

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

TESE DE DOUTORADO

**Modelagem Automatizada e Ensaio de Estabilidade *In Silico*
sobre Complexos Peptídeo:MHC-I**

MAURÍCIO MENEGATTI RIGO

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

Orientador: Prof. Dr. Gustavo Fioravanti Vieira

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

PORTO ALEGRE

JUNHO DE 2015

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

Apoio financeiro

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

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

*“Dedico este trabalho
àqueles que moldaram o
meu ser e que tanto amo:
meus pais.”*

AGRADECIMENTOS

Agradeço, primeiramente, à Universidade Federal do Rio Grande do Sul, a qual tem me beneficiado com um ensino gratuito e de qualidade. Também agradeço ao Programa de Pós-graduação em Genética e Biologia Molecular da UFRGS, pelo auxílio acadêmico, estrutural e financeiro. Agradeço, em especial, ao **Elmo Cardoso**, pessoa de caráter honesto, gentil e incansável, o qual sempre me auxiliou nos assuntos relacionados ao doutorado e ao programa de pós-graduação.

Agradeço pela rapidez e eficiência no parecer desta Tese, bem como pelos elogios, da relatora professora **Vera Gaiesky**. Agradeço à banca de avaliação, composta pela professora **Cristina Bonorino** (PUC-RS) e pelos professores **José Artur Bogo Chies** (UFRGS) e **Roney Santos Coimbra** (FIOCRUZ), por dedicar um tempo precioso à leitura e análise deste trabalho, bem como pelas sugestões e críticas. Um agradecimento especial ao professor José Artur, pois nos deu a oportunidade de crescer como grupo e sempre nos auxiliou nas questões imunológicas, desde as mais simples até as mais complexas.

Agradeço ao meu orientador, professor **Gustavo Fioravanti Vieira**, por ter aberto as portas para um mundo no qual, hoje, tenho verdadeira satisfação de trabalhar, por ter acreditado no meu potencial, por ter guiado meu trabalho e por ter estado presente durante o desenvolvimento desta Tese. Contigo aprendi a sempre buscar o meu melhor, em prol da melhor ciência que eu pudesse fazer. Obrigado pela orientação e pela amizade.



Agradeço do fundo do meu coração a minha coorientadora, professora **Marialva Sinigaglia**, por toda a atenção, carinho e comprometimento com o laboratório e com o trabalho desenvolvido nesta Tese. Sempre me deu força para não desistir e acreditar nos meus resultados. És uma pessoa admirável e singular, exemplo de pesquisadora, batalhadora e perseverante.



Agradeço aos colegas do Laboratório de Imunogenética e do Laboratório NBLI, em especial ao **Bragatte, Francis, Marcus, Martiela, Matheus e Renata**. Um

agradecimento especial ao Marcus, que apesar do sangue corintiano, é um verdadeiro amigo que tive a felicidade de conhecer durante o doutorado. Não apenas um ótimo colega de laboratório, sempre engajado e pronto a ajudar, como um companheiro dos almoços, cafés e das brincadeiras um pouco “ácidas” aos olhos alheios. Também agradeço à Martiela por todo o esforço, comprometimento e dedicação ao DockTope.



*Agradeço à família da minha noiva; **Igor, Jana, Mariana, Moacir, Rodrigo, vó Teresa e Zilma**, por me agraciarem com momentos gostosos ao longo do doutorado e pela força que sempre me deram.*



*Um agradecimento especial a três amigos: **Biel, Guaporé e Matheus**. Vocês são, para mim, como irmãos. Pessoas com quem eu sempre contei e sempre vou contar, nos melhores e piores momentos. Biel, tua amizade foi fundamental, especialmente em um dos meus piores momentos da graduação, quando achava que não teria força nem para terminar o curso. Matheus, não só pela amizade, mas por me ajudar a abrir portas importantes na minha formação acadêmica. Guaporé, pela amizade que começou antes mesmo do vestibular! Obrigado pelo ombro amigo e pelos momentos musicais. Um dia a gente ainda vai tocar violão mais do que 2 vezes por ano!! Se nada der certo, “bora” montar uma dupla musical! Só precisamos escolher nomes mais comerciais!*

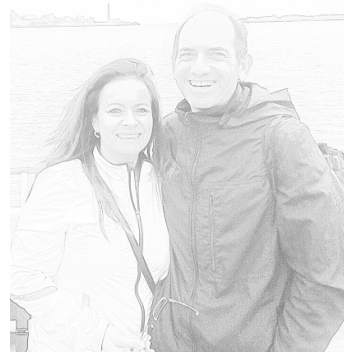


*Não poderia deixar de dedicar um parágrafo para o **Dinler**, um cara que, definitivamente, teve uma parcela significativa de contribuição para a realização deste trabalho. Mais do que um amigo que conheci na faculdade, és um irmão, que esteve sempre ao meu lado durante o TCC, mestrado e doutorado.*

Sempre me ajudaste, tanto pelo lado pessoal quanto pelo lado profissional. Obrigado pela paciência e pela motivação, por me ajudar a ver o copo meio cheio, por me auxiliar a contornar os problemas e achar a melhor solução. Hoje a distância de 8.136 km nos separa, mas a nossa amizade não tem fronteiras e tenho certeza que ela se manterá, sabe por quê? Porque ela é do tamanho do Texas!



*Agradeço de todo coração a minha vó, aos meus tios, tias, primos e primas. Em especial, agradeço ao meu tio “CÓ” e a minha tia **Ana**, pelo apoio emocional e financeiro, pelo carinho, pela confiança e por me proporcionarem momentos gostosos sempre que os visito. Agradeço a minha vó **Ignês** querida, pelo amor, por ter me ensinado a jogar cartas (com e sem truques) e pelo apoio que sempre me destes. Agradeço a minha prima*



***Solange**, uma pessoa que admiro e que sempre me deu força, especialmente quando estive em SP para fazer o visto. Sem tua ajuda, provavelmente teria sido muuuuuito mais difícil ter conseguido fazer uma das viagens mais importantes do meu doutorado.*

Por fim, gostaria de agradecer a quatro pessoas, mais do que especiais:

Agradeço a uma garota que conheci lá em 2011, no início do doutorado, e por quem eu sou apaixonado até hoje. Uma garota que em



*breve terei a sorte e o orgulho de chamar de minha esposa. **Chele**, amor das minhas vidas, linda e soberana, obrigado pelas palavras, pelo companheirismo, pela amizade, pelo teu amor. Obrigado por partilhar comigo momentos gostosos de lazer, de muita risada e de gordices supremas! Tu SEMPRE esteves ao meu lado, me auxiliando nas decisões e*



me impulsionando para que eu explorasse todo o meu potencial. Tu és tudo que eu sonhei e mais um pouco! Essa Tese, definitivamente, não teria o mesmo valor sem o teu apoio. Te amo demais!!!!!!!

*Agradeço a minha irmã, **Daniela**, que apesar de já ser bem grandinha eu*



sempre vou ver como a minha maninha de 10 anos! Talvez por isso eu seja tão chato às vezes contigo! Ao longo do doutorado, apesar da distância, percebi o quanto amadureceste como pessoa e como profissional. Tenho muito orgulho de te ter como minha irmã!

*Por fim, agradeço a minha mãe, **Marilde Menegatti Rigo**, e ao meu pai, **Jorge Luiz Corso Rigo**. Vocês se dedicaram a vida inteira para fornecer o melhor para teus filhos. Abriram mão de muita coisa para que eu*



pudesse estar aqui hoje. Forneceram os alicerces da minha educação. Me



ensinaram a ser um homem de valores. Me ajudaram nos momentos mais difíceis e nunca desistiram de acreditar em mim. Pai, teus conselhos, embora um pouco “tortos” por vezes, são ensinamentos que não encontro em livro algum. Sempre admirei a tua inteligência e tua resiliência. Mãe, não existe pessoa igual a ti. Tu és a melhor mãe que eu poderia ter em

um milhão de vidas! Sempre preocupada, mantendo a minha geladeira cheia com lasanha, massas e outras gordices. Amo vocês!



*“Passar no vestibular é
fácil. Quero ver o depois.”*

Jorge Rigo, 2004

SUMÁRIO

LISTA DE ABREVIATURAS.....	3
RESUMO.....	5
ABSTRACT.....	6
Capítulo I - Introdução e Objetivos	7
Introdução	8
O Sistema Imunológico - Imunidade inata e Imunidade adquirida.....	8
A via de processamento de antígenos endógenos	11
O MHC-I e a apresentação de antígenos	14
Imunoinformática: a bioinformática aplicada ao estudo da Imunologia.....	19
Imunoinformática e o MHC-I	21
Objetivos.....	26
Objetivo Geral.....	26
Objetivos Específicos.....	26
Capítulo II - MHC, Viral Infection and Immunoinformatics.....	27
Capítulo III - DockTope: a Web-based tool for automated pMHC-I modelling45
Capítulo IV - Peptide:MHC-I Stability Assayed through in silico Molecular Dynamics	82

Capítulo V - Discussão Geral	113
Referências Complementares (Capítulos I e V)	126
Anexos	135
<i>CrossTope: a curate repository of 3D structures of immunogenic peptide:MHC complexes.....</i>	<i>136</i>
<i>Evolutionary pattern in the OXT-OXTR system in primates: Coevolution and positive selection footprints.....</i>	<i>145</i>
<i>New Insights into the In Silico Prediction of HIV Protease Resistance to Nelfinavir.....</i>	<i>184</i>
<i>Lessons from molecular modeling human α-L-iduronidase.....</i>	<i>186</i>
<i>Improved structural method for T-cell cross-reactivity prediction</i>	<i>188</i>
<i>Peptide:MHC structural similarity as a probability for cross-reactive T cell responses</i>	<i>189</i>

LISTA DE ABREVIATURAS

ANN _ Redes Neurais Artificiais (do inglês, *Artificial Neural Network*).

BCR _ Receptor de Células B (do inglês, *B Cell Receptor*).

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

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

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

D1-EM-D2 _ *Docking 1-Energy Minimization-Docking 2*.

DM _ Dinâmica Molecular.

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

IFN- γ _ Interferon gama (do inglês, *Interferon- γ*).

IL _ Interleucina.

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

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

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

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

pMHC _ Complexo peptídeo:MHC.

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

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

RMSF _ Flutuação Quadrática Média (do inglês, *Root Mean Square Fluctuation*).

SVM _ Máquina de Suporte Vetorial (do inglês, *Support Vector Machine*).

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

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

TCR:pMHC _ Complexo TCR:peptídeo:MHC.

TNF- α _ Fator de Necrose Tumoral alfa (do inglês, *Tumor Necrosis Factor- α*).

RESUMO

O sistema imunológico é formado por um conjunto de células e moléculas envolvidas primariamente na defesa do organismo contra agentes infecciosos. Neste sentido, destaca-se a ação das proteínas codificadas pela região do MHC-I, as quais são responsáveis pela apresentação de pequenos peptídeos (normalmente entre 8 e 12 aminoácidos), gerados a partir da via de processamento endógeno, para Receptores de Células T (TCRs) citotóxicas. As bases moleculares subjacentes à apresentação de peptídeos no contexto das moléculas de MHC-I ainda não foram totalmente esclarecidas, principalmente em função da complexidade inerente a esse processo biológico. Sabe-se, no entanto, que padrões moleculares presentes na área de interação do pMHC-I com o TCR, bem como a estabilidade dos complexos pMHC-I na superfície celular, são fatores determinantes para a imunogenicidade de diferentes complexos. Com o nascimento e desenvolvimento da bioinformática, uma grande evolução foi observada na área, especialmente no que tange ao estudo de sequências lineares de genes e proteínas. Entretanto, o estudo a nível estrutural ainda é lento, especialmente devido ao limitado número de estruturas tridimensionais resolvidas experimentalmente e à falta de modelos acurados. Assim sendo, esta tese se propôs a automatizar e otimizar uma ferramenta de modelagem de complexos pMHC-I (técnica *D1-EM-D2*) e, posteriormente, criar um protocolo para avaliar a estabilidade *in silico* de complexos pMHC-I, com foco na interação entre epitopo e MHC-I. Como resultado, criamos uma nova ferramenta, a qual chamamos DockTope, validada sobre um conjunto mais amplo de dados experimentais (135 estruturas em relação às 46 estruturas do estudo anterior). Esta ferramenta permite que qualquer usuário modele seu próprio pMHC-I, através de uma interface *web* acessível. Ainda, um estudo via dinâmica molecular realizado com base nos complexos pMHC-I modelados a partir do DockTope nos permitiu inferir algumas das variáveis envolvidas na estabilização dos complexos. Os resultados apresentados aqui contemplam várias áreas de conhecimento da área biológica, favorecendo o desenvolvimento de vacinas mais racionalizadas e alavancando o conhecimento e a compreensão de eventos imunológicos.

ABSTRACT

The immunologic system is constituted by a group of cells and molecules involved mainly in the organism defense against infectious agents. In this way, the action of proteins encoded by the MHC-I region should be highlighted, since they are responsible for the presentation to T Cell Receptors (TCRs) of small peptides (normally 8 to 12 amino acids in length) generated by the endogenous processing pathway. The molecular bases underlying the presentation of peptides in the context of MHC-I molecules are not fully understood, especially because the complexity inherent to this process. However, it is known that molecular patterns presented in the TCR-interaction surface of each pMHC-I, as well as complex stability on cellular surface, are pivotal factors to determine immunogenicity. The bioinformatics development brings together a great evolution in the field, especially in respect to the analysis of genes and proteins sequences. Yet, the structural study is still delayed, mainly because we have a limited number of three-dimensional structures experimentally resolved and there is a lack of accurate models. In this way, we aimed to automatize and optimize a bioinformatics tool to pMHC-I modeling (*D1-EM-D2* approach) and, posteriorly, to create an *in silico* protocol to evaluate the pMHC-I complex stability, focusing on the interaction between epitope and MHC-I. As a result, we created a new tool, which we called DockTope, validated over a wider group of experimental data (135 structures in respect to the 46 structures previously published). This tool allows any user to model its own pMHC-I complex through a user-friendly web interface. Also, a molecular dynamics study performed using modeled pMHC-I complexes through DockTope allowed us to infer important variables involved on complex stabilization. The results presented here encompass several knowledge areas, favoring the vaccine development and propelling the knowledge and understanding of immunologic events.

Capítulo I

Introdução e Objetivos

Introdução

O Sistema Imunológico - Imunidade inata e Imunidade adquirida

Falar sobre o sistema imunológico em uma simples seção de introdução de tese, abordando suas características mais intrínsecas, suas redes de interações moleculares e a sua função regulada de forma minuciosa, seria de uma enorme presunção deste autor para com o leitor. Entretanto, com a humildade e a simplicidade sobre as quais este trabalho foi pautado, expressarei nas próximas linhas os conceitos básicos deste complexo sistema.

O sistema imunológico, especificamente o encontrado em vertebrados, é composto por um conjunto de células e moléculas que, em conjunto, atuam mantendo a homeostase da resposta imunológica de forma simultânea através da (i) defesa do organismo contra o não-próprio (patógenos, como agentes infecciosos virais, por exemplo) e (ii) da tolerância ao próprio (estruturas moleculares que pertencem ao organismo). Para entender o *modus operandi* do sistema imunológico, didaticamente, os autores o subdividem em sistema imunológico inato (ou natural) e sistema imunológico adaptativo (ou adquirido). Como o próprio nome sugere, ambas as subdivisões carregam um caráter implícito que as situa dentro de um panorama evolutivo.

O sistema imunológico inato (imunidade inata ou natural) pode ser considerado a primeira linha de defesa do organismo. Este sistema é compreendido por barreiras físicas (como a camada epitelial que recobre o corpo do organismo), células fagocitárias (como macrófagos e neutrófilos) e células NK (do inglês, *Natural Killer*) e proteínas presentes no sangue (como as proteínas do sistema complemento e as citocinas) com funções que variam desde o reconhecimento de padrões específicos dos patógenos até recrutamento de outras células do sistema imunológico. Evolutivamente, este sistema surgiu anteriormente ao sistema imunológico adaptativo, sendo que em invertebrados, fungos e plantas ele é encontrado de forma exclusiva no desempenho das funções de defesa (Schulenburg *et al.*, 2004). A imunidade inata é baseada

principalmente no reconhecimento de PAMPs (do inglês, *Pathogen-associated Molecular Patterns*) através de receptores de reconhecimento de padrões, chamados PRRs (do inglês, *Pattern Recognition Receptors*) (Iwasaki & Medzhitov, 2010), tais como TLRs (do inglês, *Toll-like Receptors*) (Botos *et al.*, 2011), PGRPs (do inglês, *Peptidoglycan Recognition Proteins*) (Royet *et al.*, 2011) e NLRs (do inglês, *NOD-like Receptors*) (Ting *et al.*, 2010).

Já o sistema imunológico adaptativo começa a surgir a partir dos vertebrados inferiores, em peixes chamados mandibulados (Intrafilo Gnathostomata), há aproximadamente 500 milhões de anos. Acredita-se que o sistema imunológico adaptativo surgiu nesses peixes através de dois eventos macroevolucionários principais, com o (i) surgimento do gene RAG (do inglês, *Recombination-activating Gene*) e de (ii) duas duplicações genômicas (Flajnik & Kasahara, 2010). Este tema está constantemente em foco, sendo que o conhecimento sobre ele evolui de forma bastante dinâmica. Recentemente, por exemplo, observou-se que, na verdade, em peixes não-mandibulados (Intrafilo Agnatha), uma duplicação genômica em relação ao seu ancestral comum aos gnatostomados deu origem, por evolução convergente, a um sistema de defesa bastante similar ao sistema imunológico adaptativo. Inclusive, com o surgimento de duas linhagens de células linfoides (Pancer *et al.*, 2004).

Ao longo de mais de 500 milhões de anos de coevolução, um aspecto chave é a cooperação funcional entre células e moléculas dos sistemas imunológicos inato e adaptativo, direcionando a resposta imune (Iwasaki & Medzhitov, 2010). Tão incrível quanto, é observar como algumas dessas moléculas que estão na interface entre os dois sistemas, permitindo a sua atuação simultânea, já estavam presentes em vertebrados mais basais (Boehm *et al.*, 2012). Como exemplo, podemos citar a IL-17 (Tsutsui *et al.*, 2007), molécula crucial para o desenvolvimento da resposta pró-inflamatória, e a IL-8 (Najakshin *et al.*, 1999), recrutador de leucócitos para o sítio de infecção, já encontradas em vertebrados não mandibulados.

O cerne do sistema imunológico adaptativo é composto por células denominadas linfócitos B e linfócitos T. Essas células possuem um progenitor

linfoide comum (linfoblasto), o qual é oriundo de uma célula hematopoiética pluripotente, na medula óssea. A partir da medula óssea, os linfócitos B recém gerados são direcionados para a corrente sanguínea, enquanto que os precursores dos linfócitos T, antes de entrarem no sistema circulatório para desempenhar suas funções específicas, precisam passar por um processo de seleção no timo (Yates, 2014). Linfócitos B e T possuem receptores de reconhecimento de antígenos em sua superfície denominados BCR e TCR, respectivamente. O BCR, também denominado imunoglobulina de membrana, pode ser encontrado na superfície da membrana de linfócitos B ou, após encontro com o patógeno, na forma solúvel (anticorpos). Já o TCR é encontrado apenas na superfície de linfócitos T e sua ativação é dependente da apresentação de peptídeos no contexto do MHC, como discutiremos adiante.

Tanto os linfócitos B como os precursores dos linfócitos T estão em um estado inativo no momento em que são gerados, ou seja, ainda não existe um direcionamento celular específico, coordenado por um conjunto de genes e proteínas, para a eliminação de um micro-organismo específico. Após o encontro com o patógeno, entretanto, eles passam por um processo de expansão clonal e de diferenciação, tornando-se células efetoras específicas e promovendo a eliminação do patógeno. Após a finalização da resposta imunológica, parte das células que participaram do processo de defesa entra em apoptose, enquanto parte das células permanece viável. Essas células remanescentes são denominadas células de memória e ajudarão o organismo a responder de forma mais rápida e eficaz em um segundo desafio contra o mesmo patógeno (Gaide *et al.*, 2015; Kurosaki *et al.*, 2015).

Traçando-se um paralelo entre o sistema imunológico inato e o sistema imunológico adaptativo, suas principais diferenças residem no fato de que a imunidade natural é específica para estruturas moleculares conservadas de micro-organismos, possui limitada diversidade de células e proteínas e, em geral, não possui memória imunológica. Já no sistema imunológico adquirido, a resposta é específica não apenas para micro-organismos, mas também para antígenos não-microbianos. Além disso, há uma grande diversidade proteica, especialmente

no que tange aos receptores de membrana BCR e TCR e há memória imunológica.

Portanto, em um contexto infeccioso, a resposta imunológica é desencadeada mediante o reconhecimento do antígeno por moléculas e células do sistema imunológico. No contexto da imunidade adquirida, o reconhecimento desse antígeno pode ser realizado através de anticorpos (resposta humoral), os quais são secretados por linfócitos B diferenciados, também chamados plasmócitos. Entretanto, anticorpos são capazes de reconhecer apenas os antígenos presentes em patógenos circulantes, ou seja, uma vez que eles adentrem nas células, não há como os mesmos serem neutralizados através da resposta humoral. Assim, um segundo mecanismo de defesa entra em ação, através do reconhecimento pelo TCR de peptídeos apresentados por células infectadas (resposta celular). Esses peptídeos podem ser apresentados na superfície celular por duas proteínas principais, dependendo da origem do patógeno: o MHC de classe I (MHC-I) e o MHC de classe II (MHC-II). Enquanto o MHC-II apresenta preferencialmente peptídeos oriundos de patógenos extracelulares, o MHC-I apresenta peptídeos oriundos de patógenos intracelulares. Aqui, daremos enfoque para a via de processamento e apresentação de peptídeos endógenos do MHC-I.

A via de processamento de antígenos endógenos

A infecção celular por um patógeno intracelular (por exemplo, um agente infeccioso viral) desencadeia um mecanismo de defesa dependente do MHC-I. Entretanto, uma série de passos ocorre desde a infecção até a apresentação do peptídeo viral na superfície celular.

Em um primeiro momento, proteínas citoplasmáticas, entre elas aquelas oriundas de patógenos, são direcionadas para um sistema proteassomo-ubiquitina (UPS, do inglês *Ubiquitin-Proteasome System*). O objetivo do sistema UPS é a marcação da proteína com ubiquitinas (pequenas proteínas com até 76 aminoácidos) e posterior direcionamento para o proteassomo. A marcação com

ubiquitinas, também chamada poliubiquitinação, ocorre através de uma cascata de eventos direcionada por três enzimas principais: E1 (Enzima ativadora da ubiquitina), E2 (Enzima transportadora da ubiquitina) e E3 (Ubiquitina ligase) (Choi *et al.*, 2012) (Figura 1).

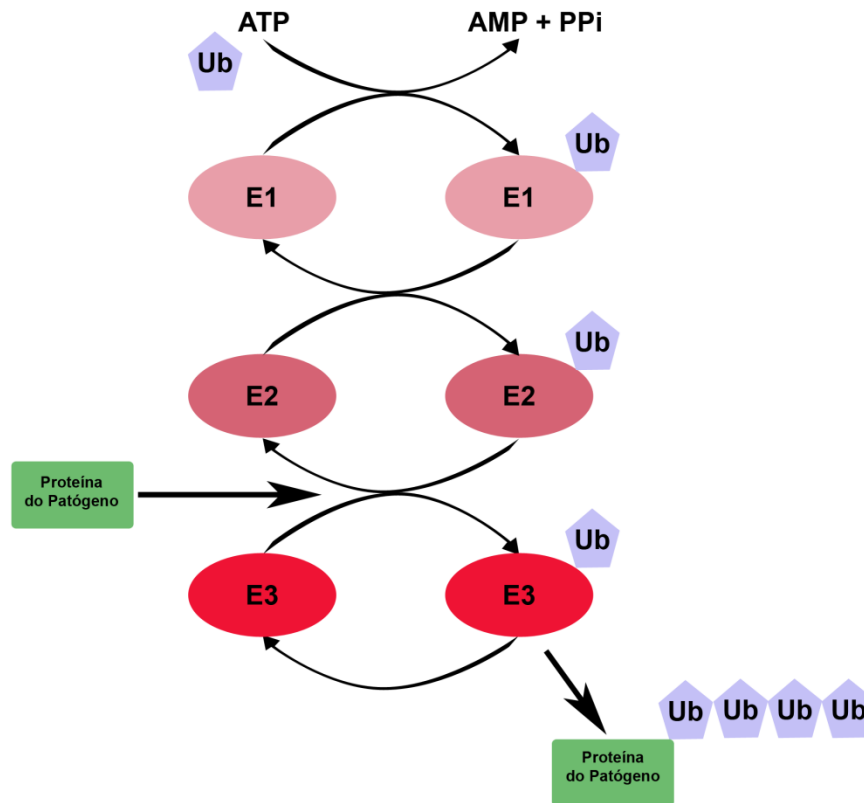


Figura 1. Via de poliubiquitinação de proteínas. A poliubiquitinação inicia com a liberação de energia através da quebra de uma molécula de ATP, o que permite que a enzima E1 ative a ubiquitina. Após, ocorre a passagem da ubiquitina para a enzima E2, deixando a enzima E1 livre para ativar novas ubiquitinas. A seguir, a ubiquitina é transferida para a proteína do patógeno com o auxílio da enzima E3. Finalmente, a proteína poliubiquitinada é direcionada para o proteassomo. Modificado de Choi *et al.*, 2012.

A ligação das múltiplas ubiquitinas à proteína é mediada principalmente por resíduos de lisina presentes na ubiquitina (Pickart & Fushman, 2004).

A proteína recém poliubiquitinada é direcionada para um complexo multienzimático de aproximadamente 2,5 MDa chamado proteassomo. O proteassomo humano é constituído por uma subunidade central, chamada subunidade 20S, e duas subunidades 19S localizadas nas extremidades da subunidade 20S. Juntas, essas subunidades formam o proteassomo 26S (van de Weijer *et al.*, 2015; Voges *et al.*, 1999). A subunidade 19S participa do reconhecimento das poliubiquitinas. Ali, seis subunidades com atividade de ATPase catalisam o desdobramento proteico e promovem a abertura do canal de entrada da subunidade 20S. Esta subunidade, por sua vez, possui sítios catalíticos denominados $\beta 1$, $\beta 2$ e $\beta 5$ responsáveis pela quebra da ligação peptídica envolvendo aminoácidos hidrofóbicos, ácidos ou básicos (Nathan *et al.*, 2013; Rock & Goldberg, 1999). O proteassomo contendo as subunidades catalíticas $\beta 1$, $\beta 2$ e $\beta 5$ é denominado proteassomo constitutivo, uma vez que é encontrado na maioria das células. Entretanto, a liberação de citocinas pró-inflamatórias em um contexto infeccioso, como IFN- γ e TNF- α , promove a expressão de subunidades denominadas $\beta 1i/LMP2$, $\beta 2i/MECL1$ e $\beta 5i/LMP7$, as quais substituem as subunidades $\beta 1$, $\beta 2$ e $\beta 5$, respectivamente, do proteassomo constitutivo, transformando-o em um imunoproteassomo (Nathan *et al.*, 2013). O imunoproteassomo é encontrado em células de atividade imune relevante (Sijts & Kloetzel, 2011) e acredita-se que seu padrão de clivagem diferenciada (cortes mais precisos na extremidade C-terminal da proteína, gerando extremidades com aminoácidos hidrofóbicos ou básicos), é importante para promover a geração de peptídeos com maior afinidade por TAP e pelo MHC-I (Driscoll *et al.*, 1993; McCarthy & Weinberg, 2015).

A clivagem proteica que ocorre no imunoproteassomo gera pequenos peptídeos que usualmente variam entre 3 e 30 aminoácidos (Crews, 2003), os quais são gradativamente liberados no citoplasma, próximo ao retículo endoplasmático rugoso (RER). Parte desses peptídeos são degradados por endopeptidases e aminopeptidases presentes no citoplasma (Lecker *et al.*, 2006; Saveanu *et al.*, 2002), enquanto outra parte consegue se ligar a um transportador associado à apresentação de antígenos (TAP) presente na membrana do RER. O transportador TAP é uma proteína heterodimérica constituída pelas subunidades

TAP1 e TAP2. As duas subunidades são proteínas transmembrana que formam uma abertura na parte citosólica, permitindo que peptídeos de 8 a 16 aminoácidos se liguem e, conseqüentemente, sejam transportados para a parte interna do RER (Parcej & Tampe, 2010). A proteína TAP, em conjunto com as proteínas calreticulina, ERp57 e tapasina formam o complexo de carregamento peptídico (PLC, do inglês *Peptide Loading Complex*) no MHC-I (van de Weijer *et al.*, 2015).

O MHC-I e a apresentação de antígenos

O Complexo Principal de Histocompatibilidade (MHC) foi descoberto há mais de 50 anos (Vandiedonck & Knight, 2009). Entre 1936 e 1937, Peter Gorer foi o primeiro pesquisador a identificar um antígeno, o qual denominou *Antígeno II*, a partir da análise de aglutinação de eritrócitos quando em contato com soro de coelhos. Mais tarde, na década de 40, em murinos, George Snell deu prosseguimento aos trabalhos realizados por Gorer, identificando um sistema de antígenos que estaria envolvido com a rejeição a enxertos e, conseqüentemente, histocompatibilidade (Snell, 1986). Por se tratar do mesmo antígeno identificado por Gorer, esse sistema passou a se chamar H-2 (do inglês, *Histocompatibility-2*), correspondendo ao MHC murino. O estudo de um sistema equivalente em humanos começou no início da década de 50, a partir de uma hipótese do pesquisador Jean Dausset, o qual afirmava que um sistema de antígenos de histocompatibilidade similar ao observado em murinos poderia ser encontrado na superfície de leucócitos humanos. Entretanto, apenas em 1958 sua teoria foi validada (Dausset, 1958). A partir de então, o MHC humano também passou a ser chamado Antígeno Leucocitário Humano (HLA, do inglês *Human Leukocyte Antigen*). Com a evolução dos estudos genéticos, especialmente no que tange aos métodos de mapeamento e sequenciamento genômico, descobriu-se que os genes envolvidos na via de apresentação de antígenos se localizavam no braço curto do cromossomo 6 humano (6p21.3), sítio este denominado região do MHC (Figura 2).

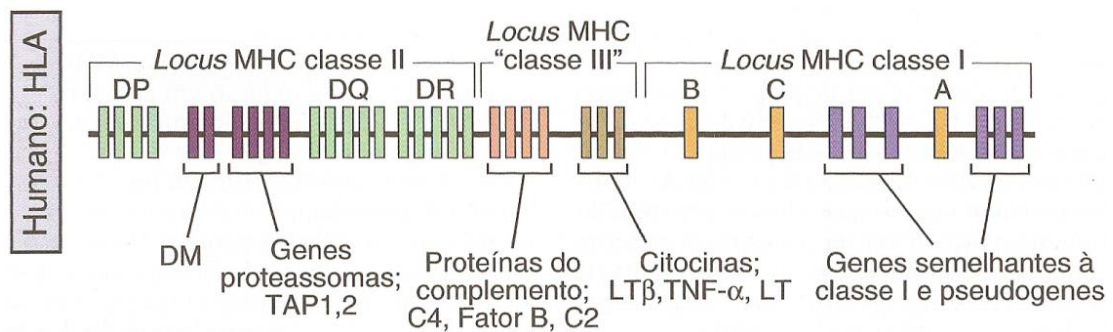


Figura 2. Região do MHC humano. A Região do MHC, encontrada no braço curto do cromossomo 6, subdividida em três *loci* principais: o MHC de classe I, o MHC de classe II e o MHC de classe III. Modificado de *Imunologia Celular e Molecular*, 5ª edição (Abbas & Lichtman, 2004).

A região do MHC é exclusiva de genomas de vertebrados mandibulados, não existindo em vertebrados não mandibulados (Kasahara, 1999). Essa região é subdividida em três *loci* principais (MHC de classe I, MHC de classe II e MHC de classe III), os quais codificam diferentes proteínas envolvidas na resposta imunológica. Além da elevada densidade gênica (em alguns trechos são encontrados até 8,5 genes a cada 100 Kb), essa região é a mais polimórfica do genoma humano (Vandiedonck & Knight, 2009).

O *locus* do MHC-I, em especial, codifica uma proteína (também chamada alótipo (Bordner & Abagyan, 2006)) composta pelos domínios $\alpha 1$, $\alpha 2$ e $\alpha 3$ (cadeia pesada), os quais interagem de forma não-covalente com um domínio $\beta 2$ -microglobulina (codificado a partir do cromossomo 15 humano). Os domínios $\alpha 1$ e $\alpha 2$ interagem formando uma região de fenda, onde o peptídeo/epitopo trazido do citosol através da TAP, com o auxílio do PLC, pode ser inserido (Figura 3). Já a $\beta 2$ -microglobulina auxilia a manter a molécula estável (Berko *et al.*, 2005). A

interação entre epitopo e MHC-I se dá através de ligações de hidrogênio, principalmente em posições de ancoragem do MHC-I, e forças de Van der Waals.

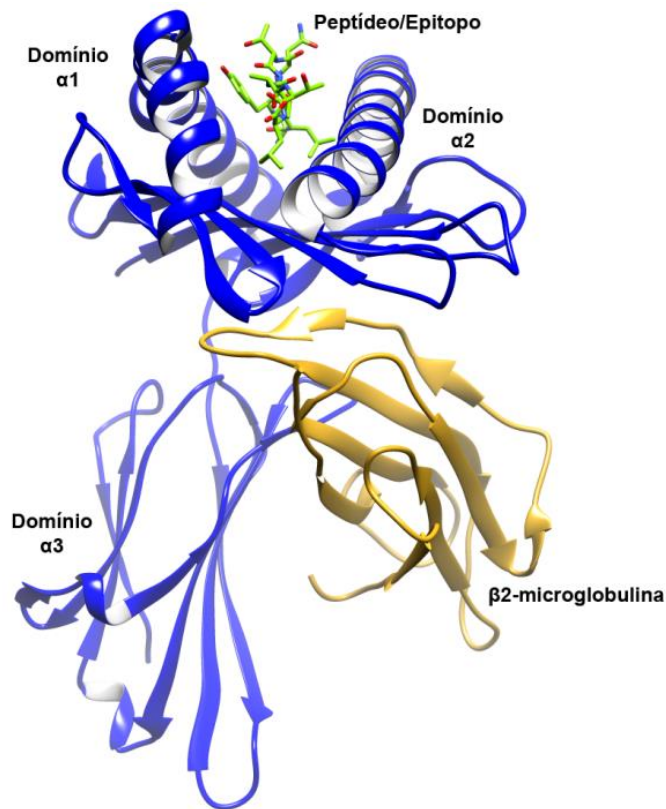


Figura 3. Representação da estrutura tridimensional do MHC-I. Em representação *ribbon*, os domínios $\alpha 1$, $\alpha 2$ e $\alpha 3$ (azul) e a $\beta 2$ -microglobulina (em amarelo). Em representação *stick*, o epitopo (verde).

O maior grau de polimorfismo ocorre na região do MHC-I que expressa os domínios $\alpha 1$ e $\alpha 2$ (Sette & Sidney, 1998), o que gera uma grande diversidade alélica e, por conseguinte, permite que uma ampla variedade de peptídeos possa interagir com o MHC-I. Cada indivíduo pode apresentar até 6 alelos de MHC-I (3 vindos da mãe e 3 vindos do pai) e essa herança se dá na forma de haplótipos (Zuniga *et al.*, 2013). Até o momento, já foram identificados pouco mais de 9.000 alelos compreendendo os *loci* HLA-A, HLA-B e HLA-C. Destes, cerca de 6.000 expressam um alótipo completo (<http://hla.alleles.org/nomenclature/stats.html>). Cada alótipo possui um conjunto de resíduos característicos que compõe a fenda

do MHC-I. De acordo com esses resíduos, cada alótipo é identificado por um nome, o qual contém basicamente um prefixo, o gene expresso, o grupo alélico e a proteína específica. Por exemplo, no alótipo HLA-A*02:01, “HLA” representa o prefixo, “A” representa o gene expresso, “02” representa o grupo alélico e “01” representa a proteína específica. Para mais informações sobre a nomenclatura completa dos alelos de MHC-I, veja o “*Capítulo II - MHC, Viral Infection and Immunoinformatics*”.

A enorme variedade de alótipos confere à molécula de MHC-I a capacidade de se ligar a uma grande diversidade de peptídeos. Estima-se que o conjunto de moléculas de MHC-I expressas por um único indivíduo possui a capacidade de ligar até $2,5 \times 10^6$ peptídeos nãoâmeros (peptídeos com 9 aminoácidos) (Welsh *et al.*, 2010; Wucherpfnig *et al.*, 2007). Considerando-se que o MHC-I possui a capacidade de ligar usualmente peptídeos entre 8 e 14 aminoácidos (Ekeruche-Makinde *et al.*, 2012), esse número pode ser ainda mais expressivo. Entretanto, o estudo a nível molecular da interação peptídeo:MHC-I (pMHC-I) pode se tornar bastante custoso em um universo com mais de 6.000 alótipos de MHC-I, cada um com a capacidade de ligar uma quantidade ainda maior de epitopos. Assim, em meados da década de 90, iniciou-se a caracterização de supertipos (ou supermotivos) de MHC-I baseado em propriedades físico-químicas de grupos de aminoácidos presentes na região de fenda (Sette & Sidney, 1998; Sidney *et al.*, 1996). Esses grupos de resíduos se arranjam formando pequenos “bolsões” (denominados *pockets*), os quais interagem de forma específica com diferentes partes do epitopo. As moléculas de MHC-I possuem um conjunto de *pockets*, nomeados de A até F, cuja constituição e afinidade por resíduos do ligante varia entre os diferentes alótipos. Por exemplo, o *pocket* B de HLA-A*02:01, o qual interage com o aminoácido da posição 2 do epitopo, é compreendido pelos resíduos 7, 9, 24, 34, 45, 63, 66, 67, 70, e 99, os quais possuem a característica de serem pequenos e alifáticos (A, T, S, V, L, I, M e Q), da mesma forma que o alótipo HLA-A*68:27. Dessa forma, ambos os alótipos são agrupados em um mesmo supertipo chamado A02 (Sidney *et al.*, 2008).

Uma vez na superfície da célula, o pMHC-I passará pelo escrutínio do Receptor de Célula T (TCR), presente em linfócitos T CD8⁺ (Figura 4). O reconhecimento do pMHC-I se dá através de contatos específicos entre resíduos do epítopo, resíduos do MHC-I e resíduos do TCR. A maior parte dos linfócitos encontrados no nosso corpo envolvidos na resposta imune adaptativa possuem em sua superfície TCRs compostos pelos domínios proteicos α e β . Na região de interação de cada domínio são encontradas três Regiões Determinantes de Complementariedade, denominadas CDRs (CDR1, CDR2 e CDR3). As CDR1 e CDR2, normalmente, reconhecem resíduos do MHC-I, enquanto a CDR3 está envolvida no reconhecimento de aminoácidos do epítopo, mais especificamente os resíduos da região central do epítopo (Rossjohn *et al.*, 2015). O reconhecimento do pMHC-I pelo TCR se dá de forma degenerada, de forma que um único TCR pode reconhecer até 1.000.000 de peptídeos diferentes ligados ao mesmo MHC-I (Wooldridge *et al.*, 2011). Ao processo de reconhecimento e geração de uma resposta imunológica de mais de um pMHC-I pelo mesmo TCR se dá o nome de reatividade cruzada (Baker *et al.*, 2012; Motozono *et al.*, 2014; Sinigaglia *et al.*, 2013).

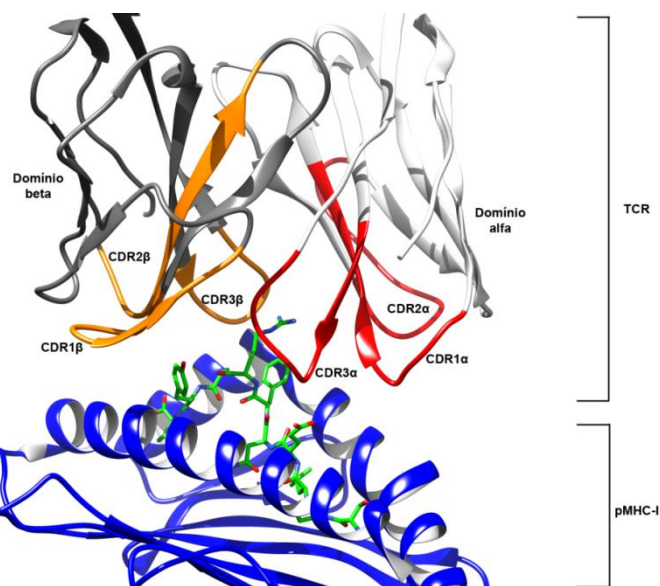


Figura 4. Interação entre pMHC-I e TCR. O reconhecimento do complexo pMHC-I (MHC-I em azul e epítopo em verde) ocorre através da interação com as regiões determinantes de complementariedade (CDRs) presentes no TCR. É possível observar o contato mais próximo entre CDR3 α /CDR3 β com o epítopo.

Para que ocorra o reconhecimento, com desencadeamento da resposta imunológica, é importante que alguns fatores sejam levados em conta, como a complementariedade de topografia e cargas entre pMHC-I e TCR, bem como a estabilidade do pMHC-I na superfície celular por determinado tempo (Harndahl *et al.*, 2012; He *et al.*, 2014; Jorgensen *et al.*, 1992; Wooldridge *et al.*, 2012). Um ponto de discussão interessante que vem ganhando importância, especialmente na última década, diz respeito ao papel combinado da afinidade do epitopo pelo MHC-I e da estabilidade do complexo pMHC-I com a imunogenicidade. Estudos mostram que existe uma relação direta entre a afinidade e a estabilidade, ou seja, é mais provável que um complexo pMHC-I formado com um peptídeo com elevada afinidade apresente maior estabilidade. Entretanto, o fato de um peptídeo não possuir elevada afinidade não significa que o complexo, necessariamente, não se manterá estável o suficiente para permitir o reconhecimento pelas células T (Harndahl *et al.*, 2012; Hoppes *et al.*, 2014; Narzi *et al.*, 2012; van der Burg *et al.*, 1996; Wu *et al.*, 2012; Yanaka *et al.*, 2014). Para obter um melhor entendimento de como características moleculares podem influenciar na imunogenicidade de diferentes complexos pMHC-I, ferramentas de imunoinformática são comumente empregadas.

Imunoinformática: a bioinformática aplicada ao estudo da Imunologia

A bioinformática é uma área de estudo interdisciplinar que começou a surgir durante o desenvolvimento do Projeto Genoma Humano. Embora ferramentas computacionais já fossem aplicadas de forma limitada ao estudo de moléculas biológicas, no início da década de 90 houve uma grande mobilização de recursos financeiros e a criação de novas demandas computacionais, como consequência da corrida pela publicação do primeiro rascunho do sequenciamento do genoma humano pelos pesquisadores Francis Collins (consórcio público) e Craig Venter (consórcio privado). Assim, se fez necessário o desenvolvimento de ferramentas de informática que pudessem auxiliar a estocar, analisar e processar os dados obtidos a partir do sequenciamento de DNA (Ikekawa & Ikekawa, 2001).

O grande volume de dados começou a dar origem a diferentes áreas de pesquisa, as chamadas ciências “ômicas”, como a genômica (estudo do genoma), a transcriptômica (estudo do transcriptoma), a proteômica (estudo do proteoma), entre outras. Nesse contexto, como era de se esperar, surgiu também a imunômica, que nada mais é do que o estudo do imunoma humano, ou seja, dos genes e proteínas envolvidas em reações imunológicas (Tomar & De, 2010). Nesse contexto, em 1989, é criado o IMGT[®] (do inglês, *the international ImMunoGeneTics information system*), o qual marca o surgimento de uma nova área na interface entre a imunogenética e a bioinformática, a chamada imunoinformática.

A imunoinformática está ligada a diferentes áreas dentro da imunologia, como o estudo da interação patógeno-hospedeiro, a predição de epitopos de células T e células B, vacinação *in silico*, identificação de genes ligados ao sistema imunológico, criação de bancos de dados imunômicos, entre outros (Tomar & De, 2010). Dentro desse contexto, a imunoinformática foi extremamente importante para o desenvolvimento de uma área chamada vacinologia reversa (Rappuoli, 2000). Métodos tradicionais de desenvolvimento de vacinas iniciam-se com o uso de técnicas bioquímicas, sorológicas e microbiológicas para identificação de componentes imunogênicos presentes em patógenos. Após várias etapas de caracterização e isolamento do alvo, pode-se chegar ao sequenciamento e clonagem do gene de interesse, para posterior expressão da proteína e testes de imunização *in vivo*. Entretanto, nem sempre é possível aplicar essas técnicas, seja por dificuldades de cultivar o patógeno *in vitro*, seja pela ampla variabilidade apresentada pelos patógenos, mesmo aqueles relacionados. A vacinologia reversa, como o próprio nome sugere, faz uso da informação genômica e imunômica para análise e identificação, *in silico*, de genes candidatos que, possivelmente, codificam antígenos envolvidos em eventos imunogênicos. Dessa forma, a informação advinda da imunoinformática é utilizada para o desenvolvimento e aplicação de novas vacinas (Davies & Flower, 2007; Donati & Rappuoli, 2013).

Imunoinformática e o MHC-I

A área da imunoinformática tem crescido bastante nos últimos anos, sendo que novas ferramentas, técnicas e bancos de dados tem sido desenvolvidos, a grande maioria direcionados para a área da vacinologia (Khalili *et al.*, 2014; Khan *et al.*, 2015; Schubert *et al.*, 2015). Tem havido grande interesse especificamente no estudo da via de processamento de peptídeos endógenos e interação com o MHC-I. Entre os programas, podemos citar (i) o NetChop 3.1, um servidor online baseado em redes neurais e dados experimentais com o objetivo de prever a clivagem proteica realizada pelo proteassomo (Nielsen *et al.*, 2005), o (ii) TAPPred, ferramenta utilizada para prever a afinidade de peptídeos gerados pelo proteassomo à proteína TAP (Bhasin & Raghava, 2004) e (iii) o NetMHC e o BIMAS, ambos direcionados para a predição de afinidade de ligação entre epitopo e MHC-I (Lundegaard *et al.*, 2008; Parker *et al.*, 1994). Também existem servidores que permitem a análise conjunta de todos os passos da via de processamento de antígenos endógenos, como é o caso do IEDB (*Immune Epitope Database and Analysis Resource*) (Vita *et al.*, 2014). Mais do que uma plataforma para ferramentas de imunoinformática, o IEDB também é um dos principais bancos de dados experimentais relacionados a epitopos de células T e B.

Além das ferramentas de análise de sequência, outras ferramentas aplicadas ao estudo estrutural das moléculas podem ser utilizadas, como por exemplo, a dinâmica molecular (DM). A DM se baseia na resolução de um conjunto de equações da mecânica clássica (contemplando as Leis de Newton), as quais são utilizadas para prever a movimentação dos átomos que compõe uma molécula ao longo do tempo de simulação. Um dos primeiros estudos realizados na área biológica utilizando a DM foi para averiguar o dobramento de uma proteína, mais especificamente, um inibidor de tripsina pancreática bovina (Levitt & Warshel, 1975). Isso foi na década de 70, quando as restrições computacionais não permitiam o estudo de grandes moléculas. Entretanto, as possíveis aplicações da dinâmica molecular já começavam a ser exploradas amplamente. Três pesquisadores de destaque na área (Martin Karplus, Michael Levitt e Arieh Warshel), os quais firmaram várias das bases para o estudo de predição de

interações atômicas na década de 70 utilizando dinâmica molecular, receberam em 2013 o Prêmio Nobel em Química (http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/). No campo da imunologia, a DM pode ser utilizada para uma série de estudos, como a avaliação do comportamento de moléculas quando em uma solução aquosa, o grau de flexibilidade de resíduos, a interação entre receptores e ligantes, entre outros.

Basicamente, a simulação de dinâmica molecular começa com a construção de uma caixa virtual, a qual irá conter a proteína de interesse, íons que estabilizarão o sistema e água (Figura 5). Após, escolhe-se um campo de força, o qual conterá as informações necessárias para a realização dos cálculos energéticos do sistema. O campo de força deve ser capaz de calcular a energia proveniente a partir de átomos que interagem através de ligações covalentes e não-covalentes. Ao final da simulação, é possível extrair os dados que refletem o comportamento da molécula no ambiente em que ela foi posicionada.

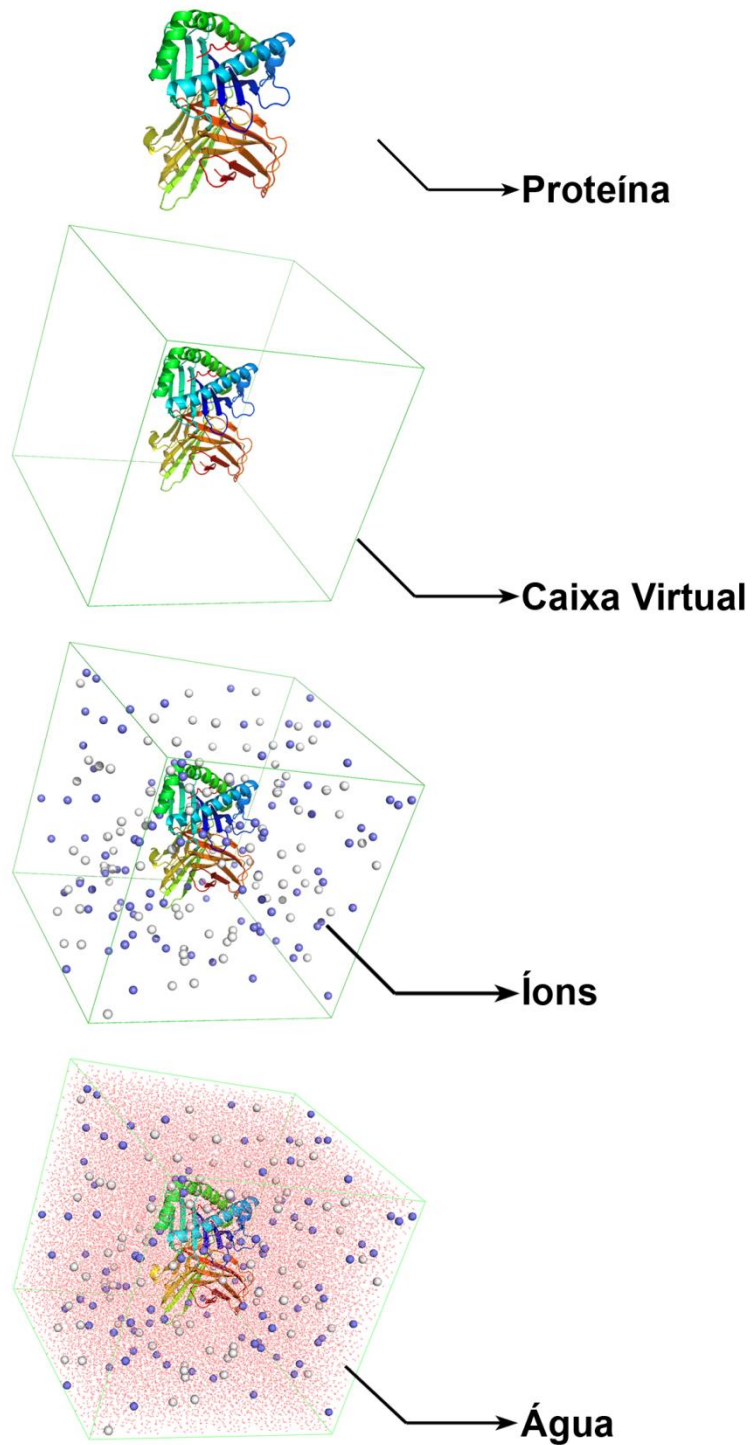


Figura 5. Construção de um sistema para realização de dinâmica molecular. Primeiramente, escolhe-se a proteína de interesse, preparando-a individualmente. Após, o sistema é preparado dentro de uma caixa virtual criada pelo programa que irá rodar a simulação, acrescentando-se íons e água.

Considerando a importância da imunoinformática para a resolução de problemas relacionados à via de processamento de antígenos endógenos, em 2010, nosso grupo empregou esforços para o desenvolvimento de uma técnica baseada em padrões alelo-específicos de MHC-I para prever o modo de ligação de epitopos diversos (Antunes *et al.*, 2010). Brevemente, a técnica inicia a partir do momento em que o usuário fornece uma sequência linear de um peptídeo de interesse. Essa sequência é transformada em uma estrutura tridimensional, a qual varia de acordo com o alótipo escolhido. Em seguida, o programa AutoDock Vina (Trott & Olson, 2009), específico para ancoragem molecular entre ligante e receptor, procura pela melhor conformação do epitopo dentro da fenda do MHC-I. A melhor conformação é escolhida e o complexo pMHC-I passa por uma minimização de energia. Após, é realizada mais uma rodada de ancoragem molecular, visando o refinamento da estrutura final. A técnica, denominada *D1-EM-D2*, foi validada utilizando mais de 40 complexos pMHC-I do PDB, principal banco de dados com estruturas tridimensionais resolvidas experimentalmente. A partir desta técnica, conseguimos modelar complexos com características específicas para cada alótipo permitido (HLA-A*02:01, HLA-B*27:05, H-2-Db e H-2-Kb), cada um contendo epitopos com diferentes composições de aminoácidos e diferentes tamanhos. Essa técnica permitiu, conseqüentemente, a criação de um banco de dados com estruturas tridimensionais modeladas a partir da técnica *D1-EM-D2*, o qual chamamos de CrossTope (Sinigaglia *et al.*, 2013).

A concepção sobre a necessidade de se criar uma ferramenta para modelagem de complexos pMHC-I surge a partir do momento em que observamos (i) a importância no estudo de estruturas tridimensionais para responder questões imunológicas e (ii) a falta de ferramentas adequadas para modelagem e estudo das interações entre epitopo e MHC-I. Comparando-se o número de sequências proteicas depositadas no NCBI com o número de proteínas com estruturas tridimensionais resolvidas experimentalmente depositadas no PDB, observamos que o número de sequências excede em mais de 450 vezes o número de estruturas. Assim, sentimos a necessidade de criar uma técnica automatizada, validada sobre um amplo conjunto de dados e, principalmente, acessível para qualquer usuário interessado em modelar

complexos pMHC-I. Ainda, a partir da modelagem desses complexos, também vislumbramos a criação de uma ferramenta de predição de estabilidade de complexos pMHC-I que, combinada com outras ferramentas, possa ser utilizada para avaliar imunogenicidade.

Objetivos

Considerando que o desencadeamento da resposta imunológica através de células T CD8+ está intimamente relacionado com elementos estruturais de complexos pMHC-I, bem como com a estabilidade destes complexos na superfície das células, este trabalho teve os seguintes objetivos:

Objetivo Geral

Realizar a predição estrutural e inferir a estabilidade de complexos pMHC-I através de uma abordagem computacional.

Objetivos Específicos

- i. Desenvolver um método automatizado para construção de complexos pMHC-I.
 - Encadear, parametrizar e programar os passos da técnica *D1-EM-D2*;
 - Revalidar a técnica *D1-EM-D2* após automatização, com inclusão de novos parâmetros de escolha do melhor modelo e um conjunto maior de dados experimentais;
 - Desenvolver uma interface *web* para a ferramenta (DockTope) a fim de permitir que usuários sem conhecimento computacional prévio sejam capazes de modelar complexos pMHC-I de interesse.

- ii. Desenvolver uma nova abordagem para avaliar a estabilidade de complexos pMHC-I utilizando dinâmica molecular.
 - Criar um protocolo utilizando simulações de dinâmica molecular com diferentes condições sobre as estruturas analisadas;
 - Analisar e selecionar os melhores parâmetros indicadores de estabilidade, a fim de utilizá-los posteriormente para análise de um conjunto maior de dados.

Capítulo II

MHC, Viral Infection and Immunoinformatics

(Capítulo publicado no livro "*Major Histocompatibility Complex: Biology, Functions and Roles in Disease*")

Chapter 6

MHC, VIRAL INFECTION AND IMMUNOINFORMATICS

Maurício Menegatti Rigo, Dinler Amaral Antunes, Marialva Sinigaglia, José Artur Bogo Chies and Gustavo Fioravanti Vieira

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

ABSTRACT

The immunologic surveillance is performed by several molecules (e.g. antibodies and complement system) and cells (e.g. T lymphocytes, Natural Killer cells) which are triggered depending on the type of infection. In the specific context of the viral infection, a special set of molecules, named Major Histocompatibility Complex of class I (MHC-I), which is represented by the Human Leukocyte Antigen (HLA) genes in humans, perform an essential role. These molecules are responsible for 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. Regarding MHC-I presentation it is known that topologies and electrostatics patterns of the peptide:MHC (pMHC) interacting region are important for TCR recognition. Also, the affinity and stability between the peptide (epitope) and the MHC (receptor) play a crucial role, especially on what concerns electrostatics and non-hydrophobic interactions between the MHC and anchor residues of the peptide. All these features stand for the peptide immunogenicity. The knowledgment of issues regarding peptide affinity by MHC cleft and the stability of resulting complex is pivotal for vaccine development field. Nowadays, there are several immunoinformatics tools available to assist the resolution of this task. For instance, peptide binding affinity could be predicted through several computational methods (i.e. using tools that consider amino acid sequence composition and preferential anchor residues occurrence). There are also bioinformatics tools that are not specific for immunological issues, but can be used for this purpose, such as molecular docking and molecular dynamics. The latter can be used to assess pMHC complex stability, one of the main factors involved in peptide presentation. In the present chapter we intend to review basic molecular aspects regarding the MHC-I molecule and its interaction with different epitopes. Also, the use of

Capítulo III

DockTope: a Web-based tool for automated pMHC-I modelling

(Artigo completo submetido para a revista “*Scientific Reports*”)

Title

DockTope: a Web-based tool for automated pMHC-I modelling.

Author list and affiliations

Maurício Menegatti Rigo^{1,+,*}

Dinler Amaral Antunes^{1,2,+}

Martielva Vaz de Freitas¹

Marcus Fabiano de Almeida Mendes¹

Lindolfo Meira³

Marialva Sinigaglia¹

Gustavo Fioravanti Vieira^{1,*}

¹ Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Postcode 91501-970, Brazil;

² Department of Computer Science, Rice University, Houston, Texas, 77005, USA.

³ CESUP - Centro Nacional de Supercomputação, Universidade Federal do Rio Grande do Sul, Postcode 90035-190, Brazil;

⁺ These authors contributed equally to this work.

^{*} Corresponding author #1: gusfioravanti@yahoo.com.br

^{*} Corresponding author #2: mauriciomr1985@gmail.com

Abstract

The immune system is constantly challenged, being required to protect the organism against a wide variety of infectious pathogens and, at the same time, to avoid autoimmune disorders. One of the most important molecules involved in these events is the Major Histocompatibility Complex class I (MHC-I), which is responsible for bind and present small peptides from intracellular environment to CD8⁺ T cells. The study of peptide:MHC-I (pMHC-I) molecules at a structural level is crucial to understand the molecular mechanisms underlying the immunologic responses. Unfortunately, there are few structures currently available on the Protein Data Bank (PDB) and pMHC-I modelling approaches are not fully available. Here, we presented the DockTope, a free and reliable web-based tool for pMHC-I modelling, based on crystal structures from PDB. DockTope is fully automated and allows any researcher to construct its own pMHC-I complex in an efficient way. In order to validate our new methodology we reproduced a dataset of 135 non-redundant pMHC-I structures from PDB (C α RMSD below 1 Å). The pMHC-I modelling is remarkably important since it takes the immunology study to the next level, contributing to the knowledge of important events such as cross-reactivity, autoimmunity, cancer therapy, transplantation and rational vaccine design.

Capítulo IV

Peptide:MHC-I Stability Assayed through *in silico* Molecular Dynamics

(Artigo completo a ser submetido para a revista “*PLOS ONE*”)

Full Title: Peptide:MHC-I Stability Assayed through *in silico* Molecular Dynamics

Short Title: *In silico* Inference of pMHC-I stability

Maurício Menegatti Rigo^{1,*}, Dinler Amaral Antunes^{1,2}, Marialva Sinigaglia¹,
Gustavo Fioravanti Vieira^{1,*}

¹ Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

² Department of Computer Science, Rice University, Houston, Texas, USA

*Corresponding authors

E-mail: gusfioravanti@yahoo.com.br

E-mail: mauriciomr1985@gmail.com

Abstract

The immune system is constantly challenged by a variety of pathogens, being required to protect the organism against these invaders. One of the defence mechanisms employs a set of molecules called MHC-I, which are able to bind short peptides (epitopes), forming a final complex called peptide:MHC-I (pMHC-I). The pMHC-I is displayed at the cell surface where it will pass through the scrutiny of a T Cell Receptor (TCR). The interaction between TCR and pMHC-I will define the immunologic response, and this encounter will also depend on the pMHC-I structural stability. Here, we aimed to define a protocol based on molecular dynamics of two epitopes presented by the MHC-I HLA-A*02:01 in order to establish a new way to assess *in silico* pMHC-I stability. To do that, we first modelled two independent epitopes, one with recognized immunogenicity (CVNGVCWTV) and an Altered Peptide Ligand (APL) with modified anchor positions (CKNGVCWTS). Each pMHC-I was submitted to a set of molecular dynamics protocols using different constraints over specific MHC-I atoms groups. We were able to identify a protocol capable to differentiate the two assayed pMHC-I complexes based on stability patterns, such as root mean square deviation, root mean square fluctuation and hydrogen bonding between key residues. We also investigated at molecular level which residues are being responsible for this differential stability. The findings here can be used to determine novel targets for vaccine approaches before *in vitro/in vivo* testing. Moreover, this study also sets the foundation for the development of a more comprehensive tool to further investigation on immunogenicity of pMHC-I complexes.

Capítulo V

Discussão Geral

Discussão Geral

Este trabalho teve como objetivo a otimização e automatização de uma ferramenta online, a qual denominamos DockTope, voltada para a modelagem de complexos tridimensionais de pMHC-I, tendo como arquivo de entrada apenas a sequência FASTA do epitopo a ser modelado. Além disso, também propomos a criação de uma ferramenta baseada em dinâmica molecular para avaliar a estabilidade *in silico* de complexos pMHC-I.

Em um momento de intensa investigação de sequências de DNA, durante a década de 90, a imunoinformática surgiu para preencher uma lacuna e auxiliar na identificação e análise de dados relativos ao sistema imunológico (Lefranc, 2014). O contínuo crescimento do sequenciamento, dando origem às diferentes ciências “ômicas”, aliado às ferramentas de imunoinformática, possibilitou que se firmassem as bases para o surgimento da vacinologia reversa (Rappuoli, 2000). A vacinologia reversa traz consigo duas grandes vantagens: (i) a possibilidade de se analisar proteomas inteiros, buscando alvos que até então não poderiam ser identificados de maneira eficaz através de experimentos *in vitro/in vivo* e (ii) a possibilidade de busca de alvos vacinais contra patógenos difíceis de se cultivar *in vitro* (Donati & Rappuoli, 2013). Dentro desse contexto, a imunoinformática avança concomitantemente com a quantidade de dados gerados, e ferramentas de predição começam a surgir, baseadas especialmente na estrutura primária de proteínas. Um conjunto significativo dessas ferramentas aborda a via de apresentação de antígenos endógenos, uma vez que esta é a principal via envolvida na apresentação de peptídeos oriundos de patógenos intracelulares para o sistema imunológico, como é o caso dos agentes infecciosos virais.

Existe atualmente uma série de ferramentas baseadas em sequências envolvidas na predição de diferentes partes da via de processamento de antígenos endógenos, desde a predição de corte pelo proteassomo, até a predição de afinidade do epitopo pelo MHC-I. Por exemplo, o programa NetChop é um programa baseado em ANN que utiliza dados *in vitro* advindos da clivagem de β -caseína bovina e enolase de levedura pelo proteassomo 20S para predizer a clivagem de qualquer proteína, desde que fornecida a sua sequência FASTA

(Nielsen *et al.*, 2005). Já o TAPPred é um programa baseado em SVM que utiliza métodos de aprendizagem estatística. Este programa foi treinado com um grupo de mais de 400 peptídeos, os quais demonstraram diferentes graus de afinidade pelo transportador TAP *in vitro* (Bhasin & Raghava, 2004). Já no contexto da apresentação de antígenos na superfície celular, existem vários programas direcionados para avaliar a afinidade do epitopo pelo MHC-I. Dois deles, o NetMHC (Lundegaard *et al.*, 2008) e o BIMAS (Parker *et al.*, 1994), são amplamente utilizados. O NetMHC foi criado com base em dados oriundos do IEDB (Vita *et al.*, 2014), SYFPEITHI (Rammensee *et al.*, 1999) e dados de afinidade obtidos no próprio laboratório do grupo que criou a ferramenta. Da mesma forma que o NetChop, ele é baseado em ANN. Já o servidor BIMAS foi construído com base em um compilado de dados da literatura e se baseia em uma matriz de distribuição da probabilidade de se encontrar determinado aminoácido em cada posição do epitopo.

A análise de sequências proteicas baseada na estrutura linear é importante, entretanto, deixa de responder algumas questões que poderiam ser mais bem acessadas através da estrutura terciária da proteína. No âmbito imunológico, a utilização de dados estruturais pode permitir (i) a avaliação da afinidade do epitopo pelo MHC-I com dados de área acessível ao solvente, volume ocupado na fenda do MHC-I, presença ou não de moléculas de água intermediando ligações de hidrogênio, ângulos de ligação entre pares de aminoácidos; não se restringindo apenas à análise de suas características físico-químicas, (ii) avaliação de modos de ligação entre TCR e pMHC, (iii) desenho de fármacos e análogos peptídicos para melhorar ligação do epitopo com MHC, entre outros (Bordner & Abagyan, 2006; Hoppes *et al.*, 2014).

Entretanto, o uso em larga escala de estruturas tridimensionais é limitado pela baixa disponibilidade dessas estruturas, que para serem resolvidas experimentalmente precisam do emprego de técnicas como Cristalografia e Difração de Raios-X e Ressonância Magnética Nuclear, as quais exigem tempo, dinheiro e mão de obra altamente qualificada. Uma consulta ao PDB, principal banco de dados de proteínas tridimensionais, nos mostra que atualmente há

cerca de 109.000 estruturas depositadas (<http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>). A título de comparação, esse valor é mais de 450 vezes menor do que o número de sequências proteicas depositadas no UniProtKB, banco de dados de sequências proteicas curadas, que totaliza cerca de 50.000.000 sequências (<http://www.uniprot.org/uniprot/>). Restringindo-se a análise, uma busca utilizando as palavras-chave “MHC”, “HLA” e “Homo sapiens” retorna cerca de 50.000 resultados do UniProtKB, enquanto apenas cerca de 700 resultados do PDB (<http://www.rcsb.org/pdb/home/home.do>).

Nesse sentido, uma abordagem utilizando a modelagem molecular se torna interessante e bastante viável, haja visto que atualmente encontra-se disponível um poderio computacional adequado para cumprir estas tarefas num tempo razoável e softwares dedicados para essa função. A modelagem molecular de epitopos no contexto do MHC-I, entretanto, deve ser feita com cautela. Há cerca de 6.000 alelos de MHC-I, alguns codificam alótipos com diferenças em apenas um único aminoácido, mas que pode interferir de maneira significativa no modo de ligação de cada epitopo. Por exemplo, os alótipos HLA-B*27:05 e HLA-B*27:09 diferem em apenas um aminoácido localizado na região de ligação ao epitopo (ocorre a troca de um Ácido Aspártico por uma Histidina na posição 116). Essa diferença já é suficiente para que ocorra a ligação de um conjunto diferente de peptídeos (Ramos *et al.*, 2002). Este tipo de apresentação diferencial de alelos de MHC-I distintos pode tanto promover um ganho na capacidade de responder a um novo patógeno como também pode levar a efeitos danosos como a predisposição a patologias autoimunes, como é o caso da espondilite anquilosante nos indivíduos que expressam o HLA-B*27:09 (Narzi *et al.*, 2012). Além disso, outros obstáculos estão presentes na modelagem de epitopos no contexto do MHC-I, pois alguns peptídeos, mesmo que diferentes entre si, mantêm uma energia de ligação ao MHC-I parecida, o que torna difícil distingui-los através de funções energéticas utilizadas pelos programas de modelagem. Além disso, o elevado número de ligações flexíveis nos aminoácidos do epitopo e do MHC-I amplia de forma exponencial o espaço conformacional a ser analisado pelos programas (Bordner & Abagyan, 2006).

É nesse contexto que desenvolvemos uma técnica de modelagem molecular, chamada *D1-EM-D2* (Figura 1 do Capítulo III) (Antunes *et al.*, 2010), para peptídeos no contexto do MHC-I (mais especificamente para os alótipos humanos HLA-A*02:01 e HLA-B*27:05 e murinos H-2-Db e H-2-Kb). No presente trabalho essa técnica foi aprimorada, validada sobre um conjunto maior de dados e disponibilizada em uma versão online, a qual denominamos DockTope. Existem três técnicas principais atualmente utilizadas para modelagem de pMHC-I: (i) o MHCsim (Todman *et al.*, 2008), (ii) o pDOCK (Khan & Ranganathan, 2010) e (iii) um protocolo de ancoragem molecular utilizando Método de Monte Carlo (Bordner & Abagyan, 2006). O MHCsim, foi o primeiro servidor automatizado e disponibilizado *online* voltado para a construção de complexos pMHC. O método utilizado, basicamente, lê as sequências de entrada (MHC e epitopo) e realiza a busca em um banco de dados interno pela estrutura de pMHC, escolhendo as estruturas que apresentam maior similaridade em relação às sequências de entrada. Após, os resíduos do pMHC são simplesmente substituídos pelos resíduos da sequência de entrada, gerando o modelo tridimensional. Dessa maneira, o DockTope apresenta algumas vantagens, uma vez que ele não é baseado apenas na similaridade de sequências, mas sim em dados estruturais obtidos a partir de padrões alelo-específicos apresentados pelos epitopos recuperados de cristais do PDB. Além disso, é possível construir complexos pMHC-I no DockTope com epitopos de diferentes tamanhos, não apenas nonâmeros como no caso do MHCsim. Outro ponto importante a ser destacado é que o DockTope permite a construção tanto de complexos humanos quanto de murinos e o MHCsim apenas de humanos. Isso é um ganho interessante, uma vez que trabalhos de prospecção de antígenos são primeiramente realizados em modelos animais e os dois alótipos (H-2-Db e H-2-Kb) são amplamente frequentes em linhagens isogênicas de camundongos C57BL/6 (Abbas & Lichtman, 2004).

Por sua vez, o pDOCK é uma metodologia utilizada para modelagem de complexos pMHC baseada principalmente em um algoritmo de otimização global chamado ICM (do inglês, *Internal Coordinate Mechanics*), amostragem Monte Carlo e minimização local. Os autores validaram a técnica sobre um grupo de 186

estruturas pMHC-I e pMHC-II com valores de RMSD (C α) abaixo de 1 Å. Esse baixo valor de RMSD, entretanto, pode ser atribuído há maneira como a validação foi feita, uma vez que os autores utilizaram *re-docking*, ao invés de *cross-docking* (método utilizado para validar o DockTope). Enquanto no *re-docking* o receptor e o ligante são separados para, posteriormente, o ligante ser ancorado no mesmo receptor, no *cross-docking* o ligante a ser ancorado não estava complexoado àquele receptor especificamente. Ainda, um dos grandes problemas desta técnica é o fato dela não estar disponível para os usuários, sendo mantida como um protocolo exclusivo do grupo que a desenvolveu. O método publicado por Bordner & Abagyan em 2006 também é baseado em um algoritmo ICM de otimização global, assim como modelagem por homologia (para gerar as estruturas tridimensionais) e SVM (para treinar a função de escore da ancoragem molecular). Os autores validaram a técnica através de *cross-docking* obtendo valores de RMSD (C α) abaixo de 1 Å para 14 estruturas de HLA-A*02:01 e 9 estruturas de H-2-Kb. As grandes desvantagens dessa ferramenta constituem-se no fato de que a validação foi feita sobre um pequeno grupo de estruturas, além de ela não ser disponibilizada online.

O DockTope surge como uma plataforma online que permite que qualquer usuário consiga modelar um complexo pMHC-I, independentemente do seu grau de conhecimento de ferramentas de computação. Ao contrário das técnicas anteriormente citadas, aqui foram utilizados dados estruturais baseados no padrão conformacional adotado pelo epitopo quando complexoado a cada um dos alótipos. Os padrões foram obtidos a partir da análise direta de 135 estruturas cristalografadas, sendo 68 de HLA-A*02:01, 10 de HLA-B*27:05, 38 de H-2-Db e 19 de H-2-Kb. Os dados foram validados através de *cross-docking* com uma média de RMSD (C α) e RMSD (todos os átomos) de 0.882 Å e 1.964 Å, respectivamente.

Quando se fala em validação de dados de modelagem molecular de proteínas, buscamos alcançar um valor de RMSD abaixo de 2 Å ao comparar a estrutura modelada com o seu alvo (Bagaria *et al.*, 2011; Madurga *et al.*, 2005). Esse valor é utilizado como padrão e foi obtido a partir de uma média geral

baseada na resolução de estruturas cristalografadas. Para melhorar a acurácia da nossa validação, entretanto, nós resgatamos os valores de resolução apenas dos cristais envolvidos na análise (Figura 3 do Capítulo III). Para cada alótipo, obtivemos um valor médio de resolução. Observamos que os valores de validação, comparando-se o RMSD ($C\alpha$) entre o modelo e o seu alvo, foram significativamente menores do que os valores de resolução dos cristais.

Uma das razões para os bons resultados apresentados pelo DockTope está na forma como as estruturas são geradas e escolhidas (Figura 6). O programa AutoDock Vina (Trott & Olson, 2009) é responsável por gerar um conjunto de conformações a partir do arquivo de entrada, a cada rodada de ancoragem molecular. Primeiramente, as melhores conformações são ranqueadas de acordo com a energia de ligação (calculada pelo programa). No caso do DockTope, programamos uma linha de código que permite que o programa rode 20 vezes, resultando na geração de até 1000 conformações. Após cada rodada, recuperamos a melhor conformação, levando-se em consideração valores de energia de ligação. Posteriormente, realiza-se um cálculo para recuperar a conformação média, ou seja, aquela que apresenta menor desvio quadrático médio (RMSD) em relação às 20 conformações escolhidas inicialmente (Capítulo III para mais detalhes). Esse cálculo foi desenvolvido especialmente para o DockTope e leva em consideração não apenas a energia de ligação, mas também a conformação média do grupo mais representativo.

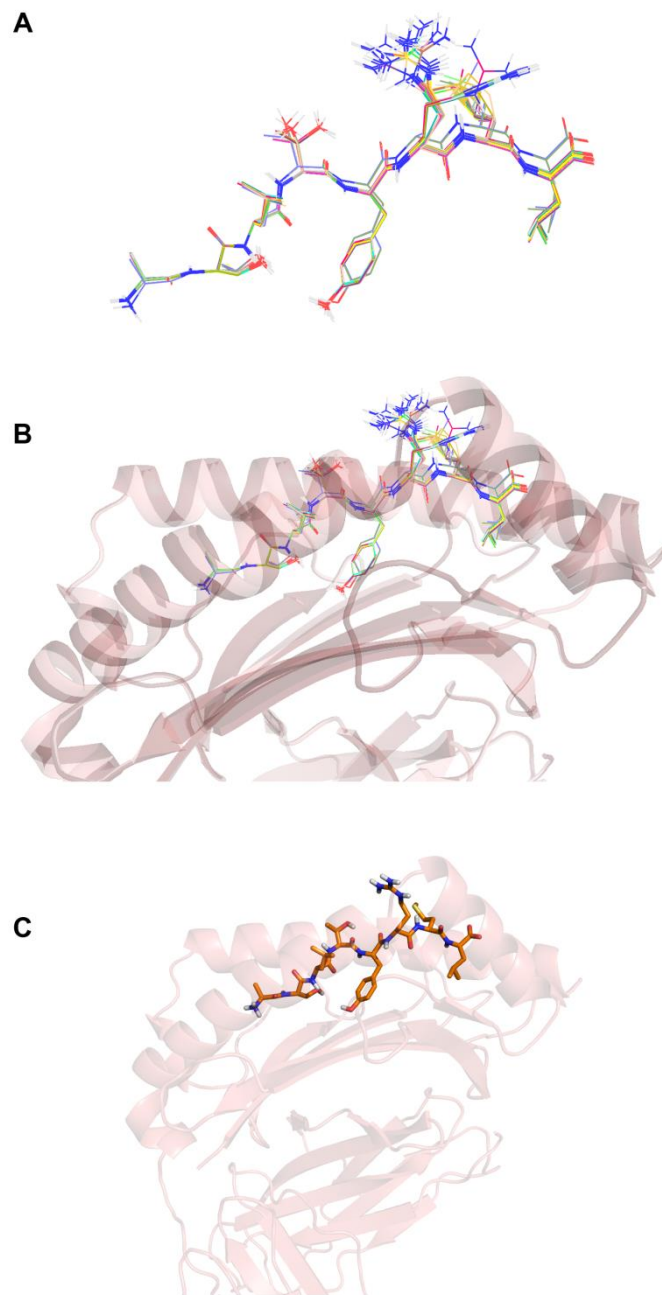


Figura 6. Escolha da melhor conformação do epitopo. (A) Após rodar 20 vezes, o AutoDock Vina ordena as melhores conformações de acordo com a energia de ligação de cada uma, gerando um total de 20 conformações. (B) As 20 conformações são mostradas no contexto do MHC-I, cada uma com pequenas diferenças no modo de interação entre cada par de átomos. (C) A melhor conformação é escolhida através de um cálculo baseado na energia de ligação e a conformação média do grupo mais representativo.

Analisando-se os dados de validação com mais cautela, nos chamou a atenção que um dos valores de RMSD ($C\alpha$) de uma estrutura modelada de HLA-A*02:01 encontrava-se com um valor bastante superior (3,129 Å) às outras estruturas modeladas do mesmo grupo. O epitopo referente a essa estrutura (LAGIGILTV), na verdade, é oriundo da proteína MART-1/Melan-A (Kawakami *et al.*, 1994). Esse epitopo representa uma variante de um epitopo decâmico (ELAGIGILTV), o qual é reconhecido por células T reativas contra MART-1 (Borbulevych *et al.*, 2007). Esse epitopo decâmico adota uma posição dentro da fenda do MHC-I de forma que a porção central se mantém mais voltada para a área externa (Figura 5 do Capítulo III). O epitopo nonâmico segue o mesmo padrão de ligação na fenda do MHC-I, onde a posição 1 se mantém como resíduo de ancoragem. Esse padrão não é observado normalmente em epitopos nonânicos ligados ao HLA-A*02:01 e, por essa razão, não pode ser reproduzido com exatidão pelo DockTope. Portanto, epitopos com conformações aberrantes ou não usuais podem não ser modelados com elevada acurácia.

Os modelos gerados pelo DockTope podem ser utilizados em abordagens variadas onde a simples análise da sequência não seria suficiente para responder questões biológicas complexas. É o caso, por exemplo, do estudo da reatividade cruzada entre complexos pMHC-I. Como mencionado anteriormente, a reatividade cruzada é um fenômeno onde dois complexos pMHC-I podem ser reconhecidos por um mesmo TCR. Isso acontece devido à semelhança estrutural entre os dois pMHC-I na área de interação com o TCR. Tais semelhanças, entretanto, não dependem necessariamente apenas da similaridade entre sequências ou da similaridade físico-química entre os aminoácidos do epitopo, mas sim, de variáveis que podem ser extraídas a partir de dados estruturais do complexo pMHC-I, como topografia e potencial eletrostático. Em um estudo recente nosso grupo avaliou 55 complexos pMHC-I incluindo epitopos de diferentes proteínas virais apresentados no contexto do mesmo MHC-I (Antunes *et al.*, 2011). Através de inferências estruturais sobre os complexos pMHC-I modelados, foi possível encontrar indícios de reatividade cruzada entre o alvo HCV-NS3₁₀₇₃ e os alvos de proteínas não correlatas EBV-LMP2₃₂₉, HIV-Gag₇₇ e IV-NA₂₃₁. Esses dados foram confirmados *in vitro* recentemente (Zhang *et al.*, 2015) demonstrando a validade

da técnica de modelagem utilizada pelo DockTope. Interessantemente, a sequência linear dos aminoácidos do EBV-LMP2₃₂₉ não apresenta qualquer similaridade de sequência com o HCV-NS3₁₀₇₃ e apenas 33% de similaridade físico-química. Entretanto, observando-se a superfície de interação com o TCR observamos uma incrível semelhança nos padrões estruturais apresentada por estes complexos (Figura 6 do Capítulo III), mais uma vez demonstrando a importância de se conhecer a estrutura tridimensional do complexo pMHC-I para realizar inferências de reatividade cruzada.

Outra área de interesse onde os modelos gerados a partir do DockTope podem ser utilizados é relativa à avaliação da estabilidade de complexos pMHC-I. O estudo da estabilidade dos complexos pMHC-I começou a ser realizado já na década de 90, quando van der Burg e cols. lançaram a hipótese de que a imunogenicidade de um determinado complexo pMHC-I dependeria da sua estabilidade na superfície celular, e não apenas da afinidade do peptídeo pelo MHC-I (van der Burg *et al.*, 1996). No artigo, os autores ainda discutem o fato de um peptídeo decâmero (ADLMGYIPLV) oriundo de uma proteína de nucleocapsídeo de HCV (HCV-core₁₃₁) apresentar menor afinidade pelo MHC-I, mas maior imunogenicidade, se comparado a um peptídeo nonâmero relacionado (DLMGYIPLV), oriundo da mesma proteína (HCV-core₁₃₂). Os autores atribuem isso ao fato de o HCV-core₁₃₁ apresentar maior estabilidade. Corroborando esses dados, no mesmo artigo, foi demonstrado que um grupo substancial de peptídeos com afinidade intermediária ou baixa pelo MHC-I, mas com valores de estabilidade adequados, figuram entre os peptídeos imunogênicos testados. Em consonância com esses dados, já a partir de 2012, vários trabalhos apontando para a importância da estabilidade de complexos pMHC-I começaram a surgir (Fridman *et al.*, 2012; Harndahl *et al.*, 2012; Hoppes *et al.*, 2014; Jorgensen *et al.*, 2013; Narzi *et al.*, 2012; Trolle & Nielsen, 2014; Wu *et al.*, 2012; Yanaka *et al.*, 2014). Fridman e cols. analisaram mais de 225 peptídeos imunogênicos provenientes de três proteínas tumorais diferentes (antígeno carcinoembrionário, receptor do fator de crescimento epidérmico humano 2 e a transcriptase reversa de telomerase). O estudo desses epitopos mostrou que para realizar uma predição mais acurada da imunogenicidade, deve-se levar em consideração a

estabilidade do pMHC-I formado, e não apenas a afinidade do peptídeo. Ainda, algoritmos de predição que se baseiam apenas na afinidade, podem ter uma taxa significativa de falsos negativos (em relação à imunogenicidade), pois eles negligenciam o fato de que complexos com uma afinidade baixa podem ser estáveis (Fridman *et al.*, 2012).

Encontramos na literatura apenas um programa voltado para a predição de estabilidade de complexos pMHC-I até o momento, o NetMHCstab (Jorgensen *et al.*, 2013). Esse programa foi lançado em 2013 e se baseia em um conjunto de dados de estabilidade advindos de uma técnica denominada Ensaio de Cintilação por Proximidade (ECP), dados estes que são fornecidos como entrada para o treinamento e teste de redes neurais artificiais. Entretanto, o NetMHCstab possui um viés. Como os próprios autores mencionam no artigo, para que o ECP possa ser realizado, é necessário que se utilize ligantes com alta afinidade pelo MHC-I. Embora não seja a regra, conforme exposto acima, existe uma relação direta entre afinidade e estabilidade. Sendo assim, haverá um viés no sentido de que o ensaio retornará mais dados positivos (complexos estáveis) do que negativos (complexos instáveis). Isso é um problema quando falamos em redes neurais, pois para realizar o treinamento das mesmas são necessários dados positivos e negativos. Como solução, os autores utilizaram como dados negativos as informações geradas a partir de um preditor de afinidade de ligação do epitopo ao MHC-I. Com isso, entretanto, o índice de falsos-negativos aumenta, aumentando a chance de erro do programa.

Assim, no âmbito da imunoinformática, percebemos a necessidade de se criar uma ferramenta que pudesse prever a estabilidade de ligantes de MHC-I, onde pudéssemos avaliar também as modificações estruturais envolvidas no processo de estabilização. Para isso, começamos a elaborar um ensaio *in silico* baseado em dinâmica molecular. Assim, a partir da análise da trajetória da molécula ao longo do tempo, poderíamos procurar por variáveis que pudessem ser utilizadas em uma ferramenta combinada. Entre essas variáveis, algumas se destacam, como o Desvio Quadrático Médio (RMSD), a Flutuação Quadrática Média (RMSF), a análise de distância entre o epitopo e a fenda do MHC-I e a

análise das interações intermoleculares. Utilizando um protocolo de restrição dos carbonos alfa da cadeia principal do MHC-I, mantendo as cadeias laterais dos aminoácidos flexíveis, conseguimos observar diferenças ao simular um peptídeo imunogênico e outro com as posições de ancoragem desfavorecidas. Observamos grande variação especialmente na extremidade amino-terminal do epitopo, especificamente nos resíduos 1, 2 e 3 do epitopo. Esses dados vão ao encontro de dados de predição de estabilidade já publicados (Harndahl *et al.*, 2012).

Outro achado interessante utilizando o nosso método foi com relação ao conjunto de interações realizadas entre os diferentes peptídeos testados e o MHC-I. Como já discutido em outros trabalhos, a rede de interações através de ligações de hidrogênio entre epitopo e MHC-I é essencial para a manutenção da estabilidade do complexo pMHC-I (Narzi *et al.*, 2012). Fazendo uma ampla busca no PDB, percebemos que alguns resíduos estão sempre envolvidos em uma ou mais ligações de hidrogênio com o epitopo. Com relação ao HLA-A*02:01, dois resíduos essenciais estão localizados nas posições 7 (Tyr7) e 99 (Tyr99) da fenda do MHC-I. Em nossas análises de dinâmica molecular, percebemos que ambos resíduos diminuem significativamente a quantidade de interações ao longo do tempo com o epitopo que teve as posições de ancoragem modificadas. Esses dois resíduos do MHC-I estão presentes no *pocket B*, responsável por interações com o aminoácido da 2^a posição do epitopo (Sette & Sidney, 1998), o que explicaria a maior variação observada na extremidade amino-terminal.

Embora não tenha sido descrito em pormenores nesta tese, também começamos a avaliar computacionalmente o possível papel de glicanas na estabilização de complexos pMHC-I. Dados publicados mostram que a glicosilação ocorre em sítios altamente conservados (Ryan & Cobb, 2012a) e, a priori, não interferem na interação entre pMHC-I e TCR (Parham, 1996). No caso do supertipo HLA-A2, a N-glicosilação ocorre em um resíduo de Asparagina, na posição 86 (Ryan & Cobb, 2012b). Estudos iniciais realizados em nosso laboratório mostram que a glicana associada ao pMHC-I exibe um comportamento

em favor da associação com a β 2-microglobulina, o que poderia estar diretamente relacionado com a estabilização do complexo pMHC-I (Figura 7).

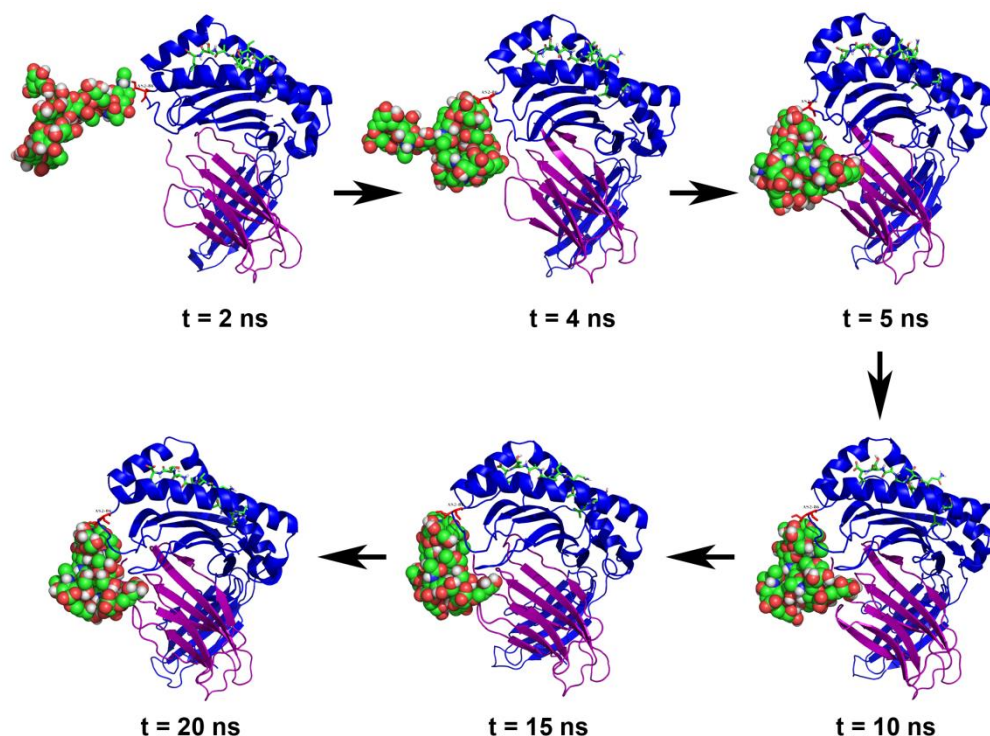


Figura 7. Estudo *in silico* do de pMHC-I ligado a uma glicana. O alótipo representado é o HLA-A*02:01 ligado ao peptídeo SLYNTVATLL (PDB ID: 2V2W), contendo uma N-glicosilação no resíduo de asparagina 86. A dinâmica molecular de 20 ns evidencia o comportamento da glicana, a qual parece se estabilizar a partir dos 5 ns ao lado do domínio β 2-microglobulina.

Com o presente trabalho, finalizamos a automatização e otimização de uma ferramenta para construção de complexos pMHC-I *in silico*. Esta ferramenta está disponível online (dirac.cesup.ufrgs.br/bio/home.php) para que qualquer usuário possa construir complexos pMHC-I de maneira personalizada, fácil e relativamente rápida. Com as modelagens advindas dessa ferramenta, também criamos um método capaz de identificar ligantes com estabilidade diferenciada, o qual poderá ser utilizado em combinação com outras ferramentas para realizar a predição de imunogenicidade.

Referências Complementares (Capítulos I e V)

- Abbas AK and Lichtman AH (2004) Cellular and Molecular Immunology. 5 edition. Elsevier Science Limited, San Francisco, 580 pp.
- Antunes DA, Rigo MM, Silva JP, Cibulski SP, Sinigaglia M, Chies JA and Vieira GF (2011) Structural in silico analysis of cross-genotype-reactivity among naturally occurring HCV NS3-1073-variants in the context of HLA-A*02:01 allele. *Molecular Immunology* 48:1461-1467.
- Antunes DA, Vieira GF, Rigo MM, Cibulski SP, Sinigaglia M and Chies JA (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.
- Bagaria A, Jaravine V, Huang YJ, Montelione GT and Guntert P (2011) Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci* 21:229-238.
- Baker BM, Scott DR, Blevins SJ and Hawse WF (2012) Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism. *Immunol Rev* 250:10-31.
- Berko D, Carmi Y, Cafri G, Ben-Zaken S, Sheikhet HM, Tzehoval E, Eisenbach L, Margalit A and Gross G (2005) Membrane-anchored beta 2-microglobulin stabilizes a highly receptive state of MHC class I molecules. *J Immunol* 174:2116-2123.
- Bhasin M and Raghava GP (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 13:596-607.
- Boehm T, Iwanami N and Hess I (2012) Evolution of the immune system in the lower vertebrates. *Annu Rev Genomics Hum Genet* 13:127-149.
- Borbulevych OY, Insaído FK, Baxter TK, Powell DJ, Jr., Johnson LA, Restifo NP and Baker BM (2007) Structures of MART-126/27-35 Peptide/HLA-A2

complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J Mol Biol* 372:1123-1136.

Bordner AJ and Abagyan R (2006) Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* 63:512-526.

Botos I, Segal DM and Davies DR (2011) The structural biology of Toll-like receptors. *Structure* 19:447-459.

Choi AG, Wong J, Marchant D and Luo H (2012) The ubiquitin-proteasome system in positive-strand RNA virus infection. *Rev Med Virol* 23:85-96.

Crews CM (2003) Feeding the machine: mechanisms of proteasome-catalyzed degradation of ubiquitinated proteins. *Curr Opin Chem Biol* 7:534-539.

Dausset J (1958) [Iso-leuko-antibodies]. *Acta Haematol* 20:156-166.

Davies MN and Flower DR (2007) Harnessing bioinformatics to discover new vaccines. *Drug Discov Today* 12:389-395.

Donati C and Rappuoli R (2013) Reverse vaccinology in the 21st century: improvements over the original design. *Ann N Y Acad Sci* 1285:115-132.

Driscoll J, Brown MG, Finley D and Monaco JJ (1993) MHC-linked LMP gene products specifically alter peptidase activities of the proteasome. *Nature* 365:262-264.

Ekeruche-Makinde J, Miles JJ, van den Berg HA, Skowera A, Cole DK, Dolton G, Schauenburg AJ, Tan MP, Pentier JM, Llewellyn-Lacey S et al. (2012) Peptide length determines the outcome of TCR/peptide-MHCI engagement. *Blood* 121:1112-1123.

Flajnik MF and Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* 11:47-59.

Fridman A, Finnefrock AC, Peruzzi D, Pak I, La Monica N, Bagchi A, Casimiro DR, Ciliberto G and Aurisicchio L (2012) An efficient T-cell epitope discovery

strategy using in silico prediction and the iTopia assay platform. *Oncoimmunology* 1:1258-1270.

Gaide O, Emerson RO, Jiang X, Gulati N, Nizza S, Desmarais C, Robins H, Krueger JG, Clark RA and Kupper TS (2015) Common clonal origin of central and resident memory T cells following skin immunization. *Nat Med*.

Harndahl M, Rasmussen M, Roder G, Dalgaard Pedersen I, Sorensen M, Nielsen M and Buus S (2012) Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur J Immunol* 42:1405-1416.

He L, De Groot AS, Gutierrez AH, Martin WD, Moise L and Bailey-Kellogg C (2014) Integrated assessment of predicted MHC binding and cross-conservation with self reveals patterns of viral camouflage. *BMC Bioinformatics* 15 Suppl 4:S1.

Hoppes R, Oostvogels R, Luimstra JJ, Wals K, Toebes M, Bies L, Ekkebus R, Rijal P, Celie PH, Huang JH et al. (2014) Altered peptide ligands revisited: vaccine design through chemically modified HLA-A2-restricted T cell epitopes. *J Immunol* 193:4803-4813.

Ikekawa A and Ikekawa S (2001) Fruits of human genome project and private venture, and their impact on life science. *Yakugaku Zasshi* 121:845-873.

Iwasaki A and Medzhitov R (2010) Regulation of adaptive immunity by the innate immune system. *Science* 327:291-295.

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.

Jorgensen KW, Rasmussen M, Buus S and Nielsen M (2013) NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* 141:18-26.

Kasahara M (1999) Genome dynamics of the major histocompatibility complex: insights from genome paralogy. *Immunogenetics* 50:134-145.

- Kawakami Y, Eliyahu S, Sakaguchi K, Robbins PF, Rivoltini L, Yannelli JR, Appella E and Rosenberg SA (1994) Identification of the immunodominant peptides of the MART-1 human melanoma antigen recognized by the majority of HLA-A2-restricted tumor infiltrating lymphocytes. *J Exp Med* 180:347-352.
- Khalili S, Jahangiri A, Borna H, Ahmadi Zanoos K and Amani J (2014) Computational vaccinology and epitope vaccine design by immunoinformatics. *Acta Microbiol Immunol Hung* 61:285-307.
- Khan JM and Ranganathan S (2010) pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Research* 6 Suppl 1:S2.
- Khan MA, Hossain MU, Rakib-Uz-Zaman SM and Morshed MN (2015) Epitope - based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *Scand J Immunol*.
- Kurosaki T, Kometani K and Ise W (2015) Memory B cells. *Nat Rev Immunol* 15:149-159.
- Lecker SH, Goldberg AL and Mitch WE (2006) Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J Am Soc Nephrol* 17:1807-1819.
- Lefranc MP (2014) Immunoglobulin and T Cell Receptor Genes: IMGT((R)) and the Birth and Rise of Immunoinformatics. *Front Immunol* 5:22.
- Levitt M and Warshel A (1975) Computer simulation of protein folding. *Nature* 253:694-698.
- 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.

- Madurga S, Belda I, Llorca X and Giralt E (2005) Design of enhanced agonists through the use of a new virtual screening method: application to peptides that bind class I major histocompatibility complex (MHC) molecules. *Protein Sci* 14:2069-2079.
- McCarthy MK and Weinberg JB (2015) The immunoproteasome and viral infection: a complex regulator of inflammation. *Front Microbiol* 6:21.
- Motozono C, Yokoyama M, Sato H and Ueno T (2014) Cross-reactivity analysis of T cell receptors specific for overlapping HIV-1 Nef epitopes of different lengths. *Microbes Infect* 16:320-327.
- Najakshin AM, Mechetina LV, Alabyev BY and Taranin AV (1999) Identification of an IL-8 homolog in lamprey (*Lampetra fluviatilis*): early evolutionary divergence of chemokines. *Eur J Immunol* 29:375-382.
- Narzi D, Becker CM, Fiorillo MT, Uchanska-Ziegler B, Ziegler A and Bockmann RA (2012) Dynamical characterization of two differentially disease associated MHC class I proteins in complex with viral and self-peptides. *J Mol Biol* 415:429-442.
- Nathan JA, Spinnenhirn V, Schmidtke G, Basler M, Groettrup M and Goldberg AL (2013) Immuno- and constitutive proteasomes do not differ in their abilities to degrade ubiquitinated proteins. *Cell* 152:1184-1194.
- 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.
- Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL and Cooper MD (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430:174-180.
- Parcej D and Tampe R (2010) ABC proteins in antigen translocation and viral inhibition. *Nat Chem Biol* 6:572-580.

- Parham P (1996) Functions for MHC class I carbohydrates inside and outside the cell. *Trends Biochem Sci* 21:427-433.
- Parker KC, Bednarek MA and Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* 152:163-175.
- Pickart CM and Fushman D (2004) Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol* 8:610-616.
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA and Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213-219.
- Ramos M, Paradela A, Vazquez M, Marina A, Vazquez J and Lopez de Castro JA (2002) Differential association of HLA-B*2705 and B*2709 to ankylosing spondylitis correlates with limited peptide subsets but not with altered cell surface stability. *J Biol Chem* 277:28749-28756.
- Rappuoli R (2000) Reverse vaccinology. *Curr Opin Microbiol* 3:445-450.
- Rock KL and Goldberg AL (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu Rev Immunol* 17:739-779.
- Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI and McCluskey J (2015) T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 33:169-200.
- Royet J, Gupta D and Dziarski R (2011) Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat Rev Immunol* 11:837-851.
- Ryan SO and Cobb BA (2012a) Host glycans and antigen presentation. *Microbes Infect* 14:894-903.
- Ryan SO and Cobb BA (2012b) Roles for major histocompatibility complex glycosylation in immune function. *Semin Immunopathol* 34:425-441.

- Saveanu L, Fruci D and van Ender P (2002) Beyond the proteasome: trimming, degradation and generation of MHC class I ligands by auxiliary proteases. *Mol Immunol* 39:203-215.
- Schubert B, Brachvogel HP, Jurges C and Kohlbacher O (2015) EpiToolKit-a web-based workbench for vaccine design. *Bioinformatics*.
- Schulenburg H, Kurz CL and Ewbank JJ (2004) Evolution of the innate immune system: the worm perspective. *Immunol Rev* 198:36-58.
- Sette A and Sidney J (1998) HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol* 10:478-482.
- Sidney J, Grey HM, Kubo RT and Sette A (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. *Immunol Today* 17:261-266.
- Sidney J, Peters B, Frahm N, Brander C and Sette A (2008) HLA class I supertypes: a revised and updated classification. *BMC Immunology* 9:1.
- Sijts EJ and Kloetzel PM (2011) The role of the proteasome in the generation of MHC class I ligands and immune responses. *Cell Mol Life Sci* 68:1491-1502.
- Sinigaglia M, Antunes DA, Rigo MM, Chies JA and Vieira GF (2013) CrossTape: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database (Oxford)* 2013:bat002.
- Snell GD (1986) Some recollections of Peter Gorer and his work on this fiftieth anniversary of his discovery of H-2. *Immunogenetics* 24:339-340.
- Ting JP, Duncan JA and Lei Y (2010) How the noninflammasome NLRs function in the innate immune system. *Science* 327:286-290.
- Todman SJ, Halling-Brown MD, Davies MN, Flower DR, Kayikci M and Moss DS (2008) Toward the atomistic simulation of T cell epitopes automated construction of MHC: peptide structures for free energy calculations. *J Mol Graph Model* 26:957-961.

- Tomar N and De RK (2010) Immunoinformatics: an integrated scenario. *Immunology* 131:153-168.
- Trolle T and Nielsen M (2014) NetTepi: an integrated method for the prediction of T cell epitopes. *Immunogenetics* 66:449-456.
- Trott O and Olson AJ (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:455-461.
- Tsutsui S, Nakamura O and Watanabe T (2007) Lamprey (*Lethenteron japonicum*) IL-17 upregulated by LPS-stimulation in the skin cells. *Immunogenetics* 59:873-882.
- van de Weijer ML, Luteijn RD and Wiertz EJ (2015) Viral immune evasion: Lessons in MHC class I antigen presentation. *Semin Immunol*.
- van der Burg SH, Visseren MJ, Brandt RM, Kast WM and Melief CJ (1996) Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J Immunol* 156:3308-3314.
- Vandiedonck C and Knight JC (2009) The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief Funct Genomic Proteomic* 8:379-394.
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A et al. (2014) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 43:D405-412.
- Voges D, Zwickl P and Baumeister W (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* 68:1015-1068.
- Welsh RM, Che JW, Brehm MA and Selin LK (2010) Heterologous immunity between viruses. *Immunol Rev* 235:244-266.

- Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, Dolton G, Clement M, Llewellyn-Lacey S, Price DA et al. (2011) A single autoimmune T cell receptor recognizes more than a million different peptides. *J Biol Chem* 287:1168-1177.
- Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, Dolton G, Clement M, Llewellyn-Lacey S, Price DA et al. (2012) A single autoimmune T cell receptor recognizes more than a million different peptides. *J Biol Chem* 287:1168-1177.
- Wu X, Xu X, Gu R, Wang Z, Chen H, Xu K, Zhang M, Hutton J and Yang T (2012) Prediction of HLA class I-restricted T-cell epitopes of islet autoantigen combined with binding and dissociation assays. *Autoimmunity* 45:176-185.
- Wucherpfennig KW, Allen PM, Celada F, Cohen IR, De Boer R, Garcia KC, Goldstein B, Greenspan R, Hafler D, Hodgkin P et al. (2007) Polyspecificity of T cell and B cell receptor recognition. *Semin Immunol* 19:216-224.
- Yanaka S, Ueno T, Shi Y, Qi J, Gao GF, Tsumoto K and Sugase K (2014) Peptide-dependent conformational fluctuation determines the stability of the human leukocyte antigen class I complex. *J Biol Chem* 289:24680-24690.
- Yates AJ (2014) Theories and quantification of thymic selection. *Front Immunol* 5:13.
- Zhang S, Bakshi RK, Suneetha PV, Fytilli P, Antunes DA, Vieira GF, Jacobs R, Klade CS, Manns MP, Kraft AR et al. (2015) Frequency, private specificity and cross-reactivity of pre-existing HCV-specific CD8+ T cells in HCV seronegative individuals: implication for vaccine responses. *J Virol*.
- Zuniga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, Acuna-Alonzo V, Yunis M, Granados-Montiel J, Cruz-Lagunas A et al. (2013) HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations. *PLoS One* 8:e74442.

Anexos

Durante o desenvolvimento desta tese, participei de outros trabalhos na área acadêmico-científica, alguns dos quais publicados e outros em vias de publicação. Estes trabalhos estão citados nesta seção uma vez que, de alguma forma, influenciaram no desenvolvimento desta tese. Os dois primeiros trabalhos estão incluídos na íntegra. O primeiro deles, relacionado ao banco de dados CrossTope, possui uma ligação intrínseca com o DockTope, uma vez que essa ferramenta permitirá o desenvolvimento e inclusão de novos complexos. O segundo trabalho, envolvendo o estudo de ocitocinas com diferentes mutações, foi importante para a melhor aprendizagem das ferramentas de dinâmica molecular e incluiu a criação de um protocolo único para o estudo dessas moléculas.

CrossTope: a curate repository of 3D structures of immunogenic peptide:MHC complexes

Sinigaglia M, Antunes DA, Rigo MM, Chies JAB, Vieira GF

O CrossTope é um banco de dados de estruturas pMHC-I tridimensionais, curado manualmente, o qual contém estruturas modeladas através da técnica *D1-EM-D2* ou obtidas do PDB. Além das estruturas serem disponibilizadas para download no formato *pdb*, o usuário encontra informações adicionais sobre o complexo, como a proteína de origem daquele epitopo, data de deposição do complexo, patógeno de origem, entre outras informações. Também é possível fazer o download da imagem que representa a área de interação do pMHC-I com o TCR, já com o potencial eletrostático calculado. Até então, as estruturas eram modeladas de forma não automatizada, o que exigia tempo, até porque as estruturas precisavam ser cuidadosamente checadas para que não contivessem erros. Com o DockTope, é possível modelar complexos de forma direta, automatizada e com menor chance de erro. Não apenas minimizando a interação entre servidor e usuário, mas permitindo que um número maior de complexos possam ser montados para compor o banco de dados do CrossTope.

Original article

CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes

Marialva Sinigaglia, Dinler Amaral Antunes, Maurício Menegatti Rigo, José Artur Bogo Chies and Gustavo Fioravanti Vieira*

NBLI - Núcleo de Bioinformática do Laboratório de Imunogenética, Department of Genetics, Universidade Federal do Rio Grande do Sul, 9500 Bento Gonçalves Avenue, Bldg 43323, Rm 225, 91501-970 Porto Alegre, RS, Brazil

*Corresponding author: Email: fioravanti.vieira@ufrgs.br or gusfioravanti@yahoo.com.br

Citation details: Sinigaglia, M., Antunes, D.A., Rigo, M.M. et al. CrossTope: a curate repository of three-dimensional structures of immunogenic peptide: MHC complexes. *Database* (2013) Vol. 2013: article id bat002; doi: 10.1093/database/bat002

Submitted 18 July 2012; Revised 26 November 2012; Accepted 8 January 2013

The CrossTope is a highly curate repository of three-dimensional structures of peptide:major histocompatibility complex (MHC) class I complexes (pMHC-I). The complexes hosted by this databank were obtained in protein databases and by large-scale *in silico* construction of pMHC-I structures, using a new approach developed by our group. At this moment, the database contains 182 'non-redundant' pMHC-I complexes from two human and two murine alleles. A web server provides interface for database query. The user can download (i) structure coordinate files and (ii) topological and charges distribution maps images from the T-cell receptor-interacting surface of pMHC-I complexes. The retrieved structures and maps can be used to cluster similar epitopes in cross-reactivity approaches, to analyse viral escape mutations in a structural level or even to improve the immunogenicity of tumour antigens.

Database URL: <http://www.crosstope.com.br>

Introduction

A great challenge in immunology is the discovery of new targets to the development of vaccines against even common infectious diseases. The choice of these targets depends on the identification of elements responsible for the stimulation of immune responses. Evidence for the sharing of these elements was first provided by cross-reactivity studies (1–4). Cross-reactivity is defined by the ability of a given T-cell population to recognize different peptide:major histocompatibility complex (MHC) class I complexes (pMHC-I). This phenomenon plays an important role in antiviral cellular immunity, and there are several described cases of heterologous immunity involving completely unrelated viruses. Additionally, it has also been suggested that knowledge about the molecular features

responsible for cross-reactivity can be applied to the development of a new generation of wide-spectrum viral vaccines (5).

Classical studies use epitope sequence data for selection of targets to be subsequently used in *in vivo* experiments. However, these data lack information about elements that are crucial for the stimulation of an appropriate immunity, such as topology and charges distribution of the molecule. These targets (peptides) are presented to the immune system in the context of MHC-I molecules, and essential structural information can be obtained from crystals of these pMHC-I complexes. These structures are determined mainly by X-ray diffraction and nuclear magnetic resonance, providing a general view of the pMHC-I surface that interacts with the T-cell receptor (TCR). Such complexes can be used for comparison between different epitopes

Evolutionary pattern in the OXT-OXTR system in primates: Coevolution and positive selection footprints

Pedro Vargas-Pinilla P, Paixão-Côrtés VR, Paré P, Tovo-Rodrigues L, Vieira CMAG, Xavier A, Comas D, Pissinatti A, Sinigaglia M, Rigo MM, Vieira GF, Lucion AB, Salzano FM, Bortolini MC

O hormônio ocitocina (OXT) é possui 9 aminoácidos e está envolvido em uma ampla gama de funções, tanto fisiológicas como de comportamento. Neste trabalho, os autores reportaram três novas formas desse hormônio em macacos do Novo Mundo (OXT-3Val-8Pro, OXT-8Ala, OXT-8Thr), bem como uma distribuição mais abrangente de uma variante previamente descrita (OXT-8Pro). Para um melhor entendimento do comportamento dessas variantes, realizamos um estudo estrutural *in silico*, partindo desde a modelagem de cada variante até a simulação por dinâmica molecular. A simulação por dinâmica molecular envolveu o desenvolvimento de um protocolo único para o estudo dessas moléculas. A partir dos dados gerados computacionalmente, conseguimos inferir um comportamento padrão entre todas as moléculas estudadas, o que corroborou estudos anteriores.

Evolutionary pattern in the OXT-OXTR system in primates: Coevolution and positive selection footprints

Pedro Vargas-Pinilla^{a,1}, Vanessa Rodrigues Paixão-Côrtes^{a,1}, Pamela Paré^a, Luciana Tovo-Rodrigues^a, Carlos Meton de Alencar Gadelha Vieira^a, Agatha Xavier^a, David Comas^b, Alcides Pissinatti^c, Marialva Sinigaglia^a, Maurício Menegatti Rigo^a, Gustavo Fioravanti Vieira^a, Aldo B. Lucion^d, Francisco Mauro Salzano^{a,2}, and Maria Cátira Bortolini^{a,2}

^aDepartamento de Genética, Instituto de Biotecnologia, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil; ^bInstitut de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain; ^cCentro de Primatologia do Rio de Janeiro, 20940-200 Rio de Janeiro, RJ, Brazil; and ^dDepartamento de Fisiologia, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, 90050-170 Porto Alegre, RS, Brazil

Contributed by Francisco Mauro Salzano, November 26, 2014 (sent for review July 11, 2014; reviewed by Guido Barbujani and Rafal Slusarz)

Oxytocin is a nonapeptide involved in a wide range of physiologic and behavioral functions. Until recently, it was believed that an unmodified oxytocin sequence was present in all placental mammals. This study analyzed oxytocin (OXT) in 29 primate species and the oxytocin receptor (OXTR) in 21 of these species. We report here three novel OXT forms in the New World monkeys, as well as a more extensive distribution of a previously described variant (Leu8Pro). In structural terms, these OXTs share the same three low-energy conformations in solution during molecular dynamic simulations, with subtle differences in their side chains. A consistent signal of positive selection was detected in the Cebidae family, and OXT position 8 showed a statistically significant ($P = 0.013$) correlation with litter size. Several OXTR changes were identified, some of them promoting gain or loss of putative phosphorylation sites, with possible consequences for receptor internalization and desensitization. OXTR amino acid sites are under positive selection, and intramolecular and intermolecular coevolutionary processes with OXT were also detected. We suggest that some New World monkey OXT-OXTR forms can be correlated to male parental care through the increase of cross-reactivity with its correlated vasopressin system.

OXT | OXTR | primates | coevolution | behavior

Oxytocin has crucial functions related to physiological processes and social behaviors in primates and other placental mammals. A nonapeptide (Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly) (1), oxytocin (OXT-8Leu) is both a neurotransmitter released by neuronal cells in synapses and a hormone, activating receptors distant from the site of its synthesis through the circulatory system (2). In mammals, OXT acts as a hormone in uterine contraction during parturition and in milk ejection while lactating. It is also a key central nervous system neurotransmitter, regulating/modulating complex social and reproductive behaviors (i.e., pair bonding and parental care) (3–7).

Until recently, it was believed that the OXT amino acid chain was the same in all placental mammals. However, Lee and colleagues (8) reported a T > C change in four New World monkeys (NWms), *Saimiri sciureus*, *Cebus apella*, *Callithrix jacchus*, and *Aotus nancimae*, substituting leucine to proline at position 8 (OXT-8Pro). This form was also found in *Tupaia belangeri*, a tree shrew species of Southeast Asia (8). OXT differs from its paralog vasopressin (AVP; Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly) at positions 3 and 8. Variation at position 8 also identifies nonplacental OXT/AVP-like nonapeptides, such as mesotocin, present in some marsupials (7, 9). These findings dispel the notion of a universal OXT amino acid sequence for placental mammals. They also suggest that residue variability at position 8, in some cases associated with variations at positions 2–5, may be connected with the recognition, binding, and activation of receptors, potentially leading to species-specific functional changes (7, 10).

OXT activity depends on adequate interaction with its unique receptor, OXTR, although it can also bind to the vasopressin receptors (AVPR1a, AVPR1b, and AVPR2) with lower affinity (11–13). Similar to other receptors that use G proteins as transducer signals across the cell membranes, OXTR is composed of seven transmembrane (TM1–TM7), four extracellular (N-terminal tail-ECL3), and four intracellular (ICL1-C-terminal tail) domains. ECL and ICL are important for the interaction with OXT and G proteins, respectively, whereas TMs are connected with both functions (7, 11).

In contrast to what is observed for placental mammal OXT, OXTR presents hundreds of variants in regulatory and coding regions, including at the intraspecific level. In humans, OXTR single-nucleotide polymorphisms have been associated with several social behavioral phenotypes (14).

The presence of OXT-OXTR-related systems throughout the animal kingdom indicates that their typical roles in placental mammals are likely exaptations of ancient functions, such as regulation of fluid balance and egg-laying (15, 16). Studies have attempted to investigate both the interaction of OXT-OXTR-like systems and their coevolution (11, 17). However, our knowledge about this nonapeptide-receptor system, including the extent of its variability in the primate order, remains limited.

Significance

It was previously believed that placental mammals present no variability in oxytocin (OXT). The present study reports novel data on the diversity of OXT and its receptor (OXTR) in primate species, including New World monkeys. Contrary to prior expectations, we found three novel OXT forms and several OXTR nonsynonymous changes not previously described. In the Cebidae family, signals of positive selection were found for an OXT variant at position 8, which is associated with larger litter sizes. We detected positive selection for OXTR forms and report a coevolutionary process between changes in OXT and OXTR.

Author contributions: P.V.-P., V.R.P.-C., C.M.d.A.G.V., F.M.S., and M.C.B. designed research; P.V.-P., V.R.P.-C., C.M.d.A.G.V., and A.X. performed research; A.P. contributed new reagents/analytic tools; P.V.-P., V.R.P.-C., P.P., L.T.-R., M.S., M.M.R., G.F.V., and M.C.B. analyzed data; A.P. provided information about the animals investigated; and P.V.-P., V.R.P.-C., P.P., D.C., A.B.L., F.M.S., and M.C.B. wrote the paper.

Reviewers included: G.B., Università di Ferrara; and R.S., University of Gdansk.

The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession nos. [KM186262](#) to [KM186289](#)).

¹P.V.-P. and V.R.P.-C. contributed equally to this work.

²To whom correspondence may be addressed. Email: francisco.salzano@ufrgs.br or maria.bortolini@ufrgs.br.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1419399112/-DCSupplemental.

New Insights into the *In Silico* Prediction of HIV Protease Resistance to Nelfinavir

Antunes, DA, Rigo MM, Sinigaglia M, de Medeiros R, Junqueira DM, Almeida SEM, Vieira GF

A protease do Vírus da Imunodeficiência Humana tipo 1 (HIV-1) é um alvo comumente utilizado para o desenvolvimento de fármacos, devido ao seu importante papel na manutenção da viabilidade viral. Assim, inibidores de protease são amplamente estudados no âmbito da pesquisa farmacêutica. Um desses alvos é denominado Nelfinavir, o qual possui a capacidade de se ligar ao sítio catalítico da protease, inibindo a sua ação. Entretanto, mutações não-sinônimas, acarretando na expressão de variantes de protease, podem influenciar na ligação desse inibidor e, conseqüentemente, na sua ação. Neste estudo nós utilizamos ferramentas de bioinformática, especialmente a dinâmica molecular, para avaliar o impacto de mutações presentes na protease (D30V e V32E) sobre a interação com o Nelfinavir. Ambas mutações apresentaram características estruturais que, em última análise, indicam uma possível resistência ao Nelfinavir.

New Insights into the *In Silico* Prediction of HIV Protease Resistance to Nelfinavir

Dinler A. Antunes^{1,3}, Maurício M. Rigo^{1,3}, Marialva Sinigaglia^{1,3}, Rúbia M. de Medeiros^{2,3}, Dennis M. Junqueira^{2,3}, Sabrina E. M. Almeida², Gustavo F. Vieira^{1,3*}

1 Núcleo de Bioinformática do Laboratório de Imunogenética (NBLI), Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **2** Technological and Scientific Development Center (CDCT), State Foundation in Production and Health Research (FEPPS), Porto Alegre, Rio Grande do Sul, Brazil, **3** Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

Abstract

The Human Immunodeficiency Virus type 1 protease enzyme (HIV-1 PR) is one of the most important targets of anti-retroviral therapy used in the treatment of AIDS patients. The success of protease-inhibitors (PIs), however, is often limited by the emergence of protease mutations that can confer resistance to a specific drug, or even to multiple PIs. In the present study, we used bioinformatics tools to evaluate the impact of the unusual mutations D30V and V32E over the dynamics of the PR-Nelfinavir complex, considering that codons involved in these mutations were previously related to major drug resistance to Nelfinavir. Both studied mutations presented structural features that indicate resistance to Nelfinavir, each one with a different impact over the interaction with the drug. The D30V mutation triggered a subtle change in the PR structure, which was also observed for the well-known Nelfinavir resistance mutation D30N, while the V32E exchange presented a much more dramatic impact over the PR flap dynamics. Moreover, our *in silico* approach was also able to describe different binding modes of the drug when bound to different proteases, identifying specific features of HIV-1 subtype B and subtype C proteases.

Citation: Antunes DA, Rigo MM, Sinigaglia M, de Medeiros RM, Junqueira DM, et al. (2014) New Insights into the *In Silico* Prediction of HIV Protease Resistance to Nelfinavir. PLoS ONE 9(1): e87520. doi:10.1371/journal.pone.0087520

Editor: Andrea Cavalli, University of Bologna & Italian Institute of Technology, Italy

Received: September 16, 2013; **Accepted:** December 22, 2013; **Published:** January 31, 2014

Copyright: © 2014 Antunes et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fioravanti.vieira@ufrgs.br

Introduction

Human immunodeficiency virus type 1 protease (HIV-1 PR) is a catalytic protein that cleaves the Gag and Gag-Pol viral polyproteins, allowing the virus to efficiently infect new host cells. The HIV-1 PR exists as an aspartyl homodimeric enzyme composed by symmetrical subunits of 99 amino acids each. The access of the substrate to the active site of PR is regulated by two mobile flaps that shift from an open to a closed conformation to bind and cleave the substrate.

The HIV-1 protease is one of the most important targets of antiretroviral therapy used in the treatment of AIDS patients due to its critical role in the viral replication cycle. Protease inhibitors (PI) were developed to inhibit cleavage function of HIV-1 protease by mimicking the reaction intermediates that arises during the hydrolysis of the substrate, disabling the enzyme. The current success of PIs is frequently limited by the emergence of protease gene mutations that confer resistance to this drug class. By changing the structure of the substrate-binding cavity, mutations directly or indirectly interfere with the binding of inhibitors, resulting in viral resistance to PIs.

According to the International AIDS Society, 23 mutations in 16 codons of the protease gene related to major drug-resistance to PIs were identified by phenotypic resistance assays [1]. In addition, it is currently known that polymorphisms in some codons not

previously related to major drug-resistance could affect the viral fitness in the presence of the drug. Previous studies demonstrated that the viability to the arising of resistance mutations is generally dependent on the genetic background. Therefore, the genetic context in which the evolutionary variations arise in the protease gene may affect the efficacy of the treatment.

In this context, codons in the protease gene related to major drug resistance to a specific protease inhibitor can provide clues on the important sites to the interaction between drug and target, and it is possible that unusual changes in these same sites can also affect the interaction with the drug. For instance, D30N mutation causes high-level resistance to Nelfinavir (NF) [1,2] and V32I is associated to reduced susceptibility to all PIs, except Saquinavir [1,3]. However, the effect of the presence of alternative amino acids in these same sites is still unclear.

Due to the elevated costs and the extensive time required for *in vitro* analysis, it is still impractical to use these conventional methods to evaluate the effect of each mutation in view of the genetic background of HIV-1 protease. Thus, computational methods can improve the screening analyzes revealing the role of individual mutations and its impact on the protein function [4–7]. In the present study, we used molecular dynamics and other bioinformatics tools aiming to identify structural features that could indicate the NF-resistance effect of the unusual mutations D30V and V32E, and to evaluate the influence of the HIV-1

Lessons from molecular modeling human α -L-iduronidase

Figueiredo DF, Antunes DA, Rigo MM, Mendes MFA, Silva JP, Mayer FQ, Matte U, Giugliani R, Vieira GF, Sinigaglia M

A enzima α -L-iduronidase humana (IDUA) encontra-se nos lisossomos, onde atua realizando o catabolismo de glicosaminoglicanas (GAGs), heparan sulfato (HS) e dermatan sulfato (DS). Mutações nessa enzima são responsáveis por uma doença denominada Mucopolissacaridose tipo I (MPS I). Apesar do grande interesse em determinar e estudar a estrutura desta enzima, a falta de uma grande identidade de sequência entre modelos estruturais, entre outros detalhes técnicos, tem desafiado bioinformatas e cristalógrafos. Durante o desenvolvimento deste trabalho, entretanto, foi publicada uma estrutura cristalografada de IDUA (PDB: 4JXP), o que nos levou a analisar modelos existentes e comparar com novas técnicas de modelagem. Assim, criamos quatro modelos de IDUA (criados anteriormente à publicação do cristal), utilizando quatro protocolos diferentes, e comparamos com a estrutura 4JXP, buscando a melhor técnica de modelagem. Uma análise combinada com dinâmica molecular apontou para os pontos fortes e fracos dos diferentes protocolos, todos baseados em um modelo com baixa identidade de sequência.



Lessons from molecular modeling human α -L-iduronidase



Danieli Forgiarini Figueiredo^{a,d,1}, Dinler A. Antunes^{a,d,1}, Maurício M. Rigo^{a,d},
 Marcus F.A. Mendes^{a,d}, Jader P. Silva^a, Fabiana Q. Mayer^b, Ursula Matte^{c,d},
 Roberto Giugliani^{c,d}, Gustavo F. Vieira^{a,d}, Marialva Sinigaglia^{a,d,*}

^a NBLI – Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

^b Molecular Biology Laboratory, Instituto de Pesquisas Veterinárias Desidério Finamor, Fundação Estadual de Pesquisa Agropecuária Porto Alegre, Rio Grande do Sul, Brazil

^c Gene Therapy Center, Experimental Research Center, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil

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

ARTICLE INFO

Article history:

Accepted 8 October 2014

Available online 18 October 2014

Keywords:

Homology modeling

Low identity template

α -L-Iduronidase (IDUA)

Model evaluation tools

Molecular dynamics

Secondary structure assessment

ABSTRACT

Human α -L-iduronidase (IDUA) is a member of glycoside hydrolase family and is involved in the catabolism of glycosaminoglycans (GAGs), heparan sulfate (HS) and dermatan sulfate (DS). Mutations in this enzyme are responsible for mucopolysaccharidosis I (MPS I), an inherited lysosomal storage disorder. Despite great interest in determining and studying this enzyme structure, the lack of a high identity to templates and other technical issues have challenged both bioinformaticians and crystallographers, until the recent publication of an IDUA crystal structure (PDB: 4JXP). In the present work, four alternative IDUA models, generated and evaluated prior to crystallographic determination, were compared to the 4JXP structure. A combined analysis using several viability assessment tools and molecular dynamics simulations highlights the strengths and limitations of different comparative modeling protocols, all of which are based on the same low identity template (only 22%). Incorrect alignment between the target and template was confirmed to be a major bottleneck in homology modeling, regardless of the modeling software used. Moreover, secondary structure analysis during a 50 ns simulation seems to be useful for indicating alignment errors and structural instabilities. The best model was achieved through the combined use of Phyre 2 and Modeller, suggesting the use of this protocol for the modeling of other proteins that still lack high identity templates.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Classical studies with globins provided the first example of a “molecular disease” [1,2], the first clues regarding the impact of amino acid exchanges on protein structure/function, and their consequences to human health [3–5]. The globin family also provided evidence that protein structures evolve more conservatively than protein sequences [6], and the tertiary structure conservancy observed among homologues enables comparative structure predictions. Since then, accurate prediction of the 3D structure of

proteins related to human diseases has been a major goal in structural bioinformatics [7].

Great progress has occurred over the past few decades, with improvements in the algorithms and computational resources. The Critical Assessment of protein Structure Prediction (CASP) competitions, which biannually challenge researchers to provide accurate models of an unreleased crystal structure, provide a wealth of information to this field [8,9]. These competitions also highlight the evolution of structure prediction and the persistent limitations and bottlenecks for comparative modeling [10,11].

In the present work, we focus on the homology modeling of the human α -L-iduronidase (IDUA, E.C: 3.2.1.76), a member of the glycoside hydrolase family [12]. Mutations in this enzyme are responsible for an inherited lysosomal storage disorder, called MPS I (Mucopolysaccharidosis I, OMIM #607014, #607015, #607016) [13,14]. The enzyme has 653 amino acids and is synthesized on the rough endoplasmic reticulum (ER), presenting a signal peptide with

* Corresponding author at: Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, Building 43323, Room 225, Brazil. Tel.: +555133089938

E-mail address: msinigaglia@gmail.com (M. Sinigaglia).

¹ These authors contributed equally to this work.

Improved structural method for T-cell cross-reactivity prediction

Mendes, MFA, Antunes DA, Rigo MM, Sinigaglia M, Vieira GF

Abstract

Cytotoxic T-Lymphocytes (CTLs) are the key players of adaptive cellular immunity, being able to identify and eliminate infected cells through the interaction with peptide-loaded Major Histocompatibility Complexes class I (pMHC-I). Despite the high specificity of this interaction, a given lymphocyte is actually able to recognize more than just one pMHC-I complex, a phenomenon referred as cross-reactivity. In the present work we describe the use of pMHC-I structural features as input for multivariate statistical methods, to perform standardized structure-based predictions of cross-reactivity among viral epitopes. Our improved approach was able to successfully identify cross-reactive targets among 28 naturally occurring Hepatitis C Virus (HCV) variants and among 8 epitopes from the four Dengue Virus serotypes. In both cases, our results were supported by multiscale bootstrap resampling and by data from previously published in vitro experiments. The combined use of data from charges and Accessible Surface Area (ASA) of selected residues over the pMHC-I surface provided a powerful way of assessing the structural features involved in triggering cross-reactive responses. Moreover, the use of an R package (pvclust) for assessing the uncertainty in the hierarchical cluster analysis provided a statistical support for the interpretation of results. Taken together, these methods can be applied to vaccine design, both for the selection of candidates capable of inducing immunity against different targets, or to identify epitopes that could trigger undesired immunological responses.

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

Antunes DA, Rigo MM, Freitas MV, Mendes MFA, Sinigaglia M, Selin LK, Cornberg M, Vieira GF

Abstract

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