

UNIVERSIDADE ESTADUAL PAULISTA

Instituto de Geociências e Ciências Exatas

Campus de Rio Claro

**TERMODINÂMICA DO ENOVELAMENTO DE CADEIAS
HETEROPOLIMÉRICAS ATRAVÉS DO ALGORITMO DE
WANG-LANDAU**

Fabio Bresighello Beig

Orientador: Prof. Dr. Makoto Yoshida

Dissertação de Mestrado elaborada junto
ao Programa de Pós-Graduação em Física
– Área de Concentração em Física
Aplicada, para obtenção do Título de
Mestre em Física.

Rio Claro - SP

2006

621 Beig, Fábio Bresighello
B422t Termodinâmica do enovelamento de cadeias heteropoliméricas através do algoritmo de Wang-Landau / Fábio Bresighello Beig. – Rio Claro : [s.n.], 2006
90 f. : il., gráfs., tabs.

Dissertação (mestrado) – Universidade Estadual Paulista, Instituto de Geociências e Ciências Exatas
Orientador: Makoto Yoshida

1. Física aplicada. 2. Enovelamento. 3. Método Wang-Landau. 4. Modelo HP. 5. Proteínas. I. Título.

Ficha Catalográfica elaborada pela STATI – Biblioteca da UNESP
Campus de Rio Claro/SP

Comissão Examinadora

Prof. Dr. Makoto Yoshida

Prof. Dr. Valter Luiz Líbero

Prof. Dr. João Ruggiero Neto

Fábio Bresighello Beig

Rio Claro, 21 de Fevereiro de 2006.

Resultado: ***Aprovado***

À minha amada imortal, por ser metade do que sou, por trazer a tona o que há de melhor em mim e por sempre ser uma companheira, mesmo nos momentos de dificuldade pelos quais passei.

Agradecimentos

À minha companheira e esposa Juliana, por batalhar a cada dia pelo meu melhor, por estar sempre do meu lado, e por ter o sorriso mais belo que me acalentou nos momentos mais difíceis. Agradeço também pelas discussões e reflexões, pois lapidaram a minha alma e enterneceram meu ser;

Aos meus pais, Irineo Fernando Beig e Maria Luisa Moura Bresighello pela enorme paciência e compreensão;

Aos meus irmãos André e Beatriz;

Aos meus avós, Irineo Beig e Dulce;

Aos meus gatos Schrödinger e Persa (em memória);

Aos meus padrinhos Noelia e Nelson (em memória);

Aos grandes amigos da Física, Diogo, William, Igor, Wadley, Juliana, Kung, Humberto, Amanda, Gui, Ramirez, Otto e Madu (em memória), pelas conversas, discussões e desafios. Obrigado pela grande amizade que cada um de vocês pôde partilhar comigo durante todos esses anos;

Ao Prof. Dr. Makoto Yoshida pela orientação deste trabalho, pelos incontáveis ensinamentos, confiança, pelas palavras de ânimo, pelas conversas e todo apoio durante a realização deste trabalho;

À Profa. Dra. Alzira por todo o apoio, conversas e muitos ensinamentos durante a graduação e mesmo durante o mestrado;

A todos professores de graduação e pós-graduação pelas muitas sementes plantadas nestes anos;

À CAPES pela concessão da bolsa durante a pesquisa.

Apoio CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior)

SUMÁRIO

Índice	i
Resumo	iii
Abstract	iv
Lista de Tabelas	v
Lista de Figuras	vi
Capítulo I – Introdução	01
Capítulo II – Método de Wang-Landau	06
Capítulo III – Enovelamento de Proteínas	30
Capítulo IV – Resultados e Discussão	56
Conclusões	86
Sugestões para trabalhos futuros	87
Referências Bibliográficas	88

ÍNDICE

Capítulo I – Introdução	01
Capítulo II – Método de Wang-Landau	06
2.1 Introdução	06
2.2 O Método de Monte Carlo e o Algoritmo de Metropolis	07
2.2.1 Equação Mestra e Princípio do Balanço Detalhado	08
2.2.2 O Algoritmo de Metropolis	10
2.2.3 Amostragem por Importância	12
2.2.4 Aplicação do Algoritmo de Metropolis	14
2.3 O Algoritmo de Wang-Landau	18
Capítulo III – Enovelamento de Proteínas	30
3.1 Introdução	30
3.2 Proteínas	31
3.2.1 Aminoácidos	31
3.2.2 Monômeros	34
3.2.3 Estrutura de uma proteína	35
3.3 O Enovelamento de Proteínas	37
3.4 Modelos	38
3.5 Mecânica Estatística de Cadeias Poliméricas	40
3.5.1 O Algoritmo de Metropolis	42

3.5.2 O Algoritmo de Wang-Landau	47
3.5.3 Condições de Contorno	49
3.5.4 Grandezas Termodinâmicas	50
3.6 Resultados Exatos	53
Capítulo IV – Resultados e Discussão	56
4.1 Introdução	56
4.2 Modelo HP	58
4.3 Modelo HP Misto	67
4.4 Modelo AB	72
4.5 Modelo de 20 Aminoácidos	77
4.6 Modelo de Interações Independentes	82
Conclusões	86
Sugestões para trabalhos futuros	87
Referências Bibliográficas	88

Resumo

Estudamos o enovelamento de proteínas através da termodinâmica de uma cadeia heteropolimérica. Para isso, adaptamos um método de Monte Carlo introduzido por Wang e Landau para o estudo de transições de fase em sistemas magnéticos para estudarmos a mecânica estatística dessa cadeia. Trata-se de um método que se vale do passeio randômico no espaço de energias para determinação da densidade de estados acessíveis à cadeia e com ela determinar grandezas termodinâmicas do sistema em estudo. Para obtermos essa densidade de estados, adotamos modelos de rede para gerarmos todas as conformações geométricas possíveis de uma cadeia de 27 monômeros em uma rede cúbica. Estudamos várias seqüências de monômeros adotando os modelos de rede mais relevantes utilizados para a investigação do enovelamento de proteínas tais como o modelo HP e modelos mais elaborados que utilizam as interações de Miyazawa-Jernigan entre monômeros. Calculamos diversas grandezas termodinâmicas essenciais para a compreensão do enovelamento de proteínas e com isso mostramos a eficiência do método Wang-Landau.

Palavras-chave: enovelamento, método Wang-Landau, modelo HP, proteínas

Abstract

We studied the protein folding through investigating the thermodynamics of a heteropolymeric chain. For this purpose, a Monte Carlo method introduced by Wang and Landau to study the phase transitions in magnetic systems, was adapted to investigate the statistical mechanics of this chain. The method is based on the random walk in the energy space to determine the density of states of the chain and important thermodynamics quantities. For this purpose, we adopted the lattice model and generate all geometric conformations of a 27-monomer chain in a cubic lattice. We studied several sequences of monomers adopting the most relevant lattice models, such as the HP model and more elaborated models considering the Miyazawa-Jernigan interactions. We computed several thermodynamics quantities to help us to understand the protein folding and show the efficiency of the Wang-Landau method.

Key words: folding, Wang-Landau method, HP model, proteins

Lista de Tabelas

- Tabela 3.1** A tabela abaixo apresenta cada um dos 20 aminoácidos naturais (monômeros), bem como sua classificação (quarta coluna) entre polar (*P*) e hidrofóbico (*H*). Nas duas primeiras colunas, são apresentadas as abreviações mais conhecidas e usadas para cada um dos aminoácidos apresentados. Na terceira coluna estão os nomes em português de cada um dos monômeros. 34
- Tabela 3.2** Tabela que apresenta na primeira linha, os valores exatos da seqüência em estudo, enquanto na segunda, traz as respectivas densidades de estado. 54
- Tabela 3.3** Tabela que apresenta na segunda e na terceira linha as entropias exata e obtida numericamente em função da energia *E*. 54
- Tabela 4.1** Na Tabela, a primeira coluna servirá de referência para a identificação da cadeia, a segunda coluna determina o número de monômeros, já a última mostra a seqüência da cadeia. 57

Lista de Figuras

- Figura 2.1** Comparação da densidade de estados obtida pelo algoritmo em estudo para o modelo de Ising bidimensional e os resultados exatos calculados pelo método da Ref. [38]. Erros relativos $\epsilon(\log(g(E)))$ são mostrados no *inset*. Figura extraída da Ref. [29]. 27
- Figura 2.2** Comparação da energia livre de Gibbs por local da rede calculada diretamente da densidade de estados da simulação para o modelo de Ising $L=256$ e as soluções exatas de Ref. [39]. Os erros relativos $\epsilon(F)$ que são mostrados no *inset*. A densidade de estados foi obtida através de caminhos aleatórios com apenas 6.1×10^6 passos de Monte Carlo. Figura extraída da Ref. [29]. 28
- Figura 2.3** Calor específico para o modelo de Ising bidimensional em uma rede de 256×256 em uma larga região de temperatura. O erro $\epsilon(C)$ relativo é mostrado no *inset*. Figura extraída da Ref. [29]. 29
- Figura 3.1** Um aminoácido típico e sua estrutura tetragonal. 32
- Figura 3.2** Reação entre dois aminoácidos e formação de uma ligação peptídica. 33
- Figura 3.3** Sequência de aminoácidos em um peptídeo. A cadeia na cor púrpura é a cadeia principal e as de cor verde são as laterais 33
- Figura 3.4** A figura mostra duas representações da Mioglobina em sua estrutura terciária. Em (a) a proteína é apresentada por um modelo de bolas e palitos e, em (b.), ela é representada por um modelo de fitas. 35

- Figura 3.5** Porção de uma cadeia helicoidal apresentada através de três representações. Em (a) a cadeia α -hélice é representada por um modelo de bolas e palitos. Em (b), a cadeia é representada por seu esqueleto, que se obtém ligando-se os carbonos- α da seqüência e por fim, em (c) um modelo da cadeia representada por fitas. 36
- Figura 3.6** Uma cadeia de monômeros segundo o modelo HP. As linhas pretas grossas representam ligações peptídicas. 41
- Figura 3.7** A figura representa os possíveis movimentos de uma cadeia. Movimento de manivela (a), movimento de canto (b) e movimento de ponta (c). 42
- Figura 3.8** Apresenta em vermelho os resultados obtidos com o método, enquanto que as linhas apresentam os resultados exatos. No gráfico (a) é apresentada a energia interna, no (b) o calor específico, no (c) a energia livre e no (d) a curva da entropia. Todas as curvas como função da temperatura ($k_b T = \mathcal{J}$). 55
- Figura 4.1** A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se um gráfico do número de contatos, como função da energia. 59
- Figura 4.2** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 60
- Figura 4.3** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 60
- Figura 4.4** O gráfico principal, apresenta a entropia como função da temperatura. No *inset* dessa figura, tem-se a entropia do método WL. 61
- Figura 4.5** Probabilidade de se encontrar o sistema em uma dada energia E , para cada uma das temperaturas indicadas. A temperatura crítica vale $T=0,60$. 62
- Figura 4.6** A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se um gráfico do número de contatos, como função da energia. 63

- Figura 4.7** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 63
- Figura 4.8** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 64
- Figura 4.9** No gráfico principal, tem-se a entropia como função da temperatura. Já no *inset* dessa figura, a entropia do método WL. 65
- Figura 4.10** Probabilidade de se encontrar o sistema em uma dada energia E , para cada uma das temperaturas indicadas, onde $T=0.133$, é a temperatura crítica. Neste gráfico foram apenas consideradas as quatro menores energias do sistema. 66
- Figura 4.11** A figura (a) representa um estado desnaturado qualquer enquanto que a (b) representa uma das conformações do estado fundamental. H representam os monômeros hidrofóbicos e P os polares. 67
- Figura 4.12** A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se o gráfico do número de contatos, como função da energia. 68
- Figura 4.13** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 69
- Figura 4.14** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 70
- Figura 4.15** A figura apresenta uma das conformações do estado fundamental, onde H são monômeros hidrofóbicos e P polares. 71
- Figura 4.16** No gráfico principal é apresentada a entropia como função da temperatura. Já no *inset* dessa figura, a entropia do método WL. 71
- Figura 4.17** O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=0,334$, no caso a temperatura crítica do sistema. 72

- Figura 4.18** A linha preta representa a energia, enquanto que a vermelha o número de contatos, ambas em função da temperatura. No *inset* se tem o gráfico do número de contatos como função da energia. 73
- Figura 4.19** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 74
- Figura 4.20** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 75
- Figura 4.21** No gráfico principal esta a entropia como função da temperatura. No “inset” dessa figura, a entropia do método WL. 75
- Figura 4.22** O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=1,54$, no caso a temperatura crítica do sistema. 76
- Figura 4.23** A figura (a) representa um estado desnaturado qualquer, enquanto que a (b) representa o estado nativo. 77
- Figura 4.24** A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se o gráfico do número de contatos, como função da energia. 78
- Figura 4.25** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 79
- Figura 4.26** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 80
- Figura 4.27** Entropia S como função da temperatura T . O *inset* dessa figura mostra a entropia do método WL. 80
- Figura 4.28** O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=0,11635$, no caso a temperatura crítica do sistema. 81

- Figura 4.29** A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* está um gráfico do número de contatos, como função da energia. 83
- Figura 4.30** A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia. 83
- Figura 4.31** A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura. 84
- Figura 4.32** No gráfico principal, tem-se a entropia como função da temperatura. No *inset* dessa figura, se tem a entropia do método WL. 84
- Figura 4.33** O gráfico acima representa a probabilidade de se encontrar o sistema em uma dada em energia, para cada uma das temperaturas indicadas, onde $T=0,595$, é a temperatura crítica. 85

Capítulo I

Introdução

As proteínas são cadeias poliméricas longas sintetizadas nas células na forma de uma cadeia linear, sendo esta a estrutura conhecida como primária, a qual rapidamente evolui para uma conformação geométrica funcional, uma estrutura tridimensional conhecida como estado nativo. Na maioria dos casos as proteínas executam as funções fisiológicas para as quais foram programadas quando se encontram nessa conformação especial também conhecida como estrutura terciária. As proteínas maiores, antes da conformação nativa, são constituídas de estruturas internas intermediárias, denominadas estruturas secundárias.

Com a recente conclusão do seqüenciamento do genoma humano [1,2], os pesquisadores passaram a compreender como um gene especifica exatamente a seqüência de aminoácidos em uma proteína a ser sintetizada. Entretanto, não se compreende ainda como e por que diferentes seqüências de aminoácidos nas proteínas, codificadas em genes diferentes, podem formar uma mesma estrutura tridimensional básica. As investigações para responder a essas perguntas são intensas, particularmente nas áreas de biologia molecular, bioquímica, química, física e computação. Este problema constitui-se hoje no que se conhece por problema do enovelamento de proteínas [3-5].

Como consequência desses seqüenciamentos de genes (genoma humano, por exemplo), o estudo e a solução do enovelamento de proteínas tornam-se urgentes já que este processo também pode ser considerado como um tradutor do código genético e completa a transferência das mensagens do DNA para a proteína no estado funcional. Portanto, é

fundamental descobrir como a seqüência dos aminoácidos codificada nos genes determina a forma da estrutura nativa da proteína e qual a correspondência dessa estrutura com as funções atribuídas à proteína [6-7].

Considera-se também importante para a investigação da causa de algumas doenças e, conseqüentemente, para a escolha de um tratamento racional das mesmas. São casos como o mal de Alzheimer [11], de fibroses císticas e algumas formas de câncer, aparentemente não correlacionadas, mas que tem uma natureza comum: o enovelamento defeituoso de proteínas [8]. Nesses casos, as proteínas não executam as funções vitais esperadas. Um importante passo seria dado em direção ao tratamento dessas doenças com a descoberta do por que certas proteínas evoluem para uma forma enovelada incorreta [9].

Após o seqüenciamento do genoma humano [2], a solução do problema do enovelamento de proteínas também passa a ser fundamental para a Biotecnologia. Hoje seria possível substituir genes no DNA, através da manipulação genética, com o intuito de modificar uma proteína ou sintetizar uma nova proteína com novas funções [10]. Para isso, é necessário realizar um estudo detalhado com a finalidade de compreender como a proteína, a ser sintetizada, iria se enovelar, bem como a sua forma nativa estaria associada às funções fisiológicas.

Há décadas o problema de enovelamento de proteínas tem sido alvo de investigações e tem se mostrado extremamente complexo. Atualmente, as atividades concentram-se na solução de quatro problemas básicos [12-16], que descrevo brevemente a seguir.

O primeiro deles consiste na determinação da estrutura de proteínas propondo-se, inicialmente, uma seqüência de aminoácidos. Considerando as forças de interação entre os monômeros da cadeia, os graus de liberdade desse sistema e argumentos termodinâmicos, procura-se simular a proteína e determinar numericamente a estrutura terciária ou a sua conformação nativa [4,17].

Em uma segunda frente de investigação, tenta-se resolver um antigo problema na área de enovelamento de proteínas. Há uma grande preocupação em se conhecer os mecanismos que levam uma proteína da sua estrutura primária para a terciária em um tempo relativamente curto quando comparado ao tempo que essa proteína, hipoteticamente, iria gastar para procurar aleatoriamente, entre todas as conformações geométricas possíveis, aquela que corresponderia à conformação nativa (paradoxo de Levinthal) [18]. Ainda nesta linha,

procura-se simular numericamente a dinâmica de enovelamento de proteínas considerando as interações entre monômeros e os graus de liberdade da proteína [19-21].

Numa terceira frente de trabalho, a qual pode ser considerada como uma investigação inversa ao do enovelamento, há um grande esforço numa linha de pesquisa que tem como objetivo o “desenho” de proteínas. A idéia consiste em “desenhar” uma proteína partindo de sua estrutura terciária, a princípio conhecida, para se determinar a seqüência de aminoácidos da proteína investigada [5,22-24].

Finalmente, sabe-se que existem proteínas que não se enovelam espontaneamente, mas necessitam da ajuda de outras para atingir o estado nativo. Nesta frente de investigação, há um imenso interesse em se compreender as interações entre proteínas com o objetivo de elucidar o processo de enovelamento de proteínas mediado por proteínas.

Nos últimos 10 anos houve grande progresso na compreensão do enovelamento de proteínas, graças a um grande volume de resultados teóricos obtidos nessa área [14-16], onde os principais focos da atenção dos pesquisadores foram:

- as seqüências de aminoácidos necessárias para que a proteína se enovele rapidamente e permaneça termodinamicamente estável no estado nativo;
- os mecanismos termodinâmicos para a estabilização da proteína e quais os mecanismos dinâmicos do enovelamento de proteínas.

Sob ponto de vista teórico, alguns dos mecanismos básicos que levam uma proteína a se enovelar foram compreendidos através de simulações numéricas utilizando modelos simplificados de proteínas. Modelos deste tipo são conhecidos na literatura como modelos na rede [12], quando a simulação ocorre em uma rede discreta de pontos, e fora da rede [26] quando a simulação é feita no espaço contínuo. Trata-se de modelos onde uma cadeia de aminoácidos, às vezes convenientemente vista como uma cadeia polimérica homogênea ou heterogênea, é substituída por sistema de pequenas esferas atadas ao longo de um fio, tal qual contas em um fino cordão [26]. Para que esses modelos sejam caricaturas fiéis de proteínas ou de uma cadeia polimérica, deve-se introduzir os potenciais de interação entre as esferas, compatíveis com o comportamento observado: o enovelamento da cadeia.

Intuitivamente, a idéia de se tratar proteínas como uma cadeia polimérica é bastante tentadora, pois sabe-se que uma cadeia polimérica pode sofrer um colapso através do qual passa de uma fase com grande número de conformações para uma fase de poucas conformações [14]. Apesar de simples, esses modelos teóricos apresentam um enorme espaço de conformações acessíveis à cadeia, inclusive a de menor energia, que corresponde ao estado nativo ou de cadeia enovelada e possuem os ingredientes básicos que permitem descrever de maneira razoável, o enovelamento de proteínas através da mecânica estatística [12,17,26]. Através de grandezas termodinâmicas, pode-se verificar as fases enoveladas (ordenadas) ocorrendo a temperaturas abaixo de uma temperatura crítica de enovelamento, e a desnaturada, ocorrendo acima dessa temperatura. Cada uma dessas fases (estado nativo – estado desnaturado) pode ser atingida com a variação da temperatura. Modelos simplificados, como os modelos na rede, ainda oferecem boas perspectivas para o estudo do enovelamento de proteínas [26].

Existem modelos mais realistas [27-28] que podem ser adotados para o estudo do enovelamento e certamente a qualidade das informações obtidas é bem maior. Mas isso exige uma estrutura computacional de alto desempenho, assim a utilização desses modelos ficaria restrita a cadeias pequenas devido à complexidade das estruturas das proteínas.

Neste trabalho de dissertação tem-se como objetivo investigar um problema que se enquadra naquela categoria de problemas em que a seqüência de aminoácidos é conhecida *a priori* e procura-se identificar os mecanismos que levam uma cadeia heteropolimérica ao enovelamento através da investigação de suas propriedades termodinâmicas. São problemas nos quais, em passado recente, foram utilizadas simulações numéricas tais como métodos de Monte Carlo, amplamente aplicados a estudos de sistemas magnéticos em Mecânica Estatística [25].

É também objetivo desse trabalho seguir a linha de pesquisa que consiste em introduzir novas técnicas numéricas de investigação do enovelamento de proteínas. Recentemente, Wang e Landau [29] introduziram um método de Monte Carlo muito promissor para simulação de sistemas magnéticos, particularmente útil para investigação de transições de fase de primeira ordem. A técnica foi derivada das idéias do método Multicanônico [30] ou *Entropic Sampling* [31] e é aplicada para determinar-se a densidade de estados do sistema estudado em função da energia. Um aspecto interessante é que o método permite o cálculo da entropia do sistema de forma direta, sem que haja a necessidade de integração do calor específico, um procedimento necessário quando se utiliza o tradicional

algoritmo de Metropolis [32]. Outro fator relevante do método é a possibilidade de dividir o cálculo da densidade de estados em intervalos independentes de energia, o que é muito conveniente, pois possibilita o processamento numérico paralelo. Assim, verificam-se as vantagens deste método sobre o de Metropolis tradicional.

No Capítulo II será introduzido o método de Monte Carlo de Wang e Landau, tal como ele foi proposto para o estudo do modelo de Ising. O método original será discutido detalhadamente ainda nesse capítulo.

No Capítulo III serão providas informações básicas sobre o problema do envelhecimento de proteínas, além de discutidas a adaptação do método Wang-Landau para o cálculo de propriedades termodinâmicas de uma cadeia heteropoliméricas.

No Capítulo IV serão apresentados os resultados obtidos com o método. Nessa dissertação foram estudados os modelos de rede mais relevantes e frequentemente utilizados na área. E finalmente, no Capítulo V, serão apresentadas algumas conclusões obtidas durante o desenvolvimento desta dissertação.

Capítulo II

Método de Wang-Landau

2.1 Introdução

O início da utilização científica do computador, ocorrida durante o século passado, permitiu que vários problemas de solução extremamente complexa, ou ainda ditos insolúveis, encontrassem um caminho alternativo para a solução através de métodos numéricos. Assim, viu-se o surgimento de um novo ramo para a ciência. Pode-se, entretanto, dizer que a implementação de métodos computacionais teve um papel muito maior do que simplesmente ser uma nova técnica para a resolução de problemas.

Os métodos computacionais abriram todo um leque de possibilidades. Uma nova maneira de se fazer ciência entrava em vigor. Com a disseminação de recursos computacionais, as experiências numéricas passaram a ocupar um lugar de importância na Física contemporânea [33].

Uma das vantagens das simulações computacionais está na possibilidade da obtenção da informação exata para modelos bem definidos, na qual os erros, a princípio, podem ser tão pequenos quanto se queira. Em sistemas estatísticos, a resolução dos problemas com obtenção de resultados precisos é determinada pela quantidade de iterações, ou seja, processos repetitivos para obtenção de resultados com o objetivo de melhorar o cálculo de médias. No caso de problemas analíticos teóricos, o que ocorre é bem diferente. Tal exatidão é encontrada

muito raramente e para sistemas bem específicos, uma vez que a grande maioria dos problemas requer pelo menos uma aproximação, principalmente quando relacionados com a Física Estatística.

O interesse, neste trabalho, está na Física do enovelamento de proteínas. Estudar-se-á um sistema que pode ser classificado como sistema complexo e certamente a opção pela utilização do computador como instrumento de trabalho é bem mais atraente. Portanto, este é claramente um problema que se enquadra dentro da Física Computacional.

A Física Computacional pode ser dividida em dois fortes ramos de pesquisa. De um lado, estudos referentes aos processos determinísticos, geralmente associados à resolução das equações de onda e equações de movimento de Newton, bem como de integradores. De outro, estudos relacionados à resolução de problemas que se enquadram como processos estocásticos, como é o caso do problema do enovelamento de proteínas que será estudado nesta dissertação. Assim, será inicialmente introduzido o método de Monte Carlo e discutido o algoritmo de Metropolis, talvez a técnica de simulação numérica mais utilizada nesta recente história das simulações. Logo após, será discutido o algoritmo de Wang-Landau, a técnica de simulação que será utilizada de fato nesta dissertação, para simulação da Mecânica Estatística de equilíbrio de uma proteína. No entanto, para facilitar a exposição dos algoritmos de Metropolis e Wang-Landau, será utilizado, como tradicionalmente é feito nos textos introdutórios, o modelo de Ising como o sistema em estudo.

2.2 O Método de Monte Carlo e o Algoritmo de Metropolis

O método de Monte Carlo é um método de cálculo que se destina a estudar problemas que se enquadram no ramo dos problemas estocásticos. Pode-se dizer de uma maneira simplificada que é dado o nome de método de Monte Carlo a qualquer simulação computacional que se valha da utilização de uma seqüência de números gerados aleatoriamente para simular o comportamento de um sistema físico qualquer. Tal nome foi atribuído por Metropolis, provavelmente referindo-se à capital mundial do jogo.

Para o estudo de transições de fase e fenômenos críticos, um dos principais ramos de pesquisa da Física da Matéria Condensada, a simulação computacional via Monte Carlo

mostrou ser uma ferramenta extremamente poderosa [25,33]. Boa parte dos trabalhos desenvolvidos nesta área durante o último meio-século se deve quase que exclusivamente ao algoritmo de Metropolis [32], um dos algoritmos de Monte Carlo de maior importância para a Física Computacional.

O método de Monte Carlo pode ser aplicado em dois grupos de sistemas: os estáticos (determinação de uma integral, etc...) e os dinâmicos (uma rede de spins, enovelamento de uma proteína). Entretanto, o método é mais intensamente aplicado no estudo de processos dinâmicos, onde num dado sistema, uma seqüência de estados sucessivos é estudada.

O problema estudado neste trabalho se enquadra no segundo grupo. Nas secções que se seguem, as propriedades e conseqüências de tais seqüências serão discutidas com mais detalhes, bem como sua relação com a simulação propriamente dita, onde se faz necessária a introdução do algoritmo de Metropolis, para posteriormente ser apresentado o método de fato utilizado durante a pesquisa, o algoritmo de Wang-Landau [29].

2.2.1 Equação Mestra e o Princípio do Balanço Detalhado

Como o método de Monte Carlo se enquadra como um processo estocástico, onde uma sucessão de estados é determinado numa cadeia de eventos de modo a construir um espaço amostral, é essencial definir-se as regras que regem esta dinâmica.

Primeiramente, pode-se dizer que o método de Monte Carlo é definido como uma cadeia Markoviana, ou seja, uma seqüência de eventos em que a probabilidade de ocorrência de qualquer evento depende apenas da probabilidade de ocorrência do evento anterior [34]. Portanto, uma questão que se mostra de grande relevância é a determinação da probabilidade de se encontrar um evento em um determinado estado. O conhecimento de tal probabilidade se mostra essencial para o desenvolvimento da dinâmica da simulação, pois a mesma será utilizada para determinar se a mudança de estado será efetuada ou não, na transição para a configuração seguinte.

Partindo-se de uma configuração inicial, o Método de Monte Carlo para um sistema dinâmico permite que novas configurações sejam elaboradas sucessivamente baseadas em seu estado anterior, ou seja, gera-se uma cadeia Markoviana. A seqüência que se permite construir

mediante tal método, segue um caminho no espaço das configurações em que o tempo da dinâmica do processo não é o tempo real e sim o “Tempo de Monte Carlo”. No caso de uma rede de spins, tal tempo será o tempo médio para que todos os possíveis sítios sejam visitados, sendo importante observar-se que o mesmo será discreto.

Uma vez que cada estado na cadeia Markoviana depende do anterior, numa dada seqüência de eventos, a probabilidade de se encontrar tal sistema em uma determinada configuração (i) também dependerá da probabilidade inicial de encontrar-se o sistema na mesma configuração. Outrossim, pode-se encontrar a nova probabilidade após um intervalo de tempo, somando-se a probabilidade do instante anterior com a variação da probabilidade (ΔP_i) durante este intervalo de tempo. Logo

$$P_i(t_0 + 1) = P_i(t_0) + \Delta P_i(t_0 \rightarrow t_0 + 1). \quad (2.1)$$

A variação de probabilidade pode ser encontrada somando-se todas as possíveis transições para o estado desejado e subtraindo-se as possíveis transições que deixam o mesmo estado. Tal dedução heurística leva à equação mestra para um tempo discreto

$$\Delta P_i(t_0 \rightarrow t_0 + 1) = \sum_j (W_{ji} P_j(t_0 + 1) - W_{ij} P_i(t_0 + 1)), \quad (2.2)$$

onde W_{ij} é a taxa de transição do estado i para o estado j por unidade de tempo. Para um tempo contínuo, a equação mestra é dada por

$$\frac{dP_i(t)}{dt} = \sum_j (W_{ji} P_j(t) - W_{ij} P_i(t)). \quad (2.3)$$

Entretanto, sempre que se queira investigar um sistema em equilíbrio térmico, deve-se considerar que não há variação temporal da probabilidade. Como consequência, tem-se que

$$\frac{\partial P}{\partial t} = 0 \quad (2.4)$$

e portanto,

$$W_{ji}P_j(t) = W_{ij}P_i(t). \quad (2.5)$$

Tal equação é conhecida como o princípio do balanço detalhado [34]. A importância deste princípio está no fato de ser ele um dos principais fatores a determinar se uma simulação pode ou não ser considerada de Monte Carlo.

2.2.2 O Algoritmo de Metropolis

Um grande problema que se pode encontrar ao gerar uma nova configuração é referente à probabilidade de transição. Para tanto, dever-se-ia conhecer todos os possíveis estados, para assim ser possível o cálculo do fator de normalização da probabilidade e, posteriormente, determinar a mudança ou não de estado. Entretanto, mesmo para um sistema pequeno, por exemplo, uma rede quadrada de spins (N^2) com 32X32 sítios, teria 2^{1024} configurações, cálculo que mesmo para os super computadores de hoje seria inviável. Fica assim aparentemente impossível determinar a escolha ou não da mudança.

Em 1953, Metropolis, Teller e Rosenbluth [32], quando estavam em Los Alamos, encontraram uma solução para o problema. Como a probabilidade de transição nada mais é do que a razão entre as probabilidades de estar no estado antigo (i) e no estado novo (j), como expresso na equação

$$P(i \rightarrow j) = \frac{P_j}{P_i}, \quad (2.6)$$

valendo-se do balanço detalhado adotaram que

$$\frac{P_i}{P_j} = \frac{W_{ij}}{W_{ji}} \Rightarrow P(i \rightarrow j) = \frac{W_{ij}}{W_{ji}}, \quad (2.7)$$

pois perceberam que não era preciso saber a probabilidade, mas apenas as taxas de transição.

Para uma cadeia de Markov, existem inúmeras possibilidades para o valor de W , cabendo escolher apenas algum que estivesse de acordo com a equação do balanço detalhado e com a ergodicidade (garantia de que todos os possíveis estados podem ser visitados e de que o sistema não ficará preso em um determinado estado). Utilizando então a distribuição de Boltzmann, escolheram valores definidos como

$$W_{ij} = \begin{cases} e^{-\frac{\Delta E}{kT}}, & \Delta E > 0 \\ 1, & \Delta E \leq 0 \end{cases}, \quad (2.8)$$

onde define-se a variação de energia como

$$\Delta E = E_j - E_i. \quad (2.9)$$

Portanto pode-se calcular a probabilidade de transição para ambos os casos. Para $E_j > E_i$, ou seja $\Delta E > 0$,

$$P(i \rightarrow j) = \frac{W_{ij}}{W_{ji}} = \frac{e^{-\frac{\Delta E}{kT}}}{1} = e^{-\frac{\Delta E}{kT}} \quad (2.10)$$

e para o caso em que $E_j \leq E_i$, tem-se

$$P(i \rightarrow j) = \frac{W_{ij}}{W_{ji}} = \frac{1}{e^{\frac{-\Delta E}{kT}}} = e^{\frac{\Delta E}{kT}}. \quad (2.11)$$

Entretanto, como a variação de energia é negativa, ter-se-ia uma probabilidade maior do que 1, o que significa que a transição com certeza será efetuada. Portanto, pode-se dizer que a probabilidade pode ser dada pela escolha do menor valor entre 1 e o valor encontrado acima, o que se expressa como

$$P(i \rightarrow j) = \min\left(1, e^{\frac{\Delta E}{kT}}\right). \quad (2.12)$$

Sendo assim, fica resolvido o problema sobre a probabilidade de transição, estando a solução de acordo com o princípio do balanço detalhado. Tal procedimento para a determinação da escolha ou não de um novo estado é conhecido como o Algoritmo de Metropolis.

Partindo de um estado inicial, e valendo-se da probabilidade de transição acima determinada, pode-se construir uma seqüência de estados e a partir da mesma estudar o comportamento do sistema e as suas propriedades termodinâmicas através do método da amostragem por importância.

2.2.3 Amostragem por Importância

Sempre que se utiliza o método de Monte Carlo existe interesse no cálculo de médias de algumas grandezas Físicas, como algumas flutuações. Para o cálculo e obtenção de qualquer propriedade termodinâmica de um sistema é utilizado o conceito de valor médio da Física Estatística. A média de uma determinada grandeza G é calculada através da equação

$$\langle G \rangle = \frac{\sum_i^N G_i e^{-\frac{E_i}{kT}}}{Z}, \quad (2.13)$$

onde E_i é a energia no estado i , k a constante de Boltzmann, T a temperatura, e Z o valor conhecido como função de partição, ou seja, a soma com peso sobre todos os possíveis estados do sistema e dado por

$$Z = \sum_k c_k e^{-\frac{E_k}{kT}}. \quad (2.14)$$

Entretanto, como visto anteriormente, mesmo para um sistema simples, é impossível que todos os estados acessíveis ao sistema sejam visitados. Para calcular então a média, o método de Monte Carlo vale-se do conceito de amostragem por importância já que o mesmo é um método para a obtenção de valores médios de uma dada grandeza [33]. A importância deste método está, não somente em obter resultados precisos de uma aproximação de um espaço amostral complexo de um caso analítico, mas também na sua aplicação para a obtenção de resultados convincentes de um todo, baseado no estudo de uma parte.

No estudo do comportamento de spins em sólidos reais seria exigida uma rede, por assim dizer, infinita. Entretanto, o mesmo não se faz possível para métodos computacionais, muito menos para um estudo analítico. Tomam-se então, redes finitas, donde pode-se obter resultados adequados e construir-se um espaço amostral. A partir desse espaço amostral local, pode-se então inferir as propriedades do sólido como um todo. O espaço amostral de uma determinada simulação, necessário para a amostragem por importância, é construído mediante uma sucessão de eventos transitando através de determinados estados, onde o número de vezes que cada estado é visitado será o peso deste estado.

Em um espaço amostral que obedece a distribuição de Boltzmann, tem-se que a probabilidade do sistema estar num estado i é dada por

$$P_i = \frac{c_i e^{-\frac{E_i}{kT}}}{Z}, \quad (2.15)$$

onde c_i é o peso da probabilidade e a função partição será a soma sobre todos os estados do espaço amostral construído

$$Z = \sum_k c_k e^{-\frac{E_k}{kT}}. \quad (2.16)$$

Portanto, segundo este método, o valor médio de uma grandeza G , pode ser determinado mediante

$$\langle G \rangle = \sum_i^N G_i P_i. \quad (2.17)$$

Partindo-se de um espaço amostral construído durante a simulação, pode-se, portanto encontrar qualquer valor médio, bem como algumas flutuações dos mesmos.

2.2.4 Aplicação do Algoritmo de Metropolis

Um sistema apropriado para se estudar com o método de Monte Carlo, bem como para se compreender melhor a aplicação do mesmo, é o sistema magnético descrito pelo modelo de Ising. A grande vantagem de tal escolha está em tratar-se de um problema resolvido e de ampla divulgação, sempre sendo usado como referência. Este algoritmo, além de fornecer inúmeros resultados, mostra-se adequado para uma demonstração do método.

Na tentativa de se explicar o ferromagnetismo em sólidos, o modelo pioneiro que permitiu explicar a magnetização espontânea foi o modelo de Heisenberg [35], muito

estudado ainda. O modelo de Ising é apenas um caso particular deste modelo mais geral, constituindo-se de uma rede de spins clássicos em contato com uma fonte térmica a temperatura constante, em que cada sítio da rede abriga um spin. Sendo um modelo de dois estados para spin 1/2, cada spin pode ter sua orientação para “cima” ou para “baixo”. Sobre o sistema, pode existir a influência de um campo magnético externo. Outro fator de interação será a energia de interação de cada spin com os seus vizinhos mais próximos. Esta energia é calculada pela equação

$$E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j, \quad (2.18)$$

onde o par $\langle i,j \rangle$ designa que a soma só será efetuada entre os vizinhos mais próximos e J representa um parâmetro positivo para o caso de se estudar alinhamento ferromagnético.

O momento magnético de cada sítio da rede será representado por σ , podendo assumir dois valores: $\sigma = +1$ (spin para “cima”, ou seja, paralelo ao campo) e $\sigma = -1$ (spin para “baixo”, ou seja, antiparalelo). Com base nessas informações a Hamiltoniana do problema de Ising na presença de um campo magnético externo B , é dada por

$$H = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - \mu B \sum_i \sigma_i. \quad (2.19)$$

A primeira parte da equação, como visto anteriormente, refere-se à energia de correlação entre os spins, enquanto que, a segunda parte, se refere à energia de interação entre o campo externo e cada um dos spins da rede; μ é o momento magnético do átomo já que cada sítio só pode ser ocupado por apenas um átomo. Para o caso bidimensional, pode-se calcular a energia de um dado sítio (i,j) , através de

$$E_{ij} = -J \sigma_{i,j} \cdot (\sigma_{i+1,j} + \sigma_{i-1,j} + \sigma_{i,j+1} + \sigma_{i,j-1}) - \mu B \sigma_{i,j}. \quad (2.20)$$

Entretanto, sabe-se que segundo o algoritmo de Metropolis, a Hamiltoniana não será necessária para o cálculo da probabilidade de transição, mas apenas a variação da mesma.

Como é tratado um sítio de cada vez, pode-se dizer que para cada iteração do programa, só será de interesse, o conhecimento da variação da energia do sítio em questão. Logo, define-se a variação da energia como

$$\Delta E = E_{i,j}^{novo} - E_{i,j}, \quad (2.21)$$

onde o $E_{i,j}^{novo}$ será a energia do sistema supondo que o spin do sítio em estudo mude seu sentido. O valor do mesmo pode ser encontrado segundo a equação

$$E_{i,j}^{novo} = -J\sigma_{i,j}^{novo} \cdot (\sigma_{i+1,j} + \sigma_{i-1,j} + \sigma_{i,j+1} + \sigma_{i,j-1}) - \mu B \sigma_{i,j}^{novo}. \quad (2.22)$$

Como

$$\sigma_{i,j}^{novo} = -\sigma_{i,j}, \quad (2.23)$$

tem-se

$$E_{i,j}^{novo} = +J\sigma_{i,j} \cdot (\sigma_{i+1,j} + \sigma_{i-1,j} + \sigma_{i,j+1} + \sigma_{i,j-1}) + \mu B \sigma_{i,j}. \quad (2.24)$$

Mas o lado direito da expressão é exatamente a energia de um spin com o sinal trocado. Logo encontra-se que

$$E_{i,j}^{novo} = -E_{i,j} \quad (2.25)$$

e portanto, a variação da energia fica definida simplesmente como

$$\Delta E = 2E_{i,j}^{novo}. \quad (2.26)$$

Com base nas informações descritas acima, fica possível o estabelecimento de uma seqüência de passos baseados no algoritmo de Metropolis, o que permitirá a resolução numérica do problema de Ising que consiste principalmente no cálculo de algumas médias termodinâmicas, bem como algumas flutuações (o que leva à susceptibilidade magnética e ao calor específico). As etapas para o cálculo dos valores desejados devem obedecer a ordem abaixo, que por uma razão didática, será apresentado na forma de um esboço do algoritmo:

- 1 Considere uma rede bidimensional com $N = L \times L$ sítios
- 2 Alinhe todos o momentos magnéticos para um mesmo sentido, caso todos sejam orientados para cima tem-se que $\sigma_{i,j} = +1$. (não é obrigatório utilizar essa configuração inicial de spins)
- 3 Através de um sorteio, um dos sítios da rede deve ser escolhido.
- 4 O sítio sorteado é submetido ao teste de Metropolis, ou seja, inverte-se o sentido do mesmo para permitir o calculo da variação de energia.
- 5 Se a variação de energia for negativa, vira-se o momento magnético.
- 6 Se a variação de energia for positiva, calcule a probabilidade de transição e submeta o evento “virar” a um sorteio. Se o número randômico for menor que o fator de Boltzmann, vire o momento. Caso contrário, deixa-se o momento como estava.
- 7 Em cada iteração, devem ser efetuados N sorteios, o que corresponde a um sorteio para cada sítio.

- 8 A condição de contorno deve ser observada com cuidado. No caso de uma condição de contorno periódica, deve-se assegurar que

para o índice i ,

$$\sigma_{L+1,j} = \sigma_{1,j}$$

e

$$\sigma_{0,j} = \sigma_{L,j}$$

para o índice j ,

$$\sigma_{i,L+1} = \sigma_{i,1}$$

e

$$\sigma_{i,0} = \sigma_{i,L}$$

- 9 Após N momentos terem sido submetidos ao sorteio, calcula-se a energia e a magnetização da configuração obtida. Tal procedimento deve ser repetido tantas vezes quanto se queira. Quanto mais efetuado, mais precisos serão os resultados.

- 10 Calcula-se então as médias termodinâmicas sobre todos os valores encontrados no passo anterior, onde é importante lembrar que essas médias são apenas condizentes com a temperatura da fonte de calor. Para nova temperatura, deve-se repetir todo o processo.

2.3 O Algoritmo de Wang-Landau

Recentemente, novos e eficientes algoritmos começaram a ser implementados, permitindo que as simulações atingissem resoluções cada vez melhores para a localização e caracterização das transições de fase. Para reduzir a desaceleração crítica, que ocorre próximo a transições de segunda ordem, o algoritmo flip-flop de aglomerados de spin, inicialmente desenvolvido por Swendsen e Wang [36], mostrou-se válido, bem como o método de Wolf [37], uma implementação do mesmo.

Outro grande problema que só pode ser superado com a implementação de outro método, o Multicanônico [30,31], foi à superação da barreira de tunelamento entre as fases coexistentes em transições de primeira ordem. Esta aproximação também se mostrou eficiente em sistemas com um vasto espectro de energia. É importante dizer que o método de Wang-Landau tem sua origem baseada no método Multicanônico. Atualmente, entretanto, o algoritmo de Wang-Landau começa a tomar corpo no meio científico devida a sua grande versatilidade e eficiência. Esses dois motivos, juntamente com outros que ainda serão citados, foi o que levou à escolha deste como o algoritmo principal a ser utilizado durante a pesquisa a ser apresentada nesta dissertação.

Dizer que o algoritmo de Wang-Landau nada mais é do que uma implementação para o método de Metropolis, consistindo a mesma basicamente na mudança das escolhas efetuadas no espaço da energia para o espaço da entropia, é não só uma inverdade, como também uma visão indevida sobre o método. Na verdade, o método de WL (Wang-Landau) é um novo algoritmo Monte Carlo que produz resultados altamente precisos e com uma grande redução do esforço computacional.

Basicamente, o método consiste em efetuar caminhos aleatórios no espaço das energias, dentro de um determinado intervalo estabelecido, enquanto a densidade de estados é continuamente reajustada até que se possa obter histogramas localmente planos (quando o histograma para todas as energias do intervalo estejam com o mesmo número de visitas, dentro de uma determinada porcentagem de erro). Desta maneira, é possível o acesso direto, tanto da entropia, como da energia livre, o que faz do método uma excelente ferramenta. Ele é mesmo independente da temperatura e extremamente eficiente para os estudos das transições de fase, tanto de primeira ordem, quanto de segunda ordem.

Outra grande vantagem deste método, além do acesso à entropia, é que também permite o cálculo de quantidades que não estejam diretamente relacionadas com a energia, pois, como foi visto, o caminho aleatório não está restrito ao espaço das energias. Sendo assim, tal algoritmo pode ser aplicado a qualquer outro espaço de parâmetros, bastando-se apenas o conhecimento da Hamiltoniana. O algoritmo então poderá ser aperfeiçoado para o cálculo da densidade de estados.

O algoritmo pode ainda, diferente de métodos Monte Carlo convencionais, gerar diretamente uma distribuição canônica a uma determinada temperatura através de

$$P(E, T) = \frac{1}{Z} g(E) e^{\frac{-E}{kT}}, \quad (2.27)$$

onde $g(E)$ é a densidade de estados, ou seja, o número de estados por intervalo de energia ΔE . A finalidade deste algoritmo está justamente em determinar $g(E)$ com precisão através da técnica do caminho aleatório que produz um histograma plano no espaço das energias.

A efetuação de múltiplos caminhos aleatórios pode fazer com que o método seja melhorado e resultados mais precisos possam ser obtidos. Tais caminhos podem ser efetuados para cada diferente intervalo de energia serialmente, ou seja, para todo o espectro de energia ou caminhos múltiplos podem ainda ser efetuados de maneira paralela, onde cada movimento do caminho será validado apenas em um determinado intervalo de energia, de forma que movimentos fora do intervalo são rejeitados. Em ambos os casos, paralelo e serial, os pedaços resultantes da densidade de estados podem ser juntados e introduzidos para produzir médias canônicas para o cálculo de quantidades termodinâmicas.

Uma grande vantagem do método é que a temperatura não precisa ficar limitada aos valores de uma fonte térmica constante, como ocorre nos métodos de Monte Carlo convencionais.

Para demonstrar a eficiência e precisão do método, o algoritmo será apresentado com o auxílio do modelo de Ising em uma rede quadrada bidimensional, tal como foi apresentado por Wang e Landau [29]. Tal problema mesmo que contendo transições de segunda ordem, ainda será válido, já que permitirá a comparação com os resultados obtidos anteriormente através do método de Metropolis.

É sabido por observação que quando se toma um caminho aleatório no espaço das energias, com a taxa de transição proporcional à densidade de estados, existe uma tendência de se gerar um histograma “plano” em tal espaço [29]. Então o algoritmo consiste em realizar modificações na densidade de estados de um modo sistemático para produzir um histograma “plano” sobre o intervalo de energia permitido. Simultaneamente, a densidade de estados converge para o valor real.

Ao iniciar o método, tal densidade de estados é ainda desconhecida, então se deve ajustar a densidade para todas as energias dentro do intervalo desejado, como sendo igual a 1; ($g(E)=1 \forall E \subset \text{Intervalo}$).

Como visto anteriormente para o caso da amostragem por importância, a função partição é calculada por

$$Z = \sum_k c_k e^{-\beta E_k} . \quad (2.28)$$

Sabe-se, porém, que o peso c_k , nada mais é do que a densidade de estados $g(E)$, que tem relação direta com a entropia $S(E)$ através de

$$S(E) = k_b \ln[g(E)] . \quad (2.29)$$

onde será adotado $k_b = 1$.

Tanto a densidade de estados como a entropia, são dadas em função da energia interna do sistema. A função partição pode ser reescrita como

$$Z = \sum_E g(E) e^{-\beta E} . \quad (2.30)$$

Tomando a equação acima em termos da entropia tem-se

$$Z = \sum_E e^{S(E) - \beta E} . \quad (2.31)$$

Com base em tal distribuição pode-se escolher as taxas de transição, assim como feito com o algoritmo de Metropolis. Entretanto, mesmo que os caminhos aleatórios sejam efetuados no espaço da energia, o referencial não será mais a variação desta, e sim a variação da entropia. Tal mudança acarreta o uso da amostragem entrópica no lugar da amostragem por importância, utilizada no método de Metropolis. A implementação de tal técnica se deve a

Jooyoung Lee [31], mesmo que Berg e Neubaus a tivessem intuído anteriormente no método do conjunto multicanônico [30], já que Lee foi o primeiro a introduzir a devida formalização teórica da mesma.

Para determinar as probabilidades de transição, serão introduzidas as escolhas

$$W_{ij} = \begin{cases} e^{-\Delta S}, & \Delta S > 0 \\ 1, & \Delta S \leq 0 \end{cases}, \quad (2.32)$$

onde define-se a variação da entropia como

$$\Delta S = S(E_j) - S(E_i). \quad (2.33)$$

Como visto anteriormente, a probabilidade de transição é dada por

$$P(i \rightarrow j) = \frac{W_{ij}}{W_{ji}}. \quad (2.34)$$

Portanto pode-se calcular a probabilidade de transição para ambos os casos. Para $\Delta S > 0$,

$$P(i \rightarrow j) = \frac{e^{-\Delta S}}{1} = e^{-S(E_j) + S(E_i)} = \frac{e^{S(E_i)}}{e^{S(E_j)}}. \quad (2.35)$$

Como é sabido que a entropia está relacionada com a densidade de estados segundo a equação 2.29, tem-se que a probabilidade de transição para o caso supracitado é dada por

$$P(i \rightarrow j) = \frac{g(E_i)}{g(E_j)}. \quad (2.36)$$

Já para o caso em que $\Delta S \leq 0$,

$$P(i \rightarrow j) = \frac{1}{e^{-\Delta S}} = \frac{1}{e^{S(E_j) - S(E_i)}} = \frac{e^{S(E_i)}}{e^{S(E_j)}}. \quad (2.37)$$

Entretanto, como a variação da entropia é dada por $\Delta S \leq 0$, tem-se que $g(E_j) \leq g(E_i)$, o que implica numa probabilidade maior ou igual a 1 e significa que a transição com certeza será efetuada. Portanto, pode-se dizer que a probabilidade de transição pode ser dada pela escolha do menor valor entre 1 e a razão encontrada na equação 2.27, o que pode ser expresso por

$$P(i \rightarrow j) = \min\left(1, \frac{g(E_i)}{g(E_j)}\right). \quad (2.38)$$

Esta será a probabilidade que determinará se o spin irá virar ou não. Entretanto, diferente do método de Metropolis, a taxa de transição será ajustada cada vez que um nível de energia E for visitado já que o método atualiza a correspondente densidade de estados multiplicando o valor existente por um fator de modificação f (fator que permite a correção interativa da densidade de estados), que deve ser maior do que 1, ou seja

$$g(E)^{(novo)} = g(E) \cdot f. \quad (2.39)$$

Inicialmente, o fator de modificação pode ser tão grande quanto $f_0 = 2.71828$, ou seja, $f = e$. Tal escolha inicial é importante pois permite que se alcance todos os possíveis níveis de energia muito depressa, mesmo para sistemas grandes. O programa continuará executando

caminhos aleatórios no espaço da energia e modificando a densidade de estados até que o histograma acumulado $H(E)$ esteja “plano”. Até este ponto, entretanto, não se pode perceber grande diferença entre este método e o Método Multicanônico. O que de fato faz este método um novo algoritmo, consiste na percepção que Wang e Landau [29] tiveram para determinar quantas vezes o histograma “plano” deveria ser encontrado e como deveria ser ajustado o fator de modificação, o que remete diretamente a precisão dos resultados obtidos.

Após alcançado o histograma “plano”, deve-se reduzir o fator de modificação, visando-se obter um ajuste mais fino na densidade de estados, seguindo a relação dada por

$$f_1 = \sqrt{f_0} . \quad (2.40)$$

A próxima etapa consiste em se reajustar o histograma em $H(E)=0$, para qualquer E dentro do intervalo. Um novo nível dos caminhos aleatórios será construído até se obter novamente um histograma “plano”. Entretanto, o fator de correção ao fim desta etapa será f_1 . Depois de alcançado o histograma “plano” como anteriormente, deve-se novamente zerar o histograma ($H(E)=0$) e reajustar o fator de modificação através de

$$f_2 = \sqrt{f_1} . \quad (2.41)$$

Tal condição leva a uma fórmula geral para o fator de modificação que será dada por

$$f_{i+1} = \sqrt{f_i} \quad (2.42)$$

onde os passos de MC necessários para determinar f podem aumentar a medida que é refinado o fator de modificação.

Uma pergunta que se pode fazer a esta altura é referente a quantidade de níveis de refinamento no fator de modificação que devem ser efetuados até que o valor gerado pela

densidade de estados seja devidamente preciso. O processo de simulação será finalizado quando o fator de modificação for menor que algum valor final pré-definido. Por exemplo,

$$f_{final} = \exp(10^{-8}) = 1,00000001. \quad (2.43)$$

pois a partir desse valor a correção na densidade de estados é desprezível.

O fator de modificação final terá como papel definir a precisão da densidade de estados final, o que por consequência está associado ao número de histogramas “planos” que devem ser encontrados. Para cada etapa, o fator de modificação terá relação direta com o número de iterações já que define a precisão necessária àquele momento para a densidade de estados.

Outro fator que tem relação direta com a precisão do programa é a determinação de quão plano o histograma deve ser, já que um histograma verdadeiramente plano é impossível de se obter. Em geral, se o histograma não possui, para qualquer valor de energia no intervalo desejado, um valor inferior a 80% do valor médio do histograma $\langle H(E) \rangle$, está se obtendo um resultado de alta precisão. Portanto, cabe a esta porcentagem, em torno do valor médio do histograma, uma segunda forma de controle da precisão do mesmo.

A densidade de estados, entretanto, só poderá ser considerada encontrada após o término da simulação. Mas vale lembrar que tal densidade é apenas relativa já que pode ser uma aproximação em escala da densidade de estados real. Para se obter uma densidade de estados adequada, deve-se conhecer o número de estados fundamentais de modo a normalizar a densidade de estados encontrada com o programa.

Para o modelo de Ising, sabe-se que o número de estados fundamentais é 2 (todos os spins para cima ou para baixo). Já para o caso em que múltiplos caminhos forem efetuados dentro de diferentes intervalos de energia, eles devem ser adequados aos limites de energia.

Um aparente problema consiste no fato de que, durante o caminho aleatório (especialmente para as iterações iniciais), o algoritmo não satisfaz o princípio do balanço detalhado já que a densidade de estados é constantemente modificada durante o caminho aleatório no espaço das energias. Porém, depois de muitas iterações, a densidade de estados

converge para o valor real muito depressa conforme o fator de modificação se aproxima de 1. Da Eq. (2.38), tem-se

$$\frac{1}{g(E_i)} p(i \rightarrow j) = \frac{1}{g(E_j)} p(j \rightarrow i), \quad (2.44)$$

onde $1/g(E_i)$ é a probabilidade do nível de energia E_i , enquanto que $p(i \rightarrow j)$ é a probabilidade de transição de E_i para E_j do caminho aleatório. Pode-se então concluir que o princípio do balanço detalhado é satisfeito alterando-se a proporção da precisão para $\ln(f)$ [29].

A convergência e precisão deste algoritmo podem ser testadas para um sistema com transição de segunda ordem. A rede quadrada de Ising $L \times L$ com acoplamento dos vizinhos mais próximos que é geralmente considerado como um banco de dados ideal para teorias novas [16] e algoritmos de simulação [7,12].

Serão apresentados a seguir os resultados obtidos por Wang e Landau [29] tanto para redes pequenas (para as quais os resultados exatos são acessíveis) como para uma rede em que $L=256$ (para a qual a enumeração exata é impossível). Na Figura abaixo, as densidades de estados calculadas são mostradas junto com os resultados exatos obtidos pelo método proposto por Beale [38].

Pela figura 2.1, nenhuma diferença é visível, motivo pelo qual é mostrado o erro relativo $\varepsilon(\log(g(E)))$. Pode-se definir o mesmo pela seguinte equação:

$$\varepsilon(X) = \left| \frac{X_{simulado} - X_{exato}}{X_{exato}} \right|, \quad (2.45)$$

onde X é uma grandeza geral qualquer. Em uma rede de 32×32 com 7×10^5 varreduras pode-se obter um erro médio tão pequeno quanto 0.035%.

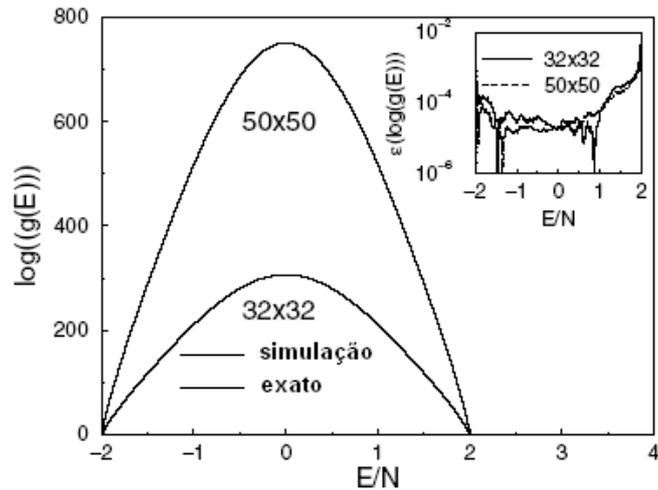


Figura 2.1 – Comparação da densidade de estados obtida pelo algoritmo em estudo para o modelo de Ising bidimensional e os resultados exatos calculados pelo método da Ref. [38]. Erros relativos $\varepsilon(\log(g(E)))$ são mostrados no *inset*. Figura extraída da Ref. [29].

Como dito anteriormente, uma vantagem deste algoritmo está em permitir prontamente o cálculo da energia livre de Gibbs e da entropia, quantidades que não são diretamente obtidas em simulações convencionais de Monte Carlo. Com a densidade de estados, a energia livre de Gibbs pode ser calculada por

$$F(T) = -kT \ln(Z) = -kT \ln\left(\sum_E g(E)e^{-\beta E}\right). \quad (2.46)$$

Na figura 2.2, são comparados os dados da simulação e as soluções exatas para a energia livre de Gibbs como uma função da temperatura. A concordância é excelente e um teste mais refinado da precisão mostra que o erro relativo $\varepsilon(F)$ é menor que 0.0008% para a região de temperatura T no intervalo $[0, 8]$.

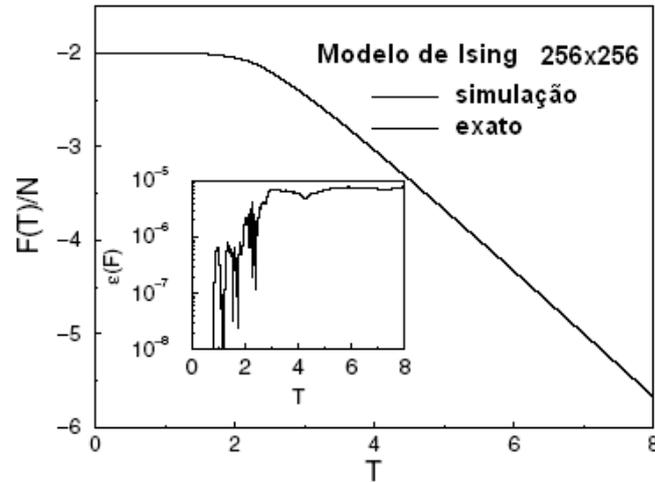


Figura 2.2 – Comparação da energia livre de Gibbs por sítio diretamente da densidade de estados da simulação para o modelo de Ising $L=256$ e as soluções exatas de Ref. [39]. Os erros relativos $\varepsilon(F)$ são mostrados no *inset*. A densidade de estados foi obtida através de caminhos aleatórios com apenas 6.1×10^6 passos de Monte Carlo. Figura extraída da Ref. [29].

Apesar da entropia ser uma das quantidades termodinâmicas de grande importância, não pode ser calculada diretamente em simulações Monte Carlo convencionais. Nesses métodos ela só pode ser calculada integrando-se sobre outras quantidades termodinâmicas, como, por exemplo, o calor específico. Porém tais cálculos não são tão confiáveis já que o calor específico em si não é tão facilmente calculado com precisão.

Com uma acurada densidade de estados calculada pelo método WL, a entropia pode ser calculada facilmente por

$$S(T) = \frac{U(T) - F(T)}{T} . \quad (2.47)$$

Onde a energia interna é dada por

$$U(T) = \langle E \rangle_T = \frac{\sum_E g(E) E e^{-\beta E}}{Z} . \quad (2.48)$$

Os erros encontrados numa rede quadrada de $L=256$ são menores que 1.2% em todas as regiões de temperatura. Um teste mais estrito da precisão da densidade de estados é o cálculo do calor específico definido pela expressão de flutuação

$$C(T) = \frac{\langle E^2 \rangle_T - \langle E \rangle_T^2}{T^2}. \quad (2.49)$$

A figura 2.3 mostra os dados numéricos na rede de tamanho-finito comparados com a solução exata obtida por Ferdinand e Fisher [39]. Um teste mais refinado da precisão é mostrado no *inset* da figura 2.3, que mostra o erro relativo $\varepsilon(C)$. O erro médio sobre todo o intervalo $T \in [0.4, 8]$ apenas usando um total de 6.1×10^6 varreduras de MC é de 0.39%. O erro relativo não é nem sequer maior que 4.5% mesmo com a delicada escala próxima de T_c .

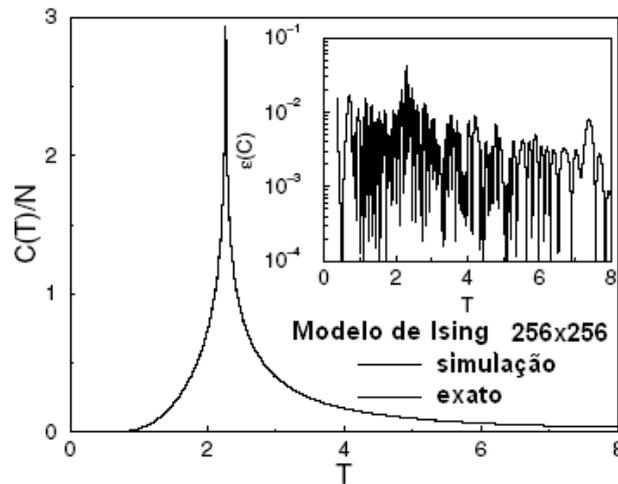


Figura 2.3 – Calor específico para o modelo de Ising bidimensional em uma rede de 256 x 256 em um largo intervalo de temperatura. O erro $\varepsilon(C)$ relativo é mostrado no *inset* da figura. Figura extraída da Ref. [29].

O algoritmo de Wang-Landau não só reduziu drasticamente o esforço computacional evitando simulações múltiplas para temperaturas diferentes perto da transição, mas também superou as cinéticas lentas a baixa temperatura ou perto de T_c para as transições de fase de primeira e segunda ordem já que o caminho aleatório não depende da temperatura.

Capítulo III

Enovelamento de Proteínas

3.1 Introdução

Neste capítulo será apresentada a definição de uma proteína e a sua composição básica, assim como as questões que tornam o enovelamento de proteína um problema fundamental. Além dos problemas relacionados à dinâmica de enovelamento, o estudo da Mecânica Estatística de equilíbrio de uma cadeia, tal qual será apresentada nesta dissertação, é ainda um aspecto importante na investigação do problema do enovelamento.

Com a finalidade de se compreender as propriedades de equilíbrio térmico serão descritas na seqüência, de forma detalhada e sistemática, as etapas básicas da implementação da técnica de simulação numérica. Por conveniência e para uma melhor apresentação didática, foi escolhida a introdução do método de simulação de Monte Carlo do enovelamento de proteínas, utilizando uma cadeia simplificada, como a de uma cadeia heteropolimérica em uma rede quadrada.

Do método de Monte Carlo serão explorados especificamente os algoritmos de Metropolis e de Wang-Landau discutidos no capítulo II. A idéia é elaborar um programa de computador para simular os movimentos de uma cadeia polimérica em contato com uma fonte de calor e, com ele, calcular grandezas termodinâmicas como a energia interna e o calor específico da cadeia em função da temperatura.

No caso da cadeia heteropolimérica citada, alguns resultados exatos podem ser obtidos e isso é bastante conveniente, pois pode-se testar o programa assim como os resultados numéricos fornecidos por ele. Inicialmente, será implementado o algoritmo de Metropolis, que neste trabalho de dissertação, apesar da sua grande importância para estudos em Mecânica Estatística, tem finalidades mais pedagógicas do que práticas. E finalmente, em uma etapa posterior, todo o procedimento poderá ser rapidamente adaptado para a implementação do algoritmo de Wang-Landau.

3.2 Proteínas

O objetivo nesta seção é apresentar apenas uma idéia do que é uma proteína, sua constituição física e algumas estruturas geométricas básicas. Não serão discutidas aqui suas funções biológicas ou detalhes sobre a química das proteínas. É apenas necessário o conhecimento básico que permita a elaboração de um modelo simples para mimetizar uma proteína.

3.2.1 Aminoácidos

Proteínas são cadeias poliméricas longas constituídas por combinações de aminoácidos. Uma cadeia pode conter de dezenas a milhares de aminoácidos. Entre muitas das suas funções vitais, pode-se citar o transporte e armazenamento de oxigênio, a de catalisação de reações bioquímicas e a de defesa do organismo, atuando como anticorpos.

São macromoléculas sintetizadas nas células seguindo instruções codificadas no DNA. Esse código, mesmo que já conhecido anteriormente, teve uma efetiva compreensão graças ao seqüenciamento do genoma no projeto Genoma Humano.

Os aminoácidos, por sua vez, são os blocos básicos com os quais as proteínas são sintetizadas. Eles são constituídos de um carbono central (C^α), denominado carbono alfa,

ligado a um grupo amina (NH), a um grupo carboxila (COOH), a um átomo de hidrogênio (H), e a um apêndice denominado grupo R, que diferencia um aminoácido do outro, em um grupo de 20 aminoácidos que ocorrem na natureza. Através deste grupo, pode-se classificar os aminoácidos segundo seu tamanho, carga, reatividade, e particularmente importante para esses estudos, segundo a sua hidrofobicidade ou afinidade pela água. A estrutura de um aminoácido é tetragonal como ilustrado na figura 3.1.

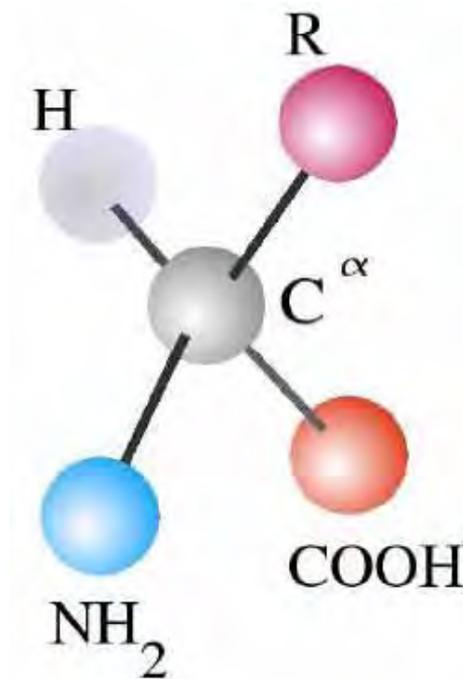


Figura 3.1 – Um aminoácido típico e sua estrutura tetragonal.

As proteínas são sintetizadas em forma de uma cadeia linear, a estrutura primária, constituída de aminoácidos atados por ligações peptídicas. São ligações covalentes entre os grupos α -carboxila de um aminoácido com o α -amina de outro aminoácido. A figura 3.2 ilustra a formação de uma ligação peptídica.

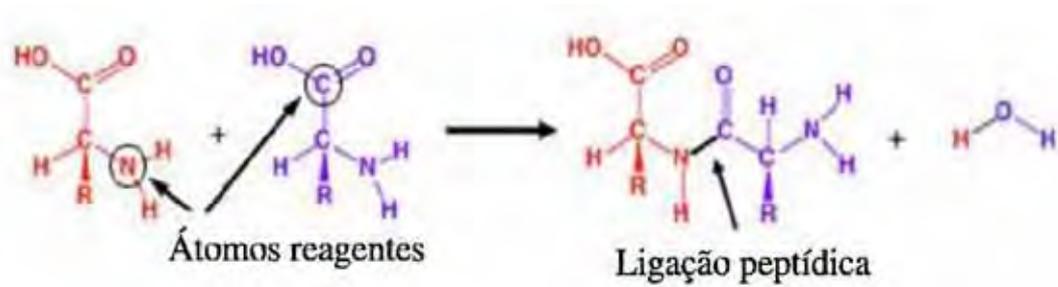


Figura 3.2 – Reação entre dois aminoácidos e formação de uma ligação peptídica.

Uma cadeia peptídica é uma série de aminoácidos ligados desta forma e cada aminoácido da cadeia é denominado resíduo. Cadeias contendo poucos resíduos são denominadas simplesmente peptídeos e as cadeias maiores, que possuem de 50 a 2000 resíduos, são denominadas proteínas. A cadeia formada pelas porções da proteína que se repetem com regularidade ao longo da cadeia é denominada cadeia principal ou esqueleto da proteína. A figura 3.3 mostra a seqüência formada por Leucina, Alanina e Glutamato em uma cadeia peptídica.

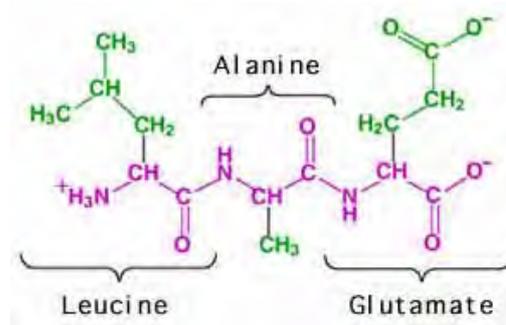


Figura 3.3 – Seqüência de aminoácidos em um peptídeo. A cadeia na cor púrpura é a cadeia principal e as de cor verde são as laterais

A estrutura mostrada na cor púrpura é a cadeia principal ou o esqueleto do peptídeo e as estruturas apresentadas na cor verde são as cadeias laterais. As proteínas são cadeias deste tipo, porém, contendo um grande número de aminoácidos e a cadeia principal pode ser encontrada em estruturas bastante complexas como as apresentadas na próxima seção.

3.2.2 Monômeros

Dentre as proteínas produzidas por seres vivos, de uma maneira geral, apenas 20 tipos diferentes de aminoácidos são encontrados. A estes aminoácidos mais comuns encontrados nas proteínas produzidas por seres vivos são denominados monômeros. Na tabela abaixo estão nomeados os monômeros, que de uma forma bem geral são divididos entre Hidrofóbicos (*H*) e Polares (*P*).

Tabela 3.1 – A tabela abaixo apresenta cada um dos 20 aminoácidos naturais (monômeros), bem como sua classificação (quarta coluna) entre polar (*P*) e hidrofóbico (*H*). Nas duas primeiras colunas, são apresentadas as abreviações mais conhecidas e usadas para cada um dos aminoácidos apresentados. Na terceira coluna estão os nomes em português de cada um dos monômeros.

Letra	Sigla	Nome	Tipo
A	Ala	Alanina	H
C	Cys	Cisteína	H
D	Asp	Ácido Aspártico	P
E	Glu	Ácido Glutâmico	P
F	Phe	Fenilalanina	H
G	Gly	Glicina	H ou P
H	His	Histidina	P
I	Ile	Isoleucina	H
K	Lys	Lisina	P
L	Leu	Leucina	H
M	Met	Metionina	H
N	Asn	Asparagina	P
P	Pro	Prolina	H
Q	Gln	Glutamina	P
R	Arg	Arginina	P
S	Ser	Serina	P
T	Thr	Treonina	P
V	Val	Valina	H
W	Trp	Triptofano	H
Y	Tyr	Tirosina	H

3.2.3 Estrutura de uma proteína

Uma cadeia de aminoácidos na forma linear é conhecida como a estrutura primária. Entretanto, a estrutura funcional de uma proteína, aquela na qual executa as suas funções biológicas, depende da seqüência de aminoácidos na cadeia e essa seqüência depende das instruções codificadas no DNA. A estrutura funcional é conhecida como estrutura nativa ou estado nativo. As proteínas executam suas funções biológicas quando encontram-se em sua estrutura nativa, uma estrutura compacta e com forma globular.

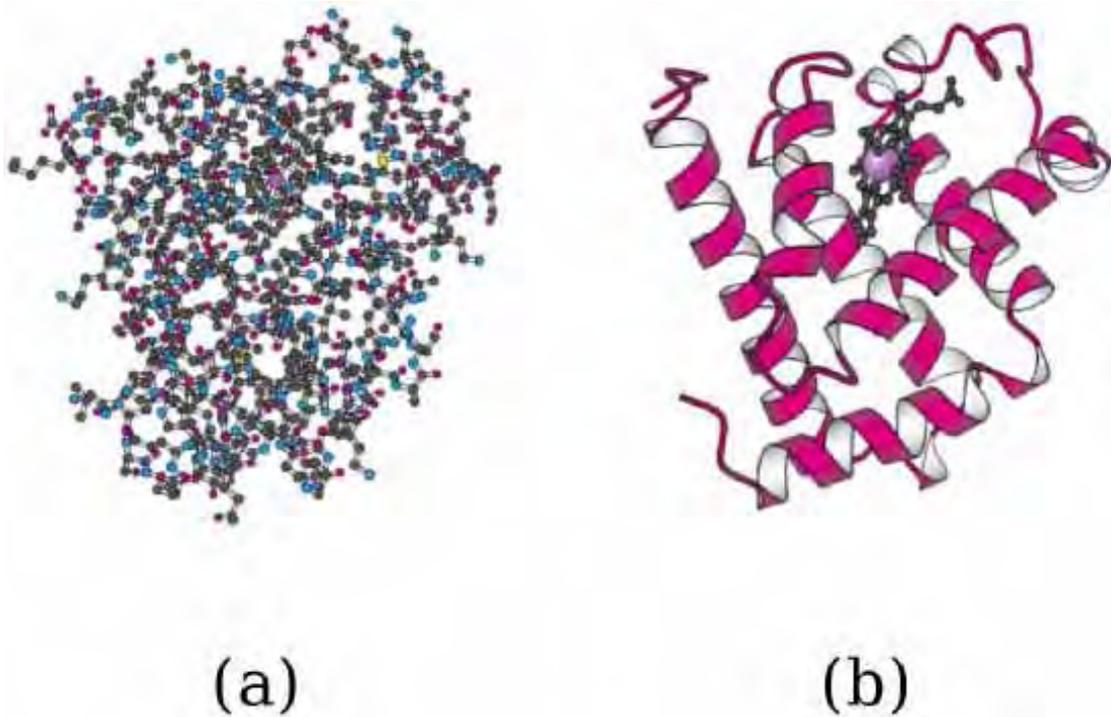


Figura 3.4 – A figura mostra duas representações da Mioglobina em sua estrutura terciária. Em (a) a proteína é apresentada por um modelo de bolas e palitos e, em (b.), ela é representada por um modelo de fitas.

A estrutura completa de uma proteína, como a do estado nativo, é denominada de estrutura terciária. A figura 3.4 mostra a Mioglobina através de duas representações. Na figura 3.4 (a) é apresentada através do modelo construído com bolas e palitos. Já em 3.4 (b) apresenta-se a Mioglobina através do modelo de fitas. Nesta, pode-se ver claramente várias cadeias helicoidais formando a estrutura terciária da proteína que aqui se encontra na forma

globular. Assim, a estrutura terciária pode ser constituída de estruturas intermediárias conhecidas como estrutura secundária como a estrutura α -hélice, mostrada na figura 3.5, uma estrutura secundária típica. A figura 3.4 (b) mostra a estrutura terciária da Mioglobina composta por várias estruturas α -hélice. Na figura 3.5 são mostradas três formas da representação da estrutura α -hélice. Em 3.5 (a) tem-se a representação por bolas e palitos. Em 3.5 (b) tem-se o esqueleto da estrutura α -hélice que se obtém ligando em seqüência os carbonos- α da cadeia. Na figura 3.5 (c) tem-se apresentação da mesma cadeia em um modelo de fitas.

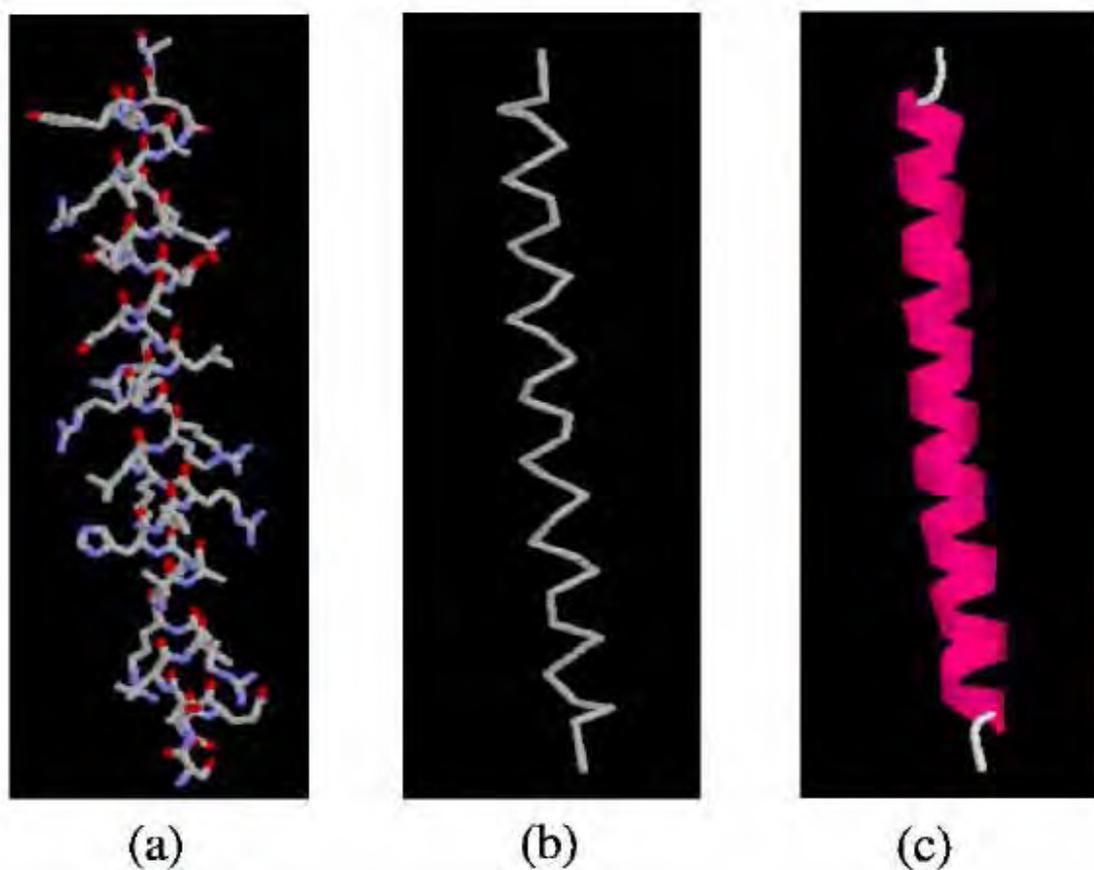


Figura 3.5 – Porção de uma cadeia helicoidal apresentada através de três representações. Em (a) a cadeia α -hélice é representada por um modelo de bolas e palitos. Em (b), a cadeia é representada por seu esqueleto, que se obtém ligando-se os carbonos-alpha da seqüência e por fim, em (c) um modelo da cadeia representada por fitas.

3.3 O Enovelamento de Proteínas

Como dito anteriormente, a proteína executa suas funções biológicas quando encontra-se em sua estrutura nativa, sendo caracterizada por ser compacta e única e, geralmente, imersa em solução aquosa com pH neutro. Entretanto, quando submetida a condições não fisiológicas como pH ácido ou básico, variação de temperatura ou ainda pela modificação do meio com a introdução de outros agentes, a proteína toma outra forma, uma cadeia mais aberta. Nestas condições ela não executa suas funções. No entanto, se as condições iniciais forem restabelecidas, pode voltar para a sua forma nativa, assim como retoma as suas atribuições biológicas.

Podem existir casos em que o processo seja não reversível. Este comportamento da proteína foi verificado experimentalmente por Afinsen [40]. Esta experiência levou a uma nova conjectura sobre as mesmas. A proteína passa de uma forma não nativa (cadeia aberta) para a estrutura nativa em um intervalo de tempo curto, da ordem de 1 segundo, o que leva à pergunta:

Como a proteína consegue em um intervalo tão curto de tempo encontrar seu estado nativo?

Esta é uma questão ainda não compreendida e aberta a novas contribuições. Esta questão ficou conhecida como o paradoxo de Levinthal [41]. O procedimento adotado pela proteína para se enovelar até atingir o estado nativo está codificado na seqüência de aminoácidos na cadeia da proteína e este código, por sua vez, é estabelecido pelo DNA. Esta correspondência entre a estrutura nativa e a seqüência de aminoácidos é conhecida também como o segundo código genético.

As investigações para se desvendar esta correspondência são intensas. Este desafio consiste, a grosso modo, em investigar quatro problemas básicos:

- dada uma seqüência de aminoácidos, determinar a estrutura nativa da cadeia, compreendendo ainda como esta estrutura permanece estável sob ponto de vista termodinâmico;
- dada uma estrutura nativa, determinar a seqüência de aminoácidos da cadeia, desenhando uma proteína que se enovele para um estado nativo conhecido

(este é o problema inverso do enovelamento de proteína e investigações nesta linha interessam bastante à indústria farmacêutica);

- dada uma seqüência, elucidar como uma proteína se enovela para o estado nativo em um intervalo de tempo tão curto e permanece estável;

- estudo do enovelamento de uma proteína assistida por outra proteína.

3.4 Modelos

Sob o ponto de vista teórico, alguns dos mecanismos básicos que levam uma proteína ao enovelamento foram compreendidos através de simulações numéricas utilizando modelos simplificados de proteínas. Modelos deste tipo são conhecidos na literatura como modelos na rede [12], quando a simulação ocorre em um espaço discreto representado por uma rede de pontos, e modelos fora da rede quando o espaço é contínuo. Trata-se de modelos onde uma cadeia de aminoácidos, às vezes convenientemente vista como uma cadeia polimérica homogênea ou heterogênea, é substituída por um sistema de pequenas esferas atadas umas às outras em seqüência por fios de comprimentos iguais. Nesta imitação de uma cadeia de aminoácidos, o fio que liga uma esfera à outra representa uma ligação covalente e cada esfera um aminoácido.

No modelo de rede, as esferas ocupam os pontos da rede mantendo o comprimento da cadeia constante. Para que esses modelos sejam caricaturas fiéis de proteínas ou de uma cadeia polimérica, deve-se introduzir os potenciais de interação entre as esferas, compatíveis com o comportamento observado. Os aminoácidos podem ser classificados como hidrofóbicos e polares e, conseqüentemente, quando em soluções aquosas a força dominante para o enovelamento é a hidrofóbica. Em soluções aquosas, os aminoácidos hidrofóbicos de uma cadeia tendem a se aglutinar formando um caroço enquanto que os aminoácidos polares tendem a formar uma capa externa envolvendo os hidrofóbicos [3]. As forças hidrofóbicas são então as responsáveis pelo enovelamento da proteína para este modelo. A idéia de se tratar

uma proteína como uma cadeia polimérica é bastante tentadora, pois sabe-se que essas cadeias sofrem um colapso, passando de uma conformação aberta para uma conformação compacta, tal como ocorre com uma proteína. Esta transição ocorre tanto com uma cadeia polimérica homogênea como em uma cadeia heterogênea. Adotando a cadeia polimérica como um modelo de proteína em uma rede, a energia de uma cadeia pode ser dada pela expressão

$$E = \sum_{i>j}^N \varepsilon_{\alpha_i, \alpha_j} \Delta(r_i - r_j), \quad (3.1)$$

onde N é o número de monômeros (aminoácidos), o índice i indica a posição de um monômero na seqüência (não na rede). O parâmetro $\varepsilon_{\alpha_i, \alpha_j}$ representa a energia de interação entre os monômeros α_i e α_j . Já a função $\Delta(r)$ é introduzida convenientemente para se representar energias de interação de curto alcance, como são os casos de forças hidrofóbicas. Assim

$$\Delta(r_i - r_j) = \begin{cases} 1 & \text{se } |r_i - r_j| = a \\ 0 & \text{para qualquer outro caso,} \end{cases} \quad (3.2)$$

onde a é o parâmetro de rede. Se $\Delta(r_i - r_j) = 1$, diz-se que o monômero α_i está em contato com o α_j .

Quando dois monômeros encontram-se um ao lado do outro, como neste caso, o contato entre eles é denominado contato nativo. Isto porque é o tipo de contato que ocorre entre dois monômeros quando a proteína se encontra em seu estado nativo, em uma forma compacta quando a grande maioria dos aminoácidos, senão todos, estão um ao lado do outro. Exclui-se desta categoria de contatos aqueles estabelecidos por dois monômeros subseqüentes na cadeia, unidos por uma ligação peptídica, pois não será alterado durante o processo.

Apesar de simples, modelos deste tipo permitem que a cadeia possa tomar formas em um espaço com um grande número de conformações possíveis, inclusive a de menor energia, que corresponde ao estado nativo e, portanto, possuem os ingredientes básicos que permitem descrever de maneira razoável o enovelamento de proteínas através da Mecânica Estatística [12,17,26]. Calculando grandezas termodinâmicas pode-se investigar as fases enoveladas

(ordenadas) ocorrendo a uma temperatura abaixo de uma temperatura crítica e a fase não nativa, ocorrendo acima desta temperatura. Modelos na rede ainda oferecem boas perspectivas para o estudo do enovelamento de proteínas [26].

3.5 Mecânica Estatística de Cadeias Poliméricas

As propriedades termodinâmicas de uma cadeia polimérica em contato com uma fonte de calor podem ser determinadas se todas as conformações possíveis da cadeia em uma rede de pontos forem conhecidas. Salvo alguns casos simples, essa contagem não pode ser feita analiticamente e, inevitavelmente, um computador precisará ser utilizado. O método mais comum para isso tem sido o método de Monte Carlo e, de uma forma mais freqüente, o algoritmo de Metropolis.

As transições de fase de uma cadeia, em que ocorre a passagem da cadeia de uma conformação não nativa para a nativa, tem sido compreendidas com sucesso através dessas simulações. Embora simulações como estas sejam já bastante conhecidas, a investigação do enovelamento de proteínas tem sido um campo fértil para aplicações de novos algoritmos para simulações numéricas. Pode-se citar vários exemplos desse tipo de iniciativa. Em 1995, Socci e Onuchic [42] aplicaram a técnica de Ferrenberg-Sweendsen [43]. Scheraga [44] utilizou o método Entropic Sampling [31]. Mais recentemente Wang [45] implementou o algoritmo flat-histogram [46] e Janke [47] implementou o método Multicanônico para modelos de rede.

Embora o método de Wang-Landau tenha sido aplicado em situações bem mais complexas [48], será implementado o método para estudar modelos de rede, situações bem mais simples, porém muito ilustrativas. Por conveniência, inicialmente será implementado o algoritmo de Metropolis para se estudar uma cadeia polimérica. Este procedimento torna a implementação do método de Wang-Landau mais simplificado.

Para esta discussão, será adotado um modelo simplificado de uma cadeia peptídica em uma rede quadrada conhecido como modelo HP (hidrofóbico-polar) introduzido por Dill [3]. Neste modelo, a cadeia é constituída por apenas dois tipos de monômeros; os hidrofóbicos e os polares. A energia desta cadeia pode ser obtida da expressão da Eq. 3.1. Como há somente dois tipos de monômeros, o parâmetro $\epsilon_{\alpha_i, \alpha_j}$ pode ser $\epsilon_{H,H}$, $\epsilon_{P,P}$ e $\epsilon_{H,P}$, que no modelo HP,

para priorizar as relações hidrofóbicas frente as interações P-P e H-P desprezíveis. Geralmente são escolhidos $\varepsilon_{H,H} = -1$ e $\varepsilon_{P,P} = \varepsilon_{H,P} = 0$.

Será considerada uma seqüência de $N=20$ monômeros, indexados de 1 a 20. Cada monômero da seqüência está atado a dois monômeros adjacentes dessa seqüência por uma ligação rígida, formando-se assim uma cadeia linear de monômeros de comprimento fixo. Os monômeros das extremidades fogem à regra e possuem apenas um vizinho ligado. A cadeia assemelha-se a um cordão com N nós regularmente espaçados. Cada monômero ocupa um sítio de uma rede quadrada, com parâmetro de rede $a = 1$. A figura 3.6 ilustra a esta cadeia heteropolimérica.

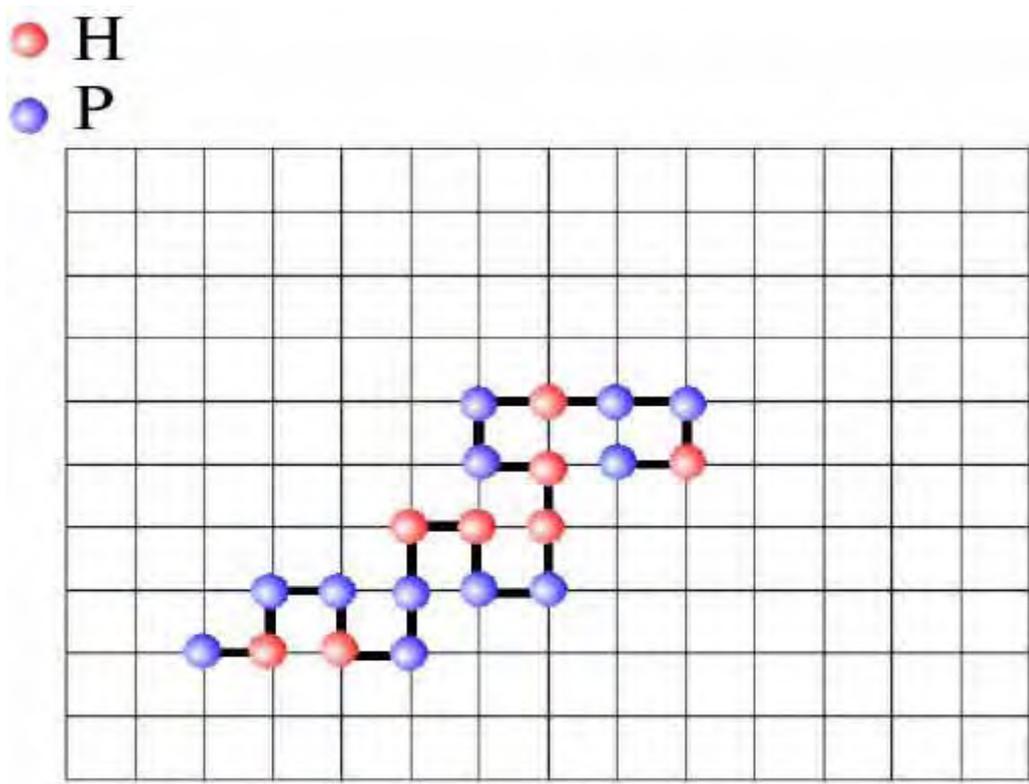


Figura 3.6 – Uma cadeia de monômeros segundo o modelo HP. As linhas pretas grossas representam ligações peptídicas.

3.5.1 O Algoritmo de Metropolis

As propriedades termodinâmicas desta cadeia podem ser calculadas se for possível a construção do espaço dos estados acessíveis à cadeia. O algoritmo de Metropolis permite que se construa sistematicamente todas as conformações da cadeia.

Considere-se uma conformação genérica da cadeia, como a apresentada na figura 3.6, por exemplo. Identifique as posições de cada monômero x_i, y_i e defina os vetores posição $\mathbf{r}_i = x_i\hat{x} + y_i\hat{y}$, para $i=1,2,\dots,N$. Uma nova conformação é gerada, a partir desta, modificando a posição de um ou um conjunto de monômeros desde que a cadeia conserve o seu comprimento e os sítios da rede não sejam duplamente ocupados. Com esse procedimento, é possível construir todas as conformações da cadeia.

Embora um conjunto de três ou mais monômeros possam ser deslocados simultaneamente como um grupo ou uma sub-cadeia para gerar uma nova conformação, há evidências concretas, como será visto, de que basta considerar a possibilidade de deslocamento simultâneo de um grupo de no máximo 2 monômeros para gerar conformações estatisticamente relevantes, isto é, que possuem probabilidades relevantes no cálculo de médias termodinâmicas.

Para gerar essas conformações, os movimentos dos monômeros serão realizados segundo os critérios abaixo.

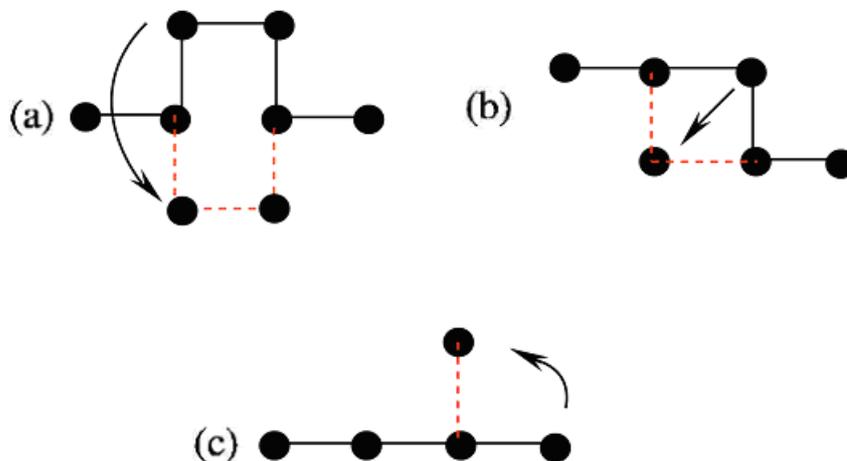


Figura 3.7 – A figura representa os possíveis movimentos de uma cadeia. Movimento de manivela (a), movimento de canto (b) e movimento de ponta (c).

1. Monômeros situados nas extremidades da cadeia podem realizar movimentos onde se desloca $1/4$ de arco de circunferência de raio unitário e centrado na posição de seu vizinho adjacente. Pode-se dizer, para simplificar, que eles podem realizar movimentos de 90° (figura 3.7 (c)).

2. Monômeros situados em porções retilíneas não conseguem fazer qualquer movimento, pois o comprimento da cadeia deve ser conservado.

3. Monômeros situados em porções da cadeia que formam vértice, doravante monômeros de cantos, podem realizar movimentos em que se desloca ao longo de uma diagonal que liga um canto ao canto oposto de um quadrado do reticulado. O monômero realiza um movimento que será denominado de movimento de canto, conforme ilustrado na figura 3.7 (b).

4. Monômeros situados em uma sub-cadeia com dois monômeros, que fazem parte de uma sub-cadeia maior com a forma de uma manivela, realizam movimentos de manivela. Em uma cadeia no plano, o movimento é de rotação de 180° . Note-se que neste tipo de movimento, dois monômeros são deslocados simultaneamente (figura 3.7 (a)).

Além destas condições, os movimentos serão realizados segundo um sorteio estabelecido pelo algoritmo de Metropolis. Os movimentos dependem também da temperatura do meio em que a cadeia se encontra. Imagina-se que em temperaturas altas a probabilidade da cadeia estar em uma conformação de cadeia aberta é bem maior que a probabilidade de encontrá-la em uma conformação compacta. Se o caso for de baixas temperaturas, as chances de se encontrar a cadeia em uma conformação compacta é bem maior. O algoritmo de Metropolis irá monitorar esses movimentos, onde as energias das conformações serão dadas pela equação 3.1.

Considere-se que a cadeia esteja em equilíbrio em um banho térmico a temperatura T . Os movimentos são feitos através de um sorteio via algoritmo de Metropolis. Ao ser realizado o movimento de um monômero, a cadeia passa de uma conformação de energia E_i para uma final de energia E_f com variação de energia ΔE . No equilíbrio, a probabilidade de ocorrência dessa transição é

$$p = \min\left(1, e^{-\frac{\Delta E}{kT}}\right). \quad (3.3)$$

Uma nova conformação para a cadeia pode ser gerada seguindo sistematicamente o seguinte procedimento:

- Sorteio de um monômero. Verificar se o monômero é um monômero em uma manivela, um canto fora da manivela, ou do extremo da cadeia.
- Se forem monômeros das pontas, a identificação é trivial: ele tem coordenada \mathbf{r}_1 ou \mathbf{r}_N . Considere-se inicialmente o caso \mathbf{r}_1 . Para se saber as possíveis novas posições do monômero, introduza-se o vetor unitário auxiliar $\hat{r} = \mathbf{r}_2 - \mathbf{r}_1$. Caso \hat{r} seja \hat{x} ou $-\hat{x}$, então o monômero pode ser deslocado para $\mathbf{r}_1^{novo} = \mathbf{r}_1 + \hat{r} \pm \hat{y}$. Caso \hat{r} seja \hat{y} ou $-\hat{y}$, então $\mathbf{r}_1^{novo} = \mathbf{r}_1 + \hat{r} \pm \hat{x}$. Para o caso do monômero em \mathbf{r}_N , basta considerar o mesmo procedimento. Defina-se o vetor unitário $\hat{r} = \mathbf{r}_N - \mathbf{r}_{N-1}$. Caso \hat{r} seja \hat{x} ou $-\hat{x}$, a nova posição do monômero é dada por $\mathbf{r}_N^{novo} = \mathbf{r}_{N-1} \pm \hat{y}$. Caso \hat{r} seja \hat{y} ou $-\hat{y}$, a nova posição do monômero é dada por $\mathbf{r}_N^{novo} = \mathbf{r}_{N-1} \pm \hat{x}$.
- Para verificar se o monômero é de canto, pode-se definir um critério simples: seja o monômero sorteado o de índice i , situado em \mathbf{r}_i . Os monômeros vizinhos estão situados em \mathbf{r}_{i-1} e \mathbf{r}_{i+1} . Se o produto escalar

$$(r_i - r_{i-1}) \cdot (r_{i+1} - r_i) = 0 \quad (3.4)$$

então o monômero encontra-se em um canto. Se não se anular, então trata-se de um monômero em uma porção retilínea da cadeia. Se o monômero for de canto, ele pode se mover, através de um sorteio, para o ponto

$$r_i^{novo} = r_i + \Delta r_i, \quad (3.5)$$

onde

$$\Delta r_i = (r_{i+1} - r_i) - (r_i - r_{i-1}) = r_{i+1} - 2r_i + r_{i-1}. \quad (3.6)$$

Portanto, se não estiver ocupada por outro monômero, a nova posição do monômero será

$$r_i^{novo} = r_{i+1} - r_i + r_{i-1}. \quad (3.7)$$

Uma vez efetuado o movimento, deve-se fazer $\mathbf{r}_i = 0$ para indicar que o sítio está vazio a partir desse momento. Entretanto, se esse sítio já estiver ocupado, é preciso verificar se ele está ocupado por um monômero de índice $i + 2$ ou $i - 2$. Se isso ocorrer, o monômero i é um monômero de canto e faz parte de uma sub-cadeia da forma de manivela. Assim, o movimento desse monômero deve ser o de manivela.

- Para se identificar um monômero de índice i que faz parte de uma sub-cadeia da forma de manivela, é preciso primeiro identificar se ele é um monômero de canto, como foi feito no item anterior. Logo após, é preciso verificar também se a posição especificada pelo vetor

$$s = r_{i+1} - r_i + r_{i-1} \quad (3.8)$$

que indica a posição, ao longo da linha diagonal, do sítio oposto ao vértice ocupado pelo monômero i , está ocupada por um monômero de índice $i + 2$ ou $i - 2$. Se não ocorre nenhum desses casos, então o monômero é um monômero de canto apenas. Entretanto, se a posição indicada por \mathcal{K} estiver ocupada pelo monômero $i + 2$ ou $i - 2$, então trata-se de monômero em uma sub-cadeia do tipo manivela. Esses monômeros ocorrem aos pares. Se ocorrer o monômero $i - 2$, então o parceiro do monômero i é o monômero $i - 1$. Caso o monômero $i + 2$ esteja ocupando o sítio indicado por \mathcal{K} , então o parceiro do monômero i é o monômero $i + 1$. Esses pares de monômeros movem-se segundo o movimento de manivela. Se a cadeia se encontra em um plano, o movimento do par i e $i + 1$ (ou $i - 1$ e i) consiste de uma rotação de

180 em torno de um eixo imaginário que passa pelos sítios localizados em \mathbf{r}_{i-1} e \mathbf{r}_{i+2} (ou \mathbf{r}_{i-2} e \mathbf{r}_{i+1}). Nesse caso, as novas posições do par i e $i+1$ serão

$$r_i^{novo} = r_{i+2} - r_{i+1} + r_{i-1} \quad (3.9)$$

e

$$r_{i+1}^{novo} = 2r_{i+2} - r_{i+1} \cdot \quad (3.10)$$

Caso o par de monômero for $i-1$ e i , suas novas posições serão

$$r_{i-1}^{novo} = r_{i+1} - r_i + r_{i-2} \quad (3.11)$$

e

$$r_i^{novo} = 2r_{i+1} - r_i \cdot \quad (3.12)$$

Uma vez efetuado o movimento, deve-se fazer $\mathbf{r}_i = 0$ e $\mathbf{r}_{i+1} = 0$ se o par for i e $i+1$ e $\mathbf{r}_{i-1} = 0$ e $\mathbf{r}_i = 0$ se o par for $i-1$ e i para indicar que esses sítios estarão vazios a partir desse momento.

Esse procedimento deve ser repetido à exaustão até que a quantidade de conformações geradas seja suficiente para que seja possível calcular médias termodinâmicas com boa precisão. Para o cálculo dessas médias a uma dada temperatura T , será registrado o número de vezes que uma conformação de energia E_k foi visitada e calculado

$$\langle G \rangle = \frac{\sum_k G(E_k) e^{-\frac{E_k}{kT}}}{\sum_k e^{-\frac{E_k}{kT}}}, \quad (3.13)$$

onde G é a grandeza física a ser calculada.

3.5.2 O Algoritmo de Wang-Landau

A implementação do método de Wang-Landau pode ser feita repetindo todo o procedimento adotado para se executar o movimento de um monômero, exceto que agora o sorteio para se fazer isso depende da variação da entropia. A probabilidade agora é dada por

$$P = \min(1, e^{-\Delta S}), \quad (3.14)$$

onde

$$\Delta S = S(E_j) - S(E_i) \quad (3.15)$$

é variação da entropia quando uma conformação da cadeia com energia E_i passa para outra conformação com energia E_j . Se a mudança da conformação diminui a entropia, aceita-se o movimento do monômero que alterou a conformação. Caso contrário, faz-se um sorteio para se decidir se o movimento da cadeia é aceito ou não. Toda vez que uma conformação seja visitada, a entropia é corrigida através de

$$S(E) = S(E) + \ln f \quad (3.16)$$

onde E é a energia da conformação e f é um número que controla a correção da entropia. O método ainda requer que se calcule paralelamente o histograma

$$H(E) = H(E) + 1, \quad (3.17)$$

que registra o número de vezes que o estado de energia E é visitado durante a simulação. Com esse critério de correção gradativa da entropia e visitado todas as conformações possíveis da cadeia, em dado momento, ocorrerá a convergência do histograma para um valor constante, isto é

$$H(E) \rightarrow \text{constante, para todo } E, \quad (3.18)$$

ou em outras palavras, um histograma plano. Significa que todas as energias E possíveis para a cadeia foram igualmente visitadas. Quando isso ocorrer, é atribuído um novo valor para f , por exemplo

$$f \rightarrow \sqrt{f} \quad (3.19)$$

e todo o procedimento é repetido novamente. O histograma deve ser zerado antes de se recommençar as novas iterações. A simulação termina quando f atingir um valor tal que (ver eq. 2.43)

$$\ln f = 10^{-8} \quad (3.20)$$

e a correção da entropia seja desprezível. A quantidade $S(E)$ obtida é utilizada para se obter

$$g(E) = e^{S(E)}, \quad (3.21)$$

que fornece a densidade de estados acessíveis à cadeia polimérica.

Quando o método é aplicado ao modelo de Ising, é conhecido por outro meio o número total de estados 2^N . Com uma normalização apropriada, $g(E)$ de fato fornece a densidade de estados absoluta. No caso de uma cadeia polimérica, salvo alguns casos em que se conheça a degenerescência do estado fundamental, não é conhecido o número total de conformações possíveis da cadeia e, conseqüentemente, $g(E)$ não será a densidade de estados de fato mas $c \times g(E)$, onde c independe da energia. No entanto, o fator c pode ser fatorado e eliminado sem prejuízo para o cálculo das médias termodinâmicas.

Finalizando a seção, deve-se deixar claro que, embora tenha sido adotado um modelo simplificado e em uma rede quadrada para expor o método, ele pode ser aplicado exatamente da mesma forma para qualquer outro modelo na rede e em casos de rede cúbica como será efetuado no próximo capítulo. A única diferença é que se tem mais graus de liberdade para o movimento dos monômeros.

3.5.3 Condições de Contorno

Para efeitos práticos, convém ainda ressaltar que, a medida que os movimentos dos monômeros são realizados, a cadeia polimérica se locomove no espaço. Em determinado instante atinge as bordas da caixa dentro da qual serão geradas as conformações possíveis. Para evitar problemas, como uma possível alteração da distribuição de probabilidades, é conveniente utilizar as condições de contorno periódicas. Quando um monômero atinge a borda da caixa, deixe o monômero sair, mas faça o entrar novamente na caixa pelo lado oposto dessa caixa. Uma outra forma de evitar esse problema, é transladar toda a cadeia para o centro da caixa quando um monômero for tocar uma borda da caixa. Assim, não há necessidade de se trabalhar com uma caixa grande, o que na prática pode significar gasto em tempo de processamento.

3.5.4 Grandezas Termodinâmicas

Para se investigar uma cadeia polimérica, algumas grandezas termodinâmicas são fundamentais. Neste trabalho, será calculado um conjunto de grandezas e com essas tentar obter o maior número de informações importantes para analisar o enovelamento de proteínas. Além das grandezas usuais como energia interna $U(T)$ e o calor específico, pode-se calcular a energia livre de Helmholtz $F(T)$ e a entropia $S(T)$. Pode-se ainda introduzir outras quantidades que seriam equivalentes aos parâmetros de ordem, o que auxiliará na identificação das fases enoveladas e compactas e fases não nativas. O número de contatos nativos $Q(T)$ em uma cadeia pode ser bastante útil para verificar o quão compacto pode ser um estado fundamental de uma cadeia polimérica. Uma outra quantidade importante é o raio de giração de uma cadeia $R_g(T)$. Esta quantidade dá uma idéia do tamanho médio do aglomerado formado pelos monômeros quando a cadeia se encontra em uma forma compacta ou quando se encontra em uma conformação não nativa a uma temperatura maior. Outra quantidade útil é a distância média $R_{ee}(T)$ entre os monômeros das extremidades da cadeia. Eventualmente, as derivadas do raio de giração ou da distância entre as pontas da cadeia pode auxiliar a definir melhor ocorrências de alguma transição de fase da cadeia.

A função de partição é calculada numericamente através de

$$\tilde{Z} = \sum_E e^{\frac{\tilde{S}(E)-E}{kT}} = e^{\frac{\tilde{S}(E_0)}{kT}} \sum_E e^{\frac{S(E)-E}{kT}}, \quad (3.22)$$

onde $S(E) = \tilde{S}(E) - \tilde{S}(E_0)$ e E_0 é a energia do estado fundamental. O valor médio de uma grandeza termodinâmica será obtida de

$$\langle O \rangle(T) = \frac{\sum_E c \times g(E) O(E) e^{-\frac{E}{kT}}}{\sum_E c \times g(E) e^{-\frac{E}{kT}}} = \frac{\sum_E g(E) O(E) e^{-\frac{E}{kT}}}{\sum_E g(E) e^{-\frac{E}{kT}}} \quad (3.23)$$

onde $g(E) = e^{S(E)}$ e $c = e^{\tilde{S}(E_0)}$. A função de partição utilizada será

$$Z = \sum_E g(E) e^{-\frac{E}{kT}}. \quad (3.24)$$

Portanto, pode-se obter a energia interna através de

$$U(T) = \frac{\sum_E g(E) E e^{-\frac{E}{kT}}}{\sum_E g(E) e^{-\frac{E}{kT}}}, \quad (3.25)$$

e o calor específico através de

$$C_v = \frac{\partial U}{\partial T} = \frac{1}{T^2} (\langle E^2 \rangle - \langle E \rangle^2). \quad (3.26)$$

Conhecendo $g(E)$ tem-se

$$F(T) = -kT \ln \sum_E g(E) e^{-\frac{E}{kT}}. \quad (3.27)$$

Logo, é calculada a entropia por meio da expressão

$$S(T) = \frac{1}{T} (U - F(T)). \quad (3.28)$$

Já o número de contatos nativos é calculado via

$$Q(T) = \frac{\sum_E g(E)Q(E)e^{-\frac{E}{kT}}}{\sum_E g(E)e^{-\frac{E}{kT}}}, \quad (3.29)$$

onde $Q(E)$ pode ser obtido da média

$$Q(E) = \sum_k \frac{\Omega_k^{(Q)}}{H(E)}. \quad (3.30)$$

A quantidade $\Omega_k^{(Q)}(E)$ é o número de diferentes valores de Q obtidos quando o estado de energia E foi visitado $H(E)$ vezes. Esta média é necessária, pois pode ocorrer que diferentes conformações possam ter a mesma energia, mas não possuam o mesmo número de contatos nativos. Médias como essa podem ser calculadas separadamente desde que $S(E)$ tenha sido calculada com boa precisão.

A distância média entre as pontas da cadeia pode ser calculada através de

$$R_{ee} = \frac{\sum_E g(E)r_{ee}(E)e^{-\frac{E}{kT}}}{\sum_E g(E)e^{-\frac{E}{kT}}}, \quad (3.31)$$

onde $r_{ee} = |r_N - r_1|$ e $r_{ee}(E)$ é dada pela média

$$r_{ee}(E) = \sum_k \frac{\Omega_k^{(r)}}{H(E)}, \quad (3.32)$$

onde $\Omega_k^{(r)}(E)$ é o número de diferentes valores de r_{ee} obtidos quando o estado de energia E foi visitado $H(E)$ vezes. Já o raio de giração é obtido através de

$$R_g = \frac{\sum_E g(E) r_g(E) e^{-\frac{E}{kT}}}{\sum_E g(E) e^{-\frac{E}{kT}}}, \quad (3.33)$$

onde

$$r_g = \left\{ \frac{1}{N} \sum_i (r_i - r_0)^2 \right\} \quad (3.34)$$

e

$$r_0 = \frac{1}{N} \sum_j r_j \quad (3.35)$$

é o centro de massa da cadeia. Note que r_g também pode ter vários valores para uma dada energia E .

3.6 Resultados Exatos

Após esta apresentação do método de Wang-Landau, é possível a título de ilustração, comparar os resultados obtidos numericamente utilizando o método com os resultados exatos conhecidos na literatura [12]. Será considerado o modelo HP com uma seqüência de 20 monômeros em uma rede quadrada, como a cadeia apresentada na Figura 3.6. A seqüência em questão é PHPPHPPHHPPHHPPHPPHP. As energias das conformações são dadas por

$$H = \sum_{i>j}^N \varepsilon_{\mu_i, \mu_j} \Delta(r_i - r_j), \quad (3.36)$$

onde $\Delta(r_i - r_j) = 1$ caso r_i e r_j sejam sítios adjacentes na rede, mas não sítios adjacentes da seqüência, isto é, sítios atados pelas ligações rígidas. Os índices μ_i representam os diferentes monômeros, podendo ser $\mu_i = H$ ou $\mu_i = P$ e o parâmetro $\varepsilon_{\mu_i, \mu_j}$ pode ser $\varepsilon_{H,H}$, $\varepsilon_{P,P}$, ou $\varepsilon_{H,P}$. Para qualquer outro caso, $\Delta(r_i - r_j) = 0$. Neste exemplo em particular, $\varepsilon_{H,H} = 1$ e $\varepsilon_{P,P} = \varepsilon_{H,P} = 0$. Essas energias são dadas em unidades do parâmetro J , que equivale a aproximadamente $0,6 \text{ kcal/mol}$.

No caso desta seqüência, a densidade de estados é conhecida exatamente [12] de modo que as propriedades termodinâmicas podem ser calculadas facilmente.

Tabela 3.2 – Tabela que apresenta na primeira linha, os valores exatos da seqüência em estudo, enquanto na segunda, traz as respectivas densidades de estado.

E	-8	-7	-6	-5	-4	-3	-2	-1	0
g(E)	1	6	457	5766	82065	779973	4526737	15348238	21146335

Com os dados da Tabela 3.2, pode-se calcular a entropia exatamente através de

$$S = k_b \ln g(E). \quad (3.37)$$

Por conveniência, será considerado $k_b = 1$. Será calculada também a entropia numericamente para essa seqüência, pelo método de Wang-Landau onde o resultado, representado por $S(E)$, é apresentado na Tabela 3.3, comparando-o com a entropia $S_e(E)$ exata.

Tabela 3.3 – Tabela que apresenta na segunda e na terceira linha as entropias exata e obtida numericamente em função da energia E .

E	-8	-7	-6	-5	-4	-3	-2	-1	0
$S_e(E)$	0.0	1.80	6.12	8.65	11.31	13.56	15.32	16.54	16.86
$S(E)$	0.0	1.80	6.10	8.63	11.30	13.55	15.31	16.53	16.85

Para cadeias pequenas como essa, os resultados são muito bons; os resultados numéricos são praticamente exatos. Este fato pode ser confirmado com o cálculo de algumas grandezas termodinâmicas e os resultados numéricos e exatos são comparados nos gráficos da figura. As curvas contínuas são os resultados exatos e os pontos representam os resultados numéricos.

Na figura 3.8, o gráfico (a) permite identificar bem a região ordenada e desordenada. Como a energia é obtida com a contagem dos contatos H-H (excetuando as ligações peptídicas) nativas, ter-se-á uma fase ordenada e a cadeia enovelada quando a energia interna é mínima. A estas temperaturas, o número de contatos nativos H-H é máximo, pois a cadeia encontra-se enovelada. Quando a energia é próxima de zero, tem-se a região desordenada. O gráfico da entropia (item (d) da figura 3.8) também mostra claramente estes estados de ordem. No item (b) se tem o calor específico da cadeia. O pico de calor específico permite avaliar a temperatura crítica de transição que, no próximo capítulo, será identificado como uma transição similar à transição de segunda ordem em sistemas magnéticos como os descritos pelo modelo de Ising.

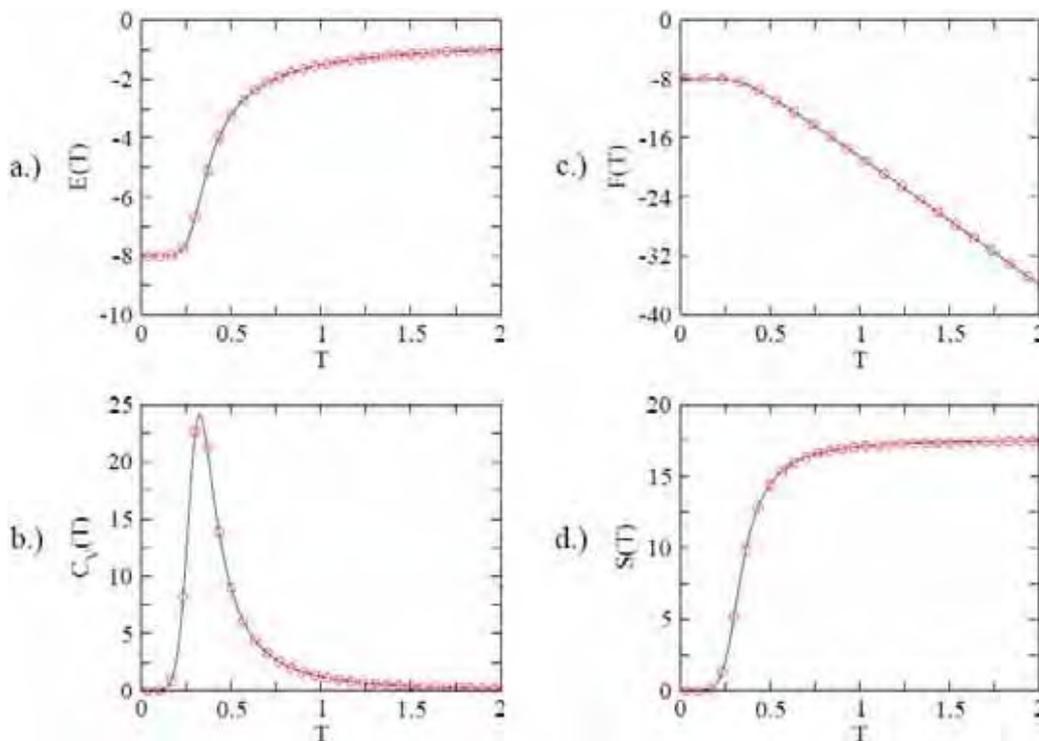


Figura 3.8 – Apresenta em vermelho os resultados obtidos com o método, enquanto que as linhas apresentam os resultados exatos. No gráfico (a) é apresentada a energia interna, no (b) o calor específico, no (c) a energia livre e no (d) a curva da entropia. Todas as curvas como função da temperatura ($k_b T = J$).

Capítulo IV

Resultados e Discussão

4.1 Introdução

Durante esta seção, será discorrido acerca de cada um dos modelos em rede utilizados durante a pesquisa, bem como a apresentação das quantidades termodinâmicas encontradas com os mesmos. Todos os resultados foram obtidos com a implementação do algoritmo de Wang-Landau. Deve-se lembrar, entretanto, que todos esses modelos são idealizados, (*toy models*), que fazem a mímica de uma cadeia de monômeros, onde o mais conhecido deles, é o modelo HP, que pode ser classificado como um Modelo Exato Simplificado [3].

Para a pesquisa, foram escolhidas cadeias com 27 monômeros em uma rede cúbica. Tal tamanho de cadeia é considerado como modelo canônico para o estudo do enovelamento e desenho de proteínas [49], já que para cadeias maiores é exigido um grande esforço computacional. Entretanto, como no espaço de 27 monômeros, o número de contatos é semelhante ao apresentados em pequenas proteínas (de 60 a 80 monômeros) [49], esse modelo pode ser considerado mais compatível com a realidade.

As seqüências utilizadas em cada modelo podem ser encontradas na tabela abaixo. As duas primeiras seqüências correspondem ao modelo HP, ou seja, *H* representa um monômero do tipo hidrofóbico, e *P* um monômero do tipo polar. A segunda seqüência, corresponde ao modelo AB, onde os monômeros são divididos em dois grupos, A e B. Já a ultima seqüência

corresponde ao modelo com interações dadas pela matriz de Miyazawa e Jernigan (este é um modelo mais realista), onde cada letra corresponde a um aminoácido. A energia será dada em unidades de $J = 0,6 \text{ kcal/mol}$ [52] e por conveniência a constante $k_b = 1$. Assim, também a temperatura será dada em unidades de J .

Tabela 4.1 – Na Tabela, a primeira coluna servirá de referência para a identificação da cadeia, a segunda coluna determina o número de monômeros, já a ultima mostra a seqüência da cadeia.

Nº	Monômeros	Seqüência
4.1	27	HPHPHPHHHPHPHPPPPHHPHPPHPHH
4.2	27	PHPPHPHPPHHPHHHPPPHHPHPPHPH
4.3	27	ABABBBBBABBABABAAABBAAAAAAB
4.4	27	DCSATYNFVPAGLSQHMRTIEIGWVKL

É importante ressaltar que todos os modelos foram trabalhados numa rede cúbica através do algoritmo de Wang-Landau e os programas escritos em FORTRAN foram executados num Pentium IV de 3Ghz. Inicialmente serão apresentados os resultados obtidos com o modelo HP (modelo da cadeia em solução aquosa) que prioriza as relações hidrofóbicas, com o objetivo de nortear os resultados obtidos com as outras cadeias heteropoliméricas.

Em seguida, serão discutidas duas variações do modelo HP; o modelo introduzido por Wingreen e sua equipe [50], onde não apenas as relações hidrofóbicas serão consideradas, mas também as interações entre H e P , e entre P e P ; e o modelo AB, que prioriza a interação entre monômeros do mesmo gênero, e começa a deixar o foco das relações hidrofóbicas.

Outros dois modelos, mais realistas, serão finalmente apresentados. O primeiro é o modelo de 20 letras (referente às 20 diferentes siglas de monômeros) com interações dadas pela matriz de Miyazawa-Jernigan (MJ) [49], que por apresentar uma interação diferente para cada par de monômeros o torna um modelo bem mais realista que os anteriores. O segundo é o Modelo de Interações Independentes [49], onde a interação entre dois monômeros será sorteada com base numa distribuição gaussiana de energias. Pode-se dizer que este último

modelo é mais real que o modelo HP, pois sorteia-se aleatoriamente as interações entre os monômeros, ou seja, se está na realidade simulando diferentes seqüências de aminoácidos.

Finalizando, durante a apresentação dos resultados, observaremos a presença de transições de fase na rede polimérica. Somos tentados a fazer uma analogia dessas transições com as que ocorrem em sistemas magnéticos, no limite termodinâmico. No caso de cadeias poliméricas, claramente este limite não se aplica. Assim, não podemos, por exemplo, identificar a presença de uma descontinuidade nas curvas dos parâmetros de ordem em função da temperatura e distinguir categoricamente dois tipos de transições. Apesar dessas limitações, utilizaremos freqüentemente os termos transições de primeira e segunda ordem para facilitar a discussão e apresentação dos resultados.

4.2 Modelo HP

Como visto, o modelo HP se caracteriza como uma cadeia de monômeros dos tipos hidrofóbico e polar, imersa numa substância aquosa, já que prioriza a interação hidrofóbica como a força dominante no enovelamento. Portanto, cada contato entre um monômero do tipo H com um de seus vizinhos mais próximos também do tipo H (desconsiderando as ligações da cadeia, ou seja, as pontes peptídicas), implicará numa redução da energia. Interações desse tipo são responsáveis pelo colapso da cadeia, cujo objetivo é minimizar o contato dos monômeros hidrofóbicos com o meio. Esse modelo é apresentado como um modelo de referência no estudo de enovelamento de proteínas na rede, assim como o modelo de Ising é para o estudo do magnetismo em sólidos.

Inicialmente serão apresentados os resultados com a seqüência 4.1 definida na tabela 4.1 (tempo de CPU de 15,9 min), que apresenta uma transição de segunda ordem do estado desnaturado para o estado compacto. Na figura 4.1 é possível observar que as curvas da energia e do número de contatos, ambos parâmetros de ordem do sistema, mostram a mesma curvatura quando analisadas em função da temperatura. No *inset* dessa figura, pode-se observar que a curva se aproxima muito de uma reta, o que indica uma forte dependência entre a energia e o parâmetro de ordem Q .

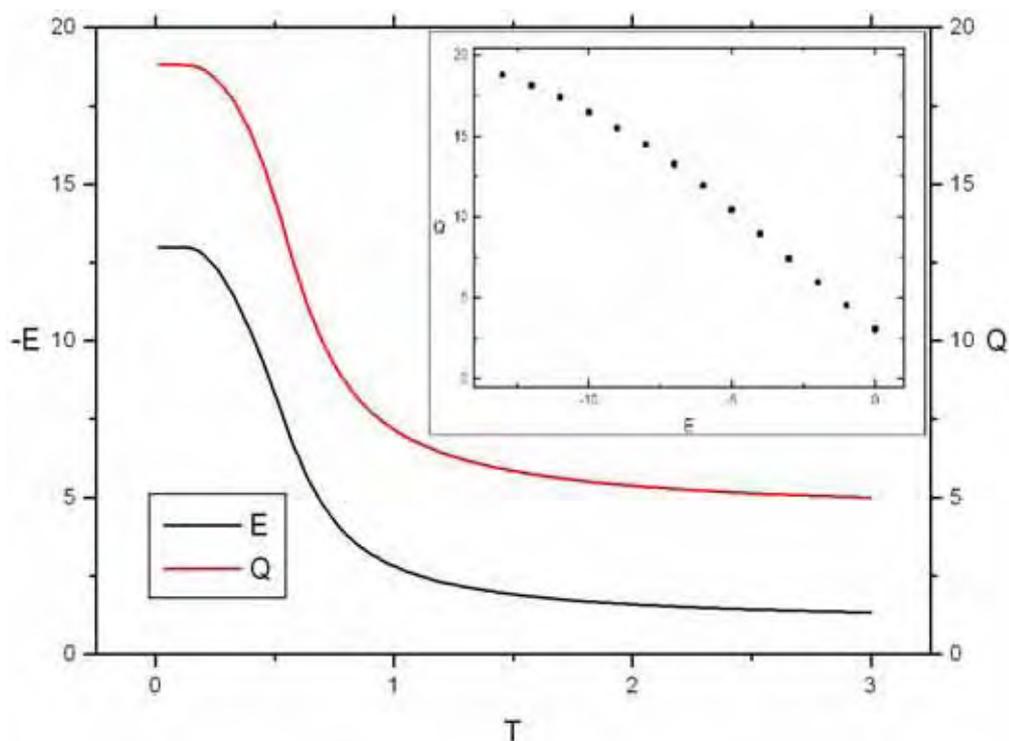


Figura 4.1 – A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se um gráfico do número de contatos, como função da energia.

O colapso da cadeia pode ser observado através do raio de giração e pela distância entre as duas pontas da cadeia, que diminuem com a diminuição da temperatura e da energia. Para temperaturas mais baixas a cadeia busca estados de menor energia, como visto na figura 4.1.

Na figura 4.2 pode-se perceber que tanto o raio de giração como a distância entre pontas possui dois patamares em relação à temperatura. Em temperaturas mais altas, têm-se os estados desnaturados de maior energia e a cadeia está aberta. Enquanto que a baixas temperaturas têm-se estados mais compactos. Entre estes patamares, tem-se a região de temperatura crítica, onde há ocorrência das conformações compactas e desnaturadas.

No *inset* da figura 4.2 pode-se observar que resultados obtidos para o raio de giração e distância entre pontas tendem a ficar cada vez menores a medida que se diminui a energia. Isso mostra que a cadeia colapsou para formas mais compactas possíveis, de modo que os contatos dos monômeros H com o meio foram minimizados. Tal resultado é ainda confirmado quando percebe-se que a energia média assume valores cada vez menores, enquanto que o número máximo de contatos se torna cada vez maior.

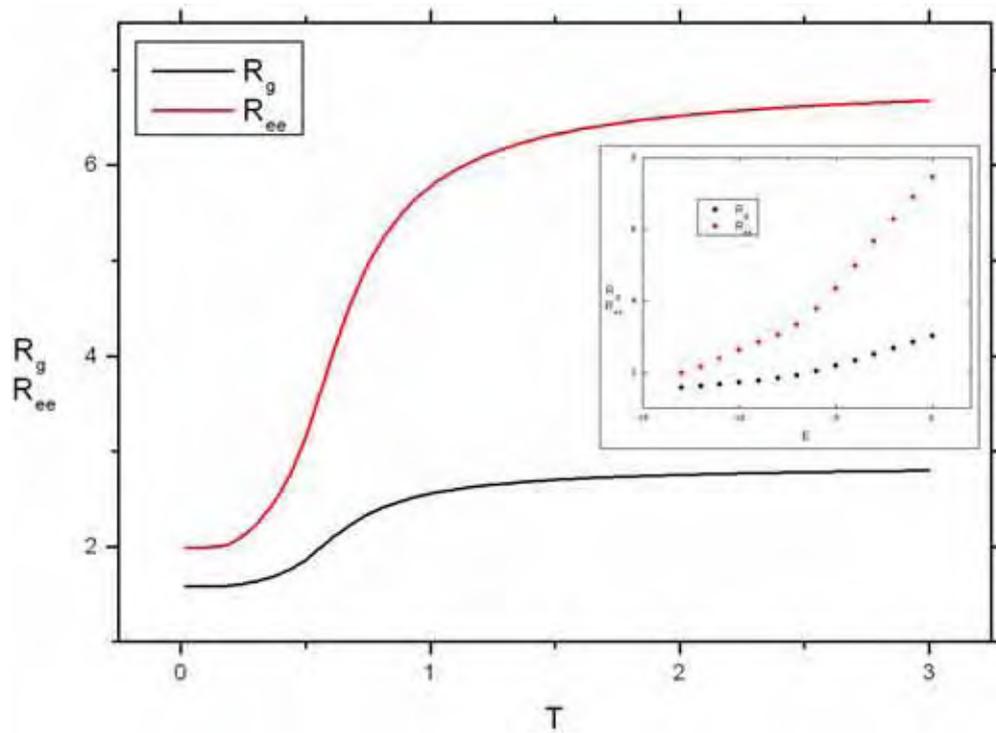


Figura 4.2 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

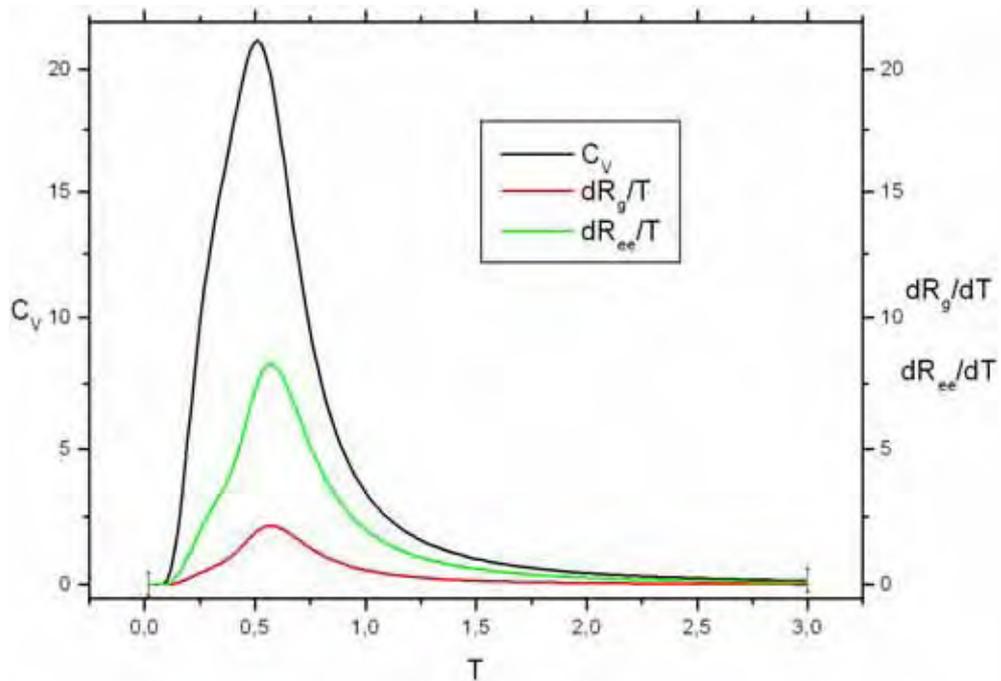


Figura 4.3 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

Os resultados seguintes ajudarão a entender como se deu essa compactação e a identificar entre uma transição de primeira ou de segunda ordem. Na figura 4.3, pode-se perceber que tanto o calor específico como as derivadas em função da temperatura do raio de giração e da distância entre pontas possuem seus picos à mesma temperatura.

Nenhuma das curvas da figura 4.3 apresenta um pico muito acentuado, o que parece indicar uma transição de fase de segunda ordem. O fato das curvas só possuírem um pico indica também apenas a presença de dois grupos de estados ou fases de ordem no sistema. Tem-se então uma fase ordenada e compacta e uma fase desordenada e desnaturada.

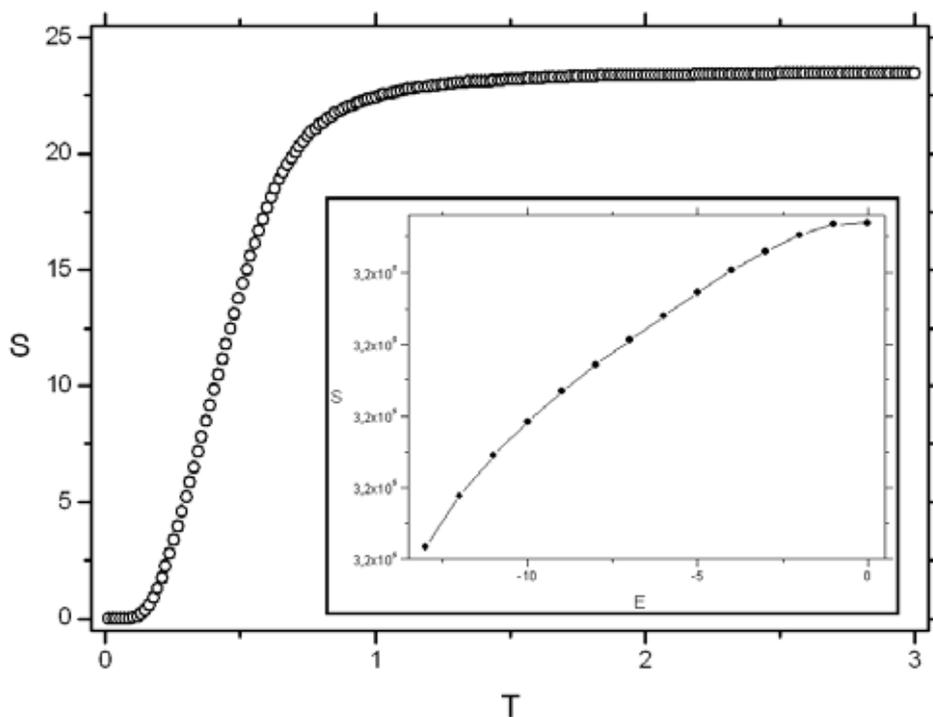


Figura 4.4 – O gráfico principal, apresenta a entropia como função da temperatura. No *inset* dessa figura, tem-se a entropia do método WL.

Para se ter uma noção adequada da ordem do sistema, pode-se observar a figura 4.4, onde é mostrado o comportamento da entropia em função da temperatura. Nesse gráfico, tem-se que a entropia também apresenta dois patamares estáveis, o que leva definitivamente à conclusão de que a seqüência em estudo tem somente dois “estados” de ordem no sistema.

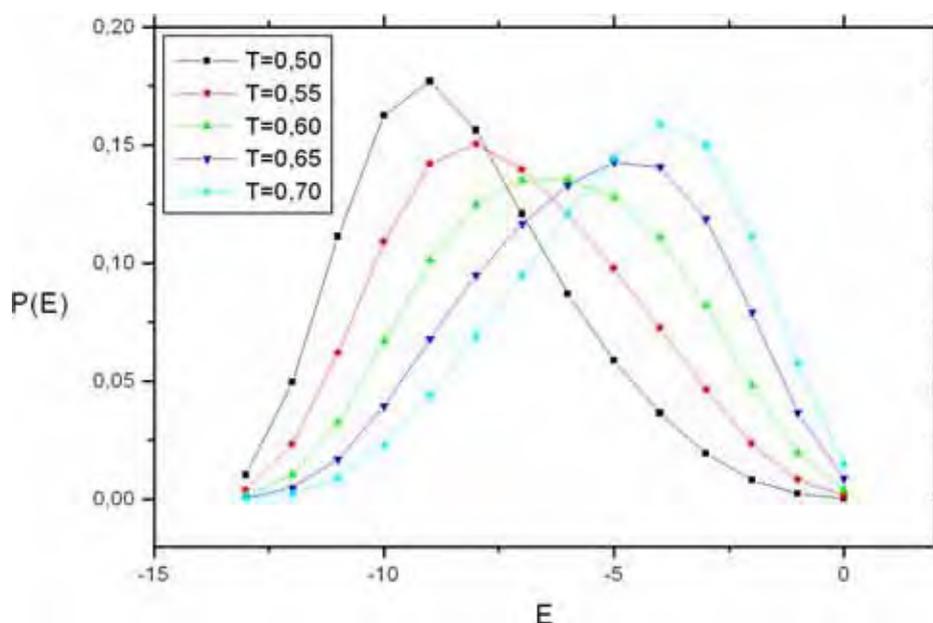


Figura 4.5 – Probabilidade de se encontrar o sistema em uma dada energia E , para cada uma das temperaturas indicadas. A temperatura crítica vale $T=0,60$.

Um último fator que ainda pode ser considerado é a probabilidade de se encontrar o sistema em uma determinada energia. Na figura 4.5 percebe-se que conforme se aumenta a temperatura, a distribuição vai se deslocando para a direita. Na temperatura crítica o gráfico tem a menor amplitude indicando que a curva apresentada na figura 4.5 é típica de transições de segunda ordem.

Concluí-se que a cadeia da seqüência 4.1, tem duas “fases”, ou grupos de estados principais (compactos e desnaturados), entretanto, devido ao curto comprimento da cadeia (27 monômeros), fica difícil caracterizar o tipo de transição analisando apenas parâmetros de ordem.

Para a seqüência 4.2 (tempo de CPU de 89.3 min), como será visto com os gráficos subseqüentes, é possível inferir que a mesma apresenta uma transição de primeira ordem, do estado desnaturado para o estado compacto. Também será possível observar que a mesma, diferente da seqüência anterior, possui três “fases” de ordem no sistema.

Há uma pequena diferença entre os resultados da figura 4.6 e os da figura 3.1. Nota-se a presença de uma rápida variação nos valores de Q e E na região de baixas temperaturas. Embora seja ainda um modelo HP, nota-se a presença de uma nova fase, intermediária, entre os estados compactos e os de conformação aberta. Como será visto adiante, a presença de um pico extra na curva do calor específico confirma esta suspeita.

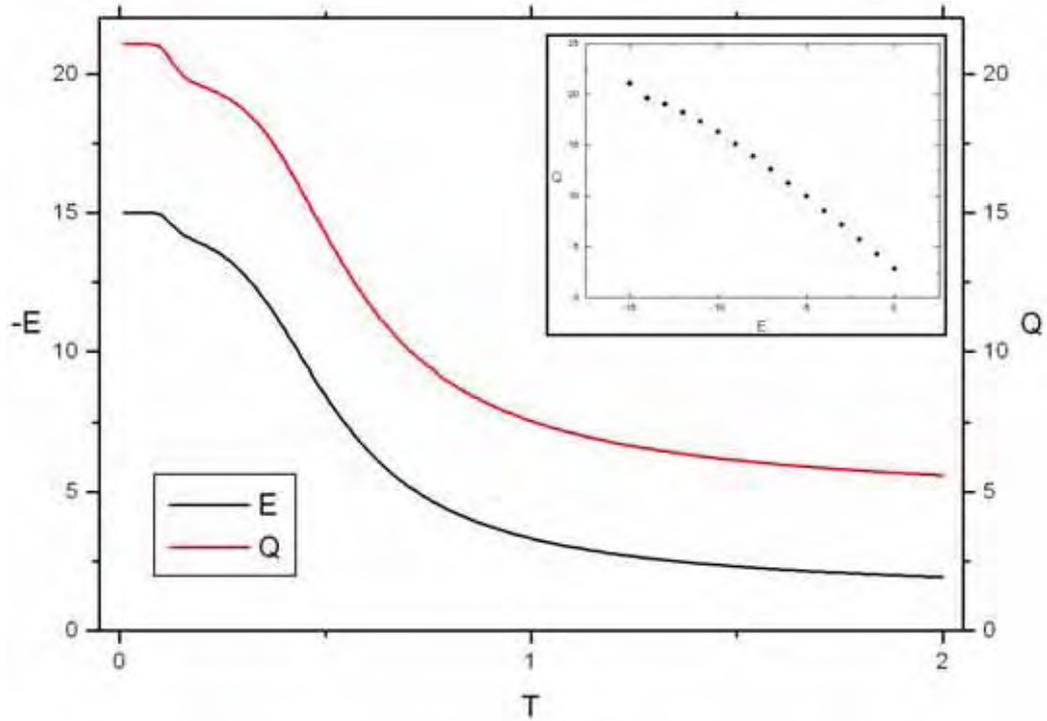


Figura 4.6 – A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se um gráfico do número de contatos, como função da energia.

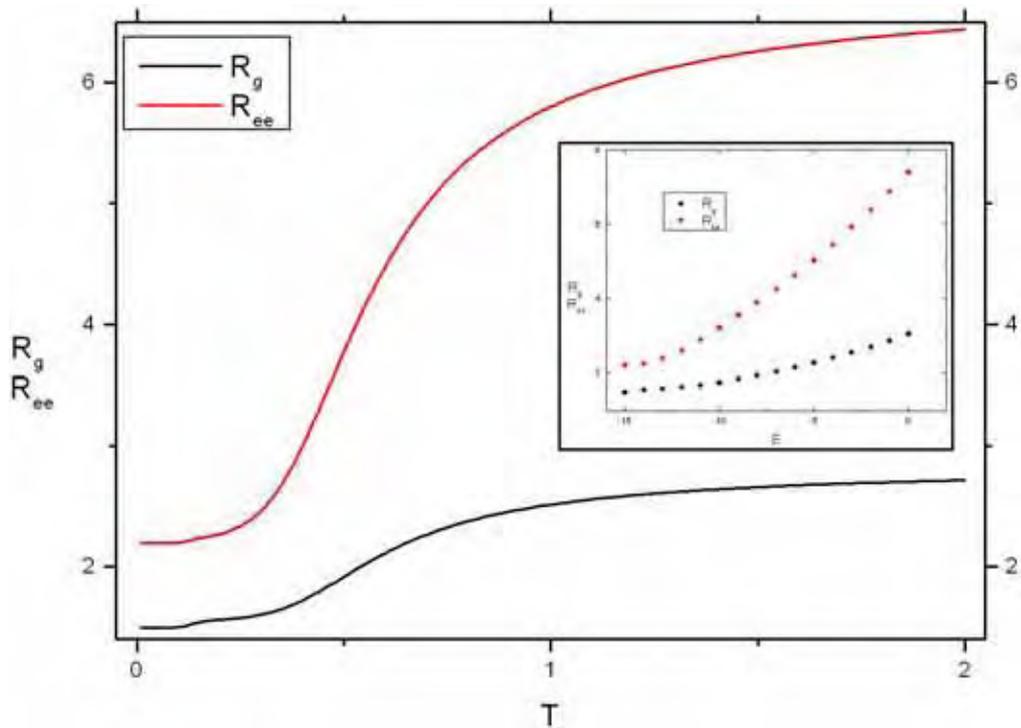


Figura 4.7 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

Assim como para a seqüência anterior, continua-se a observar o colapso da cadeia (figura 4.7). Assim como na figura 4.2, tanto o raio de giração como a distância entre pontas possui dois grandes patamares em relação à temperatura. Entretanto, existe também nestas curvas um pequeno patamar intermediário.

No *inset* da figura 4.7, assim como para a outra cadeia do modelo HP, pode-se observar que resultados obtidos para o raio de giração e distância entre pontas, tornaram-se cada vez menores com a diminuição da temperatura, onde a diminuição é mais acentuada para a distância entre as duas pontas. A figura 4.7 como um todo, mostra novamente que a cadeia, de fato, tende a buscar conformações mais compactas tanto para baixas temperaturas, como para menores energias.

Como já observado, o calor específico é bem mais ilustrativo para essa seqüência de monômeros. A figura 4.8, se mostra bem diferente da figura 4.3 basicamente por dois aspectos: a presença de dois picos e a intensidade do primeiro pico. Entretanto, como no caso da seqüência anterior, as três curvas possuem seus picos aproximadamente à mesma temperatura, o que reforça a hipótese de que a transição de energia está diretamente associada com a compactação da cadeia para o modelo HP.

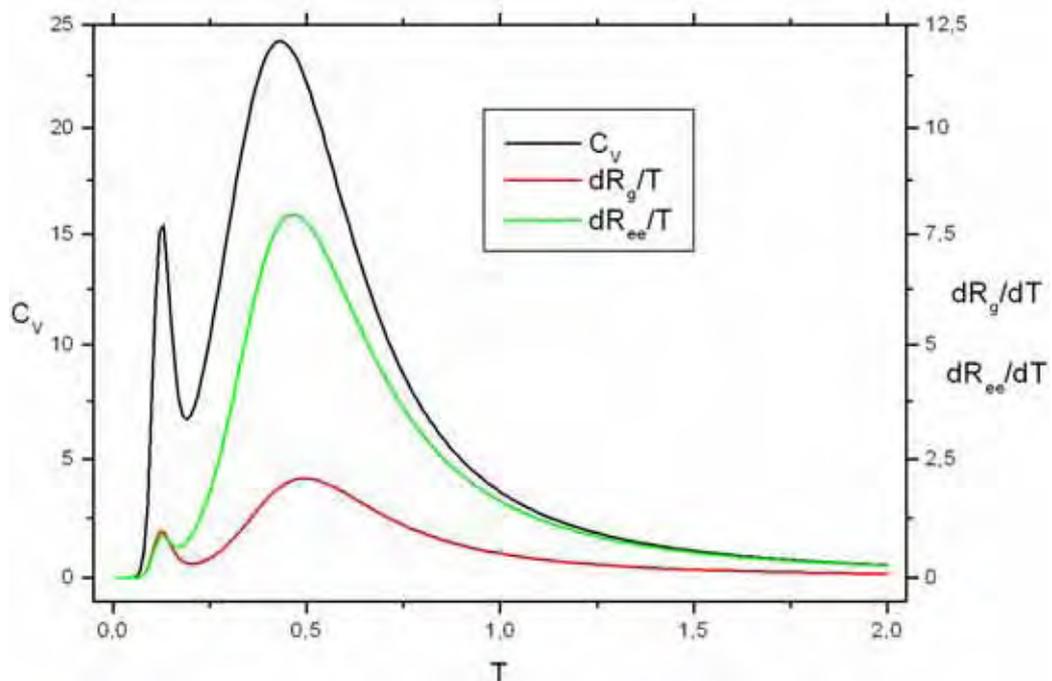


Figura 4.8 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

O fato das três curvas possuírem dois picos indica a presença de três grupos de estados ou “fases de ordem” no sistema. Tem-se então uma fase ordenada e compacta, uma fase intermediária, e uma fase desordenada, onde se tem conformações da cadeia bem abertas. Nenhuma das curvas da figura acima apresenta o segundo pico muito acentuado, o que indica uma transição de fase supostamente de segunda ordem. Entretanto, o foco do estudo, é a transição para o estado compacto, e, em tal transição, é observado um pico bem agudo para o calor específico, o que é característico de transições de primeira ordem.

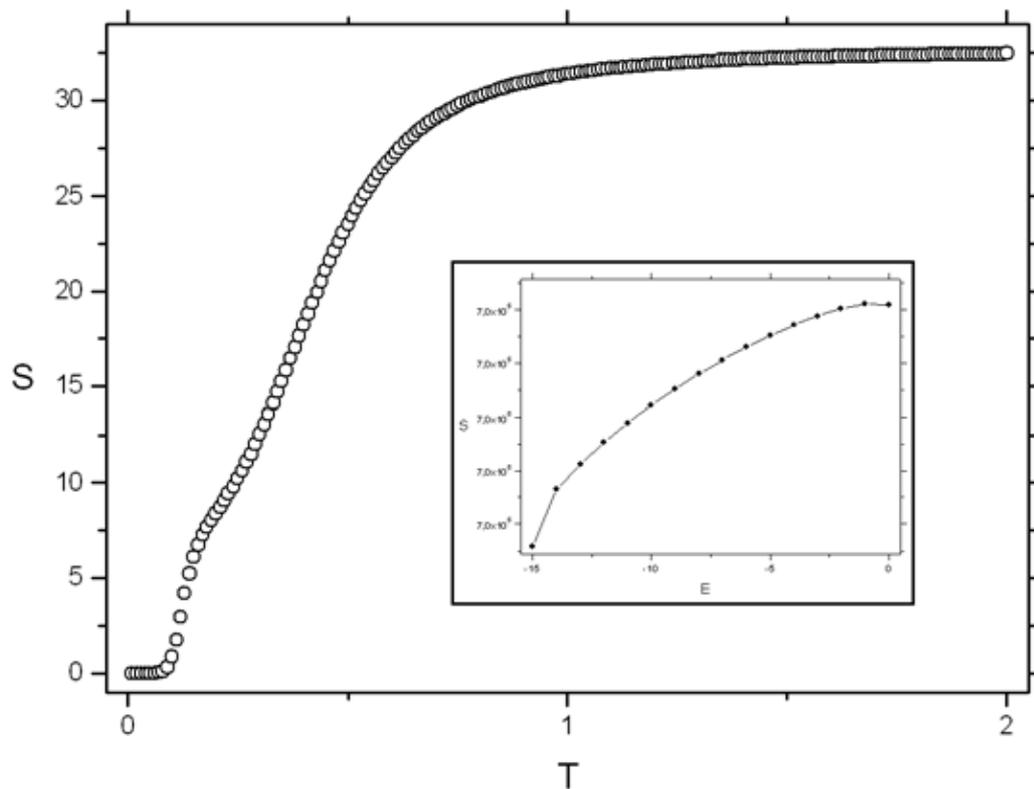


Figura 4.9 – No gráfico principal, tem-se a entropia como função da temperatura. Já no *inset* dessa figura, a entropia do método WL.

A curva da entropia também ilustra bem a ordem do sistema. Na figura 4.9, assim como para a seqüência anterior (figura 4.4), têm-se os dois patamares de ordem e desordem. Entretanto, pode-se observar que a curva apresenta uma pequena perturbação, entre as duas fases, o que indica a presença de um estado metaestável, e semi-ordenado, indicando a presença de três fases no sistema.

Finalmente a distribuição de probabilidade poderia reforçar as suspeitas. Conforme as curvas da figura 4.10, as conformações com energia -15 e -14 , são igualmente prováveis quando a temperatura está entre $0,13$ e $0,14$. Isto poderia indicar transição de primeira ordem. Entretanto, as conformações com essas energias são muito parecidas de modo que a distribuição de probabilidade não ajuda muito para a identificação da transição de fase.

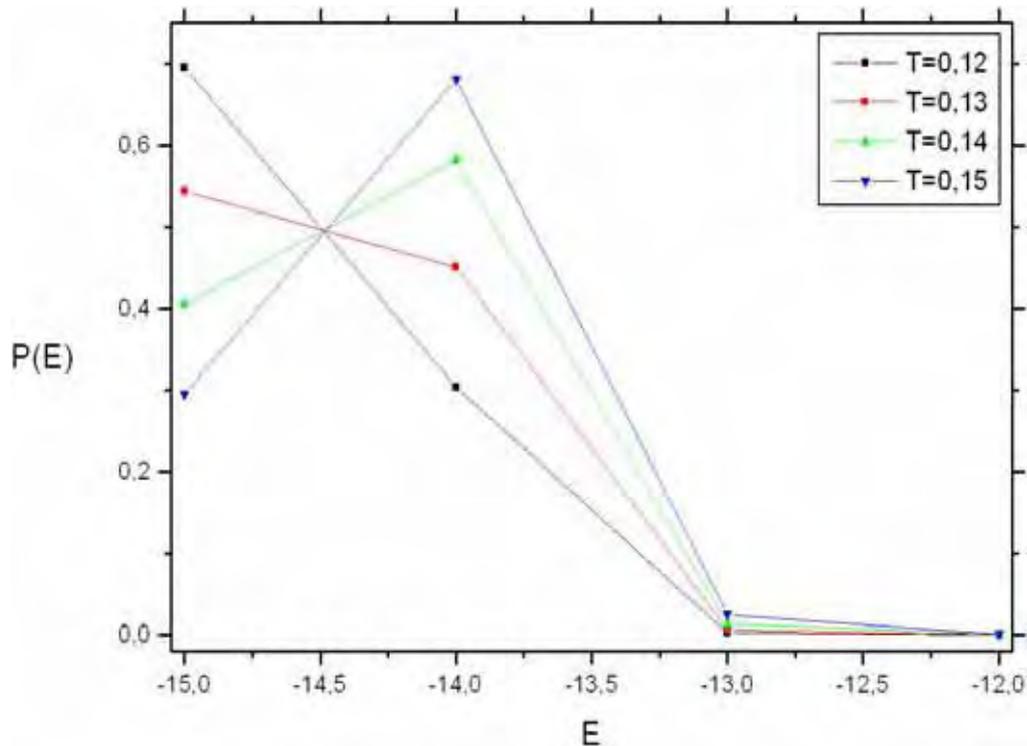


Figura 4.10 – Probabilidade de se encontrar o sistema em uma dada energia E , para cada uma das temperaturas indicadas, onde $T=0,133$, é a temperatura crítica. Neste gráfico foram apenas consideradas as quatro menores energias do sistema.

Com o modelo HP foi possível observar que em ambas as seqüências a diminuição da cadeia está diretamente associada à diminuição da energia, ou seja, a estrutura tende a uma conformação compacta para os estados de menor energia, mesmo que os eles sejam ainda altamente degenerados. Na figura 4.11 é possível observar a diferença da compactação entre a conformação de um estado desnaturado com a de uma conformação para o estado de menor energia. Pode-se ainda observar que uma cadeia descrita pelo modelo HP pode apresentar transições de primeira e segunda ordem, dependendo apenas da seqüência de monômeros considerada.

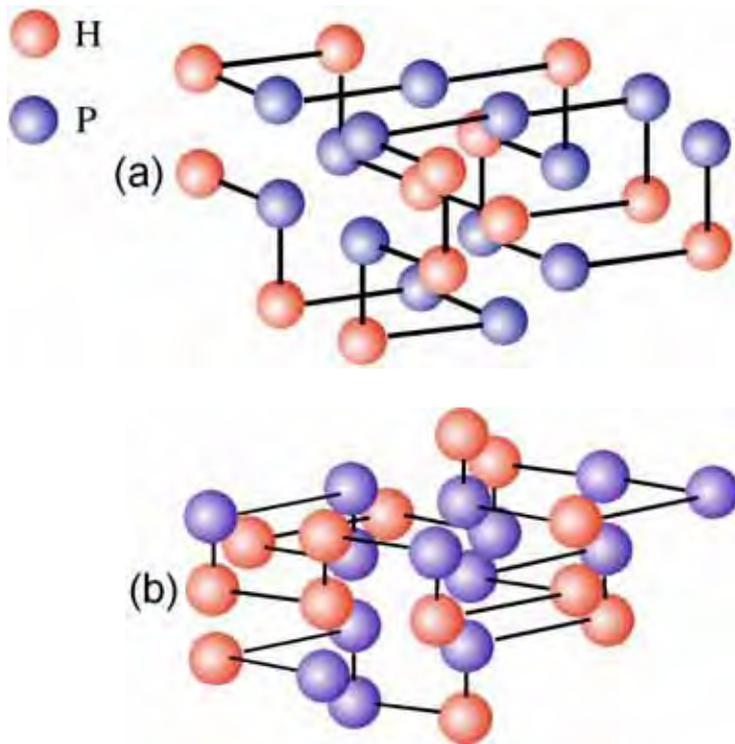


Figura 4.11 – A figura (a) representa um estado desnaturado qualquer enquanto que a (b) representa uma das conformações do estado fundamental. H representam os monômeros hidrofóbicos e P os polares.

4.3 Modelo HP Misto

O modelo que será discutido agora difere um pouco do modelo HP. A interação hidrofóbica continua sendo a principal responsável pela compactação da cadeia. Porém, considera-se também as interações HP e PP obedecendo as seguintes condições:

$$\varepsilon_{HH} < \varepsilon_{HP} < \varepsilon_{PP} \quad (4.1)$$

que garante a ação de forças hidrofóbicas e

$$\varepsilon_{HP} > \frac{\varepsilon_{HH} + \varepsilon_{PP}}{2}. \quad (4.2)$$

que introduz segregação entre monômeros de espécies diferentes.

Para o estudo deste modelo, foram utilizados os seguintes valores para as interações (a escolha de tais parâmetros se deve à experimentação de diferentes valores, já que para alguns parâmetros o tempo computacional se mostrou muito extenso)

$$\begin{cases} \varepsilon_{HH} = -2,5 \\ \varepsilon_{HP} = -1,0 \\ \varepsilon_{PP} = 0 \end{cases} \quad (4.3)$$

Com o objetivo de verificar que mudanças são provocadas por estas novas interações, foi utilizada a seqüência 4.2, a segunda seqüência apresentada no modelo HP. Para este modelo, entretanto, houve um maior tempo computacional (tempo de CPU de 475.3 min), conseqüência direta de um modelo mais complexo.

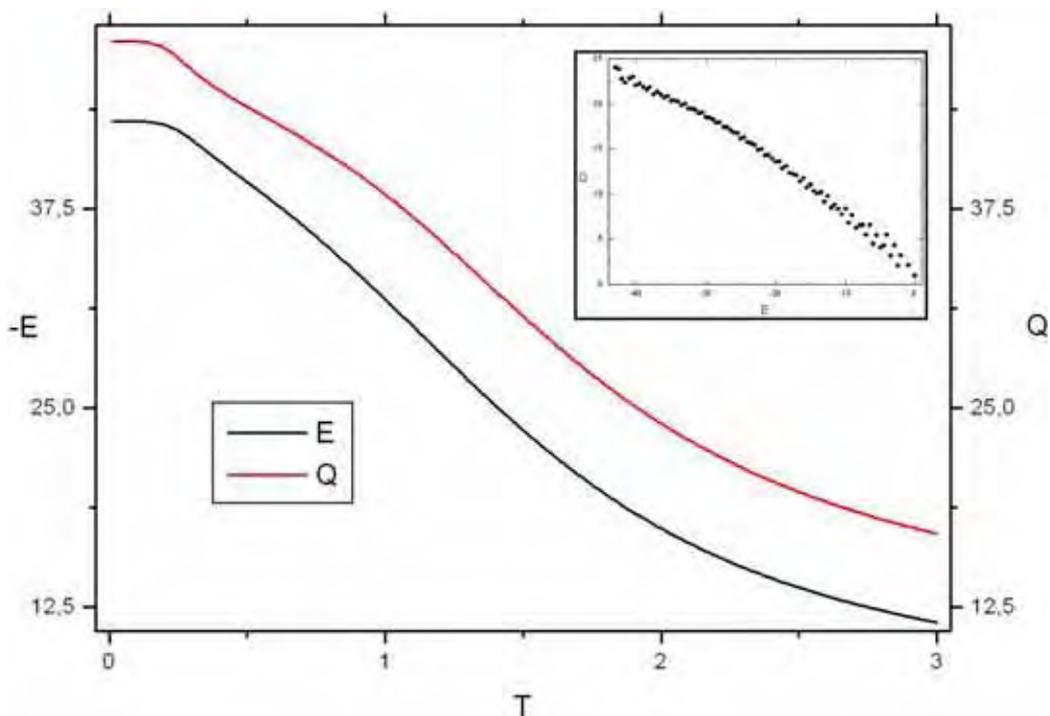


Figura 4.12 – A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se o gráfico do número de contatos, como função da energia.

No *inset* da figura 4.12 é observada uma dispersão dos dados onde pode-se perceber que os dados tendem mais para uma curva que para uma reta, diferente do modelo HP. Entretanto, ainda existe grande congruência entre as curvas da energia E e número de contatos Q . Assim como para o modelo HP, é possível observar que, para baixas temperaturas, os dois parâmetros de ordem apresentam patamares de ordem e desordem. Essas curvas ainda lembram bem as apresentadas na figura 4.6.

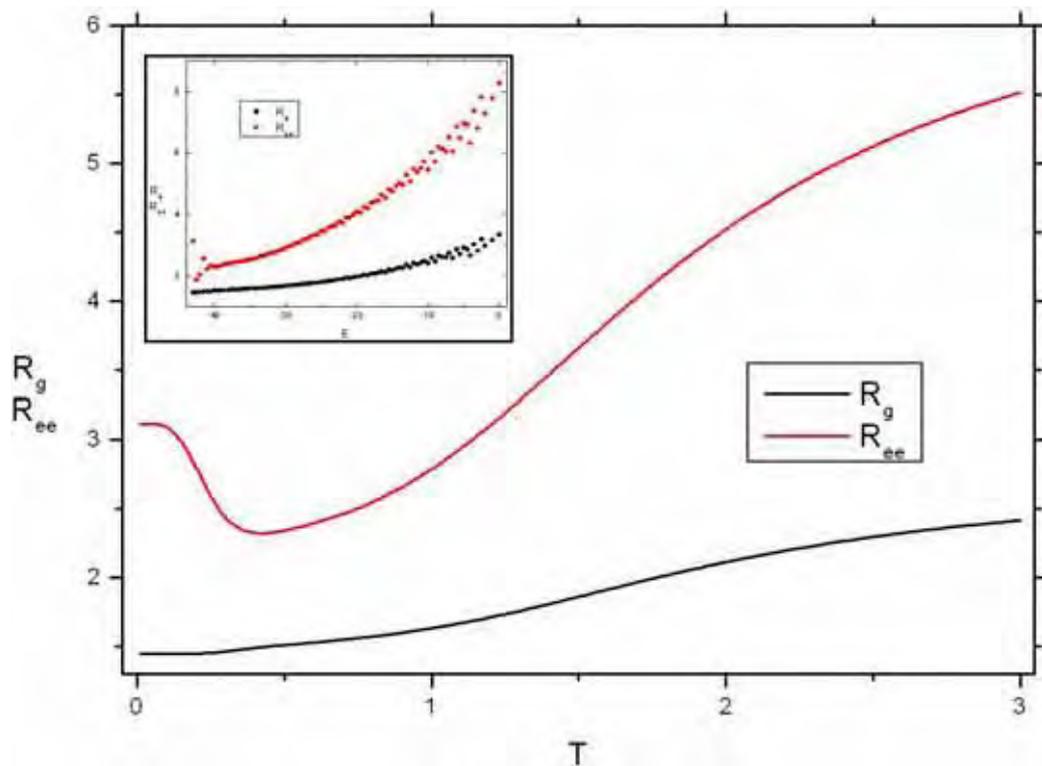


Figura 4.13 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

O *inset* da figura 4.13 começa a apresentar certa dispersão das curvas, com grande variação para a curva da distância entre pontas em relação aos modelos anteriores. No gráfico principal a curva também se mostra diferente, a distância entre pontas diminui até um mínimo e, depois, volta a aumentar com a diminuição da temperatura.

O comportamento do calor específico, figura 4.14, mostra, a exemplo do que ocorreu anteriormente, que o sistema apresenta três fases de ordem já que as três curvas, apresentam dois picos cada uma. Pode-se inferir também que aparentemente a transição entre as duas fases menos ordenadas parece ser uma transição de segunda ordem já que os segundos picos são bem alargados. Entretanto o dado que mais difere das outras seqüências e modelos trabalhados é o da derivada da distância entre pontas, o qual apresenta um pico invertido. Tal configuração da curva indica que nesse trecho houve um aumento significativo no valor da distância entre as pontas.

Quanto ao primeiro pico, ao tomar-se o calor específico fica difícil inferir qualquer coisa a respeito da ordem de transição do sistema, entretanto, o pico extremamente intenso da distância entre pontas pode indicar uma transição de primeira ordem. Entretanto, será necessária a análise de outros dados termodinâmicos antes de apresentar-se qualquer solução.

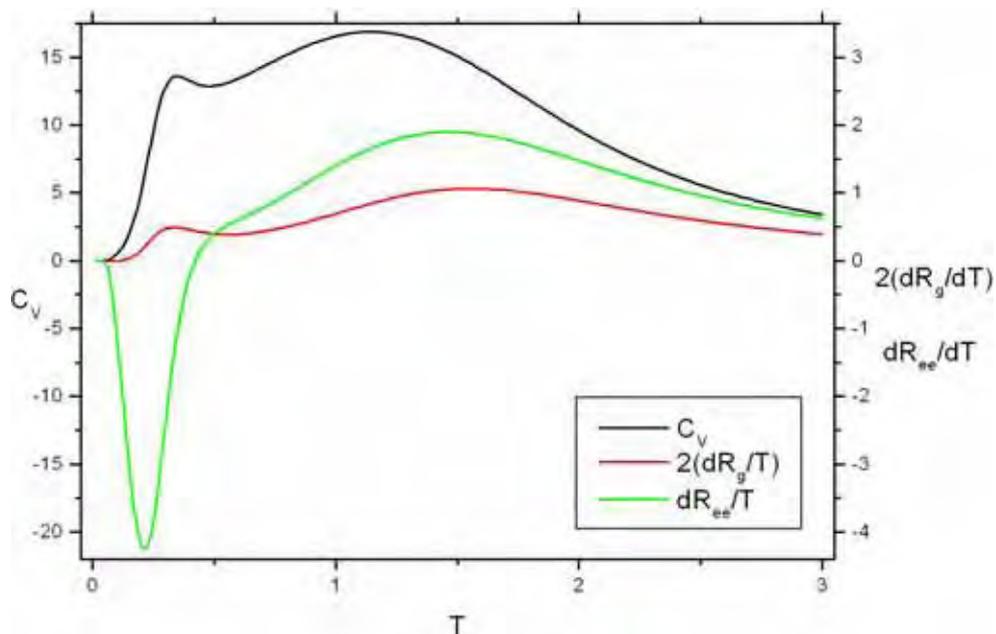


Figura 4.14 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

Na figura 4.15, pode-se perceber que o estado de menor energia encontrado, de fato, assume uma estrutura compacta, o que indica ordem no sistema.

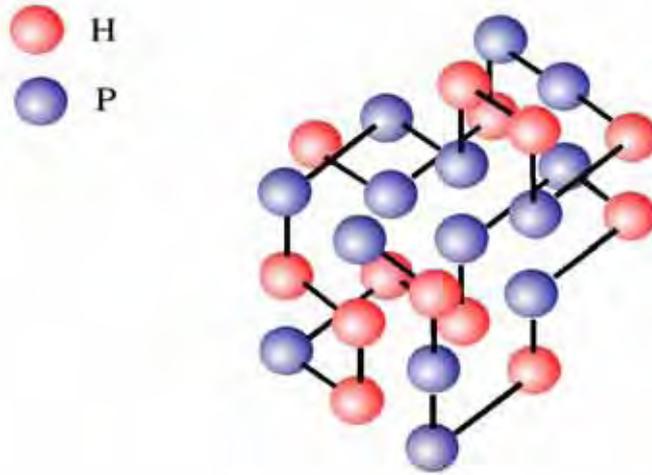


Figura 4.15 – A figura apresenta uma das conformações do estado fundamental, onde H são monômeros hidrofóbicos e P polares.

A curva da figura 4.16 confirma um estado de ordem a baixas temperaturas já que é possível observar dois patamares de ordem, muito bem definidos, sendo um deles a baixas temperaturas. Tal curva, entretanto, não permite em si, grande suporte para a hipótese de três fases de ordem para o sistema, onde se pode perceber apenas duas fases bem definidas, mesmo que a curva apresente uma pequena “barriga” entre os dois patamares.

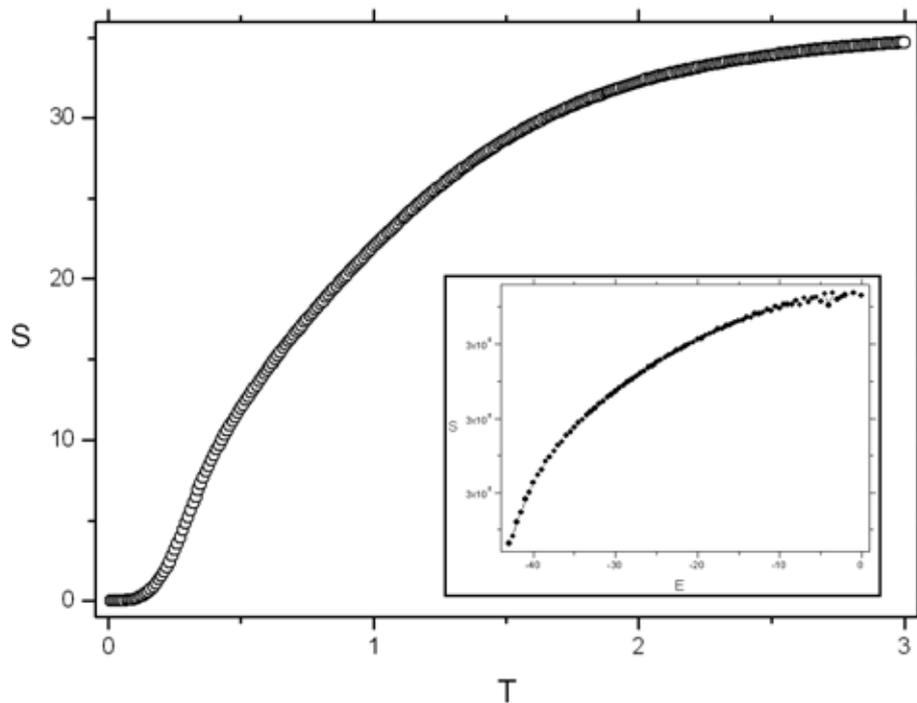


Figura 4.16 – No gráfico principal é apresentada a entropia como função da temperatura. Já no *inset* dessa figura, a entropia do método WL.

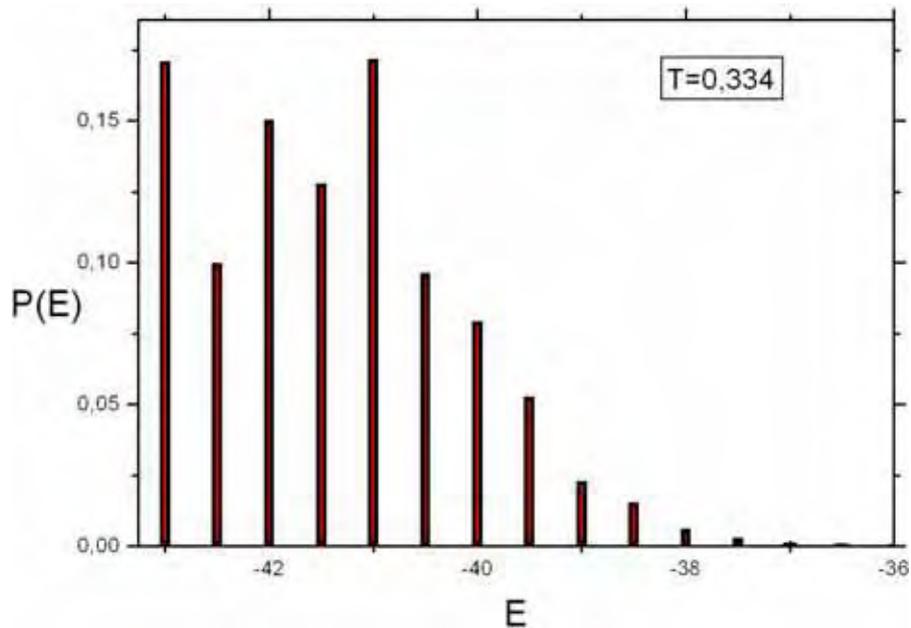


Figura 4.17 – O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=0,334$, no caso a temperatura crítica do sistema.

A figura 4.17 apresenta a distribuição de probabilidades na temperatura crítica. Ao contrario da distribuição apresentada na figura 4.10, as curvas são mais definidas. É possível perceber uma distribuição típica de primeira ordem, já que temos dois grupos de estados igualmente prováveis.

4.4 Modelo AB

O modelo AB constitui-se em mais uma variação do modelo HP, no entanto, sem caracterizar as interações como hidrofóbicas. Os monômeros A e B são genéricos. Para este modelo, os monômeros iguais irão se atrair, enquanto que os de tipos diferentes, ou se repelem, ou se atraem com menor intensidade que os iguais como mostra a equação

$$\mathcal{E}_{AA} = \mathcal{E}_{BB} < \mathcal{E}_{AB} \cdot \quad (4.4)$$

Com base nisso, pode-se dizer que não importará se um monômero for do tipo hidrofóbico ou polar, mas apenas se o contato se dá entre monômeros do mesmo tipo ou de tipos diferentes. Assim são nomeados os aminoácidos das seqüências utilizadas para esse

modelo com as letras A e B (Tabela 4.1). Para o estudo deste modelo, foram utilizados os seguintes valores para as interações

$$\begin{cases} \varepsilon_{AA} = -3,0 \\ \varepsilon_{AB} = -1,0 \\ \varepsilon_{BB} = -3,0 \end{cases} \quad (4.5)$$

Neste modelo, o tempo de CPU foi de 1981 min (cadeia 4.3 da tabela 4.1). É claramente observável, portanto, que o tempo de CPU irá aumentar conforme a complexidade do modelo. Esta seqüência foi estudada por Socci e Onuchic [51] e o propósito da escolha da mesma está na comparação com os resultados obtidos por esses autores. Pode-se adiantar que os resultados foram muito bons.

A característica principal do modelo é fazer com que a cadeia no seu estado fundamental adote uma conformação cúbica, onde os monômeros de mesmo tipo tendem a se aglomerar, como será visto na figura 4.23.

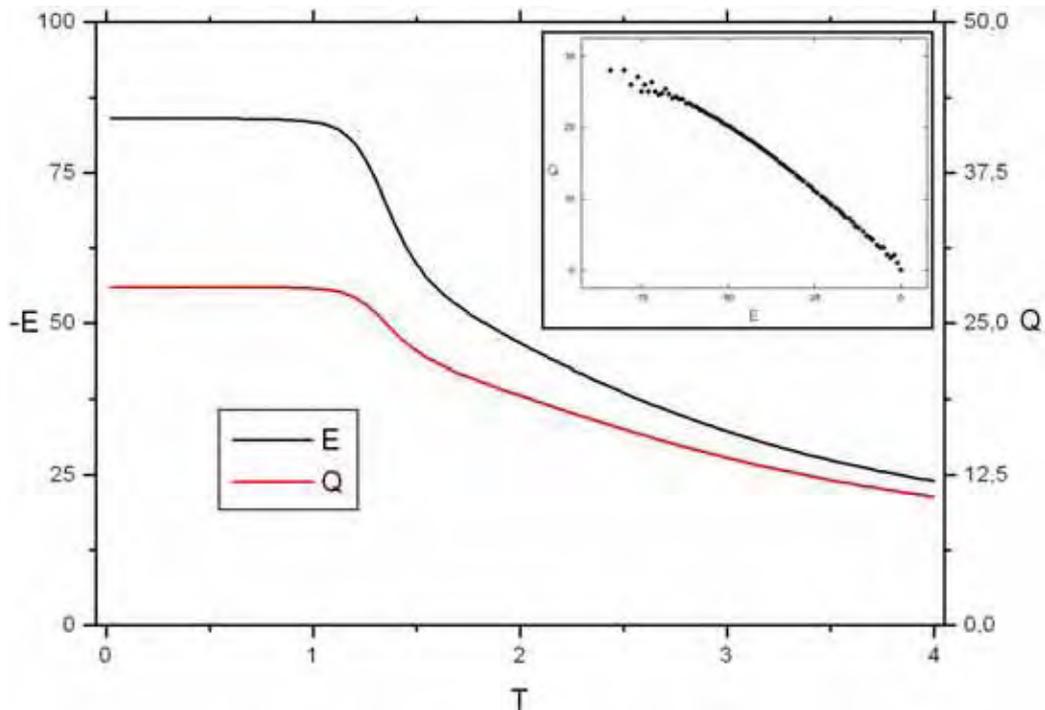


Figura 4.18 – A linha preta representa a energia, enquanto que a vermelha o número de contatos, ambas em função da temperatura. No *inset* se tem o gráfico do número de contatos como função da energia.

Na figura 4.18 a curva $Q(T)$ mostra que o estado de menor energia possui 28 contatos nativos mostrando que a estrutura é cúbica, ou seja, a mais compacta possível. Esta é a primeira vez que o estado fundamental, neste caso nativo, se apresenta como estrutura cúbica. A figura 4.19 mostra a estrutura nativa dessa seqüência. Este indício de que a estrutura nativa seja cúbica é confirmada pelo valor do raio de giração a temperatura $T = 0$, que vale $R_g = \sqrt{2}$.

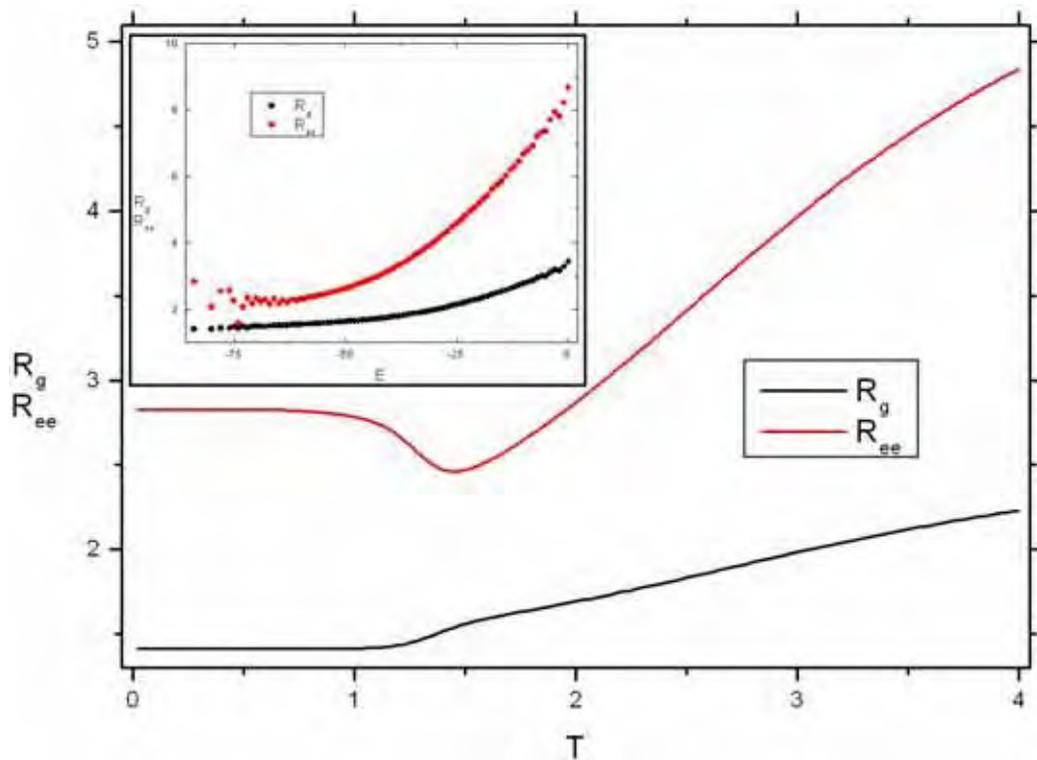


Figura 4.19 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

As curvas de calor específico e das derivadas do raio de giração e distância entre pontas apresentam picos na mesma temperatura e indicam a região de temperatura crítica. Chama atenção a forma aguda do pico do calor específico, assim como a sua altura. É sabido que curvas deste tipo são características de transição de primeira ordem.

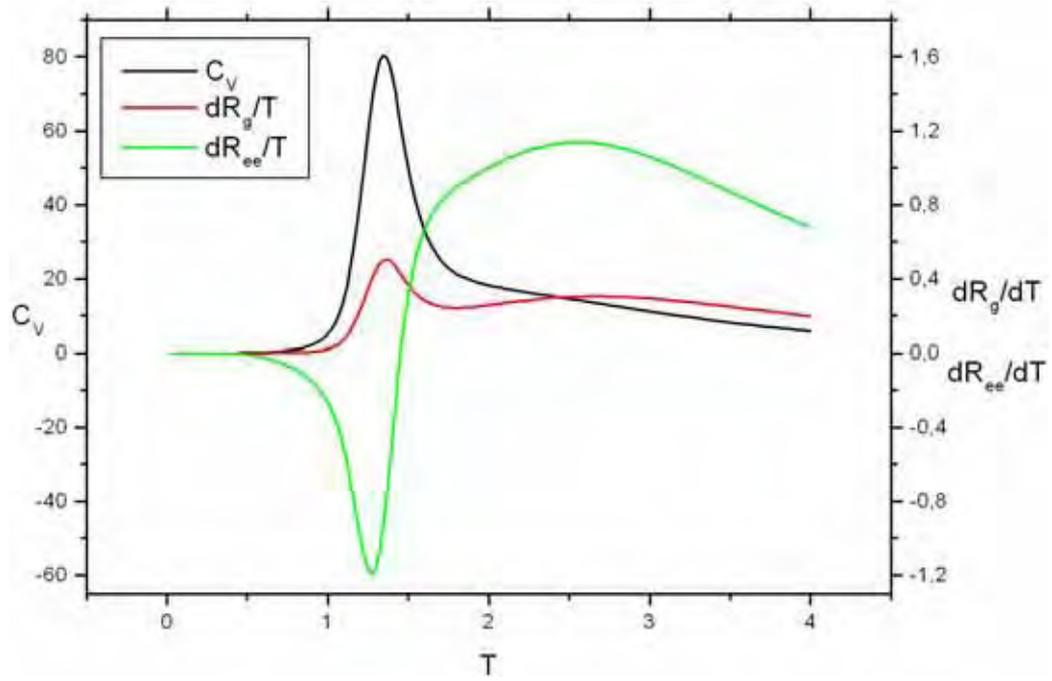


Figura 4.20 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

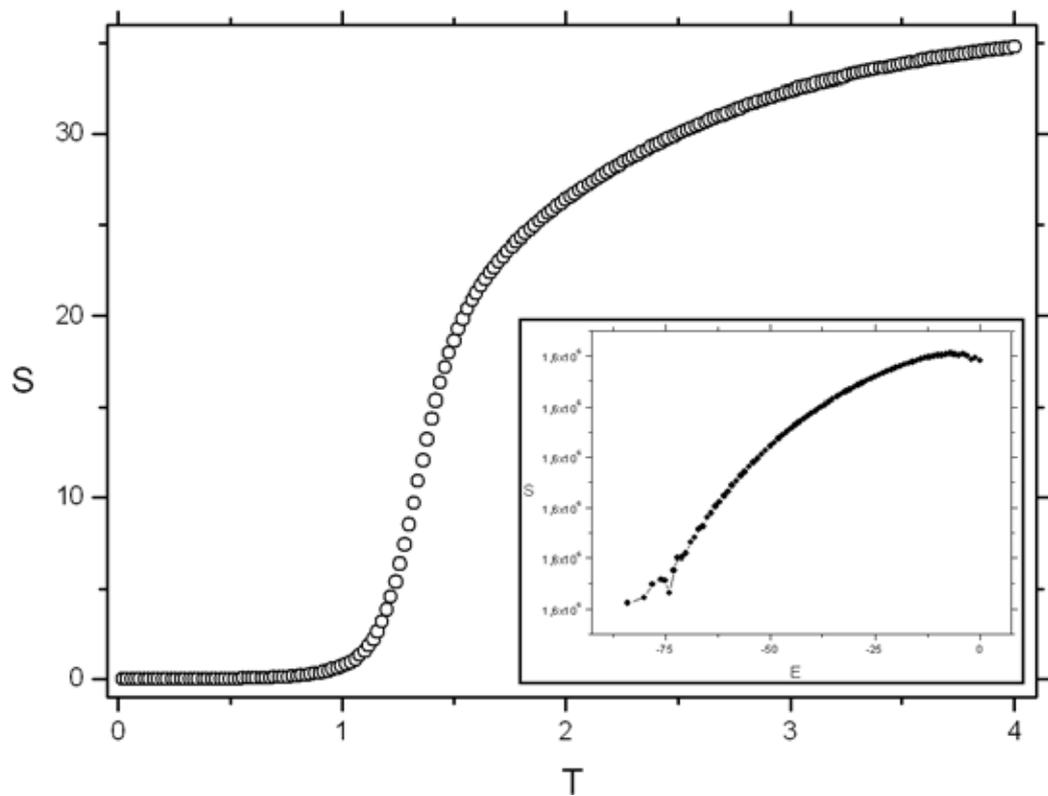


Figura 4.21 – No gráfico principal esta a entropia como função da temperatura. No “inset” dessa figura, a entropia do método WL.

A entropia também apresenta modificações se comparada com as calculadas para o modelo HP. A variação da entropia ocorre de maneira mais brusca neste caso.

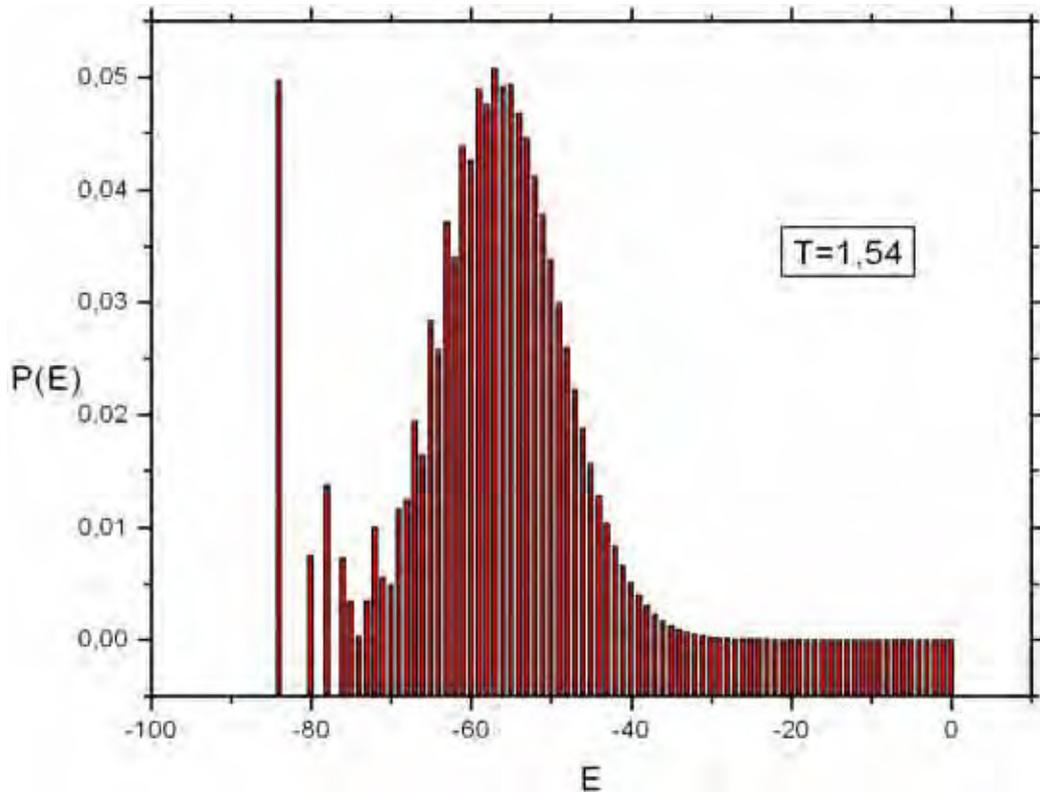


Figura 4.22 – O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=1,54$, no caso a temperatura crítica do sistema.

Entretanto, o resultado que confirma a transição de primeira ordem é a distribuição de probabilidades na temperatura crítica, como mostrada na figura 4.22. Um fato importante que ocorre nesse caso é que o estado fundamental e o primeiro estado excitado estão separados por um pequeno intervalo de energia. Isto indica que o estado nativo é relativamente estável. As flutuações térmicas não desestabilizam a cadeia enovelada, o que é uma característica fundamental para uma estrutura nativa de uma proteína.

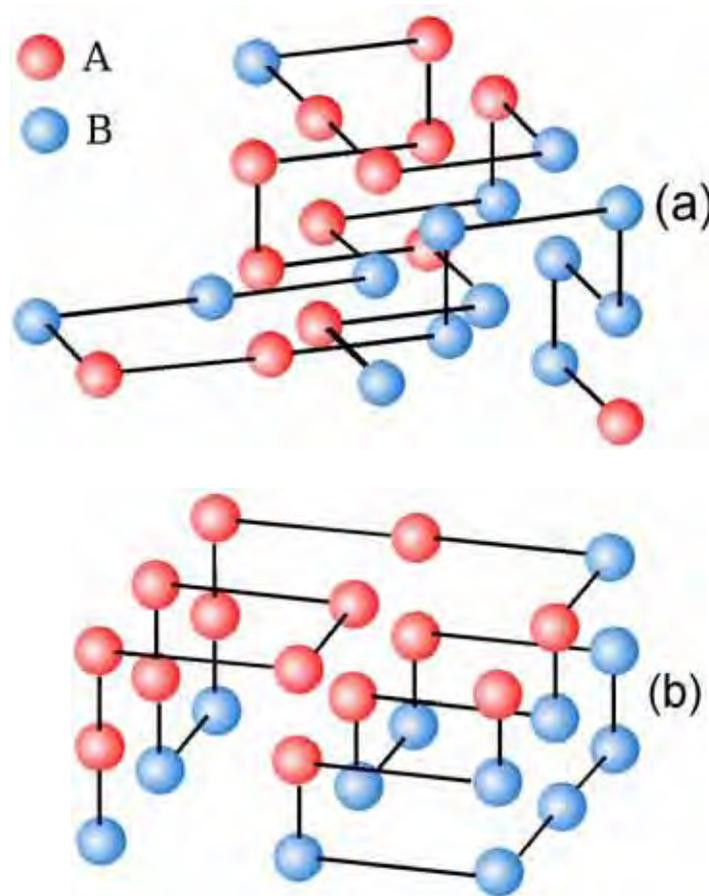


Figura 4.23 – A figura (a) representa um estado desnaturado qualquer, enquanto que a (b) representa o estado nativo.

O estado nativo desta cadeia pode ser observado na figura 4.23 (b), onde é possível observar a estrutura cúbica prevista. Outro fator interessante é que os monômeros estão divididos claramente em dois grupos distintos, consequência do modelo AB, como também afirmado anteriormente. A figura 4.23 (a) mostra uma estrutura desnaturada para fins de comparação.

4.5 Modelo de 20 Aminoácidos

Nesta seção será estudada uma cadeia que possui uma seqüência com 20 aminoácidos diferentes, onde a seqüência foi escolhida de maneira aleatória. Este modelo de cadeia heteropolimérica é bem mais real do que as cadeias que foram estudadas até o momento. Neste modelo, as interações entre os aminoácidos foram calculadas e catalogadas em uma

tabela criada por Miyazawa e Jernigan [52]. As interações são de curto alcance e a energia da cadeia dada por

$$H = - \sum_{\langle i,j \rangle} \varepsilon_{i,j} \Delta(r_i - r_j) . \quad (4.6)$$

Onde os $\varepsilon_{i,j}$ são as interações MJ, e $\Delta(r_i - r_j) = 1$ se os aminoácidos i e j estiverem em contato e $\Delta(r_i - r_j) = 0$ se $r_i - r_j$ for maior que o parâmetro de rede.

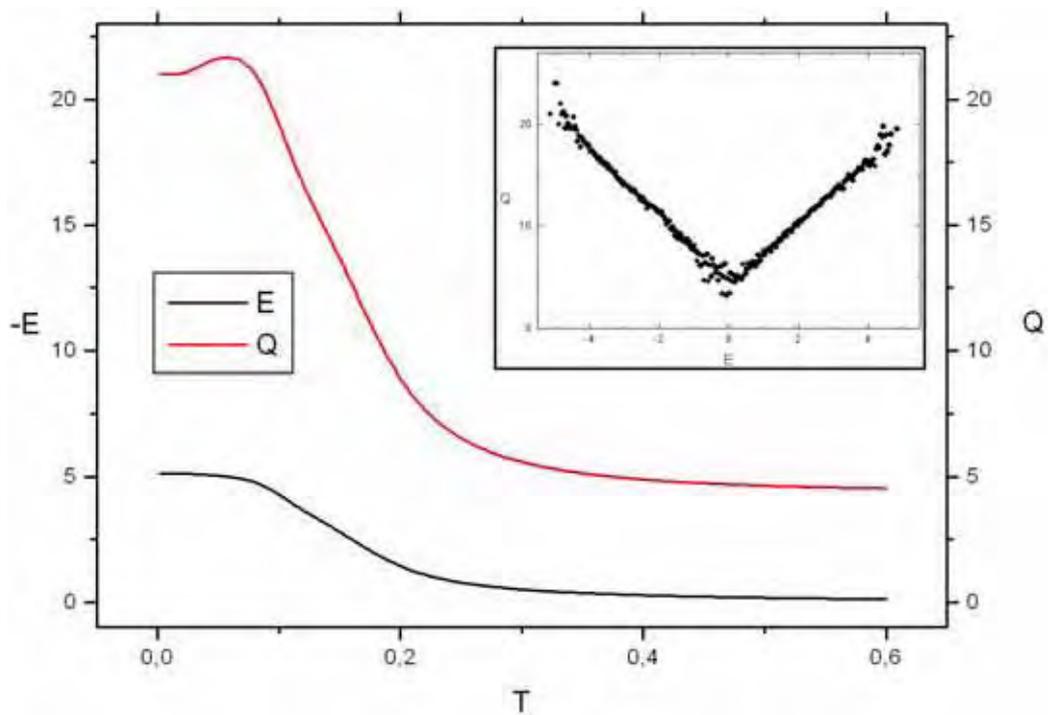


Figura 4.24 – A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* tem-se o gráfico do número de contatos, como função da energia.

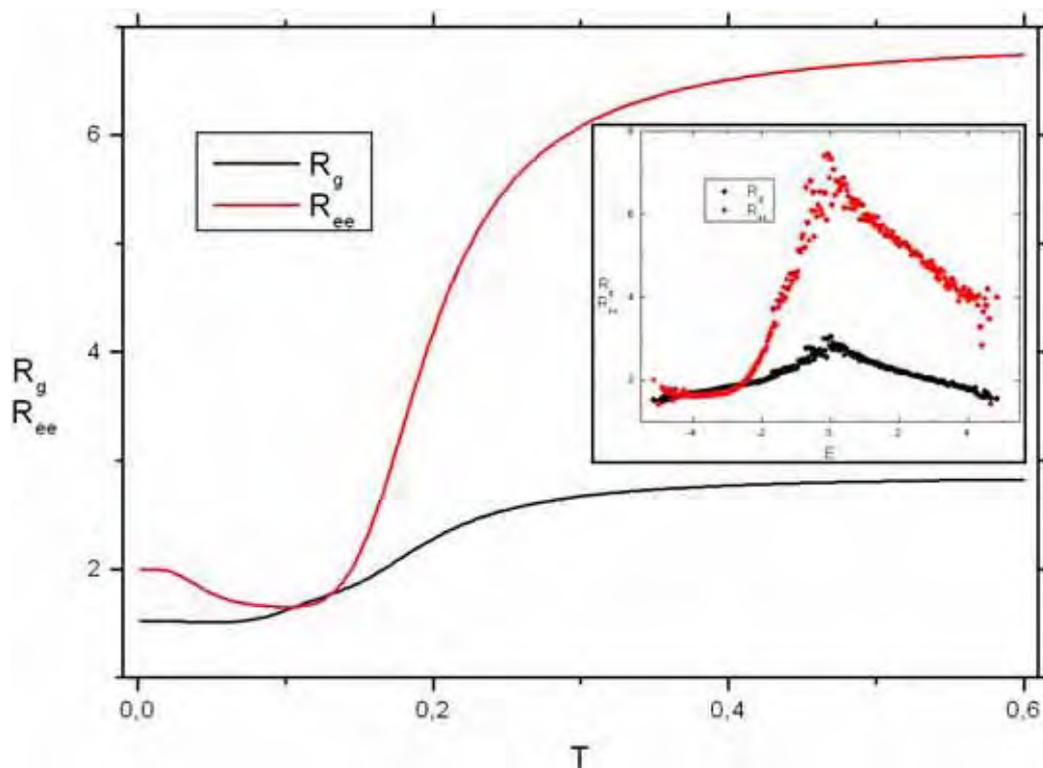


Figura 4.25 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

De modo geral, o comportamento das quantidades termodinâmicas é muito semelhante ao obtido com o modelo HP misto para as curvas em função da temperatura. Porém, como os modelos anteriores apenas consideraram valores de energia menores do que zero existe uma diferença nos gráficos onde as grandezas são apresentadas em relação a energia. Entretanto, considerando-se apenas os valores para as energias negativas, de fato as quantidades termodinâmicas são muito semelhantes.

Tem-se, novamente, uma situação que aponta para um transição de primeira ordem. Portanto, na temperatura crítica, há coexistência de duas conformações com diferentes energias. As conformações com energia maiores poderiam ser consideradas como conformações intermediárias entre o estado fundamental e as conformações de cadeias abertas.

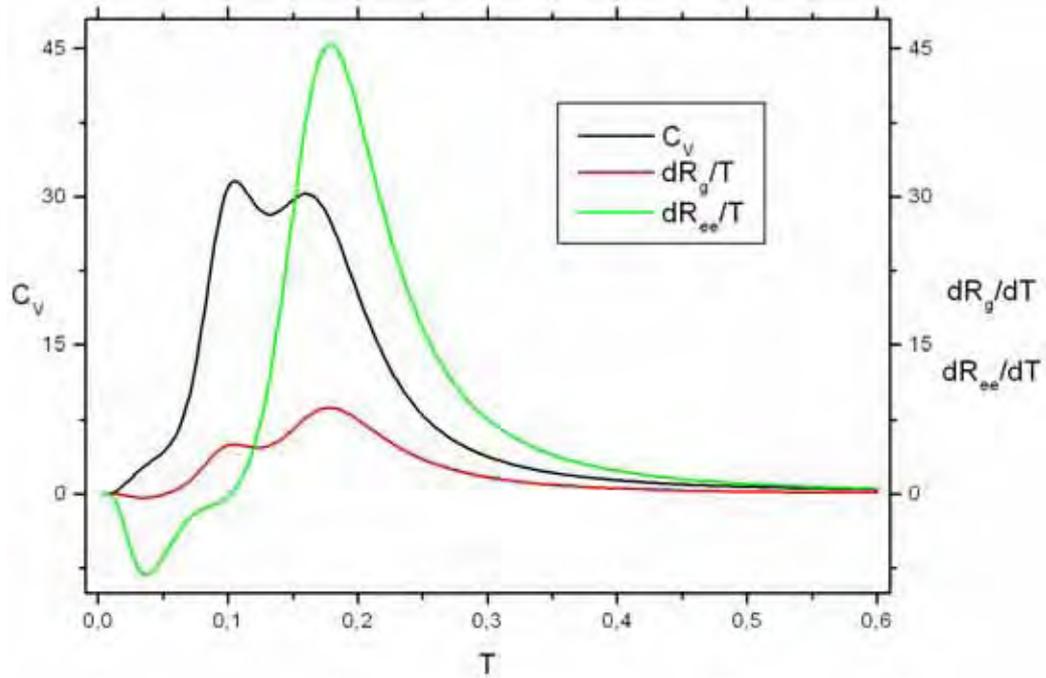


Figura 4.26 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

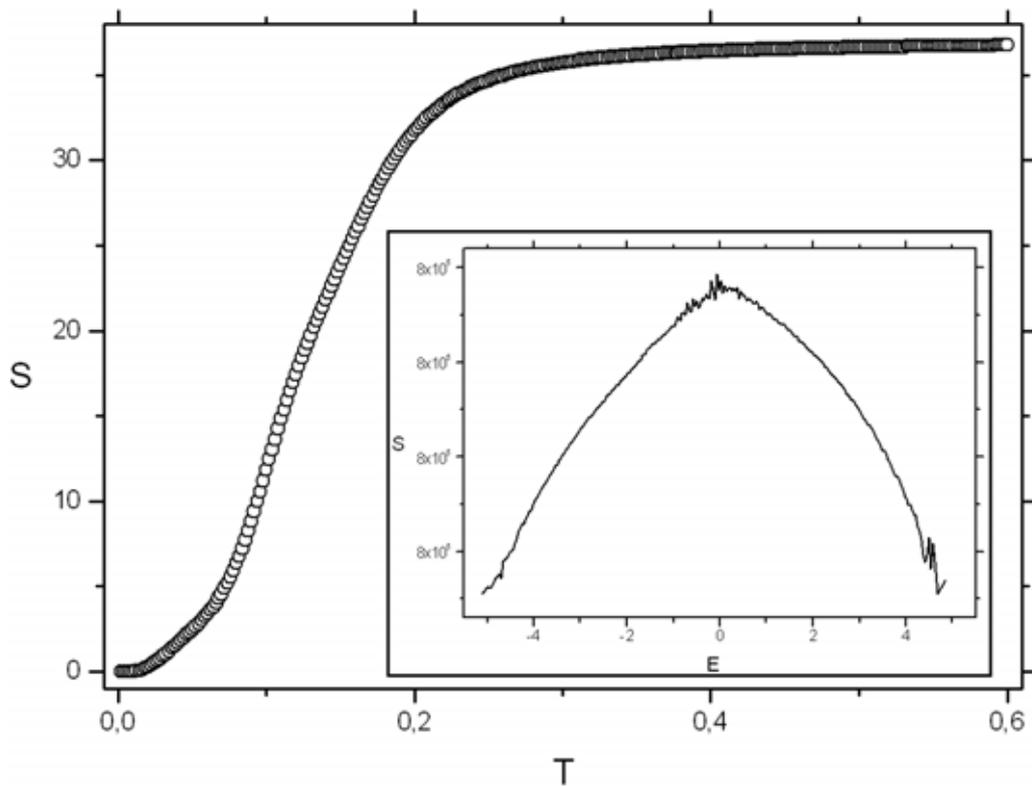


Figura 4.27 – Entropia S como função da temperatura T . O *inset* dessa figura mostra a entropia do método WL.

A hipótese de transição de primeira ordem é confirmada novamente pela distribuição de probabilidades quando o sistema se encontra na temperatura crítica.

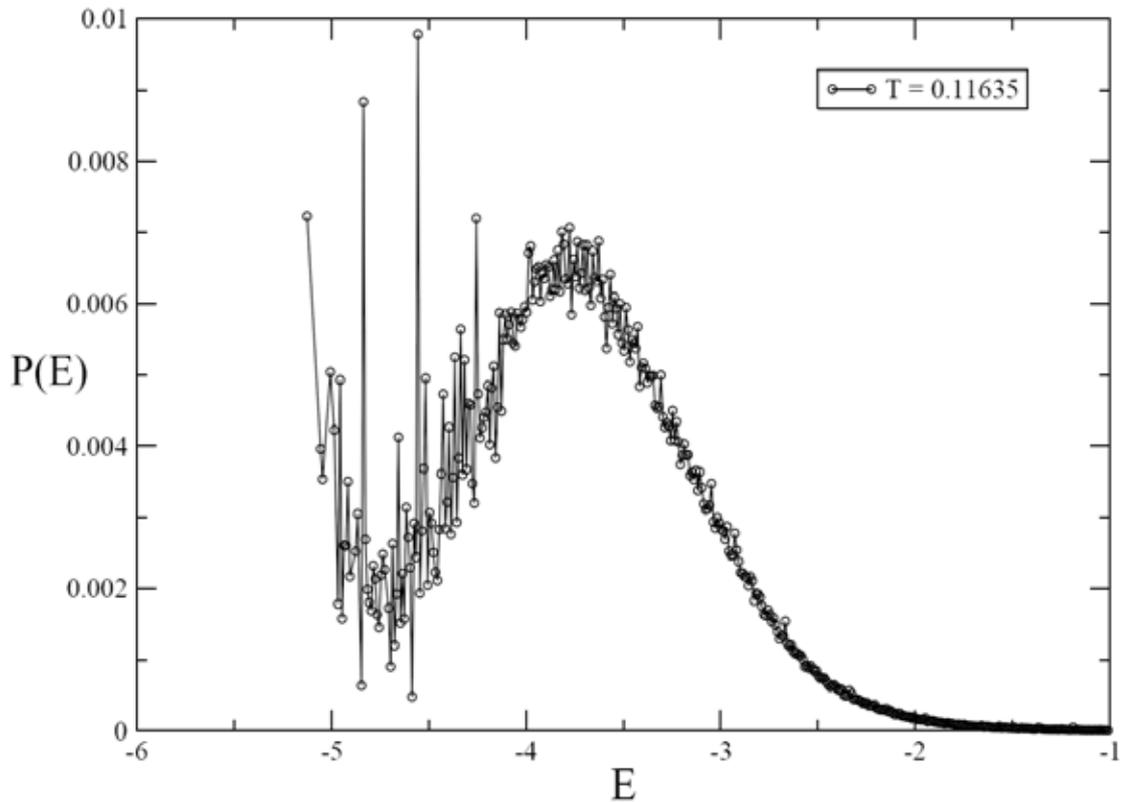


Figura 4.28 – O gráfico acima representa a probabilidade de se encontrar o sistema em cada uma das energias para uma dada temperatura $T=0,11635$, no caso a temperatura crítica do sistema.

Para o estudo deste modelo foi utilizada a seqüência 4.4 da Tabela 4.1, onde cada letra diferente corresponde a um determinado aminoácido (tabela 3.1). Para os dados acima, o tempo de CPU foi da ordem de 4 dias. Mesmo sendo um tempo relativamente grande, comparado com as seqüências anteriores, foi um tempo relativamente curto dada a complexidade do modelo.

4.6 Modelo de Interações Independentes

O modelo de interações independentes foi um dos primeiros modelos a ser utilizado para o estudo do colapso de cadeias heteropoliméricas [53]. Este também é um modelo mais realista e muito semelhante ao modelo de 20 aminoácidos. A principal diferença entre os dois modelos é que, no modelo das interações independentes, as interações entre os monômeros são tomadas aleatoriamente segundo uma distribuição gaussiana de números aleatórios. Cada interação entre um par de monômeros terá seu valor extraído de uma distribuição gaussiana que pode ser calculada mediante a fórmula

$$P(\varepsilon_{ij}) = \frac{1}{\sqrt{2\pi\sigma_B}} e^{-\frac{1}{2}\left(\frac{\varepsilon_{ij}-\varepsilon_0}{\sigma_B}\right)^2}, \quad (4.7)$$

onde os índices i e j , serão tão grandes quanto a extensão da cadeia, ou seja, para uma cadeia de 27 monômeros, será construída uma tabela de interações considerando 27 tipos diferentes de aminoácidos.

A execução do programa com este modelo teve seu tempo de CPU de 1110 min, menos até que o modelo AB, apesar da semelhança com o modelo de 20 aminoácidos.

Como esperado, os resultados são semelhantes aos obtidos com o modelo de 20 aminoácidos, demonstrado na seção anterior, exceto pela curva da figura 4.31 que não apresenta de maneira expressiva picos de calor específico. Já as outras quantidades termodinâmicas se comportam de maneira muito semelhante ao do modelo de 20 letras, como pôde ser visto, o que levaria as mesmas conclusões obtidas anteriormente.

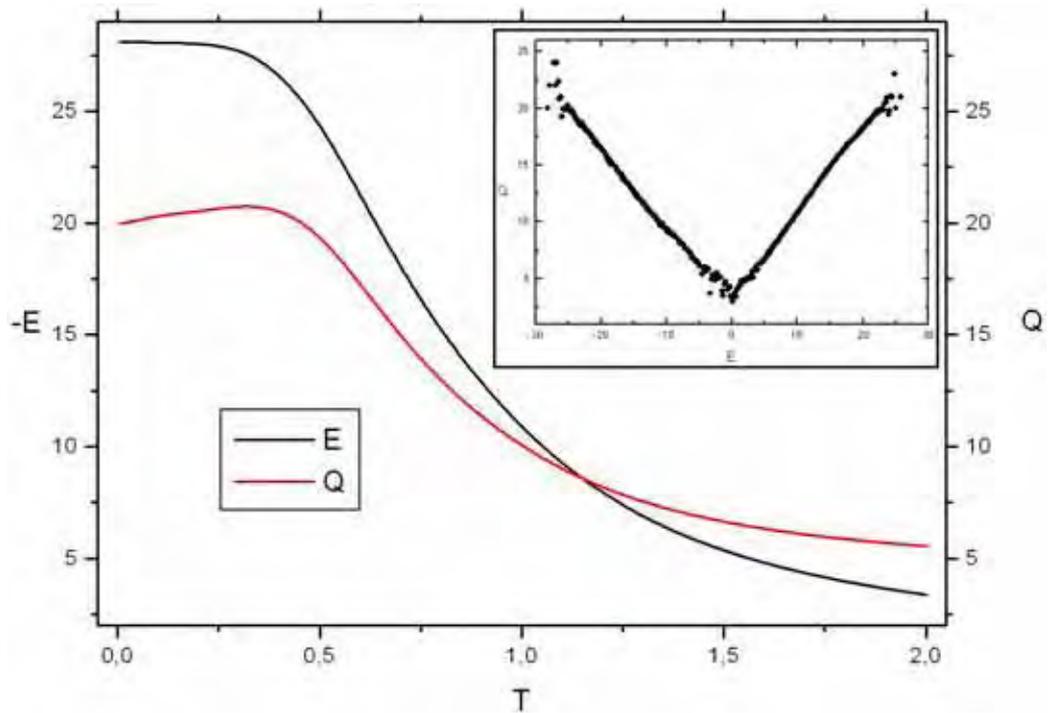


Figura 4.29 – A linha preta representa a energia, enquanto que a vermelha, o número de contatos, ambas em função da temperatura. No *inset* está um gráfico do número de contatos, como função da energia.

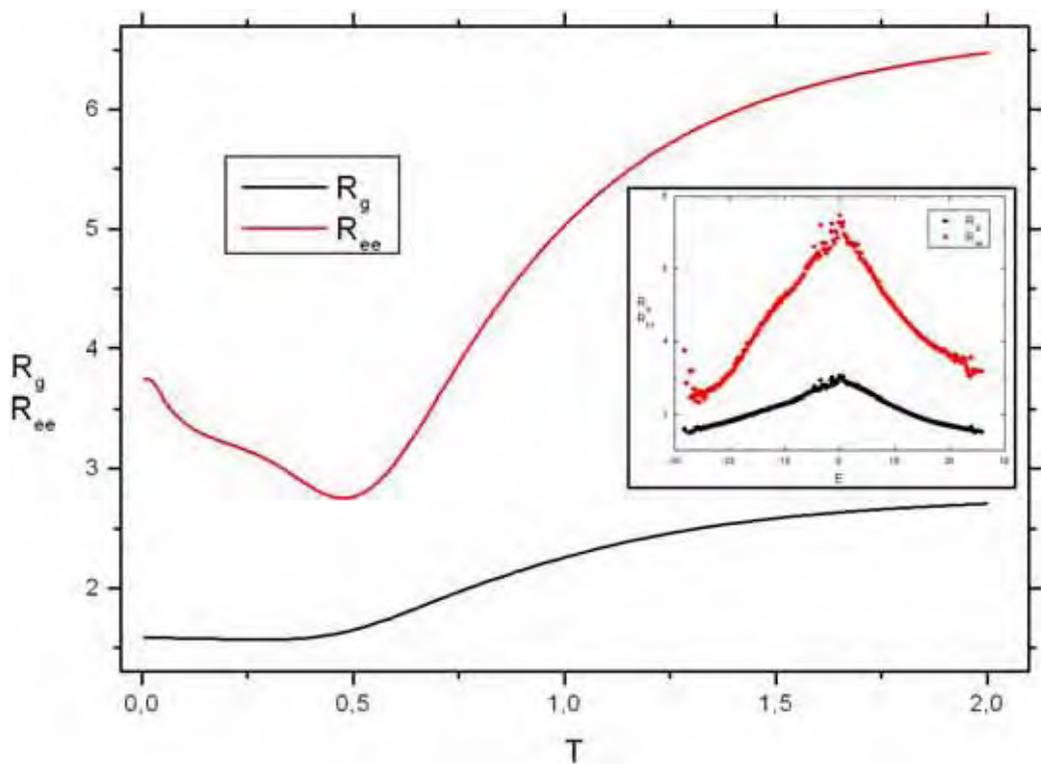


Figura 4.30 – A linha preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas da cadeia, ambas em função da temperatura. No *inset* a cor preta representa o raio de giração, enquanto que a vermelha a distância entre as pontas, ambas as grandezas em função da energia.

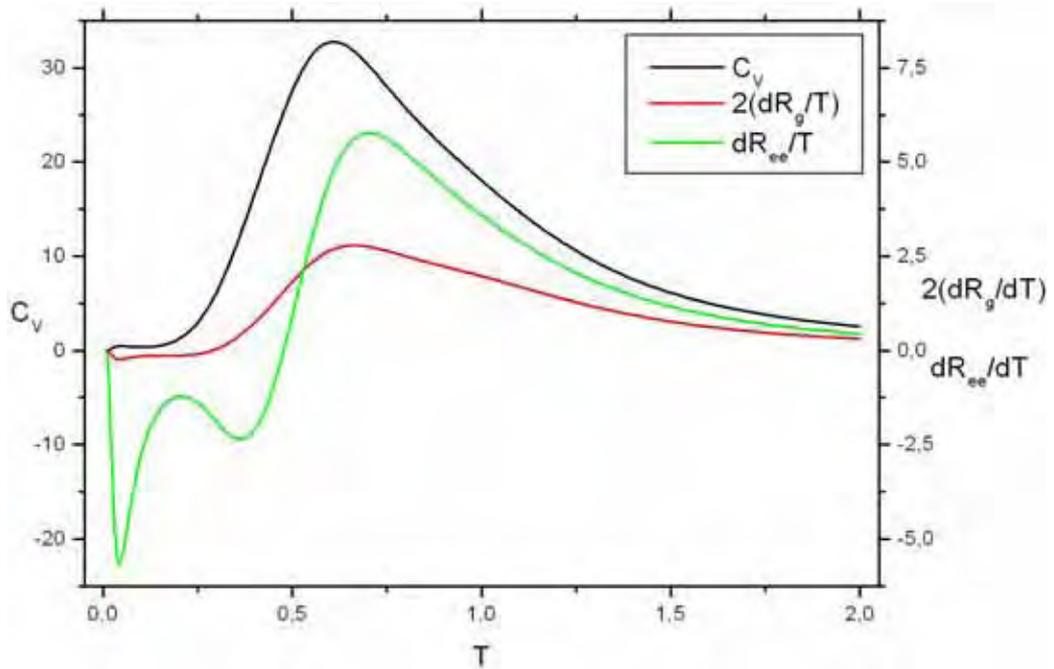


Figura 4.31 – A linha preta representa o calor específico, a vermelha representa a derivada do raio de giração pela temperatura, enquanto que a linha verde, a derivada da distância entre as pontas pela temperatura.

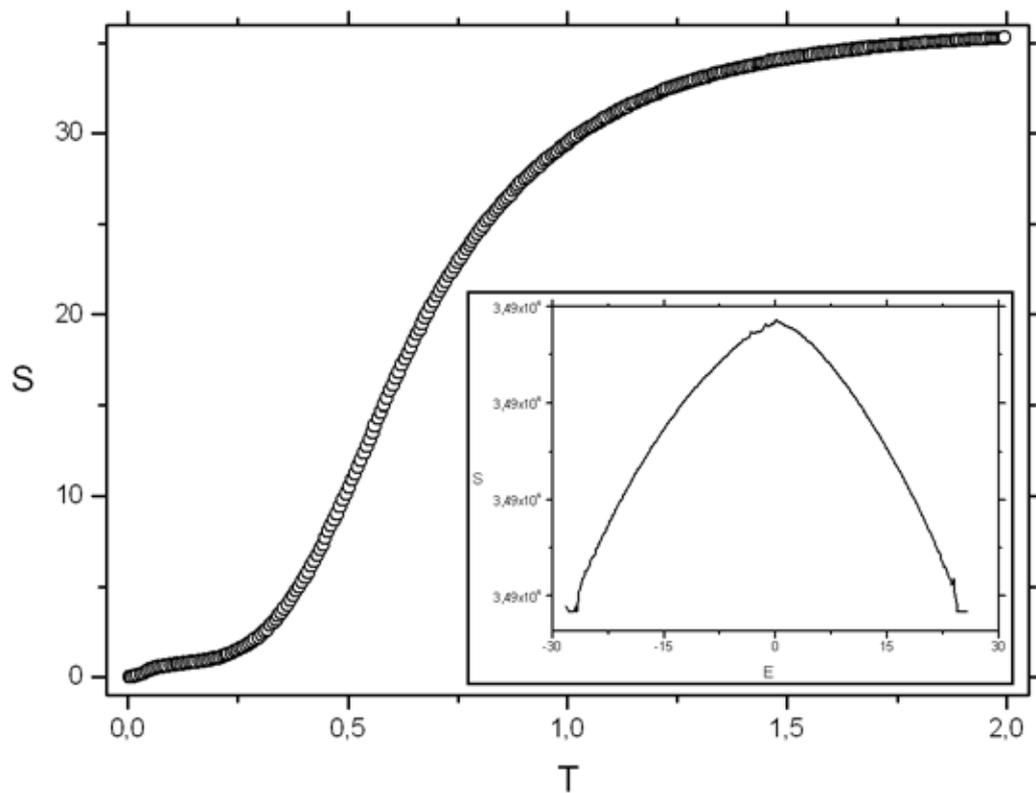


Figura 4.32 – No gráfico principal, tem-se a entropia como função da temperatura. No *inset* dessa figura, se tem a entropia do método WL.

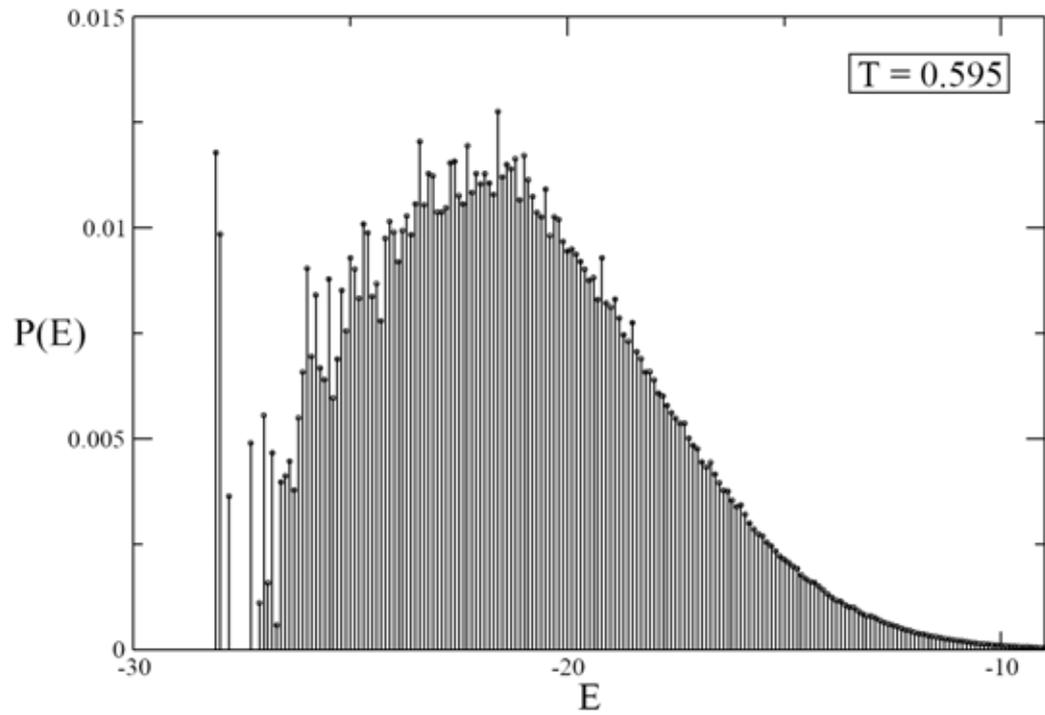


Figura 4.33 – O gráfico acima representa a probabilidade de se encontrar o sistema em uma dada energia, para cada uma das temperaturas indicadas, onde $T=0,595$, é a temperatura crítica.

Conclusões

Ao longo deste trabalho, procurou-se demonstrar a eficiência do método Wang-Landau para o estudo da termodinâmica do enovelamento de cadeias heteropoliméricas. Pôde-se mostrar essa eficiência através dos modelos de rede mais relevantes na área de enovelamento de proteínas, dos quais foram obtidas as grandezas termodinâmicas fundamentais de cadeias de 27 monômeros em uma rede cúbica.

A análise dos resultados permitiu classificar as transições de fase (entendida para sistemas finitos) da cadeia polimérica, verificar a compactação da mesma em consequência da diminuição da temperatura, bem como identificar as fases ordenadas e desordenadas de cada seqüência. Foi possível, portanto, obter as informações termodinâmicas suficientes para um bom entendimento do processo de enovelamento de proteínas.

Apesar de toda a eficiência, o método Wang-Landau não permitiu que a degenerescência do estado fundamental fosse calculada. A razão disso é que o método fornece a densidade de estados relativa e não a absoluta. Como consequência, a entropia apresentada nos resultados numéricos é também relativa.

Essa dificuldade requer que um método independente de se obter a degenerescência do estado fundamental seja introduzido. Seria então um método complementar ao de Wang-Landau, no qual a termodinâmica do enovelamento de proteínas poderia ser obtida completamente.

O método de Wang-Landau mostrou ser, para os atuais computadores pessoais, um método eficiente e rápido pelo menos para cadeias de até 27 monômeros em uma rede cúbica.

Sugestões para trabalhos futuros

- Para melhorar a eficiência do método de Wang-Landau, a densidade de estados poderia ser calculada em função de outros parâmetros de rede, além da energia como o número de contatos nativos, por exemplo.
- A introdução de um método independente de se identificar o estado fundamental seria também uma boa sugestão. Este método complementaria os resultados obtidos pelo método de Wang-Landau.
- Uma sugestão final seria a de implementar futuramente o método de Wang-Landau para modelos fora da rede, permitindo assim uma visão muito mais realista do processo de envelhecimento em si.

Referências Bibliográficas

1. J.M. Thornton, C.A. Orengo, A.E. Todd e F.M.G. Pearl, *J. Mol. Biol.* **293**, 333 (1999).
2. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
3. H.S. Chan e K.A. Dill, *Physics Today* **46**, 24 (1993).
4. H. Frauenfelder e P.G. Wolynes, *Physics Today* **47**, 58 (1994).
5. V.S. Pande, A.Y. Grosberg e T. Tanaka, *Rev. Mod. Phys.* **72**, 259 (2000).
6. Sali, E. Shakhnovich e M. Karplus, *Nature* **369**, 248 (1994).
7. D. Baker, *Nature* **405**, 39 (2000).
8. D.J. Selkoe, *Nature* **426**, 900 (2003).
9. F.E. Cohen e J.W. Kelly, *Nature* **426**, 905 (2003).
10. C.M. Dobson, *Nature* **426**, 884 (2003).
11. B.M. Pepys et al., *Nature* **417**, 254 (2002).
12. K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas e H.S. Chan, *Protein Science* **4**, 561 (1995).
13. E.I. Shakhnovich, *Curr. Opin. Struct. Biol.* **7**, 29 (1997).
14. V.S. Pande, A.Y. Grosberg, T. Tanaka e D. S. Rokhsar, *Curr. Opin. Struct. Biol.* **8**, 68 (1998).

15. J.N. Onuchic, H. Nymeyer, A.E. Garcia, J. Chahine e N.D. Socci, *Adv. Protein Chem* **53**, 87 (2000).
16. L. Mirny e E. Shakhnovich, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 361 (2001).
17. D. Gront, A. Kolinski e J. Skolnick, *J. Chem. Phys.* **115**, 1569 (2001).
18. M. Karplus, *Folding & Design* **2**, s69 (1997).
19. S. Govindarajan e R.A. Goldstein, *Proc. Natl. Acad. Sci. USA* **95**, 5545 (1998).
20. V. Muñoz e W.A. Eaton, *Proc. Natl. Acad. Sci. USA* **96**, 11311 (1999).
21. J. Schonbrun e K.A. Dill, *Proc. Natl. Acad. Sci. USA* **100**, 12678 (2003).
22. E.I. Shakhnovich, *Phys. Rev. Lett* **72**, 3907 (1994).
23. F. Seno, M. Vesdruscolo, A. Maritan e J. R. Banavar, *Phys. Rev. Lett.* **77**, 1901 (1996).
24. R.I. Dima, J. R. Banavar, M. Cieplak e A. Maritan, *Proc. Natl. Acad. Sci. USA* **96**, 4904 (1999).
25. K. Binder, *The Monte Carlo Method in Condensed Matter Physics*, 2ns edn, Topics in Applied Physics, vol 71 (Springer, 1991).
26. O. Kolinski e J. Skolnick, *Polymer* **45**, 511 (2003).
27. Y.Peng, U.H.E. Hansmann e N.A. Alves, *J. Chem. Phys.* **118**, 2374 (2003).
28. A.E. Garcia e J.N. Onuchic, *Proc. Natl. Acad. Sci. USA* **100**, 13898 (2003).
29. F. Wang e D.P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
30. B.A. Berg e T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
31. J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993)., J.R. Drugowich de Felício e V. Libero, *Am. J. Phys.* **64**, 1281 (1996).
32. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.N. Teller e E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
33. D.P.Landau e K.Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press.
34. S.R.A. Salinas, *Introdução Mecânica Estatística*, Edusp, (1999).

35. N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, HOLT, RINEHART AND WINSTON, (1976).
36. R.H. Sweendsen e J.S. Wang, *Phys. Rev. Lett* **58**, 86 (1987).
37. U. Wolff, *Phys. Rev. Lett.* **62**, 361(1989).
38. P. D. Beale, *Phys. Rev. Lett.* **76**, 78 (1996).
39. A. E. Ferdinand e M. E. Fisher, *Phys. Rev.* **185**, 832 (1969).
40. C.B. Afinsen, *Science* **181**, 223 (1973).
41. C. Levinthal, *J. Chem. Phys.* **65**, 44 (1968).
42. N.D. Socci e J.N. Onuchic, *J. Chem. Phys.* **103**, 4732 (1995).
43. A.M. Ferrenberg e R.H. Swendsen, *Phys.Rev.Lett.* **63**, 1195 (1989).
44. M. Hao e H.A. Scheraga, *J. Phys. Chem* **98**, 4940 (1994).
45. L.W. Lee e J.S. Wang, *Phys.Rev.E* **64**, 056112
46. P.M.C. de Oliveira, T.J. Penna e H.J. Hermann, *Braz. J. Phys.***26**, 677(1996).
47. M. Bachman e W. Janke, *J.Chem.Phys.***120**, 6779 (2004).
48. N. Rathore e J.J. de Pablo, *J.Chem.Phys.* **116**, 7225 (2002).
49. V.S. Pande, A.Y. Grosberg e T. Tanaka, *Biophys. J.* **73**, 3192 (1997).
50. H. Li, R.H. Helling, C. Tang e N. Wingreen, *Science.* **273**, 666 (1996).
51. N.D. Socci e J.N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).
52. S. Miyazawa e R.L. Jernigan, *Macromolecules* **18**, 534 (1985).
53. E.I. Shakhnovich e A. Gutin, *Biophys. Chem.* **34**, 187 (1989).