Universidade de São Paulo

Instituto de Química

**Desenvolvimento e automatização de um método teórico
para a avaliação quantitativa da seletividade de proteínas do
MHC por diferentes antígenos**

Jackson Gois da Silva

Dissertação apresentada ao
Instituto de Química da USP
para a obtenção do Título de
Mestre em Bioquímica.

Orientadora: Prof^a Dr^a Wladia Viviani.

São Paulo, 25 de junho de 2004

**"Desenvolvimento e automatização de um método
teórico para a avaliação quantitativa da seletividade
de proteínas do MHC por diferentes antígenos"**

JACKSON GÔIS DA SILVA

Dissertação de Mestrado submetida ao Instituto de
Química da Universidade de São Paulo como parte dos
requisitos necessários à obtenção do grau de Mestre em
Ciências - Área: Bioquímica.

Reprovado por:

Profa. Dra. WLADIA VIVIANI
IQ - USP
(Orientadora e Presidente)

Aprovado por:

Prof. Dr. SHAKER CHUCK FARAH
IQ - USP

Aprovado por:

Prof. Dr. MARCO SERRÓU DO AMARAL
UFMS

SÃO PAULO
25/06/2004

Agradecimentos

A Deus, pela vida.

À minha Esposa, Eliane, por fazer a vida valer a pena.

Aos meus Pais, pelo apoio e exemplo.

Aos amigos, que tornam alguns momentos da vida inesquecíveis.

À Prof^a Dr^a Aline Maria da Silva, pelo apoio e conselhos, sem os quais este trabalho não teria sido concluído.

Ao Instituto de Química da USP, em seus professores e funcionários, por auxiliarem na realização deste trabalho.

Ao CNPq, pelo suporte financeiro.

Sumário	Página
Agradecimentos	2
Resumo	4
Abstract	5
Lista de Abreviaturas	6
Lista de Tabelas	6
Lista de Figuras	7
1 - Introdução	8
1.1 - O Complexo Maior de Histocompatibilidade	8
1.1.1 - Estrutura, propriedades e reconhecimento de peptídeos antigênicos.	8
1.1.2 - Aplicações terapêuticas do estudo de moléculas do MHC.	12
-Vacinas.	12
-Doenças auto-imunes.	13
1.1.3 - Métodos de Predição da Seletividade de moléculas do MHC.	14
-Estágio atual do conhecimento.	14
-O método MHCcalc.	16
-Preparo do arquivo de entrada.	17
1.2 - Cálculos empíricos de Energia Potencial: uma Revisão.	19
1.2.1 - Determinação da Energia de uma Estrutura	21
-Campo de Força	22
1.2.2 - Minimização de Energia e o problema dos Múltiplos Mínimos.	28
-Algoritmos de Minimização Utilizados	30
2 - Objetivos	35
3 - Metodologia	35
4 - Resultados e Discussão	36
4.1 - Desenvolvimento e implementação de MHCcalc.	36
4.2 - Sistema Molecular Estudado.	38
4.3 - Teste de Aplicação do Programa MHCcalc na proteína HLA-DR1.	39
4.4 - Descrição de HLA-DR1	40
4.5 - Posicionamento inicial das cadeias laterais no peptídeo inicial.	42
4.6 - Estudo do bolsão 1.	42
5 - Conclusão	64
6 - Referências Bibliográficas	65
7 - Curriculum Vitae	73

Resumo

Realizamos neste trabalho a automatização da nova metodologia MHCcalc, desenvolvida anteriormente em nosso laboratório, que teve como resultado o programa homônimo escrito em linguagem C para ambiente GNU/Linux, o qual avalia quantitativamente a seletividade de uma determinada proteína MHC. Esta avaliação quantitativa possibilita o estabelecimento de uma escala de preferência dos resíduos de ocorrência natural para cada uma das posições de interação da fenda de apresentação do MHC; por permutação destes resíduos, podem-se derivar regras de composição para peptídeos reconhecíveis em cada alelo estudado, inclusive na ausência de dados experimentais. A metodologia desenvolvida em nosso laboratório se baseia na avaliação da seletividade de cada bolsão independentemente, através da energia de interação do mesmo com cada aminoácido de ocorrência natural.

O programa MHCcalc utiliza o pacote de programas THORMM [Moret, 1998] de modelagem molecular para otimização geométrica das estruturas descritas, e tem como entrada unicamente um arquivo de coordenadas em formato PDB contendo as coordenadas de um complexo MHC/peptídeo, tendo como saída uma tabela de dados contendo os resíduos de ocorrência natural em ordem de preferência para cada bolsão. Testamos o programa MHCcalc para o complexo formado pela molécula HLA-DR1 (DRA DRB1*0101) e o peptídeo de Hemaglutinina, obtido no *Brookhaven Protein Data Bank* com a entrada 1DLH, sendo este sistema escolhido por ser altamente estudado e com abundância de dados experimentais e teóricos para comparação de resultados. Até o presente momento obtivemos os dados referentes ao bolsão 1, os quais estão em pleno acordo com a literatura.

Abstract

In this work we automated the new methodology MHCcalc, developed previously in our laboratory, which resulted in the computational program with the same name, written in C language to GNU/Linux environment. The program MHCcalc evaluates quantitatively the selectivity of a given MHC protein. This quantitative evaluation allows the establishment of a preference scale of the naturally occurring residues for each pocket of a MHC molecule. From such a study it may be derived composition rules to peptides recognizable in each allele studied, even in the absence of experimental data. The methodology developed in our laboratory is based on the selectivity evaluation of each pocket independently, through its interaction energy with each naturally occurring amino acid.

The program MHCcalc uses the molecular modeling package THORMM [Moret, 1998] to optimize the geometry of the structures described below, and needs as only entry the PDB like file with the coordinates of the MHC/peptide complex. The program finishes printing out a data file containing the naturally occurring residues in preference order for each pocket and the data used to order these residues. We tested the program MHCcalc to the complex molecule HLA-DR1 (DRA DRB1*0101) with the Hemagglutinin peptide, obtained at the *BrookHaven Protein Data Bank* (1DLH entry). This system was chosen because it has been highly studied and so offers abundant experimental and theoretical data to compare our results. We obtained data referring to pocket 1 so far, which is in full agreement with the literature.

Lista de Abreviaturas

Abreviatura	Significado
APC	Célula apresentadora de Antígeno (<i>Antigen Presenting Cell</i>)
MHC	Complexo Maior de Histocompatibilidade (<i>Major Histocompatibility Complex</i>)
HLA	MHC humano (<i>Human Leucocitary Associated</i>)
TCR	Receptor de Célula T (<i>T Cell Receptor</i>)
DR1	DRA DRB1*0101
HIV	Vírus da Imunodeficiência Humana (<i>Human Immunodeficiency Virus</i>)
IDDM	Diabetes Mellitus Insulino Dependente (<i>Insulin Dependent Diabetes Mellitus</i>)
PDB	<i>Protein Data Bank</i>
QSAR	Relações quantitativas entre estrutura e atividade (<i>Quantitative Structure-Activity Relationships</i>)
RMN	Ressonância Magnética Nuclear

Lista de Tabelas

Tabela	Página
<u>Tabela 1</u> – Comparação entre a ordem de preferência de cadeias laterais de aminoácidos de ocorrência natural para o bolsão 1 de DR1 obtida neste trabalho e a ordem de preferência de outros autores para o mesmo bolsão deste MHC.	43
<u>Tabela 2</u> – Dados obtidos com a utilização do programa MHCcalc, para o bolsão 1 de DR1, no mesmo formato em que são devolvidos pelo programa.	44
<u>Tabela 3</u> – Distâncias entre LYS do peptídeo complexado ao MHC e resíduos com cargas negativas das proximidades.	45
<u>Tabela 4</u> – Distâncias entre ASP do peptídeo complexado ao MHC e resíduos com cargas positivas das proximidades.	51
<u>Tabela 5</u> – Distância entre centrômeros de cadeias laterais de resíduos aromáticos do bolsão 1 e os centrômeros das cadeias laterais aromáticas dos peptídeos com resíduos de TYR, TRP e PHE no bolsão 1.	53

Lista de Figuras

Figura	Página
Figura 1 – Representação de MHC de classe 1.	9
Figura 2 – Representação de MHC de classe 2.	9
Figura 3 – Representação do complexo MHC/peptídeo/TCR.	9
Figura 4 – Representação do peptídeo entre MHC e TCR.	9
Figura 5 – Representação das cadeias laterais entre MHC e TCR.	9
Figura 6 – Dímero de DR1 conforme obtido do Brookhaven Protein Data Bank.	18
Figura 7 – Pontos de secção no monômero DR1.	18
Figura 8 – Representação dos domínios α_1 e β_1 utilizados neste trabalho.	18
Figura 9 – Superfície acessível ao solvente aquoso de DR1/hemaglutinina.	18
Figura 10 – Superfície acessível ao solvente aquoso de DR1/polialanina.	18
Figura 11 – Representação gráfica da equação 11.	29
Figura 12 – Representação gráfica de minimização com linha de procura.	31
Figura 13 – Representação gráfica de minimização sem linha de procura.	33
Figura 14 – Cadeias laterais que formam os bolsões de DR1.	41
Figura 15 – Ligação de enxofre intramolecular em DR1.	41
Figura 16 – Posicionamento das cadeias laterais após minimização.	41
Figura 17 – Posicionamento das cadeias laterais antes da minimização.	41
Figura 18 – Cadeias laterais do bolsão 1.	46
Figura 19 – Posicionamento da cadeia lateral de lisina no bolsão 1.	46
Figura 20 – Representação da superfície da cadeia α do bolsão 1.	46
Figura 21 – Representação da superfície da cadeia β do bolsão 1.	46
Figura 22 – Comparação dos posicionamentos de lisina e alanina.	46
Figura 23 – Posicionamento da cadeia lateral de leucina no bolsão 1.	50
Figura 24 – Posicionamento da cadeia lateral de aspartato no bolsão 1.	50
Figura 25 – Comparação do posicionamento de aspartato e leucina.	50
Figura 26 – Comparação do posicionamento de aspartato e asparagina.	50
Figura 27 – Conformação do peptídeo de partida.	54
Figura 28 – Comparação dos posicionamentos de aromáticos.	54
Figura 29 – Comparação dos posicionamentos de histidina e fenilalanina.	54
Figura 30 – Comparação do posicionamento de isoleucina.	56
Figura 31 – Comparação do posicionamento de valina.	56
Figura 32 – Posicionamento da cadeia lateral de treonina no bolsão 1.	56
Figura 33 – Comparação do posicionamento de treonina e alanina.	56
Figura 34 – Posicionamento da cadeia lateral de metionina no bolsão 1.	60
Figura 35 – Posicionamento da cadeia lateral de asparagina no bolsão 1.	60
Figura 36 – Posicionamento da cadeia lateral de cisteína no bolsão 1.	60
Figura 37 – Posicionamento da cadeia lateral de serina no bolsão 1.	60
Figura 38 – Comparação dos posicionamentos de serina e cisteína.	60
Figura 39 – Posicionamento da cadeia lateral de prolina no bolsão 1.	62
Figura 40 – Comparação do posicionamento de lisina e arginina.	62
Figura 41 – Representação do volume ocupado por lisina.	62
Figura 42 – Representação do volume ocupado por arginina.	62
Figura 43 – Posicionamento da cadeia lateral de glutamato no bolsão 1.	63
Figura 44 – Comparação do posicionamento de aspartato e glutamato.	63
Figura 45 – Posicionamento da cadeia lateral de glutamina no bolsão 1.	63
Figura 46 – Comparação do posicionamento de glutamina e glutamato.	63
Figura 47 – Posicionamento da cadeia lateral de glicina no bolsão 1.	63

1 - Introdução

O sistema imunológico tem, como característica principal, a capacidade de distinguir o próprio do não-próprio. Nos mamíferos este sistema costuma ser desafiado por patógenos, tais como vírus, fungos, bactérias, protozoários, vermes e toxinas, e também por órgãos transplantados de doadores não relacionados [Peakman, 1999].

Componentes celulares patrulham constantemente o sangue, espaço intercelular e líquido cérebro-espinhal contra a invasão de microorganismos. Algumas células, como os macrófagos, têm a capacidade de reconhecer e fagocitar o agente invasor. Estas células, juntamente com células dendríticas e linfócitos B, são conhecidas como células apresentadoras de antígeno (APC – *Antigen Presenting Cell*), pois têm a capacidade de apresentar em sua superfície celular, através de complexo mecanismo de processamento em compartimentos intracelulares, peptídeos potencialmente antigênicos denominados epítomos. Estes epítomos são apresentados a receptores de linfócitos T, desencadeando processos que resultam em respostas à invasão antigênica por parte do sistema imunitário [Germain, 1993].

1.1 – O complexo maior de histocompatibilidade.

1.1.2 – Estrutura, propriedade e reconhecimento de peptídeos antigênicos.

A apresentação de epítomos na superfície de células é feita por glicoproteínas transmembrana da superfamília das imunoglobulinas, codificadas por um conjunto de genes conhecidos como Complexo Maior de Histocompatibilidade (MHC – *Major Histocompatibility Complex*). O MHC humano é conhecido como HLA (*Human Leucocitary Associated*) e os genes são encontrados no cromossomo seis (figuras 1 e 2) [Peakman, 1999].

O MHC de classe I é expresso constitutivamente na superfície externa de virtualmente todas as células nucleadas dos vertebrados, enquanto que APC's expressam em sua superfície, além do MHC de classe I, o MHC de classe II. Em condições fisiológicas o MHC de classe I apresenta antígenos endógenos processados a partir de proteínas citoplasmáticas. Em uma

Figura 1

Figura 2

Figura 3

Figura 4

Figura 5

Fig. 1 e 2. Representação de Complexos formados por um peptídeo, em verde, e pelas cadeias α e β , em azul e laranja respectivamente, de MHC de classe I (figura 1) e II (figura 2) [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

Fig. 3, 4 e 5. Representação do contato entre o peptídeo, em verde, os domínios $\alpha 2$ e $\beta 2$ do MHC, em laranja e azul respectivamente, e o Receptor de Célula T, em cinza [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

infecção viral, as proteínas virais são sintetizadas endogenamente usando a maquinaria das células do hospedeiro. Portanto peptídeos virais podem, também, ser apresentados através deste caminho endógeno. Contrariamente o MHC de classe II apresenta peptídeos exógenos, processados a partir de antígenos fagocitados por macrófagos ou internalizados por anticorpos na superfície externa de células B. Os peptídeos apresentados através deste caminho exógeno são uma variedade de moléculas provenientes do meio externo, como por exemplo, proteínas derivadas de bactérias, do meio plasmático ou de superfícies celulares. O processo de endocitose e a apresentação de antígenos estranhos às células T são os principais eventos de iniciação das respostas imunológicas [Demotz, 1999].

A característica mais marcante destas proteínas é seu alto grau de polimorfismo entre indivíduos da mesma espécie: trata-se dos genes mais polimórficos em vertebrados superiores. Uma consequência disso é a baixa probabilidade de dois indivíduos apresentarem o mesmo conjunto de genes MHC [Peakman, 1999].

O tamanho típico para peptídeos ligados ao MHC de classe I consiste em um segmento de 8 a 10 resíduos. Ocasionalmente, o peptídeo precisa dobrar-se para se encaixar na fenda, fechada em ambas as extremidades (fig. 1). As moléculas da classe I são construídas em um ponto onde uma cadeia α , uma β_2 microglobulina e um peptídeo estejam disponíveis, de forma que há pouca ou nenhuma oportunidade para o peptídeo selecionado e fortemente ligado ser substituído durante o transporte para a superfície da célula, o que faz do peptídeo um componente integral da molécula da classe I. A partir de análises das características dos peptídeos eluídos das moléculas de classe I, foi possível delinear os requerimentos mínimos para a ligação do peptídeo, os quais tem sido chamados de motivos de ligação peptídica. Tais previsões têm auxiliado na identificação dos componentes críticos de uma proteína que leva a uma resposta antimalárica em indivíduos com uma determinada molécula de HLA, e tal conhecimento deverá auxiliar no planejamento de uma vacina [Meister, 1995].

As características da ligação do peptídeo das moléculas da classe II estão bem menos definidas do que aquelas associadas à classe I. Em média os peptídeos encontrados nas fendas das moléculas da classe II possuem entre 10 e 34 aminoácidos. Isto ocorre porque a fenda

apresentadora na molécula de classe II é aberta em ambas as extremidades, o que permite que o peptídeo se estenda nas duas direções além da fenda (fig. 2). Com isso, os peptídeos têm maior variedade em comprimento, com uma região central de 7 a 10 aminoácidos que interagem diretamente com a fenda apresentadora. Além disso, o mesmo peptídeo pode se ligar a mais de uma molécula de classe II, e a mesma molécula pode ligar diferentes peptídeos. Esta promiscuidade torna mais difícil a previsão dos motivos de ligação na fenda apresentadora [Fridkis-Hareli, 1998; Rudensky, 1991].

A região responsável pela apresentação dos antígenos é formada por uma base plana, ou ligeiramente abaulada, de fitas β antiparalelas, delimitada por duas hélices α ; o receptáculo resultante acomoda peptídeos em conformação quase estendida (hélice poliprolina tipo II) [Dessen, 1997; Ghosh, 1995; Jardetzky, 1996]. Os peptídeos que se estendem além das extremidades do receptáculo têm sua conformação apenas parcialmente ordenada [Murthy, 1997]. Ligações de hidrogênio entre regiões conservadas do receptáculo, em especial resíduos com dupla capacidade de estabelecer ligações de hidrogênio como asparagina e glutamina, e o esqueleto carbônico do peptídeo, parecem desempenhar um papel universal na ligação de peptídeos de diferentes seqüências [Stern, 1994; Sundberg, 2002], de forma que a maioria dos agrupamentos polares contata a cadeia principal do peptídeo, e a maioria dos agrupamentos hidrofóbicos contata as cadeias laterais do peptídeo [Murthy, 1997].

A interação entre o MHC de classe II e o peptídeo parece cobrir aproximadamente metade da superfície molecular deste. O receptor de célula timocitária (TCR) se ligará no complexo MHC/peptídeo, reconhecendo a face apresentada pelas hélices protéicas intercalada pelo antígeno [Chicz, 1992].

A conformação em hélice estendida do peptídeo resulta no posicionamento de algumas de suas cadeias laterais (tipicamente quatro) apontando para o interior da fenda (figs. 3, 4 e 5), para bolsões ou reentrâncias na superfície interna do receptáculo [Nelson, 1993]. Ancorando-se nestes bolsões, as cadeias laterais peptídicas garantem a seletividade da ligação (geralmente atribuída a ligações de hidrogênio de sua cadeia principal) com a proteína do MHC [Siklodi, 1998; Verreck, 1996]. As posições características de ancoragem variam para as diferentes

proteínas do MHC: tratando-se, por exemplo, dos resíduos p2, p3, p6 e p9 para o H2-K de camundongo [Voet, 1995], de p1, p4, p6 e p9 para o HLA (Human Leucocitary Antigen, proteína do MHC humano) – DR1 [Brown, 1993], e de p1, p4, p5 e p8 para HLA-DQ7 [Cunha-Neto, 1996]. Por outro lado, o tamanho e a natureza das cadeias laterais que delimitam estes bolsões definem um número restrito de peptídeos que cada molécula do MHC pode ligar. Estes peptídeos devem então possuir, nas posições correspondentes às de ancoragem, resíduos que tenham características complementares às especificidades individuais da fenda apresentadora (tais como forma, propriedades eletrostática e hidropática). Estima-se de fato que cada polimorfo de MHC – classe I possa ligar menos que 1% dos diversos peptídeos de 8-9 resíduos a que venha a ser exposto [Voet, 1995].

1.1.2 – Aplicações terapêuticas do estudo de moléculas do MHC

Vacinas

Nos últimos anos, vacinas baseadas em epítomos têm recebido considerável atenção como um meio possível de desenvolvimento de vacinas profiláticas e estratégias imunoterapêuticas. A seleção de epítomos apropriados de células B e T permitiriam ao sistema imune se focalizar em epítomos conservados de patógenos caracterizados por alta variabilidade de seqüência, tais como HIV, vírus da hepatite C e malária [Cunha-Neto, 1996; Meister, 1995].

Vacinas baseadas em epítomos também oferecem a oportunidade de incluir na vacina epítomos construídos para modular a potencia, ou por aumento da afinidade de ligação do epítomo ao MHC, ou por alteração de seus resíduos que contatam o TCR, ou ambos [Cunha-Neto, 1999].

Uma vez que epítomos determinantes apropriados tenham sido definidos, eles podem ser sorteados e lançados por vários meios, incluindo lipopeptídeos, vetores virais de envio, partículas de origem viral ou sintética, adjuvantes, lipossomos e por DNA cíclico [Peakman, 1999].

Antes que epítopos apropriados possam ser definidos, entretanto, um obstáculo maior deve ser transposto: o grau de polimorfismo extremamente alto expresso na população humana. Mais de 200 tipos de moléculas de HLA classe I e II já foram identificados. Entretanto, já foi demonstrado que os peptídeos capazes de se ligarem às moléculas HLA classe I podem ser identificados, e mais de 60 % das moléculas HLA classe I podem ser agrupadas em 4 principais supertipos, caracterizados por especificidades similares de ligação de peptídeo, também conhecido como supermotifs [Peakman, 1999].

No caso das moléculas de classe II é bem conhecido o fato de existirem peptídeos capazes de se ligarem a múltiplos tipos de HLA [Stern, 1994]. Até o presente momento, uma metodologia geral para a identificação de tais peptídeos ainda não foi desenvolvida. Isto é, pelo menos em parte, um reflexo do fato de que ensaios de ligação quantitativa de peptídeos são trabalhosos, e que um grande número de alelos devem ser considerados.

Doenças Auto-imunes

Estudos epidemiológicos indicam que certos polimorfos de MHC estão associados com maior ou menor susceptibilidade a certas doenças infecciosas e auto-imunes. Na genética de doenças auto-imunes, o locus de papel preponderante é o Complexo de Maior Histocompatibilidade. Os mecanismos desencadeantes das doenças auto-imunes são pouco claros – ainda que certos efeitos ambientais como exposição a certas infecções, e efeitos genéticos, tenham sido associados com o aparecimento das patologias. Em especial podemos citar a associação entre proteínas de classe II e doenças imunológicas: DR2, DR3, DQ2, DQ8 com IDDM [Peakman, 1999; Voet, 1995; Sanjeevi, 1995; Routsias, 1995; Auger, 1997]; ou DQ7, DQ2 com a Cardiopatia Chagásica Crônica [Cunha-Neto, 1996], entre outros.

A susceptibilidade à doença auto-imune parece estar relacionada com o reconhecimento e apresentação seletiva de auto-antígenos por parte de certos alelos de MHC, burlando o sistema de autotolerância e levando a uma reação imunológica cruzada. Cabe notar que para haver mimetismo entre dois epítopos, a identidade ou alta similaridade de sequência primária não é um dos fatos necessários [Quaratino, 1995]. Basta haver identidade entre 3 resíduos

dentre 15, acompanhada pelo TCR, para a reação cruzada ser desencadeada. Já nos casos de associação negativa entre alelos de MHC Classe II e doenças auto-imunes, o alelo protetor pode estar envolvido com seleção negativa intratímica dos TCR que ligariam o auto-antígeno e o antígeno patogênico mimético [Cunha-Neto, 1996].

A identificação precisa dos peptídeos-alvo de reconhecimento imunológico em uma doença auto-imune é primordial para estudar sua patogênese, levando a compreensão da mesma e eventualmente ao controle da doença. No entanto a identificação desses peptídeos-alvo na espécie humana apresenta um obstáculo de ordem ética, devido à impossibilidade de se efetuarem experimentos envolvendo transferência intencional da doença entre humanos. Disso decorre a necessidade de um método preditivo que permita identificar, na ausência de tais experimentações e indivíduos, o segmento peptídico de um antígeno que pode se ligar a um dado alelo de HLA.

1.1.3 – Métodos de predição da seletividade de moléculas do MHC.

Estágio Atual do Conhecimento.

O problema da predição dos peptídeos que se ligam a uma dada molécula de MHC tem sido extensivamente investigado nas últimas duas décadas [Mamitsuka, 1998; Cunha-Neto, 1999]. Entretanto, abordagens experimentais para a determinação da habilidade de ligação de peptídeos requerem passos que exigem muito tempo e recursos financeiros, como por exemplo, síntese de peptídeos e quantificação de sua habilidade de ligação ao MHC. Esta característica dos ensaios experimentais tem trazido dificuldades extremas para pesquisadores determinarem minuciosamente as seqüências de peptídeos que se ligam a determinado MHC [Gulukota, 1997].

Recentemente têm sido propostos alguns algoritmos de predição baseados em análise de dados já disponíveis. Meister [1995] propõe um algoritmo destinado a reconhecer, em uma seqüência peptídica ou protéica, um motivo previamente identificado como ligante: a utilidade deste método não seria tanto a predição, mas a localização de epítomos putativos. Mallios [1997]

e Borrás-Cuesta [2000], independentemente, propõem algoritmos de análise estatística, a saber, a ocorrência dos diversos resíduos de aminoácido em cada posição de ancoragem do MHC estudado, para vários peptídeos que ligam a este MHC. A própria porcentagem de acertos, apesar de alta (<90%), se refere à classificação de peptídeos que já são parte do espaço amostral utilizado para a análise estatística, não provando qual a capacidade preditiva do método. Southwood [1998] propõe um algoritmo baseado na ligação independente das cadeias laterais, onde a ocorrência de aminoácidos em determinadas posições do peptídeo de ligação gera um coeficiente que determina sua parcela de contribuição de energia livre, coeficiente este determinado a partir de uma biblioteca de peptídeos já conhecidos por sua capacidade de ligação. A aprendizagem por redes neurais tem sido apontada como uma maneira efetiva de abordar o problema [Brusic, 1998; Mamitsuka, 1998; Savoie, 1999]. Assim como para o caso da análise estatística, a capacidade de predição depende do número de complexos MHC/peptídeo utilizados para análise, sendo pouco indicado para o caso de alelos pouco estudados.

Com o objetivo de alcançar maior eficiência em tais investigações, esforços têm sido feitos para a utilização de computadores na predição de peptídeos que se ligarão em uma determinada molécula de MHC, e o que se deseja no momento é uma abordagem computacional precisa para a resolução deste problema.

Abordagens computacionais podem ser classificadas de uma maneira geral em dois tipos: 1) métodos que usam um número de peptídeos nos quais a habilidade de ligação é determinada experimentalmente; e 2) métodos que calculam a afinidade do peptídeo pelo MHC. Este último caso pode ser baseado na representação da superfície molecular ou na minimização da energia de interação entre o peptídeo e o MHC. Métodos baseados na representação da superfície molecular consomem menos recursos computacionais uma vez que o número de possibilidades dos modos de ligação pode ser grandemente reduzido pela utilização de um modelo simplificado das formas do receptor e do ligante. Baseados nesta idéia estão os trabalhos de Lee e Richard [1971], Connolly [1983], Bacon e Moulton [1992], Jiang e Kim [1991], Kuntz [1982] e colaboradores. No entanto, pelo próprio fato de ligante e receptor terem formas fixas, desconsideram possíveis mudanças de conformação na interação entre os mesmos.

Métodos baseados na otimização da energia de interação, por outro lado, representam uma maneira mais precisa na determinação da capacidade de ligação, porém requerem maior esforço computacional. Devido a isto a maioria das abordagens é baseada em simulações tipo Monte Carlo Metropolis, tais como os trabalhos de Goodsell e Olson [1990], Hart e Read [1992], e Caflisch [1992] e colaboradores. Nestes trabalhos já é considerada a mudança de conformação do ligante durante sua interação, porém com um receptor de forma estática. Além disso, não foram obtidos bons resultados para ligantes maiores que o sítio de interação, onde apenas parte do ligante interage com o receptor. Rosenfeld [1993] e colaboradores apresentam um estudo baseado na ligação de peptídeos com seleção randômica e minimização de energia potencial de estruturas de peptídeos, no entanto o método foi projetado para MHC's de classe I, onde os peptídeos têm tamanhos pré-determinados. Mais recentemente otimização por programação dinâmica é usada para a otimização da energia livre global, baseado num algoritmo de junção de fragmentos, e simulação por dinâmica molecular foi também utilizada para o estudo da afinidade de ligação da proteína HLA-B*2705, também de classe I. Outro algoritmo visando avaliar a energia de ligação de peptídeo/MHC, com base em resultados experimentais de energia de ligação e considerações estruturais, foi proposto por Altuvia [1997]. Apesar de obter resultados satisfatórios para MHC com fendas de apresentação predominantemente hidrofóbicas, o método falha para o caso de bolsões de ancoragem de caráter hidrofílico, ou com cargas.

Todas as abordagens propostas identificam a importância de uma predição acurada, o que leva à necessidade do estabelecimento de meios eficientes e sistemáticos de predição das estruturas globais energeticamente mais favoráveis [Mamitsuka, 1998].

O método MHCcalc

O método MHCcalc, desenvolvido em nosso laboratório, está baseado na avaliação energética do complexo MHC/peptídeo. Com este método é possível avaliar, quantitativamente, a contribuição da interação entre o bolsão do MHC e a cadeia lateral do aminoácido que será colocada no bolsão, na energia de interação total entre o peptídeo e o MHC. Conhecendo a

contribuição individual de cada aminoácido por bolsão, é possível estabelecer uma escala de preferência dos aminoácidos em cada bolsão.

Neste método, o MHC em estudo tem seu peptídeo original mutado para um peptídeo geometricamente equivalente de polialanina, e a energia do complexo resultante é minimizada. São também minimizadas, separadamente, as energias referentes ao peptídeo de polialanina e ao MHC, de forma que a energia de interação, ΔE_{inter} , entre o peptídeo e o MHC é definida como:

$$\Delta E_{inter} = E_{MHC+peptídeo} - [E_{MHC} + E_{peptídeo}] \quad (1)$$

Em seguida, no peptídeo do complexo MHC/peptídeo, em posições correspondentes aos bolsões, a cadeia lateral do aminoácido de alanina é substituída sucessivamente pelas outras 19 cadeias laterais de ocorrência natural, o que gera 19 estruturas, monosubstituídas, por bolsão estudado. Cada uma destas 19 estruturas é submetida à otimização geométrica, bem como também, separadamente, seus respectivos aminoácidos. Desta forma, é calculada a energia de interação, ΔE_{inter} , para cada aminoácido de ocorrência natural no bolsão. Esta energia de interação é recalibrada:

$$\Delta E_{mut} = \Delta E_{inter} - \Delta E_{inter}^0 \quad (2)$$

onde :

$$\Delta E_{inter}^0 = \Delta E_{inter}^{ALA} \quad (3)$$

sendo ΔE_{mut} a variação de energia potencial, chamada de energia de substituição, decorrente da substituição de um resíduo de alanina por um outro resíduo. Desta forma, ao invés de trabalhar com a energia de interação do peptídeo inteiro com o HLA, isolamos a parcela de energia correspondente à contribuição individual de cada resíduo estudado. Através de ΔE_{mut} pode-se estabelecer uma escala de preferência dos aminoácidos de ocorrência natural para cada bolsão estudado.

Preparo do arquivo de entrada

A estrutura molecular obtida do banco de dados de proteínas (figura 6) [Brookhaven Protein Data Bank; Stern, 1994], é composta por um dímero de proteínas HLA-DR1, onde cada

Figura 6

Figura 7

Figura 8

Figura 9

Figura 10

Fig. 6. Representação do dímero de proteínas DR1 conforme obtida do banco de dados de proteínas [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

Fig. 7. Estrutura quaternária da Proteína DR1 com peptídeo de hemaglutinina. As setas indicam em que ponto as cadeias α e β foram seccionadas [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

Fig. 8. Representação dos Domínios α_1 e β_1 de DR1 que formam o receptáculo para o peptídeo de hemaglutinina que foram utilizados como estrutura-base neste trabalho [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

Fig. 9. Superfície acessível ao solvente aquoso com geometria otimizada dos domínios α_1 e β_1 de DR1 com peptídeo de hemaglutinina.

Fig. 10. Superfície acessível ao solvente aquoso com geometria otimizada dos domínios α_1 e β_1 de DR1 com peptídeo de polialanina, utilizada como estrutura de trabalho do programa MHCalc.

monômero se encontra complexado com um peptídeo de hemaglutinina. Os monômeros são estruturalmente equivalentes, por isso um deles foi removido do arquivo original PDB.

Como o nosso interesse é o estudo da interação MHC/peptídeo, resolvemos retirar os domínios tipo imunoglobulina α_2 (THR83 a ALA182) e β_2 (GLN92 a ALA190) (figura 7). De fato estes domínios não entram em contato com o peptídeo a ser apresentado para o TCR e por isso não contribuem para o reconhecimento do peptídeo pela fenda apresentadora do MHC. Isto diminui significativamente o tempo de cálculo do programa THORMM [Moret, 1998], uma vez que a quantidade de átomos será aproximadamente metade em relação à proteína original, de forma que o número de interações interatômicas a serem avaliadas é drasticamente menor. Decidimos deixar estas novas extremidades na forma não ionizada a fim de simular a continuidade da cadeia principal da proteína. A extremidade carboxiterminal foi modificada para – COOH e a extremidade aminoterminal para – NH₂.

Sem os domínios α_2 e β_2 o receptáculo formado pelos domínios α_1 e β_1 (figura 8) tende a se desenovelar devido a separação das cadeias [Cunha-Neto, 1996]. Para reproduzir o efeito de estabilização devido à interação entre os domínios α_2 e β_2 , responsável pela manutenção da estabilidade da proteína, fixamos as coordenadas dos aminoácidos terminais, de forma que não pudessem ser modificadas durante a otimização geométrica, na totalidade das simulações efetuadas neste trabalho.

O complexo resultante formado pelos domínios α_1 , β_1 e peptídeo de hemaglutinina, após otimização de sua geometria (figura 9), teve seu peptídeo substituído por uma polialanina, aproveitando o esqueleto carbônico do peptídeo original, e esta nova estrutura teve sua geometria otimizada novamente (figura 10). A proteína obtida com novas coordenadas foi adotada como estrutura básica de trabalho do programa MHCcalc.

1.2 –Cálculos Empíricos de Energia Potencial: uma Revisão.

Nos últimos anos, cientistas, ligados a atividades acadêmicas e industriais, têm criado e validado novas ferramentas através da engenharia de software, as quais integram a Bioquímica

e a Biologia Molecular com as tecnologias de hardware e software disponíveis na atualidade [Fletcher, 1980; Frenkel, 1996].

Vários campos de estudo têm surgido e ferramentas específicas tornam possível aos cientistas a tomada de decisões dirigidas pelo conhecimento em suas pesquisas, as quais evitam sínteses onerosas e testes de compostos químicos com provável ineficiência, tornando mais rápida a resolução de problemas específicos na descoberta de novas drogas e compostos na indústria farmacêutica, agroquímica e de biotecnologia [Grant, 1995].

A utilização de softwares que exploram bancos de dados químicos e biológicos permite a procura de compostos por sub-estrutura, similaridade de propriedades e estereoisomeria, os quais auxiliam no estabelecimento de prioridades de síntese e compra de matérias primas em grandes indústrias. Outras ferramentas computacionais, como por exemplo, na área da Bioinformática, têm ajudado pesquisadores a extrair informações úteis de grandes volumes de dados gerados por tecnologias de ponta que caracterizam os métodos modernos de pesquisa.

Softwares também são utilizados em análise, visualização e processamento de dados de RMN multidimensional para visualização do espectro e da estrutura molecular obtidos. A construção de modelos gráficos e estatísticos que relatam as propriedades de moléculas e suas estruturas e subestruturas são utilizadas para predição de propriedades ou atividade de novos compostos (QSAR).

No campo da Modelagem Molecular têm sido criadas ferramentas para visualização, manipulação, análise, comparação e construção de estruturas químicas, bem como ferramentas de alinhamento de estruturas primárias de proteínas e de alças. Outras ferramentas auxiliam na identificação de sítios de ligação no interior ou na superfície de macromoléculas através da identificação de propriedades como hidrofobicidade e área acessível ao solvente. A seleção de possíveis ligantes também pode ser feita através de softwares que determinam a energia relativa de ligação receptor-ligante a partir de um banco de dados próprio ou comercial.

Através de cálculos de energia potencial pode ser feito o refinamento do posicionamento tridimensional de estruturas obtidas através de vários métodos, como por exemplo, por cristalografia de raios-X, homologia, ou ainda estruturas criadas a partir de modificações de

estruturas tridimensionais anteriormente conhecidas. A simulação de movimentos moleculares permite a obtenção de possíveis conformações e propriedades termodinâmicas. O *annealing* simulado permite o estudo dos possíveis confôrmeros e a energia mínima associada aos mesmos [Kirkpatrick, 1983].

No nosso caso, a estrutura de partida, obtida no *Brookhaven Protein Data Bank*, foi determinada por cristalografia de raios-X. Apesar de estas estruturas serem resolvidas a baixas temperaturas (-170°C) há uma movimentação dos átomos devido à energia térmica ainda existente. Isto significa que os dados obtidos de posicionamento tridimensional de cada átomo refletem uma média de posições num dado intervalo de tempo, e que não refletem, necessariamente, o posicionamento energeticamente mais estável do átomo em relação à estrutura. Desta forma, quando uma estrutura é determinada, geralmente ela precisa ser refinada, e trazida para uma conformação estável e estericamente aceitável. O processo de refinamento é conhecido como otimização geométrica por minimização de energia potencial, e é um processo iterativo no qual as coordenadas dos átomos são ajustadas de forma a trazer a energia da estrutura para um mínimo. A estrutura de mais baixa energia é considerada como o arranjo mais estável, e, por definição, de geometria otimizada [Insight II *User Guide*, 1995].

1.2.1 - Determinação da Energia Potencial de uma Estrutura.

A avaliação energética de um sistema molecular é feita através de uma equação de energia potencial em função das coordenadas de seus átomos [Karplus, 1990]. A descrição matemática completa da energia de uma molécula inclui efeitos relativísticos e quânticos devido ao movimento dos núcleos e elétrons, caracterizado por grandes velocidades e pequenas escalas. A mecânica e a dinâmica moleculares, no entanto, estão baseadas na mecânica newtoniana, onde dados empíricos incorporam implicitamente todos os efeitos quânticos e relativísticos. Assim, as coordenadas reais de uma molécula, combinadas com valores empíricos, criam a função energia, ou função alvo. Como exemplo, no caso simples da molécula de água, vamos considerar a seguinte equação, a qual pode ser usada para descrever sua energia:

$$U(R) = K_{oh} (b - b_{oh}^0)^2 + K_{oh} (b' - b_{oh}^0)^2 + K_{hoh} (\theta - \theta_{hoh}^0)^2 \quad (4)$$

onde K_{oh} , b_{oh}^0 , K_{hoh} e θ_{hoh}^0 são parâmetros empíricos tabelados e que constituem o Campo de Força, o qual é consultado nas avaliações energéticas. Os outros valores corresponderiam às coordenadas da molécula: "b" representa o comprimento de ligação efetivo de uma ligação OH, "b'" o comprimento da outra ligação OH, e θ o ângulo HOH [Insight II *User Guide*, 1995].

Esta simples equação de quatro variáveis tem como resolução uma superfície, denominada de superfície de energia potencial. Considerando uma molécula com maior número de átomos, ou mesmo se fossem considerados maior número de parâmetros para a molécula de água, como efeito do solvente, por exemplo, não haveria um equivalente gráfico para o conjunto multidimensional de dados obtidos como solução. Desta forma, o conjunto solução para uma equação que define a energia potencial de uma molécula constitui uma Hipersuperfície, pois o valor de energia potencial da molécula depende de múltiplas variáveis para o caso de macromoléculas de importância biológica [Karplus, 1990].

A resolução da equação energia de potencial, utilizando um ajuste empírico na superfície de energia potencial, é chamada Mecânica Molecular [Grant, 1995].

Campo de Força

O campo de força define as coordenadas usadas, a forma matemática das equações envolvendo as coordenadas e os parâmetros adequados no ajuste empírico da superfície de energia potencial. As funções vão de simples formas quadráticas até funções de Morse, expansões de Fourier, potenciais de Lennard-Jones, etc. Em seu nível mais simples a equação inclui estiramentos das distâncias de ligação, ângulos de curvatura de valência, torção, e termos de interação não-ligante. Todos estes termos associam uma penalidade energética à estrutura, tomando como base os desvios de uma geometria idealizada de equilíbrio [Insight II *User Guide*, 1995].

Utilizamos o campo de força GROMOS96 para a obtenção dos dados que serão apresentados, sendo que o mesmo é livre para uso acadêmico. Este campo de força foi

parametrizado para o uso em sistemas biomoleculares, como proteínas, ácidos nucleicos, oligossacarídeos e lipídeos em fase condensada [van Gunsteren, 1996].

À semelhança do que anteriormente já apresentamos de maneira geral, a expressão da equação energia potencial, utilizando um ajuste empírico na superfície de energia potencial, no campo de força GROMOS96 tem a seguinte forma:

$$\begin{aligned}
 U(r_i) = U(r_1, r_2, \dots, r_N) = & \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n} [b_n^2 - b_{0_n}^2]^2 + \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} [\cos \theta_n - \cos \theta_{0_n}]^2 + \\
 & + \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} [\xi_n - \xi_{0_n}]^2 + \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)] + \\
 & + \sum_{ij} \left\{ \left[\frac{C_{12}(i, j)}{r_{ij}^6} - C_6(i, j) \right] \frac{1}{r_{ij}^6} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \left[\frac{1}{r_{ij}} - \frac{\frac{1}{2} C_{rf} r_{ij}^2}{R_{rf}^3} - \frac{1 - \frac{1}{2} C_{rf}}{R_{rf}} \right] \right\} \quad (5)
 \end{aligned}$$

onde a energia potencial (U) é dependente das coordenadas cartesianas $\mathbf{r}_i$ das N partículas e de um número de parâmetros discutidos em detalhes abaixo.

Os quatro primeiros termos da equação acima se referem à interação entre os átomos covalentemente ligados, sendo calculados como a somatória de termos referentes a todas as ligações na estrutura, na qual a energia potencial associada a cada ligação é avaliada com base na diferença entre valor efetivo e valor de equilíbrio da variável. Esta se encontra multiplicada por uma constante, a constante de força da ligação, a qual aumenta para um dado desvio do valor ideal. Estas quatro diferentes parcelas são definidas respectivamente como:

1. Comprimento da ligação
2. Ângulo da ligação
3. Ângulo diedral impróprio (harmônico)
4. Ângulo diedral próprio (trigonométrico)

Desta forma, no primeiro termo, o qual representa a interação de estiramento da ligação covalente, a somatória geralmente é efetuada sobre todas $n = 1, \dots, N_b$ ligações covalentes do

sistema molecular. O comprimento de ligação real da ligação n entre os átomos i e j é descrita como

$$b_n = r_{ij} \equiv [r_{ij}^p \cdot r_{ij}^q]^{1/2} \quad r_{ij}^p \equiv r_i^p - r_j^p$$

Os parâmetros K_{bn} e b_{0n} e suas atribuições aos fragmentos moleculares são especificados em uma tabela, a qual é consultada durante as avaliações energéticas.

No segundo termo, o qual representa a interação do ângulo da ligação covalente, o valor real do ângulo da ligação n entre os átomos i, j e k ($i-j-k$) é denotado por

$$\theta_n = \arccos \left(\frac{r_{ij}^p \cdot r_{kj}^p}{|r_{ij}^p| \cdot |r_{kj}^p|} \right), \quad 0 \leq \theta_n \leq \pi$$

Da mesma forma que anteriormente, os parâmetros $K_{\theta n}$ e θ_{0n} e suas atribuições aos fragmentos moleculares são especificados em uma tabela.

No terceiro termo, representando a assim chamada interação de ângulo diedral imprópria (configuração fora de plano ou de tetraedro), a somatória é efetuada sobre um grupo de ângulos diedros $n = 1, \dots, N_\xi$, os quais são selecionados para manter grupos de átomos próximos a uma configuração espacial específica. Por exemplo, em um resíduo de aminoácido os átomos CA, C, O e N são mantidos próximos a uma configuração planar pela definição do ângulo diedral impróprio CA-N-C-CB com $\xi_0 = 0^\circ$ (eclipsado). O valor real do ângulo diedral impróprio definido pelos átomos i, j, k e l ($i-j-k-l$) é definido como:

$$\xi_n = \text{sign}(\xi_n) \arccos \left(\frac{r_{mj}^p \cdot r_{nk}^p}{|r_{mj}^p| \cdot |r_{nk}^p|} \right), \quad -\pi \leq \xi_n \leq \pi \quad (6)$$

onde

$$r_{mj}^p = r_{ij}^p \times r_{kj}^p \quad r_{nk}^p = r_{kj}^p \times r_{kl}^p$$

No quarto termo, o qual representa a interação trigonométrica de ângulo diedral, os ângulos diedrais são escolhidos da seguinte forma:

1. Para cada ligação entre átomos j e k somente um quarteto de átomos $i-j-k-l$ é escolhido para formar o diedro φ_n .

2. Ligações $j-k$ que são parte de uma estrutura planar rígida (TRP, PHE, etc.) não são incluídas, mas tratadas como diedros impróprios.
3. Ângulos de torção ao redor de ligações $j-k$ onde j e k denotam átomos de carbono em anéis de açúcares, ou na presença de um átomo de fósforo são tratados de maneira especial, com uma constante $K_{\varphi n}$ específica para cada caso.

O valor real de torção de ângulo diedral definido pelos átomos i , j , k , e l ($i-j-k-l$) é denotado

$$\varphi_n = \text{sign}(\varphi_n) \arccos \left(\frac{\mathbf{r}_{im'} \cdot \mathbf{r}_{ln'}}{|\mathbf{r}_{im'}| \cdot |\mathbf{r}_{ln'}|} \right), \quad -\pi \leq \varphi_n \leq \pi \quad (7)$$

onde

$$\mathbf{r}_{im'} = \mathbf{r}_{ij} - \frac{(\mathbf{r}_{ij} \cdot \mathbf{r}_{kj})}{r_{kj}^2} \mathbf{r}_{kj} \quad (8) \quad \mathbf{r}_{ln'} = -\mathbf{r}_{kl} + \frac{(\mathbf{r}_{kl} \cdot \mathbf{r}_{kj})}{r_{kj}^2} \mathbf{r}_{kj} \quad (9)$$

e

$$\text{sign}(\varphi_n) = \text{sign}(\mathbf{r}_{ij} \cdot (\mathbf{r}_{kj} \times \mathbf{r}_{kl}))$$

Os valores de equilíbrio idealizados de distâncias e ângulos de ligação, mais as constantes de força, os raios de van der Waals e constantes associadas requeridas para calcular as interações não-ligantes, são estocados em um arquivo, onde serão buscados quando o cálculo estiver em andamento. A combinação destes parâmetros com as formas funcionais dos termos individuais de energia é conhecida como campo de força. É importante salientar que os valores ideais de ligações e ângulos não podem ser baseados puramente no tipo de elemento químico. Podemos citar como exemplo que uma distância de ligação simples carbono-carbono é maior que em uma ligação dupla; e ainda, a dupla ligação da carbonila amídica devido ao seu caráter de dupla ligação parcial, possui um valor intermediário. Por esta razão, cada átomo em uma estrutura deve ter um potencial atribuído dependendo de seu estado de hibridização e ambiente de cada átomo. Os parâmetros de campo de força estão estocados em um arquivo de parâmetros de acordo com o tipo de potencial.

O último termo da equação se refere à parcela de interação não-ligante, e inclui:

1. Atrações de van der Waals;
2. Repulsões de van der Waals;
3. Interações eletrostáticas.

Estes três termos são somados para todos os pares de átomos da estrutura que estão além da formação do ângulo diedro, e que, portanto, influenciam pouco a polarização da ligação, ou seja, átomos que estão a mais de quatro ligações de distância. Nos pares atômicos nos quais um dos átomos está ligado, ou envolvido em um ângulo de interação, o outro membro deste par atômico é excluído do cálculo da energia não-ligante. O termo eletrostático requer que seja atribuída uma carga parcial a cada átomo, a qual para qualquer átomo pode ser atrativa ou repulsiva dependendo do sinal da carga envolvida. O número de pares não-ligantes cresce com o quadrado do número de átomos, e assim o termo de interação não-ligante domina o tempo de computação requerido para avaliação energética em cálculos para estruturas maiores. Por esta razão é comum limitar a avaliação energética não-ligante para pares com uma certa distância chamada raio de corte, assumindo que pares de átomos separados por grandes distâncias contribuem com uma parcela desprezível na energia total.

Desta forma, o termo representando as interações não-ligantes é descrito como o termo de van der Waals (Lennard-Jones), somado com a parte de reação do campo de Coulomb e Poisson-Boltzmann, onde:

$$C_{ij} = \frac{(2\epsilon_1 - 2\epsilon_2)(1 + \kappa R_{ij}) - \epsilon_2 (\kappa R_{ij})^2}{(\epsilon_1 - 2\epsilon_2)(1 + \kappa R_{ij}) + \epsilon_2 (\kappa R_{ij})^2} \quad (10)$$

Esta soma é efetuada a princípio sobre todos os pares de átomos i e j . No entanto, geralmente um certo número de pares é excluído da soma:

- pares atômicos de vizinhos excluídos;
- pares atômicos não inclusos na lista de pares de grupos de cargas (raio de corte R_{cp}) ou na interação não-ligante de longa distância (raio de corte R_{cl}).

Átomos i e j que estão ligados covalentemente são chamados primeiros vizinhos, e, se cada um deles está ligado covalentemente a mais átomos vizinhos, os mesmos são chamados

segundos vizinhos. Devido à curta distância r_{ij} entre o primeiro e o segundo vizinhos, a interação não-ligante entre os vizinhos i e j será muito grande. Desta forma, primeiros e segundos vizinhos são excluídos da somatória da interação não-ligante. Terceiros vizinhos são excluídos somente em casos especiais: entre átomos pertencentes ou anexados a anéis aromáticos e entre átomos pertencentes a grupos fosfato. De uma forma geral, átomos de moléculas de solvente são tratados como vizinhos excluídos.

A versão utilizada do campo de força GROMOS96 trata carbonos alifáticos com seus hidrogênios como átomo unido, ou seja, como apenas um átomo. Quando os parâmetros de van der Waals para os átomos unidos são aplicados a átomos que são separados por três ligações covalentes, os então-chamados terceiros vizinhos, estes átomos são induzidos a uma grande repulsão em conformação *gauche*. Com o objetivo de evitar este efeito, um grupo menor de parâmetros de van der Waals são usados na obtenção de parâmetros para terceiros vizinhos envolvendo átomos unidos. Em contraste a amplamente divulgada prática de obtenção da interação de terceiros vizinhos não-ligantes pela aplicação de um fator de escala, a interação não-ligante normal, no campo de força GROMOS96 a interação de terceiros vizinhos é independente da interação normal de van der Waals.

A interação de Coulomb na equação do campo de forças é inversamente proporcional a distância r_{ij} entre átomos i e j , o que torna esta interação de longa distância. Uma vez que os vizinhos não-ligantes crescem proporcionalmente ao cubo do raio de corte, a aplicação de um raio de corte de interação não-ligante diminui dramaticamente o tempo de computação requerido. Esta situação pode ser evitada pela aplicação de dois conceitos.

1. A definição de grupos de cargas torna a interação eletrostática uma interação dipolar, que tem uma dependência de $1/r^3$ da distância.
2. A negligência de componentes de alta frequência da parte de longa distância das interações de Coulomb permite uma considerável redução do tempo de computação requerido.

No campo de força GROMOS96 a permissividade elétrica relativa ϵ_r é 1 dentro da esfera de raio de corte interatômica. Quando a soma das cargas parciais de um grupo atômico resulta

em exatamente zero, o termo de interação eletrostática entre estes dois grupos é de caráter dipolar. Assim, a faixa de interação eletrostática pode ser consideravelmente reduzida quando átomos são agrupados nos então-chamados grupos de cargas, os quais têm uma carga total igual a zero, e para os quais a interação eletrostática com outros grupos atômicos ou é calculada para todos os átomos do grupo de carga, ou para nenhum. O campo de força GROMOS96 faz uso deste conceito de grupo de cargas. As cargas atômicas q_i dos átomos i e as definições dos grupos de cargas são definidos em tabelas e consultados. Quando um raio de corte é utilizado, a distância entre dois grupos de cargas deve ser definida. A posição de um grupo de cargas de átomos do soluto é considerada como seu centro geométrico. Uma molécula de solvente é considerada como formando um grupo de cargas, e a posição de seu primeiro átomo considerado como a posição do grupo de cargas.

1.2.2 – Minimização de Energia e o problema dos Múltiplos Mínimos.

Para ser iniciado o processo de minimização de energia potencial da estrutura tridimensional de uma molécula, a equação que define a energia potencial do sistema em função de suas coordenadas deve ter sido previamente definida, bem como o critério que definirá a convergência da estrutura a um valor de energia potencial mínimo. Este critério é definido através do número de casas decimais de precisão que se deseja no cálculo da energia potencial da estrutura em questão. Quando a diferença entre as avaliações energéticas for menor que um dado valor previamente definido, este valor será julgado como a energia potencial mínima do sistema. Após a primeira avaliação energética da estrutura a ser estudada, a conformação é ajustada de forma que um novo valor de energia potencial é encontrado para a função alvo através de nova avaliação energética. Se o novo valor encontrado para a função alvo for menor que o valor anterior, o ajuste feito na conformação permanece e novos ajustes serão feitos na mesma direção iniciada. Caso o novo valor de energia potencial encontrado seja maior que o inicial, uma nova direção de ajuste na função é escolhida e seguida. Um valor de mínimo pode

ser encontrado após apenas um ajuste, ou pode requerer muitos milhares de iterações, dependendo da natureza do algoritmo de otimização, da forma da função alvo e do tamanho da molécula em estudo. A eficiência da minimização é, desta forma, julgada pelo tempo necessário para a avaliação da função alvo e do número de ajustes estruturais (iteraões) necessários para a convergência a um mínimo [Grant, 1995].

Para a descrição do processo de minimização utilizaremos como exemplo uma superfície elíptica em duas dimensões descrita pela equação:

$$E(x,y) = x^2 + 5y^2 \quad (11)$$

Esta simples função ilustra as propriedades dos algoritmos de minimização e capta a essência matemática das formulações. Toda minimização começa com uma equação análoga a (11). Para a definição da superfície de energia potencial um conjunto de coordenadas iniciais – um ponto de partida – deve ser providenciado.

Figura 11 – Representação gráfica da equação 11, onde o ponto de energia mínima pode ser deduzido visualmente.

Obviamente o mínimo desta função simples é trivial e pode ser deduzido visualmente como (0,0).

Dada uma função que define a superfície de energia potencial e um conjunto de coordenadas iniciais, um minimizador deve determinar tanto a direção a ser seguida para o mínimo, quanto à distância para o mínimo naquela direção. Uma boa direção inicial é simplesmente a inclinação, ou derivada, da função na posição atual. As derivadas da equação (11) resultam em um vetor bidimensional:

$$dE = 2x + 10y \quad (12)$$

onde as derivadas são proporcionais às coordenadas, de forma que, quanto mais longe do mínimo, maior a derivada.

As derivadas, no entanto, meramente apontam direção abaixo, e não necessariamente para o mínimo. Assim, conforme o movimento na direção indicada pelas derivadas iniciais, as novas derivadas mudam e apontam para uma nova direção. Com o objetivo de melhorar a eficiência, os algoritmos mais sofisticados, tais como *conjugate gradients* [Press, 1992], utilizam informações de como as derivadas, em função de cada coordenada, mudam, para determinar a nova direção a ser seguida.

As funções que descrevem macromoléculas de importância biológica, no entanto, por terem múltiplas variáveis que definem uma hipersuperfície de energia potencial como conjunto solução, possuem, não um, mas vários pontos de energia mínima [Frauenfelder, 1991; Wales, 2001]. O mínimo global é a conformação de mais baixa energia, e os mínimos restantes são chamados locais [Karplus, 1990]. Em nosso caso estamos interessados em encontrar o ponto de energia mínima a partir da estrutura biologicamente ativa, não importando ser este o mínimo local ou global. Supomos, no entanto, estar pelo menos próximos ao mínimo global, uma vez que a conformação biologicamente ativa é estável em condições fisiológicas.

Algoritmos de Minimização

A maior parte dos algoritmos de minimização incorpora o conceito de linha de procura. Estes algoritmos utilizam linhas de procura para modificar as coordenadas iniciais para uma nova estrutura de mais baixa energia potencial [Insight II *User Guide*, 1995]. Como ilustração, podemos considerar a figura 12, na qual a direção gradiente de um ponto inicial arbitrário foi sobreposta na função elíptica. O ponto inicial (x_0, y_0) é definido como ponto “a”.

Em termos simples, as modificações conformacionais feitas pelo algoritmo de minimização, através de uma linha de procura, equivalem a uma minimização unidimensional ao longo do vetor direção, determinado a cada iteração. Para o caminho mostrado na figura 15, isto seria ao longo do vetor $(2x, 10y)$, e a superfície unidimensional pode ser expressa parametricamente em termos de uma coordenada α :

$$(x', y') = \left[x_0 + \alpha \frac{\partial E}{\partial x_{x_0, y_0}}, y_0 + \alpha \frac{\partial E}{\partial y_{x_0, y_0}} \right] \quad (13)$$

onde (x', y') são coordenadas ao longo da linha do ponto atual na direção das derivadas em (x_0, y_0) . Se a energia referente a estes novos pontos fosse calculada e plotada em função de α , a curva na figura 12 seria obtida.

Figura 12 – Representação gráfica da procura do ponto de energia mínima com a utilização de linhas de procura.

O mínimo na direção α , “c”, coincide com o ponto no qual a linha de procura é tangencial ao contorno de energia. Pelo fato de a direção de máxima derivada ser perpendicular a linha de procura neste ponto, cada linha de procura é ortogonal à precedente. Esta é uma importante propriedade das linhas de procura, que será discutida com mais profundidade na descrição do algoritmo *conjugate gradients* [Press, 1992], logo a seguir.

Linhas de procura não dependem do algoritmo que gera o vetor direção. A estratégia geral é simplesmente determinar o mínimo unidimensional entre dois pontos de maior energia, por exemplo, pontos “b” e “d” na figura 12. Assim, através de sucessivas iterações, pode ser conseguido por aproximação o ponto de mínimo verdadeiro. Linhas de pesquisa amplas são atrativas, pois elas extraem toda a informação de como a energia potencial varia em uma direção antes de ir para próxima. Além disso, o fato de as novas derivadas serem sempre perpendiculares as anteriores produz um caminho eficiente para o mínimo através de superfícies aproximadamente quadráticas. Na prática, no entanto, linhas de pesquisa são caras, em termos de número de avaliações da função energia potencial que devem ser realizadas. A energia deve

ser avaliada de 3 a 10 vezes para localizar precisamente o mínimo unidimensional, e, desta forma, linhas de pesquisa amplas são ineficientes [Insight II *User Guide*, 1995].

Para maior eficiência da discussão apresentada abaixo, uma iteração de minimização deve ser explicitamente definida. Desta forma, uma iteração é completada quando o vetor direção é atualizado. Para minimizadores utilizando linhas de procura, cada linha de procura completada é uma iteração. Iterações não devem ser confundidas com avaliações da função de energia potencial.

No método *steepest descent* [Wiberg, 1965] (descida mais íngreme), a direção da linha de procura é definida pela direção do gradiente de descida local $-\Delta E(x_i, y_i)$. A figura 12 mostra o caminho de minimização seguido por um algoritmo *steepest descent* para a função quadrática simples. Como esperado, cada linha de procura produz uma nova direção que é perpendicular ao gradiente prévio; entretanto, as direções oscilam ao longo do caminho para o mínimo. Este comportamento ineficiente é característico deste método, especialmente em superfícies de energia contendo vales estreitos.

O que aconteceria se a linha de procura rigorosa fosse eliminada e a posição fosse atualizada simplesmente se o ponto de tentativa ao longo do gradiente tivesse uma energia menor? A vantagem seria que o número de avaliações da função por iteração decresceria dramaticamente. Além disso, devido as constantes mudanças na direção para acertar o gradiente, oscilações ao longo do caminho da minimização poderiam ser diminuídos. O resultado da minimização seria como mostrado na figura 13. A minimização começa do mesmo ponto que na figura 12, mas cada linha de procura usa, no máximo, duas avaliações de função (se o ponto de tentativa tem uma energia maior, o tamanho do passo é ajustado e um novo ponto de tentativa é gerado). Note que os passos são mais erráticos aqui, mas o mínimo é alcançado em aproximadamente o mesmo número de iterações. O aspecto crítico, entretanto, é que, ao evitar linhas de procura extensas o número de avaliações da função é apenas 10-20 % daquele utilizado em linhas de procura rigorosas. A confiança exclusiva do método *steepest descent* [Wiberg, 1965] sobre métodos gradientes é ao mesmo tempo sua fraqueza e sua força. A

convergência é lenta próxima ao mínimo, pois o gradiente é próximo a zero, mas o método é extremamente robusto, mesmo para sistemas que estão longe de serem harmônicos.

Figura 13 - Representação gráfica da procura do ponto de energia mínima sem a utilização de linhas de procura.

Este é o método que melhor gera uma estrutura de mais baixa energia, não importando qual seja a função nem onde ela começa. Desta forma, o *steepest descent* [Wiberg, 1965] é freqüentemente usado quando o gradiente é grande e a conformação esta longe de um mínimo. Este é comumente o caso para o relaxamento inicial de estruturas cristalográficas fracamente refinadas e estruturas construídas graficamente. De fato, como será explicado a seguir, algoritmos mais avançados são construídos para começar com direcionamento *steepest descent* [Insight II User Guide, 1995].

A razão pela qual o método *steepest descent* [Wiberg, 1965] converge lentamente próximo do mínimo é que cada segmento do caminho tende a reverter o progresso feito na iteração anterior. Por exemplo, na figura 13, cada linha de procura se desvia de alguma forma em relação à direção ideal para o mínimo. Sucessivas linhas de procura corrigem este desvio, mas elas não podem corrigir eficientemente, pois cada direção deveria ser ortogonal a direção prévia. Assim, o caminho oscila e é corrigido continuamente para escolhas pobres de direção em passos anteriores.

Seria preferível prevenir o próximo vetor direção de desfazer o progresso já feito. Isto significa a utilização de um algoritmo que produz um conjunto completo de direções mutuamente conjugadas, de forma que cada passo sucessivo refine continuamente a direção para o mínimo. Se estas direções conjugadas verdadeiramente atravessarem o espaço da superfície de energia,

a minimização ao longo de cada direção por sua vez deve, por definição, terminar chegando em um mínimo [Grant, 1995].

O algoritmo *conjugate gradients* [Press, 1992] (gradientes conjugados) constrói e segue um conjunto de direções. Neste método, h_{i+1} , o novo vetor direção vindo do ponto $i+1$, é computado pela adição do gradiente no ponto $i+1$, g_{i+1} , a direção prévia h_i escalada pela constante γ_i :

$$h_{i+1} = g_{i+1} + \gamma_i h_i$$

γ_i é um escalar que pode ser definido de duas formas. No método Polak-Ribiere, γ_i é definido como

$$\gamma_i = \frac{g_{i+1} \cdot g_{i+1}}{g_i \cdot g_i}$$

E no método Fletcher-Reeves, γ_i é definido como :

$$\gamma_i = \frac{(g_{i+1} - g_i) \cdot g_{i+1}}{g_i \cdot g_i}$$

Embora os dois métodos de gradiente conjugado tenham características similares, o método Fletcher-Reeves costuma ser mais robusto, e é o método utilizado nos cálculos que serão apresentados logo mais.

Esta direção é então usada em lugar do gradiente na equação e uma nova linha de procura é então conduzida. Esta construção tem a propriedade de que o próximo gradiente, g_{i+1} , é ortogonal a todos os gradientes anteriores, e a próxima direção é conjugada a todas as direções anteriores. Desta forma, o termo *conjugate gradients* [Press, 1992] é de alguma forma mal colocado. O algoritmo produz um conjunto de gradientes mutuamente ortogonais e um conjunto de direções mutuamente conjugadas [Insight II *User Guide*, 1995].

Este método é o escolhido para grandes sistemas, pois somente os 3N gradientes e direções prévias precisam ser estocados. Entretanto, para assegurar que as direções são mutuamente conjugadas, minimizações com linhas de procura mais completas devem ser realizadas ao longo de cada direção. Estas linhas de procura consomem várias avaliações de função por procura, o tempo por iteração é maior para *conjugate gradients* [Press, 1992] do que

para *steepest descent* [Wiberg, 1965]. Isto é mais do que compensatório pela convergência mais eficiente em direção ao mínimo conseguida por conjugate gradients.

Para a obtenção dos nossos resultados utilizamos a opção OPT3 de cálculo no programa THORMM [Moret, 1998] que utiliza inicialmente o algoritmo *steepest descent* [Wiberg, 1965] e, em seguida, o algoritmo *conjugate gradients* [Press, 1992]. Isto diminui o número de iterações para atingir o critério de convergência estabelecido, diminuindo assim o tempo de computação requerido, mesmo quando o gradiente da estrutura for grande e a conformação estiver longe de um mínimo de energia potencial [Frenkel 1996].

2 – Objetivos

Este trabalho tem como objetivo a automatização de uma metodologia desenvolvida em nosso laboratório. A metodologia em questão, MHCcalc, avalia quantitativamente a seletividade de proteínas do MHC para diferentes antígenos.

A partir da metodologia podem-se obter regras de seletividade para cada um dos bolsões que compõe o MHC estudado. O método utiliza minimização de energia potencial de complexos MHC/peptídeo como ferramenta de estudo.

Pretendemos obter, como resultado final, um programa que tenha como entrada um arquivo em formato PDB com dados referentes a uma proteína MHC de classe II complexada a um peptídeo. Este arquivo pode ter sido obtido por cristalografia de raios-x ou por modelagem molecular por homologia. A saída deste programa deverá fornecer um arquivo contendo a ordem de preferência de cada bolsão que compõe o MHC, para os vinte aminoácidos de ocorrência natural.

Também pretendemos fornecer dados referentes à aplicação do programa resultante da automatização do método MHCcalc na proteína HLA-DR1 e comparar os resultados com a literatura disponível.

3 – Metodologia

A visualização e construção dos modelos moleculares foram efetuadas em estações IBM RS6000-3AT e IBM 43P, utilizando o package Insight II [Insight II *User Guide*, 1995]. O

refinamento dos modelos construídos, assim como os cálculos de Mecânica Molecular, necessários para a avaliação quantitativa das parcelas individuais de energia de interação, foram efetuados em um computador PC Pentium III 1GHz 256RAM, utilizando o pacote de programas THORMM [Moret, 1998].

O algoritmo desenvolvido, bem como a interface entre este e o programa THORMM, foi implementado utilizando-se a linguagem de programação C [Kernighan, 1988], e compilando-se os códigos em ambiente GNU/Linux (distribuições Debian 3.0 e RedHat 9.0).

Para os cálculos empíricos de otimização de geometria do sistema foi utilizada a opção de cálculo OPT3 que emprega os algoritmos minimizadores *steepest descent* [Wiberg, 1965] e, em seguida, *conjugate gradients* [Press, 1992]. O critério de convergência final adotado foi de $\Delta E < 0.05 \text{ Kcal/mol}\cdot\text{\AA}$. Todos os cálculos foram efetuados usando-se o campo de força GROMOS96 [van Gunsteren, 1996], considerando uma constante dielétrica relativa ϵ igual a 80 para representar implicitamente o solvente aquoso.

4 – Resultados e Discussão

4.1 – Desenvolvimento e implementação de MHCalc

A automatização da metodologia MHCalc resultou no programa homônimo. Este programa constrói automaticamente os arquivos de entrada para o programa THORMM [Moret, 1998], correspondentes aos complexos mono-substituídos, a partir de um arquivo inicial em formato PDB. O programa também retrabalha as energias resultantes das minimizações energéticas dos complexos mono-substituídos a fim de recalcular a observável ΔE_{mut} . O programa tem cinco sub-rotinas, correspondente a cinco etapas.

A primeira sub-rotina, LETRA, padroniza o formato dos arquivos de entrada, que deve ser um arquivo tipo PDB, podendo este ter sido obtido diretamente do Banco de dados *Brookhaven Protein Data Bank* (<http://www.rcsb.org/pdb>) ou produzido pelo usuário através da construção de um modelo por homologia. A sub-rotina LETRA quantifica as cadeias existentes no arquivo original e atribui-lhes identificação, caso não possuam.

A segunda sub-rotina, AMINO, gera os arquivos de entrada para o programa

PDBTHORBF, fornecido juntamente com o pacote THORMM [Moret, 1998], a partir do arquivo precedente. Resolvemos utilizar o programa PDBTHORBF, pois este, além de auxiliar no preparo dos arquivos de entrada para THORMM, ao detectar alteração de aminoácido no arquivo de seqüências, procede à substituição do mesmo no arquivo de coordenadas, resultando em uma mutação virtual. O programa PDBTHORBF necessita: do arquivo em formato PDB gerado pela sub-rotina LETRA, de onde serão obtidas as coordenadas dos átomos; de arquivos separados contendo as diferentes seqüências de aminoácidos, em código de três letras, de cada cadeia que constitui o complexo em estudo, gerados por AMINO; de um arquivo INPUT.IN, contendo a relação dos arquivos com os quais o programa irá trabalhar, também gerado por AMINO.

Uma terceira sub-rotina, SUBSTITUICOES, gera um peptídeo de polialanina no lugar do peptídeo original proveniente do arquivo PDB. Para isto o programa detecta entre os arquivos de seqüências, aquele correspondente ao peptídeo e gera um novo arquivo com seqüência de polialanina. O final deste programa evoca o programa PDBTHORBF, o qual gera os arquivos de coordenadas correspondentes no formato aceito pelo programa THORMM.

A quarta sub-rotina consiste em uma modificação do programa PDBTHORBF, para que o mesmo produzisse os arquivos de opções de cálculo para o programa THORMM conforme o protocolo por nós selecionado para o tratamento do MHC: campo de força GROMOS96; constante dielétrica simulando água implicitamente; estratégia de minimização/algoritmos utilizados e critérios sucessivos de convergência para as minimizações.

Neste ponto o programa MHCcalc evoca o programa THORMM, o qual efetua a otimização de geometria do complexo MHC/polialanina. A geometria resultante, salva em arquivo com formato PDB, será utilizada como ponto de partida para as próximas etapas.

Os dados do novo complexo MHC/polialanina são novamente separados em arquivos individuais através da subrotina AMINO. A sub-rotina SUBSTITUICOES efetua as $n \times 19$ mutações sucessivas, sendo n o número de aminoácidos do peptídeo, no arquivo com código de três letras correspondente ao peptídeo polialanina. Esta mesma sub-rotina evoca, a cada mutação, PDBTHORBF, o qual gera os arquivos em formato próprio para o programa THORMM.

A vantagem deste algoritmo é que os novos complexos a serem otimizados terão apenas um aminoácido de diferença em relação ao arquivo original, de forma que a nova otimização de geometria será feita em todo o complexo, mas só há mudança sensível de conformação e necessidade de refinamento nas proximidades do local onde ocorreu a substituição.

Neste ponto, o programa MHCcalc inicia outro programa, RODA, o qual gerencia a submissão dos arquivos, já gerados, à minimização de energia por THORMM. Caso o programa RODA seja interrompido por motivo qualquer, o mesmo pode ser evocado novamente. Neste caso o programa busca entre os arquivos aqueles que já foram otimizados, e reinicia a submissão a partir do ponto onde havia sido interrompido.

Finalmente, após as minimizações energéticas, uma última sub-rotina ANALISA é automaticamente iniciada. Esta última, ao buscar nos arquivos de saída do programa THORMM a energia potencial resultante, verifica se o cálculo convergiu apropriadamente para cada complexo. De posse destes dados a sub-rotina ANALISA retrabalha as energias de acordo com as equações (1), (2) e (3) (pág 17) e as apresenta em forma de tabelas, já em ordem decrescente de preferência pelos diferentes resíduos, sendo uma tabela para cada bolsão do MHC.

No total foram escritas 1175 linhas de código fonte para o conjunto de programas descritos acima, os quais constituem o programa MHCcalc (anexo 1). Utilizamos o compilador gcc [GCC, 1999], fornecido com as distribuições GNU/Linux utilizadas (distribuições Debian 3.0 e RedHat 9.0), com a opção `-lm`, conforme exemplo abaixo, para linkagem da biblioteca `math.h`, utilizada pelas sub-rotinas.

```
$ gcc -lm mhcalc.c -o mhcalc
```

4.2 - Sistema Molecular Estudado

Para testar a eficiência do programa MHCcalc, foi utilizada a molécula HLA-DR1 (DRA, DRB1 0101). Das várias estruturas presentes no banco de dados escolhemos aquela complexada com um peptídeo de hemaglutinina de *Influenza* (código 1DLH no *Brookhaven Protein Data Bank*) [Stern, 1994] (figuras 6-10), porque há abundante quantidade de dados

experimentais e teóricos [O'Sullivan, 1991; O'Sullivan, 1990; Brown, 1993; Hammer, 1994; Hammer, 1993; Hammer, 1992; Stern, 1994; Androulakis, 1997; De Oliveira, 2000; Jardetzky, 1996; Murthy, 1997; Burley, 1986; Burley, 1985; Serrano, 1991; Douguerty, 1996; Georges, 2000] para comparação com nossos resultados. Outro motivo pelo qual foi escolhido o peptídeo de hemaglutinina para o nosso estudo é que o mesmo, quando ligado a DR1, representa uma conformação comum a vários outros peptídeos [Jardetzky, 1996; Murthy, 1997], em hélice tipo poliprolina II.

4.3 - Teste de Aplicação do Programa MHCcalc na proteína HLA-DR1

Após obtermos o arquivo correspondente à proteína HLA-DR1 no *Brookhaven Protein Data Bank* fizemos as modificações necessárias, conforme descrito no item 4.4, página 40 (Descrição de HLA-DR1). Este arquivo foi submetido ao programa MHCcalc, o qual gerenciou a geração dos arquivos de entrada para o programa THORMM com mutações em todos os aminoácidos, uma posição por vez, com todos os outros 19 aminoácidos de ocorrência natural, no peptídeo de polialanina. O programa MHCcalc em seguida submeteu, seguidamente e um por vez, os arquivos à otimização geométrica, para, em seguida, retrabalhar as energias obtidas e apresentá-las em forma de tabela.

Obtivemos um primeiro conjunto de dados que correspondem a um teste executado com a finalidade de confirmação da funcionalidade do programa MHCcalc. Este teste levou 2340 horas para ser executado em um computador PC Pentium III 1GHz 256RAM, o que equivalem a 97 dias e meio. O teste confirmou a funcionalidade do programa MHCcalc, e o mesmo foi então executado novamente, agora com a finalidade de gerar os dados e serem trabalhados. Os dados resultantes do teste citado não puderam ser utilizados devido a um erro contido nos arquivos de entrada do programa, o que resultou em carga líquida inicial diferente da apresentada pela proteína em questão.

Apresentamos na tabela 1 (pág. 44) os resultados obtidos com a aplicação do programa MHCcalc no bolsão 1, juntamente com a comparação destes dados com os resultados

encontrados na literatura.

4.4 - Descrição de HLA-DR1

O reconhecimento de peptídeos exógenos por moléculas MHC tem papel fundamental na iniciação e regulação da resposta imune. Segundo Stern [1994], em apenas quatro aminoácidos, dos 13 possíveis do peptídeo de hemaglutinina, as interações peptídeo/MHC seriam mais significativas, devido ao contato maximizado entre as cadeias laterais dos peptídeos e os aminoácidos que compõe os bolsões. De fato pudemos observar a complementaridade entre bolsões e o MHC (figs. 9 e 10). Portanto resolvemos realizar o estudo pelas posições correspondentes aos bolsões: 1, 4, 6 e 9 (fig. 14). As demais posições trazem pouca contribuição para a especificidade de HLA-DR1 [O'Sullivan, 1991; Hammer, 1992; Stern, 1994; Murthy, 1997]. A descrição detalhada de cada bolsão será feita conforme a apresentação dos resultados obtidos para cada um dos mesmos.

Como nosso estudo se refere à interação entre o peptídeo e o receptáculo do MHC, os domínios α_2 e β_2 tipo imunoglobulina, os quais não entram em contato com o peptídeo, foram retirados do complexo original. Além disso as extremidades das cadeias polipeptídicas foram fixadas para reproduzir o efeito de estabilização devido à interação entre os domínios α_2 e β_2 , sem os quais a proteína tende a se desenovelar, conforme descrito em "Preparo do arquivo de entrada" (pág. 14), de forma que descreveremos as interações que dizem respeito à interação do receptáculo, formado pelos domínios α_1 e β_1 , com o peptídeo.

Além das interações com as cadeias laterais do peptídeo de Hemaglutinina, há também formação de ligações de Hidrogênio entre as cadeias laterais das cadeias α_1 e β_1 e a cadeia principal do peptídeo, num total de treze ligações de hidrogênio. Após a substituição do peptídeo de hemaglutinina pelo peptídeo de polialanina, verificamos a presença de nove ligações de hidrogênio.

Também verificamos a presença de uma ligação dissulfeto intramolecular entre CYS β 15 e CYS β 79, unindo a α -hélice formada por parte do domínio β_1 com sua mesma porção β -pregueada (fig. 15). Esta ligação foi mantida para todas as estruturas substituídas em todos os bolsões.

Figura 14

Figura 15

Figura 16

Figura 17

Fig. 14. Posicionamento das cadeias laterais que formam os bolsões de DR1.

Fig. 15. Ligação dissulfeto intramolecular entre CYS 79 e CYS 15 da cadeia β unindo a α -hélice e a folha β -pregueada. Esta ligação foi mantida em todos os cálculos de mecânica molecular efetuados neste trabalho.

Fig. 16. Comparação dos posicionamentos relativos de cadeias laterais no interior do bolsão 1 após a otimização de geometria. As cadeias laterais se acomodam no interior do bolsão de acordo com seus tamanhos e geometria.

Fig. 17. Comparação dos posicionamentos relativos de cadeias laterais no interior do bolsão 1 antes da otimização de geometria. As cadeias laterais tem geometrias muito parecidas porque são obtidas da biblioteca de ângulos e distâncias ideais do campo de forças GROMOS96.

4.5 - Posicionamento inicial das cadeias laterais no peptídeo original.

No programa MHCcalc é utilizado o campo de força GROMOS96, conforme descrito em metodologia. Assim, quando os arquivos para um novo complexo são gerados, ou seja, na substituição de um aminoácido no peptídeo de polialanina, o posicionamento de todo o aminoácido, até o carbono beta (exceção feita a glicina), é mantido de acordo com o posicionamento obtido da minimização energética da estrutura com polialanina.

Na substituição de, por exemplo, alanina por glutamina, o posicionamento dos novos átomos a serem incorporados é obtido da biblioteca de ângulos e distâncias de ligação ideais do campo de força GROMOS96. Assim, a otimização geométrica deverá ter como ponto de partida estruturas idênticas, com exceção da cadeia lateral substituída. A otimização geométrica deverá inicialmente modificar o posicionamento dos átomos ao redor da nova cadeia lateral, bem como os próprios átomos da nova cadeia lateral. Em seguida, a modificação do posicionamento dos átomos será propagada pela proteína. Nas figuras 16 e 17 temos o posicionamento inicial das cadeias laterais e o posicionamento obtido após a minimização energética, respectivamente.

4.6 - Estudo do 1º Bolsão.

Considerando o terceiro resíduo do peptídeo de polialanina, que corresponde ao resíduo que interage com o primeiro bolsão do receptáculo, a cadeia lateral está completamente enterrada num bolsão formado pelos domínios α_1 e β_1 do MHC. Este bolsão, chamado P1, já foi descrito por vários autores [O'Sullivan, 1991; Brown, 1993; Hammer, 1992; Stern, 1994; Androulakis, 1997; De Oliveira, 2000; Jardetzky, 1996; Murthy, 1997; Serrano, 1991; Douguerthy, 1996; Georges, 2000] como preferindo resíduos grandes e hidrofóbicos (fig. 18). Além disso, devido ao fato de ser o bolsão mais profundo de todo o receptáculo, parece contribuir com a maior parcela na especificidade entre o peptídeo e a fenda apresentadora e está diretamente ligado ao aparecimento de patologias auto-imunes como diabetes tipo um [Androulakis, 1997] e também artrite reumatóide [De Oliveira, 2000]. Selecionando todos os resíduos que estão dentro

Tabela 1 - Comparação entre a ordem de preferência no primeiro bolsão para DR1 obtidos neste trabalho (MHCalc) e os resultados de ensaios de competição (IC-50) de peptídeos de outros autores.

MHCalc	De Oliveira (2000)	Androulakis (1997) #	Monos (1997)	O' Sullivan (1991)	Hammer (1993)	Geluk (1998)	Marshall (1994)	Jurcevic (1996)	Kobayashi (2000)	Borras-Cuesta (2000)
LYS		TYR	TRP							
LEU	TYR	PHE					PHE			
ASP		TRP	TYR							
TYR		GLN		PHE	TYR	TRP		PHE		
TRP		MET	PHE						#TYR	#TYR
PHE	TRP/PHE	ASN					TYR		#TRP	#PHE
ILE		THR	ILE						#LEU	#TRP
THR		LEU							#ILE	#ILE
VAL		ILE	LEU							#LEU
HIS	ILE	SER					LEU			#VAL
MET		CYS	VAL							
ASN		VAL		TYR	PHE	TYR		LEU		
CYS		ALA	SER							
SER	VAL	GLY					TRP			
ALA		GLU	THR	*GLU		*ARG		*ALA		
PRO		ASP	GLU	*ALA		*ASP		*THR		
ARG		HIS	ASP	*SER				*LYS		
GLN	*ALA	LYS	ARG	*LYS			VAL	*GLN		
GLU		ARG	HIS	*THR				*ASP		
GLY				*GLN						

Resíduos considerados como preferenciais para o bolsão 1, porém sem ordem de preferência.

Resultados de ordem crescente de minimização completa de energia de interação.

* Resíduos com afinidade relativa (IC-50) desprezível no bolsão 1 de DR1, e sem ordem preferência.

de um raio de 5 Å dos átomos do ligante, cujos dados de origem cristalográfica (tirosina) utilizamos para este trabalho, o primeiro bolsão é composto por ILE α 31, TRP α 43, SER β 53, VAL β 85, PHE β 89, ASN β 82, THR β 90, PHE α 32, ALA α 52, PHE α 54, GLY α 86, PHE α 24, GLU α 55 [Androulakis, 1997].

A seguir apresentamos o conjunto de dados obtidos para o bolsão 1, já em ordem crescente de energia de mutação, conforme equações (1), (2) e (3) (página 17).

Resíduo	Complexo	Peptídeo	E _{Int}	E _{Mut}	RMSD cad. A	RMSD cad. B	RMSD cad. C	RMSD BB	RMSD CA
LYS	-1369,04	53,67	-250,94	-139,26	0,26	0,25	0,41	0,27	0,28
LEU	-1356,70	51,52	-236,44	-124,76	0,76	0,25	0,37	0,55	0,56
ASP	-1303,03	50,17	-181,43	-69,75	0,18	0,13	0,39	0,18	0,18
TYR	-1289,07	56,85	-174,15	-62,47	0,38	0,47	0,19	0,42	0,41
TRP	-1285,95	55,98	-170,16	-58,48	0,40	0,45	0,18	0,41	0,40
PHE	-1280,71	57,12	-166,07	-54,39	0,37	0,46	0,17	0,40	0,38
ILE	-1274,90	53,30	-156,43	-44,75	1,07	0,46	0,80	0,81	0,83
THR	-1261,35	50,35	-139,93	-28,25	0,93	0,45	0,77	0,79	0,81
VAL	-1259,13	51,74	-139,10	-27,42	0,91	0,40	0,72	0,75	0,78
HIS	-1252,69	51,76	-132,69	-21,01	0,69	0,79	0,53	0,73	0,76
MET	-1246,36	51,32	-125,91	-14,23	0,30	0,11	0,37	0,24	0,25
ASN	-1238,98	49,91	-117,12	-5,44	0,18	0,12	0,43	0,19	0,19
CYS	-1234,94	51,48	-114,64	-2,96	0,19	0,13	0,41	0,19	0,20
SER	-1231,19	52,44	-111,86	-0,18	0,12	0,05	0,06	0,09	0,09
PRO	-1219,41	59,67	-107,31	4,37	0,69	0,15	0,35	0,48	0,48
ARG	-1225,73	53,37	-107,32	4,36	1,04	0,59	0,92	0,84	0,85
GLN	-1225,38	51,07	-104,68	7,00	1,00	0,35	0,56	0,72	0,75
GLU	-1223,27	50,98	-102,48	9,20	1,00	0,35	0,55	0,72	0,75
GLY	-1206,97	53,92	-89,12	22,56	1,40	0,60	0,90	1,05	1,07

Tabela 2 – Os dados desta tabela são relativos ao bolsão 1, no mesmo formato em que são devolvidos pelo programa MHCcalc. A primeira coluna indica o resíduo aos quais os dados da respectiva linha se referem. A segunda e terceira colunas são os resultados obtidos pelo programa MHCcalc com a otimização de geometria, em kJ/mol, para os complexos com peptídeo de polialanina substituídos com os respectivos resíduos. As colunas restantes são os dados calculados pelo programa MHCcalc, a partir dos resultados de otimização de geometria. A quarta e quinta colunas são a energia de interação e de mutação, respectivamente, calculadas pelo programa MHCcalc conforme equações (1), (2) e (3). As colunas restantes são os RMSD (Root Mean Square Deviation), também obtidas automaticamente, que podem ser úteis para verificar se as estruturas resultantes estão dentro de uma conformação média, próximas à estrutura de partida. Os RMSD foram calculados em relação à estrutura otimizada com polialanina, relativos a cadeia carbônica da cadeia alfa do MHC (coluna 6), cadeia carbônica da cadeia beta do MHC (coluna 7), cadeia carbônica do peptídeo complexado ao MHC (coluna 8), cadeia carbônica do complexo MHC e peptídeo (coluna 9) e carbonos alfa do complexo MHC e peptídeo (coluna 10).

Lisina

De acordo com os nossos resultados o resíduo que tem menor energia de mutação para o bolsão 1 é lisina. A cadeia lateral deste resíduo possui uma grande porção hidrofóbica, a qual

permanece estendida ao longo do bolsão 1 (fig. 19 e 22). Nesta conformação obtida a parte hidrofóbica da cadeia lateral interage com o bolsão, também hidrofóbico. Verificamos que não existem cargas negativas na superfície do bolsão 1 que possam interagir com a extremidade da cadeia lateral carregada positivamente (fig. 20 e 21). No entanto, notamos que existem sete grupos (GLU α 30, ASP α 29, GLU α 47, ASP α 25, GLU α 55, ASP α 27, GLU β 87) carregados negativamente (fig. 19) nas proximidades (11,33 Å de distância média – Tabela 3) da porção positivamente carregada da cadeia lateral, além da carbonila da cadeia principal de ASN β 80 a uma distância de 4,06Å, os quais formam uma espécie de esfera negativa ao redor da porção positivamente carregada de lisina. Dentro deste raio médio não existem cadeias laterais carregadas positiva ou negativamente. Comparando as distâncias destes resíduos negativamente carregados em relação a lisina peptídica, com as distâncias relativas (Distância * X Distância Relativa *** - tabela 3) que os mesmos ocupam quando o bolsão 1 é ocupado por um resíduo negativamente carregado (ASP), verificamos que a presença de um resíduo com carga positiva neste bolsão aproxima os resíduos negativamente carregados que circundam o bolsão 1, sendo o inverso também verdadeiro, ou seja, a presença de um resíduo negativamente carregado no interior deste bolsão afasta os resíduos negativamente carregados que circundam o bolsão 1.

Resíduo	Distância * (Å)	Distância Relativa ***(Å)
GLU α 30	10,98	11,06
ASP α 29	12,52	12,73
GLU α 47	12,73	12,75
ASP α 27	11,45	11,70
ASP α 25	10,19	10,51
GLU α 55	10,22	10,34
GLU β 87	11,21	11,36
ASN β 80	4,06 **	4,08

Tabela 3 – Distâncias entre lisina (bolsão 1) do peptídeo complexado ao MHC e resíduos com cargas negativas das proximidades.

* As distâncias são consideradas entre o centro de cargas negativas do resíduo apresentado na tabela e o centro de cargas positivas (nitrogênio da cadeia lateral) de lisina peptídica.

** Distância entre o nitrogênio da cadeia lateral de lisina e o oxigênio da carbonila da cadeia carbônica de ASN β 80.

*** Obtivemos estas distâncias sobrepondo o peptídeo do complexo com lisina sobre o complexo obtido com ácido aspártico, de forma que ambos os peptídeos (com ácido aspártico e com lisina) estivessem no MHC obtido com ácido aspártico. As distâncias foram medidas em relação ao centro de carga positiva da cadeia lateral de lisina.

A proximidade relativa do conjunto de cargas negativas em relação à porção positiva-

Figura 18

Figura 19

Figura 20

Figura 21

Figura 22

Fig. 18. Representação do posicionamento relativo das cadeias laterais que compõem o bolsão 1, onde boa parte das cadeias laterais é aromática [Brookhaven Protein Data Bank, entrada 1DLH; Stern, 1994].

Fig. 19. Posicionamento relativo da cadeia lateral de lisina do peptídeo polialanina no bolsão 1. Ao redor da extremidade positivamente carregada de lisina peptídica (rosa), existem cadeias laterais negativamente carregadas (tabela 3) formando uma espécie de esfera.

Fig. 20 e 21. Representação da cadeia lateral de LYS do peptídeo de polialanina em relação à superfície acessível ao solvente do bolsão 1 das cadeias α e β , respectivamente. Em ambas as superfícies do bolsão 1 não existem cargas positivas ou negativas que possam interagir com a extremidade positivamente carregada de lisina.

Fig. 22. Posicionamento relativo das cadeias laterais de lisina e alanina no bolsão 1. O posicionamento do carbono α praticamente não é modificado pela otimização de geometria (tabela 2) porque este bolsão é amplo e pode acomodar cadeias laterais grandes (Stern, 1994).

mente carregada da cadeia lateral de lisina peptídica, e também a interação entre a porção hidrofóbica da cadeia lateral peptídica e o bolsão, aparentemente contribuem para a estabilização desta cadeia lateral em relação a alanina. A contribuição energética de pares iônicos presentes no interior de uma proteína pode ser estimada a partir da expressão matemática que resulta da lei de Coulomb. Pelo fato de as cadeias laterais de peptídeos que interagem com o bolsão 1 ficarem completamente enterradas no mesmo, e devido ao fato de este bolsão ser considerado hidrofóbico [Stern, 1994], consideraremos a constante dielétrica no interior do bolsão aproximadamente constante. Considerando o valor de k igual a $9 \cdot 10^9 \text{ J.m.C}^{-2}$, a carga elétrica elementar $1,6 \cdot 10^{-19} \text{ C}$, e a constante dielétrica no interior de uma proteína igual a 4 [Burley, 1988; Clothia, 1984; Voet, 1995], para uma distância média de $11,33 \text{ \AA}$ teremos uma contribuição na energia potencial de $-30,6 \text{ kJ/mol}$. Se considerarmos os sete grupos negativamente carregados, chegamos a um valor de $-214,2 \text{ kJ/mol}$. Este valor ultrapassa a energia de interação obtida para lisina em relação a alanina ($-139,26 \text{ kJ/mol}$), obviamente por não estarmos levando em consideração interações repulsivas entre as cadeias laterais negativamente carregadas em questão, e que cada carga negativa pode estar interagindo com outras cargas positivas, além de interações com grupos polares das proximidades. O posicionamento do carbono α obtido para lisina se mantém em relação a alanina, (fig. 22; tabela 2), o que indica que, após a substituição, não houve necessidade de acomodação de todo o peptídeo, mas apenas da cadeia lateral peptídica de lisina no bolsão.

Este resultado não está de acordo com os dados da literatura [O'Sullivan, 1991; Jurcevic, 1996; Androulakis, 1997; tabela 1], onde a presença de lisina em um peptídeo nesta posição impede que o mesmo seja complexado ao MHC. Nossos resultados são uma medida da energia potencial relativa a uma cadeia lateral de alanina e não levam em conta a cinética de complexação do peptídeo ao MHC; indicam que a cadeia lateral de lisina tem melhor energia de interação que alanina e a melhor energia de mutação entre os aminoácidos de ocorrência natural, mas não fazem qualquer referência a fatores cinéticos que porventura desempenhem um papel na ligação das duas espécies. Desta forma os nossos dados não levam em conta a barreira energética necessária para fazer a extremidade positivamente carregada desta cadeia

lateral sair do solvente aquoso e passar por todo o bolsão 1, de caráter hidrofóbico, e chegar ao fundo do mesmo. Assim, quando a cadeia lateral de lisina peptídica se complexar ao bolsão 1, deve ser levado em consideração o fato de a carga positiva da extremidade da cadeia lateral de lisina ter melhor interação com o solvente aquoso do que com o interior do bolsão hidrofóbico, o que dificultaria a entrada da cadeia lateral no bolsão 1, explicando a diferença entre nossos resultados e os resultados de ensaios de competição da literatura. Androulakis [1997] afirma que resíduos carregados positivamente têm energias relativas de interação mais altas que os demais resíduos, colocando resíduos com carga positiva em últimos lugares em sua escala de preferência para o bolsão 1 (tabela 1). Esta diferença se deve ao fato de o campo de forças utilizado para tais cálculos (ECEPP/3) considerar explicitamente o que no campo de forças GROMOS96, por nós utilizado, está incluso implicitamente, como por exemplo ligações de hidrogênio e energia de solvatação. Além disso, o campo de forças GROMOS96 utiliza o modelo do átomo unido, primando pela simplicidade do cálculo, o que o torna utilizável em computadores tipo PC e em programas de dinâmica e mecânica molecular para sistemas operacionais de código aberto, o que não é possível com a grande maioria dos outros programas e campos de força, como o utilizado por Androulakis. Campos de força mais complexos não indicam melhor qualidade de dados [van Gunsteren, 1992]. A questão da qualidade geral de um campo de força não pode ser respondida facilmente, pois isto depende do tipo de sistema molecular estudado e se o campo de força descreve corretamente os tipos atômicos utilizados no cálculo. Outra diferença crucial entre os nossos dados e os apresentados por Androulakis é o critério de minimização de energia escolhido. O autor relata que seus dados são resultado de minimização global de energia potencial, enquanto que os nossos se referem a mínimos locais obtidos a partir da conformação biologicamente ativa, conforme descrito na introdução (itens 1.2.1 e 1.2.2). Mínimos globais, como o próprio Androulakis relata, podem não ter conformação próxima à biologicamente ativa, enquanto que mínimos locais obtidos a partir da conformação biologicamente ativa asseguram maior proximidade entre a conformação inicial e a final.

Leucina

Em segundo lugar na escala de preferência do bolsão 1 temos leucina, que possui cadeia lateral hidrofóbica (figura 23). Uma boa energia de mutação para este resíduo já era esperada, uma vez que o bolsão é também predominantemente hidrofóbico. Este resultado está de acordo com os dados obtidos por Marshall [1994], Kobayashi [2000], Borrás-Cuesta [2000], Jurcevic [1996] e Monos [1997], que colocam este resíduo entre os cinco primeiros resíduos preferenciais em suas escalas de preferência. Os autores da literatura citada não concordam entre si quanto ao posicionamento exato desta cadeia lateral nas escalas de preferência para o bolsão 1, devido às próprias limitações do método utilizado para o estabelecimento da escala de preferência e também devido à utilização de diferentes peptídeos para a realização dos ensaios de competição de ligação com a fenda apresentadora. Androulakis [1997] coloca este resíduo em oitavo lugar em sua escala de preferência, o que novamente evidencia as diferenças entre os campos de forças e critérios de minimização de energia utilizados por este autor e por nós, sendo que a comparação com os dados experimentais aponta a nosso favor.

Ácido Aspártico

Surpreendentemente, em terceiro lugar na escala de preferência obtivemos ácido aspártico (figura 24), que possui cadeia lateral negativamente carregada. Verificamos que não existem cargas positivas na superfície do bolsão 1 que pudessem interagir diretamente com a carga negativa da cadeia lateral de ácido aspártico do peptídeo (figuras 20 e 21). Comparamos a conformação desta cadeia lateral com as obtidas para as cadeias laterais de leucina (figura 25) e asparagina (figura 26), que possuem tamanhos equivalentes. Verificamos o posicionamento do carbono α de ambas as cadeias laterais e observamos que a cadeia lateral de ASP foi acomodada um pouco mais para dentro do bolsão 1 que as cadeias laterais de LEU e ASN, através da otimização de geometria executada pelo programa THORMM [Moret, 1998]. Como leucina possui cadeia lateral hidrofóbica, e o bolsão também é hidrofóbico, sua cadeia lateral deveria estar acomodada mais para o interior do bolsão do que a cadeia lateral de ASP, que possui cadeia lateral negativamente carregada, e que, portanto, deveria preferir o contato com o solvente aquoso. Além disso era esperado comportamento semelhante para as cadeias laterais

Fig. 23. Representação do posicionamento do resíduo de LEU do peptídeo polialanina no bolsão 1. A cadeia lateral deste resíduo é hidrofóbica, assim como o bolsão 1.

Figura 24

Fig. 24. Posicionamento relativo da cadeia lateral de ASP, presente no peptídeo de polialanina substituído, no interior do bolsão 1. Nesta figura é mostrado o posicionamento relativo de resíduos com cadeia lateral positivamente carregadas, bem como a distância entre eles. A cadeia lateral de ASP, negativamente carregada, se situa no centro de uma espécie de "círculo" de cargas positivas.

Fig. 25. Posicionamento relativo da cadeia lateral de ASP em comparação com a cadeia lateral de LEU, obtidas com o programa MHCalc para o bolsão 1. A cadeia lateral de aspártico está posicionada mais para o interior do bolsão que a cadeia lateral de leucina.

Fig. 26. Posicionamento relativo da cadeia lateral de ASP em comparação com a cadeia lateral de ASN, obtidas com o programa MHCalc para o bolsão 1. A cadeia lateral de aspártico está posicionada mais para o interior do bolsão que a cadeia lateral de asparagina.

de ASN e ASP, por serem ambas mais hidrofílicas que a cadeia lateral de leucina. Verificamos então a presença de resíduos com cadeias laterais positivamente carregadas nas proximidades do bolsão 1 com distância média de 14 Å do centro de carga negativa (carbono da carboxila de ASP peptídica), que podem interagir atrativamente com cadeias laterais negativas colocadas neste bolsão (figura 24), formando uma espécie de esfera positiva ao redor do bolsão 1. Comparando as distâncias destes resíduos positivamente carregados com as distâncias relativas (Distância * X Distância Relativa ** - tabela 4) que os mesmos ocupam quando o bolsão 1 é ocupado por um resíduo positivamente carregado (LYS), verificamos que a presença de um resíduo com carga negativa neste bolsão aproxima os resíduos positivamente carregados que circundam o bolsão 1, sendo o inverso também verdadeiro, ou seja, a presença de um resíduo positivamente carregado no interior deste bolsão afasta os resíduos positivamente carregados que circundam o bolsão 1.

Resíduo	Distância * (Å)	Distância Relativa ** (Å)
HIS α 33	13,29	13,56
ARG α 50	14,72	14,81
ARG α 44	16,67	17,04
HIS β 81	7,17	7,31
HIS β 5	14,72	14,80
ARG β 71	15,77	15,99
ARG β 80	15,55	15,75

Tabela 4 – Distâncias entre aspartato do peptídeo complexado ao MHC e resíduos com cargas positivas das proximidades.

* As distâncias são consideradas entre o centro de carga positiva do resíduo apresentado na tabela e o centro de cargas negativas (carbono carboxílico da cadeia lateral) de aspártico peptídico.

** Obtivemos estas distâncias sobrepondo o peptídeo do complexo com ácido aspártico sobre o complexo obtido com lisina, de forma que ambos os peptídeos (com ácido aspártico e com lisina) estivessem no MHC obtido com lisina. As distâncias foram medidas em relação ao centro de carga negativa da cadeia lateral de ácido aspártico.

Considerando, novamente, o valor da constante dielétrica no interior de uma proteína como 4 [Burley, 1988; Clothia, 1984; Voet, 1995], o valor de cada carga q como $1,6 \cdot 10^{-19}$ C e o valor da constante de proporcionalidade k como $9 \cdot 10^9$ J.m.C⁻², para um par iônico a uma distância de 14 Å, a contribuição na energia potencial será de -24,6 kJ/mol. Para as 7 interações (tabela 4) teremos um valor de -172,2 kJ/mol, o que ultrapassa o valor de -69,75 kJ/mol obtido para esta cadeia lateral. No interior da esfera formada por resíduos positivamente carregados (14 Å) existe uma esfera formada por resíduos carregados negativamente (11,33 Å – tabela 4). Apenas GLU α 47 e ARG α 50 estão próximos o suficiente (6,10 Å) para terem atração eletrostática

que contribua para a energia de interação. Os outros resíduos, considerando um positivo e outro negativo, estão a distâncias superiores a 16 Å uns dos outros, de forma que há pouca interação entre resíduos com carga positiva e negativa. A contribuição da esfera formada por resíduos com carga negativa é maior (-214,2 kJ/mol) do que a da esfera formada por resíduos positivamente carregados (-172,2 kJ/mol), pelo fato de os resíduos negativamente carregados estarem mais próximos do centro de cargas considerado em cada caso. Por este mesmo motivo lisina tem melhor valor na escala de preferência do bolsão 1 do que aspartato. O valor calculado ultrapassa o valor de energia obtido porque não estamos levando em consideração as repulsões entre as cargas positivas que compõe o bolsão 1, e as interações destas cargas com outras cargas negativas em suas respectivas proximidades, e também o fato de todas as cargas positivas estarem interagindo com a mesma carga negativa. Em nossa estimativa também estamos desconsiderando a penalidade energética advinda da interação desfavorável de uma carga negativa dentro de um bolsão hidrofóbico. Os dados obtidos para ASP no bolsão 1 estão em desacordo com os resultados apresentados pela literatura (tabela 1). Os resultados da literatura obtidos através de ensaios de ligação competitiva de peptídeos [Jurcevic, 1996, Geluk, 1998, Monos, 1997] indicam que a presença de aspártico em um peptídeo correspondente ao bolsão 1 faz com que o mesmo não seja complexado a DR1. No entanto, como para o resultado obtido para LYS, nossos dados são uma medida de energia da cadeia lateral do resíduo já dentro do bolsão, não fazendo qualquer referência a cinética de complexação do peptídeo ao MHC. Desta forma, a cadeia lateral de ASP em solução aquosa certamente terá maior afinidade pelo solvente aquoso, devido à carga negativa de sua cadeia lateral, do que pelo bolsão hidrofóbico. Porém, uma vez que a cadeia lateral de ASP já esteja no interior do bolsão, a mesma apresentará energia de interação relativa menor que ALA, de acordo com os nossos resultados. Androulakis [1997] (tabela 1) apresenta dados de minimização global de energia potencial onde a cadeia lateral de ASP esta em décima sexta colocação na escala de preferência do bolsão 1. A diferença entre os nossos resultados e os de Androulakis se devem, primeiro a utilização de diferentes campos de força, que utilizam diferentes parâmetros para o cálculo da energia potencial do sistema estudado, e segundo aos diferentes critérios de convergência adotados,

sendo o nosso uma medida de mínimos locais a partir da conformação biologicamente ativa, e o de Androulakis uma medida do mínimo global. Conforme já discutido para os resultados de lisina, mínimos globais podem não ter a conformação próxima à biologicamente ativa.

Tirosina , Triptofano e Fenilalanina

Em quarto, quinto e sexto lugares na escala de preferência do bolsão 1 obtivemos tirosina, triptofano e fenilalanina, respectivamente, os quais possuem cadeias laterais volumosas e hidrofóbicas. As conformações obtidas para estas cadeias laterais são semelhantes entre si e também semelhantes à conformação da estrutura de origem cristalográfica de partida (figuras 27, 28 e 29). Considerando como distância ótima de interação entre cadeias laterais aromáticas a faixa de 5.5 a 8 Å [Serrano, 1991], nas conformações obtidas, as cadeias laterais de TYR, TRP e PHE do peptídeo interagem principalmente com as cadeias laterais de PHE α 24, PHE α 32, TRP α 43, PHE α 48, PHE α 54 e PHE β 89 (tabela 5) com interações tipo "T", as quais contribuem em média com -5.4 kJ/mol cada interação [Serrano, 1991], perfazendo um total de -32.4 kJ/mol para estas seis interações.

Resíduos do Bolsão 1	TYR	*TRP	PHE
PHE α 24	5.05	5.45	4.93
PHE α 32	6.06	5.40	5.63
*TRP α 43	6.44	6.49	6.21
PHE α 48	8.60	8.30	8.76
PHE α 54	7.66	7.79	6.75
PHE β 89	6.20	6.04	6.25

Tabela 5 – Distância (Å) entre os centrômeros de cadeias laterais de resíduos aromáticos do bolsão 1 e os centrômeros das cadeias laterais aromáticas dos peptídeos com resíduos de TYR, TRP e PHE no bolsão 1.

* Considerando o centrômero do anel indólico de 6 membros.

Esta contribuição de -34,2 kJ/mol das interações tipo "T" na energia de interação representa uma parcela importante no total de energia de interação (-58 kJ/mol em média) calculada para as cadeias laterais aromáticas. Além destas interações tipo "T" existem ainda as interações hidrofóbicas entre a cadeia lateral e o bolsão, que contribuem em torno de -0,3 kJ/mol para cada contato interatômico, e que são responsáveis pelo restante da energia de interação

Fig. 27. Comparação entre o posicionamento relativo do peptídeo original (Brookhaven Protein Data Bank; Stern, 1994), em verde, e do peptídeo de polialanina substituído com PHE, em vermelho, obtido neste trabalho com o programa MHCalc. O posicionamento relativo da cadeia lateral de fenilalanina no bolsão 1 é parecido com o posicionamento da cadeia lateral de tirosina da estrutura de origem cristalográfica.

Fig. 28. Posicionamento relativo da cadeia lateral de TYR do peptídeo original (estrutura obtida no Protein Data Bank), em verde, com o posicionamento das cadeias laterais de TRP (rosa), PHE (vermelho) e TYR (azul) obtidas com o programa MHCalc, através de otimização de geometria. Todas as cadeias laterais aromáticas obtidas neste trabalho resultaram em um posicionamento semelhante a estrutura de origem cristalográfica (Brookhaven Protein Data Bank; Stern, 1994).

Fig. 29. Comparação entre os posicionamentos das cadeias laterais de PHE (azul) e HIS (vermelho) do peptídeo polialanina, obtidos com o programa MHCalc. Os posicionamentos relativos destas cadeias laterais são diferentes porque HIS possui carga positiva em sua cadeia lateral.

obtida para estas cadeias laterais. Os dados da literatura obtidos com ensaios de competição (IC-50) apontam TYR, TRP e PHE como preferenciais em relação aos demais resíduos na escala de preferência do bolsão 1 [De Oliveira, 2000; Monos, 1997; O'Sullivan, 1991; Hammer, 1993; Geluk, 1998; Marshall, 1994; Jurcevic, 1996; Kobayashi, 2000; Borrás-Cuesta, 2000; Boots, 1997] (ver Tabela 1). Esses dados experimentais não estão em acordo quanto à posição exata de TYR, TRP e PHE na escala de preferência, devido ao fato de os autores terem utilizado diferentes peptídeos nos ensaios de competição, e também devido a pequenas diferenças na metodologia utilizada. Androulakis apresenta resultados de minimização global de energia potencial onde TYR, PHE e TRP aparecem como os três primeiros resíduos em sua escala de preferência para o bolsão 1. Apesar das pequenas diferenças quanto ao posicionamento exato na escala de preferência entre os autores citados, estes resíduos aparecem entre os seis primeiros resíduos preferenciais para o bolsão 1, de forma que os nossos resultados estão em pleno acordo com a literatura.

Isoleucina

A cadeia lateral de isoleucina no bolsão 1 resultou na sétima posição na escala de preferência deste bolsão (figuras 30 e 31). O posicionamento obtido para esta cadeia lateral difere das demais do mesmo bolsão. Isto ocorreu devido a ramificação no carbono β da cadeia lateral, o que adiciona um fator estérico e dificulta a acomodação da cadeia lateral no processo de otimização de geometria. Após a minimização energética o posicionamento resultante da cadeia lateral está profundamente modificado em relação aos outros resíduos no mesmo bolsão (figuras 30 e 31), resultando em maior acomodação das cadeias α , β e do próprio peptídeo (RMSD BB de 0,83). Apesar destas grandes alterações conformacionais o resíduo de ILE pode ser considerado como um dos preferenciais para o bolsão 1 por possuir cadeia lateral hidrofóbica. Os dados da literatura [Borrás-Cuesta, 2000, Kobayashi, 2000, De Oliveira, 2000, Monos, 1997] (tabela 1) indicam que ILE está entre os resíduos preferenciais para o bolsão 1, apesar de não haver concordância quanto a posição exata na escala de preferência, devido a diferenças na utilização do método (IC-50) e também devido ao fato de os autores terem utiliza-

Figura 30

Fig. 30. Comparação entre os posicionamentos relativos das cadeias laterais de LYS, GLU, ASP, ILE, e ALA do peptídeo polialanina no bolsão 1. Pode ser notada a grande diferença no posicionamento da cadeia lateral de ILE (azul claro) em relação às demais cadeias laterais (ALA em vermelho, LYS em rosa, ASP em azul escuro e GLU em verde).

Fig. 31. Comparação entre os posicionamentos relativos das cadeias laterais de VAL (verde), ILE (laranja) e ALA (azul), mostrando que a presença de ramificação no carbono β interfere no posicionamento da cadeia lateral dentro do bolsão 1.

Fig. 32. Posicionamento relativo da cadeia lateral de THR do peptídeo poliALA substituído no bolsão 1, mostrando a aproximação da cadeia lateral de THR α 41 devido a interações dipolo-dipolo.

Fig. 33. Posicionamento relativo da cadeia lateral de ALA do peptídeo poliALA no bolsão 1, para comparação com a fig P1-22a. Pode-se notar o reposicionamento da cadeia lateral de THR α 41 devido a atrações dipolo-dipolo.

do diferentes peptídeos para a realização dos ensaios de competição, conforme já discutido anteriormente.

Treonina

Em oitavo lugar na escala de preferência obtivemos treonina, um resíduo com cadeia lateral média e polar. A conformação obtida para este resíduo apresenta a cadeia lateral direcionada para THR α 41, de forma que os dois resíduos de treonina podem interagir atrativamente com interações dipolo-dipolo. A distância entre o hidrogênio hidroxílico da cadeia lateral de THR peptídico está a uma distância de 5.75 Å do oxigênio da cadeia lateral de THR α 41, e o hidrogênio hidroxílico desta última está a uma distância de 6.68 Å do oxigênio da cadeia lateral do peptídeo. Verificamos que ocorre um reposicionamento da cadeia lateral de THR α 41 em relação a substituição do peptídeo por alanina, de forma que a parte polar é posicionada, durante a otimização de geometria, em direção a cadeia lateral do peptídeo substituído com THR em P1 (figuras 32 e 33). Da mesma forma, a minimização de energia direciona a hidroxila da cadeia lateral peptídica para THR α 41. Comparando o resultado obtido para esta cadeia lateral, com o resultado obtido para valina (tabela 2), verificamos que treonina está apenas uma posição acima na escala de preferência. A cadeia lateral de treonina difere da cadeia lateral de valina apenas pela substituição de um agrupamento metila por um agrupamento hidroxila. Desta forma, inicialmente há uma perda na energia de interação com o bolsão pelo fato de o mesmo ser hidrofóbico [Stern, 1994] e não possuir cargas em sua superfície (figuras 20 e 21) e o agrupamento hidroxila ser polar. Entretanto a interação dipolo permanente-dipolo permanente dupla com THR α 41 aparentemente superou esta perda na energia de interação com alguma vantagem. Os dados da literatura [O'Sullivan, 1991, Jurcevic, 1996] (ver Tabela 1) indicam que a presença desta cadeia lateral em posição correspondente ao bolsão 1 em um peptídeo faz com que o mesmo tenha baixa afinidade relativa pela fenda apresentadora, de forma a impedir a complexação do peptídeo ao MHC. Monos [1997] relata sua escala de preferência para o bolsão 1 onde THR aparece entre os últimos 5 colocados, de forma que este resultado não está em acordo com os dados da literatura. Como já dissemos

anteriormente, ao compararmos nossos resultados com ensaios de competição devemos considerar que nossos dados não consideram barreiras energéticas próprias da cinética de complexação de uma cadeia lateral hidrofóbica, em solução aquosa, com um bolsão hidrofóbico, de forma que a cadeia lateral teria maior afinidade pela solução do que pelo bolsão. Nossos dados são uma medida da energia de interação desta cadeia lateral, considerando que a mesma já se encontra no bolsão 1. Androurakis [1997] relata que THR ocupa o sétimo lugar em sua escala de preferência, calculada através de minimização global de energia de interação. Este resultado está em acordo com nossos dados, apesar de o autor ter utilizado tanto campo de força quanto critério de convergência diferentes.

Valina

Para o bolsão 1 obtivemos valina, que possui cadeia lateral pequena e hidrofóbica, em nono lugar na escala de preferência. A conformação obtida para esta cadeia lateral é parecida com a conformação obtida para ILE (figura 31). As conformações obtidas para ILE e VAL são diferentes devido a ramificação no carbono β , que introduz um fator estérico na acomodação da cadeia lateral no bolsão 1. De acordo com os resultados da literatura [De Oliveira, 2000, Borrascueta, 2000, Monos, 1997, Marshall, 1994] obtidos com ensaios de competição de peptídeos, esta cadeia lateral é considerada como uma das preferenciais para o bolsão 1, não havendo concordância entre os mesmos quanto posição na escala de preferência. Androurakis [1997] apresenta resultados de minimização global de energia de interação, onde VAL aparece em décimo segundo lugar na escala de preferência deste mesmo bolsão. Portanto, nosso resultado para VAL está em pleno acordo com os resultados da literatura.

Histidina, Metionina, Asparagina, Cisteína, Serina, Alanina, Prolina, Arginina, Glutamina, Ácido Glutâmico e Glicina

O restante das cadeias laterais de aminoácidos de ocorrência natural são consideradas como não-preferenciais na escala de preferência do bolsão 1. A maior parte dos autores que

realizam ensaios de competição de peptídeos nem mesmo incluem estes resíduos nos testes, uma vez que já são conhecidos por reduzirem a capacidade do peptídeo de se ligar à fenda apresentadora do MHC, considerando o bolsão 1. Nossos resultados estão em acordo com os dados da literatura pois, de acordo com ambos, estes resíduos são os últimos na escala de preferência deste bolsão. Em especial, podemos citar que os mesmos se encontram abaixo, na escala de preferência por nós apresentada, de resíduos como VAL, ILE, PHE, TYR e TRP, considerados preferenciais tanto neste estudo quanto na literatura citada.

A conformação obtida para histidina diferiu das conformações obtidas para as outras cadeias laterais cíclicas (TYR, PHE e TRP) devido ao fato desta cadeia lateral apresentar carga positiva, e a mesma ser volumosa, diferentemente de LYS que também tem carga positiva (figuras 28 e 29).

Como resultado para metionina obtivemos a conformação representada pela figura 34. Esta cadeia lateral é hidrofóbica e de tamanho médio e está completamente enterrada no bolsão 1. Cadeias laterais menores do que esta, porém com ramificação no carbono β (VAL e ILE) resultaram em um melhor posicionamento na escala de preferência devido ao fato de o bolsão ser volumoso [Stern, 1994], de forma que cadeias laterais lineares, como metionina, têm menor interação do tipo van der Waals.

A conformação obtida para asparagina está representada na figura 35. Nesta conformação o oxigênio amídico da cadeia lateral está a distância de 6.60 Å do centro de cargas positivas de HIS β 81, ocorrendo interação atrativa entre a carga positiva de HIS β 81 e o centro de carga parcial negativa da cadeia lateral de ASN presente no oxigênio amídico. A figura 26 representa a comparação das conformações obtidas para ASN e ASP. Observamos nesta figura comparativa que a cadeia lateral de ASN está em um posicionamento mais externo que ASP, provavelmente devido à atração entre ASP e as cargas positivas próximas à mesma, conforme discussão para a cadeia lateral de ácido aspártico (página 49).

Os resultados obtidos para cisteína e serina estão representados pelas figuras 36, 37 e 38. As cadeias laterais destes resíduos interagem com a cadeia lateral de ASN β 82 com interação tipo dipolo-dipolo. A distância entre o oxigênio da carbonila da cadeia lateral de

Figura 34

Figura 35

Figura 36

Figura 37

Figura 38

Fig. 34. Representação do posicionamento do resíduo de MET do peptídeo de polialanina no bolsão 1. Distância entre o enxofre da cadeia lateral de MET peptídica e um dos nitrogênios de HIS81.

Fig. 35. Posicionamento relativo da cadeia lateral de ASN do peptídeo polialanina no bolsão

Fig. 36. Distância entre o hidrogênio do grupo tiol da cadeia lateral de CYS do peptídeo polialanina no interior do bolsão 1 e o oxigênio da carbonila da cadeia lateral de ASN81.

Fig. 37. Distância entre o hidrogênio hidroxílico da cadeia lateral de SER do peptídeo polialanina no interior do bolsão 1 e o oxigênio da carbonila da cadeia lateral de ASN81.

Fig. 38. Comparação entre os posicionamentos das cadeias laterais de CYS e SER no bolsão 1.

ASN β 82 e o hidrogênio hidroxílico de SER peptídica é de 3.67 Å, enquanto que a distância entre o oxigênio da carbonila da cadeia lateral de ASN β 82 e o hidrogênio do agrupamento tiol da cadeia lateral de CYS é de 4.42 Å. Estas cadeias laterais resultaram em conformações muito parecidas (figura 38), devido as suas propriedades e tamanho também parecidos.

O resultado obtido para prolina, o décimo quinto lugar na escala de preferência, já era esperado uma vez que sua cadeia lateral possui restrição conformacional, ficando a mesma impedida de interagir livremente com o bolsão 1 (figura 39).

Para arginina obtivemos o décimo sexto posicionamento na escala de preferência do bolsão 1. A diferença de posicionamento na escala de preferência em relação a LYS (figura 40), que também possui cadeia lateral grande com carga positiva, pode ser atribuído ao volume do agrupamento guanidino (figuras 41 e 42) da cadeia lateral de arginina. O RMSD médio para esta cadeia lateral é quatro vezes maior do que para LYS, especialmente para a cadeia α do MHC (tabela 2), com quem as cadeias laterais colocadas neste bolsão interagem principalmente [Stern, 1994].

O resultado obtido para ácido glutâmico está representado na figura 43. Nesta conformação observamos a proximidade desta cadeia lateral negativamente carregada a HIS β 81, uma cadeia lateral positivamente carregada. A distância entre os centros de cargas de ambas as cadeias laterais é de 8.31 Å. Apesar da existência deste tipo de interação atrativa e que diminui a energia potencial do complexo, foi obtido uma energia de mutação de 9,20 kJ/mol em relação a ALA, o que indica que a presença desta cadeia lateral aumenta a energia potencial da estrutura em relação a ALA. Para verificarmos porque esta cadeia lateral aumentou a energia de mutação, enquanto ASP, uma cadeia lateral também com carga negativa porém de tamanho menor, diminuiu a energia potencial do complexo, comparamos o posicionamento relativo de ambas (figura 44). Observamos que a cadeia lateral de ASP se encontra mais enterrada no bolsão do que GLU, comparando-se o posicionamento do carbono alfa de ambas as cadeias laterais. Também observamos (figura 22) que o tamanho da cadeia lateral não determina se um resíduo ficará mais ou menos enterrado neste bolsão, pois o bolsão 1 é profundo e amplo [Stern, 1994]. Desta forma o grupo iônico de GLU tem tendência a ficar na mesma altura que o grupo

Fig. 39. Posicionamento relativo da cadeia lateral de PRO do peptídeo polialanina no bolsão 1. A décima-quinta colocação na escala de preferência para o bolsão 1 se deve a restrição conformacional da cadeia lateral, que restringe a interação entre a mesma e o bolsão.

Figura 40

Fig. 40. Comparação entre os posicionamentos relativos das cadeias laterais de LYS e ARG do peptídeo polialanina no bolsão 1. Ambas as cadeias laterais estão estendidas ao longo do bolsão alcançam o fundo do mesmo.

Figura 41

Figura 42

Figs. 41 e 42. Representação do volume ocupado pelas cadeias laterais de LYS e ARG no bolsão 1, respectivamente. Observamos que a cadeia lateral de ARG é muito mais volumosa em sua extremidade, o que causa maior modificação conformacional do bolsão.

Figura 43

Figura 44

Figura 45

Figura 46

Figura 47

Fig. 43. Posicionamento relativo da cadeia lateral de GLU, do peptídeo de poliALA mono-substituído, no interior do bolsão 1. Esta cadeia lateral parece interagir com HISβ81 com interações atrativas tipo dipolo-dipolo.

Fig. 44. Posicionamento comparativo entre as cadeias laterais de ASP e GLU, do peptídeo de poliALA mono-substituído. A cadeia lateral de GLU é empurrada para fora do bolsão na tentativa de posicionar o agrupamento negativamente carregado na mesma posição que ASP.

Fig. 45. Posicionamento relativo da cadeia lateral de GLN, do peptídeo de poliALA mono-substituído, o qual interage com ASNβ82.

Fig. 46. Posicionamento comparativo das cadeias laterais de GLU e GLN, do peptídeo de poliALA mono-substituído. O posicionamento das cadeias laterais praticamente coincidem.

Fig. 47. Representação do posicionamento de GLY do peptídeo polialanina no bolsão 1. Este resíduo não possui cadeia lateral, o que resultou em ultimo lugar na escala de preferência para este resíduo.

iônico de ASP otimizado, pois este último é um posicionamento energeticamente favorável, conforme já discutido anteriormente em aspártico. No entanto a cadeia principal do peptídeo tem tendência em ficar na mesma posição em que já se encontra, devido às interações entre esta e o receptáculo do MHC. Isto causa um conflito de geometria, que é resolvido com um posicionamento intermediário e que resulta em uma das maiores energias de interação, sendo que esta cadeia lateral tem melhor colocação na escala de preferência apenas em relação a glicina, um aminoácido sem cadeia lateral.

A conformação obtida para glutamina esta representada na figura 45. Nesta conformação a cadeia lateral de GLN interage atrativamente com a cadeia lateral de ASN β 82, com interações tipo dipolo-dipolo, a uma distância de 4.75 Å. Comparamos a conformação desta cadeia lateral com a cadeia lateral de GLU (figura 46) e observamos que a conformação é praticamente coincidente. Desta forma, na conformação com GLU obtivemos um oxigênio carboxílico a uma distância de 4.75 Å de um oxigênio carbonílico, que é uma interação repulsiva que aumenta a energia potencial do complexo protéico em estudo.

O resíduo de glicina (figura 47) foi o último colocado na escala de preferência do bolsão 1, com uma energia de mutação positiva de 22,56 kJ/mol, o que demonstra que é energeticamente preferível que o bolsão tenha alguma cadeia lateral em seu interior, mesmo que seja apenas a cadeia lateral de alanina com um agrupamento metil. Esta interação com o bolsão ocorre através das forças de dispersão de London (van der Waals), que apesar de fracas (-0,3 kJ/mol) resultam em um efeito hidrofóbico, o qual leva substâncias apolares a minimizarem seu contato com o solvente aquoso.

5 - Conclusão.

Com exceção das cadeias laterais de LYS e ASP nossos resultados estão em acordo com a literatura que apresenta dados experimentais sobre a preferência dos resíduos de aminoácido de ocorrência natural pelo bolsão 1 da proteína DR1. Deve ser considerado que nesses dados não estão inclusas as barreiras energéticas que as cadeias laterais encontram ao passar da solução aquosa para o bolsão hidrofóbico, o que torna nossos resultados uma medida

relativa da interação entre o bolsão e a cadeia lateral do resíduo ali colocado. De acordo com nossos resultados LYS e ASP, mesmo possuindo cargas em um bolsão hidrofóbico, terão energia de mutação em relação a ALA favoráveis, uma vez que a presença destas cadeias laterais no bolsão 1 diminui a energia de interação do complexo. Para verificar se estas cadeias laterais permanecem no bolsão 1 ao serem colocadas em seu interior, devem ser feitos estudos por dinâmica molecular, onde é considerada a movimentação dos átomos durante determinado espaço de tempo.

O programa MHCcalc foi útil e funcional para a determinação da escala de preferência do bolsão 1 da proteína DR1 para os aminoácidos de ocorrência natural. O programa foi executado uma vez para teste e foi verificado que não houve interrupção do mesmo desde a entrada dos dados, a saber, desde chamada do programa a partir da linha de comando com o nome do arquivo contendo DR1, até a término da execução do mesmo com a devolução da escala de preferência de cada bolsão na forma de tabelas contidas em um arquivo final. Apenas os dados do bolsão 1 foram apresentados pois houve um erro na carga líquida da proteína neste primeiro teste o que impossibilitou a utilização dos dados, de forma que nova execução do programa teve que ser feita para a geração de novos dados, execução esta que não foi concluída para os outros bolsões. Com a obtenção dos dados dos outros bolsões será possível completar o estudo da proteína DR1. Com o restante dos dados poderá ser avaliada a necessidade de incluir no programa MHCcalc a execução automática de alguns passos de dinâmica molecular para verificação da estabilidade de cadeias laterais no interior dos bolsões. As opções de cálculo e o campo de forças utilizado foram apropriados para a determinação de uma escala de preferência para o bolsão 1, mas somente o estudo completo de todos os bolsões de MHC-DR1 poderá, de maneira definitiva, ratificar ou indicar a necessidade de modificação destes parâmetros.

6 – Referências Bibliográficas

Altuvia, Y., Settle, A., Sidney, J., Southwood, S., Margalit, H., A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets, *Human Immunology* 58(1): 1-11, 1997.

- Androulakis**, I.P., Nakayak, N.N., Ieraperitrou, M.G., Monos, D.S., Floudas, C.A., A predictive method for the evaluation of peptide binding in pocket 1 of HLA-DRB1 via global minimization of energy interactions, *Proteins* 29: 87-102, 1997.
- Auger**, I., Roudier, J., Autoimmunity in arthritis rheumatoid, *Journal of Rheumatology* 24: 227-228, 1997.
- Bacon**, D.J., Moulton, J., Docking by least-square fitting of molecular surface patterns, *Journal of Molecular Biology* 225: 849-858, 1992.
- Boots**, A.M.H., Verhaijden, G.F.M., Schoningh, R., Staveren, C.J., Bos, E., Elewaut, D., Keyser, F., Veys, E., Joosten, I., Rijnders, A.W.M., Selection of self-reactive peptides within human aggrecan by use of a HLA-DRB1*0401 peptide binding motif, *Journal of Autoimmunity* 10: 569-578, 1997.
- Borras-Cuesta**, F., Golvano, J., Garcia-Granero, M., Sarobe, P., Riezu-Boj, J., Huarte, E., Lasarte, J., Specific and general HLA-DR binding motifs: comparison of algorithms, *Human Immunology* 61: 266-278, 2000.
- Brookhaven Protein Data Bank**, código de entrada 1DLH. <http://www.rcsb.org>.
- Brown**, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., Wiley, D.C., Three dimensional structure of the human class 2 histocompatibility antigen HLA-DR1, *Nature* 364: 33, 1993.
- Brusic**, V., Rudy, G., Honeyman, M., Hammer, J., Harrison, L., Prediction of MHC class 2-binding peptides using an evolutionary algorithm and artificial neural network, *Bioinformatics* 14(2): 121-130, 1998.
- Burley**, S.K., Petzko, G.A., Weakly polar interactions in proteins, *Advances in Protein Chemistry* 39: 125-189, 1988.
- Burley**, S.K., Petzko, G.A., Amino-aromatic interactions in proteins, *FEBS* 203: 139-143, 1986.
- Burley**, S.K., Petzko, G.A., Aromatic-aromatic interaction: a mechanism of protein structure stabilization., *Science* 229: 23-28, 1985.
- Cafilisch**, A., Niederer, P., Anliker, M., Monte Carlo docking of oligopeptides to proteins, *Proteins* 13: 223-230, 1992.

- Chicz**, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., Strominger, J.L., Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size, *Nature* 358: 764-768, 1992.
- Clothia**, C., Principles that determine the structures of proteins, *Annual Review of Biochemistry* 53: 537-572, 1984.
- Connolly**, M.L., Solvent-accessible surfaces of proteins and nucleic acids, *Science* 221: 709, 1983.
- Cunha-Neto**, E., Coelho, V.P.C., Guilherme, L., Fiorelli, A., Stolf, N., Kalil, J., Autoimmunity in Chagas' disease: identification of cardiac myosin-B13 *Trypanosoma cruzi* protein crossreactive T cell clones in heart lesions of a chronic Chagas' cardiomyopathy patient, *Journal of Clinical Investigation* 98(8): 1709-1712, 1996.
- Cunha-Neto**, E., MHC restricted antigen presentation and recognition: constraints on gene, recombinant and peptide vaccines in humans, *Brazilian Journal of Medical and Biological Research* 32: 1-7, 1999.
- De Oliveira**, D.B., Harfouch-Hammoud, E., Otto, H., Papandreou, N. A., Stern, L.J., Cohen, H., Boehm, B.O., Bach, J.M., Caillat-Zucman, S., Walk, T., Jung, G., Eliopoulos, E., Papadopoulos, G.K., Endert, P.M., Structural analysis of two HLA-DR-presented autoantigenic epitopes: crucial role of peripheral but not central peptide residues for T-cell receptor recognition, *Molecular Immunology* 37: 813-825, 2000.
- Demotz**, S., The relative abundance of two measles virus fusion protein peptide-DR1 complexes expressed by B cells is independent of the form of the antigen, *Cellular Immunology* 193: 202-208, 1999.
- Dessen**, A., Laurence, C.M., Cupo, S., Zaller, D.M., Wiley, D.C., X-Ray crystal structure of HLA-DR4 complexed with a peptide from collagen 2, *Immunity* 7: 473-481, 1997.
- Dougherty**, D.A., Cation- π interactions in chemistry and biology: a new view of benzene, PHE, TYR and TRP, *Science* 271: 163-168, 1996.
- Fletcher**, R., *Practical Methods of Optimization*, Wiley, 56-110, 1980.

- Frauenfelder**, H., Sligar, S.G., Wolynes, P.G., The energy landscapes and motions of proteins, *Science* 254: 1598, 1991.
- Frenkel**, D., Smit, B., *Understanding Molecular Simulations: from algorithms to applications*, Academic Press, San Diego, 9-15, 1996.
- Fridkis-Hareli**, M., Strominger, J.L., Promiscuous binding of synthetic copolymer 1 to purified HLA-DR molecules, *Journal of Immunology* 160: 4386-4397, 1998.
- Geluk**, A., Meijgaarden, K.E., Roep, B.O., Ottenhof, T.H.M., Altered peptide ligands of islet autoantigen imogen 38 inhibit antigenic specific T cell reactivity in human type-1 diabetes, *Journal of Autoimmunity* 11: 353-361, 1998.
- Georges**, B., Loing, E., Neveu, R., Melnyk, O., Gras-Masse, H., Auriault, C., Structural diversity of human class 2 histocompatibility molecules induced by peptide ligands, *FEBS* 481: 249-254, 2000.
- Germain**, R.N., Magulies, D.H., The biochemistry and cell biology of antigen processing and presentation, *Annual Review in Immunology* 11: 403-450, 1993.
- Goldstein**, M., Viscous liquids and the glass transition: a potential energy barrier picture, *Journal of Chemical Physics* 51: 3728, 1969.
- Goodsell**, L., Olson, G., Automated docking of substrates to proteins by simulated annealing, *Proteins* 8: 195-202, 1990.
- Ghosh**, P., Amaya, M., Mellis, E., Wiley, D.C., The structure of an intermediate in class 2 MHC maturation: CLIP bound to HLA-DR3, *Nature* 378: 457-462, 1995.
- Grant**, G.H., Richards, W.G., *Computational Chemistry*, Oxford University Press, Oxford, 32-46, 1995.
- Gulukota**, K., Sidney, J., Sette, A., Delisi, C., Two complementary methods for predicting peptides binding major histocompatibility complex molecules, *Journal of Molecular Biology* 267: 1258-1267, 1997.
- Hammer**, J., Belunis, C., Bolin, D., Papadopoulos, J., Walsky, R., Higelin, J., Danhos, W., Sinigaglia, F., Nagy, Z., High-affinity binding of short peptides to major histocompatibility complex class 2 molecules by anchor combinations, *Immunology* 91: 4456-4460, 1994.

Hammer, J., Valsasini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B., Sinigaglia, F., Promiscuous and allele specific anchors in HLA-DR-binding peptides, *Cell* 74: 197-203, 1993.

Hammer, J., Takacs, B., Sinigaglia, F., Identification of a motif for HLA-DR1 binding peptides using M13 display libraries, *Journal of Experimental Medicine* 176: 1007-1013, 1992.

Hart, A., Read, B.N., A multiple-start Monte-Carlo docking method, *Proteins* 13: 206-222, 1992.

Insight II User Guide, Outubro de 1995. San Diego: Biosym/MSI, 1995.

Jardetzky, T.S., Brown, J. H., Gorga, J.C., Stern, L.J., Urban, R.G., Storminger, J.L., Wiley, D.C., Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline 2-like conformation for bound peptides, *Proc. Natl. Acad. Sci. USA* 93: 734-738, 1996.

Jiang, F., Kim, S.H., Soft docking: Matching of molecular surface cubes, *Journal of Molecular Biology* 219: 79-102, 1991.

Jurcevic, S., Travers, P.J., Hills, A., Agrewala, J.N., Moreno, C., Ivanyi, J., Distinct conformations of a peptide bound to HLA-DR1 or DRB5*0101 suggested by molecular modeling, *International Immunology* 8(11): 1807-1814.

Karplus, M., Petsko, G.A., Molecular dynamics simulations in biology, *Nature* 347: 631, 1990.

Kernighan, B., Ritchie, D., *The C Programming Language*, Prentice Hall, 1988.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., Optimization by simulated annealing, *Science* 220: 671, 1983.

Kobayashi, H., Wood, M., Song, Y., Appella, E., Celis, E., Defining promiscuous MHC class 2 helper T cell epitopes for the HER2/neu tumor antigen, *Cancer Research* 60: 5228-5236, 2000.

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E., A geometrical approach to macromolecule-ligand interactions, *Journal of Molecular Biology* 161:269-288, 1982.

Lee, B., Richards, F.M., The interpretation of protein structures: estimation of static accessibility, *Journal of Molecular Biology* 55: 379-400, 1971.

Mallios, R.R., An interactive algorithm for converting a class 2 MHC binding motif into a quantitative predictive model, *Comput. Appl. Biosciences* 13(3): 211-215, 1997.

Mamitsuka, H., Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models, *Proteins* 33(4):460-474, 1998.

Marshall, K.W., Liu, A.F., Canales, J., Perahia, B., Jorgensen, B., Gantzios, R.D., Aguilar, B., Devaux, B., Rothbard, J.B., Role of the polymorphic residues in HLA-DR molecules in allele specific binding of peptide ligands, *Journal of Immunology* 152: 4946-4957, 1994.

Meister, G.E., Roberts, C.G., Berzofsky, J.A., De Groot, A.S., Two novel T cell epitope prediction algorithms based on MHC-binding motif comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences, *Vaccine* 13(6): 581-591, 1995.

Monos, D., Soulika, A., Argyris, E., Gorga, J., Stern, L., Magafa, V., Cordopatis, P., Androulakis, I.P., Floudas, C.A., HLA diversity: function and medical implications, *Proc. Int. Histocomp. World Conf.* 12: 1997.

Moret, M.A., Pascutti, P.G., Bisch, P.M., Munding, K.C., Stochastic molecular optimization using generalized simulated annealing, *Journal of Computational Chemistry* 19: 647-657, 1998.

Murthy, V.L., Stern, L.J., The class 2 MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding, *Structure* 5: 1385-1396, 1997.

Nelson, C.A., Petzold, S.J., Unanue, E.R., Identification of 2 distinct properties of class 2 major histocompatibility complex associated peptides, *Proc. Natl. Acad. Sci. USA* 90: 1227-1231, 1993.

O'Sullivan, D., Sidney, J., Appella, E., Walker, L., Phillips, L., J., Colon, S.M., Miles, C., Chesnut, R.W., Sette, A., Characterization of the specificity of peptide binding to four DR haplotypes, *Journal of Immunology* 145: 1799-1808, 1990.

O'Sullivan, D., Sidney, J., Del Guercio, M.F., Colon, S.M., Sette, A., Truncation analysis of several DR binding epitopes, *Journal of Immunology* 146: 1240-1246, 1991.

Peakman, M., Vergani, D., *Imunologia Básica e Clínica*, Ganabara Koogan, Rio de Janeiro, 34-69, 1999.

Press, W.H., Teukolsky, S.A., Vetterling W.T., Flannery B.P., *Numerical Recipes in Fortran*, Cambridge University Press, 1992.

- Quaratino, S.,** Thorpe, C.J., Travers, P.J., Londei, M., Similar antigenic surfaces, rather than sequence homology, dictate T cell epitope molecular mimicry, *Proc. Natl. Acad. Sci. USA* 92: 10398-100402, 1995.
- Rosenfeld, R.,** Zheng, Q., Vajda, S., Delisi, C., Computing the structure of bound peptides: applications to antigen recognition by class I major histocompatibility complex receptors, *Journal of Molecular Biology* 234: 515-521, 1993.
- Routsias, J.,** Papadopoulos, G.K., Polymorphic structural features of modeled HLA-DQ segregate according to susceptibility or resistance to IDDM, *Diabetology* 38: 1251-1261, 1995.
- Rudensky A.,** Hurlburt P., Hong, S.C., Barlow, A., Janeway Jr., C.A., Sequence analysis of peptides bound to MHC class 2 molecules, *Nature* 353: 622-627, 1991.
- Sanjeevi, C.B.,** Lybrand, T.P., DeWeese, C., Landin-Olsson, M., Kockum, I., Dahlquist, G., Sundkvist, G., Stenger, D., Lernmark, A., Polymorphic amino acid variations in HLA-DQ are associated with systematic physical property changes and occurrence of IDDM, *Diabetes* 44: 125-131, 1995.
- Savoie, C.J.,** Kamikawaji, N., Sasazuki, T., Kuhara, S., Use of bonsai decision trees for the identification of potential MHC class peptide epitope motifs, *Pac. Sump. Biocomp.* 64: 182-189, 1999.
- Serrano, L.,** Bycroft, M., Fersht, A.R., Aromatic-aromatic interactions and protein stability: investigation by double-mutant cycles, *Journal of Molecular Biology* 218: 465-457, 1991.
- Siklodi, B.,** Vogt, A.B., Kropshofer, H., Falcioni, F., Molina, M., Bolin, D.R., Campbell, R., Hammerling, G.J., Nagy, Z.A., Binding affinity independent contribution of peptide length to the stability of peptide-HLA-DR complexes in live antigen presenting cells, *Human Immunology* 59: 463-471, 1998.
- Southwood, S.,** Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Several common HLA-DR1 types share largely overlapping peptide binding repertoires, *Journal of Immunology* 160: 3363-3373, 1998.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Storminger, J.L., Wiley, D.C., Crystal structure of the human class 2 MHC protein HLA-DR1 complex with an influenza virus peptide, *Nature* 368:215-221, 1994.

Sundberg, E.J., Sawicki, M.W., Southwood, S., Andersen, P.S., Sette, A., Mariuzza, R.A., Minor structural changes in a mutated human melanoma antigen correspond to dramatically enhanced stimulation of a CD4⁺ tumor-infiltrating lymphocyte line, *Journal of Molecular Biology* 319: 449-461, 2002.

van Gunsteren, W.F., Billeter, R.S., Eising, A.A., Hunenberger, P.H., Krueger, P., Mark, A.E., Scott, W.R.P., Tironi, I.G., *Biomolecular Simulation : The GROMOS96 Manual and User Guide*, df Hochschulverlag AG an der ETH Zurich, Zurich, 1996.

van Gunsteren, W.F., Mark, A.E., On the interpretation of biochemical data by molecular dynamics computer simulation, *European Journal of Biochemistry* 204: 947-961, 1992.

Verreck, F.A., Vermeulen, C., Poel, A.V., Jorristma, P., Amons, R., Coligans, J.E., Drijfhout, J.W., Koning, F., The generation of SDS-stable HLA-DR dimmers is independent of efficient peptide binding, *International Immunology* 8: 397-404, 1996.

Voet, D., Voet, J.G., *Biochemistry*, Wiley & Sons Inc., New York, 144-192, 1995.

Wales, D.J., A microscopic basis for the global appearance of energy landscapes, *Science* 293: 2067-2070, 2001.

Wiberg, K.B., A scheme for strain energy minimization. Application to the cycloalkanes, *Journal of American Chemical Society*, 87(5): 1070-1078, 1965.