

Introdução ao problema de enovelamento de proteínas: uma abordagem utilizando modelos computacionais simplificados

Introduction to the Protein Folding Problem: An Approach Using Simplified Computational Models

Vinícius de Godoi Contessoto^{*1,2,3}, Antonio Bento de Oliveira Junior², Jorge Chahine², Ronaldo Junio de Oliveira⁴, Vitor Barbanti Pereira Leite²

¹Centro Nacional de Pesquisa em Energia e Materiais, Laboratório Nacional de Ciência e Tecnologia do Bioetanol, Campinas, SP, Brasil

²Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, Departamento de Física, São José do Rio Preto, SP, Brasil

³Rice University, Bioscience Research Collaborative, Center for Theoretical Biological Physics, Houston, TX, USA

⁴Universidade Federal do Triângulo Mineiro, Instituto de Ciências Exatas, Naturais e Educação, Departamento de Física, Laboratório de Biofísica Teórica, Uberaba, MG, Brasil

Recebido em 06 de Março, 2018. Revisado em 11 de Abril, 2018. Aceito em 17 de Abril, 2018.

Este trabalho tem como objetivo introduzir estudantes de física e áreas afins ao problema de enovelamento de proteína. A proteína é uma macromolécula biológica de grande importância para os seres vivos. Entender como esta macromolécula encontra uma configuração tridimensional e funcional, tomando como a priori a informação presente em uma sequência linear de diferentes aminoácidos, é uma questão relevante e envolve pesquisadores de diversas áreas do conhecimento. Neste trabalho, é apresentada uma introdução sobre a importância das proteínas e suas principais características, além de mostrar um breve histórico do estudo de enovelamento de proteínas e como esses trabalhos convergem na construção da teoria de Superfície de Energia. São descritos alguns modelos teóricos e computacionais, desde as primeiras abordagens em modelos analíticos até a construção de modelos baseados em estruturas para simulações. Também são apresentados e discutidos alguns resultados esperados em cada modelo utilizado.

Palavras-chave: Enovelamento de proteínas, Superfície de energia, Modelos baseados em estrutura, Modelo de rede, Biofísica.

This work aims to introduce students of physics and related areas to the problem of protein folding. Proteins are biological macromolecules of great importance to living things. The understanding of how this macromolecule finds a three-dimensional and functional structure, using the information present in a linear sequence of different amino acids, is a relevant question which involves researchers from different areas of knowledge. This work discusses the importance and the main properties of proteins and presents a brief history of protein folding studies and how these works converge to the Energy Landscape theory. There are described some theoretical and computational models from analytical approaches to the development of the Structure-Based Models for simulations. Here it is also presented and discussed some expected results for each used model.

Keywords: Protein Folding, Energy Landscape, Structure-Based Models, Lattice Model, Biophysics.

1. Introdução: aspectos biofísicos das proteínas

A compreensão dos mecanismos biomoleculares tem proporcionado um grande desafio para os cientistas de diversas áreas do conhecimento. Uma das classes de macromoléculas biológicas de grande importância para os seres vivos são as proteínas. Elas constituem a maior parte da massa celular seca e são responsáveis por inúmeras funções em meio fisiológico. A construção celular, a re-

gulação de entrada e de saída de íons, o transporte e a recepção de sinais, são exemplos de algumas funções realizadas pelas proteínas [1]. Algumas doenças estão diretamente relacionadas com o enovelamento defeituoso das proteínas e na falha dos mecanismos de controle de qualidade da célula [2], como por exemplo, o mal de Alzheimer, a doença de Huntington, a encefalopatia espongiforme (doença da vaca louca) e o diabetes tipo II. A grande importância das proteínas para a maquinaria biológica e para a fisiologia dos organismos vivos torna esta classe de macromoléculas um alvo interessante de estudo.

*Endereço de correspondência: vinicius.contessoto@unesp.br.

De uma maneira simplificada, a proteína pode ser associada à um heteropolímero linear em que cada monômero de sua cadeia é constituído por um resíduo de aminoácido. Existem, em geral, 20 diferentes tipos de aminoácidos que podem compor a cadeia proteica. Cada aminoácido possui um carbono na cadeia principal denominado carbono α (C_α). Este carbono realiza quatro ligações covalentes; uma com um átomo de hidrogênio, outra com um grupo amina, outra com um grupo carboxila, e outra com um grupo radical (também conhecido como a cadeia lateral do aminoácido). O grupo radical diferencia todos os aminoácidos encontrados na natureza quanto à estrutura, tamanho, carga elétrica e hidrofobicidade. Cada aminoácido é ligado ao seu vizinho por uma ligação covalente, e uma sequência com as combinações dos diferentes aminoácidos formam cada proteína [3, 4].

Existem quatro níveis de organização que descrevem estruturalmente as proteínas que são apresentados na figura 1A e 1B pela proteína Ubiquitina [5]. A estrutura primária consiste na representação linear da sequência de resíduos de aminoácidos simbolizadas nas diferentes letras na figura 1B. A estrutura secundária se refere às

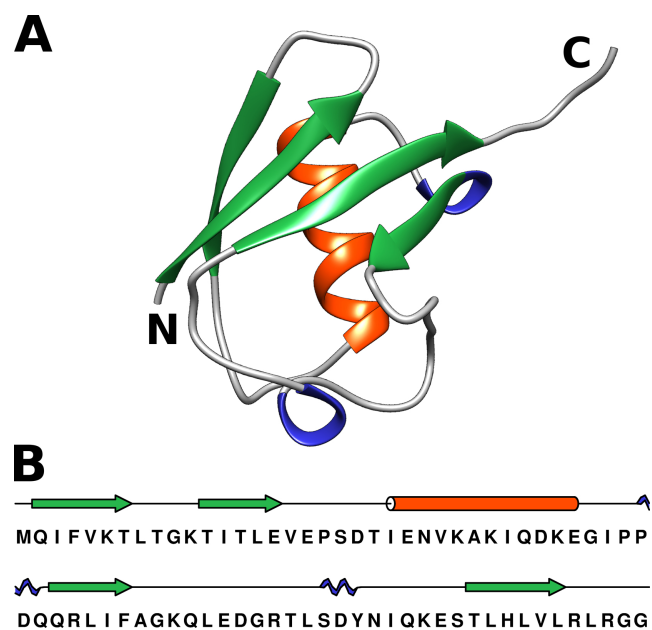


Figura 1: A) Representação estrutural do estado nativo da proteína Ubiquitina. A informação sobre a estrutura é obtida pelo PDB (*Protein Data Bank*, www.rcsb.org) com o código de identificação 1UBQ [5]. As letras N e C indicam as terminações da proteína com início no N-terminal e final no C-terminal. As cores indicam os diferentes tipos de estruturas secundárias. Em laranja é representada a estrutura de hélice α , em verde a estrutura de folhas β , em azul as voltas ou *loops* e em cinza claro as regiões sem estruturas definidas, ditas como aleatórias ou *random coil*. B) Representação da estrutura primária da proteína Ubiquitina. As cores e as figuras geométricas indicam a representação unidimensional das estruturas secundárias. As diferentes letras indicam os diferentes tipos de resíduos de aminoácidos que compõem essa proteína.

estruturas locais da proteína, como por exemplo, hélice α e folhas β (em laranja e em verde, respectivamente, na figura 1A). A estrutura terciária representa o arranjo tridimensional de todos os átomos em suas estruturas secundárias, como representado na figura 1A. A estrutura quaternária é o arranjo espacial de subunidades de uma proteína formando um complexo proteico [1, 4].

Cada proteína possui uma sequência unidimensional particular de resíduos de aminoácidos em sua formação. O processo em que esta cadeia linear encontra seu estado tridimensional nativo e funcional é chamado de enovelamento ou dobramento de proteínas. Os estudos sobre o problema do enovelamento procuram entender como a informação presente na sequência unidimensional de resíduos de aminoácidos faz com que a proteína encontre sua conformação nativa. A compreensão das interações envolvidas e dos princípios físico-químicos envolvidos no processo de enovelamento é um grande desafio da ciência atual e envolve pesquisadores de diversas áreas do conhecimento.

2. Perspectiva histórica

O início dos estudos sobre o problema do enovelamento de proteína ocorreu na década de 50 com a determinação das primeiras estruturas tridimensionais de proteínas por meio de técnicas de cristalografia e de difração de raios-X [6]. Na década de 60, os estudos de Anfinsen e colaboradores tiveram um papel fundamental para o estudo do enovelamento proteico quando mediram, por meio de reações físico-químicas, a atividade catalítica da enzima ribonuclease bovina [7].

Os experimentos demonstraram que a ribonuclease, depois de desenovelada quimicamente pela adição de um agente desnaturante, poderia se enovelar espontaneamente após serem reestabelecidas as condições fisiológicas (retirada do agente desnaturante) e, desta forma, retornar ao seu estado nativo. Anfinsen concluiu neste estudo que a sequência de resíduos de aminoácidos contém a informação necessária para caracterizar o estado nativo de uma proteína e guiar seu processo de enovelamento [8].

Com os estudos sobre a desnaturação da ribonuclease, Anfinsen também estabeleceu a qual foi chamada de “Hipótese Termodinâmica”. Esta hipótese sugere que a proteína no seu estado nativo deve possuir o valor mínimo de energia livre de Gibbs [8]. Assim, o estudo do enovelamento de proteínas começou a ser visto, não apenas como uma questão biológica, mas como uma questão de minimização de energia livre.

Abordando a questão cinética do enovelamento, Levinthal argumentou ser inviável uma proteína vasculhar todo o seu espaço conformacional em busca do menor valor de energia livre para se enovelar, e encontrar seu estado nativo em um curto intervalo de tempo para exercer sua função biológica [9]. Para exemplificar essa afirmação, pode-se tomar uma proteína com N resíduos de aminoácidos que podem assumir μ orientações no espaço.

Desta forma, o número de conformações distintas para uma proteína seria μ^N . Para uma estimativa simples, considera-se que cada resíduo de aminoácido possua duas possibilidades de orientação. O número de conformações possíveis para uma proteína com 100 resíduos de aminoácidos seria igual a $2^{100} \approx 10^{30}$ conformações. Supondo que um picossegundo seja requerido para acessar uma conformação, o tempo necessário para explorar todas as 10^{30} conformações seria de aproximadamente 10^{18} segundos. Este tempo equivale a mais de 10^{10} anos, um tempo com ordem de grandeza da idade do universo.

Desta forma, uma proteína não poderia se enovelar a tempo suficiente de encontrar seu mínimo global de energia, vasculhando todas as conformações possíveis, visto que, o tempo característico de enovelamento de uma proteína típica não leva mais do que alguns segundos. Esta contradição ficou conhecida como o “paradoxo de Levinthal” [9–11].

Uma possível explicação para este paradoxo seria que a conformação final da proteína deveria ser alcançada por um caminho cinético preferencial, e o mecanismo de enovelamento seria dependente de um caminho específico. Uma sequência bem definida de eventos guiaria a proteína a partir de um estado aleatório desenovelado para seu estado nativo [9, 11]. Com isso, entre as décadas de 70 e 80, os estudos foram direcionados na procura de estados intermediários que identificassem a existência de um caminho preferencial de enovelamento [12–14].

Um ponto de vista alternativo foi introduzido no final da década de 80 [15, 16]. Nesta abordagem, é sugerida uma visão global sobre a superfície de energia para se entender os processos que governam os mecanismos de enovelamento de uma proteína. O enovelamento de proteína seria organizado por um conjunto de estruturas similares pela superfície de energia, ao invés da existência de um rota preferencial passando por estados intermediários sucessivos. Sendo assim, uma abordagem estatística do processo de enovelamento pode ser empregada para descrever a superfície de energia da proteína.

3. A teoria de superfície de energia do enovelamento de proteínas

A teoria de superfície de energia do enovelamento de proteínas foi introduzida por Onuchic e Wolynes e desenvolvida também por outros pesquisadores durante as últimas três décadas [17–21]. Esta teoria procura descrever de maneira simplificada os princípios gerais que governam os mecanismos de enovelamento [22–26], utilizando os princípios da mecânica estatística [27–30]. Uma classe de sistemas bem estudados em mecânica estatística são os vidros de spins, que são sistemas de complexas interações e que ocorrem em um espaço de alta dimensionalidade [31, 32]. Estes serviram de base para o entendimento do problema de enovelamento de proteínas. Um dos conceitos subjacentes a estes sistemas é o de frustração, que pode ser entendido como a impossibilidade de

um sistema em satisfazer todos os vínculos do problema ao mesmo tempo. Sistemas magnéticos e heteropolímeros aleatórios são exemplos de vidro de spin com interações conflitantes que geram múltiplos estados de energia mínima local e que apresentam superfície de energia rugosa. Eles são caracterizados por uma temperatura de transição, chamada temperatura de *glass* (T_g), que está associada à rugosidade da superfície de energia (ΔE). Abaixo de T_g , o sistema fica armadilhado em mínimos locais e tem uma dinâmica lenta, chamada de vidro.

Ao relacionar as proteínas a sistemas complexos como os vidros de spin, vemos que elas também possuem uma superfície de energia rugosa. No entanto, ao contrário de heteropolímeros aleatórios, as proteínas encontradas na natureza não são aleatórias, ou seja, elas possuem sequências e estruturas nativas associadas às suas funções, que foram selecionadas ao longo da evolução. As proteínas seguem o Princípio da Frustração Mínima [15], em que evoluíram para sistemas minimamente frustrados, que em concordância com a hipótese de Anfinsen, garante que a energia do estado nativo é o mínimo global de energia [16, 28, 30, 33–35]. Seguindo este princípio, as proteínas conseguem satisfazer a maioria de suas interações energéticas simultaneamente, o que facilita atingir seu estado nativo. Uma maneira de analisar a complexa superfície intrinsecamente multidimensional do enovelamento é projetá-la em um único parâmetro de ordem ou coordenada de reação. A coordenada de reação é a grandeza que descreve o processo ao longo da superfície de energia [36] e a sua escolha é fundamental para uma boa abordagem do problema estudado [37]. Dado que a energia do estado nativo corresponde ao mínimo global de energia, este estado é um bom referencial para compará-lo com outras conformações. O grau de similaridade estrutural de uma dada conformação é uma boa medida da sua energia potencial, sendo portanto um parâmetro de ordem natural para a descrição do enovelamento. Por esta razão, uma coordenada de reação amplamente utilizada é a fração de contatos nativos (Q). Esta coordenada é a razão entre o número de contatos nativos formados em uma dada conformação (D) e o número de contatos nativos presentes na estrutura enovelada (N). Os valores de Q variam entre 0 e 1, onde 0 significa que a proteína está totalmente desenovelada e 1 significa que a proteína está no seu estado nativo. A coordenada de reação Q é baseada na estrutura do estado nativo da proteína e descreve de maneira satisfatória o processo de enovelamento [36]. A superfície resultante efetiva de energia apresenta uma topografia afunilada com um gradiente de energia direcionado para a região do estado nativo da proteína, conhecida como funil de enovelamento (figura 2A) [21, 38–40]. Nesta abordagem, o mecanismo de enovelamento pode ser entendido como um processo de difusão ao longo da superfície de energia quando se acompanha a coordenada de reação [41–43]. A figura 2A apresenta a coordenada de reação Q como o parâmetro de ordem para se acompanhar o enovelamento.

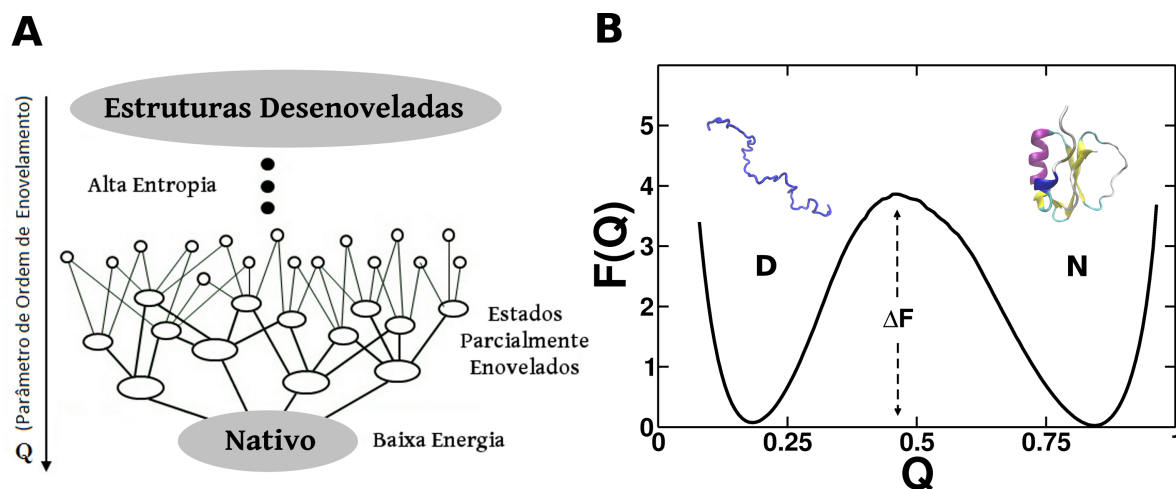


Figura 2: A) Ilustração do funil de estruturas do processo de enovelamento de uma proteína ao longo da coordenada de reação que é a fração de contatos nativos (Q). B) Representação do perfil de energia livre em função da coordenada de reação ($F(Q)$). São destacados os estados desenovelado (D) e nativo (N) e a barreira de energia no estado de transição (ΔF) existente entre D e N . A curva $F(Q)$ foi obtida por meio da simulação computacional do modelo C_α para a proteína CI2 (*Chymotrypsin Inhibitor 2*) com o código PDB 1YPA [44]. A obtenção de F é descrita com mais detalhes na seção de modelos baseados em estrutura em [45]. A figura A) é adaptada de [18, 30].

No topo, estão descritas as estruturas desenoveladas de alta entropia. O conjunto de estruturas similares são agrupadas ao longo do processo de enovelamento onde a proteína passa por estados parcialmente enovelados e quantidades de estruturas cada vez menores [20, 26, 30, 36]. Este processo ocorre até que a proteína atinja seu mínimo de energia na parte de baixo do funil de estruturas, atingindo seu estado nativo. Pela descrição do funil de estruturas não há um caminho preferencial no processo de enovelamento, mas uma multiplicidade de possíveis rotas [18]. Durante o processo de enovelamento existe uma competição entre a contribuição energética ($E(Q)$) e a contribuição entrópica ($S(Q)$), que são dependentes da coordenada de reação Q . O processo pode ser mapeado por um potencial termodinâmico efetivo que leva em consideração a competição entre essas duas grandezas. O potencial termodinâmico dado pela energia livre de Helmholtz ($F(Q)$) é utilizado para descrever esta competição, e é dado pela equação

$$F(Q) = E(Q) - TS(Q) \quad (1)$$

com T sendo a temperatura do sistema.

A figura 2B apresenta um perfil de energia livre ($F(Q)$) construído em função da coordenada de reação Q para uma proteína típica com mecanismo de enovelamento de dois estados. A curva de energia livre destaca três regiões principais: o estado desenovelado D , o estado nativo N e a barreira de energia ΔF que existe entre os dois estados. Na figura 2B, F está na temperatura de transição conhecida como temperatura de enovelamento (T_f) na qual a probabilidade da proteína se encontrar no estado desenovelado é a mesma probabilidade da proteína se encontrar no estado nativo e, por isso, os dois mínimos de F em D e N se coincidem. Experimentalmente, T_f é

conhecida como a temperatura de *melting* (T_m) e pode ser obtida por experimentos de calorimetria através de uma curva de desnaturação térmica ou de calor específico [46–48]. T_f está associada à estabilidade do estado nativo da proteína e é proporcional à variação de energia do estado nativo (E_{nat}) e da energia média do sistema (\bar{E}),

$$\delta E = \bar{E} - E_{nat} \quad (2)$$

Por outro lado, a superfície de energia das proteínas apresenta uma rugosidade intrínseca associada à frustração [49–55], que, como no caso de vidros de spin, é caracterizada por ΔE . Enquanto ΔE fica associado à propensão do sistema ficar armadilhado em mínimos locais, δE fornece um gradiente de energia na direção do mínimo global associado ao estado nativo como ilustrado na figura 3. Um critério de estabilidade de uma proteína está relacionado como a razão $\delta E/\Delta E$, que é proporcional à razão T_f/T_g . [15, 16]. Na temperatura de vidro, a proteína se apresenta armadilhada em mínimos locais da superfície, não possuindo energia suficiente para acessar diferentes configurações e chegar ao estado nativo. Quanto maior for a razão T_f/T_g , maior a possibilidade da proteína ter um enovelamento estável, sem ficar armadilhada em mínimos locais. Sendo assim, as condições ideais para o enovelamento correspondem às temperaturas na região $T_g < T < T_f$.

Com base no princípio de frustração mínima, foram desenvolvidos alguns modelos computacionais para o estudo do enovelamento de proteínas. Estes modelos, apesar de simples, possuem algumas características fundamentais para o estudo como, por exemplo, a conectividade da cadeia polipeptídica, a flexibilidade, o volume excluído, a dependência com a sequência dos resíduos de aminoácidos e as interações inter- e intra-cadeia [56–58]. Tais modelos

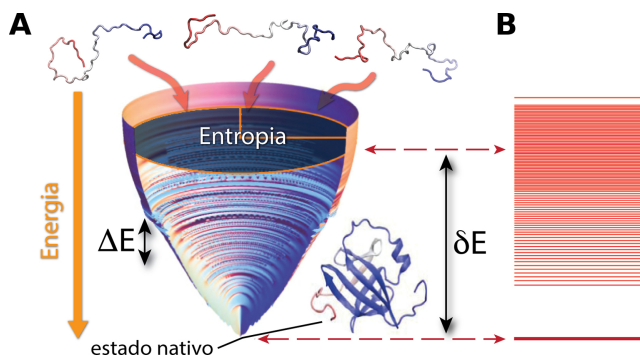


Figura 3: A) Superfície de energia com topografia afunilada. A área do funil da seção horizontal representa a entropia configuracional que está relacionada ao número de configurações para cada nível do funil. O eixo vertical representa a energia total associada ao funil de energia e também a uma coordenada de reação que organiza em cada nível o grau de similaridade com o estado nativo (Q). A construção desta superfície é feita a partir de modelos computacionais e traz informações sobre o enovelamento da proteína P13, código PDB 1QTU. B) Diagrama de níveis de energia para os diferentes estados de uma proteína ao longo de seu enovelamento. Em destaque está o estado nativo com energia mínima afastada dos demais estados não enovelados. Em A) e B), ΔE representa a flutuação energética ao longo da superfície e δE representa o intervalo energético entre a média do estado desordenado e o estado nativo. Figura adaptada de [21].

são capazes de reproduzir aspectos característicos do enovelamento como, por exemplo, o tempo de enovelamento, a identificação dos caminhos que levam à conformação nativa, a descrições de propriedades cinéticas e termodinâmicas do enovelamento e características estruturais do estado de transição por análise de valores- Φ [59–63].

4. Modelos simplificados de proteína

Nesta seção são discutidos alguns aspectos sobre a abordagem teórica utilizada no estudo de enovelamento de proteínas. Esta abordagem procura fazer uso de modelos simplificados com intuito de entender os princípios gerais que envolvem o mecanismo de enovelamento. Em alguns casos, o enovelamento pode ser entendido como uma transição de fase de primeira ordem [30,64], em que o estado enovelado (ou desenovelado) é mais estável de acordo com os parâmetros do sistema, tais como temperatura, sequência primária e hidrofobicidade. Esta abordagem sugere uma analogia com outros sistemas físicos como os vidros de spin [15,31,65].

A escolha do modelo utilizado está diretamente relacionada com o tipo de pergunta que se espera responder. Uma maior quantidade de detalhes envolvidos na construção dos modelos podem levar a um aumento no custo computacional, o que pode ser desnecessário para abordar determinadas questões. Por outro lado, a escolha de um modelo com muitas simplificações pode não ser suficiente para explicar alguns aspectos observáveis ex-

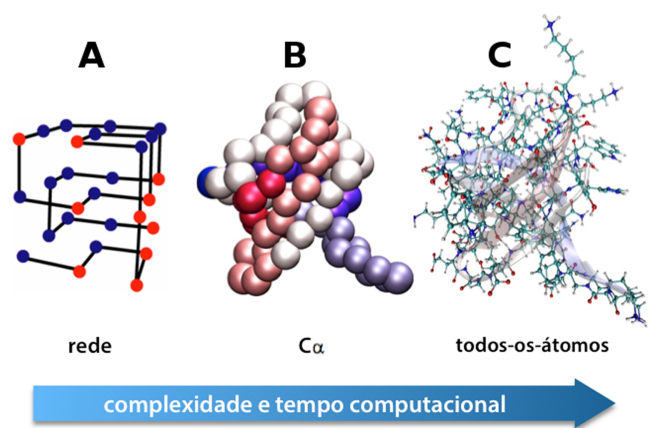


Figura 4: Representação de uma proteína para diferentes abordagens computacionais para o estudo do enovelamento empregando modelos simplificados. A) Modelo de rede cúbica com as cores podendo representar dois tipos de resíduos: em vermelho, os resíduos polares e em azul, os resíduos hidrofóbicos. B) Modelo baseado na estrutura da proteína representado somente com os carbonos alfa (C_α). C) Modelo baseado na estrutura com todos-os-átomos pesados (sem hidrogênios). A complexidade e o tempo computacional aumentam de A) para C) conforme aumentam os detalhes de cada modelo.

perimentalmente (figura 4). Sendo assim, a escolha do modelo está diretamente relacionada aos detalhes e características da questão investigada. Detalhes de alguns modelos computacionais para o estudo do enovelamento de proteínas são descritos nas próximas seções.

4.1. Modelo de rede

No modelo de rede (do inglês *lattice models*), a proteína é representada por um heteropolímero composto por monômeros dispostos em uma rede de duas ou de três dimensões. Os monômeros representam os resíduos de aminoácidos e a ligação entre eles representam uma ligação covalente. Cada monômero pode ocupar somente um sítio na rede (condição de volume excluído) [17]. Nesta seção será abordado com maior detalhe o modelo de rede cúbica com 27 monômeros (figura 4A).

A estrutura maximamente compacta (enovelada) é um cubo $3 \times 3 \times 3$, com um número máximo de contatos igual a 28. Neste modelo, um contato é definido quando dois monômeros estão na distância de primeiro vizinho e não estão ligados covalentemente.

Shakhnovich e Gutin [66] enumeraram todas as conformações maximamente compactas para um heteropolímero com 27 monômeros, encontrando um número total de 103346 diferentes cubos. Esta quantidade de diferentes estados é uma vantagem para a utilização deste modelo, pois é possível gerar todos os diferentes cubos com um baixo custo computacional quando comparado com outros modelos mais complexos.

Para representar uma proteína, este heteropolímero deve conter um estado maximamente compacto, de me-

nor energia, não degenerado e cineticamente acessível, características semelhantes às de uma proteína globular típica. Para isso, é escolhido um potencial de interação entre os monômeros que favorece essas características, descrito como

$$E = \sum_{i=1}^N \sum_{j>i+1}^N H_{i,j}. \quad (3)$$

Na equação 3, o valor da energia E é dado pela soma das interações ($H_{i,j}$) entre todos os pares de monômeros i e j que sejam considerados primeiros vizinhos na rede, e que não estejam ligados de forma covalente, ou seja, $|i - j| \neq 1$.

O número de tipos de monômeros é variável, dependendo da pergunta. No caso de um homopolímero, existe apenas um tipo de monômero/interação, e o sistema apresenta configurações degeneradas em energia. A heterogeneidade mínima necessária para gerar estados nativos não-degenerados é o modelo com dois tipos de monômeros, também conhecido como modelo hidrofóbico-polar, por representar monômeros divididos apenas segundo essas duas classes [56,67]. Como um exemplo detalhado, as interações entre os monômeros para um sistema de três tipos diferentes de monômeros (representados por A, B e C) são descritas segundo a matriz de interação:

$$H_{i,j} = \begin{matrix} & \begin{matrix} A & B & C \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} E_l & E_u & E_u \\ E_u & E_l & E_u \\ E_u & E_u & E_l \end{pmatrix} \end{matrix} \quad (4)$$

onde E_l é a energia de interação entre dois monômeros iguais (E_{like}) e E_u é a energia de interação entre dois monômeros diferentes (E_{unlike}) exemplificados na figura 5A. Os parâmetros E_l e E_u podem ser ajustados para representar diferentes condições do sistema como, por exemplo, os regimes de alta e baixa hidrofobicidade [42]. A, B e C representam três tipos diferentes de monômeros utilizados na equação 4, que poderiam ser ajustados para um número arbitrário de tipos de monômeros.

No modelo de rede, a dinâmica do enovelamento dependerá dos tipos de movimentos adotados para a cadeia [68,69]. Esse conjunto de movimentos consiste em deslocamentos locais que preservam a ligação covalente e a propriedade de volume excluído (figura 5B). Esses movimentos são efetuados seguindo uma simulação de Monte Carlo com o critério de Metropolis [70]. Os monômeros são escolhidos de forma aleatória e a simulação persiste até que um parâmetro de parada seja alcançado. Durante a simulação é possível amostrar os estados enovelados e desenovelados da proteína neste modelo e obter a estatística necessária para o cálculo dos parâmetros termodinâmicos como o calor específico, a energia total do sistema, a fração de proteína enovelada e a energia livre do sistema apresentados na figura 6.

Outra grandeza obtida das simulações do modelo de rede é o tempo de enovelamento (figura 7). Para este

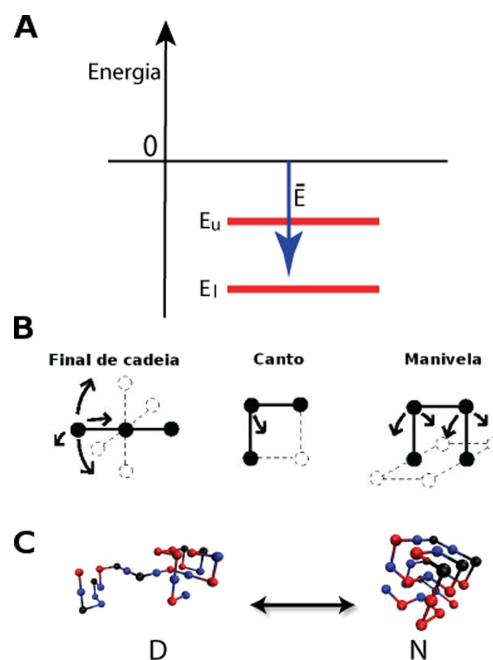


Figura 5: A) Diagrama de energia para o modelo de rede. E_l é a energia de interação entre dois monômeros iguais e E_u é a energia de interação entre dois monômeros diferentes. B) Três tipos de movimentos possíveis mais utilizados em simulações do modelo de rede cúbica. Na esquerda, é apresentado o movimento de final de cadeia, no movimento de canto no centro, e na direita, o movimento de manivela. C) Modelo de rede cúbica $3 \times 3 \times 3$ com 27 monômeros. Cada cor representa um monômero diferente, D é uma configuração inicial gerada aleatoriamente, e que representa o heteropolímero no estado desenovelado, e N é a configuração no estado nativo com os 28 contatos formados. Figura adaptada de [68].

cálculo é verificado a média do tempo de primeira passagem ou do inglês *mean first passage time* (mfpt), como também é conhecido. Neste caso, a simulação é iniciada com a proteína em um estado desenovelado aleatório. A simulação permanece em execução até que o estado nativo da proteína seja alcançado e a quantidade de passos de Monte Carlo é anotada. São realizadas várias simulações na mesma temperatura para se realizar a média do tempo de enovelamento nessa temperatura.

4.2. Modelos baseado em estrutura

O modelo de rede apresentado na seção anterior é capaz de responder questões gerais sobre o processo de enovelamento. No entanto, as aproximações realizadas no modelo de rede não caracterizam uma proteína real. Neste sentido, alguns modelos fora-da-rede (*off-lattice models*) foram desenvolvidos no intuito de abordar questões mais específicas sobre o processo de enovelamento. Os modelos fora-da-rede que utilizam a informação experimental da estrutura tridimensional da proteína são conhecidos como modelos baseados na estrutura (*structure-based models - SBM*) [71–73]. Os potenciais desses modelos simplificados possuem um custo computacional mais ele-

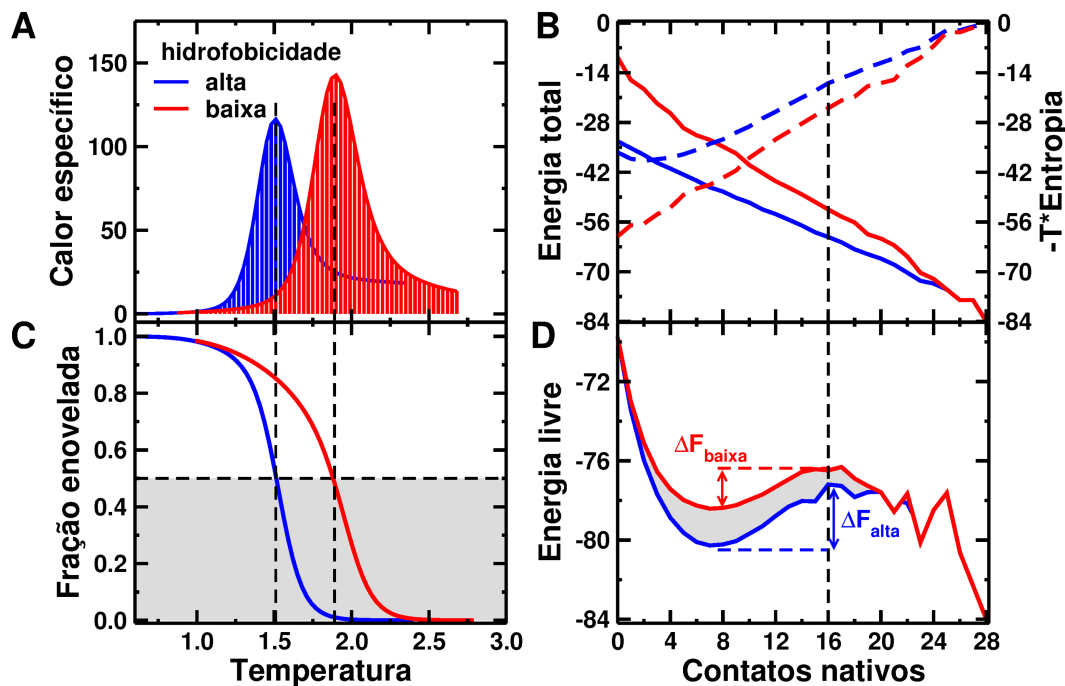


Figura 6: Análise termodinâmica obtida por simulações do modelo de rede para os regimes de alta e baixa hidrofobicidade, em azul e vermelho, respectivamente. A) Calor específico em função da temperatura em unidade reduzidas da simulação. O pico do valor do calor específico corresponde ao valor da temperatura de enovelamento (T_f). B) Em linhas contínuas é apresentada a energia total do heteropolímero em função dos contatos nativos. A energia total tem seu valor mínimo para o estado nativo da proteína que corresponde ao maior valor de contatos nativos formados. Em linhas tracejadas é apresentado o negativo do produto entre a temperatura e a entropia do sistema, que chega a zero no estado nativo devido ao único estado enovelado, cubo com 28 contatos nativos formados. C) Fração de proteína enovelada em função da temperatura. O valor de 0,5 de fração enovelada corresponde ao estado de transição e ao valor da temperatura de transição T_f , coincidindo com o valor encontrado pelo pico do calor específico. D) Perfil de energia livre em função dos contatos nativos. O parâmetro ΔF destaca a barreira de energia livre entre o estado desenovelado e o estado de transição da proteína modelo.

vado que as simulações de modelo de rede. As simulações SBMs ainda possuem custos computacionais menores que simulações de dinâmica molecular tradicionais e permitem uma ampla caracterização do espaço conformacional de proteínas. Os modelos baseados em estruturas descrevem alguns mecanismos do enovelamento e estão em acordo com resultados experimentais como, valores- Φ (mais detalhes na seção a seguir), barreiras de energia, taxas de enovelamento e coeficientes de difusão [74–84]. A aplicação do modelo baseado na estrutura dos carbonos alfa (SBM- C_α) se caracteriza por duas aproximações; a primeira se refere à cadeia polipeptídica onde são considerados apenas os carbonos alfa (C_α) dos resíduos de aminoácidos e a segunda aproximação se refere ao potencial de interação entre os carbonos [23, 71]. Esse potencial é construído a partir da estrutura nativa da proteína. A informação necessária com as características da proteína é previamente conhecida por estruturas depositadas no PDB (*Protein Data Bank*) no endereço <http://www.rcsb.org> [85]. A estrutura retirada do banco de dados é considerada como estado nativo para a construção do modelo. Desta forma, o mínimo de energia é atribuído a esse estado [86]. As simulações computacionais são realizadas por dinâmica molecular e o potencial

de interação para uma conformação Γ qualquer de uma proteína, com base na conformação nativa Γ_o é dado por

$$\begin{aligned}
 V(\Gamma, \Gamma_o) = & \sum_{bonds} \epsilon_r (r - r_o)^2 + \sum_{angles} \epsilon_\theta (\theta - \theta_o)^2 \\
 & + \sum_{dihedrals} \epsilon_\phi \{ [1 - \cos(\phi - \phi_o)] \\
 & + \frac{1}{2} [1 - \cos(3(\phi - \phi_o))] \} \\
 & + \sum_{contacts} \epsilon_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] \\
 & + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \quad (5)
 \end{aligned}$$

Na equação 5, os valores dos parâmetros são: $\epsilon_r = 100\epsilon_0$, $\epsilon_\theta = 20\epsilon_0$, $\epsilon_\phi = \epsilon_0$, $\epsilon_C = \epsilon_0$, $\epsilon_{NC} = \epsilon_0$, $\sigma_{NC} = 4.0 \text{ \AA}$ e ϵ_0 é a energia de interação por contato ($\epsilon_0 = 1$).

O primeiro termo do potencial (*bonds*) representa a ligação entre dois C_α adjacentes, formando um potencial harmônico. O parâmetro r_o é a distancia entre dois C_α da conformação nativa ligados entre si. O segundo termo (*angles*) se refere ao potencial harmônico angular formado por três C_α consecutivos, em que θ_o é o ângulo formado

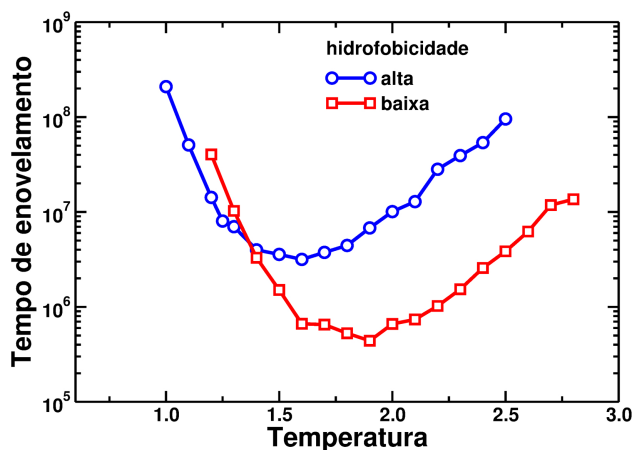


Figura 7: Tempos de envelhecimento obtidos por simulações do modelo de rede para os regimes de alta e baixa hidrofobicidade, em azul e vermelho, respectivamente. O valor do tempo de envelhecimento no modelo de rede é a média da quantidade de passos de Monte Carlo que o heteropolímero leva para formar os 28 contatos nativos, iniciando a simulação numa configuração aleatória desenovelada. São realizadas 1000 simulações para cada temperatura.

pelos resíduos i , $i + 1$ e $i + 2$ da conformação nativa. O terceiro termo (*dihedrals*) contabiliza a torção realizada na cadeia e está relacionado com a orientação da cadeia lateral (não expresso de maneira explícita no modelo). Este termo utiliza quatro C_α em sequência e o valor ângulo ϕ_o é retirado da estrutura nativa. O quarto termo (*contacts*) representa a interação entre os C_α i e j não ligados covalentemente mas que realizam um contato na estrutura nativa. Para esta informação é atribuído um potencial 10–12. O parâmetro d_{ij} é o valor da distância entre estes carbonos que realizam um contato nativo. A informação das interações entre os C_α é dada por um mapa de contato. A construção do mapa de contato pode ser realizada utilizando diferentes critérios e abordagens [87, 88]. Todos os outros carbonos que não realizam um contato nativo interagem por um potencial repulsivo, representado pelo último termo (*non-contacts*).

De maneira similar ao modelo de rede, durante a simulação é possível amostrar os estados envelhecidos e desenovelados da proteína e obter a estatística necessária para o cálculo dos parâmetros termodinâmicos. As análises termodinâmicas são realizadas utilizando o método de análise dos múltiplos histogramas (*weighted histogram analysis method - wham*), que faz uso de um conjunto de histogramas oriundos de várias simulações em diferentes temperaturas. Desta forma, é possível obter uma melhor amostragem das estruturas e da dinâmica da proteína durante o envelhecimento. Os parâmetros termodinâmicos como o calor específico, a energia total do sistema, a fração de proteína envelhecida e a energia livre utilizando o modelo C_α são apresentados na figura 8.

O tempo de envelhecimento das proteínas é outra quantidade que pode ser obtida pelo conjunto de simulações e

é apresentado na figura 9. Semelhante à parte cinética do modelo de rede, calcula-se o tempo de primeira passagem da proteína do estado desenovelado para o estado nativo. A simulação é iniciada com a proteína em um estado desenovelado aleatório e ela permanece em execução até que o estado nativo da proteína seja alcançado e a quantidade de passos de integração de dinâmica molecular é anotada.

Existem outros modelos baseados em estruturas além do modelo C_α que não serão discutidos aqui. Estes modelos, como por exemplo, C_β e todos-os-átomos (*all-atoms*), levam em consideração as informações explícitas sobre a cadeia lateral dos resíduos de aminoácidos. Os detalhes sobre as simulações dos modelos baseados em estrutura e a preparação dos arquivos para a simulação podem ser encontrados no servidor *Structure-based Models for Biomolecules (SMOG)* no endereço <http://smog-server.org> [89].

5. Caracterização do estado de transição por análise de valores- Φ

É possível alterar taxas cinéticas do envelhecimento de uma proteína e também a sua estabilidade termodinâmica realizando mutações de resíduos de aminoácidos por técnicas de engenharia de proteínas [74]. Para uma proteína que se envelhece via mecanismo de dois estados [90] ou via estados intermediários [91], pode-se caracterizar o conjunto estrutural do estado de transição destes mecanismos aplicando a análise dos valores- Φ [78]. Os valores- Φ mapeiam as estruturas dos estados de transição e de estados intermediários, não somente no envelhecimento de proteína, mas também em reações de catálise, de ligantes e de mudanças conformacionais no nível de resíduos individuais [78]. Experimentalmente, o conjunto estrutural do estado de transição em proteínas pode ser estudado medindo variações da estabilidade nativa e das taxas de envelhecimento/desenovelamento por mutações pontuais de resíduos específicos, e é dado por [74, 75]

$$\Phi \equiv \frac{-RT \ln k_{mut}/k_{wt}}{\Delta\Delta G^0}, \quad (6)$$

onde k_{mut} e k_{wt} são as taxas de envelhecimento das proteínas modificada (ou mutante) e nativa (ou tipo-selvagem), respectivamente, R é a constante dos gases ideais, T é a temperatura absoluta e $\Delta\Delta G^0$ é a mudança na estabilidade do estado nativo após a mutação.

Como mencionado nas seções anteriores, o envelhecimento de proteínas pode ser descrito como um processo difusivo sobre uma barreira de energia determinada pelo perfil de energia livre, e as taxas de envelhecimento podem ser calculadas pela equação de Kramers do tipo [23, 41]

$$k = k_0 \exp(\Delta G^{TS}/RT) \quad (7)$$

com k_0 sendo o pré-fator dependente da curvatura da barreira e do coeficiente de difusão configuracional do sistema. Considerando que k_0 tem uma variação insignificante com pequenas perturbações causadas por uma

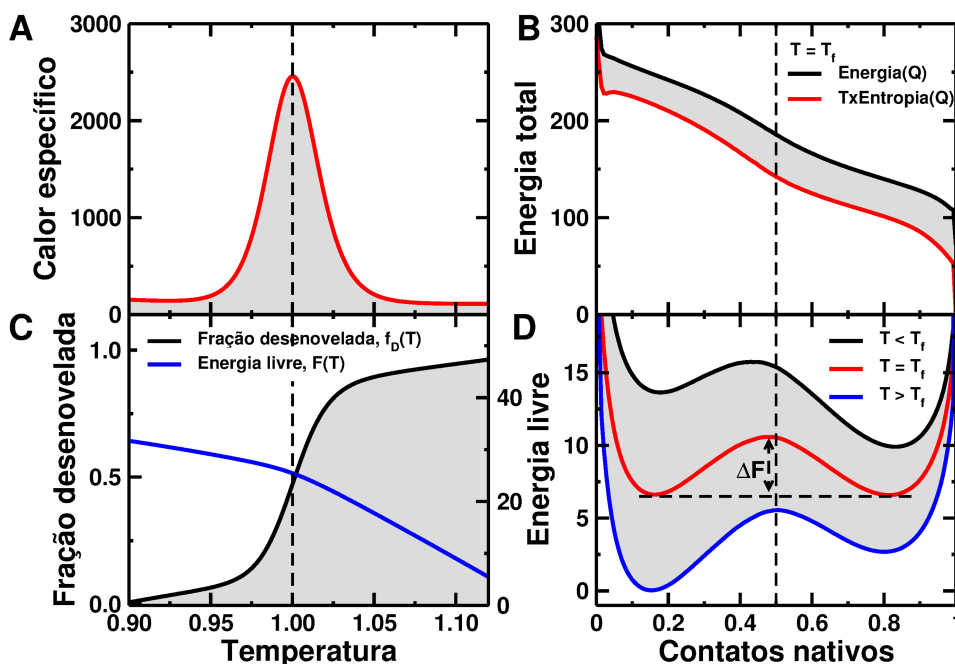


Figura 8: Parâmetros termodinâmicos obtidos por simulações do modelo C_α para a proteína *cold shock* (código PDB 1G6P) [92]. A) Calor específico em função da temperatura em unidades reduzidas da simulação normalizada pela temperatura de enovelamento (T_f). B) A linha em preto representa a energia total da proteína modelo ($E(Q)$) em função dos contatos nativos (Q). A energia total tem seu valor mínimo para o estado nativo da proteína que corresponde ao maior valor de contatos nativos formados. A linha vermelha representa o produto entre a temperatura e a entropia ($TS(Q)$) do sistema em função de Q . As curvas foram obtidas por simulações em T_f . C) A linha em preto representa a fração de proteína enovelada (f_D) em função da temperatura. O valor de 0,5 de fração enovelada corresponde ao estado de transição e ao valor da temperatura de transição, coincidindo com o valor encontrado pelo pico do calor específico em A). A linha em azul representa o valor da energia livre $F(T)$ em função da temperatura. D) Perfil de energia livre em função dos contatos nativos ($F(Q)$). O parâmetro ΔF destaca a barreira de energia livre entre o estado desenovelado e o estado de transição da proteína. A curva em preto apresenta um perfil de energia livre em uma temperatura abaixo da temperatura de enovelamento da proteína $T < T_f$, privilegiando o estado nativo e tornando-o mais estável. A curva vermelha apresenta o perfil obtido na temperatura de enovelamento $T = T_f$. Neste caso, os estados desenovelado e nativo possuem a mesma estabilidade e são acessados com mesma probabilidade. Em azul é apresentado o perfil de F para uma temperatura maior que a temperatura de enovelamento da proteína $T > T_f$. Nesse caso, o estado nativo é menos estável e o estado desenovelado é acessado mais vezes. As grandezas teóricas obtidas estão diretamente relacionadas com resultados experimentais em estudos de enovelamento e estabilidade de proteínas [43, 82].

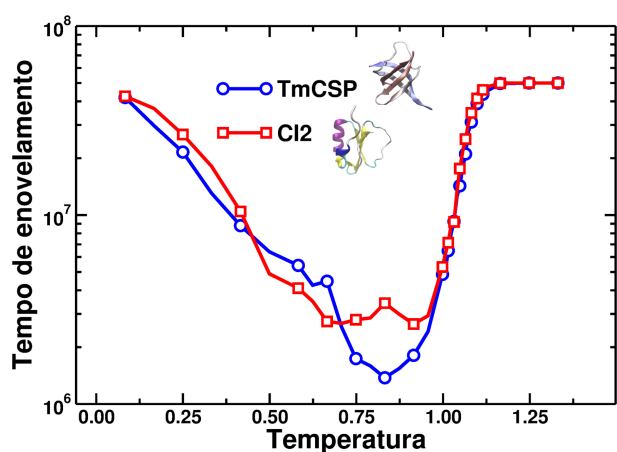


Figura 9: Tempos de enovelamento obtidos por simulações do modelo C_α para as proteínas *cold shock* (TmCSP) e *chymotrypsin inhibitor 2* (CI2), em azul e vermelho, respectivamente. O valor do tempo de enovelamento é determinado por uma média de números de passos de dinâmica molecular que uma proteína desenovelada leva para atingir o estado nativo pela primeira vez. As médias são calculadas com 1000 simulações para cada temperatura.

mutação, e substituindo a equação 7 em 6, Φ será a variação da barreira de energia livre termodinâmica do estado de transição com relação ao estado desenovelado ($\Delta\Delta G^{TS-U}$) sobre a mudança na estabilidade do estado nativo ($\Delta\Delta G^{F-U}$), ambos após a mutação, com

$$\Phi = \frac{\Delta\Delta G^{TS-U}}{\Delta\Delta G^{F-U}} \tag{8}$$

Essas variações em energia livre nos estados nativo (F), desenovelado (U) e de transição (TS) são evidenciadas na figura 10, nos quais aqui consideramos as siglas em inglês dos estados para fácil comparação com a vasta literatura acerca do assunto.

A figura 11 representa a faixa de valores que a equação 8 resulta para Φ pela variação da energia livre após a mutação de um resíduo de aminoácido da proteína. A figura 11A corresponde ao caso em que $\Phi = 0$ e, localmente, o estado de transição é idêntico ao estado desenovelado. A figura 11B corresponde à $\Phi = 1$ e, localmente, o estado de transição é idêntico ao estado enovelado. Φ será próximo de zero quando a região em torno de um resíduo for similar à região em torno do resíduo no estado dese-

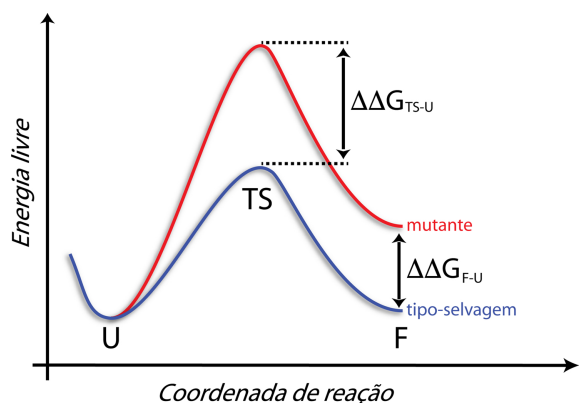


Figura 10: Análise dos valores- Φ para o caso de uma proteína que enovela via mecanismo de dois estados mapeado por uma simples coordenada de reação. Uma mutação num resíduo de aminoácido pode mudar a estabilidade do estado de transição quando comparado com o estado desenovelado ($\Delta\Delta G_{TS-U}$) e a estabilidade do estado enovelado nativo quando comparado com o estado desenovelado ($\Delta\Delta G_{F-U}$). A razão entre essas duas quantidades ($\Phi = \Delta\Delta G_{TS-U} / \Delta\Delta G_{F-U}$), determina o valor- Φ da fração da diferença de energia da proteína do tipo selvagem com relação a proteína mutante.

novelado e próximo de um quando for similar à região no estado enovelado. Sendo assim, valores- Φ entre zero e uma unidade indicam que a mutação contribuiu parcialmente para alterar a interação existente no estado de transição. No entanto, a equação 8 pode exibir desvios da faixa $0 \leq \Phi \leq 1$ os quais são considerados valores- Φ anômalos que podem ser sinais de frustrações de interações energéticas ou restrições topológicas patológicas [93].

Computacionalmente, calcular variações em ΔG causadas por mutação é menos custoso do que computar taxas de enovelamento (k). No caso de simulações de modelo baseado na estrutura de proteínas, pode-se ainda aproximar as variações em ΔG pela formação de contatos entre pares de resíduos i e j pela equação [71]

$$\Phi_{ij} = \frac{\Delta\Delta G_{TS-U}}{\Delta\Delta G_{F-U}} \approx \frac{P_{ij}^{TS} - P_{ij}^U}{P_{ij}^F - P_{ij}^U} \quad (9)$$

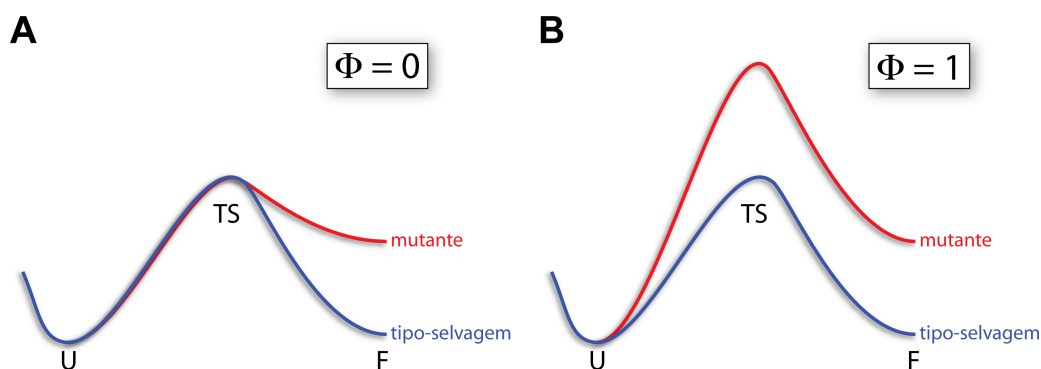


Figura 11: Valores- Φ para os casos em que, após a mutação de um resíduo de aminoácido, A) os estados desenovelado e de transição são similares ($\Phi = 0$), e B) os estados de transição e nativo são diferentes ($\Phi = 1$).

com P_{ij}^X sendo a probabilidade de contatos entre resíduos ij formados nos estados X (com X sendo F , TS ou U) [36, 82, 94]. O valor- Φ por resíduo é dado pela média envolvendo os contatos nativos de um resíduo i ($\Phi_i = \sum_j \Phi_{ij}$). De acordo com a equação 9, Φ mede o conteúdo do estado nativo ainda presente no estado de transição de cada resíduo numa escala de 0 a 1, com 1 sendo um resíduo no estado de transição pertencente ao conjunto de estruturas do estado nativo, e 0 sendo um resíduo com nenhuma informação do estado nativo quando no estado de transição. Os valores- Φ obtidos computacionalmente por utilização do SBM- C_α tiveram grande concordância com os valores obtidos experimentalmente [36, 71, 82, 94] o que contribuiu para a grande aceitação dos modelos computacionais simplificados de proteínas [95].

Com isso, com o cálculo dos valores- Φ obtidos para uma proteína, pode-se investigar a influência de determinados resíduos de aminoácidos para o conjunto de estruturas e suas energias que determinam o estado de transição e se, estes mesmos resíduos, são do conjunto de estruturas nativas ou não-nativas. Pode-se alterar a superfície de energia de uma proteína de modo a alterar rotas de enovelamento, estabilidade térmica, taxas de reação de catálise e outros mecanismos relacionadas às atividades proteicas.

6. Comentários finais

Neste trabalho foram introduzidos os principais aspectos teóricos sobre o processo de enovelamento de proteínas. Foram discutidas as principais características biofísicas e bioquímicas das proteínas e qual a sua importância para os seres vivos. Os principais resultados científicos sobre esse assunto foram apresentados dentro de um contexto histórico até a teoria de Superfície de Energia com o uso de modelos computacionais simplificados. Os principais resultados esperados com as simulações dos modelos simples são discutidos ao longo do trabalho, junto com a apresentação dos detalhes dos métodos. O trabalho procurou introduzir discentes de física e áreas afins ao problema do enovelamento de proteínas. Alguns tópicos

mais elaborados como a modificação dos potenciais utilizados, ou a adição de outros potenciais associados com a presença de frustração em proteínas, ou a adição de interação não-nativas eletrostáticas não discutidos em detalhes aqui. Existem um conjunto amplo de trabalhos na literatura abordando os efeitos da adição desses potenciais na estabilidade e na dinâmica da proteína [45, 54, 55, 82, 96], inclusive, mais recentemente, com simulações de dinâmica molecular com pH constante [97–100]. Outra abordagem utilizando os modelos simplificados baseados em estrutura é o estudo de mudanças conformacionais envolvendo duas estruturas com o mínimo de energia. Esta abordagem conhecida como *Dual-Basin SBM* é empregada para o estudo da transição conformacional entre as duas estruturas, e o mapeamento das estruturas intermediárias [101–103]. Mais recentemente, o entendimento da dinâmica em sistemas maiores e mais complexos, como ribossomos, também são estudados utilizando os modelos baseados em estruturas [104–107]. Mesmo com os avanços recentes no entendimento do problema do enovelamento de proteínas, essa é uma área em constante desenvolvimento e novas questões relevantes surgem o tempo todo. Um exemplo sobre esses tópicos são as proteínas que fazem nós durante o processo de enovelamento [108–111]. Outro exemplo são as proteínas intrinsecamente desordenadas *IDPs* (*Intrinsically Disordered Proteins*), ou seja, não possuem um estado nativo bem estruturado. Essas proteínas possuem estados parcialmente estruturados e que podem se enovelar ao interagirem com outras macromoléculas [112, 113]. Essas, entre outras questões, estão em aberto e tornam a área do estudo de enovelamento de proteínas um campo amplo e relevante para ser explorado ainda mais.

7. Agradecimentos

VGC agradece à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - proc. 2016/13998-8 e proc. 2017/09662-7 e também agradece à NSF (National Science Foundation) Grant PHYS-1427654. VBPL agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à FAPESP proc. 2014/06862-7 e proc. 2016/19766-1. RJO agradece à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). ABOJ agradece à CAPES. Os autores agradecem os recursos computacionais disponibilizados pelo Núcleo de Computação Científica (NCC/GridUNESP) da Universidade Estadual Paulista (UNESP) e pela Rede Mineira de Química (RQ-MG).

8. Material suplementar

O seguinte material suplementar está disponível online:
Apêndice

Referências

- [1] D. Voet, J. G. Voet e C. W. Pratt, *Principles of biochemistry* (J. Wiley & sons, New Jersey, 2008), 1328 p.
- [2] C. M. Dobson, *Nature* **426**, 884 (2003).
- [3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts e P. Walter, *Biologia Molecular da Célula* (Artmed, Porto Alegre, 2004), 4^a ed.
- [4] A. L. Lehninger, D. L. Nelson e M. M. Cox, *Lehninger principles of biochemistry* (W.H. Freeman, New York, 2005).
- [5] S. Vijay-Kumar, C. E. Bugg e W. J. Cook, *J. Mol. Biol.* **194**, 531 (1987).
- [6] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff e D. C. Phillips, *Nature* **181**, 662 (1958).
- [7] C. B. Anfinsen, E. Harber, M. Sela e F. White Jr, *Proc. Natl. Acad. Sci. USA* **47**, 1309 (1961).
- [8] C. B. Anfinsen, *Science* **181**, 223 (1973).
- [9] C. Levinthal, *J. Chem. Phys.* **65**, 44 (1968).
- [10] R. Zwanzig, A. Szabo e B. Bagchi, *Proc. Natl. Acad. Sci. USA* **89**, 20 (1992).
- [11] C. Levinthal, in *Mössbaun Spectroscopy in Biological Systems Proceedings*, Urbana, 1968 (University of Illinois Urbana-Champaign, Urbana, 1969), 22-24.
- [12] N. J. Darby, C. P. M. van Mierlo e T. E. Creighton, *FEBS Letters* **279**, 61 (1991).
- [13] J. P. Staley e P. S. Kim, *Proc. Natl. Acad. Sci. USA* **89**, 1519 (1992).
- [14] A. L. Fink, *Annu. Rev. Biophys. Biomol. Struct* **24**, 495 (1995).
- [15] J. D. Bryngelson e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84**, 7524 (1987).
- [16] J. D. Bryngelson e P. G. Wolynes, *The Journal of Physical Chemistry* **93**, 6902 (1989).
- [17] P. E. Leopold, M. Montal e J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **89**, 8721 (1992).
- [18] J. N. Onuchic, Z. Luthey-Schulten e P. G. Wolynes, *Annual Review of Physical Chemistry* **48**, 545 (1997).
- [19] K. A. Dill, S. B. Ozkan, M. S. Shell e T. R. Weikl, *Annual Review of Biophysics* **37**, 289 (2008).
- [20] D. Thirumalai, E. P. O'Brien, G. Morrison e Changbong Hyeon, *Annual Review of Biophysics* **39**, 159 (2010).
- [21] J. Wang, R. J. Oliveira, X. Chu, P. C. Whitford, J. Chahine, W. Han, E. Wang, J. N. Onuchic e V. B. P. Leite, *Proc. Natl. Acad. Sci. USA* **109**, 15763 (2012).
- [22] H. Frauenfelder, S. G. Sligar e P. G. Wolynes, *Science* **254**, 1598 (1991).
- [23] J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine e N. D. Socci, *Advances in Protein Chemistry* **53**, 87 (2000).
- [24] S. S. Plotkin e J. N. Onuchic, *Quarterly Reviews of Biophysics* **35**, 111 (2002).
- [25] L. L. Chavez, J. N. Onuchic e C. Clementi, *Journal of the American Chemical Society* **126**, 8426 (2004).
- [26] D. J. Wales, *Current Opinion in Structural Biology* **20**, 3 (2010).
- [27] E. I. Shakhnovich e A. M. Gutin, *Biophysical Chemistry* **34**, 187 (1989).
- [28] J. D. Bryngelson e P. G. Wolynes, *Biopolymers* **30**, 177 (1990).

- [29] R. A. Goldstein, Z. A. Luthey-Schulten e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **89**, 4918 (1992).
- [30] J. D. Bryngelson, J. N. Onuchic, N. D. Socci e P. G. Wolynes, *Proteins: Structure, Function, and Bioinformatics* **21**, 167 (1995).
- [31] D. Sherrington e S. Kirkpatrick, *Physical Review Letters* **35**, 1792 (1975).
- [32] M. Mezard, G. Parisi e M.A., *Virasoro em Spin glass theory and beyond* (World Scientific, Singapore, 1987).
- [33] L. Sutto, J. Lätzer, J. A. Hegler, D. U. Ferreira e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **104**, 19825 (2007).
- [34] P. F. N. Faisca, A. Nunes, R. D. M. Travasso e E. I. Shakhnovich, *Protein Science* **19**, 2196 (2010).
- [35] Y. Sun e D. Ming, *PLoS ONE* **9**, e87719 (2014).
- [36] S. S. Cho, Y. Levy e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **103**, 586 (2006).
- [37] H. A. Scheraga, M. Khalili e A. Liwo, *Annual Review of Physical Chemistry* **58**, 57 (2007).
- [38] P. G. Wolynes, J. N. Onuchic e D. Thirumalai, *Science* **267**, 1619 (1995).
- [39] J. Chahine, H. Nymeyer, V. B. P. Leite, N. D. Socci e J. N. Onuchic, *Physical Review Letters* **88**, 168101 (2002).
- [40] A. B. Oliveira Junior, F. M. Fatore, F. V. Paulovich, O. N. Oliveira Junior e V. B. P. Leite, *PLoS ONE* **9**, e100861 (2014).
- [41] N. D. Socci, J. N. Onuchic e P. G. Wolynes, *The Journal of Chemical Physics* **104**, 5860 (1996).
- [42] J. Chahine, R. J. Oliveira, V. B. P. Leite e J. Wang, *Proc. Natl. Acad. Sci. USA* **104**, 14646 (2007).
- [43] R. J. Oliveira, P. C. Whitford, J. Chahine, V. B. P. Leite e J. Wang, *Methods* **52**, 91 (2010).
- [44] Y. Harpaz, N. Elmasry, A. R. Fersht e K. Henrick, *Proc. Natl. Acad. Sci. USA* **91**, 311 (1994).
- [45] V. G. Contessoto, *Estudo do efeito da adição de frustração no enovelamento de proteínas utilizando modelos baseados em estruturas*. Dissertação de Mestrado, Universidade Estadual Paulista, São Paulo (2012).
- [46] J. M. Sanchez-Ruiz, *Annual Review of Physical Chemistry* **62**, 231 (2011).
- [47] J. Wen, K. Arthur, L. Chemmalil, S. Muzammil, J. Gabrielson e Y. Jiang, *Journal of Pharmaceutical Sciences* **101**, 955 (2012).
- [48] C. M. Johnson, *Archives of Biochemistry and Biophysics* **531**, 100 (2013).
- [49] H. Nymeyer, A. E. García e J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **95**, 5921 (1998).
- [50] S. S. Plotkin, *Proteins: Structure, Function, and Bioinformatics* **45**, 337 (2001).
- [51] C. Clementi e S. S. Plotkin, *Protein Science* **13**, 1750 (2004).
- [52] L. C. Oliveira, R. T. H. Silva, V. B. P. Leite e J. Chahine, *The Journal of Chemical Physics* **125**, 084904 (2006).
- [53] D. U. Ferreira, J. A. Hegler, E. A. Komives e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **104**, 19819 (2007).
- [54] V. G. Contessoto, D. T. Lima, R. J. Oliveira, A. T. Bruni, J. Chahine e V. B. P. Leite, *Proteins: Structure, Function, and Bioinformatics* **81**, 1727 (2013).
- [55] P. R. Mouro, V. G. Contessoto, J. Chahine, R. J. de Oliveira e V. B. P. Leite, *Biophysical Journal* **111**, 287 (2016).
- [56] H. S. Chan e K. A. Dill, *Annual Review of Biophysics and Biophysical Chemistry* **20**, 447 (1991).
- [57] H. S. Chan e K. A. Dill, *Proteins* **24**, 335 (1996).
- [58] K. A. Dill e H. S. Chan, *Nat Struct Mol Biol* **4**, 10 (1997).
- [59] R. I. Dima, J. R. Banavar, M. Cieplak e A. Maritan, *Proc. Natl. Acad. Sci. USA* **96**, 4904 (1999).
- [60] K. A. Dill, S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas e H. S. Chan, *Protein Science* **4**, 561 (1995).
- [61] H. Li, R. Helling, C. Tang e Ned Wingreen, *Science* **273**, 666 (1996).
- [62] D. K. Klimov e D. Thirumalai, *Folding & Design* **3**, 127 (1998).
- [63] H. Cejtin, J. Edler, A. Gottlieb, R. Helling, H. Li, J. Philbin, N. Wingreen e C. Tang, *The Journal of Chemical Physics* **116**, 352 (2002).
- [64] N. D. Socci e J. N. Onuchic, *The Journal of Chemical Physics* **103**, 4732 (1995).
- [65] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach et al., *Physical Review Letters* **62**, 1916 (1989).
- [66] E. Shakhnovich e A. Gutin, *The Journal of Chemical Physics* **93**, 5967 (1990).
- [67] A. Irbäck e C. Troein, *J. Biol. Phys.* **28**, 1 (2002).
- [68] N. D. Socci e J. N. Onuchic, *Journal of Chemical Physics* **101**, 23 (1994).
- [69] V. B. P. Leite, J. N. Onuchic, G. Stell e J. Wang, *Biophysical Journal* **87**, 3633 (2004).
- [70] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller e E. Teller, *The Journal of Chemical Physics* **21**, 1087 (1953).
- [71] C. Clementi, H. Nymeyer e J. N. Onuchic, *Journal of Molecular Biology* **298**, 937 (2000).
- [72] P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu e J. N. Onuchic, *Proteins* **75**, 430 (2009).
- [73] J. K. Noel e J. N. Onuchic, in: *Computational Modeling of Biological Systems*, editado por N. V. Dokholyan (Springer, Londres, 2012).
- [74] A. Matouschek, J. T. Kellis, L. Serrano e A. R. Fersht, *Nature* **340**, 122 (1989).
- [75] A. R. Fersht, *Current Opinion in Structural Biology* **5**, 79 (1995).
- [76] K. W. Plaxco, K. T. Simons e D. Baker, *Journal of Molecular Biology* **277**, 985 (1998).
- [77] N. Koga e S. Takada, *Journal of Molecular Biology* **313**, 171 (2001).
- [78] A. R. Fersht e S. Sato, *Proc. Natl. Acad. Sci. USA* **101**, 7976 (2004).
- [79] J. Kubelka, J. Hofrichter e W. A. Eaton, *Current Opinion in Structural Biology* **14**, 76 (2004).
- [80] C. D. Snow, E. J. Sorin, Y. M. Rhee e V. S. Pande, *Annual Review of Biophysics and Biomolecular Structure* **34**, 43 (2005).
- [81] R. B. Best e G. Hummer, *Proc. Natl. Acad. Sci. USA* **107**, 1088 (2010).
- [82] R. J. Oliveira, P. C. Whitford, J. Chahine, J. Wang, J. N. Onuchic e V. B. P. Leite, *Biophysical Journal* **99**, 600 (2010).
- [83] F. Polotto, E. Drigo Filho, J. Chahine e R. J. de Oliveira, *Physica A: Statistical Mechanics and its Applications* **493**, 286 (2018).

- [84] M. R. de Mendonça, L. G. Rizzi, V. Contessoto, V. B. P. Leite e N. A. Alves, *Proteins: Structure, Function, and Bioinformatics* **82**, 119 (2014).
- [85] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov e P. E. Bourne, *Nucleic Acids Research* **28**, 235 (2000).
- [86] H. Taketomi, Y. Ueda e N. GÖ, *International Journal of Peptide and Protein Research* **7**, 445 (1975).
- [87] V. Sobolev, A. Sorokine, J. Prilusky, E. E. Abola e M. Edelman, *Bioinformatics* **15**, 327 (1999).
- [88] J. K. Noel, P. C. Whitford e J. N. Onuchic, *J. Phys. Chem. B.* **116**, 8692 (2012).
- [89] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu e J. N. Onuchic, *Nucleic Acids Research* **38**, W657 (2010).
- [90] S. E. Jackson e A. R. Fersht, *Biochemistry* **30**, 10428 (1991).
- [91] A. R. Fersht, A. Matouschek e L. Serrano, *J. Mol. Biol.* **224**, 771 (1992).
- [92] W. Kremer, B. Schuler, S. Harrieder, M. Geyer, W. Gronwald, C. Welker, R. Jaenicke e H. R. Kalbitzer, *European Journal of Biochemistry* **268**, 2527 (2001).
- [93] M. Oliveberg e P. G. Wolynes, *Q Rev Biophys* **38**, 245 (2005).
- [94] J. E. Shea, J. N. Onuchic e C. L. Brooks, *Proc. Natl. Acad. Sci. USA* **96**, 12512 (1999).
- [95] P. C. Whitford, K. Y. Sanbonmatsu e J. N. Onuchic, *Reports on Progress in Physics* **75**, 076601 (2012).
- [96] F. B. da Silva, V. G. Contessoto, V. M. Oliveira, J. Clarke e Vitor B. P. Leite, *bioRxiv*: 232116 (2017).
- [97] V. G. Contessoto, V. M. de Oliveira, S. J. de Carvalho, L. C. Oliveira e V. B. P. Leite, *Journal of Chemical Theory and Computation* **12**, 3270 (2016).
- [98] V. G. Contessoto, *Estudo de interações eletrostáticas no processo de enovelamento e na estabilidade de proteínas utilizando modelos simplificados*. Tese de Doutorado, Universidade Estadual Paulista, São Paulo (2016).
- [99] M. A. Coronado, I. P. Caruso, V. M. De Oliveira, V. G. Contessoto, V. B. P. Leite, L. A. Kawai, R. K. Arni e R. J. Eberle, *Protein and Peptide Letters* **24**, 358 (2017).
- [100] V. M. de Oliveira, V. G. Contessoto, F. B. da Silva, D. L. Z. Caetano, S. J. de Carvalho e V. B. P. Leite, *Biophysical Journal* **114**, 65 (2018).
- [101] P. C. Whitford, S. Gosavi e J. N. Onuchic, *Journal of Biological Chemistry* **283**, 2042 (2008).
- [102] A. Schug, P. C. Whitford, Y. Levy e J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **104**, 17674 (2007).
- [103] Y. Levy, S. S. Cho, T. Shen, J. N. Onuchic e P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **102**, 2373 (2005).
- [104] P. C. Whitford, P. Geggier, R. B. Altman, S. C. Blanchard, J. N. Onuchic e K. Y. Sanbonmatsu, *RNA* **16**, 1196 (2010).
- [105] P. C. Whitford, J. N. Onuchic e K. Y. Sanbonmatsu, *Journal of the American Chemical Society* **132**, 13170 (2010).
- [106] P. C. Whitford, S. C. Blanchard, J. H. D. Cate e K. Y. Sanbonmatsu, *PLOS Computational Biology* **9**, e1003003 (2013).
- [107] J. Noel, J. Chahine, V. P. Leite e P. C. Whitford, *Biophysical Journal* **107**, 2881 (2014).
- [108] M. Jamroz, W. Niemyska, E. J. Rawdon, A. Stasiak, K. C. Millett, Piotr Sulkowski e J. I. Sulkowska, *Nucleic Acids Research* **43**, D306 (2015).
- [109] J. I. Sulkowska, E. J. Rawdon, K. C. Millett, J. N. Onuchic e A. Stasiak, *Proc. Natl. Acad. Sci. USA* **109**, E1715 (2012).
- [110] F. Ziegler, N. C. H. Lim, S. S. Mandal, B. Pelz, W. Ng, M. Schlierf, S. E. Jackson e M. Rief, *Proc. Natl. Acad. Sci. USA* **113**, 7533 (2016).
- [111] P. F.N. Faísca, *Computational and Structural Biotechnology Journal* **13**, 459 (2015).
- [112] P. E. Wright e H. J. Dyson, *Nat. Rev. Mol. Cell. Biol.* **16**, 18 (2015).
- [113] C. W. Lawrence, S. Kumar, W. G. Noid e Sc. A. Showalter, *The Journal of Physical Chemistry Letters* **5**, 833 (2014).