



UNIVERSIDADE FEDERAL DE MATO GROSSO  
INSTITUTO DE FÍSICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

FABRÍCIO DE SÁ HORA SANTOS

SIMULAÇÃO COMPUTACIONAL DO MODELO HP GENERALIZADO  
PARA PROTEÍNAS

CUIABÁ

2016

FABRÍCIO DE SÁ HORA SANTOS

SIMULAÇÃO COMPUTACIONAL DO MODELO HP GENERALIZADO PARA  
PROTEÍNAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Mato Grosso, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Magnetismo e Transição de Fase.

Orientador: Prof. Dr. Paulo Henrique Lana Martins.

Coorientador: Prof. Dr. Michael Bachmann.

CUIABÁ  
2016

FABRÍCIO DE SÁ HORA SANTOS

SIMULAÇÃO COMPUTACIONAL DO MODELO HP GENERALIZADO PARA  
PROTEÍNAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Física da Universidade Federal de Mato Grosso, como requisito parcial para a obtenção do Título de Mestre em Física. Área de Concentração: Magnetismo e Transição de Fase.

Aprovada em 31/03/2016.

BANCA EXAMINADORA

---

Prof. Dr. Paulo Henrique Lana Martins (Orientador)  
Universidade Federal de Mato Grosso (UFMT)

---

Prof. Dr. Marcio Fernando Cornélio  
Universidade Federal de Mato Grosso (UFMT)

---

Prof. Dr. Dorílson Silva Cambuí  
Secretaria de Educação do Estado de Mato Grosso  
(SEDUC-MT)

Dados Internacionais de Catalogação na Publicação  
Universidade Federal de Mato Grosso  
Biblioteca do Curso de Física

---

A000p SANTOS, F.S.H.  
Simulação computacional do modelo HP generalizado para proteínas / Fabrício de Sá Hora Santos. – Cuiabá, 2016.  
47.:il.

Dissertação (mestrado) - Universidade Federal de Mato Grosso, Instituto de Física, Cuiabá, 2016.

Área de Concentração: Magnetismo e Transição de Fase.  
Orientação: Prof. Dr. Paulo Henrique Lana Martins.

1. Monte Carlo. 2. HP. 3. PERM. 4. Transições de Fase. I. Título.

CDD:000.0

---

*A minha família*

## AGRADECIMENTOS

Gostaria de agradecer primeiramente a minha família pelos suportes, incentivos e sugestões. Agradecer também aos meus amigos que estavam sempre presentes nos momentos difíceis e de diversão. Um agradecimento a Universidade Federal de Mato Grosso em particular aos professores do Instituto de Física, que através deles, pude ter uma boa formação. E por fim, gostaria de agradecer ao meu orientador prof Paulo e meu coorientador prof Bachmann, pelo respeito e confiança.

## RESUMO

Consideramos o modelo hidrofóbico-polar generalizado para proteínas nas redes quadrada e cúbica. Além da atração entre monômeros hidrofóbicos não ligados, o atual modelo também leva em conta uma interação entre unidades hidrofóbicas e polares. Utilizando o algoritmo PERM (*pruned-enriched Rosenbluth method*), investigamos uma sequência particular composta por 42 monômeros e que tem sido proposta para simular as propriedades físicas da  $\beta$ -hélice paralela de *pectato liase C*. Para cada temperatura, o número total de cadeias geradas varia de  $10^6$  à  $10^7$ . Observáveis físicos como o calor específico, a energia total, a distância de ponta a ponta, o raio de giração, e o número médio de interações do tipo hidrofóbico-hidrofóbico e hidrofóbico-polar, são avaliados para diferentes valores do parâmetro  $s$ , que corresponde à razão entre a energia de ligação entre um monômero hidrofóbico e um polar e a energia de ligação entre dois monômeros hidrofóbicos. Um pseudo diagrama de fases no espaço da temperatura  $T$  e do parâmetro  $s$  é construído.

**Palavras-chave:** Monte Carlo. HP. PERM. Transições de Fase.

## ABSTRACT

We consider a generalized hydrophobic-polar model for proteins on square and cubic lattices. Besides the attraction between nonbonded hydrophobic monomers, the present model also takes into account an interaction between hydrophobic and polar units. By using the pruned-enriched Rosenbluth method (PERM), we investigate a specific polymer sequence composed of 42 monomers that has been proposed to simulate the physical properties of the parallel  $\beta$ -helix of pectate lyase *C*. For each temperature, the total number of generated chains varies from  $10^6$  to  $10^7$ . Physical observables such as specific heat, total energy, end-to-end distance, radius of gyration, and the average number of hydrophobic-hydrophobic and hydrophobic-polar contacts, are evaluated for different values of the parameter  $s$ , which corresponds to the ratio between the hydrophobic-polar and the hydrophobic-hydrophobic contact energies. Eventually, a pseudo-phase diagram in the space of temperature and the ratio of contact energy scales is constructed.

**Keywords:** Monte Carlo. HP. PERM. Phase Transitions.

## LISTA DE FIGURAS

Figura 1 – Exemplo da estrutura geral de um aminoácido. . . . .	14
Figura 2 – Enovelamento de uma proteína globular. . . . .	16
Figura 3 – Exemplos de conformações do modelo HP em 2D. . . . .	18
Figura 4 – Amostragem tendenciosa do caminho aleatório para $n = 50$ e $10^5$ amostras e com probabilidade de $p = 0,2, 0,5$ e $0,8$ , de se deslocar para esquerda. $S(x)$ corresponde ao número de amostras com $x$ passos para esquerda. . . . .	23
Figura 5 – Diagrama de fases de um fluido simples no espaço $P - T$ . . . . .	27
Figura 6 – Isotermas de um fluido simples no diagrama $P - V$ . . . . .	27
Figura 7 – Representação de um homopolímero com 64 monômeros. (a) fase cristalina; (b) globular; (c) estendida. Figura retirada da Ref. [24]. . . . .	28
Figura 8 – Exemplo do <i>Self Avoiding Walk</i> em 2D. . . . .	31
Figura 9 – Demonstração da eficiência do método de Rosenbluth em relação à amostragem simples do SAW. . . . .	33
Figura 10 – Representação vetorial de uma cadeia polimérica linear. . . . .	34
Figura 11 – Energia potencial de diferentes gases em função da distância em $\text{\AA}$ ( $\text{\AA}$ ). . . . .	35
Figura 12 – Calor específico em função da temperatura para $0 \leq s \leq 1$ . . . . .	37
Figura 13 – Calor específico (à esquerda) e energia média (à direita) como funções da temperatura para alguns valores do parâmetro $s$ . As barras de erro, se não forem visíveis, são menores do que os símbolos. . . . .	38
Figura 14 – Esquerda: número médio de contatos do tipo hidrofóbico-hidrofóbico $n_{HH}$ . Direita: número médio de interações hidrofóbico-polar $n_{HP}$ . As barras de erro são da ordem do tamanho dos símbolos. . . . .	38
Figura 15 – A proteína com 18 contatos $HH$ e 3 contatos $HP$ . . . . .	39
Figura 16 – Valores médios da distância de ponta a ponta (esquerda) e do raio de giração (direita). Para melhor visualização, barras de erro são mostradas somente para $s = 0,0$ . Para os outros valores de $s$ , as barras de erro têm tamanhos semelhantes. . . . .	40
Figura 17 – A proteína com um núcleo hidrofóbico e uma “blindagem” polar. . . . .	40

Figura 18 – Pseudo diagrama de fases no espaço $s - T$ . Nas conformações mostradas, um monômero H é representado por um círculo de cor vermelha, enquanto monômeros P são representados em cor preta. . . . .	41
Figura 19 – Calor específico (à esquerda) e energia média (à direita) como funções de temperatura para alguns valores do parâmetro $s$ . As barras de erro, se não forem visíveis, são menores do que os símbolos. . . . .	42
Figura 20 – Esquerda: número médio de contatos do tipo hidrofóbico-hidrofóbico $n_{HH}$ . Direita: número médio de interações hidrofóbico-polar $n_{HP}$ . As barras de erro são da ordem do tamanho símbolos. . . . .	43
Figura 21 – Valores médios da distância de ponta a ponta (esquerda) e do raio de giração (direita). Para melhor visualização, barras de erro são mostradas somente para $s = 0, 0$ . Para os outros valores de $s$ , as barras de erro têm tamanhos semelhantes. . . . .	43
Figura 22 – Pseudo diagrama de fases no espaço $s - T$ . . . . .	44
Figura 23 – Esboço do diagrama de fases no espaço $s - T$ , no cenário contendo 3 fases distintas. . . . .	44
Figura 24 – Esboço do diagrama de fases no espaço $s - T$ , no cenário contendo 4 fases distintas. . . . .	45

## SUMÁRIO

1	INTRODUÇÃO . . . . .	12
1.1	Motivação . . . . .	12
1.2	Objetivos . . . . .	13
1.3	Estrutura da Dissertação . . . . .	13
2	FUNDAMENTAÇÃO TEÓRICA . . . . .	14
2.1	Conceitos biológicos . . . . .	14
2.1.1	<i>Proteína</i> . . . . .	14
2.2	Modelo hidrofóbico-polar (HP) . . . . .	15
2.3	Conceitos de mecânica estatística . . . . .	19
2.3.1	<i>Estados microscópicos do sistema</i> . . . . .	19
2.3.2	<i>Espaço de fase</i> . . . . .	19
2.3.3	<i>Ensemble estatístico</i> . . . . .	19
2.3.4	<i>Média no ensemble</i> . . . . .	19
2.3.5	<i>Distribuição de Boltzmann e função de partição</i> . . . . .	20
2.3.6	<i>Energia interna e calor específico</i> . . . . .	21
2.3.7	<i>Média de ensemble por amostragem</i> . . . . .	22
2.3.7.1	<i>Amostragem simples e amostragem por importância</i> . . . . .	22
2.4	Transições de fase . . . . .	24
2.4.1	<i>Fases termodinâmicas</i> . . . . .	24
2.4.2	<i>Transições configuracionais em polímeros</i> . . . . .	28
3	METODOLOGIA . . . . .	30
3.1	Simulações de Monte Carlo . . . . .	30
3.2	PERM . . . . .	30
3.3	Grandezas analisadas . . . . .	33
3.3.1	<i>Distância Ponta a Ponta</i> . . . . .	33
3.3.2	<i>Raio de giração</i> . . . . .	34
3.3.3	<i>Interações HH e HP</i> . . . . .	34
3.4	Detalhes das simulações . . . . .	36
4	RESULTADOS . . . . .	37
4.1	Em duas dimensões . . . . .	37
4.2	Em três dimensões . . . . .	40
5	CONCLUSÃO . . . . .	46

REFERÊNCIAS . . . . .	47
-----------------------	----

# 1 INTRODUÇÃO

## 1.1 Motivação

Mais de 35 milhões de pessoas estão contaminadas pelo *human immunodeficiency virus* (HIV)<sup>1</sup> e estima-se que 240 milhões de pessoas estejam cronicamente infectadas com hepatite B (HBV)<sup>2</sup>. Estes são apenas dois exemplos de epidemias que atingem todo o mundo devido a infecções por vírus. Tipicamente, os vírus consistem de um ácido nucleico enovelado em uma forma bem compacta, com uma casca proteica externa. Como as proteínas nessa casca fazem parte do processo de fusão do vírus com a célula hospedeira, o estudo do funcionamento das proteínas torna-se relevante para o entendimento desse processo.

Há fortes evidências de que a estrutura configuracional de uma proteína está relacionada com sua função biológica. Muitas vezes, se uma proteína sofre uma mudança em sua conformação (ocasionada por exemplo, por uma variação de temperatura), ela perde sua funcionalidade. Algumas doenças neurodegenerativas, como a de Alzheimer, têm sido associadas a alterações estruturais de alguma proteína, com a formação de agregados que se tornam neurotóxicos [1].

De um ponto de vista mais fundamental, a famosa frase de Francis Crick<sup>3</sup> chamada de Dogma Central da Biologia, traduz bem a importância das proteínas:

*“DNA faz RNA, RNA faz proteína e as proteínas nos fazem”.*

De fato, as proteínas desempenham um papel crucial em processos biológicos e, por isso, as pesquisas biomoleculares vem tendo bastante destaque nos últimos anos. Deve-se ressaltar que as proteínas são estruturas complexas, formadas por uma cadeia linear principal contendo tipicamente entre 50 e 3000 aminoácidos, além de cadeias laterais conectadas à cadeia principal. Entretanto, modelos simplificados são, em alguns casos, bastante úteis para o entendimento de aspectos gerais dessas transições estruturais. Neste trabalho, analisamos as transições configuracionais no modelo hidrofóbico-polar generalizado, que será discutido oportunamente.

---

<sup>1</sup>UNAIDS/OMS *AIDS Epidemic Update: July 2014*.

<sup>2</sup>OMS Fact Sheet No. 204 (2015).

<sup>3</sup>Francis Crick foi um cientista conhecido por co-descobrir a molécula de DNA em 1953 e também chegou a ganhar o prêmio Nobel de Fisiologia ou Medicina de 1962.

## 1.2 Objetivos

O objetivo desta dissertação é obter, através do método de Monte Carlo, com a implementação do algoritmo de crescimento PERM, as “transições de fase” de uma cadeia particular composta por 42 monômeros. O modelo considerado é uma generalização do modelo HP (hidrofóbico-polar) que inclui interações entre dois monômeros hidrofóbicos e ainda interações entre um hidrofóbico e um polar. Apesar de simplificado, esse modelo tem sido utilizado para investigar aspectos gerais das transições estruturais. O diagrama de fases é construído a partir da análise do comportamento de grandezas como calor específico, energia total, distância de ponta a ponta, raio de giração e número médio de interações HH e HP entre os monômeros.

## 1.3 Estrutura da Dissertação

Esta dissertação encontra-se dividida em capítulos, iniciando com esta introdução (**1 INTRODUÇÃO**).

Em seguida, no capítulo **2 FUNDAMENTAÇÃO TEÓRICA**, discutem-se inicialmente alguns conceitos biológicos relevantes para essa dissertação e, logo após, explica-se o modelo HP. Alguns fundamentos da Mecânica Estatística são apresentados e o capítulo se encerra com uma discussão sobre transições de fases.

Continuando, no capítulo **3 METODOLOGIA**, explica-se sobre a simulação de Monte Carlo, o algoritmo utilizado no trabalho (PERM), as quantidades físicas obtidas e como foram estimados os erros.

No capítulo **4 RESULTADOS**, apresentam-se os resultados obtidos. Grandezas como calor específico, energia média, raio de giração etc. são mostradas para vários valores da temperatura  $T$  e do parâmetro  $s$ , que corresponde à razão entre as energias de interação HP e da interação HH. Essas quantidades são utilizadas para a construção de um diagrama de fases no planos  $s - T$ , possibilitando identificar, para um dado valor do conjunto de parâmetros  $(s, T)$ , qual a fase estrutural correspondente. De acordo com nosso conhecimento, esses resultados são inéditos para essa versão generalizada do modelo HP.

Finalmente, no capítulo **5 CONCLUSÃO**, são apresentadas as conclusões, onde são evidenciados os resultados obtidos e a perspectiva para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Conceitos biológicos

#### 2.1.1 Proteína

O nome “proteína” vem do grego *proteios*, que significa “primeiro”, devido à importância fundamental que estes compostos têm para a vida. As proteínas são responsáveis por uma ampla gama de funções, tais como: controle dos processos de transporte, construção e estabilização de estruturas celulares, formação de enzimas que catalisam as reações bioquímicas e *etc* [2]. Proteínas são moléculas lineares de polímero formadas por moléculas de aminoácidos. Na natureza há apenas 20 tipos de aminoácidos dos quais as proteínas são formadas [1, 3]. Portanto, para compreender a natureza das proteínas, é importante entender a natureza dos aminoácidos [1, 4].

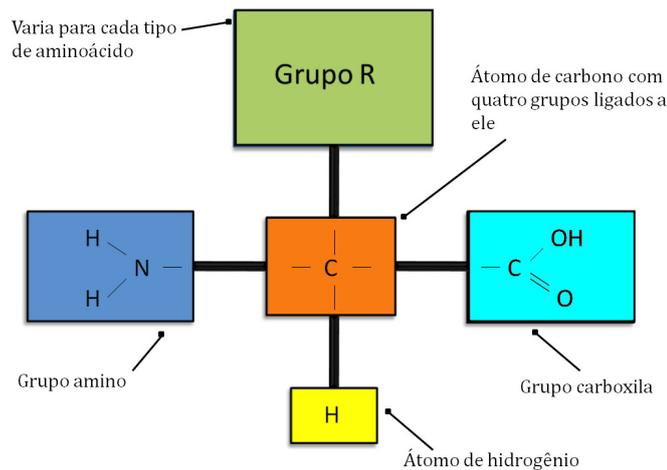


Figura 1: Exemplo da estrutura geral de um aminoácido.

Um exemplo de aminoácido é ilustrado pela Fig. 1. Todo aminoácido é constituído de, no mínimo, dois grupos funcionais: amina ( $-NH_2$ ) e o grupo carboxila ( $-COOH$ ) ligados ao átomo de carbono [4]. O grupo R mostrado na Fig. 1, representa a continuação do aminoácido, isto é, o que muda de um aminoácido para outro. Um exemplo, é a diferença entre os aminoácidos Glicina ( $H - CH(NH_2) - COOH$ ) e Alanina ( $CH_3 - CH(NH_2) - COOH$ ). Assim pode-se verificar que as diferenças entre eles são  $H-$  (o primeiro termo da fórmula estrutural da Glicina) e  $CH_3$  (o primeiro termo da fórmula estrutural da Alanina), que corresponde justamente ao grupo R.

Para que esses se liguem para formar uma proteína ocorrem uma pequena modificação em suas composições devido a união entre os aminoácidos adjacentes. No momento que a proteína está sendo sintetizada, isto é, os aminoácidos estão se juntando

a proteína, o radical carboxila perde um agrupamento  $-OH$ , deixando uma ligação livre. E simultaneamente, o radical amina de outro aminoácido perde um átomo de hidrogênio  $-H$ , ficando também com uma ligação livre. Assim o aminoácido se junta a proteína através de um outro aminoácido adjacente. Este tipo de ligação covalente recebe o nome de ligação peptídica.

As proteínas são uma classe de polímeros. Os polímeros, por sua vez, são formados por unidades menores chamadas monômeros (que são os aminoácidos). Se em uma molécula os monômeros possuem características semelhantes, chamamos de homopolímero, em contraste com o heteropolímero, que é formado por tipos diferentes de monômeros [5].

Alterações na forma das proteínas podem ser causadas por fatores como temperatura. Quando uma proteína está submetida a uma alta temperatura, fala-se em desnaturação da proteína e, conseqüentemente, ela se torna inativa [6]. Assim, quando uma pessoa está com febre, os médicos preferem inicialmente abaixar a febre do que descobrir a causa. Com a alta temperatura, ocorre a desnaturação de enzimas que realizam funções vitais, como enzimas que auxiliam no processo respiratório [7].

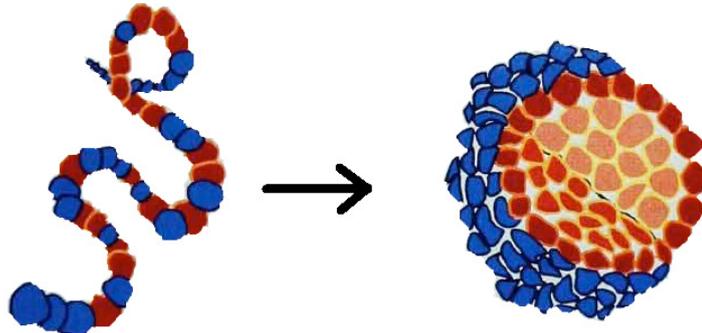
## 2.2 Modelo hidrofóbico-polar (HP)

Os termos hidrofóbico e polar (ou hidrofílico) referem-se à distribuição de cargas em uma molécula. Se não há regiões de alta ou baixa densidade eletrônica na molécula, ela é chamada de hidrofóbica. Esse termo é adequado, uma vez que essas moléculas não se dissolvem na água. Por outro lado, se há regiões com uma carga parcial positiva ou negativa, a molécula é dita polar ou hidrofílica, uma vez que se dissolve facilmente na água. Como as proteínas geralmente se encontram em um meio aquoso (que denominamos solvente), as interações com as moléculas de água são relevantes. Portanto, considerando a hidrofobicidade, podemos agrupar os 20 tipos existentes de aminoácidos em duas classes: hidrofóbica e polar. Essa simplificação deu origem ao modelo HP, proposto por Lau e Dill em 1989 [8], que é capaz de descrever qualitativamente as transições estruturais observadas.

Os aminoácidos hidrofóbicos de uma proteína em um solvente aquoso tendem a ficar afastados do solvente formando um núcleo hidrofóbico. No exterior da molécula de proteína, os aminoácidos polares tendem a formar um escudo de blindagem que ficará em contato com o solvente para minimizar a exposição dos seus monômeros  $H$  (hidrofóbicos) [9, 10]. Este fenômeno é conhecido como efeito hidrofóbico [11]. O modelo HP é, portanto, um protótipo para imitar o fenômeno do efeito hidrofóbico e modelar de forma simples as transições configuracionais das proteínas.

Cada proteína apresenta uma sequência pré-determinada de aminoácidos, como exemplificado na Fig. 2. Os monômeros representados em azul são os que apresentam características resultantes polares (P) e em vermelho os monômeros que apresentam características hidrofóbicas (H). A Fig. 2 mostra, ainda, duas conformações bem distintas. À esquerda, tem-se uma forma estendida, chamada de *random-coil*, enquanto à direita mostra-se uma configuração enovelada. Essas conformações, bem como as transições entre elas, serão discutidas em uma seção posterior.

Figura 2: Enovelamento de uma proteína globular.



Fonte: DILL, Ken; CHAN, H. Sun. *Physics today*, v.46, n.72, 24-32, 1993.

A forma mais simples do modelo HP, entretanto, apresenta alguns problemas. Em geral, o estado fundamental de uma sequência HP é altamente degenerado, o que parece não ocorrer com proteínas reais. Com a inclusão de interações além daquelas entre monômeros hidrofóbicos, algumas simetrias são quebradas e a degenerescência se torna menor. Isso significa que o número de sequências que apresentam um único estado fundamental (conhecidas por *designing sequences*) aumenta [12]. Baseado no trabalho de Miyazawa-Jernigan [13], que apresenta as energias de contato entre diferentes aminoácidos reais, Tang propôs, em 2000, uma versão modificada do modelo HP, mais realista, que leva em consideração uma interação atrativa entre monômeros H e monômeros P [14]. As intensidades das interações (ambas atrativas) sugeridas por Tang são, respectivamente  $\epsilon_{HH} = 1$  e  $\epsilon_{HP} \approx 0.435$ . Esses valores resultaram de uma análise das energias entre aminoácidos hidrofóbicos e entre hidrofóbicos e polares, de acordo com os resultados de Miyazawa e Jernigan [13].

No presente trabalho, generalizamos o modelo HP de forma a incluir o parâmetro  $s$ , que corresponde à razão entre as energias dos contatos:

$$s \equiv \frac{\epsilon_{HP}}{\epsilon_{HH}}, \quad (2.1)$$

onde  $\epsilon_{HP}$  e  $\epsilon_{HH}$  são as intensidades da energia de interação  $HP^1$  e  $HH^2$ , respectivamente. Para mantermos a característica original do modelo HP de que a interação hidrofóbica é a força dominante para o enovelamento de proteínas, mantivemos o parâmetro  $s$  no intervalo  $0 \leq s \leq 1$ . O limite  $s = 0$  equivale ao modelo HP usual, enquanto  $s = 1$  significa que as interações  $HH$  e  $HP$  possuem a mesma intensidade.

Diante dessas considerações, resumimos abaixo as principais características do modelo investigado.

- Com base na hidrofobicidade, os aminoácidos são agrupados em duas classes: hidrofóbicos e polares.
- Uma cadeia é formada por uma dada sequência de aminoácidos, o que significa que a ordem em que os monômeros estão dispostos é importante. Assim, diferentes sequências correspondem a cadeias distintas.
- Os aminoácidos ocupam os sítios de uma rede regular. Neste trabalho, consideramos as redes quadrada e cúbica.
- As interações ocorrem entre monômeros adjacentes no espaço (que ocupam sítios vizinhos na rede), mas que não estão ligados de forma covalente (isto é, que não são monômeros consecutivos da sequência).
- A interação hidrofóbica ( $\epsilon_{HP} \equiv 1$ , em unidade de energia) é a força dominante no enovelamento das proteínas [15], uma vez que consideramos o parâmetro  $s$  no intervalo entre 0 e 1.

A energia total de uma dada cadeia é expressa por

$$E = -\epsilon_{HH}n_{HH} - \epsilon_{HP}n_{HP}, \quad (2.2)$$

onde  $n_{HH}$  e  $n_{HP}$  são, respectivamente, o número de interações  $HH$  e  $HP$ . Em termos de  $s$ , podemos reescrevê-la como

$$E = -(n_{HH} + s n_{HP}). \quad (2.3)$$

Observemos que, para  $s = 0$ , que corresponde ao modelo HP original, o estado fundamental ocorre quando  $n_{HH}$  é máximo. À medida que o valor de  $s$  aumenta, as ligações  $HP$  se tornam relevantes.

---

<sup>1</sup>Nesta dissertação utiliza-se o termo  $HP$  em itálico para indicar **interação** entre um monômero hidrofóbico e um polar.

<sup>2</sup>Nesta dissertação utiliza-se o termo  $HH$  em itálico para indicar **interação** entre dois monômeros hidrofóbicos.

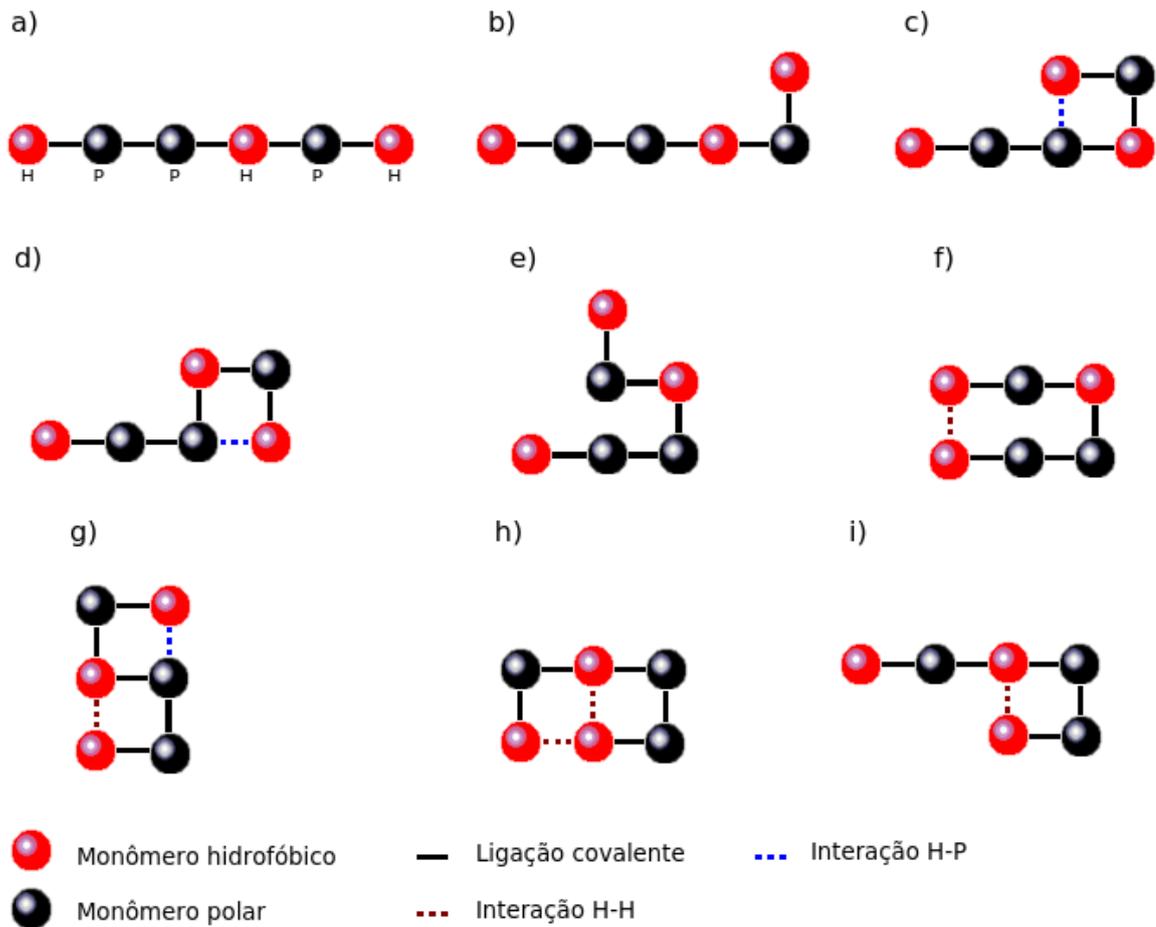


Figura 3: Exemplos de conformações do modelo HP em 2D.

Como exemplo, a Fig. 3 mostra algumas possíveis conformações de uma proteína composta por 6 monômeros em uma rede quadrada. A cor<sup>3</sup> vermelha representa um monômero do tipo hidrofóbico e a preta do tipo polar. A sequência dessa molécula específica é dada por HPPHPH. Para as conformações *a)* e *b)*, temos a molécula estendida, com um pequena mudança na posição do último monômero da direita. Ambas as configurações apresentam energia  $E = 0$ , uma vez que  $n_{HH} = 0$  e  $n_{HP} = 0$ . As conformações *c)* e *d)* têm uma interação do tipo *HP*. Se considerarmos  $s = 1$ , a energia total dessa configuração será  $E = -1$ , devido à interação *HP* (ilustrada com uma linha pontilhada azul). Por outro lado, para  $s = 0.4$ , teríamos  $E = -0.4$ . A configuração *e)* tem  $E = 0$ , enquanto as conformações *f)* e *i)* apresentam uma única interação do tipo *HH* (representada pela linha pontilhada vermelha). Por fim, as conformações *g)* e *h)* têm a mesma energia total se  $s = 1$ . Nesse caso, ambas configurações são estados fundamentais da molécula corresponde à sequência HPPHPH.

<sup>3</sup>Define-se a partir de agora, que as cores vermelha e preta representam, respectivamente, monômeros do tipo hidrofóbico e polar.

## 2.3 Conceitos de mecânica estatística

### 2.3.1 *Estados microscópicos do sistema*

Um sistema físico de partículas é governado pelas leis da mecânica, que fornecem os meios para a especificação de um estado microscópico. Dependendo do modelo investigado e do nível de detalhes que estamos interessados, devemos levar em conta se uma descrição clássica é suficiente ou se é necessária uma descrição quântica. Na Mecânica Clássica, um estado microscópico (ou microestado) é caracterizado pelo conjunto de valores, num dado instante, de todas as coordenadas generalizadas e momentos conjugados dos constituintes microscópicos da amostra. Na Mecânica Quântica, o microestado é descrito pela função de onda do sistema [16, 17].

### 2.3.2 *Espaço de fase*

O espaço de fase é um conceito da Mecânica Clássica que corresponde ao espaço formado pelo conjunto de todos os pontos cujas coordenadas são as coordenadas generalizadas e seus momentos conjugados [18].

### 2.3.3 *Ensemble estatístico*

O conjunto dos pontos do espaço de fase, acessíveis a um sistema clássico, ou o conjunto dos auto-estados de um sistema quântico, constituem um *ensemble estatístico*. Uma interpretação alternativa de um ensemble é considerar um número muito grande  $N$  de cópias do sistema, cada uma dessas cópias representando um microestado acessível. No presente trabalho, construímos o ensemble gerando aleatoriamente diferentes configurações em que a cadeia polimérica pode ser encontrada. Essas configurações são obtidas a partir de uma distribuição de probabilidade (por exemplo, a distribuição de Boltzmann). Assim, o ensemble estatístico é constituído pelo conjunto dos estados microscópicos, aos quais são associados determinados pesos probabilísticos [17].

### 2.3.4 *Média no ensemble*

Considerando o espaço de fases clássico, obtém-se o valor médio de uma determinada quantidade  $A$ , denotado por  $\langle A \rangle$ , como a integral sobre todos os pontos do espaço de fase [16, 17],

$$\langle A \rangle = \frac{\int A(q, p) \rho(q, p) d\Gamma}{\int \rho(q, p) d\Gamma}, \quad (2.4)$$

onde  $d\Gamma = dq_1 dq_2 \dots dq_s dp_1 dp_2 \dots dp_s$ ,  $A(q, p)$  é o valor de  $A$  no ponto  $(q, p)$  do espaço de fase e  $\rho(q, p)$  é a densidade de pontos no espaço de fase, de forma que  $\rho(q, p) dq dp$  corresponde ao número de pontos com coordenadas entre  $q$  e  $q + dq$  e  $p$  e  $p + dp$ . Para o caso discreto,

a média é obtida a partir da soma sobre todos os  $N$  elementos constituintes do ensemble:

$$\langle A \rangle = \frac{\sum_{\alpha=1}^N A(\alpha)\rho(\alpha)}{\sum_{\alpha=1}^N \rho(\alpha)}. \quad (2.5)$$

sendo  $\rho(\alpha)$  o peso atribuído ao elemento  $\alpha$ .

### 2.3.5 Distribuição de Boltzmann e função de partição

Consideremos um sistema termodinâmico  $S$  em contato com um reservatório térmico, de paredes diatérmicas, fixas e impermeáveis. O reservatório térmico  $R$  é muito grande em relação ao sistema  $S$ . Quando o sistema composto  $R + S$  estiver em equilíbrio térmico, terá energia total  $E_0$ . Portanto, a probabilidade  $P_n$  de encontrar o sistema  $S$  em um estado microscópico  $n$  será [18],

$$P_n = c \Omega_R(E_0 - E_n), \quad (2.6)$$

onde  $c$  é uma constante de normalização,  $E_n$  é a energia interna do sistema  $S$  no estado microscópico  $n$  e  $\Omega_R$  é o número de estados microscópicos acessíveis ao reservatório térmico  $R$  com energia  $E$ . Como o reservatório  $R$  é muito maior que  $S$ , então a energia  $E_n$  deve ser muito menor que  $E_0$ . Desta forma, pode-se justificar a expansão em série de Taylor da Eq. (2.6). Inicialmente considerando

$$\ln P_n = \ln c + \ln \Omega_R(E_0 - E_n) \quad (2.7)$$

e utilizando a série de Taylor

$$f(x) = \sum_{k=0}^{\infty} \frac{\partial^k f}{\partial x^k} \Big|_{x=a} \frac{(x-a)^k}{k!} \quad (2.8)$$

na Eq. (2.7), tem-se

$$\ln P_n = \sum_{k=0}^{\infty} \frac{\partial^k}{\partial E^k} \left( \frac{\ln c + \ln \Omega_R(E)}{k!} \right) \Big|_{E=E_0} (E - E_0)^k, \quad (2.9)$$

com  $E = E_0 - E_n$ . Logo

$$\ln P_n = \ln c + \ln \Omega_R(E_0) + \left[ \frac{\partial \ln \Omega_R(E)}{\partial E} \right]_{E=E_0} (-E_n) + \frac{1}{2} \left[ \frac{\partial^2 \ln \Omega_R(E)}{\partial E^2} \right]_{E=E_0} (-E_n)^2 + \dots \quad (2.10)$$

Utilizando a definição de entropia,  $S = k_B \ln \Omega$ , e sabendo que

$$\frac{\partial S}{\partial E} = \frac{1}{T} \quad (2.11)$$

temos

$$\left. \frac{\partial \ln \Omega_R(E)}{\partial E} \right|_{E=E_0} = \frac{1}{k_B T} \quad (2.12)$$

onde  $T$  é a temperatura do reservatório. Também se deve ter,

$$\left. \frac{\partial^2 \ln \Omega_R(E)}{\partial E^2} \right|_{E=E_0} = \frac{\partial}{\partial E} \left( \frac{1}{k_B T} \right) \Big|_{E=E_0} \mapsto 0, \quad (2.13)$$

para um verdadeiro reservatório térmico em que a temperatura está fixa. Portanto a Eq. (2.10) resulta em,

$$\ln P_n = \text{constante} - \beta E_n, \quad (2.14)$$

onde  $\beta = \frac{1}{k_B T}$ , que leva a

$$P_n = \frac{\exp(-\beta E_n)}{\sum_n \exp(-\beta E_n)}, \quad (2.15)$$

sendo esta equação chamada de distribuição de Boltzmann [17]. O denominador da Eq. (2.15),

$$Z = \sum_n \exp(-\beta E_n), \quad (2.16)$$

é chamado de função de partição [17], que é a soma sobre todos os estados microscópicos e está associada à normalização da probabilidade  $P_n$ .

### 2.3.6 Energia interna e calor específico

A energia interna de um sistema pode ser obtida pela média de ensemble das energias dos estados microscópicos [17],

$$U \equiv \langle E \rangle = \frac{\sum_n E_n \exp(-\beta E_n)}{\sum_n \exp(-\beta E_n)}. \quad (2.17)$$

A partir desta definição se obtém a relação entre a energia interna e a derivada logarítmica da função partição:

$$\langle E \rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\frac{\partial \ln Z}{\partial \beta}, \quad (2.18)$$

que também pode ser escrita como:

$$\langle E \rangle = k_B T^2 \frac{\partial \ln Z}{\partial T}, \quad (2.19)$$

onde utilizamos  $\frac{\partial T}{\partial \beta} = -k_B T^2$ . De forma análoga, podemos escrever o valor médio  $\langle E^2 \rangle$  como

$$\langle E^2 \rangle = \frac{1}{Z} \sum_n E_n^2 \exp(-\beta E_n) = \frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2}. \quad (2.20)$$

Utilizando as equações (2.18) e (2.20), a variância da energia,  $\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2$ , é dada por

$$\sigma_E^2 = \frac{1}{Z} \frac{\partial^2 Z}{\partial \beta^2} - \frac{1}{Z^2} \left( \frac{\partial Z}{\partial \beta} \right)^2 = \frac{\partial}{\partial \beta} \left( \frac{1}{Z} \frac{\partial Z}{\partial \beta} \right) = -\frac{\partial \langle E \rangle}{\partial \beta}. \quad (2.21)$$

Por outro lado, o calor específico a volume constante,  $C_V$ , é definido por

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = -\frac{1}{k_B T^2} \left( \frac{\partial \langle E \rangle}{\partial \beta} \right)_V. \quad (2.22)$$

Combinando (2.21) e (2.22) chega-se finalmente a

$$C_V = \frac{1}{k_B T^2} \sigma_E^2, \quad (2.23)$$

que é a forma utilizada neste trabalho para o cálculo do calor específico.

### 2.3.7 Média de ensemble por amostragem

O método de Monte Carlo (que será discutido posteriormente) refere-se ao procedimento computacional de buscar a solução de um problema por amostragem que envolve o sorteio de uma grande quantidade de números aleatórios e o cálculo de médias apropriadas. Vimos na subseção 2.3.4 como obter a média no ensemble das grandezas termodinâmicas. Computacionalmente, o ensemble é representado por um número  $M$  de amostras que, no presente trabalho, correspondem às diferentes conformações que a cadeia pode assumir. Cada uma dessas conformações obtidas de forma aleatória no decorrer da simulação, representa um microestado do sistema. Usando a Eq. (2.5), em vez de somar sobre todos os  $N$  componentes do ensemble (que é impraticável para  $N \rightarrow \infty$ ), soma-se sobre os  $M$  elementos da amostra [16]:

$$\langle A \rangle = \frac{1}{M} \sum_{n=1}^M A(n). \quad (2.24)$$

#### 2.3.7.1 Amostragem simples e amostragem por importância

Consideremos um caminhante aleatório (ou um bêbado<sup>4</sup>) [17, 19] que se desloca em uma linha reta a partir da origem, dando passos de comprimento  $l$ . Ele se desloca dando passos com probabilidade  $p$  para esquerda ou se desloca com probabilidade  $q = 1 - p$  para a direita. No caso simétrico, temos  $p = q = 0,5$ . Há  $2^n$  possíveis caminhos aleatórios, onde  $n$  é o número total de passos. Assim, o problema consiste em obter a probabilidade de se encontrar o caminhante aleatório a uma distância  $m = 2k - n$  da origem, sendo  $k$  o número de passos para a direita. Considerando  $l = 1$ , a probabilidade é dada pela

---

<sup>4</sup>Muitos textos que falam do caminho aleatório utilizam um bêbado como exemplo de caminhante aleatório.

distribuição binomial [17, 20],

$$P_{n,k} = \frac{1}{2^n} \binom{n}{k}, \text{ para } k = 0, 1, \dots, n. \quad (2.25)$$

Também pode-se verificar que  $P_{n,0} = P_{n,n} = 2^{-n}$ , ou seja, a probabilidade de obter um caminhante aleatório nas extremidades da distribuição binomial decai exponencialmente, ficando inviável obter as configurações próximas destas regiões [19]. Então para se extrair informações das extremidades da distribuição binomial são necessárias muitas amostras, implicando a necessidade de mais recurso computacional. Portanto, utilizar a amostragem tendenciosa (ou amostragem por importância) é uma das formas de conseguir informações sobre estas regiões. Nesta amostragem, pode-se introduzir um “vício”, aumentando a probabilidade de ir para a direita ou para esquerda, conforme se pode verificar na Fig. 4. O efeito disto é que os passos não são gerados com a mesma probabilidade. Considerando  $p$  a probabilidade de saltar para a esquerda e  $(1 - p)$  a probabilidade de saltar para a direita, a probabilidade de encontrar um caminhante na posição  $2k - n$  é dada por,

$$P_{n,k} = \binom{n}{k} p^{n-k} (1-p)^k, \text{ para } k = 0, 1, \dots, n. \quad (2.26)$$

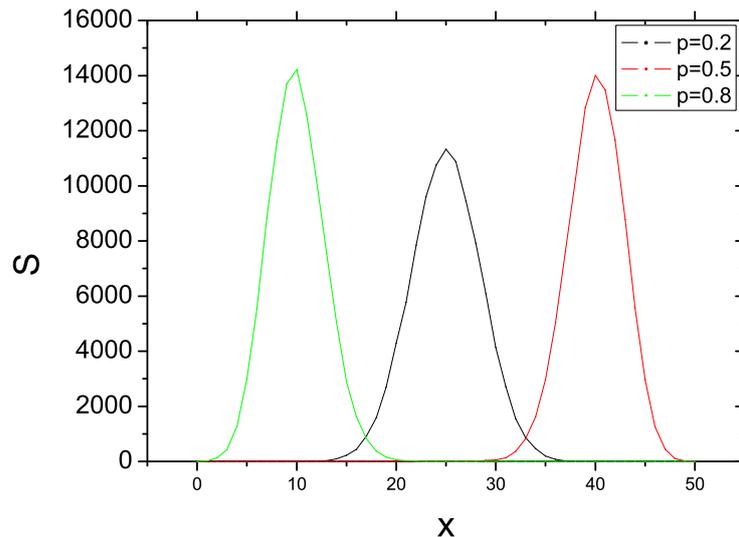


Figura 4: Amostragem tendenciosa do caminho aleatório para  $n = 50$  e  $10^5$  amostras e com probabilidade de  $p = 0, 2, 0, 5$  e  $0, 8$ , de se deslocar para esquerda.  $S(x)$  corresponde ao número de amostras com  $x$  passos para esquerda.

Para corrigir esse viés, o peso estatístico precisa ser corrigido, o que natural-

mente leva à amostragem por importância. No algoritmo, isto pode ser convenientemente realizado alterando o peso estatístico do caminho, isto é, multiplicando por um fator de  $1/2p$  ou  $1/2(1-p)$  para passos para a esquerda ou para a direita.

A Fig. 4 mostra o número de caminhadas  $S$  em função de  $x$ , que por sua vez é o número de passos para esquerda. As três curvas foram criadas utilizando diferentes probabilidades de ir para esquerda ou direita, sendo a curva do meio a única gerada com a mesma probabilidade ( $p = q = 0,5$ ). Observando os valores máximos nas três curvas do gráfico, nota-se que, para a primeira (à esquerda) e última (à direita) curva, foram geradas aproximadamente 14000 caminhadas com  $x = 10$  e  $x = 40$ . Para a curva do meio foram geradas aproximadamente 11000 cadeias com  $x = 25$ . Assim adicionando os “vícios”, ou seja, alterando a probabilidade de ir para esquerda (e conseqüentemente para a direita) é possível obter mais informações para as regiões mais difíceis de ser acessadas.

## 2.4 Transições de fase

### 2.4.1 Fases termodinâmicas

Um sistema termodinâmico pode existir em diferentes fases que, macroscopicamente, podem ter um comportamento drasticamente diferente.

Ao variarmos os parâmetros intensivos de um sistema (por exemplo, temperatura e pressão), suas propriedades como a densidade, compressibilidade, calor específico *etc.* se alteram, em geral de forma contínua [20]. Porém, em determinadas condições, pode haver variações súbitas (descontinuidades ou divergências) de algumas destas propriedades. Pode-se verificar estas variações súbitas considerando inicialmente um sistema descrito por sua energia  $U = U(S, V, N)$  [17, 21], sendo  $S$ ,  $V$  e  $N$  respectivamente a entropia, o volume e o número de partículas. Assim, pode-se escrever a forma diferencial da equação fundamental na representação da energia como

$$dU = \left(\frac{\partial U}{\partial S}\right)_{V,N} dS + \left(\frac{\partial U}{\partial V}\right)_{S,N} dV + \left(\frac{\partial U}{\partial N}\right)_{S,V} dN, \quad (2.27)$$

que descreve um processo termodinâmico quase-estático [17]. Por outro lado, a conservação de energia [17, 21], pode ser escrita como

$$\begin{aligned} dU &= dQ + dW_{mec} + dW_{qui} \\ &= TdS - PdV + \mu dN, \end{aligned} \quad (2.28)$$

onde  $Q$  é o calor adicionado,  $W_{mec}$  o trabalho mecânico,  $W_{qui}$  o trabalho químico,  $p$  a pressão,  $\mu$  o potencial químico e  $T$  a temperatura. Comparando as Eqs. (2.27) e (2.28)

obtêm-se as definições dos parâmetros intensivos:

$$\begin{aligned} T &= \left( \frac{\partial U}{\partial S} \right)_{V,N}, \\ P &= - \left( \frac{\partial U}{\partial V} \right)_{S,N}, \\ \mu &= \left( \frac{\partial U}{\partial N} \right)_{S,V}. \end{aligned} \quad (2.29)$$

Essas funções fornecem as equações de estado na representação de energia. Considerando que a função  $U(S, V, N)$  é uma função homogênea de primeiro grau, isto é,

$$U(\lambda S, \lambda V, \lambda N) = \lambda U(S, V, N), \quad (2.30)$$

podemos derivar os dois lados da Eq. (2.30) em relação a  $\lambda$ , obtendo

$$\frac{\partial U(\lambda S, \lambda V, \lambda N)}{\partial(\lambda S)} S + \frac{\partial U(\lambda S, \lambda V, \lambda N)}{\partial(\lambda V)} V + \frac{\partial U(\lambda S, \lambda V, \lambda N)}{\partial(\lambda N)} N = U(S, V, N). \quad (2.31)$$

Considerando  $\lambda = 1$ , chegamos a relação de Euler da termodinâmica,

$$TS - PV + \mu N = U. \quad (2.32)$$

Com os campos termodinâmicos dados pelas Eqs. (2.29) e com a relação de Euler mostrada na Eq. (2.32), constrói-se a energia livre de Helmholtz [22] através da transformada de Legendre da energia  $U = U(S, V, N)$  [17],

$$F(T, V, N) = U - TS, \quad (2.33)$$

Diferenciando a Eq. (2.32)

$$dU = TdS + SdT - PdV - VdP + \mu dN + Nd\mu, \quad (2.34)$$

e comparando com a Eq. (2.28), chega-se a

$$0 = SdT - VdP + Nd\mu, \quad (2.35)$$

que é conhecida como relação de Gibbs-Duhem [17]. Para a energia livre de Helmholtz (2.33), temos

$$dF = dU - TdS - SdT = -SdT - PdV + \mu dN. \quad (2.36)$$

Assim,

$$\left( \frac{\partial F}{\partial T} \right)_{V,N} = -S(T, V, N).$$

$$\begin{aligned}\left(\frac{\partial F}{\partial V}\right)_{T,N} &= -P(T, V, N), \\ \left(\frac{\partial F}{\partial N}\right)_{T,V} &= \mu(T, V, N).\end{aligned}\tag{2.37}$$

Transições de fase podem ser classificadas como contínuas ou descontínuas, conforme as primeiras derivadas da energia livre sejam contínuas ou descontínuas. As transições de primeira ordem são caracterizadas pela coexistência de duas ou mais fases na temperatura de transição  $T_0$  [1]. Considerando duas fases (I e II), a diferença entre as respectivas entropias perto do ponto de transição multiplicada pela temperatura de transição define o calor latente [1, 18]

$$\Delta Q_{latente} = \lim_{t \rightarrow 0} T_0 [S_{II}(T_0 + t) - S_I(T_0 - t)].\tag{2.38}$$

Há transição de primeira ordem quando  $\Delta Q_{latente} \neq 0$ .

Para que ocorra uma transição de segunda ordem, deve-se ter  $\Delta Q_{latente} = 0$  e, nesse caso, a entropia é contínua na temperatura de transição. A segunda derivada de  $F$  em relação a  $T$  é escrita como

$$\left(\frac{\partial^2 F}{\partial T^2}\right)_{N,V} = -\left(\frac{\partial S}{\partial T}\right)_{N,V} = -\frac{1}{T}C_V(T),\tag{2.39}$$

onde  $C_V$  é a capacidade térmica [1, 23].

Nas proximidades dos chamados pontos críticos, diversas grandezas como calor específico, compressibilidade ou suscetibilidade magnética possuem um comportamento peculiar, com divergências assintóticas que são caracterizadas por meio de expoentes críticos [17, 22].

Uma das formas mais simples de mostrar o comportamento peculiar das grandezas termodinâmicas nas vizinhanças do ponto crítico, é considerar um sistema simples. Um sistema simples é definido como um sistema macroscopicamente homogêneo, isotrópico, descarregado, quimicamente inerte e suficientemente grande. Como exemplo, analisaremos abaixo o comportamento de um *fluido puro*, que é um sistema simples composto por um único componente [17].

Considere inicialmente o diagrama de fases de um fluido puro em termos dos campos termodinâmicos pressão e temperatura, que é representado na Fig. 5. As linhas cheias da Fig. 5 representam a coexistência de fases, com densidades diferentes, mas com os mesmos valores de campos termodinâmicos. Temos, neste processo, uma transição de primeira ordem. Os símbolos S, L e G no diagrama de fases representam respectivamente os estados da matéria: sólido, líquido e gasoso. Os símbolos S-G, S-L e L-G, representam

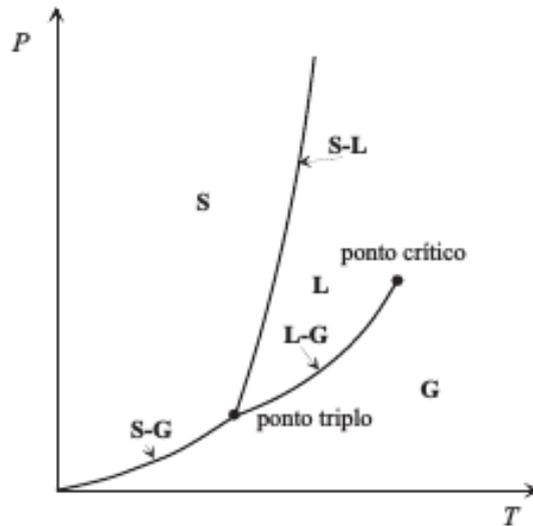


Figura 5: Diagrama de fases de um fluido simples no espaço  $P - T$ .

as coexistências de fase nos respectivos segmentos representados na Fig. 5. No ponto triplo  $(T_t, P_t)$  [17, 21], há uma coexistência de três fases distintas. Por exemplo, o ponto triplo da água é dado por  $T_t = 273,16K$  e  $P_t = 4,58mmHg$  [17] e neste ponto a água coexiste nos três estados: sólido, líquido e gasoso. O ponto crítico  $(T_c, p_c)$  representa o final da curva de coexistência de fase. A temperatura da água no ponto crítico é de  $650K$ , por outro lado, no monóxido de carbono a temperatura crítica é inferior a  $150K$ . Percorrendo a curva de coexistência entre as fases líquida e gasosa ( $L - G$ ) no diagrama de fase  $P - T$  em direção ao ponto crítico, a diferença entre as densidades do líquido e do gás diminuem. Como a transição é crítica, ela é portanto, contínua ou de segunda ordem [1, 17]. Verifica-se que para  $T < T_c$ , temos transição de primeira ordem e em  $T = T_c$  temos transição de segunda ordem.

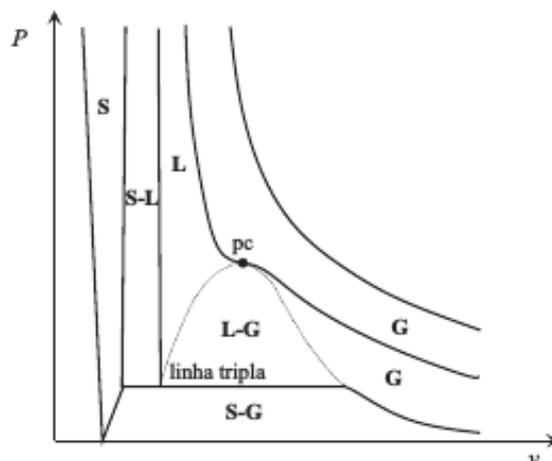


Figura 6: Isotermas de um fluido simples no diagrama  $P - v$ .

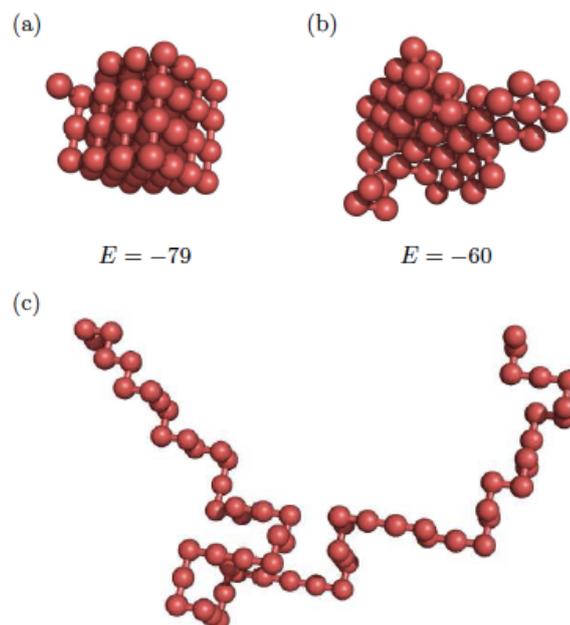
Na Fig. 6, discutimos agora o diagrama no plano  $P - v$ . Como na figura

anterior, as regiões S, L e G representam respectivamente as fases sólida, líquida e gasosa. O ponto crítico é representado por  $pc$  e a linha tripla indica a coexistência das três fases S, L e G. Em relação ao patamar na curva L-G, considere o volume específico na fase líquida  $v_l$  e na fase gasosa  $v_g$ . À medida que a temperatura aumenta a diferença  $\psi = v_g - v_l$  diminui, anulando-se no ponto crítico. Isso ocorre devido ao calor latente, que se aproxima de zero quanto mais próximo do ponto crítico [17].

### 2.4.2 Transições configuracionais em polímeros

Uma transição de fase termodinâmica ocorre apenas no limite termodinâmico, isto é, em sistemas de tamanho infinitamente grande. A transição de fase em que estamos interessados é configuracional, que ocorre quando a cadeia proteica muda sua conformação espacial. Vale ressaltar que proteínas típicas são constituídas de apenas algumas centenas de monômeros (o tamanho médio de uma proteína humana é 480 aminoácidos), portanto o caráter intrinsecamente finito das cadeias deve ser levado em conta. Por esse motivo, essas transições em polímeros não são transições termodinâmicas em seu sentido estrito. Apesar disso, alguns aspectos dessas transições conformacionais se assemelham às transições de fase usuais. Por exemplo, mostraremos no capítulo 4 que, ao analisarmos o comportamento do calor específico com a variação da temperatura, percebemos picos que são indicadores das transições configuracionais.

Figura 7: Representação de um homopolímero com 64 monômeros. (a) fase cristalina; (b) globular; (c) estendida. Figura retirada da Ref. [24].



Fonte: VOGEL, T.; BACHMANN, M.; JANKE, W. *Physical Review E*. v.76, n.061803, 2007.

A Fig. 7 mostra possíveis conformações para um homopolímero (polímero com um único tipo de monômero) na rede cúbica. A altas temperaturas, temos uma forma dissolvida, em que os monômeros estão afastados uns dos outros (Fig. 7c). Essa forma é conhecida por *random coil* e, nesse texto, a chamaremos de *estendida*. Em analogia a um fluido puro, essa fase seria a gasosa. Com a diminuição da temperatura, a macromolécula passa para uma forma *globular*, representada na Fig. 7b. Nessa fase, as conformações são condensadas, mas ainda não apresentam uma estrutura regular, assemelhando-se ao estado líquido de um fluido puro. A transição da fase estendida para a globular é conhecida como transição  $\Theta$  (ou colapso *coil-globule*) e ocorre a uma temperatura bem definida,  $T_{\Theta}$ . Essa transição ocorre de maneira contínua, sem coexistência das fases envolvidas, sendo portanto análoga a uma transição usual de segunda ordem. Outras transições contínuas também são observadas em polímeros sintéticos, proteínas e ácidos nucleicos [25]. Diminuindo ainda mais a temperatura, chega-se à fase *crystalina*, caracterizada por um ordenamento mais regular, como pode ser observado na Fig. 7a. Essa cristalização de proteínas, à temperatura  $T_m$ , ocorre descontinuamente, apresentando coexistência das fases cristalina e globular, assemelhando-se a uma transição de primeira ordem convencional.

As flutuações da energia, expressas pelo calor específico, podem ser sinal de atividade termodinâmica. Assim, picos do calor específico como função da temperatura são indicadores de transição entre estados macroscópicos fisicamente diferentes. Neste trabalho, o calor específico foi utilizado para obtermos a localização dessas transições configuracionais na cadeia considerada.

## 3 METODOLOGIA

### 3.1 Simulações de Monte Carlo

O Monte Carlo [16, 26] é um método de simulação computacional que utiliza números aleatórios. O objetivo é gerar um conjunto representativo de configurações para acessar quantidades termodinâmicas sem a necessidade de resolver o sistema analiticamente ou de realizar uma enumeração exata [27].

Consideremos uma porção macroscópica de matéria constituída de um sistema mecânico de  $N$  partículas microscópicas, com  $N \approx 10^{23}$ . Se tentássemos descrever o comportamento dinâmico por meio da Mecânica Clássica, teríamos que considerar cada átomo ou molécula como um corpo sólido, assim obteríamos  $6N$  equações diferenciais acopladas a serem resolvidas. Para cada partícula, temos três coordenadas de posição e três de momento. Assim, seria uma tarefa praticamente impossível de se resolver. Precisaríamos de um computador suficientemente excepcional para resolver o sistema de equações diferenciais. A Mecânica Estatística pode resolver esta tarefa sem resolver as equações de movimento dos constituintes microscópicos, permitindo a obtenção dos valores das grandezas macroscópicas a partir de médias realizadas no ensemble considerado. Com o intuito de avaliar quantidades como energia, calor específico etc., utilizamos o método de Monte Carlo através do algoritmo PERM, descrito a seguir.

### 3.2 PERM

Algoritmos refinados têm sido objeto de intensa pesquisa visando à melhoria da eficiência. Em 1997, Grassberger criou um algoritmo conhecido como *Pruned and Enriched Rosenbluth method*, também chamado de PERM [1, 28]. O PERM é um algoritmo de Monte Carlo bastante utilizado para descrever a transição  $\Theta$  na qual, sob variação de temperatura, a conformação da molécula passa de uma fase estendida (a alta temperatura) para uma fase globular (a baixa temperatura), sendo o ponto  $T_\Theta$  a temperatura crítica de transição.

Os algoritmos baseados no crescimento de cadeia (conhecidos como *chain-growth*) são uma promissora alternativa para simular as redes de polímeros e proteínas. O PERM veio como uma variação do método de crescimento de cadeia de Rosenbluth [29], melhorando a eficiência dos algoritmos anteriores e aumentando a amostragem de Rosenbluth com estratégias de poda e de enriquecimento. Aqui os termos “poda” e “enriquecimento” se referem respectivamente à exclusão e a multiplicação das cadeias

criadas [1, 26].

Para compreender melhor a lógica do PERM escolheu-se discutir inicialmente a caminhada aleatória auto-excludente com interações (também conhecida como *Self Avoiding Walk* - SAW) [5, 30]. SAW consiste em um caminho ou trajetória, obtida de forma aleatória sem que haja qualquer interceptação com o caminho já percorrido, ou seja, é um algoritmo de caminho aleatório (chamado de *random walk*), em que não é permitido passar por um ponto que foi visitado anteriormente.

No passo inicial do SAW na rede quadrada, existem 4 primeiros vizinhos (são os possíveis lugares para se dar o passo conforme se verifica na Fig. 8), e após o primeiro passo existe uma possibilidade a menos para se mover, pois voltar para onde estava significaria contrariar a restrição da auto-exclusão e, neste caso, a caminhada é interrompida. Fig. 8 é uma ilustração do SAW em 2D. Os círculos em vermelho são os monômeros formados em cada passo aleatório e as circunferências de centro transparente são as possíveis posições do próximo passo do crescimento da molécula.

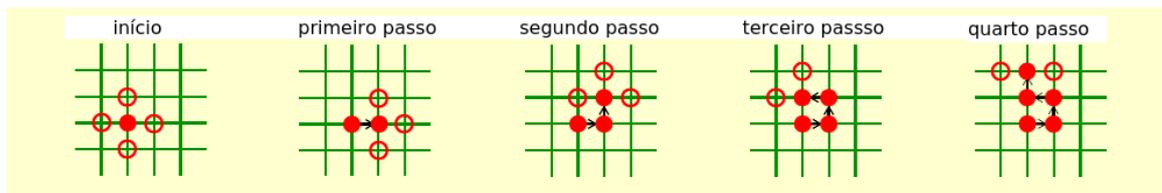


Figura 8: Exemplo do *Self Avoiding Walk* em 2D.

A performance do algoritmo para simular o SAW pode ser melhorada com a implementação do algoritmo criado por Rosenbluth em 1956 [29]. A ideia básica para o Rosenbluth é evitar as auto-interseções. Desta forma, o algoritmo apenas encerra a caminhada quando estiver preso em um beco sem saída e não puder continuar a crescer. A amostragem de Rosenbluth pode produzir configurações substancialmente maiores do que a amostragem simples.

Enquanto na amostragem simples geram-se todas as configurações com igual probabilidade, as configurações geradas com o método de Rosenbluth são obtidas com diferentes probabilidades. O algoritmo de Rosenbluth utiliza o conceito de atmosfera [1], que é entendida como o número de maneiras em que uma configuração pode continuar crescendo. Para o início da caminhada aleatória unidimensional a atmosfera é sempre dois (ou seja,  $a = 2$ ). Em duas dimensões, na rede quadrada, a atmosfera é quatro ( $a = 4$  como se verifica na Fig. 8) e em três dimensões (na rede cúbica) a atmosfera é seis ( $a = 6$ ).

No exemplo mostrado na Fig. 8 para uma rede quadrada, após o primeiro e segundo passos, a atmosfera é  $a = 3$ . Continuando a caminhada, no terceiro passo a atmosfera é  $a = 2$ , e por último, no quarto passo, a atmosfera é  $a = 3$  (para este

último, a figura mostra apenas duas possíveis trajetórias do crescimento da molécula, mas são três possibilidades, pois na Fig. 8 não aparece a possibilidade superior). No entanto, dependendo da forma com que a molécula cresce, pode ocorrer que a atmosfera seja igual a zero, portanto, a configuração está presa e não pode ser continuada. Neste trabalho, denotamos a atmosfera por  $a$  como foi mencionado anteriormente. Se existe uma configuração de atmosfera  $a$ , isto significa que existem  $a$  diferentes caminhos para dar continuidade ao crescimento, e cada uma delas pode ser selecionada com probabilidade  $p = 1/a$ . Uma caminhada de  $n$  passos, crescida pelo método de Rosenbluth, tem probabilidade e peso dados respectivamente por:

$$P_n = \left( \prod_{l=2}^n a_l \right)^{-1} \quad (3.1)$$

$$W_n^R = p_n^{-1}. \quad (3.2)$$

Assim, a cadeia é continuada desde que  $a \geq 1$ , ou seja, tenha pelo menos um primeiro vizinho livre.

A Fig. 9 mostra uma comparação, em termos de eficiência, entre o crescimento de Rosenbluth e a amostragem simples, sendo  $S_N$  o número de tentativas e  $N$  o número de monômeros da cadeia. O gráfico demonstra que a amostragem simples é menos eficiente que a de Rosenbluth, pois a cadeia é construída levando em conta que os primeiros vizinhos são escolhidos com a mesma probabilidade. Assim, quando os primeiros vizinhos estão ocupados, a cadeia é interrompida durante sua construção, tornando a amostragem simples pouco eficiente. No método de Rosenbluth, a escolha do próximo vizinho a ser ocupado é feita apenas entre os sítios vazios. Isso significa que a cadeia é descartada menos vezes que na amostragem simples. A Fig. 9 mostra o número de vezes que a cadeia atingiu um determinado comprimento  $N$ .

Para compensar o viés do *Pruned and Enriched Rosenbluth method*, é usado um fator peso dado por:

$$W_n^{PERM} = \prod_{l=2}^n a_l e^{-(E_l - E_{l-1})/k_B T} \quad (3.3)$$

onde  $T$  é a temperatura,  $E_l$  é a energia parcial da cadeia com comprimento  $l$  e  $k_B$  é a constante de Boltzmann. O PERM utiliza o peso para o controle populacional das cadeias geradas. O algoritmo favorece as cadeias que são consideradas “boas” e desfavorece as cadeias que são consideradas “ruins” [1, 26]. Se uma cadeia alcança um comprimento  $n$ , o peso  $W_n^{PERM}$  é calculado e comparado adequadamente com outros dois parâmetros empiricamente parametrizados chamados de parâmetro superior  $W_n^>$  e inferior  $W_n^<$ . Para

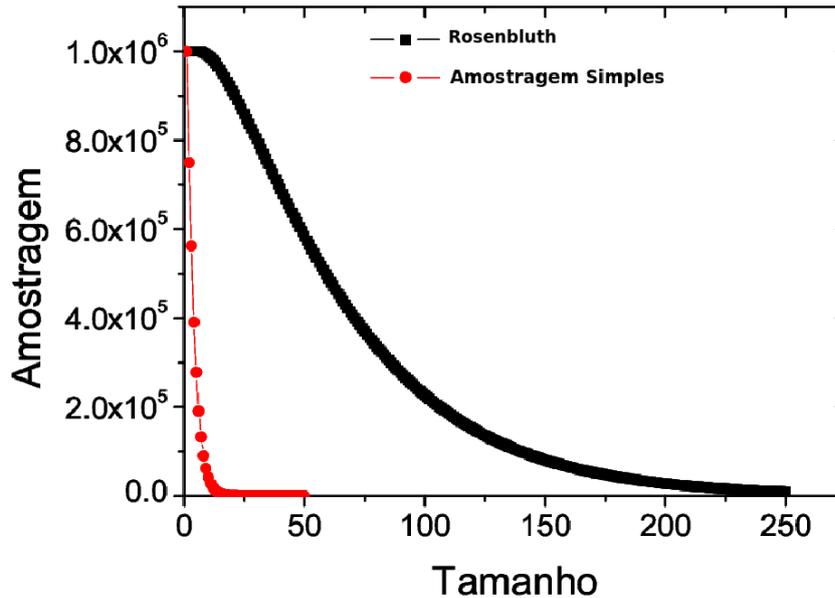


Figura 9: Demonstração da eficiência do método de Rosenbluth em relação à amostragem simples do SAW.

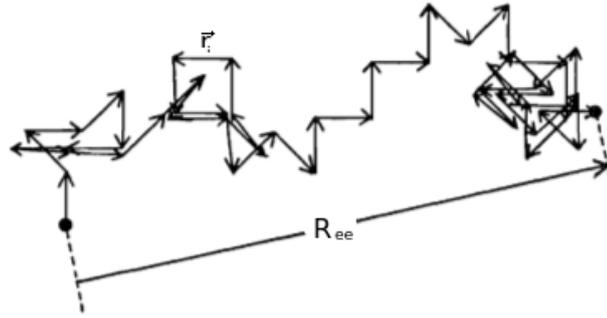
$W_n^{PERM} > W_n^>$ , idênticas cópias são criadas, e chama-se este processo de enriquecimento ou *enrichment*. A partir dessa replicação, as cópias crescem de forma independente, sendo o peso  $W_n^{PERM}$  dividido igualmente entre as cópias. Se  $W_n^{PERM} < W_n^<$ , então a cadeia é descartada (*pruned*) com uma certa probabilidade ( $1/2$ , por exemplo). Na etapa de enriquecimento, o peso é dividido igualmente para as cópias criadas e o inverso ocorre para o processo da poda. Para valores entre os limiares  $W_n^>$  e  $W_n^<$ , a cadeia continua seu crescimento. Quando os limites superior e inferior são respectivamente  $W_n^> = \infty$  e  $W_n^< = 0$ , o PERM (desconsiderando a temperatura) torna-se idêntico ao algoritmo de Rosenbluth, não ocorrendo *pruning* nem *enrichment*. A escolha dos limites superior e inferior é importante para a eficiência do algoritmo, mas os resultados não são alterados por essa escolha.

### 3.3 Grandezas analisadas

#### 3.3.1 Distância Ponta a Ponta

Consideremos uma cadeia hipotética polimérica de sequência linear com ligações covalentes de comprimento  $l$  entre os monômeros, sem qualquer restrição com relação ao ângulo descrito pelos monômeros consecutivos. Conforme mostrado na Fig. 10, o início e final de cada vetor  $\vec{r}_i$  representam monômeros sucessivos e  $R_{ee}$  representa a distância ponta a ponta [5].

Figura 10: Representação vetorial de uma cadeia polimérica linear.



Fonte: FLORY, P.J. *Principles of Polymer Chemistry*, 1971.

A distância ponta a ponta  $R_{ee}$  em três dimensões é obtida através da relação,

$$(R_{ee})_i^2 = (x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2, \quad (3.4)$$

sendo  $(x_o, y_o, z_o)$  a origem do sistema (início da molécula) e  $(x_i, y_i, z_i)$  a posição final da molécula. Considerando  $(x_o, y_o, z_o) = (0, 0, 0)$ , obtém-se,

$$(R_{ee})_i^2 = (x_i)^2 + (y_i)^2 + (z_i)^2. \quad (3.5)$$

Será utilizado  $\langle \rangle$  para designar a média de uma variável aleatória portanto,

$$\langle (R_{ee})_i^2 \rangle = \frac{1}{M} \sum_{i=1}^M (R_{ee})_i^2, \quad (3.6)$$

onde  $M$  é o número total de amostras (cadeias geradas).

### 3.3.2 Raio de giração

O raio de giração refere-se à distribuição dos componentes de um objeto à volta de um eixo. Considerando a mesma molécula hipotética da seção anterior, então o valor médio do raio de giração pode ser escrito como

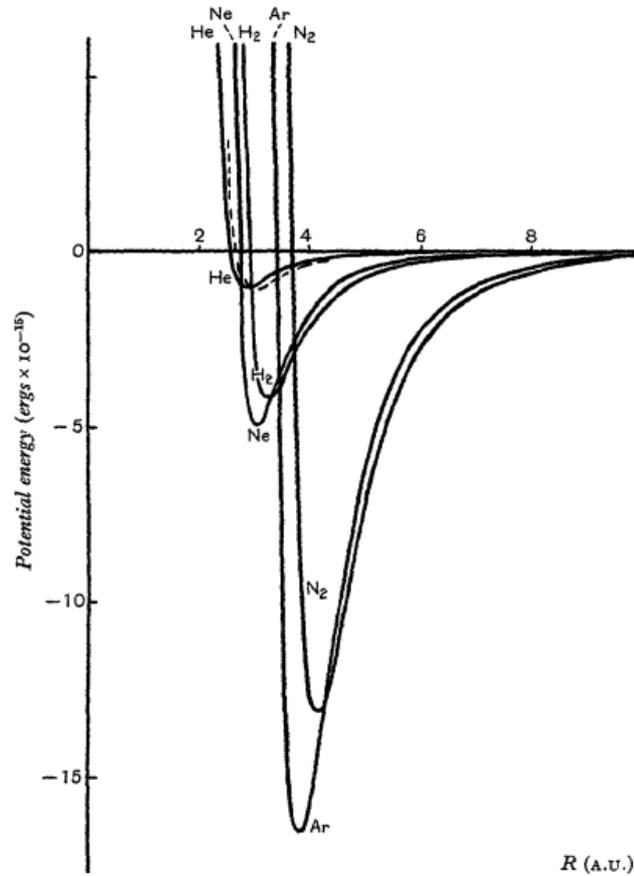
$$\langle R_g \rangle = \frac{1}{M} \sum_{j=1}^M \left[ \frac{1}{N} \sum_{i=1}^N (|\vec{r}_i - \vec{r}_{cm}|)^2 \right]_j^{1/2}. \quad (3.7)$$

sendo  $N$  o número de monômeros,  $\vec{r}_i$  a posição do monômero  $i$  relativa à origem das coordenadas e  $\vec{r}_{cm}$  a posição do centro de massa.

### 3.3.3 Interações HH e HP

Dois átomos ou moléculas são sujeitos a uma força que pode ser atrativa ou repulsiva, dependendo da maior ou menor separação entre eles, conforme ilustrado na Fig. 11. Para distâncias maiores (o que é representado na Figura 11 pela parte à direita

Figura 11: Energia potencial de diferentes gases em função da distância em ângström (Å).



Fonte: LENNARD-JONES J. *Proceedings of the Physical Society*. v.43, p.461-482, 1931.

dos pontos de mínimo) a força é atrativa (van der Waals). Para distâncias menores (à esquerda dos pontos de mínimo), a força repulsiva resulta da superposição devida ao princípio de exclusão de Pauli [1].

As interações entre monômeros ligados covalentemente ou por potencial não são conexões rígidas. Sistemas de muitas partículas governadas pela força de van der Waals são tipicamente descritos pelo potencial de Lennard-Jones [31],

$$U_{LJ}(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right], \quad (3.8)$$

sendo  $\sigma$  a distância na qual o potencial interpartícula é zero,  $\varepsilon$  é a profundidade do potencial, o termo  $(\frac{\sigma}{r})^{12}$  descreve a repulsão apresentada pelo lado esquerdo do ponto mínimo da Fig. 11 e o termo  $(\frac{\sigma}{r})^6$  descreve a atração mostrada pelo lado direito do ponto mínimo da Fig. 11.

### 3.4 Detalhes das simulações

Apesar da variedade de aminoácidos presentes na natureza, modelos simplificados têm sido propostos a fim de compreender melhor as transições conformacionais de uma cadeia de polímero. Desta forma, neste trabalho usou-se o modelo hidrofóbico-polar (HP), que é um dos modelos de rede mais simples para descrever, qualitativamente, o comportamento do enovelamento de proteínas.

Inicialmente, para a construção da molécula de proteína, utilizou-se uma cadeia de 42 monômeros e uma sequência dada por:  $PH_2PHPH_2PHPH_2H_3PHPH_2PHPH_3-P_2HPHPH_2PHPH_2P$ , que foi projetada para simular propriedades físicas do  $\beta$ -hélice paralelo de pectato liase *C*. Uma das aplicações desta proteína é a utilização nas indústrias de sucos de frutas, pois, a adição dessas enzimas resulta na diminuição da viscosidade, aumentando o rendimento e diminuindo as dificuldades de filtrar os sucos [32]. A sequência com 42 monômeros dessa molécula foi escolhida para este trabalho por ser utilizada nos estudos realizados por Bachmann e Janke [33]. No artigo, eles estudam a densidade de estados das proteínas no modelo HP, assim foi possível comparar resultados como aqueles da referência [33].

Foi utilizado o algoritmo *Pruned and Enriched Rosenbluth method* (PERM), implementado na linguagem de programação *C*. Esta simulação foi realizada por meio do *cluster* do Instituto de Física - UFMT. O tempo de simulação foi de aproximadamente 25 dias.

Iniciou-se o trabalho em uma rede quadrada e o crescimento da molécula foi feito a partir da origem do sistema de coordenadas  $P(0,0)$ . O gerador de números aleatórios utilizado foi o Marsaglia [34].

O intervalo de temperaturas considerado foi entre  $T = 0,01$  e  $T = 2,0$ , com passos de  $\Delta T = 0,02$ . O número total de cadeias simuladas foi entre  $10^6$  e  $10^7$  para cada valor da temperatura. Observáveis físicos como calor específico, energia total, distância ponta a ponta, raio de giração e número médio de interações  $HH$  e  $HP$  foram calculados através do algoritmo. As quantidades físicas foram calculadas para diferentes valores de  $s$  (verificar a Eq. 2.1) no intervalo  $0 \leq s \leq 1$ . Erros estatísticos foram avaliados usando o desvio padrão.

Para cada valor de  $s$ , a temperatura de transição foi obtida a partir dos picos do calor específico. Com isso, pudemos construir o diagrama de fases.

## 4 RESULTADOS

### 4.1 Em duas dimensões

Considerando a cadeia HP na rede quadrada, foram simulados e analisados vários valores da razão entre as energias,  $s = \epsilon_{HP}/\epsilon_{HH}$  no intervalo  $0 \leq s \leq 1$ . A Fig. 12 mostra as curvas do calor específico em função da temperatura para 11 valores de  $s$ . Pode-se notar que, para  $s = 0,0$ , é observado um único pico na região de baixa temperatura, enquanto que para  $s \geq 0,2$ , aparecem dois distintos picos, correspondentes a duas transições conformacionais.

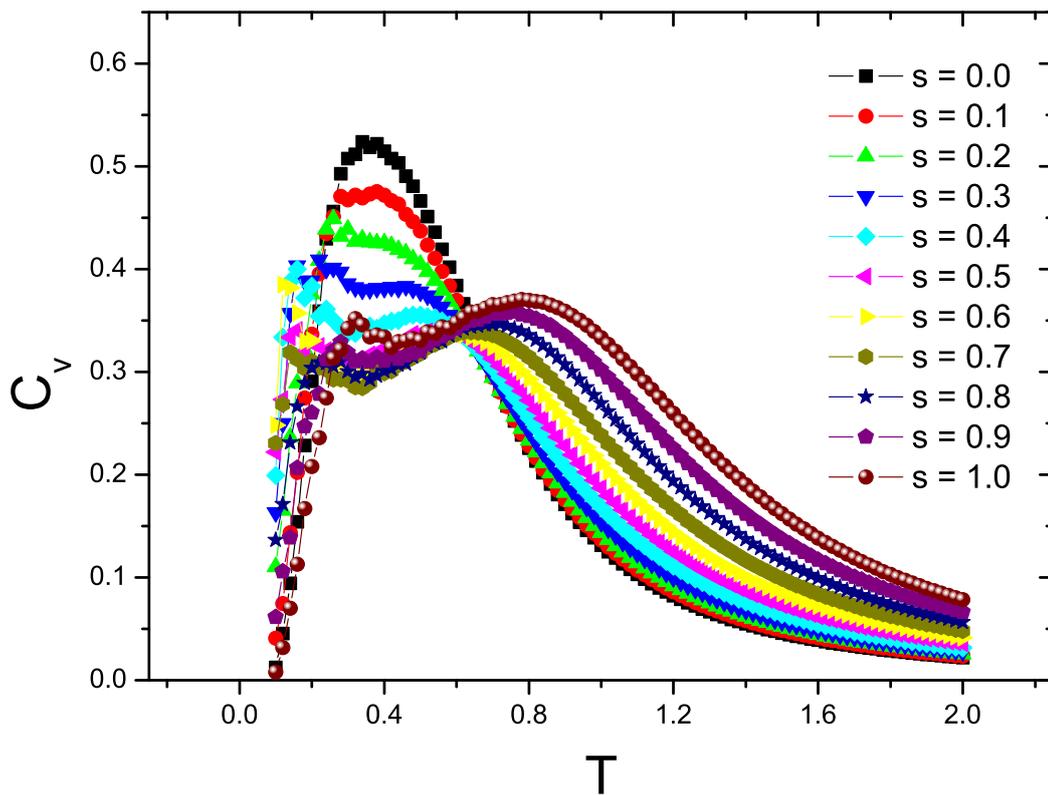


Figura 12: Calor específico em função da temperatura para  $0 \leq s \leq 1$ .

A Fig. 13 mostra o calor específico e a energia média para 4 valores de  $s$  (para facilitar a visualização). No lado esquerdo tem-se o calor específico  $C_v$  como função da temperatura  $T$ . A média da energia é traçada para os mesmos valores de  $s$ , podendo ser verificada na Fig. 13 (direita). Como esperado, quando a temperatura aumenta, a energia

se aproxima de zero, que indica a fase estendida da proteína.

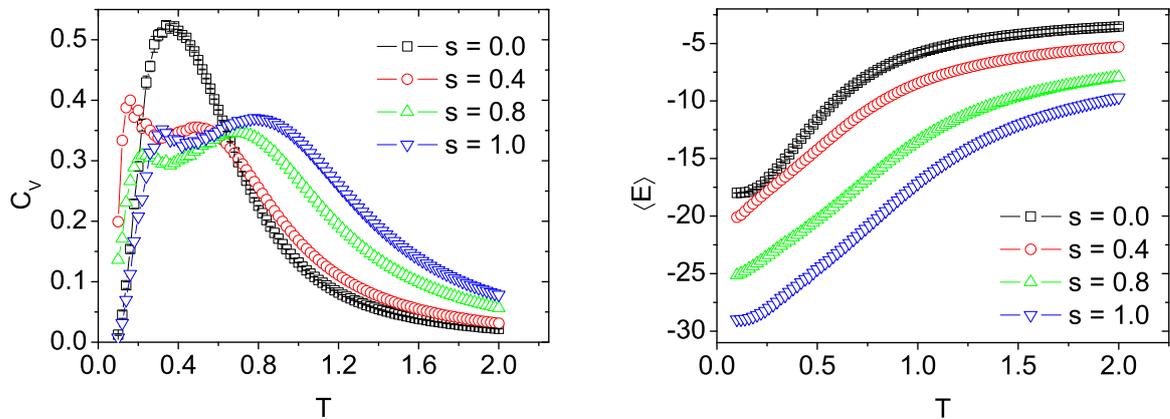


Figura 13: Calor específico (à esquerda) e energia média (à direita) como funções da temperatura para alguns valores do parâmetro  $s$ . As barras de erro, se não forem visíveis, são menores do que os símbolos.

Na Fig. 14 são apresentados os números médios de interações  $HH$  e  $HP$  (respectivamente,  $n_{HH}$  e  $n_{HP}$ ). Para pequenos valores de  $s$  (0,0 e 0,4, por exemplo), pode-se observar que as conformações de baixa temperatura possuem um número significativo de interações  $HH$ , enquanto que o número de interações  $HP$  é pequeno. Quando o parâmetro  $s$  se aproxima de 1,0 (este limite significa que as intensidades das interações  $HH$  e  $HP$  são iguais),  $n_{HP}$  fica maior enquanto  $n_{HH}$  diminui, conforme pode ser verificado na Fig. 14. Um exemplo de conformação obtida nas simulações é mostrada na Fig. 15.

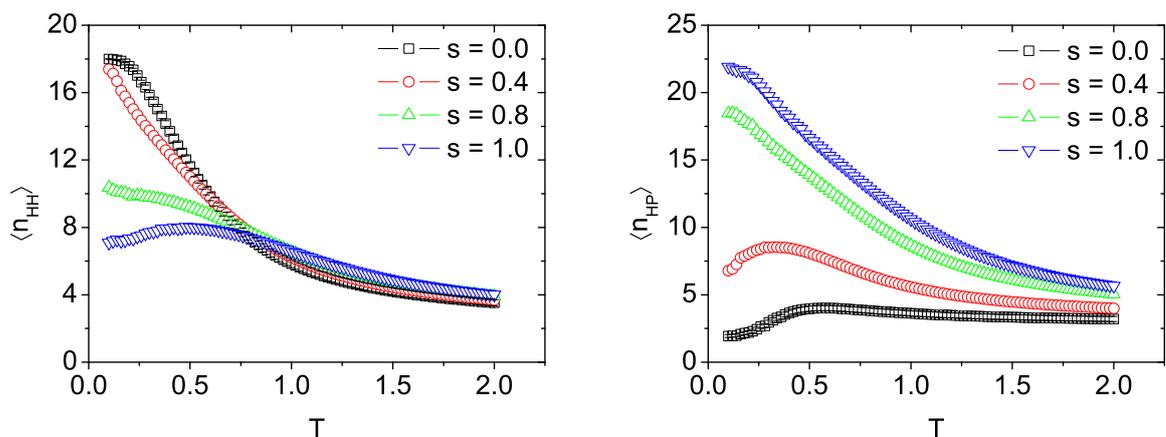


Figura 14: Esquerda: número médio de contatos do tipo hidrofóbico-hidrofóbico  $n_{HH}$ . Direita: número médio de interações hidrofóbico-polar  $n_{HP}$ . As barras de erro são da ordem do tamanho dos símbolos.

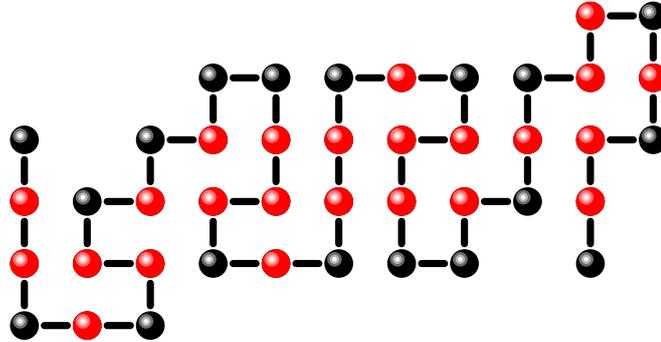


Figura 15: A proteína com 18 contatos  $HH$  e 3 contatos  $HP$ .

Para se ter uma ideia de como as conformações se parecem sob diferentes condições térmicas, analisamos quantidades estruturais, tais como a distância ponta a ponta  $R_{ee}$  e o raio de giração  $R_g$ . Os valores médios de ambas as quantidades são mostrados na Fig. 16. Um comportamento interessante pode ser notado para  $s = 0,0$ . Neste caso, a energia mais baixa que foi obtida é  $-19$ . Em contraste com o caso tridimensional [33], o estado fundamental em  $2D$  parece ser altamente degenerado. Para maximizar  $n_{HH}$  (e, conseqüentemente, minimizar a energia total), o polímero forma um núcleo hidrofóbico que também inclui alguns monômeros polares em uma forma compacta. Outras conformações do estado fundamental possuem um *backbone* hidrofóbico rodeado por monômeros polares, mas estas estruturas têm um raio de giração muito maior. Um exemplo de conformação que apresenta um núcleo hidrofóbico e uma blindagem polar é mostrado na Fig. 17. Com o aumento da temperatura,  $R_{ee}$  e  $R_g$  diminuem e chegam a um valor mínimo em  $T \approx 0,25$ . Para maiores valores de  $s$  (especialmente para  $s \geq 0,3$ ), nos quais os contatos  $HP$  se tornam mais relevantes, esse comportamento não é mais observado.

A Fig. 18 resume os resultados desta dissertação. Ela retrata o diagrama de fases no espaço de  $s-T$  e inclui algumas típicas conformações em cada região do diagrama de fases.

Os pontos de transição foram obtidos a partir das posições dos picos do calor específico. Na região de baixa temperatura, encontramos conformações compactas dominantes para grandes valores de  $s$ , ao passo que as conformações são menos compactas para pequenos valores de  $s$  (nesse caso, as interações  $HH$  são favorecidas). Para um valor fixo

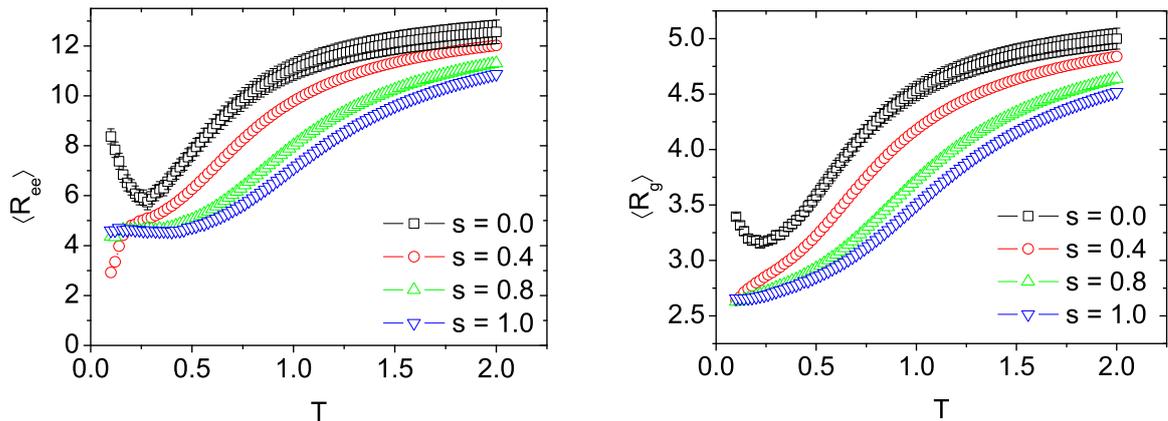


Figura 16: Valores médios da distância de ponta a ponta (esquerda) e do raio de giração (direita). Para melhor visualização, barras de erro são mostradas somente para  $s = 0,0$ . Para os outros valores de  $s$ , as barras de erro têm tamanhos semelhantes.

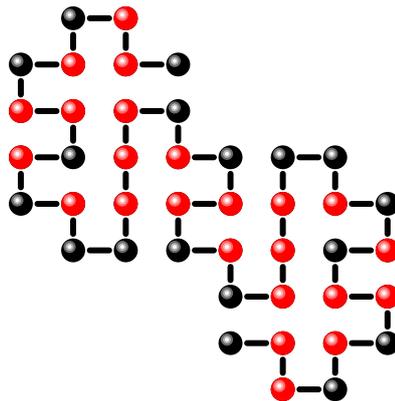


Figura 17: A proteína com um núcleo hidrofóbico e uma “blindagem” polar.

de  $s$  ( $s \lesssim 0,20$ ) e temperatura  $T \gtrsim 0,38$ , a forma estrutural estendida é caracterizada pelos grandes valores de  $R_{ee}$  e  $R_g$ . Para  $s \gtrsim 0,25$ , há uma fase globular intermediária, na qual tanto o número de contatos  $HH$  quanto de  $HP$  são relevantes.

## 4.2 Em três dimensões

Considerando o modelo em três dimensões, também foram simulados e analisados valores da razão entre as energias,  $s = \epsilon_{HP}/\epsilon_{HH}$  no intervalo  $0 \leq s \leq 1$ . O intervalo de temperaturas considerado foi entre  $T = 0,1$  e  $T = 1,0$ , com passos de  $\Delta T = 0,01$ . O número total de cadeias simuladas foi de  $10^8$  para cada valor de  $T$  e  $s$ .

O lado esquerdo da Fig. 19, mostra o calor específico em função da temperatura

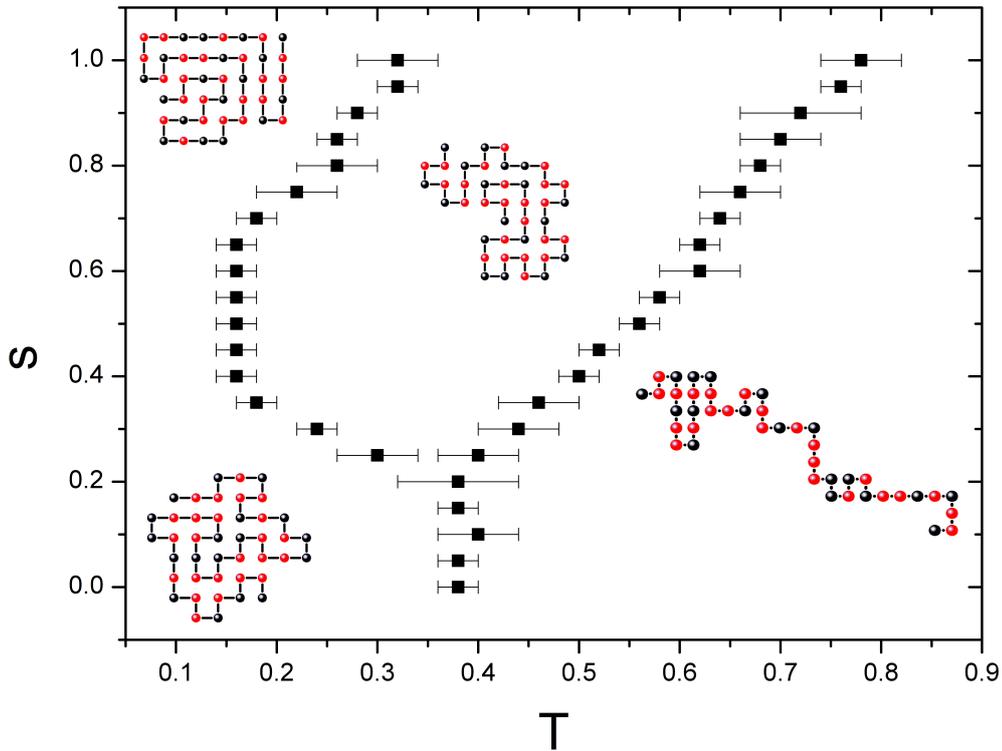


Figura 18: Pseudo diagrama de fases no espaço  $s - T$ . Nas conformações mostradas, um monômero H é representado por um círculo de cor vermelha, enquanto monômeros P são representados em cor preta.

e o lado direito a energia média em função da temperatura. Para o calor específico, é possível verificar, na mesma função, um ou dois pontos de máximos. Por exemplo, para  $s = 0,8$  é possível verificar dois picos no calor específico, que ocorre para  $T = 0,23$  e  $T = 0,80$ . Para as energias médias, os resultados estão de acordo com o esperado, assim como foi para o modelo em 2D. Os valores absolutos das energias médias diminuem à medida que a temperatura aumenta, já que os números de contatos  $n_{HH}$  e  $n_{HP}$  diminuem.

A Fig. 20 representa, à esquerda, o número de ligações hidrofóbicas ( $n_{HH}$ ) e, à direita, o número de ligações hidrofóbico-polar ( $n_{HP}$ ), ambos em função da temperatura. Esses resultados são semelhantes aos apresentados em 2D. Em geral, para baixas temperaturas,  $n_{HH}$  e  $n_{HP}$  são maiores que em temperaturas altas.

A distância ponta a ponta e o raio de giração, respectivamente à esquerda e à direita da Fig. 20, nos auxiliam a prever a forma da conformação da molécula. Se por exemplo desejamos encontrar as conformações que representam o estado fundamental das moléculas, ou seja, conformações geradas em baixas temperaturas, é possível selecionar conformações mais prováveis de serem estados fundamentais com os auxílios da distância

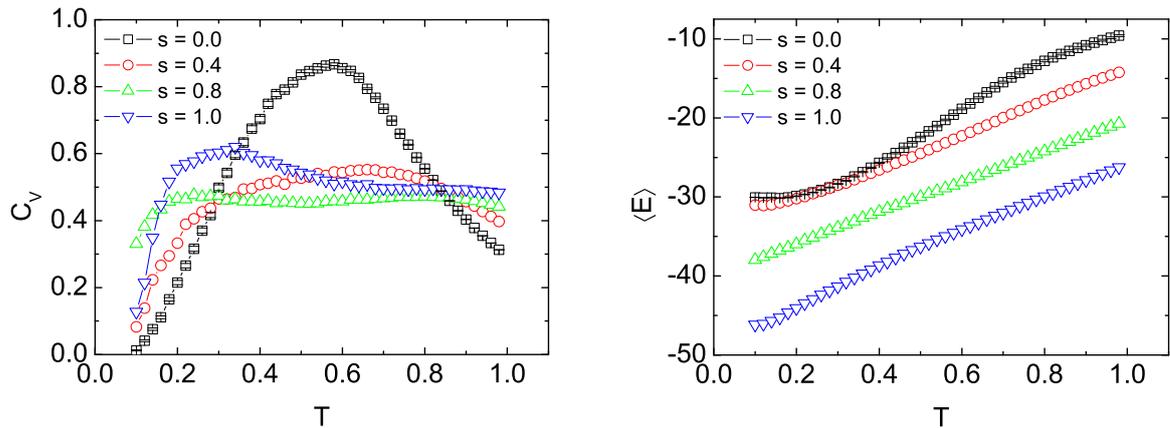


Figura 19: Calor específico (à esquerda) e energia média (à direita) como funções de temperatura para alguns valores do parâmetro  $s$ . As barras de erro, se não forem visíveis, são menores do que os símbolos.

ponta a ponta e o raio de giração. Os gráficos mostram que para baixas temperaturas as conformações são compactas. Esse comportamento não é verificado em 2D (para  $s = 0, 0$ ), conforme foi mostrado na Fig. 16. Portanto, conclui-se que no modelo HP em 2D, para  $s = 0, 0$ , os estados fundamentais não são tão compactos quanto em 3D.

Da mesma forma feita para o modelo bidimensional no espaço  $s-T$ , o diagrama de fases para a rede cúbica pode ser obtido através dos pontos de máximo do calor específico, em suas respectivas temperaturas. Em trabalhos posteriores, pretendemos refinar os resultados em 2D, o que significa analisar outros valores do parâmetro  $s$  e, assim, obter um diagrama de fases mais completo. Entretanto, de acordo com os resultados apresentados, é possível construir um esboço do diagrama, mostrado na Fig. 22. Além disso, esperamos que a forma final do diagrama seja semelhante aos formatos descritos pelas Figs. 23 e 24. Os contornos desenhados são as possíveis curvas de coexistência de fase. Para a Fig. 23, teríamos a existência de três fases configuracionais delimitadas pelas curvas de coexistência: uma no canto superior esquerdo, outra no canto inferior direito e uma terceira fase entre as duas anteriores. A Fig. 24, representa a possibilidade do diagrama de fases mostrar quatro fases configuracionais.

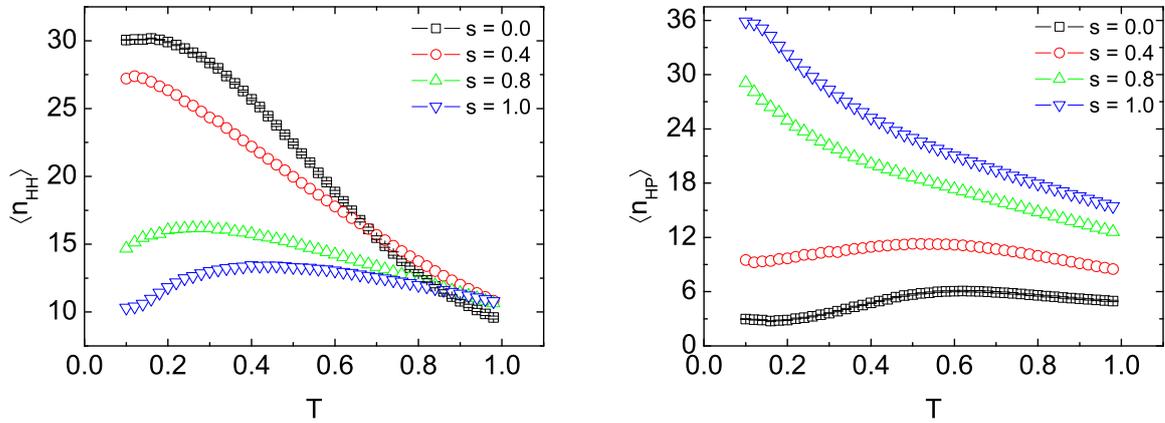


Figura 20: Esquerda: número médio de contatos do tipo hidrofóbico-hidrofóbico  $n_{HH}$ . Direita: número médio de interações hidrofóbico-polar  $n_{HP}$ . As barras de erro são da ordem do tamanho símbolos.

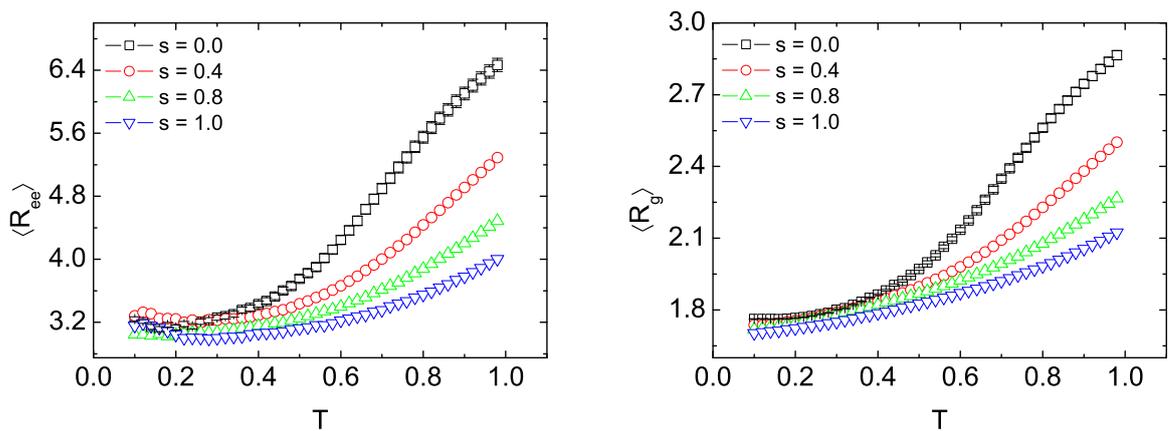


Figura 21: Valores médios da distância de ponta a ponta (esquerda) e do raio de giração (direita). Para melhor visualização, barras de erro são mostradas somente para  $s = 0,0$ . Para os outros valores de  $s$ , as barras de erro têm tamanhos semelhantes.

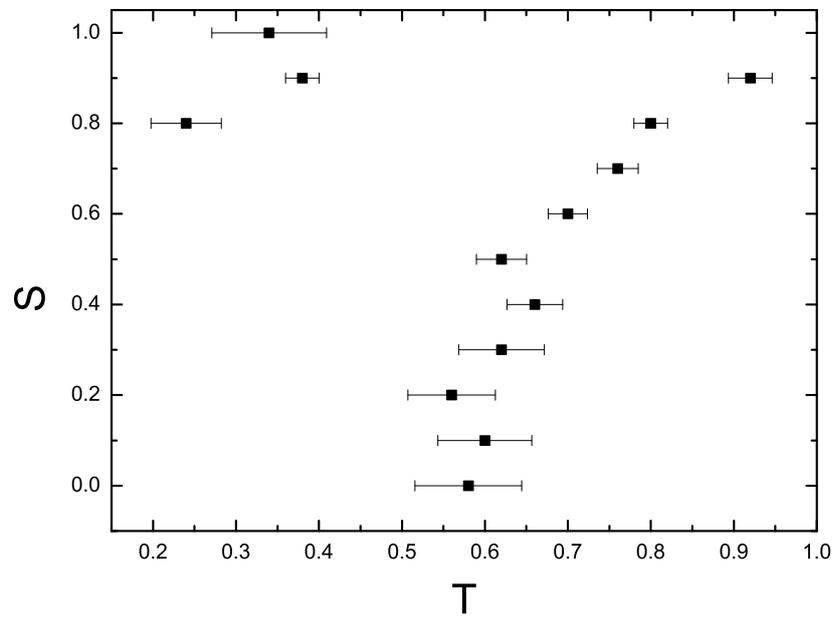


Figura 22: Pseudo diagrama de fases no espaço  $s - T$ .

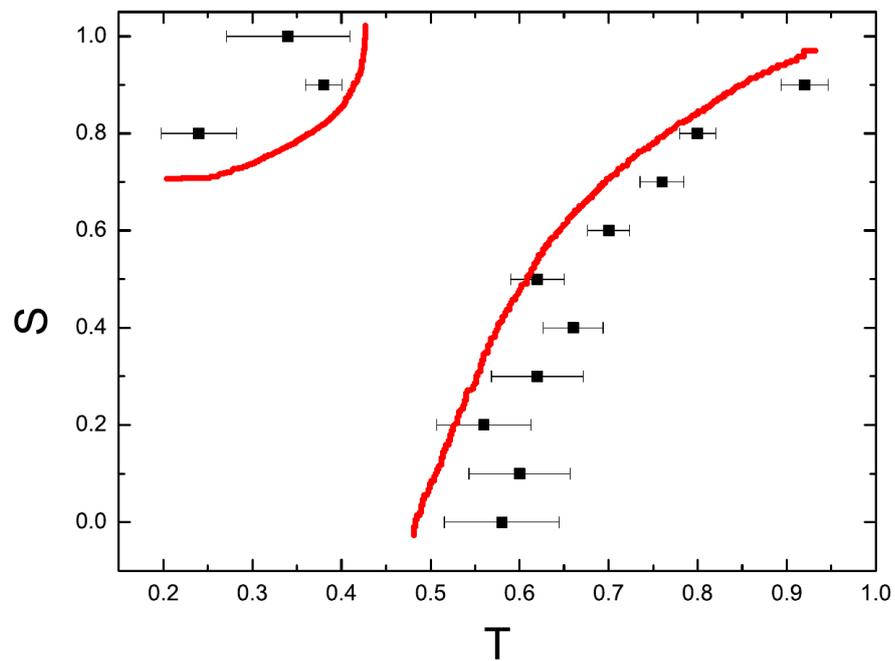


Figura 23: Esboço do diagrama de fases no espaço  $s - T$ , no cenário contendo 3 fases distintas.

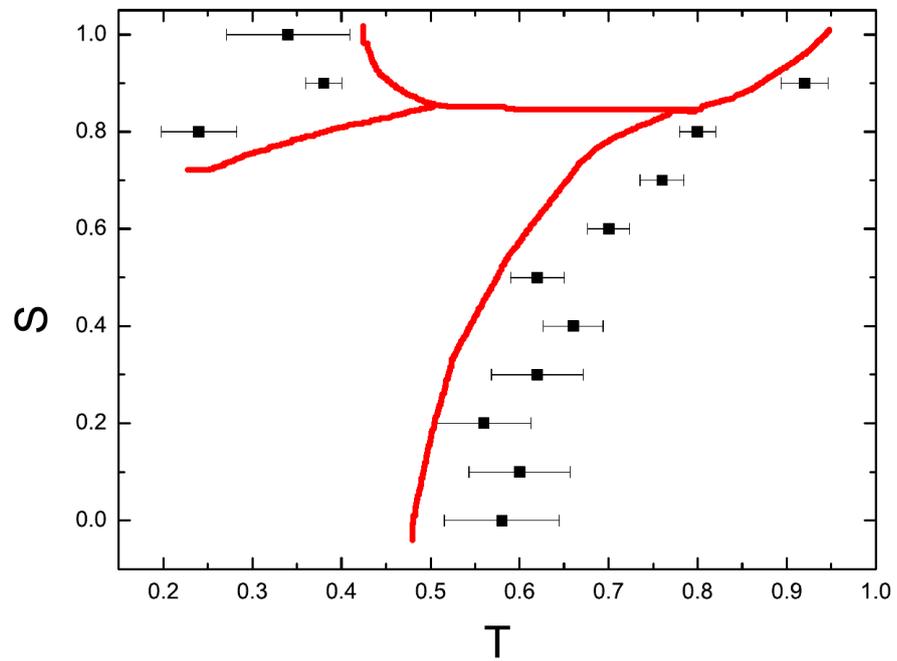


Figura 24: Esboço do diagrama de fases no espaço  $s - T$ , no cenário contendo 4 fases distintas.

## 5 CONCLUSÃO

O objetivo do trabalho era obter as “transições de fases” de uma cadeia polimérica na rede quadrada. Pelas razões expostas ao longo do texto, optou-se por se utilizar o modelo HP em uma sequência pré-determinada de 42 monômeros.

Para gerar as configurações, utilizou-se o algoritmo de crescimento de cadeia PERM (*pruned-enriched Rosenbluth method*). O PERM é baseado no algoritmo de Rosenbluth, porém favorece as “boas” cadeias e evita as “ruins”. Boas cadeias são aquelas que possuem grande probabilidade de atingirem o comprimento desejado, enquanto as ruins tem grandes chances de serem descartadas. Portanto gera-se um viés que é corrigido através de pesos diferentes atribuídos a cada cadeia. O algoritmo executa um procedimento complexo de replicação e descarte de cadeias, até que se atinja o número desejado de amostras.

Analisamos o número de contatos entre monômeros hidrofóbicos e entre hidrofóbicos e polares, obtendo a energia média do sistema. Para 21 valores de  $s$ , no intervalo  $0, 0 \leq s \leq 1, 0$ , foram obtidas as quantidades físicas: calor específico, energia média, número de contatos hidrofóbico-hidrofóbico, número de contatos hidrofóbico-polar, raio de giração e distância ponta a ponta.

Através dos picos do calor específico, construiu-se o diagrama de fases no espaço  $s - T$ . O resultado mais interessante do nosso estudo foi que, para  $s \geq 0, 2$ , o calor específico tem dois picos e conseqüentemente o diagrama de fases apresenta uma bifurcação. Esta bifurcação corresponde a uma fase intermediária, de estrutura globular, localizada entre as fases compacta e estendida.

Os resultados do caso bidimensional foram publicados em *Journal of Physics: Conference Series* [35].

Estudos preliminares do modelo em três dimensões também foram apresentados. Como trabalho futuro, pretendemos simular outros valores do parâmetro  $s$  no intuito de refinar o diagrama de fases.

## REFERÊNCIAS

- [1] BACHMANN, M. *Thermodynamics and Statistical Mechanics of Macromolecular Systems*. New York: Oxford University Press Inc., 2014.
- [2] STRYER, L. *Biochemistry*. New York: Freeman & Company W. H. 2012.
- [3] DILL, K.A. Principles of protein folding - A perspective from simple exact model. *Protein Science*. v.4, p.561-602, 1995.
- [4] PELCZAR, J.M.; CHAN, E.C.S.; NOEL, R.K. *Microbiologia*. 2<sup>a</sup> ed. São Paulo: Pearson, 1997.
- [5] FLORY, P.J. *Principles of Polymer Chemistry*. Ithaca: Cornell University Press, 1971.
- [6] LOPES, S. *Biologia*. São Paulo: Saraiva, 2005.
- [7] GUYTON, A.C.; HALL, J.E. *Tratado de Fisiologia Médica*. Rio de Janeiro: Elsevier, 2006.
- [8] LAU, K.F.; DILL, K.A. A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins. *Macromolecules*. v.22, p.3986–3997, 1989.
- [9] DILL, K.A. Theory for the Folding and Stability of Globular Protein. *Biochemistry*. v.24, p.1501-1509, 1985.
- [10] DILL, K.A.; SHORTLE, D. Denatured states of proteins. *Annual Review of Biochemistry*. v.60, p.795-825, 1991.
- [11] LI, Y.W.; WÜST, T.; LANDAU, D.P. Surface adsorption of lattice HP protein: Thermodynamics and structural transitions using Wang-Landau sampling. *Journal of Physics: Conference Series*. v.402, n.012046, 2011.
- [12] SCHIEMANN, R.; BACHMANN, M.; JANKE, W. Exact sequence analysis for three-dimensional hydrophobic-polar lattice proteins. *Chemical Physics*. v.122, n.114705, 2005.
- [13] MIYAZAWA, S.; JERNIGAN, R.L. Residue–Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *Journal of Molecular Biology*. v.256, p.623-644, 1996.
- [14] TANG, C. Simple models of the protein folding problem. *Physica A*. v.288, p.31-48, 2000.
- [15] DILL, K.A. Dominant forces in protein folding. *Biochemistry*. v.29, p.7133-7155, 1990.
- [16] SCHERER, C. *Métodos computacionais da Física*. São Paulo: Editora Livraria da Física, 2005.
- [17] SALINAS, S.R.A. *Introdução à Física Estatística*. 2 ed. São Paulo: Editora da Universidade de São Paulo, 2008.

- [18] PATHRIA, R.K.; BEALE, P.D. *Statistical Mechanics*. 3 ed. New York: Elsevier Ltd, 2011.
- [19] PRELLBERG, T. *From Rosenbluth Sampling to PERM - rare event sampling with stochastic growth algorithms, Lecture given at the International Summer School Modern Computational Science*. 2012.
- [20] CASQUILHO, J.P.; TEIXEIRA, P.I.C. *Introdução à Física Estatística*. São Paulo: Editora Livraria da Física, 2012.
- [21] OLIVEIRA, M.J. *Termodinâmica*. São Paulo: Editora Livraria da Física, 2005.
- [22] STANLEY, H.E. *Introduction to Phase Transition and Critical Phenomena*. New York: Oxford University Press Inc., 1987.
- [23] GREINER, W.; NEISE, L.; STOCKER, H. *Thermodynamics and Statistical Mechanics*, New York: Springer, 1997.
- [24] VOGEL, T.; BACHMANN, M.; JANKE, W. Freezing and collapse of flexible polymers on regular lattices in three dimensions. *Physical Review E*. v.76, n.061803, 2007.
- [25] FINKELSTEIN, A.V.; PTITSYN, O.B. *Protein Physics*. New York: Academic Press, 2002. 359p.
- [26] LANDAU, D.P.; BINDER, K. *A Guide to Monte-Carlo Simulations in Statistical Physics*. 3 ed. New York: Cambridge University Press., 2009.
- [27] BACHMANN, M.; SCHIEMANN, R.; JANKE, W. *Computer Physics Communications*. Exact enumeration of three-dimensional lattice proteins. v.166, p.8-16, 2005.
- [28] GRASSBERGER, P. Pruned-enriched Rosenbluth method: Simulation of  $\theta$  polymer of chain length up to 1000000. *Physical Review E*. v.3, n.3682, 1997.
- [29] ROSENBLUTH, M.N. Monte Carlo Calculation of the Average Extension of Molecular Chains. *The Journal of Chemical Physics*. v.23, n.2, p.356-359, 1955.
- [30] GENNES P.G. *Scaling Concepts in Polymer Physics*. Ithaca: Cornell University Press, 1988.
- [31] LENNARD-JONES, J.E. Cohesion. *Proceedings of the Physical Society*. v.43, n.240, p.461-482, 1931.
- [32] UENOJO, M.; PASTORE, G.M. Biotransformação de Limoneno: uma revisão das principais rotas metabólicas. *Química Nova*. v.30, n.2, p.388-394, 2007.
- [33] BACHMANN, M.; JANKE, W. Density of States for HP Lattice Proteins *Acta Physica Polonica B*. v.34, n.10, 2003.
- [34] HARMON, M.G.; BAKER, T.P. *ACM SIGADA Ada Letters*. An Ada implementation of Marsaglia's "universal" random number generator. v.8, p.110-112, 1988.
- [35] SANTOS, F.S.H.; MARTINS, P.H.L.; BACHMANN, M. Chain-growth simulations of the HP model for proteins. *Journal of Physics: Conference Series*. v.686, n.012003. 2016.