinigaglia, Gustavo F. Vieira

Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

mauriciomr1985@gmail.com

Abstract— The Major Histocompatibility Complex of class I (MHC-I), which is represented by the Human Leukocyte Antigen (HLA) genes in humans, perform an essential role in the presentation of small peptides from pathogens (8-10 amino acids) to the immune system. The peptide:MHC (pMHC) complex interacts with the T Cell Receptor (TCR) which will or will not generate an immune response. One of the factors that influence this response is the pMHC stability. Aiming to characterize molecular aspects underlying this phenomenon, we employed a molecular dynamics approach in good and bad ligands anchored in the HLA-A*02:01 molecule. In this work we were able to observe a greater instability in pMHC complexes presenting bad ligands on the first 10ns, pointing to the efficiency of molecular dynamics to differentiate good from bad HLA peptide binders.

Keywords - Immunoinformatics; Molecular dynamics; HLA; pMHC stability.

I. INTRODUCTION

The human organism is constantly challenged by several pathogens and must defend itself but, at the same time, it should be able to differentiate "self" from "nonself" molecules. This difficult and complex task is driven by a cluster of cells and molecules that, combined, constitute the human immune system. One of these molecules with pivotal importance for response against pathogens is the Major Histocompatibility Complex (MHC) - also called Human Leukocyte Antigen (HLA) in humans. These molecules are responsible by the presentation of small peptides from viruses, parasites, bacteria or even from human proteins to the T Cell Lymphocytes, which will or will not trigger an immune response, through the T Cell Receptor (TCR) interaction. There are two classes of MHC: MHC-I and MHC-II. The MHC-I is responsible for the presentation of small peptides (8-12 amino acids long) generated by endogenous processing, while the MHC-II presents longer peptides (up to 30 amino acids) from exogenous environment [1].

It is known that topologies and electrostatics patterns of the peptide:MHC (pMHC) interacting region are important in the TCR recognition [2]. Also the affinity and stability between the peptide (epitope) and the MHC (receptor) plays a crucial role [3], especially on what concerns electrostatics and nonhydrophobic interactions between the MHC and anchor residues of the peptide [4]. All these features stands for the peptide immunogenicity.

The knowledgment of issues regarding the peptide affinity by MHC cleft and the stability of resulting complex is pivotal for vaccine development field [5]. Nowadays, there are several immunoinformatics tools available to assist the resolution of this task [6]. For instance, it is possible to predict peptide binding affinity through several computational methods (i.e. using tools that consider amino acid sequence composition and preferential anchor residues occurrence) [7, 8].

There are also bioinformatics tools that are not specific for immunological issues, but can be employed for this purpose, such as molecular docking and molecular dynamics [9,10,11]. The latter can be used to assess pMHC complex stability, one of the main factors involved in peptide presentation.

In the present work, our goal was to characterize the molecular basis for stability of good and bad ligands in the context of the HLA-A*02:01 MHC-I allele, through molecular dynamics (MD) approach.

II. MATERIAL AND METHODS

*A. Choose of good and bad HLA-A*02:01 ligands*

Five HLA-A*02:01 optimal ligands (G1-G5) of nine amino acids length (9-mer) were obtained from literature [12,13,14,15], based on in vivo/in vitro experiments. The ligand's affinity was confirmed by immunoinformatics prediction tools (T Cell Epitope Prediction Tools from IEDB [8], NetMHC [16] and SYFPEITHI [7]) (Table 1). In contrast, five matching bad ligands (B1-B5), were designed by the inclusion of alternative anchor residues, and the impairment of MHC binding affinity was also verified through in silico analysis.

TABLE 1. Binding scores from different immunoinformatics programs.

Epitopes	**Binding predictors**		
	SYFPEITH I[a]	**NetMHC IC50 (nM)**	**IEDB IC50 (nM)**
Good binders			
KLLEPVLLL (G1)	31	10	10.3
ILMEHIHKL (G2)	32	4	7.5
VLADQVWTL (G3)	30	5	9.2
ILDDIGHGV (G4)	25	8	12.7
YLIELIDRV (G5)	31	3	7.0
Bad binders			
KPLEPKLLH (B1)	7	48435	24787.3
IPMEHIHKH (B2)	12	45629	24526.8
VPADQVWTH (B3)	10	45902	24550.6
IPDDIGHGH (B4)	5	47853	24851.5
YPIELIDRH (B5)	11	45442	24559.9

a. The SYFPEITHI score is proportional to the MHC:peptide binding degree.

B. pMHC complexes construction

Since there are no crystallographic structures of these ligands complexed to HLA-A*02:01 molecule, we performed the construction of each complex (pMHC) based on a technique (D1-EM-D2) implemented by our group [17]. Briefly, each epitope was fitted with the structure of a typical epitope presented by HLA-A*02:01 allele. The receptor structure was obtained from Protein Data Bank (PDB ID: 2v2w) and the ligand docking was performed using the program AutoDock Vina v.1.1.1 [11]. A thousand conformations were generated but only the best conformation was chosen,

based on the binding energy and the root mean square deviation (RMSD) of the epitope. The complex was submitted to an energy minimization step with the GROMACS v4.0.7. package [9], and a subsequent molecular docking to refine the complex. Finally, the best conformation was obtained and used for the following analysis.

C. pMHC molecular dynamics

The resulting peptide:HLA-A*02:01 complexes were submitted to a molecular dynamics simulation using the GROMACS v4.0.7 package. The simulation was set in a periodic box including SPC water model and using the GROMOS56a6 force field for each complex. Briefly, each molecule was submitted to one energy minimization step using the steepest descent algorithm, one molecular dynamics step using position restraints for all heavy atoms, five molecular dynamics steps to gradually heat the system from 50K to 300K over 0.03 ns and a molecular dynamics of 30 ns.

D. Root Mean Square Deviation analysis

GROMACS software package contains several programs that could be employed to analyze the generated data from molecular dynamics simulation. One of them is the g_rms program, which allows the analysis of the Root Mean Square Deviation (RMSD) of each epitope along the simulation time. In order to do that, the program calculates the RMSD of the epitope between each point of the simulation and the initial position. From these data we could infer the stability of the pMHC complex along the time and compare each group of epitopes among them.

III. RESULTS

Root mean square deviation (RMSD) values were statistically higher for the bad ligands, considering the peptide structure variation along the first 5ns, indicating that these complexes present a more unstable behavior (Figure 1).

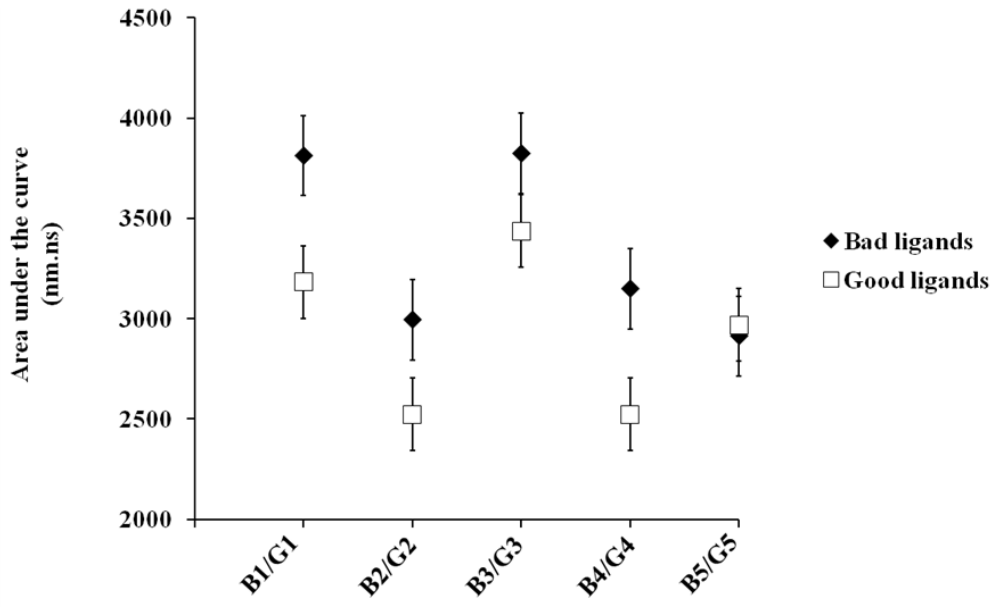

Figure 1. Area under the curve (AUC) based on RMSD values for epitopes (ligands) 1 to 5 (G1-G5 and B1-B5). Student's t-test; $p < 0.05$.

The only exception was the pair B5/G5. Although we could not observe statistically differences regarding RMSD analysis, structurally the complex B5:HLA-A*02:01 presented alfa-helix dissociation and epitope detachment earlier than G5:HLA-A*02:01 (Figure 2).

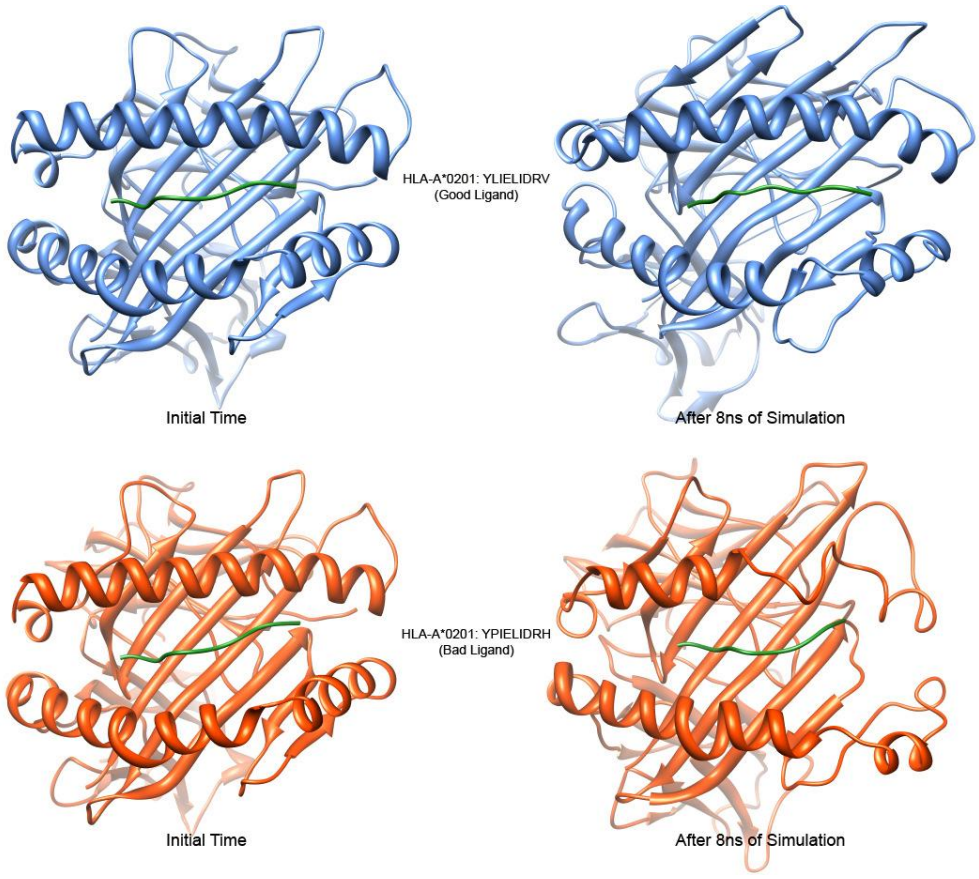

Figure 2. Qualitative analysis of the pMHC complexes could be performed using the molecular dynamics approach. In this figure, we could observe that the MHC bounded with a bad binder resulted in the destabilization of the complex, especially in the TCR recognition region.

In fact, when we performed a qualitative structural analysis of all the complexes, the partial detachment of the peptide from the HLA-A*02:01 cleft was more pronounced in the pMHC complexes containing bad epitopes. Alfa-helix, an essential feature for the TCR interaction, dissociated before the end of the simulation in all complexes. However, alfa-helix dissociation for bad epitopes occurred earlier than for the good ones.

IV. DISCUSSION

There are many tools for epitope binding prediction for the MHC cleft considering mainly the primary sequence of the epitopes. This kind of approach only show a static view of the

process, especially considering the presence of appropriate anchor residues at specific positions in the epitope sequence and the result is limited to categorizations such as binders or non-binders. We propose a more realistic approach, evaluating the structural behavior of the epitope in the MHC cleft in a specific period of time. It will allow us to determine not only if the epitope is a good ligand, but if this epitope is stable in the MHC cleft for an adequate presentation to the T cell receptor.

As presumed, we were able to observe epitope detachment and a greater instability in pMHC complexes presenting bad ligands on the first 10ns (although we performed a 30ns molecular dynamics). Along the simulation, it was also observed that pMHC complexes of G1-G5 epitopes also presented dissociation. This is expected, since none pMHC complex lasts forever on the cell surface [18]. What should be noted is that pMHC complexes of B1-B5 epitopes presented dissociation features earlier than G1-G5 epitopes (Figure 1 and Figure 2).

The simulation time can not be compared with the real time since molecular dynamics simulation involves an active molecular environment. This means that modifications could be observed in short times (in this case, after nanoseconds) when compared with in vivo/in vitro experiments, which usually takes longer times. These preliminary results point to the efficiency of molecular dynamics to differentiate good from bad HLA peptide binders, opening a new field on vaccine development using advanced computational tools.

ACKNOWLEDGMENT

Funded by CNPq, CAPES and a grant from Bill & Melinda Gates foundation through the Grand Challenges Exploration Initiative.

REFERENCES

[1] A. K. Abbas, A. H. Lichtman on Cellular and Molecular Immunology, 5th ed. Elsevier, 2005.

[2] H. W. Kessels, K. E. de Visser, F. H. Tirion, M. Coccoris, A. M. Kruisbeek and T. N. Schumacher, "The Impact of Selftolerance on the Polyclonal CD8+ T-cell Repertoire," J. Immunol., vol. 172, Feb. 2004, pp. 2324–2331.

[3] S. H. van der Burg, M. J. Visseren, R. M. Brandt, W. M. Kast and C. J. Melief, "Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability," J. Immunol., vol. 156, May. 1996, pp. 3308-3314.

[4] J. L. Jorgensen, U. Esser, B. Fazekas de St Groth, P. A. Reay and M. M. Davis, "Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics," Nature, vol. 355, Jan. 1992, pp. 224-230, doi:10.1038/355224a0.

[5] R. H. McMahan, J. A. McWilliams, K. R. Jordan, S. W. Dow, D. B. Wilson and J. E. Slansky, " Relating TCR-peptide- MHC affinity to immunogenicity for the design of tumor vaccines," J. Clin. Invest., vol. 116, Sep. 2006, pp. 2543- 2551, doi: 10.1172/JCI26936.

[6] B. Korber, M. LaBute and K. Yusim, "Immunoinformatics Comes of Age," PLoS Comput Biol., vol. 2, Jun. 2006, e71, doi: 10.1371/journal.pcbi.0020071.

[7] H. Rammensee, J. Bachmann, N. N. Emmerich, O. A. Bachor and S. Stevanovic, " SYFPEITHI: Database for MHC Ligands and Peptide motifs," Immunogenetics, vol. 50, Nov. 1999, pp. 213-219, doi: 10.1007/s002510050595.

[8] R. Vita, L. Zarebski, J. A. Greenbaum, H. Emami, I. Hoof, N. Salimi, R. Damle, A. Sette and B. Peters, "The immune epitope database 2.0," Nucleic Acids Res., vol. 38, Nov 2009, pp. 854-862, doi: 10.1093/nar/gkp1004.

[9] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, "GROMACS: Fast, Flexible, and Free," J. Comput. Chem., vol. 26, Dec. 2005, pp. 1701–1718, doi: 10.1002/jcc.20291.

[10] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, "Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function," J. Comput. Chem., vol. 19, Jan. 1999, pp. 1639–1662, doi: 10.1002/(SICI)1096- 987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

[11] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading," J. Comput. Chem., vol. 31, Jan. 2010, pp. 455-461, doi: 10.1002/jcc.21334.

[12] D. F. Hunt, H. Michel, T. A. Dickinson, J. Shabanowitz, A. L. Cox, K. Sakaguchi, E. Appella, H. M. Grey and A. Sette, "Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad," Science, vol. 256, Jun. 1992, pp. 1817-1820, doi: 10.1126/science.1319610.

[13] M. Diehl, C. Münz, W. Keilholz, S. Stevanović, N. Holmes, Y. W. Loke and H. G. Rammensee, "Nonclassical HLA-G molecules are classical peptide presenters," Curr Biol., vol. 6,
Mar. 1996, pp. 305-314, doi: doi:10.1016/S0960- 9822(02)00481-5.

[14] A.O. Weinzierl, C. Lemmel, O. Schoor, M. Müller, T. Krüger, D. Wernet, J. Hennenlotter, A. Stenzl, K. Klingel, H. G. Rammensee and S. Stevanovic, "Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface," Mol. Cell. Proteomics, vol. 6, Jan. 2007, pp. 102- 113, doi: 10.1074/mcp.M600310-MCP200.

[15] V. Ramakrishna, M. M. Ross, M. Petersson, C. C. Gatlin, C. E. Lyons, C. L. Miller,

H. E. Myers, M. McDaniel, L. R. Karns, R. Kiessling, G. Parmiani and D. C. Flyer, "Naturally occuring peptides associaated with HLA-A2 in ovarian cancer cell lines identified by mass spectrometry are targets of HLAA2- restricted cytotoxic T cells," Int. Immunol., vol. 15, Jun 2003, pp. 751-763, doi: 10.1093/intimm/dxg074.

[16] C. Lundegaard, O. Lund and M. Nielsen, " Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers," Bioinformatics, vol. 24, Jun. 2008, pp. 1397-1398, doi:10.1093/bioinformatics/btn128.

[17] D. A. Antunes, G. F. Vieira, M. M. Rigo, S. P. Cibulski, M. Sinigaglia, J. A. B. Chies, " Structural Allele-Specific Patterns Adopted by Epitopes in the MHC-I Cleft and Reconstruction of MHC:peptide Complexes to Cross-Reactivity Assessment," Plos One, vol. 5, Apr. 2010, pp. e10353, doi:10.1371/journal.pone.0010353.

[18] B. R. Clark, " Fate of intercellular MHC-peptide-T-cell receptor complexes during T-cell activation," J. Mol. Recognit., vol. 8, Apr. 1995, pp. 63-66, doi: 10.1002/jmr.300080111.