

**ESTUDO COMPARATIVO DAS VARIABILIDADES
DE PEPTÍDEOS ANTIGÊNICOS E DE PROTEÍNAS
HUMANAS NORMAIS E TUMORAIS E
PROTEÍNAS VIRAIS**

FUNDAÇÃO ANTÔNIO PRUDENTE
ANA MARIA R. S. MARTINI
Coordenadora de Pós-Graduação

MARIÂNGELA DAMETTO

Dissertação apresentada à Fundação Antônio Prudente
para a obtenção do título de Mestre em Ciências.

Área de concentração: Oncologia

Orientador: Dr. Saul Gdanski Jacchieri

São Paulo

2002

EXEMPLAR
ESPECIAL

FICHA CATALOGRÁFICA

Preparada pela Biblioteca do Centro de Tratamento e Pesquisa
Hospital do Câncer A.C. Camargo

Dametto, Mariângela

Estudo comparativo das variabilidades de peptídeos antigênicos e de proteínas humanas normais e tumorais e proteínas virais / Mariângela

Dametto -- São Paulo, 2002

34p.

Dissertação(mestrado)--Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências-Área de concentração: Oncologia.

Orientador: Saul Gdanski Jacchieri.

Descritores: 1. GENES CLASSE I DO COMPLEXO DE HISTOCOMPATIBILIDADE (MHC). 2. EPITOPOS DE LINFÓCITO T/química
3. SUBSTITUIÇÃO DE AMINOÁCIDOS.

Dedico este trabalho a meus pais e irmãos pelo amor, incentivo e toda força e coragem que eles me proporcionaram sempre.

Dedico também à Patrícia, à Cláudia e ao Pedro por tudo que eles nem imaginam que representam para mim.



AGRADECIMENTOS

Ao Dr. Saul Gdanski Jacchieri: pela orientação e aprendizado.

Ao Dr. Paulo Boschov: por todas as ajudas despendidas durante o desenvolvimento deste mestrado.

À Dra. Luísa Lina Villa: pelo acompanhamento de meu trabalho.

À Dra. Zulma Fernandes Peixinho: pelas sugestões dadas a meu trabalho.

Ao Dr. Humberto Torloni: pelas conversas extra-classe.

Ao Dr. Luís Fernando Lima Reis: pela disposição para auxiliar em todas as ocasiões.

Ao pessoal da biblioteca: pela prontidão e pela vontade de ajudar sempre.

À Ana Maria e à Márcia da secretaria: pela ajuda prestada em todos os momentos.

Aos amigos e funcionários do Instituto Ludwig (filial São Paulo): pelo apoio, pela ajuda e amizade.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo): pela bolsa de estudo concedida.

RESUMO

Dametto M. **Estudo comparativo das variabilidades de peptídeos antigênicos e de proteínas humanas normais e tumorais e proteínas virais.** São Paulo; 2002. [Dissertação de Mestrado - Centro de Tratamento e Pesquisa Hospital do Câncer].

Objetivo: Este trabalho tem como objetivo estudar a interação entre o peptídeo e a molécula de MHC de classe I por meio de uma abordagem teórica, tendo como base informações experimentais disponíveis. **Metodologia:** O método utilizado neste trabalho para o estudo da interação peptídeo-MHC foi a comparação das variabilidades dos peptídeos apresentados pela molécula MHC de classe I com as variabilidades das proteínas humanas normais e tumorais e das virais. Para alcançar esse propósito, foram construídas uma matriz de substituição para os peptídeos apresentados por moléculas MHC de classe I, uma matriz de mutação para as proteínas virais, uma matriz de mutação para as proteínas humanas normais e uma matriz de mutação para as proteínas humanas tumorais. Esses cálculos também foram realizados após a divisão dos peptídeos de acordo com suas origens.

Resultados: Os resultados obtidos antes da divisão dos peptídeos segundo suas origens mostram que, geralmente, há maior semelhança entre o padrão de substituição dos peptídeos e o padrão de mutação das proteínas humanas normais, em relação aos padrões de mutação das proteínas virais e das proteínas humanas tumorais. Ou seja, as substituições encontradas mais freqüentemente nos peptídeos também são, de um modo geral, as mutações mais freqüentemente ocorridas nas proteínas humanas normais. Após a classificação quanto à origem dos peptídeos, a semelhança das variabilidades dos peptídeos e das proteínas humanas normais não se mostra tão evidente. **Conclusões:** Os resultados da comparação entre as variabilidades dos peptídeos (sem divisão específica) apresentados por moléculas de MHC de classe I e das proteínas humanas normais e das virais mostram haver uma seleção por parte da molécula de MHC na escolha do peptídeo a ser apresentado, ou seja, peptídeos que contêm determinadas substituições de aminoácidos são preferencialmente apresentados e peptídeos que contêm outras substituições não são apresentados. E as substituições mais encontradas nos peptídeos são, na maioria das vezes, as mesmas mutações observadas em proteínas humanas normais. Os resultados das comparações feitas após a divisão dos peptídeos segundo suas origens não permite uma conclusão precisa, pois provavelmente o reduzido número de peptídeos provenientes humanas normais e tumorais pode estar interferindo com os resultados.

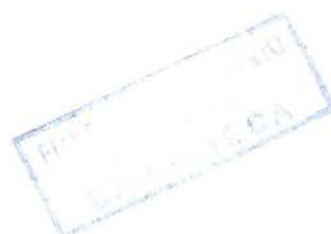
FUNDAÇÃO ANTONIO PRUDENTE
ANA MARIA R. A. KUMAR
Coordenadora de Pós-Graduação

SUMMARY

Dametto M. **Estudo comparativo das variabilidades de peptídeos antigênicos e de proteínas humanas normais e tumorais e proteínas virais [Comparative study of sequence variabilities in antigenic peptides, in normal and tumoral human proteins and in viral proteins]**. São Paulo; 2002. [Dissertação de Mestrado - Centro de Tratamento e Pesquisa Hospital do Câncer].

Objective: A theoretical study of the interaction between peptides and class I MHC molecule that makes use of publicly available experimental information.

Metodology: Comparisons between the rates of mutation found in class I MHC bound peptides and in normal and tumoral human proteins and in viral proteins are established. In order to carry out such comparisons, mutation matrices were calculated in databases containing the above described proteins. These calculations were repeated considering that peptides are originated in specific proteins. **Results:** Previous to the separation of peptides in accordance with their origins in proteins, the results have shown that there is greater resemblance between normal human proteins and peptides. This resemblance indicates that the rates of mutation in peptides and normal human proteins follow similar trends. **Conclusions:** The presently described results indicate that there are mutations frequently found in viral and tumoral proteins that do not correspond to a high frequency in antigenic peptides. On the other hand, rate mutations in these proteins may correspond to high mutation rates in peptides bound to class I MHC molecules.



ÍNDICE

1	INTRODUÇÃO	1
2	OBJETIVO	9
3	MATERIAIS E MÉTODOS	
	Bancos de dados	10
	Proteínas utilizadas	12
	Alinhamento das seqüências	13
	Criação das matrizes de substituição/mutação	13
	Estabelecimento das comparações (gráficos)	16
	Divisão dos peptídeos de acordo com suas origens	17
	Correlação entre taxa de substituição e acessibilidades dos aminoácidos dos peptídeos	17
	Equipamentos utilizados	18
4	RESULTADOS	
	Comparação das variabilidades sem divisão dos peptídeos de acordo com suas origens	22
	Comparação das variabilidades após divisão dos peptídeos de acordo com suas origens	25
5	DISCUSSÃO	27
6	CONCLUSÕES	32
7	REFERÊNCIAS	33
	ANEXOS	
	Anexo 1	A1-1
	Anexo 2	A2-1

1 INTRODUÇÃO

Proteínas existem nas mais diversas formas e possuindo funções bastante variadas, sendo que a função de uma proteína está diretamente relacionada à sua forma. A estrutura de uma proteína é definida por mecanismos que ocorrem durante e após a síntese dessa macromolécula. A sua grande variabilidade também é decorrente das muitas combinações diferentes que podem existir entre as unidades constituintes das proteínas, os aminoácidos. Existem vinte tipos de aminoácidos naturais mais freqüentemente encontrados e uma proteína hipotética que contenha, por exemplo, duzentos aminoácidos poderia ter um número enorme de seqüências. Contudo, o número dessas combinações é reduzido devido a restrições que estão relacionadas a propriedades intrínsecas das cadeias moleculares. Apesar disso, a variabilidade é muito grande e toda essa variabilidade permite que as proteínas possuam um número elevado de funções diferentes, contribuindo para que sejam realizadas as mais diversas atividades necessárias para a existência dos diferentes organismos.

Mas as proteínas não agem sozinhas, sendo que a interação entre elas é imprescindível para a ocorrência dos complexos processos biológicos. Existe um fenômeno conhecido como reconhecimento molecular que é a base para a especificidade da interação entre proteínas (DESIRAJU 2001; LEHN 1993). Ele ocorre em muitos processos biológicos como a interação entre anticorpo e antígeno, enzimas e seus substratos e inibidores, hormônios e seus receptores, uma droga medicinal e seu alvo

biológico e as duas fitas de DNA, entre outros. O reconhecimento molecular ocorre entre duas moléculas que são geometricamente complementares e que podem se ligar por meio de forças não-covalentes, como pontes de hidrogênio e interações eletrostáticas e hidrofóbicas.

A interação entre diferentes proteínas possibilita o funcionamento do organismo como um todo. A compreensão de como as proteínas diferentes interagem é, então, fundamental para que se adquira conhecimento sobre os mecanismos envolvidos nos processos referentes à manutenção da vida. Uma dessas interações ocorre entre a molécula de MHC (*major histocompatibility complex*) de classe I e um peptídeo (8-11 aminoácidos), oriundo de proteínas próprias e não-próprias (proteínas de patógenos, por exemplo) ao organismo (ver figura 1, abaixo).

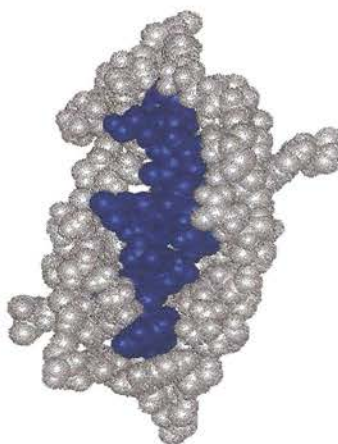


Figura 1. Esta figura representa o complexo formado pelo peptídeo (azul) e a molécula de MHC (cinza).

A ligação da molécula de MHC de classe I com o peptídeo está envolvida em processos que mantêm a homeostase do organismo, como eliminação de células que possuem proteínas sintetizadas de forma incorreta (STOLTZE et al. 2000), por exemplo. Essa ligação está associada, também, a um tipo de resposta imunológica celular, importante na eliminação de patógenos do organismo. O conhecimento sobre essa interação possibilita um aumento da compreensão da resposta imunológica associada, gerando, portanto, maior conhecimento das doenças que envolvem esse tipo de resposta, como, por exemplo, as doenças infecciosas. Dessa forma, torna-se possível o estudo do controle e tratamento dessas e de outras doenças, como o câncer, visto que uma das tentativas para controlá-las pode ser por meio de respostas imunológicas mais acentuadas (potencializadas e específicas), eliminando células infectadas por um agente agressor ou células tumorais.

Diversos trabalhos têm sido realizados no sentido de melhorar a compreensão da interação entre a molécula de MHC e o peptídeo (ZHANG et al. 1998; MICHIELIN et al. 2000). Isso gerou enorme quantidade de dados, como determinação de seqüências dos peptídeos que se ligam a moléculas de MHC e determinação e modelagem de estruturas do complexo peptídeo-MHC. Esses dados têm possibilitado a análise da interação, permitindo verificar quais são as características do bolso da molécula de MHC, quais resíduos do peptídeo são importantes para que ocorra a interação. O presente trabalho tem como objetivo estudar a interação entre o peptídeo e a molécula de MHC de classe I, visando aumentar a compreensão dessa interação.

A molécula de MHC é uma glicoproteína que possui três regiões: uma pequena região citoplasmática, uma região transmembrânica e uma região extracelular, na qual pode se ligar o peptídeo. Essa molécula apresenta peptídeos que são, geralmente, sintetizados endogenamente (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998). Dessa forma, para que o peptídeo fique localizado extracelularmente, ele se liga à molécula de MHC dentro da célula, no retículo endoplasmático (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998). Há alguns eventos que antecedem essa ligação, como a degradação da proteína que origina esse peptídeo. Essa degradação ocorre numa estrutura conhecida como proteassomo, que contém um complexo de proteases que atuam em proteínas, gerando vários peptídeos (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998). Esses peptídeos são transportados até o retículo endoplasmático por proteínas conhecidas como TAP (transportador associado com o processamento de antígeno). No retículo, os peptídeos se ligam a moléculas MHC de classe I e, posteriormente, uma série de eventos ocorre para que haja o transporte do complexo MHC-peptídeo até a posição específica que ocupa na célula (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998).

Uma vez que o peptídeo está ligado na molécula de MHC, esse complexo pode ou não ser reconhecido pelo linfócito T, para desencadeamento de resposta imunológica. Esse processo está envolvido com a imunogenicidade do peptídeo apresentado (YEWDELL e BENNINK 1999). Se o linfócito T reconhecer como estranho esse peptídeo, haverá ativação do linfócito e a resposta imunológica culminará com a destruição da célula portadora desse peptídeo. Um dos mecanismos de destruição da

FUNDAÇÃO DE AMPLIAR O CONHECIMENTO
INSTITUTO DE CIÊNCIAS E TECNOLOGIA
GABINETE DO REITOR

célula é por meio de lise celular, ou seja, o linfócito T ativado libera o conteúdo de seus grânulos, macromoléculas como perforina e granzimas, que promovem a formação de poros na membrana da célula a ser lisada e a morte desta por apoptose.

O peptídeo que se liga à molécula MHC de classe I pode ser proveniente de proteínas próprias do organismo (grande maioria das vezes) ou de proteínas de vírus e outros parasitas, como o *Plasmodium falciparum*, por exemplo. As proteínas virais são, geralmente, apresentadas por moléculas MHC de classe I (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998) e tem sido observado que fragmentos de proteínas tumorais também podem ser apresentados por esse tipo de molécula (SCHULTZE e VONDERHEIDE 2001). Essas são proteínas sintetizadas dentro da célula, seja uma proteína de tumor que é produzida por uma célula tumoral, seja uma proteína viral que é produzida após o vírus ter infectado a célula e utilizado a maquinaria dessa célula para se reproduzir, sintetizando, então, com o auxílio da célula, as proteínas necessárias para sua replicação. Proteínas de bactérias, em geral, são apresentadas por moléculas MHC de classe II, pois as bactérias, sendo fagocitadas, entram em outra via de degradação dentro da célula e se ligam a moléculas MHC de classe II, sendo que estas moléculas não são objeto de estudo deste trabalho.

A molécula de MHC pode se ligar a um conjunto grande de peptídeos porque essas moléculas possuem resíduos de aminoácidos muito polimórficos (os genes do complexo de histocompatibilidade principal são altamente polimórficos) e esses resíduos formam bolsos que acomodam as cadeias laterais de diversos aminoácidos. Entretanto, cada produto de alelo de MHC pode se ligar a um conjunto mais restrito de peptídeos,

sendo que cada molécula de MHC, originada de um determinado alelo, forma bolsos que contribuem para a discriminação dos peptídeos que se ligam nesse alelo. As características dos bolsos como hidrofobicidade, carga e tamanho têm sido descritas como importantes na determinação da ligação de um resíduo de aminoácido específico (MAFFEI e HARRIS 1998). Os resíduos do peptídeo que se ligam a esses bolsos são conhecidos como resíduos-âncora e permanecem, então, mais profundamente ligados à molécula de MHC, estando localizados no terminal amino do peptídeo (posições P2 e P3), no terminal carboxi (P Ω) ou ainda podem estar na região central do peptídeo (posição P5) (ver figura 2, abaixo) (BATALIA e COLLINS 1997; MAFFEI e HARRIS 1998; RUDOLPH e WILSON 2002). Os resíduos de aminoácidos da região central não são, de uma forma geral, resíduos âncora e podem estar posicionados em direção à superfície mais exposta do bolso de ligação para permitir a interação com o receptor do linfócito T citotóxico.

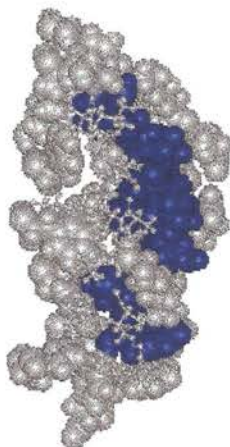


Figura 2. Esta figura mostra mais claramente os resíduos-âncora do peptídeo (azul) ligados mais profundamente à molécula de MHC (cinza), posicionados nos terminais do bolso de ligação, e os resíduos da região central do peptídeo posicionados de uma forma que permanecem mais expostos do bolso.

A proposta deste trabalho é contribuir para o aumento do conhecimento sobre a interação entre a molécula MHC de classe I e o peptídeo, realizando o estudo dessa interação por meio de uma abordagem teórica, tendo em vista dados experimentais disponíveis. A molécula de classe I foi escolhida para este estudo visto que há um maior número de dados experimentais sobre as seqüências dos peptídeos ligados a esse tipo de molécula e também há um número maior de estruturas terciárias do complexo formado entre peptídeos e moléculas de MHC de classe I.

O caminho escolhido para o estudo da interação peptídeo-MHC foi a comparação entre as variabilidades dos peptídeos e das proteínas humanas normais e tumorais e das proteínas virais. Essas comparações foram realizadas com o propósito de verificar se havia algum tipo de seleção na escolha do peptídeo a ser apresentado pela molécula de MHC, visto que os peptídeos apresentados podem ser provenientes de proteínas humanas normais e tumorais e de proteínas virais, mas não é qualquer fragmento dessas proteínas que se liga à molécula de MHC. Assim, essas comparações foram feitas com o intuito de verificar se as substituições que ocorrem nos peptídeos seguem as mesmas regras das mutações de aminoácidos que ocorrem em proteínas humanas normais e tumorais e em proteínas virais. Com isso, é possível analisar quais substituições são importantes para a apresentação de peptídeos pelas moléculas de MHC, ou seja, quais substituições são favoráveis para que os peptídeos sejam apresentados e quais substituições são desfavoráveis à estabilidade da ligação peptídeo-MHC.

As acessibilidades de cada aminoácido presente nos peptídeos também foram calculadas para verificar a existência de correlação entre o número de substituições de cada aminoácido presente no peptídeo e a área exposta desses aminoácidos.



2 OBJETIVO

Estudar a interação entre a molécula de MHC de classe I e o peptídeo apresentado por essa molécula por meio de uma abordagem teórica, utilizando informações experimentais disponíveis.

2.1 OBJETIVOS ESPECÍFICOS

Geração de matrizes de mutação referentes às proteínas humanas normais e tumorais e às proteínas virais e de matriz de substituição de aminoácidos referente aos peptídeos apresentados por moléculas MHC de classe I utilizados neste trabalho.

Analisar a correlação entre os peptídeos e as proteínas consideradas neste trabalho, por meio da construção de gráficos em que se estabelecem as comparações e do cálculo do coeficiente de correlação linear para cada caso.

3 METODOLOGIA

Inicialmente, as seqüências das proteínas utilizadas neste trabalho foram extraídas de três bancos de dados (disponíveis na rede).

- ♦ PDB (BERMAN et al. 2000) - contém estruturas protéicas e foi utilizado para cálculos envolvendo as estruturas terciárias do complexo peptídeo-MHC e para a criação do banco de proteínas humanas normais.
- ♦ SWISS-PROT (<http://www.ebi.ac.uk/swissprot/>) - possui mais de 300000 seqüências de proteínas diversas. Foi utilizado na criação dos bancos de proteínas humanas normais e tumorais e de proteínas virais.
- ♦ MHCPEP (BRUSIC et al. 1998) - formado somente por seqüências de peptídeos que se ligam a moléculas de MHC.

Na fase seguinte, foram construídos os quatro bancos de dados específicos, utilizados neste trabalho: um banco para proteínas humanas normais, um banco para proteínas de vírus, um banco para os peptídeos ligados a moléculas de MHC e um banco de dados para as proteínas humanas tumorais. Essa construção foi feita por meio de um programa de computador desenvolvido para este trabalho e consistiu na separação e no agrupamento de seqüências semelhantes, para permitir a comparação dessas seqüências feita posteriormente (ver, no anexo 1, o quadro com a descrição do programa utilizado

para este passo do trabalho e os arquivos com exemplos da saída de dados desse programa).

O banco de dados de peptídeos é formado por 3412 seqüências, o (banco de dados) de proteínas virais possui 1068 seqüências, o banco de proteínas humanas normais contém 1029 seqüências e o banco de proteínas humanas tumorais contém 369 seqüências. Foram utilizados alguns tipos de proteínas humanas normais, de proteínas do envelope de alguns tipos diferentes de vírus e de peptídeos ligados a moléculas de diferentes alelos de MHC, mostrados no Quadro 1. As proteínas humanas normais e as proteínas virais foram escolhidas baseadas no trabalho realizado por Dayhoff (DAYHOFF et al. 1972), sendo que no presente trabalho foram incluídas algumas das proteínas utilizadas naquele trabalho para que pudesse haver um termo de comparação deste novo trabalho com o anterior.

FLA 123 45 67 89 1011 1213 1415 1617 1819 2021 2223 2425 2627 2829 3031 3233 3435 3637 3839 4041 4243 4445 4647 4849 5051 5253 5455 5657 5859 6061 6263 6465 6667 6869 7071 7273 7475 7677 7879 8081 8283 8485 8687 8889 9091 9293 9495 9697 9899 10000

Quadro 1 - Proteínas humanas normais, vírus e produtos de alelos de HLA utilizados neste trabalho.

Produtos de alelos de HLA		Proteínas humanas normais	Vírus	Proteínas humanas tumorais
A1	B35	insulina	astrovírus	bcr/abl
A11	B37	lisozima	rotavírus	cdc27 mutada
A2	B38	hemoglobina	HIV	cdk4 mutada
A23	B39	citocromo c	vacciniavírus	ctcl
A24	B42	mioglobina	vírus hepatite C e A	gage
A25	B44		adenovírus	hst-2
A26	B45		rinovírus	lage
A28	B46		vírus papiloma humano	mage
A29	B5		poliovírus	mart-1
A3	B51			psa
A30	B52			psm
A31	B53			rage
A32	B54			sage
A33	B55			
A62	B56			
A68	B57			
A69	B58			
B13	B60			
B14	B61			
B15	B62			
B17	B7			
B18	B73			
B27	B8			

Após as seqüências terem sido separadas e agrupadas, foi feito o alinhamento de cada grupo de seqüências. Esse alinhamento foi feito com o auxílio de um programa de computador, Clustalw (THOMPSON et al. 1994), que está disponível na rede. Esse alinhamento foi feito para possibilitar o cálculo da taxa de substituição dos aminoácidos dos peptídeos e das taxas de mutação das proteínas humanas normais e tumorais e das proteínas virais.

O passo seguinte realizado foi a comparação das seqüências das proteínas e peptídeos deste estudo para que se obtivesse o número de substituições/mutações dos aminoácidos referente a cada um dos quatro grupos formados para este trabalho. Era necessário, então, criar um programa que comparasse todas as seqüências (presentes em um dos quatro grupos) entre si, encontrasse as trocas de aminoácidos ocorridas nessas seqüências e contasse todas as trocas encontradas para cada par de aminoácidos (esse procedimento foi feito separadamente para cada um dos quatro grupos deste trabalho). Dessa forma, com o programa criado, foi possível verificar quais substituições de aminoácidos ocorrem em maior ou menor número em cada grupo. Isso foi feito para possibilitar as três comparações feitas neste trabalho: a comparação entre a taxa de mutação dos aminoácidos das proteínas humanas normais e a taxa de substituição dos aminoácidos dos peptídeos ligados a moléculas de MHC, a comparação entre a taxa de substituição dos peptídeos e a taxa de mutação de proteínas virais e a comparação entre a taxa de substituição dos peptídeos e a taxa de substituição das proteínas humanas tumorais. Essas comparações estão explicadas mais adiante.

O programa criado, então, realiza os seguintes passos:

- ◆ Passo 1) lê cada linha do arquivo de seqüências alinhadas gerado no Clustalw (ver, no anexo 1, arquivos com exemplos da saída de dados desse programa), desconsiderando o código da seqüência. As seqüências são lidas como

```
GDRSTDYGIFQINSRYWCNDGKTP-GAVNACHLSCSALLD-----
GDRSTDYGIFQINSRYWCNDGKTP-GAVNACHLSCSKLLD-----
```

- ◆ Passo 2) compara cada seqüência lida com todas as outras seqüências do arquivo. Por exemplo, comparando as quatro seqüências abaixo, vê-se que, na posição marcada com um asterisco (*), existem seis substituições: há uma substituição de leucina (L) (1ª seqüência) por metionina (M) (2ª seqüência), de leucina (1ª seqüência) por valina (V) (3ª seqüência), de leucina (1ª seqüência) por cisteína (C) (4ª seqüência). Também há substituição de metionina (2ª seqüência) por valina (3ª seqüência) e por cisteína (4ª seqüência) e há substituição de valina (3ª seqüência) por cisteína (4ª seqüência).

*

```
1ª seqüência AEKLFNQDVDAAVRGI LRNAKLKP-VYDSLDAVRRALLINMVFQ-----M
2ª seqüência AEKLFNQDVDAAVRGI LRNAKLKP-VYDSLDAVRRAMLINMVFQ-----M
3ª seqüência AEKLFNQDVDAAVRGI LRNAKLKP-VYDSLDAVRRAVLINMVFQ-----M
4ª seqüência AEKLFNQDVDAAVRGI LRNAKLKP-VYDSLDAVRRACLINMVFQ-----M
```


O programa, então, lê e soma essas substituições para cada par de aminoácidos. Dessa forma, são obtidos os números de trocas parciais de cada par de aminoácidos. Para formar a matriz de substituição, esses valores parciais são somados e colocados em forma de matriz (Tabelas 1, 2, 3, 4, 5, 6 e 7 no anexo 2).

Na matriz da tabela 1, por exemplo, estão somadas todas as substituições encontradas nas 3412 seqüências de peptídeos. Essa soma foi feita da seguinte forma (considerando-se somente os alelos A1 e B15 como exemplo):

- ◆ compararam-se somente as seqüências de peptídeos ligados ao produto final do alelo A1 (suponha que foram encontradas 20 substituições do aminoácido lisina pelo aminoácido arginina).
- ◆ essa comparação também foi feita para seqüências de peptídeos ligados ao produto final do alelo B15 (suponha que foram encontradas 35 substituições de lisina por arginina).
- ◆ são somadas, então, as substituições de cada alelo (neste caso, alelos A1 e B15, somando $20+35=55$ substituições de lisina por arginina) para formar o termo da matriz.

Para as proteínas humanas normais, essa soma foi feita da mesma forma (considerando-se somente as hemoglobinas e o citocromo c como exemplo):

- ◆ compararam-se somente as seqüências de hemoglobina entre si (suponha que foram encontradas 35 mutações de lisina por arginina)
- ◆ compararam-se somente as seqüências de citocromo c entre si (suponha que foram encontradas 15 mutações de lisina por arginina)

- ♦ as mutações de cada proteína são somadas (neste caso, hemoglobina e citocromo c, somando $35+15=50$ mutações de lisina por arginina) para formar o termo da matriz.

O mesmo procedimento foi utilizado para as proteínas virais e as proteínas humanas tumorais.

Por fim, foram obtidas quatro matrizes (uma para cada um dos quatro grupos deste estudo): uma matriz de substituição para peptídeos apresentados por moléculas de MHC, uma matriz de mutação para proteínas humanas, uma matriz de mutação para proteínas virais e uma matriz de substituição para as proteínas tumorais.

Para estabelecer as três comparações desejadas neste trabalho, foram construídos gráficos e calculados os coeficientes de correlação linear dos mesmos. Para a comparação entre peptídeos ligados a moléculas de MHC e proteínas humanas normais, por exemplo, foram construídos 20 gráficos, referentes a cada um dos 20 aminoácidos. Cada gráfico representa todas as trocas de um determinado aminoácido, ou seja, as trocas desse aminoácido por todos os outros 19 aminoácidos. Por exemplo, considerando-se o aminoácido isoleucina, foi feito um gráfico contendo os valores do número de substituições de isoleucina por todos os outros 19 aminoácidos nos peptídeos (eixo y) e os valores do número de mutações de isoleucina por todos os outros 19 aminoácidos nas proteínas humanas normais (eixo x). Todos os gráficos foram construídos com os valores logarítmicos ($\log_{10}$) das substituições/mutações dos aminoácidos. Esse mesmo procedimento foi feito para a comparação dos peptídeos e das proteínas de vírus e para a comparação dos peptídeos e das proteínas humanas tumorais.

Posteriormente, os peptídeos ligados a moléculas de MHC foram divididos em peptídeos provenientes de proteínas humanas normais (505 seqüências), peptídeos provenientes de proteínas virais (1591 seqüências) e peptídeos provenientes de proteínas humanas tumorais (310 seqüências). Assim, puderam ser realizadas as comparações mais especificamente entre proteínas humanas normais e peptídeos provenientes de proteínas humanas normais, entre proteínas de vírus e peptídeos provenientes de proteínas virais e entre proteínas humanas tumorais e peptídeos provenientes de proteínas humanas tumorais. E todo o procedimento descrito anteriormente foi feito para a construção das matrizes e dos gráficos necessários para a realização dessas novas comparações.

Neste trabalho também foi analisada a correlação entre a taxa de substituição dos aminoácidos dos peptídeos ligados a moléculas de MHC e a acessibilidade (área exposta) média desses aminoácidos. As acessibilidades foram calculadas utilizando-se as estruturas já determinadas do complexo formado pelo peptídeo e pela molécula de MHC de classe I (depositadas no PDB) e o programa What_if (VRIEND 1990). Essas acessibilidades foram calculadas para cada uma das 77 estruturas do complexo presentes no PDB (até o momento da realização deste trabalho) e, posteriormente, a média da área exposta de cada aminoácido foi calculada somando-se as acessibilidades de um determinado aminoácido encontradas em cada uma dessas estruturas e dividindo-se pelo número de vezes que esse aminoácido estava presente nos peptídeos. Dessa forma, foram obtidas as áreas expostas de todos os aminoácidos que faziam parte dos peptídeos deste estudo. Essas áreas foram utilizadas na construção dos gráficos em que se

comparou a taxa de substituição dos aminoácidos dos peptídeos e as áreas expostas médias desses aminoácidos. Foram gerados dois gráficos, pois se separou os aminoácidos hidrofóbicos dos aminoácidos hidrofílicos visto que se considera que as interações hidrofóbicas são importantes para a estabilidade da ligação do peptídeo com a molécula de MHC (MAFFEI e HARRIS 1998).

Os cálculos deste trabalho foram feitos utilizando-se um computador Silicon Origin 200, um microcomputador com processador Pentium II 300 Hz e um com processador Pentium III 1,1 GHz.

4 RESULTADOS

As matrizes deste trabalho foram construídas com a finalidade de utilizá-las na comparação entre as taxas de substituição dos aminoácidos dos peptídeos ligados a moléculas de MHC e as taxas de mutação das proteínas humanas normais e tumorais e das virais. Essas comparações foram possíveis com a geração dos gráficos descritos a seguir. A análise dos coeficientes de correlação linear de todos os gráficos gerados neste trabalho está colocada na seção de conclusões e discussão.

Os gráficos deste trabalho foram, então, gerados com o intuito de serem estabelecidas as comparações desejadas, ou seja, as comparações entre as taxas de substituição dos aminoácidos dos peptídeos deste estudo e as taxas de mutação das proteínas utilizadas neste trabalho. Dessa forma, cada gráfico representa a comparação entre a taxa de substituição de um determinado aminoácido dos peptídeos e a taxa de mutação do aminoácido correspondente observada nas proteínas humanas normais ou nas proteínas humanas tumorais ou ainda nas proteínas virais. Cada ponto dos gráficos representa as substituições observadas de determinado aminoácido por outro aminoácido, havendo, portanto, 19 pontos em cada gráfico (um ponto para cada um dos 19 aminoácidos pelos quais o aminoácido em questão foi substituído). Os resultados deste trabalho mostram que a correlação existente entre as variabilidades dos peptídeos ligados a moléculas de MHC-I e das proteínas humanas normais é maior do que a correlação entre as variabilidades dos peptídeos e das outras proteínas deste estudo.

A tabela 1 (ver anexo 2) mostra a matriz de substituição para os peptídeos ligados a moléculas de MHC. As matrizes de mutação referente às proteínas humanas normais e às proteínas virais estão mostradas nas tabelas 2 e 3 do anexo 2, respectivamente. A tabela 4 (anexo 2) representa a matriz de substituição das proteínas humanas tumorais. Nessas matrizes pode ser visto que existem algumas substituições que são mais frequentes que outras. Por exemplo, na matriz referente aos peptídeos, o número de substituições do aminoácido triptofano (W) pelo aminoácido cisteína (C) é o mais baixo encontrado nos peptídeos. O número de substituições de leucina (L) por isoleucina (I) é um dos mais altos. Nas matrizes das proteínas humanas normais, das proteínas virais e das proteínas humanas tumorais, a mutação de triptofano (W) por cisteína (C) também é a menos frequente, embora a mutação de leucina (L) por isoleucina (I) não seja a mais frequente.

As figuras de 1 a 12 (ver anexo 2) representam alguns dos gráficos gerados para as três comparações iniciais (considerando-se todos os peptídeos e não os peptídeos separados de acordo com suas origens) feitas neste trabalho. Os gráficos deste trabalho foram construídos com os logaritmos ($\log_{10}$) dos valores das substituições/mutações de aminoácidos. Os gráficos das figuras 1, 3, 5 e 7 foram construídos com os valores das substituições de determinados aminoácidos dos peptídeos e das proteínas humanas normais, ou seja, o gráfico da figura 1 mostra as substituições do aminoácido triptofano (W) referentes aos peptídeos e às proteínas humanas normais. A distribuição dos pontos nesse gráfico mostra que o padrão de substituição do aminoácido triptofano nos

peptídeos apresenta certa semelhança com o padrão de mutação desse aminoácido nas proteínas humanas normais.

O gráfico da figura 3 mostra as substituições do aminoácido isoleucina (I) referentes aos peptídeos e às proteínas humanas normais. Nesse gráfico pode-se observar que há forte linearidade na relação entre as substituições de isoleucina nos peptídeos e as mutações desse aminoácido nas proteínas humanas normais, ou seja, a variabilidade do aminoácido isoleucina apresenta grande semelhança quando se comparam os peptídeos com proteínas humanas normais.

Os gráficos das figuras 5 e 7 mostram as comparações entre peptídeos e proteínas humanas normais para os aminoácidos valina (V) e tirosina (Y), respectivamente. Esse tipo de gráfico foi, então, construído para cada aminoácido específico e também foram feitos os gráficos comparando-se os peptídeos com as proteínas virais (figuras 2, 4, 6 e 8 – anexo 2) e com as proteínas humanas tumorais (figuras 9 a 12). Os coeficientes de correlação linear de todos os gráficos construídos para essas três comparações estão mostrados na tabela 2. Nessa tabela, que mostra os coeficientes de correlação linear para a análise feita utilizando-se os peptídeos como um todo, verifica-se que há um número maior de correlações estatisticamente significativas quando se comparam as proteínas humanas normais e os peptídeos apresentados por moléculas MHC de classe I.

Tabela 2 - Valores dos coeficientes de correlação linear (r) calculados para os gráficos referentes às três comparações em que não há divisão específica da origem dos peptídeos (na designação xH, por exemplo, x representa qualquer um dos 19 aminoácidos naturais pelos quais o aminoácido H pode ser substituído). Nível de significância (p) – correlação é significativa (*) com $p < 0,05$.

Coeficientes de correlação linear entre proteínas humanas normais e peptídeos		Coeficientes de correlação linear entre proteínas virais e peptídeos		Coeficientes de correlação linear entre proteínas humanas tumorais e peptídeos	
r	(p)	r	(p)	r	(p)
0,46 - xH	0,048 *	0,15 - xH	0,540	-0,09 - xF	0,722
0,57 - xL	0,012 *	0,20 - xW	0,405	0,76 - xM	0,756
0,57 - xN	0,011 *	0,30 - xM	0,210	0,10 - xC	0,680
0,59 - xW	0,008 *	0,38 - xN	0,114	0,14 - xP	0,560
0,62 - xR	0,005 *	0,38 - xY	0,112	0,22 - xA	0,371
0,62 - xC	0,004 *	0,39 - xK	0,098	0,26 - xH	0,287
0,67 - xY	0,002 *	0,47 - xQ	0,042 *	0,28 - xY	0,242
0,68 - xM	0,001 *	0,47 - xF	0,040 *	0,32 - xN	0,177
0,75 - xP	0,000 *	0,50 - xV	0,030 *	0,40 - xG	0,089
0,76 - xK	0,000 *	0,56 - xA	0,013 *	0,42 - xR	0,070
0,77 - xA	0,000 *	0,56 - xT	0,012 *	0,43 - xD	0,066
0,79 - xS	0,000 *	0,57 - xC	0,012 *	0,43 - xW	0,070
0,80 - xT	0,000 *	0,57 - xD	0,011 *	0,44 - xS	0,057
0,81 - xD	0,000 *	0,60 - xI	0,007 *	0,46 - xL	0,047 *
0,81 - xV	0,000 *	0,60 - xL	0,007 *	0,54 - xQ	0,018 *
0,82 - xF	0,000 *	0,65 - xP	0,003 *	0,58 - xK	0,009 *
0,83 - xQ	0,000 *	0,65 - xS	0,003 *	0,63 - xT	0,004 *
0,85 - xI	0,000 *	0,68 - xG	0,001 *	0,74 - xI	0,000 *
0,85 - xG	0,000 *	0,68 - xR	0,001 *	0,75 - xE	0,000 *
0,86 - xE	0,000 *	0,73 - xE	0,000 *	0,84 - xV	0,000 *

Esses coeficientes mostram que, de um modo geral, as proteínas humanas normais apresentam padrões de mutação mais parecido com o padrão de substituição dos peptídeos ligados a moléculas de MHC. As proteínas de vírus apresentam um padrão mais diferente do padrão dos peptídeos assim como as proteínas humanas de tumor.

Os resultados da análise feita com a divisão específica dos peptídeos em peptídeos originados de proteínas humanas normais e tumorais e de proteínas de vírus estão mostrados nas tabelas 5, 6 e 7 e figuras 13 a 24 (anexo 2). Nas tabelas 5, 6 e 7 estão representadas as matrizes de substituição para peptídeos provenientes de proteínas humanas normais, de proteínas virais e de proteínas humanas tumorais, respectivamente.

O gráfico da figura 13 representa a variabilidade do aminoácido isoleucina (I) quando é estabelecida uma comparação entre proteínas humanas normais e peptídeos originados de proteínas humanas normais. O gráfico da figura 15 mostra a taxa de substituição do aminoácido triptofano (W) nas proteínas humanas normais e nos peptídeos originados de proteínas humanas normais. As figuras 17 e 19 representam as comparações entre peptídeos originados de proteínas humanas normais e proteínas humanas normais referentes aos aminoácidos valina e tirosina, respectivamente.

A figura 14 representa o resultado da comparação entre proteínas de vírus e peptídeos originados de proteínas virais referente ao aminoácido isoleucina (I) e o gráfico 16 mostra a comparação referente ao aminoácido triptofano (W). Os gráficos 18 e 20 são referentes às comparações das variabilidades dos aminoácidos valina e tirosina, respectivamente, entre peptídeos provenientes de proteínas virais e proteínas de vírus.

Os gráficos das figuras 21 a 24 (anexo 2) mostram os resultados da comparação entre proteínas humanas tumorais e os peptídeos provenientes de proteínas humanas tumorais referentes ao aminoácido isoleucina (I), triptofano (W), valina (V) e tirosina (Y), respectivamente.

Na tabela 3 estão colocados todos os coeficientes das comparações feitas em que se separou os peptídeos de acordo com suas origens. A análise desses coeficientes não permite dizer que há maior semelhança entre as variabilidades dos aminoácidos das proteínas humanas normais com os peptídeos originados de proteínas humanas normais, como foi observado na comparação inicial feita.



Tabela 3 - Valores dos coeficientes de correlação linear calculados para os gráficos referentes às três comparações feitas em que os peptídeos foram divididos de acordo com sua origem (na designação xA, por exemplo, x representa qualquer um dos 19 aminoácidos naturais pelos quais o aminoácido A pode ser substituído). Nível de significância (p) – correlação é significativa (*) com $p < 0,05$.

Coeficientes de correlação linear entre proteínas humanas normais e peptídeos originados de proteínas humanas normais		Coeficientes de correlação linear entre proteínas virais e peptídeos originados de proteínas virais		Coeficientes de correlação linear entre proteínas humanas tumorais e peptídeos originados de proteínas humanas tumorais	
r	(p)	r	(p)	r	(p)
0,03 - xI	0,919	-0,02 - xH	0,928	-0,28 - xP	0,253
0,03 - xS	0,905	0,09 - xT	0,700	-0,09 - xA	0,728
0,12 - xQ	0,636	0,17 - xY	0,495	-0,04 - xC	0,871
0,17 - xK	0,490	0,31 - xM	0,203	0,00 - xF	0,991
0,23 - xC	0,340	0,31 - xN	0,200	0,03 - xL	0,889
0,26 - xF	0,287	0,32 - xA	0,184	0,03 - xN	0,893
0,27 - xR	0,263	0,36 - xQ	0,135	0,08 - xK	0,760
0,37 - xM	0,116	0,38 - xF	0,109	0,11 - xD	0,665
0,41 - xP	0,080	0,39 - xI	0,103	0,13 - xG	0,587
0,42 - xN	0,076	0,39 - xP	0,096	0,16 - xH	0,519
0,43 - xY	0,068	0,39 - xV	0,099	0,17 - xS	0,497
0,46 - xD	0,048 *	0,41 - xC	0,085	0,18 - xR	0,462
0,46 - xW	0,047 *	0,46 - xK	0,049 *	0,22 - xM	0,384
0,49 - xV	0,034 *	0,47 - xG	0,043 *	0,25 - xE	0,295
0,53 - xE	0,020 *	0,51 - xS	0,026 *	0,31 - xY	0,196
0,54 - xT	0,018 *	0,52 - xW	0,022 *	0,32 - xQ	0,178
0,57 - xG	0,011 *	0,56 - xD	0,013 *	0,37 - xW	0,129
0,64 - xL	0,005 *	0,73 - xE	0,000 *	0,46 - xI	0,049 *
0,72 - xA	0,001 *	0,73 - xL	0,000 *	0,46 - xT	0,050
0,74 - xH	0,000 *	0,74 - xR	0,000 *	0,56 - xV	0,010 *

Os gráficos referentes à análise da correlação entre a taxa de substituição dos aminoácidos dos peptídeos e as médias das áreas expostas desses aminoácidos estão mostrados nas figuras 25 e 26 (ver anexo 2). O gráfico 25 refere-se aos aminoácidos hidrofóbicos, que apresentam um coeficiente de correlação linear de -0,95. No gráfico 26 é apresentada essa correlação para os aminoácidos hidrofílicos (coeficiente linear é 0,36). Esses resultados mostram que há uma forte linearidade (negativa) da relação entre média da área exposta e a taxa de substituição dos aminoácidos dos peptídeos somente em relação aos aminoácidos hidrofóbicos. Na tabela 6 (anexo 1) estão representados os valores das médias das áreas expostas para cada aminoácido.

5 DISCUSSÃO

Nos gráficos construídos inicialmente neste trabalho (coeficientes de correlação linear da tabela 2 mostrada nos resultados), pode-se observar que os coeficientes variam de 0,46 a 0,86 nos gráficos proteínas humanas normais x peptídeos e variam de 0,15 a 0,73 nos gráficos proteínas virais x peptídeos. A linearidade relativamente alta encontrada, por exemplo, no gráfico referente à variabilidade de isoleucina nas proteínas humanas normais e nos peptídeos (figura 3, ver anexo 2) mostra que as substituições mais freqüentes de isoleucina nos peptídeos estão relacionadas às mutações mais freqüentes desse mesmo aminoácido observadas nas proteínas humanas normais. Isso significa que a mutação de isoleucina por outro aminoácido, em geral, não interfere com a funcionalidade da proteína humana normal, pois essa mutação é bastante freqüente, e também não deve interferir com a apresentação do peptídeo, visto que essa substituição também é bastante freqüente no peptídeo apresentado pela molécula de MHC.

Os coeficientes dos gráficos em que se comparam as proteínas virais e os peptídeos são, de uma forma geral, mais baixos que os coeficientes dos gráficos feitos na comparação entre proteínas humanas normais e peptídeos. Isso significa que existem algumas mutações que, em geral, não alteram a função da proteína do vírus, pois essa mutação é bastante freqüente nas proteínas virais, mas interferem na apresentação do peptídeo porque a substituição correspondente ocorre com uma freqüência menor nos peptídeos. Dessa forma, existem certas substituições que podem estar sendo importantes

na escolha do peptídeo a ser apresentado pela molécula de MHC, ou seja, uma proteína que apresentar determinada mutação com alta frequência pode não ser apresentada porque a substituição correspondente não é observada tão frequentemente nos aminoácidos dos peptídeos. Essa maior diferença na variabilidade encontrada na comparação entre proteínas virais e peptídeos pode contribuir para que os vírus escapem da apresentação de suas proteínas para o sistema imunológico e, assim, os vírus estariam menos sujeitos à eliminação no organismo infectado. Uma diferença ainda maior foi observada entre as variabilidades dos peptídeos e de proteínas humanas tumorais, gerando uma dificuldade maior ainda para que determinadas mutações dessas proteínas sejam apresentadas por moléculas MHC para serem eliminadas pelo sistema imunológico. Dessa forma, considerando-se os gráficos de aminoácidos que apresentam uma linearidade relativa alta, parece não haver seletividade pela molécula de MHC na apresentação de peptídeos que contêm esses aminoácidos, ou seja, peptídeos que contêm essas substituições são apresentados. Considerando-se as substituições de aminoácidos cujos gráficos apresentam linearidade relativa mais baixa, os peptídeos que contêm essas substituições parecem sofrer uma seleção pela molécula de MHC, fazendo com que sejam menos apresentados por tais moléculas, visto que essas substituições não apresentam a mesma frequência de ocorrência nos peptídeos e na proteína que está sendo considerada.

Após a divisão dos peptídeos considerando as fontes dos mesmos, a semelhança entre proteínas humanas e peptídeos diminuiu e a comparação entre as proteínas de vírus e os peptídeos originados de proteínas de vírus manteve, de uma maneira geral, o mesmo

padrão da comparação inicial. Quando foi feita a separação dos peptídeos, observou-se que havia um número grande de peptídeos que eram provenientes de proteínas virais, sendo que das 3412 seqüências de peptídeos totais usadas no início, foi possível obter 505 seqüências provenientes de proteínas humanas normais, 310 de proteínas humanas tumorais e 1591 que eram originadas de vírus (há mais seqüências que faziam parte das 3412 seqüências e que não foram separadas porque eram provenientes outros organismos - exemplo, *Plasmodium falciparum*). Essa maior quantidade da amostra de peptídeos que têm como fonte proteínas virais pode fazer com que a comparação entre as proteínas de vírus e os peptídeos provenientes de proteínas virais apresente um resultado mais consistente do que as outras duas comparações. Dessa forma, para uma melhor comparação seria necessário que se obtivessem mais seqüências de peptídeos originados de proteínas humanas normais e tumorais. Contudo, na comparação inicial, mesmo tendo considerado todos os peptídeos e sendo a maioria desses peptídeos provenientes de proteínas virais, houve maior semelhança na variabilidade observada entre os peptídeos e as proteínas humanas normais. Esse dado mostra que, apesar de terem sido originados de proteínas de vírus, os peptídeos que são apresentados pela molécula de MHC contêm substituições de aminoácidos que se assemelham às mutações ocorridas em proteínas humanas normais. Essa semelhança reforça o fato já discutido aqui de que as substituições de aminoácidos que ocorrem nos peptídeos seguem a mesma tendência das mutações observadas em proteínas humanas normais.

A correlação encontrada entre o número de substituições ocorridas nos aminoácidos dos peptídeos e as áreas expostas médias desses aminoácidos mostra que os

aminoácidos hidrofóbicos que apresentam maior número de substituições são os que têm menor acessibilidade média. Considerando-se que as interações hidrofóbicas são importantes para a estabilidade da ligação entre moléculas, os aminoácidos hidrofóbicos que apresentam menor área exposta média e uma alta taxa de substituição aparecem bastante freqüentemente em posições-âncora, como a leucina. O fato de ser um resíduo-âncora explica o dado de que esse aminoácido apresenta pequena acessibilidade, mas o dado da alta taxa de substituição não era esperado, pois sendo um resíduo-âncora e, portanto, importante na ligação entre o peptídeo e a molécula de MHC, ele deveria ser um resíduo conservado e não apresentar substituições freqüentes. Mas o fato é que a leucina também pode estar afastada das posições-âncora, o que elevaria sua taxa de substituição. Contudo, um aspecto do trabalho que não foi considerado anteriormente, e que deve ser considerado em estudos posteriores a este trabalho, está relacionado à verificação da taxa de substituição dos aminoácidos dos peptídeos em cada posição específica. Dessa forma, de posse da informação de quais aminoácidos são mais substituídos em cada posição específica do peptídeo e quais são as acessibilidades desses aminoácidos, poderia-se analisar quais substituições ocorrem mais freqüentemente em cada posição e quais são as acessibilidades mais aceitas desses aminoácidos no bolso de ligação.

Um ponto que pode ser analisado com o resultado da comparação entre área exposta média e taxa de substituição dos aminoácidos dos peptídeos é com relação a resíduos que não são considerados âncoras, por exemplo, a metionina. Esse resíduo apresenta grande acessibilidade e pequena taxa de substituição, podendo ser importante

na ligação com o linfócito T e, assim, peptídeos que possuem esse aminoácido em regiões de ligação à célula T podem ter preferência na apresentação pela molécula de MHC.

Os resultados mostrando que os aminoácidos hidrofílicos não apresentam correlação entre acessibilidade média e taxa de substituição dos aminoácidos dos peptídeos (ver figura 26, no anexo 2) significa que essa relação não deve ser importante e resíduos de aminoácidos que têm alta taxa de substituição podem ter alta ou baixa área exposta.



6 CONCLUSÕES

- Foi possível gerar matrizes de mutação para as proteínas humanas normais e tumorais e proteínas virais e de substituição para os peptídeos apresentados por moléculas de MHC de classe I.
- Analisando os peptídeos como um todo, verificou-se que o número de correlações significativas é maior quando a comparação é realizada com proteínas humanas normais.
- Analisando separadamente os peptídeos de acordo com suas origens, o padrão de correlações significativas foi diferente.



7 REFERÊNCIAS BIBLIOGRÁFICAS

Batalia MA, Collins EJ. Peptide binding by class I and class II MHC molecules. **Biopolymers** 1997; 43:281-302.

Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. **Nucleic Acids Res** 2000; 28:235-42.

Brusic V, Rudy G, Harrison LC. MHCPEP, a database of MHC-binding peptides: update 1997. **Nucleic Acids Res** 1998; 26:368-71.

Dayhoff MO, Eck RV, Park CM. **Atlas of protein sequence and structure**. Maryland: Silver Spring The National Biomedical Research Foundation; 1972. A model of evolutionary change in proteins; p.89-99.

Desiraju GR. Chemistry beyond the molecule. **Nature** 2001; 412:397-400.

Lehn JM. Supramolecular chemistry. **Science** 1993; 260:1762-3.

Maffei A, Harris PE. Peptides bound to major histocompatibility complex molecules. **Peptides** 1998; 19:179-98.

Michielin O, Luescher I, Karplus M. Modeling of the TCR-MHC-peptide complex. **J Mol Biol** 2000; 300:1205-35.

Rudolph MG, Wilson IA. The specificity of TCR/pMHC interaction. **Curr Opin Immunol** 2002; 14:52-65.

Schultze JL, Vonderheide RH. From cancer genomics to cancer immunotherapy: toward second-generation tumor antigens. **Trends Immunol** 2001; 22:516-23.

Stoltze L, Nussbaum AK, Sijts A, Emmerich PN, Kloetzel PM, Schild H. The function of the proteasome system in MHC class I antigen processing. **Immunol Today** 2000; 21:317-9.

Swiss-prot. Disponível em <URL:<http://www.ebi.ac.uk/swissprot>> [2002 Jul 20].

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res** 1994; 22:4673-80.

Vriend G. WHAT IF: a molecular modeling and drug design program. **J Mol Graph** 1990; 8:52-6.

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

Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. **J Mol Biol** 1998; 281: 929-47.

ANEXO 1

Quadro 1. Descrição dos programas de computador utilizados para os cálculos referentes às proteínas humanas.

Arquivos de entrada de dados	Nome do programa	Descrição	Arquivos de saída de dados
allcyto_c.fasta allhemo.fasta allinsul.fasta allyzo.fasta allmyogl.fasta	cadeias.f (*)	Separa as seqüências de todas as cadeias presentes em cada proteína específica.	allcyto_cA.dat até allcyto_cZ.dat allhemoA.dat até allhemoZ.dat allinsulA.dat até allinsulZ.dat allyzoA.dat até allyzoZ.dat allmyoglA.dat até allmyoglZ.dat
allcyto_cA.dat allcyto_cB.dat allcyto_cC.dat allhemoA.dat allinsulA.dat allinsulB.dat allinsulD.dat allyzoA.dat allmyoglA.dat	Clustalw	Alinha seqüências	cyto_cA.aln cyto_cB.aln cyto_cC.aln hemo.aln insulA.aln insulB.aln insulD.aln lyzo.aln myoglA.aln
cyto_cA.aln cyto_cB.aln cyto_cC.aln hemo.aln insulA.aln insulB.aln insulD.aln lyzo.aln myoglA.aln	matriz_h.f	Calcula a matriz de mutação	mtcyto_a.dat mtcyto_b.dat mtcyto_c.dat mthemo.dat mtinsuA.dat mtinsuB.dat mtinsuC.dat mtlyzo.dat mtmyogl_a.dat
mtcyto_a.dat mtcyto_b.dat mtcyto_c.dat mthemo.dat mtinsuA.dat mtinsuB.dat mtinsuC.dat mtlyzo.dat mtmyogl_a.dat	total_h.f	Soma as matrizes calculadas para cada proteína específica, calculando o número total de mutações para as proteínas humanas.	total_h.dat

(*) após a separação das cadeias existentes em cada proteína por meio do programa 'cadeias.f', foram utilizadas somente algumas dessas cadeias, como mostram os arquivos de entrada de dados para o programa Clustalw. Essa separação foi necessária para viabilizar o alinhamento e a contagem do número de trocas de aminoácidos feitos posteriormente. Foram gerados arquivos de saída com cadeias de A até Z porque não era conhecido quais cadeias existiam em cada proteína no banco de dados original. Daí, utilizaram-se os arquivos que continham maior número de seqüências (ver arquivos de entrada de dados para o programa Clustalw).

Quadro 2. Programas de computadores usados para os cálculos referentes às proteínas virais.

Arquivos de entrada de dados	Nome do programa	Descrição	Arquivos de saída de dados
piv.swiss (adenovírus)	Clustalw	Alinha seqüências	ade.aln
piv.fasta (adenovírus)			
astrovir.swiss			astrovir.aln
envhiv1.fasta			envhiv.aln
hepagenpoli.swiss			hepa.aln
papVL1.fasta			papVL1 fa.aln
papVL1.swiss			papVL1.aln
papVL2.swiss			papVL2.aln
polioVP2.fasta			polioVP2.aln
rinogenpoli.swiss			rino.aln
rotaVP4.swiss			rotaVP4.aln
rotaVP7.swiss			rotaVP7.aln
vaccienv.fasta			vaccienv.aln
ade.aln	matriz_v.f	Calcula a matriz de mutação	mtade.dat
astrovir.aln			mtast.dat
envhiv.aln			methiv.dat
hepa.aln			methpa.dat
papVL1 fa.aln			mtpapL1 f.dat
papVL1.aln			mtpapL1.dat
papVL2.aln			mtpapL2.dat
polioVP2.aln			mtpoliP2.dat
rino.aln			mtrino.dat
rotaVP4.aln			mtrotaP4.dat
rotaVP7.aln			mtrotaP7.dat
vaccienv.aln			mtvac.dat
mtade.dat	total_v.f	Soma as matrizes calculadas para cada proteína específica, calculando o número total de mutações para as proteínas humanas.	total_v.dat
mtast.dat			
methiv.dat			
methpa.dat			
mtpapL1 f.dat			
mtpapL1.dat			
mtpapL2.dat			
mtpoliP2.dat			
mtrino.dat			
mtrotaP4.dat			
mtrotaP7.dat			
mtvac.dat			

Quadro 3. Programas de computador usados para os cálculos referentes aos peptídeos ligados a moléculas de MHC.

Arquivos de entrada de dados	Nome do programa	Descrição	Arquivos de saída de dados		
mhcpep.txt	mhcpep3.f	Separa e agrupa as seqüências dos peptídeos que estão ligados ao mesmo alelo de MHC.	pepa1.dat	pepb13.dat	pepb53.dat
			pepa11.dat	pepb14.dat	pepb54.dat
			pepa2.dat	pepb15.dat	pepb55.dat
			pepa23.dat	pepb17.dat	pepb56.dat
			pepa24.dat	pepb18.dat	pepb57.dat
			pepa25.dat	pepb27.dat	pepb58.dat
			pepa26.dat	pepb35.dat	pepb60.dat
			pepa28.dat	pepb37.dat	pepb61.dat
			pepa29.dat	pepb38.dat	pepb62.dat
			pepa3.dat	pepb39.dat	pepb7.dat
			pepa30.dat	pepb42.dat	pepb73.dat
			pepa31.dat	pepb44.dat	pepb8.dat
			pepa32.dat	pepb45.dat	
			pepa33.dat	pepb46.dat	
			pepa62.dat	pepb5.dat	
			pepa68.dat	pepb51.dat	
			pepa69.dat	pepb52.dat	
pepa1.dat	Clustalw	Alinha seqüências	pepa1.aln		
pepa11.dat			pepa11.aln		
pepa2.dat			pepa2.aln		
pepa23.dat			pepa23.aln		
pepa24.dat			pepa24.aln		
⋮	matriz_p.f	Calcula a matriz de mutação	⋮		
pepb8.dat			pepb8.aln		
pepa1.aln			mtpepa1.dat		
pepa11.aln			mtpepa11.dat		
pepa2.aln			mtpepa2.dat		
pepa23.aln			mtpepa23.dat		
pepa24.aln			mtpepa24.dat		
⋮			⋮		
pepb8.aln			mtpepb8.dat		
mtpepa1.dat	total_p.f	Soma as matrizes dos grupos de peptídeos ligados a cada alelo de MHC, calculando o número total de substituições para os peptídeos.	total_p.dat		
mtpepa11.dat					
mtpepa2.dat					
mtpepa23.dat					
mtpepa24.dat					
⋮					
mtpepb8.dat					

Quadro 4. Programas de computador criados para a construção dos gráficos e cálculo dos coeficientes lineares.

Arquivos de entrada de dados	Nome do programa	Descrição	Arquivos de saída de dados
total_h.dat total_p.dat	pontgra2.f	Coloca os valores das matrizes dos arquivos de entrada de dados a num formato aceito pelo programa que gera os gráficos.	grafhum.dat
total_v.dat total_p.dat			grafvir.dat
total_at.dat total_p.dat			grftaa.dat
total_h.dat total_p.dat	coef_hp.f	Calcula os coeficientes dos gráficos referentes aos valores das matrizes dos arquivos de entrada de dados.	coef_hp.dat
total_v.dat total_p.dat			coef_vp.dat
total_at.dat total_p.dat			coef_atp.dat



Quadro 5. Programas de computador usados para auxiliar a construção dos gráficos referentes às acessibilidades dos aminoácidos dos peptídeos ligados a moléculas de MHC e ao número de substituições ocorridas entre esses aminoácidos.

Arquivos de entrada de dados	Nome do programa	Descrição	Arquivos de saída de dados
grafhum.dat (tabela 5)	ordsubpp.f	Soma todas as substituições ocorridas para cada aminoácido específico.	ordsubpp.dat
ac_todas.txt	areas.f	Calcula a média das áreas expostas de cada aminoácido presente nos peptídeos das estruturas do complexo peptídeo-MHC retiradas do PDB.	areas.dat

OBS: Com os arquivos 'ordsubpp.dat' e 'areas.dat', foram gerados os arquivos 'acsupphi.txt' (referente aos aminoácidos hidrofílicos presentes nos peptídeos) e 'acsupphi.txt' (referente aos aminoácidos hidrofóbicos presentes nos peptídeos). Esses dois arquivos foram utilizados no programa Maple-V para construção dos respectivos gráficos.

Tabela 6. Valores das acessibilidades médias calculadas para cada aminoácido.

Aminoácido	Acessibilidade média (Å ²)
A	5.3576
C	6.1778
D	9.5840
E	14.0929
F	10.6721
G	0.0000
H	16.5191
I	6.7885
K	13.2616
L	5.1025
M	16.5533
N	14.9025
P	8.5162
Q	7.7519
R	15.3811
S	8.3380
T	8.4134
V	8.7409
W	12.8301
Y	12.0705

**EXEMPLOS DE ALGUNS ARQUIVOS DE ENTRADA E SAÍDA DE
DADOS DOS PROGRAMAS UTILIZADOS NESTE TRABALHO**

OS ARQUIVOS EXEMPLIFICADOS AQUI SÃO REFERENTES ÀS PROTEÍNAS HUMANAS, MAS EXATAMENTE O MESMO PROCEDIMENTO FOI FEITO PARA AS PROTEÍNAS VIRAIS E PARA OS PEPTÍDEOS.

ARQUIVO: allhemo.fasta (arquivo de entrada de dados para o programa 'cadeias.f')

```
>1HV4:A HEMOGLOBIN ALPHA-A CHAIN
VLAAADKTNVKGVFSGKISGHAEYGAETLERMFTAYPQTKTYFPHFDLQHGSAQIKAHGK
KVVAALVEAVNHIDDIAGALSLSLHAQKLKRVDPVNFKFLGHCFLVVVAIHHPSALTAE
VHASLDKFLCAVGTVLTAKYR
>1HV4:B HEMOGLOBIN BETA CHAIN
VHWSAEKQLITGLWGKVNVDGAEALARLLIVYPWTQRFFSSFGNLSSPTAILGNPMV
RAHGKKVLTSFGDAVKNLNDNIKNTFAQLSELHCDKLHVDPENFRLLGDILIIVLAHFAK
EFTPDCAAWQKLVRVVAHALARKYH
>1HV4:C HEMOGLOBIN ALPHA-A CHAIN
VLAAADKTNVKGVFSGKISGHAEYGAETLERMFTAYPQTKTYFPHFDLQHGSAQIKAHGK
KVVAALVEAVNHIDDIAGALSLSLHAQKLKRVDPVNFKFLGHCFLVVVAIHHPSALTAE
VHASLDKFLCAVGTVLTAKYR
>1BIJ:A HEMOGLOBIN A
VLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
>1BIJ:B HEMOGLOBIN A
VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
KAHGKKVLGAFSDGLAHLNKLGTFTLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK
EFTPPVQAAYQKVVAGVANALAHKYH
>1BIJ:C HEMOGLOBIN A
VLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
>1HGC:A HEMOGLOBIN (T STATE, ALPHA-OXY) - CHAIN A
VLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
>1HGC:B HEMOGLOBIN (T STATE, ALPHA-OXY) - CHAIN B
VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
KAHGKKVLGAFSDGLAHLNKLGTFTLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK
EFTPPVQAAYQKVVAGVANALAHKYH
>1HGC:C HEMOGLOBIN (T STATE, ALPHA-OXY) - CHAIN C
VLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
```


ARQUIVOS: allhemoA.dat até allhemoZ.dat (arquivos de saída de dados do programa 'cadeias.f' e entrada para o programa Clustalw - estão exemplificados somente algumas seqüências dos arquivos 'allhemoA.dat' e 'allhemoD.dat')

ARQUIVO: allhemoA.dat

```
>1HV4:A HEMOGLOBIN ALPHA-A CHAIN
VLSAADKTNVKGVFSGHAEYGAETLERMFTAYPQTKTYFPHFDLQHGSAQIKAHGK
KVVAALVEAVNHIDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFLVVVAIHHPALTAE
VHASLDKFLCAVGTVLTKYR
>1BIJ:A HEMOGLOBIN A
VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNAVAHVDDMPNALSALSDLHAHKLVRDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
>1HGC:A HEMOGLOBIN (T STATE, ALPHA-OXY) - CHAIN A
VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGK
KVADALTNAVAHVDDMPNALSALSDLHAHKLVRDPVNFKLLSHCLLVTLAAHLPAEFTPA
VHASLDKFLASVSTVLTSKYR
```

ARQUIVO: allhemoD.dat

```
>1HV4:D HEMOGLOBIN BETA CHAIN
VHWSAEEKQLITGLWGKVNVDGAEALARLLIVYPWTQRFFSSFGNLSSPTAILGNPMV
RAHGKKVLTSFGDAVKNLDNINKNTFAQLSELHCDKLHVDPENFRLLGDILIIIVLAAHFAK
EFTPDCAAWQKLVRVVAHALARKYH
>1BIJ:D HEMOGLOBIN A
VHLTPEEKSAVTALWGKVNVDGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKV
KAHGKKVLGAFSDGLAHLNLLKGTFAITSELHCDKLHVDPENFRLLGNVLCVLAHHFGK
EFTPPVQAAYQKVVAGVANALAHKYH
>1HGC:D HEMOGLOBIN (T STATE, ALPHA-OXY) - CHAIN D
VHLTPEEKSAVTALWGKVNVDGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKV
KAHGKKVLGAFSDGLAHLNLLKGTFAITSELHCDKLHVDPENFRLLGNVLCVLAHHFGK
EFTPPVQAAYQKVVAGVANALAHKYH
```

ARQUIVO: hemo.aln (saída do programa Clustalw e entrada para o programa 'matriz_h.f')

```

1HV4_A      -----VLS--AADKTNVKGVSFSG-----HAEYEG
1HGC_A      -----VLS--PADKTNVKAAGWK-VGA-----HAGEYG
1BIJ_A      -----VLS--PADKTNVKAAGWK-VGA-----HAGEYG

1HV4_A      AETLERMFTAYPQTKTYFP-----HF--DLQH-----
1HGC_A      AEALERMFLSFPTTKTYFP-----HF--DLSH-----
1BIJ_A      AEALERMFLSFPTTKTYFP-----HF--DLSH-----

1HV4_A      -----GSAQIK-----AH--GKKVVAALVEAVNHIDD--IAGALSKLS----DLHAQ
1HGC_A      -----GSAQVK-----GH--GKKVADALTNAVAHVDD--MPNALSALS----DLHAH
1BIJ_A      -----GSAQVK-----GH--GKKVADALTNAVAHVDD--MPNALSALS----DLHAH

1HV4_A      KLRVD-----PVNFKFLGHCF---LVVVAIHHPG-----A-LTAE--VHASLDKFL
1HGC_A      KLRVD-----PVNFKLLSHCL---LVTLAHLPA-----E-FTPA--VHASLDKFL
1BIJ_A      KLRVD-----PVNFKLLSHCL---LVTLAHLPA-----E-FTPA--VHASLDKFL

1HV4_A      CAVGTVLTAKYR-----
1HGC_A      ASVSTVLTAKYR-----
1BIJ_A      ASVSTVLTAKYR-----

```

ANEXO 2

A	C	D	E	F	G	H	I	K	L
30187.00									
5285.00	661.00								
12733.00	1788.00	6072.00							
17088.00	2323.00	7223.00	18249.00						
28793.00	4152.00	6920.00	9063.00	24399.00					
31624.00	2834.00	10552.00	18218.00	12788.00	32476.00				
8919.00	1068.00	2692.00	5234.00	6085.00	5793.00	1956.00			
25058.00	3285.00	6224.00	8505.00	17850.00	13238.00	5020.00	24300.00		
32080.00	4153.00	9785.00	16246.00	14638.00	32068.00	6431.00	15258.00	51779.00	
43502.00	6469.00	11054.00	15903.00	34142.00	21532.00	9572.00	53481.00	28864.00	77288.00
7041.00	973.00	1673.00	2536.00	6289.00	3144.00	1512.00	7636.00	3710.00	14125.00
11508.00	1500.00	3991.00	7552.00	6467.00	6581.00	2615.00	5303.00	8223.00	12208.00
21077.00	2607.00	7532.00	7875.00	11862.00	15192.00	3950.00	10127.00	13135.00	21464.00
14774.00	1708.00	6475.00	8300.00	9305.00	13584.00	3864.00	9785.00	10324.00	14222.00
26459.00	2851.00	7680.00	12403.00	13387.00	29324.00	10512.00	14069.00	45222.00	31584.00
26708.00	3506.00	6641.00	11283.00	13011.00	21360.00	6218.00	19498.00	15993.00	24927.00
26473.00	2861.00	8737.00	14357.00	15451.00	16813.00	6441.00	20152.00	19104.00	28254.00
33244.00	4770.00	6636.00	10166.00	18588.00	14424.00	5832.00	28506.00	14480.00	49239.00
4106.00	528.00	1310.00	1939.00	5887.00	2912.00	1146.00	3955.00	2998.00	8037.00
12191.00	1544.00	3609.00	4433.00	17850.00	6898.00	3282.00	11121.00	14228.00	25086.00

M	N	P	Q	R	S	T	V	W	Y
1891.00									
1837.00	2215.00								
2768.00	5695.00	60055.00							
2133.00	4399.00	7518.00	12287.00						
3551.00	6805.00	9847.00	15647.00	112465.00					
3559.00	6151.00	16453.00	9056.00	19311.00	18362.00				
5142.00	8822.00	12302.00	8949.00	20406.00	18197.00	29619.00			
7884.00	6585.00	14229.00	10499.00	15330.00	18170.00	22327.00	35759.00		
1100.00	1298.00	2850.00	4244.00	2514.00	2413.00	2578.00	3205.00	1267.00	
3553.00	3143.00	5455.00	4303.00	10356.00	6605.00	10543.00	10997.00	8128.00	36288.00

abela 1. Matriz referente às trocas observadas entre os aminoácidos dos peptídeos deste estudo (sem isão dos peptídeos segundo suas origens). Cada termo da matriz representa as substituições ocorridas entre dois aminoácidos especificados em cada um dos dois eixos da matriz. A matriz foi dividida em duas partes a maior clareza de leitura.



A	C	D	E	F	G	H	I	K	L
1137056.00									
17655.00	55616.00								
105145.00	14431.00	164187.00							
69885.00	6056.00	60712.00	136718.00						
37658.00	8204.00	27071.00	21902.00	122060.00					
133433.00	18329.00	61409.00	57773.00	24414.00	262576.00				
29393.00	3845.00	18223.00	13686.00	10842.00	24953.00	87784.00			
67936.00	9657.00	27532.00	17227.00	35220.00	67037.00	13694.00	118149.00		
88906.00	9411.00	57816.00	44905.00	18773.00	67755.00	64134.00	34925.00	216513.00	
112151.00	22973.00	46683.00	262156.00	50408.00	55351.00	23062.00	96799.00	80648.00	350165.00
22734.00	13386.00	12943.00	13398.00	10006.00	14328.00	5293.00	17411.00	16073.00	43182.00
77499.00	11687.00	60117.00	28326.00	19220.00	52202.00	128396.00	33849.00	61265.00	48134.00
42132.00	8940.00	31135.00	25677.00	9838.00	32534.00	9500.00	13737.00	36470.00	25148.00
62480.00	6058.00	33621.00	24840.00	16470.00	35477.00	13748.00	23099.00	43921.00	68092.00
59849.00	33153.00	33313.00	39947.00	21198.00	76809.00	13749.00	26213.00	67912.00	61349.00
92986.00	17232.00	48313.00	34122.00	20604.00	63355.00	15568.00	29568.00	61193.00	81201.00
80780.00	16545.00	46251.00	47471.00	34839.00	76656.00	16267.00	43996.00	58561.00	55835.00
98067.00	12222.00	59821.00	30348.00	33543.00	53255.00	16024.00	88292.00	52529.00	94286.00
27387.00	2858.00	18870.00	4120.00	11679.00	16850.00	6628.00	7874.00	10016.00	20231.00
48443.00	8098.00	30532.00	14307.00	26430.00	23921.00	9981.00	28963.00	42135.00	32449.00
M	N	P	Q	R	S	T	V	W	Y
47532.00									
19454.00	151808.00								
11111.00	30499.00	94407.00							
6005.00	32766.00	15503.00	87315.00						
18548.00	52493.00	19371.00	45113.00	167486.00					
13985.00	48151.00	39337.00	24704.00	51272.00	152257.00				
11427.00	71634.00	34279.00	50790.00	48327.00	45753.00	164475.00			
22087.00	38574.00	23296.00	26271.00	50590.00	41559.00	47594.00	192778.00		
3022.00	12962.00	4656.00	13273.00	24727.00	13499.00	11846.00	13896.00	55082.00	
13330.00	30332.00	10517.00	25625.00	21449.00	25679.00	39630.00	39740.00	16765.00	103610.00

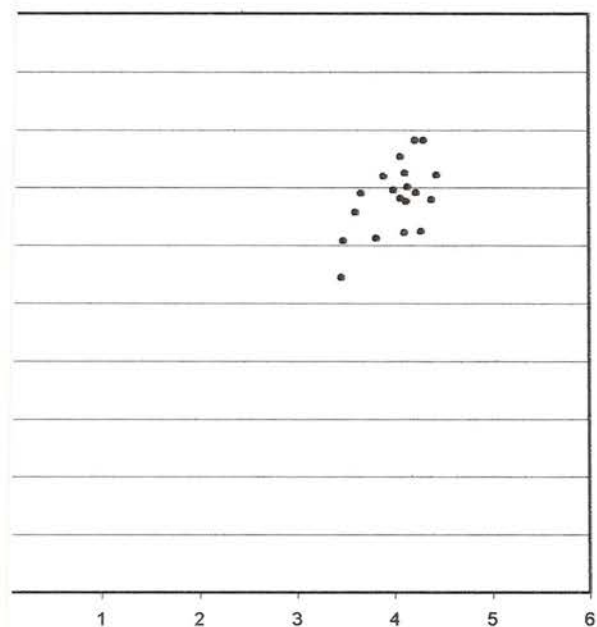
bela 2. Esta matriz representa as mutações entre os aminoácidos colocados nos eixos da matriz ocorridas proteínas humanas normais. A matriz foi dividida em duas partes para uma melhor visualização.

A	C	D	E	F	G	H	I	K	L
222220.00									
52924.00	171856.00								
42246.00	11072.00	181613.00							
62449.00	3680.00	79494.00	289349.00						
28986.00	29083.00	8745.00	34605.00	186342.00					
51585.00	16832.00	29276.00	80385.00	29194.00	294635.00				
38131.00	28788.00	9634.00	21903.00	10224.00	17892.00	88161.00			
36843.00	29002.00	46735.00	71360.00	30051.00	57265.00	28514.00	327927.00		
68572.00	39622.00	53030.00	66443.00	27911.00	39864.00	8580.00	62899.00	194293.00	
45488.00	37880.00	27403.00	35133.00	74379.00	44103.00	13693.00	104414.00	81070.00	581364.00
10124.00	6455.00	7833.00	30090.00	8848.00	15301.00	17449.00	38783.00	17455.00	52377.00
51070.00	102912.00	88926.00	67051.00	73001.00	68621.00	32954.00	66291.00	105814.00	75606.00
50980.00	44698.00	30534.00	37908.00	12357.00	48131.00	10785.00	51301.00	40841.00	38811.00
38365.00	22567.00	27243.00	53212.00	12572.00	19345.00	27802.00	35634.00	38723.00	57092.00
38452.00	14228.00	23290.00	26476.00	8613.00	45162.00	8881.00	62710.00	91844.00	35993.00
91133.00	38360.00	53955.00	49841.00	47728.00	73341.00	21266.00	60686.00	51582.00	61251.00
133370.00	61052.00	58390.00	76059.00	27372.00	77349.00	39080.00	109292.00	88086.00	94817.00
65921.00	45228.00	32703.00	94420.00	23651.00	31178.00	20050.00	169562.00	67863.00	78156.00
16601.00	2454.00	17973.00	13286.00	5528.00	19636.00	3572.00	8438.00	73322.00	28808.00
28634.00	26771.00	10395.00	12481.00	66265.00	23616.00	24262.00	15678.00	20668.00	31773.00
M	N	P	Q	R	S	T	V	W	Y
126421.00									
39286.00	349608.00								
3872.00	37288.00	310828.00							
16691.00	42854.00	29615.00	142803.00						
13609.00	39319.00	33690.00	35720.00	233493.00					
7597.00	154036.00	47725.00	46209.00	48275.00	267026.00				
20045.00	181234.00	67526.00	61292.00	62091.00	171655.00	357705.00			
28825.00	67103.00	85168.00	53583.00	35755.00	85048.00	123149.00	352834.00		
16918.00	37258.00	7731.00	14376.00	11718.00	16191.00	23366.00	40680.00	102228.00	
2196.00	28944.00	5917.00	9299.00	21227.00	24834.00	19372.00	40322.00	5506.00	149581.00

abela 3. Matriz referente às mutações observadas em proteínas virais. Cada termo representa a mutação observada entre os dois aminoácidos específicos. Essa matriz também foi dividida em duas partes para uma melhor visualização dos termos.

A	C	D	E	F	G	H	I	K	L
334621.00									
43985.00	90893.00								
17245.00	27349.00	30635.00							
5105.00	1870.00	2902.00	39706.00						
51922.00	2444.00	7752.00	10201.00	28833.00					
22878.00	1845.00	8590.00	9346.00	15546.00	49212.00				
56326.00	13589.00	2513.00	1168.00	13505.00	3486.00	104261.00			
1777.00	456.00	751.00	1178.00	405.00	1122.00	485.00	8229.00		
7761.00	514.00	3578.00	2676.00	767.00	1675.00	3156.00	648.00	17726.00	
34849.00	5002.00	1784.00	3360.00	2088.00	3036.00	4023.00	2110.00	23751.00	24480.00
59265.00	19523.00	400.00	809.00	44686.00	849.00	13546.00	454.00	491.00	1492.00
60659.00	4087.00	2473.00	3472.00	3446.00	18226.00	1645.00	514.00	1191.00	1241.00
17247.00	8247.00	47938.00	2345.00	5322.00	2607.00	4683.00	1014.00	5262.00	13320.00
2592.00	1594.00	1710.00	2768.00	2142.00	2057.00	975.00	1238.00	1886.00	2539.00
9143.00	1616.00	5767.00	2953.00	11369.00	2667.00	14966.00	1029.00	2041.00	7733.00
31029.00	2914.00	3241.00	3521.00	1465.00	3920.00	18427.00	1717.00	3678.00	3771.00
11242.00	976.00	3803.00	2248.00	1291.00	2220.00	2375.00	1333.00	1756.00	2913.00
3972.00	716.00	1243.00	1970.00	1051.00	2146.00	617.00	1457.00	3557.00	5597.00
475.00	122.00	207.00	313.00	162.00	489.00	243.00	156.00	378.00	818.00
1657.00	524.00	1774.00	818.00	509.00	758.00	622.00	483.00	594.00	1153.00
M	N	P	Q	R	S	T	V	W	Y
141972.00									
5314.00	446068.00								
22372.00	4442.00	95899.00							
677.00	1956.00	2912.00	12034.00						
891.00	1219.00	5123.00	2671.00	39764.00					
1498.00	2350.00	4580.00	2718.00	3300.00	19972.00				
902.00	2479.00	3054.00	1592.00	2613.00	2990.00	24918.00			
996.00	660.00	1828.00	1339.00	1757.00	2070.00	1632.00	39940.00		
67.00	205.00	295.00	154.00	892.00	485.00	489.00	319.00	2689.00	
246.00	422.00	1227.00	475.00	1702.00	1245.00	1910.00	870.00	270.00	8764.00

abela 4. Matriz referente às mutações observadas em proteínas humanas tumorais. Cada termo representa a mutação observada entre os dois aminoácidos específicos. Essa matriz também foi dividida em duas partes para uma melhor visualização dos termos.



a 1. Gráfico mostrando o número de substituições do ácido triptofano nas proteínas humanas normais (eixo das abscissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,59

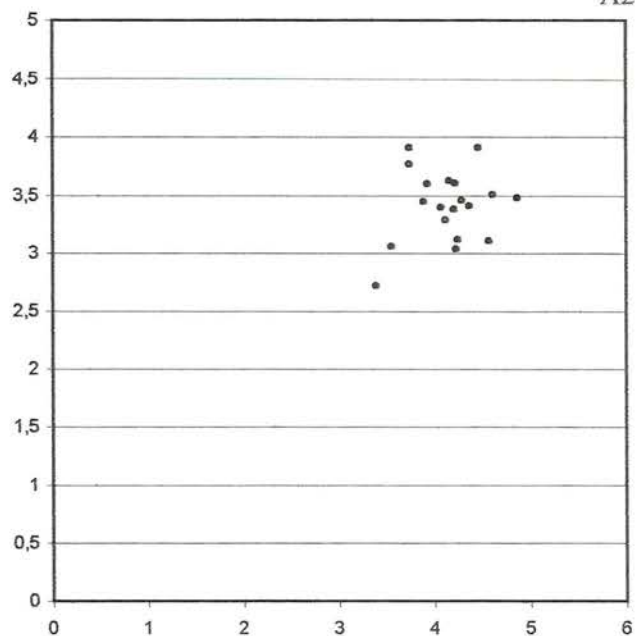


Figura 2. Gráfico mostrando o número de substituições do aminoácido triptofano das proteínas virais (eixo das abscissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,20

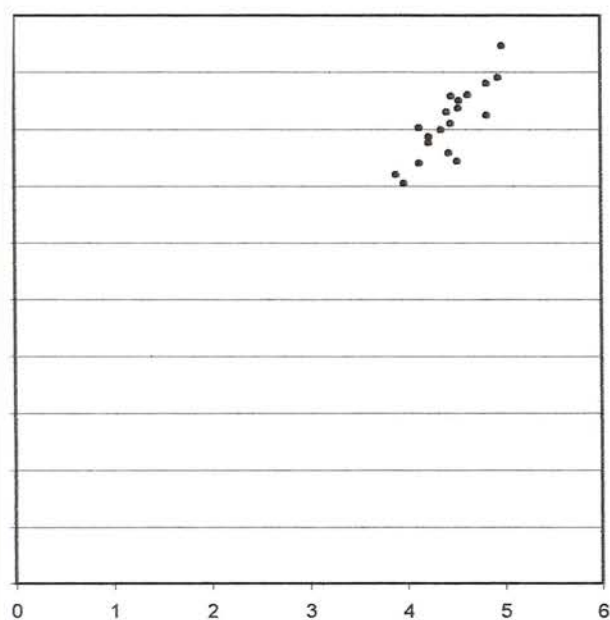


Figura 3. Gráfico mostrando o número de substituições do aminoácido isoleucina das proteínas humanas normais (eixo das abscissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,85

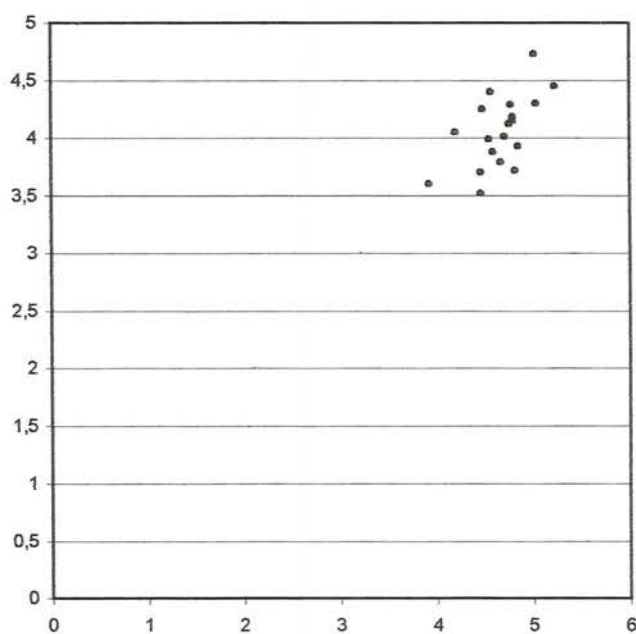


Figura 4. Gráfico mostrando o número de substituições do aminoácido isoleucina das proteínas virais (eixo das abscissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,60

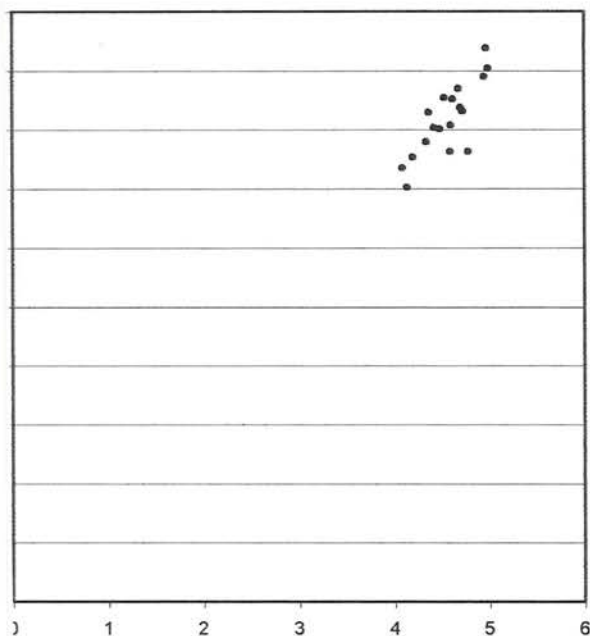


Figura 5. Gráfico mostrando o número de substituições do aminoácido valina das proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,81

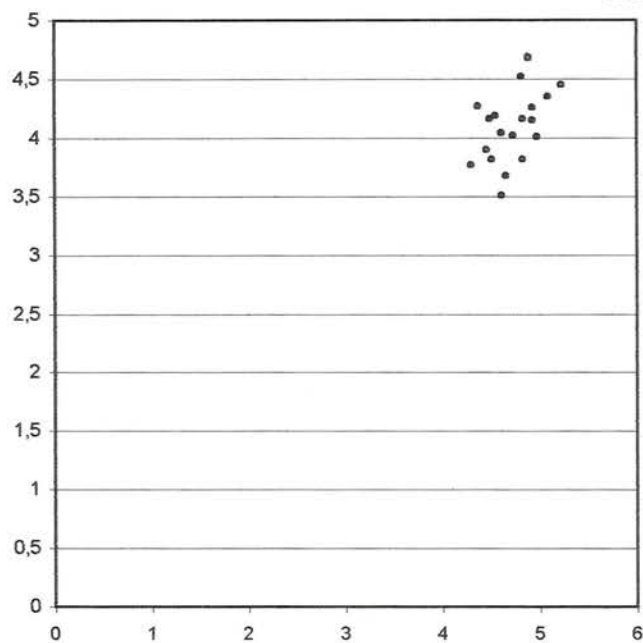


Figura 6. Gráfico mostrando o número de substituições do aminoácido valina das proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,50

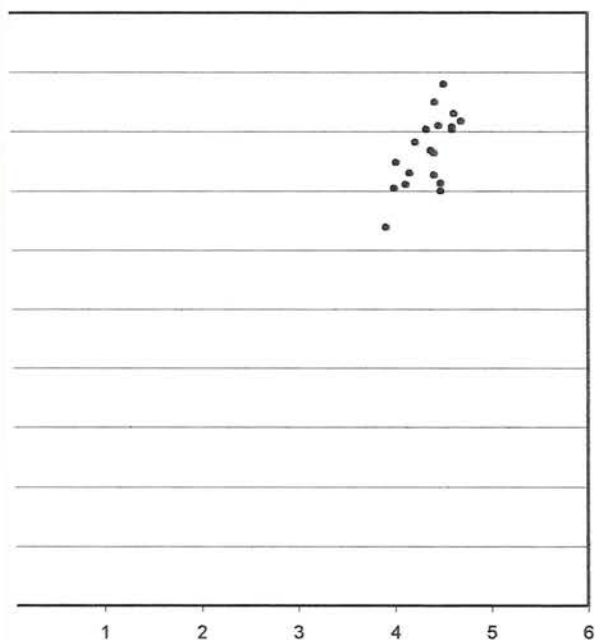


Figura 7. Gráfico mostrando o número de substituições do aminoácido tirosina das proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,67

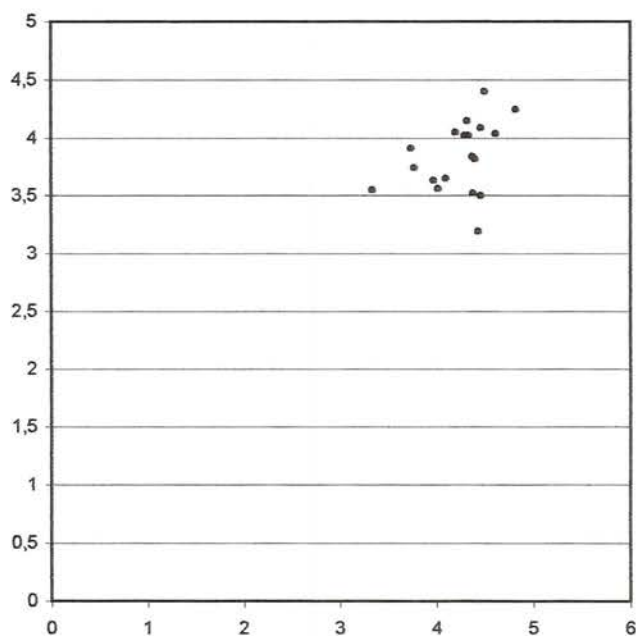


Figura 8. Gráfico mostrando o número de substituições do aminoácido tirosina das proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,38

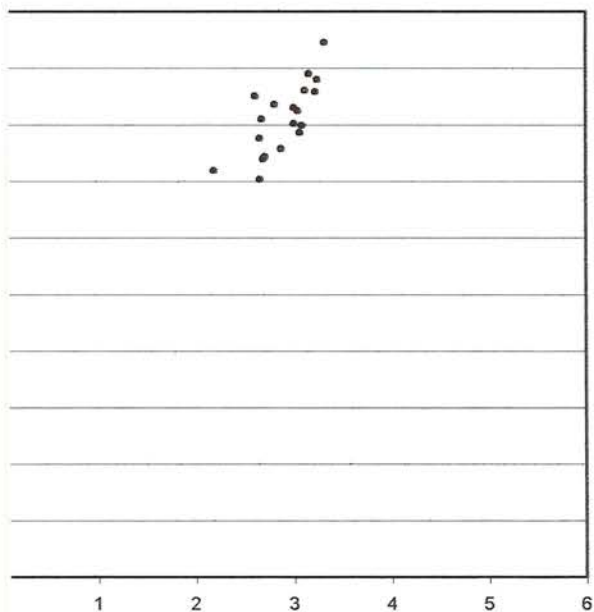


Figura 9. Gráfico mostrando o número de substituições do ácido isoleucina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo ácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,74

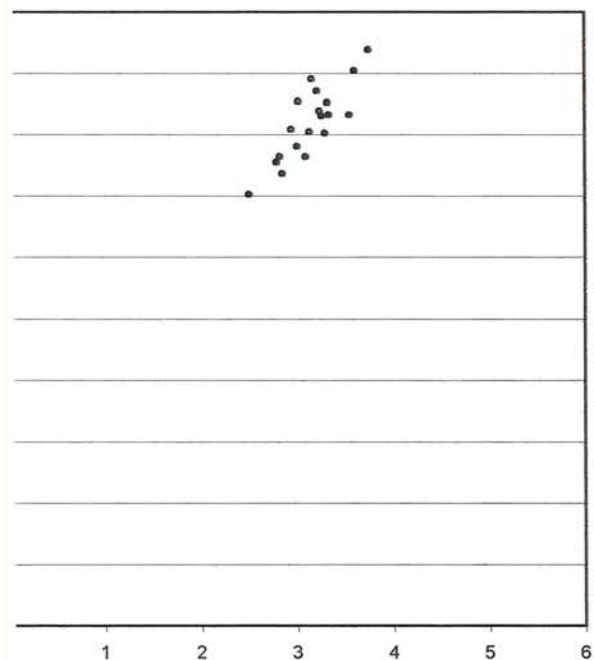


Figura 11. Gráfico mostrando o número de substituições do ácido valina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo ácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,84

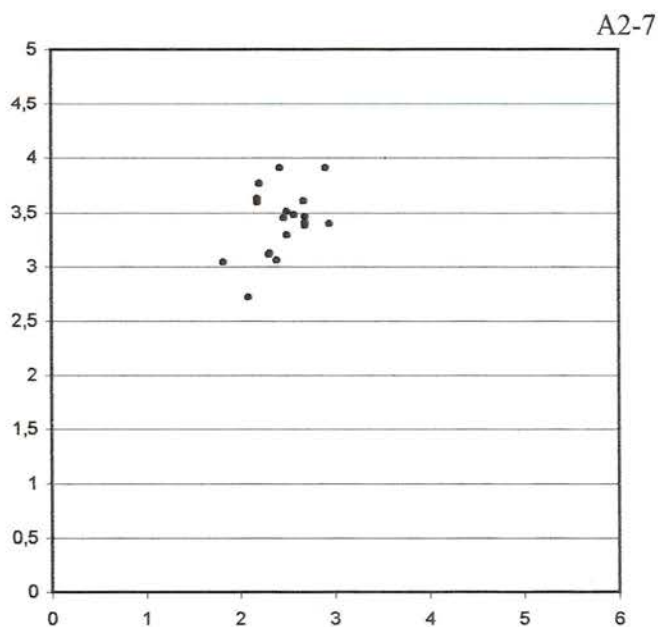


Figura 10. Gráfico mostrando o número de substituições do aminoácido triptofano nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,43

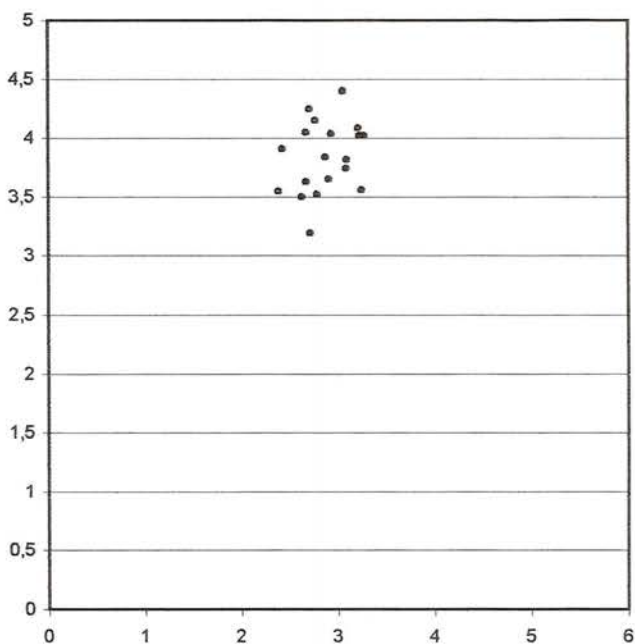


Figura 12. Gráfico mostrando o número de substituições do aminoácido tirosina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos ligados a MHC (eixo das ordenadas). Coeficiente de correlação linear = 0,28

A	C	D	E	F	G	H	I	K	L
5540.00									
66.00	89.00								
533.00	17.00	229.00							
817.00	252.00	251.00	6953.00						
500.00	148.00	157.00	399.00	808.00					
1824.00	39.00	576.00	410.00	402.00	1594.00				
733.00	156.00	112.00	374.00	404.00	420.00	558.00			
410.00	48.00	99.00	244.00	419.00	300.00	310.00	332.00		
1624.00	253.00	448.00	952.00	334.00	1040.00	368.00	336.00	7842.00	
1027.00	339.00	407.00	849.00	1166.00	1057.00	568.00	904.00	668.00	9394.00
270.00	11.00	129.00	214.00	72.00	315.00	179.00	132.00	181.00	833.00
432.00	24.00	150.00	267.00	241.00	448.00	168.00	180.00	281.00	558.00
876.00	100.00	352.00	591.00	240.00	989.00	418.00	247.00	754.00	684.00
1075.00	138.00	321.00	898.00	308.00	447.00	182.00	315.00	596.00	712.00
1517.00	238.00	431.00	1595.00	344.00	507.00	881.00	537.00	1788.00	1563.00
1802.00	172.00	764.00	871.00	983.00	1865.00	613.00	579.00	767.00	1392.00
1180.00	149.00	261.00	569.00	488.00	705.00	702.00	688.00	645.00	867.00
860.00	131.00	218.00	451.00	292.00	500.00	576.00	596.00	751.00	1384.00
31.00	9.00	25.00	155.00	176.00	59.00	70.00	50.00	145.00	236.00
362.00	17.00	105.00	230.00	580.00	314.00	118.00	302.00	423.00	519.00
M	N	P	Q	R	S	T	V	W	Y
239.00									
196.00	320.00								
332.00	416.00	676.00							
172.00	266.00	931.00	6526.00						
387.00	285.00	763.00	1131.00	18933.00					
219.00	588.00	729.00	716.00	917.00	7565.00				
166.00	325.00	614.00	445.00	789.00	709.00	6837.00			
250.00	275.00	404.00	382.00	595.00	842.00	560.00	1769.00		
8.00	69.00	31.00	130.00	166.00	329.00	145.00	40.00	50.00	
411.00	107.00	263.00	188.00	312.00	446.00	265.00	298.00	48.00	8946.00

Figura 5. Matriz referente às trocas observadas entre os aminoácidos dos peptídeos provenientes de células humanas normais. Cada termo da matriz representa as substituições ocorridas entre os dois aminoácidos especificados em cada um dos dois eixos da matriz. A matriz foi dividida em duas partes para maior clareza de leitura.



A	C	D	E	F	G	H	I	K	L
12633.00									
4159.00	1100.00								
7928.00	3102.00	5603.00							
9993.00	2837.00	5234.00	5472.00						
14568.00	3037.00	5673.00	7556.00	49939.00					
20063.00	3438.00	9102.00	11438.00	30466.00	52973.00				
4353.00	1338.00	2171.00	3261.00	3817.00	3176.00	1127.00			
25862.00	4053.00	8871.00	11639.00	24949.00	26342.00	3965.00	35999.00		
20580.00	5168.00	12648.00	18551.00	26714.00	35791.00	6754.00	25509.00	49020.00	
20157.00	6419.00	11447.00	13872.00	36535.00	27363.00	7373.00	37863.00	35423.00	69800.00
6077.00	917.00	1683.00	5097.00	8036.00	8854.00	1516.00	11564.00	11552.00	19686.00
6167.00	1890.00	3672.00	5978.00	8850.00	9831.00	2120.00	7635.00	15929.00	11615.00
12998.00	3152.00	6458.00	6535.00	17852.00	17162.00	2536.00	15943.00	16283.00	24765.00
14179.00	4119.00	10774.00	14048.00	18305.00	24330.00	4507.00	14803.00	33464.00	19617.00
19050.00	3633.00	9412.00	9543.00	13299.00	19791.00	4047.00	20480.00	29283.00	25559.00
13908.00	4695.00	10660.00	13258.00	13184.00	18106.00	5406.00	16442.00	27374.00	26560.00
15170.00	3964.00	7812.00	11032.00	15662.00	17034.00	4815.00	16485.00	29882.00	33444.00
16632.00	3555.00	6106.00	10019.00	22738.00	17981.00	4672.00	23346.00	23057.00	31576.00
6631.00	1769.00	2122.00	5974.00	7641.00	5294.00	2590.00	7103.00	11039.00	11988.00
15169.00	2252.00	4255.00	7540.00	17734.00	17602.00	2235.00	26034.00	16788.00	23827.00
M	N	P	Q	R	S	T	V	W	Y
7108.00									
3904.00	3312.00								
4900.00	5931.00	16181.00							
5488.00	13036.00	12874.00	15530.00						
6802.00	6245.00	13775.00	15365.00	14014.00					
8198.00	13079.00	14759.00	25434.00	15575.00	17225.00				
16795.00	5685.00	11370.00	10975.00	19751.00	13521.00	54034.00			
13404.00	7765.00	13722.00	14051.00	18129.00	17497.00	21621.00	46753.00		
4478.00	3628.00	6317.00	8140.00	7198.00	7174.00	10558.00	9199.00	3375.00	
9706.00	4610.00	10728.00	9018.00	13212.00	8967.00	14041.00	16022.00	6071.00	10173.00

ela 6. Essa matriz representa as trocas observadas entre os aminoácidos dos peptídeos originados de eínas virais. Cada termo da matriz representa as substituições ocorridas entre os dois aminoácidos cificados em cada um dos dois eixos da matriz. A matriz foi dividida em duas partes para maior clareza itura.

A	C	D	E	F	G	H	I	K	L
714.00									
9.00	23.00								
113.00	0.00	206.00							
214.00	9.00	87.00	399.00						
357.00	26.00	459.00	166.00	581.00					
412.00	7.00	78.00	107.00	274.00	929.00				
133.00	8.00	39.00	272.00	45.00	210.00	536.00			
665.00	3.00	294.00	435.00	576.00	398.00	544.00	932.00		
81.00	36.00	30.00	62.00	218.00	104.00	45.00	383.00	213.00	
825.00	35.00	400.00	731.00	853.00	636.00	912.00	1042.00	226.00	1900.00
87.00	0.00	163.00	90.00	324.00	62.00	100.00	183.00	3.00	337.00
94.00	6.00	21.00	102.00	12.00	18.00	85.00	45.00	6.00	373.00
346.00	19.00	67.00	180.00	121.00	426.00	75.00	172.00	34.00	305.00
100.00	3.00	58.00	155.00	59.00	28.00	22.00	92.00	27.00	292.00
84.00	4.00	14.00	143.00	186.00	73.00	316.00	415.00	263.00	554.00
417.00	55.00	87.00	664.00	227.00	533.00	130.00	281.00	9.00	1155.00
772.00	44.00	60.00	483.00	262.00	795.00	135.00	477.00	40.00	1135.00
686.00	22.00	204.00	224.00	568.00	695.00	57.00	662.00	61.00	1030.00
80.00	0.00	6.00	65.00	49.00	24.00	6.00	54.00	38.00	224.00
162.00	17.00	107.00	56.00	630.00	124.00	69.00	736.00	651.00	653.00

M	N	P	Q	R	S	T	V	W	Y
87.00									
14.00	83.00								
42.00	512.00	1503.00							
28.00	58.00	70.00	55.00						
42.00	174.00	687.00	40.00	279.00					
44.00	284.00	811.00	185.00	277.00	563.00				
33.00	174.00	555.00	119.00	84.00	851.00	596.00			
90.00	124.00	404.00	159.00	120.00	621.00	741.00	786.00		
44.00	90.00	311.00	49.00	110.00	162.00	100.00	94.00	61.00	
61.00	2.00	83.00	15.00	611.00	69.00	143.00	149.00	86.00	833.00

ela 7. Matriz representando as trocas observadas entre os aminoácidos dos peptídeos originados de eínas humanas tumorais. Cada termo da matriz representa as substituições ocorridas entre os dois ioácidos especificados em cada um dos dois eixos da matriz. A matriz foi dividida em duas partes para or clareza de leitura.

FUNDAÇÃO ANTONIO PRUDENTE
 ANA MARIA R. PINNARI
 Coordenadora de Pós-graduação

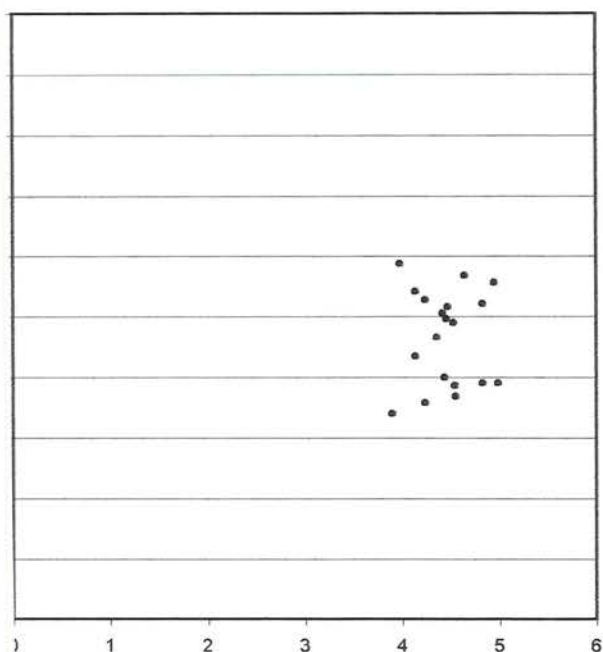


Figura 13. Gráfico mostrando o número de substituições do aminoácido isoleucina nas proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas normais (eixo das ordenadas).

Coefficiente de correlação linear = 0,03

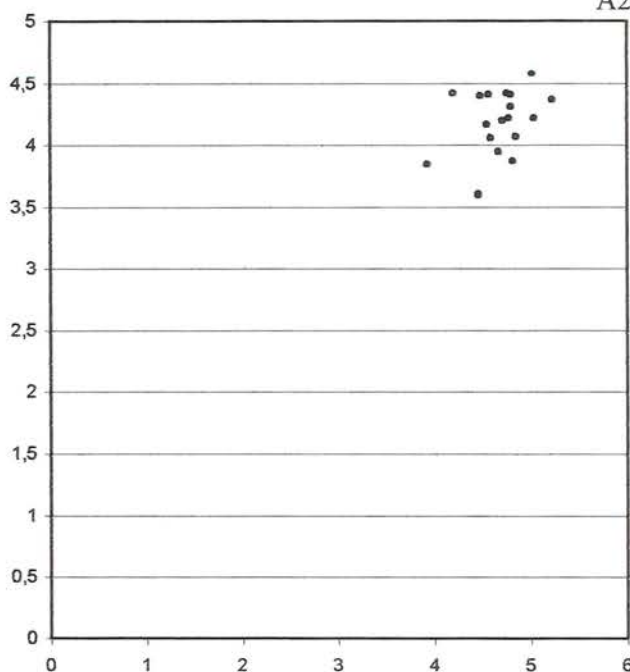


Figura 14. Gráfico mostrando o número de substituições do aminoácido isoleucina nas proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas virais (eixo das ordenadas). Coeficiente de correlação linear = 0,39

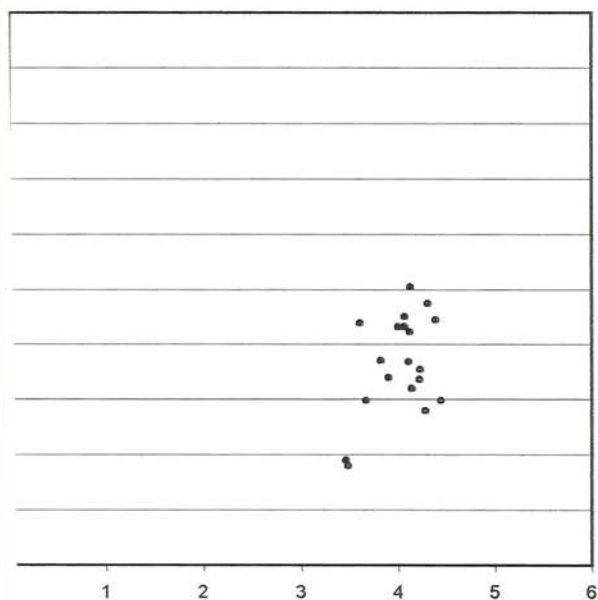


Figura 15. Gráfico mostrando o número de substituições do aminoácido triptofano nas proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas normais (eixo das ordenadas). Coeficiente de correlação linear = 0,46

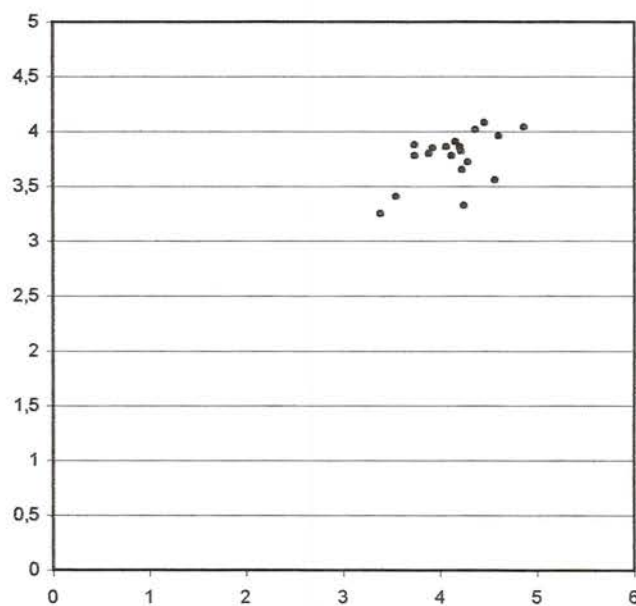


Figura 16. Gráfico mostrando o número de substituições do aminoácido triptofano nas proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas virais (eixo das ordenadas). Coeficiente de correlação linear = 0,52

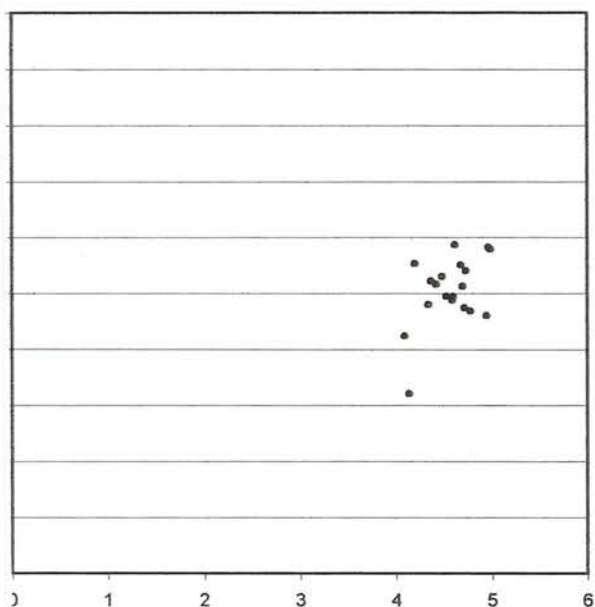


Figura 17. Gráfico mostrando o número de substituições do aminoácido valina nas proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas normais (eixo das ordenadas). Coeficiente de correlação linear = 0,49

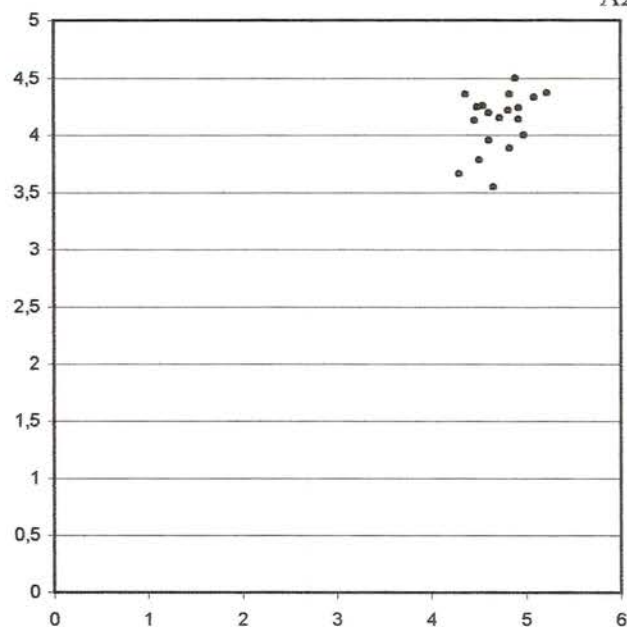


Figura 18. Gráfico mostrando o número de substituições do aminoácido valina nas proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas virais (eixo das ordenadas). Coeficiente de correlação linear = 0,39

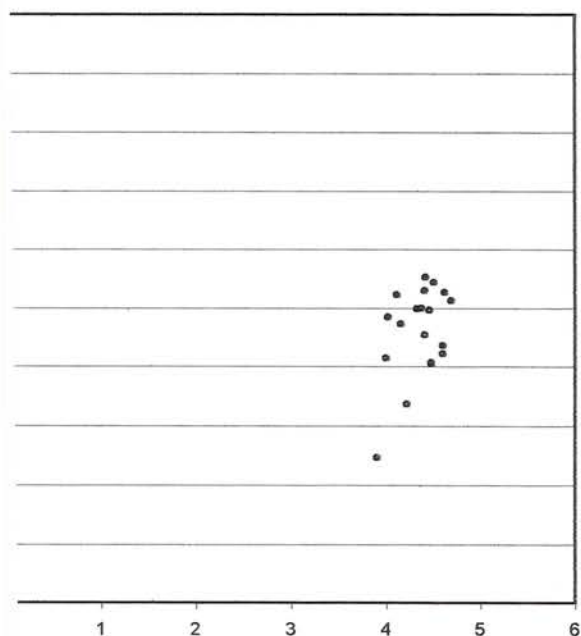


Figura 19. Gráfico mostrando o número de substituições do aminoácido tirosina nas proteínas humanas normais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas normais (eixo das ordenadas). Coeficiente de correlação linear = 0,43

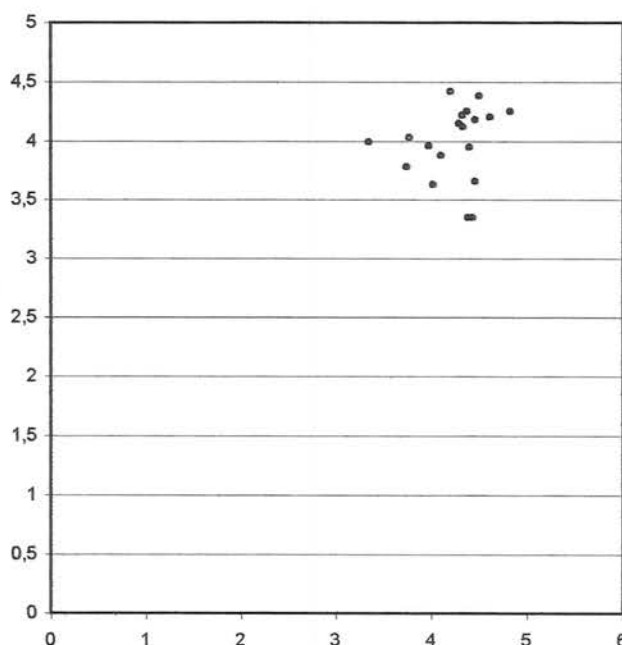


Figura 20. Gráfico mostrando o número de substituições do aminoácido tirosina nas proteínas virais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas virais (eixo das ordenadas). Coeficiente de correlação linear = 0,17

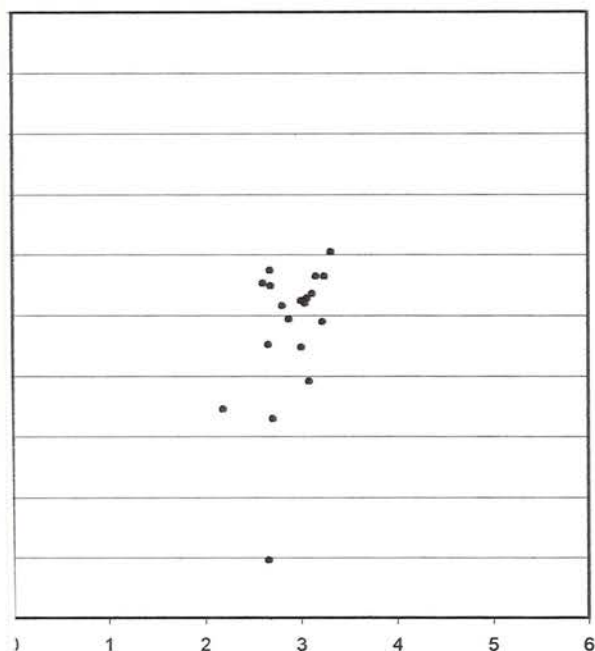


Figura 21. Gráfico mostrando o número de substituições do aminoácido isoleucina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas tumorais (eixo das ordenadas). Coeficiente de correlação linear = 0,46

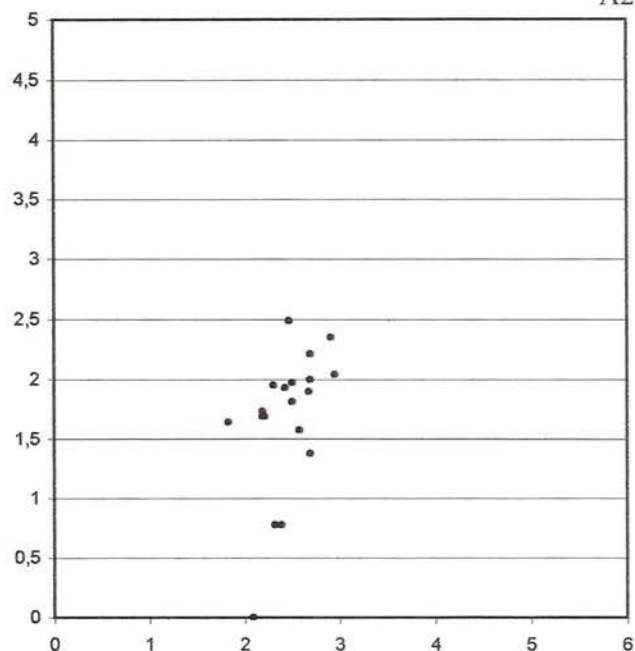


Figura 22. Gráfico mostrando o número de substituições do aminoácido triptofano nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas tumorais (eixo das ordenadas). Coeficiente de correlação linear = 0,37

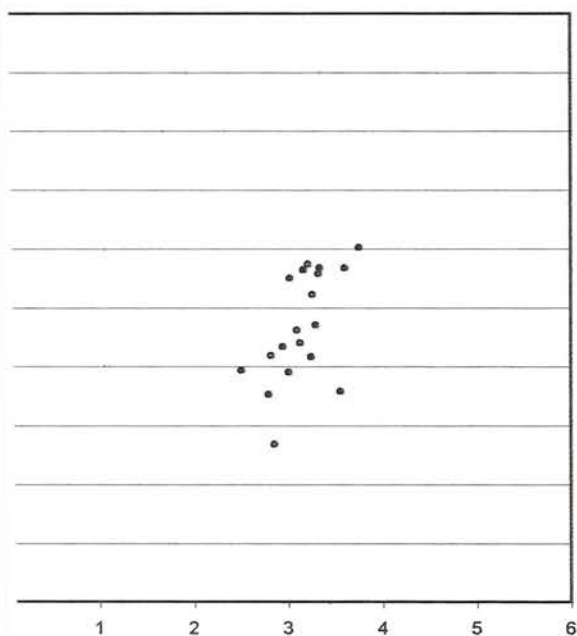


Figura 23. Gráfico mostrando o número de substituições do aminoácido valina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas tumorais (eixo das ordenadas). Coeficiente de correlação linear = 0,56

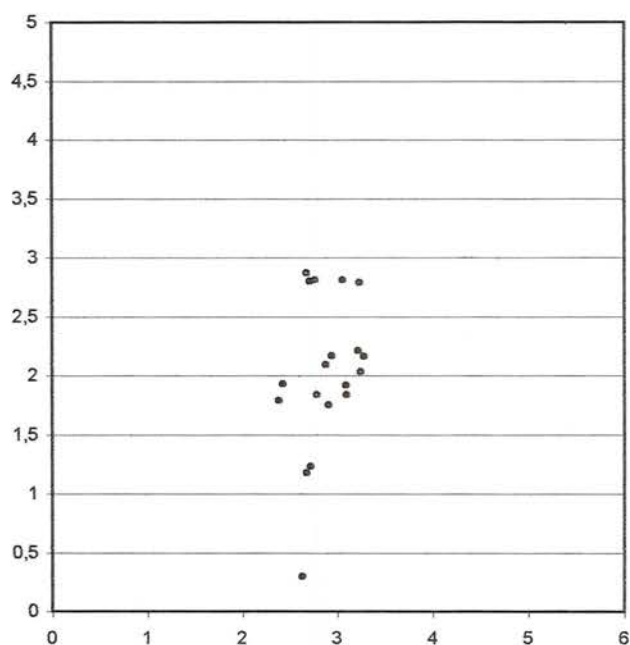
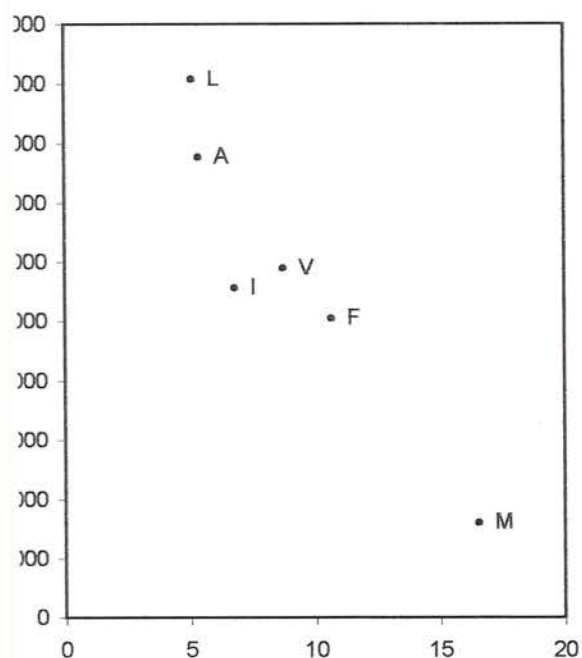


Figura 24. Gráfico mostrando o número de substituições do aminoácido tirosina nas proteínas humanas tumorais (eixo das abcissas) e o número de substituições desse mesmo aminoácido nos peptídeos provenientes de proteínas humanas tumorais (eixo das ordenadas). Coeficiente de correlação linear = 0,31



a 25. Gráfico mostrando a correlação entre o número de substituições observadas nos aminoácidos hidrofóbicos dos peptídeos (eixo das ordenadas) e as áreas expostas desses aminoácidos (eixo das abcissas). Coeficiente de correlação linear = -0,95.

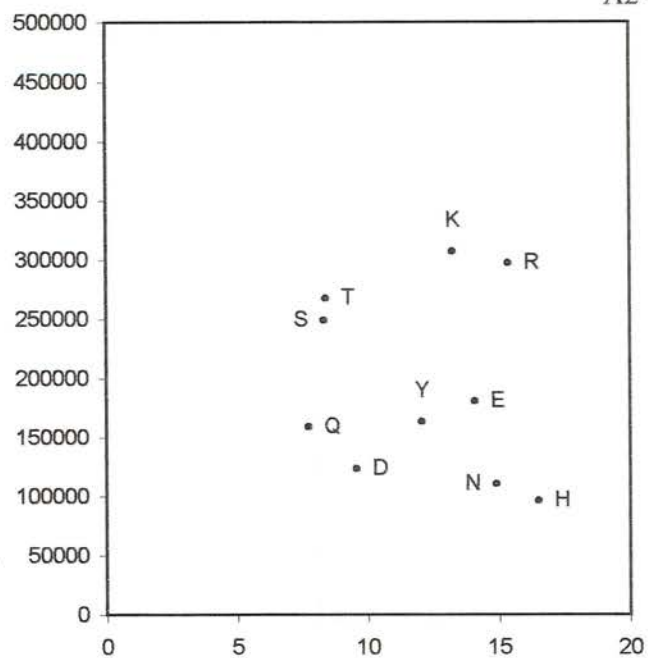


Figura 26. Gráfico mostrando a correlação entre o número de substituições observadas nos aminoácidos hidrofílicos dos peptídeos (eixo das ordenadas) e as áreas expostas desses aminoácidos (eixo das abcissas). Coeficiente de correlação linear = 0,36.