



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**

ANDER FRANCISCO PEREIRA

**EFEITOS DE SOLVENTE NA TERMODINÂMICA DO ENovelAMENTO DE
PROTEÍNAS**

CAMPINAS

2025

ANDER FRANCISCO PEREIRA

**EFEITOS DE SOLVENTE NA TERMODINÂMICA DO ENOVELAMENTO DE
PROTEÍNAS**

**Tese apresentada ao Instituto de Química da Universidade
Estadual de Campinas como parte dos requisitos exigidos
para a obtenção do título de Doutor em Ciências.**

Orientador: Prof. Dr. Leandro Martínez

**ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO
ALUNO ANDER FRANCISCO PEREIRA E ORIENTADA PELO PROF. DR. LEANDRO
MARTÍNEZ.**

CAMPINAS

2025

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Química
Simone Luiz Alves - CRB 8/9094

P414e Pereira, Ander Francisco, 1996-
Efeitos de solvente na termodinâmica do enovelamento de proteínas /
Ander Francisco Pereira. – Campinas, SP : [s.n.], 2025.

Orientador: Leandro Martínez.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP),
Instituto de Química.

1. Enovelamento de proteínas. 2. Modelos baseados na estrutura. 3.
Cossolventes. 4. Funções de distribuição de mínima distância. 5. Teoria de
soluções de Kirkwood-Buff. I. Martínez, Leandro, 1979-. II. Universidade
Estadual de Campinas (UNICAMP). Instituto de Química. III. Título.

Informações complementares

Título em outro idioma: Solvent effects on protein folding thermodynamics

Palavras-chave em inglês:

Protein folding

Structure-based models

Cosolvents

Minimum-distance distribution functions

Kirkwood–Buff solution theory

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora:

Leandro Martínez [Orientador]

Munir Salomão Skaf

René Alfonso Nome Silva

Ari Fernando Zeida Camacho

Rodrigo Leandro Silveira

Data de defesa: 28-03-2025

Programa de Pós-Graduação: Química

Objetivos de Desenvolvimento Sustentável (ODS)

Não se aplica

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0001-6844-5610>

- Currículo Lattes do autor: <http://lattes.cnpq.br/0909231391640212>

BANCA EXAMINADORA

Prof. Dr. Leandro Martínez (Orientador)

Prof. Dr. Munir Salomão Skaf (Instituto de Química - UNICAMP)

Prof. Dr. René Alfonso Nome Silva (Instituto de Química - UNICAMP)

Prof. Dr. Ari Fernando Zeida Camacho (Departamento de Bioquímica, Facultad de Medicina - Universidad de la República)

Prof. Dr. Rodrigo Leandro Silveira - (Instituto de Química - Universidade Federal do Rio de Janeiro)

A Ata da defesa assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

“The important thing is not to stop questioning. Curiosity has its own reason for existing.”

Albert Einstein

Agradecimentos

Esta seção não poderia começar sem antes agradecer a Deus, por sua bondade, por iluminar meus passos, permitir-me sonhar e dar-me coragem para perseguir meus objetivos ao longo desses anos.

Agradeço especialmente ao meu orientador, professor Leandro, pela excelente orientação ao longo deste período, essencial para meu desenvolvimento pessoal e profissional. O professor Leandro é sempre solícito e didático. Espero seguir seu exemplo como pessoa e pesquisador. Obrigado, professor, por me dar a oportunidade de fazer parte do grupo M3G.

Agradeço aos atuais membros do grupo (Ana Beatriz, Camila, Eduard, Felipe, Jéssica, Lucas, Pamella e Vinicius) e aos antigos (Brenda, Diego e Guilherme) pela convivência e pelas trocas de conhecimento. De modo especial, agradeço à Pamella e ao Vinicius pela companhia no laboratório, pelas conversas e também pelos momentos de descontração fora do ambiente de trabalho. Um agradecimento adicional ao Vinicius pela colaboração e ajuda, especialmente quando ingressei no grupo.

Sou grato ao professor José N. Onuchic pela oportunidade de fazer parte do *Center for Theoretical Biological Physics* e pela orientação durante um ano do meu estágio sanduíche nos EUA. Durante esse período, convivi grande parte do tempo com alguns membros do grupo, e agradeço à Aleksandra, Ananya, Antônio, Erdong, Esteban, Hana, Matheus, Mengyan, Murilo, Ronaldo, Steven e Vinicius pela convivência e pelas ótimas conversas. Agradeço ainda ao Vinicius pela colaboração no trabalho desenvolvido ao longo do estágio.

Agradeço ao meu amigo Francisco, pela amizade desde a graduação e, coincidentemente, pela convivência durante um ano nos EUA.

Agradeço imensamente à minha mãe, Maria, e ao meu pai, Benedito, pelo apoio incondicional, pela confiança e pelos ensinamentos constantes. Agradeço também à minha irmã, Amanda, e aos meus sobrinhos, Maria Clara, Antonella e Amílcar, que tornam minha vida mais feliz.

Reconheço e valorizo o trabalho dos desenvolvedores de software livre, que disponibilizam ferramentas essenciais para a pesquisa científica e a produção acadêmica.

Agradeço ao CNPq pelo financiamento parcial do meu doutorado (processo: 140853/2020-0).

Agradeço à FAPESP pela minha bolsa de doutorado e pelo financiamento de parte da minha pesquisa nos EUA (processos: 2020/04549-0 e 2023/02071-4).

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

Por fim, agradeço ao CEPID (Centro de Pesquisa, Inovação e Difusão) CCE&S (*Center for Computing in Engineering and Sciences*), apoiado financeiramente pela FAPESP (processos: 2010/16947-9, 2018/24293-0, 2013/08293-7, 2018/14274-9, 2019/17874-0 e 2020/04549-0). Os cálculos realizados nesta tese foram desenvolvidos com o auxílio do KAHUNA, do CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho em São Paulo) e do supercomputador Coaraci.

Muito obrigado a todos que contribuíram para essa jornada!

Resumo

Estruturas proteicas e seus mecanismos de enovelamento são dependentes do ambiente molecular. Proteínas podem adotar conformações distintas - ativas, inativas ou com diferentes funcionalidades - em função da temperatura, pressão ou características do solvente no qual estão inseridas. Embora as proteínas desempenhem suas funções em meios predominantemente aquosos, mesmo no ambiente celular, há uma variedade de espécies, como cossolventes e íons, que influenciam sua estrutura e função. Além disso, alguns cossolventes são usados *in vitro* para estabilizar proteínas e melhorar sua função em aplicações biotecnológicas. Os cossolventes estão presentes em organismos vivos e em diversas formulações, mas o mecanismo pelo qual afetam a estabilidade de proteínas, seja na sua forma nativa ou nas conformações não nativas, muitas vezes não é claro. Nesta tese, abordamos sobre o enovelamento de proteínas, e como diferentes solventes afetam sua termodinâmica. No Capítulo 3, investigamos o enovelamento do peptídeo (AAQAA)₃ em água e em soluções de 2,2,2-Trifluoroetanol (TFE), com o objetivo de elucidar os mecanismos estabilizadores do TFE. Utilizamos simulações de Dinâmica Molecular com troca de réplicas para melhorar a amostragem do peptídeo. No Capítulo 4, analisamos como o enovelamento do domínio B da proteína A (BdpA) está associado à formação de sua estrutura secundária e às estruturas de hidratação. Neste caso, utilizamos simulações computacionais com os modelos baseados na estrutura (SBMs, *Structure-Based Models*). Nossos resultados permitiram agrupar as estruturas parcialmente enoveladas da BdpA, possibilitando a compreensão dos efeitos de osmólitos na estabilidade relativa dos conjuntos de estruturas de enovelamento, que são abordados com detalhes no Capítulo 5. Neste Capítulo, discutimos como os osmólitos ureia e TMAO atuam no mecanismo de estabilização dos diferentes conjuntos de enovelamento do domínio SH3 e da proteína BdpA, que são proteínas com motivos estruturais distintos e modelos para estudos de enovelamento. Os efeitos do solvente foram abordados a partir das funções de distribuição de mínima-distância (MDDFs, *Minimum-distance distribution functions*) e da teoria de soluções de Kirkwood-Buff. Finalmente, no Capítulo 6 abordamos aspectos sobre o mecanismo de enovelamento da proteína MAD2 (*Mitotic spindle assembly checkpoint protein*). A MAD2 é uma proteína que se enovela de forma reversível em dois estados nativos distintos: o inativo e o ativo. O modelo desenvolvido neste trabalho captura a transição entre esses dois estados.

Abstract

Protein structures and their folding mechanisms are dependent on the molecular environment. Proteins can adopt distinct conformations - active, inactive, or with different functionalities - depending on temperature, pressure, or the characteristics of the solvent in which they are immersed. Although proteins perform their functions primarily in aqueous environments, even within the cellular milieu, various species, such as cosolvents and ions, influence their structure and function. Additionally, some cosolvents are used *in vitro* to stabilize proteins and enhance their function in biotechnological applications. Cosolvents are present in living organisms and various formulations, but the mechanism by which they affect protein stability - whether in the native state or non-native conformations - is often unclear. This thesis explores protein folding and how different solvents affect its thermodynamics. In Chapter 3, we investigate the folding of the (AAQAA)₃ peptide in water and 2,2,2-Trifluoroethanol (TFE) solutions to elucidate TFE's stabilizing mechanisms. We employ Replica Exchange Molecular Dynamics simulations to enhance peptide sampling. In Chapter 4, we analyze how the folding of the B domain of protein A (BdpA) is associated with its secondary structure formation and hydration structures. In this case, we use simulations with Structure-Based Models (SBMs). Our results allowed us to group partially folded BdpA structures, enabling an understanding of the effects of osmolytes on the relative stability of folding ensembles, which are discussed in detail in Chapter 5. Chapter 5 examines how the osmolytes urea and TMAO contribute to the stabilization mechanisms of different folding ensembles of the SH3 domain and BdpA, proteins with distinct structural motifs widely adopted as models for folding studies. Solvent effects are analyzed using Minimum-Distance Distribution Functions (MDDFs) and Kirkwood-Buff theory. Finally, in Chapter 6, we address aspects of the folding mechanism of the MAD2 protein (Mitotic Spindle Assembly Checkpoint Protein). MAD2 undergoes reversible folding into two distinct native states: inactive and active. The model developed in this work captures the transition between these two states.

Lista de Figuras

- 1.0.0.1 Representação dos níveis de estrutura nas proteínas. A estrutura primária consiste em uma sequência de aminoácidos unidos por ligações peptídicas, incluindo quaisquer pontes dissulfeto. A estrutura secundária destaca a formação das hélices- α e das folhas- β . A estrutura terciária mostra o arranjo tridimensional da proteína, enquanto a estrutura quaternária evidencia a associação de subunidades proteicas. 27
- 1.2.0.1 Representação do funil de energia do enovelamento da proteína SH3.¹⁶ O processo de enovelamento ocorre por meio da organização progressiva de conjuntos de estruturas (à esquerda) em um funil de energia livre (à direita). No topo do funil, encontram-se as estruturas desenoveladas, com alta entropia conformacional, a qual é compensada pela energia livre adquirida à medida que as interações nativas são formadas. Figura adaptada de Brooks III et al. (2001).¹⁶ 32
- 1.3.0.1 Representação estrutural de osmólitos compatíveis (nomeados em azul) e não compatíveis (nomeados em vermelho). As três maiores classes de osmólitos compatíveis são: 1) aminoácidos e seus derivados, como a prolina, taurina e o ácido aspártico; 2) sais de metilamônio, como o TMAO, colina e glicina betaína; e os 3) açúcares e poliálcoois, como a glicose, sorbitol e o glicerol. Os osmólitos não compatíveis mais comuns são a ureia e o GdmCl. 33
- 1.3.0.2 Ilustração esquemática dos mecanismos mais prováveis de estabilização de proteínas por osmólitos protetores (lado esquerdo) e de desestabilização por desnaturantes (lado direito). Figura adaptada de Rani & Venkatesu (2018).⁶ . . . 35

2.1.1.1	Representação da MDDF de um solvente em relação à uma proteína em função da distância r . A figura mostra a região da distância de exclusão, das camadas de solvatação e da região de <i>bulk</i> . Na região <i>bulk</i> , o soluto não interfere mais na estrutura do solvente, e a MDDF converge para 1, como indicado pela linha cinza.	37
2.1.2.1	Ilustração do experimento de diálise no equilíbrio. A figura mostra duas soluções separadas por uma membrana permeável (retângulo laranja) ao cossolvente (triângulos pretos) mas impermeável à proteína (círculos cinzas). Figura adaptada de Pinger et al. (2017). ³⁴	40
2.1.3.1	Representação do ciclo termodinâmico a partir do modelo de transferência de Tanford. ^{39,41} Esse ciclo ilustra como as energias livres de enovelamento entre os estados nativo (N) e desenovelado (U), na presença de cossolvente ($N_{\text{cos}} \rightleftharpoons U_{\text{cos}}$, representado por $\Delta\mu_{N \rightarrow U}^{\text{cos}}$), diferem daquelas em um tampão aquoso ($N_{\text{aq}} \rightleftharpoons U_{\text{aq}}$, representado por $\Delta\mu_{N \rightarrow U}^0$). Medidas diretas das constantes de equilíbrio com e sem o cossolvente (reações na horizontal) permitem o cálculo do efeito do cossolvente na energia livre de enovelamento, $\partial\Delta\mu_{p, tr} = \Delta\mu_{N \rightarrow U}^{\text{cos}} - \Delta\mu_{N \rightarrow U}^0$. Alternativamente, as energias livres de transferência para os estados nativo ($\Delta\mu_{p, tr}^N$) e desenovelado ($\Delta\mu_{p, tr}^U$) podem ser obtidas (reações na vertical), e $\partial\Delta\mu_{p, tr} = \Delta\mu_{p, tr}^U - \Delta\mu_{p, tr}^N$.	43
2.2.1.1	Representação esquemática do algoritmo de Leap-Frog para o cálculo das posições e velocidades. No algoritmo de Leap-Frog, a velocidade é explicitamente calculada no intervalo de tempo $\Delta t/2$.	51
2.2.3.1	Esquema da simulação de troca de réplicas Hamiltonianas com têmpera de soluto (REST2). ⁷⁰	55

3.0.0.1	Modos de ligação de hidrogênio da A) água e do B) TFE com a cadeia principal do peptídeo. As moléculas de água atuam tanto como doadoras quanto comoceptoras de ligações de hidrogênio, causando o rompimento das ligações de hidrogênio intramoleculares do peptídeo. Por outro lado, o TFE é um forte doador de ligações de hidrogênio, mas um fracoceptor. ⁷⁴ Como resultado, foi proposto que o hidrogênio ácido do TFE pode interagir com o oxigênio da cadeia principal, ao mesmo tempo que protege as ligações de hidrogênio intramoleculares do peptídeo (N-H...O) que estabilizam a hélice. ^{71,88} Figura adaptada do artigo de revisão de Vincenzi et al. (2019). ⁸⁸	58
3.2.1.1	Propriedades estruturais do peptídeo AAQAA ₃ em água (azul) e em soluções aquosas de TFE, calculadas a partir das simulações: A) Espectros de CD. O TFE aumenta o conteúdo de α -hélice do peptídeo, como mostrado pelo aumento da intensidade da banda em \sim 190 nm e pelo duplo aumento entre 200-230 nm. B) Prevalência média de α -hélice por resíduo. C) Densidade de probabilidade do conteúdo helicoidal do peptídeo. D) Elipticidade média do peptídeo em cada solução. As barras de erro nas figuras indicam o erro padrão da média das quantidades calculadas.	62
3.2.2.1	MDDFs A) da água e B) do TFE em função da concentração do cossolvente. As IKBs para água e TFE são mostradas em C) e D), respectivamente.	64
3.2.2.2	MDDF do TFE em relação ao peptídeo AAQAA ₃ e contribuições dos grupos, na solução de 60% vol/vol. A) Contribuição dos átomos de TFE. B) Contribuições das cadeias laterais do peptídeo. C) Contribuições dos átomos da cadeia principal. A única contribuição dos átomos da cadeia principal para a MDDF do TFE é devido ao átomo de oxigênio, que forma ligações de hidrogênio com o hidrogênio hidroxílico do TFE.	67
3.2.3.1	Propriedades estruturais do peptídeo AAQAA ₃ em água (azul) e em 80% de TFE, computadas a partir da simulação padrão (verde) e com potencial modificado (verde tracejada): A) Espectros de CD. B) Prevalência média de α -hélice por resíduo.	68

3.2.4.1 MDDFs A) da água e B) do TFE em função da concentração do cossolvente para as simulações com potenciais modificados. As IKBs para água e TFE são mostradas em C) e D), respectivamente.	70
3.2.5.1 MDDFs entre o peptídeo e A) água e B) em 20% de TFE para os conjuntos helicoidal (vermelho) e não helicoidal (preto). O conjunto desenovelado forma mais ligações de hidrogênio com a água e apresenta menos interações não específicas com TFE.	71
4.0.0.1 Representação em cartoon do domínio B da proteína A (BdpA), PDB: 1BDD. ¹³⁶ A Hélice I (¹⁰ QQNAFYEILH ¹⁹), a Hélice II (²⁵ EEQRNGFIQSLKD ³⁷) e a Hélice III (⁴² SANLLAEAKKLNDAQ ⁵⁶) são representadas em azul, verde e vermelho, respectivamente. Na parte inferior, tem-se a sequência da BdpA colorida de acordo com a representação em cartoon.	75
4.1.4.1 Pipeline para estudar o enovelamento da BdpA. As etapas iniciais (preto) validaram o enovelamento da BdpA com base em trabalhos anteriores, enquanto que os passos incrementais (verde) permitiram investigar a relação entre o enovelamento da BdpA, a formação de sua estrutura secundária e a estrutura do solvente em torno de estados parcialmente enovelados.	78
4.2.1.1 Caracterização do enovelamento da BdpA. A) Calor específico (C_v) em função da temperatura, permitindo a identificação da temperatura de enovelamento ($T_f = 0,97$ unidades reduzidas). B) Energia livre em função da fração de contatos nativos (Q). A partir da simulação realizada em T_f : C) Fração de contatos nativos (Q) em função do tempo da simulação (τ em unidades reduzidas). D) Mapas de contorno da densidade de probabilidade (PD, <i>Probability Density</i>) em função de Q e RMSD.	80
4.2.1.2 Projeção do conjunto de enovelamento da BdpA obtida com o ELViM. Cada estrutura é representada como um ponto, com uma cor associada à fração de contatos nativos (Q): regiões em amarelo e em roxo representam os estados nativos (Q alto) e desenovelados (Q baixo), respectivamente.	81

4.2.2.1	Conteúdo de α -hélice projetado nas estruturas de enovelamento. A cor em cada gráfico é uma função do conteúdo de: A) α -Hélice total da proteína, B) α -Hélice I (¹⁰ QQNAFYEILH ¹⁹), C) α -Hélice II (²⁵ EEQRNGFIQSLKD ³⁷), D) α -Hélice III (⁴² SANLLAEAKKLND ⁵⁶).	82
4.2.2.2	Mapas de contorno da PD das estruturas da BdpA com A) >75%-100%, B) >50%-75%, C) >25%-50% e D) 0%-25% de α -hélice. O conteúdo de α -hélice total é calculado a partir da soma das três hélices da BdpA.	83
4.2.2.3	Mapas de contorno da PD dos estados em função dos conteúdos de A) α -Hélice I e α -Hélice II, B) α -Hélice I e α -Hélice III, e A) α -Hélice II e α -Hélice II da BdpA.	84
4.2.3.1	A) MDDFs da água para os subconjuntos de estruturas. B) Conjuntos de estruturas obtidas a partir da Figura 4.2.2.2. C) IKBs correspondentes para água. D) Diferença na densidade das MDDFs em torno dos estados nativo (N_{84}^{B0}) e desenovelado (U_{37}^{B11}). As cores em vermelho indicam uma maior densidade da água em torno do estado nativo (N_{84}^{B0}), enquanto que as cores em azul estão associadas a maiores densidades de água em torno do estado desenovelado (U_{37}^{B11}). O último caso destaca as interações do solvente com resíduos predominantemente hidrofóbicos, que estão protegidos do solvente no estado nativo.	85

5.2.1.1 Conjunto de enovelamento da SH3 e da BpdA. A) e B) Representação em cartoon do domínio SH3 (PDB: 1FMK,¹⁷⁶ resíduos de T84 à S140) e do domínio B da proteína A (BdpA, PDB: 1BDD).¹³⁶ C) e D) Conjuntos de enovelamento SBM caracterizados por: i) a projeção ELViM, ii) o calor específico (C_v) em função da temperatura, e iii) a fração de contatos nativos (Q) em função do tempo, em unidades reduzidas (τ). Cada ponto em i) representa uma estrutura colorida em uma escala de Q, variando de 0 (estado desnaturado) a 1 (estado enovelado). E) Projeção do enovelamento da SH3 colorida pelo conteúdo de estrutura secundária e definição dos subconjuntos conformacionais. Há uma alta correlação entre Q (painel C) e o conteúdo de estrutura secundária. F) Dimensão da estrutura secundária na projeção 2D para BpdA, necessária para classificar os estados desnaturados.¹⁷⁷ Nas figuras E) e F), os rótulos são atribuídos aos conjuntos dos estados nativos (N) e desenovelados (U), com um subscrito indicando o conteúdo médio de estrutura secundária e um sobrescrito indicando a ordem decrescente da fração média de contatos nativos. 95

5.2.2.1 Estruturas de solvatação do conjunto de enovelamento do domínio SH3 em soluções aquosas de ureia 0,5 mol L⁻¹. A) MDDFs da água e da ureia para os conjuntos N_{80}^{50} e $U_{1.0}^{512}$. B) e C) Picos mais proeminentes na MDDF da ureia e da água. D) IKBs e E) Parâmetros de interação preferencial (Γ). A interação preferencial da ureia pelos estados desenovelados é maior do que pelo estado nativo (N_{80}^{50}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF da ureia na vizinhança do estado N_{80}^{50} e do estado desenovelado ($U_{1.0}^{512}$). Regiões vermelhas indicam maior densidade de ureia próxima ao estado N_{80}^{50} , enquanto regiões azuis mostram maior densidade ao redor do estado $U_{1.0}^{512}$ 98

- 5.2.2.2 Parâmetros de integração preferencial (Γ) em função de SASA (A e C) e de Q (B e D). Os pontos representados em vermelho e azul correspondem aos parâmetros das proteínas SH3 e BdpA, respectivamente. Um ajuste quadrático captura de forma eficaz a correlação entre Q e Γ (coeficiente de correlação de Pearson, $R = 0,966$), o conteúdo de folhas- β e Γ ($R = 0,947$) e a SASA e Γ ($R = 0,966$) nos *ensembles* do domínio SH3. Como esperado, Q e SASA, assim como Q e o conteúdo de folhas- β , também estão correlacionados ($R = 0,983$ e $R = 0,993$, respectivamente). 100
- 5.2.2.3 Estruturas de solvatação do conjunto de enovelamento do domínio SH3 em soluções aquosas de TMAO 0,5 mol L⁻¹. A) MDDFs da água e do TMAO para os conjuntos N₈₀^{S0} e U_{0.9}^{S9}. B) e C) Picos mais proeminentes na MDDF do TMAO e da água. D) IKBs e E) Parâmetros de interação preferencial. A exclusão do TMAO nos estados desenovelados é maior do que no estado nativo (N₈₀^{S0}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF do TMAO na vizinhança do estado N₈₀^{S0} e do estado desenovelado (U_{0.9}^{S9}). Regiões vermelhas indicam maior densidade de TMAO próxima ao estado N₈₀^{S0}, enquanto regiões azuis mostram maior densidade ao redor do estado U_{0.9}^{S9}. 101
- 5.2.2.4 MDDF total do TMAO e contribuições de grupo para os *ensembles* N₈₀^{S0} e U_{0.9}^{S9}. A) e C) mostram as contribuições dos átomos e grupos de átomos para a MDDF do TMAO nos *ensembles* N₈₀^{S0} e U_{0.9}^{S9}. B) e D) mostram as respectivas contribuições da cadeia principal (amarelo) e da cadeia lateral (verde) para a MDDF total do TMAO. 103
- 5.2.3.1 Estruturas de solvatação do conjunto de enovelamento da BdpA em soluções aquosas de ureia 0,5 mol L⁻¹. A) MDDFs de água e ureia para os *ensembles* N₈₄^{B0} e U₃₇^{B11}. B) e C) picos mais proeminentes nas MDDFs de ureia e água. D) IKBs e E) parâmetros de interação preferencial. A interação preferencial da ureia para os estados desenovelados é maior do que para o estado nativo (N₈₄^{B0}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF da ureia por resíduo de N₈₄^{B0} e U₆₃^{B1}. Vermelho indica maior densidade ao redor de N₈₄^{B0} e azul ao redor de U₆₃^{B1}. 105

5.2.3.2	Estruturas de solvatação da BdpA em soluções aquosas de TMAO 0,5 mol L ⁻¹ . A) MDDFs de água e ureia para os <i>ensembles</i> N ₈₄ ^{B0} e U ₃₇ ^{B11} . B) e C) mostram os picos mais proeminentes nas MDDFs do TMAO e da água. D) IKBs e E) parâmetros de interação preferencial. A interação preferencial do TMAO para os estados desenovelados é menor do que para o estado nativo (N ₈₄ ^{B0}) e diminui nos estados desenovelados. F) Mapa da diferença na densidade da MDDF por resíduo de N ₈₄ ^{B0} e U ₃₇ ^{B11} . Vermelho indica maior densidade ao redor de N ₈₄ ^{B0} e azul ao redor de U ₃₇ ^{B11} .	107
5.2.4.1	$\Delta\Gamma$ para cada conjunto conformacional de estruturas das proteínas SH3 e BdpA, em relação ao <i>ensemble</i> nativo. Os painéis A) e B) mostram os valores de $\Delta\Gamma$ para cada conjunto conformacional da SH3 em ureia e TMAO a 0,5 mol L ⁻¹ . Da mesma forma, os painéis C) e D) mostram os valores de $\Delta\Gamma$ para cada conjunto conformacional da BdpA nas mesmas condições de cossolvente.	109
5.2.5.1	Variação das energias livres de enovelamento ($\partial\Delta\mu_{p,tr}$) em relação à água pura, em função da concentração de cossolvente, para estados desenovelados selecionados. A) e B) mostram $\partial\Delta\mu_{p,tr}$ para os modelos de SH3 em ureia e TMAO, respectivamente. Esta análise focou nos conjuntos desenovelados representativos U ₄₆ ^{S1} (em azul) e U _{1.0} ^{S12} (em vermelho). C) e D) mostram $\partial\Delta\mu_{p,tr}$ para BdpA para o equilíbrio de enovelamento associado aos conjuntos U ₆₃ ^{B1} (em azul) e U ₃₇ ^{B11} (em vermelho) de BdpA.	112
6.1.2.1	A) Representação em cartoon da MAD2 inativa alinhada à MAD2 ativa, com as regiões de mudança conformacional destacadas em azul e vermelho, respectivamente. Os painéis i., ii. e iii. mostram as perspectivas frontal e lateral, rotacionadas conforme indicado. B) Raiz da flutuação quadrática média (RMSF) da estrutura inativa comparada à ativa após alinhamento com o LovoAlign. ²¹⁹ Todos os resíduos com RMSF superior a 2,5 Å estão destacados em azul. C) Representação em cartoon da estrutura inativa colorida em função do RMSF.	120

- 6.2.1.1 A proteína MAD2 nas formas A) inativa (O-MAD2, PDB: 1DUJ)²¹¹ e B) ativa (C-MAD2, PDB: 1S2H).²¹² Às regiões responsáveis pelo mecanismo de *fold-switching* em A) e B) estão coloridas em azul e vermelho, respectivamente. C) Mapas de contato para as estruturas inativa (canto inferior direito) e ativa (canto superior esquerdo). Os contatos comuns e exclusivos de cada estrutura são representados por pontos amarelos, azuis e vermelhos. D) Comparação da estrutura secundária das conformações inativa (superior) e ativa (inferior). 124
- 6.2.2.1 Caracterização do enovelamento da proteína MAD2. A) Calor específico (C_v) em função da temperatura, permitindo a identificação da temperatura de mudança de conformação ($T_s = 0,895T_f$) e da temperatura de enovelamento ($T_f = 1,0$, por definição). B) Energia livre em função da fração de contatos formados (Q/Q^*) em T_s , T_f . Q e Q^* representam o número de contatos e a média do número de contatos nativos ativos e inativos, respectivamente. Densidade de estados em função de Q_{inativo} e Q_{ativo} em: C) $T_s = 0,895T_f$, D) T_f e E) $T > T_f$ ($T = 1,05T_f$). F) Estruturas representativas para cada bacia em T_s (ou T_f). 126
- 6.2.2.2 A) Agrupamento das estruturas ativas, $\beta 1$ -ativas e inativas. B) Transições entre as bacias de enovelamento em função do tempo (τ , em unidades reduzidas). C) Percentuais de transições visando o estado inativo (esquerda) e originadas do estado inativo (direita). Mais de 90% das transições envolvem o estado $\beta 1$ -ativo como intermediário. Mapas de contato para as estruturas inativa (triângulo inferior) e ativa (triângulo superior) coloridos pela probabilidade de contatos formados em $T_s = 0,895T_f$. Esses mapas correspondem às bacias D) ativas, E) $\beta 1$ -ativas e F) inativas das estruturas mostradas no painel A. As áreas sombreadas correspondem a regiões de contatos exclusivos de qualquer conformação nativa. Contatos sombreados em marrom são satisfeitos e em cinza não são satisfeitos em cada bacia: o intermediário exhibe contatos N-terminais da conformação inativa e contatos C-terminais da conformação ativa. 128

6.2.3.1	Análise de coevolução da proteína MAD2. Os 50, 150 e 250 principais contatos previstos (pontos pretos) são comparados com os contatos calculados a partir das estruturas experimentais inativas (parte inferior direita, PDB: 1DUJ) ²¹¹ e ativas (parte superior esquerda, PDB: 1S2H) ²¹² nas figuras A), B) e C), respectivamente. Contatos comuns entre MAD2 inativa e ativa são mostrados em amarelo, enquanto os contatos exclusivos para cada uma são representados em azul e vermelho, respectivamente. D) Representações em estilo cartoon das conformações ativa e inativa são mostradas nos lados esquerdo e direito. Os 250 principais contatos previstos entre resíduos nas regiões de mudança de conformação são indicadas por linhas pretas.	130
6.2.4.1	Análises de frustração na proteína MAD2. A) Distribuição do SRFI para as conformações ativa (barras vermelhas) e inativa (barras azuis). Linhas verticais pretas indicam o raio de corte usado para definir resíduos minimamente frustrados (MF, $SRFI > 0,55$), neutros ($-1 < SRFI < 0,55$) ou altamente frustrados (HF, $SRFI < -1$). B) Correlação entre o SRFI nas conformações inativa (eixo y) e ativa (eixo x). Pontos verdes representam os resíduos MF (29,73%), enquanto os pontos vermelhos representam resíduos HF (5,95%) em ambas as conformações. Pontos cinzas representam resíduos neutros em ambas as conformações ou neutros em uma e MF na outra (54,6%). SRFIs para resíduos HF presentes apenas na estrutura inativa (ativa) são mostrados em magenta (laranja). Resíduos HF envolvidos nas regiões de <i>fold-switching</i> (FS) são destacados com um contorno preto. Os painéis C) e D) mostram a representação em estilo cartoon e o gráfico de barras dos resíduos HF no estado ativo e inativo, respectivamente.	132
I.0.0.1	Estruturas representativas de cada <i>ensemble</i> das proteínas SH3 e BdpA.	166

Lista de Tabelas

3.1.1.1 Composições dos sistemas simulados com diferentes concentrações de 2,2,2-Trifluoroetanol (TFE).	60
3.2.2.1 Concentrações efetivas das soluções de TFE, IKBs para o TFE (G_{pc}) e água (G_{pw}) em relação ao peptídeo AAQAA ₃ , e parâmetro de interação preferencial (Γ_{pc}). As simulações com os potenciais modificados são destacadas por um asterisco	66
1.0.0.1 Detalhes das caixas de simulação para cada <i>ensemble</i> da proteína SH3 em soluções de ureia e TMAO 0,5 mol L ⁻¹	161
1.0.0.2 Detalhes das caixas de simulação para cada <i>ensemble</i> da proteína BdpA em soluções de ureia e TMAO 0,5 mol L ⁻¹	162
1.0.0.3 Detalhes das caixas de simulação para cada estrutura representativa da proteína SH3 em soluções de 0,1; 0,2; 0,3; 0,4; 0,5; 0,75; e 1,0 mol L ⁻¹ de ureia e TMAO.	163
1.0.0.4 Detalhes das caixas de simulação para cada estrutura representativa da proteína BdpA em soluções de 0,1; 0,2; 0,3; 0,4; 0,5; 0,75; e 1,0 mol L ⁻¹ de ureia e TMAO.	164
1.0.0.5 Fração de contatos nativos (Q), área superficial acessível ao solvente (SASA, <i>Solvent-Accessible Surface Area</i>), conteúdo de estrutura secundária (folha- β (SH3) e α -hélice (BdpA)), e parâmetros de interação preferencial da ureia e do TMAO para cada <i>ensemble</i> das proteínas SH3 e BdpA.	165

Lista de Abreviaturas

AAs	Aminoácidos.
ELT	Em inglês, <i>Energy Landscape Theory</i> , ou teoria da paisagem energética.
TMAO	Em inglês, <i>Trimethylamine N-oxide</i> , ou N-óxido de trimetilamina.
GdmCl	Em inglês, <i>Guanidinium Chloride</i> , ou cloreto de guanidina.
MDDF	Em inglês, <i>Minimum-Distance Distribution Function</i> , ou funções de distribuição de mínima distância.
RDF	Em inglês, <i>Radial Distribution Function</i> , ou funções de distribuição radial.
IKB	Integral de Kirkwood-Buff.
Γ	Parâmetro de interação preferencial.
$\Delta\Gamma$	Parâmetro da relação de ligação de Wyman.
DM	Dinâmica Molecular.
AMBER	Em inglês, <i>Assisted Model Building and Energy Refinement</i> .
CHARMM	Em inglês, <i>Chemistry at HARvard Macromolecular Mechanics</i> .
OPLS	Em inglês, <i>Optimized Potentials for Liquid Simulations</i> .
SBM	Em inglês, <i>Structure-Based Model</i> , ou modelos baseados na estrutura.
WHAM	Em inglês, <i>Weighted Histogram Analysis Method</i> .
ELViM	Em inglês, <i>Energy Landscape Visualization Method</i> .

TREM	Em inglês, <i>Temperature Replica Exchange Method</i> , ou método de troca de réplicas por temperatura.
REST2	Em inglês, <i>Replica Exchange with Solute Tempering</i> , ou método de troca de réplicas Hamiltonianas com têmpera de soluto.
TFE	2,2,2-Trifluoroetanol.
PME	Em inglês, <i>Particle mesh Ewald</i> .
VMD	Em inglês, <i>Visual Molecular Dynamics</i> .
CD	Em inglês, <i>Circular Dichroism</i> , ou dicroísmo circular.
DSSP	Em inglês, <i>Dictionary of Secondary Structure in Proteins</i> .
RMSD	Em inglês, <i>Root Mean Square Deviation</i> .
RMSF	Em inglês, <i>Root Mean Square Fluctuation</i> .
Q	Fração de contatos nativos.
SASA	Em inglês, <i>Solvent-Accessible Surface Area</i> , ou área superficial acessível ao solvente.
PD	Em inglês, <i>Probability Density</i> .
C_v	Calor específico à volume constante.
T_f	Em inglês, <i>Folding Temperature</i> .
T_s	Em inglês, <i>Fold-switch Temperature</i> .
MSA	Em inglês, <i>Multiple Sequence Alignment</i> , ou alinhamento múltiplo de sequências.
DCA	Em inglês, <i>Direct coupling analysis</i> , ou análise de acoplamento direto
HMM	Em inglês, <i>Hidden Markov Model</i> , ou modelo oculto de Markov
FS	Em inglês, <i>Fold-switching</i> .
smFRET	Em inglês, <i>Single-molecule Fluorescence Resonance Energy Transfer</i> .

Sumário

1	Contextualização	26
1.1	Enovelamento de proteínas	28
1.2	Teoria do enovelamento de proteínas	29
1.3	Interações entre osmólitos e proteínas	32
2	Metodologia	36
2.1	Caracterização termodinâmica das estruturas do solvatação	36
2.1.1	Funções de Distribuição de Mínima-Distância	36
2.1.2	Teoria de soluções de Kirkwood-Buff	38
2.1.3	Energia livre de transferência	43
2.2	Simulações de Dinâmica Molecular	46
2.2.1	Integração das equações de movimento	49
2.2.2	Simulações com os Modelos Baseados na Estrutura	52
2.2.3	Simulações de Dinâmica Molecular por troca de réplicas	54
3	Efeitos do 2,2,2-Trifluoroetanol em α-hélices	57
3.1	Metodologia	59
3.1.1	Simulações de Dinâmica Molecular	59
3.1.2	Cálculo da elipticidade do peptídeo	61
3.1.3	Funções de Distribuição de Mínima-Distância e termodinâmica de solvatação	61
3.2	Resultados e discussão	61
3.2.1	TFE estabiliza a estrutura helicoidal do peptídeo AAQAA ₃	61

3.2.2	Interações diretas e interação preferencial por TFE	63
3.2.3	Rompimento da interação entre o TFE e a cadeia principal do peptídeo leva à estabilização da α -hélice	68
3.2.4	Interação preferencial na ausência da ligação de hidrogênio entre o TFE e a cadeia principal	69
3.2.5	Interações do TFE com os <i>ensembles</i> helicoidais e não helicoidais .	71
3.3	Conclusões	72
4	Enovelamento da BdpA e estruturas de hidratação	74
4.1	Metodologia	75
4.1.1	Simulações com os Modelos Baseados na Estrutura e análises . . .	75
4.1.2	Enovelamento da BdpA no espaço de fase 2D	77
4.1.3	Análise da estrutura secundária a partir das estruturas contendo todos os átomos	77
4.1.4	Base molecular da hidratação	78
4.2	Resultados e discussão	79
4.2.1	Enovelamento da BdpA em um espaço de fase 2D	79
4.2.2	Formação de hélices no conjunto de enovelamento da BdpA	81
4.2.3	Hidratação em estados enovelados e desenovelados	84
4.3	Conclusões	87
5	Efeitos de osmólitos no enovelamento de proteínas	88
5.1	Metodologia	90
5.1.1	Simulações com os Modelos Baseados na Estrutura e análises . . .	90
5.1.2	Espaço de fase do enovelamento de proteínas	91
5.1.3	Simulações atomísticas e análises	91
5.1.4	Caracterização da estrutura e termodinâmica do solvente	92
5.2	Resultados	93
5.2.1	Paisagens de enovelamento de proteínas modelo	93
5.2.2	Efeitos dos osmólitos nos estados enovelados e desenovelados da SH3	97
5.2.3	Efeitos dos osmólitos nos estados enovelado e desenovelado da BdpA105	

5.2.4	Efeitos nas constantes de equilíbrio de desnaturação induzidos por osmólitos	109
5.2.5	Dependência das energias livres de desnaturação em relação aos osmólitos para estados desnaturados selecionados	110
5.2.6	Discussão	113
6	Mecanismos de mudança conformacional da proteína MAD2	116
6.1	Metodologia	118
6.1.1	Modelo baseado na estrutura	118
6.1.2	Construção do potencial <i>dual-basin</i> SBM	118
6.1.3	Detalhes da simulação	120
6.1.4	Alinhamento Múltiplo de Sequências e Análise de Acoplamento Direto	121
6.1.5	Análises de Frustração	122
6.2	Resultados e Discussão	122
6.2.1	Estrutura da MAD2 e características do <i>dual-basin</i> SBM	123
6.2.2	Paisagem energética da proteína MAD2 e o intermediário β 1-ativo	124
6.2.3	Contatos coevolutivos suportam um núcleo conservado e uma troca lábil da região N-terminal	128
6.2.4	Frustração na proteína MAD2	130
6.3	Conclusões	133
7	Conclusões gerais	134
	Bibliografia	136
		161
I	Apêndice I	161

Capítulo 1

Contextualização

As proteínas constituem uma classe de biomoléculas essenciais para a manutenção dos processos biológicos, participando de praticamente todos os mecanismos responsáveis pela sobrevivência dos organismos vivos.¹ Essas biomoléculas são constituídas por sequências de aminoácidos (AAs), definidas pela evolução, que se auto-organizam formando as estruturas secundárias, como as hélices- α e as folhas- β . Entretanto, em muitos casos, a função biológica da proteína está intimamente relacionada à sua estrutura terciária, originada pelo arranjo tridimensional de todos os átomos no espaço.² Esse arranjo tridimensional funcional é denominado de estado nativo que, de acordo com a hipótese de Anfinsen, corresponde ao mínimo de energia livre de Gibbs.³ Entretanto, algumas estruturas terciárias se agregam a outras subunidades proteicas para encontrar o seu estado nativo. Nesses casos, a formação de dímeros ou outras estruturas complexas compõem a estrutura quaternária da proteína (Figura 1.0.0.1).

Tendo em vista que, na maioria das vezes, as funções das proteínas estão intimamente relacionadas à sua estrutura tridimensional, um problema central na biologia molecular é o entendimento do processo pelo qual uma sequência de aminoácidos adquire sua estrutura enovelada. A resolução desse problema físico-químico é complexa e exige a compreensão detalhada dos mecanismos envolvidos nessa transição conformacional, bem como das interações não covalentes envolvidas. Essas interações dividem-se em específicas - predominantemente eletrostáticas - e não específicas, como as interações hidrofóbicas e de van der Waals, que desempenham papéis fundamentais no processo de enovelamento.⁴ Todavia, a natureza e magnitude dessas interações dependem de condições físicas - como pressão e temperatura - e químicas - como concentrações de cossolventes - nas quais as estruturas estão envolvidas.

Os cossolventes são predominantes moléculas orgânicas de baixo peso molecular, que estão presentes até mesmo na célula, e podem auxiliar tanto na modulação da estabilidade quanto no rompimento das interações que mantêm a estrutura.^{1,5} Em certos contextos, os cossolventes são denominados de osmólitos, tendo em vista que são utilizados pelas células de muitos organismos para conter o estresse osmótico.⁶ Muitos organismos experienciam o estresse osmótico de diferentes formas, como desidratação celular, alta concentração extracelular de sal e variações de pH, temperatura e pressão. Dadas essas particularidades dos cossolventes, uma descrição detalhada dos seus efeitos no enovelamento de proteínas pode guiar a compreensão dos fatores termodinâmicos envolvidos nesse mecanismo, visando otimizar suas aplicações na indústria farmacêutica e biotecnológica.

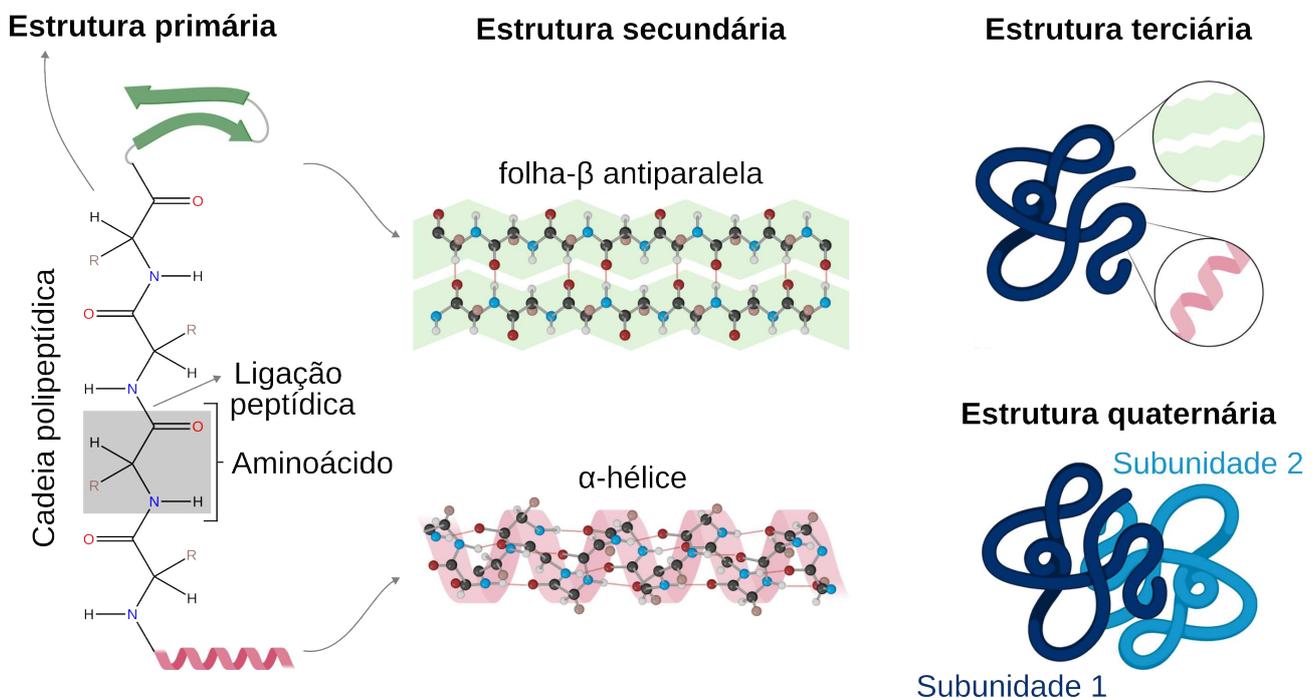


Figura 1.0.0.1: Representação dos níveis de estrutura nas proteínas. A estrutura primária consiste em uma sequência de aminoácidos unidos por ligações peptídicas, incluindo quaisquer pontes dissulfeto. A estrutura secundária destaca a formação das hélices- α e das folhas- β . A estrutura terciária mostra o arranjo tridimensional da proteína, enquanto a estrutura quaternária evidencia a associação de subunidades proteicas.

1.1 Enovelamento de proteínas

Proteínas são macromoléculas que desempenham um papel fundamental em praticamente todos os processos biológicos. A forma funcional da proteína está, na maioria das vezes, associada a uma estrutura tridimensional bem definida, conhecida como estado nativo. Do ponto de vista termodinâmico, a estabilidade de uma proteína é determinada pela diferença de energia livre entre os diferentes estados que ela pode assumir, dependendo das condições ambientais.⁴ Em um modelo simples, no qual apenas os estados nativo (N, *native*) e desenovelado (U, *unfolded*) estão em equilíbrio,



a diferença de energia livre associada ao processo de desenovelamento é dada pela Equação 1.1.0.2.

$$\Delta G = -RT \ln \left(\frac{U}{N} \right) = -RT \ln(K_{\text{eq}}) = \Delta H - T\Delta S \quad (1.1.0.2)$$

em que ΔG , ΔH e ΔS são as mudanças de energia livre, entalpia e entropia, respectivamente. R é a constante universal dos gases, T é a temperatura (em Kelvin) e K_{eq} é a constante de equilíbrio entre os estados nativo e desenovelado.

A variação de energia livre é uma medida quantitativa que define a população dos estados nativo e desenovelado, cujo equilíbrio entre eles está na ordem de alguns $\text{kcal}\cdot\text{mol}^{-1}$.¹ As contribuições entálpicas e entrópicas são da ordem de centenas de $\text{kcal}\cdot\text{mol}^{-1}$, mas contribuem para ΔG em direções opostas.^{1,6} A variação de entalpia é primeiramente associada às ligações de hidrogênio, interações eletrostáticas, forças de van der Waals, interações hidrofóbicas, e ligações de dissulfeto. Portanto, essas interações são mais prevalentes no estado enovelado do que no estado desenovelado, resultando em uma variação de entalpia negativa para a reação de desenovelamento. Por outro lado, a entropia conformacional é maior no estado desenovelado, resultando em uma contribuição positiva para a energia livre.

Além disso, a estabilização da proteína se estende por interações entre a cadeia polipeptídica e o solvente. No estado enovelado, os resíduos hidrofóbicos formam interações de

van der Waals favoráveis e se protegem da água (as interações em si são fracas), enquanto que os resíduos carregados e polares encontram-se predominantemente na superfície da proteína. Uma exceção são os grupos polares da cadeia principal da proteína (CO e NH), que foram ligações de hidrogênio intracadeia e dão origem aos elementos de estrutura secundária, como as hélices- α e as folhas- β . No estado desenovelado, a proteína expõe a área superficial de resíduos não polares das cadeias laterais para as moléculas de água, e diminui a rede de ligações de hidrogênio formadas com a água, resultando em um estado energético desfavorável. Para minimizar esse efeito, as moléculas de água precisam se reorientar para formar ligações de hidrogênio entre elas. No geral, isso resulta em uma perda de energia do sistema e de entropia em torno da proteína.¹ Uma forma de neutralizar a penalização entrópica deste efeito se dá pelas interações favoráveis entre os resíduos da proteína. A consequência dessas interações é o empacotamento da cadeia polipeptídica em torno dos núcleos hidrofóbicos por meio de forças dispersivas e o rompimento de interações com as moléculas de água previamente organizadas em torno desses núcleos. Esse fenômeno, conhecido como efeito hidrofóbico,⁴ é essencialmente de natureza entrópica e está correlacionado com a quantidade da área superficial acessível ao solvente com o enovelamento.

1.2 Teoria do enovelamento de proteínas

Na década de 60, os trabalhos de Christian Anfinsen e colaboradores tiveram um papel importante para o estudo do enovelamento de proteínas.^{7,8} Um de seus estudos, que envolvia uma enzima chamada ribonuclease, mostrou que a adição de agentes desnaturantes remove completamente a atividade da enzima.³ Entretanto, eles perceberam que após a remoção dos agentes desnaturantes (por diálise), a atividade enzimática era recuperada. Ou seja, a enzima se enovelou espontaneamente após serem restabelecidas as condições fisiológicas. A partir desse estudo, Anfinsen concluiu que a sequência de aminoácidos contém a informação necessária para caracterizar o estado nativo. Além disso, estabeleceu-se a “hipótese termodinâmica”, sugerindo que o estado nativo é o estado com energia livre mínima no ambiente fisiológico.³ Este trabalho foi reconhecido com o Prêmio Nobel de Química em 1972.

Embora o enovelamento de proteínas seja um processo físico-químico complexo, esse mecanismo ocorre em uma escala de tempo relativamente curta (microsegundos a segundos),

comparada com as inúmeras possíveis conformações que, supostamente, a proteína poderia adotar. Em 1969, Levinthal⁹⁻¹¹ argumentou que não era possível uma proteína vasculhar todo o espaço conformacional na busca pelo menor valor de energia livre para que se enovelasse. Por exemplo, vamos supor que a proteína tem N resíduos e, para cada um dos resíduos, μ orientações no espaço. Logo, o número de conformações distintas que podem ser obtidas é igual a μ^N . Considerando um caso genérico de uma proteína com 100 aminoácidos, e que cada resíduo tem apenas três possibilidades de conformação, o número de conformações totais seria igual a $3^{100} = 5 \times 10^{47}$. Além disso, vamos supor que a mudança de conformação da cadeia pode ocorrer na escala de tempo de picossegundos. Mesmo assim, um único evento de enovelamento levaria aproximadamente $5 \times 10^{47} \times 1 \text{ ps} \approx 2.5 \times 10^{50}$ anos. Esse argumento deixa claro que a busca aleatória pela conformação nativa não é correta.

A resolução desse paradoxo foi proposta a partir da análise cinética do processo de enovelamento. Em 1992, Zwanzig, Szabo e Bagchi demonstraram, por meio de um modelo estatístico, que ao introduzir um viés energético contra as conformações incorretas, os tempos de enovelamento se tornavam compatíveis com os observados experimentalmente.¹¹ Posteriormente, uma outra abordagem baseada nesse argumento, porém com conceitos matemáticos mais simples, captura os aspectos fundamentais do tempo de enovelamento e fornece uma explicação para o limite de tamanho das proteínas reais.¹⁰ Essas abordagens sugerem que o enovelamento não é uma busca aleatória, mas sim um processo guiado por um custo energético que direciona a proteína à conformação nativa.

Uma outra explicação para o paradoxo de Levinthal é a teoria da paisagem energética (ELF, *Energy Landscape Theory*), desenvolvida no final da década de 1980 por Wolynes, Onuchic e colaboradores.¹² Essa teoria busca descrever os princípios que governam os mecanismos de enovelamento de proteínas, no contexto de modelos simplificados da mecânica estatística de polímeros, teoria de vidros de spin e simulações computacionais.^{13,14} Vidros de spin, como sistemas magnéticos e heteropolímeros aleatórios, são sistemas com uma paisagem energética rugosa, com várias estruturas diferentes apresentando energias muito baixas, mas próximas entre si. De forma semelhante aos vidros de spin, as proteínas também apresentam uma superfície de energia rugosa, mas com a diferença de que a sequência de aminoácidos é selecionada ao longo da evolução.

A ELT incorpora o princípio de frustração mínima¹² para explicar a eficiência do

enovelamento de proteínas. Ou seja, a sequência de aminoácidos evoluiu de modo que a maioria das interações são “minimamente frustradas” ou, em outras palavras, “maximamente consistentes”.¹⁴ Embora interações não nativas estejam presentes, durante o enovelamento, elas são minimizadas para impedir armadilhas cinéticas. Essa condição é necessária para obter uma paisagem energética afunilada,¹⁵ que fornece uma visão global dos estados envolvidos no enovelamento.

O funil de energia é uma representação utilizada para mostrar as possíveis rotas de enovelamento (Figura 1.2.0.1).¹⁵ O eixo y (altura) do funil representa a energia livre de uma única conformação, e o eixo x (largura do funil) a entropia conformacional. No topo do funil, encontram-se as estruturas desenoveladas, cuja entropia conformacional é máxima. À medida que interações consistentes vão sendo formadas, a energia livre do sistema, em média, diminui, até atingir o estado nativo. No estado nativo, a sequência de aminoácidos corresponde à conformação de menor energia livre e entropia conformacional. Desta forma, o enovelamento ocorre por uma organização de um conjunto de estruturas, em vez de intermediários específicos. Consequentemente, existem muitas rotas específicas, e não apenas um caminho, que levam ao estado nativo. Isso satisfaz a hipótese termodinâmica de Anfinsen de forma direta e rápida.

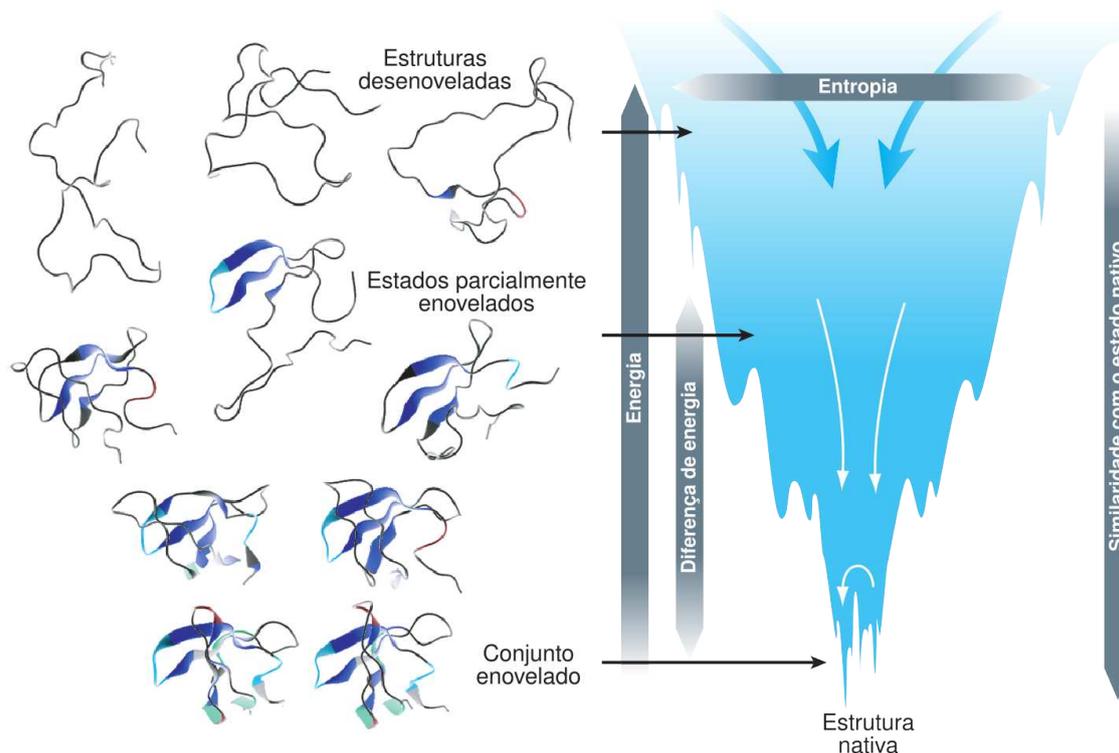


Figura 1.2.0.1: Representação do funil de energia do enovelamento da proteína SH3.¹⁶ O processo de enovelamento ocorre por meio da organização progressiva de conjuntos de estruturas (à esquerda) em um funil de energia livre (à direita). No topo do funil, encontram-se as estruturas desenoveladas, com alta entropia conformacional, a qual é compensada pela energia livre adquirida à medida que as interações nativas são formadas. Figura adaptada de Brooks III et al. (2001).¹⁶

1.3 Interações entre osmólitos e proteínas

O enovelamento de proteínas, bem como os diferentes conjuntos de estruturas, são fortemente dependentes do meio em que se encontram. Em resposta ao estresse osmótico, muitos organismos utilizam osmólitos, que atuam como cossolventes, a fim de atenuar os efeitos no equilíbrio entre os estados nativo e desenovelado da proteína, mantendo-a funcional. O estresse osmótico pode resultar de condições químicas - como a alta concentração extracelular de sal, desidratação celular, e valores extremos de pH - e físicas - como variações de temperatura e pressão. Esses fatores influenciam no tipo de osmólito que é mais comumente utilizado pelos organismos.^{6,17}

As três maiores classes de osmólitos são os 1) aminoácidos e seus derivados, 2) sais de metilamônio, e os 3) açúcares e poliálcoois.⁶ Alguns exemplos de osmólitos que compõem essas classes são 1) prolina, taurina e ácido aspártico; 2) N-óxido de trimetilamina (TMAO), colina e glicina

betaína; e 3) glicose, sorbitol e glicerol, respectivamente. Uma outra classificação, baseada no tipo de efeito causado na estrutura da proteína, trata-se de osmólitos compatíveis e não compatíveis. Os osmólitos compatíveis aumentam a estabilidade da proteína contra condições desnaturantes (por exemplo, o TMAO). Por outro lado, os osmólitos não compatíveis atuam desestabilizando a estrutura da proteína. Os exemplos mais conhecidos para a classe dos osmólitos não compatíveis são a ureia e o cloreto de guanidina (GdmCl). Na Figura 1.3.0.1 encontra-se a representação estrutural desses osmólitos.

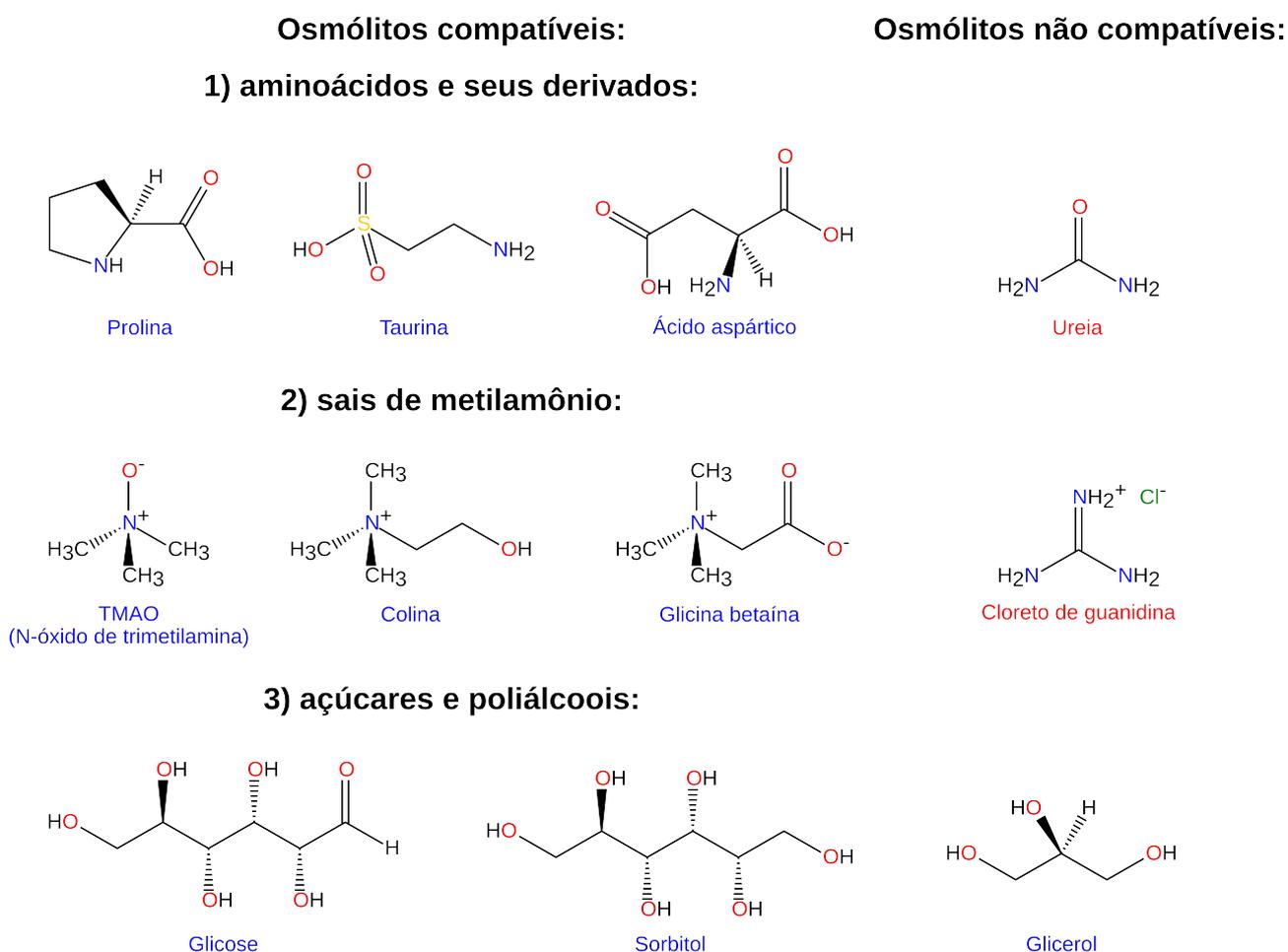


Figura 1.3.0.1: Representação estrutural de osmólitos compatíveis (nomeados em azul) e não compatíveis (nomeados em vermelho). As três maiores classes de osmólitos compatíveis são: 1) aminoácidos e seus derivados, como a prolina, taurina e o ácido aspártico; 2) sais de metilamônio, como o TMAO, colina e glicina betaína; e os 3) açúcares e poliálcoois, como a glicose, sorbitol e o glicerol. Os osmólitos não compatíveis mais comuns são a ureia e o GdmCl.

Os osmólitos foram primeiramente observados na natureza e, posteriormente,

desenvolvidos para serem aplicados na indústria, a fim de otimizar formulações biológicas. Estima-se que os osmólitos mais utilizados em formulações disponíveis atualmente são derivados de aminoácidos (41%), como glicina e histidina, e de carboidratos (36%),¹⁸ como sacarose e o manitol. A importância de osmólitos nos estágios do desenvolvimento de vacinas, como nas etapas de produção, purificação, estabilidade da proteína, também tem sido destacada.¹⁹ Estudos indicam que, na etapa de produção de vacinas, osmólitos podem estimular a taxa de crescimento dos microrganismos cultivados e, portanto, a produção de antígenos. Na etapa de purificação de vírus para vacinas, os osmólitos atuam como agentes floculantes, promovendo a agregação das partículas virais e facilitando sua remoção do meio. Esse efeito contribui para a redução de custos em comparação com métodos tradicionais, como ultracentrifugação e cromatografia, e é especialmente vantajoso em grandes escalas de produção.¹⁹

Osmólitos atuam como estabilizadores ou desestabilizadores em função da sua interação com proteínas. Os estabilizadores apresentam interações desfavoráveis com a proteína, e estabilizam estruturas mais compactas pelo mecanismo de exclusão preferencial, ou seja, hidratação preferencial.²⁰ Esse mecanismo é desfavorável, pois aumenta a energia livre da proteína. Para reduzir essa energia, a proteína permanece enovelada. Em nível microscópico, a concentração de osmólitos em regiões próximas da proteína é menor do que a que existiria nesta mesma solução na ausência de proteína no meio, ou seja, na ausência de interações proteína-osmólito. Por outro lado, os osmólitos desestabilizadores apresentam interações favoráveis com a proteína, e estabilizam estruturas de maior superfície exposta ao solvente pelo mecanismo de interação preferencial. Neste caso, ocorre um acúmulo de osmólitos ao redor da proteína, de tal forma que a concentração do osmólito nessa região é maior do que a concentração no *bulk* da solução, onde a interação proteína-osmólito é negligenciada (Figura 1.3.0.2).

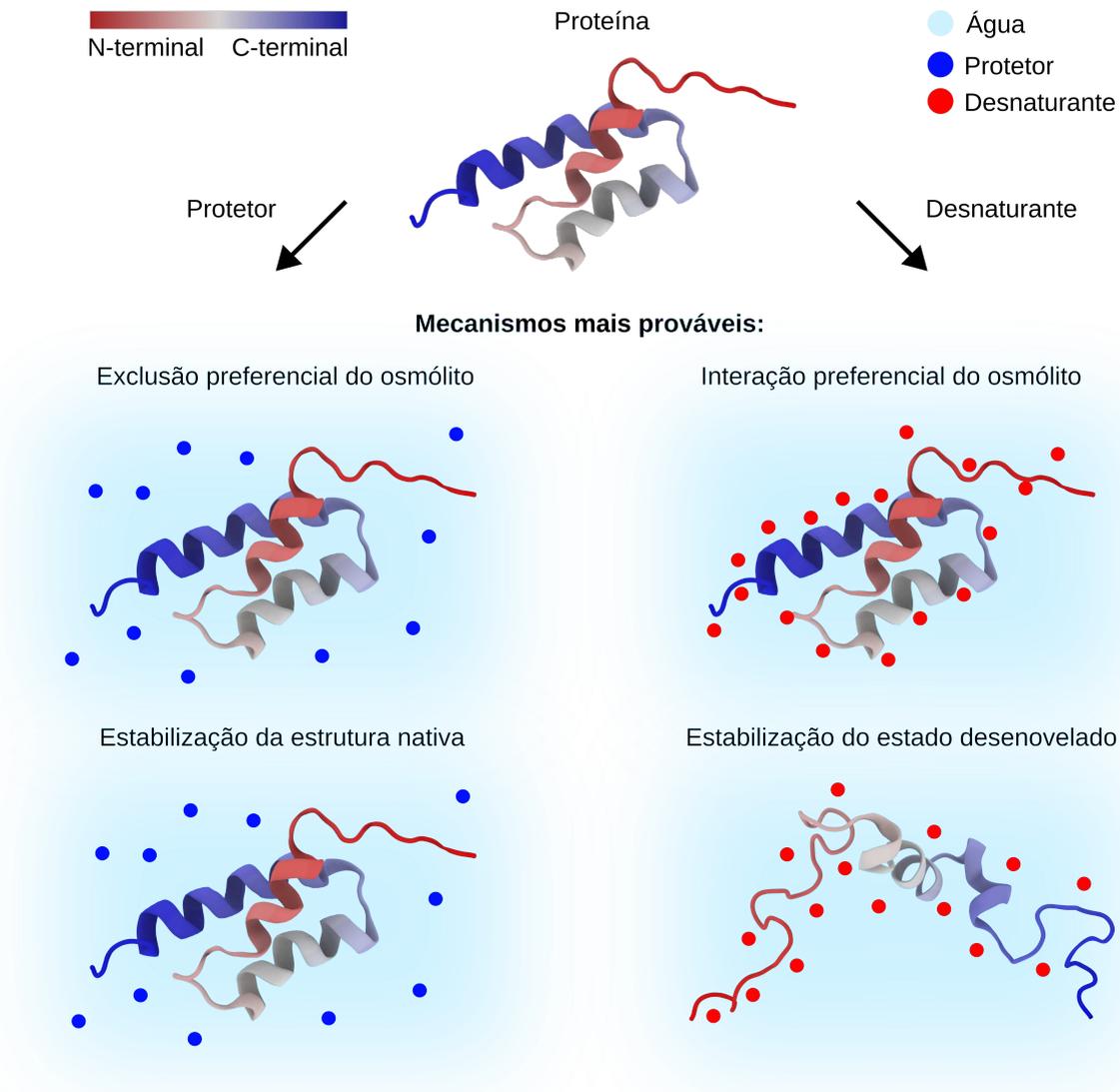


Figura 1.3.0.2: Ilustração esquemática dos mecanismos mais prováveis de estabilização de proteínas por osmólitos protetores (lado esquerdo) e de desestabilização por desnaturantes (lado direito). Figura adaptada de Rani & Venkatesu (2018).⁶

O papel de osmólitos/cossolventes na modulação da estabilidade e enovelamento de proteínas é evidente, mas as interações entre essas moléculas e a superfície heterogênea da proteína é fraca e com diferentes afinidades. Compreender essas interações é essencial para elucidar como os cossolventes influenciam os caminhos de enovelamento de proteínas, tendo em vista que essas moléculas podem estabilizar ou desestabilizar intermediários estruturais específicos. Esse entendimento pode ser alcançado diretamente a partir da teoria de soluções de Kirkwood–Buff (KB).²¹ A teoria de KB é uma teoria exata válida para qualquer solução isotrópica em equilíbrio termodinâmico, e será abordada na seção 2.1.2.^{21,22}

Capítulo 2

Metodologia

2.1 Caracterização termodinâmica das estruturas do solvatação

A forma em que as moléculas de solvente se distribuem na solução contendo um soluto, como uma proteína, pode ser obtida por meio de funções de distribuição. Esta seção revisa o formalismo e métodos computacionais para obter as funções de distribuição de mínima-distância (MDDFs, *Minimum distance distribution-functions*), as integrais de Kirkwood-Buff (IKBs) e os parâmetros de interação preferencial (Γ s), calculados com o pacote ComplexMixtures.jl.²³ Ao longo do texto, as notações p , w , c , e s serão usadas para referir ao peptídeo (ou proteína), água, cossolvente, e qualquer componente do solvente (água ou cossolvente), respectivamente.

2.1.1 Funções de Distribuição de Mínima-Distância

As MDDFs²⁴ fornecem uma visão clara das estruturas de solvatação e apresentam vantagens em relação às funções de distribuição radial (RDF),²⁵ especialmente para solutos não esféricos (como as proteínas, e, particularmente os estados desenovelados). Isso é possível pois, em vez de computar as distâncias entre os centros de massa do soluto e do solvente (comum no cálculo das RDFs), as MDDFs adotam a distância mínima entre átomos do solvente e os átomos do soluto. Especificamente, a “distância mínima” refere-se a menor distância entre qualquer átomo de cada molécula do solvente e qualquer átomo do soluto na solução. Com isso, as MDDFs consideram automaticamente a forma das moléculas de soluto e solvente, resultando em funções de distribuição facilmente interpretáveis do ponto de vista das interações físico-químicas específicas.²⁴

As MDDFs, aqui representadas por $g_{ps}(r)$, são descritas em termos da densidade

numérica média de átomos do solvente, $n_s(r)$, relativa à densidade em uma distribuição de gás ideal, $n_s^*(r)$,

$$g_{ps}(r) = \frac{n_s(r)}{n_s^*(r)} \quad (2.1.1.1)$$

onde r é a distância entre a proteína e o solvente.

A Figura 2.1.1.1 representa uma MDDF genérica de um solvente em relação à proteína, caracterizada por três regiões principais: a distância de exclusão, as camadas de solvatação, e a região conhecida como *bulk* da solução. A distância de exclusão consiste na região ao redor da proteína que é inacessível às moléculas do solvente. Em distâncias (r) maiores, é possível observar a formação da primeira camada de solvatação, em decorrência de interações favoráveis entre a proteína e o solvente. Posteriormente, a MDDF é caracterizada por regiões de depleção e acúmulo de moléculas de solvente ao redor da proteína, resultantes das sucessivas camadas de solvatação. Em distâncias longas, geralmente a partir de 10 Å no caso da água, a MDDF converge para 1, como consequência dos valores de $n_s(r)$ e $n_s^*(r)$ serem praticamente iguais. Nessa região (denominada de região *bulk*), o soluto não interfere mais na estrutura do solvente.

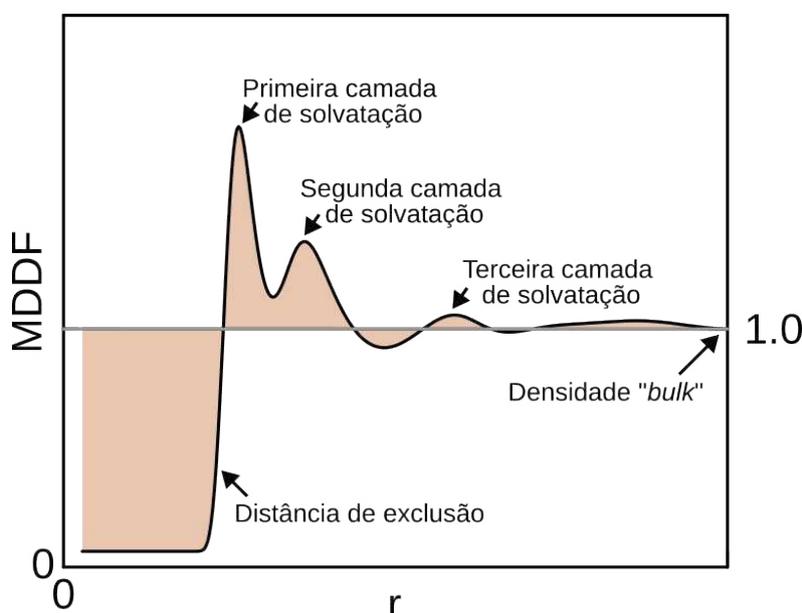


Figura 2.1.1.1: Representação da MDDF de um solvente em relação à uma proteína em função da distância r . A figura mostra a região da distância de exclusão, das camadas de solvatação e da região de *bulk*. Na região *bulk*, o soluto não interfere mais na estrutura do solvente, e a MDDF converge para 1, como indicado pela linha cinza.

Após serem devidamente normalizadas, as MDDFs podem ser utilizadas para calcular as IKBs^{21,26} e parâmetros de interação preferencial, com base na teoria de soluções de Kirkwood-Buff (seção 2.1.2).^{21,26,27} Dessa forma, torna-se possível conectar a visão microscópica da solvatação com parâmetros termodinâmicos.²⁰

2.1.2 Teoria de soluções de Kirkwood-Buff

Em 1951, Kirkwood e Buff publicaram uma teoria, baseada na mecânica estatística, que relaciona funções de distribuição molecular para um sistema multicomponente no *ensemble* grande canônico (μVT) com quantidades termodinâmicas macroscópicas.^{21,26,27} A teoria de Kirkwood-Buff (KB) permite obter grandezas termodinâmicas como a compressibilidade e volume molar parcial, que podem ser expressas por meio de integrais de volume das funções de correlação dos pares ($g_{ij}(r)$). Essas integrais, são conhecidas como integrais de Kirkwood-Buff (IKBs).

Tendo em vista que a teoria de KB é baseada no ensemble μVT , as integrais de KB são definidas para sistemas abertos. Por outro lado, simulações de DM são mais convenientemente realizadas nos ensembles canônico (NVT), isotérmico-isobárico (NPT) ou micro-canônico (NVE), que correspondem a sistemas fechados. N, V, P, E, e T correspondem ao número de partículas, volume, pressão, energia e temperatura do sistema. Por esse motivo, as funções de distribuição para simulações μVT e NPT diferem por um termo da ordem de $1/N$, em que N é o número de partículas do sistema.²⁸ Rigorosamente, a convergência da função de distribuição para um, em longas distâncias, seria violada. No entanto, para sistemas suficientemente grandes, os cálculos convergem para resultados corretos, independentemente do ensemble utilizado.²⁸ Por conta disso, é possível definir a IKB para um sistema fechado, de tamanho finito.

As IKBs, de um solvente em relação à uma proteína, por exemplo, podem ser calculadas a partir de $n_s(r)$ e $n_s^*(r)$ ²⁴ (definidos na seção anterior):

$$G_{ps} = \frac{1}{\rho_s} \int_0^{\infty} [n_s(r) - n_s^*(r)] S(r) dr \quad (2.1.2.1)$$

em que s é a concentração do solvente no *bulk* da solução e $S(r)$ é a superfície definida pela distância mínima r em relação a qualquer átomo do soluto. Devido à contagem de distâncias mínimas, $S(r)$

depende do formato do soluto.

Ao integrar a Equação 2.1.2.1 para um subvolume finito do sistema, obtém-se uma aproximação da IKB, expressa na Equação 2.1.2.2.

$$G_{ps}(R) \approx \frac{1}{\rho_s} [N_{ps}(R) - N_{ps}^*(R)] \quad (2.1.2.2)$$

em que $N_{ps}(R)$ e $N_{ps}^*(R)$ são o número de distâncias mínimas entre a proteína e o solvente menores que R , e o número de distâncias mínimas equivalentes dentro de R em um sistema com distribuição de gás ideal (ou seja, na ausência de interações proteína-solvente), respectivamente.^{24,29,30}

As IKBs representadas nas equações 2.1.2.1 e 2.1.2.2 podem ser interpretadas como uma medida de acúmulo ou depleção de moléculas de um determinado componente do solvente na região integrada (R), em relação ao número de moléculas que estariam presentes em uma distribuição de gás ideal – ou seja, na ausência de interações com a proteína. Essa região, denominada “domínio da proteína”, corresponde ao volume da solução em que a presença da proteína afeta a distribuição dos componentes do solvente, como água ou cossolvente. Assim, as IKBs quantificam o excesso ou a deficiência de determinado componente ao redor da proteína, refletindo a afinidade entre esse componente e o soluto.²²

Em um sistema ternário, contendo soluto, solvente e cossolvente, é possível definir o parâmetro de interação preferencial (Γ). O Γ é uma propriedade termodinâmica que permite quantificar a acumulação relativa das moléculas de solvente e cossolvente em torno do soluto.^{17,31,32} A partir de (Γ), é possível determinar qual componente da solução interage preferencialmente com o soluto. Parâmetros de interação preferencial para sistemas biomoleculares podem ser obtidos por experimentos de diálise no equilíbrio.³³ De uma forma resumida, duas soluções tamponadas – apenas uma delas contendo o cossolvente – são separadas por meio de uma membrana permeável ao cossolvente, mas impermeável à biomolécula. Ao adicionar a proteína em um dos compartimentos, o cossolvente move-se livremente entre as soluções, concentrando-se mais em uma delas, dependendo da sua afinidade com a proteína (Figura 2.1.2.1).³⁴

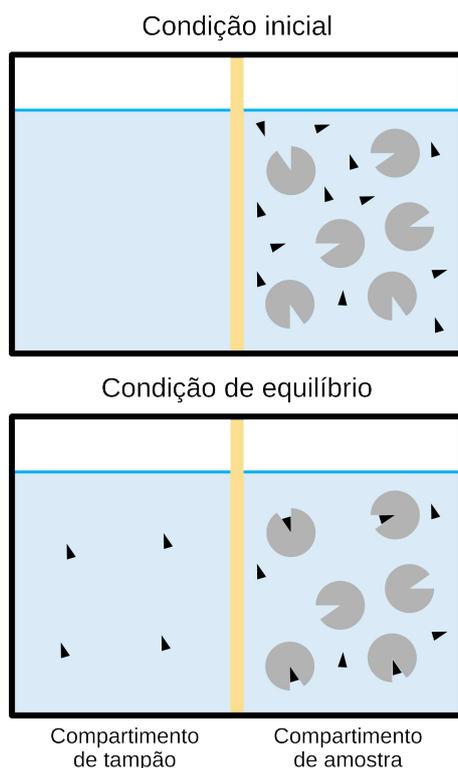


Figura 2.1.2.1: Ilustração do experimento de diálise no equilíbrio. A figura mostra duas soluções separadas por uma membrana permeável (retângulo laranja) ao cossolvente (triângulos pretos) mas impermeável à proteína (círculos cinzas). Figura adaptada de Pinger et al. (2017).³⁴

Após o equilíbrio, a concentração do cossolvente livre no compartimento de tampão é medida para determinar, de forma indireta, a quantidade de cossolvente ligado à proteína e, conseqüentemente, o Γ .³⁵ Matematicamente, Γ é definido de acordo com a Equação 2.1.2.3.

$$\Gamma = \left(\frac{\partial m_c}{\partial m_p} \right)_{T, \mu_s, \mu_c} \quad (2.1.2.3)$$

em que m_c e m_p são a molalidade do cossolvente e da proteína; T é a temperatura do sistema; s e c representam o potencial químico do solvente (usualmente a água) e do cossolvente, respectivamente.

É razoável supor que manter μ_s , μ_c e T constantes equivale a manter μ_c , T e P constantes.³⁶ Assim, pode-se escrever:

$$\Gamma = \left(\frac{\partial m_c}{\partial m_p} \right)_{T, P, \mu_c} \quad (2.1.2.4)$$

em que μ é o potencial químico e m é a concentração molal das espécies.

Γ representa a mudança na concentração do cossolvente necessária para manter o potencial químico μ_c constante quando a proteína é introduzida na solução.^{5,37} Se a adição da proteína na solução levar a um acúmulo do cossolvente no domínio do soluto, Γ_{pc} será positivo. Logo, será necessário uma adição de cossolvente (aumento de m_c) para manter o seu potencial químico. Por outro lado, um valor de Γ_{pc} negativo significa que a concentração de cossolvente no domínio da proteína é menor que a sua concentração no *bulk*. Nesse caso, a adição da proteína na solução requer uma diminuição da concentração de cossolvente para manter μ_c constante. Em casos específicos, o Γ_{pc} pode ser igual a zero. Nesse caso, o domínio da proteína é ocupado por moléculas de água ou de cossolvente com afinidades iguais, de tal forma que a composição do cossolvente no domínio da proteína e no *bulk* será a mesma.³⁸ Resumidamente, Γ_{pc} consiste no número de moléculas do cossolvente em excesso ($\Gamma_{pc} > 0$), ou déficit ($\Gamma_{pc} < 0$), no domínio da proteína, considerando o volume molecular do cossolvente no *bulk* da solução.

Dado que a regra cíclica para derivadas parciais de três variáveis independentes A, B, e C é dada por:

$$\left(\frac{\partial A}{\partial B}\right)_C \left(\frac{\partial B}{\partial C}\right)_A \left(\frac{\partial C}{\partial A}\right)_B = -1 \quad (2.1.2.5)$$

pode-se aplicá-la à Equação 2.1.2.4, e perceber que o Γ também pode ser expresso como a variação no potencial químico da proteína com o potencial químico do cossolvente na solução.

$$\left(\frac{\partial m_c}{\partial m_p}\right)_{T,P,\mu_c} \left(\frac{\partial m_p}{\partial \mu_c}\right)_{T,P,m_c} \left(\frac{\partial \mu_c}{\partial m_c}\right)_{T,P,m_p} = -1 \quad (2.1.2.6)$$

Logo,

$$\Gamma = \left(\frac{\partial m_c}{\partial m_p}\right)_{T,P,\mu_c} = -\frac{\left(\frac{\partial \mu_c}{\partial m_p}\right)_{T,P,m_c}}{\left(\frac{\partial \mu_c}{\partial m_c}\right)_{T,P,m_p}} = -\frac{\left(\frac{\partial \mu_p}{\partial m_c}\right)_{T,P,m_p}}{\left(\frac{\partial \mu_c}{\partial m_c}\right)_{T,P,m_p}} = \left(\frac{\partial \mu_p}{\partial \mu_c}\right)_{T,P,m_p} \quad (2.1.2.7)$$

Em uma solução contendo a proteína (soluto) idealmente diluída, o parâmetro de interação preferencial do cossolvente à proteína (Γ_{pc}) pode ser calculado a partir da diferença das IKBs dos componentes do solvente, de acordo com a Equação 2.1.2.8.

$$\Gamma_{pc} \approx \rho_c [G_{pc}(R) - G_{pw}(R)] \quad (2.1.2.8)$$

Similarmente, o parâmetro de hidratação preferencial é definido como:

$$\Gamma_{pw} \approx \rho_w [G_{pw}(R) - G_{pc}(R)] \quad (2.1.2.9)$$

Os parâmetros p_c e p_w apresentam sinais opostos. Portanto, se Γ_{pc} for maior do que zero, Γ_{pw} será negativo, isto é, a proteína é efetivamente desidratada.

Conforme discutido anteriormente, os cossolventes podem influenciar o processo de enovelamento de proteínas. Dado que os Γ_{pc} para os estados enovelado e desenovelado sejam conhecidos, é possível compreender a direção para a qual um cossolvente irá favorecer a reação em uma dada concentração de cossolvente utilizando a relação de ligação de Wyman.^{20,39,40} A relação de ligação de Wyman afirma que a variação da constante de equilíbrio para a reação de desnaturação (K) com a atividade do cossolvente (a_c) é igual à diferença entre os parâmetros de interação preferencial do cossolvente nos estados desenovelado (Γ_{Uc}) e nativo (Γ_{Nc}) (Equação 2.1.2.10).⁴⁰

$$\left(\frac{\partial \ln K}{\partial \ln(a_c)} \right)_{T,P,a_p} = \Delta\Gamma = \Gamma_{Uc} - \Gamma_{Nc} \quad (2.1.2.10)$$

em que T , P e a_p são a temperatura (em Kelvin), a pressão e a atividade da proteína. Se $\Delta\Gamma$ for positivo, o estado desenovelado é estabilizado em relação ao estado nativo pela adição do cossolvente. Caso contrário, $\Delta\Gamma < 0$, o estado nativo é estabilizado em relação ao estado desenovelado. Se $\Delta\Gamma$ for zero, o cossolvente não tem efeito na reação em uma dada composição de cossolvente.

2.1.3 Energia livre de transferência

Para muitas proteínas, o processo de desnaturação é reversível. Isso implica que os estados nativo e desenovelado existem em equilíbrio termodinâmico, e que a diferença de energia livre entre esses estados é de apenas algumas quilocalorias por mol, como discutido na seção 1.1. Evidentemente, no ambiente nativo, esse estado deve possuir uma energia livre inferior a todas as possíveis conformações desenoveladas. Da mesma forma, cada estado desenovelado específico, deve ser energeticamente o mais estável nas condições experimentais em que é predominante. Cerca de 60 anos atrás Tanford e Nozaki iniciaram o desenvolvimento de um modelo quantitativo para explicar o processo de desnaturação de proteínas por ureia e cloreto de guanidina (GdmCl), que ficou conhecido como Modelo de Transferência de Tanford.^{39,41-43} O Modelo de Transferência de Tanford é representado em um ciclo termodinâmico (Figura 2.1.3.1), que compara a diferença de energia livre ($\partial\Delta\mu$) nos estados nativo ($\Delta\mu_{p,tr}^N$) e desenovelado ($\Delta\mu_{p,tr}^U$) da proteína, ao serem transferidos da água (ou tampão) para uma solução aquosa de cossolvente, sem mudarem de conformação.

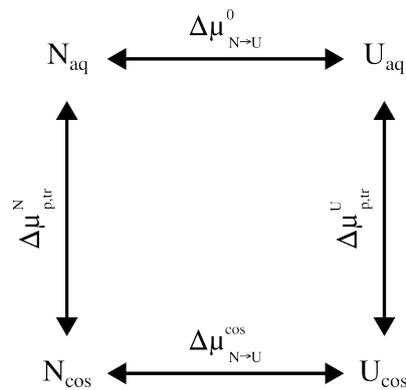


Figura 2.1.3.1: Representação do ciclo termodinâmico a partir do modelo de transferência de Tanford.^{39,41} Esse ciclo ilustra como as energias livres de enovelamento entre os estados nativo (N) e desenovelado (U), na presença de cossolvente ($N_{cos} \rightleftharpoons U_{cos}$, representado por $\Delta\mu_{N \rightarrow U}^{cos}$), diferem daquelas em um tampão aquoso ($N_{aq} \rightleftharpoons U_{aq}$, representado por $\Delta\mu_{N \rightarrow U}^0$). Medidas diretas das constantes de equilíbrio com e sem o cossolvente (reações na horizontal) permitem o cálculo do efeito do cossolvente na energia livre de enovelamento, $\partial\Delta\mu_{p,tr} = \Delta\mu_{N \rightarrow U}^{cos} - \Delta\mu_{N \rightarrow U}^0$. Alternativamente, as energias livres de transferência para os estados nativo ($\Delta\mu_{p,tr}^N$) e desenovelado ($\Delta\mu_{p,tr}^U$) podem ser obtidas (reações na vertical), e $\partial\Delta\mu_{p,tr} = \Delta\mu_{p,tr}^U - \Delta\mu_{p,tr}^N$.

Tendo em vista que esse modelo é um representado por um ciclo termodinâmico, segue que:

$$\partial\Delta\mu = \Delta\mu_{p,tr}^U - \Delta\mu_{p,tr}^N = \Delta\mu_{N\rightarrow U}^{\cos} - \Delta\mu_{N\rightarrow U}^0 \quad (2.1.3.1)$$

em que $\Delta\mu_{p,tr}^N$, $\Delta\mu_{p,tr}^U$, $\Delta\mu_{N\rightarrow U}^{\cos}$, $\Delta\mu_{N\rightarrow U}^0$ são as energias livres correspondentes às reações:

- a) $N_{aq} \rightleftharpoons N_{cos}$,
- b) $U_{aq} \rightleftharpoons U_{cos}$,
- c) $N_{cos} \rightleftharpoons U_{cos}$, e
- d) $N_{aq} \rightleftharpoons U_{aq}$, respectivamente.

A energia livre de transferência do estado nativo, $\Delta\mu_{p,tr}^N$, por exemplo, é definida como $\Delta\mu_{p,tr}^N = \mu_p^N$ (em cossolvente) - μ_p^N (em água)). No entanto, experimentalmente, determina-se $\Delta\mu_{p,tr}^X$, e não os valores individuais de μ_p . A energia livre de transferência ($\Delta\mu_{p,tr}^X$) para os conjuntos nativo e desenovelado é determinada pela Equação 2.1.3.2:

$$\Delta\mu_{p,tr}^X = \int_0^{m_c} \left(\frac{\partial\mu_p}{\partial m_c} \right)_{T,P,m_p} dm_c \quad (2.1.3.2)$$

em que X é o estado nativo ou o desenovelado e m_i é a molalidade do componente i . $\frac{\partial\mu_p}{\partial m_c}$ é a variação do potencial químico da proteína com a concentração de cossolvente. A integral é realizada a partir da água pura até a concentração desejada m_c de cossolvente.

A adição de um cossolvente à uma solução aquosa de proteína leva à variação dos potenciais químicos da proteína (μ_p) e do cossolvente (μ_c), de modo que o termo $\left(\frac{\partial\mu_p}{\partial m_c} \right)_{T,P,m_p}$ da Equação 2.1.3.2 pode ser definido como:²⁰

$$\left(\frac{\partial\mu_p}{\partial m_c} \right)_{T,P,m_p} = \left(\frac{\partial\mu_c}{\partial m_p} \right)_{T,P,m_c} \quad (2.1.3.3)$$

É possível aplicar a regra cíclica (Equação 2.1.2.5) ao termo $\left(\frac{\partial\mu_c}{\partial m_p} \right)_{T,P,m_c}$ para obter:

$$\left(\frac{\partial\mu_c}{\partial m_p} \right)_{T,P,m_c} \left(\frac{\partial m_p}{\partial m_c} \right)_{T,P,\mu_c} \left(\frac{\partial m_c}{\partial \mu_c} \right)_{T,P,m_p} = -1 \quad (2.1.3.4)$$

Ao isolar o termo $\left(\frac{\partial \mu_c}{\partial m_p}\right)_{T,P,m_c}$, obtém-se:

$$\left(\frac{\partial \mu_c}{\partial m_p}\right)_{T,P,m_c} = - \left(\frac{\partial m_c}{\partial m_p}\right)_{T,P,\mu_c} \left(\frac{\partial \mu_c}{\partial m_c}\right)_{T,P,m_p} \quad (2.1.3.5)$$

Note que o termo $\left(\frac{\partial m_c}{\partial m_p}\right)_{T,P,\mu_c}$ é o igual ao parâmetro de interação preferencial definido na Equação 2.1.2.4. Por via de regra, é conveniente comparar o parâmetro de interação preferencial, expresso em função da molalidade, com unidades de concentração expressa em gramas do componente i por gramas de água. Essas unidades podem ser facilmente convertidas, tendo como base a equação da molalidade:

$$m_i = \frac{n}{\text{kg do solvente}} = \frac{g_i}{MM_p \cdot \text{kg do solvente}} \quad (2.1.3.6)$$

Logo,

$$\Gamma_{T,P,\mu_c} = \left(\frac{\partial m_c}{\partial m_p}\right)_{T,P,\mu_c} = \left(\frac{\partial g_c}{MM_c \cdot \text{kg do solvente}}\right) / \left(\frac{\partial g_p}{MM_p \cdot \text{kg do solvente}}\right) \quad (2.1.3.7)$$

em que g_i é o valor em gramas do componente i e MM_p a massa molar da proteína.

Por uma simples manipulação da Equação 2.1.3.7, obtém-se:

$$\Gamma_{T,P,\mu_c} = \left(\frac{\partial g_c}{\partial g_p}\right)_{T,P,\mu_c} \left(\frac{MM_p}{MM_c}\right) \quad (2.1.3.8)$$

Vamos voltar ao termo $\left(\frac{\partial \mu_c}{\partial m_c}\right)_{T,P,m_p}$ da Equação 2.1.3.5. Este termo é uma correção para a não idealidade do cossolvente. Dado que o potencial químico do cossolvente é igual a:

$$\mu_c = \mu_c^0(T) + RT \ln(a_c) \quad (2.1.3.9)$$

em que $\mu_c^0(T)$ é o potencial químico padrão e a_c é a atividade do cossolvente ($a_c = m_c \gamma_c$, sendo m_c e γ_c a molalidade e o coeficiente de atividade do cossolvente). De forma equivalente à Equação

2.1.3.10, tem-se que:

$$\mu_c = \mu_c^0(T) + RT \ln(m_c) + RT \ln(\gamma_c) \quad (2.1.3.10)$$

A derivada da Equação 2.1.3.10 em relação a molalidade do cossolvente (m_c) é igual a:

$$\left(\frac{\partial \mu_c}{\partial m_c} \right)_{T,P,m_p} = -RT \left[\frac{1}{m_c} + \left(\frac{\partial \ln \gamma_c}{\partial m_c} \right)_{T,P,m_p} \right] \quad (2.1.3.11)$$

Finalmente, $\left(\frac{\partial \mu_p}{\partial m_c} \right)_{T,P,m_p}$ (Equação 2.1.3.3), em uma concentração específica de cossolvente, e a T , P e m_p constantes, é definido de acordo com a Equação 2.1.3.12. Dessa forma, os valores de $\Delta \mu_{p,tr}^X$ podem ser obtidos a partir dos valores de Γ_{pc} do estado X (nativo ou desenovelado), juntamente com valores de atividade do cossolvente em diferentes concentrações.

$$\left(\frac{\partial \mu_c}{\partial m_c} \right)_{T,P,m_p} = -\Gamma_{T,P,\mu_c} \left(\frac{RT \cdot MM_p}{MM_c} \right) \left(\frac{1}{m_c} + \frac{\partial \ln \gamma_c}{\partial m_c} \right)_{T,P,m_p} \quad (2.1.3.12)$$

em que MM_p e MM_c são as massas moleculares da proteína e do cossolvente, respectivamente; γ_c é o coeficiente de atividade do cossolvente.

Uma energia livre de transferência positiva indica que o estado considerado (nativo ou o desenovelado) é menos favorável em cossolvente do que em água. Por outro lado, uma energia livre de transferência negativa específica que o estado considerado é mais favorável em cossolvente do que em água.⁴⁴ Nos casos em que o cossolvente é inerte, as energias livres de transferência, $\left(\frac{\partial \mu_c}{\partial m_p} \right)_{T,P,\mu_c}$ e $\left(\frac{\partial \mu_p}{\partial m_c} \right)_{T,P,m_p}$ serão iguais a zero para todas concentrações de cossolvente.^{38,45}

2.2 Simulações de Dinâmica Molecular

A ciência do movimento de moléculas de um sistema, com uma resolução temporal da ordem de femtossegundos, pode ser estudada por meio de simulações de dinâmica molecular (DM). Um elemento essencial na descrição das moléculas do sistema é o conhecimento dos parâmetros e potenciais de interações intra e intermoleculares de cada componente do sistema, a partir do qual

as forças podem ser calculadas, bem como das equações que governam o movimento das partículas. Equações de movimento de Newton (2.2.0.1) são adequadas para prever a trajetória das partículas de um sistema biomolecular ao longo do tempo.

$$\vec{F}_i = m_i \vec{a}_i = \frac{\partial V_{TOT}(r_1, r_2, \dots, r_n)}{\partial r_n} \quad (2.2.0.1)$$

em que os parâmetros \vec{F}_i , m_i , \vec{a}_i , r_i são a força, a massa atômica, a posição, e a aceleração da partícula i no tempo t . V_{TOT} é o potencial total do sistema.

Essas equações são resolvidas simultaneamente por integração numérica, resultando em um conjunto de equações de movimento para cada partícula do sistema. Ao final das simulações, obtêm-se informações estrutural e dinâmica, com resolução atômica, e na escala de tempo considerada. A escolha do potencial de interação é uma etapa essencial para a descrição precisa do sistema de interesse. O potencial total (V_{TOT}) determina as forças que atuam sobre cada partícula, e a forma com que o sistema irá evoluir no tempo para gerar as trajetórias. V_{TOT} corresponde à soma de todas as interações intra e intermoleculares presentes no sistema. O conjunto de parâmetros e funções matemáticas utilizados para definir essas interações é denominado campo de força.

As interações intramoleculares são geralmente descritas por potenciais harmônicos, que modelam as vibrações das ligações (V_L) e ângulos (V_A), e por potenciais periódicos, que representam os movimentos dos diedros (V_D). Já as interações intermoleculares incluem o potencial de Lennard-Jones (V_{LJ}), responsável por descrever as forças repulsivas e atrativas, acrescido do potencial de Coulomb (V_C), que captura as interações eletrostáticas (Equação 2.2.0.2). Campos de força amplamente utilizados, como CHARMM, OPLS e AMBER, empregam funções de energia potencial semelhantes às descritas pela Equação 2.2.0.2.

$$V_{TOT} = V_S + V_A + V_D + V_{LJ} + V_C \quad (2.2.0.2)$$

em que,

$$V_S = \sum_{\text{ligação}} k_b (b - b_0)^2 \quad (2.2.0.3)$$

$$V_A = \sum_{\text{ângulo}} k_\theta (\theta - \theta_0)^2 \quad (2.2.0.4)$$

$$V_D = \sum_{\text{diedro}} k_\phi [1 + \cos(n\phi - \delta)] \quad (2.2.0.5)$$

$$V_{LJ} = \sum_{LJ} 4\epsilon_{ij} \left[\frac{\sigma^{12}}{r_{ij}^{12}} - \frac{\sigma^6}{r_{ij}^6} \right] \quad (2.2.0.6)$$

$$V_C = \sum_{\text{Coulomb}} = \left[\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \quad (2.2.0.7)$$

Nas Equações 2.2.0.3 e 2.2.0.4 os parâmetros b_0 e θ_0 correspondem, respectivamente, ao comprimento e ângulo de equilíbrio da ligação, enquanto k_b e k_θ são as constantes de força do oscilador harmônico. Os quatro parâmetros k_ϕ , n , ϕ , e δ no potencial diédrico (Equação 2.2.0.5) são a 1) constante de força que define a altura da barreira de rotação, 2) o número de mínimos para uma rotação completa, 3) o ângulo diedro e 4) o ângulo de fase ajustado para alinhar os mínimos de energia do potencial, respectivamente. As posições, assim como a velocidade de cada partícula, devem ser fornecidos inicialmente em um DM. A estrutura de proteínas, por exemplo, podem ser obtidas no Protein Data Bank (PDB, <http://www.rcsb.org>).

Os termos de Lennard-Jones entre átomos distintos são calculados usando regras de combinação. Alguns exemplos são as regras de Lorentz-Berthelot (Equação 2.2.0.8), empregadas nos campos de força AMBER e CHARMM, ou a regra geométrica (Equação 2.2.0.9), utilizada no OPLS.⁴⁶

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad \text{e} \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (2.2.0.8)$$

$$\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \quad \text{e} \quad \epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \quad (2.2.0.9)$$

Note que na DM clássica, as moléculas são tratadas como uma coleção de partículas unidas por forças harmônicas ou elásticas. Sendo assim, não é possível observar formação e ruptura de ligações, interações entre os orbitais, transferência de carga ou qualquer outro efeito eletrônico. Entretanto, a ruptura de uma ligação, especialmente para biomoléculas, não é relevante para muitos fenômenos estruturais na temperatura e escala de tempo consideradas em uma simulação de DM. Os efeitos quânticos podem ser incorporados dentro da descrição clássica. Parâmetros de potenciais clássicos, como definidos anteriormente, podem ser obtidos por cálculos quânticos e refinados por métodos semi-empíricos, a fim de representar com precisão as interações moleculares em uma determinada condição termodinâmica.

2.2.1 Integração das equações de movimento

Uma etapa importante em uma simulação de DM é a resolução da Equação 2.2.0.1. As equações de Newton não podem ser resolvidas analiticamente devido à interação de inúmeras partículas. Na prática, a resolução dessas equações deve ser realizada por integração numérica com a utilização de algoritmos específicos, nos quais a integração é feita em pequenos intervalos de tempo (dt), geralmente em $dt = 2 \text{ fs}$. O cálculo é realizado assumindo que a força permaneça constante no intervalo de tempo dt .

Por exemplo, a partir das posições e velocidades de cada partícula do sistema, determinadas inicialmente, a velocidade e a posição do sistema em um intervalo de tempo $t + \Delta t$ poderiam ser obtidas a partir das Equações 2.2.1.1 e 2.2.1.2. A velocidade inicial de cada partícula do sistema é usualmente atribuída de acordo com a distribuição gaussiana de Maxwell-Boltzmann na temperatura desejada.

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a}(t)\Delta t \quad (2.2.1.1)$$

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2} \quad (2.2.1.2)$$

Dessa forma, conhecendo as posições e velocidades no instante de tempo $t + \Delta t$, torna-se possível determinar as novas forças e acelerações, permitindo o cálculo das posições subsequentes. A repetição deste ciclo possibilita a obtenção da trajetória, que pode ser usada para obter parâmetros cinéticos e termodinâmicos do sistema. No entanto, esse método apresenta limitações, uma vez que as Equações 2.2.1.1 e 2.2.1.2 são equações truncadas da expansão de Taylor de 2ª ordem, com os termos de ordem superiores sendo ignorados. Isso leva a desvios significativos nas posições e velocidades ao longo do tempo, comprometendo a simulação.

Uma alternativa para resolver as equações de movimento foi introduzida por Verlet em 1967.⁴⁷ No método de Verlet, as posições das partículas em um instante posterior ($t + \Delta t$) e anterior ($t - \Delta t$) a t , são obtidas pela expressão em série de Taylor de 4ª ordem.

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \frac{d\vec{x}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{x}}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3\vec{x}}{dt^3}\Delta t^3 + O(\Delta t^4) \quad (2.2.1.3)$$

$$\vec{x}(t - \Delta t) = \vec{x}(t) - \frac{d\vec{x}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\vec{x}}{dt^2}\Delta t^2 - \frac{1}{6}\frac{d^3\vec{x}}{dt^3}\Delta t^3 + O(\Delta t^4) \quad (2.2.1.4)$$

A soma das Equações 2.2.1.3 e 2.2.1.4 permite determinar novas posições atômicas, conforme a Equação 2.2.1.5.

$$\vec{x}(t + \Delta t) = 2\vec{x}(t) + \frac{d^2\vec{x}}{dt^2}\Delta t^2 - \vec{x}(t - \Delta t) + O(\Delta t^4) \quad (2.2.1.5)$$

Perceba que no algoritmo de Verlet a velocidade não aparece de forma explícita nas equações que determinam a trajetória do sistema, mas o seu valor deve ser conhecido no cálculo da energia cinética do sistema. A velocidade média pode ser determinada pela diferença entre as posições em $t + \Delta t$ e $t - \Delta t$ (Equação 2.2.1.6):

$$\vec{v}(t + \Delta t) = \frac{\vec{x}(t + \Delta t) - \vec{x}(t - \Delta t)}{2\Delta t} \quad (2.2.1.6)$$

Uma consideração importante é que a expressão da velocidade possui um erro de segunda ordem, enquanto que a da posição possui um erro de quarta ordem. O algoritmo de Leap-Frog é uma

alternativa mais precisa que o algoritmo de Verlet, pois a velocidade é calculada de forma explícita na metade do intervalo de tempo ($\Delta t/2$), conforme representada na Figura 2.2.1.1 e definida na Equação 2.2.1.7.

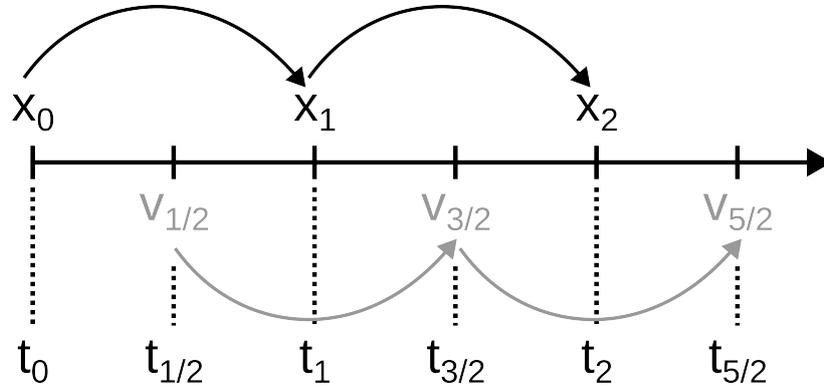


Figura 2.2.1.1: Representação esquemática do algoritmo de Leap-Frog para o cálculo das posições e velocidades. No algoritmo de Leap-Frog, a velocidade é explicitamente calculada no intervalo de tempo $\Delta t/2$.

$$\vec{v}\left(t + \frac{1}{2}\Delta t\right) = \vec{v}\left(t - \frac{1}{2}\Delta t\right) + \vec{a}(t)\Delta t \quad (2.2.1.7)$$

A partir da velocidade, calcula-se a posição da partícula conforme a Equação 2.2.1.8.

$$\vec{x}(t + \Delta t) = \vec{x} + \vec{v}\left(t + \frac{1}{2}\Delta t\right) \Delta t \quad (2.2.1.8)$$

Neste trabalho, simulações de DM foram amplamente utilizadas para caracterizar as estruturas de solvatação nos conjuntos de enovelamento de proteínas (veja os capítulos 4 e 5 para mais detalhes). Por meio das simulações atômicas foi possível capturar o comportamento dinâmico do solvente em um meio onde, essencialmente, o soluto manteve-se próximo da conformação inicial. No entanto, o entendimento de eventos biofísicos lentos, dentre eles o enovelamento de proteínas, geralmente está além das simulações clássicas de DM. Tais processos envolvem grandes mudanças conformacionais e são dificilmente bem amostrados, pois mínimos locais podem ser separados por altas barreiras energéticas. Nesse contexto, modelos *coarse-grained* ou algoritmos de amostragem ampliada devem ser considerados.

Na seção 2.2.2, serão discutidos os modelos baseados na estrutura (SBMs,

Structured-Based Models), desenvolvidos com base no princípio de frustração mínima e no conceito de uma paisagem energética afunilada.¹⁴ Os modelos SBMs incluem explicitamente dados estruturais da estrutura nativa no Hamiltoniano. Embora sejam modelos simplificados, têm-se mostrado eficazes para obter o conjunto de enovelamento de proteínas compatíveis com resultados experimentais.^{48,49} Nos capítulos 4 e 5 uma metodologia computacional foi desenvolvida, empregando simulações com os modelos SBMs e simulações atomísticas, para estudar o papel de diferentes solventes em toda superfície de enovelamento de proteínas. No Capítulo 6, os SBMs padrão^{48,50} foram também utilizados, e um novo potencial (*dual-basin*)⁵¹ foi desenvolvido para estudar o enovelamento de uma proteína que assume dois estados nativos distintos.

Na seção 2.2.3, será discutido brevemente sobre o método de amostragem ampliada por trocas de réplicas hamiltonianas, utilizado para estudar o enovelamento de um peptídeo à base de alanina na presença de trifluoroetanol, cujos resultados serão discutidos no Capítulo 3.

2.2.2 Simulações com os Modelos Baseados na Estrutura

O fundamento teórico dos modelos baseados na estrutura baseia-se em dois conceitos principais: a) no princípio de frustração mínima e no b) conceito de uma paisagem energética em formato de funil, conforme discutido na seção 1.2. Um potencial baseado na estrutura reduz drasticamente a complexidade do Hamiltoniano do sistema, tornando as simulações mais eficientes do ponto de vista computacional. Além disso, os modelos SBMs têm-se mostrado capazes de reproduzir o conjuntos de estados de transição de várias proteínas.^{48,49} Interações espacialmente próximas ao estado nativo são estabilizantes, enquanto que interações não nativas são repulsivas. Portanto, qualquer barreira de energia durante o enovelamento surge da competição energética e entrópica dos múltiplos caminhos de enovelamento que a proteína pode seguir.¹⁴

O potencial SBM é composto por três contribuições: 1) um potencial que inclui as interações que mantêm os resíduos da estrutura ligados covalentemente (V_{ligado}), 2) um potencial repulsivo ($V_{\text{repulsivo}}$), que impõe o volume de exclusão e previne a sobreposição das cadeias, e 3) um potencial atrativo (V_{atrativo}), cujas interações são representadas em um mapa de contatos.⁵² Como o próprio nome sugere, todos os três potenciais são obtidos com base em uma estrutura de referência.

$$V = \underbrace{(V_{\text{ligado}} + V_{\text{repulsivo}})}_{\text{geometria principal}} + \underbrace{V_{\text{atrativo}}}_{\text{estrutura terciária}} \quad (2.2.2.1)$$

O potencial usado nesta tese é o modelo baseado na estrutura de carbonos alfa ($C\alpha$ -SBM).⁴⁸ O $C\alpha$ -SBM é uma representação *coarse-grained* em que apenas o carbono $C\alpha$ de cada resíduo da proteína é utilizado para representar individualmente os resíduos da proteína na simulação. O potencial $C\alpha$ -SBM é dado pela Equação 2.2.2.2.

$$\begin{aligned} V(X, X_0) = & \sum_{\text{ligação}} \epsilon_r (r - r_0)^2 + \sum_{\text{ângulo}} \epsilon_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{diedro}} \epsilon_\phi \left[(1 - \cos(\phi - \phi_0)) + \frac{1}{2} (1 - \cos(3(\phi - \phi_0))) \right] \\ & + \sum_{\text{contato}} \epsilon_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] \\ & + \sum_{\text{não-contato}} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \end{aligned} \quad (2.2.2.2)$$

onde r e r_0 representam as distâncias entre dois resíduos subsequentes da configuração X e da configuração do estado nativo (X_0); de forma análoga, θ (θ_0) e ϕ (ϕ_0) representam os ângulos e diedros de cada configuração. As interações não ligadas entre pares de resíduos i e j são dadas pelo potencial de Lennard-Jones 10-12, conforme mostrado no quarto termo da Equação 2.2.2.2. O parâmetro d_{ij} é a distância entre dois $C\alpha$ na estrutura de referência (usualmente obtida a partir do PDB), onde $j > i + 3$. É importante ressaltar que, embora apenas os $C\alpha$ sejam explicitamente representados na simulação, o mapa de contato é calculado considerando a estrutura contendo todos os átomos pesados. O último potencial na Equação 2.2.2.2 é usado para evitar sobreposição de átomos, onde os resíduos possuem um raio aparente de 4 angstroms ($\sigma_{NC} = 4 \text{ \AA}$). Os parâmetros ϵ_r , ϵ_θ , ϵ_ϕ , ϵ_C e ϵ_{NC} são as constantes associadas a cada tipo de interação, sendo $\epsilon_r = 20000 \epsilon_0$, $\epsilon_\theta = 40 \epsilon_0$, $\epsilon_\phi = \epsilon_0$, $\epsilon_C = \epsilon_0$, $\epsilon_{NC} = \epsilon_0$, e ϵ_0 é a energia de interação entre os contatos ($\epsilon_0 = k_B T = 1$ unidade reduzida).

Os potenciais SBM podem ser expandidos para simular o enovelamento de proteínas cuja sequência pode se enovelar em mais de um estado nativo.⁵³⁻⁵⁶ Nesse caso, as informações

estruturais e topológicas dos estados nativos devem ser adicionadas em um único potencial, que é genericamente conhecido como um potencial com múltiplas bacias (*multiple-basin*).⁵¹ No Capítulo 6 iremos abordar o enovelamento da proteína MAD2, uma proteína importante no processo de divisão celular que se enovela em dois estados nativos distintos, os estados inativo e ativo, em resposta de estímulos celulares. Neste estudo, foi possível capturar a interconversão entre os estados inativo e ativo da proteína.

2.2.3 Simulações de Dinâmica Molecular por troca de réplicas

Vários algoritmos têm sido desenvolvidos para melhorar a amostragem em dinâmica molecular.⁵⁷⁻⁶⁰ O método de troca de réplicas por temperatura (TREM, *Temperature Replica Exchange Method*), em particular, é reconhecido por sua praticidade conceitual e por ser altamente paralelizável.⁶¹ No método TREM, várias réplicas do sistema são executadas em paralelo ao longo de uma faixa de temperaturas, variando da temperatura desejada até temperaturas mais altas. As réplicas em temperaturas elevadas podem superar altas barreiras energéticas e amostrar estados conformacionais dificilmente amostrados em uma simulação com temperatura baixa. Durante a execução simultânea das réplicas do sistema, em intervalos de tempo predefinidos, pode haver uma troca entre as configurações das réplicas vizinhas. Essa troca é aceita ou rejeitada com base no critério de aceitação de Metropolis.⁶² Assim, sempre haverá uma troca quando a energia do sistema da réplica inferior for menor que a energia do sistema da réplica superior ($\Delta E < 0$). Quando a energia do sistema da réplica superior é maior ($\Delta E > 0$), pode haver troca com probabilidade r .

$$r \leq \exp\left(-\frac{\Delta E}{k_b T}\right) \quad (2.2.3.1)$$

em que r é um número aleatório entre 0 e 1.

Uma etapa importante no método TREM é definir uma distribuição de temperatura adequada para se obter taxas de aceitação homogêneas entre os pares de réplicas adjacentes. Tal distribuição pode ser obtida quando as temperaturas são distribuídas exponencialmente,⁶³ como na Equação 2.2.3.1:

$$\lambda_i = \frac{T_0}{T_i} = \exp \left[\frac{-i}{n-1} \ln \left(\frac{T_m}{T_0} \right) \right] \quad (2.2.3.2)$$

em que T_0 , T_i , n e T_m são a temperatura de interesse, a temperatura da réplica i , o número total de réplicas e a temperatura máxima. T_m é escolhido de tal forma que a réplica correspondente possa rapidamente ultrapassar as barreiras de energia potencial.

Uma limitação clara do método ocorre com o aumento do tamanho do sistema. No método TREM todo o sistema é aquecido, e um grande número de réplicas são necessárias para criar uma distribuição de probabilidade da energia potencial que tenha uma sobreposição eficiente. Isso é necessário para obter uma taxa de aceitação adequada, geralmente em torno de 0,3.^{64,65} O número de réplicas necessárias para obter uma amostragem representativa aumenta linearmente com a raiz quadrada do número total de graus de liberdade no sistema.⁶⁶ A fim de contornar esses desafios, métodos por trocas de réplicas Hamiltonianas foram desenvolvidos.⁶⁷⁻⁶⁹

Em particular, o método de troca de réplicas Hamiltonianas com t mpera de soluto (REST2, *Replica Exchange with Solute Tempering*)⁷⁰ permite que as réplicas do sistema sejam simuladas na mesma temperatura, mas com potenciais variados para que se obtenha uma melhor amostragem (Figura 2.2.3.1).

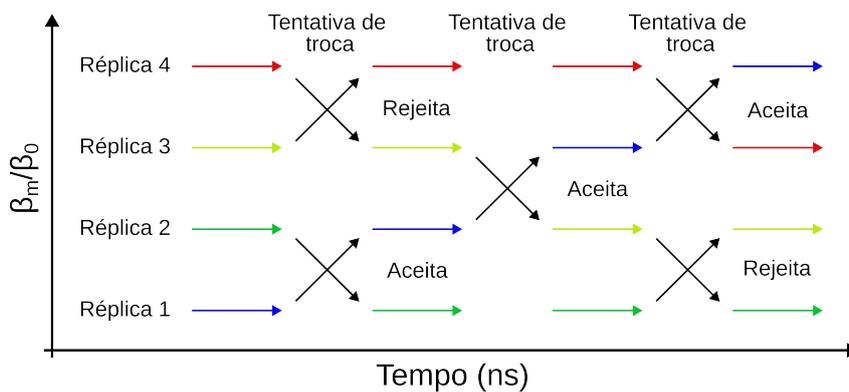


Figura 2.2.3.1: Esquema da simula o de troca de r plicas Hamiltonianas com t mpera de soluto (REST2).⁷⁰

Como consequ ncia, o sistema   menos perturbado, e o n mero de r plicas necess rias para que ocorra uma sobreposi o efetiva de energia entre as r plicas vizinhas   menor, comparado com o m todo TREM. A energia potencial para cada r plica m   escalonada de acordo com a

Equação 2.2.3.3.

$$E_{\text{REST}}(X) = \frac{\beta_m}{\beta_0} E_{pp}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{ps}(X) + E_{ss}(X) \quad (2.2.3.3)$$

em que E_{pp} é a energia intramolecular da proteína, E_{ps} é a energia de interação entre a proteína e o solvente, e E_{ss} é a energia de interação solvente-solvente. X representa a configuração de todo o sistema; $\beta_m = 1/k_B T_m$ e T_0 é a temperatura de interesse.

Uma outra vantagem do método REST2 é a possibilidade escolher de forma seletiva apenas uma parte do sistema para ser escalonada. Além disso, tendo em vista que todas as réplicas estão na mesma temperatura, as velocidades não precisam ser reescaladas durante a troca de configurações.

Capítulo 3

Efeitos do

2,2,2-Trifluoroetanol em

α -hélices

O 2,2,2-Trifluoroetanol (TFE) é conhecido por induzir a formação de hélices em peptídeos e proteínas. No entanto, após muitas investigações, nenhum consenso foi alcançado sobre o mecanismo de estabilização do TFE. O TFE altera a estrutura da água⁷¹⁻⁷³ e afeta suas propriedades dielétricas,⁷⁴ mas também compete com ela por ligações de hidrogênio com as cadeias principal e laterais das proteínas. Assim, mecanismos diretos e indiretos de atividade do TFE têm sido propostos. O modo direto de interação é particularmente atraente: o TFE estabelece ligações de hidrogênio com as carbonilas da cadeia principal da proteína, eliminando a água e, aparentemente, protegendo as ligações de hidrogênio intra-hélices (Figura 3.0.0.1). No entanto, medições de partição isotópica de hidrogênio em equilíbrio indicaram que as ligações de hidrogênio intra-helicoidais não são fortalecidas pela presença do TFE.⁷⁵ Assim, o mecanismo de estabilização da hélice pelo TFE também foi racionalizado em termos da interação preferencial com as estruturas helicoidais ou pela perturbação indireta da estrutura da água.^{72,76-78} Ambos os mecanismos, direto e indireto, podem cooperar para estabilizar as hélices. Tendo em vista que essas interações ocorrem simultaneamente com outras mudanças na estrutura da solução, é difícil separar a contribuição dos processos diretos *versus* indiretos para o efeito estabilizador do TFE.

Neste trabalho, realizamos simulações extensivas de amostragem ampliada do peptídeo Ace-AAQAAAQAAAQAA-NH₂ (aqui denominado como AAQAA₃) em água pura e misturas TFE em várias concentrações. Este peptídeo exibe uma estrutura helicoidal modelo em água e foi utilizado em investigações experimentais e computacionais sobre o papel dos cossolventes na estabilidade da hélice.⁷⁹⁻⁸⁵ As MDFFs e a teoria de soluções de KB são usadas para entender a base molecular e termodinâmica do mecanismo de estabilização de α -hélice pelo TFE. As simulações confirmam o

papel estabilizador do TFE a partir do cálculo do conteúdo de α -hélice do peptídeo, e que as estruturas são preferencialmente solvatadas pelo TFE.^{86,87} O TFE interage efetivamente com a cadeia principal da proteína, excluindo a água, em concordância com o mecanismo direto de estabilização.

As simulações permitem a realização de experimentos alquímicos, e assim modificamos as interações intermoleculares entre a cadeia principal e o TFE para evitar a ligação de hidrogênio supostamente estabilizadora. Surpreendentemente, o conteúdo de α -hélice do peptídeo aumentou, mostrando que esses contatos diretos têm uma contribuição efetiva desnaturante. Ao mesmo tempo, os parâmetros de interação preferencial permanecem basicamente constantes na ausência das ligações de hidrogênio entre o TFE e a cadeia principal. Portanto, o mecanismo de estabilização de hélices pelo TFE, baseado na proteção das ligações de hidrogênio intrapeptídicas por meio da ligação de hidrogênio do TFE com a carbonila da cadeia principal, não é suportado por evidências. Finalmente, mostramos que as interações não específicas do TFE com as conformações helicoidais são mais fortes do que com os estados sem qualquer conteúdo de α -hélice, excluindo as moléculas de água.

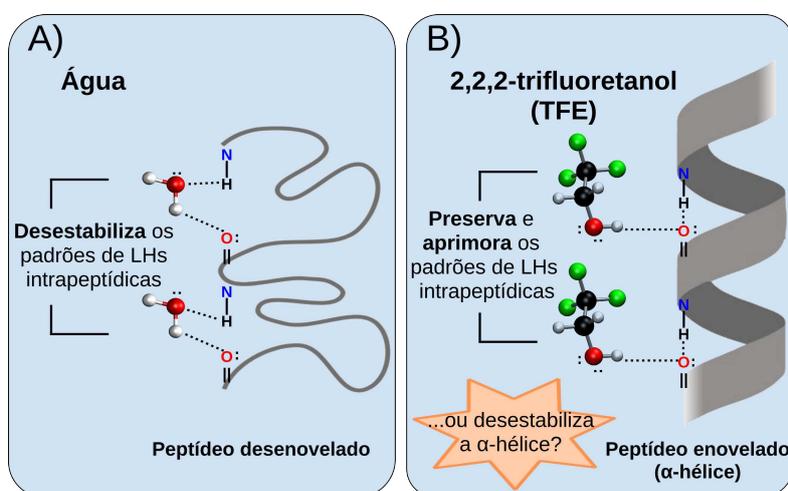


Figura 3.0.0.1: Modos de ligação de hidrogênio da A) água e do B) TFE com a cadeia principal do peptídeo. As moléculas de água atuam tanto como doadoras quanto comoceptoras de ligações de hidrogênio, causando o rompimento das ligações de hidrogênio intramoleculares do peptídeo. Por outro lado, o TFE é um forte doador de ligações de hidrogênio, mas um fraco receptor.⁷⁴ Como resultado, foi proposto que o hidrogênio ácido do TFE pode interagir com o oxigênio da cadeia principal, ao mesmo tempo que protege as ligações de hidrogênio intramoleculares do peptídeo (N-H...O) que estabilizam a hélice.^{71,88} Figura adaptada do artigo de revisão de Vincenzi et al. (2019).⁸⁸

3.1 Metodologia

3.1.1 Simulações de Dinâmica Molecular

Neste trabalho, utilizamos simulações de dinâmica molecular (DM) por troca de réplicas com temperatura de soluto (REST2, seção 2.2.3)⁷⁰ para estudar o efeito do TFE na estabilidade conformacional do AAQAA₃. Todas as simulações foram realizadas no software GROMACS (v2019.4),⁸⁹ compilado com o PLUMED (v2.5.5),⁹⁰ à temperatura de 300 K. No total, 10 réplicas foram adotadas para cada sistema, e E_{pp} e E_{ps} foram escalados por um fator de β_m/β_0 e $\sqrt{\beta_m/\beta_0}$, respectivamente, com β_m/β_0 variando de 1 para 0.5245.

Cada sistema foi inicialmente minimizado em 1000 passos utilizando o método *Steepest Descent*, seguido de duas equilibrações de 1 ns nos *ensembles* NVT e NPT. Após as etapas de equilíbrio, cada réplica foi simulada no *ensemble* NPT por 500 ns (totalizando 5 μ s para cada sistema, considerando as réplicas), com tentativas de troca a cada 400 passos de DM. O intervalo de perturbações foi ajustado de modo a obter uma taxa de troca acima de 30%. As estruturas foram salvas a cada 500 ps. As simulações foram realizadas a uma pressão constante de 1 bar, utilizando o algoritmo de Parrinello-Rahman⁹¹ com um tempo de relaxação de 2 ps e uma compressibilidade isotérmica de $4,5 \times 10^{-5} \text{ bar}^{-1}$. Um termostato de reescalonamento de velocidades foi utilizado para controlar a temperatura com um período de 0,1 ps.⁹²

Condições periódicas de contorno foram aplicadas, e um corte de 1,2 nm foi usado para interações de curto alcance. As interações eletrostáticas de longo alcance foram calculadas pelo método da soma de Ewald, Particle-Mesh Ewald (PME),⁹³ com interpolação de quarta ordem e um *grid spacing* de 0,16 nm. Todas as ligações envolvendo átomos de hidrogênio foram restritas com o algoritmo LINCS.⁹⁴ Para integrar as equações de movimento, o algoritmo leap-frog foi utilizado com um intervalo de tempo de 2 fs. Para as análises, apenas a réplica 0 foi utilizada.

Uma configuração inicial do AAQAA₃ foi construída com o pacote *Visual Molecular Dynamics* (VMD).⁹⁵ Em seguida, o AAQAA₃ foi solvatado em caixas cúbicas de 56 Å com o Packmol,^{96,97} contendo diferentes concentrações de água e TFE de acordo com a Tabela 3.1.1.1. Todos os sistemas foram simulados com o modelo de água TIP4P/2005⁹⁸ e o campo de força amber03w⁹⁹ para o peptídeo. As moléculas de cossolvente (TFE) foram simuladas com um modelo previamente desenvolvido para reproduzir as propriedades termodinâmicas do TFE puro e misturas

de água-TFE.¹⁰⁰ O mesmo conjunto de parâmetros foi utilizado e validado em trabalhos anteriores para entender o efeito do TFE na estabilidade da hélice.⁸⁷ As simulações subestimaram ligeiramente o conteúdo helicoidal do peptídeo em água, comparado com o resultado de Shalango et al. (1994),⁸⁴ e esta é uma limitação conhecida do campo de força para proteínas na presença do modelo de água atual.¹⁰¹ No entanto, a compatibilidade dos campos de força da água e do cossolvente foi preferida.⁸⁷

Tabela 3.1.1.1: Composições dos sistemas simulados com diferentes concentrações de 2,2,2-Trifluoroetanol (TFE).

Sistemas	Número de moléculas de água	Número de moléculas de TFE
AAQAA ₃ - 0% v/v	5805	0
AAQAA ₃ - 20% v/v	4703	289
AAQAA ₃ - 40% v/v	3584	578
AAQAA ₃ - 50% v/v	3003	723
AAQAA ₃ - 60% v/v	2417	868
AAQAA ₃ - 80% v/v	1214	1158
AAQAA ₃ - 100% v/v	0	1448

Finalmente, realizamos simulações semelhantes do peptídeo em soluções de TFE, nas quais as ligações de hidrogênio intermoleculares entre o oxigênio hidroxílico do TFE e o oxigênio da cadeia principal foram inibidas aumentando o raio de exclusão de van der Waals associado apenas a esses pares de átomos. Especificamente, σ_{ij} foi multiplicado por um fator de escala γ de 1,5 dentro da regra de combinação de Lorentz-Berthelot ($\sigma_{ij} = \gamma((\sigma_i + \sigma_j)/2)$), onde σ_i é o parâmetro do oxigênio hidroxílico do TFE (tipo de átomo oh) e σ_j é o parâmetro do átomo do oxigênio carbonílico do peptídeo (tipo de átomo O). Esses átomos foram escolhidos porque σ é zero para o átomo de hidrogênio do TFE no campo de força.

A partir das simulações com o potencial modificado, exploramos o papel das ligações de hidrogênio entre TFE e a cadeia principal na estabilidade da hélice, usando protocolos semelhantes aos descritos acima. No total, 13 sistemas foram simulados, totalizando 65 μs de simulações. A convergência das simulações foi avaliada calculando erros padrão das estimativas e as funções de autocorrelação do conteúdo médio de α -hélice do peptídeo em todas as simulações. O conteúdo de α -hélice foi descorrelacionado após ~ 50 ns nas simulações, e as médias calculadas a partir de blocos não se desviam das estimativas da amostragem completa o suficiente para invalidar o efeito estabilizador do TFE. Como as simulações REST2 foram realizadas por 500 ns, acreditamos que os

resultados estão bem convergidos para os propósitos das análises atuais.

3.1.2 Cálculo da elipticidade do peptídeo

A atribuição da estrutura secundária para cada resíduo do peptídeo nas simulações foi realizada com o software DSSP (versão 3.0.0).¹⁰² Os espectros teóricos de Dicroísmo Circular (CD, *Circular Dichroism*) do peptídeo AAQAA₃ em água foram calculados com o software SESCA¹⁰³ e comparados com os resultados experimentais obtidos por Shalongo et al. (1994).⁸⁴ Os espectros teóricos foram calculados com conjuntos de bases recomendados para três algoritmos (DSSP,¹⁰² DISICL,¹⁰⁴ e HbSS)¹⁰³, que avaliam a contribuição da cadeia principal (conjuntos de bases DSSP-T, DISICL-dT e HBSS-3) e das cadeias laterais (conjuntos de bases DSSP-1SC3, DISICL-dTSC3 e HBSS-3SC1) na previsão dos espectros de CD. Após avaliar a concordância entre os espectros teóricos e experimentais do peptídeo em água, adotamos o conjunto de bases HBSS-3SC1 para todos os cálculos.

3.1.3 Funções de Distribuição de Mínima-Distância e termodinâmica de solvatação

As MDDFs foram calculadas com o pacote ComplexMixtures.jl²³ usando uma versão discretizada da Equação 2.1.1.1, na qual a densidade foi computada a partir do número médio de distâncias mínimas em cada intervalo de 0,1 Å. As IKBs e os parâmetros de interação preferencial para o cossolvente foram calculados de acordo com as equações 2.1.2.2 e 2.1.2.8, e o parâmetro de hidratação preferencial de acordo com a Equação 2.1.2.9. Utilizamos $R = 15$ Å, que é maior do que o normalmente usado para calcular as IKBs em sistemas de proteína-solvente.¹⁰⁵ Com esse raio de corte, obtivemos integrais convergidas na maioria das simulações, e os parâmetros de interação preferencial foram claros.

3.2 Resultados e discussão

3.2.1 TFE estabiliza a estrutura helicoidal do peptídeo AAQAA₃

Na Figura 3.2.1.1, fornecemos uma visão geral das propriedades estruturais do peptídeo AAQAA₃ conforme obtido a partir de simulações de DM com amostragem ampliada. A Figura

3.2.1.1A mostra os espectros de CD calculados do peptídeo em água (azul) e em soluções aquosas de TFE. É evidente que o conteúdo de α -hélice do peptídeo é sempre maior na presença de TFE do que em água pura. No espectro do peptídeo em água, há uma banda positiva em 184 nm e uma banda negativa em 202 nm, em concordância com a observação experimental.⁸⁴ A segunda banda negativa em torno de 220 nm, característica de α -hélices, não está bem definida, mas à medida que a concentração de TFE aumenta na solução, é possível observar o aumento da intensidade da banda em 221 nm.

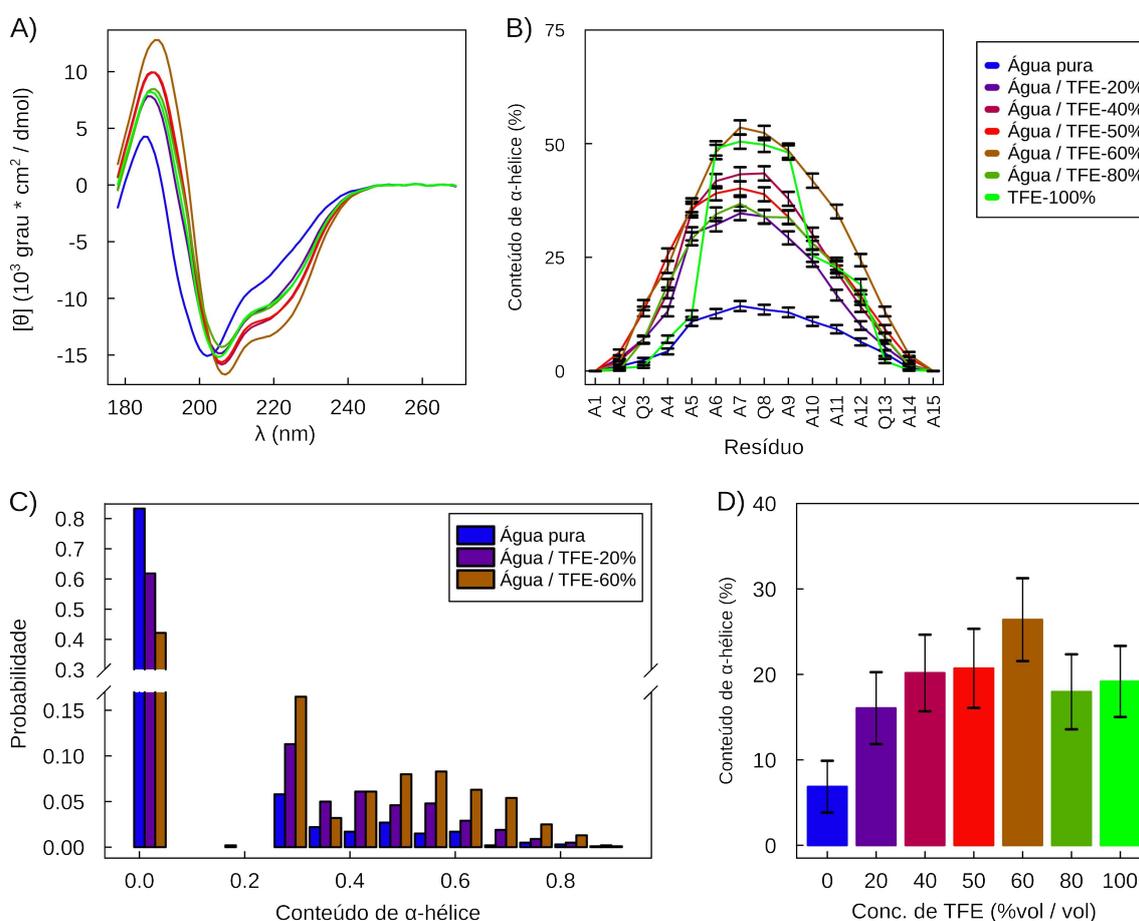


Figura 3.2.1.1: Propriedades estruturais do peptídeo AAQAA₃ em água (azul) e em soluções aquosas de TFE, calculadas a partir das simulações: A) Espectros de CD. O TFE aumenta o conteúdo de α -hélice do peptídeo, como mostrado pelo aumento da intensidade da banda em ~ 190 nm e pelo duplo aumento entre 200-230 nm. B) Prevalência média de α -hélice por resíduo. C) Densidade de probabilidade do conteúdo helicoidal do peptídeo. D) Elipticidade média do peptídeo em cada solução. As barras de erro nas figuras indicam o erro padrão da média das quantidades calculadas.

Conforme mostrado na Figura 3.2.1.1B, o conteúdo de α -hélice por resíduo do peptídeo em TFE é consideravelmente maior do que em água pura, mesmo nas menores concentrações do

cossolvente. A propensão helicoidal dos resíduos no centro do peptídeo é a maior. Em TFE puro, os resíduos centrais mantêm seu comportamento helicoidal, mas os resíduos mais próximos dos grupos terminais perdem sua elipticidade. É importante destacar que em todas as simulações ocorreram flutuações, de modo que o peptídeo visitou conformações com altos e baixos conteúdos helicoidais. De qualquer forma, a população dos estados com conteúdo nulo de α -hélice em água é máxima e mínima na solução de TFE a 60% vol/vol.

A elipticidade máxima do peptídeo é observada em TFE a 60% vol/vol, conforme mostrado na Figura 3.2.1.1D. Esse resultado está de acordo com o fato de que o conteúdo de α -hélice de peptídeos ricos em alanina,¹⁰⁶⁻¹⁰⁸ assim como de peptídeos e proteínas de diferentes sequências,¹⁰⁹⁻¹¹² é máximo em soluções aquosas de TFE, e não no cossolvente puro. A justificativa para a presença desse máximo na estabilidade helicoidal em função da concentração ainda não é bem compreendida,¹¹³ especialmente por ser dependente da sequência de aminoácidos.^{74,107} Alguns autores sugerem que a diminuição da elipticidade com o aumento da concentração de TFE está relacionada à quebra de componentes hidrofóbicos das interações que mantêm estruturas secundárias e terciárias.^{114,115} Alternativamente, à medida que a concentração de TFE aumenta, a solução torna-se mais homogênea, reduzindo a formação de aglomerados do cossolvente, o que pode estar associado à estabilização de α -hélices pelo efeito de volume excluído.^{114,116}

Em resumo, nesta seção mostramos que as simulações são qualitativamente consistentes com a estabilização experimental da hélice do peptídeo AAQAA₃ pelo TFE, tanto por promover a estabilidade helicoidal em concentrações mais baixas de TFE, quanto por induzir um máximo de elipticidade em uma determinada concentração.

3.2.2 Interações diretas e interação preferencial por TFE

Nesta seção, analisaremos a estrutura de solvatação do peptídeo em soluções de água e TFE. As MDDFs para água e TFE em relação ao peptídeo são mostradas nas Figuras 3.2.2.1A e 3.2.2.1B, respectivamente, e apresentam dois picos bem definidos. Esses picos seguem tendências semelhantes à medida que as soluções se tornam mais concentradas. Os primeiros picos (em $\sim 1,8$ Å) indicam a presença de interações específicas (ligações de hidrogênio) com o peptídeo, enquanto os segundos picos (em $\sim 2,6$ Å) são característicos da segunda camada de solvatação e interações não

específicas. Com o aumento da concentração de TFE, os primeiros picos das MDDFs aumentam, enquanto os segundos picos geralmente diminuem. O aumento no primeiro pico provavelmente está associado ao aumento da hidrofobicidade da solução, que favorece interações hidrofílicas diretas com a proteína. Como o aumento da densidade local está inversamente associado ao potencial de força média, a estabilidade das ligações de hidrogênio entre a água e o TFE com a proteína aumenta quando o TFE é adicionado à solução.

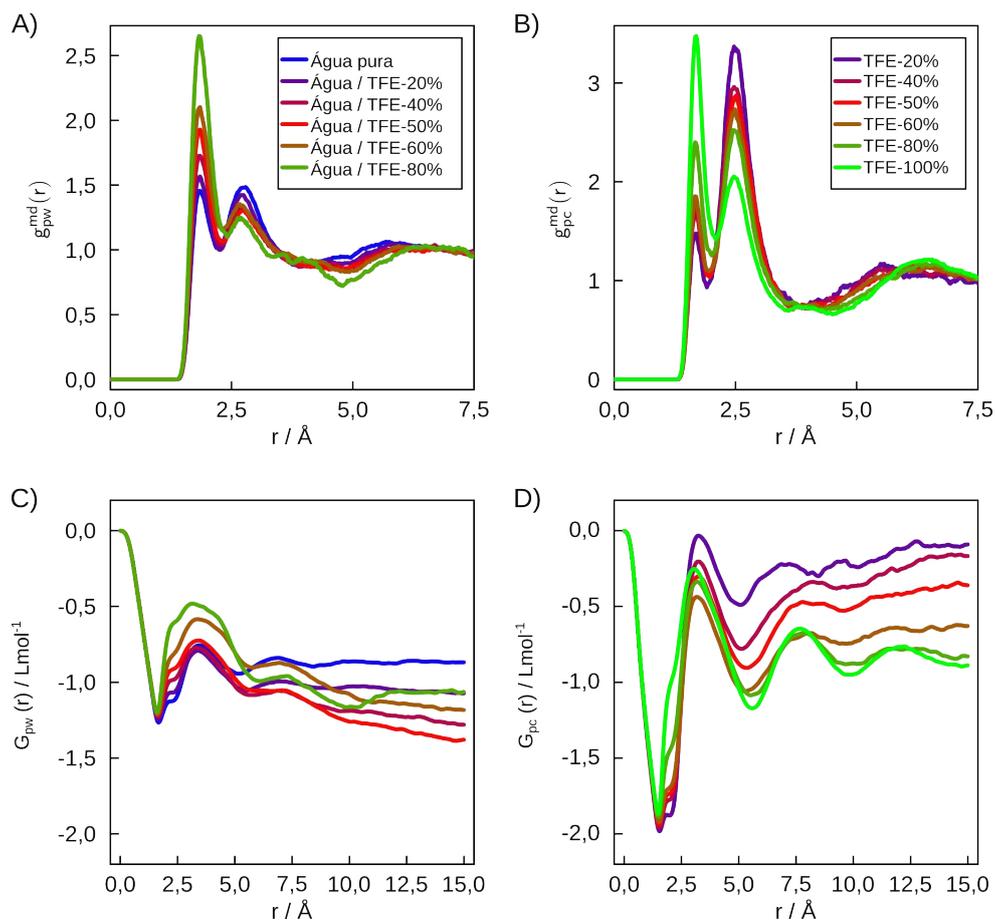


Figura 3.2.2.1: MDDFs A) da água e B) do TFE em função da concentração do cossolvente. As IKBs para água e TFE são mostradas em C) e D), respectivamente.

É importante notar que o aumento no primeiro pico da MDDF da água não significa que o número de ligações de hidrogênio entre as moléculas de água e o peptídeo aumenta. Na verdade, o TFE substitui as ligações de hidrogênio entre o peptídeo e a água. Além disso, a concentração de água é muito diferente nos sistemas contendo TFE (Tabela 3.2.2.1). Por exemplo, a concentração em água pura ($55,1 \text{ mol L}^{-1}$) é cerca de 4,8 vezes maior do que a concentração de água na solução com 80% de TFE ($11,6 \text{ mol L}^{-1}$, Tabela 3.2.2.1). Portanto, para que o número de ligações de

hidrogênio seja o mesmo em ambos os sistemas, o pico em $\sim 1,8 \text{ \AA}$ na solução de 80% de TFE deveria ter uma integral 4,8 vezes maior. De fato, à medida que a solução se torna mais concentrada, as moléculas de TFE são encontradas a distâncias mais curtas e competem com as moléculas de água por ligações de hidrogênio com o peptídeo. Paralelamente, o número de ligações de hidrogênio entre TFE e o peptídeo aumenta progressivamente com a adição de TFE à solução.

Apesar dessa estabilização das moléculas de água em distâncias de ligação de hidrogênio, o TFE exclui as moléculas de água da segunda camada de solvatação e de distâncias maiores (de $\sim 2,0 \text{ \AA}$ até $\sim 6,5 \text{ \AA}$), conforme mostrado na Figura 3.2.2.1A. As interações não específicas entre as moléculas de TFE e o peptídeo são mais pronunciadas do que as interações específicas, particularmente em concentrações menores de TFE.

As Figuras 3.2.2.2C e 3.2.2.2D mostram as IKBs para a água e o TFE, respectivamente. As IKBs determinam se há excesso ou exclusão de cada componente do solvente ao redor do soluto. Em todos os casos, em distâncias curtas, $r < 1,5 \text{ \AA}$, as IKBs exibem valores muito negativos. Essa diminuição acentuada reflete o volume de exclusão entre o peptídeo e as moléculas do solvente, que pode ser compensado por interações soluto-solvente favoráveis, levando à acumulação em regiões mais distantes da solução. A diminuição dos valores das IKBs do TFE nessa distância é maior do que a da água porque o TFE é uma molécula mais volumosa.¹¹⁷

Apenas uma parte do volume de exclusão é compensado pela acumulação de moléculas de água. As IKBs para a água aumentam após $\sim 1,5 \text{ \AA}$, como mostrado na Figura 3.2.2.2C, mas ainda são negativas em distâncias maiores (a tendência qualitativa das IKBs é consistente, apesar de sua convergência não ser completamente satisfatória em todas as simulações). Na Figura 3.2.2.2D, observa-se que as moléculas de TFE se acumulam de maneira mais eficaz na superfície da proteína do que as moléculas de água e, embora as IKBs finais também sejam negativas, são maiores do que as da água. Por exemplo, na solução com 20% de TFE, as IKBs do TFE convergem para valores próximos de zero, embora essa densidade relativa de acumulação diminua conforme a solução se torna mais concentrada.

A interação preferencial do peptídeo pode ser quantificada pelo parâmetro de interação preferencial, Γ_{pc} , calculado a partir das IKBs da água e do cossolvente (Equação 2.1.2.8). Os valores de Γ_{pc} encontrados aqui estão expressos na Tabela 3.2.2.1. Γ_{pc} é positivo para todas as concentrações. Dessa forma, o TFE solvata preferencialmente a proteína, de forma consistente

com os resultados publicados anteriormente.^{86,87} Esse resultado é interessante, pois mostra que a acumulação do cossolvente na superfície da proteína nem sempre favorece a desnaturação, como geralmente observado, por exemplo, em soluções de ureia e cloreto de guanidina.¹¹⁸ Pelo contrário, observamos que o TFE interage preferencialmente com o peptídeo e estabiliza a estrutura da hélice.

Tabela 3.2.2.1: Concentrações efetivas das soluções de TFE, IKBs para o TFE (G_{pc}) e água (G_{pw}) em relação ao peptídeo AAQAA₃, e parâmetro de interação preferencial (Γ_{pc}). As simulações com os potenciais modificados são destacadas por um asterisco

Sistemas	[Água] (mol L ⁻¹)	[TFE] (mol L ⁻¹)	G_{pc} (L mol ⁻¹)	G_{pw} (L mol ⁻¹)	Γ_{pc}
AAQAA ₃ - 20%	45,268	2,757	-0,0862	-1,075	2,727
AAQAA ₃ - 20%*	45,381	2,726	1,100	-1,321	6,601
AAQAA ₃ - 40%	34,504	5,505	-0,162	-1,280	6,154
AAQAA ₃ - 40%*	34,571	5,501	0,00874	-1,462	8,088
AAQAA ₃ - 50%	28,933	6,901	-0,366	-1,384	7,023
AAQAA ₃ - 50%*	28,880	6,912	-0,562	-1,220	4,549
AAQAA ₃ - 60%	23,212	8,294	-0,617	-1,186	4,716
AAQAA ₃ - 60%*	23,158	8,311	-0,874	-0,944	0,578
AAQAA ₃ - 80%	11,617	11,057	-0,822	-1,060	2,630
AAQAA ₃ - 80%*	11,491	11,091	-1,113	0,141	-13,911

As MDDFs podem ser decompostas em suas contribuições atômicas e nas contribuições de cada tipo de resíduo, proporcionando uma interpretação molecular detalhada das interações soluto-solvente.^{23,24} A decomposição das MDDFs, em essência, revela a frequência com que cada átomo (ou grupo de átomos) está mais próximo de qualquer átomo do soluto em cada distância, com a soma de todas as curvas de contribuição sendo igual à MDDF total. As interações do peptídeo com o solvente foram ilustradas usando a MDDF da solução com 60% de TFE, mas os resultados observados são qualitativamente representativos de todas as concentrações. Como mostrado na Figura 3.2.2.2A, as interações específicas (ligações de hidrogênio) entre o cossolvente e o peptídeo são totalmente determinadas pela contribuição do hidrogênio hidroxílico (cinza escuro - H_{TFE}). O H_{TFE} pode interagir por ligações de hidrogênio tanto com os resíduos polares da cadeia lateral (resíduos de glutamina) (Figura 3.2.2.2B) quanto com o oxigênio amídico da cadeia principal

(Figura 3.2.2.2C). Não há interação específica entre as moléculas de TFE e o nitrogênio amídico (Figura 3.2.2.2C). Por outro lado, as interações não específicas envolvem predominantemente os grupos trifluorometil e os hidrogênios alifáticos das moléculas de TFE com as cadeias laterais não polares do peptídeo (Figuras 3.2.2.2A e 3.2.2.2B).

As interações específicas entre o TFE e o peptídeo estão de acordo com o mecanismo direto proposto por Rajan & Balaram (2009)⁷¹ (Figura 3.0.0.1) para explicar o efeito estabilizador do TFE em hélices. De acordo com o mecanismo direto, as moléculas de TFE interagem com o oxigênio da cadeia principal via ligações de hidrogênio. No entanto, essa interação pode ocorrer sem a ruptura das ligações de hidrogênio intramoleculares do peptídeo (O_{backbone} e H_{backbone}), que desestabilizariam a hélice.⁷¹ Ao mesmo tempo, o grupo trifluorometil impediria a aproximação e interação de moléculas de água com o nitrogênio amídico da cadeia principal. Na próxima seção, avaliaremos o efeito dessas ligações de hidrogênio peptídeo-TFE e mostraremos que, inesperadamente, elas favorecem conformações desenoveladas do peptídeo.

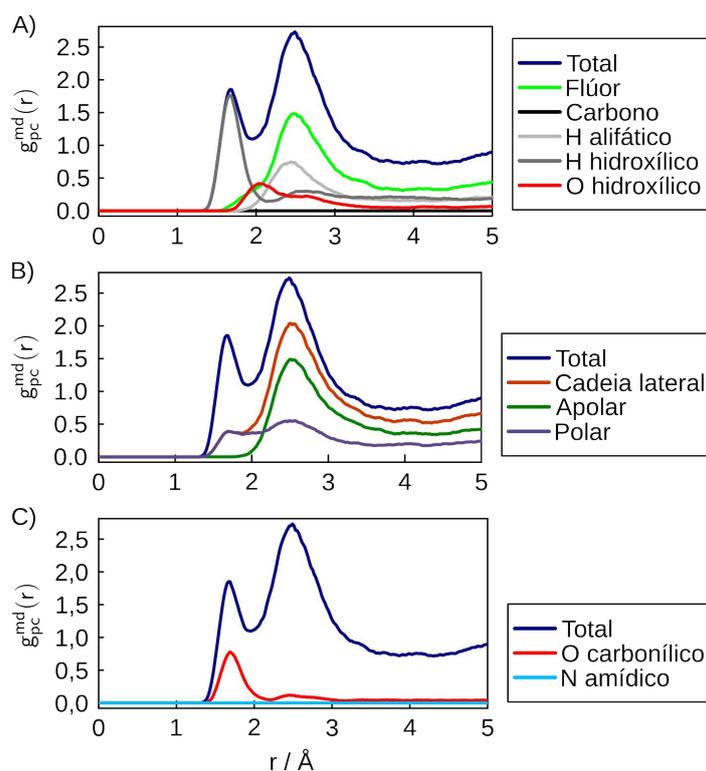


Figura 3.2.2.2: MDDF do TFE em relação ao peptídeo AAQAA₃ e contribuições dos grupos, na solução de 60% vol/vol. A) Contribuição dos átomos de TFE. B) Contribuições das cadeias laterais do peptídeo. C) Contribuições dos átomos da cadeia principal. A única contribuição dos átomos da cadeia principal para a MDDF do TFE é devido ao átomo de oxigênio, que forma ligações de hidrogênio com o hidrogênio hidroxílico do TFE.

3.2.3 Rompimento da interação entre o TFE e a cadeia principal do peptídeo leva à estabilização da α -hélice

Os resultados apresentados até agora mostram claramente o papel estabilizador da hélice do TFE (Figuras 3.2.1.1A e 3.2.1.1B). Os modos de interação do TFE com o peptídeo, conforme determinado pela caracterização da estrutura do solvente com as MDDFs, apoiam que mecanismos diretos e indiretos poderiam estar envolvidos na estabilização da hélice.

Para investigar o papel das interações diretas entre o TFE e a cadeia principal na estabilização da hélice, aumentamos os raios associados especificamente à interação entre o hidrogênio polar do TFE e os oxigênios carbonílicos da cadeia principal do peptídeo. Nenhuma outra interação foi modificada. Com isso, evitamos a formação de ligações de hidrogênio entre esses grupos e suprimimos o possível papel dessa interação na exclusão local de moléculas de água da cadeia principal. Esperávamos demonstrar que a água se tornaria um ligante mais forte para a cadeia principal, competindo com as ligações de hidrogênio intramoleculares do peptídeo, e, assim, a hélice seria desestabilizada.

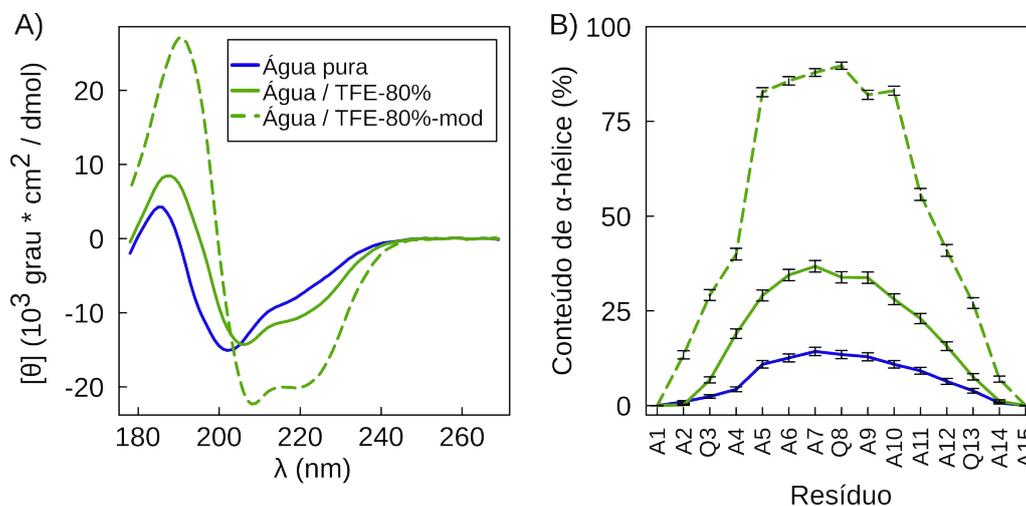


Figura 3.2.3.1: Propriedades estruturais do peptídeo AAQAA₃ em água (azul) e em 80% de TFE, computadas a partir da simulação padrão (verde) e com potencial modificado (verde tracejada): A) Espectros de CD. B) Prevalência média de α -hélice por resíduo.

Surpreendentemente, após modificar o potencial de interação para evitar interações diretas entre o TFE e a cadeia principal do peptídeo, observamos que o conteúdo de α -hélice do peptídeo aumenta significativamente. Os espectros de CD e o conteúdo de α -hélice por resíduo

das simulações com os potenciais modificados estão mostrados nas Figuras 3.2.3.1A e 3.2.3.1B, para a solução de 80% vol/vol de TFE em água. A estabilização da hélice foi observada em todas as concentrações, devido ao rompimento da ligação de hidrogênio entre o TFE e a cadeia principal. Está claro, então, que as interações diretas do TFE com a cadeia principal estabilizam conformações desenoveladas do peptídeo, em vez de suportar a formação das hélices.

3.2.4 Interação preferencial na ausência da ligação de hidrogênio entre o TFE e a cadeia principal

As estruturas de solvatação das simulações com o potencial modificado foram caracterizadas a partir das MDDFs para um melhor entendimento dos fatores que contribuem para a estabilização do peptídeo e o aumento do conteúdo de α -hélice. Novamente, observamos dois picos distintos nas funções de distribuição de água e TFE (Figura 3.2.4.1A e 3.2.4.1B). No entanto, as densidades relativas das moléculas de água e TFE nas primeiras e segundas camadas de solvatação mudam. A distâncias curtas ($\sim 1,8$ Å), a densidade relativa das moléculas de água aumenta progressivamente com o acréscimo da concentração de TFE, como observado com o potencial original. Por outro lado, o pico associado às moléculas de TFE permanece quase constante, em nítido contraste com as distribuições observadas na Figura 3.2.2.1B para as interações originais. Essa observação indica que, com o potencial padrão, o aumento da concentração de TFE desloca as moléculas de água da cadeia principal do peptídeo. Esse efeito é suprimido nas simulações com o potencial modificado, e as interações específicas restantes (com a cadeia lateral dos resíduos de glutamina) ocorrem proporcionalmente à concentração de TFE.

Como as moléculas de TFE não competem mais com a água por interações específicas com a cadeia principal, a densidade das moléculas de água nessa região é maior do que nas simulações com os potenciais padrão (Figuras 3.2.2.1A e 3.2.4.1A). Assim, a afinidade da água pela cadeia principal aumenta, como esperado, mas com o resultado inesperado de estabilizar a hélice.

As Figuras 3.2.4.1C e 3.2.4.1D mostram as IKBs para água e TFE, respectivamente, obtidas nas simulações com potenciais modificados. A tendência geral das curvas é semelhante à das simulações com o potencial padrão (Figuras 3.2.2.1C e 3.2.4.1D), mas as afinidades das moléculas do solvente (água e TFE) para o peptídeo diferem. Em particular, as IKBs da água e TFE estão mais próximas umas das outras, implicando que a interação preferencial pelo TFE foi reduzida para

a maioria dos sistemas, como mostrado na Tabela 3.2.2.1. Novamente, isso é esperado, já que uma das interações específicas do TFE com o peptídeo foi removida. No entanto, para todos, exceto para a solução de 80%, os parâmetros de interação preferencial são positivos e ainda implicam que o TFE solvata preferencialmente o peptídeo. Entretanto, os conjuntos de estruturas do peptídeo não são os mesmos nas simulações com os potenciais padrão e modificado, uma vez que, nas últimas, o conteúdo helicoidal do peptídeo é maior em todas as simulações, e os peptídeos apresentam áreas superficiais médias menores.

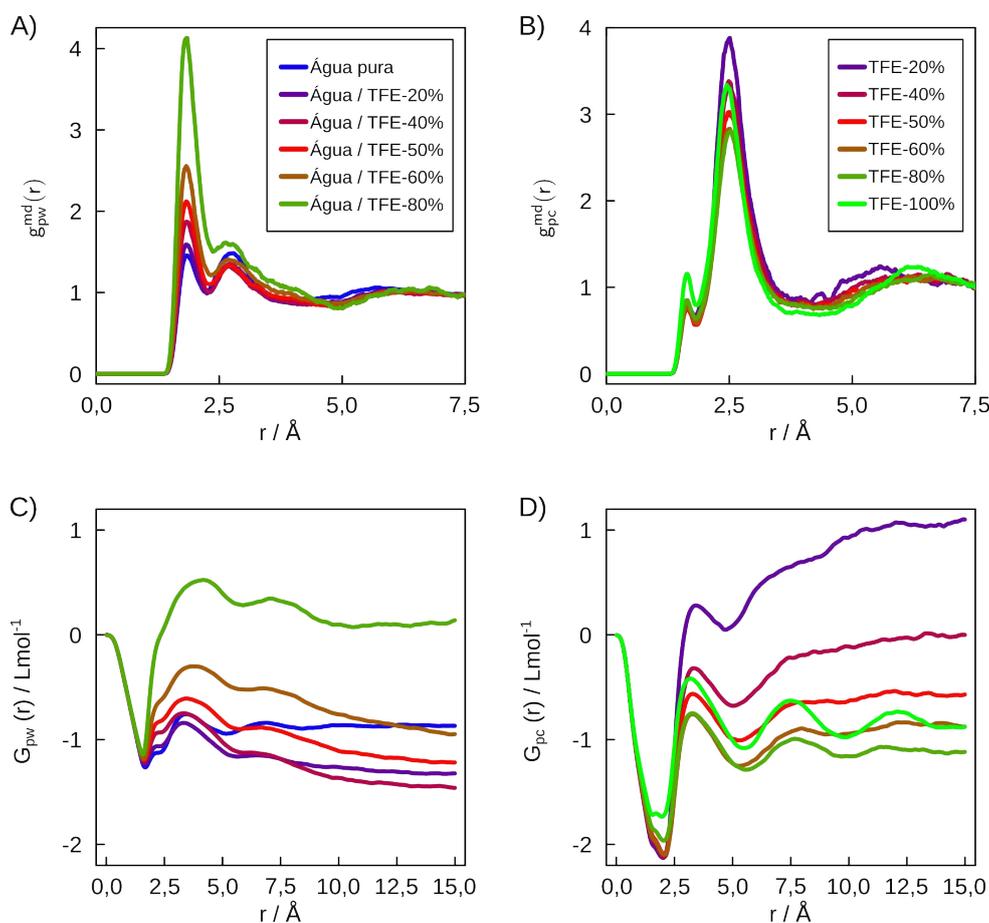


Figura 3.2.4.1: MDDFs A) da água e B) do TFE em função da concentração do cossolvente para as simulações com potenciais modificados. As IKBs para água e TFE são mostradas em C) e D), respectivamente.

Em resumo, as simulações com o potencial modificado levaram a um aumento no conteúdo de α -hélice do peptídeo, em contradição com a hipótese do efeito estabilizador da interação entre o TFE e a cadeia principal. O peptídeo continua a ser preferencialmente solvatado por TFE em todas as soluções aquosas, exceto na mais concentrada de TFE, na ausência da

ligação de hidrogênio com a cadeia principal. Portanto, as interações não específicas, em sua maioria apolares, entre TFE e o peptídeo são responsáveis pela interação preferencial e pela proteção da estrutura helicoidal (veja a seção 3.2.5). Nesse sentido, nossos resultados apoiam aqueles de Roccatano et al. (2002),⁷² que mostraram que as interações apolares têm um papel na estabilização das hélices. O TFE apresenta interações fracas com resíduos apolares e não perturba significativamente os contatos hidrofóbicos intra-peptídicos. Paralelamente, também foi relatado que a presença de TFE na solução causa uma diminuição na constante dielétrica local, o que promove a formação de hélices ao fortalecer as ligações de hidrogênio intra-peptídicas, e tais efeitos não podem ser desconsiderados.¹¹⁹

3.2.5 Interações do TFE com os *ensembles* helicoidais e não helicoidais

Nas seções anteriores, discutimos que as interações não específicas com o peptídeo são responsáveis pelo efeito protetor das hélices. Aqui, agrupamos as estruturas das simulações com os potenciais padrão em conjuntos helicoidais e não helicoidais e caracterizamos as estruturas do solvente em cada conjunto. Note, na Figura 3.2.1.1C, que os sistemas sem qualquer conteúdo helicoidal representam cerca de 60% da amostragem, enquanto o agrupamento de estruturas com elipticidade compreende os outros 40% e apresenta, em média, cerca de 40% dos resíduos em conformação helicoidal. Assim, uma quantidade significativa de peptídeo não estruturado ainda está presente no conjunto helicoidal.

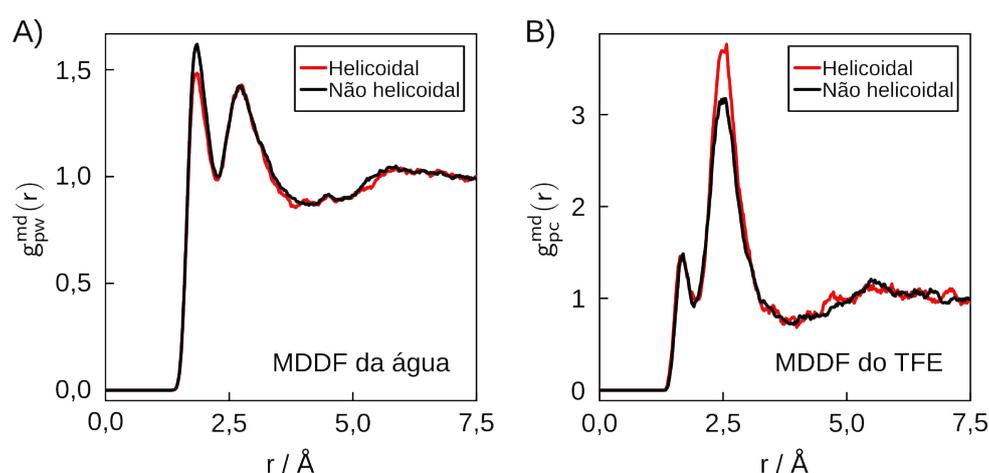


Figura 3.2.5.1: MDDFs entre o pept\u00eddeo e A) \u00e1gua e B) em 20% de TFE para os conjuntos helicoidal (vermelho) e n\u00e3o helicoidal (preto). O conjunto desenovelado forma mais liga\u00e7\u00f5es de hidrog\u00eanio com a \u00e1gua e apresenta menos intera\u00e7\u00f5es n\u00e3o espec\u00edficas com TFE.

As Figuras 3.2.5.1A e 3.2.5.1B mostram as MDDFs da água e do TFE para os conjuntos helicoidal e não helicoidal no sistema com 20% de TFE (as MDDFs dos outros sistemas seguem aproximadamente a mesma tendência). A Figura 3.2.5.1A mostra que os estados desenovelados, sem estrutura helicoidal, formam ligações específicas ligeiramente mais fortes com a água do que o conjunto de estruturas com algum conteúdo helicoidal, apoiando o papel das interações específicas com a água na desestabilização das hélices. Isso pode ser observado pelo aumento na densidade de moléculas de água no primeiro pico da distribuição ($\sim 1,8 \text{ \AA}$). As interações não específicas (segundo pico) permanecem aproximadamente constantes.

A Figura 3.2.5.1B, por outro lado, mostra o efeito oposto para o TFE: o primeiro pico permanece constante, mas a distribuição de moléculas de TFE diminui no segundo pico, indicando que os estados desenovelados interagem menos favoravelmente com o TFE. Como resultado da diminuição na força das interações não específicas entre os peptídeos e o TFE no conjunto não helicoidal, moléculas de água se acumulam a curtas distâncias. Além disso, as interações não específicas entre o TFE e o peptídeo são mais fortes no conjunto helicoidal, corroborando sua importância na estabilização das hélices.

3.3 Conclusões

Investigamos os efeitos do 2,2,2-Trifluoroetanol (TFE) no equilíbrio conformacional do peptídeo AAQAA₃ usando simulações de DM de amostragem ampliada com o método de troca de réplicas REST2 para obter uma amostragem exaustiva do equilíbrio de enovelamento. A elipticidade do peptídeo é sempre maior na presença de TFE, conforme mostrado pelos espectros teóricos de CD e pelo cálculo direto do conteúdo de α -hélice. Todas essas descobertas são consistentes com os efeitos estabilizadores observados do TFE.^{75,86-88}

As MDDFs demonstraram que o TFE interage com o oxigênio da cadeia principal do peptídeo por meio de ligações de hidrogênio e exibe interações não específicas com suas cadeias laterais por meio do grupo trifluormetil e hidrogênios alifáticos. O TFE compete com as moléculas de água por interações específicas com o peptídeo. Além disso, o peptídeo é preferencialmente solvatado pelo TFE em todas as concentrações. Conformações helicoidais exibem interações não específicas mais fortes com o TFE do que os estados desestruturados.

As ligações de hidrogênio entre o TFE e a carbonila sugerem um mecanismo pelo qual

o TFE restringiria o acesso das moléculas de água à cadeia principal, impedindo a competição com as ligações de hidrogênio intra-hélice. No entanto, ao modificar o potencial de interação entre o oxigênio da carbonila na cadeia principal e o hidrogênio polar do TFE, investigamos o papel dessa interação na estabilidade das hélices. Inesperadamente, o conteúdo de α -hélice aumentou significativamente, implicando que a interação do TFE com a cadeia principal estabiliza o estado desenovelado do peptídeo. Mesmo na ausência dessa interação polar forte, a presença de TFE desidrata o peptídeo. Portanto, o papel estabilizador do TFE na hélice está mais provavelmente associado ao "revestimento" da hélice, que exclui as moléculas de água da segunda camada de solvatação da cadeia principal, e da primeira camada de solvatação das cadeias laterais do peptídeo. Mecanismos indiretos associados a perturbações na estrutura da solução não podem ser descartados.

Capítulo 4

Enovelamento da BdpA e estruturas de hidratação

O domínio B da proteína A (BdpA) é um alvo atraente para estudos de enovelamento de proteínas por possuir uma topologia simples contendo três hélices (Figura 4.0.0.1)¹²⁰ e uma cinética de enovelamento rápida.^{121,122} O enovelamento da BdpA tem sido investigado por meio de simulações atômicas e *coarse-grained*,^{123–125} bem como por métodos experimentais,^{126,127} demonstrando concordância do conjunto de enovelamento com uma descrição teórica baseada na teoria da paisagem energética (seção 1.2).^{124–130} O mecanismo de enovelamento da BdpA pode ser representado por uma cinética de dois estados, na qual os estados nativo (N) e desenovelado (U) coexistem em equilíbrio sem uma população significativa de estados intermediários.^{126,127,131,132} Estados intermediários de curta duração, naturalmente, existem e foram caracterizados recentemente.^{122,133}

No que diz respeito à caracterização da variabilidade estrutural dos estados N e U, simulações em alta temperatura¹²⁵ e experimentos recentes de troca H/D mostraram que a estrutura residual observada nos estados desnaturados está associada à hélice III.¹³³ O estudo experimental também sugeriu a possibilidade de uma ponte salina entre os resíduos Lys50 e Asp54 como responsável pela estabilidade da hélice III, além da hipótese de que essa hélice poderia atuar como um sítio de iniciação do enovelamento do BdpA.¹³³ Ao mesmo tempo, outros dados experimentais da análise de valores Φ mostram que a hélice III está pouco formada no estado de transição, enquanto a hélice II está bem formada.^{126,127,134,135} Portanto, não há um consenso claro sobre os mecanismos subjacentes à formação/deformação das hélices da BdpA durante o enovelamento.

Neste trabalho, investigamos o enovelamento da BdpA combinando simulações *coarse-*

grained usando os modelos baseados na estrutura do $C\alpha$ ($C\alpha$ -SBM)⁴⁸ e simulações atomísticas. As simulações com os $C\alpha$ -SBM permitiram uma amostragem exaustiva da superfície de enovelamento, incluindo os estados de transição. Por sua vez, simulações atomísticas de cada estrutura enovelada permitiram obter estruturas equilibradas para uma análise detalhada de seu conteúdo de α -hélice e a interação com o solvente. Os resultados mostram que a diversidade conformacional do estado nativo é devida principalmente à variabilidade conformacional da hélice I. As hélices I, II e III ocorrem de maneira fracamente correlacionada, com coeficientes de correlação de Spearman de 0,15 (I e II), 0,12 (I e III) e 0,26 (II e III). Esses resultados sugerem a maior cooperatividade entre as hélices II e III. Nossos resultados permitem o agrupamento de estruturas parcialmente enoveladas da BdpA com base em sua estrutura secundária, possibilitando a compreensão dos efeitos de solvente na estabilidade relativa dos conjuntos de estruturas de enovelamento. Aqui, essa estabilidade é ilustrada pela dependência estrutural das camadas de hidratação da proteína, conforme calculado pelas MDDFs.

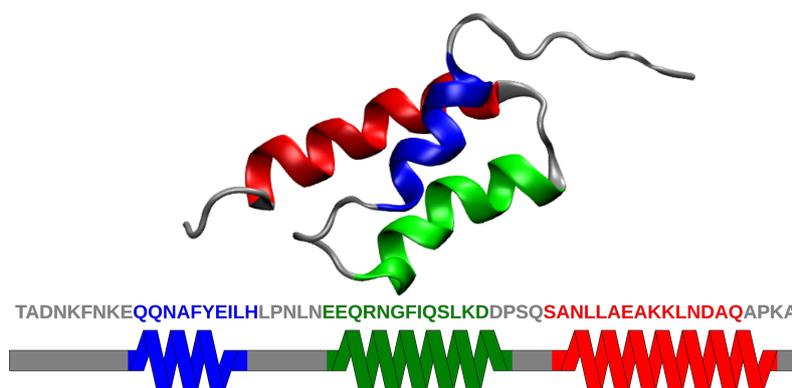


Figura 4.0.0.1: Representação em cartoon do domínio B da proteína A (BdpA), PDB: 1BDD.¹³⁶ A Hélice I (¹⁰QQNAFYEILH¹⁹), a Hélice II (²⁵EEQRNGFIQSLK³⁷) e a Hélice III (⁴²SANLLAEAKKLNDAAQ⁵⁶) são representadas em azul, verde e vermelho, respectivamente. Na parte inferior, tem-se a sequência da BdpA colorida de acordo com a representação em cartoon.

4.1 Metodologia

4.1.1 Simulações com os Modelos Baseados na Estrutura e análises

Inicialmente, obtivemos o conjunto de enovelamento da BdpA (PDB: 1BDD)¹³⁶ usando os Modelos Baseados na Estrutura dos carbonos alfa ($C\alpha$ -SBM, seção 2.2.2).^{48,50} A BdpA é uma proteína modelo para estudos de enovelamento e tem sido amplamente

estudada.^{128-130,137,138} As simulações com os SBMs permitem a amostragem exaustiva do espaço conformacional da proteína, fornecendo, dentro das aproximações do modelo, um conjunto completo de estados de enovelamento de pequenas proteínas como a BdpA. A convergência do conjunto pode ser demonstrada pela obtenção de múltiplas transições entre os estados enovelados e desenovelados na temperatura de transição (conhecida como temperatura de enovelamento, T_f), no caso de um modelo de enovelamento de dois estados, como o estudado aqui.

O mapa de contatos da BdpA foi determinado com o algoritmo *Contact of Structural Units* (CSU).¹³⁹ O $C\alpha$ -SBM foi gerado usando o servidor web SMOG (<https://smog-server.org/>).¹⁴⁰ Todas as simulações com o $C\alpha$ -SBM foram realizadas com o Gromacs 4.6.7.¹⁴¹ Para determinar a temperatura de enovelamento, dois conjuntos de simulações foram realizados. O primeiro conjunto utilizou temperaturas entre 80 e 160K no Gromacs (as unidades reais não são significativas para os SBMs) com um incremento de temperatura de 10K. Uma vez identificada aproximadamente a temperatura de calor específico máximo, C_v , um novo conjunto de simulações com temperaturas variando entre 113 e 119K com incrementos de 1K foi realizado para localizar a temperatura de enovelamento dentro de uma precisão de ~ 1 K. Em unidades de temperatura reduzida, as simulações foram realizadas entre 0,66 e 1,33 com passos de 0,083 e entre 0,94 e 0,99 com passos de 0,0083 unidades de temperatura para o primeiro e segundo conjuntos de simulações, respectivamente. Essas simulações foram analisadas com o *Weighted Histogram Analysis Method* (WHAM)¹⁴², conforme implementado no SMOG2,¹⁴³ para determinar a dependência da temperatura do calor específico, $C_v(T)$, e o potencial de força média, $F(Q)$, em função da fração de contatos nativos (Q).

Cada simulação SBM consistiu em 5×10^8 passos com um intervalo de tempo de 0,0005 unidades reduzidas, a temperatura constante. Um contato foi considerado nativo se a distância entre os átomos de $C\alpha$ correspondentes não fosse maior que 20% da observada na estrutura experimental de RMN.¹³⁶ A temperatura de calor específico máximo é a temperatura de enovelamento da proteína (T_f). Na T_f , os estados enovelado e desenovelado têm populações similares. A simulação SBM realizada na T_f foi usada para estudar o conjunto de enovelamento. A convergência estatística dos valores de Q na temperatura de enovelamento foi confirmada pela média em blocos (calculado com o pacote MolSimToolkit.jl - versão 1.3.4), que mostra que o tempo de correlação característico de Q é muito inferior ao tempo total de simulação.

4.1.2 Enovelamento da BdpA no espaço de fase 2D

O enovelamento da BdpA foi visualizado com o *Energy Landscape Visualization Method* (ELViM).^{124,144–149} Dado que o enovelamento de proteínas ocorre em um espaço multidimensional, sua visualização depende de uma redução de dimensionalidade. O método ELViM utiliza as matrizes de distâncias internas das conformações para definir uma métrica robusta de similaridade entre estruturas, sem uma definição prévia de uma coordenada de reação. A matriz de similaridades entre as conformações é projetada em um espaço bidimensional, otimizando as distâncias entre os pontos de modo que estas sejam correlacionadas com a dissimilaridade entre as estruturas. Para essa projeção, aplicamos a técnica *Force Scheme*,¹⁵⁰ conforme implementada no ELViM.¹²⁴

4.1.3 Análise da estrutura secundária a partir das estruturas contendo todos os átomos

Para calcular as estruturas secundárias, as representações de todos os átomos de 5.000 estruturas da simulação com o $C\alpha$ -SBM, realizada na T_f , foram reconstruídas usando o software Pulchra.¹⁵¹ As caixas de simulação foram construídas com uma distância mínima de 12,0 Å dos extremos da proteína, utilizando o Packmol.^{96,97} Isso resultou em volumes de caixa variando de $1,428 \times 10^5 \text{ \AA}^3$ a $4,231 \times 10^5 \text{ \AA}^3$, com o número de moléculas de água variando entre 4401 e 13775. A estrutura da proteína foi restringida aplicando potenciais harmônicos com constantes de força de 25 kcal mol⁻¹ nos átomos de $C\alpha$, de modo que a topologia SBM fosse preservada, permitindo ao mesmo tempo a relaxação dos átomos reconstruídos. O campo de força CHARMM36¹⁵² para a proteína e o modelo de água TIP3P¹⁵³ foram usados. Todas as simulações atomísticas foram realizadas no Gromacs 2021.2¹⁵⁴ a 298,15 K, 1 atm, e com intervalo de tempo de 2 fs. Inicialmente, o sistema foi minimizado em até 20.000 passos utilizando o método *Steepest Descent* e equilibrado por 1 ns no *ensemble* de volume e temperatura constantes (NVT), seguido por 1 ns de simulação de volume e pressão constantes (NPT). A temperatura e a pressão foram controladas usando o termostato de Berendsen¹⁵⁵ modificado e o barostato de Parrinello-Rahman.⁹¹ Finalmente, simulações de produção de 10 ns foram realizadas no conjunto NPT para cada sistema, totalizando 50 μs de simulação. O conteúdo de estrutura secundária das estruturas foi calculado com o método DSSP, usando o pacote ProteinSecondaryStructures.jl.¹⁵⁶ Os resíduos pertencentes às hélices I, II e III (Figura 4.0.0.1) foram atribuídos de acordo com a literatura.^{126,127,134,157} A atribuição do conteúdo de α -hélice com o DSSP

exigiu estruturas de todos os átomos equilibradas, uma vez que as estimativas obtidas diretamente dos modelos reconstruídos com Pulchra falharam. O conteúdo de α -hélice obtido do conjunto estruturas de enovelamento foi mapeado na projeção ELViM para obter gráficos de contorno com a função *histogram2dcontour* do pacote PlotlyJS.jl v0.18.10. O coeficiente de correlação de postos de Spearman foram calculados para identificar as correlações entre as hélices durante o enovelamento.

4.1.4 Base molecular da hidratação

A hidratação de cada estrutura foi estudada usando as MDDFs e a teoria de soluções de KB, descritas na seção 2.1. As MDDFs, IKBs e mapas de solvatação foram calculados com o pacote ComplexMixtures.jl.¹⁵⁸ Calculamos as MDDFs e as IKBs para os subconjuntos dos *ensembles* com diferentes características de enovelamento de forma independente, compreendendo médias de dezenas ou centenas de estruturas em cada caso.

A metodologia descrita acima está integrada em um pipeline abrangente (Figura 4.1.4.1). As etapas iniciais, destacadas em preto, foram realizadas para validar o enovelamento da BdpA com base em trabalhos anteriores. As etapas sucessivas, destacadas em verde, foram implementadas neste trabalho para investigar a associação entre o enovelamento da BdpA, a formação de sua estrutura secundária e a estrutura do solvente ao redor de estados parcialmente enovelados.

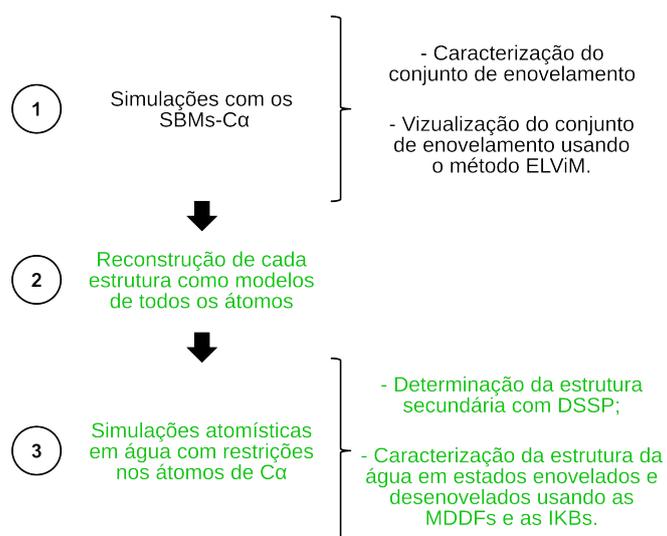


Figura 4.1.4.1: Pipeline para estudar o enovelamento da BdpA. As etapas iniciais (preto) validaram o enovelamento da BdpA com base em trabalhos anteriores, enquanto que os passos incrementais (verde) permitiram investigar a relação entre o enovelamento da BdpA, a formação de sua estrutura secundária e a estrutura do solvente em torno de estados parcialmente enovelados.

4.2 Resultados e discussão

Os resultados são divididos em três partes: 1) a validação do conjunto de enovelamento obtido; 2) a análise da elipticidade da estrutura; 3) *insights* sobre as estruturas de solvatação dos estados parcialmente enovelados. Dessa forma, na seção 4.2.1, mostramos as propriedades do conjunto simulado e sua consistência com simulações anteriores e resultados experimentais. Na seção 4.2.2, caracterizamos a formação das hélices da proteína BdpA com um nível de detalhe sem precedentes, revelando sua correlação com a natureza heterogênea dos conjuntos enovelados e desenovelados, e a correlação entre os conteúdos de α -hélice de cada elemento da estrutura. Finalmente, na seção 4.2.3, mostramos como as estruturas de solvatação da proteína variam em cada um dos conjuntos desenovelados em relação ao estado enovelado.

4.2.1 Enovelamento da BdpA em um espaço de fase 2D

Proteínas com mecanismos de enovelamento de dois estados exibem uma temperatura de enovelamento bem definida, pela qual a distribuição de moléculas em uma propriedade mensurável é bimodal. O pico agudo no perfil de $C_v(T)$ na Figura 4.2.1.1A mostra que o modelo de dois estados é uma boa representação do enovelamento da BdpA a partir do SBM. Não há estados intermediários estáveis neste caso, mas sim um conjunto de estados intermediários de curta duração que são distintos entre si.

Também é possível identificar os conjuntos nativo (N) e desenovelado (U) a partir do perfil de energia livre em função da fração de contatos nativos $F(Q)$ em T_f (Figura 4.2.1.1B). Existe uma única barreira de energia bem definida conectando os estados N ($Q \sim 0,8$) e U ($Q \sim 0,3$). Na Figura 4.2.1.1C, os valores de Q em função dos intervalos de tempo na temperatura de enovelamento (T_f) mostram claramente várias transições entre os estados N e U, indicando uma boa amostragem da transição. A proteína BdpA não visita estados completamente estendidos ($Q \sim 0$), o que foi sugerido como consequência de alguns contatos nativos de alta afinidade. A formação de contatos e a distância média entre pares de átomos revelaram três contatos com a maior probabilidade de serem preservados (38-42, 83,52%; 38-45, 84,18% e 20-31, 86,00%), envolvendo principalmente a volta I (resíduos de 19-25) e a volta II (resíduos de 37-42). Além disso, as distâncias médias desses pares de átomos são próximas às observadas na estrutura experimental.

Na Figura 4.2.1.1D, ilustramos o conjunto de enovelamento mapeando a densidade de

probabilidade do conjunto em função dos RMSDs (eixo y), calculados em relação à estrutura nativa (PDB: 1BDD), e à fração de contatos nativos (eixo x). Os estados enovelados abrangem uma faixa de valores de Q de aproximadamente 0,5 a 0,9 e valores de RMSD de 0,2 a 0,6 nm. O conjunto desenovelado exibe valores de Q entre 0,2 e 0,4, e RMSDs maiores entre aproximadamente 0,9 e 1,6 nm.

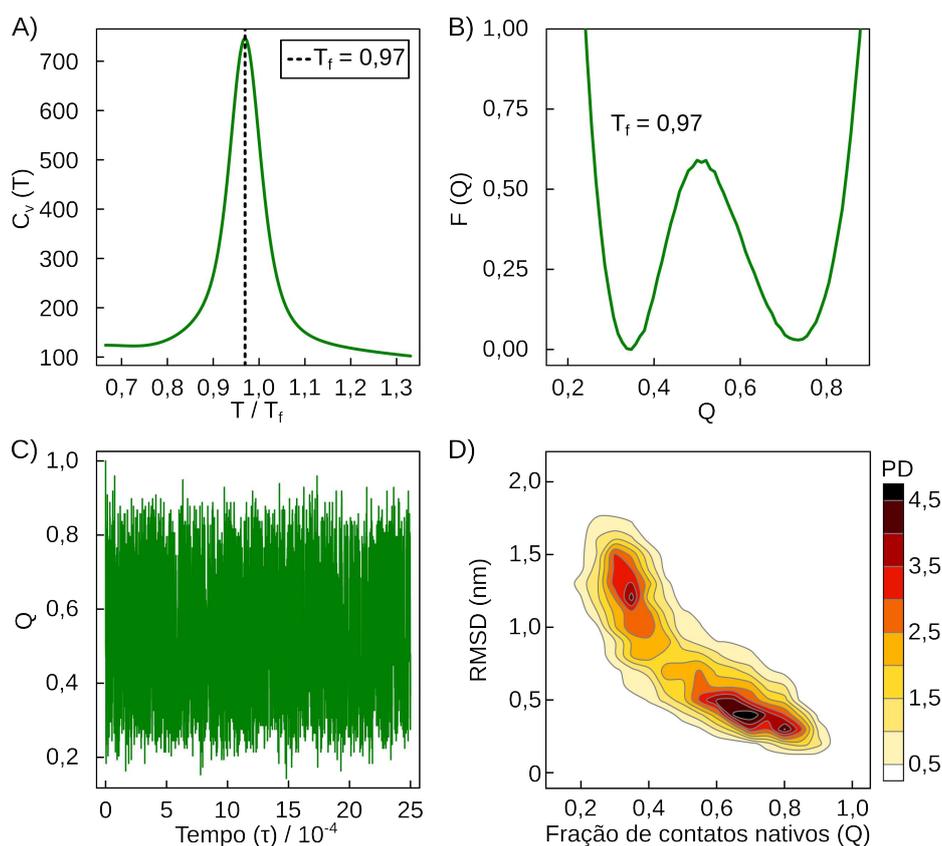


Figura 4.2.1.1: Caracterização do enovelamento da BdpA. A) Calor específico (C_v) em função da temperatura, permitindo a identificação da temperatura de enovelamento ($T_f = 0,97$ unidades reduzidas). B) Energia livre em função da fração de contatos nativos (Q). A partir da simulação realizada em T_f : C) Fração de contatos nativos (Q) em função do tempo da simulação (τ em unidades reduzidas). D) Mapas de contorno da densidade de probabilidade (PD, *Probability Density*) em função de Q e RMSD.

Embora seja possível classificar os estados nativos e desenovelados na Figura 4.2.1.1, a diversidade conformacional de cada estado e dos intermediários está oculta nessas representações. Portanto, utilizamos aqui um método independente de coordenadas, o método ELViM,¹³⁷ para obter uma visualização detalhada do conjunto de enovelamento da proteína.

A Figura 4.2.1.2 mostra uma projeção 2D do espaço de fase obtida com o ELViM,

colorida com a fração de contatos nativos (Q). Cada ponto na figura representa uma estrutura das simulações com os SBMs. A projeção tenta mapear a dissimilaridade estrutural entre as estruturas à distância euclidiana na projeção. Assim, pontos próximos indicam estruturas semelhantes, enquanto pontos distantes indicam estruturas diferentes. Muitas estruturas dissimilares apresentam valores de Q semelhantes, ilustrando a variabilidade conformacional do que é definido como o conjunto desenovelado. O mapa obtido reproduz os mapas previamente calculados.^{123,159} A variabilidade estrutural do conjunto desenovelado está associada à formação das hélices da BdpA, conforme será discutido.

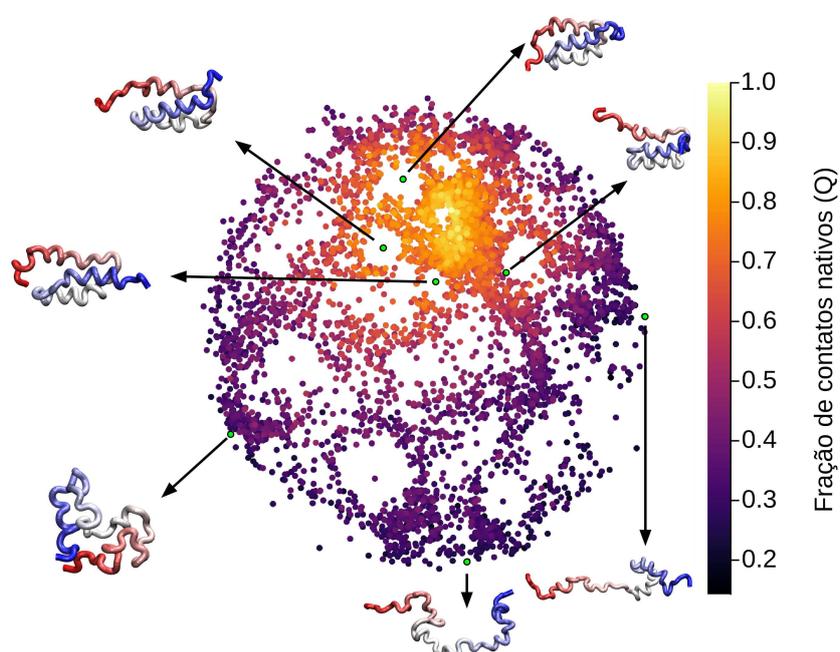


Figura 4.2.1.2: Projeção do conjunto de enovelamento da BdpA obtida com o ELViM. Cada estrutura é representada como um ponto, com uma cor associada à fração de contatos nativos (Q): regiões em amarelo e em roxo representam os estados nativos (Q alto) e desenovelados (Q baixo), respectivamente.

4.2.2 Formação de hélices no conjunto de enovelamento da BdpA

A Figura 4.2.2.1 mostra as projeções do ELViM das estruturas de enovelamento, mas coloridas de acordo com a elipticidade da proteína. A elipticidade total é mostrada na Figura 4.2.2.1A, e, em comparação com a Figura 4.2.1.2, fica claro que as regiões de maior conteúdo de α -hélice são aquelas associadas a uma maior fração de contatos nativos, ou seja, ao conjunto enovelado. As Figuras 4.2.2.1B-D mostram o conteúdo de α -hélice da proteína nas regiões correspondentes às

hélices I, II e III, respectivamente, sobre a projeção do ELViM. É possível perceber nas Figuras 4.2.2.1A-C que a formação da hélice I é um pior indicativo do estado enovelado do que a formação das hélices II e III. Assim, os mapas sugerem que a hélice I é menos estável (menos populada nos estados nativos) do que as hélices II e III, em concordância com dados experimentais.^{121,126,127} A probabilidade da hélice I estar completamente desenovelada (conteúdo de α -hélice de $\leq 0,25$) é de 43%. Por outro lado, as probabilidades de as hélices II e III estarem desenoveladas são menores, 17% e 16%, respectivamente. Da mesma forma, simulações anteriores mostram a maior instabilidade da hélice I e sugerem ainda que ela é particularmente instável na ausência de contatos com a hélice II.¹²³

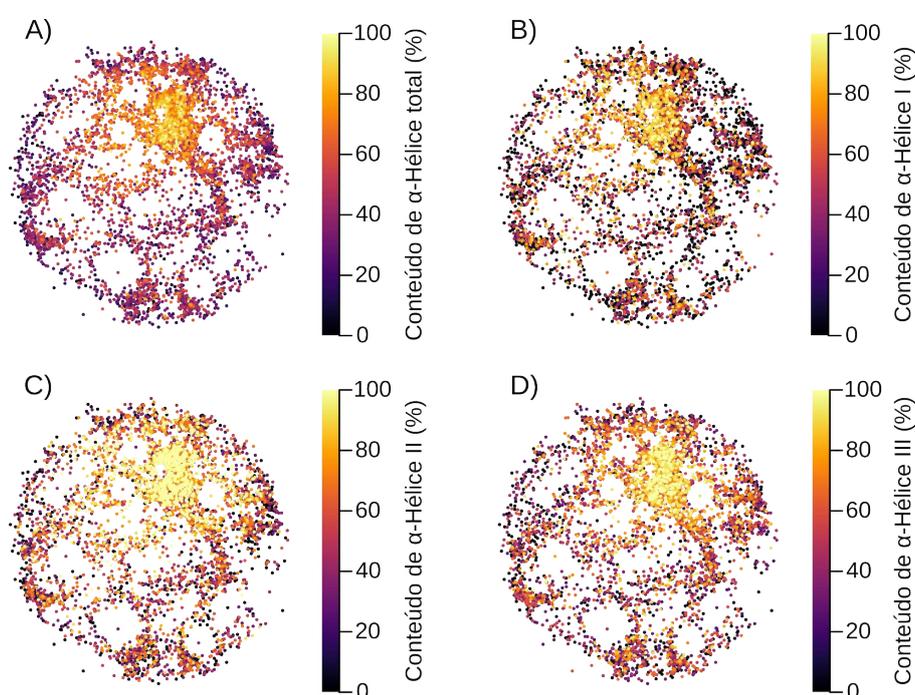


Figura 4.2.2.1: Conteúdo de α -hélice projetado nas estruturas de enovelamento. A cor em cada gráfico é uma função do conteúdo de: A) α -Hélice total da proteína, B) α -Hélice I (¹⁰QQNAFYELH¹⁹), C) α -Hélice II (²⁵EEQRNGFIQSLKD³⁷), D) α -Hélice III (⁴²SANLLAEAKKLNDQA⁵⁶).

Os histogramas de ocorrência das estruturas de enovelamento podem fornecer uma perspectiva sobre a ocupação dos espaços de fase em função do conteúdo de α -hélice da BdpA. As Figuras 4.2.2.2A-D mostram as projeções do ELViM dos conjuntos de enovelamento da proteína BpdA para diferentes extensões da formação de hélices. Os estados desenovelados serão nomeados como U_H^Q , onde o sobrescrito H indica o conteúdo médio de α -hélice do conjunto, e o subscrito Q refere-se a classificação do conjunto em função dos valores decrescentes de Q. Por exemplo, N_{84}^{B0}

(Figura 4.2.2.2A) representa o conjunto nativo de BdpA, que exibe, em média, 84% de α -hélice do modelo de RMN.¹³⁶ Alternativamente, U_{37}^{B11} , na Figura 4.2.2.2C, é um conjunto desenovelado da BdpA com 37% de sua estrutura secundária nativa, sendo o 11º conjunto em ordem decrescente de contatos nativos.

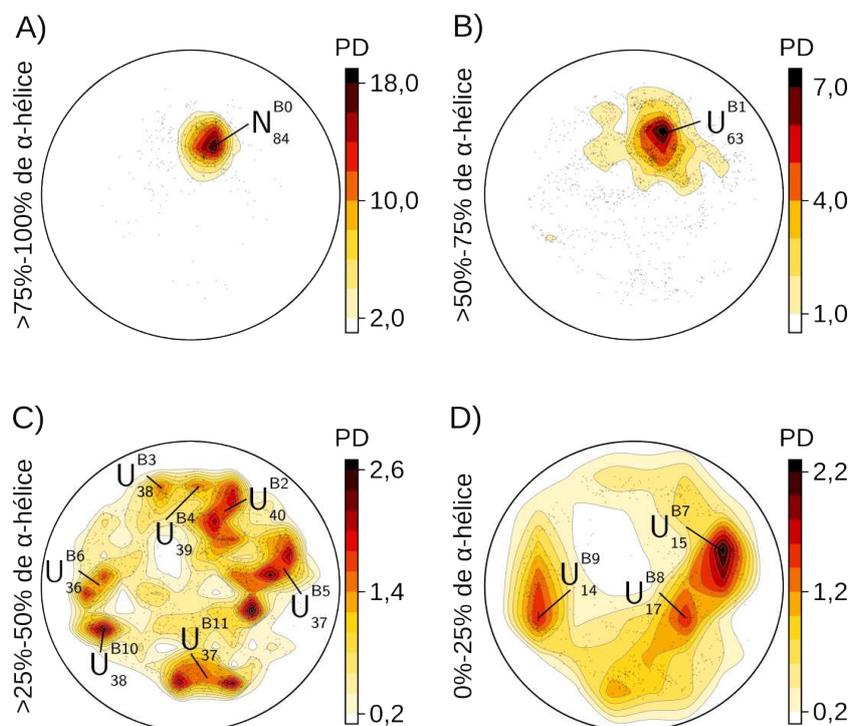


Figura 4.2.2.2: Mapas de contorno da PD das estruturas da BdpA com A) >75%-100%, B) >50%-75%, C) >25%-50% e D) 0%-25% de α -hélice. O conteúdo de α -hélice total é calculado a partir da soma das três hélices da BdpA.

A Figura 4.2.2.2D mostra pelo menos três regiões distintas (U_{14}^{B9} , U_{17}^{B8} , and U_{15}^{B7}) no espaço de fase onde a BdpA desenovelada é mais provável de ser encontrada com uma elipticidade baixa (<25%). Essas regiões representam estados desenovelados distintos. Na Figura 4.2.2.2C, vários conjuntos dissimilares de estruturas com densidade significativa têm conteúdo de α -hélice entre 25%-50%. Segundo Garcia e Onuchic (2003),¹²³ o estado de transição da BdpA contém pelo menos 40% de α -hélice, estando, portanto, presente no conjunto de estruturas da Figura 4.2.2.2C. Por outro lado, muitas estruturas com conteúdo de α -hélice maior que 50-75% (Figura 4.2.2.2B) já apresentam características de estruturas nativas, pois ocupam quase exclusivamente a região do estado nativo (região N_{84}^{B0} , Figura 4.2.2.2A).

Na Figura 4.2.2.3 avaliamos o coeficiente de correlação de postos de Spearman entre

cada hélice durante o enovelamento da BdpA. O coeficiente de correlação de postos de Spearman é particularmente útil aqui, dado que as correlações entre os conteúdos de α -hélices, ou entre as hélices e a proteína completa, são aparentemente não lineares (dados não mostrados). A Figura 4.2.2.3 mostra os histogramas de ocorrência de cada hélice em função do conteúdo de α -hélice de cada uma das outras hélices, a fim de ilustrar as possíveis correlações entre sua formação em todo o *ensemble* de enovelamento. As Figuras 4.2.2.3A e 4.2.2.3B mostram que há uma baixa correlação entre as hélices I e II, e entre as hélices I e III: as hélices II e III podem apresentar uma ampla variedade de estados de estruturação enquanto a hélice I está desenovelada. Por outro lado, na Figura 4.2.2.3C, vemos que os estados mais prováveis envolvendo as hélices II e III são aqueles em que ambas têm um baixo conteúdo de α -hélice ($\leq 30\%$) ou estão estruturadas (conteúdo de α -hélice $\geq 70\%$). Os coeficientes de correlação para a formação das hélices I e II (Figura 4.2.2.3A), I e III (Figura 4.2.2.3B), e II e III (Figura 4.2.2.3C) são 0.1539, 0.1259 e 0.2561, confirmando que as hélices II e III estão de certa forma correlacionadas, ao menos em um grau maior que os outros pares de hélices.

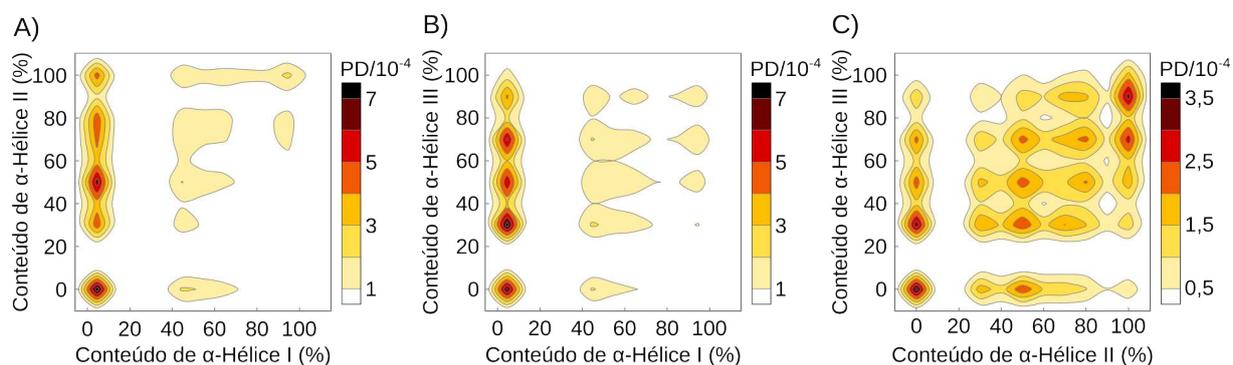


Figura 4.2.2.3: Mapas de contorno da PD dos estados em função dos conteúdos de A) α -Hélice I e α -Hélice II, B) α -Hélice I e α -Hélice III, e C) α -Hélice II e α -Hélice III da BdpA.

4.2.3 Hidratação em estados enovelados e desenovelados

Como ilustrado na Figura 4.2.2.2, obtivemos uma classificação de vários estados desenovelados da BdpA, em função de suas distâncias na projeção 2D e de seu conteúdo de estrutura secundária. Para cada estrutura reconstruída com todos os átomos, de cada conjunto representado na Figura 4.2.2.2, realizamos simulações em água para fornecer uma caracterização detalhada de como o estado de enovelamento da proteína afeta sua estrutura de solvatação.

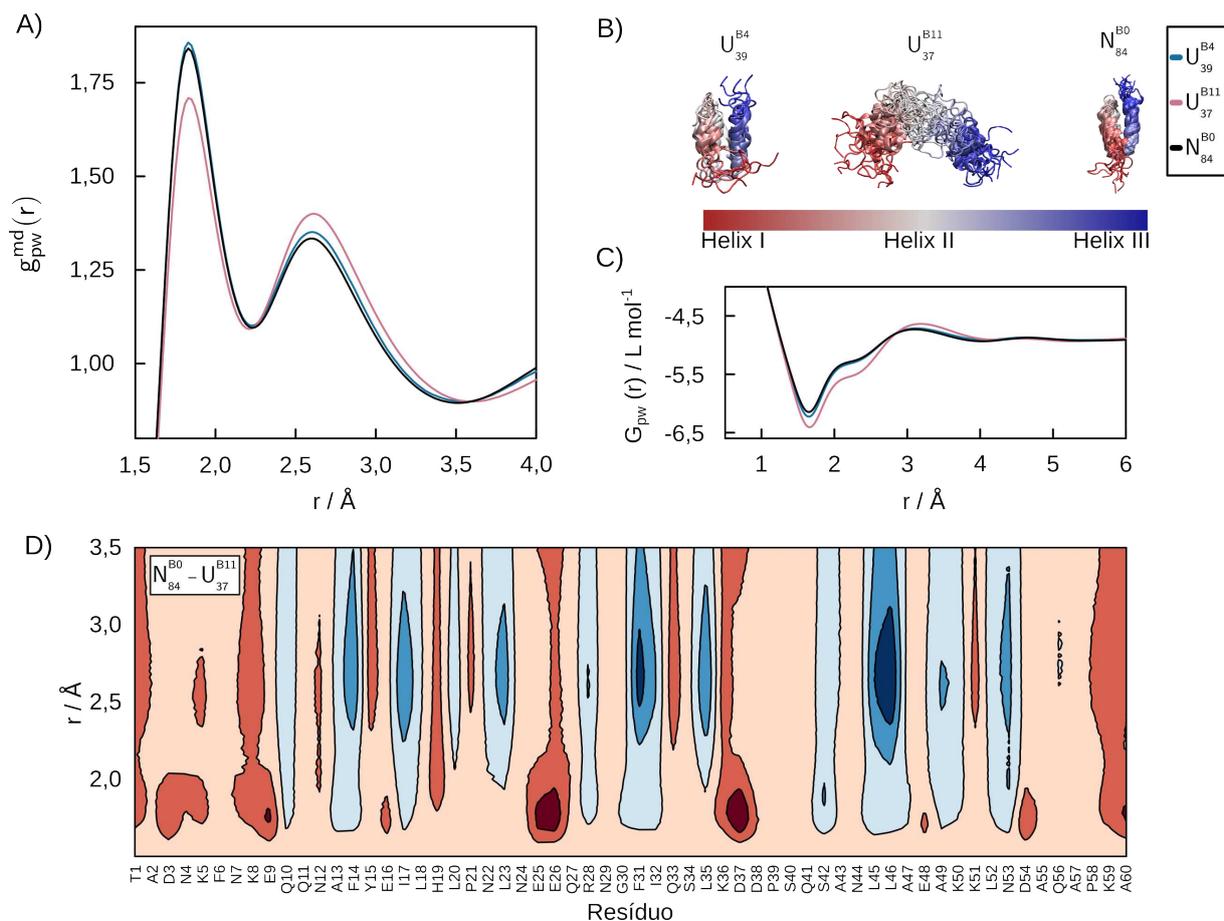


Figura 4.2.3.1: A) MDDFs da água para os subconjuntos de estruturas. B) Conjuntos de estruturas obtidas a partir da Figura 4.2.2.2. C) IKBs correspondentes para água. D) Diferença na densidade das MDDFs em torno dos estados nativo (N_{84}^{B0}) e desenovelado (U_{37}^{B11}). As cores em vermelho indicam uma maior densidade da água em torno do estado nativo (N_{84}^{B0}), enquanto que as cores em azul estão associadas a maiores densidades de água em torno do estado desenovelado (U_{37}^{B11}). O último caso destaca as interações do solvente com resíduos predominantemente hidrofóbicos, que estão protegidos do solvente no estado nativo.

A Figura 4.2.3.1A mostra as MDDFs médias da água para conjuntos selecionados de estados enovelados e desenovelados. Aqui, escolhemos ilustrar a solvatação de três conjuntos: o estado nativo (N_{84}^{B0}) e dois estados desenovelados (U_{39}^{B4} e U_{37}^{B11}) que apresentaram os picos de ligação de hidrogênio máximos e mínimos entre os conjuntos classificados na Figura 4.2.2.2. Na Figura 4.2.3.1B, a sobreposição das estruturas desses conjuntos é mostrada. O conjunto U_{39}^{B4} tem um arranjo terciário que se assemelha ao do estado enovelado, mas com hélices parcialmente formadas, enquanto as estruturas do estado U_{37}^{B11} estão desenoveladas e quase completamente estendidas.

De maneira geral, a densidade de moléculas de água aumenta a curtas distâncias e

diminui na segunda camada de solvatação à medida que a proteína se enovela. Os estados nativo N_{84}^{B0} e o desenovelado U_{39}^{B4} exibem funções de distribuição (Figura 4.2.3.1A - preto e azul) e IKBs (Figura 4.2.3.1C) semelhantes. Isso ilustra que os estados desenovelados podem interagir com o solvente de maneira semelhante ao estado nativo, se as conformações expuserem área superficial e resíduos semelhantes. Por outro lado, o conjunto U_{37}^{B11} exibe um pico de ligação de hidrogênio mais curto (em $\sim 1,9$ Å) e um pico de hidratação maior associado as interações não específicas (em $\sim 2,7$ Å).

A diminuição inicial nos valores das IKBs (Figura 4.2.3.1C) está associada aos volumes de exclusão da proteína. O conjunto de estruturas nativas é o mais compacto (diminuição inicial mais rasa). O conjunto U_{37}^{B11} tem o mínimo mais profundo, implicando em um maior volume de exclusão. No entanto, o aumento da densidade de água em torno do segundo pico compensa a exclusão inicial, e as IKBs convergem para praticamente o mesmo valor, o que implica que, neste caso, os volumes molares aparentes da proteína são semelhantes, independentemente do estado de enovelamento.

Na Figura 4.2.3.1D, tem-se as variações na densidade de moléculas de água na superfície de resíduos individuais tanto no estado nativo quanto no estado desenovelado, pelo cálculo da contribuição de cada resíduo para as MDDFs. O aspecto intrigante dessa análise é que ela mostra como as moléculas de água interagem com os resíduos da proteína, dependendo dos estados conformacionais e, conseqüentemente, da exposição ao solvente. Regiões marcadas em vermelho denotam maior densidade de água no estado nativo, enquanto aquelas em azul indicam maior densidade de água no estado desenovelado. Notavelmente, dentro da primeira camada de solvatação, a densidade de água é significativamente maior no estado nativo (N_{84}^{B0}) em comparação com o estado desenovelado (U_{37}^{B11}), especialmente em regiões próximas aos resíduos carregados (D3, K5, K8, E9, E25, E26, K36, D37, D38). Na segunda camada, a densidade é geralmente maior no estado desenovelado, particularmente ao redor de resíduos hidrofóbicos (F14, I17, L23, F31, I32, L35, L45, L46), que são expostos quando a proteína sofre desnaturação. A abordagem apresentada, portanto, pode identificar quais resíduos são expostos ou protegidos do solvente durante a desnaturação da proteína, de uma maneira dependente da camada de solvatação.

4.3 Conclusões

Neste trabalho, combinamos simulações $C\alpha$ -SBM e atômicas para fornecer *insights* sobre o equilíbrio das hélices da BdpA nos conjuntos enovelado e desenovelado. Simulações com os $C\alpha$ -SBM e o método ELViM permitiram obter uma visão detalhada dos conjuntos de enovelamento da BdpA, com múltiplos estados enovelados e, particularmente, estados desenovelados. Além disso, identificamos muitos conjuntos de estruturas dissimilares com conteúdos semelhantes de α -hélice nas regiões de maior densidade de probabilidade nos espaços de fase 2D da BdpA. A formação das hélices individuais foi investigada a partir dos resultados das simulações atômicas, que forneceram estruturas equilibradas cruciais para determinar o conteúdo de estrutura secundária.

A Hélice I é a mais instável das hélices e é responsável pela heterogeneidade do estado nativo da BdpA, enquanto a Hélice II é a mais estável. Observamos um enovelamento gradual das hélices em direção à estrutura nativa, consistente com o cenário de funil. No entanto, também identificamos que as hélices estão fracamente, mas positivamente correlacionadas, com coeficientes de correlação de Spearman iguais a 0,15 (Hélices I e II), 0,12 (Hélices I e III) e 0,26 (Hélices II e III). Esse resultado indica que as hélices se formam e se deformam inúmeras vezes de maneira fracamente correlacionada.

Finalmente, com a caracterização precisa do conjunto estrutural, conseguimos descrever as estruturas de hidratação de cada conjunto de estrutura usando as MDFFs. A estrutura de hidratação muda do estado enovelado para os desenovelados, reduzindo a importância relativa das ligações de hidrogênio e aumentando a densidade de água em distâncias associadas as interações não específicas. O volume molar aparente das estruturas, no entanto, acaba sendo semelhante para todos os estados. Uma análise detalhada da contribuição de cada resíduo para as estruturas de hidratação permitiu identificar os resíduos que são expostos à água durante a desnaturação. Esses resultados abrem caminho para a análise dos efeitos de osmólitos na estrutura e no equilíbrio do enovelamento de proteínas, que será abordado no Capítulo 5.

Capítulo 5

Efeitos de osmólitos no enovelamento de proteínas

Compreender como os osmólitos influenciam o enovelamento de proteínas é fundamental para a biologia e a biotecnologia.¹⁶⁰⁻¹⁶² O TMAO, um osmólito protetor bem conhecido, encontrado em espécies marinhas e nos rins de mamíferos,¹⁶³ é preferencialmente excluído da superfície de proteínas,^{164,165} e aumenta tanto a temperatura de *melting* (T_m) quanto a energia livre de desenovelamento da proteína (ΔG_U^0).^{44,166,167} Por outro lado, a ureia, um desnaturante comum, interage preferencialmente com as proteínas, e reduz tanto T_m quanto ΔG_U^0 .^{167,168}

O efeito dos osmólitos na estabilidade proteica pode ser analisado, em um primeiro nível de aproximação, a partir de suas interações com a proteína no estado nativo: a exclusão do osmólito da superfície da proteína favorece estruturas compactas, normalmente associadas as formas enoveladas.^{169,170} Por outro lado, interações preferenciais entre a proteína e o osmólito promovem a exposição dos resíduos da proteína e a desnaturação.^{171,172} Esse raciocínio é uma simplificação excessiva para proteínas, pois a desnaturação altera a natureza química geral da sua superfície. Assim, as interações observadas no estado nativo não capturam completamente a complexa relação entre cossolventes e o conjunto desenovelado.¹⁷³ Caracterizar o estado desenovelado é difícil, tanto experimental quanto computacionalmente, devido à sua ampla diversidade conformacional.¹⁷¹ Além disso, compreender de forma consistente a estrutura de solvatação e a termodinâmica em uma grande variedade de conformações desnaturadas é um grande desafio.^{173,174}

A obtenção de parâmetros de interação preferencial para múltiplas conformações de proteínas (de forma experimental ou computacional) permite uma análise mais abrangente do impacto dos cossolventes no equilíbrio de enovelamento.^{172,173,175} A diferença na solvatação

preferencial entre dois estados está relacionada à dependência de suas populações relativas em função da concentração do cossolvente.³⁹ Além disso, determinar o efeito absoluto do cossolvente no equilíbrio de enovelamento requer a medição direta das constantes de equilíbrio tanto em solvente puro quanto em soluções com cossolvente, ou a determinação das energias livres de transferência dos estados nativo e desenovelado.¹⁶⁷

A caracterização experimental do equilíbrio de enovelamento é desafiadora, pois as conformações dos conjuntos nativo e desenovelado podem diferir em condições protetoras e desnaturantes, e os experimentos dependem de uma variável reduzida para classificar os conjuntos: a atividade da proteína, o conteúdo de estrutura secundária, um sinal específico de RMN, etc. Do ponto de vista computacional, os desafios são igualmente significativos: a amostragem adequada do enovelamento da proteína, a visualização precisa das superfícies de enovelamento, métodos confiáveis para identificar as camadas de solvatação ao redor do soluto e a obtenção da convergência das propriedades termodinâmicas exigem precisão metodológica.

Neste trabalho, desenvolvemos uma abordagem computacional para investigar os efeitos estruturais e termodinâmicos da ureia e do TMAO ao longo da paisagem energética de duas proteínas modelo com motivos estruturais distintos: o domínio SH3, rico em folhas- β ,¹⁷⁶ e o domínio B da proteína A (BdpA), predominantemente formado por α -hélices.¹³⁶ Combinamos os SBMs⁵² com simulações atomísticas para gerar os conjuntos de enovelamento, e utilizamos as MDDFs²⁴ para caracterizar as estruturas de solvatação, os parâmetros de interação preferencial e as energias livres de transferência dos intermediários.

Nossos resultados demonstram que os efeitos de solvatação variam entre conformações desenoveladas, influenciando a paisagem energética e potencialmente alterando as vias de desenovelamento. Além disso, mostramos que o efeito desnaturante da ureia decorre de seu acúmulo preferencial em locais de interações dispersivas, que são mais prevalentes nos estados desenovelados. A decomposição das MDDFs em contribuições residuais revelou detalhes inéditos sobre a correlação entre exposição ao solvente e intermediários de enovelamento. Os efeitos dos cossolventes nas energias livres de enovelamento estão de acordo com os dados experimentais dentro de $1 \text{ kcal}\cdot\text{mol}^{-1}$, validando os campos de força e os métodos de análise. Esses resultados oferecem uma nova perspectiva sobre a interação entre osmólitos e proteínas, com implicações para o *design* de solventes e biomoléculas.

5.1 Metodologia

5.1.1 Simulações com os Modelos Baseados na Estrutura e análises

Os conjuntos de enovelamento dos domínios SH3 (PDB: 1FMK)¹⁷⁶ e BdpA (PDB: 1BDD)¹³⁶ foram gerados a partir de simulações com os modelos baseados na estrutura dos $C\alpha$ ($C\alpha$ -SBMs, seção 2.2.2). O mapa de contatos das estruturas nativas foi determinado usando o algoritmo *Contact of Structural Units* (CSU),¹³⁹ resultando em 140 e 98 contatos, respectivamente. Os $C\alpha$ -SBMs foram gerados utilizando o servidor web SMOG (<https://smog-server.org/>) e as simulações foram realizadas com o Gromacs 4.6.7.¹⁴¹

O conjunto de enovelamento descrito para BdpA no Capítulo 4¹⁷⁷ foi utilizado neste trabalho. Um protocolo semelhante foi adotado aqui para obter o conjunto de enovelamento da SH3. Para a proteína SH3, as simulações foram conduzidas em temperaturas variando de 110 a 170 unidades, com um incremento de temperatura de 10 unidades. As simulações a temperaturas constantes consistiram em 5×10^8 passos com um intervalo de tempo de 0,0005 unidades reduzidas. Após identificar aproximadamente a temperatura de calor específico máximo, C_v , um novo conjunto de simulações foi realizado com temperaturas variando entre 135 e 139, com incrementos de 1 unidade, para localizar com precisão de ~ 1 unidade a temperatura de enovelamento (T_f). Em unidades de temperatura reduzida, as simulações foram conduzidas entre 0,91 e 1,41 com incrementos de 0,083 e entre 1,12 e 1,15 com incrementos de 0,0083 unidades de temperatura para o primeiro e o segundo conjuntos de simulações, respectivamente. A simulação na T_f foi estendida dez vezes para aumentar a amostragem das transições conformacionais.

A dependência da temperatura do calor específico ($C_v(T)$) e o potencial de força média em função da fração de contatos nativos ($F(Q)$) dessas simulações foram analisados com o WHAM,¹⁴² conforme implementado no SMOG2.¹⁴³ Um contato foi classificado como nativo se a distância entre os átomos de $C\alpha$ correspondentes não excedesse 20% do valor observado nos modelos determinados experimentalmente. A convergência estatística dos valores de Q na T_f foi confirmada por média em blocos usando o pacote MolSimToolkit.jl versão 1.3.4.

5.1.2 Espaço de fase do enovelamento de proteínas

O método ELViM (seção 4.1.2) foi utilizado para representar o espaço de fases do enovelamento das proteínas SH3 e BdpA. As estruturas secundárias foram determinadas pelo método DSSP,^{102,178} utilizando o pacote ProteinSecondaryStructures.jl.

5.1.3 Simulações atomísticas e análises

Um total de 5.000 modelos SBM para cada proteína, extraídos das simulações realizadas na temperatura de enovelamento (T_f), foram reconstruídas como modelo de todos os átomos utilizando o software Pulchra.¹⁵¹ Em seguida, estudamos a solvatação de cada estrutura reconstruída em soluções de ureia 0,5 mol L⁻¹ e TMAO 0,5 mol L⁻¹.

Caixas de simulação ortorrômicas foram construídas com uma distância mínima de 12,0 Å a partir das extremidades da proteína, utilizando o software Packmol.^{96,97} Os detalhes de cada sistema, incluindo o volume da caixa e o número de moléculas de solvente, estão apresentados nas Tabelas Suplementares I.0.0.1-I.0.0.4 do Apêndice I. Cada conformação foi restringida pela aplicação de potenciais harmônicos com constantes de força de 25 kcal·mol⁻¹ nos átomos de C α , de modo que a topologia C α -SBM fosse preservada, permitindo a relaxação dos átomos reconstruídos e do solvente. O campo de força CHARMM36¹⁵² foi utilizado para a proteína, juntamente com o modelo de água TIP3P.¹⁵³ As moléculas de ureia foram modeladas pela combinação do CGenFF^{179,180} e das cargas do campo de força KBFF, desenvolvido para reproduzir as IKBs das moléculas de ureia e água.¹⁸¹ O TMAO foi descrito pelo campo de força de Netz,¹⁸² com a correção proposta por Shea e colaboradores,¹⁸³ projetada para reproduzir com precisão as IKBs do TMAO em soluções ternárias contendo água, TMAO e ribonuclease T1 (RNase T1).¹⁶⁷

Todas as simulações atomísticas foram realizadas no GROMACS 2021.21⁵⁰ a 298,15 K, 1 atm, e com um passo de integração de 2 fs. Inicialmente, os sistemas foram minimizados por até 20.000 passos do método *Steepest Descent* e equilibrados por 1 ns no *ensemble* de volume e temperatura constantes (NVT), seguido de 1 ns de simulação no *ensemble* de volume e pressão constantes (NPT). A temperatura e a pressão foram controladas pelo termostato de Berendsen modificado¹⁵⁵ e pelo barostato de Parrinello-Rahman.⁹¹ Finalmente, foram realizadas simulações de produção de 10 ns no *ensemble* NPT para cada sistema, totalizando 240 μ s de simulação, utilizadas para obter as MDDFs, IKBs e os parâmetros de interação preferencial dos sistemas.

Simulações em ureia e TMAO a 0,1, 0,2, 0,3, 0,4, 0,5, 0,75 e 1,0 mol L⁻¹, com seis estruturas representativas selecionadas dos conjuntos da SH3 e da BdpA, foram realizadas para determinar as energias livres de transferência durante a transição de enovelamento (N (nativa) \rightleftharpoons U (desenovelada)) das proteínas. Foram realizadas 20 réplicas de 25 ns para cada sistema, totalizando 42 μ s de simulação. As estruturas representativas foram selecionadas utilizando a função *most_representative_structure* do pacote MolSimToolkit.jl (<https://github.com/m3g/MolSimToolkit.jl>), versão 1.23.0. A estrutura mais representativa foi definida como aquela com o menor RMSD em relação à estrutura média de cada conjunto.

5.1.4 Caracterização da estrutura e termodinâmica do solvente

Este estudo emprega um sistema de notação para maior clareza: w (água), p (proteína), c (cossolvente) e s (água ou cossolvente). O pacote ComplexMixtures.jl²³ foi utilizado para caracterizar as estruturas de solvatação a partir das MDDFs e da teoria de solvatação de Kirkwood-Buff (seção 2.1). Especificamente, o pacote fornece as funções de distribuição, as IKBs e as decomposições atômica, de grupo e de resíduos necessárias para uma compreensão molecular da solvatação. Mapas de densidade, úteis para identificar a contribuição de cada resíduo da proteína para a MDDF total, foram gerados com a função *ResidueContributions* do ComplexMixtures.jl.

A partir das IKBs, é possível obter os parâmetros de interação preferencial (Γ). Dado o Γ_{pc} para dois estados, a equação de ligação de Wyman (Equação 2.1.2.10) é usada para determinar a direção em que um cossolvente irá conduzir a reação em uma dada concentração. As energias livres de transferência são calculadas usando a integração numérica da Equação 2.1.3.2, com a Equação 2.1.3.12 como o integrando, a partir de um conjunto finito de simulações dos estados considerados em diferentes concentrações do cossolvente. O cálculo das energias livres de transferência seguiu o protocolo descrito por Lin e Timasheff.¹⁶⁷ A média dos parâmetros de interação preferencial de 20 simulações, calculada em sete concentrações de cossolvente, é usada para determinar $\partial\Delta\mu_{p,tr}^{U-N} = \Delta\mu_{p,tr}^U - \Delta\mu_{p,tr}^N$, ou seja, a diferença entre as energias livres de transferência associadas à transição N \rightleftharpoons U para cada estado desenovelado. Os códigos estão disponíveis no repositório <https://m3g.github.io/PereiraMartinez2024.jl/dev/>. A combinação de MDDFs, Γ , $\Delta\Gamma$ e $\Delta\mu$ (seção 2.1) permite uma caracterização abrangente das estruturas de

solvatação nos conjuntos de enovelamento das proteínas.

5.2 Resultados

5.2.1 Paisagens de enovelamento de proteínas modelo

Entender como cossolventes afetam a superfície de energia livre do enovelamento de uma proteína requer a caracterização da sua variabilidade conformacional. A teoria da paisagem energética^{14,184} fornece um arcabouço para modelar os mecanismos de enovelamento proteico e serve de base para simulações utilizando modelos baseados na estrutura (SBMs). Os SBMs demonstram ser capazes de fornecer representações *coarse-grained* das superfícies de energia do enovelamento proteico e dos conjuntos estruturais de proteínas modelo.^{124,128–130,138,185–189} Neste trabalho, os SBMs foram inicialmente utilizados para amostrar a superfície conformacional da SH3 (um modelo de enovelamento em folhas- β , Figura 5.2.1.1A) e da BdpA (um domínio helicoidal, Figura 5.2.1.1B).¹⁷⁷ Os mecanismos de enovelamento desses modelos foram amplamente caracterizados por métodos experimentais e computacionais, incluindo simulações com os SBMs.^{124,128–130,138,177}

As simulações com os SBMs podem ser realizadas na temperatura de enovelamento do modelo, permitindo múltiplas transições entre os estados enovelado e desenovelado, o que possibilita a reconstrução da superfície de energia livre do enovelamento. Para a SH3, as propriedades gerais da amostragem via SBM são mostradas na Figura 5.2.1.1C. Para a BdpA, a caracterização da superfície de enovelamento é apresentada na Figura 5.2.1.1D, e mais detalhes foram recentemente publicados.¹⁷⁷ Perfis de $C_v(T)$ com picos únicos e bem definidos (Figuras 5.2.1.1C e 5.2.1.1D) indicam mecanismos de enovelamento em dois estados, embora a transição de enovelamento seja mais nítida para SH3 do que para BdpA. Nas Figuras 5.2.1.1C e 5.2.1.1D, painéis *iii*), observa-se a amostragem exaustiva da fração de contatos nativos (Q) ao longo do tempo, indicando a correta amostragem das paisagens conformacionais.

A visualização da superfície de enovelamento requer técnicas de redução de dimensionalidade. Assim, para cada simulação SBM, as conformações amostradas foram projetadas em um espaço bidimensional utilizando o ELViM.^{124,149} Nos painéis i) das Figuras 5.2.1.1C e 5.2.1.1D, cada ponto representa uma conformação obtida na simulação SBM, e as distâncias entre os pontos na projeção são otimizadas para correlacionar com a dissimilaridade dos

modelos.¹²⁴ As cores dos pontos, do azul ao vermelho, representam conformações completamente desenoveladas ($Q = 0$) até completamente enoveladas ($Q = 1$).

Nas Figuras 5.2.1.1C e 5.2.1.1F, rótulos foram atribuídos aos conjuntos de conformações: N e U representam a bacia nativa e os subconjuntos desenovelados, respectivamente, com um subscrito indicando o conteúdo médio de estrutura secundária do conjunto e um sobrescrito representando por S (SH3) ou B (BdpA), seguido da classificação do conjunto em função dos valores decrescentes de Q . Essa nomenclatura será utilizada ao longo do texto para referenciar os subconjuntos da superfície de enovelamento dos modelos. Caracterizações estruturais detalhadas de cada conjunto conformacional são apresentadas na Figura Suplementar I.0.0.1 e na Tabela I.0.0.5.

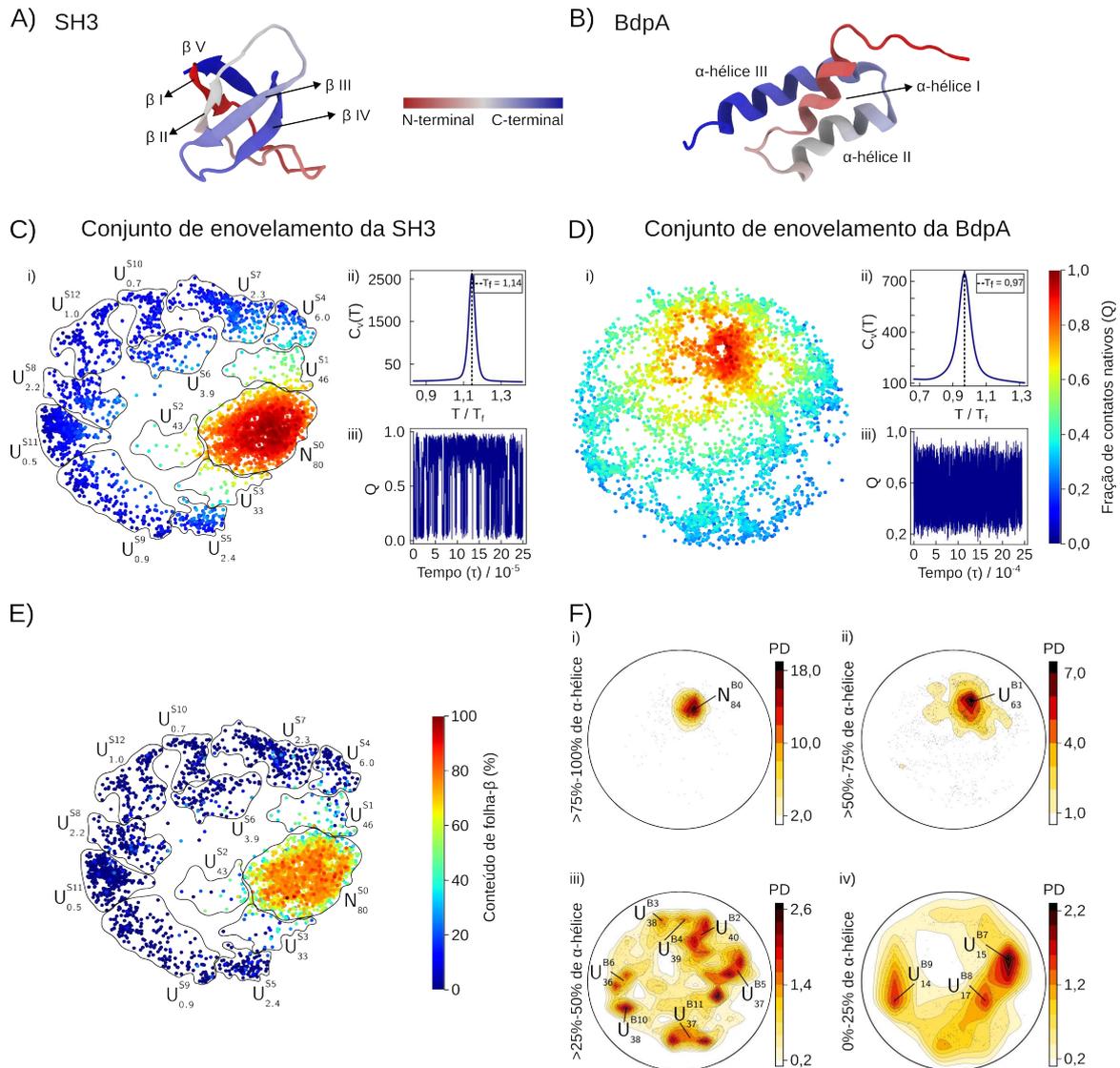


Figura 5.2.1.1: Conjunto de enovelamento da SH3 e da BpdA. A) e B) Representação em cartoon do domínio SH3 (PDB: 1FMK,¹⁷⁶ resíduos de T84 à S140) e do domínio B da proteína A (BdpA, PDB: 1BDD).¹³⁶ C) e D) Conjuntos de enovelamento SBM caracterizados por: i) a projeção ELViM, ii) o calor específico (C_v) em função da temperatura, e iii) a fração de contatos nativos (Q) em função do tempo, em unidades reduzidas (τ). Cada ponto em i) representa uma estrutura colorida em uma escala de Q, variando de 0 (estado desnaturado) a 1 (estado enovelado). E) Projeção do enovelamento da SH3 colorida pelo conteúdo de estrutura secundária e definição dos subconjuntos conformacionais. Há uma alta correlação entre Q (painel C) e o conteúdo de estrutura secundária. F) Dimensão da estrutura secundária na projeção 2D para BpdA, necessária para classificar os estados desnaturados.¹⁷⁷ Nas figuras E) e F), os rótulos são atribuídos aos conjuntos dos estados nativos (N) e desenovelados (U), com um subscrito indicando o conteúdo médio de estrutura secundária e um sobrescrito indicando a ordem decrescente da fração média de contatos nativos.

As Figuras 5.2.1.1C e 5.2.1.1D mostram que os conjuntos enovelados (em vermelho) estão agrupados, como esperado, devido à baixa variabilidade conformacional do estado nativo.

Por outro lado, os estados desenovelados estão dispersos na projeção, refletindo sua diversidade estrutural. A variabilidade do enovelamento da SH3 é capturada pela projeção em 2D: há uma bacia nativa bem definida (N_{80}^{50}), separada de uma variedade de estados desenovelados ($U_{6.0}^{54} - U_{1.0}^{512}$) por apenas três conjuntos intermediários de conformações (U_{46}^{51} , U_{43}^{52} , e U_{33}^{53}). Além disso, há uma alta correlação entre a fração de contatos nativos (Figura 5.2.1.1C) e o conteúdo de estrutura secundária (Figura 5.2.1.1E).

Em contraste, o espaço de enovelamento da BdpA exibe múltiplos caminhos conectando os estados nativo (N) e desenovelado (U) por meio de estruturas com Q semelhantes (Figura 5.2.1.1D). Além disso, estados parcialmente desenovelados podem ser encontrados na bacia nativa, e o conteúdo de estrutura secundária precisa ser adicionado como uma dimensão adicional para diferenciar corretamente as conformações.¹⁷⁷ Assim, na Figura 5.2.1.1F, as projeções foram divididas em faixas de elipticidade da proteína. Os painéis i), ii), iii) e iv) representam os conjuntos conformacionais dentro das faixas de 75-100%, 50-75%, 25-50% e 0-25% de conteúdo de α -hélice, respectivamente. O estado U_{63}^{B1} , no painel ii) da Figura 5.2.1.1F, é particularmente interessante: trata-se de um estado parcialmente desnaturado, caracterizado pela perda parcial da estrutura secundária da hélice I de BdpA, mas com a maioria dos contatos nativos preservados (Figura I.0.0.1). A desnaturação da hélice I é amplamente reconhecida como fundamental para o equilíbrio de enovelamento de BdpA em água.^{123,177,190} As interações notáveis do estado U_{63}^{B1} e sua desestabilização na presença de ureia serão discutidas em detalhes.

A caracterização dos conjuntos conformacionais da SH3 e BdpA é essencial para o principal objetivo deste estudo: investigar minuciosamente as interações da ureia e do TMAO, bem como seus efeitos termodinâmicos nas paisagens de enovelamento de cada proteína. Para isso, realizamos reconstruções atomísticas de 5000 conformações a partir das simulações SBM, e cada um desses modelos foi simulado em soluções aquosas de ureia ou TMAO a $0,5 \text{ mol L}^{-1}$. Após a equilibração os sistemas solvatados foram simulados por 10 ns, sempre com as coordenadas dos $C\alpha$ restritas. Para cada estado, conforme classificado na Figura 5.2.1.1, isso resultou em tempos de simulação variando de 210 a 34.130 ns (21 a 3413 estruturas em cada subconjunto), a partir dos quais foram obtidas as estruturas médias e a termodinâmica dos solventes.

5.2.2 Efeitos dos osmólitos nos estados enovelados e desenovelados da SH3

Primeiramente, focamos no efeito dos osmólitos ureia e TMAO no conjunto conformacional da enovelamento em folha- β do domínio SH3, que apresenta uma paisagem conformacional mais simples. A ureia é um desnaturante conhecido, enquanto o TMAO é um estabilizador que contrabalança o efeito desnaturante da ureia.¹⁹¹ A caracterização estrutural das interações proteína-solvente foi realizada utilizando MDDFs e a teoria de solvatação de Kirkwood-Buff. O pacote ComplexMixtures.jl²³ foi utilizado para realizar essas análises. A termodinâmica das interações proteína-solvente foi caracterizada por meio da medição de parâmetros de interação preferencial, da relação de ligação de Wyman e das energias livres de transferência (seção 2.1).^{20,44,172}

As MDDFs (seção 2.1.2) são funções de distribuição úteis para caracterizar a estrutura de soluções contendo solutos e solventes de formas complexas,²⁴ especialmente os estados desnaturados de proteínas.¹⁷³ As MDDFs são obtidas calculando as menores distâncias entre átomos do soluto e átomos das moléculas do solvente, o que permite uma interpretação clara da camada de solvatação, independentemente da complexidade das estruturas envolvidas. Demonstramos que uma normalização adequada dessas distribuições possibilita o cálculo prático das IKBs a partir da contagem de distâncias mínimas,²⁴ conectando assim a visão microscópica da solvatação às interações preferenciais macroscópicas e às energias livres de solvatação.

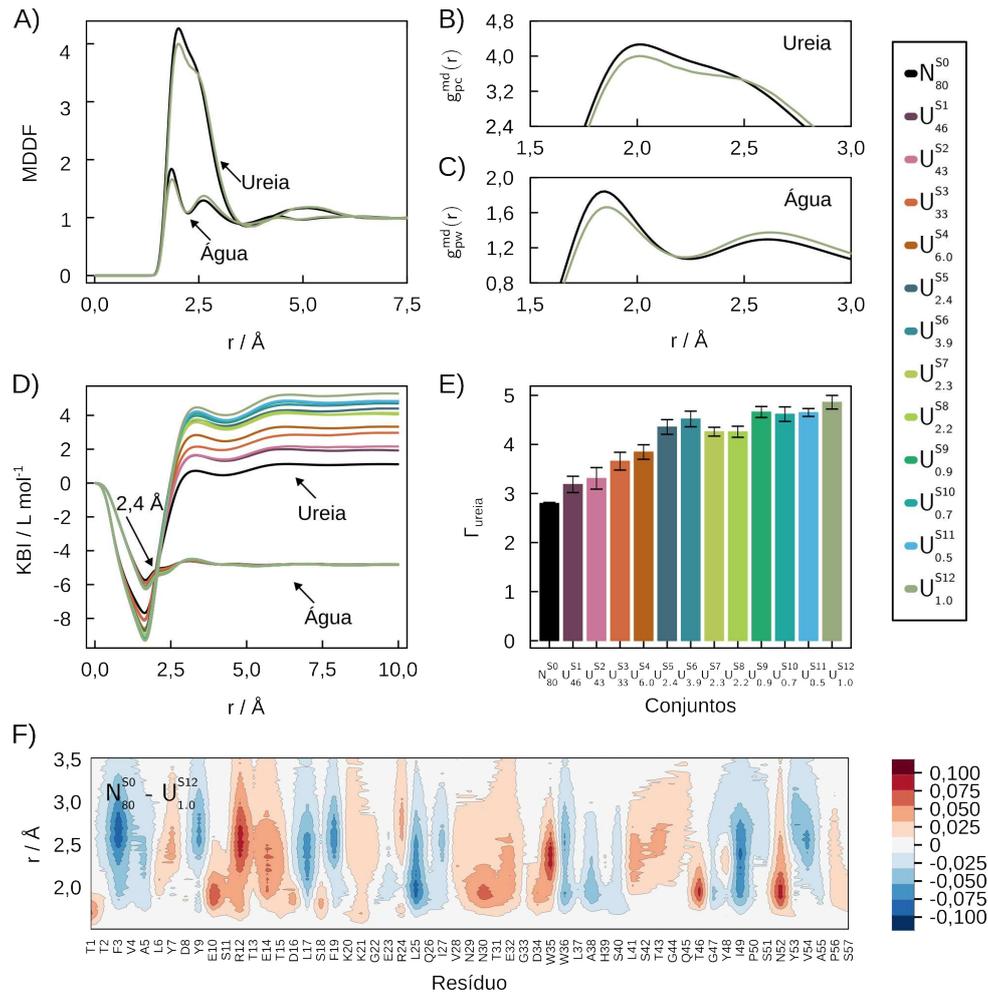


Figura 5.2.2.1: Estruturas de solvatação do conjunto de enovelamento do domínio SH3 em soluções aquosas de ureia 0,5 mol L⁻¹. A) MDDFs da água e da ureia para os conjuntos N₈₀^{S0} e U_{1.0}^{S12}. B) e C) Picos mais proeminentes na MDDF da ureia e da água. D) IKBs e E) Parâmetros de interação preferencial (Γ). A interação preferencial da ureia pelos estados desenovelados é maior do que pelo estado nativo (N₈₀^{S0}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF da ureia na vizinhança do estado N₈₀^{S0} e do estado desenovelado (U_{1.0}^{S12}). Regiões vermelhas indicam maior densidade de ureia próxima ao estado N₈₀^{S0}, enquanto regiões azuis mostram maior densidade ao redor do estado U_{1.0}^{S12}.

As Figuras 5.2.2.1A-C mostram as MDDFs da água e da ureia em relação à proteína nos estados enovelado (N₈₀^{S0}) e desenovelado (U_{1.0}^{S12}), sendo as Figuras 5.2.2.1B e 5.2.2.1C inserções dos picos mais proeminentes dessas distribuições. O estado U_{1.0}^{S12} foi escolhido por apresentar as interações preferenciais mais intensas com a ureia. A Figura 5.2.2.1A revela um aumento claro da densidade de água e ureia nas proximidades da proteína. Tanto a água quanto a ureia formam ligações de hidrogênio com a proteína (evidenciadas pelos picos em $\sim 1,9$ Å). Esses picos de ligações de

hidrogênio são mais intensos no estado nativo. Por outro lado, as MDFFs para o estado desnaturado aumentam a distâncias maiores ($>2,5 \text{ \AA}$), onde predominam interações não específicas. As IKBs (seção 2.1.2), que quantificam o acúmulo global de solvente ao redor da proteína, fornecem uma visão quantitativa da importância das interações em cada distância.

As IKBs da ureia (Figura 5.2.2.1D), são maiores do que os da água, o que implica na desidratação da proteína. De fato, os parâmetros de interação preferencial com a ureia são sempre positivos (Figura 5.2.2.1E). A dependência das IKBs com a distância, além disso, fornece informações sobre as contribuições das interações em cada camada de solvente para a formação das IKBs finais: a queda inicial (Figura 5.2.2.1D) está associada ao volume de exclusão entre a proteína e a ureia. Em distâncias menores que $2,5 \text{ \AA}$, as IKBs dos estados desnaturados tornam-se menores do que as do estado nativo, o que é consistente com a observação de que interações específicas com o estado enovelado são mais fortes, como mostrado nas Figuras 5.2.2.1A e 5.2.2.1B). Em seguida, as IKBs da ureia aumentam rapidamente, superando as da água, e atingem aproximadamente o valor final de convergência em 5 \AA . Portanto, o acúmulo de ureia em distâncias correspondentes às interações não específicas é a característica determinante que leva à desidratação da proteína. Além disso, essas interações dispersivas são as únicas responsáveis pelo aumento da solvatação preferencial por ureia nos estados desnaturados em relação ao estado nativo. Experimentos e simulações que sustentam o fortalecimento dessas interações dispersivas com os estados desnaturados já foram encontrados com ureia¹⁹² e outros desnaturantes.¹⁹³

Os parâmetros de interação preferencial (Γ_{ureia}) aumentam monotonicamente com o grau de desnaturação da proteína. Por exemplo, o estado N_{80}^{S0} (enovelado) apresenta um Γ_{ureia} menor do que os outros estados conformacionais, sendo 1,74 vezes menor do que o maior Γ_{ureia} observado no estado $U_{1.0}^{S12}$. As interações preferenciais da ureia correlacionam-se com a fração média de contatos nativos (Q), o conteúdo de folha- β e a área superficial acessível ao solvente (SASA) (Figura 5.2.2.2).

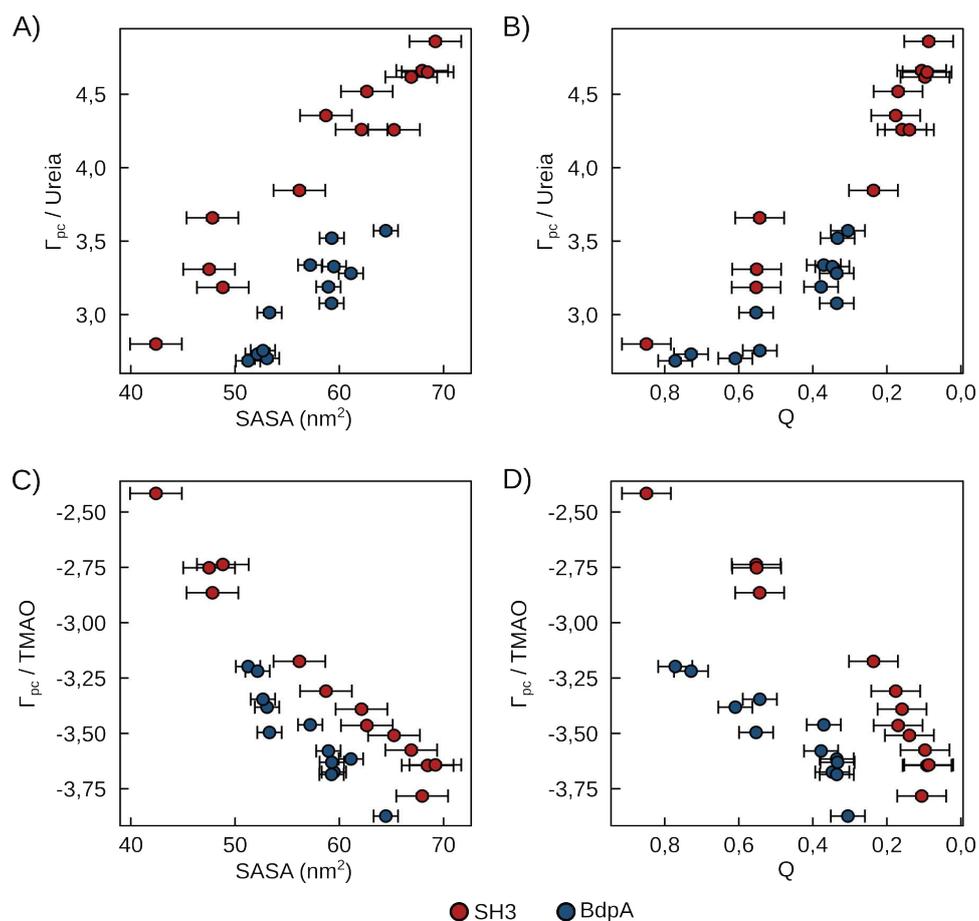


Figura 5.2.2.2: Parâmetros de integração preferencial (Γ) em função de SASA (A e C) e de Q (B e D). Os pontos representados em vermelho e azul correspondem aos parâmetros das proteínas SH3 e BdpA, respectivamente. Um ajuste quadrático captura de forma eficaz a correlação entre Q e Γ (coeficiente de correlação de Pearson, $R = 0,966$), o conteúdo de folhas- β e Γ ($R = 0,947$) e a SASA e Γ ($R = 0,966$) nos *ensembles* do domínio SH3. Como esperado, Q e SASA, assim como Q e o conteúdo de folhas- β , também estão correlacionados ($R = 0,983$ e $R = 0,993$, respectivamente).

Uma visão molecular detalhada do acúmulo de ureia na superfície da SH3 nos diferentes estados conformacionais pode ser obtida pelo mapa da diferença na densidade das MDDFs (Figura 5.2.2.1F). Esse mapa mostra a diferença na densidade local de ureia entre os estados N_{80}^{50} e $U_{1.0}^{512}$ em torno de cada resíduo da proteína. Regiões vermelhas indicam maior densidade de ureia no estado nativo, enquanto regiões azuis indicam maior densidade no estado desenovelado. No estado N_{80}^{50} , as moléculas de ureia se acumulam na primeira camada de solvatação ao redor de resíduos polares (treoninas T1, T31 e T46, serinas S18 e S57, e asparaginas N30 e N52), carregados (glutamato E10, arginina R12, lisina K21 e aspartatos D16 e D34) e nos grupos N- e C-terminais. No estado $U_{1.0}^{512}$, a ureia se acumula a distâncias maiores, preferencialmente ao redor de resíduos hidrofóbicos, como fenilalanina (F3 e F19), tirosina (Y9), leucina (L17 e L25), triptofano (W36), isoleucina (I27

e I49) e valina (V54). Um comportamento semelhante é observado para a água, sugerindo uma competição entre água e ureia por interações específicas e não específicas nos estados enovelado e desenovelado da proteína (dados não mostrados).

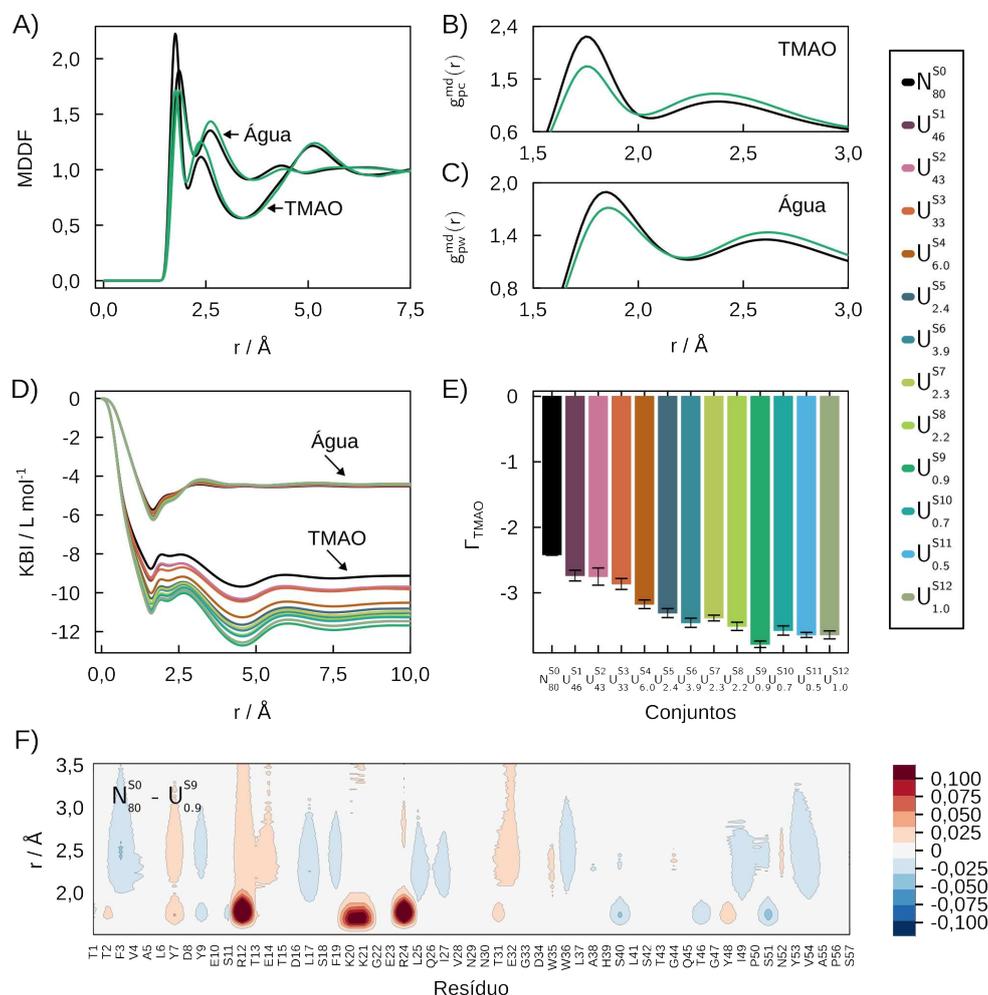


Figura 5.2.2.3: Estruturas de solvatação do conjunto de enovelamento do domínio SH3 em soluções aquosas de TMAO 0,5 mol L⁻¹. A) MDDFs da água e do TMAO para os conjuntos N₈₀^{S0} e U_{0,9}^{S9}. B) e C) Picos mais proeminentes na MDDF do TMAO e da água. D) KBIs e E) Parâmetros de interação preferencial. A exclusão do TMAO nos estados desenovelados é maior do que no estado nativo (N₈₀^{S0}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF do TMAO na vizinhança do estado N₈₀^{S0} e do estado desenovelado (U_{0,9}^{S9}). Regiões vermelhas indicam maior densidade de TMAO próxima ao estado N₈₀^{S0}, enquanto regiões azuis mostram maior densidade ao redor do estado U_{0,9}^{S9}.

A Figura 5.2.2.3 apresenta a análise das estruturas de solvatação da SH3 em soluções aquosas de TMAO. O TMAO é um osmólito estabilizante, uma propriedade associada à exclusão preferencial (ou, equivalentemente, hidratação preferencial).^{167,194} Nas Figuras 5.2.2.3A e 5.2.2.3B,

observamos que tanto o TMAO quanto a água formam interações específicas com a proteína²⁴ (ligações de hidrogênio em $\sim 1,75$ Å), mas os picos do TMAO são muito menores do que os da ureia nessas distâncias (Figuras 5.2.2.3A e 5.2.2.3B). Além disso, há uma depleção significativa centrada em $\sim 3,5$ Å nas MDDFs do TMAO. Ao compararmos as MDDFs do estado nativo (N_{80}^{50} - preto) com o estado desenovelado com maior exclusão de TMAO ($U_{0,9}^{59}$ - verde), observamos que a densidade local de TMAO no estado desenovelado é menor em distâncias de ligação de hidrogênio, mas maior em distâncias associadas a interações não específicas.

A maior contribuição de interações não específicas no estado desenovelado é, assim como ocorre com a ureia, uma característica esperada associada à exposição do núcleo da proteína. As ligações de hidrogênio com a proteína envolvem o átomo de oxigênio do TMAO (Figuras 5.2.2.4A e 5.2.2.4C). A segunda camada da MDDF do TMAO surge devido à sua natureza anfífilica: o TMAO possui três grupos metila, que interagem favoravelmente com as regiões hidrofóbicas dos conjuntos desenovelados (Figura 5.2.2.4). É importante destacar que tanto as interações específicas quanto as não específicas entre o TMAO e a SH3 envolvem principalmente as cadeias laterais da proteína, com interações desprezíveis com a cadeia principal (Figuras 5.2.2.4B e 5.2.2.4D). Isso é consistente com o fato de que as interações do TMAO, assim como de outros osmólitos protetores, com a cadeia principal da proteína favorecem a desnaturação.^{174,195}

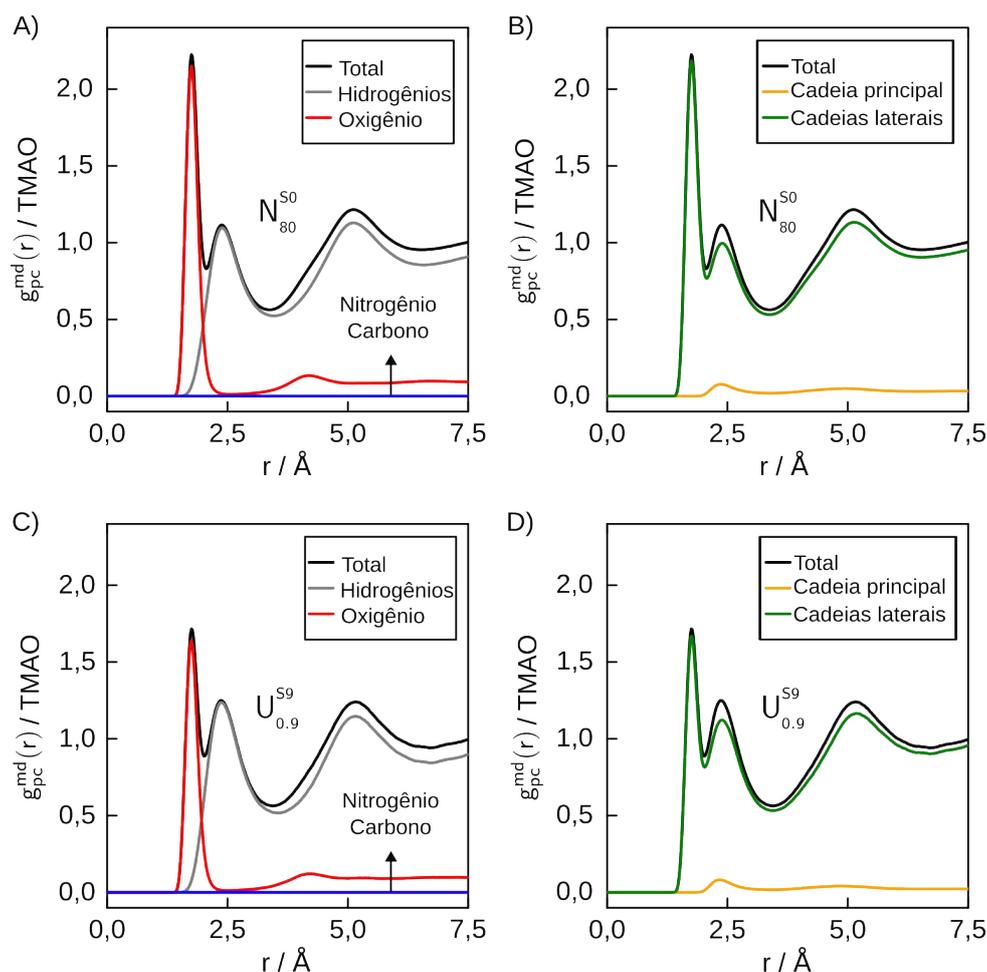


Figura 5.2.2.4: MDDF total do TMAO e contribuições de grupo para os *ensembles* N_{80}^{S0} e $U_{0.9}^{S9}$. A) e C) mostram as contribuições dos átomos e grupos de átomos para a MDDF do TMAO nos *ensembles* N_{80}^{S0} e $U_{0.9}^{S9}$. B) e D) mostram as respectivas contribuições da cadeia principal (amarelo) e da cadeia lateral (verde) para a MDDF total do TMAO.

A Figura 5.2.2.3D exibe as IKBs da água e do TMAO para todos os estados da SH3, em função da distância. As IKBs da água são semelhantes para todos os conjuntos, mas as IKBs do TMAO diminuem à medida que a proteína se desenvela. A exclusão do TMAO em distâncias curtas ($<2,0$ Å) e intermediárias ($\sim 2-5$ Å) essencialmente determina os valores convergentes das IKBs. As IKBs menores do TMAO em relação à água implicam que todos os conjuntos apresentam hidratação preferencial ($\Gamma_{\text{TMAO}} < 0$) (Figura 5.2.2.3E), e a exclusão do TMAO aumenta com a desnaturação. Isso é consistente com o mecanismo em que osmólitos protetores são excluídos da superfície da proteína.^{167,194} Uma característica interessante observada aqui é que os conjuntos mais desenovelados ($U_{6.0}^{S4}$ à $U_{1.0}^{S12}$) apresentam maior exclusão de TMAO, independentemente do caráter mais hidrofóbico do núcleo proteico exposto.

A Figura 5.2.2.3F mostra a diferença na densidade da MDDF por resíduo do TMAO na vizinhança de N_{80}^{50} e $U_{0.9}^{59}$. O vermelho e o azul indicam maior densidade de TMAO ao redor de N_{80}^{50} e $U_{0.9}^{59}$, respectivamente. A densidade de TMAO ao redor de N_{80}^{50} e do estado desenovelado $U_{0.9}^{59}$ é semelhante para a maioria dos resíduos. As interações nas primeiras camadas de solvatação de R12, R24, K20 e K21 são responsáveis pelo maior pico da MDDF em $\sim 1,75$ Å no estado nativo (Figura 5.2.2.3B). O aumento na densidade da segunda camada de solvatação em $U_{0.9}^{59}$ está ligado principalmente a resíduos não polares (F3, L17, L25, I27, W36, I49 e V54). Os mapas diferenciais de densidade da água se assemelham aos observados nas soluções de ureia (dados não mostrados). Assim, a água se acumula ao redor dos mesmos resíduos, independentemente do cossolvente, embora com uma densidade maior nas soluções de TMAO.

5.2.3 Efeitos dos osmólitos nos estados enovelado e desenovelado da BdpA

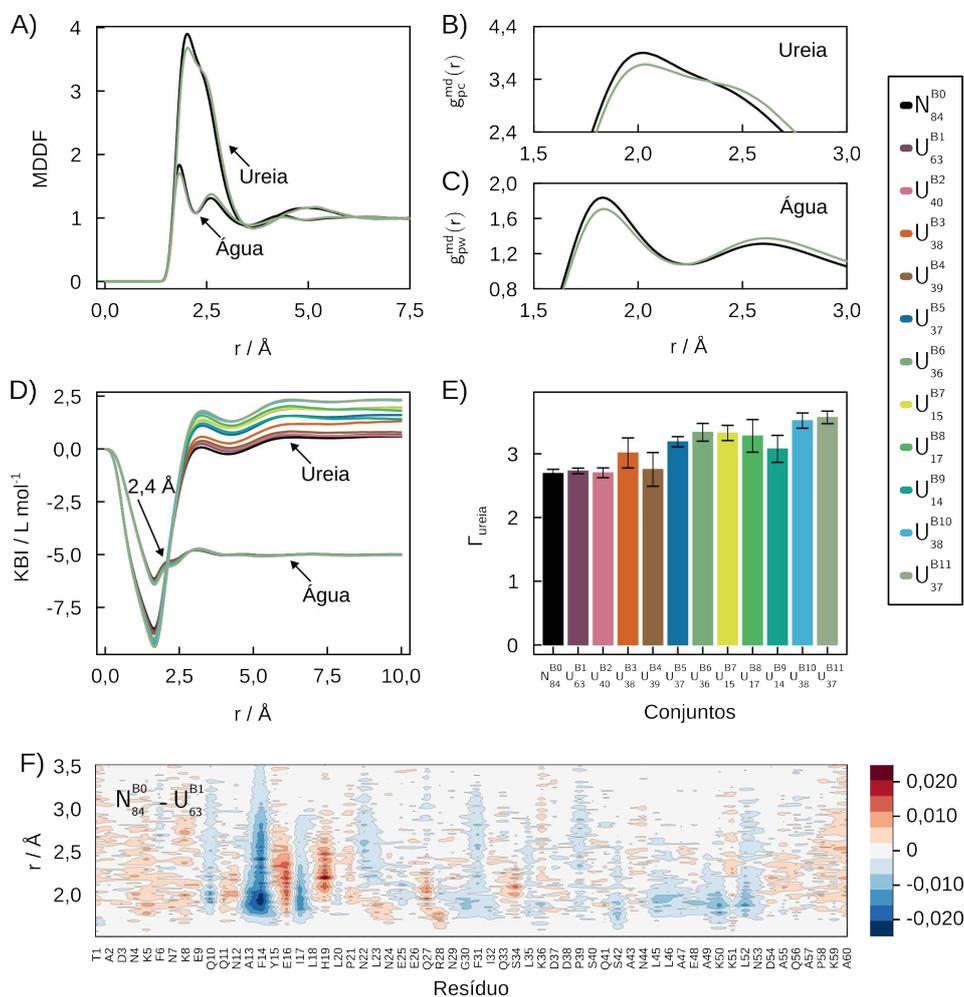


Figura 5.2.3.1: Estruturas de solvatação do conjunto de enovelamento da BdpA em soluções aquosas de ureia $0,5 \text{ mol L}^{-1}$. A) MDDFs de água e ureia para os *ensembles* N_{84}^{B0} e U_{37}^{B11} . B) e C) picos mais proeminentes nas MDDFs de ureia e água. D) IKBs e E) parâmetros de interação preferencial. A interação preferencial da ureia para os estados desenovelados é maior do que para o estado nativo (N_{84}^{B0}) e aumenta entre os estados desenovelados. F) Mapa da diferença na densidade da MDDF da ureia por resíduo de N_{84}^{B0} e U_{63}^{B1} . Vermelho indica maior densidade ao redor de N_{84}^{B0} e azul ao redor de U_{63}^{B1} .

Nesta seção, discutiremos a estrutura de solvatação da BdpA, uma proteína modelo em α -hélice, na presença de ureia e TMAO. Por um lado, as propriedades gerais de solvatação da BdpA são semelhantes às da SH3: nas Figuras 5.2.3.1A e 5.2.3.1B, é possível observar aumentos de densidade da água e da ureia em distâncias curtas da proteína. A Figura 5.2.3.1B mostra que os picos

de ligações de hidrogênio são maiores no estado nativo em relação ao estado desenovelado, mas as interações não específicas são fortalecidas nos estados desenovelados. Perfis qualitativamente semelhantes de MDDF são observados para o TMAO nas Figuras 5.2.3.2A e 5.2.3.2B. A ureia apresenta IKBs maiores do que a água (Figura 5.2.3.1D) e solvata preferencialmente o domínio da BdpA (Figura 5.2.3.1E). O TMAO, por outro lado, é preferencialmente excluído (Figuras 5.2.3.2D e 5.2.3.2E), principalmente devido ao acúmulo em distâncias superiores a $2,4 \text{ \AA}$. A ureia é um melhor solvente para a SH3 do que para a BdpA, e o TMAO é um solvente pior, conforme indicado pelos parâmetros de interação preferencial. Isso é consistente com a observação de que a ureia predominantemente desnatura estruturas em folha- β .^{168,196,197}

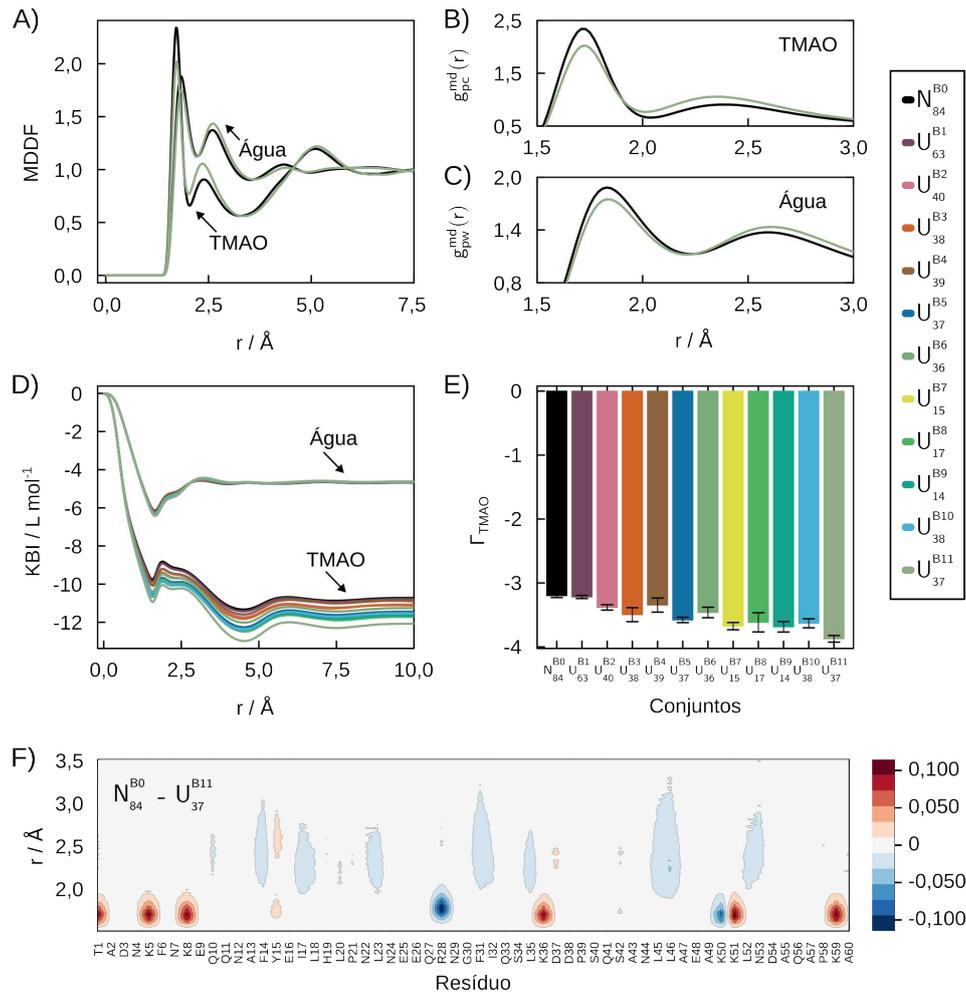


Figura 5.2.3.2: Estruturas de solvatação da BdpA em soluções aquosas de TMAO $0,5 \text{ mol L}^{-1}$. A) MDDFs de água e ureia para os *ensembles* N_{84}^{B0} e U_{37}^{B11} . B) e C) mostram os picos mais proeminentes nas MDDFs do TMAO e da água. D) IKBs e E) parâmetros de interação preferencial. A interação preferencial do TMAO para os estados desenovelados é menor do que para o estado nativo (N_{84}^{B0}) e diminui nos estados desenovelados. F) Mapa da diferença na densidade da MDDF por resíduo de N_{84}^{B0} e U_{37}^{B11} . Vermelho indica maior densidade ao redor de N_{84}^{B0} e azul ao redor de U_{37}^{B11} .

Os quatro primeiros subconjuntos desenovelados da BdpA, U_{63}^{B1} , U_{40}^{B2} , U_{38}^{B3} , e U_{39}^{B4} , exibem interações preferenciais aproximadamente equivalentes às do estado nativo N_{84}^{B0} dentro das flutuações de simulação, notavelmente com a ureia (Figura 5.2.3.1E), mas também, em menor grau, com o TMAO (Figura 5.2.3.2E). O estado U_{63}^{B1} é, efetivamente, apenas parcialmente desenovelado: ele exibe 63% do conteúdo de α -hélice experimental e 72% dos contatos nativos, contra 84% e 77% do conjunto N_{84}^{B0} da bacia enovelada (Tabela Suplementar I.0.0.5). A desnaturação parcial do U_{63}^{B1} é, por essa razão, percebida apenas quando a projeção de enovelamento é dividida em composições de estrutura secundária (Figura 5.2.1.1F). No entanto, U_{63}^{B1} é um conjunto conformacional importante

da BdpA: é caracterizado pela desnaturação parcial da hélice I (Figura Suplementar I.0.0.1). A hélice I é conhecida por ser instável na BdpA, fornecendo variabilidade estrutural ao conjunto enovelado e levando as primeiras etapas da desnaturação. Em contraste, o conjunto U_{39}^{B4} , com 39% de conteúdo helicoidal e 54% de contatos nativos, está muito mais desnaturado. No entanto, sua área superficial acessível ao solvente (SASA), de $52,7 \text{ nm}^2$ (Tabela Suplementar I.0.0.5), é semelhante à do conjunto nativo, de $51,2 \text{ nm}^2$, indicando que manteve um enovelamento globular apesar do rearranjo da estrutura secundária e dos contatos terciários. Esses primeiros quatro conjuntos desnaturados da BdpA têm SASAs dentro de 4% do conjunto nativo, enquanto os outros conjuntos desnaturados (U_{37}^{B5} a U_{37}^{B11}) têm áreas de superfície acessível pelo menos 11% maiores que as de N_{84}^{B0} . Portanto, os valores de SASA determinam os parâmetros de interação preferenciais resultantes com a ureia.

A Figura 5.2.3.1F mostra a diferença do mapa de densidade U_{63}^{B1} e N_{84}^{B0} por resíduo. A exposição de F14 e do resíduo vizinho à ureia destaca a ligação entre a desnaturação da hélice I e a estrutura de solvatação. A exposição pronunciada de F14 à ureia é uma característica comum de todos os conjuntos desnaturados, exceto do conjunto compacto e desnaturado U_{39}^{B4} , onde ele permanece no interior do núcleo hidrofóbico da proteína. As variações na densidade de TMAO ao redor de cada resíduo são menores devido à sua exclusão preferencial. No entanto, nos estados desenovelados, o TMAO interage mais com resíduos hidrofóbicos e menos com resíduos carregados, particularmente as lisinas (Figura 5.2.3.2F).

Em resumo, as interações da BdpA com ureia e TMAO são qualitativamente semelhantes às da SH3, mas dois tipos de *ensembles* podem ser distinguidos: i) estados parcialmente desnaturados com um leve aumento na área superficial interagem com o solvente de maneira semelhante ao estado nativo; ii) estados desnaturados com maior área superficial são desidratados pela ureia e preferencialmente hidratados na presença de TMAO. Esses comportamentos contrastantes contribuem para uma dependência multifacetada do solvente no equilíbrio de desnaturação da BdpA, que será discutido quantitativamente abaixo.

5.2.4 Efeitos nas constantes de equilíbrio de desnaturação induzidos por osmólitos

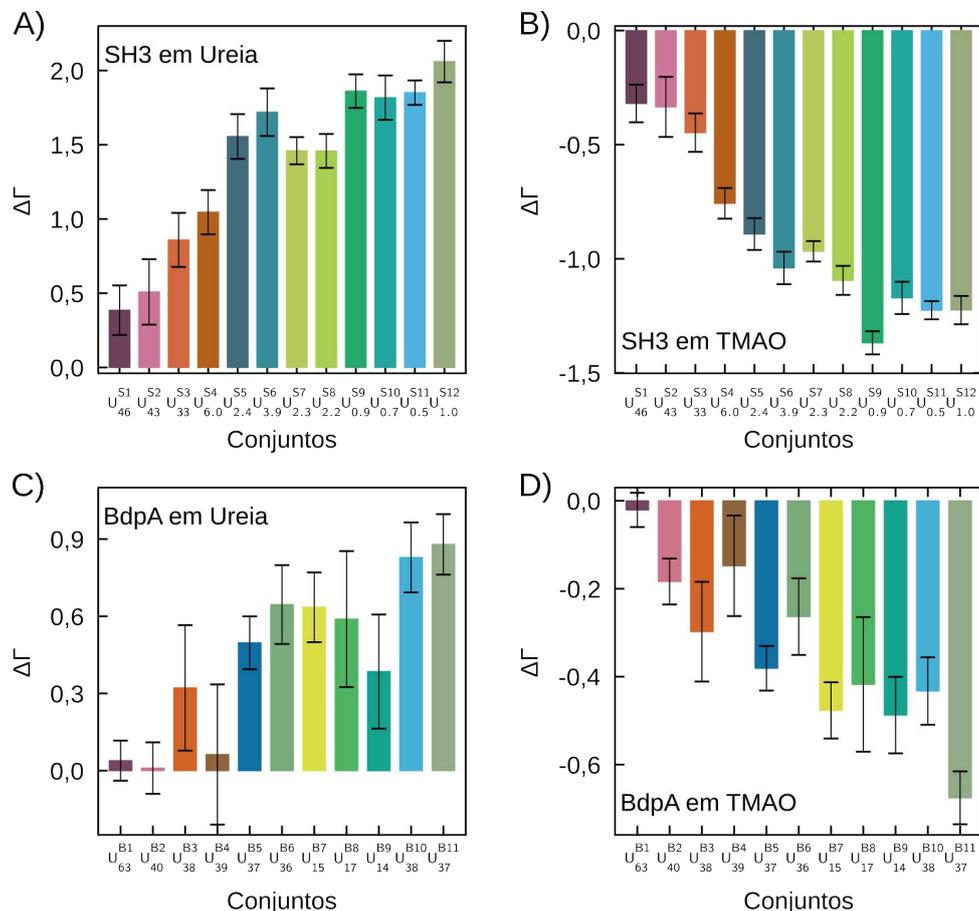


Figura 5.2.4.1: $\Delta\Gamma$ para cada conjunto conformacional de estruturas das proteínas SH3 e BdpA, em relação ao *ensemble* nativo. Os painéis A) e B) mostram os valores de $\Delta\Gamma$ para cada conjunto conformacional da SH3 em ureia e TMAO a $0,5 \text{ mol L}^{-1}$. Da mesma forma, os painéis C) e D) mostram os valores de $\Delta\Gamma$ para cada conjunto conformacional da BdpA nas mesmas condições de cossolvente.

A Figura 5.2.4.1 exibe $\Delta\Gamma$ para SH3 e BdpA em ureia e TMAO a $0,5 \text{ mol L}^{-1}$, considerando o equilíbrio entre os estados nativos e cada um dos estados desenovelados. Os estados desenovelados da SH3 são sempre estabilizados pela adição de ureia (Figura 5.2.4.1A - $\Delta\Gamma$ positivo) e desestabilizados pelo TMAO (Figura 5.2.4.1B - $\Delta\Gamma$ negativo). Para BdpA, $\Delta\Gamma$ também é sempre positivo para a ureia, embora dentro das incertezas do modelo para os quatro primeiros estados parcialmente desenovelados. Por fim, o TMAO favorece o estado nativo de BdpA ($\Delta\Gamma$ negativo) em relação a todos os estados desenovelados, com provável exceção de U₆₃^{B1}.

Em todos os casos, os efeitos dos osmólitos se tornam mais proeminentes à medida que o grau de desnaturação aumenta.

Assim, de modo geral, a ureia favorece os estados desenovelados, enquanto o estado nativo é favorecido pelo TMAO, de forma consistente com os papéis amplamente conhecidos desses osmólitos na estabilidade da proteína. No entanto, essa regra não se aplica a estados desenovelados que apresentam área superficial acessível ao solvente semelhante à do estado nativo. A relação de ligação, Equação 2.1.2.10, não permite, contudo, obter a dependência absoluta da constante de equilíbrio de desnaturação (ou, equivalentemente, da energia livre de desnaturação) em função da concentração do cossolvente. Para determinar o efeito líquido do cossolvente no equilíbrio de enovelamento, é necessário integrar a dependência da energia livre de cada estado com a concentração, desde a água pura até a solução ternária final. Esse cálculo foi realizado para estados selecionados, conforme descrito a seguir.

5.2.5 Dependência das energias livres de desnaturação em relação aos osmólitos para estados desnaturados selecionados

O efeito absoluto de um cossolvente no equilíbrio de enovelamento da proteína pode ser determinado pela diferença entre as constantes de equilíbrio de enovelamento em cada concentração do solvente, em relação à água. Isso corresponde à variação da energia livre de enovelamento com a adição do cossolvente. Experimentalmente, essas variações foram medidas por RMN com diversos osmólitos, incluindo ureia e TMAO, para o domínio Drk-SH3 (*Drosophila signal transduction protein*), que possui o mesmo enovelamento do modelo SH3 estudado aqui.

Alternativamente, se os estados de enovelamento puderem ser estabilizados na escala de tempo do experimento, as energias livres de transferência de cada estado podem ser determinadas, permitindo obter a mesma dependência da energia livre de enovelamento com a concentração de cossolvente. Essas duas abordagens correspondem aos pares de reações do ciclo termodinâmico de Transferência de Tanford, mostrado na Figura 2.1.3.1. Como a modelagem computacional do equilíbrio de enovelamento mesmo para pequenas proteínas, como SH3 e BdpA, é difícil, o efeito dos cossolventes no equilíbrio de enovelamento será estimado por meio do cálculo das energias livres de transferência. Esse cálculo simula o experimento de adição de um cossolvente a uma solução aquosa de uma proteína, preservando aproximadamente seu estado de enovelamento, seja ele enovelado ou

em uma conformação desenovelada.

Nesta análise, focamos em três estados de enovelamento para cada proteína: os estados nativos (N_{80}^{S0} para SH3 e N_{84}^{B0} para BdpA) e dois estados desenovelados. Os estados desenovelados foram escolhidos a partir dos extremos de $\Delta\Gamma$ em ureia ou TMAO a $0,5 \text{ mol L}^{-1}$ (Figura 5.2.4.1). Para SH3, selecionamos U_{46}^{S1} e $U_{1.0}^{S12}$, enquanto para BdpA, U_{63}^{B1} e U_{37}^{B11} foram escolhidos (Figura 5.2.1.1). A Figura 5.2.5.1A mostra que a variação na energia livre de desenovelamento dos estados da SH3 é negativa para todas as concentrações de ureia, indicando que os estados desenovelados são favorecidos. $U_{1.0}^{S12}$ é um estado altamente desnaturado, com grande área superficial, e, de forma consistente, torna-se mais estabilizado pela ureia do que o modelo mais compacto U_{46}^{S1} . Na presença de TMAO (Figura 5.2.5.1B), um $\partial\Delta\mu_{p,tr}$ positivo implica na desestabilização dos estados desenovelados da SH3, consistente com o "efeito osmofóbico".¹⁷⁰ O $\partial\Delta\mu_{p,tr}$ para U_{46}^{S1} é menor do que para $U_{1.0}^{S12}$, pois o TMAO favorece estruturas compactas.

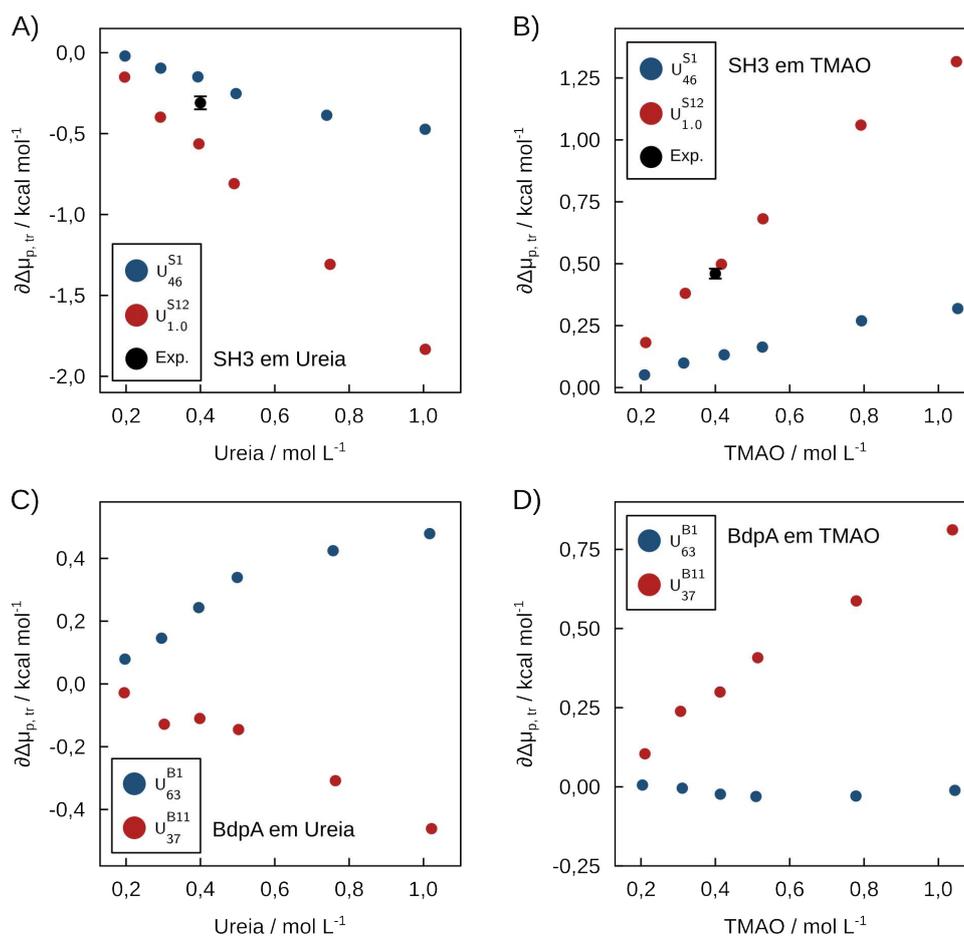


Figura 5.2.5.1: Variação das energias livres de enovelamento ($\partial\Delta\mu_{p,tr}$) em relação à água pura, em função da concentração de cossolvente, para estados desenovelados selecionados. A) e B) mostram $\partial\Delta\mu_{p,tr}$ para os modelos de SH3 em ureia e TMAO, respectivamente. Esta análise focou nos conjuntos desenovelados representativos U_{46}^{S1} (em azul) e $U_{1.0}^{S12}$ (em vermelho). C) e D) mostram $\partial\Delta\mu_{p,tr}$ para BdpA para o equilíbrio de enovelamento associado aos conjuntos U_{63}^{B1} (em azul) e U_{37}^{B11} (em vermelho) de BdpA.

Experimentalmente, foi demonstrado que a energia livre de equilíbrio de enovelamento da Drk-SH3 é afetada pela ureia, na concentração de 0,4 mol L⁻¹, em $-0,31 \pm 0,02$ kcal mol⁻¹. Nesta concentração, as variações calculadas na energia livre de desenovelamento da SH3 após a adição de ureia foram de -0,15 e -0,56 kcal mol⁻¹ para os estados U_{46}^{S1} e $U_{1.0}^{S12}$, respectivamente. De maneira semelhante, à variação experimental determinada na energia livre de enovelamento em soluções de TMAO a 0,4 mol L⁻¹ foi de $0,46 \pm 0,02$ kcal mol⁻¹, e aqui obtivemos 0,13 e 0,50 kcal mol⁻¹. Assim, os valores experimentais estão dentro dos limites inferior e superior das estimativas computacionais para o efeito dos cossolventes no equilíbrio de enovelamento. Esse acordo qualitativo na faixa de energia de sub-kcal é uma validação notável do conjunto dos estados de enovelamento obtido pelas

simulações, dos campos de força utilizados para representar as interações solvente-proteína e dos métodos usados para análise da estrutura do solvente. Os valores experimentais de $\partial\Delta\mu_{p,tr}^{U-N}$ são mostrados como pontos pretos na Figura 5.2.5.1.

Uma análise semelhante dos efeitos dos cossolventes nos equilíbrios de enovelamento dos conjuntos de BdpA, U_{63}^{B1} e U_{37}^{B11} , é apresentada nas Figuras 5.2.5.1C e 5.2.5.1D. U_{63}^{B1} apresenta um $\partial\Delta\mu_{p,tr}^{U-N}$ positivo na ureia e praticamente zero na presença de TMAO. Para o estado U_{37}^{B11} , que apresenta maior grau de desnaturação e área superficial exposta, as energias livres de enovelamento confirmam os efeitos desnaturante da ureia e protetor do TMAO (Figuras 5.2.5.1C e 5.2.5.1D - pontos vermelhos). Na seção anterior, demonstramos que as características de solvatação do conjunto U_{63}^{B1} são semelhantes às do estado nativo de BdpA. Aqui, mostramos que este é um caso particular onde a ureia exerce um efeito desestabilizador sobre um estado parcialmente desenovelado, enquanto em TMAO, o equilíbrio de desenovelamento envolvendo esse estado é essencialmente independente da concentração do cossolvente. Essa observação só poderia ser parcialmente inferida a partir das relações de ligação de Wyman nas Figuras 5.2.4.1C e 5.2.4.1D, considerando o erro padrão dos valores de $\Delta\Gamma$.

Não podemos obter a razão experimental entre as populações de estados similares ao U_{63}^{B1} em relação aos estados nativos a partir de nossas simulações, mas em nosso modelo SBM, U_{63}^{B1} foi duas vezes mais populoso que N_{84}^{B0} . Isso implica uma diferença de energia livre de aproximadamente $-0,4 \text{ kcal mol}^{-1}$ a 300K. O efeito da ureia no equilíbrio de enovelamento, calculado aqui, está na mesma ordem de magnitude, mas favorecendo o estado nativo. Assim, esse efeito pode reduzir significativamente a população da hélice I parcialmente desnaturada, redirecionando a desnaturação por novos caminhos.

5.2.6 Discussão

Neste trabalho, desenvolvemos uma abordagem computacional para caracterizar o efeito de osmólitos nos conjuntos de enovelamento de proteínas. Nossa abordagem combina os modelos baseados na estrutura (SBMs) e simulações atômicas para modelar as paisagens de enovelamento das proteínas SH3 e BdpA, e as interações soluto-solvente de cada conformação. Para elucidar as estruturas de solvatação e quantificar a estabilização promovida por ureia e TMAO, empregamos as MDFFs, que são cruciais para a caracterização da solvatação de estruturas complexas. Usando essas

ferramentas, aplicamos a teoria de Kirkwood-Buff para obter interações preferenciais e a dependência das energias livres de enovelamento em relação aos osmólitos.

Nossos resultados confirmam a interação preferencial da ureia e a exclusão preferencial do TMAO das superfícies das proteínas, ao mesmo tempo em que adicionam uma nova dimensão sobre como essas interações variam ao longo da paisagem de enovelamento da proteína. Diferentes estados desnaturados exibem interações distintas com a água e os cossolventes, correlacionando-se significativamente com a área superficial acessível ao solvente. Isso sugere que estados globulares desnaturados podem ser apenas marginalmente afetados por osmólitos, com a ureia possivelmente promovendo sua desestabilização. As MDDFs revelam que água, ureia e TMAO formam tanto ligações de hidrogênio quanto interações não específicas com a superfície da proteína (Figuras 5.2.2.1A-B e 5.2.2.3A-B). As interações específicas são mais fortes nos estados nativos, enquanto as interações não específicas são mais prevalentes nos estados desnaturados. Esses padrões se mantêm em diferentes composições de solvente, tipos de proteínas e motivos estruturais, refletindo o aumento da exposição de resíduos do núcleo hidrofóbico após a desnaturação.

Interações dispersivas entre ureia e a proteína governam as interações preferenciais, como mostrado pela análise quantitativa das IKBs (Figuras 5.2.2.1D e 5.2.2.3D). À medida que essas interações se fortalecem com o aumento da exposição da superfície da proteína, os estados desnaturados se tornam mais desidratados. Em contraste, tanto as interações específicas quanto as não específicas do TMAO são fracas, levando à hidratação preferencial, que é ainda mais aprimorada em estados desnaturados.

A exposição ao solvente, a fração de contatos nativos e os elementos da estrutura secundária (folha- β e α -hélice) são determinantes chave das interações preferenciais. No caso da SH3, para o qual essas propriedades são altamente correlacionadas, as interações preferenciais e as energias livres de enovelamento variam de forma monótona com o aumento da desnaturação. No caso da BdpA, por outro lado, o conteúdo de estrutura secundária não se correlaciona diretamente com a área superficial acessível ao solvente ou com a fração de contatos nativos. Assim, estados desnaturados como U_{63}^{B1} , U_{40}^{B2} , e U_{39}^{B4} exibem propriedades de solvatação semelhantes ao estado nativo, apesar de terem um menor conteúdo α -hélice.

Simular estruturas representativas de estados de enovelamento selecionados em várias concentrações de TMAO e ureia nos permitiu calcular o efeito absoluto dos osmólitos nas energias

livres de enovelamento. Surpreendentemente, a dependência experimental da energia livre de enovelamento em relação ao TMAO para uma proteína homóloga à SH3 foi encontrada dentro dos limites calculados, validando os campos de força, a estratégia de modelagem e os métodos de análise de solvatação. Ao mesmo tempo, a solvatação de um único estado desnaturado não é suficiente para inferir quantitativamente o papel dos cossolventes no equilíbrio de enovelamento, de forma consistente com a variabilidade do *ensemble* desnaturado desempenhando um papel chave. Para SH3, a dependência da concentração das energias livres de enovelamento é consistente com a desnaturação induzida por ureia e proteção pelo TMAO. Em contraste, BdpA exhibe estados desnaturados que podem ser desestabilizados pela ureia ou não afetados pelo TMAO, ilustrando a complexidade dos efeitos dos osmólitos em proteínas com múltiplos caminhos de desnaturação.

Capítulo 6

Mecanismos de mudança conformacional da proteína MAD2

Proteínas conhecidas como *fold-switching* adotam espontaneamente duas ou mais conformações nativas, que se interconvertem em resposta a estímulos celulares.¹⁹⁸⁻²⁰¹ Cerca de 40% das proteínas identificadas como *fold switchers* passam por mudanças conformacionais reversíveis ao converter α -hélices em folhas- β (ou vice-versa).²⁰²⁻²⁰⁴ Outro subconjunto aumenta o conteúdo de estrutura secundária²⁰⁵ ou forma dímeros.^{206,207} Um grupo menor, aproximadamente ~6%, está envolvido na formação de fibrilas, um processo tipicamente irreversível.²⁰⁸ Quase 100 proteínas no banco de dados de proteínas (PDB, *Protein Data Bank*) foram validadas como exemplos de *fold switchers*, mas estima-se que 0,5-4% das proteínas do PDB pertencem a essa classe.¹⁹⁸

As conformações dessas proteínas geralmente estão associadas a papéis distintos em processos biológicos.^{209,210} Por exemplo, a proteína MAD2 se enovela em dois estados nativos distintos: *Open-MAD2*²¹¹ e *Closed-MAD2*²¹², sendo a primeira inativa e a segunda ativa. Esses dois estados são essenciais durante a divisão celular.²¹⁰ A proteína MAD2 atua na transição da metáfase para a anáfase impedindo a separação dos cromossomos duplicados até que todas as cromátides irmãs estejam corretamente ligadas aos polos opostos dos microtúbulos do fuso mitótico.²¹⁰ A segregação prematura dos cromossomos leva à aneuploidia.^{210,213} As duas conformações compartilham uma subestrutura comum composta por três α -hélices (αA , αB e αC) cercadas por uma β -*hairpin* ($\beta 2$ - $\beta 3$) e três folhas- β ($\beta 4$, $\beta 5$ e $\beta 6$), conforme mostrado nas Figuras 6.2.1.1A e 6.2.1.1B. No entanto, elas diferem estruturalmente nos resíduos das regiões N-terminal (I11 a S16) e C-terminal (K159 a I201). No estado ativo, a MAD2 se liga à proteína Cdc20 por meio da exposição da $\beta 6$. Por outro lado, no estado inativo, essa região de interação é coberta

CAPÍTULO 6. MECANISMOS DE MUDANÇA CONFORMACIONAL DA PROTEÍNA MAD2 117

pela $\beta 8$. Além disso, o estado inativo apresenta uma fita β N-terminal ($\beta 1$), que é rearranjada em uma α -hélice na forma ativa (Figura 6.2.1.1B). O equilíbrio de mudança conformacional da MAD2 ocorre em condições fisiológicas,^{210,212} mas com uma escala de tempo de várias horas.²¹² Essa transição é facilitada pela autodimerização e pela ligação ao alvo Cdc20,²¹⁴ embora esses fatores não sejam necessários para a mudança conformacional.^{212,213}

Experimentos de transferência de energia de fluorescência de molécula única (smFRET) sobre o enovelamento da MAD2 na presença de cloreto de guanidínio (GdmCl), um desnaturante, revelaram a existência de um intermediário de enovelamento.²¹⁵ Esse intermediário se enovela novamente para a forma ativa da MAD2 após a diluição do GdmCl, sugerindo uma semelhança estrutural maior com a forma ativa. Estados intermediários que retêm características de ambos os estados também foram identificados por RMN.²¹⁶ Mais recentemente, estudos de RMN com resolução temporal mostraram que, durante a ativação, a região N-terminal da MAD2, incluindo a fita $\beta 1$, é a última a interconverter da forma inativa para a ativa.²¹⁷ Os autores hipotetizaram que a ativação ocorre inicialmente pelo desenovelamento parcial das fitas $\beta 7$ e $\beta 8$ da C-terminal, expondo a superfície de ligação ao Cdc20.

Neste trabalho, investigamos o mecanismo de enovelamento da proteína MAD2 e a transição entre os estados inativo e ativo usando simulações com os SBMs, coevolução de aminoácidos e análise de frustração estrutural. A MAD2 alterna entre as conformações ativa e inativa enquanto mantém a estabilidade do núcleo. Os nossos dados indicam a presença de um intermediário de ativação no qual as fitas- β C-terminais expõem a região de ligação ao Cdc20, evoluindo para a forma totalmente ativada por meio de um rearranjo conformacional da porção N-terminal da MAD2, de forma consistente com experimentos recentes de RMN.^{216,217} Demonstramos que o rearranjo N-terminal ocorre em um equilíbrio relativamente rápido com o estado ativo, sendo um intermediário crucial para as transições de ativação ou inativação. A análise da coevolução de resíduos apoia a existência de um núcleo conservado e de uma conformação ativa dominante. Por fim, resíduos altamente frustrados estão presentes, especialmente nos segmentos responsáveis pela mudança de conformação e nas regiões de ligação da proteína.

6.1 Metodologia

6.1.1 Modelo baseado na estrutura

A MAD2 existe em equilíbrio dinâmico entre duas formas: inativa (também conhecida como *Open-MAD2*) e ativa (*Closed-MAD2*).²¹⁰ A primeira estrutura determinada, MAD2 inativa (PDB: 1DUJ),²¹¹ corresponde aos resíduos 11-195 (Δ NC10). A MAD2 inativa foi expressa sem a região N-terminal flexível (M1 a G10). Além disso, a região C-terminal também é flexível em solução e foi excluída na determinação da estrutura. No entanto, mesmo na ausência desses doze resíduos, MAD2 inativa mantém sua estrutura terciária.²¹¹ Por outro lado, a conformação alternativa completa (ativa) está disponível no PDB (PDB: 1S2H).²¹² Os dez resíduos N-terminais (M1 a G10) e quatro C-terminais (P202 a D205) desordenados na MAD2 ativa, que não são relevantes para o enovelamento, foram removidos para construir o modelo SBM. Em resumo, o modelo final considerado neste trabalho para MAD2 inativa e ativa foi Δ NC10 (resíduos I11 a S195) e Δ N10C4 (resíduos I11 a I201), respectivamente (Figura 6.2.1.1D). Os diagramas da estrutura secundária da MAD2, mostrados na Figura 6.2.1.1D, foram realizados com o software SSDraw.²¹⁸ As representações em cartoon da conformação ao longo do texto foram preparadas com o programa VMD 1.9.4a55.⁹⁵

6.1.2 Construção do potencial *dual-basin* SBM

O modelo baseado na estrutura, mencionado acima, foi expandido para estudar o enovelamento da proteína MAD2, explicitamente combinando as informações das conformações inativa e ativa em um único Hamiltoniano, chamado *dual-basin* SBM. Primeiro, avaliamos as similaridades estruturais entre as conformações inativa e ativa usando o LovoAlign,^{219,220} um pacote que obtém superposições estruturais que ponderam favoravelmente as regiões estruturalmente conservadas e destacam as dissimilaridades estruturais. Em seguida, a partir dos valores da raiz da flutuação quadrática média (RMSF, *Root Mean Square Fluctuation*) da estrutura inativa alinhada com a estrutura ativa, foi possível identificar elementos estruturais estáveis ($\text{RMSF} \leq 2,5 \text{ \AA}$) e não estáveis ($\text{RMSF} > 2,5 \text{ \AA}$) (Figura 6.1.2.1B).

Para as regiões estruturais estáveis, que correspondem essencialmente à maior parte das regiões em comum nas conformações inativa e ativa (regiões amarelas no cartoon da Figura 6.1.2.1A, painel i), adotamos os parâmetros de ligações, ângulos e diedros do estado ativo, conforme obtido

CAPÍTULO 6. MECANISMOS DE MUDANÇA CONFORMACIONAL DA PROTEÍNA MAD2 119

originalmente pelo SMOG2.^{143,221} Para a região de alta variabilidade (essencialmente as regiões de mudança conformacional e de volta (N- e C-terminais), destacadas em azul no cartoon da Figura 6.1.2.1B, painel ii), os parâmetros de ligações, ângulos e diedros também foram adotados da conformação ativa, mas com algumas modificações: neste caso, as constantes associadas aos ângulos e diedros (ϵ_θ e ϵ_ϕ) foram escaladas por 0,25, para permitir uma melhor amostragem do estado inativo com uma estabilização adequada do estado ativo. As diferenças absolutas nos comprimentos de ligação de equilíbrio entre dois $C\alpha$ dos estados nativos não são superiores a 0,0756 Å. Por outro lado, isso não ocorre com os ângulos e ângulos diedros, que variam nos estados inativo e ativo. No entanto, evitamos combinar ângulos ou diedros, pois isso poderia deslocar o mínimo de energia potencial para a média entre eles nas duas estruturas e desestabilizar ambos os estados nativos.^{54,222}

Os contatos únicos de cada estrutura, envolvendo as regiões de mudança conformacional N- (resíduos I11 a S16) e C-terminal (resíduos K159 a I201), foram combinados em um único mapa de contatos. Os contatos envolvendo os resíduos A17 a D158 foram descartados. Esses contatos resultam de variações estruturais observadas nos modelos de RMN de ambas as conformações, inativa e ativa. Além disso, todos os contatos comuns foram definidos com base nas distâncias de pares na estrutura ativa. Notavelmente, as distâncias entre a maioria dos contatos comuns são semelhantes nas conformações inativa e ativa. Apenas $\sim 3\%$ desses contatos têm distâncias que diferem mais de 25%, o que nos permite manter os contatos compartilhados de apenas uma das estruturas.

A combinação dos parâmetros ligados e não ligados de cada estado nativo em um único Hamiltoniano não garantiu a observação de ambos os estados. Isso ocorre devido às diferentes estabilidades das conformações inativa e ativa. Para facilitar uma transição reversível entre os estados nativos da MAD2, ajustamos o parâmetro (ϵ_C) do potencial de contato para o estado ativo em um intervalo de 0,8 a 0,9. Notavelmente, quando $\epsilon_C = 0,82$, as probabilidades das conformações inativa e ativa são quase iguais. Portanto, utilizamos simulações a partir deste potencial para discutir os resultados apresentados na seção 6.2.

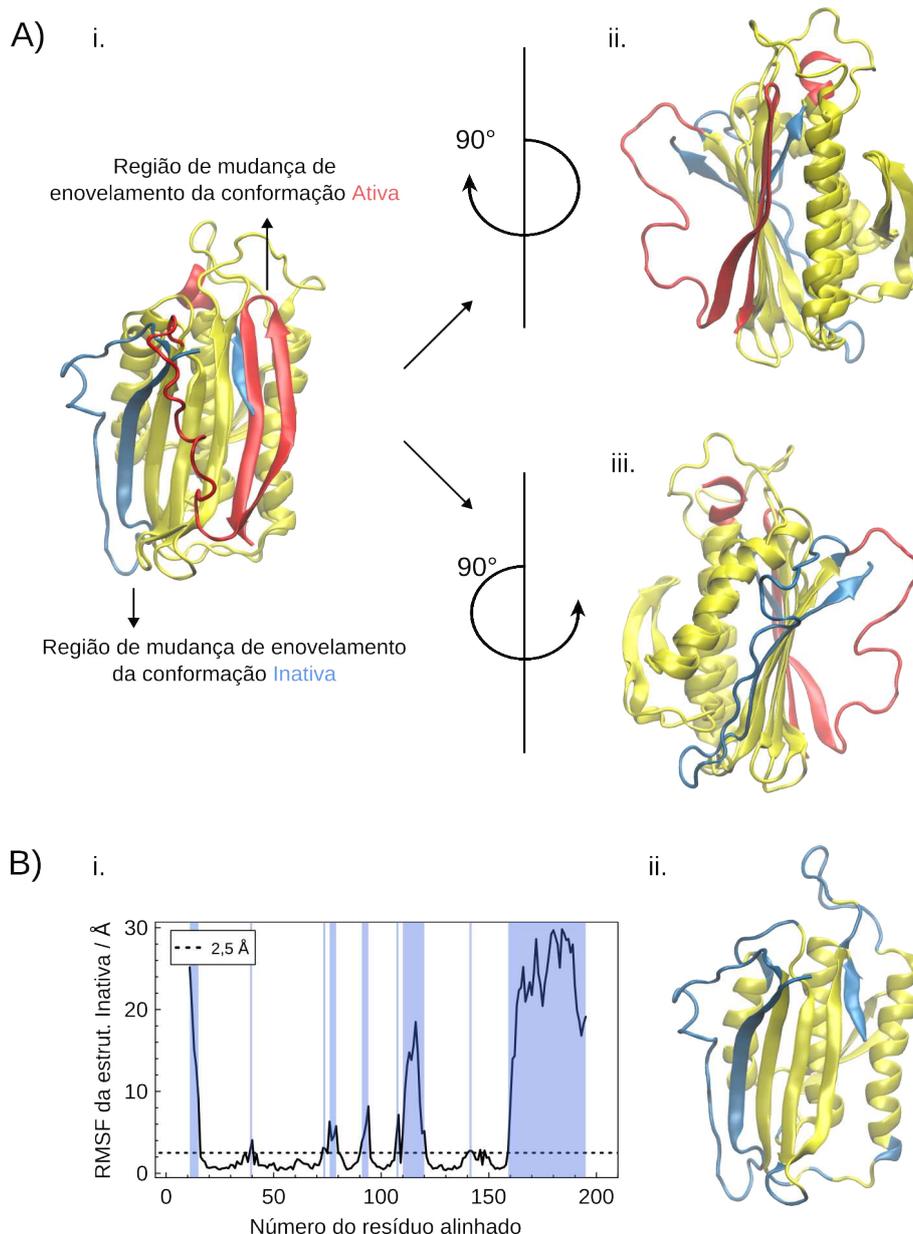


Figura 6.1.2.1: A) Representação em cartoon da MAD2 inativa alinhada à MAD2 ativa, com as regiões de mudança conformacional destacadas em azul e vermelho, respectivamente. Os painéis i., ii. e iii. mostram as perspectivas frontal e lateral, rotacionadas conforme indicado. B) Raiz da flutuação quadrática média (RMSF) da estrutura inativa comparada à ativa após alinhamento com o LovoAlign.²¹⁹ Todos os resíduos com RMSF superior a 2,5 Å estão destacados em azul. C) Representação em cartoon da estrutura inativa colorida em função do RMSF.

6.1.3 Detalhes da simulação

Os modelos SBM para cada estrutura determinada experimentalmente por RMN foram gerados utilizando a versão mais recente do SMOG2,^{143,221} disponível em

<https://smog-server.org/smog2/>. Os pares de contatos nativos para as conformações inativa e ativa foram determinados com o mapa de contatos Shadow.¹³⁸ As simulações foram realizadas com o módulo OpenSMOG²²¹ do OpenMM²²³ (que suporta aceleração com GPU), iniciando com as coordenadas da estrutura ativa. Dois conjuntos de simulações foram realizados para determinar as temperaturas de transição no enovelamento da MAD2. A primeira série de temperaturas, em unidades reduzidas,¹⁴³ variou entre 0,799 e 1,247 com incrementos de 0,0416. Após identificar aproximadamente a temperatura máxima do calor específico (C_v), uma nova série de simulações foi conduzida com temperaturas variando de 0,798 a 0,948 e de 1,014 a 1,047, usando incrementos de 0,000831. Essa abordagem teve como objetivo localizar, dentro de aproximadamente 0,000831, a temperatura correspondente a cada transição da proteína. As simulações foram analisadas com o *Weighted Histogram Analysis Method* (WHAM),¹⁴² conforme implementado no SMOG2,¹⁴³ para a determinação da dependência da temperatura do calor específico ($C_v(T)$) e do potencial de força média ($F(Q)$) em função da fração de contatos nativos (Q). Cada simulação SBM consistiu em $1,5 \times 10^9$ passos com um intervalo de tempo de 0,0005 unidades reduzidas a uma temperatura constante, seguindo os protocolos estabelecidos.¹³⁰ As estruturas foram salvas a cada 20.000 passos, resultando em 75.000 estruturas para análises subsequentes. Finalmente, a simulação ao redor da temperatura de *fold-switching* (T_s) foi estendida dez vezes para aumentar a amostragem das transições conformacionais.

O cálculo da fração total de contatos nativos da proteína consistiu no rastreamento dos contatos exclusivos nas regiões N- e C-terminais da estrutura inativa (190 círculos azuis - resíduos I11 a S16 e K159 a S195, Figura 6.2.1.1C), da estrutura ativa (184 círculos vermelhos - resíduos I11 a S16 e K159 a I201, Figura 6.2.1.1C), e dos contatos em comum (299 círculos amarelos - resíduos A17 a D158, Figura 6.2.1.1C). Um contato foi considerado nativo se a distância entre os átomos $C\alpha$ correspondentes não fosse superior a 20% da observada nos modelos de RMN reportados experimentalmente.

6.1.4 Alinhamento Múltiplo de Sequências e Análise de Acoplamento Direto

As sequências de proteínas foram obtidas do Uniprot com o nome "MAD2", resultando em 14.965 sequências. Essas sequências foram usadas como entrada para o JackHMMER²²⁴ para

gerar o alinhamento múltiplo de sequências (MSA, *Multiple Sequence Alignment*). O JackHMMER itera um modelo oculto de Markov (HMM, *Hidden Markov Model*) contra sequências de proteínas no banco de dados UniProt.²²⁵ Na primeira interação, o JackHMMER constrói um perfil de HMM a partir da sequência desejada e pesquisa um banco de dados de sequências. Subsequentemente, o perfil de HMM é refinado com base nas sequências homólogas identificadas, e esse processo iterativo é repetido para melhorar gradualmente a sensibilidade e a especificidade da busca. A execução do JackHMMER foi iterada até a convergência do MSA ou até que dez iterações fossem concluídas. Os parâmetros usados para os valores -incdomE e -incE foram definidos como $10E^{-20}$ e $10E^{-20}$, respectivamente. Em seguida, o MSA foi filtrado removendo todas as sequências contendo >25% de posições com lacunas e quaisquer colunas com >75% de lacunas.²²⁶ Finalmente, 6608 sequências foram usadas nas Análises de Acoplamento Direto Gaussiano (gDCA)²²⁷ utilizando parâmetros padrão. As 50, 150 e 250 melhores previsões do gDCA foram comparadas com o mapa de contatos das estruturas de RMN. A sobreposição entre os contatos previstos e os contatos do mapa Shadow foi definida como ocorrendo com ± 3 resíduos.

6.1.5 Análises de Frustração

As análises de frustração para a MAD2 inativa (PDB: 1DUJ)²¹¹ e ativa (PDB: 1S2H)²¹² foram realizadas utilizando a versão online do AWSEM-MD Frustratometer (<http://frustratometer.qb.fcen.uba.ar/>).²²⁸ Essa análise foi essencial para investigar quais regiões são mais ou menos frustradas nas conformações da MAD2, e como a frustração pode estar relacionada à sua função biológica. De forma geral, a frustração energética pode ser calculada comparando a energia dos resíduos individuais (ou pares de resíduos) na proteína nativa com as energias de um conjunto de estados alternativos. O índice de frustração por resíduo único (SRFI, *Single Residue Frustration Index*) foi classificado como altamente frustrado (SRFI > -1), neutro ($-1 < \text{SRFI} < 0,55$) ou minimamente frustrado (SRFI > 0,55), conforme proposto por Ferreira et al.^{228,229}

6.2 Resultados e Discussão

Os resultados deste trabalho serão discutidos em quatro partes:

- 1) diferenças estruturais entre as conformações inativa e ativa;

2) análises dos conjuntos estruturais provenientes das simulações e sua correlação com estruturas sugeridas por experimentos de smFRET e RMN;

3) informações estruturais da MAD2 obtidas por meio da coevolução de aminoácidos; e, por fim,

4) índice de frustração nos estados nativos da proteína MAD2.

6.2.1 Estrutura da MAD2 e características do *dual-basin* SBM

A MAD2 é uma proteína humana que se enovela em dois estados nativos distintos, composta por 205 resíduos, cujo núcleo é formado por três folhas- β estáveis (β_4 , β_5 e β_6) sustentadas por três hélices- α (α_A , α_B e α_C), conforme mostrado nas Figuras 6.2.1.1A e 6.2.1.1B. A folha- β_6 (resíduos C149 a D158) é um sítio de ligação ao Cdc20, que está bloqueada pelas folhas β_7 e β_8 (Figura 6.2.1.1A, em azul) na estrutura inativa e exposto na estrutura ativa (Figura 6.2.1.1B). Diferenças estruturais entre as conformações inativa e ativa também são observadas na região N-terminal. A estrutura inativa apresenta uma folha- β (β_1) estabilizada por interações com o núcleo da proteína, enquanto na estrutura ativa esses resíduos fazem parte da hélice α_C . A Figura 6.2.1.1D mostra uma comparação adicional entre o conteúdo de estrutura secundária de cada conformação.

Os mapas de contato de ambas as conformações são apresentados na Figura 6.2.1.1C. Os pontos amarelos representam os contatos comuns (299 contatos), enquanto os pontos azuis (269) e vermelhos (263) indicam os contatos exclusivos das conformações inativa e ativa, respectivamente. Os contatos comuns correspondem a 52% do total de contatos nativos no estado inativo (ou 53% no estado ativo) e persistem durante a mudança conformacional (detalhes serão discutidos na próxima seção). Os contatos mutuamente exclusivos estão localizados principalmente (70%) nas regiões N- e C-terminal (resíduos I11 a S16 e K159 a I201, respectivamente), que estão envolvidas na mudança conformacional da MAD2. A forma ativa apresenta 184 contatos nessas regiões, enquanto a conformação inativa forma 190 contatos. Alguns contatos exclusivos envolvendo resíduos que não participam da região de mudança conformacional (pontos azuis e vermelhos na faixa de A17 a D158) resultam de rearranjos internos dos motivos de estrutura secundária em cada conformação.

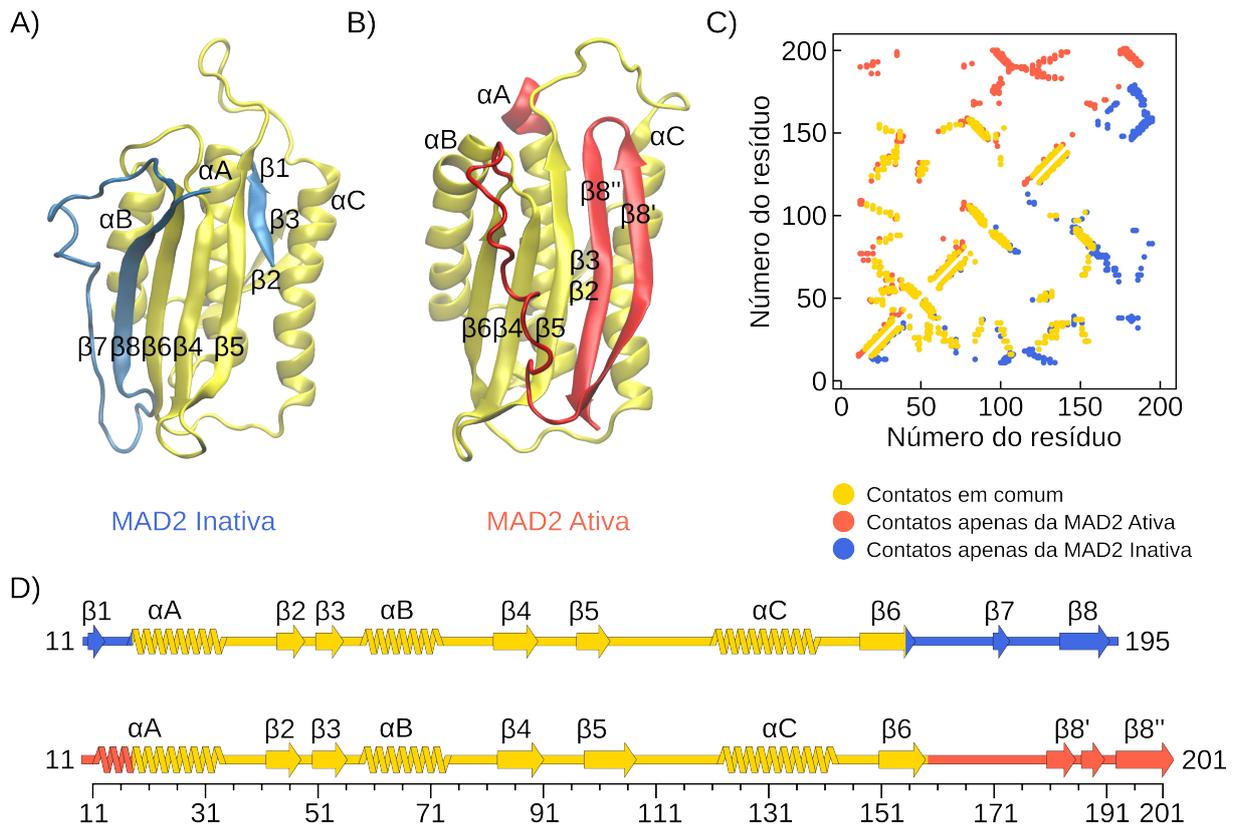


Figura 6.2.1.1: A proteína MAD2 nas formas A) inativa (O-MAD2, PDB: 1DUJ)²¹¹ e B) ativa (C-MAD2, PDB: 1S2H).²¹² Às regiões responsáveis pelo mecanismo de *fold-switching* em A) e B) estão coloridas em azul e vermelho, respectivamente. C) Mapas de contato para as estruturas inativa (canto inferior direito) e ativa (canto superior esquerdo). Os contatos comuns e exclusivos de cada estrutura são representados por pontos amarelos, azuis e vermelhos. D) Comparação da estrutura secundária das conformações inativa (superior) e ativa (inferior).

6.2.2 Paisagem energética da proteína MAD2 e o intermediário

$\beta 1$ -ativo

A transição entre as duas conformações nativas estáveis da MAD2 foi modelada por meio de simulações com os SBMs em diferentes temperaturas. A Figura 6.2.2.1A mostra o calor específico (C_v) em função da temperatura (em unidades da temperatura de enovelamento, T_f). O pico associado ao calor específico define a T_f , onde as populações dos estados enovelado e desenovelado são semelhantes.^{128,230,231} Um leve aumento na curva de C_v é observado em $T = 0,780T_f$, seguido por um pico amplo em $T_s = 0,895T_f$, onde cerca de 84% dos contatos comuns aos estados ativo e inativo são preservados. Essa transição foi associada à transição entre as estruturas

ativa e inativa (discutida posteriormente). O pico em T_s indica uma transição de segunda ordem. Além disso, o menor calor específico em T_s é consistente com uma transição envolvendo rearranjos estruturais parciais.^{216,232} Esses resultados estão alinhados com experimentos de RMN, que indicam que o núcleo da proteína permanece estável durante a mudança de conformação.²¹⁶

A Figura 6.2.2.1B apresenta o perfil de energia livre em função da fração de contatos formados (definida pelo número de contatos, Q , dividido pela média do número de contatos nativos das estruturas ativa e inativa, Q^*). Em T_s , o mínimo de energia livre ocorre em $Q/Q^* \approx 0,8$, uma fração de contatos nativos compatível com ambas as conformações, ativa e inativa. Por outro lado, em $T > T_f$, os estados enovelados com $Q/Q^* \approx 0,18$ são favorecidos. Em T_f , há uma mistura de estados enovelados e desenovelados separados por uma barreira de aproximadamente $4k_B T$.

Os perfis de energia livre bidimensionais das Figuras 6.2.2.1C-E fornecem detalhes adicionais sobre o equilíbrio conformacional da MAD2, onde a população de cada estado é plotada em função da fração de contatos nativos ativos (Q_{ativo}) e inativos ($Q_{inativo}$). Em T_s , há três bacias de energia com estabilidade semelhante. A primeira bacia consiste em flutuações térmicas ao redor do estado nativo ativo ($Q_{ativo} = 0,80$ e $Q_{inativo} = 0,56$).

Com uma fração ligeiramente menor de contatos ativos ($Q_{ativo} = 0,76$) e uma maior de contatos inativos ($Q_{inativo} = 0,60$), observa-se uma segunda bacia. Esta bacia está associada, conforme mostrado na Figura 6.2.2.1F, a um intermediário onde a fita $\beta 1$ N-terminal está formada e posicionada como na forma inativa, mas a estrutura está majoritariamente enovelada no estado ativo, expondo a superfície de interação com alvo Cdc20 na região C-terminal. Assim, nomeamos esse intermediário como $\beta 1$ -ativo. Esse intermediário pode estar relacionado àquele observado em solução sob condições moderadas de desnaturação ($\sim 3-3,5$ M GdmCl),²¹⁵ e ao intermediário da transição de ativação observado por RMN com resolução temporal.²¹⁷ Experimentalmente, as características estruturais desse intermediário não são claras, mas após a diluição do tampão, ele se reenova para o estado ativo. Portanto, o intermediário pode ser semelhante ao estado $\beta 1$ -ativo encontrado na simulação.

A terceira bacia na Figura 6.2.2.1C, onde $Q_{ativo} \approx 0,54$ e $Q_{inativo} \approx 0,76$, está associada ao estado inativo. Por fim, na temperatura de transição de enovelamento (T_f), as bacias associadas aos estados ativo e $\beta 1$ -ativo tornam-se indistinguíveis e estão em equilíbrio com os estados inativo e desenovelado (Figura 6.2.2.1D). Em T_f , a forma inativa é favorecida em relação ao estado ativo.

Quando a temperatura é superior a T_f , apenas a bacia do estado desenovelado é observada (Figura 6.2.2.1E).

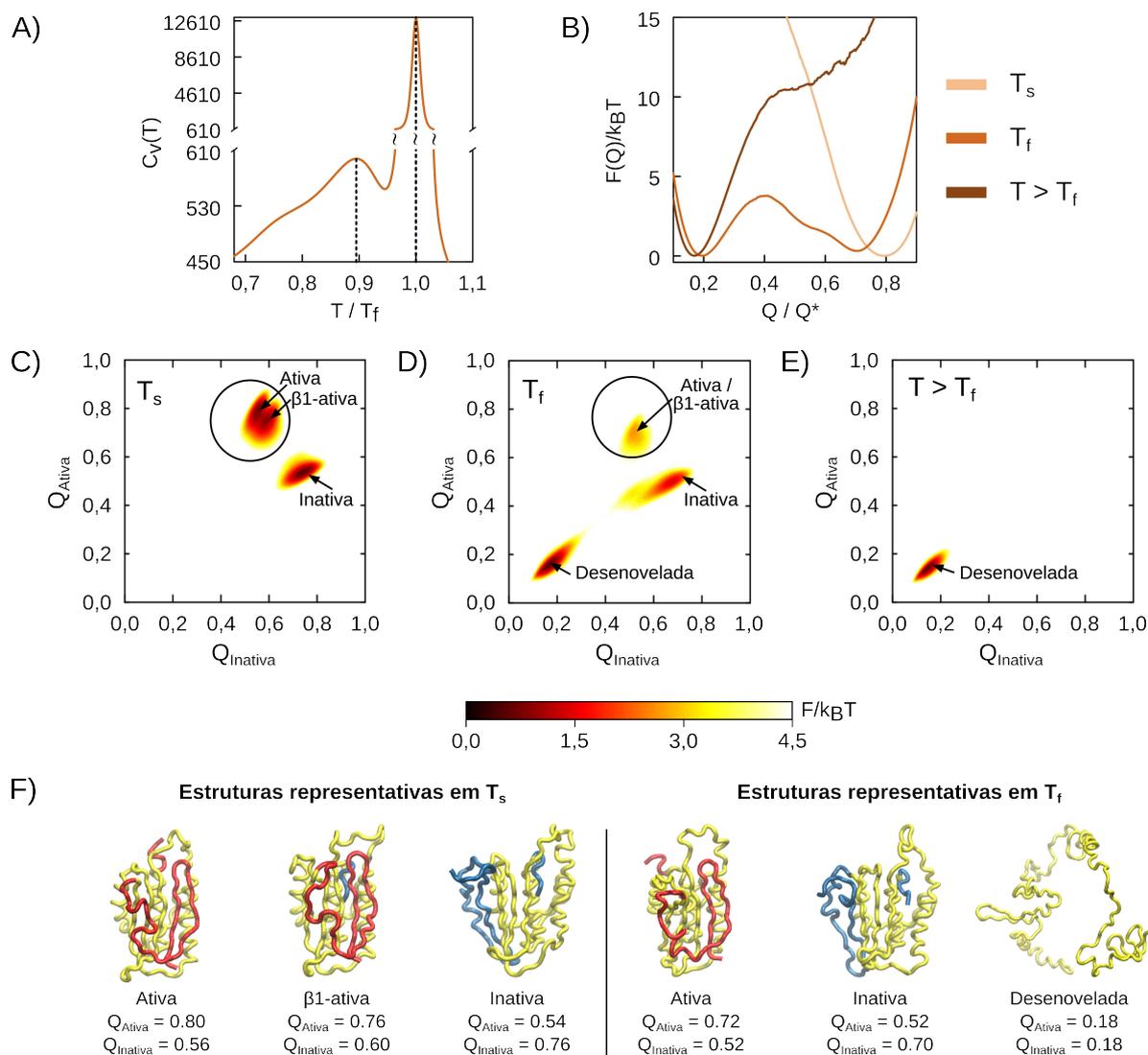


Figura 6.2.2.1: Caracterização do enovelamento da proteína MAD2. A) Calor específico (C_v) em função da temperatura, permitindo a identificação da temperatura de mudança de conformação ($T_s = 0,895T_f$) e da temperatura de enovelamento ($T_f = 1,0$, por definição). B) Energia livre em função da fração de contatos formados (Q/Q^*) em T_s , T_f . Q e Q^* representam o número de contatos e a média do número de contatos nativos ativos e inativos, respectivamente. Densidade de estados em função de $Q_{inativo}$ e Q_{ativo} em: C) $T_s = 0,895T_f$, D) T_f e E) $T > T_f$ ($T = 1,05T_f$). F) Estruturas representativas para cada bacia em T_s (ou T_f).

O papel do estado $\beta 1$ -ativo no mecanismo de mudança de conformação foi investigado pelas probabilidades de transições envolvendo o estado inativo, mostrado na Figura 6.2.2.2C. Primeiramente, agrupamos as estruturas em cada bacia e acompanhamos as transições entre as

bacias na T_s (6.2.2.2A). Transições que visam o estado inativo, originadas no estado β 1-ativo (β 1A \rightarrow I), têm ~ 10 vezes mais probabilidade (90,8% contra 9,2%) de ocorrer do que aquelas originadas no estado ativo (A \rightarrow I) (Figuras 6.2.2.2B e 6.2.2.2C). De maneira semelhante, sair do estado inativo para o estado β 1-ativo também é o caminho mais provável. Portanto, o estado β 1-ativo é um intermediário importante na transição de mudança conformacional da MAD2, tanto nos caminhos diretos quanto inversos.

Para caracterizar ainda mais o estado β 1-ativo, a estabilidade dos contatos dos estados ativo e inativo formados em cada bacia foi acompanhada no T_s . A Figura 6.2.2.2D mostra a extensão em que os contatos são observados na bacia ativa. As porções triangulares superiores do mapa de contato se referem aos contatos ativos e as inferiores aos contatos das formas inativas. As áreas sombreadas (regiões C- e N-terminais) exibem contatos exclusivos de qualquer enovelamento. Contatos sombreados em marrom são satisfeitos e os sombreados em cinza não são satisfeitos em cada bacia. Os contatos comuns das formas ativa e inativa (regiões centrais do mapa de contato: resíduos A17 a D158) apresentam frequências semelhantes em todas as bacias. Por outro lado, na Figura 6.2.2.2D é mostrado que os contatos específicos para o estado inativo têm baixa probabilidade de estar presentes na bacia do estado ativo (resíduos I11 a S16 e K159 a I201). Da mesma forma, na Figura 6.2.2.2F, confirmamos que os contatos específicos para o estado ativo não são formados na bacia inativa.

Finalmente, como mostrado na Figura 6.2.2.2E em relação à 6.2.2.2D, a formação de β 1-ativo a partir do estado ativo ocorre pela formação dos contatos inativos da região N-terminal, com a concomitante perda dos contatos da forma ativa. No entanto, os contatos C-terminais da forma ativa são preservados. Em resumo, a conformação β 1-ativa coexiste com o estado ativo e se difere dele no fato de que a região N-terminal assume a estrutura secundária e a posição observada no estado inativo. Ao mesmo tempo, a região C-terminal está na conformação aberta, expondo a interface para a dimerização e associação com Cdc20.

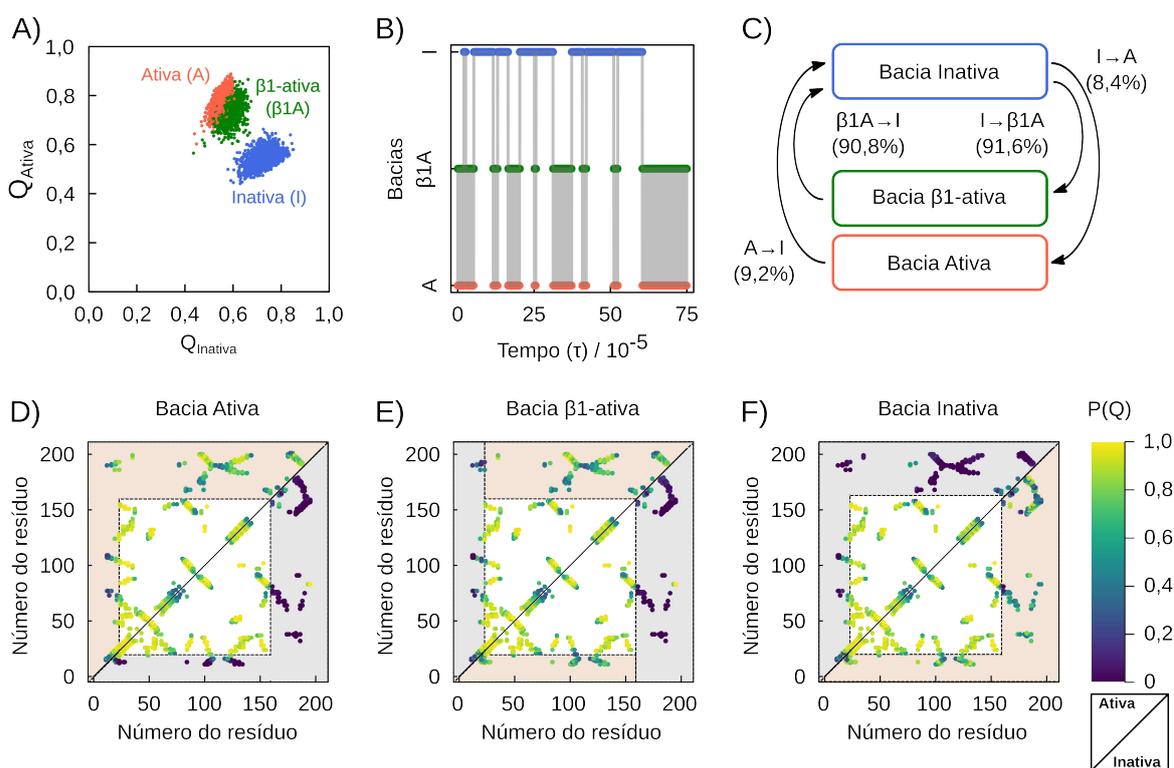


Figura 6.2.2.2: A) Agrupamento das estruturas ativas, $\beta 1$ -ativas e inativas. B) Transições entre as bacias de enovelamento em função do tempo (τ , em unidades reduzidas). C) Percentuais de transições visando o estado inativo (esquerda) e originadas do estado inativo (direita). Mais de 90% das transições envolvem o estado $\beta 1$ -ativo como intermediário. Mapas de contato para as estruturas inativa (triângulo inferior) e ativa (triângulo superior) coloridos pela probabilidade de contatos formados em $T_s = 0,895T_f$. Esses mapas correspondem às bacias D) ativas, E) $\beta 1$ -ativas e F) inativas das estruturas mostradas no painel A. As áreas sombreadas correspondem a regiões de contatos exclusivos de qualquer conformação nativa. Contatos sombreados em marrom são satisfeitos e em cinza não são satisfeitos em cada bacia: o intermediário exibe contatos N-terminais da conformação inativa e contatos C-terminais da conformação ativa.

6.2.3 Contatos coevolutivos suportam um núcleo conservado e uma troca lábil da região N-terminal

A evolução de enovelamento duplo na proteína MAD2 foi investigada por meio de análise de coevolução. A MAD2 pertence à família de proteínas do domínio HORMA (Hop1, Rev7 e MAD2).²³³ A característica estrutural comum entre as proteínas do domínio HORMA é de um núcleo composto pelas folhas $\beta 4$, $\beta 5$ e $\beta 6$, sustentadas pelas hélices αA , αB e αC (Figuras 6.2.1.1A-B).

As Figuras 6.2.3.1A-C mostram o mapa de contatos obtido a partir dos modelos

CAPÍTULO 6. MECANISMOS DE MUDANÇA CONFORMACIONAL DA PROTEÍNA MAD2 129

experimentais (anteriormente mostrado na Figura 6.2.1.1C), sobreposto pelos 50, 150 e 250 principais contatos previstos da análise de coevolução (pontos pretos). Os contatos comuns correspondem a 76,0%, 62,66% e 52,8% de todos os contatos previstos. Essa alta porcentagem sugere que a análise de coevolução é particularmente eficiente em identificar contatos dentro dos elementos estruturais estáveis da proteína. Além disso, os contatos previstos específicos para a estrutura ativa correspondem a 14,0%, 20,67% e 18,8% dos 50, 150 e 250 principais contatos. Por sua vez, os contatos inativos correspondem a apenas 0,0%, 2,0% e 7,2%, respectivamente, indicando que a conformação ativa é um estado enovelado dominante.

Notavelmente, as previsões de contatos na região N-terminal da MAD2 são desafiadoras. Nossa observação de que essa região é lábil na forma ativa, ao passar para a conformação β 1-ativa, pode explicar as correlações coevolutivas limitadas nesta região. Ao mesmo tempo, alguns contatos das regiões de troca C-terminal, exclusivas de qualquer forma da MAD2, são evidentes nos mapas derivados da coevolução.

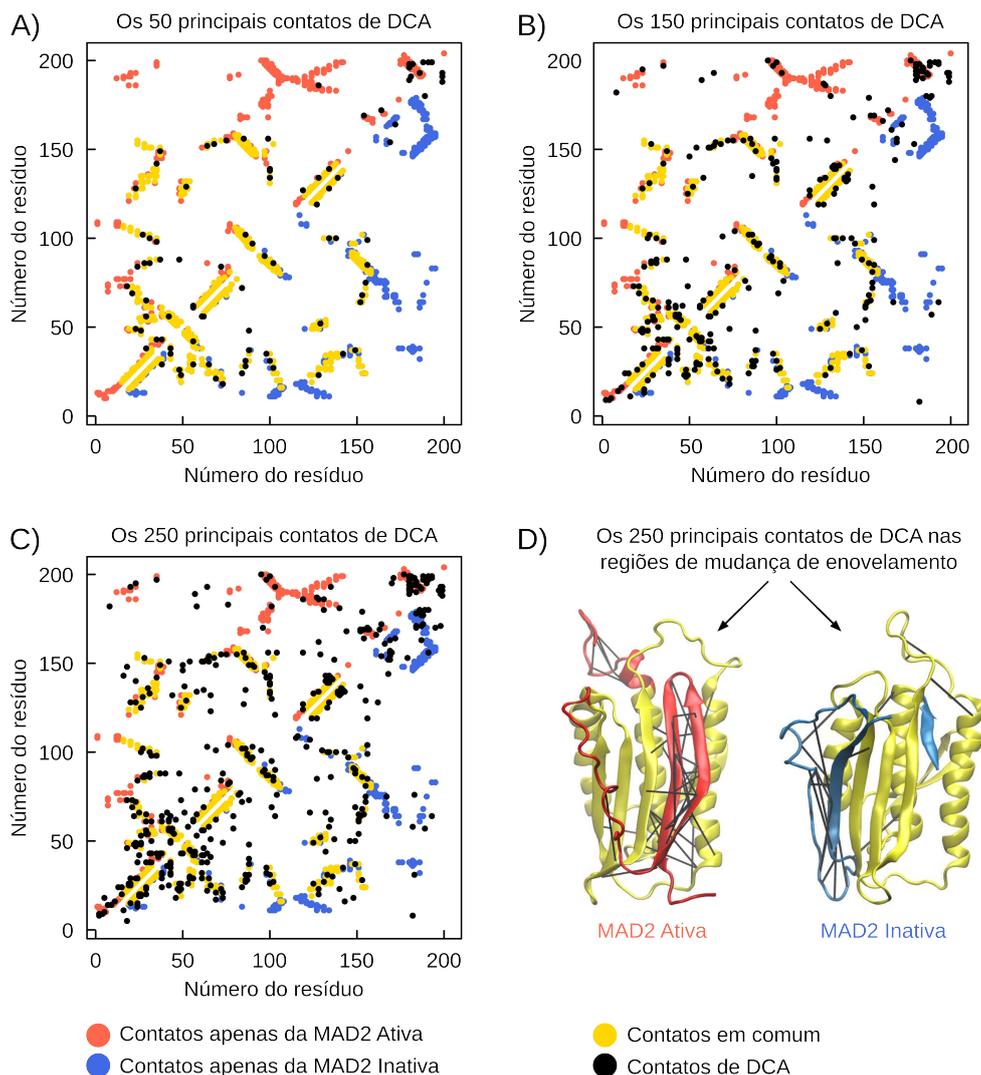


Figura 6.2.3.1: Análise de coevolução da proteína MAD2. Os 50, 150 e 250 principais contatos previstos (pontos pretos) são comparados com os contatos calculados a partir das estruturas experimentais inativas (parte inferior direita, PDB: 1DUJ)²¹¹ e ativas (parte superior esquerda, PDB: 1S2H)²¹² nas figuras A), B) e C), respectivamente. Contatos comuns entre MAD2 inativa e ativa são mostrados em amarelo, enquanto os contatos exclusivos para cada uma são representados em azul e vermelho, respectivamente. D) Representações em estilo cartoon das conformações ativa e inativa são mostradas nos lados esquerdo e direito. Os 250 principais contatos previstos entre resíduos nas regiões de mudança de conformação são indicadas por linhas pretas.

6.2.4 Frustração na proteína MAD2

Regiões de proteínas propensas à alosteria e mudanças conformacionais significativas frequentemente exibem uma alta prevalência de interações altamente frustradas (ou seja, excepcionalmente desestabilizadoras).²³⁴ Em contraste, elementos estruturais estáveis em proteínas tendem a carregar interações minimamente frustradas.²³⁵ Para abordar o papel da frustração na

CAPÍTULO 6. MECANISMOS DE MUDANÇA CONFORMACIONAL DA PROTEÍNA MAD2 131

troca conformacional da proteína MAD2, calculamos o índice de frustração por resíduo único (SRFI) para cada conformação.²²⁸

A Figura 6.2.4.1A mostra a distribuição do SRFI para as conformações inativa (barras azuis) e ativa (barras vermelhas). As distribuições são semelhantes, com 42,7% de resíduos neutros ($-1 < \text{SRFI} < 0,55$) ou minimamente frustrados (29,73%) ($\text{SRFI} > 0,55$). Os pontos verdes na Figura 6.2.4.1B representam os resíduos minimamente frustrados das estruturas inativa e ativa. Os pontos cinzas indicam resíduos neutros ou minimamente frustrados em qualquer conformação (54,6% dos resíduos). 5,95% dos resíduos (11 ao todo) são altamente frustrados em ambas as conformações (pontos vermelhos na Figura 6.2.4.1B). A maioria desses resíduos são carregados (E22, E105, D112 e D160) ou polares (Q34, Q47, N66, Q71 e W167). Esses resíduos estão localizados no loop (D112) e em regiões envolvidas no mecanismo de mudança de conformação (Y38, Q71, E105, D160, P164 e W167). Os resíduos Y38, Q71 e E105 não fazem parte das sequências de troca N-terminal (I11 a S16) e C-terminal (K159 a I201), mas interagem diretamente com resíduos que estabilizam as conformações inativa ou ativa.

Alguns resíduos são altamente frustrados em uma estrutura, mas não na outra. Esses resíduos, representados por pontos magenta e laranja na Figura 6.2.4.1B, correspondem a 4,32% e 5,41% de todos os resíduos nas conformações inativa e ativa, respectivamente. As Figuras 6.2.4.1C (ou 6.2.4.1D) mostram os resíduos que são altamente frustrados apenas em uma das estruturas. Vale ressaltar que, embora esses resíduos estejam presentes ao longo de toda a cadeia da proteína, a maioria deles é encontrada em partes biologicamente críticas da MAD2. Por exemplo, D152 e Y156 são resíduos altamente frustrados da $\beta 6$, que é o local de ligação ao Cdc20²¹⁰ ou de interações com a região de mudança de conformação C-terminal na MAD2 inativa. N177 (Figura 6.2.4.1C), E165 e K166 (Figura 6.2.4.1D) sofrem mudanças conformacionais significativas durante a mudança de conformação. Nossa hipótese é que K108, R117, K119 e Q121 (Figura 6.2.4.1C) são altamente frustrados na conformação ativa porque interagem com a região N-terminal da proteína, que sofre mudança conformacional para o estado $\beta 1$.

Note que há também resíduos altamente frustrados dentro dos elementos estruturais estáveis da MAD2. R133, por exemplo, pertence à hélice αC , é altamente frustrado no estado inativo e é necessário para a dimerização da MAD2.²³⁶ Curiosamente, R133 não está mais frustrado após a formação do dímero (Índice de Frustração R133: -0.801, PDB: 3GMH).²³⁷ Resíduos frustrados

na superfície dos elementos estruturais estáveis - como E22 e Q34 na αA , N66 na αB e E41, Q47 e Y49 na $\beta 2-\beta 3$ - podem ser minimamente frustrados ao interagir com proteínas do ponto de checagem do fuso (Spindle Assembly Checkpoint, SAC).^{237,238}

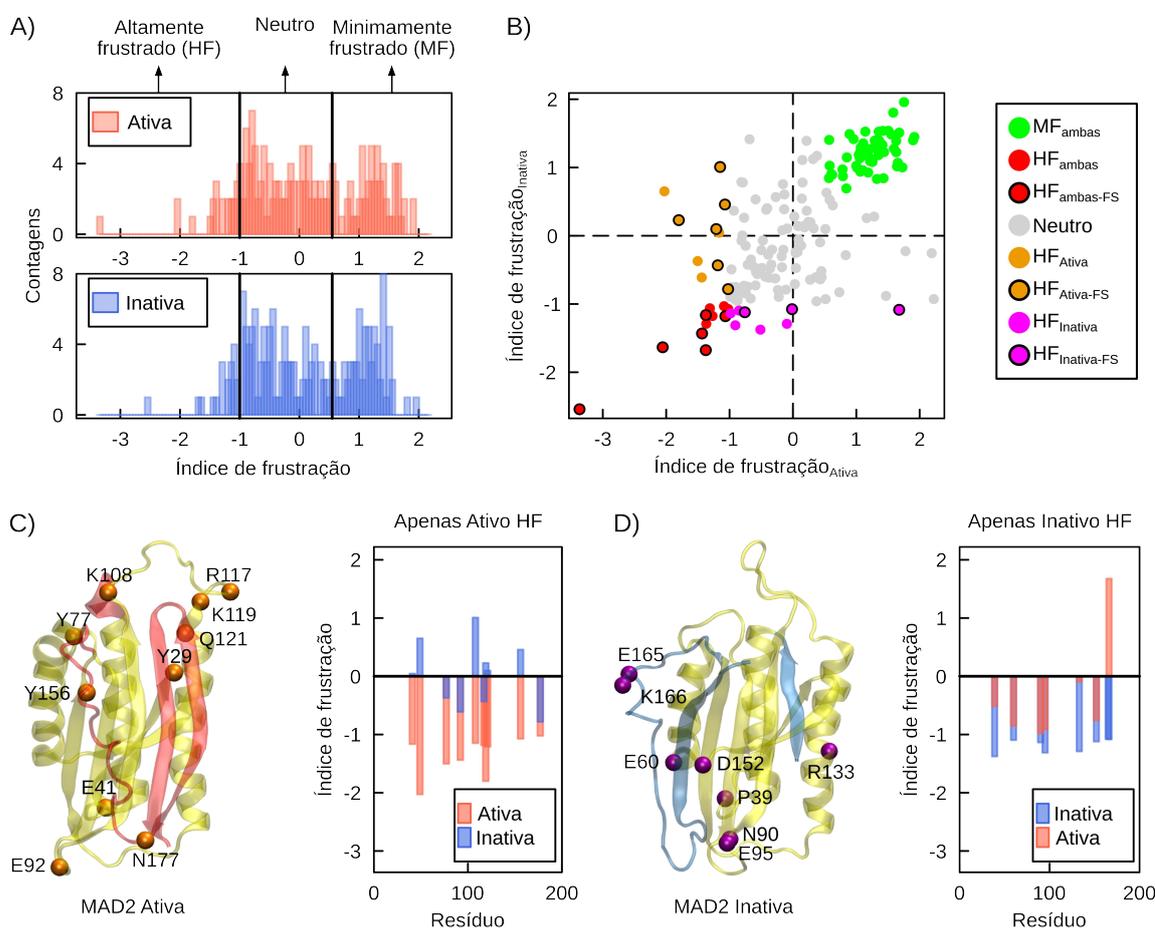


Figura 6.2.4.1: Análises de frustração na proteína MAD2. A) Distribuição do SRFI para as conformações ativa (barras vermelhas) e inativa (barras azuis). Linhas verticais pretas indicam o raio de corte usado para definir resíduos minimamente frustrados (MF, SRFI > 0,55), neutros (-1 < SRFI < 0,55) ou altamente frustrados (HF, SRFI < -1). B) Correlação entre o SRFI nas conformações inativa (eixo y) e ativa (eixo x). Pontos verdes representam os resíduos MF (29,73%), enquanto os pontos vermelhos representam resíduos HF (5,95%) em ambas as conformações. Pontos cinzas representam resíduos neutros em ambas as conformações ou neutros em uma e MF na outra (54,6%). SRFIs para resíduos HF presentes apenas na estrutura inativa (ativa) são mostrados em magenta (laranja). Resíduos HF envolvidos nas regiões de *fold-switching* (FS) são destacados com um contorno preto. Os painéis C) e D) mostram a representação em estilo cartoon e o gráfico de barras dos resíduos HF no estado ativo e inativo, respectivamente.

6.3 Conclusões

O mecanismo de enovelamento da proteína MAD2 pode ser capturado usando um novo modelo baseado na combinação da paisagem energética de ambas as conformações da proteína. As simulações capturam a formação de um intermediário no qual a fita- $\beta 1$ é formada ($\beta 1$ -ativo), mas o núcleo da proteína mantém a forma ativa. Isso é consistente com as observações experimentais sobre a transição de ativação, que envolve, inicialmente, o rearranjo C-terminal e, posteriormente, uma troca N-terminal.²¹⁷ Neste trabalho, mostramos que a mudança estrutural N-terminal ocorre rapidamente se a MAD2 estiver na forma ativa, e que esse intermediário constitui o principal caminho para as transições de ativação ou inativação.

Dados recentes de smFRET também revelam a existência de um estado intermediário estável em soluções moderadamente desnaturantes de GdmCl (~ 3 - $3,5$ mol L⁻¹).²¹⁵ Após a diluição do GdmCl, esse intermediário se converte exclusivamente para a forma ativa da MAD2, sugerindo uma maior semelhança estrutural com essa conformação, assim como o identificado nas simulações com os SBMs ($\beta 1$ -ativo). Portanto, os resultados das simulações permitirão investigar os efeitos do GdmCl na estabilização dos diferentes estados conformacionais da MAD2, buscando elucidar os mecanismos de solvatação que suportem as observações experimentais.

Finalmente, a análise de acoplamento direto prevê contatos que correspondem aos elementos estruturais estáveis na proteína MAD2, predominantemente para o estado ativo. Além disso, a análise de frustração revela que os estados nativos da MAD2 são minimamente frustrados, o que contribui para sua capacidade de acessar estados funcionais distintos enquanto mantêm a estabilidade. Resíduos altamente frustrados estão predominantemente localizados nos segmentos de mudança conformacional e nas regiões de ligação da proteína.

Capítulo 7

Conclusões gerais

O principal objetivo desta tese foi entender como diferentes solventes afetam a termodinâmica do processo de enovelamento de proteínas, utilizando técnicas computacionais. O enovelamento de proteínas ocorre na escala de tempo de microssegundos a segundos, exigindo métodos que permitem avaliar as transições conformacionais em escalas de tempo compatíveis com os experimentos. Nos trabalhos apresentados, empregamos tanto simulações de DM com troca de réplicas quanto as simulações com os modelos baseados na estrutura para caracterizar de forma satisfatória a dinâmica conformacional da proteína em cada estudo.

No Capítulo 3, exploramos os mecanismos de interação e estabilização do TFE no equilíbrio conformacional do peptídeo AAQAA₃, cujo trabalho está publicado na revista "*Journal of Molecular Liquids*".¹⁷⁴ Neste trabalho, utilizamos simulações de DM com troca de réplicas com o método REST2 para explorar o conjunto conformacional do peptídeo. O trabalho confirma o papel estabilizador de α -hélices do TFE e fornece uma visão detalhada dos mecanismos de estabilização do cossolvente, obtida pela modificação dos potenciais de interação entre o oxigênio da carbonila na cadeia principal e o hidrogênio polar do TFE.

No Capítulo 4, estudamos o enovelamento do domínio B da proteína A (BdpA)¹⁷⁷ utilizando simulações com os C α -SBM, que são potenciais parametrizados a partir da estrutura nativa. Esses potenciais são amplamente utilizados na determinação do mecanismo de enovelamento de proteínas como a BdpA, que apresenta uma cinética de dois estados. Neste trabalho, além de caracterizar o enovelamento da BdpA, determinamos a correlação na formação das hélices, agrupamos as estruturas de enovelamento e estudamos os modos de interação entre as

moléculas de água e todo o conjunto de estruturas de enovelamento. Os resultados apresentados no Capítulo 4, publicados na revista “*Journal of Chemical Information and Modeling*”, possibilitam análises dos efeitos de cossolventes no equilíbrio do enovelamento de proteínas, abordado com detalhes no Capítulo 5.

Especificamente, o Capítulo 5, disponível como *preprint*,²³⁹ aborda um desafio fundamental na biofísica, que é o de entender como osmólitos modulam o enovelamento e a estabilidade de proteínas. O estudo investiga os efeitos da ureia e do TMAO na superfície de energia das proteínas modelo SH3 (rica folhas- β) e BdpA (rica em α -hélices), que são amplamente utilizadas em estudos de enovelamento. Neste trabalho, aplicamos a metodologia introduzida no Capítulo 4 para caracterizar os efeitos dos osmólitos em vários estados enovelados e desenovelados das proteínas, incluindo cálculos de energia livre de transferência. Os valores de energia livre estão de acordo com resultados experimentais e, portanto, validam os campos de força e métodos de análises utilizados. Esses resultados fornecem uma nova perspectiva dos efeitos de osmólitos nos caminhos de enovelamento de proteínas, com possíveis aplicações biotecnológicas.

Finalmente, no Capítulo 6, discutimos as características do enovelamento da proteína MAD2. A MAD2 é uma proteína que pertence à classe das *fold-switching* que, por definição, podem se enovelar de forma reversível em mais de um estado nativo. Neste trabalho, desenvolvemos um potencial, a partir de modelos $C\alpha$ -SBM, considerando os estados ativo e inativo da MAD2. A partir do novo potencial, caracterizamos os diferentes estados nativos da MAD2 e, inclusive, um estado intermediário com características estruturais sugeridas por experimentos de smFRET na presença do desnaturante GdmCl. Esses resultados permitirão investigações dos efeitos do GdmCl na estabilização dos diferentes estados conformacionais da MAD2, que poderão corroborar as observações experimentais.

Bibliografia

1. A. Mohan, V. Puppala e I. Roy. "Protein Folding Pathways in the Presence of Osmolytes". Em: *Cellular Osmolytes: From Chaperoning Protein Folding to Clinical Perspectives*. Ed. por L. R. Singh, T. A. Dar e K. Kumari. Singapore: Springer Nature Singapore, 2024, pp. 1–16.
2. D. L. Nelson e M. M. Cox. "Princípios de Bioquímica de Lehninger". Em: 6ª ed. Porto Alegre: Artmed Editora, 2014. Cap. 3. ISBN: 9781429234146.
3. C. B. Anfinsen. "Principles that govern the folding of protein chains". Em: *Science* 181.4096 (1973), pp. 223–230. DOI: <https://doi.org/10.1126/science.181.4096.223>.
4. C. Gomes e P. Faísca. "Protein Folding: An Introduction". Em: *Protein Folding*. SpringerBriefs in Molecular Science. Springer, Cham, 2019, pp. 1–63. DOI: https://doi.org/10.1007/978-3-319-00882-0_1.
5. D. R. Canchi e A. E. García. "Cosolvent effects on protein stability". Em: *Annual Review of Physical Chemistry* 64.1 (2013), pp. 273–293. DOI: <https://doi.org/10.1146/annurev-physchem-040412-110156>.
6. A. Rani e P. Venkatesu. "Changing relations between proteins and osmolytes: a choice of nature". Em: *Physical Chemistry Chemical Physics* 20.31 (2018), pp. 20315–20333. DOI: <https://doi.org/10.1039/C8CP02949K>.
7. C. B. Anfinsen et al. "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain". Em: *Proceedings of the National Academy of Sciences* 47.9 (1961), pp. 1309–1314. DOI: <https://doi.org/10.1073/pnas.47.9.1309>.
8. E. Haber e C. B. Anfinsen. "Side-chain interactions governing the pairing of half-cystine residues in ribonuclease". Em: *Journal of Biological Chemistry* 237.6 (1962), pp. 1839–1844. DOI: [https://doi.org/10.1016/S0021-9258\(19\)73945-3](https://doi.org/10.1016/S0021-9258(19)73945-3).

9. C. Levinthal. "How to Fold Graciously". Em: *Mössbauer Spectroscopy in Biological Systems*. Ed. por J. T. P. DeBrunner e E. Munck. Urbana, IL: University of Illinois Bulletin, 1969, pp. 22–24.
10. L. Martínez. "Introducing the Levinthal's Protein Folding Paradox and Its Solution". Em: *Journal of Chemical Education* 91.11 (2014), pp. 1918–1923. DOI: <https://doi.org/10.1021/ed300302h>.
11. R. Zwanzig, A. Szabo e B. Bagchi. "Levinthal's paradox". Em: *Proceedings of the National Academy of Sciences* 89.1 (1992), pp. 20–22. DOI: <https://doi.org/10.1073/pnas.89.1.20>.
12. J. D. Bryngelson e P. G. Wolynes. "Spin glasses and the statistical mechanics of protein folding". Em: *Proceedings of the National Academy of Sciences* 84.21 (1987), pp. 7524–7528. DOI: <https://doi.org/10.1073/pnas.84.21.7524>.
13. J. N. Onuchic, Z. Luthey-Schulten e P. G. Wolynes. "Theory of protein folding: the energy landscape perspective". Em: *Annual Review of Physical Chemistry* 48.1 (1997), pp. 545–600. DOI: <https://doi.org/10.1146/annurev.physchem.48.1.545>.
14. J. N. Onuchic e P. G. Wolynes. "Theory of protein folding". Em: *Current Opinion in Structural Biology* 14.1 (2004), pp. 70–75. DOI: <https://doi.org/10.1016/j.sbi.2004.01.009>.
15. K. A. Dill et al. "The protein folding problem". Em: *Annual Review of Biophysics* 37.1 (2008), pp. 289–316. DOI: <https://doi.org/10.1146/annurev.biophys.37.092707.153558>.
16. 3. Brooks C. L., J. N. Onuchic e D. J. Wales. "Statistical thermodynamics. Taking a walk on a landscape". Em: *Science* 293.5530 (2001), pp. 612–613. DOI: <https://doi.org/10.1126/science.1062559>.
17. D. Harries e J. Rösgen. "A practical guide on how osmolytes modulate macromolecular properties". Em: *Methods in Cell Biology* 84 (2008), pp. 679–735. DOI: [https://doi.org/10.1016/S0091-679X\(07\)84022-2](https://doi.org/10.1016/S0091-679X(07)84022-2).
18. S. R. Wlodarczyk et al. "Influence and effect of osmolytes in biopharmaceutical formulations". Em: *European Journal of Pharmaceutics and Biopharmaceutics* 131 (2018), pp. 92–98. DOI: <https://doi.org/10.1016/j.ejpb.2018.07.019>.
19. S. Giri et al. "Harnessing the Power of Osmolytes for Industrial and Pharmaceutical Applications". Em: *Cellular Osmolytes*. Ed. por L. Singh, T. Dar e K. Kumari. Springer, Singapore, 2024. DOI: https://doi.org/10.1007/978-981-97-6001-5_8.

20. S. N. Timasheff. "Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components". Em: *Proceedings of the National Academy of Sciences* 99.15 (2002), pp. 9721–9726. DOI: <https://doi.org/10.1073/pnas.122225399>.
21. J. G. Kirkwood e F. P. Buff. "The statistical mechanical theory of solutions. I". Em: *The Journal of Chemical Physics* 19.6 (1951), pp. 774–777. DOI: <https://doi.org/10.1063/1.1748352>.
22. A. Ben-Naim. "Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system". Em: *The Journal of Chemical Physics* 67.11 (1977), pp. 4884–4890. DOI: <https://doi.org/10.1063/1.434669>.
23. L. Martínez. "ComplexMixtures.jl: Investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective". Em: *Journal of Molecular Liquids* 347 (2022), p. 117945. DOI: <https://doi.org/10.1016/j.molliq.2021.117945>.
24. L. Martínez e S. Shimizu. "Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions". Em: *Journal of Chemical Theory and Computation* 13.12 (2017), pp. 6358–6372. DOI: <https://doi.org/10.1021/acs.jctc.7b00599>.
25. W. Song, R. Biswas e M. Maroncelli. "Intermolecular interactions and local density augmentation in supercritical solvation: A survey of simulation and experimental results". Em: *The Journal of Physical Chemistry A* 104.30 (2000), pp. 6924–6939. DOI: <https://doi.org/10.1021/jp000888d>.
26. J.-M. Simon et al. "Kirkwood-Buff integrals: From fluctuations in finite volumes to the thermodynamic limit". Em: *The Journal of Chemical Physics* 157.13 (2022), p. 130901. DOI: <https://doi.org/10.1063/5.0106162>.
27. A. Ben-Naim. "Molecular Theory of Solutions". Em: Oxford: Oxford University Press, 2006. Cap. 4.
28. J. Abildskov, R. Wedberg e J. P. O'Connell. "Fluctuation Solution Theory Properties from Molecular Simulation". Em: *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*. Ed. por P. E. Smith, E. Matteoli e J. P. O'Connell. Elsevier, 2013. Cap. 6, p. 136.

29. V. Piccoli e L. Martínez. "Correlated counterion effects on the solvation of proteins by ionic liquids". Em: *Journal of Molecular Liquids* 320 (2020), p. 114347. DOI: <https://doi.org/10.1016/j.molliq.2020.114347>.
30. I. P. de Oliveira e L. Martínez. "The shift in urea orientation at protein surfaces at low pH is compatible with a direct mechanism of protein denaturation". Em: *Physical Chemistry Chemical Physics* 22.1 (2019), pp. 354–367. DOI: <https://doi.org/10.1039/C9CP05196A>.
31. I. L. Shulgin e E. Ruckenstein. "A protein molecule in an aqueous mixed solvent: fluctuation theory outlook". Em: *The Journal of Chemical Physics* 123.5 (2005), p. 054909. DOI: <https://doi.org/10.1063/1.2011388>.
32. I. L. Shulgin e E. Ruckenstein. "A protein molecule in a mixed solvent: the preferential binding parameter via the Kirkwood-Buff theory". Em: *Biophysical Journal* 90.2 (2006), pp. 704–707. URL: [https://www.cell.com/biophysj/fulltext/S0006-3495\(06\)72249-4](https://www.cell.com/biophysj/fulltext/S0006-3495(06)72249-4).
33. E. F. Casassa e H. Eisenberg. "Thermodynamic analysis of multicomponent solutions". Em: *Advances in Protein Chemistry* 19 (1964), pp. 287–395. DOI: [https://doi.org/10.1016/S0065-3233\(08\)60191-6](https://doi.org/10.1016/S0065-3233(08)60191-6).
34. C. W. Pinger, A. A. Heller e D. M. Spence. "A Printed Equilibrium Dialysis Device with Integrated Membranes for Improved Binding Affinity Measurements". Em: *Analytical Chemistry* 89.14 (2017), pp. 7302–7306. DOI: <https://doi.org/10.1021/acs.analchem.7b01848>.
35. P. E. Smith. "Equilibrium dialysis data and the relationships between preferential interaction parameters for biological systems in terms of Kirkwood-Buff integrals". Em: *The Journal of Physical Chemistry B* 110.6 (2006), pp. 2862–2868. DOI: <https://doi.org/10.1021/jp056100e>.
36. S. N. Timasheff. "Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated". Em: *Advances in Protein Chemistry* 51.51 (1998), pp. 355–432. DOI: [https://doi.org/10.1016/s0065-3233\(08\)60656-7](https://doi.org/10.1016/s0065-3233(08)60656-7).
37. J. Record M. T. e C. F. Anderson. "Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a two-domain model". Em: *Biophysical Journal* 68.3 (1995), pp. 786–794. DOI: [https://doi.org/10.1016/S0006-3495\(95\)80254-7](https://doi.org/10.1016/S0006-3495(95)80254-7).

38. S. N. Timasheff. "Water as ligand: preferential binding and exclusion of denaturants in protein unfolding". Em: *Biochemistry* 31.41 (1992), pp. 9857–9864. DOI: <https://doi.org/10.1021/bi00156a001>.
39. C. Tanford. "Protein denaturation: Part C. Theoretical models for the mechanism of denaturation". Em: *Advances in Protein Chemistry* 24 (1970), pp. 1–95. DOI: [https://doi.org/10.1016/S0065-3233\(08\)60241-7](https://doi.org/10.1016/S0065-3233(08)60241-7).
40. J. Wyman J. "Linked Functions and Reciprocal Effects in Hemoglobin: A Second Look". Em: *Advances in Protein Chemistry* 19 (1964), pp. 223–286. DOI: [https://doi.org/10.1016/S0065-3233\(08\)60190-4](https://doi.org/10.1016/S0065-3233(08)60190-4).
41. C. Tanford. "Isothermal unfolding of globular proteins in aqueous urea solutions". Em: *Journal of the American Chemical Society* 86.10 (1964), pp. 2050–2059. DOI: <https://doi.org/10.1021/ja01064a028>.
42. Y. Nozaki e C. Tanford. "The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions". Em: *Journal of Biological Chemistry* 245.7 (1970), pp. 1648–1652. DOI: [https://doi.org/10.1016/S0021-9258\(19\)77141-5](https://doi.org/10.1016/S0021-9258(19)77141-5).
43. Y. Nozaki e C. Tanford. "The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions". Em: *Journal of Biological Chemistry* 238 (1963), pp. 4074–4081. DOI: [https://doi.org/10.1016/S0021-9258\(18\)51830-5](https://doi.org/10.1016/S0021-9258(18)51830-5).
44. D. W. Bolen e G. D. Rose. "Structure and energetics of the hydrogen-bonded backbone in protein folding". Em: *Annual Review of Biochemistry* 77.1 (2008), pp. 339–362. DOI: <https://doi.org/10.1146/annurev.biochem.77.061306.131357>.
45. S. N. Timasheff. "Protein hydration, thermodynamic binding, and preferential hydration". Em: *Biochemistry* 41.46 (2002), pp. 13473–13482. DOI: <https://doi.org/10.1021/bi020316e>.
46. M. G. Martin. "Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor–liquid coexistence curves and liquid densities". Em: *Fluid Phase Equilibria* 248.1 (2006), pp. 50–55. DOI: <https://doi.org/10.1016/j.fluid.2006.07.014>.
47. L. Verlet. "Computer"experiments"on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules". Em: *Physical Review* 159.1 (1967), pp. 98–103. DOI: <https://doi.org/10.1103/PhysRev.159.98>.

48. C. Clementi, H. Nymeyer e J. N. Onuchic. "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". Em: *Journal of Molecular Biology* 298.5 (2000), pp. 937–953. DOI: <https://doi.org/10.1006/jmbi.2000.3693>.
49. M. W. Grunst et al. "Structure and inhibition of SARS-CoV-2 spike refolding in membranes". Em: *Science* 385.6710 (2024), pp. 757–765. DOI: <https://doi.org/10.1126/science.adn5658>.
50. H. Taketomi, Y. Ueda e N. Gō. "Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions". Em: *International Journal of Peptide and Protein Research* 7.6 (1975), pp. 445–459. DOI: <https://doi.org/10.1111/j.1399-3011.1975.tb02465.x>.
51. K.-I. Okazaki et al. "Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations". Em: *Proceedings of the National Academy of Sciences* 103.32 (2006), pp. 11844–11849. DOI: <https://doi.org/10.1073/pnas.0604375103>.
52. J. K. Noel e J. N. Onuchic. "The many faces of structure-based potentials: From protein folding landscapes to structural characterization of complex biomolecules". Em: *Computational Modeling of Biological Systems*. Boston, MA: Springer, 2012, pp. 31–54. DOI: https://doi.org/10.1007/978-1-4614-2146-7_2.
53. S. Bazmi, B. Seifi e S. Wallin. "Simulations of a protein fold switch reveal crowding-induced population shifts driven by disordered regions". Em: *Communications Chemistry* 6.1 (2023), p. 191. DOI: <https://doi.org/10.1038/s42004-023-00995-2>.
54. C. A. Ramírez-Sarmiento et al. "Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape". Em: *PLOS Computational Biology* 11.7 (2015), e1004379. DOI: <https://doi.org/10.1371/journal.pcbi.1004379>.
55. P. Galaz-Davison, D. U. Ferreira e C. A. Ramírez-Sarmiento. "Coevolution-derived native and non-native contacts determine the emergence of a novel fold in a universally conserved family of transcription factors". Em: *Protein Science* 31.6 (2022), e4337. DOI: <https://doi.org/10.1002/pro.4337>.

56. M. Rivera et al. "Dimer dissociation is a key energetic event in the fold-switch pathway of KaiB". Em: *Biophysical Journal* 121.6 (2022), pp. 943–955. DOI: <https://doi.org/10.1016/j.bpj.2022.02.012>.
57. A. Laio e M. Parrinello. "Escaping free-energy minima". Em: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. DOI: <https://doi.org/10.1073/pnas.202427399>.
58. C. Tsallis e D. A. Stariolo. "Generalized simulated annealing". Em: *Physica A* 233.1-2 (1996), pp. 395–406. DOI: [https://doi.org/10.1016/S0378-4371\(96\)00271-3](https://doi.org/10.1016/S0378-4371(96)00271-3).
59. Y. Sugita e Y. Okamoto. "Replica-exchange molecular dynamics method for protein folding". Em: *Chemical Physics Letters* 314 (1999), pp. 141–151. DOI: [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
60. S. e. a. Izrailev. "Steered Molecular Dynamics". Em: *Computational Molecular Dynamics: Challenges, Methods, Ideas*. Ed. por P. Deuffhard et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 39–65.
61. Y. M. Rhee e V. S. Pande. "Multiplexed-replica exchange molecular dynamics method for protein folding simulation". Em: *Biophysical Journal* 84.2 (2003), pp. 775–786. DOI: [https://doi.org/10.1016/S0006-3495\(03\)74897-8](https://doi.org/10.1016/S0006-3495(03)74897-8).
62. N. Metropolis et al. "Equation of state calculations by fast computing machines". Em: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092.
63. D. A. Kofke. "On the acceptance probability of replica-exchange Monte Carlo trials". Em: *The Journal of Chemical Physics* 117.15 (2002), pp. 6911–6914. DOI: <https://doi.org/10.1063/1.1507776>.
64. X. Li et al. "An improved replica-exchange sampling method: temperature intervals with global energy reassignment". Em: *The Journal of Chemical Physics* 127.16 (2007), p. 164116. DOI: <https://doi.org/10.1063/1.2780152>.
65. R. Zhou. "Replica exchange molecular dynamics method for protein folding simulation". Em: *Methods in Molecular Biology* 350 (2007), pp. 205–223. DOI: <https://doi.org/10.1385/1-59745-189-4:205>.
66. H. Fukunishi, O. Watanabe e S. Takada. "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction". Em:

- The Journal of Chemical Physics* 116.20 (2002), pp. 9058–9067. DOI: <https://doi.org/10.1063/1.1472510>.
67. P. Liu et al. “Replica exchange with solute tempering: a method for sampling biological systems in explicit water”. Em: *Proceedings of the National Academy of Sciences* 102.39 (2005), pp. 13749–13754. DOI: <https://doi.org/10.1073/pnas.0506346102>.
68. X. Huang et al. “Replica exchange with solute tempering: efficiency in large scale systems”. Em: *The Journal of Physical Chemistry B* 111.19 (2007), pp. 5405–5410. DOI: <https://doi.org/10.1021/jp068826w>.
69. L. Wang, R. A. Friesner e B. J. Berne. “Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2)”. Em: *The Journal of Physical Chemistry B* 115.30 (2011), pp. 9431–9438. DOI: <https://doi.org/10.1021/jp204407d>.
70. L. Wang, R. A. Friesner e B. J. Berne. “Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2)”. Em: *The Journal of Physical Chemistry B* 115.30 (2011), pp. 9431–9438. DOI: <https://doi.org/10.1021/jp204407d>.
71. R. Rajan e P. Balaram. “A model for the interaction of trifluoroethanol with peptides and proteins”. Em: *International Journal of Peptide and Protein Research* 48.4 (2009), pp. 328–336. DOI: <https://doi.org/10.1111/j.1399-3011.1996.tb00849.x>.
72. D. Roccatano et al. “Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study”. Em: *Proceedings of the National Academy of Sciences* 99.19 (2002), pp. 12179–12184. DOI: <https://doi.org/10.1073/pnas.182199699>.
73. N. Rezaei-Ghaleh, M. Amininasab e M. Nemat-Gorgani. “Conformational changes of alpha-chymotrypsin in a fibrillation-promoting condition: a molecular dynamics study”. Em: *Biophysical Journal* 95.9 (2008), pp. 4139–4147. DOI: <https://doi.org/10.1529/biophysj.108.132407>.
74. M. Buck. “Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins”. Em: *Quarterly Reviews of Biophysics* 31.3 (1998), pp. 297–355. DOI: <https://doi.org/10.1017/S003358359800345X>.
75. A. Kentsis e T. R. Sosnick. “Trifluoroethanol promotes helix formation by destabilizing backbone exposure: desolvation rather than native hydrogen bonding defines the kinetic

- pathway of dimeric coiled coil folding". Em: *Biochemistry* 37.41 (1998), pp. 14613–14622. DOI: <https://doi.org/10.1021/bi981641y>.
76. M. Fioroni et al. "Solvation Phenomena of a Tetrapeptide in Water/Trifluoroethanol and Water/Ethanol Mixtures: A Diffusion NMR, Intermolecular NOE, and Molecular Dynamics Study". Em: *Journal of the American Chemical Society* 124.26 (2002), pp. 7737–7744. DOI: <https://doi.org/10.1021/ja0259335>.
77. M. D. Díaz et al. "Evidence of complete hydrophobic coating of bombesin by trifluoroethanol in aqueous solution: an NMR spectroscopic and molecular dynamics study". Em: *Chemistry* 8.7 (2002), pp. 1663–1669. DOI: [https://doi.org/10.1002/1521-3765\(20020402\)8:7%3C1663::AID-CHEM1663%3E3.0.CO;2-P](https://doi.org/10.1002/1521-3765(20020402)8:7%3C1663::AID-CHEM1663%3E3.0.CO;2-P).
78. R. Walgers, T. C. Lee e A. Cammers-Goodwin. "An Indirect Chaotropic Mechanism for the Stabilization of Helix Conformation of Peptides in Aqueous Trifluoroethanol and Hexafluoro-2-propanol". Em: *Journal of the American Chemical Society* 120.20 (1998), pp. 5073–5079. DOI: <https://doi.org/10.1021/ja973552z>.
79. J. A. Vila, D. R. Ripoll e H. A. Scheraga. "Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides". Em: *Proceedings of the National Academy of Sciences* 97.24 (2000), pp. 13075–13079. DOI: <https://doi.org/10.1073/pnas.240455797>.
80. Y.-P. Pang. "FF12MC: A revised AMBER forcefield and new protein simulation protocol". Em: *Proteins* 84.10 (2016), pp. 1490–1516. DOI: <https://doi.org/10.1002/prot.25094>.
81. V. Muñoz e L. Serrano. "Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence". Em: *Journal of Molecular Biology* 245.3 (1995), pp. 297–308. DOI: <https://doi.org/10.1006/jmbi.1994.0024>.
82. S. A. Celinski e J. M. Scholtz. "Osmolyte effects on helix formation in peptides and the stability of coiled-coils". Em: *Protein Science* 11.8 (2002), pp. 2048–2051. DOI: <https://doi.org/10.1110/ps.0211702>.
83. E. A. Ploetz, S. Karunaweera e P. E. Smith. "Kirkwood-Buff-Derived Force Field for Peptides and Proteins: Applications of KBFF20". Em: *Journal of Chemical Theory and Computation* 17.5 (2021), pp. 2991–3009. DOI: <https://doi.org/10.1021/acs.jctc.1c00076>.

84. W. Shalongo, L. Dugad e E. Stellwagen. "Distribution of helicity within the model peptide acetyl(AAQAA)₃amide". Em: *Journal of the American Chemical Society* 116.18 (1994), pp. 8288–8293. DOI: <https://doi.org/10.1021/ja00097a039>.
85. J. Huang e J. MacKerell A. D. "Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)₃ peptide". Em: *Biophysical Journal* 107.4 (2014), pp. 991–997. DOI: <http://dx.doi.org/10.1016/j.bpj.2014.06.038>.
86. H. Ohgi et al. "Two different regimes in alcohol-induced coil-helix transition: effects of 2,2,2-trifluoroethanol on proteins being either independent of or enhanced by solvent structural fluctuations". Em: *Physical Chemistry Chemical Physics* 23.10 (2021), pp. 5760–5772. DOI: <https://doi.org/10.1039/D0CP05103A>.
87. J. Vymětal, L. Bednářová e J. Vondrášek. "Effect of TFE on the Helical Content of AK17 and HAL-1 Peptides: Theoretical Insights into the Mechanism of Helix Stabilization". Em: *The Journal of Physical Chemistry B* 120.6 (2016), pp. 1048–1059. DOI: <https://doi.org/10.1021/acs.jpcc.5b11228>.
88. M. Vincenzi, F. A. Mercurio e M. Leone. "About TFE: Old and New Findings". Em: *Current Protein and Peptide Science* 20.5 (2019), pp. 425–451. DOI: <https://doi.org/10.2174/1389203720666190214152439>.
89. M. Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". Em: *SoftwareX* 1-2 (2015), pp. 19–25. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>.
90. M. Bonomi et al. "PLUMED: A portable plugin for free-energy calculations with molecular dynamics". Em: *Computer Physics Communications* 180.10 (2009), pp. 1961–1972. DOI: <https://doi.org/10.1016/j.cpc.2009.05.011>.
91. M. Parrinello e A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". Em: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190. DOI: <https://doi.org/10.1063/1.328693>.
92. G. Bussi, D. Donadio e M. Parrinello. "Canonical sampling through velocity rescaling". Em: *The Journal of Chemical Physics* 126.1 (2007), p. 014101. DOI: <https://doi.org/10.1063/1.2408420>.

93. T. Darden, D. York e L. Pedersen. "Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems". Em: *The Journal of Chemical Physics* 98 (1993), pp. 10089–10092. DOI: <http://dx.doi.org/10.1063/1.464397>.
94. B. Hess et al. "LINCS: A linear constraint solver for molecular simulations". Em: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H).
95. W. Humphrey, A. Dalke e K. Schulten. "VMD: Visual molecular dynamics". Em: *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
96. J. M. Martínez e L. Martínez. "Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking". Em: *Journal of Computational Chemistry* 24.7 (2003), pp. 819–825. DOI: <https://doi.org/10.1002/jcc.10216>.
97. L. Martínez et al. "PACKMOL: a package for building initial configurations for molecular dynamics simulations". Em: *Journal of Computational Chemistry* 30.13 (2009), pp. 2157–2164. DOI: <https://doi.org/10.1002/jcc.21224>.
98. J. L. F. Abascal e C. Vega. "A general purpose model for the condensed phases of water: TIP4P/2005". Em: *The Journal of Chemical Physics* 123.23 (2005), p. 234505. DOI: <https://doi.org/10.1063/1.2121687>.
99. R. B. Best e J. Mittal. "Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse". Em: *The Journal of Physical Chemistry B* 114.46 (2010), pp. 14916–14923. DOI: <https://doi.org/10.1021/jp108618d>.
100. J. Vymětal e J. Vondrášek. "Parametrization of 2,2,2-trifluoroethanol based on the generalized AMBER force field provides realistic agreement between experimental and calculated properties of pure liquid as well as water-mixed solutions". Em: *The Journal of Physical Chemistry B* 118.45 (2014), pp. 10390–10404. DOI: <https://doi.org/10.1021/jp505861b>.
101. P. Robustelli, S. Piana e D. E. Shaw. "Developing a molecular dynamics force field for both folded and disordered protein states". Em: *Proceedings of the National Academy of Sciences* 115.21 (2018), E4758–E4766. DOI: <https://doi.org/10.1073/pnas.1800690115>.

102. W. Kabsch e C. Sander. "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". Em: *Biopolymers* 22.12 (1983), pp. 2577–2637. DOI: <https://doi.org/10.1002/bip.360221211>.
103. G. Nagy et al. "SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures". Em: *Journal of Chemical Theory and Computation* 15.9 (2019), pp. 5087–5102. DOI: <https://doi.org/10.1021/acs.jctc.9b00203>.
104. G. Nagy e C. Oostenbrink. "Dihedral-based segment identification and classification of biopolymers I: proteins". Em: *Journal of Chemical Information and Modeling* 54.1 (2014), pp. 266–277. DOI: <https://doi.org/10.1021/ci400541d>.
105. B. M. Baynes e B. L. Trout. "Proteins in Mixed Solvents: A Molecular-Level Perspective". Em: *The Journal of Physical Chemistry B* 107.50 (2003), pp. 14058–14067. DOI: <https://doi.org/10.1021/jp0363996>.
106. A. Jasanoff e A. R. Fersht. "Quantitative determination of helical propensities from trifluoroethanol titration curves". Em: *Biochemistry* 33.8 (1994), pp. 2129–2135. DOI: <https://doi.org/10.1021/bi00174a020>.
107. P. Luo e R. L. Baldwin. "Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water". Em: *Biochemistry* 36.27 (1997), pp. 8413–8421. DOI: <https://doi.org/10.1021/bi9707133>.
108. A. Chakrabarty et al. "Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities". Em: *Biochemistry* 32.21 (1993), pp. 5560–5565. DOI: <https://doi.org/10.1021/bi00072a010>.
109. Y. Fezoui e D. B. Teplow. "Kinetic studies of amyloid beta-protein fibril assembly. Differential effects of alpha-helix stabilization". Em: *Journal of Biological Chemistry* 277.40 (2002), pp. 36948–36954. DOI: <https://doi.org/10.1074/jbc.M204168200>.
110. D. Hamada et al. "High helical propensity of the peptide fragments derived from beta-lactoglobulin, a predominantly beta-sheet protein". Em: *Journal of Molecular Biology* 254.4 (1995), pp. 737–746. DOI: <https://doi.org/10.1006/jmbi.1995.0651>.
111. M. Ndao et al. "Probing the self-association, intermolecular contacts, and folding propensity of amelogenin". Em: *Protein Science* 20.4 (2011), pp. 724–734. DOI: <https://doi.org/10.1002/pro.603>.

112. C. M. Othon et al. "Solvation in protein (un)folding of melittin tetramer-monomer transition". Em: *Proceedings of the National Academy of Sciences* 106.31 (2009), pp. 12593–12598. DOI: <https://doi.org/10.1073/pnas.0905967106>.
113. C. A. Rohl, A. Chakrabartty e R. L. Baldwin. "Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol". Em: *Protein Science* 5.12 (1996), pp. 2623–2637. DOI: <https://doi.org/10.1002/pro.5560051225>.
114. B. Chaubey et al. "Assessment of the Role of 2,2,2-Trifluoroethanol Solvent Dynamics in Inducing Conformational Transitions in Melittin: An Approach with Solvent F Low-Field NMR Relaxation and Overhauser Dynamic Nuclear Polarization Studies". Em: *The Journal of Physical Chemistry B* 124.28 (2020), pp. 5993–6003. DOI: <https://doi.org/10.1021/acs.jpcc.0c03544>.
115. A. Cammers-Goodwin et al. "Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol". Em: *Journal of the American Chemical Society* 118.13 (1996), pp. 3082–3090. DOI: <https://doi.org/10.1021/ja952900z>.
116. R. M. Culik et al. "Experimental validation of the role of trifluoroethanol as a nanocrowder". Em: *The Journal of Physical Chemistry B* 118.39 (2014), pp. 11455–11461. DOI: <https://doi.org/10.1021/jp508056w>.
117. A. R. V. Buuren e H. J. Berendsen. "Molecular dynamics simulation of the stability of a 22-residue alpha-helix in water and 30% trifluoroethanol". Em: *Biopolymers* 33.8 (1993), pp. 1159–1166. DOI: <https://doi.org/10.1002/bip.360330802>.
118. W. Zheng et al. "Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment". Em: *Journal of the American Chemical Society* 138.36 (2016), pp. 11702–11713. DOI: <https://doi.org/10.1021/jacs.6b05443>.
119. J. A. Vila, D. R. Ripoll e H. A. Scheraga. "Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides". Em: *Proceedings of the National Academy of Sciences* 97.24 (2000), pp. 13075–13079. DOI: <https://doi.org/10.1073/pnas.240455797>.
120. P. G. Wolynes. "Latest folding game results: protein A barely frustrates computationalists". Em: *Proceedings of the National Academy of Sciences* 101.18 (2004), pp. 6837–6838. DOI: <https://doi.org/10.1073/pnas.0402034101>.

121. T. Otsu et al. "Highly Heterogeneous Nature of the Native and Unfolded States of the B Domain of Protein A Revealed by Two-Dimensional Fluorescence Lifetime Correlation Spectroscopy". Em: *The Journal of Physical Chemistry B* 121.22 (2017), pp. 5463–5473. DOI: <https://doi.org/10.1021/acs.jpccb.7b00546>.
122. C. M. Davis, A. K. Cooper e R. B. Dyer. "Fast helix formation in the B domain of protein A revealed by site-specific infrared probes". Em: *Biochemistry* 54.9 (2015), pp. 1758–1766. DOI: <https://doi.org/10.1021/acs.biochem.5b00037>.
123. A. E. García e J. N. Onuchic. "Folding a protein in a computer: an atomic description of the folding/unfolding of protein A". Em: *Proceedings of the National Academy of Sciences* 100.24 (2003), pp. 13898–13903. DOI: <https://doi.org/10.1073/pnas.2335541100>.
124. J. Oliveira A. B. et al. "Distinguishing Biomolecular Pathways and Metastable States". Em: *Journal of Chemical Theory and Computation* 15.11 (2019), pp. 6482–6490. DOI: <https://doi.org/10.1021/acs.jctc.9b00704>.
125. D. O. Alonso e V. Daggett. "Staphylococcal protein A: unfolding pathways, unfolded states, and differences between the B and E domains". Em: *Proceedings of the National Academy of Sciences* 97.1 (2000), pp. 133–138. DOI: <https://doi.org/10.1073/pnas.97.1.133>.
126. S. Sato et al. "Testing protein-folding simulations by experiment: B domain of protein A". Em: *Proceedings of the National Academy of Sciences* 101.18 (2004), pp. 6952–6956. DOI: <https://doi.org/10.1073/pnas.0401396101>.
127. S. Sato, T. L. Religa e A. R. Fersht. "Phi-analysis of the folding of the B domain of protein A using multiple optical probes". Em: *Journal of Molecular Biology* 360.4 (2006), pp. 850–864. DOI: <https://doi.org/10.1016/j.jmb.2006.05.051>.
128. H. Lammert, A. Schug e J. N. Onuchic. "Robustness and generalization of structure-based models for protein folding and function". Em: *Proteins* 77.4 (2009), pp. 881–891. DOI: <https://doi.org/10.1002/prot.22511>.
129. J. Noel et al. "Mirror images as naturally competing conformations in protein folding". Em: *The Journal of Physical Chemistry B* 116.23 (2012), pp. 6880–6888. DOI: <https://doi.org/10.1021/jp212623d>.
130. P. Whitford et al. "An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields". Em: *Proteins* 75.2 (2009), pp. 430–441. DOI: <https://doi.org/10.1002/prot.22253>.

131. J. K. Myers e T. G. Oas. "Preorganized secondary structure as an important determinant of fast protein folding". Em: *Nature Structural Molecular Biology* 8 (2001), pp. 552–558. DOI: <https://doi.org/10.1038/88626>.
132. Y. Bai et al. "Absence of a stable intermediate on the folding pathway of protein A". Em: *Protein Science* 6.7 (1997), pp. 1449–1457. DOI: <https://doi.org/10.1002/pro.5560060709>.
133. S. Yanaka et al. "The B domain of protein A retains residual structures in 6 M guanidinium chloride as revealed by hydrogen/deuterium-exchange NMR spectroscopy". Em: *Protein Science* 32.3 (2023), e4569. DOI: <https://doi.org/10.1002/pro.4569>.
134. S. Sato e A. R. Fersht. "Searching for multiple folding pathways of a nearly symmetrical protein: temperature dependent phi-value analysis of the B domain of protein A". Em: *Journal of Molecular Biology* 372.1 (2007), pp. 254–267. DOI: <https://doi.org/10.1016/j.jmb.2007.06.043>.
135. M. C. Baxa, K. F. Freed e T. R. Sosnick. "Quantifying the structural requirements of the folding transition state of protein A and other systems". Em: *Journal of Molecular Biology* 381.5 (2008), pp. 1362–1381. DOI: <https://doi.org/10.1016/j.jmb.2008.06.067>.
136. Gouda, H. et al. "Three-dimensional solution structure of the B domain of staphylococcal protein A: comparisons of the solution and crystal structures". Em: *Biochemistry* 31.40 (1992), pp. 9665–9672. DOI: <https://doi.org/10.1021/bi00155a020>.
137. J. Oliveira A. B. et al. "Distinguishing Biomolecular Pathways and Metastable States". Em: *Journal of Chemical Theory and Computation* 15.11 (2019), pp. 6482–6490. DOI: <https://doi.org/10.1021/acs.jctc.9b00704>.
138. J. K. Noel, P. C. Whitford e J. N. Onuchic. "The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function". Em: *The Journal of Physical Chemistry B* 116.29 (2012), pp. 8692–8702. DOI: <https://doi.org/10.1021/jp300852d>.
139. V. Sobolev et al. "Automated analysis of interatomic contacts in proteins". Em: *Bioinformatics* 15.4 (1999), pp. 327–332. DOI: <https://doi.org/10.1093/bioinformatics/15.4.327>.
140. J. K. Noel et al. "SMOG@ctbp: simplified deployment of structure-based models in GROMACS". Em: *Nucleic Acids Research* 38.suppl_2 (2010), W657–61. DOI: <https://doi.org/10.1093/nar/gkq498>.

141. B. Hess et al. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation". Em: *Journal of Chemical Theory and Computation* 4.3 (2008), pp. 435–447. DOI: <https://doi.org/10.1021/ct700301q>.
142. S. Kumar et al. "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method". Em: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. DOI: <https://doi.org/10.1002/jcc.540130812>.
143. J. K. Noel et al. "SMOG 2: A Versatile Software Package for Generating Structure-Based Models". Em: *PLOS Computational Biology* 12.3 (2016), e1004794. DOI: <https://doi.org/10.1371/journal.pcbi.1004794>.
144. A. B. Oliveira Junior et al. "Exploring Energy Landscapes of Intrinsically Disordered Proteins: Insights into Functional Mechanisms". Em: *Journal of Chemical Theory and Computation* 17.5 (2021), pp. 3178–3187. DOI: <https://doi.org/10.1021/acs.jctc.1c00027>.
145. R. V. R. Dias et al. "Unveiling Metastable Ensembles of GRB2 and the Relevance of Interdomain Communication during Folding". Em: *Journal of Chemical Information and Modeling* 63.20 (2023), pp. 6344–6353. DOI: <https://doi.org/10.1021/acs.jcim.3c00955>.
146. M. N. Sanches et al. "Resolving the fine structure in the energy landscapes of repeat proteins". Em: *QRB Discovery* 3 (2022), e7. DOI: <https://doi.org/10.1017/qrd.2022.4>.
147. R. G. Viegas et al. "Characterizing the Folding Transition-State Ensembles in the Energy Landscape of an RNA Tetraloop". Em: *Journal of Chemical Information and Modeling* 63.17 (2023), pp. 5641–5649. DOI: <https://doi.org/10.1021/acs.jcim.3c00426>.
148. F. B. da Silva et al. "Exploring the folding landscape of leptin: Insights into threading pathways". Em: *Journal of Structural Biology* 216.1 (2023), p. 108054. DOI: <https://doi.org/10.1016/j.jsb.2023.108054>.
149. R. G. Viegas et al. "ELViM: Exploring Biomolecular Energy Landscapes through Multidimensional Visualization". Em: *Journal of Chemical Information and Modeling* 64.8 (2024), pp. 3443–3450. DOI: <https://doi.org/10.1021/acs.jcim.4c00034>.
150. C. Hardin et al. "Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins". Em: *Proceedings of the National Academy of Sciences* 100.4 (2003), pp. 1679–1684. DOI: <https://doi.org/10.1073/pnas.252753899>.

151. P. Rotkiewicz e J. Skolnick. "Fast procedure for reconstruction of full-atom protein models from reduced representations". Em: *Journal of Computational Chemistry* 29.9 (2008), pp. 1460–1465. DOI: <https://doi.org/10.1002/jcc.20906>.
152. A. D. MacKerell et al. "All-atom empirical potential for molecular modeling and dynamics studies of proteins". Em: *The Journal of Physical Chemistry B* 102.18 (1998), pp. 3586–3616. DOI: <https://doi.org/10.1021/jp973084f>.
153. W. L. Jorgensen et al. "Comparison of simple potential functions for simulating liquid water". Em: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: <https://doi.org/10.1063/1.445869>.
154. D. Van Der Spoel et al. "GROMACS: fast, flexible, and free". Em: *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718. DOI: <https://doi.org/10.1002/jcc.20291>.
155. H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". Em: *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690. DOI: <https://doi.org/10.1063/1.448118>.
156. L. Martínez e M. Matthes. *ProteinSecondaryStructures.jl: This package parses STRIDE and DSSP secondary structure prediction outputs, to make them convenient to use from Julia, particularly for the analysis of MD simulations*. Version 0.3.2. 2023. DOI: 10.5281/zenodo.8192529. URL: <https://doi.org/10.5281/zenodo.8192529>.
157. K. Itoh e M. Sasai. "Flexibly varying folding mechanism of a nearly symmetrical protein: B domain of protein A". Em: *Proceedings of the National Academy of Sciences* 103.19 (2006), pp. 7298–7303. DOI: <https://doi.org/10.1073/pnas.0510324103>.
158. L. Martínez. "ComplexMixtures.jl: Investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective". Em: *Journal of Molecular Liquids* 347 (2022), p. 117945. DOI: <https://doi.org/10.1016/j.molliq.2021.117945>.
159. Z. Guo, 3. Brooks C. L. e E. M. Boczko. "Exploring the folding free energy surface of a three-helix bundle protein". Em: *Proceedings of the National Academy of Sciences* 94.19 (1997), pp. 10161–10166. DOI: <https://doi.org/10.1073/pnas.94.19.10161>.
160. T. Hasan et al. "Osmolytes in vaccine production, flocculation and storage: a critical review". Em: *Human Vaccines Immunotherapeutics* 15.2 (2019), pp. 514–525. DOI: <https://doi.org/10.1080/21645515.2018.1526585>.

161. W. Wang e S. Ohtake. "Science and art of protein formulation development". Em: *International Journal of Pharmaceutics* 568 (2019), p. 118505. DOI: <https://doi.org/10.1016/j.ijpharm.2019.118505>.
162. G. Rabbani e I. Choi. "Roles of osmolytes in protein folding and aggregation in cells and their biotechnological applications". Em: *International Journal of Biological Macromolecules* 109 (2018), pp. 483–491. DOI: <https://doi.org/10.1016/j.ijbiomac.2017.12.100>.
163. M. T. Velasquez et al. "Trimethylamine N-Oxide: The Good, the Bad and the Unknown". Em: *Toxins (Basel)* 8.11 (2016). DOI: <https://doi.org/10.3390/toxins8110326>.
164. M. M. Boob et al. "TMAO: Protecting proteins from feeling the heat". Em: *Biophysical Journal* 122.7 (2023), pp. 1414–1422. DOI: <https://doi.org/10.1016/j.bpj.2023.03.008>.
165. M. Pepelnjak et al. "In situ analysis of osmolyte mechanisms of proteome thermal stabilization". Em: *Nature Chemical Biology* 20.8 (2024), pp. 1053–1065. DOI: <https://doi.org/10.1038/s41589-024-01568-7>.
166. D. R. Canchi et al. "Molecular mechanism for the preferential exclusion of TMAO from protein surfaces". Em: *The Journal of Physical Chemistry B* 116.40 (2012), pp. 12095–12104. DOI: <https://doi.org/10.1021/jp304298c>.
167. T. Y. Lin e S. N. Timasheff. "Why do some organisms use a urea-methylamine mixture as osmolyte? Thermodynamic compensation of urea and trimethylamine N-oxide interactions with protein". Em: *Biochemistry* 33.42 (1994), pp. 12695–12701. DOI: <https://doi.org/10.1021/bi00208a021>.
168. B. J. Bennion e V. Daggett. "The molecular basis for the chemical denaturation of proteins by urea". Em: *Proceedings of the National Academy of Sciences* 100.9 (2003), pp. 5142–5147. DOI: <https://doi.org/10.1073/pnas.0930122100>.
169. S. Khan et al. "Osmolytes: Wonder molecules to combat protein misfolding against stress conditions". Em: *International Journal of Biological Macromolecules* 234 (2023), p. 123662. DOI: <https://doi.org/10.1016/j.ijbiomac.2023.123662>.
170. D. W. Bolen e I. V. Baskakov. "The osmophobic effect: natural selection of a thermodynamic force in protein folding". Em: *Journal of Molecular Biology* 310.5 (2001), pp. 955–963. DOI: <https://doi.org/10.1006/jmbi.2001.4819>.
171. X. Peng et al. "Prediction and Validation of a Protein's Free Energy Surface Using Hydrogen Exchange and (Importantly) Its Denaturant Dependence". Em: *Journal of Chemical Theory*

- and Computation* 18.1 (2022), pp. 550–561. DOI: <https://doi.org/10.1021/acs.jctc.1c00960>.
172. S. N. Timasheff. “Solvent stabilization of protein structure”. Em: *Methods in Molecular Biology* 40 (1995), pp. 253–269. DOI: <https://doi.org/10.1385/0-89603-301-5:253>.
173. V. Piccoli e L. Martínez. “Ionic liquid solvation of proteins in native and denatured states”. Em: *Journal of Molecular Liquids* 363 (2022), p. 119953.
174. A. F. Pereira, V. Piccoli e L. Martínez. “Trifluoroethanol direct interactions with protein backbones destabilize α -helices”. Em: *Journal of Molecular Liquids* 365 (2022), p. 120209. DOI: <https://doi.org/10.1016/j.molliq.2022.119953>.
175. A. E. Rydeen, E. M. Brustad e G. J. Pielak. “Osmolytes and protein-protein interactions”. Em: *Journal of the American Chemical Society* 140.24 (2018), pp. 7441–7444. DOI: <https://doi.org/10.1021/jacs.8b03903>.
176. W. Xu, S. C. Harrison e M. J. Eck. “Three-dimensional structure of the tyrosine kinase c-Src”. Em: *Nature* 385.6617 (1997), pp. 595–602. DOI: <https://doi.org/10.1038/385595a0>.
177. A. F. Pereira e L. Martínez. “Helical Content Correlations and Hydration Structures of the Folding Ensemble of the B Domain of Protein A”. Em: *Journal of Chemical Information and Modeling* 64.9 (2024), pp. 3350–3359. DOI: <https://doi.org/10.1021/acs.jcim.3c01822>.
178. R. Joosten et al. “A series of PDB related databases for everyday needs”. Em: *Nucleic Acids Research* 39.suppl_1 (2011), pp. D411–9. DOI: <https://doi.org/10.1093/nar/gkq1105>.
179. K. Vanommeslaeghe et al. “CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields”. Em: *Journal of Computational Chemistry* 31.4 (2009), pp. 671–690. DOI: <https://doi.org/10.1002/jcc.21367>.
180. W. Yu et al. “Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations”. Em: *Journal of Computational Chemistry* 33.31 (2012), pp. 2451–2468. DOI: <https://doi.org/10.1002/jcc.23067>.
181. S. Weerasinghe e P. E. Smith. “A KirkwoodBuff derived force field for mixtures of urea and water”. Em: *The Journal of Physical Chemistry B* 107.16 (2003), pp. 3891–3898. DOI: <https://doi.org/10.1021/jp022049s>.

182. E. Schneck, D. Horinek e R. R. Netz. "Insight into the molecular mechanisms of protein stabilizing osmolytes from global force-field variations". Em: *The Journal of Physical Chemistry B* 117.28 (2013), pp. 8310–8321. DOI: <https://doi.org/10.1021/jp400790f>.
183. P. Ganguly et al. "Cosolvent Exclusion Drives Protein Stability in Trimethylamine -Oxide and Betaine Solutions". Em: *The Journal of Physical Chemistry Letters* 13.34 (2022), pp. 7980–7986. DOI: <https://doi.org/10.1021/acs.jpcllett.2c01692>.
184. J. N. Onuchic et al. "The energy landscape theory of protein folding: insights into folding mechanisms and scenarios". Em: *Advances in Protein Chemistry* 53 (2000), pp. 87–152. DOI: [https://doi.org/10.1016/S0065-3233\(00\)53003-4](https://doi.org/10.1016/S0065-3233(00)53003-4).
185. C. Demakis, M. C. Childers e V. Daggett. "Conserved patterns and interactions in the unfolding transition state across SH3 domain structural homologues". Em: *Protein Science* 30.2 (2021), pp. 391–407. DOI: <https://doi.org/10.1002/pro.3998>.
186. J. M. Borreguero et al. "Multiple folding pathways of the SH3 domain". Em: *Biophysical Journal* 87.1 (2004), pp. 521–533. DOI: <https://doi.org/10.1529/biophysj.104.039529>.
187. J. M. Borreguero et al. "Thermodynamics and folding kinetics analysis of the SH3 domain from discrete molecular dynamics". Em: *Journal of Molecular Biology* 318.3 (2002), pp. 863–876. DOI: [https://doi.org/10.1016/S0022-2836\(02\)00136-5](https://doi.org/10.1016/S0022-2836(02)00136-5).
188. J. Gsponer e A. Caflisch. "Molecular dynamics simulations of protein folding from the transition state". Em: *Proceedings of the National Academy of Sciences* 99.10 (2002), pp. 6719–6724. DOI: <https://doi.org/10.1073/pnas.092686399>.
189. J. Tsai, M. Levitt e D. Baker. "Hierarchy of structure loss in MD simulations of src SH3 domain unfolding". Em: *Journal of Molecular Biology* 291.1 (1999), pp. 215–225. DOI: <https://doi.org/10.1006/jmbi.1999.2949>.
190. M. Khalili, A. Liwo e H. A. Scheraga. "Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains". Em: *Journal of Molecular Biology* 355.3 (2006), pp. 536–547. DOI: <https://doi.org/10.1016/j.jmb.2005.10.056>.
191. Q. Zou et al. "The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea". Em: *Journal of the American Chemical Society* 124.7 (2002), pp. 1192–1202. DOI: <https://doi.org/10.1021/ja004206b>.

192. L. Hua et al. "Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding". Em: *Proceedings of the National Academy of Sciences* 105.44 (2008), pp. 16928–16933. DOI: <https://doi.org/10.1073/pnas.0808427105>.
193. Y. Okuno et al. "Atomic view of cosolute-induced protein denaturation probed by NMR solvent paramagnetic relaxation enhancement". Em: *Proceedings of the National Academy of Sciences* 118.34 (2021). DOI: <https://doi.org/10.1073/pnas.2112021118>.
194. E. S. Courtenay et al. "Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of 'osmotic stress' experiments in vitro". Em: *Biochemistry* 39.15 (2000), pp. 4455–4471. DOI: <https://doi.org/10.1021/bi9928871>.
195. A. Wang e D. W. Bolen. "A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation". Em: *Biochemistry* 36.30 (1997), pp. 9101–9108. DOI: <https://doi.org/10.1021/bi970247h>.
196. S. Roy e B. Bagchi. "Comparative study of protein unfolding in aqueous urea and dimethyl sulfoxide solutions: surface polarity, solvent specificity, and sequence of secondary structure melting". Em: *The Journal of Physical Chemistry B* 118.21 (2014), pp. 5691–5697. DOI: <https://doi.org/10.1021/jp5037348>.
197. S. Parui e B. Jana. "Relative Solvent Exposure of the Alpha-Helix and Beta-Sheet in Water Determines the Initial Stages of Urea and Guanidinium Chloride-Induced Denaturation of Alpha/Beta Proteins". Em: *The Journal of Physical Chemistry B* 123.42 (2019), pp. 8889–8900. DOI: <https://doi.org/10.1021/acs.jpccb.9b06859>.
198. L. L. Porter e L. L. Looger. "Extant fold-switching proteins are widespread". Em: *Proceedings of the National Academy of Sciences* 115.23 (2018), pp. 5968–5973. DOI: <https://doi.org/10.1073/pnas.1800168115>.
199. A. G. Murzin. "Metamorphic proteins". Em: *Science* 320.5884 (2008), pp. 1725–1726. DOI: <https://doi.org/10.1126/science.1158868>.
200. A. F. Dishman e B. F. Volkman. "Unfolding the Mysteries of Protein Metamorphosis". Em: *ACS Chemical Biology* 13.6 (2018), pp. 1438–1446. DOI: <https://doi.org/10.1021/acscchembio.8b00276>.
201. A. K. Kim e L. L. Porter. "Functional and Regulatory Roles of Fold-Switching Proteins". Em: *Structure* 29.1 (2021), pp. 6–14. DOI: <https://doi.org/10.1016/j.str.2020.10.006>.

202. D. Giganti et al. "Secondary structure reshuffling modulates glycosyltransferase function at the membrane". Em: *Nature Chemical Biology* 11.1 (2015), pp. 16–18. DOI: <https://doi.org/10.1038/nchembio.1694>.
203. S. Mishra, L. L. Looger e L. L. Porter. "A sequence-based method for predicting extant fold switchers that undergo α -helix β -strand transitions". Em: *Biopolymers* 112.10 (2021), e23471. DOI: <https://doi.org/10.1002/bip.23471>.
204. B. Burmann et al. "An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor". Em: *Cell* 150.2 (2012), pp. 291–303. DOI: <http://dx.doi.org/10.1016/j.cell.2012.05.042>.
205. M. Podobnik et al. "Crystal structure of an invertebrate cytolysin pore reveals unique properties and mechanism of assembly". Em: *Nature Communications* 7.1 (2016), p. 11598. DOI: <https://doi.org/10.1038/ncomms11598>.
206. N. Sung et al. "Mitochondrial Hsp90 is a ligand-activated molecular chaperone coupling ATP binding to dimer closure through a coiled-coil intermediate". Em: *Proceedings of the National Academy of Sciences* 113.11 (2016), pp. 2952–2957. DOI: <https://doi.org/10.1073/pnas.1516167113>.
207. Y. Liu e D. Eisenberg. "3D domain swapping: as domains continue to swap". Em: *Protein Science* 11.6 (2002), pp. 1285–1299. DOI: <https://doi.org/10.1110/ps.0201402>.
208. Q. Cao et al. "Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils". Em: *Nature structural molecular biology* 27.7 (2020), pp. 653–659. DOI: <https://doi.org/10.1038/s41594-020-0435-3>.
209. R. Tuinstra et al. "Interconversion between two unrelated protein folds in the lymphotactin native state". Em: *Proceedings of the National Academy of Sciences* 105.13 (2008), pp. 5057–5062. DOI: <https://doi.org/10.1073/pnas.0709518105>.
210. X. Luo e H. Yu. "Protein metamorphosis: the two-state behavior of Mad2". Em: *Structure* 16.11 (2008), pp. 1616–1625. DOI: <https://doi.org/10.1016/j.str.2008.10.002>.
211. X. Luo et al. "Structure of the Mad2 spindle assembly checkpoint protein and its interaction with Cdc20". Em: *Nature Structural Molecular Biology* 7.3 (2000), pp. 224–229. DOI: <https://doi.org/10.1038/73338>.

212. X. Luo et al. "The Mad2 spindle checkpoint protein has two distinct natively folded states". Em: *Nature Structural Molecular Biology* 11.4 (2004), pp. 338–345. DOI: <https://doi.org/10.1038/nsmb748>.
213. Y. Gu, A. Desai e K. D. Corbett. "Evolutionary dynamics and molecular mechanisms of HORMA domain protein signaling". Em: *Annual Review of Biochemistry* 91.1 (2022), pp. 541–569. DOI: <https://doi.org/10.1146/annurev-biochem-090920-103246>.
214. M. Mapelli et al. "The Mad2 conformational dimer: structure and implications for the spindle assembly checkpoint". Em: *Cell* 131.4 (2007), pp. 730–743. DOI: <https://doi.org/10.1016/j.cell.2007.08.049>.
215. X. Hu et al. "Direct Observation and Real-Time Tracking of an Extraordinarily Stable Folding Intermediate in Mitotic Arrest Deficient Protein 2 Folding by Single-Molecule Fluorescence Resonance Energy Transfer". Em: *The Journal of Physical Chemistry Letters* 14.3 (2023), pp. 763–769. DOI: <https://doi.org/10.1021/acs.jpcllett.2c03181>.
216. S. Jain e A. Sekhar. "Transient excited states of the metamorphic protein Mad2 and their implications for function". Em: *Proteins* 93.1 (2024), pp. 302–319. DOI: <https://doi.org/10.1002/prot.26667>.
217. C. Yu et al. "Molecular mechanism of Mad2 conformational conversion promoted by the Mad2-interaction motif of Cdc20". Em: *Protein Science* 34.4 (2024), e70099. DOI: <https://doi.org/10.1002/pro.70099>.
218. E. A. Chen e L. L. Porter. "SSDraw: Software for generating comparative protein secondary structure diagrams". Em: *Protein Science* 32.12 (2023), e4836. DOI: <https://doi.org/10.1002/pro.4836>.
219. L. Martínez, R. Andreani e J. M. Martínez. "Convergent algorithms for protein structural alignment". Em: *BMC Bioinformatics* 8 (2007), pp. 1–15. DOI: <https://doi.org/10.1186/1471-2105-8-306>.
220. R. Andreani et al. "Low Order-Value Optimization and applications". Em: *Journal of Global Optimization* 43 (2008), pp. 1–22. DOI: <https://doi.org/10.1007/s10898-008-9280-3>.
221. A. de Oliveira Jr et al. "SMOG 2 and OpenSMOG: Extending the limits of structure-based models". Em: *Protein Science* 31.1 (2022), pp. 158–172. DOI: <https://doi.org/10.1002/pro.4209>.

222. I. Retamal-Farfán et al. “Exploring the structural acrobatics of fold-switching proteins using simplified structure-based models”. Em: *Biophysical Reviews* 15.4 (2023), pp. 787–799. DOI: <https://doi.org/10.1007/s12551-023-01087-0>.
223. P. Eastman et al. “OpenMM 7: Rapid development of high performance algorithms for molecular dynamics”. Em: *PLOS Computational Biology* 13.7 (2017), e1005659. DOI: <https://doi.org/10.1371/journal.pcbi.1005659>.
224. L. S. Johnson, S. R. Eddy e E. Portugaly. “Hidden Markov model speed heuristic and iterative HMM search procedure”. Em: *BMC Bioinformatics* 11 (2010), pp. 1–8. DOI: <https://doi.org/10.1186/1471-2105-11-431>.
225. U. Consortium. “UniProt: the universal protein knowledgebase in 2021”. Em: *Nucleic Acids Research* 49.D1 (2021), pp. D480–D489. DOI: <https://doi.org/10.1093/nar/gkaa1100>.
226. J. W. Schafer e L. L. Porter. “Evolutionary selection of proteins with two folds”. Em: *Nature Communications* 14 (2023), p. 5478. DOI: <https://doi.org/10.1038/s41467-023-41237-2>.
227. C. Baldassi et al. “Fast and accurate multivariate Gaussian modeling of protein families: predicting residue contacts and protein-interaction partners”. Em: *PLoS One* 9.3 (2014), e92721. DOI: <https://doi.org/10.1371/journal.pone.0092721>.
228. D. U. Ferreira et al. “Localizing frustration in native proteins and protein assemblies”. Em: *Proceedings of the National Academy of Sciences* 104.50 (2007), pp. 19819–19824. DOI: <https://doi.org/10.1073/pnas.0709915104>.
229. M. Freiberger et al. “Local energetic frustration conservation in protein families and superfamilies”. Em: *Nature Communications* 14.1 (2023), p. 8379. DOI: <https://doi.org/10.1038/s41467-023-43801-2>.
230. P. C. Whitford, K. Y. Sanbonmatsu e J. N. Onuchic. “Biomolecular dynamics: order-disorder transitions and energy landscapes”. Em: *Reports on Progress in Physics* 75.7 (2012), p. 076601. DOI: <https://doi.org/10.1088/0034-4885/75/7/076601>.
231. L. Sun et al. “Connecting thermal and mechanical protein (un)folding landscapes”. Em: *Biophysical Journal* 107.12 (2014), pp. 2950–2961. DOI: <http://dx.doi.org/10.1016/j.bpj.2014.10.021>.

232. J. J. Skinner et al. "The Mad2 partial unfolding model: regulating mitosis through Mad2 conformational switching". Em: *Journal of Cell Biology* 183.5 (2008), pp. 761–768. DOI: <https://doi.org/10.1083/jcb.200808122>.
233. L. Aravind e E. V. Koonin. "The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair". Em: *Trends in Biochemical Sciences* 23.8 (1998), pp. 284–286. DOI: [https://doi.org/10.1016/s0968-0004\(98\)01257-2](https://doi.org/10.1016/s0968-0004(98)01257-2).
234. D. U. Ferreira et al. "On the role of frustration in the energy landscapes of allosteric proteins". Em: *Proceedings of the National Academy of Sciences* 108.9 (2011), pp. 3499–3503. DOI: <https://doi.org/10.1073/pnas.1018980108>.
235. D. U. Ferreira, E. A. Komives e P. G. Wolynes. "Frustration in biomolecules". Em: *Quarterly Reviews of Biophysics* 47.4 (2014), pp. 285–363. DOI: <https://doi.org/10.1017/S0033583514000092>.
236. L. Sironi et al. "Mad2 binding to Mad1 and Cdc20, rather than oligomerization, is required for the spindle checkpoint". Em: *The EMBO Journal* 20 (2001), pp. 6371–6382. DOI: <https://doi.org/10.1093/emboj/20.22.6371>.
237. M. Hara et al. "Structure of an intermediate conformer of the spindle checkpoint protein Mad2". Em: *Proceedings of the National Academy of Sciences* 112.36 (2015), pp. 11252–11257. DOI: <https://doi.org/10.1073/pnas.1512197112>.
238. A. D. McAinsh e G. J. P. L. Kops. "Principles and dynamics of spindle assembly checkpoint signalling". Em: *Nature Reviews Molecular Cell Biology* 24.8 (2023), pp. 543–559. DOI: <https://doi.org/10.1038/s41580-023-00593-z>.
239. A. F. Pereira e L. Martinez. "Dispersive interactions determine Urea and TMAO thermodynamic effects across the protein folding landscape". Em: *bioRxiv* (2025). DOI: <https://doi.org/10.1101/2025.02.07.637036>.

Apêndice I

Apêndice I

Tabela I.0.0.1: Detalhes das caixas de simulação para cada *ensemble* da proteína SH3 em soluções de ureia e TMAO 0,5 mol L⁻¹.

<i>Ensembles</i>	Volume médio da caixa (nm ³) e SE	Número médio de moléculas de água e SE	Número médio de moléculas de cossolvente e SE
N ₈₀ ^{S0}	162,1 ± 0,2	4576 ± 6	42,78 ± 0,05
U ₄₆ ^{S1}	182 ± 3	5207 ± 86	48,7 ± 0,8
U ₄₃ ^{S2}	180 ± 3	5132 ± 82	48,0 ± 0,8
U ₃₃ ^{S3}	177 ± 3	5035 ± 86	47,1 ± 0,8
U _{6.0} ^{S4}	207 ± 2	5957 ± 78	55,7 ± 0,7
U _{2.4} ^{S5}	228 ± 3	6616 ± 108	62 ± 1
U _{3.9} ^{S6}	250 ± 3	7296 ± 101	68,2 ± 0,9
U _{2.3} ^{S7}	249 ± 3	7268 ± 82	67,9 ± 0,8
U _{2.2} ^{S8}	284 ± 4	8353 ± 129	78 ± 1
U _{0.9} ^{S9}	330 ± 7	9806 ± 234	92 ± 2
U _{0.7} ^{S10}	282 ± 4	8286 ± 113	78 ± 1
U _{0.5} ^{S11}	322 ± 4	9549 ± 112	89 ± 1
U _{1.0} ^{S12}	313 ± 4	9268 ± 124	87 ± 1

Tabela I.0.0.2: Detalhes das caixas de simulação para cada *ensemble* da proteína BdpA em soluções de ureia e TMAO 0,5 mol L⁻¹.

<i>Ensembles</i>	Volume médio da caixa (nm ³) e SE	Número médio de moléculas de de água e SE	Número médio de moléculas de cossolvente e SE
N ₈₄ ^{B0}	208,1 ± 0,9	5985 ± 28	55,9 ± 0,3
U ₆₃ ^{B1}	207,1 ± 0,8	5953 ± 23	55,7 ± 0,2
U ₄₀ ^{B2}	203 ± 1	5822 ± 37	54,4 ± 0,3
U ₃₈ ^{B3}	191 ± 2	5467 ± 68	51,1 ± 0,6
U ₃₉ ^{B4}	192 ± 3	5496 ± 94	51,2 ± 0,9
U ₃₇ ^{B5}	242 ± 2	7023 ± 55	65,7 ± 0,5
U ₃₆ ^{B6}	221 ± 2	6392 ± 72	59,8 ± 0,7
U ₁₅ ^{B7}	235 ± 2	6818 ± 74	63,7 ± 0,7
U ₁₇ ^{B8}	251 ± 4	7324 ± 120	68 ± 1
U ₁₄ ^{B9}	223 ± 3	6446 ± 104	60 ± 1
U ₃₈ ^{B10}	240 ± 4	6973 ± 110	65 ± 1
U ₃₇ ^{B11}	302 ± 4	8909 ± 105	83 ± 1

Tabela I.0.0.3: Detalhes das caixas de simulação para cada estrutura representativa da proteína SH3 em soluções de 0,1; 0,2; 0,3; 0,4; 0,5; 0,75; e 1,0 mol L⁻¹ de ureia e TMAO.

Estrutura representativa	Concentração (mol L ⁻¹)	Volume da caixa (nm ³)	Número de moléculas de água	Número de moléculas de cossolvente
N ₈₀ ^{S0}	0,1	228,54	6826	12
N ₈₀ ^{S0}	0,2	227,73	6772	25
N ₈₀ ^{S0}	0,3	227,79	6722	37
N ₈₀ ^{S0}	0,4	227,60	6667	50
N ₈₀ ^{S0}	0,5	227,71	6617	62
N ₈₀ ^{S0}	0,75	228,00	6488	93
N ₈₀ ^{S0}	1	228,11	6359	124
U ₄₆ ^{S1}	0,1	254,66	7674	14
U ₄₆ ^{S1}	0,2	255,27	7615	28
U ₄₆ ^{S1}	0,3	254,78	7557	42
U ₄₆ ^{S1}	0,4	255,14	7499	56
U ₄₆ ^{S1}	0,5	255,04	7440	70
U ₄₆ ^{S1}	0,75	255,39	7298	104
U ₄₆ ^{S1}	1	254,65	7152	139
U _{1.0} ^{S12}	0,1	322,87	9851	18
U _{1.0} ^{S12}	0,2	322,07	9776	36
U _{1.0} ^{S12}	0,3	323,07	9701	54
U _{1.0} ^{S12}	0,4	322,02	9630	71
U _{1.0} ^{S12}	0,5	322,17	9555	89
U _{1.0} ^{S12}	0,75	322,43	9367	134
U _{1.0} ^{S12}	1	322,49	9179	179

Tabela I.0.0.4: Detalhes das caixas de simulação para cada estrutura representativa da proteína BdpA em soluções de 0,1; 0,2; 0,3; 0,4; 0,5; 0,75; e 1,0 mol L⁻¹ de ureia e TMAO.

Estrutura representativa	Concentração (mol L ⁻¹)	Volume da caixa (nm ³)	Número de moléculas de água	Número de moléculas de cossolvente
N ₈₄ ^{B0}	0,1	327,229	10000	18
N ₈₄ ^{B0}	0,2	327,344	9924	36
N ₈₄ ^{B0}	0,3	327,15	9849	54
N ₈₄ ^{B0}	0,4	327,72	9770	73
N ₈₄ ^{B0}	0,5	327,294	9695	91
N ₈₄ ^{B0}	0,75	327,065	9507	136
N ₈₄ ^{B0}	1	327,14	9319	181
U ₆₃ ^{B1}	0,1	305,938	9262	17
U ₆₃ ^{B1}	0,2	306,058	9191	34
U ₆₃ ^{B1}	0,3	306,046	9124	50
U ₆₃ ^{B1}	0,4	305,645	9053	67
U ₆₃ ^{B1}	0,5	306,011	8983	84
U ₆₃ ^{B1}	0,75	306,166	8807	126
U ₆₃ ^{B1}	1	305,406	8632	168
U ₃₇ ^{B11}	0,1	396,974	11673	21
U ₃₇ ^{B11}	0,2	397,009	11585	42
U ₃₇ ^{B11}	0,3	396,789	11493	64
U ₃₇ ^{B11}	0,4	398,107	11406	85
U ₃₇ ^{B11}	0,5	396,861	11318	106
U ₃₇ ^{B11}	0,75	397,379	11097	159
U ₃₇ ^{B11}	1	397,272	10876	212

Tabela I.0.0.5: Fração de contatos nativos (Q), área superficial acessível ao solvente (SASA, *Solvent-Accessible Surface Area*), conteúdo de estrutura secundária (folha- β (SH3) e α -hélice (BdpA)), e parâmetros de interação preferencial da ureia e do TMAO para cada *ensemble* das proteínas SH3 e BdpA.

Proteína SH3						
<i>Ensembles</i>	Q médio	SASA médio (nm ²)	Conteúdo médio de folha- β	Γ_{TMAO} médio	Γ_{ureia} médio	
N ₈₀ ^{S0}	0,850	42,413	80,018	-2.416	2,800	
U ₄₆ ^{S1}	0,553	48,828	45,584	-2.737	3,185	
U ₄₃ ^{S2}	0,552	47,512	42,650	-2.752	3,309	
U ₃₃ ^{S3}	0,544	47,831	33,044	-2.865	3,659	
U _{6.0} ^{S4}	0,236	56,175	5,999	-3.175	3,846	
U _{2.4} ^{S5}	0,176	58,718	2,437	-3.309	4,355	
U _{3.9} ^{S6}	0,170	62,640	3,932	-3.464	4,519	
U _{2.3} ^{S7}	0,159	62,128	2,258	-3.391	4,260	
U _{2.2} ^{S8}	0,139	65,246	2,220	-3.509	4,258	
U _{0.9} ^{S9}	0,106	67,949	0,892	-3.783	4,661	
U _{0.7} ^{S10}	0,097	66,900	0,689	-3.575	4,617	
U _{0.5} ^{S11}	0,091	68,472	0,502	-3.645	4,651	
U _{1.0} ^{S12}	0,087	69,213	1,040	-3.642	4,860	
Proteína BdpA						
<i>Ensembles</i>	Q médio	SASA médio (nm ²)	Conteúdo médio de α -hélice	Γ_{TMAO} médio	Γ_{ureia} médio	
N ₈₄ ^{B0}	0,772	51,248	84,495	-3,198	2,685	
U ₆₃ ^{B1}	0,729	52,156	62,579	-3,219	2,731	
U ₄₀ ^{B2}	0,609	53,069	40,224	-3,382	2,702	
U ₃₈ ^{B3}	0,553	53,303	37,951	-3,496	3,014	
U ₃₉ ^{B4}	0,543	52,673	38,889	-3,346	2,755	
U ₃₇ ^{B5}	0,378	58,953	36,962	-3,579	3,189	
U ₃₆ ^{B6}	0,370	57,210	35,996	-3,461	3,337	
U ₁₅ ^{B7}	0,348	59,481	15,067	-3,675	3,327	
U ₁₇ ^{B8}	0,336	61,114	17,224	-3,616	3,281	
U ₁₄ ^{B9}	0,335	59,255	13,642	-3,685	3,077	
U ₃₈ ^{B10}	0,333	59,279	37,873	-3,631	3,520	
U ₃₇ ^{B11}	0,305	64,460	36,861	-3,874	3,571	

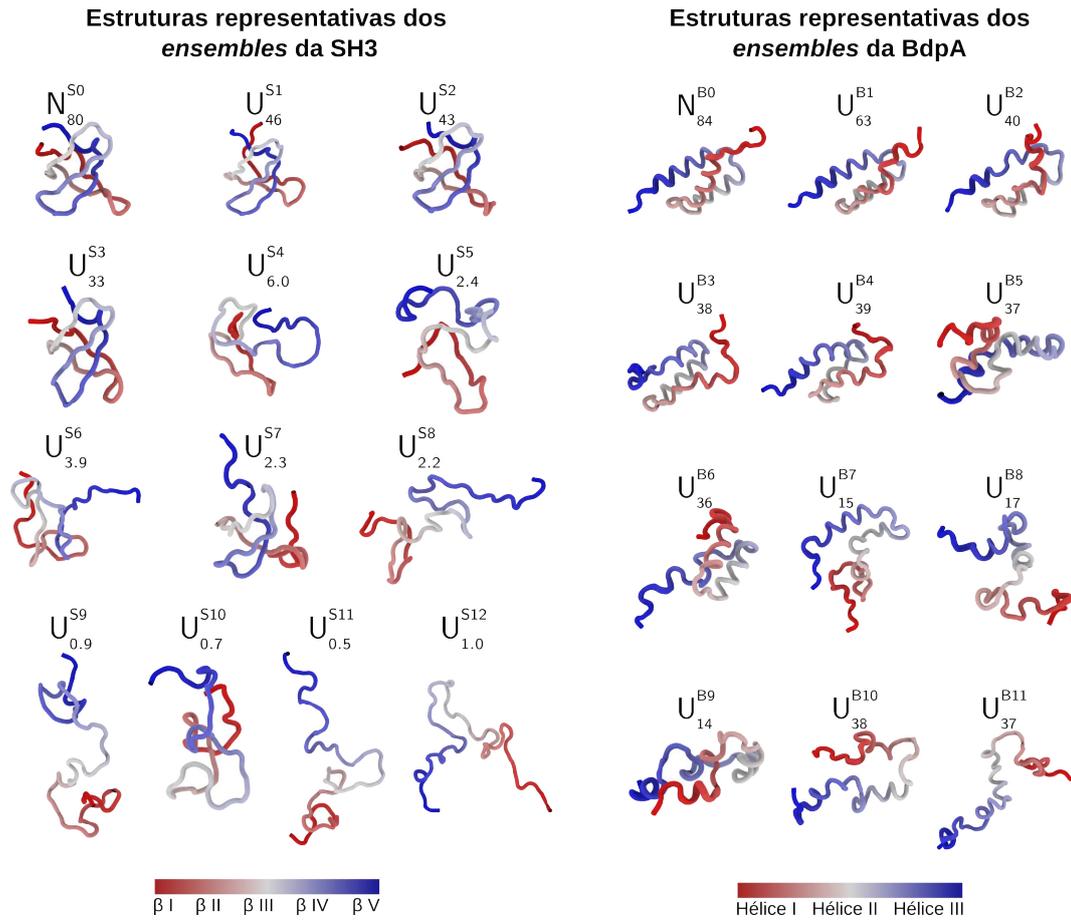


Figura I.0.0.1: Estruturas representativas de cada *ensemble* das proteínas SH3 e BdpA.