



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

Fernando Bruno da Silva

Estudo das Interações Não-Nativas,
Eletrostáticas e Hidrofóbicas,
no Processo de Enovelamento de Proteínas
Utilizando Modelos Minimalistas

São José do Rio Preto

2021

Fernando Bruno da Silva

Estudo das Interações Não-Nativas,
Eletrostáticas e Hidrofóbicas,
no Processo de Enovelamento de Proteínas
Utilizando Modelos Minimalistas

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

Financiadora: CNPq - Proc. 141715/2017-0

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

São José do Rio Preto
2021

S586e

Silva, Fernando Bruno da

Estudo das interações não-nativas, eletrostáticas e hidrofóbicas, no processo de enovelamento de proteínas utilizando modelos minimalistas / Fernando Bruno da Silva. -- São José do Rio Preto, 2021

115 p.

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

Orientador: Vitor Barbanti Pereira Leite

1. Biologia molecular. 2. Biofísica molecular. 3. Proteínas - Estruturas. 4. Enovelamento de proteínas. 5. Dinâmica molecular. I. Título.

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

Essa ficha não pode ser modificada.

Fernando Bruno da Silva

**Estudo das Interações Não-Nativas,
Eletrostáticas e Hidrofóbicas,
no Processo de Enovelamento de Proteínas
Utilizando Modelos Minimalistas**

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

Financiadora: CNPq - Proc. 141715/2017-0

Comissão Examinadora

Prof. Dr. Vitor Barbanti Pereira Leite
UNESP – Câmpus São José do Rio Preto
Orientador

Profa. Dra. Ana Lígia Scott
UFABC – Universidade Federal do ABC

Prof. Dr. Leandro Martínez
UNICAMP – Universidade Estadual de Campinas

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

Prof. Dr. Laurent Emmanuel Dardenne
LNCC – Laboratório Nacional de Computação Científica

**São José do Rio Preto
20 de maio de 2021**

*Dedico este trabalho a toda minha família,
principalmente a minha mãe e aos meus ir-
mãos que sempre me apoiaram e me incenti-
varam nesta minha jornada.*

Agradecimentos

Quero agradecer, primeiramente, a toda minha família, por acreditar em meu potencial e principalmente por me apoiar. Agradeço exclusivamente a minha mãe, Marta Bruno Prudêncio, e aos meus irmãos, Aline Bruno da Silva e Diego Bruno da Silva, pelo companheirismo e carinho, pois sem vocês eu não teria conseguido chegar onde cheguei.

Agradeço ao Prof. Dr. Vitor Barbanti Pereira Leite, por me orientar durante esses anos e por confiar em meu trabalho. Muito obrigado pela dedicação e confiança.

Agradeço ao pessoal do grupo de pesquisa Paulo Mouro, Thiago Tombonis, Vinícius de Godoi Contessoto, Vinícius Martins de Oliveira, Antonio Bento de Oliveira e Murilo Sanches pela ajuda nos momentos de dificuldade e pelos bons momentos que passamos juntos.

Aos meus amigos do departamento de Física e do programa de pós-graduação em Biofísica Molecular, agradeço pelos bons momentos que passamos juntos, por todo o companheirismo e ao acolhimento, vocês foram essenciais em minha trajetória.

Aos professores do Departamento de Física e do Programa de Pós-Graduação em Biofísica Molecular, muito obrigado por todo ensinamento compartilhado.

Agradeço aos membros titulares Profa. Dra. Ana Lígia Scott, Prof. Dr. Jorge Chahine, Prof. Dr. Leandro Martínez e Prof. Dr. Laurent Emmanuel Dardenne por aceitarem em participar desta banca de defesa de doutorado, deixo aqui meu muito obrigado.

Por fim e não menos importante, agradeço ao IBILCE/UNESP por fornecer toda a infraestrutura necessário para o desenvolvimento do projeto proposto em meu doutorado. CNPq pelo apoio financeiro fornecido em meu doutorado e ao GridUNESP pela disponibilização dos recursos computacionais imprescindíveis para a realização deste trabalho.

*“What you do makes a difference, and you have to
decide what kind of difference you want to make”*

Dr. Jane Goodall [1]

Resumo

O estudo do enovelamento proteína nesse trabalho tem como base o uso do modelo baseado em estrutura, modelo no qual apenas as interações nativas são consideradas como favoráveis durante o enovelamento. Entretanto, as interações ditas como não favoráveis são consideradas pelo modelo. Dessa forma, o objetivo geral desse trabalho buscou averiguar os efeitos das interações não-nativas durante o processo enovelamento. Os potenciais não-nativos adotados são: potencial eletrostático (Elec), hidrofóbico (HP) e de dessolvatação (db ou dsb). Esses potenciais foram combinados de diferentes formas e adicionados no modelo SBM-C α . Portanto, o trabalho desenvolvido buscou compreender como as interações não-nativas e os estados metaestáveis podem estar correlacionados. O primeiro trabalho desenvolvido envolve a proteína CI2, onde foram analisados diferentes grupos mutacionais para obter uma proteína mais termoestável. O modelo computacional adotado foi o SBM-C α + Elec + HP. O segundo trabalho foi com os domínios R15, R16 e R16M₅ da α -espectrina, esse estudo é continuação dos trabalhos realizadas anteriormente com os domínios R15, R16 e R17. O modelo computacional utilizado foi o mesmo proposto para CI2 e tem como objetivo analisar os fatores mutacionais envolvidos na redução do tempo de enovelamento devido a um grupo mutacional específico. Além disso, por meio do método ELViM, foram averiguados os perfis energéticos dessas proteínas. Por fim, no terceiro trabalho foi investigado os estados intermediários das proteínas Im7 e Im9, pois em diferentes valores de pH, esses estados apresentam diferentes comportamentos. Portanto, o modelo computacional adotado foi o SBM-C α + Elec + HP + db. Afim de reproduzir o sistema em diferentes valores de pH, foram alterados os valores de carga dos resíduos ionizáveis.

Palavras-chave: Modelo Baseado em Estrutura. Enovelamento de Proteína. Dinâmica Molecular. Efeito Hidrofóbico. Interações Eletrostáticas. Potencial de dessolvatação. Estados Metaestáveis. Estados Intermediários.

Abstract

The protein folding studies in this work are based on the structure-based model. Structure-based model are usually used to explore the folding process, but this model consists only the native interactions are favorable to the protein folding. However, unfavorable interactions may also help the process too. The non-native potentials, that incorporate these interactions, it was added to the structure-based C α model (the electrostatic (Elec), the hydrophobic (HP), and the desolvation potential (db or dsb)). These potentials were combined in different ways into the SBM-C α model and, as a result, we tried to understand how non-native interactions and the metastable states could be associated. In the present study with the CI2 protein, a group of mutants was selected to obtain a new version of the CI2 protein with a different thermostability. The simulation was performed with the SBM-C α + Elec + HP. The second project includes two domains from α -spectrin, R15 and R16, and a version of R16 with five mutations, R16M₅. This study is part II of our previous work with the domains R15, R16, and R17. The computational model was the same applied for CI2 and the main goal of this study was to understand the mutation effects in the folding time when we have the group of five mutations. The central idea was to explore the presence of a metastable state that may or may not support the folding process when considering the five mutations. In addition, using the ELViM method, the energy profiles of these proteins were analyzed. Finally, in the third project, the Im7 and Im9 intermediate states were investigated. The intermediate states of these proteins stabilized under different pH values. The simulations were carried out with the SBM-C α + Elec + HP + db model. In order to reproduce the system at different pH values, the charge values of the ionizable residues were changed, according to their theoretical pKa.

Keywords: Strucuture-Based Model. Protein Folding. Molecular Dynamics. Hydrophobic Effects. Electrostatic Interactions. Desolvation Potential. Metastable States. Intermediate States.

Lista de Figuras

- 1.1 Representação simplificada de uma ligação peptídica entre dois aminoácidos diferentes. O grupo carboxílico libera um OH, enquanto que o grupo amino cede um H. Posteriormente, essas duas espécies irão formar uma molécula de água, H_2O . A ligação CO-NH resultante é denominada ligação peptídica. 20
- 1.2 Diagrama de uma proteína (PCNA com PDB ID: 1AXC) em diferentes níveis hierárquicos. Estrutura primária (A), estrutura secundária (B), estrutura terciária (C) e por fim, estrutura quaternária (D). 21
- 1.3 Representação do funil de energia para uma proteína qualquer. No eixo horizontal, tem-se a entropia, $S(Q)$, em função da coordenada de reação. No eixo vertical, a representação da energia, $E(Q)$, lado esquerda, e a representação da coordenada de reação, lado direita, fração dos contatos nativos, Q 23
- 3.1 Representação do potencial eletrostático, V_{ele} , de Coulomb puro e blindado em função da distância de separação, r , entre dois resíduos carregados em diferentes concentrações da sal. 31
- 3.2 Representação do potencial de barreira de dessolvatação, $U(r) = V_{db}$, em função da distância de separação entre dois resíduos de aminoácidos. Esse potencial é caracterizado por um máximo, ϵ_{db} , e por um mínimo, ϵ_{ssm} , que representam as barreiras de dessolvatação. As variáveis r_{nat} , r_{db} e r_{ssm} representam, respectivamente, as distâncias de separação entre os dois resíduos e entre os resíduos e as moléculas de água. 32
- 3.3 Exemplo de duas estruturas (k e l) onde são indicadas as distâncias internas ($r_{i,j}$) entre os aminoácidos (1, 7, 13, 33 e 60). As diferenças estruturais estão relacionadas com as variações destas distâncias entre todos os aminoácidos. 37

3.4	Exemplo de visualização em 2D. No lado esquerdo da figura, tem-se a visualização da superfície de enovelamento da proteína <i>ww domain</i> . Cada ponto neste gráfico refere-se a uma conformação. No lado direito, estão representadas quatro conformações (em vermelho), e suas respectivas distâncias multidimensionais ($\delta^{k,l}$). Uma estrutura exemplo de cada uma delas está representada sobreposta com a estrutura nativa da proteína (em verde).	39
3.5	Histogramas produzidos em função da energia potencial total para o domínio R15 da α -espectrina em diferentes temperaturas. Os valores das temperaturas variam de 70(A) a 160(J), com intervalos iguais a 10. As simulações foram realizadas com o modelo SBM-C α	40
4.1	Representação da estrutura terciária (A), secundária e sequência primária (B) da proteína CI2. As folhas β_1 , β_2 , β_3 e β_4 estão representadas em cor vermelha, α_1 -hélice em cor azul e as regiões não estruturadas em cinza. Os resíduos destacados por uma esfera de cor branca (K19, D24, R49 e F51), são os resíduos de aminoácidos que quando substituídos por resíduos hidrofóbicos ou carregados aceleram o tempo de enovelamento. Já os resíduos indicados por uma seta na sequência primária, correspondem aos mesmos resíduos destacados na estrutura terciária. O desenho da estrutura tridimensional foi obtido pelo software CHIMERA e a estrutura secundária e a sequência primária pelo software ALINE	44
4.2	Energia livre eletrostática, ΔG_{qq} , para a proteína CI2-WT (A) e CI2-M _D (K19R/D24A/R49F/F51L) (B). $\Delta G_{qq} > 0$ indica que o resíduo apresenta energia livre eletrostática desfavorável para estabilizar o estado nativo, porém seu valor de SASA $< 50\%$. As setas vermelhas indicam quais são os candidatos que sofrerão mutações. Todos os resultados foram calculados para o pH = 7,0 e obtidos pelo software TKSA-MC	48
4.3	Energia livre eletrostática, ΔG_{qq} , para todas as mutações realizadas na CI2. A sequência das figuras segue a ordem decrescente, em módulo, da energia: K19R, R49W, R49Y, R49F/F51L, M _B - K19R/R49F/F51L, D24A, M _A - K19R/D24A, M _C - D24A/R49F/F51L. Os resultados foram obtidos para o pH = 7,0. $\Delta G_{qq} > 0$ indica que o resíduo apresenta energia livre eletrostática desfavorável para estabilizar o estado nativo, com SASA $< 50\%$. Já as barras em vermelho, apresentam SASA $\geq 50\%$, tornando esses resíduos candidatos a sofrerem mutações	49

- 4.4 Efeitos mutacionais na estabilidade, ΔG_T , e no tempo de enovelamento, τ^* , da proteína CI2. **(A)** Energia livre eletrostática total (ΔG_T) em função da temperatura de enovelamento (T_f^* - em unidades reduzidas) para cada grupo mutacional da CI2. **(B)** Tempo de enovelamento para cada versão da CI2, τ^* , normalizado pelo tempo de enovelamento da tipo selvagem (CI2-WT), τ^{wt} . A correlação linear esta representada pela linha continua em cor vermelha. **(C)** Tempo de enovelamento (τ^*) para as mutações propostas (M_A , M_B , M_C , e M_D) comparado com o tempo de enovelamento da R49Y e WT. As barras de cor azul são dados computacionais e as barras de cor laranja e cinza, são dados experimentais. Todos dados experimentais foram extraídos de Ladurner e Lawrence . . . 51
- 4.5 Análise da energia média eletrostática e hidrofóbica, assim como da diferença de probabilidade de formação dos contatos nativos no *ensemble* ETS e TS para CI2- M_D e CI2-WT. **(A)** Energia média eletrostática, $\langle E^{el} \rangle$, e hidrofóbica, $\langle E^{hp} \rangle$, para a CI2 selvagem (CI2-WT em cor azul) e para a mutação M_D (CI2- M_D em cor vermelha). Energia computada para o estado não-nativo (Unf - *unfolded*), pré-estado de transição (ETS - *early transtion state*), estado de transição (TS - *transition state*), pós-estado de transição (LTS - *late transition state*) e estado nativo (Fold - *folded*) para CI2-WT e CI2- M_D . **(B)** Diferença percentual (ΔP_{ij} (%)) entre os mapas de contato da CI2- M_D e CI2-WT para o ETS (triângulo superior) e para TS (triângulo inferior). A estrutura secundária está representada em cor cinza e a barra está associada à diferença de probabilidade do mapa de contato. **(C)** e **(D)** são as estruturas representativas da CI2- M_D e CI2-WT, respectivamente, no *esemble* ETS. A região destacada em vermelho na CI2- M_D corresponde a mesma região em vermelho do triângulo superior (ETS) da figura B. 53
- 5.1 Alinhamento das sequências dos domínios R15, R16 e R17 e a representação estrutural do domínio R16 em que os aminoácidos da primeira α -hélice que serão substituídos estão destacados. Na Figura A, representação estrutural da R16 destacando os resíduos Glu14 em vermelho, Glu15 em verde, Ile18 em azul, Lys21 em laranja e Val25 em preto. Na Figura B, alinhamento sequencial dos domínios R15, R16 e R17 com sua respectiva representação tubular, α -hélices, da estrutura secundária. . . 57

5.2	(A) Visualização da trajetória de enovelamento da C(R15WT, R16WT) no espaço de fase 2D obtido pelo ELViM. A barra de cor está associada à fração de contatos nativos, Q , normalizada de 0 a 1. (B) e (C) mapa de densidade bidimensional para os domínios R15WT e R16WT, respectivamente. A barra de cor corresponde a densidade de estados, ρ . Em (A), (B) e (C), cada ponto representa uma conformação e a distância euclidiana entre os pontos está associada a uma distância relativa entre as estruturas. (D) e (E), <i>ensemble</i> de estruturas obtidas a partir do pontos indicados em (B) e (C), onde o estado desenovelado e nativo estão indicado em (1) e (4).	61
5.3	(A) Visualização da trajetória de enovelamento da C(R16WT, R16M ₅) no espaço de fase 2D obtido pelo ELViM. A barra de cor está associada à fração de contatos nativos, Q , normalizada de 0 a 1. (B) e (C) mapa de densidade bidimensional para os domínios R16WT e R16M ₅ , respectivamente. A barra de cor corresponde a densidade de estados, ρ . Em (A), (B) e (C), cada ponto representa uma conformação e a distância euclidiana entre os pontos está associada a uma distância relativa entre as estruturas. (D) e (E), <i>ensemble</i> de estruturas obtidas a partir do pontos indicados em (B) e (C), onde o estado desenovelado e nativo estão indicado em (1) e (4).	64
5.4	Distribuição bidimensional do número dos contatos nativos versus o número dos contatos não-nativos (NN) hidrofóbicos (A,B e C) e carregados (D, E e F) para as proteínas R15WT, R16WT e R16M ₅ . Os histogramas correspondentes aos contatos não-nativos e aos contatos nativos encontram-se nos eixos verticais e horizontais, respectivamente	65
5.5	A e B, número de contatos não-nativos hidrofóbicos e carregados (η_{HP} e η_{Elec}) em função dos contatos nativos. C e D, energia média hidrofóbica e eletrostática ($\langle E_{HP} \rangle$ e $\langle E_{DH} \rangle$) em função dos contatos nativos. Todos os cálculos foram realizados para os domínios R15WT, R16WT e R16M ₅ que serão representadas pelas cores, vermelho, preto e verde, respectivamente.	66

5.6	Em A, mapa de densidade para os domínios R16WT e R16M ₅ , respectivamente, com seus estados intermediários destacados por uma circunferência. Número de estados selecionados correspondem a 531 e 667 para os domínios R16WT e R16M ₅ , respectivamente. Em B, C e D, probabilidade de formação de contatos nativos e não-nativos hidrofóbicos e carregados, respectivamente.	68
5.7	A distância do centro de massa das hélices. R16WT (A, B e C) e R16M ₅ (D, E e F) exibem a distância do centro de massa (Å) entre α -hélice 1 e 2, 1 e 3, e 2 e 3, respectivamente, em função dos contatos nativos com sua distribuição correspondente (vertical e horizontal). As barras coloridas representam a densidade dos estados, ρ	69
5.8	Histograma da energia eletrostática (A) e da energia hidrofóbica (B) para as proteínas R16WT e R16M ₅ , nas cores preto e azul, respectivamente.	70
6.1	Representação estrutural das proteínas Im7 e Im9. Em A, Im7 (azul) e em B, Im9 (verde), ambas apresentam 86 resíduos de aminoácidos e apresentam o RMSD $\approx 2,0$ Å. As proteínas Im7 e Im9 foram extraídas do PDB 1AYI e 1IMQ, respectivamente.	74
6.2	(A) Perfil de energia livre, $F(Q)$, em função da fração dos contatos nativos, Q . (B) Calor específico, C_v , em função temperatura em unidades reduzidas, T . As linhas nas cores preto, vermelho e verde correspondem aos potenciais dsb+HP, dsb+HP+Elec1, e dsb+HP+Elec2, respectivamente.	80
6.3	(A) Perfil de energia livre, $F(Q)$, em função dos contatos nativos em diferentes temperaturas (parte superior) e rota de enovelamento, $R(Q)$, em função dos contatos nativos, calculada na temperatura de enovelamento, $T = 0,79$ (parte inferior). (B) Número de estruturas presentes dentro do intervalo de $80 < Q < 100$, para o estado intermediário da Im7 em diferentes valores de temperatura.	82

6.4	(A) Alinhamento da estrutura nativa, em cinza, com a estrutura representativa do <i>cluster</i> -1 formado a partir da trajetória na temperatura T_2 , com valor de RMSD igual a 1,3 Å. (B) Alinhamento das estruturas representativas selecionadas a partir do <i>cluster</i> -1 nas temperaturas T_2 e T_5 , para as cores rosa e amarelo, respectivamente, com valor de RMSD igual a 1,7 Å. (C) Probabilidade da formação dos contatos nativos formados para estruturas presentes na região destacada no perfil de energia livre para a temperatura T_2	83
6.5	(A) Perfil de energia livre, $F(Q)$, em função dos contatos nativos para diferentes temperaturas. (B) Rota de enovelamento, $R(Q)$, em função dos contatos nativos, calculada na temperatura de enovelamento, $T_2 = 0,79$. (C) Triângulo superior, probabilidade de formação dos contatos nativos para o conjunto de estruturas selecionadas a partir da região destacada no perfil de energia livre para a temperatura T_2 . Triângulo inferior, diferença na probabilidade de formação dos contatos nativos entre as estruturas representativas dos sistemas Elec1 e Elec2. (D) Diferença percentual das rotas de enovelamento para os sistemas Elec1 e Elec2.	85
6.6	Cálculo do número de contatos não-nativos, hidrofóbicos (A) e eletrostáticos (B), em função do número de contatos nativos. O sistema Elec1 e Elec2 estão representados pelas linhas sólidas nas cores vermelho e verde, respectivamente.	87
6.7	Cálculo da energia livre ($K_B T$) em função do número de contatos nativos e do RMSD (Å). (A) Para o sistema Elec1, o RMSD foi calculado utilizando a estrutura nativa da Im7 como referência. (B e C) Para os sistemas Elec1 e Elec2, foi utilizado no cálculo do RMSD* a estrutura do estado intermediário (<i>cluster</i> – 1) exibida na Figura 6.4A. (D) Para o sistema ausente do potencial eletrostático, foi utilizado o cálculo do RMSD* tendo a estrutura do estado intermediário como referência (6.4A).	88
C.1	Grupo de estruturas correspondente aos domínios R15WT e R16WT obtidas pelo ELViM (vide Figuras 5.2B e 5.2C)	107
C.2	Grupo de estruturas correspondente aos domínios R16WT e R16M ₅ obtidas pelo ELViM (vide Figuras 5.3B e 5.3C)	108

Lista de Tabelas

3.1	Energia de interação entre aminoácidos hidrofóbicos dada pelo potencial de Miyazawa-Jernigan.	29
6.1	Diferentes valores de carga para as histidinas 40 e 47 da Im7 e para as histidinas 05, 39 e 46 da Im9 em diferentes valores de pH.	78
B.1	Energia de contato entre os aminoácidos segundo Miyazawa e Jernigan	106

Lista de Abreviaturas e Siglas

DNA	<i>Deoxyribonucleic acid</i> (ácido desoxirribonucleico)
RNA	<i>Ribonucleic acid</i> (ácido ribonucleico)
ELViM	<i>Energy Landscape Visualization Method</i>
SBM	<i>Structure-Based Model</i>
WHAM	<i>Weighted Histogram Analysis Method</i>
Unf	<i>Unfold</i>
ETS	<i>Early Transition State</i>
TS	<i>Transition State</i>
LTS	<i>Late Transition State</i>
MFPT	<i>Mean First Passage Time</i>
ML	<i>Machine Learning</i>
ALINE	<i>Alignment Editor</i>
RMSD	<i>Root-Mean-Square Deviation</i>
HP	<i>Hydrophobic</i>
Elec	<i>Electrostatic</i>
MJ	<i>Miyazawa-Jernigan</i>
dsb	<i>Desolvation Barrier</i>
IDP	<i>Intrinsically disordered protein</i>
TKSA-MC	<i>Tanford-Kirkwood Surface Accessibility Monte Carlo</i>

Sumário

1	Introdução	19
2	Organização da Tese	25
3	Metodologia	27
3.1	Modelo Baseado em Estrutura: SBM- $C\alpha$	27
3.2	Potencial Hidrofóbico	28
3.3	Potencial Eletrostático	30
3.4	Potencial de Dessolvatação	31
3.5	Modelo TKSA-MC	34
3.6	Cálculo da Rota de Enovelamento	35
3.7	Probabilidade de Formação de Contatos	36
3.8	Método de Visualização da Superfície de Energia Potencial	36
3.8.1	Projeção multidimensional	38
3.9	Método dos Múltiplos Histogramas	39
3.10	Detalhes da Simulação	41
3.10.1	Análise das Simulações	41
4	Efeitos Mutacionais e Otimização das Interações Não Nativas na Proteína CI2	43
4.1	Introdução	43
4.2	Objetivo	46
4.3	Resultados e Discussão	47

4.3.1	CI2 - Mutações	47
4.3.2	Interação Eletrostática e a Estabilidade da Proteína CI2	47
4.3.3	Efeitos Mutacionais na Cinética de Enovelamento	49
4.3.4	Comportamento Antagônico na CI2-M _D	51
4.4	Conclusão	55
5	Estudo do Perfil Energético dos Domínios R15, R16 e R16M₅ da α-Espectrina	56
5.1	Introdução	56
5.2	Objetivo	59
5.3	Resultados e Discussão	60
5.3.1	Rota de enovelamento da R15WT e R16WT	60
5.3.2	Otimização das interações não-nativas e a diminuição no tempo de enovelamento do domínio R16MM ₅	63
5.3.3	Densidade de contatos não-nativos do R15WT, R16WT e R16M ₅	65
5.3.4	Formação de contatos específicos devido as interações não-nativas	67
5.4	Conclusão	71
6	Estabilização do Estado Intermediário da Im7 e Im9 devido à Mudança de Carga	73
6.1	Introdução	73
6.2	Objetivo	77
6.3	Resultados e Discussão	78
6.3.1	Efeito da Mudança de Carga nas Histidinas 40 e 47 da Im7	79
6.3.2	Estado intermediário para o sistema denominado Elec1: H40(+1) e H47(0)	81
6.3.3	Estado intermediário para o sistema denominado Elec2: H40(0) e H47(0)	84
6.3.4	Densidade de contatos não-nativos, hidrofóbicos e eletrostáticos, para os sistemas Elec1 e Elec2	86
6.4	Conclusão	90

7	Considerações Finais	92
	Referências	94
	Apêndice A Dinâmica Molecular	104
	Apêndice B Table geral de Miyazawa e Jernigan	106
	Apêndice C Representação das estruturas obtidas pelo ELViM	107
	Apêndice D Produção Acadêmica	109

Capítulo 1

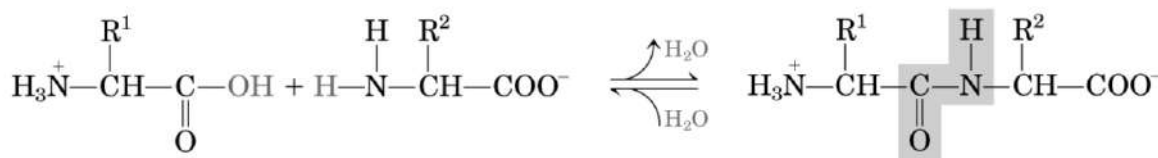
Introdução

Em nosso organismo, quase todas as funções celulares vitais para o seu funcionamento são realizadas por moléculas de proteínas. As proteínas são biopolímeros constituídos por pequenas moléculas de aminoácidos, sendo de grande importância para a vida e estando presente em todos os seres vivos. As proteínas são componentes estruturais elementares das células e apresentam diversas funções, tais como: regulatória (controlando as condições intra e extracelulares), transportadora, receptora de sinais, elaboradora de estruturas celulares, assim como, são uma parte crucial dos músculos e outros tecidos, convertendo energia química em energia mecânica [2, 3].

O que diferencia uma proteína da outra são as possíveis combinações que podem ser formadas devido a grande variedade de aminoácidos, estruturas e funções. Existem 22 tipos de aminoácidos essenciais e naturais, sendo Alanina, Asparagina, Ácido glutâmico, Ácido aspártico, Glutamina, Glicina, Ornitina, Prolina, Taurina, Cistina, Isoleucina, Leucina, Valina, Histidina, Lisina, Metionina, Fenilalanina, Treonina, Triptofano, Arginina, Cisteína, Tirosina, Serina. Os aminoácidos são conhecidos por aminoácidos α , pois o seu carbono central, usualmente denominado carbono α , realiza quatro ligações covalentes: i) uma com o grupo amino ($-\text{NH}_2$); ii) uma com um hidrogênio (H); iii) uma com um grupo carboxílico ($-\text{COOH}$) e; iv) uma com um radical ($-\text{R}$), sendo este o responsável por distinguir um aminoácido do outro, conforme representado na Figura 1.1 [2, 3].

Sabe-se que a sequência de nucleotídeos do DNA é a responsável por dirigir a síntese proteica. Por consequência, alterações no material genético (mutações/substituições) podem produzir proteínas com funções diferenciadas e, ademais, as funções das proteínas estão intrinsecamente relacionadas com suas estruturas. Portanto, a estrutura de uma proteína pode ser resumida em diferentes níveis hierárquicos: primário, secundário, terciário e quaternária [2, 3].

Figura 1.1: Representação simplificada de uma ligação peptídica entre dois aminoácidos diferentes. O grupo carboxílico libera um OH, enquanto que o grupo amino cede um H. Posteriormente, essas duas espécies irão formar uma molécula de água, H₂O. A ligação CO-NH resultante é denominada ligação peptídica.

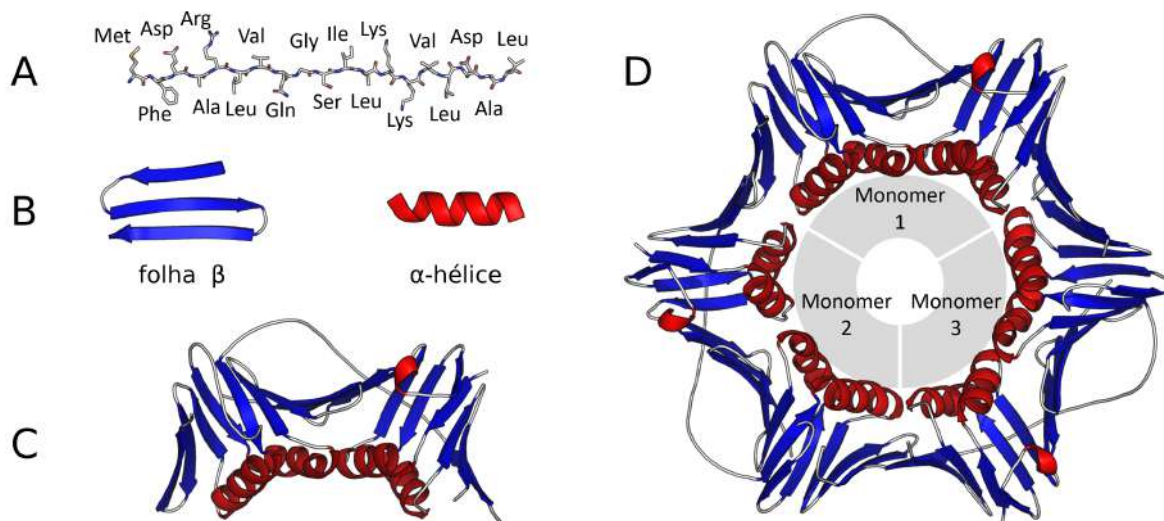


Fonte: Extraído e modificado da Ref. [2]

A estrutura primária de uma proteína, denominada pela sua sequência (Figura 1.2A), é o nível mais simples da estrutura da proteína. A estrutura secundária (Figura 1.2B) de uma proteína refere-se à estruturação local que se forma dentro de uma cadeia polipeptídica. Essas regiões tipicamente se formam devido as ligações de hidrogênio entre os resíduos e os elementos estruturais secundários, onde em uma molécula são conhecidos como α -hélices e as folhas β . A estrutura terciária (Figura 1.2C) descreve o dobramento dos elementos estruturais secundários e especifica o arranjo tridimensional de todos os átomos que compõem a proteína, caracterizando, dessa forma, sua configuração global. A estrutura terciária é estabilizada por meio de ligações não covalentes, incluindo ligações de hidrogênio, ligações iônicas, interações dipolo-dipolo e forças hidrofóbicas. No próximo nível hierárquico encontra-se a estrutura quaternária (Figura 1.2D), no qual envolve o arranjo e as interações entre os domínios da proteína ou entre duas ou mais proteínas. A estrutura quaternária descreve essencialmente como diferentes proteínas encontram-se agregadas [2, 3].

A função da proteína depende criticamente de sua estrutura tridimensional determinada por sua sequência de aminoácidos (estrutura primária), assim como, apresenta uma correlação com sua dinâmica. Este paradigma de estrutura-função, uma ideia central da biologia, sugere que as proteínas recém sintetizadas enovelam em uma única forma tridimensional antes de realizar sua função biológica. No entanto, muitas exceções à regra foram recentemente encontrados na forma das chamadas proteínas intrinsecamente desordenadas (IDPs). As IDPs não apresentam uma forma tridimensional fixa, mas apresentam funções celulares altamente essenciais [5–7]. Exemplos das funções realizadas por IDPs: reconhecimento molecular de proteínas regulatórias, DNAs e RNAs, aceleração enzimática de reações químicas, facilitador de modificações pós-traducional, splicing alternativo e inserções/deleções [8, 9]. Em geral, as atividades biológicas das IDPs desempenham um papel crucial nas células vivas, frequentemente complementando as das proteínas estruturadas. O trabalho desenvolvido nessa tese

Figura 1.2: Diagrama de uma proteína (PCNA com PDB ID: 1AXC) em diferentes níveis hierárquicos. Estrutura primária (A), estrutura secundária (B), estrutura terciária (C) e por fim, estrutura quaternária (D).



Fonte: Extraído e modificado da Ref. [4].

não envolve IDPs, mas foi mencionada devido a sua relevância no cenário científico.

Assim, compreender o mecanismo biomolecular que leva uma proteína do seu estágio inicial (estrutura primária) ao seu estágio final (estrutura terciária ou quaternária), tem sido um enorme desafio durante décadas. Ao que se pode observar, está é uma área da ciência que vem incentivando e gerando discussões entre diferentes pesquisadores. Apesar dos avanços das técnicas experimentais e teóricas, encontra-se em aberto o entendimento exato do processo que leva uma proteína a sair de sua estrutura primária e alcançar a sua estrutura terciária ou quaternária. Este processo é conhecido na literatura como enovelamento de proteína [2, 3].

O enovelamento de proteínas é um processo indispensável que ocorre em todos os seres vivos. O foco deste estudo não se restringe somente ao papel que o enovelamento desempenha no organismo, mas também aos processos físico-químicos que o governam. O entendimento do enovelamento pode responder a problemas fundamentais em diversas áreas da ciência que ainda encontram-se em aberto como, por exemplo, a causa das doenças de Alzheimer, Huntington e Parkinson [10, 11].

A formação de uma estrutura tridimensional única pode ocorrer espontaneamente sem a necessidade de ajuda adicional de outras moléculas como, por exemplo, as proteínas chaperonas. As interações dos aminoácidos dentro da cadeia polipeptídica e com moléculas do solvente são as responsáveis pela auto-organização espontânea de uma molécula de proteína para obter a conformação enovelada, frequentemente referida

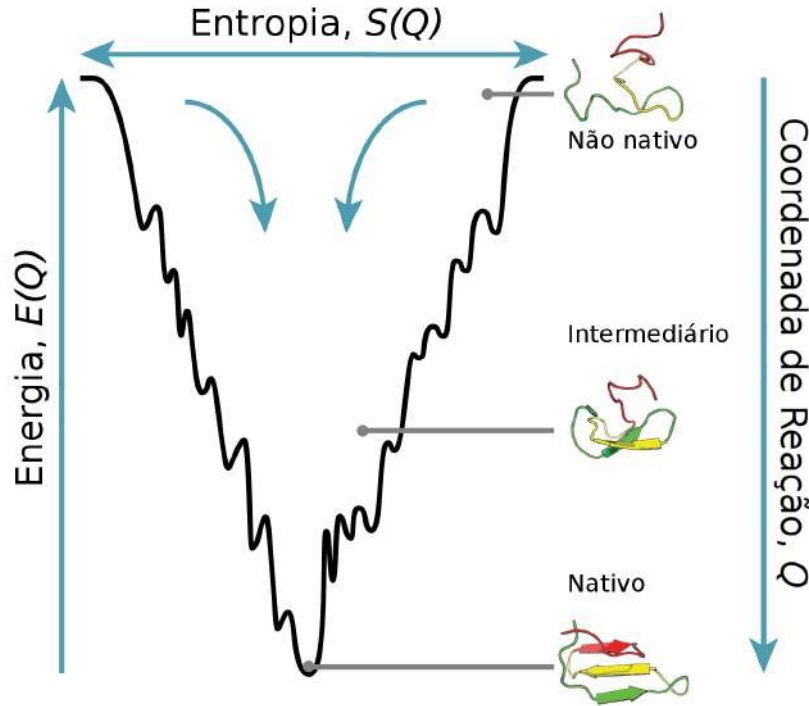
como estado nativo.

O número de configurações tridimensionais disponíveis para um polipeptídeo cresce exponencialmente com o tamanho da proteína. Explorar todo espaço de fase para encontrar uma conformação particular levaria muito tempo, tornando uma busca incompatível com os requisitos biológicos, sendo esse fenômeno conhecido como paradoxo de Levinthal [12]. De acordo com Levinthal, se levar um picossegundo para amostrar um estado conformacional particular. O estado nativo de uma proteína por meio de uma busca aleatória levaria uma escala de tempo que é comparável à idade do universo. Isso, evidentemente, sugere que as moléculas de proteína limitam o espaço fase de modo que seu enovelamento ocorra através de um número limitado de caminhos para o estado nativo.

A teoria que foi e ainda é capaz de descrever o enovelamento de proteína é conhecida como a Teoria de Superfície de Energia do inglês *Energy Landscape Theory*. [13–15]. A teoria de superfície de energia propõe a existência de uma seleção natural de vias de enovelamento, que resultam em uma superfície energética afunilada e rugosa, onde a mais profunda energia do funil corresponde a conformação nativa da proteína, conforme representado na Figura 1.3. As rugosidades correspondem aos mínimos locais de energia, que funcionam como armadilhas para o enovelamento. Para que a proteína possa exercer seu papel biológico, a reação de enovelamento deve acontecer sem que ela fique presa em mínimos locais e, por esse motivo, a inclinação do funil de energia deve ser grande o bastante para vencer esses mínimos. Essa teoria mostrou-se uma ferramenta fundamental para a compreensão qualitativa e quantitativa dos resultados teóricos e experimentais relacionados ao enovelamento [16–20]. Por meio de análises estatísticas, essa teoria também fornece uma forte base teórica para o entendimento de fenômenos de oligomerização [21] e transições funcionais em proteínas [22–25].

De acordo com a teoria de superfície de energia, o mecanismo de enovelamento apresenta a forma de um funil e pode ser descrito por um pequeno número de parâmetros, denotados como Q , que medem a similaridade de uma conformação para o estado nativo. A largura do funil representa a entropia e no eixo vertical tem-se a representação da coordenada de reação, Q , e entalpia, $E(Q)$. Para valores de $Q \approx 0$, tem-se o conjunto em sua forma não nativa e para $Q \approx 1$, o conjunto de estruturas do estado nativo. A entropia, $S(Q)$, e a energia, $E(Q)$, dependem da semelhança com a conformação nativa medida por Q . No topo do funil a entropia é grande, pois existem muitas conformações sem semelhanças com o estado nativo. A energia desses estados não estruturados também é maior em comparação com a energia do estado nativo. Conforme Q aumenta, a entropia conformacional reduz, pois o conjunto de estruturas apresenta alta similaridade com o estado nativo e a energia torna-se mais estável. A

Figura 1.3: Representação do funil de energia para uma proteína qualquer. No eixo horizontal, tem-se a entropia, $S(Q)$, em função da coordenada de reação. No eixo vertical, a representação da energia, $E(Q)$, lado esquerda, e a representação da coordenada de reação, lado direita, fração dos contatos nativos, Q .



Fonte: Extraído e modificado da Ref. [26].

correlação da energia de uma dada conformação e a similaridade do estado nativo é conhecida como o "princípio da frustração mínima" [27, 28].

Em condições fisiológicas, o estado nativo de uma proteína corresponde aos mínimos globais da energia livre. A energia livre, $F(Q)$, da proteína pode ser formulada por sua entalpia, $E(Q)$, e entropia, $S(Q)$, na temperatura de equilíbrio, T , assumindo pressão e volume constante, conforme representado pela Equação 1.1:

$$F(Q) = E(Q) - TS(Q), \quad (1.1)$$

onde Q é o parâmetro de ordem que descreve a enovelamento. Dependendo das condições a entropia e a energia não se compensam completamente conforme a proteína caminha para o estado nativo, o que resulta em um perfil de energia livre com uma barreira que passa a controlar a taxa de enovelamento em um valor intermediário de Q . As propriedades estruturais do conjunto conformacional dentro deste estado de transição

indicaram o mecanismo de enovelamento da proteína.

De acordo com a Equação 1.1, a competição entre entalpia e entropia define o estado nativo e não nativo da proteína. No estado nativo as interações intramoleculares são responsáveis pela estabilização estrutural, o que caracteriza uma baixa entropia conformacional. A baixa entropia conformacional resulta no aumento da entropia do solvente, pois os resíduos hidrofóbicos encontram-se menos expostos ao solvente no estado nativo. A entalpia da proteína também é baixa, pois as interações favoráveis são satisfeitas. A conformação nativa pode ser perturbada pela adição de desnaturante, aumento da temperatura ou pela alteração do nível de pH. O estado não nativo ou desnaturado apresenta alta entropia conformacional e alta entalpia devido à perda de interações favoráveis. Além disso, as interações entre o solvente e os resíduos hidrofóbicos reduzem a entropia do solvente em relação ao estado nativo. Portanto, o aumento da entropia também é considerado como uma força que auxilia a proteína para estado nativo [29, 30].

O estudo do enovelamento de proteínas a partir da teoria do funil de energia tem levado a contribuições relevantes para a Biofísica Molecular, assim como para áreas afins. O comportamento cinético e termodinâmico de sistemas com uma superfície energética irregular (ou rugoso) diferencia-se daqueles com uma superfície regular (ou lisa), sendo que, em casos onde a superfície de energia for irregular, há uma grande competição entre as interações locais e não locais, caracterizando assim o nível de frustração do sistema [16, 31, 32].

A frustração existe em biomoléculas e, às vezes, é de grande importância em determinadas funções biológicas. Do ponto de vista biofísico, as interações não locais também podem ser energeticamente favoráveis [33], como as interações locais. Mediante o aprimoramento de técnicas experimentais, as interações não-nativas passaram a apresentar um importante papel durante o processo de enovelamento [34, 35] e o entendimento da presença ou da ausência dessas interações é fundamental para compreender os seus efeitos sobre o processo de enovelamento [36].

Com base em toda teoria apresentada, foram explorados três sistemas com o objetivo de estudar o papel dessas interações não locais no processo de enovelamento. Os sistemas de interesse foram: i) efeitos mutacionais na proteína CI2; ii) papel das interações não nativas nos domínios R15, R16 e R16M₅ da α -espectrina; e iii) estado intermediário das proteínas Im7 e Im9. Na próxima seção encontram-se descritos como foram divididos os capítulos dos sistemas que serão abordados, assim como, os métodos e as análises estatísticas e termodinâmicas aplicadas.

Capítulo 2

Organização da Tese

Os trabalhos que serão apresentados como parte dos requisitos para a obtenção do título de Doutor em Biofísica Molecular estão organizados da seguinte forma:

No capítulo 3 será abordado o modelo computacional utilizado para realizar as simulações, onde o modelo baseado em estrutura, SBM- $C\alpha$, foi utilizado. Introduziremos os potenciais não-nativos que serão adicionados, posteriormente, ao potencial do modelo adotado (potencial não-nativo hidrofóbico, eletrostático e de dessolvatação). Também serão descritos detalhadamente os métodos de análises, sendo utilizado o modelo TKSA-MC, cálculo da rota de enovelamento, $R(Q)$, cálculo da probabilidade de formação de contatos nativos e não-nativos, $\Delta P(\%)$, método ELViM e, por fim, o método para o cálculo das propriedades termodinâmicas, WHAM, assim como, os detalhes da simulação.

No capítulo 4 serão estudados os efeitos mutacionais e a otimização das interações não-nativas na proteína CI2. Nesse trabalho buscamos compreender por meio de simulações computacionais como certas mutações afetam de forma direta a estabilidade da proteína e, como também, certas combinações de mutações pré-existentes podem aumentar a estabilidade e a taxa de enovelamento. Esse estudo foi realizado por meio de simulações com o modelo baseado em estrutura com adição de potenciais não-nativos, eletrostático e hidrofóbico.

No capítulo 5 serão apresentados os resultados obtidos para os domínios R15, R16 e R16M₅ da *alpha*-espectrina. Esse trabalho é continuação dos estudos realizados para os domínios R15, R16 e R17, porém com o objetivo de compreender e analisar detalhadamente os estados metaestáveis que esses domínios apresentam com as mutações que aumentam a taxa de enovelamento. Para melhor entender esses estados, foram calculados os mapas bidimensionais pelo método ELViM e, posteriormente, aqueles com maior densidade foram analisados. As simulações foram realizadas utilizando

o modelo baseado em estrutura com adição dos potenciais não-nativos, eletrostático e hidrofóbico.

No capítulo 6 focamos nossos estudos nas proteínas Im7 e Im9. Essas proteínas apresentam estados intermediários estáveis em diferentes condições físico-químicas, sendo que tais estados já foram observados experimentalmente. Do ponto de vista computacional, esses estados não foram reportados considerando o efeito eletrostático ao modelo SBM-C α , mas, tão somente, o efeito hidrofóbico. Portanto, as simulações aqui realizadas utilizam o modelo SBM-C α , onde o potencial de Lennard-Jones foi substituído pelo potencial de dessolvatação. Também foram acoplados os potenciais não-nativos, eletrostático e hidrofóbico.

No capítulo 7 apresentaremos as considerações finais e concluiremos o trabalho de forma geral ressaltando as pontes que existem entre os projetos aqui abordados. O modelo computacional aqui utilizado foi crucial para que pudessemos conectar nossos resultados, pois apesar dos projetos possuírem propostas diferentes, eles se complementam.

Capítulo 3

Metodologia

Neste capítulo serão detalhados os modelos adotados neste trabalho. Na seção 3.1, encontra-se a descrição do modelo baseado em estrutura SBM- $C\alpha$ e dos potenciais não-nativos acoplados ao potencial do modelo SBM- $C\alpha$. Esses potenciais contabilizam interações hidrofóbicas, eletrostáticas e de dessolvatação e encontram-se descritos nas seções 3.2, 3.3 e 3.4. Nas próximas seções, encontram-se os métodos utilizados nos três projetos estudados nesse trabalho. Na seção 3.5, apresenta-se o cálculo da energia eletrostática da interação carga-carga dos resíduos de aminoácidos utilizando o modelo de Tanford-Kirkwood. Na seção 3.6, o cálculo da rota de enovelamento, que permite uma análise detalhada de todo o processo. Na seção 3.7, o cálculo da probabilidade de formação de contatos nativos e não-nativos. Seção 3.8, a descrição do método de visualização da superfície de energia potencial. Na seção 3.9, o método dos múltiplos histogramas (WHAM), que foi utilizado para o cálculo das variáveis termodinâmicas. Os detalhes computacionais adotados estão descritos na seção 3.10.

3.1 Modelo Baseado em Estrutura: SBM- $C\alpha$

As simulações computacionais de dinâmica molecular (Apêndice A) com modelo baseado em estrutura (SBM) tem seu potencial definido de tal maneira que o mínimo de energia total corresponde à estrutura nativa da proteína. Este modelo tem sido eficiente na investigação de propriedades cinéticas e termodinâmicas, como também, no processo de enovelamento [19, 37]. Nesse trabalho, foi adotado o modelo SBM- $C\alpha$, no qual os aminoácidos são representados por esferas rígidas localizadas na posição do carbono $C\alpha$. Além disso, essa representação considera que todos os carbonos $C\alpha$ possuem a mesma massa e diâmetro. A equação do modelo SBM- $C\alpha$ é dado pelos

termos da Equação 3.1:

$$\begin{aligned}
 V_T(\Gamma, \Gamma_o) = & \sum_r \epsilon_r (r - r_o)^2 + \sum_\theta \epsilon_\theta (\theta - \theta_o)^2 \\
 & + \sum_\phi \epsilon_\phi \left\{ [1 - \cos(\phi - \phi_o)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_o))] \right\} \\
 & + \sum_{LJ} \epsilon_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{rep} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12}.
 \end{aligned} \tag{3.1}$$

onde r é definido como a distância entre dois resíduos subsequentes, θ pelos ângulos formados por três resíduos subsequentes e ϕ pelos ângulos diédricos formados por quatro resíduos subsequentes. Já o termo zero subscrito indica que o parâmetro é retirado da estrutura nativa, r_o , θ_o e ϕ_o . O termo de contato está representado pelo potencial de Lennard-Jones com os termos elevados a 10-12 em vez de 6-12, essa modificação está ligado ao fato da representação do modelo escolhido. Assim, d_{ij} é a distância nativa entre os pares de contatos i e j , sendo essa distância determinada pelo algoritmo *Shadow Contact Map* [38] e r_{ij} representa a distância desses contatos ao longo da dinâmica. Os parâmetros ϵ_r , ϵ_θ e ϵ_ϕ descrevem as forças de ligação, angular e diédrica, respectivamente, com valores iguais a $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$, $\epsilon_\phi = \epsilon_C$ e $\epsilon_{NC} = \epsilon_C$, ϵ_C igual a 1 unidade (unidades reduzidas do Gromacs). Por fim, o último termo, termo de exclusão, considera as interações de repulsão entre pares de resíduos que não estão contabilizados no mapa de contato nativo [19, 39–42].

3.2 Potencial Hidrofóbico

O potencial do modelo SBM-C α contabiliza apenas as interações favoráveis, interações nativas, durante o processo de enovelamento. Essas interações já foram tratadas como sendo únicas durante todo o processo, negligenciando outras interações favoráveis ou desfavoráveis para o sistema. Com base em observações experimentais, as interações não nativas são verificadas em maior incidência em um sistema conformacional não estruturado. Dessa forma, elas facilitam o acesso a estruturas localizadas em mínimos locais na funil de energia, ou seja, essas interações possibilitam varrer um espaço conformacional que apresenta baixa frequência quando apenas as interações nativas são contabilizadas.

Com a adição das interações não-nativas no potencial SBM-C α e com base

nos resultados que constam na literatura, sabe-se que a cinética de enovelamento sofrerá alterações, ocasionando mudanças e perturbações no estado de transição. Com os resultados da src-SH3 publicados por Clementi et al. [43], foi possível estabelecer uma conexão entre resultados computacionais e experimentais. Esses resultados mostraram um aumento na taxa de enovelamento quando as interações não-nativas foram adicionadas. No caso dos domínios R15, R16 e R17 da α -espectrina [44], pode-se observar um compartimento semelhante ao contabilizar as interações não-nativas hidrofóbicas durante o processo de enovelamento.

Portanto, as interações não-nativas entre os aminoácidos hidrofóbicos para o modelo abordado serão contabilizadas apenas para os aminoácidos a seguir: Ala (A), Val (V), Leu (L), Ile (I), Met (M), Trp (W) e Phe (F). A energia de interação entre um par de aminoácidos hidrofóbico é dado pela tabela de potencial Miyazawa-Jernigan (Apêndice B)[45], que está reproduzida abaixo.

Tabela 3.1: Energia de interação entre aminoácidos hidrofóbicos dada pelo potencial de Miyazawa-Jernigan.

	ALA	GLY	ILE	LEU	MET	PHE	PRO	TRP	VAL
ALA	-0,12								
GLY	-0,08	-0,29							
ILE	-0,37	-0,13	-0,74						
LEU	-0,38	-0,16	-0,81	-0,84					
MET	-0,27	-0,17	-0,66	-0,70	-0,70				
PHE	-0,36	-0,19	-0,73	-0,80	-0,83	-0,88			
PRO	0,15	0,02	-0,05	-0,12	-0,13	-0,19	0,11		
TRP	-0,27	-0,25	-0,60	-0,62	-0,73	-0,68	-0,37	-0,64	
VAL	-0,32	-0,15	-0,67	-0,74	-0,51	-0,67	-0,05	-0,51	-0,65

Fonte: Dados extraídos da Referência [45]

O potencial que contabiliza as interações não-nativas entre os resíduos hidrofóbicos é dada pela Equação 3.2 [36]. O termo r_{ij}^{HP} representa a distância entre os aminoácidos hidrofóbicos, i e j , o número total de resíduos hidrofóbicos é igual M e σ representa volume excluído dos átomos. As interações não-nativas para modelo homogêneo, assumi-se $\kappa_{ij} = -1$ e $K_{HP} = 1$, independentemente do tipo de resíduo [46], enquanto que no modelo aqui apresentado, distingui-se as interações entre os aminoácidos hidrofóbicos, o valor de $\kappa_{ij} = \Delta\epsilon_{ij}$, onde $\Delta\epsilon_{ij}$ representa a energia de interação entre os contatos i e j que consta na tabela do Miyazawa e Jernigan, $\sigma = 5 \text{ \AA}$ e K_{hp} é um parâmetro ajustável. Dois aminoácidos são considerados como contatos não-nativos hidrofóbicos se, e somente se, $|i - j| > 3$, e $r_{ij} < 8 \text{ \AA}$.

$$V_{hp} = \sum_i^M \sum_{j=i+4}^M K_{hp} \kappa_{ij} \exp[-(r_{ij}^{HP} - \sigma)^2]. \quad (3.2)$$

3.3 Potencial Eletrostático

Em um sistema soluto-solvente, a distribuição das cargas do soluto, assim como, as condições físico-químicas que o sistema apresenta, determinarão a distribuição e a coordenação das moléculas de água. Em um sistema *in vitro*, adiciona-se sais à solução para balanceá-la e torná-la neutra. O processo de balanceamento é essencial para o sistema, pois em um meio neutro ocorre a maximização da atividade proteica. Já em simulações computacionais com solvente explícito, o mesmo procedimento pode ser realizado, adicionam-se contra-íons a solução e como resultado, obtém-se um sistema eletricamente neutro. Em simulações computacionais com solvente implícito, a vantagem que esse sistema apresenta encontra-se no baixo custo computacional quando comparado com simulações com solvente explícito. O modelo que incorpora as interações eletrostáticas na presença de íons foi introduzida ao potencial através da teoria de Debye-Hückel [47].

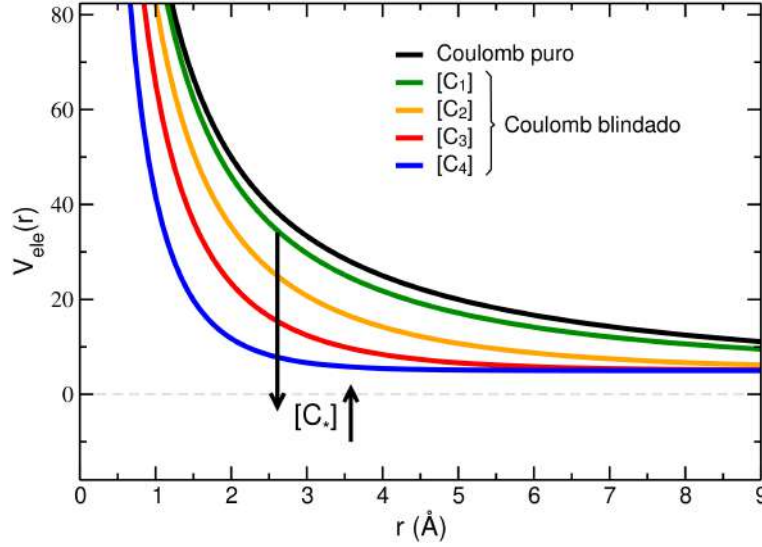
O potencial proposto por Peter Debye e Erich Hückel explica o comportamento de soluções de eletrólitos. Esse potencial é obtido a partir da solução linearizada da equação de Poisson-Boltzmann. O potencial de Debye-Huckel apresenta um termo exponencial, ocasionando uma diminuição no alcance das interações quando comparado ao alcance das interações com o potencial de Coulomb puro (vide a Figure 3.1). Essa alteração no alcance das interações está ligado ao fato das cargas blindarem o sistema, pois os íons de sal (cátions e ânions) desempenham essa função.

Conforme mencionado, o baixo custo computacional está diretamente ligado ao fato da utilização desse sistema, assim, a energia de interação entre dois resíduos carregados é dada pela seguinte expressão:

$$V_{ele} = \sum K_{elec} \frac{q_i q_j \exp(-\kappa r_{ij}^{elec})}{\epsilon_K r_{ij}^{elec}}, \quad (3.3)$$

onde r_{ij}^{elec} é a distância entre quaisquer dois resíduos carregados i e j , a carga dos resíduos são representadas por q_i e q_j , K_{elec} é igual a $332 \text{ kcal}\text{\AA}(\text{mole}^2)$, a constante dielétrica, ϵ_K , é igual a 80 e κ^{-1} é o comprimento de Debye [47] e ele corresponde à

Figura 3.1: Representação do potencial eletrostático, V_{ele} , de Coulomb puro e blindado em função da distância de separação, r , entre dois resíduos carregados em diferentes concentrações de sal.



Fonte: Elaborada pelo autor.

distância em que a energia de interação entre dois resíduos carregados é igual à energia térmica $k_B T$, conforme Equação 3.4:

$$\kappa^{-1} = \sqrt{\frac{\epsilon_o \epsilon_K \kappa_B T}{2 N_A e^2 I}} \equiv \frac{0,304}{\sqrt{I}}. \quad (3.4)$$

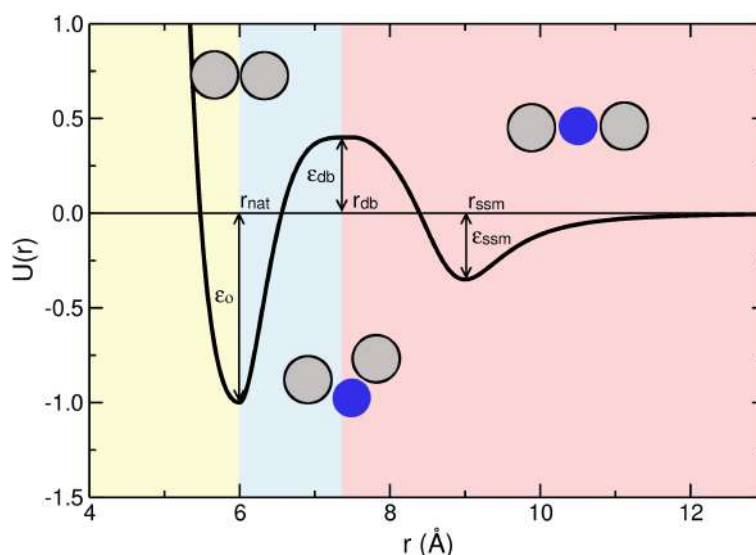
Na equação acima, o termo ϵ_o é a constante de permissividade no vácuo, κ_B é a constante de Boltzmann, N_A é o número de Avogadro, e é a carga elementar e I é a força iônica da solução, ou seja, é a medida da concentração de íons presentes na solução. Portanto, o alcance das interações eletrostáticas do sistema será modulado com o aumento ou a diminuição da concentração de sal presente no sistema, conforme pode ser visto na Figura 3.1.

3.4 Potencial de Dessolvatação

As interações intermoleculares proteína-água são essenciais na formação das estruturas, nas dinâmicas e nas funções biomoleculares que esses complexos podem exercer [48]. Durante o processo de enovelamento, para que a proteína alcance seu estado nativo, ocorrem não só interações intramoleculares, onde os resíduos da pro-

teína interagem entre si, mas também interações intermoleculares, onde os resíduos da proteína interagem com as moléculas externas, como, por exemplo, as moléculas de água. A cinética de formação desses contatos (resíduos da proteína e moléculas de água) influencia a barreira de dessolvatação/solvatação, pois ela está associada de forma direta com a região parcialmente excluída entre as camadas de solvatação. Na Figura 3.2, tem-se a exemplificação da barreira de dessolvatação/solvatação entre dois resíduos interagindo em função da distância de separação entre eles.

Figura 3.2: Representação do potencial de barreira de dessolvatação, $U(r) = V_{db}$, em função da distância de separação entre dois resíduos de aminoácidos. Esse potencial é caracterizado por um máximo, ϵ_{db} , e por um mínimo, ϵ_{ssm} , que representam as barreiras de dessolvatação. As variáveis r_{nat} , r_{db} e r_{ssm} representam, respectivamente, as distâncias de separação entre os dois resíduos e entre os resíduos e as moléculas de água.



Fonte: Elaborada pelo autor.

Conforme os resíduos aproximam-se, uma molécula de água será expulsa, ou seja, as interações favoráveis soluto-solvente sofrerem uma diminuição, ao mesmo tempo, essa diminuição será compensada pelas interações solvente-solvente, que passam a interagir fortemente em curtas distâncias. Portanto, a barreira de dessolvatação é a barreira de energia que surge na formação de interações favoráveis. Do ponto de vista computacional, simulações que possuem solvente explícito apresentam esse processo de dessolvatação. Já em simulações com solvete implícito, o processo para retirar uma molécula de água pode ser realizado com a adição de um potencial que apresente uma barreira que contabilize o processo de solvatação/dessolvatação.

A utilização do potencial com uma barreira de dessolvatação vem sendo amplamente empregado em diversos modelos computacionais [36, 49]. Contudo, sua

primeira aplicação foi realizada em um modelo de rede, ou seja, em um modelo super minimalista com um potencial de poço quadrado simples que apresenta um núcleo repulsivo infinito, representando, assim, o potencial de força média de agregação de grupos hidrofóbicos em solução [50]. No entanto, o potencial de dessolvatação amplamente empregado no modelo SBM-C α consiste em três funções distintas [51] que são acopladas aos potenciais de ligação, angular e diedral. Portanto, na Equação 3.5 tem-se a sua representação:

$$V_{db} = \begin{cases} \epsilon_o Z(r)[Z(r) - 2], & \text{se } r < r_{nat}, \\ CY(r)^n[Y(r)^n/2 - (r_{db} - r_{nat})^{2n}]/2n + \epsilon_{db}, & \text{se } r_{nat} \leq r < r_{db}, \\ -B[Y(r) - h_1]/[Y(r)^m + h_2], & \text{se } r \geq r_{db}, \end{cases} \quad (3.5)$$

onde

$$\begin{aligned} Z(r) &= (r_{nat}/r)^k, \\ Y(r) &= (r - r_{db})^2, \\ C &= 4n(\epsilon_o + \epsilon_{db})/(r_{db} - r_{nat})^{4n}, \\ B &= m\epsilon_{ssm}(r_{ssm} - r_{db})^{2(m-1)}, \\ h_1 &= (1 - 1/m)(r_{ssm} - r_{db})^2/(\epsilon_{ssm}/\epsilon_{db} + 1), \\ h_2 &= (m - 1)(r_{ssm} - r_{db})^{2m}/(1 + \epsilon_{db}/\epsilon_{ssm}). \end{aligned} \quad (3.6)$$

O potencial de dessolvatação descrito na Equação 3.5 é aplicado somente nos contatos nativos que a proteína realiza. Esse potencial apresenta dois mínimos, ϵ_o e ϵ_{ssm} , e um máximo que corresponde à barreira de dessolvatação/solvatação, ϵ_{db} . Os termos r_{nat} , r_{db} e r_{ssm} correspondem à distância nativa entre dois resíduos de aminoácidos, ao máximo da barreira de solvatação e ao mínimo da interação entre dois contatos nativos na presença de uma molécula de água, respectivamente. Nesse trabalho, consideramos $r_{ssm} = r_{nat} + 3 \text{ \AA}$, sendo 3 \AA valor aproximado do diâmetro da molécula de água, $r_{db} = (r_{ssm} + r_{nat})/2$ e k, m, n correspondem a 6, 3 e 2, respectivamente.

O primeiro, segundo e terceiro termo da Equação 3.5 correspondem às cores amarela, azul e vermelha, respectivamente, na Figura 3.2. O primeiro termo do potencial ($r < r_{nat}$), região amarela, corresponde ao potencial de Lennard-Jones. O segundo termo, em cor azul, corresponde à parte atrativa do potencial ($r_{nat} \leq r < r_{db}$) e o terceiro termo, em cor rosa, corresponde ao mínimo mediado pela água.

3.5 Modelo TKSA-MC

Os cálculos da interação carga-carga dos resíduos de aminoácidos serão realizados utilizando o modelo de Tanford-Kirkwood [52]. Esse modelo permite avaliar o efeito eletrostático de cada resíduo ionizável para a estabilidade térmica da proteína e, nesse estudo, essa análise será realizada por meio do software TKSA-MC do inglês *Tanford-kirkwood Surface Accessibility Monte Carlo* [53]. Este servidor usa a abordagem Tanford-Kirkwood com uma correção de acessibilidade do solvente [54]. Em suma, o método TKSA consiste em considerar a proteína como uma esfera de baixa constante dielétrica centrada no seu centro de massa. Todas as cargas estão dentro desta esfera e são colocadas na posição dos átomos carregados. Fora da esfera, a constante dielétrica é igual a 78,7 (constante dielétrica da água) e o sal é tratado implicitamente no modelo [55–57]. A energia de interação carga-carga de um par de resíduos tituláveis é dada por:

$$U_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}), \quad (3.7)$$

onde e é a carga elementar; A_{ij} , B_{ij} e C_{ij} são parâmetros da solução de Tanford-Kirkwood relacionados à posição dos resíduos ionizáveis, força iônica e constante dielétrica, respectivamente [52, 58, 59]; b é o raio da proteína; a é a distância de aproximação mínima dos íons; e SA_{ij} é a área da superfície acessível ao solvente de um par de resíduos carregados [56]. Por fim, os estados de protonação dos resíduos tituláveis são amostrados pelo algoritmo de Metropolis [60].

Uma vez determinada a energia eletrostática entre os resíduos tituláveis, é possível calcular a energia livre, $\Delta G_N(\chi)$, para a proteína no estado nativo em um determinado estado de protonação, χ :

$$\Delta G_N(\chi) = -RT(\ln 10) \sum_{i=1}^n (q_i + x_i) pK_{a,Ref,i} + \frac{1}{2} \sum_{i,j=1}^n U_{ij}(q_i + x_i)(q_j + x_j), \quad (3.8)$$

com R sendo a constante universal dos gases, T é a temperatura, $q_{i,j}$ é a carga do resíduo ionizável, $x_{i,j}$ é igual a 0, 1 ou -1 (isso dependerá da natureza do resíduo ionizável) e o $pK_{a,ref,i}$ é o valor do pK_a do resíduo analisado.

Então, a probabilidade da proteína estar em seu estado nativo com um determinado estado de protonação, $\rho_N(\chi)$, é dada por:

$$\rho_N(\chi) = \frac{1}{Z_N} \exp \left[-\frac{\Delta G_N(\chi)}{RT} - v(\chi)(\ln 10)pH \right], \quad (3.9)$$

sendo $v(\chi)$ igual ao número de resíduos ionizáveis que são protonados no estado de protonação χ e Z_N é a função de partição para a proteína em seu estado nativo. Não foi contabilizado que estruturas desenoveladas contribuem para as interações eletrostáticas. Os estados de protonação χ são calculados via Monte Carlo utilizando o algoritmo de Metrópolis. A energia eletrostática total média, $\langle W_{qq} \rangle$, é calculada pela média desses estados de protonação, conforme a equação abaixo:

$$\langle W_{qq} \rangle = \sum_{\chi} \left[\frac{1}{2} \sum_{i,j=1}^n U_{ij}(q_i + x_i)(q_j + x_j) \right] \rho_N(\chi). \quad (3.10)$$

Por fim, foi demonstrado por Ibarra-Molero e Makhatadze que a contribuição da energia livre eletrostática para a energia livre, $\Delta G_N(\chi)$, pode ser descrita pelo negativo da energia eletrostática total média do estado nativo, $\Delta G_{qq}(\chi) \approx -\langle W_{qq} \rangle$.

3.6 Cálculo da Rota de Enovelamento

O cálculo da rota de enovelamento é uma ferramenta que permite a análise mais detalhada do processo de enovelamento ao longo do funil de energia [61]. A rota indica a especificidade de um caminho dentre todos os outros caminhos possíveis que podem levar uma proteína para o seu estado nativo. Seu cálculo é realizado a partir de um *ensemble* de amostras da simulação. Sendo assim, determina-se a razão das configurações acessíveis com todas as outras possibilidades conformacionais que apresentam a mesma quantidade de contatos nativos [19].

A função que descreve o cálculo da rota de enovelamento é dada pela Equação 3.11:

$$R(Q) = \sum_{i=1}^{N_{nat}} \frac{\left\langle \left(\langle Q_i \rangle_Q - Q \right)^2 \right\rangle_Q}{MQ(1-Q)}, \quad (3.11)$$

onde Q_i , $i = 1, \dots, N_{nat}$, representa os contatos nativos. Para qualquer configuração, Q_i é igual a 1 somente se o contato i estiver formado, caso contrário, Q_i é igual a 0. O cálculo da média $\langle Q_i \rangle_Q$ é realizado sobre todas as configurações que apresentam o mesmo número de contatos nativos. Em seguida, o cálculo é normalizado pelo número máximo de rotas possíveis. Dessa forma, $R(Q)$ apresentará valores entre 0 e 1. Para

valores próximos de 1, $R(Q)$ sinaliza alta especificidade na formação de contatos nativos, levando a algumas rotas possíveis para o enovelamento. Em outras palavras, o funil de energia do enovelamento exibe uma alta rugosidade. Em contrapartida, valores próximos de zero indicam que a formação de diferentes contatos nativos ocorrem com a mesma probabilidade. Isto significa que não existe um caminho favorável para o enovelamento [62].

3.7 Probabilidade de Formação de Contatos

A probabilidade de formação dos contatos nativos, P_{ij} , foi calculada considerando todo o processo de enovelamento da proteína. A formação de dois contatos nativos, i e j , para um estado qualquer (P_{ij}^*) é calculada em relação ao estado desenovelado (P_{ij}^U - do inglês *unfolded*) e enovelado (P_{ij}^F - *folded*). Portanto, P_{ij} é um parâmetro estatístico calculado a partir da trajetória da dinâmica, conforme a Equação 3.12. Na Equação 3.13, ΔP_{ij} é a diferença entre dois mapas de contatos nativos para diferentes estados, sendo P_{ij}^{E1} para o estado um e P_{ij}^{E2} para o estado dois. Logo, nesse trabalho, apresentaremos a diferença percentual entre esses mapas. Quando um conjunto de resíduos são substituídos por outros grupos de aminoácidos, o ΔP_{ij} reflete essa alteração na maioria dos casos.

$$P_{ij} = \frac{P_{ij}^* - P_{ij}^U}{P_{ij}^F - P_{ij}^U}. \quad (3.12)$$

$$\Delta P_{ij} = (100\%) \times \frac{P_{ij}^{E2} - P_{ij}^{E1}}{P_{ij}^{E1}}. \quad (3.13)$$

3.8 Método de Visualização da Superfície de Energia Potencial

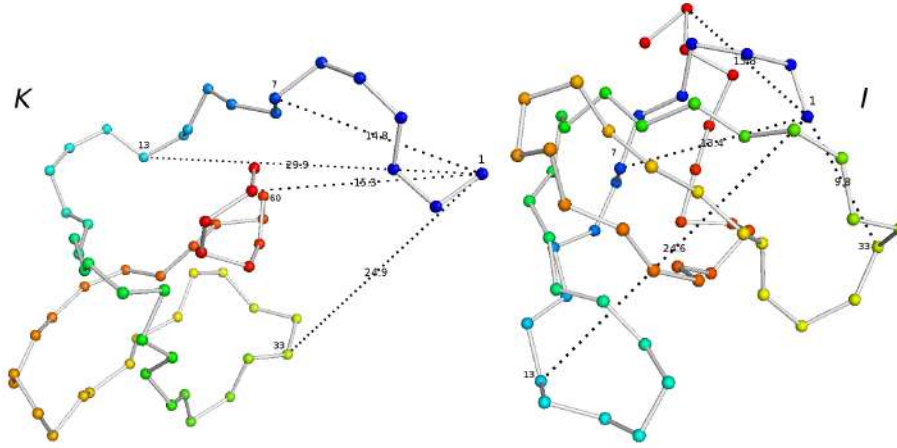
O método de visualização da superfície de energia potencial (*Energy Landscape Visualization Method* - ELViM) foi desenvolvido por Oliveira Junior [63] e permite ter uma descrição do funil de enovelamento. Para obter a visualização bidimensional, deve-se, primeiro, estabelecer a medida de dissimilaridade entre as conformações. Observa-se que as coordenadas dos resíduos são essenciais para a construção dos dados de dissimilaridade, pois diferentes combinações de coordenadas resultam em diferentes estruturas e, conseqüentemente, em diferentes conformações. Dessa forma, a medida

de comparação entre duas estruturas, k , l , pode ser escrita como:

$$q^{k,l} = \frac{1}{N} \sum_N \exp \frac{-(r_{i,j}^k - r_{i,j}^l)^2}{\sigma_{i,j}^2}, \quad (3.14)$$

onde $r_{i,j}$ é a distância entre os aminoácidos i e j das conformações k e l , N é o fator de normalização para que o valor de q seja um número entre 0 e 1 e $\sigma_{ij} = \sigma_0|i - j|^\epsilon$, com σ_0 e ϵ parâmetros ajustáveis [64]. Na Figura 3.3 é apresentado um esquema de como as distâncias são calculadas para duas conformações diferentes, k e l .

Figura 3.3: Exemplo de duas estruturas (k e l) onde são indicadas as distâncias internas ($r_{i,j}$) entre os aminoácidos (1, 7, 13, 33 e 60). As diferenças estruturais estão relacionadas com as variações destas distâncias entre todos os aminoácidos.



Fonte: Extraído da Ref. [65].

Para encontrar o valor de dissimilaridade, utiliza-se a expressão:

$$\delta^{k,l} = 1 - q^{k,l}, \quad (3.15)$$

sendo $\delta^{k,l}$ a dissimilaridade entre as conformações k e l , com valores na faixa entre 0 e 1. Os valores próximos de zero significam que as duas conformações são muito similares, ao passo que, valores próximo de 1, as estruturas são diferentes. A medida de dissimilaridade é calculada para todos os pares de estruturas k e l de uma dada trajetória, resultando em uma matriz triangular $n \times n$, onde n é o número total de conformações analisadas.

Após o cálculo da matriz de dissimilaridades, realiza-se a projeção muti-dimensional. Devido ao tamanho da matriz (na ordem de 5×10^9 estruturas), o cálculo da projeção multidimensional fica inviável, uma vez que esse tipo de tratamento tem

um grande custo computacional por necessitar de uma grande quantidade de memória. Para contornar esse problema, antes do cálculo da matriz dissimilaridade, foram selecionadas 5×10^4 estruturas. Dessa forma, diminui-se de forma considerável o tempo computacional para realizar o cálculo de δ .

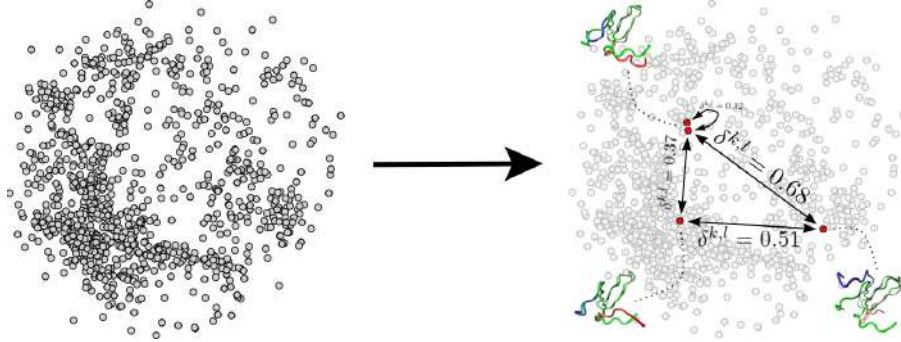
3.8.1 Projeção multidimensional

A etapa mais complexa para a criação da projeção é a redução do espaço multidimensional original dos dados para um espaço dimensional compreensível, no máximo, três dimensões. Idealmente, o modelo visual criado deve se assemelhar ao modelo visual proposto pela teoria do funil de enovelamento. Para isso, deve-se manter, ao máximo, a correspondência entre os dados antes e depois da redução dimensional. Técnicas de projeção multidimensional tratam essa questão, pois mapeiam os dados em espaços p -dimensionais, com $p = \{1, 2, 3\}$, preservando, ao máximo, a informação sobre as relações de distâncias ou dissimilaridade entre os dados [66] .

Em uma projeção, a proximidade dos pontos indica a semelhança entre os objetos que os representam. Pontos próximos indicam conformações semelhantes de acordo com a medida de distância δ . Intuitivamente, pontos distantes representam objetos com pouca similaridade. Assim, a questão principal para a construção de uma boa projeção está diretamente relacionada com a forma com que as distâncias entre os objetos multidimensionais (δ) são calculados. Dentre as várias técnicas de projeção que podem ser utilizadas, adotou-se, neste trabalho, a técnica *Force Scheme* [66], pois ela propõe um balanceamento entre precisão e desempenho computacional. Esta técnica estabelece um sistema de forças, onde, inicialmente, posicionam-se os objetos de forma aleatória ou por meio de alguma heurística e, em seguida, forças de atração e repulsão entre os objetos levam o sistema a um estado de equilíbrio.

Ao término das interações, cada objeto sofreu um deslocamento na direção de cada outro objeto, aproximando a distância entre eles com a distância calculada entre as conformações que eles representam. As interações são repetidas sucessivamente até o sistema atingir o equilíbrio. O número de interações pode ser definido arbitrariamente ou, então, pode-se parar o sistema através de algum critério específico. Aqui, estabeleceu-se como critério de parada a verificação de quanto o sistema está em um estado próximo ao equilíbrio, comparando interações sucessivas, ou seja, é comparada a quantidade de deslocamentos dos objetos entre interações subsequentes. Se esse valor for pequeno, abaixo de um limiar estabelecido (neste trabalho, adotou-se 10^{-5} como limiar), o sistema é considerado em equilíbrio e a projeção final é obtida, como no exemplo da Figura 3.4.

Figura 3.4: Exemplo de visualização em 2D. No lado esquerdo da figura, tem-se a visualização da superfície de enovelamento da proteína *ww domain*. Cada ponto neste gráfico refere-se a uma conformação. No lado direito, estão representadas quatro conformações (em vermelho), e suas respectivas distâncias multidimensionais ($\delta^{k,l}$). Uma estrutura exemplo de cada uma delas está representada sobreposta com a estrutura nativa da proteína (em verde).



Fonte: Extraído da Ref. [65].

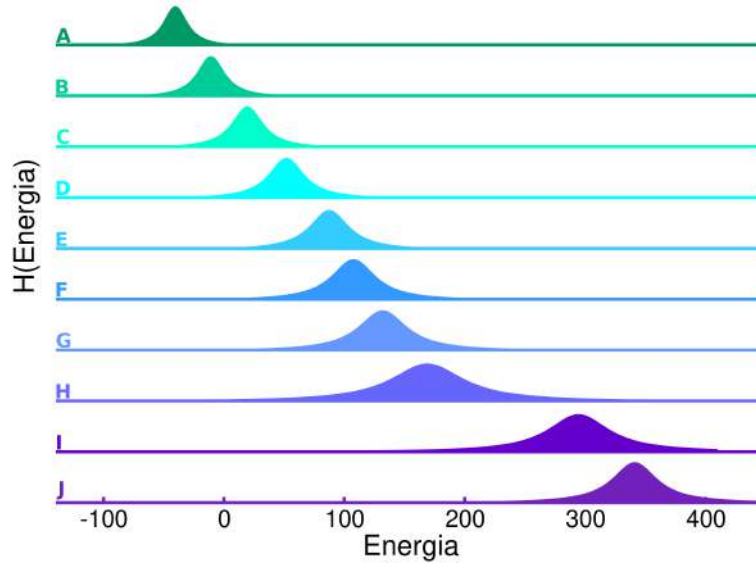
3.9 Método dos Múltiplos Histogramas

O método dos múltiplos histogramas (WHAM) é um método estatístico que tem como foco levantar as informações termodinâmicas sobre o sistema em análise. Foi originalmente desenvolvido por Ferrenberg e Swendsen [67] e têm sido amplamente utilizado em vários sistemas para o cálculo da energia livre [68, 69]. A estrutura básica do WHAM tem a temperatura como parâmetro variável, com isso, um conjunto de conformações são gerados para cada temperatura, onde cada uma irá receber o índice n . Posteriormente, os histogramas $H_n(E)$ são construídos, para exemplificar, aqui a energia foi utilizada como coordenada de reação.

Na Figura 3.5, têm-se os histogramas da proteína R15 da α -espectrina em diferentes temperaturas. Verifica-se que com o aumento da temperatura, aumenta-se a sobreposição desses histogramas, ao passo que, o mesmo comportamento é visto em baixas e altas temperaturas. Já em temperaturas intermediárias, nota-se um aumento na distribuição e na sobreposição desses histogramas. Esse comportamento indica que nessas temperaturas ocorre uma mudança de estado termodinâmico, ou seja, a proteína encontra-se próxima da temperatura de transição. Assim, vê-se que o método WHAM é uma técnica de combinação de histogramas, onde requer que haja uma sobreposição significativa dos histogramas vizinhos.

Para um determinado conjunto N de simulações com C_n configurações para a temperatura $T_n = (k_B \beta_n)^{-1}$, a probabilidade da distribuição de energia é dado pela Equação 3.16:

Figura 3.5: Histogramas produzidos em função da energia potencial total para o domínio R15 da α -espectrina em diferentes temperaturas. Os valores das temperaturas variam de 70(A) a 160(J), com intervalos iguais a 10, em unidades do Gromacs. As simulações foram realizadas com o modelo SBM-C α .



Fonte: Elaborada pelo autor.

$$P(E, \beta) = \frac{\sum_{n=1}^N H_n(E) \exp(-\beta_n E)}{\sum_{n=1}^N C_n(E) \exp(-\beta_n E + f_n)} \quad (3.16)$$

o termo f_n é a energia livre adimensional no inverso da temperatura, $\beta_n = (k_B T_n)^{-1}$, para a n -ésima simulação. Logo, se $P(E, \beta)$ é conhecido, a energia livre pode ser calculada conforme Equação 3.17:

$$\exp(-f_n) = \sum_E P(E, \beta_n). \quad (3.17)$$

Como a expressão para $P(E, \beta_n)$ envolve as energias livres, f_n , a Equação 3.17 precisa ser resolvido de forma autoconsistente para os valores de energia livre, f_n para cada valor de T_n . Usando as energias livres, a densidade dos estados pode ser calculada de acordo com Equação 3.18:

$$\Omega(E) = H_n(E) \exp(\beta_n E - f_n), \quad (3.18)$$

onde os termo $H_n(E)$ representa os histogramas em cada temperatura T_n .

Em seguida, a partir da combinação dos histogramas, é possível obter a

função de partição:

$$Z = \sum_E \Omega(E) \exp(\beta E). \quad (3.19)$$

Por fim, com a função de partição (Equação 3.19), pode-se calcular a energia livre de Helmholtz:

$$F = -k_B T \ln(Z). \quad (3.20)$$

Cada configuração simulada contribui para a propriedade média calculada. Além disso, este formalismo permite calcular as propriedades termodinâmicas em qualquer temperatura dentro da faixa simulada (não apenas para os valores do conjunto simulado). Isto é muito útil porque muitas vezes não conhecemos a temperatura de interesse. No entanto, o método WHAM pode ser usado para calcular o calor específico e a temperatura de interesse pode ser determinada pelo pico de sua curva. Então, usando WHAM novamente, podemos calcular as propriedades de equilíbrio do sistema, embora as simulações não estejam na temperatura desejada. Vale lembrar, que a sobreposição dos histogramas vizinhos devem ocorrer, por isso, o conjunto de temperaturas deve ser escolhido cuidadosamente.

3.10 Detalhes da Simulação

Todas as simulações foram realizadas utilizando o pacote de dinâmica molecular Gromacs versão 4.5.5, com os arquivos de entrada obtidos no software *SMOG webtool* [38] e o mapa nativo da proteína calculado pelo método *Shadow Contact Map* [70]. O intervalo de tempo foi definido em 0,5 fs e a simulação foi realizada com 5×10^9 passos, no qual uma conformação foi salva a cada 5000 passos de simulação. Para manter a simulação acoplada a um banho térmico, foi empregado o algoritmo do termostato de Berendsen, com a constante de acoplamento igual a 1 ps [71].

3.10.1 Análise das Simulações

Para acompanhar o processo de enovelamento é necessário escolher uma coordenada de reação e, neste trabalho, foi escolhido a fração dos contatos nativos, (Q). Portanto, dois contatos nativos, i e j , são formados quando a distância entre eles é menor que $1,2d_{ij}$, com d_{ij} sendo a distância entre dois C α na estrutura nativa,

com $j > i + 3$. O cálculo das propriedades termodinâmicas foi realizado por meio do conjunto de várias simulações realizadas sob diferentes temperaturas, obtendo, assim, a energia livre, $F(Q)$, em função dos contatos nativos, e o calor específico, C_v , por meio do Método dos Múltiplos Histogramas (do inglês *Weighted Histograms Analysis Method* - WHAM) [68]. Para o cálculo do tempo médio de primeira passagem (do inglês *Mean First Passage Time* - MFPT), as simulações foram inicializadas em temperaturas onde a proteína encontra-se desenovelada. Foram necessários 25×10^3 passos e, a cada 50 passos, salva-se uma estrutura da proteína. Em seguida, foram realizadas novas simulações que utilizaram como estrutura inicial as coordenadas da última conformação salva da simulação anterior. As simulações nessa etapa foram realizadas em temperaturas inferiores, onde a simulação foi interrompida quando a proteína atingiu 80% dos contatos nativos. Os parâmetros utilizados nesta simulação foram 10^8 passos de simulação, sendo que a cada 2500 passos uma estrutura foi salva. Para cada temperatura, foram realizadas 1000 simulações, onde cada simulação foi realizada na temperatura de enovelamento da proteína selvagem de interesse. O cálculo do tempo de enovelamento é realizado de acordo com a equação 3.21:

$$\tau = P_i \times P_{traj} \times 0,0005, \quad (3.21)$$

onde P_i é quantidade de passos simulados ao atingir 80% dos contatos nativos, P_{traj} é quantidade de passos estabelecidos para salvar uma trajetória e 0,0005 corresponde ao incremento no intervalo de tempo.

Capítulo 4

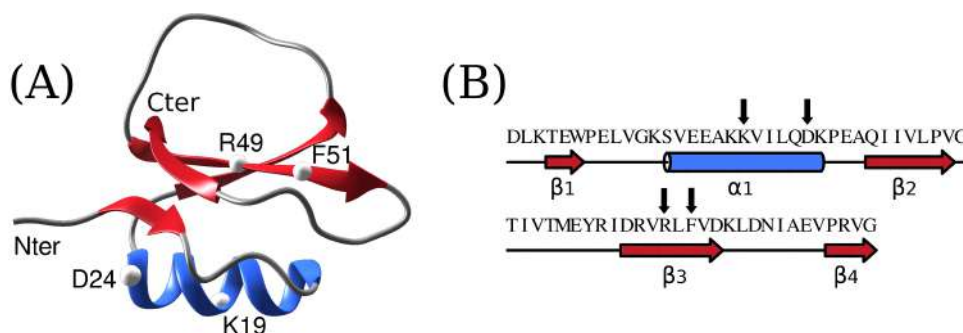
Efeitos Mutacionais e Otimização das Interações Não Nativas na Proteína CI2

4.1 Introdução

O aumento da estabilidade das proteínas por meio de mutações é uma tarefa difícil e vem desafiando cientistas há décadas. A sua aplicação é ampla e pode ser empregada em diversas áreas, desde o campo médico ao industrial [72–75]. O desenvolvimento de novos fármacos à base de proteínas e o uso de enzimas com uma termostabilidade aprimorada para seu emprego, ou seja, para realizar sua catálise em melhores condições, são alguns exemplos dessas aplicações [74–78]. Essa relevância no cenário científico, médico e industrial motivou o desenvolvimento de preditores de mutações que variam desde simulações moleculares baseadas em campos de força a algoritmos baseados em aprendizado de máquina (do inglês *machine learning* - ML) [53, 79–84]. Embora existam avanços nessas ferramentas preditivas, questões como o ajuste excessivo de dados nessas previsões ainda são recorrentes e árduos, destacando o fato de que alguns efeitos mutacionais permanecem obscuros e de difícil entendimento [85, 86]. Portanto, nesse cenário, entender a influência de cada aminoácido no enovelamento e na estabilidade da proteína é fundamental para o desenvolvimento de novas ferramentas predicionais, assim como, possibilitar um melhor entendimento na engenharia de proteínas.

No estudo realizado, pelo autor, com três domínios específicos da espectrina α (R15, R16 e R17), foi observado como as interações não-nativas modulam as taxas de enovelamento desses domínios [44]. Os potenciais adicionados ao modelo SBM- C_α foram

Figura 4.1: Representação da estrutura terciária (A), secundária e sequência primária (B) da proteína CI2. As folhas β_1 , β_2 , β_3 e β_4 estão representadas em cor vermelha, α_1 -hélice em cor azul e as regiões não estruturadas em cinza. Os resíduos destacados por uma esfera de cor branca (K19, D24, R49 e F51), são os resíduos de aminoácidos que quando substituídos por resíduos hidrofóbicos ou carregados aceleram o tempo de enovelamento. Já os resíduos indicados por uma seta na sequência primária, correspondem aos mesmos resíduos destacados na estrutura terciária. O desenho da estrutura tridimensional foi obtido pelo software CHIMERA [87] e a estrutura secundária e a sequência primária pelo software ALINE [88]



Fonte: Elaborada pelo autor e extraído da ref. [89].

aqueles que levam em consideração os efeitos hidrofóbicos e eletrostáticos. Esse tipo de abordagem é, atualmente, amplamente estudado por grupos científicos para investigar diferenças no processo de enovelamento quando realiza-se uma mutação, ou seja, uma substituição de um resíduo carregado por um hidrofóbico, como também, o contrário [90–93]. Portanto, proteínas com mutações específicas que podem ocasionar ou não um desempenho favorável nas propriedades termodinâmicas e cinéticas tornaram-se alvo de estudo para a implementação do modelo. Nesse caso, mais precisamente, o inibidor da quimotripsina 2 (CI2) foi o escolhido, pois essa proteína é uma das mais estudadas na literatura e sua gama de resultados experimentais (termodinâmico e cinético) torna possível a comparação direta com os resultados computacionais. Além disso, existem muitos resultados experimentais sobre mutações para acelerar o enovelamento da CI2, nos quais mutações únicas diminuíram a estabilidade da proteína apenas substituindo grupos carregados ou hidrofóbicos.

A CI2 é uma proteína monomérica de 65 resíduos com um domínio único, no qual a estrutura terciária consiste em uma α -hélice hidrofóbica e quatro folhas β (vide Figura 4.1) [94]. O processo de enovelamento da CI2 ocorre por meio de dois estados (desenovelado e enovelado) com um estado de transição, no qual a α -hélice está quase intacta, enquanto as folhas β estão desestruturadas [95, 96]. Estudos de engenharia de proteínas demonstraram que os efeitos da mutação também podem aumentar ou

diminuir sua termoestabilidade [97, 98]. Além disso, essas mutações intensificaram a interação local na α -hélice, eliminando algumas interações desfavoráveis em uma região específica, incluindo um grupo de argininas que encontram-se nas folhas β três e quatro. No presente estudo, investigamos esses detalhes com simulações de dinâmica molecular, utilizando o modelo baseado em estrutura $C\alpha$ (SBM- $C\alpha$), com os potenciais não-nativos, hidrofóbico e eletrostático. As mutações que aceleram o enovelamento (K19R, D24A, R49F e F51L, vide Figura 4.1) foram combinadas entre si para fornecer diferentes CI2s. Portanto, um exame aprofundado dessas substituições esclarecerá questões que ainda não foram respondidas com relação aos efeitos cooperativos, discutidas em termos de termoestabilidade e parâmetros cinéticos.

4.2 Objetivo

A proteína CI2 na sua forma selvagem, sem nenhuma mutação, é uma proteína termodinamicamente estável. Na literatura é possível encontrar diversas mutações que aumentam sua taxa de enovelamento. Entretanto, no que diz respeito a sua estabilidade com as respectivas mutações, poucas informações encontram-se disponíveis. Portanto, nesse trabalho, buscamos analisar a termoestabilidade de um grupo de mutações selecionadas, que apresentam dados experimentais disponíveis na literatura, tendo como finalidade estudar sua termoestabilidade. Essas análises foram realizadas por meio do software TKSA-MC e simulações computacionais.

A busca por novas mutações, com alta eficiência, é um desafio na otimização de enzimas. Dessa forma, não iremos gerar novas mutações, mas sim, combinações de mutações já existentes que apresentam um alto desempenho, com isso, novas versões de CI2s poderão ser geradas e, posteriormente, analisadas por meio de simulações computacionais e métodos estatísticos para obter parâmetros termodinâmicos.

As simulações serão realizadas com o modelo SBM-C α com adição dos potenciais não-nativos, eletrostático e hidrofóbico. Vale lembrar que as mutações envolvem resíduos carregados e hidrofóbicos, mesma abordagem realizada no estudo dos domínios da α -espectrina. Entretanto, o foco principal será averiguar se o mesmo comportamento é observado na CI2, se com a adição de interações não-nativos, o efeito observado será cooperativo, em outras palavras, investigar se com a diminuição no tempo de enovelamento, será obtido um aumento na termoestabilidade.

4.3 Resultados e Discussão

4.3.1 CI2 - Mutações

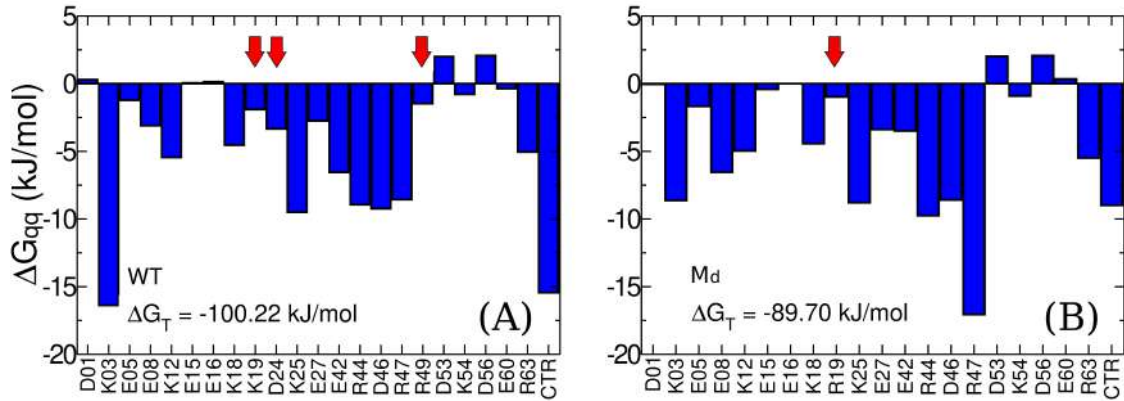
Neste trabalho, foram realizados certas combinações que envolvem quatro mutações (K19R, D24A e R49F/F51L) que aceleram o tempo de enovelamento da CI2. Todos as CI2s mutantes projetadas (M_A , M_B , M_C e M_D) foram geradas combinando certas mutações mencionadas anteriormente [97, 98]. Nessas mutações, foram substituídos grupos carregados aminoácidos, positivo ou negativo, como também, carregados por hidrofóbicos e vice-versa. Assim, têm-se: M_A - K19R/D24A, M_B - K19R/R49F/F51L, M_C - D24A/R49F/F51L e M_D - K19R/D24A/R49F/F51L. Todas as mutações foram realizadas por meio do software Modeller [99].

4.3.2 Interação Eletrostática e a Estabilidade da Proteína CI2

Dentre todas as interações responsáveis por manter a termoestabilidade da proteína, a interação eletrostática é a dominante em aumentar, manter ou diminuir a sua estabilidade [93, 100–103]. Com o auxílio da ferramenta TKSA-MC, foi possível calcular a contribuição da energia livre eletrostática de cada resíduo ionizável da proteína, ΔG_{qq} . Na Figura 4.2, os perfis de energia livre eletrostáticos da CI2-WT e sua forma mutada, CI2- M_D , são apresentados. A forma mutada apresenta menor energia livre eletrostática total, com $\Delta G_T = -89,70$ kJ/mol. O ΔG_T é igual ao somatório da contribuição eletrostática de todos os resíduos, ΔG_{qq} , e quanto mais negativo esse valor, mais estável a proteína tende a ser. Logo, espera-se que a temperatura de enovelamento aumente conforme esse valor se torne mais negativo. Já na Figura 4.3, são apresentados todos os perfis eletrostáticos, ΔG_T , seguindo a ordem da mais para o menos estável.

A CI2-WT já é uma proteína eletrostaticamente estável, com quase todos os resíduos ionizáveis contribuindo para estabilizar sua estrutura nativa (vide Figura 4.2A). Os únicos resíduos que apresentam uma contribuição positiva significava são os resíduos D53 e D56. Contudo, em sua versão mutada, a CI2- M_D possui o maior valor de ΔG_T dentre todas as mutações geradas neste trabalho, apresentando uma diferença de 11,0 kJ/mol quando comparada com a CI2-WT. Esse aumento na energia livre eletrostática é causado pela substituição de três resíduos carregados, K19, D24 e R49, por um outro resíduo, K19R, e por dois resíduos hidrofóbicos, D24A e R49F. Portanto, pode-se observar que antes das substituições, esses resíduos contribuíam de forma favorável para manter a estabilidade da CI2-WT, ao passo que, após serem

Figura 4.2: Energia livre eletrostática, ΔG_{qq} , para a proteína CI2-WT (A) e CI2-M_D (K19R/D24A/R49F/F51L) (B). $\Delta G_{qq} > 0$ indica que o resíduo apresenta energia livre eletrostática desfavorável para estabilizar o estado nativo, porém seu valor de SASA < 50%. As setas vermelhas indicam quais são os candidatos que sofrerão mutações. Todos os resultados foram calculados para o pH = 7,0 e obtidos pelo software TKSA-MC [53].



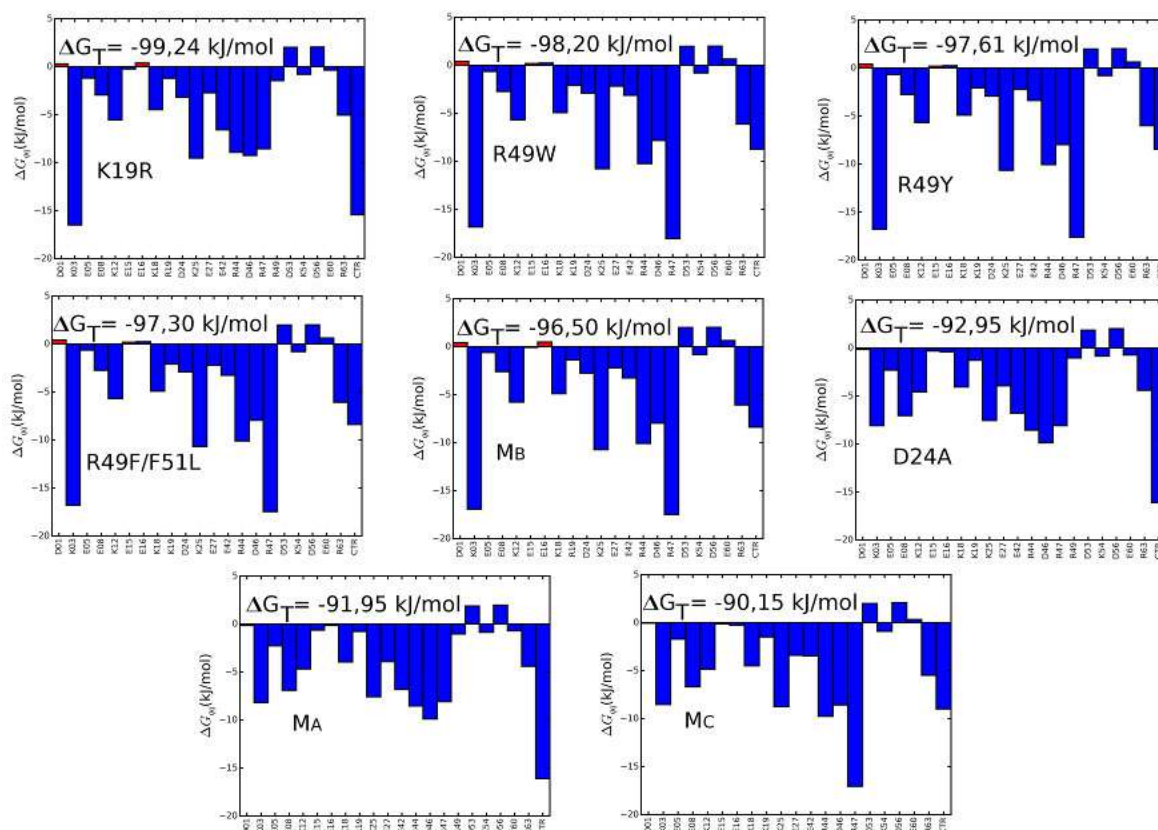
Fonte: Elaborada pelo autor e extraído da Referência [89].

substituídos, ocorre uma diminuição na estabilidade do mutante, CI2-M_D, como pode ser visto na Figura 4.2B. Os resíduos carregados antes e depois das mutações estão indicados por setas vermelhas nas Figuras 4.2A e 4.2B.

Na Figura 4.4A, encontram-se os valores da energia livre eletrostática, ΔG_T , em função da temperatura de enovelamento, T_f , para CI2-WT e suas respectivas mutações. Devido à falta de dados experimentais, análises referentes à energia livre eletrostática versus temperatura de enovelamento não serão apresentados neste trabalho. Portanto, apenas os resultados teóricos computacionais constam na Figura 4.4A. As mutações propostas e estudadas experimentalmente encontram-se em cor laranja e as mutações propostas estão na cor azul. O valor de T_f computacional de cada proteína pode ser determinado pelo pico do calor específico obtido pelo método WHAM. Este método utiliza um conjunto de simulações, que nesse trabalho, foram realizadas com o modelo SBM-C α com a adição de potenciais não-nativos, eletrostático e hidrofóbico, Capítulo 3.

Foi observado que o aumento de ΔG_T tende a promover uma diminuição em T_f , com coeficiente de correlação igual a -0,91 (vide Figure 4.4A). Contudo, o método TKSA-MC leva em consideração apenas o efeito eletrostático na estrutura nativa da proteína, enquanto nossas simulações consideram duas interações não-nativas, hidrofóbicas e eletrostáticas, que são calculadas em todo o processo de enovelamento. A notável correlação entre ΔG_T e T_f , indica que a interação eletrostática pode ser

Figura 4.3: Energia livre eletrostática, ΔG_{qq} , para todas as mutações realizadas na CI2. A sequência das figuras segue a ordem decrescente, em módulo, da energia: K19R, R49W, R49Y, R49F/F51L, M_B - K19R/R49F/F51L, D24A, M_A - K19R/D24A, M_C - D24A/R49F/F51L. Os resultados foram obtidos para o pH = 7,0. $\Delta G_{\text{qq}} > 0$ indica que o resíduo apresenta energia livre eletrostática desfavorável para estabilizar o estado nativo, com SASA < 50%. Já as barras em vermelho, apresentam SASA $\geq 50\%$, tornando esses resíduos candidatos a sofrerem mutações



Fonte: Elaborada pelo autor e extraído da Referência [89].

compreendida como a interação dominante para mantém a termoestabilidade da CI2 e a substituição de resíduos carregados por hidrofóbicos causaram a sua desestabilização.

4.3.3 Efeitos Mutacionais na Cinética de Enovelamento

Segundo os dados apresentados por Ladurner [97], os resíduos K19 e D24, que estão localizados na região C terminal da α -hélice, são resíduos essenciais para manter a estabilidade da estrutura nativa, porém o resíduo D24 apresentou o principal efeito desestabilizador na α -hélice isolada. Por esse motivo, os resíduos K19 e D24 foram substituídos por arginina (R) e alanina (A), respectivamente. Os resulta-

dos dessas mutações preservaram a formação da α -hélice e aceleraram seu tempo de enovelamento. Por outro lado, eles diminuíram a estabilidade de toda a proteína.

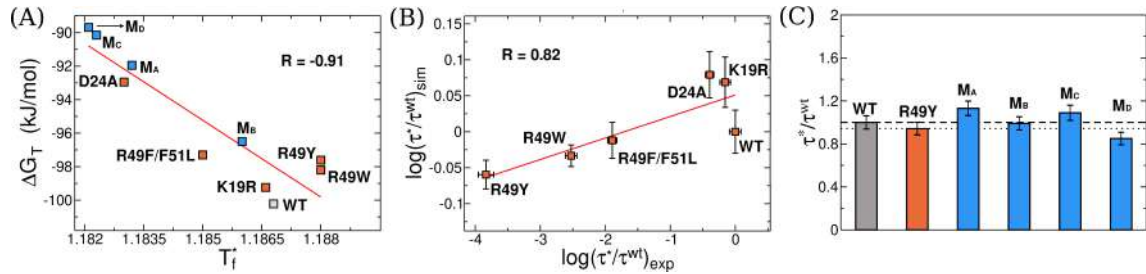
Outra região crítica na proteína CI2-WT inclui várias argininas (R47, R49 e R63), onde seu arranjo promove uma repulsão eletrostática entre seus grupos carregados. No entanto, Ladurner [97] e Lawrence [98] mostraram em seus estudos que essa região é altamente conservada, esta análise foi feita a partir de um grupo de CI2s obtidas de diferentes organismos. Com isso, mutações foram realizadas nessa região para o resíduo R49, o substituindo por grandes grupos hidrofóbicos. O resultado dessa substituição acarretou em uma notável redução no tempo de enovelamento [97, 98]. Além disso, o R49 forma interações relevantes com outros resíduos, em especial, no estado de transição com o resíduo F51.

Devido a esses efeitos particulares observados na proteína CI2, foram investigados os parâmetros cinéticos com foco no tempo de enovelamento para a CI2-WT e suas mutações: K19R, D24A, R49F/F51L, R49W e R49Y. Essa investigação foi realizada por meio de simulações computacionais utilizando o modelo descrito na Seção 3.1 com adição de potenciais não-nativos, hidrofóbico e eletrostático. Na Figura 4.4B, tem-se o logaritmo da razão do tempo de enovelamento para os dados experimentais versus os dados computacionais. A comparação entre os dados, experimental e computacional, apresentaram uma correlação linear de 0,82. Portanto, os resultados evidenciam que nossa abordagem encontra-se em boa concordância qualitativa, sendo que apenas dois pontos foram observados fora de correlação, WT e K19R.

O resultado mencionado no parágrafo anterior foi realizado apenas para os dados experimentais disponíveis na literatura. As mutações propostas nesse trabalho não apresentam dados experimentais, contudo, os dados cinéticos para essas mutações, M_A , M_B , M_C , e M_D , são discutidos em comparação com os dados experimentais de cinética da proteína CI2-WT e da mutação com melhor desempenho, R49Y. Na Figura 4.4C, encontram-se os resultados do tempo de enovelamento, no qual as mutações propostas, M_A e M_C , não apresentaram resultados significativos, ou seja, não aceleram o seu enovelamento, mas o desaceleram. Diferente dos resultados de M_A e M_C , a mutação M_B apresenta aproximadamente o mesmo tempo de enovelamento da CI2-WT. Já o cálculo do tempo de enovelamento da M_D apresentou uma diminuição considerável dentre todas as mutações projetadas, como também exibiu um valor inferior ao da mutação R49Y, mutação que apresenta o melhor desempenho.

O conjunto de mutações que foram apresentadas nesse trabalho forneceram informações cruciais a respeito das combinações de mutações em relação ao tempo de enovelamento e sua termoestabilidade. Também foi observado, no conjunto de versões

Figura 4.4: Efeitos mutacionais na estabilidade, ΔG_T , e no tempo de enovelamento, τ^* , da proteína CI2. **(A)** Energia livre eletrostática total (ΔG_T) em função da temperatura de enovelamento (T_f^* - em unidades reduzidas) para cada grupo mutacional da CI2. **(B)** Tempo de enovelamento para cada versão da CI2, τ^* , normalizado pelo tempo de enovelamento da tipo selvagem (CI2-WT), τ^{wt} . A correlação linear esta representada pela linha continua em cor vermelha. **(C)** Tempo de enovelamento (τ^*) para as mutações propostas (M_A , M_B , M_C , e M_D) comparado com o tempo de enovelamento da R49Y e WT. As barras de cor azul são dados computacionais e as barras de cor laranja e cinza, são dados experimentais. Todos dados experimentais foram extraídos de Ladurner [97] e Lawrence [98].



Fonte: Elaborada pelo autor e extraído da Referência [89].

de CI2s, uma termoestabilidade mais baixa promovida pelo aumento da energia livre eletrostática. Esse comportamento não foi observado em estudos anteriores, o tornando ainda mais interessante. Com base nessa premissa, calculamos alguns parâmetros para analisar melhor os efeitos eletrostáticos e hidrofóbicos da mutação M_D .

4.3.4 Comportamento Antagônico na CI2- M_D

A proteína CI2 com a respectiva mutação M_D foi a versão projetada que apresentou o menor tempo de enovelamento dentre todas as mutações propostas, como também, em relação às mutações experimentais. Vale ressaltar que o modelo de simulação aqui proposto, SBM- $C\alpha$, leva em consideração apenas a contribuição das interações nativas durante o processo de enovelamento. Contudo, algumas limitações do modelo são observadas ao estudar os efeitos mutacionais. Por esse motivo, foram adicionados dois potenciais não-nativos ao modelo SBM- $C\alpha$, com a finalidade de mimetizar esses efeitos (eletrostáticos e hidrofóbicos).

A metodologia aplicada nesse trabalho nos permite observar os efeitos mencionados acima que, conseqüentemente, serão calculados para verificar seu respectivo peso energético. As energias hidrofóbicas e eletrostáticas foram calculadas em diferentes estágios durante o processo de enovelamento. O cálculo foi realizado para CI2-WT e M_D nos seguinte estágios: estado desenovelado ($Unf - 19 < Q < 38$), estado de

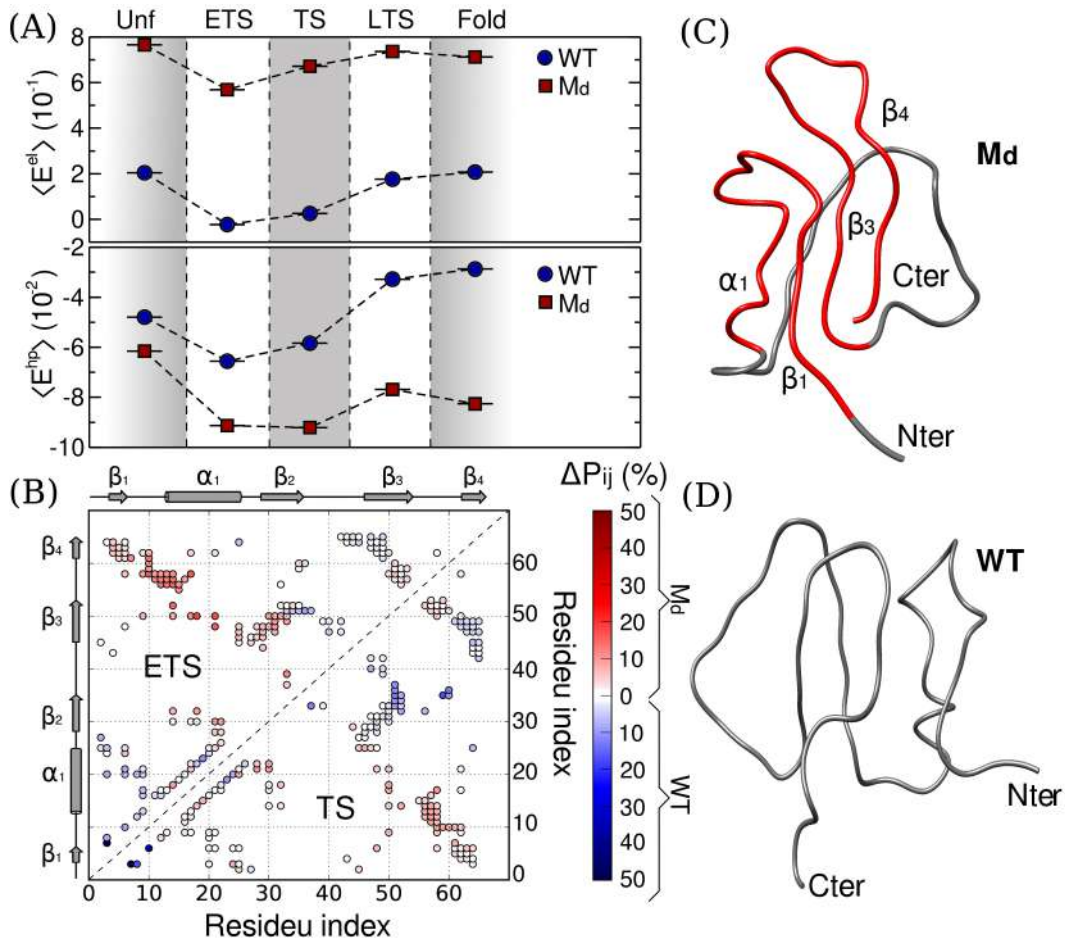
pré-transição (ETS - $39 < Q < 56$), estado de transição (TS - $57 < Q < 82$), estado pós-transição (LTS - $83 < Q < 112$) e estado enovelado (Fold - $113 < Q < 141$), como pode ser visto na Figura 4.5A.

Do ponto de vista eletrostático, a proteína CI2-WT apresenta menor energia em todos os estágios durante o processo de enovelamento quando comparada com a mutação proposta, M_D . Por outro lado, a mutação M_D apresenta menor energia hidrofóbica em todos os estágios, quando comparado com a CI2-WT. Portanto, observou-se que a energia livre eletrostática, ΔG_T (Figura 4.2), e a energia eletrostática média, $\langle E^{ele} \rangle$ (Figura 4.5A), da CI2-WT ao longo de todo o processo de enovelamento, apresentaram valores inferiores aos da mutação M_D . Desse modo, vê-se que a substituição de resíduos carregados por hidrofóbicos acarreta a diminuição da estabilidade proteica.

Com os dados obtidos a partir da Figura 4.5A, tem-se a diferença energética entre CI2-WT e M_D , $\Delta\Delta G_T = -10$ kJ/mol, e seus respectivos T_{fs} , iguais a 1,187 e 1,182 em unidades reduzidas, o que corresponde 0,6% de diferença. A barreira de energia livre da mutação M_D é menor que a barreira da CI2-WT em aproximadamente 0,4 kT, mas a combinação dos efeitos geram um tempo de enovelamento 20% menor, como pode ser visto na Figura 4.4C. A energia eletrostática no estado ETS para o LTS tende a aumentar em ambas proteínas, Figura 4.5A. Contudo, para a CI2-WT, a energia hidrofóbica aumenta monotonicamente do ETS para o TS, enquanto que para o M_D , a variação da energia hidrofóbica, $\langle E^{hp} \rangle$, oscila ao longo do processo de enovelamento. Devido a esse comportamento distinto entre estágios específicos, a probabilidade de formação de contato nativo foi calculado para aqueles estágios que apresentaram uma diferença mais evidente, ou seja, ETS e TS.

Na Figura 4.5B, tem-se a diferença de probabilidade de formação de contatos nativos, ΔP_{ij} , entre M_D e CI2-WT. No ETS (triângulo superior) e no TS (triângulo inferior), a mutante M_D apresenta seus contatos nativos em tons de vermelho. Por outro lado, em tons de azul, tem-se a probabilidade de formação para os contatos nativos da CI2-WT. Uma diferença significativa pode ser observada na probabilidade de formação da M_D entre ETS e TS. A região em vermelho em ETS corresponde ao local em que as mutações foram realizadas, α -hélice um (α_1) e folha β três (β_3). Já no triângulo inferior, TS, em sua maioria, os contatos nativos apresentam valores similares na probabilidade de formação para ambas as proteínas. Tais resultados destacam uma importância evidente para o ETS, já que o mesmo apresenta uma diferença significativa na formação dos contatos. Portanto, nossos resultados indicam que o ETS é o estágio inicial em que o processo de enovelamento da M_D sofreu um decréscimo em seu tempo de enovelamento comparado com o CI2-WT.

Figura 4.5: Análise da energia média eletrostática e hidrofóbica, assim como da diferença de probabilidade de formação dos contatos nativos no *ensemble* ETS e TS para CI2-M_D e CI2-WT. **(A)** Energia média eletrostática, $\langle E^{el} \rangle$, e hidrofóbica, $\langle E^{hp} \rangle$, para a CI2 selvagem (CI2-WT em cor azul) e para a mutação M_D (CI2-M_D em cor vermelha). Energia computada para o estado não-nativo (Unf - *unfolded*), pré-estado de transição (ETS - *early transition state*), estado de transição (TS - *transition state*), pós-estado de transição (LTS - *late transition state*) e estado nativo (Fold - *folded*) para CI2-WT e CI2-M_D. **(B)** Diferença percentual (ΔP_{ij} (%)) entre os mapas de contato da CI2-M_D e CI2-WT para o ETS (triângulo superior) e para TS (triângulo inferior). A estrutura secundária está representada em cor cinza e a barra está associada à diferença de probabilidade do mapa de contato. **(C)** e **(D)** são as estruturas representativas da CI2-M_D e CI2-WT, respectivamente, no *ensemble* ETS. A região destacada em vermelho na CI2-M_D corresponde a mesma região em vermelho do triângulo superior (ETS) da figura B.



Fonte: Elaborada pelo autor e extraído da Referência [89].

De acordo com a Figura 4.5B, sabe-se que a proteína com a mutação M_D apresenta maior probabilidade de formação de contatos nativos no ETS para certas regiões em sua estrutura. Por esse motivo, foi selecionada a estrutura representativa

desse grupo para ambas as proteínas. Desse *ensemble* de estruturas, foram selecionadas uma estrutura representativa de cada proteína, CI2-WT e da M_D , como visto nas figuras 4.5C e 4.5D, respectivamente. De acordo com a região em vermelho no mapa de contato nativo, destacamos essa região na estrutura representativa da M_D . Essa estrutura exibe uma α -hélice não estruturada, enquanto as folhas β um, três e quatro encontram-se emparelhadas, como na estrutura nativa (vide Figura 4.1). Em contraste com o observado na M_D , a estrutura representativa do tipo selvagem no ETS também apresenta uma α -hélice não estruturada, mas suas folhas β não encontram-se emparelhadas. O N e o C terminal da CI2-WT estão distantes, enquanto N e C terminal da M_D estão compactados com o restante de sua estrutura.

Com base nos dados experimentais de Ladurner [97], sabe-se que a remoção de algumas interações eletrostáticas aumenta o empacotamento das proteínas, mesmo diminuindo o tempo de enovelamento. Nossos resultados computacionais estão em concordância com dados experimentais. Uma diminuição de energia foi observada na M_D , bem como um estágio favorável no seu enovelamento foi determinado por simulação, na qual descreve um dos principais motivos pelos quais o tempo de dobragem acelera. Desse modo, a substituição de resíduos carregados por hidrofóbicos não apenas diminui o tempo de dobragem, mas também é responsável por tornar o CI2- M_D mais compacta no ETS, quando comparada com a CI2-WT.

A adição de interações não-nativas em diferentes sistemas apresentam resultados semelhantes aos apresentados aqui. Essas interações podem acelerar o enovelamento diminuindo a barreira de energia livre [43, 104]. A quantidade ideal de interações não-nativas em uma proteína depende de sua estrutura e topologia [105]. Qualquer valor além do valor ótimo de interações não-nativas leva uma proteína a um processo de enovelamento mais lento, aumentando o número de mínimos locais no ETS [106]. Esse conceito está relacionado à frustração nas proteínas e a mutação M_D é um exemplo notório desse comportamento. A remoção de algumas interações eletrostáticas pode diminuir a frustração e limitar a competição entre interações nativas e não-nativas.

4.4 Conclusão

Conhecer com maior rigor as especificidades de cada resíduo de aminoácido durante o processo de enovelamento é fundamental para o desenvolvimento e o desenho racional em engenharia de proteínas. No presente trabalho, usamos métodos computacionais que combinam interações eletrostáticas e hidrofóbicas, que foram acopladas ao modelo baseado em estrutura, com objetivo de analisar os efeitos mutacionais, cinéticos e termodinâmicos. Esse método foi e é amplamente empregado em diversos sistemas biológicos de interesse, como em nosso recente trabalho, onde foram estudados os efeitos cooperativos das interações não-nativas no processo de enovelamento dos domínios R15, R16 e R17 da α -espectrina [44].

No presente estudo, foi verificado um comportamento distinto daquele mencionado anteriormente (domínios R15, R16 e R17 da α -espectrina) ao estudar as proteínas CI2WT e CI2M_D. Algumas mutações propostas para a CI2 não apresentaram um efeito cooperativo não-nativo, mas um efeito oposto que promoveu um comportamento diferente ao se analisar sua termostabilidade e a seu tempo de enovelamento. Portanto, na mutação M_D, foi observada uma diminuição em seu tempo de enovelamento, em compensação, ocorreu um aumento no valor de ΔG_T . Esse efeito é contraintuitivo, pois mutações que melhoram a termoestabilidade geralmente estão associadas à aceleração do processo de enovelamento [93, 107].

Nossos resultados também mostraram que as substituições de resíduos carregados por hidrofóbicos tendem a promover a formação de estruturas secundárias com maior probabilidade no estágio ETS. Essas substituições também alteraram o perfil de energia eletrostática do estado nativo, o que causou a diminuição em sua termoestabilidade. Portanto, a metodologia empregada nesse trabalho foi não só capaz de fornecer informações valiosas sobre os efeitos das mutações, como também, identificar os estágios mais críticos do enovelamento relacionados à termostabilidade e a velocidade.

Capítulo 5

Estudo do Perfil Energético dos Domínios R15, R16 e R16M₅ da α -Espectrina

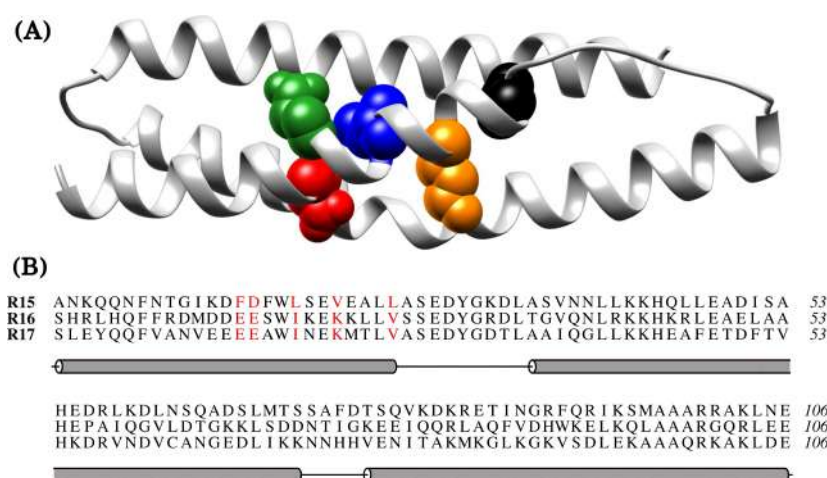
5.1 Introdução

A teoria do funil de energia tem sido amplamente empregada no estudo do enovelamento de proteínas. Por meio de análises estatísticas, a teoria fornece informações cruciais de fatores importantes do processo do enovelamento, como por exemplo: transições de fase, estabilidade termodinâmica e cinética [27, 28, 108, 109]. Modelos baseados nas ideias do funil de energia foram úteis para caracterizar o processo de enovelamento de forma quantitativa e realizar a conexão entre a teoria e o experimento [18, 19, 110, 111]. Assumindo que toda informação esteja codificada na estrutura nativa da proteína, os modelos baseados em estrutura (SBM) podem descrever informações sobre o mecanismo de enovelamento e função, assim como obter valores de barreiras de energia livre e valores ϕ , sendo este uma medida experimental de como a cinética de enovelamento das proteínas pode ser alterada realizando mutações. No entanto, existem exceções quanto ao uso desses modelos que ainda desafiam nossa compreensão e que, geralmente, servem como motivação para aprimorar nosso conhecimento.

No estudo do enovelamento de três domínios específicos da α -espectrina, foram evidenciadas diferenças significativas em suas taxas de enovelamento [112–118]. Nesse trabalho, cada domínio da α -espectrina pode ser estudado de maneira independente e os domínios analisados são identificados como R15, R16 e R17. Embora os domínios possuam alta similaridade estrutural, com RMSD < 1 Å, o mecanismo cinético e o processo de enovelamento são distintos [113], onde, notavelmente, a taxa de

enovelamento do R15 é três ordens de grandeza maior do que a de seus homólogos, com o domínio R17 sendo o mais lento de todos. A origem desta ampla diferença nas taxas de enovelamento foi atribuída às rugosidades no perfil de energia e, principalmente, ao atrito interno [118, 119].

Figura 5.1: Alinhamento das sequências dos domínios R15, R16 e R17 e a representação estrutural do domínio R16 em que os aminoácidos da primeira α -hélice que serão substituídos estão destacados. Na Figura A, representação estrutural da R16 destacando os resíduos Glu14 em vermelho, Glu15 em verde, Ile18 em azul, Lys21 em laranja e Val25 em preto. Na Figura B, alinhamento sequencial dos domínios R15, R16 e R17 com sua respectiva representação tubular, α -hélices, da estrutura secundária.



Fonte: Elaborada pelo autor.

No projeto desenvolvido anteriormente com esses domínios [44], a origem da diferença no tempo de enovelamento pode ser compreendida em termos de interações não-nativas, mais precisamente, em razão de um efeito cooperativo não-nativo que emerge das interações hidrofóbicas e eletrostáticas. Vale lembrar que vários grupos de pesquisadores já realizaram estudos envolvendo esses três domínios, porém o modelo proposto pelo autor desse trabalho foi original para o caso dos domínios R15, R16 e R17 dentre todos modelos abordados na literatura. Dessa forma, o estudo realizado capturou as diferenças observadas na taxa de enovelamento, como também, obteve uma boa correlação com os dados experimentais ao realizar determinadas mutações que foram cruciais para aumentar a taxa de enovelamento dos domínios R16 e R17 [112, 115].

Assim, de forma simultânea, foram realizadas as mutações E14F, E15D, I18L, K21V e V25L nos domínios R16 e R17. Estas mutações não só aumentaram a taxa de enovelamento, como também, influenciaram na redução da rugosidade do perfil de energia. Ademais, o mecanismo de enovelamento dos domínios R16 e R17, após as mutações, apresentaram características similares ao mecanismo de enovelamento do

domínio R15. Além disso, os valores ϕ obtidos experimentalmente para os domínios R16 e R17 foram aproximadamente iguais aos do domínio R15 [113, 120–122].

O modelo proposto de simulação foi capaz de observar tais efeitos, pois as mutações substituem resíduos carregados e hidrofóbicos por outros grupos de aminoácidos, além de que, o modelo contabiliza as interações eletrostáticas e hidrofóbicas, diferente do modelo SBM- $C\alpha$ sem adição das interações não-nativas. Devido à correlação obtida, foram utilizadas as mesmas trajetórias para o cálculo do mapa bidimensional, onde foi estudado os estado metaestáveis durante o processo de enovelamento desses domínios.

5.2 Objetivo

As diferenças na taxa de enovelamento dos domínios R15, R16 e R117 foram estudadas anteriormente pelo autor desse trabalho, onde os efeitos não-nativos, eletrostáticos e hidrofóbicos, foram evidenciados como os responsáveis pela diferença na taxa de enovelamento. Com base nesse resultado, o objetivo do trabalho será analisar, detalhadamente, as mudanças geradas no espaço de fase dos domínios R15, R16 e R16M₅ (M₅ - E14F, E15D, I18L, K21V e V25L) utilizando o modelo computacional proposto (SBM-C α +HP+Elec) para as proteínas em sua forma selvagem e com as mutações que aumentaram a taxa de enovelamento.

Para averiguar as mudanças no espaço de fase, o método ELViM será utilizado, pois com as trajetórias obtidas é possível calcular, por meio desse método, o mapa bidimensional, assim como, a densidade de probabilidade onde investigaremos a possível existência de um estado metaestável que pode ser interpretado como uma região de armadilhamento, bem como, uma região facilitadora para o enovelamento.

Como já mencionado, as simulações foram realizadas por meio do modelo SBM-C α , com a adição de dois ponteciais não-nativos, hidrofóbico (HP) e eletrostático (Elec). Dessa forma, será calculado o mapa de contato nativo e não-nativo com o objetivo de estudar os efeitos cooperativos e não cooperativos que as interações podem apresentar, mediante uma região de interesse evidenciada.

5.3 Resultados e Discussão

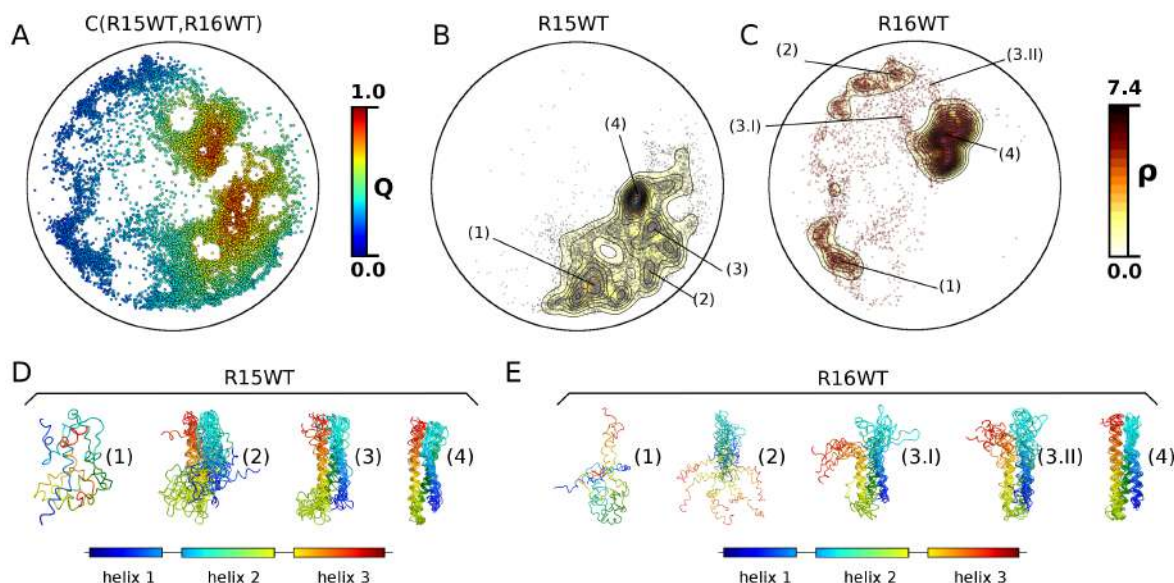
A partir das simulações com o modelo SBM- $C\alpha$ com adição dos potenciais não-nativos, eletrostático e hidrofóbico, foram calculadas as projeções bidimensionais por meio do software ELViM (Seção 3.8) para os domínios R15, R16 e R16M₅. Primeiramente, as projeções foram realizadas para cada domínio com diferentes pontenciais, por exemplo: com o potencial SBM- $C\alpha$, SBM- $C\alpha$ + Elec (potencial eletrostático), SBM- $C\alpha$ + HP (potencial hidrofóbico) e, por fim, com o modelo SBM- $C\alpha$ + HP + Elec. Contudo, somente serão apresentadas as projeções com o último modelo mencionado (SBM- $C\alpha$ +HP+Elec), pois conforme observado por Silva, et al. [44], os resultados do cálculo do tempo de enovelamento corroboraram de forma considerável com os dados experimentais disponíveis na literatura.

A trajetória selecionada para o cálculo da projeção foi a trajetória que encontra-se na temperatura de enovelamento. No total, cada simulação apresenta 10^6 estruturas, mas foi considerado apenas 10^4 estruturas. Essa redução no número de estruturas foi devido ao alto custo computacional para obter as projeções. Com a finalidade de observar as diferenças existentes na rota de enovelamento dos domínios, foi realizado o cálculo da projeção combinando duas trajetórias de interesse, como por exemplo, as trajetórias das proteínas na sua forma selvagem, R15WT e R16WT, e com a mutação, R16WT e R16M₅, respectivamente. No caso onde combinamos duas trajetórias, se houver diferenças entre elas, o ELViM irá apresentar, possivelmente, diferentes distribuições no espaço bidimensional. Caso contrário, quando duas trajetórias são muito parecidas ou idênticas, elas irão compartilhar, aproximadamente, do mesmo espaço de fase.

5.3.1 Rota de enovelamento da R15WT e R16WT

Na Figura 5.2A, tem-se a projeção dos domínios R15WT e R16WT, com a escala de cor correspondendo à fração dos contatos nativos. Além disso, a fração dos contatos nativos foi normalizada, variando de 0 (estrutura totalmente desenovelada) até 1 (estrutura totalmente enovelada). Com a projeção combinando duas trajetórias, C(R15WT, R16WT), foi possível analisar se ambas compartilham ou não o mesmo espaço de fase. A projeção obtida para esses dois domínios exibiram um conjunto de estruturas que encontram-se próximas no espaço de fase no estado desenovelado, $Q < 0,2$, em lados opostos no espaço de fase na transição, $0,4 < Q < 0,8$, e próximas no estado nativo, $Q > 0,8$. Vale ressaltar que o conjunto de pontos na parte superior da projeção corresponde à trajetória do R16WT e na parte inferior ao do R15WT.

Figura 5.2: (A) Visualização da trajetória de enovelamento da C(R15WT, R16WT) no espaço de fase 2D obtido pelo ELViM. A barra de cor está associada à fração de contatos nativos, Q , normalizada de 0 a 1. (B) e (C) mapa de densidade bidimensional para os domínios R15WT e R16WT, respectivamente. A barra de cor corresponde a densidade de estados, ρ . Em (A), (B) e (C), cada ponto representa uma conformação e a distância euclidiana entre os pontos está associada a uma distância relativa entre as estruturas. (D) e (E), *ensemble* de estruturas obtidas a partir dos pontos indicados em (B) e (C), onde o estado desenovelado e nativo estão indicado em (1) e (4).



Fonte: Elaborada pelo autor.

Com o espaço de fase 2D, pode-se analisar cada conjunto de dados separadamente, como mostrado nas Figuras 5.2B e 5.2C para os domínios R15WT e R16WT, respectivamente, com o cálculo do mapa de densidade. Com esse cálculo foi possível observar detalhadamente as rotas de enovelamento que cada domínio apresenta. Para o domínio R15WT, com o espaço de fase bidimensional e o mapa de densidade, quatro clusters principais foram identificados e seu conjunto de estruturas encontram-se na Figura 5.2D. O conjunto de estruturas em (1) representam o estado desenovelado, (2) têm-se o estado de transição, caracterizado pela formação simultânea de interações secundárias e terciárias. Após o estado de transição (3), ocorre a estruturação correta das hélices, assim como, uma notável flutuação na região não estruturada com o N e C terminal da proteína. Por fim, o domínio R15WT atinge seu estado nativo (4). De acordo com os resultados experimentais [123], o R15WT enovela-se por meio de um mecanismo denominado *nucleation – condensatio*, onde apenas os centros das duas hélices (hélices 1 e 3) encontram-se estruturados e ancorados. A partir do conjunto de estruturas obtidas (Figura 5.2D), nossos resultados computacionais revelaram que

o mecanismo de enovelamento do domínio R15WT seguiu o mesmo comportamento observado experimentalmente.

Por outro lado, o domínio R16WT enovela-se com três ordens de grandeza mais lento do que o domínio R15WT [44, 113] e apresenta um mecanismo de enovelamento conhecido como *framework – like*, onde ocorre o encaixe das hélices 1 e 3 parcialmente pré-formadas. Além disso, esse mecanismo está associado a um enovelamento mais rápida [124], mas este não é o caso do domínio R16WT. Na Figura 5.2C, com o mapa de densidade do R16WT, observa-se que sua trajetória encontra-se distribuída de forma diferente da R15WT, Figure 5.2B. Ademais, foram observadas duas rotas possíveis que leva esse domínio do estado desenovelado ao estado nativa. Esses caminhos moleculares apresentam diferentes conjuntos de estruturas, como por exemplo: (i) *ensemble* (1) e (2), estados desenovelados; (ii) estado de transição, indicado em (3.I) e (3.II), duas rotas possíveis; (iii) estado nativo, *ensemble* (4). As conformações em (3.I) e (3.II) apresentaram estruturas diferentes, sendo que em (3.I) a hélice 3 não encontra-se em sua posição nativa e a região entre as hélices 1 e 2 apresenta um alto grau de flutuação. Já as conformações em (3.II) apresentaram conformações próximas da conformação nativa.

Os mecanismos de enovelamento e a rotas assumidas pelos domínios são significativamente diferentes, resultando no enovelamento mais rápido para o R15WT em relação ao R16WT. Em geral, a diferença na taxa de enovelamento entre o R15WT e o R16WT surge dos efeitos cooperativos das interações não-nativas, como consta em nosso estudo anterior [44]. Nas Figuras 5.2B e 5.2C, como resultado desses efeitos cooperativos, as diferenças no espaço de fase bidimensional exibem aspectos intrínsecos durante o processo de enovelamento desses domínios. O R16WT pode assumir duas caminhos possíveis no estado de transição, um mais estruturada que a outra. Nesse caso, uma das rotas pode ser interpretada como uma armadilha (Figura 5.2E (3.I)), fazendo com que a proteína gaste mais tempo procurando o encaixe correto das hélices, conforme mostrado nas Figuras 5.2C e 5.2E. No entanto, o R15WT enovela-se através de uma rota direcionada, caracterizada por um espaço de fase bidimensional mais compactado, assumindo configurações menos aleatórias, como pode ser visto nas Figuras 5.2B e 5.2D.

Portanto, nesse estudo foi possível observar, do ponto de vista computacional, como as interações não-nativas afetam de forma direta o processo de enovelamento desses domínios, onde o domínio R16WT apresenta um espaço de fase bidimensional com regiões mais visitadas, enquanto que o domínio R15WT exibe um espaço de fase mais reduzido, o que caracteriza sua baixa barreira de energia livre durante seu processo de enovelamento.

5.3.2 Otimização das interações não-nativas e a diminuição no tempo de enovelamento do domínio R16MM₅

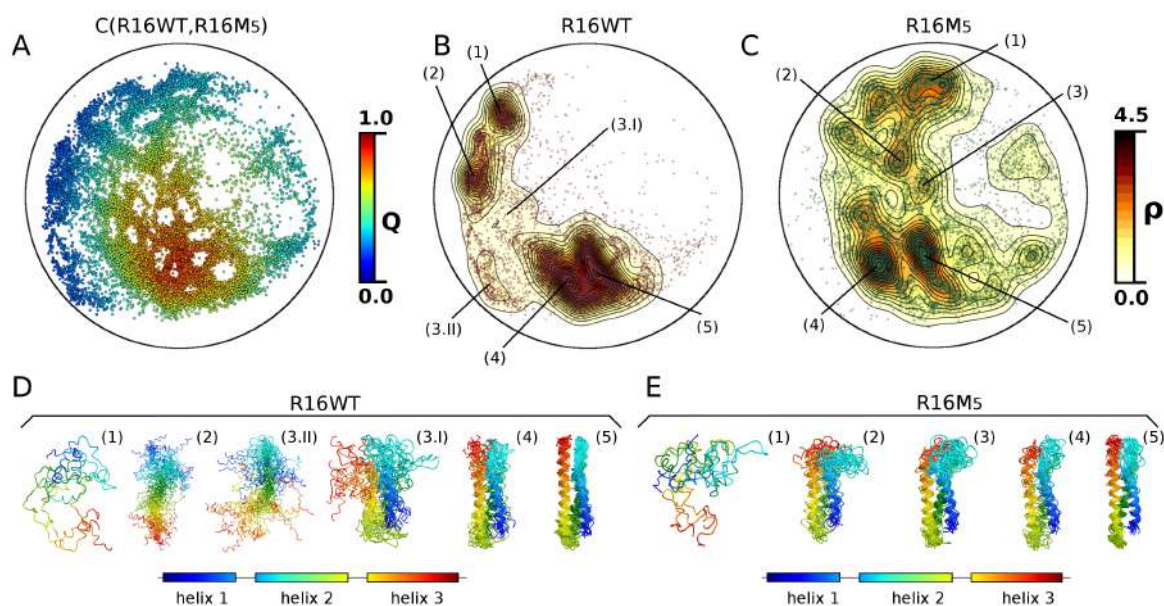
De acordo com os dados experimentais, foram projetados cinco mutações, resíduos da R15WT foram introduzidos na R16WT a fim de acelerar o seu enovelamento. As mutações foram realizadas substituindo resíduos hidrofóbicos e carregados, $R16M_5 = E18F + E19D + I22L + K25V + V29L$. A partir das mutações propostas, foi constatado que o R16M₅ diminuiu significativamente o seu tempo de enovelamento. Além disso, resultados experimentais demonstraram que o mecanismo de enovelamento do R16M₅ foi alterado promovendo o encaixe correto da hélice 3.

Na Figura 5.3, as trajetórias de enovelamento dos domínios R16WT e R16M₅ foram combinadas, C (R16WT, R16M₅). Segundo o espaço de fase bidimensional e o mapa de densidade individual de cada domínio, foi evidenciado diferenças cruciais a partir das substituições mencionadas no parágrafo anterior. Mesmo que o R16WT e o R16M₅ apresentem uma superposição no espaço de fase 2D (vide Figura 5.3A), o mapa de densidade de cada proteína revela uma variação de densidade significativa entre os estados desenovelado e nativo, como visto nas Figuras 5.3B e 5.3C. Os conjuntos de estruturas visitadas pelo R16WT é caracterizado por dois *ensembles* desenovelados e um nativo, ambos altamente populados.

Nas Figuras 5.3B e 5.3D, a partir de (1) e (2), têm-se o *ensemble* de estruturas correspondentes a essa região, onde sua estruturação segue com a formação das estruturas secundárias e posteriormente, ocorre as interações intercadeias. O conjunto (3.II) foi identificado igualmente ao estado observado na Figura 5.2C. As conformações características do estado de transição, Figura 5.3B e 5.3D em (3.I), corresponde ao conjunto de estruturas observado na Figura 5.2C (3.I). Esse resultado demonstra um conjunto representativo de estruturas que podem fazer parte de uma armadilha durante o seu processo enovelamento. Em ambos os casos, a hélice 3 e a região entre as hélices 1 e 2 encontram-se altamente frustradas. O estado enovelado foi caracterizado por dois grupos estruturais, onde o grupo (4) apresenta uma alta flutuação quando comparado ao grupo (5).

De acordo com os dados experimentais, os resultados computacionais estão em concordância com o valor do tempo de enovelamento, como também, nas mudanças cruciais em seu mecanismo enovelamento, caso R16M₅. Na Figura 5.3, o mapa de densidade do R16M₅ exibe um aumento de densidade significativo em seu estado de transição, indicado por (2) e (3), Figura 5.3C. De acordo com o processo de enovelamento do R16M₅, o estado desenovelado está representado em (1), com o estado de pré-transição e transição em (2) e (3), respectivamente, e o estado nativo em (4) e

Figura 5.3: (A) Visualização da trajetória de enovelamento da C(R16WT, R16M₅) no espaço de fase 2D obtido pelo ELViM. A barra de cor está associada à fração de contatos nativos, Q , normalizada de 0 a 1. (B) e (C) mapa de densidade bidimensional para os domínios R16WT e R16M₅, respectivamente. A barra de cor corresponde a densidade de estados, ρ . Em (A), (B) e (C), cada ponto representa uma conformação e a distância euclidiana entre os pontos está associada a uma distância relativa entre as estruturas. (D) e (E), *ensemble* de estruturas obtidas a partir dos pontos indicados em (B) e (C), onde o estado desenovelado e nativo estão indicado em (1) e (4).



Fonte: Elaborada pelo autor.

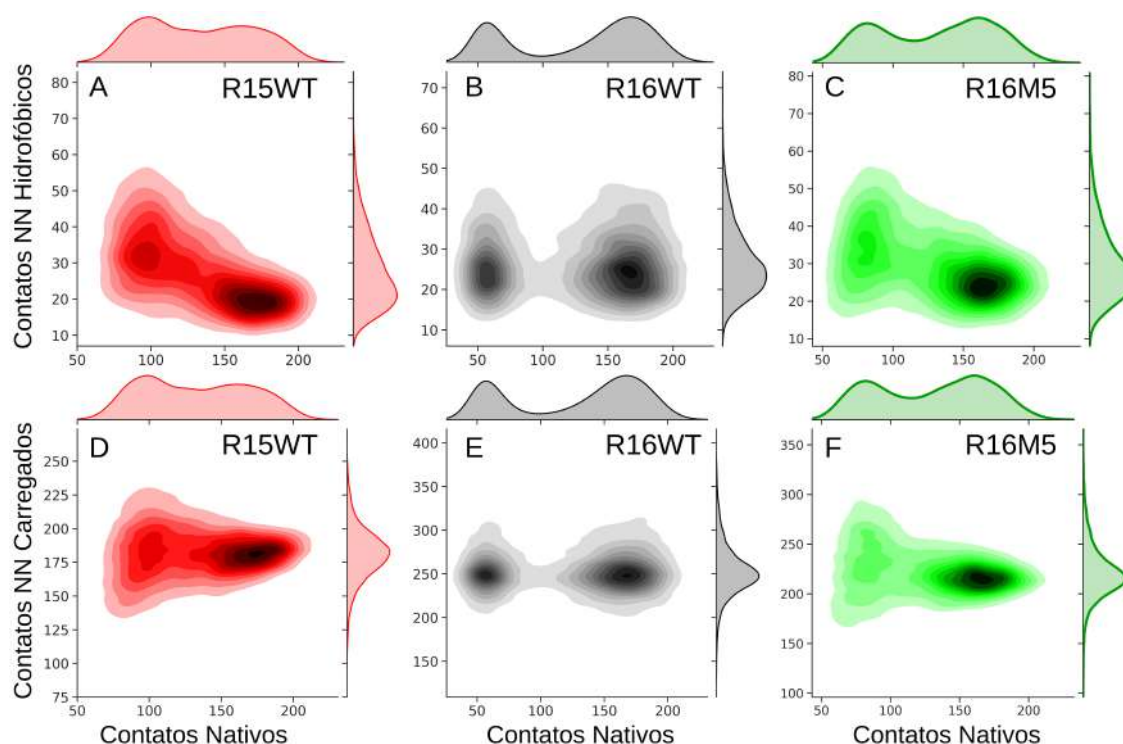
(5). As mudanças no mecanismo de enovelamento ocorrem pela substituição de cinco resíduos, promovendo uma conformação altamente colapsada nos estágios iniciais do processo de enovelamento, como pode ser visto na Figura 5.3E.

Portanto, as proteínas em sua forma não mutada, R15WT e R16WT, assumem diferentes mecanismos de enovelamento, *nucleation–condensation* e *framework–like* mecanismo, respectivamente. No primeiro caso, as hélices 1 e 3 da R15WT estabeleceram o encaixe correto. Enquanto que no segundo caso, a hélice 3 do domínio R16WT, encontra-se em uma busca frustrada pelo encaixe correto de sua hélice devido as rugosidades no perfil energético que esse domínio apresenta. Nesse sentido, as interações nativas e não-nativas são responsáveis por essa busca aleatória, tornando-se a principal fonte de frustração na perfil energética. Desse modo, o conjunto de mutações no R16M₅ foi crucial e responsável em reduzir a rugosidade, como também, promover eventos cooperativos não-nativos que aceleram o processo de dobramento.

5.3.3 Densidade de contatos não-nativos do R15WT, R16WT e R16M₅

Na Figura 5.4, encontram-se as distribuições bidimensionais dos contatos nativos versus os contatos não-nativos (NN), hidrofóbicos e carregados, para os domínios R15WT, R16WT e R16M₅. Os domínios R15WT e R16M₅ apresentam o mesmo comportamento na distribuição de formação dos contatos hidrofóbicos e carregados, conforme consta nas Figuras 5.4A, 5.4C, 5.4D e 5.4F. O comportamento que essas distribuições apresentam segue um decaimento ao longo do processo de enovelamento e, no caso dos contatos não-nativos carregados, eles apresentam uma maior dispersão no estágio inicial quando comparado com o estado nativo.

Figura 5.4: Distribuição bidimensional do número dos contatos nativos versus o número dos contatos não-nativos (NN) hidrofóbicos (A,B e C) e carregados (D, E e F) para as proteínas R15WT, R16WT e R16M₅. Os histogramas correspondentes aos contatos não-nativos e aos contatos nativos encontram-se nos eixos verticais e horizontais, respectivamente.



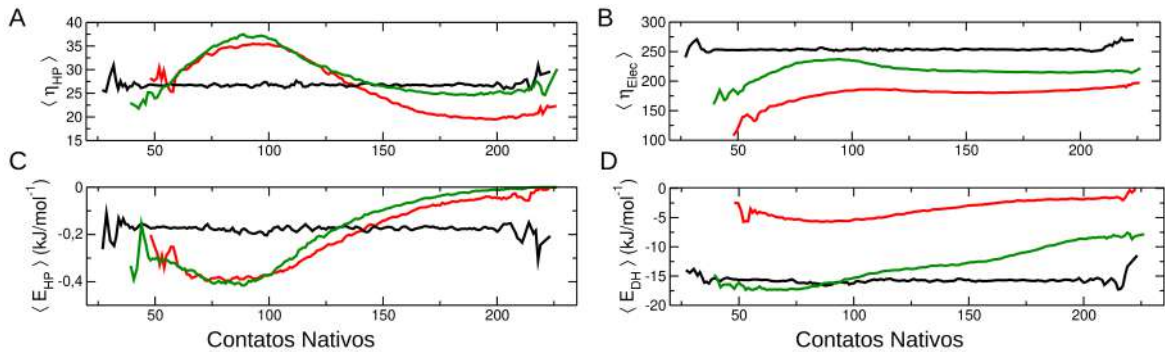
Fonte: Elaborada pelo autor.

Já o número de contatos não-nativos do domínio R16WT sofre, em ambos os resultados (vide Figuras 5.4B e 5.4E), um decaimento considerável no estado de transição. Esse decaimento pode estar diretamente ligado ao fato desse domínio ter

um enovelamento mais lento, quando comparado com R15WT e R16M₅. Portanto, podemos concluir que com as adições dos pontenciais não-nativos ao modelo SBM-C α , o sistema passou a ser retratado de forma mais realística e foi possível compreender os efeitos cooperativos das interações não-nativas durante processo de enovelamento.

Com o cálculo da formação média dos contatos não-nativos hidrofóbicos e carregados, $\langle \eta_{HP} \rangle$ e $\langle \eta_{Elec} \rangle$, para os domínios R15WT (vermelho), R16WT (preto) e R16M₅ (verde) e com o cálculo da energia média hidrofóbica e eletrostática, $\langle E_{HP} \rangle$ e $\langle E_{DH} \rangle$, foi possível analisar o peso dessas variáveis durante o processo de enovelamento. Do ponto de vista hidrofóbico, verifica-se que o número médio dos contatos e a energia média permanecem constantes para R16WT, enquanto que para o R15WT e R16M₅, a energia média tende a zero no estado nativo. Para os caso dos contatos não-nativos carregados e a energia eletrostática, tem-se um número de contatos superior para o domínio R16WT, seguido do R16M₅ e por fim, R15WT. Comportamento semelhante ao cálculo da energia média eletrostática.

Figura 5.5: A e B, número de contatos não-nativos hidrofóbicos e carregados (η_{HP} e η_{Elec}) em função dos contatos nativos. C e D, energia média hidrofóbica e eletrostática ($\langle E_{HP} \rangle$ e $\langle E_{DH} \rangle$) em função dos contatos nativos. Todos os cálculos foram realizados para os domínios R15WT, R16WT e R16M₅ que serão representadas pelas cores, vermelho, preto e verde, respectivamente.



Fonte: Elaborada pelo autor.

De forma geral, todo o processo de enovelamento foi abordado para os domínios R15WT, R16WT e R16M₅ por meio da técnica ELViM, como também o cálculo a distribuição dos contatos não-nativos e as energias hidrofóbicas e eletrostática, tendo como finalidade averiguar regiões onde essas interações atuam de forma predominante. Na próxima seção discutiremos e analisaremos certas regiões de interesse destacadas na projeção dos domínios R16WT e R16M₅.

5.3.4 Formação de contatos específicos devido as interações não-nativas

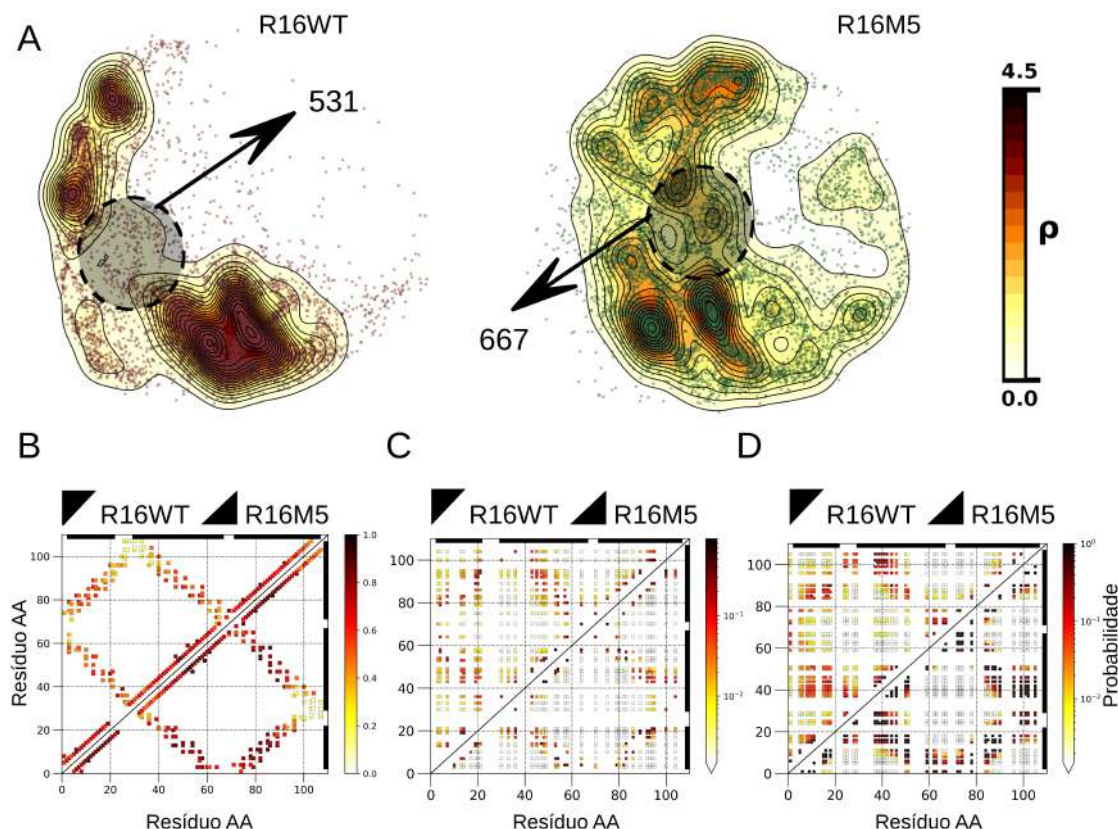
As estruturas presentes nas regiões mencionadas na Seção 5.3.2 para as projeções bidimensionais dos domínios R16WT e R16M₅ foram selecionadas para análise (vide a Figura 5.6A). Também foram calculadas as probabilidades de formação de contatos nativos e não-nativos, conforme constam nas Figuras 5.6B, 5.6C e 5.6D. Ao total, foram selecionadas 531 estruturas para o domínio R16WT e 667 estruturas para o domínio R16M₅ e, em seguida, foram computadas as probabilidades de formação dos contatos nativos. Na Figura 5.6B, os contatos nativos para formar as α -hélices um, dois e três encontram-se em sua totalidade 100% formados. No entanto, existem algumas diferenças na probabilidade de formação quando são analisadas as interações intercadeia. Nesse caso, as interações entre as α -hélices apresentam maior probabilidade para o R16M₅ do que para o R16WT, conforme visto na Figura 5.6B.

Nas Figuras 5.6C e 5.6D, foram calculadas as probabilidades de formação dos contatos não-nativos, hidrofóbicos e eletrostáticos, respectivamente. Observa-se que na coluna vertical e na coluna horizontal, no intervalo de 0 a 20, há uma diferença considerável na probabilidade de formação desses contatos. As interações entre as α -hélices para o domínio R16M₅ ocorrem com maior probabilidade, mas em menor número; enquanto que para o domínio R16WT, esses contatos ocorrem em maior número, porém com menor probabilidade.

Conforme apresentado por Silva, et al. [44], o aumento na taxa de enovelamento do domínio R16WT pode estar ligado a uma busca frustrada pelo encaixe correto da α -hélice um. Com os resultados aqui apresentados, podemos reforçar a ideia de que o aumento na taxa de enovelamento do domínio R16M₅ ocorre devido a certos contatos não-nativos se formarem com alta probabilidade, enquanto que, para o domínio R16WT, esses contatos ocorrem em maior número, porém de forma desordenada, caracterizando, dessa forma, a busca frustrada dessa α -hélice durante o processo de enovelamento.

Outro fator importante foi o aumento no mapa de densidade entre os estados desenovelado e nativo para o domínio R16M₅. Um número considerável de estruturas passaram a popular uma certa região do espaço de fase, evidenciando que certas conformações passaram a ser formadas e consideradas para o aumento na taxa de enovelamento desse domínio. Portanto, a α -hélice um desempenha um papel importante no processo de enovelamento. Dessa forma, uma análise mais aprofundada será realizada em torno dessa região.

Figura 5.6: Em A, mapa de densidade para os domínios R16WT e R16M₅, respectivamente, com seus estados intermediários destacados por uma circunferência. Número de estados selecionados correspondem a 531 e 667 para os domínios R16WT e R16M₅, respectivamente. Em B, C e D, probabilidade de formação de contatos nativos e não-nativos hidrofóbicos e carregados, respectivamente.

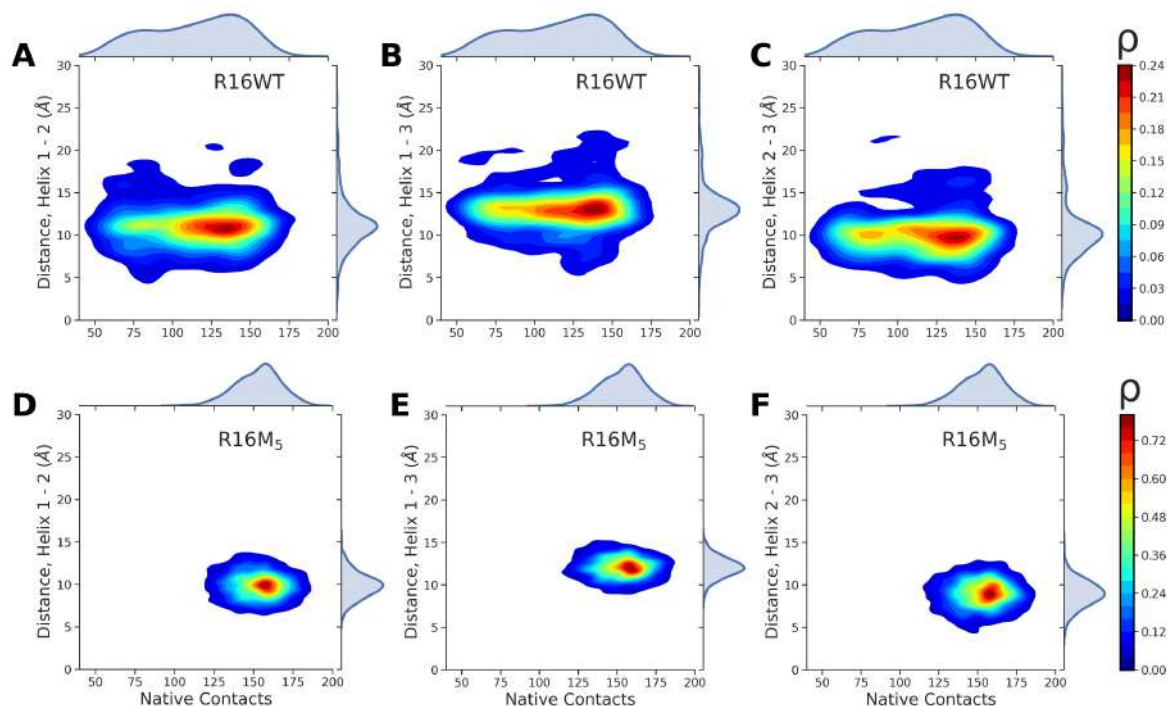


Fonte: Elaborada pelo autor.

O número de contatos nativos para o R16WT das estruturas selecionadas variam de 50 a 175. A distância entre a hélice 1 e 2, com Q em torno de 75, exibe uma flutuação significativa, variando de 5 a 20 Å, conforme mostrado na Figura 5.7A. No entanto, enquanto o número de contatos nativos aumentam, a distância entre a hélice 1 e 2 diminui, com $125 < Q < 150$, a hélice atinge aproximadamente sua distância nativa, 11 Å. Entretanto, foi observado um comportamento oposto entre as hélices 1 e 3 e as hélices 2 e 3, apresentando uma pequena flutuação próximo de $Q = 75$ e uma alta flutuação para $Q > 125$ (Figuras 5.7B e 5.7C). Como visto, foi sugerido que a busca pelo encaixe correto da hélice 1 e 2 promove o encaixe incorreto da hélice 3 com as hélices 1 e 2. O R16WT se dobra através de um mecanismo onde a formação das hélices antecede o empacotamento das hélices. Nossos resultados corroboram com o processo de estruturação da formação das hélices, onde a hélice 1 empacota com a hélice

2, logo a hélice 3 deve realizar uma bruca frustrada pelo encaixe correto, sendo assim, o armadilha cinética emerge de uma busca frustrada devido a eventos não-nativos.

Figura 5.7: A distância do centro de massa das hélices. R16WT (A, B e C) e R16M₅ (D, E e F) exibem a distância do centro de massa (Å) entre α -hélice 1 e 2, 1 e 3, e 2 e 3, respectivamente, em função dos contatos nativos com sua distribuição correspondente (vertical e horizontal). As barras coloridas representam a densidade dos estados, ρ .

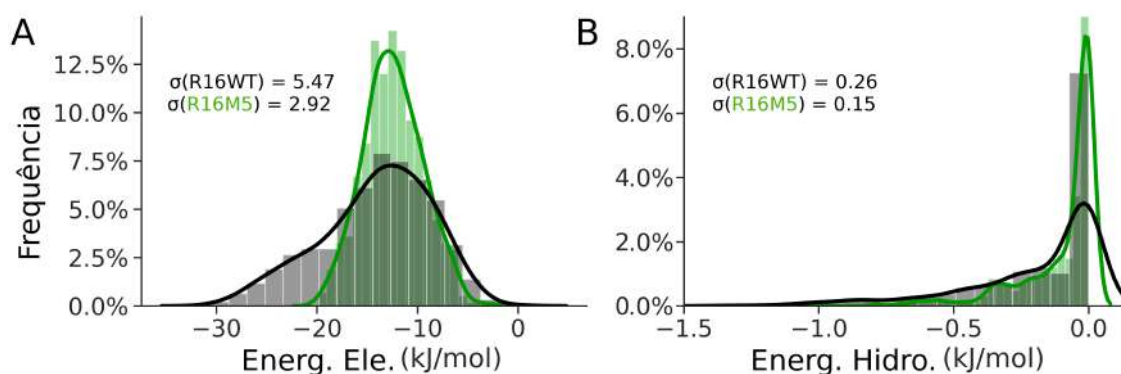


Fonte: Elaborada pelo autor.

Em contraste com os resultados do R16WT, o R16M₅ apresenta um conjunto de conformações mais estruturadas, $120 < Q < 180$. De acordo com a probabilidade formação dos contatos (Figura 5.6), a distância entre as hélices exibem uma alta densidade localizada em valores próximos à distância nativa, 9.9 Å (Figuras 5.7D), 11.9 Å (Figuras 5.7E) e 8.9 Å (Figuras 5.7F). O conjunto de estruturas selecionadas para R16M₅ encontram-se do pré-estado de transição até o pós-estado de transição. Esta região corresponde aproximadamente ao mesmo estágio de R16WT, mas como resultado das cinco mutações, E18F + E19D + I22L + K25V + V29L, o mecanismo de dobramento, as rugosidade no perfil energético e os eventos não-nativos foram afetados de forma a acelerar o processo de enovelamento do R16M₅.

Por fim, foram calculadas as energias, eletrostáticas e hidrofóbicas, para os grupos de conformações selecionadas (vide a Figura 5.6). Nas Figuras 5.8A e 5.8B são apresentados os histogramas de energia para as proteínas R16WT e R16M₅, nas

Figura 5.8: Histograma da energia eletrostática (A) e da energia hidrofóbica (B) para as proteínas R16WT e R16M₅, nas cores preto e azul, respectivamente.



Fonte: Elaborada pelo autor.

cores preto e verde, respectivamente. Em ambos os cálculos, o valor das energias, em média, são aproximadamente iguais, porém o valor do desvio padrão para o R16WT foi superior ao valor do desvio padrão do R16M₅. Também foi observado regiões no histograma que podem ser interesse para melhor compreender o enovelamento desses domínios.

Portanto, de acordo com os resultados das Figura 5.7 e 5.8, foi possível identificar que a R16M₅ apresenta, em sua trajetória de enovelamento, uma busca direcionada das α -hélices para o seu correto encaixe. Nossa hipótese sustenta o fato de que o R16WT não apresenta nenhum *cluster* ou um possível estado intermediário, mas sim, um conjunto de estruturas que passa a visitar no espaço do fase que não o direciona para seu estado nativo, mas sim para outros mínimos locais em seu perfil energético. O fato do direcionamento das α -hélices para o R16M₅ pode ser um dos fatores que levaram seu aumento na taxa de enovelamento.

5.4 Conclusão

No seção anterior foram apresentados os resultados que tiveram como foco estudar, analisar e explorar os efeitos cooperativos das interações não nativas no processo de enovelamento dos domínios R15WT, R16WT e R16M₅. Com base no método de visualização ELViM, foi possível explorar todo o perfil energético e verificar se os domínios compartilhavam do mesmo espaço de fase, ou seja, se eles apresentavam a mesma trajetória de enovelamento. Dessa forma, foi observado que os domínios R15WT e R16WT apresentaram trajetórias distintas, onde o domínio R15WT apresentou uma alta densidade de probabilidade em seu estado de transição, o que caracteriza sua baixa barreira de energia para enovelar, conforme nossos resultados publicados anteriormente [44]. Já o domínio R16WT apresentou baixa densidade de probabilidade para a mesma região, o que acarretou no aumento da barreira de energia (vide Referência [44]). Para o R16WT e R16M₅, foi observado uma região de alta relevância, que foi explorada a fundo com a finalidade de melhor entender e aferir razões para o aumento considerável na taxa de enovelamento do R16M₅. Outro fator importante foi o cálculo da distribuição dos contatos não-nativos ao longo da trajetória de enovelamento. Esse resultado salientou que as interações eletrostáticas são as interações predominantes no sistema e as interações hidrofóbicas podem ser compreendidas como as interações compensatórias ou estabilizadoras.

Devido a essa região evidenciada no mapa densidade de probabilidade para o R16M₅, foi possível examinar e calcular o mapa de contato nativo e não-nativo para esse grupo de estruturas. Dessa forma, no caso do R16M₅, ocorreu a formação de contatos entre as α -hélices com alta probabilidade, evidenciando a estruturação de uma conformação específica. Esse resultado exibiu o alto grau de importância da adição dos potenciais não-nativos e da escolha da coordenada de reação, pois com os potenciais não-nativos, foi possível obter uma boa concordância com os dados experimentais ao comparar os tempos de enovelamento. Já com a escolha adequada da métrica, ou seja, as distâncias entre as α -hélices, foi possível melhor compreender e entender a busca frustrada pelo encaixe correto que as α -hélices apresentam. Nesse caso, o encaixe da α -hélice um com a α -hélice três.

Por fim, os resultados corroboram com os resultados e as hipóteses apresentadas na literatura no que diz respeito à diferença no tempo de enovelamento dos domínios R15WT e R16WT e ao aumento na taxa de enovelamento ao realizar as substituições de resíduos carregados e hidrofóbicos. Não apenas em corroborar resultados, mas foi apresentado um possível estado responsável pelo aumento na taxa de enovelamento do R16M₅, como também, fatores que podem ser compreendidos e estarem

contribuindo para esse aumento. Vale lembrar que outras regiões foram observadas, como aquelas apresentadas na Figura 5.8, que serão estudadas posteriormente. Como trabalho futuro, iremos caracterizar o estado indicado para o R16M₅, como também determinar, por meio de outras medidas, os estados citados para o R16WT na Figura 5.8.

Capítulo 6

Estabilização do Estado Intermediário da Im7 e Im9 devido à Mudança de Carga

6.1 Introdução

O enovelamento de proteínas é um processo intrínseco que ocorre em todos os seres vivos. O foco deste estudo não se restringe somente ao papel que o enovelamento desempenha no organismo, mas também aos processos físico-químicos que o governam. O entendimento do enovelamento pode responder a problemas fundamentais em diversas áreas da ciência [10]. Atualmente, a teoria que melhor explica o processo de enovelamento de proteínas é a Teoria da Superfície de Energia.

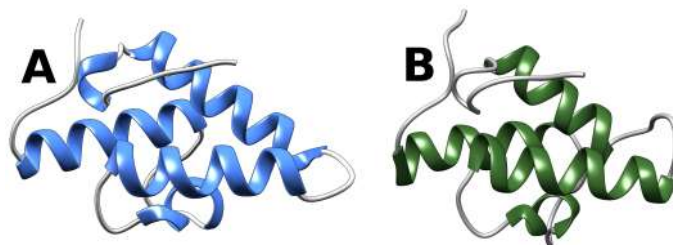
Essa teoria propõe a existência de uma seleção natural de sequências e conformações da proteína que resultam em uma superfície energética afunilada e intrinsecamente rugosa, onde a energia mais profunda do funil corresponde à conformação nativa da proteína. As rugosidades correspondem aos mínimos locais de energia, que funcionam como armadilhas para o enovelamento. Para que a proteína possa exercer seu papel biológico a tempo, seu enovelamento deve ocorrer sem que ela fique presa em armadilhamentos (mínimos locais). Por esse motivo, a inclinação do funil de energia deve ser grande o bastante para vencer esses mínimos locais. Essa teoria é uma ferramenta fundamental para a compreensão qualitativa e quantitativa dos resultados teóricos e experimentais relacionados ao enovelamento [16–20].

O estudo do enovelamento de proteínas, a partir da teoria descrita acima, tem levado à contribuições relevantes para a Biofísica, assim como para áreas afins. O

comportamento cinético e termodinâmico de sistemas com uma superfície energética irregular (ou rugoso) diferencia-se daqueles com uma superfície regular (ou lisa), sendo que, em casos nos quais as superfícies de energias encontram-se irregulares, há uma grande competição entre as interações locais e não locais. Esta competição é chamada de frustração [16, 31, 32]. A frustração existe em biomoléculas e, às vezes, são de grande importância em determinadas funções biológicas. Do ponto de vista biofísico, as interações não locais e não-nativas também podem ser energeticamente favoráveis [33–35] e o entendimento da sua presença ou da sua ausência é fundamental para compreender os seus efeitos sobre o processo de enovelamento [36].

Para os casos onde a superfície energética é altamente irregular, pode ocorrer a formação de estados intermediários bem populados e significativos para o enovelamento. Esses intermediários podem levar a proteína até sua estrutura nativa ou, ainda, podem levar à formação de agregados proteicos. Muitas vezes, a formação desses agregados resultam em doenças amiloidogênicas que são de grande interesse na área de saúde pública. A doença de Alzheimer, doença da vaca louca e outras amiloidoses são exemplos de doenças causadas pela formação de agregados.

Figura 6.1: Representação estrutural das proteínas Im7 e Im9. Em A, Im7 (azul) e em B, Im9 (verde), ambas possuem 86 resíduos de aminoácidos e apresentam o RMSD $\approx 2,0$ Å. As proteínas Im7 e Im9 foram extraídas do PDB 1AYI e 1IMQ, respectivamente.



Fonte: Elaborada pelo autor.

Essas ocorrências podem estar relacionadas a fatores positivos ou negativos para o enovelamento. Em alguns casos, essas regiões são bem caracterizadas, ou seja, é conhecida a estrutura do estado intermediário. No entanto, para outros casos, essas estruturas são desconhecidas. Como objeto de estudo, selecionamos as proteínas Im7 e Im9 (PDB ID: 1AYI e 1IMQ), pois essas proteínas apresentam estados intermediários, tendo o estado intermediário da Im7 bem caracterizado [125]. Em contrapartida, esses estados intermediários não foram verificados por simulações computacionais utilizando modelos minimalistas convencionais com adição de potenciais não-nativos, como mencionado nos capítulos anteriores. As proteínas Im7 e Im9 são altamente si-

milares e apresentam exatamente 86 resíduos de aminoácidos e sua estrutura nativa é constituída por quatro α -hélices, como pode ser visto na Figura 6.1.

De acordo com os dados publicados por Gorski [126], os experimentos realizados em diferentes faixas de pH com as proteínas Im7 e Im9 apresentaram um comportamento distinto na estabilização do estado intermediário. De acordo com o cálculo das taxas de enovelamento e desenovelamento e a partir dos dados de *chevron plot* disponíveis [126], foi possível observar que, para os valores de pH iguais a 6,0, 6,5, 7,0 e 8,0, ocorre a formação do estado intermediário na Im7. Entretanto, para a Im9, seu estado intermediário pode ser observado para os valores de pH iguais a 5,0, 6,0, 6,5 e 7,0, porém com maior estabilidade para $\text{pH} = 6,0$. Os resultados demonstram que as proteínas apresentam mecanismo diferentes de enovelamento quando estão em diferentes ambientes químicos.

O mecanismo de enovelamento da proteína Im7 ocorre da seguinte maneira: i) existe uma barreira de energia a ser vencida entre o estado desenovela e intermediário (estado de transição 1 - TS1). Nesse estágio ocorre a formação e o empacotamento das α -hélices um e dois e sua estabilização é resultado das interações hidrofóbicas; ii) no estado intermediário as α -hélices um, dois e quatro encontram-se estruturadas e a α -hélice quatro interage de forma não nativa com as α -hélices um e dois para auxiliar o encaixe correto da α -hélice três; iii) após vencer o estado de transição 2 (TS2), os resíduos da α -hélice três passam a ocupar no espaço sua posição nativa e, dessa forma, ocorre a estruturação da α -hélice três e o correto encaixe da α -hélice quatro, chegando, assim, ao seu estado nativo [125, 127]. O mecanismo de enovelamento da Im9 não foi reportado na literatura. Logo, acredita-se que, no caso da Im9, seu estado intermediário deva apresentar o mesmo mecanismo da Im7, pois as duas proteínas são altamente homólogas e apresentam alta similaridade estrutural e sequencial.

Não apenas para diferentes valores de pH os estados intermediários apresentam modificações, mas alterações no valor da concentração de sal no ambiente acarreta mudancas na estabilidade do estado intermediário. Para o valor de pH igual a 5,5 e em uma concentração de NaCl igual a 0,4 M, a proteína Im9 apresenta um estado intermediário estável. No entanto, para o valor de 0,9 M, não ocorre a formação do estado intermediário. Já para o caso da Im7, em pH igual a 7,0 e em 0,4 M NaCl, o estado intermediário encontra-se estável. Contudo, para o valor igual a 1,0 M, essa proteína não apresenta estado intermediário. Portanto, do ponto de vista experimental, existem diferentes formas de controlar a estabilização do estado intermediário, seja por alterações no valor do pH, como também nos valores das concentrações do ambiente químico.

Os resultados computacionais disponíveis na literatura abordam esse sistema apenas ponderando os efeitos hidrofóbicos. Os resultados publicados por Chen [90] foram realizados com o modelo $\text{SBM-C}\alpha + \text{dsb} + \text{HP}$ (dsb ou db: potencial de dessolvatação; HP: potencial hidrofóbico), onde o estado intermediário da Im7 foi observado, mas apenas a formação do núcleo hidrofóbico foi mencionada para sua estabilização. Os efeitos eletrostáticos não foram observados devido ao modelo utilizado não ponderar essa variável. O modelo proposto nesse trabalho mimetizará as variações de pH, assim como, o efeito hidrofóbico na formação dos estados intermediários para a Im7 e Im9.

6.2 Objetivo

Os resultados computacionais disponíveis na literatura com as proteínas Im7 e Im9 que abordam o processo de enovelamento e o estudo dos estados intermediários, apenas retratam esses sistemas considerando as interações hidrofóbicas como as responsáveis por estabilizar esse estado. No momento atual, poucos estudos encontram-se disponíveis na literatura que abordam esse sistema. Portanto, o trabalho desenvolvido abordará o modelo já reportado na literatura, porém com a adição do potencial eletrostático. Afim de reproduzir o sistema para diferentes valores de pH, será calculado o valor do pKa de cada resíduo ionizável e, assim, estimar o valor da carga para o pH fixado experimentalmente, com base no valor de pKa obtido teoricamente.

Com as simulações reliazadas, será estudada a probabilidade de formação dos contatos nativos e não-nativos durante o processo de enovelamento, tendo como finalidade determinar sua influência no processo de estabilização do estado intermediário. A simulação será realizada com o modelo SBM-C α , onde o potencial dos contatos não locais, potencial de Lennard-Jones, será substituído pelo potencial de dessolvatação (Seção 3.4). Os potenciais não-nativos, eletrostático e hidrofóbico, também serão adicionados ao sistema. Dessa forma, torna-se possível o estudo dos efeitos eletrostáticos e hidrofóbicos.

6.3 Resultados e Discussão

Os resultados que serão apresentados nessa seção foram produzidos a partir de simulações de dinâmica molecular com o modelo SBM-C α . Esse modelo, conforme apresentado na Seção 3.1, contabiliza as interações não locais pelo potencial de Lennard-Jones. Nesse estudo, o potencial de Lennard-Jones será substituído pelo potencial de dessolvatação (Seção 3.4). Essa modificação foi realizada para melhor reproduzir o sistema e para observar os efeitos de dessolvatação no processo de enovelamento. Devido aos resultados experimentais evidenciarem uma certa importância para a formação do núcleo hidrofóbico, dessa forma torna-se inviável sua deleção no modelo, sendo que o estado intermediário da Im7, em alguns estudos, foi amplamente sustentado por esse efeito. Além dessa modificação, foram introduzidos ao modelo, os potenciais não-nativos, eletrostático e hidrofóbico.

Com a adição do potencial eletrostático ao sistema, foi possível estudar o efeito da mudança de carga na estabilização do estado intermediário das proteínas Im7 e Im9. De acordo com os dados experimentais disponíveis na literatura, a estabilização desse estado ocorre devido à mudança que acontece no ambiente químico. Dessa forma, a proteína Im7 apresenta um estado intermediário estável próximo de pH = 7,0, enquanto que a proteína Im9, um possível estado intermediário em pH = 5,5. Portanto, os resultados para as proteínas Im7 e Im9 serão estudados e analisados de maneira individual, devido ao comportamento distinto que ambos apresentam.

Tabela 6.1: Diferentes valores de carga para as histidinas 40 e 47 da Im7 e para as histidinas 05, 39 e 46 da Im9 em diferentes valores de pH.

Proteína	Aminoácidos	pKa	pH > pKa(HIS-40)	pH < pKa(HIS-40)
Im7	HIS-40	6,60	carga = 0,0	carga = 1,0
	HIS-47	4,60	carga = 0,0	carga = 0,0

Proteína	Aminoácidos	pKa	pH > pKa(HIS-05)	pH < pKa(HIS-36)
Im9	HIS-05	6,19	carga = 0,0	carga = 1,0
	HIS-39	6,10	carga = 0,0	carga = 1,0
	HIS-46	2,19	carga = 0,0	carga = 0,0

Fonte: Valores de pKa obtidos por meio do software H++

Nos capítulos anteriores, nos estudos realizados com a CI2 e com os domínios R15, R16 e R16M₅, as simulações foram realizadas contabilizando o efeito eletrostático e hidrofóbico. Para a parte eletrostática, foram adicionadas cargas aos resíduos de aminoácidos carregados, conforme introduzido na Seção 3.3. Entretanto, para o estudo abordado nesse capítulo, não serão adicionadas cargas em todos os resíduos carregados.

De acordo com Tabela 6.1 para a Im7, de todos os resíduos carregados, as histidinas serão os únicos aminoácidos que seguiram a ordem apresentada, assumindo carga +1 para a arginina e lisina e carga -1 para o ácido glutâmico e o ácido aspártico. Portanto, serão realizadas duas simulações com diferentes valores de cargas para as histidinas e, posteriormente, serão analisados os efeitos dessa mudança.

6.3.1 Efeito da Mudança de Carga nas Histidinas 40 e 47 da Im7

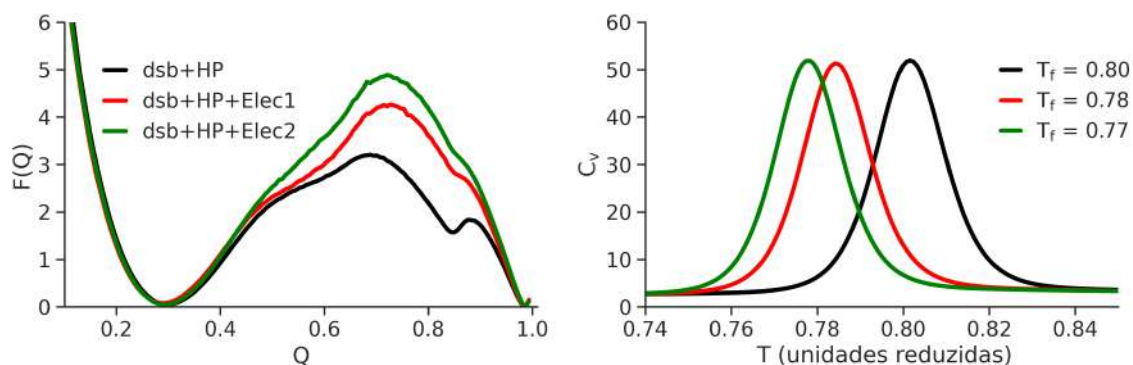
A partir das análises realizadas com os resultados experimentais, foi possível determinar que o estado intermediário da Im7 pode ser estabilizado alterando as condições químicas do ambiente [126]. Do ponto de vista computacional, não existem muitos resultados disponíveis na literatura que exploram esse sistema da forma que ele é proposto nesse trabalho, onde foram adicionados potenciais não-nativos ao modelo SBM- $C\alpha$. Esse modelo de simulação retrata o sistema de forma mais realística, assim como é capaz de realizar certas observações que são observadas do pontos de vista experimental.

Com o modelo adotado, SBM- $C\alpha$ + dessolvatação + eletrostático + hidrofóbico, foi investigado, por meio de simulações computacionais, o enovelamento da proteína Im7. Conforme mencionado, foram adotados valores diferentes de cargas para a histidina 40 (vide a Tabela 6.1). Nas Figuras 6.2A e 6.2B, encontram-se os perfis de energia livre, $F(Q)$, e o calor específico, C_v , respectivamente. Como controle, foi realizado uma simulação considerando apenas o modelo SBM- $C\alpha$ + dessolvatação + hidrofóbico (dsb + HP(MJ)), pois esse resultado já foi demonstrado na literatura por outros pesquisadores [90].

O perfil de energia livre da simulação controle, pode-se observar que para valores de Q acima de 0,8 e abaixo de 0,95, um mínimo local é observado e uma barreira de aproximadamente 0,5 $k_B T$ deve ser vencida para a proteína enovelar. Esse mesmo mínimo foi constatado nos resultados publicados Chen e Chan [90]. Durante o processo de enovelamento, com o potencial de dessolvatação, antes dos contatos nativos se formarem, eles devem vencer uma barreira de energia, ϵ_{db} , ou seja, a barreira para retirar uma molécula de água, conforme a Figura 3.2 da Seção 3.4. Assim, o mínimo local observado no perfil de energia livre representa o estágio do enovelamento onde ocorre esse processo de retirar uma molécula de água entre dois resíduos que interagem de forma nativa.

Com o potencial eletrostático acoplado ao sistema, o mínimo local obser-

Figura 6.2: (A) Perfil de energia livre, $F(Q)$, em função da fração dos contatos nativos, Q . (B) Calor específico, C_v , em função temperatura em unidades reduzidas, T . As linhas nas cores preto, vermelho e verde correspondem aos potenciais dsb+HP, dsb+HP+Elec1, e dsb+HP+Elec2, respectivamente.



Fonte: Elaborada pelo autor.

vado anteriormante desapareceu e a barreira de energia livre sofreu um aumento de aproximadamente $1,2 k_B T$ para o sistema Elec1 e $2 k_B T$ para o sistema Elec2. O sistema Elec1 representa o modelo para o valor de $pH < pKa(HIS40)$, ou seja, quando a HIS40 apresenta carga igual a $+1$ e a HIS47 está com carga zero. Já o sistema Elec2, para o valor de $pH > pKa(HIS40)$, têm-se as histidinas 40 e 47 com cargas iguais a zero. Nossos resultados corroboram com os dados experimentais, pois para os valores de pH iguais a 6,0, 6,5, 7,0 e 8,0, têm-se ΔG_{un} , barreiras de desnaturação, iguais a $16,1 \pm 0,5$, $19,7 \pm 0,5$, $23,5 \pm 0,6$, $27,3 \pm 1,1$, respectivamente, em unidade de $kJ mol^{-1}$. Portanto, os valores da barreira de desnaturação e do pH apresentam um crescimento linear, assim como as barreiras de energia livre de nossas simulações, o que pode ser verificado na Figura 6.2A.

Na Figura 6.2B, o pico do calor específico corresponde à temperatura de enovelamento que cada modelo apresenta. Portanto, ao adicionar carga aos resíduos de aminoácidos, nosso sistema caminhou para valores menores de temperatura. Os valores de temperaturas aqui apresentados estão em unidades reduzidas do Gromacs. O modelo controle foi o que apresentou maior temperatura de enovelamento, seguido das temperaturas correspondentes para os modelos Elec1 e Elec2, respectivamente. Com os modelos que possuem cargas, podemos realizar uma comparação qualitativa com os dados experimentais, em relação às barreiras, já que esses resultados estão disponíveis na literatura.

Devido às diferenças encontradas nas barreiras de energia livre para os modelos Elec1 e Elec2, uma análise mais detalhada será realizada. De acordo com a

Figura 6.2A, para o modelo Elec1, a barreira de energia para o processo de dessolvatação sofreu uma redução considerável. Contudo, para os valores próximos de $Q=0,5$, observa-se o que chamamos de "ombro" no perfil de energia livre, que aparece no estado de pré-transição. Ao comparar esse resultado com a simulação controle, nota-se um comportamento distinto, onde tem-se um estado de transição com maior amplitude. Em vista disso, os sistemas Elec1 e Elec2 serão analisados de forma individual.

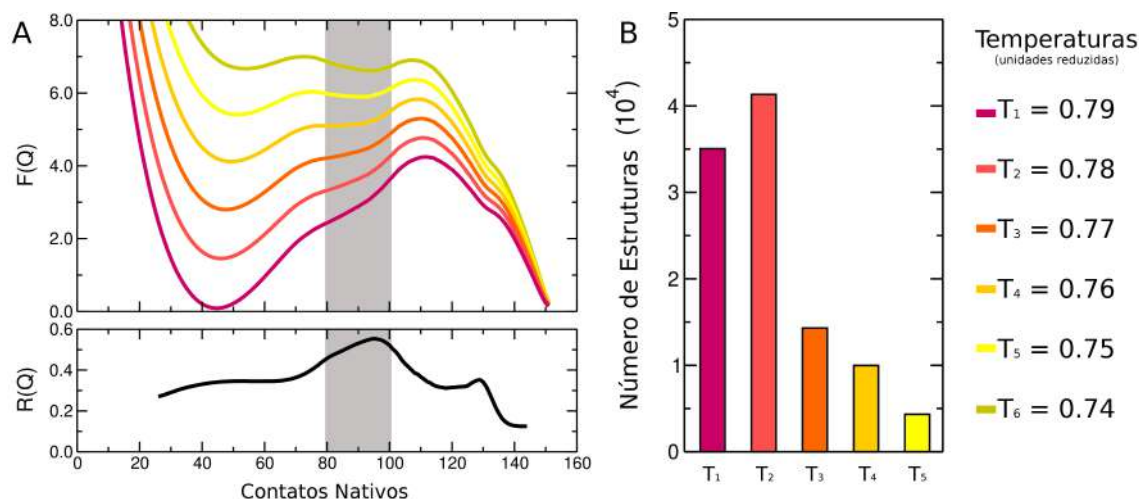
6.3.2 Estado intermediário para o sistema denominado Elec1: H40(+1) e H47(0)

Na Figura 6.3A, encontram-se os perfis de energia livre para o sistema Elec1 em diferentes temperaturas, onde os perfis selecionadas estão abaixo da temperatura de enovelamento, T_1 . Conforme caminha-se para valores menores da temperatura, $T^* < T_1$, ocorre uma estabilização e um aumento na barreira de energia para vencer o estado de pré-transição, região em destaque no perfil de energia livre. Foram selecionados, ao todo, seis perfis de energia, onde o último, $T_6 = 0,76$, foi o que apresentou dois picos consideráveis com, aproximadamente, mesma altura. Ademais, o estado desenovelado tornou-se menos estável, pois o valor mínimo de energia aumenta conforme se diminui o valor da temperatura. Outro fator importante foi que o estado de transição principal não sofreu nenhum deslocamento no eixo horizontal, conforme os perfis de energia livre foram obtidos.

Com o cálculo da rota de enovelamento, $R(Q)$, para a trajetória na temperatura de enovelamento, T_1 , conforme apresentado na Figura 6.3B, foi possível observar que existe uma região, sendo a mesma região destacada no perfil de energia de livre, que apresenta uma alta especificidade na formação dos contatos nativos. Esse perfil de rota demonstra que existe uma certa formação de contatos específicos no estado de pré-transição. Ademais, para a região próxima do valor de $Q = 130$, pode-se observar um pico que corresponde, aproximadamente, a região onde os resíduos precisam extrair uma molécula de água que impede que os contatos nativos formados alcancem seu mínimo de energia.

Devido à região destacada no perfil de energia livre ter uma correspondência direta no perfil de rota obtido, foi calculado o número de estruturas presentes nesse intervalo para extrair a estrutura representativa desse *cluster* e, assim, caracterizar o estado intermediário da proteína Im7 no modelo Elec1. Portanto, na Figura 6.3C, tem-se o número de estruturas selecionadas no intervalo $80 < Q < 100$ e para as temperaturas T_1, T_2, T_3, T_4, T_5 e T_6 . O número de estruturas selecionadas em cada temperatura foi maior para a temperatura T_2 , que corresponde a temperatura abaixo

Figura 6.3: (A) Perfil de energia livre, $F(Q)$, em função dos contatos nativos em diferentes temperaturas (parte superior) e rota de enovelamento, $R(Q)$, em função dos contatos nativos, calculada na temperatura de enovelamento, $T = 0,79$ (parte inferior). (B) Número de estruturas presentes dentro do intervalo de $80 < Q < 100$, para o estado intermediário da Im7 em diferentes valores de temperatura.



Fonte: Elaborada pelo autor.

da temperatura de enovelamento. Por esse motivo, seguiremos nossas análises com esse conjunto de estruturas, pois foi o grupo que apresentou um maior número de conformações nessa região.

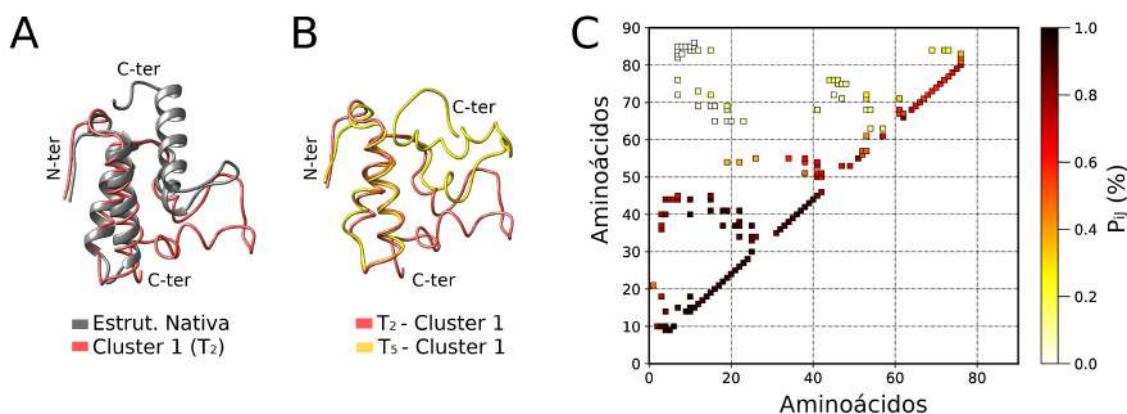
Com as estruturas selecionadas de T_2 e com auxílio da ferramenta *gcluster* do Gromacs, foi selecionado a estrutura representativa do primeiro *cluster* obtido, onde 45% das estruturas selecionadas em T_2 encontram-se no *cluster*-1. Já com a estrutura representativa selecionada, calculou-se o RMSD dessa estrutura com a estrutura nativa da proteína Im7, conforme Figura 6.4A. O valor obtido de RMSD foi 1,3 Å e as estruturas quando alinhadas, apresentam as α -hélices um, dois e três sobrepostas, enquanto que a α -hélice 4 encontra-se em posição distinta.

De acordo com os dados experimentais, durante o enovelamento da Im7, existe um estado intermediário estável acessado por essa proteína. Foi reportado que nesse estágio do enovelamento ocorre a formação de um núcleo hidrofóbico para acomodar de forma correta a hélice três. Devido à formação desse núcleo hidrofóbico, ocorre o encaixe não nativo da α -hélice quatro com as α -hélices um e dois. Nesse momento, em sua totalidade, as α -hélices não apresentam todos as interações nativas. Após vencer a barreira de energia do estado intermediário, ocorre a formação das interações favoráveis para a estruturação da α -hélice três, seguido do encaixe correto da α -hélice quatro

com as α -hélices um e dois. Portanto, de acordo com a estrutura representativa do *cluster*-1 (vide a Figura 6.4A), observa-se similaridades entre a estrutura selecionada e a estrutura mencionada do ponto de vista experimental.

Também foi selecionada a estrutura representativa do *cluster*-1 para a temperatura T_5 . O objetivo em selecionar as estruturas representativas em outras temperaturas foi para averiguar se a mesma estruturação foi formada, pois de acordo com os dados de rota para T_1 , existe uma alta especificidade na região que foi destacada na Figura 6.3A, ou seja, existe uma estrutura "única" sendo visitada com alta frequência. Então, foi calculado o RMSD das estruturas representativas selecionadas das temperaturas T_2 e T_5 , conforme apresentado na Figura 6.4B. O valor do RMSD foi igual a 1,7 Å e foi observada a mesma estruturação: α -hélices um e dois parcialmente estruturadas e sobrepostas, enquanto que as α -hélices três e quatro encontram-se em sua forma não nativa. Além disso, a α -hélice quatro pode estar formando interações não favoráveis com as α -hélices um e dois.

Figura 6.4: (A) Alinhamento da estrutura nativa, em cinza, com a estrutura representativa do *cluster*-1 formado a partir da trajetória na temperatura T_2 , com valor de RMSD igual a 1,3 Å. (B) Alinhamento das estruturas representativas selecionadas a partir do *cluster*-1 nas temperaturas T_2 e T_5 , para as cores rosa e amarelo, respectivamente, com valor de RMSD igual a 1,7 Å. (C) Probabilidade da formação dos contatos nativos formados para estruturas presentes na região destacada no perfil de energia livre para a temperatura T_2 .



Fonte: Elaborada pelo autor.

Para validar os dados apresentados no parágrafo anterior, foi calculado a probabilidade de formação dos contatos nativos para a proteína Im7 com o conjunto de estruturas selecionadas na região destacada no perfil de energia livre para a temperatura T_2 . Portanto, conforme consta na Figura 6.4C, com a probabilidade de formação dos contatos nativos, foi possível verificar que as α -hélices três e quatro apresentam menor

probabilidade de formação quando comparada com a probabilidade de formação das α -hélices um e dois. Também foi verificado que a α -hélice quatro apresenta probabilidade abaixo de 0,1 para os contatos nativos entre as α -hélices um e dois. Já a probabilidade para os contatos nativos entre α -hélice três e quatro encontram-se na faixa de 0,2 a 0,4, enquanto que probabilidade entre as α -hélices um e dois encontra-se acima de 0,75.

A estrutura selecionada por meio das simulações computacionais com o modelo SBM-C α + dsb + HP + Elec1 apresentou as mesmas características estruturais daquela reportada experimentalmente. Vale lembrar que os resultados computacionais disponíveis na literatura não abordam o modelo de simulação aqui proposto e nem reproduzem o modelo de estrutura aqui evidenciado. Relembrando que o modelo Elec1 é o modelo no qual as histinas 40 e 47 possuem cargas iguais a +1 e zero, respectivamente. Nesse modelo, mimetizamos um sistema onde o valor de pH é menor que o valor pKa da histina 40, sendo ele igual a 6,0. Do ponto de vista experimental, para valores de pH < 7, o estado intermediário torna-se mais estável, enquanto que para valores acima, menos estável.

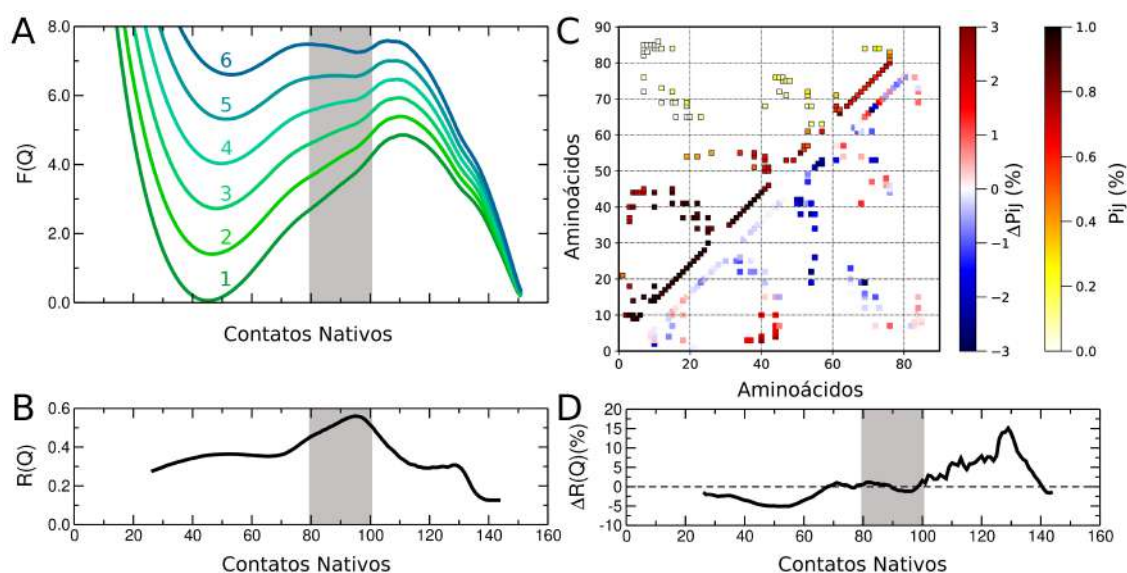
No sistema Elec2, as histinas 40 e 47 apresentam cargas iguais a zero (vide a Tabela 6.1) e, de acordo com a Figura 6.2A, a barreira de energia livre é maior para esse sistema. As mesmas análises que foram realizadas para o sistema Elec1 serão realizadas para o sistema Elec2 com a finalidade de observar e determinar o estado de pré-transição ou estado intermediário.

6.3.3 Estado intermediário para o sistema denominado Elec2: H40(0) e H47(0)

Na Figura 6.5A, têm-se os perfis de energia livre para o sistema Elec2 em diferentes temperaturas. Foram apresentados seis perfis de energia livre, sendo que o primeiro, T₁, corresponde à temperatura de enovelamento para esse sistema. Conforme foi destacada uma região de interesse no perfil de energia para o sistema Elec1, no sistema Elec2, essa mesma região foi observada e selecionada para fins de comparação entre os dois sistemas. Portanto, foram selecionadas as estruturas presentes dentro dessa região para estudos posteriores. Na Figura 6.5B, tem-se cálculo da rota de enovelamento, $R(Q)$, para a trajetória completa na temperatura T₁. O mesmo perfil de rota foi observado no intervalo entre 80 e 100 contatos nativos, onde um pico foi notado para essa faixa e, para valores próximos de 130, um novo aumento no $R(Q)$ é verificado. Devido aos dois sistemas apresentarem comportamentos semelhantes para $F(Q)$ e $R(Q)$, o passo seguinte será analisar as diferenças entre os perfis de rota e as probabilidades de formação dos contatos nativos.

Com a probabilidade de formação dos contatos nativos e a diferença da probabilidade entre os modelos, pode-se ter uma ideia das mudanças estruturais que ocorrem quando a Im7 encontra-se em diferentes valores de pH. Na Figura 6.5C, triângulo superior, é apresentada probabilidade de formação dos contatos nativos. Os contatos entre as α -hélices apresentam baixa probabilidade de formação entre os contatos da α -hélice quatro com os das α -hélices um e dois. Comportamento semelhante foi constatado para o sistema Elec1. Ainda na Figura 6.5C, porém no triângulo inferior, foi calculada a diferença de probabilidade entre os contatos nativos do sistema Elec1 e Elec2. Portanto, para valores negativos, assumi-se que contatos são formados para o sistema Elec1, enquanto que os valores positivos representam os contatos formados no sistema Elec2. Nota-se que a menor e maior diferença fica entorno de 3%, valor muito pequeno para assumir que nos sistemas adotados ocorreu uma mudança estrutural, ou seja, que foi alterado o mecanismo de enovelamento.

Figura 6.5: (A) Perfil de energia livre, $F(Q)$, em função dos contatos nativos para diferentes temperaturas. (B) Rota de enovelamento, $R(Q)$, em função dos contatos nativos, calculada na temperatura de enovelamento, $T_2 = 0,79$. (C) Triângulo superior, probabilidade de formação dos contatos nativos para o conjunto de estruturas selecionadas a partir da região destacada no perfil de energia livre para a temperatura T_2 . Triângulo inferior, diferença na probabilidade de formação dos contatos nativos entre as estruturas representativas dos sistemas Elec1 e Elec2. (D) Diferença percentual das rotas de enovelamento para os sistemas Elec1 e Elec2.



Fonte: Elaborada pelo autor.

Na 6.5D é exibido o cálculo da diferença percentual entre as rotas de enovelamento para os sistemas Elec1 e Elec2. Logo, não foi observado nenhuma diferença significativa no intervalo de Q mencionado anteriormente, reforçando mais uma vez que

o estado de pré-transição manteve-se o mesmo para os dois sistemas adotados. Contudo, nota-se que, para valores próximos de $Q = 130$, há uma diferença em $\Delta R(Q)$, onde o sistema Elec2, apresenta uma especificidade superior para esta região quando comparado com o sistema Elec1. Podemos relacionar esse estado pós-transição com o estado pré-transição, pois ao observar o estado pré-transição do sistema Elec2, nota-se que esse estado apresenta menor barreira de energia livre nas temperaturas inferiores que a temperatura de enovelamento, quando comparado com a barreira de energia livre do sistema Elec1.

Com o resultado da diferença percentual da rota, $\Delta R(Q)$, é possível inferir que o estado pré-transição do sistema Elec1 pode auxiliar o enovelamento, pois este sistema apresenta menor barreira de energia no estado de transição, como também, pode auxiliar a vencer mais facilmente a barreira de energia para dessolvatar o sistema. Além disso, os resultados computacionais corroboram qualitativamente com os dados experimentais, onde foi possível determinar a estrutura de referência que possivelmente representa o estado de pré-transição ou estado intermediário mencionado por dados experimentais.

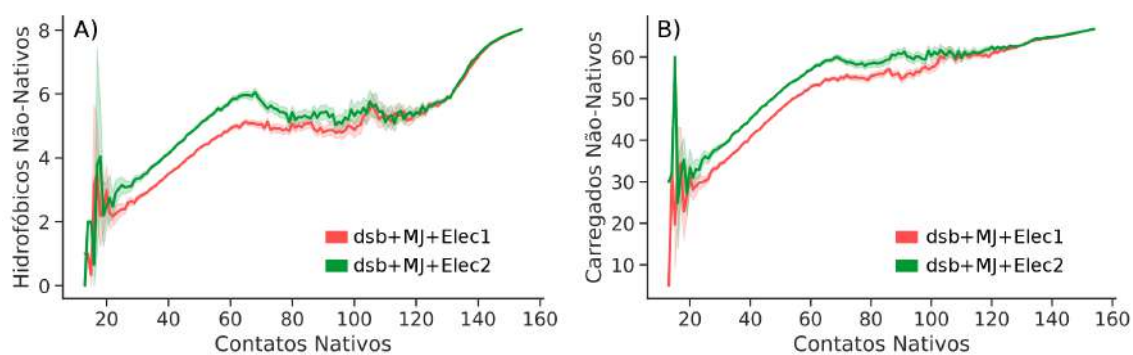
6.3.4 Densidade de contatos não-nativos, hidrofóbicos e eletrostáticos, para os sistemas Elec1 e Elec2

Conforme descrito na Figura 6.2A, o efeito da adição de carga aos resíduos carregados causou um aumento significativo na barreira de energia livre. Por outro lado, o sistema ausente do efeito eletrostático (dsb+HP), apresentou uma barreira desolvatação durante o processo de enovelamento, barreira que não foi observada nos sistemas Elec1 e Elec2. Portanto, o efeito eletrostático, como já mencionado, desempenha um papel importante no empacotamento das hélices devido a uma possível alteração no percurso do enovelamento após o estado de transição. Além do mais, o efeito da mudança de carga nas histidinas apresentaram diferentes valores na barreira de energia livre. Assim, existe um efeito a ser compreendido entre essas duas forças, hidrofóbica e eletrostática, durante o processo de enovelamento da Im7.

Na Figura 6.6, tem-se o número de contatos não-nativos, hidrofóbicos (Figure 6.6A) e carregados (Figure 6.6B), que foram formados entre a hélice quatro com as hélices um, dois e três. O cálculo foi realizado utilizando a hélice quatro como referência, pois a mesma não apresenta contatos nativos com alta probabilidade de formação no estágio de interesse, estado intermediário (ver Figura 6.4 e 6.5). Conforme pode-se acompanhar pelo coordenação de reação que foi adotado para descrever o processo de enovelamento, contatos nativos (Q), vê-se que o número de contatos não-nativos

aumentam de forma uniforme, chegando ao estado nativo com valores idênticos. Vale ressaltar que para o sistema Elec2, foi observado uma densidade superior no número de contatos não-nativo quando comparado com o sistema Elec1 em determinadas regiões. Sugerimos nesse caso, que o aumento na barreira de energia livre deve-se a essa diferença no número de contatos não-nativos, $Q < 110$. Entretanto, para valores $Q > 110$, ambos apresentaram números iguais, ponto no qual a proteína vence a barreira energia do estado de transição.

Figura 6.6: Cálculo do número de contatos não-nativos, hidrofóbicos (A) e eletrostáticos (B), em função do número de contatos nativos. O sistema Elec1 e Elec2 estão representados pelas linhas sólidas nas cores vermelho e verde, respectivamente.

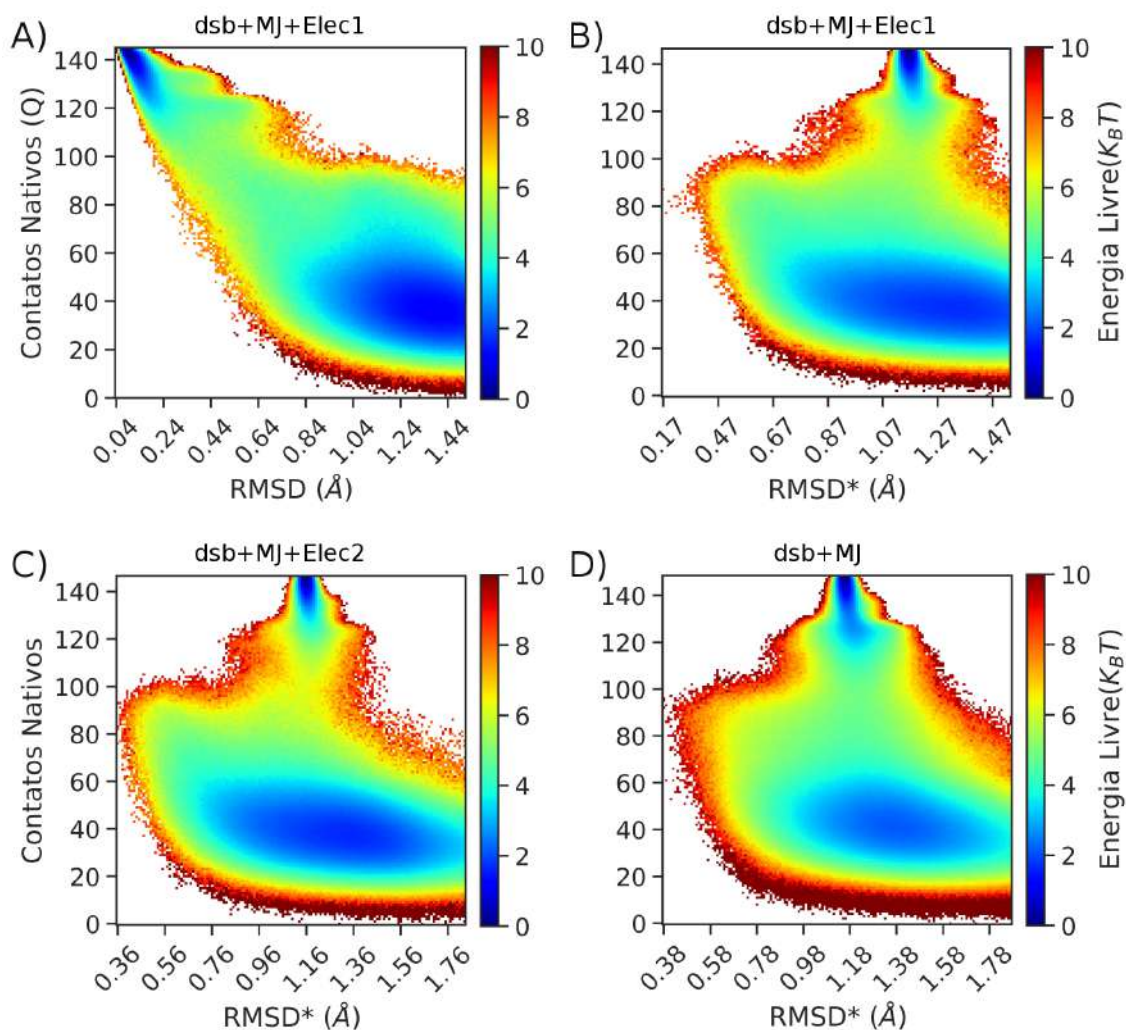


Fonte: Elaborada pelo autor.

De acordo com os resultados experimentais publicados por Gsponer [128], as interações não-nativas, principalmente aquelas formadas a partir da hélice quatro, desempenham um papel de estabilização do estado intermediário da Im7. Vale ressaltar que as mudanças nos valores de carga foram realizadas na H40 e H47, seguindo os valores de pH dos dados experimentais [128]. Esses resultados exibem alta correlação com os dados de Gsponer, pois os valores de pH escolhidos alteram os valores das cargas das histidinas 40 e 47, resíduos localizados na hélice dois. O *ensemble* de estruturas do estado intermediário apresentado por Gsponer, exibem estruturas no qual as hélices dois e quatro interagem via interações não-nativas. Dessa forma, levando a proteína para uma estruturação diferente da nativa. Logo, os dados apresentados nesse trabalho corroboram com os resultados experimentais, onde a estrutura representativa do estado intermediário apresenta padrões estruturais similares as estruturas obtidas experimentalmente.

Afim de verificar que o efeito eletrostático foi o responsável pela estabilização do estado intermediário, foram calculados os RMSDs ao longo do processo de enovelamento e em seguida realizado o cálculo da energia livre em função dos contatos nativos

Figura 6.7: Cálculo da energia livre ($K_B T$) em função do número de contatos nativos e do RMSD (Å). (A) Para o sistema Elec1, o RMSD foi calculado utilizando a estrutura nativa da Im7 como referência. (B e C) Para os sistemas Elec1 e Elec2, foi utilizado no cálculo do RMSD* a estrutura do estado intermediário (*cluster* – 1) exibida na Figura 6.4A. (D) Para o sistema ausente do potencial eletrostático, foi utilizado o cálculo do RMSD* tendo a estrutura do estado intermediário como referência (6.4A).



Fonte: Elaborada pelo autor.

e do RMSD. No cálculo dos RMSDs foram utilizadas duas estruturas de referência, estrutura nativa da proteína Im7 e estrutura representativa do estado intermediário do sistema Elec1. Na Figura 6.7, têm-se os mapas de densidade com a escala de cor correspondendo aos valores de energia livre. Para o sistema Elec1, Figuras 6.7A e 6.7B, caso onde a estrutura nativa foi utilizada como referência, regiões significativas com alta densidades não ocorreram próximas ao estado de transição, mas sim, apenas para o estado desenovelado e nativo. Por outro lado, foi observado comportamento diferente

para o sistema no qual foi utilizado a estrutura do estado intermediário. Portanto, além dos estados desenovelado e nativo, a região correspondente ao estado intermediário apresenta um desvio significativo, no qual corresponde ao seu estado pré-transição.

Para o sistema Elec2 e para o sistema ausente dos efeitos eletrostáticos (dsb+MJ), Figuras 6.7C e 6.7D, respectivamente, o perfil de distribuição apresentou padrões similares entre os sistemas Ele1 e Elec2 e padrões diferentes em relação ao sistema dsb+MJ. Segundo os resultados apresentados nas seções anteriores, o sistema Elec2 apresenta os mesmos padrões estruturais apresentados pelo sistema Elec1. Contudo, para o sistema Elec2, ocorreu um aumento na barreira de energia livre, assim como, uma diminuição na densidade do número de estruturas do estado intermediário, corroborando dessa forma com os dados experimentais, no qual o aumento do valor pH, diminuiu a estabilização do estado intermediário. No caso do sistema onde as interações eletrostáticas estão ausentes, não foi observado um número considerável de estruturas correspondentes aos mesmos estados destacados no sistema Elec1 e Elec2. Portanto, as simulações realizadas com o último modelo, dsb+MJ, não apresentaram padrões similares estruturais com alta densidade que caracterizaram o estados intermediário dos sistemas Elec1 e Elec2.

Em suma, os sistemas abordados nesse capítulo foram capazes de reproduzir dados importantes que antes não foram evidenciados em simulações computacionais. A adição dos potenciais não-nativos tornou possível uma análise mais detalhada do processo de enovelamento, já que os dados experimentais levantaram questões importantes de pontos que envolvem os efeitos eletrostáticos e hidrofóbicos. Interessante notar que com a adição do potencial eletrostático ocorreu um aumento considerável na barreira de energia, enquanto que com a mudança no valor das cargas, foi observado uma redução na densidade de conformações do estado de pré-transição. Esse comportamento passa a ser importante para o sistema estudado, pois torna-se evidente como pequenas alterações podem acarretar mudanças significativas para o sistema.

6.4 Conclusão

O modelo proposto para estudar o enovelamento da proteína Im7 foi o modelo computacional que abordou três potenciais que mimetizam as interações envolvidas em um sistema *in vitro*. Os potenciais acoplados ao modelo SBM-C α foram os potenciais de dessolvatação, eletrostático e hidrofóbico. Com o potencial de dessolvatação, foi possível analisar sua influência no perfil de energia livre, onde ocorreu o surgimento de uma barreira na energia livre para a proteína alcançar seu estado nativo. Lembrando que essa barreira surgiu devido ao sistema ter que retirar uma molécula de água presente entre dois contatos nativos. Esse modelo de simulação foi realizado em conjunto com o potencial hidrofóbico. Já com a adição do potencial eletrostático, em conjunto com o potencial de dessolvatação e hidrofóbico, verificou-se a formação de um estado pré-transição e um aumento na barreira de energia livre ao considerar diferentes valores de cargas para as histinas. Além disso, ocorreu uma diminuição na barreira de energia do pré-estado nativo quando foi considerado apenas os potenciais de dessolvatação e hidrofóbico.

Com os diferentes valores de carga para as histidinas, foi possível obter uma similaridade entre as barreiras de energia livre com as barreiras de desnaturação obtidas experimentalmente para diferentes valores de pH. O sistema Elec1 foi o modelo que representou o sistema para valores de pH abaixo de 6,0. Esse sistema, do ponto de vista experimental, é o sistema que apresenta uma estabilidade do estado intermediário. Em relação às simulações realizadas, foi possível determinar um estado pré-transição, que foi e continuará sendo estudado com a finalidade de caracterizá-lo como o estado intermediário reportado experimentalmente. Já o sistema Elec2 corresponde ao sistema que apresenta pH acima de 6,6 e corresponde a um sistema, do ponto de vista experimental, com menor estabilidade do estado intermediário. Os resultados obtidos com as simulações computacionais para esse sistema apresentou um estado pré-transição menos estável, quando comparado com o estado citado anteriormente.

Além da determinação do estado de pré-transição, foram obtidas as estruturas correspondentes a esse estado, para o pH abaixo de 6,6, assim como sua probabilidade de formação de contatos nativos. Com esses resultados, foi verificado a existência de um estado correspondente ao citado experimentalmente, onde ocorreu a estruturação não nativa da α -hélice quatro com as α -hélices um e dois. Essa formação ocorre para que a α -hélice três possa encaixar-se corretamente e, assim, posteriormente, a α -hélice quatro consiga se estruturar de forma nativa com o restante da proteína. Resultado muito importante que mostrou que os potenciais não-nativos passam a capturar um estado intermediário que ocorre com baixa frequência e que, até o momento, não tinha

sido observado com as técnicas computacionais empregadas.

Os trabalhos em andamento visam explorar o estado intermediário determinado experimentalmente para a proteína Im9. As simulações já foram realizadas com os mesmos parâmetros utilizadas para Im7. Entretanto, foram empregados diferentes valores de carga, pois a proteína Im9 apresenta mais do que duas histinas em sua sequência. O estado intermediário detectado para a Im9 foi determinado em outra faixa de pH, do ponto de vista experimental. Já em nossas simulações, foi determinado um estado de pré-transição, seguindo o mesmo estudo realizado para a Im7. Contudo, esses resultados não foram apresentados no trabalho, pois devem ser analisados e estudados detalhadamente. Como última análise proposta para esse trabalho, será calculado o tempo de enovelamento, para a construção do *chevron plot*, pois essa análise nos possibilitará visualizar se o estado de pré-transição proposto nesse trabalho poderá corresponder ao estado intermediário descrito experimentalmente.

Em conclusão, vê-se que o modelo porposto de simulação nesse capítulo foi capaz de capturar e obter os resultados observados experimentalmente. Esse modelo não foi empregado em nenhum estudo realizado anteriormente e sua validação ainda está em processo. Por isso, até o momento, foram apresentados os resultados com a proteína Im7, pois seu estado intermediário é obtido de forma estável e, por isso, acreditamos que venha ser mais fácil de obtê-lo do ponto de vista computacional. Assim, como citado anteriormente, todos os resultados obtidos para Im7 e Im9 serão apresentados em uma próxima versão desse trabalho.

Capítulo 7

Considerações Finais

O estudo desenvolvido com a proteína CI2 foi realizado seguindo o mesmo modelo de simulação proposto para os domínios R15, R16 e R17 da α -espectrina. Os dados computacionais para esses domínios revelaram que as diferenças nos tempos de enovelamento podem ser compreendidas em termos das interações não-nativas. Com a finalidade de investigar os efeitos mutacionais na termoestabilidade e no tempo de enovelamento da CI2, foi adequada a escolha do modelo, pois as mutações foram realizadas em resíduos carregados e hidrofóbicos e o modelo utilizado capturou os efeitos decorrentes dessas interações. Além disso, os dados computacionais revelaram que o comportamento cooperativo observado nos domínios da α -espectrina não foi o mesmo verificado na CI2 com as mutações. Acredita-se que as mutações que melhoram a termoestabilidade, geralmente, estão associadas à aceleração do processo de enovelamento. Entretanto, os resultados obtidos foram contraintuitivos. Enquanto um dos efeitos diminuiu a termoestabilidade, o outro foi responsável pela alteração no tempo de enovelamento.

No caso dos domínios R15WT, R16WT e R16M₅ da α -espectrina, os efeitos oriundos dos potenciais não-nativos foram capazes de elucidar as diferenças nas taxas de enovelamento, assim como, obter, aproximadamente, a mesma ordem de grandeza nos tempos de enovelamento dos domínios R15WT e R16M₅. Devido a esse resultado, foi investigado o perfil energético das trajetórias de enovelamento com a finalidade de observar regiões de interesse que pudessem caracterizar os efeitos não-nativos. Análise semelhante foi realizada no estudo da CI2, onde uma região se destacou e, assim, foi interpretada como sendo uma região crucial durante o processo de enovelamento.

As simulações realizadas com as proteínas Im7 e Im9 foram executadas considerando um potencial a mais no modelo utilizado para as proteínas CI2, R15WT, R16WT e R16R16M₅. Contudo, antes de chegar a esse modelo, foram realizadas si-

mulações combinando diferentes técnicas, onde buscou-se investigar os estados intermediários dessas proteínas. Do ponto de vista experimental, os estados intermediários são estabilizados em diferentes valores de pH e, do ponto de vista computacional, foram realizadas simulações com carga constante para mimetizar os sistemas experimentais. O estado intermediário da Im7 foi observado quando quando foi incorporado, no modelo computacional, o potencial de dessolvatação. Os efeitos desses potenciais encontram-se em diferentes ordens de grandeza, mas devido a essas pequenas perturbações/contribuições, foi possível observar estados com baixa população, porém exercendo um papel importante durante o processo de enovelamento. Esse mesmo comportamento também é verificado com as proteínas CI2 e R16M₅. Logo, com a adição dos potenciais não-nativos, estados visitados com baixa frequência passaram a ser caracterizados por diferentes métodos, o que pode ser averiguado nos resultados apresentados neste trabalho.

Referências Bibliográficas

- 1 GOODALL, J. *Jane Goodall Quotes (Author of In the Shadow of Man)*. Goodreads. Acesso em: 18 june 2021. Disponível em: <https://www.goodreads.com/author/quotes/18163.Jane_Goodall>.
- 2 VOET, D. *Fundamentos de Bsiocquímica. A Vida em Nível Molecular (Em Português do Brasil)*. [S.l.]: Artmed, 2014.
- 3 NELSON, D. L.; COX, M. M. *Lehninger Principles of Biochemistry, Fourth Edition*. [S.l.]: W. H. Freeman, 2004.
- 4 PROTEIN structure – Wikimedia Commons, the free media repository. Acesso em: 23 nov. 2020. Disponível em: <[https://commons.wikimedia.org/w/index.php?title=File:Protein_structure__\(full\).png&oldid=480491567](https://commons.wikimedia.org/w/index.php?title=File:Protein_structure__(full).png&oldid=480491567)>.
- 5 KIM, P. M. et al. The role of disorder in interaction networks: a structural analysis. *Mol. Syst. Biol.*, v. 4, n. 1, p. 179, 2008.
- 6 WRIGHT, P. E.; DYSON, H. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, v. 293, n. 2, p. 321 – 331, 1999.
- 7 GALEA, C. A. et al. Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. *Biochemistry*, v. 47, n. 29, p. 7598–7609, 2008.
- 8 GSPONER, J. et al. Tight regulation of unstructured proteins: From transcript synthesis to protein degradation. *Science*, v. 322, n. 5906, p. 1365–1368, 2008.
- 9 BABU, M. M. et al. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.*, v. 21, n. 3, p. 432 – 440, 2011.
- 10 DOBSON, C. M. Protein folding and misfolding. *Nature*, v. 426, n. 6968, p. 884–890, 2003.
- 11 HATTERS, D. M. Protein misfolding inside cells: The case of huntingtin and huntington’s disease. *IUBMB Life*, v. 60, n. 11, p. 724–728, 2008.
- 12 LEVINTHAL, C. Are there pathways for protein folding? *J. Chem. Phys.*, v. 65, n. 1, p. 44–45, 1968.
- 13 ONUCHIC, J. N. et al. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios. In: . [S.l.: s.n.], 2000. v. 53, p. 87 – 152.

-
- 14 SHOEMAKER, B.; WANG, J.; WOLYNES, P. G. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. U.S.A.*, v. 94, n. 3, p. 777–782, 1997.
- 15 DILL, K. A.; CHAN, H. S. From levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.*, v. 4, n. 1, p. 10–19, 1997.
- 16 CHAVEZ, L. L.; ONUCHIC, J. N.; CLEMENTI, C. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.*, v. 126, n. 27, p. 8426–8432, 2004.
- 17 NETTELS, D. et al. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U.S.A.*, v. 104, n. 8, p. 2655–2660, 2007.
- 18 BRYNGELSON, J. D. et al. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Bioinf.*, v. 21, n. 3, p. 167–195, 1995.
- 19 CLEMENTI, C.; NYMEYER, H.; ONUCHIC, J. N. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, v. 298, n. 5, p. 937–953, 2000.
- 20 WHITFORD, P. C. et al. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Struct., Funct., Bioinf.*, v. 75, n. 2, p. 430–441, 2009.
- 21 LU, Q.; LU, H. P.; WANG, J. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Physical Review Letters*, v. 98, n. 12, 2007.
- 22 WHITFORD, P. C. et al. Conformational transitions of adenylate kinase: Switching by cracking. *J. Mol. Biol.*, v. 366, n. 5, p. 1661–1671, 2007.
- 23 BEST, R. B.; CHEN, Y.-G.; HUMMER, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure*, v. 13, n. 12, p. 1755–1763, 2005.
- 24 ZUCKERMAN, D. M. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *J. Phys. Chem. B*, v. 108, n. 16, p. 5127–5137, 2004.
- 25 LU, Q.; WANG, J. Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *J. Am. Chem. Soc.*, v. 130, n. 14, p. 4772–4783, 2008.
- 26 FOLDING funnel schematic – Wikimedia Commons, the free media repository. Acesso em: 23 nov. 2020. Disponível em: <https://commons.wikimedia.org/w/index.php?title=File:Folding_funnel_schematic.svg&oldid=509924168>.
- 27 BRYNGELSON, J. D.; WOLYNES, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, v. 84, n. 21, p. 7524–7528, 1987.

-
- 28 BRYNGELSON, J. D.; WOLYNES, P. G. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, v. 93, n. 19, p. 6902–6915, 1989.
- 29 SALI, A.; SHAKHNOVICH, E.; KARPLUS, M. How does a protein fold? *Nature*, v. 369, n. 6477, p. 248–251, 1994.
- 30 WOLYNES, P. G. Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie*, v. 119, p. 218 – 230, 2015.
- 31 SNOW, C. D. et al. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Bioph. Biom.*, v. 34, n. 1, p. 43–69, 2005.
- 32 MOURO, P. R. *Estudo da Interação não-nativa no enovelamento de proteínas*. Dissertação (Mestrado em Biofísica Molecular) — UNESP, São José do Rio Preto, 2014.
- 33 CHAN, H. S.; DILL, K. A. Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.*, v. 100, n. 12, p. 9238–9257, 1994.
- 34 BROCKWELL, D. J.; RADFORD, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.*, v. 17, n. 1, p. 30–37, 2007.
- 35 BARTLETT, A. I.; RADFORD, S. E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.*, v. 16, n. 6, p. 582–588, 2009.
- 36 CHEN, T.; SONG, J.; CHAN, H. S. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.*, v. 30, p. 32–42, 2015.
- 37 TAKETOMI, H.; UEDA, Y.; GÖ, N. Studies on protein folding, unfolding and fluctuations by computer simulation: I. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.*, v. 7, n. 6, p. 445–459, 1975.
- 38 NOEL, J. K. et al. Smog 2: A versatile software package for generating structure-based models. *PLoS Comput. Biol.*, v. 12, n. 3, p. 1–14, 03 2016.
- 39 SUN, Y.; MING, D. Energetic frustrations in protein folding at residue resolution: A homologous simulation study of Im9 proteins. *PLoS One*, v. 9, n. 1, p. e87719, 2014.
- 40 CONTESSOTO, V. G. et al. Analyzing the effect of homogeneous frustration in protein folding. *Proteins: Struct., Funct., Bioinf.*, v. 81, n. 10, p. 1727–1737, out. 2013.
- 41 MENDONÇA, M. R. de et al. Inferring a weighted elastic network from partial unfolding with coarse-grained simulations. *Proteins: Struct., Funct., Bioinf.*, v. 82, n. 1, p. 119–129, 2014.

-
- 42 TRIPATHI, S.; GARCÍA, A. E.; MAKHATADZE, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J. Phys. Chem. B*, v. 119, n. 41, p. 13103–13112, out. 2015.
- 43 CLEMENTI, C.; PLOTKIN, S. S. The effects of nonnative interactions on protein folding rates: Theory and simulation. *Protein Sci.*, v. 13, n. 7, p. 1750–1766, 2004.
- 44 SILVA, F. Bruno da et al. Non-native cooperative interactions modulate protein folding rates. *J. Phys. Chem. B*, v. 122, n. 48, p. 10817–10824, 2018.
- 45 MIYAZAWA, S.; JERNIGAN, R. L. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins: Struct., Funct., Bioinf.*, v. 34, n. 1, p. 49–68, 1999.
- 46 ZHANG, Z.; CHAN, H. S. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc. Natl. Acad. Sci. U. S. A.*, v. 107, n. 7, p. 2920–2925, 2010.
- 47 ULLNER, M.; WOODWARD, C. E.; JÖNSSON, B. A Debye-Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.*, v. 105, p. 2056–2065, 1996.
- 48 PAPOIAN, G. A.; ULANDER, J.; WOLYNES, P. G. Role of water mediated interactions in protein–protein recognition landscapes. *J. Am. Chem. Soc.*, v. 125, n. 30, p. 9170–9178, 2003.
- 49 LIU, Z.; CHAN, H. S. Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Physical Biology*, v. 2, n. 4, p. S75–S85, 2005.
- 50 HILLSON, N.; ONUCHIC, J. N.; GARCIA, A. E. Pressure-induced protein-folding/unfolding kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, v. 96, n. 26, p. 14848–14853.
- 51 CHEUNG, M. S.; GARCÍA, A. E.; ONUCHIC, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.*, v. 99, n. 2, p. 685–690, 2002.
- 52 TANFORD, C.; KIRKWOOD, J. G. Theory of protein titration curves. i. general equations for impenetrable spheres. *J. Am. Chem. Soc.*, v. 79, n. 20, p. 5333–5339, 1957.
- 53 CONTESSOTO, V. G. et al. Tksa-mc: A web server for rational mutation through the optimization of protein charge interactions. *Proteins: Struct., Funct., Bioinf.*, v. 86, n. 11, p. 1184–1188, 2018.
- 54 HAVRANEK, J. J.; HARBURY, P. B. Tanford–kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. U.S.A.*, v. 96, n. 20, p. 11145–11150, 1999.
- 55 SHIRE, S.; HANANIA, G.; GURD, F. N. Electrostatic effects in myoglobin. hydrogen ion equilibria in sperm whale ferrimyoglobin. *Biochemistry*, v. 13, n. 14, p. 2967–2974, 1974.

-
- 56 TANFORD, C.; ROXBY, R. Interpretation of protein titration curves. application to lysozyme. *Biochemistry*, v. 11, n. 11, p. 2192–2198, 1972.
- 57 ORTTUNG, W. H. Proton binding and dipole moment of hemoglobin. refined calculations. *Biochemistry*, v. 9, n. 12, p. 2394–2402, 1970.
- 58 KIRKWOOD, J. G. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.*, v. 2, n. 7, p. 351–361, 1934.
- 59 SILVA, F. L. B. D.; JÖNSSON, B.; PENFOLD, R. A critical investigation of the tanford-kirkwood cheme by means of monte carlo simulations. *Protein Sci.*, v. 10, n. 7, p. 1415–1425, 2001.
- 60 METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, v. 21, n. 6, p. 1087–1092, jun. 1953.
- 61 PLOTKIN, S. S.; ONUCHIC, J. N. Investigation of routes and funnels in protein folding by free energy functional methods. *Proc. Natl. Acad. Sci. U. S. A.*, v. 97, n. 12, p. 6509–6514, jun. 2000.
- 62 LAMMERT, H.; NOEL, J. K.; ONUCHIC, J. N. The dominant folding route minimizes backbone distortion in sh3. *PLoS Comput. Biol.*, v. 8, n. 11, p. 1–12, 11 2012.
- 63 JUNIOR, A. B. O. et al. Distinguishing biomolecular pathways and metastable states. *J. Chem. Theory Comput.*, 2019.
- 64 HARDIN, C. et al. Associative memory hamiltonians for structure prediction without homology: / proteins. *Proc. Natl. Acad. Sci. U.S.A.*, v. 100, n. 4, p. 1679–1684, 2003.
- 65 JUNIOR, A. B. de O. *Métodos para visualização de superfície de energia do enovelamento de proteínas*. Dissertação (Doutorado em Biofísica Molecular) — UNESP, São José do Rio Preto, 2018.
- 66 TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multi-dimensional data sets. *Inform. Visual.*, v. 2, n. 4, p. 218–231, 2003.
- 67 FERRENBURG, A. M.; SWENDSEN, R. H. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.*, v. 61, p. 2635–2638, 1988.
- 68 KUMAR, S. et al. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comp. Chem.*, v. 13, n. 8, p. 1011–1021, 1992.
- 69 KUMAR, S. et al. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, v. 16, n. 11, p. 1339–1350, 1995.

-
- 70 NOEL, J. K.; WHITFORD, P. C.; ONUCHIC, J. N. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *J. Phys. Chem. B*, v. 116, n. 29, p. 8692–8702, 2012.
- 71 BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, v. 81, n. 8, p. 3684–3690, 1984.
- 72 YANG, H. et al. Rational design to improve protein thermostability: Recent advances and prospects. *Chem. Bio. Eng. Rev.*, v. 2, n. 2, p. 87–94, 2015.
- 73 SEELIGER, D. et al. Boosting antibody developability through rational sequence optimization. *mAbs*, v. 7, n. 3, p. 505–515, 2015.
- 74 FENG, R. et al. Rational design of xylose dehydrogenase for improved thermostability and its application in development of efficient enzymatic biofuel cell. *Enzyme Microb. Technol.*, v. 84, p. 78–85, 2016.
- 75 THOMPSON, P. et al. Rational design, biophysical and biological characterization of site-specific antibody-tubulysin conjugates with improved stability, efficacy and pharmacokinetics. *J. Controlled Release*, v. 236, p. 100–116, 2016.
- 76 GALDIERO, S.; GOMES, P. A. C. Peptide-based drugs and drug delivery systems. *Molecules*, v. 22, n. 12, p. 2185, 2017.
- 77 SACHDEVA, S.; LOBO, S.; GOSWAMI, T. What is the future of noninvasive routes for protein- and peptide-based drugs? *Ther. Delivery*, v. 7, n. 6, p. 355–357, 2016.
- 78 MOHANDESI, N. et al. Catalytic efficiency and thermostability improvement of suc2 invertase through rational site-directed mutagenesis. *Enzyme Microb. Technol.*, v. 96, p. 14–22, 2017.
- 79 KELLOGG, E. H.; LEAVER-FAY, A.; BAKER, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Struct., Funct., Bioinf.*, v. 79, n. 3, p. 830–838, 2011.
- 80 YIN, S.; DING, F.; DOKHOLYAN, N. V. Modeling backbone flexibility improves protein stability estimation. *Structure*, v. 15, n. 12, p. 1567–1576, 2007.
- 81 MASSO, M.; VAISMAN, I. I. Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics*, v. 24, n. 18, p. 2002–2009, 2008.
- 82 HUANG, L.-T.; GROMIHA, M. M. Reliable prediction of protein thermostability change upon double mutation from amino acid sequence. *Bioinformatics*, v. 25, n. 17, p. 2181–2187, 2009.
- 83 CAPRIOTTI, E.; FARISELLI, P.; CASADIO, R. I-mutant2. 0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, v. 33, n. suppl_2, p. W306–W310, 2005.

-
- 84 CHENG, J. et al. Scratch: A protein structure and structural feature prediction server. *Nucleic Acids Res.*, v. 33, n. suppl_2, p. W72–W76, 2005.
- 85 JIA, L.; YARLAGADDA, R.; REED, C. C. Structure based thermostability prediction models for protein single point mutations with machine learning tools. *PloS one*, v. 10, n. 9, p. e0138022, 2015.
- 86 FANG, J. A Critical Review of Five Machine Learning-Based Algorithms for Predicting Protein Stability Changes Upon Mutation. *Briefings Bioinf.*, v. 00, n. 00, p. 1–8, 2019.
- 87 PETTERSEN, E. F. et al. Ucsf chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.*, v. 25, n. 13, p. 1605–1612, 2004.
- 88 BOND, C. S.; SCHÜTTELKOPF, A. W. *ALINE*: A WYSIWYG Protein-Sequence Alignment Editor for Publication-Quality Alignments. *Acta Crystallogr., Sect. D: Struct. Biol.*, v. 65, n. 5, p. 510–512, 2009.
- 89 SILVA, F. B. da et al. Rational design of chymotrypsin inhibitor 2 by optimizing non-native interactions. *J. Chem. Inf. Model.*, v. 60, n. 2, p. 982–988, 2020.
- 90 CHEN, T.; CHAN, H. S. Native contact density and nonnative hydrophobic effects in the folding of bacterial immunity proteins. *PLoS Comput. Biol.*, v. 11, n. 5, p. 1–27, 05 2015.
- 91 WU, J. et al. The low populated folding intermediate of a mutant of the fyn sh3 domain identified by a simple model. *Phys. Chem. Chem. Phys.*, v. 19, p. 22321–22328, 2017.
- 92 YRAZU, F. M.; PINAMONTI, G.; CLEMENTI, C. The effect of electrostatic interactions on the folding kinetics of a 3- α -helical bundle protein family. *J. Phys. Chem. B*, v. 122, n. 49, p. 11800–11806, 2018.
- 93 TZUL, F. O.; SCHWEIKER, K. L.; MAKHATADZE, G. I. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. U. S. A.*, v. 112, n. 3, p. E259–E266, 2015.
- 94 MCPHALEN, C.; JAMES, M. Crystal and molecular structure of the serine proteinase inhibitor ci-2 from barley seeds. *Biochemistry*, v. 26, n. 1, p. 261–269, 1987.
- 95 ITZHAKI, L. S.; OTZEN, D. E.; FERSHT, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.*, v. 254, n. 2, p. 260–288, 1995.
- 96 JACKSON, S. E.; FERSHT, A. R. Folding of chymotrypsin inhibitor 2. 1. evidence for a two-state transition. *Biochemistry*, v. 30, n. 43, p. 10428–10435, 1991.
- 97 LADURNER, A. G. et al. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, v. 95, n. 15, p. 8473–8478, 1998.

-
- 98 LAWRENCE, C. et al. Investigation of an anomalously accelerating substitution in the folding of a prototypical two-state protein. *J. Mol. Biol.*, v. 403, n. 3, p. 446–458, 2010.
- 99 WEBB, B.; SALI, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinf.*, v. 54, n. 1, p. 5.6.1–5.6.37, 2016.
- 100 CONTESSOTO, V. G. et al. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.*, v. 12, n. 7, p. 3270–3277, 2016.
- 101 CORONADO, M. A. et al. Cold shock protein A from *Corynebacterium pseudotuberculosis*: Role of electrostatic forces in the stability of the secondary structure. *Protein Pept. Lett.*, v. 24, n. 4, p. 358–367, 2017.
- 102 OLIVEIRA, V. M. de et al. Effects of pH and salt concentration on stability of a protein g variant using coarse-grained models. *Biophys. J.*, v. 114, n. 1, p. 65–75, 2018.
- 103 GRIBENKO, A. V. et al. Rational stabilization of enzymes by computational redesign of surface charge–charge interactions. *Proc. Natl. Acad. Sci. U. S. A.*, v. 106, n. 8, p. 2601–2606, 2009.
- 104 PLOTKIN, S. S. Speeding protein folding beyond the $g\bar{o}$ model: How a little frustration sometimes helps. *Proteins: Struct., Funct., Bioinf.*, v. 45, n. 4, p. 337–345, 2001.
- 105 CONTESSOTO, V. G. et al. Analyzing the effect of homogeneous frustration in protein folding. *Proteins: Struct., Funct., Bioinf.*, v. 81, n. 10, p. 1727–1737, 2013.
- 106 MOURO, P. R. et al. Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophys. J.*, v. 111, n. 2, p. 287–293, 2016.
- 107 ZARRINE-AFSAR, A. et al. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the fyn sh3 domain. *Proteins: Struct., Funct., Bioinf.*, v. 80, n. 3, p. 858–870, 2012.
- 108 SOCCI, N. D.; ONUCHIC, J. N.; WOLYNES, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.*, v. 104, n. 15, p. 5860–5868, 1996.
- 109 LEOPOLD, P. E.; MONTAL, M.; ONUCHIC, J. N. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. U. S. A.*, v. 89, n. 18, p. 8721–8725, 1992.
- 110 KUBELKA, J.; HOFRICHTER, J.; EATON, W. A. The protein folding ‘speed limit’. *Curr. Opin. Struct. Biol.*, v. 14, n. 1, p. 76–88, 2004.
- 111 FERSHT, A. R.; MATOUSCHEK, A.; SERRANO, L. The folding of an enzyme. *J. Mol. Biol.*, v. 224, n. 3, p. 771–782, 1992.

-
- 112 SCOTT, K. A.; RANGLES, L. G.; CLARKE, J. The folding of spectrin domains ii: Phi-value analysis of R16. *J. Mol. Biol.*, v. 344, n. 1, p. 207–221, 2004.
- 113 SCOTT, K. A. et al. The folding of spectrin domains i: Wild-type domains have the same stability but very different kinetic properties. *J. Mol. Biol.*, v. 344, n. 1, p. 195–205, 2004.
- 114 SCOTT, K. A.; CLARKE, J. Spectrin R16: Broad energy barrier or sequential transition states? *Protein Sci.*, v. 14, n. 6, p. 1617–1629, 2005.
- 115 SCOTT, K. A. et al. The folding pathway of spectrin R17 from experiment and simulation: Using experimentally validated MD simulations to characterize states hinted at by experiment. *J. Mol. Biol.*, v. 359, n. 1, p. 159–173, 2006.
- 116 WENSLEY, B. G. et al. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature*, v. 463, n. 7281, p. 685–688, Feb 2010.
- 117 WENSLEY, B. G. et al. Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. *Proc. Natl. Acad. Sci. U. S. A.*, v. 109, n. 44, p. 17795–17799, 2012.
- 118 BORGIA, A. et al. Localizing Internal Friction Along the Reaction Coordinate of Protein Folding by Combining Ensemble and Single-Molecule Fluorescence Spectroscopy. *Nat. Commun.*, v. 3, p. 1195, 2012.
- 119 BEST, R. B. How well does a funneled energy landscape capture the folding mechanism of spectrin domains? *J. Phys. Chem. B*, v. 117, n. 42, p. 13235–13244, 2013.
- 120 PASCUAL, J. et al. The spectrin repeat folds into a three-helix bundle in solution. *FEBS Lett.*, v. 383, n. 3, p. 201–207, 1996.
- 121 PASCUAL, J. et al. Solution structure of the spectrin repeat: A left-handed antiparallel triple-helical coiled-coil. *J. Mol. Biol.*, v. 273, n. 3, p. 740–751, oct 1997.
- 122 KUSUNOKI, H. et al. Independent movement, dimerization and stability of tandem repeats of chicken brain α -spectrin. *J. Mol. Biol.*, v. 344, n. 2, p. 495–511, 2004.
- 123 KWA, L. G. et al. The folding of a family of three-helix bundle proteins: Spectrin r15 has a robust folding nucleus, unlike its homologous neighbours. *Journal of Molecular Biology*, v. 426, n. 7, p. 1600–1610, 2014.
- 124 WENSLEY, B. G. et al. Protein folding: Adding a nucleus to guide helix docking reduces landscape roughness. *J. Mol. Biol.*, v. 423, n. 3, p. 273–283, 2012.
- 125 CAPALDI, A. P.; KLEANTHOUS, C.; RADFORD, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.*, n. 9, p. 209–2016, 2002.
- 126 GORSKI, S. A. et al. Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J. Mol. Biol.*, v. 312, n. 4, p. 849–863, 2001.

- 127 WHITTAKER, S. B.-M. et al. NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol.*, v. 366, n. 3, p. 1001–1015, 2007.
- 128 GSPONER, J. et al. Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U.S.A.*, v. 103, n. 1, p. 99–104, 2006.
- 129 ALLEN, M.; TILDESLEY, D. *Computer simulation of liquids Oxford Science*. [S.l.]: London, 1990.
- 130 SCHERER, C. *Métodos Computacionais da Física*. [S.l.]: Editora Livraria da Física, 2005.

Apêndice A

Dinâmica Molecular

A Dinâmica Molecular é um método de simulação computacional que resolve as equações de movimento de átomos e moléculas. Este método considera as partículas como sendo puntiformes, onde interagem com outras partículas do sistema e, possivelmente, com partículas externas. À medida em que os átomos variam seus estados, caracterizados pelas posições e velocidades, calculam-se as variáveis relevantes do sistema, tais como, temperatura e energia, em função da posição [129, 130].

Devido ao enorme número de átomos que um sistema pode apresentar, torna-se inviável computacionalmente a descrição das interações do sistema pela Mecânica Quântica, com isso recorre-se a Mecânica Clássica, pois esta contorna tais empecilhos. Esta técnica faz uso de aproximações, onde as partículas estão unidas por forças e são representadas por suas coordenadas, r_i , e momentos, p_i . Portanto, a Hamiltoniana, H , de um sistema clássico pode ser escrita como a soma das energias cinética e potencial [129], conforme a equação A.1,

$$H = \sum_{i=1}^N \frac{(p_i)^2}{2m_i} + V(r_i) \quad (\text{A.1})$$

onde p_i representa o momento da partícula i , r_i representa a coordenada da partícula i , m_i a massa da partícula i e N o número de partículas. O primeiro termo da equação A.1 refere-se a energia cinética do sistema e o segundo termo refere-se a energia potencial, sendo este último termo representado pelas equações descritas na seção de métodos deste trabalho.

Com a Hamiltoniano da equação A.1, é possível obter as equações que descrevem a evolução temporal do sistema, com isso as equações de Hamilton comumente

são escritas conforme a equação A.2 e A.3, chegando as equações de Newton,

$$\dot{r} = \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i} = v_i \quad (\text{A.2})$$

$$\dot{p} = -\frac{\partial H}{\partial q_i} = m_i \ddot{r} = -\nabla_i V(r_i) = -F_i \quad (\text{A.3})$$

onde $\dot{q}$ e $\ddot{r}$ são as velocidades e as acelerações da partícula i ,m respectivamente, e F_i é a força sobre cada partícula i . Portanto, a dinâmica molecular clássica consiste na resolução das equações A.2 e A.3.

Antes de inicializar uma dinâmica molecular clássica,deve-se seguir alguns passos obrigatoriamente: i) definir a posição, velocidade e aceleração para cada partícula e; ii) determinar o intervalo de tempo, δt , este valor deve ser cuidadosamente escolhido, pois seu sistema será melhor detalhada para determinados valores.

Sistemas de muitos corpos (partículas) tornam a solução do sistema quase impossível, por essa razão o método utiliza de algoritmos para obter esta solução. O algoritmo que vem sendo muito utilizado é o *Velocity-Verlet*, pois é necessário que as posições da partículas variem suavemente com o tempo, devido as equações potencias serem caracterizadas como funções contínuas. Portanto, o algoritmo *Velocity-Verlet*, usa, nesta ordem, as seguintes operações:

$$r(t + \delta t) = r(t) + \dot{r}(t)\delta t + \frac{1}{2}F(t)\delta t^2 \quad (\text{A.4})$$

$$\dot{r}(t + \delta t) = \dot{r}(t) + \frac{1}{2}\{F(t + \delta t) + F(t)\}\delta t \quad (\text{A.5})$$

Observa-se que a atualização da velocidade ocorre logo após a atualização das coordenadas, isto porque a força usada para atualizar a velocidade é a média dos valores no início e no fim do intervalo, portanto, usar $r(t + \delta t)$ para calcular $F(t + \delta t)$ e, a seguir, $\dot{r}(t + \delta t)$. Ao final deste processo, inicia-se o ciclo novamente e com o final de simulação é obtido a trajetória do sistema, bem como as coordenadas de todas as partículas.

Table geral de Miyazawa e Jernigan

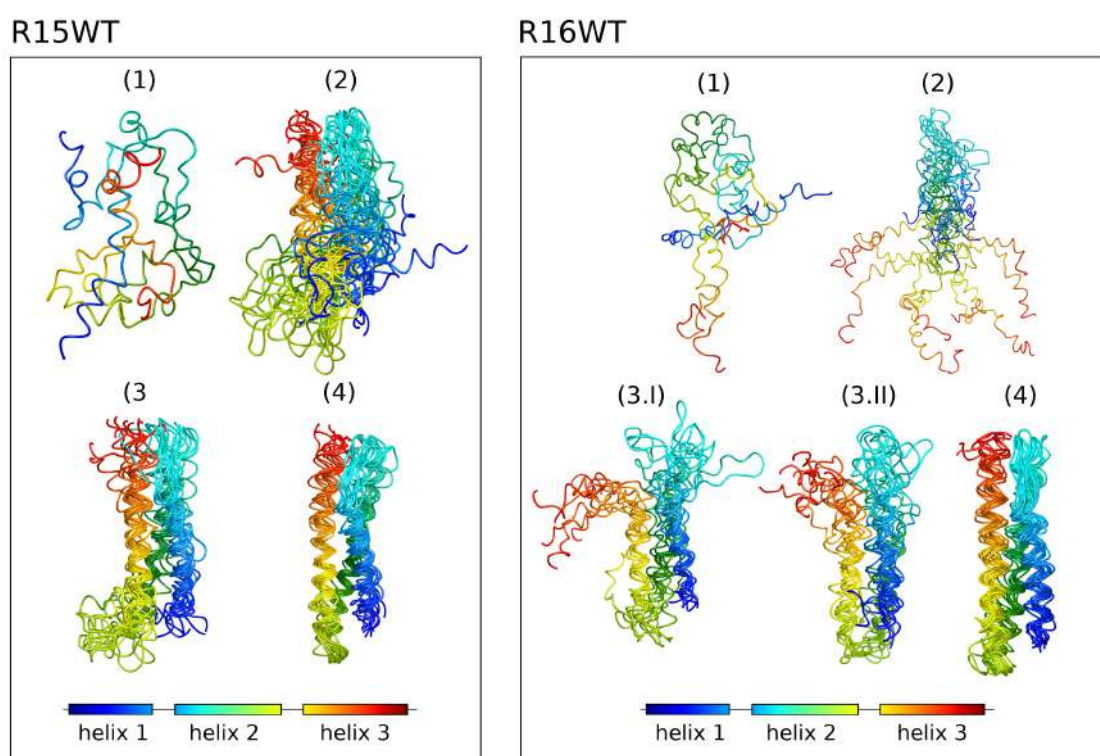
ALA	-0.12																					
GLY	-0.08	-0.29																				
ILE	-0.37	-0.13	-0.74																			
LEU	-0.38	-0.16	-0.81	-0.84																		
MET	-0.27	-0.17	-0.66	-0.70	-0.70																	
PHE	-0.36	-0.19	-0.73	-0.80	-0.83	-0.88																
PRO	0.15	0.02	-0.05	-0.12	-0.13	-0.19	0.11															
TRP	-0.27	-0.25	-0.60	-0.62	-0.73	-0.68	-0.37	-0.64														
VAL	-0.32	-0.15	-0.67	-0.74	-0.51	-0.67	-0.05	-0.51	-0.65													
ASN	0.15	-0.01	0.14	0.04	0.04	-0.01	0.18	-0.10	0.12	-0.06												
CYS	-0.33	-0.31	-0.64	-0.65	-0.61	-0.67	-0.18	-0.66	-0.59	-0.01	-1.19											
GLN	0.22	0.13	-0.01	-0.04	-0.06	-0.11	0.17	-0.02	0.08	0.06	-0.07	0.20										
SER	0.10	-0.01	0.03	-0.02	0.05	-0.12	0.20	-0.01	0.04	0.09	-0.13	0.22	0.05									
THR	0.04	-0.04	-0.15	-0.15	-0.11	-0.15	0.13	-0.02	-0.07	0.04	-0.15	0.12	0.04	0.03								
TYR	-0.20	-0.22	-0.49	-0.55	-0.56	-0.58	-0.25	-0.49	-0.38	-0.11	-0.39	-0.14	-0.08	-0.09	-0.45							
ARG	0.24	0.09	0.00	-0.04	0.03	-0.05	0.17	-0.21	0.08	0.10	0.08	0.09	0.16	0.11	-0.25	0.19						
HIS	0.07	0.00	-0.13	-0.18	-0.29	-0.34	0.01	-0.37	-0.06	0.00	-0.36	0.15	0.04	-0.03	-0.30	0.05	-0.40					
LYS	0.41	0.29	0.24	0.22	0.29	-0.19	0.47	0.09	0.29	0.22	0.33	0.28	0.36	0.33	-0.05	0.66	0.38	0.76				
ASP	0.27	0.11	0.22	0.27	0.30	0.18	0.33	0.07	0.36	0.02	0.12	0.24	0.010	0.11	-0.07	-0.24	-0.10	-0.01	0.29			
GLU	0.38	0.32	0.17	0.17	0.12	0.14	0.37	-0.00	0.26	0.12	0.20	0.27	0.18	0.16	-0.08	-0.22	0.00	-0.06	0.44	0.46		
	ALA	GLY	ILE	LEU	MET	PHE	PRO	TRP	VAL	ASN	CYS	GLN	SER	THR	TYR	ASP	GLU	ARG	HIS	LYS		

Fonte: Dados extraídos da Referência [45].

Apêndice C

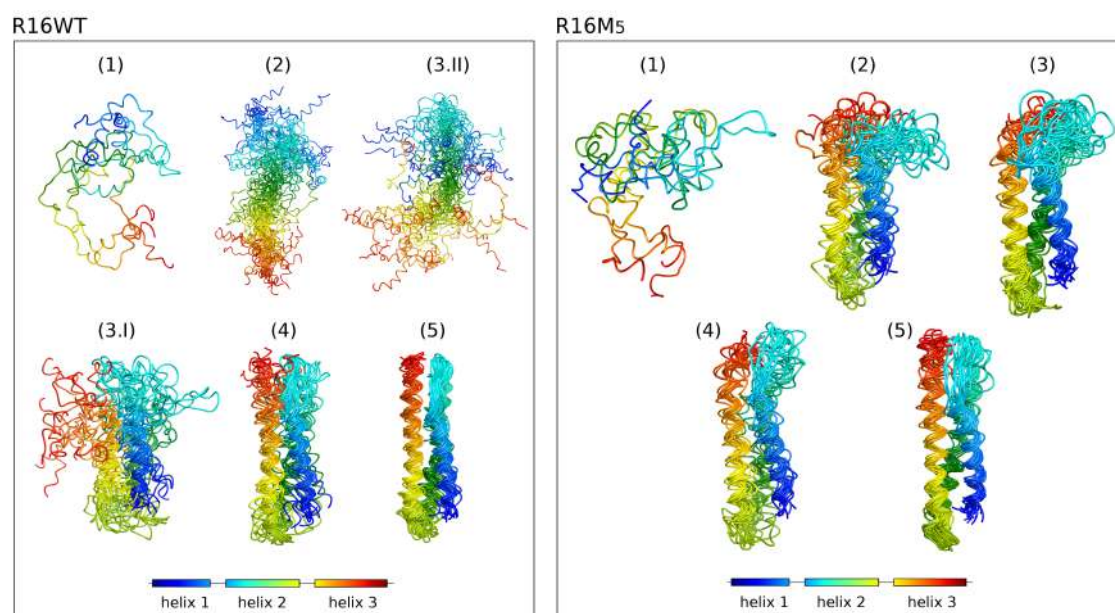
Representação das estruturas obtidas pelo ELViM

Figura C.1: Grupo de estruturas correspondente aos domínios R15WT e R16WT obtidas pelo ELViM (vide Figuras 5.2B e 5.2C).



Fonte: Elaborada pelo autor.

Figura C.2: Grupo de estruturas correspondente aos domínios R16WT e R16M₅ obtidas pelo ELViM (vide Figuras 5.3B e 5.3C)



Fonte: Elaborada pelo autor.

Apêndice D

Produção Acadêmica

Neste apêndice encontram-se listados os trabalhos realizados e os que estão em andamento durante o Doutorado do autor deste trabalho.

Artigos Publicados:

- 1 NGO, K., **SILVA, F.B.**, Leite, V.B.P., Contessoto, V.G., Onuchic J.N. **Improving the Thermostability of Xylanase A from *Bacillus Subtilis* by Combining Bioinformatics and Electrostatic Interactions Optimization.** The Journal of Physical Chemistry B. 2021.
- 2 Oliveira, V.M.; Caetano, D.L.Z.; **SILVA, F.B.**; Mouro, P.R., Antonio, A.B., Carvalho, S.J.; Leite, V.B.P. **pH and Charged Mutations Modulate Cold Shock Protein Folding and Stability: A Constant pH Monte Carlo Stud.** Journal of Chemical Theory and Computation, 16 (1), 765-772, 2020.
- 3 **SILVA, F.B.**, Oliveira, V.M., Contessoto, V.G., Leite, V.B.P. **Rational Design of Chymotrypsin Inhibitor 2 by Optimizing Non-Native Interaction.** Journal of Chemical Information and Modeling, 2019.
- 4 **SILVA, F.B.**, Contessoto, V.G., Oliveira, V.M., Clarke, J., Leite, V.B.P. **Non-Native Cooperative Interactions Modulate Protein Folding Rates.** The Journal of Physical Chemistry B, 122 (48), 10817-10824, 2018.

-
- 5 Oliveira, V.M.; Contessoto, V.G.; **SILVA, F.B.**; Caetano, D.L.Z.; Carvalho, S.J.; Leite, V.B.P. **Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models.** Biophysical Journal 114, 65-75, 2018.

Em Fase de Formatação:

- 1 **Fernando B. da Silva**, Vinícius M. de Oliveira, Vinícius G. Contessoto, Antonio B. de Oliveira, Vitor B. P. Leite. **Using the ELViM technique to identify how specific mutations speed up the folding rate and modulate the metastable state of R16 from α -spectrin.**
- 2 Vinícius M. de Oliveira, Marieli M. G. Dias, Thayná M. Avelino, Natália B. Videira¹, **Fernando B. da Silva**, Tabata R. Doratioto, Paul C. Whitford, Vitor B. P. Leite, and Ana Carolina M. Figueira. **The Activation Process of the Estrogen Receptor α : Combined Effects of pH and the Breast Cancer Recurrent Mutation D538G.**
- 3 Vinícius M. de Oliveira, **Fernando B. da Silva**, Natália B. Videira, Marieli M. G. Dias, Vitor B. P. Leite, and Ana Carolina M. Figueira. **Exploring the Molecular Pathways of the Activation Process on Recurrent Cancer Mutants of PPAR γ .**

Improving the Thermostability of Xylanase A from *Bacillus subtilis* by Combining Bioinformatics and Electrostatic Interactions Optimization

Published as part of *The Journal of Physical Chemistry virtual special issue "Dave Thirumalai Festschrift"*.

Khoa Ngo, Fernando Bruno da Silva, Vitor B. P. Leite, Vinícius G. Contessoto,* and José N. Onuchic*



Cite This: <https://doi.org/10.1021/acs.jpcb.1c01253>



Read Online

ACCESS |



Metrics & More

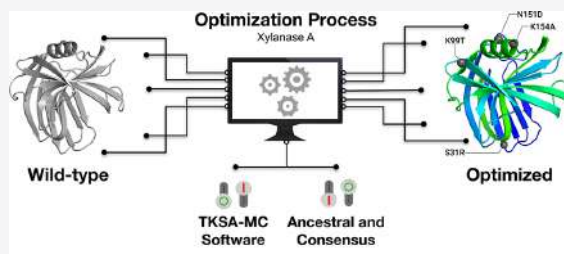


Article Recommendations



Supporting Information

ABSTRACT: The rational improvement of the enzyme catalytic activity is one of the most significant challenges in biotechnology. Most conventional strategies used to engineer enzymes involve selecting mutations to increase their thermostability. Determining good criteria for choosing these substitutions continues to be a challenge. In this work, we combine bioinformatics, electrostatic analysis, and molecular dynamics to predict beneficial mutations that may improve the thermostability of XynA from *Bacillus subtilis*. First, the Tanford–Kirkwood surface accessibility method is used to characterize each ionizable residue contribution to the protein native state stability. Residues identified to be destabilizing were mutated with the corresponding residues determined by the consensus or ancestral sequences at the same locations. Five mutants (K99T/N151D, K99T, S31R, N151D, and K154A) were investigated and compared with 12 control mutants derived from experimental approaches from the literature. Molecular dynamics results show that the mutants exhibited folding temperatures in the order K99T > K99T/N151D > S31R > N151D > WT > K154A. The combined approaches employed provide an effective strategy for low-cost enzyme optimization needed for large-scale biotechnological and medical applications.



1. INTRODUCTION

Advances in protein engineering contribute to a vast and growing number of biotechnology applications. These applications include DNA manipulation, disease treatment, natural and pharmaceutical compound synthesis, and many others. In particular, the bioenergy industry has adopted enzymes for the hydrolysis of lignocellulosic biomass to produce soluble sugars that undergo a fermentation process for bioethanol production.¹ Improving the thermostability of these enzymes would allow them to work longer under the conditions prescribed by the biomass degradation process. Moreover, the ability to use high temperatures for such procedures might not only reduce microbial contamination but also improve reaction rates.² There are several strategies for improving the thermostability of enzymes. Directed evolution, for example, mimics the natural evolution cycle under specific laboratory conditions to guide the proteins toward adapting to the new environments through natural selection.³ In rational design, the prediction of mutations is based on known information about the enzyme sequence and/or structure. By using such information, computational analyses are performed to suggest good candidates to be used in synthetic work. These computational techniques, combined with wet-labo-

ratory approaches, accelerate the protein design and make the protein suitable for different applications.⁴

Despite recent advances in protein engineering, it remains a challenging process to discover beneficial mutations. In this work, we explore different techniques to improve the thermostability of XynA from *Bacillus subtilis*. Xylanases are enzymes that catalyze the hydrolysis of the β -1,4 glycosidic linkages of xylans, present in many plant cell walls.⁵ XynA was chosen for our study due to the availability of experimental data in the literature regarding its thermostability and mutant variants; therefore, it is a great test system for our theoretical design approach. We propose to perform rational residue substitution by combining several existing theoretical tools: electrostatic analysis, ancestral sequence reconstruction strategy, and consensus sequence strategy. First, the Tanford–Kirkwood surface accessibility–Monte Carlo (TKSA–MC)⁶ method is used to identify candidate residues in the wild-type

Received: February 10, 2021

Revised: April 10, 2021



ACS Publications

© XXXX American Chemical Society

A

<https://doi.org/10.1021/acs.jpcb.1c01253>
J. Phys. Chem. B XXXX, XXX, XXX–XXX

pH and Charged Mutations Modulate Cold Shock Protein Folding and Stability: A Constant pH Monte Carlo Study

Vinícius M. de Oliveira,^{†,||} Daniel L. Z. Caetano,^{‡,||} Fernando B. da Silva,^{‡,||} Paulo R. Mouro,[‡] Antonio B. de Oliveira, Jr.,[‡] Sidney J. de Carvalho,[‡] and Vitor B. P. Leite^{*,‡,§}

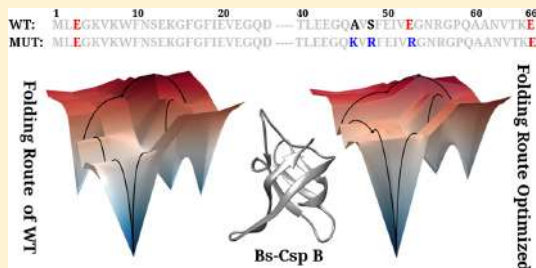
[†]Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, LNBio/CNPEM, Campinas, São Paulo, 13083-970, Brazil

[‡]Department of Physics, São Paulo State University (UNESP), Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo, 15054-000, Brazil

[§]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: The folding and stability of proteins is a fundamental problem in several research fields. In the present paper, we have used different computational approaches to study the effects caused by changes in pH and for charged mutations in cold shock proteins from *Bacillus subtilis* (Bs-CspB). First, we have investigated the contribution of each ionizable residue for these proteins to their thermal stability using the TKSA-MC, a Web server for rational mutation via optimizing the protein charge interactions. Based on these results, we have proposed a new mutation in an already optimized Bs-CspB variant. We have evaluated the effects of this new mutation in the folding energy landscape using structure-based models in Monte Carlo simulation at constant pH, SBM-CpHMC. Our results using this approach have indicated that the charge rearrangements already in the unfolded state are critical to the thermal stability of Bs-CspB. Furthermore, the conjunction of these simplified methods was able not only to predict stabilizing mutations in different pHs but also to provide essential information about their effects in each stage of protein folding.



1. INTRODUCTION

The protein folding process, although theoretically characterized by the presence of an infinity of possible partially folded structures, can exhibit some conformations which have a higher probability of being formed during the transition stages, known as folding nuclei.^{1–3} The preference for certain structures shows that the folding of a protein can be governed by the sequence of amino acid residues⁴ and by the environment in which that protein is folded.^{5–9} Local and nonlocal interactions between amino acid residues can optimize the folding of a protein, with high folding rate, and/or giving the native structure greater thermostability.^{10–12} Both the amino acid residue sequence and the environment are responsible for assigning variations in the protein charge distribution, and this variation can be considered one of the causes in the change of the folding process mechanism.⁹

The protein charge distribution depends on its conformational state and the protonation states of the amino acid residues that compose it. Variations in ionic strength and the pH of the environment also influence the charge distribution.^{13–18} According to the charge distribution, the electrostatic interactions between amino acid residues are responsible for influencing both the stability of the native protein and its

function.^{19–24} Given the importance of charged residues, such interactions have been massively studied in the folding process, in both theory and in vitro experiments.^{15,25,26} The search for more thermostable enzymes with potential application in the industry has attracted interest for the folding process of variant proteins. Through point mutations, it is sought to change the charge distribution and to optimize protein stability without affecting its function.^{27,28} Moreover, stability changes upon mutations are crucial features for genomic diseases such as cancers.^{29–31} Therefore, the understanding of this process can provide useful information for medicine and the development of new cancer therapies.

Computational simulations that take into account the energy-landscape theory have been successful in explaining several aspects regarding the protein folding process.^{3,17,32–35} Although electrostatic interactions are directly related to how proteins fold, there are few computational works that study the folding process in different pH conditions, allowing a protonation/deprotonation process during the folding.^{13,17} In this work, we discuss the effects caused by pH variations in the

Received: September 6, 2019

Published: November 22, 2019

Rational Design of Chymotrypsin Inhibitor 2 by Optimizing Non-Native Interactions

Fernando B. da Silva,[†] Vinícius M. de Oliveira,[‡] Murilo N. Sanches,[†] Vinícius G. Contessoto,[§] and Vitor B. P. Leite^{*,†}

[†]Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (UNESP), São José do Rio Preto, São Paulo 15054-000, Brazil

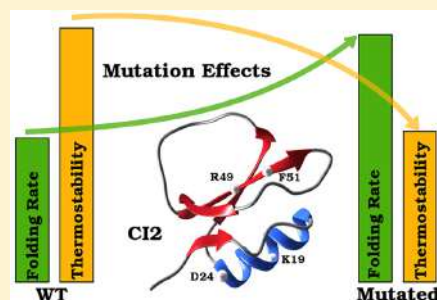
[‡]Brazilian Biosciences National Laboratory, Brazilian Center for Research in Energy and Materials, LNBio/CNPEN, Campinas, São Paulo 13083-970, Brazil

[§]Brazilian Biorenewables National Laboratory - LNBR, Brazilian Center for Research in Energy and Materials - CNPEM, Campinas, São Paulo 13083-100, Brazil

^{*}Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: Rational design of proteins via mutagenesis is crucial for several biotechnological applications. A significant challenge of the computational strategies used to predict optimized mutations is to understand the influence of each amino acid during the folding process. In the present work, chymotrypsin inhibitor 2 (CI2) and several of its designed mutants have been simulated using a non-native hydrophobic and electrostatic potential as a structure-based $C\alpha$ model. Through these simulations, we could identify the most critical folding stage to accelerate CI2 and also the charged residues responsible for providing its thermostability. The replacement of ionizable residues for hydrophobic ones tended to promote the formation of the CI2 secondary structure in the early transition state, which speeds up folding. However, this same replacement destabilized the native structure, and there was a decrease in the protein thermostability. Such a simple method proved to be capable of providing valuable information about thermodynamics and kinetics of CI2 and its mutations, thus being a fast alternative to the study of rational protein design.



INTRODUCTION

The improvement of protein stability through rational mutations is a significant scientific challenge, and protein engineering has a wide range of applications in medical and industrial areas.^{1–4} The development of new protein-based drugs and the use of enzymes with improved thermostability to perform catalysis in better conditions are some examples of these applications.^{3–7} Such relevance in scientific and industrial scenario motivated the development of rational mutations predictors that range from force-field based molecular simulations to algorithms based on machine learning (ML).^{8–14} Although there are advances in these predictor tools, issues such as data overfitting in these predictions are still recurrent, highlighting the fact that some mutation effects remain unclear.^{15,16} In such a scenario, understanding the influence of each amino acid in protein folding and stability is fundamental to developing protein engineering.

Recent works by our research group have explored the addition of non-native potentials in structure-based models (SBM), which can be a powerful method to rationalize thermodynamic and kinetic effects of specific mutations in different proteins.^{17–20} Using these simple models, we were

able to explore the entire folding process and quantify the importance of each residue in this process using less computing power than other molecular dynamics approaches. Furthermore, even when coarse-grained models were employed, we could achieve good agreement with experimental results regarding the thermostability and folding speed.^{17–20}

In our previous work, we have studied how non-native interactions modulate the protein folding rates,²⁰ in which hydrophobic and electrostatic potentials were added in the SBM. This kind of approach is widely studied by scientific groups to investigate different biological systems.^{21–24} Therefore, proteins with mutations that might induce or not an outstanding performance, in terms of thermodynamics and kinetics parameters, become an attractive target for our studies. In this case, more precisely, chymotrypsin inhibitor 2 (CI2) was chosen because this protein is one of the most studied proteins. Moreover, there are many experimental results about

Special Issue: Molecular Simulation in Latin America: Coming of Age

Received: October 1, 2019

Published: December 3, 2019



Non-Native Cooperative Interactions Modulate Protein Folding Rates

Fernando Bruno da Silva,[†] Vinícius G. Contessoto,[‡] Vinícius M. de Oliveira,[†] Jane Clarke,[¶] and Vitor B. P. Leite^{*,†}

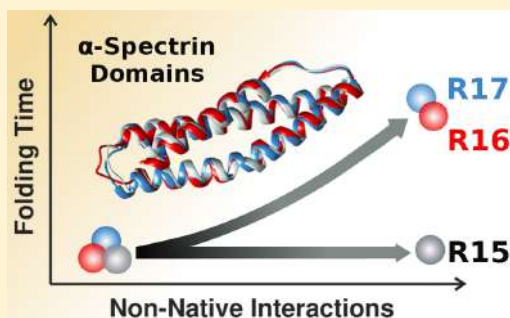
[†]Department of Physics, Institute of Biosciences, Humanities and Exact Sciences, São Paulo State University (UNESP), São José do Rio Preto - São Paulo 15054-000, Brazil

[‡]Brazilian Bioethanol Science and Technology Laboratory - CTBE, Campinas - São Paulo 13083-100, Brazil

[¶]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: The energy landscape theory and the funnel description have had remarkable success in describing protein folding mechanisms and function. However, there are experimental results that are not understood using this approach. Among the puzzling examples are the α -spectrin results, in which the R15 domain folds 3 orders of magnitude more rapidly than the homologous R16 and R17, even though they are structurally very similar to each other. Such anomalous observations are usually attributed to the influence of internal friction on protein folding rates, but this is not a satisfactory explanation. In this study, this phenomenon is addressed by focusing on non-native interactions that could account for this effect. We carried out molecular dynamics simulations with structure-based C_α models, in which the folding process of α -spectrin domains was investigated. The simulations take into account the hydrophobic and electrostatic contributions separately. The folding time results have shown qualitative agreement with the experimental data. We have also investigated mutations in R16 and R17, and the simulation folding time results correlate with the observed experimental ones. We suggest that the origin of the internal friction, at least in this case, might emerge from a cooperativity effect of these non-native interactions.



■ INTRODUCTION

The energy landscape theory has been successful for addressing protein folding and function. Through a statistical mechanics approach, the theory provides a framework for understanding many aspects of protein folding, such as order–disorder phase transitions, thermodynamic stability, and kinetics.^{1–4} Models based on the landscape ideas have been useful to characterize the folding process in quantitative detail, and have successfully bridged theory and experiment.^{5–8} Since the native structure of a protein encode much of its information regarding folding mechanism and function, structure-based models (SBM) describe overall protein features, such as free-energy barriers, ϕ -values, and folding times, which correlate well with experimental results.⁹ However, there are exceptions to the prediction of these models that challenge our understanding of the problem, and they usually serve as a probe to advance our knowledge. One of the puzzling problems occurs with the three specific domains from α -spectrin, in which they were found to display significant differences in their folding rates.^{10–16} The R15, R16, and R17 domains of α -spectrin are composed of an elongated three-helix bundle and display similar structures, thermodynamic

stabilities, and β -Tanford values.^{11,17–19} Clarke and co-workers reported that R15 folds and unfolds 3 orders of magnitude faster than its homologues R16 and R17.¹¹ This result is counterintuitive with the notion that similar structures would yield comparable folding times. A fraction of this difference in folding rate could be ascribed directly to differences in internal friction. Early computational efforts show a good agreement of the Φ -values with experiments,²⁰ but the reason for the differences in folding rates remains unclear.

The concept of internal friction in the folding processes has been extensively studied by experimental and computational groups.^{20–26} Some experimental measurements have identified a deviation in the expected relationship between the folding rate and solvent viscosity.^{21,27} The suggestion is that the folding process is also influenced by internal collisions within the protein, which could explain the dependence of the folding rates on low viscosity.²¹ The internal friction generated by collisions within the protein may be interpreted as an energy

Received: September 13, 2018

Revised: November 8, 2018

Published: November 8, 2018

Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models

Vinícius Martins de Oliveira,¹ Vinícius de Godoi Contessoto,^{1,2} Fernando Bruno da Silva,¹ Daniel Lucas Zago Caetano,¹ Sidney Jurado de Carvalho,¹ and Vitor Barbanti Pereira Leite^{1,*}

¹São Paulo State University (UNESP), Institute of Biosciences, Humanities and Exact Sciences (Ibilce), Campus São José do Rio Preto, Brazil and ²Brazilian Bioethanol Science and Technology Laboratory- (CTBE), Campinas, Brazil

ABSTRACT The importance of charge-charge interactions in the thermal stability of proteins is widely known. pH and ionic strength play a crucial role in these electrostatic interactions, as well as in the arrangement of ionizable residues in each protein-folding stage. In this study, two coarse-grained models were used to evaluate the effect of pH and salt concentration on the thermal stability of a protein G variant (1PGB-QDD), which was chosen due to the quantity of experimental data exploring these effects on its stability. One of these coarse-grained models, the TKSA, calculates the electrostatic free energy of the protein in the native state via the Tanford-Kirkwood approach for each residue. The other one, CpHMD-SBM, uses a Coulomb screening potential in addition to the structure-based model C_{α} . Both models simulate the system in constant pH. The comparison between the experimental stability analysis and the computational results obtained by these simple models showed a good agreement. Through the TKSA method, the role of each charged residue in the protein's thermal stability was inferred. Using CpHMD-SBM, it was possible to evaluate salt and pH effects throughout the folding process. Finally, the computational pK_a values were calculated by both methods and presented a good level of agreement with the experiments. This study provides, to our knowledge, new information and a comprehensive description of the electrostatic contribution to protein G stability.

INTRODUCTION

A knowledge of the factors that affect the thermal stability of proteins is of fundamental importance in understanding the basic principles that govern the behavior of these macromolecules as well as in the rational development of biotechnological applications (1–3). It is widely known that a variation in pH and salt concentration can affect the thermal stability of proteins (4–6), indicating the importance of electrostatic interactions. Indeed, studies have shown the thermal stability of proteins can be enhanced by optimizing charge-charge interactions (7–9). Establishing a relationship between the distribution of ionizable residues and the features of folding and thermal stability can therefore provide valuable information for protein engineering.

Several levels of approximation have been proposed for protein models and for evaluating the electrostatic contributions of unfolding free energy (10–13). In several of these

studies, versions of structure-based models that include a fixed charge have been adopted to study the contribution of the electrostatic interaction in the unfolded state and the interaction between proteins and charged macromolecules (7,14–18). Recently, to study pH effects on the folding dynamics of the N-terminal domain of ribosomal protein L9 (19), the constant-pH molecular dynamics (MD) method (CpHMD) (19–21) was implemented and the results showed a good level of agreement with the experiments. Another alternative to evaluate the charge-charge contribution to protein stability is the Tanford-Kirkwood model (TK) (22), which enables the individual electrostatic interaction of each ionizable residue to be calculated (23). Makhatazde and colleagues (24–27) have managed to optimize proteins by the mutation of specific residues, which are predicted by the TK approach with a solvent accessibility modification, the Tanford-Kirkwood Solvent Accessibility (TKSA) method. In this study, both methods, CpHMD and TKSA, are used to study the importance of pH and salt concentration in 1PGB-QDD stability.

1PGB-QDD is a variant of the B1 domain of protein G with mutations T2Q, N8D, and N37D (Fig. 1 A). Protein G is a multidomain protein present in the cell wall of *G. streptococcus* containing the immunoglobulin G binding

Submitted September 1, 2017, and accepted for publication November 13, 2017.

*Correspondence: vleite@sjrp.unesp.br

Vinícius Martins de Oliveira and Vinícius de Godoi Contessoto contributed equally to this work.

Editor: Nathan Baker.

<https://doi.org/10.1016/j.bpj.2017.11.012>

© 2017

