



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

Fernando Bruno da Silva

Compreendendo a Cinética de Enovelamento
da α -Espectrina por meio
de Interações Não Nativas

São José do Rio Preto
2017

Fernando Bruno da Silva

**Compreendendo a Cinética de Enovelamento
da α -Espectrina por meio
de Interações Não Nativas**

Dissertação apresentada, como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular do Instituto de Biociências, Letras e Ciências Exatas “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Financiadora: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

São José do Rio Preto, 22 de Março de 2017.

Silva, Fernando Bruno da.

Compreendendo a cinética de enovelamento da α -Espectrina por meio de interações não nativas / Fernando Bruno da Silva. -- São José do Rio Preto, 2017

74, [27] f. : il., tabs.

Orientador: Vitor Barbanti Pereira Leite

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Biofísica. 3. Proteínas – Estrutura.
4. Enovelamento de proteínas. 5. Dinâmica molecular. 6. Viscosidade.
I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. II. Título.

CDU – 577.112

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Fernando Bruno da Silva

**Compreendendo a Cinética de Enovelamento
da α -Espectrina por meio
de Interações Não Nativas**

Dissertação apresentada, como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular do Instituto de Biociências, Letras e Ciências Exatas “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

Financiadora: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Comissão Examinadora

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

Prof. Dr. Antonio Francisco Pereira de Araujo
UNB – Brasília – DF

Prof. Dr. Leandro Cristanti de Oliveira
UNESP – São José do Rio Preto – SP

São José do Rio Preto, 22 de Março de 2017.

Dedico este trabalho a toda minha família, principalmente a minha mãe Marta, à minha irmã Aline e ao meu irmão Diego, que sempre me apoiaram e me incentivaram nesta minha jornada.

Agradecimentos

Quero agradecer primeiramente a toda minha família, por acreditar em meu potencial e principalmente por me apoiar. Agradeço exclusivamente minha mãe Marta Bruno Prudêncio, a meu pai Claudemir Ap. da Silva 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 (Chapado), Vinícius Contessoto, Vinícius de Oliveira (Góias), Antonio Bento (Tiburcio) pela ajuda nos momentos de dificuldade.

Aos amigos Denis Bruno, Ingrid Bernardes, Carolina Tatiani, Willian Derek, Ives Oliveira, Janine Contro, Alex Miller, Taisa Viegas, Nayara Alcântara, Daniel Zago, Alexandre (Jesus), Gabriela de Araujo, Paulo Henrique, Ricardo, Monique e as pessoas que aqui não foram mencionadas, agradeço pela amizade e o companheirismo.

Agradeço ao Prof. Dr. Antonio Francisco de Araújo e ao Prof. Dr. Leandro Cristante de Oliveira por aceitarem participar desta banca de defesa de mestrado e também aos professores suplentes Prof. Dr. Alexandre Suman de Araújo e ao Prof. Dr. Luis Gustavo Dias.

Aos professores do Departamento de Física por toda ajuda prestada, ao programa de pós-graduação em Biofísica Molecular pelo acolhimento e à UNESP pela infra-estrutura fornecida.

Por fim e não menos importante, agradeço a CAPES pelo apoio financeiro.

“Só aqueles que desistiram de viver acham que os sonhos são impossíveis.”

Hyoga de Cisne

Resumo

O objetivo deste trabalho foi caracterizar o atrito interno que são evidenciados nos domínios R15, R16 e R17 da *chicken brain* α -Espectrina. A Espectrina é uma proteína que faz parte do citoesqueleto e foi descoberta primeiramente em eritrócitos, sendo importante para manter a estabilidade, estrutura e forma da membrana celular. Esta proteína é composta de duas subunidades, α e β , sendo que cada um dos domínios presentes nestas subunidades apresentam 106 aminoácidos. Além disso, estes domínios apresentam alta similaridade estrutural ($\text{RMSD} < 1 \text{ \AA}$) e os resultados experimentais indicam que o domínio R15 apresenta um tempo de enovelamento que é três ordens de grandeza menor que seus homólogos (R16 e R17), esta observação foi atribuída aos efeitos do atrito interno na taxa de enovelamento. Portanto, neste trabalho foi estudado o enovelamento do R15, R16 e R17 utilizando o modelo baseado em estrutura C_α . Em nossas simulações, além das interações homogêneas para os contatos nativos (conhecido como modelo *vanilla*), foi adicionado um potencial atrativo para as interações não nativas hidrofóbicas (conhecido como modelo *flavored*) utilizando a tabela do potencial de interação Miyazawa-Jernigan, bem como simulações com carga fixa e a pH constante, onde foram analisadas as interações eletrostáticas no processo de enovelamento. Por fim, realizaram-se simulações com o acoplamento para melhor descrever o sistema estudado.

Palavras-chave: Modelo Baseado em Estrutura. Enovelamento de Proteína. Dinâmica Molecular. Método de Dinâmica Molecular a pH Constante. Atrito Interno. Interações Hidrofóbicas. Interações Eletrostáticas.

Abstract

The main aim of this work is characterized the internal friction evident in the three chicken brain α – spectrin domains, known as R15, R16, and R17. Spectrin is a cytoskeletal protein that was first discovered in erythrocytes and it is important for maintaining the stability, structure, and shape of the cell membrane. The protein is composed of a modular structure of α and β subunits, which typically contain 106 residues amino acid sequence motifs called "spectrin repeats". They are very similar in terms of structure and stability (RMSD < 1 Å). However, experimental results indicate that R15 folds and unfolds 3 orders of magnitude faster than its homologs. Such baffling observation is usually attributed to the influence of "internal friction" on protein folding kinetics. However, this effect cannot be satisfactorily corroborated by computer simulation. In the present work, the folding of these three domains is addressed using C_α structure-based models. We carried out simulations using homogeneous interactions on native contacts (known as a vanilla model) and has been added an attractive potential in the non-native hydrophobic interactions with Miyazawa-Jernigan potential (known as flavored model), as well as simulations with fixed charge and at constant pH, where the electrostatic interactions were analyzed in the folding. Finally, simulations were performed with the coupling of the potential (flavored and fixed load) to better describe the system.

Keywords: Structure-Based Models. Protein Folding. Molecular Dynamics. Constant pH Molecular Dynamic. Internal Friction. Hydrophobic Interactions, Eletrostatic Interactions.

Lista de Figuras

1.1	Representação simplificada de uma ligação peptídica entre dois aminoácidos diferentes.	15
1.2	Diagrama esquemático do relevo da superfície de energia	16
1.3	Modelo representativo da Espectrina (cadeias α e β) e dos domínios R15, R16 e R17 da <i>chicken brain</i> α -Espectrina	19
1.4	Representação do alinhamento estrutural dos domínios R15, R16 e R17.	20
1.5	Alinhamento das sequências dos domínios R15, R16 e R17 e representação estrutural da R16 destacando os aminoácidos da primeira hélice que serão mutados.	21
1.6	Taxa de relaxamento na mudança conformacional da mioglobina em função da viscosidade do solvente	24
5.1	Calor específico C_v em função da temperatura reduzida T^* para as simulações realizadas com o modelo <i>vanilla</i>	38
5.2	Energia livre em função da fração dos contatos nativos para o modelo <i>vanilla</i>	40
5.3	Rota de enovelamento em função da coordenada de reação, Q , e razão do tempo de enovelamento, τ , para os domínios R15, R16 e R17 calculados pelo modelo <i>vanilla</i>	41
5.4	Calor específico, C_v , em função da temperatura reduzida, T^* , e energia livre em função da coordenada de reação, Q , para as simulações realizadas pelo modelo <i>flavored</i>	44
5.5	Rota de enovelamento em função da coordenada de reação, Q , e razão do tempo de enovelamento, τ , para os domínios R15, R16 e R17 calculados pelo modelo <i>flavored</i>	45

5.6	Energia livre em função da fração dos contatos nativos, Q , para o modelo com carga fixa	46
5.7	Rota de enovelamento em função da fração dos contatos nativos, Q , e o tempo de enovelamento, τ , para os domínios R16 e R17 calculados no modelo com carga fixa	47
5.8	Calor específico, C_v , em função da temperatura reduzida, T^* , e energia livre em função da coordenada de reação, Q , para as simulações realizadas pelo modelo <i>flavored</i> + carga fixa	49
5.9	Comparação entre os tempos de enovelamento de todos os modelos de simulações adotados para os domínios R15, R16 e R17.	50
5.10	Calor específico C_v em função da temperatura reduzida T^* para as simulações realizadas com as mutações.	52
5.11	Energia livre em função da fração dos contatos nativos, Q , para as mutações com o modelo <i>flavored</i> + carga fixa	52
5.12	Comparação entre os valores do tempo enovelamento experimental e computacional para os domínios R16 e R17 na sua forma regular e mutada.	53
5.13	Cálculo da rota de enovelamento, $R(Q)$, em função da fração dos contatos nativos, Q , para os domínios R16 e R17 em sua forma regular e com as mutações.	55
D.1	Comparação dos calores específicos e dos perfis de energia livre para as simulações realizadas no GROMACS e ESPREesSo.	72
D.2	Trajetória de enovelamento e estados de protonação da histidina em função dos passos de simulação para o domínio R16.	74

Lista de Tabelas

3.1	Energia de contato, $\Delta\epsilon_{ij}$, entre os aminoácidos hidrofóbicos	31
3.2	Valores de pK_a para os aminoácidos que possuem grupos carregados. .	32
5.1	Informações relacionadas ao número de aminoácidos, número de contatos nativos e similaridade estrutural(RMSD) e sequencial(FASTA e BLAST) dos domínios estudados.	37
5.2	Informações a respeito dos domínios estudados para diferentes valores de K_{HP}	42
B.1	Energia de contato entre os aminoácidos em unidades RT	68

Lista de Símbolos

Q	Fração dos contatos nativos.
pH	potencial Hidrogeniônico.
η_s	Viscosidade do solvente.
T	Temperatura do sistema.
$k(T, \eta_s)$	Taxa de relaxamento.
k^{-1}	Tempo de relaxamento.
k_B	Constante de Boltzmann.
ΔE_a	Variação da energia de ativação.
σ	Viscosidade interna ou atrito interno.
A	Termo independente da viscosidade do solvente e temperatura.
Γ	Estrutura qualquer da proteína.
Γ_o	Estrutura nativa da proteína.
$V(\Gamma, \Gamma_o)$	Energia potencial.
E_T	Energia potencial total.
$E_{G\bar{o}}$	Energia potencial do modelo baseado em estrutura C_α .
E_{HP}	Energia potencial do modelo flavored.
E_{elec}	Energia potencial eletrostático.
$\Delta\epsilon_{ij}$	Energia de contato dos aminoácidos.
ΔE	Variação da energia eletrostática decorrente a mudança de estado.
$R(Q)$	Rota de enovelamento em função da coordenada de reação Q .
τ	Tempo de primeira passagem.
T_F	Temperatura de transição para o estado enovelado.
$F(Q)$	Energia livre em função da coordenada de reação Q .
$\Delta F(Q)$	Variação da energia livre em função da coordenada de reação Q .
C_v	Calor específico.

Sumário

1	Introdução	14
1.1	α – Espectrina	19
1.2	Atrito Interno	23
2	Objetivos	26
3	Metodologia	28
3.1	Modelo Baseado em Estrutura - <i>Vanilla</i>	28
3.2	Interações Hidrofóbicas - <i>Flavored</i>	30
3.3	Dinâmica Molecular a pH Constante - CpHMD	31
3.3.1	Critério de Metropolis para a Protonação ou Desprotonação	33
4	Métodos de Análise	34
4.1	Cálculo da Rota de Enovelamento	34
4.2	Propriedades Termodinâmicas	35
4.3	Detalhes das Simulações	35
5	Resultados e Discussão	37
5.1	Modelo <i>Vanilla</i>	38
5.2	Modelo <i>Flavored</i>	41
5.3	Modelo com Carga Fixa	45
5.4	Modelo <i>Flavored</i> + Carga Fixa	47
5.4.1	Análise das Mutações	51

6	Conclusão	56
	Referências Bibliográficas	58
A	Dinâmica Molecular	66
B	Potencial Miyazawa e Jernigan	68
C	Método dos Múltiplos Histogramas - WHAM	69
D	Resultados Complementares	71
E	Artigos: 1PGB-DQQ e α -Espectrina	75

Capítulo 1

Introdução

As proteínas são biopolímeros constituídos por pequenas moléculas de aminoácidos, sendo de grande importância para a vida, estando presente em todos os seres vivos. Além disso, 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 e elaboradora de estruturas celulares, entre outros [1].

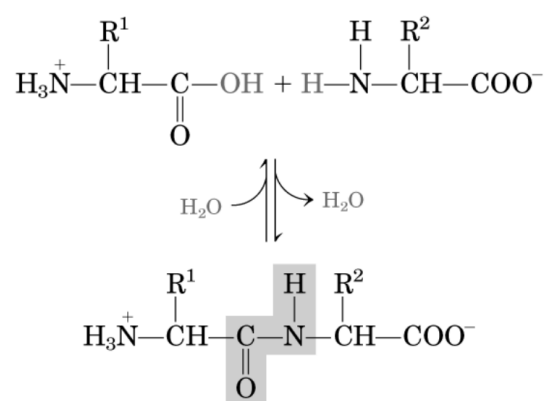
O que diferencia uma proteína da outra são as possíveis combinações que podem ser formadas devido a grande variedade de aminoácidos, como também sua estrutura e função. Existem ao todo vinte tipos de aminoácidos essenciais onde, geralmente, as proteínas não possuem todos em sua cadeia. Os aminoácidos 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) um com um radical ($-\text{R}$), que distingue um aminoácido do outro 1.1 [1, 2].

Sabe-se que a sequência de nucleotídeos do DNA é a responsável por dirigir a síntese proteica. Por consequência disso, alterações no material genético (mutações) podem produzir proteínas com funções diferenciadas e, ademais, as funções das proteínas estão intrinsecamente relacionadas com suas estruturas. Portanto, as etapas para que uma sequência linear de aminoácidos alcance sua estrutura tridimensional são descritas em quatro níveis de organização: estrutura primária, secundária, terciária e quaternária [1, 2].

A estrutura primária consiste na sequência linear dos aminoácidos da cadeia polipeptídica que é formada a partir de uma reação de condensação, também denominada ligação peptídica entre dois aminoácidos (conforme descrito na Figura 1.1). Na estrutura secundária têm-se o surgimento das conformações locais da proteína, como as

hélices α e as folhas β . A estrutura terciária descreve o dobramento dos elementos estruturais secundários e especifica o arranjo tridimensional de todos os átomos que compõe a proteína, caracterizando, dessa forma, a configuração global da proteína. Geralmente as proteínas que apresentam massa molecular acima de 100 kDa são constituídas por mais de uma cadeia polipeptídica, sendo que essas subunidades associam-se em grupos específicos, onde seu arranjo espacial tridimensional é conhecido como estrutura quaternária [1, 2].

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, H_2O . A ligação CO-NH resultante é denominada ligação peptídica.



Fonte: Elaborada pelo autor.

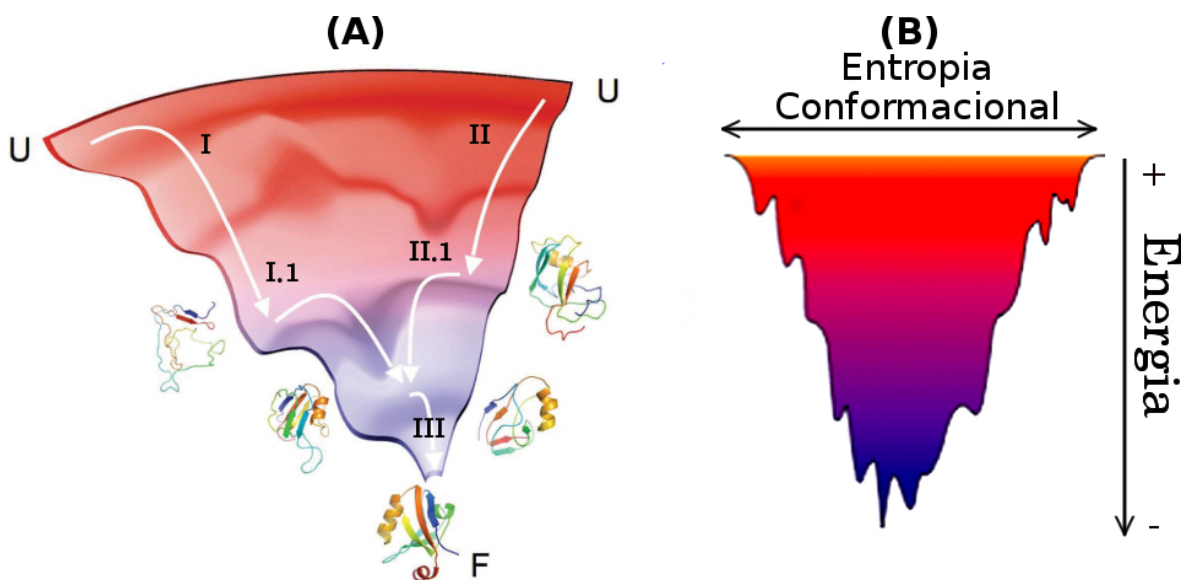
Compreender o mecanismo biomolecular que leva uma proteína do seu estágio inicial (estrutura primária) ao seu estágio final (estrutura terciária ou quaternária), tem sido um enorme desafio durante décadas. Ao que se pode observar, esta é uma área da ciência que vem incentivando e gerando discussões entre 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 [1, 2].

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 [3, 4].

Em meados de 1961, Anfinsen et al. [5], por meio de estudos experimentais, apresentou resultados de grande impacto científico, indicando que, ao se conhecer a sequência de aminoácidos, é possível definir o estado de mínima energia global da proteína, sendo este o seu estado nativo funcional. Atualmente, este resultado é conhecido na literatura como Hipótese Termodinâmica. Nesta hipótese, os resultados evidenciaram que toda informação necessária para o processo correto do enovelamento está contida na sequência primária de aminoácidos da proteína [5, 6].

Anos mais tarde, em 1968, Levintal [7] levantou uma questão a respeito do tempo necessário para que uma proteína efetue seu enovelamento. Considerando todos os microestados¹ que uma proteína pode visitar, calculou-se o tempo necessário para que ela atinja uma estrutura que apresente o estado de mínima energia, de acordo com sua sequência de aminoácidos [7, 8]. Portanto, para uma proteína com 100 aminoácidos e apenas duas conformações para cada ligação peptídica, o tempo necessário para que ela se enovele seria da ordem de 10^{18} segundos, tempo compatível com a idade do universo.

Figura 1.2: Diagrama esquemático do relevo da superfície de energia. Na Figura A, tem-se a representação multidimensional da superfície de energia, onde I e II representam as diferentes vias de enovelamento; I.1 e II.1 representam os estados intermediários e; III representa o estado nativo da proteína. Na Figura B, tem-se a representação do funil de energia para uma proteína qualquer.



Fonte: Figura extraída de Brunori et al.[9] e adaptada pelo autor.

Este resultado é conhecido na literatura como *Paradoxo de Levinthal*. O próprio Levinthal conclui que é necessário conhecer rotas específicas de enovelamento,

¹Diferentes maneiras que uma cadeia polipeptídica pode arranjar-se no espaço.

nas quais, a proteína não visitará todas as suas possibilidades conformacionais. Por exemplo, uma proteína com 100 resíduos de aminoácidos, cada resíduo experimentando 3 posições, 3^{100} , onde o tempo para testar cada posição corresponde 10^{-13} seg, logo o tempo total para a proteína realizar todos os movimentos seria igual, $3^{100} \times 10^{-13}$ seg $= 10^{87}$ seg, o que corresponde a $1,6 \times 10^{27}$ anos. Portanto, o tempo necessário para uma proteína de 100 aminoácidos enovelar-se 3 de todas as possibilidades conformacionais possíveis, ou seja randomicamente, seria maior do que o tempo do universo, $4,3 \times 10^{17}$ seg, ou seja, 13,7 bilhões de anos, sendo que em *vitro*, o enovelamento ocorre em menos de um segundo. Contudo, se conhecendo as rotas específicas, ocorrerá uma redução no tempo devido ao fato de que, agora, o processo de enovelamento não seria mais realizado por uma busca aleatória, mas sim por um processo guiado [7]. Sendo assim, somente após a introdução de uma nova teoria foi possível resolver o paradoxo de Levinthal. Esta teoria é conhecida como a teoria de superfície de energia² [10–12].

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 (ver Figuras 1.2A e 1.2B). 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. 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 [13–17]. Por meio de análises estatísticas, essa teoria também tem fornecido uma forte base teórica para o entendimento de fenômenos de oligomerização [18] e transições funcionais em proteínas [19–22].

Outro fator adotado neste trabalho foi a escolha da coordenada de reação. A coordenada utilizada está difundida na literatura, sendo conhecida como Q. Esta coordenada representa a fração dos contatos formados pela estrutura nativa da proteína. Durante o processo de enovelamento esta coordenada mede a fração da similaridade entre a estrutura desenovelada e a nativa e, uma vez que a proteína encontra-se em seu estado desenovelado, a fração dos contatos nativos apresentará valores Q próximos de zero ($Q \approx 0$). Já para o caso em que a proteína encontra-se praticamente enovelada, Q apresentará valores próximos da unidade igual a 1 ($Q \approx 1$). Logo, a faixa de valores que Q pode assumir varia de 0 até 1. Deste modo, a escolha da coordenada Q representa de forma aceitável o processo de enovelamento [23].

²Do inglês *Energy Landscape Theory*.

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 no qual a superfície de energia é irregular, há uma grande competição entre as interações locais e não locais. Esta competição é chamada de frustração [13, 24, 25].

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 também podem ser energeticamente favoráveis [26], 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 [27, 28]. Portanto, o entendimento da presença ou da ausência destas interações é fundamental para compreender os seus efeitos sobre o processo de enovelamento [29].

Um exemplo que pode ser citado é o caso da Im7 e Im9, duas proteínas que apresentam estruturas praticamente idênticas, porém mecanismo de enovelamento distintos. No caso da Im9, seu enovelamento é caracterizado por apresentar dois estados: o estado desenovelado e o enovelado [30]. Entretanto, a Im7 é uma proteína de três estados: estado desenovelado, intermediário e enovelado [31]. Segundo Chen e Chan [32], aminoácidos hidrofóbicos identificados por resultados experimentais podem estar envolvidos em interações não nativas durante o enovelamento da Im7. Portanto, suas observações sustentaram a perspectiva experimental de que o estado intermediário da Im7 é amplamente apoiado por interações não nativas envolvendo aminoácidos hidrofóbicos. Além disso, suas simulações sugeriram que os efeitos não nativos na Im7 são facilitados por uma menor densidade local de contatos nativos quando comparada com as da Im9 [32].

Tendo em vista a importância dos contatos citados anteriormente para o enovelamento da Im7, foi estudada, neste trabalho, a influência das interações não nativas para os domínios R15, R16 e R17 da α -Espectrina (seção 1.1) durante seu processo de enovelamento. Tal fato baseia-se nas diferenças significativas nas taxas de enovelamento observadas para essas proteínas. Além disso, outras interações poderiam estar envolvidas nesse processo, como, por exemplo, interações eletrostáticas.

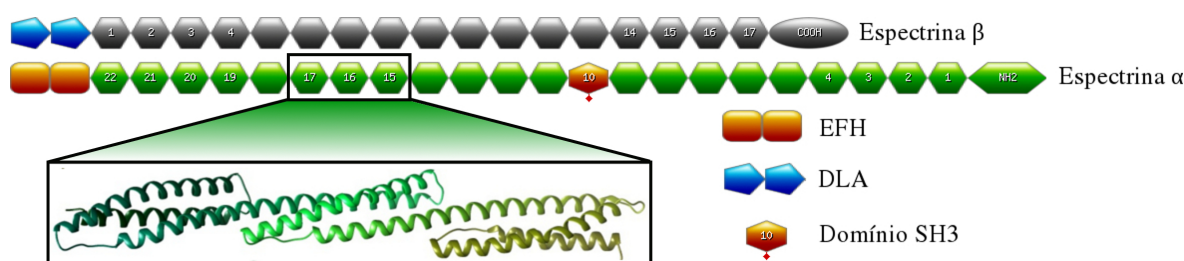
As interações eletrostáticas durante o processo de enovelamento de proteínas têm sido exaustivamente investigadas por pesquisadores experimentais e teóricos [33–36]. Os aminoácidos carregados também podem desempenhar um papel importante na estabilidade e função das proteínas, especialmente quando estão relacionados ao

reconhecimento do sítio de ligação [37–39]. A distribuição da carga em uma proteína é definida pelo tipo de aminoácido presente em sua estrutura e pelas condições do ambiente. Alguns aspectos que podem alterar o estado de protonação de um aminoácido ionizável são: a área da superfície acessível ao solventes; a distância entre esses grupos carregados; a força iônica e as variações de pH, que são responsáveis pelas modificações nas características do ambiente [35]. De todos esses aspectos, a variação no pH da solução é geralmente a que mais afeta o grau de ionização dos aminoácidos.

1.1 α – Espectrina

A Espectrina é uma proteína que está associada ao citoesqueleto e foi descoberta inicialmente em eritrócitos. Ela desempenha um importante papel na manutenção da estabilidade, do arcabouço e da conformação da membrana celular, conferindo-a flexibilidade[40]. A Espectrina foi isolada pela primeira vez quando estava associada à membrana de glóbulos vermelhos e, devido a estas células estarem desprovidas de hemoglobina, passaram a ser denominadas de eritrócito - "fantasma" [41, 42]. Por muito tempo, a existência das Espectrinas foi considerada restrita somente aos eritrócitos, porém é possível encontrá-las em outros tipos celulares [43, 44].

Figura 1.3: Modelo representativo da Espectrina (cadeias α e β) e dos domínios R15, R16 e R17 da *chicken brain* α -Espectrina (PDB: 1U4Q) na sua estrutura tridimensional. DLA é o domínio de ligação da actina e EFH é o domínio de ligação do Ca^{+2} . A subunidade α da Espectrina apresenta um domínio SH3, que está relacionado aos processos de adesão celular.



Fonte: Elaborada pelo autor.

A Espectrina é composta por duas cadeias polipeptídicas semelhantes, a subunidade α , de 280 kDa e a subunidade β , de 246 kDa. As subunidades α e β apresentam, respectivamente, 22 e 17 domínios, conforme ilustrado na Figura 1.3. Os domínios de cada subunidade apresentam, aproximadamente, 106 aminoácidos. Esses grandes polipeptídeos encontram-se entrelaçados fracamente para formar um dímero com 1.000 Å de comprimento. Quando dois desses dímeros se unem é formado um

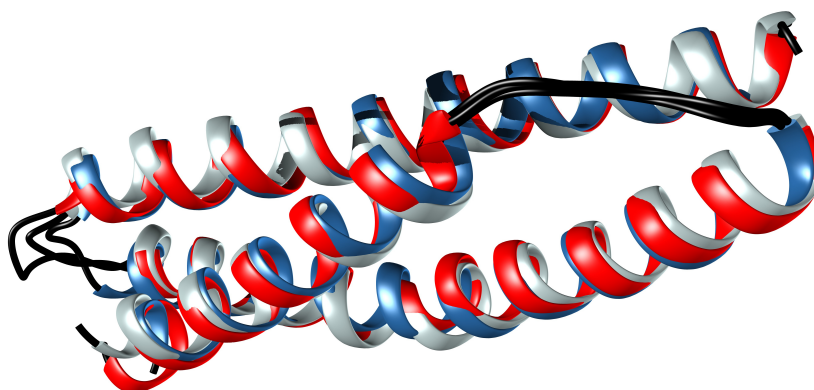
heterodímero [1].

Em uma célula existem cerca de 100.000 heterodímeros de espectrina, os quais estão interligados em ambas as extremidades por conexões formadas a partir de outras proteínas do citoesqueleto, tais como a actina, a ancurina e a anquirina. O conjunto dessas proteínas forma uma rede irregular que está disposta por toda a membrana eritrocítica.

A função da Espectrina está diretamente relacionada com sua estrutura. Apesar de apresentarem estruturas muito semelhantes, Espectrinas de diferentes tipos celulares possuem sítios funcionais distintos. Essas diferenças fazem com que Espectrinas de eritrócitos sejam mais flexíveis, já as Espectrinas de outros tipos celulares tendem a conferir uma maior estabilidade [1].

Neste trabalho, foi analisada somente a α -Espectrina, que é composta apenas por estruturas helicoidais (Figura 1.3). Cada domínio da α -Espectrina pode ser estudado de maneira independente. Logo, neste estudo, os domínios que são analisados são identificados como R15, R16 e R17. Embora os domínios possuam alta similaridade estrutural, com $\text{RMSD} < 1 \text{ \AA}$ (Figura 1.4), o mecanismo cinético e o processo de enovelamento são distintos [45]. Mais notavelmente, a taxa de enovelamento do R15 é de três ordens de grandeza maior do que o de seus homólogos, sendo o domínio R17 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 [46, 47].

Figura 1.4: Representação do alinhamento estrutural dos domínios R15 (vermelho), R16 (cinza) e R17 (azul). Os domínios possuem aproximadamente 106 resíduos de aminoácidos e apresentam um $\text{RMSD} < 1 \text{ \AA}$.



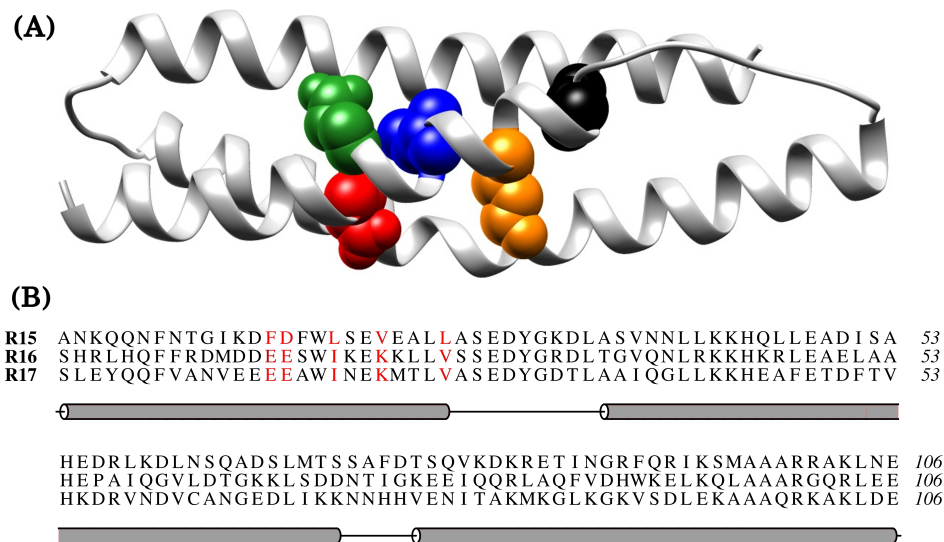
Fonte: Elaborada pelo autor.

Muito se discute acerca dos efeitos do atrito interno durante o processo de

enovelamento. Em 2012, Jane Clarke e Benjamin Schuler, realizaram análises experimentais com os domínios mencionados no parágrafo anterior. De acordo com suas averiguações, pôde-se identificar os mesmos efeitos que Ansari et al.[48] observaram para a mioglobina. Identificaram também uma região em que o atrito interno afeta de forma direta esses domínios. Nesta exata região ocorre uma desaceleração na dinâmica, estando o atrito interno localizado pouco antes do estado de transição [46].

Em 2012, Wensley et al.[49] realizaram mutações em todos os aminoácidos dos domínios R16 e R17 e, por meio de duas análises, valores de ϕ e curvas de *chevron plot*, concluíram que apenas cinco das mutações realizadas na primeira estrutura helicoidal exibiram resultados significativos. Portanto, constatou-se que a substituição dos aminoácidos desta hélice pelos aminoácidos correspondentes do domínio R15 (Figura 1.5), alterou de forma significativa as taxas de enovelamento e desenovelamento da R16 e R17.

Figura 1.5: Alinhamento das sequências dos domínios R15, R16 e R17 e a representação estrutural da R16 destacando os aminoácidos da primeira hélice que serão mutados. Na Figura A, representação estrutural da R16 destacando os aminoácidos, 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, hélices α , da estrutura secundária.



Fonte: Elaborada pelo autor.

Para as mutações E15D, I18L e V25L realizadas na R16, nenhum resultado significativo foi obtido mediante as análises citadas anteriormente. Nota-se que apenas foram realizadas mutações entre aminoácidos que compartilham das mesmas propriedades: aminoácidos carregados negativamente, E15D e aminoácidos hidrofóbicos, I18L e V25L. Já para as mutações E14F e K21V, onde aminoácidos que apresentam carga

negativa e positiva são substituídos por aminoácidos hidrofóbicos, um aumento na taxa de enovelamento e desenovelamento foi verificado [49].

De forma simultânea, foram realizadas cinco mutações (E14F, E15D, I18L, K21V e V25L) nos domínios R16 e R17. Estas mutações não só aumentaram, de forma direta, 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 essas mutações, apresentaram o mesmo mecanismo de enovelamento do domínio R15. Resultado comprovado ao se comparar o perfil dos valores de ϕ desses domínio com os valores do domínio R15.

De acordo com o parágrafo anterior, em sua forma regular, os domínios R15, R16 e R17 apresentam padrões distintos em seus mecanismo de enovelamento [49]. O domínio R15 apresenta um padrão conhecido como *nucleation condensation model*, que consiste na formação das estruturas secundárias e terciárias ao mesmo tempo. O núcleo hidrofóbico da proteína colapsa de forma aleatória, permitindo que os elementos da estrutura secundária se formem e que a proteína obtenha a sua conformação nativa [50, 51]. Já os domínios R16 e R17 apresentam o mecanismo conhecido como *framework-like mechanism*. Tal mecanismo consiste em uma montagem hierárquica da estrutura onde elementos estruturais secundários são guiados por contatos locais, que são formados independentemente da estrutura terciária. A fusão dos blocos estruturais secundários preexistentes resulta na formação de uma estrutura globular intermediária, sendo que, por fim, um ajuste da estrutura intermediária é realizado. Como resultado, tem-se a estrutura nativa [52, 53].

Afim de averiguar os mesmos efeitos mutacionais encontrados na R16 e R17, realizaram mutações no domínio R15 com o objetivo de verificar quais mutações alterariam tanto o mecanismo como a taxa de enovelamento. Portanto, por meio das curvas de *chevron plot* [54], foi observado que maioria das mutações elevaram a taxa de enovelamento. Entretanto, para a mutação V21K, que substitui um aminoácido hidrofóbico por um de carga positiva, verificou-se uma diminuição na taxa de enovelamento quando comparada com a do domínio R15 em sua forma regular. Além disso, observaram-se mudanças anômalas no enovelamento para esta mutação. Tais alterações podem estar relacionadas tanto com o aumento na rugosidade do perfil de energia como com a mudança do mecanismo de enovelamento. Afim de corrigir tais anomalias, foi realizada uma segunda mutação, D15E, que foi analisada anteriormente de forma individual e que não apresentou mudanças significativas nos valores de ϕ e nas taxas, porém foi capaz de estabilizar o domínio R15 contendo a mutação V21K [54].

Por conseguinte, foram constatados os mesmos valores de ϕ para a primeira

e para a terceira hélice ao se comparar a R15 regular com a R15 que sofreu duas mutações, D15E e V21K. Tais resultados indicam que as duas proteínas apresentam o mesmo mecanismo de enovelamento. Além disso, esses mesmos valores comprovam que o estado de transição não sofreu nenhuma alteração significativa. Logo, conclui-se que a R15 é um domínio que não apresente susceptibilidade às mutações e, por fim, no que diz respeito aos domínios R16 e R17, chega-se à conclusão que as variações nos valores de ϕ , nas taxas e na alteração do mecanismo de enovelamento, são decorrentes da perda da cadeia lateral carregada positivamente ou negativamente [49, 54].

É também de fundamental importância a realização de uma análise computacional, devido à deficiência e às lacunas que uma análise experimental pode apresentar. Computacionalmente é possível realizar uma varredura minuciosa das origens do efeito do atrito interno. Robert B. Best identificou, em seu trabalho, a origem molecular do efeito do atrito interno sobre a taxa de enovelamento para os domínios R15, R16 e R17. Tal origem corresponde à torção estrutural dos elementos da proteína, o que ocasiona uma busca frustrada para o encaixe correto das hélices parcialmente pré-formadas [47]. Além disso, assim como apontado por Best, Scott, Randles e Clarke [47, 55], a formação das ligações de hidrogênio que estabilizam as α - hélices dos domínios R15, R16 e R17, torna o processo de enovelamento desfavorável até certo ponto, elucidando, por fim, que o atrito interno é mais efetivo em proteínas pequenas e ricas em α - hélice.

1.2 Atrito Interno

Os efeitos da viscosidade do solvente na dinâmica do enovelamento e desenovelamento foram estudados por meio de métodos experimentais e por simulações computacionais [47, 48, 56, 57]. Em 1992, Ansari e colaboradores foram os pioneiros no estudo do relaxamento conformacional da mioglobina [48], analisando o efeito da viscosidade do solvente, η_s , sobre a taxa de relaxamento, k . Suas medições mostraram que apesar da taxa de relaxamento ser inversamente proporcional à viscosidade (teoria de Kramer, $k \propto 1/\eta_s$ [58]) no regime de alta viscosidade, o mesmo não se verifica para o regime de baixa viscosidade, $\eta_s \rightarrow 0$ (Figura 1.6). Ainda assim, os autores sugeriram uma expressão em que a taxa de relaxamento da mioglobina pode ser descrita:

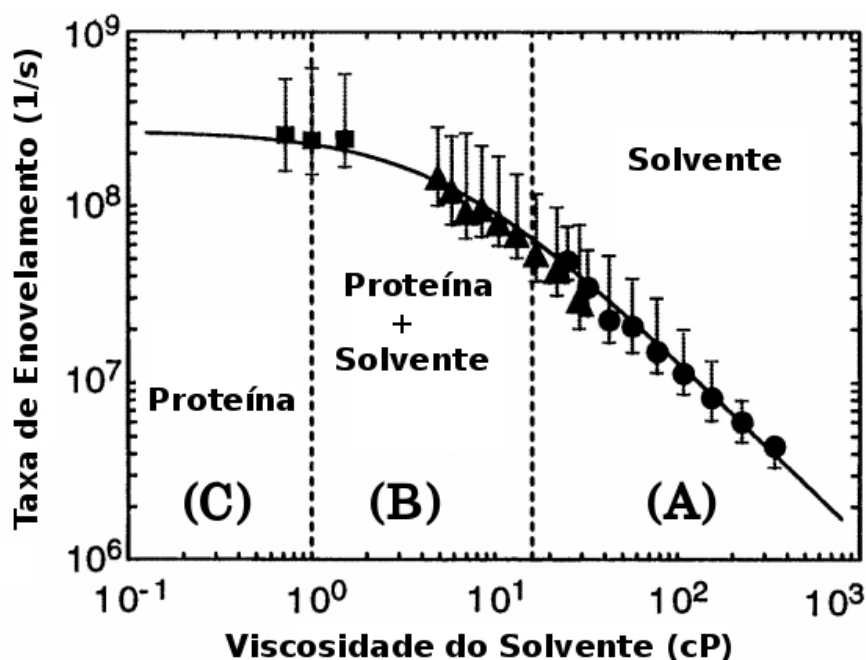
$$k(T, \eta_s) = \frac{A}{\sigma + \eta_s} \exp\left(\frac{-\Delta E_a}{k_B T}\right) \quad (1.1)$$

onde k_B é a constante de Boltzmann; T é a temperatura; η_s e σ são as contribuições totais para o atrito; A é o termo independente da viscosidade e da temperatura; e ΔE_a representa a barreira de energia de ativação, ou seja, é a energia mínima que um dado

composto precisa para iniciar a reação química. Esta energia mínima é necessária para a formação do complexo ativado. Quanto maior a energia de ativação, mais lenta é a reação devido ao aumento da dificuldade para que o processo ocorra. Quanto menor é a energia de ativação, menor a barreira de energia e mais colisões efetivas ocorrem. Portanto, uma reação mais rápida é verificada.

A Figura 1.6 exibe dados ajustados pela equação 1.1 para um valor fixo de $\sigma = (4.1 \pm 1.3)$ cP e em diferentes concentrações. Verificou-se que o sistema dividiu-se em três regiões: i) em valores abaixo de 1 cP (região C), ocorreu uma pequena contribuição do atrito proveniente do solvente para diminuição na taxa de relaxamento; ii) em valores entre 1 e 15 cP (região B), duas fontes de atrito cooperam para a diminuição da taxa, sendo que uma provém da proteína e outra do solvente e; iii) em valores acima de 15 cP (região A), somente uma única fonte contribui para a diminuição da taxa, a qual corresponde ao atrito do solvente.

Figura 1.6: Taxa de relaxamento na mudança conformacional da mioglobina em função da viscosidade do solvente. (A) Regime de alta viscosidade. Para esta região, o atrito provém unicamente da viscosidade do solvente. A taxa de relaxamento e a concentração do solvente são inversamente proporcionais. (B) Regime intermediário. Nesta região ocorre uma contribuição mútua para o atrito, oriunda da viscosidade do solvente e da própria proteína. (C) Regime de extrapolação, onde a fonte do atrito provém exclusivamente da proteína. Círculos e triângulos correspondem a 79% e 56% de glicerol em água, respectivamente, à uma faixa de temperatura que varia de -5°C à 35°C , com um intervalo de 5°C , enquanto que quadrados representam a água a uma temperatura de 5° , 20° e 35°C .



Fonte: Figura extraída de Ansari et al.[48] e adaptada pelo autor.

A variável σ pode ser interpretada como a viscosidade interna da proteína, também denominada de atrito interno. Este termo, em conjunto com a viscosidade do solvente, η_s , limita fortemente a dinâmica do enovelamento [59–61]. Em 2012, Soranno et. al. [62] quantificaram o atrito interno, independente do solvente, em proteínas desenoveladas e intrinsecamente desordenadas, através da técnica experimental *single-molecule spectroscopy*. No caso dos domínios R15, R16 e R17, que são as proteínas alvo deste trabalho, Wensley et. al. [49] quantificaram, através da técnica experimental citada anteriormente, o atrito interno de cada domínio. Seus resultados exibiram um valor mais elevado de σ para o domínio R17, um valor intermediário para o domínio R16 e um valor inferior comparado aos demais para o domínio R15. O atrito interno pode ser interpretado como uma alteração conformacional que é caracterizada por uma troca de energia e pelas colisões dos átomos que estão envolvidos no processo de enovelamento [63]. O atrito encontrado em proteínas provém de duas fontes: a primeira delas está associada ao movimento das moléculas do solvente, enquanto que a segunda fonte se deve à reorganização no interior da proteína [48].

Capítulo 2

Objetivos

Os efeitos do atrito interno sobre os domínios R15, R16 e R17 da α -Espectrina já estão estabelecidos na literatura, como descrito na seção 1.2. Portanto, sua verdadeira origem é um fato desconhecido pela comunidade científica. Logo, este trabalho foi instigado pelas complicações geradas pelo atrito interno no processo de enovelamento desses domínios.

Portanto, este trabalho tem como objetivo analisar, com base nos diferentes modelos empregados, a origem intrínseca do atrito interno nos domínios R15, R16 e R17 da α -Espectrina. Os modelos de simulações adotados foram:

Modelo *Vanilla* - As simulações baseadas neste modelo têm como objetivo verificar se apenas interações de caráter homogêneo podem caracterizar os efeitos do atrito interno a partir das análises apresentadas na seção 4. Este modelo está descrito, em detalhes, na seção 3.1.

Modelo *Flavored* - Com a possibilidade de distinguir as interações entre os aminoácidos, as simulações estudadas neste modelo irão verificar se a influência dos contatos não nativos hidrofóbicos contribuem de forma positiva ou negativa para o enovelamento das proteínas estudadas. A descrição do modelo *flavored* encontra-se na seção 3.2.

Modelo CpHMD e Carga Fixa - A protonação/desprotonação dos aminoácidos ao decorrer da simulação pode descrever, de forma mais realística, o enovelamento da proteína. Por alternativo, serão realizadas simulações pelo *software* ESPResSO, versão 3.2.0, onde os aminoácidos podem variar sua carga durante a simulação. Também serão realizadas simulações pelo *software* GROMACS, versão 4.5.5, em que a carga dos aminoácidos serão fixadas e não sofrerão nenhuma alteração. Tais

simulações serão realizadas com o intuito de verificar de que forma a inserção de carga pode influenciar o enovelamento da proteína. Os detalhes deste modelo estão transcritos na seção 3.3.

Modelo *Flavored* + Carga Fixa - Estes dois potenciais serão acoplados ao potencial do modelo *vanilla*, com o objetivo de averiguar os balanços energéticos dentre esses modelos e a sua influência para o enovelamento.

Capítulo 3

Metodologia

Neste capítulo serão apresentadas as informações sobre os modelos adotados neste trabalho. Na seção 3.1, encontra-se a descrição do modelo baseado em estrutura (SBM)¹, sendo que, neste trabalho, empregou-se o modelo C_α . A partir da seção 3.2, o modelo utilizado distingue as interações individuais entre os aminoácidos, diferenciando-o de uma simulação "convencional", onde essas individualidades não são consideradas. Dessa forma, as interações entre os aminoácidos baseiam-se na tabela do potencial de Miyazawa e Jernigan (Tabela B). Um modo de aumentar a complexidade do sistema, visando torná-lo mais realístico, é adotar simulações a pH contante. Por este motivo, adicionou-se carga aos aminoácidos, de acordo com a classificação de sua cadeia lateral (seção 3.3).

3.1 Modelo Baseado em Estrutura - *Vanilla*

A representação de um modelo que aborda de forma realística todas as interações atuantes em um sistema proteico é, de fato, quase impossível, pois o alto grau de complexidade desses sistemas resultaria em simulações com um alto custo computacional. Levando-se isso em consideração, utilizam-se modelos minimalistas, ou seja, modelos que não englobam todas as interações. Tais modelos representam de forma aceitável alguns fenômenos e, ainda assim, são capazes de elucidar fundamentos básicos que governam o processo de enovelamento. Dentre todos os modelos minimalistas, o modelo C_α foi escolhido para este trabalho devido ao seu baixo custo computacional.

Simulações computacionais de dinâmica molecular (ver Apêndice A) para o modelo C_α levam em conta a realização de algumas modificações, tais como: todos

¹Do inglês *Structure-based Model (SBM)*

os resíduos de um aminoácido são substituídos por uma esfera centrada na posição do carbono alfa correspondente e o potencial de interação entre os carbonos alfa é modificado, pois este modelo considera apenas os termos de ligações e ângulos para manter a geometria da molécula. Os contatos não locais são incluídos via potencial Lennard-Jones 12-10 ao invés de 12-6, usualmente empregado no do modelo com todos os átomos².

Os modelos de potenciais que se baseiam na estrutura nativa da proteína são conhecidos como modelos $G\bar{o}$ [64]. Estes modelos são implementados a partir das informações extraídas de um banco de dados de proteínas³ [65], em que são depositados os valores baseados na resolução estrutural de proteínas, como cristalografia de raio-X e ressonância magnética nuclear (RMN).

A energia potencial do modelo $C\alpha$, para uma conformação qualquer Γ de uma proteína, considera como base uma estrutura de referência, ou seja, a estrutura nativa Γ_o . Assim, a energia potencial, $V(\Gamma, \Gamma_o)$, do modelo $C\alpha$ é dada por:

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

Conforme a equação 3.1, o potencial para o modelo baseado em estrutura (modelo C_α) apenas considera termos referentes às interações que atuam na estrutura nativa. O primeiro termo refere-se ao potencial harmônico que é formado pela ligação entre dois carbonos α adjacentes, onde r_o é a distância entre dois carbonos α na estrutura nativa. Já o segundo termo refere-se ao potencial harmônico angular que é formado por três carbonos α consecutivos, onde θ_o representa o ângulo formado por três resíduos i , $i + 1$ e $i + 2$ na estrutura nativa. O terceiro termo da equação 3.1 refere-se à torção realizada na cadeia, onde ϕ_o é o ângulo diedro formado por quatro carbonos α consecutivos. O valor d_{ij} do potencial 12-10 utilizado no penúltimo termo da equação 3.1 é calculado através do mapa de contatos entre os aminoácidos que pertencem à estrutura nativa. Para este cálculo utiliza-se o servidor SMOG@ctbp [66] com o algoritmo *Shadow Contact Map* (SCM) [67]. Já o último termo da equação 3.1 contabiliza a repulsão entre os contatos que não estão presentes na estrutura nativa da proteína. A aproximação máxima entre os carbonos α caracteriza o volume ocupado

²Do inglês *All-atom model*

³Do inglês *Protein Data Bank* (PDB)

por um aminoácido que é de $\sigma_{NC} = 4 \text{ \AA}$.

As constantes ϵ_r , ϵ_θ , ϵ_ϕ e ϵ_{NC} , valem 100, 20, 1 e 1, respectivamente, e são dadas em unidades de ϵ_C , que é a energia de interação entre os contatos, apresentando um valor aproximadamente igual a 1 kcal/mol [25, 68].

3.2 Interações Hidrofóbicas - *Flavored*

As interações nativas já foram tratadas como sendo as únicas favoráveis para o processo de enovelamento, já que as não nativas eram considerada apenas como neutras ou repulsivas. Entretanto, descobriu-se através de estudos experimentais e teóricos, que as interações não nativas apresentam um papel imprescindível no processo de enovelamento [32, 69–71]. Tendo por base a importância da formação desses contatos, como, por exemplo, a interação entre resíduos hidrofóbicos, torna-se inviável sua deleção no modelo adotado, ou seja, no modelo C_α . Devido a esta importância, adicionou-se ao modelo anterior um termo de hidrofobicidade (modelo também denominado como *flavored*), de acordo com a equação 3.2.

$$E_T = E_{G\bar{o}} + E_{HP} \quad (3.2)$$

Para este modelo as interações entre os aminoácidos são de caráter heterogêneo e hidrofóbico. Logo, apenas as energias dos contatos não nativos que apresentam esta determinada característica são contabilizadas. Portanto, nesse trabalho apenas os aminoácidos hidrofóbicos, tais como Ala (A), Val (V), Leu (L), Ile (I), Met (M), Trp (W) e Phe (F) [69] são considerados neste modelo.

De acordo com a equação 3.3, o termo r_{ij} representa a distância entre os aminoácidos i e j da cadeia C_α , sendo M o número total de aminoácidos e σ o volume excluído dos átomos. Entretanto, para as interações não nativas no modelo homogêneo assumi-se $\kappa_{ij} = -1$ e $K_{HP} = 1$, independentemente do tipo de resíduo [73], enquanto que para o modelo aqui apresentado, considera-se apenas aminoácidos hidrofóbicos, têm-se o valor de $\kappa_{ij} = \Delta\epsilon_{ij}$, onde $\Delta\epsilon_{ij}$ representa a energia de interação entre os contatos i e j que constam na tabela do Miyazawa e Jernigan [72], $\sigma = 5 \text{ \AA}$ e K_{HP} é um parâmetro ajustável. Dois aminoácidos são considerados como contatos não nativos hidrofóbicos se, e somente se, $|i - j| > 3$ e $r_{ij} < 8 \text{ \AA}$.

$$E_{HP} = \sum_i^M \sum_{j=i+4}^M K_{HP} \kappa_{ij} \exp[-(r_{ij} - \sigma)^2] \quad (3.3)$$

Tabela 3.1: Energia de contato, $\Delta\epsilon_{ij}$, entre os aminoácidos hidrofóbicos em unidades RT. As energias de contato em destaque fazem referência à menor e à maior energia de contato entre os aminoácidos iguais, respectivamente.

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
	ALA	GLY	ILE	LEU	MET	PHE	PRO	TRP	VAL

Fonte: Dados extraídos de Miyazawa e Jernigan [72]

No apêndice B, encontra-se a tabela de energia de contato entre todos os aminoácidos. Já na tabela 3.1, são apresentados somente as energias de contato entre os aminoácidos hidrofóbicos.

3.3 Dinâmica Molecular a pH Constante - CpHMD

Dinâmica Molecular (DM) é uma das técnicas mais utilizadas para o estudo de macromoléculas. Dentre todos os métodos existentes na literatura (Método de Monte Carlo e Dinâmica Browniana), esta técnica trata de forma mais realística os movimentos de um corpo e as causas referentes a estes movimentos. Isto se deve ao fato do método se valer de potenciais clássicos para descrever a evolução temporal de um sistema [74, 75].

Um dos fatores externos que influencia a dinâmica de uma proteína é o pH. O pH relaciona-se à atividade dos íons H^+ em solução e seu valor indica a acidez ou alcalinidade do sistema, variando de 0 a 14. Soluções que apresentam valores de pH abaixo de 7 são ácidas, ao passo que soluções que apresentam pH acima de 7 são alcalinas (básicas). Por essa razão, este termo exibe alta relevância para o ambiente que uma macromolécula se encontra [1, 2, 76].

Um modo de quantificar o grau de dissociação de um ácido é analisar os seus valores de K_a e pK_a ($pK_a = -\log_{10}K_a$). É direta a relação entre K_a e a acidez, pois quanto maior o valor de K_a , mais forte o ácido. Portanto, para determinar o grau de protonação de um aminoácido, em um dado pH, é muito importante que sejam

conhecidos seus respectivos valores de pK_a (Tabela 3.2).

Tabela 3.2: Valores de pK_a para os aminoácidos que possuem grupos carregados.

Resíduo	pK_a
Cter	3.6
ASP	4.0
GLU	4.5
HIS	6.3
Nter	7.5
LYS	10.6
ARG	12.0

Fonte: Dados extraídos de Contessoto et al.[77].

Em simulações convencionais de DM, parâmetros como protonação são estabelecidos previamente e mantidos fixos ao longo de toda dinâmica [78]. No entanto, os resíduos carregados de uma proteína podem variar seu estado de protonação. Devido não só ao valor do pH do ambiente como também à própria distância entre os grupos ionizáveis.

Em resposta a este problema, surgiram os métodos de DM a pH constante⁴ [79, 80], que permitem alterações no estado de protonação durante uma simulação de DM. Em vez de definir os estados fixos de protonação durante toda dinâmica para um dado pH, os resíduos carregados podem variar sua carga durante este processo.

Com esse método de simulação, pode-se calcular as energias eletrostáticas de acordo com a teoria de Debye-Hückel e, devido à simplicidade do modelo baseado em estrutura, as interações carga-carga são representadas por um potencial de Coulomb blindado (equação 3.4). Este potencial é adicionado à equação 3.1, onde os termos q_i e q_j referem-se aos resíduos carregados i e j respectivamente, $K_{\text{eletrostático}} = 332 \text{ kcalÅ}/(\text{mol e}^2)$, k é o inverso do comprimento de Debye e a constante dielétrica é $\epsilon_k = 80$.

$$E_{\text{elec}} = \sum_{\text{eletrostático}} K_{\text{eletrostático}} \frac{q_i q_j \exp(-kr_{ij})}{\epsilon_k r_{ij}} \quad (3.4)$$

As considerações deste modelo reduzem a complexidade do sistema. Entretanto, o método é aceito pela comunidade científica e seus resultados corroboram com os dados experimentais. Além disso, é possível obter resultados difíceis de serem medidos experimentalmente [34, 77].

⁴Do inglês *Constant-pH Molecular Dynamic Method* (CpHMD)

3.3.1 Critério de Metropolis para a Protonação ou Desprotonação

No decorrer da simulação CpHMD, em um determinado momento, realizam-se, através do critério de Metropolis, os cálculos para verificar as probabilidades de um aminoácido com característica ácida ou básica ser protonado ou desprotonado.

A partir da equação 3.5, o critério é implementado por meio da variação da energia eletrostática ΔE_{elec} , que é calculada a partir da equação 3.4, e dos valores já estabelecidos de pH e pK_a (Tabela 3.2), onde os sinais positivos e negativos são usados, respectivamente, para a protonação e desprotonação. A constante ξ pode valer +1 ou -1 para os resíduos ácidos ou básicos, na devida ordem [77, 81]. Além de que, o primeiro termo da equação representa a contribuição dos próton em solução a um determinado pH; o segundo termo refere-se aos valores de pK_a que cada aminoácido possui e; por fim, o terceiro termo representa a variação da energia eletrostática decorrente a mudança no estado de protonação de um aminoácido titulável.

$$\Delta E = \pm \xi k_B T \ln(10)(pH - pK_a) + \Delta E_{elec} \quad (3.5)$$

De acordo com o critério de Metropolis, se um passo da simulação é aceito, o aminoácidos titulável em questão irá alterar seu estado de protonação ou desprotonação e a dinâmica seguirá seu curso. Entretanto, se o novo estado não for aceito, não ocorrerá nenhuma mudança e a simulação continua com o aminoácido sem sofrer nenhuma alteração em seu estado de protonação/desprotonação.

Os resultados que encontram-se no apêndice D para o modelo CpHMD, foram extraídos a partir das simulações realizadas pelo o programa desenvolvido pelo Prof. Dr. Sidney Jurado de Carvalho e Prof. Dr. Leandro Cristante de Oliveira, no qual foram realizadas pequenas modificações pelos alunos de doutorado Vinicius Contessoto e Vinicius Martins de Oliveira, onde ambos encontram-se no mesmo instituto que o autor deste trabalho.

Capítulo 4

Métodos de Análise

4.1 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 [82]. 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 [16].

A função que descreve o cálculo da rota de enovelamento é dada pela equação 4.1

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

onde Q_i , $i = 1, \dots, M$, 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 [83].

4.2 Propriedades Termodinâmicas

Para tratar os sistemas recorre-se à Mecânica Estatística, a qual permite calcular as propriedades termodinâmicas de interesse, sendo a energia livre uma das mais importantes dessas propriedades. No entanto, calcular o perfil de energia livre dos domínios estudados por meio das simulações computacionais não é trivial. Isso se deve ao fato que não é possível obter a entropia desse sistema apenas pela média de algumas propriedades. O método do Múltiplo Histograma (*Weighted Histograms Analysis Method* - WHAM - ver Apêndice C) tem-se mostrado eficiente para sistemas com as mesmas características definidas neste projeto e um meio de se contornar a dificuldade citada anteriormente. Além disso, com esse método, pode-se maximizar as informações extraídas das simulações.

Resumidamente, esse método consiste em determinar a densidade de estados, ω , em uma dada temperatura T , com a soma ponderada de densidades de estado estimadas, ω_i , obtidas por meio de diversas simulações com temperaturas T_i diferentes de T . Os pesos, usados para o cálculo da soma ponderada, são obtidos pela minimização dos erros estatísticos $\delta^2\omega$, por meio da sobreposição das distribuições de probabilidade Ω de cada simulação executada em uma temperatura T_i .

4.3 Detalhes das Simulações

Primeiramente, foi extraído as informações estruturais das proteínas estudadas, R15, R16 e R17, no banco de dados PDB [65], onde encontram-se as informações referentes as coordenadas espaciais de cada átomo. Já com a estrutura nativa, é realizado a montagem do mapa de contato e para este cálculo utiliza-se o servidor SMOG@ctbp [66] com o algoritmo *Shadow Contact Map* [67], onde também foram adquiridos os arquivos de topologia.

As simulações foram realizadas com recursos computacionais do Shiva Cluster do departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da UNESP e do Grid UNESP, onde a dinâmica molecular foi efetuada pelo pacote GROMACS versão 4.5.5 utilizando integrador estocástico e o acoplamento térmico de Berendsen. Foi utilizado 10^9 passos de simulação, onde a cada 5000 passos foi salvo as coordenadas e as energias, totalizando 200000 *frames*, com o intervalo de tempo igual a 0.0005 ps. Para acompanhar a dinâmica foi escolhido a coordenada de reação Q , onde representa a fração dos contatos nativos e para um contato nativo ser aceito ele deve estar abaixo de $1,2d_{ij}$, sendo que d_{ij} representa a distância entre os contatos i e j da

estrutura nativa. Ao final das simulações, as grandezas termodinâmicas são calculadas, tais como, energia livre, entropia e calor específico, sendo estes calculados pelo método WHAM (*Weighted Histograms Analysis Method*).

Simulações com outros parâmetros iniciais foram realizadas com o objetivo de calcular o tempo de enovelamento. Portanto, iniciou-se as simulações em temperaturas altíssimas para desenovelar a proteína, foram necessários 25000 passos e a cada 50 salvou-se uma trajetória. Em seguida, foram realizadas simulações, que não iniciou-se das coordenadas da estrutura nativa, mas sim das coordenadas da última trajetória salva, sendo estas 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 para a simulação, sendo que a cada 2500 passos a trajetória foi salva. Para cada temperatura foram realizadas 100 simulações, onde cada simulação foi realizada em uma temperatura diferente. O cálculo do tempo de enovelamento é realizado de acordo com a equação 4.2

$$\tau = P_i \times P_{traj} \times 0,0005 \quad (4.2)$$

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 5

Resultados e Discussão

Neste capítulo encontram-se os resultados obtidos pelos métodos e pelas simulações mencionadas nas seções 3.1, 3.2 e 3.3. Inicialmente, são apresentadas algumas informações e dados importantes sobre os domínios estudados. De acordo com a Tabela 5.1, nota-se que esses domínios apresentam alta similaridade estrutural, assim como foi mencionado na seção 1.1. Já no que diz respeito à similaridade sequencial há uma pequena concordância entre as sequências. Estas análises foram realizadas por meio de dois *softwares*, FASTA e BLAST.

Tabela 5.1: Informações dos domínios R15, R16 e R17 da α -Espectrina: número de aminoácidos (#a.a), número de contatos nativos (M), similaridade estrutural (RMSD) e similaridade sequencial.

Proteína(PDB)	#a.a	M	RMSD(Å)	FASTA	BLAST
R15(1U5P)	107	232	0,96 (R15-R16)	32,0% (R15-R16)	33,0% (R15-R16)
R16(1AJ3)	107	231	0,90 (R16-R17)	34,2% (R16-R17)	35,0% (R16-R17)
R17(1U4Q)	104	221	0,83 (R15-R17)	32,7% (R15-R17)	33,0% (R15-R17)

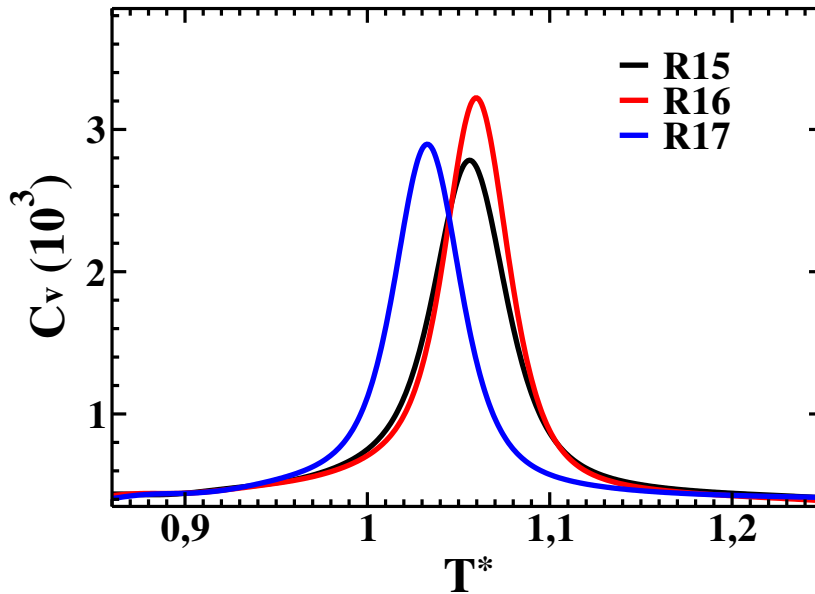
A partir dos valores de RMSD para todas as combinações realizadas, os domínios R15 e R17 apresentaram o menor valor (0,83 Å), seguido pelo RMSD de 0,90 Å para os domínios R16 e R17 e, finalmente, pelo RMSD igual a 0,96 Å referente aos domínios R15, R16 e R17. Com base nas análises de similaridade sequencial, verificou-se que os domínios apresentaram uma reduzida similaridade. Portanto, de modo geral, torna-se evidente que a informação crucial a respeito da origem do atrito interno está contida na sequência de aminoácidos que cada domínio possui.

5.1 Modelo *Vanilla*

Foram calculadas determinadas grandezas termodinâmicas, a partir dos dados gerados pelas simulações, tais como: energia livre, entropia e calor específico. Estas grandezas foram obtidas pelo método dos múltiplos histogramas (*Weighted Histograms Analysis Method* - WHAM) para as simulações realizadas com o modelo baseado em estrutura, também conhecido como modelo *vanilla* cujos aminoácidos são representados por um esfera rígida centrada na posição do carbono α . Uma descrição detalhada do método WHAM pode ser encontrada no Apêndice A.

A temperatura de transição, também denominada temperatura de enovelamento, T_F , pode ser definida como a temperatura na qual a população das configurações enoveladas é igual às populações em todas as outras configurações [84]. Já o valor do T_F é obtido a partir do pico do calor específico, C_v . Na Figura 5.1, tem-se o calor específico para cada domínio.

Figura 5.1: Calor específico, C_v , em função da temperatura reduzida, T^* , para as simulações realizadas com o modelo *vanilla*. Os domínios R15, R16 e R17 são representados pelas curvas preta, vermelha e azul, respectivamente.



Fonte: Elaborada pelo autor.

A partir da Figura 5.1, nota-se que os domínios R15 e R16 apresentaram aproximadamente, os mesmos valores de T_F , 1,055 e 1,06, respectivamente. Já o domínio R17 apresentou valor igual a 1,033 para sua temperatura de enovelamento. A partir dos dados experimentais, obtidos por Kusunoki et al.[85], para a temperatura de *melting*, T_m , temperatura na qual a proteína encontra-se parcialmente desenove-

lada e sem nenhuma atividade, o domínio R15 apresenta uma temperatura de *melting* igual a $(58,1 \pm 0,2)^\circ\text{C}$, R16 igual a $(49,4 \pm 0,1)^\circ\text{C}$ e R17 igual a $(31,0 \pm 0,1)^\circ\text{C}$. Com base nos dados experimentais, nota-se que o domínio R15 possui a maior temperatura de *melting*, seguido pelo domínio R16 e, por fim, pelo domínio R17. Em relação à estabilidade dos domínios, o R15 possui a maior estabilidade, quando comparado com seus homólogos, já que o mesmo apresentou a maior temperatura de *melting*.

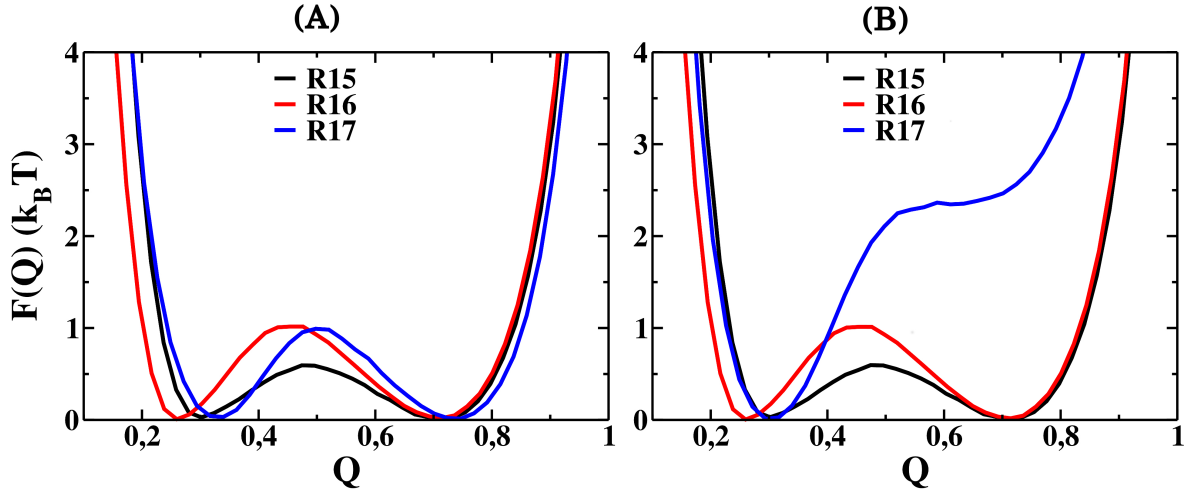
Os resultados obtidos pelas simulações computacionais, para a temperatura de enovelamento, exibiram uma diferença ínfima entre os valores do T_F para os três domínios estudados. Dessa forma, torna-se inviável realizar a análise citada no parágrafo anterior, já que nenhuma diferença significativa foi verificada. Em vista disso, uma análise mais rebuscada, baseada não só nos perfis de energia livre, mas também em suas respectivas barreiras, pode descrever se o domínio apresenta um enovelamento rápido ou lento. Caso a barreira de energia seja baixa, o enovelamento é rápido. Em contrapartida, se a barreira de energia é alta, o enovelamento é lento. A barreira de energia livre foi determinada a partir da diferença entre as energias do estado de transição, F^{TS} , e do estado desenovelado, F^D . Logo, $\Delta F(Q) = F^{TS} - F^D$.

De acordo com a Figura 5.2, os domínios R15, R16 e R17 estão representados pelas cores preta, vermelha e azul, respectivamente. Na Figura 5.2A, têm-se as curvas de energia livre para os três domínios em seus respectivos valores de T_F . Observou-se que os domínios R16 e R17 possuem os mesmos valores de $\Delta F(Q) = 1,0 k_B T$. Entretanto, o domínio R15 apresentou uma barreira de energia livre menor que a de seus homólogos, ou seja, $\Delta F(Q) = 0,5 k_B T$. No que diz respeito à estabilidade dos domínios em relação aos estados desenovelado, de transição e nativo, constatou-se que apenas o domínio R16 possui um estado desenovelado mais estável em relação aos seus homólogos, ao passo que, para os outros estados, verificou-se uma convergência de estabilidade.

Vale a pena enfatizar que os parâmetros termodinâmicos calculados experimentalmente foram obtidos em uma temperatura fixa, ou seja, os dados coletados foram realizados em uma temperatura única. A temperatura utilizada foi sempre abaixo da menor temperatura de *melting* dos domínios R15, R16 e R17. Por exemplo, no trabalho desenvolvido por Wensley et al.[49], os experimentos foram realizados à 25°C . Dessa forma, na Figura 5.2B, foram plotadas as curvas de energia livre na temperatura de enovelamento do domínio R15.

Com base na Figura 5.2A, notou-se que os domínios R15 e R16 não apresentaram nenhuma alteração em seus perfis de energia livre. Todavia, o domínio R17 possui uma T_F inferior a 1,055, que é a temperatura de enovelamento do domínio

Figura 5.2: Energia livre em função da fração dos contatos nativos, para o modelo *vanilla*. Na Figura A, têm-se os perfis de energia livre nos respectivos valores de T_F dos domínios R15, R16 e R17. Na Figura B, têm-se os perfis de energia livre dos domínios R15, R16 e R17 no valor de T_F do domínio R15. As cores preto, vermelha e azul corresponde aos domínios R15, R16 e R17, respectivamente.



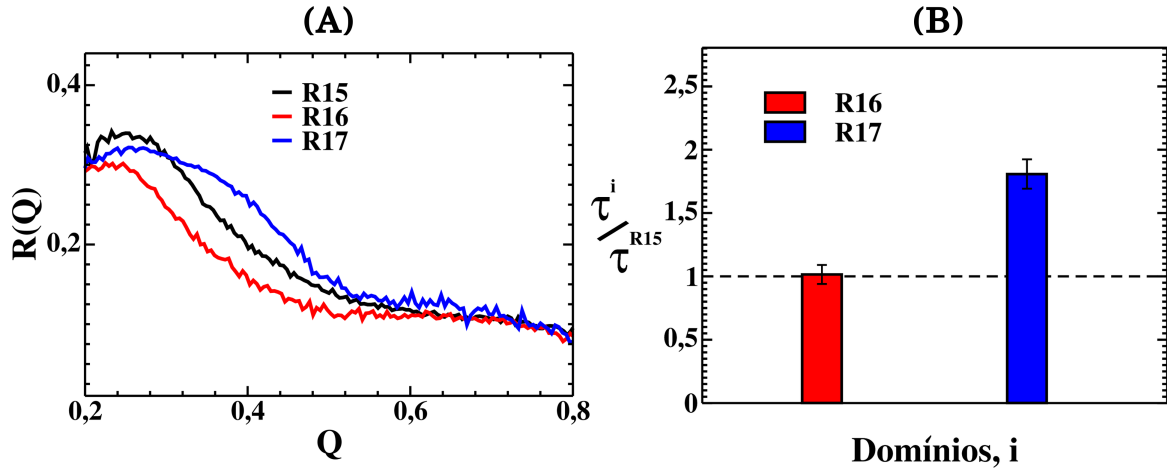
Fonte: Elaborada pelo autor.

R15. Consequentemente, o domínio R17 apresentou o estado enovelado desestabilizado, como pode ser visto na Figura 5.2B. No que diz respeito ao estado desenovelado, os domínios R15 e R17 apresentaram valores de Q iguais.

O R17 foi o único domínio que apresentou alterações em seu perfil de energia. Este domínio apresentou uma barreira de, aproximadamente, $1,0 k_B T$ quando analisado em seu valor de T_F . Entretanto, quando este domínio é analisado no T_F do domínio R15, sua barreira aumentou para valores próximos de $2,5 k_B T$. Tendo em vista este aumento de $1,5 k_B T$, foram calculadas e analisadas não só as rotas, $R(Q)$, como também os tempos de enovelamento para os três domínios no valor de T_F do R15.

A partir dos valores de $R(Q)$, é possível determinar a especificidade da rota e dos contatos nativos que devem ser formados durante o processo de enovelamento. Na Figura 5.3A, têm-se a rota de enovelamento para os domínios R15, R16 e R17. Todas as curvas foram plotadas na temperatura de enovelamento do domínio R15. Logo, observou-se que todos os domínios apresentaram, aproximadamente, o mesmo perfil de rota. Para o intervalo $0.20 < Q < 0.35$, verificou-se que o domínio R15 apresentou a mais alta especificidade, ou seja, o maior valor de $R(Q)$, seguido dos domínios R17 e R16. Entretanto, no intervalo $0.35 < Q < 0.50$, constatou-se que o domínio R17 apresentou o maior valor de $R(Q)$, seguido dos domínios R15 e R16, e para os valores acima de 0.50 , ocorreu uma convergência nas rotas de enovelamento.

Figura 5.3: Na Figura A, têm-se a rota de enovelamento, $R(Q)$, em função da coordenada de reação, Q . As rotas de enovelamento dos domínios R15, R16 e R17 estão representados pelas cores preta, vermelha e azul, respectivamente. Na Figura B, têm-se a razão dos tempos de enovelamento, τ , para os domínios R16 e R17, onde as linhas tracejadas correspondem à razão no tempo de enovelamento do domínio R15.



Fonte: Elaborada pelo autor.

A partir do cálculo do tempo de enovelamento, exibido na Figura 5.3B, verificou-se que o R16 apresenta o mesmo tempo de enovelamento que o R15. Em contrapartida, o R17 apresenta um tempo de enovelamento maior. De acordo com a literatura, o R15 é o domínio que enovela-se mais rapidamente, com duas e com três ordens de grandeza mais rápido que seus homólogos, R16 e R17, respectivamente. No entanto, as simulações realizadas para os domínios, por intermédio do modelo *vanilla*, não exibiram uma diferença significativa no cálculo do tempo de enovelamento, quando comparados com os resultados experimentais [49].

5.2 Modelo *Flavored*

O modelo *flavored* é caracterizado pela adição de um potencial gaussiano atrativo (equação 3.3) ao potencial do modelo baseado em estrutura (modelo *vanilla*). Além das interações nativas, considera-se favorável, para o processo de enovelamento da proteína, as interações não nativas hidrofóbicas (vide seção 3.2). A partir da equação 3.3, foram realizadas simulações para três diferentes valores de K_{HP} , 0,10, 0,20 e 0,30. Na Tabela 5.2 e na Figura 5.4, encontram-se as informações a respeito dos valores das temperaturas de enovelamento, dos valores de Q para os estados desenovelado, de transição e enovelamento (nativo), assim como, dos valores da barreira e os perfis de energia livre.

Tabela 5.2: Informações a respeito dos valores da temperatura de enovelamento, T_F ; fração dos contatos nativos formados no estado desenovelado, Q_D ; estado transição, Q_{TS} ; estado nativo, Q_N ; e barreira de energia livre, $\Delta F(Q)$ em diferentes valores de K_{HP} (0,10, 0,20 e 0,30) para os domínios R15, R16 e R17.

Domínios	K_{HP}	T_F	Q_D	Q_{TS}	Q_N	$\Delta F(Q)(k_B T)$
R15	0.10	1,060	0,38	0,53	0,70	0,28
	0.20	1,050	0,40	0,55	0,74	0,43
	0.30	1,026	0,43	0,62	0,81	1,04
R16	0.10	1,066	0,34	0,48	0,71	0,46
	0.20	1,060	0,37	0,47	0,74	0,26
	0.30	1,046	0,40	0,48	0,80	0,22
R17	0.10	1,040	0,34	0,56	$\sim 0,68$	$\sim 1,58$
	0.20	1,038	0,37	0,60	0,67	1,15
	0.30	1,020	0,40	0,58	0,77	1,12

Fonte: Elaborada pelo autor.

Com o aumento no valor de K_{HP} , observou-se uma diminuição nos valores de T_F . Para $K_{HP} = 0,10$, notou-se que os domínios R15 e R16 apresentaram temperaturas de enovelamento aproximadamente iguais (ver Figura 5.4A), sendo que para $K_{HP} = 0,30$, não observou-se o mesmo comportamento que em 0,10. Contudo, verificou-se uma proximidade nos valores de T_F dos domínios R15 e R17 (ver Figura 5.4C). O domínio R15 apresentou a maior variação na temperatura de enovelamento quando comparado a seus homólogos, enquanto que os domínios R16 e R17 apresentaram variações idênticas. As variações foram calculadas para os valores iguais a 0,10 e 0,30.

Com a diminuição nos valores de T_F , constatou-se que os domínios sofreram mudanças em seus respectivos estados. O domínio R15, por exemplo, apresentou $Q_D = 0,38$ em $K_{HP} = 0,10$ e $Q_D = 0,43$ em $K_{HP} = 0,30$. Este aumento corresponde a, aproximadamente, 10 contatos nativos, ou seja, o estado desenovelado tornou-se menos estável. Já os estados de transição e nativo sofreram pequenas alterações. Contudo, a barreira de energia livre apresentou um aumento significativo. Para $K_{HP} = 0,10$, a barreira de energia é igual $0,28 k_B T$, enquanto que, para K_{HP} igual 0,30, tem-se $1,04 k_B T$ para barreira de energia livre.

Já os domínios R16 e R17 não apresentaram variações significativas em seus estados desenovelado, de transição e nativo. Todavia, identificou-se que as barreiras de energia livre, para estes dois domínios, não apresentaram o mesmo comportamento que seu homólogo, o domínio R15, conforme citado no parágrafo anterior. Esses domínios, R16 e R17, apresentaram um decréscimo em suas respectivas barreiras de energia livre,

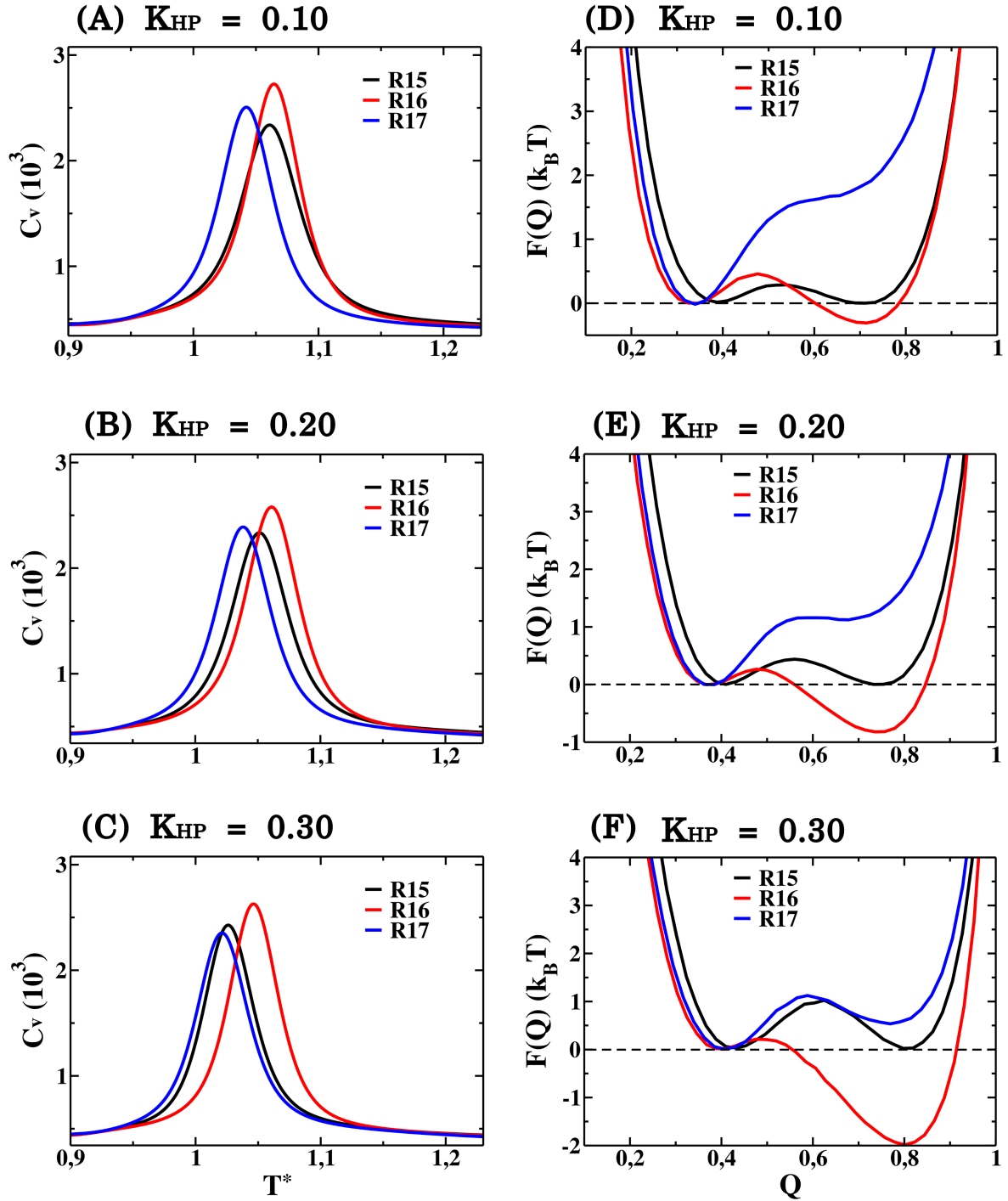
devido ao aumento no valor de K_{HP} . Portanto, para $K_{HP} = 0,10$, o R16 e o R17, apresentaram valores iguais a 0,46 e 1,58 $k_B T$ (ver Figura 5.4 D), respectivamente. Para este mesmo valor de K_{HP} , o domínio R16 possui o estado nativo mais estável dentre seus homólogos, com energia igual a -0.5 $k_B T$. Já o domínio R17 exibiu uma energia livre de, aproximadamente, 2 $k_B T$.

Para os demais valores de K_{HP} , 0,20 e 0,30 (ver Figuras 5.4 E e F), verificou-se que a barreira de energia livre do domínio R16 é menor que a do domínio R15, enquanto que o domínio R17 passa a assumir um perfil aproximadamente idêntico ao do R15, tanto em relação à barreira como nas posições dos estados de transição e nativo. Dessa forma, constatou-se que o aumento no valor de K_{HP} torna o domínio R16 mais estável, confirmando que as interações hidrofóbicas aumentaram a estabilidade deste domínio quando comparado com seus respectivos homólogos. Já para o domínio R17 constatou-se que, a partir do valor de $K_{HP} = 0,20$, ocorreu um aumento em sua estabilidade. Contudo, este aumento em sua estabilidade não foi capaz de tornar o domínio R17 mais estável do que o R15.

Comparando os resultados deste modelo, com os do modelo anterior (modelo *vanilla*), constatou-se que apenas os domínios R16 e R17 apresentaram uma redução nas barreiras de energia livre com a adição do potencial não nativo hidrofóbico. Para o domínio R16, as interações não nativas hidrofóbicas aumentaram sua estabilidade conforme altera-se o valor de K_{HP} . Entretanto, para os valores K_{HP} iguais a 0,20 e 0,30, as barreiras de energia livre desviaram-se consideravelmente dos resultados encontrados na literatura, qual relatam que o domínio R16 possui um valor intermediário aos seus homólogos, R15 e R17. Logo, o cálculo da rota e do tempo de enovelamento foram realizados apenas para o valor de $K_{HP} = 0,10$, pois este foi o único em que os resultado de barreira de energia livre corroboraram qualitativamente com aqueles mencionais por Best [47].

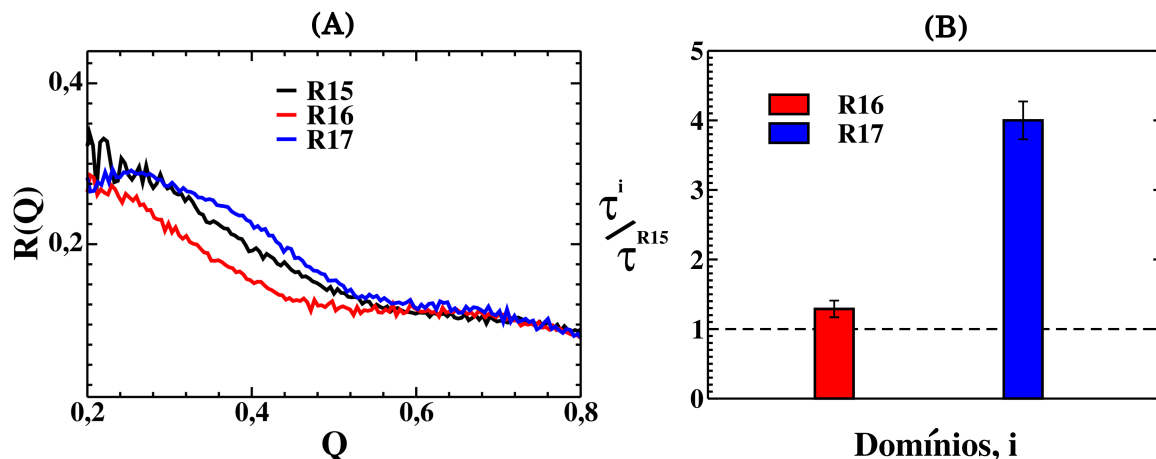
Na Figura 5.5, são exibidos os perfis de rota (ver Figura 5.5A) e os tempos de enovelamento para os domínios R16 e R17 (ver Figura 5.5B). Em ambas as figuras, os dados foram calculados na temperatura de enovelamento do domínio R15. Verificou-se, a partir do cálculo da rota de enovelamento, que os domínios não apresentaram diferenças nem mudanças significativas em seus respectivos perfis de rota. Entretanto, com base no cálculo do tempo de enovelamento, nota-se que o domínio R17 apresentou um aumento significativo em seu tempo de enovelamento. Todavia, o domínio R16 não apresentou um aumento considerável ao ponto de sustentar a ideia de que as interações não nativas hidrofóbicas afetam seu tempo de enovelamento, assim como foi notado para o domínio R17.

Figura 5.4: Nas Figuras A, B e C, têm-se as curvas de calor específico, C_v , em função da temperatura reduzida, T^* , e nas Figuras D, E e F, têm-se os perfis de energia livre em função da coordenada de reação, Q . Neste modelo de simulação adotou-se os valores K_{HP} iguais a 0,10 (A e D), 0,20 (B e E) e 0,30 (C e F). Os domínios R15, R16 e R17 estão representados pelas curvas preta, vermelha e azul, respectivamente.



Fonte: Elaborada pelo autor.

Figura 5.5: Na Figura A, têm-se a rota de enovelamento, $R(Q)$, em função da coordenada de reação, Q . As rotas de enovelamento dos domínios R15, R16 e R17 estão representados pelas cores preta, vermelha e azul, respectivamente. Na Figura B, têm-se a razão dos tempos de enovelamento, τ , para os domínios R16 e R17, onde as linhas tracejadas correspondem à razão no tempo de enovelamento do domínio R15. A rota e o tempo de enovelamento foram calculados para o valor de $K_{HP} = 0,10$.



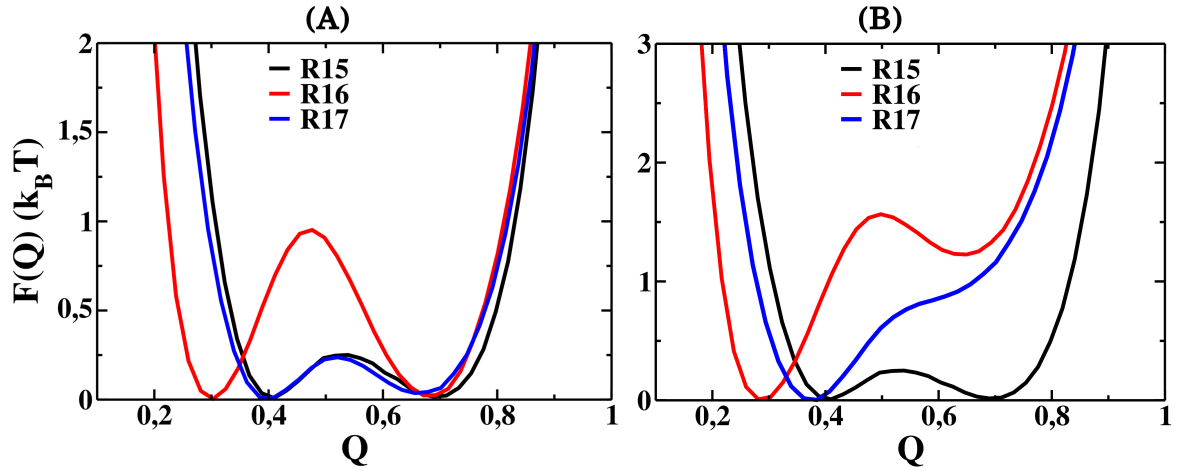
Fonte: Elaborada pelo autor.

Com este resultado, entende-se que a adição das interações não nativas hidrofóbicas afetam mais o domínio R17 do que o domínio R16, como pode ser visto na figura 5.5B. Este comportamento sugere que as interações atuam de maneiras diferentes nos domínios R16 e R17 e, com isso, foram realizadas simulações onde adicionou-se cargas nos respectivos aminoácidos, ácidos e básicos. Como objetivo averiguar como as interações eletrostáticas afetam o processo de enovelamento dos domínios estudados.

5.3 Modelo com Carga Fixa

Nesta seção, serão apresentados resultados que foram obtidos por meio de simulações realizadas no GROMACS, onde os resíduos apresentam carga fixa durante a dinâmica, ou seja, simulações que não permitem a desprotonação dos resíduos tituláveis. Os aminoácidos com carga positiva (Arginina, Histidina e Lisina) e com carga negativa (Ácido Aspártico e Ácido Glutâmico) nas simulações realizadas no GROMACS são iniciados com carga +1 e -1, respectivamente. A comparação entre resultados obtidos pelo GROMACS e ESPResSO, ou seja, simulações que permitem a protonação/desprotonação dos resíduos tituláveis, podem ser encontrada no Apêndice D.

Figura 5.6: Energia livre em função da fração dos contatos nativos, Q . Na Figura A, têm-se os perfis de energia livre nos respectivos valores de T_F dos domínios R15, R16 e R17. Na Figura B, têm-se os perfis de energia livre dos domínios R15, R16 e R17 no valor de T_F do domínio R15. As cores preto, vermelha e azul corresponde aos domínios R15, R16 e R17, respectivamente.



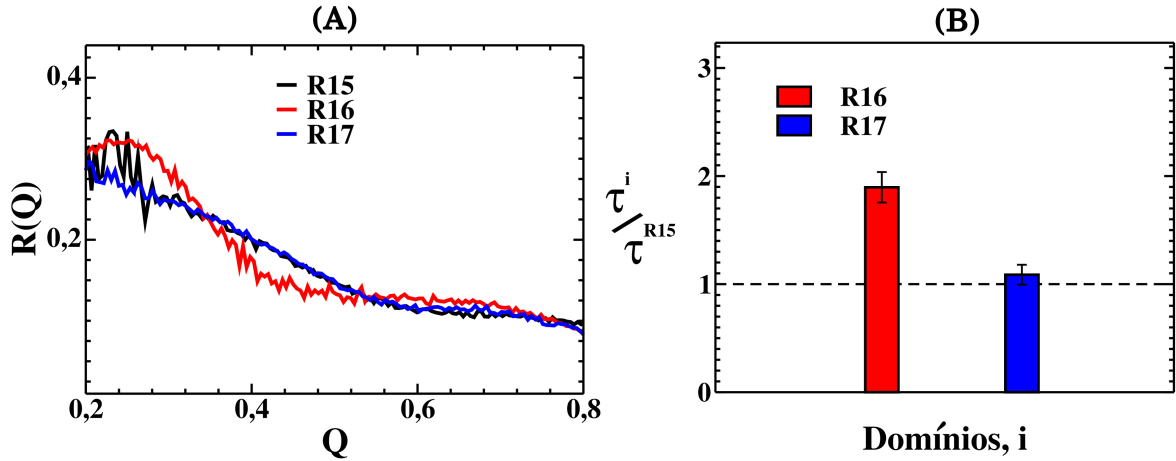
Fonte: Elaborada pelo autor.

Para os perfis de energia livre representados na temperatura de enovelamento do domínio R15 (Figura 5.6B), verificou-se que o domínio R16 possui a maior barreira de energia livre, seguido pelo domínio R17 e, por fim, pelo R15. Conforme mencionado na seção anterior, em que as interações não nativas hidrofóbicas afetam mais o processo de enovelamento do domínio R17, enquanto que, nesta seção, os resultados demonstram que as interações eletrostáticas afetam mais o processo de enovelamento do domínio R16, uma vez que este apresentou a barreira mais elevada.

As afirmações acima são sustentadas pelo cálculo do tempo de enovelamento. Na Figura 5.7, encontram-se os perfis de rota, e os tempos de enovelamento calculados para os domínios R15, R16 e R17. Baseando-se no cálculo da rota de enovelamento (ver Figura 5.7 A), não verificou nenhuma alteração significativa, baseando-se o mesmo comportamento obtido nas seções anteriores. Já para o cálculo do tempo de enovelamento (ver Figura 5.7 B), constatou-se que o domínio R16 necessita de um tempo maior para enovelar, enquanto que o domínio R17 apresentou um tempo de enovelamento próximo ao do domínio R15.

Como discutido na seção anterior, as interações hidrofóbicas afetam mais o enovelamento do domínio R17. Entretanto, nesta seção, foi verificado que, com a adição do termo eletrostático, o domínio R16 apresentou o tempo de enovelamento mais elevado, demonstrando que as interações eletrostáticas atuam mais fortemente neste domínio do que nos outros dois (R15 e R17).

Figura 5.7: Rota de enovelamento em função da fração dos contativos nativos, Q , e o tempo de enovelamento, τ , para os domínios R16 e R17 calculados no modelo com carga fixa. Na Figura A, têm-se a rota de enovelamento no respectivo T_F do R15, estando os domínios R15, R16 e R17 representados pelas cores preto, vermelho e azul, respectivamente. Na Figura B, têm-se a razão dos tempos de enovelamento para os domínios R16 e R17, e as linhas tracejadas correspondem à razão no tempo de enovelamento do domínio R15.



Fonte: Elaborada pelo autor.

Combinado os resultados desta seção com os da seção anterior, é possível concluir que, durante o processo de enovelamento, os domínios R16 e R17 apresentam comportamentos distintos, ou seja, as interações não nativas hidrofóbicas afetam mais a dinâmica do domínio R17, enquanto que as interações eletrostáticas alteram o domínio R16. Com o intuito de averiguar como as interações não nativas alteram o processo de enovelamento dos domínios estudados, acoplou-se os dois potenciais, hidrofóbico e eletrostático (equações 3.3 e 3.4, respectivamente) ao potencial do modelo baseado em estrutura (equação 3.1). Os resultados para este acoplamento de potenciais serão apresentados na próxima seção.

5.4 Modelo *Flavored* + Carga Fixa

Neste modelo de simulação foram utilizados os mesmos parâmetros dos modelos *Flavored* e Carga Fixa (seções 5.2 e 5.3) nas simulações realizadas com o acoplamento dos potenciais. Na Figura 5.8, têm-se o calor específico e os perfis de energia livre obtidos para os domínios R15, R16 e R17, os quais estão representados pelas cores preta, vermelha e azul, respectivamente.

Nas Figuras 5.8A, 5.8B e 5.8C, os valores T_F foram obtidos a partir do pico

do calor específico calculado para diferentes valores de K_{HP} , assim como realizado na seção 5.2. Portanto, para $K_{HP} = 0,10$, nota-se que os valores de T_F dos domínios R15 e R16 são aproximadamente iguais. Entretanto, o domínio R17 apresentou um valor inferior ao de seus homólogos. Logo, para o valor $K_{HP} = 0,20$, o domínio R15 apresenta o maior T_F , seguido pelo domínio R16 e, por fim, pelo R17. Enquanto que para $K_{HP} = 0,30$, os três domínios apresentaram valores de T_F aproximadamente iguais.

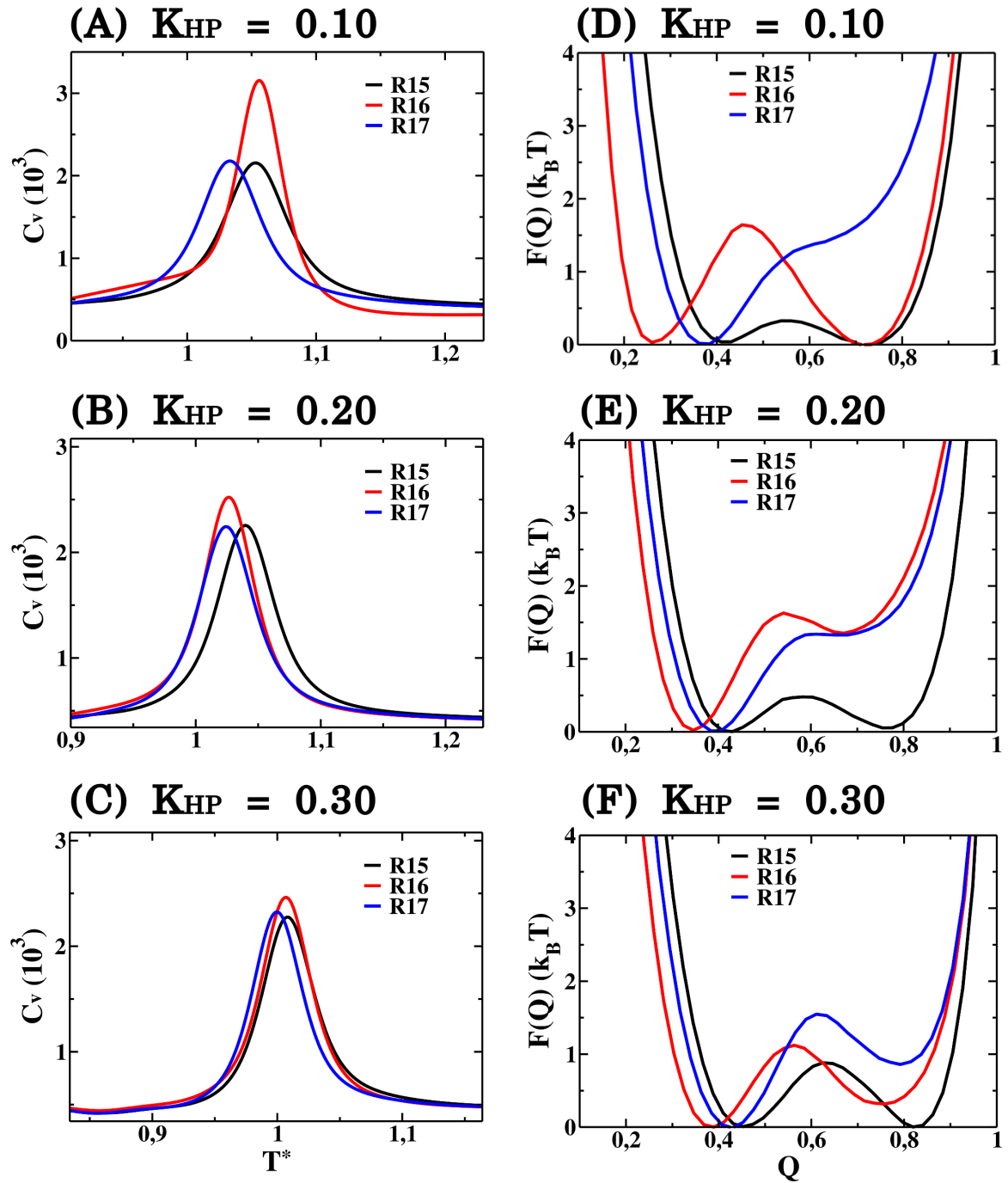
Nas Figuras 5.8D, 5.8E e 5.8F, têm-se os perfis de energia livre em função da fração do contatos nativos, onde todas as curvas encontram-se na temperatura de enovelamento do domínio R15. Para o valor de $K_{HP} = 0,10$, o domínio R16 apresenta a barreira de energia mais elevada e o estado desenovelado mais estável, enquanto que o domínio R17 possui o estado nativo menos estável. Já para $K_{HP} = 0,20$, o R16 e R17 apresentaram aproximadamente o mesmo grau de estabilidade no estado nativo e comparando os resultados de 0,20 com o 0,10, verifica-se que o estado desenovelado do domínio R16 convergiu para valores próximos aos dos domínios R15 e R17. Contudo, para o valor de $K_{HP} = 0,30$, o domínio R17 apresentou a barreira de energia mais elevada, enquanto que o R16 e R15 possuem valores inferiores. Comparou-se os resultados de $K_{HP} = 0,30$ com os de 0,20 e 0,10 e notou-se que os domínios R15 e R17 tiveram um aumento em sua barreira de energia livre, enquanto que o domínio R16 apresentou uma diminuição.

A partir deste ponto este trabalho somente apresentará resultados obtidos para o valor de $K_{HP} = 0,10$ pois, para o valor de 0,20, o domínio R16 apresenta a barreira de energia livre mais elevada, enquanto que para 0,30, os estados desenovelados estão próximos de Q igual a 0,50, indicando que neste valor de K_{HP} , os domínios estão sobre forte colapso hidrofóbico.

Na Figura 5.9, têm-se a comparação dos tempos de enovelamento dos modelos que foram adotados, como também, o tempo de enovelamento experimental. Conforme as interações não nativas foram adicionadas, observou-se efeitos distintos nos domínios R16 e R17. Com o modelo *Flavored* (HP^{HP}), as interações não nativas hidrofóbicas atuam mais efetivamente no domínio R17 e, conseqüentemente, o tempo de enovelamento deste domínio é superior a seus homólogos. Já no modelo com carga fixa (Elec) observou-se o contrário, onde o domínio R16 apresenta o tempo mais elevado quando foi adicionado a ele o potencial eletrostático. Assim, com o acoplamento dos potenciais ($HP^{HP} + Elec$), os domínios R16 e R17 apresentaram uma ordem de grandeza no tempo quando comparados com o domínio R15, onde o R17 possui o tempo mais elevado, seguido pelo R16.

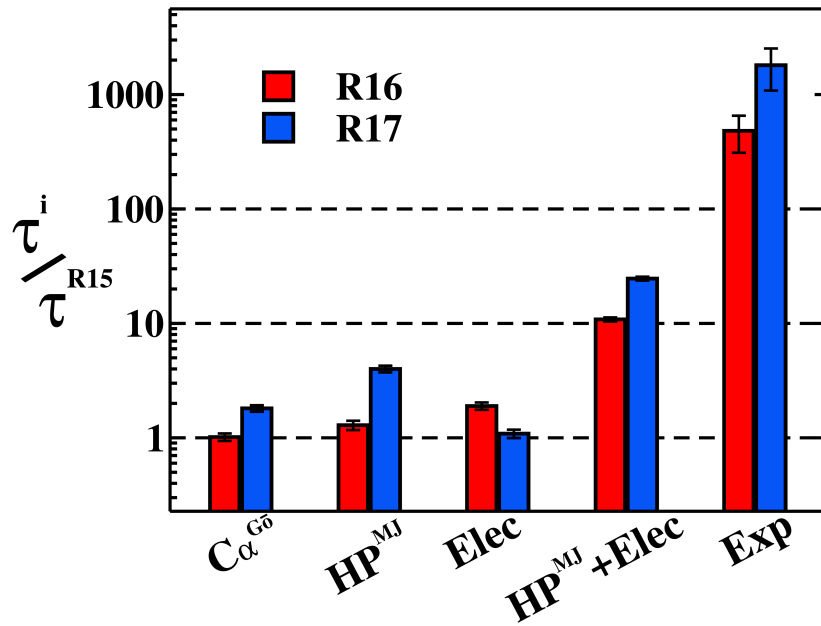
Neste trabalho foi adotado um modelo de simulação simplificado, onde resul-

Figura 5.8: Nas Figuras A, B e C, têm-se as curvas de calor específico, C_v , em função da temperatura reduzida, T^* , e nas Figuras D, E e F, têm-se os perfis de energia livre em função da coordenada de reação, Q . Neste modelo de simulação adotou-se os valores K_{HP} iguais a 0,10 (A e D), 0,20 (B e E) e 0,30 (C e F). Os domínios R15, R16 e R17 estão representados pelas curvas preta, vermelha e azul, respectivamente.



Fonte: Elaborada pelo autor.

Figura 5.9: Comparação entre os tempos de enovelamento de todos os modelos de simulações adotados para os domínios R15, R16 e R17. Representação dos modelos: *vanilla* ($C_\alpha^{G\ddot{o}}$), *flavored* (HP^{MJ}), carga fixa (Elec), *flavored* + carga fixa ($HP^{MJ} + Elec$) e experimental (Exp). Os domínios R16 e R17 estão representados pelas cores vermelha e azul, e as linhas tracejadas igual um correspondem à razão no tempo de enovelamento do domínio R15. Os dados experimentais foram extraídos de Scott et al.[45].



Fonte: Elaborada pelo autor.

tados obtidos apresentaram uma ordem de grandeza entre os tempos de enovelamento. Assim, com o acoplamento dos potenciais, constatou-se a origem do atrito interno nos domínios R16 e R17 e, deste modo, verificou-se como estes atuam de maneiras distintas nos domínios. Com o objetivo de validar o modelo de simulação adotado nesta seção, realizou-se simulações para as mutações mencionadas na introdução deste trabalho, onde verificou-se que o tempo de enovelamento dos domínios R16 e R17 diminuem conforme as mutações foram realizadas.

5.4.1 Análise das Mutações

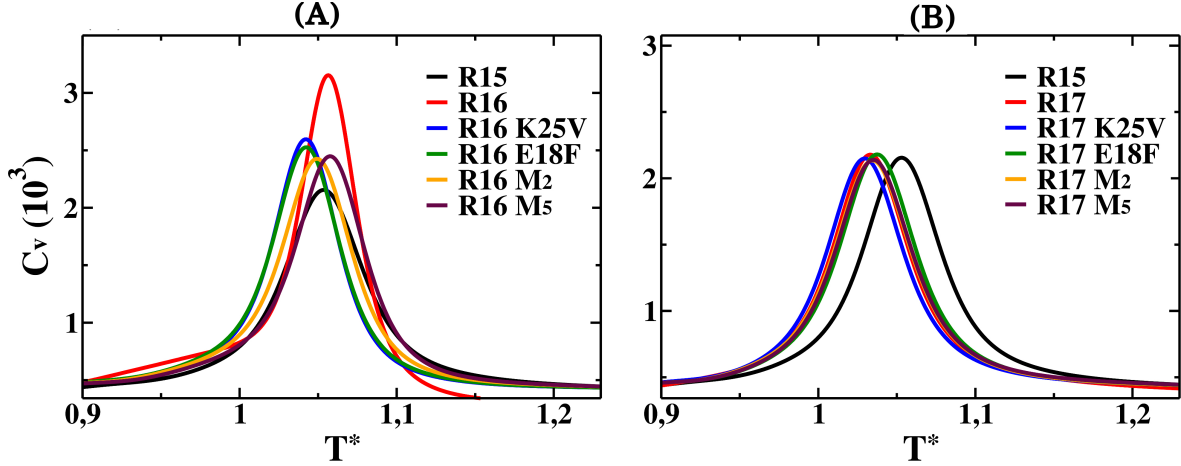
De acordo com a literatura, as mutações K25V, E18F, M₂ (E18F e K25V) e M₅ (E18F, E19D, I22L, K25V e V30L), alteram o mecanismo de enovelamento dos domínios R16 e R17, como também aumentam a taxa de enovelamento. Os resultados obtidos por simulações tem como objetivo verificar se o modelo adotado é capaz de corroborar com os dados experimentais.

Nas Figuras 5.10A e 5.10B, encontram-se as curvas de calor específico em função da temperatura reduzida para os domínios R15, R16 e R17. Observou-se que para as mutações na R16, K25V e E18F, ocorreu uma redução na temperatura de enovelamento, e logo em seguida, um aumento nas temperaturas para as mutações M₂ e M₅. A mutação, M₅, foi a que apresentou valor de T_F aproximadamente igual ao domínio R16 em sua forma regular. Contudo, para o domínio R17, as temperaturas de enovelamento para as mutações são aproximadamente iguais ao de sua forma regular.

Na Figura 5.11 encontram-se os perfis de energia livre para domínios em sua forma regular e com suas mutações, sendo que todas as curvas estão na temperatura de enovelamento do domínio R15. Para o domínio R16 (ver Figura 5.11A), conforme as mutações são realizadas verifica-se uma diminuição na barreira de energia, sendo que a mutação R16 M₅ apresenta a maior diminuição, dentre todas, e seu estado nativo foi mais estável. Já o domínio R17 (ver Figura 5.11B) apresentou um pequeno decaimento no valor da barreira de energia.

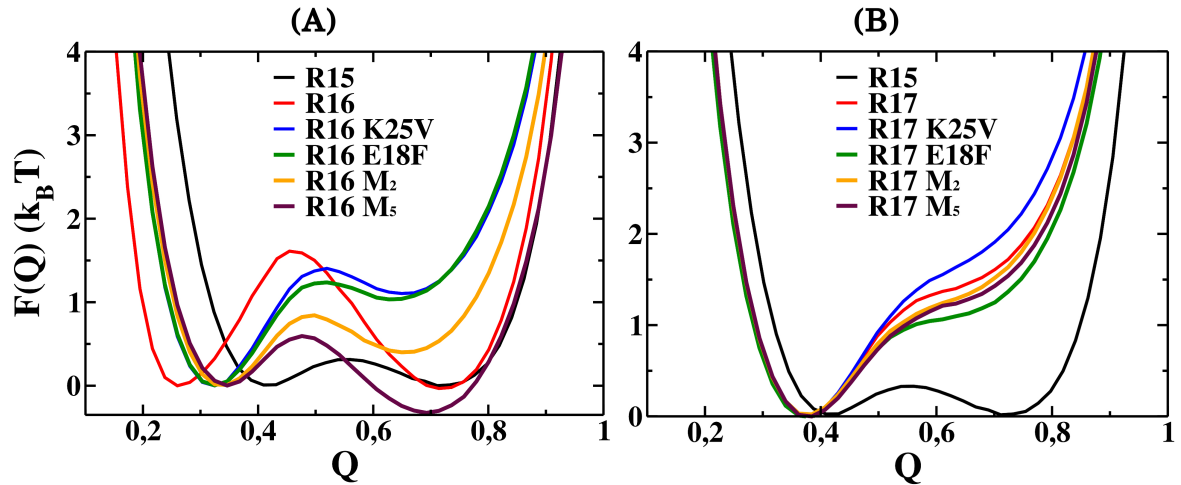
De acordo com a literatura, essas mutações são responsáveis por alterarem a taxa de enovelamento de ambos os domínios, R16 e R17, sendo que, nos resultados obtidos por simulações computacionais para a energia livre, apenas o domínio R16 apresentou grandes variações. Sabe-se que quanto maior a barreira, menor a taxa de enovelamento e, consecutivamente, maior o tempo para a proteína se enovelar. Dessa forma, calcula-se o tempo de primeira passagem (Mean First Passage Time - MFPT) de ambos os domínios com as mutantes.

Figura 5.10: Calor específico, C_v , em função da temperatura reduzida, T^* , para as simulações realizadas com as mutações K25V, E18F, M₂ e M₅. Na Figura A, têm-se as curvas de calor específico para o domínio R16 em sua forma regular e com as mutações. Na Figura B, têm-se as curvas de calor específico para o domínio R17 em sua forma regular e com as mutações.



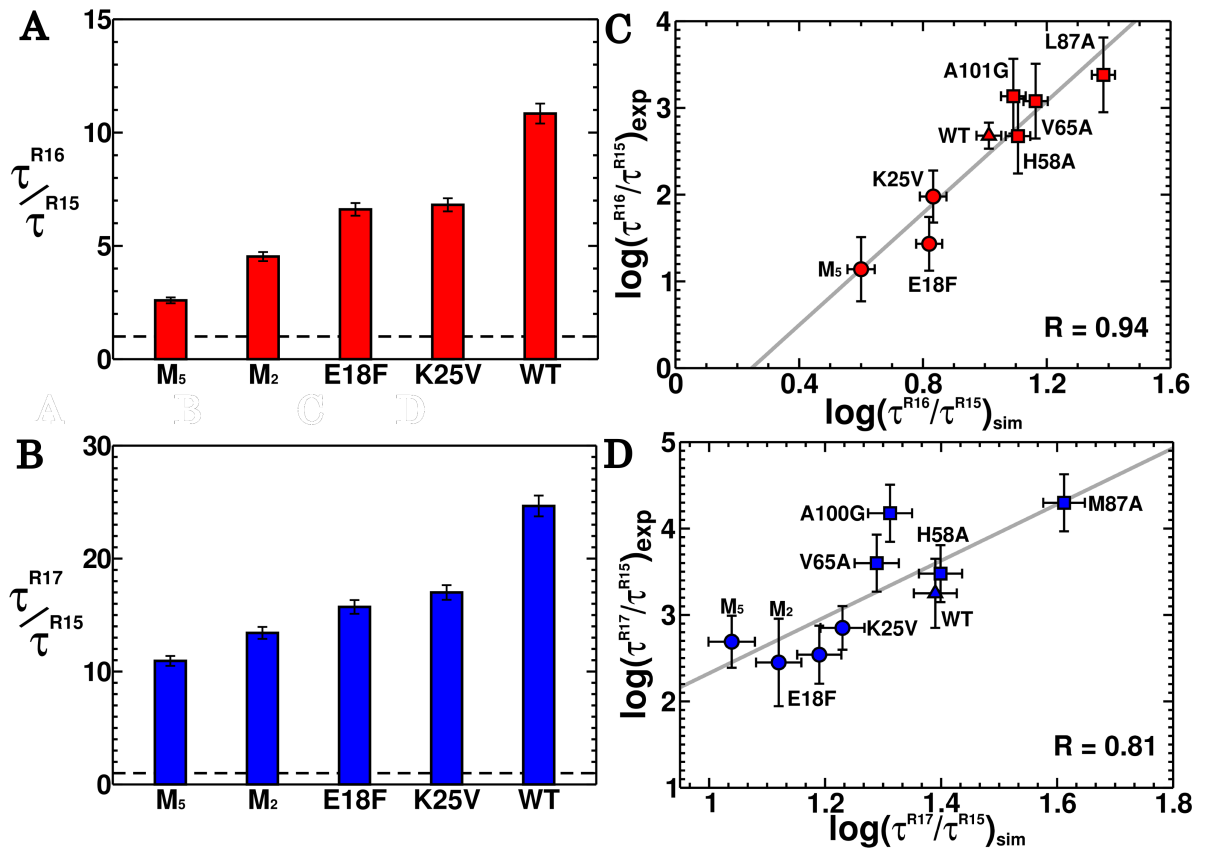
Fonte: Elaborada pelo autor.

Figura 5.11: Energia livre em função da fração dos contatos nativos, Q , para as mutações com o modelo *flavored* + carga fixa. Na Figura A, têm-se os perfis de energia livre para os domínios R15 e R16 na forma regular e para o R16 com as mutações K25V, E18F, M₂ e M₅. Na Figura B, têm-se os perfis de energia livre dos domínios R15 e R17 na forma regular regular e para o R17 com as mutações K25V, E18F, M₂ e M₅. Ambas as figuras foram plotadas na temperatura de enovelamento do domínio R15.



Fonte: Elaborada pelo autor.

Figura 5.12: Comparação entre os valores do tempo enovelamento experimental e computacional para os domínios R16 e R17 na sua forma regular e mutada. Nas Figuras A e B, têm-se os valores obtidos computacionalmente para os domínios R15, R16 e R17 em sua forma regular e para os domínios R16 e R17 com as mutações K25V, E18F, M₂ e M₅. Nas Figuras C e D, têm-se os logaritmos das razões dos tempos de enovelamento experimental e computacional para as mutações que aumentam a taxa de enovelamento (K25V, E18F, M₂ e M₅) e para aquelas que diminuem (H58A, V65A, M87A, L87A, A100G e A101G). O coeficiente correlação, R, foi obtido pelo ajuste linear (Figuras C e D). Os dados experimentais para os domínios em sua forma regular foram extraídos de Wensley et al.[49], para as mutações que aumentam a taxa de enovelamento de Scott, Randles e Clarke[55] e para as mutações que diminuem a taxa de Scott et al.[86].



Fonte: Elaborada pelo autor.

Nas Figuras 5.12A e 5.12B, encontram-se os tempos calculados para as mutações citadas anteriormente. Para o domínios R16 (ver Figura 5.12A), verificou-se um decaimento no tempo de enovelamento conforme as mutações foram realizadas, chegando a valores próximos de um, ou seja, o R16 apresentou valores no tempo de enovelamento aproximadamente igual ao R15. No domínio R17 (ver Figura 5.12B) também verificou-se um decaimento no tempo de enovelamento conforme as mutações foram realizadas, mas este decaimento não foi tão abrupto quanto ao do domínio R16. Com o cálculo do tempo de enovelamento, notou-se que nossos resultados corroboraram qualitativamente com os resultados da literatura, onde as respectivas mutações aumentaram o tempo de enovelamento dos domínios.

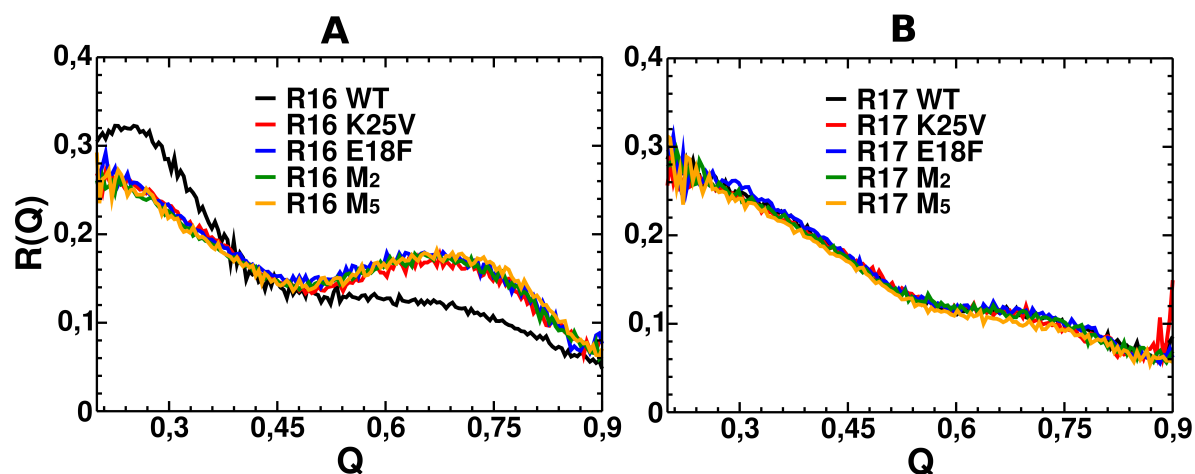
Foram realizadas simulações não só para as mutações que diminuem o tempo de enovelamento, mas também para aquelas mutações que aumentaram o tempo de enovelamento, sendo elas: H58A, V65A, M87A, L87A, A100G e A101G (ver Figura 5.12 C e D). Ao final de cada simulação são comparados os resultados obtidos computacionalmente com os resultados experimentais. Na Figura 5.12C, o coeficiente de correlação, R , do domínio R16 foi de 0,94, ou seja, nossos resultados apresentaram o mesmo comportamento que os dados da literatura e os resultados estão altamente correlacionados. Para o domínio R17 o sistema apresentou comportamento semelhante, onde o coeficiente de correlação foi de 0,81 e, apesar desse valor ser inferior que o R do R16, o sistema apresenta estar altamente correlacionado, onde apenas a mutação A100G encontra-se fora do ajuste linear.

Por fim, calculou-se a rota de enovelamento e verificou-se uma alteração no perfil da rota do domínio R16 no momento que as mutações são realizadas (ver Figura 5.13A). Em sua forma regular, o domínio R16 apresentou alta especificidade para valores de Q abaixo de 0,40, indicando que é necessária a formação de contatos específicos para este domínio enovelar-se. Assim, após a ocorrência das mutações, o perfil da rota alterou-se indicando agora uma alta especificidade depois do estado de transição. Para o caso do domínio R17 não foi verificado nenhuma alteração no perfil da rota de enovelamento, sendo que todos os mutantes apresentam perfis idênticos ao R17 em sua forma regular.

Com o cálculo da rota de enovelamento verificou-se que as mutações além de alterarem a taxa e o mecanismo de enovelamento, também modificaram o perfil da rota do domínio R16 e os tempos de enovelamento do R16 e R17. Além do mais, com a adição dos potenciais foi possível caracterizar o atrito interno nos domínios estudados, onde o R16 é caracterizado pelas interações eletrostáticas e o R17 pelas interações hidrofóbicas. Com o acoplamento dos potenciais houve uma melhor descrição do sistema estudado e uma explicação mais abrangente sobre as amplas diferenças

encontradas no tempo de enovelamento para os domínios estudados.

Figura 5.13: Cálculo da rota de enovelamento, $R(Q)$, em função da fração dos contatos nativos, Q , para os domínios R16 e R17 em sua forma regular e com as mutações. Na Figura A, têm-se as perfis de rota para o domínios R16 em sua forma regular e com as mutações (K25V, E18F, M_2 e M_5). Na Figura B, têm-se as perfis de rota para o domínios R17 em sua forma regular e com as mutações (K25V, E18F, M_2 e M_5). Ambas as figuras foram calculadas na temperatura de enovelamento do domínio R15 em sua forma regular.



Fonte: Elaborada pelo autor.

Capítulo 6

Conclusão

Este trabalho procurou compreender as contribuições das interações hidrofóbicas e eletrostáticas no processo de enovelamento dos domínios R15, R16 e R17 da α -Espectrina. Foram aplicados diferentes modelos de simulação, onde primeiro foi utilizado o modelo *vanilla*, seguido do modelo *flavored*, posteriormente simulações mantendo a carga fixa e por fim o modelo de acoplamento dos potenciais, modelo *flavored* + carga fixa.

Com o modelo *vanilla*, que trata de forma homogênea as interações, os resultados não apresentam diferenças significativas nas análises realizadas, como por exemplo, no cálculo da rota de enovelamento, onde os domínios apresentam perfis equivalentes. Já para o cálculo do tempo de enovelamento, não observou-se o mesmo comportamento encontrado na literatura [46].

O modelo *flavored* é caracterizado por distinguir as interações, sendo que neste trabalho foi adicionado um potencial gaussiano atrativo nas interações não nativas hidrofóbicas, de modo que os resultados obtidos por este modelo evidenciaram que estas interações não alteram em nenhum aspecto os domínios R15 e R16. Entretanto, o domínio R17 apresenta um acréscimo em seu tempo de enovelamento, como também em sua barreira de energia, logo entende-se que essas interações caracterizam o atrito interno para este domínio.

No modelo carga fixa, as interações eletrostáticas foram mediadas por um potencial de Coulomb blindado e os resultados obtidos por este modelo demonstram que os domínios R15 e R17 não apresentam alterações significativas, do ponto de vista termodinâmico e cinético. Por outro lado, o domínio R16 foi o que mais apresentou alterações, como pôde-se verificar no perfil de energia livre e no cálculo do tempo de enovelamento. Portanto, verificou-se que as interações eletrostáticas afetam mais o domínio R16.

Com o acoplamento dos potenciais, o cálculo do tempo de enovelamento foi uma ordem de grandeza maior para os domínios R16 e R17, quando comparados com o do R15. Além disso, com os resultados obtidos para as mutações, valida-se o modelo de simulação, pois estes foram concordantes com os da literatura. Ademais, a rota de enovelamento para o domínio R16 em sua forma regular e com as mutações foi modificada, indicando que este domínio contornou os efeitos do atrito interno. Tais resultados foram comprovados a partir do cálculo do tempo de enovelamento.

Por fim, todos os objetivos propostos no início deste trabalho foram alcançados, além de que, os resultados aqui obtidos apresentam grande relevância para área de estudo do atrito interno, sendo que o tratamento realizado neste estudo não foi evidenciado por nenhuma outra abordagem sobre os domínios presentes.

A partir dos resultados aqui obtidos, pode-se ter como perspectiva a realização dos cálculos de ϕ , afim de explorar certas regiões onde ocorreram o decaimento e o acréscimo nos perfis de energia livre, como também nos perfis de rota de enovelamento. Este cálculo permitirá uma análise mais rebuscada dessas regiões, onde exibirão quais contatos e quais aminoácidos são imprescindíveis para o processo de enovelamento, bem como na explicação dos efeitos do atrito interno.

Referências Bibliográficas

- 1 VOET, D. *Fundamentos de Bioquímica. A Vida em Nível Molecular (Em Portuguesa do Brasil)*. [S.l.]: Artmed, 2014. ISBN 8582710658.
- 2 NELSON, D. L.; COX, M. M. *Lehninger Principles of Biochemistry, Fourth Edition*. [S.l.]: W. H. Freeman, 2004.
- 3 DOBSON, C. M. Protein folding and misfolding. *Nature*, Nature Publishing Group, v. 426, n. 6968, p. 884–890, dec 2003.
- 4 HATTERS, D. M. Protein misfolding inside cells: The case of huntingtin and huntington's disease. *IUBMB Life*, Wiley Subscription Services, Inc., a Wiley company, v. 60, n. 11, p. 724–728, 2008.
- 5 ANFINSEN, C. B. et al. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, v. 47, p. 1309–1314, 1961.
- 6 ANFINSEN, C. B. Principles that govern the folding of protein chains. *Science*, American Association for the Advancement of Science, v. 181, n. 4096, p. 223–230, 1973.
- 7 LEVINTHAL, C. Are there pathways for protein folding? *J. Chem. Phys.*, v. 65, n. 1, p. 44–45, 1968.
- 8 ZWANZIG, R.; SZABO, A.; BAGCHI, B. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*, v. 89, n. 1, p. 20–22, 1992.
- 9 BRUNORI, M. et al. Morphogenesis of a protein: folding pathways and the energy landscape. *Biochemical Society Transactions*, Portland Press Ltd., v. 40, n. 2, p. 429–432, 2012.
- 10 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.
- 11 SHOEMAKER, B.; WANG, J.; WOLYNES, P. G. Structural correlations in protein folding funnels. *Proceedings of the National Academy of Sciences*, v. 94, n. 3, p. 777–782, 1997.
- 12 DILL, K. A.; CHAN, H. S. From levinthal to pathways to funnels. *Nature Structure Molecular Biology*, v. 4, n. 1, p. 10–19, 1997.

-
- 13 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.
- 14 NETTELS, D. et al. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 104, n. 8, p. 2655–2660, 2007.
- 15 BRYNGELSON, J. D. et al. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Genetics*, Wiley-Blackwell, v. 21, n. 3, p. 167–195, 1995.
- 16 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 proteins¹. *Journal of Molecular Biology*, v. 298, n. 5, p. 937 – 953, 2000.
- 17 WHITFORD, P. C. et al. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Structure, Function, and Bioinformatics*, Wiley-Blackwell, v. 75, n. 2, p. 430–441, 2009.
- 18 LU, Q.; LU, H. P.; WANG, J. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Physical Review Letters*, American Physical Society (APS), v. 98, n. 12, 2007.
- 19 WHITFORD, P. C. et al. Conformational transitions of adenylate kinase: Switching by cracking. *Journal of Molecular Biology*, Elsevier BV, v. 366, n. 5, p. 1661–1671, 2007.
- 20 BEST, R. B.; CHEN, Y.-G.; HUMMER, G. Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure*, Elsevier BV, v. 13, n. 12, p. 1755–1763, dec 2005. Disponível em: <https://doi.org/10.1016/j.str.2005.08.009>.
- 21 ZUCKERMAN, D. M. Simulation of an ensemble of conformational transitions in a united-residue model of calmodulin. *The Journal of Physical Chemistry B*, American Chemical Society (ACS), v. 108, n. 16, p. 5127–5137, 2004.
- 22 LU, Q.; WANG, J. Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *Journal of the American Chemical Society*, American Chemical Society (ACS), v. 130, n. 14, p. 4772–4783, 2008.
- 23 CHO, S. S.; LEVY, Y.; WOLYNES, P. G. P versus q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc Natl Acad Sci U S A*, v. 103, n. 3, p. 586–591, 2006.
- 24 SNOW, C. D. et al. How well can simulation predict protein folding kinetics and thermodynamics? *Annual Review of Biophysics and Biomolecular Structure*, Annual Reviews, v. 34, n. 1, p. 43–69, 2005.

-
- 25 MOURO, P. R. *Estudo da Interação não-nativa no enovelamento de proteínas*. Dissertação (Mestrado em Biofísica Molecular) — UNESP, São José do Rio Preto, 2014.
- 26 CHAN, H. S.; DILL, K. A. Transition states and folding dynamics of proteins and heteropolymers. *The Journal of Chemical Physics*, AIP Publishing, v. 100, n. 12, p. 9238–9257, 1994.
- 27 BROCKWELL, D. J.; RADFORD, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Current Opinion in Structural Biology*, Elsevier BV, v. 17, n. 1, p. 30–37, 2007.
- 28 BARTLETT, A. I.; RADFORD, S. E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nature Structural & Molecular Biology*, Springer Nature, v. 16, n. 6, p. 582–588, 2009.
- 29 CHEN, T.; SONG, J.; CHAN, H. S. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Current Opinion in Structural Biology*, Elsevier BV, v. 30, p. 32–42, feb 2015.
- 30 FRIEL, C. T.; BEDDARD, G. S.; RADFORD, S. E. Switching two-state to three-state kinetics in the helical protein im9 via the optimisation of stabilising non-native interactions by design. *Journal of Molecular Biology*, Elsevier BV, v. 342, n. 1, p. 261–273, 2004.
- 31 CAPALDI, A. P.; KLEANTHOUS, C.; RADFORD, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nature Structural Biology*, Springer Nature, 2002.
- 32 CHEN, T.; CHAN, H. S. Native contact density and nonnative hydrophobic effects in the folding of bacterial immunity proteins. *PLoS Comput Biol*, Public Library of Science, v. 11, n. 5, p. 1–27, 05 2015.
- 33 CHO, J.-H. et al. Electrostatic interactions in the denatured state ensemble: Their effect upon protein folding and protein stability. *Archives of Biochemistry and Biophysics*, Elsevier BV, v. 469, n. 1, p. 20–28, 2008.
- 34 AZIA, A.; LEVY, Y. Nonnative electrostatic interactions can modulate protein folding: Molecular dynamics with a grain of salt. *Journal of Molecular Biology*, v. 393, n. 2, p. 527 – 542, 2009.
- 35 ZARRINE-AFSAR, A. et al. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the fyn SH3 domain. *Proteins: Structure, Function, and Bioinformatics*, Wiley-Blackwell, v. 80, n. 3, p. 858–870, 2011.
- 36 CONTESSOTO, V. G. et al. NTL9 folding at constant pH: The importance of electrostatic interaction and pH dependence. *Journal of Chemical Theory and Computation*, American Chemical Society (ACS), v. 12, n. 7, p. 3270–3277, 2016.
- 37 GIVATY, O.; LEVY, Y. Protein sliding along DNA: Dynamics and structural characterization. *Journal of Molecular Biology*, Elsevier BV, v. 385, n. 4, p. 1087–1097, 2009.

-
- 38 DAVIS, S. J. et al. The role of charged residues mediating low affinity protein–protein recognition at the cell surface by cd2. *Proceedings of the National Academy of Sciences*, v. 95, n. 10, p. 5490–5494, 1998.
- 39 GANGULY, D.; ZHANG, W.; CHEN, J. Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins. *PLoS Computational Biology*, Public Library of Science (PLOS), v. 9, n. 11, p. e1003363, 2013.
- 40 ELGSAETER, A. et al. The molecular basis of erythrocyte shape. *Science*, American Association for the Advancement of Science, v. 234, n. 4781, p. 1217–1223, 1986.
- 41 MARCHESI, V. T.; STEERS, E. Selective solubilization of a protein component of the red cell membrane. *Science*, American Association for the Advancement of Science, v. 159, n. 3811, p. 203–204, 1968.
- 42 SCHWOCH, G.; PASSOW, H. Preparation and properties of human erythrocyte ghosts. *Molecular and Cellular Biochemistry*, v. 2, n. 2, p. 197–218, 1973.
- 43 LEVINE, J.; WILLARD, M. Fodrin: axonally transported polypeptides associated with the internal periphery of many cells. *The Journal of Cell Biology*, v. 90, n. 3, p. 631–642, set. 1981.
- 44 KANDA, K.; TANAKA, T.; SOBUE, K. Calspectin (fodrin or nonerythroid spectrin)-actin interaction: A possible involvement of 4.1-related protein. *Biochemical and Biophysical Research Communications*, v. 140, n. 3, p. 1051 – 1058, 1986.
- 45 SCOTT, K. A. et al. The folding of spectrin domains i: Wild-type domains have the same stability but very different kinetic properties. *Journal of Molecular Biology*, v. 344, n. 1, p. 195 – 205, 2004.
- 46 BORGIA, A. et al. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nature Communications*, Nature Publishing Group, v. 3, p. 1195, 2012.
- 47 BEST, R. B. How well does a funneled energy landscape capture the folding mechanism of spectrin domains? *The Journal of Physical Chemistry B*, American Chemical Society (ACS), v. 117, n. 42, p. 13235–13244, 2013.
- 48 ANSARI, A. et al. The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, American Association for the Advancement of Science, v. 256, n. 5065, p. 1796–1798, 1992.
- 49 WENSLEY, B. G. et al. Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 109, n. 44, p. 17795–17799, 2012.
- 50 BALDWIN, R. L. How does protein folding get started? *Trends in Biochemical Sciences*, Elsevier BV, v. 14, n. 7, p. 291–294, 1989.

-
- 51 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. *Journal of Molecular Biology*, Elsevier BV, v. 254, n. 2, p. 260–288, 1995.
- 52 PTITSYN, O. Protein folding: Hypotheses and experiments. *Journal of Protein Chemistry*, Springer Nature, v. 6, n. 4, 1987.
- 53 LIN, C. C.-J.; CHANG, J.-Y. Pathway of oxidative folding of bovine α -interferon: Predominance of native disulfide-bonded folding intermediates. *Biochemistry*, American Chemical Society (ACS), v. 46, n. 12, p. 3925–3932, 2007.
- 54 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.
- 55 SCOTT, K. A.; RANGLES, L. G.; CLARKE, J. The folding of spectrin domains II: Phi-value analysis of R16. *Journal of Molecular Biology*, v. 344, n. 1, p. 207 – 221, 2004.
- 56 AIOANEI, D. et al. Single-molecule-level evidence for the osmophobic effect. *Angewandte Chemie International Edition*, Wiley-Blackwell, v. 50, n. 19, p. 4394–4397, 2011.
- 57 ZAGROVIC, B.; PANDE, V. Solvent viscosity dependence of the folding rate of a small protein: Distributed computing study. *Journal of Computational Chemistry*, Wiley-Blackwell, v. 24, n. 12, p. 1432–1436, 2003.
- 58 KLIMOV, D. K.; THIRUMALAI, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.*, American Physical Society, v. 79, p. 317–320, Jul 1997.
- 59 JACOB, M. et al. Diffusion control in an elementary protein folding reaction. *Proceedings of the National Academy of Sciences*, v. 94, n. 11, p. 5622–5627, 1997.
- 60 PLAXCO, K. W.; BAKER, D. Limited internal friction in the rate-limiting step of a two-state protein folding reaction. *Proceedings of the National Academy of Sciences*, v. 95, n. 23, p. 13591–13596, 1998.
- 61 SILOW, M.; OLIVEBERG, M. High concentrations of viscosogens decrease the protein folding rate constant by prematurely collapsing the coil. *Journal of Molecular Biology*, v. 326, n. 1, p. 263 – 271, 2003.
- 62 SORANNO, A. et al. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 109, n. 44, p. 17800–17806, 2012.
- 63 FRAUENFELDER, H. *The Physics of Proteins: An Introduction to Biological Physics and Molecular Biophysics (Biological and Medical Physics, Biomedical Engineering)*. [S.l.]: Springer, 2010.

-
- 64 TAKETOMI, H.; UEDA, Y.; GO, 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. *International Journal of Peptide and Protein Research*, v. 7, n. 6, p. 445–459, 1975.
- 65 BERMAN, H. M. et al. The protein data bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242, 2000.
- 66 NOEL, J. K. et al. Smog 2: A versatile software package for generating structure-based models. *PLoS Comput Biol*, Public Library of Science, v. 12, n. 3, p. 1–14, 03 2016.
- 67 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. *The Journal of Physical Chemistry B*, v. 116, n. 29, p. 8692–8702, 2012.
- 68 CONTESSOTO, V. G. *Estudo do efeito da adição de frustração no enovelamento de proteínas utilizando modelos baseados em estruturas*. Dissertação (Mestrado em Biofísica Molecular) — UNESP, São José do Rio Preto, 2012.
- 69 ZARRINE-AFSAR, A. et al. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proceedings of the National Academy of Sciences*, v. 105, n. 29, p. 9999–10004, 2008.
- 70 CHAN, H. S. et al. Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. *Annual Review of Physical Chemistry*, v. 62, p. 301–326, 2011.
- 71 TREPTOW, W. L. et al. Non-native interactions, effective contact order, and protein folding: A mutational investigation with the energetically frustrated hydrophobic model. *PROTEINS: Structure, Function, and Genetics*, v. 47, p. 167–180, 2002.
- 72 MIYAZAWA, S.; JERNIGAN, R. L. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *PROTEINS: Structure, Function, and Genetics*, v. 34, p. 49–68, 1999.
- 73 ZHANG, Z.; CHAN, H. S. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proceedings of the National Academy of Sciences*, v. 107, n. 7, p. 2920–2925, 2010.
- 74 ALDER, B. J.; WAINWRIGHT, T. E. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, v. 27, n. 5, p. 1208–1209, 1957.
- 75 KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct Biol.*, Nature Publishing Group, v. 9, n. 9, p. 646–652, 2002.
- 76 MYERS, R. J. One-hundred years of pH. *J. Chem. Educ.*, American Chemical Society (ACS), v. 87, n. 1, p. 30–32, jan 2010.

-
- 77 CONTESSOTO, V. G. et al. NTL9 folding at constant pH: The importance of electrostatic interaction and pH dependence. *J. Chem. Theory Comput.*, American Chemical Society (ACS), v. 12, n. 7, p. 3270–3277, 2016.
- 78 TRIPATHI, S.; GARCÍA, A. E.; MAKHATADZE, G. I. Alterations of nonconserved residues affect protein stability and folding dynamics through charge–charge interactions. *The Journal of Physical Chemistry B*, American Chemical Society (ACS), v. 119, n. 41, p. 13103–13112, 2015.
- 79 KHANDOGIN, J.; BROOKS, C. L. Constant pH molecular dynamics with proton tautomerism. *Biophysical Journal*, Elsevier BV, v. 89, n. 1, p. 141–157, 2005.
- 80 GOH, G. B. et al. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins*, Wiley-Blackwell, v. 82, n. 7, p. 1319–1331, 2014.
- 81 MONGAN, J.; CASE, D. A. Biomolecular simulations at constant ph. *Current Opinion in Structural Biology*, v. 15, n. 2, p. 157 – 163, 2005. Theory and simulation/Macromolecular assemblages.
- 82 PLOTKIN, S. S.; ONUCHIC, J. N. Investigation of routes and funnels in protein folding by free energy functional methods. *Proceedings of the National Academy of Sciences*, v. 97, n. 12, p. 6509–6514, 2000.
- 83 LAMMERT, H.; NOEL, J. K.; ONUCHIC, J. N. The dominant folding route minimizes backbone distortion in sh3. *PLoS Comput Biol*, Public Library of Science, v. 8, n. 11, p. 1–12, 11 2012.
- 84 BRYNGELSON, J. D. et al. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 21, n. 3, p. 167–195, 1995.
- 85 KUSUNOKI, H. et al. Independent movement, dimerization and stability of tandem repeats of chicken brain α -spectrin. *Journal of Molecular Biology*, v. 344, n. 2, p. 495 – 511, 2004.
- 86 SCOTT, K. et al. The folding pathway of spectrin r17 from experiment and simulation: Using experimentally validated md simulations to characterize states hinted at by experiment. *Journal of Molecular Biology*, Academic Press Inc., v. 359, n. 1, p. 159–173, 5 2006.
- 87 ALLEN, M.; TILDESLEY, D. *Computer simulation of liquids Oxford Science*. [S.l.]: London, 1990.
- 88 SCHERER, C. *Métodos Computacionais da Física*. [S.l.]: Editora Livraria da Física, 2005.
- 89 KUMAR, S. et al. THE weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, Wiley-Blackwell, v. 13, n. 8, p. 1011–1021, oct 1992.

90 KUMAR, S. et al. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, Wiley-Blackwell, v. 16, n. 11, p. 1339–1350, 1995.

Apêndice A

Dinâmica Molecular

Dinâmica Molecular é um método de simulação computacional de resolver 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 vão variando seus estados, caracterizados pelas posições e velocidades, calculam-se as variáveis relevantes do sistema, tais como, temperatura e energia, em função da posição [87, 88].

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 [87], 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 da 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.

Potencial 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 de Miyazawa e Jernigan[72].

Apêndice C

Método dos Múltiplos Histogramas - WHAM

O método dos múltiplos histogramas (WHAM) é um método autoconsistente que tem como foco levantar as informações sobre o sistema. Para o estudo do enovelamento de proteína, simulações realizadas em uma única temperatura, não pode representar um conjunto de estados conformacionais, por isso são realizadas simulações em várias temperaturas, para que a amostragem do sistema seja melhor. Portanto, com uma melhor amostragem é possível calcular a partir do método do WHAM as grandezas termodinâmicas, tais como: calor específico e energia livre na coordenada de reação Q . [89, 90].

Para um determinado conjunto N de simulações com C_n configurações para a temperatura $T_n = \frac{1}{k_B \beta_n}$, a densidade de estado $\Omega(Q)$ em função da fração dos contatos nativos Q , pode ser dado por:

$$\Omega(Q) = H_n(Q) \exp(\beta_n E - f_n) \quad (C.1)$$

onde os termos $H_n(Q)$ representa os histogramas em função da coordenada Q e f_n a energia livre adimensional.

Os valores de f_n são dados arbitrariamente, e ao longo do método são realizados interações que testam os valores até que se atinja a convergência esperada.

A probabilidade para que o sistema assuma uma determinada configuração Q em uma dada temperatura T , é dado por:

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

A partir da probabilidade calculada da equação C.2, pôde-se calcular a energia livre do sistema e as demais grandezas termodinâmicas:

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

Apêndice D

Resultados Complementares

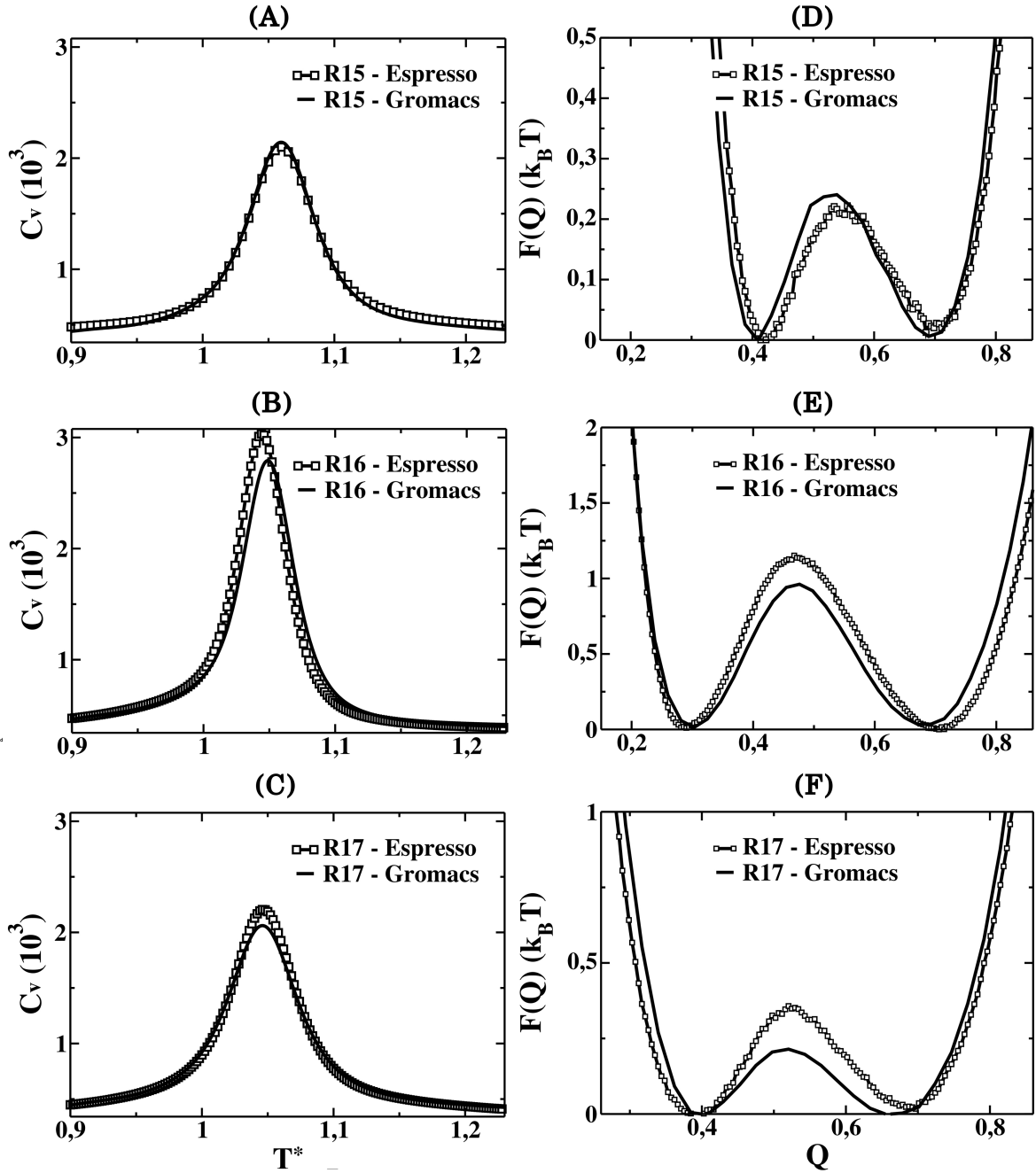
Nesta seção, serão comparados resultados que foram obtidos por meio de dois modelos de simulação: simulações realizadas no GROMACS e no ESPResSo. As simulações realizadas no ESPResSo, permite que os resíduos carregados protonem ou desprotonem e as simulações são iniciadas com aminoácidos carregados positivamente com carga +1 e, os aminoácidos com carga negativa, igual a zero, sendo que no decorrer da dinâmica sua carga pode variar, aminoácidos ácidos entre -1 e 0, e os aminoácidos básicos entre 0 e +1.

Na Figura D.1, as temperaturas de enovelamento, assim como os perfis de energia livre para os domínios R15, R16 e R17 são analisado. Para o domínio R15, Figura D.1A e D.1D, foram obtidos valores idênticos para as temperaturas de enovelamento, assim como para o perfis de energia livre, ou seja, os resultados do R15 não apresentam nenhuma dependência em relação ao método de simulação (GROMACS ou ESPResSo). Contudo, os domínios R16 (Figura D.1B e D.1E) e R17 (Figura D.1C e D.1F), apresentaram pequenos desvios em relação à barreira de energia livre. O R16 apresentou uma diferença na barreira de energia livre, $\Delta\Delta F(Q)$, de 0,2 $k_B T$, enquanto que para o domínio R17 essa diferença foi de 0,15 $k_B T$.

O principal motivo desta ampla diferença para os domínios R16 e R17, está relacionado aos parâmetros utilizados pelo GROMACS e ESPResSo. As simulações realizadas pelo GROMACS utiliza o integrador *Leap-Frog*, onde as velocidades são atualizadas em tempos intermediários aqueles em que são atualizadas as posições. Já o ESPResSo utiliza o integrador *Velocity-Verlet* (Apêndice A), onde as velocidades são atualizadas no mesmo instante que as posições.

Sabe-se que durante a dinâmica as energias cinéticas podem variar devido as colisões das partículas, bem como a temperatura do sistema. O GROMACS e o ESPResSo utilizam uma série de algoritmos para controlar e corrigir estes desvios, tais

Figura D.1: Comparação dos calores específicos e dos perfis de energia livre para as simulações realizadas no GROMACS e no ESPResSo. Simulações realizadas pelo ESPResSo permite a protonação/desprotonação ao longo da simulação, já para o GROMACS essa função não é possível. As Figuras A, B e C, apresentaram uma diferença muito pequena nos picos dos calores específicos e as Figuras D, E e F os perfis de energia livre, como as barreiras de energia livre, $\Delta F(Q)$.

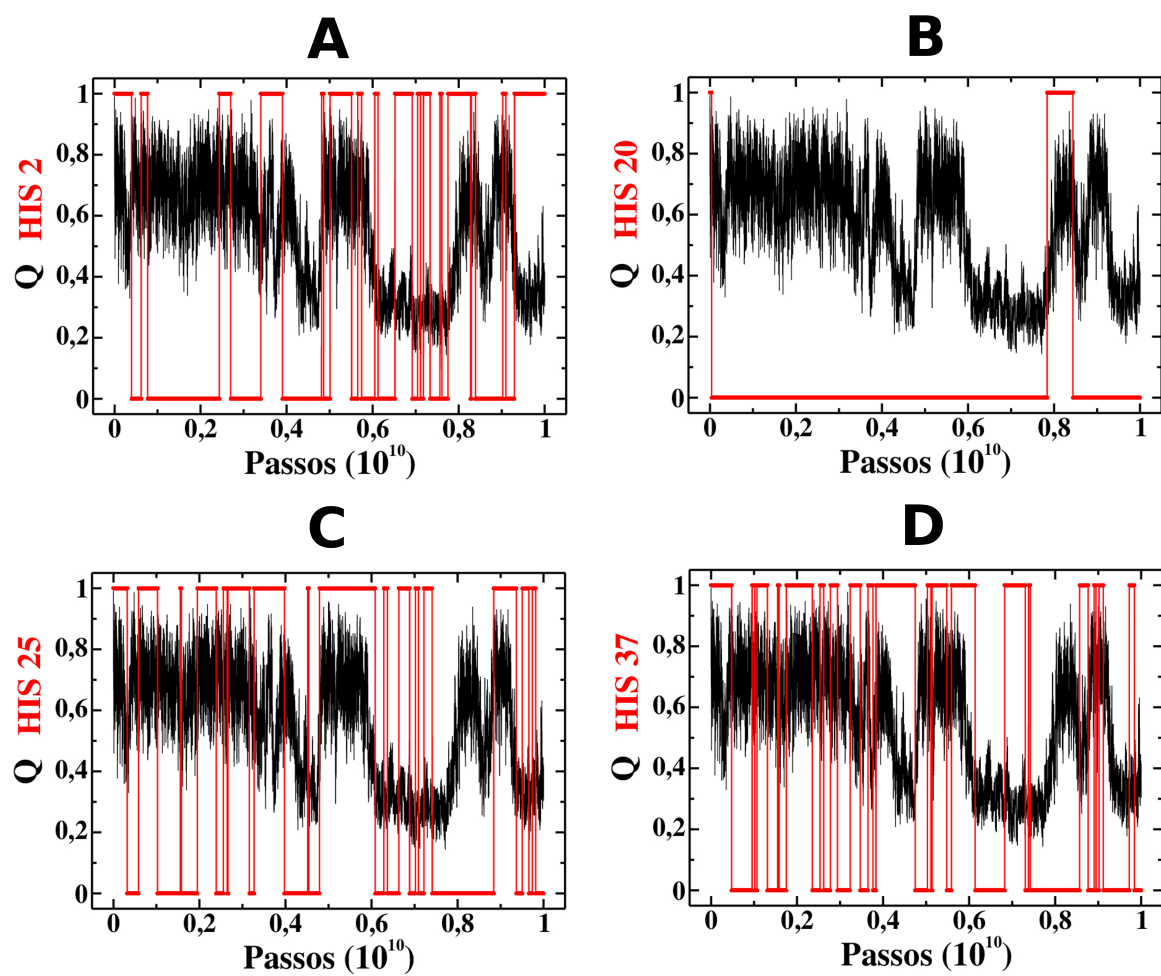


Fonte: Elaborada pelo autor.

algoritmos são conhecidos como termostato. As simulações realizadas pelo GROMACS utiliza o termostato de Berendsen, onde o reescalonamento é feito acoplando o sistema a um banho térmico com temperatura fixa. As velocidades a cada passo são renormalizadas de forma que a taxa de mudança de temperatura seja proporcional as diferenças de temperaturas do sistema e do banho utilizado. Já as simulações realizadas pelo ESPResSo utiliza o termostato de Langevin, onde a cada passo da simulação, todas as partículas recebem uma força aleatória e têm suas velocidades reduzidas por meio da aplicação de um atrito. A média das forças e do atrito estão relacionadas, o que garante que o teorema de flutuação–dissipação não seja violado, garantindo assim a identidade do *ensemble* NVT.

Outra diferença identificada foi para os resíduos tituláveis. As simulações realizadas no ESPResSo permite que os resíduos ácidos e básicos alterem seus estados. As simulações foram realizadas a $\text{pH} = 7.0$, logo a histidina foi o único aminoácido que apresentou mudanças no estado de protonação, pois esta apresenta o ponto isoelétrico, pI , próximo de 7,0. Na Figura D.2, verifica-se que as histidinas do domínio R16 sofreram mudanças em seus estados ao longo da dinâmica, sendo que somente a histidina da posição 20, possui ao longo de toda a simulação carga zero, sendo que nas simulações realizadas no GROMACS, todas as histidinas possuem carga +1 durante toda a simulação.

Figura D.2: Trajetória de enovelamento e estados de protonação da histidina em função dos passos de simulação para o domínio R16..



Fonte: Elaborada pelo autor.

Apêndice E

Artigos: 1PGB–DQQ e α –Espectrina

No artigo da 1PGB–DQQ, o autor participou de forma integral da elaboração, das discussões e das análises, onde foram estudados os efeitos da concentração de sal e do pH para a estabilidade da proteína. Este artigo foi submetido a JCTC.

O artigo referente a α –Espectrina está em fase de elaboração.

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

Vinícius M. de Oliveira,^{†,¶} Vinícius G. Contessoto,^{‡,¶} Fernando B. da Silva,[†]
Daniel L. Z. Caetano,[†] Sidney J. de Carvalho,[†] and Vitor B. P. Leite^{*,†}

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

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

[¶]*These authors contributed equally to this work*

E-mail: vleit@sjrp.unesp.br

Phone: +55 (17) 3221-2240

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 the present 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 on account of 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 remarkable 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. The present study provides new information and a comprehensive description of the electrostatic contribution to protein G stability.

1 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,⁴⁻⁷ indicating the importance of electrostatic interactions. Indeed, studies have shown the thermal stability of proteins can be enhanced by optimizing charge-charge interactions.⁸⁻¹⁰ 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 model proteins and for evaluating the electrostatic contributions of unfolding free energy.¹¹⁻¹⁴ 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.^{8,15,16} Recently, in order to study pH effects on the folding dynamics of the N-terminal domain of ribosomal protein L9 (NTL9),¹⁷ the

constant-pH molecular dynamic method (CpHMD)¹⁷⁻¹⁹ 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),²⁰ which enable the individual electrostatic interaction of each ionizable residue to be calculated.²¹ Makhatadze and colleagues 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.²²⁻²⁵ In the present 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.1A). Protein G is a multi-domain protein present in the cell wall of *G. streptococcus* containing the immunoglobulin G binding domains denoted as B1 and B2.²⁶⁻³⁰ These proteins differ by six amino acids and do not have disulfide bonds.^{31,32} Also, experimental data showed that the protein folds/unfolds as a two-state model without kinetic intermediates.³² PGB1-QDD has 56 amino acids (including the NH₂-terminal Met) and a regular α/β structure. The fold consists of a four β -sheet and one α -helix bundled against the sheet. The charged residues are almost entirely surface exposed³³ and display distinct pH-dependence protein stability.³¹ Another important factor is that PGB1-QDD displays the highest stability at low salt concentrations.⁴ Due to its dependence on salt concentration and pH, it is an excellent model for studying electrostatic interactions.

Salt concentration and pH dependence in 1PGB-QDD stability were tested using the TKSA and CpHMD methods, and the results are in remarkable qualitative agreement with the experiments. The contribution of each ionizable residue to this protein stability was also calculated using the TKSA. From CpHMD simulations, it was possible to evaluate the dependence of the free energy unfolding barrier and thermal denaturation on pH. Electrostatic interaction influence at each folding stage was also analyzed. Finally, the pK_a values were calculated for the folded protein (using CpHMD and TKSA) and the unfolded protein

(using only CpHMD). The computational pK_a values are in excellent agreement with the experimental values, considering the simplicity of the models. Thus, these coarse-grained models were found to be a computationally fast and cheap alternative to explore and provide useful insight into biological systems.

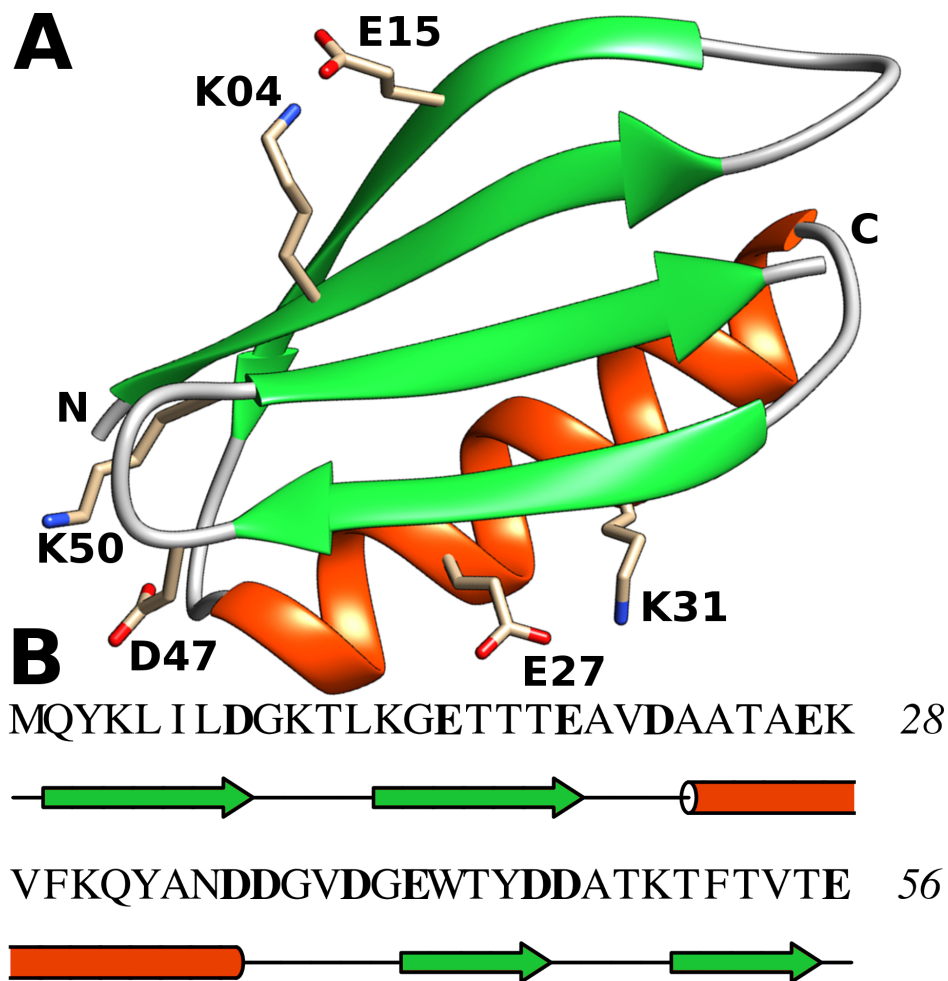


Figure 1: (A) Structure of PGB1-QDD built from wild-type protein with PDB ID:1PGB. The side-chains are shown for residues which have major contributions to protein stability. (B) Primary sequence of PGB1-QDD with highlighted acid residues, which have the highest charge variations in the studied pH range, and provide a useful insight into biological systems.

2 Methods

2.1 The Tanford-Kirkwood Model with Solvent Accessibility - TKSA

In the model proposed by Tanford and Kirkwood (TK), the protein is modeled by a spherical cavity with dielectric constant, ϵ_p and radius b , surrounded by an electrolyte solution treated according to the Debye-Hückel theory.²⁰ A solvent static accessibility parameter for each charged residue was incorporated by Shire *et al* into the TK model.³⁴ This modification was introduced to overcome the uncertainty of an adjustable charge-burial parameter beneath the dielectric interface and to allow for the irregular protein-solvent interface. With this change, the model proposed by Tanford and Kirkwood is now referred to as the Tanford-Kirkwood model with a solvent accessibility (TKSA).³⁴⁻³⁶ The interaction energy, U_{ij} , between two charged residues is given by

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

where e is the elementary charge; b is the radius of the protein; a is the closest possible approach distance of an ion; A_{ij} , B_{ij} and C_{ij} are parameters obtained from the analytical solution of the Tanford and Kirkwood model, which are functions of the distance between charged residues, the dielectric constants, and the salt concentration;^{20,37,38} and SA_{ij} is the average of the solvent accessible surface area of residues i and j .^{34,35,39}

Once the electrostatic energy of the ionizable residues of the protein is determined, it is necessary to sample all the possible protonation states.^{21,40} For a protein with n ionizable residues, there are $\chi = 2^n$ different protonation states. The free energy, $\Delta G_N(\chi)$, for a protein in its native state and for given state of protonation χ is given by

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

where R is the ideal gas constant; T is the temperature; q_i is the charge of the ionizable

residue i in its deprotonated state; x_i is 0 or 1, depending on the protonation state of the residue i ; and $pK_{int,i}$ is the intrinsic pK_a of the i th ionizable residue.

The probability that the protein is in its native state with a particular state of protonation, $\rho_N(\chi)$, is

$$\rho_N(\chi) = \frac{1}{Z_N} \exp \left[-\frac{\Delta G_N(\chi)}{RT} - \nu(\chi)(\ln 10)pH \right], \quad (3)$$

where $\nu(\chi)$ is the number of ionizable residues that are protonated in the state of protonation χ , and Z_N is the partition function for the protein in its native state.

The average total electrostatic energy, $\langle W_{qq} \rangle$ is given by the average over all protonation states, taking into account Eq. 3:

$$\langle W_{qq} \rangle = \sum_{\chi} \left[\frac{1}{2} \sum_{i,j=1}^n U_{ij}(q_i + x_i)(q_j + x_j) \right] \rho_N(\chi). \quad (4)$$

Ibarra-Molero and Makhatadze have shown that the contribution of electrostatic interaction energy to free energy, ΔG_{qq} , can be described by the negative of the average total electrostatic energy of the native state, that is, $\Delta G_{qq} \approx -\langle W_{qq} \rangle$.²¹

2.2 Constant-pH Molecular Dynamics Simulations

In the Structure-Based Model (SBM), the residues of a protein are represented by individual beads centered in C_α position.^{41–45} The energy of the protein is given by a Hamiltonian equation in which only the native interactions based on its native topology are taken into account.⁴¹ The first simplification of this model does not take into account the charged residues in the protein; the electrostatic interactions are therefore inserted by adding charged points at beads, which represent acid/basic residues.^{8,15,17,46} The interaction between charged residues is given by a Coulomb screening potential. The energy in a given configuration Γ

with regard to configuration Γ_o , the native structure, is given by:

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

where r_o represents the distance between two subsequent residues, θ_o and ϕ_o are the angles formed by three and four subsequent residues, respectively. All of them are taken from the native structure. Factor d_{ij} represents the distance between pairs of residues i and j ($i < j - 3$), $\epsilon_C = 1.0$ kcal/mol, $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$, $\epsilon_\phi = \epsilon_{NC} = \epsilon_C$, $\sigma_{NC} = 4\text{\AA}$ ⁴¹ and $K_{electrostatic} = 332 \text{ kcal}\text{\AA}/(\text{mol } e^2)$.^{8,16,17} The charges of residues i and j are q_i and q_j respectively, κ is the inverse of Debye length⁴⁷ and the dielectric constant is $\epsilon_K = 80$.

2.2.1 Protonation/Deprotonation Criterion

The constant-pH molecular dynamics method (CpHMD) was utilized in this study. It associates a standard molecular dynamics (MD) simulation with the Metropolis Monte Carlo method (MC) for sampling protein protonation states.^{17,19,48–50} During the simulation, the molecular dynamic is regularly halted, a MC step is taken, and a random titratable residue is considered to change its protonation. The transition energy, ΔE , for this step is appraised according to Equation 6, in which the positive and negative signs mean protonation and deprotonation respectively. The acidic or basic residues are represented by ξ , where acidic residues present $\xi = -1$ and basic residues $\xi = +1$.

$$\Delta E = \pm \xi k_B T \ln(10)(pH - pK_a) + \Delta E_{elec} \tag{6}$$

In Equation 6, k_B is the Boltzmann constant, T is the temperature, pK_a is the value of the reference compound (reference pKa values are 3.6 for C-terminal, 4.0 for Aspartic acid, 4.5 for Glutamic acid, 6.3 for Histidine, 7.5 for N-terminal, 10.6 for Lysine and 12.0 for Arginine). The first term is the contribution due to the presence of a proton bath at a given concentration determined by pH; the second one is the contribution of the compound model; and the last one, ΔE_{elec} , is the electrostatic energy variation for the protonation state change of the titratable residue. The protonation/deprotonation is determined by the Metropolis criterion.⁵¹ Thus, if the MC step is accepted, the protonation/deprotonation of the residue will change to the new state, and the MD will continue. However, if the new state is not accepted, there is no need for updating, and the residue continues with its unchanged protonation state.

2.3 Simulation Details

The CpHMD-SBM algorithm was implemented using Espresso Package version 3.2 in NVT ensemble and Langevin thermostat.^{52,53} In the folding simulations, the protein was initialized in its native state with all titratable residues protonated. An equilibration process was executed with 10^7 steps and 10^9 steps to generate the trajectory with a time step of 0.5 fs. The data was stored every 1000 MD steps, and the Monte Carlo titration procedure was carried out every 2000 steps.¹⁷ The free energy and specific heat profiles were obtained using the weighted histogram method (WHAM),⁵⁴⁻⁵⁶ calculated from different simulations at 25 temperatures for pH ranging from 1.0 to 12.0 for high and low salt concentrations. The reaction coordinate used to describe the folding was defined as the sum of native contacts (Q) in a structure Γ and a time t . A native contact was considered formed when the distance between the residue i and j ($i < j - 3$) was shorter than $1.2 d_{ij}$ from the native structure. The native contact map was determined by contact of structural units software (CSU).⁵⁷ The mutant protein PGB1-QDD was built from PGB1 protein with PDB ID:1PGB using Modeller software version 9.17.⁵⁸ The protein 3D structure graphic in Figure 1A was created

with the UCSF Chimera package⁵⁹ and the protein sequence in Figure 1B was generated with Aline software.⁶⁰ The TKSA method was implemented in Python 2.7, and the solvent accessible surface area (SASA) was calculated with Surface Racer© software.⁶¹ The TKSA method was performed for each of the 19 charged residues in PGB1-QDD and was simulated for pH from 1.0 to 12.0 at intervals of 0.5.

3 Results and Discussion

3.1 Electrostatic energy calculations – TKSA

In this section the contribution of electrostatic energy to protein free energy was calculated with the TKSA method. The sum of electrostatic energy of each charged residue is here referred to as ΔG_{elec} . The ΔG_{elec} of PGB1-QDD was calculated in a range of pH from 1.0 to 12.0, and the stability pH-dependent profile is presented in Figure 2.

The data presented in Figure 2 of ΔG_{elec} as a function of pH have similar behavior with slight differences for high and low salt concentrations in simulations via the TKSA method. The TKSA simulation results indicate that high levels of salt concentration do not eliminate the electrostatic interaction of surface charges and the pH dependence of protein native state stability is still strong.⁴ The peak of TKSA curves indicates that the protein PGB1-QDD native state is more stable around its isoelectric point, which is close to pH 4.5 for both high and low salt concentrations. The TKSA result has a level of good agreement with experimental data⁴ for low salt concentrations shown inset in Figure 2, and the correlation value between TKSA and experimental data is 0.83. On the other hand, for high salt concentrations, there is not a strong correlation between the TKSA and experimental results and the correlation value is 0.44. The weak correlation is due to the differences between the results for lower pH values. The behavior of experimental data for lower pH values does not present a significant stability difference from 2M of NaCl⁴ shown inset in Figure 2. The correlation between TKSA and experimental data became stronger if only values of pH

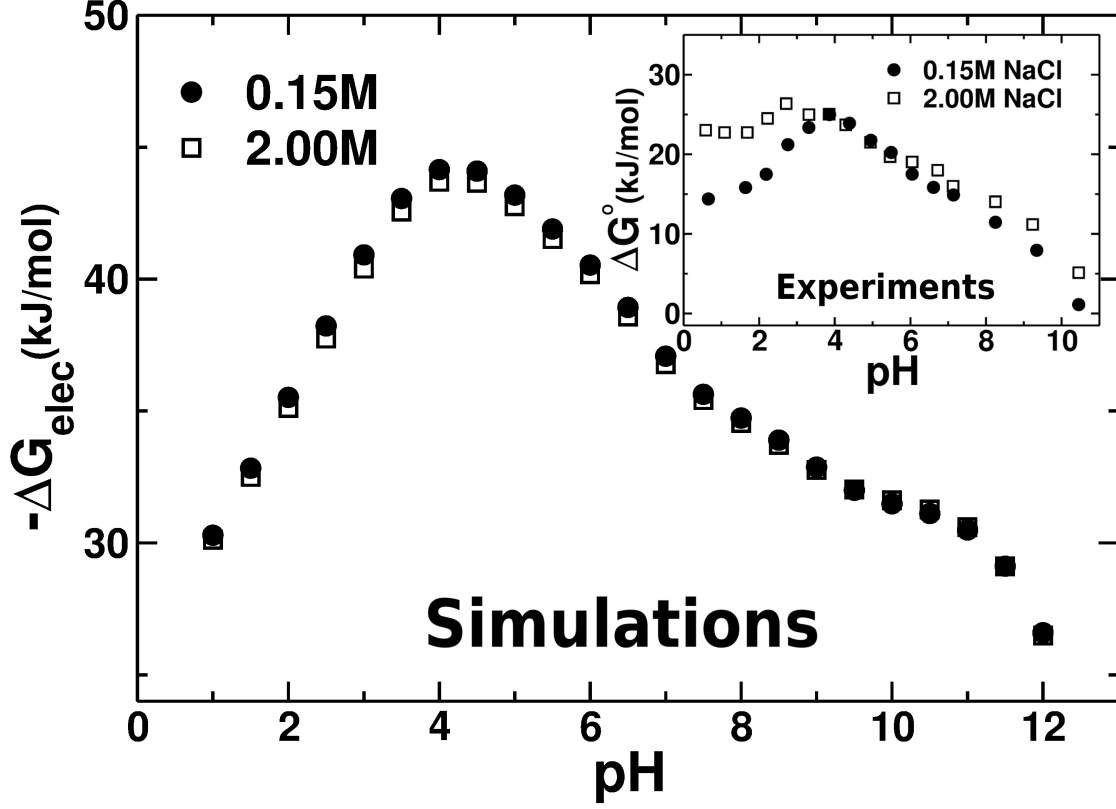


Figure 2: Electrostatic energy contribution to free energy native state stability ΔG_{elec} as a function of pH in kJ/mol. The values were calculated from TKSA simulation analysis for 23 different pH values from 1.0 to 12.0. The black circles represent values for 0.15M of a monovalent salt concentration in simulation and the white filled squares represent values of 2.0M of salt concentration. The inset plot shows the experimental results of the free energy difference between the unfolded and the native state ΔG° from 1PGB-QDD as a function of pH in kJ/mol. The black circles represent values for 0.15M of NaCl, and the white filled squares for 2.00M of NaCl (adapted from ref.⁴)

higher than 3.0 are taken into account and the new correlation value is then 0.95.

Another result obtained from TKSA simulation is the contribution of each charged residue to protein stability.^{21,25,62} Figure 3 shows the TKSA analysis for pH values 2.5, 4.5, 7.5 and 10.0 for low salt concentrations. For pH 2.5 in Figure 3A there are eight residues with $\Delta G_{qq} < 0.0$ which contribute favorably to protein stability. The other eleven residues have energy values close to zero with $\Delta G_{qq} \approx 0.0$. The total electrostatic free energy ΔG_{elec} for pH 2.5 is $-38.22 \text{ kJ.mol}^{-1}$ and is calculated as the sum of ΔG_{qq} of all charged residues.^{21,22,62} The major contribution to stability comes from six residues: K04, E15, E27, K31, D47,

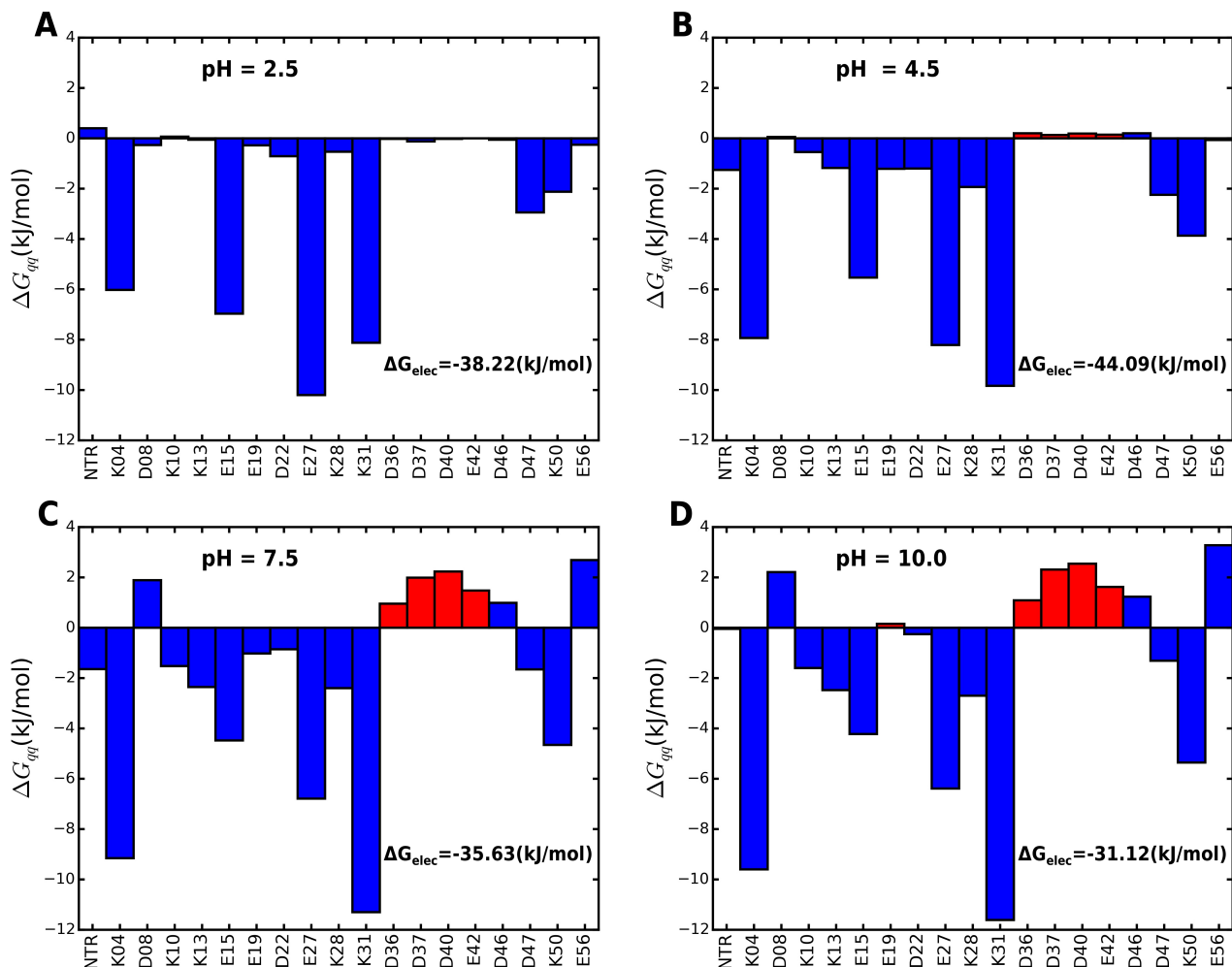


Figure 3: Charge-charge interaction energy ΔG_{qq} calculated by the TKSA model for each ionizable residue. A, B, C, and D represent the energy profiles of pH values 2.5, 4.5, 7.5 and 10.0, respectively. The red bars indicate the residues with the side chain exposed to solvent with $\text{SASA} \geq 50\%$ and positive energy contribution to native state stability.

and K50. These six residues make an important energetic contribution to all pH values and they are responsible for $86\% \pm 6.2\%$ of all favorable interaction for pH values 2.5, 4.5, 7.5, and 10.0 shown in Figure 3. The side chains of these residues are not fully exposed to solvent, with $\text{SASA} \leq 55\%$. However, the short distance between their ionizable groups (less than $2.95\text{\AA} \pm 0.56\text{\AA}$) promotes the high energetic contribution of these groups to protein stability (the side chains of K04, E15, E27, K31, D47, and K50 are highlighted in Figure 1A). At pH 4.5 in Figure 3B, similar to pH 2.5 analysis, the major electrostatic energy contribution comes from the six residues mentioned above. At pH 4.5

there are twelve residues with $\Delta G_{qq} < 0.0$ and the other seven residues have energy values close to zero, with $\Delta G_{qq} \approx 0.0$. The total electrostatic free energy ΔG_{elec} for pH 4.5 is -44.09 kJ.mol⁻¹, which is the pH which maximizes protein stability in experiments⁴ and also in the TKSA simulation shown in Figure 2. For pH 7.5 shown in Figure 3C there are also 12 residues with $\Delta G_{qq} < 0.0$ similar to pH 4.5. There are seven residues with $\Delta G_{qq} > 0.0$ contributing with unfavorable electrostatic interactions to native state stability. These seven residues: D08, D36, D37, D40, E42, D46, and E56 are all acidic residues in their deprotonated state with charge -1. The protonation/deprotonation events are detailed in the pK_a comparison section. The positive ΔG_{qq} values of these seven residues make the greatest contribution to decreasing the native state stability of PGB1-QDD, and the total electrostatic free energy ΔG_{elec} for pH 7.5 is -35.63 kJ.mol⁻¹. In Figure 3C the red bars highlight the residues whose side chains are exposed to solvent more than 50% with SASA = 73% $\pm$ 14% for residues D36, D37, D40, and E42. The positive ΔG_{qq} energy contribution and the side chain exposed more than 50% to solvent indicate that mutations in these residues should lead the protein to a more stable native state based on optimization of the electrostatic interactions on the protein surface.^{21-25,62-64} The TKSA result for pH 10.0 is shown in Figure 3D with 9 residues $\Delta G_{qq} < 0.0$, 3 residues with ΔG_{qq} close to zero and 7 residues with $\Delta G_{qq} > 0.0$. The discussion of the favorable and unfavorable electrostatic energy contribution of each residue in pH 10.0 is similar to what has been discussed for pH 7.5. There is an increase of ΔG_{qq} for acid residues. The positive energy contribution passes from 12.2 kJ.mol⁻¹ in pH 7.5 to 14.5 kJ.mol⁻¹ in pH 10.0. The residue E19 had a significant variation from negative value $\Delta G_{qq} = -1.07$ kJ.mol⁻¹ to a slightly positive contribution $\Delta G_{qq} = 0.15$ kJ.mol⁻¹. The N-terminal also had a significant variation due to the fact that it becomes deprotonated with charge 0 in pH 10.0 and its energy contribution is close to zero. The total electrostatic free energy ΔG_{elec} for pH 10.0 is -31.12 kJ.mol⁻¹

3.2 pH and ionic strength effects on 1PGB-QDD folding – CpHMD-SBM- C_α

The previous section indicates that the drive interaction responsible for PGB1-QDD stability is electrostatic. However, the TKSA model takes into account only the static native structure of the protein in its calculations, while experimental data suggests that the unfolded state of PGB1-QDD is also important to its stability.⁴ Thus, in this section, CpHMD-SBM simulations were utilized, which can explore the effects of the unfolded and transition states on protein stability. 1PGB-QDD was therefore simulated at four different pHs (2.5, 4.5, 7.5 and 10.0) and three ionic strengths (Low salt ≈ 0.01 M, intermediate salt ≈ 0.1 M and high salt ≈ 0.4 M).

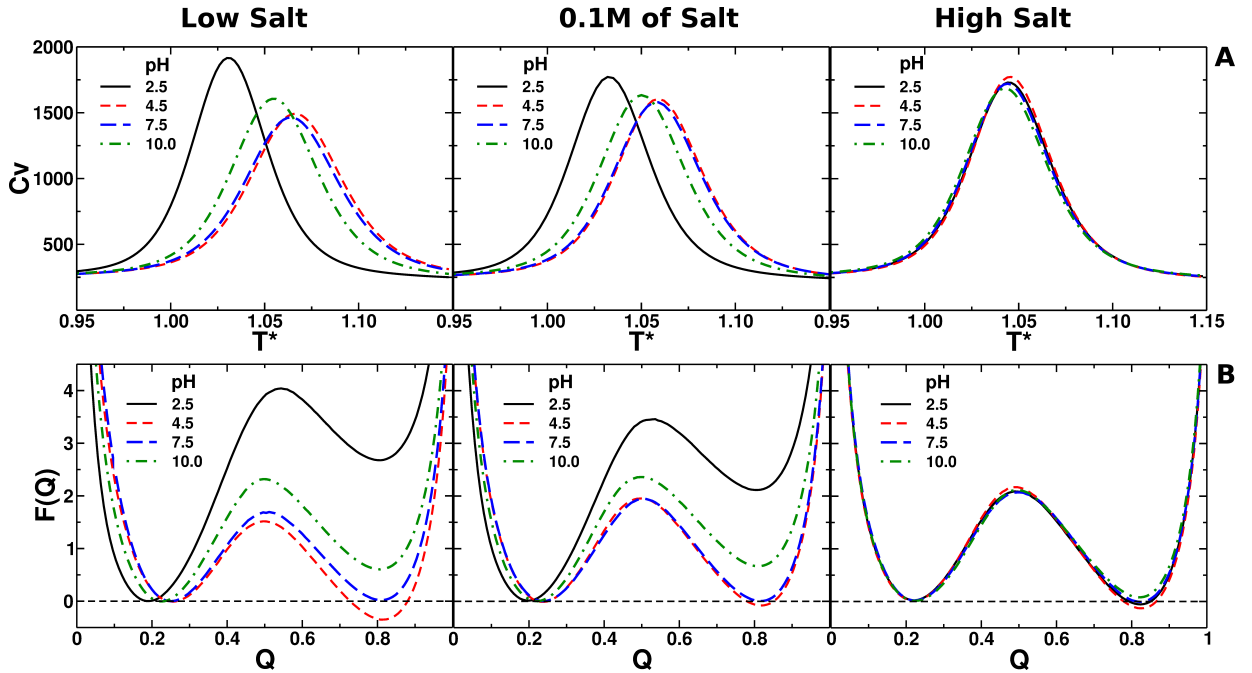


Figure 4: Thermodynamic properties of 1PGB-QDD folding. **A:** Heat capacities at constant volume (C_v) in low, intermediate, and high salt concentrations. Solid black lines for pH 2.5, dashed red lines for pH 4.5, dashed blue lines for pH 7.5, and dashed/dotted green lines for pH 10.0. **B:** Free energy curves as a function of the reaction coordinate Q for the same three salt concentrations. All systems at T close to T_M of pH 7.5 in each ionic strength.

Heat capacity (C_v) and free energy profiles of PGB1-QDD were calculated to explore the role of pH and ionic strength on protein folding thermodynamics (Figure 4). In these

computational results, electrostatic screening at the high salt concentration is too strong, making the pH effect on protein stability almost insignificant. Even so, it is possible to see a slight increase in 1PGB-QDD stability at pH 4.5. As expected, the pH effect becomes more evident in the protein at low and intermediate salt concentrations, but the higher stability remains at pH 4.5. At the low salt concentration, the difference between the C_v peaks of each pH curve was maximized. The transition barrier between the folded and unfolded state in protein near T_M was not affected by salting, with the barrier close to $2.0 \epsilon_C$. In general, the thermodynamic properties calculated by CpHMD-SBM are in good agreement with the experimental results.

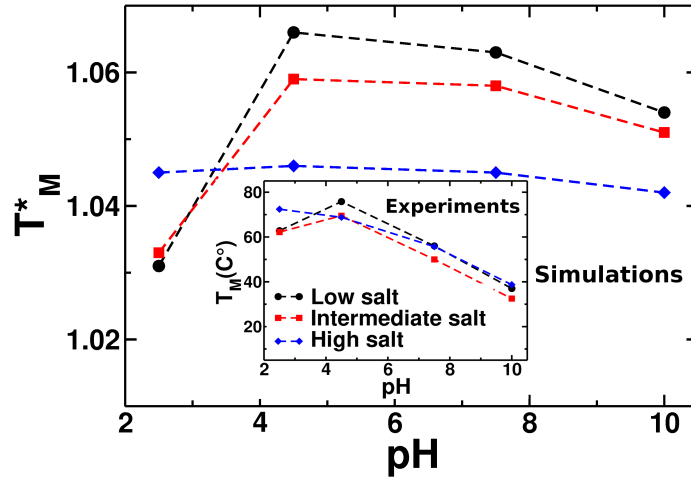


Figure 5: Values of melting temperature T_M^* in reduced temperature as a function of pH. Black circles for low salt, red squares for an intermediate ionic strength, and blue diamonds for high salt concentration. Inset graphic presents the experimental results of T_M as a function of pH in similar salt conditions to simulation (adapted from ref.⁴). Dashed lines connecting symbols help guide the eye.

The comparison between experimental and computational T_M is presented in Figure 5. The general behavior T_M for computational results in each salt concentration shows a good level of agreement with the experimental data, with the T_M peak at pH 4.5 for all systems. At low pH, there is a decrease of T_M mostly in low and intermediate salt concentrations. In this case, both experiments and simulations agree that a high ionic strength contributes to 1PGB-QDD stability. However, when the protein has a pH higher than 4.5, the highest

stability in the simulations occurs at low salt concentration, followed by intermediate and high salt concentrations. On the other hand, experimental results at pH 7.5 and 10.0 indicate that the electrostatic screening caused by a high salt concentration helps to stabilize 1PGB-QDD. This difference between experimental and computational results is caused by the electrostatic energy calculated in the simulation, and this interaction will be explored below.

The effect of electrostatic interaction and charge regulation on each folding stage of 1PGB-QDD can be evaluated by an analysis of Figure 6. This figure presents a two dimensional map of the distribution of electrostatic energy as a function of Q . The net charge q in each highly populated region of distribution of electrostatic energy was also calculated and included in the Figure 6.

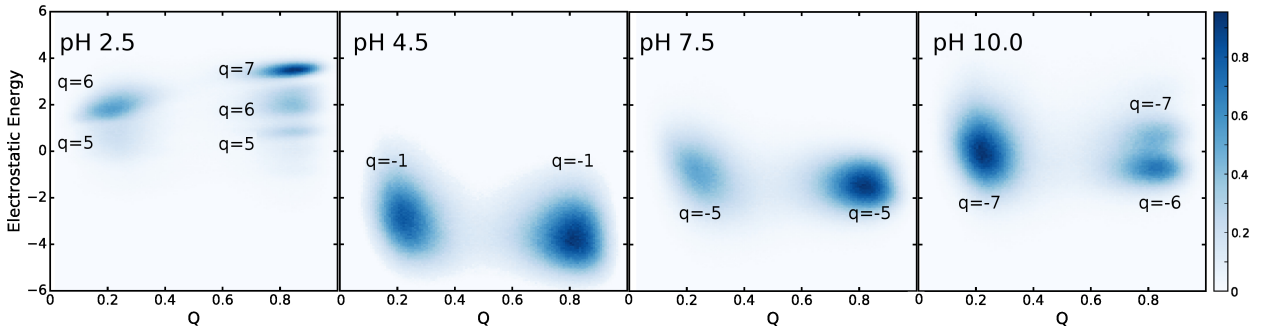


Figure 6: Two-dimensional map of the distribution of electrostatic energy as a function of the reaction coordinate Q for pH 2.5, 4.5, 7.5 and 10.0. The color map represents the probability distribution of the electrostatic energy and Q , normalized by its highest value. All the systems at $T \approx T_M$ and intermediate salt concentration. q represents the net charge of 1PGB-QDD in each minimum.

Previous results showed that the highest stability of 1PGB-QDD occurs at pH 4.5. This is because the protein in this condition has the lowest mean electrostatic energy $\langle E_{elec} \rangle$ compared to the protein at pH 2.5, 7.5 and 10.0. For both folded and unfolded states, the $\langle E_{elec} \rangle$ is near $-4 \epsilon_C$ and net charge $q = -1$, not favoring any state in comparison to each other. A similar discussion applies to the protein at pH 7.5: the folded and unfolded states have $\langle E_{elec} \rangle$ close to $-2 \epsilon_C$ and $q = -5$ (experimental net charge⁴ is -6). An increase in mean electrostatic energy promotes a loss of stability and is caused by this negative q , although, even with a high negative net charge, $\langle E_{elec} \rangle$ is kept negative. This favorable

energy explains the fact that low salt concentration promotes more stability in 1PGB-QDD at pH 7.5 than intermediate and high salt concentrations (Figure 5), which does not agree with the experimental results. Although the net charge is similar (experimental $q = -6$ and computational $q = -5$), the arrangement of charges in the C_α model can be different from the more realistic case, and creates more regions of favorable interaction. At pH 10.0 the same thing occurs: the protein has a high negative $q = -7$ or $q = -6$ and a negative value of $\langle E_{elec} \rangle$ close to $-1 \epsilon_C$.

Figure 6 also shows that the pH effect at pH 4.5 and 7.5 is strong enough to keep the protein net charge, irrespective of its conformational state. However, at the extreme pHs 2.5 and 10.0, some differences in electrostatic energy and net charge arise, depending on the folding stages. In these cases, the charge regulation is more evident, and the interaction between charged groups becomes fundamental for their protonation states. With the protein at pH 2.5, the 2D map presents two populated regions in the unfolded state, one with $\langle E_{elec} \rangle$ close to $1 \epsilon_C$ ($q = 5$) and other near $2.0 \epsilon_C$ ($q=6$), the second being the most populated. In the folded state, in this same pH condition, three populated regions arise, the most populated with the highest $\langle E_{elec} \rangle = 4 \epsilon_C$ and $q = 7$, an intermediate region in $\langle E_{elec} \rangle = 2\epsilon_C$ and $q = 6$, and the lowest populated region with $\langle E_{elec} \rangle = 1\epsilon_C$ and $q = 5$. The loss of stability at this low pH condition may be associated with the fact that $\langle E_{elec} \rangle$ of the unfolded protein is lower than $\langle E_{elec} \rangle$ in folded state. Thus, the electrostatic interaction favors the unfolded state.

3.3 Comparison between computational and experimental pK_a

In this final section, the experimental pK_a and the pK_a calculated by TKSA (for folded state), and CpHMD (for folded and unfolded states) were compared. The computational pK_a was calculated fitting individual titration curves with the Henderson-Hasselbalch equation¹⁷ and the results are presented in Table 1. The errors in pK_a values were calculated using a set of three different simulations and an error lower than 0.1 pK_a unit was obtained for each

ionizable residue. Thus, these error values were omitted from Table 1.

Table 1: Comparison between experimental⁶⁵ and computational pK_a values for the acidic residues in folded and unfolded 1PGB-QDD protein.

Residue	Experimental ^a		CpHMD ^b		TKSA ^c
	pK_a Folded	pK_a Unfolded	pK_a Folded	pK_a Unfolded	pK_a Folded
D08	4.9	3.7	3.5	3.5	4.2
E15	4.6	4.4	4.0	4.1	1.8
E19	3.9	4.5	4.0	4.4	4.0
D22	3.0	3.9	3.7	3.9	3.4
E27	4.8	4.4	3.5	3.7	0.6
D36	4.2	4.1	4.1	4.2	4.1
D37	6.5	4.2	4.1	4.3	4.0
D40	4.1	4.2	4.0	4.1	4.3
E42	4.8	4.4	4.6	4.7	5.0
D46	3.8	4.0	4.0	4.1	4.0
D47	3.1	3.9	3.6	3.8	2.9
E56	3.8	4.7	4.6	4.4	6.5
C-ter	3.1	3.6	**	**	3.5

^aThe experimental pK_a values for the acidic residues measured by Lindman et al.;⁶⁵

^bThe computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation, simulation using CpHMD.

^cThe computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation, simulation using TKSA.

Only differences higher than 1.0 pK_a unit between experimental and computational values were considered significant.¹⁷ Comparing the experimental pK_a of folded 1PGB-QDD protein and the values calculated using TKSA, only four residues presented a difference higher than 1.0 pK_a unit. This is a good level of agreement considering the simplicity of the model, which can output the results in less than one hour. E15 and E27 have a downshift in computational pK_a that may be caused by the overrated interaction calculated in this model

that these residues make with K04 and K21, respectively. D37 has a large upshift in its experimental pK_a , which may be caused by desolvation effects associated with side-chain burial⁶⁵ that TKSA is not taking into account sufficiently. Finally, the residue E56 has two charged groups that cause a large electrostatic repulsion in the TKSA model; this repulsion may explain the high upshift in its computational pK_a value.

Using CpHMD, the comparison with the experimental results is even better, with only three residues of folded 1PGB-QDD presenting more than 1.0 pK_a unit of difference between the simulation and the experimental pK_a values. D27 has a similar downshift in pK_a values to that obtained using TKSA; the reason for this difference is the high interaction that this residue makes with K04. In the case of D08, the downshift of 1.4 pK_a units obtained in computational results could be caused by a model limitation. Considering the native structure of the protein, the distance between the C_α of D08 and C_α K13 is near 5.1Å, whereas the distance between the carboxyl group of D08 and ammonium group of K13 is close to 8.3Å. This difference in the distances promotes a more favorable interaction between these residues in CpHMD, resulting in this disagreement between experimental and simulation results. In residue D37, as previously stated, desolvation effects have a strong impact on its experimental pK_a value and the CpHMD model does not treat this effect adequately. When the pK_a of the unfolded protein was considered, the agreement between experimental values and CpHMD values is remarkable: in no case does the difference between these values exceed 1.0 pK_a unit.

4 Conclusions

In the present study, the protein 1PGB-QDD was studied by two coarse-grained models with the objective of evaluating the pH and salting effect on protein stability. The results of this study were compared with experimental results. The first model tested here was TKSA, which analyses the effect of each ionizable residue in the electrostatic free energy

of the protein in its native state. Using this simple model, it was possible to quantify the importance of each charged residue in 1PGB-QDD stability. The comparison between TKSA calculations and experimental results also indicates that the driving force of this stability is the electrostatic interaction. The second model used in this study was CpHMD-SBM, which was able to explore the protein folding process in constant pH.¹⁷ With CpHMD-SBM simulations, the folded and unfolded states and the transition between these states could be studied. The pH effect in 1PGB-QDD stability explored by simulation was found to be in good agreement with the experimental data. Similarly, the salt effect was evaluated, and the simulations obtained the same general behavior in protein stability as the experiments. Finally, pK_a was calculated using TKSA in the folded protein and CpHMD-SBM for folded and unfolded 1PGB-QDD. The pK_a values calculated by the two models are near experimental values. Therefore, both methods were able to describe the effects of charge-charge interaction, pH and ionic strength on protein, despite the high level of simplification.

5 Acknowledgments

VMO was supported by the National Council for Scientific and Technological Development (CNPq - Grant Process No. 141985/2013-5). VGC was funded by Grant 2016/13998-8, São Paulo Research Foundation (FAPESP) and Higher Education Personnel Improvement Coordination (CAPES). FBS was supported by the Higher Education Personnel Improvement Coordination (CAPES). DLZC acknowledges the São Paulo Research Foundation (FAPESP, Grant Process No. 2013/13151-7) for the financial support. VBPL was supported by the National Council for Scientific and Technological Development (CNPq) and FAPESP Grant 2014/06862-7 and 2016/19766-1. We also thank the Center for Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) for computational resources.

References

- (1) Roca, M.; Liu, H.; Messer, B.; Warshel, A. *Biochemistry* **2007**, *46*, 15076–15088.
- (2) Bye, J.; Platts, L.; Falconer, R. *Biotech. Lett.* **2014**, *36*, 869–875.
- (3) Huber, K.; Olek, K.; Müller, A.; Tan, C.; Bennett, K.; Colinge, J.; Superti-Furga, G. *Nature Methods* **2015**, *12*, 1055–1057.
- (4) Lindman, S.; Xue, W.-F.; Szczepankiewicz, O.; Bauer, M. C.; Nilsson, H.; Linse, S. *Biophys. J.* **2006**, *90*, 2911–2921.
- (5) Yamasaki, M.; Yano, H. *Int. J. Biol. Macr.* **1990**, *12*, 263–268.
- (6) Kim, K.; Kim, S.; Yang, H.; Kwon, D. *Int. J. Food Sci. Tech.* **2004**, *39*, 385–393.
- (7) Bye, J.; Falconer, R. *Prot. Sci.* **2013**, *22*, 1563–1570.
- (8) Tripathi, S.; García, A. E.; Makhatadze, G. I. *J. Phys. Chem. B* **2015**, *119*, 13103–13112.
- (9) Wunderlich, M.; Martin†, A.; Schmid, F. *J. Mol. Biol.* **2005**,
- (10) Papaleoa, E.; Olufsena, M.; De Gioiab, L.; Brandsdala, B. *J. Mol. Graph. Mod.* **2007**, *26*, 93–103.
- (11) Date, M. S.; Dominy, B. N. *Commun. Comput. Phys.* **2013**, *13*, 90–106.
- (12) Luitz, M. P.; Bomblies, R.; Ramcke, E.; Itzen, A.; Zacharias, M. *Sci Rep.* **2016**, *6*, 19896.
- (13) Su, J.-G.; Han, X.-M.; Zhao, S.-X.; Hou, Y.-X.; Li, X.-Y.; Qi, L.-S.; Wang, J.-H. *J. Mol. Mod.* **2016**, *22*, 91.
- (14) Coronado, M.; Caruso, I.; De Oliveira, V.; Contessoto, V.; Leite, V.; Arni, R.; Eberle, R. *Protein Pept. Lett.* **2017**,

- (15) Azia, A.; Levy, Y. *J. Mol. Biol.* **2009**, *393*, 527–542.
- (16) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087–1097.
- (17) Contessoto, V. G.; de Oliveira, V. M.; de Carvalho, S. J.; Oliveira, L. C.; Leite, V. B. P. *J. Chem. Theory Comput.* **2016**, *12*, 3270–3277.
- (18) Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89*, 141–157.
- (19) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163.
- (20) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.
- (21) Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I.; Sanchez-Ruiz, J. M. *Biochemistry* **1999**, *38*, 8138–8149.
- (22) Makhatadze, G. I.; Loladze, V. V.; Ermolenko, D. N.; Chen, X.; Thomas, S. T. *J. Mol. Biol.* **2003**, *327*, 1135–1148.
- (23) Permyakov, S. E.; Makhatadze, G. I.; Owenius, R.; Uversky, V. N.; Brooks, C. L.; Permyakov, E. A.; Berliner, L. J. *Protein Eng., Des. Sel.* **2005**, *18*, 425–433.
- (24) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. *Biochemistry* **2006**, *45*, 2761–2766.
- (25) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2601–2606.
- (26) Björck, L.; Kastern, W.; Lindahl, G.; Widebäck, K. *Mol. Immunol.* **1987**, *24*, 1113–1122.
- (27) Sjöbring, U.; Björck, L.; Kastern, W. *Mol. Immunol.* **1989**, *3*, 319–327.
- (28) Björck, L.; Kronvall, G. *J. Immunol.* **1984**, *133*, 969–974.
- (29) Reis, K. J.; Ayoub, E. M.; Boyle, M. D. *J. Immunol.* **1984**, *132*, 3091–3097.

- (30) Kmiecik, S.; Kolinski, A. *Biophys. J.* **2008**, *94*, 726–736.
- (31) Alexander, P.; Fahnestock, S.; Lee, T.; Orban, J.; Bryan, P. *Biochemistry* **1992**, *31*, 3597–3603.
- (32) Alexander, P.; Orban, J.; Bryan, P. *Biochemistry* **1992**, *31*, 7243–7248.
- (33) Gronenborn, A.; Filpula, D.; Essig, N.; Achari, A.; Whitlow, M.; Wingfield, P.; Clore, G. *Science* **1991**, *253*, 657–661.
- (34) Shire, S.; Hanania, G.; Gurd, F. N. *Biochemistry* **1974**, *13*, 2967–2974.
- (35) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192–2198.
- (36) Orttung, W. H. *Biochemistry* **1970**, *9*, 2394–2402.
- (37) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351–361.
- (38) Da Silva, F. L. B.; Jönsson, B.; Penfold, R. *Protein Sci.* **2001**, *10*, 1415–1425.
- (39) Havranek, J. J.; Harbury, P. B. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11145–11150.
- (40) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556–9561.
- (41) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937–953.
- (42) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res.* **2010**,
- (43) Contessoto, V. G.; Lima, D. T.; Oliveira, R. J.; Bruni, A. T.; Chahine, J.; Leite, V. B. P. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 1727–1737.
- (44) de Mendonca, M. R.; Rizzi, L. G.; Contessoto, V.; Leite, V. B. P.; Alves, N. A. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 119–129.
- (45) Mouro, P. R.; de Godoi Contessoto, V.; Chahine, J.; Junio de Oliveira, R.; Pereira Leite, V. B. *Biophys. J.* **2016**, *111*, 287–293.

- (46) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E259–E266.
- (47) Ullner, M.; Woodward, C. E.; Jönsson, B. *J. Chem. Phys.* **1996**, *105*, 2056–2065.
- (48) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (49) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (50) Williams, S. L.; Blachly, P. G.; McCammon, J. A. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3381–3388.
- (51) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (52) Arnold, A.; Lenz, O.; Kesselheim, S.; Weeber, R.; Fahrenberger, F.; Roehm, D.; Košovan, P.; Holm, C. In *Meshfree Methods for Partial Differential Equations VI*; Griebel, M., Schweitzer, M. A., Eds.; Lecture Notes in Computational Science and Engineering 89; Springer Berlin Heidelberg, 2013; pp 1–23.
- (53) Limbach, H. J.; Arnold, A.; Mann, B. A.; Holm, C. *Comput. Phys. Commun.* **2006**, *174*, 704–727.
- (54) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635.
- (55) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (56) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
- (57) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. *Bioinformatics* **1999**, *15*, 327.
- (58) Webb, B.; Sali, A. *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2002.

- (59) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (60) Bond, C. S.; Schüttelkopf, A. W. *Acta Crystallogr.* **2009**, *65*, 510–512.
- (61) Tsodikov,; Record, M.; Sergeev, Y. *J. Comput. Chem.* **2002**, *23*, 600–609.
- (62) Ibarra-Molero, B.; Sanchez-Ruiz, J. M. *J. Phys. Chem. B* **2002**, *106*, 6609–6613.
- (63) Loladze, V. V.; Makhatadze, G. I. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3494–3499.
- (64) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. *J. Mol. Biol.* **2004**, *336*, 929–942.
- (65) Lindman, S.; Bauer, M. C.; Lund, M.; Diehl, C.; Mulder, F. A.; Akke, M.; Linse, S. *Biophys. J.* **2010**, *99*, 3365–3373.

Results:

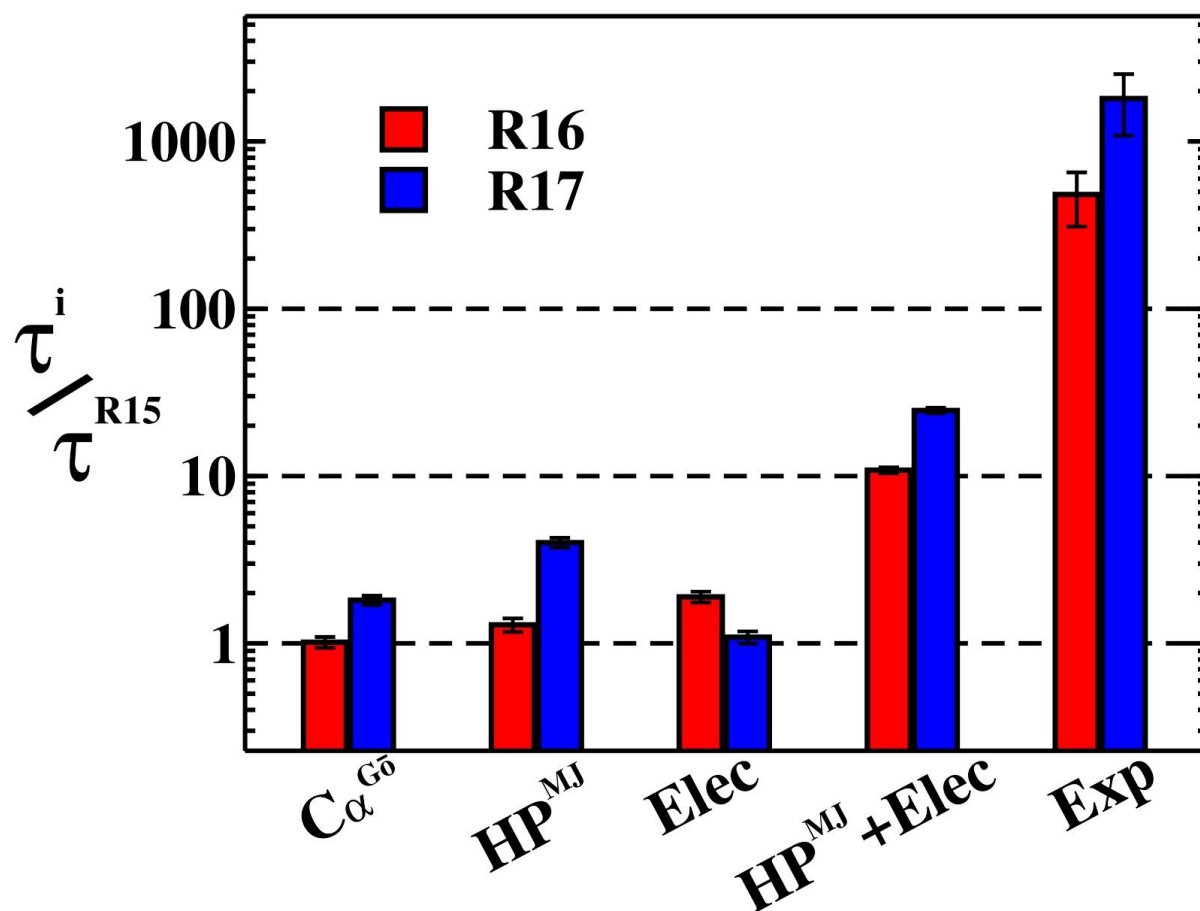


Figure 1. The ratio folding time of wild-type spectrin domains in different computational methods and compared with experimental data (Exp). The bar graphs in red and blue represent the R16 and R17, respectively and the dotted line equal 1 represent the ratio folding time for the R15. The C_{α}^{Go} model show a small effect on the folding time in the spectrin domains. The HP^{MJ} model (nonnative hydrophobic interaction) has a more significant effect on the R17 and display a small effect on the R16. The bar graphs in the Elec model (electrostatic interaction) gives the information that electrostatic interaction has a more effect on the R16 than R17 domain. The folding time data from $HP^{MJ} + Elec$ model show for R16 and R17 one order of magnitude slower than R15 domain. The experimental data, the last bar graphs, show that R16 has two orders and R17 has three orders of magnitude slower than R15. Experimental data of wild-type is take from Scott et. al. (1).

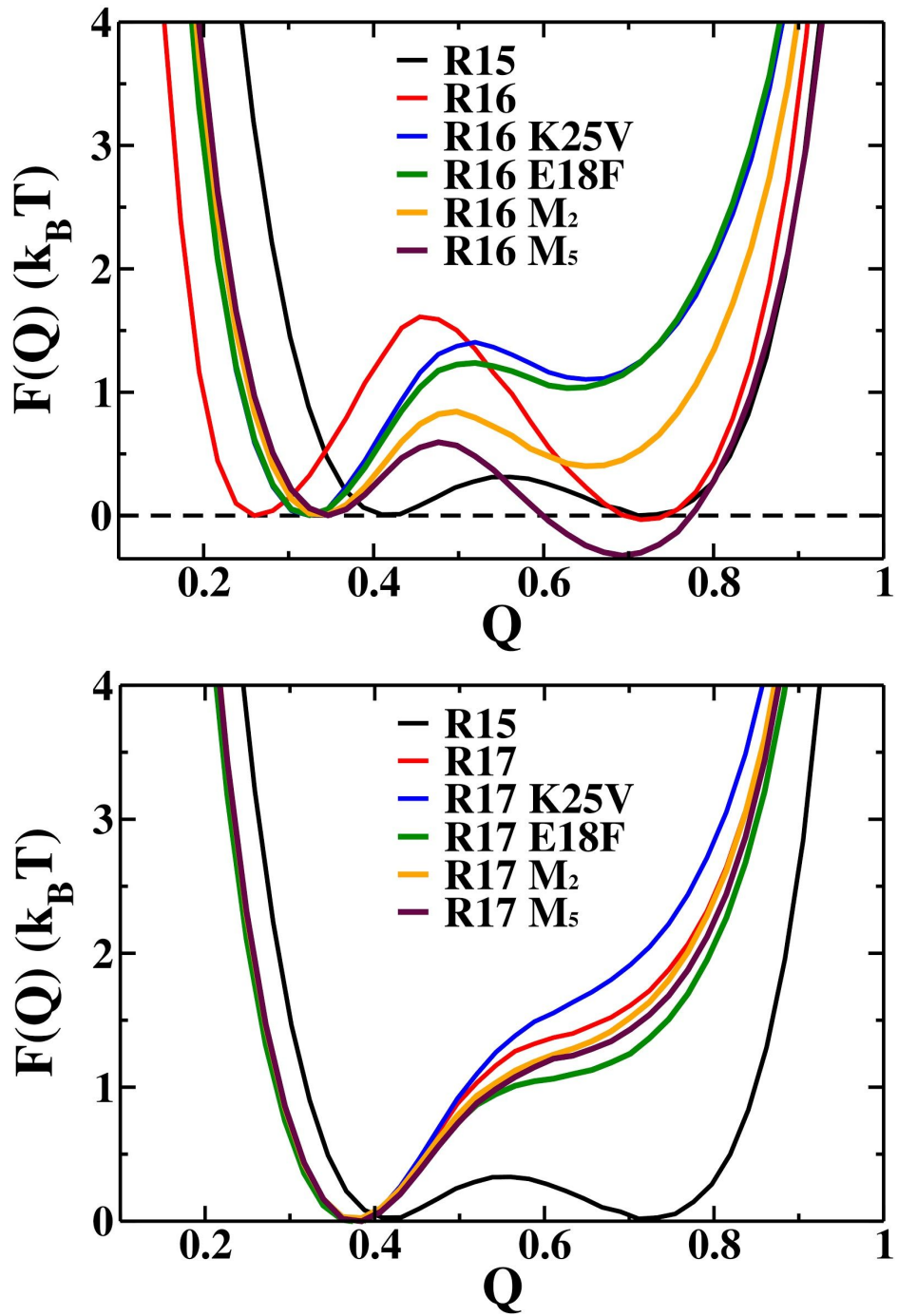


Figure 2. Folding thermodynamics of R15, R16, and R17 for the wild-type and mutations. The free energy curves $F(Q)$ for the spectrin domains are plotted as a function of a fraction of native contacts, Q . The M_2 and M_5 it is means a double mutations (E18F / K25V) and five sequential mutation (E18F / E19D / I22L / K25V / V29L), respectively. All the plots are made at the corresponding T_f of R15 wild-type.

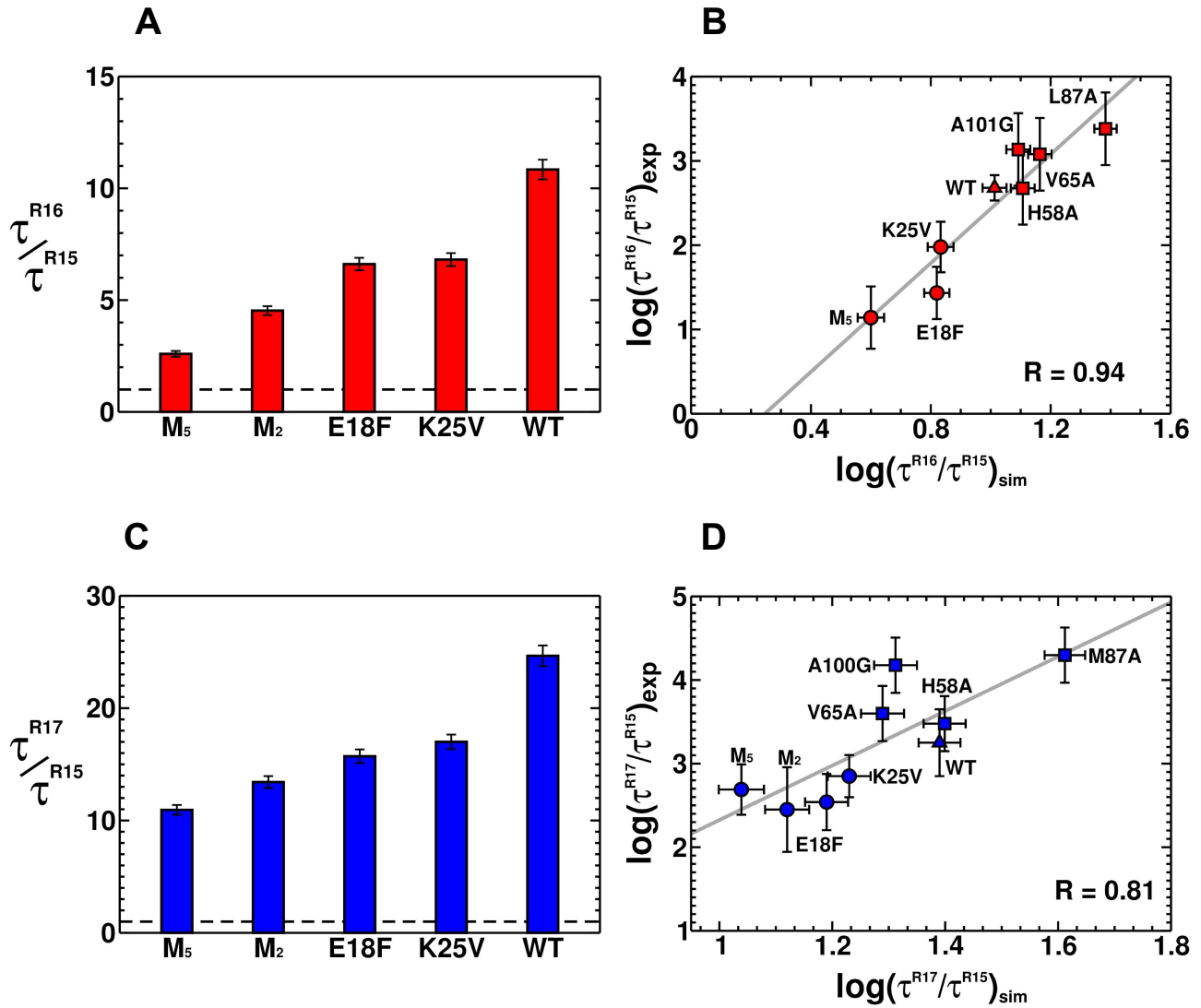


Figure 3. The bar graphs **A** and **C** display the ratio folding time for the R16 and R17 (wild-type and mutations) and the dotted line equal 1 represent the ratio of R15. The ratio folding time decrease as the mutations are made, K25V, E18F, M_2 and M_5 , respectively. The figures **B** and **D** show the relationship between the ratio of the folding time of R16 and R15 and R17 and R15, respectively. The coordinate axes “y” are represented the data for the experimental values and the coordinate axes “x” are represented the data for the computational results. The data are compared in a plot with a best-fit line; R is correlation coefficient. Therefore, the R16 domain show a correlation coefficient of 0.94 and R17 show a correlation coefficient of 0.84. The experimental values for a better mutation (E18F, K25V, M_2 and M_5) is take from Wensley et. al. (2) and for a worse mutation is take from SCOTT et. al. (3) for the R16 (H58A, V65A, A101G and L87A) and it is take from SCOTT et. al. (4) for the R17 (H58A, V65A, A100G and M87A).

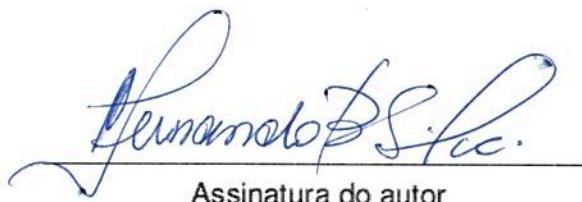
Bibliography

1. SCOTT, K. A. et al. The folding of spectrin domains I: Wild-type domains have the same stability but very different kinetic properties. *Journal of Molecular Biology*, v. 344, n. 1, p. 195 – 205, 2004.
2. WENSLEY, B. G. et al. Separating the effects of internal friction and transition state energy to explain the slow, frustrated folding of spectrin domains. *Proceedings of the National Academy of Sciences*, *Proceedings of the National Academy of Sciences*, v. 109, n. 44, p. 17795–17799, 2012.
3. SCOTT, K. A. et. al. The folding of spectrin domains II: Phi-value analysis of R16. *Journal of Molecular Biology*, v. 344, n. 1, p. 207 – 221, 2004.
4. 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, *Journal of Molecular Biology*, v. 359, n. 1, p. 159-173, 2006.

TERMO DE REPRODUÇÃO XEROGRÁFICA

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

São José do Rio Preto, 03, 04, 2017


Assinatura do autor