UNIVERSIDADE ESTADUAL PAULISTA

"Júlio de Mesquita Filho"

Instituto de Geociências e Ciências Exatas – IGCE

Campus de Rio Claro

Método de Wang-Landau para sequenciamento de aminoácidos em estrutura nativa de proteínas em modelos de rede.

Renato Luciano Cagnin

Orientador : Prof. Dr. Makoto Yoshida

Dissertação de Mestrado elaborada junto ao Programa de Pós-Graduação em Física Área de Física Aplicada para obtenção do título de Mestre em Física

RIO CLARO – SP

2010

621 Cagnin, Renato Luciano
C131m Método de Wang-Landau para sequenciamento de aminoácidos em estrutura nativa de proteínas em modelos de rede / Renato Luciano Cagnin. - Rio Claro : [s.n.], 2010
79 f. : il., figs., gráfs., tabs.

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

1. Física aplicada. 2. Desenho de proteínas. 3. Enovelamento. 4. Método Wang-Landau. 5. Sequenciamento de aminoácidos. 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. Edson Denis Leonel

Renato Luciano Cagnin

Rio Claro, 15 de abril de 2010.

Resultado : **Aprovado**

À minha namorada Priscila e à minha família por me darem o apoio necessário para a conclusão deste trabalho.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por todas as experiências positivas e negativas que passei.

À minha namorada Priscila Arjona, pelos conselhos, palavras de motivação, e por ser a pessoa que mais me amparou em momentos de dificuldades com seu amor puro e verdadeiro.

Aos meus pais Vlademir Armando Cagnin e Clair Aparecida Volet Cagnin por tudo que fizeram ao longo da minha caminhada nesta vida.

Ao meu irmão Adriano Rogério Cagnin por ser uma de minhas figuras de admiração, por sua capacidade e inteligência.

À minha irmã Patrícia Fernanda por toda a ajuda que me deu ao longo da minha caminhada.

Ao meu grande amigo e praticamente irmão Vinícius Soares por sua amizade e companheirismo ao longo de todos estes anos.

Ao meu grande amigo Tommy por todos estes anos de amizade.

Aos grandes amigos Julio César e Cínthia pelos bons momentos que passamos.

Aos amigos William, Renato, Márcio e Vinícius pela troca de conhecimento e companheirismo.

Aos amigos Rafael, Bruno, Luís Felipe, Leandro pelos tempos de república e aos demais colegas de graduação e pós-graduação.

Aos amigos do curso de Ciência da Computação e professores.

Ao Prof. Dr. Makoto Yoshida pela orientação deste trabalho e experiências passadas.

À Prof. Alzira pelos conselhos e por toda ajuda prestada na elaboração deste trabalho.

Aos demais professores do Departamento de Física que me auxiliaram durante minha graduação e pós-graduação, meus mais sinceros agradecimentos.

À CAPES (Coordenação de aperfeiçoamento de pessoal de nível superior) pela concessão da bolsa ao longo desse trabalho.

APOIO CAPES (Coordenação de aperfeiçoamento de pessoal de nível superior)

SUMÁRIO

ÍNDICE

RESUMO:

Neste trabalho de dissertação, apresentamos uma técnica de se construir sistematicamente sequências de aminoácidos que, ao serem dispostas ao longo de uma cadeia previamente conhecida, resultam em cadeias que se comportam como proteínas. Cada cadeia de aminoácidos, a uma dada temperatura, deve assumir uma forma funcional denominada estrutura nativa, não degenerada, na qual sua energia é a de menor valor possível. A técnica está baseada em um método Monte Carlo, introduzido por Wang e Landau, para se estudar transição de fases em sistemas magnéticos e que neste trabalho foi adaptada e aplicada para se desenhar proteínas. Para se verificar a eficiência do método, foi adotado o modelo de rede para proteínas, onde as cadeias são compostas por 27 monômeros interagindo através do potencial de Miyazawa-Jernigan e 20 tipos de aminoácidos. Um elevado número de sequências foram sintetizadas e todas foram sistematicamente testadas para verificar se cumpriam os requisitos de proteína. Com os resultados obtidos pôde-se verificar o sucesso da implementação da técnica. Trata-se então de uma ferramenta muito interessante e eficiente para o estudo do problema do enovelamento de proteínas.

Palavras – Chave : Desenho de proteínas, Enovelamento, Método de Wang-Landau, Sequenciamento de aminoácidos.

ABSTRACT:

In this dissertation, we present a technique to search and order sequences of amino acids placed along a known chain to build one that behaves as a protein. At a given temperature, each designed sequence should fold to a special nondegenerated conformation known as native state. The energy of the sequence in this state is the lowest one. The technique is based on a Monte Carlo method, introduced by Wang and Landau, to study phase transition of magnetic systems and in this work was adapted and applied to protein design. We adopted the lattice model protein composed of 27 monomers interacting through the Miyazawa-Jernigan potencial and with 20 types of different amino acids. Many sequences were synthesized and all of them were systematically verified if they fulfilled the protein requirements and to check the efficiency of this method. The obtained results showed the success of the implemantation of this technique. Therefore, it is one more very interesting tool to efficiently study the protein folding problem.

Key words : Protein design, Folding, Wang-Landau method, Amino acid sequencing.

Lista de Tabelas

Lista de figuras

Capítulo I - INTRODUÇÃO

As proteínas são biopolímeros de alto peso molecular sintetizados pelas células. Geralmente são estruturadas em longas cadeias contendo de centenas a milhares de aminoácidos e executam uma diversidade de funções biológicas essenciais para quaisquer organismos vivos. Quando transcritas a partir do RNA das células, as cadeias proteicas encontram-se numa forma não funcional, geralmente uma cadeia polimérica aberta, caracterizada somente por sua composição e sequência de aminoácidos. Tal estrutura recebe o nome de estrutura primária [1]. Suas funções somente serão executadas quando, após um enovelamento, induzido ou espontâneo, as cadeias adquirem uma conformação tridimensional única, cuja forma está relacionada à sequência de aminoácidos. Esta é a estrutura nativa na qual as proteínas se encontram aptas para a execução de suas funções. É observado que depois do enovelamento ou dobramento, as cadeias protéicas encontram-se na estrutura de menor energia e de maior estabilidade termodinâmica [1-3]. Contudo, ainda não se compreende como ocorre o processo de enovelamento, e muito esforço tem sido concentrado para se resolver esse problema nas mais diversas áreas do conhecimento como a Física, Química, Biologia, Bioengenharia e Computação. Esse problema, constitui-se hoje no chamado **problema do enovelamento de proteínas** [2-3]. As principais questões envolvendo o enovelamento de proteínas, são:

- dada uma sequência de aminoácidos, como predizer a estrutura funcional da proteína, conhecida como **estrutura nativa** da proteína?
- como a sequência de aminoácidos determina a estrutura nativa? A sequência de aminoácidos em uma proteína é na verdade um código que determina a forma da estrutura nativa. Como decodificar e identificar a correlação entre a sequência de aminoácidos e a estrutura nativa?
- como a proteína encontra o caminho para se enovelar e atingir a estrutura nativa? Por que, contrariando a expectativa, o tempo de enovelamento é assim tão rápido apesar da complexidade da estrutura nativa?
- como as interações entre os aminoácidos e a interação com o meio aquoso (pressão hidrofóbica) contribuem para o processo de enovelamento ?

Para responder a essas perguntas, existem atualmente duas linhas de investigação do problema do enovelamento. Aquela conhecida como a do **enovelamento direto** [2,3], tenta predizer a estrutura nativa partindo de uma sequência de aminoácidos conhecida. Considerando apropriadamente as interações entre os monômeros da cadeia e entre os monômeros e o meio aquoso no qual a proteína está diluída, tenta-se investigar como ela encontra espontaneamente o caminho para se enovelar e assumir a sua forma funcional. Esta estrutura funcional, conhecida como a estrutura nativa, é única.

A segunda maneira de investigar esses mesmos problemas, mas igualmente importante, consiste no método do **enovelamento inverso** ou do **desenho de proteínas** [4,5]. A idéia consiste em adotar inicialmente o esqueleto de uma estrutura nativa conhecida de uma proteína, denominada estrutura alvo, e preencher o esqueleto com sequências de aminoácidos. Naturalmente esse preenchimento não pode ser feito arbitrariamente, mas essas sequências deverão ser sistematicamente construídas ou desenhadas. O procedimento consiste em escolher uma composição de aminoácidos conhecida, e considerando apropriadamente as interações entre os monômeros da proteína e deles com o meio onde se encontra a proteína, construir todas as sequências que poderiam ser dispostas ao longo do esqueleto da estrutura nativa. Logo após, deve-se verificar quais destas sequências produziriam cadeias poliméricas que possuem propriedades de uma proteína e que, ao se enovelarem, assumam a forma da estrutura nativa proposta inicialmente. Uma informação vital é que, esta proteína desenhada, quando na estrutura nativa, esteja em uma conformação de menor energia e seja única. Esta

linha de desenhar a proteína, também é intensivamente adotada atualmente [5,6,7]. Na realidade, o enovelamento direto de proteínas e o desenho de sequências ou enovelamento inverso, são duas diferentes formulações de um mesmo problema.

Além das motivações pela busca do conhecimento básico para a compreensão do processo de enovelamento, existem as de interesse prático e tecnológico. Se por um lado as proteínas são responsáveis pela existência da vida, elas também podem ser a causa de diversas anomalias e doenças atribuídas às falhas no enovelamento. Erros no enovelamento podem provocar o agregamento das proteínas causando doenças neuro-degenerativas, como mal de Alzheimer [8] e mal de Parkinson. Inclui-se também entre as doenças causadas por falhas no enovelamento, a doença da vaca louca (Encefalopatia Espongiforme Bovina–BSE) [9], certo tipo de enfisema pulmonar e alguns tipos de câncer. Assim, a investigação através do enovelamento direto e a investigação do enovelamento através do desenho de proteínas, são fundamentais no desenvolvimento de novos medicamentos para o tratamentos desses males. Outras aplicações podem ser apontadas no ramo da biotecnologia, que incluem a produção de novos agentes catalíticos biológicos e químicos, bio-sensores, hormônios e agentes reguladores biológicos [10].

O objetivo desta dissertação é investigar o problema de enovelamento de proteínas através da linha do **enovelamento inverso** ou **desenho de proteínas**. O enovelamento direto, já foi objeto de estudo na dissertação de Fábio Beig [47]. Vamos propor uma técnica de se construir ou desenhar sequências de aminoácidos em modelos de rede para a proteína, baseada no método de Wang-Landau, um método de Monte Carlo originalmente aplicado para o estudo de transições de fase em sistemas magnéticos [11]. São modelos simples de proteínas, bastante idealizadas, mas que se comportam satisfatoriamente como uma proteína [12]. Ainda que o modelo seja simples, o estudo do enovelamento continua bastante complexo e se enquadra na classe de problemas de muitos corpos correlacionados, bastante semelhantes aos problemas de momentos magnéticos interagentes em redes, como os descritos pelo modelo de Ising , intensivamente estudados na área da física estatística.

A dissertação está estruturada de forma que, no capítulo II, será apresentada uma síntese do que é uma proteína, sua composição e classificação quanto às suas formas estruturais. Apresenta-se também uma revisão de métodos de Monte Carlo, usualmente aplicados em investigações em mecânica estatística e que serão utilizados ao longo desse

trabalho, tanto no estudo do enovelamento como no desenho de proteínas. Inicia-se com o algoritmo de Metropolis, que será aplicado na investigação de aspectos dinâmicos do enovelamento da cadeia assim como na seleção de sequências de aminoácidos no processo de desenhar uma proteína. Será apresentado um modelo de rede para a proteína, e nele serão explicitados os tipos de interações existentes entre os monômeros da cadeia, assim como as interações dos diferentes monômeros com o meio onde ocorre o processo de enovelamento. Através desse modelo, será apresentado em detalhes o método para simular o movimento espacial da cadeia em contacto com um a fonte de calor utilizando o algoritmo de Metropolis. Utilizando este mesmo modelo de rede, também será apresentado em detalhes o método introduzido por Gutin e Shakhonovich [13,14] para desenhar uma proteína utilizando do algoritmo de Metropolis. No capítulo III, será apresentado o método de Wang-Landau [11] acompanhado de uma discussão detalhada e a extensão e adaptação do método para se estudar o enovelamento de proteínas [4,5]. Nele, será também apresentada a contribuição desta dissertação. Baseado no método de Wang-Landau, propõe-se um método para selecionar sequências de aminoácidos e desenhar uma proteína em modelo de rede. Na construção destas sequências, as interações entre os monômeros da cadeia assim como a interação deles com o meio serão devidamente consideradas. Para completar, também será apresentada a adaptação do método de Wang-Landau para o cálculo de propriedades termodinâmicas de uma cadeia de aminoácidos em contacto com uma fonte de calor. Estes cálculos serão necessários para caracterizar as proteínas desenhadas e testar a estabilidade termodinâmica delas.

No capítulo IV, serão apresentados os resultados obtidos das simulações e no capítulo V serão apresentadas as conclusões assim como sugestões para trabalhos futuros.

Capítulo II – ENOVELAMENTO DE PROTEÍNAS E O MÉTODO DE MONTE CARLO

Neste capítulo, descreveremos as proteínas, sua composição, classificações e algumas de suas principais propriedades. Apresentaremos proteínas em modelo de rede, faremos uma revisão dos métodos utilizados em simulações numéricas do enovelamento e do processo estocástico usado para desenhar sequências de aminoácidos nas proteínas. Os métodos de Monte Carlo apresentados neste capítulo são baseados no algoritmo de Metropolis.

2.1 Proteínas

Proteínas naturais são macromoléculas lineares encontradas em todas as partes de todas as células, uma vez que são fundamentais às funções celulares. Existem muitas espécies diferentes de proteínas, cada uma sintetizada para uma função biológica. Além disso, a maior parte das informações genéticas são expressas pelas proteínas. Todas contêm carbono, hidrogênio, nitrogênio e oxigênio, e quase todas contêm enxofre. Algumas proteínas contêm elementos adicionais, particularmente fósforo, ferro, zinco e cobre. Seu peso molecular é extremamente elevado. A função de uma proteína está relacionada em especial com sua forma e as proteínas são classificadas em três grandes grupos: (i) estruturais; (ii) de membrana e (iii) globulares [1,15].

As proteínas estruturais ou fibrosas, são constituídas de feixes ou placas de longas cadeias lineares de aminoácidos, formando micro filamentos e micro tubos, exercendo a função de sustentação e estruturação em meio biológico. As proteínas de membrana são responsáveis pela intermediação de substâncias que entram e saem da célula, agindo como receptoras de substâncias para a função celular, atuando como enzimas para catalisar reações na superfície da membrana e como marcadoras das células provenientes de outros organismos. Finalmente, as proteínas globulares constituem a quase totalidade das estruturas espaciais conhecidas e depositadas no PDB (*protein data base*) [16] , e é a classe de proteínas que se pretende simular neste trabalho. Desempenham funções diversificadas, como ação enzimática, transporte, função reguladora (hormônios) e como fator de crescimento FGF (Fibroblast Growth Factor). Funcionam também como anticorpos, atuam na coagulação e na produção de energia e fazem parte do material cromossômico. De modo geral, a forma, a regulação, a preservação e a reprodução dos seres vivos são controladas pelas proteínas globulares.

Os componentes das proteínas são os aminoácidos. Existem somente 20 tipos de aminoácidos naturais e estes são classificados dentre várias características, segundo seu caráter hidrofóbico (apolar) ou polar. Todas as proteínas, independentemente de sua função ou espécie de origem, são construídas a partir desse conjunto básico, arranjados em várias sequências específicas.

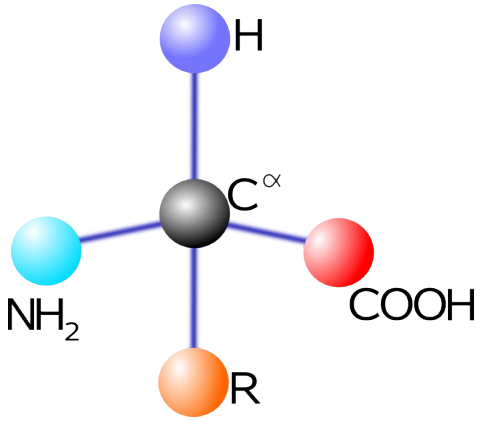


Fig. 2.1 – Estrutura geral de um aminoácido. Em geral, todo aminoácido é composto por um carbono central (C_α) ligado a um átomo de hidrogênio (H), um grupo amina (NH_2), um grupo carboxila (COOH) e a um radical (R), que define suas propriedades físico-químicas.

Todo aminoácido é constituído por um átomo de carbono, chamado carbono-alfa (C_α), ligado a quatro grupos químicos: (i) grupo carboxila (COOH), (ii) grupo amina (NH_2), (iii) a um átomo de hidrogênio (H) e (iv) a um radical (R) (Fig. 2.1), também denominado cadeia lateral do aminoácido; o radical é o que distingue os 20 aminoácidos naturais entre si. Os 20 aminoácidos naturais das proteínas são Glicina, Alanina, Valina, Leucina, Isoleucina, Serina, Treonina, Cisteína, Metionina, Prolina, Ácido Aspártico, Asparagina, Ácido Glutâmico, Glutamina, Arginina, Lisina, Histidina, Fenilalanina, Tirosina, Triptofano (Tabela. 2.1).

Os vinte aminoácidos que compõem as proteínas			
Aminoácido	**Abreviação**	**Símbolo**	**Polaridade**
Alanina	Ala	A	Apolar
Cisteína	Cys	C	Polar
Ácido aspártico	Asp	D	Polar
Ácido glutâmico	Glu	E	Polar
Fenilalanina	Phe	F	Apolar
Glicina	Gly	G	Apolar
Histidina	His	H	Polar
Isoleucina	Ile	I	Apolar
Lisina	Lys	K	Polar
Leucina	Leu	L	Apolar
Metionina	Met	M	Apolar
Asparagina	Asn	N	Polar
Prolina	Pro	P	Apolar
Glutamina	Gln	Q	Polar
Arginina	Arg	R	Polar
Serina	Ser	S	Polar
Treonina	Thr	T	Polar
Valina	Val	V	Apolar
Tirosina	Tyr	Y	Polar
Triptofano	Trp	W	Apolar

Tabela. 2.1 – Os 20 aminoácidos naturais classificados de acordo com sua polaridade.

Nas proteínas, os aminoácidos são unidos por ligações consecutivas entre o carbono da carboxila de um aminoácido e o nitrogênio do grupo amino do aminoácido seguinte, formando cadeias lineares. Esta ligação que une os aminoácidos é conhecida por ligação peptídica e é o resultado da condensação do grupo carboxila de um aminoácido com o grupo amino de outro, e a eliminação de uma molécula de água (Fig. 2.2).

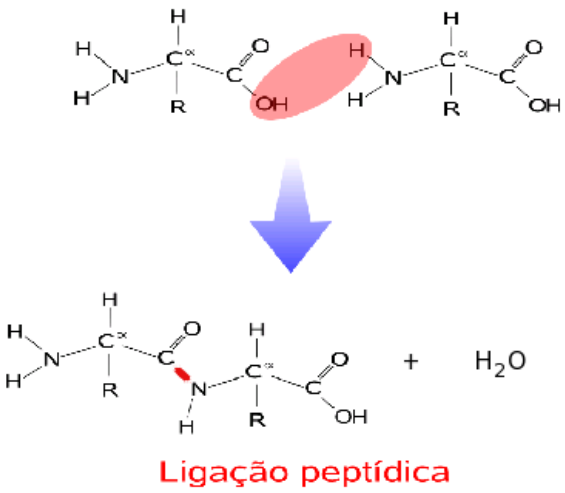


Fig. 2.2 – A ligação peptídica que une os aminoácidos ao longo da cadeia é uma ligação covalente resultante da reação de condensação entre dois aminoácidos.

Uma proteína específica é identificada por sua sequência particular de aminoácidos, da qual provém as informações específicas para a determinação de sua estrutura 3-D particular e sua funcionalidade. Em geral, as proteínas são descritas através de quatro níveis estruturais: da estrutura primária à quaternária. A estrutura "primária" corresponde à sequência de aminoácidos da proteína propriamente dita, isto é, descreve sua estrutura linear ou unidimensional. Ela pode variar em três aspectos fundamentais: número de aminoácidos (ou tamanho da cadeia), sequência de aminoácidos e composição (natureza dos aminoácidos). É o nível estrutural mais simples e mais importante, pois dele deriva todo o arranjo espacial da molécula. As estruturas secundárias não serão abordadas nas simulações deste trabalho. Referem-se a padrões estruturais da cadeia peptídica caracterizadas pelos contatos locais entre aminoácidos. Corresponde às cadeias na conformação de alfa hélices e folhas beta. As alfa

hélices são estruturas que possuem seus aminoácidos contactantes separados por poucas unidades ao longo da cadeia. Estas interações são intermediadas por pontes de hidrogênio que contribuem na estabilização da hélice. Já as folhas beta são associações lado a lado de diversas partes da cadeia polipeptídica estendidas em forma de fitas e unidas também por ligações de hidrogênio entre os grupos amino e carboxila. Geralmente as proteínas globulares podem ser formadas por alfa hélices somente, ou somente por folhas-beta, ou ainda da forma mista [1,15].

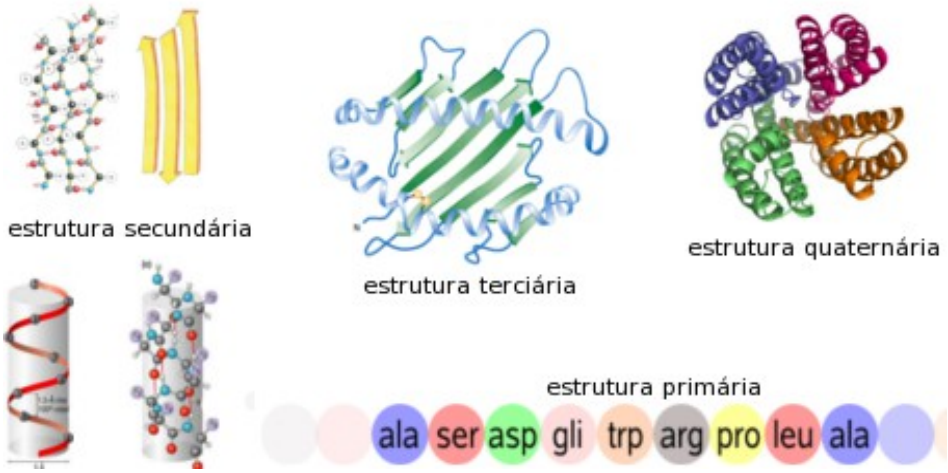


Fig. 2.3 – Organização estrutural em proteínas

As estruturas terciárias são definidas como a disposição dos aminoácidos no espaço, descrevendo como as estruturas secundárias se arranjam na conformação globular final da proteína. A forma das proteínas está relacionada com sua estrutura terciária. Esta estrutura resulta de interações físicas e químicas que se estabelecem entre as cadeias laterais dos aminoácidos e dessas com o meio aquoso. Essas interações estabilizam termodinamicamente a estrutura de forma apropriada, condição esta necessária para a efetiva atividade biológica das proteínas. A proteína somente exercerá alguma de suas funções biológica se estiver em sua forma tridimensional nativa. A estrutura terciária está relacionada com as torções e dobraduras da cadeia protéica sobre si mesma e ocorrem nas proteínas globulares mais complexas estrutural e funcionalmente. Este tipo de estrutura determina a funcionalidade da proteína, estabilizada por pontes de hidrogênio, interações hidrofóbicas entre aminoácidos e pontes dissulfeto.

Por fim, a estrutura quaternária refere-se ao modo pelo qual duas ou mais cadeias polipeptídicas interagem.

2.2 O enovelamento de proteína

Depois de transcrita a partir do RNA, a nova proteína dobra-se sobre si mesma num processo físico, causado pela interação com o solvente, até conformar-se numa estrutura compacta e tornar-se funcional e ativa. Este estado no qual a proteína executa sua função biológica é denominado de estado nativo. O estado nativo de uma proteína globular possui algumas propriedades importantes. É altamente compacto e os monômeros hidrofóbicos tendem a se concentrar no centro da proteína formando o denominado núcleo hidrofóbico, indicando que as interações hidrofóbicas levam à compactação da proteína. Finalmente, o estado nativo é geralmente termodinamicamente estável.

Já no ano de 1936, investigações eram feitas para explicar esse processo. Sabe-se que a distribuição de forças que atuam durante o enovelamento da proteína para levá-la à sua única estrutura nativa compacta está codificada em sua sequência de aminoácidos. Mirsky e Pauling [15] propuseram em 1936 que as pontes de hidrogênio fossem as forças dominantes no enovelamento. Contudo, por volta de 1950, Kauzmann [17] apontou que as pontes de hidrogênio não favoreceriam o estado nativo pois elas seriam formadas não somente entre elementos da cadeia, mas estariam presentes também na interação da cadeia com o solvente, e seriam tão fortes quanto as primeiras. Portanto, as interações hidrofóbicas seriam as forças dominantes no enovelamento de proteínas. Assim, quando a cadeia é imersa em água, a ação das forças entre os monômeros da cadeia e destes com a água, obriga os monômeros a se acomodarem de tal forma que, ao atingirem o equilíbrio, os monômeros hidrofóbicos ocupem uma região com pouco contacto com a água (núcleo hidrofóbico) e os hidrofílicos ocupem a região de maior contacto com a água, formando uma camada que isola os monômeros hidrofóbicos do solvente. Esse conjunto de forças resultante desse processo é comumente denoninado de pressão hidrofóbica. Esse processo cinético de acomodação dos aminoácidos até atingir o equilíbrio denomina-se enovelamento de proteína.

A partir desta descrição qualitativa sobre o processo de enovelamento é possível inferir que a distribuição de forças que atuam durante o enovelamento da proteína está

codificada em sua sequência de aminoácidos, ou seja, a sequência de aminoácidos que formam a estrutura primária bem como sua composição estão correlacionados com essa distribuição de forças. Entretanto, essa correlação entre a sequência de aminoácidos e a forma da estrutura nativa, ainda não é bem compreendida, já que várias sequências podem se enovelar na mesma estrutura nativa. Uma segunda questão fundamental é conhecida como o paradoxo de Levinthal [18]. Contrariando a expectativa para uma estrutura extremamente complexa, uma cadeia protéica se enovela em um intervalo de tempo muito curto. O que se sabe é que o processo de enovelamento ocorre em duas fases, a primeira é caracterizada pela rápida compactação da cadeia e uma segunda, mais lenta quando a proteína busca a estrutura de menor energia. Entretanto, é um problema que continua sob intensa investigação.

2.3 O Método de Monte Carlo e o Algoritmo de Metropolis

O método de Monte Carlo caracteriza-se por ser um método numérico que se utiliza de uma sequência de números gerados aleatoriamente com o objetivo de simular o comportamento de um sistema físico. Diferentes fenômenos físicos podem ser explorados usando esse método, e propriedades de sistemas de partículas interagentes tais como modelos para materiais magnéticos, ligas metálicas, superfícies absorventes, polímeros, fluidos simples e complexos, entre outros, têm sido exaustivamente estudadas através do método de Monte Carlo [19]. Na realidade, o uso do método não se restringe somente à Física, e pode ser encontrado em diversas áreas do conhecimento. De maneira geral, ele aplica-se tanto a problemas tradicionalmente tratados como determinísticos, quanto aqueles de natureza estocástica, contudo em ambos os casos, há sempre a presença de aleatoriedade, característica natural destes métodos. O nome do método vem de uma referência à capital mundial dos jogos de azar, no principado de Mônaco.

Os métodos de Monte Carlo são extremamente úteis no estudo de sistemas que envolvem muitos graus de liberdade e acoplamentos entre seus constituintes visto que uma solução analítica para tais sistemas torna-se impraticável. Particularmente no estudo de problemas de transição de fase e de criticalidade, o método tem sido largamente utilizado, demonstrando ser uma ferramenta extremamente poderosa [20].

Um algoritmo clássico e talvez o mais utilizado em simulações através do método de Monte Carlo é o algoritmo de Metropolis [21]. É um algoritmo que propicia a construção

sistemática de estados acessíveis a um sistema em equilíbrio com um banho térmico, sem que haja a necessidade de visitar a enorme quantidade de estados possíveis do sistema estudado. Neste procedimento, procura-se assegurar que os estados sejam visitados de acordo com suas respectivas importâncias determinadas pelo fator de Boltzmann. Cada estado gerado depende apenas do estado anterior, e esse processo de evolução temporal do sistema é denominado processo Markoviano [22]. Se *P(i,t)* é a probabilidade de se obter um estado *i* num instante t, e *w(i → j)* a taxa de transição para que o sistema passe do estado microscópico *i* para *j*, então em um processo do tipo Markoviano, a evolução de *P(i,t)* é determinada pela equação mestra

$$P(i,t+\Delta t)=P(i,t)+\left[\sum_j P(j,t).w(j\to i)-\sum_j P(i,t).w(i\to j)\right]\Delta t. \qquad (2.1)$$

No estado estacionário, quando $t\to\infty$, espera-se que $P(i,t) \to P_i$, ou seja, que a probabilidade de ocorrência de um particular estado *i*, no equilíbrio, não dependa mais do tempo, e assim obtemos a condição

$$\sum_{(j)} P(j,t).w(j\to i)-\sum_{(j)} P(i,t).w(i\to j)=0\,, \qquad (2.2)$$

e para que a reversibilidade microscópica seja obedecida deve-se exigir a condição de simetria

$$P(j,t).w(j\to i)=P(i,t).w(i\to j)\,, \qquad (2.3)$$

conhecida como o princípio do balanço detalhado. Uma forma válida e simplificada de reescrever a equação acima é considerar a situação de equilíbrio, onde se define $P_{eq}(i,t) \equiv P_i$,

$$P_j.w(j\to i)=P_i.w(i\to j), \qquad (2.4)$$

o que implica em

$$w(i\to j)=w(j\to i).\frac{P_j}{P_i}. \qquad (2.5)$$

Também, no equilíbrio de um sistema termodinâmico, caracterizado pelo ensemble

canônico, tem-se que $P_i = exp(- E_i / k_B T) / Z$, onde Z é a função partição do sistema. Por meio desta igualdade, torna-se explícito que a mudança de um estado para outro não depende do conhecimento prévio da função de partição Z do sistema, isto é

$$\frac{w(i \to j)}{w(j \to i)} = e^{-\Delta E / k_B T}, \tag{2.6}$$

onde $\Delta E = E_j - E_i$. Sempre que a mudança de estado reduz a energia do sistema, a nova configuração é aceita com probabilidade igual a 1. Baseado no equilíbrio detalhado, o algoritmo de Metropolis é descrito pelos seguintes passos:

(i) gera-se uma configuração inicial qualquer do sistema;

(ii) verifica-se se a configuração gerada viola alguma condição física do sistema; no caso afirmativo, a configuração é rejeitada e retorna-se ao passo anterior, senão, vai para o passo seguinte;

(iii) uma nova configuração é gerada e é aceita ou não, de acordo com as possibilidades:

$$w(i \to j) = \begin{cases} e^{-\Delta E / k_B T}, & se \ \Delta E > 0 \\ 1, & se \ \Delta E \leq 0 \end{cases}. \tag{2.7}$$

Uma condição para a aplicação deste algoritmo é que w seja sempre positivo ou nulo para evitar problemas de ergodicidade (a frequência de um determinado evento é a mesma em qualquer instante de tempo), que podem interferir nos valores médios das observáveis a serem estimadas.

2.4 Simulações do Enovelamento de Proteína

O problema do enovelamento de proteínas pode ser tratado como um problema de mecânica estatística, de maneira semelhante aos problemas de transições de fase em sistemas magnéticos estudados através do modelo de Ising. Considera-se uma cadeia polimérica imersa em um solvente a uma dada temperatura e as informações físicas da proteína serão obtidas através do cálculo de algumas quantidades termodinâmicas no equilíbrio térmico. Todas as

relações entre as variáveis termodinâmicas estão discutidas e apresentadas no apêndice A.1. A cinética de enovelamento também pode ser investigada considerando a evolução temporal do enovelamento em um caso fora do equilíbrio. Investiga-se a relaxação de quantidades como a energia ou número de contatos nativos submetendo o sistema a uma brusca variação de temperatura.

2.4.1 Modelo de rede

Uma proteína real é formada por uma longa cadeia de aminoácidos, interagindo entre si através de forças de curto alcance. Trata-se de um sistema com muitos graus de liberdade, que pode ser classificado como um sistema complexo e um problema de muitos corpos correlacionados. É um sistema com estruturas espaciais extremamente complexas. Uma investigação do enovelamento através de simulações, considerando todos os graus de liberdades de cada aminoácido da cadeia e todas as conformações espaciais da proteína, mesmo hoje, com vários recursos computacionais disponíveis, ainda é impraticável. No entanto, o progresso na investigação de enovelamento tem sido obtido com a adoção e estudo de modelos básicos e simplificados de proteínas. O modelo mais simplificado e amplamente adotado é o modelo de rede para a proteína. É baseado em duas aproximações. Primeiro, a estrutura molecular interna de aminoácidos é completamente desprezada e cada monômero é descrito como uma partícula. Uma segunda aproximação consiste em representar a proteína como sendo uma cadeia de esferas rígidas, representando os aminoácidos, unidas por hastes rígidas, representando as ligações covalentes (ligações peptídicas). Os aminoácidos são colocados nos sítios da rede cúbica de modo que o comprimento das hastes que ligam os aminoácidos seja igual ao parâmetro de célula da rede cúbica. A ocupação dos sítios é feita de tal forma que o comprimento das hastes seja conservado assim como o comprimento total da cadeia. Não deve haver mais do que um aminoácido por sítio (Fig. 2.4).

Apesar dessas aproximações, o caráter polimérico da proteína e a heterogeneidade das interações entre aminoácidos ainda são preservados e espera-se que a adoção de tais modelos para o estudo de proteínas seja plenamente adequada.

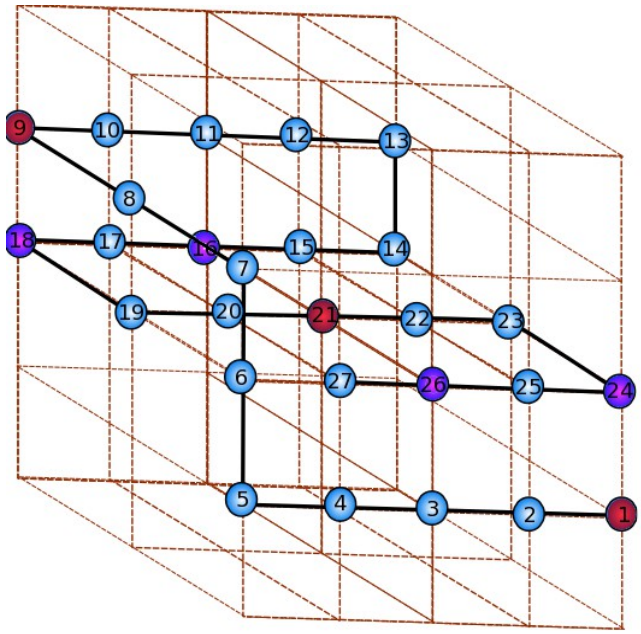


Fig. 2.4 – Representação de uma proteína segundo o modelo minimalista de rede. Os monômeros da cadeia (aminoácidos) são representados por esferas rígidas e suas ligações (ligação peptídica) por hastes rígidas. Cada sítio da rede deve ser ocupado apenas por um único monômero. A posição de cada monômero na sequência é indexada para facilitar seu reconhecimento. Em destaque, três monômeros ocupando posições variadas ao longo da cadeia (em vermelho) e seus respectivos vizinhos de rede que contribuem para formação de ligações locais (em roxo).

A energia total de uma proteína depende de sua conformação espacial, de sua composição e do solvente. As interações entre aminoácidos são de curto alcance. O modelo de rede procura preservar estas características das interações considerando que dois monômeros interagem entre si somente quando eles se encontram em sítios adjacentes na rede. Os monômeros conectados pelas ligações peptídicas também encontram-se em sítios adjacentes, mas neste caso as interações entre eles são desprezadas já que as conexões são consideradas rígidas e as distâncias entre esses monômeros nunca são alteradas (Fig. 2.4). A energia de uma conformação de uma proteína é dada por

$$E = \sum_{i>j}^{N} \varepsilon_{\alpha_i,\alpha_j} \Delta\left(r_i - r_j\right), \tag{2.8}$$

onde N é o número de monômeros (aminoácidos), o índice *i* indica a posição de um

monômero na sequência (não na rede). O parâmetro $\varepsilon_{\alpha_i,\alpha_j}$ representa a energia de interação entre os monômeros α_i no sítio *i* e α_j no sítio *j*. Já a função $\Delta(r)$ é introduzida para se representar convenientemente as energias de curto alcance e é definida como

$$\Delta(r_i - r_j) = \begin{cases} 1, & se\ |r_i - r_j| = a \\ 0, & caso\ contrário \end{cases}, \qquad (2.9)$$

onde a é o parâmetro de rede (tamanho de uma célula) e será considerado igual a uma unidade.

A energia de interação entre os pares de aminoácidos possíveis depende da aproximação adotada. Caso sejam considerados apenas dois tipos de monômeros onde um conjunto é formado por monômeros hidrofóbicos e os restantes polares , como é o caso do modelo HP, então $\varepsilon_{H,H} = -1.0$, $\varepsilon_{P,P} = \varepsilon_{H,P} = 0$ [2,3]. A interação com o solvente já está sendo considerada nestes parâmetros. Estas energias do potencial de contato estão dadas em unidades da constante universal RT_0 , que equivale a aproximadamente a 0,6 kcal/mol. Onde T_0 é dada na temperatura ambiente (T_0 = 300 K). Nos casos mais gerais em que são considerados 20 tipos de monômeros, as energias são dadas pela matriz de Miyazawa-Jernigan [23]. No modelo de rede, a estrutura nativa é a mais compacta e a de menor energia. Em uma cadeia com 27 monômeros, a estrutura é um cubo onde os 27 monômeros são dispostos nos sítios de uma rede 3x3x3.

Modelos desse tipo têm sido muito utilizados pois são capazes de reproduzir aspectos característicos do processo, como o tempo de enovelamento, identificar a conformação nativa, e descrever propriedades termodinâmicas com detalhes. Alguns dos mecanismos básicos que levam uma proteína a se enovelar foram compreendidos através de simulações utilizando modelos simplificados de proteínas. Finalmente, modelos de rede tem sido utilizados com grande sucesso para aplicações e teste de novos algoritmos. Pode-se citar Socci e Onuchic [24] que aplicaram a técnica de Ferrenberg-Sweendsen [25], Scheraga [26] utilizou o método Entropic Sampling [27], Wang [28] implementou o algoritmo flat-histogram [29] e Janke [30] implementou o método multicanônico para modelos de rede.

2.4.2 Algoritmo de Metropolis para modelos de rede

As propriedades termodinâmicas de uma cadeia polimérica podem ser determinadas se as conformações possíveis da cadeia forem conhecidas. As conformações da proteína podem ser obtidas através do algoritmo de Metropolis que será detalhado a seguir.

Para ilustração, será tomada uma cadeia genérica bidimensional. É importante deixar claro que este trabalho foi desenvolvido a partir de cadeias representadas em redes tridimensionais.

As grandezas vetoriais estão escritas em negrito. Considera-se uma conformação genérica da cadeia, como a apresentada na Fig. 2.5. As posições de cada monômero x_i e y_i são definidas pelo vetor posição $\boldsymbol{r}_i = x_i \hat{\boldsymbol{x}} + y_i \hat{\boldsymbol{y}}$, para i = 1,2,...,N. Uma nova conformação é gerada, a partir desta, modificando a posição dos monômeros desde que a cadeia conserve o seu comprimento e os sítios da rede não sejam duplamente ocupados. Embora um conjunto de três ou mais monômeros possam ser deslocados como um grupo ou sub-cadeia, há evidências concretas de que basta considerar a possibilidade de um grupo de no máximo 2 monômeros para gerar conformações estatísticas relevantes, isto é, que possuem probabilidades relevantes no cálculo termodinâmico.

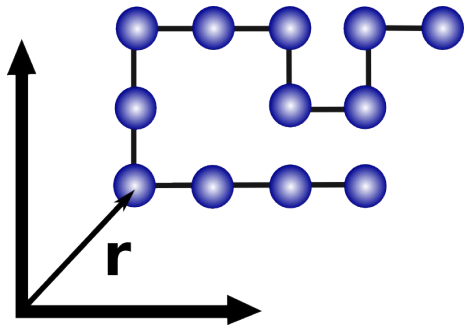


Fig 2.5 – Representação de uma cadeia a partir de um sistema de coordenadas bidimensional. Cada monômero na cadeia pode ser localizado na rede a partir de um vetor posição **(r)**. Para simplificar, a representação da rede bidimensional foi omitida do desenho.

Neste trabalho, adotou-se cadeia com 27 monômeros. As conformações são construídas utilizando uma combinação de movimentos de monômeros de fim da cadeia, ou movimento de ponta, movimento de monômeros em cantos da cadeia e movimentos de monômeros em segmentos de cadeia em forma de manivela (Fig. 2.6) obedecendo critérios tais como

- Monômeros situados nas extremidades da cadeia podem realizar movimentos onde se desloca ¼ de arco de circunferência de raio unitário e centrado na posição de seu vizinho adjacente.
- Monômeros situados em porções retilíneas não conseguem fazer qualquer movimento, pois o comprimento da cadeia deve ser conservado.
- Monômeros situados em porções da cadeia que formam vértice, podem realizar movimentos em que se desloca ao longo de uma diagonal que liga um canto ao canto oposto de um quadrado reticulado.
- 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^{o} . Nota-se que neste tipo de movimento, dois monômeros são deslocados simultaneamente.

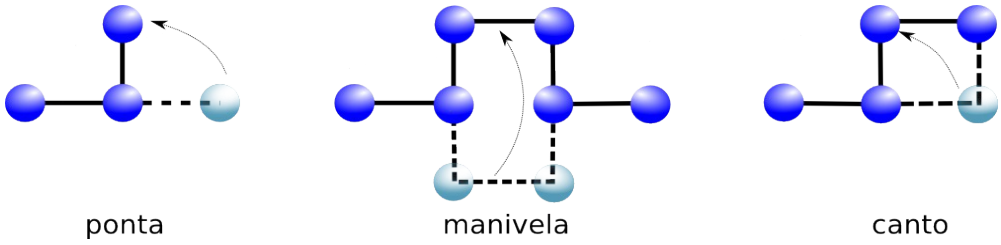


Fig. 2.6 – Possíveis movimentos para o estudo da mecânica estatística de cadeias heteropoliméricas em uma rede bidimensional.

Este conjunto de movimentos permite que a cadeia visite todas as conformações possíveis, o que garante a condição de volume excluído e a ergodicidade do sistema. Os movimentos serão realizados segundo um sorteio estabelecido pelo algoritmo de Metrópolis e eles dependem da temperatura do meio em que a cadeia se encontra assim como das energias das conformações apresentadas na Eq. (2.8).

Considerando-se que a cadeia esteja em equilíbrio térmico em um banho térmico a uma 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

$$p = min(1, \exp(-\beta \Delta E)), \qquad (2.10)$$

onde $\beta = 1/k_B T$.

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

1) Sorteio de um monômero. Verificar se o monômero está em uma manivela, um canto fora da manivela, ou extremo da cadeia.
2) Se forem das pontas, a identificação é trivial: ele tem coordenada r_1 ou r_N . Considere-se inicialmente o caso r_1. Para saber as possíveis novas posições do monômero, introduza-se o vetor unitário auxiliar $\hat{\boldsymbol{r}} = \boldsymbol{r}_2 - \boldsymbol{r}_1$. Caso $\hat{\boldsymbol{r}}$ seja $\hat{\boldsymbol{x}}$ ou - $\hat{\boldsymbol{x}}$, então o monômero pode ser deslocado para $\boldsymbol{r}_{1^{novo}} = \boldsymbol{r}_1 + \hat{\boldsymbol{r}} \pm \hat{\boldsymbol{y}}$. Caso $\hat{\boldsymbol{r}}$ seja $\hat{\boldsymbol{y}}$ ou - $\hat{\boldsymbol{y}}$, então $\boldsymbol{r}_{1^{novo}} = \boldsymbol{r}_1 + \hat{\boldsymbol{r}} \pm \hat{\boldsymbol{x}}$. Para o caso do monômero em r_N , basta considerar o mesmo procedimento. Define-se o vetor unitário $\hat{\boldsymbol{r}} = \boldsymbol{r}_N - \boldsymbol{r}_{N-1}$. Caso $\hat{\boldsymbol{r}}$ seja $\hat{\boldsymbol{x}}$ ou - $\hat{\boldsymbol{x}}$, a nova posição do monômero é dada por $\boldsymbol{r}_{N^{novo}} = \boldsymbol{r}_1 \pm \hat{\boldsymbol{y}}$. Caso $\hat{\boldsymbol{r}}$ seja $\hat{\boldsymbol{y}}$ ou - $\hat{\boldsymbol{y}}$, então $\boldsymbol{r}_{N^{novo}} = \boldsymbol{r}_{N-1} \pm \hat{\boldsymbol{x}}$.

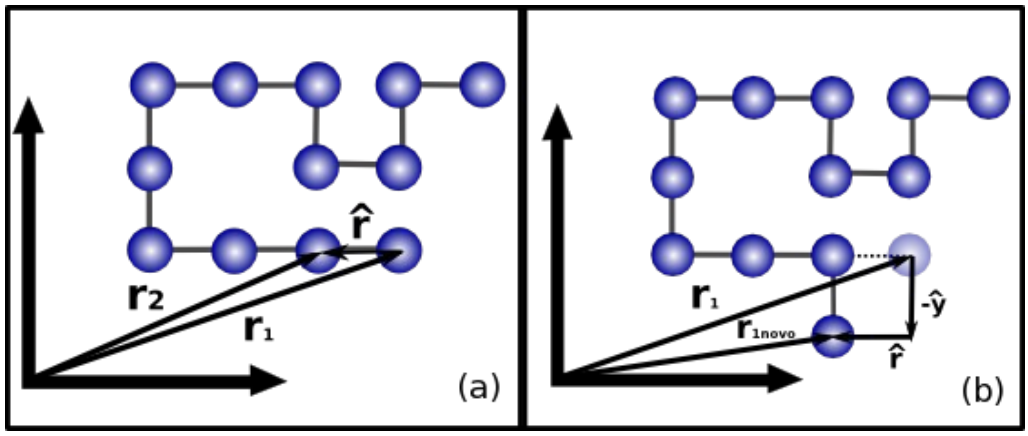


Fig. 2.7 – Movimento de ponta. a) Cálculo do versor auxiliar $\hat{\boldsymbol{r}}$. b) Cálculo da nova posição do monômero 1, após a execução do movimento.

3) 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

$$(\boldsymbol{r}_i - \boldsymbol{r}_{i-1}).(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i) = 0, \tag{2.11}$$

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

$$\boldsymbol{r}_{i^{novo}} = \boldsymbol{r}_i + \Delta \boldsymbol{r}_i, \tag{2.12}$$

onde

$$\Delta \boldsymbol{r}_i = (\boldsymbol{r}_{i+1} - \boldsymbol{r}_i) - (\boldsymbol{r}_i - \boldsymbol{r}_{i-1}) = \boldsymbol{r}_{i+1} - 2\mathbf{r}_i + \boldsymbol{r}_{i-1}. \tag{2.13}$$

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

$$\boldsymbol{r}_{i^{novo}} = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i + \mathbf{r}_{i-1}. \tag{2.14}$$

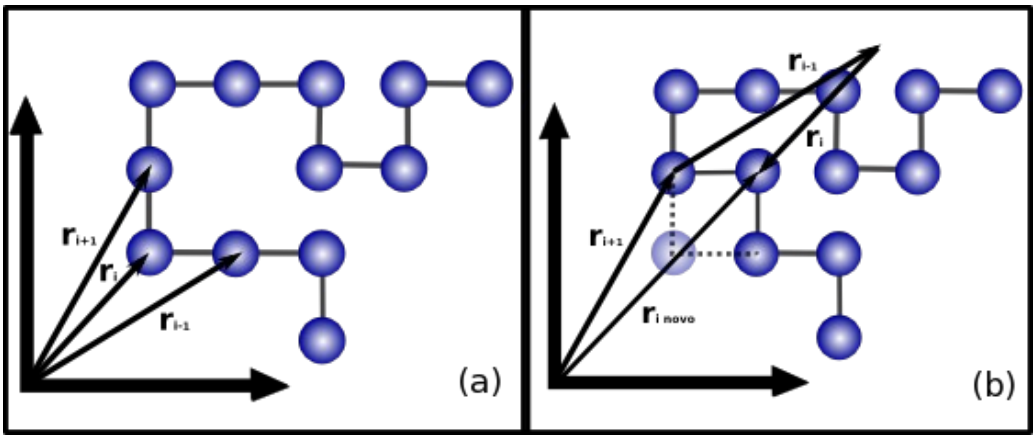


Fig. 2.8 – Movimento de canto. a) Identicação de um possível movimento de canto para o monômero i. Observa-se a participação dos monômeros adjacentes i+1 e i-1. b) Movimento de canto efetuado, cálculo vetorial para a nova posição do monômero i na rede bidimensional.

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 feito no item anterior. Logo após, é preciso verificar também se a posição especificada pelo vetor

$$\boldsymbol{s}=\boldsymbol{r}_{i+1}-\boldsymbol{r}_i+\boldsymbol{r}_{i+1}, \tag{2.15}$$

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 ocorrer nenhum desses casos, então o monômero é um monômero de canto apenas. Entretanto se a posição identificada por $\boldsymbol{s}$ 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 $\boldsymbol{s}$, 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 em uma rotação de 180 em torno de um eixo imaginário que passa pelos sítios localizados em $\boldsymbol{r}_{i-1}$ e $\boldsymbol{r}_{i+2}$ (ou $\boldsymbol{r}_{i-2}$ e $\boldsymbol{r}_{i+1}$). Nesse caso, as novas posições do par i e i+1 serão

$$\boldsymbol{r}_{i^{novo}}=\boldsymbol{r}_{i+2}-\boldsymbol{r}_{i+1}+\boldsymbol{r}_{i-1}, \tag{2.16}$$

e

$$\boldsymbol{r}_{i^{novo}}=2\boldsymbol{r}_{i+2}-\boldsymbol{r}_{i+1}. \tag{2.17}$$

Caso o par de monômeros seja i-1 e i, suas novas posições serão

$$\boldsymbol{r}_{i-1^{novo}} = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i + \boldsymbol{r}_{i-2}, \tag{2.18}$$

e

$$\boldsymbol{r}_{inovo} = 2\boldsymbol{r}_{i+1} - \boldsymbol{r}_i. \tag{2.19}$$

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

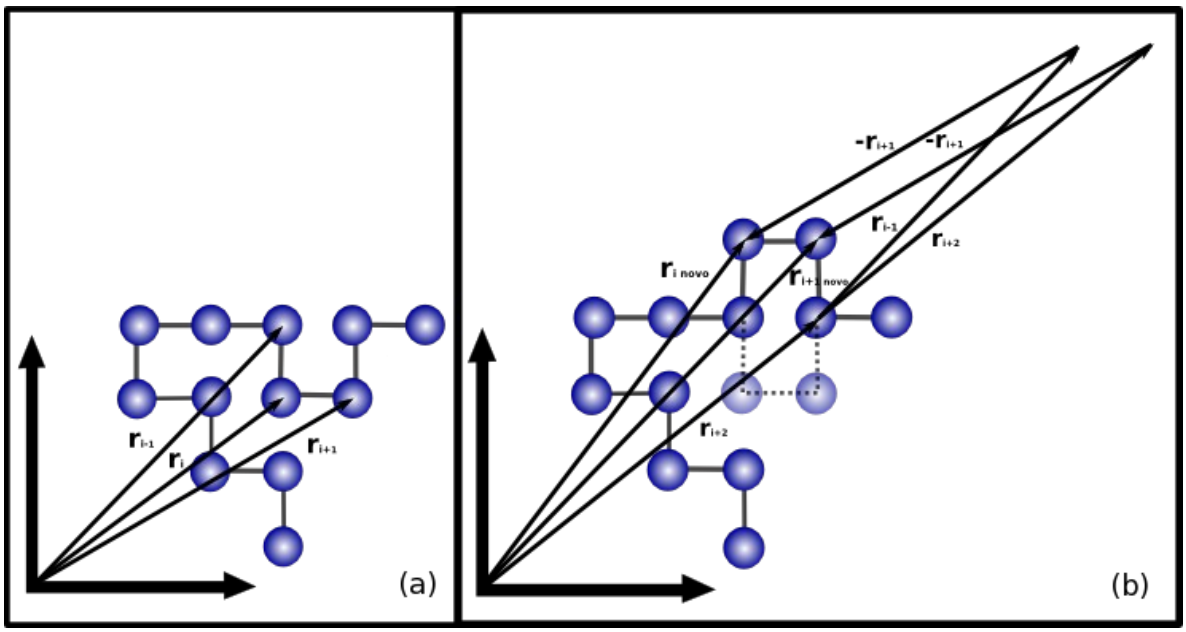


Fig. 2.9 – Movimento de manivela. a) Identificação de um possível movimento de manivela para o monômero i. Este tipo de movimento envolve a participação de dois monômeros, no exemplo, i e i+1. b) Cálculo vetorial das novas posições dos monômeros i e i+1, depois de efetuado o movimento.

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.

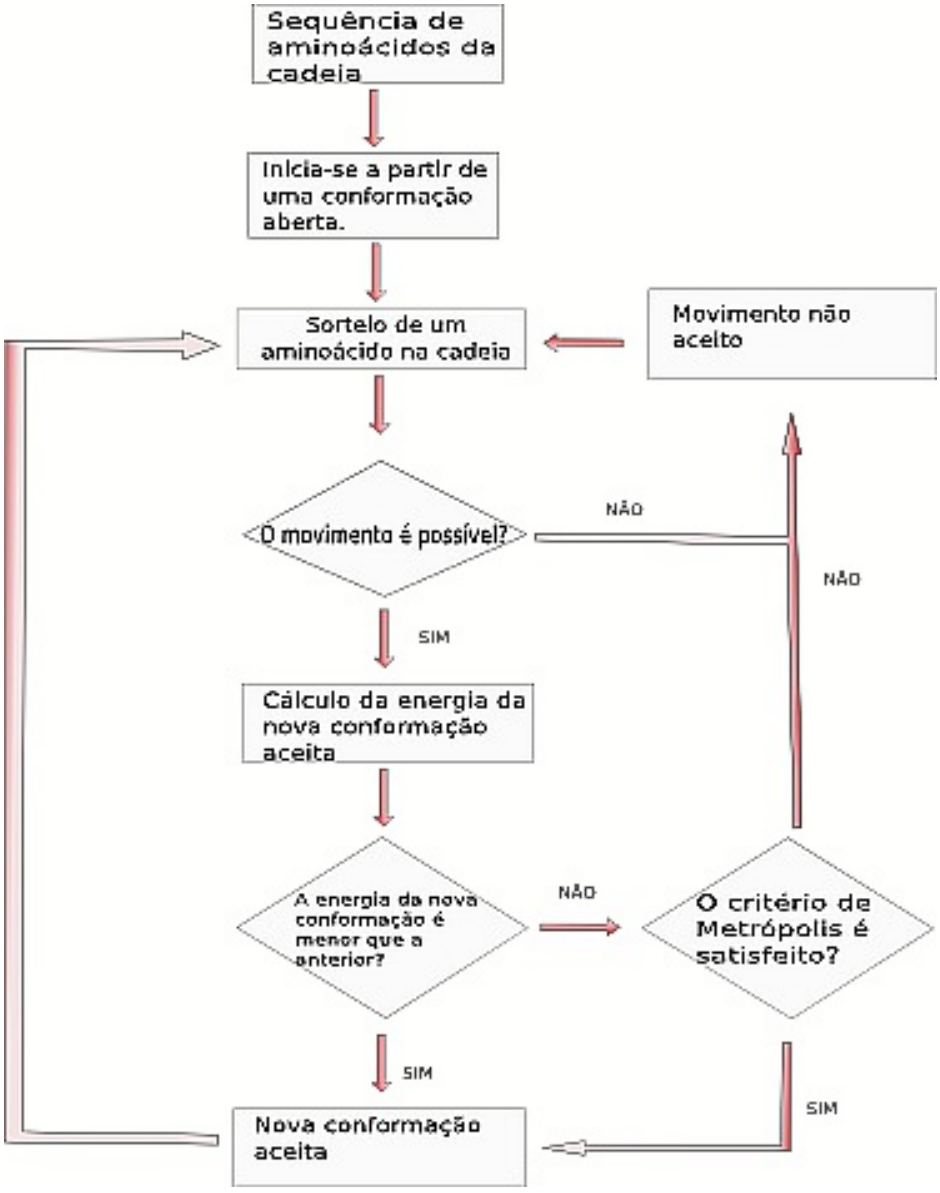


Fig. 2.10 – Fluxograma do método de metrópolis para o problema do enovelamento de proteínas.

2.5 Cinética do enovelamento

Em 1968, Levinthal [18] apontou problemas na hipótese termodinâmica para explicar o enovelamento. Por esta hipótese, a estrutura nativa é adotada pela cadeia por ser a estrutura mais estável sob ponto de vista termodinâmico. A interpretação de Levinthal sugere que, devido ao número muito grande de graus de liberdade num polipeptídeo, a molécula pode passar por um número astronômico de possíveis conformações até atingir a estrutura nativa. Se a proteína deve buscar sua conformação nativa única, a expectativa é de que o faça em um tempo muito grande devido à complexidade das conformações. Entretanto, muitas proteínas pequenas se enovelam espontaneamente num escala de milisegundos ou microsegundos sugerindo que deve haver um mecanismo muito especial que guia as cadeias ao longo de sua trajetória no espaço das conformações para que a proteína atinja a estrutura nativa.

Este paradoxo vem sendo investigado há décadas. Chan e Dill [31-33] explorando o modelo HP (apenas monômeros hidrofóbicos e polares) em uma rede cúbica, estudaram aspectos dinâmicos de cadeias curtas durante o processo de enovelamento. Apesar do modelo HP ser considerado muito simples, foram obtidas informações interessantes a respeito dos mecanismos cinéticos de enovelamento. Este modelo apresentou um processo similar ao encontrado em proteínas reais. Há um rápido colapso para conformações compactas, seguido de uma reconfiguração lenta dessas cadeias até a estrutura nativa. Camacho e Thirumalai [34] estudaram sistemas em redes bidimensionais a partir de diferentes potenciais de interação. Eles encontraram duas temperaturas de transição; uma temperatura de colapso na qual a cadeia forma uma estrutura compacta e uma temperatura de enovelamento na qual a estrutura nativa é formada. Estudos com modelos de heteropolímeros têm mostrado que sequências que se enovelam em um intervalo de tempo curto, como esperado, apresentam um “gap” de energia significativo entre o estado nativo e a energia das demais conformações. Onuchic e Socci [35] também estudaram esse problema do enovelamento observando seus aspectos dinâmicos, mas desta vez, a partir do potencial de interação do modelo AB com dois tipos de monômeros, onde monômeros iguais se atraem e monômeros diferentes se repelem. Os resultados encontrados são similares.

As simulações da cinética do enovelamento são baseadas no algoritmo de Metropolis e

no modelo de rede. Inicialmente, uma sequência qualquer numa cadeia com uma conformação cuja energia é 0, ou seja, a cadeia aberta, é escolhida como a conformação inicial. Permite-se que a cadeia passe a se enovelar, seguindo o critério de aceitação do método de Metropolis até que esta atinja a energia e a conformação do estado nativo. O tempo de todo o processo (número de passos de simulação) é registrado e uma nova conformação com energia 0, diferente da anterior é escolhida como conformação inicial, o número de passos de Monte Carlo é zerado e uma nova simulação é iniciada. Quando todas as conformações são avaliadas, muda-se a temperatura do sistema e calcula-se o tempo médio de enovelamento. A simulação é interrompida caso o número de passos de Monte Carlo exceda a tolerância desejada.

A partir desse método do estudo dinâmico da cadeia, busca-se determinar o tempo de enovelamento apresentado pelas cadeias. Este tempo calculado numa determinada temperatura é dado pela média dos tempos de enovelamentos de uma cadeia polimérica, ou seja:

$$t_{folding} = \frac{1}{M} \sum_{i=1}^{M} t_i , \qquad (2.20)$$

onde M é o número de vezes que a cadeia se enovelou para o estado nativo, t_i é o tempo que a cadeia levou para atingir o estado nativo na i-ésima tentativa.

Outras grandezas podem ser calculadas em função do tempo tal como serão apresentadas neste trabalho.

2.6 Sequenciamento de aminoácidos via método de Monte Carlo

O sequenciamento de aminoácidos, também conhecido como "*design*" de proteínas, é outra formulação do problema do enovelamento de proteínas. Enquanto no "*folding*" ou enovelamento direto, o objetivo é a busca pelo estado nativo de uma sequência previamente escolhida, o design visa encontrar as sequências que se enovelam para uma dada estrutura conhecida, a estrutura alvo. Dessa forma, o "*design*" constitui um método para sequênciamento de aminoácidos numa estrutura nativa pré-definida [4,5].

O objetivo do “design” pode ser alcançado se a energia da estrutura alvo for reduzida ao máximo com a seleção adequada de sequências de aminoácidos. As energias das demais conformações não são alteradas significativamente com a troca de sequências. Por exemplo, em uma conformação não compacta da cadeia, pode-se permutar a ordem dos aminoácidos na sequência sem que a energia seja significativamente alterada. Isso tornará o espectro de energia mais largo e maximizará o “*gap*” de energia existente entre o estado nativo e o conjunto de estados não nativos [5,6,7]. Energias de conformações aleatórias de proteínas correspondem então à parte contínua do espectro dependendo somente da composição de aminoácidos (e não das sequências). Por outro lado, conformações de baixa energia têm um melhor ajuste dos contatos e sua energia depende somente da sequência de aminoácidos.

Em 1993, Shakhnovich [4,5] propôs um método estocástico para desenhar sequências baseado no algoritmo de Metropolis. Para isso, ele estabeleceu uma analogia entre sistemas magnéticos descritos pelo modelo de Ising e cadeias poliméricas contendo dois tipos de monômeros. Em uma rede de spins, a uma dada temperatura *T*, uma fração n_1 de spins são “up” e n_2 são “down” satisfazendo $n_1+n_2=N$ onde N é o número total de spins.

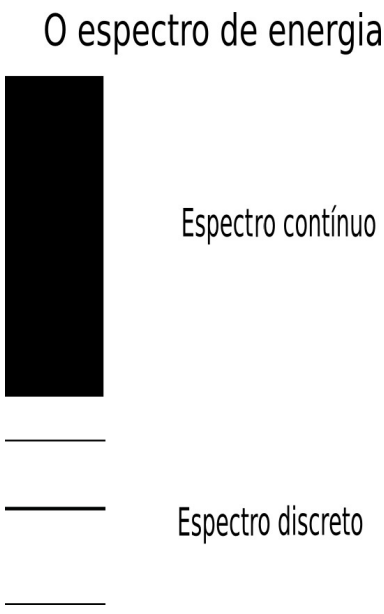


Fig. 2.11-Energias de conformações aleatórias de proteínas correspondem então à parte contínua do espectro de energia dependendo somente da composição de aminoácidos (e não das sequências). Por outro lado, conformações de baixa energia têm um melhor ajuste dos contatos e sua energia depende somente da sequência de aminoácidos. Essas conformações de baixa energia se caracterizam pelo espectro discreto.

Dentre estas configurações, considere aquelas em que n_1 e n_2 sejam fixos; elas totalizam $N!/(n_1!n_2!)$ configurações. Em uma cadeia polimérica contendo dois tipos de monômeros, n_1 polares e n_2 não-polares, distribuídos em uma estrutura previamente conhecida, também há $N!/(n_1!n_2!)$ maneiras de distribuí-los na cadeia de N sítios. Na verdade, são todas as sequências possíveis de serem construídas. Isto ocorre também se n_1 e n_2 puderem variar, mantendo N constante. No caso da cadeia de monômeros, isso significaria modificar a composição de monômeros da cadeia. Shakhnovich observou que tanto o sistema magnético quanto os monômeros na cadeia obedecem a mesma estatística e propôs um método para determinar sequências baseado no algoritmo de Metropolis a exemplo de métodos aplicados para sistemas magnéticos. A mecânica estatística de sequências é análoga à mecânica estatística para configurações de spin do modelo Ising e ambos seguem a estatística de Boltzmann [36]. Dessa forma, para o caso do método de determinar sequências, é necessário selecionar uma temperatura, a temperatura evolucionária, para o processo de optimização de Monte Carlo.

Considere o exemplo do modelo de proteínas composto por dois tipos de monômeros, os polares e apolares. Os monômeros apolares se atraem e os polares não interagem entre si e também não interagem com os monômeros apolares. As sequências podem ser representadas na forma $\{\sigma_i\}$ com $\sigma_i = 1$ se o grupo i é apolar e $\sigma_i = 0$ se for polar, similares às configurações de spins do modelo de Ising representadas por $\{\sigma_i\}$ onde $\sigma_i = \pm 1$. A estrutura alvo pode ser definida pelo conjunto $\{ r_{i^0} \}$ das coordenadas de seus resíduos. A energia da cadeia formada por uma sequência disposta na estrutura alvo é dada por

$$E_0(\{\sigma_i\}) = \frac{1}{2}\sum_{ij}^{N} U_0 \sigma_i \sigma_j \Delta(r_{i^0} - r_{j^0}) \quad , \qquad (2.21)$$

onde N é o número total de resíduos, $U_0 < 0$ é o parâmetro de interação entre monômeros apolares, Δ define o alcance do potencial de contatos entre resíduos ou seja, $\Delta(r) = 1$ se dois monômeros forem vizinhos na rede mas não vizinhos ao longo da cadeia. Se o sistema for tratado através de ensemble canônico, a probabilidade de se encontrar uma dada sequência $\{\sigma_i^0\}$ é escrita como

$$P\left\{\sigma_{i^0}\right\}=\frac{1}{\tilde{Z}}e^{\frac{-E_0\left(\left\{\sigma_{i^0}\right\}\right)}{T_{evol}}}\delta\left(\sum_i^N \sigma_{i^0}-N_\alpha\right) \quad , \qquad (2.22)$$

onde

$$\tilde{Z}=\sum_{ij} e^{\frac{-E_0\left(\left\{\sigma_{i^0}\right\}\right)}{T_{evol}}}\delta\left(\sum_i^N \sigma_{i^0}-N_\alpha\right) \quad , \qquad (2.23)$$

é a função de partição, T_{evol} é a temperatura e N_α é o número total de resíduos apolares. As funções-δ nas Eqs. (2.22) e (2.23) permitem apenas sequências com uma composição fixa de aminoácidos.

A tabela 2.2 ilustra a correspondência entre os modelos de spin e proteínas.

Modelo de Ising	**Design de proteínas**
Temperatura	Temperatura evolucionária
Direção do spin (up or down)	Polaridade do resíduo
Momento magnético total	Hidrofobicidade da proteína
Energia	Energia
Entropia numa dada energia	Log (número de sequências tendo uma dada energia)
Magnetização espontânea	Separação do núcleo hidrofóbico e da superfície polar

Tabela 2.2 – Analogia entre modelo de Ising e o "*design*" de proteínas. Observa-se a correspondência das grandezas apresentadas por ambos modelos.

Algumas considerações devem ser mencionadas. O número de monômeros apolares da composição da cadeia no modelo utilizado para "*design*" é uma medida da hidrofobicidade e, corresponde à magnetização no modelo de Ising, com a restrição de que ela permaneça conservada.

A simulação computacional do enovelamento de uma cadeia polimérica é um processo complexo devido à frustração geométrica. As barreiras de energia tornam este processo lento [37,38]. Ao contrário do processo de enovelamento, não há presença de frustração no

processo de "*design*" de modo que desenhar sequências deve ser um processo mais rápido em uma simulação.

Outra importante diferença entre o enovelamento de proteínas e o "*design*" é que no problema do enovelamento, ocorre a busca por uma estrutura única e estável em temperaturas fisiológicas. Entretanto, o problema do design de proteínas pode não ter uma única solução - muitas sequências podem acabar ajustadas para a estrutura alvo e energia. Isso garante uma superfície de energia-livre suave enquanto que no caso do enovelamento direto as superfícies geralmente são rugosas.

Finalmente, modelos de ferromagnetismo a baixas temperaturas exibem uma transição de fase. No espaço de sequências isso significa que a baixas T_{evol} haverá separação entre resíduos polares e apolares formando uma conformação de cadeia com caroço hidrofóbico e uma superfície hidrofílica.

O método de Shakhnovich para desenhar sequências é baseado no algoritmo de Metropolis. O processo tem início com a escolha aleatória de uma sequência . Colocando esta sequência na estrutura alvo, obtemos a energia correspondente. Uma nova sequência é gerada permutando dois monômeros da sequência tomados aleatoriamente. Caso a energia da nova sequência seja menor que a da sequência anterior, aceita-se a nova sequência. No entanto, se a variação de energia for positiva, deve-se submeter a aceitação da nova sequência a um sorteio.

O método de "design" proposto neste trabalho baseia-se nas idéias descritas acima, entretanto, utilizando-se do método de Wang-Landau. Esta não é a unica forma de realizar o sequenciamento de aminoácidos. No apêndice A.2, encontra-se um método alternativo para o "design" de proteínas, realizado a partir da teoria de campos médios.

Capítulo III – SELEÇÃO DE SEQUÊNCIAS DE AMINOÁCIDOS DE PROTEÍNAS ATRAVÉS DO MÉTODO DE WANG-LANDAU

Neste capítulo, será apresentado um novo método para seleção de sequências de aminoácidos de uma proteína, cuja conformação nativa e composição sejam previamente conhecidas. É baseado no método de Wang-Landau, introduzido para estudar transições de fase em sistemas magnéticos. Inicialmente, abrindo o capítulo, este método será apresentado de forma detalhada. A seguir, será apresentada a adaptação do método para seleção ("*design*") de sequências de aminoácidos em modelo de rede. Finalizando, também será apresentado a adaptação do método de Wang-Landau para o cálculo de grandezas termodinâmicas de proteínas utilizando o modelo de rede. O cálculo destas grandezas auxiliará na caracterização das proteínas desenhadas. Dirá se as sequências desenhadas dão origem a cadeias poliméricas que se comportam como proteínas, que se caracterizam pela rápida compactação e com estabilidade termodinâmica.

3.1 O método de Wang-Landau

O algoritmo de Metropolis talvez seja o algoritmo mais utilizado em métodos de Monte Carlo para simulações de sistemas em equilíbrio térmico. Entretanto, quantidades

como entropia não podem ser calculados de maneira direta e não são resultados imediatos neste tipo de simulação. Sistemas cuja superfície de energia-livre apresenta “rugosidades”, impõem dificuldades para a aplicação do algoritmo pois o sistema pode ser capturado por algum vale da superfície por um longo período de tempo e inutilizar a simulação. Apesar desse algoritmo possibilitar a geração de uma distribuição canônica $g(E)e^{-E/kT}$ em dada temperatura *T*, a distribuição é tão estreita que múltiplas iterações são necessárias para o cálculo de quantidades termodinâmicas de interesse em uma faixa de temperatura significativa [11]. Na distribuição canônica a densidade de estados independe da temperatura. Se a densidade de estados puder ser calculada com precisão, a função de partição ($Z = \Sigma g(E)e^{-E/kT}$) do sistema pode ser determinada, e assim o problema estaria resolvido, pois as demais quantidades podem ser derivadas dela (apêndice A.2 para uma revisão de Mecânica Estatística).

Recentemente, Wang e Landau [11,39,40] propuseram um método de simulação Monte Carlo para sistemas clássicos que se utiliza do passeio aleatório no espaço das energias para a obtenção da densidade de estados do sistema *g(E)*. Este método é baseado na observação do comportamento do histograma de visitas *H(E)* durante o passeio aleatório no espaço de energia. Quando o passeio é realizado com probabilidade de transição para um estado de energia *E* proporcional a *1/g(E)*, verifica-se que o histograma produzido é plano. Ao final, todos os estados são igualmente visitados.

A função de partição contém todas as informações essenciais de um sistema clássico [41,42,43] e é dada por

$$Z=\sum_{E} g(E)e^{\frac{-E}{k_B T}}, \qquad (3.1)$$

onde *g(E)* é a densidade de estados acessíveis ao sistema, e cuja relação com a entropia é

$$S(E)=k_B \ln\left[g(E)\right]. \qquad (3.2)$$

Para um sistema em equilíbrio termodinâmico, a probabilidade de ocorrência de um estado com energia *E* a temperatura *T*, pode ser escrita na forma

$$P(E,T)=\frac{1}{Z}g(E)e^{\frac{-E}{k_B T}}, \tag{3.3}$$

e por conveniência, doravante será considerado k_B=1. Tomando a equação (3.1) em termos da entropia pode-se reescrevê-la como

$$Z=\sum_E e^{S(E)-\frac{E}{T}}. \tag{3.4}$$

Com base nestas considerações e no princípio do balanço detalhado, pode-se introduzir o algoritmo sugerido por Wang e Landau (WL). Para isso, pode-se escolher convenientemente a taxa de transição $W_{i,j}$, a probabilidade por unidade de tempo para que um estado de energia E_i passe para um estado de energia E_j , tal como feito na introdução do algoritmo de Metropolis. Entretanto, como o passeio aleatório se dá no espaço das energias, pode-se propor que a taxa de transição seja dada por

$$W_{ij}=\begin{cases} e^{-\Delta S}, & \Delta S>0 \\ 1, & \Delta S\leq 0 \end{cases}, \tag{3.5}$$

onde a variação de entropia é definida como

$$\Delta S=S(E_j)-S(E_i). \tag{3.6}$$

Deste modo, a probabilidade de transição passa a ser dada por

$$p(i\to j)=\frac{W_{ij}}{W_{ji}}. \tag{3.7}$$

Para $\Delta S>0$, a probabilidade de transição será então

$$p(i \to j) = \frac{e^{-\Delta S}}{1} = e^{-S(E_j)+S(E_i)} = \frac{e^{S(E_i)}}{e^{S(E_j)}}, \tag{3.8}$$

e consequentemente

$$p(i \to j) = \frac{g(E_i)}{g(E_j)}. \tag{3.9}$$

Já para o caso $\Delta S \leq 0$

$$p(i \to j) = \frac{1}{e^{-\Delta S}} = \frac{1}{e^{-S(E_j)+S(E_i)}} = \frac{e^{S(E_j)}}{e^{S(E_i)}}. \tag{3.10}$$

Neste caso, tem-se $g(E_j) \leq g(E_i)$ o que corresponde a uma probabilidade maior ou igual a 1 e significa que a transição com certeza será efetuada. Concluindo, o passeio aleatório no espaço de energia é efetuada com probabilidade de transição

$$p(i \to j) = min\left[\frac{g(E_i)}{g(E_j)}, 1\right]. \tag{3.11}$$

No início de uma simulação do método WL, a densidade de estados é desconhecida e é ajustada como $g(E) = 1$ para todos os valores de energia. Ao decorrer da simulação, a probabilidade de transição para um estado de energia *E* é dada por *1/g(E)*. Cada vez que um estado de energia *E* é visitado, modifica-se a densidade de estados por um fator *f>1*; ou seja

$$g(E) = g(E).f, \tag{3.12}$$

e adiciona-se a *H(E)* uma visita a mais, isto é, $H(E) = H(E) + 1$. Se o sistema se mantém no mesmo estado de energia, a densidade de estados *g(E)* correspondente é modificada da

mesma forma assim como *H(E)*. O valor inicial de f é $f_0 = e \approx 2.71828$. Acumula-se os valores de visitas no histograma *H(E)* até que este se torne plano, o fator de modificação seja reduzido para um valor mais refinado através de uma função do tipo $f_{i+1}=\sqrt{f_i}$. Então o histograma H(E) é zerado e um novo passeio inicia-se. Na verdade qualquer função que reduza o fator de modificação monotonicamente pode ser utilizada; em geral

$$f_{i+1}=f_i^{1/n}, \tag{3.13}$$

satisfaz essa condição. A simulação continua até que f ≈ 1, geralmente utiliza-se $f \approx e^{10^{-8}}$. A precisão da densidade de estados não depende somente do fator de modificação, mas também de outros parâmetros, como tamanho e complexidade do sistema, critério adotado para histograma plano e detalhes de implementação. Como é impossível obter um histograma totalmente plano, tolera-se variações em H(E) da ordem de 5 a 15%, dependendo do sistema estudado.

3.2. O desenho de proteínas com o método de Wang-Landau

A aplicação do método de Wang-Landau para selecionar sequências de aminoácidos a partir de uma composição fixa, é uma adaptação do método de Shakhnovich previamente discutida. Como será visto, este método tem a vantagem de não depender da escolha de uma temperatura evolucionária para gerar as sequências de interesse. Este método permite gerar diretamente as sequências , classificadas em ordem crescente de energia. Tal como no método de Shakhnovich, é necessário inicialmente escolher o esqueleto da estrutura alvo. Pode ser, por exemplo, uma das várias catalogadas no Protein Data Bank [16]. Além da estrutura alvo, é necessário escolher os aminoácidos da composição da proteína que se quer desenhar, além da energia potencial de interação entre aminoácidos ou potencial de contacto. Neste trabalho, serão considerados proteínas de modelos de rede, contendo 27 monômeros, onde as interações de contacto serão dadas pela matriz de Miyazawa-Jernigan [23]. Na composição da proteína será considerada a presença dos 20 tipos de aminoácidos naturais. Deve-se observar que neste processo de desenhar sequências, tanto a composição da proteína quanto a estrutura alvo, tida como nativa, não se modifica durante a simulação, apenas a ordem na qual os aminoácidos aparecem na cadeia.

O objetivo da simulação consiste em determinar a densidade

$$\rho = \sum_{Seq} \delta(H_{Seq} - E), \tag{3.14}$$

onde H_{seq} são as energias de todas as sequências possíveis. O cálculo desta densidade é iniciado com a escolha da composição e do esqueleto da estrutura alvo. Os 27 aminoácidos da composição são dispostos aleatoriamente nos sítios do esqueleto e considerando a energia potencial de interação entre os aminoácidos, a energia da cadeia é calculada. As sequências são geradas permutando dois aminoácidos da cadeia escolhidos aleatoriamente. Contudo esta permutação deve ser feita segundo a probabilidade de transição

$$p(i \to j) = min\left[\frac{\rho(E_i)}{\rho(E_j)}, 1\right], \tag{3.15}$$

onde E_i e E_j são as energias da sequência *i* e *j*, e $\rho(E)$ a densidade de estados dada na Eq. (3.14). Para cada sequência visitada, a densidade é refinada através de $\rho(E) = \rho(E).f$ e o histograma de visitas representado por *H(E)* deve ser corrigido para *H(E) = H(E) + 1*. Esse processo de permutação é repetido um grande número de vezes até que $\rho(E) \, tende \, a \, 1$.

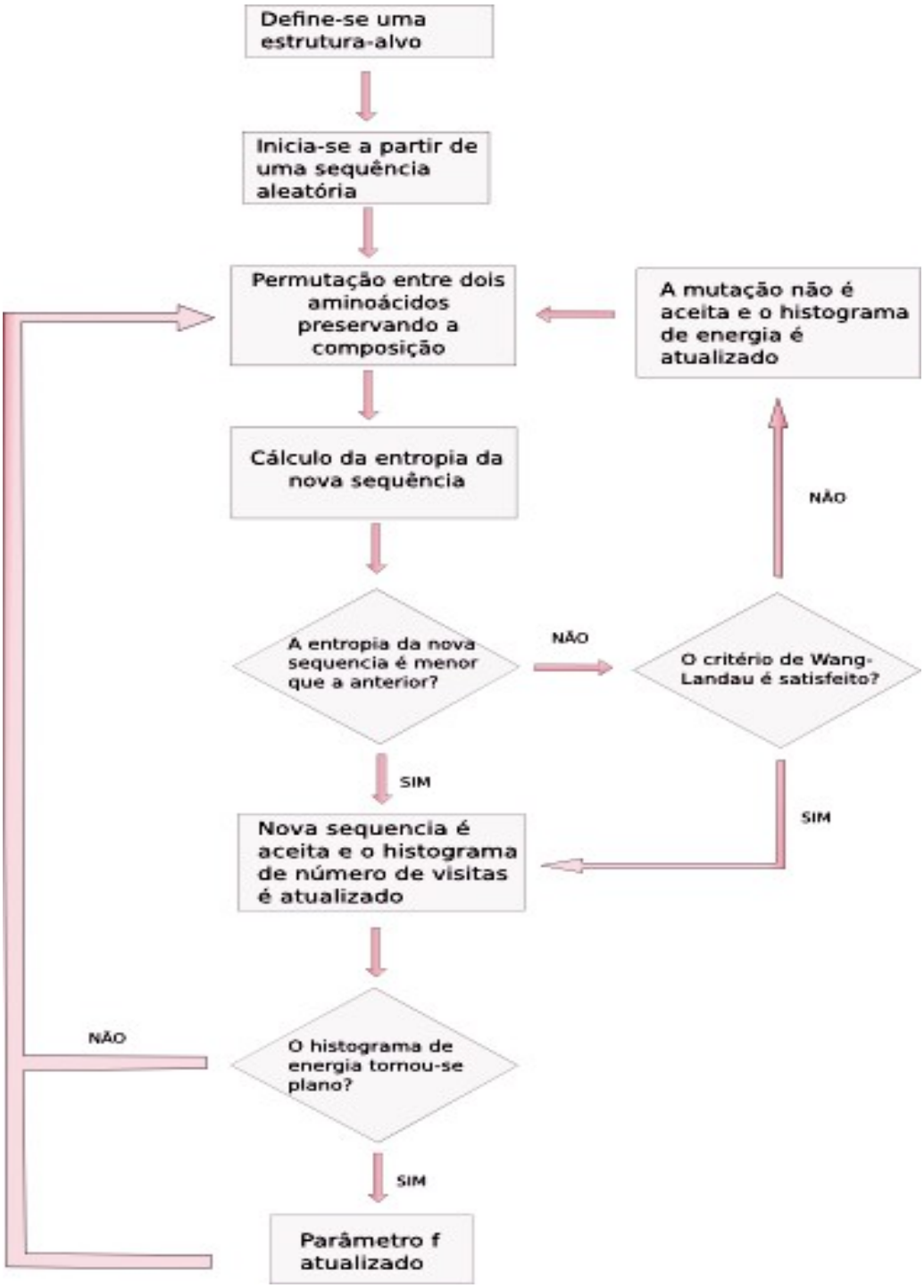


Fig. 3.1 – Fluxograma do método de sequênciamento de proteínas a partir de Wang-Landau.

Ao final desse processo todas as sequências possíveis estão armazenadas e classificadas segundo uma ordem crescente de energia. Espera-se que as sequências de menor energia sejam boas candidatas para constituir uma cadeia de aminoácidos com propriedades de uma proteína. Porém, somente uma caracterização sistemática dirá se a cadeia obtida é uma proteína. O teste natural é abrir a cadeia a uma temperatura escolhida convenientemente e verificar se ela se enovela para a estrutura nativa. Esta estrutura nativa deverá ser a estrutura previamente adotada como estrutura alvo. Se isto ocorrer, o desenho da proteína foi bem sucedido. No entanto, pode ocorrer que a estrutura de menor energia não seja a estrutura alvo. Neste caso, a sequência deve ser eliminada.

3.3. Similaridade com a estrutura-alvo

Um método de avaliar o sucesso da seleção de sequências, consiste em determinar a similaridade da conformação nativa de uma cadeia desenhada com a conformação da estrutura-alvo. Se ambas coincidirem, a sequência desenhada foi bem sucedida. A similaridade é avaliada através de um parâmetro definido como parâmetro de similaridade. Ele é calculado como produto de duas matrizes denominadas matrizes de contato. A matriz tem como objetivo indicar quais aminoácidos da cadeia em modelo de rede estão em contato entre si ou são vizinhos próximos, excetuando os aminoácidos conectados pela ligação peptídica. Estes contatos podem ser colocados em uma matriz booleana de adjacências que indica se um dado monômero i, encontra-se em contato com outro monômero j. Se i está em contato com j, o elemento de matriz correspondente é $S_{i,j}=1.$ Se não estiverem em contacto, recebem o valor zero.

Seja S_0, a matriz de adjacências correspondente à estrutura-alvo, e S a matriz de adjacências calculada para o estado nativo da sequência desenhada. A similaridade entre as conformações pode ser calculada como

$$\sigma=\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} S_{0,i,j}S_{i,j}}{N_q}, \tag{3.16}$$

onde N_q é o número de contatos máximos possíveis para a estrutura-alvo. No caso das

estruturas-alvo utilizadas, o número de máximo de contatos é 28.

Se duas conformações são similares, então $\sigma = 1$.

3.4 – Cálculo de quantidades termodinâmicas

Para análise das sequências desenhadas, algumas quantidades termodinâmicas são fundamentais [44,45,46]. Estas grandezas podem ser calculadas através das simulações de proteínas em equilíbrio térmico, utilizando também o método de Wang-Landau. A implementação do método de Wang-Landau para o o cálculo de propriedades termodinâmicas não é o objetivo desta dissertação. Mas sua implementação será muito útil para análise das proteínas desenhadas pelo método proposto. A implementação do método de Wang-Landau pode ser feita repetindo todo o procedimento adotado para simular uma proteína em banho térmico através do algoritmo de Metropolis e apresentado no capítulo II. A diferença é que agora o sorteio para fazer a mudança de conformação da proteína depende da variação da entropia. A variação da entropia

$$\Delta S = S(E_j) - S(E_i), \tag{3.17}$$

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

$$S(E) = S(E) + \ln f, \tag{3.18}$$

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

$$H(E) = H(E) + 1, \tag{3.19}$$

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, em dado momento ocorrerá a convergência do histograma para um valor constante, isto é

$$H(E) \rightarrow \textit{constante para todo } E\,, \tag{3.20}$$

significando que obteve-se um histograma plano, ou seja, todas as energia E possíveis para a cadeia foram igualmente visitadas. Quando isso ocorrer, é atribuído um novo valor para f, por exemplo

$$f \rightarrow \sqrt{f}\,, \tag{3.21}$$

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

$$\ln f = 10^{-8}\,, \tag{3.22}$$

e a correção da entropia seja desprezível. O valor da entropia é usado para o cálculo da densidade de estados acessíveis ao sistema.

É importante observar que a medida que os movimentos dos monômeros são executados, a cadeia se move no espaço. Por isso, deve-se tratar esse problema aplicando uma das seguintes condições de contorno periódica. Toda vez, que a cadeia atingir a borda da rede (observa-se que no caso do modelo computacional, o espaço é discretizado e finito), deve-se transladar toda a cadeia para o centro da rede, ou se um monômero tende a sair do espaço da rede, deve retornar pela borda oposta.

Algumas propriedades termodinâmicas são fundamentais para o estudo desses modelos. Além das grandezas usuais como energia interna e o calor específico, pode-se calcular a entropia. Pode-se ainda introduzir outras quantidades que seriam equivalentes aos parâmetros de ordem no ferromagnetismo, o que auxiliará na identificação das fases enoveladas e compactas e fases não nativas. O número de contatos nativos em uma cadeia pode ser de grande valia na verificação de quão compacto encontra-se a cadeia.

A função de partição do sistema é calculada a partir de

$$\tilde{Z}=\sum_E e^{\tilde{S}(E)-\frac{E}{k_B T}}=e^{\tilde{S}(E_0)}\sum_E e^{S(E)-\frac{E}{k_B T}}, \qquad (3.23)$$

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

$$\langle O\rangle(T)=\frac{\sum_E c\,g(E)O(E)e^{\frac{-E}{k_B T}}}{\sum_E c\,g(E)e^{\frac{-E}{k_B T}}}=\frac{\sum_E g(E)O(E)e^{\frac{-E}{k_B T}}}{\sum_E g(E)e^{\frac{-E}{k_B T}}}, \qquad (3.24)$$

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

$$Z=\sum_E g(E)e^{\frac{-E}{k_B T}}. \qquad (3.25)$$

Portanto, pode-se obter a energia interna através de

$$U(T)=\frac{\sum_E g(E)Ee^{\frac{-E}{k_B T}}}{\sum_E g(E)e^{\frac{-E}{k_B T}}}, \qquad (3.26)$$

e o calor específico através de

$$c_v(T)=\frac{\partial U}{\partial T}=\frac{1}{T^2}\left(\langle E^2\rangle-\langle E\rangle^2\right). \tag{3.27}$$

Conhecendo g(E) tem-se

$$F(T)=-k_B T \ln\sum_E g(E)e^{\frac{-E}{k_B T}}, \tag{3.28}$$

logo é calculada a entropia por meio da expressão

$$S(T)=\frac{1}{T}\left(U-F(T)\right). \tag{3.29}$$

Já o número de contatos nativos é dado via

$$Q(T)=\frac{\sum_E g(E)Q(E)e^{\frac{-E}{k_B T}}}{\sum_E g(E)e^{\frac{-E}{k_B T}}}, \tag{3.30}$$

onde Q(E) pode ser obtido da média

$$Q(E)=\sum_k \frac{\Omega_{k(Q)}}{H(E)}. \tag{3.31}$$

A quantidade $\Omega_{k(Q)}$ 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.

Capítulo IV - RESULTADOS E DISCUSSÃO

Neste capítulo vamos apresentar os principais resultados obtidos com a técnica de construção de sequências introduzida no capítulo III. Trata-se da implementação do método de Wang-Landau para seleção de sequências de aminoácidos e síntese de uma cadeia protéica e que é a principal proposta desta dissertação. O principal objetivo é mostrar que as cadeias poliméricas desenhadas com o método possuem o comportamento esperado de uma cadeia protéica. Assim, em uma primeira etapa, as propriedades termodinâmicas das cadeias desenhadas serão investigadas sistematicamente para verificar se elas se comportam como uma "proteína". Em outra etapa, as cadeias desenhadas serão submetidas a inúmeras simulações de enovelamento para investigação da cinética desse processo . Com isso, espera-se que as cadeias desenhadas se enovelem para uma estrutura única e principalmente para aquela sobre a qual a sequência de monômeros foi organizada e ordenada. Uma vez neste estado, seja estável.

Para este fim, será adotado o modelo de rede para as proteínas, cujas cadeias serão constituídas de 27 monômeros. Os monômeros ocupam os sítios de uma rede cúbica e são conectados por ligações rígidas constituindo cadeias com comprimento fixo, tal como apresentadas no Capítulo II. Estudaremos os casos de cadeias com composição com 20 tipos de aminoácidos e com 15 tipos de aminoácidos. A idéia é comprovar a eficiência da técnica

de desenhar “proteínas” em todos estes casos. Além disso, cadeias com este comprimento e composições semelhantes já foram bastante estudadas e alguns resultados podem ser encontrados na literatura. Os modelos estudados nesta dissertação são considerados modelos canônicos para o estudo do enovelamento e do design de proteínas.

O estudo das propriedades termodinâmicas das cadeias desenhadas será feito utilizando também o método de Wang-Landau, recentemente adaptado para cálculo de quantidades termodinâmicas de proteínas em modelo de rede [47]. Como já apresentado no capítulo III, o método permite o cálculo da densidade de estados g(E) acessíveis à cadeia protéica e dela todas as quantidades termodinâmicas de interesse podem ser determinadas. Já para a investigação da cinética de enovelamento das cadeias desenhadas, será utilizado o algoritmo de Metropolis, tal como descrito no capítulo II.

Serão considerados 2 conjuntos de cadeias caracterizadas pelas estruturas alvo (pretensa estrutura nativa) escolhidas, pela composição de aminoácidos pelo número de aminoácidos na cadeia. O primeiro conjunto é constituído por cadeias formadas por composições com 20 tipos de aminoácidos. O segundo conjunto é formado por cadeias formadas por composições com 15 tipos de aminoácidos. A composição dos aminoácidos em todos os casos foi escolhida ao acaso, sem um critério específico. Todos os resultados foram obtidos numericamente através de programas escritos em linguagem FORTRAN e executados em três Pcs (computadores de mesa).

A seguir, serão apresentados os resultados referentes aos modelos de 20 letras e 15 letras, demonstrando a eficiência do algoritmo do design para estes casos estudados.

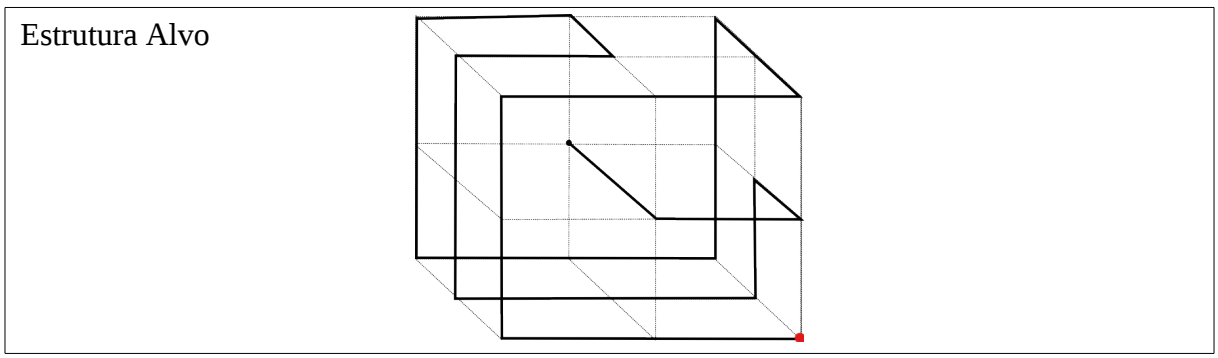


Fig 4.1 – Estrutura-alvo utilizada nas simulações de 20 e 15 Destaca-se (em vermelho) o sítio que marca o início da cadeia (sítio 1).

4.1 - Modelo com 20 letras

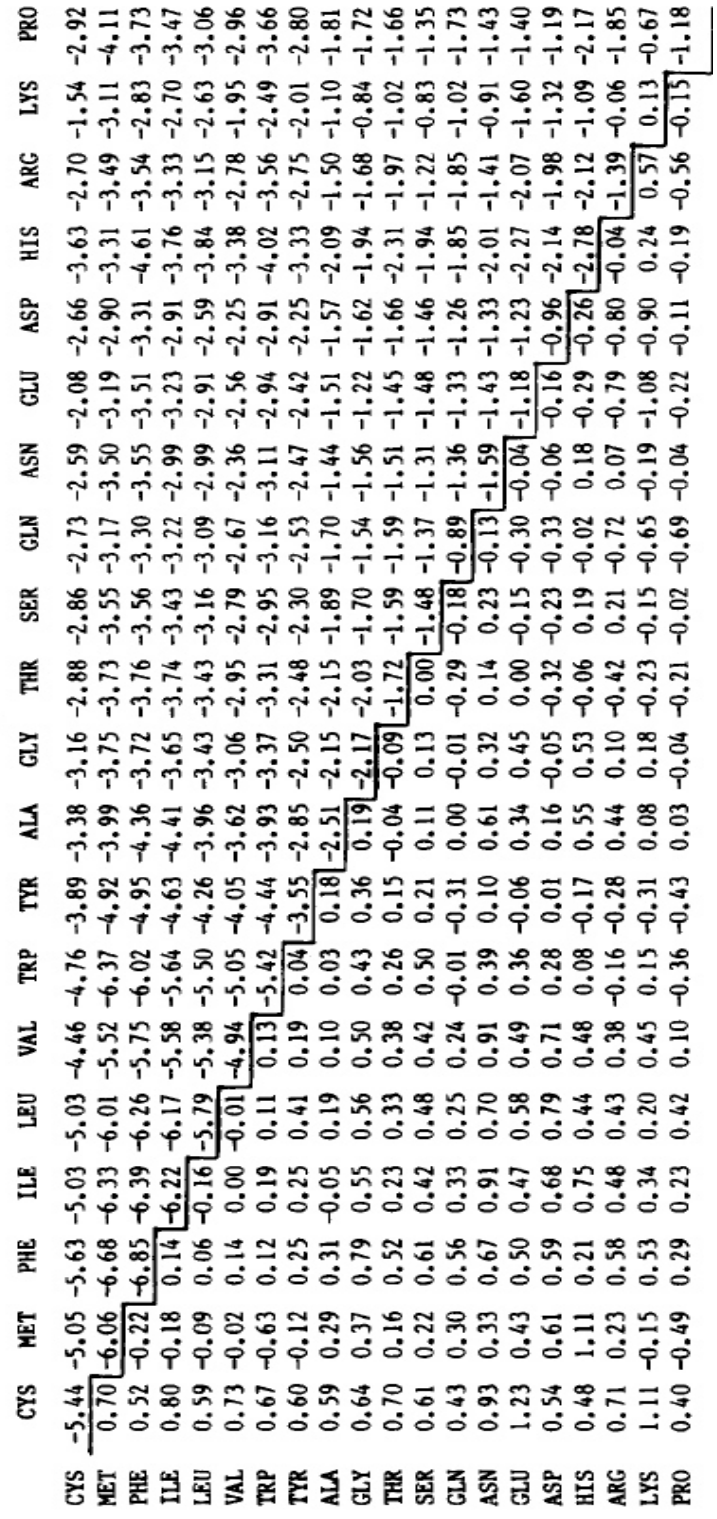
Este é o caso em que todos os tipos de aminoácidos serão considerados na cadeia protéica. Embora o modelo seja de rede e a cadeia seja relativamente curta, espera-se que os resultados reflitam o comportamento mais próximo de uma proteína. Neste caso, a estrutura alvo utilizada é a apresentada na Fig. 4.1, e foi apresentada pela primeira vez por Li[48-51]. Os aminoácidos da composição escolhida, serão distribuídos nos 27 sítios da cadeia e todas as possíveis sequências serão classificadas em ordem crescente de energia. Haverá uma imensa quantidade de sequências geradas pelo método. No entanto, espera-se que apenas um conjunto de cadeias construídas apresentem uma caraterística fundamental esperada: as cadeias irão se enovelar para a estrutura nativa desejada que é exatamente a estrutura alvo. Esta característica das cadeias credenciam-nas a serem candidatas a "proteína". Outros aspectos também são importantes. O estado nativo é a estrutura de menor energia. Uma vez que a composição de aminoácidos é sempre fixa, uma característica importante que também ajuda na escolha das cadeias candidatas a proteínas, é a de que as energias das cadeias construídas sejam as menores do espectro de energia. Assim sendo, apenas um conjunto de cadeias construídas, as de menores energias, foram escolhidas para serem testadas e serem confirmadas como "proteínas".

A energia de uma cadeia de aminoácidos é dada por

$$E=-\sum_{i,j}\epsilon_{i,j}\Delta(r_j-r_i), \tag{4.1}$$

onde os coeficientes $\epsilon_{i,j}$ são as energias de contacto entre os aminoácidos nos sítios r_i e r_j respectivamente. A função $\Delta(r_i-r_j)$ é dada por $\Delta(r_i-r_j)=1$ 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. As energias de contacto $\epsilon_{i,j}$ foram calculadas por Miyazawa e Jernigan [23] e estão disponíveis na próxima página (Fig. 4.2).

Estas energias do potencial de contato estão dadas em unidades da constante universal RT_0, que equivale a aproximadamente a 0,6 kcal/mol. Devido a isto, k_BT deve ser também dada em unidades de RT_0. Entretanto, apenas por conveniência, será considerado k_B=1 de modo que, além das energias, as temperaturas também serão dadas em termos de T_0. Onde T_0 é dada na temperatura ambiente (T_0 = 300 K).



	CYS	MET	PHE	ILE	LEU	VAL	TRP	TYR	ALA	GLY	THR	SER	GLN	ASN	GLU	ASP	HIS	ARG	LYS	PRO
CYS	-5.44	-5.05	-5.63	-5.03	-5.03	-4.46	-4.76	-3.89	-3.38	-3.16	-2.88	-2.86	-2.73	-2.59	-2.08	-2.66	-3.63	-2.70	-1.54	-2.92
MET	0.70	-6.06	-6.68	-6.33	-6.01	-5.52	-6.37	-4.92	-3.99	-3.75	-3.73	-3.55	-3.17	-3.50	-3.19	-2.90	-3.31	-3.49	-3.11	-4.11
PHE	0.52	-0.22	-6.85	-6.39	-6.26	-5.75	-6.02	-4.95	-4.36	-3.72	-3.76	-3.56	-3.30	-3.55	-3.51	-3.31	-4.61	-3.54	-2.83	-3.73
ILE	0.80	-0.18	0.14	-6.22	-6.17	-5.58	-5.64	-4.63	-4.41	-3.65	-3.74	-3.43	-3.22	-2.99	-3.23	-2.91	-3.76	-3.33	-2.70	-3.47
LEU	0.59	-0.09	0.06	-0.16	-5.79	-5.38	-5.50	-4.26	-3.96	-3.43	-3.43	-3.16	-3.09	-2.99	-2.91	-2.59	-3.84	-3.15	-2.63	-3.06
VAL	0.73	-0.02	0.14	0.00	-0.01	-4.94	-5.05	-4.05	-3.62	-3.06	-2.95	-2.79	-2.67	-2.36	-2.56	-2.25	-3.38	-2.78	-1.95	-2.96
TRP	0.67	-0.63	0.12	0.19	0.11	0.13	-5.42	-4.44	-3.93	-3.37	-3.31	-2.95	-3.16	-3.11	-2.94	-2.91	-4.02	-3.56	-2.49	-3.66
TYR	0.60	-0.12	0.25	0.25	0.41	0.19	0.04	-3.55	-2.85	-2.50	-2.48	-2.30	-2.53	-2.47	-2.42	-2.25	-3.33	-2.75	-2.01	-2.80
ALA	0.59	0.29	0.31	-0.05	0.19	0.10	0.03	0.18	-2.51	-2.15	-2.15	-1.89	-1.70	-1.44	-1.51	-1.57	-2.09	-1.50	-1.10	-1.81
GLY	0.64	0.37	0.79	0.55	0.56	0.50	0.43	0.36	0.19	-2.17	-2.03	-1.70	-1.54	-1.56	-1.22	-1.62	-1.94	-1.68	-0.84	-1.72
THR	0.70	0.16	0.52	0.23	0.33	0.38	0.26	0.15	-0.04	-0.09	-1.72	-1.59	-1.59	-1.51	-1.45	-1.66	-2.31	-1.97	-1.02	-1.66
SER	0.61	0.22	0.61	0.42	0.48	0.42	0.50	0.21	0.11	0.13	0.00	-1.48	-1.37	-1.31	-1.48	-1.46	-1.94	-1.22	-0.83	-1.35
GLN	0.43	0.30	0.56	0.33	0.25	0.24	-0.01	-0.31	0.00	-0.01	-0.29	-0.18	-0.89	-1.36	-1.33	-1.26	-1.85	-1.85	-1.02	-1.73
ASN	0.93	0.33	0.67	0.91	0.70	0.91	0.39	0.10	0.61	0.32	0.14	0.23	-0.13	-1.59	-1.43	-1.33	-2.01	-1.41	-0.91	-1.43
GLU	1.23	0.43	0.50	0.47	0.58	0.49	0.36	-0.06	0.34	0.45	0.00	-0.15	-0.30	-0.04	-1.18	-1.23	-2.27	-2.07	-1.60	-1.40
ASP	0.54	0.61	0.59	0.68	0.79	0.71	0.28	0.01	0.16	-0.05	-0.32	-0.23	-0.33	-0.06	-0.16	-0.96	-2.14	-1.98	-1.32	-1.19
HIS	0.48	1.11	0.21	0.75	0.44	0.48	0.08	-0.17	0.55	0.53	-0.06	0.19	-0.02	0.18	-0.29	-0.26	-2.78	-2.12	-1.09	-2.17
ARG	0.71	0.23	0.58	0.48	0.43	0.38	-0.16	-0.28	0.44	0.10	-0.42	0.21	-0.72	0.07	-0.79	-0.80	-0.04	-1.39	-0.06	-1.85
LYS	1.11	-0.15	0.53	0.34	0.20	0.45	0.15	-0.31	0.08	0.18	-0.23	-0.15	-0.65	-0.19	-1.08	-0.90	0.24	0.57	0.13	-0.67
PRO	0.40	-0.49	0.29	0.23	0.42	0.10	-0.36	-0.43	0.03	-0.04	-0.21	-0.02	-0.69	-0.04	-0.22	-0.11	-0.19	-0.56	-0.15	-1.18

Fig 4.2 –Matriz de potenciais de interação de Miyazawa e Jernigan retirada da referência [23]. Na diagonal superior, o valor das energias de contato entre os 20 aminoácidos naturais são dadas em termos de RT_0.

Para desenhar as sequências, foi escolhida a composição fixa dada na tabela 4.1. Alguns aminoácidos como L, M,N,P,Q, R e Y apresentarão dois representantes ao longo da cadeia. Estes monômeros foram dispostos nos sítios da estrutura alvo e foram submetidas ao algoritmo de seleção de sequências baseado no método Wang-Landau apresentado no capítulo III. Uma grande quantidade de sequências foram determinadas e classificadas. Dessas, uma quantidade de 120 sequências de menores energias foram submetidas a intensivos testes para verificar se as cadeias enovelariam para o estado nativo estabelecido pela estrutura alvo. Todas enovelam para o estado nativo como esperado. Entretanto, apenas o conjunto de cadeias com a menor energia, de -97,2, foram classificadas na tabela 4.2 e serão estudados neste capítulo.

Tipo de aminoácido	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Número de aminoácidos presente na composição	1	1	1	1	1	2	1	1	1	1	1	2	2	2	2	2	2	1	1	1

Tabela 4.1 : Composição utilizada no modelo composto por 20 letras. Cada aminoácido natural está representado por seu símbolo característico.

Mais adiante, será mostrado um gráfico típico do parâmetro de similaridade em função do tempo (Fig.4.11) que ilustra o enovelamento dessas cadeias para o estado nativo. Além de enovelarem-se para a estrutura nativa desejada, um levantamento do espectro de estados acessíveis a essas cadeias desenhadas revelam que o espectro de energia apresenta um “gap” entre a energia do estado nativo e a borda inferior da região contínua do espectro. Este “gap” pode ser claramente visto no histograma apresentado na Fig. 4.3. A presença de gaps no espectro é um indicativo de que a cadeia de aminoácidos apresenta características de proteína pois o “gap” pode acelerar o processo de enovelamento. Há uma hipótese de que o “gap” de energia seja resultado de efeito cooperativo entre os monômeros da cadeia. Entretanto não há uma conclusão definitiva quanto a isso [52]. O histograma apresentado na figura 4.3 refere-se apenas à sequência S7.20. Mas os histogramas das 20 cadeias foram construídos e os “gaps correspondentes assim como as temperaturas de transição foram determinadas e apresentadas na tabela 4.2.

	Sequências	Gap de energia	Temperatura de transição
S1.20	EPNPSHDGQVETQASTCGWNMKLRYFI	9,40	1,79
S2.20	EPQPSHDGQVETNASTCGWNMKLRYFI	8,70	1,79
S3.20	RQNHTPSPDAKVQGSTMGCNLEWEYFI	7,30	1,78
S4.20	RQNHTPKPDASVQGSTMGCNLEWEYFI	7,50	1,80
S5.20	RQQHTPKPDASVNGSTMGCNLEWEYFI	7,90	1,77
S6.20	GQNHDPSPNAQVKSETMRCELTWGYFI	8,40	1,75
S7.20	GQNHDPSPQANVKSETMRCELTWGYFI	9,40	1,75
S8.20	EHNQSPEPSADGQVTTMGLNCRWKYFI	8,50	1,75
S9.20	EHQQSPEPSADGNVTTMGLNCRWKYFI	7,70	1,73
S10.20	GHDQKPSPQANTEVSTMELRCNWGYFI	8,10	1,76
S11.20	GHEQSPTPQANSDVKTMELRCNWGYFI	7,80	1,78
S12.20	THDQKPNPQASGEVSTMELRCGWNYFI	7,70	1,78
S13.20	GHEQKPNPQASTDVSTMELRCGWNYFI	7,70	1,80
S14.20	GHKQDPSPNAQSNVETMRLECTWGYFI	8,30	1,78
S15.20	GHKQDPSPQANSNVETMRLECTWGYFI	8,10	1,77
S16.20	GHKQEPSPNAQSNVDTMRLECTWGYFI	7,50	1,76
S17.20	GHSQEPNPQATSKVDTMRLECGWNYFI	8,40	1,77
S18.20	GHKQEPNPQATSSVDTMRLECGWNYFI	8,90	1,76
S19.20	GHKQDPNPQASSNVETMRLECGWTYFI	8,90	1,76
S20.20	GHKQEPNPQASSNVDTMRLECGWTYFI	7,80	1,76

Tabela 4.2 : As 20 sequências de menor energia geradas pelo algoritmo de “design” via método de Wang-Landau. Na terceira coluna da tabela apresenta-se os “gaps” de energia entre o estado nativo e o espectro contínuo de energias. Na última coluna, são apresentadas as temperaturas de transição.

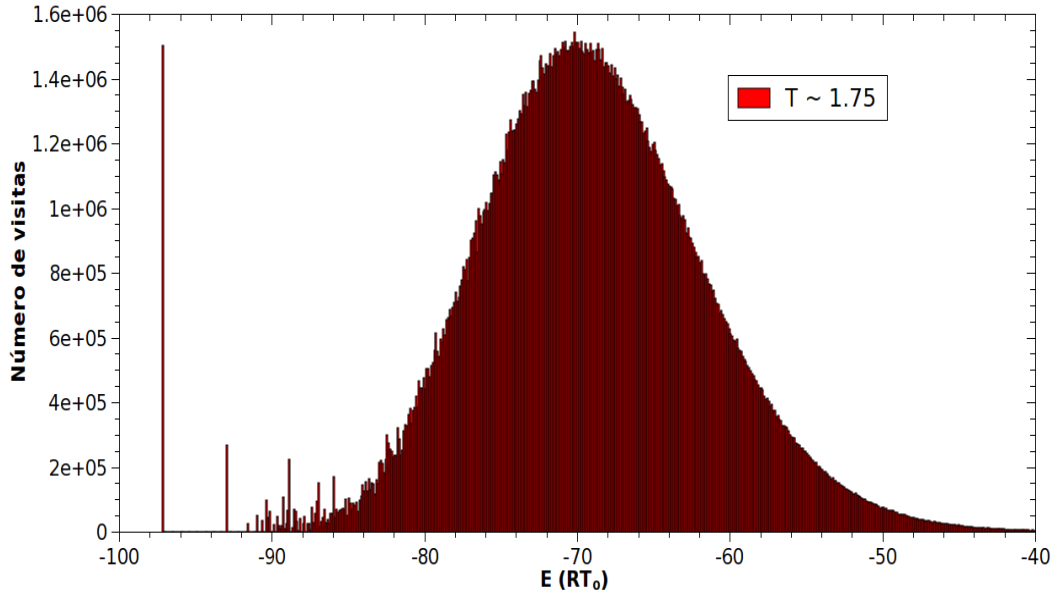


Fig 4.3 – Histograma de energia para a sequência S7.20 do modelo composto por 20 aminoácidos próximo à temperatura crítica. Observa-se duas regiões bem definidas no espectro, uma região composta por valores de energia muito próximos entre si, e outra região composta por valores discretos de energia. A diferença entre o estado de menor energia e o limite do espectro contínuo, constitui o chamado "*gap*".

Este histograma foi construído utilizando o algoritmo de Metropolis fixando-se a temperatura em um valor escolhido convenientemente e registrando o número de visitas efetuadas pela cadeia a cada estado acessível a essa temperatura. Nota-se que a probabilidade da cadeia estar no estado de menor energia é a mesma para que ela seja encontrada a um estado de energia correspondente a um estado desnaturado (cadeia aberta). O histograma revela que a cadeia de aminoácidos pode passar da fase ordenada para a fase desnaturada através de uma transição de fase similar às transições de primeira ordem em sistemas magnéticos. Esta característica apresentada pelas cadeias desenhadas indica que elas se comportam de maneira muito similar às proteínas globulares [53].

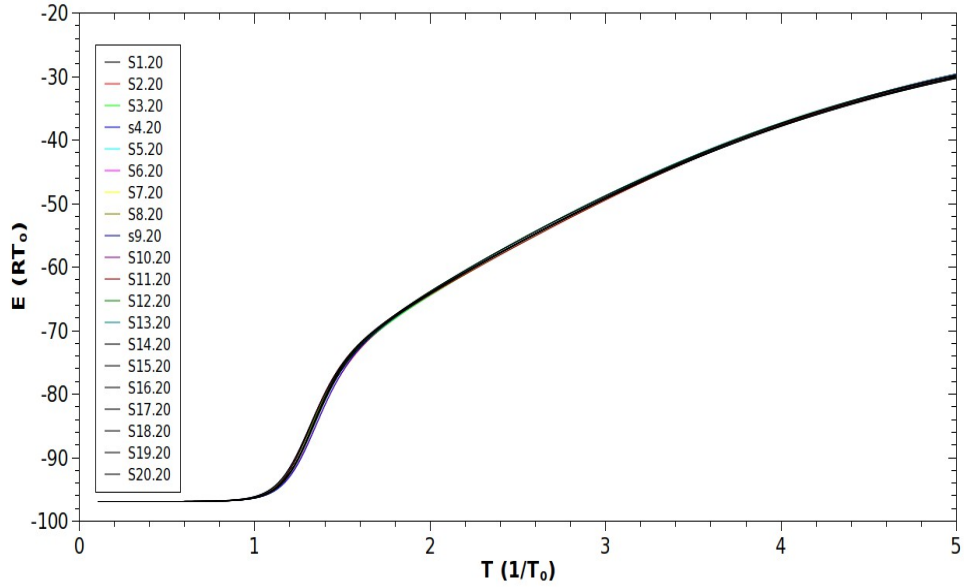


Fig4.4 - Energia média em função da temperatura para as vinte sequências escolhidas a partir do modelo de 20 letras desenhadas para estrutura-alvo.

As sequências classificadas na tabela 4.2 apresentaram propriedades termodinâmicas e cinéticas similares. Todas apresentaram “gaps” de energia com valores semelhantes e as temperaturas de transição são praticamente as mesmas. As quantidades termodinâmicas como a energia interna média, calor específico, número de contatos nativos e a entropia, dadas respectivamente nas figuras 4.4, 4.5, 4.6 e 4.7, todas calculadas em função da temperatura mostram isso. Estas quantidades foram calculadas para todas as cadeias com as sequências apresentadas na tabela 4.2. Todas as quantidades foram obtidas da densidade de estados g(E) calculada aplicando o método de Wang-Landau para estudo de propriedades termodinâmicas. A energia média e o calor específico mostram claramente duas fases. Na região de baixas temperaturas tem-se o mínimo da energia média e o fato dela não se modificar nesse intervalo de temperatura, mostra que a cadeia se encontra no estado nativo. Na região de temperatura acima de 1.6 , tem-se a fase das cadeias desnaturadas. O calor específico com um pico agudo e acentuado sugere que a transição entre as fases enovelada e a desnaturada seja similar à transição de fase de primeira ordem, como já revelado

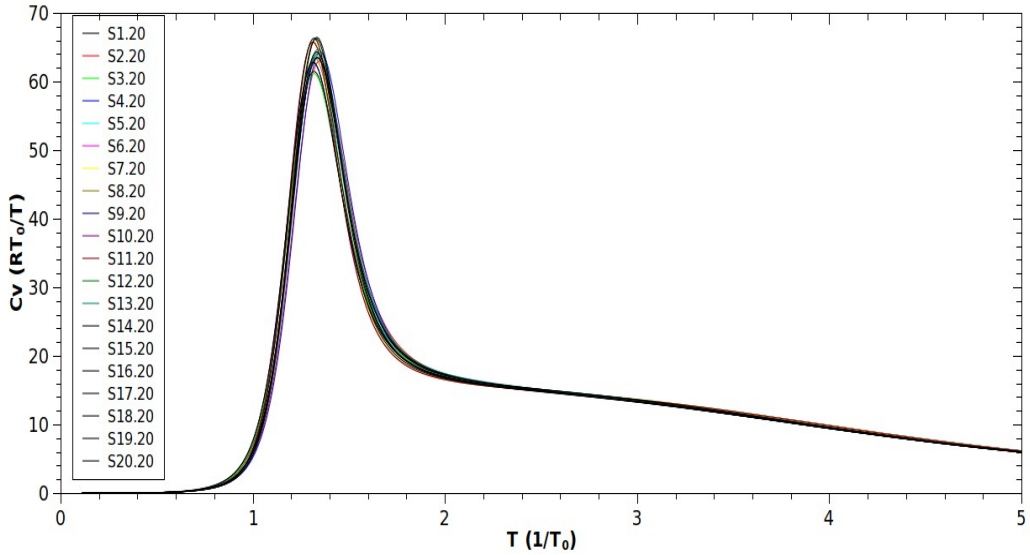


Fig. 4.5 - Gráfico do calor específico para o conjunto de sequências compostas por 20 letras

pelo histograma apresentado na Fig. 4.3. Mas é através do número médio de contatos nativos que se pode ver claramente que, na região de baixa temperatura, as cadeias encontram-se nos estados de estrutura cúbica. Não podemos afirmar, somente com esta informação, que a cadeia encontra-se no

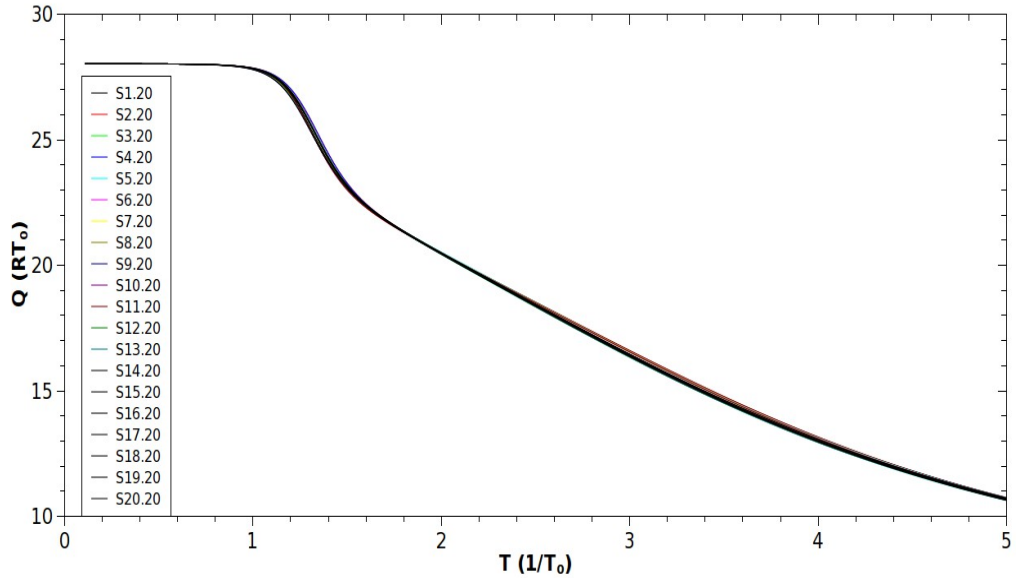


Fig. 4.6 - Gráfico do número de contatos médio em função da temperatura para o conjunto de sequência de 20 monômeros.

estado nativo ainda que o número de contatos seja 28 . Mais adiante quando estivermos estudando a cinética de enovelamento, teremos a comprovação de que se trata do estado nativo. As cadeias nas conformações desnaturadas, como esperado, o número de contatos nativos é pequeno.

Finalmente, a entropia média das cadeias, que mostra claramente as fases ordenadas (enovelada) e desordenadas (desnaturadas) e confirmam as conclusões obtidas nesta discussão. Mais ainda, um fato importante: a entropia vai a zero a uma temperatura não nula. Assim, para todas as cadeias desenhadas, a entropia mostra que existe um único estado de menor energia para cada cadeia, e reforça a idéia de que estes estados possam ser exatamente os estados nativos de cada cadeia. Outro fato verificado com os cálculos das quantidades termodinâmicas, é todas as cadeias desenhadas comportam-se de maneira muito semelhante entre si. Se as cadeias são constituídas pelo mesmo conjunto de aminoácidos, e essas cadeias se enovelam para a mesma estrutura nativa, podemos afirmar que a permutação de dois ou mais aminoácidos em uma cadeia, pode gerar uma nova proteína com propriedades muito semelhantes às da proteína original. A investigação desse fenômeno pode ser importante no estudo da mutação em proteínas, e mostra a importância do método de seleção de sequências proposto neste trabalho.

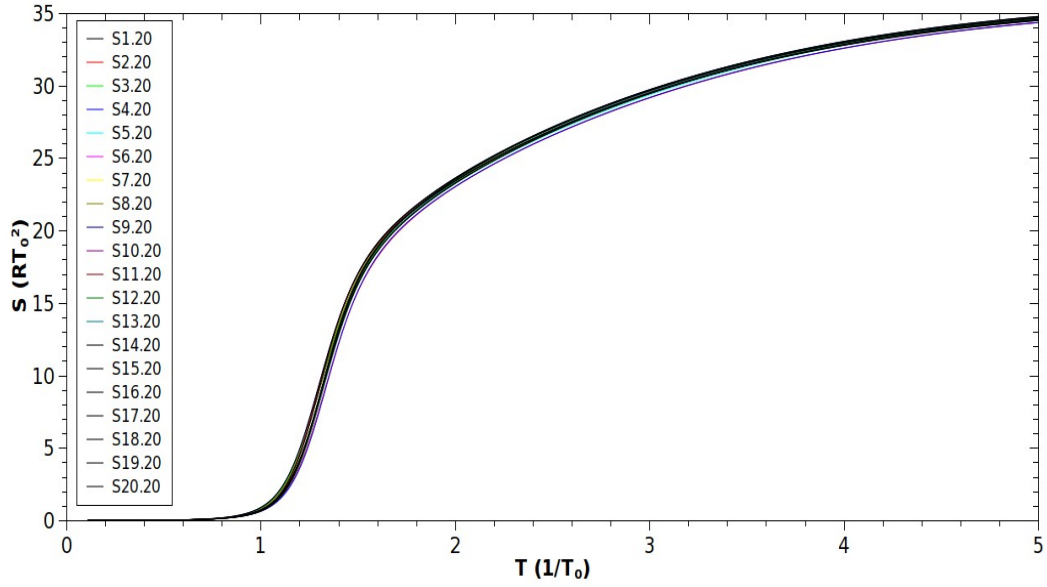


Fig. 4.7 - Entropia em função a temperatura para o conjunto de sequências do modelo de vinte letras.

As cadeias também foram submetidas às simulações de cinética de não-equilíbrio utilizando o algoritmo de Metropolis. As cadeias foram inicialmente tomadas nas formas desnaturadas, e deixadas para evoluir no espaço das conformações. Esperava-se que as cadeias evoluissem em sua "trajetória" de enovelamento e em um dado instante atingisse a conformação nativa. A idéia foi registrar o tempo gasto para que uma cadeia saisse de uma conformação aberta e atingisse a conformação nativa. Mas este tempo, que será denominado de tempo de enovelamento, depende da temperatura. E isto é confirmado pelos resultados mostrados na figura 4.8. Neste gráfico está registrado o tempo de enovelamento das cadeias em função da temperatura. Existe uma faixa de temperatura em que as cadeias enovelaram-se gastando tempos menores. Em regiões de baixa temperatura, o tempo de enovelamento foi maior. A razão disso é que a probabilidade para que uma cadeia passe de uma conformação para outra a baixa temperaturas é pequena. Já no outro extremo também o tempo de enovelamento aumenta pois o número de conformações acessíveis é muito grande fazendo com que a probabilidade para que a conformação nativa seja visitada seja pequena.

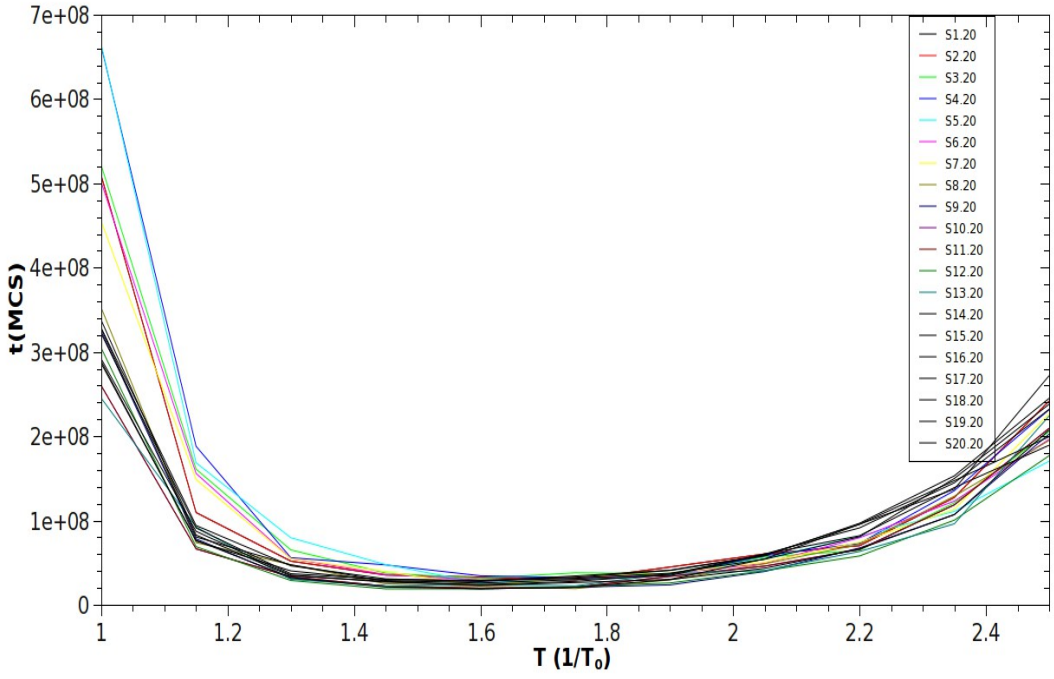


Fig 4.8 - Gráfico de tempo de enovelamento por temperatura para o conjunto de sequências compostas por 20 tipos de aminoácidos.

Continuando, aspectos da evolução temporal de enovelamento também foram investigados. Fixando a temperatura em valores para os quais o tempo de enovelamento fosse mínimo, foi registrada a evolução do número de contatos nativos, da energia e do parâmetro de similaridade das conformações, apresentadas respectivamente nas figuras 4.9, 4.10 e 4.11. Uma vez que as cadeias desenhadas possuem comportamento semelhantes, foi escolhida a cadeia com a sequência S7.20 para ser investigada. Em todos os gráficos, pode-se ver que a cadeia compacta-se rapidamente, assumindo uma forma em que o número de contatos nativos é 25. Na forma mais compacta, que é a nativa, o número de contatos é 28. Assim, aparentemente a cadeia assume uma forma próxima da conformação nativa. Mas como se pode ver no gráfico do parâmetro de similaridade, embora neste instante a forma seja compacta, a conformação ainda está longe de ser a da nativa. Esta compactação rápida é conhecida na literatura como colapso hidrofóbico e é resultado do efeito cooperativo entre os monômeros. Logo após o colapso, a cadeia passa um longo tempo neste regime de cadeia compacta mas em um processo de reordenamento. Pode-se observar pelo número de

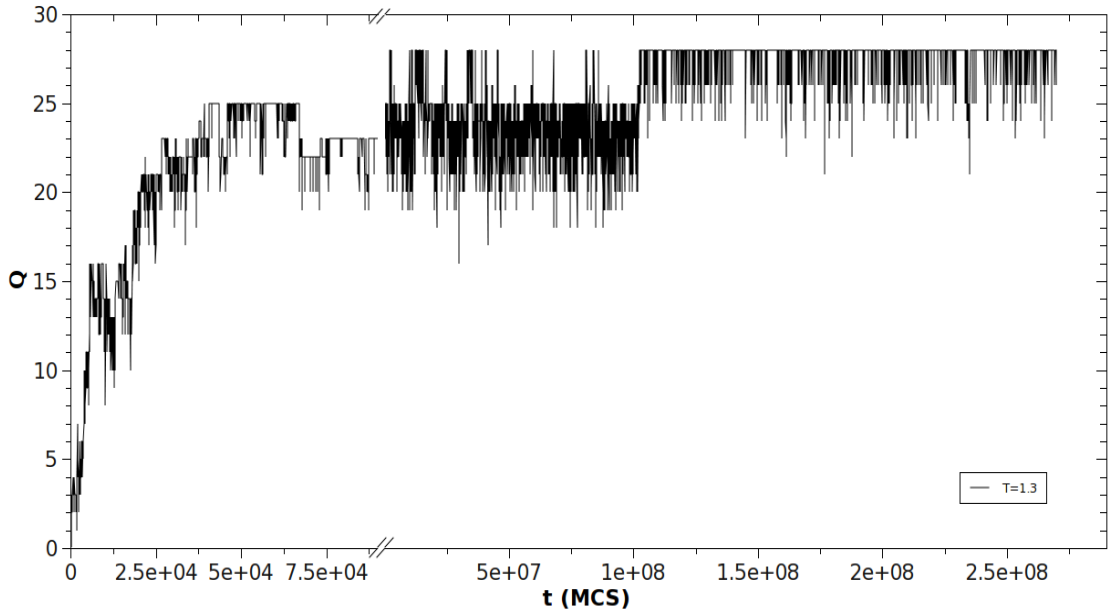


Fig 4.9 – Número de contatos em função do tempo de enovelamento em *Monte Carlo Steps* para a sequência S7.20. Observa-se duas fases características (compactação e reconfiguração) apresentada pela cadeia durante o processo de enovelamento.

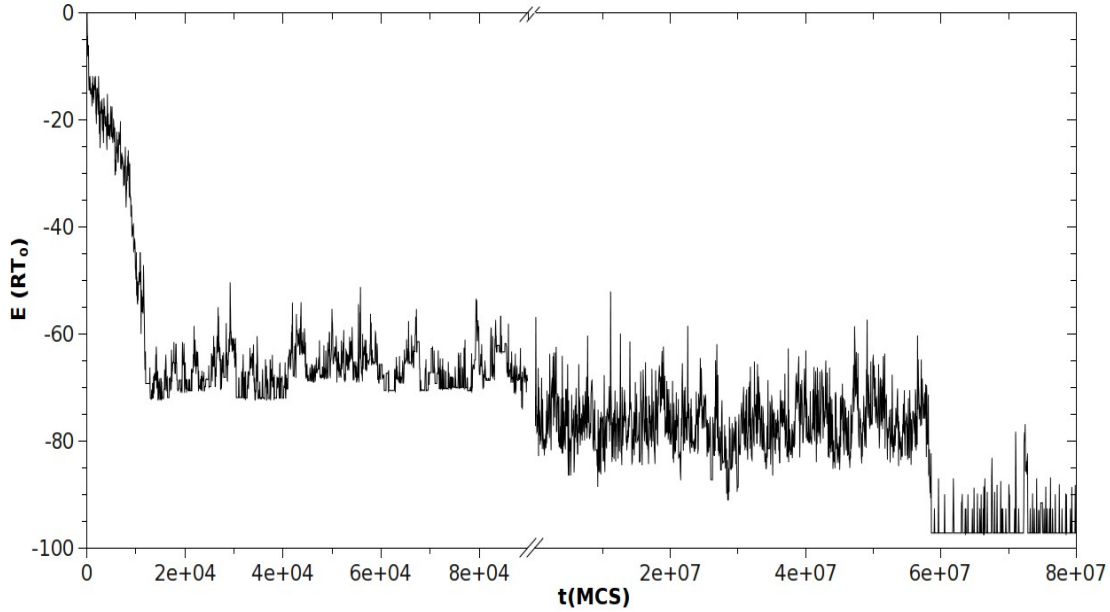


Fig 4.10 – Energia em função do tempo de enovelamento em *Monte Carlo steps* para a sequência S7.20 a temperatura $T = 1{,}3$.

contatos nativos, que durante este período de reordenamento da cadeia, ela assume várias vezes uma conformação de estrutura cúbica. No entanto, como a energia correspondente não é a menor energia, não se trata do estado nativo, apenas uma das conformações possíveis para a cadeia durante este regime. O parâmetro de similaridade confirma que a conformação atingida é a conformação de estrutura nativa.

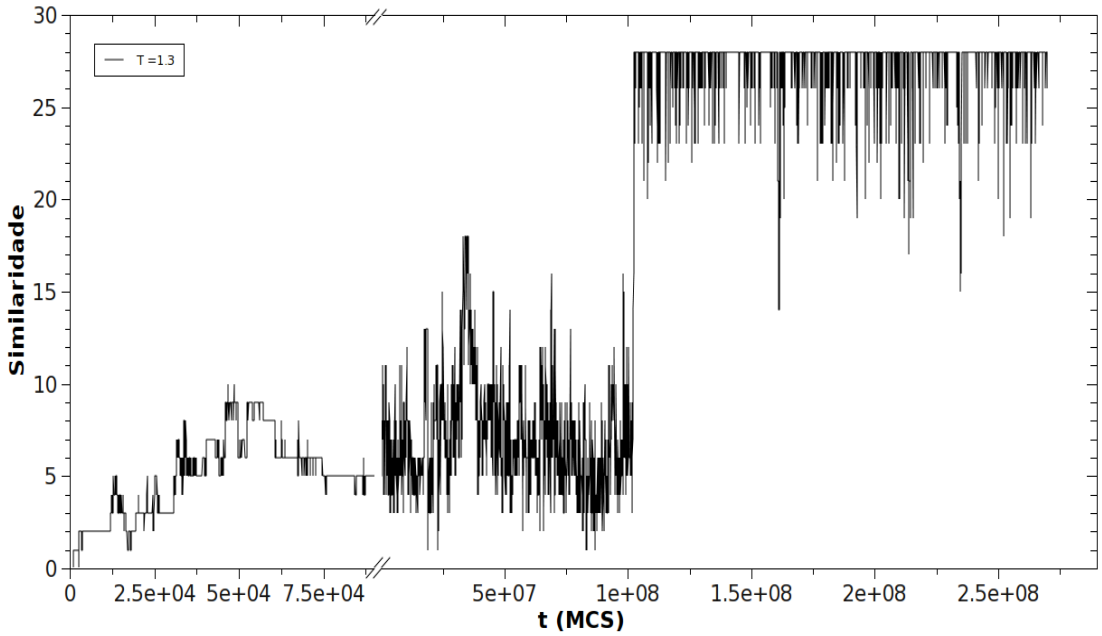


Fig 4.11 – Parâmetro de similaridade em função do tempo em *Monte Carlo steps* para a sequência S7.20.

Finalizando, todas as sequências apresentadas na tabela 4.2 foram submetidas às simulações de enovelamento. A evolução temporal do número de contatos nativos, da energia e do parâmetro de similaridade foram acompanhadas e registradas. Todas as cadeias com as sequências selecionadas na tabela 4.2 enovelaram-se para o estado nativo e comportaram-se como proteínas. Em nenhum caso, foi detectado duas ou mais estruturas diferentes com a mesma energia do estado nativo, o que inviabilizaria a cadeia correspondente como uma proteína. Frente a estes resultados, o algoritmo proposto para selecionar sequências é bastante eficiente e uma ferramenta bastante útil no estudo do enovelamento de proteínas.

4.2 - Modelo de 15 letras

Dando continuidade ao desenho de sequências, foi considerado também uma composição com menor número de aminácidos. A idéia foi verificar que tipo de alterações a diminuição de monômeros acarretaria no processo de sintetização da cadeia e, como as propriedades termodinâmicas e cinéticas dessas cadeias seriam afetadas pela diminuição do

números de monômeros. Novamente a composição foi escolhida aleatoriamente, sem nenhum critério em particular. Os aminoácidos D, H, K, P e R não participam da composição. A estrutura alvo é a mesma adotada no caso do modelo com 20 aminoácidos. Novamente as sequências com as menores energias foram classificadas em ordem crescente. Mas apenas as sequências de menor energia, $E_{min} = -123{,}30$, classificadas na tabela 4.4, foram sistematicamente estudadas. Ao todo foram 15 sequências. Quanto às propriedades termodinâmicas, todas as cadeias responderam de modo semelhante as cadeias com 20 letras, apenas os valores numéricos das grandezas foram modificadas. Por exemplo, como já vimos, a energia da estrutura nativa diminuiu e o valor da temperatura de transição aumentou. Mas as curvas de energia média, calor específico, número de contatos e entropia são muito semelhantes as obtidas com o modelo de 20 letras.

Tipo de aminoácido	A	C	E	F	G	I	L	M	N	Q	S	T	V	W	Y
Número de aminoácidos presente na composição	2	1	2	2	2	2	2	1	2	1	2	2	2	2	2

Tabela 4.3 : Composição das sequências desenhadas para a estrutrura-alvo, apresentando 15 tipos de aminoácidos.

Na tablea 4.4 estão apresentadas as 15 sequências testadas. Os valores dos “gaps” de energia e os valores de temperatura de transição foram determinados através da construção de histogramas de estados acessíveis, como o apresentado na figura 4.12. que corresponde ao histograma da sequência S2.15. Estes histogramas são típicos de sistemas que passam de uma fase ordenada para a desordenada (e vice-versa) através de uma transição semelhante à de primeira ordem.

	sequências	Gap de energia	Temperatura de transição
S1.15	GLNVEASAQVTLNCEYWSFYIGWTIMF	8,9	2,02
S2.15	GLQVEASANVTLNCEYWSFYIGWTIMF	7,0	2,03
S3.15	GLQLEASANVTVNCEYWSFYIGWTIMF	9,0	2,02
S4.15	GVNLEASAQVTCNLEYWSIYFGWTIMF	9,6	2,02
S5.15	GLNVEATAQVSLNCEYWSFYITWGIMF	7,3	2,02
S6.15	GLNLEATAQVSVNCEYWSFYITWGIMF	8,7	2,02
S7.15	GLNLEATANVSVQCEYWSFYITWGIMF	7,1	2,01
S8.15	GLEVNASANVTLECQYWYFSIGWTIMF	7,8	2,02
S9.15	GLELNASANVTVECQYWYFSIGWTIMF	8,8	2,02
S10.15	GLNLEATAQVSCNVEYWSIYFTWGIMF	6,7	2,04
S11.15	GLNLEATANVSCQVEYWSIYFTWGIMF	8,6	2,03
S12.15	GLEVNATANVSLECQYWYFSITWGIMF	8,5	2,13
S13.15	GLELQATANVSCELNYWYISFTWGIMF	8,9	2,09
S14.15	GLELQATANVSCEVNYWYISFTWGIMF	7,0	2,03
S15.15	GLELNATANVSCEVQYWYISFTWGIMF	6,3	2,14

Tabela 4.4 : Estudo das sequências de menor energia escolhidas para o modelo de 15 letras, obtidas a partir do sequênciamento da estrutura-alvo e utilizando-se na composição 15 aminoácidos diferentes.

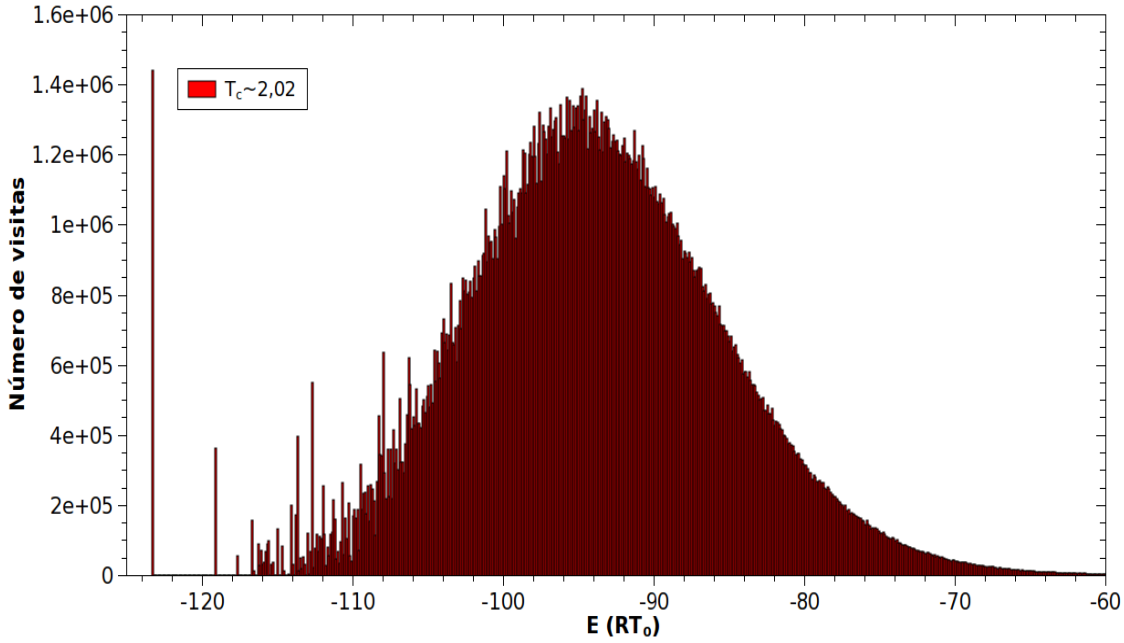


Fig 4.12 - Histograma de número de visitas por energia para a sequência S2.15 do modelo de 15 letras.

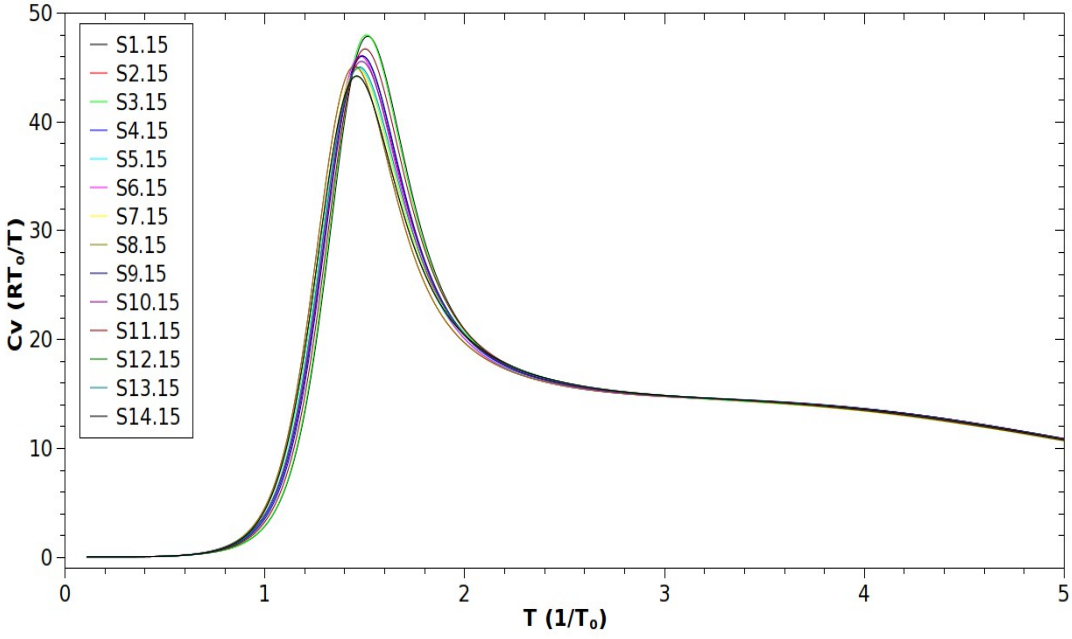


Fig 4.13 – Calor específico versus temperatura para todas as 15 sequências selecionadas do modelo de 15 letras.

Quanto às grandezas termodinâmicas, será suficiente a apresentação do calor específico e da entropia já que as cadeias de 15 letras e de 20 letras apresentam resultados muito semelhantes. O calor específico e a entropia caracterizam bem o sistema quanto às fases possíveis. Novamente a entropia se anula em temperatura não nula o que indica que o estado nativo é não degenerado.

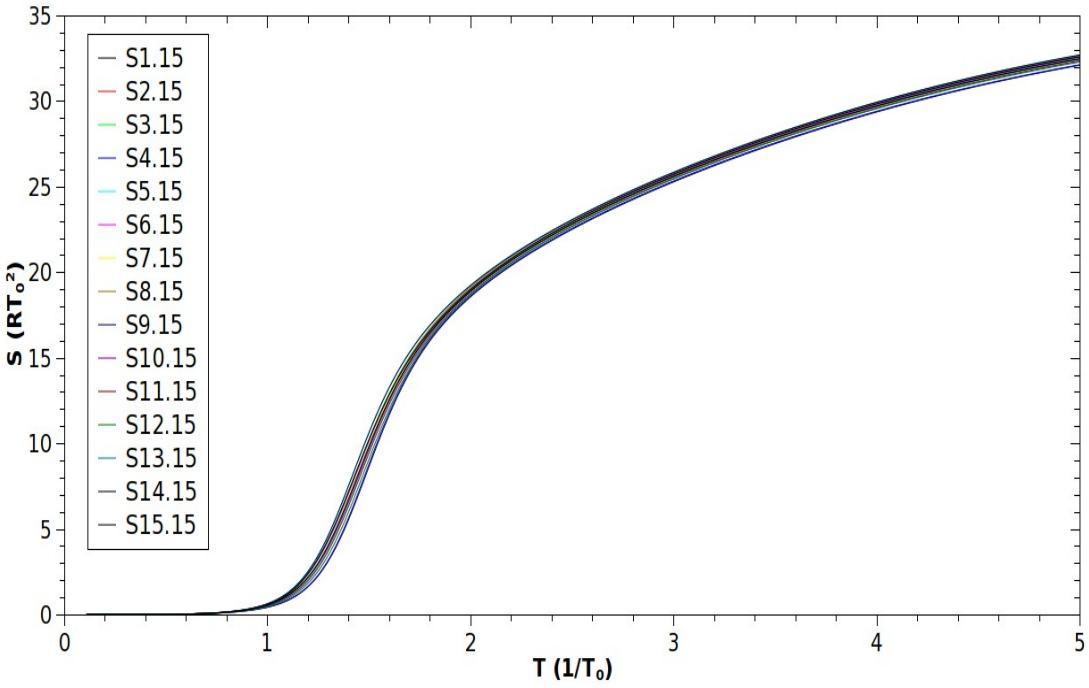


Fig 4.14 – Entropia versus temperatura para todas as 15 sequências selecionadas do modelo de 15 letras.

Já na investigação dos aspectos cinéticos do enovelamento das cadeias de 15 letras, existem algumas diferenças em relação às cadeias de 20 letras. A primeira constatação é a de que o tempo de enovelamento aumenta para todas as cadeias. E a faixa de temperatura onde o tempo de enovelamento é mínimo, é deslocado para uma faixa de temperatura um pouco maior, como ilustrado na figura 4.15. Se comparado com as cadeias de 20 letras, algumas cadeias do modelo de 15 letras demoram mais tempo para se enovelarem para o estado nativo porque há aumento da frustração acarretada pelo aprisionamento da cadeia em conformações

com energias próximas do valor mínimo. Isto torna-se mais evidente a baixas temperaturas, quando a probabilidade de modificação da conformação da cadeia é muito pequena.

Finalizando, a título de ilustração, são apresentadas nas figuras 4.16, 4.17 e 4.18, a evolução temporal do número de contatos nativos, da energia e do parâmetro de similaridade da cadeia com

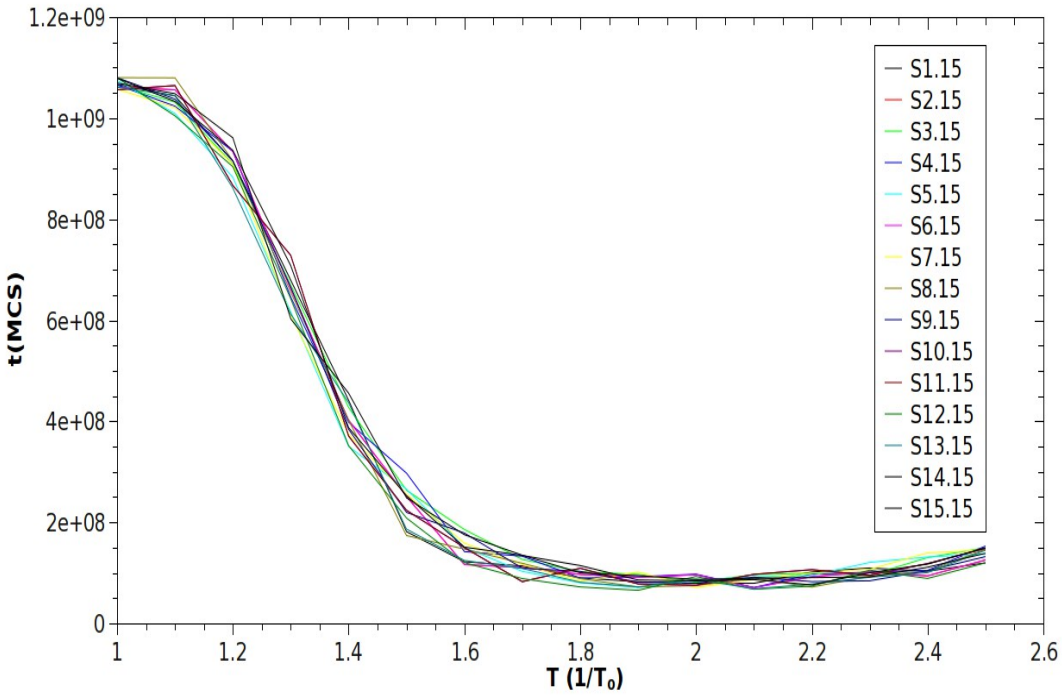


Fig 4.15 – Tempo de enovelamento das 15 cadeias com as sequências de monômeros selecionadas em função da temperatura para o modelo de 15 letras.

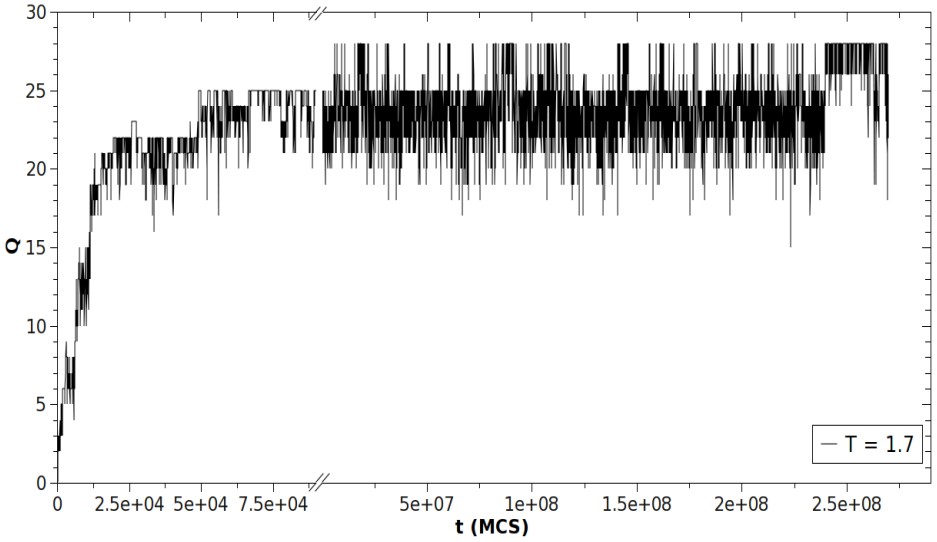


Fig 4.16 – Evolução do número de contatos nativos da cadeia com a sequência S2.15

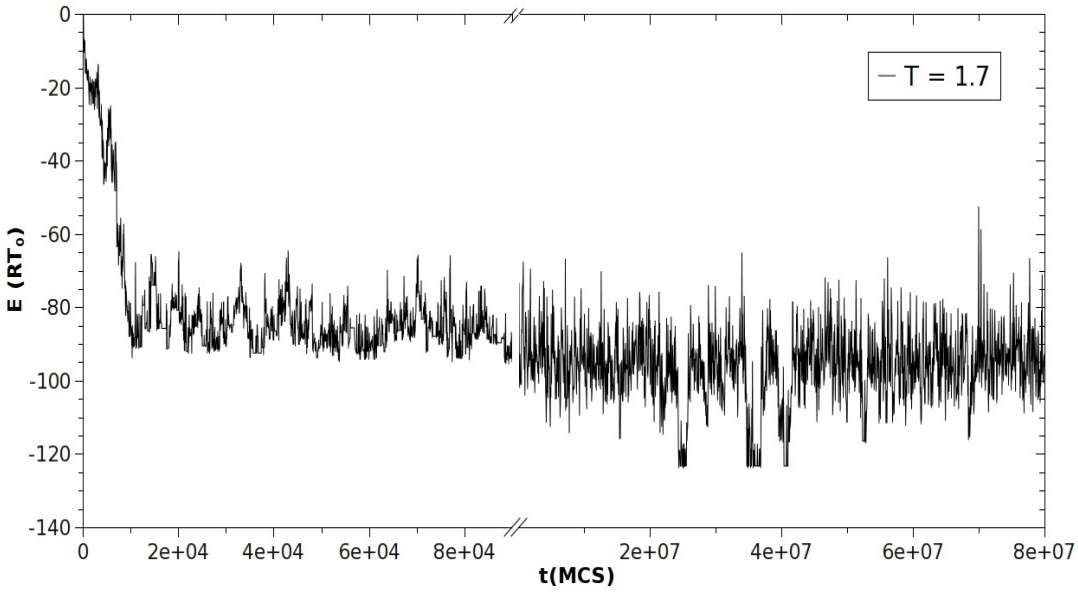


Fig 4.17 – Evolução da energia durante o enovelamento da cadeia com a sequência S2.15

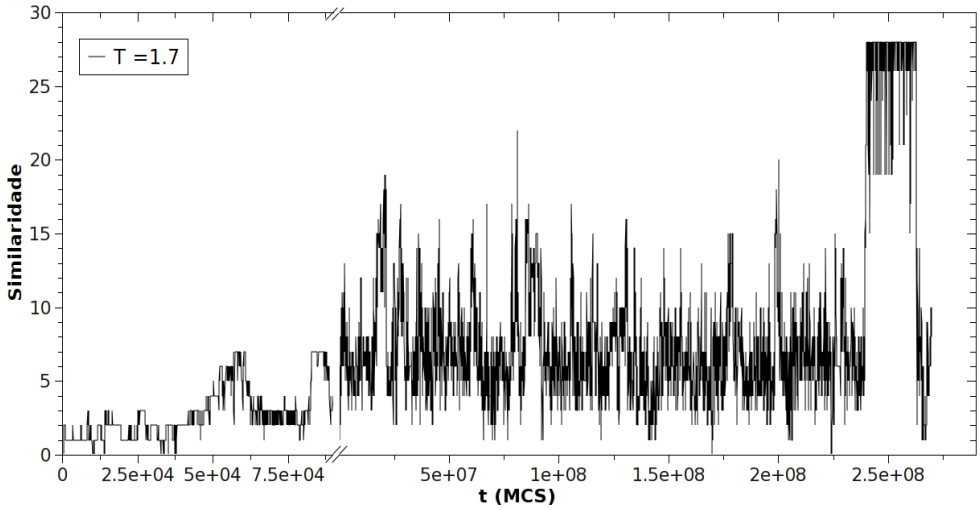


Fig 4.18 – Evolução do parâmetro de sililaridade durante o enovelamento da cadeia S2.15.

a sequência S2.15. Para se saber se a sequência atinge o estado nativo, é necessário analisar o comportamento destas três quantidades. Tal qual as cadeias do modelo de 20 letras, pode-se observar que o número de contatos nativos atinge o valor 28 um número elevado de vezes durante o processo enovelamento. Isto quer dizer que a cadeia assumiu a conformação cúbica várias vezes. Mas ao se verificar a energia nestes instantes, constata-se que seus valores não são os valores mínimos. Assim, não são as conformações nativas. Ainda que seja um fato curioso, não há nada de errado nisso. O que não poderia ocorrer, seria uma dessas conformações compactas ter a mesma energia que o estado nativo e o parâmetro de similaridade ser diferente de 28. Neste caso a cadeia não teria propriedades de proteína.

Encerrando a discussão, o comportamento das cadeias com 15 monômeros é praticamente o mesmo que o apresentado pelas cadeias do modelo com 20 letras. Neste caso do modelo de 15 letras, todas as sequências foram testadas sistematicamente para confirmar se as cadeias se enovelam para o estado nativo desenhado. Assim, excetuando os casos em que há frustrações, as cadeias estudadas poderiam ser candidatas a proteínas..

Capítulo V-CONCLUSÃO

Neste trabalho, foi proposto um método numérico para selecionar sequências de aminoácidos e colocá-las em uma estrutura alvo com a finalidade de desenhar cadeias protéicas. O objetivo é sintetizar uma cadeia que tenha a estrutura alvo como o estado nativo. A idéia é simples. Conhecido esqueleto de uma proteína em seu estado nativo e conhecida a composição de aminoácidos da cadeia, o método proposto neste trabalho permite selecionar sequências de aminoácidos que podem ser dispostos ao longo do esqueleto e sintetizar uma proteína. O método é baseado no método de Wang e Landau [11], introduzido inicialmente para estudar transições de fase em sistemas magnéticos. Ao longo deste trabalho, procurou-se demonstrar a eficiência do método proposto neste trabalho, aplicando-o para sintetizar cadeias poliméricas compostas por 27 monômeros em modelo de rede. Todas as cadeias com as sequências selecionadas com o método satisfizeram as condições para que estas cadeias pudessem ser consideradas candidatas a proteínas.

Estas cadeias foram submetidas a uma série sistemática de testes envolvendo cálculo de grandezas termodinâmicas assim como a análise da cinética de enovelamento destas cadeias. As cadeias desenhadas apresentaram os resultados esperados para uma proteína. As grandezas termodinâmicas calculadas revelaram as transições de fase esperadas (entendida para sistemas finitos) de cadeia polimérica, e permitiram identificar as fases ordenadas e desordenadas de cada sequência. Da cinética de enovelamento, pôde-se determinar a evolução

temporal da energia, número de contatos nativos e do parâmetro de similaridade e como esperado, com elas caracterizar todas as cadeias desenhadas como sendo cadeias protéicas.

Sob ponto de vista de processamento numérico, o método também mostrou-se eficiente. Todos os cálculos deste trabalho, necessários para selecionar sequências, foram efetuados em um computador pessoal comum, com processador pentium IV. Para os modelos de rede com 27 monômeros e 20 tipos de aminoácidos, os resultados podem ser obtidos rapidamente, em questão de minutos.

Concluindo, para o estudo completo de se “desenhar” uma proteína, é preciso utilizar o método proposto neste trabalho apenas como uma ferramenta de sequenciar aminoácidos. Mas para caracterização da cadeia composta por essas sequências, são precisos testes sistemáticos e a utilização de outras técnicas. Ainda assim, trata-se de um método numérico bastante eficiente, demanda computacional aceitável, e sugere boas perspectivas de sucesso em estudos de cadeias mais realistas.

SUGESTÕES PARA TRABALHOS FUTUROS

- Durante a seleção de sequências para o modelo de 20 letras, pôde-se observar que determinados aminoácidos sempre ocupam alguns sítios preferenciais da estrutura alvo. Assim, uma sugestão é estudar qual a relação que este fato tem com a estabilidade da estrutura nativa e mutações.
- Tentar generalizar e implementar o método para modelos contínuos. Ainda que ambiciosa, é bastante atraente a idéia desenhar modelos de proteínas mais realistas. Mesmo que sejam cadeias curtas.

Apêndice

A.1. Mecânica Estatística

O objetivo da mecânica estatística é estudar sistemas compostos de diversos graus de liberdade [41-43] derivando propriedades macroscópicas, expressas em termos de quantidades termodinâmicas, a partir do comportamento microscópico das partículas constituintes do sistema. De modo geral é possível escrever as equações de movimento dessas partículas, porém, a quantidade dessas equações necessárias para descrever todo o sistema, torna impraticável sua resolução.

O interesse deste trabalho é descrever um sistema a partir mecânica estatística. Isso pode ser feito por meio da chamada função de partição do sistema. Esta contém toda a informação essencial sobre o sistema considerado. De modo geral, a função de partição para um sistema clássico pode ser descrita como segue:

$$Z = \sum_{i=1}^{\Omega} e^{\frac{-H_i}{k_B T}} \qquad \text{(A.1.1)}$$

onde H é a Hamiltoniana, k_B é a constante de Boltzmann, T é a temperatura e Ω são os estados acessíveis do sistema. Observa-se a dependência de Z com o tamanho do sistema e com o número de graus de liberdade por partículas na somatória da expressão (A.1.1). Para sistemas com poucas partículas interagentes, a função de partição pode ser escrita de forma exata e, conseqüentemente, as propriedades podem ser calculadas de forma direta.

A partir da função de partição é possível calcular a probabilidade de um determinado estado do sistema ocorrer. A probabilidade do sistema estar em uma configuração ou estado i é

$$p_i = \frac{e^{-H_i/k_B T}}{Z}. \qquad \text{(A.1.2)}$$

É importante mencionar a possibilidade de estabelecer uma conexão direta entre a função de partição e as quantidades termodinâmicas. Isso é possível diferenciando a energia livre de Helmholtz, que pode ser escrita por

$$F = -k_B T \ln Z. \qquad \text{(A.1.3)}$$

Esta relação provê a conexão entre a mecânica estatística e a termodinâmica. A obtenção da energia interna de um sistema através da energia livre se dá via

$$U = -T^2 \frac{\partial (F/T)}{\partial T}. \qquad \text{(A.1.4)}$$

Muitas vezes é conveniente trocar variáveis extensivas tais como S (entropia), V (volume), N (número de moles) etc. por suas conjugadas, variáveis intensivas. Para isso, existem potenciais termodinâmicos definidos por transformadas de Legendre da energia interna.

Dentre outros:

$$F = U - TS\ , \tag{A.1.5}$$

$$H = U + PV\ , \tag{A.1.6}$$

$$G = U - TS + PV\ , \tag{A.1.7}$$

onde F e a energia livre de Helmholtz, H e a entalpia e G a energia livre de Gibbs. Em particular, as quantidade F e U serão calculadas mais adiante a partir do método de Wang-Landau.

A.2. O método de Wolynes para o problema do design

Wolynes [54] faz uso da análise combinatória para o design de proteínas. Estes podem revelar princípios generalizados sobre as forças que estabilizam estruturas de proteínas.

Uma aproximação muito usada em física da matéria condensada é a teoria do campo médio [55-57], onde considera-se as energias médias locais associadas com cada sítio que é determinado pelo ambiente local do sítio. A teoria de campo médio têm uma aplicação extensiva ao problema do enovelamento de proteínas e a exploração de conformações. Aqui as variáveis internas do sistema não são estados conformacionais dos monômeros mas, ao invés, o tipo de aminoácido que está presente em cada posição da sequência. A teoria fornece um modo de estimar não apenas o número de sequências para um dada energia total mas também a probabilidade de que cada posição da sequência seja ocupada por um dado tipo de monômero. A teoria fornece meio convenientes para calcular diferentes estratégias de design combinatórias, onde pode-se comparar os efeitos de mudar o padrão de resíduos e a estrutura alvo.

A.2.1 Teoria para sequências compatíveis com a estrutura escolhida

Seja E a energia de uma sequência quando esta assume a conformação alvo ou enovelada σ(E) é o número de sequências apresentando energia E na estrutura escolhida, e σ_s é o número total de sequências. Ambos σ(E) e σ_s satisfazem quaisquer restrições a respeito ao número total de cada tipo de monômero ou as identidades de monômeros em sítios particulares.

$$S(E)=k_B \ln \Omega_s(E). \qquad \text{(A.2.1)}$$

A entropia de sequências S(E) é definida na forma que é exatamente análoga à equação de Boltzmann para a entropia.

Ao calcular o número de sequências de energia particular, foca-se na estimativa de S(E), aproveitando-se as ferramentas da mecânica estatística. Para baixas energias, S(E) é uma função crescente de E. Como E cresce, a frustração presente na estrutura enovelada também aumenta. Como o número de interações não favoráveis aumenta, há mais maneiras de distribuir os monômeros, e a entropia de sequência aumenta.

Para encontrar a probabilidade de encontrar uma sequência com energia menor que E na estrutura alvo, integra-se o σ(E) até a energia encolhida E.

$$f(E)=\int_{E=-\infty}^{E} dE' \Omega_s(E')/\Omega_s. \qquad \text{(A.2.2)}$$

Alternativamente, se as energia permitidas são discretas, como num modelo de redes, soma-se sobre as energia

$$f(E)=\sum_{E=-\infty}^{E} \Omega_s(E')/\Omega_s. \qquad \text{(A.2.3)}$$

Geralmente, para sistemas que englobam muitos sítios interagentes, calcular a entropia microcanônica exatamente não é algo trivial. Se a energia do sistema é a soma das energias individuais de cada componente, portanto, podem ser obtidas diretamente. No caso do design, os elementos interagentes são resíduos individuais, como ocorrem na estrutura alvo. Funções de energia efetivas, dependem apenas da identidade e localização de um resíduo numa

estrutura particular. Portanto estes tipos de funções de energia são puramente da forma de energia de um corpo. Para funções mais complicadas envolvendo muitos corpos, usa-se a teoria de campo médio. Para tal função, a energia de uma sequência particular é dada por

$$E=\sum_{i=1}^{N}\epsilon_i(\alpha_i). \qquad \text{(A.2.4)}$$

Aqui N é o comprimento do polímero em unidades de monômeros, e ε_i é a energia efetiva de um corpo no sítio i na estrutura. A sequência é denotada por uma lista ordenada de tipos de monômeros $\{\alpha_1...\alpha_N\}$, onde α_i é o tipo de monômero presente na posição da sequência i. O índice i rotula tanto a posição de um monômero na sequência quanto sua posição tridimensional na estrutura alvo.

A entropia de sequência é obtida ao maximizar a entropia S(E) com respeito à quaisquer parâmetros sem restrição. Desde de que a energia interna pode ser escrita como uma soma de termos individuais de um corpo, S(E) pode ser expressa como

$$S(E)/k_B=-\sum_{i=1}^{N}\sum_{\alpha_i=1}^{m}w_i(\alpha_i)\ln w_i(\alpha_i). \qquad \text{(A.2.5)}$$

Onde k_B é a constante de Boltzmann, e m é o número de tipos de monômeros. Nos casos de peptídeos, m é apenas o número de aminoácidos diferentes usados para sintetizar as sequências. Aqui wi(α) é a probabilidade que o resíduo do tipo α esteja na posição i na estrutura. A entropia de sequência S(E) é maximizada mediante à restrição de que a energia total seja conservada. Incorpora-se a restrição ao restringir o valor da energia interna

$$U=\sum_{i=1}^{N}\sum_{\alpha_i=1}^{m}\epsilon_i(\alpha_i)w_i(\alpha_i). \qquad \text{(A.2.6)}$$

Uma restrição adicional é que a soma das probabilidades sobre cada sítio é um, ou seja, cada sítio deve ser ocupado por pelo menos um monômero.

$$1=\sum_{\alpha=1}^{m}w_i(\alpha). \qquad \text{(A.2.7)}$$

Se um monômero α é impedido de ocupar o sítio i, então tem-se restrições adicionais da forma $w_i(\alpha)=0$, se α não é permitido no sítio i. Há m-m_i restrições desse tipo para cada posição i da sequência, onde m_i é o número de tipos de resíduos permitidos na posição i.

Este é o caso em que o número de tipo de monômeros é o mesmo para cada sequência. A localização de cada monômeros na sequência pode, entretanto, variar livremente. Seja n(α) o número total de monômeros do tipo α. Lembrando que $w_i(\alpha)$ é a probabilidade de que um sítio i tenha um tipo de resíduo α. Para cada sequência a soma das probabilidades individuais sobre as posições da sequências deve ser igual ao número de monômeros daquele tipo.

$$n(\alpha)=\sum_{i=1}^{N} w_i(\alpha). \qquad \text{(A.2.8)}$$

Lembrando que se um resíduo do tipo α não é permitido no sítio i $w_i(\alpha)=0$.

Como anteriormente, maximiza-se a entropia com respeito à restrições da conservação de energia e da probabilidade, mas agora introduz-se mais uma restrição a de que o sistema apresenta composição constante. Para cada tipo de monômero, acrescenta-se um correspondende multiplicador de Lagrange $\beta\mu_\alpha$.

$$S(E)/k_B=\beta U+\sum_{i=1}^{N}\ln z_i+\beta\sum_{\alpha=1}^{m}\mu_\alpha n(\alpha), \qquad \text{(A.2.9)}$$

onde

$$z_i=\sum_{\alpha=1}^{m}\delta_i(\alpha)\exp(-\beta(\epsilon_i(\alpha)+\mu_\alpha)). \qquad \text{(A.2.10)}$$

E as probabilidades individuais se tornam

$$w_i(\alpha)=z^{-1}\sum_{\alpha=1}^{m}\delta_i(\alpha)\exp(-\beta(\epsilon_i(\alpha)+\mu_\alpha)). \qquad \text{(A.2.11)}$$

Usando a restrição da composição fixa, pode-se escrever uma equação que μ_α deve satisfazer.

$$\beta\mu_\alpha=\ln\sum_{i=1}^{N} z_i^{-1}\exp(-\beta\epsilon_i(\alpha))-\ln n(\alpha). \qquad \text{(A.2.12)}$$

Pode construir uma analogia com a termodinâmica estatística. β^{-1} tem as propriedades de uma temperatura efetiva. A composição constante requer que o número de cada tipo de monômero seja constante. As variáveis termodinâmicas conjugadas dos números de cada componente são seus potenciais efetivos μ_α. O potencial químico efetivo μ_α é uma energia livre efetiva por partículas para cada tipo de monômero. Para uma composição total, alguns tipos de monômeros podem ser mais favoráveis que outros na média. Os potenciais quìmicos efetivos reajustam para manter a composição constante em tais casos.

Quando o número de tais tipos de monômeros é predeterminado, o número total de sequências é dado simplesmente pelo coeficiente multicanônico correspondente.

$$\Omega_s = N!/\prod_{\alpha=1}^{m} n(\alpha), \qquad \text{(A.2.13)}$$

onde N é, como já dito, o comprimento da cadeia em unidades monoméricas, m é o número total de monômeros disponíveis, e n(α) é o número de resíduos do tipo α.

A entropia de sequências microcanônica S(E) é obtida a partir das eqs. 3.13. Essa entropia de sequência é normalizada de acordo com o número total de possíveis sequências. Em termos termodinâmicos, é o equivalente à escolher uma temperatura para o sistema de tal forma que a energia livre seja igual à energia em interesse. Para uma dada estrutura alvo, a teoria retorna a probabilidade de que um sítio i na sequência seja ocupado por um tipo de monômero específico numa dada energia. A teoria também fornece o logaritmo do número de sequências que têm uma energia particular na estrutura alvo.

A teoria pode ser usada para calcular uma quantidade que reflete a variabilidade mutacional de cada sítio, esta quantidade é conhecida como entropia local de sequências

$$s_i = -\sum_{\alpha=1}^{m} w_i(\alpha)\ln w_i(\alpha), \qquad \text{(A.2.14)}$$

onde a soma é sobre todos os resíduos no sítio i. Dependendo da estrutura, alguns sítios são mais prováveis a ter um tipo particular de monômero que as demais. Para um dado valor de energia E, a entropia local mede o quanto mutações são suscetíveis numa posição particular da sequência. Nota-se que se apenas um tipo é permitido no sítio i, então s_i=0. Se todos os

tipos são permitidos, no sitio i, então $s_i=\ln(m_i)$. A uma dada energia E, um sítio tendo um valor pequeno de si tem um tipo de monômero que é mais conservado ao longo de diferentes sequências.

Referências

1. CHAMPE P.C., HARVEY R.A., FERRIER D.F., *Bioquímica Ilustrada*, 3ª edição, editora Artmed, 2005,capítulo 2, p. 13-25.

2. CHAN H.S.; DILL K.A., The Protein Folding Problem, *Physics Today* 46, 24-32, fev. 1993.

3. FRAUENFELDER H.; WOLYNES P. G., Biomolecules: Where the Physics of complexity and simplicity meet, *Physics Today* 47, 58-64, fev. 1994.

4. PANDE V.S., GROSBERG A.Y. e TANAKA T.,Heteropolymer freezing and design: Towards physical models of protein folding, *Reviews of Modern Physics*, vol. 72, 259-314, Jan. 2000.

5. SHAKHNOVICH E.I., Protein Design: a perspective from simple tractable models, *Folding & Design* vol. 3 N. 3, R45-R58, Jun. 1998.

6. SHAKHNOVICH E.I.; GUTIN A.M., A new approach to the design of stable proteins, *Protein Engeneering* vol.6 N. 8, 793-800 Jun. 1993.

7. SHAKHNOVICH E.I. e GUTIN A.M., Engeneering of stable and fast-folding sequences of model proteins, *Proceedings of National Academy of Sciences*, vol. 90, 7195-7199 Ago.1993.

8. PEPYS B.M. et al., Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis, *Nature* 417, 254-259, Mai. 2002.

9. HILL A. F.; DESBRUSLAIS M.; JOINER S.; SIDLE K. C. L.; GOWLAND I.;

COLLINGE J.; DOEY L. J. ; LANTOS P., The same prion strain causes vCJD and BSE, *Nature* 389, 448-450, Out. 1997.

10. BROGLIA R.A.; TIANA G.; PROVASI D., Simple models of protein folding and of non-conventional drug design, *Journal of Physics : Condensed Matter* N. 16, R111 – R144, Jan. 2004.

11. WANG F. e LANDAU D.P., Efficient, Multiple range random walk algortihm to calculate the density of stated, *Physical Review Letters*, vol. 86 N.10, 2050-2053, Mar. 2001.

12. DILL K. A.; BROMBERG S.; YUE K.; FIEBIG K. M.; YEE D. P.; THOMAS P. D.; CHAN H. S., Principles of protein folding - A perspective from simple exact models, *Protein Science* vol. 4, 561-602, Abr. 1995.

13. SHAKHNOVICH E.I. e GUTIN A.M.,Formation of unique structure in polypeptide-chains theorical investigation with the aid of a replica approach, *Biophysical Chemistry*, vol. 34, 187-199, Nov.1989.

14. SHAKHNOVICH E.I. e GUTIN A.M., Implications of thermodynamics of protein folding for evolution of primary sequences, *Nature* 346, 773-775, Ago. 1990.

15. SCHULZ G.E., SCHIRMER R.H., *Principles of protein structure*, 1ª edição, editora Halliday Litograph, 1979.

16. SUSSMAN J. L.; LIN D.; JIANG J.; MANNING N. O.; PRILLUSKY J.; RITTER O.; ABOLA E. E., Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules, *Acta Crystallographica Section D, v*ol. 54, N 1, 1078-1084, Jul. 1998.

17. FRENSDORFF H. K.; WATSON M. T.; KAUZMANN W., The kinetics of protein denaturation, *Journal of the American Chemical Society*, vol . 75, 5152-5172, Nov. 1953.

18. LEVINTHAL C., *Journal of Chemical Physics*, vol .65, 44, 1968.

19. LANDAU D.P. e BINDER K., *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press.

20. BINDER K., *The Monte Carlo Method in Condensed Matter Physics*, 2ns edn, Topics in Applied Physics, vol 71 (Springer, 1991).

21. METROPOLIS N., ROSENBLUTH A.W., ROSENBLUTH M.N., TELLER A.N., TELLER E., Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, vol. 21, 1087-1093, Dez. 1953.

22. MEYN S. P.; TWEEDIE R. I. *Markov chains and stochastic stability*, Springer, 1993.

23. MYAZAWA S. e JERNIGAN R.L., Residue – residue potencials with a favorable contact pair term and unfavorable high packing density term, for simulation and threading, *Journal of Molecular Biology*, vol. 256, 623-644, Mar. 1996.

24. SOCCI N.D. e ONUCHIC J.N., Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique, *Journal of Chemical Physics*, vol. 103, 4732-4745 ,Set. 1995.

25. FERRENBERG A.M.; SWENDSEN R.H., Optimized Monte Carlo data analysis, *Physical Review Letters*, vol. 63, 1195-1198, Fev. 1989.

26. HAO M.; SCHERAGA H.A., Monte Carlo Simulation of a First-Order Transition for Protein Folding, *Journal of Physics Chemistry*, vol. 98, 4940-4948, Mai. 1994.

27. LEE J., New Monte Carlo algorithm: Entropic sampling, *Physical Review Letters*, vol. 71, 211-214, Abr. 1993.

28. LEE L.W. e WANG J.S., Flat histogram simulation of lattice polymer systems, *Physical Review E*, vol. 64, 056112, Out. 2001.

29. DE OLIVEIRA P.M.C., PENNA T.J. E HERMANN H.J., Broad Histogram Method, *Brazilian Journal of Physics*, vol. 26, 677-683, Dez. 1996.

30. BACHMAN M. e JANKE W., Thermodynamics of lattice heteropolymers, *Journal of Chemical Physics*, vol. 120, 6779 - 6792, Abr. 2004.

31. CHAN H.S. e DILL K.A., Polymer Principles in Protein Structure and Stability, *Annual Review of Biophysics and Biophysical Chemistry*, vol. 20, 447-490, Jun. 1991.

32. MILLER R., DANKO C. A., FASOLKA M.J., BALAZ A.C., CHAN H.S., e DILL K.A, Folding kinetics of proteins and copolymers, *Journal of Chemical Physics*, vol. 96, 768-781, Jan. 1992.

33. CHAN H.S. e DILL K.A., Energy landscapes and the collapse dynamics of homopolymers, *Journal of Chemical Physics*, vol. 99, 2116-2128, Ago. 1993.

34. CAMACHO C. J. e THIRUMALAI D.,Kinetics and thermodynamics of folding in model proteins, *Proc. Natl. Acad. Sci. USA* vol. 90, 6369-6372, Jul. 1993.

35. SOCCI N.D. e ONUCHIC J.N., Kinetic and thermodynamic analysis of proteinlike heteropolymers : Monte Carlo histogram technique, *Journal of Chemical Physics*, vol. 103, 4732-4745, Set. 1995.

36. HUANG K. *Statistical Mechanics*, John Wiley & Sons, NY (1966).

37. BRYNGELSON J.D.; WOLYNES P.G.,Spin glasses and the statistical mechanics of

protein folding, *Proceedings of the National Academy of Sciences USA*, vol. 84, 7524 - 7528, Nov. 1987.

38. BINDER K.; YOUNG A. P., Spin glasses: Experimental facts, theoretical concepts, and open questions, *Review of Modern Physics*, vol. 58, 801-976, Out. 1989.

39. ZHOU C.; BHATT R. N., Understanding and improving the Wang-Landau algorithm, *Physical Review E*, vol. 72, 025701 - 025705, Ago. 2005.

40. WANG F.; LANDAU P., Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram, *Physical Review E*, vol. 64, 056101-056117, Out. 2001.

41. PATHIA R.K., *Statistical Mechanics*, Butterworth-Heinemann, 1996.

42. REIF F., *Fundamentals os Statistical and Thermal Physics*, New York : McGraw-Hill-Book, 1965.

43. SALINAS R.A., *Introdução à Mecânica Estatística*, Edusp, 1999.

44. SHAKHNOVICH E. I., Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet, *Chemical Reviews*, vol. 106, 1559-1568, Mai. 2006.

45. SHAKHNOVICH E. I. ,Proteins with selected sequence fold into a unique native conformation, *Physical Review Letters*, vol. *72, N. 24*, 3907-3911, Jun. 1994.

46. SALI A. ; SHAKHNOVICH E. I.; KARPLUS M., How does a protein fold?, *Nature* vol. 369, 248-251, Mai. 1994.

47. BEIG F. B., *Termodinâmica do enovelamento de cadeias heteropoliméricas através do algoritmo de Wang-Landau.* 2006. 90f. Dissertação (Mestrado em Física Aplicada) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro. 2006.

48. LI H., TANG C. E WINGREEN N.S.,Designability of protein structure: A lattice study using the Miyazawa-Jernigan matrix, *Proteins: Structure, Function and Genetics*, vol. 49, 403-412, Jul. 2002.

49. LI H., HELLING R., TANG C., WINGREEN N., Emergence of Preferred Structures in a Simple Model of Protein Folding, *Science* vol. 273, 666-669, Ago. 1996.

50. LI H., TANG C., WINGREEN N., Are proteins fold atypical?, *Proceedings of National Academy of Sciences U.S.A.*, vol. 95, 4987-4990, Abr. 1998.

51. MÉLIN R., LI H., TANG C., WINGREEN N., Designability, thermodynamic stability, and dynamics in protein folding: A lattice model study, *Journal of Chemical Physics*, vol. 110, 1252-1263, Jan. 1999.

52. SHAKHNOVICH E. I., Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry and Biology Meet, *Chemical Reviews*; vol 106, 1559-1588, Mai 2006.

53. SHAKHNOVICH E. I., FINKELSTEIN A. V., Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers*.; vol. 281667–1680, Out.1989.

54. SAVEN J.G.; WOLYNES P.G., Statistical Mechanics of the combinatorial synthesis and analysis of folding Macromolecules, *Journal of Physical Chemistry B,* N.101, 8375-8389, Jul. 1997.

55. FINKELSTEIN A. E REVA B., A search for the most stable folds of protein chai ns,*Nature*, vol. 351, 497-500, Jun. 1991.

56. KOECH P. E DELARUE M., Mean-field minimization methods for biological macromolecules , *Current Opinion in Structural Biology*, vol. 6, 222-226, Abr. 1996.

57. REVA B.A., FILKENSTEIN A.V., RYKUNOV D.S., OLSON A.J., Building self-avoiding lattice models of proteins using a self-consistent field optimization, *Proteins* vol. 26, 1-8, Set. 1996.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo