

Ricardo Hildebrand Theodoro da Silva

**Comportamento Termodinâmico e a Cinética
de Mutações de Proteínas Otimamente
desenhadas em modelos de rede**



1910001079



COMPORTAMENTO TERMODINÂMICO E A
CINÉTICA DE MUTAÇÕES DE PROTEÍNAS
OTIMAMENTE DESENHADAS EM
MODELOS DE REDE

Ricardo Hildebrand Theodoro da Silva

Dissertação de Mestrado
Pós-Graduação em Biofísica Molecular

Comportamento termodinâmico e a cinética de mutações de
proteínas otimamente desenhadas em modelo de rede

Ricardo Hildebrand Theodoro da Silva

Silva, Ricardo Hildebrand Theodoro da.

Comportamento termodinâmico e a cinética de mutações de
proteínas otimamente desenhadas em modelo de rede / Ricardo
Hildebrand Theodoro da Silva – São José do Rio Preto : [s.n.], 2002
51 f. : il. ; 30 cm.

Orientador: Vitor Barbanti Pereira Leite
Dissertação (mestrado) – Universidade Estadual Paulista. Instituto
de Biociências, Letras e Ciências Exatas

1. Proteínas. 2. Biopolímeros. 3. Proteínas – Mutações. I. Leite,
Vitor Barbanti Pereira II. Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.112

Comportamento Termodinâmico e a Cinética de Mutações de Proteínas Otimamente Desenhadas em Modelo de Rede

Ricardo Hildebrand Theodoro da Silva

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto, São Paulo, para a obtenção do título de Mestre em Biofísica Molecular.

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

Co-orientador: Prof. Dr. Jorge Chahine



T 1.079/02

São José do Rio Preto

Terminado



Agradecimentos

À minha esposa Lúcia Paula, que muito me ajudou na preparação do trabalho.

À minha amiga Tânia, pelas dicas e sugestões na hora de redigir.
À minha colega de sala Luciana e Silvana, por me auxiliarem nos momentos mais difíceis, e por me ajudarem sempre que pude.

À minha amiga Lúcia P. que, ao discursar sobre o trabalho, me ajudou a entender os conceitos do trabalho.

À Faculdade, por ajudar sempre que pude.

Às colegas de pós-graduação: Rosângela, Alexandre, Juliana, Valquíria, Bruna, Larissa, Mariana, Flávia, Tânia, Sérgio, Leonardo, Roberto, Ilvânia, Fabiana, Cássia e André, que de alguma forma me ajudaram no trabalho.

À minha amiga da Faculdade de São Carlos.

Às professoras: José Roberto, João e Rosângela, que me ajudaram no trabalho.

Às professoras: José Roberto, João e Rosângela, que me ajudaram no trabalho.

Às professoras da pós-graduação: José Roberto, que me ajudou a entender o trabalho e a escrever o trabalho.

À minha amiga da Faculdade de São Carlos.

Às professoras: José Roberto, João e Rosângela, que me ajudaram no trabalho.

Às professoras da pós-graduação: José Roberto, que me ajudou a entender o trabalho e a escrever o trabalho.

Dedico este trabalho
à minha esposa, Lucimara
e à minha filha Sofia

Agradecimentos

Ao meu amigo Luís Paulo, que muito me ajudou no aprendizado da linguagem C.

Ao meu amigo Plínio, pelas discussões na hora do café.

Às minhas colegas de sala Luciana e Sabrina, por me suportarem nos momentos mais difíceis, e por me ajudarem sempre que puderam.

Ao meu colega Luíz F. com as discussões muito frutíferas sobre enovelamento de proteína, nos corredores do departamento.

À Fernanda, por ajudar sempre que pode.

Aos colegas de pós-graduação Renato, Alexandre, Janaína, Carla, Valmir, Bernadete, Tereza, Fátima, Flávio, Fausto, Sérgio, Lessandra, Ricardo, Hamilton, Rubens, Cecília e Neide, que de alguma forma sempre me ajudaram.

Aos meus amigos da Federal de São Carlos.

Aos professores, José Roberto, João e Augusto, com os quais fiz disciplina.

Ao ex-coordenador da pós-graduação José Roberto, que me recebeu muito bem quando cheguei de São Carlos para conhecer o departamento.

Ao atual coordenador de pós graduação João.

Ao Prof. Dr. Jorge Chahine pela co-orientação, paciência nas discussões sobre este trabalho.



Ao Prof. Dr. Vitor Barbanti Pereira Leite, pela orientação, incentivo e paciência na elaboração deste trabalho, e todos professores e funcionários do departamento de Física que de alguma forma contribuíram para a elaboração deste trabalho.

Aos meus pais e meu irmão que sempre me apoiam.

Aos amigos pós-graduandos.

À CAPES, pelo auxílio financeiro.

A Deus, por tudo.



Resumo

Neste trabalho foram avaliados os efeitos termodinâmicos e cinéticos que pequenas mutações em uma proteína otimamente desenhada, de 27 monômeros em um modelo de rede cúbico para os regimes de alta e baixa hidrofobicidade. Este trabalho mostra que, pequenas mutações provocam uma diminuição do valor da temperatura crítica de enovelamento, bem como um aumento do tempo de enovelamento. Estes efeitos foram verificados serem mais acentuados no regime de baixa hidrofobicidade.

Palavras-chave: mutação, modelo de rede, enovelamento de proteína



Sumário

Abstract

In this work it was evaluated the thermodynamic and kinetic effects of small mutations in a well designed protein of 27 monomers in a cubic lattice model for high and low hydrophobicity. This work shows that small mutations can lead to a decrease of critical temperature of folding value, as well as an increase of folding time. Was verified that these effects are more accentuated in low hydrophobicity.

Keywords: mutation, lattice model, protein folding



Sumário

1	Introdução	2
1.1	Importância e Bioquímica	2
1.2	Enovelamento de Proteína	4
1.2.1	Introdução Histórica	4
1.2.2	Interações do Enovelamento	7
1.2.3	Modelos de Vidros de Spins	9
1.2.4	Modelo de Energia Aleatória	11
1.2.5	Modelo Bryngelson-Wolynes	13
1.2.6	Funil	18
1.2.7	Motivação	21
2	Modelo de proteína na rede	25
2.1	Estruturas compactas e Energética	25
3	Resultados	34
4	Conclusão	45
	Referências Bibliográficas	48

Capítulo 1

Introdução

1.1 Importância e Bioquímica

As proteínas são componentes essenciais aos sistemas vivos. Elas estão presentes na estrutura de vários organismos e realizam a maioria das tarefas necessárias para manter tais organismos vivos. As proteínas são formadas por longas cadeias de aminoácidos unidas por ligações peptídicas. Os aminoácidos são as unidades básicas de uma proteína. Todas as proteínas, de todas as espécies, da bactéria mais simples ao homem são construídas do mesmo conjunto de vinte aminoácidos que ocorrem naturalmente [1]. A proteína é um polímero, constituído pelos vinte tipos de aminoácidos encontrados na natureza. O polímero pode ser pensado pictoricamente como um colar, onde as contas deste colar são os aminoácidos, que podem ser de diferentes tamanhos e formas [2]. Os vinte aminoácidos que ocorrem na natureza basicamente são constituídos por um carbono que está na cadeia principal da proteína denominado de C_α , que possui ligados covalentemente a ele, um grupo amina, um grupo carboxílico um átomo de hidrogênio e uma cadeia lateral ou um radical **R**. O que diferencia um aminoácido do outro é a sua cadeia lateral ou radical **R**, que varia em tamanho, forma, car-



ga líquida, reatividade química e capacidade de realizar pontes de hidrogênio e pontes covalentes [1]. Os vinte tipos de aminoácidos que ocorrem naturalmente podem ser genericamente classificados em polares e hidrofóbicos. Na proteína os aminoácidos estão ligados entre si por ligações peptídicas. As ligações peptídicas são formadas através da reação de condensação entre um grupo amina NH_3^+ , e outro do $-O$ do grupo carboxílico, liberando uma molécula de água. A ligação peptídica possui uma estrutura planar pelo caráter da dupla ligação devido à ressonância da ligação $C=O$ e $N-C$, conferindo rigidez à sequência $C_\alpha-N-C-C_\alpha$, o conjunto de sequências $C_\alpha-N-C-C_\alpha$ compõem o esqueleto da proteína [5].

Existem quatro níveis de estrutura que freqüentemente são citadas em discussões de arquitetura de proteína. A **estrutura primária** consiste na sequência de aminoácidos da sua cadeia polipeptídica. A **estrutura secundária** das proteínas inclui os padrões estruturais locais regulares de proteínas, como as alfa hélices e as folhas beta pregueadas e as voltas. A **estrutura terciária** nos informa sobre a topologia global do arranjo espacial dos elementos estruturais secundários, especificando as posições de cada átomo, incluindo as das cadeias laterais. A maioria das proteínas, particularmente as com massas maiores do que 100kD são constituídas por mais de uma cadeia polipeptídica que associam-se com uma geometria específica. O arranjo especial dessas subunidades é conhecido como **estrutura quaternária**[5].

A maioria das transformações químicas em sistemas vivos são catalisadas por proteínas, que estão envolvidas na execução de tarefas específicas [6]. As proteínas para se tornarem biologicamente ativas devem, a partir de uma dada estrutura, primária ou sequência da proteína, atingir uma conformação particular (nativa). A conformação de uma proteína biologicamente ativa tem que ter certas características importantes. A conformação deve ser única e estável do ponto de vista termodinâmico, além do fato de que esta conformação deve ser atingida rapidamente para a proteína desempenhar sua tarefa específica [5]. As questões que se tornam relevantes a partir destas características são: Dada a sequência



de aminoácidos como a proteína atinge sua conformação nativa ? Quais funções são definidas por uma configuração tridimensional particular para cada espécie de proteína? E quais os mecanismos que existem para que isto ocorra?

1.2 Enovelamento de Proteína

1.2.1 Introdução Histórica

Pauling e seu grupo durante 15 anos estudaram pequenas moléculas compostas de pontes de hidrogênio, e mostraram em 1951 [7], que tais moléculas têm certas estruturas locais regulares, as quais eles denominaram de α (alfa) hélice e folha β (beta). Watson e Crick [8] em seu trabalho de 1953 conseguiram prever as estruturas globais dos ácidos nucleicos, conseguindo elucidar o processo de hereditariedade. A importância deste trabalho vem do fato do paradigma DNA $\rightarrow$ RNA $\rightarrow$ proteína, que é o seguinte: A informação contida em uma fita de DNA é transcrita em uma fita de RNA, que por sua vez é traduzido como uma proteína específica. Estas duas publicações que iniciaram a biologia molecular fizeram com que os cientistas compreendessem que os processos fundamentais da biologia molecular deveriam ser elucidados por uma abordagem físico-química [2].

Utilizando uma abordagem físico-química Christian B. Anfinsen no início dos anos 60 publicou um trabalho [9] muito importante para o desenvolvimento do problema do enovelamento de proteína. Neste trabalho utilizando a ribonuclease, uma proteína, Anfinsen media a atividade da proteína para várias condições físico-químicas. Para condições desfavoráveis, a proteína se encontrava em seu estado desnaturado, para condições fisiológicas, a proteína se encontrava em seu estado nativo (quando a proteína desempenha seu papel biológico) [9]. Anfinsen mostrou que não somente as estruturas locais, mas também as estruturas globais, poderiam ser atingidas sem a existência de um maquinário biológico [10]. Tal tra-



balho sugere claramente que as sequências de aminoácidos das proteínas contém toda a informação necessária para a proteína atingir seu estado nativo [11]. Anfinsen mostrou também que o processo, em que a proteína parte de seu estado desnaturado e atinge seu estado nativo é reversível, ou seja, que uma proteína ora pode encontrar-se no estado nativo, ora em seu estado desnaturado. Foi observado também que, quando a proteína se encontrava em seu estado desnaturado, a proteína quebrava cinco pontes de dissulfeto, quando a proteína se renaturava ela formava novamente essas 5 pontes de sulfeto. Através deste resultado Anfinsen então postulou que o estado nativo de uma proteína (conformação esta que a proteína é biologicamente ativa) é o mínimo global dentre suas energias livres acessíveis. A partir daí, entender o modo pelo qual uma sequência unidimensional de aminoácidos de uma proteína atinge uma estrutura tridimensional tem fascinado gerações de biólogos-moleculares, físico-químicos, matemáticos e físicos e é denominado o **Problema do Enovelamento de Proteína** [2].

Depois de formulado o problema do enovelamento de proteína pela hipótese termodinâmica de Anfinsen, os cientistas começaram a tentar compreender os mecanismos pelo qual uma proteína passa até atingir seu estado nativo. O problema antes de mais nada não é de fácil solução, porque as proteínas são sistemas complexos, tornando difícil idealizar um modelo para as proteínas [12]. Além disso o problema tornou-se ainda mais complicado, quando no final dos anos 60 Cyrus Levinthal em [13], levantou uma questão, que depois passou a ser chamado de **Paradoxo de Levinthal** [14]. Levinthal mostrou que o espaço de todas as conformações possíveis de uma proteína é muito grande para que a proteína visite todas conformações [11]. Admitindo que cada um dos N aminoácidos possa assumir γ orientações, então o número de conformações possíveis distintas que uma proteína pode assumir é $M = \gamma^N$ [11]. Se supusermos por simplicidade que temos um polímero com $N = 100$, sendo que cada monômero pode assumir somente duas orientações, então o número total de conformações diferentes é $M = 2^{100} \approx 10^{30}$. Se ainda supusermos que neste espaço de aproximadamente



10^{30} conformações fosse necessário 10^{-12} (pico segundo) para se pesquisar uma única conformação, então o tempo necessário para explorar todo o espaço das conformações da proteína seria de 10^{18} segundos aproximadamente, o que dá mais que 10^{10} anos. Levinthal argumentou com esta estimativa que uma proteína nunca atingiria a conformação de mínimo global de energia visitando todas as conformações possíveis [15]. Usando argumentos semelhantes Levinthal levantou um problema cinético, durante o processo de enovelamento, que deveria ser resolvido. Para solucionar esse problema, Levinthal postulou que a conformação do estado enovelado de uma proteína fosse encontrada por um caminho. Durante mais de 25 anos essa idéia foi dominante na formulação do enovelamento de proteína. Com estes conceitos produziu-se uma grande quantidade de experimentos cinéticos e termodinâmicos, mas de modo geral foi muito difícil relacionar tais resultados a um referencial com alguma consistência estrutural e energética [2].

Uma visão alternativa começou a surgir no início da década de 80. Este ponto de vista sustenta que o mais importante para entender o processo de enovelamento é olharmos para o panorama global da superfície de energia da proteína. Esta visão global do problema é útil se o enovelamento ocorre através da organização de um conjunto de estruturas que pode ser descrito por umas poucas variáveis, e desta forma pensar em uma descrição estatística do relevo de energia. De posse da descrição estatística podemos assim construir ferramentas baseadas na mecânica estatística de sistemas minimamente frustrados, e de transições de fase de sistemas finitos [2].

A teoria do relevo de energia e o conceito do funil estão no centro do referencial teórico necessário para se entender quantitativamente a questão do enovelamento. Esta visão do mecanismo do enovelamento advém de uma alternativa à idéia inicial de que devia existir um caminho para o evento do enovelamento com estados intermediários claramente definidos [2].



1.2.2 Interações do Enovelamento

A forma do relevo de energia é determinada pelas interações relevantes ao processo de enovelamento [4]. Em estudos cristalográficos e de modelos de peptídios Mirsky e Pauling [16] e Eyring e Stearn [17] propuseram que as pontes de hidrogênio fossem as interações dominantes no enovelamento. Walter Kauzmann [18], em um trabalho 1959, conseguiu mostrar que as pontes de hidrogênio não seriam fortes o suficientes para serem dominantes no processo de enovelamento, porque as pontes de hidrogênio das cadeias laterais das proteínas realizadas com as águas vizinhas têm a mesma magnitude das pontes de hidrogênio intracadeias encontradas no estado nativo da proteína. De modo que Kauzmann argumentou que a interação importante para o enovelamento de proteínas estava na tendência de aminoácidos hidrofóbicos associarem-se entre si, e os aminoácidos polares em se associarem com as moléculas de água, tal efeito foi denominado de **efeito hidrofóbico** [19]. Depois de Kauzmann ter publicado seu trabalho sobre o efeito da hidrofobicidade no processo de enovelamento, passou-se a analisar a influência da hidrofobicidade da proteína no enovelamento, e apesar de muitas pesquisas feitas para se compreender as interações hidrofóbicas os detalhes deste tipo de interação ainda não são completamente compreendidos. Em 20 anos de observação da regra de Kauzmann, a visão que emerge é que a não especificidade das interações hidrofóbicas favorecem a formação de estruturas compactas, deixando para as pontes de hidrogênio e as interações locais a responsabilidade pelos detalhes internos da estrutura enovelada e por sua unicidade [4].

Apesar da diversidade das proteínas, elas formam uma classe de moléculas relativamente homogênea, já que todas as proteínas são formadas pelos mesmos vinte aminoácidos que ocorrem naturalmente [3]. Se considerarmos os graus de apolaridade e polaridade dos aminoácidos podemos dividir os vinte tipos diferentes em dois tipos básicos. O tipo hidrofóbico **H** e o tipo polar **P**, então desta forma podemos imaginar que as proteínas são constituídas basicamente de



monômeros hidrofóbicos **H** e monômeros polares **P** (Modelo **HP**). Neste modelo as interações dos monômeros hidrofóbicos com a água são desfavorecidas, e as interações entre os monômeros hidrofóbicos, bem como as interações dos monômeros polares com as moléculas de água são favorecidas. As relações entre esses dois tipos de monômeros bem como a relação de cada tipo de monômero com as moléculas de água são responsáveis pela compactação da cadeia [4].

Se se admitir por simplicidade um modelo **HP**, no enovelamento os monômeros hidrofóbicos tentam minimizar sua área de contato com a água formando um núcleo hidrofóbico causando o colapso da cadeia em estruturas compactas. Entretanto existem menos conformações compactas do que conformações expandidas indicando assim que a entropia das conformações compactas é menor do que a entropia das conformações expandidas. A maior entropia das conformações expandidas tenderá a se opor ao colapso da cadeia em uma estrutura compacta. Neste processo existe uma competição entre a minimização da energia de interação entre os monômeros e a perda de entropia pelo colapso da cadeia da proteína[4].

Quando ocorre a compactação da cadeia da proteína por causa do efeito hidrofóbico, dois monômeros hidrofóbicos que estão distantes na sequência, se aproximam um do outro para formarem o núcleo hidrofóbico e minimizar a energia da cadeia da proteína. Só que estes monômeros distantes podem trazer consigo monômeros polares e por causa da conectividade da cadeia da proteínas, esses monômeros polares competirão posições com os monômeros hidrofóbicos que minimizam a energia, esta competição é denominada de **frustração**. O termo frustração foi introduzido em 1977 por Toulouse [20] para um modelo de vidro de spin para antiferromagnetos. Eles consideraram a interação de três spins. Assumindo que a energia de dois spins antiparalelos, ($\uparrow\downarrow$) é menor do que a interação destes dois spins quando estão paralelos entre si ($\uparrow\uparrow$) ou ($\downarrow\downarrow$). No entanto quando se introduz um terceiro spin junto do par de spins antiparalelos, qualquer orientação que tenha este spin, para cima $\uparrow$ ou para baixo $\downarrow$, não se consegue

minimizar todas as interações do sistema e então diz-se que o sistema é frustrado [12]

1.2.3 Modelos de Vidros de Spins

Vidros de spin são sistemas que podem ser considerados como análogos a sistemas magnéticos, em que as interações entre os momentos magnéticos estão em conflito umas com as outras devido ao “congelamento” estrutural do sistema. Tais sistemas exibem uma transição de vidro em que o sistema se encontra armadilhado em poucas configurações, dizemos que o sistema está na fase vidro quando a temperatura do sistema for $T < T_g$ (que está definido em 1.9). Apenas acima de T_g os tempos de relaxação tem um largo espectro. Quando $T \rightarrow T_g$ alguns tempos de relaxação começam a ser muitas ordens de magnitudes maiores que qualquer tempo microscópico característico. O intervalo das correlações no espaço também aumentam quando o sistema atinge a temperatura crítica T_g . Sendo mais preciso, considere, para spins de Ising e a função correlação $\langle S_i S_j \rangle_t$. Para vidros de spins o sinal desta correlação será aleatório, de modo que é conveniente discutir o quadrado da função correlação $\langle S_i S_j \rangle_t^2$. Esta função decai a zero, geralmente de forma exponencial para uma primeira aproximação, quando $|R_i - R_j|$ aumenta. O intervalo sobre o qual ocorre o decaimento é denominado de comprimento de correlação de vidro de spin ξ_{SG} , de forma que temos a função correlação dada por

$$\langle S_i S_j \rangle_t^2 \propto \exp(-|R_i - R_j|/\xi_{SG}) \quad (1.1)$$

De fato, observa-se que ξ_{SG} aumenta quando o sistema aproxima-se de T_g , mostrando que importantes efeitos coletivos ocorrem quando $T \leq T_g$. O aumento de ξ_{SG} parece ser menos rápido do que o aumento dos tempos característicos. Na fase de vidro ou os tempos e comprimentos de correlação divergem abaixo de T_g , ou simplesmente tornam-se significativamente grandes para poderem ser medidos. Na fase vidro tempos de relaxação e comprimento de correlação dependem



das condições iniciais em que o sistema foi colocado. Portanto tais efeitos provocam a quebra de ergodicidade do sistema de modo que é necessário definirmos um parâmetro de ordem diferente do adotado para $T > T_g$. Se existe uma transição de fase pode-se calcular a média temporal pela média da mecânica estatística e assim definir um parâmetro de ordem para o vidro de spin. O ordenamento de spins em vidros de spins é um problema de materiais estruturalmente desordenados, diferente de sistemas mais convencionais regulares tais como os cristais ideais [22].

A física do vidro de spin advém de problemas sobre questões fundamentais e por isso foi pesquisada longamente pela física da matéria condensada, de forma que foram apresentados vários tipos de modelos para os mais variados tipos de vidros de spin. No entanto os vários vidros de spins estão divididos em dois tipos básicos: modelos triviais e modelos não triviais. Os vidros de spins triviais têm seu estado fundamental degenerado, e possui um relevo de energia com muitos vales, mas todos eles com barreiras de energia entre os vales de mesmo tamanho. Além deste fato, quando estes sistemas encontram a temperatura de vidro o eles se congelam em muitas configurações diferentes. Outro tipo de vidro de spins são os modelos não triviais. Para estes modelos o estado fundamental é não degenerado com um relevo de energia rugoso que também possui muitos mínimos, porém estes mínimos estão separados entre si por barreiras de energia de tamanhos diferentes. E quando o sistema encontra-se na fase vidro ele se congela em um número pequeno de configurações [22]. Tal classe de modelos lançou luz sobre o problema de enovelamento de proteína. Para se descrever o relevo de energia rugoso associado a tais sistemas os modelos precisam de dois ingredientes importantes, que devem ser considerados conjuntamente, que são: a aleatoriedade e a frustração. Estes dois termos tomados juntamente são responsáveis pela rugosidade do relevo de energia. A rugosidade do relevo de energia refere-se a competição (frustração) entre os termos de interação da Hamiltoniana. Gross e Mezard [23] generalizaram esta classe de modelos e mostraram que a Hamiltoniana para este tipo de sistema



poderia ser descrita por:

$$H = - \sum_{1 \leq i_1 < i_2 \dots < i_p \leq N} J_{i_1 i_2 \dots i_p} S_{i_1} S_{i_2} \dots S_{i_p} \quad (1.2)$$

onde as interações de spin $J_{i_1 i_2 \dots i_p}$ são dadas pela distribuição:

$$P(J_{i_1 i_2 \dots i_p}) = \left[\frac{N^{p-1}}{\pi p!} \right]^{1/2} \exp \left[- \frac{(J_{i_1 i_2 \dots i_p})^2}{(\Delta J)^2 p!} N^{p-1} \right] \quad (1.3)$$

sendo que para $p=1$ e N spins livres temos o modelo Edwards-Anderson, para $p=2$ interações e N spins temos o modelo Sherrington-Kirkpatrick e para $p \rightarrow \infty$ este modelo torna-se equivalente ao modelo de energia aleatória (**R**andom **E**nergy **M**odel-**R**EM (REM) de Derrida em [24].

1.2.4 Modelo de Energia Aleatória

O modelo de energia aleatória (**R**andom **E**nergy **M**odel-**R**EM) é o limite de uma família de sistemas de vidros de spin, em que a correlação entre os níveis de energia pode ser desprezada. O modelo exibe uma transição de fase que tem uma temperatura T_g na qual abaixo dela o sistema se encontra congelado. O modelo de Derrida devido as suas características é solúvel. O modelo de energia aleatória para um sistema de N spins de Ising tem as seguintes propriedades (REM) [25]:

(i) O sistema tem 2^N níveis de energia E_i .

(ii) Os níveis de energia estão distribuídos de acordo com a distribuição de probabilidade

$$P(E) = \frac{1}{\sqrt{(2\pi N(\Delta E)^2)}} \exp \left[\frac{-(E - \bar{E})^2}{2N(\Delta E)^2} \right] \quad (1.4)$$

(iii) As energias E_i são variáveis aleatórias.

As duas primeiras propriedades são características de muitos modelos de vidros de spin (modelos não triviais), especialmente definidos pela Hamiltoniana

(1.2) A terceira característica do modelo é uma simplificação adotada por Derrida para que o modelo proposto por ele possa ser resolvido exatamente [25].

O número médio de estados com energias E e $E + dE$ no modelo de energia aleatória é

$$\langle n(E) \rangle = 2^N P(E) dE. \quad (1.5)$$

Quando $\langle N(E) \rangle \gg 1$ é possível aproximar $\langle \log(n(E)) \rangle$ por $\log \langle n(E) \rangle$, e assim a entropia é dada por

$$S(E) \simeq \log \langle n(E) \rangle \quad (1.6)$$

ou seja,

$$S(E) \simeq N \log 2 - \frac{(E - \bar{E})^2}{2N(\Delta E)^2}. \quad (1.7)$$

Desta equação encontramos uma energia crítica E_c , dada por

$$E_c = \{\bar{E} - \Delta E N(2 \log 2)^{1/2}\}, \quad (1.8)$$

onde a entropia se anula e a aproximação (1.6) não é válida. Quando o sistema encontra tal energia crítica E_c , o sistema fica então congelado em um pequeno número de estados de baixa energia. No limite termodinâmico para $E < E_c$ ao invés da equação (1.6), $S(E) = 0$. A temperatura desta transição ocorre a [25]

$$\frac{1}{T_g} = \left. \frac{\partial S}{\partial E} \right|_{E=E_c} \quad (1.9)$$

o que dá

$$\frac{1}{T_g} = \frac{(2 \log 2)^{1/2}}{\Delta E} \quad (1.10)$$

A energia livre do sistema é dado

$$F(T) = \begin{cases} \bar{E} - \frac{(\Delta E)^2}{2T} - TS & \text{se } T > T_c \text{ e } S = N(2 \log 2)^{1/2} \\ E_c = \bar{E} - \Delta E N(2 \log 2)^{1/2} & \text{se } T \leq T_c \end{cases} \quad (1.11)$$

1.2.5 Modelo Bryngelson-Wolynes

Alguns experimentos de enovelamento de proteína sugerem que elas tenham um relevo de energia consideravelmente rugoso. Contudo outros experimentos mostram que o relevo de energia das proteínas é suave. Através de uma visão fenomenológica pode se afirmar que o enovelamento de proteína ocorre em um relevo de energia que fica entre o suave e o rugoso. O modelo Bryngelson-Wolynes interpola estes dois limites utilizando o modelo de **REM**. Tal modelo está baseado em dois postulados [15].

Primeiro Bryngelson e Wolynes assumem que a energia de contato entre dois resíduos de uma proteína, que não ocorrem na estrutura nativa, pode ser considerada como variável aleatória, ou seja, as interações não nativas podem ser tomadas como variáveis aleatórias. Segundo o modelo quando uma parte da proteína se encontra em uma estrutura secundária correta as contribuições energéticas tendem a estabilizar a estrutura terciária. Além disso, quando um contato correto é realizado, em média os contatos possíveis diminuem o valor da energia. Deste modo adota-se a medida de similaridade de estrutura, que é usada para medir a distância que uma proteína se encontra no enovelamento em relação ao estado nativo. A proteína pode sofrer algumas flutuações energéticas, mas a medida que a similaridade de estrutura em relação ao estado nativo aumenta, pode-se garantir que a proteína está se aproximando do estado nativo e que o valor da energia está diminuindo [15].

O modelo Bryngelson-Wolynes introduziu o conceito de **frustração mínima**. Tal conceito corresponde a propensão da estrutura secundária não estar em conflito com a estrutura terciária do estado nativo da proteína. Neste modelo, a medida que a proteína se enovela, ela está minimiza sua frustração. Bryngelson e Wolynes utilizaram o termo **frustração mínima** para descrever o postulado da suavidade. A suavidade do relevo de uma proteína aparece por causa da se-

leção das proteínas realizada pela evolução. A evolução escolhe sequências que maximizem a habilidade da proteína em se enovelar rapidamente, para que se tornem biologicamente ativa [15]. A suavidade dos relevo do enovelamento surge da seleção de sequências de proteínas através da evolução. Se há a necessidade de se enovelar rapidamente então o relevo suave será o panorama adequado. Por outro lado, existem outros fatores que podem influenciar na evolução. A evolução pode não conseguir remover totalmente as interações frustradas de proteínas que podem ocorrer naturalmente. De fato a evolução neutra pode sugerir que aleatoriedade e a frustração podem se estender até o ponto adequado de estabilidade e enovelabilidade cinética [15].

A **frustração mínima** em proteínas pode ser facilmente mostrada. Examinando o cristal de raios-X de uma macromolécula biológica se pode notar que as cadeias laterais são mudadas pela evolução para tornarem-se coerentes à **estrutura quaternária** ou **super estrutura**. Exemplos deste tipo de observação podem ser encontrados nas *zippers* de luecinas e barris betas anfifílicos. Sequências simétricas como elas normalmente têm baixa frustração. A consistência entre estrutura secundária e estrutura terciária é denominada de **princípio da consistência estrutural** que foi enunciado por Gō[15].

Uma forma de medir a proximidade da configuração é através da similaridade da estrutura, pela contagem de contatos nativos corretos formados na configuração da estrutura nativa, encontrados na configuração atual, este é um bom parâmetro do seu grau de enovelamento. Esta medida pode ser considerada como uma coordenada de reação Q , onde Q é uma variável normalizável, variando de 0 a 1. Quando a estrutura da proteína tem uma completa similaridade com relação ao estado nativo $Q \approx 1$. Quando $Q \approx 0$ não existe nenhuma similaridade com relação a estrutura nativa da proteína. Para o modelo Bryngelson-Wolynes a energia de um dado estado aparece devido a contribuições de muitos termos aleatórios, tal que para cada valor de Q a distribuição de probabilidade é uma Gaussiana centrada na energia média dada por [2]:



$$P(Q, E) = \frac{1}{\sqrt{2\pi \Delta E^2(Q)}} \exp \left[\frac{-|E - E(Q)|^2}{2\Delta E^2(Q)} \right]. \quad (1.12)$$

Através do modelo de energia aleatória pode-se determinar a energia livre do enovelamento, em analogia a equação (1.11, para $T > T_c$)

$$F(Q, T) = \bar{E}(Q) - \frac{\Delta E^2(Q)}{2k_b T} - T S_0(Q), \quad (1.13)$$

onde $S_0(Q)$ é a entropia configuracional, k_b é a constante de Boltzmann. Para altas temperaturas os estados desenovelados são favorecidos. No entanto a temperaturas baixas as configurações enoveladas serão favorecidas. A função da energia potencial possuirá dois mínimos. A temperatura de enovelamento é a temperatura em que o processo de enovelamento começa a ocorrer. Uma forma de definir a temperatura de enovelamento T_f é a temperatura em que os estados desnaturado ($Q \approx 0$) e o estado nativo ($Q \approx 1$) estão igualmente populados [2].

Muitas, senão quase todas reações químicas são complexas e envolvem a transposição de um número grande de barreiras de energia. Exemplos deste tipo de reação são as reações de isomerização e transporte difusional em vidros, ligantes sendo ligadas a hemoglobinas e o enovelamento de proteína. O modelo Bryngelson-Wolynes trata a dinâmica do enovelamento como uma reação complexa do tipo: $D \rightleftharpoons N$. Entretanto, a maioria das idéias dinâmicas advém de reações químicas simples, realizadas na fase gasosa ou de isomerização da fase condensada em líquidos. O problema é que não se pode descrever reações químicas complexas através de reações simples. Uma reação simples pensada através da Teoria de Transição Padrão tem quatro pontos importantes. Em primeiro lugar o solvente em reações simples ou é ausente ou tem um comportamento passivo. Em segundo lugar o estado inicial e final bem como todos os estados intermediários têm estruturas bem definidas de tal forma que considerações a cerca da entropia não são necessárias. Terceiro existe uma única coordenada de reação bem definida. E por último o coeficiente de difusão varia muito pouco desde o estado inicial até seu estado final [15].

O enovelamento como vimos acima está longe de ser considerado uma reação simples pela Teoria de Estado de Transição. O enovelamento é uma reação complexa que possui características muito peculiares. Em primeiro lugar ao contrário das reações químicas padrões, no enovelamento o solvente é muito importante para a estabilização da estrutura enovelada da proteína. Em segundo o estado desnaturado, o estado enovelado e todos os estados de transição referem-se a conjuntos de estruturas e desta forma se faz necessária uma descrição que considere os efeitos da entropia. Em terceiro lugar, diferentemente das reações químicas padrões, o enovelamento tem várias coordenadas de reações e diversos caminhos ou rotas que podem atingir o estado nativo. Por último existem mudanças qualitativas durante o enovelamento, em particular se o movimento térmico de uma cadeia aberta for maior do que as estruturas colapsadas, além disso o coeficiente de difusão pode variar significativamente entre o estado desnaturado e o estado enovelado [15].

Reações químicas complexas como o enovelamento possuem relevos de energia que precisam ser descritos por um grande número de parâmetros de ordem por possuírem um grande número de pontos de sela e muitos mínimos, tornando a descrição cinética do enovelamento difícil. No modelo de Bryngelson-Wolynes a cinética é formulada da seguinte forma: a proteína se move de um estado a outro por estados dinamicamente conectados. A proteína que está com uma dada similaridade Q com a estrutura nativa pode se mover para um estado com uma similaridade $Q + \delta q$ ou $Q - \delta q$. Entretanto cada estado está conectado a $\mathcal{S}$ microestados, onde $\mathcal{S}$ é bem maior do que 1 e bem menor do que o número de estados da proteína. O estado de uma proteína é determinado pelo mínimo local dos $\mathcal{S}$ microestados de similaridade Q . Esta aproximação é boa quando o número de aminoácidos que mudam simultaneamente é bem menor do que o número total de estados da proteína. A discretização dos estados simplifica bastante a matemática usada na análise estatística do relevo de energia. No entanto, apesar da discretização dos estados, Bryngelson e Wolynes mantém uma



coordenada coletiva que mede o progresso do enovelamento definida pelo relevo de energia potencial [26].

O modelo Bryngelson-Wolynes estimou o tempo de enovelamento e o tempo de vida de um estado em mínimo local. No entanto, quando a temperatura do sistema está sendo abaixada a proteína pode encontrar um temperatura T_g análogo a T_c da equação (??)

$$\frac{1}{T_g} = \sqrt{\frac{2k_b S_0(Q)}{\Delta E(Q)^2}}, \quad (1.14)$$

na qual o tempo de vida de um estado pode se tornar comparável ao tempo enovelamento. A análise de Bryngelson e Wolynes mostrou que para $T > T_g$ a proteína pode ser descrita por um ensemble estatístico e as equações de tempo de enovelamento e tempo de vida de microestados são válidas. Para $T < T_g$ a proteína está na fase vidro, ficando armadilhada em muitos mínimos locais e o enovelamento completo da proteína em sua estrutura nativa pode não ocorrer. Um ensemble estatístico é um conjunto de um número grande de $\mathcal{N}$ de sistemas (ensemble), sendo que cada um desses sistemas pode ser pensado como a cópia em nível termodinâmico do sistema a ser estudado. As estruturas encontradas na fase de vidro não podem ser representada por ensemble estatístico, porque quando a proteína atinge esta fase o sistema passa a depender das condições iniciais em que foi colocada. Nesta fase as estimativas para o tempo de enovelamento, e tempo de vida de um estado em um mínimo local, feitas no modelo Bryngelson-Wolynes deixam ser válidas [15].

Sequências boas para o enovelamento são sequências que se enovelam rapidamente e que sejam dominadas por um forte gradiente de energia livre, o que faz com que a proteína atinja a estrutura do estado fundamental (nativo). Para que a proteína alcance a estrutura do estado nativo, ela precisa separar em energia a conformação nativa das conformações que são estruturalmente similares a ela da maioria das outras conformações que não tem nenhuma similaridade com a estrutura nativa. Utilizando as idéias do modelo Bryngelson-Wolynes, Socci

[27] mostrou que sequências boas para o enovelamento são sequências que além de possuírem um forte gradiente de energia, minimizem a frustração da proteína maximizando o valor da razão T_f/T_g [28]. Goldstein em [?] utilizando mais de um método seguro de se avaliar parâmetros diferentes avaliou certas grandezas de interesse para o enovelamento e percebeu que o gap de energia (diferença entre o estado nativo e o primeiro estado excitado) está sendo maximizado quando a razão T_f/T_g está sendo maximizado. Neste trabalho então, Goldstein et al. mostraram que o gap de energia pode estar associado à estabilidade termodinâmica, pois quando a grandeza relativa a estabilidade termodinâmica T_f/T_g está sendo maximizada o gap de energia também está sendo maximizado. No modelo Bryngelson-Wolynes, a maximização do gap de energia de uma proteína é uma consequência tautológica do maior grau de estabilidade das interações presentes na estrutura nativa por causa do princípio de frustração mínima. Os resultados obtidos por Goldstein em [29] confirmaram a evidência do modelo Bryngelson-Wolynes de que sequências boas são as sequências que maximizem a razão T_f/T_g .

1.2.6 Funil

Leopold et al. em 1992[35] formularam o conceito de funil sugerindo que a minimização da energia, durante o processo de enovelamento, pode ser acompanhada por um modelo cinético que otimiza o processo. O trabalho inicial afirma que o estado nativo é o único que possui as condições de ser cineticamente acessível, e termodinamicamente estável com respeito a flutuações locais em condições fisiológicas. As condições de unicidade do estado nativo, de acessibilidade cinética, e de estabilidade termodinâmicas são necessárias e suficientes segundo Leopold para se descrever o processo do enovelamento [35].

Simplificações são feitas para a descrição do modelo cinético proposto pelo conceito de funil de enovelamento. O trabalho apresenta um modelo de rede, em que cada monômero da sequência é colocado em um sítio da rede e



considerando a proteína que seja um polímero de 27 monômeros, tendo vinte tipos de monômeros. O problema do enovelamento de proteína foi tratado em um regime de alta hidrofobicidade. Neste regime as interações entre monômeros iguais são mais atrativas do que as interações entre monômeros diferentes. Em alta hidrofobicidade a proteína se enovela de um estado desnaturado aleatório rapidamente para uma região de colapso e daí a proteína se reconfigura lentamente até atingir o estado nativo. A dinâmica do modelo do funil, para enovelamento, pode ser pensada como uma coleção de acontecimentos discretos entre os mínimos locais. Se a dinâmica for governada pelas mesmas interações entre os resíduos, que as dirigem ao colapso, então se pode dividir o movimento da proteína em dois tipos: um relativamente rápido estericamente restringido ao movimento da cadeia que não envolve quebra de muitos contatos entre os resíduos, e um outro tipo de movimento que é lento por atravessar barreiras de energias associadas a grandes flutuações conformacionais quebrando muitos contatos da cadeia e fazendo outros contatos entre os resíduos [35].

No trabalho original sobre o funil é sugerido que o enovelamento deve ocorrer através de um rápido colapso da proteína de seu estado desnaturado aleatório e posteriormente uma lenta reconfiguração da proteína até que esta atinja seu estado nativo. No funil esta reconfiguração deve ocorrer entre conformações similares compactas, de modo que somente as interconversões locais são permitidas depois do colapso da cadeia [35]. Desde que o conceito de funil esteja acoplado ao fato de uma proteína poder se enovelar, então o funil de enovelamento dependerá da estrutura primária da proteína, então se pode afirmar que quando o funil de uma certa sequência de aminoácido atinge uma conformação única estável dizemos que esta proteína é enovelável. Quando uma certa proteína exhibe múltiplos funis ou um funil que não consegue atingir um estado único que seja termodinamicamente estável dizemos que esta proteína é não enovelável [35]

A idéia de enovelamento usando o conceito de funil, é que a conformação colapsada de uma proteína sofre flutuações entre uma série de estados conforma-



cionais densos que estão na rota para que a proteína atinja seu estado nativo, a cada momento a proteína está mais próxima de sua conformação nativa. O enovelamento é um processo de difusão entre estados compactos, que podem ser medidos pelo número de contatos nativos que existem em determinado tempo. Se os estados compactos estudados forem os maximamente compactos, então se pode descrever os estados da dinâmica pelas projeções em estados maximamente compactos. A conclusão do trabalho sobre o modelo funil, é que este modelo propôs a chave termodinâmica e cinética que o enovelamento exige: convergência de famílias de caminhos difusivos para um único estado localmente estável e cineticamente acessível. Este trabalho mostrou que a idéia proposta por Levinthal de que deveria existir um único caminho para o enovelamento não é verdadeira, ao invés disso Leopold conseguiu mostrar que existe uma multiplicidade de caminhos que a proteína pode tomar para atingir seu estado nativo [35].

O trabalho original sobre o funil de Leopold tratou de proteínas que se encontravam em regime de alta hidrofobicidade, no entanto o conceito de funil também é aplicável para proteínas que estão em regime de baixa hidrofobicidade. No regime de baixa hidrofobicidade as interações entre os monômeros iguais são atrativas, ao passo que as interações entre os monômeros diferentes são repulsivas. Neste regime a proteína se enovela diretamente em seu estado nativo, sem passar por um colapso da cadeia da proteína. Aplicar o conceito de funil para o relevo de energia do enovelamento de proteína pode ser descrito por uma visão cinética, que considera o processo de enovelamento como a organização progressiva de um ensemble de estruturas parcialmente enoveladas, as quais a proteína passa até atingir seu estado nativo. Globalmente o relevo de energia assemelha-se a um funil extensamente rugoso. A rugosidade local reflete as armadilhas transientes das proteínas em mínimos locais. A forma afunilada do relevo de energia é necessária para que a proteína vença o paradoxo de Levinthal. Tardamente no processo de enovelamento a proteína pode encontrar um caminho, mas neste estágio a proteína já está quase enovelada de modo que a pesquisa conformacional torna-se



limitada [2].

1.2.7 Motivação

O problema de enovelamento de proteínas tem sido intensamente abordado pela comunidade científica, causando avanços sobre a questão. Exemplos são o Modelo Bryngelson-Wolynes, que busca descrever o processo de enovelamento estatisticamente, bem como o conceito de funil. Modelos simples, tais como os modelos de rede têm desempenhado um papel fundamental na elucidação dos mecanismos associados ao problema do enovelamento de proteína. No modelo de rede os monômeros de uma proteína ocupam os sítios da rede. O volume excluído é introduzido nestes modelos considerando que um sítio da rede é ocupado por um único monômero. O modelo de proteína mais usado é o modelo **HP**, que considera que o polímero tem somente dois tipos diferentes de monômeros, um que é o polar **P** e o hidrofóbico **H**. O modelo de rede é uma aproximação simplificada de proteínas. A grande vantagem em se utilizar este tipo de modelo é que ele possibilita uma análise mais profunda de aspectos essenciais, aspectos estes que seriam de difícil abordagem com a utilização de modelos mais complexos (realísticos). Por exemplo usando modelos de rede em duas dimensões Chan e Dill [36, 37] demonstraram que o enovelamento ocorre em um processo com duas fases distintas: primeiro um rápido colapso da cadeia e depois uma fase lenta em que a cadeia da proteína se reconfigura até atingir o estado nativo [27].

Conforme mencionado, quando é maximizada a razão T_f/T_g estamos maximizando o gap de energia. Shakhovich e Gutin [38] estudaram na rede sequências de 27 monômeros bem projetadas. Na rede eles [38] restringiram seus estudos a conformações maximamente compactas. A escolha feita por Shakhovich e Gutin [38] não foi gratuita porque uma conformação maximamente compacta não pode se interconverter em outra conformação maximamente compacta, de



modo que como vimos anteriormente o gap de energia ,calculado nos estados maximamente compactos, está associado a estabilidade da proteína, ou de outra forma a maximização da razão T_f/T_g . Primeiro Shaknovich e Gutin [38], escolheram uma conformação maximamente compacta. Eles adotaram um código de duas letras para o polímero, em que o potencial é dado por:

$$\begin{matrix} & A & B \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} E_l & E_u \\ E_u & E_l \end{pmatrix} \end{matrix}, \quad (1.15)$$

onde $E_l = -3$ e $E_u = -1$ (caso de alta hidrofobicidade). Fixando também que as sequências tenham 14 monômeros do tipo A e 13 monômeros do tipo B, buscou-se sequências com estado nativo único, e que tivesse a menor quantidade de contatos entre monômeros diferentes (impróprios), ou seja com menor frustração no estado nativo. A melhor sequência de duas letras foi a sequência bem projetada é ABABBBBBBABBABABAAABBAAAAAAB (sequência 002). É Importante salientar que esta sequência tem seus 28 contatos na rede todos próprios(contatos entre monômeros iguais), um gap de energia igual a 4. Esta sequência 002 é uma sequência bem projetada por não possuir em seu estado nativo nenhum contato entre tipos diferentes de monômeros [38].

Socci e Onuchic [27] confirmaram o princípio de frustração mínima do modelo Bryngelson-Wolynes, mostrando que sequências boas eram sequências que não tinham nenhum contato impróprio (entre monômeros diferentes) . Eles trabalharam com seis sequências diferentes, entre elas estava a sequência 002 de Shaknovich. Socci e Onuchic mostraram que as sequências boas para o e-novelamento eram as sequências com maior razão T_f/T_g . Para os parâmetros de $E_l = -3$ e $E_u = -1$ a sequência 002 possui a razão T_f/T_g com valor próximo de $= 1.3$ [28]

Posteriormente Socci e Onuchic[28] redesenharam a sequência 002 com três tipos diferentes de monômeros, e obtiveram uma sequência bem projetada



ABABBBBCBACBABABACACBACAACAB (sequência 012), que também não tem nenhum contato impróprio no estado nativo. Esta sequência (012) possui 11 monômeros do tipo A, 10 monômeros do tipo B e 6 monômeros do tipo C. O potencial de interação adotado foi semelhante ao adotado por Shaknovich que foi dado por

$$\begin{matrix} & A & B & C \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} E_l & E_u & E_u \\ E_u & E_l & E_u \\ E_u & E_u & E_l \end{pmatrix} \end{matrix} \quad (1.16)$$

onde $E_l = -3$ e $E_u = -1$. O estado nativo da sequência 012 é mostrado na figura (1.1)

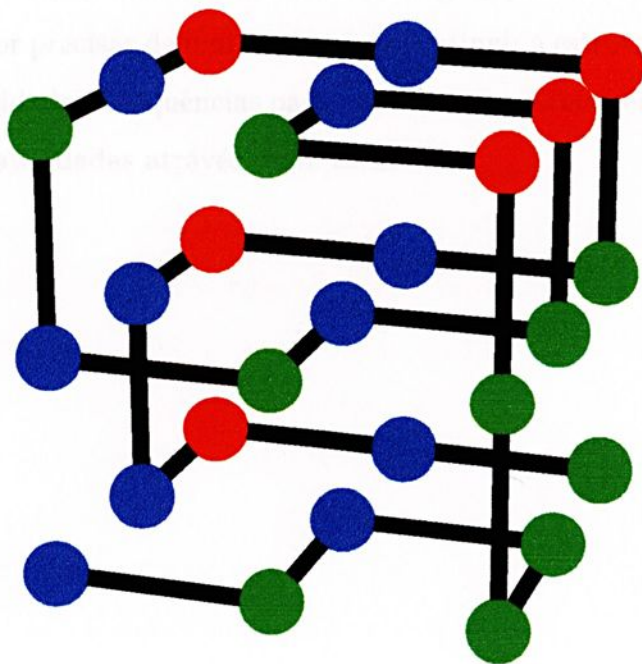


Figura 1.1: Estrutura Nativa da sequência 012

Com a mudança do código de 2 para 3 tipos distintos de monômeros

houve um aumento da razão T_f/T_g de 1.3 para 1.6 indicando que a sequência 012 deve ser menos frustrada e ainda mais bem projetada do que a da sequência 002. Dada uma sequência otimamente desenhada, como a sequência 012, nosso trabalho investiga o efeito das mutações em tais sequências. As questões abordadas são: Como perturbações na sequência primária afetam a estabilidade termodinâmica e cinética destas proteínas? Estes efeitos variam com o grau de hidrofobicidade do sistema?

Do ponto de vista evolutivo é importante que as proteínas utilizadas na biologia apresentem um enovelamento robusto às perturbações na sua estrutura primária. Caso contrário, proteínas que sofram pequenas mutações poderiam comprometer seriamente a sua atividade biológica, por não se enovelar satisfatoriamente, ou por precisar de muito tempo para atingir a estrutura nativa. Questões como a diversidade de sequências para uma mesma proteína em diferentes espécies também são abordadas através deste estudo.



Capítulo 2

Modelo de proteína na rede

2.1 Estruturas compactas e Energética

Os polímeros estudados são todos do tamanho de 27 monômeros. Neste trabalho a proteína foi considerada como um polímero em que seus monômeros estão posicionados em uma rede cúbica tridimensional. Os monômeros estão conectados aos seus vizinhos mais próximos na sequência de aminoácidos covalentemente [39]. Cada monômero ocupa somente um sítio (volume excluído) da rede cúbica. Trabalhando em um modelo de rede, não é fácil enumerar todas as conformações compactas de um polímero, entretanto é factível enumerar as conformações maximamente compactas de um polímero. As conformações maximamente compactas são cubos com todos os vinte e sete sítios ocupados pelos monômeros do polímero. Nestes cubos o número total de contatos na rede (desconsiderando os contatos devido às ligações covalentes) é de 28 contatos. O número total de cubos compactos é 103346. Se escolhermos um potencial de interação que favoreça a formação de contatos então a conformação nativa de uma dada sequência pode ser encontrada pela enumeração dos cubos compactos, e o estado nativo será dado pelo estado maximamente compacto com menor ener-



gia. Trabalhando no espaço das conformações compactas podemos controlar a degenerescência do estado nativo. Como estamos interessados em polímeros que tenham comportamento de proteínas, estaremos interessados em polímeros que tenham o estado nativo não degenerado para satisfazer a condição de que o estado nativo deve ser único. Neste trabalho estudamos polímeros no espaço das configurações maximamente compactas na rede $3 \times 3 \times 3$ monômeros.

Em nossos estudos utilizamos sequências de 27 monômeros com três tipos diferentes de monômeros. As sequências estudadas mantêm constante a proporção 11:10:6 para os monômeros A, B e C respectivamente. A energia da sequência é dada pela interação de contato entre os vizinhos mais próximos na rede (desconsiderando os contatos covalentes da sequência do polímero). A energia entre dois monômeros iguais é dada por E_l , e a energia de contato entre dois monômeros diferentes é dada por E_u , de modo que a função energia é dada por

$$E = N_l E_l + N_u E_u, \quad (2.1)$$

onde N_l é o número de contatos entre monômeros iguais e N_u é o número de contatos entre monômeros diferentes. O potencial de interação é dado por

$$\begin{array}{c} A \quad B \quad C \\ A \left(\begin{array}{ccc} E_l & E_u & E_u \\ B \left(\begin{array}{ccc} E_u & E_l & E_u \\ C \left(\begin{array}{ccc} E_u & E_u & E_l \end{array} \right) \end{array} \right) \end{array} \right) \end{array} \quad (2.2)$$

Em nosso trabalho utilizamos dois tipos diferentes de potenciais. Um potencial altamente hidrofóbico, como visto anteriormente favorece um enovelamento, onde primeiro o polímero se colapsa rapidamente em estruturas compactas e depois lentamente este polímero se reconfigura até atingir o seu estado nativo. O outro tipo de potencial utilizado é um potencial de baixa hidrofobicidade. Neste regime, a proteína vai se enovelando em sua conformação nativa, sem passar pelo colapso inicial da cadeia, que ocorre no caso de um potencial de alta hidrofobi-

cidade. Do ponto de vista prático a análise do comportamento termodinâmico do sistema pode ser realizada facilmente nos dois regimes pois se estivermos em um certo regime hidrofóbico e quisermos avaliar as grandezas termodinâmicas em outro regime, é só reescalarmos os valores das energias E_l e E_u . Por exemplo, se quisermos saber o valor de energia do estado nativo, que foi calculado no regime de alta hidrofobicidade, agora no regime de baixa hidrofobicidade, precisamos saber os valores de N_l e N_u , e mudarmos os valores de E_l e E_u , e assim recalcula-se o valor da energia do estado nativo para baixa hidrofobicidade. Para alta hidrofobicidade $E_l = -3$, e $E_u = -1$, para baixa hidrofobicidade, E_l tem o mesmo valor que para alta hidrofobicidade $E_l = -3$, e $E_u = 3$.

Como vimos é muito fácil expressar a energia de um polímero em termos de E_l e E_u . Entretanto algumas propriedades deste modelo são mais claras se considerandos os conjuntos de parâmetros E_{avg} , Δ e κ . Para um polímero com dois tipos de monômeros, com igual abundância na cadeia do polímero, definimos: $E_{avg} = \frac{1}{2}(E_l + E_u)$ e $\Delta = (E_u - E_l)$, onde E_{avg} indica a compactação da cadeia. Se E_{avg} for menor que zero o colapso da cadeia do polímero será favorecida. Δ indica a rugosidade do relevo[40]. Para cadeias de polímeros com mais de dois tipos de aminoácidos, que possuam os diferentes tipos distribuídos igualmente ao longo da cadeia, as definições acima apresentadas deixam de ser apropriadas por estarem definidas para polímeros que tenham somente dois tipos de monômeros. Chahine et al. [41], mostraram que para cadeias com m tipos diferentes de monômeros os parâmetros que descrevem bem o modelo são a energia média (não específica) entre monômeros não ligados pode ser dada

$$\bar{E}_{ns} \cong \sum_{i=1}^m [f_i^2 E_l + f_i(1 - f_i) E_u], \quad (2.3)$$

para indicar a compactação da cadeia, onde f_i , é a frequência relativa do monômero i na cadeia do heteropolímero. O valor de $\bar{E}_{ns}$, nos informa o quanto os contatos são favorecidos, ou sobre o grau de compactação da cadeia, se o valor de $\bar{E}_{ns}$ for menor do que zero a formação de estruturas compactas colapsadas será favore-

cida. A rugosidade do relevo, segundo Chahine et al. [41], para polímeros de m tipos diferentes de monômero pode ser expressada aproximadamente por:

$$\Delta \cong \left\{ \sum_{i=1}^m [f_i^2 (E_l - \bar{E}_{ns})^2 + f_i(1 - f_i)(E_u - \bar{E}_{ns})^2] \right\}^{1/2}, \quad (2.4)$$

e a razão

$$\kappa = \frac{-\bar{E}_{ns}}{\Delta} \quad (2.5)$$

é um parâmetro que mede o grau de hidrofobicidade. Em nosso trabalho temos $m = 3$ monômeros distintos para polímeros de 27 monômeros que mantêm a proporção 11:10:6 respectivamente para os monômeros do tipo A, B e C, indicando assim que a ocorrência destes monômeros é 11/27, 10/27 e 6/27 respectivamente para os monômeros do tipo A, B e C. Obtemos para o regime de alta hidrofobicidade HH , $E_{ns}(HH) = -1.70$, indicando que a compactação da cadeia está sendo fortemente favorecida, e o valor de $\Delta(HH) = 0.95$ o que dá um valor de $\kappa(HH) = 1.78$. Para o regime de baixa hidrofobicidade LH obtém-se que $E_{ns}(LH) = 0.94$, este valor não assegura o colapso da cadeia. O valor de $\Delta(LH)$ para baixa hidrofobicidade é de 2.85 o que nos dá um $\kappa = -0.33$. Grandezas como o número de contatos Z foram determinado pelo método histograma de Monte Carlo, o qual será descrito mais adiante. A temperatura crítica de enovelamento T_f é determinada usando o critério de Gutin dado em [39], que define a temperatura crítica de enovelamento como a temperatura na qual metade do número total de contatos nativos são formados. Depois determinada a temperatura crítica de enovelamento é realizada uma análise cinética dos polímeros através do Método de Monte Carlo padrão, com especial interesse em medirmos os tempos de enovelamento e de colapso para os polímeros pesquisados.

Devemos agora especificar a dinâmica de nosso modelo. Para sistemas em modelos de rede não existe uma única maneira de definir o conjunto de movimentos que podem ser adotados, sendo que cada conjunto de movimento pode nos

fornecer comportamentos cinéticos diferentes. Chan e Dill em [36, 37] mostraram que medidas como o tempo de colapso dependem do tipo de conjunto de movimentos adotados. A idéia é adotarmos um conjunto de movimentos que seja do ponto de vista dinâmico o mais realista possível. Tem-se que tomar cuidado com o tipo de informação a ser extraída da dinâmica em rede. Uma questão inapropriada de se fazer quando adota-se um conjunto de movimentos na rede, por exemplo, é desejarmos analisar os detalhes específicos que influenciam na formação da estrutura secundária de um polímero, para tal tipo de interesse precisaríamos de um conjunto de movimentos mais realista. No entanto comportamentos gerais do colapso e do enovelamento não devem depender dramaticamente do conjunto de movimentos adotados [27].

O conjunto de movimentos neste modelo consiste de movimentos locais que preservam as ligações covalentes do polímero e impõem que cada sítio da rede seja ocupado por somente um monômero. Este conjunto de movimentos foi desenvolvido pela dinâmica da física de polímeros. Estes movimentos padrões são constituídos por três tipos de movimento como na figura 2.1,

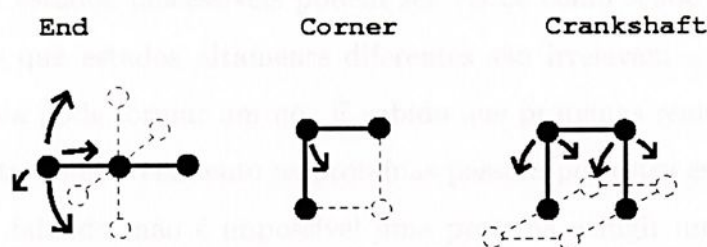


Figura 2.1: tipos de movimentos

O movimento de fim consiste do movimento dos fins da cadeia movem-se para sítios adjacentes vazios. O movimento de canto, que consiste do movimento de um monômero e o movimento de eixo de manivela em que dois monômeros são movimentados simultaneamente como na figura 2.1. Movimentos mais complexos, em que mais de dois monômeros se movimentam simultaneamente podem

ocorrer. Mas para tais tipos de movimento devem existir taxas diferentes, isto é, o movimento de cinco monômeros simultaneamente ocorrerá mais lentamente que o movimento de um único monômero. Se isto não for levado em conta, as escalas de tempo serão distorcidas para movimentos diferentes, o que quer dizer que encontraremos tempos “físicos” diferentes, porque estes tempos dependerão do tipo de movimento que o polímero sofreu. Para corrigirmos isto nós utilizamos uma probabilidade diferente para cada tipo de movimento. Como no nosso caso os polímeros são pequenos nós movimentamos somente um ou dois monômeros por vez [27].

Existe um comentário que deve ser feito sobre o conjunto dos movimentos escolhido para a dinâmica na rede, eles são não ergódigos. Em particular, não é possível que uma cadeia atinja uma conformação compacta de nó a partir de uma conformação estendida. A questão é se a não ergodicidade dos movimentos cria algum problema. A resposta simples para o problema é que estamos interessados na dinâmica do enovelamento, quando o estado nativo da proteína é acessível, não existindo problemas devido a existência de estados inacessíveis. Mais precisamente, os estados inacessíveis podem ser vistos como sendo irrelevantes do mesmo modo que estados altamente diferentes são irrelevantes para proteínas reais. A cadeia pode formar um nó. É sabido que proteínas reais não tem esses nós, entretanto no enovelamento as proteínas passam por estes estados com nós. Estritamente falando, não é impossível uma proteína atingir um estado de nó, isto é apenas diferente. Devido aos vínculos da rede, o estado de nó é inacessível mais que diferente. Entretanto a existência de estados que tenham nós não afetará as propriedades gerais do enovelamento. Uma maneira de argumentar é que estes estados inacessíveis não afetam os cálculos termodinâmicos. Em particular, qual seria o efeito pelo fato do movimento da cadeia estar restrito ao subespaço ergódigo, ao invés de considerar todo o espaço? De que maneira dependerá os tamanhos relativos do espaço excluído? Na prática para pequenas cadeias de proteínas o erro introduzido pela não ergodicidade do conjunto de movimento é



menor do que o erro estatístico [27].

Um movimento é realizado usando o algoritmo Metropolis Monte Carlo. Um monômero da cadeia é escolhido de forma aleatória. Se o monômero selecionado for um monômero do fim da cadeia, então seleciona-se aleatoriamente um ponto vizinho. Se for verificado que este ponto não faz parte do fim da cadeia, então ele pode ter um movimento de canto ou de eixo de manivela, dependendo da posição dos vizinhos ao longo da cadeia. Se algum dos movimentos permitidos violar o vínculo do volume excluído, colocando dois monômeros em um mesmo sítio, então o polímero retorna a configuração anterior e um novo monômero é escolhido. Se um movimento é possível, ou seja, se o sítio da rede é vazio, então a energia da conformação é calculada e comparada com a energia original. Se a energia diminuir o movimento é aceito incondicionalmente. Se a energia aumenta, então o movimento é aceito com a probabilidade usual de Boltzmann:

$$P = \exp[-(E_{nova} - E_{velha})/T] \quad (2.6)$$

onde E_{nova} e E_{velha} são respectivamente as energias nova e antiga. Note que foi escolhido que a constante de Boltzmann como um $k_b = 1$. Se a energia do movimento aumentar, calcula-se a probabilidade P dada pela equação acima e gera-se um número aleatório entre zero e um, se o número aleatório gerado for menor do que o valor da probabilidade P , então o movimento é aceito caso contrário é rejeitado [27].

O método de histograma de Monte Carlo é uma amostragem do Monte Carlo. O histograma de Monte Carlo é uma técnica que corre várias simulações para um certo intervalo de temperatura, ou seja, é calculado um histograma utilizando o algoritmo de Monte Carlo. Este histograma em uma temperatura T , a mede a probabilidade da energia E ocorrerem $h(E, T')$. Isto é igual a média térmica da densidade de estados

$$h(E, T') = \frac{n(E)e^{-E/T'}}{Z(T')} \quad (2.7)$$

onde $Z(T')$ é a função de partição para a temperatura T'

$$Z(T') = \sum_E n(E)e^{-E/T'} \quad (2.8)$$

$$n(E) = h(E, T')e^{E/T'} Z(T') \quad (2.9)$$

onde $Z(T')$ é uma constante multiplicativa desconhecida. Para grandezas intensivas a média térmica é calculada usando

$$\langle O \rangle (T) = \frac{\sum_E O(E)n(E)e^{-E/T}}{\sum_E n(E)e^{-E/T}}, \quad (2.10)$$

ou de outra forma

$$\langle O \rangle (T) = \frac{\sum_E O(E)n(E)e^{-E/T+E/T'}}{\sum_E h(E, T')e^{-E/T'+E/T'}}. \quad (2.11)$$

Se o interesse é em calcular grandezas extensivas como a energia livre e a entropia, torna-se necessário determinar a constante $Z(T')$. Para nosso sistema é possível calcular a função de partição $Z(T')$ e obter a densidade de estados. Para determinar a constante $Z(T')$ precisa-se conhecer a multiplicidade de estado energético. Por exemplo, as sequências estudadas aqui neste trabalho são sequências com estado nativo não degenerado. Isto significa que $n(E_{ef}) = 1$, onde E_{ef} é a menor energia dos 103346 cubos maximamente compactos. $Z(T')$ pode ser determinada e a energia livre então pode ser calculada usando $F = -T \log Z$.

Há um limite para o intervalo de temperatura sobre o qual a equação (2.11) é válida. Para temperaturas que estão muito longe da temperatura de simulação (enovelamento), os erros no cálculo da densidade de estados tornam-se significativos usando a equação (2.9). Para uma dada temperatura o sistema somente é amostrado para uma dada energia do espaço de fase. Para altas temperaturas de simulação, o estado nativo nunca é investigado, por outro lado para

temperaturas muito baixas um estado com energia alta nunca será alcançado. Consequentemente, a densidade de estados não será correta para regiões não amostradas de forma apropriada. Os limites do intervalo de temperatura podem ser explorados por qualquer simulação, que monitore os erros no cálculo da densidade de estados. Uma solução para este problema é o método de múltiplos histogramas. A idéia é usar várias simulações interligando os histogramas, para que estes histogramas sejam analisados juntamente.

Para os nossos polímeros de 27 monômeros usados neste trabalho, um histograma simples nos fornece resultados satisfatórios sobre um intervalo de temperatura de interesse. A razão é que a largura do histograma de energia tem uma escala como $1/\sqrt{N}$, onde N é o tamanho do sistema. No nosso caso o sistema é pequeno o suficiente para assegurar que os histogramas são largos e uma grande região do espaço de fase é amostrada para uma dada temperatura. Em nosso trabalho realizamos a simulação Monte Carlo na temperatura crítica de enovelamento T_f .

Capítulo 3

Resultados

Para analisar o efeito de mutações sobre uma sequência bem projetada com relação a estabilidade termodinâmica e cinética, neste trabalho foi escolhida uma sequência bem projetada com um código de três letras ABABB-BCBACBABABACACBACAACAB (sequência 012). Esta sequência foi escolhida para nosso trabalho porque ela é uma sequência otimamente projetada, que possui uma grande esbabilidade, com uma razão $T_f/T_g = 1.6$ maximizada, garantindo que, como vimos anteriormente que, a sequência bem projetada tenha um grande gap de energia para o modelo de rede 3x3x3, de modo que garanta que este gap esteja associado à estabilidade da proteína. Sobre esta sequência foram realizadas mutações simples e duplas. Uma mutação simples permitida em nosso trabalho, pode ser entendida como a troca de posição de dois monômeros diferentes na sequência do polímero bem projetado, ou seja, toda vez que na sequência bem projetada encontrarmos dois monômeros diferentes, então se troca de posição estes dois monômeros e se considera que esta mutação simples gerou uma nova sequência mutada. Uma mutação dupla é a troca de posição de dois pares distintos de monômeros diferentes na sequência do polímero. As mutações simples ou duplas sempre mantém a proporção da sequência bem projetada (sequência 012) de 11:10:6 monômeros respectivamente para A, B e C.



Todas as sequências mutadas, com uma ou duas mutações diferentes, percorreram o espaço das conformações maximamente compactas para um polímero de 27 monômeros em uma rede cúbica 3x3x3. Percorrendo este espaço para todas as sequências diferentes se procurou saber quais destas sequências tinham estado nativo não degenerado, para que as sequências mutadas estudadas tenham um comportamento de uma proteína. Para as sequências com estado nativo não degenerado, avaliou-se os valores da energia do estado fundamental da proteína bem como o gap de energia

$$gap = E_1 - E_0 \quad (3.1)$$

onde E_0 é a energia do estado fundamental (nativo) e E_1 é a energia do primeiro estado excitado. O cálculo do gap de energia é realizado mantendo-se as configurações das duas menores energias, sendo que a menor energia (estado fundamental) E_0 necessariamente não pode ser degenerada e o Zscore

$$Zscore = \frac{\langle E \rangle - E_0}{\sigma} \quad (3.2)$$

onde $\langle E \rangle$ é a energia média das conformações maximamente compactas e é dado por

$$\langle E \rangle = \frac{\sum_{i=1}^M f_i E_i}{\sum_{i=1}^M f_i} \quad (3.3)$$

, onde $\sum_{i=1}^M f_i = M$, que o número total de conformações maximamente compactas para uma rede cúbica 3x3x3, ou seja $M = 10346$ e σ é o desvio padrão que é dado por: $\sigma = \sqrt{\langle E^2 \rangle - \langle E \rangle^2}$, onde $\langle E^2 \rangle$ é a média da soma das energias ao quadrado. Estas duas grandezas são duas grandezas relativas à estabilidade da proteína muito debatidas na literatura. Alguns dizem que o gap de energia associado ao valor da energia do estado fundamental é a melhor forma de se medir a estabilidade da proteína, outros afirmam que o Zscore é a melhor forma[42], por isto é que neste trabalho medimos as duas grandezas para avaliar a estabilidade da proteína. A sequência 012 no espaço dos cubos 3x3x3 tem a energia do estado nativo $E_0 = -84$, $gap = 8$ $Zscore = 6.75$.

O espaço de sequências com somente uma sequência sobre a sequência bem projetada é de 236. Este número é obtido fazendo-se todas as mutações simples sobre a sequência original, de modo que a cada mutação simples permitida se considere uma sequência de polímero diferente. Uma forma simples de se entender como são realizadas as mutações é a seguinte: toda vez que dois monômeros da sequência de aminoácidos diferentes forem encontrados, trocam-se de posição estes monômeros e se considera que esta permutação gera uma nova sequência da proteína mutada. De todo o espaço de sequências com uma mutação, que tem 236, 204 delas têm o estado nativo não degenerado. Das 204 sequências diferentes com estado nativo não degenerado, 196 (96% das sequências mutadas com estado fundamental não degenerado) sequências têm a mesma estrutura nativa da sequência original (sequência 012) vide figura (??). No caso de duas mutações realizadas sobre a sequência bem projetada estão envolvidos quatro monômeros, ou melhor dois pares de monômeros. Na mutação dupla, toda vez que se encontrava dois pares de monômeros diferentes na sequência original, realizava-se a troca de posição na sequência dos dois pares de monômeros, aqui a mutação dupla, ou de outra forma a permutação dupla pode ser entendida como duas mutações simples que ocorrem simultaneamente. A cada mutação dupla considera-se uma nova sequência de polímero. Quando obtida uma nova sequência com mutação dupla se verificava se esta nova sequência já não tinha sido obtida para o caso de uma mutação simples com estado nativo não degenerado. Para o caso de duas mutações possíveis obtém-se 20076 sequências mutadas diferentes, desse total foram encontradas 9224 sequências diferentes com estado nativo não degenerado. Do total de 9224 sequências diferentes com estado nativo não degenerado, 4214 sequências (46% das sequências com estado nativo não degenerado) têm a mesma estrutura nativa da sequência original. Tais resultados indicam que quando vamos do caso de uma mutação simples para o caso de uma mutação dupla, a porcentagem de sequência com estado nativo único que tem a mesma estrutura da sequência original cai de 96% para 46%. Isto mostra que as mutações

duplas podem causar grandes mudanças de comportamento termodinâmico nas sequências mutadas. Como este trabalho está interessado em sequências mutadas que mantenham um comportamento termodinâmico semelhante ao da sequência original, então decidimos estudar as sequências, ou seja as 204 sequências com estado nativo não degenerado, que têm somente uma única mutação em relação a sequência original.

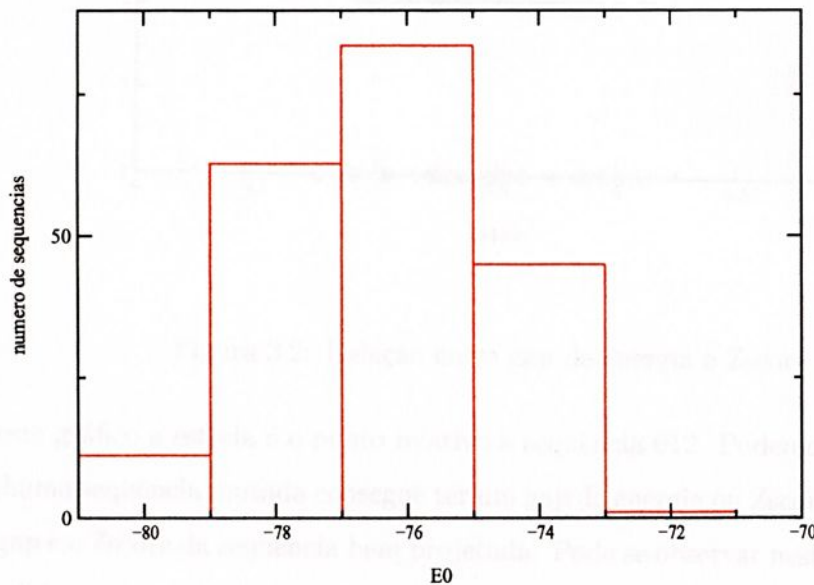


Figura 3.1: Histograma das energias das sequências mutadas não degeneradas

O histograma 3.1 acima nos fornece as energias das sequências mutadas com estado fundamental não degenerado. Das 204 sequências mutadas somente uma sequência possui energia -72 , 45 têm energia -74 , 84 têm energia -76 , 63 possuem energia -78 , 11 sequências têm energia -80 , mostrando que nenhuma sequência mutada consegue ter um estado nativo não degenerado, ou seja, um estado nativo sem contatos impróprios.

A relação entre o gap de energia com o Zscore é dada na figura(3.2).

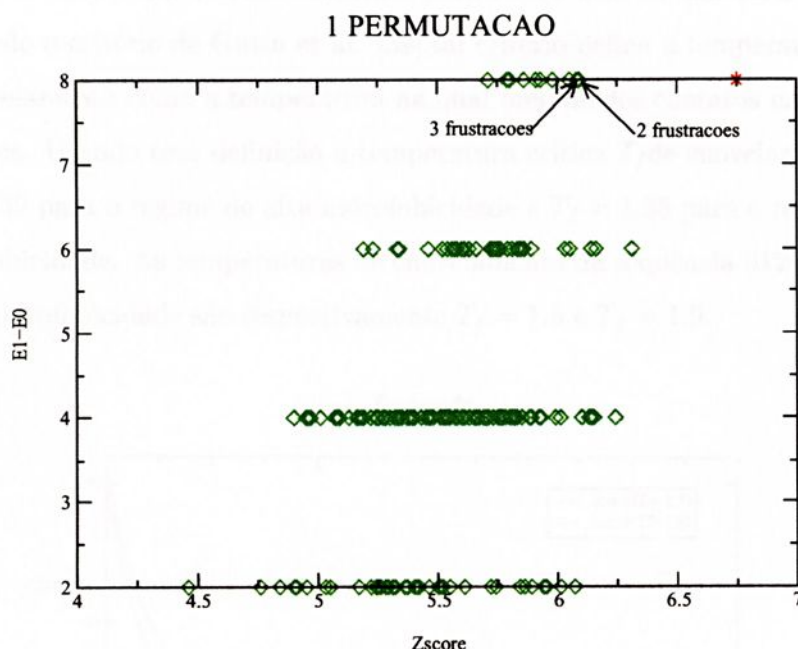


Figura 3.2: Relação entre gap de energia e Zscore

Neste gráfico a estrela é o ponto relativo a sequência 012. Podemos observar que nenhuma sequência mutada consegue ter um gap de energia ou Zscore maior do que o gap e o Zscore da sequência bem projetada. Pode se observar neste gráfico que à medida que o gap de energia e o Zscore crescem o número de sequências mutadas diminuem. Das sequências mutadas com gap de energia igual a oito escolheu-se inicialmente a sequência com maior Zscore que possuía a maior frustração, que é dada por ABABBBBCBABBABACACACBACAACAB para comparar seu comportamento cinético com a da sequência bem projetada (012). A sequência mutada têm três contatos impróprios, ou seja, a energia do estado nativo desta sequência é $E_0 = -78$ no regime de alta hidrofobicidade e $E_0 = -66$ para o regime de baixa hidrofobicidade, enquanto que a energia do estado nativo da sequência bem projetada não possui nenhum contato impróprio, ou seja, a energia do estado fundamental é $E_0 = -84$.

A temperatura de enovelamento para a sequência mutada foi determinada utilizando o critério de Gutin et al. [39], tal critério define a temperatura crítica de enovelamento como a temperatura na qual metade dos contatos nativos estão formados. Usando esta definição a temperatura crítica T_f de enovelamento como $T_f = 1.32$ para o regime de alta hidrofobicidade e $T_f = 1.35$ para o regime baixa hidrofobicidade. As temperaturas de enovelamento da sequência 012 para alta e baixa hidrofobicidade são respectivamente $T_f = 1.5$ e $T_f = 1.9$.

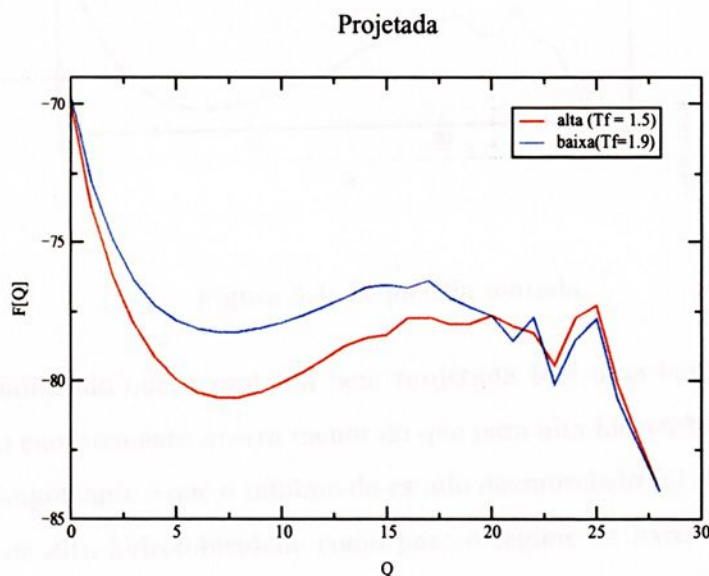


Figura 3.3: Sequência bem projetada

Os comportamentos das energias livres em função dos contatos nativos Q para os regimes de alta e baixa hidrofobicidade da sequência bem projetada, estão apresentados na figura (3.3). A curva vermelha refere-se ao comportamento da proteína bem projetada (012) em alta hidrofobicidade, e a curva azul refere-se ao comportamento da proteína bem projetada em baixa hidrofobicidade. As curvas da energia livres para os regimes de alta e baixa hidrofobicidade têm comportamentos superficiais parecidos. Porém examinando as duas curvas mais detalhadamente percebe-se a curva de energia livre em função do número de contatos

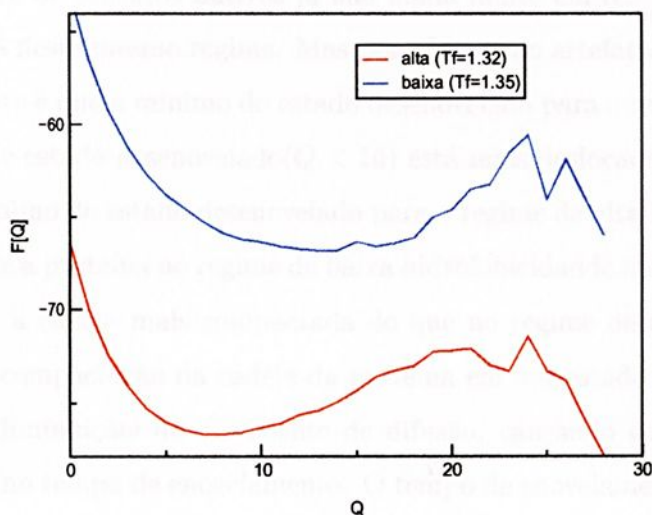


Figura 3.4: Sequência mutada

nativos, indicando que a proteína bem projetada tem uma barreira de energia para que o enovelamento ocorra menor do que para alta hidrofobicidade. Um outro ponto importante é que o mínimo do estado desenovelado ($Q < 16$) tanto para o regime de alta hidrofobicidade como para o regime de baixa hidrofobicidade estão próximos. O tempo de enovelamento para a proteína bem projetada em alta hidrofobicidade é aproximadamente cinco vezes mais lento do que para baixa hidrofobicidade

Os comportamentos das energias livres em função dos contatos nativos Q para os regimes de alta e baixa hidrofobicidade da sequência mutada, estão apresentados na figura (3.4) . A curva azul refere-se ao comportamento da proteína em baixa hidrofobicidade, e a curva vermelha refere-se ao comportamento da proteína mutada em alta hidrofobicidade. A sequência mutada tem ainda um comportamento superficial global semelhante para os regimes de alta e baixa hidrofobicidade. Nota-se que a curva de energia livre para baixa hidrofobicidade já não é tão suave quanto no caso da sequência bem projetada. Ao passo que

para alta hidrofobicidade o comportamento da curva de energia livre em função do número de contatos nativos já não muda muito em relação à sequência bem projetada neste mesmo regime. Mas tais efeitos são artefatos da rede. Um ponto importante é que o mínimo do estado desenovelado para o regime de baixa hidrofobicidade estado desenovelado ($Q < 16$) está mais deslocado para a esquerda do que o mínimo do estado desenovelado para o regime de alta hidrofobicidade. Isto indica que a proteína no regime de baixa hidrofobicidade no estado desenovelado está com a cadeia mais compactada do que no regime de alta hidrofobicidade. A maior compactação da cadeia da proteína em seu estado desenovelado provoca uma diminuição do coeficiente de difusão, causando consequentemente um aumento no tempo de enovelamento. O tempo de enovelamento para a sequência mutada em alta hidrofobicidade é aproximadamente 3 vezes mais rápida do que para baixa hidrofobicidade.

Agora fazendo um gráfico do número de contatos pelo número de contatos nativos temos o gráfico (3.5) Neste gráfico as curvas vermelhas se referem ao regime de alta hidrofobicidade e as curvas azuis se referem ao regime de baixa hidrofobicidade. As curvas contínuas são designadas para a sequência mutada e as curvas pontilhadas para a sequência bem projetada. Observando o gráfico percebe-se que no regime de alta hidrofobicidade, a formação de contatos em função dos contatos nativos da proteína mutada e da proteína bem projetada são bem parecidos. E a formação dos contatos nativos das proteínas mutada e bem projetada, para baixa hidrofobicidade, têm comportamentos bem diferentes se os compararmos com os comportamentos das proteínas mutada e bem projetada para alta hidrofobicidade, indicando talvez que a estabilidade termodinâmica da proteína mutada em alta hidrofobicidade varie menos do que quando a proteína é mutada em baixa hidrofobicidade. Na região desenovelada ($Q < 16$), o número médio de contatos $\langle Z \rangle$ em função do número de contatos nativos Q na sequência mutada aumenta em relação à sequência bem projetada, o que indica que para um dado valor de $Q < 16$ a proteína se en-



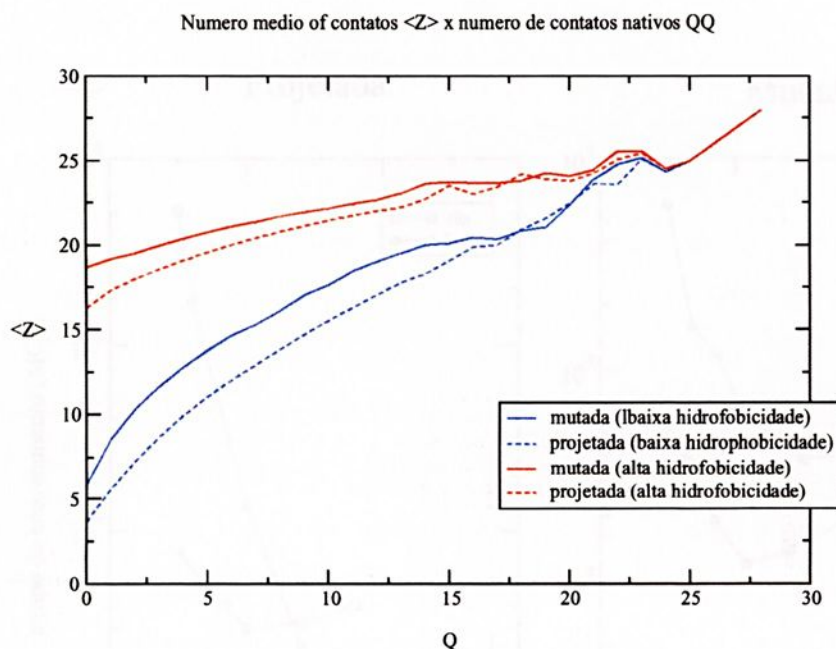


Figura 3.5: Z contra Q

contra mais compactada em seu estado desenovelado. Quando a proteína em seu estado desenovelado está mais compactada, a dinâmica torna-se mais lenta causando assim um aumento no tempo de enovelamento[39]. Desta forma a medida que $\langle Z(Q) \rangle$ aumenta a dinâmica torna-se mais lenta. Este efeito é mais pronunciado no regime de baixa hidrofobicidade, fato em concordância com os resultados cinéticos, que demonstram que neste regime o aumento nos tempos de enovelamento na sequência mutada é mais acentuado.

O resumo dos efeitos de pequenas mutações em uma proteína bem projetada sobre o comportamento termodinâmico e a cinética são apresentados no gráfico (3.6). Nesta figura estão lançados os tempos de enovelamento, medidos em passos de Monte Carlo, contra as temperatura. As curvas vermelhas se referem ao regime de alta hidrofobicidade, as curvas azuis se referem ao regime de baixa hidrofobicidade. O gráfico da esquerda nos mostra os tempos de enovelamento

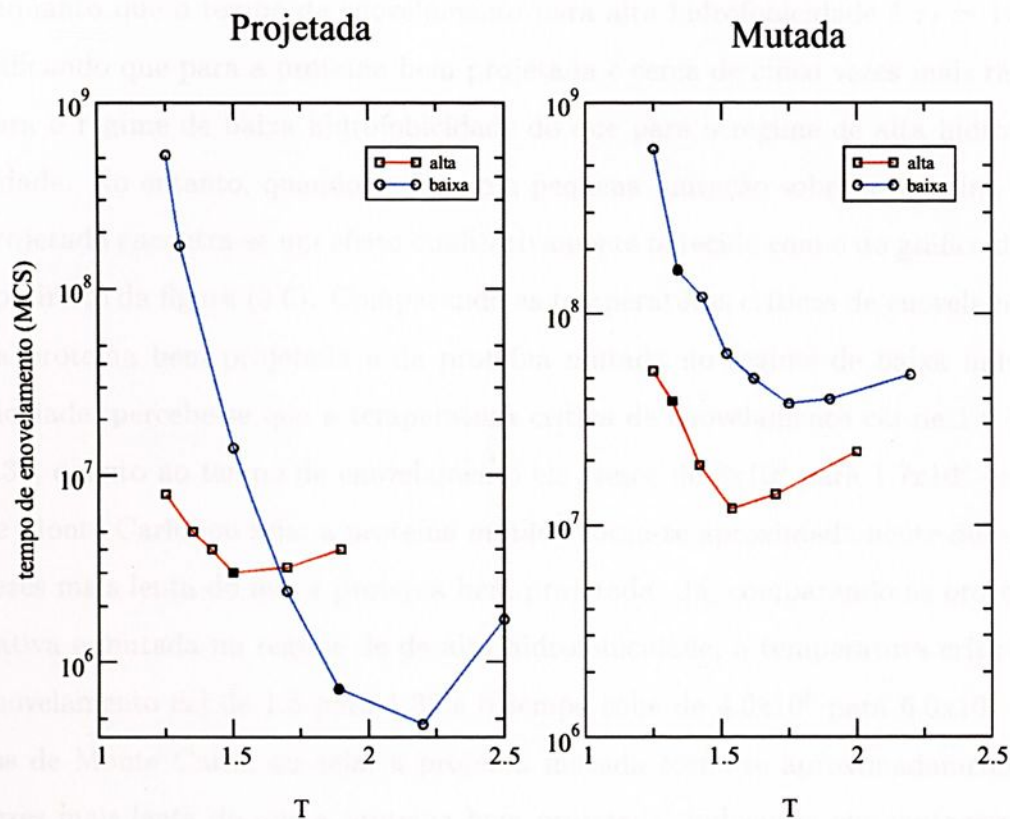


Figura 3.6: Tempo de enovelamento vs. temperatura

contra a temperatura para a proteína bem projetada. Os pontos cheios nos informam qual é a temperatura crítica de enovelamento T_f e o tempo de enovelamento τ_f . O gráfico da direita nos mostra os tempos de enovelamento contra a temperatura para a proteína mutada. O quadrado nos informa qual é a temperatura crítica de enovelamento T_f . Analisando primeiro o gráfico da esquerda se pode notar que a proteína bem projetada para o regime de baixa hidrofobicidade tem uma temperatura crítica de enovelamento $T_f = 1.9$ superior a temperatura crítica de enovelamento $T_f = 1.5$ para o regime de alta hidrofobicidade. Além do fato de no regime de baixa hidrofobicidade o tempo de enovelamento é $\tau_f \simeq 8 \times 10^6$,

enquanto que o tempo de enovelamento para alta hidrofobicidade é $\tau_f \simeq 4 \times 10^7$, indicando que para a proteína bem projetada é cerca de cinco vezes mais rápida para o regime de baixa hidrofobicidade do que para o regime de alta hidrofobicidade. No entanto, quando se faz uma pequena mutação sobre a proteína bem projetada encontra-se um efeito qualitativamente parecido com o do gráfico do lado direito da figura (3.6). Comparando as temperaturas críticas de enovelamento da proteína bem projetada e da proteína mutada no regime de baixa hidrofobicidade, percebe-se que a temperatura crítica de enovelamento cai de 1.9 para 1.35, quanto ao tempo de enovelamento ele cresce de 8×10^5 para 1.7×10^8 passos de Monte Carlo, ou seja, a proteína mutada torna-se aproximadamente duzentas vezes mais lenta do que a proteína bem projetada. Já, comparando as proteínas nativa e mutada no regime de alta hidrofobicidade, a temperatura crítica de enovelamento cai de 1.5 para 1.32 e o tempo sobe de 4.0×10^6 para 6.0×10^7 passos de Monte Carlo, ou seja, a proteína mutada torna-se aproximadamente 15 vezes mais lenta do que a proteína bem projetada, indicando que mutações em proteínas bem projetadas causam mudanças consideráveis no comportamento e na cinética da proteína e mais que as mudanças são mais acentuadas no regime de baixa hidrofobicidade do que no regime de alta hidrofobicidade.



Capítulo 4

Conclusão

O efeito de permutações na estrutura primária de uma proteína otimamente desenhada foi investigado em diferentes regimes de hidrofobicidade. A motivação para o estudo destas perturbações foi simular o efeito de mutação em sequências bem projetadas. Supondo que no processo evolutivo somente as boas mutações são preservadas, o comportamento cinético e termodinâmico da sequência bem projetada foi comparado com o comportamento de uma "boa mutação". A seleção de uma boa mutação seguiu um critério que em princípio maximiza a estabilidade termodinâmica e otimiza a cinética do processo de enovelamento, neste critério são utilizadas sequências que maximizem o gap e o Zscore.

No nosso modelo a mutação consiste na permutação da posição de dois monômeros na cadeia. A sequência bem desenhada não apresenta contatos impróprios no seu estado nativo, e qualquer sequência mutada apresenta contatos impróprios no estado nativo. No caso da sequência escolhida, que apresenta o maior gap e maior Zscore, ocorrem 3 contatos impróprios. A introdução dos contatos impróprios torna a sequência mais frustrada. Em princípio uma boa sequência mutada não deveria apresentar uma grande diferença no comportamento termodinâmico e cinético quando comparado à sequência original. No entanto a frustração introduzida é responsável por uma grande alteração no seu enovelamento. À medida que a frustração ocorre, a energia do estado nativo é



maior, sua estabilidade termodinâmica é menor, bem como o tempo de enovelamento. Estes efeitos são fortemente amplificados à medida que hidrofobicidade da cadeia diminui. Isto porque os contatos impróprios são mais energeticamente desfavoráveis em regime de baixa hidrofobicidade.

Nossos resultados demonstram de forma direta a sugestão de Gutin et al [39], que no regime de baixa hidrofobicidade as sequências de uma proteína enovelável tem que ser fortemente otimizada, e que no regime de alta hidrofobicidade sequências enoveláveis podem ser menos otimizadas. Mais que isso, nosso trabalho mostra como perturbações em sequências otimizadas influenciam o enovelamento em diferentes cenários hidrofóbicos, o que leva a importantes implicações na evolução de proteínas. Uma proteína que se enovela em um regime de baixa hidrofobicidade deve ser altamente seletiva a mutações ao longo de sua evolução biológica, ou seja, somente mutações que mantenham a sequência bem otimizada são preservadas. Neste sentido espera-se que tal proteína tenha alta homologia quando comparada entre diferentes espécies. Para uma proteína que se enovela em um regime altamente hidrofóbico o efeito é reverso, ou seja, a proteína apresenta uma maior probabilidade de enovelamento satisfatório mesmo diante de mutações não otimizadas. Desta forma, nesta classe de proteínas espera-se encontrar uma homologia significativamente menor quando sequências de uma mesma proteína em diferentes espécies são comparadas.

A proteína mutada analisada foi considerada uma "boa mutação" seguindo o critério de maximização do gap e Zscore. No entanto, conforme observamos esta sequência apresenta uma considerável frustração (3 contatos impróprios). Resta investigar se ao realizarmos uma mutação que seja desfavorável ao gap e Zscore, mas que não haja uma variação significativa da energia fundamental do estado nativo, as considerações feitas a respeito do regime hidrofóbico das proteínas continuam válidas. Este estudo já se encontra em andamento.

Quanto a trabalhos futuros, sugerimos a validação destas idéias em proteínas reais. Este estudo pode ser realizado investigando o grau de homologia de proteínas



de entre diferentes espécies nas proteínas depositadas no Protein Data Bank. O grau de homologia entre proteínas de diferentes espécies deverá apresentar direta relação com o grau de hidrofobicidade da proteína.

Referências Bibliográficas

- [1] ATKYNS L. *Biochemistry*, New York, Oxford, 1984, 1089 p.
- [2] DRUCKER, J.N., LUTNEY-SCHULTEN Z., WOLYNES, P.G. *Ann. Rev. of Phys. Chem.*, vol. 49, p. 535-554, 1987.
- [3] CRITCHFIELD, T. E. *Protein structure and Molecular properties*, New York, W.H. Freeman and Company 1982, 347 p.
- [4] CHAN, S.P., DILL, K.A. *Physica Today*, 34-35, FEBRUARY, 1991.
- [5] VOET D., VOET J., *Biochemistry*, New York, John Wiley & Sons, 1990, 722 p.
- [6] BROOKS, C.H., KARPLUS M., DILL, S.M. *Protein: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, New York, J. Wiley & Sons, 1995, p. 199.
- [7] PAULING L., COREY R.H., BRANSON H.R. *Proc. Natl. Acad. Sci. USA*, vol. 37, p. 255-261, 1951.
- [8] WATSON J.D., CRICK F.H.C. *Nature*, vol. 177, p. 398, 1963.
- [9] ANFENSEN C.B., HARPER J., SELA M., WHITE F.J. *Proc. Natl. Acad. Sci. USA*, vol. 47, p. 1209, 1961.
- [10] ANFENSEN, C.B. *Science*, vol. 181, p. 224-230, 1973.

Referências Bibliográficas

- [1] STRYER L., **Biochimistry**, New York, Frerman, 1988, 1089 p.
- [2] ONUCHIC, J.N., LUTHEY-SCHULTEN Z., WOLYNES,P.G.,*Annu. Rev of Phys. Chem.*, vol. 48,p. 539-594, 1997.
- [3] CREIHHTON, T. E. **Proteins:strutures and Molecuar properties**, New York, W.H. Freman and Company,1993, 507 p.
- [4] CHAN, H.S., DILL, K.A., *Physics Today*p. 24-32,FEBRAURY, 1993.
- [5] VOET D., VOET J., **Biochimistry**, New York, Jonh Wilie & Sons,1990,1223 p.
- [6] BROOKS, C.H., KARPLUS M., Pettitt B.M. **Proteins: A Theoretical Perspective of Dynmamics, Struture and Termodynamimics**,New York,J. Wiley & Sons,295 p. 1993.
- [7] PAULING L., COREY R.B., BRANSON H.B. *Proc. Natl. Acad. Sci. USA* vol.37, p. 205-211, 1951
- [8] WATSON J.D., CRICK,F.H.C., *Nature*, vol. 177, p. 964 , 1953
- [9] ANFINSEN C. B., HABER E., SELA M., WHITE F Jr., *Proc. Natl. Acad. Sci. USA*vol. 47, p. 1309, 1961
- [10] ANFINSEN, C.B.*Science*, vol 181, p. 223-230, 1973



- [11] PANDE, V.S., GROSBERG A.Y. *Rev. Mod. Phys.*, vol.72, p. 259-314, 2000
- [12] FRAUNBELDER H., WOLYNES, P.G., *Physics Today*, February, p-58-64, 1994.
- [13] LEVINTHAL C., *J. Chim. Phys*, vol.65, p. 44-45, 1968.
- [14] DILL K. A. *Protein Science*, vol. 8, p.1166-1180, 1999.
- [15] BRYNGELSON J.D., P.G. WOLYNES, *PROTEINS: Stru., Func. and Gene.*, vol. 21, p. 167-195, 1995.
- [16] MISKY A.E., Pauling, L., *Proc. Natl. Acad. Sci USA*. vol. 22, p. 439-447, 1936.
- [17] EYRING H., Stearn, A.E., *Chem. Rev.*, vol.24, p. 253-270, 1939.
- [18] KAUMANN W., *Adv. Protein Chem.*, vol.16 p. 1-16, 1959.
- [19] DILL K.A., BROMBERG, S., YUE K., *Protein Science*, vol. 4, p. 561-602, 1995.
- [20] TOULOUSE, G., *Commun. Phys.*, vol. 2, p. 115, 1977.
- [21] STEIN, D.L., *Proc. Natl. Acad. Sci USA*, vol. 82, p. 3670, 1985.
- [22] BINDER K., YONG A.P., *Rev. Mod. Physics* vol.58, número 4, p. 801-976, 1986.
- [23] GROSS D.J., MÉZARD, J., *Nucl. Phys. B*, vol 240, p. 431, 1984.
- [24] DERRIDA B. *The Physical Review Letters*, vol. 45, número 2, p. 79-82, 1980.
- [25] DERRIDA B., *Phys. Rev B*, vol. 34, número 5, p. 2613-2626, 1981.
- [26] BRYNGELSON J.D., P.G. WOLYNES, *The Journal Phys. Chem.*, vol. 93, p. 6902-6915, 1989.



- [27] SOCCI N.D., ONUCHIC, J.D., *J. Chem. Phys.*, vol. 101, p 4732-4744, 1995.
- [28] SOCCI N.D., ONUCHIC J.D., WOLYNES P.G., *J. Chem* , vol.104, p. 5860-5868, 1996.
- [29] GOLDSTEIN R. A., LUTHEY-SCHULTEN, Z. A., WOLYNES, P.G., *Proc. Natl. Acad. Sci. USA* vol.89, p. 4918-4922, 1992.
- [30] MCCAMMON J. A., HARVEY S. C., **Dynamics of Proteins and Nucleic Acids**, Cambridge, Cambridge University Press, 1987.
- [31] HONEYCUTT J.D., THIRUMALAI D., *Proc. Natl. Acad. Sci USA*, vol. 87, p. 3526-3529, 1990.
- [32] KARPLUS M., MacCammon, J. A., *Ann. Rev. Biochem.* , vol. 53, p. 263-300, 1983.
- [33] FRAUNBELDE H., PARAK F., YONG, R. D., *Ann. Rev. Biophys. Biophys. Chem.*, vol.17, p. 451-479, 1988.
- [34] GODSTEIN R. A. A., LUTHEY-SCHULTEN, Z. A., WOLYNES, P.G., *Proc. Natl. Acad. Sci. USA* vol.89, p. 9029-9033, 1992.
- [35] , LEOPOLD P.E., MONTAL M., ONUCHIC, J.N., *Proc Natl. Acad.Sci. USA*. vol. 89 p. 8721-8725. 1992.
- [36] CHAN H. S., DILL K. A., *Macromolecules*, vol. 22, p. 4559-4573, 1989b.
- [37] CHAN H. S., DILL K. A., *Macromolecules*, vol. 22, p. 4559-4573, 1989b.
- [38] SHAKHNOVICH E.I., GUTIN A., *Proc Natl. Sci USA* , vol. 90, p.7195-7200, 1993.
- [39] GUTIN A. M., ABKEVICH A. B., SHAKHNOVICH E. I. *Biochemistry*, vol. 34, p. 3066-3076, 1995.



- [40] SOCCI N. D., ONUCHIC, J. D. *J. Chem. Phys.*, vol. 101, p. 1519-1528, 1994.
- [41] CHAIHINE J., NYMEYER H., LEITE V.B.P. SOCCI, N. D., ONUCHIC J. N., aceito *Phys. Rev. Letters*
- [42] DIMA I.A., BANAVAR J.B., CIEPLACK M., MURITAN A., *Proc Natl. Acad. Sci. USA*, vol. 96, p. 4904-4907, 1999.



