

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO” UNESP

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE MATERIAS
POSMAT**

Paula Martins da Silva

**SIMULAÇÃO COMPUTACIONAL DO
ENOVELAMENTO DE PROTEÍNAS
UTILIZANDO O MODELO HP.**

BAURU

2009

Paula Martins da Silva

Simulação Computacional do Enovelamento de Proteínas utilizando o Modelo HP.

Dissertação apresentada como requisito à obtenção do título de Mestre à Universidade Estadual Paulista “Júlio de Mesquita Filho” Programa de Pós-graduação Institucional em Ciência e Tecnologia de Materiais, sob orientação do Prof. Dr. Aguinaldo Robinson de Souza.

BAURU
2009

DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO UNESP – BAURU

SILVA, Paula Martins da.

Simulação Computacional do Enovelamento de
Proteínas utilizando o Modelo HP.

Paula Martins da Silva, 2007

81 f.

Orientador: Prof. Dr. Aguinaldo Robinson de Souza

Dissertação (Mestrado) – Universidade Estadual

Paulista. Faculdade de Ciências, Bauru, 2009.

*Dedico está dissertação ao meu pai
Nilson Bento da Silva, que foi graças aos seus ensinamentos,
e pautando-me em seus princípios, em sua ética
e em seus exemplos, que tive força e coragem
para lutar e chegar até esta etapa..*

AGRADECIMENTO ESPECIAL

Em primeiro lugar, a Deus que está em tempo integral ao meu lado, iluminando o meu caminho, dando-me saúde, alegria e forças.

Aos meus pais, pela compreensão quando não estive presente em momentos importantes da família.

Ao Professor Aguinaldo Robinson de Souza agradeço a pronta disponibilidade para esclarecimentos de dúvidas e os conselhos científicos muito importantes para a compreensão do problema em estudo, pela amizade conquistada; pela oportunidade de trabalhar em conjunto, com liberdade e de aprender ainda mais; pelo crescimento profissional; pela confiança em mim depositada; pela paciência.

A Rondes Ferreira da Silva, por seu companheirismo. Por me mostrar a cada dia que podemos contar com verdadeiros amigos, não só nos momentos bons e alegres, mas, sobretudo nos momentos de adversidades e dissabores, de tristeza, de desânimo, de dificuldades.

A todos aqueles que não foram citados diretamente aqui, nesses agradecimentos e que tiveram extrema importância na minha formação acadêmica para que eu pudesse cumprir este objetivo pessoal. Sem eles, não teria sido possível realizar este trabalho.

SILVA, P. M. da. **Simulação Computacional do Enovelamento de Proteínas utilizando o Modelo HP** 2009. 81f. Dissertação (Mestrado em Ciência e Tecnologia de Materiais) – UNESP, Faculdade de Ciências, Bauru, 2009.

RESUMO

Simulação Computacional do Enovelamento de Proteínas utilizando o Modelo HP.

Compreender como a seqüência de aminoácidos de uma proteína determina a sua funcionalidade biológica é de grande importância conceitual e prática como, por exemplo, no projeto de novas drogas. As proteínas consistem em polipeptídios composta por 20 L- aminoácidos diferentes, (formados pelos átomos de hidrogênio, carbono, nitrogênio e oxigênio) ligados por ligações peptídicas. O que torna as proteínas diferentes não é o número de aminoácidos, mas a seqüência deles na cadeia polipeptídica. No presente trabalho, apresentamos uma simulação computacional, com um modelo simples, para a enumeração exata de todas as conformações possíveis em uma rede quadrada para $n = 1$ a 17 monômeros. Resultados mostram que 2.155.667 conformações diferentes são possíveis. Neste trabalho, confirmamos estes resultados e atribuímos aos sítios da rede quadrada os monômeros no modelo que foram usados para computar o número de contatos entre os monômeros, distribuição da energia configuracional, distância em relação a frequência relativa de todas as conformações, contato topológico de longo alcance, aplicamos e comparamos os resultados das simulações do modelo com seqüências de Epitopos.

Palavra Chave: Enovelamento de Proteínas, Modelo HP, Simulação Computacional.

ABSTRACT

The understanding of how the amino acids sequence determine the protein biological functionality is one of the most practical and conceptual challenge as, for example, in the development of new drugs. The protein molecule consist of a long chain of polypeptides composed of 20 different amino acids (essentially formed by carbon, oxygen, nitrogen, and hydrogen atoms) linked together by peptide chemical bonds. What makes proteins different is not the number of amino acids in the chain but the amino acid sequence along the chain. In the present work we present a computer simulation of a simple protein model, using the technique of exact enumeration over all the possible conformations in the square lattice to a sequence of $n = 1$ to 17 monomers. The results indicate that we can found 2.155.667 different possible conformations. We present the results for the Cartesian coordinates of all the monomers along the chain and computed the monomer-monomer contacts, the configurational energy, the long-range topological contact and a possible application of the HP model for a variety of epitopes sequences.

Keywords: protein folding, HP model, computer simulation

LISTA DE FIGURAS

Figura 1 - Estrutura geral dos aminoácidos, com destaque para o carbono alfa e para a cadeia lateral "R" característica para cada aminoácido.....	1
Figura 2 – Família dos 20 aminoácidos e suas classificações.	18
Figura 3 - Esquema representativo dos diferentes níveis de estrutura protéica da molécula de hemoglobina.	19
Figura 4 – Demonstração da estrutura primária.	19
Figura 5 – Representação de uma letra, no Modelo HP, dos 20 aminoácidos classificados como hidrofóbicos e hidrofílicos.	24
Figura 6 – Vizinhos topológicos em vermelho.	25
Figura 7 – Representação da movimentação permitida aos monômeros.	29
Figura 8 - Duas conformações diferentes da mesma seqüência diferenciando-se somente por uma reorganização do núcleo dos aminoácidos.....	29
Figura 9 - Representação gráfica da proteína no modelo.....	31
Figura 10 - Uma imagem de um linfócito feita a partir de um microscópio eletrônico.....	34
Figura 11 - Uma imagem de um fagócito.....	34
Figura 12 - Mecanismos de defesa e seus principais mediadores.....	34
Figura 13 – Representação dos três graus de liberdade.....	40
Figura 14 - Diversas formas de crescimento no algoritmo, sempre o ponto (0,0) estará fixo.....	43
Figura 15 - Configurações com problemas de rotação e simetria.....	43
Figura 16- Crescimento de uma cadeia do Algoritmo: Na ilustração à esquerda, das três possibilidades para a adição de uma unidade extra, somente duas delas (as duas mais à esquerda) satisfazem a condição de configuração <i>Self Avoid Walk</i>	47
Figura 17 – Saída da implementação do algoritmo para se obter os vizinhos topológicos.....	54
Figura 18 – Energia configuracional total do sistema.....	55
Figura 19 – Distribuição total da distância em relação a frequência.....	59
Figura 20 - Distribuição das populações de 2.155.667 configurações possíveis para 17 monômeros, onde utilizamos a nomenclatura de Número Total de Configurações (NTC) em função da ordem de contato, que fornece o número	

de monômeros existentes entre os pares dos primeiros vizinhos (de monômeros) em contato topológico, chamado de Número Total de Contatos Vizinhos Topológicos (NTCVT).....	60
Figura 21 – Gráfico da frequência da distância 3 em função do número de monômeros na cadeia.	61
Figura 22 – Representação da distância 3 para os monômeros [1,4] nas cadeias com:.....	62
Figura 23 – Gráfico na forma logarítmica da frequência da distância 3 em função do número de monômeros na cadeia.....	63
Figura 24 – Distância total em relação à frequência da cadeia de L= 16.	64
Figura 25 – Distância total em relação à frequência da cadeia 17.....	64
Figura 26 – Contatos de longo alcance, entre resíduos distantes da seqüência.	65
Figura 27 - (a) Tabela da distância em função da frequência ao longo da cadeia. (b) Gráfico da distância ao longo da cadeia em relação a frequência. Inserido no gráfico representamos a equação de melhor ajuste aos pontos e o coeficiente de correlação.....	66
Figura 28 – Distribuição total das 2.155.667 configurações em função dos contatos topológicos dos primeiros vizinhos na rede quadrada.	67
Figura 29 - Banco de Dados Immune Epitope Database and Analysis Resource	68
Figura 30 – Modelo HP aplicado a Epitopo.	69
Figura 31 – Modelo HP aplicado ao Epitopo AALFMYAK	70
Figura 32 – Distância em relação a frequência e energia configuracional.	70
Figura 33 – Seqüência do Epitopo possui H-H apenas nos contatos [1,4].....	71
Figura 34 – Interações P-P e H-P de acordo com o Modelo não são computados.	71
Figura 35 – Contato topológico de longo alcance, obtemos na distância [1,8] .	72
Figura 36 – Modelo HP aplicado ao Epitope de seqüência AALIEEGQR	73

SUMÁRIO

Introdução	12
Capítulo 1	15
1. Proteínas.....	15
1.1 Aminoácidos.....	17
1.2 Estrutura da Proteína	19
1.3 Modelos de Proteínas e Biomateriais.....	22
1.4 Enovelamento de Proteínas.....	23
1.5 Paradoxo de Levinthal	27
1.6 Modelo HP.....	28
1.7 Modelo de Energia Livre.....	30
Capítulo 2.....	32
2.1 <i>Sistema Imunológico</i>	32
2.2 <i>Antígeno</i>	35
2.3 <i>Epítomos</i>	35
2.3.1 <i>Epítomos de célula T</i>	36
2.3.2 <i>Antígeno T independente</i>	36
2.3.3 <i>Interação antígeno-anticorpo</i>	36
Capítulo 3.....	38
Modelagem Computacional – Algoritmo	38
3.1 Modelagem Computacional	38
3.2 A Técnica de Enumeração Exata	40
3.3 Algoritmo Self Avoid Walk	42
3.3.1 O algoritmo aleatório auto-excludente:	44
3.3.2 Implementação:.....	44

3.3.3	Considerações preliminares sobre algoritmo:	45
3.4	Algoritmo Vizinho Topológicos	48
3.5	Hardware & Software:	49
Capítulo 4	50
4.	Análise dos Resultados.....	50
4.1	Resultado do Algoritmo Self Avoid Walk	51
4.2	Resultados do Algoritmo Vizinhos Topológicos	53
4.3	Distribuição da energia segundo parâmetros topológicos.....	55
4.4	Distribuição da distância entre os contatos topológicos.....	56
4.5	Freqüências dos Vizinhos Topológicos Distância [1,4],.....	61
	<u>em função do número de cadeias.....</u>	61
4.6	Contato Topológico de Longo Alcance	65
4.7	Mapa de Contato Topológico.....	67
4.8	Modelo aplicado a Epitopo.....	68
Considerações Finais	74
Referências Bibliográficas	75

INTRODUÇÃO

Motivação – Objetivo – Organização da dissertação

Existem caminhos para as proteínas envelerem?

Levinthal, 1968.

Este trabalho utiliza um modelo computacional, (ZHAO, 2008), (CROUT, 2009), (CHANDRU, DATTASHARMA, KUMAR, 2003), para tratar de um sistema complexo e facilitar o entendimento das possíveis conformações adotadas para uma proteína. (QIAN, RAMAN,DAS,BRADLEY, MCCOY, READ, BAKER, 2007) através da representação da cadeia polipeptídica com 17 aminoácidos, conectados linearmente e restritos a ocuparem os sítios em um espaço configuracional bidimensional (CALDARELLI, RIOS, 2007).

As estruturas nativas de uma proteína globular, no modelo computacional, são representadas por configurações em uma rede quadrada contendo somente contatos de primeiros vizinhos topológicos. Nestas configurações, são consideradas somente interações energéticas entre elementos da cadeia apolares (H-H), desprezando todas as outras interações polar-polar (P-P), e apolar-polar (H-P). As interações assim, definidas representam a energia hidrofóbica do sistema cadeia-solvente, a qual, conforme constatação da literatura, é a responsável pela força dominante no processo de envelamento, (SALI, SMITH, DOBSON, KARPLUS, 2000), (WHITFORD, NOEL, GOSAVI, SCHUG, SANBONMATSU, ONUCHIC, 2009), (IRBACK, TROEIN, 2001). A representação da função energia é simples:

associa-se energia -1 (unidades arbitrárias) para cada contato H-H direto entre ocupando vizinhos não adjacentes em um espaço configuracional de uma rede quadrada e 0 para as demais interações. Apesar da simplicidade do modelo HP, o processo apresenta um enovelamento semelhante com o processo do sistema real (ZHAO, 2008), sendo utilizada para a formulação de novas hipóteses de conformação estrutural da proteína, isto é, um conjunto discreto de possíveis conformações tornando possível enumerar e considerar todas elas. A energia da conformação pode ser descrita a partir da escolha possível de resíduos H ou P monômeros ao longo da cadeia, (WROE, BAUER, CHANZ, 2005).

Este modelo é tratado por modelagem computacional utilizando um programa em código Fortran 90, implementado com diversas sub-rotinas especificamente desenvolvidas.

O método utilizado é a enumeração exata, com a dinâmica dada pelo algoritmo de *self-avoiding-walk*, SAW, isto é, cada proteína é representada por uma configuração (estrutura) contendo duas matrizes ordenadas de acordo com as coordenadas cartesianas (DETHRIDGE, GUTTMANN, 2008), (WILLIAMS, 1996). A energia de cada configuração na estrutura da proteína de todas as 2155667 configurações possíveis para uma cadeia de 17 aminoácidos é calculada e armazenada (CALDARELLI, RIOS, 2007). A partir daí, obtemos os resultados relacionando os primeiros vizinhos topológicos com a distribuição da energia. Também é verificado, por meio de simulação Monte Carlo, a frequência relativa na configuração nativa, dada pelo número relativo de vezes em que a cadeia visita esta configuração (CALIRI, 1997). Finalmente, aplicamos o modelo a diversas seqüências de Epítopes.

Esta dissertação está dividida em quatro capítulos, sendo a introdução um breve relato sobre a motivação, o objetivo e a estrutura da pesquisa.

O primeiro capítulo abordará o polímero natural, proteína, o modelo computacional, o conceito de vizinho topológico e o estado nativo das proteínas. Apresentamos um levantamento bibliográfico sobre as concepções de espaço bidimensional e o conceito de um modelo simplificado para heteropolímeros. Inicialmente, definiremos o modelo que utilizamos neste trabalho, e, em seguida, discutiremos os trabalhos existentes na literatura, que utilizam esses modelos para estudar o processo de enovelamento.

Apresentaremos, no segundo capítulo, o sistema imunológico, antígeno, *epítope* e conceitos para compreender como o sistema imunológico funciona e, assim, aplicar, posteriormente, o modelo HP.

O terceiro capítulo é dedicado a descrição do funcionamento do algoritmo. O sistema computacional tem sido usado no esforço para se deduzir maneiras de entender o enovelamento.

O quarto capítulo apresenta e discute os resultados das simulações, onde aplicamos o modelo HP em um epítope, analisamos a frequência dos contatos nativos obtidos durante a simulação, isto é, onde há maior probabilidade de enovelamento com a distância em relação a sua frequência de todas as configurações.

Finalmente, apresentaremos a conclusão e as referências bibliográficas.

CAPÍTULO 1

Proteínas, Modelo, Método, Vizinho Topológico e Estado Nativo.

Ciência é um processo de representação do mundo, sempre sujeito a reformulações. (Veit;Mors;Teodoro, 2002)

1. Proteínas

A palavra proteína deriva da palavra grega “proteios”, que significa primeiro (ZHAO, 2008). Esta designação tem origem na elevada importância das proteínas na vida, uma vez que as proteínas têm um elevado peso na constituição dos organismos vivos (RADFORD, 2000).

Proteína é um polímero natural formado por aminoácidos, que são ligados uns aos outros através de ligações peptídicas (ZHAO, 2008), (RADFORD, 2000). A função que uma determinada proteína exerce em um organismo está intrinsecamente ligada à sua estrutura tridimensional que, por sua vez, depende da seqüência de aminoácidos que a compõe (CHARK, 2008), (DILL, OZKN, WEIKL, CHODERA, VOELZ, 2007).

Desempenhando funções biológicas fundamentais para o equilíbrio dos organismos vivos (LEHNINGER, 1991) as proteínas dividem-se em duas grandes classes: proteínas fibrosas, insolúveis em água, e proteínas globulares, solúveis em água ou em soluções aquosas (DILL, 1985). Esta

classificação está relacionada com a forma das moléculas. As proteínas fibrosas são estruturas alongadas e filamentosas (LEHNINGER, 1991). As proteínas globulares apresentam estruturas moleculares compactas, consiste de uma ou mais cadeias polipeptídicas lineares, e cada uma dessas, sob condições fisiológicas, é enovelada na forma de uma estrutura complexa tridimensional (LI,TANG, WINGREEN, 1997), (ANANIKIAN, ANANIKYAN, ARTUSO, 2006). A esse arranjo final da estrutura enovelada, dá-se o nome de conformação nativa da proteína; acredita-se que a funcionalidade de uma proteína é determinada primariamente por sua conformação nativa (RADFORD, 2000). Desta forma, é possível realizar uma previsão de qual função uma determinada proteína desempenhará num organismo. (SILVA, 2002).

Contudo, pode acontecer que, devido a alteração no meio (pH, temperatura ou presença de agentes químicos etc.) podem levar à destruição (desnaturação) (SALVI, MOBERT, RIOS, 2002), (CALDARELLI, RIOS, 2007). Desta forma, a proteína perde a estrutura tridimensional que lhe permite desempenhar as suas funções no organismo (RADFORD, 2000). Um exemplo simples de desnaturação ocorre com o ovo, isto é, as proteínas da clara do ovo, quando sujeitas a temperaturas elevadas ou sob agitação, desnaturam e precipitam. A perda da estrutura tridimensional de uma proteína leva à perda das suas funções biológicas; em alguns casos, leva a formação de agregados fibrilares, os quais se depositam nos tecidos, podendo originar diversas doenças (RIOS, GALDARELLI, 2000).

1.1 Aminoácidos

As proteínas são constituídas por aminoácidos (CRESCENZI, GOLDMAN, PAPADIMITRIOU, PICCOLBONI, YANNAKAKIS, 1998). Cada uma dessas unidades é um ácido orgânico COOH – no qual o carbono próximo do grupo (chamado carbono alfa) está também ligado a um grupo NH_2 – . O carbono alfa também se liga a uma cadeia lateral (R), que é diferente para cada aminoácido (Figura 1), isto é, as características de cada aminoácido variam consoante a composição química da cadeia lateral (THALHEIM, MERKLE, MIDDENDORF, 2006), (CHANDRU, DATTASHARMA, KUMAR, 2003), (ZHAO, 2008). Os vinte aminoácidos existentes na natureza foram classificados, quanto à sua cadeia lateral, em quatro grupos: ácidos, básicos, polares e não polares (CALDARELLI, RIOS, 2007),

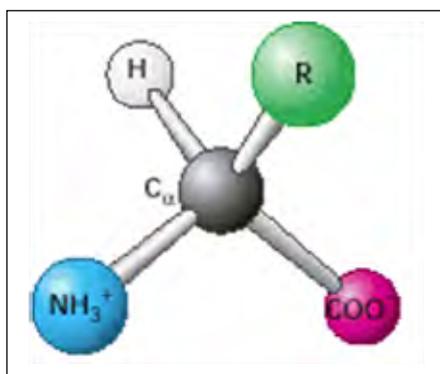


Figura 1 - Estrutura geral dos aminoácidos, com destaque para o carbono alfa e para a cadeia lateral “R” característica para cada aminoácido.

Aminoácidos hidrofóbicos ou não polares possuem cadeias laterais compostas de hidrocarbonetos e tendem a se manter juntos na presença de água para minimizar seu contato com água. Aminoácidos hidrofílicos ou polares possuem grupos apolares (com oxigênio ou nitrogênio) em suas cadeias

laterais e podem estar em contato com água (CHEN, WEBER, HARRISON, 2008).

Na (Figura 2), é possível observar as famílias de aminoácidos, sendo esta classificação importante para a análise dos dados, e assim compararmos alguns dos resultados obtidos.

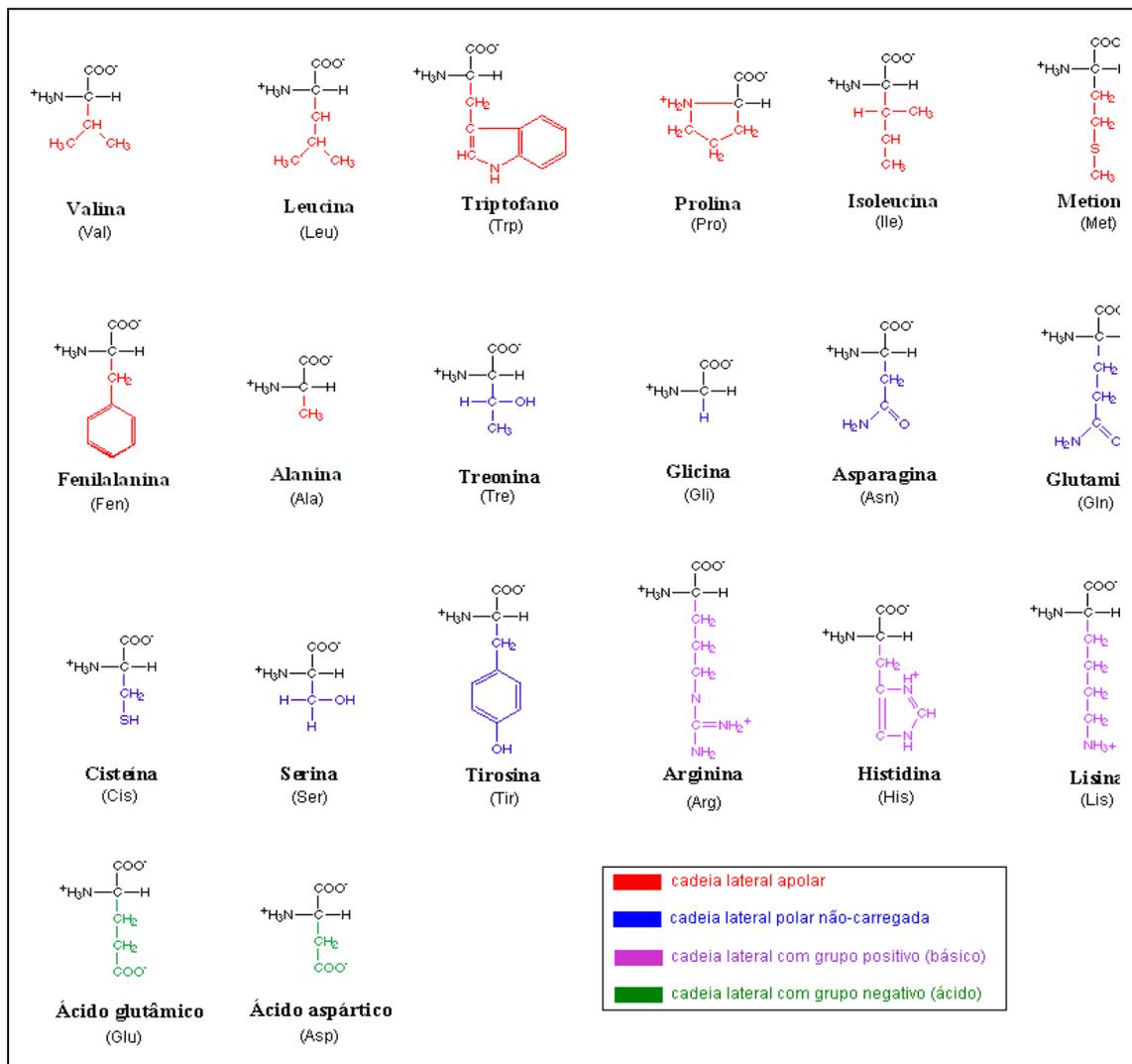


Figura 2 – Família dos 20 aminoácidos e suas classificações.

1.2 Estrutura da Proteína

A estrutura tridimensional de uma proteína pode ser descrita em quatro níveis de complexidade (SILVA, 2002), (Figura 3).

Fonte: QUÍMICA NOVA NA ESCOLA N° 24, NOVENBRO 2006. Proteínas como tema para o ensino de Química.



Figura 3 - Esquema representativo dos diferentes níveis de estrutura proteica da molécula de hemoglobina.

- **Estrutura primária:** Consiste na seqüência linear aleatória de aminoácidos que formam uma cadeia polipeptídica (Figura 4) . A seqüência de aminoácidos tem elevada importância, pois basta a troca de dois aminoácidos, ou a substituição de um qualquer para produzir importantes alterações biológicas (SILVA, 2002), (TARRAGÓ, 2003), (DILL, BROMBERG, FIEBIG, YEE, THOMAS, CHAN, 1995).



Figura 4 – Demonstração da estrutura primária.

- **Estrutura secundária:** é o resultado do arranjo espacial entre aminoácidos próximos na cadeia peptídica. Dois dos principais elementos de estrutura secundária são:

- α - *hélice*: tipo mais comum da estrutura secundária que se caracteriza por uma estrutura helicoidal, estabilizada por ligações de hidrogênio estabelecidas entre os grupos CO – e NH – de aminoácidos que distam entre si 4 resíduos.

- *folha β* : estrutura estendida, constituída por associação lateral de segmentos peptídicos numa orientação paralela ou anti-paralela, mantida por ligações de hidrogênio (SILVA, 2002), (TARRAGÓ, 2003)..

- **Estrutura terciária:** A estrutura terciária está relacionada com as dobras da cadeia protéica sobre ela mesma. Esta estrutura descreve um arranjo espacial global dos aminoácidos numa proteína, podendo estes estar bastante separados entre si na seqüência linear (SILVA, 2005).

- **Estrutura quaternária:** Certas proteínas são constituídas por mais de uma cadeia protéica. A estrutura quaternária é o resultado da disposição espacial relativa de subunidades protéicas em proteínas complexas, formadas por duas ou mais cadeias polipeptídicas. (TARRAGÓ, 2003), (SILVA, 2005).

Cada proteína tem uma estrutura primária e tridimensional única e, como tal, cada proteína desempenha uma função biológica específica (RADFORD, 2000). Segundo Kauzmann, 25 a 40% dos aminoácidos de uma cadeia peptídica são hidrofóbicos. Este tipo de aminoácido tem pouca afinidade

com a água; logo, espera-se que as proteínas tendam a enovelar-se de tal forma que esses aminoácidos fiquem no interior da molécula. As ligações de hidrogênio desempenham um importante papel no enovelamento das cadeias polipeptídicas e na formação da estrutura secundária (CHEN, WEBER, HARRISON, 2008).

1.3 Modelos de Proteínas e Biomateriais

O estudo de modelos de proteínas, tanto do ponto de vista experimental como de simulação computacional, tem o potencial de diversas aplicações na área de biomateriais ou materiais com interesse na biologia.

Uma das possíveis aplicações diz respeito ao problema da compatibilidade entre materiais de implantes e os tecidos biológicos. Conhecer o comportamento das biomoléculas como, por exemplo, proteínas, ácidos nucléicos e enzimas, pode levar à um maior entendimento de como os efeitos do meio podem influenciar na atividade biológica destas moléculas.

O modelo aqui apresentado pode ser estendido também para o estudo de polímeros sintéticos na busca do entendimento da variação de suas propriedades conformacionais como a distância final-final e o raio de giração em função tanto do número de monômeros com dos efeitos ambientais.

1.4 Enovelamento de Proteínas

Para poder exercer uma função, a proteína enovela-se, cujo processo pelo qual a informação linear contida na seqüência de aminoácidos de um polipeptídeo dá origem à conformação tridimensional bem definida da proteína funcional (HART, ISTRAIL, 1996). Neste estado, a proteína está com o máximo de sua organização e demonstra atividade biológica (catálise, transporte, defesa etc).

Computacionalmente, para reproduzir a estrutura de uma proteína, é necessário ter um modelo preciso que abstraia a formação da estrutura real em um nível de detalhes desejado. (CALDARELLI, RIOS, 2007), (QIAN, RAMAN,DAS,BRADLEY, MCCOY, READ, BAKER, 2007).

Um modelo é relevante se reflete alguma propriedade da formação da estrutura da proteína no sistema real; no entanto, poderíamos buscar alguma equivalência entre a formação da estrutura da proteína no modelo e a formação da estrutura no sistema real. (ZHAO, 2007). Como as leis da termodinâmica afirmam que o estado nativo de uma proteína é o estado de menor energia livre, (SALVI, MOLBERT, RIOS, 2002), o sistema real é freqüentemente modelado pela busca das conformações de uma proteína nativa que minimizem a energia configuracional ao longo de um conjunto de conformações (DILL, OZKAN, WEIKL, CHODERA, VOELZ, 2007).

O modelo proposto por Dill, (1989) é um modelo de energia livre das conformações importantes para a conformação nativa de uma proteína. Devido às interações entre os aminoácidos hidrofóbicos e hidrofílicos, ocorre a formação de um núcleo envolvendo o solvente e os aminoácidos. No modelo HP, uma proteína é descrita como uma seqüência de aminoácidos hidrofóbicos

e hidrofílicos. (CHANDRU, DATTASHARMA, KUMAN, 2003). Apesar de alguns aminoácidos não poderem ser classificados como sendo claramente hidrofílicos ou hidrofóbicos, o modelo ignora este fato (ZHAO, 2007). O modelo é normalmente referido como o modelo HP, em que H representa os aminoácidos hidrofóbicos, ou apolares e P os polares que são os hidrofílicos, (CALDARELLI, RIOS, 2001) (Vide Figura 5).

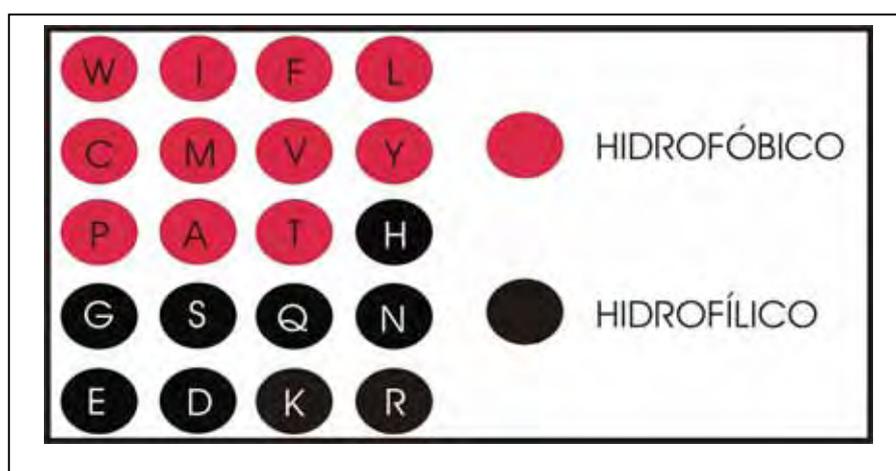


Figura 5 – Representação de uma letra, no Modelo HP, dos 20 aminoácidos classificados como hidrofóbicos e hidrofílicos.

Na rede bidimensional, os aminoácidos que são adjacentes na seqüência ocupam pontos na rede, (ZHAO, 2008) e nenhum ponto na rede é ocupado por mais de um aminoácido (DILL, OZKAN, WEIKL, 2007). A Figura 6, mostra a conformação no Modelo 2D, onde são representados nove aminoácidos não-adjacentes hidrofóbicos na rede (THALHEM, MERKLE, MIDDENDORF, 2006).

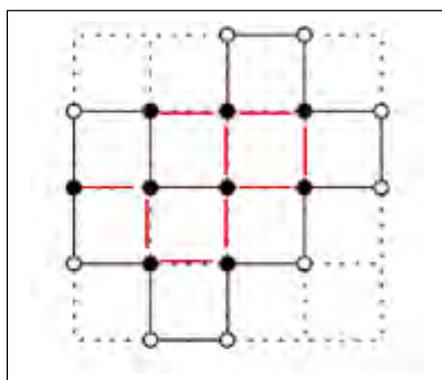


Figura 6 – Vizinhos topológicos em vermelho.

Apesar da simplicidade do modelo HP, o processo de enovelar tem semelhanças com o processo de enovelar no sistema real (CALDARELLI, RIOS, 2003). O modelo tem sido usado em vários estudos para avaliar a hipótese de formação da estrutura protéica. Devemos salientar que o estudo do enovelamento de proteínas tanto na rede quadrada como na rede cúbica buscam o entendimento do processo em termos de suas características gerais (KAUZMANN, 1964), (RADFORD, 2000), (TARRAGÓ, 2003), (CHANDRU, DATTASHARMA, KUMAR, 2003).

Muitas tentativas têm sido feitas para prever a conformação nativa, ou seja, a conformação de menor energia livre de uma proteína no Modelo HP. Mais interessante este foi o primeiro modelo relevante para o enovelamento da proteína com um algoritmo para a predição do problema, ou seja, algoritmo que encontrasse uma conformação perto do estado nativo (CALDARELLI, RIOS, 2007), (RIOS, CALDARELLI, 2000).

Durante algum tempo acreditava-se que a previsão da estrutura no modelo HP seria resolvido em tempo polinomial. Recentemente este problema tem se mostrado como de ordem NP-completo (ARNALDI, LIBERTI, MAFFIOLI, MACULAN, 2009), (CHANDRU, DATTASHARMA, KUMAR, 2003).

Na década de 50 do século XX, Christian Anfinsen procurou dar resposta às seguintes questões: Por que a proteína enovela para a estrutura nativa? Será ajudada por outras moléculas? Por que é que a estrutura nativa é única? (CLARK, 2008).

Para responder a estas questões, Anfinsen realizou várias experiências com a proteína RNaseA que deram origem à hipótese termodinâmica. Logo, o estado mais estável no enovelamento da proteína é o estado nativo, porque corresponde ao estado de energia mais baixo.

Segundo esta hipótese, o estado nativo é único porque é determinado por uma seqüência de aminoácidos única, e a procura do estado nativo é aleatória.

Em 1968, Cyrus Levinthal mostrou que a hipótese termodinâmica não resolvia todas as questões no enovelamento proteico..

1.5 **Paradoxo de Levinthal**

A hipótese de que o processo de enovelamento ocorre de forma aleatória através de todas as conformações possíveis, até encontrar a conformação correta foi refutada em 1968 por Cyrus Levinthal por meio de um argumento simples.

Considere-se uma proteína com 100 aminoácidos e cada aminoácido com 2 conformações possíveis (porque dois é o número mínimo de conformações por aminoácido), o que se obteriam seriam $2^{100} \sim 10^{30}$ conformações possíveis, dentre as quais está a que corresponde ao estado nativo. Considere-se, ainda, que a proteína levaria 1 picosegundo a tentar cada combinação, que é o tempo de vibração térmica, ~~então~~, o tempo necessário para explorar todo o espaço conformacional seria 10^{18} s, que é aproximadamente 32 bilhões de anos (maior do que a idade do Universo, cerca de 13,7 bilhões de anos). Levinthal concluiu que a hipótese termodinâmica não explica a escala de tempo do processo de enovelamento das proteínas (LEVINTHAL, 1968), e para resolver o problema, propôs a existência de “caminhos específicos” para a estrutura nativa, o que corresponde à Hipótese Cinética, na qual o estado nativo não corresponde necessário ao estado termodinâmico mais estável, mas sim, ao estado de energia mínima que é o mais acessível do ponto de vista cinético (DILL, CHAN, 1997). Segundo Levinthal, na década de 90, verificou-se que esta hipótese por si só também não resolvia a questão e surgiu uma teoria unificada, que se baseia na natureza estatística do processo (HONIG, 1999), (LEVINTHAL, 1968), (DILL, CHAN, 1997), (KARPLUS, 1997), (ZHAO, 2008).

1.6 Modelo HP.

Trata-se de um modelo simplificado ou discreto, (ZHAO, 2008), (CROUT, 2009), (CHANDRU, DATTASHARMA, KUMAR, 2003) em que a visão sobre as proteínas torna-se diferente, pois são modeladas como polímeros, onde o solvente é a água, (CHEN, WEBER, HARRISON, 2008) no qual cada monômero representa um aminoácido, reduzindo extremamente o número de conformações acessíveis (LU, LIANG, 2007), (YAP, FAWZI, GORDON, 2008).

O Modelo HP é a abstração mais simples possível do problema de enovelamento de proteínas.; a propriedade química e eletrostática dos aminoácidos também foi simplificada (ALBRECHT, SKALIOTIS, STEINHOFEL, STOCHASTIC, 2008). As interações principais neste modelo, que estabiliza a proteína globular, advêm da parte hidrofóbica dos aminoácidos não-polares e a interação dos aminoácidos com as moléculas de água (ZHAO, 2008).

Conseqüentemente, as propriedades importantes dos aminoácidos são reduzidas para duas: polar (P) ou não-polar (H), (BUI, SUNDARRAJ, 2005). Assim, o modelo é considerado simplificado, pois todos os detalhes dos aminoácidos são descartados, o espaço foi discretizado (rede quadrada) e somente dois tipos de aminoácidos são considerados: hidrofóbicos e os hidrofílicos.

Vários métodos têm sido propostos para estudar as propriedades de cadeias heteropoliméricas, (CALDARELLI, RIOS, 2001), contudo, no modelo HP, são estudadas propriedades internas da cadeia. Uma proteína é modelada como um heteropolímero, (DILL, 1985) no qual cada monômero representa um aminoácido como um todo. Uma configuração do polímero é uma seqüência definida e isolada numa rede quadrada, (SILVA,2005), (TARRAGÓ, 2003).

Obtemos vantagem em estudar modelos simples; há possibilidade de análises mais rigorosas que determinam caminhos detalhados, pois o espaço das configurações pode ser totalmente conhecido (LAU, DIU, 1989). No modelo HP, 17 monômeros enovelam-se em um tempo pequeno, comparado ao tempo de enovelamento das proteínas reais.

Para esta rede, o enovelamento é representado como uma seqüência de comprimento $L-1$, onde L é o total de aminoácidos (ZHAO, 2008). Cada coordenada depende de sua coordenada anterior, e o primeiro aminoácido da cadeia é sempre fixo (Figura 7) (REGO, LI, GLOVER, 2006), (WILLIAMS, 1996).

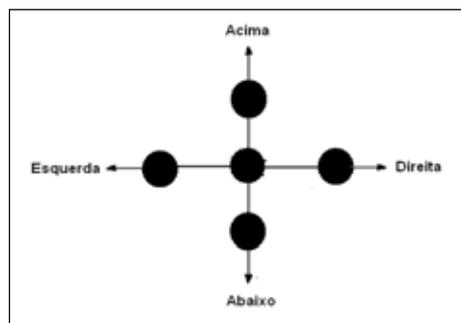


Figura 7 – Representação da movimentação permitida aos monômeros.

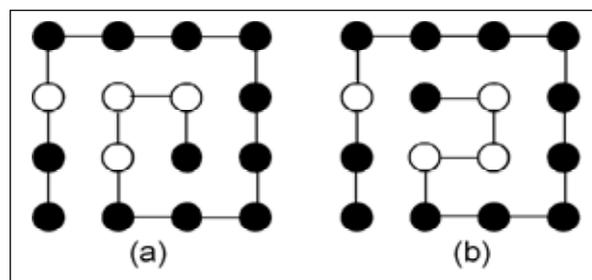


Figura 8 - Duas conformações diferentes da mesma seqüência diferenciando-se somente por uma reorganização do núcleo dos aminoácidos.

As interações ocorrem entre monômeros vizinhos mais próximos não consecutivos ao longo das seqüências, chamados de vizinhos topológicos,

(THALHEIN, MERKLE, MIDDENDORF, 2006). Entretanto, fazendo uma analogia com proteínas reais, as interações possíveis ocorrem entre vizinhos mais próximos, não ligados covalentemente (PEDERSEN, LYNGS, 2000). Contudo, neste modelo, a energia livre de uma conformação é proporcional ao número de contatos topológicos não locais na seqüência dada (Figura 8).

1.7 Modelo de Energia Livre

A idéia básica para se definir a energia livre está relacionada a escolha do modelo. Cada modelo possui uma função de partição onde se calcula a energia livre do sistema (SALI, SMITH, DOBSON, KARPLUS, 2000). A Figura 9 mostra um exemplo de uma mesma proteína onde o nível de detalhes de sua representação é decrescente da esquerda para a direita.

A proteína, da esquerda está sendo representada cristalograficamente; a do meio, não leva em consideração os átomos dos resíduos com um todo, por sua vez, a da direita, além de não considerar os átomos dos resíduos, limita as ligações em coordenadas cartesianas em uma rede quadrada. A função de partição do modelo de energia livre, sugerida:

$$\sum_{(i,j)} (i+j)$$

Onde:

- $i+j = 1$, se os aminoácidos i, j formam uma ligação não consecutiva.
- $i+j = 0$, caso contrário

Dependendo da característica do contato entre os aminoácidos, (WHITFORD, NOEL, GOSAVI, SCHUG, SANBONMATSU, ONUCHIC, 2009), (IRBACK, TROEIN, 2001) a energia será calculada de acordo com as interações correspondendo aos contatos H-H, H-P ou P-P, respectivamente. (TANG, 2000).

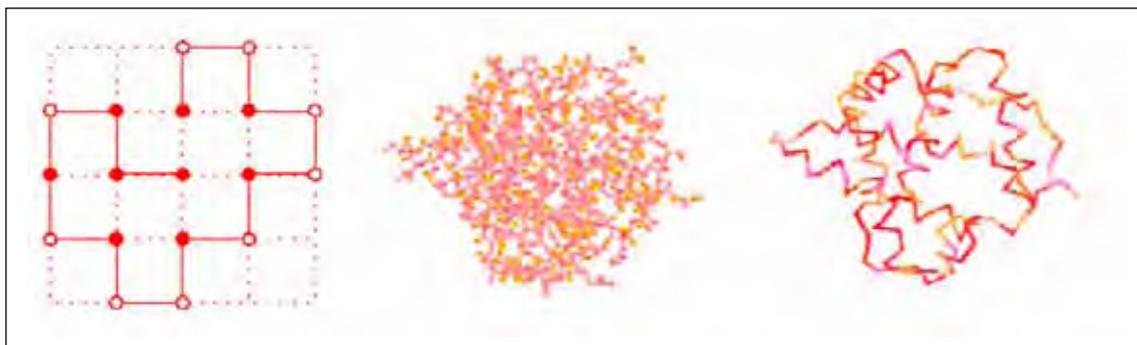


Figura 9 - Representação gráfica da proteína no modelo.

Este modelo permite-nos analisar comportamentos semelhantes entre os resultados encontrados e as conformações reais das proteínas, não sendo tão usado para comparação de uma proteína real, pois o grau de simplicidade é muito grande. Existem outros modelos discretos menos conhecidos e que podem ser aplicados no enovelamento de proteínas como o Modelo LPE, Modelo CGE, Modelo PH, Modelo HP-hélico, Modelo HP SCM, Modelo HP TSSCM.

CAPÍTULO 2

Sistema Imunológico - Antígeno – Epítomos – T-Cell

Mente e saúde . . . O homem é o único ser vivo capaz de ser seu próprio agente agressor, através de seus pensamentos, angústias e inseguranças diante da vida.

(Bonamin, 1994)

2.1 Sistema Imunológico

O Sistema Imunológico natural é um mecanismo complexo responsável pela defesa do organismo contra o ataque de microorganismos externos que podem causar doenças (bactérias, fungos, vírus, entre outros), chamados de antígenos (BARROZO, YANG, 2006).

As células responsáveis pela imunidade são os linfócitos (Figura 10) e os fagócitos (Figura 11). Os linfócitos podem apresentar-se como linfócitos T ou linfócitos B (estes são responsáveis pela produção de anticorpos¹). As células T citotóxicas (CD8) destroem células infectadas por vírus, e os linfócitos T auxiliares (CD4) coordenam as respostas imunes, as defesas internas; em contra partida, existem, também, defesas externas. As defesas externas são a primeira barreira contra muitos organismos agressores. No entanto, muitos

¹ Anticorpos são proteínas produzidas por Linfócitos B.

agressores conseguem penetrar, ativando, assim, as defesas internas do organismo. (FARMER, PACKARD, PERELSON, 1986).

Os linfócitos B e linfócitos T tem a capacidade de produzir uma resposta específica para um dado antígeno. Os linfócitos possuem uma memória imunológica, pois reconhecem o reaparecimento de antígeno e evitam uma nova contaminação pela doença, aperfeiçoando-se a cada encontro com um invasor ou similar (CASTRO, ZUBEN, 2000).

O sistema imune pode sofrer um desequilíbrio que se apresenta como imunodeficiência, hipersensibilidade ou doença auto-imune; portanto, antígenos são quaisquer moléculas que possam ser reconhecidas pelo sistema imune adaptativo. O reconhecimento do antígeno é a base principal de todas as respostas imunes adaptativas (BARROZO, YANG, 2006).

O ponto essencial a ser considerado com relação ao antígeno é que a estrutura é a força iniciadora e condutora de todas as respostas imunes. O sistema imune evoluiu com a finalidade de reconhecer os antígenos e destruir e eliminar a sua fonte (CASTRO, ZUBEN, 2001). Quando o antígeno é eliminado, o sistema imune é desligado. Possui apenas regiões que podem ser reconhecidas por um anticorpo (Figura 12); essas regiões são chamadas de *epítome*, sendo que um anticorpo possui uma região de reconhecimento de antígeno (paratopo²) e também uma região que pode ser reconhecida por outro anticorpo (idiotopo) (SCOTT, SMITH, 2009).

² Paratopo é a parte do anticorpo que reconhece corpo estranho.

Fonte: <http://visualsonline.cancer.gov/details.cfm?imageid=1944>

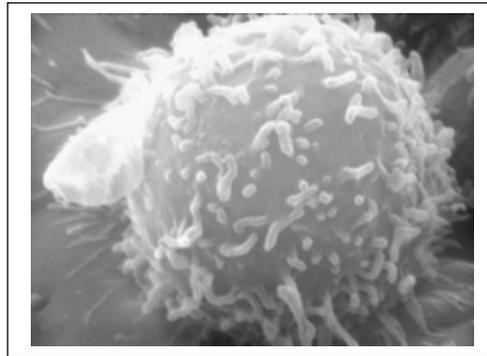


Figura 10 - Uma imagem de um linfócito feita a partir de um microscópio eletrônico

Fonte: www.edu.xunta.es/.../salud/debuxos/fagocito.gif

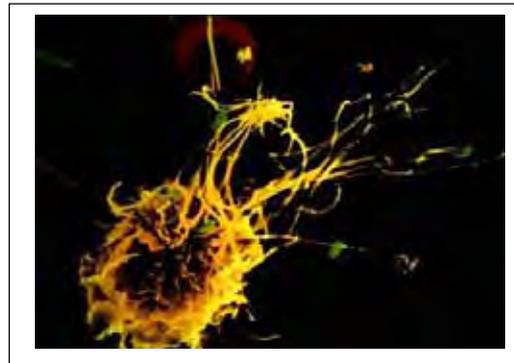


Figura 11 - Uma imagem de um fagócito

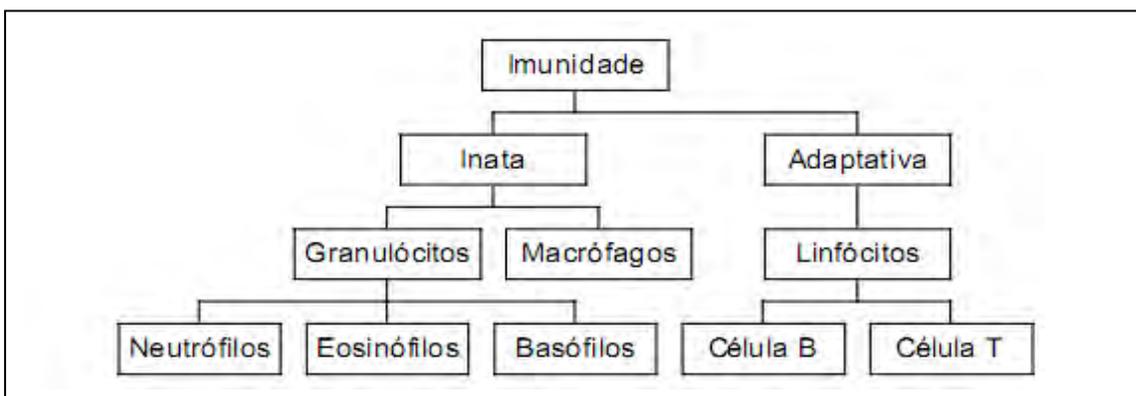


Figura 12 - Mecanismos de defesa e seus principais mediadores.

2.2 Antígeno

Substâncias que podem ser reconhecidas pelas células T, células B por meio de receptores, representados por anticorpo ou TCR (receptor de célula T) particular (DREHER, 1995).

- Antígeno completo ou imunógeno: é capaz de ativar uma resposta imune.
- Antígeno incompleto: não é capaz de ativar uma resposta imune.

2.3 Epítomos

Conhecidos como determinantes antigênicos, os epítomos são aglomerados de antígeno que reúnem aspectos físicos e químicos que favorecem o reconhecimento a regiões específicas dos anticorpos ou TCR's. Uma única molécula antigênica normalmente possui vários epítomos diferentes (GERMAM, 1994,1995).

Epítomos lineares são aqueles formados por resíduos dispostos seqüencialmente, de maneira linear num antígeno protéico ou polissacarídico. Não são afetados por nenhum tratamento que altere a estrutura tridimensional da substância (TONEGAWA, 1983,1985)

Epítomos conformacionais são aqueles formados pelas estruturas secundária, terciária ou quaternária de uma proteína, ou pelo dobramento tridimensional normal de um polissacarídeo. Eles perdem suas funções de epítomos se desnaturados (PERELSON, WEISHBUCH, 1997).

2.3.1 Epítomos de célula T

Os TCR só se ligam a epítomos que formam complexos com moléculas de MHC (Complexo Principal de Histocompatibilidade) (GERMAIN, 1994, 1995) podendo, dessa forma, sofrer processamento e, nesse mecanismo, ser apresentados por uma APC. Para ser imunogênica, uma molécula deve ter, pelo menos, um epítomo de célula T. Resíduos imunodominantes são aqueles que, dentro de determinado epítomo, interagem com maior afinidade de ligação e que podem induzir uma resposta mais forte (TIZARD, 1995)

2.3.2 Antígeno T independente

São aqueles que possuem a capacidade de estimular células B a produzirem anticorpos sem a necessidade da ativação da célula T auxiliar, que normalmente dá o segundo sinal para a deflagração da resposta imune (o primeiro sinal é dado pelo antígeno). Em geral, são polímeros com numerosos determinantes antigênicos repetidos e não produzem memória imunológica (MANSON, 1998).

2.3.3 Interação antígeno-anticorpo

Os antígenos possuem estruturas químicas que favorecem a complementaridade com o anticorpo, por meio de ligações não-covalentes. Essas interações são semelhantes ao que acontece com reações envolvendo enzimas; Portanto, são reversíveis e possuem afinidades diferentes com diversas substâncias (DREHER, 1995).

Como um anticorpo pode se relacionar com antígenos com afinidades diversas, ele pode ligar-se com um que não seja o seu antígeno de melhor

complementaridade por meio de ligações mais fracas com regiões semelhantes, mas não idênticas, àquele que o induziu. Esta ligação é chamada de reação cruzada.

CAPÍTULO 3

Modelagem Computacional – Algoritmo

As aplicações práticas de uma ciência geralmente antecedem o desenvolvimento da própria ciência. (Jerne, 1985).

3.1 Modelagem Computacional

A Modelagem Computacional é um artifício utilizado para representar uma situação real, isto é, necessidade de uma melhor compreensão da realidade (BANKES, 2009).

A Teoria da Complexidade Computacional possui tipologia dos problemas em algoritmos, que podem ser classificados como: intratáveis e tratáveis (MERTENS, 2000).

Problemas intratáveis, são aqueles em que não existe algoritmo de complexidade polinomial que o resolva, contudo, problemas tratáveis podem ser resolvidos. De acordo com MERTENS (2000), os problemas intratáveis poderiam levar séculos para computar dados.

A matemática fornece uma melhor compreensão da essência dos problemas de otimização, porém eles representam os objetivos que se pretende alcançar com a resolução do problema (MANSON, 2001).

O problema de otimização combinatória (PCV), percebe-se que o uso da enumeração exata é totalmente inviável devido à explosão computacional

gerada. (REEVEST, 1995), (MARCON, CALAZANS, MORAES, RIES, HESSEL, 2002) .

O PCV é um exemplo clássico que não apresenta, até o presente momento, algoritmo de complexidade polinomial. O aumento significativo do tempo quando há um crescimento de uma unidade é chamado de explosão computacional (MERTENS, 2000). Desta forma, foram surgindo outros métodos como, por exemplo, as heurísticas, para solução de problemas clássicos considerados intratáveis como, por exemplo, problema de coloração de grafos (AGUIAR, HONORATO, SANTOS, OCHI, 2005), quadrático de alocação (NETO, 2007) e outros. Os algoritmos heurísticos buscam o equilíbrio entre o desempenho computacional e a qualidade da solução (AGUIAR, HONORATO, SANTOS, OCHI, 2005). A variação do algoritmo heurístico é a técnica denominada de metaheurísticas, não existe uma formalização literária para este termo, contudo torna-se cada vez mais aceito no que se refere a uma estratégia de algoritmos heurísticos (AGUIAR, HONORATO, SANTOS, OCHI, 2005). Metaheurísticas caracterizam-se por redes neurais artificiais e algoritmos genéticos (WHITLEY, 1994). Os problemas de seqüenciamento são muito complexos e com resolução difícil chamado NP-Completo, (ADAMI, 2002), (CAI, CHEN, 1997), sendo aquele em que se tem um conjunto de tarefas que serão executadas em processadores obedecendo a certas restrições com o objetivo de otimizar algum critério, (AARONSON, 2005). A técnica de enumeração exata, portanto, utiliza os métodos aproximativos para otimizar e solucionar os problemas de seqüenciamento.

3.2 A Técnica de Enumeração Exata

O ponto central da enumeração exata é que devem ser contadas uma única vez todas as configurações do sistema (LAVOR, LIBERTI, MACULAN, 2008). Esta condição impõe limitações quanto ao tamanho máximo do sistema que pode ser estudado por essa técnica, uma vez que o número de configurações do sistema cresce muito rápido com o aumento do tamanho do mesmo (DILL, BROMBERG, YUE, FIEBIG, YEE, THOMAS, CHAN, 1995).

Para tornar clara a técnica e, ao mesmo tempo, apresentar como foi implementada no nosso caso, a questão mais difícil de resolver, ao fazer uma enumeração exata, é garantir que todas as configurações do sistema sejam geradas uma única vez (CALDARELLI, RIOS, 2007). O algoritmo que utilizamos para isso depende intimamente do método com que descrevemos a configuração do polímero na rede (WROE, BAUER, CHANZ, 2005).

O processo de construção da solução consiste em elaborar modelos conformacionais (situações de enovelamento) de acordo com a leitura de cadeias formadas sob o Alfabeto $\{H, P\}^+$. No caso do modelo 2D-HP, possuímos três graus de liberdade (Figura 13) de movimento: seguir em frente (S), dobrar `a esquerda (L) ou `a direita (R) (WROE, BAUER, CHANZ, 2005).

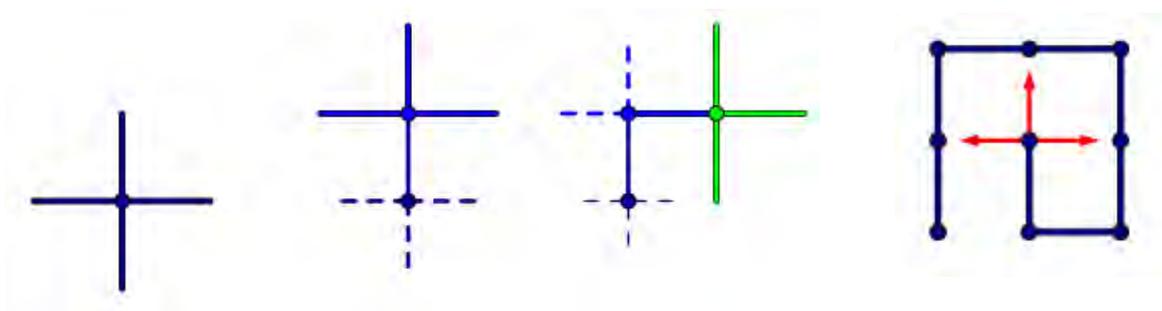


Figura 13 – Representação dos três graus de liberdade.

No trabalho atual, executamos uma simulação computacional, com um modelo simples, na linguagem de programação Fortran 90, para a enumeração exata de todas as conformações possíveis em uma rede quadrada para $n = 1$ a 17 monômeros. Os resultados prévios mostram que existem 2.155.667 conformações diferentes possíveis (CALDARELLI, RIOS, 2007).

Enquanto a cadeia de entrada, constrói a solução pelo espaço de busca matricial, a memória associa para cada componente da cadeia, uma alocação de entrada e uma posição da cadeia de movimentação. Esta cadeia de movimentação é formada pelo alfabeto $\{S,L,R\}^+$. A determinação do movimento é essencial para que possa definir a vizinhança dos componentes da cadeia de solução. A definição da vizinhança é utilizada para definir o número de contatos não-adjacentes entre os aminoácidos hidrofóbicos, para que seja possível calcular o valor da energia livre.

Quanto mais contatos $H-H$ adjacentes houver, menor o valor da energia livre final (WROE, BAUER, CHANZ, 2005).

3.3 Algoritmo Self Avoid Walk

“*Self Avoid Walk*” também chamado de caminho aleatório é um dos exemplos mais simples e mais clássicos na teoria de probabilidades (SPITZER, 2001). Contudo, apesar de sua simplicidade, ele oferece comportamento extremamente rico, e serve como um exemplo de paradigma para ilustração da maioria das leis probabilísticas (WERNER, 2005). Ao mesmo tempo, várias modificações de caminho aleatório, tais como caminho aleatório auto-excludente (*self-avoiding random walk*) e caminho aleatório com voltas apagadas (*loop erased random walk*) ainda são pouco entendidos (SCHRAMM, 2000), (ALBERTS, SHEFFIELD, 2007). Os estudos desses caminhos levaram a um maior progresso na teoria de probabilidade, em particular, o estudo do limite de escala de caminho aleatório com voltas apagadas levou Oded Shramm a descoberta da Evolução Estocástica de Loewner (SLE), que interligou a probabilidade com análise, geometria e sistemas dinâmicos (BAUER, BERNARD, 2003).

O *self-avoiding walk* é um caminho a partir de um ponto fixo que nunca se intercepta. Esses caminhos (Figura 15), em uma rede quadrada nunca passa duas vezes no mesmo ponto.

Este algoritmo apresenta problema rotacional e translacional (Figura16), conhecido como sobreposição, para resolver-lo, implementou no código um critério para que isto não ocorra.

O primeiro algoritmo adotado, no presente trabalho, é o caminho aleatório auto-excludente (DETHRIDGE, GUTTMANN, 2008),(WILLIAMS, 1996), fornecido pelo Prof. Dr. Caliri, A. (USP-Ribeirão Preto) e adaptado pelo grupo de simulação molecular (UNESP-Bauru). O segundo algoritmo para

encontrar os vizinhos topológicos e verificar sua frequência foi desenvolvido para este trabalho em específico.

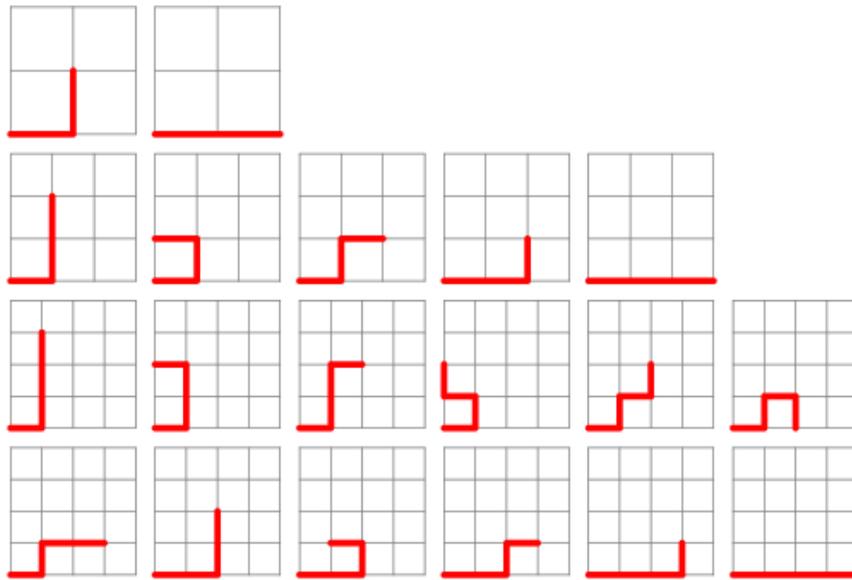


Figura 14 - Diversas formas de crescimento no algoritmo, sempre o ponto (0,0) estará fixo.

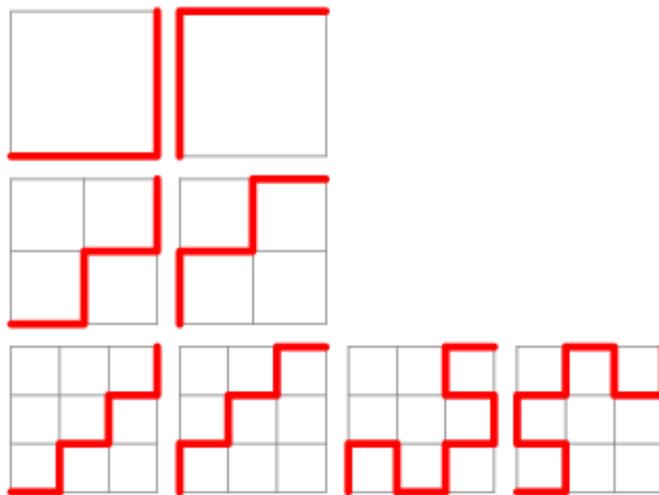


Figura 15 - Configurações com problemas de rotação e simetria.

3.3.1 O algoritmo aleatório auto-excludente:

Sítios já visitados não são mais ocupados. Essa correlação introduz a idéia de volume excluído ao sistema. Em uma rede bidimensional o quadrado possui 8 simetrias (rotações por 0, $\pi/2$, π , $3\pi/2$ e as combinações de cada um destes quatro com uma reflexão sobre o eixo-x).

3.3.2 Implementação:

A sintaxe é elaborada para obter todas as configurações possíveis de uma cadeia com L unidades, os quais ocupam sítios de uma rede quadrada.

As propriedades gerais das configurações consistem em unidades da cadeia, que por sua vez, devem ocupar sucessivos e distintos sítios da rede.

Assim, em uma configuração válida a cadeia passa em cada um dos L sítios uma única vez. Uma configuração que difere de outra apenas por uma reflexão ou rotação (ângulo de 0, $\pi/2$, π , $3\pi/2$), são equivalentes, portanto, são contadas uma única vez. A estratégia é a enumeração sucessiva, ou seja, uma vez conhecida todas as configurações para a cadeia de tamanho L-1 (cadeia contendo L-1 unidades), parte-se delas para gerar as configurações de cadeias de tamanho L.

Limite superior (número máximo de configurações devido ao crescimento da cadeia por uma unidade):

Se cadeias de tamanho L-1 possuem N' configurações distintas, então o número N de configurações distintas para cadeias de tamanho L satisfaz:

$$N \leq 4 N'$$

Visto que partindo do final da cadeia, temos quatro possíveis sítios para localizá-lo a nova unidade da cadeia (devido ao crescimento da cadeia de L-1 para L unidades).

Contudo, como um mesmo sítio deve ser ocupado por uma e somente uma unidade da cadeia, uma das quatro possibilidades, como comentada acima, fica automaticamente proibida. Portanto, podemos escrever:

$$N \leq 3 N'$$

3.3.3 Considerações preliminares sobre algoritmo:

Cadeia com L = 1 - A unidade 1 no ponto (0,0) da rede

$$\# \text{ configurações}^3 = 1$$

Cadeia com L = 2

$$\text{Limite superior} = 3$$

configurações = 1: seguimento de linha ligando os pontos (0,0)+(1,0).

Note que 4 possíveis configurações, obtidas ligando-se o ponto (0,0), respectivamente os pontos (1,0); (0,1); (-1,0), e (0,-1), são todas equivalentes, pois 3 delas são obtidos da rotação, digamos, da configuração definida pelos pontos (0,0) e (1,0); (ligando o ponto (0,0) ao ponto (1,0) da rede quadrada).

Cadeia com L = 3

$$\text{Limite superior: } N = 3N', \text{ isto é } N=1 \times 3 = 3$$

$$\# \text{ configurações} = 2 [(0,0)+(1,0)+(1,0); (0,0)+(1,0)+(0,1)]$$

³ Nota: o símbolo # representa aqui a palavra "numero".

Note que a configuração $(0,0)+(1,0)+(0,-1)$ é equivalente à $(0,0)+(1,0)+(0,1)$, pois a primeira pode ser obtida do segundo por reflexão do eixo-y.

Cadeia com $L = 4$

Limite superior. $N = 3N'$, isto é $N=3 \times 2 = 6$

configurações = 5

[$(0,0)+(1,0)+(1,0)+(1,0)$; $(0,0)+(1,0)+(1,0) + (0,1)$];

[$(0,0)+(1,0)+(0,1)+/(1,0)$; $(0,0)+(1,0)+(0,1)+/(0,1)$;

$(0,0)+(1,0)+(0,1)+ (0,-1)$]

As 5 configurações acima, para cadeia de tamanho $L= 4$, formam a base para a obtenção de todas configurações distintas quanto a rotações e reflexões, para cadeias de tamanho $L = 5, 6...17$.

A partir de $L=4$, será incluída uma condição para não haver a sobreposição e assim gerar as configurações automaticamente.

Para armazenar as configurações para as cadeias de tamanho L . As seguintes matrizes serão necessárias: $X(i,j)$ e $Y(i,j)$ onde "i" é o índice da /i-ésima/ configuração da cadeia de tamanho L , e "j" é o índice da /j-ésima/ unidade da cadeia. Assim, o par de coordenadas ($X(i,j)$; $Y(i,j)$) identifica no espaço a /j-ésima/ unidade da /i-ésima/ configuração. Lembrando que o número de configurações $N(L)$ de cadeias distintas de tamanho L satisfaz

$$N(L) \leq 3N(L-1).$$

A sobreposição da cadeia ocorre, quando:

$X(i,j) = X(k,l)$ ao mesmo tempo que

$Y(i,j) = Y(k,l)$, para “ k ” diferente de “ j ”

Assim, quando a nova unidade é adicionada na cadeia (no processo da construção da cadeia de L unidades), verifica se há superposição para todas as unidades anteriores, isto é para unidade 1, 2, ..., até “ $L-1$ ” (Figura 16).

Obtemos os resultados de todas as configurações possíveis independentes de rotação e reflexão para a cadeia até 17 monômeros, resultando em $N=2155667$ configurações (RIOS, L. P.; CALDARELLI, G, 1996).

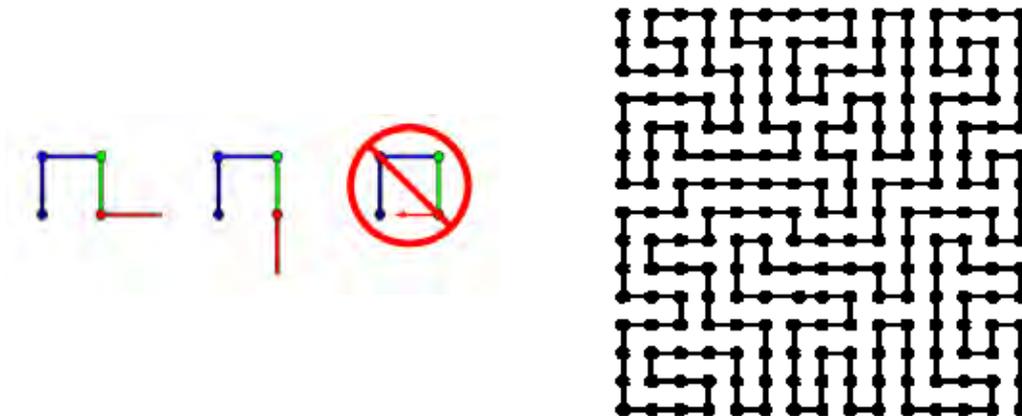


Figura 16- Crescimento de uma cadeia do Algoritmo: Na ilustração à esquerda, das três possibilidades para a adição de uma unidade extra, somente duas delas (as duas mais à esquerda) satisfazem a condição de configuração *Self Avoid Walk*.

3.4 Algoritmo Vizinho Topológicos

Este algoritmo lê o arquivo gerado no algoritmo anterior e procura o seu vizinho topológico, utilizando tabelas. Ao reservar o espaço destinado as tabelas, com alocação dinâmica, podem controlar o espaço utilizado, evitando consumo excessivo de memória, para encadeamentos. Após os dados serem lidos, efetua a análise.

O procedimento utilizado: Dado um monômero, verifica se este possui outros elementos a sua volta. Caso possua, verifica se estes outros monômeros são pertencentes à cadeia, e em caso negativo, então se trata de um vizinho topológico. Assim, os vizinhos topológicos são gravados em com a identificação do seu monômero correspondente.

3.5 Hardware & Software:

No presente trabalho a determinação do algoritmo dependeu do Hardware e Software utilizados para gerar as configurações e rastrear os pares de contatos, bem como manipular os dados que se obteve na simulação.

Verificando que para gerar as configurações é essencial o hardware possuir no mínimo 4 Gb de RAM, principalmente nos cálculos referente ao 17-éssimo monômero.

Hardware utilizado:

Processador Intel Core 2 Quad a Q6700 2.66 GHz;
Memória RAM de 4Gb e 667 MHz;
Disco rígido SATA DE 250 Gb;

Softwares utilizados:

- Sistema Operacional Linux 64 bits, Fedora;
- Linguagem: Fortran 90,
- Aplicativos: Excel utilizando VBA;
- Compilador: gfortran;
- Gerenciador de banco de dados: Oracle Express v.10.0, com SQL.

CAPÍTULO 4

4. Análise dos Resultados

Este capítulo tem por objetivo apresentar os procedimentos de classificação dos dados coletados e os seus resultados. As informações iniciais são referentes às representações como a seqüência do aminoácido de uma proteína determina sua funcionalidade biológica, e como os monômeros são representados para a enumeração exata de todas as conformações possíveis em uma rede quadrada para cadeias de tamanho $L = 1$ a 17 monômeros. Confirmamos estes resultados e atribuímos variáveis aos sítios de todos os monômeros no modelo de proteína, onde apresenta as saídas deste algoritmo e os deslocamentos realizados em um campo espacial com o seu crescimento evolutivo. Após os resultados serem apresentados com as configurações válidas, seguem os resultados das representações gráficas de algumas saídas de contatos topológicos, distribuição da distância em relação a freqüência, mapa de contatos topológicos, energia configuracional em relação aos contatos, contato topológicos de longo alcance e aplicação do modelo HP a seqüências de Epitopos, com suas respectivas energias e a freqüência dos contatos nativos obtidos durante a simulação. Finalmente, será exposto as considerações finais entre o modelo e uma seqüência de um Epítope.

4.1 Resultado do Algoritmo Self Avoid Walk

Considerações preliminares sobre o resultado:

A apresentação de configurações, na próxima página, ilustra somente o caso de uma cadeia de 5 aminoácidos.

A saída, gerada pelo programa codificado em Fortran 90, especifica:

L = número de monômeros

$N(L)$ = número de configurações possíveis.

- Configurações para cadeia que nomeamos por L .
- Para $L = 5$, obtemos $N(L) = 13$ configurações.
- Primeira coluna descreve as coordenadas do x e a segunda coluna as coordenadas do y de cada cadeia.
- Portanto, na cadeia 1, possuímos as coordenadas cartesianas:

0,0

1,0

2,0

3,0

4,0

- Descrevendo uma configuração possível e assim sucessivamente até obtermos para a cadeia $L = 5$, 13 configurações possíveis.
- Segue a representação gráfica ao lado.

4.2 Resultados do Algoritmo Vizinhos Topológicos

Serão apresentadas Figura 17 representações gráficas das interações dos vizinhos topológicos, que, por sua vez, são saídas do algoritmo dado, disposto:

- primeira coluna: o número que identifica a configuração
- segunda coluna: o número correspondente ao monômero
- terceira coluna: posição da coordenada cartesiana do eixo x
- quarta coluna: posição da coordenada cartesiana do eixo y
- quinta coluna: onde encontramos os contatos dos primeiros vizinhos.

Exemplificando a representação abaixo na cadeia 4 o monômero 2, seu contato topológico é o monômero 5 e vice e versa.

Representando graficamente a mesma, verificamos que cada aminoácidos com suas coordenadas cartesianas possuem uma alocação na memória para armazenar seus respectivos números.

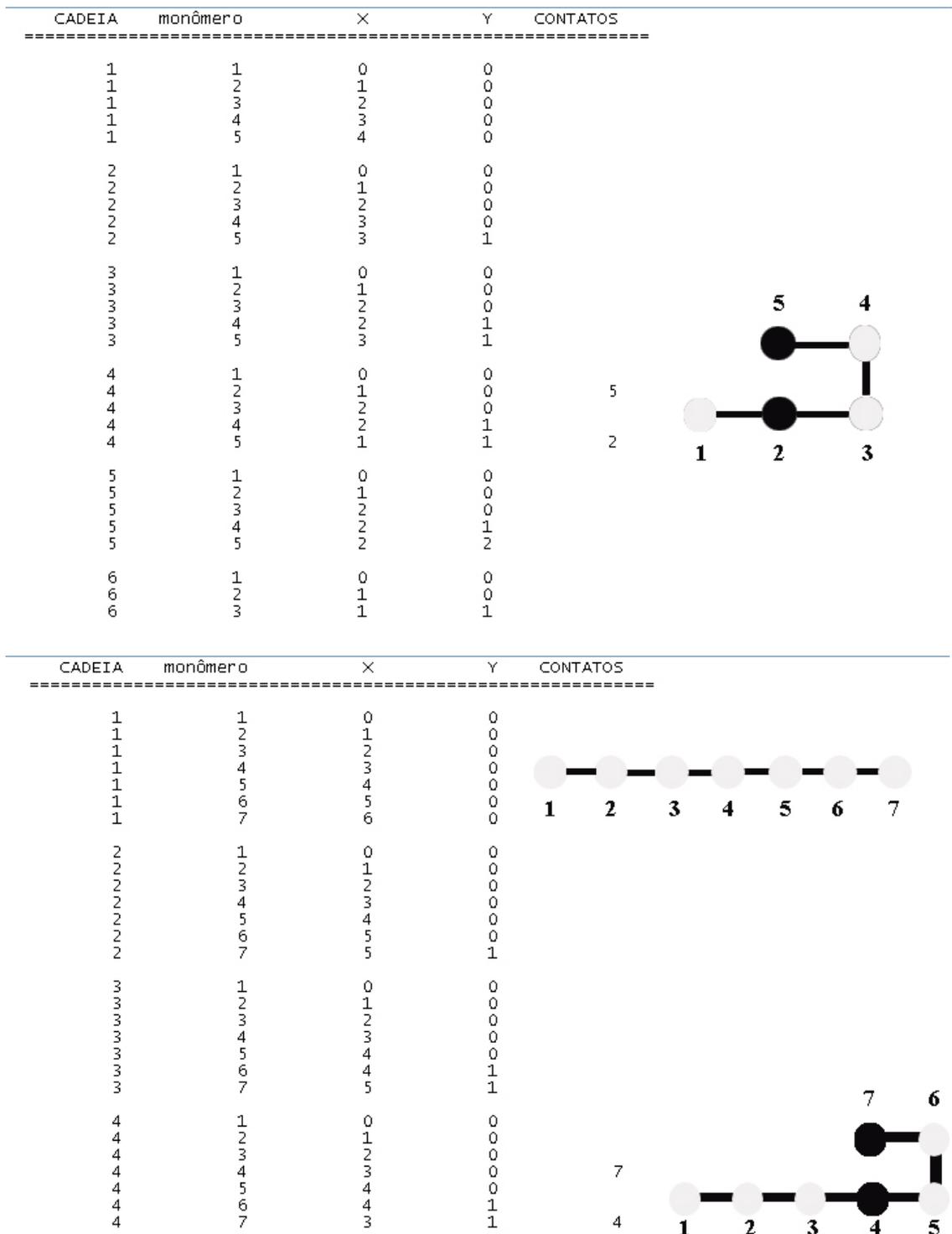


Figura 17 – Saída da implementação do algoritmo para se obter os vizinhos topológicos.

4.3 Distribuição da energia segundo parâmetros topológicos

A representação abaixo Figura 18 demonstra como obter a energia livre do sistema em relação aos contatos topológicos.

Assim, para algumas configurações da cadeia 16, onde a distribuição das seqüências corresponde ao modelo HP, apresenta o exemplo de evolução da energia para a configuração nativa (enovelamento).

Cadeia	Sequência	Energia
16	phhphpphphpphph	-1
16	phhphpphphpphph	-2
16	phhphpphphpphph	-3
16	phhphpphphpphph	-4
16	phhphpphphpphph	-5

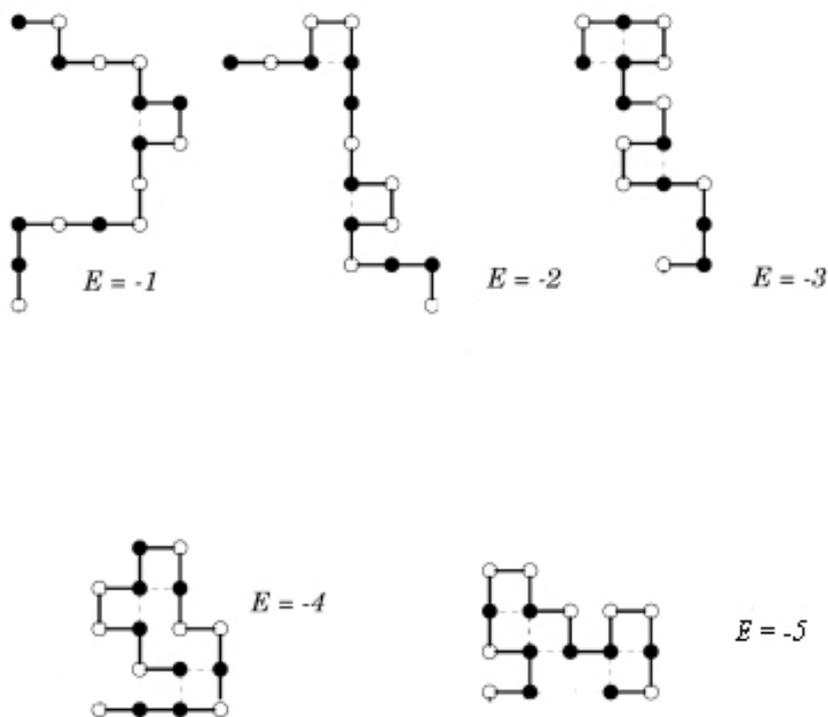


Figura 18 – Energia configuracional total do sistema.

4.4 Distribuição da distância entre os contatos topológicos

Na Figura 19, apresentamos a distribuição (frequência) de distâncias ao longo da cadeia em que os contatos topológicos ocorrem. Para isso são consideradas o conjunto de todas as configurações. Distâncias ao longo da cadeia no intervalo de 4 a 17 monômeros são consideradas.

Poderemos verificar que o número total de contatos (NTC) cresce com o número de monômeros, onde confirmamos o número obtido em outros estudos para o número de conformações possíveis para $N=17$ monômeros (temos 2.155.667 conformações).

Apresentamos, também, o número total de contatos para vizinhos topológicos (NTCVT). Para $N=17$ temos 4892837 possíveis contatos topológicos.

A distribuição das frequências das distâncias foi representada por $[a,b]^c$, onde a e b referem-se aos monômeros ao longo da cadeia e o índice c , a frequência com que esta distância aparece. Por exemplo, no caso $[1,4]^2$, esta distância aparece duas vezes.

Cadeia	NTC	NTCVT	Distribuição total da Distância em relação a Frequência						
4	1	1	[1,4] ¹						
5	13	4	[1,4] ²	[2,5] ²					
6	36	18	[1,4] ⁶	[1,6] ²	[2,5] ⁴	[3,6] ⁶			
7	98	59	[1,4] ¹⁶	[1,6] ³	[2,5] ¹⁰	[2,7] ⁵	[3,6] ¹¹	[4,7] ¹⁴	
8	272	225	[1,4] ³⁸	[1,6] ¹⁵	[1,8] ¹⁴	[2,5] ³⁰	[2,7] ⁶	[3,6] ³¹	
9	740	740	[1,4] ¹²⁴	[1,6] ³⁹	[1,8] ²²	[2,5] ⁸³	[2,7] ²¹	[2,9] ²²	
10	2034	2433	[1,4] ³⁴⁶	[1,6] ¹¹¹	[1,8] ⁶¹	[1,10] ⁷⁰	[2,5] ²²⁷	[2,7] ⁵⁶	
11	5513	7371	[1,4] ⁹³⁴	[1,6] ²⁹⁹	[1,8] ¹⁴⁵	[1,10] ⁹⁸	[2,5] ⁶²¹	[2,7] ¹⁵³	
12	15037	21393	[1,4] ²¹¹⁴	[1,6] ⁷⁷⁶	[1,8] ³⁸⁷	[1,10] ²²⁸	[1,12] ³²¹	[2,5] ¹⁶⁹¹	
13	40617	72116	[1,4] ⁷³⁷⁶	[1,6] ²³⁷¹	[1,8] ¹¹⁴⁹	[1,10] ⁶⁸⁷	[1,12] ⁵³³	[2,5] ⁴⁵⁹⁴	
14	110188	203978	[1,4] ¹⁸⁸⁴⁸	[1,6] ⁶⁰⁹⁵	[1,8] ³⁰²¹	[1,10] ¹⁸⁰⁹	[1,12] ¹³¹⁰	[1,14] ¹⁹²²	
15	296806	597253	[1,4] ⁵⁰⁷⁸²	[1,6] ¹⁶⁴³¹	[1,8] ⁸⁹⁰⁹	[1,10] ⁵⁴¹³	[1,12] ³⁵⁷¹	[1,14] ³²³⁹	
16	802075	1496559	[1,4] ¹¹³⁰⁷	[1,6] ³⁸⁵⁷⁴	[1,8] ²¹⁵¹⁶	[1,10] ¹²⁷⁵⁵	[1,12] ⁸¹⁶⁶	[1,14] ⁶²⁶⁰	
17	2155667	4892837	[1,4] ⁷⁷⁹⁴¹⁹	[1,6] ⁷¹¹⁶²	[1,8] ⁵³²⁶³	[1,10] ²⁷⁵⁹³	[1,12] ¹⁷²⁸⁶	[1,14] ¹¹⁶³⁵	

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225	[3,8]	16	[4,7]	30	[5,8]	45						
9	740	740	[3,6]	81	[3,8]	21	[4,7]	81	[4,9]	39	[5,8]	83	[6,9]	124
10	2034	2433	[2,9]	33	[3,6]	228	[3,8]	57	[3,10]	61	[4,7]	215	[4,9]	56
11	5513	7371	[2,9]	79	[2,11]	109	[3,6]	609	[3,8]	144	[3,10]	79	[4,7]	591
12	15037	21393	[2,7]	314	[2,9]	167	[2,11]	138	[3,6]	1601	[3,8]	379	[3,10]	176
13	40617	72116	[2,7]	1240	[2,9]	620	[2,11]	395	[2,13]	582	[3,6]	4601	[3,8]	1125
14	110188	203978	[2,5]	12462	[2,7]	3095	[2,9]	1564	[2,11]	962	[2,13]	901	[3,6]	12358
15	296806	597253	[2,5]	33675	[2,7]	8355	[2,9]	4218	[2,11]	2593	[2,13]	1906	[2,15]	3239
16	802075	1496559	[1,16]	10338	[2,5]	91057	[2,7]	16983	[2,9]	8776	[2,11]	5585	[2,13]	3897
17	2155667	4892837	[1,16]	10000	[2,5]	3E+05	[2,7]	145451	[2,9]	42358	[2,11]	26612	[2,13]	19161

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433	[5,8]	228	[5,10]	111	[6,9]	227	[7,10]	346				
11	5513	7371	[4,9]	144	[4,11]	159	[5,8]	591	[5,10]	153	[6,9]	609	[6,11]	299
12	15037	21393	[3,12]	281	[4,7]	1562	[4,9]	340	[4,11]	201	[5,8]	1567	[5,10]	353
13	40617	72116	[3,10]	506	[3,12]	383	[4,7]	4393	[4,9]	1046	[4,11]	493	[4,13]	722
14	110188	203978	[3,8]	2926	[3,10]	1312	[3,12]	795	[3,14]	1427	[4,7]	11715	[4,9]	2681
15	296806	597253	[3,6]	33218	[3,8]	7847	[3,10]	3481	[3,12]	1954	[3,14]	1906	[4,7]	31749
16	802075	1496559	[2,15]	4348	[3,6]	85946	[3,8]	18394	[3,10]	8194	[3,12]	4666	[3,14]	3390
17	2155667	4892837	[2,15]	13552	[2,17]	13552	[3,6]	107654	[3,8]	63574	[3,10]	43994	[3,12]	23643

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371	[7,10]	621	[8,11]	934								
12	15037	21393	[5,12]	409	[6,9]	1529	[6,11]	381	[7,10]	1599	[7,12]	758	[8,11]	1634
13	40617	72116	[5,8]	4565	[5,10]	1067	[5,12]	621	[6,9]	4452	[6,11]	1122	[6,13]	1203
14	110188	203978	[4,11]	1217	[4,13]	962	[5,8]	11955	[5,10]	2730	[5,12]	1312	[5,14]	2593
15	296806	597253	[4,9]	7229	[4,11]	3246	[4,13]	1954	[4,15]	3571	[5,8]	31972	[5,10]	7208
16	802075	1496559	[3,16]	7219	[4,7]	82740	[4,9]	16988	[4,11]	7581	[4,13]	7581	[4,15]	4307
17	2155667	4892837	[3,14]	17565	[3,16]	13669	[4,7]	191709	[4,9]	69169	[4,11]	23091	[4,13]	15686

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371												
12	15037	21393	[9,12]	2487										
13	40617	72116	[7,10]	6588	[7,12]	1234	[8,11]	4592	[8,13]	2231	[9,12]	4701	[10,13]	6924
14	110188	203978	[6,9]	11641	[6,11]	2681	[6,13]	1564	[7,10]	11955	[7,12]	2926	[7,14]	3241
15	296806	597253	[5,12]	3246	[5,14]	2593	[6,9]	31632	[6,11]	7208	[6,13]	3481	[6,15]	5410
16	802075	1496559	[5,8]	82277	[5,10]	16310	[5,12]	7311	[5,14]	4553	[5,16]	8987	[6,9]	79803
17	2155667	4892837	[4,15]	10427	[4,17]	13928	[5,8]	241571	[5,10]	49412	[5,12]	24112	[5,14]	13012

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371												
12	15037	21393												
13	40617	72116												
14	110188	203978	[8,11]	11715	[8,13]	3095	[9,12]	12358	[9,14]	6095	[10,13]	12462	[11,14]	18848
15	296806	597253	[7,10]	31632	[7,12]	7229	[7,14]	4218	[8,11]	31972	[8,13]	7847	[8,15]	8909
16	802075	1496559	[6,11]	16118	[6,13]	7348	[6,15]	6354	[7,10]	81160	[7,12]	16362	[7,14]	7905
17	2155667	4892837	[5,16]	10390	[6,9]	2E+05	[6,11]	60143	[6,13]	28422	[6,15]	15469	[6,17]	17799

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371												
12	15037	21393												
13	40617	72116												
14	110188	203978												
15	296806	597253	[9,12]	31749	[9,14]	8355	[10,13]	33218	[10,15]	16431	[11,14]	33675	[12,15]	50782
16	802075	1496559	[7,16]	13322	[8,11]	79043	[8,13]	16258	[8,15]	9992	[9,12]	80757	[9,14]	17555
17	2155667	4892837	[7,10]	217628	[7,12]	57911	[7,14]	28431	[7,16]	17969	[8,11]	214272	[8,13]	59139

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371												
12	15037	21393												
13	40617	72116												
14	110188	203978												
15	296806	597253												
16	802075	1496559	[9,16]	21352	[10,13]	79489	[10,15]	19400	[11,14]	82484	[11,16]	38744	[12,15]	86569
17	2155667	4892837	[8,15]	28373	[8,17]	26807	[9,12]	215390	[9,14]	59665	[9,16]	32073	[10,13]	216968

Cadeia	NTC	NTCVT												
4	1	1												
5	13	4												
6	36	18												
7	98	59												
8	272	225												
9	740	740												
10	2034	2433												
11	5513	7371												
12	15037	21393												
13	40617	72116												
14	110188	203978												
15	296806	597253												
16	802075	1496559	[13,16]	130538										
17	2155667	4892837	[10,15]	62740	[10,17]	47207	[11,14]	221040	[11,16]	65845	[12,15]	224904		

Cadeia	NTC	NTCVT						
4	1	1						
5	13	4						
6	36	18						
7	98	59						
8	272	225						
9	740	740						
10	2034	2433						
11	5513	7371						
12	15037	21393						
13	40617	72116						
14	110188	203978						
15	296806	597253						
16	802075	1496559						
17	2155667	4892837	[12,17]	74996	[13,16]	241147	[14,17]	278385

Figura 19 – Distribuição total da distância em relação a frequência.

Na Figura 18 apresentamos na forma logarítmica a distribuição das populações para as 2155667 configurações obtidas. Poderemos observar que até 8 monômeros $NTC > NTCVT$; para $N = 9$ $NTC = NTCVT$ e para $N > 9$, $NTC < NTCVT$. Vemos também que tanto NTC como $NTCVT$ seguem uma lei de crescimento exponencial em relação ao número de monômeros na cadeia.

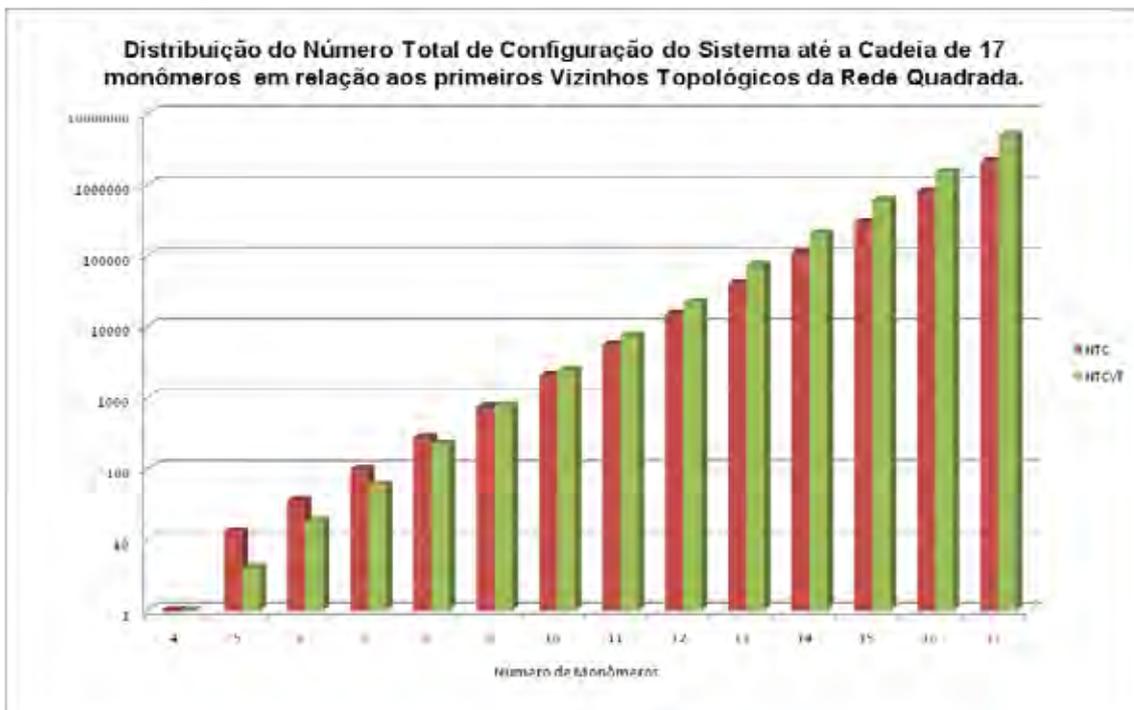


Figura 20 - Distribuição das populações de 2.155.667 configurações possíveis para 17 monômeros, onde utilizamos a nomenclatura de Número Total de Configurações (NTC) em função da ordem de contato, que fornece o número de monômeros existentes entre os pares dos primeiros vizinhos (de monômeros) em contato topológico, chamado de Número Total de Contatos Vizinhos Topológicos (NTCVT).

4.5 *Freqüências dos Vizinhos Topológicos Distância [1,4], em função do número de cadeias.*

Na Figura 21 apresentamos, a freqüência das distâncias (NTCVT), para os monômeros [1,4] em função do número total de monômeros na cadeia. Para $N=4$ até $N=13$, neste gráfico não é possível visualizar qualquer tipo de tendência visto que a freqüência das distâncias é muito baixa. Podemos verificar, no entanto que para $N>13$ ocorre um aumento exponencial na freqüência da distância 3. Na Figura 22 buscamos a visualização da tendência de crescimento da freqüência das distâncias em função do numero de monômeros na cadeia. Podemos ver que em todo o intervalo de número de monômeros na cadeia o crescimento da freqüência das distâncias segue exponencialmente.

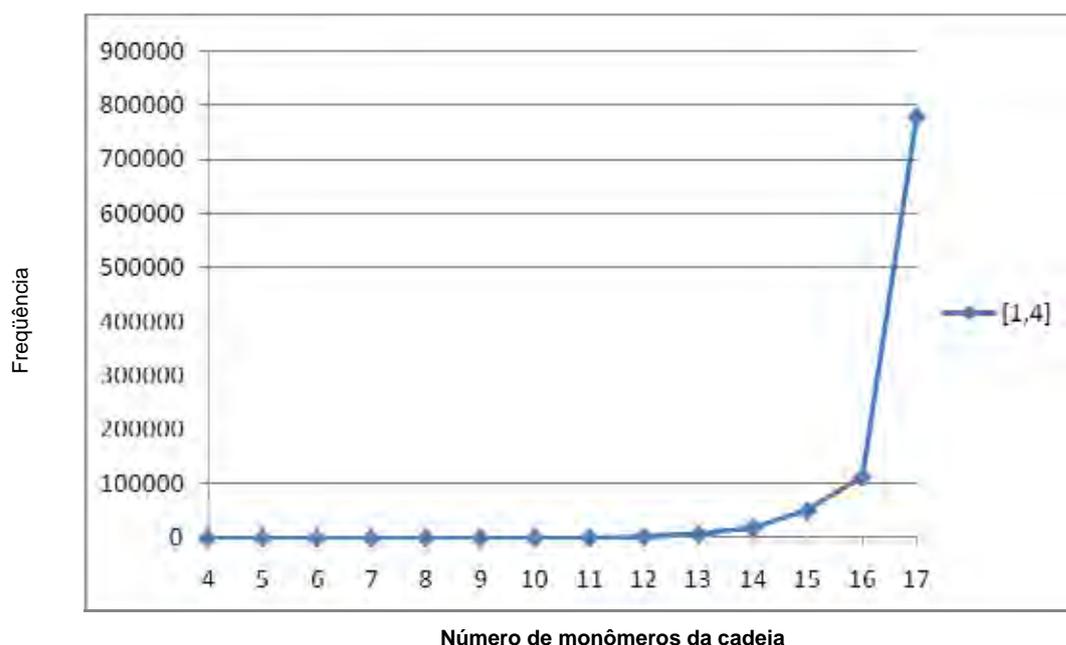
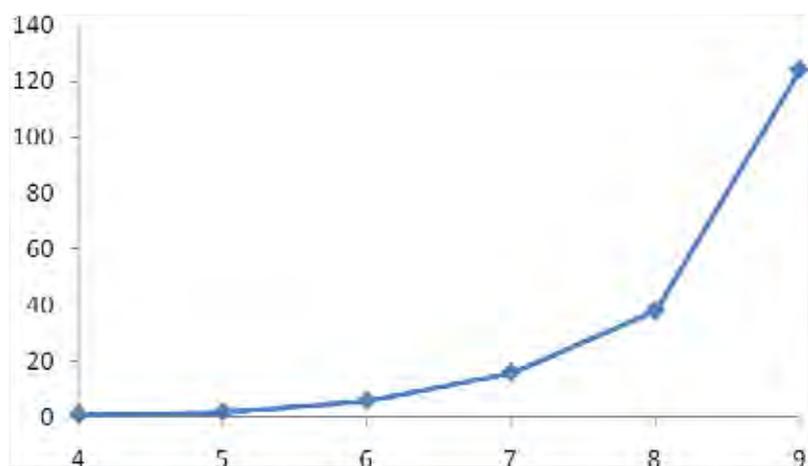
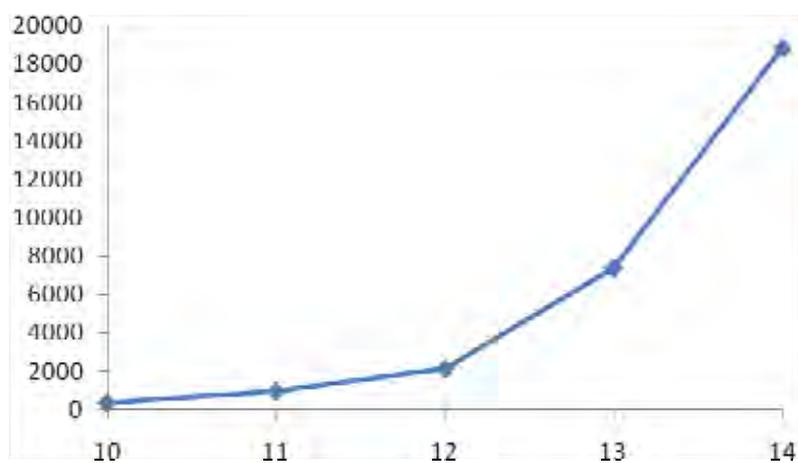


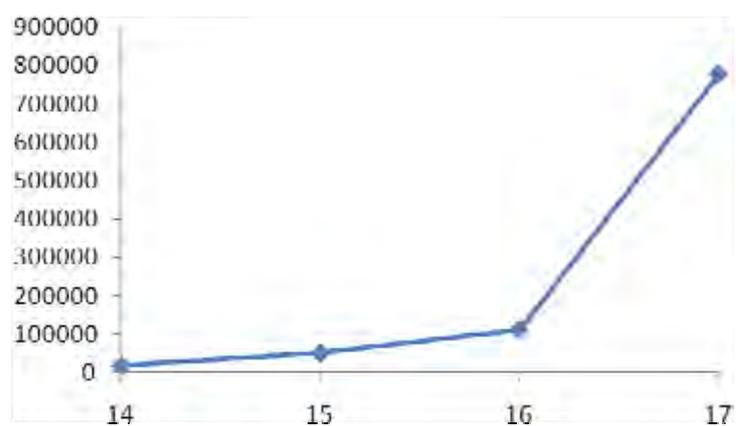
Figura 21 – Gráfico da freqüência da distância 3 em função do número de monômeros na cadeia.



(a)



(b)



(c)

Figura 22 – Representação da distância 3 para os monômeros [1,4] nas cadeias com: (a) 4 a 9 monômeros, (b) 10 a 14 monômeros, (c) 14 a 17 monômeros.

Na Figura 23 apresentamos, em forma logarítmica a distribuição da frequência para distância 3 em função do número de monômeros da cadeia. Podemos visualizar neste gráfico todo o intervalo de número de monômero na cadeia e confirmar o crescimento exponencial da frequência das distâncias.

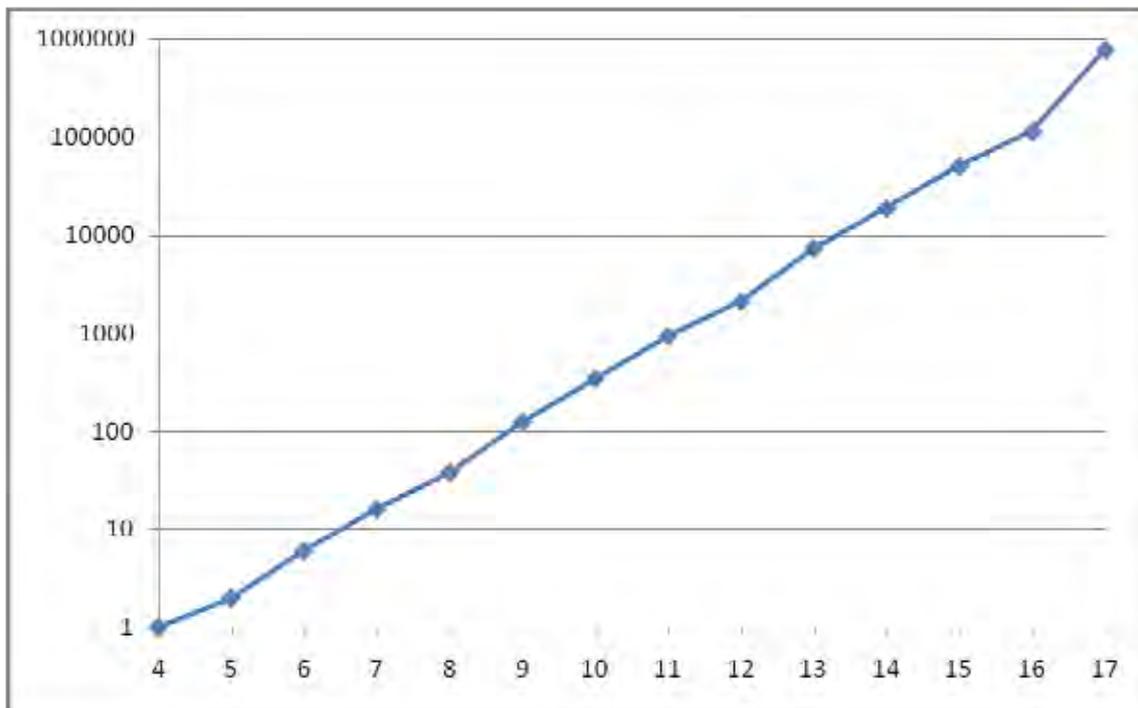


Figura 23 – Gráfico na forma logarítmica da frequência da distância 3 em função do número de monômeros na cadeia.

A Figura 24 ilustra a distribuição total de frequência em relação à cadeia com 16 monômeros. Podemos verificar que as distâncias que apresentam as maiores frequências, são: [1,6], [2,5], [3,6], [4,7], [5,8], [6,9], [6,15], [7,10], [8,11], [9,12], [10,13], [11,14], [12,15] e [13,16].

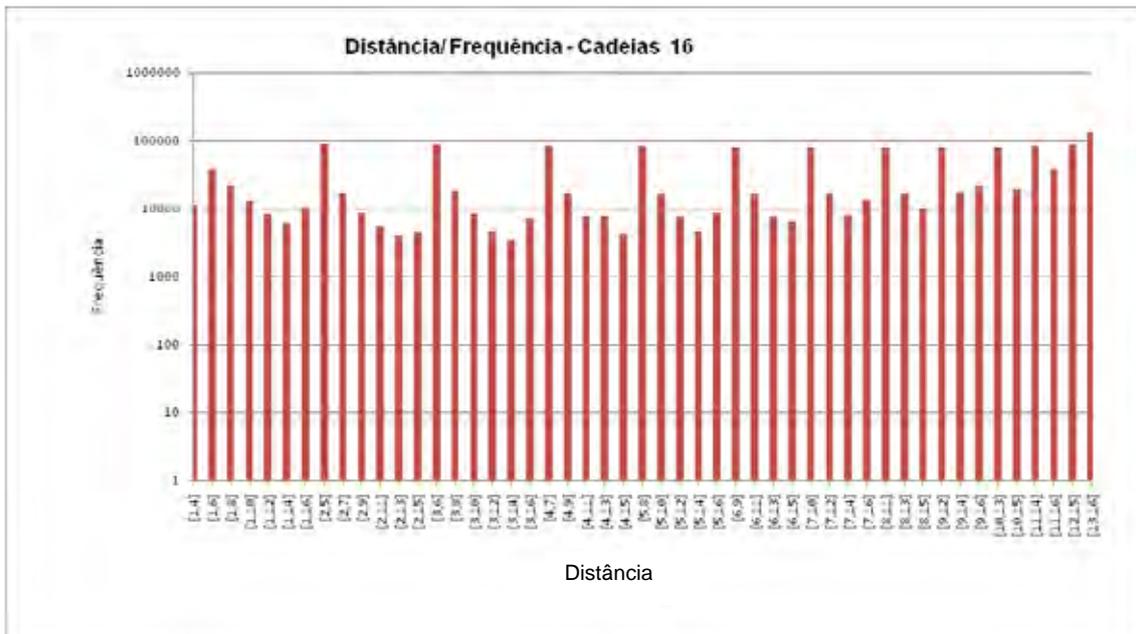


Figura 24 – Distância total em relação à frequência da cadeia de L= 16.

A Figura 25 as frequências para N=17 são analisadas também em relação ao número de monômeros. Podemos ver que as distâncias que apresentam as maiores frequências são:

- [1,4], [2,5], [4,7], [5,8], [6,9], [7,10], [8,11], [9,12], [10,13], [11,14], [12,15], [13,16],[14,17].

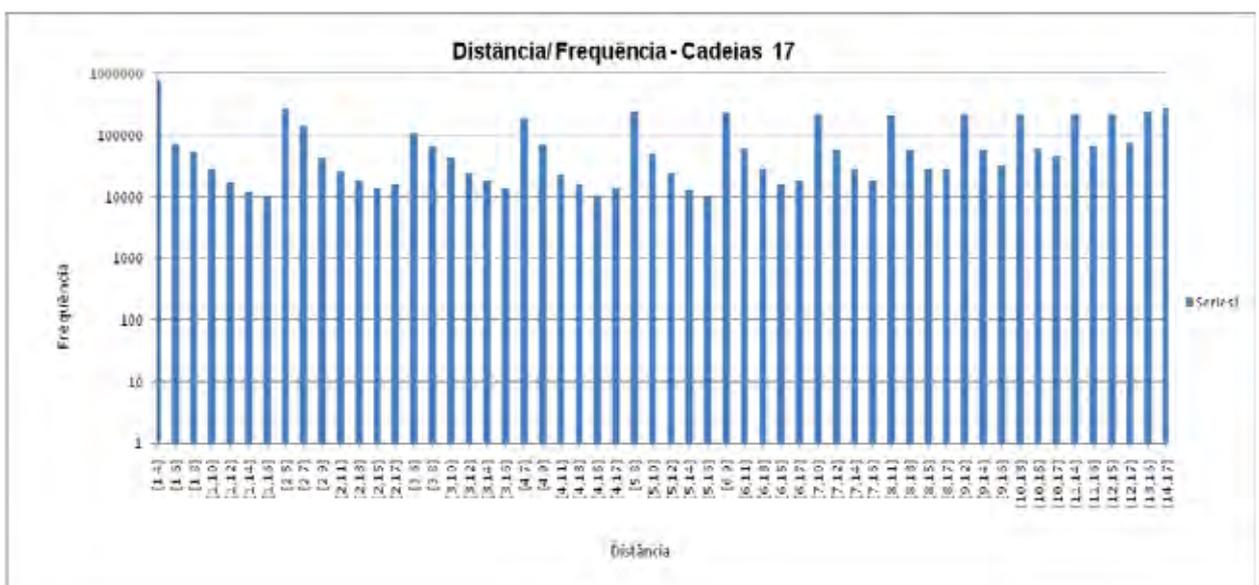


Figura 25 – Distância total em relação à frequência da cadeia 17

4.6 Contato Topológico de Longo Alcance

Contato Topológico de longo alcance (CTLA) Figura 26, é o número total de contatos ao longo alcance dividido pelo número de resíduos da cadeia formando assim contatos topológicos que estão separados, ao longo da seqüência, por mais de 7 monômeros, portanto, estão próximo no espaço, porém longe na cadeia.

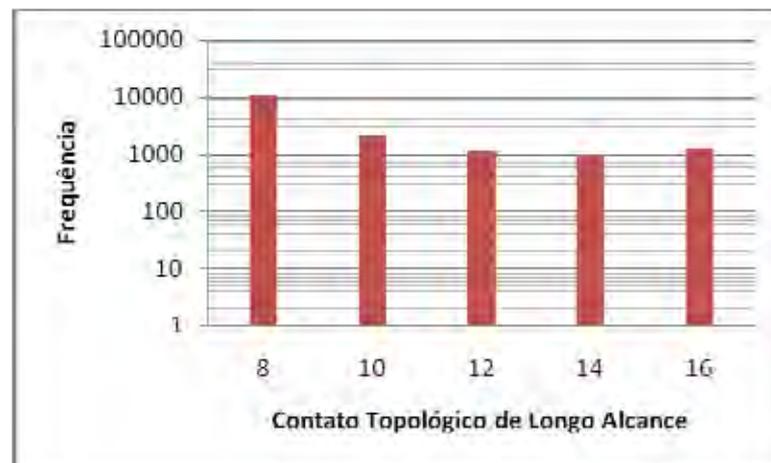


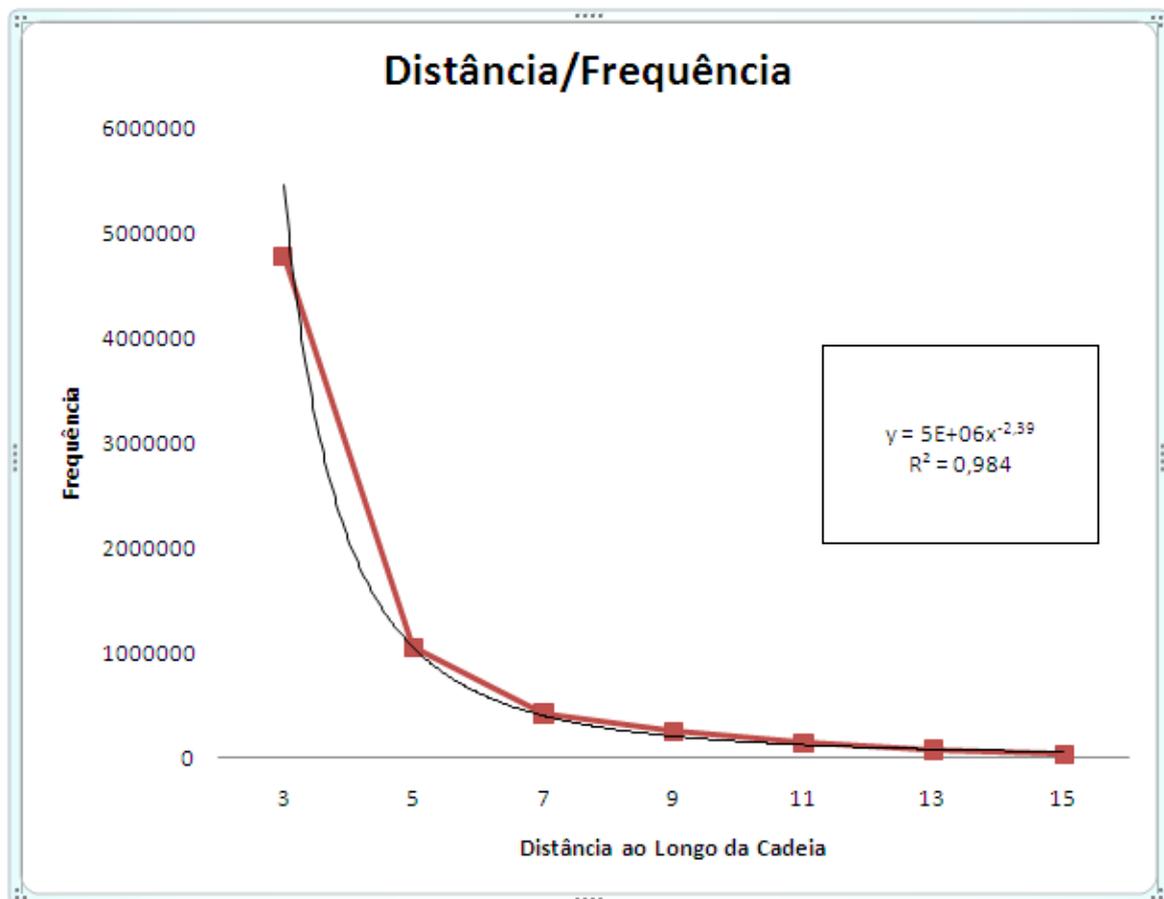
Figura 26 – Contatos de longo alcance, entre resíduos distantes da seqüência.

Na Figura 27 (a), distância ao longo da cadeia, isto é, na primeira coluna se encontra as distâncias encontradas até 17 monômeros, segunda coluna a frequência em relação a distância sem especificar qual o monômero.

Figura 27 (b), representação gráfica determinada pela lei de potência.

Distância ao longo da cadeia	
Distância	Frequência
3	4781501
5	1050896
7	421553
9	251181
11	137403
13	79011
15	35560

(a)



(b)

Figura 27 - (a) Tabela da distância em função da frequência ao longo da cadeia. (b) Gráfico da distância ao longo da cadeia em relação a frequência. Inserido no gráfico representamos a equação de melhor ajuste aos pontos e o coeficiente de correlação.

4.7 Mapa de Contato Topológico

A interação do presente modelo se dá entre os primeiros sítios vizinhos, portanto obteve-se o mapa de contatos dos inter-resíduos de 17 monômeros, Figura 28, onde verifica o número de monômero do eixo x, ligado ao número de monômero do eixo y, formando assim os pares de contatos topológicos.

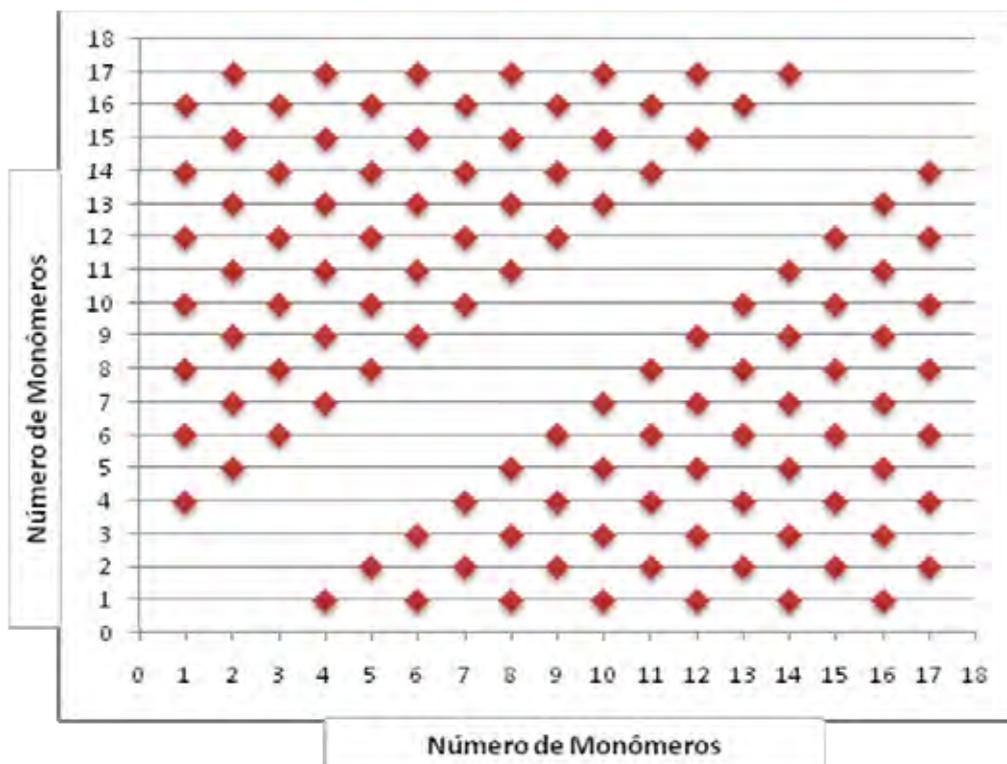
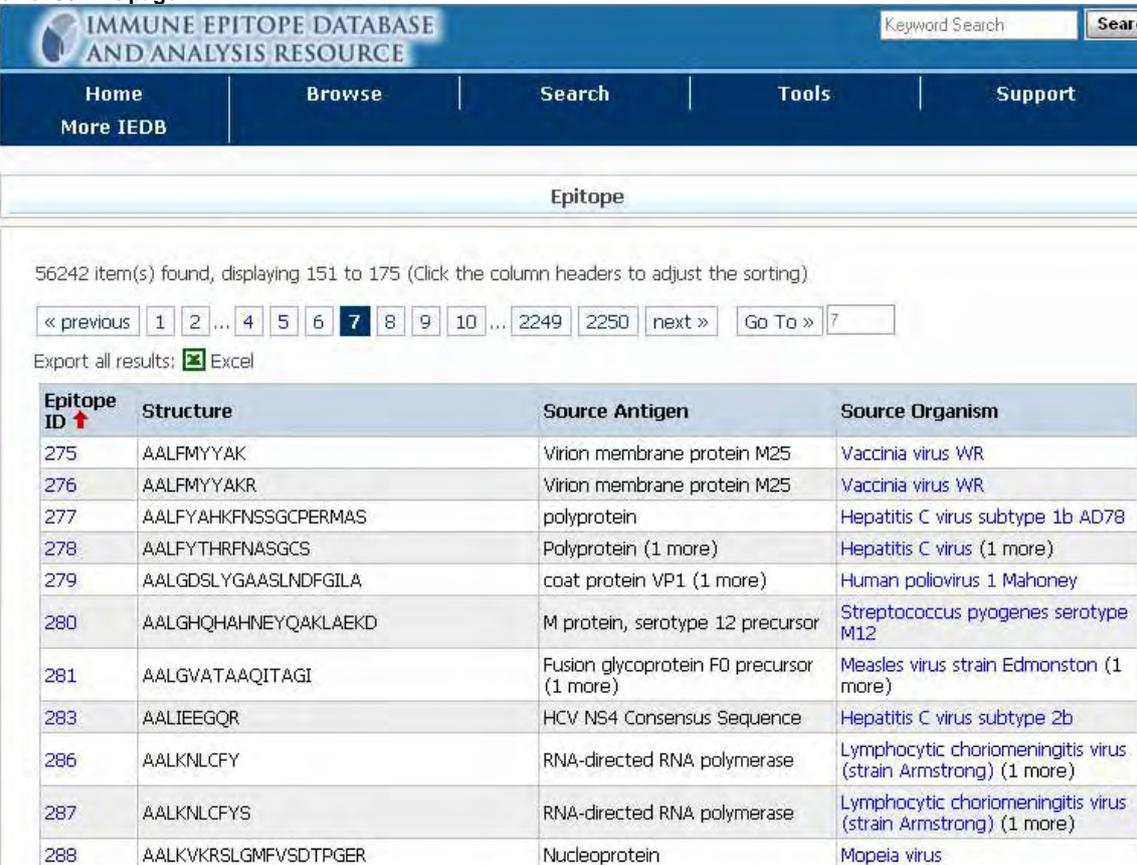


Figura 28 – Distribuição total das 2.155.667 configurações em função dos contatos topológicos dos primeiros vizinhos na rede quadrada.

4.8 Modelo aplicado a Epitopo

Na Figura 29, no banco de dados Immune Epitope Database and Analysis Resource escolhemos seqüência de epitopes para aplicar ao modelo HP. O critério utilizado para escolher as seqüências foi definido pela estrutura que deverá conter até 17 monômeros. Apresentamos na Figura 30 algumas seqüências de aminoácidos presentes em alguns epitopos encontrados nas proteínas de algumas espécies de organismos.

Fonte: http://www.immuneepitope.org/epitope_list.php?structureCategory=Peptidic&response=Positive&total_rows=56242&page=7



Epitope ID ↑	Structure	Source Antigen	Source Organism
275	AALFMYYAK	Virion membrane protein M25	Vaccinia virus WR
276	AALFMYYAKR	Virion membrane protein M25	Vaccinia virus WR
277	AALFYAHKFNSSGCPERMAS	polyprotein	Hepatitis C virus subtype 1b AD78
278	AALFYTHRFNAGCS	Polyprotein (1 more)	Hepatitis C virus (1 more)
279	AALGDSLYGAASLNDFGILA	coat protein VP1 (1 more)	Human poliovirus 1 Mahoney
280	AALGHQHAHNEYQAKLAEKD	M protein, serotype 12 precursor	Streptococcus pyogenes serotype M12
281	AALGVATAAQITAGI	Fusion glycoprotein F0 precursor (1 more)	Measles virus strain Edmonston (1 more)
283	AALIEEGQR	HCV NS4 Consensus Sequence	Hepatitis C virus subtype 2b
286	AALKNLCFY	RNA-directed RNA polymerase	Lymphocytic choriomeningitis virus (strain Armstrong) (1 more)
287	AALKNLCFYS	RNA-directed RNA polymerase	Lymphocytic choriomeningitis virus (strain Armstrong) (1 more)
288	AALKVKRSLGMFVSDTPGER	Nucleoprotein	Mopeia virus

Figura 29 - Banco de Dados Immune Epitope Database and Analysis Resource

Epitopo	Estrutura	Antigeno	Organismo	Modelo HP	NTC	NTCVT	CTLA > 7	Energia
275	AALFMYYAK	Virion membrane protein M25	Vaccinia virus WR	HHHHHHHHP	740	[1,4] ¹²⁴ , [1,6] ³⁹ , [1,8] ²² , [2,5] ⁸³ , [2,7] ²¹ , [2,9] ⁰ , [3,6] ⁸¹ , [3,8] ²¹ , [4,7] ⁸¹ , [4,9] ⁰ , [5,8] ⁸³ , [6,9] ⁰	2,7	-555
276	AALFMYYAKR	Virion membrane protein M25	Vaccinia virus WR	HHHHHHHHHP	2034	[1,4] ³⁴⁶ , [1,6] ¹¹¹ , [1,8] ⁶¹ , [1,10] ⁰ , [2,5] ²²⁷ , [2,7] ⁵⁶ , [2,9] ³³ , [3,6] ²²⁸ , [3,8] ⁵⁷ , [3,10] ⁰ , [4,7] ²¹⁵ , [4,9] ⁰ , [5,8] ²²⁸ , [5,10] ⁰ , [6,9] ⁰ , [7,10] ⁰	7,6	-1529
283	AALIEEGQR	HCV NS4 Consensus Sequence	Hepatitis C virus subtype 2b	HHHHPPPPP	740	[1,4] ¹²⁴ , [1,6] ⁰ , [1,8] ⁰ , [2,5] ⁰ , [2,7] ⁰ , [2,9] ⁰ , [3,6] ⁰ , [3,8] ⁰ , [4,7] ⁰ , [4,9] ⁰ , [5,8] ⁰ , [6,9] ⁰	0	-124
287	AALKNLCFYS	RNA-directed RNA polymerase	Lymphocytic choriomeningitis virus (strain Armstrong) (1 more)	HHHPHHP	2034	[1,4] ⁰ , [1,6] ¹¹¹ , [1,8] ⁶¹ , [1,10] ⁰ , [2,5] ⁰ , [2,7] ⁵⁶ , [2,9] ³³ , [3,6] ²²⁸ , [3,8] ⁵⁷ , [3,10] ⁰ , [4,7] ⁰ , [4,9] ⁰ , [5,8] ⁰ , [5,10] ⁰ , [6,9] ²²⁷ , [7,10] ⁰	7,6	-773
293	AALLQSAR	lipoprotein, 15 kDa (tpp15)	Treponema pallidum subsp. pallidum str. Nichols	HHHHPPHP	272	[1,4] ³⁸ , [1,6] ⁰ , [1,8] ⁰ , [2,5] ⁰ , [2,7] ⁶ , [3,6] ⁰ , [3,8] ⁰ , [4,7] ³⁰ , [5,8] ⁰	0	-74

Figura 30 – Modelo HP aplicado a Epitopo.

Figura 31, sendo computadas apenas as interações H-H, portanto obtemos a visualização das distâncias, os pares de contatos em relação a frequência e a energia configuracional do sistema.



Figura 31 – Modelo HP aplicado ao Epitopo AALFMYYAK

A Figura 32 observa a energia configuracional do sistema uma vez que está relacionada com a frequência de contatos.



Figura 32 – Distância em relação a frequência e energia configuracional.

Analisando a Figura 33, verificamos um energia alta na distância [1,4] diminuindo e mantendo um estado com maior probabilidade de desnaturar.

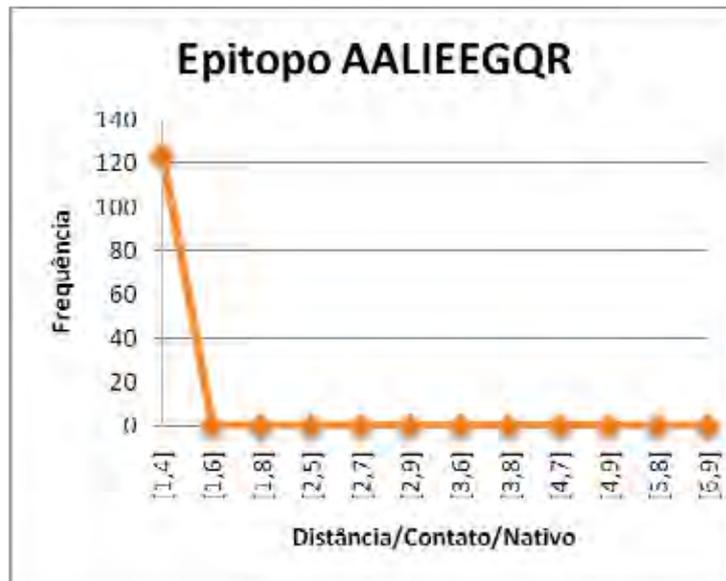


Figura 33 – Seqüência do Epitopo possui H-H apenas nos contatos [1,4].

Na Figura 34, distâncias [1,4], [1,10], [2,5], [3,10], [4,7], [4,9], [5,8], [5,10], [7,10], possui freqüência 0, devido sua interação não ser H-H.

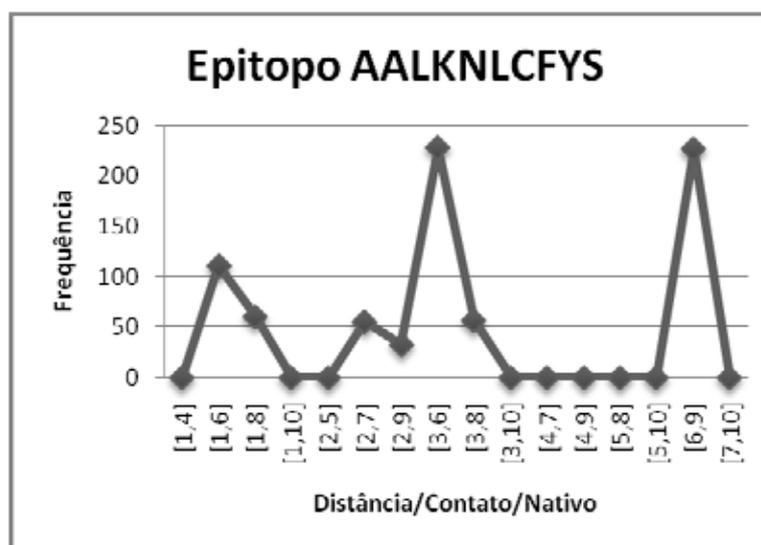


Figura 34 – Interações P-P e H-P de acordo com o Modelo não são computados.

Observando a distância [1,8], na Figura 35 verifica-se que está separado pela distância de 7 monômeros na cadeia, com uma frequência de 14 pares de contatos para esta seqüência de Epitopo .

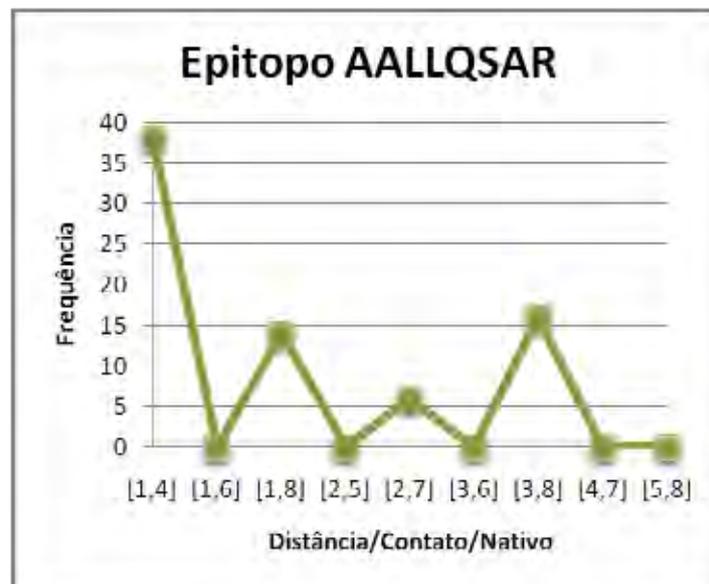


Figura 35 – Contato topológico de longo alcance, obtemos na distância [1,8]

Na Figura 36 a energia hidrofóbica do sistema está correlacionado a ordem de contato, assim, no modelo HP 740 configurações da cadeia 9, possui 740 possíveis contatos topológicos, ao aplicar ao Epitopo que possui esta seqüência AALIEEGQR, lembrando que a energia é somente computada com a interação H-H, obtemos 124 configurações que possuem contatos da cadeia 9, sendo descartadas as demais interações.

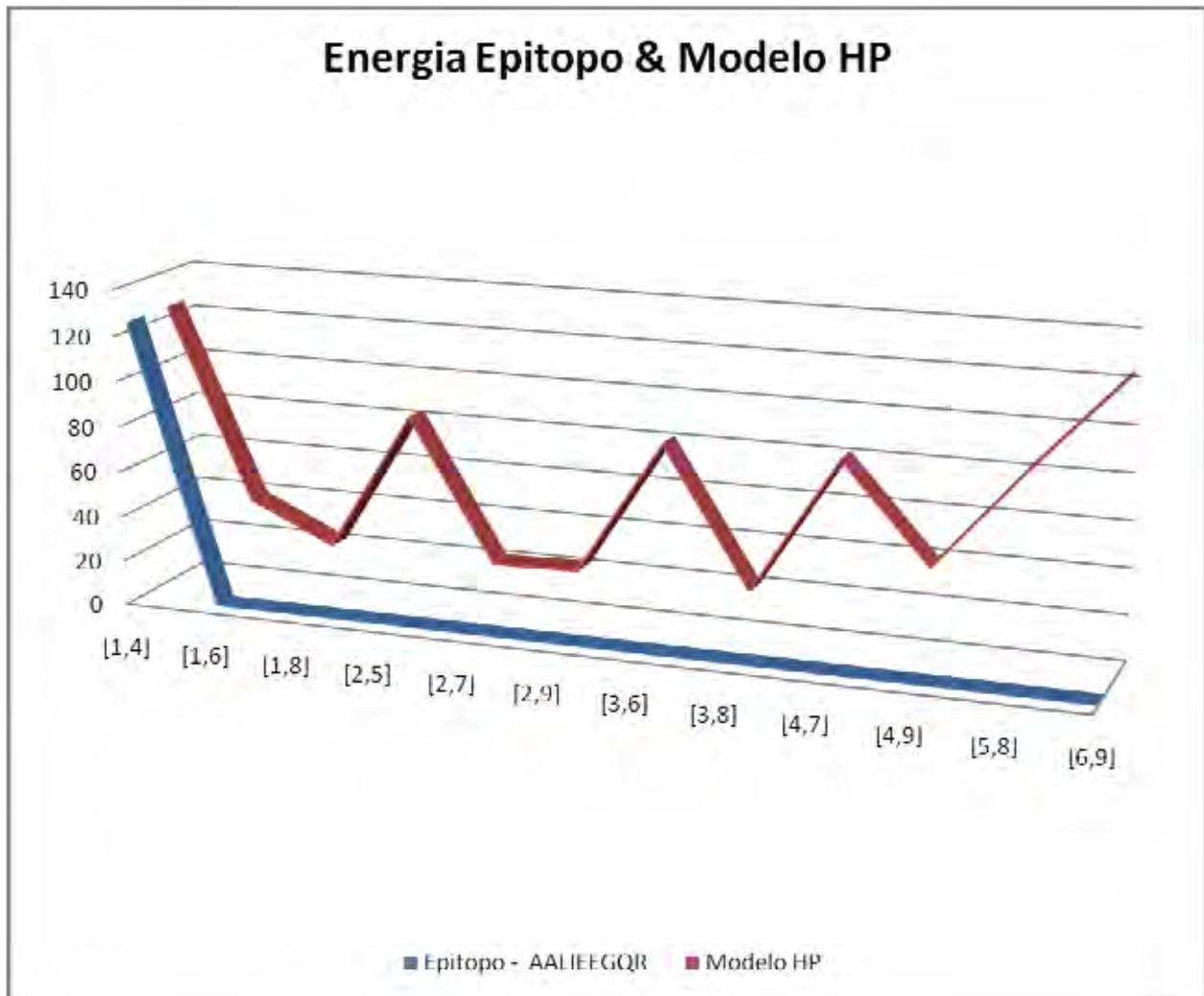


Figura 36 – Modelo HP aplicado ao Epitope de seqüência AALIEEGQR .

CONSIDERAÇÕES FINAIS

Neste trabalho confirmamos o número de possíveis configurações adotadas por um modelo de proteína na rede quadrada utilizando o modelo HP.

As vantagens deste modelo, é a facilidade de definir as seqüências de aminoácidos para cadeias de 17 monômeros, para estudar o comportamento e aplicar a qualquer estrutura com seqüências menores, tal qual o trabalho apresenta sua aplicação aos epitopos, onde distribui as distâncias em relação a freqüência de contatos, registrando a evolução de energia onde o índice de maior freqüência é encontrado, verificando os pares de contatos topológicos de longo alcance e levantando o histograma total de todas as configurações em relação ao seus contatos topológicos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AARONSON, S.; NP-complete Problems and Physical Reality, **Institute for advanced Study, Princeton**, NJ, 2005.
2. ADAMI, C.; What is complexity, **Review articles**, 2002.
3. AGUIAR, N. F.; HONORATO, C. S. G.; SANTOS, G. H.; OCHI, S. L.; Metaheurística busca tabu pra o problema de coloração de grafos, **Apresentado e publicado nos Anais do XXXVII SBPO – Gramado**, pg 2497-2504, 2005.
4. ALBERTS, T.; SHEFFIELD, S., Hausdor dimension of the SLE curve intersected with the real line, **Mathematics Subject Classification**, arXiv:0711.4070v1 [math.PR] 26 Nov, 2007.
5. ALBRECHT, A. A.; SKALIOTIS, A.; STEINHOFEL, K.; STOCHASTIC. Protein folding simulation in the three-dimensional HP-Model, **Science Direct Applied Soft Computing** 8, 1029–1040, 2008.
6. ANANIKIAN, N.; ANANIKYAN, L.; ARTUSO, R., Multi-dimensional mapping and folding properties for non-classical helix-stabilizing interactions of proteins, **Physical. A**, 361 255–262, 2006.
7. ARNALDI, E., LIBERTI, L., MAFFIOLI, F., MACULAN, N.; Edge-swapping algorithms for the minimum fundamental cycle basis problem, **Math Meth Oper Res** 69:205-233, 2009.
8. BANKES, Models as lab equipment: science from computational experiments, **Comput Math Organ Theory** ,15:8-10, 2009.
9. BARROSO, S., YANG. M. H.; Mecanismos da interação antígeno-anticorpo em uma resposta primária célula T. mediada, Sociedade Brasileira de Matemática aplicada e Computacional, 2006.
10. BAUER. B.; BERNARD, d.; Conformal field theories of stochastic loewner evolutions, **Commun. Math. Phys.** 239, 493–521, 2003.

11. BONAMIN, L. V.; "O Estresse e as Doenças", **Ciência Hoje**, 17(99), pp. 25-30, 1994.
12. BUI, N. T.; SUNDARRAJ, G.; An Efficient Genetic Algorithm for Predicting Protein Tertiary Structures in the 2D HP Model, **Gecco'05, june 25-29, 2005, Washington, DC, USA**, 2005.
13. CAI, L.; CHEN. J.; On fixed-parameter tractability and approximability of NP Optimization Problems, **Journal of computer and system sciences**, 54, 465-474, 1997.
14. CALDARELLI.G; RIOS, P. Cold and Warm Denaturation of Proteins, **Biophysical Journal** Volume 93 December, 2007.
15. CALIRI, A.; SILVA, A. A. M.; Geometrical effect on folding of macromolecules; **J. Chemphys** 106(18), 1997.
16. CASTRO, N. L., ZUBEN, V. J. F.; An Evolutionary Immune Network for Data Clustering, **Brazilian Symposium on Artificial Neural Networks**, pp.84-89, 2000.
17. CASTRO, N. L., ZUBEN, V. J. F.; Learning and Optimization Using the Clonal Selection Principle, **IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems**, 2001.
18. CEJTIN, H.; EDLER, J.; GOTTLIEB, A.; HELLING, R.; Li, H.; PHILBIN, J.; WINGREEN, N.; TANG, C. "Fast tree search for enumeration of a lattice model of protein folding." **The Journal of chemical physics** Vol. 116, N°1, 2002: 352-359.
19. CHANDRU, V.; DATTASHARMA, A.; KUMAR, A.; The algorithmics of folding proteins on lattices; **Discrete applied Mathematics** 127, 145-161, 2003.
20. CHARK, 2008.; Protein Folding: Are we there yet?; Archives of biochemistry and biophysics, 469, **Science Direct**, 2008 Editorial.
21. CHEN, X.; WEBER, I.; HARRISON, W.R.; Hidration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures; **J. Phys. Chem. B**, 113, 2008.

22. CRESCENZI, P. GOLDMAN, D. PAPADIMITRIOU, C. PICCOLBONI. A., YANNAKAKIS. M.; On the complexity of protein folding. **Journal of Computational Biology**, v. 5, p. 423–465, 1998.
23. CROUT, J.M.N.; Is my model too complex? Evaluating model formulation using model reduction, **Elsevier Environmental Modelling & Software**, 24 1-7, 2009.
24. DETHRIDGE, C. J.; GUTTMANN, J. A., Prudent Self-Avoiding Walks, 10, 309-318; DOI: 10.3390/e10030309, **Entropy**, 2008.
25. DILL, A. K.; BROMBERG, S.; YUE, K.; FIEBIG, M. K.; YEE, P. DAVID.; THOMAS, D. P.; CHAN, S. H.; Principles of protein folding – A perspective from simple exact models, **Protein Science**, 1995
26. DILL, A. K.; OZKN, B. S.; WEIKL, R. T.; CHODERA, D. J.; VOELZ, A.V.; The protein folding problem: When will it be solved?; **Science Direct**, 17:342-346, 2007.
27. DILL, A. K.; Theory for the folding and stability of globular proteins; **Biochemistry**, 24, 1501-1509, 1985.
28. DILL, K. A. ;CHAN, H. S. “From Levinthal to pathways to funnels.” **Nature Structure Biology** N^o4, 1997: 10-19.
29. DINNER, R. A.; SALI, A.; SMIDTH, J. L.; DOBSON, M. C.; KARPLUS, M.; Understanding protein folding via free-energy surfaces from theory and experiment, **FEATURE**, TIBS 25, 2000.
30. DREHER H.; The Immune Power Personality, Penguin Books, 1995.
31. FARMER, D. J.; PACKARD, H.N.; PERELSON, S. A.; The immune system, adaptation, and machine learning, *Physica* 22D, 187-204, 1986.
32. GERMAIN, R. N.; “MHC-Associated Antigen Processing, Presentation, and Recognition Adolescence, Maturity and Beyond”, **The Immunologist**, 3/5-6, pp. 185-190, 1995.
33. HART, W. E.; ISTRAIL, S. Fast protein folding in the Hydrophobic-Hydrophilic model within three-eighths of optimal. **Journal of Computational Biology**, v. 3, p. 53–96, 1996.

34. HONIG, B. "Protein Folding: From the Levinthal Paradox to Structure Prediction." **Journal Molecular Biology**, 283-293, 1999.
35. IRBACK, A., TROEIN, C.; Enumerating designing sequences in the HP Model, **Complex Systems Division, Department of theoretical Physics**, 2001.
36. JERNE, N. K.; "The Generative Grammar of the Immune System", **Science**, 229, pp. 1057-1059, 1985.
37. KARPLUS, M. The Levinthal paradox: yesterday and today, **Folding & Design** 01 Aug 1997, 2:S69–S75, 1997.
38. KAUZMANN, W. "The three Dimensional Structures of Proteins.", **Biophysical Journal** 4,1964: 43-54.
39. LAU, F.K. DILL, A.K. A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins. . **Macromolecules**,22,3986-3997, 1989.
40. LAVOR. C.; LIBERTI. L.; MACULAN, N.; The discretizable Molecular Distance Geometry Problem, **Department of Applied Mathematics**, 2008.
41. LEHNINGER, A. L. **Princípios de Bioquímica**. 1ª ed. São Paulo: Sarvier, 1991.
42. LEVINTHAL, C. Are there pathways for protein folding? **Journal of Chimie Physique et de Physico – Chimie Biologique**, 65, 1, 1968.
43. LI, H.; TANG, C.; WINGREEN, N. S. Nature of driving force for protein folding: a result from analyzing the statistical potential. **Physical Review Letters**, v. 79, p. 765–768, 1997.
44. LU, M. H., LIANG, J.; A model study of protein nascent chain and cotranslational folding using hydrophobic-polr residues, **InterScience**, 2007.
45. MANSON, M. S.; Simplifying complexity: a review of complexity theory; **Geoforum**, 32 304-414, 2001.

46. MARCON, CALAZANS, MORAES, RIES, HESSEL; Modelagem e descrição de sistemas computacionais – Um estudo de caso de comparação das linguagens VHDL e SDL, **Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)**, 2002.
47. MASON, D.; “Antigen Cross-Reactivity: Essential in the Function of TCRs”, **The Immunologist**, 6/6, pp. 220-222. 1998.
48. MERTENS, S.; Computational Complexity for Physicists, **Lecture given at the DPG school on “Quantum Computing”**, Bad Honnef, Oct. 9 – 13, 2000.
49. NETO, P. R. A.; Procedimento para a geração de arranjos físicos distribuídos baseado em otimização da simulação de algoritmos genéticos, **XXVII Encontro Nacional de Engenharia de Produção**, 2007.
50. OYLUMLUOGLU, G.; BUYUKKILIÇ, F.; DEMIRHAN, D. Investigation of hydration effect of the proteins by Phenomenological thermostistical methods, **Journal Physical. Chem. B** (2008).
51. P. De Los RIOS, G. CALDARELLI, Puttin proteins back into water, **Physical Review E**, 62, 8449 2000.
52. PACE C. N.; SHIDLEY, B. A.; MACNUTT, M.; GAJIWALA, K. “Forces contributing to the conformational stability of Proteins.” **The FASEB Journal** 10 (1996): 75-83.
53. PEDERSEN, S. N. C.; LYNGS, B. R.; Protein folding in the 2D HP Model, Department of computer Science, 2000.
54. PERELSON, A. S. , WEISBUCH, G.; Immunology for Physicists, **Rev. of Modern Physics**, 69(4), pp. 1219-1267, 1997.
55. QIAN, B.; RAMAN, S.; DAS, R.; BRADLEY, P.; MCCOY, J. A.; READ, J.R.; BAKER, D.; High-Resolution structure prediction and the crystallographic phase problem; **Nature Publishing Group**, Vol. 450, 2007.
56. RADFORD, E. S; Protein folding: progress made and promises ahead, **Reviews**, 2000.

-
57. REEVES, R. C.; A genetic algorithm for flowshop sequencing, **Computers Ops Res.** Vol. 22, no. 1, PP. 5-13, 1995.
58. REGO, C., LI, H., GLOVER, F.; A Filter-and fan approach to the 2D HP Model of the proteins folding problem, **Latest Revision**; 2006
59. RIOS, L. P.; CALDARELLI, G. Puttin proteins back into water, **Physical Review E**, 62, 8449, 2000.
60. SALVI, G.; RIOS, L. P.; Effective Interactions Cannot Replace Solvent Effects in a Lattice Model of Proteins. **Physical Review Letters**, vol. 91, number 25, 2003.
61. SALVI, G.; MOLBERT, S.; RIOS, L.D. P.; Design of lattice proteins with explicit; **Physical Review E** 66, 061911, 2002.
62. SCHRAMM, O.; Scaling Limits of Loop-Erased Random Walks and Uniform Spanning Trees, **Israel Journal of Mathematics**, 221-288, 2000.
63. SCOTT, K. J., SMITH, P. G.; Searching for peptide ligands with an Epitope Library, **Downloaded from www.sciencemag.org** on July 17, 2009.
64. SHIRLEY, B. A.; STANSSENS, P.; HAHN, U.; PACE, C. N. "Contribution of hydrogen bonding to the conformational stability of ribonuclease T1." **Biochemistry** 31, 1992: 725-732.
65. SILVA, A. R. **Enovelamento determinístico de macromoléculas**. 2002. 143 p. Dissertação (Doutorado em Ciências) - Faculdade de Filosofia Ciências e Letras, Universidade de São Paulo, Ribeirão Preto.
66. SILVA, R. I. **Enovelamento protéico: fatores topológicos**. 2005. 74 p. Dissertação (Doutorado em Ciências) - Faculdade de Filosofia Ciências e Letras, Universidade de São Paulo, Ribeirão Preto.
67. SPITZER, F. - Principles of Random walk, Second Edition, **Springer** 2001.
68. TANG, C. Simple models of the protein folding problem. **Physica A**, v. 288, p. 31-48, 2000.

69. TARRAGÓ, P. E. M. **Potencial estéreo-hidrofóbico e propriedades topológicos no enovelamento de proteínas**. 2003. 158 p. Dissertação (Doutorado em Ciências) - Faculdade de Filosofia Ciências e Letras, Universidade de São Paulo, Ribeirão Preto.
70. THALHEIM, T.; MERKLE, D.; MIDDENDORF, M.; Proteins folding in the HP-Model solved with a hybrid population based ACO algorithm; *International Journal of computes science*, 35:3, IJCS-35-3, 2006.
71. TIZARD, I. R.; Immunology An Introduction, Saunders College Publishing, 4th Ed. 1995.
72. TONEGAWA, S.; Somatic Generation of Antibody Diversity, **Nature**, 302, pp. 575- 581, 1983.
73. TONEGAWA, S.; The Molecules of the Immune System, **Scientific American**, 253(4), pp. 104-113, 1985.
74. VEIT, E. A.; MORS, P. M ; TEODORO, V. D. Ilustrando a segunda lei de Newton no século XXI. **Revista Brasileira de Ensino de Física**, São Paulo, v.24, n.2, p. 176-184, 2002.
75. WERNER, W.; Conformal restriction and related Questions, **Probability Surveys**, Vol. 2 145-190, 2005.
76. WHITFORD, C. P., NOEL, K. J., GOSAVI, S., SHUG, A., SANBONMATSU, Y. K., ONUCHIC, N. J.; An all-atom structure-based potencial for proteins; Bridging minimal models with all-atom empirical forcefields, **Proteins**, 75: 430-441, 2009.
77. WHITLEY, D.; A genetic algorithm tutorial, Statistics and Computing, 1994.
78. WILLIAMS, K. L.; Enumerating Up-Side Self-Avoiding Walks on Integer Lattices, **Harvard University Cambridge**, MA 02138, 1996.
79. WROE, R.; BAUER, B. E.; CHAN, S. H.; Comparing Folding Codes in Simple Heteropolymer Models of Protein Evolutionary Landscape: Robustness of the Superfunnel Paradigm, **Biophysical Journal**, vol. 88, 2005

80. YAP, H. E., FAWZI, L. N., GORDON, H. T.; A coarse-grained carbon protein model with anisotropic hydrogen – bonding, InterScience, **Proteins**, 70: 626-638, 2008.
81. ZHAO X., Advances on protein folding simulations based on the lattice HP models with natural computing, **Science Direct**, 2008.