

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS

HELOISA DOS SANTOS MUNIZ

**Estudo computacional da difusão térmica  
em proteínas termoestáveis**

São Carlos

2013



HELOISA DOS SANTOS MUNIZ

**Estudo computacional da difusão térmica  
em proteínas termoestáveis**

Dissertação apresentada ao programa de Pós-graduação em Física do Instituto de Física de São Carlos de Universidade de São Paulo, para a obtenção do título de Mestre em ciências.

Área de concentração: Física Aplicada  
Opção: Física Biomolecular  
Orientado: Prof. Dr. Leandro Martínez

Versão Corrigida  
(versão original disponível na unidade que aloja o programa)

São Carlos  
2013

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,  
com os dados fornecidos pelo(a) autor(a)

dos Santos Muniz, Heloisa  
Estudo computacional da difusão térmica em  
proteínas termoestáveis / Heloisa dos Santos Muniz;  
orientador Leandro Martínez - versão corrigida -- São  
Carlos, 2013.  
113 p.

Dissertação (Mestrado - Programa de Pós-Graduação em  
Física Aplicada Biomolecular) -- Instituto de Física  
de São Carlos, Universidade de São Paulo, 2013.

1. Difusão de energia vibracional. 2. Dinâmica  
molecular. 3. Estabilidade térmica. I. Martínez,  
Leandro, orient. II. Título.





# AGRADECIMENTOS

Agradecimentos ao meu orientador, Leandro Martínez, pela oportunidade, por todo conhecimento passado e por tornar tudo mais fácil.

Aos meus colegas de grupo, Luciano Censoni e Mariana Bunoro, por estarem compartilhando o mesmo momento e também por eventuais discussões.

Ao meu pai, Ronaldo Muniz, pelo incentivo e apoio desde o começo.

À minha mãe, M. Cristina N. dos Santos, pela força psicológica e por acreditar sempre.

À toda minha família, que mesmo não entendendo muito bem o que faço me apoiam e se esforçam para lembrar o nome da minha graduação.

Por último, ao Wagner R. Correr, por me ajudar de todas formas possíveis e impossíveis, e é claro, pelo apoio moral e até científico.



*It is not enough to have knowledge, one must also apply it. It is not enough to have wishes, one must also accomplish.*

*–Johann Wolfgang von Goethe*



# RESUMO

MUNIZ, H.S. **Estudo computacional da difusão térmica em proteínas termoestáveis.** 2013. 113p. Dissertação (Mestrado em ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Mecanismos de difusão de energia vibracional em biomoléculas têm sido relacionadas a função, alosterismo e sinalização intramolecular. Neste trabalho nós utilizamos uma metodologia computacional para analisar o fluxo de energia em proteínas. Simulações de Dinâmica Molecular são utilizadas para o estudo de difusão térmica, provendo artifícios que não são possíveis experimentalmente: a proteína é esfriada a baixas temperaturas e apenas um resíduo é aquecido através do acoplamento de um banho térmico. Consequentemente, o calor flui do aminoácido aquecido para a proteína, revelando os caminhos da difusão da energia vibracional. Pelo fato de que proteínas termoestáveis possam ter particulares mecanismos de relaxação, distribuição e dissipação da energia vibracional, elas são sistemas interessantes para serem utilizadas por este método. Um padrão de difusão de calor de uma proteína termofílica pode ser identificado e comparado com outro de uma proteína homóloga mesofílica. Aqui estudamos um conjunto de proteínas em particular, as pertencentes à família 11 de Xilanases. O mapa de difusão térmica obtido da proteína no vácuo mostrou diferenças entre xilanases mesofílica e termofílica, e termofílica e hipertermofílica: qualquer que seja o resíduo aquecido, aminoácidos específicos respondem com alta temperatura. Esta resposta em alta energia de certas regiões é decorrente de processos de relaxação. Simulações adicionais e outras análises, como da mobilidade de cada resíduo, levam à hipótese que estas regiões de superfície possuem grande flexibilidade e uma importante interação estrutural com a água. Mapas de difusão térmica para duas proteínas homólogas, diferindo em apenas 7 mutações, sendo 6 delas no N-terminal, apresentam-se diferentes nesta região. Em especial, a mutação Ser35Glu se destaca tanto no mapa quanto em outras medidas realizadas, apresentando-se na proteína mais estável com um maior nível de mobilidade, solvatação e energia de interação. Simulações em água não resultaram em padrões de difusão diferentes, por não apresentarem processos de relaxação. Entretanto, elas evidenciaram as mesmas regiões frias para cinco xilanases da família 11, especialmente o núcleo e a região de ligação do substrato, sugerindo uma possível característica funcional de difusão de calor. Por fim, evidenciado pelos mapas, observou-se que a região do cordão de xilanases termoestáveis, em especial a hipertermoestável, é maior se comparado à proteína mesofílica. Desta forma, através de comparações entre mapas de difusão e estruturas de proteínas similares, esta metodologia pode sugerir novas abordagens em engenharia racional de proteínas com estabilidade modulada.

Palavras-chave: Difusão de energia vibracional. Dinâmica molecular. Estabilidade térmica.



# ABSTRACT

MUNIZ, H.S **Computational study of thermal diffusion in thermostable proteins.** 2013. 113p. Dissertação (Mestrado em ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.

Vibrational energy dissipation in biomolecules have been related to function, in particular allosterism and intramolecular signaling. In this work, we use a computational method to analyze the energy flux through proteins. Molecular Dynamics simulations are used to study thermal diffusion in protein structure, in an artificial way which is not accessible experimentally: the protein is cooled down to very low temperatures and a single residue is heated by coupling a thermal bath to it. Heat flows from the heated residue to the rest of the protein, revealing the paths of vibrational energy dissipation. Since thermostable proteins may have specific mechanisms for vibrational energy relaxation, dissipation and distribution, they are interesting subjects for the application of the present methodologies. The heat dissipation patterns of thermophilic proteins can be compared to the ones of less stable structures. Here, we focus in a specif set of proteins known as Xylanases of Family 11. The thermal diffusion maps obtained for hiperthermostable, thermostable and mesphilic xylanases in vacuum were different, some of them displaying apparent thermal responses whatever the heated residue. These high temperature regions appeared because of differential structural relaxation processes in each structure. The analysis of the mobility of the structures in equilibrium simulations revealed that these regions are mobile and belong to the surface of the proteins, thus interacting significantly with water molecules. Thermal diffusion maps for homologous proteins differing in only 7 residues, being 6 of them at the N-terminal region of the proteins, were different in this region. One particular mutation was determined to be more mobile in the less thermostable protein, as well as displaying a higher solvation and stronger interaction energies with remaining protein structure. Thermal diffusion simulations in water were not able to discern any difference between the structures, particularly because the relaxation processes observed in vacuum were suppressed. Nevertheless, these maps reveal that every Xylanase display the same cold regions, which were observed to belong the protein core and the catalytic site, suggesting that thermal diffusion may have some functional role. Finally, we observed that a loop which relaxed systematically in thermostable protein resulting in high temperatures is larger than in non-thermostable structures. The addition of the loop to non-thermostable proteins make the maps equivalent. Therefore, the comparison of the thermal diffusion maps of similar structures highlight important structural differences which may be useful for providing insights into the design of proteins with modulated thermal stability.

Keywords: Vibrational energy diffusion. Molecular dynamics. Thermal stability.



# LISTA DE FIGURAS

Figura 1 - Estrutura das xilanases da família 11. Em forma de mão direita, possui regiões com nomes próprios: A) polegar, B) palma, C) dedos, D) cordão e E) $\alpha$ -hélice. . . . .	30
Figura 2 - Possível unidade repetitiva do polímero de xilano, composta em sua cadeia principal por três moléculas de D-xilano e alguns possíveis sítios de ramificação representados por números. . . . .	31
Figura 3 - Alinhamento global da sequência de aminoácidos entre as cinco proteínas utilizadas nesse estudo, regiões conservadas apontadas pelo gráfico abaixo do alinhamento. Realizado no ClustalX. . . . .	35
Figura 4 - Alinhamento estrutural das 5 proteínas aqui estudadas. Apresentam a região catalítica bem conservada em relação à cadeia principal (esquerda), com o ligante da xilanase EnXyn11A da mesma família (código PDB 2VGD). As variações aparecem no comprimento de alguns loops e fitas- $\beta$ (direita). . . . .	35
Figura 5 - Esquema do método ATD retirado de (1). A estrutura inicialmente a 10 K recebe um estímulo térmico através de um aminoácido. A resposta de cada resíduo da estrutura é apresentada no MDT. Na segunda linha, realiza-se a mutação do resíduo aquecido pela glicina, por exemplo. A resposta total da proteína mutada é subtraída da nativa, computando assim a CCL. . . . .	42
Figura 6 - Mapa de contatos e MDT médio obtido de estruturas retiradas de 3 simulações de 40 ns distintas para a) 2VUJ e b) 2VUL. . . . .	48
Figura 7 - a) Mutações em destaque nas estruturas cristalográficas 2VUJ (azul) e 2VUL (vermelha) alinhadas. Resíduo 34 em destaque nas duas proteínas: b) glutamina em 2VUJ está próxima de outra glutamina, possibilitando ligações de hidrogênio, enquanto c) leucina em 2VUL é um resíduo hidrofóbico em um ambiente hidrofílico. . . . .	49
Figura 8 - ATD nas proteínas 2VUJ (acima) e 2VUL (abaixo). Os mapas de difusão térmica (MDT) são de <i>20 frames</i> para 3 simulações de equilíbrio distintas. . . . .	50
Figura 9 - Comparação linhas quentes do MDT para a) 1XNB, b) 2VUJ e c) 2VUL. . . . .	51
Figura 10 - Alinhamento entre 1XNB e a sequência das termoestáveis na região do cordão. Em vermelho está a sequência que foi adicionada. . . . .	52

Figura 11 - a) Estruturas alinhadas, 2VUJ, 2VUL e 1XNB (cinza). As proteínas mais estáveis possuem o loop maior, o que gerou uma linha quente no MDT, diferente de 1XNB; b) Novo MDT para a proteína mesofílica com a inserção do loop, sendo destacada a linha quente formada em relação à proteína nativa. . . . .	52
Figura 12 - ATD em vácuo para as cinco proteínas a) 1XNB, b) 1F5J, c) 1M4W, d) 2VUJ e e) 2VUL; mapa contatos para 2VUJ. . . . .	53
Figura 13 - MDT para 1F5J. A diferença entre os dois mapas é somente a origem dos <i>frames</i> : a) e b) são mapas médios de <i>frames</i> retirados de simulações de equilíbrio distintas. . . . .	54
Figura 14 - Mapas de difusão térmica para as cinco proteínas em água, a) 1XNB, b) 1F5J, c) 1M4W, d) 2VUJ e e) 2VUL. Em preto, as regiões mais frias (abaixo de 45 K) são semelhantes em todas proteínas. . . . .	56
Figura 15 - Estrutura da xilanase 2VUL com o ligante da xilanase 2VGD em vermelho, também da família 11; em destaque, em azul, as regiões que permaneceram frias independentemente do aminoácido aquecido no MDT em água. a) Representação “surf” no VMD e b) “new cartoon”, mostrando que a região fria encontra-se principalmente na palma da mão. . . . .	57
Figura 16 - MDT para a proteína a) 1F5J, b) 2VUJ e c) 2VUL no vácuo e em água, respectivamente. A escala é 45 K a 90 K. . . . .	58
Figura 17 - Gráfico de CCL para a proteína 2VUJ utilizando a mutação para glicina. À esquerda, a) a média temporal da temperatura, e à direita, b) a temperatura final. Na primeira linha estão os gráficos com a resposta da proteína nativa, na segunda, a resposta da proteína mutada na posição do aminoácido acoplado ao banho, e por último a diferença entre os dois, ou seja, a contribuição da cadeia lateral. Todos em função do aminoácido aquecido. . . . .	60
Figura 18 - CCL médio de <i>frames</i> da a) primeira ou b) segunda metade de uma simulação de equilíbrio de 40 ns para 2VUL. Ainda, c) a diferença entre as CCL de cada metade. . . . .	61
Figura 19 - Temperatura final em função do resíduo aquecido para a proteína 2VUL a) nativa e b) mutada para glicina; média de 10 <i>frames</i> de duas metades (20ns) de uma simulação de equilíbrio (gráficos da primeira e segunda linhas), e a diferença entre elas (terceira linha). c) RMSD da proteína com alinhamento nos carbonos- $\alpha$ em relação à estrutura média no tempo. . . . .	62

Figura 20 - CCL para ATD realizado em a) solvente e b) na ausência deste, para a proteína 2VUJ, utilizando mutação ALA. Na última linha o gráfico da CCL normalizado para seus máximos e mínimos de cada gráfico. . . . .	64
Figura 21 - CCL para 1XNB, 2VUJ e 2VUL, nesta ordem. Em amarelo, as mutações que distinguem as duas últimas. Os resíduos foram mutados para glicina. . . . .	66
Figura 22 - Histogramas para CCL de aminoácidos individuais utilizando a mutação GLY. Em ordem, a)-g) as sete mudanças de 2VUJ e 2VUL, h) o aminoácido 105 que é a região do cordão, onde as duas proteínas possuem uma linha vermelha no MDT. A indicação dos aminoácidos correspondentes à 2VUJ está à esquerda, e à 2VUL, à direita no número. Temperatura média no tempo. . . . .	68
Figura 23 - Histogramas para CCL aminoácidos individuais utilizando a mutação ALA. Em ordem, as sete mudanças de 2VUJ e 2VUL; temperatura média no tempo. . . . .	69
Figura 24 - A Arg 126, previamente apresentada com diferente CCL para as proteínas mesofílica (1XNB), termofílica (2VUJ) e hipertermofílica (2VUL) quando representado o CCL em função do resíduo aquecido. O resíduo equivalente em 1XNB sera a Arg 112. Foi utilizado a média da temperatura. . . . .	70
Figura 25 - Histogramas da CCL para aminoácidos individuais acoplados ao banho térmico. Mutação para Gly, média no tempo. . . . .	71
Figura 25 - Gráficos de relaxamento e à direita os MDTs, ambos em vácuo. a) 1XNB, b) 1XNB com a inserção no cordão, c) 1F5J, d) 2VUJ e e) 2VUL.	75
Figura 26 - Relaxamento em água para as três proteínas a) 1XNB, b) 2VUJ e c) 2VUL. Todos foram médias de 3 frames, enquanto que para a última foi feita a média em 20 frames, por isso o gráfico é mais suave. . . . .	76
Figura 27 - Relaxamento no vácuo para as estruturas anteriores à termalização de 10 K , para 2VUL. Média em 20 frames. . . . .	77
Figura 26 - Mutações aquecidas e a resposta (temperatura) de toda proteína em função do tempo. 2VUJ (esquerda) e 2VUL (direita) com os aminoácidos que receberam o estímulo térmico no ATD. . . . .	81
Figura 27 - Gráfico de ATD em água (esquerda) e no vácuo (direita), em função do tempo para um frame. Neste caso a proteína 1F5J teve os aminoácidos a) 86 e b) 156 aquecidos. . . . .	82

Figura 28 - RMSD para as cinco simulações em função do tempo de simulação para a) 2VUJ e b) 2VUL (as 5 simulações), c) 1F5J e d) 1XNB (estrutura nativa e a com o cordão maior). O alinhamento foi feito em relação ao C- $\alpha$ , tendo a estrutura média no tempo como referência. . . . .	84
Figura 29 - Estrutura de 2VUJ no tempo da simulação 2, representando o movimento do polegar, de acordo com a variação do RMSD. a) e b) são estruturas como polegar fechado, no tempo 20 ns, enquanto que c) e d) representam o polegar aberto em 30 ns. À esquerda: visualização de estruturas secundárias (modo <i>New Cartoon</i> no VMD). À direita: representação da superfície da estrutura (modo <i>Surf</i> no VMD). . . . .	85
Figura 30 - RMSD em função do aminoácido para as simulações de equilíbrio de 2VUJ e 2VUL. . . . .	86
Figura 30 - Quantidade de águas de solvatação em uma esfera de raio igual a 4 Å do aminoácido. Gráficos a)-g) individuais, e h) a soma entre todas as mutações. . . . .	89
Figura 31 - Gráficos das energias eletrostática e de Van der Waals como função do tempo de simulação para as termoestáveis 2VUJ e 2VUL. . . . .	91
Figura 32 - Energia variável de 2VUL é devido às variações das interações do ácido glutâmico 35 com asparagina (a) e arginina (b). . . . .	92

# LISTA DE TABELAS

Tabela 1 - Mutações que diferenciam as estruturas de 2VUJ e 2VUL. . . . .	34
Tabela 2 - Xilanases e suas características. . . . .	37
Tabela 3 - Número de frames iniciais utilizado para cada proteína em função do experimento de ATD, para os sistemas em água ou em vácuo. . . . .	43
Tabela 4 - Quantidade de ligações de hidrogênio média no tempo para cada simulação de 2VUJ e 2VUL e o desvio padrão. . . . .	95



# LISTA DE ABREVIATURAS

ATD	<i>Anisotropic Thermal Diffusion</i>
MDT	Mapa de Difusão Térmica
CCL	Contribuição da Cadeia Lateral
PDB	<i>Protein Data Bank</i>
VDW	Van Der Waals
$T_m$	Temperatura de <i>melting</i>
$t_{1/2}$	Tempo de meia-vida
$T_{CO}$	Temperatura de crescimento ótimo
$T_{AO}$	Temperatura de atividade ótima
VMD	<i>Visual Molecular Dynamics</i>
DM	Dinâmica molecular
RMSD	<i>Root mean square deviation</i>



# SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>23</b>
1.1	Dinâmica Molecular . . . . .	23
1.2	Termoestabilidade em proteínas . . . . .	25
1.2.1	Definições . . . . .	26
1.2.2	Estrutura e termoestabilidade . . . . .	27
1.3	Fluxo de energia em proteínas . . . . .	28
1.3.1	ATD - <i>Anisotropic Thermal Diffusion</i> . . . . .	29
1.4	Sistema de estudo: xilanases da família 11 . . . . .	30
1.4.1	Importância e função . . . . .	31
1.4.2	Termoestabilidade em xilanases . . . . .	31
1.4.3	Proteínas 2VUJ e 2VUL . . . . .	33
1.4.4	Alinhamentos das xilanases . . . . .	34
<b>2</b>	<b>Metodologia</b>	<b>37</b>
2.1	Estruturas . . . . .	37
2.2	Simulações de equilíbrio . . . . .	38
2.2.1	Análises simulação de equilíbrio . . . . .	39
2.3	ATD - <i>Anisotropic Thermal Diffusion</i> . . . . .	41
2.4	Relaxamento . . . . .	44
2.5	Vídeo . . . . .	44
2.6	<i>Scripts</i> e programas . . . . .	45
<b>3</b>	<b>Resultados e Discussão</b>	<b>47</b>
3.1	ATD no vácuo . . . . .	47
3.1.1	2VUJ <i>versus</i> 2VUL . . . . .	48
3.1.2	Mesofílica <i>versus</i> termofílicas . . . . .	50
3.1.3	As outras proteínas . . . . .	53
3.1.4	Conclusões . . . . .	54
3.2	ATD na presença de solvente . . . . .	55
3.2.1	Comparação ATD em água versus vácuo . . . . .	56
3.2.2	Conclusões . . . . .	58

3.3	Contribuição da Cadeia Lateral . . . . .	59
3.3.1	Análise geral . . . . .	59
3.3.2	Comparação mesofílica, termofílica e hipertermofílica . . . . .	64
3.3.3	CCL individual . . . . .	67
3.3.4	Conclusões . . . . .	72
3.4	Relaxamento . . . . .	72
3.4.1	Relaxamento em vácuo . . . . .	73
3.4.2	Relaxamento em água . . . . .	75
3.4.3	Relaxamento sem termalização . . . . .	76
3.4.4	Conclusões . . . . .	77
3.5	ATD em função do tempo . . . . .	78
3.5.1	Mutações no banho térmico . . . . .	78
3.5.2	Comparação água <i>versus</i> vácuo . . . . .	81
3.5.3	Conclusões . . . . .	83
3.6	Análises da simulação de equilíbrio . . . . .	83
3.6.1	RMSD . . . . .	83
3.6.2	Solvatação . . . . .	87
3.6.3	Energia entre átomos não ligados . . . . .	90
3.6.4	Ligações de hidrogênio . . . . .	93
<b>4</b>	<b>Conclusões</b>	<b>97</b>
4.1	2VUJ e 2VUL . . . . .	98
4.2	Sugestões . . . . .	99

## REFERÊNCIAS

103

## Apêndice A - Tabela de aminoácidos

113

# Capítulo 1

## Introdução

### 1.1 Dinâmica Molecular

A simulação por Dinâmica Molecular (DM) é uma técnica computacional que realiza cálculos em modelos atomísticos, resolvendo suas equações de movimento. A partir das posições e velocidades iniciais e um potencial estipulado, a resolução das equações de movimento da segunda lei de Newton (Equação 1) torna-se possível. A variação do potencial ( $V(\vec{r})$ ) no espaço ( $\vec{r}_i$ ) resulta no negativo da força ( $\vec{F}_i$ ) que atua na partícula  $i$  (Equação 2), sendo que o potencial é dado pelos campos de força. Apartir desta informação, calcula-se a posição e velocidade através da Equação 1, onde  $m_i$  é a massa da partícula  $i$ . A resolução dessas equações gera a evolução temporal das posições e velocidades, ou seja, é obtido o movimento de todos os átomos no tempo.

$$\frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i} \quad (1)$$

$$\vec{F}_i = -\nabla V(\vec{r}_i) \quad (2)$$

As interações entre as partículas são determinadas pelos campos de força, que são parametrizações das interações intra- e intermoleculares. Existem vários campos de força que foram desenvolvidos e validados para diversas aplicações. Em especial, os mais conhecidos aplicados em biomoléculas são AMBER (2), CHARMM (3), GROMOS (4) e OPLS (5). De modo geral, estes são compostos pelos potenciais de átomos ligados e não-ligados (Equação 3). São exemplos de potencial não-ligado os potenciais de Lennard-Jones e de Coulomb, Equação 4, onde  $r_{ab}$  é a distância entre os átomos  $a$  e  $b$ ,  $q_a$  e  $q_b$  são as cargas parciais e  $\varepsilon_{ab}$  e  $\sigma_{ab}$  são os parâmetros dependentes dos tipos de átomos envolvidos,  $a$  e  $b$ . No potencial de Lennard-Jones,  $\varepsilon_{ab}$  é o valor do potencial no mínimo (ou seja, na distância ótima) e  $\sigma_{ab}$  é a separação não infinita na qual o potencial é nulo.

O potencial entre átomos ligados pode ser descrito como na Equação 5, onde a variação do comprimento e ângulo das ligações são representados como movimentos harmônicos nos primeiro e segundo termo, sendo  $K_r$  e  $K_\theta$  as constantes de força e  $r_0$  e  $\theta_0$  a distância e ângulo de equilíbrio. O terceiro termo representa torções, denominadas deformações de ângulo

diedro. Nesta equação,  $\varphi$  representa o ângulo da diedro e  $\gamma_n$  sua fase.

$$V_{Total} = V^{Ligado} + V^{Não-ligado} \quad (3)$$

$$V^{Não-ligado} = \sum_{a,b} 4\epsilon_{ab} \left[ \left( \frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left( \frac{\sigma_{ab}}{r_{ab}} \right)^6 \right] + \frac{q_a q_b}{4\pi\epsilon_0 r_{ab}} \quad (4)$$

$$V^{Ligado} = \sum_{ligações} K_r(r - r_0)^2 + \sum_{ângulos} K_\theta(\theta - \theta_0)^2 + \frac{1}{2} \sum_{n,diedro} V_n [1 - (-1)^n \cos(n\varphi + \gamma_n)] \quad (5)$$

Dois principais tipos de *ensembles* são utilizados em dinâmica molecular. O microcanônico (NVE) fixa valores tanto para o número de átomos quanto para o volume do sistema, além da energia ser mantida constante. Há também o *ensemble* canônico (NVT), no qual o número de átomos e volume também são mantidos contantes, entretanto neste *ensemble* a temperatura é mantida constante.

Algoritmos de dinâmica molecular geram um sistema inicial a uma temperatura T através da velocidade inicial dos átomos, normalmente escolhida de forma randômica através da distribuição de Maxwell-Boltzmann (Equação 6). Esta distribuição descreve a probabilidade para um átomo  $i$  ter as velocidades  $v_{ix}$ ,  $v_{iy}$  e  $v_{iz}$ , nas direções x, y e z, dada uma temperatura T. Sendo  $k_B$  a constante de Boltzmann e  $m_i$  a massa do átomo  $i$ .

$$P(v_{ix}, v_{iy}, v_{iz}) = \left( \frac{m_i}{2\pi k_B T} \right)^{1/2} e^{-\frac{1}{2} \frac{m_i(v_{ix}^2 + v_{iy}^2 + v_{iz}^2)}{k_B T}} \quad (6)$$

De forma reversa, a temperatura pode ser calculada através do teorema da equipartição, que a relaciona com energia cinética:

$$\frac{1}{2} \sum_{i=1}^N m_i v_i^2 = \frac{3}{2} N k_B T \quad (7)$$

onde  $m_i$  e  $v_i$  são a massa e velocidade da partícula i, respectivamente, e N o número total de partículas.

## Termostato de Langevin

Em sistemas isotérmicos (NVT e NPT), é necessário empregar uma estratégia a fim de manter a temperatura constante. A dinâmica de Langevin é comumente utilizada, de forma que a temperatura do sistema é mantida constante através da interação com um banho

térmico virtual. Esta dinâmica, aplicada ao sistema, seja em todo ou em parte dele, é denominada termostato em analogia aos termostatos reais, devido à sua propriedade de manter a temperatura constante. Outro nome empregado é banho térmico, em analogia com reservatórios térmicos, os quais possuem alto calor específico e podem manter a temperatura constante de um sistema termodinâmico. Este termostato, em termos físicos, corresponde à adição de duas componentes na equação da segunda lei de Newton: uma força de atrito proporcional à velocidade e contrária ao movimento das partículas sujeitas ao banho e uma força estocástica que simula impactos com as moléculas do banho virtual. Então, a equação de movimento fica:

$$m_i \vec{a}_i = \vec{F}_i - \gamma \vec{v}_i + \vec{R}_i(t) \quad (8)$$

$$\langle \vec{R}_i(t) \rangle = 0 \quad (9)$$

onde  $m_i$ ,  $v_i$  e  $a_i$  são respectivamente a massa, velocidade e aceleração do átomo  $i$ .  $\gamma$  é o coeficiente de fricção e pode ser expresso em termos da frequência de colisão  $\zeta = m_i \gamma$ . Por outro lado,  $\vec{R}_i(t)$  representa a força estocástica aplicada no átomo  $i$  em função do tempo e deve obedecer a condição de expressa na equação 9. Assim, na prática a temperatura é mantida constante através do controle de  $\gamma$  e  $\vec{R}_i(t)$ .

## Reescalonamento de velocidades

Outra metodologia utilizada para equilibrar o sistema a uma nova temperatura é o reescalonamento de velocidades. Como visto anteriormente, a temperatura do sistema pode ser expressa por meio das velocidades dos átomos que o compõem. Assim, as velocidades podem ser reescaladas com uma certa frequência para que o sistema atinja a temperatura desejada. Desta forma, a transição de temperatura é feita de forma gradativa e não chega a ser tão brusca como se uma nova fosse atribuída de um passo a outro dentro de um *loop* de simulação.

## 1.2 Termoestabilidade em proteínas

Microorganismos extremófilos\* vêm sido estudados nos últimos 40 anos (6). A capacidade de adaptação às condições destes organismos é refletida nas condições de atividade e estabilidade de suas enzimas. Desta maneira as enzimas destes organismos também podem ser denomina-

---

\*Aqueles que vivem em condições extremas como alta ou baixa temperatura ou pH, ambientes salinos ou ainda altas pressões.

das extremófilas, já que atuam em condições extremas de pH, temperatura ou concentrações salinas, entre outros.

A faixa de temperatura em que a maioria das proteínas atua é estreita, motivo pelo qual este é um fator de grande impacto sobre a atividade enzimática. Recentemente, com o aprimoramento das técnicas de biotecnologia molecular e a possibilidade de geração de proteínas recombinantes, o setor industrial voltou seu interesse para o uso de enzimas em processos químicos industriais. Entretanto, as enzimas ainda são mais caras e com menor tempo de vida útil se comparadas aos processos tradicionais e isto tem representado um desafio no emprego destas na indústria. Por outro lado, as enzimas são ótimos catalisadores e o uso híbrido de técnicas enzimáticas e processos tradicionais pode ser uma alternativa promissora para a indústria usufruir das suas vantagens (7). Processos industriais normalmente requerem altas temperaturas, e a viabilização de enzimas para utilização nestes casos pode gerar melhorias como redução de dejetos químicos e menor custo. Proteínas termoestáveis como amilase, xilanase e pectinase são de grande importância industrial (8).

### 1.2.1 Definições

Muitas vezes a proteína pode ter uma boa estabilidade térmica mas não possui atividade elevada à altas temperaturas. Para isso normalmente duas medidas estão relacionadas com essas duas propriedades: temperatura de *melting* ( $T_m$ ) e temperatura de atividade ótima. A  $T_m$  é a temperatura em que 50% da proteína está desnaturada, e uma das formas que pode ser medida é por dicroísmo circular, através da variação da estrutura em função da temperatura. A temperatura de atividade ótima é o valor de temperatura em que a curva de atividade em função da temperatura é máxima. Outra medida para avaliar a atividade é o tempo de meia-vida ( $t_{1/2}$ ), no qual o gráfico de atividade é avaliado em função do tempo depois do aquecimento a uma temperatura escolhida. A temperatura de crescimento ótimo ( $T_{CO}$ ) é uma medida que pode ser utilizada de certa forma tanto para atividade quanto para estabilidade térmica, onde mede-se a melhor temperatura de crescimento do microorganismo originário da proteína.

Todas essas medidas possuem muitas variáveis, e por isso fica praticamente impossível utilizá-las para comparar proteínas tratadas em condições diferentes. Então, aqui as proteínas serão tratadas pela classificação que as separa em três níveis de termoestabilidade, mesofílicas (estabilidade próxima à temperatura ambiente), termofílicas ( $T_{CO}$  entre 50-80 °C) e hipertermofílicas\* ( $T_{CO}$  maiores que 80 °C), em ordem crescente de termoestabilidade (9).

---

\*Ou também, mesófilas, termófilas e hipertermófilas.

## 1.2.2 Estrutura e termoestabilidade

Assim como a função de uma proteína depende de sua estrutura tridimensional, sua termoestabilidade está relacionada às interações intra- e intermoleculares, por exemplo, interações com a própria proteína e com a água. Portanto, vários estudos são feitos no intuito de relacionar a estrutura (de forma dinâmica ou estática) à sua termoestabilidade. Porém, ainda não é muito claro quais são as propriedades moleculares que contribuem para a estabilidade térmica de proteínas. Provavelmente, esses mecanismos são complexos e devem ser levados em conta não só um, mas vários fatores.

A comparação entre estruturas mesofílicas com suas homólogas termofílicas levou à conclusão que os mecanismos desenvolvidos nos reinos Archea e Eubacteria em sua adaptação a ambientes termicamente extremos são diferentes (10). Portanto, os mecanismos pelos quais as proteínas adquirem maior estabilidade térmica devem ser estudados levando em conta a origem da proteína. Isso também evidencia a complexidade desta tarefa, uma vez que há diferentes mecanismos que levam ao mesmo fim (a estabilidade térmica), através de biomoléculas compostas pelos mesmos aminoácidos fundamentais.

As interações hidrofóbicas no núcleo são relacionadas à termoestabilidade, uma vez que o maior empacotamento da estrutura a torna menos suscetível ao desenovelamento. Ainda, a minimização de interação entre o núcleo hidrofóbico e a água colabora para maior estabilidade (efeito hidrofóbico). Assim, um maior empacotamento, que está relacionado principalmente às interações hidrofóbicas no núcleo, leva a uma maior rigidez da estrutura como um todo, e portanto, à uma maior termoestabilidade. Porém, a relação entre rigidez e estabilidade vem sendo questionada. Através de experimentos de troca de hidrogênio por deutério, foi mostrado que a proteína rubredoxina do organismo hipertermófilo *Pyrococcus furiosus*\* apresenta uma alta taxa de abertura conformacional para acesso ao solvente (11). Essa inesperada flexibilidade, dada a estabilidade da proteína, mostra que a relação rigidez-termoestabilidade de fato deve ser repensada. Além disso, em (12), são reunidos vários estudos em que proteínas termoestáveis apresentam alta flexibilidade estrutural.

Em geral, as interações eletrostáticas tendem a estabilizar a estrutura. Deve-se destacar a existência de resíduos carregados próximos, pela possibilidade de formação de pontes salinas. De fato, as pontes salinas são mais abundantes em hipertermófilas quando comparado com suas proteínas homólogas menos estáveis (13, 14). Também, o conjunto de ligações de hidrogênio na proteína como um todo é considerado um fator de estabilidade. As pontes dissulfeto formam ligações covalentes entre resíduos de cisteína e normalmente estabilizam estruturas

\* Até a data do estudo, 1999, era a proteína caracterizada mais termoestável, com a taxa de desenovelamento de  $10^{-6} s^{-1}$  a 100 °C.

terciárias. Vários outros fatores são frequentemente citados quando são comparadas duas estruturas de diferentes termoestabilidades: a composição de aminoácidos (15, 16), a quantidade relativa de estrutura secundária presente na estrutura (15, 17), o encurtamentos de *loops* (18, 19), entre outros.

### 1.3 Fluxo de energia em proteínas

O fluxo de energia através de biomoléculas tem recebido atenção em estudos computacionais, experimentais e teóricos. Este tem grande importância, pois pode ser relacionado com allosterrismo, transdução de sinal, atividade catalítica, entre outros, exemplificados adiante. O fluxo de energia em uma proteína globular é bem descrito pela teoria de transporte em um *cluster* percolante\*, no qual a energia flui facilmente por contatos fortes, e em sítios fracamente conectados, a energia flui mais devagar (20). Adicionalmente, a transferência de energia dentro da proteína também é descrita por modos normais, onde o acoplamento anarmônico entre os modos vibracionais permite o fluxo de energia (21). Após uma grande revisão, (22) concluiu que toda proteína globular possui canais de transporte de energia, caracterizando fluxos anisotrópicos.

Tratando-se de forma experimental, algumas metodologias foram aplicadas no estudo do fluxo de calor dentro de uma biomolécula. Experimentos de ressonância magnética nuclear e mutagênese mostraram o acoplamento energético de certos resíduos no domínio proteico PDZ (23). Espectroscopia ultra-rápida foi amplamente utilizada com este propósito. Como exemplo, o estímulo energético do N-terminal em uma hélice de peptídeos possibilitou a visualização da propagação do calor e cálculos da constante de difusão térmica (24). Através de estudos de efeitos de mutações, a comunicação entre dois sítios de acoplamento entre a proteína interleucina-1 $\beta$  e seu receptor foi atribuída a um resíduo, sendo de importância para a estabilidade e função (25).

Por fim, o estudo computacional dá uma maior abrangência de possibilidades de observação deste fenômeno. Simulações de DM são largamente utilizadas, algumas vezes em conjunto com técnicas experimentais ou cálculos teóricos (26–30). Como um exemplo interessante, o resfriamento do grupo heme na mioglobina foi estudado com DM. Os tempos de mudança de temperatura da redondeza foram medidos, e assim, concluiu-se que uma possível canalização do calor acontece do heme para a água (31). O domínio proteico PDZ foi bastante estudado por possuir um mecanismo particular de transferência de energia. Através de um mapeamento termodinâmico, foi encontrado caminho de conectividade energética dentro destes

---

\*Um cluster percolante pode ser definido como uma rede de pontos interconectados. Em proteínas, os pontos seriam os átomos ou resíduos, e as interações químicas (ligações) conectam os pontos entre si, formando a rede.

domínios, indicando que esta pode ser uma proteína sinalizadora (32). Mais estudos sobre a energia anisotrópica neste domínio, desta vez mapeando o mecanismo físico do caminho existente nessa proteína confirmaram essa sinalização intramolecular (33).

O tempo de difusão de energia em uma proteína é restrinido pelo seu tamanho. Considerando o modelo de difusão vibracional em fractais aplicado a proteínas, Leitner faz a estimativa: os modos normais de baixa frequência de uma proteína possuem velocidade da ordem da velocidade do som ( $10 \text{ \AA}/\text{ps}$ ) e o livre caminho médio é da ordem de  $1 \text{ \AA}$ . Portanto, o limite inferior para o tempo de difusão de energia é da ordem de  $0,1 \text{ ps}$  (22).

Canais de propagação de energia, podem servir para o rápido resfriamento, direcionar o transporte de energia para aumentar a velocidade de reação, sinais de comunicação e alosterismo. Portanto, o estudo da termoestabilidade de proteínas através da indução e monitoramento do fluxo anisotrópico de energia apresenta-se com um grande potencial. Aqui estudaremos a possibilidade de encontrar mecanismos de propagação de energia relacionados com a adaptação de proteínas a ambientes de altas temperaturas.

### 1.3.1 ATD - *Anisotropic Thermal Diffusion*

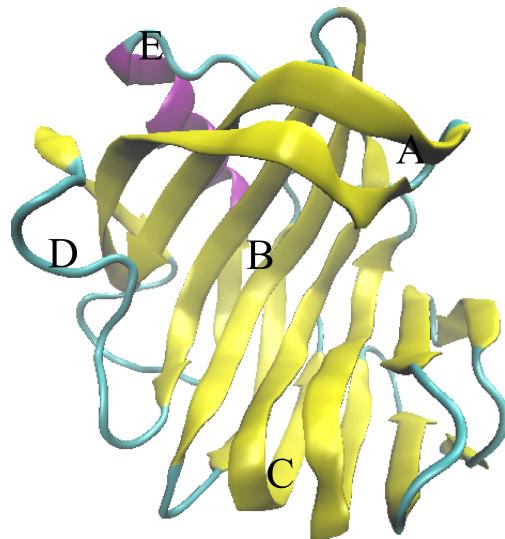
A difusão anisotrópica de calor, ou ATD, é um método que simula um estímulo energético localizado e analisa a resposta do resto do sistema a nível atômico. Tal simulação só é possível, até o presente momento, por dinâmica molecular. Aqui, o estímulo é um banho térmico localizado em somente um aminoácido, e a resposta da proteína é monitorada através da temperatura individual da cada resíduo. O método requer que o sistema esteja inicialmente em baixas temperaturas, aumentando assim a razão sinal-ruído. No sistema proteico, espera-se que a difusão térmica seja anisotrópica, como inicialmente proposto pelo pioneiro do método (33): Ota e sua equipe estudaram a difusão de calor em uma proteína membro da família domínio PDZ, revelando uma via sinalizadora previamente proposta. Logo após, nosso grupo encontrou que resíduos de arginina conservados possuem maior contribuição na difusão térmica, os quais também possuem um papel chave na atividade de receptores de hormônio da tireoide (1). Entretanto, a metodologia foi ligeiramente modificada em relação ao trabalho de Ota et al.: ao invés de analisar a difusão originadas somente em resíduos considerados importantes, o banho térmico é acoplado a todos resíduos, e, portanto, a difusão é observada para toda proteína. Adicionalmente, é feito o monitoramento da temperatura para cada aminoácido (energia cinética), ao invés do desvio quadrático médio (RMSD) como foi empregado no trabalho original\*. De maneira que o método de ATD permite obter a resposta térmica de uma proteína inteira quando o calor é fornecido localmente.

\*A metodologia será explicada mais detalhadamente na Parte 2.

Embora o método ATD já tenha sido utilizado para observar sinalização intramolecular e a importância de argininas na difusão térmica, ainda não foi proposto relacioná-lo à termoestabilidade de proteínas. Portanto, partindo da premissa que uma proteína globular possui fluxo anisotrópico de energia, o presente trabalho propõe-se a estudar uma possível relação entre termoestabilidade e mecanismos de difusão térmica anisotrópica em proteínas e correlacioná-los com sua estrutura.

## 1.4 Sistema de estudo: xilanases da família 11

As xilanases, ou endo-1,4- $\beta$ -xylanase, são glicosidases que catalizam a hidrólise do polímero de açúcar xilano, clivando ligações 1,4- $\beta$ -D-xilosídicas. Inicialmente eram divididas em duas principais famílias nomeadas F e G ou também como 10 e 11 (34), de acordo com o peso molecular e o ponto isoelétrico. Porém, recentemente, novas xilanases caracterizadas não poderiam se encaixar em nenhum dos grupos, resultando em uma nova classificação. Então, foi proposta a classificação de xilanases baseando-se na semelhança da sequência primária do domínio catalítico, o que incluiu outras proteínas glicosidases (35). Hoje existem 17 famílias glicosil hidrolase (9).



**Figura 1** – Estrutura das xilanases da família 11. Em forma de mão direita, possui regiões com nomes próprios: A) polegar, B) palma, C) dedos, D) cordão e E)  $\alpha$ -hélice.

Dentro da grande variedade de xilanases extremófilas, serão tratadas aqui as termoestáveis da família 11, que compõe um vasto conjunto de estruturas já caracterizadas e de diferentes termoestabilidades. Mais especificamente, serão utilizadas cinco proteínas (Tabela 2), sendo uma mesofílica, uma hipertermofílica e três termofílicas.

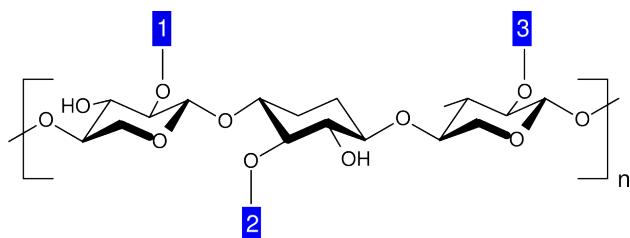
Sua estrutura é composta por uma hélice- $\alpha$  e duas folhas- $\beta$ , formando o sanduíche- $\beta$ . Descrita desde 1994 como uma mão direita (36), possui regiões nomeadas como polegar, palma, dedos, cordão e hélice- $\alpha^*$  (Figura 1).

### 1.4.1 Importância e função

A xilanase é uma proteína bastante estudada pela importância industrial. Ela é utilizada no processo de fabricação de papel, ração animal, setor alimentício (como sucos) (9) e mais recentemente adquiriu importância na terceira geração do bioetanol (37, 38). A reação catalisada é de quebra de polímeros de açúcares (xilano) em dímeros, monômeros ou trímeros. A parede celular de várias plantas é composta principalmente por celulose, hemicelulose e lignina. Acredita-se que a interface entre a lignina e a celulose seja importante para integridade da parede celular e rigidez (39). Como o xilano constitui a maior porção da hemicelulose, a xilanase apresenta-se como uma das opções para a degradação da parede celular e separação de seus constituintes.

#### Xilano

O polissacarídeo degradado pela xilanase é o xilano. Monômeros de D-xilose são ligados através de ligações  $\beta$ -1,4-glicosil, e são essas as ligações clivadas pela xilanase. Estes podem ser substituídos por grupos 4-O-metil-D-glucuronosil,  $\alpha$ -L-arabinofuranosil, acetil, feruloil, glucuronopiranossil e p-coumaroil (9, 39, 40), formando polissacarídeos complexos.



**Figura 2** – Possível unidade repetitiva do polímero de xilano, composta em sua cadeia principal por três moléculas de D-xilano e alguns possíveis sítios de ramificação representados por números.

### 1.4.2 Termoestabilidade em xilanases

Os determinantes estruturais que levam ao aumento de termoestabilidade normalmente são dependentes do grupo de proteínas em estudo. Na Família 11 das Xilanases, variados fatores

\*Ou no original em inglês, *thumb*, *palm*, *fingers*, *cord* e  *$\alpha$ -helix*, respectivamente.

já foram apontados e serão enumerados a seguir. Pontes dissulfetos não são apontadas como elementos para explicar a termoestabilidade, dada a análise de 82 sequências de aminoácidos desta família; a frequência de resíduos de cisteína também é baixa dentro dessas proteínas (41). Apesar da introdução desse elemento já ter sido usado com sucesso para a melhora da estabilidade térmica: a introdução de uma ponte dissulfeto ligando a hélice- $\alpha$  e uma fita- $\beta$  aumentou a termoestabilidade, porém reduziu a atividade em alta temperatura (42, 43). Em outro caso, essa modificação gerou aumento tanto na termoatividade quanto no tempo de meia-vida (44). Neste caso, a ponte dissulfeto ligou o N- e C-terminal, reduzindo a mobilidade das extremidades da proteína.

O N-terminal é uma região de grande importância para a estabilidade térmica em xilanases da família 11. Este já foi alvo de estudo e mutações, que geraram variantes com aumentada termoestabilidade (45–48). De fato, foi observado em simulações de DM que essa região inicia o processo de desnaturação em algumas xilanases, enquanto que a região da palma é apontada como a área mais resistente ao desenovelamento (49). Xilanases termofílicas geralmente possuem uma fita- $\beta$  a mais no N-terminal, enquanto que mesofílicas, em sua maioria, possuem um loop longo e irregular nessa posição (44, 49). Das duas principais estruturas aqui utilizadas (2VUJ e 2VUL), a variante hipertermoestável possui 6 de 7 mutações no N-terminal.

Diversos outros fatores foram identificados por conferir termoestabilidade em xilanases, como aumento na fração de área de superfície polar (50, 51), aumento do números de ligações de hidrogênio (51–55), interações hidrofóbicas (54, 56–58), oligomerização (59, 60), entre outros.

Modificações em estruturas de xilanases a fim de aumentar a termoestabilidade foram abordadas tanto através de evolução dirigida\* (técnicas experimentais) quanto por design racional (técnicas computacionais). Estudos de modelagem por homologia geraram novas variantes termoestáveis da xilanase de *Bacillus subtilis* (61). Assim como essa mesma metodologia aumentou a termoestabilidade da xilanase de *Neocallimastix Patriciarum* (62). E a xilanase A de *Aspergillus niger* teve termoestabilidade aumentada pela substituição de seu N-terminal pela mesma região da xilanase do organismo *Thermomonospora fusca* (46). Outro caso na família 11 de xilanases abordou o problema através de alinhamento estrutural entre a proteína termofílica e sua homóloga mesofílica, gerando uma variante com o aumento de 10 °C na  $T_{AO}$  e 9 °C na  $T_m$  da proteína Xyl1 de *Streptomyces sp.*. A mutação T11Y gerou uma nova interação aromática, similar à da proteína comparada mais estável, o que estabilizou a estrutura, aumentando assim a termoestabilidade (63).

Em especial, para a xilanase de *Bacillus circulans* (código PDB 1XNB) que foi utilizada

---

\*Técnica que simula a evolução natural de proteínas de maneira acelerada.

neste estudo, algumas variantes de maior termoestabilidade foram desenvolvidas. A utilização de modelo computacional, que aponta regiões instáveis através de análises de flutuações térmicas, tornou possível a construção de um triplo mutante\* com aumento na  $T_m$  de 48 °C à 50,1 °C e  $t_{1/2}^{\dagger}$  de 2 min para 30 min (64). Em 2011, com o aprimoramento da técnica, um quádruplo mutante<sup>‡</sup> levou ao aumento da  $T_m$  para 52,2 °C e o  $t_{1/2}$  para 60 min. As mutações realizadas são em resíduos de superfície na face oposta ao sítio catalítico. De certa forma, estes resíduos não representam uma melhora representativa frente à outras xilanases termoestáveis existentes, mas mostram que é possível o design inteligente de proteínas, especialmente na família de interesse de estudo deste trabalho.

### 1.4.3 Proteínas 2VUJ e 2VUL

Uma atenção especial é dada à comparação de duas xilanases. Esse sistema em questão é interessante pois uma proteína hipertermoestável (2VUL) foi gerada a partir de uma outra termoestável (2VUJ), gerando uma diferença na  $T_m$  de 25 °C (65). O processo experimental realizado por Dumon e sua equipe é uma evolução compulsória, que é extremamente trabalhoso e caro. Ele é a combinação das metodologias GSSM (*gene site saturation mutagenesis*) e *GeneReassemblyTM*. Primeiramente a GSSM gerou todas as possíveis substituições individuais de aminoácido em cada posição na sequência primária. Com essa imensa biblioteca de mutantes, ensaios de atividade residual após o aquecimento a 76 °C restringiram os *hits* com o mínimo de 150% de atividade em relação à proteína nativa. Em sequência, estes *hits* foram ranqueados através de outros ensaios com maiores temperaturas. Por fim, essas principais estruturas possuindo mutações individuais geraram uma biblioteca de aproximadamente 12.000 sequências com todas combinações entre elas. Novos ensaios de atividade residual dependente da temperatura e do tempo selecionaram a estrutura 2VUL.

A estrutura tridimensional destas proteínas alinhadas é extremamente semelhante, tendo somente 7 aminoácidos modificados (Tabela 1). Somente foi compreendida a razão do aumento da termoestabilidade para dois dos sete resíduos mutados: de acordo com o trabalho original, a mutação Thr13Phe aumentaria as interações hidrofóbicas, e Ser9Pro trancaria a conformação do *loop* (65). Portanto, aderindo ao desafio que o próprio artigo lança, este é mais um motivo para desenvolver novas metodologias a fim de estudar o comportamento, em nível molecular, de proteínas termoestáveis.

\* Mutações Phe48Tyr/Thr50Val/Thr147Leu

<sup>†</sup> Tempo de meia vida à 50 °C.

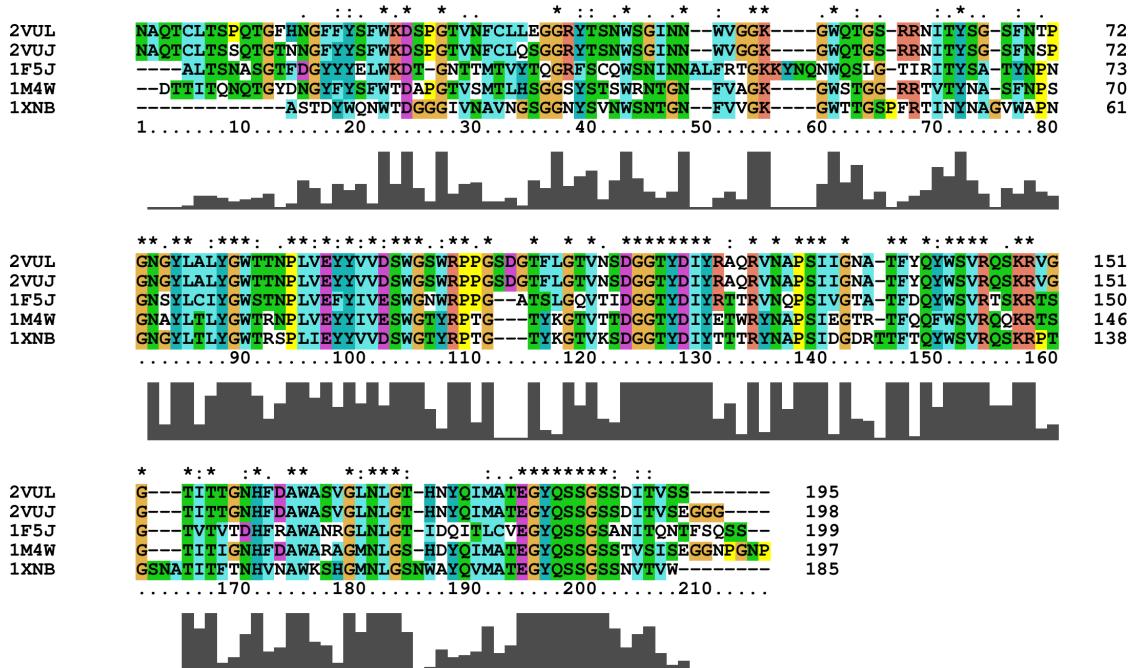
<sup>‡</sup> Mutações Phe48Tyr/Thr50Val/Asn52Tyr/Thr147Leu

**Tabela 1** – Mutações que diferenciam as estruturas de 2VUJ e 2VUL.

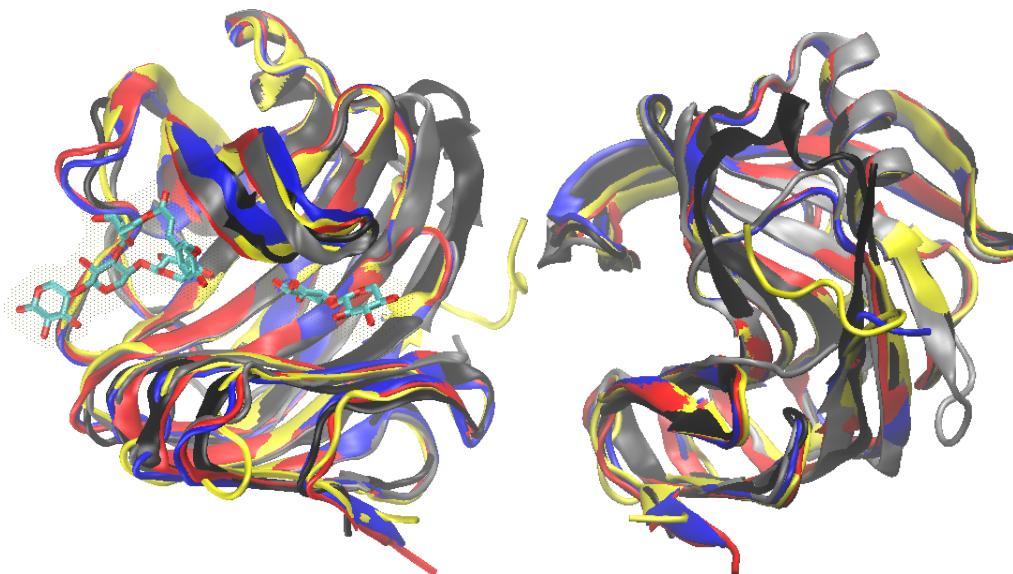
	2VUJ		2VUL	
	Aminoácido	Estrutura	Aminoácido	Estrutura
9	SERINA		PROLINA	
13	TREONINA		FENILALANINA	
14	ASPARAGINA		HISTIDINA	
18	TIROSINA		FENILALANINA	
34	GLUTAMINA		LEUCINA	
35	SERINA		AC. GLUTÂMICO	
71	SERINA		TREONINA	

#### 1.4.4 Alinhamentos das xilanases

As 5 proteínas deste estudo, representadas por seus códigos PDB 1M4W, 1F5J, 1XNB, 2VUJ e 2VUL, por estarem na mesma família, possuem estruturas tridimensionais muito semelhantes, como é mostrado no alinhamento estrutural na Figura 4. Uma análise do alinhamento sequencial (Figura 3) aponta o grau de conservação para cada aminoácido equivalente.



**Figura 3** – Alinhamento global da sequência de aminoácidos entre as cinco proteínas utilizadas nesse estudo, regiões conservadas apontadas pelo gráfico abaixo do alinhamento. Realizado no ClustalX.



**Figura 4** – Alinhamento estrutural das 5 proteínas aqui estudadas. Apresentam a região catalítica bem conservada em relação à cadeia principal (esquerda), com o ligante da xilanase EnXyn11A da mesma família (código PDB 2VGD). As variações aparecem no comprimento de alguns loops e fitas- $\beta$  (direita).



# Capítulo 2

## Metodologia

### 2.1 Estruturas

O conjunto de proteínas estudadas são pertencentes à família 11 de xilanases. Foram utilizadas as estruturas do domínio catalítico obtidas no banco de dados *Protein Data Bank* (PDB). Para facilitar a leitura e a distinção, as proteínas sempre serão nomeadas pelo seu código PDB.

**Tabela 2** – Xilanases e suas características.

Proteína	Código PDB	Organismo	Termoestabilidade	Resolução	Número de Aminoácidos <sup>1</sup>
XynA	1XNB	<i>Bacillus circulans</i>	$t_{1/2} = 100'$ à 45 °C; $t_{1/2} = 2'$ à 50 °C (64); $T_{AO} = 45^\circ\text{C}$ (66).	1,49	185
XynB	1F5J	<i>Dictyoglomus thermophilum</i>	90% de atividade após 8 h à 80 °C (67); $T_{AO} = 75^\circ\text{C}$ (41)	1,80	199
Xyn11A	1M4W	<i>No nomuraea flexuosa</i>	$t_{1/2} = 273'$ à 80 °C; $t_{1/2} = 28'$ à 100° C; $T_{AO}$ entre 65–80 °C (68).	2,10	197
<i>EvXyn11</i>	2VUJ	Isolada do ambiente	50% de atividade após 15' à 66 °C; $T_{AO} = 65^\circ\text{C}$ ; $T_m = 76,2^\circ\text{C}$ (65).	1,80	195
<i>EvXyn11<sup>TS</sup></i>	2VUL	Geneticamente modificada de 2VUJ	50% de atividade após 15' à 91 °C; $T_{AO} = 75^\circ\text{C}$ ; $T_m = 101,1^\circ\text{C}$ (65).	1,90	193

<sup>1</sup>da estrutura PDB, que representa o domínio catalítico.

Todas estruturas utilizadas possuem boa resolução e não foi observado nenhum átomo de carbono faltando. O comprimento das estruturas varia de 185 a 199 aminoácidos. Antes de qualquer simulação cada estrutura foi preparada como segue. Somente uma cadeia proteica foi utilizada. No caso de resíduos com dupla conformação, foi escolhida aquela de maior probabilidade da conformação, através de uma biblioteca de rotâmeros e o ambiente estereoquímico. Cada histidina foi inspecionada e sua protonação escolhida de forma a favorecer possíveis interações. Também foram removidos os ligantes, outras moléculas presentes no cristal e as águas cristalográficas. Não foram levadas em conta possíveis glicosilações, sendo utilizada somente estruturas proteicas.

## Modificação de 1XNB

Foram feitas modificações na estrutura da proteína mesofílica 1XNB. A estrutura da mesma foi alinhada à 2VUJ e 2VUL por meio do programa MultiProt (69). Foi localizada a região de cordão em ambas termoestáveis e na mesofílica, a qual apresentou comprimento menor que as termoestáveis. A sequência de aminoácidos desta região alinhada para as três proteínas pode ser vista na Figura 10.

Para o experimento de inserto, a estrutura de 1XNB recebeu literalmente um recorte das outras estruturas: foram retirados os resíduos 91 e 92, e substituídos por 102 à 106 de 2VUL ou 2VUJ\*. Portanto, a retirada de dois aminoácidos e a inserção de cinco tornou o loop três resíduos mais longo. O arquivo PDB foi modificado manualmente, e depois refeita a numeração de átomos e resíduos.

## 2.2 Simulações de equilíbrio

O experimento realizado nesta dissertação é feito com simulações de dinâmica molecular utilizando o programa NAMD 2.8, com o campo de força CHARMM. Simulações de equilíbrio foram geradas no intuito de compor uma amostragem de estruturas iniciais para serem aplicadas no método ATD. O objetivo é gerar uma simulação longa levando em conta as condições biológicas, pois ainda que o método ATD use baixas temperaturas, a estrutura proteica estará equilibrada próxima a condições mais amenas.

As estruturas retiradas do PDB com as modificações explicadas no item anterior (Seção 2.1) foram colocadas em uma caixa cúbica com água de arestas entre 72 e 77,5 Å, dependendo da maior dimensão da proteína. Nestas condições, o sistema total é composto por 35 à 45 mil átomos. As cargas das proteínas são neutralizadas pela adição de íons  $Na^+$  e  $Cl^-$ , a uma concentração de 0,15 M, valor razoavelmente próximo de um meio fisiológico. As simulações foram feitas no *ensemble NVT*, com pressão e temperaturas mantidas pela dinâmica de Langevin a 1 atm e 300 K, respectivamente. Todas as simulações deste estudo foram realizadas em condições periódicas de contorno e com o passo de integração de 2 fs.

Nestas condições, os sistemas foram minimizados e equilibrados por 100 ps, e na sequência, realizadas simulações de equilíbrio por 40 ns.<sup>†</sup> Além destes 40 ns, para as proteínas 2VUJ e 2VUL foram feitos mais 20 ns de simulação. Destas simulações de equilíbrio foram retirados 20 *frames* igualmente espaçados no tempo (30 frames no caso das simulações de 60 ns). Estes *frames* compõem a amostragem utilizada no método ATD, ou seja, após serem termalizadas,

\*uma vez que na região 102-106 os resíduos são os mesmos para ambas proteínas

<sup>†</sup>Cada simulação desta demorou por volta de 12 dias em um nó com 12 processadores.

estas serão as estruturas iniciais para cada simulação de ATD.

## Simulações adicionais

A fim de aumentar a amostragem, foram utilizadas também outras simulações de equilíbrio realizadas pelo grupo para três proteínas: uma simulação de 40 ns para 1F5J, e quatro simulações de 40 ns tanto para 2VUJ quanto para 2VUL. As estruturas foram preparadas de forma semelhante.

## Termalização

O método ATD exige que a proteína esteja em baixa temperatura, diminuindo assim bruscamente o nível de ruído. *Frames* retirados das simulações de equilíbrio a 300 K são então termalizados a 10 K por reescalonamento de velocidades. O tempo de 2 ps de simulação foi suficiente para levar o sistema, ainda em água, à temperatura desejada, reescalando a velocidade a todo passo de integração.

### 2.2.1 Análises simulação de equilíbrio

As simulações de equilíbrio não servem apenas como fonte de amostragem estrutural ao método ATD, mas também compõem material para análises mais usuais em DM. Foram feitas análises de solvatação, ligação de hidrogênio, RMSD e energia de átomos não ligados (van der Waals e eletrostática) utilizando o programa LM-MDAnalysis (70).

#### Energia entre átomos não ligados

O programa LM-MDAnalysis calcula a energia de átomos não ligados, ou seja, a energia de van der Waals (VDW) e a eletrostática. Para isso, o programa baseia no cálculo de distâncias entre átomos em função do tempo de simulação.

Foram medidas as energias entre a cadeia lateral de cada aminoácido mutado de 2VUJ e 2VUL e o restante da proteína, dentro de um raio de 5 Å.

#### RMSD

O valor de desvio quadrático médio (*root-mean-square deviation*, RMSD) é uma grandeza comumente medida em experimentos de DM, e, de modo geral, expressa a variação das posições dos átomos em relação a uma estrutura escolhida. Aqui, dois tipos de RMSD foram

calculados, o RMSD computado para toda proteína, em função do tempo de simulação e o RMSD em função do aminoácido. A estrutura escolhida é a estrutura média da simulação e representa uma estrutura virtual. Para computar o RMSD, primeiro realiza-se um alinhamento entre a estrutura de referência e a estrutura do frame da simulação, ou seja, as duas estruturas são mantidas rígidas e sofrem deslocamentos e rotações de forma a minimizar a distância entre os átomos. Esse tipo alinhamento é feito somente entre os carbonos- $\alpha$ . Depois de alinhado, o RMSD é calculado para todos os átomos exceto hidrogênios. O RMSD entre as estruturas u e w contendo N átomos, é calculado através da Equação 10, encontrando as diferenças das posições  $r_x$ ,  $r_y$  e  $r_z$  do átomo i (equivalente nas duas estruturas).

$$RMSD(u, w) = \sqrt{\frac{1}{N} \sum_i^N \delta_i^2} = \sqrt{\frac{1}{N} \sum_i^N [(r_{ix}^u - r_{ix}^w)^2 + (r_{iy}^u - r_{iy}^w)^2 + (r_{iz}^u - r_{iz}^w)^2]} \quad (10)$$

Para calcular o RMSD nas proteínas 2VUJ e 2VUL, uma espécie de filtro foi aplicada. O primeiro cálculo de RMSD foi analisado, e foram apontadas 3 regiões mais móveis\*, localizadas entre os resíduos 2 ao 7, 12 ao 18 e 125 ao 139. Então, um novo alinhamento foi feito em relação aos C- $\alpha$  de todos os resíduos exceto essas 3 regiões. Esta técnica melhora o alinhamento, pois acredita-se que a porção mais rígida da estrutura é mais representativa no alinhamento, já que as regiões mais móveis tendem a variar muito no tempo.

Ainda, foi calculado também o RMSD para cada aminoácido, levando em conta todas as conformações de cada *frame*, ainda com as mesmas condições de alinhamento. Ou seja, ao invés de somar o desvio quadrático para todos os átomos da proteína, isto é feito para todos os átomos do resíduo individual, e no fim tira-se a média do RMSD por resíduo no tempo de simulação.

## Solvatação

Medir o quanto um resíduo está solvatado pode dar informações úteis tais como a acessibilidade do solvente ou a interação com o mesmo. Para cada uma das sete mutações entre 2VUJ e 2VUL, foram computados o número de moléculas de água dentro de um raio de 4 Å de cada átomo. As medidas foram graficadas em função do tempo de simulação, para 3 simulações de 40 ns, e suavizadas com filtro de médias a cada 20 pontos.

---

\*Regiões que apresentaram 60% ou mais do valor máximo de RMSD em função do resíduo.

## Ligações de hidrogênio

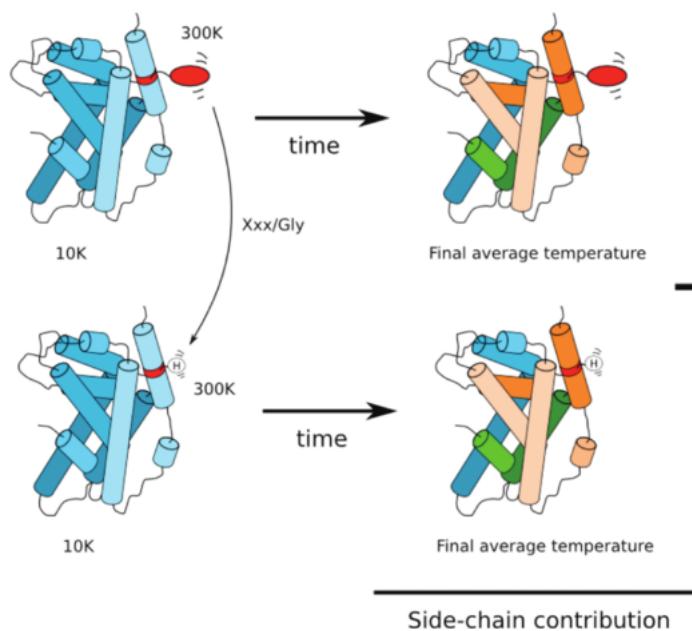
Ligações de hidrogênio são interações intermoleculares importantes para a estabilização das estruturas proteicas. Aqui focou-se em computar as ligações realizadas entre os aminoácidos mutados de 2VUJ e 2VUL e moléculas de água. Utilizando como definição de ligação de hidrogênio um ângulo de  $20^\circ$  e distância de 3,5 Å entre os átomos doador e receptor. Para todas as 5 simulações com 40 ns de 2VUJ e 2VUL foi computado o número de ligações a cada instante de tempo. Isso gerou uma distribuição, que foi apresentada como a média e o desvio padrão.

## 2.3 ATD - *Anisotropic Thermal Diffusion*

O método ATD consiste em acoplar um banho térmico a um resíduo e observar a propagação da energia térmica em todos os outros. Esse experimento é possível somente através de simulações de DM. Entretanto, para uma boa diferenciação do estímulo aplicado, é necessário ter um sistema em baixas temperaturas. As estruturas submetidas ao ATD são previamente resfriadas a 10 K (como explicado em Subseção 2.2 - Termalização) e, então, para cada resíduo da proteína um banho térmico é acoplado. Assim, se uma proteína possui 200 aminoácidos, serão realizadas 200 simulações de ATD, onde cada uma terá um resíduo diferente acoplado ao banho térmico. Em geral, o tempo de uma simulação de ATD é 30 ps.

O termostato de Langevin é aplicado a fim de manter a temperatura constante do aminoácido estimulado, fornecendo a energia necessária para tanto. A temperatura da proteína é calculada a partir das velocidades (Equação 7), em função do tempo e dos aminoácidos. Nos gráficos, utiliza-se tanto temperatura no tempo final quanto a temperatura média temporal. Foram utilizadas dois tipos de medida, o mapa de difusão térmica (MDT) e a contribuição da cadeia lateral (CCL).

O MDT reúne todas as simulações de ATD para todos os aminoácidos da proteína acoplados ao banho térmico. Onde uma linha vertical no mapa corresponde aos dados de uma simulação de ATD, isto é, cada ponto das abscissas representa o aminoácido aquecido e das ordenadas, a resposta cada aminoácido de toda a proteína. A representação da temperatura é feita em escala de cores, onde as cores frias representam temperaturas baixas e as cores quentes, altas. Tal escala foi dimensionada de acordo com a temperatura máxima e mínima escolhidas para melhor visualização (normalmente 45 K e 90 K). Junto ao MDT foram gerados os mapas de contados das proteínas, o qual representa a distância par a par entre todos os aminoácidos em uma escala de cinza. Esse mapa permite analisar a difusão mais previsível, a qual é dada pela menor distância.



**Figura 5** – Esquema do método ATD retirado de (1). A estrutura inicialmente a 10 K recebe um estímulo térmico através de um aminoácido. A resposta de cada resíduo da estrutura é apresentada no MDT. Na segunda linha, realiza-se a mutação do resíduo aquecido pela glicina, por exemplo. A resposta total da proteína mutada é subtraída da nativa, computando assim a CCL.

Para medir a CCL, além da simulação já explicada, uma nova simulação de ATD é realizada: o aminoácido aquecido é modificado para glicina ou alanina, e então realiza-se a simulação de ATD (Figura 5). Esse procedimento também é repetido para todos aminoácidos da proteína. Para cada simulação de ATD, computa-se a temperatura total da proteína, ao invés de monitorar a temperatura de cada aminoácido. Para um mesmo aminoácido aquecido, a resposta (temperatura) de toda a proteína nativa menos a resposta de proteína mutada representa a contribuição da cadeia lateral para o aquecimento total da estrutura. Ao fim, gráficos de CCL são representados como a temperatura em função do aminoácido acoplado ao banho.

Foram gerados também histogramas de CCL para alguns aminoácidos específicos. Ou seja, foram coletados vários pontos de CCL cujo resíduo acoplados ao banho térmico pertence ao grupo das 7 mutações de 2VUJ e 2VUL, além de outros casos especiais, como o resíduo 105 (localizado no cordão de 2VUJ e 2VUL) e a arginina 126 destas duas proteínas e a arginina 112 equivalente em 1XNB. Os histogramas foram gerados usando o programa Matlab®.

## Amostragem

As simulações de ATD foram feitas a partir de frames retirados de simulações de equilíbrio (Seção 2.2). Os gráficos apresentados são a média dessas medidas realizadas a partir de vários frames. O número de *frames* usados é dependente da proteína, já que algumas destas foram analisadas com maior detalhe. O número de frames iniciais utilizados em cada experimento é mostrado na tabela seguinte em função da proteína.

**Tabela 3** – Número de frames iniciais utilizado para cada proteína em função do experimento de ATD, para os sistemas em água ou em vácuo.

		Estrutura	2VUJ e 2VUL	1XNB	1F5J	1M4W
Sistema	Medida					
Vácuo	MDT		70	20	40	20
			20	20	-	-
	CCL	Ala	400	-	-	-
		Gly	3000 e 2000 <sup>1</sup>	2000 <sup>2</sup>	-	-
	CCL local	MDT	20	20	20	20
		CCL	20	-	-	-
Água	Ala		-	-	-	-
		Gly	300	-	-	-

<sup>1</sup> 3000 para as sete mutações e 2000 para resíduos 105 e 126

<sup>2</sup> Somente para o resíduo 112

## ATD em função do tempo

Para o ATD em função do tempo, foram rodadas simulações de 60 ps para todas proteínas, salvando as velocidades a cada 0,4 ps. Para vizualizar a trajetória, foram gerados alguns arquivos tipo *dcd*.\* Exceto por isso, todas as outras configurações são mesmas que as das simulações anteriores de ATD. Em água, apenas uma ou duas simulações foram realizadas para cada proteína, apenas para visualização. O aminoácido que foi acoplado ao banho térmico em cada simulação possui alguma particularidade no MDT. Já em vácuo, foram rodados as simulações dos aminoácidos mutados em 2VUJ e 2VUL (9, 13, 14, 18, 34, 35 e 71), usando 10 *frames* das simulações de equilíbrio como pontos iniciais.

\*Serão utilizadas no vídeo, Seção 2.5.

## 2.4 Relaxamento

A estapas de preparação das simulações de ATD podem gerar instabilidades locais nas estruturas, uma vez que a estrutura não foi reequilibrada após a retirada de água, nem após a diminuição da temperatura. A não equilibração foi uma escolha para que a estrutura da proteína preserve ao máximo sua estrutura da simulação de equilíbrio. Portanto, para analisar o estado em que se encontram as estruturas exatamente antes da simulação de ATD, simulações de relaxamento foram realizadas.

As estruturas foram simuladas em condições semelhantes ao ATD (como temperatura inicial de 10 K), porém, nenhum banho térmico foi introduzido, ou seja, agora é um sistema NVE. Simulações foram feitas com e sem moléculas de água. O resultado é apresentado na forma de temperatura dos aminoácidos em função do tempo, da mesma forma que o MDT. As velocidades (que serão convertidas a temperatura) foram gravadas a cada 2 ps. Foi realizada a média em 17 simulações de *frames* iniciais distintos em vácuo para as proteínas 1XNB e sua variante construída neste trabalho, 1F5J, 2VUJ e 2VUL. Como o MDT em água não apresentou grandes diferenças entre as proteínas, uma média de simulações a partir de 20 *frames* iniciais foi feita somente para 2VUL, e de 3 *frames* para 1XNB e 2VUJ.

Ainda, a fim de analisar o efeito da mudança da temperatura de 300 a 10 K na estrutura, foi rodado um relaxamento sem termalização. Foi realizado somente para a estrutura 2VUL, utilizando 20 *frames* retirados da simulação de equilíbrio de 40 ns. Neste caso, todas as configurações da simulação foram iguais às dos outros relaxamentos, exceto pelo fato de que as estruturas não passaram pelo processo de termalização à 10 K, mas somente foram inicializadas nesta temperatura. O sistema não possui água, então o possível deslocamento no potencial pela retirada do solvente ainda estará presente.

## 2.5 Vídeo

Um vídeo de uma simulação de ATD foi gerado, utilizando cores para as temperaturas em função do tempo. Uma simulação gerada como na Seção 2.3, porém com maior frequência na geração de posições e temperaturas, forneceu uma simulação de 60 ps. A estrutura, juntamente com os arquivos DCD de posição e velocidade são iniciados no programa VMD, e através de um *script* em tcl, cada resíduo fica colorido de acordo com sua temperatura, em função do frame. Foi usada a escala de cor BWR, ajustada para melhor visualização. Como é feito o escalonamento, não é necessário calcular a temperatura real, portanto, somente é calculada a energia cinética.

O vídeo se encontra no link: <http://www.youtube.com/watch?v=GknBvCpmue0>

## 2.6 *Scripts* e programas

Para tratar este conjunto grande de dados, pequenos *scripts* foram desenvolvidos utilizando as linguagens e programas TLC, BASH e Matlab® (71). Para rodar o experimento de ATD, *scripts* previamente escritos em TCL e FORTRAN por Leandro Martínez (1) foram adaptados. Gráficos aqui exibidos foram gerados nos programas Matlab® ou Grace, e normalmente foram aplicados filtros para suavizar as curvas, como a média dentro de 30 pontos. As figuras de estruturas foram feitas no programa VMD. Todos os alinhamentos estruturais, com intenção de visualização e não de cálculos, foram feitos através do programa online Multiprot (69), com as configurações *default*.



# Capítulo 3

## Resultados e Discussão

### 3.1 ATD no vácuo

O mapa de contatos é a distância entre pares de aminoácidos representada em escala de cinza, onde quanto mais próximos dois aminoácidos maior a intensidade da cor cinza. O mapa de difusão térmica é apresentado em escala colorida de temperatura. Dessa maneira, as regiões em comum entre ambos os mapas são correspondentes à difusão mais previsível, onde a energia térmica é transmitida por ligações covalentes e outras interações entre resíduos próximos, como ligação de hidrogênio, por exemplo. Nesta difusão simples, um aminoácido aquecido (abscissa) irá aquecer os aminoácidos vizinhos (diagonal principal), podendo se estender às regiões mais próximas, neste caso evidenciadas no mapa de contatos.

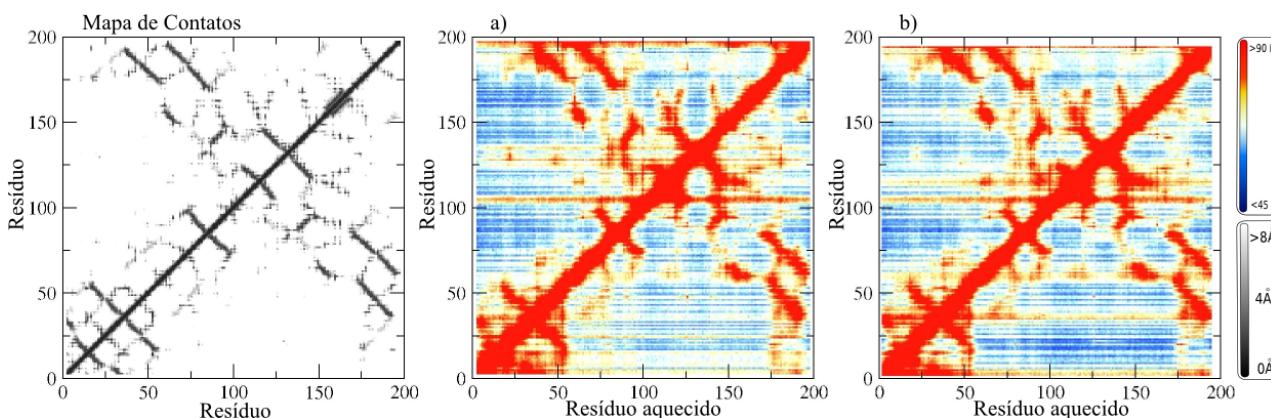
A escala do mapa de difusão térmica (MDT) foi escolhida de forma que as regiões de menor interesse, isto é, os aminoácidos próximos ao aquecido e, portanto no regime de difusão por contato, ficasse saturadas em relação à temperatura máxima. Dessa forma, o limite superior da escala de cores no MDT não corresponde à temperatura máxima global, mas sim à maior temperatura que não envolve as regiões de difusão por contato. Isto se deve ao fato que as regiões aquecidas por contato possuem temperaturas em média três vezes superiores ao resto do sistema (a proteína toda está inicialmente a 10 K, enquanto o banho térmico é de 300 K).

Em uma inspeção rápida dos mapas de difusão térmica, a primeira característica marcante é a existência de linhas horizontais, as quais mostram que, independente do aminoácido aquecido, a resposta da estrutura (eixo das ordenadas) é sempre similar. Dentre essas linhas que surgem no mapa de ATD, as quentes são as que mais se destacam, o que pode ser interpretado como um fenômeno de regiões acopladas, como proposto para sinalização intramolecular em uma proteína de domínio PDZ (33). Entretanto trata-se de relaxamento da estrutura, como efeito das etapas de dessolvatação e termalização precedentes ao ATD. Uma discussão mais detalhada a respeito das linhas quentes e sua relação com o relaxamento da estrutura pode ser encontrada na Seção 3.4. Assim, o calor está fluindo ou concentrando-se em determinadas regiões, as quais não aparecem no mapa de contato, e, portanto não se enquadram na classificação de difusões por contato. Então, pode haver proteínas com estruturas muito semelhantes, porém com linhas horizontais diferentes entre si, de forma que toda análise e

comparação entre estruturas estarão focadas neste aspecto.

Nesta secção serão mostrados os mapas de difusão térmica resultantes do ATD no vácuo para todas as proteínas deste estudo. Em especial, uma comparação entre as proteínas referência 2VUJ e 2VUL, e entre estas e a mesofílica. As proteínas 2VUJ e 2VUL são referência neste trabalho pois compõe um sistema interessante, uma vez que as pequenas diferenças entre suas estruturas geram grandes diferenças na estabilidade térmica. Detalhes de todas proteínas utilizadas neste trabalho podem ser revistos na Seção 2.1 da Parte 2.

### 3.1.1 2VUJ *versus* 2VUL

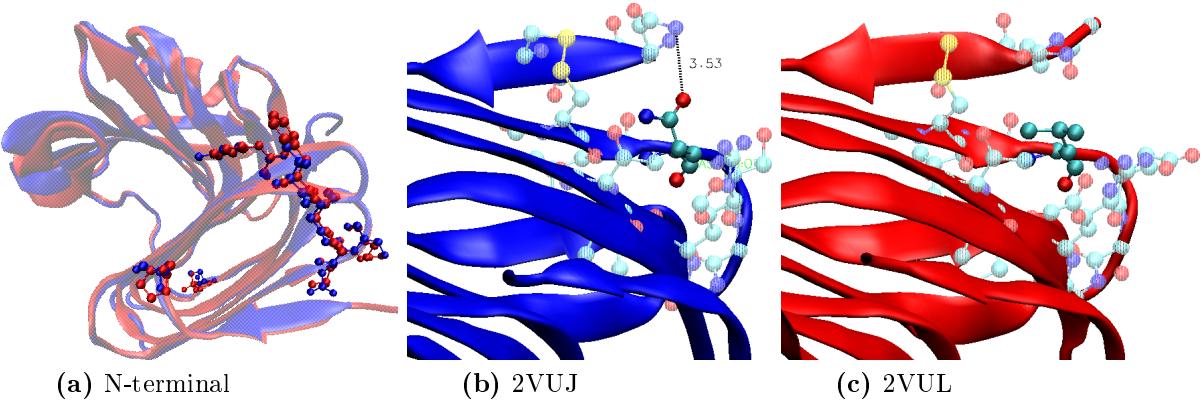


**Figura 6** – Mapa de contatos e MDT médio obtido de estruturas retiradas de 3 simulações de 40 ns distintas para a) 2VUJ e b) 2VUL.

Mesmo com estruturas e sequências muito parecidas, as proteínas 2VUJ e 2VUL apresentaram mapas de difusão térmica (MDT) no vácuo com sutis diferenças. Em geral, os MDT para essas duas proteínas são muito semelhantes, já que estas diferem entre si em apenas 7 aminoácidos. A 2VUJ possui um C-terminal com 3 resíduos de glicina mais longo e a 2VUL um N-terminal com um resíduo de alanina a mais.

A diferença no comprimento das regiões terminais pode explicar a alta temperatura, já que são regiões mais flexíveis e com menos impedimentos estereoquímicos, sofrendo processos de relaxamento rápidos. Nota-se este fenômeno principalmente na região C-terminal da proteína 2VUJ. Já o N-terminal de 2VUL com somente um resíduo a mais possui uma resposta mais quente no MDT. Como a diferença é de apenas um resíduo, parece mais razoável atribuir esta resposta no N-terminal às mutações, que estão muito próximas do mesmo (Figura 7). Em especial, a mutação Gln34Leu pode ter uma grande contribuição. A glutamina está próxima e suficiente de outro resíduo de mesmo tipo, na posição 3, com grande potencial para formar ligação de hidrogênio e estabilizar a ponta do N-terminal. Por outro lado, a leucina é apolar

e fica impedida de estabilizar o N-terminal por ligações de hidrogênio, tornando a região potencialmente móvel na estrutura 2VUL.



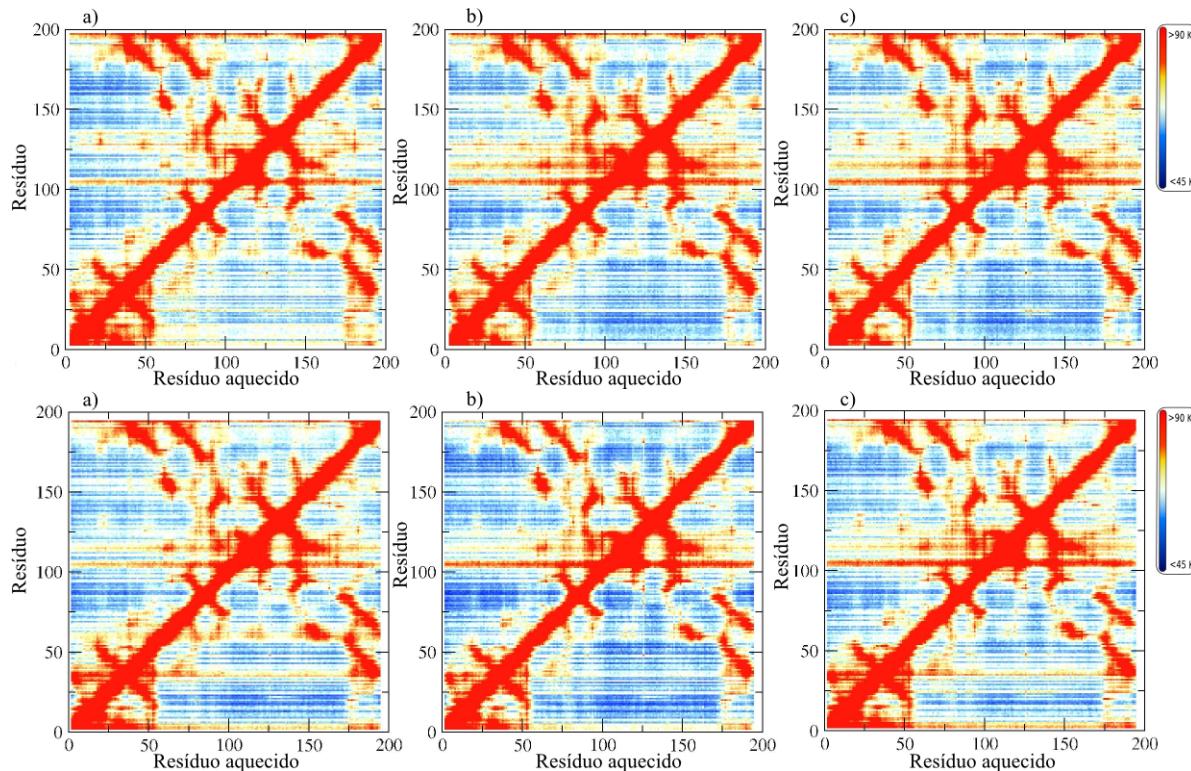
**Figura 7** – a) Mutações em destaque nas estruturas cristalográficas 2VUJ (azul) e 2VUL (vermelha) alinhadas. Resíduo 34 em destaque nas duas proteínas: b) glutamina em 2VUJ está próxima de outra glutamina, possibilitando ligações de hidrogênio, enquanto c) leucina em 2VUL é um resíduo hidrofóbico em um ambiente hidrofílico.

Além da mutação citada, a mutação Ser35Glu destaca-se nos MDT e relaciona-se às mutações do N-terminal. No mapa, a região em volta desta mutação forma uma linha horizontal mais quente em 2VUL. De fato, este resíduo está em um loop e a mutação para um resíduo de cadeia mais longa e carregada permitiu certas interações, discutidas em 3.6. Além destas duas principais diferenças devido a mutações, há também linhas quentes distintas acima do resíduo 100; as linhas de 112, e as duas depois de 125 aparecem, em ambas as proteínas, mas com intensidades diferentes. Não foi possível relacioná-las diretamente com as mutações e como será visto a seguir, a linha do resíduo 112 não é presente em todas simulações da amostragem (Figura 8).

Analizando também a média do ATD de *frames* retirados de simulações de equilíbrio distintas é possível analisar áreas que realmente são discrepantes entre essas duas proteínas das que não são. Uma maior amostragem diferencia o que é flutuação estatística do que não é, de forma que a média pode oscilar em torno de um ponto de temperatura mais alta ou mais baixa, ou seja, a média pode ser deslocada conforme se adicionam dados à distribuição.

A Figura 8 é formada por MDT médios para *frames* retirados de três simulações de equilíbrio. O padrão de linhas quentes tem uma certa semelhança entre si, mas algumas linhas praticamente somem em alguns mapas. Por exemplo, a linha quente por volta do 112 não aparece no primeiro gráfico de 2VUJ (Figura 8a). Análise dos gráficos individuais de cada *frame* (não mostrado) confirma a tendência representada na média, por toda a simulação. Ou seja, aquela simulação realmente não amostrou uma estrutura que geraria aquela linha

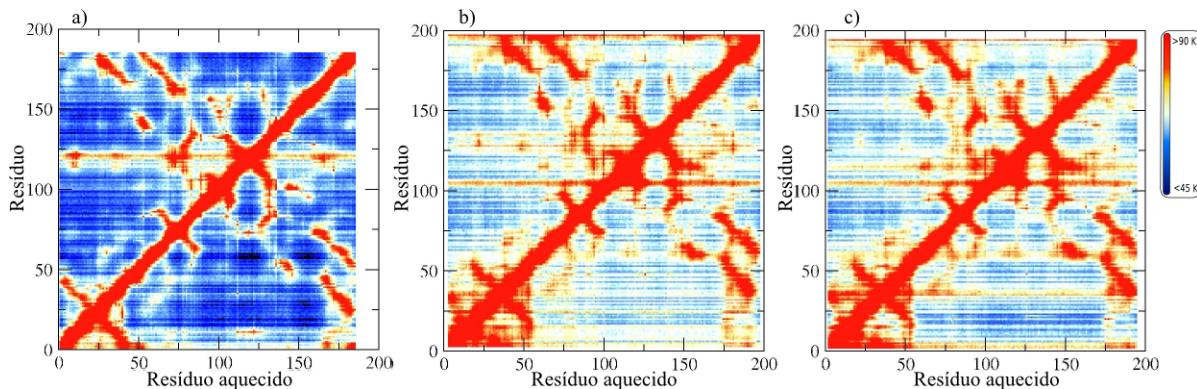
horizontal quente, resultando em um padrão diferente para a mesma proteína. Isso indica a importância da amostragem ser gerada não somente de uma simulação longa de equilíbrio para este método.



**Figura 8** – ATD nas proteínas 2VUJ (acima) e 2VUL (abaixo). Os mapas de difusão térmica (MDT) são de 20 frames para 3 simulações de equilíbrio distintas.

### 3.1.2 Mesofílica *versus* termofílicas

No intuito de localizar diferenças dependentes da estrutura para explicar diversos níveis de termoestabilidade entre proteínas da mesma família, o método é aplicado à proteína mesofílica 1XNB. Esta é comparada às mais termicamente estáveis (Figura 9). Pelo fato desta proteína ter menos resíduos, o alinhamento estrutural torna-se necessário para a análise das diferenças. O mapa da mesofílica apresenta algumas linhas quentes que podem ser relacionadas com as observadas em 2VUJ e 2VUL: o N-terminal é semelhante ao de 2VUJ, e além disso, a região 112 a 124 em 1XNB, equivalentes à 126 ao 137 nas outras proteínas, parece decrescer de temperatura com o aumento de termoestabilidade.



**Figura 9** – Comparação linhas quentes do MDT para a) 1XNB, b) 2VUJ e c) 2VUL.

Entretanto uma linha quente observável em ambas proteínas termoestáveis (em torno do aminoácido 104) não é visível em 1XNB, na qual deveria ser vista por volta do resíduo 91. Esta região de *loop*, nomeado *cordão*, é mais longa nas duas proteínas mais estáveis. A possibilidade de um *loop* maior contribuir para a termoestabilidade é incomum, já que geralmente se aceita que a diminuição dos *loops* aumenta a estabilidade das proteínas (72). Nas termoestáveis, esse *loop* possui três aminoácidos a mais, além de um aminoácido substituído para prolina (91 em 1XNB). Um *loop* mais extenso, normalmente significa uma maior flexibilidade, e a maior energia cinética apresentada nos mapas corrobora essa afirmação. Apesar do conceito geral de que proteínas mais termoestáveis são menos flexíveis, essa hipótese não é consensual (12, 73). Simulações com uma estrutura nativa e suas variantes termoestáveis sugerem que o mecanismo de estabilização é devido ao aumento de pequenas flutuações em conjunto com a diminuição dos movimentos de maior amplitude (74). Ainda neste trabalho, várias mutações que geraram maiores flutuações em relação à nativa estão em regiões de *loops*, corroborando a existência de *loops* mais flexíveis em proteínas mais estáveis. Mutações no cordão da xilanase XynR8 que substituíram o ácido aspártico (o qual fazia interação com uma glutamina) para glicina ou para asparagina (que deixaram de fazer essa interação), aumentaram a termoestabilidade (45). Aparentemente, a flexibilidade também foi aumentada, já que estas mutações permitiram um maior grau de liberdade para o loop. Extrapolando, essa flexibilidade local poderia ser alvo de um início de desnaturação, mas, possuindo uma cinética de enovelamento com mais que dois estados, o intermediário pode ser estável. Em relação a isso, a ideia de *molten globule* está sendo cada vez mais aceita e a flexibilidade que ela implica é um fator que contribui com a estabilização estrutural (75). Ou seja, a viabilidade de um estado semi-enovelado pode ser o mecanismo de estabilização térmica das proteínas mais estáveis (2VUJ e 2VUL).

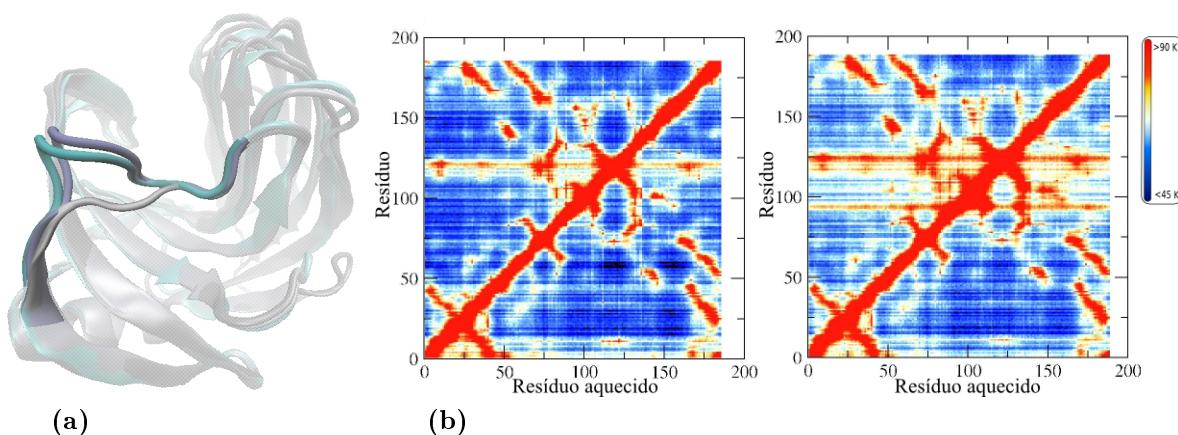
Um padrão no MDT que aparece em uma xilanase termoestável e também em outra hipertermoestável, pode indicar uma região importante para a estabilidade térmica. Esse loop

maior na região do cordão, cuja temperatura no ATD é maior que a média, também aparece em outra xilanase termoestável de *Thermomyces lanuginosus* (76). Na proteína 1F5J (termoestável) esse loop possui um tamanho intermediário, assim como esta última. Entretanto, a linha do MDT para 1F5J não é tão facilmente observada nesta região, provavelmente devido ao fato do loop não ser longo o suficiente. Dessa forma, essa diferença estrutural foi testada no método: o loop maior pertencente à termofílica foi inserido na estrutura mesofílica e em seguida foi feito o estudo de ATD. Na Figura 10, estão representadas as sequências originais e substituídas, alinhadas.

1XNB	90/PRO	<b>91/THR</b>	-	-	-	-	92/GLY
2VUL	101/PRO	<b>102/PRO</b>	<b>103/GLY</b>	<b>104/SER</b>	<b>105/ASP</b>	<b>106/GLY</b>	

**Figura 10** – Alinhamento entre 1XNB e a sequência das termoestáveis na região do cordão. Em vermelho está a sequência que foi adicionada.

A proteína mesofílica com a inserção do loop mais longo apresentou o MDT como o esperado. Uma linha horizontal característica aparece em torno do resíduo 91 (Figura 11). Assim, o resultado ajuda a fortalecer a hipótese da relação entre a presença deste loop e a estabilidade térmica. Se confirmado experimentalmente que esta estrutura modificada possui termoestabilidade maior, este é um método candidato para ser utilizado no design racional de proteínas termoestáveis.

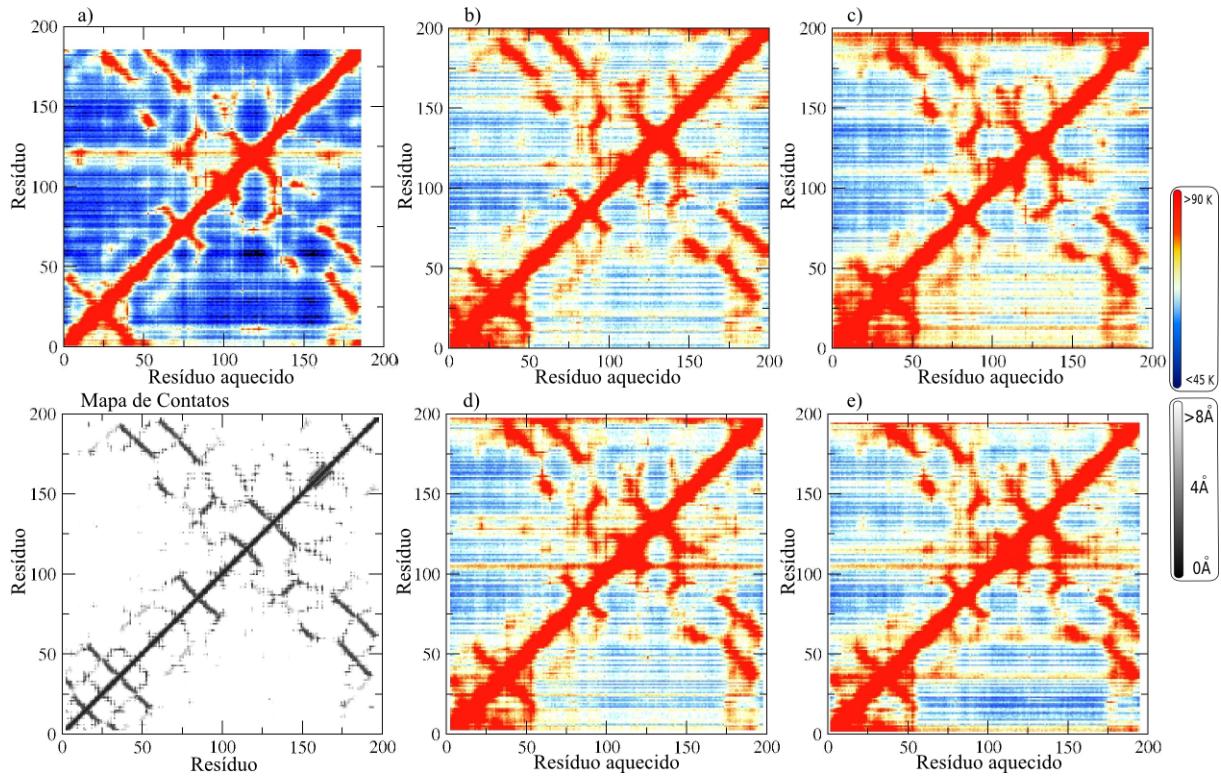


**Figura 11** – a) Estruturas alinhadas, 2VUJ, 2VUL e 1XNB (cinza). As proteínas mais estáveis possuem o loop maior, o que gerou uma linha quente no MDT, diferente de 1XNB; b) Novo MDT para a proteína mesofílica com a inserção do loop, sendo destacada a linha quente formada em relação à proteína nativa.

### 3.1.3 As outras proteínas

O mapa de contatos de proteínas estruturalmente similares é, naturalmente, similar. O MDT também o é, mas possui informações adicionais. As linhas horizontais se apresentam de forma particular para cada proteína (Figura 12).

As medidas experimentais de estabilidade térmica não são totalmente comparáveis, uma vez que os estudos empregaram diferentes métodos e diferentes critérios, já citados na Subseção 1.2.1. Então, aqui as proteínas são comparadas nos grandes grupos: mesofílica, termofílica e hipertermofílica (classificadas de acordo com a literatura - Tabela 2).

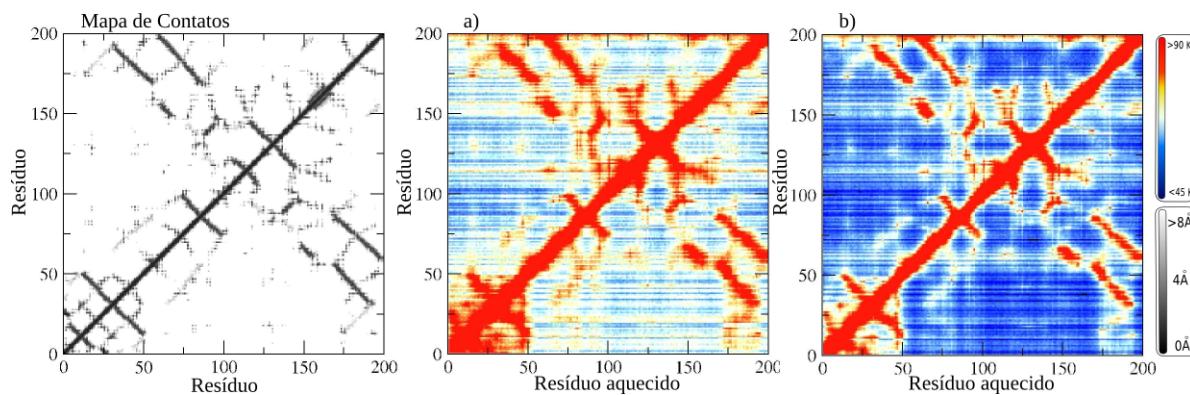


**Figura 12** – ATD em vácuo para as cinco proteínas a) 1XNB, b) 1F5J, c) 1M4W, d) 2VUJ e) 2VUL; mapa contatos para 2VUJ.

1F5J até pouco tempo era considerada a xilanase da família 11, de estrutura caracterizada, com maior termoestabilidade (77). Dois MDT para *frames* de duas simulações de equilíbrio apresentaram temperaturas médias discrepantes. Ou seja, na figura 13, nota-se que um mapa é mais azul e outro mais vermelho. Na verdade isso não representa um problema, já que essa diferença acredita-se ser proveniente da energia inicial contida na estrutura anterior ao ATD. Uma análise relativa à vizinhança dentro do próprio mapa resolveria o problema. De qualquer forma, os mapas para esta estrutura são pouco conclusivos, não apresentando nenhuma particularidade especial.

A região C-terminal da proteína termoestável 1M4W respondeu de forma quente no mapa, assim como 2VUJ. De fato, a cadeia carbônica final continua se distanciando da proteína em forma de *loop*, apresentando-se de forma coerente no MDT como uma região de alta energia cinética, dada a alta mobilidade. Seu N-terminal no mapa é muito semelhante à 2VUJ, com várias linhas quentes.

O N-terminal de todas as proteínas termoestáveis aqui analisadas possuem o mesmo comportamento, com exceção da hipertermoestável. Ao invés de várias linhas quentes, como apresentado em 2VUJ, 1M4W e 1F5J, 2VUL apresenta duas principais linhas quentes (no início e em 34), e a região intermediária é mais fria. O N-terminal é uma importante região para termoestabilidade em xilanases, sendo alvo de mutações (46). Uma menor temperatura nessa região nos MDT pode estar relacionada à estabilização térmica da proteína.



**Figura 13** – MDT para 1F5J. A diferença entre os dois mapas é somente a origem dos *frames*: a) e b) são mapas médios de *frames* retirados de simulações de equilíbrio distintas.

### 3.1.4 Conclusões

A simulação das estruturas no vácuo revelou diferenças entre estruturas muito semelhantes. As observações são consistentes entre as diferentes simulações de uma mesma estrutura e, geralmente, entre *frames* de uma mesma simulação (dentro de 100 ns de simulação de equilíbrio a 300 K, foram retiradas 70 estruturas igualmente espaçadas no tempo).

As regiões que apresentaram linhas horizontais quentes estão localizadas na superfície da proteína e, por isso, possuem maior liberdade para mover-se, aparecendo como regiões com maior temperatura nos mapas de MDT.

Nesta seção foi dado o primeiro passo para distinguir diferenças entre estruturas, associadas à termoestabilidade. Os resultados apontam para que as proteínas exibam diferentes respostas quando sujeitas às variações de temperatura e retirada do equilíbrio (no caso de retirada das águas). Tais respostas são dependentes da estrutura tridimensional da proteína

em questão. Assim, partindo da premissa em que a informação para diferir duas proteínas baseando-se em suas estabilidades térmicas está em suas estruturas, o resultado aqui obtido pode ser um indício e uma ferramenta para avaliar a termoestabilidade em proteínas estruturalmente semelhantes. Há também a possibilidade de relacionar o conteúdo proteico com a atividade dentro do ATD, como já mostrado no trabalho com receptores do hormônio da tireoide (1). Uma vez que não há impedimentos para que estas características estejam correlacionadas entre si e com a estrutura, o ATD mostra-se uma ferramenta flexível no estudo de propriedades proteicas dependentes de estrutura.

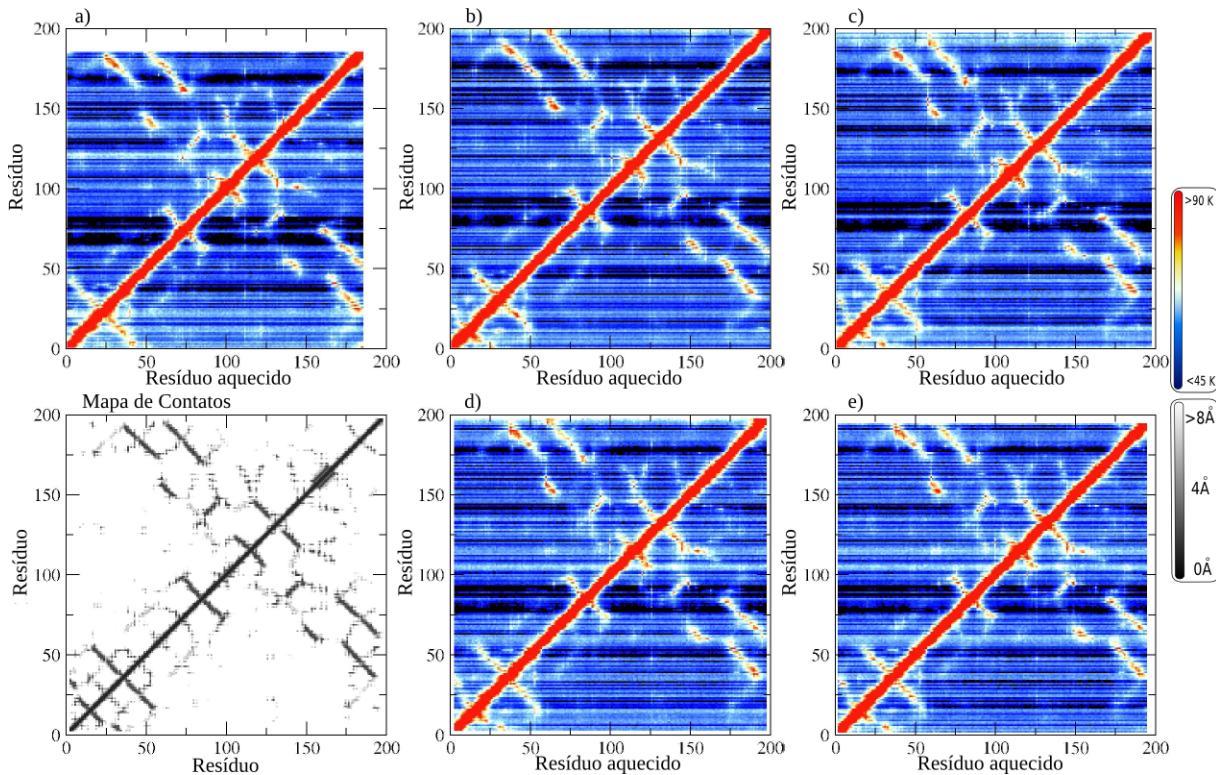
Ainda, o ATD apresenta-se como potencial instrumento para design inteligente de proteínas termoestáveis. Evidentemente, há a necessidade de maior desenvolvimento, mas como exemplo, uma estrutura proteica de maior termoestabilidade e relativa identidade estrutural pode ser submetida ao método paralelamente à proteínas menos estáveis. A partir da comparação dos MDT e suas estruturas, mutações podem ser sugeridas.

### 3.2 ATD na presença de solvente

Aqui mostraremos os resultados das cinco proteínas submetidas ao ATD na presença de água. Vale lembrar que a adição de água aumenta drasticamente o número total de átomos, portanto, o número de cálculos e por consequência o tempo computacional da simulação. Ainda assim, a metodologia do ATD repete a simulação para todos os aminoácidos acoplados ao banho térmico, o que leva a um aumento de cerca de 22 vezes no tempo computacional.

Os gráficos desta seção monitoram somente a temperatura dos aminoácidos. Em geral, as temperaturas dos MDT em solvente são menores que as temperaturas de mapas obtidos no vácuo, já que a água é um componente extra que atua na dissipação de energia térmica. A Figura 14 mostra uma certa convergência destes mapas e, ao contrário do ATD no vácuo, as linhas horizontais quentes não se destacam, enquanto que linhas horizontais frias, comuns a todas proteínas podem ser notadas. Provavelmente o comportamento em água é convergente devido à rápida difusão do calor para esta, ao invés de acumular a energia somente na estrutura proteica. O resultado é coerente com o obtido em vácuo, um vez que a maioria das linhas horizontais ainda são quentes se comparadas às regiões vizinhas (mais detalhes Seção 3.2.1).

### 3.2.1 Comparação ATD em água versus vácuo

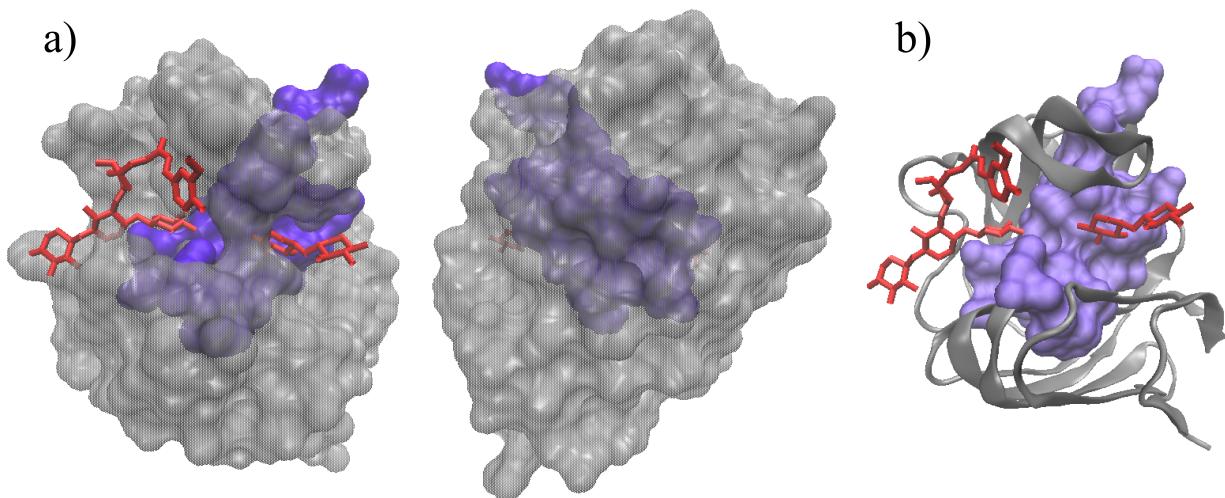


**Figura 14** – Mapas de difusão térmica para as cinco proteínas em água, a) 1XNB, b) 1F5J, c) 1M4W, d) 2VUJ e e) 2VUL. Em preto, as regiões mais frias (abaixo de 45 K) são semelhantes em todas proteínas.

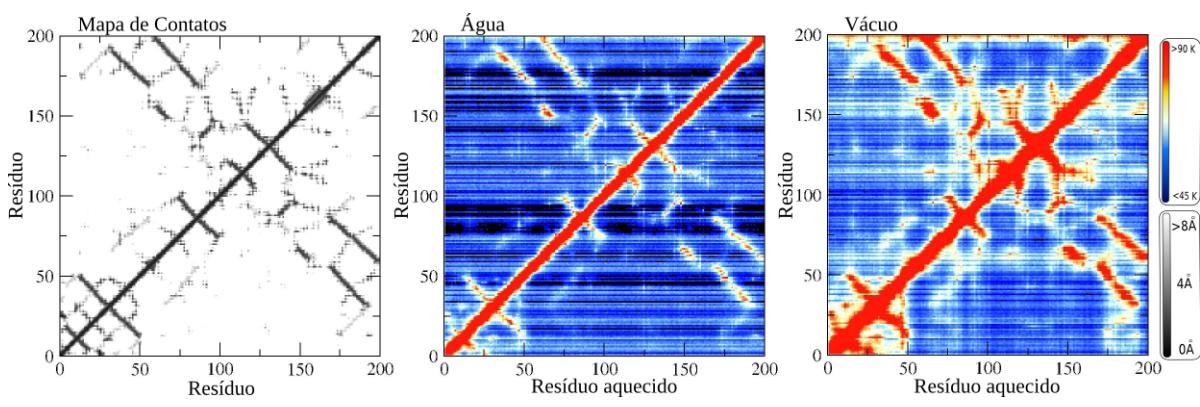
A introdução da água neste regime de não equilíbrio térmico promove o surgimento de algumas linhas horizontais frias (Figura 14), ou seja, regiões que sempre respondem ao estímulo térmico da mesma forma: adquirindo menos energia cinética que outras regiões. A região fria é comum a todas as proteínas e compreende o núcleo, contorna boa parte da superfície do sítio ativo e é formada por no mínimo uma parte da superfície da proteína (Figura 15), sendo esta última região variável entre as proteínas. É esperado que o núcleo de uma proteína globular possua uma estrutura mais compacta e relativamente imóvel, o que auxilia na estruturação proteica. Assim em um experimento como o ATD, é compreensível que o núcleo adquira menor energia cinética se comparado à superfície. No vácuo a energia não possui outras moléculas para onde fluir, ficando distribuída pela proteína. Porém com a adição de água ao sistema, a diferença entre temperaturas do núcleo e superfície é enfatizada , já que a água atua como um dissipador de energia na superfície. Dessa forma, a estrutura do sítio ativo é protegida do estímulo térmico, o que é bastante razoável.

A maioria das linhas quentes resultantes dos MDTs em vácuo ainda são encontradas nos

MDTs desta seção. Porém em água, a energia total da proteína é menor já que neste sistema há dissipação do calor da proteína para água (o que foi deduzido da menor temperatura nos MDTs). Por outro lado, no vácuo não há dissipação pois o tratamento clássico dos átomos aqui utilizado não reproduz a radiação térmica. Portanto, as linhas horizontais ficam menos evidentes, já que a variação de temperatura no mapa é menor (isto é melhorado se mudarmos a escala do mapa). Na Figura 16 são comparados lado a lado os MDTs em vácuo e em água para algumas proteínas. As linhas são coincidentes se fizermos uma análise relativa , ou seja, uma linha relativamente quente em vácuo ainda é mais quente que sua região vizinha no MDT em solvente.



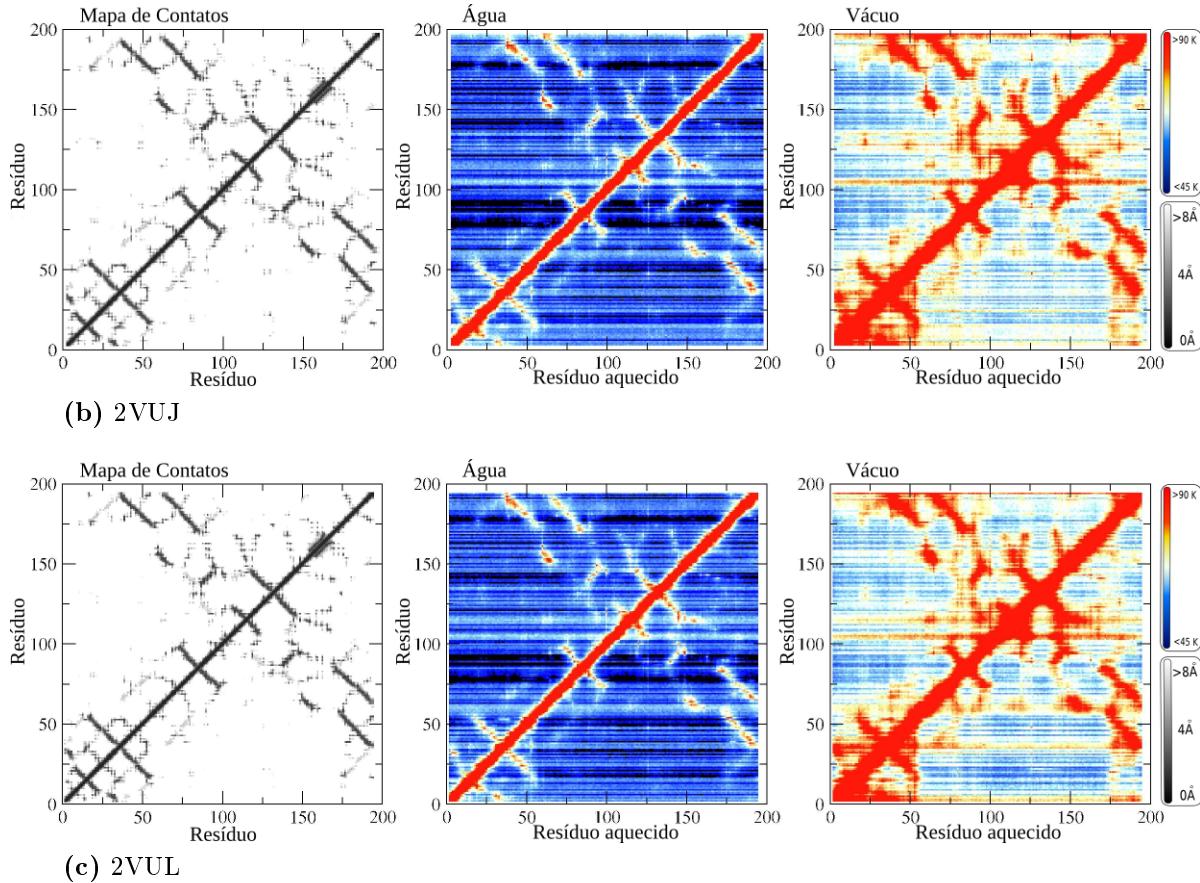
**Figura 15** – Estrutura da xilanase 2VUL com o ligante da xilanase 2VGD em vermelho, também da família 11; em destaque, em azul, as regiões que permaneceram frias independentemente do aminoácido aquecido no MDT em água. a) Representação “surf” no VMD e b) “new cartoon”, mostrando que a região fria encontra-se principalmente na palma da mão.



(a) 1F5J

(continua)

(continuação)



**Figura 16** – MDT para a proteína a) 1F5J, b) 2VUJ e c) 2VUL no vácuo e em água, respectivamente. A escala é 45 K a 90 K.

### 3.2.2 Conclusões

Apesar do ATD em água não ser capaz de apontar diferenças estruturais relacionadas à difusão térmica, este é capaz de mostrar a ausência de aquecimento em regiões específicas. Para as proteínas estudadas, o ATD em água indicou que a região do núcleo possui temperatura mais baixa, sugerindo uma menor mobilidade. O fato dessa região ser destacada somente no ATD em água sugere que essas proteínas (mesofílica, termofílica e hipertermofílica) podem usar este mecanismo de dissipação para água a fim de manter o núcleo proteico relativamente livre de perturbações térmicas e, portanto, mais estável.

Uma outra análise pode ser feita aumentando o tempo de simulação no ATD com água, pois talvez seja possível visualizar com clareza linhas que em vácuo são quentes, e tornam-se frias quando adicionada água ao sistema. Diferentemente do ATD em vácuo, o tempo do ATD em água não foi longo suficiente a ponto que no fim da simulação praticamente todos

os resíduos atingissem a temperatura superior determinada na escala (90 K ou acima). Uma discussão mais detalhada a respeito do tempo de simulação pode ser vista nos mapas do ATD em função do tempo (Seção 3.5).

### 3.3 Contribuição da Cadeia Lateral

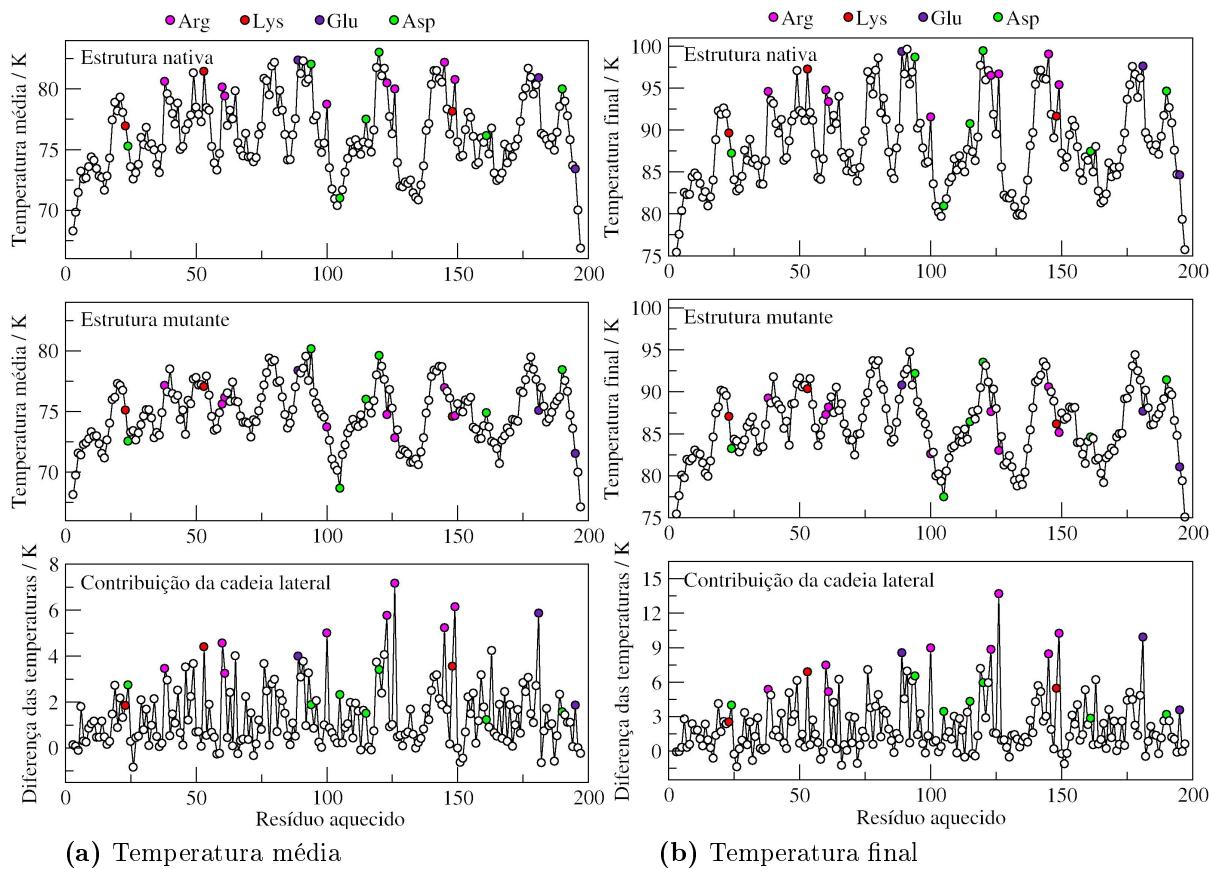
Para analisar a contribuição da cadeia lateral (CCL) na difusão térmica, uma mutação para alanina ou glicina é introduzida antes de acoplar o banho térmico para cada resíduo (1). Comparamos então as simulações de ATD da estrutura nativa e mutada: para cada aminoácido acoplado ao banho térmico, a temperatura final da proteína nativa é subtraída da proteína mutada. O resultado é então um gráfico de temperatura da proteína em função do aminoácido acoplado ao banho térmico. Vale ressaltar que quando a mutação é para glicina somente um hidrogênio estará no lugar da cadeia lateral, ou seja, é observada somente a resposta da cadeia principal; quando mutado para alanina, a cadeia lateral é um grupo metil, o que equivale ao carbono  $\beta$  presente em todos aminoácidos exceto a glicina. A diferença nas estruturas das duas simulações (nativa e mutada para Ala ou Gly) é a cadeia lateral, e, assim, é razoável dizer que a diferença entre temperaturas seja a contribuição da cadeia lateral.

Em destaque estão os resíduos arginina, lisina, ácido aspártico e ácido glutâmico: aminoácidos carregados e de cadeia longa. A seguir são apresentados os gráficos de análise da CCL para as proteínas deste estudo, em especial a comparação das proteínas 2VUJ e 2VUL.

#### 3.3.1 Análise geral

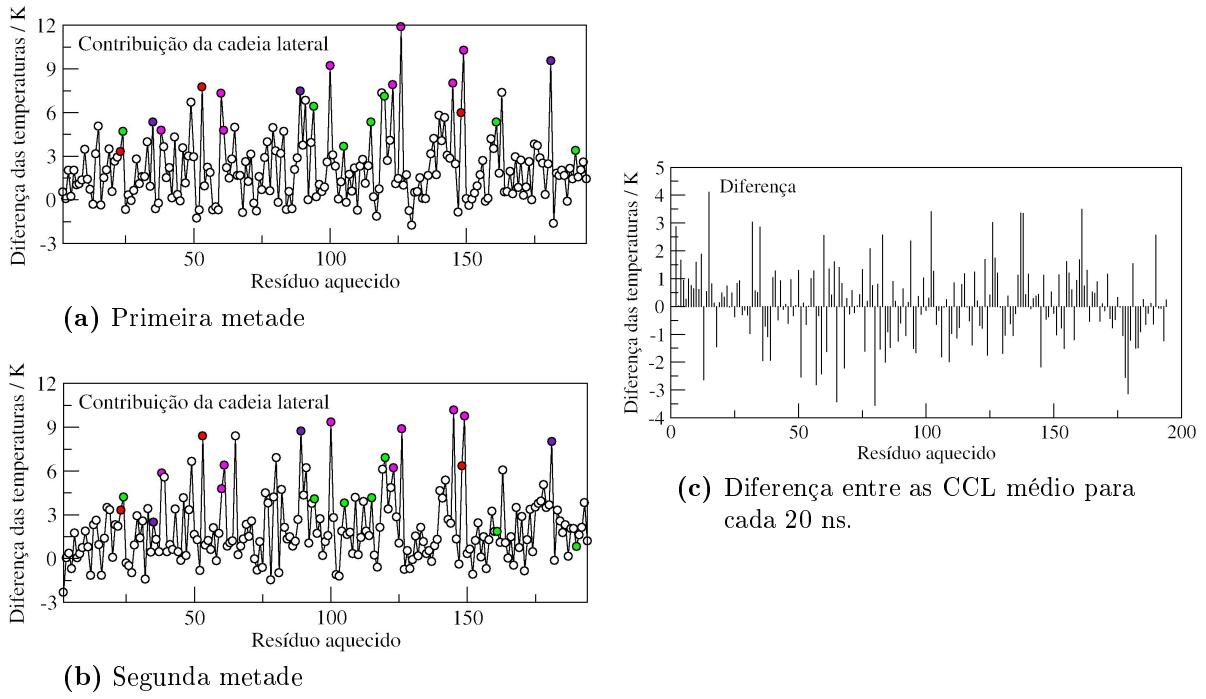
##### Comportamento temporal

Os gráficos possuem comportamento semelhante tanto para a temperatura final quanto temperatura média no tempo, exceto que no gráfico de temperatura final há uma maior separação dos pontos entre as proteínas. Por esta razão, serão apresentados somente os de temperatura final para todas as proteínas. Na Figura 17, onde mostra-se os casos de temperatura média no tempo ou temperatura final, unicamente como exemplo, pode-se notar que tanto a curva da proteína nativa quanto a da mutada para Gly ou Ala possuem a mesma forma em ambas as situações, porém deslocada para um valor maior nos gráficos de temperatura final, pois conforme o tempo aumenta, a proteína recebe mais energia e, assim, sua temperatura final fica cada vez maior.



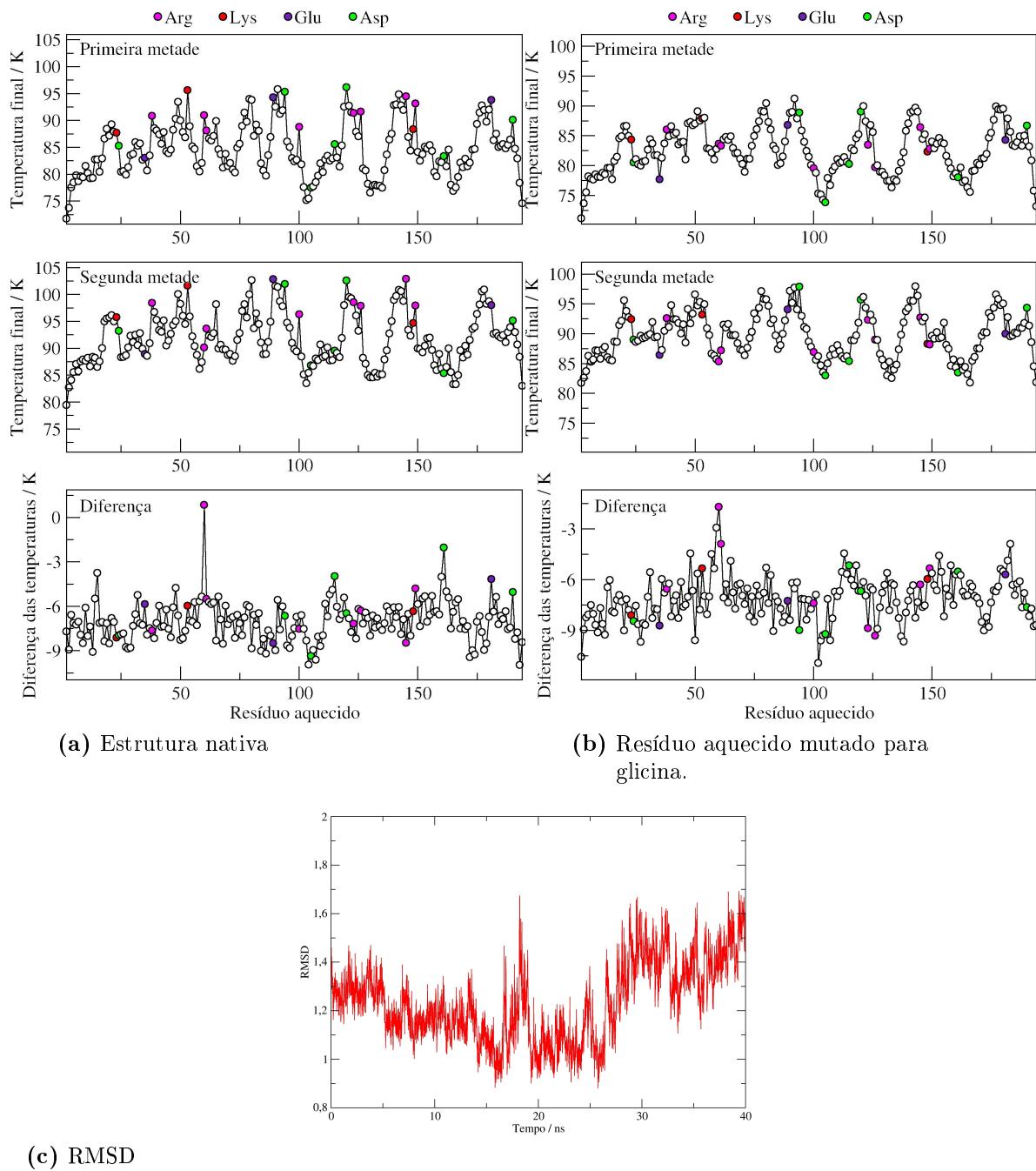
**Figura 17** – Gráfico de CCL para a proteína 2VUJ utilizando a mutação para glicina. À esquerda, a) a média temporal da temperatura, e à direita, b) a temperatura final. Na primeira linha estão os gráficos com a resposta da proteína nativa, na segunda, a resposta da proteína mutada na posição do aminoácido acoplado ao banho, e por último a diferença entre os dois, ou seja, a contribuição da cadeia lateral. Todos em função do aminoácido aquecido.

A análise de CCL reduz uma dimensão em relação ao MDT, quando, ao invés de apontar a temperatura para cada aminoácido dado um acoplamento, este retrata a temperatura média da proteína. Ou seja, uma linha vertical do MDT é reduzido a um ponto. Por isso, não é possível observar o comportamento das linhas horizontais dos MDTs aqui. É também muito dependente da estrutura inicial: aquelas que possuem RMSD um pouco maior em relação à média vão oscilar em torno de uma temperatura maior. Dessa maneira, é necessário que a comparação entre estruturas seja feita de forma normalizada. É válido ressaltar que qualquer normalização linear, feita em ambos os gráficos seria indiferente na subtração. Porém, a subtração exclui as contribuições iniciais relativas à temperatura, já que as duas estruturas são iguais, exceto pela mutação. O resultado final pode ser interpretado como CCL e comparado para diferentes proteínas.



**Figura 18** – CCL médio de *frames* da a) primeira ou b) segunda metade de uma simulação de equilíbrio de 40 ns para 2VUL. Ainda, c) a diferença entre as CCL de cada metade.

As Figuras 18 e 19 apresentam uma estimativa de oscilação da medida. Foi feita a média da CCL para *frames* de duas partes de uma simulação de 40 ns, separadamente. Assim em cada 20 ns, foram usados dez *frames* para o cálculo de CCL. Uma primeira análise da resposta da estrutura ao aquecimento é feita pelos gráficos 19a e 19b. O perfil da média para cada conjunto de *frames* de uma mesma proteína é semelhante, porém com temperaturas deslocadas. A diferença entre cada conjunto de *frames* (gráficos terceira linha) aponta um aumento de aproximadamente  $7,5 \pm 1,5$  K para a maioria dos resíduos. Mesmo com esse deslocamento, algumas respostas oscilam bastante e é necessário ter cautela no cálculo de CCL antes de fazer comparações com outras proteínas. O aumento na temperatura para o gráfico que utiliza a amostragem dos últimos 20 ns da simulação de equilíbrio pode ser devido a uma conformação de maior energia, já que este possui estruturas que mais se distanciam da estrutura média (Figura 19c). O RMSD dos 20 ns finais apresenta-se levemente maior que os 20 ns iniciais.



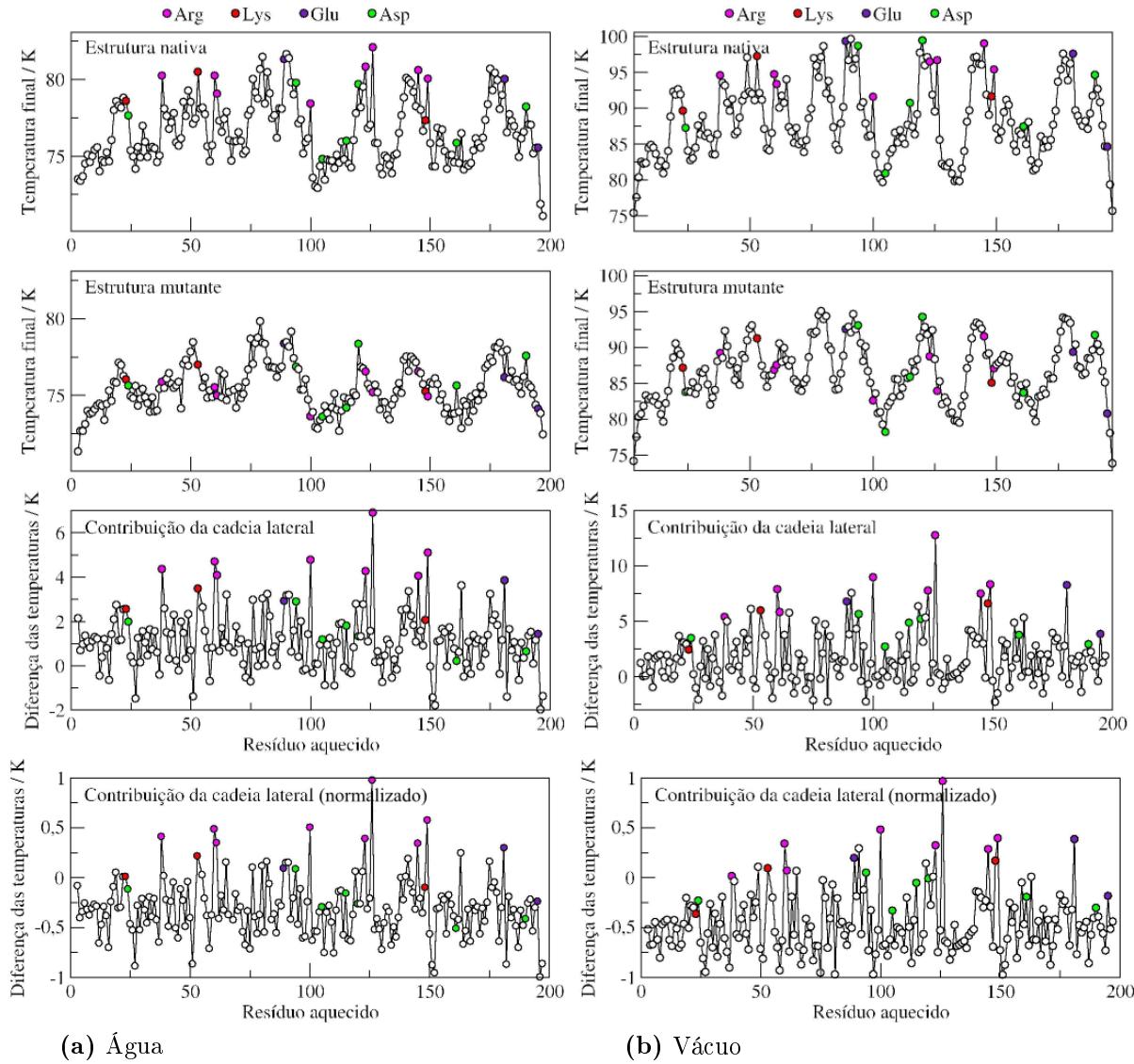
**Figura 19** – Temperatura final em função do resíduo aquecido para a proteína 2VUL a) nativa e b) mutada para glicina; média de 10 frames de duas metades (20ns) de uma simulação de equilíbrio (gráficos da primeira e segunda linhas), e a diferença entre elas (terceira linha). c) RMSD da proteína com alinhamento nos carbonos- $\alpha$  em relação à estrutura média no tempo.

Os gráficos acima mostram que essa medida é dependente da estrutura inicial, e a oscilação desta medida para uma proteína pode ser maior que a diferença entre a mesma e

outras proteínas. A Figura 18 mostra que a CCL de cada metade não convergiu em 20 ns de amostragem. Portanto, para comparar duas proteínas é necessário que a diferença entre elas seja maior que esta oscilação. Adiante (Seção 3.3.2), é visto que a CCL para 2VUJ e 2VUL são muito semelhantes. Por outro lado, alguns resíduos individuais apresentaram medidas de CCL semelhantes em todas as cinco proteínas aqui estudadas (não mostrado): os resíduos de Arg (destacados em magenta nos gráficos) possuem sempre maior valor de CCL, sugerindo que este aminoácido é um bom condutor térmico. Isso está de acordo com dados de CCL para as proteínas TR $\alpha$  e TR $\beta$  (1), onde estes resíduos também foram encontrados como sendo os de maior contribuição. Isso provavelmente deve-se ao fato de argininas possuírem uma cadeia lateral longa e carregada, com 3 grupos nitrogenados, permitindo maiores interações eletrostáticas, além de uma maior superfície de interação. A lisina, que possui menos nitrogênios, também apresenta CCL acima da média (Figura 18), porém sempre abaixo das argininas. Em um estudo de 1992, foi concluído que mutações lisina → argininas aumentaram a estabilidade da proteína xilose isomerase (78).

### Comportamento em água

Comparando os sistemas em água e em vácuo (Figura 20), depois de 30 ps, a temperatura total da proteína no ATD é sempre maior se o sistema está no vácuo. Entretanto, o perfil das curvas nesses diferentes cenários são bem semelhantes, principalmente quando normalizadas em relação ao seus máximos e mínimos. Isso indica que, de modo geral, o papel da água foi retirar energia de forma proporcional e sem preferências de regiões ou direções. De outra forma, poderíamos esperar que aminoácidos que entram em contato com a água pudessem levar a uma menor temperatura para a proteína, pois o aquecimento do mesmo poderia levar a um potencial processo de dissipação de energia térmica à água. Entretanto não é possível afirmar que as diferenças entre os gráficos são representativas dentro da oscilação da medida.



**Figura 20** – CCL para ATD realizado em a) solvente e b) na ausência deste, para a proteína 2VUJ, utilizando mutação ALA. Na última linha o gráfico da CCL normalizado para seus máximos e mínimos de cada gráfico.

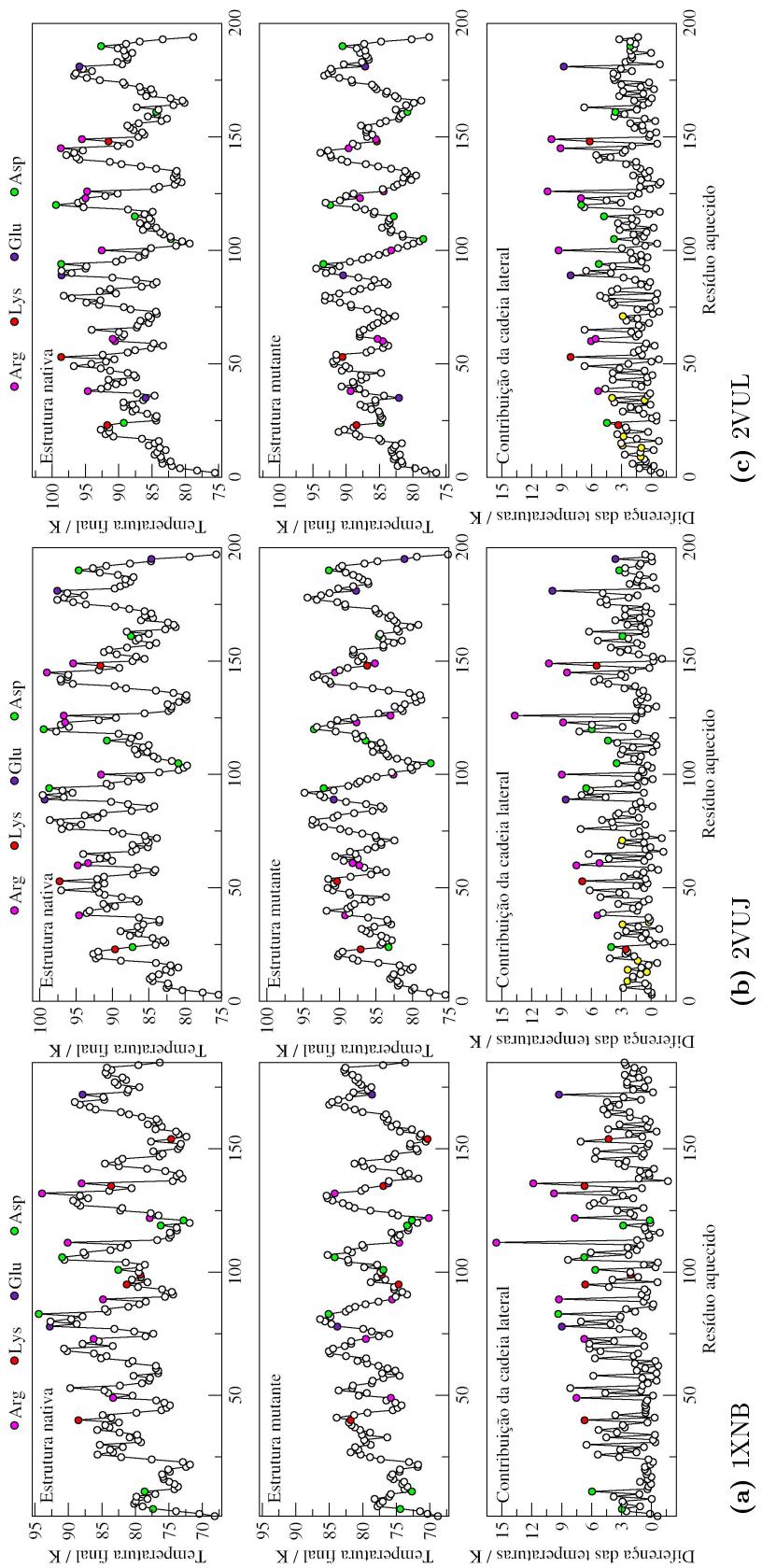
### 3.3.2 Comparação mesofílica, termofílica e hipertermofílica

A comparação novamente é baseada nas proteínas 2VUJ e 2VUL, já que a segunda foi desenvolvida a partir da primeira, além da proteína mesofílica 1XNB. Para verificar a contribuição da cadeia lateral, foram feitas mutações para glicina e alanina. Também foi feita a mutação para isoleucina, já que este é um aminoácido com maior frequência em proteínas termoestáveis. Porém nenhuma das três mutações proporcionaram padrões de CCL discrepantes entre estas proteínas. Os experimentos com isoleucina não foram bem sucedidos, já que, como

este possui cadeia lateral grande e não há equilibração após a mutação, muitas simulações resultaram em erro de convergência de cálculos e seus resultados foram prejudicados. Para alanina e glicina, nota-se que os gráficos são muitos semelhantes. Na Figura 21 mostra-se uma comparação entre 1XNB, 2VUJ e 2VUL, utilizando CCL com mutação para glicina.

O gráfico de CCL para as duas proteínas mais estáveis são bem semelhantes (Figura 21). Com exceção da Arg 126, que na hipertermoestável possui menor CCL, com uma diferença de 3,5 K com 2VUJ. Na mesofílica, este resíduo é equivalente ao 112, e também possui alta CCL. Ou seja, a CCL parece ser decrescente em função do aumento da termoestabilidade para este resíduo. Contudo uma maior amostragem seria necessário para confirmar essa afirmação. Na Subseção 3.3.3, histogramas com 2.000 pontos são feitos para a CCL do resíduo 126.

Esta medida possui uma oscilação natural, e a diferença entre estas duas proteínas é menor que esta oscilação. Portanto, não faz sentido fazer a comparação entre elas utilizando a CCL, já que esta não possui sensibilidade suficiente para a diferenciação com base na termoestabilidade. A sensibilidade alcançada através do ATD foram as linhas horizontais e uma vez que a medida de CCL é a temperatura total da proteína e não a temperatura em função dos resíduos, essa diminuição de dimensão limita o conteúdo das informações. A quantização da medida da oscilação foi feita de forma grosseira nesta etapa, porém, histogramas foram construídos na próxima Subseção 3.6.2, permitindo melhor visualização e comparação.



**Figura 21 – CCL para 1XNB, 2VUJ e 2VUL, nesta ordem. Em amarelo, as mutações que distinguem as duas últimas. Os resíduos foram mutados para glicina.**

### 3.3.3 CCL individual

Uma forma de analisar medidas que oscilam entre valores próximos é fazendo um histograma, porém é necessário uma grande quantidade de dados. Desta forma, para melhor interpretar o resultado de CCL, várias simulações foram feitas focando principalmente os resíduos diferentes entre 2VUJ e 2VUL. São usados de 300 a 3.000 pontos na amostragem para cada resíduo acoplado ao banho térmico\*, possibilitando ver distribuições e suas respectivas variâncias. A CCL individual é medida tanto como temperaturas finais e médias, e apenas um resíduo por vez é analisado, permitindo quantificar a oscilação e comparar duas proteínas com maior precisão. Relembmando que o estímulo do ATD é local: um banho térmico é acoplado somente ao aminoácido em questão, simulando um estímulo externo totalmente localizado.

A seguir serão discutidos os histogramas da CCL individual para os resíduos 9, 13, 14, 18, 34, 35 e 71 (mutações entre 2VUJ e 2VUL), para uma região de linha quente comum para as duas proteínas no MDT (resíduo 105) e também para Arg 126, que aparentemente apresenta comportamento diferente nas proteínas.

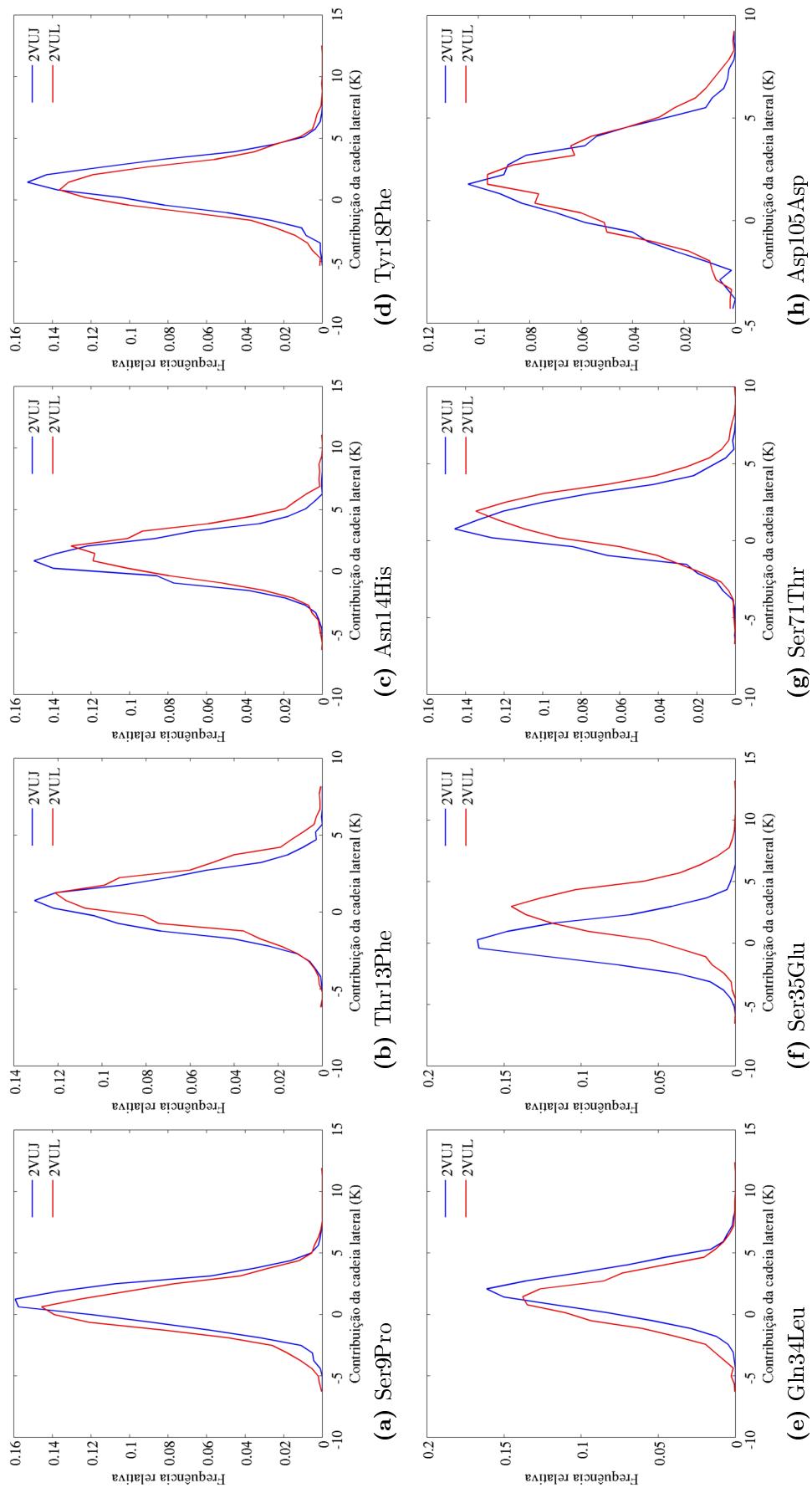
#### Contribuições das cadeias laterais em vácuo

Quando avaliado a semelhança com medida já feita (mutações para Gly), decidiu-se por uma amostragem menor para Ala, já que apenas 400 pontos nos histogramas com esta mutação foram suficientes para obter uma boa representação da distribuição da CCL individual. Observa-se que independentemente de ser feita mutação para Ala (Figura 23) ou Gly (Figura 22), as CCL individuais são equivalentes e em geral concordam entre si.

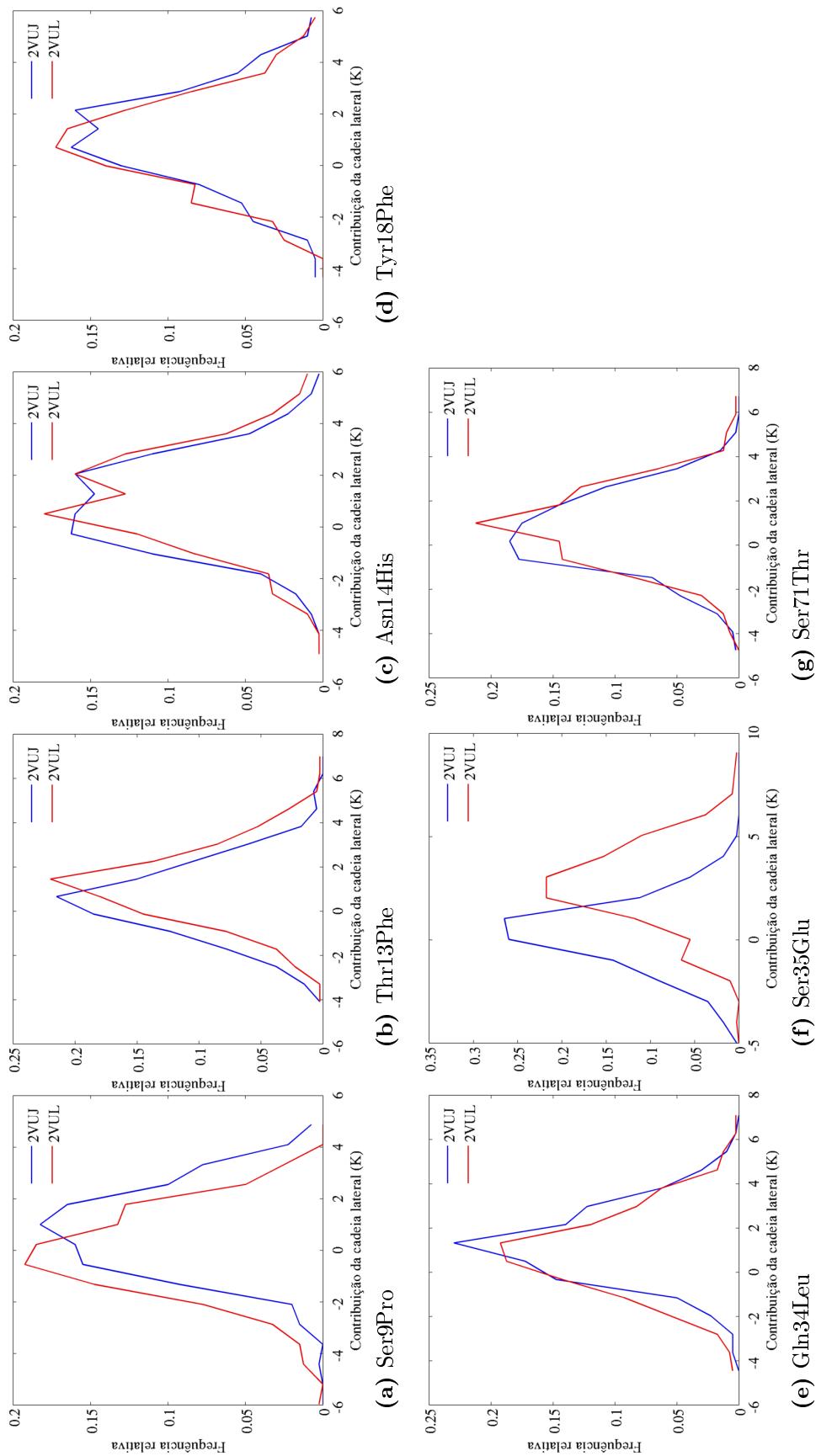
Cada aminoácido mutante de 2VUJ e 2VUL foi acoplado ao banho térmico do ATD e a CCL foi obtida. Também, o resíduo 105, localizado no cordão, que se apresenta quente nos mapas de ambas proteínas é utilizado como controle, pois é esperada a mesma resposta nas duas proteínas. Somente os histogramas associados aos resíduos na posição 35 são claramente distinguíveis. Na proteína hipertermoestável, a CCL deste aminoácido é maior. Ou seja, a troca de serina para ácido glutâmico (Glu), em uma região de loop, proporciona um aumento da energia cinética total na proteína, através de estímulo térmico neste resíduo. Isso corrobora os MDTs, que apresentam uma linha quente na região do resíduo 35 somente para 2VUL. Postiores análises da simulação de equilíbrio na Seção 3.6 dão mais detalhes sobre o comportamento do resíduo 35 nas duas proteínas.

---

\*Neste caso, quatro mil simulações de ATD foram feitas para cada resíduo a fim de gerar os dois mil pontos de CCL, por exemplo. Utilizando duas mil estruturas iniciais diferentes, retiradas de diferentes *frames* da simulação.

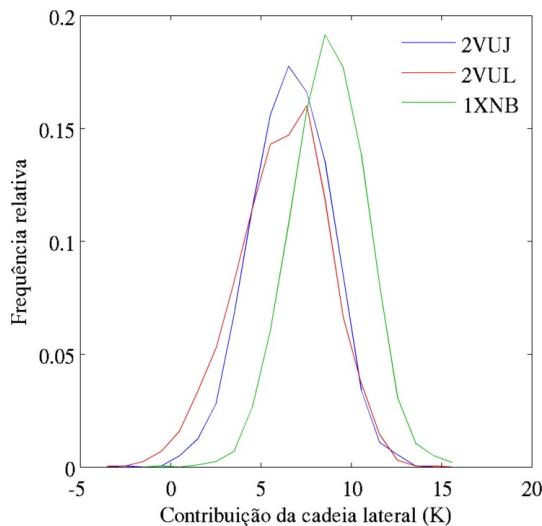


**Figura 22 – Histogramas para CCL de aminoácidos individuais utilizando a mutação GLY. Em ordem, a)-g) as sete mudanças de 2VUJ e 2VUL, h) o aminoácido 105 que é a região do cordão, onde as duas proteínas possuem uma linha vermelha no MDT. A indicação dos aminoácidos correspondentes à 2VUJ está à esquerda, e à 2VUL, à direita no número. Temperatura média no tempo.**



**Figura 23** – Histogramas para CCL aminoácidos individuais utilizando a mutação ALA. Em ordem, as sete mudanças de 2VUJ e 2VUL; temperatura média no tempo.

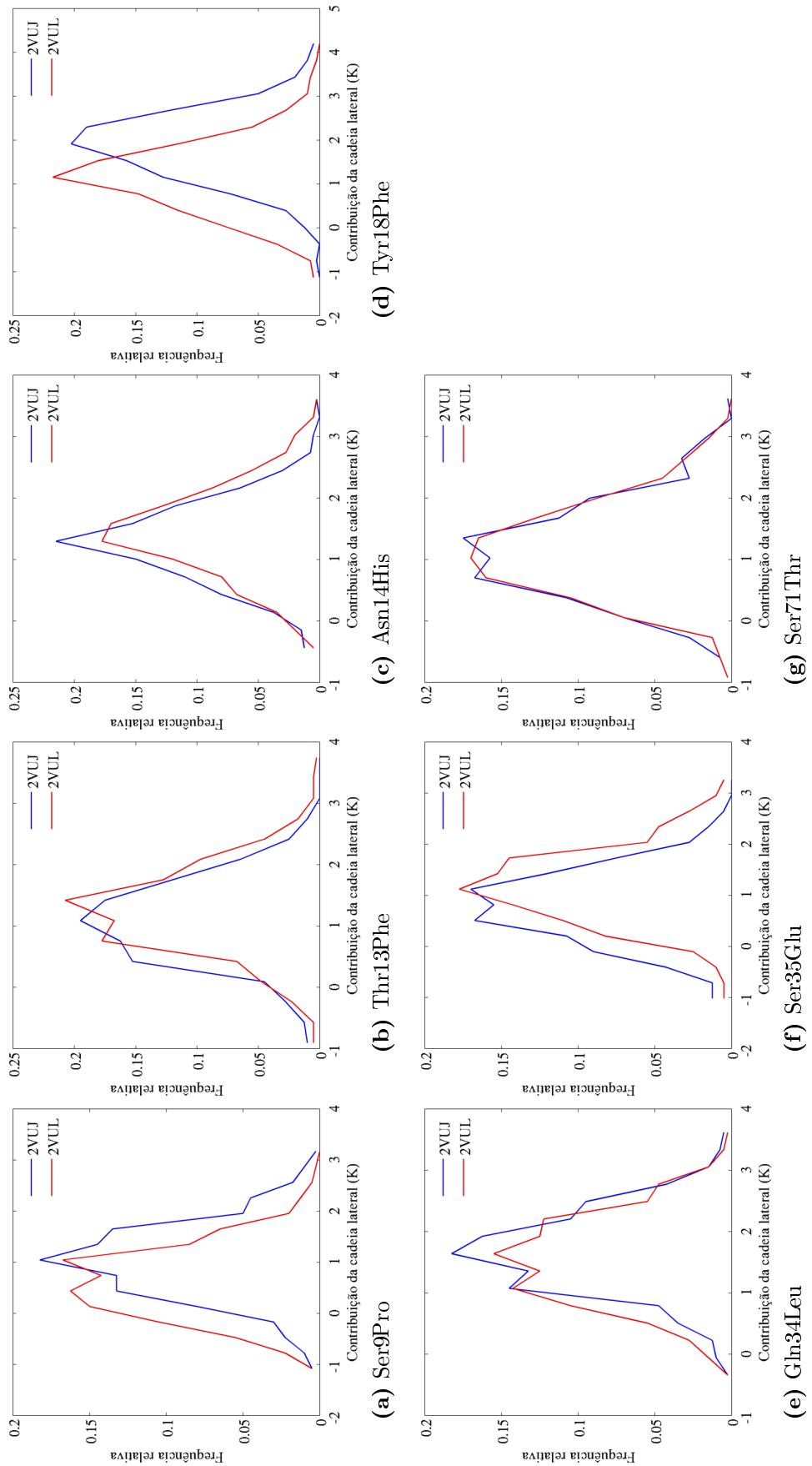
Ainda, a Arg 126 que previamente parecia ter diferentes contribuições para cada proteína (Subseção 3.3.2), apresentou histogramas sobrepostos para 2VUJ e 2VUL. Esse é um ponto que apresentou diferenças consideráveis entre 2VUJ e 2VUL nos gráficos de CCL em função do resíduo aquecido, mas não no histograma, indicando que a maior amostragem eliminou as diferenças. Porém, este realmente apresenta uma maior CCL para 1XNB (Figura 24) .



**Figura 24** – A Arg 126, previamente apresentada com diferente CCL para as proteínas mesofílica (1XNB), termofílica (2VUJ) e hipertermofílica (2VUL) quando representado o CCL em função do resíduo aquecido. O resíduo equivalente em 1XNB sera a Arg 112. Foi utilizado a média da temperatura.

### Contribuições das cadeias laterais em água

Devido ao alto custo computacional, somente 300 pontos para cada resíduo foram obtidos em água e somente com a mutação para glicina. Como já observado na Subseção 3.3.1, a CCL em água é mais atenuada, provavelmente devido à dissipação de energia para a água. Os histogramas de CCL para a mutação 35 apresentam uma leve diferença, principalmente na temperatura final (não apresentado), concordando com o CCL em vácuo. Porém diferentemente do que foi observado no CCL em vácuo, há uma diferença significativa entre os histogramas da mutação Tyr18Phe. A fenilalanina apresenta menor CCL que a tirosina quando o sistema é simulado em água. Isso indica que na presença de água esses aminoácidos reagem de modo diferente do que no vácuo. Em comparação com a fenilalanina, a tirosina possui um grupo hidroxila extra exposto ao solvente, e pode formar ligações de hidrogênio com moléculas de água, provavelmente levando a uma maior CCL. Esta conclusão está de acordo com o resultado calculado na Subseção 3.6.2.



**Figura 25** – Histogramas da CCL para aminoácidos individuais acoplados ao banho térmico. Mutação para Gly, média no tempo.

### 3.3.4 Conclusões

A medida da contribuição da cadeia lateral para o aquecimento da proteína foi obtida para as estruturas 2VUJ e 2VUL. Estas proteínas são muito semelhantes entre si, tendo RMSD do alinhamento de  $C\alpha$  de apenas 0,24 Å. Porém estas proteínas possuem diferença de 25 °C em suas respectivas temperaturas de *melting* (65). Os gráficos de CCL em função do resíduo aquecido não possuíram sensibilidade para diferenciar as proteínas, pois não foi possível distinguir as oscilações das medidas das suas diferenças entre as proteínas. Contudo, com uma amostragem maior foi possível construir histogramas para os aminoácidos mutados, e a mutação Ser35Glu gerou uma distribuição de CCL significativamente distinta entre as duas proteínas em vácuo. Em água, além dessa mutação, a Tyr18Phe também apresentou histogramas levemente separados, podendo indicar um mecanismo de interação com a água importante para a termoestabilidade. O significado da CCL para a termoestabilidade não está muito claro, pois somente duas de sete mutações apresentaram distribuições significativamente diferentes, sendo em um caso com diferença negativa e o outro positiva.

Para as cinco proteínas estudadas, todos os valores de CCL para resíduos de Arginina foram acima da média, assim como encontrado previamente em proteínas não termoestáveis (1). Esse fato não parece estar claramente relacionado com a termoestabilidade da proteína, apesar de todas as argininas equivalentes mostrarem temperatura de CCL decrescente em função do aumento da termoestabilidade. Os histogramas para a arginina 126 apresentaram-se equivalentes em 2VUJ e 2VUL, porém com maiores valores de CCL para 1XNB.

A medida de CCL apresentou maior custo computacional em relação ao MDT por dois motivos. Devem ser duas simulações ao invés de uma, pois a CCL é a subtração da resposta da estrutura nativa da mutada. Além disso para observar a resposta em um único resíduo, uma amostragem de no mínimo 300 pontos foi necessária para computar a distribuição de CCL, ao passo que para construir um MDT, um conjunto com 20 estruturas iniciais já é suficiente.

## 3.4 Relaxamento

Este experimento de relaxamento é semelhante ao de ATD, porém sem o acoplamento térmico em resíduos. De forma que a simulação da estrutura anterior ao ATD, livre de estímulos térmicos, pode fornecer informações sobre a qualidade da estrutura. Vale ressaltar que a estrutura retirada a partir da simulação a 300 K é uma representação da mesma em condições naturais e a não equilíbrio a 10 K é uma escolha proposital visando a menor variação possível da estrutura nativa. A maneira que a mesma é processada até o ATD, como a reti-

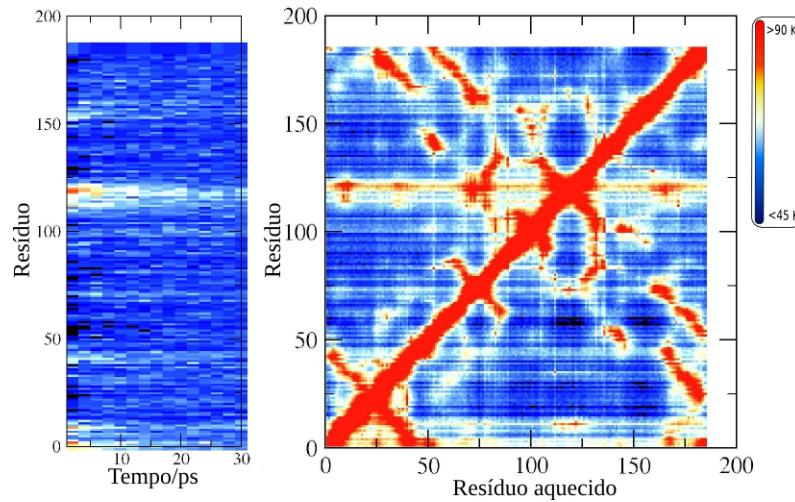
rada das águas e a diminuição da temperatura, ambos feitos sem uma equilíbrio posterior, podem levar à instabilidade de alguns elementos estruturais. Assim, se por um lado a diminuição da temperatura é um artifício indispensável para melhor visualização da resposta ao calor difundido, já que há diminuição da energia total da proteína, por outro lado a não equilíbrio e a retirada da águas podem gerar regiões de alto potencial. Durante a análise ambos fatores devem ser levados em conta a fim de evitar uma interpretação errônea dos resultados.

Cada aminoácido da mesma estrutura usada no ATD, seja em água ou em vácuo, teve sua temperatura monitorada em função do tempo. A fim de avaliar a influência da termalização na estrutura, foram também utilizadas estruturas que não foram termalizadas. Questões fundamentais surgem a respeito da análise dos MDT, e essas situações ajudam a explicar particularidades do ATD, principalmente do MDT. De forma que se pode questionar: a origem da energia cinética representada como temperatura no mapa é puramente do banho térmico ou energia potencial da estrutura? Mesmo através da tentativa de minimizar a energia inicial, este experimento demonstra que a forma de minimização escolhida para gerar estruturas iniciais pode carregar regiões não equilibradas.

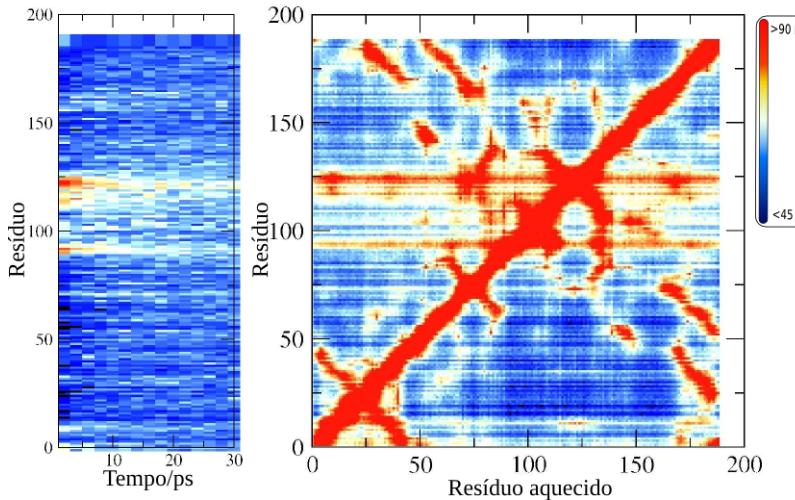
### 3.4.1 Relaxamento em vácuo

Foram gerados gráficos de temperatura de cada aminoácido em função do tempo para as proteínas termofílicas 1F5J, 2VUJ e 2VUL, e para a mesofílica 1XNB. A fim de garantir comparação entre estes e os gráficos de ATD, a mesma escala de cores foi usada. Na Figura 25 os gráficos de relaxamento para cada proteína foram colocados lado a lado com os de ATD.

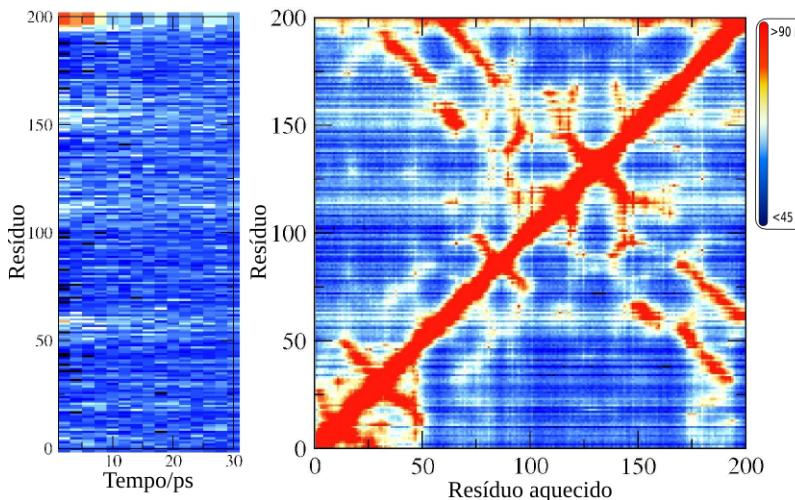
Logo no início dos gráficos de relaxamento há regiões de maior temperatura que estão localizadas exatamente nas mesmas regiões de linhas horizontais quentes dos MDTs. Portanto isso indica que aquelas linhas com alta temperatura nos MDTs não são consequentes do banho térmico no ATD. Principalmente dois fatores são atribuídos como responsáveis pelo surgimento dessas regiões quentes. Primeiro, a retirada das águas exatamente anterior à simulação de ATD faz com que a estrutura possua regiões em altas energias potenciais, as quais relaxam em forma de energia cinética e equilibram-se em menores potenciais. Segundo, a estrutura sofre uma mudança brusca de temperatura. Então, a nova energia cinética de cada átomo não é mais consistente com a energia potencial local. Embora acredite-se que este segundo efeito seja menos influente, uma vez que os efeitos de alta energia inicial não são observados nos experimentos de relaxamento em água (Subseção 3.4.2).



(a) 1XNB



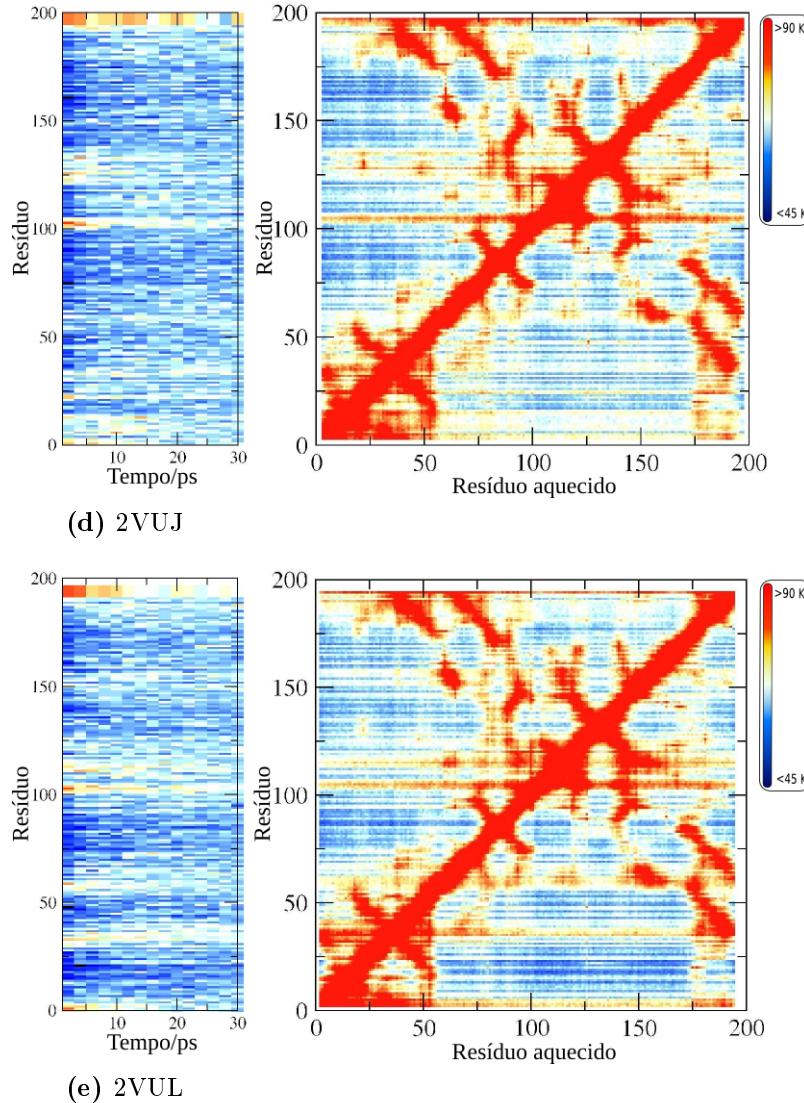
(b) 1XNB com a inserção no cordão



(c) 1F5J

(continua)

(continuação)



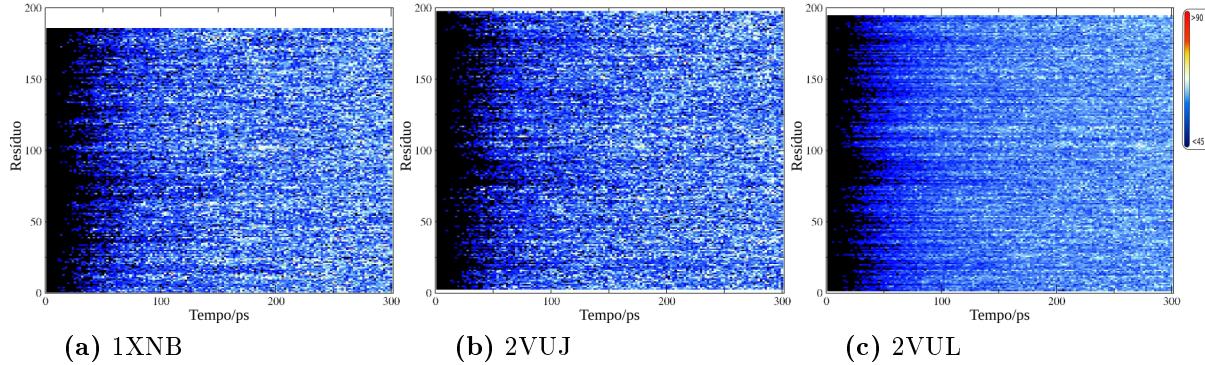
**Figura 25** – Gráficos de relaxamento e à direita os MDTs, ambos em vácuo. a) 1XNB, b) 1XNB com a inserção no cordão, c) 1F5J, d) 2VUJ e e) 2VUL.

### 3.4.2 Relaxamento em água

Na Figura 26 estão os gráficos de relaxamento em água para as proteínas 1XNB, 2VUJ e 2VUL. Nestes gráficos, a estrutura simulada é a anterior ao ATD em água, isto é, já termalizada a 10 K e ainda com as moléculas de água.

Em oposição ao resultado em vácuo, o relaxamento em água não mostrou regiões quentes no começo da simulação. Todos os resíduos começam com uma temperatura menor que 45 K e após 20 ps a temperatura média oscila em  $52 \pm 10$  K. De fato, as linhas horizontais quentes não são marcantes no MDT em água, como visto na Seção 3.2. Dessa maneira, as

linhas quentes observadas no ATD em vácuo provavelmente são originárias de regiões com alto potencial que surgiram pela retirada da água.



**Figura 26** – Relaxamento em água para as três proteínas a) 1XNB, b) 2VUJ e c) 2VUL.

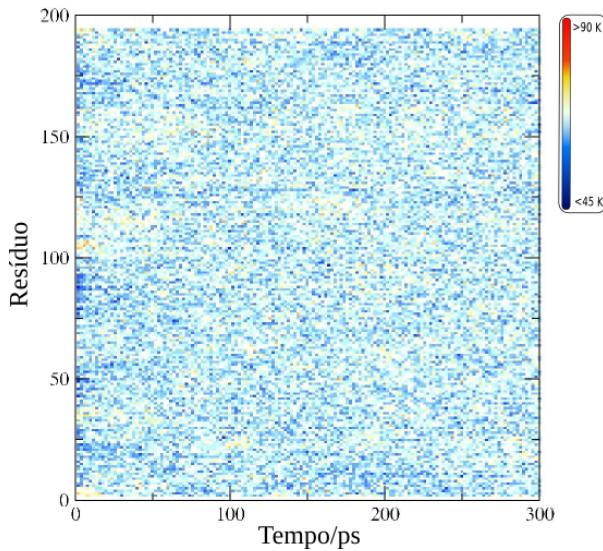
Todos foram médias de 3 *frames*, enquanto que para a última foi feita a média em 20 *frames*, por isso o gráfico é mais suave.

As regiões quentes do relaxamento em vácuo exibem, desde o início da simulação, temperatura diferente do restante da proteína. Entretanto para as simulações em água, o cenário é um pouco diferente. As linhas frias não aparecem como regiões de temperatura inicial diferente, porém estas permanecem a uma menor temperatura por mais tempo enquanto outras regiões aumentam de temperatura. A região em que isto é mais claro é a contida entre os resíduos 75 e 90. Essa região pertence ao núcleo da proteína, como visto na Seção 3.2 na página 55, corroborando que estas regiões de menor flexibilidade são menos propensas a adquirirem energia cinética.

### 3.4.3 Relaxamento sem termalização

Este experimento foi feito para avaliar o efeito da termalização no processo do ATD. No caso sem termalização, a estrutura passa por uma diminuição abrupta de temperatura, ou seja, ela foi retirada da simulação de equilíbrio à 300 K, para uma situação em que a velocidade inicial dos átomos é atribuída tal que o sistema esteja a 10 K (pela distribuição de Boltzman, como explicado na Seção 1.1). Esse experimento não conta com o reescalonamento de velocidades (termalização), ou seja, a mudança gradual da temperatura do sistema. A temperatura inicial apenas foi atribuída para 10 K. O sistema é em vácuo e, portanto, carrega as consequências da retirada das águas.

A Figura 27 mostra que ainda existem as regiões com maior temperatura inicial. Portanto, o reescalonamento de velocidades não é o responsável pelo surgimento destas.



**Figura 27** – Relaxamento no vácuo para as estruturas anteriores à termalização de 10 K , para 2VUL. Média em 20 frames.

### 3.4.4 Conclusões

O relaxamento da estrutura mostrou que as linhas horizontais quentes dos MDTs não são geradas por uma resposta térmica ao aquecimento que independe do aminoácido aquecido, mas sim pela relaxação da estrutura anterior ao ATD. Esse resultado não altera as conclusões feitas no MDT em vácuo, já que este não deixa de ser uma resposta sensível para estruturas semelhantes que sofreram os mesmos processos de simulações.

Atribui-se à retirada da água sem posterior equilíbrio da estrutura como sendo a principal razão para existirem regiões de maior energia inicial. Esta atribuição está baseada no fato do relaxamento em água não apresentar regiões iniciais com temperaturas maiores que o resto da proteína. Outra possibilidade poderia ser a descontinuidade da temperatura, isto é, passagem abrupta de 300 K para 10 K. Contudo, o relaxamento no vácuo sem termalização já apresenta regiões quentes. De fato, as regiões de linhas quentes estão localizadas na superfície da proteína e possuem maior liberdade para mover-se, fazendo interações diretamente com a água. Também, estas regiões provavelmente são as mais móveis da superfície da proteína, dado pela maior energia cinética em relação ao restante. Maiores detalhes sobre a mobilidade podem ser encontrados na Seção 3.6.1, na qual foram realizadas medidas de RMSD.

Também, é possível concluir que a energia da proteína após a retirada da água é bem maior que em água, sendo mais um argumento de que a retirada da água aumentou a energia potencial da estrutura, a qual relaxou através da transformação da potencial em energia cinética.

## 3.5 ATD em função do tempo

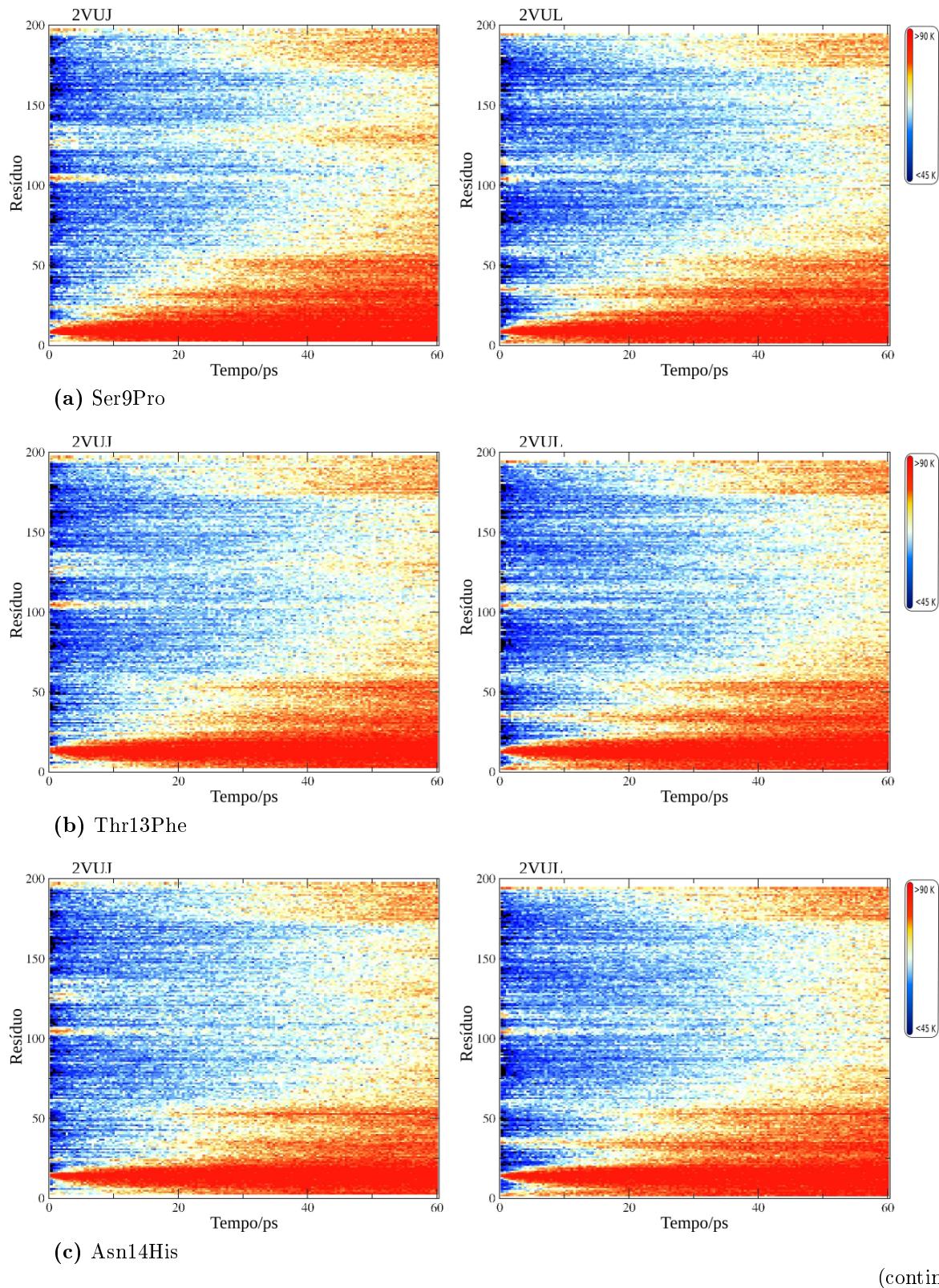
Até agora, os mapas gerados para o ATD são valores médios ou somente um ponto no tempo. A fim de avaliar o ATD em função do tempo uma dimensão deve ser abandonada. Dessa forma, são gerados gráficos de simulações individuais de ATD utilizando somente um resíduo aquecido e observando a resposta dos demais em função do tempo. Estes gráficos permitem obter estimativas temporais, como por exemplo o aquecimento de regiões vizinhas ao banho térmico com o passar do tempo. É possível também distinguir quais regiões aparecem quentes no MDT por possuir energia inicial daquelas que recebem energia do banho térmico. O tempo para os gráficos desta seção é duas vezes maior que o tempo de simulação no ATD e permite a visualização do momento em que os resíduos atingem 90 K ou mais, a temperatura máxima usada na escala de cores.

Os gráficos de ATD no tempo (Figura 26) devem ser interpretados com algumas premissas. A linha mais quente, cuja existência pode ser observada desde o início, corresponde ao aminoácido acoplado ao banho térmico. Além disso, regiões não necessariamente próximas ao resíduo aquecido já possuem uma energia inicial maior e, portanto, não apresentam-se mais quentes pela transferência de energia do banho térmico, como já discutido na Seção 3.4.

### 3.5.1 Mutações no banho térmico

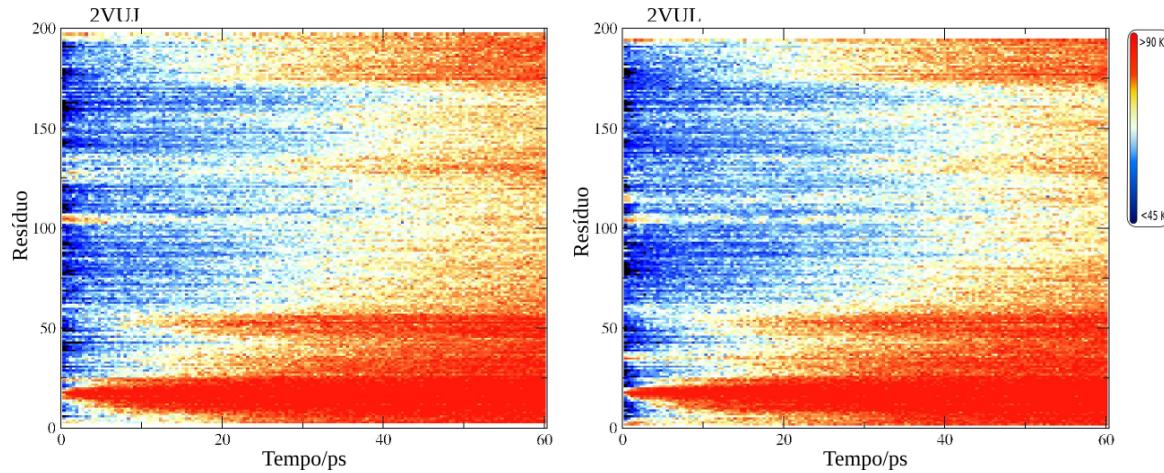
Os resíduos escolhidos para serem aquecidos foram as mutações entre as proteínas 2VUJ e 2VUL e a resposta de toda a proteína em função do tempo foi monitorada. Assim, através da análise da resposta temporal das proteínas quando suas mutações são excitadas, espera-se compreender a diferença entre a estabilidade térmica destas duas proteínas. Na Figura 26 são colocados lado a lado os gráficos de ambas as proteínas. Excluindo as regiões inicialmente a maior temperatura, a difusão de calor partindo dos aminoácidos mutados é praticamente a mesma para as duas proteínas.

Se os primeiros picossegundos de simulação, os quais apresentam linhas quentes devido à energia inicial da estrutura, forem ignorados, não há diferença significativa entre os gráficos de ATD no tempo para as diferentes proteínas. Para a mutação no resíduo 71, as regiões de alta energia inicial ficam mais evidentes, já que o resíduo 71 encontra-se distante destas e as outras mutações encontram-se mais próximas. Assim, na mutação 71 (Figura 26g) é possível observar diversas linhas quentes iniciais no N-terminal: em 2VUJ há energia inicial por volta do resíduo 24, 14 e nos primeiros resíduos; em 2VUL há regiões de alta energia nos resíduos iniciais e por volta de 35.

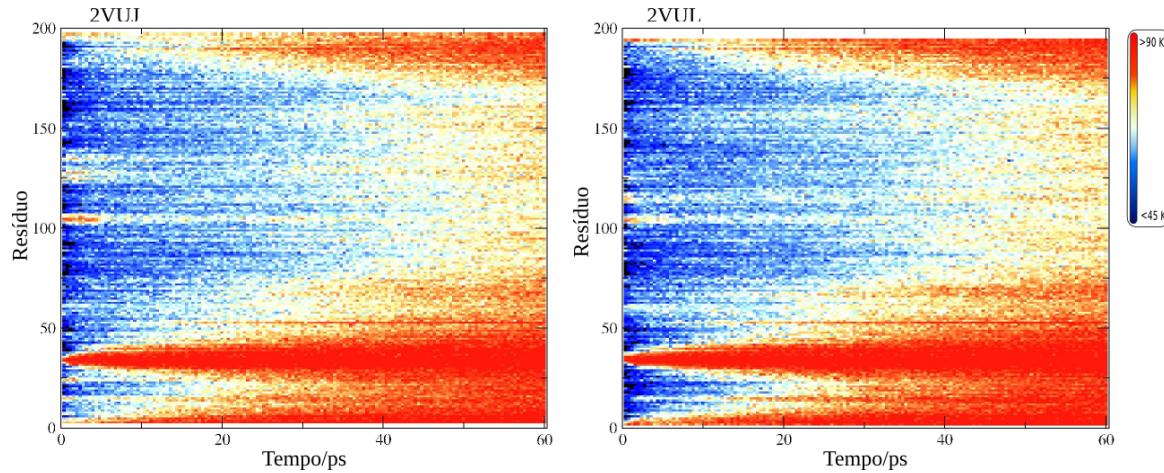


(continua)

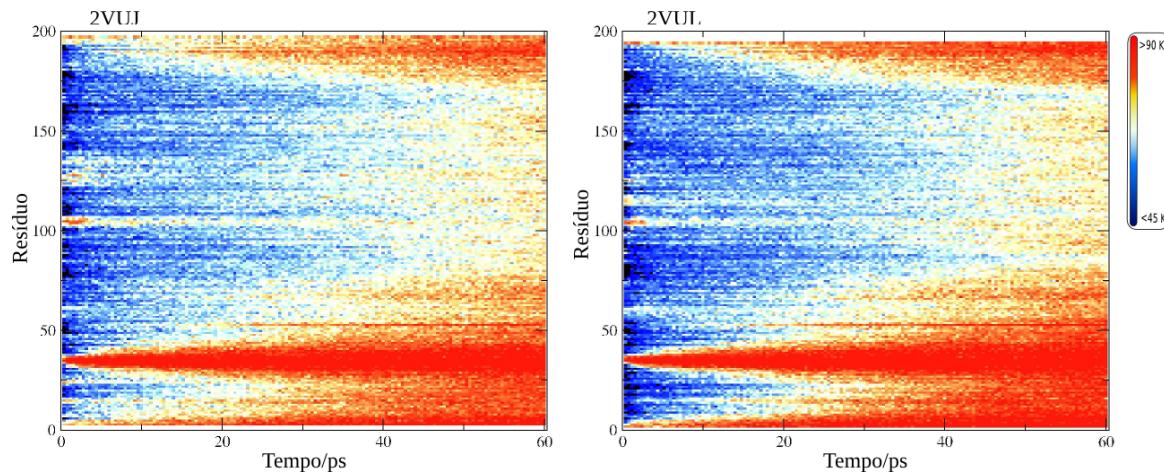
(continuação)



(d) Tyr18Phe



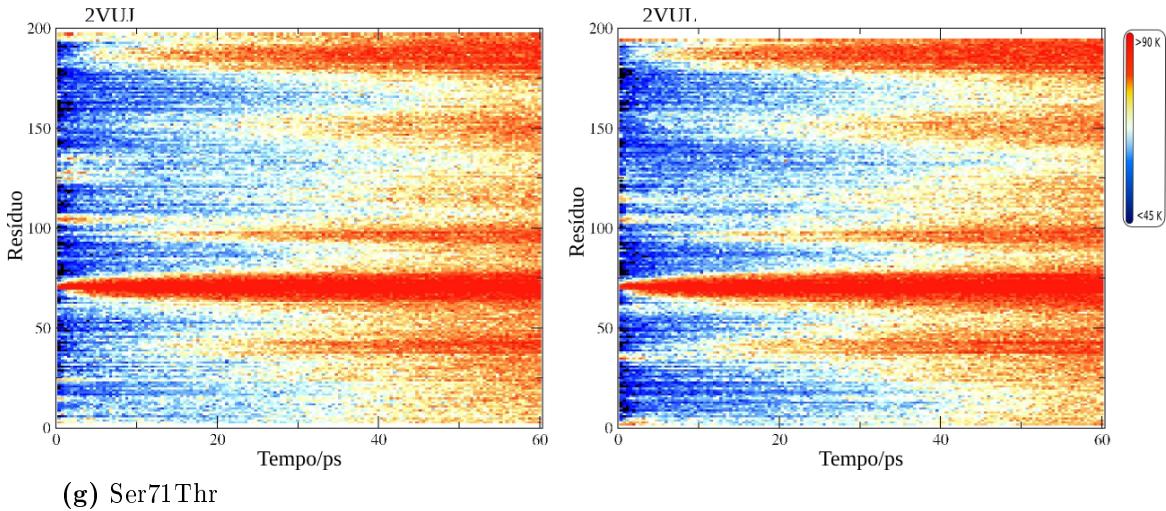
(e) Gln34Leu



(f) Ser35Glu

(continua)

(continuação)



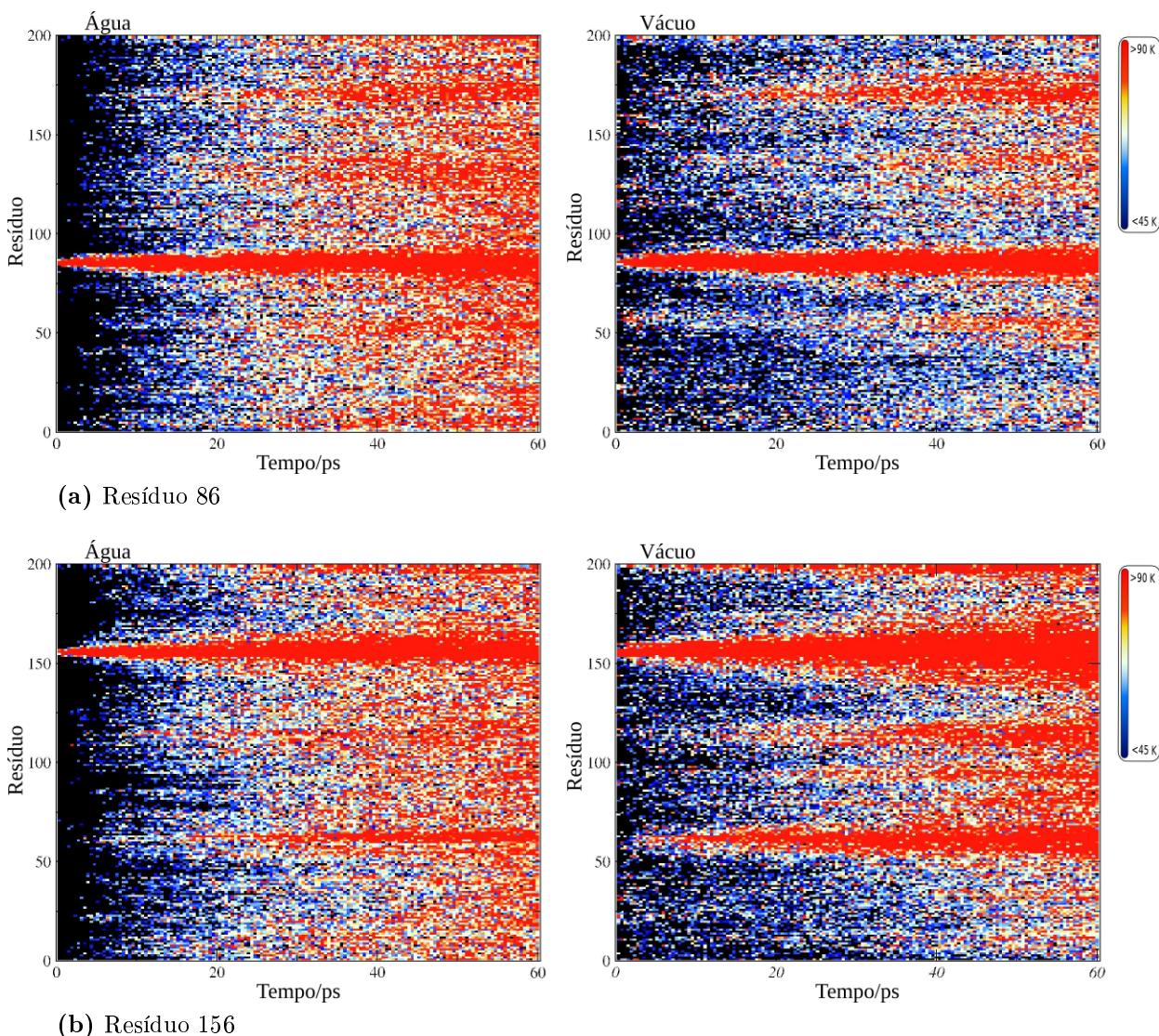
**Figura 26** – Mutações aquecidas e a resposta (temperatura) de toda proteína em função do tempo. 2VUJ (esquerda) e 2VUL (direita) com os aminoácidos que receberam o estímulo térmico no ATD.

### 3.5.2 Comparação água *versus* vácuo

Foram gerados poucos gráficos de ATD em água como função do tempo, já que é necessário realizar novas simulações de alto custo computacional devido ao aumento no tamanho do sistema. Desta maneira, foram comparados os resultados de água e vácuo de apenas uma simulação no tempo por proteína e não a média de várias simulações. Como visto anteriormente, o resultado em água diferencia-se do resultado em vácuo principalmente pelo início não exibir regiões com maior energia inicial. Ao contrário: é necessário um tempo para que a proteína comece a apresentar temperaturas maiores que 40 K. Isso corrobora os resultados obtidos na Seção 3.4 - Relaxamento.

Nestas simulações, a proteína permanece com 40 K ou menos até por volta de 10 ps e geralmente a temperatura do sistema em água é menor neste intervalo de tempo. No vácuo, as regiões quentes já começam a surgir no instante inicial. Adicionalmente, a temperatura média em água é maior, visto que após cerca de 40 ps praticamente não é possível distinguir entre as regiões quentes, como ocorre no vácuo. Este fato não é observado no MDT já que este é a média temporal apenas dos primeiros 30 ps. Provavelmente a energia introduzida no sistema leva um tempo para se dissipar. Assim, a água presente absorve energia e redistribui para boa parte da proteína, aumentando a flutuação de temperatura. No vácuo, a transmissão dessa energia é feita somente através da proteína, portanto as regiões mais próximas são aquecidas em primeiro lugar. Como a quantidade de energia introduzida no sistema não é sempre a mesma (método de Langevin), é possível que em água uma maior quantia de energia é

introduzida para manter a temperatura do banho, dado o maior número de átomos no sistema (dado não calculado). Então, a energia do banho térmico é primeiramente dissipada na água e nas regiões mais próximas. Após alguns picossegundos de simulação a água redistribui essa energia adicional para o restante da proteína. Isto é observado nas figuras 27a e 27b. Por isso que em água, por volta de 40 ps praticamente todos os aminoácidos já estão próximos à temperatura máxima da escala (90 K), diferente do MDT em vácuo, no qual para o mesmo tempo de simulação as proteínas são mais frias.



**Figura 27** – Gráfico de ATD em água (esquerda) e no vácuo (direita), em função do tempo para um *frame*. Neste caso a proteína 1F5J teve os aminoácidos a) 86 e b) 156 aquecidos.

### 3.5.3 Conclusões

O ATD em função do tempo corrobora os resultados obtidos até agora, como a energia inicial maior em algumas regiões e também a preferência de transferir energia para regiões vizinhas ao resíduo aquecido. Não foi possível encontrar diferenças entre os gráficos de ATD no tempo para as proteínas 2VUJ e 2VUL, além das regiões de maior energia inicial. Já para o ATD em água simulado no dobro do tempo do original observa-se que o sistema é aquecido mais rapidamente que o sistema em vácuo, pois a água drena mais energia do banho térmico. Por isso, aumentar o tempo de simulação do ATD em água não é a solução para aumentar a sensibilidade do MDT em solvente.

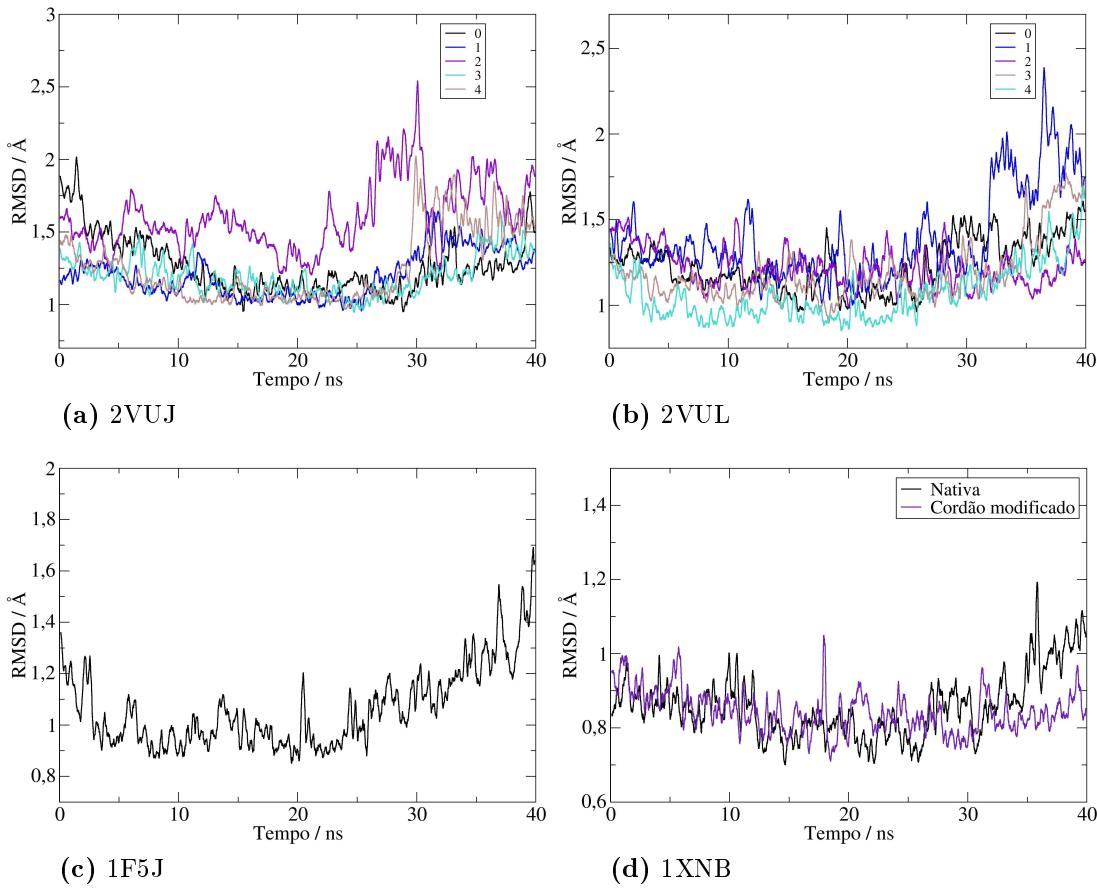
## 3.6 Análises da simulação de equilíbrio

Adicionando às análises comparativas, as quais visam justificar o aumento de termoestabilidade entre as proteínas 2VUJ e 2VUL, foram feitas análises com metodologias mais usuais, em simulações de equilíbrio. Foram calculadas energias eletrostática e de Van der Waals para átomos não ligados, moléculas de solvatação, deslocamento quadrático médio (do inglês *Root Mean Square Displacement*, RMSD) e ligações de hidrogênio, baseando-se principalmente nas diferenças da sequência de aminoácidos entre as duas proteínas. Vale relembrar que tais mutações são Ser9Pro, Thr13Phe, Asn14His, Tyr18Phe, Gln34Leu, Ser35Glu e Ser71Thr, e suas estruturas químicas estão representadas na Tabela 1.

### 3.6.1 RMSD

Podemos inspecionar a variação dos átomos em relação à estrutura média, como função do tempo, dessa forma obtendo um único valor de deslocamento quadrático médio para a proteína em cada instante de tempo de simulação. Ou ainda, pode-se calcular o valor de RMSD em função do resíduo ao longo de toda simulação.

Os cálculos de RMSD como função do tempo de simulação podem fornecer informações a respeito da qualidade de equilibração e mostrar mudanças conformacionais ocorrendo, já que o desvio em relação à média tende a se deslocar quando estas ocorrem. A Figura 28 mostra o RMSD como função do tempo para as simulações de equilíbrio das proteínas usadas neste estudo, indicando que estas não sofrem grandes mudanças conformacionais ao longo das simulações.

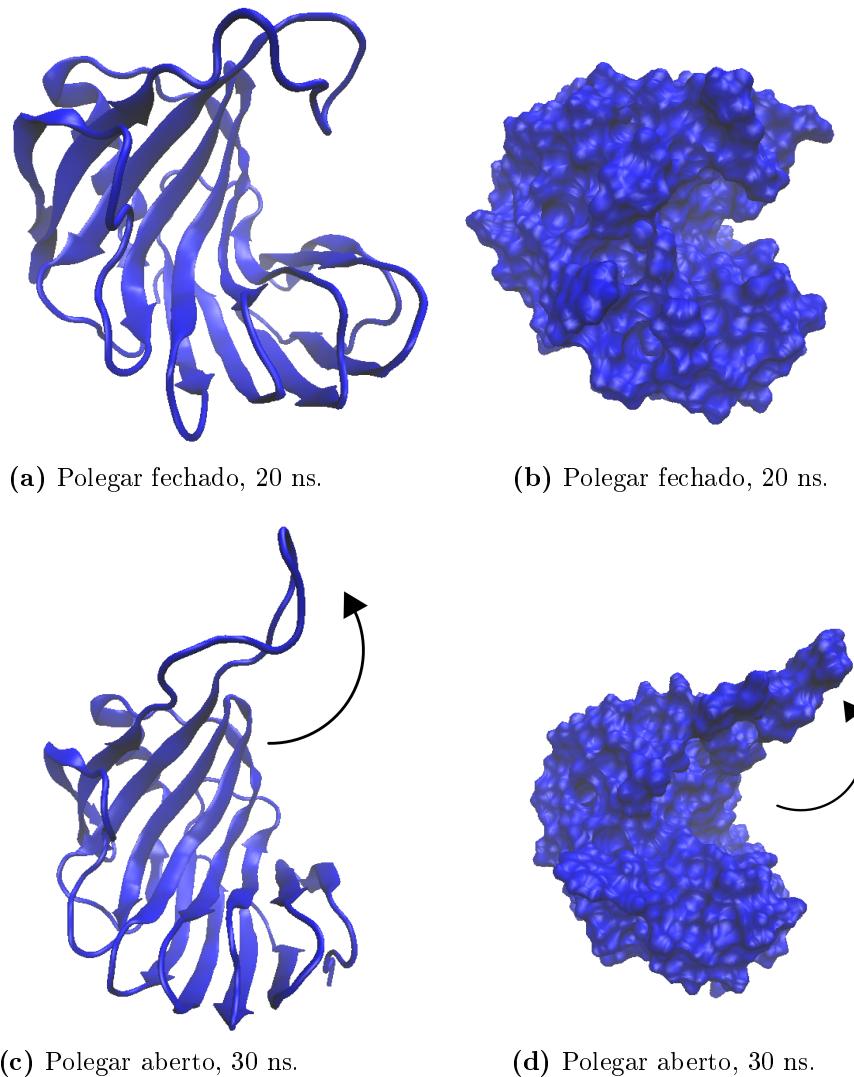


**Figura 28** – RMSD para as cinco simulações em função do tempo de simulação para a) 2VUJ e b) 2VUL (as 5 simulações), c) 1F5J e d) 1XNB (estrutura nativa e a com o cordão maior). O alinhamento foi feito em relação ao C- $\alpha$ , tendo a estrutura média no tempo como referência.

Os maiores valores de RMSD da curva 2 da Figura 28a e curva 1 de 28b, respectivamente das proteínas 2VUJ e 2VUL, sugerem que alguma mudança conformacional esteja ocorrendo. Quando as estruturas destas proteínas são analisadas, é notado que a região do polegar (Figura 1, na Parte ) realiza um movimento de abertura, expondo o sítio catalítico nos instantes em que o RMSD aumenta. De fato, esse movimento já foi observado e acredita-se ser o mecanismo responsável para a entrada do xilano no sítio catalítico e saída do produto (79).

O movimento do polegar foi observado em apenas uma entre cinco simulações de 40 ns, portanto é razoável supor que a frequência deste movimento seja menor se comparada à escala de tempo das simulações de ATD. Uma estimativa grosseira da frequência de abertura do polegar seria 1 evento a cada 200 ns. A região do cordão foi observada nos MDTs com alta energia cinética e gerou a hipótese de que esta região é bastante móvel. Já o movimento do polegar acontece em uma escala de tempo muito maior. Portanto, tentar correlacionar

esse movimento de abertura com os resultados de ATD é incoerente, uma vez que o ATD é observado no intervalo de alguns picossegundos e a abertura do polegar acontece tipicamente em uma escala três ordens de grandeza acima.

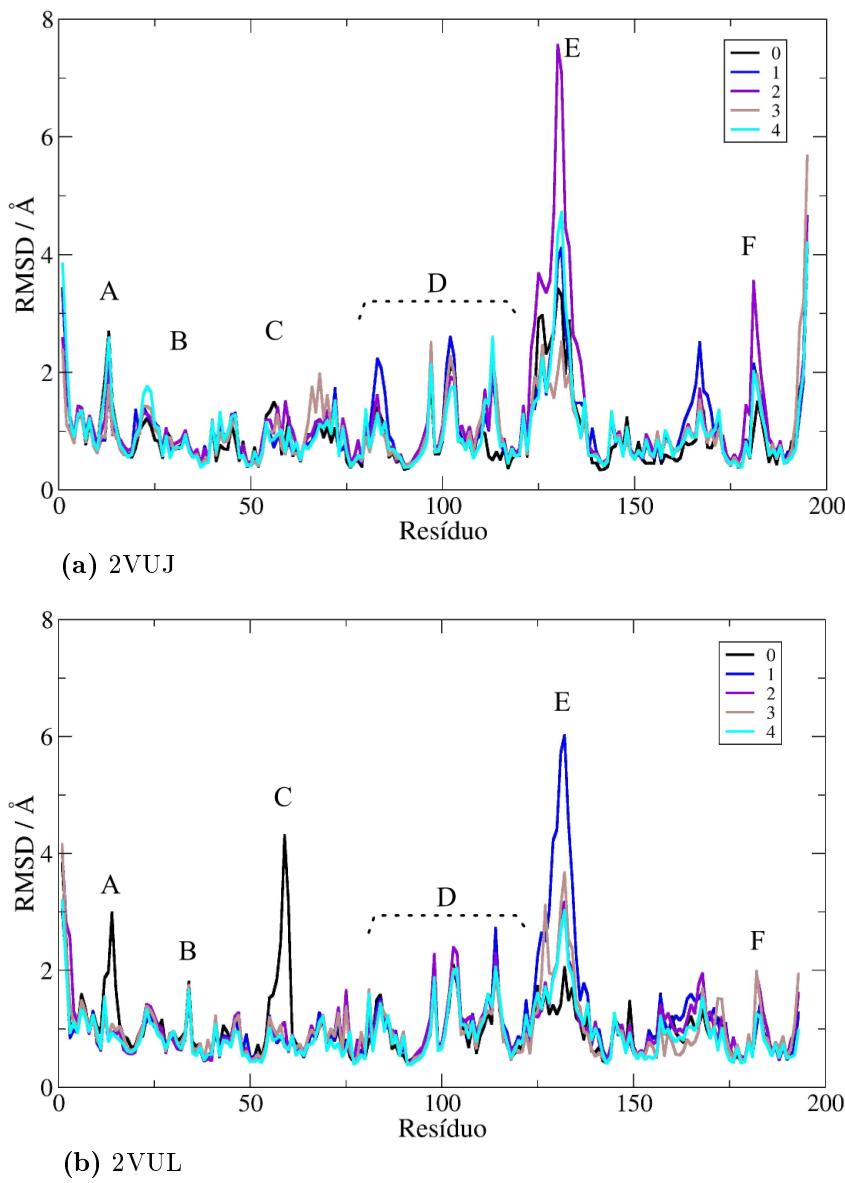


**Figura 29** – Estrutura de 2VUJ no tempo da simulação 2, representando o movimento do polegar, de acordo com a variação do RMSD. a) e b) são estruturas como polegar fechado, no tempo 20 ns, enquanto que c) e d) representam o polegar aberto em 30 ns. À esquerda: visualização de estruturas secundárias (modo *New Cartoon* no VMD). À direita: representação da superfície da estrutura (modo *Surf* no VMD).

A Figura 29 mostra as estruturas de 2VUJ da simulação em que foi observada a abertura do polegar. Aos 20 ns a estrutura ainda se encontra com o polegar fechado e, em 30 ns, o polegar apresenta-se na forma mais esticada, permanecendo nesta configuração até o final da

simulação (40 ns).

A análise de RMSD em função do resíduo (Figura 30) gerou uma compreensão mais detalhada em relação à mobilidade local nas estruturas 2VUJ e 2VUL, permitindo correlacionar esta mobilidade local com os resultados de ATD.



**Figura 30** – RMSD em função do aminoácido para as simulações de equilíbrio de 2VUJ e 2VUL.

O pico que mais se destaca representa o polegar (região E). Apresenta grande RMSD para ambas proteínas, confirmando a hipótese anterior da sua alta mobilidade. As outras regiões apresentam em sua maioria um RMSD bem menor que do polegar, porém alguns máximos locais podem ser destacados. Na região A, o RMSD atinge 3 Å entre resíduos 13 e 18 em

todas simulações de 2VUJ, porém em uma só de 2VUL (Figura 30a e Figura 30b). Esta é uma região de loop, portanto é razoável que apresente uma mobilidade maior que a média. Três mutações estão nesta região, e se tratando da média nas simulações, pode-se dizer que 2VUL apresenta um RMSD menor, com quase 2 Å de diferença. A região B está centralizada na mutação 35 e apesar de não ter um pico relativamente alto, apresenta quase 1 Å de diferença entre as proteínas. Este é um indicativo que essa região possui maior mobilidade na proteína hipertermoestável, corroborando os resultados de ATD, os quais mostram que no MDT em vácuo há somente linha quente nesta região para a proteína 2VUL (Subseção 3.1.1). A região C, entre os resíduos 53 e 63, apresenta um valor alto de RMSD apenas para 2VUL e somente em uma simulação. Esta, também é uma região de loop e sua linha quente no MDT apresenta-se ligeiramente mais intensa no mapa de 2VUL. A região D é composta por quatro picos semelhantes para ambas proteínas. Os resíduos que geram tais picos de RMSD formam regiões de loop e são observadas no MDT linhas quentes para os mesmos, com exceção do primeiro pico por volta do resíduo 85. O pico central é formado pelos resíduos que compõem o cordão, possuindo RMSD por volta de 2,5 Å nas duas proteínas. Finalmente, a região F tem seu máximo em um resíduo de glutamina cuja cadeia lateral é longa e carregada e está localizado na superfície da proteína. Porém, em uma simulação de 2VUJ essa região, e também o pico anterior, atigem valores 1 Å mais altos, aparentando uma maior mobilidade na proteína termofílica. A respeito das extremidades da cadeia proteica, o N-terminal apresenta altos valores de RMSD para ambas as proteínas e o C-terminal só possui alto RMSD em 2VUJ, o que é observado no MDT como uma linha quente.

### 3.6.2 Solvatação

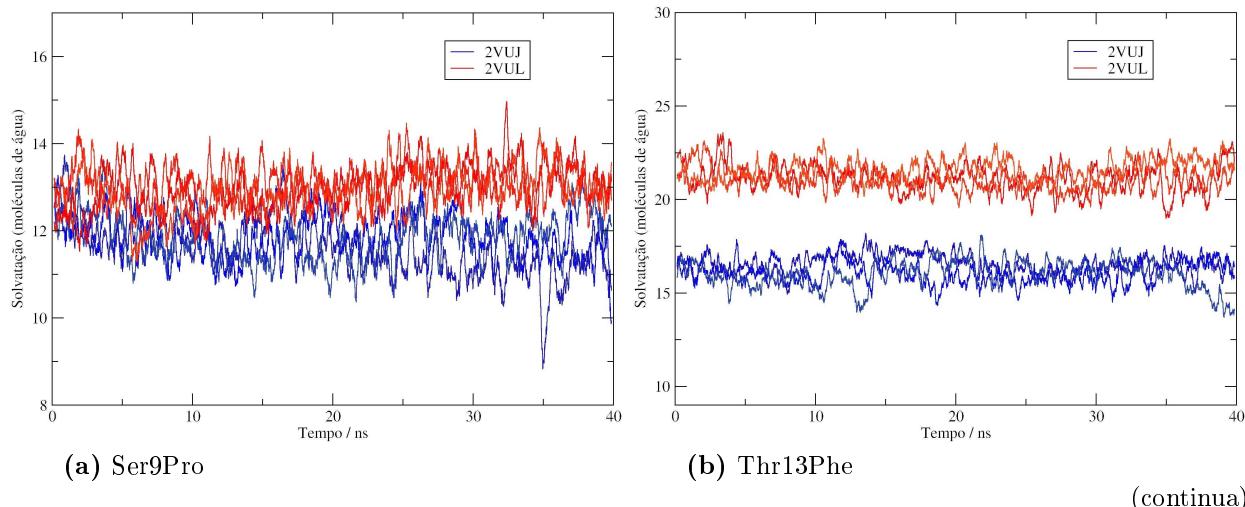
O número de moléculas de água contidas dentro do raio pré-estabelecido de 4 Å (primeira camada de solvatação) para cada resíduo é representado em função do tempo de simulação (Figura 30). Cores semelhantes representam a mesma proteína em diferentes simulações de 40 ns realizadas. Em geral, somando-se todas contribuições das mutações, a proteína mais termoestável 2VUL é mais solvatada do que 2VUJ, com uma diferença de aproximadamente 10 moléculas de água. Analisando individualmente:

Ser9Pro Em média, a prolina de 2VUL possui uma molécula de água de solvatação a mais. A prolina possui um carbono a mais que a serina, o que gera um aumento na superfície de solvatação, mesmo sendo um resíduo apolar. Estes resíduos se encontram nas superfícies das proteínas.

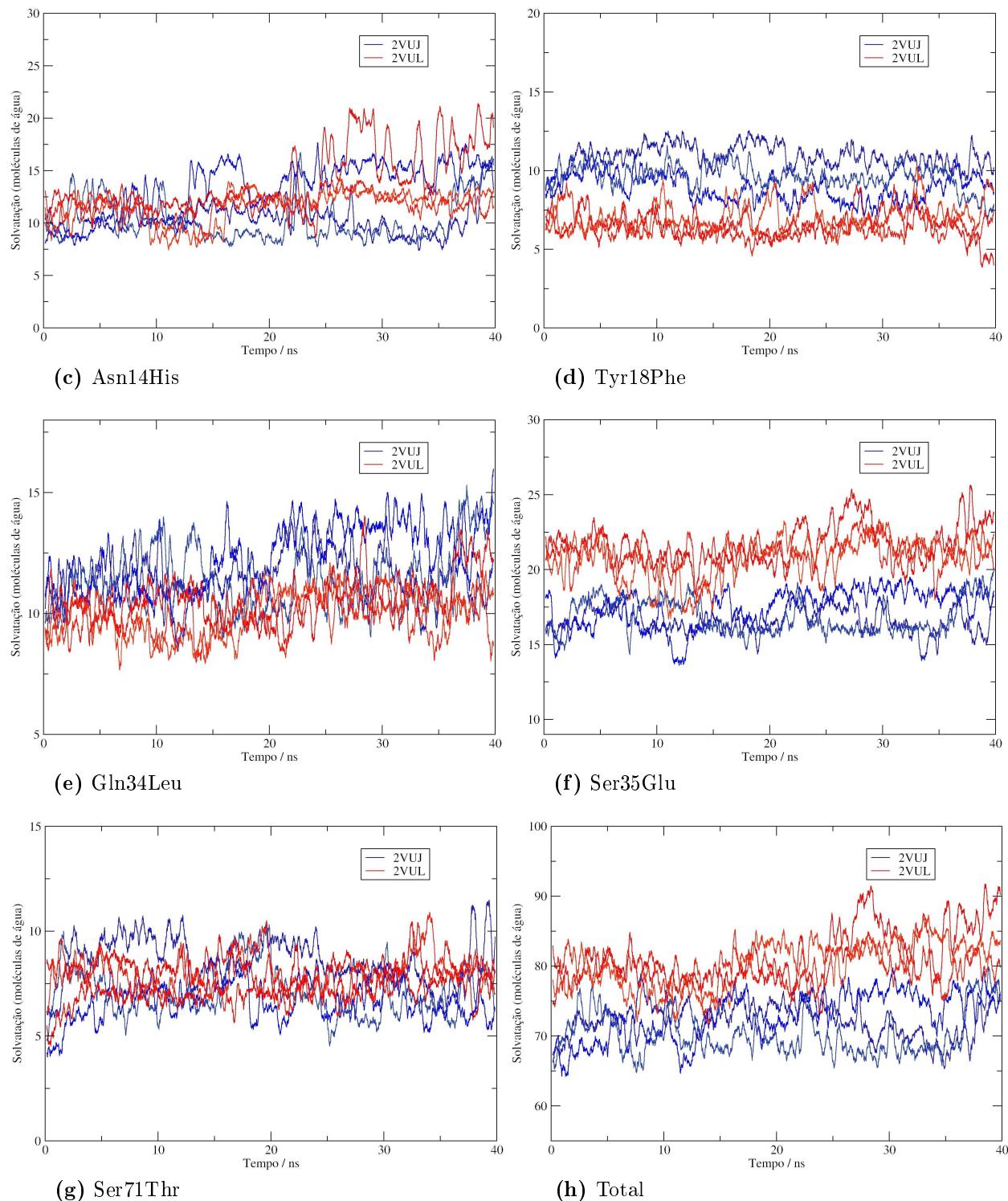
Thr13Phe Novamente uma mutação que envolve o aumento da hidrofobicidade. Uma diferença significativa de cerca de 5 moléculas de água de solvatação a mais para

2VUL. Mesmo sendo inicialmente um resíduo de caráter polar, a troca por um apolar, porém com maior quantidade de átomos, aumentou a superfície de solvatação na proteína mais termoestável. A localização destes aminoácidos é na superfície.

- Asn14His A mutação de um resíduo de asparagina para um histidina não levou a significativas diferenças entre a solvatação das duas proteínas segundo este experimento. A quantidade de águas de solvatação oscilou significativamente em função do tempo. Como essa variação é vista em mais de uma simulação, em ambas as proteínas, essa pode ser uma região sujeita a variações estruturais de escalas de tempo maiores.
- Tyr18Phe Apresentou uma diferença por volta de 3 moléculas, mas desta vez 2VUJ está mais solvatada.
- Gln34Leu Mesmo sendo um resíduo de superfície, a mutação perde aproximadamente 2 águas de hidratação, provavelmente por ser um resíduo hidrofóbico. Discutido mais detalhadamente a seguir, em *Ligações de Hidrogênio*.
- Ser35Glu Essa posição faz com que a cadeia lateral fique voltada ao solvente. Sendo o ácido glutâmico maior, este possui maior área para ser solvatado. Assim, cerca de cinco moléculas de água extras estão presentes na 2VUL. Isto é coerente com o aumento dos números de ligações de hidrogênio vistas na Subseção 3.6.4.
- Ser71Thr As curvas praticamente não se distinguem, oscilando em torno de sete águas de hidratação. Ambos resíduos são hidrofóbicos, porém, sua cadeia lateral é direcionada ao centro da proteína.



(continuação)



**Figura 30** – Quantidade de águas de solvatação em uma esfera de raio igual a 4 Å do aminoácido. Gráficos a)-g) individuais, e h) a soma entre todas as mutações.

### 3.6.3 Energia entre átomos não ligados

No intuito de avaliar a diferença no ambiente estereoquímico, energias eletrostática e de Van der Waals foram computadas para os resíduos mutados entre as proteínas 2VUJ e 2VUL. São analisadas as interações entre a cadeia lateral do resíduo e a proteína em um raio de 5 Å. Não foram analisadas interações com moléculas de água, ficando reservado às análises de solvatação e ligações de hidrogênio (Subseções 3.6.2 e 3.6.4). Os resultados são listados a seguir.

**Ser9Pro** 2VUL tem a energia de VDW levemente positiva (por volta de 2 kcal/mol), provavelmente pela conformação rígida do aminoácido prolina; enquanto que a eletrostática oscila em zero. A energia total\* de 2VUJ apresenta-se mais negativa, tendo como principal contribuição a energia eletrostática por volta de - 7 kcal/mol. Essa diferença na energia eletrostática é devido ao caráter dos átomos envolvidos; no caso da prolina, são somente carbonos, diferente da serina que possui o grupo hidroxila com carga de dipolo negativa. Portanto, neste caso o aminoácido de 2VUJ está mais estabilizado, se tratando em energias de átomos não ligados.

**Thr13Phe** Em ambas proteínas, a energia total oscila em torno de -4 kcal/mol, com as contribuições da duas energias variando em função do tempo. Ou seja, a contribuição de átomos não ligados para a estabilização neste resíduo é praticamente a mesma para as duas proteínas.

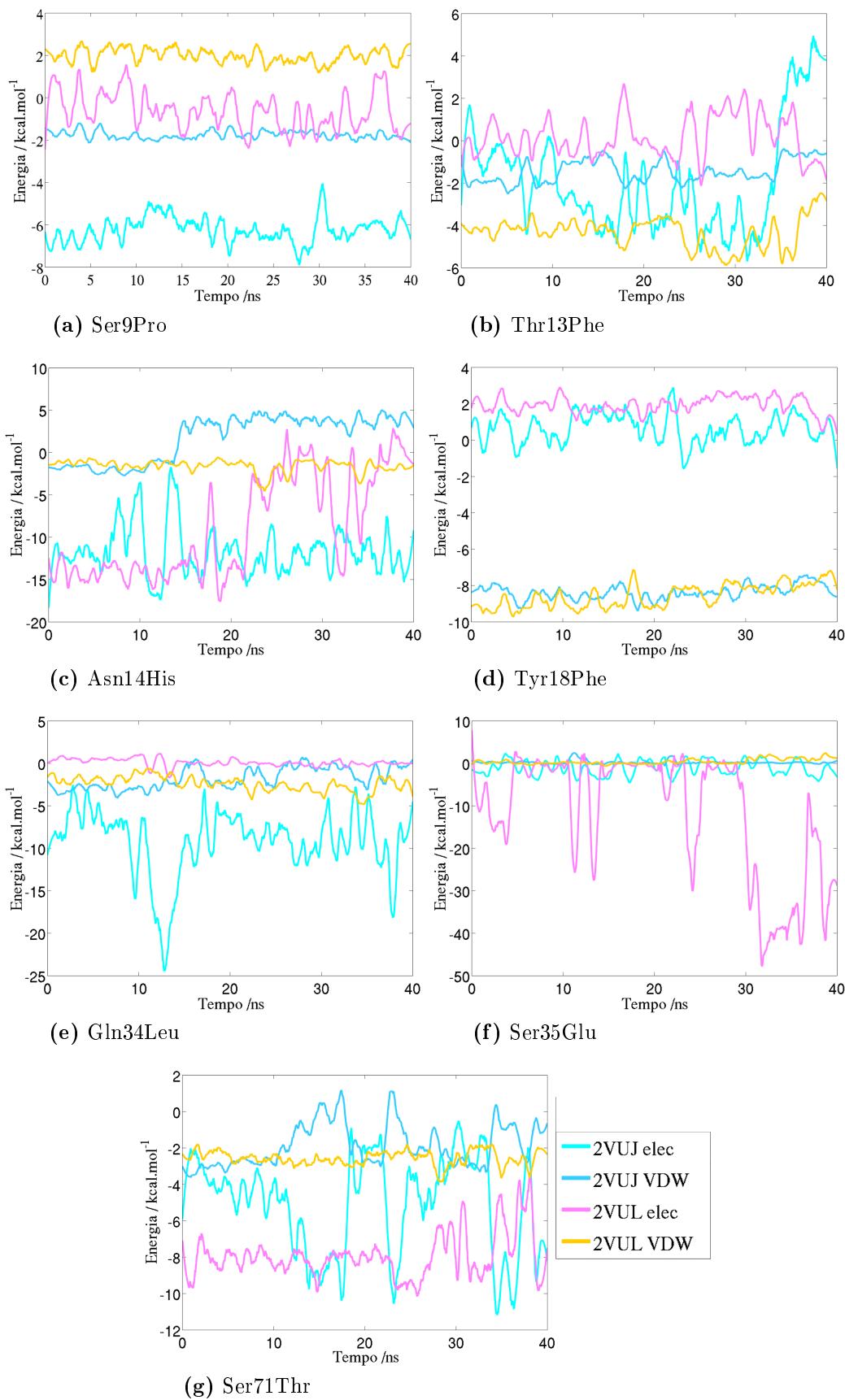
**Asn14His** A energia de VDW da asparagina oscila no início com aproximadamente o mesmo valor que a histidina e após de 14 ns desloca-se e oscila em torno de uma energia positiva em torno de 5 kcal/mol. A energia eletrostática, por sua vez, aumenta no final da simulação em 2VUL.

**Tyr18Phe** Praticamente não há diferença na energia de VDW das duas proteínas. Na eletrostática, a tirosina possui sempre menor ou igual energia que fenilalanina, cuja diferença chega a 3 kcal/mol. Isso provavelmente porque o grupo hidroxila na extremidade da tirosina pode interagir com a glutamina próxima.

**Gln34Leu** 2VUJ apresentou menor energia de interação, principalmente pela eletrostática. De fato, a troca de um aminoácido polar (glutamina) por um apolar (leucina) diminui as interações polares. Além de que, o ambiente ao redor deste resíduo é composto por aminoácidos polares e/ou carregados, como Arg, Gln e Thr.

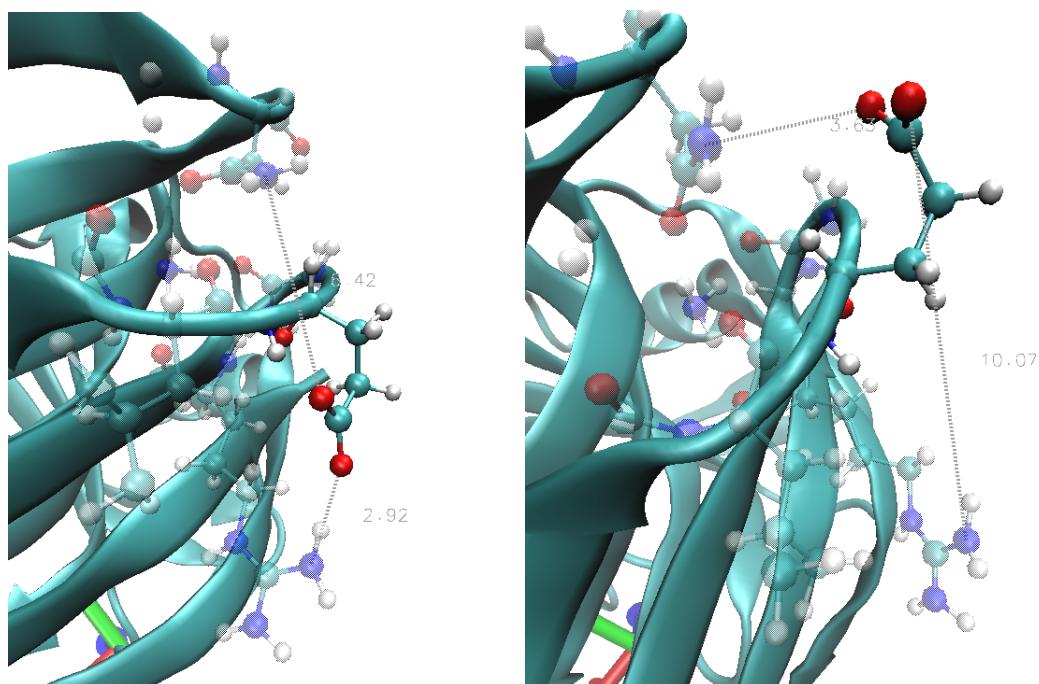
---

\*Energia total nesta Subseção representa a soma das energias eletrostática e VDW, ou seja, a energia de átomos não ligados.



**Figura 31** – Gráficos das energias eletrostática e de Van der Waals como função do tempo de simulação para as termoestáveis 2VUJ e 2VUL.

- Ser35Glu** A energia de VDW permanece a mesma para ambas, oscilando em torno de zero. Portanto, a energia total é composta praticamente pela eletrostática. Para 2VUL, é observada que a energia eletrostática oscila para valores negativos altos. O ácido glutâmico é um aminoácido carregado negativamente em contraste com a serina, um aminoácido de cadeia curta e com um grupo hidroxila. Assim, a razão da oscilação da energia eletrostática do ácido glutâmico reside na relativa ausência de impedimento estérico que este aminoácido de superfície possui, levando à grande movimentação de sua extremidade (composta por oxigênios carregados). De fato, durante a simulação são observadas interações deste resíduo com outros próximos. Para valores de energia de cerca de -20 kcal/mol, observa-se que a carboxila de Glu35 está próxima do nitrogênio de Asn15, e para valores próximos a -40 kcal/mol, esta interage com Arg38, formando ponte salina (Figura 32). Para valores próximos de zero, a cadeia lateral está entre esses dois estados, com a extremidade oxigenada apontando para fora, longe dos átomos de interação (entre 6 e 9 Å de distância).
- Ser71Thr** 2VUL possui menor energia total de interação, sendo a maior parte contribuição da energia eletrostática. Serina e treonina são aminoácidos semelhantes, possuindo um grupo hidroxila e um carbono a mais na treonina.



**Figura 32** – Energia variável de 2VUL é devido às variações das interações do ácido glutâmico 35 com asparagina (a) e arginina (b).

O aminoácido 35 é destacado tanto no MDT (como uma linha horizontal quente), quanto na CCL individual. Aqui, vimos que este apresenta fortes interações com 2 aminoácidos adjacentes, caracterizando um movimento oscilatório entre esses dois estados (Figura 32). Tendo em vista que o padrão de linhas quentes no MDT é devido a energia inicial, a alta mobilidade de Glu35 justifica sua temperatura nos padrões de MDT, e as interações eletrostáticas (inexistentes em 2VUJ) pode explicar a maior CCL deste resíduo de 2VUL, encontrada na Subseção 3.3.3.

### 3.6.4 Ligações de hidrogênio

Ligações de hidrogênio entre as moléculas de água e aminoácidos podem dar informações como acessibilidade do solvente ou estabilização da estrutura através do mesmo. Como visto anteriormente, a água pode atuar como um dissipador térmico e conduzir o calor para fora da cadeia proteica. Ou até mesmo, a água pode servir como importante fator estrutural, como por exemplo, na proposição de que a retirada das águas seria responsável pelo aumento de energia local em regiões supostamente móveis. É relevante, então, analisar estruturalmente as diferenças entre as ligações de hidrogênio entre a água e os resíduos mutados para ambas proteínas. A tabela 4 sumariza os valores médios e seus respectivos desvios para todas simulações realizadas para 2VUJ e 2VUL em função do resíduo observado.

Apesar do alto desvio relativo observado para cada simulação, a média é condizente entre as simulações. Normalmente na simulação são observadas entre 0 a 3 ligações de hidrogênio entre o resíduo e moléculas de água. O desvio padrão também apresenta valores bem próximos, indicando que tanto a média quanto o desvio convergem.

Em 2VUJ, os resíduos Tyr18 e Gln34 fazem em média uma ligação de hidrogênio com água. Porém em 2VUL (resíduos 18Phe e 34Leu) essa interação praticamente desaparece, já que as cadeias laterais de fenilalanina e leucina não possuem aceptores ou doadores para a ligação de hidrogênio, além do que a cadeia principal não possui acesso ao solvente. Em contraposição, na mutação Thr13Phe, a cadeia lateral não mostra influência no número de ligações de hidrogênio, uma vez que é o oxigênio da cadeia principal que realiza interações de hidrogênio, pois este é acessível à água. De forma similar, em Ser71Thr, apesar de ambos possuírem um grupo hidroxila, é o oxigênio da cadeia principal que acessa o solvente, e portanto esta mutação não influencia muito no números de ligações de hidrogênio. Na mutação Ser9Pro, a serina possui dois sítios para ligação de hidrogênio com a água, enquanto a prolina possui somente um, de forma que o valor observado desta última é praticamente metade da primeira.

A interação mais interessante é a do resíduo de Ser35Glu. Em 2VUJ, a serina faz em

média entre uma e duas interações de hidrogênio com a água. Quando esta é substituída para ácido glutâmico em 2VUL, a mesma região passa a formar por volta de 5 ligações. O ácido glutâmico, com sua cadeia lateral longa e carregada, e com dois aceptores de hidrogênio, fica livre se movimenta com alto grau de liberdade, como já visto em 3.6.3. Como este está direcionado para o solvente, realiza frequentes interações com a água. Este aminoácido foi visto no ATD como bom difusor de calor e mais flexível em sua estrutura, quando comparado com a proteína nativa. Aqui, observa-se também que este pode fazer várias interações de hidrogênio com a água.

**Tabela 4** – Quantidade de ligações de hidrogênio média no tempo para cada simulação de 2VUJ e 2VUL e o desvio padrão.

Mutação	2VUJ				2VUL			
	0	1	2	3	0	1	2	3
Ser9Pro	1,45±0,98	1,44±0,97	1,44±0,95	1,47±0,98	1,39±0,97	0,77±0,70	0,78±0,71	0,79±0,71
Thr13Phe	1,86±1,07	1,90±1,07	1,90±1,07	1,85±1,08	1,88±1,07	1,28±0,91	1,29±0,93	1,31±0,91
Asn14His	1,10±0,95	1,05±0,92	0,97±0,92	1,00±0,91	0,94±0,88	1,16±0,99	0,74±0,74	1,03±0,85
Tyr18Phe	0,89±0,73	1,02±0,82	0,90±0,72	0,91±0,77	0,86±0,73	0,04±0,20	0,03±0,17	0,03±0,17
Gln34Leu	1,24±1,00	0,95±0,86	1,00±0,91	0,96±0,84	1,18±0,96	0,12±0,38	0,07±0,26	0,07±0,27
Ser35Glu	1,67±1,05	1,65±1,04	1,74±1,06	1,70±1,06	1,66±1,07	5,16±1,62	5,19±1,66	4,9±1,66
Ser71Thr	1,19±0,90	1,03±0,86	1,01±0,86	1,18±0,91	1,12±0,86	0,82±0,76	0,87±0,77	0,95±0,83



# Capítulo 4

## Conclusões

O ATD foi utilizado na tentativa de desvendar mecanismos de termoestabilidade dentro da família 11 de xilanases. Concluiu-se que as diferenças mais evidentes nos mapas de difusão térmica foram resultantes das etapas anteriores às simulações de ATD. Ou seja, o que diferenciou as proteínas nos MDTs, foram regiões que apresentaram temperatura mais alta possivelmente pelo processo de dessolvatação e mudança de temperatura. O objetivo principal, que era observar possíveis diferenças na termoestabilidade através da difusão térmica anisotrópica, ficou em segundo plano quanto aos mapas, já que somente foi vista a difusão através de átomos próximos. Porém, o MDT ainda se apresenta como uma metodologia para diferenciar as proteínas, localizando e quantificando as regiões mais sensíveis à processos de relaxamento apresentados no método.

Possivelmente, as regiões que apresentaram maior energia inicial são regiões de superfície, com alta mobilidade e com uma grande contribuição da água em sua estruturação. De fato, análises de RMSD por resíduo confirmam que as regiões que apresentam-se quentes nos MDTs em vácuo são regiões mais móveis. Através do principal sistema avaliado (estruturas 2VUJ e 2VUL), o MDT mostrou que diferentes regiões de estruturas muito parecidas se comportam de maneira diferente quando submetidas pelos processos de ATD. Frente a essa sensibilidade no MDT, comparações entre proteínas termoestáveis com sua homóloga mesofílica apontaram a região do cordão como uma possível modificação para o aumento da termoestabilidade. Por si só, o fato observado em que constata a região cordão maior na proteína hipertermoestável (e na proteína nativa) corrobora recentes estudos que apontam para a relevância de regiões mais flexíveis em proteínas mais estáveis (12).

Os MDTs em água apresentaram-se semelhantes em todas as proteínas, já que efeitos de relaxamento não foram observados neste caso\*. Isso corrobora a hipótese de que regiões com maiores valores de energia inicial formaram-se pela retirada das moléculas de água. Interessantemente, regiões semelhantes de todas as proteínas apresentaram-se mais frias nos MDTs em água. No relaxamento, essas regiões são vistas como as que levam maior tempo para se aquecerem e, na estrutura proteíca, estão contidas principalmente no núcleo e em parte do sítio ativo. De fato, estas regiões se conservam entre proteínas de mesma família, e uma menor temperatura no mapa pode indicar um comportamento típico para manter tanto

---

\* analisados através dos gráficos de ATD no tempo e de relaxamento

o núcleo da estrutura quanto o sítio catalítico protegidos.

Já na contribuição da cadeia lateral, a energia inicial de regiões observada no MDT não deve apresentar influência, já que é contabilizada a subtração das respostas das estruturas, e portanto, contribuições iniciais se anulam. Portanto, este experimento representa de melhor forma a difusão anisotrópica. Entretanto, não ficou claro o papel da contribuição da CCL na termoestabilidade, uma vez que somente dois resíduos de sete mutações entre 2VUJ e 2VUL apresentaram distribuições não sobrepostas. Como foram necessárias gerar distribuições para melhor visualização dos dados, essa medida requereu uma maior amostragem, ou seja, uma maior demanda computacional.

O comportamento de CCL de resíduos de arginina confirmam resultados anteriores, no qual estes apresentam-se como bons condutores térmicos (1). Este comportamento pode estar relacionado com as características estereoquímicas do aminoácido, já que este possui uma longa cadeia lateral carregada. Adicionalmente, apesar da relação deste fato com a termoestabilidade não ter sido muito abordada, foi observado que a CCL de resíduos de Arg equivalentes em diferentes estruturas possuírem decrescentes CCL em função do aumento da termoestabilidade, fato que pode ser explorado no futuro devido à relevância destes resíduos neste fenômeno.

## 4.1 2VUJ e 2VUL

A região N-terminal das Xilanases da família 11 é bastante estudada e considerada como uma importante região para a termoestabilidade (45, 46, 63), sendo apontada como uma região de início de desnaturação em estudos usando simulações de DM (49). Esta foi a região mais destacada nas diferenças entre os MDTs, em especial na região das mutações Gln34Leu e Ser35Glu. Em particular, o resíduo Glu35 de 2VUL se destaca nas análises da simulação de equilíbrio: apresenta-se mais solvatado e realiza mais ligações de hidrogênio com a água em relação à Ser35 de 2VUJ. Ainda, através de análises de energia, foi observado que esse ácido glutâmico realiza um movimento oscilatório, ora interagindo com uma asparagina, ora com uma arginina e no meio tempo interage com a água.

Apesar da mutação Ser71Thr apresentar menor energia em 2VUL, o MDT não se diferencia para as duas proteínas nessa região. Experimentos em que somente essa mutação foi introduzida em 2VUJ mostram que esta estrutura apresenta a menor das  $T_m$  se comparada às estruturas com apenas uma das outras seis mutações, praticamente não diferindo sua  $T_m$  em relação à estrutura nativa (76,8 °C e 76,2 °C, respectivamente) (65). Isto poderia indicar uma mutação não tão importante para estabilidade térmica, mas possivelmente importante para a atividade.

Um resultado um pouco isolado do assunto, mas não irrelevante, foi o movimento de abertura do polegar encontrado em uma de cinco simulações de 40 ns para ambas as proteínas. De fato, esse movimento já foi observado e acredita-se ser o mecanismo responsável para a entrada do xilano no sítio catalítico e saída do produto (79).

## 4.2 Sugestões

A metodologia aqui utilizada já foi empregada na confirmação e vizualização de sinalização intramolecular proteica, na descoberta de que argininas e outros aminoácidos que são bons condutores de calor tem uma importância para a atividade das proteínas TR $\alpha$  e TR $\beta$  (1), e agora foi relacionado à termoestabilidade e estrutura em xilanases. Partindo de um conjunto inicial de estruturas para constituir a amostragem, este método demanda relativamente pouco tempo computacional. De um princípio simples, o método ainda possui um grande potencial para ser aplicado a outras situações. Em geral, mecanismos protéicos que estejam ligados a algum estímulo energético ou que sejam de caráter anisotrópicos são potenciais alvos. Por exemplo, proteínas alostéricas e suas modificações estruturais. Em particular a esse sistema, algumas sugestões são listadas abaixo.

- Para o ATD em água, outro solvente com menor capacidade térmica pode ser testado, com a idéia de que a difusão para o solvente não seja tão intensa, possibilitando observar a difusão para proteína juntamente a um fator dissipativo. Somente aumentar o tempo de simulação do ATD talvez não seja a solução, pois no ATD em função do tempo foi visto que, no dobro do tempo, a estrutura em água se aquece mais rápido como um todo, aumentando o ruído da medida.
- Ao invés de 2 fs, utilizar 1 fs como o tempo de integração, para adicionar ao resultado a vibração do átomo de hidrogênio, analisando assim essa contribuição na fluência do calor. Entretanto, isso também aumentaria o tempo computacional.
- Fazer o aquecimento do ligante e monitorar a resposta de toda proteína para assim ter uma idéia de difusão térmica relacionada ao sítio catalítico, ou mesmo a atividade da proteína termoestável. Essa idéia pode se estender a outros sistemas. Possivelmente, este experimento obtenha informações de aminoácidos importantes para função, já que tanto a difusão de calor em proteínas quanto a interação proteína-ligante estão relacionadas com as interações intermoleculares.
- Desenvolver melhor a idéia das regiões instáveis na retirada de água, pois através deste que foi possível notar diferenças entre as proteínas. De primeira análise, parece razoável regiões de proteínas termoestáveis apresentarem-se mais passíveis à este estímulo,

como discutido na Subseção 3.1.2 - Mesofílica *versus* termofílicas. Testes experimentais de estabilidade térmica da proteína 1XNB com a região do cordão modificada podem ser realizados no intuito de confirmar a hipótese. Por fim, o melhor desenvolvimento do método pode direcionar ao desenvolvimento racional de proteínas para termoestabilidade, usando como premissa a comparação entre estruturas similares de diferente estabilidades térmicas.





## REFERÊNCIAS

- 1 MARTINEZ, L.; FIGUEIRA, A. C. M.; WEBB, P.; POLIKARPOV, I.; SKAF, M. S. Mapping the intramolecular vibrational energy flow in proteins reveals functionally important residues. *The Journal of Physical Chemistry Letters*, v. 2, p. 1–16, Aug. 2011.
- 2 PONDER, J. W.; CASE, D. A. Force fields for protein simulations. *Advances in Protein Chemistry*, v. 66, p. 27–85, Jan. 2003.
- 3 BROOKS, B. R. et al. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, v. 30, n. 10, p. 1545–614, July 2009.
- 4 BERENDSEN, H.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, v. 91, n. 1-3, p. 43–56, Sept. 1995.
- 5 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, Washington, v. 118, n. 45, p. 11225–11236, Jan. 1996.
- 6 MADIGAN, M. T.; MARRS, B. L. Extremophiles. *Scientific American*, v. 276, n. 4, p. 82–7, Apr. 1997.
- 7 SCHMID, A.; HOLLMANN, F.; PARK, J. B.; BüHLER, B. The use of enzymes in the chemical industry in Europe. *Current Opinion in Biotechnology*, v. 13, n. 4, p. 359–366, Aug. 2002.
- 8 GOMES, E.; GUEZ, M.; MARTIN, N.; DA SILVA, R. Enzimas termoestáveis: fontes, produção e aplicação industrial. *Química Nova*, v. 30, n. 1, p. 136–145, 2007.
- 9 COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiology Reviews*, v. 29, n. 1, p. 3–23, Jan. 2005.
- 10 MIZUGUCHI, K.; SELE, M.; CUBELLIS, M. V. Environment specific substitution tables for thermophilic proteins. *BMC bioinformatics*, v. 8 Suppl 1, p. S15, Jan. 2007.

- 11 HERNANDEZ, G.; JENNEY, F. E.; ADAMS, M. W.; LEMASTER, D. M. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 97, n. 7, p. 3166–70, Mar. 2000.
- 12 SEN, S.; NILSSON, L. *Thermostable Proteins - Structural Stability and Design*. 1a edição. ed. CRC Press, 2012.
- 13 KARSHIKOFF, A.; LADENSTEIN, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends in Biochemical Sciences*, v. 26, n. 9, p. 550–6, Sept. 2001.
- 14 BLEICHER, L.; PRATES, E. T.; GOMES, T. C. F.; SILVEIRA, R. L.; NASCIMENTO, A. S.; ROJAS, A. L.; GOLUBEV, A.; MARTÍNEZ, L.; SKAF, M. S.; POLIKARPOV, I. Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermstable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. *The Journal of Physical Chemistry. B*, v. 115, n. 24, p. 7940–9, June 2011.
- 15 SZILÁGYI, A.; ZÁVODSZKY, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure (London, England : 1993)*, v. 8, n. 5, p. 493–504, May 2000.
- 16 GREAVES, R. B.; WARWICKER, J. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. *BMC Structural Biology*, v. 7, p. 18, Jan. 2007.
- 17 CHAKRAVARTY, S.; VARADARAJAN, R. Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Letters*, v. 470, n. 1, p. 65–9, Mar. 2000.
- 18 RUSSELL, R. J.; FERGUSON, J. M.; HOUGH, D. W.; DANSON, M. J.; TAYLOR, G. L. The crystal structure of citrate synthase from the hyperthermophilic archaeon *pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry*, v. 36, n. 33, p. 9983–94, Aug. 1997.
- 19 THOMPSON, M. J.; EISENBERG, D. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology*, v. 290, n. 2, p. 595–604, July 1999.

- 20 YU, X.; LEITNER, D. M. Anomalous diffusion of vibrational energy in proteins. *The Journal of Chemical Physics*, v. 119, n. 23, p. 12673, 2003.
- 21 UZER, T.; MILLER, W. Theories of intramolecular vibrational energy transfer. *Physics Reports*, v. 199, n. 2, p. 73–146, Jan. 1991.
- 22 LEITNER, D. M. Energy flow in proteins. *Annual Review of Physical Chemistry*, v. 59, p. 233–59, Jan. 2008.
- 23 FUENTES, E. J.; GILMORE, S. A.; MAULDIN, R. V.; LEE, A. L. Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *Journal of Molecular Biology*, v. 364, n. 3, p. 337–51, Dec. 2006.
- 24 BOTAN, V.; BACKUS, E. H. G.; PFISTER, R.; MORETTO, A.; CRISMA, M.; TONILO, C.; NGUYEN, P. H.; STOCK, G.; HAMM, P. Energy transport in peptide helices. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 104, n. 31, p. 12749–54, July 2007.
- 25 HEIDARY, D. K.; ROY, M.; DAUMY, G. O.; CONG, Y.; JENNINGS, P. A. Long-range coupling between separate docking sites in interleukin-1beta. *Journal of Molecular Biology*, v. 353, n. 5, p. 1187–98, Nov. 2005.
- 26 GNANASEKARAN, R.; AGBO, J. K.; LEITNER, D. M. Communication maps computed for homodimeric hemoglobin: computational study of water-mediated energy transport in proteins. *The Journal of Chemical Physics*, v. 135, n. 6, p. 065103, Aug. 2011.
- 27 LEITNER, D. M. Frequency-resolved communication maps for proteins and other nanoscale materials. *The Journal of Chemical Physics*, v. 130, n. 19, p. 195101, May 2009.
- 28 KUBO, M.; SHIOMITSU, E.; ODAI, K.; SUGIMOTO, T.; SUZUKI, H.; ITO, E. Picosecond dynamics of the glutamate receptor in response to agonist-induced vibrational excitation. *Proteins*, v. 54, n. 2, p. 231–6, Feb. 2004.
- 29 ISHIKURA, T.; YAMATO, T. Energy transfer pathways relevant for long-range intramolecular signaling of photosensory protein revealed by microscopic energy conductivity analysis. *Chemical Physics Letters*, Amsterdam, v. 432, n. 4-6, p. 533–537, Dec. 2006.

- 30 WANG, Q.; WONG, C. F.; RABITZ, H. Simulating energy flow in biomolecules: application to tuna cytochrome c. *Biophysical Journal*, v. 75, n. 1, p. 60–9, July 1998.
- 31 SAGNELLA, D. E.; STRAUB, J. E. Directed Energy "Funneling" Mechanism for Heme Cooling Following Ligand Photolysis or Direct Excitation in Solvated Carbonmonoxy Myoglobin. *The Journal of Physical Chemistry B*, v. 105, n. 29, p. 7057–7063, July 2001.
- 32 LOCKLESS, S. W. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science*, Washington, v. 286, n. 5438, p. 295–299, Oct. 1999.
- 33 OTA, N.; AGARD, D. A. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *Journal of Molecular Biology*, v. 351, n. 2, p. 345–354, Aug. 2005.
- 34 WONG, K. K.; TAN, L. U.; SADDLER, J. N. Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications. *Microbiological Reviews*, v. 52, n. 3, p. 305–17, Sept. 1988.
- 35 HENRISSAT, B.; CLAEYSSENS, M.; TOMME, P.; LEMESLE, L.; MORNON, J. P. Cellulase families revealed by hydrophobic cluster analysis. *Gene*, v. 81, n. 1, p. 83–95, Sept. 1989.
- 36 TÖRRÖNEN, A.; HARKKI, A.; ROUVINEN, J. Three-dimensional structure of endo-1, 4-beta-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *The EMBO Journal*, v. 13, n. 11, p. 2493–2501, 1994.
- 37 SAHA, B. C. Hemicellulose bioconversion. *Journal of Industrial Microbiology & Biotechnology*, v. 30, n. 5, p. 279–91, May 2003.
- 38 MIELENZ, J. R. Ethanol production from biomass: technology and commercialization status. *Current Opinion in Microbiology*, v. 4, n. 3, p. 324–9, June 2001.
- 39 BEG, Q. K.; KAPOOR, M.; MAHAJAN, L.; HOONDAL, G. S. Microbial xylanases and their industrial applications: a review. *Applied Microbiology and Biotechnology*, v. 56, n. 3-4, p. 326–338, Aug. 2001.
- 40 KULKARNI, N.; SHENDYE, A.; RAO, M. Molecular and biotechnological aspects of xylanases. *FEMS Microbiology Reviews*, v. 23, n. 4, p. 411–56, July 1999.

- 41 SAPAG, A.; WOUTERS, J.; LAMBERT, C.; DE IOANNES, P.; EYZAGUIRRE, J.; DEPIEREU, E. The endoxylanases from family 11: computer analysis of protein sequences reveals important structural and phylogenetic relationships. *Journal of Biotechnology*, v. 95, n. 2, p. 109–31, May 2002.
- 42 WAKARCHUK, W. W.; SUNG, W. L.; CAMPBELL, R. L.; CUNNINGHAM, A.; WATSON, D. C.; YAGUCHI, M. Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. *Protein Engineering, Design and Selection*, v. 7, n. 11, p. 1379–1386, 1994.
- 43 TURUNEN, O.; ETUAHO, K.; FENEL, F.; VEHMAANPERÄ, J.; WU, X.; ROUVINEN, J.; LEISOLA, M. A combination of weakly stabilizing mutations with a disulfide bridge in the alpha-helix region of *Trichoderma reesei* endo-1,4-beta-xylanase II increases the thermal stability through synergism. *Journal of Biotechnology*, v. 88, n. 1, p. 37–46, June 2001.
- 44 PAËS, G.; O'DONOHUE, M. J. Engineering increased thermostability in the thermostable GH-11 xylanase from *Thermobacillus xylanilyticus*. *Journal of Biotechnology*, v. 125, n. 3, p. 338–50, Sept. 2006.
- 45 XUE, H.; ZHOU, J.; YOU, C.; HUANG, Q.; LU, H. Amino acid substitutions in the N-terminus, cord and  $\alpha$ -helix domains improved the thermostability of a family 11 xylanase XynR8. *Journal of industrial microbiology & biotechnology*, v. 39, n. 9, p. 1279–88, Sept. 2012.
- 46 SUN, J.-Y.; LIU, M.-Q.; XU, Y.-L.; XU, Z.-R.; PAN, L.; GAO, H. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. *Protein Expression and Purification*, v. 42, n. 1, p. 122–30, July 2005.
- 47 GAO, S.-J.; WANG, J.-Q.; WU, M.-C.; ZHANG, H.-M.; YIN, X.; LI, J.-F. Engineering hyperthermostability into a mesophilic family 11 xylanase from *Aspergillus oryzae* by in silico design of N-terminus substitution. *Biotechnology and bioengineering*, Oct. 2012.
- 48 ZHANG, S.; ZHANG, K.; CHEN, X.; CHU, X.; SUN, F.; DONG, Z. Five mutations in N-terminus confer thermostability on mesophilic xylanase. *Biochemical and Biophysical Research Communications*, v. 395, n. 2, p. 200–6, Apr. 2010.
- 49 PURMONEN, M.; VALJAKKA, J.; TAKKINEN, K.; LAITINEN, T.; ROUVINEN, J. Molecular dynamics studies on the thermostability of family 11 xylanases. *Protein Engineering, Design & Selection : PEDS*, v. 20, n. 11, p. 551–9, Nov. 2007.

- 50 DELBONI, L. F.; MANDE, S. C.; RENTIER-DELRUE, F.; MAINFROID, V.; TURLEY, S.; VELLIEUX, F. M.; MARTIAL, J. A.; HOL, W. G. Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. *Protein Science : a Publication of the Protein Society*, v. 4, n. 12, p. 2594–604, Dec. 1995.
- 51 VOGT, G.; ARGOS, P. Protein thermal stability: hydrogen bonds or internal packing? *Folding & design*, v. 2, n. 4, p. S40–6, Jan. 1997.
- 52 MACEDO-RIBEIRO, S.; DARIMONT, B.; STERNER, R. Structural features correlated with the extreme thermostability of 1[4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. *Biological chemistry*, v. 378, n. 3-4, p. 331–6.
- 53 YIP, K. S.; STILLMAN, T. J.; BRITTON, K. L.; ARTYMIUK, P. J.; BAKER, P. J.; SEDELNIKOVA, S. E.; ENGEL, P. C.; PASQUO, A.; CHIARALUCE, R.; CONSALVI, V. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure (London, England : 1993)*, v. 3, n. 11, p. 1147–58, Nov. 1995.
- 54 KNEGTEL, R. M.; WIND, R. D.; ROZEBOOM, H. J.; KALK, K. H.; BUITELAAR, R. M.; DIJKHUIZEN, L.; DIJKSTRA, B. W. Crystal structure at 2.3 Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermonanaerobacterium thermosulfurigenes* EM1. *Journal of Molecular Biology*, v. 256, n. 3, p. 611–22, Mar. 1996.
- 55 SALMINEN, T.; TEPLYAKOV, A.; KANKARE, J.; COOPERMAN, B. S.; LAHTI, R.; GOLDMAN, A. An unusual route to thermostability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases. *Protein Science : a Publication of the Protein Society*, v. 5, n. 6, p. 1014–25, June 1996.
- 56 VOORHORST, W. G.; WARNER, A.; DE VOS, W. M.; SIEZEN, R. J. Homology modelling of two subtilisin-like proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri*. *Protein Engineering*, v. 10, n. 8, p. 905–14, Aug. 1997.
- 57 WALLON, G.; KRYGER, G.; LOVETT, S. T.; OSHIMA, T.; RINGE, D.; PETSKO, G. A. Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart

- from *Thermus thermophilus*. *Journal of Molecular Biology*, v. 266, n. 5, p. 1016–31, Mar. 1997.
- 58 HENNIG, M.; STERNER, R.; KIRSCHNER, K.; JANSONIUS, J. N. Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. *Biochemistry*, v. 36, n. 20, p. 6009–16, May 1997.
- 59 HARRIS, G. W.; PICKERSGILL, R. W.; CONNERTON, I.; DEBEIRE, P.; TOUZEL, J. P.; BRETON, C.; PéREZ, S. Structural basis of the properties of an industrially relevant thermophilic xylanase. *Proteins*, v. 29, n. 1, p. 77–86, Sept. 1997.
- 60 VILLERET, V.; CLANTIN, B.; TRICOT, C.; LEGRAIN, C.; ROOVERS, M.; STALON, V.; GLANSDORFF, N.; Van Beeumen, J. The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 95, n. 6, p. 2801–6, Mar. 1998.
- 61 OMPRABA, G.; VELMURUGAN, D.; ARUMUGAM, P.; GOVINDASAMY, V.; KALAICHELVAN, P. T. Homology Model of a Novel Thermostable Xylanase from *Bacillus subtilis*-AK1. *Journal of Biomolecular Structure and Dynamics*, v. 25, n. 3, p. 311–319, Dec. 2007.
- 62 YOU, C.; HUANG, Q.; XUE, H.; XU, Y.; LU, H. Potential hydrophobic interaction between two cysteines in interior hydrophobic region improves thermostability of a family 11 xylanase from *Neocallimastix patriciarum*. *Biotechnology and Bioengineering*, v. 105, n. 5, p. 861–70, Apr. 2010.
- 63 GEORIS, J.; de Lemos Esteves, F.; LAMOTTE-BRASSEUR, J.; BOUGNET, V.; DEVREESE, B.; GIANNOTTA, F.; GRANIER, B.; FRÈRE, J. M. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. *Protein Science : a Publication of the Protein Society*, v. 9, n. 3, p. 466–75, Mar. 2000.
- 64 JOO, J. C.; POHKREL, S.; PACK, S. P.; YOO, Y. J. Thermostabilization of *Bacillus circulans* xylanase via computational design of a flexible surface cavity. *Journal of Biotechnology*, v. 146, n. 1-2, p. 31–9, Mar. 2010.

- 65 DUMON, C. et al. Engineering hyperthermostability into a GH11 xylanase is mediated by subtle changes to protein structure. *The Journal of Biological Chemistry*, v. 283, n. 33, p. 22557–64, Aug. 2008.
- 66 SRINIVASAN, M.; RELE, M. Microbial xylanases for paper industry. *Current Science*, v. 77, n. 1, p. 137–142, 1999.
- 67 MORRIS, D. D.; GIBBS, M. D.; CHIN, C. W.; KOH, M. H.; WONG, K. K.; ALLISON, R. W.; NELSON, P. J.; BERGQUIST, P. L. Cloning of the xynB gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on kraft pulp. *Applied and Environmental Microbiology*, v. 64, n. 5, p. 1759–65, May 1998.
- 68 HAKULINEN, N.; TURUNEN, O.; JANIS, J.; LEISOLA, M.; ROUVINEN, J. Three-dimensional structures of thermophilic beta-1,4-xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa*. Comparison of twelve xylanases in relation to their thermal stability. *European Journal of Biochemistry*, v. 270, n. 7, p. 1399–1412, Apr. 2003.
- 69 SHATSKY, M.; NUSSINOV, R.; WOLFSON, H. J. A method for simultaneous alignment of multiple protein structures. *Proteins*, v. 56, n. 1, p. 143–56, July 2004.
- 70 MARTÍNEZ, L. LM-MDanalysis. Disponível em:  
<<http://lm-mdanalysis.googlecode.com>>. Acesso em: 22 jan. 2013.
- 71 MATLAB. *version 7.8 (r2009a)*. Natick, Massachusetts: The MathWorks Inc., 2010.
- 72 KUMAR, S.; NUSSINOV, R. How do thermophilic proteins deal with heat? *Cellular and Molecular Life Sciences*, v. 58, n. 9, p. 1216–1233, Aug. 2001.
- 73 JAENICKE, R. Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 97, n. 7, p. 2962–4, Mar. 2000.
- 74 WINTRODE, P. L.; ZHANG, D.; VAIDEHI, N.; ARNOLD, F. H.; GODDARD, W. A. Protein Dynamics in a Family of Laboratory Evolved Thermophilic Enzymes. *Journal of Molecular Biology*, v. 327, n. 3, p. 745–757, Mar. 2003.
- 75 BROOKS, C. L. Molecular simulations of peptide and protein unfolding: in quest of a molten globule. *Current Opinion in Structural Biology*, v. 3, n. 1, p. 92–98, Feb. 1993.

- 76 GRUBER, K.; KLINTSCHAR, G.; HAYN, M.; SCHLACHER, A.; STEINER, W.; KRATKY, C. Thermophilic xylanase from *Thermomyces lanuginosus*: high-resolution X-ray structure and modeling studies. *Biochemistry*, v. 37, n. 39, p. 13475–85, Sept. 1998.
- 77 MCCARTHY, A. A.; MORRIS, D. D.; BERGQUIST, P. L.; BAKER, E. N. Structure of XynB, a highly thermostable  $\beta$ -1,4-xylanase from *Dictyoglomus thermophilum* Rt46B.1, at 1.8 Å resolution. *Acta Crystallographica Section D Biological Crystallography*, v. 56, n. 11, p. 1367–1375, Nov. 2000.
- 78 MRABET, N. T.; Van den Broeck, A.; VAN DEN BRANDE, I.; STANSSENS, P.; LAROCHE, Y.; LAMBEIR, A. M.; MATTHIJSSENS, G.; JENKINS, J.; CHIADMI, M.; VAN TILBEURGH, H. Arginine residues as stabilizing elements in proteins. *Biochemistry*, v. 31, n. 8, p. 2239–53, Mar. 1992.
- 79 POLLET, A.; VANDERMARLIERE, E.; LAMMERTYN, J.; STRELKOV, S. V.; DELCOUR, J. A.; COURTIN, C. M. Crystallographic and activity-based evidence for thumb flexibility and its relevance in glycoside hydrolase family 11 xylanases. *Proteins*, v. 77, n. 2, p. 395–403, Nov. 2009.



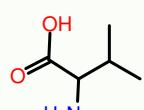
# Apêndice A - Tabela de aminoácidos

## Apolares

Alanina - ALA - A



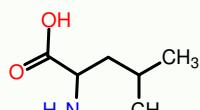
Valina - VAL - V



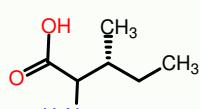
Metionina - MET - M



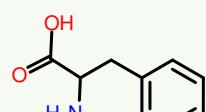
Leucina - LEU - L



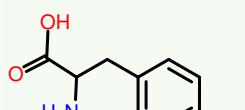
Isoleucina - ILE - I



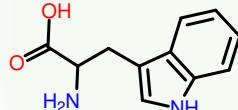
## Aromáticos



Fenilalanina - PHE - F



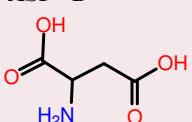
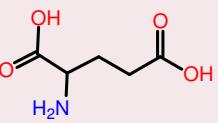
Tirosina - TYR - Y



Triptofano - TRP - W

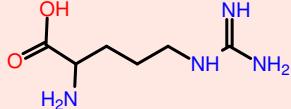
## Polares

### Negativos

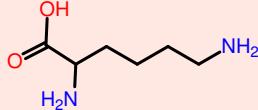
Ácido Aspártico  
ASP - DÁcido Glutâmico  
GLU - E

### Positivos

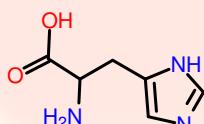
Arginina - ARG - R



Lisina - LYS - K



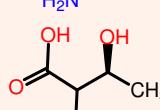
Histidina - HIS - H



Serina - SER - S



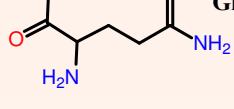
Treonina - THR - T



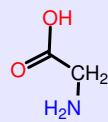
Asparagina - ASN - N



Glutamina - GLU - Q



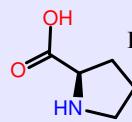
## Casos especiais



Glicina - GLY - G



Cisteína - CYS - C



Prolina - PRO - P