

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

DAVID LEANDRO PALOMINO SALCEDO

Simulações computacionais na proteína TM1030 da bactéria hipertermófila
Thermotoga maritima

São Carlos
2015

DAVID LEANDRO PALOMINO SALCEDO

Simulações computacionais na proteína TM1030 da bactéria hipertermófila
Thermotoga maritima

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular
Orientador: Prof. Dr. Eduardo Horjales Reboredo

Versão Original

São Carlos
2015

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

Palomino Salcedo, David Leandro
Simulações computacionais na proteína TM1030 da
bactéria hipertermófila *Thermotoga maritima* / David
Leandro Palomino Salcedo; orientador Eduardo
Horjales Reboredo -- São Carlos, 2015.
91 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Biomolecular) -- Instituto de Física de São
Carlos, Universidade de São Paulo, 2015.

1. Dinâmica molecular de proteínas. 2. Análises de
modos normais. 3. *Thermotoga maritima*. 4. TM1030. 5.
Fatores de temperatura cristalográficos. I. Horjales
Reboredo, Eduardo, orient. II. Título.

FOLHA DE APROVAÇÃO

David Leandro Palomino Salcedo

Dissertação apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências. Área de Concentração: Física Aplicada - Opção: Física Biomolecular.

Aprovado(a) em: 19/01/2016

Comissão Julgadora

Dr(a). Eduardo Horjales Reboredo

Instituição: (IFSC/USP)

Dr(a). Mauricio Garcia de Souza Costa

Instituição: (FIOCRUZ/Rio de Janeiro)

Dr(a). Guilherme Menegon Arantes

Instituição: (IQ/ USP)

A todas aquelas pessoas que sempre me apoiam
em novas aventuras. Obrigado.

AGRADECIMENTOS

Ao Prof. Dr. Eduardo Horjales Reboredo pelo apoio como orientador deste trabalho.

Ao Instituto de Física de São Carlos, pela oportunidade de realização do curso de mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela concessão da bolsa de mestrado.

A minha família por sempre me apoiar e acompanhar em todos meus propósitos.

Aos meus amigos e colegas pelo tempo que compartilhamos ainda digitalmente.

A todas as pessoas que ajudarem revisando a redação deste texto.

À banca avaliadora pelo tempo dedicado à leitura da dissertação.

E finalmente um agradecimento muito especial a María José que com seu carinho e ajuda conseguiu que este texto fosse terminado.

“Ciência é tudo aquilo que admite sempre discussão”
José Ortega y Gasset (1883-1955)

RESUMO

PALOMINO SALCEDO, D. L. Simulações computacionais na proteína TM1030 da bactéria hipertermófila *Thermotoga marítima*. 2015. 91 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

A *Thermotoga marítima* (Tm) é uma bactéria que vive em temperaturas na faixa dos 65 até 90°C, com temperatura ótima do redor dos 80°C. A proteína TM1030 de Tm, é um regulador transcricional da família TetR (Tetracycline repressor protein) reguladores da expressão gênica das proteínas TetA e TetB (Tetracycline resistance protein). Neste trabalho se rodaram 200ns de trajetória de dinâmica molecular a três temperaturas (293, 323 e 353K) da proteína TM1030 (PDB-1Z77) usando o pacote GROMACS com o potencial Amber99 e solvente explícito numa caixa cúbica com 90Å de comprimento, observando que RMSD da estrutura média da trajetória é menor em relação à estrutura cristalográfica, além disso que num primer momento esse RMSD tem uma mudança grande e que se estabiliza com uma maior velocidade nas maiores temperaturas. Também foi feita uma análise de modos normais na mesma estrutura usando o mesmo potencial, mas com solvente implícito, usando o modelo GBSA, minimizando a estrutura até ter um coeficiente de força média de $6,4 \times 10^{-8} \text{J} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ que assegura um bom mínimo local. Das trajetórias simuladas a partir das 6 menores frequências se achou uma relação com os movimentos observados nas dinâmicas moleculares e os esperados na transição alostérica entre as duas estruturas cristalográficas. Finalmente se calculam os fatores de temperatura das três trajetórias de dinâmica molecular, observando que seus esses fatores de temperatura aumentam com o aumento da temperatura, contrario do esperado da cristalografia onde diminuem com o aumento da temperatura do sistema.

Palavras-chave: Dinâmica molecular de proteínas. Análises de modos normais. *Thermotoga marítima*. TM1030. Fatores de temperatura cristalográficos. RMSD.

ABSTRACT

PALOMINO SALCEDO, D. L. Computational simulations at TM1030 protein of hyperthermophile *Thermotoga maritima* bacterium. 2015. 91 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2015.

The *Thermotoga maritima* (Tm) is a bacterium who can live at temperatures of 65 to 90°C, with optimum temperature around of 80°C. The TM1030 protein of Tm is a transcriptional regulator from TetR family (Tetracycline repressor protein) regulators of gene expression of the TetA and TetB protein (Tetracycline resistance protein). In this work 200ns of molecular dynamics trajectory was run at three temperatures (293, 323 and 353K) of TM1030 protein (PDB-1Z77) using GROMACS package with Amber99 potential and explicit solvent in a cubic box with length 90Å, noting that RMSD of the average structure of the trajectory is smaller with respect to the crystallographic structure, in addition, in a first time this RMSD have a large change and stabilizes at a higher speed at higher temperatures. There was also an analysis of normal modes on the same structure using the same potential, but with implicit solvent, using the GBSA model, minimizing the structure to have a medium force coefficient of $6,4 \times 10^{-8} \text{J} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ which ensures a good local minimum. Of the trajectories simulated from 6 lower frequencies was found a relationship with the movements observed in molecular dynamics and expected the allosteric transition between the two crystal structures. Finally was calculate the temperature factor of the three trajectories of molecular dynamics, observing their temperature factors increase with increasing temperature, contrary to expectations of crystallography which decrease with the increase of the system temperature.

Keywords: Protein molecular dynamics. Normal mode analysis. *Thermotoga maritima*. TM1030. Crystallographic temperature factors. RMSD.

LISTA DE FIGURAS

Figura 1 - Esquerda, TM1030 com o ligante alostérico(PDB-1ZKG). Direita, TM1030 sem o ligante alostérico (PDB-1Z77).....	29
Figura 2 - Domínios da proteína TM1030 com o ligante alostérico (PDB-1ZKG).....	29
Figura 3 - Diagrama de fita da estrutura cristalina do monômero da proteína TM1030.....	30
Figura 4 - Potencial do tipo Lennard-Jones.....	32
Figura 5 - Geometria de uma cadeia molecular simples, ilustrando a definição das distancias interatômicas r_{23}, ângulos entre ligações θ_{234}, e os ângulos de torção ϕ_{1234}.....	32
Figura 6 - Superfície de potencial real e aproximação harmônica.....	36
Figura 7 - RMSD em função do tempo em 200ns de trajetória das dinâmicas da estrutura 1Z77 nas temperaturas de 293, 323 e 353K, comparadas com a estrutura cristalográfica 1Z77.....	48
Figura 8 - RMSD em função do tempo em 200ns de trajetória das dinâmicas da estrutura 4I76 nas temperaturas de 293, 323 e 353K, comparadas com a estrutura cristalográfica 1Z77.....	49
Figura 9. Comparação dos fatores de temperatura anisotrópicos nas dinâmicas moleculares executadas a 293K (azul), 323K (verde) e 353K (vermelho)...	51
Figura 10 - Alinhamentos das estruturas médias das trajetórias a 293K (azul), 323K (verde) e 353K (vermelho) em relação à estrutura 1Z77 (amarelo)...	52
Figura 11 - Distancia entre as hélices H3 do domínio HTH em função do tempo em 200ns de trajetória das dinâmicas da estrutura 1Z77 nas temperaturas de 293, 323 e 353K.....	53
Figura 12 - Alinhamentos das estruturas médias da trajetória 353K na estrutura 1Z77, entre os 40 e 60 ns (azul) e entre os 180 e 200 ns (magenta).....	54
Figura 13 - Vistas do modo vibracional 7 desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	56
Figura 14 - Vistas do modo vibracional 8 desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	57

Figura 15 - Modo vibracional 9 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	58
Figura 16 - Modo vibracional 10 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	59
Figura 17 - Modo vibracional 11 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	60
Figura 18 - Modo vibracional 12 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.....	61
Figura 19 - Os modos normais restantes não tem deslocamentos maiores da cadeia principal.....	62
Figura 20 - Modo normal 9 com o ligante na posição de ligação alostérica.....	63
Figura 21 - No desenho a) se observa os deslocamentos do Modo normal 10, no b) a proteína TM1030 no estado ativo (amarelo) e no estado inativo (laranja).....	64
Figura 22 - Comparação desde diferentes posições da soma dos modos normais 7-12 (roxo) com a média da dinâmica a 353K (vermelho).....	66
Figura 23 - Comparação desde diferentes posições da soma dos modos normais 7-9, 11 e 12 (azul) com a média da dinâmica a 323K (verde).....	67
Figura 24 - Fatores de temperatura cada 1ns das dinâmicas da proteína TM1030 (PDB 1Z77) nas três temperaturas.....	68
Figura A1 - Gel SDS da expressão e lise a da ElrR, na coluna (a) meio 2XYT 3h solúvel, (b) 2XYT 3h insolúvel, (c) LB 3h insolúvel, (d) LB 3h solúvel, (e) LB 4h solúvel e (f) LB 4h insolúvel.....	81
Figura A2 - Absorbancia a 280nm na Gel filtración.....	82
Figura A3 - (a) Cristal do tipo octaédrico. (b) Cristal do Tipo Placa.....	83
Figura A4 - Ejemplo de las difracciones obtenidas para un cristal octaédrico con grupo espacial P41212.....	84

SUMÁRIO

1	INTRODUÇÃO.....	21
2	OBJETIVOS.....	23
3	BASES TEÓRICO/CONCEPTUAIS.....	25
3.1	CONCEPTOS BÁSICOS.....	25
3.1.1	Organismos extremófilos	25
3.1.2	<i>Thermotoga marítima</i>.....	26
3.1.3	Proteínas.....	26
3.1.4	Proteínas reguladoras de expressão, família TetR.....	28
3.1.5	TM1030.....	28
3.2	BASES TEÓRICAS.....	30
3.2.1	Dinâmica molecular.....	31
3.2.2	RMSD.....	33
3.2.3	Fatores de temperatura isotrópicos	34
3.2.4	Fatores de temperatura anisotrópicos	35
3.2.5	Modos normais.....	35
4	METODOLOGIA.....	39
4.1	Dinâmica molecular.....	39
4.1.1	Escolha da estrutura no PDB.....	39
4.1.2	Escolha do campo de força e preparação da estrutura.....	40
4.1.3	Criação da caixas de moléculas de água e adição dos íons.....	40
4.1.4	Minimização e equilíbrio do sistema.....	41
4.1.5	Dinâmica molecular em três temperaturas.....	42
4.1.6	Alinhamentos das trajetórias.....	42
4.1.7	Outras dinâmicas.....	43

4.2	RMSD.....	43
4.2.1	Cálculo do RMSD.....	43
4.3	Fatores de temperatura.....	44
4.3.1	Fatores de temperatura e estrutura média das trajetórias.....	44
4.3.2	Fatores de temperatura por faixas de tempo na trajetória.....	44
4.4	Modos normais.....	44
4.4.1	Minimização da estrutura.....	45
4.4.2	Criação e linearização da matriz hessiana.....	45
4.4.3	Trajетórias virtuais dos autovetores.....	45
5.	RESULTADOS E ANALISES.....	47
5.1	Dinâmica molecular.....	47
5.1.1	RMSD com respeito da estrutura cristalográfica.....	47
5.1.2	Fatores de temperatura anisotrópicos das trajetórias.....	49
5.1.3	Estrutura média das dinâmicas moleculares.....	52
5.2	Modos normais.....	54
5.2.1	Descrição dos modos normais representativos.....	54
5.2.2	Modos normais restantes.....	62
5.3	Comparação entre os modos normais, a dinâmica molecular e as estruturas cristalográficas.....	62
5.3.1	Modo normal 9 e o movimento do sítio alostérico.....	63
5.3.2	Modo normal 10 e o deslocamento dos domínios HTH.....	64
5.3.3	A soma dos modos normais se aproxima aos movimentos observados nas dinâmicas moleculares.....	65
5.4	Análises dos fatores de temperatura.....	67
5.4.1	Fatores de temperatura isotrópicos nas trajetórias.....	68
5.4.2	Porque não diminuem os fatores de temperatura?.....	69

6	CONCLUSÕES.....	71
7	PERSPECTIVAS.....	73
	REFERÊNCIAS.....	75
	ANEXOS.....	79
	ANEXO A - Relatório Cristalização da proteína ElrR.....	81
	ANEXO B - Inputs rodados no pacote GROMACS.....	87

1. INTRODUÇÃO

Estudar proteínas e enzimas termófilas é de grande interesse para a indústria, assim como outras moléculas extremófilas adaptadas já seja a valores limite de temperatura, pH ou pressão, devido a seu potencial para melhorar alguns processos industriais (1), por exemplo tratamentos para limpar a água, preparação de massa, processamento de amidos, diferentes tipos de catalises para bio-transformações, entre outros (2). Essas proteínas têm um especial potencial porque conservam sua flexibilidade, estabilidade e função biológica ainda em condições extremas em que outras proteínas e enzimas se desenovelariam. Ainda que a estabilidade da proteína não seja um fator determinante para sua aplicação comercial, o seu enovelamento em altas temperaturas requer de adaptação, e poder caracterizar o tipo de mudanças estruturais necessárias para se adaptar seria de grande importância para o desenho futuro de moléculas biotecnológicas que possam trabalhar em condições extremas. Nesta dissertação iremos estudar, através de métodos de simulação computacional, a proteína alostérica TM1030. A TM1030 é uma proteína reguladora de expressão pertencente à família das proteínas TetR (repressores de tetraciclina), da bactéria¹ *Thermotoga marítima*, uma bactéria gram-negativa, anaeróbica e hipertermófila. Existem alguns trabalhos sobre alosterismo em proteínas hipertermófilas como o de De la Vega-Ruíz et. al. de 2003 (3), mas não se tem ainda um consenso do que faz que as proteínas deste tipo sejam estáveis em altas temperaturas, uma das possibilidades é que estas estejam mais empacotadas, ou seja, que contenham uma maior quantidade de aminoácidos hidrofóbicos do que suas ortólogas mesófilas. A TM1030 foi escolhida por ter sido cristalizada e determinada a estrutura a três temperaturas diferentes 277, 310 e 323K. Também porque são conhecidas as estruturas de ambas as conformações alostéricas (T e R). Análises computacionais de processos bioquímicos tem-se convertido numa grande ferramenta para conhecer detalhes destes processos a nível molecular, detalhes que experimentalmente são muito difíceis de observar. Um método que vem tomado grande relevância desde a segunda metade do século passado é a dinâmica molecular de proteínas; com este método tem-se conseguido elucidar algumas das suas propriedades, por exemplo, estabilidade macromolecular, propriedades conformacionais e alostéricas (4), o papel da flexibilidade molecular na atividade enzimática, no reconhecimento molecular e nas propriedades de complexos, no enovelamento de proteínas, e

¹ Nesta dissertação a *Thermotoga marítima* será referenciada como uma bactéria, ainda que ao ter cruzamento de genes com arqueias compartilhe algumas proteínas com este domínio que gera dúvidas em sua classificação. Em todos os artigos usados como referencia para este trabalho, a *Thermotoga marítima* sempre foi classificada como bactéria.

na hidratação proteica. Outro método que também se tem tornado de grande interesse desde suas primeiras aplicações em 1982 (5) (6) é o de análises de modos normais de proteínas, técnica com a qual se podem encontrar os movimentos coletivos harmônicos dos aminoácidos da proteína e determinar a influência destes nas funções biológicas da molécula. Usando estas duas técnicas em nossa proteína de interesse, nós iremos analisar as mudanças de algumas propriedades como os fatores de temperatura e o RMSD (Root Mean Square Deviation), ao ir aumentando sua temperatura, no capítulo 3 das bases teórico/conceituais se detalham tanto as técnicas como as propriedades que se medirão. Experimentalmente tem-se achado que poderia existir um comportamento que é contrário ao predito pela termodinâmica, que ao aumentar a temperatura de cristalização na proteína TM1030 o fator de temperatura B diminui (1). Este fenômeno poderia ser constatado através de dinâmicas moleculares a diferentes temperaturas? Essa é uma das perguntas chave que procuramos responder nesta dissertação. Também se espera que a estabilidade e flexibilidade das proteínas de organismos extremófilos sejam maiores em suas condições ótimas de sobrevivência (7), por isso é interessante fazer análises do RMSD da dinâmica com respeito à estrutura cristalográfica e dos parâmetros de deslocamento atômico em diferentes faixas de tempo. Para o caso dos modos normais os movimentos coletivos nos indicariam propriedades da molécula como a variação da conformação do sítio alostérico e a relação dos movimentos entre domínios que lhe permitem se ligar ao DNA e realizar sua função biológica.

As análises feitas nesta dissertação poderiam nos aproximar cada vez mais à determinação das propriedades que fazem às proteínas e enzimas extremófilas diferentes de suas homólogas mesófilas. E assim, talvez nos ajudar no futuro, a desenhar moléculas que possam realizar trabalhos específicos em condições extremas.

2 OBJETIVOS

Nesta seção vamos expor os objetivos deste trabalho de pesquisa. Será dividida em objetivos gerais e objetivos específicos de cada uma das linhas de análises empregadas.

2.1 Objetivo geral

Observar o efeito do aumento da temperatura nas simulações de dinâmica molecular de uma proteína de um organismo hipertermófila. Usar estas trajetórias de dinâmica molecular para analisar a estabilidade e flexibilidade da proteína com o aumento da temperatura e comprovar se os fatores de temperatura diminuem com este aumento como foi achado experimentalmente. Além disso, realizar uma análise de modos normais e comparar com as trajetórias de dinâmica molecular.

2.1.1 Objetivos específicos

- Calcular trajetórias de dinâmica molecular na estrutura cristalográfica da proteína TM1030 em três temperaturas (293, 323 e 353K). Comparar o comportamento observado com a estrutura cristalográfica ativa e com o esperado experimentalmente.
- Comparar através do cálculo do RMSD as mudanças conformacionais das três trajetórias em relação à estrutura cristalográfica no estado alostérico considerado como inativo.
- Encontrar uma estrutura média de cada trajetória e analisar as diferenças e similaridades com a estrutura cristalográfica no estado alostérico considerado como inativo.
- Representar os deslocamentos atômicos da molécula nas três trajetórias como fatores de temperatura anisotrópicos e observar as diferenças nestes movimentos com a mudança da temperatura.
- Realizar o cálculo de modos normais da estrutura cristalográfica da proteína TM1030 para analisar os modos de baixa frequência e comparar tanto com o comportamento da dinâmica molecular, quanto com as estruturas cristalográficas nos dois estados conformacionais.
- Fazer os cálculos de criação e diagonalização da matriz hessiana da estrutura minimizada da TM1030 usando o pacote GROMACS para conseguir representar as trajetórias que representam os modos normais de menor frequência.

- Comparar os movimentos harmônicos encontrados dos modos normais com os encontrados nas dinâmicas moleculares e também com as mudanças conformacionais necessárias para passar entre as duas conformações alostéricas.
- Analisar o comportamento dos fatores de temperatura isotrópicos calculados a partir das três trajetórias de dinâmica molecular.
- Verificar se os fatores de temperatura diminuem com o aumento da temperatura, como encontrado na cristalografia. Não sendo assim, tentar formular uma explicação do porquê isso pode acontecer baseados nos diferentes resultados obtidos.

3 BASES TEÓRICO/CONCEITUAIS

Por ser interdisciplinar, este projeto demanda conhecer alguns aspectos de bioquímica, biologia, física e matemática. Por isso, nas seguintes páginas se faz uma descrição resumida dos conceitos básicos e teorias que vão ser desenvolvidas no decorrer do trabalho. Primeiro são apresentados os conceitos biológicos desde o tipo de organismos com que se vai trabalhar até a proteína de interesse, seguido de uma pequena introdução às teorias que servem de base para as simulações computacionais.

3.1 CONCEITOS BÁSICOS

Os conceitos básicos necessários para compreender o problema proposto nesta dissertação são principalmente biológicos. Sendo assim, se faz uma revisão sobre os organismos extremófilos e suas características particulares, seguida da descrição da bactéria extremófila *Thermotoga marítima*, a qual a proteína TM1030, usada neste trabalho, pertence. Uma pequena exposição das propriedades das proteínas é feita na continuação. Uma vez tendo estes conceitos, é abordada a família da proteína alvo, a TetR (proteínas reguladoras de expressão gênica). Por fim, faz-se uma breve discussão da proteína TM1030.

3.1.1 Organismos extremófilos

Tanto o crescimento como a sobrevivência de todos os organismos vivos estão regulados por fatores físicos e químicos, dos tipos bióticos e abióticos. (8) Para diferenciar os organismos que vivem em condições extremas (extremófilos) dos organismos que vivem em condições ‘normais’, se deve definir uma faixa de valores de temperatura, pH e salinidade, entre os quais esses últimos vivem. No caso do habitat bacteriano a temperatura pode variar de 4° até 40°C, o pH de 5 até 8.5 e a salinidade da água de fresca até salgada. (8)

Levando em conta que organismos considerados ‘normais’ podem sobreviver fora dessa faixa de valores, um organismo extremófilo se definiria como todo organismo onde seu crescimento e sobrevivência ótimos estão fora dessa faixa de condições ‘normais.’ (8) Nem todos os tipos de metabolismo ocorrem em todos os biótipos como, por exemplo, a fotossíntese, que não ocorre em altas temperaturas. Contudo, todas as reações primárias de absorção de energia usadas por organismos vivos são realizadas por reações de oxidação-redução.

Nessa dissertação, iremos trabalhar com um organismo extremófilo que vive em elevadas temperaturas, chamado de organismo hipertermófilo. Os organismos termófilos têm um crescimento ótimo entre 50 e 80°C, já os organismos hipertermófilos têm um crescimento ótimo em temperaturas entre os 80 e os 100°C. (9)

3.1.2 *Thermotoga marítima*

A *Thermotoga marítima* é uma bactéria gram-negativa, anaeróbica e hipertermófila, que vive numa temperatura entre 55 e 90°C, com temperatura ótima em 80°C, condições de pH entre 5.5 e 9.0, e salinidade na faixa de 0.23 a 3.75% de NaCl. Esta Bactéria foi isolada do solo geotérmico quente marinho do Vulcão de Ischia (Nápoles - Itália), no sistema vulcânico Reykjanes (Isafjörður - Islândia) e da Riviera Quente (nas ilhas Açores). (10) O DNA desta bactéria é constituído de aproximadamente 46% conteúdo de bases GC.

As células desta bactéria têm forma de bastão, podendo aparecer sozinhas ou em pares, rodeadas por uma estrutura revestida em forma de balão, localizada no final dos bastões. Ela é estritamente anaeróbica e está adaptada para metabolizar glucose, ribose, xilose, galactose, sacarose, maltose, amido e glicogênio. A bactéria não pode crescer a temperaturas inferiores a 55 °C. (10)

3.1.3 Proteínas

Todas as proteínas são polímeros formados por aminoácidos. (11) Com algumas raras exceções, como as de alguns aminoácidos de enzimas especializadas, todas as formas vivas da Terra têm proteínas compostas pelos mesmos 20 α -aminoácidos. (12) Cada uma das células de um organismo possui milhões de proteínas, que são cópias de milhares de proteínas diferentes, as quais ajudam a realizar uma diversidade de funções necessárias para sobreviver. Dentre essas funções estão a de transportar e armazenar moléculas pequenas, contribuir na organização estrutural das células e dos tecidos ou, em funções ainda mais específicas como coagular o sangue. (11) Cada proteína possui sua própria estrutura funcional, assim como determinadas características comuns com outras proteínas.

A conformação geral dos aminoácidos é simples, possuindo um grupo amina (-NH₂), um grupo carboxilo (-COOH), um átomo de hidrogênio e uma cadeia lateral variável (denotada como -R), onde todos esses grupos estão ligados ao mesmo átomo de carbono α (carbono assimétrico). (11) No pH fisiológico (~7,0), os grupos amina e carboxilo se encontram

ionizados nas formas $-\text{NH}_3^+$ e $-\text{COO}^-$, respectivamente, gerando assim uma espécie zwitteriônica. Existem duas exceções à definição anterior: no aminoácido glicina, o carbono α não é um carbono assimétrico, já que o grupo $-\text{R}$ é outro átomo de hidrogênio, e enquanto no aminoácido prolina o grupo amino não é primário, mas sim secundário ($-\text{NH}_2^+$, em pH 7,0) pois sua cadeia lateral está ligada ao seu N-terminal, formando um anel de pirrolidina.

As ligações ao redor do carbono α formam um tetraedro com dois possíveis estereoisômeros distinguíveis, que são imagens especulares não superponíveis uma a outra. Eles são conhecidos como os enantiômeros levogiro (L) e dextrogiro (D). (11) Todos os aminoácidos, com exceção da glicina, podem existir na forma L e D, pois o carbono α é quiral para todos os casos. Contudo, os aminoácidos na natureza, que fazem parte dos processos bioquímicos de transcrição e tradução que sintetizam as proteínas nos seres vivos, existem na forma L.

Os 20 α -aminoácidos codificados no DNA podem se dividir em dois grandes grupos em função de seu grau de atração pela água: hidrofílicos e hidrofóbicos. No primeiro grupo temos os aminoácidos de cadeia lateral polar neutra ou polar carregada (positiva ou negativamente) e no grupo dos hidrofóbicos, os aminoácidos de cadeia lateral neutra apolar. Existem muitas outras formas de classificar os aminoácidos, mas nenhuma é totalmente satisfatória, podendo ser classificados em função de diferentes características.

Os aminoácidos se unem por meio de ligações peptídicas, onde um grupo amino se liga ao grupo carboxilo do outro aminoácido perdendo uma molécula de água. Esta união de dois aminoácidos é denominada como dipeptídeo, (13) ao somar outro aminoácido nós teríamos um tripeptídeo e a partir de quatro são conhecidos como polipeptídeos. Um aminoácido dentro da cadeia polipeptídica é chamado de resíduo de aminoácido ou simplesmente resíduo. Normalmente, polipeptídeos formados por até 10 resíduos se denominam oligopeptídeos e cadeias polipeptídicas grandes (mais de 50 ou 100 resíduos de acordo com o autor) recebem o nome de proteínas.

As proteínas se caracterizam por ter uma sequência de aminoácidos que determina sua estrutura tridimensional e sua função. (11) Proteínas pequenas podem ter cadeias de algumas centenas de resíduos e as grandes de alguns milhares. Proteínas com funções similares de organismos diferentes podem ter sequências e estruturas similares, conhecidos como famílias de proteínas e são o resultado de mutações evolutivas. No tópico seguinte falaremos de uma família específica, a família TetR na qual a proteína central desta dissertação pertence.

3.1.4 Proteínas reguladoras de expressão, Família TetR

As proteínas que regulam a transcrição de sequências particulares do DNA têm propriedades específicas em seu domínio de ligação com o DNA, onde esses domínios geralmente são pequenos. (14) Um dos domínios mais estudados é o domínio HTH (helix-turn-helix), com um tamanho que varia de 60 a 90 resíduos de aminoácidos. São muitas as famílias de reguladores que possuem o domínio HTH com pequenas modificações, podendo ser formada por três hélices α , quatro hélices α (podendo estar a quarta hélice no começo ou no final), três hélices α e uma fita β , entre outras. (14)

O primeiro regulador transcricional pertencente à família dos TetR (repressor da tetraciclina) foi isolado no ano 1982. (15) As proteínas TetR são reguladores da expressão gênica das proteínas TetA e TetB (Tetracycline resistance protein). As TetR são homodímeros onde cada monômero tem um domínio N-terminal HTH que liga o DNA e um domínio C-terminal que faz parte da dimerização e liga o antibiótico tetraciclina. Os reguladores transcpcionais são importantes na adaptação dos microrganismos ao ambiente. (16) A família TetR tem uma similaridade alta suficiente para assegurar a conservação do enovelamento na sequência do domínio N-terminal, por volta dos 50 aminoácidos.

3.1.5 TM1030

A TM1030 é uma proteína da bactéria *Thermotoga marítima* que pertence à família TetR, tendo 200 resíduos de aminoácidos, um peso molecular de 25kDa e um ponto isoelétrico de 6,25. (16) Seu domínio N-terminal (aproximadamente 50 resíduos de aminoácidos) mostra uma alta similaridade com outros membros da família TetR, mas a região C-terminal (em torno de 150 resíduos de aminoácidos), não mostra similaridade com outras proteínas da família. (17) A proteína TM1030 é alostérica, onde no PDB (Protein Data Bank) é possível encontrar a estrutura nos dois estados conformacionais (1ZKG e 1Z77, Fig 1), e uma ligada ao seu DNA específico.

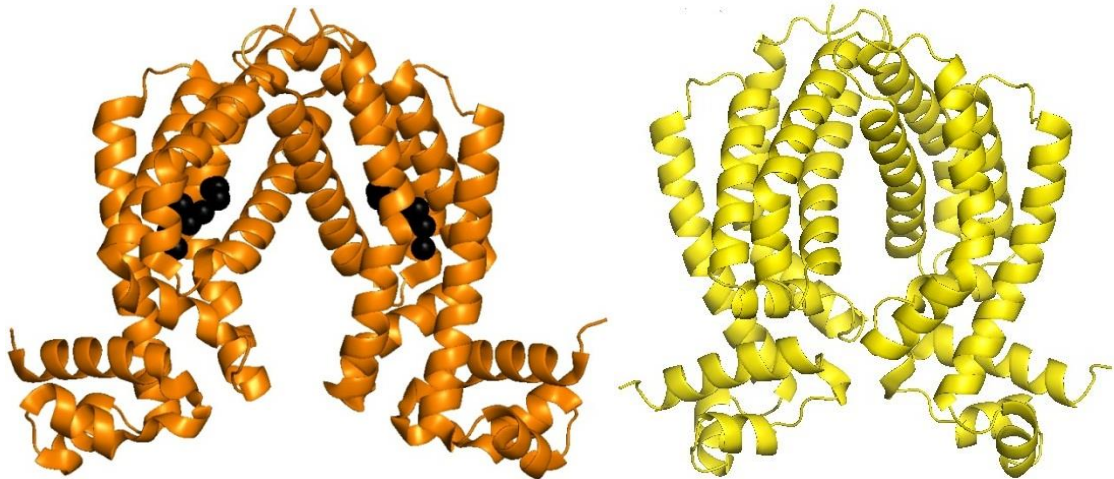


Figura 1 - Esquerda, TM1030 com o ligante alostérico (PDB-1ZKG).
Direita, TM1030 sem o ligante alostérico (PDB-1Z77).

Fonte: Elaborada pelo autor.

A proteína TM1030 está dividida em três regiões ou domínios, o domínio HTH, o domínio de dimerização e o domínio do sítio alostérico, Figura 2. É composta totalmente por hélices, onde 10 delas são α -hélices (H1-H7, H7A-H9) e uma hélice é 3_{10} (H6A). O domínio N-terminal possui as primeiras 3 α -hélices, sendo H2 e H3 as α -hélices do domínio HTH. O domínio regulatório (sítio alostérico) está entre as H4, H5 e as H7-H9 e o domínio de dimerização é composto pelas hélices H8 e H9 (16), Figura 3.

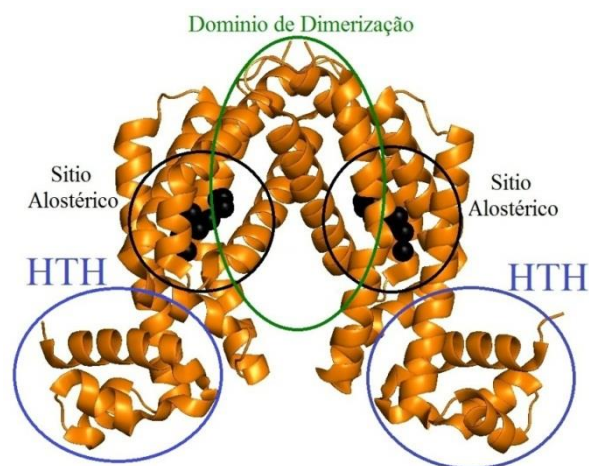


Figura 2 – Domínios da proteína TM1030 com o ligante alostérico (PDB-1ZKG).

Fonte: Elaborada pelo autor

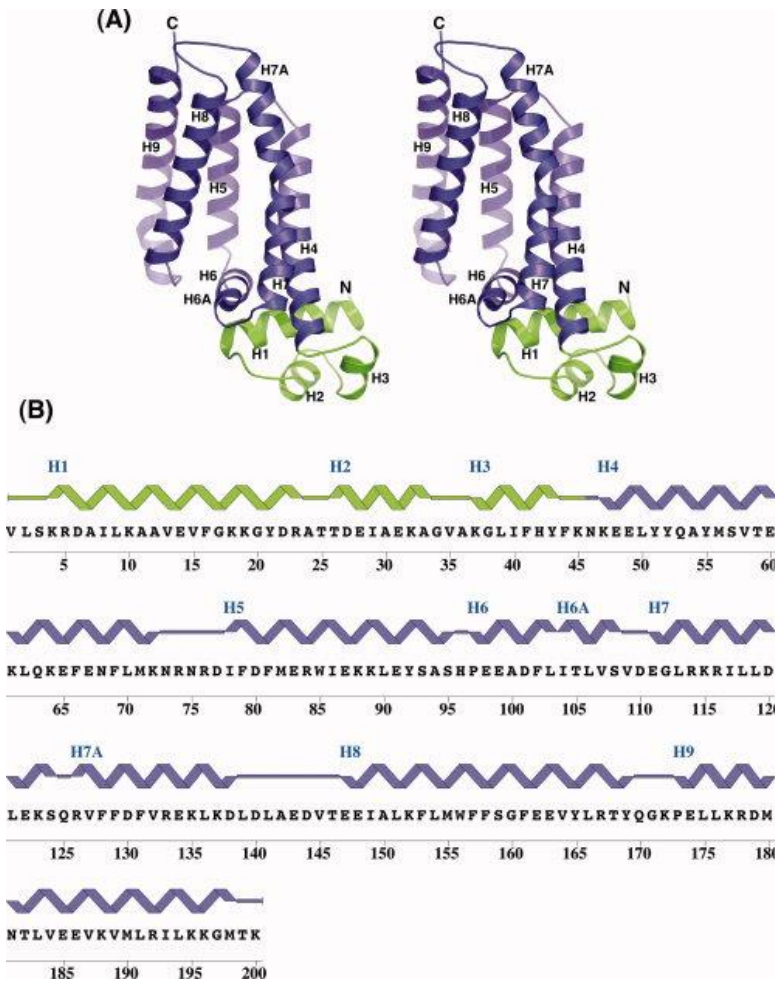


Figura 3 - Diagrama de fita da estrutura cristalina do monômero da proteína TM1030.

Fonte: PREMKUMAR (16)

3.2 BASES TEÓRICAS

Nesta seção iremos abordar as bases teóricas dos métodos de simulação e análises que vão ser usados nesta dissertação. Primeiro revisaremos a teoria por trás da simulação de dinâmica molecular. Depois serão introduzidos dois métodos usados na cristalografia, usados para analisar o comportamento da dinâmica molecular. O primeiro é o RMSD que serve para comparar duas estruturas, neste caso vai ser usado para comparar a estrutura a cada 250 passos de dinâmica molecular em relação à estrutura inicial e assim medir a deformação da estrutura com o tempo. O outro são os fatores de temperatura que servem para medir o deslocamento dos átomos dentro da estrutura.

A: Figura em estéreo que mostra o domínio de ligação de DNA e domínio regulatório estão coloridos em verde e roxo, respectivamente. As hélices assim como os N- e C-terminais estão rotulados.

B: Diagrama esquemático mostrando os elementos da estrutura secundária da TM1030 superpostos na sequência primária. As α -hélices e 3_{10} -hélice (H6A) estão indicadas [5].

3.2.1 Dinâmica Molecular

As simulações de dinâmica molecular são uma técnica computacional onde é possível estudar sistemas a níveis moleculares. As propriedades dinâmicas das proteínas são muito importantes para determinar sua função. Estas propriedades podem ser estudadas por meio da dinâmica molecular onde a proteína se representa no nível atômico. (18) As simulações de dinâmica molecular de proteínas têm escalas de tempo que vão tipicamente dos picossegundos aos microssegundos, mas atualmente existem simulações que chegam até os milissegundos, (19) Simulações de dinâmica molecular longas em proteínas servem para câmbios conformacionais que acontecem em escalas de tempo grandes, por exemplo, a mudança de configuração alostérica ou o enovelamento de uma proteína, e ajudam a entender melhor a relação estrutura função das proteínas. (19) Os tempos de cálculo neste caso são muito grandes já que é preciso fazer uma serie de contas independentes para cada átomo, sendo que uma proteína numa caixa de água pode ter na ordem de dois mil átomos.

Uma simulação de dinâmica molecular é simplesmente a solução, passo a passo, de uma equação de movimento clássica (20), representado pelas equações 1 e 2:

$$m_i \ddot{r}_i = f_i \quad (1)$$

$$f_i = -\frac{\partial}{\partial r_i} U \quad (2)$$

Onde, m_i são as massas, \ddot{r}_i são as acelerações e f_i as forças de cada átomo. U é a energia potencial da molécula. A partir disso, é preciso calcular as forças f_i que atuam nos átomos que derivam de uma energia potencial da forma $U(r^N)$, onde $r^N = (r_1, r_2 \dots r_N)$, representando um conjunto de $3N$ coordenadas atômicas (20). Este potencial é composto por dois tipos de forças de interação, as forças das interações não ligadas e das interações ligadas. Entre os potenciais de interação não ligados temos o potencial de Lennard-Jones, representado na equação 3:

$$v^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (3)$$

Onde os dois parâmetros σ e ϵ representam o diâmetro e a profundidade do poço de potencial respetivamente, como se mostra na seguinte figura:

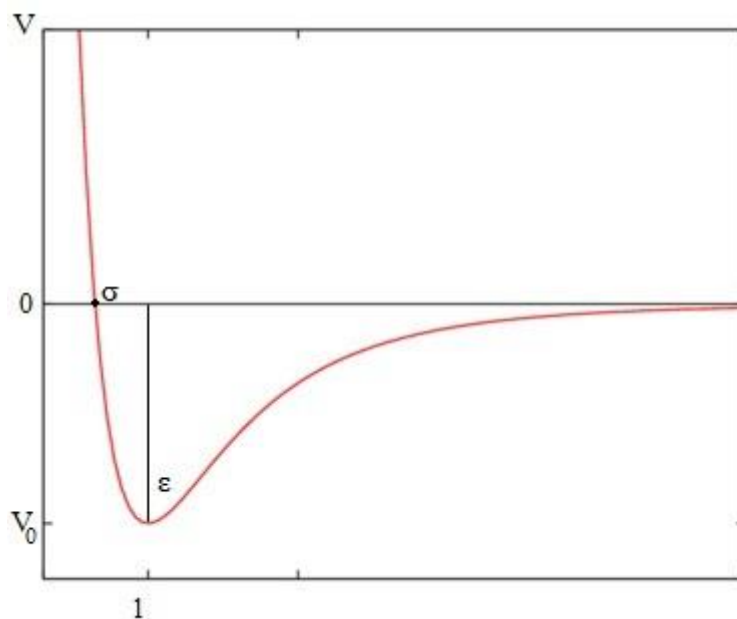


Figura 4 - Potencial do tipo Lennard-Jones

Fonte: Elaborada pelo autor.

O segundo potencial de interação não ligado é o das cargas eletrostáticas, conhecido como potencial de Coulomb, representado na equação 4:

$$v^{Coulomb}(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0 r} \quad (4)$$

Onde Q_1 e Q_2 são as cargas e ϵ_0 a permissividade no vácuo. No caso dos potenciais ligados, ou intramoleculares, temos três tipos: os que envolvem as ligações dos átomos, os que envolvem os ângulos entre essas ligações e os que envolvem ângulos de torção que definem planos entre três átomos ligados cada quatro coordenadas atômicas (20) (Fig. 3).

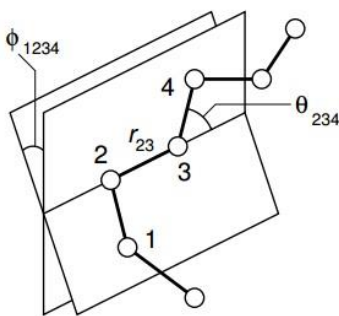


Figura 5 - Geometria de uma cadeia molecular simples, ilustrando a definição das distancias interatômicas r_{23} , ângulos entre ligações θ_{234} , e os ângulos de torção ϕ_{1234} .

Fonte: ALLEN (16)

As equações que representam estas ligações seriam representadas na equação 5:

$$\begin{aligned}
 u_{intramolecular} = & \frac{1}{2} \sum_{\text{ligações}} k_{ij}^r (r_{ij} - r_{eq})^2 \\
 & + \frac{1}{2} \sum_{\substack{\text{ângulos} \\ \text{deligação}}} k_{ijk}^\theta (\theta_{ijk} - \theta_{eq})^2 \quad (5) \\
 & + \frac{1}{2} \sum_{\substack{\text{ângulos} \\ \text{detorção}}} \sum_m k_{ijkl}^{\phi,m} (1 + \cos(m\phi_{ijkl} - \gamma_m))
 \end{aligned}$$

Os campos de força da mecânica molecular podem prever a estrutura e as propriedades de um modo muito preciso. Para o cálculo de dinâmica molecular de proteínas existem diferentes famílias de campos de força como AMBER, CHARMM ou OPLS. Os parâmetros para estes campos são determinados por uma combinação entre cálculos químico-quânticos com dados de termodinâmica, coexistência de fases e de conformação cristalográfica de di e tri peptídeos para as distâncias e ângulos (20).

Para realizar uma dinâmica é preciso que a proteína interaja com outros átomos que estariam ao redor dela num sistema real. Para isso se pode optar por duas possibilidades. A primeira é o solvente explícito, onde as moléculas de água têm seus próprios potenciais de interação com a proteína. A segunda possibilidade é o solvente implícito, no qual a vizinhança pode ser representada por um potencial contínuo que seria a soma de todas as forças externas.

3.2.2 RMSD

O procedimento mais comum para avaliar a superposição de estruturas é o root-mean-square-deviation (RMSD) (21), que mede a distância entre dois carbonos α equivalentes nas estruturas sobrepostas. A determinação do RMSD depende de dois processos críticos: o método de minimização e o procedimento para excluir regiões não equivalentes estruturalmente. Existem diferentes métodos para calcular esses dois parâmetros. A forma de calcular o RMSD é mostrada na equação 6 (22):

$$RMSD = \left(\frac{1}{n} \sum_n d_n^2 \right)^{\frac{1}{2}} \quad (6)$$

Onde n é o número de átomos e d é a distancia entre os dois átomos, sendo d_n representado pela equação 7:

$$d_n = \left[(a_{nx} - a'_{nx})^2 + (a_{ny} - a'_{ny})^2 + (a_{nz} - a'_{nz})^2 \right]^{\frac{1}{2}} \quad (7)$$

Onde $a_{nx} - a'_{nx}$, $a_{ny} - a'_{ny}$ e $a_{nz} - a'_{nz}$ são as diferencias de coordenadas de cada átomo na estrutura com relação ao átomo equivalente na outra estrutura. Nesta dissertação foram usados os programas `g_rms`, dentro do pacote GROMACS, e o Coot, do CCP4, que utilizam seus próprios algoritmos para minimizar e excluir regiões não equivalentes, mas os valores de RMSD que se obtém são muito similares.

3.2.3 Fatores de temperatura isotrópicos

O parâmetro de deslocamento atômico ou fator de temperatura (B) é uma medida da amplitude do deslocamento quadrado médio do átomo com relação à posição de equilíbrio num cristal. (23) Os átomos na rede cristalina ou dentro de uma molécula não estão fixos numa posição absolutamente rígida, podendo vibrar ao redor de uma posição media. (24) Seu deslocamento na rede pode estar relacionado à desordem, tendo uma posição ligeiramente diferente da esperada. Maiores temperaturas implicariam maior energia vibracional e, deste modo, um maior deslocamento da posição de equilíbrio. Na cristalografia isso significa uma diferença de fase adicional, causando um incremento na interferência de ondas parciais que emana dos elétrons, Isso gera uma função gaussiana adicional dependente do comprimento de onda, chamada de fator de Debye-Waller, fator B, representado na equação 8:

$$B = 8\pi^2 \langle u^2 \rangle \quad (8)$$

Onde B é o parâmetro de deslocamento isotópico o simplesmente fator B que está diretamente relacionado com o deslocamento médio quadrático isotrópico $\langle u^2 \rangle$ do átomo com respeito à posição de equilíbrio ou posição media.

3.2.4 Fatores de temperatura anisotrópicos

Os movimentos reais dos átomos da proteína não descrevem esferas perfeitas, mas sim elipsoides conhecidas também como fatores de temperatura anisotrópicos ou fator B anisotrópico (B_w). (23) Neste caso, o fator de temperatura tem a mesma forma do fator de temperatura isotrópico, sendo um tensor na forma da equação 9:

$$B_w = 8\pi^2 \langle U_w \rangle \quad (9)$$

Onde U_w representa a matriz de deslocamentos. O fator B_w , serve para gerar o fator de atenuação de escala anisotrópico (k_h) equação 10:

$$k_h = k_0 \exp\left(\frac{-1}{4} h^T O^{-1} B_w (O^{-1})^T h\right) \quad (10)$$

Onde h é o vetor do espaço recíproco expressados em termos de U . Onde k_0 é o fator de atenuação de escala anisotrópico inicial e O representa uma matriz de rotação.

3.2.5 Modos Normais

O primeiro trabalho envolvendo cálculos de modos normais a uma proteína foi publicado por Tasumi et.al. em 1982, usando um campo de força sem termos não ligados. (5,25) A proteína usada, o glucagon, é uma α -hélice com 29 resíduos. Nesse mesmo ano, Noguti e Gô analisaram as flutuações coletivas do inibidor de tripsina (58 resíduos), usando 241 ângulos de torção. (6) Em 1983 Levitt et. al., publicaram o primeiro estudo de modos normais a nível atômico (25) e finalmente Brooks e Karpus fizeram a primeira análise de modos normais usando coordenadas cartesianas (1740 no total), onde os dois trabalhos utilizaram a mesma proteína pequena: o inibidor de tripsina. (26)

A análise de modos normais é utilizada para estudar os movimentos vibracionais usando uma aproximação harmônica, onde se assume uma energia potencial igual à soma de termos quadráticos. (27) Com os coeficientes dos termos de força constante se cria uma matriz onde seus coeficientes formam a matriz de equações de modos vibracionais. Depois de achar os autovalores desta matriz, que seriam $3N$ (onde N é número de átomos), restando os 6 valores, devidos aos deslocamentos e rotações de toda a molécula, se obtém as $3N-6$

frequências fundamentais de vibração associadas aos $3N-6$ vectores que representam as direções e amplitudes relativas dos deslocamentos atômicos para cada modo. Já que as proteínas de interesse podem ter muitos átomos, são necessários métodos especiais para calcular esses valores. Modos de baixas frequências em proteínas são de particular interesse porque estão diretamente relacionados com propriedades funcionais, como se tem observado em proteínas nativas onde os modos normais de menor frequência tem uma maior contribuição nas flutuações conformacionais no equilíbrio térmico. (27)

O ponto de partida para realizar uma análise de modos normais é estabelecer uma conformação do sistema que represente um mínimo na superfície de energia potencial. Uma vez ali, se faz uma aproximação harmônica do potencial ao redor desta conformação. (28) Este passo é a aproximação central do método.

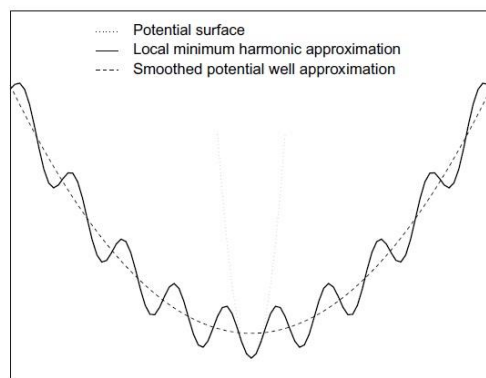


Figura 6 - Superfície de potencial real e aproximação harmônica.

Fonte: CUI; BAHAR (29)

Partindo desde o mínimo de energia potencial V . O potencial V pode ser expandido em séries de Taylor em termos da massa ponderada em coordenadas (30), equação 11:

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_0 q_i q_j \quad (11)$$

Onde q é a coordenada massa-peso que interage com potencial V , com $q = \sqrt{m_i} \Delta x_i$, onde Δx_i é o deslocamento da i -ésima coordenada em torno do mínimo de energia e m_i é a massa correspondente a cada átomo i . Como o novo potencial é quadrático, a superfície de energia se aproxima a uma parábola. Como esta é a base do cálculo, se espera que para biomoléculas em temperaturas fisiológicas os modos normais oscilem em superfícies de energia parabólica,

Isso é uma aproximação boa para deslocamentos pequenos mas isso nem sempre é verdade. A discussão seguinte se refere a como estender a aproximação às oscilações a temperatura ambiente. É conhecido que essas superfícies têm muitos mínimos com diferentes barreiras de energia. Esta segunda derivada do potencial é conhecida como matriz Hessiana. Para determinar os autovalores e autovetores desta é preciso diagonalizar a matriz como na equação 12 (30):

$$\mathbf{F}\mathbf{w}_j = \omega_j^2 \mathbf{w}_j \quad (12)$$

Onde \mathbf{w}_j é o j -ésimo autovetor e ω_j^2 é o j -ésimo autovalor. Têm-se $3N$ equações de autovetor. Cada autovetor é específico para uma coordenada de modo normal através da equação 13 (28):

$$Q_j = \sum_{i=1}^{3N} w_{ij} q_i \quad (13)$$

A soma é relacionada aos elementos de \mathbf{w}_j , com $|\mathbf{w}_j| = 1$. Note-se que as coordenadas de modos normais oscilam harmonicamente e independentemente, cada uma com frequência angular ω_j , conforme mostra a equação 14 (29):

$$Q_j = A_j \cos(\omega_j t + \varepsilon_j) \quad (14)$$

Aqui, A_j é a amplitude e ε_j é a fase. Estas coordenadas de modos normais são variáveis coletivas porque elas são combinações lineares das coordenadas cartesianas de cada átomo, como mostrado na equação 12 (30). No final cada modo normal ativado é representado na equação 15:

$$\Delta x_{ij} = \frac{W_{ij}}{\sqrt{m_i}} A_j \cos(\omega_j t + \varepsilon_j) \quad (15)$$

Onde Δx_{ij} , é o deslocamento da i -ésima coordenada em relação ao mínimo de energia e m_i , é a massa correspondente a cada átomo. Cada modo normal especifica um deslocamento atômico coletivo.

4 METODOLOGIA

No presente capítulo, iremos explicar os procedimentos utilizados para realizar as simulações de dinâmica molecular, as medições do RMSD, as medições dos fatores de temperatura e as análises de modos normais da proteína TM1030 da bactéria *Thermotoga marítima*. Todos os métodos foram baseados nos conhecimentos adquiridos durante o mestrado e nos tutoriais do pacote GROMACS. Os diferentes cálculos foram executados no supercomputador Blue Gene/P (convênio USP – Rice University). O análise foi feito usando um computador com um processador i7-2600 a 3,4 Ghz, 16 Gb de RAM e uma GPU Nvidia GT430.

4.1 Dinâmica molecular

Para realizar uma simulação de dinâmica molecular de uma proteína é necessário definir um protocolo a seguir em função do que se quer analisar, compatível com todo o estudo e que abrange desde a escolha da estrutura e o programa a ser utilizado até os parâmetros do cálculo. Neste caso, o protocolo de escolha engloba dinâmicas moleculares em três temperaturas diferentes na estrutura da proteína TM1030.

4.1.1 Escolha da estrutura no PDB

Escolher uma estrutura da proteína de interesse, depositada no Protein Data Bank (PDB²), é o primeiro passo a seguir para realizar a simulação de dinâmica molecular. Das nove estruturas disponíveis da proteína TM1030 no PDB (três com ligante: 1ZKG, ligante desconhecido; 2IEK, ligante hexaetilenoglicol e 4I76, ligante octanol. Duas sem ligante: 1Z77 e 2ID6. Três sem ligantes e cristalizadas em diferentes temperaturas: 3IH2 a 323K, 3IH3 a 310K e 3IH4 a 277K. E uma ligada a um oligonucleotídeo de 24bp de DNA: 4I6Z), optou-se pela 1Z77 por ser a única estrutura sem ligantes³ e com a sequência de aminoácidos completa.

² No PDB existem 111.385 estruturas de proteínas (onde 5.053 são mistura de proteínas e DNA, data: 19 de agosto de 2015), apesar que existam 52.494.032 sequências de proteínas depositadas no Reference Sequence Database (RSD, data: 13 de julho de 2015) do National Center for Biotechnology (NCBI). Considerando que muitas dessas estruturas são da mesma proteína, o número de estruturas únicas é ainda menor.

³ O baixo valor do RMSD entre todas as estruturas sem ligantes, com valores de 0,3 – 0,7Å, permite trabalhar com qualquer uma delas, sendo descartado o trabalho com estruturas contendo ligantes ao não ser conhecido o ligante alvo da TM1030.

No arquivo da estrutura 1Z77, obtido do PDB com formato de extensão .pdb, estão disponíveis diferentes parâmetros da estrutura, dentre eles a posição de cada átomo da proteína, necessários para rodar a dinâmica molecular.

4.1.2 Escolha do campo de força e preparação da estrutura

Cada programa de dinâmica molecular tem opções de trabalhar com diferentes campos de força e cada campo de força tem suas características próprias. No caso desta dissertação, o programa escolhido para realizar a dinâmica molecular foi o pacote GROMACS 4.5, que possui diferentes ferramentas para analisar as trajetórias simuladas. O campo de força usado foi o AMBER99SB, disponível dentro do pacote de GROMACS, o qual é o campo que nós achamos mais viável para trabalhar com solvente explícito (opção empregada na dinâmica molecular) ou com solvente implícito (opção empregada nas análises de modos normais), permitindo assim trabalhar com um único campo de força ao decorrer do estudo.

A preparação da estrutura 1Z77 consistiu em trocar na sua sequência os resíduos de selênio-metionina (SME) por resíduos de metionina (MET), já que a técnica cristalográfica usada foi a de substituir os resíduos de MET por SME, usando o programa **Coot** do CCP4. Em seguida, foram adicionados os hidrogênios não explícitos no arquivo .pdb, o que é necessário para trabalhar com o potencial escolhido. A partir deste arquivo .pdb modificado e usando o programa **pdb2gmx**, do pacote GROMACS, foram gerados os arquivos estrutural, de extensão .gro, e de topologia, de extensão .top, necessários para as dinâmicas no GROMACS. O arquivo de topologia inclui as massas dos diferentes átomos, as cargas parciais e as diferentes forças de interação e depende dos arquivos de extensão .itp que são gerados de forma automática ao somar a caixa de moléculas de água e íons, como é mostrado a continuação.

4.1.3 Criação da caixa de moléculas de água e adição dos íons

Ao ter os arquivos de estrutura e topologia da proteína prontos, é necessário definir um sistema no qual as dinâmicas moleculares serão executadas. O sistema mais comum usado neste tipo de dinâmicas é o de solvente explícito, onde a proteína é solvatada com moléculas de água e alguns íons. Dependendo do tipo de dinâmica, é possível escolher se a fronteira da região solvatada tenha ou não condições periódicas. Para nosso caso, trabalhamos com condições periódicas, inserindo a proteína numa caixa cúbica de moléculas de água com 9 nm

de comprimento, assegurando que a proteína não tem uma interação direta com suas reflexões.

Empregando o arquivo de estrutura obtido antes e o programa **editconf**, do pacote GROMACS, foi definida uma forma cúbica para a caixa de moléculas de água e a posição centrada para a proteína dentro desta, obtendo um novo arquivo .gro com esses parâmetros. Usando este novo arquivo estrutural e o arquivo de topologia, as moléculas de água necessárias para ocupar o espaço assinado foram inseridas com o programa **solvate**, do pacote GROMACS, gerando dois novos arquivos de estrutura e topologia.

Utilizando o programa **grompp**, do pacote GROMACS, se preparou um arquivo .mdp para executar no programa **genion**, do pacote GROMACS, a inserção de íons necessários para neutralizar o sistema caixa de água proteína. Usando o programa **genion** foram adicionados 12 íons Na^+ e 10 íons Cl^- em posições aleatórias na caixa de água (concentração 0,2mM), neutralizando a carga total -2 da proteína TM1030. A concentração de trabalho é uma das recomendadas para executar uma dinâmica molecular de proteína (se aconselha trabalhar com concentrações na faixa de 0,1 – 0,3mM). Como resultado, foram obtidos dois novos arquivos de estrutura e topologia com a proteína, as moléculas de água e os íons, sendo estes todos os átomos necessários para calcular a dinâmica molecular.

4.1.4 Minimização e equilíbrio do sistema

Têm-se dois importantes procedimentos de preparação do sistema necessários para executar a dinâmica molecular, sendo eles a minimização e equilíbrio do sistema. O processo de minimização busca diminuir as forças entre os átomos da proteína para que esta não se desenrole no começo da dinâmica molecular. Para fazer a minimização do sistema, o pacote GROMACS utiliza o mesmo programa usado para rodar as dinâmicas ou criar a matriz hessiana dos modos normais, o **mdrun**. Esse programa tem três diferentes métodos (algoritmos) para realizar a minimização, que são detalhados no manual do pacote.

Usando o método de gradiente conjugado⁴, foram feitos 50.000 passos de minimização para a proteína TM1030, sobre os arquivos de estrutura e topologia contendo a proteína, as moléculas de água e os íons, com um tamanho de passo inicial de 0,01nm (igual ao tutorial do GROMACS, disponível na web da Bevanlab). (31)

⁴ Este método foi usado já que é o recomendado a usar quando a configuração do sistema se encontra longe da configuração de mínima energia.

Uma vez minimizado o sistema é preciso equilibrar as moléculas de água e os íons, deixando presa a proteína, para que estas se organizem de modo que sua interação não rompa as ligações da proteína quando estas ganharem energia cinética. Para isso, o **mdrun** permite restringir o movimento da proteína e só deixar que as moléculas de água e os íons se desloquem ganhando a energia da temperatura fixada. O tutorial de GROMACS recomenda fazer duas equilibrações. A primeira para estabilizar a temperatura, com o número de partículas, a temperatura e o volume constantes (NTV). A segunda para estabilizar a pressão, com o número de partículas, a pressão e a temperatura constantes (NPT).

Usando como entrada os arquivos obtidos na minimização e dando condições periódicas nas fronteiras da caixa de água, com o programa **mdrun**, foram feitos 10.000 passos de equilibração NTV (dinâmica, 2fs por passo = 200ps total) nas temperaturas de 293K, 323K e 353K. Com os novos arquivos gerados em cada equilibração NTV foram feitas equilibrações NPT para cada temperatura.

4.1.5 Dinâmica molecular em três temperaturas

Com o sistema preparado para executar as dinâmicas moleculares, estas foram realizadas nas três temperaturas de trabalho propostas. Usando condições periódicas nas fronteiras da caixa e o programa **mdrun**, foram calculadas dinâmicas moleculares de 200ns, com intervalos de tempo de 10ns, no supercomputador BlueGene/P, nas temperaturas de 293K, 323K e 353K, guardando as configurações a cada 250 passos (0,5ps). O pacote GROMACS tem dois tipos de arquivo de trajetória: o arquivo com extensão **.trr**, que guarda junto das estruturas os valores de energia e velocidades no tempo determinado pelo usuário, e o arquivo com extensão **.xtc**, que guarda só as estruturas. Como nosso trabalho é centrado em medidas de valores estruturais, somente foram usados os arquivos **.xtc**, que são de menor tamanho.

4.1.6 Alinhamento das trajetórias

Como a proteína não foi presa no centro da caixa para executar as dinâmicas e as condições periódicas fazem com que a molécula apareça do outro lado da caixa quando se encosta num dos lados, é preciso rearmar a proteína uma vez terminados todos os cálculos e tendo as trajetórias completas. No alinhamento, são eliminadas as rotações e translações da molécula para conseguir fazer as diferentes análises, como as de mudança conformacional.

Empregando o programa **trjconv**, do pacote GROMACS, foram alinhadas as trajetórias das dinâmicas moleculares da proteína TM1030 executadas a 293K, 323K e 353K. Como a proteína é um dímero, foi preciso alinhar usando a opção `-pbc cluster` e depois fazer um novo alinhamento para fixar a proteína no centro da caixa, eliminando assim todas as rotações e translações com a opção `-fit rot+trans`. Os arquivos de saída .xtc solicitados ao programa, só contém as estruturas no tempo, eliminando as moléculas de água e os íons (pois não foram realizados cálculos com eles), passando de 76.967 átomos por arquivo para apenas 6.774 átomos da proteína (com os hidrogênios inclusos).

4.1.7 Outras dinâmicas moleculares

Foram executadas dinâmicas moleculares nas temperaturas de 293K, 323K e 353K para a estrutura 4I76, retirando o ligante octanol e seguindo o protocolo estabelecido para as dinâmicas na estrutura 1Z77. Com estas dinâmicas, esperava-se observar o que acontece com a conformação estrutural quando se minimiza e se realizam as dinâmicas moleculares. Realizaram-se comparações de todas as trajetórias com relação as estruturas cristalográficas 1Z77, 4I76 e 1ZKG.

4.2 RMSD

Uma das primeiras medidas a realizar e ajuda a observar o comportamento das trajetórias é, além da conservação da energia total, o cálculo do desvio médio da estrutura molecular no tempo. Na continuação se detalha o procedimento realizado.

4.2.1 Calculo do RMSD

Usando o programa **trjcat**, do pacote GROMACS, procedeu-se a unir todos os arquivos de trajetória .xtc obtidos a cada 10ns, para cada uma das temperaturas (293K, 323K e 353K), obtendo três arquivos únicos com os 200ns das dinâmicas moleculares executadas. Empregando estes novos arquivos e o programa **rms**, do pacote GROMACS, foram gerados gráficos dos valores de RMSD vs. tempo. Empregando os valores de RMSD global obtidos, foram realizadas as comparações da mudança nas conformações frente à estrutura da proteína na conformação sem ligantes 1Z77.

4.3 Fatores de Temperatura

Os dois tipos de fatores de temperatura, isotrópicos e anisotrópicos, foram calculados para as três trajetórias da estrutura 1Z77. A continuação se detalha o procedimento que foi seguido.

4.3.1 Fatores de temperatura e estrutura média das trajetórias

Para calcular os fatores de temperatura é preciso encontrar os deslocamentos dos átomos na trajetória com respeito a uma estrutura média. Empregando o programa **rms**, do pacote GROMACS, foram feitas as análises dos últimos 150ns de cada trajetória, sendo descartados os primeiros 50ns porque apresentarem uma grande mudança do RMSD o que se interpreta como um tempo de equilíbrio do sistema. O programa calcula uma estrutura média de cada trajetória e usa estas como referências para calcular os fatores de temperatura isotrópicos e anisotrópicos, que saem juntos num arquivo .pdb.

4.3.2 Fatores de temperatura por faixas de tempo na trajetória

Para medir os fatores de temperatura é preciso conhecer os movimentos locais dos átomos. Para isso, se devem eliminar os deslocamentos por movimentos coletivos de baixa frequência. Um modo de fazer isso é realizando as medições dos fatores de temperatura em janelas de menor tempo. Em nosso trabalho, tomamos janelas de 1ns (com as quais se obtém uma boa aproximação) das três trajetórias a diferentes temperaturas, descartando os primeiros 50ns, obtendo 150 fatores de temperatura para as 150 janelas. Empregando o programa **rms**, se desenharam esses pontos para observar o comportamento.

4.4 Modos Normais

O segundo método de simulação computacional empregado foi a análise de modos normais da proteína TM1030 da bactéria *Thermotoga marítima*, trabalhando com a estrutura 1Z77. Com este método, foram calculadas as frequências de oscilação da molécula e, usando os menores valores, foram observados os movimentos coletivos desta. Assim como na dinâmica molecular, é necessário seguir um protocolo, detalhado a continuação.

4.4.1 Minimização da estrutura

Para conseguir uma boa aproximação do potencial real usando um potencial quadrático é preciso fazer uma minimização até atingir uma força máxima de entre 10^{-4} e 10^{-12} $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-1}$, como recomendam diferentes autores. (29) Para isto, são necessários uma quantidade alta e indefinida de passos de minimização (na ordem de centenas de milhares) e o uso de diferentes algoritmos. Usando o potencial AMBER99SB, do pacote GROMACS 4.5.5 com dobro de precisão, e a técnica de solvente implícito, foi feita a minimização da estrutura 1Z77 obtida no PDB. O primeiro algoritmo de minimização empregado foi o CG, alcançando um gradiente na ordem de 10^{-3} $\text{J}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. A partir desta minimização, foi empregado o algoritmo steepest-descent, o qual fez poucos passos e manteve o gradiente na ordem de 10^{-3} $\text{J}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. No entanto, a partir da minimização com o algoritmo steepest-descent, se conseguiu executar uma segunda minimização com o algoritmo CG que diminuiu a força máxima até a ordem de 10^{-5} $\text{J}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. Por último, foi feita uma minimização com o algoritmo L-BFGS chegando a um gradiente de força de $6,4 \times 10^{-8}$ $\text{J}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$.

4.4.2 Criação e diagonalização da matriz hessiana

A matriz hessiana com as posições de todos os átomos é gerada a partir da estrutura minimizada. Usando o programa **mdrun_d**, do pacote GROMACS, geramos a matriz hessiana de $3N$ componentes, onde N é o número de átomos da proteína. Como a estrutura 1Z77 contém 6.774 átomos, a matriz gerada foi de dimensão 20.322×20.322 . Empregando o programa **nmeig**, do pacote GROMACS, a matriz foi diagonalizada obtendo os valores dos autovetores e dos autovalores. Os autovetores foram obtidos num arquivo de trajetória, com extensão **.trr**, onde cada frame representa um vetor. Os valores dos autovalores foram obtidos num arquivo de gráfico de função, com extensão **.xvg**, onde cada ponto é um autovalor.

4.4.3 Trajetórias virtuais dos autovetores

Usando o programa **mmtraj**, do pacote GROMACS, e os arquivos dos autovalores e autovetores, foram geradas as trajetórias oscilantes virtuais para cada autovetor (descartando os 6 primeiros valores por serem as rotações e translações da proteína toda) dos primeiros 10 modos normais. Empregando o programa VMD, as trajetórias foram representadas.

5 RESULTADOS E ANALISES

Todos os resultados obtidos são apresentados e discutidos neste capítulo. Primeiro iremos discutir as dinâmicas moleculares feitas na estrutura da proteína TM1030 da bactéria *Thermotoga marítima*, nas temperaturas de 293K, 323K e 353K, os deslocamentos atômicos dessas três dinâmicas (fatores de temperatura isotrópicos e anisotrópicos) e o cálculo do RMSD das três trajetórias de dinâmica molecular. Na continuação serão analisados seus modos normais, fazendo uma descrição para cada modo que apresenta um movimento significativo da cadeia principal. Por último, serão discutidas as mudanças nos fatores de temperatura isotrópicos com a variação da temperatura nas três dinâmicas moleculares.

5.1 Dinâmica molecular

A partir das dinâmicas moleculares são medidas as variações do RMSD no tempo da trajetória com relação a estrutura inicial (5.1.1), os fatores de temperatura isotrópicos e anisotrópicos (5.1.2) representados pelos deslocamentos atômicos da trajetória e as estruturas médias das dinâmicas nas três temperaturas (5.1.3). Esses dados são comparados com a estrutura cristalográfica 1Z77 da TM1030, depositada no PDB, que é o estado ativo de acordo com os autores da estrutura. (16) Os primeiros resultados a analisar são os obtidos das dinâmicas moleculares, com a exceção dos fatores de temperatura.

5.1.1 RMSD com respeito da estrutura cristalográfica

Na Figura 7 se pode ver que depois dos primeiros 50ns as mudanças nos valores do RMSD para as três trajetórias calculadas diminuem, sendo ainda mais acentuado para as dinâmicas nas maiores faixas de temperatura (linhas vermelho e azul). No caso da trajetória de menor temperatura (linha preta), observa-se que a estrutura na dinâmica molecular se afasta mais da estrutura cristalográfica devido aos maiores valores de RMSD, atingindo os 4,5Å. Isso não significa que as estruturas médias das outras duas trajetórias sejam iguais entre si, mas sim que as duas estruturas têm uma diferença estrutural média igual em relação à estrutura cristalográfica 1Z77.

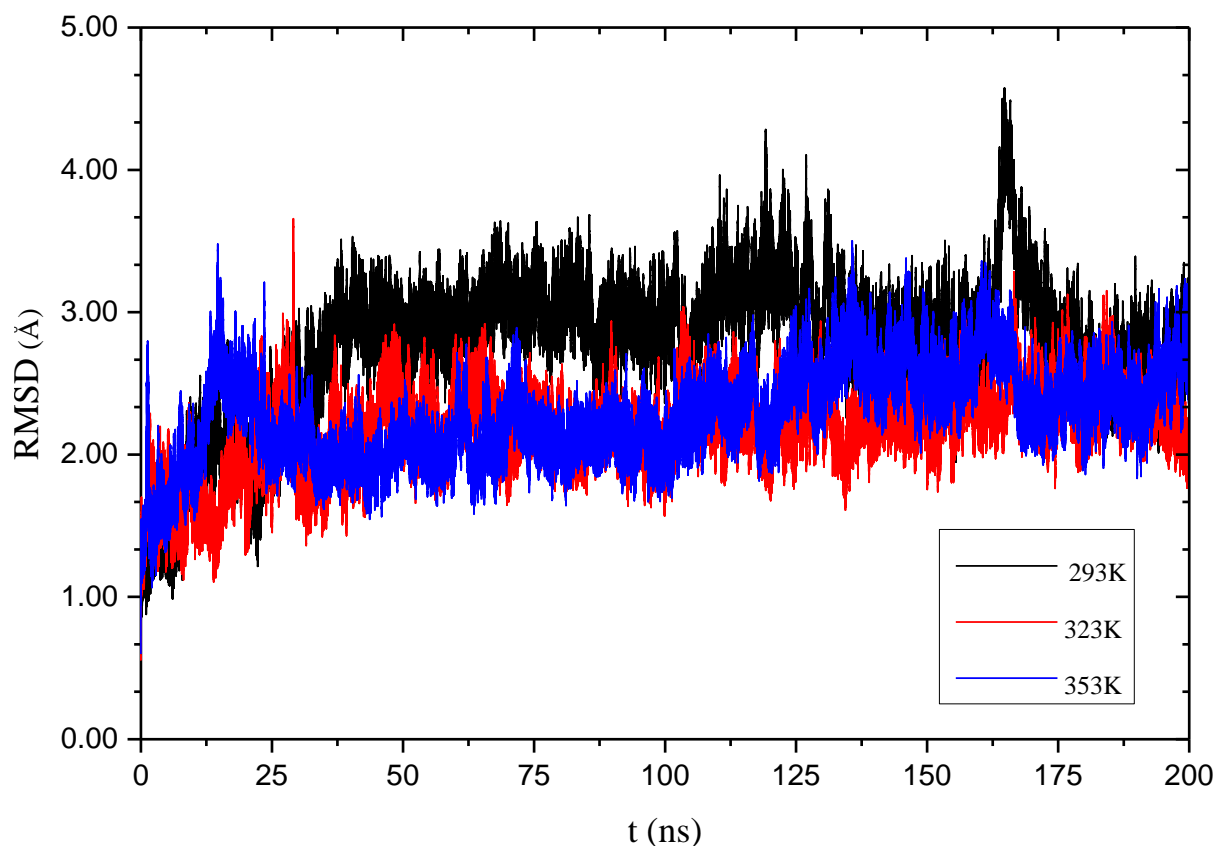


Figura 7 - RMSD em função do tempo em 200ns de trajetória das dinâmicas da estrutura 1Z77 nas temperaturas de 293, 323 e 353K, comparadas com a estrutura cristalográfica 1Z77.

Fonte: Elaborada pelo autor

No caso do RMSD calculado para as estruturas médias das três trajetórias a diferentes temperaturas é possível observar um dado interessante: os valores de RMSD diminuem com o aumento da temperatura, como mostra a Tabela 1.

Tabela 1 - RMSD médio em diferentes faixas de tempo com respeito à estrutura cristalográfica PDB 1Z77 (RMSD em Å).

Faixa de tempo da dinâmica molecular (ns)	RMSD médio vs. 1Z77 (Å)		
	293K	323K	353K
50-200	2,62	1,83	1,71
100-200	2,67	1,79	1,69
150-200	2,59	1,73	1,50

Fonte: Elaborada pelo autor

É possível observar também que, assim como na Fig. 7, os valores de RMSD nas trajetórias às temperaturas 323 e 353K estão muito próximos com diferenças na ordem de 0,1Å. Já no caso da trajetória de menor temperatura, 293K, o valor de RMSD é por volta de 1Å maior que o das outras duas trajetórias. Isso implica que há maiores mudanças conformacionais que se

afastam da estrutura cristalográfica em baixas temperaturas. Na seção 5.1.3 se faz uma análise disso.

Um comportamento similar ocorreu nas dinâmicas da estrutura 4I76, Figura 8. Ao calcular o RMSD com relação à estrutura 1Z77, observa-se uma estabilização nos primeiros 50ns que é mais rápida para a trajetória a 323K que para a trajetória a 293K (não se usaram os dados da trajetória a 353K porque a proteína se desenovelou por volta dos 70ns). Note-se também que o RMSD diminui até valores abaixo de 2Å, onde a estrutura fica mais próxima da estrutura 1Z77, determinada como ativa, do que da estrutura 4I76, considerada como inativa.

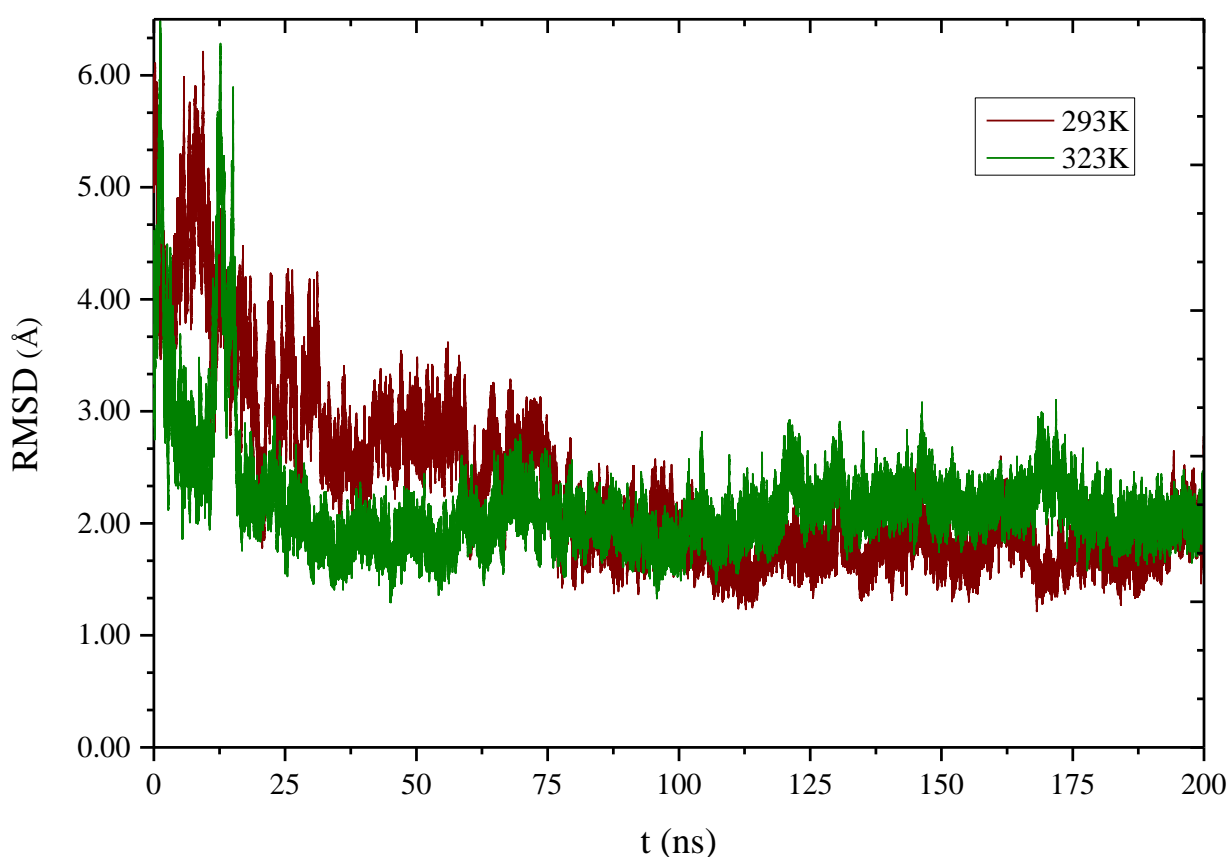


Figura 8 - RMSD em função do tempo em 200ns de trajetória das dinâmicas da estrutura 4I76 nas temperaturas de 293, 323 e 353K, comparadas com a estrutura cristalográfica 1Z77.

Fonte: Elaborada pelo autor

5.1.2 Fatores de temperatura anisotrópicos das trajetórias

Os desenhos dos fatores de temperatura anisotrópicos representam os movimentos dos átomos dentro da estrutura. Como nosso interesse é analisar os movimentos de menor frequência, os que geram mudanças conformacionais ao deslocar a cadeia principal, foram unicamente representados os fatores de temperatura anisotrópicos dos carbonos- α da proteína, Figura 9.

Para esta representação, se empregaram novamente os últimos 150ns de cada trajetória, calculando as estruturas médias a partir dos deslocamentos nesse tempo.

Assim, nas Figuras 7-9 podemos observar que para as diferentes temperaturas de trabalho os carbonos- α da proteína que apresentam maiores deslocamentos (maiores elipsoides) são os que se encontram na região de ligação do ligante alostérico, na região HTH (que se liga ao DNA) e na região mais externa que compreende os resíduos 107-138 e 171-182. Do mesmo modo, é possível observar que o aumento da temperatura leva a um incremento nos deslocamentos dos diferentes domínios, ou seja, um aumento em suas flexibilidades, mas ainda assim, como vai ser visto no próximo tópico, a estrutura média que se mantém mais próxima da estrutura cristalográfica 1Z77 é a de maior temperatura (353K). Neste resultado encontramos que não é confirmado o esperado experimentalmente na cristalografia em quanto a que a maior temperatura os fatores de temperatura desta proteína deveriam diminuir. Mas ainda assim acontece o esperado fisicamente que é um aumento dos deslocamentos dos átomos. Ao final deste capítulo se faz uma melhor abordagem deste problema usando também os fatores de temperatura isotrópicos. Note-se também que os átomos do domínio de dimerização e dos resíduos 56-72 não apresentam maiores câmbios na amplitude do deslocamento com o aumento da temperatura.

Outro fato que se pode ver é que a forma e as direções dos elipsoides são diferentes para cada uma das temperaturas e que esta diferença é ainda maior na trajetória de maior temperatura em relação às duas de menor temperatura. No desenho de maior temperatura foram apagados três aminoácidos que se desenovelarem no final da trajetória, isso se pode esperar porque estes campos de força não estão otimizados para rodar dinâmicas moleculares em temperaturas muito altas.

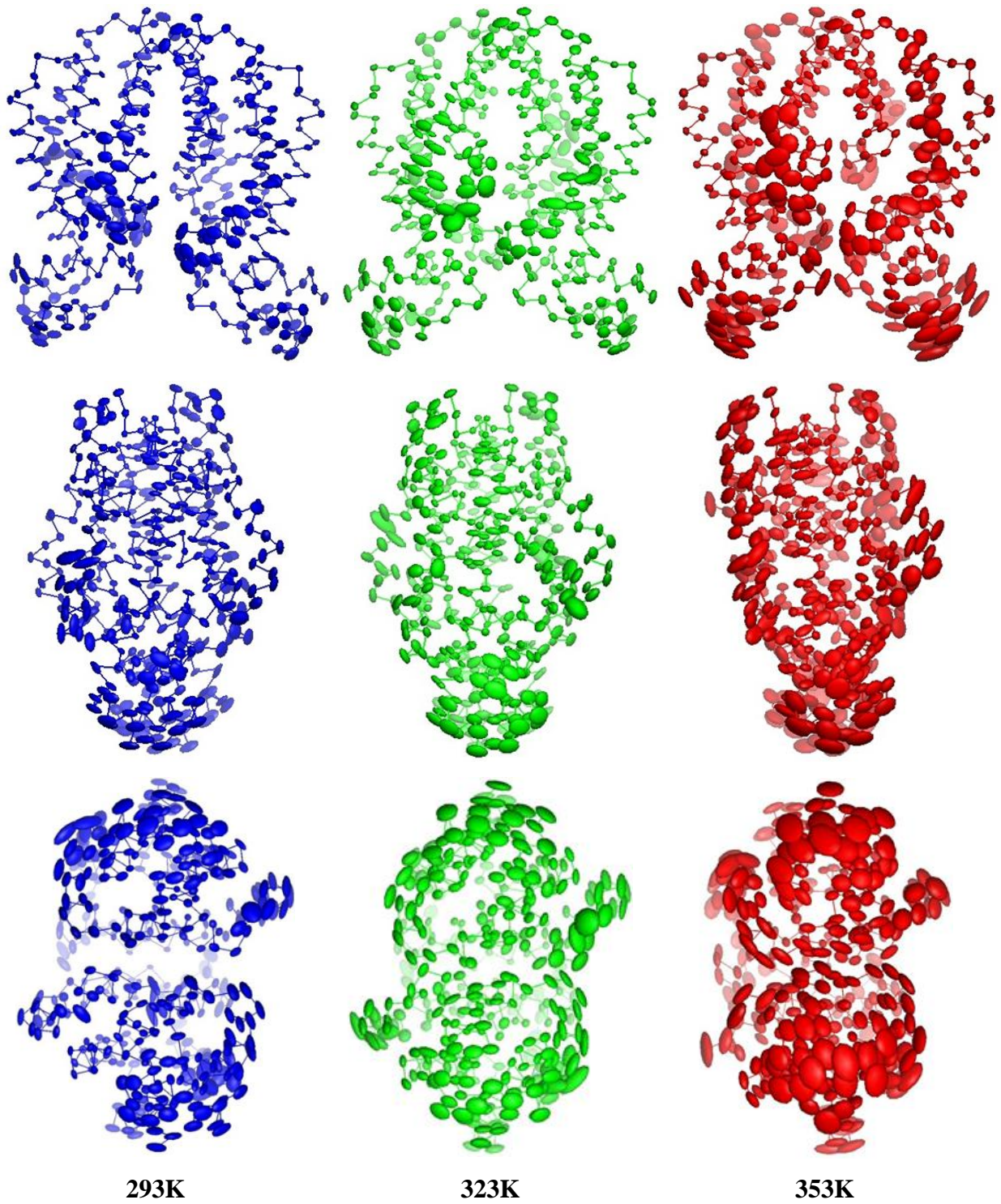


Figura 9. Comparação dos fatores de temperatura anisotrópicos nas dinâmicas moleculares executadas a 293K (azul), 323K (verde) e 353K (vermelho).

Fonte: Elaborada pelo autor.

5.1.3 Estrutura média das dinâmicas moleculares

A Figura 10 representa os alinhamentos das estruturas médias das dinâmicas moleculares executadas nas três temperaturas contra a estrutura cristalográfica 1Z77. Pode-se observar que com o aumento da temperatura algumas das hélices vão se aproximando às posições que estas apresentam na estrutura 1Z77, isto é, a estrutura média das dinâmicas tem uma geometria mais próxima da cristalográfica nestas maiores temperaturas. A estrutura média obtida a 353K é a que fica com um RMSD mais próximo da estrutura 1Z77. Note-se que as hélices do domínio HTH se vão aproximando com o aumento da temperatura, sendo que na maior das temperaturas os domínios se encontram mais próximos que na estrutura cristalográfica. Ainda assim outras hélices, da região mais externa, se alinham melhor e por isso os valores menores de RMSD.

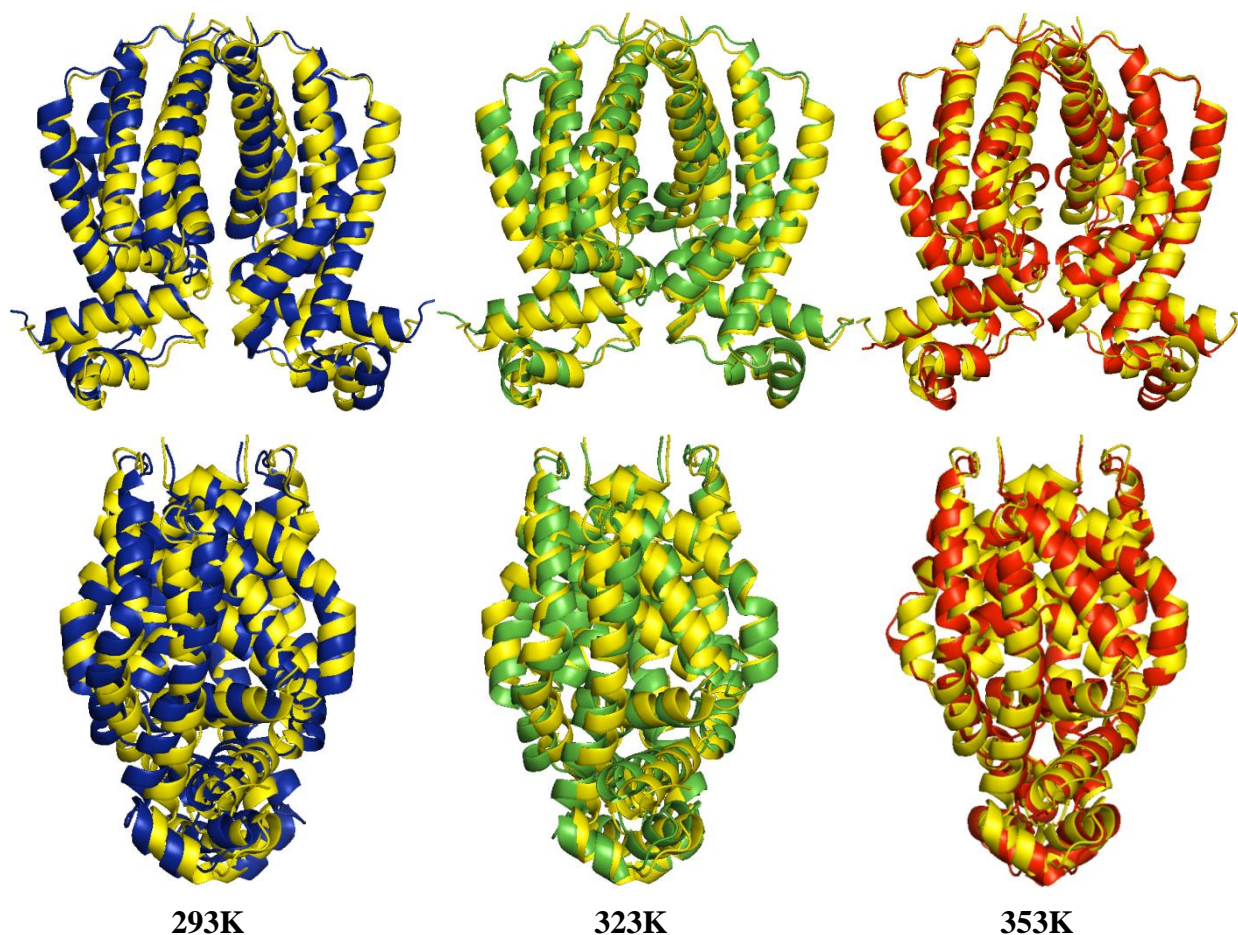


Figura 10 - Alinhamentos das estruturas médias das trajetórias a 293K (azul), 323K (verde) e 353K (vermelho) em relação à estrutura da 1Z77 (amarelo).

Fonte: Elaborada pelo autor.

Medindo a distancia entre os centros de massa das duas hélices que ligam o DNA (H3 na Figura 3) se encontrou que no estado ativo esta distancia é de 32,1 Å em quanto que para o estado inativo era de 50,2 Å. Sabendo isso se fez uma medição da distancia entre os centros de massa das hélices em cada uma das três trajetórias a diferentes temperaturas os resultados estão representados na Figura 11:

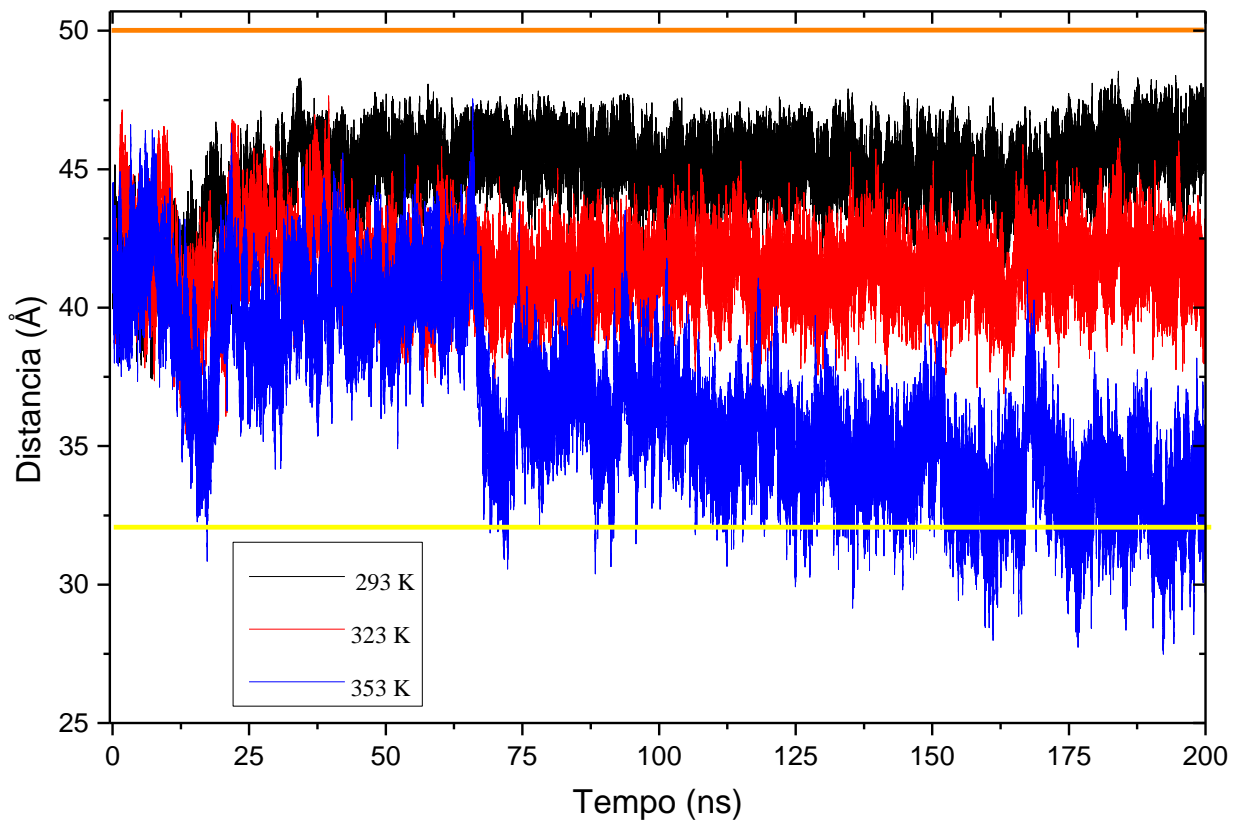


Figura 11 – Distância entre as hélices H3 em função do tempo em 200ns de trajetória das dinâmicas da estrutura 1Z77 nas temperaturas de 293, 323 e 353K.

Fonte: Elaborada pelo autor.

Note-se que assim como os RMSD (Figura 7) mostram nas três dinâmicas a proteína se mantém numa conformação intermedia e neste caso as hélices alfa também estão numa distancia intermedia entre o estado ativo e o inativo. Mas na dinâmica a 353K esta distancia vai diminuindo com o tempo até conseguir atingir um tamanho com que conseguiria se ligar no DNA. Também se pode observar que próximo dos primeiros 20 ns esta distancia é atingida, mas como mostra o RMSD antes dos 50 ns a proteína se esta ainda estabilizando e se separada rapidamente. Outra observação que se pode realizar é que próximo dos 70 ns esta distancia tem uma rápida mudança dos 40 Å para os 35 Å aproximadamente. Calculando a estrutura média antes e depois dos 70 ns na dinâmica a 353 K se observa que tem uma transição da estrutura em que os domínios HTH se aproximam como mostra a Figura 12:

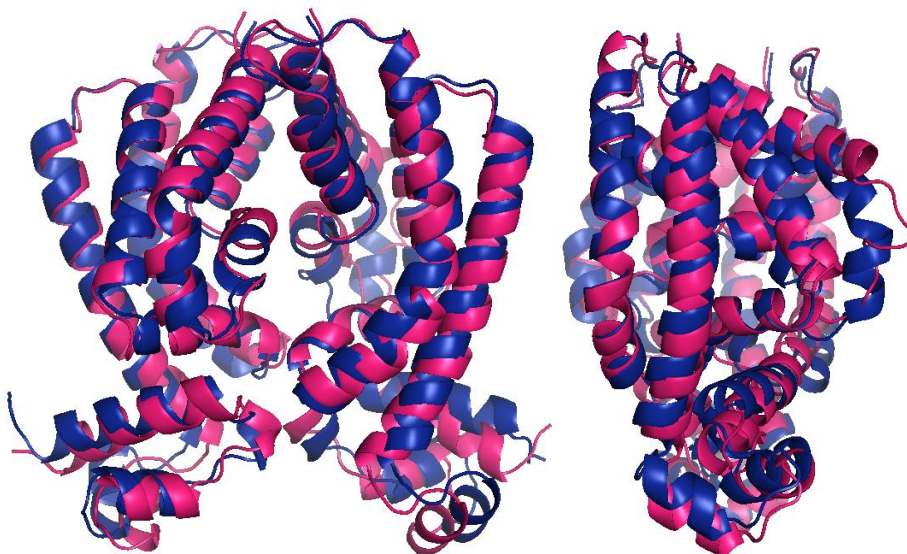


Figura 12 - Alinhamentos das estruturas médias da trajetória 353K na estrutura 1Z77, entre os 40 e 60 ns (azul) e entre os 180 e 200 ns (magenta).

Fonte: Elaborada pelo autor.

Este resultado indicaria que numa temperatura da ordem da temperatura fisiológica para a bactéria a proteína TM1030 teria uma tendência a se ativar com o que conseguiria realizar sua função biológica. Fazendo outras dinâmicas moleculares da proteína dentro da faixa dos 55 até 90 °C se poderia comprovar se esta tendência se mantém e qual é o rol da temperatura na ativação.

5.2 Modos normais

Depois de feito o cálculo de modos normais foram selecionadas as menores frequências, já que estas representam as maiores amplitudes do movimento da proteína ao ser excitada numa temperatura da ordem dos 300K. Uma explicação do comportamento de cada modo normal e sua relação com o movimento da proteína é detalhado na continuação.

5.2.1 Descrição dos modos normais representativos

Como a proteína tem 6774 átomos se obtém 20322 modos normais. Destes os primeiros 6 são descartados porque correspondem às rotações e translações da proteína como conjunto e não são de interesse para ser analisados. Em nosso caso, os modos vibracionais 7 até 12 têm um

deslocamento coletivo significativo, ou seja a cadeia principal se deforma e muda de conformação. A partir do modo 13 os deslocamentos passam a ser mais localizados, ou seja, os deslocamentos da cadeia principal são muitos menores comparados com os primeiros quando se estiver no mesmo nível de energia (menor que 2%), e são os átomos das cadeias laterais os que passam a serem mais excitados. Note-se o fato que sejam só 6 modos normais os que representam os deslocamentos da cadeia na proteína, ao serem estes modos os de menor frequência são a sua vez os de menor energia. Também reflete o fato que para poder ter a possibilidade de encontrar estas transições numa trajetória de dinâmica molecular se precisaria de tempos muito longos, pelo menos dois períodos dessa frequência e sem poder assegurar em que tempo vai ser observado o fenômeno. Para poder ter uma estatística boa é preciso atingir no mínimo 6 vezes o período, que para o caso de transições alostéricas estaria na ordem dos microssegundos.

A seguir se representará com desenhos (Figuras 13 a 18) e se fará uma descrição de cada modo normal usando setas para marcar as direções dos movimentos em três diferentes vistas da proteína. Um quarto desenho mostra as regiões de maior (vermelho) e menor (azul) movimento como se fossem fatores de temperatura isotrópicos. Usou-se o mesmo fator multiplicador (50) para as amplitudes de todos os modos normais e este foi escolhido por simples conveniência, foi o menor valor que permitia observar os deslocamentos nos gráficos.

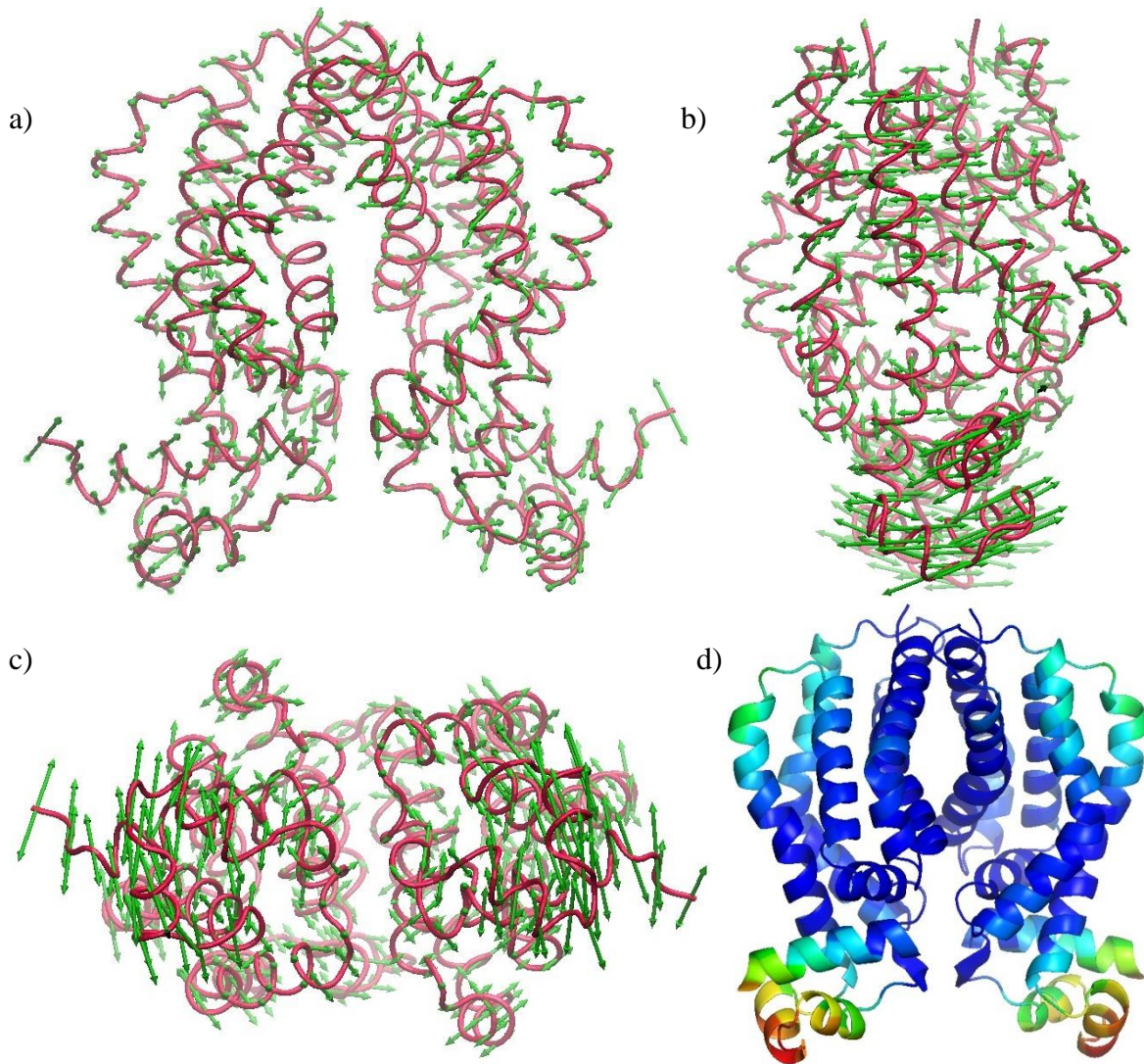


Figura 13 - Vistas do modo vibracional 7 desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho

Fonte: Elaborada pelo autor

Neste primeiro modo, o modo sete (Figura 13), se pode observar que a região com maior movimento é a do domínio HTH. O movimento dos dois monômeros tem a mesma direção o que nesta dissertação chamaremos de movimento em paralelo. Note-se que este movimento alinha ou desalinha as hélices do domínio HTH que se ligam ao DNA o que favoreceria ou desfavoreceria a ligação da proteína no sitio específico de regulação. No caso da região que compreende os aminoácidos 50 a 200 os pequenos movimentos observados nos carbonos- α são consequência do movimento do domínio HTH, já que, é preciso reconfigurar as posições para manter as forças intermoleculares.

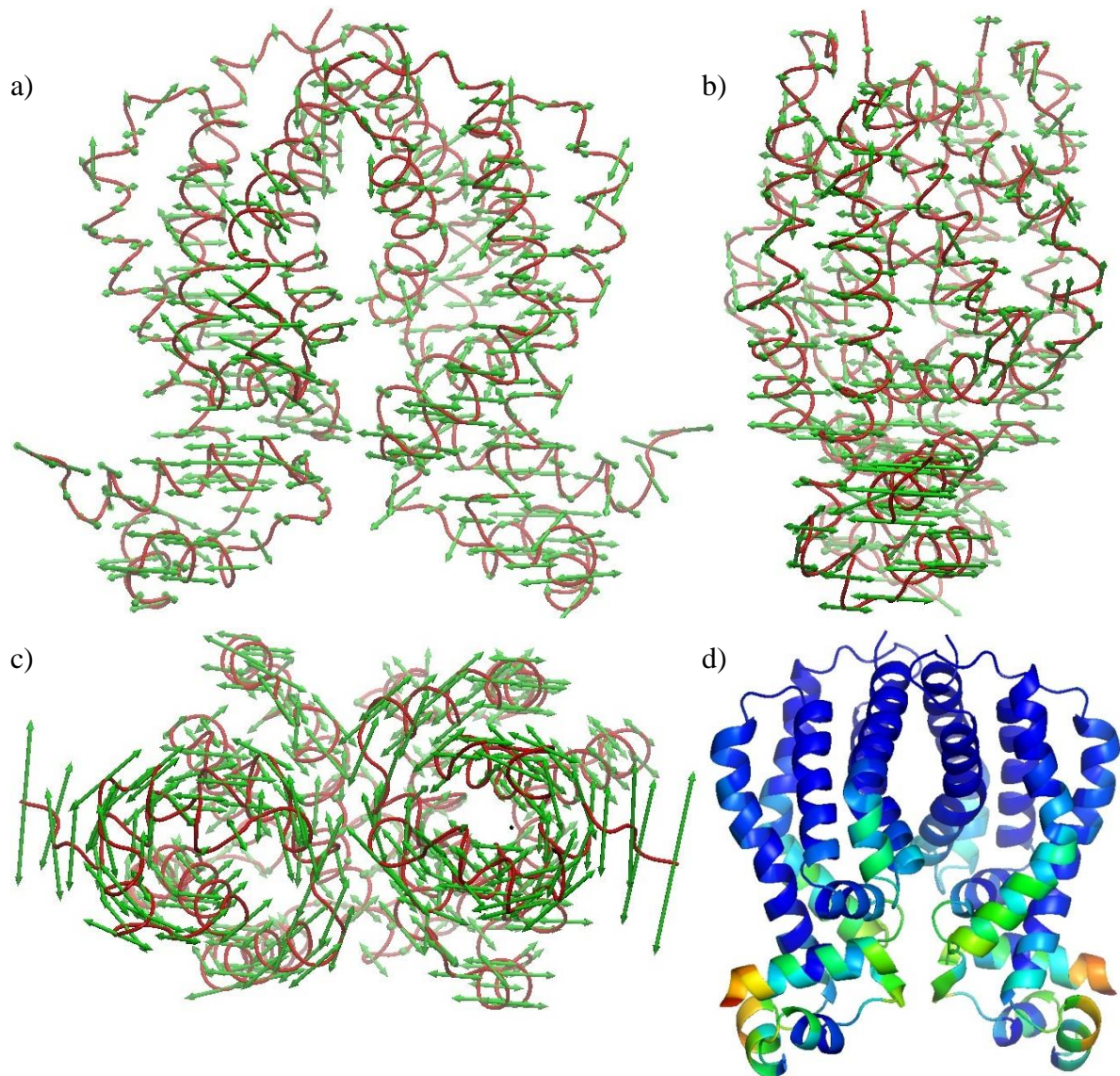


Figura 14 - Vistas do modo vibracional 8 desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho

Fonte: Elaborada pelo autor.

No segundo modo normal, o modo oito (Figura 14), se observa um movimento maior no domínio HTH em relação aos demais átomos da proteína. Este movimento tem duas diferenças significativas ao comparar com o modo anterior, a primeira é que as direções são ligeiramente diferentes, os vetores são praticamente paralelos ao eixo determinado pelo domínio de dimerização. A segunda diferença é que um monômero se move contrário em relação ao outro, ou seja, se movem de forma antiparalela. Este movimento também alinha ou desalinha os domínios HTH dos monômeros o que pode influenciar sua ligação ao DNA.

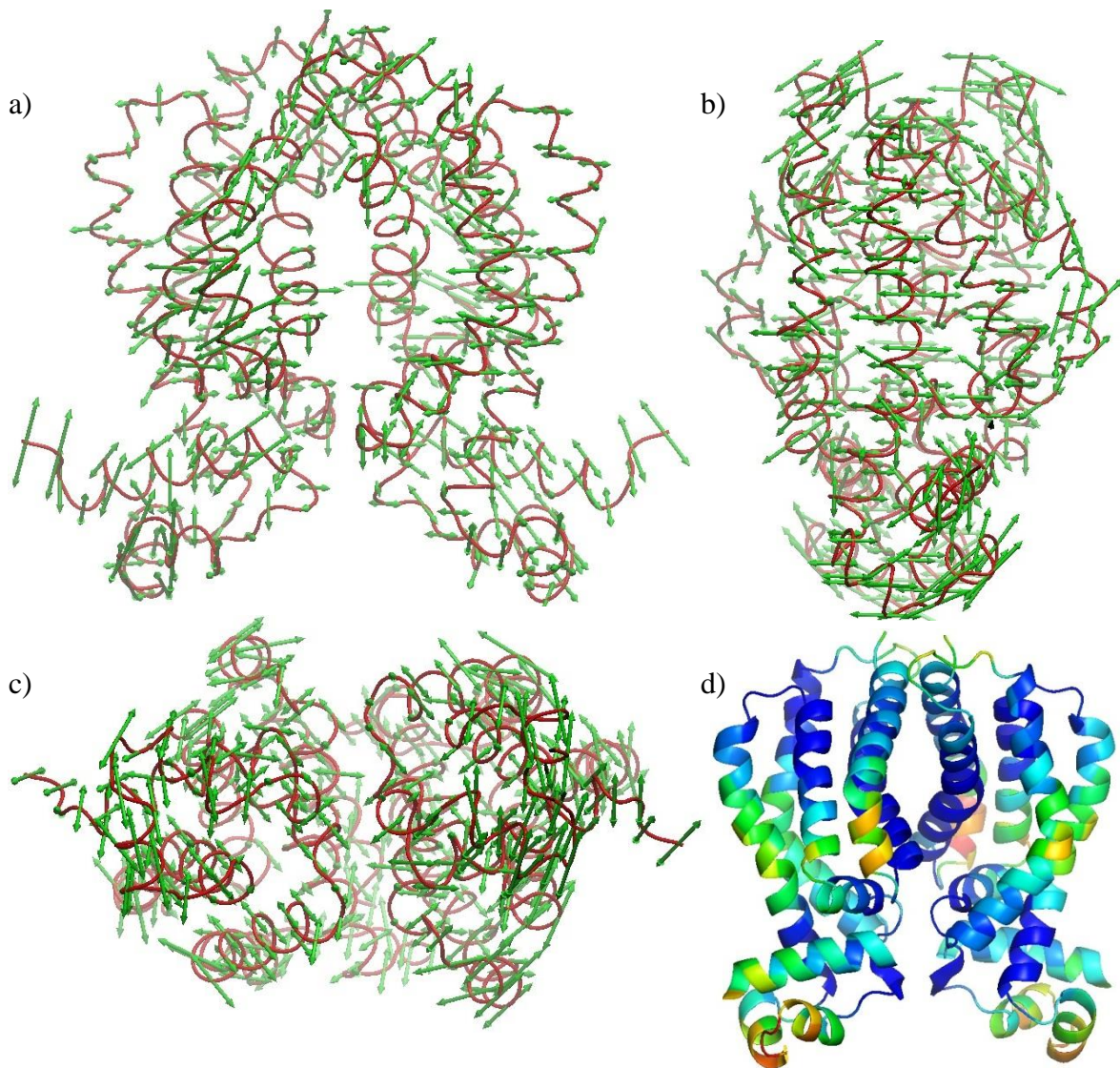


Figura 15 - Modo vibracional 9 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho

Fonte: Elaborada pelo autor.

Neste modo normal, modo nove (Figura 15), se pode ver uma diminuição do deslocamento dos átomos do domínio HTH e um incremento no movimento da região que compreende o sítio de ligação alostérico. Este modo normal seria o que está relacionado á reconfiguração destas hélices com a ligação ou desprendimento do ligante do sítio alostérico. Este movimento poderia representar um primeiro passo na transição alostérica, no qual ao se ligar o ligante alostérico os domínios HTH se deslocam, desalinhando as duas hélices que se ligam ao DNA e por tanto inativando a proteína.

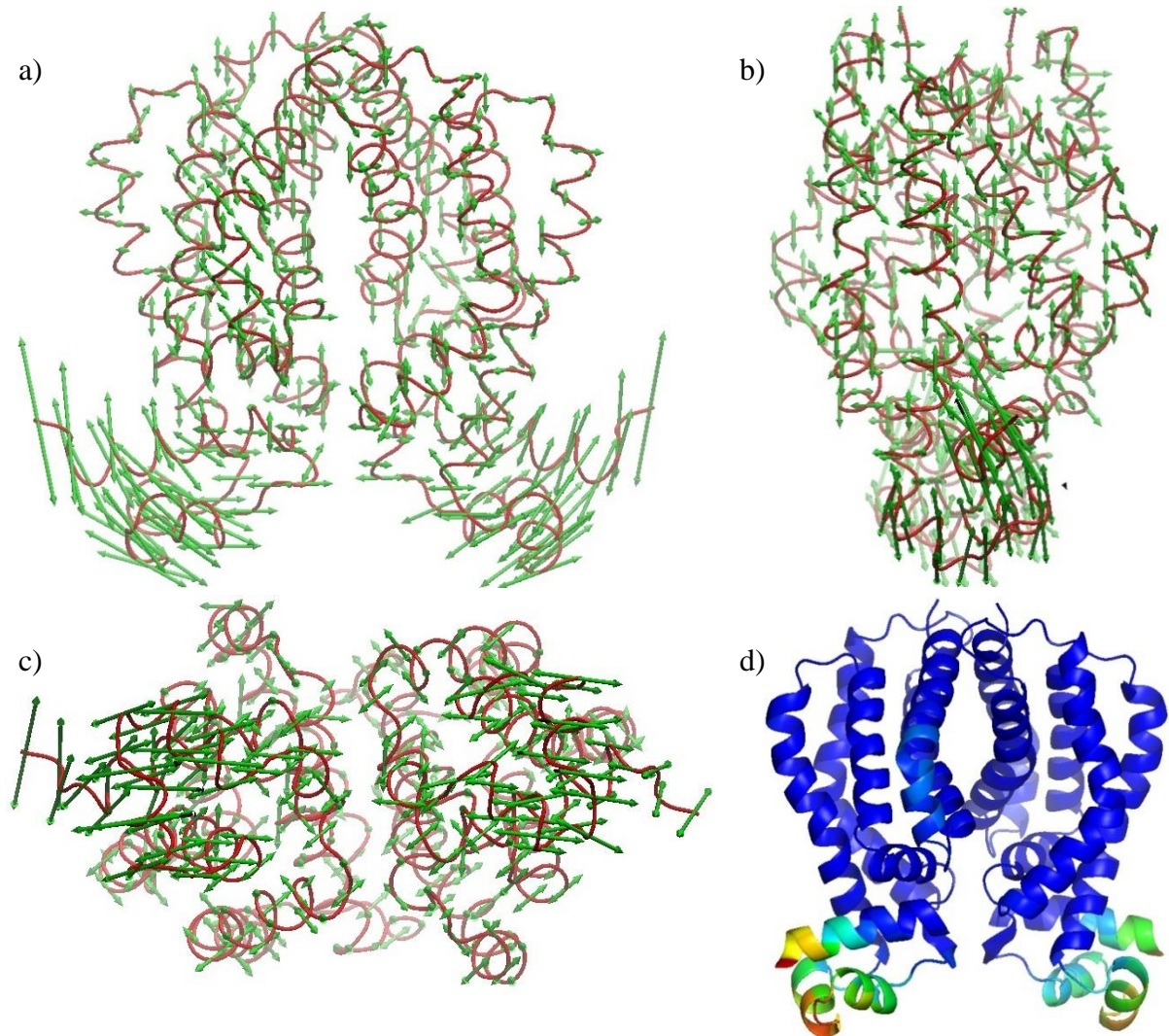


Figura 16 - Modo vibracional 10 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.

Fonte: Elaborada pelo autor.

Este modo normal, o modo dez (Figura 16), apresenta o maior dos deslocamentos do domínio HTH. O movimento realizado neste modo aumento ou diminui a distancia entre as duas hélices que ligam ao DNA o que ativa ou inativa a proteína para realizar sua função biológica. O movimento é realizado no mesmo sentido nos dos monômeros sendo este paralelo. Ainda que este movimento mostre o deslocamento dos HTH necessário para mudar entre as duas configurações alostéricas ao não ter movimento das hélices no sitio alostérico, não se pode relacionar diretamente com a transição alostérica. Pode ser que se precise de uma combinação linear com outro ou outros modos para observar este fenômeno. Uma análise disso será feita no numeral 5.3.2.

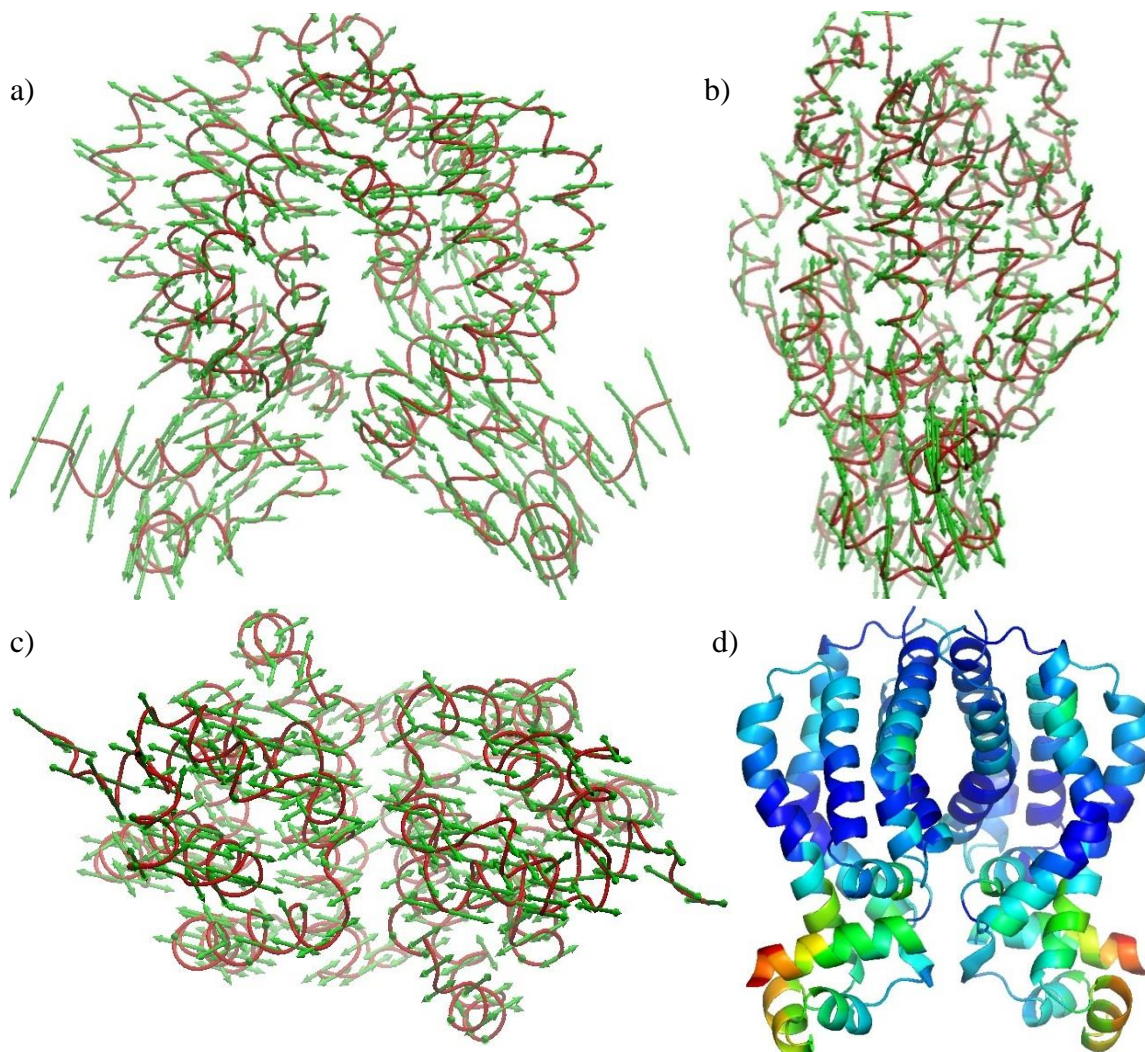


Figura 17 - Modo vibracional 11 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho.

Fonte: Elaborada pelo autor.

No modo normal onze (Figura 17) temos um movimento similar ao realizado no modo anterior, mas com três diferenças importantes. A primeira diferença é que o deslocamento dos domínios HTH é menor que a do modo anterior e a distancia entre eles se mantém fixa. A segunda que neste caso os dois monômeros se movem em sentidos contrários, ou seja, antiparalelos. E o terceiro e muito importante é que neste caso a região próxima ao sítio de ligação alostérico também tem um movimento significativo o que poderia mostrar uma abertura da região para a entrada do ligante.

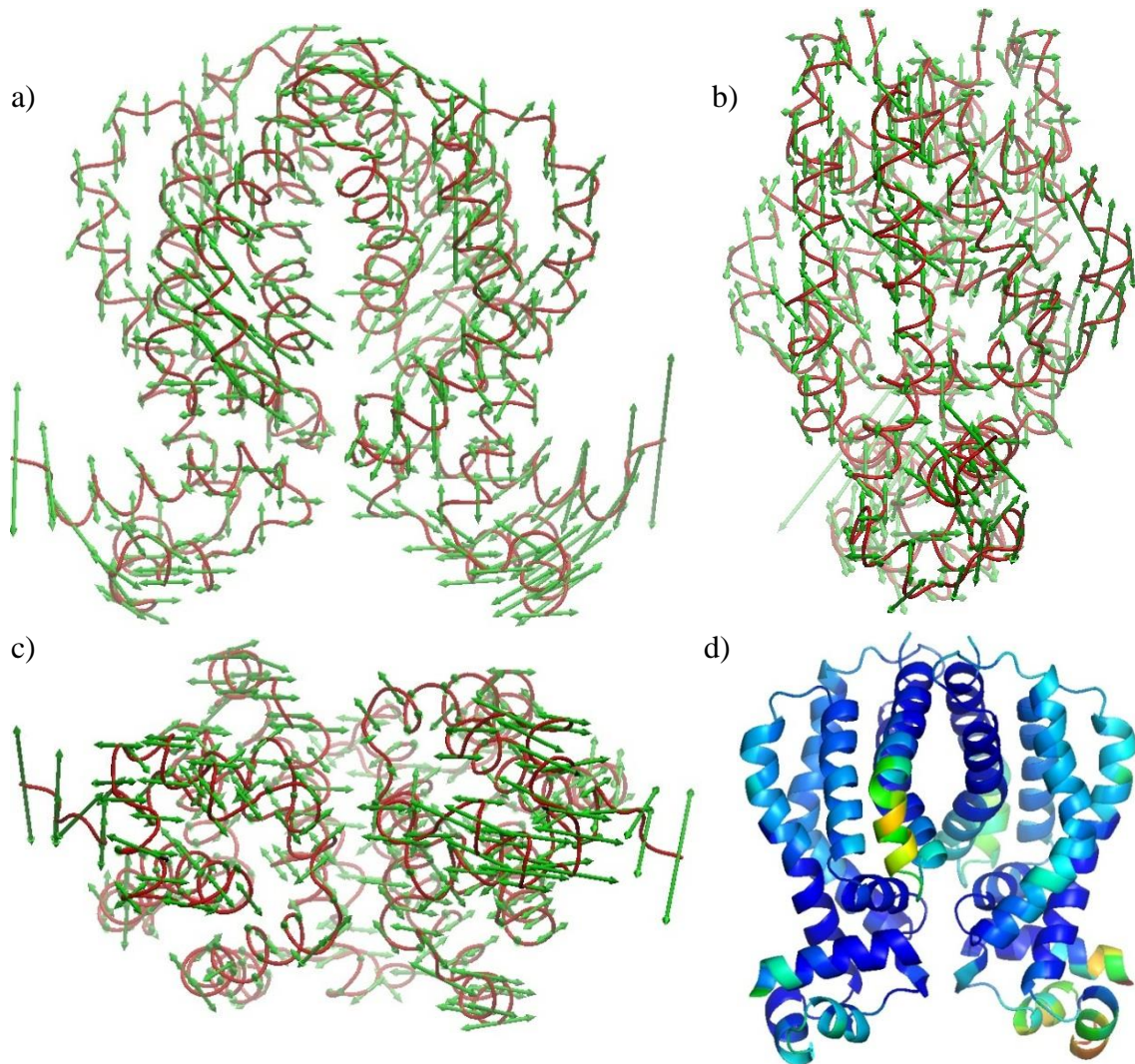


Figura 18 - Modo vibracional 12 visto desde três diferentes posições (a, b e c). No último desenho (d) se representam os deslocamentos dos átomos desde os menores em azul até os maiores em vermelho

Fonte: Autor usando o programa VMD e PyMol

Este último modo normal, modo doze (Figura 18), é praticamente idêntico ao anterior mas com uma mudança na direção de deslocamentos dos átomos sendo as novas direções perpendiculares às anteriores. De novo se observa um deslocamento importante tanto no domínio HTH, que aumenta e diminui a distância entre as hélices HTH, mas com menor amplitude que o modo dez, como na região que contém o sítio alostérico. Este modo seria um bom candidato para representar a transição alostérica, mas não possui todos os movimentos necessários das hélices e por tanto ainda se acredita que deve ser uma combinação linear a que representaria o movimento da transição alostérica.

Uma análise de como estes modos influenciam o movimento da proteína e sua relação com os movimentos observados nas trajetórias de dinâmica molecular será feito mais na frente deste texto.

5.2.8 Modos normais restantes

Os demais modos normais não apresentam um deslocamento maior dos carbonos- α , ou seja, sua cadeia principal tem um deslocamento quase imperceptível se comparado com o dos modos normais anteriores, menor de que 2% do movimento total da proteína. Os movimentos se começam a apresentar nas cadeias laterais dos aminoácidos.

Como se observa no desenho, Figura 19, os movimentos são quase imperceptíveis, já que, não se vê mais de que as pontas das setas. Desde o modo 13 para frente os movimentos atômicos ficam localizados nas cadeias laterais e por último nos hidrogênios que são os que vibram com frequências maiores na proteína.

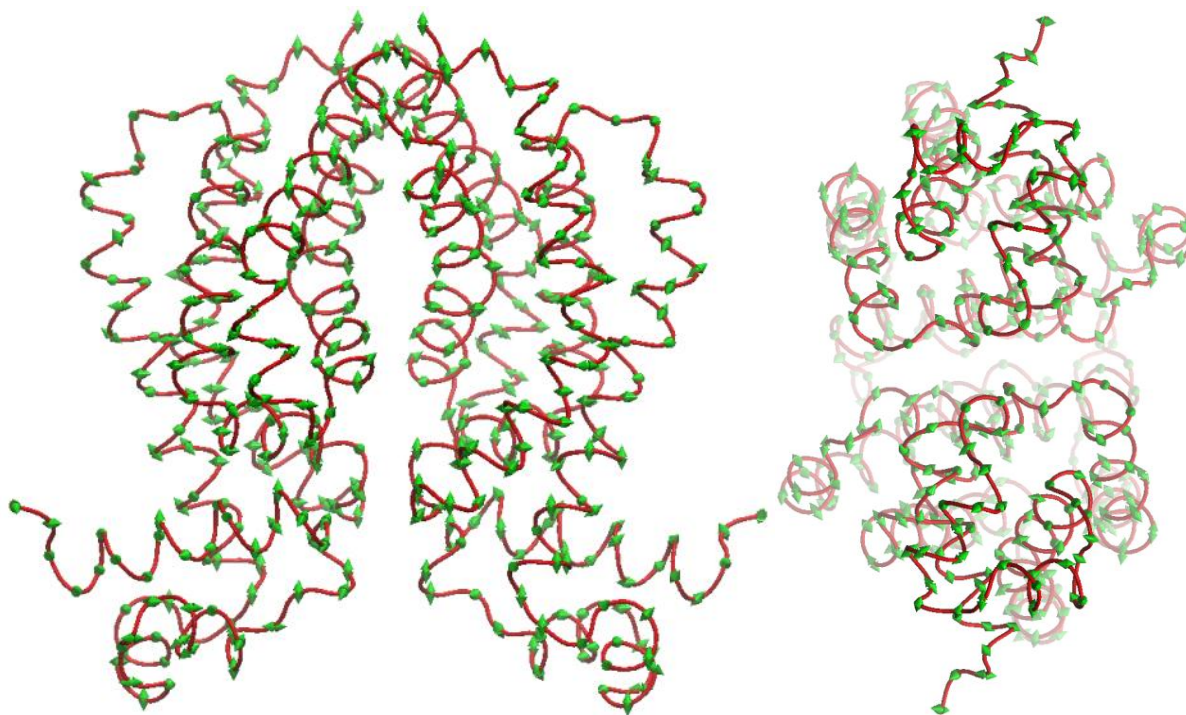


Figura 19 - Os modos normais restantes não tem deslocamentos maiores da cadeia principal
Fonte: Elaborada pelo autor.

5.3 Comparação entre os modos normais, a dinâmica molecular e as estruturas cristalográficas

A seguir, serão relacionados os movimentos encontrados nos modos normais com os deslocamentos apresentados nas dinâmicas moleculares. Cada movimento reproduz dois movimentos importantes na função do regulador alostérico: o movimento dos domínios HTH

(que permitem sua ligação ao DNA) e o movimento do sítio alostérico (que permite a transição entre o estado ativo e inativo e a estabilização do estado ativo gerada pela ligação de um determinado ligante).

5.3.1 Modo normal 9 e o movimento do sítio alostérico

Um dos movimentos mais importantes para uma proteína alostérica é o que permite seu deslocamento entre os estados inativo e ativo. Para isso é preciso que uma molécula se ligue ao sítio alostérico, o que dispara a mudança conformacional e a mantém. No modo normal 9 se observa um maior deslocamento nesta região comparado aos outros modos, pelo qual se considera que este modo é o que mais influencia o movimento descrito. Na Figura 20 se representa a estrutura do modo vibracional 9 contendo o ligante na posição de ligação alostérica.

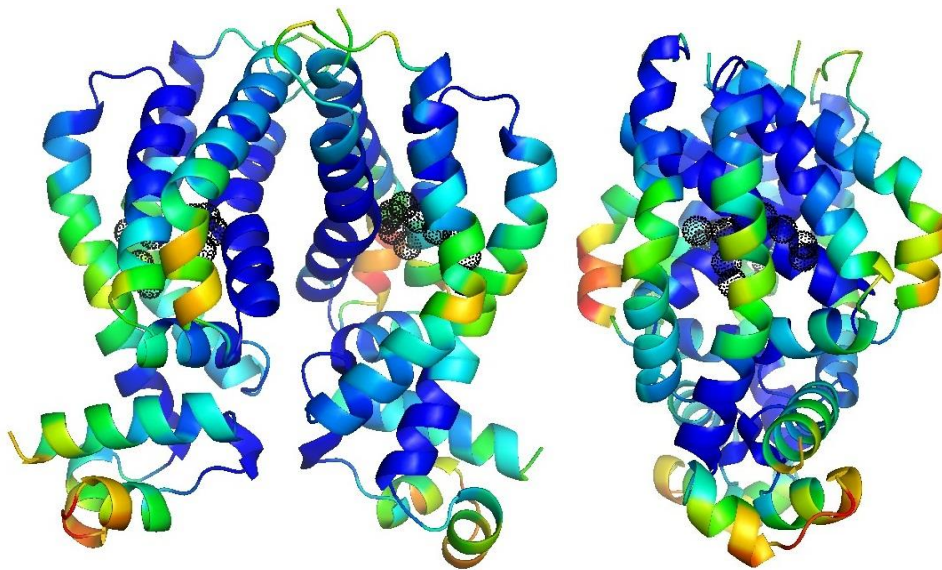


Figura 20 - Modo normal 9 com o ligante na posição de ligação alostérica

Fonte: Elaborada pelo autor

Além deste modo outros modos também mostram movimentos no domínio alostérico como o modo dez e o modo onze, mas com uma menor amplitude e não tendo uma separação maior entre as hélices do sítio alostérico. Mais na frente se analisa isso.

5.3.2 Modo normal 10 e o deslocamento dos domínios HTH

O fator que determina se a proteína TM1030 está no estado ativo ou inativo é a distância entre os domínios HTH, determinada pela configuração que o sítio alostérico apresenta quando estiver ou não ligado. O modo normal 10 tem uma grande influência neste movimento, como pode ser observado na Figura 21, que compara o movimento da proteína com os deslocamentos apresentados tanto nas estruturas cristalográficas ativa (1Z77) como na inativa (1ZKG).

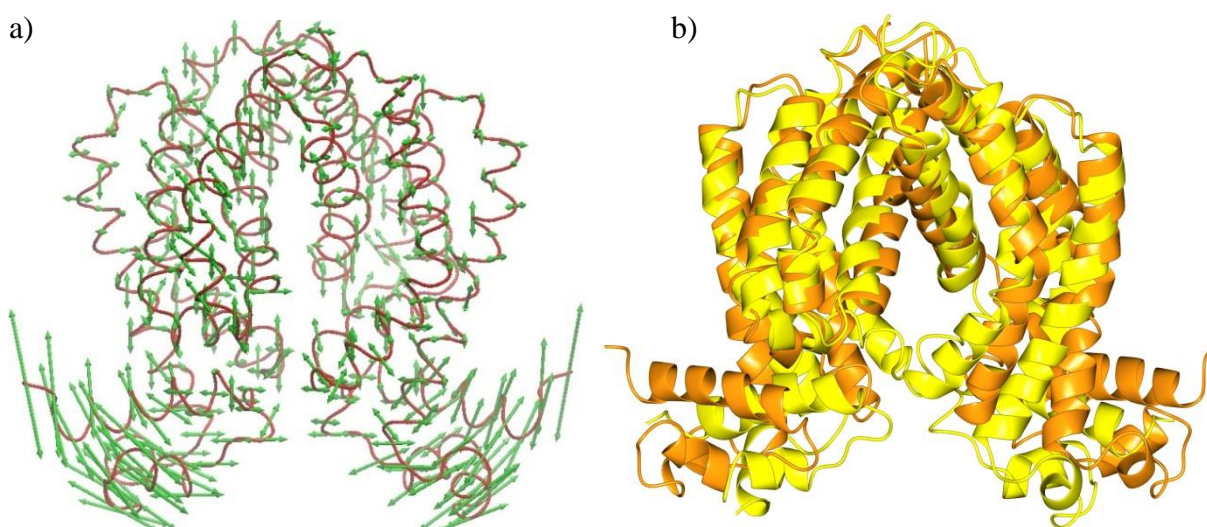


Figura 21 - No desenho a) se observa os deslocamentos do Modo normal 10, no b) a proteína TM1030 no estado ativo (amarelo) e no estado inativo (laranja).

Fonte: Elaborada pelo autor.

Mas ainda assim este modo só mostra a abertura dos domínios HTH mas os deslocamentos de outras hélices não são representados como é o caso das hélices H4 e H7, os modos normais 11 e 12 mostrariam deslocamento desta região junto a movimentos do domínio HTH, isso leva a pensar que para representar a transição alostérica a partir dos modos normais é preciso realizar uma combinação linear destes seis modos normais, já que por exemplo, os modos 7 e 8 têm movimentos do domínio HTH que podem contribuir para um melhor encaixe nos sulcos do DNA. Além do modo normal 9 que como visto anteriormente mostra o movimento das hélices no sítio alostérico. Propõe-se fazer então uma aproximação de mínimos quadrados para encontrar os pesos de cada modo normal dentro desta transição.

5.3.3 A soma dos modos normais se aproxima aos movimentos observados nas dinâmicas moleculares

Ao somar os seis modos normais representativos, que tem maior movimento da cadeia principal, usando uma amplitude igual para todos, é possível observar seu comportamento simulado para a proteína. Usamos a mesma amplitude já que não foram feitos cálculos da influência de cada modo no movimento, isso poderia ser um trabalho futuro. Comparando a soma dos modos normais com o movimento apresentado na dinâmica molecular realizada a 353K, se observa que, em geral, os movimentos simulados pelos dois métodos apresentam comportamentos similares. Ainda que algumas das direções dos movimentos não coincidam exatamente, isso poderia ser corrigido mudando as amplitudes de cada modo. Além disso, a interação com a água na dinâmica molecular pode gerar mudanças nos modos normais. Na Figura 22, que apresenta diferentes vistas da simulação, é possível observar as similaridades e diferenças entre os movimentos obtidos com a somatória dos modos normais e a estrutura média da dinâmica a 353K.

Olhando o domínio HTH, Figura 22, se pode ver que as direções dos movimentos apresentam uma similaridade. Mesmo que as amplitudes de todos os deslocamentos atômicos das trajetórias dos modos normais sejam iguais, é possível notar que algumas regiões se deslocam mais que outras, apresentando alguma concordância entre os dois métodos.

Nesta vista se observa que os átomos das regiões mais externas da proteína têm um deslocamento maior na dinâmica molecular que no análises de modos normais. No domínio HTH se vê que os movimentos de todos os átomos não têm direções exatamente iguais entre os dois métodos, mas algumas de estas direções poderiam ficar mais perto se modificarem as amplitudes de cada modo normal para fazer coincidir com os deslocamentos da dinâmica molecular.

Nesta vista do domínio HTH se pode ver que a proporção do deslocamento de alguns átomos guarda relação entre os dois métodos. Isto faria pensar que poderia existir uma relação numérica entre os modos normais e os movimentos da dinâmica.

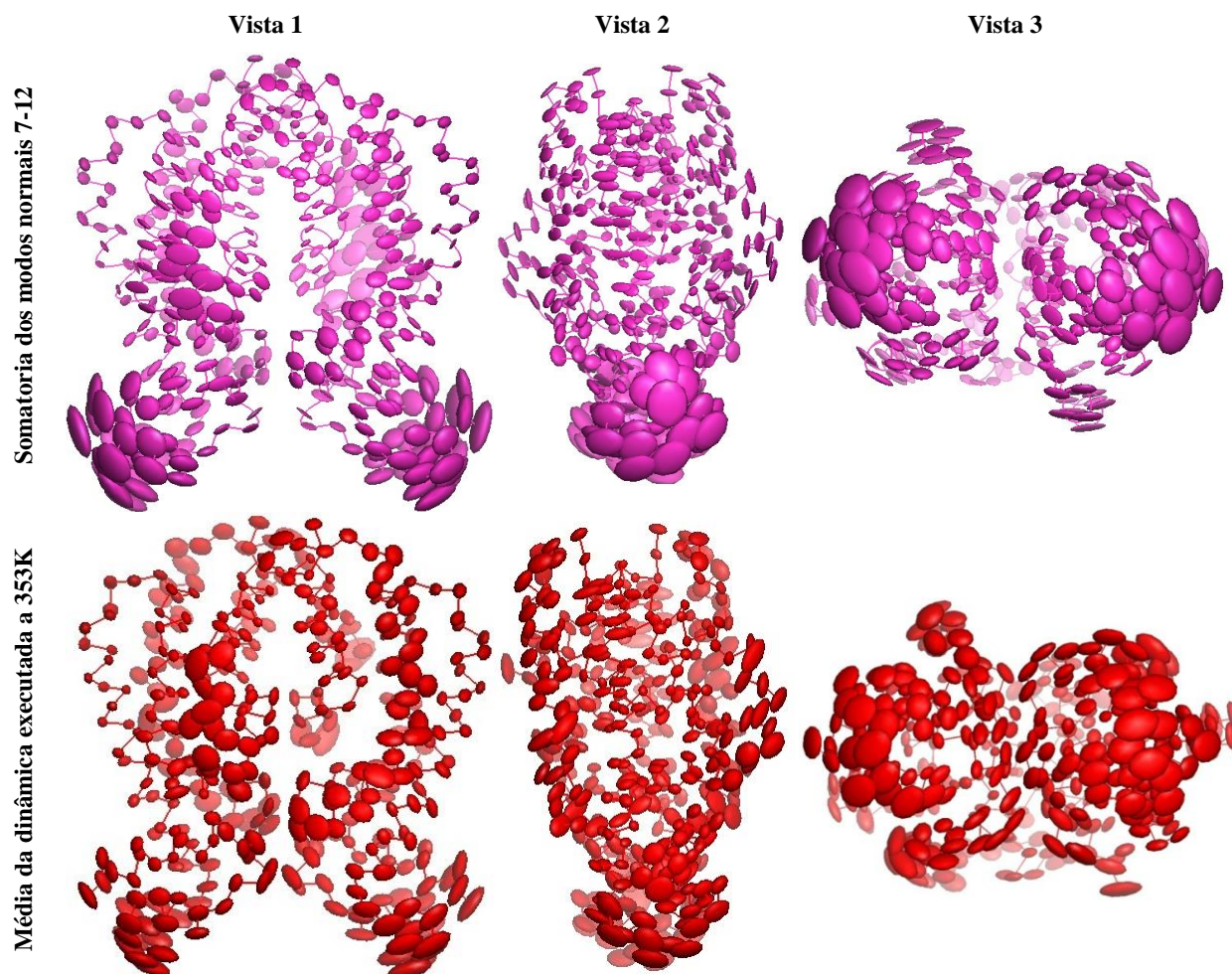


Figura 22 - Comparação desde diferentes posições da soma dos modos normais 7-12 (roxo) com a média da dinâmica a 353K (vermelho).

Fonte: Elaborada pelo autor.

No caso da dinâmica molecular realizada a 323K se observa uma relação com o movimento simulado pela soma dos modos normais menos o modo 10, Figura 23. Nestes se pode observar que as direções dos movimentos são muito similares, ainda que as amplitudes não sejam exatamente a mesmas, isso poderia se corrigir se for feita uma regressão de mínimos quadrados e for obtido o fator multiplicador de cada modo que melhoraria a relação entre os movimentos da dinâmica e os simulados na soma dos modos normais.

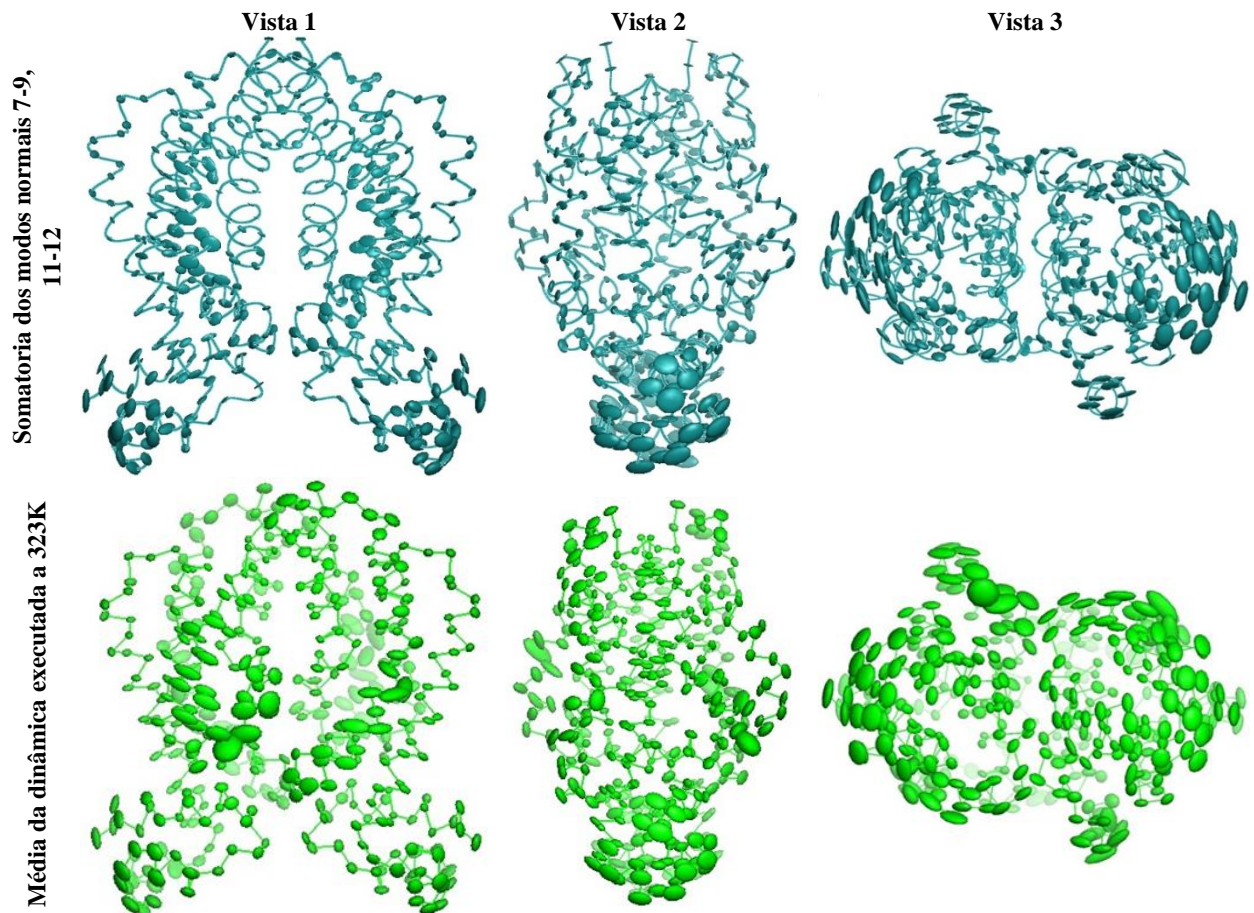


Figura 23 - Comparação desde diferentes posições da soma dos modos normais 7-9, 11 e 12 (azul) com a média da dinâmica a 323K (verde).

Fonte: Elaborada pelo autor

Das figuras 22 e 23 se pode observar que o modo normal 10 está mais representado na dinâmica de maior temperatura. Agora tomando em conta que este modo normal representa o afastamento e aproximação dos domínios HTH e que este movimento por sua vez ativa ou inativa a proteína, se pode concluir que este movimento tem uma maior probabilidade de ser realizado quando a proteína se encontra numa temperatura dentro da ótima para a sobrevivência da bactéria.

5.4 Análises dos fatores de temperatura isotrópicos

Foi observado por Koclega et. al. que ao cristalizar a TM1030 em três temperaturas 4°C, 37°C e 50°C os fatores de temperatura B diminuam nos cristais com o aumento da temperatura de cristalização (1), o que seria contrario ao esperado fisicamente. Com o intuito de revisar esse fenômeno foi proposto fazer dinâmicas moleculares a três temperaturas diferentes como foi

explicado na metodologia. Destas dinâmicas foram calculados os fatores de temperatura tanto em toda a trajetória como por faixas de 1ns de tempo. Os resultados destas medidas são apresentados a continuação.

5.4.1 Fatores de temperatura isotrópicos nas trajetórias

Na Figura 24 se pode ver que contrario ao comportamento mostrado na cristalografia os fatores de temperatura aumentam com o incremento da temperatura.

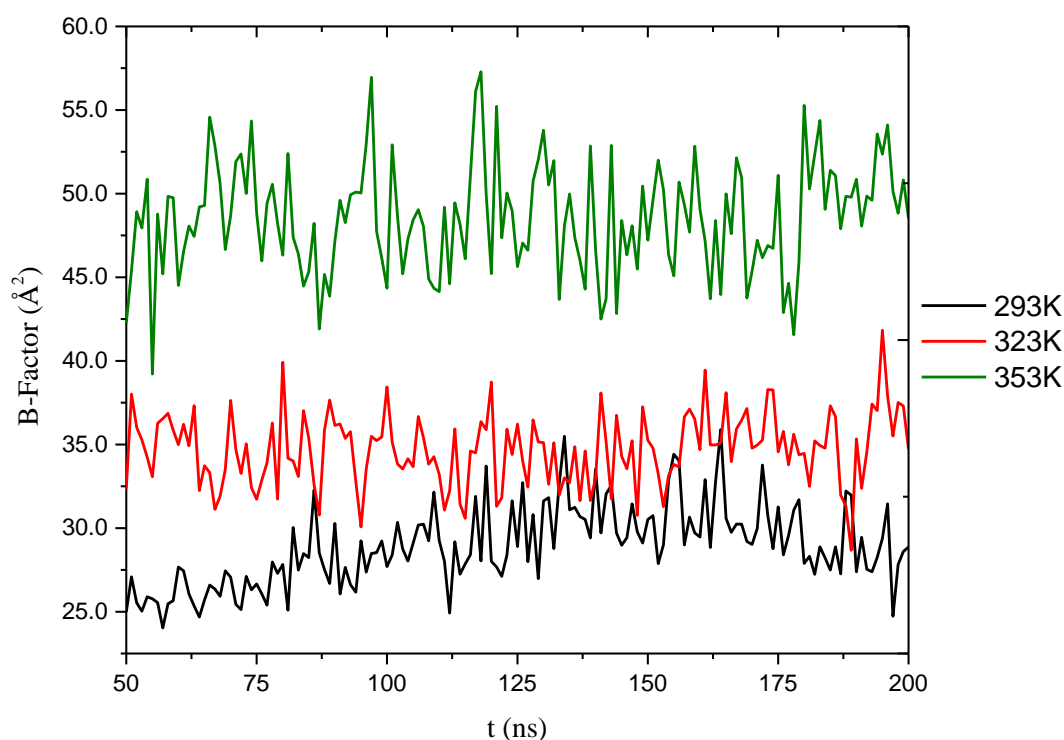


Figura 24 - Fatores de temperatura cada 1ns das dinâmicas da proteína TM1030 (PDB 1Z77) nas três temperaturas.

Fonte: Elaborada pelo autor

Pode-se observar que claramente os fatores de temperatura isotrópicos são maiores nas maiores temperaturas. Note-se também que na menor das temperaturas os valores da média mudam durante a dinâmica o que se reflete no calculo dos fatores de temperatura média para faixas de tempo de 50, 100 e 150ns que se mostram na Tabela 2.

Tabela 2 - Fatores de temperatura B (\AA^2) em diferentes faixas de tempo das diferentes dinâmicas moleculares executadas.

Faixa de tempo da dinâmica molecular (ns)	Fatores de temperatura B (\AA^2)		
	293K	323K	353K
50-200 (150)	60,8	68,8	118,0
100-200 (100)	58,2	64,1	106,2
150-200 (50)	55,8	64,2	93,6
Média das dinâmicas a cada 1ns	28,9	34,7	48,5
Cristalográfico	39,2	33,2	30,3

Fonte: Elaborada pelo autor.

Comparando os valores obtidos na Figura 24 e os da Tabela 2 se pode ver que a proteína tem uma mudança conformacional importante durante a trajetória de dinâmica molecular ao serem os fatores de temperatura maiores em faixas de tempo maiores. Observa-se também que esta diminuição tem um maior peso na dinâmica a 353K onde esta mudança é de 26% comparada com a diminuição de só 9% para a dinâmica a 293K ao comparar os fatores de temperatura dos últimos 50ns com os de os últimos 150ns. Ou seja, as vibrações de baixa frequência estão mais representadas nas dinâmicas a altas temperaturas. Estes términos de baixa frequência se refletem em maiores oscilações coletivas dos diferentes domínios da proteína. Uma prova possível para determinar a incidência destas frequências nas trajetórias seria passá-las por um filtro de passa-baixos.

5.4.2 Porque não diminuem os fatores de temperatura?

Uma primeira explicação de porque não diminuem os fatores de temperatura com o aumento da temperatura nas dinâmicas moleculares poderia ser atribuído ao fato de um possível problema metodológico no refinamento das estruturas cristalográficas. Isto porque ainda que os três fatores de temperatura são diferentes as estruturas são idênticas, se esperaria que ao mudar o fator de temperatura existira um rearme da molécula. Não sendo esse o caso uma segunda explicação seria que os contatos cristalográficos diminuem as baixas frequências no cristal ao amortecer as oscilações coletivas. Como se viu o comportamento dos RMSD médios, calculados nas trajetórias e no tempo, mostra que a estrutura da proteína vai se tornando mais próxima da estrutura cristalográfica com o aumento da temperatura. Além disso, os fatores de temperatura calculados mostram que estas baixas frequências incidem muito nos movimentos a maiores temperaturas pelo que é provável que se tenha uma diminuição maior nestes cristais. Uma forma de verificar isto seria usar o programa CHARMM e realizar uma simulação de dinâmica molecular da proteína dentro do cristal.

6 CONCLUSÕES

Nas três dinâmicas moleculares depois de passado um tempo de aproximadamente 50ns o sistema se estabiliza e a estrutura se desloca para uma conformação que tem um menor RMSD com respeito da estrutura cristalográfica, o que seria uma conformação intermedia que dependendo de fatores externos (temperatura, ligante alostérico, etc.) promoveria a proteína para o estado ativo ou inativo. A estrutura média destas trajetórias mostra que a maiores temperaturas as hélices alfa que conformam a estrutura da proteína se encaixam em posições mais próximas da estrutura cristalográfica PDB-1Z77 que seria o estado ativo de acordo com os dados experimentais. Ao medir as distancias entre as hélices H3 dos domínios HTH se pode observar que nas dinâmicas a 293 e 323 K esta distancia se mantem no tempo, mas na dinâmica a 353 K esta distancia diminui para valores próximos à distancia que tem essas hélices na estrutura cristalográfica ativa.

Para os parâmetros de deslocamento atômicos, representados nos fatores de temperatura anisotrópicos, se observa que à maior temperatura os átomos da cadeia principal se movimentam mais, o que era fisicamente esperado. Também se pode ver que as direções dos movimentos do domínio HTH são ligeiramente diferentes nas menores temperaturas em relação às maiores temperaturas, o que ao comparar com a soma de modos normais permite inferir que nestas temperaturas não se atinge alguns destes modos já que eliminando modos como o 10 os movimentos deste domínio ficam mais parecidos com os da dinâmica. Como perspectiva se poderia calcular os coeficientes necessários para encaixar os movimentos dos modos normais dentro dos que realiza a proteína na dinâmica molecular e comparar o comportamento da proteína nas duas técnicas.

O grupo de modos normais que representam os movimentos coletivos rígidos da molécula, movimento da cadeia principal, esta compostos pelos modos 7 até 12, tomando em conta que os primeiros seis modos são eliminados por serem as rotações e translações da molécula no espaço. Destes modos que representam os movimentos da cadeia principal se tem dois que são uma boa representação da transferência de um estado para outro da transição alostérica, o modo 9 que mostra a abertura do sitio de ligação alostérico e o modo 10 que mostra a abertura dos dominós HTH do dímero que é principal diferencia entre esses dois estados e condição para ligar ou não o DNA. O normal 10 estaria mais representado nas dinâmicas moleculares rodadas em maiores temperaturas. Mas nenhum dos modos representa exatamente a transição

alostérica, por tanto este movimento deveria ser uma combinação linear de algum ou todos os seis modos normais que apresentam deslocamentos grandes da cadeia principal.

No caso dos fatores de temperatura como era de se esperar fisicamente, eles aumentam com o aumento da temperatura. Não em tanto o resultado experimental que é a diminuição deste valor com o aumento da temperatura se poderia explicar como uma consequência do empacotamento da proteína no cristal que amortece os movimentos de baixa frequência.

7 PERSPECTIVAS

Um primeiro dado importante a calcular é peso que cada um dos modos tem no movimento da transição alostérica. Para isso seria preciso fazer uma minimização por mínimos quadrados e aproximar cada modo normal a través de uma única amplitude constante.

Como não se conseguiu reproduzir o que aconteceu experimentalmente com os fatores de temperatura se propõe fazer uma dinâmica reproduzindo as condições cristalográficas usando o programa CHARMM. Com isto se poderia aclarar se o que acontece experimentalmente é realmente uma propriedade do cristal ou um problema dos dados cristalográficos.

Uma prova interessante seria realizar simulações do mesmo tipo das feitas nesta dissertação numa proteína hortóloga de um organismo mesófilo para observar que comportamentos são iguais e quais são diferentes.

Fazer uma análise de componentes principais (PCA) poderia ajudar também a comparar frequências e direções dos movimentos simulados pelos modos normais com os movimentos realizados durante as dinâmicas moleculares na proteína nas três temperaturas.

REFERÊNCIAS

- 1 KOCLEGA, K. D.; CHRUSZCZ, M.; ZIMMERMAN, M. D.; BUJACZ, G.; MINOR, W. “Hot” macromolecular crystals. **Crystal Growth & Design**, v. 10, n. 2, p. 580–586, 2010.
- 2 LEHMANN, M.; PASAMONTES L.; LASSEN, S.; WYSS, M. The consensus concept for thermostability engineering of proteins. **Biochimica et Biophysica Acta (BBA)** - protein structure and molecular enzymology, v. 1543, n. 2, p. 408-415, 2000.
- 3 JOHNSEN, U.; HANSEN, T.; SCHÖNHEIT, P. Comparative Analysis of Pyruvate Kinases from the Hyperthermophilic Archaea *Archaeoglobus fulgidus*, *Aeropyrum pernix*, and *Pyrobaculum aerophilum* and the Hyperthermophilic Bacterium *Thermotoga maritima* UNUSUAL REGULATORY PROPERTIES IN HYPERTHERMOPHILIC ARCHAEA. **Journal of Biological Chemistry**, v. 278, n 28, p. 25417-25427. 2003.
- 4 ADCOCK, S. A.; MCCAMMON, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. **Chemical Reviews**, v. 106, n. 5, p. 1589-1615, 2006.
- 5 TASUMI, M.; TAKEUCHI, H.; ATAKA, S.; DWIVEDI, A. M.; KRIMM S. Normal vibrations of proteins: glucagon. **Biopolymers**, v. 21, n. 3, p. 711-714, 1982.
- 6 NOGUTI, T.; GŌ, N. Collective variable description of small-amplitude conformational fluctuations in a globular protein. **Nature**, v. 296, n. 5859, p. 776-778, 1982.
- 7 FIELDS, P. A. Review: protein function at thermal extremes: balancing stability and flexibility. **Comparative Biochemistry and Physiology Part A: molecular & integrative physiology**, v. 129, n. 2-3 p. 417-431, 2001.
- 8 KRISTJÁNSSON, J. K.; HREGGVIDSSON, G. O. Ecology and habitats of extremophiles. **World Journal of Microbiology and Biotechnology**, v. 11, n. 1, p. 17-25, 1995.
- 9 VIEILLE, C.; ZEIKUS, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. **Microbiology and Molecular Biology Reviews**, v. 65, n. 1, p. 1-43, 2001.
- 10 ROBER, H.; THOMAS, A. L.; HELMUT, K. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up 90°C. **Archives of Microbiology**, v. 144, n. 4, p 324-333, 1986.
- 11 MATHEWS, C. K.; VAN HOLDE, K. E.; AHERN, K. G. **Bioquímica**. [s. l.]: Addison-Wesley, 2002.
- 12 KURIYAN, J.; KONFORTI, B.; WEMMER, D. **The molecules of life**. New York: Garland Science, 2013.
- 13 MORENO YANES, J. A. Aminoácidos, péptidos y proteínas: teoría y problemas. In: **Conceptos Fundamentales en Bioquímica**. Caracas: Espasan de Editores, 1986.

- 14 HARRISON, S. C.; AGGARWAL, A. K. DNA recognition by proteins with the helix-turn-helix motif. **Annual Review of Biochemistry**, v. 59, p. 933-969, 1990. doi: 10.1146/annurev.bi.59.070190.004441.
- 15 HILLEN, W.; KLOCK, G.; KAFFENBERGER, I.; WRAY JR, L. V.; REZNIKOFF, W. S. Purification of the TET repressor and TET operator from the transposon Tn10 and characterization of their interaction. **Journal of Biological Chemistry**, v. 257, n. 11, p. 6605-6613, 1982.
- 16 PREMKUMAR, L.; et al. Crystal structure of TM1030 from *Thermotoga maritima* at 2.3 Å resolution reveals molecular details of its transcription repressor function. **Proteins: structure, function, and bioinformatics**, v. 68, n. 1, p. 418-424, 2007.
- 17 RAMOS, J.; et al. The TetR family of transcriptional repressors. **Microbiology and Molecular Biology Reviews**, v. 69, n. 2, p. 326-356, 2005.
- 18 HO, B. K.; PERAHIA, D.; BUCKLE, A. M. Hybrid approaches to molecular simulation. **Current Opinion in Structural Biology**, v. 22, n. 3, p 386-393, 2012.
- 19 KLEPEIS, J L., et al. Long-timescale molecular dynamics simulations of protein structure and function. **Current opinion in structural biology**. v. 19, n 2, p 120-127, 2009.
- 20 ALLEN, M. P. Introduction to molecular dynamics simulation. In: ATTIG, N.; BINDER, K.; GRUBMÜLLER, H.; KREMER, K. (Ed.). **Computational Soft Matter: from synthetic polymers to proteins**. [s.l.]: NIC, 2004. p. 1-27. (NIC Series 23)
- 21 ARMOUGOM, F.; MORETTI, S.; KEDUAS, V.; NOTREDAME, C. The iRMSD: a local measure of sequence alignment accuracy using structural information. **Bioinformatics**, v. 22, n. 14, p e35-e39, 2006.
- 22 DAMM, K. L.; CARLSON, H. A. Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. **Biophysical Journal**, v. 90, n. 12, p. 4558-4573, 2006.
- 23 SALESA, B. C.; CHAKOUMAKOSA, B. C.; MANDRUSA, D.; SHARP, J. W. Atomic displacement parameters and the lattice thermal conductivity of clathrate-like thermoelectric compounds. **Journal of Solid State Chemistry**, v. 146, n. 2, p. 528-532, 1999.
- 24 RUPP, B. **Biomolecular crystallography**. New York: Garland Science, 2010.
- 25 LEVITT, M.; SANDER, C.; STERN, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. **Journal of Molecular Biology**, v. 181, n. 3, p. 423-447, 1985.
- 26 BROOKS, B.; KARPLUS, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. **Proceedings of the National Academy of Sciences**, v.80, n. 21, p. 6571-6575, 1983.
- 27 BROOKS, B. R.; JANEŽIČ, D.; KARPLUS, M. Harmonic analysis of large systems. I. methodology. **Journal of Computational Chemistry**, v. 16, n. 12, p. 1522-1542, 1995.

28 HINSEN, K.; THOMAS, A.; FIELD, M. J. Analysis of domain motions in large proteins. **Proteins: Structure, Function, and Bioinformatics**, v. 34, n. 3, p. 369-382, 1999.

29 CUI, Q.; BAHAR, I. (Ed.). **Normal mode analysis: theory and applications to biological and chemical systems**. New York: CRC Press, 2005

30 HAYWARD, S. et al. Normal modes and essential dynamics. In: KUKOL, A. (Ed.). **Molecular modeling of proteins**. New York: Humana Press. 2008. p. 89-106. (Methods in molecular biology, 443).

31 LEMKUL, J. Lysozyme in water. 2008-2014. Disponível em: <<http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/>>. Acesso em: 22 set. 2015.

ANEXOS

ANEXO A - Relatório Cristalização da proteína ElrR

MATERIAS E MÉTODOS

Expressão e Purificação

Plasmídeos contendo o gene codificado da ElrR foram transformados em células de *E. coli* BL21 DE3, em placas com LB sólido e usando 150 µg de Kanamicina. Colônias individuais foram inoculadas em 5 mL de LB com Kanamicina e glicerinato para estocar a 190 K. Para o crescimento de células foi realizado previamente um preinoculo de 5 mL por 12 horas, depois este foi transferido a 500 mL de LB e 2XYT, incubando a 150 RPM e 310K até atingir uma DO de 1.0, nesse momento foi adicionado 1 mM de IPTG para expressar por 3 y 4 h, a 293 e 303 K, e 150 RPM. As células foram coletadas usando uma centrifuga a 4690 g e 277 K, sendo resuspensas num tampão de Tris 80 mM, NaCl 250 mM e pH 7.8. A lise celular foi feita usando o Fisher Scientific Sonic dismembrator com pulso por 30s e descanso de 30s durante 6min, e depois centrifugado de novo a 13700 g e 277 K por 30 min. A fração solúvel foi purificada usando uma coluna de afinidade de níquel (NiNTA-Superflow QIAGEN), esta foi lavada primeiro usando o mesmo tampão adicionando 30 mM de imidazol, removendo as proteínas de baixa afinidade. Por último a ElrR foi totalmente purificada uma coluna de exclusão por tamanho (Sephadex 200 16/60 Pregrade) ao atingir os 270 mM de imidazol como se mostra na Figura A1. Usando esta informação e a cromatografia em gel SDS-PAGE 12 % foi calculada a massa correspondente ao homodímero em solução.

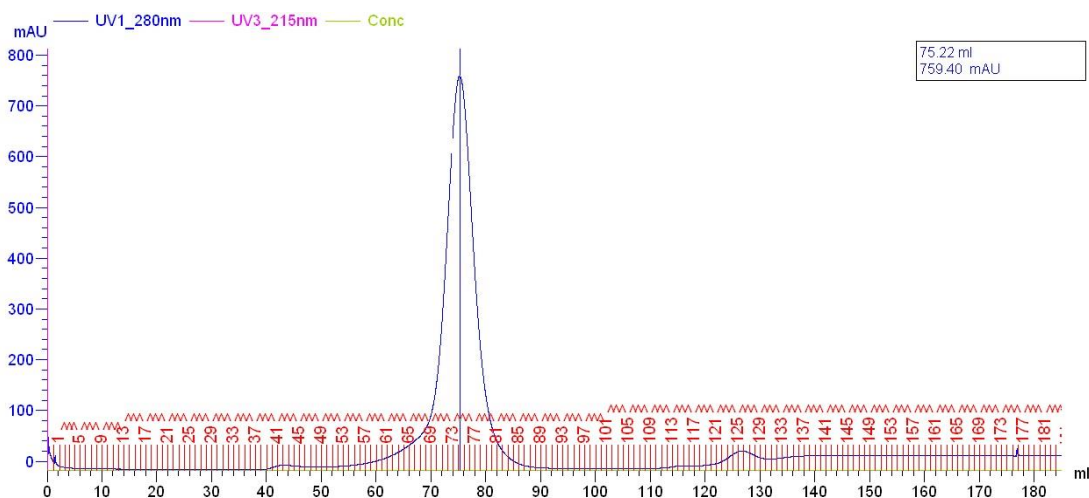


Figura A1 - Absorbância a 280nm na filtração de exclusão por tamanho.

Fonte: Autor

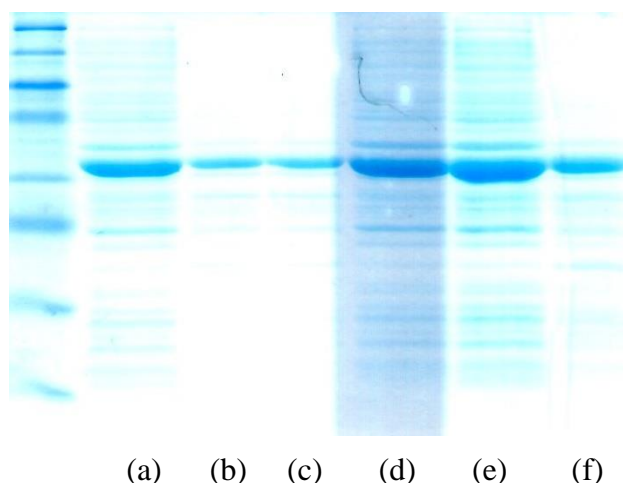


Figura A2 - Gel SDS da expressão e lise a da ElrR, na coluna (a) meio 2XYT 3h solúvel, (b) 2XYT 3h insolúvel, (c) LB 3h insolúvel, (d) LB 3h solúvel, (e) LB 4h solúvel e (f) LB 4h insolúvel.

Fonte: Autor

Cristalização

A proteína foi concentrada a 10, 20 and 30 mg/mL num concentrador Amicon de 30 Kda (Millipore), os tampão da dialise foi Tris 80 mM, NaCl 30 mM e pH 7.8,. Ensaio de cristalização foram feitos em placas de 96 poços, com a técnica de difusão de vapor. Usando o robô “Cartesian dispensing system” (Genomic solutions). Dois kits de cristalização comerciais foram usados: MORPHEUS (Molecular Dimensions) e SaltRx (Hampton Research). As placas foram deixadas numa temperatura de 291 K. Foi realizado uma variação do Tris para 30mM e NaCl 50mM mas não foram observadas mudanças nos ensaios.

RESULTADOS E DISCUSSÃO

A ElrR foi expressa em LB e 2XYT, sendo o LB onde se atingiu um maior rendimento na temperatura de 310 K e para um tempo de 3 h de expressão como se mostra na Figura 2.

Cristais começaram crescer depois de 24 h no kit Morpheus, mas depois de um tempo maior crescerem em ambos kits. Observa-se dois tipos diferentes de geometria nos cristais depois de uma semana de crescimento, sendo estes em forma de placas e de octaedros (Figura 3). Estas duas geometrias estão relacionadas com a concentração de proteína, se observa uma maior frequência de placas na concentração 20 mg/mL e maior frequência de octaedros na concentração 30mg/mL. Ainda assim as duas geometrias podem acontecer em todas as

concentrações pelo qual nós achamos que se deve provavelmente a que a proteína está nas duas conformações alostéricas nestes cristais. O imidazol também tem sido importante para a cristalização observando-se formação de cristais na presença dele.

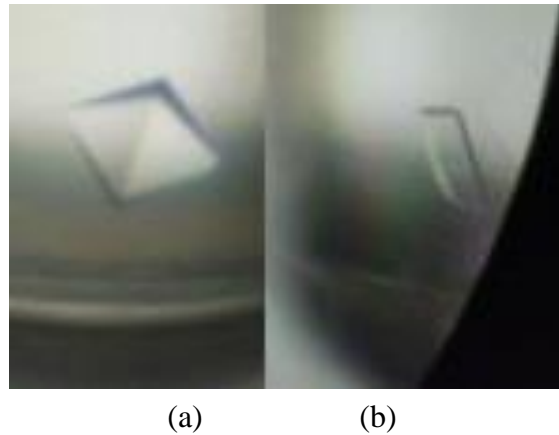


Figura A3 - (a) Cristal do tipo octaédrico. (b) Cristal do Tipo Placa.

Fonte: Autor

Estes dois tipos de cristais foram difratados com raios-X no sincrotron de Diamond Light Source (UK), e se obtiverem dois datasets diferentes para cada tipo de cristal. Para los cristales octaédricos se obtuvo una resolución de 3.25 Å com grupo espacial $P4_32_12$ numa cela de $a=b=108.60$, $c=142.30$, $\alpha=\beta=109.10$, $\gamma=142.40$. No caso dos cristais de placa os cristais difrataram com uma resolução de 2.5 Å com grupo espacial P1 numa cela de $a=70.00$, $b=76.14$, $c=86.06$, $\alpha=112.44$, $\beta=89.54$, $\gamma=103.40$. Na Figura 4A se mostra um exemplo das difrações obtidas para um dos cristais.

A difração dos cristais em P1 foi coletada em 720° numa tentativa de incrementar a multiplicidade e obter um melhor dataset. Também os cristais $P4_32_12$ foram difratados num comprimento de onda de 1.477 nm, para usar a diferencia anômala de grupo Sulfuro e tentar obter as fases usando o método de difração anômala simples (single anomalous diffraction (SAD)).

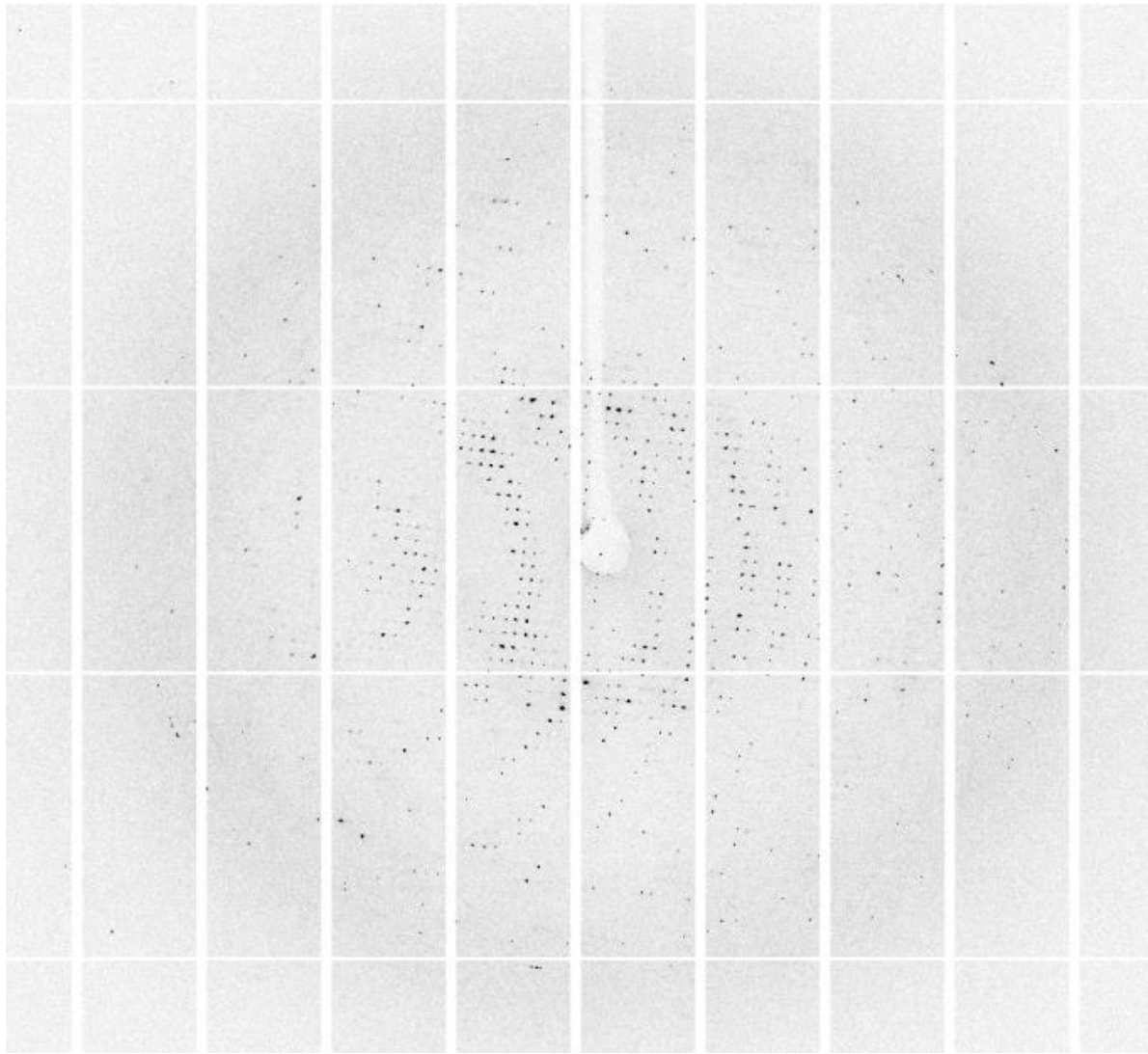


Figura A4 - Exemplo das difrações obtidas para um cristal octaédrico com grupo espacial $P4_32_12$

Fonte: Autor

Num primeiro momento se tentou resolver o problema das fases tentando encontrar um homólogo com o qual comparar, mas não foi possível achar uma proteína com homologia maior de que 12 %. Então foram tentados outro tipo de técnicas aproveitando a forma na que foram tomadas os dataset, se usou XDS e Shelx, mas pela pouca intensidade dos pontos e problemas dos eixos de difração não foi possível achar uma estrutura com a qual começar.

CONCLUSÃO Y PERPECTIVAS.

Dos resultados se pode interpretar inicialmente que temos cristais nas duas conformações possíveis da proteína alostérica. Pela falta de estruturas homologas direitas se necessita tentar resolver o problema das fases com outros métodos, entre os quais temos o de átomos pesados e o crescimento com selênio metioninas.

NOTA DE ESCLARECIMIENTO:

Este trabalho foi começado como parte de meu projeto de mestrado junto com as simulações computacionais. Chegado num ponto que se precisaria dedicar muito tempo nos dois trabalhos o professor Horjales me recomendou continuar só com um deles. Foi escolhido seguir o trabalho de simulações computacionais por ser mais interessante para mim.

ANEXO B - Inputs rodados no pacote GROMACS

Alguns dos programas do pacote GROMACS precisam de um input para poder ser rodados, neste anexo vamos mostrar os usados para realizar os cálculos desta dissertação.

```
-----1z77_minw.mdp-----  
define = -DFLEXIBLE  
constraints = none  
integrator = cg  
nsteps = 5000  
nstlist = 10  
rlist = 1.0  
coulombtype = pme  
rcoulomb = 1.0  
vdw-type = cut-off  
rvdw = 1.0  
nstenergy = 250  
emtol = 0.01  
emstep = 0.01  
-----
```

```
-----1z77_ions.mdp-----  
emtol = 1000.0  
emstep = 0.01  
nsteps = 50000  
  
nstlist = 1  
cutoff-scheme = Verlet  
ns_type = grid  
coulombtype = PME  
rcoulomb = 1.0  
rvdw = 1.0  
pbc = xyz  
-----
```

```
-----1z77_eqr.mdp-----
define = -DPOSRES
constraints = all-bonds
integrator = md
dt = 0.002
nsteps = 100000 ; total 200.0 ps
nstcomm = 10
nstxout = 250 ; collect data every 1.0 ps
nstxtcout = 250
nstvout = 250
nstfout = 0
nstlog = 100
nstenergy = 100
nstlist = 10
ns_type = grid
rlist = 1.0
coulombtype = PME
rcoulomb = 1.0
vdwtype = cut-off
rvdw = 1.4
pme_order = 4
ewald_rtol = 1e-5
optimize_fft = yes
DispCorr = no

; Berendsen temperature coupling is on
Tcoupl = v-rescale
tau_t = 0.1 0.1
tc-grps = protein non-protein
ref_t = 293 293

; Pressure coupling is on
Pcoupl = parrinello-rahman
Pcoupltype = isotropic
tau_p = 1.0
compressibility = 4.5e-5
ref_p = 1.0

; Generate velocities is on at 293 K.
gen_vel = yes
gen_temp = 293.0
gen_seed = -1
-----
```

```
-----1z77_eqf.mdp-----
define = -DFLEXIBLE
constraints = all-bonds
continuation = yes
integrator = md
dt = 0.002
nsteps = 100000 ; total 200.0 ps
nstcomm = 10
nstxout = 250 ; collect data every 1.0 ps
nstxtcout = 250
nstvout = 250
nstfout = 0
nstlog = 100
nstenergy = 100
nstlist = 10
ns_type = grid
rlist = 1.0
coulombtype = PME
rcoulomb = 1.0
vdwtype = cut-off
rvdw = 1.4
pme_order = 4
ewald_rtol = 1e-5
optimize_fft = yes
DispCorr = no

; Berendsen temperature coupling is on
Tcoupl = v-rescale
tau_t = 0.1 0.1
tc-grps = protein non-protein
ref_t = 293 293

; Pressure coupling is on
Pcoupl = parrinello-rahman
Pcoupltype = isotropic
tau_p = 1.0
compressibility = 4.5e-5
ref_p = 1.0

; Generate velocities.
gen_vel = no
-----
```

```
-----1z77_md01.mdp-----
define = -DFLEXIBLE
constraints = all-bonds
continuation = yes
integrator = md
tinit = 0
dt = 0.002
nsteps = 5000000 ; total 10.0 ns
nstxout = 250
nstxtcout = 250
nstvout = 250
nstfout = 0
nstlog = 100
nstenergy = 250
nstlist = 10
ns_type = grid
pbc = xyz
rlist = 1.0
coulombtype = PME
rcoulomb = 1.0
vdwtype = cut-off
rvdw = 1.4
pme_order = 4
ewald_rtol = 1e-5
optimize_fft = yes
DispCorr = no

; Berendsen temperature coupling is on
Tcoupl = v-rescale
tau_t = 0.1 0.1
tc-grps = protein non-protein
ref_t = 293 293

; Pressure coupling is on
Pcoupl = no
Pcoupltype = isotropic
tau_p = 1.0
compressibility = 4.5e-5
ref_p = 1.0

; Generate velocities.
gen_vel = no
-----
```

```
-----1z77_minw.mdp-----
define                = -DFLEXIBLE
constraints           = none
integrator            = cg ;steep ou l-bfgs
nsteps               = -1
nstlist              = 1.3
rlist                = 0
cutoff-scheme        = group
coulomb-modifier      = Potential-shift
coulombtype          = cut-off
pbc                  = no
rcoulomb             = 0
vdw-modifier         = Potential-shift
vdw-type             = cut-off
rvdw                 = 0
ns_type              = grid
nstenergy            = 100
emtol                = 4.16e-7
emstep               = 0.001
gen_seed             = -1

;          Implicit solvent
implicit_solvent     = GBSA
gb_algorithm         = OBC
nstgbradii          = 1
rgbradii            = 0
gb_epsilon_solvent  = 80
gb_saltconc         = 0
gb_obc_alpha        = 1
gb_obc_beta         = 0.8
gb_obc_gamma        = 4.85
gb_dielectric_offset = 0.009
sa_algorithm         = Ace-approximation
sa_surface_tension  = 2.25936
-----
```