

**ESTUDO DO COEFICIENTE DE DIFUSÃO NO
ENOVELAMENTO DE PROTEÍNAS NA REDE**

Ronaldo Júnio de Oliveira

Dissertação para Defesa de Mestrado
Pós-Graduação em Biofísica Molecular

Estudo do Coeficiente de Difusão no Enovelamento de Proteínas na Rede

Ronaldo Júnio de Oliveira

Exame apresentado como parte das exigências para obtenção do título de Mestre em Biofísica Molecular, área de concentração Biofísica Molecular do Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, câmpus de São José do Rio Preto, São Paulo, Brasil.

Orientador: **Prof. Dr. Vitor Barbanti Pereira Leite**

Co-orientador: **Prof. Dr. Jorge Chahine**

São José do Rio Preto

Março de 2007

Oliveira, Ronaldo Júnio de.

Estudo do coeficiente de difusão no enovelamento de proteínas na rede /
Ronaldo Júnio de Oliveira. - São José do Rio Preto : [s.n.], 2007.
78 f. : il. ; 30 cm.

Orientador: Vitor Barbanti Pereira Leite

Co-Orientador: Jorge Chahine

Dissertação (mestrado) - Universidade Estadual Paulista. Instituto de
Biociências, Letras e Ciências Exatas

1. Proteínas. 2. Proteínas - Enovelamento. 3. Coeficiente de difusão. 4.
Enovelamento de proteínas – Simulação por computador. 5. Método de Monte
Carlo. I. Leite, Vitor Barbanti Pereira. II. Chahine, Jorge. III. Universidade
Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. IV.
Título.

CDU - 577.112

*Aos meus pais que me
“ensinaram a pescar” ...*

Agradecimentos

Foram muitas as pessoas que contribuíram, direta ou indiretamente, para a realização desse trabalho de mestrado.

Gostaria de agradecer ao Prof. Dr. Vitor Barbanti Pereira Leite pela destreza na orientação desse trabalho. Foi de extrema valia seus ensinamentos para o meu amadurecimento profissional e pessoal.

Ao Prof. Dr. Jorge Chahine pela co-orientação e sabedoria nas longas discussões acerca do problema do envelhecimento de proteína.

Aos alunos do grupo: Leandro C. Oliveira, Luciana Melo, Mariane L. Paiva e Ricardo H. Theodoro pelas constantes discussões e companheirismo que fazem desse um ótimo grupo de pesquisa.

A todos os professores do Departamento de Física pelo ensino de qualidade que esta instituição me proporcionou. Em especial, ao Prof. Dr. João Ruggiero pela atenção especial durante o meu mestrado e pela sua participação na banca de qualificação.

A todos os amigos pós-graduandos e funcionários do Departamento de Física. Em especial, à Ana H. Casé, André L. Martins, Flávio Barboza e José Ézio Bessa pelo companheirismo e amizade formada nesses anos. Também, à Isadora P. D. Picola pela ajuda nos momentos mais difíceis.

Aos meus pais José de Oliveira e Vânia M. S. Oliveira e minha irmã Aline M. Oliveira pela paciência, dedicação e carinho por mim.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro.

“A imaginação humana, sublime por natureza, deleita-se com tudo que é remoto e extraordinário, e ela corre, sem controle, pelas mais longínquas regiões do tempo e do espaço, visando assim evitar os objetos que o costume lhe tem tornado demasiado familiares...”

Hume

Sumário

Lista de Abreviaturas	8
Lista de Símbolos	9
Lista de Figuras	12
Resumo	13
Abstract	14
1 Introdução	15
1.1 O que é uma proteína	15
1.2 O problema do enovelamento	17
1.3 O conceito de funil e superfície de energia: <i>The energy landscape</i> .	19
1.4 A equação de difusão	26
1.5 Motivação	31
2 Modelo	32
2.1 A importância dos Modelos Minimalistas	32
2.2 Modelo de rede cúbica	34
2.3 Os parâmetros de hidrofobicidade	40
2.4 Método do histograma	42
3 Resultados	45

3.1	Tempos de envelamento por simulação	45
3.2	Termodinâmica e cinética	47
3.3	Coefficiente de difusão constante	51
3.3.1	Barreira de energia em $Q = 16$	51
3.3.2	Barreira de energia em $Q = 23$	55
3.4	Coefficiente de difusão variado	59
4	Conclusão	67
	Referências Bibliográficas	70
	Apêndices	76
A	Função correlação	76

Lista de Abreviaturas

2D	Duas dimensões
3D	Três dimensões
BW	Bryngelson e Wolynes
DNA	Deoxyribose Nucleic Acid
HP	Hidrofóbico-Polar
MC	Monte Carlo
NMR	Nuclear Magnetic Resonance
RNA	Ribonucleic Acid

Lista de Símbolos

$D(Q, T)$	Coeficiente de difusão dependente da coordenada de reação e da temperatura
D_0	Coeficiente de difusão equivalente a uma superfície de energia lisa
E	Energia potencial do sistema
\bar{E}	Energia média de interação
E_{het}	Rugosidade da superfície de energia
E_{gs}	Energia do estado fundamental
E_l	Energia de interação entre contatos vizinhos iguais não-covalentes
E_u	Energia de interação entre contatos vizinhos diferentes não-covalentes
$F(Q, T)$	Energia livre dependente da coordenada de reação e da temperatura
H	Matriz de interação entre os contatos
$h(E, T)$	Histograma que mede a probabilidade da energia E ocorrer na temperatura T
k_B	Constante de <i>Boltzmann</i>
k	Grau de colapso
K	Constante de mola
$n(E)$	Densidade de estados na energia E
N_l	Número de contatos covalentes (<i>like contacts</i>)
N_u	Número de contatos não-covalentes (<i>unlike contacts</i>)
$P(Q, t)$	Probabilidade dependente da coordenada de reação e do tempo
Q	Coordenada de reação, número de contatos corretos nativos

$S(Q)$	Entropia dependente da coordenada de reação
T_g	Temperatura de vidro
T_f	Temperatura de enovelamento
β	Inverso da temperatura multiplicado pela constante de <i>Boltzmann</i>
$\Delta E^2(Q)$	Flutuação quadrática média da energia
τ_f	Tempo de enovelamento
τ_{max}	Tempo máximo de uma simulação
$Z(T)$	Função de partição na temperatura T

Lista de Figuras

1.1	Funil de estruturas proposto por Leopold	22
1.2	Esquema de conformações de uma proteína com relação ao grau de similaridade com o estado nativo	24
1.3	Esboço da energia livre para temperatura de enovelamento T_f em função da coordenada de reação	28
2.1	Estrutura aleatória inicial gerada na rede para uma seqüência de 27 monômeros	36
2.2	Estrutura maximamente compacta da seqüência de 27 monômeros	36
2.3	Tipos de movimentos permitidos na simulação	38
2.4	Diagrama dos níveis de energia do modelo	41
3.1	Tempo de enovelamento (τ_f) em função da temperatura	46
3.2	Energia livre em função do número de contatos corretos	48
3.3	Energia livre em função do parâmetro de ordem na temperatura de enovelamento T_f	50
3.4	Coefficiente de difusão D_0 em função da temperatura	53
3.5	Comparação dos tempos de enovelamento	54
3.6	Tempos de enovelamento para alta e baixa hidrofobicidade	57
3.7	Tempos de enovelamento calculados pela equação 3.3	58
3.8	Coefficientes de difusão em função da coordenada de reação	60
3.9	Coefficientes de difusão em função da coordenada de reação	62
3.10	Trajectoria de enovelamento em passos de MC	64

3.11	Coeficientes de difusão em função da coordenada de reação	65
A.1	Função correlação para várias temperaturas	77

Resumo

O enovelamento de proteínas é um problema fundamental em Biofísica Molecular. O processo de enovelamento é, em geral, mapeado através de uma equação de difusão aplicada ao longo da coordenada de reação Q , a qual descreve o grau de similaridade de uma determinada configuração com o estado nativo da proteína. Os tempos de enovelamento podem ser calculados a partir dos potenciais efetivos e do coeficiente de difusão D . Na literatura, D é assumido constante. Usando modelos de rede, mostramos neste trabalho variações desse coeficiente de difusão em função de Q e calculamos os novos tempos de enovelamento.

Palavras-chave: coeficiente de difusão, tempo de enovelamento, modelo de rede, enovelamento de proteína

Abstract

The protein folding is a fundamental problem in Molecular Biophysics. The folding process is, in general, mapped in a diffusion equation through the reaction coordinate Q , that is the similarity degree of a state with the native state. The folding times can be calculated with effective potentials and the diffusion coefficient D . In the literature, D is assumed constant. Using lattice models, we show in this work its variations in function of Q and we calculate the new folding times.

Keywords: diffusion coefficient, folding time, lattice model, protein folding

Capítulo 1

Introdução

1.1 O que é uma proteína

As proteínas fazem parte de uma classe de macromoléculas que têm papel fisiológico fundamental em toda atividade celular dos seres vivos. Algumas vezes encontram-se nas membranas celulares regulando a entrada e saída de íons, outras vezes no meio celular envolvendo macromoléculas e podem participar da formação de complexos moleculares. Enfim, essas têm inúmeras funções no meio celular e o seu mau funcionamento pode levar a célula e conseqüentemente o ser vivo a adoecer. A grande influência das proteínas sob a fisiologia dos organismos vivos torna essa classe de macromoléculas importante para o estudo.

As proteínas são formadas por seqüências de aminoácidos ligadas covalentemente por ligações peptídicas [1]. Essas seqüências podem ser curtas, peptídeo, ou longas, polipeptídeo ou proteína. Existem na natureza, em geral, 20 diferentes aminoácidos que podem ser classificados como hidrofílicos (ou polares) e hidrofóbicos (ou apolares).

Todas as proteínas de qualquer ser vivo são, em sua estrutura mais simples, uma seqüência formada por combinações destes 20 aminoácidos. Esses são

compostos por um carbono localizado na cadeia principal (C_α) que mantém quatro ligações, sendo uma com um átomo de hidrogênio, uma com o grupo amina (NH_3^+), uma com grupo carboxila (COO^-) e uma com um radical (R). Os grupos amina e carboxila estão em suas formas iônicas. O radical diferencia cada um dos 20 tipos de aminoácidos [2].

Os aminoácidos ligam-se através do grupo amina de um aminoácido com o grupo carboxila de outro liberando uma molécula de água (ligação peptídica). A junção dessas unidades peptídicas dão origem às cadeias poliméricas e, então, às proteínas. Elas são sintetizadas nos ribossomos onde o RNA mensageiro é traduzido de acordo com seu código genético, cada tripleto de nucleotídeo especifica um aminoácido correspondente [3].

Uma proteína possui quatro níveis de estrutura. A seqüência de aminoácidos na cadeia polipeptídica determina a estrutura primária ou seqüência da proteína. Grupos locais de aminoácidos que formam padrões que se repetem pela proteína como estruturas de hélices alfa, folhas beta ou voltas aleatórias formam a estrutura secundária da proteína. O arranjo tridimensional ($3D$) dessas estruturas secundárias e sua topologia com coordenadas de cada átomo incluindo as cadeias laterais formam a estrutura terciária. Cadeias polipeptídicas podem se rearranjar para formar a estrutura quaternária, ou seja, as estruturas terciárias idênticas se reagrupam para formar uma nova estrutura $3D$ [2].

A estrutura terciária (ou quaternária quando houver) de uma proteína determina a função que será exercida no meio biológico. Para que uma proteína seja biologicamente ativa, ela deve [4]:

- possuir uma conformação nativa;
- a conformação nativa deve ser estável termodinamicamente;
- atingir a conformação nativa (enovelar) a tempo de realizar sua função biológica.

Sendo assim, torna-se importante descobrir como uma proteína atinge sua forma nativa, quais são as forças que governam o enovelamento, quanto tempo uma proteína leva para atingir sua forma a fim de desempenhar seu papel biológico.

1.2 O problema do enovelamento

Entre as décadas de 1940 e 1950, iniciou-se a revolução na Biologia que deu origem à Biologia Molecular moderna. Os trabalhos de Pauling e colaboradores em 1951 [5], determinaram estruturas 3D de cristais de pequenas cadeias de aminoácidos. Com a resolução destas estruturas, Pauling descobriu que elas têm padrões que se repetem pela cadeia polipeptídica como alfa hélices e folhas beta, estruturas secundárias mantidas por ligações do tipo ponte de hidrogênio. Também, sugeriu que, devido à ressonância da dupla ligação entre os átomos de carbono-oxigênio e carbono-nitrogênio, a configuração de cada resíduo é planar, havendo somente ângulos entre os resíduos de aminoácidos na cadeia.

Logo após, em 1953, Watson e Crick [6] propuseram uma estrutura em dupla-hélice para as moléculas de DNA, inspirando um entendimento para o que foi chamado de Dogma Central da Biologia: DNA \rightarrow transcrição \rightarrow RNA \rightarrow tradução \rightarrow proteína. Dessa forma, foi compreendido o complexo maquinário envolvido na biosíntese de cadeias polipeptídicas [7] e deu-se início à novas discussões sobre essa nova área da biologia que estava emergindo: **O estudo do enovelamento e predição de estrutura de proteína.**

Os estudos pioneiros de Anfinsen e colaboradores [8] na década de 60 tiveram um papel fundamental para o desenvolvimento do estudo do enovelamento proteico. Seu estudo consistia em medir através de reações físico-químicas, a atividade de uma proteína, no caso a ribonuclease — uma cadeia polipeptídica de 104 aminoácidos com 4 pontes de sulfeto. Anfinsen mostrou que uma proteína

depois de desnaturada (desenovelada) poderia espontaneamente reenovelar-se, ou seja, reestabelecer sua forma nativa de proteína em condições fisiológicas adequadas, reativando sua função biológica. Anfinsen então, concluiu que essa seqüência continha toda a informação necessária para definir a sua estrutura tridimensional de proteína e por conseqüência, a sua função biológica [9]. Também pode-se concluir com esses experimentos que o processo de desnaturação – renaturação era reversível: sempre que se clivasse as pontes de sulfeto, quando reestabelecida as condições fisiológicas essas eram novamente formadas.

Com seus estudos sobre a desnaturação da ribonuclease, Anfinsen estabeleceu o que foi chamado de **Hipótese Termodinâmica**. Essa hipótese diz que a estrutura tridimensional de uma proteína no seu estado nativo em condições fisiológicas normais é o nível energético mais baixo da energia livre de Gibbs de todo o sistema, ou seja, a conformação nativa é determinada totalmente por interações interatômicas e, portanto, pela seqüência de aminoácidos de uma proteína no meio [9]. Dessa forma, tornou-se evidente que o enovelamento de uma proteína era “independente do caminho”, ou seja, não existiria um caminho cinético pela qual a proteína tivesse que passar para encontrar a sua forma nativa.

Em oposição às idéias de Anfinsen, Levinthal [10] argumentou ser impossível acessar todo o espaço conformacional em busca da menor energia livre de Gibbs em um tempo relativamente curto. Isso por que o número de conformações possíveis cresce exponencialmente com o número de resíduos de aminoácidos. Para ilustrar essa afirmação, Levinthal supôs uma proteína com N aminoácidos que pudessem assumir μ orientações no espaço, então o número de conformações possíveis distintas para uma proteína seria μ^N . Para uma estimativa simples, consideremos que cada aminoácido tenha somente duas possibilidades de orientação, então o número de conformações possíveis para uma proteína com 100 aminoácidos é de $2^{100} \approx 10^{30}$. Se pelo menos um picosegundo é requerido para acessar uma conformação, então o tempo necessário para explorar todas as conformações de uma proteína de 100 aminoácidos é aproximadamente

10^{18} segundos ou mais de 10^{10} anos, tempo da ordem de grandeza da idade do universo. Essa idéia ficou conhecida como o **Paradoxo de Levinthal**.

Levinthal então estimou que uma proteína não poderia se enovelar em tempo suficiente de procurar seu mínimo global de energia, sendo que a conformação final da proteína deveria ser alcançada por um caminho cinético preferencial, o enovelamento é “dependente do caminho” (*pathways*), ou seja, “uma seqüência bem definida de eventos que seguem um ao outro e então levam a proteína de um estado aleatório desenovelado para um único estado metaestável enovelado” [10].

Como pode o enovelamento de uma proteína ser dependente e independente do caminho ao mesmo tempo? Durante a década de 80, uma nova teoria começou a ser desenvolvida para resolver esse problema. Leopold e colaboradores [11, 12] introduziram o conceito de funil de energia, onde se observa a superfície de energia global da proteína (*energy landscape*) [13], como uma alternativa ao paradoxo de Levinthal e aos caminhos de enovelamento. Esse conceito será estudado mais detalhadamente na próxima subseção.

As pesquisas no enovelamento de proteína têm fascinado biólogos, físicos, químicos e matemáticos nos últimos 50 anos. Entender como ocorre esse complexo mecanismo ainda continuará sendo, por muito tempo, um problema em aberto para pesquisadores do mundo inteiro.

1.3 O conceito de funil e superfície de energia:

The energy landscape

O conceito de funil e superfície de energia, junto com a nova geração de experimentos, têm contribuído significativamente para o entendimento teórico e experimental do complexo processo de enovelamento de proteína [11, 12, 13, 14].

A abordagem é feita a partir de uma descrição estatística do enovelamento e é considerada como a “Nova Visão” [15, 16] da cinética de enovelamento de proteína. A superfície de energia (*energy landscape*) representa a energia livre interna de todos os estados conformacionais, com base energética dominante. A maior parte da topografia da superfície de energia pode ser dita afunilada e suas seqüências de proteínas são consideradas bem-desenhadas [17].

As idéias sobre o funil são utilizadas para o entendimento do mecanismo cinético de rearranjo das estruturas da proteína durante o enovelamento. Para isso, Leopold [11] considerou as seguintes propriedades para uma proteína:

- proteínas enovelam de um estado aleatório passando por um estado colapsado seguido de um rearranjo;
- o rearranjo ocorre de maneira difusiva e segue uma brusca queda de altas energias para baixas energias conformacionais;
- o rearranjo acontece depois do colapso entre conformações que são geometricamente similares.

Sendo assim, ele define um funil de enovelamento como uma coleção de estruturas colapsadas geometricamente similares, umas das quais é termodinamicamente estável em relação às outras (veja Figura 1.1). O funil de estruturas de uma proteína depende de sua seqüência primária de aminoácidos, pois é nela que está toda a informação necessária para se enovelar. Uma proteína é dita “enovelável” quando esta tem um único estado nativo termodinamicamente acessível em seu funil de enovelamento.

O potencial termodinâmico que pode descrever um processo a pressão e temperatura constante, como o enovelamento de proteína em meio biológico, é a energia livre de Gibbs [18]. Por isso o sistema proteína-solvente é descrito em função da energia livre e do espaço de conformações da proteína. Com isso, podemos obter uma descrição detalhada da superfície de energia de uma proteína

especificando sua energia livre média sobre a coordenada de reação em função das coordenadas de cada átomo na proteína. A superfície de energia apresentará mínimos locais [13] que são pequenas excitações de energias correspondentes a mudanças conformacionais locais e individuais como rotação de cadeias laterais. Essas energias são da ordem de $k_B T$, ou seja, flutuações de energia dos átomos da proteína. A proteína visita esses mínimos locais muito mais rapidamente do que visita o mínimo global de energia.

Dessa forma, a dinâmica de enovelamento pode ser considerada como sendo um processo em que o ensemble de estruturas parcialmente enoveladas caminham para a estado nativo com sua energia livre e entropia associada [19]. A superfície de energia pode ser associada a um funil parcialmente rugoso (veja Figura 1.1). Para uma superfície de energia desse tipo, temos muitas estruturas com altas energias e poucas com baixas energias. Quanto mais próximo o sistema encontra-se do estado nativo, mais baixa é a energia dessas estruturas. Uma proteína no estado desenovelado encontra seu estado nativo dinamicamente através dos caminhos presentes neste funil de energia que levam ao enovelamento [20] (veja Figuras 1.1 e 1.2). Esses caminhos são determinados pela forma e rugosidade do relevo da superfície de energia de uma proteína. À medida que a proteína se aproxima do estado nativo, mais similar é sua estrutura com a estrutura do estado nativo e menor é o número de estados acessíveis. Vários experimentos têm sido feitos estudando esses caminhos que levam a proteína a se enovelar [21]. Além disso, vários experimentos e simulações computacionais têm sido implementados para encontrar o relevo da superfície de energia de peptídeos e proteínas [22, 23]. Desse modo vencemos o Paradoxo de Levinthal [10] que é a improbabilidade estatística da proteína encontrar seu estado nativo em tempo funcional por tentativas aleatórias.

As fases termodinâmicas são definidas pela temperatura do sistema. Em altas temperaturas, as estruturas com altas energias predominam e à medida que a temperatura diminui, estados de baixa energia são ocupados. Cineticamente,

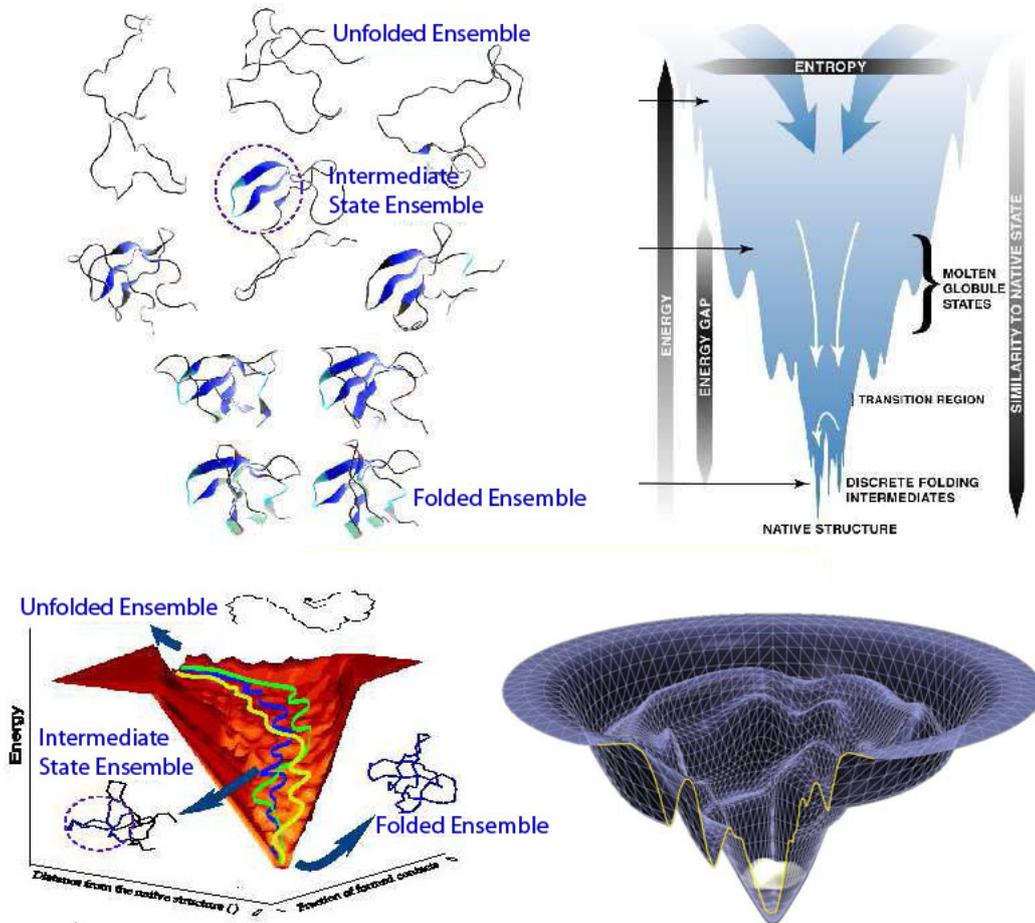


Figura 1.1: No topo a esquerda, temos o funil de estruturas proposto por Leopold [11]. Ensemble de estruturas no estado desenovelado que passam por um estado intermediário ao enovelarem antes de atingirem o ensemble de estruturas enoveladas. No topo a direita, é apresentado um esboço qualitativo da superfície de energia (*energy landscape*) de uma proteína. O eixo horizontal representa a entropia, o vertical esquerdo, a energia total e o direito, o grau de similaridade com o estado nativo. Abaixo a esquerda, temos algumas rotas de enovelamento para uma determinada proteína. Abaixo a direita, é apresentada a superfície de energia com a energia no eixo vertical e o espaço conformacional nos outros. *Figura de J. N. Onuchic apresentada no XXXVII Latin-American School of Physics, 2006.*

abaixo de uma certa temperatura, denominada de temperatura de transição, o sistema atinge certos estados de baixa energia onde fica armadilhado. Uma vez que sua energia térmica não é suficiente para ultrapassar barreiras de energia. Então podemos generalizar, para uma pequena proteína em solvente, duas fases termodinâmicas:

- fase desenovelada com alta entropia e alta energia;
- fase enovelada com baixa entropia e baixa energia.

A rugosidade da superfície de energia é determinada pelas competições nas interações da função energia de uma determinada proteína [13]. Essa competição é chamada “frustração” e é utilizada também em sistemas de vidros de spin por exemplo, um sistema magnético no qual os spins estão arranjados aleatoriamente em uma liga metálica diluída. Nesse paradigma, as energias de interação entre os spins também são distribuídas aleatoriamente. Neste caso, as interações entre os spins tenderão a ser igualmente freqüentes e os spins serão aleatoriamente ferromagnético (os spins querendo apontar para a mesma direção) e antiferromagnético (os spins querendo apontar para direções opostas). Essas duas tendências nos conflitos locais não podem ser satisfeitas em qualquer arranjo de orientações dos spins. Sendo assim, é quase que impossível favorecer um tipo de configuração, isto é, minimizar a energia do sistema satisfazendo completamente todas as interações locais e então dizemos que o sistema é energeticamente “frustrado”.

Consideremos o efeito hidrofóbico, que podemos imaginar como sendo uma interação da cadeia com o solvente que favorece a formação de contatos entre pares hidrofílicos ou pares hidrofóbicos de aminoácidos. Uma cadeia polipeptídica é formada por seqüências de aminoácidos hidrofóbicos e hidrofílicos. O polipeptídeo terá a tendência de formar pares hidrofóbicos que estão distantes na cadeia. Por causa das restrições e conectividade da cadeia, o polipeptídeo também formará pares dissimilares, e então, desfavoráveis. Proteínas reais são

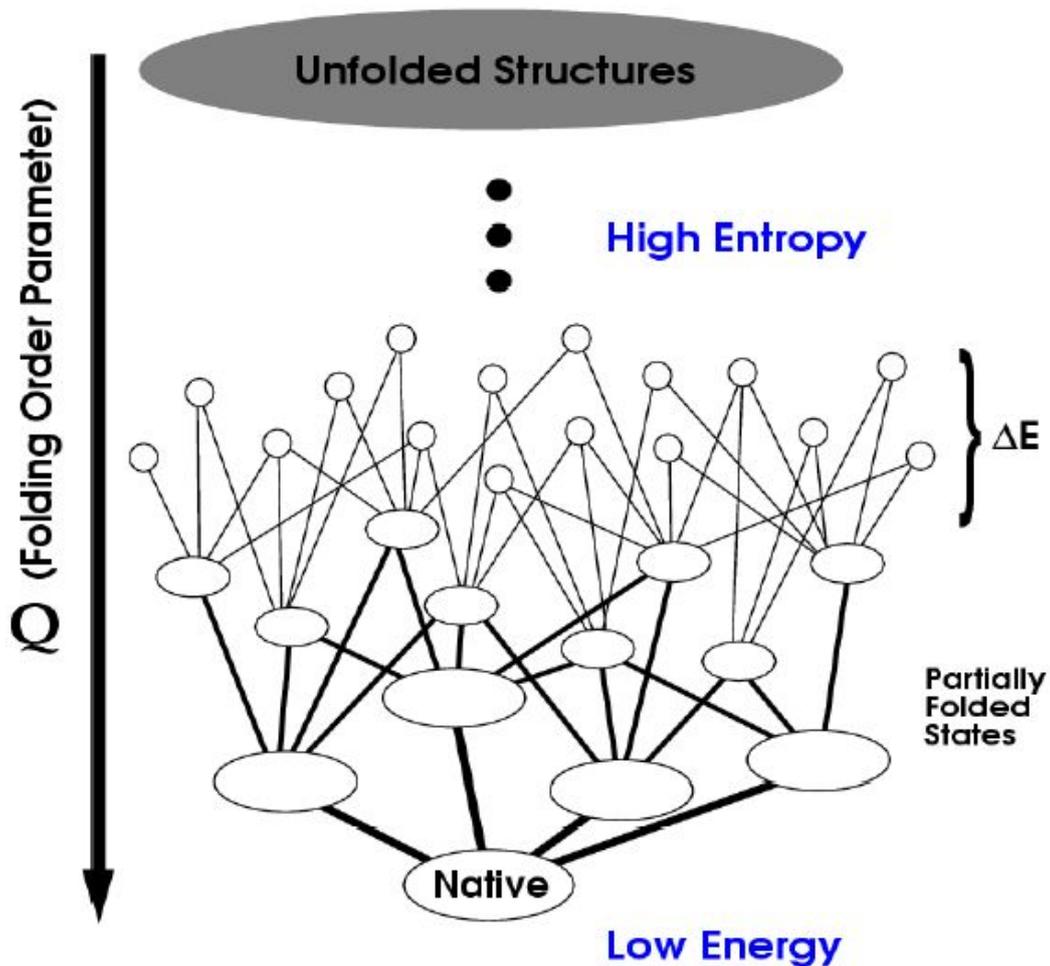


Figura 1.2: Esquema de conformações de uma proteína com relação ao grau de similaridade com o estado nativo. No eixo horizontal temos a entropia e no vertical, a coordenada de reação, ou seja, grau de similaridade da estrutura com o estado nativo. No topo da figura existe alta entropia, pois o estado desenovelado possui um ensemble de estruturas muito grande. Quando a proteína se enovela, move-se para baixo na coordenada de reação, dessa forma o espaço conformacional e a energia reduzem drasticamente e a proteína encontra seu estado nativo. *Figura de J. N. Onuchic [14].*

sistemas frustrados.

Outro tipo de frustração advém do fato de que as moléculas devem ultrapassar a barreira do potencial de energia para mudar de uma estrutura para outra. Se essa energia é muito alta para ser atravessada em um curto intervalo de tempo, dizemos que as duas estruturas não estão “dinamicamente ligadas”. Duas estruturas podem ser parecidas entre si e, mesmo possuindo mesma energia livre, uma não pode mudar para a outra em um tempo razoável. Isso acontece devido às restrições geométricas da cadeia polimérica que não permitem atravessar ela mesma, criando uma enorme barreira de energia. Nesse caso, dizemos que temos frustração geométrica [24].

Um sistema com superfície de energia rugosa exhibe transição de fase. Quando a temperatura desse sistema é diminuída ele tende a ocupar estados de baixa energia, e na temperatura de transição fica armadilhado em um desses estados. A transição para o estado armadilhado é acompanhada de uma considerável diminuição de movimentos do sistema assim que ele tenta ultrapassar a alta barreira de energia existente. Esse caso é conhecido como transição de vidro. Sistemas com superfície de energia rugosa tipicamente apresentam transição de vidro análogo ao que ocorre quando líquidos são supercongelados abaixo de seu ponto de congelamento.

Bryngelson e Wolynes (BW) [25], utilizando conceitos do modelo de energia aleatória desenvolvido por Derrida [26, 27] para estudar sistemas de vidros de spins, descreveram o enovelamento de proteína como esses sistemas. Com isso, encontraram que as proteínas também exibem transição para a fase de vidro. Essa transição ocorre em uma temperatura característica chamada de temperatura de vidro (T_g). Se o sistema está a uma temperatura abaixo de T_g então dizemos que o sistema está na sua fase de vidro. Abaixo de T_g , a entropia da cadeia é zero porque ela está armadilhada em um dos seus estados de energia mais baixos, a proteína possui acesso cinético a poucas estruturas, e o tempo para sair desse estado é muito grande. A temperatura de enovelamento, T_f , pode ser definida

como a temperatura na qual a população das configurações enoveladas é igual as populações em todas as outras configurações [13]. BW também mostraram que boas seqüências são aquelas que procuram maximizar a razão T_f/T_g , ou seja, a temperatura de enovelamento T_f deve ser maior e o mais afastada possível da temperatura de vidro T_g . Isso porque se a proteína estiver em uma temperatura próxima de T_f não haverá a possibilidade de atingir T_g e impossibilitar a proteína de atingir seu estado nativo.

A conexão entre a teoria de superfície de energia e proteínas reais já está bem estabelecida para pequenas proteínas que se enovelam rapidamente, cujo enovelamento é da ordem de milisegundos e tem domínio de enovelamento único, ou seja, elas possuem dois estados termodinâmicos com um único e bem definido funil [19]. Uma nova geração de experimentos (espectroscopia dinâmica em NMR, engenharia de proteínas, enovelamento acompanhado a laser, fenômenos ultrarrápidos, entre outros) têm contribuído para o desenvolvimento e aperfeiçoamento dessa teoria.

1.4 A equação de difusão

A teoria sobre funis de superfície de energia tem como resultado a descrição simples do processo de enovelamento de proteína. Esse processo é visto como um processo coletivo, que se auto organiza e que pode ser descrito como uma complexa reação química. Ele não ocorre por uma seqüência de passos obrigatórios (*pathways*) intermediários (micro-estados), mas por uma multiplicidade de rotas pelo funil de enovelamento descrito por macro-estados (*ensemble*) [28] como já discutido nas seções anteriores.

O processo de enovelamento de proteína é estudado como uma dinâmica no funil de energia em que a progressiva formação de um ensemble de estruturas parcialmente enoveladas é impedida de alcançar imediatamente a estrutura

nativa por ficar temporariamente aprisionada em mínimos locais da superfície de energia. Podemos acompanhar a evolução do enovelamento no funil através de parâmetros de ordem macroscópicos, ou seja, uma coordenada de reação efetiva. A coordenada de reação descreve o progresso, a cinética do enovelamento em função da conformação da proteína. São várias as discussões à respeito de qual é a melhor coordenada de reação que descreve o enovelamento (veja a referência [28] e suas referências). Uma boa coordenada de reação não deve somente ter uma proximidade com a cinética do sistema, mas deve também ser de fácil cálculo nas simulações e nas medidas em experimentos. A coordenada de reação será o número de contatos corretos (nativos) (Q) numa conformação. Esta coordenada é utilizada na teoria de superfície de energia e descreve o quão próximo da estrutura nativa está uma determinada conformação. Se Q é igual a 1 significa que a conformação é a estrutura nativa e a proteína está enovelada, se Q é 0 a proteína está totalmente desenovelada¹. Qualquer valor entre 0 e 1 significa que a conformação é de um estado intermediário no enovelamento (veja Figura 1.2). Para proteínas reais, seriam necessários mais parâmetros de ordem para quantificar o grau de similaridade com o estado nativo como grau de colapso, porcentagem de hélices na estrutura, ângulos de diedros, entre outros.

Como caracterizado na seção 1.3, a superfície de energia é determinada pela energia livre e pela entropia e o enovelamento é uma competição entre esses dois parâmetros. À medida em que a cadeia se move através da coordenada de reação, energeticamente do topo para a parte mais baixa do funil e fica mais similar a estrutura nativa, a entropia conformacional (número de estados acessíveis) é reduzida. O gradiente de energia determina a média com que a proteína percorre o funil até o estado enovelado. Isso é determinado por movimentos estocásticos cuja estatística depende dos saltos através de mínimos locais. Para uma primeira aproximação, esse processo pode ser considerado como uma difusão [25, 29]. O

¹Aqui, a coordenada de reação Q está normalizada podendo assumir qualquer valor entre 0 e 1 inclusivamente.

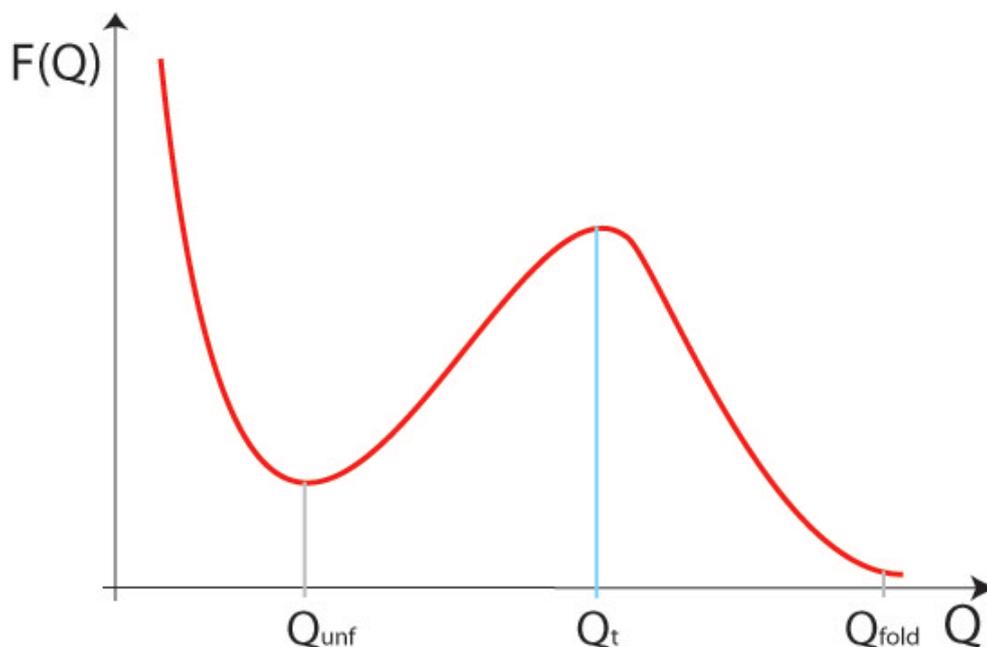


Figura 1.3: Esboço da energia livre para temperatura de enovelamento T_f em função da coordenada de reação, no qual Q é o grau de similaridade com o estado nativo. A linha vertical em Q_t corresponde à região de transição do estado desenovelado para o estado enovelado.

tempo de enovelamento, então, dependerá da rugosidade da superfície de energia e da dificuldade de superar a barreira termodinâmica de energia imposta por esta rugosidade (veja Figura 1.3).

A rugosidade da superfície de energia em qualquer estágio do enovelamento atua com várias barreiras diminuindo a velocidade com que a proteína percorre o funil e o movimento Browniano imposto, uma vez que, na temperatura de enovelamento T_f a trajetória da coordenada de reação coletiva é descrita pelo movimento Browniano. BW mostraram que para uma primeira aproximação, o tempo de enovelamento poderia ser calculado agrupando em camadas estados com mesmo valor na coordenada de reação. A equação de difusão descreve o fluxo de probabilidade através dessas camadas que é derivada assumindo que a

coordenada de reação pode mudar somente por pequenos passos relativos. Em uma determinada temperatura T , a equação de difusão que descreve a população dessas várias camadas é do tipo Fokker-Planck [30] e muda com o tempo de acordo com

$$\frac{\partial P(Q, t)}{\partial t} = \frac{\partial}{\partial Q} \left\{ D(Q, T) \left[\frac{\partial P(Q, t)}{\partial Q} + P(Q, t) \frac{\partial \beta F(Q, T)}{\partial Q} \right] \right\} \quad (1.1)$$

A direção média do fluxo é dada pelo gradiente de energia livre. O coeficiente de difusão configuracional local $D(Q)$ depende da rugosidade da superfície de energia e reflete o tempo para escapar de seus mínimos locais no sistema próximo da transição para a fase de vidro.

A energia livre $F(Q, T)$ depende da competição entre dois termos, o termo energético que diminui à medida que ocorre o envelhecimento e o termo entrópico $-TS(Q)$ que aumenta com o envelhecimento. A energia livre é fortemente dependente da temperatura. Em altas temperaturas, a função energia livre possui um vale com uma barreira de energia muito grande e o tempo de envelhecimento aumenta exponencialmente. Na temperatura de envelhecimento T_f , a energia livre possui dois estados com uma pequena barreira termodinâmica que tem origem na competição entre o termo entrópico e a energia, à medida que o sistema tende ao envelhecimento. A baixas temperaturas, o tempo de envelhecimento se comporta como uma exponencial decrescente. O coeficiente de difusão $D(Q, T)$ depende da temperatura e da coordenada de reação e, dessa forma, o tempo de envelhecimento dependerá das forças termodinâmicas que dirigem o envelhecimento.

O tempo de envelhecimento pode ser escrito como uma integral dupla [29]

$$\tau_f = \int_{Q_{unf}}^{Q_{fold}} dQ \int_0^Q dQ' \frac{e^{\beta\{F(Q)-F(Q')\}}}{D(Q)} \quad (1.2)$$

Para uma função de energia livre $F(Q)$ com uma barreira bem definida, assim como em T_f (veja Figura 1.3), Q_{unf} corresponde ao vale de $F(Q)$ onde a cadeia está desenovelada e Q_{fold} é a coordenada de reação para o outro vale onde a cadeia está enovelada. A integral dupla 1.2 pode ser aproximada pela lei de Kramers [31]

$$\tau_f = \left(\frac{2\pi}{\beta}\right)^{\frac{1}{2}} \frac{1}{D_0 \omega_{unf} \bar{\omega}_{fold}} e^{\beta\{\bar{F}(Q_t) - F(Q_{unf})\}} \quad (1.3)$$

em que

$$\bar{F}(Q) = F(Q) - T \log \left[\frac{D(Q)}{D_0} \right] \quad (1.4)$$

e ω_{unf} é a curvatura em Q_{unf} e ω_{fold} a curvatura do topo da barreira de energia. D_0 é o coeficiente de difusão efetivo equivalente a uma superfície de energia lisa.

O coeficiente de difusão depende da superfície de energia e dos movimentos locais permitidos da proteína. BW analisaram a difusão configuracional assumindo o limite de uma superfície de energia rugosa descorrelacionada e as regras de Metropolis [32]. O coeficiente de difusão dependerá também da temperatura em que a proteína deverá escapar dos mínimos locais da superfície de energia.

A temperatura de transição de vidro, no equilíbrio, será dada por

$$T_g(Q) = \frac{\Delta E^2(Q)}{[2S^*(Q)]^{\frac{1}{2}}} \quad (1.5)$$

em que $S^*(Q)$ é a entropia configuracional em Q . Na temperatura T_g , o coeficiente de difusão diminui com o número total de estados, sugerindo a possibilidade do Paradoxo de Levinthal. Abaixo de T_g , não podemos considerar a energia como uma média, pois a flutuação na energia livre é considerável e dependerá fortemente do modelo e da seqüência do heteropolímero [33]. Sendo assim, o lento processo de envelhecimento pode ser melhor explicado por poucos caminhos cinéticos predominantes do que pelas leis de médias estatísticas para eventos rápidos. A dependência de D da temperatura, da competição da entropia e da energia, leva a uma curva parabólica para os tempos de envelhecimento em função de β . Leite [34, 35] mostrou que, para uma superfície de energia de um sistema de polarização de solvente, gradualmente as flutuações dominam a cinética abaixo da temperatura T_g análogo à teoria de BW para a superfície de energia de uma proteína.

Dessa forma, os tempos de enovelamento dependem da energia livre e do coeficiente de difusão que incorpora as múltiplas difusões através da barreira de energia. A escolha de uma coordenada de reação também poderá influenciar na dependência do coeficiente de difusão e da forma da energia livre de um heteropolímero [33].

1.5 Motivação

Como discutido nas seções anteriores, o enovelamento de proteína tem sido um problema fundamental na biofísica molecular para a compreensão de estruturas e funções biológicas. Determinar os tempos de enovelamento de uma proteína não é um problema trivial. Vários têm sido os modelos e teorias que tentam explicar e calcular esses tempos de enovelamento. Na presente discussão, a dinâmica do enovelamento segue a equação de difusão do tipo Fokker-Planck e os tempos de enovelamento podem ser calculados através de potenciais efetivos e do coeficiente de difusão.

Em geral, o coeficiente de difusão é mantido constante como uma aproximação para o cálculo dos tempos de enovelamento teóricos resultando em uma discordância com os tempos de enovelamento de simulação. O coeficiente de difusão depende da coordenada de reação e da temperatura do sistema e, por isso, ocorre esta discrepância.

Através de simulação computacional e utilizando modelos simples, calcularemos o coeficiente de difusão em função da coordenada de reação e da temperatura resultando em cálculos mais precisos.

Capítulo 2

Modelo

2.1 A importância dos Modelos Minimalistas

A estrutura das proteínas, em nível atômico, é muito complexa. Essa possui milhares de átomos interagindo entre si e com o meio onde se encontra. Se olharmos para uma proteína no seu estado nativo, veremos que ela não possui somente uma estrutura, mas uma coleção de estruturas bem definidas com alto grau de similaridade em constante dinâmica. Desde os experimentos de Pauling [5], sabemos que a estrutura de uma proteína não é aleatória, essas possuem regiões periódicas (hélices e folhas) e aperiódicas (curvas aleatórias). Isso torna a proteína um sistema muito interessante e difícil de ser estudado do ponto de vista teórico. Se o sistema proteico fosse completamente ordenado, seria simples e de fácil descrição. Se fosse completamente aleatório, muitas propriedades desse tipo de sistema já estaria descritas. Porém, as proteínas estão entre esses dois tipos de sistema. Mais uma vez, percebe-se que as proteínas são sistemas complexos para serem estudados teoricamente.

Entretanto, examiná-las por meio de modelos mais simples (minimalistas), que possuem algumas das propriedades das proteínas (*protein-like*), nos tem

ajudado a entender o processo de enovelamento. Modelos simples são exatos. Isso porque, podemos calcular sua função de partição e a partir da qual propriedades físicas podem ser obtidas sem considerações ou aproximações. Esses modelos são representações rudimentares de proteínas envolvendo poucos parâmetros, nos quais se perde a representação atômica e geométrica de uma proteína real, mas se ganha na caracterização de todas as possíveis seqüências de aminoácidos (espaço de seqüências) e todas as possíveis conformações (espaço conformacional) de uma dada seqüência [28].

A simulação computacional de modelos minimalistas tem ajudado no entendimento do enovelamento de proteínas. Isso porque, a simulação de modelos de proteínas reais envolveria uma infinidade de parâmetros e interações atômicas e, como consequência, o tempo computacional tenderia à idade do universo, mesmo usando os computadores mais modernos. Os modelos minimalistas, por possuírem poucos parâmetros, gastam um tempo computacional ainda relevante, porém mensurável.

Os trabalhos teóricos em proteínas começaram, em 1975, com as simulações em modelos de rede em duas dimensões ($2D$) de $G\bar{o}$ e colaboradores [36] e, por isso, esse modelo recebeu seu nome. Esse modelo tem as propriedades de uma proteína que a faz se enovelar para uma estrutura única. O potencial de interação utilizado não é do tipo convencional, um hamiltoniano; ele pode ser dito não-físico, contém a informação necessária para enovelar para a estrutura nativa, por possuir muitas propriedades que são observadas em proteínas reais. Potenciais desse tipo são chamados de potenciais de $G\bar{o}$ e são muito utilizados no estudo do enovelamento por possuírem essas propriedades.

Logo após em 1989, Chan e Dill [37] construíram um novo modelo de rede. Criaram o modelo HP (Hidrofóbico-Polar) com um potencial físico mais realístico inspirado no modelo de $G\bar{o}$. Nesse modelo, os dois tipos de monômeros possíveis representam os aminoácidos hidrofóbicos (H) e o polares (P). Ele é capaz de simular o efeito hidrofóbico e de prever quais seqüências são capazes

de se enovelar. Em 1990, Chan e Dill [38] estenderam suas idéias para a rede $3D$ com uma cadeia cúbica maximamente compacta de 27 monômeros, com a finalidade de explorar as conseqüências de um potencial físico sem aproximações e as propriedades do espaço conformacional e de seqüências por ele gerados.

Em 1994, Socci e Onuchic [39] basearam-se no modelo de rede cúbica de Chan e Dill e nos potenciais de $G\bar{o}$ para estudar o modelo de rede cúbica de 27 monômeros. Esse modelo permite uma visão mais detalhada em simulação de proteínas e será descrito e utilizado nas próximas seções.

Modelos de rede possuem limitações ao representar a energia de superfície de proteínas reais. A superfície de energia pode ser utilizada como uma ferramenta teórica para o entendimento das propriedades do enovelamento. Porém, os modelos de rede são muito bons como modelo genérico, possuem muitas funções e propriedades e têm provado, desde sua criação até hoje, que podem prever em grande detalhe o mecanismo de enovelamento de proteína.

2.2 Modelo de rede cúbica

As idéias desenvolvidas por BW [29] serão utilizadas no modelo de rede cúbica de 27 monômeros de Socci e Onuchic, interpretando-o com base na teoria da superfície de energia e de funil de estruturas. Nesse modelo, estudaremos a dependência do coeficiente de difusão nos tempos de enovelamento de um heteropolímero em um sistema de rede cúbica. Como discutido na seção anterior, esses modelos apresentam bom embasamento teórico e uma vasta literatura [13, 14, 17, 33, 38, 39, 40, 41, 42, 43].

No nosso sistema, simplificamos uma pequena proteína globular real por um heteropolímero de 27 monômeros (equivalentes aos aminoácidos de proteínas reais) em uma rede cúbica $3D$. Os monômeros estão ligados covalentemente ao longo da cadeia. Cada monômero pode ocupar somente um sítio na rede

(condição de volume excluído) (veja Figura 2.1). A estrutura inicial desenovelada do heteropolímero é criada aleatoriamente na rede. A estrutura maximamente compacta (enovelada) é um cubo de tamanho $3 \times 3 \times 3$ com número máximo de contatos igual a 28 (veja Figura 2.2).

Shakhnovich e Gutin [44] enumeraram exaustivamente todas as conformações maximamente compactas desse polímero com 27 monômeros, encontrando um número total de 103346 cubos. Uma conformação maximamente compacta é um estado de menor energia, se escolhermos um potencial que favoreça a formação de contatos. Escolhendo tal potencial, conseguimos determinar a estrutura nativa de uma determinada seqüência enumerando os cubos. Também, a degenerescência do estado de energia mais baixa pode ser determinada. Como queremos um heteropolímero que tenha as propriedades de uma proteína (*protein-like*), as seqüências deverão possuir um único estado de mais baixa energia, um estado de energia mínima não-degenerado.

Sendo assim, devemos escolher um potencial que favoreça a formação de estados compactos, que faça a cadeia colapsar e enovelar. A força dominante no enovelamento de proteínas é o efeito hidrofóbico [45]. Esse efeito é uma interação de muitos corpos entre as cadeias laterais hidrofóbicas da proteína (resíduos de aminoácidos) e o solvente (água). O efeito hidrofóbico faz com que a cadeia sofra um colapso, ou seja, a formação de um núcleo hidrofóbico. No nosso problema, modelamos esse efeito escolhendo um potencial que favoreça a formação de contatos entre quaisquer dois monômeros. Monômeros que são vizinhos próximos na rede interagem com uma certa energia. O potencial também deve assegurar que a seqüência tenha um único estado de baixa energia. Entretanto, não queremos um homopolímero, então, adicionamos um termo que dependa do tipo de monômero que está interagindo, um termo energético de interação que distingue se os dois monômeros em contato são do mesmo tipo ou não. A energia potencial será

$$E = \sum_{\langle i,j \rangle} H_{t_i,t_j} \quad (2.1)$$

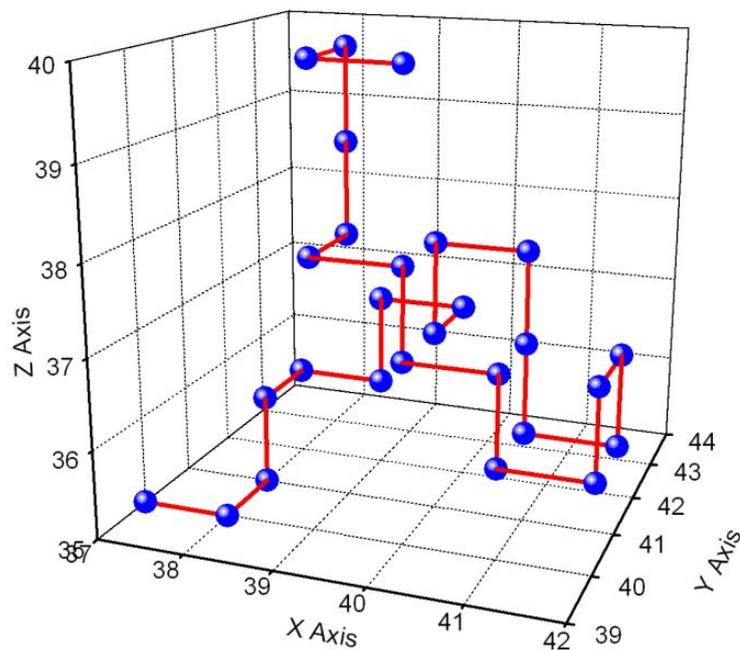


Figura 2.1: Estrutura aleatória inicial gerada na rede para uma seqüência de 27 monômeros.

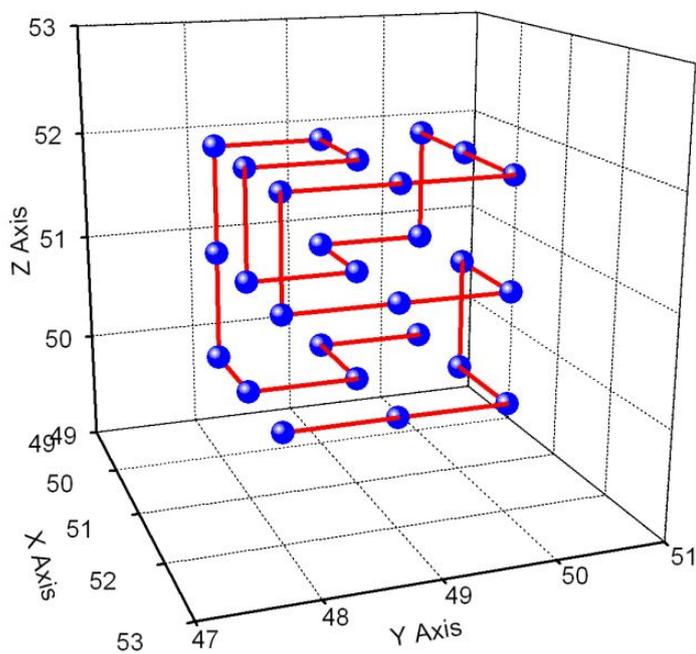


Figura 2.2: Estrutura maximamente compacta da seqüência de 27 monômeros. A estrutura é um cubo $3 \times 3 \times 3$ com máximo de 28 contatos não-covalente.

no qual a soma é efetuada sobre todos os pares vizinhos na rede, excluindo os pares ligados covalentemente. A questão central do enovelamento é determinar a estrutura da proteína a partir de sua seqüência. Para isso, precisamos de pelo menos dois tipos de monômeros. No presente estudo, como estamos utilizando um modelo simples, usaremos apenas três tipos. O monômero i é do tipo t_i que poderá ser A , B ou C . H_{t_i,t_j} é a matriz de interação entre os contatos dada por

$$H_{t_i,t_j} = \begin{matrix} & A & B & C \\ A & \left(\begin{matrix} E_l & E_u & E_u \end{matrix} \right) \\ B & \left(\begin{matrix} E_u & E_l & E_u \end{matrix} \right) \\ C & \left(\begin{matrix} E_u & E_u & E_l \end{matrix} \right) \end{matrix}. \quad (2.2)$$

E_l é a energia entre dois monômeros de mesmo tipo em contato e E_u entre monômeros diferentes.

A dinâmica de enovelamento do nosso modelo de rede dependerá dos tipos de movimentos adotados para a cadeia [39]. Diferentes tipos de movimentos implicam em diferentes comportamentos cinéticos. Os tipos de movimentos escolhidos devem ser tais que respondam à questão do tempo de enovelamento e que sejam o mais realístico possível. O conjunto de movimentos escolhido para a simulação é baseado na dinâmica de física de polímeros. Esse conjunto consiste em movimentos locais que preserve a ligação covalente e cada monômero pode ocupar somente um sítio na rede. Os movimentos permitidos movimentam o monômero para um sítio adjacente não ocupado e são três (veja Figura 2.3):

Movimento de fim de cadeia (*end move*) movimentam um monômero que se encontra no final da cadeia.

Movimento de canto (*corn move*) movimentam um monômero de canto na estrutura para um canto vago.

Movimento de manivela (*crankshaft move*) é o único a fazer o movimento de dois monômeros simultaneamente.

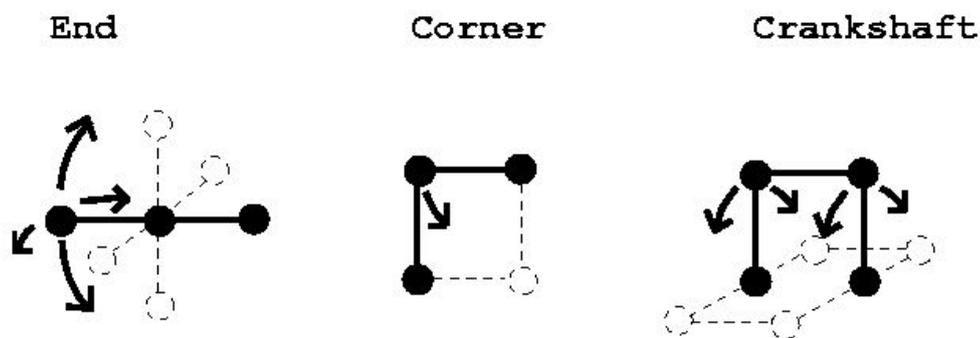


Figura 2.3: Tipos de movimentos permitidos na simulação. Esses movimentos são movimentos locais. Os círculos pontilhados representam os sítios na rede desocupados e que permitem o movimento. Os círculos cheios são monômeros ligados covalentemente. No movimento de final de cadeia e de manivela, o sítio a ser ocupado é escolhido aleatoriamente. O movimento de manivela é o único a movimentar dois monômeros simultaneamente.

Toda a discussão sobre os movimentos permitidos em modelos de rede que estamos utilizando já está bem estabelecida [33, 39, 42]. Esses movimentos são suficientes para responder apenas a questão da cinética do tempo de enovelamento. Se estivéssemos interessados em estudar, como exemplo, a dinâmica de formação de estruturas secundárias (hélices ou folhas), deveríamos utilizar movimentos mais complexos e globais. Nosso modelo é muito simples e a seqüência é muito pequena e por isso podemos fazer um movimento de cada vez.

A partir de uma estrutura crescida aleatoriamente no espaço da rede, uma configuração inicial qualquer, são realizados os movimentos locais em monômeros escolhidos aleatoriamente. A escolha dos monômeros tem o mesmo peso. Um movimento é feito usando o algoritmo de Metropolis de Monte Carlo [32, 46]. Se for um monômero no final da cadeia, então um sítio vizinho na rede é também escolhido aleatoriamente. Se não for um monômero final, então ele pode fazer um movimento de canto ou de manivela, dependendo das posições dos vizinhos

ao longo da cadeia. Se o movimento escolhido é o de manivela, a direção do movimento é escolhida aleatoriamente com mesma probabilidade para todas as direções. Pode-se escolher uma probabilidade diferente para a realização de movimentos de manivela. Se o monômero escolhido violar o vínculo de volume excluído, tentando movimentar o monômero para um sítio ocupado, a antiga configuração é restaurada e é contado esse movimento como um passo de Monte Carlo (para o cálculo da média de tempos) e então, um novo monômero é escolhido. Se o movimento é permitido, isto é, o sítio na rede está vazio, a energia da nova configuração é calculada e comparada com a energia da configuração original. Se a energia diminuir, o movimento é aceito incondicionalmente e é contado um passo. Se a energia aumentar, o movimento será aceito de acordo com a usual probabilidade de Boltzmann

$$P = \exp\left[-\frac{E_{new} - E_{old}}{k_B T}\right] \quad (2.3)$$

em que E_{new} é a energia da nova configuração, E_{old} é a energia da configuração original e T é a temperatura do sistema. As unidades são escolhidas de tal forma que a constante de Boltzmann seja unitária ($k_b = 1$). Se um número aleatório entre zero e um gerado é menor que P , o movimento é aceito. Caso contrário, nenhum movimento é realizado e a configuração original é restaurada. Um movimento sendo aceito ou não, uma unidade de tempo (passo de Monte Carlo) é considerada.

Não há nenhuma conexão direta dos passos de Monte Carlo de uma simulação (escalas de tempo computacional) com escalas de tempo físico real de uma proteína [39]. Apesar disso, os passos de Monte Carlo podem ser usados para calcular tempos pois se tornam realísticos quando comparado tempos de enovelamento (tempos curtos) com tempos de dinâmica (tempos longos). As propriedades qualitativas, como os comportamentos cinéticos e termodinâmicos, continuarão sendo as mesmas de uma proteína real.

2.3 Os parâmetros de hidrofobicidade

Numa proteína enovelada, o núcleo consiste prioritariamente de resíduos hidrofóbicos e sua superfície de resíduos hidrofílicos. Esse ordenamento ocorre devido ao efeito hidrofóbico, originário do enovelamento de proteínas em água. Variações nas propriedades do solvente terão efeitos profundos na energia de interação dos grupos hidrofóbicos da cadeia.

Como discutido na seção 2.2, no modelo de rede cúbica com uma cadeia de 27 monômeros, utilizamos um potencial que simula o efeito hidrofóbico que faz o heteropolímero colapsar e enovelar. O potencial utilizado é dado pela equação 2.1, que também pode ser reescrita na forma [40]

$$E = N_l E_l + N_u E_u, \quad (2.4)$$

em que N_l é o número de contatos não-covalentes entre monômeros do mesmo tipo (*like contacts*) e N_u é o número de contatos não-covalentes entre monômeros de tipos diferente (*unlike contacts*). Podemos analisar o comportamento do sistema através de parâmetros de hidrofobicidade [47]:

$$\bar{E} = \frac{1}{2}(E_l + E_u), \quad (2.5)$$

$$E_{het} = (E_u - E_l). \quad (2.6)$$

\bar{E} dirige a formação dos contatos ou compactação da cadeia. Se for menor que zero, a formação de contatos é favorecida e a cadeia colapsa. Ele simula o efeito hidrofóbico. E_{het} determina a rugosidade da superfície de energia, a heterogeneidade dos diferentes resíduos e dirige a seqüência a alcançar o estado único de menor energia.

Em nosso trabalho, estamos interessados no comportamento do modelo com as mudanças na força que dirige o heteropolímero ao colapso. Para isso, utilizamos um parâmetro que mede o grau de colapso k

$$k = -\frac{\bar{E}}{E_{het}}. \quad (2.7)$$

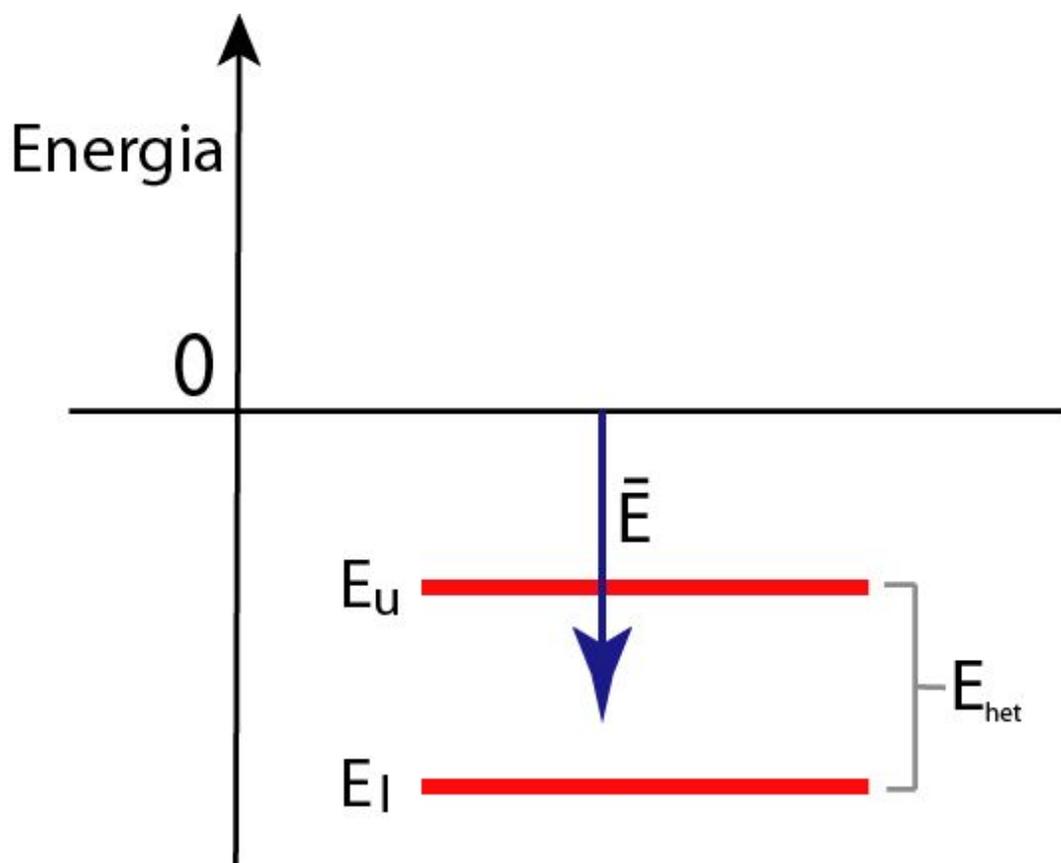


Figura 2.4: Diagrama dos níveis de energia do modelo. No diagrama está representada a energia de interação entre os monômeros vizinhos não-covalentes de mesmo tipo (E_l) e de tipos diferentes (E_u). \bar{E} dirige a cadeia à formação dos contatos, a compactação. E_{het} determina a diferença entre os níveis de energia e assegura que a cadeia seja um heteropolímero. Também, $E_l = \bar{E} - E_{het}/2$ e $E_u = \bar{E} + E_{het}/2$.

Ajustando k mudamos o grau de compactação, ou seja, a hidrofobicidade do sistema. Estamos interessados em duas regiões: o limite de alta hidrofobicidade onde $k = 1$ e o limite de baixa hidrofobicidade onde $k = 0$. No enovelamento em regime de alta hidrofobicidade, a cadeia passa por um rápido colapso seguido de um lento enovelamento até atingir o estado nativo. No regime de baixa hidrofobicidade, a compactação e o enovelamento ocorrem simultaneamente em direção ao estado nativo [47]. Proteínas reais exibem estes dois comportamentos. Chahine e colaboradores [48] estudaram como esses parâmetros energéticos e a temperatura influenciam na cinética do modelo de rede.

2.4 Método do histograma

As propriedades termodinâmicas do modelo de rede cúbica são calculadas via dinâmica de Monte Carlo. Uma simulação de Monte Carlo é feita em uma determinada temperatura e as médias termodinâmicas são calculadas. Se desejarmos obter quantidades termodinâmicas em uma outra temperatura, a simulação deve ser refeita. Para cada temperatura, uma nova simulação. Entretanto, é possível obter as grandezas termodinâmicas de outras temperaturas a partir de uma simulação em uma única temperatura. Essa técnica é chamada método do histograma de Monte Carlo [49, 50]. Usando essa técnica podemos calcular uma densidade de estados aproximada do sistema com o qual calculamos qualquer grandeza termodinâmica para uma faixa de temperatura.

Para cadeias extremamente curtas em $2D$ é possível enumerar todas as conformações e obter a função de partição com a qual calculamos qualquer grandeza termodinâmica. Já para uma cadeia de 27 monômeros em uma rede cúbica é impossível enumerar todas as conformações, somente é possível enumerar todas as conformações maximamente compactas [44]. Então, utilizamos o método do histograma para calcular a densidade de estados. Para isso, contamos quantas

vezes ocorreu uma determinada energia na simulação e montamos um histograma de energia. O histograma $h(E, T')$ mede a probabilidade da energia E ocorrer na temperatura T' . Ela é igual a média térmica da densidade de estados [51], pois o sistema é considerado um ensemble canônico com

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

em que $Z(T')$ é a função de partição na temperatura T' , que é dada por

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

na qual $n(E)$ é a densidade de estados na energia E (número de conformação com energia E). k_B é igual a 1 (veja seção 2.2) e T' a temperatura da simulação.

Rearranjando a equação 2.8 encontramos a densidade de estados

$$n(E) = h(E, T')e^{E/T'} Z(T'), \quad (2.10)$$

na qual a função de partição $Z(T')$ é uma constante que deve ser calculada. Para o nosso sistema, é possível calcular $Z(T')$ e obter a densidade de estados. Para isso, precisamos da multiplicidade de algum estado. A seqüência que estudaremos possuiu um estado não-degenerado, o estado fundamental. Isso significa que $n(E_{gs}) = 1$, em que E_{gs} é a energia do estado mais baixo. Com $Z(T')$ determinado, podemos encontrar a densidade de estados $n(E)$ e calcular uma quantidade extensiva, por exemplo, a energia livre usando a equação 2.9 e

$$F(T) = -T \log(Z). \quad (2.11)$$

Para o cálculo de quantidades intensivas, médias termais são determinadas através de

$$\langle \xi \rangle (T) = \frac{\sum_E \xi(E)n(E)e^{-E/T}}{\sum_E n(E)e^{-E/T}} \quad (2.12)$$

e substituindo a equação 2.10 obtemos

$$\langle \xi \rangle (T) = \frac{\sum_E \xi(E)h(E, T')e^{-E/T+E/T'}}{\sum_E h(E, T')e^{-E/T+E/T'}}. \quad (2.13)$$

A equação 2.13 deve ser usada sobre uma certa faixa de temperatura. Em temperaturas muito maiores ou muito menores que a temperatura de simulação os erros no cálculo da densidade de estados (equação 2.10) tornam-se significativos. O sistema é amostrado para uma dada região do espaço de fase em uma dada temperatura. Para simulações onde a temperatura é muito alta, estados de mais baixa energia não são visitados e o espaço de fase não é amostrado. Para baixas temperaturas de simulação, estados com altas energias nunca são visitados. Dessa forma, a densidade de estados não estará correta para regiões do espaço de fase não amostrado (é zero para regiões nunca visitadas). É por isso que existe um intervalo de temperatura no qual é possível fazer extrapolações para uma simulação. Uma solução seria utilizar o método do múltiplo histograma [50] em que se usa várias simulações em temperaturas diferentes para interligar em um único histograma.

No modelo de rede cúbica, para uma seqüência de 27 monômeros, o método do histograma simples satisfaz o intervalo de temperatura de interesse. Isso porque, a largura do histograma de energia é $1/\sqrt{N}$, onde N é o tamanho do sistema. Por o sistema ser pequeno o suficiente para que o histograma amostrasse uma grande região do espaço de fase numa temperatura, o método do histograma nos fornece resultados satisfatórios.

Capítulo 3

Resultados

3.1 Tempos de enovelamento por simulação

No nosso estudo do coeficiente de difusão, utilizamos uma seqüência de 27 monômeros e com código de três letras (ABABBBBCBACBABABACACBACAACAB), conhecida como 0012. Essa seqüência é muito usada na literatura [33, 40, 43] e foi otimamente projetada por Socci e colaboradores [40] com a razão $T_f/T_g = 1.6$, indicando que possui grande estabilidade. Também não possui contatos frustrados (contatos entre monômeros diferentes) na estrutura nativa (cubo $3 \times 3 \times 3$). No regime de alta hidrofobicidade, os parâmetros utilizados para 2.5 e 2.6 são, respectivamente, $\bar{E} = -2$ e $E_{het} = 2$ resultando em $E_l = -3$ e $E_u = -1$. Para o regime de baixa hidrofobicidade, usamos $\bar{E} = 0$ e $E_{het} = 6$ resultando em $E_l = -3$ e $E_u = +3$. O número total de contatos não covalentes para uma estrutura cúbica nativa é 28. Sendo assim, a energia do estado fundamental, de acordo com a função potencial 2.4, será -84 pois $N_l = 28$ e $N_u = 0$. Essa seqüência está longe de ser uma proteína real, porém possui comportamento similar e resultados que podem ser aplicados a uma pequena proteína hélice alfa desde que utilizemos uma lei de correspondência de estados [14]. Para as si-

mulações, as unidades de temperatura e energia são escolhidas convenientemente tais que $k_B = 1$. A temperatura de vidro é $T_g \approx 1$ e as simulações são feitas a partir dessa temperatura crítica. Para o cálculo do tempo de envelhecimento, são feitas 1000 corridas para cada temperatura e calcula-se a média dos tempos. Em cada corrida, é anotado o tempo em que a seqüência atinge pela primeira vez o estado nativo. Nesse momento, ou quando se atinge o tempo máximo de $\tau_{max} = 1.0 \times 10^9$ passos de Monte Carlo, a corrida é parada. τ_{max} é escolhido de tal forma que seja muito maior que o tempo de envelhecimento (cerca de 1000 vezes maior) e que minimize o tempo computacional. A faixa de temperatura em

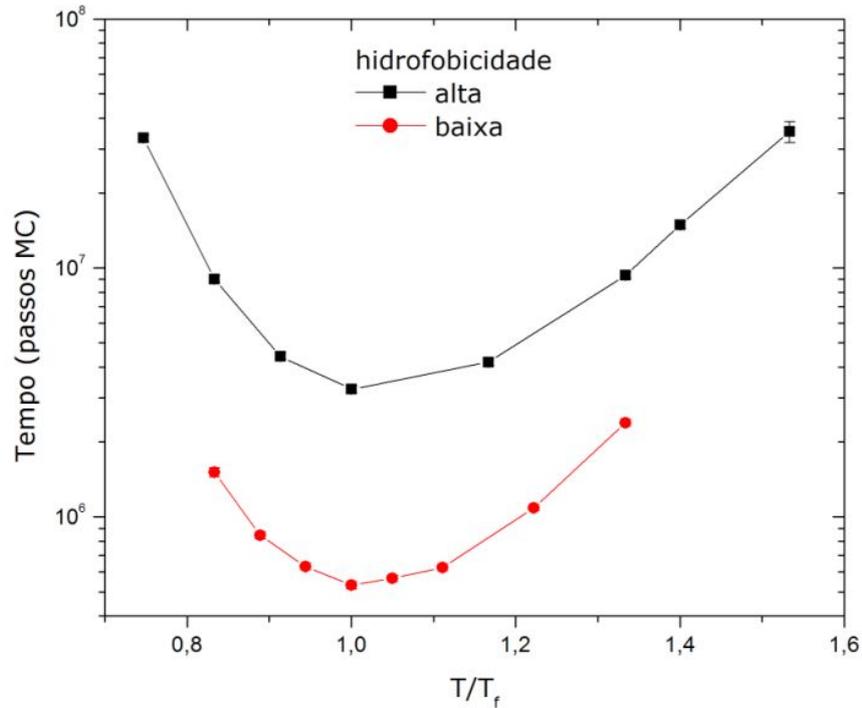


Figura 3.1: Tempo de envelhecimento (τ_f) em função da temperatura para os regimes de alta e baixa hidrofobicidade. O tempo é medido em passos de Monte Carlo da simulação onde está em escala logarítmica. A temperatura para cada regime está normalizada pela temperatura de envelhecimento T_f . O erro é calculado através do desvio médio padrão.

que fizemos as simulações foi $1.1 \leq T \leq 2.4$. A temperatura de enovelamento T_f é encontrada utilizando o método do histograma descrito na seção 2.4. Para a seqüência estudada 0012, a temperatura de enovelamento no regime de alta hidrofobicidade é $T_f = 1.51$ e no regime de baixa $T_f = 1.89$.

A figura 3.1 representa os tempos de enovelamento (τ_f) encontrados para simulações nos dois regimes, alta e baixa hidrofobicidade. A temperatura para os dois regimes está normalizada pela respectiva temperatura de enovelamento, então $T/T_f = 1$ na temperatura de enovelamento. Os tempos de enovelamento têm a forma de “U” nos dois regimes de hidrofobicidade. Os tempos são mais rápidos para baixa hidrofobicidade pois, nesse regime, a medida que vai ocorrendo o enovelamento, os contatos vão se formando de forma correta levando o heteropolímero direto para sua forma nativa. Já no limite de alta hidrofobicidade, primeiro acontece o colapso da cadeia, contatos não favoráveis (frustrados) são criados e depois, lentamente, a cadeia se rearranja para formar a estrutura nativa tendo que abri-la para romper os contatos frustrados e formar os contatos corretos.

3.2 Termodinâmica e cinética

Nesta seção veremos como ocorre a termodinâmica e a cinética do enovelamento de proteínas na rede para os dois regimes: alta e baixa hidrofobicidade. Na figura 3.2 temos a energia livre, para várias temperaturas obtidas nos dois regimes, em função do parâmetro de ordem Q . Esse parâmetro representa a medida do grau de similaridade com a estrutura nativa, número de contatos corretos ou contatos presentes nessa estrutura. As temperaturas estão acima da temperatura de vidro. A energia livre foi calculada de acordo com a densidade de estados usando o método do histograma de Monte Carlo (veja seção 2.4). Podemos notar que o comportamento dos gráficos da energia livre para a proteína nos dois

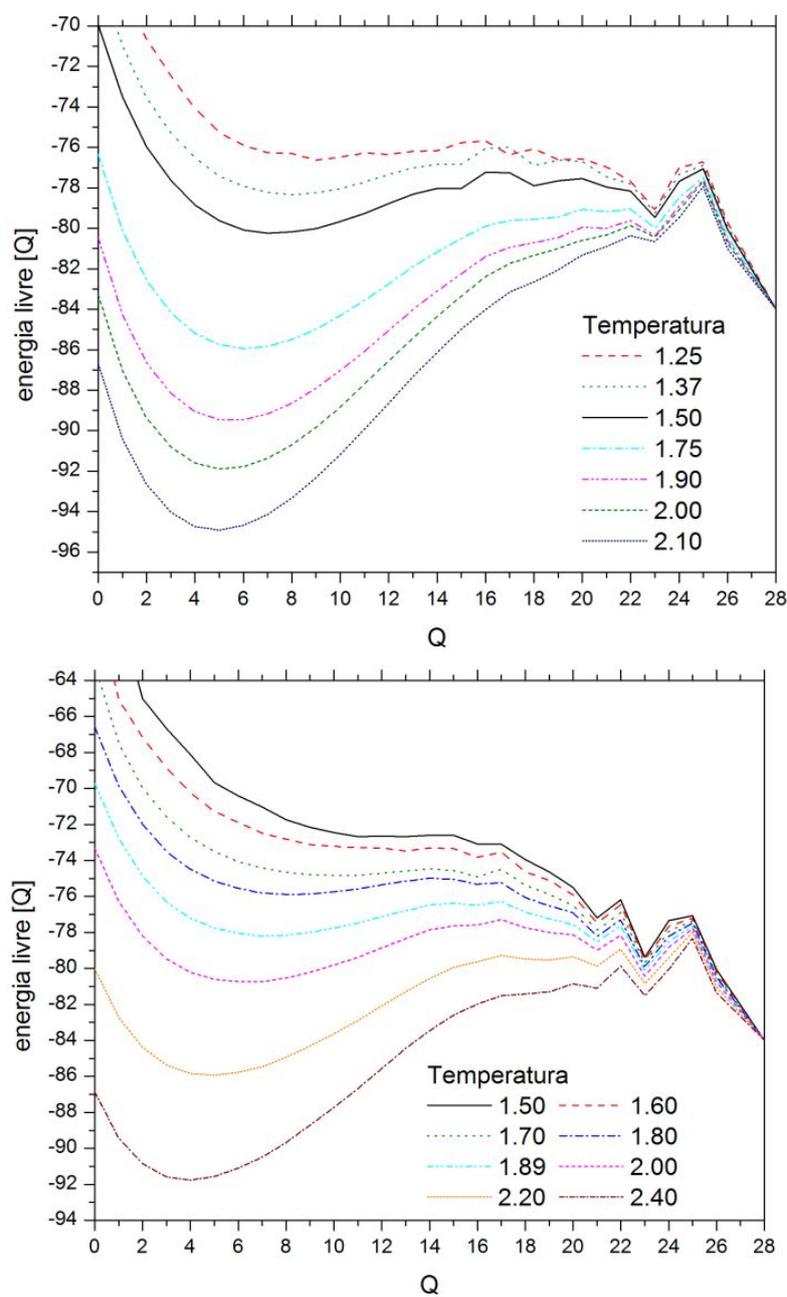


Figura 3.2: Energia livre em função do número de contatos corretos, contatos que existem no estado nativo, para várias temperaturas. No topo, temos o regime de alta hidrofobicidade e, abaixo, regime de baixa. Para os dois regimes, a energia livre tem a forma da figura 1.3 da seção 1.4.

regimes é diferente.

Para o regime de alta hidrofobicidade, há dois vales bem definidos com uma barreira de energia entre eles. O primeiro vale ocorre em $Q \approx 7$ que é onde a cadeia sofre um rápido colapso e forma uma estrutura fechada. Nesse ponto, a cadeia permanece boa parte do tempo de enovelamento a caminho do estado nativo. A forma da energia livre nessa vizinhança é quasi-harmônica e o tempo de enovelamento é dado, em sua grande parte, pelo tempo que a cadeia leva para atravessar a barreira de energia em $Q_t \approx 16$, a região de transição. Depois de atravessar a barreira, a proteína atinge o próximo vale que corresponde ao estado enovelado em $Q = 28$. A forma rugosa da energia livre próximo de $Q = 23$ se dá devido à artificialidade da rede. Nesse ponto, existem movimentos de Monte Carlo que levam a coordenada de reação direto de $Q = 23$ para $Q = 28$.

Para o regime de baixa hidrofobicidade, o gráfico da energia livre é geralmente “ladeira à baixo”, isto é, possui um vale de energia em $Q = 28$ onde corresponde ao estado enovelado. Isso é devido ao fato de que nesse regime de hidrofobicidade o colapso e o enovelamento ocorrem simultaneamente a medida que o heteropolímero se enovela. Para altas temperaturas, começa a aparecer dois vales de energia como no regime de alta hidrofobicidade.

A figura 3.3 mostra a energia livre na temperatura de enovelamento T_f para os dois regimes. Vemos que o comportamento é parecido a não ser pelo fato de que, no regime de baixa, a região em $Q = 7$ possui energia mais alta e a barreira de energia em $Q = 16$ é muito menor. Como mencionado antes, no regime de alta temos dois vales bem definidos, a proteína leva um tempo no primeiro vale para atravessar a barreira de energia e atingir o segundo vale, o estado nativo. No regime de baixa, o enovelamento é ladeira à baixo, não gasta-se muito tempo para atravessar a barreira de energia que é muito mais baixa que no regime de alta. Isso explica o fato de o tempo de enovelamento no regime de baixa ($\tau_f = (5.7 \pm 0.2) \times 10^5$ passos de MC) ser cerca de seis vezes mais rápido que no regime de alta hidrofobicidade ($\tau_f = (3.3 \pm 0.1) \times 10^6$ passos de MC).

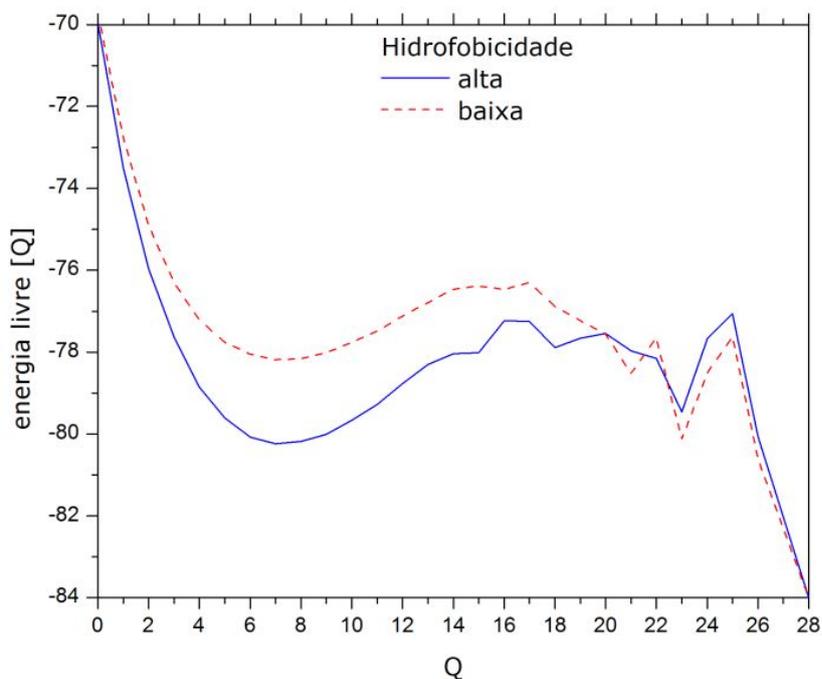


Figura 3.3: Energia livre em função do parâmetro de ordem na temperatura de enovelamento T_f . A curva contínua se refere ao regime de alta hidrofobicidade e a pontilhada, ao regime de baixa. Nos dois regimes, as curvas têm comportamentos diferentes.

A medida que aproximamos da temperatura de vidro, diminuímos a temperatura do sistema, a barreira de energia diminui. Entretanto, o tempo de enovelamento diverge pelo fato de que a proteína encontra-se armadilhada em mínimos locais de energia (rugosidade da superfície de energia). A cadeia não possui energia térmica suficiente para sair desse mínimo local e acaba permanecendo muito tempo armadilhada numa estrutura desenovelada. A entropia, nesse ponto, converge pois são poucas as estruturas em que a cadeia pode se interconverter quando na situação armadilhada.

Quando aumentamos a temperatura do sistema, a barreira de energia

aumenta consideravelmente pois a proteína possui um gradiente de energia e entropia muito grande diminuindo consideravelmente a energia livre, a energia livre do primeiro vale é muito baixa. Nesse ponto do enovelamento, a energia térmica do sistema é muito alta e a entropia diverge. A cadeia possui energia suficiente para acessar todas as estruturas de seu espaço conformacional aumentando consideravelmente o tempo de enovelamento pois a probabilidade de acessar a estrutura nativa é a mesma de todas as outras.

Sendo assim, a explicação para a forma parabólica do gráfico dos tempos de enovelamento (veja Figura 3.1) tem razões diferentes para cada lado da parábola, não há simetria nesse gráfico. O lado esquerdo, baixas temperaturas, se refere à baixa energia térmica para a cadeia sair dos mínimos locais e o tempo aumenta exponencialmente. Já o lado direito, altas temperaturas, se explica pelo fato de que a cadeia possui alta energia e alta entropia, ela acessa todas as estruturas em igual probabilidade demorando para acessar a estrutura nativa. A temperatura de enovelamento T_f é uma temperatura intermediária, pois ocorre o balanceamento dos termos entrópicos e energéticos e, com isso, o tempo de enovelamento é o mínimo possível.

3.3 Coeficiente de difusão constante

3.3.1 Barreira de energia em $Q = 16$

Considerando o enovelamento como uma reação simples, com uma única coordenada de reação Q , podemos utilizar a equação 1.2 para calcular o tempo de enovelamento τ_f . O coeficiente de difusão depende de Q , mas aqui faremos a simplificação de utilizá-lo constante. Calculamos o coeficiente de difusão D_0 através da função correlação das flutuações da coordenada de reação $\Delta Q(t)$. A difusão quasi-harmônica da função correlação decai exponencialmente para tem-

pos longos. Para uma determinada temperatura, o coeficiente de difusão pode ser calculado pelo modelo de difusão harmônica com

$$D_0(T) = \Delta Q^2(T)/\tau_{corr}(T) \quad (3.1)$$

no qual $\Delta Q^2(T)$ é a média quadrática dos valores instantâneos da flutuação da coordenada de reação na temperatura T e $\tau_{corr}(T)$ é o tempo de autocorrelação, tempo de decaimento da função correlação. $\Delta Q^2(T)$ para uma determinada temperatura T é calculado de acordo com

$$\Delta Q^2(T) = \sum_{i=0}^N Q_i^2(T)/N - \left(\sum_{i=0}^N Q_i(T)/N \right)^2 \quad (3.2)$$

sendo Q_i o valor instantâneo da coordenada de reação no tempo i da simulação que ocorre do tempo $i = 0$ até $i = N$. O cálculo de $\tau_{corr}(T)$ é explicado no apêndice A.

Esses cálculos são restritos para valores de Q antes da região de transição, ou seja, há uma barreira de energia infinita em $Q = 16$. Isso porque, para o cálculo da função correlação $\langle Q(t)Q(0) \rangle$, foi considerado qualquer configuração com $Q < 17$ na região de colapso. Essa consideração é feita por Socci e colaboradores [33] para o cálculo do coeficiente de difusão no regime de alta hidrofobicidade.

A figura 3.4 mostra o coeficiente de difusão obtido nos cálculos para várias temperaturas em alta e baixa hidrofobicidade. Podemos notar que o coeficiente de difusão para o regime de baixa hidrofobicidade possui crescimento exponencial mais ascentuado que o regime de alta. D_0 é maior no regime de baixa pois a difusão ocorre de maneira mais fácil nesse regime que no de alta, os contatos corretos vão se formando a medida que ocorre o enovelamento do heteropolímero. A difusão em alta hidrofobicidade é mais baixa, pois a cadeia deve atingir sua forma colapsada e lentamente rearranjar-se, difundir pela barreira de energia livre, para atingir sua forma enovelada.

Com o coeficiente de difusão e energia livre calculados, estimamos o tempo de enovelamento analiticamente dado pela equação 1.2. Aqui, utiliza-

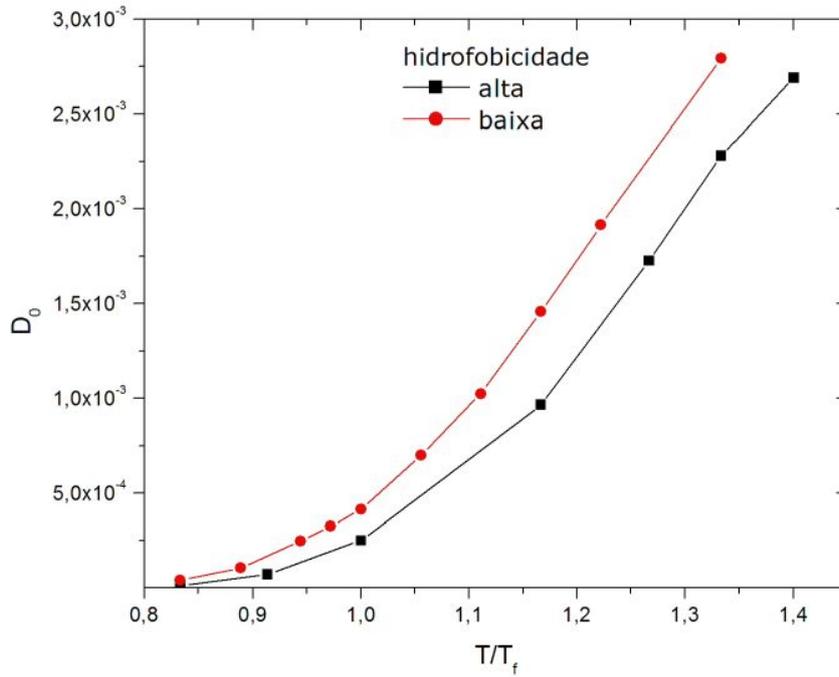


Figura 3.4: Coeficiente de difusão D_0 em função da temperatura para os regimes de alta e baixa hidrofobicidade. O gráfico do coeficiente de difusão é do tipo exponencial crescente. O tempo de decaimento da função autocorrelação τ_{corr} possui comportamento contrário, exponencial decrescente. A temperatura para cada regime está normalizada pela temperatura de enovelamento T_f .

mos a forma discreta para a integral dupla com o coeficiente de difusão constante [33]

$$\tau_f = \frac{1}{D_0} \sum_{Q=0}^{23} \sum_{Q'=0}^{Q-1} e^{\beta\{F(Q)-F(Q')\}}. \quad (3.3)$$

A soma ocorre até $Q = 23$ pois não esperamos que o coeficiente de difusão permaneça constante para $23 \leq Q \leq 28$. Nessa região, a rede possuiu uma artificialidade e a cadeia não passa por alguns valores de Q . Também, não é significativo para a soma total do tempo de enovelamento o tempo que o heteropolímero leva para sair de $Q = 23$ para chegar no estado enovelado $Q = 28$. Isso será visto com maiores detalhes nas seções posteriores.

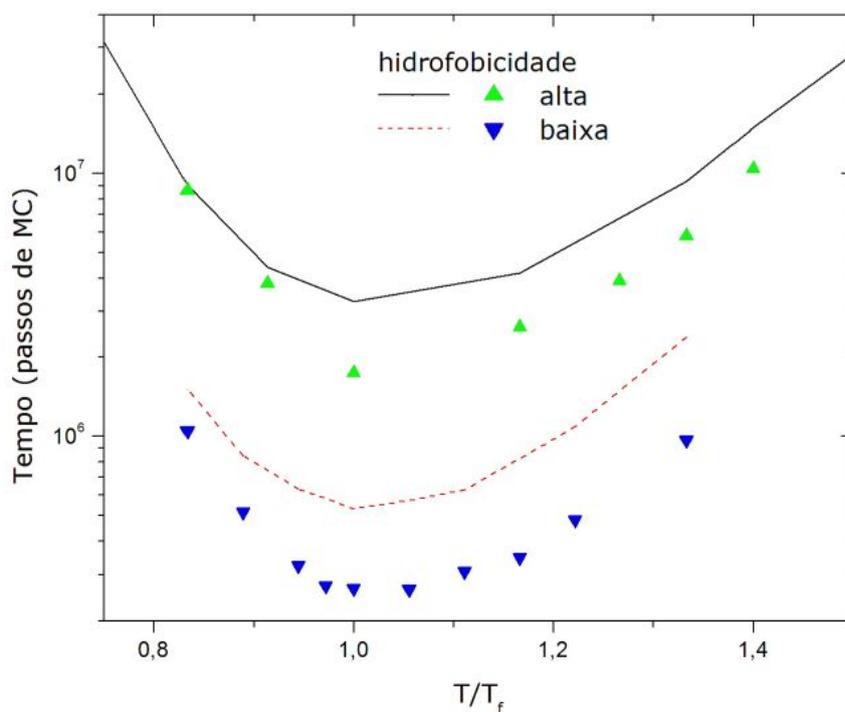


Figura 3.5: Comparação dos tempos de envelhecimento obtidos via simulação de Monte Carlo (linhas) com os tempos calculados com a equação 3.3 (pontos) para os dois regimes de hidrofobicidade. A temperatura para cada regime está normalizada pela temperatura de envelhecimento T_f .

A partir da equação 3.3, obtivemos os tempos de envelhecimento que estão representados pelos pontos na figura 3.5. As linhas são os tempos obtidos via simulação de Monte Carlo (veja seção 3.1). Calculamos os tempos de envelhecimento para alta e baixa hidrofobicidade e para várias temperaturas próximas da temperatura de envelhecimento T_f .

Os tempos calculados teoricamente em alta hidrofobicidade são da ordem de seis vezes maiores que os calculados em baixa hidrofobicidade. Isso pode ser explicado pelo fato de o cálculo do tempo de envelhecimento depender da energia livre e do coeficiente de difusão. O perfil da energia livre é diferente para os dois regimes: em alta hidrofobicidade há uma barreira de energia que aumenta

o tempo de enovelamento pois a cadeia leva um tempo para atravessá-la; em baixa, quase não há barreira e o enovelamento é mais rápido (veja Figura 3.3). Também, o tempo de enovelamento é proporcional ao inverso do coeficiente de difusão. Quanto maior o coeficiente de difusão, menor é o tempo para a proteína se enovelar. O coeficiente de difusão é maior no regime de baixa (veja Figura 3.4) o que também explica o fato de, nesse regime, a cadeia encontrar mais rápido o estado enovelado.

Os tempos de enovelamento teórico para o regime de alta hidrofobicidade concordam com os resultados obtidos por Socci e colaboradores [33]. Os tempos encontrados para o regime de baixa hidrofobicidade são qualitativamente parecidos com os tempos teóricos calculados em alta hidrofobicidade. Há uma discrepância entre os tempos calculados teoricamente com os tempos obtidos via simulação. Isso é explicado devido ao fato de utilizarmos um coeficiente de difusão constante na somatória da equação 3.3. Também, a somatória é feita na região de colapso, ou seja, há uma barreira de energia em $Q = 16$ imposta por Socci [33] com argumentos qualitativos, o que faz com que a somatória dos tempos seja restrita a essa região.

Podemos observar que a análise teórica baseada nos cálculos na região de colapso e do funil de energia livre não está quantitativamente correta, porém pode ser usado para análises qualitativas da predição do tempo de enovelamento para uma grande faixa de temperatura.

Esperamos obter melhores resultados nas próximas seções analisando a importância dessa barreira e, logo adiante, calcularemos os coeficientes de difusão em função da coordenada de reação para o cálculo dos tempos de enovelamento.

3.3.2 Barreira de energia em $Q = 23$

O critério utilizado por Socci e colaboradores [33] em colocar a barreira

de energia em $Q = 16$ não é bem explicado ou investigado quantitativamente e essa pode ser uma das razões da discrepância encontrada entre os tempos de enovelamento teórico e obtidos via simulação.

Realizamos os mesmos cálculos para o tempo de enovelamento teórico, como na seção anterior, mudando apenas a barreira de energia, que agora está em $Q = 23$. Fizemos simulações para os dois regimes de hidrofobicidade em suas respectivas temperatura de enovelamento T_f . Para essas, fixamos o parâmetro de ordem, número de contatos corretos Q , para a estrutura inicial e calculamos os tempos de enovelamento a partir desse Q inicial. Medimos o tempo de enovelamento médio entre 1000 simulações para cada Q , no intervalo $4 \leq Q \leq 24$. O resultado é apresentado na figura 3.6.

O critério utilizado para usarmos uma barreira de energia em $Q = 23$ é bem explicado pela figura 3.6. Nela podemos perceber que para valores de Q até 23 o tempo de enovelamento é praticamente o mesmo comparando com o tempo de enovelamento τ_f para cada regime, no qual $\tau_f = 5.7 \times 10^5$ passos de MC em baixa hidrofobicidade e $\tau_f = 3.3 \times 10^6$ passos de MC em alta hidrofobicidade. A partir de $Q = 23$, os tempos são pelo menos duas ordens de grandeza mais rápidos que τ_f para os dois regimes de hidrofobicidade e não influenciam muito no tempo total de enovelamento.

Depois de calcularmos os tempos de enovelamento com barreira de energia em $Q = 23$, comparamos com os tempos obtidos com barreira energética em $Q = 16$ (veja seção 3.3.1) e por simulação de MC os quais estão representados na figura 3.7. Os resultados para os tempos de enovelamento teórico com a barreira de energia em $Q = 23$ oferecem resultados mais precisos que os obtidos por Socci com barreira em $Q = 16$, apesar de ainda utilizarmos o coeficiente de difusão constante. Isso porque deve-se levar em conta a soma dos tempos até pelo menos $Q = 23$, lembrando que depois disso se espera que o coeficiente de difusão permaneça constante porque, após esse ponto, há o segundo vale de energia livre onde o enovelamento ocorre constante e direto para o estado enovelado (veja seção

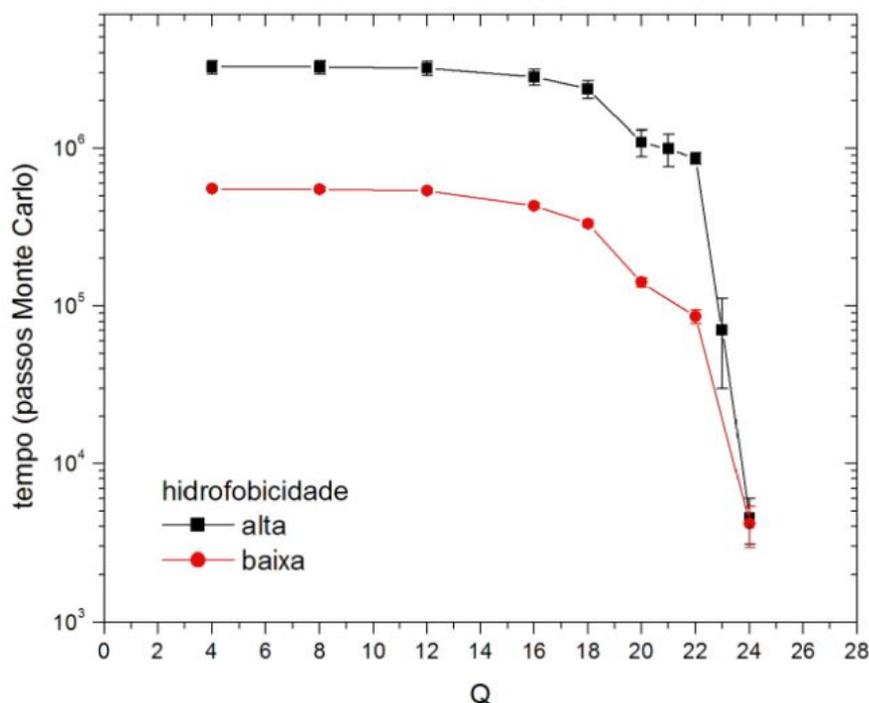


Figura 3.6: Tempos de enovelamento para alta e baixa hidrofobicidade na temperatura de enovelamento T_f . Fixamos o número de contatos nativos Q para a estrutura inicial e medimos o tempo de enovelamento médio entre 1000 simulações para cada Q .

3.2). Por isso, somamos ao tempo de enovelamento a contribuição dos tempos entre $16 \leq Q \leq 23$.

Apesar disso, o cálculo com barreira de energia em $Q = 16$ se torna importante pois foi a primeira aproximação para o cálculo analítico dos tempos de enovelamento. Esse não está totalmente errado, pois se considerou a barreira de energia na coordenada de reação em que ocorre a transição de fase (Q_t) da proteína do estado desenovelado para o estado enovelado (veja Figura 1.3).

Dessa forma, com os gráficos da figura 3.6 podemos notar que não precisamos calcular o tempo de enovelamento para diferentes Q 's, basta observarmos qual a região na coordenada de reação que mais contribui para o tempo de eno-

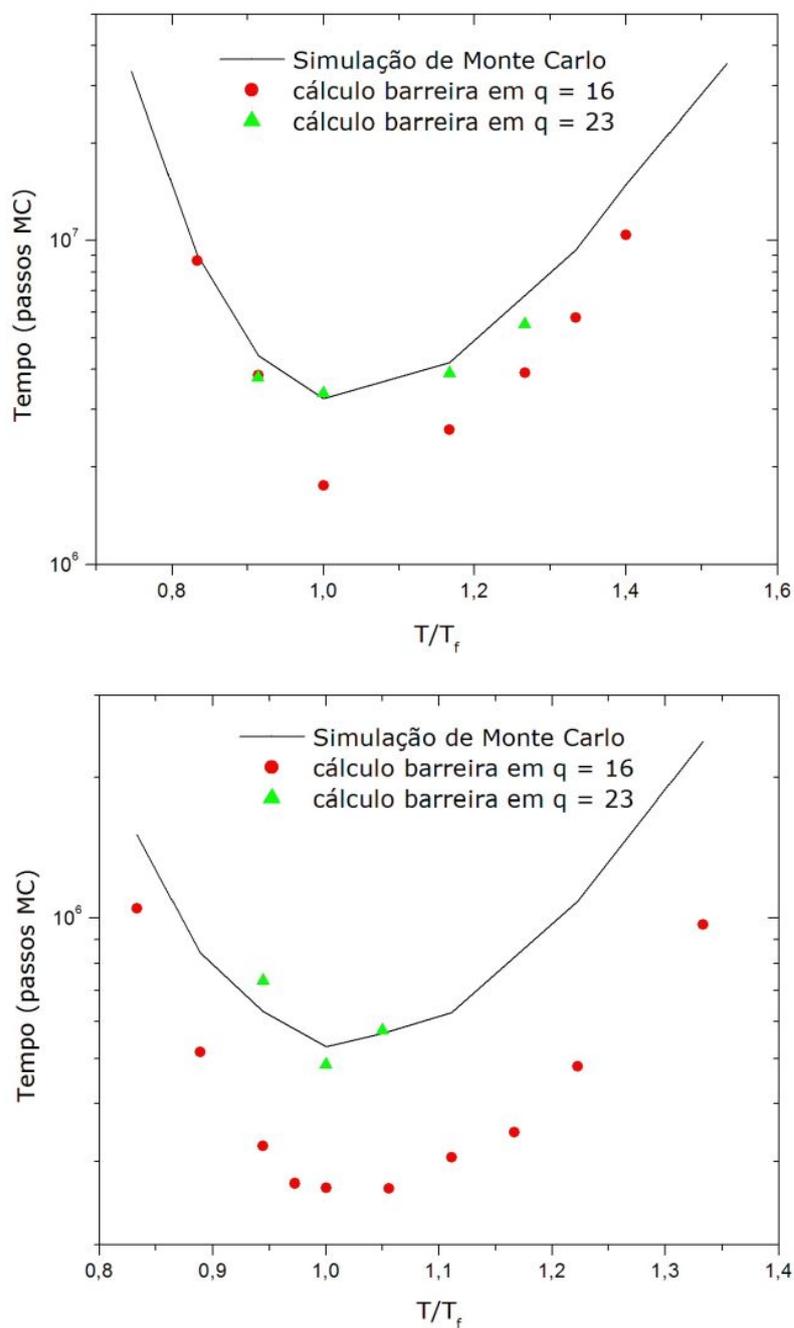


Figura 3.7: As linhas representam os tempos de enovelamento obtidos via simulação de MC. Os pontos são tempos calculados pela equação 3.3 para o regime de alta hidrofobicidade, no topo, e para baixa hidrofobicidade, abaixo. A temperatura para cada regime está normalizada por T_f .

velamento e, então, a barreira de energia pode ser posicionada nesse ponto.

3.4 Coeficiente de difusão variado

Nosso intuito é obter os tempos de enovelamento analiticamente através da equação 1.2. Nessa equação, o coeficiente de difusão D depende da coordenada de reação Q , ou seja, o coeficiente de difusão varia ao longo da coordenada de reação ($D(Q)$). Nas seções anteriores 3.3.1 e 3.3.2, utilizamos um coeficiente de difusão constante para uma primeira aproximação. Na literatura [33] é calculado os tempos com coeficiente constante para o regime de alta hidrofobicidade obtendo uma predição qualitativa dos tempos de enovelamento.

Para calcular os coeficientes de difusão ao longo da coordenada de reação, fizemos novas simulações longas adicionando um novo termo ao potencial 2.4 utilizado até aqui em nosso modelo. Esse termo deve garantir que a simulação oscile próximo de um certo valor da coordenada de reação, isto é, garantir que a cadeia visite certos valores de Q próximos de um Q médio definido [52]. Dessa forma, podemos calcular um coeficiente de difusão em função desse Q médio. O novo termo é do tipo potencial de mola e o potencial utilizado é então definido por

$$E = N_l E_l + N_u E_u + K [Q - \langle Q \rangle]^2, \quad (3.4)$$

em que K é a constante de mola, Q é a coordenada de reação da simulação e $\langle Q \rangle$ é o valor médio da coordenada de reação em torno do qual a simulação ficará oscilando. Assim, calculamos o coeficiente de difusão em função da coordenada de reação média que fizemos a simulação.

Já não há mais a necessidade de se utilizar uma barreira de energia para o cálculo do coeficiente de difusão. Com o termo de mola, impomos que a simulação deve ocorrer próximo do valor médio da coordenada de reação.

Os possíveis valores para o parâmetro K foi calculado em [52]. Em

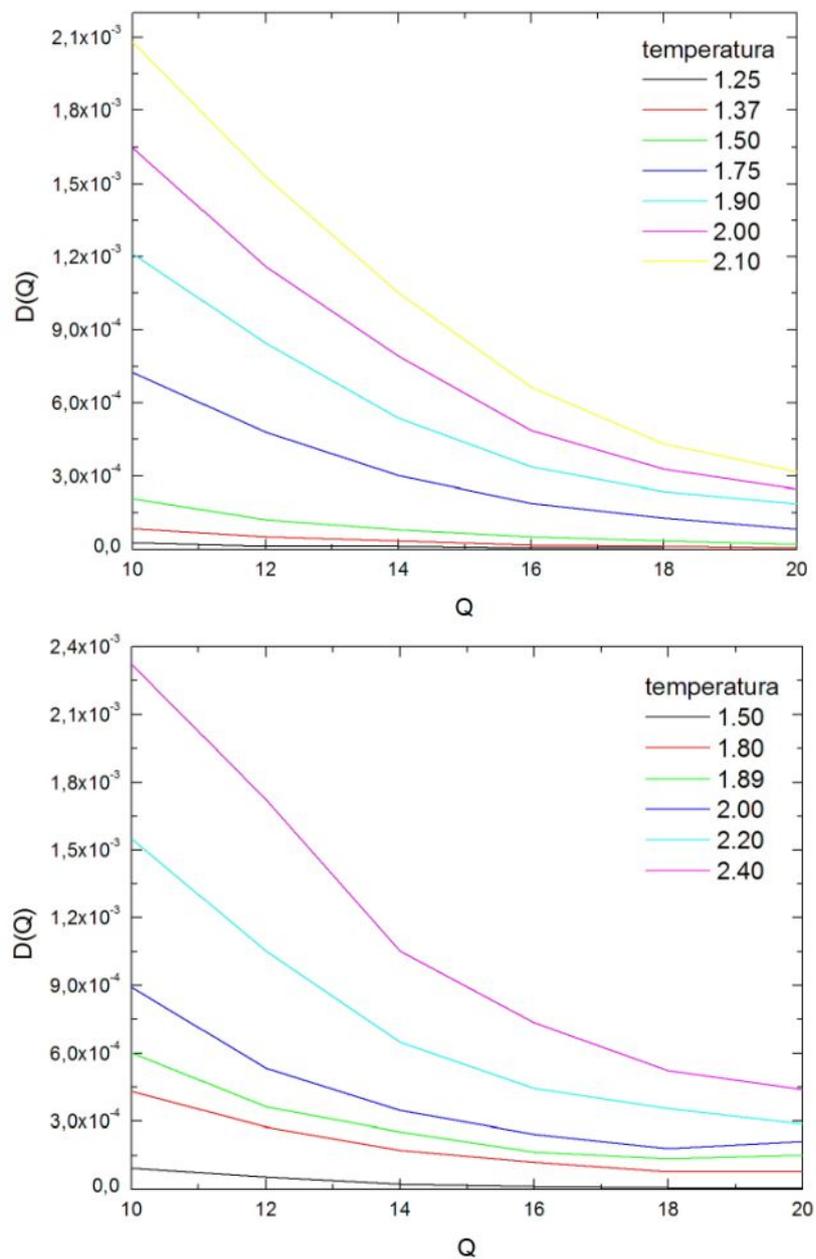


Figura 3.8: Representação dos coeficientes de difusão em função da coordenada de reação Q para várias temperaturas. No topo, temos o gráfico para o regime de alta hidrofobicidade e, abaixo, para o regime de baixa.

nossas simulações, utilizamos $K = 0.08$, pois não queremos que o termo de mola se sobreponha muito aos outros dois termos nem distorça significativamente o histograma da coordenada de reação já calculado [52]. Esse valor para K torna o termo de mola do potencial pequeno quando comparado com a energia de interação entre os monômeros ($1 \leq |E_{l,u}| \leq 3$). A cadeia não deve oscilar muito longe de $\langle Q \rangle$, não queremos alterar as características de nosso modelo.

Para cada $\langle Q \rangle$ é realizada uma simulação longa e calculado o coeficiente de difusão $D(Q)$ correspondente. Para cada temperatura, são feitas simulações entre $10 \leq \langle Q \rangle \leq 20$. Encontrar o coeficiente de difusão fora desse intervalo se torna difícil pelo fato de que, antes de $\langle Q \rangle = 10$, $\langle Q \rangle$ está muito afastado do pico da gaussiana de distribuição dos $\langle Q \rangle$'s correspondentes a uma simulação utilizando o potencial 2.4 convencional do nosso modelo, causando uma distorção no histograma de energia [52]. Já para $\langle Q \rangle > 20$, existe um movimento em $Q = 23$ que leva o heteropolímero para $Q = 28$, seu estado nativo, e então o coeficiente de difusão calculado nessa região não seria real devido a esta artificialidade da rede.

Na figura 3.8 estão representados os coeficientes de difusão em função da coordenada de reação calculados para várias temperaturas nos dois regimes de hidrofobicidade estudados. O cálculo do coeficiente de difusão é explicado na seção 3.3.1 e no apêndice A.

Podemos observar que o coeficiente de difusão da figura 3.8 varia muito ao longo da coordenada de reação. Utilizá-lo constante é uma grande aproximação e resulta em apenas análises qualitativas para a predição do tempo de enovelamento. Seu gráfico, em geral, tem a forma de uma exponencial decrescente. Inicialmente, a proteína atinge o primeiro vale de energia muito rapidamente e por isso o coeficiente de difusão é alto. Logo após, a proteína tem que atravessar uma barreira de potencial e o coeficiente de difusão diminui. Próximo do estado enovelado, o coeficiente de difusão permanece praticamente constante pelo fato de que a proteína atinge lentamente esse estado.

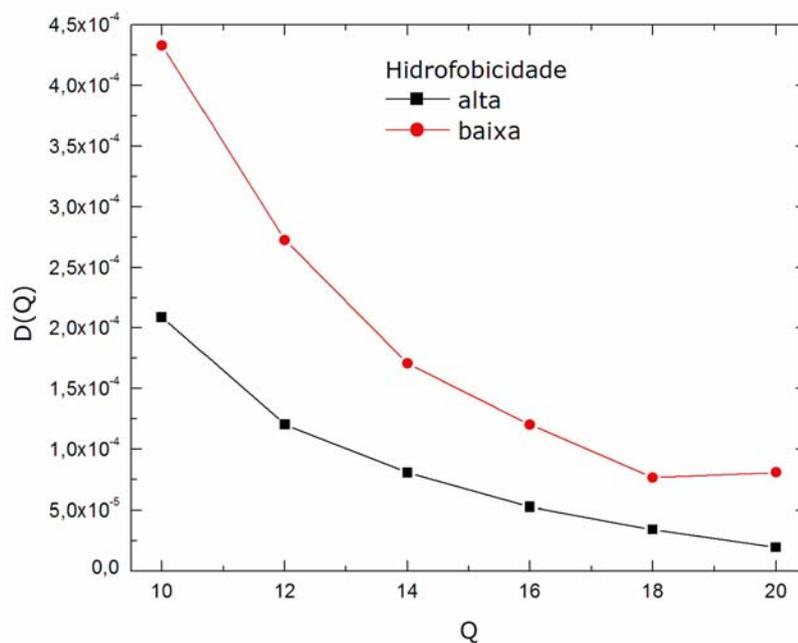


Figura 3.9: Representação dos coeficientes de difusão em função da coordenada de reação Q para os regimes de alta e baixa hidrofobicidade na sua respectiva temperatura de enovelamento T_f .

Na figura 3.9 representamos os coeficientes de difusão somente na temperatura de enovelamento para os dois regimes de hidrofobicidade, alta e baixa. Novamente, observamos que o coeficiente de difusão no regime de baixa hidrofobicidade possui valores maiores que o regime de alta confirmando o que já observamos nas seções anteriores. Em baixa hidrofobicidade, o enovelamento acontece junto com o colapso a medida que a cadeia se enovela. Isso faz com que a cadeia alcance rapidamente o estado enovelado. Já em alta hidrofobicidade, a cadeia sofre um colapso de sua estrutura e lentamente rearranja-se para atingir seu estado nativo.

Com a energia livre e os coeficientes de difusão em função de Q em mãos,

podemos calcular os tempos de enovelamento analiticamente. Utilizamos a forma discreta da equação 1.2 substituindo as integrais por somatórias que será

$$\tau_f = \sum_{Q=Q_{unf}}^{Q_{fold}} \sum_{Q'=0}^{Q-1} \frac{e^{\beta\{F(Q)-F(Q')\}}}{D(Q)}. \quad (3.5)$$

O coeficiente de difusão está em função da coordenada de reação, $D(Q)$, e é incluído dentro da somatória. Como na equação 1.2, a soma é entre Q_{unf} e Q_{fold} e que no nosso modelo corresponderá a $Q_{unf} = 7$ e $Q_{fold} = 16$, respectivamente. Utilizamos $Q_{fold} = 16$ na somatória, pois calcularemos até a região de transição.

Podemos considerar, nesse modelo, que o tempo de enovelamento é dado quase que exclusivamente pelo tempo levado pela proteína para atravessar a barreira de energia livre. Depois de $Q = 16$, o coeficiente de difusão permanece quase constante (veja Figuras 3.9 e 3.8) e a barreira de energia livre diminui consideravelmente (veja Figura 3.3). Isso traz para os cálculos da somatória dos tempos termos que podem ser descartados.

No gráfico da figura 3.10, temos a trajetória do enovelamento ($\tau(Q)$) superposto à energia livre ($F[Q]$), ambos em função da coordenada de reação [33]. Observamos que o heteropolímero gasta a maior parte do tempo (cerca de 85% de τ_f) difundindo no primeiro vale da energia livre, em $Q \approx 7$, região do colapso. Quando a barreira é ultrapassada¹, o enovelamento ocorre rapidamente (cerca de $0.03\tau_f$).

Então, a partir da equação 3.5, obtivemos os tempos de enovelamento que estão representados pelos triângulos na figura 3.11. As linhas são os tempos obtidos via simulação de Monte Carlo (veja seção 3.1). Calculamos os tempos de enovelamento para alta e baixa hidrofobicidade e para várias temperaturas próximas da temperatura de enovelamento T_f .

Utilizando o coeficiente de difusão em função da coordenada de reação, $D(Q)$, obtivemos resultados para o tempo de enovelamento melhores que os cal-

¹A cadeia atravessa e volta a barreira de energia livre inúmeras vezes confirmando que a teoria da difusão é bem aplicada para essa coordenada de reação.

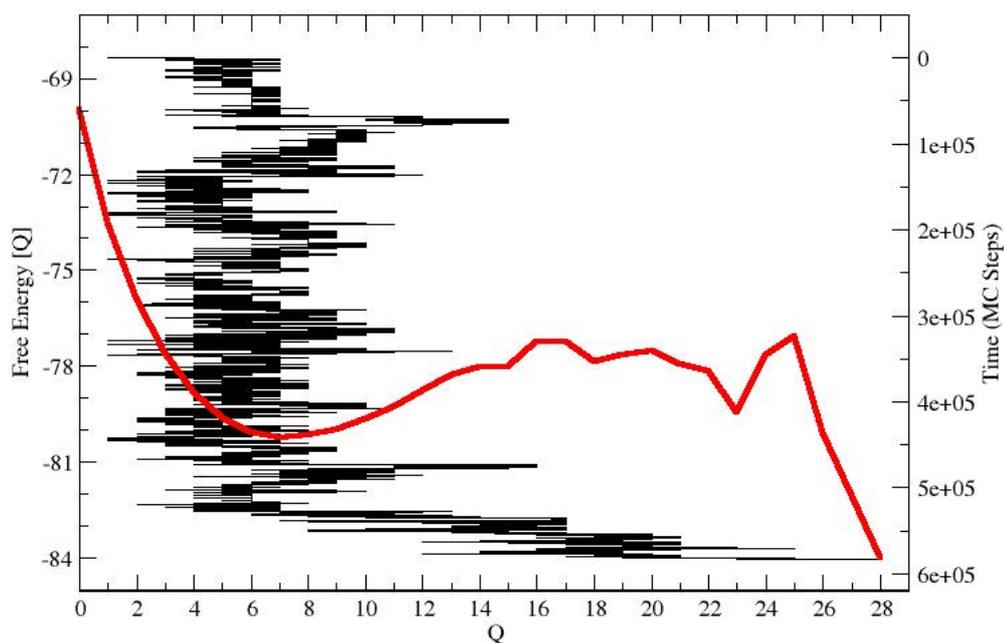


Figura 3.10: Estão representados a trajetória de envelamento em passos de MC ($\tau(Q)$) no eixo direito e, no esquerdo, a energia livre ($F[Q]$). Ambos em função da coordenada de reação Q [33]. Notemos o aspecto difusivo através da coordenada de reação. Cadeia permanece boa parte da trajetória do envelamento no primeiro vale de energia livre tentando difundir pela barreira. Após a difusão pela barreira, a cadeia segue direto para o segundo vale de energia livre e atinge sua forma envelada.

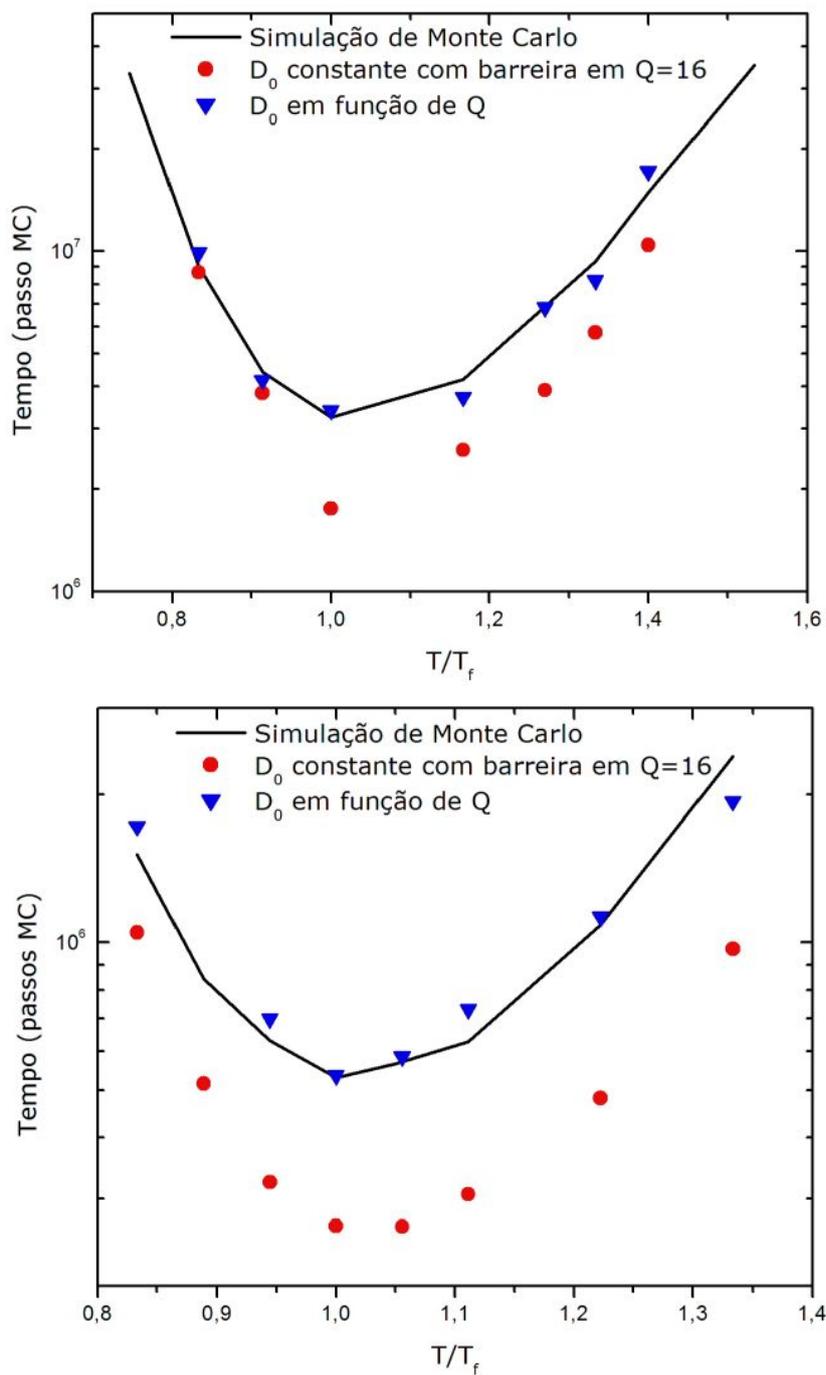


Figura 3.11: Estão representados os coeficientes de difusão em função da coordenada de reação Q para várias temperaturas. No topo, temos o gráfico para o regime de alta hidrofobicidade e, abaixo, para o regime de baixa.

culados com o coeficiente de difusão constante, tanto com a barreira de energia em $Q = 16$ quanto com a barreira de energia em $Q = 23$. Isso já era esperado, pois como mencionado nas seções anteriores, utilizar o coeficiente de difusão constante resulta em previsões qualitativas dos tempos de enovelamento, uma grande discrepância com os tempos obtidos via simulação de Monte Carlo. As simulações são mais simples de serem implementadas nessa modalidade.

O cálculo dos tempos de enovelamento, de acordo com a teoria desenvolvida por BW [29], deve ser feito com o coeficiente de difusão em função da coordenada de reação. O tempo de enovelamento com o coeficiente de difusão dependente da coordenada de reação está mais próximo dos resultados de Monte Carlo do que os tempos calculados com coeficiente de difusão constante. Com o coeficiente de difusão dependente, temos uma melhor descrição da dinâmica de enovelamento de proteína.

Capítulo 4

Conclusão

O enovelamento de proteína é um exemplo de reação química que é mais complexa que as reações de pequenas moléculas. Uma simples generalização da teoria de estado de transição é desenvolvida e usada para estimar o tempo de enovelamento para este modelo. Um Hamiltoniano estocástico é utilizado para estudar uma complexa reação química, uma aplicação em um modelo simples de proteína.

Para estudar propriedades moleculares precisamos calcular entropia, energia livre e energia total do sistema, que são derivadas da função de partição da mecânica estatística. Calcular essas quantidades de uma proteína via simulação com todos os átomos é quase impossível, visto que é enorme o grau de liberdade desse sistema. Neste trabalho, vimos que utilizar um modelo simples de proteína resolve este nosso impasse. Através da simulação computacional, conseguimos calcular com exatidão a função de partição do nosso modelo com as técnicas do histograma de Monte Carlo com o qual obtivemos todas as grandezas termodinâmicas que necessitamos. Apesar da simplicidade do modelo que é muito distante de uma proteína real, podemos correlacionar seus resultados com os obtidos experimentalmente e termos uma idéia sobre o enovelamento de proteína.

A cinética do enovelamento pode ser descrita em termos da superfície de

energia. A teoria de superfície de energia se mostrou muito útil no desenvolvimento das idéias sobre o enovelamento de proteína. Com ela, podemos entender porque uma proteína possui muitos caminhos de enovelamento ao mesmo tempo que consegue atingir seu estado enovelado em tempo útil de exercer sua atividade metabólica (Anfinsen *versus* Levinthal). No caminho do enovelamento através da superfície de energia, a proteína passa por um funil de estruturas partindo de uma estrutura inicial desenovelada até alcançar sua estrutura enovelada. A pergunta que queremos responder é: “Quanto tempo leva esse processo?”.

Utilizando um modelo simples exato aliado à simulação computacional, calculamos as médias dos tempos de enovelamento de um heteropolímero com seqüência e estrutura exaustivamente estudada na literatura. Com esses tempos obtidos via simulação, calculamos os tempos de enovelamento através da solução da equação de difusão do tipo Fokker-Plank. Essa equação, possui solução para o tempo de enovelamento o qual é dependente do coeficiente de difusão e da energia livre do sistema. Também, o enovelamento é fortemente dependente da temperatura do sistema.

O coeficiente de difusão é função de uma coordenada de reação, no nosso caso uma coordenada de reação coletiva do sistema. A cinética do enovelamento de um modelo de rede simples é bem descrita por uma única coordenada de reação coletiva sobre uma grande faixa de temperatura. Essa coordenada é suficiente para descrever a reação do enovelamento através da superfície de energia e do funil. Calculamos esses coeficientes de difusão dependente através de simulações longas de Monte Carlo em tempo computacional não muito grande.

A literatura já havia calculado esses tempos, porém utilizavam o coeficiente de difusão constante resultando em uma discrepância com os tempos obtidos via simulação. Nesse trabalho, calculamos os tempos com coeficiente de difusão em função do enovelamento, mais precisamente da coordenada de reação, sobre uma pequena faixa de temperatura. Constatamos que essa abordagem resulta em tempos mais precisos que os calculados com o coeficiente de difusão

constante.

Utilizar os coeficientes de difusão em função da coordenada de reação nos traz uma descrição mais realística da dinâmica de enovelamento de proteína, ou seja, os tempos teóricos são mais próximos dos experimentais.

Referências Bibliográficas

- [1] M. Daune. *Molecular Biophysics*. Oxford, New York, english edition, 1999.
- [2] D. Voet and J. Voet. *Biochemistry*. John Wiley & Sons, New York, 1990.
- [3] L. Stryer. *Biochemistry*. Freeman, New York, 1988.
- [4] A. L. Lehninger. *Princípios de Bioquímica*. Sarvier Editora de Livros Médicos, 1985.
- [5] L. Pauling; R. B. Corey and H. R. Branson. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 37:205–211, 1951.
- [6] J. D. Watson and H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [7] T. E. Creighton. *Protein Folding*. Freeman, New York, 1st edition, 1992.
- [8] C. B. Anfinsen; E. Harber; M. Sela and F. White Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 47:1309–1314, 1961.
- [9] C. B. Anfinsen. Principles that govern the folding of proteins chains. *Science*, 181:223–230, 1973.
- [10] C. Levinthal. Are there pathways for protein folding? *J. Chem. Phys.*, 65:44–45, 1968.

- [11] P. E. Leopold; M. Montal and J. N. Onuchic. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*, 89:8721–8725, 1992.
- [12] P. E. Leopold and E. I. Shakhnovich. Protein folding kinetics in the dense phase. *System Sciences, Proceeding of the Twenty-Sixth Hawaii International Conference*, 89:726–735, 1993.
- [13] J. D. Bryngelson; J. N. Onuchic; N. D. Socci and P. Wolynes. Funnels, pathways and the energy landscape of protein folding: A synthesis. *Proteins: Struct. Funct. Genet.*, 21:167–195, 1995.
- [14] J. N. Onuchic; Z. Luthey-Schulten and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997.
- [15] R. L. Baldwin. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR*, 5:103–109, 1995.
- [16] K. A. Dill and H. S. Chan. From levinthal to pathways to funnels. *Nat. Struct. Biol.*, 4:10–19, 1997.
- [17] S. S. Plotkin and J. N. Onuchic. Structural and energetic heterogeneity in protein folding. I. Theory. *J. Chem. Phys.*, 116(12):5263–5283, 2002.
- [18] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, New York, 2nd edition, 1985.
- [19] J. N. Onuchi; H. Nymeyer; A. E. García; J. Chahine and N. D. Socci. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. *Adv. Protein Chem.*, 53:87–152, 2000.
- [20] K. A. Dill. Polymer principles and protein folding. *Prot. Sci.*, 8:1166–1180, 1999.

- [21] A. R. Fersht. Mapping the structures of transition states and intermediates in folding: delineation of pathways at high resolution. In C. M. Dobson and A. R. Fersht, editor, *Protein Folding*, New York, 1995. Cambridge University Press.
- [22] U. H. E. Hansmann; Y. Okamoto and J. N. Onuchic. The folding funnel landscape for the peptide met-enkephalin. *Proteins: Struct. Funct. Genet.*, 34:472–483, 1999.
- [23] A. R. Dinner; A. Šali; L. J. Smith; C. M. Dobson and M. Karplus. Understanding protein folding via free-energy surfaces from theory and experiment. *TIBS*, 25:331–339, 2000.
- [24] H. Nymeyer; N. D. Socci and J. N. Onuchic. Landscape approaches for determining the ensemble of folding transition states: success and failure hinge on the degree of frustration. *Proc. Natl. Acad. Sci. USA*, 97:634–639, 2000.
- [25] J. D. Bryngelson and P. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*, 84:7524–7528, 1987.
- [26] B. Derrida. Random-energy model: Limit of a family of disordered models. *Phys. Rev. Lett.*, 45(2):79–82, 1980.
- [27] B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24(5):2613–2626, 1981.
- [28] K. A. Dill; S. Bromberg; K. Yue; K. M. Fiebig; D. P. Yee; P. D. Thomas and H. S. Chan. Principal of protein folding — A perspective from simple exact models. *Prot. Sci.*, 4:561–602, 1995.
- [29] J. D. Bryngelson and P. Wolynes. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, 93:6902–6915, 1989.

- [30] T. Tomé and M. J. de Oliveira. *Dinâmica Estocástica e Irreversibilidade*. Editora da Universidade de São Paulo, São Paulo, 2001.
- [31] H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.
- [32] N. Metropolis; A. W. Rosenbluth; M. N. Rosenbluth; A. N. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
- [33] N. D. Socci; J. N. Onuchic and P. G. Wolynes. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.*, 104:5860–5868, 1996.
- [34] V. B. P. Leite and J. N. Onuchic. Structure and dynamics of solvent landscapes in charge-transfer reactions. *J. Phys. Chem.*, 100:7680–7690, 1996.
- [35] V. B. P. Leite. Smooth landscape solvent dynamics in electron transfer reactions. *J. Chem. Phys.*, 110(20):10067–10075, 1999.
- [36] H. Taketomi; Y. Ueda and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulation. *Int. J. Peptide Res.*, 7:445–459, 1975.
- [37] H. S. Chan and K. A. Dill. Compact polymers. *Macromolecules*, 22:4559–4573, 1989.
- [38] H. S. Chan and K. A. Dill. The effects of internal constraints on the configurations of chain molecules. *J. Chem. Phys.*, 92:3118–3135, 1990.
- [39] N. D. Socci and J. N. Onuchic. Folding kinetics of proteinlike heteropolymers. *J. Chem. Phys.*, 101:1519–1528, 1994.
- [40] N. D. Socci; H. Nymeyer and J. N. Onuchic. Exploring the protein folding funnel landscape. *Physica D*, 107:366–382, 1997.

- [41] R. Mélin; H. Li; N. S. Wingreen and C. Tang. Designability, thermodynamics stability, and dynamics in protein folding: A lattice model study. *J. Chem. Phys.*, 110(2):1252–1262, 1999.
- [42] V. B. P. Leite; J. N. Onuchic; G. Stell and J. Wang. Probing the kinetics of single molecule protein folding. *Biophys. J.*, 87:3633–3641, 2004.
- [43] L. C. Oliveira; R. T. H. Silva; V. B. P. Leite and J. Chahine. Frustration and hydrophobicity interplay in protein folding and protein evolution. *J. Chem. Phys.*, 125:1–7, 2006.
- [44] E. I. Shakhnovich and A. Gutin. Enumeration of all compact conformations of copolymers with random sequence of links. *J. Chem. Phys.*, 93(8):5967–5971, 1994.
- [45] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990.
- [46] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford, New York, reprinted edition, 2001.
- [47] N. D. Socci; J. N. Onuchic and P. G. Wolynes. Protein folding mechanisms and the multidimensional folding funnel. *Proteins: Struct. Funct. Genet.*, 32:136–158, 1998.
- [48] J. Chahine; H. Nymeyer; V. B. P. Leite; N. D. Socci and J. N. Onuchic. Specific and nonspecific collapse in protein folding funnels. *Phys. Rev. Lett.*, 88(16):168101–168103, 2002.
- [49] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61:2635–2638, 1988.
- [50] A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.

- [51] N. D. Socci and J. N. Onuchic. Kinetic and thermodynamic analysis of proteinlike heteropolymers. *J. Chem. Phys.*, 103:4732–4744, 1995.
- [52] J. Chahine; V. B. P. Leite; R. J. de Oliveira and J. Wang. Configurational dependent diffusion and protein folding kinetics. *Proc. Natl. Acad. Sci. USA*, 2006. submitted.

Apêndice A

Função correlação

A função correlação será utilizada para obtermos o tempo de autocorrelação da coordenada de reação Q (número de contatos nativos) para calcular o coeficiente de difusão. Para isso, fazemos simulações longas a fim de vizitarmos todo o espaço conformacional. No nosso caso específico, uma barreira de energia foi imposta para restringirmos o espaço conformacional à região de colapso [33].

A dinâmica é feita através do algoritmo de Monte Carlo com movimentos locais até um tempo máximo (τ_{max}). Nossas simulações foram feitas até $\tau_{max} = 1.0 \times 10^9$ passos de MC. Isso porque $\tau_{max} \gg \tau_f$, cerca de 1000 vezes maior. Após a termalização do sistema, para cada passo de MC t anotamos a coordenada de reação $Q(t)$ corresponde. Para o cálculo da função correlação, utilizamos

$$C_Q(\Delta) = \frac{\langle Q(t)Q(t + \Delta) \rangle - \langle Q(t) \rangle^2}{\langle Q(t)^2 \rangle - \langle Q(t) \rangle^2}. \quad (\text{A.1})$$

A figura A.1 mostra os gráficos em escala log-log obtidos para várias temperaturas nos dois regimes estudados: alta e baixa hidrofobicidade. Nesse caso, a barreira de energia imposta está em $Q = 16$, ou seja, foi considerado qualquer configuração com $Q < 17$ na região de colapso. Podemos observar que o decaimento é multi-exponencial com um rápido decaimento seguido de um lento

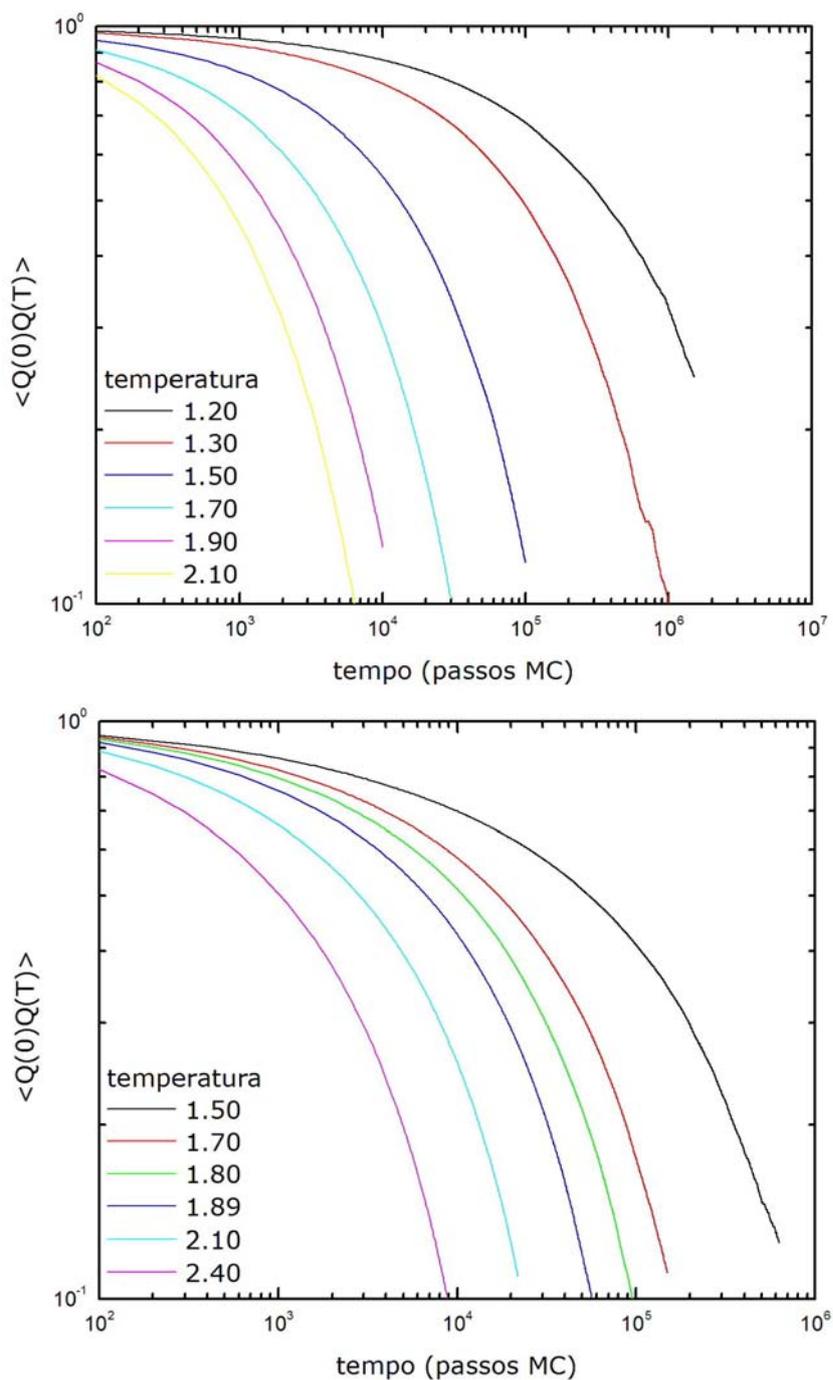


Figura A.1: Função correlação para várias temperaturas plotado em escala log-log. No topo, temos o regime de alta hidrofobicidade e, abaixo, regime de baixa. A medida que a temperatura diminui, o tempo de correlação aumenta.

e longo decaimento. O decaimento final, a região abaixo de $C_Q(\Delta) = 10^{-1}$, é devido ao erro inerente ao cálculo da função correlação em tempos muito longos. Com a diminuição da temperatura, os tempos de correlação aumentam.

Para uma determinada temperatura, uma função com 3 exponenciais foi ajustada ao gráfico da figura A.1. Utilizamos a expressão

$$y(x) = A_1 e^{-x/a_1} + A_2 e^{-x/a_2} + A_3 e^{-x/a_3} \quad (\text{A.2})$$

para o ajuste. O primeiro termo exponencial domina o ajuste em 70%, o segundo em cerca de 20% e o terceiro em 10% os quais se referem aos termos A_1 , A_2 e A_3 , respectivamente. Os expoentes são utilizados para o cálculo do tempo de autocorrelação de acordo com

$$\tau_{corr} = A_1 a_1 + A_2 a_2 + A_3 a_3. \quad (\text{A.3})$$

Estes cálculos foram feitos para várias temperaturas nos casos de alta e baixa hidrofobicidade como mostra a figura A.1 no qual foi utilizado para o cálculo do coeficiente de difusão.