**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA**

**Predição *In Silico* de Epítomos de
Microrganismos com Identidade a Autoantígenos
Humanos**

André Luis da Silva Breve

**Ribeirão Preto – SP
2010**

ANDRÉ LUIS DA SILVA BREVE

Predição *In Silico* de Epítomos de Microrganismos
com Identidade a Autoantígenos Humanos

Dissertação apresentada à Faculdade de
Medicina de Ribeirão Preto, para a
obtenção do título de Mestre em Ciências

Área de concentração: Genética

Orientadora: Prof^a. Dr^a. Silvana Giuliatti

Ribeirão Preto – SP

2010

FICHA CATALOGRÁFICA

Breve, André Luis da Silva

Predição *In Silico* de Epítomos de Microorganismos com Identidade a Autoantígenos Humanos. Ribeirão Preto, 2010.

79 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Genética.

Orientador: Giuliatti, Silvana.

1. Bioinformática. 2. Doenças Autoimunes. 3. Mimetismo Molecular. 4. Epítomos. 5. Antígenos.

Dedico essa conquista
aos meus pais, Decio e Janice
às minhas irmãs, Flávia e Camila
e à minha noiva, Juliane

Amo vocês...

Agradecimentos

Agradeço primeiramente a Deus, por tudo!

Agradeço à Prof^a. Dr^a. Silvana Giuliatti pela oportunidade de integração ao Grupo de Bioinformática, por todo o apoio, respeito e incentivo durante a realização deste trabalho.

Agradeço ao Prof. Dr. Geraldo Aleixo Passos e ao Prof. Dr. Eduardo Antonio Donadi pela colaboração e apoio para a realização deste trabalho.

Agradeço ao Departamento de Genética e à Faculdade de Medicina de Ribeirão Preto pela estrutura oferecida para minha formação na pós-graduação.

Agradeço aos funcionários do Departamento, em especial à Susie, sempre disposta a ajudar.

Agradeço à CAPES pela bolsa de Mestrado oferecida para a realização deste trabalho e todo o suporte financeiro.

Agradeço a disposição da banca para avaliar e discutir o presente trabalho.

Agradeço à FAEPA pelo apoio financeiro dado ao projeto, principalmente para a participação em congressos.

Agradeço aos meus grandes amigos: Saulo, Daniel, Luiza, Renato, Caninha e Edward pela dedicação, força, incentivos e amizade sem igual.

Agradeço aos demais amigos do Grupo de Bioinformática: Luciano, Pablo, Nilson, Gabi e Fernando pela amizade, apoio e pelo convívio durante esse anos.

Agradeço aos meus pais, Decio e Janice pela educação, pelos ensinamentos, por todo o esforço que fizeram para eu estar aqui e por toda a ajuda que me deram e dão sempre.

Agradeço às minhas irmãs Flávia e Camila, pela amizade, pelo apoio incondicional.

Agradeço à minha noiva, Juliane, por todo o amor, carinho, apoio, compreensão e paciência ao longo desses 8 anos.

Agradeço à minha sogra, D. Terezinha pela amizade, apoio e consideração de sempre.

Agradeço aos meus sobrinhos, Mari, João e Carol por todo carinho e alegria que proporcionam a esse tio orgulhoso

Agradeço ao meu cunhado Rodolfo, pela amizade, companheirismo e apoio.

A toda a minha família que sempre apoiou minhas decisões e sempre se manteve ao meu lado em todos os momentos.

Aos colegas do laboratório da Profª Catarina e Profª Elza pelo companheirismo e amizade.

Agradeço aos amigos de condomínio: Luciano, Matão, Dawit, Mixelã, Fausto, Mercadante e os demais que passaram por lá.

Agradeço ao Prof. Dr. Ademilson E. Espencer Soares pela amizade e companheirismo de sempre.

Em fim, agradeço a todos que de alguma forma, direta ou indiretamente, colaboraram na realização deste

Muito obrigado a todos!

RESUMO

BREVE, A. L. S. **Predição *In Silico* de Epítomos de Microrganismos com Identidade a Autoantígenos Humanos.** 2010. Dissertação (Mestrado). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. Ribeirão Preto, 2009.

A origem das doenças autoimunes é multifatorial, sendo que envolve condições ambientais e predisposição genética, dificultando sua identificação. Muitos pesquisadores têm estudado a associação entre agentes infecciosos e autoimunidade, a qual pode ser disparada pelo processo conhecido por mimetismo molecular. Neste caso, respostas imunes cruzadas envolvendo antígenos próprios têm sido documentadas. O presente projeto tem como objetivo a busca *in silico* por associações entre epítomos de microrganismos e autoantígenos humanos. Iniciaram-se as análises pela identificação de semelhanças de sequências de aminoácidos entre epítomos de microrganismos e autoantígenos humanos por meio do alinhamento local de sequências efetuado pelo programa BLASTP. As sequências de epítomos dos microrganismos e autoantígenos humanos foram previamente adquiridas nos bancos de dados *Immune Epitope Database and Analysis Resource* (IEDB) e no Genbank, respectivamente. Foram também realizadas modelagens de estruturas proteicas para o antígeno e o autoantígeno que obtiveram melhores valores de alinhamento, com base no valor do *E-value*, por meio dos programas Modeller e Rosetta. Por fim, a predição de epítomos foi executada, pelo uso dos softwares NetMHC e NetMHCII, para avaliar a possibilidade de epítomos de microrganismos e de autoantígenos humanos se associarem aos mesmos alelos de HLA. Como resultado, foram encontradas similaridades tanto de sequências proteicas quanto de afinidade a 4 tipos de alelos de HLA entre um epítomo do antígeno LSA-1 de *Plasmodium falciparum* e o autoantígeno de miosina, o que sugere uma associação entre eles, atingindo o objetivo deste trabalho.

Palavras-chave: Bioinformática, Doenças Autoimune, Epítomos, Mimetismo Molecular.

ABSTRACT

Breve, A. L. S. ***In Silico Prediction of Microorganism Motifs with Identity to Human Autoantigens.*** 2010. Thesis (Master). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. Ribeirão Preto, 2009.

The origin of autoimmune diseases is multifactorial. It involves environmental conditions and genetic predisposition that difficulties its identification. Several researchers have studied the association between infectious agents and autoimmunity, which can be initiated by a process named molecular mimicry. In this case, cross immune responses involving self antigens have been documented. This project aims to search *in silico* for associations between microorganisms epitopes and human autoantigens. The first step was the identification of similarities in amino acid sequences between microorganisms epitopes and human autoantigens by use of sequence local alignment performed by the program blastp. The sequences of the microorganisms epitope and the human autoantigens had been previously acquired in the Immune Epitope Database and Analysis Resource (IEDB) and Genbank, respectively. The modeling of protein structures for the antigen and autoantigen was also carried out to show the best alignment values, based on the E-value, using the programs Modeller and Rosetta. Finally, the prediction of epitopes was performed by use of NetMHC and NetMHCII softwares to evaluate the possibility of microorganisms epitopes and human autoantigens join the same HLA alleles. Similarities of protein sequences was found for both. It was possible to observe affinity of 4 HLA alleles between an epitope from LSA-1 *Plasmodium falciparum* antigen and the myosin, suggesting an association between them, reaching the goal of this work.

Key-words: Bioinformatics, Autoimmune Diseases, Epitopes, Molecular Mimicry.

LISTA DE FIGURAS

Figura 1. Reconhecimento de um complexo peptídeo-MHC pelo linfócito T. A figura ilustra uma molécula do MHC apresentando um peptídeo a um receptor de linfócito T (ABBAS & LICHTMAN, 2003).	4
Figura 2. Estrutura de uma molécula de MHC de classe I. (ABBAS et al., 2005).	5
Figura 3. Estrutura da molécula do MHC de classe II. (ABBAS et al., 2005).	6
Figura 4. Moléculas de MHC de classe I e II com peptídeos (em vermelho) ligados nas suas estruturas. Nota-se a cavidade fechada do MHC de classe I que engloba o peptídeo na figura à esquerda, diferente da estrutura do MHC de classe II, no qual a cavidade da estrutura é mais aberta, permitindo um peptídeo com maior número de resíduos. (GOLDSBY et al., 2003 - Alterado).	7
Figura 5. Esquema representativo da metodologia simplificada utilizada no trabalho	17
Figura 6. Visão geral da hierarquia do Immune Epitope Database and Analysis Resource (SATHIAMURTHY et al., 2005).	19
Figura 7. Histórico do crescimento do banco de dados Genbank medido pelo número de sequências depositadas compreendidas entre os anos de 1982 e 2008 (http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html).	21
Figura 8. A figura apresenta um gráfico contendo o número de estruturas presentes no banco de dados PDB à partir do ano de 1992 (www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100).	23
Figura 9. Parte do resultado gerado pelo BLAST para o antígeno liver-stage antigen 1 do protozoário Plasmodium falciparum. Nesta imagem estão apresentados os 6 autoantígenos humanos que obtiveram melhores valores de E-value para o alinhamento com o antígeno em questão.	40
Figura 10. Visualização dos 2 autoantígenos melhores alinhados ao liver-stage antigen 1 do protozoário Plasmodium falciparum de acordo com o valor de E-value.	41
Figura 11. Resultado do alinhamento global de sequências realizado pelo programa ClustalW2 para as sequências de rhabdomyosarcoma antigen MU-RMS-40.7B e α -miosina.	43
Figura 12. Continuação do resultado do alinhamento global de sequências realizado pelo programa ClustalW2 para as sequências de rhabdomyosarcoma antigen MU-RMS-40.7B e α -miosina	44

Figura 13. Resultado da qualificação da estrutura 3D modelada da proteína LSA-1 de Plasmodium falciparum à partir da ferramenta ModFOLD, versão 2.0, para os três modelos melhores qualificados. Observa-se que as estruturas são dispostas em ordem de score de qualidade do modelo e apresentam, além desta qualificação, o nome do modelo, o gráfico do erro por resíduo bem como a visualização da estrutura tridimensional da proteína.....	46
Figura 14. Tela de visualização dos resultados da qualificação da melhor proteína modelada (aat0000635.pdb) para o antígeno LSA-1 de Plasmodium falciparum, contendo os valores calculados para a estrutura bem como os parâmetros de qualificação da ferramenta.	47
Figura 15. Resultados apresentados pelo software de validação Procheck para a proteína de melhor estrutura (aat0000635.pdb). Nota-se que dos dez métodos de qualificação da estrutura utilizados pelo programa, oito aprovaram a mesma (representados na cor verde).	48
Figura 16. Mapa de Ramachandran resultante da execução da ferramenta Procheck para a estrutura do antígeno LSA-1 de Plasmodium falciparum.	49
Figura 17. Estrutura tridimensional do antígeno LSA-1 de Plasmodium falciparum que obteve melhores valores na validação pelos três programas utilizados, ModFOLD, ProQ e Procheck.....	50
Figura 18. Estrutura do antígeno LSA-1 de Plasmodium falciparum com ênfase na região do epítipo (na cor amarela) que alinhou com o autoantígeno rabdomiosarcoma.....	51
Figura 19. Visualização da estrutura da proteína do antígeno do rabdomiosarcoma, resultante do processo de modelagem por homologia pela ferramenta Modeller.	52
Figura 20. Visualização da superfície da estrutura da proteína do antígeno do rabdomiosarcoma, identificando na coloração amarela a região do epítipo do autoantígeno que foi alinhado ao epítipo do LSA-1 de Plasmodium falciparum.	53
Figura 21. Resultado da análise de validação da estrutura da proteína do antígeno de rabdomiosarcoma pelo programa ProQ.	54
Figura 22. Diagrama de Ramachandran para a proteína modelada do antígeno de rabdomiosarcoma. Nota-se que alguns resíduos não estão em regiões favoráveis, o que torna a estrutura de baixa qualidade.....	55
Figura 23. Resultado apresentado pela ferramenta de predição de epítopos NetMHCII. Oito epítopos foram preditos com afinidade para o alelo de HLA – DRB1_0101.....	60

LISTA DE TABELAS

Tabela 1 – Número de Antígenos relacionados a cada banco de dados.	32
Tabela 2 – Quantidade de sequências de antígenos adquiridas bem como o local de aquisição.....	33
Tabela 3 – Quantidade de sequências resultantes do alinhamento pela ferramenta BLAST de acordo com sua fonte bem como as descartadas do estudo.	36
Tabela 4 – Distribuição de epítomos de acordo com os organismos procedentes bem como melhores valores de E-value por sequência de epítomo de cada organismo.....	37
Tabela 5 - Alinhamentos mais significantes para o epítomo do antígeno de gi: 225719, referente ao liver-stage antigen 1 do protozoário Plasmodium falciparum.....	39
Tabela 6 - Afinidade de ligação do antígeno LSA-1 de Plasmodium falciparum e do autoantígeno rhabdomyosarcoma MU-RMS-40.7B aos diferentes alelos de HLA calculada pelo método de redes neurais. A região na cor cinza representa os epítomos preditos para o LSA-1 enquanto que a região na cor branca os do autoantígeno. Os valores em negrito e itálico representam possíveis epítomos. As células preenchidas pela cor preta simbolizam os epítomos preditos para os dois peptídeos com afinidade de ligação aos HLAs correspondentes na tabela.	59

LISTA DE ABREVIATURAS E SIGLAS

ANN	Redes Neural Artificial
BCR	Receptor de Linfócito B
BLAST	<i>Basic Local Alignment Search Tool</i>
CBS	<i>Center for Biological Sequence Analysis</i>
HLA	Antígenos Leucocitários Humanos
HMM	Modelos Ocultos de Markov
IEDB	<i>Immune Epitope Database and Analysis Resource</i>
LSA1	<i>Liver-Stage Antigen 1</i>
MHC	Complexo de Histocompatibilidade Principal
MQAPs	Model Quality Assessment Models
NCBI	<i>National Center for Biotechnology Information</i>
NIAID	<i>National Institute of Allergy and Infectious Diseases</i>
PDB	<i>Protein Data Bank.</i>
PERL	<i>Practical Extraction and Report Language</i>
RCSB	<i>Research Collaboratory for Structural Bioinformatics</i>
RF	Febre Reumática
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1. Introdução	2
1.1. Sistema Imune	2
1.2. Doenças Autoimunes	8
1.3. Bioinformática	11
2. Objetivos	15
2.1. Objetivo Geral.....	15
2.2. Objetivos Específicos.....	15
3. Materiais e Métodos	17
3.1. Buscas em Bancos de Dados de Sequências	18
3.1.1. Busca por Sequências de Epítomos	18
3.1.2. Formatação dos dados de Epítomos.....	20
3.1.3. Aquisição das Sequências de Antígenos	20
3.1.4. Busca e Aquisição de Sequências de Autoantígenos Humanos.....	20
3.2. Alinhamento de Sequências	21
3.2.1. Formatação do Banco de Dados do BLAST	22
3.3. Predição de Estruturas Terciárias de Antígenos.....	22
3.3.1. Busca por Estruturas Proteicas	22
3.3.2. Modelagem de Estruturas Proteicas	23
3.3.2.1. Modelagem por Homologia.....	24
3.3.2.2. Modelagem por <i>Ab initio</i>	25
3.3.3. Validação das Estruturas Protéicas.....	26
3.4. Predição de Epítomos.....	28
4. Resultados e Discussão	31
4.1. Busca por Epítomos.....	31
4.2. Busca por Autoantígenos Humanos	33

4.3. Alinhamento de Sequências	34
4.4. Modelagem de Estruturas Proteicas.....	45
4.4.1. Modelagem do antígeno por <i>ab initio</i> e Validação	45
4.4.2. Modelagem do autoantígeno por homologia e Validação	51
4.5. Predição de Epítomos.....	56
5. Conclusão	62
6. Referências Bibliográficas.....	64
7. Apêndice	73

INTRODUÇÃO

1. Introdução

1.1. Sistema Imune

O sistema imune é um excelente mecanismo de defesa devido a sua capacidade de reagir de forma rápida, específica e eficiente contra um grande número de microrganismos patogênicos. Além disso, apresenta importante papel na rejeição de tumores e também pode atuar na regulação de outros sistemas do corpo humano (PAUL, 2003).

A primeira linha de defesa é a imunidade inata, que consiste de mecanismos celulares e bioquímicos que estão aptos a responder rapidamente às infecções. Essa eficiência se dá pelo fato das células da imunidade inata possuírem a capacidade de reconhecer estruturas conservadas presentes em muitos microrganismos (MEDZHITOV; JANEWAY, 1997; TAKEDA *et al.*, 2003).

A segunda linha de defesa é composta pela resposta imune adaptativa, capaz de identificar e reagir a um grande número de compostos. Uma vez infectado, o hospedeiro prepara uma resposta imune específica para o agente infeccioso em questão e, depois da infecção resolvida, células de memória são armazenadas, permitindo uma resposta mais rápida e potente caso o agente infeccioso seja novamente encontrado no organismo (ABBAS & LICHTMAN, 2003).

A resposta imune adaptativa apresenta dois grandes ramos: a resposta imune celular de linfócitos T e a resposta imune humoral de linfócitos B (secretores de anticorpos). Em ambos os casos, a resposta imune é estimulada pelo reconhecimento de uma parte pequena e específica do antígeno conhecida como epítopo (GOLDSBY *et al.*, 2003). Um epítopo é qualquer estrutura

molecular que pode ser reconhecida pelo sistema imune e pode ser composto por proteínas, carboidratos, lipídeos, nucleotídeos ou uma combinação dos mesmos. É por meio do reconhecimento de epítomos estranhos (não pertencentes ao próprio organismo), que o sistema imune pode identificar e destruir os patógenos (MADDEN *et al.*, 1995).

O reconhecimento de antígenos pelos linfócitos B e T é bastante diferente. Linfócitos B e anticorpos geralmente reconhecem antígenos solúveis (proteínas intactas). Um epítomo de linfócito B é definido como uma região de uma proteína capaz de ser reconhecida tanto por um anticorpo como por um receptor de linfócito B (BCR) (PETERS *et al.*, 2005). Pelo fato destes antígenos estarem livres em solução, estes epítomos tendem a estar em locais altamente acessíveis, expostos na superfície do antígeno. De acordo com sua disposição, os epítomos podem ser classificados como lineares (aminoácidos contínuos) ou conformacionais (descontínuos porém dispostos juntos espacialmente no enovelamento da proteína). Epítomos descontínuos são definidos através da metagênese, por meio da co-cristalização ou modelagem de estruturas de proteínas, e *docking*.

Em contraste, os epítomos de linfócitos T são compostos por peptídeos curtos, processados a partir das proteínas antigênicas. Estes são apresentados no contexto do complexo de histocompatibilidade principal (MHC) ou, no caso dos humanos, pelos antígenos leucocitários humanos (HLA) de classes I e II (MAENAKA & JONES, 1999; TERASAKI, 2007). Os epítomos de linfócitos T não podem ser analisados separados das moléculas de MHC (Figura 1). O complexo MCH-peptídeo é então reconhecido por receptores de linfócitos T e este reconhecimento pode desencadear uma resposta imune (JENSEN *et al.*, 2007).

Figura 1. Reconhecimento de um complexo peptídeo-MHC pelo linfócito T. A figura ilustra uma molécula do MHC apresentando um peptídeo a um receptor de linfócito T (ABBAS & LICHTMAN, 2003).

O MHC corresponde a uma coleção de genes arranjados dentro do cromossomo 6 em humanos. Os genes de MHC de classe I codificam glicoproteínas que são expressas na superfície de quase todas as células nucleadas, sendo que a principal função dos produtos dos genes de classe I é a apresentação de peptídeos antigênicos a linfócitos T $CD8^+$. Já os genes MHC de classe II codificam glicoproteínas que são expressas primariamente em células apresentadoras de antígenos (macrófagos, células dendríticas e linfócitos B), que apresentam antígenos processados a linfócitos T $CD4^+$ (GOLDSBY *et al.*, 2003).

As proteínas do MHC são altamente polimórficas e cada uma se liga a um conjunto limitado de peptídeos. Portanto, a combinação particular de alelos MHC presentes num hospedeiro, limita a faixa de epítomos potenciais reconhecidos durante uma infecção. A conformação adotada do epítomo de linfócito T na

molécula do MHC é crítica para o reconhecimento do receptor de linfócito T (ABBAS & LICHTMAN, 2003; KORBER *et al.*, 2006).

As moléculas de MHC de classe I e II já foram isoladas e purificadas, e suas estruturas tridimensionais resolvidas por cristalografia de raio X (MADDEN *et al.*, 1995). Ambas as moléculas formam complexos bastante estáveis com os epítopos imunogênicos, apresentando-os na superfície da célula para o reconhecimento por linfócitos T.

As moléculas de MHC de classe I apresentam uma cadeia α associada não covalentemente à uma molécula de β_2 -microglobulina (Figura 2). A cadeia α é uma glicoproteína transmembrânica codificada por genes polimórficos dentro das regiões A, B e C do complexo MHC humano. A associação da cadeia α com a β_2 -microglobulina é necessária para a expressão das moléculas de classe I na superfície das células (WANG & REINHERZ, 2001).

Figura 2. Estrutura de uma molécula de MHC de classe I (ABBAS & LICHTMAN, 2005).

As moléculas de MHC de classe II apresentam duas cadeias polipeptídicas diferentes, uma cadeia α e uma cadeia β , que se associam por interações não-covalentes (Figura 3). Cada cadeia apresenta dois domínios externos: domínios $\alpha 1$ e $\alpha 2$, na cadeia α , e $\beta 1$ e $\beta 2$, na cadeia β . Os domínios $\alpha 2$ e $\beta 2$, proximais à membrana, apresentam similaridade de sequência com os domínios das imunoglobulinas. A porção distal da molécula de classe II é composta pelos domínios $\alpha 1$ e $\beta 1$ e formam a cavidade de ligação aos epítomos (GOLDSBY *et al.*, 2003).

Figura 3. Estrutura da molécula do MHC de classe II (ABBAS & LICHTMAN, 2005).

A similaridade estrutural na cavidade de ligação ao antígeno nas moléculas de MHC de classe I e de classe II faz com que elas exibam características comuns (LAFUENTE *et al.*, 2009). Os epítomos são mantidos em uma

conformação estendida que percorre todo o comprimento da cavidade. A cavidade das moléculas de classe I é bloqueada nas duas extremidades enquanto que a cavidade das moléculas de classe II é mantida aberta (Figura 4). Como resultado desta diferença, as moléculas de classe I se ligam a peptídeos que tipicamente apresentam de 8 a 10 resíduos de aminoácidos, enquanto que as moléculas de classe II acomodam peptídeos levemente maiores de 13 a 18 resíduos (LAFUENTE *et al.*, 2009).

MHC de Classe I

MHC de Classe II

Figura 4. Moléculas de MHC de classe I e II com peptídeos (em vermelho) ligados nas suas estruturas. Nota-se a cavidade fechada do MHC de classe I que engloba o peptídeo na figura à esquerda, diferente da estrutura do MHC de classe II, no qual a cavidade da estrutura é mais aberta, permitindo um peptídeo com maior número de resíduos. (GOLDSBY *et al.*, 2003 - Modificado).

A habilidade das moléculas de MHC se ligarem a uma grande variedade de peptídeos é devido à presença de aminoácidos similares distribuídos em posições definidas ao longo do peptídeo. Pelo fato destes aminoácidos ancorarem o epítipo na cavidade da molécula de MHC, eles são denominados resíduos de ancoramento. As cadeias laterais dos resíduos de ancoramento, nos epítipos, são complementares à superfície da cavidade das moléculas de MHC. Os

aminoácidos situados próximos aos sítios de ligação variam entre as diferentes formas alélicas dos genes do MHC e determinam a identidade dos resíduos que podem interagir com as moléculas (BARBER & PARHAM, 1993; ZHANG *et al.*, 1998; WANG & REINHERZ, 2001).

Devido a estas características conservadas entre os epítomos é possível prever quais epítomos poderiam se ligar às moléculas de MHC específicas. Durante os últimos 25 anos os métodos de predição de epítomos focaram primariamente em epítomos contínuos. São métodos dependentes de sequência, baseados em várias propriedades dos aminoácidos, tais como hidrofobicidade (HOPP *et al.*, 1981), acessibilidade do solvente (EMINI *et al.*, 1985), estrutura secundária (PELLEQUER *et al.*, 1993) entre outros. Nos últimos anos, vários métodos utilizando técnicas de aprendizado de máquinas foram introduzidos, por meio da aplicação de modelos ocultos de Markov (HMM) (LARSEN *et al.*, 2006), redes neurais artificiais (ANN) (SAHA & RAGHAVA, 2006), entre outras metodologias (SOLLNER *et al.*, 2006).

1.2. Doenças Autoimunes

As respostas imunes, como mencionado anteriormente, são de grande importância na proteção do organismo humano contra agentes infecciosos. Porém, por uma falha de autotolerância, essas respostas podem causar danos ao próprio organismo e desencadear as chamadas doenças autoimunes.

A autoimunidade é uma causa importante de doença no homem, com mais de 80 doenças autoimunes descritas até o momento. As doenças autoimunes chegam a afetar cerca de 5 a 8% da população dos Estados Unidos (NIH, 2004).

O estudo sobre a autoimunidade cresceu muito durante as duas últimas décadas, principalmente devido ao desenvolvimento de vários modelos dessas doenças em animais e à identificação de genes que possam predispor à autoimunidade. Entretanto, a etiologia da maioria das doenças autoimunes humanas continua desconhecida (ABBAS & LICHTMAN, 2005).

A origem das doenças autoimunes é multifatorial. Fatores ambientais, como infecções, aliados à predisposição genética, podem resultar em danos nos tecidos causados pela autoatividade das células T ou anticorpos (LEECH, 1998). Suspeita-se que alterações genéticas desempenhem papel importante na etiologia destas doenças. Porém, diferentemente de outras doenças genéticas, as autoimunes não são causadas pelo defeito de um único gene, e sim pela disfunção de um grupo complexo de genes (JAWAHEER & GREGERSEN, 2002; KAROPKA *et al.*, 2006). Apesar da diversidade clínica das diversas doenças autoimunes, elas apresentam similaridade quanto à disfunção no sistema imune (DAVIDSON & DIAMOND, 2001).

Os mecanismos envolvidos no desenvolvimento das doenças autoimunes ainda não foram totalmente compreendidos. Entretanto, há décadas tem-se estudado a possibilidade da associação de agentes infecciosos com a autoimunidade (BENOIST & MATHIS, 2001). O termo mimetismo molecular foi usado em 1970 para explicar a persistência de infecções virais. Foi sugerido que o MHC e os vírus codificavam sequências peptídicas similares, as quais permitiam que o organismo hospedeiro gerasse uma “infecção viral” em si mesmo, levando a uma resposta imune. Esta sugestão tem sido usada como uma hipótese para explicar as doenças autoimunes (LEECH, 1998).

O MHC é a uma das regiões mais estudadas do genoma humano devido à contribuição de seus múltiplos variantes no desenvolvimento de doenças autoimunes, doenças infecciosas, inflamatórias e transplantes (FERNANDO *et al.*, 2008). Entre os genes que se associam a autoimunidade, a ligação mais forte ocorre com os genes do MHC de classe II (NEJENTSEV *et al.*, 2007). Uma das explicações para estas associações propõe que moléculas de MHC associadas às doenças se ligam eficientemente a autoantígenos, o que pode resultar em uma resposta imune contra o autoantígeno levando à doença autoimune.

O interesse nesse conceito cresceu muito desde o reconhecimento de que uma resposta imune a uma infecção por estreptococo pudesse levar à febre reumática (RF), uma doença inflamatória que pode comprometer as articulações, o coração, o cérebro e a pele do paciente (CARAPETIS *et al.*, 2005).

A febre reumática ocorre após uma infecção não tratada da faringe por *Streptococcus* do grupo A em indivíduos suscetíveis. Duas manifestações da febre reumática são a artrite, que é a manifestação mais comum afetando cerca de 90% dos pacientes, e a cardite, que é a mais grave manifestação e afeta de 30 a 45% dos pacientes de febre reumática. Embora a artrite não cause danos permanentes às juntas, a cardite causa danos cardíacos que envolvem o pericárdio, miocárdio e endocárdio, ocasionando lesões valvulares progressivas e permanentes que levam à doença reumática cardíaca (FAÉ *et al.*, 2006; STEER *et al.*, 2009).

Uma das consequências graves da febre reumática é o dano valvular severo. Segundo essa hipótese, a partir da infecção na orofaringe, pelo *Streptococcus* β hemolítico do grupo A, antígenos estreptocócicos seriam liberados, desencadeando uma resposta imunológica sistêmica, com

sensibilização dos linfócitos T e B e consequente formação de anticorpos antiestreptocócicos com reatividade cruzada para estruturas cardíacas, desencadeando então a doença autoimune (GUILHERME *et al.*, 2002).

Em geral, ligações específicas de peptídeos antigênicos ao complexo MHC são pré-requisitos para a reatividade imunológica. As sequências de peptídeos que disparam a ativação das células imunes são classificados como epítomos imunodominantes. A resposta imune a um antígeno é focada em poucos, e, frequentemente, num único epítomo (KANDUC *et al.*, 2001).

Embora patógenos possam expressar milhares de proteínas contendo milhões de epítomos potenciais, apenas um pequeno grupo destas sequências é capaz de estimular o sistema imune. Para que aconteça a reatividade imunológica é necessário que ocorra a ligação específica do peptídeo antigênico ao complexo MHC. Por esta razão para encontrar os epítomos de alta afinidade de ligação é necessário uma boa predição. Ferramentas de predição de epítomos aceleram a descoberta das sequências de patógenos responsáveis por grande parte da resposta à infecção, demonstrando a importância da bioinformática aplicada à imunologia (MOUTAFTSI *et al.*, 2006).

1.3. Bioinformática

Inicialmente, a bioinformática foi definida como um campo interdisciplinar que envolvia a biologia, a ciência da computação, a matemática e a estatística para analisar dados de sequências biológicas, conteúdo de genomas e para a predição de função e estruturas de macromoléculas. Com o advento da era genômica, a bioinformática passou a atuar fortemente em pesquisas biológicas e

médicas, e representa um grande e crescente número de publicações anuais (MOUNT, 2004).

A Bioinformática tem crescido exponencialmente devido, principalmente, a enorme quantidade de dados gerados pela Biologia Molecular (BALDI & BRUNNAK, 2001), principalmente por projetos de sequenciamento do genoma e de outros esforços experimentais para determinar a estrutura e a função das moléculas biológicas. A demanda para interpretar estes dados está se expandindo rapidamente.

Muitos procedimentos hoje considerados como parte da bioinformática – comparação de sequências, pesquisas em bancos de dados de sequências, análise de sequências – são bem mais complexos do que apenas projetar e preencher bancos de dados. Os bioinformatas buscam inspiração em uma ampla variedade de campos quantitativos incluindo estatística, física, ciência da computação e engenharia (GIBAS & JAMBECK, 2001).

Nas últimas décadas, vários algoritmos computacionais têm sido usados na busca por sequências de aminoácidos de uma dada proteína por características que, acredita-se, sejam comuns a peptídeos imunogênicos, na localização de regiões que são prováveis de indução de resposta imune celular *in vitro* (LUCCHESE *et al.*, 2002).

Esses algoritmos podem identificar regiões proteicas que contenham epítomos e, em adição, podem ser usados para analisar alelos HLA que são mais representativos das populações geográficas ou subgrupos dentro de uma área geográfica. A pesquisa *in silico* pode ser também usada como uma ferramenta preliminar, a fim de avaliar a evolução da resposta imune de um indivíduo (LUCCHESE *et al.*, 2002).

Definir a especificidade da sequência do epítopo é um problema que pode ser estudado através da biologia computacional. Os padrões de aminoácidos preditos requerem a aplicação de métodos de reconhecimento de padrões, para avaliar diretamente das sequências proteicas, e determinar quais peptídeos possuem o papel de epítopos. A complexidade é composta pelo fato de que o reconhecimento de padrões pode não ser codificado pela sequência primária contínua, mas sim pela estrutura terciária da proteína.

Uma das áreas de atuação da bioinformática é justamente no estudo e predição das estruturas proteicas. Métodos computacionais para predição de estruturas tridimensionais de proteínas são alvos de grande interesse e focos de muitas pesquisas nos dias atuais. A predição de estruturas 3D de uma proteína a partir de sua sequência de aminoácidos remete a um problema científico fundamental e é frequentemente considerado um dos grandes desafios da química e da bioinformática (SCHWEDE & PEITSCH, 2008).

OBJETIVOS

2. Objetivos

2.1. Objetivo Geral

O presente trabalho teve como objetivo geral a busca *in silico* por associações entre epítomos de antígenos de microrganismos e autoantígenos humanos por meio da identificação de semelhanças de sequências de aminoácidos, bem como pelo cálculo de afinidade de ligação entre epítomos e alelos de HLAs.

2.2. Objetivos Específicos

Os objetivos específicos são apresentados a seguir:

- Modelagem e validação de estruturas terciárias de antígenos de microrganismos e autoantígenos humanos por duas abordagens diferentes, homologia e *ab initio*.
- Predição de epítomos a partir de sequências de antígenos de microrganismo e de sequências de autoantígenos humanos.

MATERIAIS E MÉTODOS

3. Materiais e Métodos

Neste capítulo, será discutida a metodologia empregada na realização deste trabalho. A figura 5 apresenta um esquema simplificado da mesma.

Figura 5. Esquema representativo da metodologia simplificada utilizada no trabalho

3.1. Buscas em Bancos de Dados de Sequências

A primeira etapa do trabalho consistiu na busca em bancos de dados por sequências de epítomos de antígenos de microrganismos diversos e de autoantígenos humanos. Para tal tarefa, dois bancos de dados foram explorados, o *Immune Epitope Database and Analysis Resource* (IEDB) (disponível no endereço <http://immuneepitope.org>) e o Genbank, presente no *website* do *National Center for Biotechnology Information* (NCBI) (acessado pelo endereço <http://www.ncbi.nlm.nih.gov>).

3.1.1. Busca por Sequências de Epítomos

O IEDB faz parte de um programa iniciado pelo *National Institute of Allergy and Infectious Diseases* (NIAID) cujo objetivo geral é catalogar e organizar uma grande quantidade de informação imunológica, como epítomos de células B e T de patógenos infecciosos e antígenos próprios (PETERS *et al.*, 2005).

Trata-se de um banco de dados público composto por classes tais como: referência, estrutura do epítomo, fonte do epítomo, ligação ao MHC, ligantes naturalmente processados, resposta de linfócitos T e resposta de linfócitos B. A figura 6 ilustra a hierarquia do banco em questão.

Figura 6. Visão geral da hierarquia do *Immune Epitope Database and Analysis Resource* (SATHIAMURTHY *et al.*, 2005).

Embora existam muitos outros bancos de dados relacionados, o IEDB se destaca pela precisão das informações armazenadas e por apresentar os epítomos juntamente com os contextos experimentais em que foram determinados (VITA *et al.*, 2009). Por estes motivos, essa base de dados foi selecionada como a fonte para a obtenção das sequências dos epítomos.

Todo o banco de dados de epítomos foi importado para o servidor modelo HP Proliant ML150 G2, com 2 processadores XEON 3.2GHz, 2Gb de memória RAM e 640GB de HD, disponível no laboratório de pesquisa do Grupo de Bioinformática (GBI) do Departamento de Genética da Faculdade de Medicina de Ribeirão Preto (FMRP) da Universidade de São Paulo (USP), no qual este trabalho foi desenvolvido.

O arquivo do IEDB, contendo os dados dos epítomos, foi extraído no formato XML (*Extensible Markup Language*), que é uma meta-linguagem de marcação que possui um conjunto de regras para definição de *tags* que dividem o documento em partes e identifica as diferentes partes deste documento (HAROLD, 1999). Este formato adotado pelo banco é de grande importância

tendo em vista que facilita o processo de aquisição das informações referentes às sequências de epítomos pertinentes ao projeto.

3.1.2. Formatação dos dados de Epítomos

A partir dos arquivos XML, as informações necessárias, tais como o código pelo qual a sequência do antígeno que contém o epítomo é referenciada no banco de dados, o tipo do epítomo (linear ou conformacional), o banco de dados que contém a sequência do antígeno e a própria sequência do epítomo, no caso de epítomos lineares, devem ser adquiridas. Para tal tarefa foram desenvolvidos alguns programas na linguagem de programação Perl (*Practical Extraction and Report Language*) (WALL *et al.*, 2001), uma vez que se trata de uma linguagem de programação estável e multiplataforma especialmente desenvolvida para processamentos de textos.

3.1.3. Aquisição das Sequências de Antígenos

Os arquivos do IEDB contêm as sequências dos epítomos, mas não as sequências dos antígenos aos quais esses epítomos pertencem. Portanto, a aquisição das mesmas tornou-se uma tarefa indispensável. Novamente, foram desenvolvidos *scripts* em linguagem de programação Perl para a busca dessas sequências utilizando as informações sobre os bancos de dados referenciados pelo arquivo XML. As duas bases de dados indicadas para a busca pelos antígenos foram o Genbank e o Swiss-Prot (<http://www.expasy.ch/sprot/>).

3.1.4. Busca e Aquisição de Sequências de Autoantígenos Humanos

Para a busca pelas sequências dos autoantígenos humanos, o banco de dados utilizado foi a versão *online* do GenBank, que é considerada a maior base de dados de sequências genômicas existente, mantida pelo *National Center for*

Biotechnology Information (NCBI) (<http://www.ncbi.nih.gov>) (WHEELER *et al.*, 2007). A figura 7 ilustra a quantidade de dados de sequências depositadas no Genbank até o ano de 2008.

Figura 7. Histórico do crescimento do banco de dados Genbank medido pelo número de sequências depositadas compreendidas entre os anos de 1982 e 2008 (<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>).

3.2. Alinhamento de Sequências

Para os alinhamentos entre as sequências de epítomos de microrganismos e autoantígenos humanos, o programa utilizado foi o BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL *et al.*, 1990) que é uma das ferramentas mais utilizadas para alinhamentos locais de sequências. O programa compara sequências de nucleotídeos ou aminoácidos com sequências de uma base de

dados e calcula a significância estatística destas semelhanças. O BLAST pode ser usado para inferir relacionamentos funcionais e evolucionários assim como ajudar na identificação de membros de famílias gênicas.

A ferramenta pode ser adquirida gratuitamente através da página *web* do BLAST (<http://www.ncbi.nih.gov/BLAST/>) onde também pode ser executada *on line*. Existe uma variedade de algoritmos BLAST que podem ser usados para buscas em diferentes bases de dados de sequências, sendo que, no presente trabalho, o algoritmo utilizado foi o BLASTP, que efetua o alinhamento entre uma sequência de proteína (epítipo) e um banco de dados também proteico (autoantígenos humanos). Para tal tarefa, foram utilizados valores padrões como parâmetros do programa.

3.2.1. Formatação do Banco de Dados do BLAST

Para que o conjunto de sequências de autoantígenos humanos fosse utilizado como banco de dados da ferramenta BLAST tornou-se necessária a sua formatação de acordo com as necessidades do programa. Para esta tarefa foi utilizada o programa formatdb, um *script* que acompanha o BLAST local. Ele realiza, de forma automática, a codificação do arquivo texto que contém as sequências, em um formato específico para ser lido pelo programa.

3.3. Predição de Estruturas Terciárias de Antígenos

3.3.1. Busca por Estruturas Proteicas

O processo de busca por estruturas proteicas de antígenos e autoantígenos, que obtiveram melhores alinhamentos, se deu no principal banco de dados de estruturas protéicas, o *Protein DataBank* (PDB).

O PDB, administrado pelo RCSB (*Research Collaboratory for Structural Bioinformatics*), uma organização sem fins lucrativos, é um repositório único de informações sobre estruturas tridimensionais que compreendem, principalmente, proteínas e ácidos nucleicos. Trata-se de um banco de dados público no qual cientistas de todo o mundo depositam estruturas e informações a cerca de suas proteínas e pode ser acessado *online* por meio de seu endereço eletrônico (<http://www.rcsb.org>) (PDB ANNUAL REPORT, 2009). O grande crescimento do PDB ao longo dos anos pode ser analisado pela figura 8.

Figura 8. A figura apresenta um gráfico contendo o número de estruturas presentes no banco de dados PDB à partir do ano de 1992 (www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100).

3.3.2. Modelagem de Estruturas Proteicas

No caso da não localização de algumas estruturas tridimensionais, a modelagem de proteínas tornou-se necessária.

Dentre as metodologias para a predição de estruturas proteicas existentes, optou-se pelos métodos de modelagem por homologia (para o atuoantígeno) e *Ab initio* (para o antígeno de microrganismo).

3.3.2.1. Modelagem por Homologia

A modelagem por homologia, ou modelagem comparativa, consiste em explorar similaridades estruturais entre proteínas, construindo um modelo tridimensional a partir de uma ou mais estruturas conhecidas. Para isso, é necessário escolher criteriosamente a proteína molde, sendo que a similaridade de sequências costuma ser um dos fatores mais importantes (BLEICHER, 2005).

O *software* utilizado para tal tarefa foi o Modeller 9v7, que é um programa computacional gratuito, responsável pela modelagem por homologia de estruturas tridimensionais de proteínas. O usuário promove um alinhamento entre a sequência a ser modelada e estruturas relacionadas já conhecidas e o Modeller, automaticamente, define um modelo baseado nas estruturas pré-existentes. Além dessa tarefa, o Modeller pode efetuar alinhamento múltiplo de sequências e/ou estruturas, *clustering* de proteínas e buscas em bancos de dados de sequências. O programa é escrito em linguagem de programação Fortran 90, não possui interface gráfica e é multi-plataforma, uma vez que pode ser executado em computadores com sistemas operacionais UNIX, Windows e Mac (SALI & BLUNDELL, 1993).

3.3.2.2. Modelagem por *Ab initio*

Com o aumento da capacidade computacional, o uso de métodos *ab initio* tem crescido a cada dia e auxiliado na investigação, em nível molecular, de diversos fenômenos da química.

Diferentemente do método de predição por homologia, no qual a sequência da proteína a ser modelada é comparada com sequências que já possuem suas estruturas resolvidas, os métodos *ab initio* se propõem a prever as propriedades de sistemas atômicos e moleculares usando, para isso, somente as leis fundamentais da mecânica quântica e algumas constantes físicas universais, tais como carga e massa do elétron, entre outras.

O programa que utiliza o método de predição por *ab initio* utilizado para a predição da estrutura da proteína foi o Rosetta cujo método resulta na melhor predição de domínios estruturais a partir da sequência de aminoácidos (BONNEAU *et al.*, 2002; CHIVIAN *et al.*, 2003). Este software pode ser executado localmente e é baseado numa combinação entre modelagem computacional usando análises de padrões de aminoácidos e experimentos laboratoriais de ressonância magnética nuclear (NMR) para estimar a energia associada com interações localizadas em sequências que são os passos iniciais na modelagem de proteínas (MOUNT, 2004). Além disso, o Rosetta explora as possíveis combinações de fragmentos utilizando simulações de Monte Carlo. A função de energia possui termos que refletem a compactação, fitas β emparelhadas e o isolamento interior de resíduos hidrofóbicos em proteínas. O procedimento, em seu parâmetro padrão, realiza 1000 simulações independentes, com a estrutura inicial selecionada a partir do padrão de distribuição de conformações dos fragmentos, gerados previamente.

3.3.3. Validação das Estruturas Protéicas

Uma vez modelada uma proteína, torna-se necessária a validação de sua estrutura terciária. A seleção dos modelos tridimensionais de estruturas proteicas de melhor qualidade a partir de um grande número de alternativas é um problema que compete à biologia estrutural. Existem duas principais abordagens para a classificação de modelos, Programas de Avaliação de Qualidade de Modelos (Model Quality Assessment Models – MQAPs), capazes de fornecer pontuações considerando-se somente um único modelo, e o agrupamento ou abordagem baseado em consenso, que procura ranquear um grupo composto por vários modelos baseando-se em comparações de estruturas par a par, podendo também incluir informações adicionais a cerca da confiabilidade dos modelos por servidores de reconhecimentos de dobramento ou por alinhamento de sequências (McGUFFIN, 2008).

Para tal fim, existem diversos métodos implementados sendo que no presente trabalho foram utilizados os programas de validação ModFOLD (McGUFFIN, 2008), ProQ (WALLNER & ELOFSSON, 2003) e Procheck (LASKOWSKI *et al.*, 1993). Todos são gratuitos e foram executados em suas versões *online*.

O primeiro *software* de validação executado foi o ModFOLD, que, a partir da combinação das duas abordagens descritas anteriormente, é capaz de fornecer um único *score* e um valor *P-value* relacionado à qualidade de um único modelo 3D de uma estrutura proteica predita e ranquear múltiplos modelos 3D para uma mesma proteína de acordo com a qualidade do modelo gerado e previsões da qualidade local (erros por resíduo) dentro de múltiplos modelos (McGUFFIN, 2008). No caso deste trabalho, como a execução da ferramenta de

modelagem de proteínas resultou em 1000 modelos, todos foram compactados num único arquivo e submetidos ao ModFOLD para que fossem ranqueados de acordo com a qualidade da estrutura predita. O ModFOLD pode ser acessado no endereço http://www.reading.ac.uk/bioinf/ModFOLD/ModFOLD_form_2_0.html.

A partir dos resultados do ModFOLD, as 10 melhores estruturas proteicas ranqueadas foram selecionadas para a execução do ProQ, que é um método de validação baseado em rede neural que leva em consideração uma série de características estruturais para a predição da qualidade de um modelo de proteína. Além disso, é otimizado para encontrar os modelos corretos em contraste com outros métodos que são otimizados para encontrar estruturas nativas. Para se medir a qualidade de uma estrutura, duas pontuações são utilizadas, o LGscore (CRISTOBAL *et al.*, 2001), que deve ter valor acima de 4,0 para que a estrutura seja excelente, acima de 2,5 para que seja considerada boa e acima de 1,5 para que seja razoavelmente boa. Outra pontuação adotada é o MaxSub (CRISTOBAL *et al.*, 2001), cujo valor superior a 0,8 basta para que a estrutura seja qualificada como excelente, acima de 0,5 é considerada muito boa e acima de 0,1 razoavelmente boa. A ferramenta é gratuita e está disponível no *website* do desenvolvedor (<http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi>).

Assim como no ProQ, as 10 melhores estruturas ranqueadas pelo ModFOLD foram submetidas ao Procheck, versão 3.5.4. Este software analisa a qualidade estereoquímica da estrutura protéica e produz gráficos que analisam a geometria total da proteína bem como a geometria resíduo a resíduo. Neste trabalho, em especial, o único diagrama que será levado em consideração para a avaliação da estrutura proteica gerada é o mapa de Ramachandran (MSRP), que apresenta os ângulos de torção phi-psi para todos os resíduos da estrutura

(exceto os da cadeia terminal), pois é o principal gráfico de avaliação de estruturas de acordo com a qualidade estereoquímica das mesmas. A coloração no mapa representa diferentes regiões, sendo que as áreas mais escuras representam os locais onde as combinações de valores phi-psi são mais favoráveis (MORRIS *et al.*, 1992). A porcentagem de resíduos presentes nessas regiões é responsável por definir a qualidade estereoquímica da estrutura. Idealmente, espera-se que a porcentagem de resíduos na região de melhor qualidade exceda os 90% para ser considerada favorável (LASKOWSKI *et al.*, 1998). A ferramenta pode ser acessada pelo *website* do Laboratório de Genômica e Proteômica Estrutural NIH MBI (<http://nihserver.mbi.ucla.edu/SAVES/>).

3.4. Predição de Epítomos

Para se calcular a afinidade com que alguns epítomos de um antígeno se ligam às moléculas do MHC (HLA em humanos), foram utilizadas duas ferramentas de predição de epítomos, o NetMHC Server (para alelos de MHC I) e o NetMHCII (para alelos de MHC II), na versão 3.0, presente no *website* do Center for Biological Sequence Analysis (CBS) (<http://www.cbs.dtu.dk/index.shtml>).

O NetMHC Server calcula a afinidade de ligação de peptídeos a diferentes alelos HLAs por meio do uso de redes neurais artificiais e matrizes de pontuação (BUUS *et al.*, 2003). O método de treinamento por rede neural artificial implementado no NetMHC tem sido usado para a predição de possíveis ligações peptídicas ao MHC de vários proteomas de patógenos virais, incluindo SARS (Síndrome Respiratória Aguda Grave), Influenza e HIV, resultando numa média de 75 a 80% de ligações comprovadas ao MHC (PETERS *et al.*, 2006).

A predição é baseada numa rede neural artificial treinada com 55 alelos MHC (43 humanos e 12 não humanos) e matriz de pontuação para outros 67 alelos HLA. As predições são possíveis para peptídeos de tamanhos que variam entre 8 e 11 aminoácidos para os 122 alelos. Os resultados da execução do programa podem ser analisados na forma *online* ou até mesmo salvos para análise local posterior (LUNDEGAARD *et al.*, 2008).

A ferramenta é de uso gratuito e está disponível no endereço <http://www.cbs.dtu.dk/services/NetMHC/>.

O NetMHC II possui uma estrutura semelhante, porém é responsável pela predição de epítomos de tamanhos que variam de 13 a 18 resíduos e a afinidade é calculada para alelos HLA – DR, de MHC de classe II.

RESULTADOS E DISCUSSÃO

4. Resultados e Discussão

4.1. Busca por Epítomos

Foram extraídas todas as sequências de epítomos disponíveis no banco de dados do IEDB na data em que o mesmo foi importado. Essa aquisição resultou em 7.457 arquivos no formato XML, contendo variadas informações a respeito do antígeno bem como os dados dos epítomos relacionados àquele antígeno especificamente. Algumas das informações disponíveis no arquivo são: código de referência do antígeno no banco do IEDB, resumo da publicação do artigo referente ao sequenciamento, o tipo químico da sequência (aminoácidos ou nucleotídeos), o tipo do epítopo (linear ou conformacional), o nome, a sequência (no caso de linear) ou os aminoácidos, bem como suas posições na sequência (no caso de epítomos conformacionais), o código de identificação do banco de dados ao qual o antígeno está cadastrado (Genbank ou Swiss-Prot) além de diversas outras informações não pertinentes a este estudo.

Destes 7.457 arquivos, 2.737 possuem a identificação de seus antígenos relacionadas ao banco de dados Genbank, sendo que o restante, 4.720, são referenciados pelo Swiss-Prot, um banco de dados de proteínas, do grupo UniProt, que tem como principal característica a baixa redundância adquirida através da fusão de uma grande quantidade de dados de fontes diferentes num mesmo banco. Outra característica marcante é o fato do banco ser curado manualmente, o que procura evitar a inconsistência dos dados (SCHNEIDER *et al.*, 2009). A tabela 1 apresenta a quantidade de antígenos referenciados pelos dois bancos de dados citados.

Tabela 1 – Número de Antígenos relacionados a cada banco de dados.

Banco de Dados	Quantidade	Porcentagem
Genbank	2.737	36,7%
Swiss-Prot	4.720	63,3%
Total	7.457	100%

A fase seguinte consistiu na busca pelas sequências destes antígenos no formato FASTA. Para tal tarefa, os bancos de dados foram armazenados no servidor local e as buscas efetuadas através de um *script* desenvolvido para tal fim.

Para as sequências referenciadas ao Genbank aproximadamente 4,1% das sequências não foram localizadas no banco local. Logo, tiveram que ser adquiridas através da versão online do banco. No final do processo, apenas 4 antígenos não tiveram suas sequências adquiridas, uma vez que são sequências que foram excluídas Genbank.

No caso dos antígenos referentes ao Swiss-Prot, as sequências não localizadas foram também submetidas à busca *online*, porém, essa pesquisa foi efetuada no Genbank devido à facilidade que esse banco apresenta quando se busca por um conjunto de sequências de uma só vez. Além disso, pôde-se confirmar a grande integração entre os dados dos dois bancos, uma vez que o resultado da busca foi satisfatório, tendo em vista que 97,8% dos antígenos cujas sequências de aminoácidos não haviam sido adquiridas a partir de então passaram a estar completos.

Tabela 2 – Quantidade de sequências de antígenos adquiridas bem como o local de aquisição.

Banco de Dados	Banco Local	Banco Online	Adquiridas	Não Localizadas
Genbank	2.624	109	2.733	4
Swiss-Prot	2.641	2.032	4.673	47
Total	5.265	2.141	7.406	51

A etapa seguinte consistiu na extração das sequências dos epítomos referentes a cada antígeno. Esta tarefa foi realizada utilizando-se também os arquivos XML adquiridos junto ao IEDB. Um *script* foi desenvolvido para efetuar essa extração de todas as sequências de epítomos, juntamente com os códigos de identificação dos antígenos e o tipo de epítopo. Todas essas informações foram salvas em um arquivo texto.

Somente os epítomos lineares foram selecionados para a continuidade do trabalho. Para tal, um arquivo foi gerado somente com as sequências no formato FASTA dos epítomos lineares contendo um total de 4.297 sequências de epítomos para a submissão à ferramenta de alinhamento.

4.2. Busca por Autoantígenos Humanos

Uma vez concluída a busca por epítomos dos antígenos presentes no banco do IEDB foram coletadas sequências de autoantígenos humanos diretamente no site do NCBI através de uma busca avançada no Genbank por sequências proteicas de antígenos do organismo *Homo sapiens*. Variadas buscas foram executadas, sendo que a que retornou maior quantidade de autoantígenos

foi a busca pelo nome da sequência contendo a palavra *antigen* referente ao organismo *Homo sapiens*. Essa busca retornou 7.039 sequências, que foram salvas do banco de dados online do Genbank e agrupadas em um arquivo texto também no formato FASTA.

4.3. Alinhamento de Sequências

Como a ferramenta BLAST alinha uma sequência por vez com todo um banco de dados de sequências, foi necessário criar um banco de dados estruturado de acordo com as necessidades do programa. Nesse caso, tanto poderia ser comparada cada sequência de epítopo com todo o banco de autoantígenos como o contrário, ou seja, cada sequência de autoantígeno com um banco de dados de epítotos. Optou-se pelo primeiro caso, pois dessa forma o resultado seria dado em função dos antígenos o que facilitaria no final do processo a identificação do microrganismo que resultou em alinhamentos com os autoantígenos.

A formatação do banco de dados do BLAST é feita de forma automática por meio de um *script* que acompanha a ferramenta, o **formatdb**. Logo, o arquivo contendo as sequências de autoantígenos humanos foi submetido a esse *script* e preparado para ser utilizado pelo BLAST.

Prontos os arquivos, tanto das sequências de epítotos quanto do banco de sequências de autoantígenos humanos, tornou-se necessária, mais uma vez, a criação de um *script* responsável por executar a ferramenta de alinhamento.

Finalmente, o alinhamento de sequências foi executado, retornando 1.752 arquivos de alinhamento, ou seja, das 4.297 sequências de epítotos lineares

submetidas, somente 1.752 sequências obtiveram quaisquer alinhamentos, cerca de 40% dos epítomos.

A análise dos resultados de um alinhamento executado por meio da ferramenta BLAST se dá em função de dois valores estatísticos, o *score* e, principalmente, o *E-value*. O BLAST procura identificar caracteres, no caso aminoácidos, que são semelhantes entre a sequência de entrada e a sequência do banco de dados. Uma vez identificadas regiões de semelhança, estes conjuntos de caracteres são estendidos em ambas as direções a fim de se ampliar as regiões de similaridade. A cada alinhamento entre um par de caracteres é dada uma pontuação que, quando somadas, definem o valor de *score* para o trecho similar. Para fornecer uma maior segurança de que esse *score* não seja reflexo de uma aleatoriedade, utiliza-se o *E-value* (SOUZA & LIFSCHITZ, 2007).

O valor do *E-value*, para um dado *score*, corresponde à probabilidade de se obter outro alinhamento com *score* igual ou superior para outra sequência aleatória de mesmo tamanho e composição de caracteres. Com isso, quanto mais próximo o valor de *E-value* for de zero, melhor é o resultado do alinhamento (SOUZA & LIFSCHITZ, 2007).

Como o BLAST foi executado com o valor de corte para *E-value* padrão do programa, todo valor de alinhamento que excedesse o valor 10.0 seria desconsiderado. Com isso, das 1752 sequências que obtiveram algum alinhamento, somente 89 delas obtiveram alinhamentos com valores de *E-value* abaixo de 10.0.

Examinando-se somente as 89 sequências de antígenos que obtiveram valores de *E-value* abaixo do corte, notou-se a presença de sequências que não

são de microrganismos no conjunto de dados de sequências de epítomos. Como o foco deste trabalho foi a identificação de regiões semelhantes entre sequências de proteínas de microrganismos e de auto-antígenos humanos, esses alinhamentos referentes às sequências de proteínas não pertencentes a microrganismos foram também excluídos das análises posteriores, bem como sequências que por algum problema foram excluídas do banco de dados do Genbank, totalizando 61 sequências excluídas, ou seja, aproximadamente 68,5% das sequências alinhadas, como ilustra a tabela 3.

Tabela 3 – Quantidade de sequências resultantes do alinhamento pela ferramenta BLAST de acordo com sua fonte bem como as descartadas do estudo.

Fontes das Sequências	Quantidade	Porcentagem	Descartadas do estudo
Microrganismos	28	31,5%	0
Homo sapiens	50	56,2%	50
Outros organismos	6	6,7%	6
Excluídas do Banco	5	5,6%	5
Total:	89	100%	61

A partir de então, somente restaram do alinhamento 28 sequências de microrganismos que obtiveram alinhamentos com alguma sequência do banco de dados de autoantígenos humanos de acordo com os resultados apresentados pela ferramenta BLAST, dispostos na tabela 4.

Tabela 4 – Distribuição de epítomos de acordo com os organismos procedentes bem como melhores valores de *E-value* por sequência de epítopo de cada organismo.

Microrganismo	Quantidade de epítomos	Melhor <i>E-value</i>
<i>Influenza A virus</i>	3	7.6
<i>Theiler's encephalomyelitis virus</i>	1	8.4
<i>Streptococcus pyogenes</i>	2	3.7
<i>Plasmodium vivax</i>	4	0.34
<i>Plasmodium falciparum</i>	6	0.089
<i>Human immunodeficiency vírus 1</i>	1	1e-08
<i>Clostridium botulium</i>	1	1.0
<i>Human herpesvirus</i>	5	0.20
<i>Leishmanis infantum</i>	1	1.1
<i>Hepatitis B vírus</i>	2	3.0
<i>Human T-lymphotropic virus 2</i>	1	0.26
<i>Rabies virus</i>	1	8.4

Dentre os resultados apresentados na tabela anterior, algumas sequências merecem destaque pela quantidade de epítomos presentes no resultado do alinhamento bem como pelo *E-value* resultante. Nota-se pela tabela, que os valores de *E-value* em geral são elevados, ou seja, que os alinhamentos não foram bem pontuados. Porém, quando se analisa o número de identidades e de positivos (como o exemplo da tabela 5), percebe-se que boa parte do

autoantígeno foi alinhada com o epítopo o que leva a uma contradição. Isso ocorre pelo fato de que as sequências dos epítomos são sequências curtas e o cálculo do *E-value* leva em consideração o tamanho da sequência de entrada. O valor estatístico acaba sendo elevado por considerar que sequências curtas possuem maior probabilidade de serem alinhadas, ao acaso, com sequências do banco de dados.

O microrganismo que possui o maior número de sequências de epítomos alinhadas é o *Plasmodium falciparum*, além do que, de acordo com o *E-value*, é uma das sequências que melhor foi alinhada com sequências do banco de autoantígenos humanos. Por estes motivos, a antígeno selecionado para as análises posteriores (modelagem da estrutura da proteína e predição de epítomos) foi o *liver-stage antigen 1* (LSA-1) (gi 225719), um antígeno do *Plasmodium falciparum* que é expresso somente em hepatócitos infectados no ciclo de vida do protozoário no fígado (HILLIER *et al.*, 2005).

O LSA-1, por possuir epítomos tanto de linfócitos T como de linfócitos B, é capaz de desencadear uma forte resposta imune (DOMARLE *et al.*, 1999; KURTIS *et al.*, 1999; LUTY *et al.*, 1999; JOHN *et al.*, 2004). Já foi demonstrado que camundongos imunizados com esta proteína são protegidos contra a malária. Vários estudos têm estabelecido que esta proteína é altamente imunogênica e capaz de produzir uma boa memória imunológica sendo um alvo importante para a construção de vacinas (HOLLINGDALE *et al.*, 1990; CONNELLY *et al.*, 1997; JOSHI *et al.*, 2000; MOORMANN *et al.*, 2009).

Este patógeno é de grande importância devido a sua ação como causador da malária, uma doença infecciosa responsável por mais de 500 milhões de casos clínicos por ano e morte de 3 milhões de pessoas anualmente no mundo, sendo,

por esse motivo, alvo de muitos estudos (GOLENSER *et al.*, 2006; NACHER *et al.*, 2000).

Recentemente, alguns estudos focaram na relação entre malária e autoimunidade, uma vez que já está estabelecido que durante o desenvolvimento da proteção imune contra essa doença ocorre a formação de autoanticorpos. Por meio da busca de anticorpos contra antígenos de *P. falciparum*, no soro de pacientes com lúpus eritematoso sistêmico, foi visto que há uma relação entre as duas doenças (SINGH *et al.*, 2001; ZANINI *et al.*, 2009). Foi observado que 44% dos pacientes com lúpus apresentam reação cruzada com o parasita sem infecção prévia de malária e que 81% dos pacientes com malária apresentaram reatividade contra autoantígenos (ZANINI *et al.*, 2009).

Os melhores autoantígenos aos quais o epítipo do referido microrganismo foi alinhado bem como os valores estatísticos referentes aos alinhamentos estão representados nas tabelas 5. Os resultados do alinhamento em questão podem ser visualizados nas figuras 9 e 10.

Tabela 5 - Alinhamentos mais significantes para o epítipo do antígeno de gi: 225719, referente ao *liver-stage antigen 1* do protozoário *Plasmodium falciparum*.

gi	Nome	Identities	Positivos	e-value
AAT44534.1	rhabdomyosarcoma antigen MU-RMS-40.7B	11/29 (37%)	22/29 (75%)	0.089
AAO65167.1	sarcoma antigen NY- SAR-24	11/28 (39%)	20/28 (71%)	0.44
NP_996769.2	sarcoma antigen NY- SAR-41	11/28 (39%)	20/28 (71%)	0.44

EAW97471.1	early endosome antigen	16/30 (53%)	22/30 (73%)	0.98
1				
AAD42862.1	NY-REN-6 antigen	11/32 (34%)	19/32 (59%)	1.3
BAA05025.1	golgi antigen gcp372	12/31 (38%)	19/31 (61%)	1.3

```

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query= |225719|
      (34 letters)

Database: autoantigens_db
      7039 sequences; 1,733,446 total letters

Searching.....done

Sequences producing significant alignments:
                                     Score
                                     (bits) Value
gb|AAT44534.1| rhabdomyosarcoma antigen MU-RMS-40.7B [Homo sapiens]      29   0.089
gb|AAO65167.1| sarcoma antigen NY-SAR-24 [Homo sapiens]                  27   0.44
ref|NP_996769.2| sarcoma antigen NY-SAR-41 [Homo sapiens]                27   0.44
gb|EAW97471.1| early endosome antigen 1, 162kD, isoform CRA_a [Homo...  25   0.98
gb|EAW97472.1| early endosome antigen 1, 162kD, isoform CRA_b [H...    25   0.98
ref|NP_003557.2| early endosome antigen 1, 162kD [Homo sapiens]          25   0.98

```

Figura 9. Parte do resultado gerado pelo BLAST para o antígeno *liver-stage antigen 1* do protozoário *Plasmodium falciparum*. Nesta imagem estão apresentados os 6 autoantígenos humanos que obtiveram melhores valores de E-value para o alinhamento com o antígeno em questão.

```

>gb|AAT44534.1| rhabdomyosarcoma antigen MU-RMS-40.7B [Homo sapiens]
      Length = 1179

Score = 28.9 bits (63), Expect = 0.089
Identities = 11/29 (37%), Positives = 22/29 (75%)

Query: 4  KLQEQSDLEQERLAKEKLQEQSDLEQE 32
      +++E + +LE ER A+ K+++ +SDL +E
Sbjct: 358 RIEEEEELEAERTARAKVEKLRSLSRE 386

Score = 22.3 bits (46), Expect = 8.3
Identities = 11/30 (36%), Positives = 17/30 (56%)

Query: 1  AKEKLQEQSDLEQERLAKEKLQEQSDLE 30
      A + LQ ++ + AK KL++Q DLE
Sbjct: 250 ALDDLQAEEDKVNTLTAKVKLEQQVDDLE 279

>gb|AAO65167.1| sarcoma antigen NY-SAR-24 [Homo sapiens]
      Length = 221

Score = 26.6 bits (57), Expect = 0.44
Identities = 11/28 (39%), Positives = 20/28 (71%)

Query: 5  LQEQSDLEQERLAKEKLQEQSDLEQE 32
      LQ+QQ L+QE + +L++ Q+ LE++
Sbjct: 57 LQQQQMLQQETIRNGELEDTQTKLEKQ 84

Score = 22.7 bits (47), Expect = 6.3
Identities = 12/33 (36%), Positives = 17/33 (51%)

Query: 2  KEKLQEQSDLEQERLAKEKLQEQSDLEQERL 34
      K K Q EQ + + LQ+QQ L+QE +
Sbjct: 37 KIKNHSLQETSEQNVILQHTLQQQQMLQQETI 69

```

Figura 10. Visualização dos 2 autoantígenos melhores alinhados ao *liver-stage antigen 1* do protozoário *Plasmodium falciparum* de acordo com o valor de *E-value*.

O antígeno humano que obteve melhor valor de *E-value* no alinhamento com o LSA-1 de *Plasmodium falciparum* foi o *rhabdomyosarcoma antigen* MU-RMS-40.7B. Informações sobre esse antígeno foram adquiridas no Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene>), onde foi visto que também é conhecido como β -miosina de cadeia pesada ou somente miosina.

Para confirmar que o rhabdomyosarcoma antigen corresponde, de fato, à β -miosina, um alinhamento global entre essas duas sequências foi realizado por meio da ferramenta Clustal W 2.0 (LARKIN *et al.*, 2007). Pelos resultados

apresentados, foi possível comprovar que se tratava da mesma proteína, uma vez que o alinhamento resultou em 100% de identidade entre as duas sequências. As figuras 11 e 12 mostram o resultado apresentado pelo Clustal W 2.0.

Pelo fato da miosina ser uma proteína que possui maior quantidade de informações relacionadas na literatura, optou-se por considerá-la como o antígeno alinhado ao LSA-1 do *Plasmodium falciparum*, uma vez que a região do epítopo presente no alinhamento é idêntico nas duas proteínas humanas.

Esta proteína é predominantemente expressa no ventrículo humano e também em tecidos de músculo esquelético ricos em fibras de contração lenta (BUVOLI *et al.*, 2008). Mudanças na abundância relativa desta proteína se relacionam com a velocidade de contração do músculo do coração. Mutações neste gene estão associadas com a cardiomiopatia hipertrófica familiar, miopatia de armazenamento de miosina e cardiomiopatia dilatada, sendo que a miocardite é umas das principais causas de parada cardíaca em pacientes com menos de 40 anos de idade (DRORY *et al.*, 1991).

A miosina pode ser considerada um importante autoantígeno humano porque é uma das proteínas mais abundantes no coração, e respostas imune à miosina são suficientes para induzir miocardite até mesmo na ausência de infecção (SMITH *et al.*, 1991).

gi 48476973 gb AAT44534.1	-----KFGHTKVFFKAGLLGLEEMRDERLSRIITRIQAQSRGVLARME	44
gi 201067589 gb ACH92815.1	IDHNQYKFGHTKVFFKAGLLGLEEMRDERLSRIITRIQAQSRGVLARME	800

gi 48476973 gb AAT44534.1	YKKLLERRDSSLVIQWNIRAFMGVKNWPMKLYFKIKPLLSAEREKEMA	94
gi 201067589 gb ACH92815.1	YKKLLERRDSSLVIQWNIRAFMGVKNWPMKLYFKIKPLLSAEREKEMA	850

gi 48476973 gb AAT44534.1	SMKEEFTRLKEALEKSEARRKELEEKMSVLLQEKNDLQLQVQAEQDNLAD	144
gi 201067589 gb ACH92815.1	SMKEEFTRLKEALEKSEARRKELEEKMSVLLQEKNDLQLQVQAEQDNLAD	900

gi 48476973 gb AAT44534.1	AEERCDQLIKNKIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDECSEL	194
gi 201067589 gb ACH92815.1	AEERCDQLIKNKIQLEAKVKEMNERLEDEEEMNAELTAKKRKLEDECSEL	950

gi 48476973 gb AAT44534.1	KRDIDDELTLAKVEKEKHATENKVKNLTEEMAGLDEIIAKLTKEKKALQ	244
gi 201067589 gb ACH92815.1	KRDIDDELTLAKVEKEKHATENKVKNLTEEMAGLDEIIAKLTKEKKALQ	1000

gi 48476973 gb AAT44534.1	EAHQQALDDLQAEEDKVNLTAKVKLEQQVDDLEGSLEQEKVVRMDLER	294
gi 201067589 gb ACH92815.1	EAHQQALDDLQAEEDKVNLTAKVKLEQQVDDLEGSLEQEKVVRMDLER	1050

gi 48476973 gb AAT44534.1	AKRKLEGDLKLTQESIMDLENDKQQLDERLKKKDFELNALNARIEDEQAL	344
gi 201067589 gb ACH92815.1	AKRKLEGDLKLTQESIMDLENDKQQLDERLKKKDFELNALNARIEDEQAL	1100

gi 48476973 gb AAT44534.1	GSQQLKKLQELQARIEELEELEAERTARAKVEKLRSDLSRELEEISERL	394
gi 201067589 gb ACH92815.1	GSQQLKKLQELQARIEELEELEAERTARAKVEKLRSDLSRELEEISERL	1150

gi 48476973 gb AAT44534.1	EEAGGATSVQIEMNKKREAEFQKMRRDLEEATLQHEATAALRKKHADSV	444
gi 201067589 gb ACH92815.1	EEAGGATSVQIEMNKKREAEFQKMRRDLEEATLQHEATAALRKKHADSV	1200

gi 48476973 gb AAT44534.1	AELGEQIDNLQRVKQKLEKEKSEFKLELDDVTSNMEQIIKAKANLEKMC	494
gi 201067589 gb ACH92815.1	AELGEQIDNLQRVKQKLEKEKSEFKLELDDVTSNMEQIIKAKANLEKMC	1250

gi 48476973 gb AAT44534.1	TLEDQMNEHRSKAEETQRSVNDLTSQQAQLQTENGELSRQLDEKEALISQ	544
gi 201067589 gb ACH92815.1	TLEDQMNEHRSKAEETQRSVNDLTSQQAQLQTENGELSRQLDEKEALISQ	1300

gi 48476973 gb AAT44534.1	LTRGKLTYTQQLDLKRLQLEEEVKAKNALAHALQSARHDCDLLREQYEEE	594
gi 201067589 gb ACH92815.1	LTRGKLTYTQQLDLKRLQLEEEVKAKNALAHALQSARHDCDLLREQYEEE	1350

gi 48476973 gb AAT44534.1	TEAKAELQRVLSKANSEVAQWRTKYETDAIQRTTEELEAKKKLAQRLQEA	644
gi 201067589 gb ACH92815.1	TEAKAELQRVLSKANSEVAQWRTKYETDAIQRTTEELEAKKKLAQRLQEA	1400

Figura 11. Resultado do alinhamento global de seqüências realizado pelo programa ClustalW2 para as seqüências de rhabdomyosarcoma antigen *MU-RMS-40.7B* e β -miosina

gi 48476973 gb AAT44534.1	EEAVEAVNAKCSSLEKTKHRLQNEIEDLMVDVERSNAAAAALDKKQRNFD 694
gi 201067589 gb ACH92815.1	EEAVEAVNAKCSSLEKTKHRLQNEIEDLMVDVERSNAAAAALDKKQRNFD 1450

gi 48476973 gb AAT44534.1	KILAEWKQKYEESQSELESSQKEARSLSTELFKLNAYEESLEHLETFR 744
gi 201067589 gb ACH92815.1	KILAEWKQKYEESQSELESSQKEARSLSTELFKLNAYEESLEHLETFR 1500

gi 48476973 gb AAT44534.1	ENKNLQEEISDLTEQLGSSGKTIHELEKVRKQLEAEKMELQSALEEEAEAS 794
gi 201067589 gb ACH92815.1	ENKNLQEEISDLTEQLGSSGKTIHELEKVRKQLEAEKMELQSALEEEAEAS 1550

gi 48476973 gb AAT44534.1	LEHEEGKILRAQLEFNQIKAEIERKLAEKDEEMEQAQRNHLRVVDSLQTS 844
gi 201067589 gb ACH92815.1	LEHEEGKILRAQLEFNQIKAEIERKLAEKDEEMEQAQRNHLRVVDSLQTS 1600

gi 48476973 gb AAT44534.1	LDAETRSRNEALRVKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLSLL 894
gi 201067589 gb ACH92815.1	LDAETRSRNEALRVKKMEGDLNEMEIQLSHANRMAAEAQKQVKSLSLL 1650

gi 48476973 gb AAT44534.1	KDTQIQLDDAVRANDDLKENIAIVERRNLLQAELEELRAVVEQTERSRL 944
gi 201067589 gb ACH92815.1	KDTQIQLDDAVRANDDLKENIAIVERRNLLQAELEELRAVVEQTERSRL 1700

gi 48476973 gb AAT44534.1	LAEQELIETSERVQLLHSQNTSLINQKKMDADLSQLQTEVEEAVQECRN 994
gi 201067589 gb ACH92815.1	LAEQELIETSERVQLLHSQNTSLINQKKMDADLSQLQTEVEEAVQECRN 1750

gi 48476973 gb AAT44534.1	AEEKAKKAITDAAMMAEELKKEQDTSALERMKKNMEQTIKDLQHRLEA 1044
gi 201067589 gb ACH92815.1	AEEKAKKAITDAAMMAEELKKEQDTSALERMKKNMEQTIKDLQHRLEA 1800

gi 48476973 gb AAT44534.1	EQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKEL 1094
gi 201067589 gb ACH92815.1	EQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKEL 1850

gi 48476973 gb AAT44534.1	TYQTEEDRKNLLRLQDLVDKQLKVKAYKRQAEAEQANTNLSKFRKVQ 1144
gi 201067589 gb ACH92815.1	TYQTEEDRKNLLRLQDLVDKQLKVKAYKRQAEAEQANTNLSKFRKVQ 1900

gi 48476973 gb AAT44534.1	HELDEAEERADIAESQVKNLRAKSRDIGTKGLNEE 1179
gi 201067589 gb ACH92815.1	HELDEAEERADIAESQVKNLRAKSRDIGTKGLNEE 1935

Figura 12. Continuação do resultado do alinhamento global de sequências realizado pelo programa ClustalW2 para as sequências de rhabdomyosarcoma antígeno *MU-RMS-40.7B* e β -miosina

4.4. Modelagem de Estruturas Proteicas

4.4.1. Modelagem do antígeno por *ab initio* e Validação

O antígeno selecionado para a modelagem de sua estrutura tridimensional, como mencionado anteriormente, foi o *liver-stage antigen 1* (LSA-1) (gi 225719), um antígeno do *Plasmodium falciparum*.

Inicialmente optou-se pela técnica de modelagem por homologia, porém como não foram localizadas estruturas com sequências proteicas semelhantes à do LSA-1, no banco de dados do PDB, o uso dessa metodologia foi impossibilitado. Decidiu-se então pela técnica de modelagem de proteínas *ab initio* por meio do uso do *software* Rosetta. A ferramenta de modelagem molecular foi então executada localmente e retornou 1000 modelos de proteínas preditos. Muitos modelos gerados mantiveram um padrão, caracterizada por duas alfa-hélices.

Para determinar a melhor estrutura dentre as 1000 geradas pelo Rosetta, a primeira ferramenta de validação utilizada foi o ModFOLD, que ranqueou as estruturas tridimensionais e as disponibilizou em ordem de qualidade baseada no valor do *score* calculado. Como exemplo do resultado da execução do ModFOLD para as proteínas em questão, a figura 13 apresenta as três melhores estruturas ranqueadas.

ModFOLDv2_0 results			
Model name	Model quality score	Residue error plot	3D view of residue error
aat0000635.pdb	0.6031
aat0000265.pdb	0.5538
aat0000224.pdb	0.5501

Figura 13. Resultado da qualificação da estrutura 3D modelada da proteína LSA-1 de *Plasmodium falciparum* à partir da ferramenta ModFOLD, versão 2.0, para os três modelos melhores qualificados. Observa-se que as estruturas são dispostas em ordem de score de qualidade do modelo e apresentam, além desta qualificação, o nome do modelo, o gráfico do erro por resíduo bem como a visualização da estrutura tridimensional da proteína.

A ferramenta ProQ apresentou como resultados os valores preditos de LGscore de 1.563, valor que corresponde a uma qualidade razoavelmente boa para a melhor estrutura modelada (aat0000635.pdb), e de MaxSub igual a 0.648, valor que qualifica a estrutura como sendo um modelo muito bom. A tela de resultados da ferramenta, bem como os parâmetros de avaliação podem ser visualizados na figura 14.

Figura 14. Tela de visualização dos resultados da qualificação da melhor proteína modelada (aat0000635.pdb) para o antígeno LSA-1 de *Plasmodium falciparum*, contendo os valores calculados para a estrutura bem como os parâmetros de qualificação da ferramenta.

Para finalizar a etapa de validação, o programa Procheck foi executado. Os resultados podem ser visualizados na figura 15. Como mencionado na metodologia, o mapa de Ramachandran foi o diagrama considerado para avaliar a qualidade da estrutura gerada. Analisando-se o mapa, nota-se que 98,4% dos resíduos estão situados nas regiões mais favoráveis para a combinação dos valores de phi-psi (figura 16), o que representa que a estrutura predita tem uma conformação altamente favorável, confirmando a consistência da estrutura.

Procheck		
Summary		
1	Note	Ramachandran plot: 98.4% core 1.6% allow 0.0% gener 0.0% disall [PostScript] • [PDF] • [JPG]
2	Note	All Ramachandrans: 0 labelled residues (out of 63) [PostScript] • [PDF] Images: 1 2 3
3	Note	Chi1-chi2 plots: 0 labelled residues (out of 56) [PostScript] • [PDF] Images: 1 2
4	Note	Main-chain params: 6 better 0 inside 0 worse [PostScript] • [PDF] • [JPG]
5	Note	Side-chain params: 5 better 0 inside 0 worse [PostScript] • [PDF] • [JPG]
6	Warning	+ Residue properties: Max.deviation: 10.2 Bad contacts: 0 Bond len/angle: 1.0 Morris et al class: 1 1 1 G-factors Dihedrals: 0.78 Covalent: 0.70 Overall: 0.75 [PostScript] • [PDF] • [JPG]
7	Note	G-factors Dihedrals: 0.78 Covalent: 0.70 Overall: 0.75 [PostScript] • [PDF] • [JPG]
8	Note	M/c bond lengths: 100.0% within limits 0.0% highlighted [PostScript] • [PDF] Images: 1 2
9	Note	M/c bond angles: 100.0% within limits 0.0% highlighted [PostScript] • [PDF] • [JPG]
10	Error	* Planar groups: 88.9% within limits 11.1% highlighted [PostScript] • [PDF] • [JPG]
View the interactive Ramachandran Plot		

Figura 15. Resultados apresentados pelo *software* de validação Procheck para a proteína de melhor estrutura (aat0000635.pdb). Nota-se que dos dez métodos de qualificação da estrutura utilizados pelo programa, oito aprovaram a mesma (representados na cor verde).

Figura 16. Mapa de Ramachandran resultante da execução da ferramenta Procheck para a estrutura do antígeno LSA-1 de *Plasmodium falciparum*.

A estrutura tridimensional melhor ranqueada pelo ModFOLD, aat0000635.pdb, foi confirmada como a melhor estrutura predita para a proteína do antígeno LSA-1 do *Plasmodium falciparum* pelos outros dois métodos de validação utilizados (ProQ e Procheck). A estrutura validada por técnicas computacionais da proteína que obteve maior score pode ser visualizada na figura 17, que apresenta a imagem gerada pelo programa de visualização de estruturas proteicas Accelrys Discovery Studio Visualizer 2.5 (HIRASHIMA & HUANG, 2008).

Figura 17. Estrutura tridimensional do antígeno LSA-1 de *Plasmodium falciparum* que obteve melhores valores na validação pelos três programas utilizados, ModFOLD, ProQ e Procheck.

É possível visualizar a região do epítipo de LSA-1 que alinhou com a miosina na estrutura da proteína gerada, como mostra a figura 18.

Figura 18. Estrutura do antígeno LSA-1 de *Plasmodium falciparum* com ênfase na região do epítipo (na cor amarela) que alinhou com o autoantígeno rabdomiosarcoma.

Vale lembrar que os métodos computacionais para a validação de estruturas proteicas não são suficientes para definir a real estrutura das mesmas, sendo que para tal são necessárias análises de bancada por meio de técnicas específicas como a cristalografia de raio X.

4.4.2. Modelagem do autoantígeno por homologia e Validação

A proteína submetida à modelagem por homologia foi o autoantígeno rabdomiosarcoma, de identificação AAT44534.1. Optou-se pela modelagem por homologia pois a sequência de aminoácidos da proteína possui estruturas com semelhanças de sequência no PDB, o que possibilita a modelagem pelo método em questão. Foram gerados 15 modelos, sendo que o melhor foi escolhido a partir dos *softwares* de validação. A melhor estrutura gerada pelo Modeller está apresentada na figura 19.

As estruturas que foram utilizadas para o processo de modelagem por homologia do autoantígeno foram a 1df1A, que é uma oxidoreductase de *Mus musculus* com resolução de 2,35 angstrons e 61% de identidade com a região alinhada, a 1tr2A, uma vinculina humana (proteína de adesão transmembrana), com resolução de 2,90 angstrons e 19% de identidade, e a 2fxm, que é uma estrutura de um pequeno fragmento (126 aminoácidos) de β -miosina com resolução de 2,70 angstrons com 100% de identidade.

Figura 19. Visualização da estrutura da proteína do antígeno do rabdomiosarcoma, resultante do processo de modelagem por homologia pela ferramenta Modeller.

Na figura 20, está representada na cor amarela a região do autoantígeno cuja sequência alinhou com o epítipo LSA-1. Nota-se que essa região é favorável para um epítipo, uma vez que é uma região voltada para o exterior da estrutura.

Figura 20. Visualização da superfície da estrutura da proteína do antígeno do rabdomiosarcoma, identificando na coloração amarela a região do epítipo do autoantígeno que foi alinhado ao epítipo do LSA-1 de *Plasmodium falciparum*.

Após a modelagem das proteínas, elas foram submetidas ao processo de validação pelos programas ProQ e Procheck. Como esperado, a partir do resultado da validação, percebe-se que a melhor estrutura predita ainda não é considerada de alta qualidade. O resultado da validação pelo ProQ está apresentado na figura 21. Nota-se que os valores de LGscore e MaxSub estão abaixo do valor mínimo para que a estrutura seja considerada, ao menos, moderadamente boa.

Figura 21. Resultado da análise de validação da estrutura da proteína do antígeno de rabdomiosarcoma pelo programa ProQ.

O diagrama de Ramachandran, gerado pela ferramenta de validação ProCheck, apresentou alguns aminoácidos presentes em regiões desfavoráveis, o que fez com que a pontuação da estrutura predita (85,7%) não atingisse o valor ideal de resíduos nas regiões mais favoráveis para os valores dos ângulos *phi-psi* (Figura 22), que é de 90%.

Figura 22. Diagrama de Ramachandran para a proteína modelada do antígeno de rabdomiosarcoma. Nota-se que alguns resíduos não estão em regiões favoráveis, o que torna a estrutura de baixa qualidade.

Um problema encontrado no processo de modelagem dessa proteína foi o tamanho elevado da sequência de seus aminoácidos (1179 resíduos), que, mesmo possuindo modelos de estruturas resolvidas no banco de dados com semelhança de sequências com ela, nem toda a sequência obteve alinhamento com as estruturas do banco, algumas lacunas foram formadas no alinhamento. Além disso, algumas sequências homólogas alinhadas não possuem alta qualidade em suas estruturas, o que acabou por prejudicar o modelo final.

4.5. Predição de Epítomos

Como etapa final do trabalho, foi realizada a predição de epítomos a partir dos mesmos peptídeos selecionados para a modelagem de estrutura proteica. Os programas NetMHC e NetMHCII (NIELSEN *et al.*, 2003) foram utilizados para a predição dos epítomos referentes às sequências alinhadas e para o cálculo das afinidades de ligação à diferentes alelos de HLAs.

O NetMHC foi inicialmente executado com o parâmetro padrão de 9 resíduos de peptídeos como o tamanho do epítopo a ser predito uma vez que uma maior quantidade de epítomos de 9 resíduos pode se ligar ao MHC de classe I (YEWDELL & BENNINK, 1999). Foram calculadas as afinidades para todos os HLAs A e B presentes no banco de dados do programa pela técnica de cálculo por redes neurais artificiais.

De acordo com as avaliações do NetMHC, para que um epítopo possa ser relacionado a algum alelo de HLA, ele deve possuir valores significativos de afinidade, no caso, valores de afinidade inferiores a 500 nM. Valores abaixo de 50

nM são suficientes para se concluir que a afinidade é excelente, enquanto que de 50 a 500 nM representam valores de afinidade moderada.

Os resultados da análise de afinidade tanto para o antígeno SLA-1 quanto para a miosina foram satisfatórios, uma vez que dois epítomos diferentes preditos obtiveram valores satisfatórios de afinidade para 6 alelos de HLA pelo cálculo baseado em ANNs enquanto que 6 sequências preditas de epítomos para o autoantígeno resultaram em afinidades com 14 alelos de HLA diferentes. Estes resultados podem ser vistos na tabela 6.

Além disso, para os alelos de HLA – A*0202, A*0211, A*0212 e A*0216, tanto alguns epítomos de LSA-1 quanto da miosina, obtiveram afinidades significativas, o que representa que um mesmo tipo de HLA possa se ligar aos epítomos dos dois peptídeos. Essa informação, aliada à semelhança entre as duas sequências proteicas comprovadas pelo alinhamento, pode sugerir que uma célula T, específica para o combate ao antígeno do *Plasmodium falciparum* em questão, poderia, por reação cruzada, reagir à miosina, assim como ocorre na cardite proveniente da doença autoimune febre reumática, um caso comprovado de mimetismo molecular provocado pela reação cruzada entre uma proteína cardíaca e o *Streptococcus pyogenes* (GORTON *et al.*, 2009). Os resultados desta predição também estão apresentados na tabela 6.

Esta mesma análise foi realizada alterando-se o número de resíduos de 9 para 8, 10 e 11 (valores possíveis de acordo com a ferramenta). Os resultados não foram tão bons quanto os da análise anterior, pois poucos epítomos preditos tiveram valores de afinidade altos ou moderados, sendo que nenhum epítomo predito para o LSA-1, com afinidade ao menos moderada, foi relacionado a um HLA de afinidade moderada para a miosina. Esse resultado era esperado uma

vez que epítomos de 8 resíduos podem não se ligar a alguns alelos de HLA devido ao seu menor tamanho, enquanto que epítomos de 10 e 11 resíduos estão presentes em quantidades limitadas (NIELSEN *et al.*, 2003).

Ainda utilizando o mesmo programa, porém com o uso da metodologia de matrizes de pontuação, cálculos foram efetuados para a predição de epítomos com afinidades a outros alelos de HLA – A e B, e HLA – C. Não foram encontrados valores de afinidade representativos para os alelos de HLAs – A e B e de HLA – C para o epítomo de LSA-1. Para a miosina, somente um valor de afinidade moderada foi encontrado para o alelo de HLA – Cw1601 que, por não estar associado ao do LSA-1, não consta aqui nos resultados.

Tabela 6 - Afinidade de ligação do antígeno LSA-1 de *Plasmodium falciparum* e da miosina aos diferentes alelos de HLA calculada pelo método de redes neurais. A região na cor cinza representa os epítomos preditos para o LSA-1 enquanto que a região na cor branca os do autoantígeno. Os valores em negrito e itálico representam possíveis epítomos. As células preenchidas pela cor preta simbolizam os epítomos preditos para os dois peptídeos com afinidade de ligação aos HLAs correspondentes na tabela.

Posição	Sequência	Tipo de HLA							
		A*0202	A*0204	A*0211	A*0212	A*0216	A*0301	A*1101	
0	KLQEQQSDL	73	47	59	18	256	29360	38352	
10	QERLAKEKL	29633	41400	38281	31429	30992	46393	45662	
17	KLQEQQSDL	73	47	59	18	256	29360	38352	
0	RIEEL EEL	453	4372	58	129	194	44293	40103	
2	EELEEELEA	34575	33188	19103	19062	29574	45981	45128	
6	EELEAERTA	34657	36687	28074	26729	33282	46324	45486	
7	ELEAERTAR	27644	34015	21128	18862	22207	20030	17173	
10	AERTARAKV	25060	31426	21986	13990	9648	45434	44860	
12	RTARAKVEK	24078	27993	26688	22298	24001	179	30	

Posição	Sequência	Tipo de HLA							
		A*3001	A*3101	A*3301	A*6801	B*1801	B*4001	B*4403	B*4501
0	KLQEQQSDL	29342	23670	41275	42447	38564	36055	44714	38230
10	QERLAKEKL	26300	38623	42584	42997	516	199	11627	7517
17	KLQEQQSDL	29342	23670	41275	42447	38564	36055	44714	38230
0	RIEEL EEL	28150	28752	42618	39361	32953	27335	44249	35927
2	EELEEELEA	40520	41965	42133	40664	3541	4362	327	125
6	EELEAERTA	39336	41700	41486	41530	9044	6464	1260	193
7	ELEAERTAR	20805	6351	132	133	43140	40382	44935	37741
10	AERTARAKV	8886	25858	39911	43694	417	1573	6406	169
12	RTARAKVEK	14	67	16089	64	41569	37758	42111	38029

O segundo programa utilizado para a predição de epítomos foi o NetMHC II, que, além das predições, calcula valores de afinidade para associação de epítomos a alelos de HLA – DR. Não foram preditos epítomos do LSA – 1 capazes de se ligar a nenhum alelo de HLA – DR para peptídeos de tamanho compreendido entre 13 e 18 (únicos tamanhos disponíveis para os peptídeos no programa). Para o autoantígeno, foram encontrados epítomos capazes de se ligarem a alelos de HLA específicos, principalmente para o alelo HLA – DRB1_0101, como mostra a figura 23.

```
# Input is in FSA format
NetMHCII version 2.0.

Strong binder threshold 50.00. Weak binder threshold 500.00.
```

Allele	pos	peptide	core	1-log50k(aff)	affinity(nM)	Bind Level	%Random	Identity
DRB1_0101	0	RIEEEEEEEAER	IEEEEELE	0.1272	12627.4		50.00	Sequence
DRB1_0101	1	IEEEEELEAERT	EEEELEAERT	0.1850	6751.9		50.00	Sequence
DRB1_0101	2	EEEEEELEAERTA	EEEELEAERT	0.2185	4701.4		50.00	Sequence
DRB1_0101	3	ELEELEAERTAR	EEEELEAERT	0.2549	3169.8		50.00	Sequence
DRB1_0101	4	LEEELEAERTARA	LEAERTARA	0.4719	303.0	WB	50.00	Sequence
DRB1_0101	5	EEEELEAERTARAK	LEAERTARA	0.5360	151.5	WB	50.00	Sequence
DRB1_0101	6	EELEAERTARAKV	LEAERTARA	0.6294	55.1	WB	32.00	Sequence
DRB1_0101	7	ELEAERTARAKVE	LEAERTARA	0.6503	44.0	SB	32.00	Sequence
DRB1_0101	8	LEAERTARAKVEK	LEAERTARA	0.6183	62.2	WB	32.00	Sequence
DRB1_0101	9	EAERTARAKVEKL	TARAKVEKL	0.2924	2113.5		50.00	Sequence
DRB1_0101	10	AERTARAKVEKLR	TARAKVEKL	0.3790	827.9		50.00	Sequence
DRB1_0101	11	ERTARAKVEKLRS	TARAKVEKL	0.4060	618.1		50.00	Sequence
DRB1_0101	12	RTARAKVEKLRS	TARAKVEKL	0.3772	844.5		50.00	Sequence
DRB1_0101	13	TARAKVEKLRS	TARAKVEKL	0.2693	2713.2		50.00	Sequence
DRB1_0101	14	ARAKVEKLRS	VEKLRS	0.5985	77.0	WB	32.00	Sequence
DRB1_0101	15	RAKVEKLRS	VEKLRS	0.6883	29.1	SB	32.00	Sequence
DRB1_0101	16	AKVEKLRS	VEKLRS	0.6600	39.6	SB	32.00	Sequence

```
Allele: DRB1_0101. Number of high binders 3. Number of weak binders 5. Number of peptides 17
```

Figura 23. Resultado apresentado pela ferramenta de predição de epítomos NetMHCII. Oito epítomos foram preditos com afinidade para o alelo de HLA – DRB1_0101.

Como não foram preditos epítomos com afinidades a nenhum alelo para o LSA-1, não foi possível a associação entre o antígeno em questão e a miosina por meio de um alelo em comum.

CONCLUSÃO

5. Conclusão

Diante dos resultados apresentados pelo presente trabalho, pode-se concluir que os objetivos inicialmente propostos foram atingidos, uma vez que foi identificada uma potencial associação de um epítopo de um microrganismo a um autoantígeno humano, por meio da metodologia aplicada. O epítopo do antígeno LSA-1 do *Plasmodium falciparum* obteve semelhanças de sequências proteicas com um autoantígeno referente à proteína miosina, além disso, apresentaram valores significantes de afinidade de seus epítomos preditos para alguns alelos de HLAs em comum.

Além disso, como o epítopo de LSA-1 do *P. falciparum* está relacionado à malária, e a miosina está relacionada à miocardite autoimune e essas duas proteínas têm afinidades para 4 alelos de HLA em comum, pode-se sugerir que o *P. falciparum* possa, de alguma forma, estar associado à miocardite autoimune.

Vale ressaltar que as análises *in silico* são de grande importância para o meio científico, pois, por meio de suas técnicas, podem simular muitos procedimentos de forma mais rápida e econômica se comparada a estudos de bancada. Porém, resultados de análise *in silico*, como o apresentado neste trabalho, necessitam de estudos mais avançados em bancada para que sejam comprovados.

REFERÊNCIAS BIBLIOGRÁFICAS

6. Referências Bibliográficas

- ABBAS, A. K.; LICHTMAN, A. H. **Imunologia Básica: Funções e Distúrbios do Sistema Imune**. 2. ed. Rio de Janeiro: Elsevier, 2007.
- ABBAS, A. K.; LICHTMAN, A. H. **Imunologia Celular e Molecular**. 5. ed. Rio de Janeiro: Elsevier, 2005.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W. AND LIPMAN, D. J. Basic Local Alignment Search Tool. **Journal of Molecular Biology**, v. 215, p. 401-410, 1990.
- BAIROCH, A.; APWEILER, R. The SWISS-PROT Protein Sequence Database and Its Supplement TrEMBL in 2000. **Nucleic Acids Research**, v. 28, n. 1, p. 45-48, 2000.
- BALDI, P.; BRUNNAK, S. **Bioinformatics: The Machine Learning Approach**. 2. ed., Cambridge: MIT Press, 2001.
- BARBER, L. D.; PARHAM, P. Peptide binding to major histocompatibility complex molecules. *Annu Rev Cell Biol*, v.9, p. 9: 163-206, 1993.
- BENOIST, C.; MATHIS, D. Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? **Nature Immunology**, v. 2, n. 9, p. 797-801, 2001.
- BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; WHEELER, D. L. GenBank. **Nucleic Acids Research**, v. 35, p. 25-30, 2008.
- BLAST - Basic Local Alignment Search Tool. Disponível em: <<http://www.ncbi.nih.gov/blast>>. Acesso em: 14 mar. 2008.
- BLEICHER, L. **Estudos Estruturais do Receptor do Hormônio Tireoideano (TR) e Modelagem por Homologia da Globulina de Ligação à Tiroxina (TBG)**. 2005. 87f. Dissertação (Mestrado em Física Aplicada) – Universidade de São Paulo, São Carlos, 2005.
- BONNEAU, R.; STRAUSS, C. E.; ROHL, C. A.; CHIVIAN, D.; BRADLEY, P.; MALMSTROM, L.; ROBERTSON, T.; BAKER, D. De novo Prediction of Three-dimensional structures for major protein families. **J. Mol. Biol.**, v. 322, p. 65-78, 2002.
- BUUS, S.; LAUEMOLLER, S. L.; WÖRNING, P.; KESMIR, C.; FRIMURER, T.; CORBET, S.; FOMSGAARD, A.; HILDEN, J.; HOLM, A.; BRUNAK, S. Sensitive Quantitative Predictions of Peptide-MHC Binding by a 'Query by Committee' Artificial Neural Network Approach. **Tissue Antigens**, v. 62, n. 5, p. 378-384, 2003.

BUVOLI, M.; HAMADY, M.; LEINWAND, L. A.; KNIGHT, R. Bioinformatics Assessment of β -Myosin Mutations Reveals Myosin High Sensitivity to Mutations. **Trends Cardiovasc Med**, v. 18, n. 4, p. 141–149, 2008.

CARAPETIS, J. R.; Steer, A. C.; Mulholland, E. K.; WEBER, M. The global burden of group A streptococcal diseases. **Lancet Infect Dis**, v. 5, p. 685–694, 2005.
CHANG, H. H.; SHYU, H. F.; WANG, Y. M.; SUN, D. S.; SHYU, R. H.; TANG, S. S.; HUANG, Y. S. Facilitation of cell adhesion by immobilized dengue viral nonstructural protein I (NSI): arginine-glycine-aspartic acid structural mimicry within the dengue viral NSI antigen. **The Journal of Infectious Diseases**, v. 186, n. 6, p. 743-751, 2002.

CHIVIAN, D.; KIM, D. E.; MALMSTROM, L.; BRADLEY, P.; ROBERTSON, T.; MURPHY, P.; STRAUSS, C. E.; BONNEAU, R.; ROHL, C. A.; BAKER, D. Automated Prediction of CASP-5 Structures Using the Robetta Server. **Proteins**, v. 6, p. 524-533, 2003.

CONNELLY, M.; KING, C. L.; BUCCI, K.; WALTERS, S.; GENTON, K.; ALPERS, M. P.; HOLLINGDALE, M.; KAZURA, J. W. T-cell immunity to peptide epitopes of liver-stage antigen 1 in an area of Papua New Guinea in which malaria is holoendemic. **Infect. Immun**, v. 65, p. 5082–5087, 1997.

CRISTOBAL, S.; ZEMLA, A.; FISCHER, D.; RYCHLEWSKI, L.; ELOFSSON, A. A study of quality measures for protein threading models. **BMC Bioinformatics**, v. 2, n. 5, 2001.

DAVIDSON, A., DIAMOND, B. Autoimmune diseases. **N Engl J Med**, v.345, n.5, p.340-50, 2001.

DE GROOT, A. S.; SAINT-AUBIN, C.; BOSMA, A.; SBAI, H.; RAYNER, J.; MARTIN, W. Rapid determination of HLA B*07 ligands from the west Nile virus NY99 genome. **Emerging Infectious Diseases**, v. 7, n. 4, p. 706-713, 2001.

DOMARLE, O.; MIGOT-NABIAS, F.; MVOUKANI, J. L.; LU, C. Y.; NABIAS, R.; MAYOMBO, J. Factors influencing resistance to reinfection with *Plasmodium falciparum*. **Am J Trop Med Hyg**, v. 61, n. 6, p. 926–931, 1999.

DRORY, Y.; TURETZ, Y.; HISS, Y.; LEV, B.; FISMAN, E. Z.; PINES, A.; KRAMER, M. R. Sudden unexpected death in persons less than 40 years of age. *Am J Cardiol*, v. 68, p.1388–1392, 1991.

EMINI, E. A.; HUGHES, J. V.; PERLOW, D. S.; BOGER, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. **J. Virol**, v. 55, n.3, p. 836-839, 1985.

FAÉ, K. C.; SILVA, D. D.; OSHIRO, S. E.; TANAKA, A. C.; POMERANTZEFF, P. M. A.; DOUAY, C.; CHARRON, D.; TOUBERT, A.; CUNNINGHAM, M. W.; KALIL, J.; GUILHERME, L. Mimicry in Recognition of Cardiac Myosin Peptides by Heart-Intralesional T Cell Clones from Rheumatic Heart Disease. **The Journal of Immunology**, v. 176, p. 5662-5670, 2006.

- FERNANDO, M. M. A.; STEVENS, C. R.; WALSH, E. C.; DE JAGER, P. L.; GOYETTE, P.; Plenge, M. R.; Vyse, T. V.; Rioux, J. D. Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. **PLoS Genet**, v. 4, n. 4, e1000024, 2008.
- GARZA, K. M.; TUNG, K. S. K. Frequency of molecular mimicry among T cell peptides as the basis for autoimmune disease and autoantibody induction. **The American Association of Immunologists**, v. 155, n. 11, p. 5444-54448, 1995.
- GIBAS, C.; JAMBECK, P. **Desenvolvendo Bioinformática**. 1. ed. Rio de Janeiro: Campus, 2001.
- GOLENSER, J.; MCQUILLAN, J.; HEE, L.; MITCHELL, A. J.; HUNT, N. H. Conventional and experimental treatment of cerebral malaria. **Int. J. Parasitol.**, v. 36, p. 583–593, 2006.
- GOLDSBY, R.; KINDT, T. J.; OSBORNE, P. A.; KUBY, J. **Immunology**. 5. ed. New York: W. H. Freeman, 2003.
- HAROLD, E. R. **XML Bible**. 1. ed. Foster City, CA: IDG Books Worldwide Inc., 1999.
- HILLIER, C. J.; WARE, L. A.; BARBOSA, A.; ANGOV, E.; LYON, J. A.; HEPPNER, D. G.; LANAR, D. E. Process Development and Analysis of Liver-Stage Antigen 1, a Preerythrocyte-Stage Protein-Based Vaccine for *Plasmodium falciparum*. **Infection and Immunity**, v. 73, n. 4, p. 2109-2115, 2005.
- HIRASHIMA, A.; HUANG, H. Homology modeling, agonist binding site identification, and docking in octopamine receptor of *Periplaneta americana*. **Comput. Biol. Chem.**, v. 32, p. 185-190, 2008.
- HOLLINGDALE, M. R.; AIKAWA, C. T.; ATKINSON, W. R.; BALLOU, G.; CHEN, J.; LI, J. F. G. M.; MEIS, B.; SINA, C.; WRIGHT, AND J. ZHU. Non-CS preerythrocytic protective antigens. **Immunol. Lett.** v. 25, p. 71–76, 1990.
- HOPP, T. P.; WOODS, K. R. Prediction of Protein Antigenic Determinants from Amino Acid Sequences. **Proc Natl Acad Sci**, v. 78, p. 3824-3828, 1981.
- JAWAHEER, D.; GREGERSEN, P. K. Rheumatoid arthritis: The genetic components. **Rheumatic Disease Clinics of North America**, v.28, n.1, p.1, 2002.
- JENSEN, P.E. Recent advances in antigen processing and presentation. **Nat Immunol**, v. 8, p. 1041-1048, 2007.
- JOHN, C. C.; MOORMANN, A. M.; SUMBA, P. O.; OFULLA, A. V.; PREGIBON, D. C.; KAZURA, J. W. Gamma interferon responses to *Plasmodium falciparum* liver-stage antigen and thrombospondin-related adhesive protein and their relationship to age, transmission intensity, and protection against malaria. **Infect Immun**, v. 72, p. 5135–5142, 2004.

JOSHI, S. K.; BHARADWAJ, A.; CHATTERJEE, S.; CHAUHAN, V. S. Analysis of Immune Responses against T- and B-Cell Epitopes from *Plasmodium falciparum* Liver-Stage Antigen 1 in Rodent Malaria Models and Malaria-Exposed Human Subjects in India. **Infection and Immunity**, v. 68, n. 1, p. 141–150, 2000.

KANDUC, D.; LUCCHESI, A.; MITTELMAN, A. Individuation of monoclonal anti-HPV16 E7 antibody linear peptide epitope by computational biology. **Peptides**, v. 22, n. 12, p. 1981-1985, 2001.

KAROPKA, T., FLUCK, J., MEVISSSEN, H. T. , GLASS, A. The Autoimmune Disease Database: a dynamically compiled literature-derived database. **BMC Bioinformatics**, v.7, p.325. 2006.

KORBER, B.; LABUTE, M.; YUSIM, K. Immunoinformatics comes of age. **PloS Computational Biology**, v. 2, n. 6, p. 484-492, 2006.

KURTIS, J. D.; LANAR, D. E.; OPOLLO, M.; DUFFY, P. E. Interleukin-10 responses to liver-stage antigen 1 predict human resistance to *Plasmodium falciparum*. **Infect Immun**, v. 67, p. 3424–3429, 1999.

LAFUENTE, E. M & RECHE, P. A. Prediction of MHC-Peptide Binding: A Systematic and Comprehensive Overview. **Current Pharmaceutical Design**, v. 15, p. 3209-3220, 2009.

LARKIN, M.A.; BLACKSHIELDS, G.; BROWN, N. P.; CHENNA, R.; MCGETTIGAN, P. A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I. M.; WILM, A.; LOPEZ, R.; THOMPSON, J. D.; GIBSON, T. J.; HIGGINS, D. G. Clustal W and Clustal X version 2.0. **Bioinformatics**, vol. 23 n. 21, p. 2947-2948, 2007.

LARSEN, J. E.; LUND, O.; NIELSEN, M. Improved method for predicting linear B-cell epitopes. **Immunome Res**, v. 2, n. 2, 2006.

LASKOWSKI, R. A.; MacARTHUR, M. W.; SMITH, D. K.; JONES, D. T.; HUTCHINSON, E. G.; MORRIS, A. L.; MOSS, D. S.; THORNTON, J. M. **Programs to Check the Stereochemical Quality of Protein Structures – Operating Manual**. 1998.

LEECH, S. Molecular mimicry in autoimmune disease. **Arc. Dis. Child**, v. 79, n. 5, p. 448-451, 1998.

LIN, C. F.; LEI, H. Y.; SHIAU, A. L.; LIU, C. C.; LIU, H. S.; YEH, T. M.; CHEN, S. H.; LIN, Y. S. Antibodies from dengue patient sera cross-react with endothelial cells and induce damage. **Journal of Medical Virology**, v. 69, n. 1, p. 82-90, 2003.

LUCCHESI, A.; STEVANOVIC, S.; SINHA, A. A.; MITTELMAN, A.; KANDUC, D. Role of MHC II affinity and molecular mimicry in defining anti-HER-2/neu MAb-3 linear peptide epitope. **Peptides**, v. 24, n. 2, p. 193-197, 2003.

LUNDEGAARD, C.; LAMBERTH, K.; HARND AHL, M.; BUUS, S.; LUND, O.; NIELSEN, M. NetMHC-3.0: Accurate Web Accessible Predictions of Human, Mouse

and Monkey MHC Class I Affinities for Peptides of Length 8-11. **Nucleic Acids Research**, v. 36, p. W509-W512, 2008.

LUTY, A. J.; LELL, B.; SCHMIDT-OTT, R.; LEHMAN, L. G.; LUCKNER, D.; GREVE, B. Interferon-gamma responses are associated with resistance to reinfection with *Plasmodium falciparum* in young African children. **J Infect Dis**, p. 179-980, 1999.

MADDEN, D. The three-dimensional structure of peptide-MHC complexes. **Annu. Rev. Immunol**, v.13, p. 587-622, 1995.

MAENAKA, K.; JONES, E. Y. MHC superfamily structure and the immune system. **Curr Opin Struct Biol**, v. 9, p. 745-753, 1999.

McGUFFIN, L. J. The ModFOLD Server for the Quality Assessment of Protein Structural Models. **Bioinformatics**, v. 24, n. 4, p. 586-587, 2008.

MEDZHITOV, R.; JANEWAY Jr, C. A. Innate immunity: the virtues of a nonclonal system of recognition. **Cell**, v. 91, p. 295–298, 1997.

MEIDANIS, J.; SETUBAL, J. C. Introduction to Computacional Molecular Biology. 1 ed., Boston: PWS Publishing Company, 1997.

MOORMANN, A. N.; SUMBA, P. O.; TISCH, D. J.; EMBURY, P.; KING, C. H.; KAZURA J. W.; JOHN, C. C. Stability of Interferon-Gamma and Interleukin-10 Responses to *Plasmodium falciparum* Liver Stage Antigen 1 and Thrombospondin-Related Adhesive Protein Immunodominant Epitopes in a Highland Population from Western Kenya. **Am. J. Trop. Med. Hyg**, v. 81, n. 3, p. 489–495, 2009.

MORRIS, A. L.; MACARTHUR, M. W.; HUTCHINSON, E. G.; THORNTON, J. M. Stereochemical quality of protein structure coordinates. **Proteins**, v. 12, p. 345-364, 1992.

MOUNT, D. W. **Bioinformatics: Sequences and Genome Analysis**. 2. ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2004.

NACHER, M.; GAY, F.; SINGHASIVANON, P.; KRUDSOOD, S.; TREEPRASERTSUK, S.; MAZIER, D.; I. VOULDOUKIS, I.; LOOAREESUWAN, S. *Ascaris lumbricoides* infection is associated with protection from cerebral malaria. **Parasite Immunol.**, v. 22, p. 107–113, 2000.

National Institutes of Health Autoimmune Disease Coordinating Committee Report, 2004. Bethesda (MD): The Institutes; 2004.

NCBI – National Center for Biotechnology Information. Disponível em: <<http://www.ncbi.nih.gov/>>. Acesso em: 14 mar. 2008.

NEJENTSEV, S.; HOWSON, J. M. M.; WALKER, N. M.; SZESZKO, J.; FIELD, S. F.; STEVENS, H. E.; REYNOLDS, P.; HARDY, M.; KING, E.; MASTERS, J.; HULME, J.; MAIER, L. M.; SMYTH, D.; BAILEY, R.; COOPER, J. D.; RIBAS, G.; CAMPBELL, R.D.; The Wellcome Trust Case Control Consortium, Clayton, D. G. & Todd, J. A.

Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. **Nature**, v. 450, n. 6, p. 887-893, 2007.

NIELSEN, M.; LUNDEGAARD, C.; WORNING, P.; LAUEMOLLER, S. L.; LAMBERTH, K.; BUUS, S.; BRUNAK, S.; LUND, O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. **Protein Sci.**, vol. 12, p. 1007-1017, 2003.

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS. **PDB Annual Report July 2009**. Rutgers, 2009. 16 p.

PAUL, W. E. **Fundamental Immunology**. 5. ed. Philadelphia, Pa: Lippincott Williams & Wilkins, 2003.

PELLEQUER, J. L.; WESTHOF, E.; VAN REGENMORTEL, M. H. Correlation between the location of antigenic sites and the prediction of turns in proteins. **Immunol Lett**, v. 36, n. 1, p. 83-99, 1993.

PETERS, B.; BUI, H. H.; FRANKILD, S.; NIELSON, M.; LUNDEGAARD, C.; KOSTEM, E.; BASCH, D.; LAMBERTH, K.; HARND AHL, M.; FLERI, W.; WILSON, S. S.; SIDNEY, J.; LUND, O.; BUUS, S.; SETTE, A. A community resource benchmarking predictions of peptide binding to MHC-I molecules. **PLoS Comput. Biol.**, v. 2, n. 6, 2006.

PETERS, B.; SIDNEY, J.; BOURNE, P.; BUI, H. H.; BUUS, S.; DOH, G.; FLERI, W.; KRONENBERG, M.; KUBO, R.; LUND, O.; NEMAZEE, D.; PONOMARENKO, J. V.; SATHIAMURTHY, M.; SCHOENBERGER, S. P.; STEWART, S.; SURKO, P.; WAY, S.; WILSON, S.; SETTE, A. The immune epitope database and analysis resource: from vision to blueprint. **PLoS Biol**, v. 3, n. 3, p. 379-381, 2005.

SAHA, S.; RAGHAVA, G. P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. **Proteins**, v. 65, n. 1, p. 40-48, 2006.

SALI, A.; BLUNDELL, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. **J. Mol. Biol.**, v. 234, p. 779-815, 1993.

SATHIAMURTHY, M.; PETERS, B.; BUI, H. H.; SIDNEY, J.; MOKILI, J.; WILSON, S. S.; FLERI, W.; McGUINNESS, D.; BOURNE, P. E.; SETTE, A. An Ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. **Immunome Research**, v. 1, n. 2, 2005.

SCHNEIDER, M.; LANE, L.; BOUTET, E.; LIEBERHERR, D.; TOGNOLLI, M.; BOUGUELERET, L.; BAIROCH, A. The UniProtKB/Swiss-Prot Knowledgebase and its Plant Proteome Annotation Program. **J. Proteomics**, v. 72, n. 3, p. 567-573, 2009.

SCHWEDE, T.; PEITSCH, M. **Computational Structural Biology: Methods and Applications**. 1. ed. Singapore: World Scientific Publishing Co., 2008.

SILVESTRIS, F.; WILLIAMS, R. C. JR.; DAMMACCO, F. Autoreactivity in HIV-1 infection: the role of molecular mimicry. **Clinical Immunology and Immunopathology**, v. 75, n. 3, p. 197-205, 1995.

SINGH, S.; CHATTERJEE, S.; SOHONI, R.; BADAHERE, S.; SHARMA, S. Sera from lupus patients inhibit growth of *P. falciparum* in culture. **Autoimmunity**, v. 33, n. 4, p. 253-63, 2001.

SOLLNER, J.; MAYER, B. Machine learning approaches for prediction of linear B-cell epitopes on proteins. **J Mol Recognit**, v. 19, n. 3, p. 200-208, 2006.

SOUZA, D. X.; LIFSCHITZ, S. **Avaliação do E-value para Execução do BLAST sobre Bases de Dados Fragmentadas**. 2007. Monografia – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

STEER, A. C.; KADO, J.; JENNEY, A. W.; BATZLOFF, M.; WAQATAKIREWA, L.; MULHOLLAND, E. K.; CARAPETIS, J. R. Acute rheumatic fever and rheumatic heart disease in Fiji: prospective surveillance, 2005–2007. **Med J Aust**, v. 190, n. 3, p. 133–135, 2009.

TAKEDA, K.; KAISHO, T.; AKIRA, S. Toll-like receptors. **Annu Rev Immunol**, v. 21, p. 335–376, 2003.

TERASAKI PI. A brief history of HLA. **Immunol Res**, v. 38, p.139- 148, 2007.

VITA, R.; ZAREBSKI, L.; GREENBAUM, J. A.; EMAMI, H.; HOOF, I.; SALIMI, N.; DAMLE, R.; SETTE, A.; PETERS, B. The Immune Epitope Database 2.0. **Nucleic Acids Research**, v. 38, p. 854 – 862, 2009.

WALL, L.; CHRISTIANSEN, T.; ORWANT, J. **Programação Perl**. 3. ed., Rio de Janeiro: Campus, 2001.

WALLNER, B. & ELOFSSON, A. Can correct protein models be identified? **Protein Sci**. v. 12, n. 5, p. 1073-1086, 2003.

WANG, J. H.; REINHERZ, E. Structural basis of T cell recognition of peptides bound to MHC molecules. *Mol Immunol*, v.38, p. 1039-1049, 2001.

WHEELER, D. L.; BARRETT, T.; BENSON, D. A.; BRYANT, S. H.; CANESE, K.; CHETVERNIN, V.; CHURCH, D. M.; DICUCCIO, M.; EDGAR, R.; FEDERHEN, S.; GEER, L. Y.; KAPUSTIN, Y.; KHOVAYKO, O.; LANDSMAN, D.; LIPMAN, D. J.; MADDEN, T. L.; MAGLOTT, D. R.; OSTELL, J.; MILLER, V.; PRUITT, K. D.; SCHULER, G. D.; SEQUEIRA, E.; SHERRY, S. T.; SIROTKIN, K.; SOUVOROV, A.; STARCHENKO, G.; TATUSOV, R. L.; TATUSOVA, T. A.; WAGNER, L.; YASCHENKO, E. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, v. 35, p. 5-12, 2007.

ZANINI, G.M.; CARVALHO, L. J. M.; BRAHIMI, K.; SOUZA-PASSOS, L. F.; GUIMARÃES, J.; MACHADO, E. S.; BIANCO-JUNIOR, C.; RICCIO, E. K. P.; SOUSA, M. A.; ALECRIM, M. G. C.; LEITE, N.; DRUILHE, P.; DANIEL-RIBEIRO,

C.T. Sera of patients with systemic lupus erythematosus react with plasmodial antigens and can inhibit the *in vitro* growth of *Plasmodium falciparum*. **Autoimmunity**, v. 42, n. 6, p. 545-552, 2009.

ZHANG, C.; ANDERSON, A.; DELISI, C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *J Mol Biol*, v. 281, p. 929-47, 1998.

APÊNDICE

7. Apêndice

Abaixo estão apresentados parte dos resultados dos alinhamentos dos autoantígenos que obtiveram similaridade com os referidos epítomos de microrganismos.

60459 - M2 protein [influenza A virus (A/WSN/1933 (H1N1))]

	Score (bits)	E Value
emb CAA80541.1 M130 antigen [Homo sapiens]	22	7.9
emb CAA80542.1 M130 antigen cytoplasmic variant 1 [Homo sapiens]	22	7.9
emb CAA80543.1 M130 antigen cytoplasmic variant 2 [Homo sapiens]	22	7.9
emb CAA80544.1 M130 antigen extracellular variant [Homo sapiens]	22	7.9
gb EAW88662.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88663.1 CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.9
gb EAW88664.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88665.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88666.1 CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.9
ref NP_004235.3 CD163 antigen isoform a [Homo sapiens]	22	7.9
ref NP_981961.1 CD163 antigen isoform b [Homo sapiens]	22	7.9

130523 - Theiler's encephalomyelitis virus (STRAIN BEAN 8386)

	Score (bits)	E Value
emb CAA59173.1 leucocyte antigen CD97 [Homo sapiens]	22	8.4
gb EAW84412.1 CD97 antigen, isoform CRA_a [Homo sapiens]	22	8.4
gb EAW84413.1 CD97 antigen, isoform CRA_b [Homo sapiens]	22	8.4
ref NP_001020331.1 CD97 antigen isoform 3 precursor [Homo sapiens]	22	8.4
ref NP_001775.2 CD97 antigen isoform 2 precursor [Homo sapiens]	22	8.4
ref NP_510966.1 CD97 antigen isoform 1 precursor [Homo sapiens]	22	8.4

153707 - M-like protein [Streptococcus pyogenes]

	Score (bits)	E Value
dbj BAA05025.1 golgi antigen gcp372 [Homo sapiens]	23	3.7
ref NP_996769.2 sarcoma antigen NY-SAR-41 [Homo sapiens]	23	4.8
gb AAM44457.1 AF273054_1 CTCL tumor antigen HD-CL-01 [Homo sapiens]	23	6.2
ref NP_060487.2 CTCL tumor antigen L14-2 [Homo sapiens]	23	6.2
gb EAW97471.1 early endosome antigen 1, 162kD, isoform CRA_a [H...]	22	8.1
gb EAW97472.1 early endosome antigen 1, 162kD, isoform CRA_b [H...]	22	8.1
ref NP_003557.2 early endosome antigen 1, 162kD [Homo sapiens]	22	8.1

160167 - circumsporozoite protein [Plasmodium vivax]

	Score (bits)	E Value
gb EAX02017.1 aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	25	0.70
gb EAX02018.1 aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	25	0.70

gb EAX02019.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	25	0.70
gb EAX02020.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	25	0.70
gb EAX02021.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	25	0.70

160537 - surface protein [Plasmodium falciparum]

		Score	E
		(bits)	Value
gb AAC18036.1	antigen NY-CO-1 [Homo sapiens]	23	4.8
gb EAW65742.1	serologically defined colon cancer antigen 1, iso...	23	4.8
gb EAW65743.1	serologically defined colon cancer antigen 1, iso...	23	4.8
gb EAW65744.1	serologically defined colon cancer antigen 1, iso...	23	4.8
gb EAW65746.1	serologically defined colon cancer antigen 1, iso...	23	4.8
gb EAW65747.1	serologically defined colon cancer antigen 1, iso...	23	4.8
ref NP_004704.2	serologically defined colon cancer antigen 1 [H...	23	4.8

280510 - antigen PV9 - Plasmodium vivax (fragment)

		Score	E
		(bits)	Value
gb AAI32872.1	Cutaneous T-cell lymphoma-associated antigen 1 [H...	24	2.8
ref NP_758441.2	cutaneous T-cell lymphoma-associated antigen 1 ...	24	2.8

441276 - BoNT/G [Clostridium botulinum]

		Score	E
		(bits)	Value
gb AAL37897.1	AF443295_1 tumor antigen MAGE-N [Homo sapiens]	24	1.0
gb AAN62752.1	MAGE-1 protein; melanoma antigen family A member ...	24	1.0
ref NP_004979.3	melanoma antigen family A, 1 [Homo sapiens]	24	1.0
gb AAP88791.1	melanoma antigen, family A, 1 (directs expression...	24	1.0
gb AAH17555.1	Melanoma antigen family A, 1 (directs expression ...	24	1.0
gb EAW72885.1	melanoma antigen family A, 1 (directs expression ...	24	1.0
gb AAH04105.1	Melanoma antigen family A, 10 [Homo sapiens]	21	6.6
gb EAW99421.1	melanoma antigen family A, 10, isoform CRA_a [Hom...	21	6.6
gb EAW99422.1	melanoma antigen family A, 10, isoform CRA_a [Hom...	21	6.6
ref NP_066386.1	melanoma antigen family A, 10 [Homo sapiens]	21	6.6
ref NP_001011543.1	melanoma antigen family A, 10 [Homo sapiens]	21	6.6
gb AAH50588.1	Melanoma antigen family E, 1 [Homo sapiens]	21	8.6
gb EAW98620.1	melanoma antigen family E, 1 [Homo sapiens]	21	8.6
emb CAI42970.1	melanoma antigen family E, 1 [Homo sapiens]	21	8.6
ref NP_065983.1	melanoma antigen family E, 1 [Homo sapiens]	21	8.6

495518 - circumsporozoite protein [Plasmodium vivax-like sp.]

		Score	E
		(bits)	Value
gb AAH10056.1	Melanoma antigen family F, 1 [Homo sapiens]	27	0.34
gb EAW78235.1	melanoma antigen family F, 1 [Homo sapiens]	27	0.34
ref NP_071432.2	melanoma antigen family F, 1 [Homo sapiens]	27	0.34
gb AAK84667.1	hepatocellular carcinoma-associated antigen [Homo...	23	4.9
gb AAN76720.1	hepatocellular carcinoma-associated antigen HCA10...	23	4.9
gb EAW52447.1	small nuclear ribonucleoprotein 70kDa polypeptide...	23	6.4
gb EAW52453.1	small nuclear ribonucleoprotein 70kDa polypeptide...	23	6.4

1899035 - Human herpesvirus 8

		Score	E
		(bits)	Value

gb EAX02017.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	22	8.6
gb EAX02018.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	22	8.6
gb EAX02019.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	22	8.6
gb EAX02020.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	22	8.6
gb EAX02021.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, 1...	22	8.6

2808520 - surface antigen protein [Leishmania infantum]

		Score (bits)	E Value
gb AAD31421.1	AF124440_1 MAGE tumor antigen D1 [Homo sapiens]	25	1.1
gb AAQ14483.1	AF300328_1 MAGE tumor antigen CCF [Homo sapiens]	25	1.1
gb AAH32473.1	Melanoma antigen family D, 1 [Homo sapiens]	25	1.1
gb AAH14070.1	Melanoma antigen family D, 1 [Homo sapiens]	25	1.1
gb EAW62896.1	melanoma antigen family D, 1, isoform CRA_a [Homo...	25	1.1
gb EAW62897.1	melanoma antigen family D, 1, isoform CRA_a [Homo...	25	1.1
emb CAH71560.1	melanoma antigen family D, 1 [Homo sapiens]	25	1.1
emb CAH71561.1	melanoma antigen family D, 1 [Homo sapiens]	25	1.1
ref NP_008917.3	melanoma antigen family D, 1 isoform b [Homo sa...	25	1.1
ref NP_001005332.1	melanoma antigen family D, 1 isoform b [Homo...	25	1.1
ref NP_001005333.1	melanoma antigen family D, 1 isoform a [Homo...	25	1.1
gb AAC83816.1	monocyte antigen CD14 precursor [Homo sapiens]	22	6.9
gb AAP35995.1	CD14 antigen [Homo sapiens]	22	6.9
gb EAW62037.1	CD14 antigen [Homo sapiens]	22	6.9
ref NP_000582.1	CD14 antigen precursor [Homo sapiens]	22	6.9
ref NP_001035110.1	CD14 antigen precursor [Homo sapiens]	22	6.9

3551268 - pol protein [Hepatitis B virus]

		Score (bits)	E Value
gb AAN62752.1	MAGE-1 protein; melanoma antigen family A member ...	22	3.0
ref NP_004979.3	melanoma antigen family A, 1 [Homo sapiens]	22	3.0
gb ABW06861.1	melanoma antigen family A 1 [Homo sapiens]	22	3.0
gb AAP88791.1	melanoma antigen, family A, 1 (directs expression...	22	3.0
gb AAH17555.1	Melanoma antigen family A, 1 (directs expression ...	22	3.0
gb EAW72885.1	melanoma antigen family A, 1 (directs expression ...	22	3.0

4099270 - merozoite surface protein-4 [Plasmodium falciparum]

		Score (bits)	E Value
gb EAW74511.1	CD248 antigen, endosialin [Homo sapiens]	28	0.20
gb EAX10171.1	CD93 antigen [Homo sapiens]	23	4.8
ref NP_036204.2	CD93 antigen precursor [Homo sapiens]	23	4.8
gb AAL62061.1	AF400226_1 bullous pemphigoid antigen 1 eA [Homo s...	23	6.3

4972899 - gag polyprotein [Human T-lymphotropic virus 2]

		Score (bits)	E Value
gb AAH26194.1	Cancer antigen 1 [Homo sapiens]	27	0.26
gb AAH33545.1	Cancer antigen 1 [Homo sapiens]	27	0.26
emb CAI16451.1	cancer antigen 1 [Homo sapiens]	27	0.26
ref NP_995586.1	cancer antigen 1 [Homo sapiens]	27	0.26
gb AAG34906.1	AF273046_1 CTCL tumor antigen se20-4 [Homo sapiens]	27	0.44
gb AAL99920.1	AF432213_1 CLL-associated antigen KW-13 [Homo sapi...	25	1.3
gb EAX11796.1	cytosolic ovarian carcinoma antigen 1, isoform CR...	24	2.8
gb EAX11797.1	cytosolic ovarian carcinoma antigen 1, isoform CR...	24	2.8
emb CAH71675.1	cytosolic ovarian carcinoma antigen 1 [Homo sapi...	24	2.8
emb CAH71676.1	cytosolic ovarian carcinoma antigen 1 [Homo sapi...	24	2.8
emb CAI43112.1	cytosolic ovarian carcinoma antigen 1 [Homo sapi...	24	2.8
emb CAI43113.1	cytosolic ovarian carcinoma antigen 1 [Homo sapi...	24	2.8
emb CAI43114.1	cytosolic ovarian carcinoma antigen 1 [Homo sapi...	24	2.8
gb EAW89039.1	fetal Alzheimer antigen, isoform CRA_b [Homo sapi...	23	3.7

dbj BAD92374.1	NY-REN-8 antigen variant [Homo sapiens]	23	4.9
emb CAA36988.1	CD22 antigen [Homo sapiens]	22	8.3
dbj BAF32756.1	CD22 antigen [Homo sapiens]	22	8.3
dbj BAF32757.1	CD22 antigen [Homo sapiens]	22	8.3
dbj BAF32758.1	CD22 antigen [Homo sapiens]	22	8.3

6683941 - rhopty-associated protein 1 [Plasmodium falciparum]

		Score (bits)	E Value
gb AAN04658.1	minor histocompatibility antigen HA-1 [Homo sapiens]	24	2.2
ref NP_036424.2	minor histocompatibility antigen HA-1 [Homo sap...]	24	2.2
gb AAD15273.2	T200 leukocyte common antigen precursor [Homo sap...]	23	6.3

7108613 - S antigen [Hepatitis B virus]

		Score (bits)	E Value
ref XP_945853.1	PREDICTED: similar to Galectin-3-binding protei...	23	6.4

8307811 - matrix protein 2 [Influenza A virus (A/Hong Kong/483/1997(H5N1))]

		Score (bits)	E Value
emb CAA80541.1	M130 antigen [Homo sapiens]	22	7.9
emb CAA80542.1	M130 antigen cytoplasmic variant 1 [Homo sapiens]	22	7.9
emb CAA80543.1	M130 antigen cytoplasmic variant 2 [Homo sapiens]	22	7.9
emb CAA80544.1	M130 antigen extracellular variant [Homo sapiens]	22	7.9
gb EAW88662.1	CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88663.1	CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.9
gb EAW88664.1	CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88665.1	CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.9
gb EAW88666.1	CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.9
ref NP_004235.3	CD163 antigen isoform a [Homo sapiens]	22	7.9
ref NP_981961.1	CD163 antigen isoform b [Homo sapiens]	22	7.9

12004303 - circumsporozoite protein [Plasmodium vivax]

		E (bits)	Score Value
gb EAX02017.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, l...	26	0.57
gb EAX02018.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, l...	26	0.57
gb EAX02019.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, l...	26	0.57
gb EAX02020.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, l...	26	0.57
gb EAX02021.1	aggrecan 1 (chondroitin sulfate proteoglycan 1, l...	26	0.57
gb EAW91801.1	sperm associated antigen 1, isoform CRA_b [Homo s...	25	1.6
gb EAW91802.1	sperm associated antigen 1, isoform CRA_c [Homo s...	25	1.6
gb EAW91803.1	sperm associated antigen 1, isoform CRA_d [Homo s...	25	1.6
gb EAW91804.1	sperm associated antigen 1, isoform CRA_d [Homo s...	25	1.6
gb EAW91806.1	sperm associated antigen 1, isoform CRA_b [Homo s...	25	1.6
ref NP_003105.2	sperm associated antigen 1 [Homo sapiens]	25	1.6
ref NP_757367.1	sperm associated antigen 1 [Homo sapiens]	25	1.6
gb AAG48260.1	AF308292_1 serologically defined breast cancer ant...	23	3.7

14162008 - phosphoprotein [Rabies virus]

		Score (bits)	E Value
gb AAC18041.1	antigen NY-CO-10 [Homo sapiens]	22	8.4
gb EAW51364.1	serologically defined colon cancer antigen 10, is...	22	8.4
gb EAW51365.1	serologically defined colon cancer antigen 10, is...	22	8.4

ref|NP_005860.2| serologically defined colon cancer antigen 10 [...] 22 8.4

44903266 - UL40 [Human herpesvirus 5]

	Score (bits)	E Value
gb AAC32742.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAC32743.1 MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32611.1 MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32610.1 MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32612.1 MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32613.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAD11469.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAD50823.1 AF168611_1 HLA-C class I antigen [Homo sapiens]	21	8.6
gb AAA59687.1 lymphocyte antigen [Homo sapiens]	21	8.6
gb AAM76870.1 AF196489_1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAL03994.1 AF405691_1 MHC class I antigen heavy chain [Homo s...]	21	8.6
gb AAS45436.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAS98883.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAR99590.1 MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAX18631.1 MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAX18632.1 MHC class I antigen precursor [Homo sapiens]	21	8.6
emb CAA09340.1 HLA class I antigen [Homo sapiens]	21	8.6
gb AAX94769.1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAT39989.1 MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAZ67360.1 MHC class I antigen precursor [Homo sapiens]	21	8.6
emb CAA25190.1 class I histocompatibility antigen HLA-CW3 [Homo...]	21	8.6
emb CAB38942.1 human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB41889.2 human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB65547.1 MHC class I antigen [Homo sapiens]	21	8.6
emb CAB71940.1 human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB95011.1 human leucocyte antigen Cw [Homo sapiens]	21	8.6
emb CAC05372.1 human leucocyte antigen Cw [Homo sapiens]	21	8.6
emb CAC18746.1 MHC class I antigen [Homo sapiens]	21	8.6
emb CAC19018.1 MHC class I antigen [Homo sapiens]	21	8.6

48525483 - K1 glycoprotein [Human herpesvirus 8]

	Score (bits)	E Value
gb AAS73157.1 KIR antigen 2DS4 [Homo sapiens]	28	0.20
gb AAS73158.1 KIR antigen 2DS4 [Homo sapiens]	28	0.20
gb AAS73159.1 KIR antigen 2DS4 [Homo sapiens]	28	0.20
gb AAS73161.1 KIR antigen 3DL2 [Homo sapiens]	28	0.20
gb AAS73162.1 KIR antigen 3DL2 [Homo sapiens]	28	0.20
gb ABC03226.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03227.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03228.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03229.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03230.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03231.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03232.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03233.1 KIR antigen [Homo sapiens]	28	0.20
gb ABC03234.1 KIR antigen [Homo sapiens]	28	0.20
gb AAF06354.1 AF200348_1 melanoma-associated antigen MG50 [Homo ...]	27	0.26
gb AAS73160.1 KIR antigen 3DL1 [Homo sapiens]	25	1.7

63081091 - M protein [Streptococcus pyogenes]

	Score (bits)	E Value
dbj BAA05025.1 golgi antigen gcp372 [Homo sapiens]	23	3.7
ref NP_996769.2 sarcoma antigen NY-SAR-41 [Homo sapiens]	23	4.8
gb AAM44457.1 AF273054_1 CTCL tumor antigen HD-CL-01 [Homo sapiens]	23	6.2

ref NP_060487.2	CTCL tumor antigen L14-2 [Homo sapiens]	23	6.2
gb EAW97471.1	early endosome antigen 1, 162kD, isoform CRA_a [H...	22	8.1
gb EAW97472.1	early endosome antigen 1, 162kD, isoform CRA_b [H...	22	8.1
ref NP_003557.2	early endosome antigen 1, 162kD [Homo sapiens]	22	8.1

82703961 - BOLF1 [Human herpesvirus 4]

		Score	E
		(bits)	Value
gb AAL35222.1	AF439338_1 MHC class II antigen [Homo sapiens]	29	0.032
gb AAP78473.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAQ88029.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAQ88030.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAQ88031.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAV97949.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAU87988.1	MHC class II antigen [Homo sapiens]	29	0.032
gb ABB05231.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAC27418.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAD47822.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAD89801.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAJ76977.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAL15002.1	MHC class II antigen [Homo sapiens]	29	0.032
emb CAL64838.1	MHC class II antigen [Homo sapiens]	29	0.032
gb AAC64605.1	MHC class II antigen [Homo sapiens]	27	0.16
gb AAF28315.1	AF195245_1 MHC class II antigen [Homo sapiens]	27	0.16
gb AAL77085.1	MHC class II antigen [Homo sapiens]	27	0.16
gb AAM09472.1	AF479569_1 MHC class II antigen [Homo sapiens]	27	0.16
gb AAQ88019.1	MHC class II antigen [Homo sapiens]	27	0.16
gb AAQ88020.1	MHC class II antigen [Homo sapiens]	27	0.16
gb AAQ88021.1	MHC class II antigen [Homo sapiens]	27	0.16
gb AAQ88022.1	MHC class II antigen [Homo sapiens]	27	0.16

87047829 - glycoprotein [Human herpesvirus 5]

		Score	E
		(bits)	Value
gb AAC32742.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAC32743.1	MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32611.1	MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32610.1	MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32612.1	MHC class I antigen [Homo sapiens]	21	8.6
dbj BAA32613.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAD11469.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAD50823.1	AF168611_1 HLA-C class I antigen [Homo sapiens]	21	8.6
gb AAA59687.1	lymphocyte antigen [Homo sapiens]	21	8.6
gb AAM76870.1	AF196489_1 MHC class I antigen [Homo sapiens]	21	8.6
gb AAL03994.1	AF405691_1 MHC class I antigen heavy chain [Homo s...	21	8.6
gb AAS45436.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAS98883.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAR99590.1	MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAX18631.1	MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAX18632.1	MHC class I antigen precursor [Homo sapiens]	21	8.6
emb CAA09340.1	HLA class I antigen [Homo sapiens]	21	8.6
gb AAX94769.1	MHC class I antigen [Homo sapiens]	21	8.6
gb AAT39989.1	MHC class I antigen precursor [Homo sapiens]	21	8.6
gb AAZ67360.1	MHC class I antigen precursor [Homo sapiens]	21	8.6
emb CAA25190.1	class I histocompatibility antigen HLA-CW3 [Homo...	21	8.6
emb CAB38942.1	human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB41889.2	human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB65547.1	MHC class I antigen [Homo sapiens]	21	8.6
emb CAB71940.1	human leucocyte antigen C [Homo sapiens]	21	8.6
emb CAB95011.1	human leucocyte antigen Cw [Homo sapiens]	21	8.6
emb CAC05372.1	human leucocyte antigen Cw [Homo sapiens]	21	8.6

107784976 - circumsporozoite protein [Plasmodium falciparum]

	Score	E
	(bits)	Value
dbj BAD96736.1 squamous cell carcinoma antigen recognized by T ...	24	2.9
gb AAP35283.1 squamous cell carcinoma antigen recognised by T c...	23	3.7
gb EAW74474.1 squamous cell carcinoma antigen recognised by T c...	23	3.7
gb EAW74475.1 squamous cell carcinoma antigen recognised by T c...	23	3.7
gb AAH01058.1 Squamous cell carcinoma antigen recognized by T c...	23	3.7
ref NP_005137.1 squamous cell carcinoma antigen recognized by T...	23	3.7
gb AAF36964.1 AF227899_1 breast carcinoma-associated antigen iso...	23	6.4

P10920 - RecName: Full=Matrix protein 2; AltName: Full=Proton channel protein M2. [Influenza virus]

	Score	E
	(bits)	Value
emb CAA80541.1 M130 antigen [Homo sapiens]	22	7.6
emb CAA80542.1 M130 antigen cytoplasmic variant 1 [Homo sapiens]	22	7.6
emb CAA80543.1 M130 antigen cytoplasmic variant 2 [Homo sapiens]	22	7.6
emb CAA80544.1 M130 antigen extracellular variant [Homo sapiens]	22	7.6
gb EAW88662.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.6
gb EAW88663.1 CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.6
gb EAW88664.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.6
gb EAW88665.1 CD163 antigen, isoform CRA_a [Homo sapiens]	22	7.6
gb EAW88666.1 CD163 antigen, isoform CRA_b [Homo sapiens]	22	7.6
ref NP_004235.3 CD163 antigen isoform a [Homo sapiens]	22	7.6
ref NP_981961.1 CD163 antigen isoform b [Homo sapiens]	22	7.6

Q25893 - Liver stage antigen. [Plasmodium falciparum]

	Score	E
	(bits)	Value
:		
gb AAT44534.1 rhabdomyosarcoma antigen MU-RMS-40.7B [Homo sapiens]	27	0.44
gb AAX47276.1 sperm associated antigen 9 transcript variant 1 [...]	26	0.57
gb EAW94569.1 sperm associated antigen 9, isoform CRA_a [Homo s...	26	0.57
gb EAW94570.1 sperm associated antigen 9, isoform CRA_b [Homo s...	26	0.57
gb EAW94571.1 sperm associated antigen 9, isoform CRA_c [Homo s...	26	0.57
gb EAW94572.1 sperm associated antigen 9, isoform CRA_d [Homo s...	26	0.57
gb EAW94574.1 sperm associated antigen 9, isoform CRA_f [Homo s...	26	0.57
ref NP_003962.3 sperm associated antigen 9 [Homo sapiens]	26	0.57
gb AAC18048.1 antigen NY-CO-37 [Homo sapiens]	25	1.7
gb AAC18049.1 antigen NY-CO-38 [Homo sapiens]	25	1.7
dbj BAA81740.1 autoimmune enteropathy-related antigen AIE-75 [H...	25	1.7
gb AAN07913.1 medulloblastoma antigen MU-MB-20.240 [Homo sapiens]	24	2.2
gb AAT44400.1 antigen MU-RMS-40.16A [Homo sapiens]	24	2.8
gb AAT66048.1 rhabdomyosarcoma antigen MU-RMS-40.19 [Homo sapiens]	24	2.8
gb AAD42862.1 AF155096_1 NY-REN-6 antigen [Homo sapiens]	23	3.7
gb AAO65167.1 sarcoma antigen NY-SAR-24 [Homo sapiens]	23	3.7
ref NP_996769.2 sarcoma antigen NY-SAR-41 [Homo sapiens]	23	3.7
gb AAG34903.1 AF273043_1 CTCL tumor antigen se2-1 [Homo sapiens]	23	4.9
gb AAQ93059.1 antigen MLAA-20 [Homo sapiens]	23	4.9
gb ABB04050.1 rhabdomyosarcoma antigen MU-RMS-40.12 [Homo sapiens]	23	4.9
emb CAI64497.1 tumor rejection antigen (gp96) 1 [Homo sapiens]	23	6.3
dbj BAD92771.1 tumor rejection antigen (gp96) 1 variant [Homo s...	23	6.3
ref NP_003290.1 tumor rejection antigen (gp96) 1 [Homo sapiens]	23	6.3
dbj BAB16440.1 sperm antigen [Homo sapiens]	22	8.3
gb AAG48253.1 AF308285_1 serologically defined breast cancer ant...	22	8.3
gb AAH21123.2 Sperm antigen with calponin homology and coiled-c...	22	8.3