



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Paulo Ricardo Mouro

**Estudo do enovelamento da enzima SOD1 por simulações de
dinâmica molecular**

São José do Rio Preto
2019

Paulo Ricardo Mouro

Estudo do enovelamento da enzima SOD1 por simulações de
dinâmica molecular

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador: Prof. Dr. Vitor Barbanti. Pereira Leite
Coorientador: Prof. Dr. Jorge Chahine

São José do Rio Preto
2019

M931e

Mouro, Paulo Ricardo

Estudo do enovelamento da enzima SOD1 por simulações de dinâmica molecular / Paulo Ricardo Mouro. -- São José do Rio Preto, 2019
76 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientador: Vitor Barbanti Pereira Leite

Coorientador: Jorge Chahine

1. Biologia molecular. 2. Biofísica. 3. Dobramento de proteínas. 4. Dinâmica Molecular. 5. Esclerose lateral amiotrófica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Paulo Ricardo Mouro

Estudo do enovelamento da enzima SOD1 por simulações de
dinâmica molecular

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof. Dr. Jorge Chahine
UNESP – Câmpus de São José do Rio Preto
Coorientador

Prof^a. Dr^a. Ana Lígia Scott
UFABC – Santo André

Prof. Dr. Alexandre Suman de Araujo
UNESP – Câmpus de São José do Rio Preto

Dr. Fabio Rogério de Moraes
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Richard Charles Garrat
USP – São Carlos

São José do Rio Preto
07 de março de 2019

AGRADECIMENTOS

Agradeço aos professores Jorge Chahine e Vitor Leite pela orientação e ensinamentos sem os quais este trabalho não seria realizado.

Aos membros da comissão examinadora por contribuírem para a melhoria desta tese.

A UNESP/IBILCE e ao programa de pós-graduação em Biofísica Molecular, que por meio dos professores e funcionários me deu todo o suporte para a realização desta tese.

Aos meus familiares, em especial ao meu pai Clovis, minha mãe Sílvia e minha irmã Bruna, por não hesitarem em me dar apoio, carinho e conselhos em todos os momentos.

À minha esposa Rafaela, por estar sempre perto e enfrentar comigo todos os momentos deste trabalho.

Aos meus parceiros de sala, Carolina Tatiani e Fernando Bruno, por partilharem as alegrias, confissões e desesperos ao longo destes anos.

Aos companheiros de grupo de pesquisa e aos amigos da pós-graduação.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, a qual agradeço.

RESUMO

A ausência de metais ligados (proteína Apo) e mutações na enzima superóxido dismutase humana (SOD1) tem sido associado com a doença Esclerose Lateral Amiotrófica (ELA). A ELA é uma doença neurodegenerativa, que causa a morte de neurônios motores, tanto do córtex cerebral quanto da medula espinhal, levando o portador a um estado avançado de atrofia muscular. Acredita-se que, no mundo, a prevalência de ELA seja de 3-8 portadores para cada 100 mil habitantes. Utilizando o modelo baseado em estrutura, e modificando a energia de interação entre resíduos de aminoácidos do sítio de ligação ao metal, e entre a ligação dissulfeto C57-C146, detectamos diferenças entre o enovelamento do monômero apo e holo, nas formas oxidada e reduzida, da SOD1. A presença de íons metálicos claramente diminui a barreira de energia livre e também sugere que o caminho de enovelamento sofra mudanças para atingir o estado nativo. A cinética de enovelamento das formas apo também se correlaciona com a quantidade de barreira de energia livre no processo de enovelamento. Além disso, a estabilidade do estado nativo é significativamente afetada pela ausência de íons metálicos. Quanto à ligação dissulfeto, os resultados mostram que a sua ausência diminui a estabilidade da estrutura nativa e afeta menos o estado de transição, sugerindo que esta seja feita tardiamente no processo de enovelamento. Além disso, investigamos o efeito da concentração no enovelamento do dímero, e, incluindo simulações com o método ABF, buscamos elucidar algumas questões sobre o processo de dimerização da SOD1 na presença e ausência de metais.

Palavras-chave: modelo baseado em estrutura, SOD1, ELA, Método de força adaptativa.

ABSTRACT

The absence of bound metals (Apo protein) and mutations in the Cu-Zn Human Superoxide Dismutase (SOD1) have been associated with Amyotrophic Lateral Sclerosis (ALS) disease. ALS is a neurodegenerative disease, which causes the death of motor neurons in both the cerebral cortex and the spinal cord, leading to an advanced muscle atrophy state. It is believed that in the world, the ALS prevalence is 3-8 per 100,000 inhabitants. Using a structure based-model and modifying the energy of interaction between amino acids residues in the metal-binding site and between the disulfide bond C57-C146, we detected differences between the folding of the apo and holo, oxidized and reduced monomeric SOD1. The presence of metal ions clearly decreases the free energy barrier and also suggests that the folding pathway may change to reach the native state. The kinetics of folding of the apo forms also correlates with the amount of free energy barrier in the folding process. In addition, the stability of the native state is significantly affected by the absence of the metal ions. As regarding to the disulfide bond, the results show that its absence decreases the stability of the native structure and affects less the transition state, suggesting that it is possible made late in the folding process. Furthermore, we investigated the effect of the concentration on the dimer folding process, and including all-atoms simulations and ABF method, we search to elucidate some questions about SOD1 dimerization process.

Keywords: structure-based model, SOD1, ALS, Adaptive Biasing Force.

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	13
3	METODOLOGIA	14
3.1	Dinâmica Molecular	14
3.1.1	Condições periódicas de contorno	15
3.2	Replica-Exchange (REMD)	16
3.3	Modelo Baseado em Estrutura (SBM)	17
3.3.1	Modelo C α – Vanilla	17
3.3.2	Coordenada de reação	18
3.3.3	Interações Não-Nativas	20
3.4	Adaptive Biasing Force Method (ABF)	21
3.5	Detalhes Computacionais	22
3.5.1	SBM-C α com REMD	23
3.5.2	Regiões no perfil de energia livre	26
3.5.3	ABF	26
3.6	Simulações de controle	27
3.6.1	Simulações de cinética de enovelamento com o SBM-C α	28
3.6.2	All-Atoms	28
4	ANÁLISES APLICADAS	29
4.1	Perfil de energia livre	29
4.2	Formação de contatos entre resíduos de aminoácidos	30
4.3	Tempo de enovelamento	30
4.4	Rota de enovelamento	31
5	RESULTADOS E DISCUSSÕES	32
5.1	Enovelamento do monômero utilizando SBM-Cα e REMD	32
5.1.1	Simulações de controle – cinética de enovelamento	41
5.2	Enovelamento do dímero	42
5.2.1	Utilizando SBM-C α e REMD	42
5.2.2	Simulações de controle - Cinética de enovelamento do dímero	46
5.2.3	Estabilidade dímero-monômero	47

5.3	Simulações de dimerização	48
5.3.1	Utilizando o Adaptive Biasing Force (ABF)	48
5.3.2	Simulações de controle – Sistema livre	51
6	CONCLUSÕES	56
	REFERÊNCIAS	58
	APÊNDICE A - WHAM - <i>Weighted Histograms Analysis Method</i>	67
	APÊNDICE B - Artigo publicado durante o doutorado	69

Capítulo 1

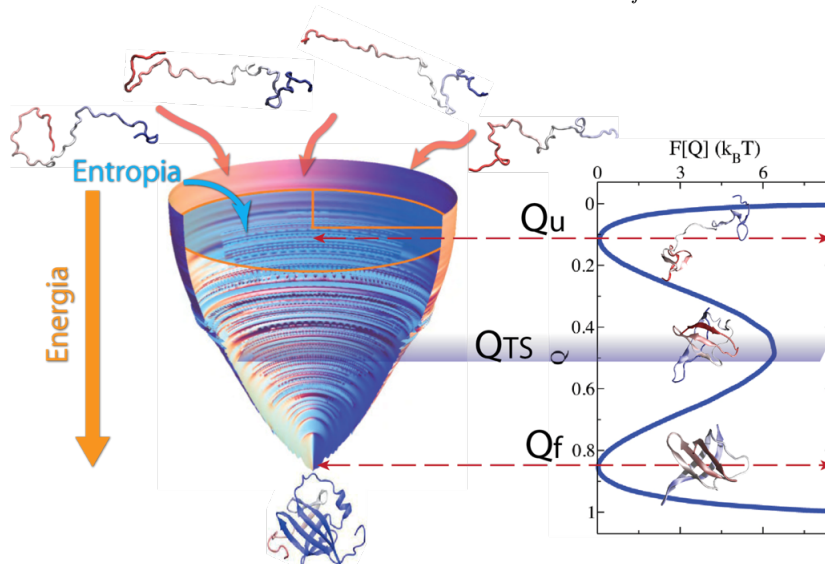
Introdução

Grande parte das transformações moleculares envolvidas no metabolismo celular são mediadas via catálise protéica. As proteínas, além de desempenharem funções catalíticas, também exercem funções estruturais e regulatórias nas células (1). É de grande valia o conhecimento da estrutura protéica para inferirmos aspectos relacionados a função desempenhada por cada proteína, porém, além dos parâmetros estruturais, também é de suma importância o conhecimento dos fatores físico-químicos que regulam o processo de estruturação das proteínas. Para desempenharem uma função biológica nas células, as proteínas devem se moldar em uma estrutura tridimensional funcional, chamada de estrutura nativa (1). O mecanismo, que as levam de uma estrutura linear primária até uma estrutura funcional, é chamado de enovelamento.

Nas últimas décadas, a teoria da superfície de energia tem obtido grande sucesso em explicar aspectos qualitativos e quantitativos envolvidos no estudo teórico e experimental do enovelamento de proteínas (2-9). Nesta teoria, conceitualmente, a superfície de energia possui uma forma afunilada, como representada na Figura 1.1. Utilizando conhecimentos termodinâmicos, considera-se que no topo deste funil de energia há uma alta entropia conformacional dada a grande quantidade de estruturas desenoveladas possíveis para uma mesma proteína. A medida que a proteína começa a se enovelar e formar contatos entre pares de resíduos de aminoácidos que diminuem a sua energia global, temos que esta começa a ser direcionada para o estado nativo, mínima energia do sistema, passando por estados parcialmente enovelados (10).

Para conseguir se enovelar em um tempo hábil para desempenhar uma função biológica, a superfície de energia deve ser íngreme o suficiente para que a proteína consiga passar por algumas rugosidades e mínimos locais (5, 12-16), ocasionados devido ao que chamamos de frustração (12). A frustração energética em proteínas acontece devido a impossibilidade de satisfazer de forma concomitante todas as interações que são

Figura 1.1: **Diagrama esquemático do relevo da superfície de energia.** Esquerda: Funil de enovelamento com alta entropia conformacional na parte superior, e baixa energia e entropia na parte inferior, onde está localizada a estrutura nativa das proteínas. Direita: Perfil de energia livre em função da fração de contatos nativos Q formados. Q_u = Conformação desenovelada. Q_{TS} = Conformação no estado de transição. Q_f = Conformação enovelada.



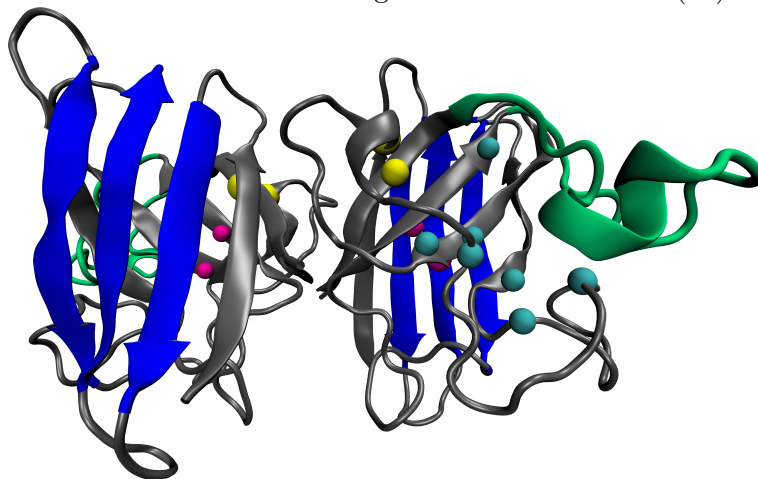
Fonte: Figura adaptada de Oliveira et al (11).

energeticamente favoráveis para o enovelamento (17-20), de tal forma que a proteína necessite fazer alguns contatos entre pares de resíduos de aminoácidos que terão de ser desfeitos ao final do processo. Estes contatos, por não estarem presentes na estrutura nativa, são chamados de contatos não-nativos. Apesar da evolução natural das proteínas ter sido responsável por fazer com que o estado nativo seja minimamente frustrado (17-19, 21, 22), estudos têm mostrado que um pouco de frustração energética pode favorecer o processo de enovelamento (23-25), principalmente nas proteínas que formam mais contatos não-nativos no estado de transição, do que nos estados desenovelados (Para maiores detalhes, veja artigo no Apêndice B).

O aprofundamento dos princípios envolvidos no processo de enovelamento das proteínas podem ajudar no fornecimento de respostas para alguns dos problemas da genômica e da biotecnologia, como por exemplo no desenho consistente de fármacos e enzimas, assim como na compreensão de mecanismos que podem ajudar no controle de doenças genéticas (26) e neurodegenerativas, incluindo as doenças de Alzheimer, Parkinson e Esclerose Lateral Amiotrófica (ELA) (24, 27, 28). A ELA, por exemplo, é uma doença neurodegenerativa que causa a perda dos neurônios motores, levando seus portadores a ficarem incapazes de realizar qualquer tipo de movimento voluntário, tal como andar, falar e engolir. Acredita-se que por ano, no mundo, surja de 2-3 novos casos de ELA a cada 100 mil habitantes (29).

Mais de 100 tipos de mutações na enzima SOD1 têm sido encontradas em pacientes portadores de ELA (27, 28, 30), e falhas no enovelamento da SOD1 têm sido relacionadas com a formação de agregados de fibras amilóides. Tais agregados são típicos e característicos de outras estruturas amiloidogênicas encontradas em outras proteínas relacionadas com doenças neurodegenerativas (31). A *Zn,Cu Human Superoxide Dismutase* (SOD1) é uma enzima homodimérica na qual cada monômero é formado por 153 aminoácidos, um íon de zinco e um de cobre (32-34), como representada na Figura 1.2. Esta enzima, em sua forma holo nativa, protege as células do *stress* oxidativo, catalisando a dismutação do radical superóxido (O_2^{-1}) em oxigênio molecular e peróxido de hidrogênio (35, 36). Os íons de Zn e Cu, além de desempenharem um papel importante na atividade enzimática da SOD1, também atuam de forma fundamental na manutenção estrutural e na estabilidade da enzima nativa: principalmente na manutenção do Loop IV e Loop VII, os quais são responsáveis por formar um gradiente eletrostático que guia o superóxido até o sítio ativo.

Figura 1.2: **Representação esquemática do homodímero da Superóxido Dismutase de humanos (SOD1).** PDB code: 1HL5. As folhas- β 2,3 e 6 estão destacadas em azul. Em verde temos a região do Loop Eletrostático (Loop VII, resíduos 121-142 de cada monômero). As esferas na cor ciano representam o $C\alpha$ dos resíduos que se ligam ao cobre e ao zinco. As esferas amarelas representam os $C\alpha$ das cisteínas 57 e 146. As esferas em rosa, indicam a localização dos resíduos 47 e 117, os quais apresentam um elevado valor- ϕ (37, 38). As estruturas secundárias foram identificadas segundo o critério SCOP (39)



Fonte: Elaborada pelo Autor.

Em sua forma nativa, holo e com a ponte dissulfeto entre a C57 e C146 oxidada, a SOD1 é altamente estável, apresentando temperaturas de enovelamento superiores a 90°C (40), assim como tem sido mostrado por experimentos envolvendo desnaturação química ou térmica. Mutantes da SOD1 que perdem a capacidade de se ligar aos metais (36, 41-44) apresentam uma estabilidade menor, reduzindo a sua temperatura de

melting para valores em torno de 52°C. Acrescido a este fato, na ausência da ponte dissulfeto entre as cisteínas 57 e 146, a SOD1 apo fica ainda menos estável, com temperatura de enovelamento da ordem de 42°C, e se mostra mais propícia a permanecer na forma monomérica (45, 46).

Assim como outras metaloenzimas ou proteínas oligoméricas, os mecanismos de enovelamento da SOD1 têm se tornado um problema complexo para se estudar devido ao seu longo processo de maturação pós-traducional (47), como por exemplo: a dimerização, a formação de pontes dissulfeto, e a inserção dos metais. Apesar de todas essas dificuldades, a superóxido dismutase humana têm sido intensamente investigada nas últimas décadas. Um grande número de publicações têm mostrado evidências que mutantes da SOD1 podem agregar e levar ao desenvolvimento da Esclerose Lateral Amiotrófica. No entanto, ainda há na literatura discussões a respeito do que pode estar causando a doença: se é a formação dos agregados, ou da perda de função da SOD1, pois, além das mutações, a ausência dos íons de cobre e zinco também têm sido associados como possíveis agentes para o comportamento patogênico da enzima (28).

Medidas de atividade enzimática indicam que cada monômero da SOD1 trabalha de forma independente um do outro quando a proteína está em sua forma dimérica nativa (48), porém, a atividade enzimática do dímero é em torno de uma ordem de grandeza maior do que quando cada monômero atua de forma isolada (49, 50). Neste caso, o fato dos monômeros conseguirem desempenhar suas funções quando estão isolados, e trabalharem de forma independente quando estão juntos no dímero, fizeram surgir algumas questões à respeito do porque a SOD1 seria uma proteína dimérica (50). Por que a natureza produziria um dímero feito de dois monômeros idênticos que trabalham de forma independente? Além disso, outras questões envolvendo: a presença da formação de pontes dissulfeto, o papel dos metais e como eles podem afetar o processo de enovelamento, assim como a presença de estados intermediários, são elementos relevantes no estudo do enovelamento da SOD1 descrito neste trabalho.

A técnica de Dinâmica Molecular têm sido utilizada nas últimas décadas para ajudar na compreensão de alguns problemas enquadrados pela Física Biológica, e o entendimento dos mecanismos de enovelamento de metaloproteínas diméricas têm sido parte deste grande desafio. Simulações computacionais tem nos ajudado a obter um maior conhecimento à respeito de como as proteínas se enovelam (15, 26, 51-53), e o uso de modelos simplificados que levam em consideração a teoria da superfície de energia, tais como o Modelo Baseado em Estrutura (SBM) (4, 54), tem sido capazes de capturar algumas características essenciais do problema de enovelamento de proteínas (11, 23, 25, 54, 55). Tais modelos, apesar de serem simplificados, apresentam boa concordância com outros modelos e métodos, analíticos ou experimentais (2-5, 12, 56-

59).

Neste estudo, utilizando uma simplificação do SBM, o SBM- C_α , nós buscamos compreender quão bem estes modelos simplificados conseguem capturar o comportamento termodinâmico e cinético envolvido no enovelamento da SOD1, elucidando diferenças entre o enovelamento do monômero na forma holo, apo, reduzida ou oxidada. Além disso, ainda utilizando o SBM- C_α , investigamos o efeito da concentração proteica no enovelamento do dímero, e utilizando técnicas de simulação computacionais mais complexas, incluindo simulações com todos os átomos e solvente explícito, analisamos o efeito da presença e da ausência dos metais no processo de dimerização.

Capítulo 2

Objetivos

Este estudo tem como objetivo compreender algumas informações a respeito do processo de enovelamento da enzima superóxido dismutase de humanos (SOD1) utilizando técnicas de dinâmica molecular, focando principalmente nas respostas para estas quatro perguntas:

- I) Qual o efeito dos metais (Zn e Cu) no enovelamento de cada monômero?;
- II) Qual o efeito da ponte dissulfeto (C57-C146) sobre o enovelamento dos monômeros?;
- III) Quais as possíveis razões para a enzima atuar na forma dimérica?;
- IV) Como a ausência dos íons metálicos de Zn/Cu afetam o processo de dimerização?

Capítulo 3

Metodologia

3.1 Dinâmica Molecular

Simulações de Dinâmica Molecular têm sido utilizada com frequência para investigação dos mais diversos comportamentos de biomoléculas em nível molecular (60, 61). A quantidade elevada de átomos de uma macromolécula biológica inviabiliza, atualmente, o seu estudo computacional pelas equações da Mecânica Quântica. A alternativa é usar a Dinâmica Molecular, que utiliza os princípios da Mecânica Clássica para descrever como variam as velocidades, posições e orientação dos átomos em função do tempo.

Para descrever o comportamento dos átomos ao longo de uma simulação computacional por dinâmica molecular, é necessário a utilização de potenciais que consigam ser simples ao ponto de viabilizar a sua utilização, mas também serem capazes de captar de forma realística o comportamento do modelo em estudo. A soma de todos estes potenciais, que descrevem diferentes graus de liberdade, é chamado de campo de força.

A forma do campo de força pode variar dependendo do nível de simplificação do modelo do sistema, mas em geral contém termos que descrevem interações entre átomos ligados e não-ligados, por meio de potenciais de longo e curto alcance, como representado pelo campo de força da equação 3.1. Os três primeiros termos do potencial descrevem interações entre átomos ligados, representando respectivamente: a ligação covalente entre dois átomos separados por uma distância r , a distorção angular θ formada por três átomos ligados sequencialmente, e a torção em torno das ligações químicas formada pelo ângulo diedral ϕ entre quatro átomos sequenciais:

$$\begin{aligned}
V = & \sum_{bonds} \epsilon_r (r - r_0)^2 + \sum_{angles} \epsilon_\theta (\theta - \theta_0)^2 \\
& + \sum_{backbone} \epsilon_\phi [1 + \cos(n\phi - \delta)] \\
& + \sum_{contacts} 4\epsilon_c \left[\left(\frac{d_{ij}}{r_{ij}} \right)^{12} - \left(\frac{d_{ij}}{r_{ij}} \right)^6 \right] + \sum_{elec} \epsilon_e \left(\frac{q_i q_j}{r_{ij}} \right)
\end{aligned} \tag{3.1}$$

Os dois últimos termos do campo de força 3.1 representam as interações entre átomos não-ligados, sendo eles respectivamente o potencial de Lennard-Jones entre átomos i e j separados por uma distância r_{ij} , e o potencial eletrostático de longo alcance, representado pela lei de Coulomb, sendo q_i e q_j as cargas parciais dos átomos.

Definido o campo de força, o próximo passo é escrever a Hamiltoniana do sistema em função dos momentos das partículas (Energia Cinética) e das energias potenciais. A partir do cálculo da Hamiltoniana é possível, por meio das equações de movimento da mecânica clássica, acompanhar a evolução temporal do sistema. Porém, por ser um cálculo complexo, há a necessidade de utilização de métodos de integração numérica para que possamos, por meio de uma condição inicial, determinar a posição e as velocidades dos átomos em um intervalo de tempo posterior. Alguns dos integradores numéricos mais utilizados são o *Velocit-Verlet* e *Leap-Frog* (62).

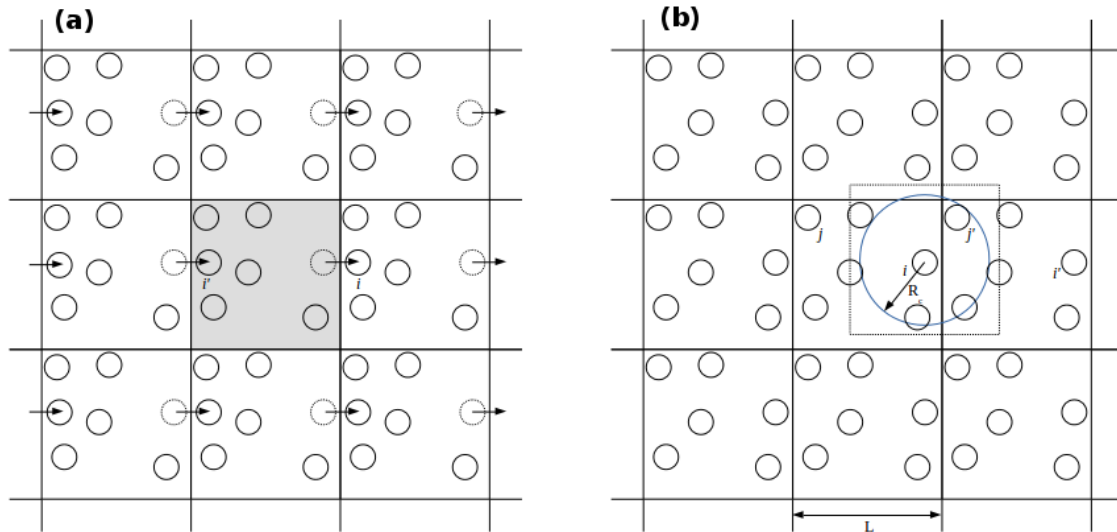
3.1.1 Condições periódicas de contorno

Sistemas reais apresentam uma quantidade de átomos cuja ordem se equivale ao número de Avogadro, porém, simular de forma computacional um sistema deste tamanho é inviável dado o poder computacional que temos hoje em dia. Deste modo, as simulações computacionais são realizadas com número relativamente reduzido de partículas que viabilizam a sua realização, no entanto, algumas condições causadas por essa limitação, tais como: efeito de borda e contorno do sistema, devem ser examinados. Para minimizar estes efeitos, utiliza-se o método de replicação das caixas de simulação por todo o espaço conformacional, e permitimos que quando uma partícula i sai da caixa de simulação, outra partícula i' idêntica, chamada de imagem, entre na caixa de simulação. Estes processos estão inseridos dentro do que chamamos de condições periódicas de contorno (PBC), ilustrada na Figura 3.1 (a).

Ao permitir a entrada de uma partícula imagem na caixa de simulação, garantimos que o número de partículas na caixa se mantenha constante ao longo do processo, porém, ao utilizar condições periódicas de contorno e permitir esta entrada e saída

de partículas, também é necessário garantir que a partícula i não interaja com a sua imagem i' . Para isto, delimita-se uma região por meio de uma esfera de interação que apresenta um Raio de Corte (R_c). Somente partículas cujo centro de massa estão dentro dessa região limitada por R_c podem interagir. Por definição, $R_c \leq L/2$, sendo L o valor de menor dimensão dos lados da caixa de simulação. A figura 3.1 (b) ilustra a aplicação do Raio de Corte.

Figura 3.1: **Representação esquemática bidimensional da replicação da caixa de simulação utilizando condições periódicas de contorno.** (a) A caixa central (em cinza) é replicada em todas as coordenadas espaciais que definem o sistema. Quando uma partícula i da caixa central se desloca para uma caixa réplica, uma partícula i' automaticamente entra na caixa central. Este mesmo processo acontece para as outras caixas, mantendo assim, a concentração inicial de partículas inalterada. (b) O Raio de Corte traçado a partir de uma partícula de interesse i , define uma região (em azul) de interação entre esta partícula e as suas vizinhas, de tal modo a impedir que ela interaja com a sua própria imagem i' , e ao mesmo tempo, com as partículas vizinhas e suas respectivas imagens.



Fonte: Figura extraída de (63)

3.2 Replica-Exchange (REMD)

Varrer o espaço conformacional de uma forma ampla, de tal modo a capturar a maior quantidade possível de estados acessados por um sistema biológico é um problema a ser enfrentado, principalmente quando se estuda conformações que estão separadas por barreiras energéticas altas, como é o caso do enovelamento da enzima SOD1 objeto de estudo deste trabalho. Para tratar este problema, utiliza-se o método

de dinâmica molecular por troca de réplicas (*REMD*) (64). Neste método, cópias de um determinado sistema são colocadas para simular de forma paralela em condições diferentes de temperatura (T), de tal forma a permitir a troca entre pares de réplicas de acordo com uma probabilidade de transição entre réplicas vizinhas, exemplificados pela Equação 3.2. Deste modo, é possível que as réplicas vizinhas i e j troquem de configuração de acordo com a probabilidade de transição baseado nas energias potenciais (U) das réplicas vizinhas, e da frequência imposta de forma externa pelo utilizador do método.

$$P(i \leftrightarrow j) = \min \left(1, \exp \left[\frac{1}{K_B T_i} - \frac{1}{K_B T_j} \right] (U_i - U_j) \right) \quad (3.2)$$

3.3 Modelo Baseado em Estrutura (SBM)

É dito um modelo baseado em estrutura (SBM) qualquer modelo que utiliza parâmetros estruturais obtidos por técnicas experimentais (cristalografia de raios-X e ressonância magnética nuclear), a fim de construir o potencial energético que define as conformações de uma dada proteína. Os parâmetros da expressão do campo de força são obtidos a partir dos dados estruturais da proteína já enovelada, de tal forma que, o mínimo de energia corresponde à conformação da estrutura nativa da proteína.

No caso dos modelos minimalistas baseados em estrutura, a primeira simplificação se refere ao potencial de interação entre os componentes do sistema. A segunda se refere aos próprios componentes do sistema: todos-os-átomos ou somente carbonos alfas (C_α). Aplica-se o modelo com todos-os-átomos quando as informações sobre as cadeias laterais são importantes para o entendimento do problema em estudo, caso contrário, pode-se utilizar o modelo C_α (54), pois, além de ser o modelo melhor testado e utilizado pela comunidade científica, permite uma economia de tempo computacional de até 10 vezes se comparado ao modelo com todos-os-átomos.

3.3.1 Modelo C_α - *Vanilla*

A simplicidade do modelo C_α está baseada na referência à cadeia de resíduos de aminoácidos como uma cadeia de esferas simples centralizadas nas posições dos carbonos alfas. Tais esferas são mantidas juntas por meio de ligações e ângulos de interação, e a geometria nativa fica contida no potencial diedro e no termo de interação não-local (54).

A expressão que define a energia de uma configuração Γ baseada na conformação nativa Γ^0 para o modelo C_α é dada por:

$$\begin{aligned}
V(\Gamma, \Gamma^0) = & \sum_{bonds} \epsilon_r (r - r_0)^2 + \sum_{angles} \epsilon_\theta (\theta - \theta_0)^2 \\
& + \sum_{backbone} \epsilon_\phi \left\{ [1 - \cos(\phi - \phi_0)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_0))] \right\} \quad (3.3) \\
& + \sum_{contacts} \epsilon_c \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{non-cont} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12}
\end{aligned}$$

De acordo com a expressão [3.3](#), o potencial que rege este modelo é calculado via a contabilização de 5 termos referentes a algumas das interações que atuam na estrutura nativa. O primeiro termo remete a um potencial harmônico representando a ligação entre dois carbonos α adjacentes, no qual r_0 é a distância entre os dois carbonos α ligados entre si na estrutura nativa. A segunda somatória também forma um potencial harmônico, mas desta vez, um potencial harmônico angular formado por três carbonos α em sequência na cadeia polipeptídica, no qual θ_0 é o ângulo formado pelos três resíduos na conformação nativa. Já o terceiro termo da expressão [3.3](#) é uma função de três mínimos, que leva em consideração a torção realizada pela cadeia, no qual, ϕ_0 é o ângulo diédrico formado por quatro carbonos α em sequência. O quarto termo contabiliza a interação entre o carbono α_i e o carbono α_j que formam um contato na estrutura nativa, para isso é utilizado um potencial 10-12, no qual d_{ij} é o valor da distância entre estes carbonos na estrutura nativa. Por fim, o último termo remete a todos os carbonos α que não realizam contato nativo algum. Este termo é utilizado para manter a distância de máxima aproximação entre os carbonos α , no qual σ_{NC} possui valor de $4\text{\AA}$ e está correlacionado com o volume ocupado por um carbono no modelo. As constantes: ϵ_r , ϵ_θ , ϵ_ϕ e ϵ_{NC} são todas dadas em unidades de ϵ_c e valem, respectivamente, 100, 20, 1 e 1 ([23](#), [54](#)).

3.3.2 Coordenada de reação

Coordenadas de reação, ou parâmetros de ordem, são utilizados para acompanhar os passos obtidos ao longo do tempo em uma simulação computacional, ou em experimentos *in vitro*. Na maioria dos estudos de enovelamento de proteína feitos computacionalmente, utiliza-se a fração de contatos nativos Q como parâmetro de ordem. Esta coordenada de reação nos diz quão próxima uma estrutura “ i ” está da estrutura nativa. Tal comparação é feita calculando ao longo da trajetória simulada a

quantidade de contatos realizados por cada resíduo de aminoácido, comparando estes contatos com um mapa pré-estabelecido obtido quando a proteína se apresenta em sua forma enovelada. A fração de contatos nativos Q é normalizada pelo número total de contatos nativos, e varia de 0 à 1, de tal forma que quanto mais próximo de 1, mais parecida está a configuração i com a conformação nativa da proteína.

A fração de contatos nativos (Q) tem tido sucesso em explicar alguns fenômenos relacionados com o enovelamento de proteínas, porém, experimentalmente, não é possível dizer quais e quantos contatos os resíduos de aminoácidos de uma determinada proteína esta realizando no tempo “ t ” do experimento. Neste caso, outras coordenadas de reação que podem ser obtidas também de forma experimental, parecem ser uma boa opção quando se busca comparar os dados computacionais com os experimentais. Neste caso, parâmetros como o Raio de Giro (R_g) e Área Acessível ao Solvente (SASA) fazem uma correlação melhor entre simulação e experimento. O β -Tanford é uma coordenada que envolve a variação dos valores de R_g , e pode ser calculada tanto de forma experimental quanto computacional. Artigos experimentais de enovelamento da SOD1 utilizam o β -Tanford para elucidar o perfil de energia livre durante o enovelamento desta enzima. A equação que remete ao cálculo desta coordenada pode ser vista a seguir:

$$\beta - Tanford = \frac{a_r + b_r(R_g^D - R_g)}{a_r + b_r(R_g^D - R_g^N)} \quad (3.4)$$

R_g^D e R_g^N são, respectivamente, os valores do raio de giro no “estado” desenovelado e nativo. R_g é o valor do raio de giro para cada conformação acessada pela proteína durante a simulação. Ao longo das nossas simulações, R_g recebe valores que depende de cada conformação acessada, já os demais termos da equação 3.4, são mantidos fixos. Assim como Q , o β -Tanford também tem seus valores entre 0 e 1, e quanto mais próximo de 1, mais enovelada está a proteína. Os termos a_r e b_r são obtidos de forma experimental dependendo do desnaturante que se utiliza para acompanhar o processo de *folding/refolding*. No nosso caso, seguimos os valores determinados por Geierhaas *et. al.* (65) que usam ureia como desnaturante.

O β -Tanford também nos permite estimar os valores- m (*m-values*), os quais são quantidades fortemente correlacionadas com a quantidade de superfície hidrofóbica enterrada no processo de enovelamento. Esses valores também quantificam a cooperatividade do enovelamento e desenovelamento na presença de desnaturante. A relação entre β -Tanford e os valores- m é dado pela equação 3.5:

$$\beta - \text{Tanford} = \frac{m_{D-TS}}{m_{D-N}} \quad (3.5)$$

em que D-TS indica a mudança conformacional entre o estado desenovelado e o estado de transição, e D-N entre o estado desenovelado e o nativo. De posse de valores-m experimentais, é possível usar este procedimento para provarmos quão bem nosso *ensemble* de transição, obtido computacionalmente, se correlaciona com os dados experimentais. Essa é uma análise importante a se fazer quando se usa potenciais que são baseados em estrutura, na tentativa de verificar se outros *ensembles*, além do nativo, está sendo amostrado de forma confiável.

3.3.3 Interações Não-Nativas

O modelo SBM- C_α tem se mostrado muito bom em acessar configurações localizadas depois do estado de transição das proteínas, porém, acredita-se que a sua representação no estado desenovelado possa ser um pouco precária. Desconsideradas no modelo SBM- C_α *Vanilla*, as interações entre pares de aminoácidos não-ligados que caracterizam um contato não-nativo, têm se mostrado importante para a compreensão de fenômenos que não são bem explicados pelos modelos que não levam em consideração a frustração no enovelamento de proteínas (23, 25).

Em estudo prévio (25), adicionando um potencial ao SBM- C_α *Vanilla* para contabilizar as interações não-nativas durante o enovelamento, obtivemos como resultado que a frustração energética, neste caso homogênea (independente de qual resíduo de aminoácido está realizando o contato não-nativo), era capaz de favorecer o enovelamento de algumas proteínas, e de prejudicar o enovelamento de outras. Tais dados foram corroborados e explicados utilizando o modelo analítico proposto por Clementi-Plotkin (24), e o artigo publicado na revista *Biophysical Journal* encontra-se no Apêndice B desta tese.

Recentemente, Silva e colaboradores (66), adicionando frustração heterogênea nos contatos não-nativos entre aminoácidos hidrofóbicos, conseguiu compreender a cinética de enovelamento da α -Espectrina, sugerindo a origem do atrito interno nestas proteínas, até então desconhecido na literatura. Neste estudo, as interações não-nativas hidrofóbicas são pesadas de acordo com os resíduos de aminoácidos que estão realizando determinado contato. O valor energético k_{ij} para a formação de cada contato não-nativo é fornecido pela tabela de Miyazawa-Jernigan (67).

3.4 Adaptive Biasing Force Method (ABF)

O método de força adaptativa (*Adaptive Biasing Force* - ABF) (68) é uma metodologia de integração termodinâmica na qual a variação de energia livre de um sistema ΔG é calculada por meio da integração da derivada da energia livre G em função de uma coordenada de reação ξ , que pode ser a distância entre dois centros-de-massa, como utilizado no caso do estudo da dimerização da SOD1 neste trabalho. Deste modo, a variação de energia livre ΔG é dada pela seguinte expressão:

$$\Delta G = G(\xi_1) - G(\xi_0) = \int_{\xi_0}^{\xi_1} \frac{\delta G}{\delta \xi} \delta \xi \quad (3.6)$$

Da mecânica estatística, temos que $G = -KT \ln Z$, no qual Z é a função de partição de um ensemble canônico. Deste modo, podemos substituir $Z = \sum_i e^{-\frac{H_i}{KT}}$, na expressão de G e na Equação 3.6. Feita as devidas substituições, torna-se possível rearranjar os termos da equação 3.6 de tal modo que a variação da energia livre se torne correlacionada com a variação da energia total média do sistema $\langle \frac{\delta H}{\delta \xi} \rangle$, e, se considerarmos a nossa simulação como sendo um processo isotérmico, a variação da energia total média se torna equivalente à variação da energia potencial em função de ξ , uma vez que nestas condições, a derivada da energia cinética tende a zero:

$$\frac{\delta G}{\delta \xi} \approx \left\langle \frac{\delta H}{\delta \xi} \right\rangle_{\xi} = \left\langle \frac{\delta V}{\delta \xi} \right\rangle_{\xi} \quad (3.7)$$

A variação da energia potencial está associada com a força resultante atuando sobre as partículas de um sistema:

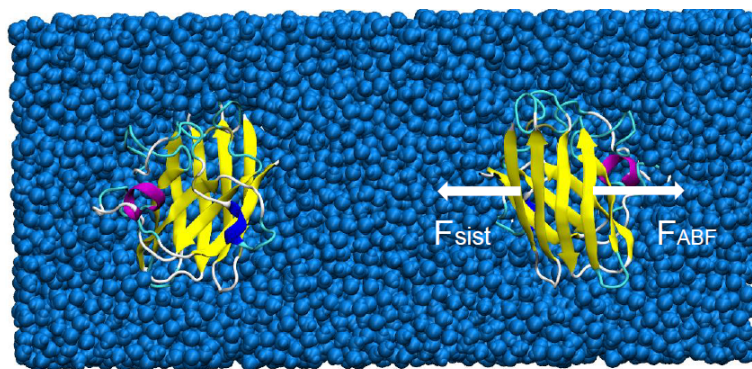
$$\frac{\delta G}{\delta \xi} \approx \left\langle \frac{\delta V}{\delta \xi} \right\rangle_{\xi} \approx - \left\langle \vec{F}_{sys} \right\rangle_{\xi} \quad (3.8)$$

Neste caso, de posse do valor da força que atua sobre o sistema, em função do parâmetro de ordem ξ , é possível encontrar o valor da energia livre também em função desta coordenada de reação. Além de ser capaz de nos informar dados termodinâmicos de um sistema biológico, o ABF tem como característica aplicar sobre o sistema a mesma força que age nele de forma natural, aqui chamada de F_{sys} , só que em sentido contrário à esta. Sendo assim, a força total que atua no sistema tende a ser nula, já que

a força aplicada pelo ABF (F_{ABF}) tem mesmo módulo que F_{sys} mas está em sentido contrário. Neste caso, o sistema começa a se comportar de modo difusivo, e estados conformacionais que provavelmente não seriam acessados pelas moléculas biológicas, começam a ser possíveis de serem visitados, garantindo uma boa amostragem do sistema, mesmo estando estes estados em regiões de transição com altas barreiras.

Esta metodologia será usada para elucidar questões à respeito de como ocorre o processo de dimerização da SOD1 e porquê esta enzima é um dímero na maioria dos organismos vivos. Uma representação esquemática de como este método atua no sistema pode ser visto na Figura 3.2. O ABF permite que várias simulações de um mesmo sistema sejam executadas de forma paralela e independente, podendo definir em cada uma destas simulações um intervalo no qual a coordenada de reação pode variar, cada um destes intervalos são chamados de “janelas”. Ao final, para construção do perfil de energia livre, pode-se concatenar os dados de todas as janelas, varrendo todo o intervalo definido por ξ .

Figura 3.2: **Representação esquemática da metodologia utilizada pelo ABF.** O método ABF atua na proteína (monômero da SOD1, na cor amarela) aplicando uma força (F_{ABF}) contrária à direção da força natural do movimento do sistema (F_{sist}). Neste *cartoon*, as esferas em azul estão representando as moléculas de água na caixa de simulação.



Fonte: Elaborada pelo Autor

3.5 Detalhes Computacionais

A estrutura nativa da Zn-Cu Superóxido Dismutase (SOD1) de humanos, utilizada neste trabalho, está depositada no *Protein Data Bank* (69) com código 1HL5. Nas simulações de monômero da SOD1, utilizamos a cadeia A do *PDB* em questão. Já para o dímero, utilizamos as cadeias A e H.

Neste trabalho, foram realizados dois tipos de simulações envolvendo dinâmica

molecular: uma para estudar o processo de enovelamento com o SBM-C α , e outra para estudar o processo de dimerização com os monômeros já enovelados. As simulações para estudo de enovelamento da SOD1 podem ser divididas em três tópicos: (i) enovelamento de um monômero isolado; (ii) enovelamento de dois monômeros para a formação do dímero; (iii) Simulações de controle. Nos itens (i) e (ii) fizemos simulações utilizando o SBM-C α e o método de *Replica-Exchange*. Já no tópico (iii) foram feitas simulações de enovelamento de um, ou dois monômeros, com o SBM, em temperatura baixa, sem o REMD.

Para investigar o fenômeno da dimerização com os monômeros já enovelados, realizamos dois tipos de simulações. Primeiramente, fizemos uso da metodologia do ABF para obtermos dados a respeito da termodinâmica dos sistemas, e com estes dados, fizemos simulações livres, sem o ABF, para controle e confirmação dos resultados obtidos pelo *Adaptive Biasing Force Method*. Maiores detalhes à respeito de cada tipo de simulação podem ser encontrados nas seções seguintes.

3.5.1 SBM-C α com REMD

Para investigar o enovelamento da SOD1 na condição de monômero e dímero, utilizamos o nível de simplificação do SBM-C α , e realizamos as simulações com o método de troca de réplicas a fim de obtermos uma melhor amostragem dos estados conformacionais acessados pela enzima ao longo de seu enovelamento. Fizemos o uso do *GROMACS versão 4.5.5* (70) como software para realizar estas simulações. Utilizamos 24 réplicas para cada sistema, tanto para o monômero quanto para o dímero, estando cada réplica em uma temperatura diferente. No total, cada réplica foi simulada em 5×10^8 passos, utilizando um passo de integração de $0.5 \times 10^{-15} s$. As conformações eram guardadas a cada 2500 passos, e as trocas entre as réplicas vizinhas eram permitidas à cada 10000 passos de simulação. Os arquivos de topologia e velocidades necessários para a simulação no *GROMACS* foram obtidos pela ferramenta online *SMOG* (71). Para simular o enovelamento do dímero, utilizamos um potencial de restrição para impedir que as duas cadeias se afastassem “muito” durante o desenovelamento. O potencial de restrição utilizado segue o seguinte padrão:

$$V(r) = \begin{cases} \frac{1}{2}k(r - r_0)^2, & r \leq r_0 \\ 0 & r_0 < r < r_1 \\ \frac{1}{2}k(r - r_1)^2, & r_1 \leq r < r_2 \\ \frac{1}{2}k(r_2 - r_1)^2(2r - r_2 - r_1), & r \geq r_2 \end{cases}$$

No nosso caso, mantivemos r_0 como sendo nulo. Já os valores de r_1 e r_2 variavam

de modo a dar uma certa liberdade de movimentação para as cadeias além da distância nativa de separação entre os vínculos (um em cada monômero). O valor de k foi mantido fixo e igual a 1kJ/mol.

A temperatura de enovelamento foi definida utilizando como padrão o perfil de energia livre, obtido pelo *WHAM*, como sendo a temperatura capaz de amostrar tanto o estado desenovelado quanto o enovelado, e neste caso, $F(Q=0) \approx F(Q=1)$. O modelo baseado em estrutura faz o uso de unidades reduzidas, neste caso, nossas temperaturas são dadas em unidades reduzidas e com valores próximos de 1.

Encontrar temperaturas reduzidas que correspondam a temperaturas reais não é uma tarefa simples. A relação mais óbvia é obtida pela temperatura de enovelamento (T_f), na qual o sistema computacional compartilha a mesma característica termodinâmica do sistema real: a estabilidade igual a 0 ($F(Q=0) \approx F(Q=1)$). Deste modo, seria simples encontrar, por relação linear (regra de três), temperaturas reduzidas correspondentes as temperaturas reais, associando o valor de T_f da simulação com o T_f (ou T_m - *temperatura de melting*) obtido pelos resultados experimentais. Utilizando a relação linear, tomando como base a correspondência entre os T_f para encontrar qual seria a temperatura ambiente em unidades reduzidas, encontramos valores muito baixos de temperatura, os quais nos davam valores extremamente altos de barreira de desenovelamento, que chegavam a ser iguais a estabilidade exagerada do sistema. Este dado não é verdadeiro, principalmente para o monômero da SOD1, o qual, em temperatura ambiente, possui barreira de desenovelamento maior do que a sua estabilidade.

Neste trabalho, para encontrar a correspondência entre as temperaturas em unidades reduzidas e as temperaturas reais, utilizamos como critério a razão entre a capacidade calorífica (C_p) em função da temperatura. A razão entre o máximo valor de C_p , o qual ocorre em temperatura próxima de T_m , com o valor de C_p em temperatura ambiente T_0 , será usado para encontrar a “temperatura ambiente” em unidades reduzidas. Ao efetuarmos a divisão entre os valores de C_p ($C_p(T_m)/C_p(T_0)$), temos como resultado um valor adimensional, livre de unidades. Este valor nos guiará para encontrarmos a respectiva “temperatura ambiente” de cada sistema, uma vez que o *WHAM* nos permite calcular os valores de C_p dos dados da simulação.

As simulações de enovelamento elaboradas neste estudo descrevem o comportamento de cinco formas monoméricas, identificadas com o prefixo “m” em nossos resultados, e duas “formas” diméricas, identificadas com o prefixo “d”. O enovelamento dos monômeros foram identificados da seguinte forma: (i) mApo_{SH}, para a apo reduzida; (ii) mApo_{conf}, para a apo com ponte dissulfeto conformacional; (iii) mApo_{SS}, para a

apo com ponte dissulfeto oxidada; (iv) mHolo_{conf}, para a holo com ponte dissulfeto conformacional; e (v) mHolo_{SS}, para a holo com ponte dissulfeto oxidada. Por outro lado, o dímero só foi estudado na situação holo e com ponte dissulfeto conformacional. A variável nas simulações de dímero é apenas a “concentração” de proteína, ou liberdade de afastamento entre os monômeros.

As simulações das formas holo foram feitas levando em consideração todos os contatos nativos, e todas as energias de interação entre os resíduos de aminoácidos que formam o sítio de ligação com os metais na forma original obtida pelo SMOG (71), utilizando o “Shadow Map” (72) como algoritmo para obtenção do mapa de contato nativo, apesar de não ter o metal de forma explícita neste modelo. Para os casos apo, devido a similaridade entre as formas cristalográficas das estruturas holo e apo, optamos por utilizar a mesma cadeia do caso holo, porém, o mapa de contato é diferente: foram removidos os contatos entre pares de resíduos que se ligam aos metais. As alterações na ponte dissulfeto são indicadas pelos índices “conf”, “SS” e “SH”. A ponte dissulfeto entre as cisteínas 57 e 146 para os casos identificados com o índice “conf” (mApo_{conf} e mHolo_{conf}) foi modelada como qualquer outro contato nativo, sendo regida pelo potencial “10-12” do campo de força do SBM-C α . Nos casos com índice “SS” (mApo_{SS} e mHolo_{SS}), a ponte dissulfeto foi modelada como sendo uma ligação covalente, regida pelo potencial de “bond” do campo de força, representando uma simulação na qual a ponte está sempre formada. Por fim, para o caso mApo_{SH}, foi retirada a energia de interação entre os resíduos 57 e 146 (e dos contatos entre resíduos de aminoácidos em comum), representando a SOD1 com a ponte dissulfeto reduzida. A tabela 3.1 resume os sistemas estudados:

Tabela 3.1: Sistemas estudados

Sistema	MS ^a	Ponte Dissulfeto ^b	Distância CM-CM ^c
mApo _{SH}	NÃO	Reduzida	—
mApo _{conf}	NÃO	Conformacional	—
mApo _{SS}	NÃO	Oxidada	—
mHolo _{conf}	SIM	Conformacional	—
mHolo _{SS}	SIM	Oxidada	—
dHolo _L	SIM	Conformacional	61Å
dHolo _H	SIM	Conformacional	81Å

^a MS indica a presença dos contatos do sítio metálico;

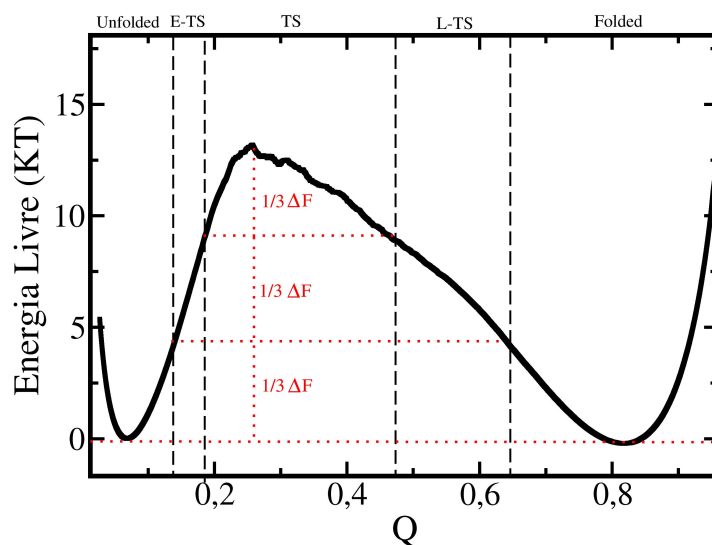
^b Indica o tipo de conformação da ligação entre as cisteínas 57 e 146;

^c Indica o valor de r_1 colocado no potencial de restrição do *GROMACS* entre os centros dos dois monômeros;

3.5.2 Regiões no perfil de energia-livre

Será comum em nossos resultados apresentarmos valores referentes à cinco regiões específicas do perfil de energia livre, referenciadas como: Unfolded, E-TS (*Early TS*), TS, L-TS (*Late TS*) e Folded. Estas regiões representam respectivamente os *ensembles* desenovelado, de pré-transição, transição, de pós-transição e enovelado. O perfil de energia livre em função da coordenada de reação é uma grandeza apropriada para distinguir cada uma dessas regiões, de acordo com suas estabilidades termodinâmicas. Com base na altura da barreira de energia livre, especificamos que cada região fosse delimitada por $\frac{1}{3}$ desta barreira, assim como exemplificado na Figura 3.5.2:

Figura 3.3: **Definição das regiões no perfil de energia livre.** Cada região, separada pelas linhas tracejadas de cor preta, foi delimitada utilizando $\frac{1}{3}$ da altura da barreira de energia livre (ΔF) como parâmetro. As regiões referenciadas como: *Unfolded*, E-TS, TS, L-TS e *folded* remetem respectivamente aos *ensembles* desenovelado, de pré-transição, transição, de pós-transição e enovelado.



Fonte: Elaborada pelo autor

3.5.3 ABF

Partindo da estrutura dimérica nativa extraída do PDB 1HL5, cadeias A e H, construímos a nossa caixa de simulação utilizando o *software VMD (Visual Molecular Dynamics)* (73). As chamadas “janelas” de simulação, cada qual mantendo um intervalo de distâncias entre os centros-de-massa de cada monômero da SOD1, foram colocadas para simular de forma paralela e independente, utilizando o método ABF

disponível no pacote de simulação do *NAMD* - (*NANoscale Molecular Dynamics*) (74). Deste modo, foi possível calcular o perfil de energia livre de dimerização para o caso apo e holo da SOD1.

Neste tipo de método, a proteína foi simulada com todos os átomos (e não só os C_α como no caso do SBM), usando o campo de força do CHARMM e solvente explícito (modelo de água TIP3P) na presença de íons de cloro e sódio (na concentração de 0.15mol/L). Na simulação da SOD1holo, devido a falta de parâmetros para os íons de cobre, substituímos este metal por um zinco. Como não estamos interessados na atividade enzimática da SOD1, que depende do estado de oxidação do cobre, acreditamos que a substituição do cobre por um átomo de zinco, não será responsável por grandes flutuações nos parâmetros estruturais. A simulação da SOD1apo foi feita utilizando as mesmas cadeias A e H, porém, na ausência de metais. Maiores detalhes à respeito dos parâmetros utilizados são encontrados na Tabela 3.2:

Tabela 3.2: Parâmetros utilizados na simulação da SOD1 com o ABF

Sistema	Minimização	Equilibração	Tamanho da Caixa (X,Y,Z)	Temperatura
APO	30000 passos	5 ns	85,95,190 (Å)	298 K
HOLO	30000 passos	5 ns	86,86,86 (Å)	298 K

O passo de integração e a pressão para ambos sistemas foi mantido em 2 fs e 1 atm, respectivamente. Para termos um controle à respeito da convergência do método, realizamos simulações sucessivas de 5ns cada, e utilizamos a diferença entre os valores de energia livre das simulações de um mesmo sistema como critério para estabelecer quando a simulação havia convergido. Neste caso, quando a diferença entre as curvas fosse menor ou igual a 0.3 kcal/mol ($\approx 0.5kT$, para $T=298K$) de energia, parávamos a simulação.

3.6 Simulações de controle

A fim de verificar a consistência dos resultados obtidos pelo método da troca de réplicas e do ABF, realizamos simulações de “controle” para ambos casos. Detalhes à respeito de como estas simulações foram realizadas pode ser encontradas nos tópicos a seguir.

3.6.1 Simulações de cinética de enovelamento com o SBM-C α

O método da troca de réplicas utilizando a simplificação do SBM-C α se mostrou útil para a construção dos perfis termodinâmicos de cada sistema. Para confirmar os dados termodinâmicos obtidos pelo REMD, realizamos 200 simulações de cinética de enovelamento, partindo de estruturas desenoveladas e descorrelacionadas, cada simulação com 2.5×10^7 passos, em temperatura baixa. Os resultados foram calculados efetuando a média das 200 simulações.

3.6.2 All-Atoms

O ABF conseguiu nos dar informações a respeito do perfil de energia livre em função da distância entre os centros-de-massa de cada sistema, apresentando alguns mínimos de energia (locais e globais). Na tentativa de verificar se a presença destes mínimos era oriunda de algum efeito ocasionado pelo método ABF ou se os estados realmente poderiam existir, realizamos simulações mais longas, de 300ns, em cada região de interesse. Elaboramos três simulações de controle neste tópico, uma com o dímero holo (dHolo_N) e duas com o dímero apo (dApo_N e dApo_I). Os índices “N” e “I”, significam, respectivamente, “nativa” e “intermediária”. As simulações referenciadas como caso “nativo” foram iniciadas com as duas cadeias separadas na distância nativa como exemplificadas em maiores detalhes na seção de Resultados e Discussão deste trabalho. Os parâmetros utilizados para a simulação de controle são os mesmos para as três simulações, e estão descritos na Tabela 3.3:

Tabela 3.3: Parâmetros utilizados na simulação da SOD1 de controle

Minimização	Equilibração	Tamanho da Caixa (X,Y,Z)	Temperatura
30000 passos	5 ns	93,93,93 (Å)	298 K

Capítulo 4

Análises aplicadas

4.1 Perfil de energia Livre

O cálculo da energia livre em simulações computacionais não é trivial. O método dos Múltiplos Histogramas (WHAM) [\(75, 76\)](#) têm sido utilizado com frequência e mostrado eficiência na determinação de propriedades termodinâmicas de sistemas biológicos, tais como as proteínas. Por meio de conhecimentos oriundos da mecânica estatística, o algoritmo do *WHAM* consegue fornecer parâmetros termodinâmicos capazes de nos mostrar propriedades importantes à respeito da estabilidade térmica das proteínas, principalmente pelo perfil de energia livre.

Para realizar tal medida, o *WHAM* calcula a densidade de estados ω , em uma dada temperatura T simulada, e faz uma média ponderada utilizando pesos que vem do erro estatístico considerando a densidade de estados das outras temperaturas simuladas. O *WHAM* é um método autoconsistente capaz de extrapolar as grandezas termodinâmicas (entropia, entalpia, energia-livre e calor específico) para temperaturas que não foram simuladas. Maiores informações à respeito do método podem ser encontrados no Apêndice A.

É por meio do perfil de energia livre obtido pelo *WHAM* que somos capazes de indentificar, em função de uma coordenada de reação (normalmente a fração de contatos nativos Q), a localização das regiões de transição no enovelamento, que será utilizado para descrição do comportamento das proteínas nas demais análises deste trabalho, como por exemplo, na verificação de quais contatos são importantes para a proteína conseguir vencer a barreira de energia livre e chegar no estado nativo.

4.2 Formação de contatos entre resíduos de aminoácidos

Para determinar a quantidade de contatos entre resíduos de aminoácidos que estão presentes na estrutura nativa, recorre-se a algoritmos que conseguem calcular por meio da distância, posição e orientação, quais resíduos de aminoácidos estão interagindo de acordo com os dados estruturais da proteína nativa extraídos do *Protein Data Bank* (PDB). Neste trabalho, utilizamos o algoritmo do *Shadow Map* (72) para cálculo do mapa de contato. Este algoritmo está disponível na ferramenta online *SMOG* (71).

Não somente a fração de contatos nativos total pode ser acompanhada ao longo do processo de enovelamento, mas também contatos específicos entre resíduos de aminoácidos que estão localizados em regiões de interesse, como por exemplo no sítio de ligação com os metais no caso da SOD1. Além de utilizar o mapa de contato nativo oriundo do *Shadow Map* como fonte para o cálculo de Q , dizemos que durante a simulação um contato nativo está formado se a distância r entre os dois aminoácidos em questão, é menor do que $1.2 r_0$, sendo r_0 a distância entre os mesmos resíduos de aminoácidos na estrutura nativa.

A definição de contato não-nativo ainda é algo que não segue um padrão na literatura. Neste trabalho, seguindo a linha de Mouro et al, definiremos contato não-nativo como sendo qualquer contato entre resíduos de aminoácidos separados por pelo menos 4 resíduos, que estão a menos de $6\text{\AA}$ de distância. É fundamental que estes contatos chamado de não-nativos não estejam presentes no mapa de contato nativo obtido pelo *Shadow Map*.

4.3 Tempo de enovelamento

Valores de taxa, ou tempo de enovelamento/desenovelamento, obtidos de forma computacional, podem nos dizer muito à respeito da cinética e da estabilidade do enovelamento das proteínas (57). Este tempo é calculado utilizando a média dos tempos de várias simulações realizadas em uma mesma temperatura. Para calcular o tempo de enovelamento (ou desenovelamento) em cada simulação, utiliza-se o método de primeira passagem, partindo de uma estrutura desenovelada aleatória (ou nativa) e medindo o tempo de simulação necessário para que a proteína atinja pela primeira vez uma porcentagem significativa de contatos nativos formados (ou desfeitos). Neste trabalho, definimos que esta porcentagem deveria ser de 0.80 do valor de Q para o estado nativo, e de 0.15 para o estado desenovelado. Tais valores foram escolhidos de acordo com as posições dos mínimos do perfil de energia livre da SOD1. Foram realizadas 50

simulações de tempo de desenovelamento para cada temperatura, tanto para o dímero da SOD1 quanto para o monômero.

4.4 Rota de Enovelamento

A medida da rota de enovelamento de uma proteína é uma ferramenta que nos permite investigar a especificidade dos contatos nativos realizados pela proteína ao se enovelar. O cálculo da rota de enovelamento pode nos dar uma idéia de quão rugoso é a superfície (ou funil) de energia de cada proteína (57, 77). A equação que define o cálculo da rota de enovelamento $R(Q)$ é dada pela seguinte expressão:

$$R(Q) = \sum_{i=1}^M \frac{\left\langle \left(\langle Q_i \rangle_Q - Q \right)^2 \right\rangle_Q}{MQ(1 - Q)} \quad (4.1)$$

M representa o número de contatos nativos, e Q_i , com i variando de 1 até M , são os próprios contatos nativos, de tal forma que $\langle Q_i \rangle_Q$ é a probabilidade do contato Q_i se formar em um determinado valor de Q . Por definição, o valor de $R(Q)$ varia entre 0 e 1, e quanto mais próximo de 1, mais específico é o caminho que a proteína faz para se enovelar, de tal forma que, no valor $R(Q)=1$ a proteína só teria um caminho possível, indicando uma superfície de energia bastante rugosa. Por outro lado, um valor de $R(Q)$ próximo de zero significa que qualquer contato nativo possui a mesma probabilidade de ser feito, ou seja, neste caso, a proteína não teria rotas preferenciais. Proteínas que apresentam valores elevados de $R(Q)$ normalmente são mais lentas para se enovelar (57).

Capítulo 5

Resultados e Discussão

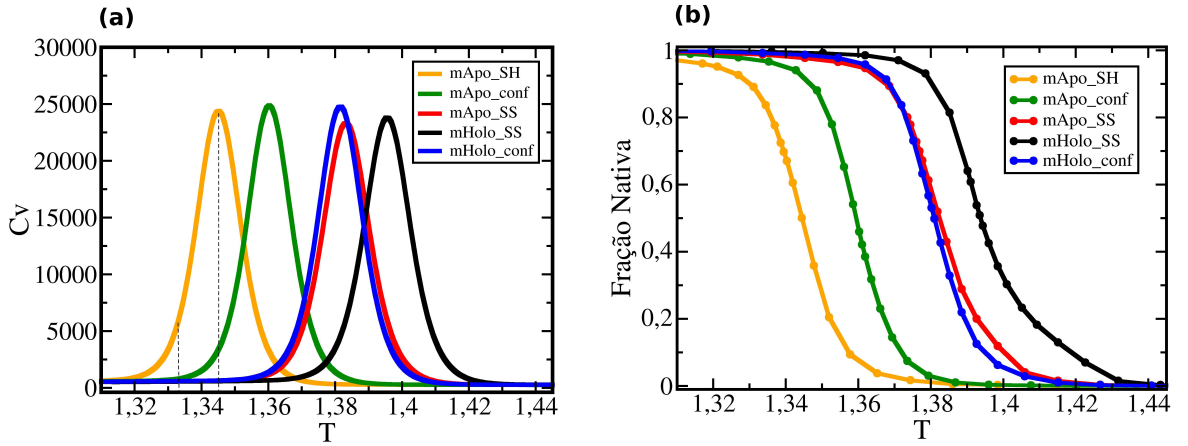
5.1 Enovelamento do Monômero utilizando SBM- C_α e REMD

Esta seção descreve os aspectos termodinâmicos e cinéticos do enovelamento do monômero, cuja estrutura foi retirada da enzima dimérica, identificada pelo código PDB “1HL5”, assim como descrito nos detalhes computacionais. Estima-se uma maior credibilidade aos resultados teórico-computacionais quando estes são validados por dados experimentais, no entanto, para compará-los, é necessário encontrar uma correspondência entre unidades reais e unidades reduzidas. Utilizando o critério das razões entre valores de C_p (descrito no capítulo de Metodologia deste trabalho), identificamos temperaturas em unidades reduzidas que podem ser correspondentes com a temperatura ambiente real, uma vez que há, na literatura, dados experimentais relevantes para três das cinco formas monoméricas estudadas neste trabalho (mApo_{SH}, mApo_{SS} e mHolo_{SS}) e para uma forma dimérica, dHolo_{SS}.

A figura 5.1 (a) nos mostra as curvas de calor específico para os cinco sistemas monoméricos modelados nesse trabalho. A localização dos picos de cada curva nos dá os valores das temperaturas de transição, as quais também correspondem, aproximadamente, com o meio da transição nas curvas de fração enoveladas da Figura 5.1 (b). Deste modo, as temperaturas dos picos de C_p (ou do meio de transição da fração nativa) são as temperaturas nas quais a energia livre, em função da coordenada de reação, indica o zero de estabilidade. Comparando as diferentes temperaturas de transição de cada um dos sistemas monoméricos estudado, é possível notar que a presença da ponte dissulfeto entre as cisteínas 57 e 146 aumenta o valor da temperatura de transição, indicando um aumento na estabilidade térmica. O mesmo acontece quando comparamos

o caso mApo_{SS} com o mHolo_{SS}, indicando que a “presença” dos metais também leva a um aumento na estabilidade termodinâmica. O caso apo reduzido (mApo_{SH}) demonstrou ser a forma menos estável, com temperatura de enovelamento em $T=1.342$ em unidades reduzidas, 2.7% menor do que o caso apo oxidado (mApo_{SS}), que apresentou um $T_f = 1.379$. Os valores determinados experimentalmente para estes dois sistemas, 315K e 325K, respectivamente, indicam uma diferença de 3%, o que corrobora os nossos dados. Acredita-se que a ponte dissulfeto conformacional seja o modelo mais próximo do que acontece *in vivo*. Nos nossos sistemas, tanto o mApo_{conf} quanto o mHolo_{conf} mostraram estabilidade intermediária, superior ao caso reduzido, mas inferior ao caso oxidado, porém, a falta de experimentos *in vivo* dificulta a confirmação de tais dados.

Figura 5.1: Calor específico e fração nativa em função da temperatura. (a) Cada uma das cinco curvas mostram o valor do calor específico C_v de uma simulação do monômero da SOD1 em função da temperatura. As curvas foram obtidas pela aplicação do WHAM nos dados oriundos do REMD com SBM- $C\alpha$. Em amarelo temos o sistema mApo_{SH}, em verde o mApo_{conf}, em vermelho o mApo_{SS}, em preto o mHolo_{SS}, e em azul o mHolo_{conf}. As linhas verticais tracejadas na curva do sistema mApo_{SH} representam a forma como foi calculada o valor da temperatura ambiente em unidades reduzidas. (b) A fração nativa de cada sistema foi calculado nas 24 réplicas de simulação. O valor de 1 diz que 100% das conformações amostradas estão dentro do ensemble enovelado. O ponto intermediário da transição (Fração Nativa = 0.5) indica a temperatura na qual os ensembles enovelado e desenovelado são igualmente amostrados. As cores seguem o mesmo padrão do item (a)



Fonte: Elaborada pelo autor

Uma forma interessante de checar quão perspicaz é o procedimento das razões entre valores de C_p , consiste em verificar se os dados da termodinâmica (C_p) se correlacionam com informações de cinética, relativas a barreiras de energia livre. Há muitos valores experimentais para as taxas de enovelamento e desenovelamento dos monômeros apo (reduzido e oxidado) e holo oxidado da SOD1, dos quais é possível extrair valores de barreira e estabilidade utilizando a teoria do estado de transição.

Os dados expostos por Kayatekin et al (78) mostram que, em 298K, a barreira para desenovelar o monômero apo oxidado é $\Delta G_u = 18 \text{ kcal/mol}$, enquanto a sua estabilidade ($\Delta G_0 = \Delta G_{(fold)} - \Delta G_{unf}$) é de $\Delta G_0 = 4.9 \text{ kcal/mol}$. A razão entre esses dois valores nos dá o número 3.7. Esta razão também pode ser facilmente obtida dos dados da simulação em temperatura reduzida. Curiosamente, o valor de 3.7 para o sistema mApo_{SS} ocorreu apenas uma vez, em uma temperatura bem próxima da encontrada pelo critério da razão entre o calor específico, sugerindo uma coerência razoável entre os dados da termodinâmica e de cinética. A Tabela 5.1 traz os valores de T_0 encontrado para cada sistema:

Tabela 5.1: Temperatura ambiente em unidades reduzidas para cada sistema

Sistema	mApo _{SH}	mApo _{SS}	mHolo _{SS}	dHolo _{SS}
$C_p(T_m)/C_p(T_0)$	1.336 ^a	1.368 ^{a,c}	1.368 ^e	1.360 ^f
$\Delta G_u/\Delta G_0$	1.333 ^b	1.366 ^d	1.364 ^b	— ^g

^a Dados extraídos de Vassal et al (79)

^b Dados extraídos de Kayatekin et al (80)

^c Dados extraídos de Broom et al (81)

^d Dados extraídos de Kayatekin et al (78)

^e Dados extraídos de Battistoni et al (82)

^f Dados extraídos de Stathopoulos et al (83)

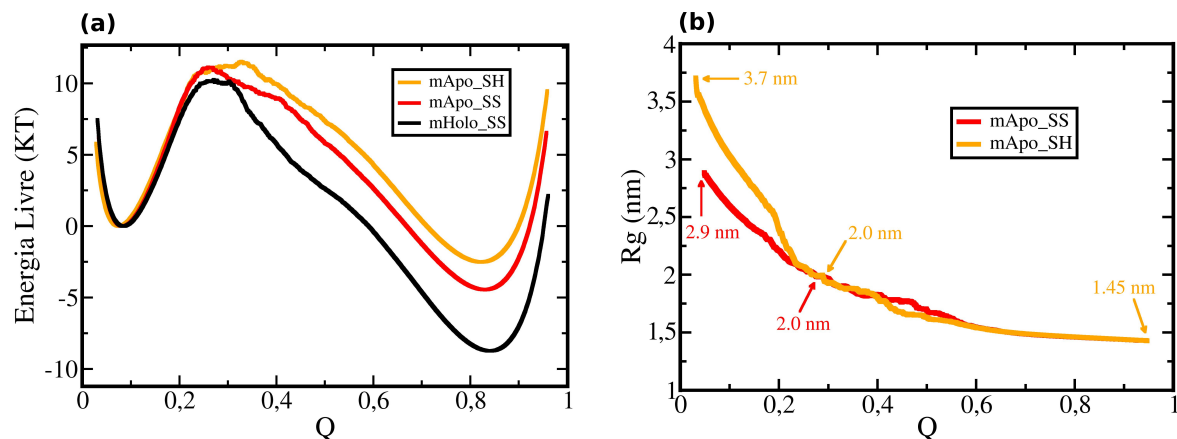
^g A presença de estados intermediários torna difícil usar a razão entre as barreiras e a estabilidade

Projetando os valores de T_0 nos dados de fração nativa em função da temperatura (Figura 5.1(b)), percebe-se que em todos os casos, a fração enovelada do sistema apresenta valores acima de 90% (e 98% para o caso mHolo_{SS}), o que correlaciona bem com uma grande quantidade de estudos experimentais (79-81), validando a nossa forma de encontrar T_0 .

Adotando o critério das razões entre C_p para determinarmos os valores da temperatura reduzida que corresponde a temperatura ambiente, somos capazes de comparar dados computacionais de energia livre, com dados obtidos experimentalmente. A Figura 5.2(a) mostra o perfil de enovelamento na temperatura ambiente da simulação para os três casos os quais podemos comparar com dados experimentais: mApo_{SH}, mApo_{SS} e mHolo_{SS}. Todos os perfis de energia livre mostram um típico processo de enovelamento de 2-estados, sem a presença de intermediários metaestáveis. O pico da barreira de energia livre mostrou estar localizado em pequenos valores de Q ($Q \approx 0.25$ para os caso mApo_{SS} e mHolo_{SS}, e $Q \approx 0.33$ para o caso mApo_{SH}), indicando uma

energia, ou custo entálpico, relativamente alto no estado de transição, já que somente 25% dos contatos nativos (capazes de abaixar a entalpia) estão feitos.

Figura 5.2: Perfis de Energia Livre e Raio de Giro em função da fração de contatos nativos (Q). (a) Cada uma das três curvas representam o perfil de energia livre de uma simulação do monômero da SOD1: mApo_{SH} em amarelo, mApo_{SS} em vermelho, e mHolo_{SS} em preto. Todas as três curvas foram calculadas na temperatura ambiente, de acordo com o critério de C_p . (b) Raio de Giro, Rg, em função da fração de contatos nativos Q, para o caso mApo_{SH} (em amarelo) e mApo_{SS} (em vermelho). As flechas indicam as posições e o valor de Rg nos ensembles: desenovelado, no estado de transição (2.0 nm) e na conformação nativa (1.45 nm).



Fonte: Elaborada pelo autor

A predominância da entalpia com relação a entropia na formação da barreira de energia livre da SOD foi observado por Kayatekin et al (78), utilizando dados experimentais. Os autores mostraram a presença de uma entropia de ativação favorável, devido as interações com o solvente durante o colapso do estado de transição, mas o alto custo entálpico foi capaz de superar essa mudança na entropia. Neste caso, diz-se que a barreira tem uma característica entálpica. Todavia, nas nossas simulações, é de se esperar que a entropia de ativação seja desfavorável, uma vez que somente a entropia conformacional está envolvida, a qual é obviamente reduzida no estado de transição ($\Delta S < 0$, comparada ao estado desenovelado). Neste caso, para nós, a barreira de energia livre possui uma característica entrópica, porém, de uma forma geral, quem vai determinar o quão alta ela é, é a entalpia. No caso da SOD, a barreira relativamente alta no estado de transição em temperatura ambiente (em torno de 11kT), é devido ao pequeno decaimento da entalpia, oriunda do baixo número de contatos nativos na região de transição.

Comparando os dois perfis de energia livre dos monômeros apo na Figura 5.2 (a), podemos ver que o monomero apo reduzido, mApo_{SH}, tem uma estabilidade de

1.5 kT menor do que o caso oxidado, $mApo_{SS}$. Além disso, o caso reduzido apresentou uma barreira de enovelamento levemente maior, cerca de 0.4 kT, comparado com o caso oxidado $mApo_{SS}$. Neste caso, temos que a ausência dos metais e da ponte dissulfeto, torna o enovelamento mais difícil e deixa a proteína menos estável. Experimentalmente, o caso apo oxidado se enovela de 2-3 vezes mais rápido do que o caso reduzido (78), o que é consistente com uma diferença pequena de barreira, em torno de 0.4-0.5 kcal/mol. O aumento de barreira de 0.4 kT, pode ser inferido devido ao aumento no custo entrópico para o caso $mApo_{SH}$ se enovelar, uma vez que a ausência da ponte dissulfeto torna a entropia do estado desenovelado maior. Neste caso, a posição do pico da barreira de transição do caso $mApo_{SH}$ em um Q ($Q \approx 0.33$) maior do que do caso oxidado, pode indicar uma tentativa da proteína de pagar este alto custo entrópico, formando mais contatos nativos. Sendo assim, nem a formação de mais contatos nativos foi suficiente para pagar o custo entrópico da falta da ponte dissulfeto no estado desenovelado, fazendo com que a barreira de enovelamento seja levemente maior.

O caso $mHolo_{SS}$ foi o que mostrou maior estabilidade, menor barreira de enovelamento, e maior barreira de desenovelamento, o que mostra a importância do sítio metálico para o processo de enovelamento, e torna nossos dados compatíveis tanto qualitativamente quanto quantitativamente com dados experimentais de monômeros da SOD na presença de zinco (84, 85). Estudos mostram que a forma oxidada e ligada ao zinco estabiliza a estrutura nativa de 7-9 kcal/mol com relação ao caso apo oxidado (80), o que não é muito diferente do valor de 4.1kT obtidos dos dados da simulação (ao comparar os mínimos nativos da curva preta com a vermelha). Podemos inferir que a diferença de estabilidade do estado nativo entre cada um dos três sistemas é de origem puramente entálpica, uma vez que estamos utilizando a mesma estrutura proteica, mas com diferentes energias de contatos, dada as deleções que foram feitas em cada um dos mapas de contatos nativos.

A Figura 5.2(b) mostra o comportamento do raio de giro R_g em função da fração de contatos nativos Q para o caso $mApo_{SH}$ (em amarelo) e $mApo_{SS}$ (em vermelho). Pode-se notar que o valor de R_g no estado de transição vale ≈ 2.0 nm para ambos. O estado desenovelado também se mostrou igual, independente da oxidação da ponte, com $R_g = 1.45$ nm. Este último valor está bem próximo do raio de giro da estrutura cristalográfica, o que é uma consequência de se usar SBMs. Já os estados desenovelados apresentaram R_g iguais a 3.7nm e 2.9nm, para o caso reduzido e oxidado, respectivamente. Utilizando os valores de R_g obtidos para o caso reduzido, e aplicando-os na equação 3.4, encontramos o valor de β -Tanford teórico, o que nos dá um valor de 0.71. Dados experimentais (86), mostram que para o monomero apo e reduzido, os valores-m são: $m_{D-TS} = 1.05$, enquanto $m_{D-N} = 1.53$. Ao confrontarmos esses dados com

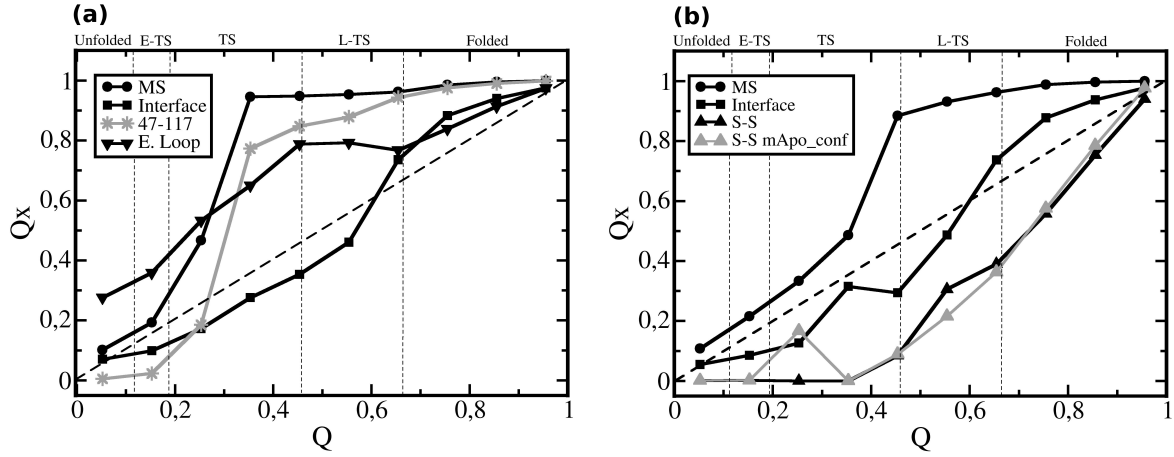
a equação 3.5, encontramos um valor de β -Tanford experimental igual a 0,69, o que é muito próximo do valor teórico. Já para o monomero apo e oxidado, utilizando os valores de R_g , podemos encontrar o valor de $m_{TS} = 0.65$, o que é bastante razoável quando comparado com o valor de 0.62 determinado por Kayatekin et al (78). Deste modo, tais resultados indicam que nosso modelo apo está sendo capaz de amostrar ensembles realísticos.

Já é de conhecimento que na ausência dos metais, além da SOD1 perder sua função enzimática mediada pelo íon de Cu, ela também tem sua estabilidade reduzida, assim como mostrado nos nossos resultados. Porém, ainda há na literatura discussões a respeito do momento da inserção dos metais na SOD1: se há a presença de chaperonas para carrear esta ligação, e, se esta ligação ocorre durante o enovelamento, em qual dos estágios ela ocorre. Tem sido proposto que, *in vivo*, a Sod1 precisa de assistência da chaperona de cobre Ccs1, que se ligando a uma forma imatura de Sod1, apo e com a ponte dissulfeto de forma reduzida, induz mudanças estruturais, aumentando a afinidade pelo zinco (87). Então, mais tarde, perto da estrutura nativa, a Ccs1 auxilia a finalização do enovelamento, ligando o cobre e oxidando a ligação dissulfeto intramolecular, além de juntar os dois monômeros para a formação do dímero.

Vimos que dificultar a formação do sítio de ligação com os metais afetou diretamente a barreira de energia livre, sendo assim, é de se esperar que provavelmente a ligação com os metais aconteça já nos primeiros estágios do enovelamento, próximo ao estado de transição. A Figura 5.3 nos informa o valor médio de formação dos contatos nativos das regiões de ligação com os metais (MS), da ponte dissulfeto entre as cisteínas 57-146 (S-S), da região de interface, do loop eletrostático, e do contato entre os resíduos 47-117, para os sistemas mHolo_{SS} em (a) e mHolo_{conf} em (b), em suas respectivas temperaturas de enovelamento. Estes resultados mostram que a formação do sítio de ligação com os metais ocorre logo nos primeiros estágios do enovelamento, iniciando antes mesmo do estado de transição, independentemente da oxidação da ponte dissulfeto. Nosso modelo é demasiadamente simples para distinguir entre a vizinhança do cobre e do zinco, e essa é a razão de mencionarmos de forma geral o sítio metálico (Zn+Cu).

Experimentalmente, os resíduos 47 e 117 mostraram ter um valor- ϕ relativamente alto (37, 38), e de acordo com o algoritmo do *shadow map*, eles estão próximos na estrutura nativa e formam um contato. O contato entre estes dois resíduos (representado por * cinza), assim como o sítio metálico, também apresenta um crescimento precoce na sua frequência de formação, o que é compatível com um elevado valor- ϕ para ambos resíduos. Os resíduos do monômero que fazem parte da interface estão representados por quadrados na Figura 5.3. Tanto em (a) quanto em (b), vemos que a

Figura 5.3: **Valor médio de formação por região nativa Q_x** (a) Para o caso mHolo_{SS} e (b) para o caso mHolo_{conf}. Cada uma das curvas representam uma região específica do monômero: o sítio metálico (MS, círculos pretos), a ponte dissulfeto entre as cisteínas 57 e 146 (S-S, triângulo para cima), a região da interface (quadrados pretos), o contato entre os resíduos 47 e 117 (* em cinza), e a região do loop eletrostático (E. Loop, triângulo para baixo, em preto). Cada uma das curvas foram calculadas utilizando os dados das simulações nas respectivas temperaturas de enovelamento de cada sistema. As siglas E-TS, TS e L-TS remetem, respectivamente, aos estágios de pré-transição, transição e pós-transição. A linha diagonal tracejada indica a formação dos contatos totais da proteína.



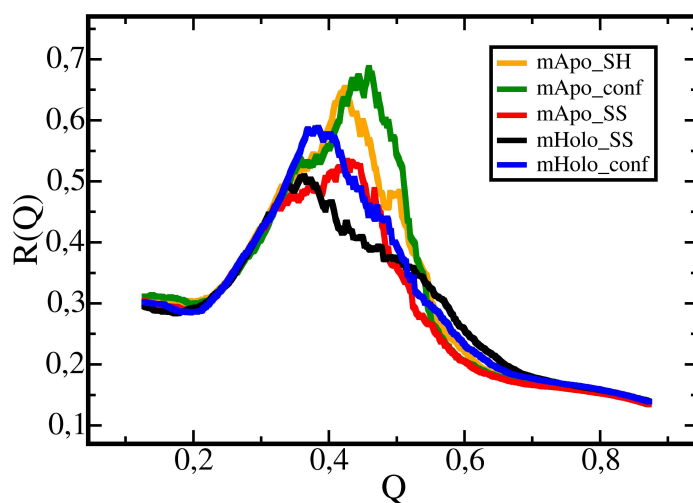
Fonte: Elaborada pelo autor

formação dessa região é gradual, e acompanha o enovelamento da cadeia, ao contrário da ponte dissulfeto nos casos conformacionais, mHolo_{conf} (triângulo preto) e mApo_{conf} (triângulo cinza) na Figura 5.3(b), a qual é formada tardiamente, tendo seu crescimento após o estado L-TS. Já a região que envolve o loop eletrostático (triângulo para baixo em (a)), se mostrou ser uma das primeiras regiões da proteína a começar a se formar, estando sempre a frente do “enovelamento da cadeia” (linha tracejada), porém, sua estruturação parece ocorrer de forma constante, estando totalmente formada apenas no final do processo.

Uma outra forma de verificar o efeito que as modificações no mapa de contato podem ocasionar no enovelamento da SOD1, é por meio do cálculo da medida da Rota de Enovelamento $R(Q)$ (57), como pode ser visto na Figura 5.4. A medida de $R(Q)$ mostra quão específico é o caminho seguido pela proteína para esta se enovelar. Neste sentido, para o caso do monômero da SOD1, é possível ver que a região que engloba o estado de transição (para os cinco casos simulados) possui valores elevados de $R(Q)$, sugerindo a formação de contatos específicos para que a proteína consiga vencer a barreira de energia livre e atingir o estado nativo. Ao comparar os nossos resultados com os obtidos por Clementi et al (57), vemos que o perfil de $R(Q)$ da

SOD1 se assemelha ao perfil da SH3, a qual também possui um aumento no valor de $R(Q)$ no estado de transição. Analogamente, a proteína SH3 também é formada em sua estrutura terciária por um “barril- β ”, o que pode evidenciar um comportamento típico de proteínas com este *motif*.

Figura 5.4: **Medida da Rota de Enovelamento $R(Q)$ para os cinco casos de monômeros da SOD1 simulados** Normalizados entre 0 e 1, valores elevados de $R(Q)$ indicam que a proteína segue um caminho bastante específico para se enovelar, formando praticamente sempre os mesmos contatos nativos em determinados estágios do enovelamento. As curvas em amarelo, verde, vermelho, preto e azul, remetem, respectivamente, às simulações do monômero $mApo_{SH}$, $mApo_{conf}$, $mApo_{SS}$, $mHolo_{SS}$ e $mHolo_{conf}$. Cada uma das curvas foram obtidas nas suas respectivas temperaturas de enovelamento.



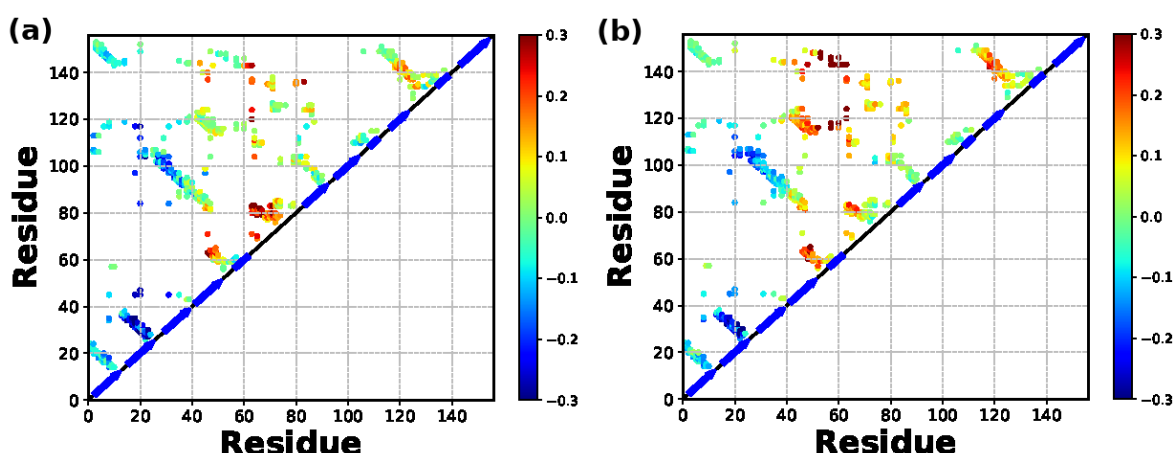
Fonte: Elaborada pelo autor

O elevado valor de $R(Q)$ na simulação da proteína Apo reduzida ($mApo_{SH}$) e na conformacional ($mApo_{conf}$), nos mostra que nestes casos, a rota de enovelamento se tornou ainda mais específica nos estados de transição. Valores elevados de $R(Q)$ podem indicar uma superfície de energia livre bastante rugosa, e a procura por um caminho para se enovelar pode acabar retardando o enovelamento. Na maioria das vezes, valores elevados de $R(Q)$ no estado de transição indicam proteínas com baixos valores de taxa de enovelamento (57), e é isso que acontece na ausência da ponte dissulfeto para o caso apo. Esta “demora”, associada com a presença de mínimos locais na superfície de energia e pela alta especificidade, pode acarretar na formação de estruturas patogênicas, como agregados amilóides. Ao forçarmos a oxidação da ponte dissulfeto desde o início do enovelamento, tanto para o caso $mHolo_{SS}$ quanto para o $mApo_{SS}$, vemos que esta causa uma diminuição na especificidade de contatos, sendo que o caso holo foi o que mostrou o menor valor de $R(Q)$, sugerindo que a “presença” dos metais pode tornar o estado de transição ainda menos específico, o que é compatível

com o aumento na taxa de enovelamento.

A medida de $R(Q)$ nos informa a respeito da especificidade dos caminhos, mas não nos conta a respeito dos contatos que os resíduos de aminoácidos estão formando. A Figura 5.5 mostra a diferença de probabilidade de formação de contatos nativos no estado de transição em (a) entre $mHolo_{conf}$ e $mApo_{conf}$ e em (b) entre $mApo_{SS}$ e $mApo_{SH}$. A figura 5.5 (a) tem como objetivo mostrar a perturbação causada pelo sítio metálico, e como esperado, ela mostra que sem alterar a ponte dissulfeto, deixando-a livre para se formar ao longo do enovelamento, a região próxima do loop (região entre os resíduos 60 e 80) que se liga ao zinco possui cerca de 30% mais probabilidade de formação no caso holo do que no apo. No entanto, a Figura 5.5 (a) também nos mostra que além de afetar a formação do sítio metálico, a estrutura Apo mostrou ter uma probabilidade maior em formar outras estruturas secundárias do que o caso holo. De forma mais específica, os resíduos de aminoácidos que aparecem formando contatos com maior probabilidade no TS da estrutura $mApo_{conf}$ estão nas folhas- β 2, 3 e 6. Neste caso, as diferenças no comportamento de $R(Q)$ não só pode estar relacionado a um crescimento da especificidade, mas também devido a uma alteração no caminho que a proteína segue para se enovelar.

Figura 5.5: **Diferença na probabilidade de formação de contatos nativos** (a) Entre os casos $mHolo_{conf}$ e $mApo_{conf}$ e (b) entre os casos $mApo_{SS}$ e $mApo_{SH}$. Todas as probabilidades foram calculadas no estado de transição. Os valores de probabilidade foram obtidos nas respectivas temperaturas de enovelamento de cada sistema. Valores positivos nas probabilidades, dados pela barra de cor à direita da figura, indicam que houve uma propensão maior do contato ser formado no caso $mHolo_{conf}$ em (a) e $mApo_{SS}$ em (b). Os eixos do gráfico remetem aos resíduos de aminoácidos do monômero da SOD1.



Fonte: Elaborada pelo autor

A Figura 5.5 (b) reflete a perturbação causada pela ponte dissulfeto 57-146 no

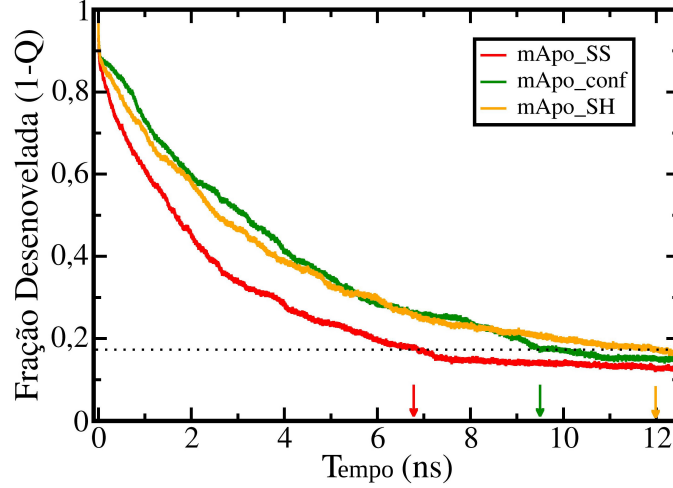
enovelamento do monomero apo. Ela mostra a diferença que a formação precoce da ponte dissulfeto pode causar no enovelamento, quando comparada com o caso no qual ela não se forma. Na figura (b), vemos que o resultado da (a) é mantido, no sentido que, agora, a forma apo reduzida forma com mais frequência os contatos entre as folhas- β : 2, 3 e 6, do que a forma apo oxidada. Porém, a oxidação da ponte dissulfeto mostrou promover um aumento na ligação entre o sítio metálico e o loop eletrostático, apresentando um alosterismo entre essas regiões no estado de transição. Recentemente, simulando um experimento de espectroscopia de força em um monômero apo da SOD, e utilizando um modelo com todos-os-átomos e solvente explícito, Habib et al (88) foram capazes de ver que as últimas folhas- β a se desenovelarem são exatamente as folhas 2, 3 e 6. Ivanova et al (89) mostraram por meio de estudo envolvendo estatística de sequencias com propensão a formar agregados de fibras amilóides, que algumas das folhas- β que mostraram se formar com maior probabilidade na nossa simulação do caso apo, e especialmente do caso apo reduzido, são mais propícias de formar estruturas com características patogênicas. Deste modo, nossos resultados corroboram com a ideia de que a ausência dos metais, acrescido da redução da ponte dissulfeto, pode favorecer o processo de “misfolding”.

5.1.1 Simulação de Controle - Cinética de Enovelamento

Até agora, todos os dados de simulação se referem a dados termodinâmicos e feitos com o REMD. Nesta seção, é descrita a cinética de enovelamento relativa ao parâmetro de ordem, utilizando a média de 200 simulações iniciadas com estruturas desenoveladas e aleatórias, em temperatura única e baixa. A Figura 5.6 mostra o valor médio da fração desenovelada (1-Q) para os três caso apo: mApo_{SS}, mApo_{SH} e mApo_{conf}. A cadeia é considerada enovelada quando ela atinge a fração desenovelada (1-Q) de 0.17, equivalente ao Q=0.83, que corresponde ao mínimo enovelado da Figura 5.2(a). Cabe ressaltar que os valores de tempo da simulação, em nanosegundos, não correspondem ao tempos reais de enovelamento, os quais são da ordem de segundos para o caso da SOD1. Essa é uma das razões na qual usaremos somente comparação relativa entre um caso e outro.

A fração desenovelada atinge o valor de 0.17 em um tempo de 12.0 ns para o caso mApo_{SH}, e 6.7 ns para o caso na qual a ponte dissulfeto está sempre formada, mApo_{SS}, indicando ter um enovelamento de 1.8 vezes mais rápido que o caso reduzido. Essa variação é consistente com o resultado apresentado na Figura 5.2 (a), a qual mostrou a presença de uma barreira de enovelamento 0.4 kT maior para o caso reduzido. Além disso, o fator de 1.8 é consistente com os resultados de cinética obtidos

Figura 5.6: **Cinética de enovelamento do monômero apo.** Fração desenovelada ($1-Q$) em função do tempo de simulação. A linha horizontal tracejada foi fixada no valor da fração desenovelada referente ao mínimo nativo ($Q \approx 0.83$) obtido pela curva de energia livre (Figura 5.2). As flechas indicam, de acordo com a cor, o tempo no qual cada sistema atinge a linha tracejada. Em vermelho temos a $mApo_{SS}$, em verde a $mApo_{conf}$ e em amarelo a $mApo_{SH}$



Fonte: Elaborada pelo autor

experimentalmente, os quais mostram que, com a ponte dissulfeto intacta, a taxa de enovelamento do monômero apo aumenta 2 vezes. Comparando com o caso conformacional $mApo_{conf}$, em verde, vemos que este novamente se comporta como um caso intermediário, tendo um enovelamento 1.4 vezes mais lento que o caso oxidado e 1.26 mais rápido que o caso reduzido. Estes resultados de certa forma corroboram os obtidos pelo método de troca de réplicas.

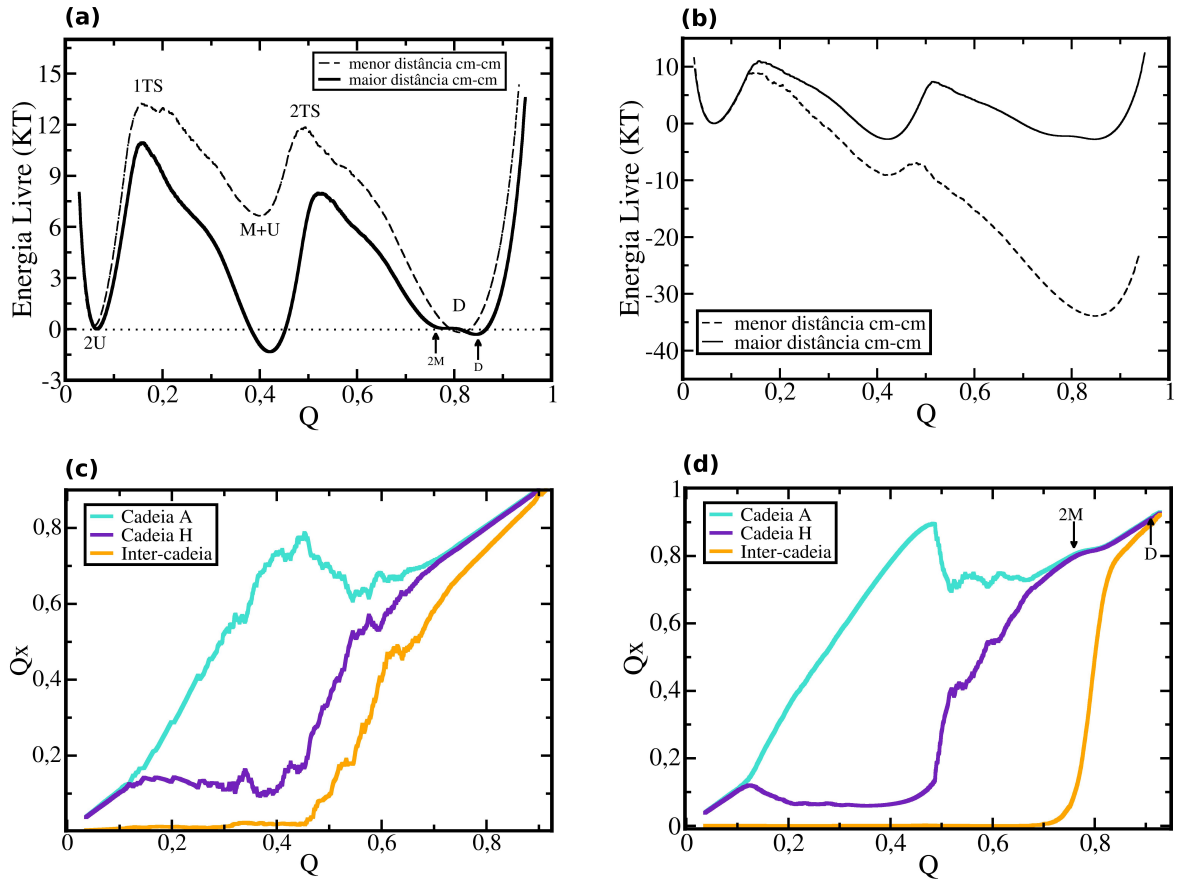
5.2 Enovelamento do Dímero

5.2.1 Utilizando SBM- C_α e REMD

Assim como feito para o caso com apenas um monômero, realizamos simulações com dois monômeros permitindo a interação entre eles. Como explicado na seção de *Detalhes Computacionais*, utilizamos um potencial de restrição para mantermos as cadeias próximas durante o enovelamento. Variando a distância de máximo afastamento entre as duas cadeias, tentamos realizar simulações que se assemelhariam a um experimento com alta e baixa concentração de SOD1, baseado no princípio de que em um meio mais concentrado, as moléculas estariam mais próximas uma das outras. O comportamento termodinâmico do enovelamento da SOD1 nestes dois tipos de simulações

pode ser visto nos perfis de energia livre da Figura 5.7.

Figura 5.7: **Perfil de energia livre para o dímero da SOD1 e formação de cada monômeros.** (a) Os dados utilizados para elaboração desta figura foram obtidos com o SBM-C α e o método WHAM, utilizando o pdb 1HL5 da SOD1 (cadeias A e H). A linha preta tracejada indica o resultado da simulação com maior concentração de SOD1 (menor distância cm-cm), e em linha preta contínua, a simulação com menor concentração de SOD1 (maior distância cm-cm). Nesta, observa-se a formação de quatro estados: 2U (duas cadeias desenoveladas), M+U (uma cadeia enovelada e a outra não), 2M (dois monômeros enovelados) e D (dímero), separados por dois estados de transição (1TS e 2TS). O perfil de energia livre foi extraído da temperatura de enovelamento de cada caso. (b) Perfil de energia livre na temperatura correspondente à temperatura ambiente. O padrão das curvas seguem o mesmo do item (a). A fração de enovelamento da cadeia A (em ciano), cadeia H (em roxo) e dos contatos entre resíduos da região inter-cadeia (em amarelo) em função da fração de contatos nativos do dímero (Q) são mostradas em (c) para o caso de maior concentração de SOD1, e em (d) para o caso de menor concentração de SOD1. As flechas em (d) indicam a formação de dois monômeros (2M) e do dímero (D), referente ao perfil de energia livre da Figura 5.7 (a)



Fonte: Elaborada pelo autor

Utilizando a fração de contatos nativos como coordenada de reação, vemos que para o caso em que limitamos a menor distância entre os centros de massa (cm-cm) dos

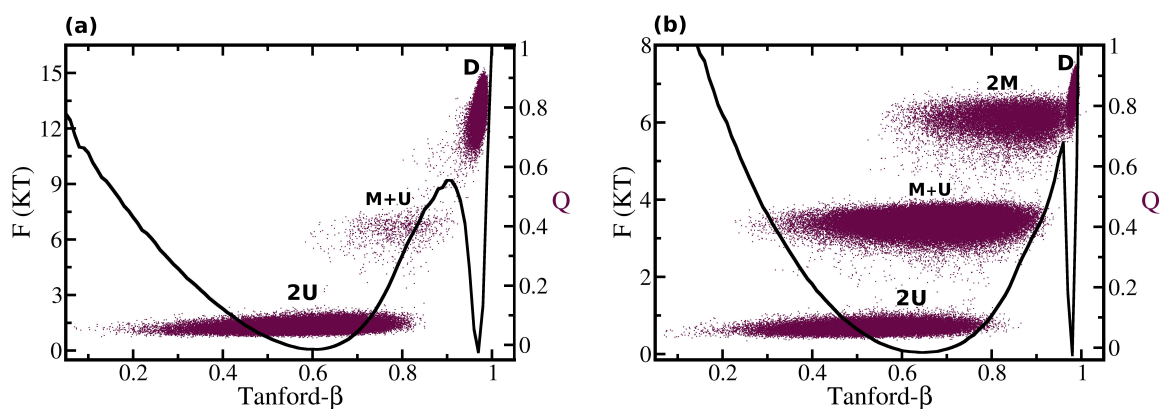
monômeros, o perfil de energia livre em T_f (Figura 5.7(a)) apresenta um processo de enovelamento sequencial em 3-estados ($2U \leftrightarrow M+U \leftrightarrow D$), sendo $2U$ composto por dois monômeros desenovelados, e D o dímero. Por outro lado, um processo de 4-estados ($2U \leftrightarrow M+U \leftrightarrow 2M \leftrightarrow D$) pode ser notado quando damos mais liberdade para os monômeros se afastarem, apresentando neste caso, um estado $2M$ pouco estável (dois monômeros formados sem a presença da interface), o qual não apareceu na simulação em que restringimos o afastamento das cadeias. O estado intermediário designado como $M+U$ foi comum nas duas simulações, e consiste em um monômero já praticamente enovelado, enquanto o outro monômero ainda está desenovelado. Neste sentido, podemos ver que no caso de “menor concentração”, o estado $M+U$ se mostra mais estável, e o desenovelamento do dímero se torna mais fácil, já que a barreira é menor. Além disso, a Figura 5.7(b) nos mostra que o dímero em um meio mais concentrado é mais estável em temperatura “ambiente” do que o dímero em menor concentração. Estes resultados apontam para a existência de uma relação entre o enovelamento e a concentração da SOD1.

A presença de um estado intermediário do tipo $M+U$ pode ser vista ao acompanharmos o enovelamento de cada monômero, como mostrado na Figura 5.7(c) e (d). De acordo com este resultado, o segundo monômero (cadeia H) parece se enovelar após a formação de 80% do primeiro monômero (cadeia A), independente da concentração. Contudo, a formação dos contatos entre resíduos da interface (inter-cadeia), no caso de maior concentração de SOD1, se inicia no mesmo ponto em que o enovelamento do segundo monômero têm o seu crescimento. Além disso, a formação da interface parece acompanhar a mesma taxa de crescimento da segunda cadeia. A formação da interface de forma concomitante com o enovelamento do segundo monômero, implica na não formação do estado $2M$, no caso de maior concentração. Além disso, aparentemente, formar a interface junto com o enovelamento do segundo monômero, parece diminuir a barreira de energia livre para este monômero enovelar, como pode ser visto, ao comparar o estado 2TS da figura 5.7(a) entre os dois casos. Esta diminuição da barreira pode estar relacionada com uma menor exposição da interface, uma vez que esta é composta em sua essência por aminoácidos hidrofóbicos. Neste caso, a formação precoce dessa região não só deve ser capaz de diminuir a barreira de energia livre, como também pode impedir a formação de agregados, uma vez que se diminui de forma nativa a área exposta ao solvente.

Levy et al (90), também realizando simulações computacionais, mostraram que o enovelamento do dímero da SOD é composto pela formação de 4-estados, sendo eles, exatamente os mesmos presentes na curva de energia livre para o caso de menor concentração de proteína. Contudo, resultados experimentais (79, 91), sugerem que

em alta concentração de SOD1, o enovelamento desta enzima seja caracterizado por um processo de 2-estados, do tipo $2U \leftrightarrow D$. Essa questão pode ser mais complexa ao se analisar o processo de enovelamento por meio de diferentes parâmetros de ordem e em diferentes concentrações de proteína. De acordo com Svensson et al, o enovelamento da SOD deveria ser de 3-estados ($2U \leftrightarrow 2M \leftrightarrow D$) apenas em concentrações menores do que $30\mu\text{M}$ de proteína (acredita-se que *in vivo* a concentração seja em torno de $100\mu\text{M}$ (92), porém, não é citado a possibilidade de existir um estado do tipo M+U. A presença do estado M+U apenas nos resultados computacionais nos leva a pensar quais são os possíveis motivos para ele aparecer em dados *in silico*, e não aparecer nos dados *in vitro*. Uma das razões pode estar associada com a ponte dissulfeto, que no nosso caso é conformacional e nos experimentos é oxidada. Entretanto, uma outra possibilidade está na forma como acompanhamos o sistema se enovelar. Os dados experimentais realizado por Svensson et al (91), utiliza o β -Tanford (65) como coordenada para acompanhar o enovelamento. Quando alteramos o nosso parâmetro de ordem Q para o β -Tanford, vemos que o que antes caracterizava um processo de 3 ou 4-estados, agora passa a ser um perfil de energia livre com 2-estados ($2U \leftrightarrow D$), como pode ser visto na Figura 5.8. A ausência do estado 2M, mesmo utilizando a coordenada β -Tanford para o caso de menor concentração, nos induz a acreditar que a nossa menor concentração ainda é maior do que a de $30\mu\text{M}$ determinada experimentalmente.

Figura 5.8: **β -Tanford como coordenada de reação** (a) Perfil de energia livre (eixo y a esquerda) para o dímero da SOD1 em alta concentração (linha preta), e a relação entre Q (eixo y a direita) e β -Tanford (pontos na cor vinho). (b) Perfil de energia livre (eixo y a esquerda) para o dímero da SOD1 com menor concentração (linha preta), e a relação entre Q (eixo y a direita) e β -Tanford (pontos na cor vinho). Ambos estão nas respectivas temperaturas de enovelamento. Os símbolos 2U, M+U, 2M e D remetem, respectivamente, à dois monômeros desenovelados, um monômero enovelado e o outro desenovelado, dois monômeros enovelados, e o dímero enovelado.



Fonte: Elaborada pelo autor

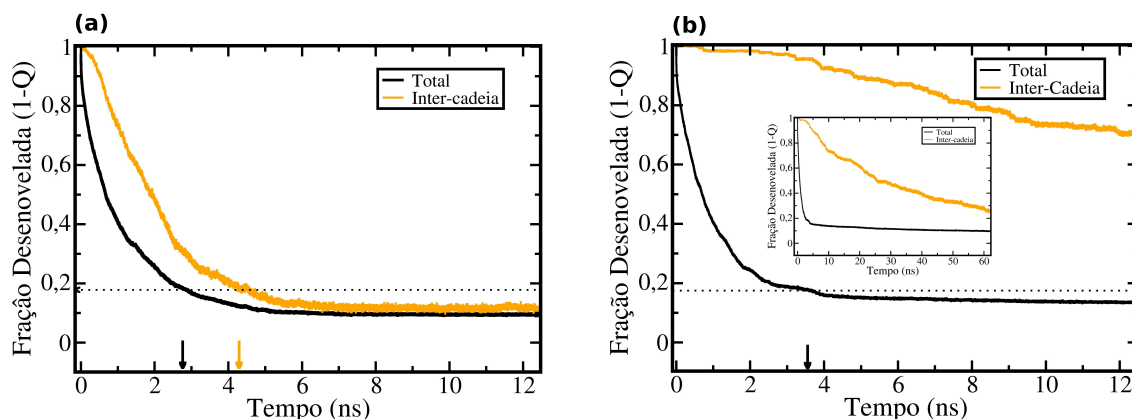
A alteração do perfil de energia livre de acordo com a coordenada de reação utilizada, nos mostra a importância da escolha deste tipo de parâmetro. Na Figura 5.8, ao confrontar os valores de β -Tanford com os valores de Q para as conformações amostradas pela simulação, é possível perceber a diferença entre estas duas coordenadas. O β -Tanford se mostra uma boa coordenada de reação para separar estruturas com baixos valores de Q, porém, estruturas com valores elevados de contatos nativos, não apresentam variações significativas nos valores de β -Tanford, sendo melhor “separadas” pela fração de contatos nativos. Neste caso, de acordo com a relação entre as duas coordenadas, as conformações do tipo M+U podem estar localizadas na região de transição (TS), ou mescladas com o estado desnaturado na Figura 5.8, um vez que o β -Tanford se mostrou bastante degenerado, podendo ser essa, uma das razões para este estado não ser evidente.

5.2.2 Simulações de controle - Cinética de enovelamento do dímero

Ao fazermos simulações de controle, procuramos investigar se a amostragem obtida pelo REMD foi suficiente para nos dar resultados confiáveis. As 200 simulações realizadas somente com o SBM- C_α para o dímero em menor e maior concentração, vão na direção de tentar explicar alguns parâmetros termodinâmicos das simulações com o REMD, por meio dos parâmetros cinéticos obtidos por esta gama de simulações. A Figura 5.9 nos mostra dados à respeito da formação dos contatos nativos totais e dos contatos inter-cadeias.

Os dados de cinética de enovelamento do dímero apontam novamente para a dependência deste com a concentração de proteína, assim como proposto por Svensson et al, a qual mostra que quanto maior a concentração de SOD, mais desnaturante é necessário para desenovelá-la, indicando um aumento na estabilidade do dímero. Além deste aumento na estabilidade, já mostrada nos gráficos de energia livre da Figura 5.7 (b), vemos pelo gráfico de cinética da Figura 5.9 (a) que a fração de desenovelamento do caso mais concentrado atinge a linha tracejada em um tempo 1.32 vezes mais rápido que o caso menos concentrado em (b), indicando a presença de uma barreira menor para enovelar os dois monômeros. Contudo, a formação dos contatos inter-cadeia, mostrou ser o fator limitante do enovelamento, sendo quase 15 vezes mais lento no caso de menor concentração do que no caso mais concentrado. Tais dados mostram que a concentração da SOD1 também afeta a cinética de enovelamento, principalmente para o processo de dimerização, indicando que quanto maior a concentração, maior a taxa de enovelamento.

Figura 5.9: **Cinética de enovelamento do Dímero.** Formação dos contatos totais dos dois monômeros (em preto) e entre os resíduos da interface inter-cadeia (em amarelo) para o dímero em (a) maior concentração de proteína, e (b) menor concentração de proteína. As linhas tracejadas, nos dois gráficos, foram fixadas no valor da fração desenovelada referente ao mínimo nativo ($Q \approx 0.83$) obtido pela curva de energia livre (Figura 5.7). As flechas indicam, de acordo com a cor, o tempo no qual cada estrutura atinge a linha tracejada. O gráfico no interior da figura (b) indica o tempo necessário para a formação dos contatos inter-cadeias.

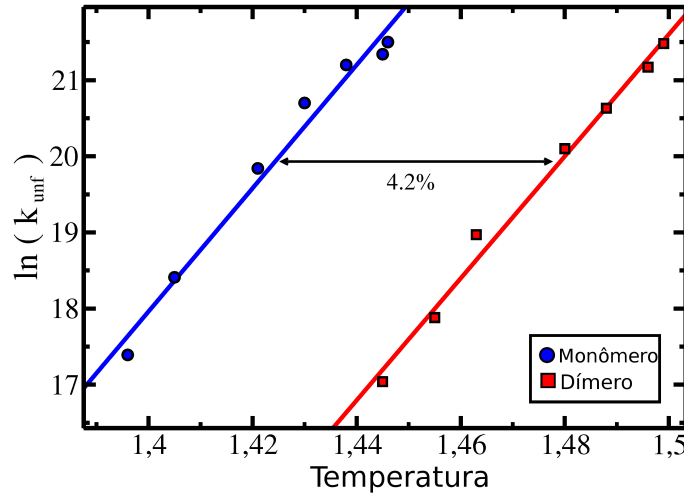


Fonte: Elaborada pelo autor

5.2.3 Estabilidade dímero-monômero

Como discutido nas seções anteriores, aparentemente não só a presença dos metais pode afetar os valores das taxas de enovelamento, mas a própria concentração de monômeros também. Os dados de cinética de enovelamento do dímero da SOD1 nos mostra que a estrutura dimérica é mais estável do que o monômero isolado. Tais resultados podem ser vistos na Figura 5.10. Quando colocamos o dímero e o monômero nas mesmas condições de desenovelamento (mesma taxa), vemos que em média, o dímero está em uma temperatura 4.2% maior, apresentando ser mais estável do que a forma monomérica da SOD1. Tal fato já nos daria um indício sobre o porque da enzima ser um dímero na maioria dos organismos. Comparando nossos dados com alguns encontrados na literatura, vemos que o monômero nativo apresenta uma temperatura de melting em torno de 66°C, enquanto o dímero, dependendo do organismo vivo, possui uma temperatura de melting que varia entre 75°C e 93°C, o que nos dá um intervalo de 2.6% à 7.9% em ganho de estabilidade, o que coloca nossos resultado dentro deste intervalo.

Figura 5.10: **Logaritmo da taxa de desenovelamento em função da temperatura.** Para cada temperatura, calculou-se os valores das taxas para que o monômero (círculos em azul) e para o dímero (círculos em vermelho) se desenovelassem. Em cada temperatura, cada sistema foi simulado 50 vezes, dizendo que o sistema havia se desenovelado quando ele atingisse 0.15 do valor de Q total. As temperaturas estão em unidades reduzidas do *Gromacs*. Em média, o dímero se mostrou 4.2% mais estável que o monômero. As retas (em azul e vermelho) foram obtidas pelo ajuste linear dos pontos



Fonte: Elaborada pelo autor

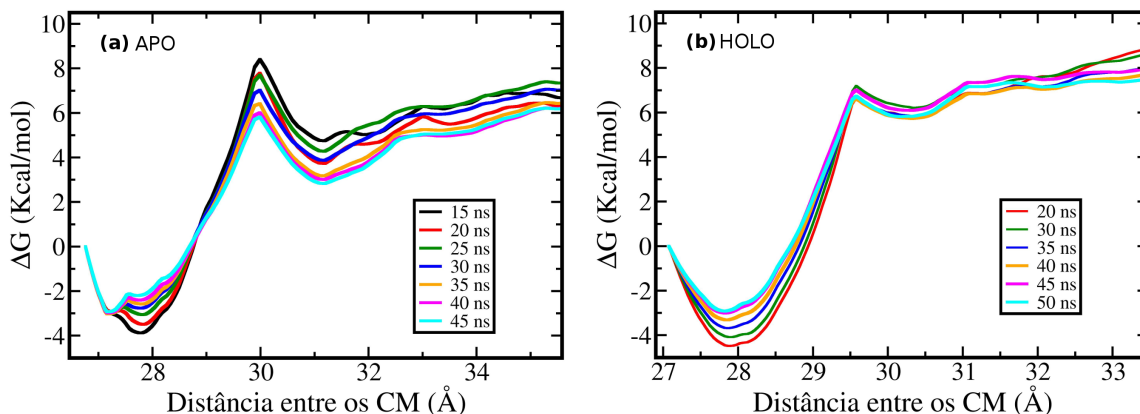
5.3 Simulações de Dimerização

5.3.1 Utilizando o Adaptive Biasing Force (ABF)

Nesta tese, utilizamos o método do ABF com o objetivo de analisar as grandezas termodinâmicas relativas ao processo de dimerização da SOD1. Neste sentido, mostraremos nesta seção os resultados das simulações nas quais afastamos as duas cadeias monoméricas já enoveladas, tanto na presença quanto na ausência de íons de Zn. Como primeiro resultado, temos os perfis de Energia Livre de dimerização em cada intervalo de tempo de simulação, os quais podem ser vistos na Figura 5.11. Cada uma das cores representa um tempo de simulação, e o maior tempo de cada gráfico indica o momento no qual ocorreu a convergência. Assim, temos que as curvas da SOD1 Apo atingiram o limiar de convergência após cada “janela” ter simulado 45ns, totalizando 540ns simulados. Já para o caso Holo, cada janela simulou 50ns, totalizando 600ns.

Quando analisamos de forma isolada somente os perfis de energia livre no tempo em que a convergência se deu, como pode ser visto na Figura 5.12, podemos inferir com afinco que a distância nativa entre os CM (27.9Å) está na região de mais baixa

Figura 5.11: **Convergência da simulação realizada com o ABF.** Cada uma das curvas remetem ao perfil de energia livre de acordo com o tempo de simulação de cada janela, em função da distância entre os centros-de-massa de cada monômero. Em (a) temos o comportamento da SOD1 Apo , enquanto que em (b) temos a simulação da SOD1 Holo(2xZn).



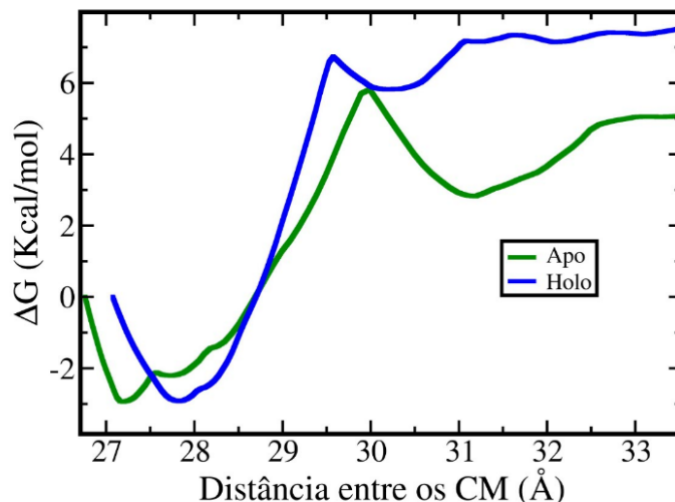
Fonte: Elaborada pelo autor

energia livre do sistema, nos dando parâmetros numéricos que ajudam a mostrar do ponto de vista termodinâmico, que a conformação mais estável da SOD1 é o dímero, e não dois monômeros isolados. Este resultado se mostrou independente da presença dos íons de Zn, ou seja, o dímero apo também se mostrou estável. Pode-se observar na Figura 5.12 que, à medida que separamos as duas cadeias, temos um aumento na energia livre, indicando que uma separação de $2\text{\AA}$ em relação a distância mais estável entre os centros de massa, já é suficiente para gerar uma barreira em torno de 9.0 kcal/mol. Deste modo, vemos que a formação da interface, independente da presença dos metais, está diretamente relacionada com a estabilização dos dois monômeros no dímero.

Apesar dos perfis de energia livre da Figura 5.12 indicarem que a barreira para desfazer o dímero Holo ser apenas 1 kcal/mol maior do que o caso Apo, as duas conformações apresentam perfis com algumas diferenças, principalmente entre as inclinações da barreira de energia livre, o que pode estar associado a fatores cooperativos na dimerização, e, pela presença de estados intermediários antes da formação do dímero nativo. Para o caso Apo, o estado intermediário, com mínimo em $31\text{\AA}$, é estabilizado por uma barreira de 3 kcal/mol. Já na presença dos metais, o ponto de mínima energia do estado intermediário em $30\text{\AA}$ indica que a SOD1 deve vencer uma barreira de 1 kcal/mol para se dimerizar.

Com os CM afastados em $34\text{\AA}$, os dois monômeros Apo se mostraram ser 2.5kcal/mol mais estável que os dois monômeros Holo. Estes valores de energia livre elucidados pelo ABF, podem nos indicar que a falta dos metais ocasiona uma

Figura 5.12: **Energia Livre de dimerização utilizando o ABF.** Cada uma das curvas representam o perfil de energia livre de dimerização da SOD1 Holo (em azul) e Apo (em verde) em função da distância entre os centros-de-massa (CM) dos monômeros. As simulações foram realizadas em 298K.



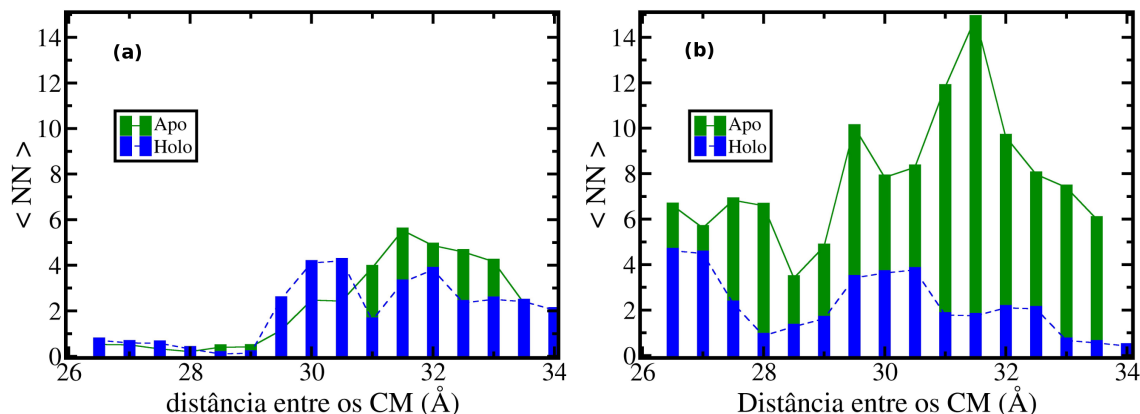
Fonte: Elaborada pelo autor

difficuldade a mais no processo de dimerização, o que vai na direção do resultado proposto por Strange et al (46), o qual afirma que na ausência dos metais, a SOD1 é mais propícia a monomerização.

O aparecimento destes estados intermediários, por serem algo ainda não relatado na literatura, nos fizeram duvidar a respeito de sua real existência, levantando o questionamento se eles surgiram pois eram ocasionados por algum artifício do próprio ABF, ou, se eles realmente eram passíveis de existir, e neste caso, propor o que poderia estar ocasionando tal existência. A primeira dúvida será respondida por meio de simulações de controle de dinâmica livre (seção 5.3.2). Já a segunda questão pode ser respondida quando analisamos a formação de contatos não-nativos durante o processo de dimerização da SOD1. A figura 5.13 nos mostra, com relação à distância entre os centros-de-massa, os valores médios nas quantidades de contatos não-nativos formados para o caso Apo e Holo. Nesta figura, analisamos a formação de contatos não-nativos realizados entre resíduos de aminoácidos de uma mesma cadeia (contatos intracadeia - Figura 5.13 (b)), e entre resíduos de aminoácidos de cadeias diferentes (contatos intercadeia - Figura 5.13 (a)).

A quantidade de contatos não-nativos também parece ser um fenômeno relacionado com a presença/ausência dos metais. Em média, se somarmos os contatos inter e intracadeia, podemos ver que o sistema Apo forma mais contatos não-nativos do que a proteína Holo. Entretanto, o pico de formação de contatos não-nativos, nos dois casos,

Figura 5.13: **Formação de contatos não-nativos em função da distância entre os centros-de-massa (CM).** (a) Contatos não-nativos intercadeias. (b) Contatos não-nativos intracadeias. Em azul temos a quantidade de contatos não-nativos da SOD1 Holo (2xZn), e em verde da SOD1 Apo. As linhas ligando as barras são apenas guias para os olhos.



Fonte: Elaborada pelo autor

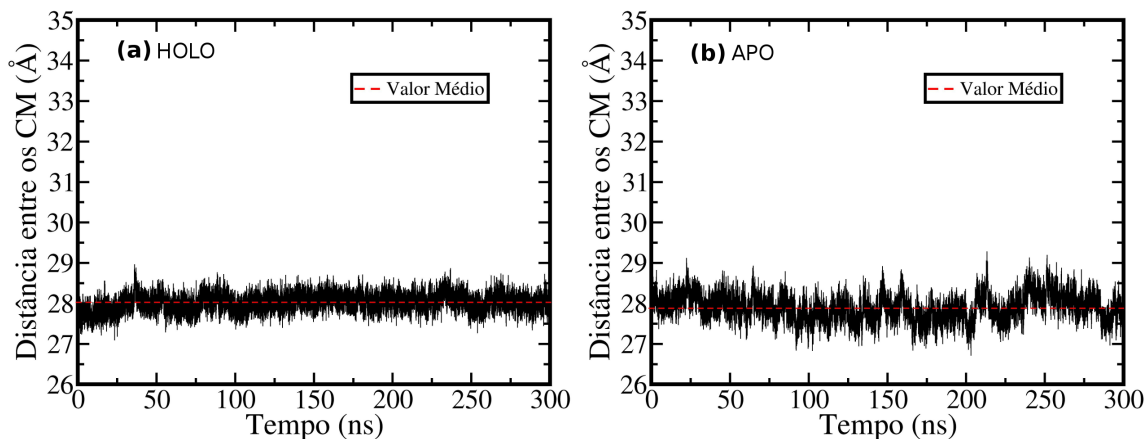
se dão em regiões próximas onde estão localizados os estados intermediários, podendo este tipo de interação ser a causa da presença de um estado metaestável antes do dímero nativo ser formado. Aparentemente, quando os CM dos monômeros estão separados em $31.5\text{\AA}$, as interações não-nativas intracadeia ajudam na estabilização da proteína na ausência dos metais, além de indicar a formação de uma interface não-nativa. Já na presença dos metais, a estabilidade da proteína em $30\text{\AA}$ parece ser ocasionada de forma igual pela formação de contatos não-nativos inter e intracadeia. Mourou et al (25) mostraram que de acordo com a quantidade de contatos não-nativos formados, as proteínas podem ter seus processos de enovelamento favorecidos ou desfavorecidos. Neste caso, a diferença na quantidade de contatos não-nativos formados na proteína Apo, somado a um aumento na dimensão da barreira de energia livre, nos dá indícios que neste caso, estas interações podem estar desfavorecendo a dimerização da SOD1 sem os íons de Zn.

5.3.2 Simulações de Controle - Sistema Livre

Simulações longas com os sistemas livres (sem o ABF) foram feitas para verificar a permanência dos monômeros/dímero nos estados em que o método do ABF apresentou como sendo possíveis pontos de mínima energia (local e global). O primeiro resultado aqui mostrado, na Figura 5.14, nos informa o valor da distância entre os centros-de-massa dos monômeros em função do tempo de simulação, para o caso dHolo_N e dApo_N. Estas simulações foram iniciadas com os dois monômeros na posição

dimérica nativa (extraída do arquivo PDB).

Figura 5.14: **Distância entre os centros-de-massa (CM) durante 300ns.** (a) Simulação da SOD1 Holo iniciada com a proteína na conformação nativa ($dHolo_N$). (b) Simulação da SOD1 Apo iniciada com a proteína na conformação da SOD1 Holo nativa. As simulações foram realizadas em 298K. As linhas tracejadas (em vermelho) correspondem ao valor médio de distância durante os 300ns de simulação.

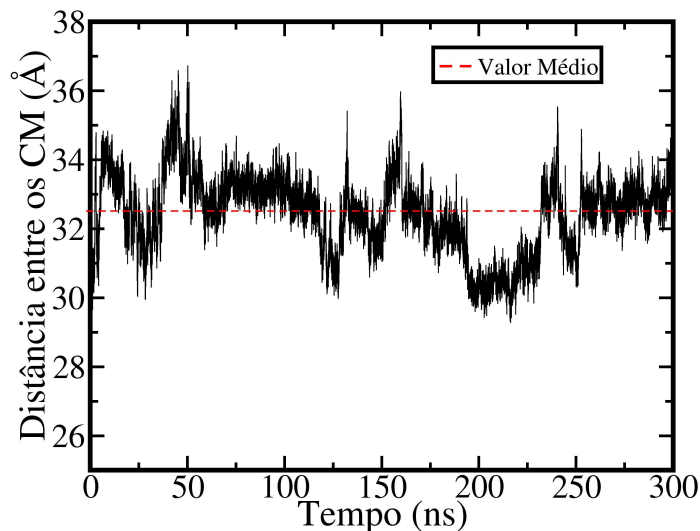


Fonte: Elaborada pelo autor

Quando iniciada com os monômeros distanciados em $27.9 \text{ \AA}$ (distância nativa entre os centros-de-massa), a proteína Holo ficou, em média, em uma distância de $28.01 \text{ \AA}$, apresentando uma flutuação em torno de $0.50 \text{ \AA}$ da distância nativa. Na ausência dos metais, a proteína Apo, ficou com os centros-de-massa afastados, em média, $27.88 \text{ \AA}$, com uma flutuação de $0.74 \text{ \AA}$. As baixas flutuações e a permanência das proteínas nestes valores de distância, nos mostram que a região de mínimo global apontada pelo gráfico de energia-livre do ABF, corresponde à conformações estáveis, o que vai na direção de afirmar que a forma dimérica é a conformação na qual a SOD1 é mais estável.

Ao iniciarmos a simulação com a proteína sem os metais, e com os centros-de-massa afastados em $31 \text{ \AA}$, vemos que a SOD1 também permanece em torno da região na qual o ABF indicou como existindo um mínimo local (estado intermediário). Neste caso, como pode ser visto na Figura 5.15, a proteína apresentou uma flutuação maior na distância entre os centros-de-massa (quando comparado com a Figura 5.14), varrendo um espaço conformacional que vai de $30 \text{ \AA}$ à $36 \text{ \AA}$. Se compararmos com a curva de energia livre do ABF (Figura 5.12), vemos que para o caso Apo, a barreira de dimerização tem seu pico em $30 \text{ \AA}$, indicando que a proteína chegou nesta região mas voltou para valores mais altos de distância entre os CM, característica que pode indicar a presença de uma barreira nesta região, e que a proteína não teve energia suficiente, durante os 300ns simulados, para ultrapassá-la.

Figura 5.15: **Distância entre os centros-de-massa (CM) para a SOD1 dApo_I** Variação entre as distâncias de centro-de-massa em função do tempo, para o sistema dApo_I. A simulação foi iniciada com os monômeros afastados em 31Å. Esta simulação foi realizada em 298K e 1atm de pressão. A linha tracejada em vermelho corresponde a média aritmética da distância mantida pela proteína durante os 300ns.



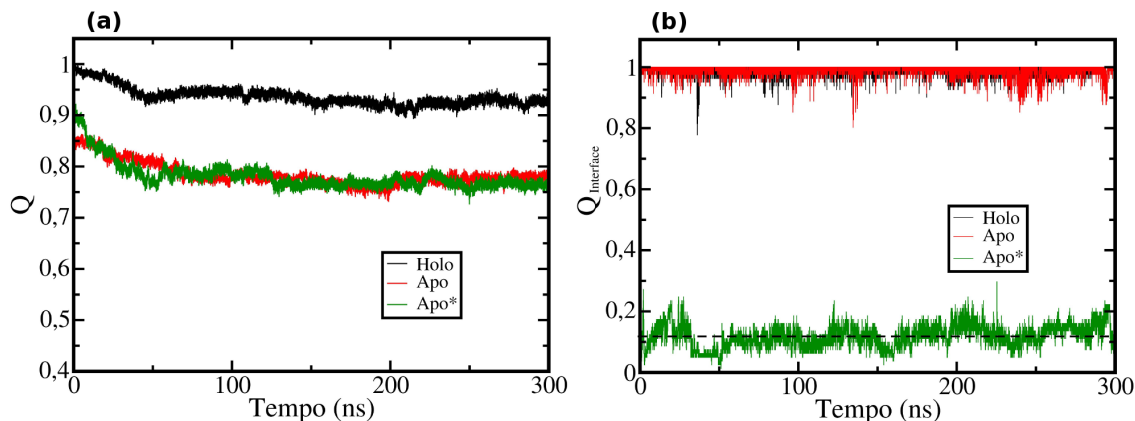
Fonte: Elaborada pelo autor

Na Figura 5.16, mostramos o valor da fração de contatos nativos mantidos pela proteína ao longo de toda a simulação. Podemos perceber na Figura 5.16 (a), que na presença dos metais, 90% dos contatos nativos são mantidos. Por outro lado, na ausência dos metais, tanto para o caso dApo_N (no mínimo global de energia) quanto para o dApo_I (no mínimo local - intermediário), praticamente 22% dos contatos nativos são perdidos durante a simulação.

Associando os resultados da Figura 5.16 (a) com os de (b), pode-se notar que a proteína Apo, quando simulada na distância de localização do mínimo global de energia, manteve a interface completamente formada, sendo assim, os 22% de contatos que foram desfeitos são formados exclusivamente entre resíduos de aminoácidos que estão numa mesma cadeia, indicando que na ausência dos metais, deve haver uma maior flutuação na posição de alguns resíduos de aminoácidos que estão afastados da região da interface, e esta flutuação está ocasionando a perda de contatos nativos. Todavia, para o caso dApo_I, a perda de contatos nativos deve ser devido tanto à uma flutuação intracadeia, quanto intercadeia, uma vez que a interface manteve, em média, apenas 11% dos contatos nativos. Maiores detalhes à respeito da flutuação residual na SOD1 será discutido na Figura 5.17

A presença de alguns contatos nativos na interface da proteína dApo_I, associado com os contatos não-nativos elucidados com o método do ABF, mostram que este

Figura 5.16: **Fração de contatos nativos em função do tempo de simulação** (a) Fração de contatos nativos totais para as simulações: dHolo_N (iniciada na conformação nativa, curva em preto); dApo_N (iniciada na conformação nativa, curva em vermelho); e da dApo_I iniciada com os centros-de-massa distânciados em 31Å (curva em verde). (b) Fração de contatos intercadeia formados na interface nativa. Os padrões de cores seguem o mesmo definido na figura (a).



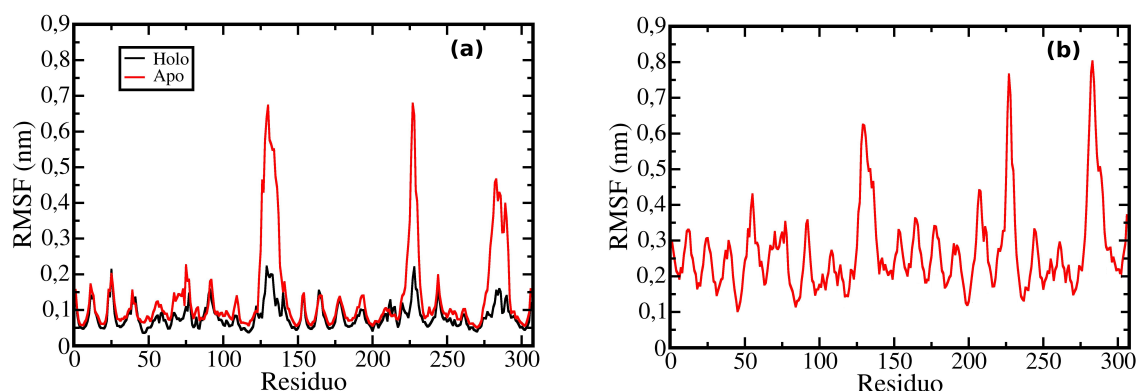
Fonte: Elaborada pelo autor

estado intermediário pode ser composto por um dímero não-nativo. A formação de dímeros não-nativos da SOD1 foi recentemente mostrado por Itsuki et al (93), que de forma experimental, investigou um comportamento anormal de mutantes da SOD1 em condições fisiológicas (37°C). Estes mutantes mantinham uma estrutura dimérica mas perdiam algumas conformações secundárias, os quais poderiam formar oligômeros por meio de pontes dissulfeto intermoleculares. Embora este dado possa trazer um respaldo para a presença deste estado intermediário metaestável, ainda é necessário uma análise mais detalhada para comprovar a sua real existência.

A perda de estruturas secundárias pode ser causada por flutuações conformacionais, como foi citado na discussão da Figura 5.16. Maiores detalhes a respeito da flutuação média de cada resíduo podem ser vistos na Figura 5.17. Quando simulada em distância nativa, vemos que ao menos 3 regiões do dímero Apo sofrem mais flutuações do que na proteína Holo. Estas três regiões no dímero estão localizadas em duas regiões específicas de cada monômero: no loop IV (loop de ligação com os metais, entre os aminoácidos 42-85 e 195-238 do dímero) e no loop VII (loop eletrostático, entre os aminoácidos 121-148 e 274-301 do dímero). Os demais aminoácidos mantêm flutuações próximas aos do dímero Holo. Recentemente, Habibi et al (88), utilizando outra técnica de simulação, obteve como resultado que os loop IV e VII dos monômeros Apo da SOD1 possuem valores elevados de flutuação, o que corrobora com os nossos resultados. Ao afastarmos os dois monômeros, podemos perceber pela Figura 5.17(b),

que o dímero Apo aumenta a flutuação de uma maneira global, mas ainda mantem os maiores valores nos loops IV e VII. Neste caso, nota-se que tanto a presença/ausência dos metais quanto a da interface nativa interfere na flutuação dos aminoácidos, sendo o dímero nativo Holo a conformação com menor flutuação.

Figura 5.17: **Valor do RMSF para cada resíduo da SOD1** (a) Os valores de RMSF para os resíduos da curva em preto remetem à simulação da SOD1 Holo iniciada na conformação nativa. Já os valores da curva em vermelho são da simulação da SOD1 Apo iniciadas nesta mesma conformação. (b) Os valores de RMSF aqui representados foram obtidos pela simulação da SOD1 apo iniciada no estado intermediário elucidado pelo ABF (Figura 5.12).



Fonte: Elaborada pelo autor

Capítulo 6

Conclusões

As simulações realizadas neste trabalho elucidaram alguns efeitos que a presença dos metais podem causar no enovelamento e dimerização da SOD1. Podemos concluir e reponder a primeira pergunta do nosso objetivo indicando que os metais, Zn e Cu, em concordância com a formação do sítio metálico, auxiliam a proteína a se enovelar, diminuindo a barreira de energia livre. Tal fato indica que eles podem já se ligar na enzima ainda desenovelada, uma vez que esta região se mostrou estar formada já nos primeiros estágios do enovelamento. Neste caso, além de auxiliar a proteína a se enovelar, os metais também são responsáveis por conferir uma maior estabilidade para o estado nativo. A formação precoce do sítio metálico mostrou interferir diretamente no caminho de enovelamento seguido por cada monômero, o qual se torna menos específico. A ausência dos metais, além de retardar o processo de enovelamento, indicou que nesta condição, é maior a probabilidade da SOD1 visitar conformações de transição que podem levar a proteína a formar estruturas patogênicas, tais como as folhas- β 2, 3, e 6.

De uma forma geral, a barreira de energia livre para o enovelamento do monômero, apesar de ter um caráter entrópico, mostrou ser dominada pela entalpia, uma vez que poucos contatos nativos estão sendo realizados no estado de transição. Tanto a entalpia, quanto a variação de entropia conformacional se mostraram importantes para elucidar o papel da ponte dissulfeto no enovelamento de cada monômero. A formação da ponte dissulfeto intracadeia entre as cisteínas 57 e 146, se mostrou importante para garantir a estabilidade do estado nativo de cada monômero, diminuindo a energia do ensemble nativo. O caso com ponte dissulfeto conformacional nos oferece a hipótese de que *in vivo* a ponte dissulfeto seja feita apenas após o estado de transição, não sendo responsável por auxiliar a proteína a superar a barreira de energia livre. Porém, a formação precoce da ponte dissulfeto foi responsável por uma pequena modificação na

posição e na altura da barreira de transição no caso apo, indicando que na presença da ponte dissulfeto oxidada, temos um menor custo entrópico de enovelamento, o qual a proteína reduzida tenta compensar formando mais contatos nativos (diminuindo a entalpia do sistema) para se enovelar. A ausência da ponte dissulfeto, associada com a falta dos metais, também pode favorecer um processo de *misfolding*, uma vez que o caso apo reduzido mostrou ter maior probabilidade de formar regiões da SOD1 com alta propensão de formar agregados. Valores experimentais de estabilidade, de compactação da cadeia, e de cinética de enovelamento, corroboram nossos dados e a forma como foi encontrada a temperatura ambiente em unidades reduzidas.

O enovelamento da SOD1 se mostra dependente das condições de concentração de proteína. O aumento da estabilidade em alta concentração de proteína pode estar associado com a não formação do estado 2M, uma vez que a região da interface, altamente hidrofóbica, reduz a sua área acessível ao solvente quando esta se enovela de forma concomitante com a cadeia. Além disso, formar a interface de forma concomitante com o enovelamento da cadeia mostrou acelerar a cinética de formação do dímero em 15 vezes, diminuindo a exposição dos resíduos da interface ao meio. Atingir o mesmo valor de taxa de desenovelamento em uma temperatura 4.2% maior, indica que o dímero da SOD1 é mais estável do que o monômero isolado, podendo ser essa uma das razões para a enzima atuar como dímero.

Os dados de taxa de desenovelamento, de perfil de energia livre obtidos pelo ABF, e os de flutuações nos resíduos das simulações livres, indicam que a SOD1, apesar de ser formada por dois monômeros que trabalham de forma independente, tem na sua forma dimérica a sua conformação mais termodinamicamente e estruturalmente estável. A formação dos contatos de resíduos de aminoácidos entre as duas cadeias, não só reduz a área acessível ao solvente nestes resíduos da interface, como também interfere na flutuação dos resíduos ao longo da cadeia, uma vez que na presença da interface nativa, a flutuação da cadeia foi menor. O dímero mostra ser a estrutura mais estável mesmo na ausência dos metais, porém, a presença dos metais, também afetam o processo de dimerização. Os metais são responsáveis por diminuir a barreira de energia livre para os dois monômeros se dimerizarem, além de dificultar o processo de separação entre as duas cadeias, dando mais estabilidade para o dímero nativo, e desestabilizar a formação de estados intermediários, os quais se mostraram formados por dímeros não-nativos.

REFERÊNCIAS

1. VOET, D.; PRATT, J.; W, C. Fundamentos de bioquímica: a vida em nível molecular. *Artmed*, v. 2 ed, 2008.
2. NYMEYER, H.; GARCIA, A. E.; ONUCHIC, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA*, v. 95, p. 5921–5928, 1998.
3. SHOEMAKER, B. A.; WANG, J.; WOLYNES, P. G. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. USA*, v. 94, p. 777–782, 1997.
4. WHITFORD, P. C. et al. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Struct. Funct. Bioinf.*, v. 75, p. 430–441, 2009.
5. OLIVEIRA, R. J. et al. Coordinate and time-dependent diffusion dynamics in protein folding. *Methods*, v. 52, n. 1, p. 91–98, 2010. ISSN 1046-2023.
6. FERSHT, A. R. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.*, v. 5, n. 1, p. 79–84, 1995.
7. Garcia-Mira, M. M. et al. Experimental identification of downhill protein folding. *Science*, v. 298, n. 5601, p. 2191–2195, 2002.
8. NETTELS, D. et al. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA*, v. 104, p. 2655–2660, 2007.
9. CHUNG, H. S.; LOUIS, J. M.; EATON, W. A. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA*, v. 106, n. 29, p. 11837–11844, 2009.
10. ONUCHIC, J. N. et al. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. In: *Adv. Protein Chem.* [S.l.]: Elsevier, 2000. v. 53, p. 87–152. ISBN 9780120342532.
11. WANG, J. et al. The topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. USA*, v. 109, n. 39, p. 15763–15768, 2012.
12. GOSAVI, S. et al. Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.*, v. 357, p. 986–996, 2006.

10. Onuchic, J. N., H. Nymeyer, A. E. Garcia, J. Chahine, and N. D. Socci. 2000. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. *In* *Adv. Protein Chem.*, volume 53. Elsevier. 87–152.
11. Wang, J., R. J. Oliveira, X. Chu, P. C. Whitford, J. Chahine, W. Han, E. Wang, and V. B. P. Leite. 2012. The topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. USA*. 109:15763–15768.
12. Gosavi, S., L. L. Chavez, P. A. Jennings, and J. N. Onuchic. 2006. Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.* 357:986–996.
13. Onuchic, J. N., P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA*. 92:3626–3630.
14. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
15. Onuchi, J., H. Nymeyer, A. Garcia, J. Chahine, and N. D. Socci. 2000. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. *Adv. Protein Chem.* 53:87–152.
16. Eaton, W. A., P. A. Thompson, C.-K. Chan, S. J. Hage, and J. Hofrichter. 1996. Fast events in protein folding. *Structure*. 4:1133–1139.
17. Shakhnovich, E. I., and A. M. Gutin. 1989. Formation of unique structure in polypeptide chains. theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34:187–199.
18. Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.
19. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes. 1992. Optimal protein-folding codes from spin-glass theory. *Proc. Natl. Acad. Sci. USA*. 89:4918–4922.
20. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.* 101:6052–6062.

21. Bryngelson, J. D., J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein-folding - A synthesis. *Proteins*. 21:167–195.
22. Oliveira, L. C., R. T. H. Silva, V. B. P. Leite, and J. Chahine. 2006. Frustration and hydrophobicity interplay in protein folding and protein evolution. *J. Chem. Phys.* 125:084904.
23. Contessoto, V. G., D. T. Lima, R. J. Oliveira, A. T. Bruni, J. Chahine, and V. B. P. Leite. 2013. Analyzing the effect of homogeneous frustration in protein folding. *Proteins: Struct. Funct. Bioinf.* 81:1727–1737.
24. Clementi, C., and S. S. Plotkin. 2004. The effects of nonnative interactions on protein folding rates: Theory and simulation. *Prot. Sci.* 13:1750–1766.
25. Mouro, P. R., V. de Godoi Contessoto, J. Chahine, R. J. de Oliveira, and V. B. P. Leite. 2016. Quantifying nonnative interactions in the protein-folding free-energy landscape. *Biophys. J.* 111:287 – 293.
26. Dobson, C. M. 2003. Protein folding and misfolding. *Nature*. 426:884–890.
27. Andersen, P. M. 2006. Amyotrophic lateral sclerosis associated with mutations in the *cuzn* superoxide dismutase gene. *Curr Neurol Neurosci Rep.* 6:37–46.
28. Banci, L., I. Bertini, M. Boca, S. Girotto, M. Martinelli, J. S. Valentine, and M. Vieru. 2008. Sod1 and amyotrophic lateral sclerosis: Mutations and oligomerization. *PLOS ONE*. 3:1–8.
29. Werneck, L. C., R. Bezerra, O. d. Silveira Neto, and R. H. Scola. 2007. A clinical epidemiological study of 251 cases of amyotrophic lateral sclerosis in the south of Brazil. *Arquivos de Neuro-Psiquiatria*. 65:189 – 195.
30. Bruijn, L. I., T. M. Miller, and D. W. Cleveland. 2004. Unraveling the mechanisms involved in motor neuron degeneration in als. *Annu. Rev. Neurosci.* 27:723–749.
31. Ding, F., Y. Furukawa, N. Nukina, and N. V. Dokholyan. 2012. Local unfolding of cu, zn superoxide dismutase monomer determines the morphology of fibrillar aggregates. *J. Mol. Biol.* 421:548 – 560. *Amyloid Structure, Function, and Molecular Mechanisms (Part II)*.
32. Das, A., and S. S. Plotkin. 2013. Sod1 exhibits allosteric frustration to facilitate metal binding affinity. *Proc. Natl. Acad. Sci. USA*. 110:3871–3876.

33. Tainer, J. A., E. D. Getzoff, K. M. Beem, J. S. Richardson, and D. C. Richardson. 1982. Determination and analysis of the 2 Å... structure of copper, zinc superoxide dismutase. *J. Mol. Biol.* 160:181 – 217.
34. Valentine, J. S., P. A. Doucette, and S. Z. Potter. 2005. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Annu Rev Biochem.* 74:563 – 593.
35. Sturtz, L. A., K. Diekert, L. T. Jensen, R. Lill, and V. C. Culotta. 2001. A fraction of yeast cu,zn-superoxide dismutase and its metallochaperone, ccs, localize to the intermembrane space of mitochondria: A physiological role for sod1 in guarding against mitochondrial oxidative damage. *J. Biol. Chem.* 276:38084–38089.
36. Hassan, H. M., and I. Fridovich. 1979. Paraquat and escherichia coli. mechanism of production of extracellular superoxide radical. *J. Biol. Chem.* 254:10846–10852.
37. Nordlund, A., and M. Oliveberg. 2006. Folding of cu/zn superoxide dismutase suggests structural hotspots for gain of neurotoxic function in als: Parallels to precursors in amyloid disease. *Proceedings of the National Academy of Sciences.* 103:10218–10223.
38. Schmidlin, T., K. Ploeger, A. L. Jonsson, and V. Daggett. 2013. Early steps in thermal unfolding of superoxide dismutase 1 are similar to the conformational changes associated with the als-associated a4v mutation. *Protein Engineering, Design and Selection.* 26:503–513.
39. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.
40. Roe, J. A., A. Butler, D. M. Scholler, J. S. Valentine, L. Marky, and K. J. Breslauer. 1988. Differential scanning calorimetry of copper-zinc-superoxide dismutase, the apoprotein, and its zinc-substituted derivatives. *Biochemistry.* 27:950–958. PMID: 2835081.
41. Biliaderis, C. G., R. J. Weselake, A. Petkau, and A. D. Friesen. 1987. A calorimetric study of human cuzn superoxide dismutase. *Biochem. J.* 248:981–984.
42. Parge, H. E., E. D. Getzoff, C. S. Scandella, R. A. Hallewell, and J. A. Tainer. 1986. Crystallographic characterization of recombinant human cuzn superoxide dismutase. *J. Biol. Chem.* 261:16215–16218.

43. Mei, G., N. Rosato, N. Silva, R. Rusch, E. Gratton, I. Savini, and A. Finazzi-Agro. 1992. Denaturation of human copper-zinc superoxide dismutase by guanidine hydrochloride: a dynamic fluorescence study. *Biochemistry*. 31:7224–7230. PMID: 1510915.
44. Ding, F., and N. V. Dokholyan. 2008. Dynamical roles of metal ions and the disulfide bond in cu, zn superoxide dismutase folding and aggregation. *Proc. Natl. Acad. Sci. USA*. 105:19696–19701.
45. Arnesano, F., L. Banci, I. Bertini, M. Martinelli, Y. Furukawa, and T. V. O’Halloran. 2004. The unusually stable quaternary structure of human cu,zn-superoxide dismutase 1 is controlled by both metal occupancy and disulfide status. *J. Biol. Chem*. 279:47998–48003.
46. Strange, R. W., C. W. Yong, W. Smith, and S. S. Hasnain. 2007. Molecular dynamics using atomic-resolution structure reveal structural fluctuations that may lead to polymerization of human cu–zn superoxide dismutase. *Proc. Natl. Acad. Sci. USA*. 104:10040–10044.
47. Rakhit, R., and A. Chakrabartty. 2006. Structure, folding, and misfolding of cu,zn superoxide dismutase in amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 1762:1025 – 1037. *Molecular Basis of Amyotrophic Lateral Sclerosis*.
48. Malinowski, D. P., and I. Fridovich. 1979. Bovine erythrocyte superoxide dismutase: diazo coupling, subunit interactions and electrophoretic variants. *Biochemistry*. 18:237–244. PMID: 420775.
49. Bertini, I., M. Piccioli, M. S. Viezzoli, C. Y. Chiu, and G. T. Mullenbach. 1994. A spectroscopic characterization of a monomeric analog of copper, zinc superoxide dismutase. *Eur. Biophys. J*. 23:167–176.
50. Banci, L., M. Benedetto, I. Bertini, R. Del Conte, M. Piccioli, and M. S. Viezzoli. 1998. Solution structure of reduced monomeric q133m2 copper, zinc superoxide dismutase (sod). why is sod a dimeric enzyme?,. *Biochemistry*. 37:11780–11791. PMID: 9718300.
51. Baldwin, R. 1995. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR*. 5:103–109.
52. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol*. 4:10–19.

53. Pande, V. S., A. Y. Grosberg, and T. Tanaka. 1997. On the theory of folding kinetics for short proteins. *Folding and Des.* 2:109–114.
54. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
55. Clementi, C., A. E. Garcia, and J. N. Onuchic. 2003. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: All-atom representation study of protein L. *J. Mol. Biol.* 326:933–954.
56. Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: A simulation study with a gö-like model. *J. Mol. Biol.* 313:171–180.
57. Chavez, L. L., J. N. Onuchic, and C. Clementi. 2004. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* 126:8426–8432.
58. Snow, C. D., E. J. Sorin, Y. M. Rhee, and V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Structure.* 34:43–69.
59. Chu, X., L. Gan, E. Wang, and J. Wang. 2013. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. USA.* :201220699.
60. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* 9:646–652.
61. Steinbach, P. J., and B. R. Brooks. 1993. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 90:9135–9139.
62. Verlet, L. 1967. Computer experiments on classical fluids. i thermodynamical properties of lennard-jones molecules. *Phys. Rev.* 159:98–103.
63. Tavares, Rafaela Magalhães. 2016. Estudos da interação de peptídeos antimicrobianos com modelo de membrana por simulações de dinâmica molecular.
64. Sugita, Y., and O. Y. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

65. Geierhaas, C. D., A. A. Nickson, K. Lindorff-Larsen, J. Clarke, and M. Vendruscolo. 2007. Bppred: A web-based computational tool for predicting biophysical parameters of proteins. *Protein Sci.* 16:125–134.
66. Bruno da Silva, F., V. G. Contessoto, V. M. de Oliveira, J. Clarke, and V. B. P. Leite. 2018. Non-native cooperative interactions modulate protein folding rates. *The Journal of Physical Chemistry B.* 122:10817–10824. PMID: 30407825.
67. Miyazawa, S., and R. L. Jernigan. 1999. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins.* 34:48–68.
68. Tuckerman, M. E. 2007. Free energy calculations: Theory and applications in chemistry and biology. *J. Am. Chem. Soc.* 129:10963–10964.
69. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.
70. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comp. Chem.* 26:1701–1718.
71. Noel, J. K., P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Research.* .
72. Noel, J. K., P. C. Whitford, and J. N. Onuchic. 2012. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B.* .
73. Humphrey, W., A. Dalke, and K. Schulten. 1996. Vmd: Visual molecular dynamics. *J. Mol. Graphics.* 14:33 – 38.
74. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable molecular dynamics with namd. *J. Comp. Chem.* 26:1781–1802.
75. Ferrenberg, A. M., and R. H. Swendsen. 1988. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638.
76. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized monte-carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

77. Plotkin, S. S., and J. N. Onuchic. 2000. Investigation of routes and funnels in protein folding by free energy functional methods. *Proceedings of the National Academy of Sciences of the United States of America*. 97:6509–6514.
78. Kayatekin, C., N. R. Cohen, and C. R. Matthews. 2012. Enthalpic barriers dominate the folding and unfolding of the human cu, zn superoxide dismutase monomer. *J. Mol. Biol.* 424:192 – 202.
79. Vassall, K. A., H. R. Stubbs, H. A. Primmer, M. S. Tong, S. M. Sullivan, R. Sobering, S. Srinivasan, L.-A. K. Briere, S. D. Dunn, W. Colón, and E. M. Meiering. 2011. Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in als. *Proc. Natl. Acad. Sci. USA*. 108:2210–2215.
80. Kayatekin, C., J. A. Zitzewitz, and C. R. Matthews. 2010. Disulfide-reduced als variants of cu, zn superoxide dismutase exhibit increased populations of unfolded species. *Journal of Molecular Biology*. 398:320 – 331.
81. Broom, Helen. 2015. Stability and aggregation studies of immature superoxide dismutase. Ph.D. thesis.
82. Battistoni, A., S. Folcarelli, L. Cervoni, F. Polizio, A. Desideri, A. Giartosio, and G. Rotilio. 1998. Role of the dimeric structure in cu,zn superoxide dismutase: ph-dependent, reversible denaturation of the monomeric enzyme from escherichia coli. *Journal of Biological Chemistry*. 273:5655–5661.
83. Stathopoulos, P. B., J. A. O. Rumfeldt, G. A. Scholz, R. A. Irani, H. E. Frey, R. A. Hallewell, J. R. Lepock, and E. M. Meiering. 2003. Cu/zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *Proc. Natl. Acad. Sci. USA*. 100:7021–7026.
84. Kayatekin, C., J. A. Zitzewitz, and C. R. Matthews. 2008. Zinc binding modulates the entire folding free energy surface of human cu,zn superoxide dismutase. *Journal of Molecular Biology*. 384:540 – 555.
85. Li, H.-T., M. Jiao, J. Chen, and Y. Liang. 2010. Roles of zinc and copper in modulating the oxidative refolding of bovine copper, zinc superoxide dismutase. *Acta Biochimica et Biophysica Sinica*. 42:183–194.
86. Lindberg, M. J., J. Normark, A. Holmgren, and M. Oliveberg. 2004. Folding of human superoxide dismutase: Disulfide reduction prevents dimerization and produces marginally stable monomers. *Proceedings of the National Academy of Sciences of the United States of America*. 101:15893–15898.

87. Stefanie D Boyd, L. B., Li Liu, and D. D. Winkler. 2018. Quantifying the interaction between copper-zinc superoxide dismutase (sod1) and its copper chaperone (ccs1). *Journal of Proteomics and Bioinformatics*. 11:1–5.
88. Habibi, M., J. Rottler, and S. S. Plotkin. 2017. The unfolding mechanism of monomeric mutant sod1 by simulated force spectroscopy. *Biochim. et Biophys. Acta*. .
89. Ivanova, M. I., S. A. Sievers, E. L. Guenther, L. M. Johnson, D. D. Winkler, A. Galaleldeen, M. R. Sawaya, P. J. Hart, and D. S. Eisenberg. 2014. Aggregation-triggering segments of sod1 fibril formation support a common pathway for familial and sporadic als. *Proc. Natl. Acad. Sci. USA*. 111:197–201.
90. Levy, Y., P. G. Wolynes, and J. N. Onuchic. 2004. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. USA*. 101:511–516.
91. Svensson, A.-K. E., O. Bilsel, E. Kondrashkina, J. A. Zitzewitz, and C. R. Matthews. 2006. Mapping the folding free energy surface for metal-free human cu,zn superoxide dismutase. *J. Mol. Biol.* 364:1084 – 1102.
92. Banci, L., I. Bertini, A. Durazo, S. Girotto, E. B. Gralla, M. Martinelli, J. S. Valentine, M. Vieru, and J. P. Whitelegge. 2007. Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: A possible general mechanism for familial als. *Proceedings of the National Academy of Sciences*. 104:11263–11267.
93. Anzai, I., E. Tokuda, A. Mukaiyama, S. Akiyama, F. Endo, K. Yamanaka, H. Misawa, and Y. Furukawa. 2017. A misfolded dimer of cu/zn-superoxide dismutase leading to pathological oligomerization in amyotrophic lateral sclerosis. *Protein Science*. 26:484–496.

Apêndice A

WHAM – *Weighted Histograms Analysis Method*

O método dos múltiplos histogramas é um método autoconsistente e tem como objetivo aumentar a informação sobre o sistema. Utilizando histogramas oriundos das simulação de um mesmo sistema em uma gama de temperaturas, fazemos com que os dados sejam melhores amostrados para que se possa ter uma boa estatística sobre as conformações adquiridas pelo sistema ao longo da simulação. Por meio deste método, somos capazes de gerar grandezas termodinâmicas tais como: perfis de calor específico em função da temperatura, e a energia livre do sistema com relação a coordenada de reação Q .

O método é baseado no fato de que o logaritmo da distribuição de probabilidade $P(Q)$ dos valores correlacionados com um certo Q em uma temperatura fixa, pode servir como uma estimativa do perfil da energia livre. Utilizando dados da mecânica estatística, para um conjunto de N simulações com C_n configurações para a temperatura $T_n = \frac{1}{k_B \beta_n}$, temos que a densidade de estados é dada por:

$$\Omega_n(Q) = H_n(Q) \exp(\beta_n E - f_n) \quad (\text{A.1})$$

no qual $H_n(Q)$ é o histograma da coordenada Q e f_n é a energia livre adimensional. Os valores para f_n são dados arbitrariamente, e ao longo do método são realizados interações que testam os valores até que se atinja a convergência esperada.

A probabilidade de o sistema adquirir a conformação Q numa dada temperatura T , é calculada da seguinte forma:

$$P(Q) = \frac{\sum_{n=1}^N H_n(Q) \exp(-\beta_n E)}{\sum_{n=1}^N C_n(Q) \exp(-\beta_n E + f_n)} \quad (\text{A.2})$$

De posse da probabilidade calculada de acordo com a equação A.2, podemos calcular a energia livre do sistema e as demais grandezas termodinâmicas:

$$\exp(-f) = \sum P(Q) \quad (\text{A.3})$$

Apêndice B

Artigo publicado durante o Doutorado

Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape

Paulo Ricardo Mouro,¹ Vinícius de Godoi Contessoto,¹ Jorge Chahine,¹ Ronaldo Junio de Oliveira,² and Vitor Barbanti Pereira Leite^{1,*}

¹Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, São Paulo, Brazil; and ²Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, Minas Gerais, Brazil

ABSTRACT Protein folding is a central problem in biological physics. Energetic roughness is an important aspect that controls protein-folding stability and kinetics. The roughness is associated with conflicting interactions in the protein and is also known as frustration. Recent studies indicate that an addition of a small amount of energetic frustration may enhance folding speed for certain proteins. In this study, we have investigated the conditions under which frustration increases the folding rate. We used a C_α structure-based model to simulate a group of proteins. We found that the free-energy barrier at the transition state (ΔF) correlates with nonnative-contact variation (ΔA), and the simulated proteins are clustered according to their fold motifs. These findings are corroborated by the Clementi-Plotkin analytical model. As a consequence, the optimum frustration regime for protein folding can be predicted analytically.

INTRODUCTION

Understanding the underlying folding mechanism of a protein to its functional compact three-dimensional structure is one of the great challenges of modern science. Failure in the process of a protein to achieve the correct folded native state can cause a series of pathological conditions, like neurodegenerative disorders, including Alzheimer's and Parkinson's diseases (1). Over the last decades, the energy-landscape theory has been a consistent framework in revealing protein-folding mechanisms (1–5). This theory states that the energy landscape of globular proteins resembles a funnel of structures progressively folded en route to the native state with its bottleneck narrowed at the transition state between unfolded and folded ensembles (6–9). The energy-landscape theory is successful in explaining, qualitatively and quantitatively, folding studies in theoretical (10–13) as well as experimental (14–17) investigations. Based on energy-surface theory, many computational models have been developed to predict folding mechanisms, rates, and stability parameters connected with experiments (9,18–22).

The protein-folding energy surface is multidimensional, and it has a funnelled topography as a function of the re-

action coordinate, usually described by the fraction of native contacts, Q . The funnel theory describes protein folding as an ensemble of conformations gradually diffusing from the unfolded state (lower Q), high entropic and energetic conformations at the top of the surface funnel, to the native state (higher Q), the lowest entropic and energetic state at its bottom (23–28). Proteins are naturally designed through evolution, so that folding pathways toward the native state are not dominated by bumps due to local energetic traps (5,29–33). Nature-designed sequences have the ability to fold completely on biological timescales.

To fold, the funnel energy slope must be steep enough to overcome the roughness and minimize local energetic trapping (13), or energetic frustration (21). Energetic frustration occurs due to the impossibility of satisfying all favorable energetic interactions simultaneously during folding events (34–37), allowing the formation of nonnative contacts (pairs of residues not in contact at the native state). Kinetically foldable proteins are naturally selected throughout the evolution process so that the native state is minimally frustrated (34–36,38,39).

Clementi and Plotkin introduced a theoretical/analytical model that takes into account the effects of nonnative interactions on the folding rate and on the folding free-energy barrier (40). This model indicates that a nonzero amount of nonnative energy interaction may enhance protein folding

Submitted December 17, 2015, and accepted for publication May 17, 2016.

*Correspondence: vleite@sjrp.unesp.br

Editor: Daniel Raleigh.

<http://dx.doi.org/10.1016/j.bpj.2016.05.041>

© 2016



rates. The theory for the folding speed limit upon the nonnative energetic frustration is well discussed in the literature (41,42), and studies have been done on specific proteins regarding nonnative interactions (26,43–45). In our previous article (46), using C_α structure-based models (SBMs) and homogeneous frustration, the maximum value of energetic frustration that can be added to the system to speed up folding rates was quantified. This maximum value was named the optimum energetic frustration (ϵ_f^{opt}) of each protein, and it was determined by analyzing the effect of the energetic frustration value (ϵ_{NN}) on free-energy barrier height and folding time. It was shown that not all proteins could have folding rates enhanced by including a small amount of frustration, i.e., ϵ_f^{opt} is zero. However, for all of the other proteins, ϵ_f^{opt} correlates with the free-energy barrier height (21) and the contact order parameter (47), clustering the proteins accordingly to their motif (48).

In this article, we seek the connection between the Clementi-Plotkin analytical model (40) and our previous computational results (46) through analysis of the underlying mechanism in nonnative-contact formation during folding. We show that the perturbed free-energy barrier variation ($\Delta\Delta F$) depends on the nonnative energetic contact parameter (ϵ_{NN}), and on the nonnative-contact variation (ΔA), in the folding transition state (Q^{ts}). These three quantities were calculated for a group of proteins using SBM simulations. The difference between the perturbed and nonperturbed SBM model, in both analytical and computational results, explains quantitatively the relevance of the nonnative-contact formation in the pretransition state for the maximal folding-speed-limit mechanisms. The former results of grouping the simulated proteins by their folding motif were also recovered.

MATERIALS AND METHODS

In this study, a group of 15 proteins (46) was used to investigate the effect of nonnative interactions in the transition state. Proteins were simulated using molecular dynamics and a coarse-grained C_α SBM (12,49). Despite the simplicity of this model, it can capture important characteristics of protein-folding mechanisms, as has been shown by computer simulations (9,49–51). The agreement between experimental and computational folding rates using the SBM is remarkable, as has been shown by Chavez et al. (21). The simulation details are explained in the [Supporting Material](#).

Effect of nonnative interactions on the free-energy barrier

Two order parameters are used by Clementi and Plotkin to map the folding process in their analytical model: the fraction of native contacts (Q) and the fraction of nonnative contacts (A) (40). These two order parameters are normalized between 0 and 1. The fraction of nonnative contacts depends on Q : the more native contacts are formed, the fewer nonnative interactions are allowed. The theoretical model does not allow nonnative contacts for $Q = 1$. The native attraction is characterized by the mean attraction energy, ϵ ($\epsilon < 0$), and two energy scales are used to analyze the nonnative contribu-

tion: the mean energy of a nonnative interaction (ϵ_{NN}) and the energetic variance of nonnative interaction (b^2). Nonnative interactions are designed to be weaker when compared to native interactions in a perturbative way. Expressions for thermal energy (E), free energy (F), and entropy (S) as a function of Q , A , and temperature (T) were normalized by the maximum number of contacts, M (Eqs. 7(a)–7(c) in (40)). The folding temperature of the model (T_F^0) was defined as the temperature at which folded and unfolded states have the same probability. Thus, at T_F^0 , $F(0, A) \approx F(1, 0)$ for $\epsilon_{\text{NN}} = b^2 = 0$. If $A^*(Q)$ is the most probable value of A in a certain Q , the difference between the most probable value in the transition state, $A^*(Q^{\text{ts}})$, and in the unfolded state, $A^*(0)$, is given by

$$\Delta A^*(Q^{\text{ts}}) \equiv A^*(Q^{\text{ts}}) - A^*(0). \quad (1)$$

The variation in free energy between the unfolded and the transition states at T_F^0 is calculated. It is shown that the free-energy barrier height (ΔF) corresponds to the sum of the free-energy barrier between these states in the absence of nonnative forces (ΔF^0) plus a term referring to energetic frustration:

$$\frac{\Delta F}{T_F^0} = \frac{\Delta F^0}{T_F^0} + M \left(\frac{\epsilon_{\text{NN}}}{T_F^0} - \frac{b^2}{2T_F^0{}^2} \right) \Delta A^*(Q^{\text{ts}}). \quad (2)$$

Since ϵ_{NN} always has negative values, Eq. 2 can be analyzed in two conditions. If $\Delta A^*(Q^{\text{ts}}) > 0$, more nonnative contacts are formed in the transition state than in the unfolded state, and the third term of Eq. 2 is negative, leading to a decrease in the free-energy barrier when compared with the barrier without nonnative interactions ($\Delta F < \Delta F^0$). This condition implies an increase in the folding rate. On the other hand, for $\Delta A^*(Q^{\text{ts}}) < 0$, the opposite happens: nonnative interactions lead to an increase in the barrier, decreasing protein folding rates.

Nonnative-contact calculation

Nonnative contacts for the C_α SBM simulations were defined according to the following criteria:

- 1) Nonnative contacts cannot be listed in the first native-contact map generated from the protein native configuration (see the [Supporting Material](#));
- 2) Nonnative contacts consist of any two nonbonded amino acid residues separated by at least four residues in the main chain and at a distance of up to 6 Å from each other;
- 3) Nonnative contacts cannot have a >30% probability of being found on folded structures ($Q > 0.9$), and if that is the case, these new contacts are included in the final native-contact list.

The difference in the number of nonnative contacts (ΔA) between the transition and the unfolded states is calculated by

$$\Delta A = \bar{A}(\text{TS}) - \bar{A}(\text{Unf}), \quad (3)$$

with $\bar{A}$ being the average of nonnative contacts formed in the transition state (TS) and in the unfolded state (Unf). The TS and the unfolded regions are defined using the free-energy profile as a function of the native contacts ($F(Q)$) for the respective protein (Fig. 1). The unfolded region is defined as the region that involves all states from $Q = 0$ to the configuration in which the protein reaches 15% of barrier height (ΔF) after the unfolded minimum. On the other hand, the transition region is defined as the region that involves all states located inside the free-energy barrier and having a free energy $F(Q) \geq 85\% \Delta F$. ΔF is calculated by the difference between the peak of the free-energy barrier ($F^* = F(Q^{\text{ts}})$) and its unfolded minimum ($F_{\text{unf}} = F(Q^{\text{u}})$), i.e., $\Delta F = F^* - F_{\text{unf}}$, as shown in Fig. 1. The values of 15% and 85%, described above, were chosen because they delimit the unfolded and folded states and allow the presence of a pretransition region for all proteins studied in this work, including those with low free-energy barriers.

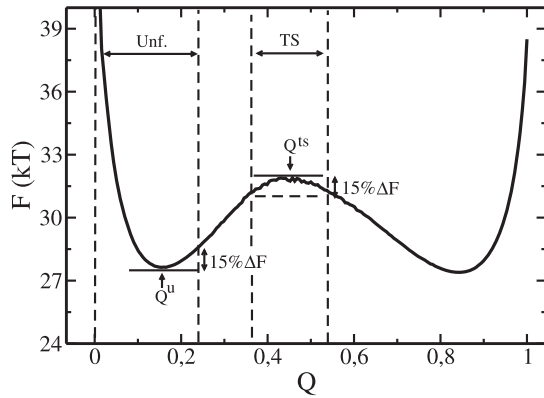


FIGURE 1 Free-energy profile (F) as a function of the native-contact fraction (Q) at the folding temperature (T_F) for a typical protein (Ubiquitin, PDB: 1UBQ). Q^u is the native-contact fraction that corresponds to the first minimum in the unfolded state and Q^{ts} corresponds to the transition state at the free energy barrier peak. The region between the first two vertical dashed lines is defined as the unfolded region (*Unf*). The region delimited by the third and fourth vertical dashed lines, which involves the free-energy barrier peak, is defined as the transition region (*TS*). ΔF is the variation in free energy between the unfolded and transition states ($\Delta F = F(Q^{ts}) - F(Q^u)$).

RESULTS AND DISCUSSION

Correlation between nonnative-contact variation and the free-energy barrier without energetic frustration

In accordance with the Clementi-Plotkin model (40) in Eq. 2, the change in folding rates is directly related to the difference in the number of nonnative contacts that are formed in the transition and unfolded states (ΔA). As described in the previous section, if ΔA is positive, the addition of an attractive interaction between nonnative contacts decreases the free-energy barrier and increases the folding rate. To test the correlation between the number of nonnative contacts and the height of the free-energy barrier without frustration (ΔF^0), the number of nonnative contacts formed during the folding process for 15 proteins was calculated (see Table 1).

TABLE 1 Data Obtained for the 15 Proteins Studied and Sorted by ΔF^0

Protein	PDB	Amino Acids	M	RCO	ΔF^0 ($k_B T$)	ΔA ($\times 10^{-3}$)
$\alpha_3 D$	2A3D	73	136	0.095	0.23	-5.44
PtABD	1BDC	60	102	0.086	0.36	-0.98
EnHD	1ENH	54	111	0.13	0.71	-9.81
IM9	1IMP	86	178	0.11	1.56	-4.44
HHCC	1HRC	104	246	0.11	1.79	-7.10
PtL	2PTL	60	136	0.18	2.13	2.05
ADA2h	1PBA	81	175	0.14	2.38	3.60
PtG	2K0P	56	139	0.17	2.95	4.17
CI2	1CIS	66	152	0.16	3.00	-1.25
SH3	1FMK	61	152	0.19	3.89	3.94
Ubiquitin	1UBQ	76	188	0.15	4.34	1.27
HPr	1HDN	85	222	0.18	5.06	1.30
CSPTm	1G6P	66	180	0.17	5.69	5.00
TWlg	1WIU	93	253	0.20	5.86	10.19
α AIT	2AIT	74	196	0.19	6.21	11.68

Data included in the table for the proteins shown are the Protein Data Bank (PDB) code, the number of amino acids, the number of native contacts (M), the relative contact order (RCO), the free-energy barrier without energetic frustration (ΔF^0), and the variation in the fraction of nonnative contacts (ΔA).

The average number of nonnative contacts formed during the folding process varies for each protein. For example, Fig. 2 shows the average number of nonnative contacts formed during the folding process for two proteins, EnHD (Fig. 2 a) and SH3 (Fig. 2 b). Fig. 2 shows that for both proteins, there is an increase in the formation of nonnative contacts up to the maximum value at the pretransition region (*P-TS*). After the maximum value, the number of nonnative contacts has a monotonic decrease. Beyond the small difference in the average number of nonnative contacts formed, these proteins can be differentiated by their values and signals of ΔA . The protein EnHD, and also the proteins $\alpha_3 D$, PtABD, EnHD, HHCC, IM9, and CI2, formed more nonnative contacts on average in the unfolded than in the transition region, i.e., $\Delta A < 0$. In this case, according to Eq. 2, it is expected that the addition of an energetic frustration in these proteins will be responsible for reducing the folding rates. On the other hand, for proteins PtL, ADA2h, PtG, SH3, Ubiquitin, HPr, TWlg, CSPTm, and α AIT, the formation of nonnative contacts in the transition state exceeds the

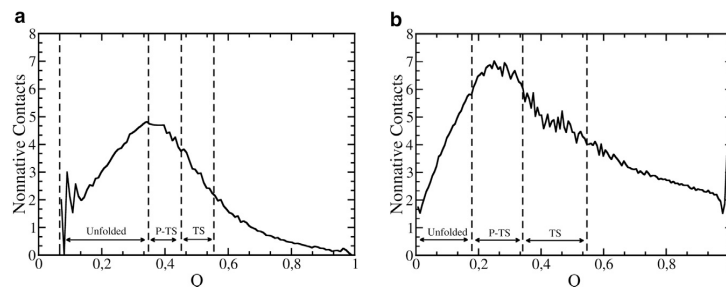


FIGURE 2 Average number of nonnative contacts ($\bar{A}$) as a function of the native-contact fraction (Q) at the folding transition temperature (T_F) for the proteins (a) EnHD and (b) SH3. Vertical dashed lines delimit the defined regions for the unfolded (*Unf*), transition (*TS*), and pretransition (*P-TS*) states. $\bar{A}$ reaches its maximum at the pre-transition state for both cases.

number formed in the unfolded state, and therefore, $\Delta A > 0$. ΔA values calculated for all 15 proteins are shown in Table 1.

The correlation between the free-energy barrier height (ΔF^0 , in kT units), without the energetic frustration term, and the variation in the number of nonnative contacts (ΔA) for the 15 proteins studied are shown in Fig. 3. The free-energy barrier correlates with the nonnative-contact variation by a factor of 0.81 (linear correlation) and it clusters the simulated proteins by their fold motifs. This result shows a strong correlation between these two variables and indicates that α -helix proteins make, on average, fewer nonnative contacts in the transition state than in the unfolded state. The opposite effect happens with proteins with β -sheets. This result appears not to be dependent on the chain length. The positive linear correlation coefficient indicates that proteins with a low variation in nonnative contacts (ΔA) have a low free-energy barrier and that the free-energy barrier increases with the increase in variation of nonnative contacts.

The product of ΔF^0 by native contact order was found to be the best parameter to correlate with optimum frustration values, according to the principal-component analysis shown by Contessoto et al. (46). It will be seen that the ΔA values for each protein are also strongly correlated with relative contact order ($\text{RCO} \times \Delta F^0$), as can be seen in Fig. S1. The linear correlation of 0.84 between these parameters indicates that free-energy barrier height multiplied by the contact order is a good parameter for predicting the effect of the energetic frustration on these proteins, as was suggested in the earlier study (46). The ΔF^0 and RCO values are shown in Table 1.

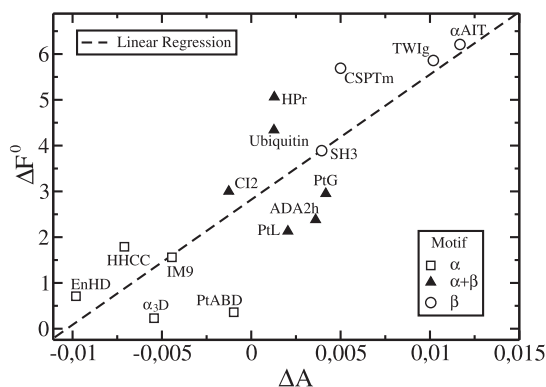


FIGURE 3 The free-energy barrier (ΔF^0) as a function of nonnative-contact-fraction variation (ΔA) for all proteins studied. Proteins are represented by their fold motif according to the SCOP database criterion (48): α (squares), β (circles), and $\alpha + \beta$ (triangles). The linear-fit correlation to the data is 0.81. ΔF^0 is strongly correlated to ΔA , and it clusters the simulated proteins by their motif in three distinct groups. The data were extracted from Table 1 using ΔF^0 (without energetic frustration).

Comparing analytical and computational results

In this section, the ΔA values and the variation in the free-energy barrier ($\Delta\Delta F = \Delta F^{\text{frus}} - \Delta F^0$) when the energetic frustration (ϵ_{NN}) is added to the C_α model are presented. The results were compared with those predicted by the Clementi-Plotkin model (Eq. 2) and the optimal energetic frustration from our previous study (46).

This study uses homogeneous energetic frustration with no sequence dependence. In our previous article (46), it was shown through principal-component analysis and partial least-squares analysis that the results seem to depend weakly on the frustration potential. In another study, a heterogeneous frustration potential was employed (9) in which energetic frustration was achieved by the addition of a random energy term characterized by a distribution width, b . An increase in folding temperature was observed for some values of b^2 and a decrease for high values. These results are analogous to the case of homogeneous frustration. More details regarding the potential used to add frustration to the C_α model can be seen in the Supporting Material. The ϵ_{NN} values used were 0.05, 0.10, and 0.20.

Nonnative-contact formation with energetic frustration

For all proteins studied here, an increase in the nonnative interaction leads to an increase in the formation of nonnative contacts during the folding process. This increase does not occur homogeneously in all ensembles; it becomes more evident in the unfolded and pretransition states. Therefore, depending on the value of ϵ_{NN} , a gradual increase in the amount of nonnative contacts is formed before the pretransition state and it can be higher than the average amount formed in the transition region. Thus, the gradual increase in the energetic frustration accounts for the reduction of ΔA values until it becomes negative, which, according to the Clementi-Plotkin model, makes the increase in frustration unfavorable to the folding process.

Fig. 4 shows the nonnative-contact variation (ΔA) as a function of the energetic frustration term (ϵ_{NN}) included in the simulation model. Each protein responds differently when a small quantity of frustration is added. Proteins formed exclusively by α -helices, such as EnHD and HHCC, shown in Fig. 4, already had a negative value of ΔA , even without the energetic frustration term, and continue to have more negative values of ΔA when ϵ_{NN} is incremented. For $\alpha + \beta$ proteins, which had, initially, a positive value of ΔA , an increase in the ϵ_{NN} values causes a gradual change of their ΔA values and signals. Ubiquitin and ADA2h had their signals of ΔA inverted when the frustration term was 0.05 and 0.1, respectively. These values of ϵ_{NN} , which inverted the signal of ΔA , correspond exactly to the optimal energetic frustration determined for these proteins by Contessoto et al. (46). Very similar behavior occurs with CSPTm and α AIT proteins formed mainly by β -sheets. Such proteins had their ΔA values decreased gradually, but only reached a

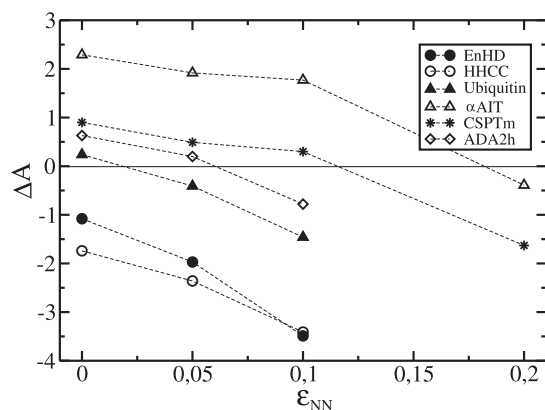


FIGURE 4 Variation in the formation of nonnative contacts (ΔA) according to the energetic frustration term (ϵ_{NN}) added in the simulation for six of our simulated proteins. Both axes are calculated at the respective folding temperature. The black line divides the positive and negative values of ΔA . The dotted lines connecting the points are a visual guide. ΔA is the difference between the numbers of nonnative contacts formed in the transition region and the unfolded region.

negative value at larger degrees of frustration ($\epsilon_{NN} = 0.2$). In all cases, when proteins formed more nonnative contacts in the unfolded state than in the transition state, ϵ_{NN} was close to the optimal energetic frustration value determined previously (46). In this case, when proteins are in their optimum frustration degree, they are at the same regime as those proteins in which the increase of frustration slows folding.

The vast majority of proteins have an unfolded dimension that scales with the protein chain length. An increase in the formation of nonnative contacts in the unfolded region can occur when an attractive energetic frustration is added to the simulation, and more unfolded compact structures can be formed, as shown in Fig. 4. Compaction of unfolded structures has been identified and studied computationally and experimentally in some proteins (52–56). This compaction is associated with the formation of hydrophobic cores or with the formation of medium/long-range contacts. These unfolded structural aspects seem to play an important role in the thermodynamics and kinetics of the folding process (52,57).

Changes in the free-energy barrier height

To compare the variation of the simulated energetic barrier with and without the frustration term, a value of energetic frustration close to the optimal values already determined (46) was used; $\epsilon_{NN} = 0.05$ for α -helix proteins and Ubiquitin, $\epsilon_{NN} = 0.1$ for PtL, PtG, ADA2h, CI2, SH3, and HPr; and $\epsilon_{NN} = 0.2$ for CSPTm, α AIT, and TWIg.

The comparison between the variation in the free-energy barrier height, $\Delta\Delta F^{\text{sim}}$, obtained by the simulations, and $\Delta\Delta F^{\text{theory}}$, given by Eq. 2, is shown in Fig. 5. $\Delta\Delta F^{\text{theory}}$ is calculated using the M , T_F^0 , and ΔA parameters calculated by simulation. Fig. 5 shows that the simulated results and

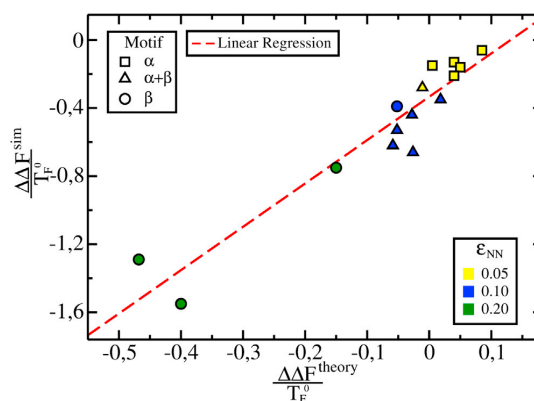


FIGURE 5 Analytical and simulated free-energy barrier variations when energetic frustration is added ($\Delta\Delta F = \Delta F^{\text{frus}} - \Delta F^0$). Each protein is represented by its respective motif according to the SCOP database criterion (48)— α (squares), β (circles), and $\alpha + \beta$ (triangles)—and colored by the energetic frustration (ϵ_{NN}) added to the simulation: 0.05, 0.1, and 0.2. Both axes are normalized by the respective folding temperature in the absence of ϵ_{NN} (T_F^0). The dotted line is the linear regression of the two data sets. The values of $\Delta\Delta F^{\text{sim}}$ are obtained by simulation using the weighted histogram analysis method with Q (the fraction of native contacts) as an order parameter. The values of $\Delta\Delta F^{\text{theory}}$ are obtained using Eq. 2 as reference with $b = 0$ and ΔA calculated by the simulation of each protein. The linear correlation to the data is 0.95. To see this figure in color, go online.

the analytical model predictions are strongly correlated, with linear fit correlation 0.95. Thus, when the proteins are folded with a small degree of frustration close to their optimal value, the computational and the analytical result have the same behavior. Also, Fig. 5 shows the results clustered according to their folding motifs; β proteins have lower $\Delta\Delta F$ and high ϵ_{NN} , $\alpha + \beta$ proteins have intermediate $\Delta\Delta F$ and ϵ_{NN} , whereas α -proteins have high $\Delta\Delta F$ and low ϵ_{NN} , which is in agreement with the analytical model (Eq. 2) and the results from Fig. 3.

CONCLUSIONS

The difference in the number of nonnative contacts formed between the unfolded and the transition states showed a strong correlation with the height of the free-energy barrier, since proteins with a low free-energy barrier were those with fewer nonnative contacts formed in the transition state than in the unfolded state. In addition, the results suggest that proteins formed mainly by α -helix structures form more nonnative contacts in the unfolded state, in such a way that the ratio between the numbers of nonnative contacts formed in the transition state makes the energetic frustration unfavorable for them.

Energetic frustration, as it has been modeled in this study, proved to be responsible for increasing the formation of nonnative contacts, especially in the unfolded state. Frustration has also proved to reach a maximum threshold as the

ΔA values is decreased, leading proteins to form fewer nonnative contacts in the transition state than in the unfolded state. The energetic frustration value, which changes the sign of ΔA for almost all proteins tested, was close to the optimal frustration determined previously by Contessoto et al. (46). This corroborates the results expected by the theory developed by Clementi and Plotkin and makes a connection with the former computational analysis. Despite all the approximations that were made for the simulation and the analytical models, both resulted in agreement in 95% of the cases, as can be seen in Fig. 5.

Although it appears to be controversial that the presence of nonnative contacts may favor the folding process, this can be explained mainly by the fact that when inserting an energetic frustration, we add an attractive interaction energy in the midst of the contacts, which according to the results obtained, leads to an increase in the nonnative collapse. The increase in the nonnative collapse may favor access to other structures, which may be responsible for bringing amino acids close to each other, which in turn favors the formation of native contacts, accelerating the kinetics of the process. One question that arises from these studies is about the nature of the changes and stabilizations observed when a nonnative potential is introduced. Such a stabilization may originate with entropic or enthalpic factors. In preliminary studies, we observed variations dependent on the protein motifs. Further studies should not only investigate different motifs, but also analyze such effects in terms of their dependence on the type of frustration used.

SUPPORTING MATERIAL

Supporting Materials and Methods and one figure are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(16\)30388-5](http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30388-5).

AUTHOR CONTRIBUTIONS

P.R.M., J.C., and V.B.P.L. designed research. P.R.M. performed research. P.R.M. and V.G.C. contributed analytical tools. Data were analyzed by P.R.M., V.G.C., R.J.O., J.C., and V.B.P.L. All authors contributed to the manuscript writing.

ACKNOWLEDGMENTS

P.R.M. and V.G.C. were supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. R.J.O. was funded by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). V.B.P.L. was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2014/06862-7). R.J.O., J.C., and V.B.P.L. were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This research was supported by resources supplied by the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP).

SUPPORTING CITATIONS

References (58–63) appear in the Supporting Material.

REFERENCES

1. Dobson, C. M. 2003. Protein folding and misfolding. *Nature*. 426:884–890.
2. Baldwin, R. L. 1995. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR*. 5:103–109.
3. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.
4. Pande, V. S., AYU Grosberg, and T. Tanaka. 1997. On the theory of folding kinetics for short proteins. *Fold. Des.* 2:109–114.
5. Onuchic, J. N., H. Nymeyer, ..., N. D. Socci. 2000. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* 53:87–152.
6. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA*. 89:8721–8725.
7. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science*. 254:1598–1603.
8. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science*. 267:1619–1620.
9. Wang, J., R. J. Oliveira, ..., V. B. P. Leite. 2012. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. USA*. 109:15763–15768.
10. Nymeyer, H., A. E. García, and J. N. Onuchic. 1998. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA*. 95:5921–5928.
11. Shoemaker, B. A., J. Wang, and P. G. Wolynes. 1997. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. USA*. 94:777–782.
12. Whitford, P. C., J. K. Noel, ..., J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins*. 75:430–441.
13. Oliveira, R. J., P. C. Whitford, ..., J. Wang. 2010. Coordinate and time-dependent diffusion dynamics in protein folding. *Methods*. 52:91–98.
14. Fersht, A. R. 1995. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* 5:79–84.
15. Garcia-Mira, M. M., M. Sadqi, ..., V. Muñoz. 2002. Experimental identification of downhill protein folding. *Science*. 298:2191–2195.
16. Nettels, D., I. V. Gopich, ..., B. Schuler. 2007. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA*. 104:2655–2660.
17. Chung, H. S., J. M. Louis, and W. A. Eaton. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA*. 106:11837–11844.
18. Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.
19. Chavez, L. L., J. N. Onuchic, and C. Clementi. 2004. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* 126:8426–8432.
20. Snow, C. D., E. J. Sorin, ..., V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69.
21. Gosavi, S., L. L. Chavez, ..., J. N. Onuchic. 2006. Topological frustration and the folding of interleukin-1 β . *J. Mol. Biol.* 357:986–996.
22. Chu, X., L. Gan, ..., J. Wang. 2013. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. USA*. 110:E2342–E2351.
23. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*. 84:7524–7528.
24. Fersht, A. R. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7:3–9.

25. Chahine, J., R. J. Oliveira, ..., J. Wang. 2007. Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proc. Natl. Acad. Sci. USA*. 104:14646–14651.
26. Oliveira, R. J., P. C. Whitford, ..., V. B. Leite. 2010. The origin of non-monotonic complex behavior and the effects of nonnative interactions on the diffusive properties of protein folding. *Biophys. J.* 99:600–608.
27. Xu, W., Z. Lai, ..., J. Wang. 2012. Configuration-dependent diffusion dynamics of downhill and two-state protein folding. *J. Phys. Chem. B*. 116:5152–5159.
28. Oliveira, A. B., Jr., F. M. Fatore, ..., V. B. P. Leite. 2014. Visualization of protein folding funnels in lattice models. *PLoS One*. 9:e100861.
29. Onuchic, J. N., P. G. Wolynes, ..., N. D. Socci. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA*. 92:3626–3630.
30. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
31. Eaton, W. A., P. A. Thompson, ..., J. Hofrichter. 1996. Fast events in protein folding. *Structure*. 4:1133–1139.
32. Eaton, W. A., V. Muñoz, ..., J. Hofrichter. 1997. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7:10–14.
33. Ozkan, S. B., I. Bahar, and K. A. Dill. 2001. Transition states and the meaning of Φ -values in protein folding kinetics. *Nat. Struct. Biol.* 8:765–769.
34. Shakhnovich, E. I., and A. M. Gutin. 1989. Formation of unique structure in polypeptide chains. Theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34:187–199.
35. Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.
36. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes. 1992. Optimal protein-folding codes from spin-glass theory. *Proc. Natl. Acad. Sci. USA*. 89:4918–4922.
37. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.* 101:6052–6062.
38. Bryngelson, J. D., J. N. Onuchic, ..., P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*. 21:167–195.
39. Oliveira, L. C., R. T. H. Silva, ..., J. Chahine. 2006. Frustration and hydrophobicity interplay in protein folding and protein evolution. *J. Chem. Phys.* 125:084904.
40. Clementi, C., and S. S. Plotkin. 2004. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* 13:1750–1766.
41. Plotkin, S. S. 2001. Speeding protein folding beyond the G(o) model: how a little frustration sometimes helps. *Proteins*. 45:337–345.
42. Plotkin, S. S., and J. N. Onuchic. 2002. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q. Rev. Biophys.* 35:111–167.
43. Zarrine-Afsar, A., S. Wallin, ..., H. S. Chan. 2008. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. USA*. 105:9999–10004.
44. Zarrine-Afsar, A., Z. Zhang, ..., H. S. Chan. 2012. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins*. 80:858–870.
45. Zhang, Z., and H. S. Chan. 2010. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc. Natl. Acad. Sci. USA*. 107:2920–2925.
46. Contessoto, V. G., D. T. Lima, ..., V. B. P. Leite. 2013. Analyzing the effect of homogeneous frustration in protein folding. *Proteins*. 81:1727–1737.
47. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
48. Murzin, A. G., S. E. Brenner, ..., C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.
49. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
50. Clementi, C., A. E. García, and J. N. Onuchic. 2003. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J. Mol. Biol.* 326:933–954.
51. de Mendonça, M. R., L. G. Rizzi, ..., N. A. Alves. 2014. Inferring a weighted elastic network from partial unfolding with coarse-grained simulations. *Proteins*. 82:119–129.
52. Kohn, J. E., I. S. Millett, ..., K. W. Plaxco. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA*. 101:12491–12496.
53. Kazmirski, S. L., K.-B. Wong, ..., V. Daggett. 2001. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. USA*. 98:4349–4354.
54. Garcia, P., L. Serrano, ..., M. Bruix. 2001. NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci.* 10:1100–1112.
55. Hodsdon, M. E., and C. Frieden. 2001. Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. *Biochemistry*. 40:732–742.
56. Meng, W., N. Lyle, ..., R. V. Pappu. 2013. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc. Natl. Acad. Sci. USA*. 110:2123–2128.
57. Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10:27–34.
58. Berman, H. M., J. Westbrook, ..., P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.
59. Sobolev, V., R. C. Wade, ..., M. Edelman. 1996. Molecular docking using surface complementarity. *Proteins*. 25:120–129.
60. Noel, J. K., P. C. Whitford, ..., J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38:W657–W661.
61. Van Der Spoel, D., E. Lindahl, ..., H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
62. Ferrenberg, A. M., and R. H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638.
63. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, 11 / 03 / 2019

Paulo Ricardo Moura

Assinatura do autor