UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

Amanda Souza Câmara

Movimentos coletivos harmônicos, suas frequências e combinações lineares, na regulação de três proteínas: na transição alostérica da DEA, na ativação por redução da MosR e na ligação da ElrR ao DNA

São Carlos

2017

Amanda Souza Câmara

Movimentos coletivos harmônicos, suas frequências e combinações lineares, na regulação de três proteínas: na transição alostérica da DEA, na ativação por redução da MosR e na ligação da ElrR ao DNA

Tese 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 Doutor em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular

Orientador: Prof. Dr. Eduardo Horjales Re-boreda

Versão corrigida
(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2017

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)

Câmara, Amanda Souza

Movimentos coletivos harmônicos, suas frequências e combinações lineares, na regulação de três proteínas: na transição alstérica da DEA, na ativação por redução da MosR e na ligação da ElrR ao DNA / Amanda Souza Câmara; orientador Eduardo Horjales - versão corrigida -- São Carlos, 2017.

173 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2017.

1. Mudanças conformacionais. 2. Movimentos coletivos. 3. Dinâmica molecular. 4. Filtro de modos harmônicos. 5. Combinação linear de modos. I. Horjales, Eduardo, orient. II. Título.

*Dedico este trabalho à minha filha,
Sofia.*

AGRADECIMENTOS

Agradeço aos meus pais, Tércio e Kátia, e à minha irmã, Katharina, pelo apoio incondicional e por sempre valorizarem meu esforço e trabalho. Um agradecimento a mais para meu pai que revisou meu texto.

Agradeço à minha filha, Sofia, por todo o carinho e por ter uma paciência desproporcional ao seu tamanho.

Agradeço ao professor Eduardo Horjales pela orientação.

Agradeço ao professor Otávio por me ajudar a lidar com minha delicada relação profissional com meu orientador.

Agradeço aos meus colegas de trabalho pelo suporte na rotina, pela hora do café, pelos conselhos e por tornar o ambiente mais familiar.

Agradeço ao meu amigo Tonho por seus ouvidos, pela compreensão e pelas discussões produtivas e existenciais.

Agradeço ao Michel por dividir comigo os fardos e angústias do grupo sendo um companheiro de batalha indispensável.

Agradeço aos meus amigos Valéria, Paty, Ana Eliza, Militar, Katarina, Larissa, Evandro e Laure por me ouvirem falar de trabalho mesmo nos finais de semana e me lembrarem de falar de outras coisas também.

Agradeço à equipe da biblioteca, em especial à Neusa, e do serviço de pós-graduação por me auxiliarem em todo o processo.

Agradeço à equipe da seção de informática, em especial ao Gilmar, por ajudar a montar e manter os computadores de alta performance e programas imprescindíveis ao meu trabalho.

Agradeço ao professor Alessandro Nascimento por me auxiliar na reta final de defesa do doutorado.

*“Nós nascemos para mudar
Às vezes consideramos isso como uma metáfora
que reflete a forma que as coisas deveriam ser.
De fato, a mudança leva tempo
Excede todas as expectativas
Requer o agora e o depois
Veja, apesar de os músicos mudarem,
A música permanece a mesma
E a verdade é que
É preciso ter coragem para mudar.”
Vinnie Jones*

RESUMO

Câmara, A. S. **Movimentos coletivos harmônicos, suas frequências e combinações lineares, na regulação de três proteínas:** na transição alostérica da DEA, na ativação por redução da MosR e na ligação da ElrR ao DNA. 2017. 173p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

Nas duas últimas décadas, houve um enorme aumento no número de estruturas proteicas resolvidas, e entre elas há uma variedade imensa de proteínas com mais de uma conformação observada. Essa quantidade incontestável de dados experimentais corroboram a hipótese de que cada proteína exista num espaço conformacional próprio, onde ela possa adotar inúmeras conformações, umas mais distintas ou estáveis que outras. Essas conformações estão distribuídas nesse espaço de acordo com sua energia potencial, que pode ser definida como uma superfície cheia de rugosidades, poços e barreiras energéticas. Duas conformações distantes nesse espaço são muito diferentes entre si, enquanto que duas conformações próximas são mais semelhantes. Da mesma forma, se distinguem os movimentos necessários para passar de uma conformação à outra. Para uma proteína passar de um estado a outro, geralmente identificados com grandes mudanças conformacionais, é necessário um movimento coletivo. Por ser de grande amplitude, esse tipo de movimento ocorre com baixa frequência, e dificilmente é observado em simulações clássicas de dinâmica molecular. Assim, existem métodos dedicados à obtenção destes movimentos, como a análise de modos normais, os modelos de redes elásticas e a análise de componentes principais. Neste trabalho, adaptamos o método de transformada de Fourier para recuperar modos harmônicos que compõem uma trajetória simulada suficientemente longa para analisar três proteínas distintas quanto a seus movimentos biológicos de importância funcional. Uma é a DEA, cuja simetria hexagonal observamos influenciar nos modos coletivos e na transição entre estados. Outra é a MosR, que simulamos em seus dois estados diferentes, oxidado ou reduzido, para encontrar como a oxidação é capaz de impedir os movimentos coletivos que levam à conformação ligada ao DNA. Nestas duas proteínas, observamos que nenhum modo por si só é responsável pela transição entre as conformações experimentais, mas que eles dependem de outros modos ou outras mudanças conformacionais ocorrendo de forma combinada. A terceira proteína analisada é um regulador transcricional, assim como a MosR, a ElrR, cuja estrutura é conhecida somente na forma apo. Neste trabalho, construímos modelos da ElrR ligada ao DNA pela combinação linear de modos harmônicos para modelar um possível ligante na nova conformação do sítio alostérico. As amplitudes usadas nessa combinação foram obtidas pelo método de mínimos quadrados, visando minimizar o desvio em relação somente às coordenadas que as hélices de reconhecimento devem apresentar para se ligar ao sítio de DNA. Este prognóstico foi feito pela análise metódica das estruturas de 27 reguladores transcricionais, homodiméricos com o motivo HTH, em

complexo com DNA. Essa análise também nos permitiu descrever a estereoquímica do encaixe das hélices de reconhecimento nos sulcos maiores do DNA com novos parâmetros geométricos, intimamente relacionados com a simetria do complexo, com a sequência de resíduos das hélices de reconhecimento e com a sequência de bases do sítio de DNA, de forma a auxiliar na modelagem de novos complexos.

Palavras-chave: Mudanças conformacionais. Movimentos coletivos. Dinâmica molecular. Filtro de modos harmônicos. Combinação linear de modos.

ABSTRACT

Câmara, A. S. **Collective harmonic motions, their frequencies and linear combinations, on the regulation of three proteins:** on the allosteric transition of DEA, on the activation by reduction of MosR and on the DNA-binding of ElrR. 2017. 173p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

There was an enormous increase in the deposited protein structures in the past two decades, among them there is a great variety of proteins with more than one observed conformation. This undeniable amount of experimental data ratify the hypothesis that each protein possesses its own conformational space, where it can adopt countless conformations, some more distinct or stable than others. These conformations are distributed in the space according to its potential energy, which maybe defined as a rough landscape filled with energetic wells and barriers. Two conformations lying apart from each other in this landscape do not carry much resemblances, while neighbouring conformations are very similar. The motions required to get one conformation to another are just as distinguishable. There must be a collective motion inbetween two states of a protein, commonly characterized by large conformation changes. This type of motion is related to large amplitudes and low frequencies, thus it is hardly seen in classical molecular dynamics simulations. Therefore, there are dedicated methods to obtain these motions, as normal modes analysis, elastic network models and essential dynamics. In this work we adapted the method of Fourier transform filtering to retrieve harmonic modes that compose a simulated trajectory and thus analyse the biological motions with functional importance of three distinct proteins. One is DEA, which hexagonal symmetry was observed to affect its collective motions and the transition between biological states. Another protein is MosR, which we simulated in two different states, oxidized or reduced, to learn how the formation of a disulphide bridge is able to preclude the collective motions that lead to a DNA-binding conformation. With these two proteins we observed that no mode by itself is responsible for the transition between experimental conformations, and they actually depend on other conformational changes occurring in a combined manner. The third protein that we analysed, ElrR, is a transcriptional regulator, like MosR, which structure is known only on its apo form. Hence in this work we built models of ElrR bound to DNA by the linear combination of harmonic modes aiming to model a ligand that would fit in the allosteric site upon the conformational changes driven by the collective motions. The amplitudes we used in this method were calculated by the least square method to minimize the deviation to the positions of the recognition helices when bound to the DNA. This prognostic of the target position of the recognition helices was made upon the methodical analysis of 27 structures of homodimeric transcriptional regulators, that present the Helix-Turn-Helix motif, complexed with DNA. This approach allowed us to describe the stereochemical fitting of the recognition helices

into the DNA major grooves with new geometrical parameters intimately related to the symmetry of the complex, the residue sequence of the recognition helices and the base sequence of the DNA site, providing thus support to model new complexes.

Keywords: Conformational changes. Collective motions. Molecular dynamics. Harmonic modes filter. Linear combination of modes.

LISTA DE ILUSTRAÇÕES

- Figura 1 – Esquema de um movimento funcionalmente importante levando a mioglobina de um estado intermediário a outro, com diferentes valores de energia de Gibbs, pela superfície de energia representada apenas na direção de uma coordenada conformacional. 35
- Figura 2 – a) Estrutura cristalográfica apo da ElrR representada em *cartoon* e com sua superfície. Em vermelho está indicado o DBD, em laranja uma região central e em amarelo uma região C-terminal. Prevê-se que o sítio alostérico se encontre entre essas duas últimas regiões. b) Esquema de um movimento que levaria as hélices de reconhecimento, indicadas em *cartoon*, a se encaixarem nos sulcos maiores do DNA. 40
- Figura 3 – a) Estrutura da MosR em sua forma reduzida e ligada ao DNA representada em *cartoon*. As cisteínas que formam uma ponte de dissulfeto em ambiente oxidativo estão representadas em *stick*. O DBD está colorido em *rainbow*. b) superposição em outra perspectiva das duas estruturas cristalográficas da MosR, ligada e não ligada ao DNA, de códigos PDB 4fx4 e 4fx0, respectivamente. A estrutura ligada está colorida como em a), enquanto que a estrutura não ligada ao DNA está colorida em lilás. 43
- Figura 4 – Esquema da transição alostérica entre os estados T e R da DEA, cujas estruturas cristalográficas possuem código de PDB 1fsf e 1lea, respectivamente, e estão desenhadas em *cartoon* com cada cadeia colorida de uma cor. Duas perspectivas são apresentadas: uma em cima e outra embaixo. Entre as conformações T e R estão representadas as mudanças conformacionais observadas indicadas por vetores para cada carbono α substituindo uma das cadeias, enquanto as demais estão representadas em *trace*. 48
- Figura 5 – Exemplo de duas hélices de reconhecimento, de mesma sequência, interagindo simetricamente com dois sulcos maiores vizinhos no DNA. Cada hélice de proteína faz as mesmas interações, mas com uma fita diferente do DNA. O exemplo mostra o complexo de código PDB 4fx4. 53

Figura 6 – Duas perspectivas das hélices de reconhecimento e ligadas ao DNA de um complexo TR-DNA (código PDB 4fx4) para ilustrar os parâmetros levantados para análise. O resíduo principal está representado em CPK em branco. O ponto principal de cada hélice de reconhecimento está representado em preto. O cilindro em preto une os dois pontos principais. Os vetores em preto representam os eixos das hélices de reconhecimento. O vetor em laranja representa o vetor ortogonal aos eixos das duas hélices de reconhecimento. Os pontos em vermelho representam os centros de massa de cada volta do DNA. O vetor em vermelho representa o eixo do DNA usado nos cálculos de λ e γ . O vetor em azul representa o eixo de DNA usado no cálculo de Dph, e o ponto em azul é a projeção do ponto principal sobre este vetor. 56

Figura 7 – a) monômero solvatado compondo um sexto de uma caixa hexagonal. b) hexagonal construído a partir do monômero simulado e suas imagens. c) as cinco operações de simetria cristalográfica que recompõem o hexâmero a partir do monômero. d) inequações geométricas usadas para excluir as moléculas de água cuja posição do oxigênio cumprisse essas características, restringindo o solvente a apenas um sexto da caixa hexagonal. 76

Figura 8 – Superposição de complexos TR-DNA por apenas um dos motivos HTH. A superposição foi feita considerando sempre o resíduo principal, os 10 resíduos anteriores a ele e os 9 posteriores a ele. No centro está representada essa região superposta de todos os complexo e também as respectivas hélices de reconhecimento completas. A hélice de reconhecimento da outra cadeia também está representada em cartoon para todos os complexos. A dupla hélice de DNA desenhada corresponde ao complexo de código PDB 4aik. Além deste estão desenhados os complexos de código PDB 4ivz, 2vz4, 3q5f, 3s8q, 3u3w, 3zkc, 3zpl, 4fth, 4fx4, 4ihs, 1z9c, 4jl3, 4l62, 2e1c, 1jt0, 2yvh, 3vok, 4i6z, 1zlk, 3a12 e 1h0m. 78

Figura 9 – Dois complexos TR-DNA exemplificando o tipo I acima (a e b) e o tipo II abaixo (c e d), sob dois pontos de vista rotados em 90° ao redor do eixo de simetria mostrado em preto. Coloridos em rainbow estão os motivos HTH. O tipo I é exemplificado pelo complexo de código PDB 4ivz, enquanto o de tipo II pelo de código PDB 2e1c. 79

Figura 10 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos do tipo I, evidenciando a distribuição de cargas dependente do valor do ângulo λ . Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. À esquerda está representado o complexo de código PDB 3q5f, enquanto que à direita está representado o de código PDB 4hf1. . 80

Figura 11 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos, evidenciando o tamanho do sítio de DNA, definido pelo número de bases entre as bases interagentes com os resíduos principais, e a distância pd , definida como a distância entre os pontos principais. Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. Os resíduos principais e as bases com que eles fazem pontes de hidrogênio estão coloridos em preto. À esquerda está representado o complexo de código PDB 1z9c, enquanto que à direita está representado o de código PDB 2yvh. 81

Figura 12 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos, evidenciando a inclinação das hélices de reconhecimento em relação ao eixo de DNA, definida por γ . Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. À esquerda está representado o complexo de código PDB 4fx4 pertencente ao tipo I, enquanto que à direita está o representado o de código PDB 2e1c pertencente ao tipo II. 82

Figura 13 – a) sequências de resíduos das hélices de reconhecimento das proteínas ElrR e NrtR, e as sequências de seus respectivos sítios de DNA. Em vermelho estão destacados resíduos que fazem ligação de hidrogênio com bases do sítio, os resíduos em verde são uma mutação conservativa e aqueles em amarelo são idênticos. As bases que obedecem uma simetria palindrômica no sítio de DNA da ElrR e da NrtR estão coloridas em amarelo à esquerda do centro de simetria, e em verde à direita do centro. b) estrutura resolvida do sítio de DNA da NrtR, colorida em verde e amarelo de acordo com as posições das bases que preservam a simetria das bases no sítio da ElrR, como indicado na tabela em a). O par de bases central está colorido em vermelho. 87

Figura 14 – a) representação elipsoidal dos ADPs para todos os átomos pesados da ElrR. b) Representação do sétimo modo normal com vetores de deslocamento atômico para os carbonos α . Nas duas figuras em a) e b) os domínios 1, 2 e 3 estão coloridos respectivamente em vermelho, laranja, e amarelo, exceto pelas regiões de conexão dos domínios, que compreendem os resíduos de 63 a 77 e 175 a 198, coloridas em branco. c) tabela indicando a divisão de domínios dinâmicos que resulta da análise feita pelo DynDom, mostrando os resíduos que compreendem cada domínio, e o RMSD pela superposição de cada domínio de duas distintas conformações da ElrR. 89

Figura 15 – a) estrutura cristalográfica da ElrR dividida em três corpos rígidos, circulos em preto. b) superposição de dois corpos rígidos da ElrR com a estrutura cristalográfica da NrtR ligada ao seu sítio de DNA (código PDB 3gz6), alinhando estruturalmente somente os carbonos α das hélices de reconhecimento. c) reconstrução dos três corpos rígidos para que as hélices de reconhecimento coincidam com as hélices alvo da NrtR e o terceiro domínio se posicione com o menor desvio possível das α -hélices 11, das duas cadeias concomitantemente. A estrutura do sítio de DNA da NrtR está representada em complexo com este modelo. Em toda a figura, as estruturas do complexo de código 3gz6 (NrtR e seu sítio de DNA) estão coloridas em azul. A ElrR está colorida seguindo a divisão de regiões descrita na legenda da Figura 14. 90

Figura 16 – a) um decoy do conjunto com RMSD em relação à estrutura original de 6,0 Å, com 35% de distorções nas distâncias entre carbonos α vizinhos. b) gráfico de distribuição dos *decoys* de acordo com o RMSD em relação à estrutura de origem, com o RMSD em relação às hélices alvo e ao número de distorções entre cada carbono α vizinho. Distorções são consideradas apenas quando a distância entre dois carbonos α vizinhos ultrapassa a faixa entre 3,6 e 4,1 Å. 91

Figura 17 – Gráfico do RMSD em relação às hélices alvo em função do tempo. As conformações foram guardadas a cada 0,5 ps durante uma simulação de dinâmica molecular de 100 ns. 92

Figura 18 – Os gráficos em a), b) e c) mostram a distribuição dos modelos de acordo com o RMSD em relação às hélices alvo e um segundo parâmetro que quantifica as distorções geradas, como o RMSD em relação à estrutura inicial, o número de distorções nas distâncias entre carbonos α vizinhos, ou a porcentagem de estrutura secundária preservada, respectivamente. Os modelos construídos pela combinação linear de modos normais estão representados em cinza, enquanto que os construídos por modos harmônicos filtrados estão representados em preto. d) modelo construído a partir de modos harmônicos filtrados para seguir nas análises. e) modelo construído a partir de modos normais para continuar com análises. Os modelos mostrados estão representados tanto em cartoon quanto em linhas, que também representa as cadeias laterais.	96
Figura 19 – a) modelo gerado pela superposição de corpos rígidos. A cadeia principal está representada em cartoon e colorida como descrito na Figura 14. A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza. b) outro ponto de vista indicando o sítio alostérico com uma seta vermelha.	97
Figura 20 – Conformação retirada da MD quando as hélices de reconhecimento mais se aproximaram do alvo. As duas perspectivas mostram os dois sítios alostéricos, indicados por uma flecha vermelha. A cadeia principal está representada em cartoon e colorida como descrito na Figura 14. A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza.	98
Figura 21 – a) e b) modelo obtido pela combinação de modos harmônicos filtrados. c) e d) modelo obtido pela combinação de modos normais. As duas perspectivas mostram os dois sítios alostéricos. A cadeia principal está representada em cartoon e colorida como descrito na Figura 14. A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza. O peptídeo resolvido em complexo com a proteína PrgX (código PDB 2axz), superposta com cada modelo, aparece em cartoon colorido em azul.	99
Figura 22 – Mapas de contato a) da conformação cristalográfica apo, b) do modelo construído pela combinação de modos harmônicos filtrados, e c) do modelo construído pela combinação de modos normais. d) Legenda de cores para as distâncias entre carbonos α . e) Conformação cristalográfica apo ressaltando os resíduos de número 45, 75, 180 e 200, também indicados nos eixos dos mapas de contato.	100

Figura 23 – Quatro conformações da MosR superpostas, ressaltando o posicionamento das hélices de reconhecimento, indicadas por uma espessura maior do *ribbons*, em duas orientações diferentes. Em verde está representada a conformação média calculada para o estado reduzido, considerando duas simulações de 400 ns, em azul está representada a conformação média calculada para o estado oxidado, considerando duas simulações de 400 ns, em rosa está representada a estrutura cristalográfica apo (código PDB 4fx0) e em amarelo está representado a estrutura cristalográfica holo (código PDB 4fx4), com seu DNA representado em laranja. 101

Figura 24 – RMSD das hélices de reconhecimento comparadas às da estrutura cristalográfica holo. As quatro trajetórias simuladas estão representadas. Em verde escuro a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4. 103

Figura 25 – Distribuição das conformações retiradas das simulações de acordo com a distância relativa entre as hélices de reconhecimento durante as trajetórias simuladas em comparação com a estrutura cristalográfica holo. No eixo das abscissas está a distância entre as hélices, cujo vetor possui uma direção normal às hélices. No eixo das ordenadas está a distância entre os centros de massa (CM), definida pela projeção do vetor diferença entre os CMs num plano paralelo às hélices. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3, em azul claro a simulação 4, e em laranja a estrutura cristalográfica holo. 104

Figura 26 – Distribuição das conformações retiradas das simulações de acordo com os ângulos entre cada hélice de reconhecimento e sua respectiva na estrutura cristalográfica holo. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4. o ponto de coordenadas (0,0) corresponderia à própria estrutura cristalográfica holo. 105

Figura 27 – ADPs isotrópicos calculados para cada simulação e separadamente para cada cadeia em diferentes janelas de tempo: a) média dos valores obtidos as janelas de 1 ns; b) média dos valores obtidos para as janelas de 50 ns. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4. . 106

Figura 28 – Mapas de contato entre o N-terminal e a cadeia oposta. Um contato é considerado quando dois carbonos α ficam a menos de 6 Å entre si. A escala de cores indica quantas vezes um contato foi encontrado durante as simulações, sendo no máximo 40.000 vezes, já que foram guardadas 40 mil conformações para cada trajetória de 400 ns. No eixo das abscissas estão indicados os números dos resíduos do N-terminal, enquanto no eixo das ordenadas estão os números dos resíduos da cadeia oposta. Cada mapa indica os contatos encontrados em cada N-terminal simulado. Em a) o N-terminal da cadeia A da simulação 1, b) o N-terminal da cadeia B da simulação 1, c) o N-terminal da cadeia A da simulação 2, d) o N-terminal da cadeia B da simulação 2, e) o N-terminal da cadeia A da simulação 3, f) o N-terminal da cadeia B da simulação 3, g) o N-terminal da cadeia A da simulação 4, e em h) o N-terminal da cadeia B da simulação 4. 108

Figura 29 – Interações do N-terminal com a α -hélice 3' da cadeia oposta. a) detalhe do mapa de contato na Figura 28e, mostrando os contatos entre os resíduos de 1 a 7 do N-terminal da cadeia A com os resíduos de 62 a 70 da cadeia oposta na simulação 3, em estado oxidado. b) detalhe do mapa de contato na Figura 28a, mostrando os contatos entre os resíduos de 1 a 7 do N-terminal da cadeia A com os resíduos de 62 a 70 da cadeia oposta na simulação 1, em estado reduzido. c) conformação retirada da simulação 3 aos 50 ns, mostrando a posição do N-terminal da cadeia A. d) conformação retirada da simulação 1 aos 50 ns, mostrando a posição do N-terminal da cadeia A. e) estruturas cristalográficas apo e holo superpostas e mostrando a posição relativa das α -hélices 1 e 3', e as diferentes ligações de hidrogênio feitas pela I66'. A estrutura holo está em amarelo, com o DNA em complexo colorido em laranja, e a estrutura apo está em rosa. 109

Figura 30 – Distância atômica intercatenária entre o nitrogênio da Ala6 e o oxigênio da Arg65 da cadeia oposta durante a) a simulação 1 e b) a simulação 3. 110

Figura 31 – Parâmetros de deslocamento atômico isotrópico calculados somente para os carbonos α como médias de janelas de 50 ns. Os valores calculados para a primeira metade da simulação 1 estão em rosa e os da segunda metade estão em laranja. 111

Figura 32 – Parâmetros de deslocamento atômico anisotrópico representados como elipsoides somente para os carbonos α , calculados considerando os 200 ns de cada metade da simulação 1. As conformações médias de cada metade da simulação estão mostradas em a) rosa para a primeira metade e b) laranja para a segunda metade. As hélices de reconhecimento de cada conformação média foram superpostas com as hélices de reconhecimento da estrutura cristalográfica holo, e estão representadas em preto, assim como o sítio de DNA ligado a elas. 111

Figura 33 – Vetores de deslocamento atômico para cada carbono α representando o modo de mais baixa frequência recuperado de a) a primeira parte da simulação 1, e b) a segunda parte da simulação 1. As conformações representadas são as primeiras do período considerado para esta análise. A da primeira parte está colorida em rosa e a da segunda parte em laranja. Cada cadeia está colorida com um tom diferente. 112

Figura 34 – Análise da simulação restringida em comparação com a primeira parte da simulação 1, ambas em estado reduzido. As conformações da simulação restringida estão representadas em marrom, as conformações da simulação 1 em verde, a estrutura cristalográfica holo em laranja e a estrutura cristalográfica apo em rosa. a) distâncias atômicas entre o nitrogênio da Ala6 e o oxigênio da Arg65 da cadeia oposta durante a simulação restringida. b) ADPs isotrópicos para as duas cadeias da MosR calculados como médias das janelas de 50 ns tanto para a simulação restringida, quanto para a primeira parte da simulação 1. c) distâncias relativas entre as duas hélices de reconhecimento. d) ângulo que cada hélice de reconhecimento da conformação simulada faz com sua respectiva hélice da estrutura holo. e) e f) duas perspectivas da conformação média da simulação restringida superposta à conformação média da primeira parte da simulação 1 e à estrutura cristalográfica apo. As hélices de reconhecimento estão em destaque. 114

- Figura 35 – Representação em *trace* de diferentes conformações da DEA. a) superposição das conformações T e R cristalográficas, T em preto e R em branco. b) A conformação cristalográfica T com os vetores de deslocamento atômico para os carbonos α representando a transição para a conformação R. c e d) hexâmero da conformação T cristalográfica com uma das cadeias representada apenas pelos vetores de deslocamento atômico dos carbonos α da conformação T para a R. Em c) o plano da figura está perpendicular ao eixo de ordem 3 da DEA, e em d) o plano da figura está paralelo a ele. e) e f) o hexâmero na conformação T cristalográfica e os vetores de deslocamento dos carbonos α no modo elástico 7. Cada cadeia está colorida por uma cor diferente. O hexâmero em e) e f) estão orientados assim como em c) e d), respectivamente. g) conformação T cristalográfica representando o modo elástico 7 por vetores de deslocamento atômico para os carbonos α . A região reguladora está circulada em vermelho, destacando o *loop* com os resíduos de 144 a 154, e ampliada em h). 116
- Figura 36 – Representação em *trace* de diferentes conformações da DEA. a), b), c) e d) representam a conformação T após a minimização necessária para ANM, com os vetores de deslocamento atômico para os carbonos α representando um modo normal. Em a) e b) está representado o hexâmero com cada cadeia colorida por uma cor diferente e os vetores indicando o modo normal 7, os planos das figuras estão orientados perpendicularmente e paralelamente ao eixo de ordem 3, respectivamente. Em c) está representada a cadeia B com os vetores indicando o deslocamento dos carbonos α no modo normal 7. Em d) está representada a cadeia C com os vetores indicando o deslocamento dos carbonos α no modo normal 9. e) e f) representam a primeira conformação da dinâmica do hexâmero considerada para o cálculo do filtro de frequências, cada cadeia está colorida por uma cor diferente e os vetores indicam o modo harmônico filtrado número 1. Os planos das figuras estão orientados como em a) e b), respectivamente. g) e h) representam a primeira conformação da dinâmica do monômero e os vetores indicam os modos harmônicos filtrados números 1 e 41, respectivamente, sendo que para o modo 41 os módulos estão multiplicados por 20. A região reguladora está circulada em vermelho, destacando o *loop* com os resíduos de 144 a 154. 120

LISTA DE TABELAS

Tabela 1 – Valores observados nos complexos de tipo I estudados para os parâmetros que levantamos (ver seção de métodos). A coluna hélice mostra os resíduos que definem a hélice de reconhecimento na sequência primária da proteína, a coluna rp remete ao resíduo principal, a coluna outros mostra outros resíduos interagentes especificamente com bases, na coluna sítio está o tamanho do sítio de DNA e wHTH remete à participação na subfamília <i>winged helix</i>	84
Tabela 2 – Valores observados nos complexos de tipo II estudados para os parâmetros que levantamos (ver seção de métodos). A coluna hélice mostra os resíduos que definem a hélice de reconhecimento na sequência primária da proteína, a coluna rp remete ao resíduo principal, a coluna outros mostra outros resíduos interagentes especificamente com bases, na coluna sítio está o tamanho do sítio de DNA e wHTH remete à participação na subfamília <i>winged helix</i>	85
Tabela 3 – Número de interações específicas encontradas nos complexos estudados distribuídas em função dos resíduos e bases interagentes.	86
Tabela 4 – Relação dos 20 modos normais e modos harmônicos filtrados com os menores valores de frequência (Freq.). Eles foram enumerados em ordem crescente de acordo com o aumento na frequência. Os primeiros 6 modos normais foram desconsiderados. Estão relacionados também o grau de coletividade (Col.) e o produto escalar (PE) com o vetor diferença entre as hélices de reconhecimento alvo, e sua posição original na conformação cristalográfica. Para os modos harmônicos filtrados também estão relacionadas as amplitudes do movimento, considerando tanto todos os carbonos α (amplitude global - Amp.global), quanto somente aqueles das hélices de reconhecimento (amplitude local - Amp.local).	94
Tabela 5 – Distâncias atômicas observadas nas quatro conformações da MosR mostradas na Figura 23, estrutura cristalográfica apo (código PDB 4fx0), estrutura cristalográfica holo (código PDB 4fx4), conformação média calculada para o estado reduzido simulado e conformação média calculada para o estado oxidado simulado. Duas distâncias são mostradas para cada conformação em relação à cadeia A e outras duas em relação à cadeia B. A primeira é a distância intracatenária da cisteína 10 à treonina 14, e a segunda é a distância intercatenária entre a cisteína 12 e asparagina 37 da cadeia oposta.	102

Tabela 6	–	Frequência (Freq.), grau de coletividade (Col.), amplitude global (Amp.global) e amplitude somente das hélices de reconhecimento (Amp.local) dos 5 primeiros modos recuperados de cada metade da simulação 1 (primeira parte e segunda parte).	112
Tabela 7	–	Relação dos 10 modos de mais baixa frequência obtidos por redes elásticas e NMA. Os modos estão relacionados por número, frequência (Freq.), grau de coletividade(Col.) e produto escalar (PE) com o vetor que descreve a transição entre as conformações cristalográficas T e R.	117
Tabela 8	–	Relação dos 20 primeiros modos obtidos pelo método de filtro de frequências da simulação do hexâmero e da do monômero. Os modos estão relacionados por número, frequência (Freq.), grau de coletividade(Col.) e produto escalar (PE) com o vetor que descreve a transição entre as conformações cristalográficas T e R.	118
Tabela 9	–	Valores do produto escalar entre o vetor que representa cada cadeia do hexâmero em cada modo listado e o vetor que representa a transição alostérica entre as conformações cristalográficas T e R. Estão listados os 20 primeiros modos harmônicos filtrados da simulação do hexâmero, além do modo normal 7 e o modo elástico 7	121

LISTA DE ABREVIATURAS E SIGLAS

RNA	Ácido Ribonucleico
DNA	Ácido Desoxirribonucleico
IPTG	Isopropil β -D-tiogalactosídeo
TR	Regulador Transcricional
HTH	Hélice-Volta-Hélice
PDB	<i>Protein Data Bank</i>
RMSD	Raiz do desvio quadrático médio
RMN	Ressonância magnética nuclear
ITC	Calorimetria de titulação isotérmica
DBD	Domínio de ligação ao DNA
SPR	Ressonância de superfície plasmônica
ENM	Modelo de redes elásticas
NMA	Análise por modos normais
PCA	Análise de componentes principais
RT-PCR	Transcrição reversa da reação em cadeia da polimerase
MWC	Modelo concertado de Monod-Wyman-Changeux
PE	Produto escalar

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Sobre regulação transcricional	29
1.2	O encaixe de reguladores transcricionais em seu sítio de DNA	30
1.3	Mudanças conformacionais e movimentos coletivos	32
1.4	A análise computacional dos movimentos biologicamente funcionais de três proteínas	37
1.4.1	ElrR	38
1.4.2	MosR	42
1.4.3	DEA	46
2	OBJETIVOS	49
3	BASE TEÓRICA E METODOLOGIA	51
3.1	O motivo HTH em Domínios de Ligação ao DNA	51
3.2	O estudo de movimentos coletivos	57
4	MÉTODOS	69
4.1	Parâmetros de análise estrutural e de flexibilidade	69
4.1.1	RMSD	69
4.1.2	Conservação de hélices α	69
4.1.3	Distorções nas distâncias entre carbonos α vizinhos	69
4.1.4	Mapa de contatos	70
4.1.5	Parâmetro de deslocamento atômico isotrópico e anisotrópico	70
4.1.6	Produto escalar, ou <i>overlap</i>	70
4.1.7	Amplitude do modo harmônico	71
4.1.8	Grau de coletividade	71
4.1.9	Alinhamento, ou superposição, estrutural	71
4.2	Parâmetros de simulação	72
4.2.1	ElrR	72
4.2.2	MosR	73
4.2.3	DEA	74
5	RESULTADOS E DISCUSSÃO	77
5.1	Sobre os complexos TR-DNA	77
5.2	Sobre a construção de um modelo para a ElrR ligada ao DNA	86
5.2.1	As hélices de reconhecimento alvo	86
5.2.2	A decomposição e a movimentação em corpos rígidos	88

5.2.3	Os <i>decoys</i>	90
5.2.4	Os modos normais e modos harmônicos filtrados de dinâmica molecular . . .	92
5.2.5	A combinação linear dos modos no método de mínimos quadrados	94
5.2.6	A análise dos modelos selecionados em cada método	97
5.3	Sobre a afinidade da MosR pelo DNA	99
5.3.1	As hélices de reconhecimento	102
5.3.2	Sobre a participação do N-terminal	106
5.3.3	Impondo uma restrição	110
5.4	Sobre os modos coletivos da DEA	113
6	CONCLUSÕES E PERSPECTIVAS	123
6.1	Sobre os complexos TR-DNA	123
6.2	Sobre o método de filtro de modos harmônicos	124
6.3	Sobre a construção do modelo da ElrR ligada ao DNA	124
6.4	Sobre a afinidade da MosR pelo DNA	125
6.5	Sobre os modos coletivos da DEA	126
6.6	Conclusões finais	127
	REFERÊNCIAS	129
	APÊNDICES	137
	APÊNDICE A – ARTIGO EM FASE DE CORREÇÃO PARA SER SUBMETIDO EM BREVE NA PLOS ONE	139

1 INTRODUÇÃO

1.1 Sobre regulação transcricional

Supunha-se que a complexidade de um organismo estivesse diretamente relacionada não só à extensão de seu genoma, mas também à quantidade de fatores de transcrição que ele possuísse. A mosca da fruta, *Drosophila*, possui não mais que 14 mil genes e por volta de mil fatores de transcrição, enquanto os humanos possuem cerca de 30 mil genes e 3 mil fatores de transcrição. O dobro de genes ou mesmo o triplo de fatores de transcrição entre drosófilas e humanos não parecem explicar a diferença de complexidade entre esses dois organismos, mas basta esse pequeno aumento no número de genes e fatores de transcrição para elevar enormemente as combinações deles dentre os mecanismos de regulação e então permitir a miríade de fenótipos que observamos.¹

A robustez dos mecanismos de regulação, sem implicar no aumento do número de genes, permite que os organismos sejam muito estáveis fenotipicamente ante perturbações genéticas e ambientais.² Ou seja, a grande redundância existente nas vias de regulação para a obtenção de determinado fenótipo ou simplesmente a combinação de mais de uma proteína para sua promoção ou inibição, permitem que o resultado final não seja alterado significativamente caso ocorra alguma falha em uma via de regulação ou função de alguma proteína.

A regulação do fenótipo já na etapa de transcrição evita a produção de RNA e proteína nas etapas seguintes, de tratamento do RNA mensageiro e tradução, poupando a célula de gastos energéticos. Assim, existem diversas formas de regulação nesta que é uma etapa primária, formando uma rede de transcrição. Nessa rede de transcrição, as proteínas reguladoras podem ser divididas em três grandes grupos: aquelas que se ligam especificamente a sítios de DNA, promovendo uma transcrição seletiva; aquelas que são componentes do complexo multiproteína que interage com a RNA polimerase e permitem o reconhecimento da região promotora; e aquelas que auxiliam a transcrição modificando e remodelando a cromatina.¹

Interessantemente, a seletividade dos fatores de transcrição, que se ligam a sítios específicos do DNA, também pode ser acompanhada de grande complexidade. Esses fatores de transcrição são sensíveis a mudanças no ambiente que alternam seu estado entre ativo ou inativo. Assim, não é necessária a produção de novas proteínas para regular a transcrição gênica, basta controlar a ativação dos fatores de transcrição.

Um exemplo é o regulador transcricional NF- κ B, que ativa uma variedade de genes relacionados a mecanismos de defesa do sistema imunológico. Ele reside no citoplasma em estado inativo quando ligado ao inibidor I κ B. Livre deste, ele adentra o núcleo e ativa a

transcrição gênica.^{3,4}

Outro clássico exemplo é o repressor lac, normalmente expresso na célula. É a presença ou ausência de um fator indutor, a alolactose ou o IPTG, que regula a função do repressor lac e indiretamente a transcrição do operon lac. Na presença do indutor, a afinidade do repressor lac pelo sítio de DNA a que se liga diminui, e, quando ele se desliga do DNA, a RNA polimerase pode iniciar a transcrição do operon lac. Essa sensibilidade do repressor lac, ou percepção às mudanças no ambiente, especificamente, à presença do indutor, é possível porque ele possui tanto um sítio de ligação ao DNA, quanto um sítio de ligação ao indutor. Assim, a interação com um dos ligantes, DNA ou indutor, afeta a afinidade pelo outro. A esse mecanismo é dado o nome de alosterismo.⁵

É nesse cenário que se introduz este estudo, com a motivação de compreender o mecanismo de regulação dos fatores de transcrição, quais características estruturais viabilizam a afinidade deles pelo sítio de DNA, e como essas características são afetadas pelas mudanças no sítio alostérico.

1.2 O encaixe de reguladores transcricionais em seu sítio de DNA

Alosterismo é uma palavra que deriva do grego, *állos* e *stereós*, que significam outro ou diverso, e sólido ou tridimensional, respectivamente. Juntos, esses morfemas passam a ideia de algo com diversas formas tridimensionais. Para uma proteína, referem-se ao conceito de que ela possua diversas conformações estruturais, cada uma delas com afinidades diferentes por ligantes diferentes, também com distintas formas tridimensionais.

Sendo os reguladores transcricionais (TR) proteínas alostéricas, eles possuem a habilidade de se ligar ao DNA e também de se combinar a outros ligantes, macros ou micros. Essa natureza combinatória permite um aumento na complexidade das redes de transcrição e, conseqüentemente, uma grande diversidade fisiológica e fenotípica.¹

É claro que por existirem conformações distintas, algumas terão maior afinidade pelo DNA que outras. Como o DNA apresenta uma estrutura tridimensional bastante uniforme em toda a sua extensão, a estrutura de hélice dupla fita, para que uma proteína interaja especificamente com suas bases é preciso que ela também apresente algumas certas características estereoquímicas.

Apenas alguns domínios de ligação ao DNA estão descritos. Em 1991, Harrison listou alguns motivos estruturais bastante conservados.⁶ A lista ainda estava incompleta, mas a maioria dos casos conhecidos até hoje cai nesta discriminação. São eles o HTH (*helix-turn-helix*); o Zn-binding; o *leucine-zipper coiled-coil*; e o β -ribbon. Os três primeiros motivos utilizam uma α -hélice para o encaixe no sulco maior do DNA, que promove a interação entre resíduos de aminoácidos e bases de DNA. Nos dois primeiros, ela é denominada de hélice de reconhecimento, e se ressalta de todo o domínio para encaixar no

DNA. Os dois últimos motivos apresentam estruturas mais flexíveis, capazes de se enrolar no DNA. No β -*ribbon*, por exemplo, o longo par de fitas β antiparalelas é que se encaixa no DNA. Cada fita pertence a uma cadeia de um homodímero, assim, por terem a mesma sequência e se posicionarem de forma antiparalela, existe no centro desse motivo um eixo de simetria que deve ser acompanhado de uma simetria também nos pares de bases do DNA.

Essa simetria proveniente da homodimerização é uma característica que prevalece em TRs mesmo quando possuem os outros domínios de ligação ao DNA listados acima. Isso porque esses motivos, como monômeros, são capazes de interagir somente com algumas poucas bases. Assim, a simetria provê um sítio maior de ligação ao DNA, não só fortalecendo a interação com a proteína, mas também aumentando a seletividade do sítio de ligação. Harrison conclui, em seu artigo de 1991, que essas condições estruturais devem ter sido mais importantes para a seleção evolutiva desses motivos que as características do próprio DNA.⁶

Até 1991, foram publicadas apenas 20 estruturas contendo domínios de ligação ao DNA, 12 delas em complexo com DNA. Hoje, 4.376 estruturas resultam da busca por “DNA binding domain” no *Protein Data Bank* (PDB),⁷ dentre elas 883 formam um complexo de proteína com DNA. Como mais da metade desses complexos foram resolvidos somente de 2010 em diante, vimos uma oportunidade para rever neste estudo os requisitos estruturais necessários à interação proteína-DNA. Essa ampla amostragem de estruturas, possível somente nos dias atuais, nos motivou a encontrar traços geométricos mais conservados. Para isso, escolhemos em 2014 estruturas depositadas no PDB correspondentes à busca por reguladores transcricionais ligados a DNA. Destas estruturas – 59 ao todo –, selecionamos apenas as 39 que possuíam o motivo HTH, o mais simples e mais populoso dos motivos listados acima, e que fossem homodiméricas (muitas estruturas eucarióticas evoluíram para heterodímeros ainda reconhecendo sítios simétricos de DNA). Dentre estas, somente 27 realmente apresentavam um sítio de ligação ao DNA com certa simetria, ou palíndromia, excluindo aquelas estruturas que representassem o mesmo regulador transcricional.

Na análise desses complexos, considerando a simetria dos dois motivos HTH presentes em cada estrutura, procuramos compreender como se dá o encaixe das duas hélices de reconhecimento nos sulcos maiores do DNA, que equivalem à segunda hélice do HTH em cada DBD. Ela recebeu esse nome por ser uma das principais estruturas a fazer ligações específicas com as bases de DNA. Ainda que outros *loops* também sejam capazes de fazer interações com bases, eles nem sempre estão presentes nos DBDs com o motivo HTH, ao contrário das hélices de reconhecimento. Além disso, alguns estudos indicam que as hélices de reconhecimento são as primeiras a interagir com o DNA no processo de associação do complexo TR-DNA.⁸ Então, na tentativa de compor uma nova descrição para essa interação, buscamos por parâmetros geométricos como ângulo entre as

hélices de reconhecimento, ângulo entre elas e o DNA, curvatura do DNA, distância entre as hélices de reconhecimento e a proximidade destas ao eixo do DNA, além de relacionar esses parâmetros a características estruturais como tamanho das hélices de reconhecimento, tamanho do sítio de DNA, e as próprias sequências de resíduos de aminoácidos e pares de bases.

Uma segunda motivação para essas análises, mas não menos importante, é a busca por parâmetros geométricos, ou variáveis coletivas, que possam distinguir conformações da proteína como viáveis ao encaixe nos sulcos do DNA ou não. Esses parâmetros facilitaram nossos estudos da MosR, uma das proteínas que analisamos, pela compreensão de como as mudanças conformacionais modificam o domínio de ligação ao DNA e alteram a afinidade da proteína por seu sítio. Com a definição desses parâmetros ainda fizemos um prognóstico de como se daria a ligação da ElrR – outra das proteínas que estudamos – com o DNA, como será discutido adiante. Compreender as mudanças alostéricas, que envolvem as mudanças relativas entre DBD, implica de antemão numa discussão sobre flexibilidade e estados conformacionais em proteínas.

1.3 Mudanças conformacionais e movimentos coletivos

Quando Monod cunhou o termo alosterismo para proteínas em 1963,⁹ ele partiu de um primeiro pressuposto de que tais proteínas alostéricas possuam pelo menos dois sítios distintos em forma e distantes entre si. Um deles é o sítio ativo, o outro foi denominado sítio alostérico. Cada um deles deve interagir com ligantes diferentes, também distintos em forma um do outro, e denominados ligantes ativo e alostérico, respectivamente. Esses ligantes não passam por nenhuma reação ou interação entre si. As alterações identificadas na atividade da proteína mediante a ligação de um ou de outro devem surgir, então, unicamente das mudanças conformacionais na proteína induzidas ou estabilizadas pela interação com o ligante alostérico. A essas mudanças conformacionais deu-se o nome de transição alostérica.

Nessa transição alostérica, a proteína transita entre dois estados conformacionalmente distintos e com diferentes afinidades pelo ligante ativo. Um detalhe fundamental reside em definir que essa transição seja induzida ou estabilizada pelo ligante alostérico. Na primeira situação, essas mudanças só ocorrem mediante a ligação deste; na segunda, essas mudanças fazem parte da dinâmica da proteína, e na presença do ligante uma conformação se torna mais estável que outra, mudando o equilíbrio dinâmico entre os estados conformacionais.

Por muito tempo, aceitou-se a compreensão de que as proteínas formavam corpos com estrutura estática. Já a hipótese do encaixe induzido e o modelo chave-fechadura admitiam certa plasticidade local pela interação com seus ligantes. Com o avanço das técnicas de RMN e com as maiores resoluções alcançadas na cristalografia, ganhou respaldo

o conceito de que as proteínas estejam se movimentando, não só rodando e transladando, mas com cada átomo seu sendo constantemente atraído e repelido por seus vizinhos próprios e pelos de outros compostos, também em movimento, que existem no mesmo ambiente repleto de micro e macromoléculas e seus multicomplexos.

Logo em 1975, Weber descreveu as proteínas como se estivessem gritando e esperneando, o que caracteriza um movimento caótico e incompreensível, mas já longe de ser estático.¹⁰ Em 1991, Frauenfelder propõe que, ainda que complexos, esses movimentos possam ser caracterizados e classificados.¹¹ Ele conclui que muitos fenômenos biológicos, assim como as regulações alostéricas, devem se desenrolar e evoluir de acordo com um espaço energético acessível.

O termo adotado originalmente em inglês para esse espaço foi “energy landscapes”, que remete à ideia de que a proteína esteja sujeita a um potencial energético cheio de vales e morros, suavidades e rugosidades, como numa paisagem com acidentes geográficos. Frauenfelder propõe que esse espaço de potencial energético também rege a dinâmica da proteína, e que seus movimentos podem ser classificados de acordo com o tipo de “relevo” em que se encontra determinada conformação dela.

Esse espaço acessível de energia e conformações é único e intrínseco a cada estrutura. Uma massa conectada a uma mola com uma extremidade fixa possui uma simples curva de energia, um potencial harmônico unidimensional, em função de apenas um parâmetro, que é a posição da massa. As distâncias que essa massa pode alcançar, as amplitudes do movimento, dependem também da energia cinética que ela possui. Uma descrição semelhante pode ser feita para proteínas num cenário mais complexo. A energia potencial de uma proteína pode ser descrita em função das suas conformações, ou seja, das posições de cada átomo num espaço de $3N$ dimensões, onde N é o número de átomos. Já a energia cinética, proporcional à temperatura, define a acessibilidade neste espaço, quão distantes são as conformações que essa proteína pode adotar. Se são necessárias altas temperaturas para que uma proteína passe de uma conformação a outra, então essas duas conformações estão separadas por uma barreira de potencial energético. Caso contrário, essas duas conformações devem se encontrar no mesmo poço de potencial.

Convencionalmente, o conceito de que duas conformações estruturais estejam separadas por uma barreira energética já era adotado para explicar o funcionamento das proteínas e suas energias de ligação. Nesse sentido, cada uma dessas conformações designava um estado diferente da proteína, ligado ou desligado de seus ligantes. E entendia-se que a cada um desses estados correspondia uma única conformação.

No entanto, Frauenfelder identificou experimentalmente que a reação de ligação da mioglobina às moléculas de O_2 ou CO acontecia em um único passo a temperatura fisiológica, enquanto que a temperaturas menores que 200 K a taxa de ligação não era exponencial, o que sugeria que a “mioglobina pudesse assumir um grande número de

estruturas levemente diferentes, cada uma com uma taxa de ligação diferente”.¹¹ Cada uma dessas estruturas levemente diferentes ele denominou de subestados conformacionais, ou microestados, que pertenciam a poços de energia potencial menores no fundo do grande poço de mesmo estado. No fundo desses subestados, outros subestados deveriam existir separados por barreiras energéticas ainda menores, e igualmente menores diferenças estruturais. Inúmeros níveis de subestados podem existir, assim como uma gama de movimentos que permitem a proteína mudar sua conformação e passar de um subestado a outro, ou de um estado a outro, devido às múltiplas dimensões de espaço conformacional acessível.

Esses movimentos também são diferentes dependendo das barreiras de potencial que precisam superar com a mesma energia cinética, e nem todos eles estão relacionados à função biológica da proteína. De forma simplificada, já em 1985, Frauenfelder dividiu os movimentos em dois tipos, flutuações em torno do equilíbrio, e movimentos funcionalmente importantes.¹² Nos do primeiro tipo, “as flutuações levam a proteína de um subestado a outro, e a energia interna e a entropia podem ser determinadas pela termodinâmica do equilíbrio”. Os do segundo tipo são “movimentos fora do equilíbrio e levam a proteína de um estado a outro”. Outras características separam esses dois tipos. Os do primeiro tipo ocorrem com alta frequência e não provocam mudanças muito grandes na conformação da proteína, apenas mudanças locais, mas podem levar a muitos dos subestados localizados na vizinhança, sendo assim isotrópicos. Os movimentos fora do equilíbrio ocorrem com baixa frequência e requerem que um grupo de átomos se mova coletivamente para levar a proteína a conformações mais distantes, em outro estado. Como não existem muitos outros estados na vizinhança, esses movimentos apresentam direções preferenciais e são, portanto, anisotrópicos.¹³ Nem todos os movimentos anisotrópicos são biologicamente, ou funcionalmente, importantes, mas todas as mudanças conformacionais biologicamente funcionais, por serem de grande amplitude, são anisotrópicas. A [Figura 1](#), retirada dos trabalhos de Frauenfelder com a mioglobina, exemplifica um movimento funcionalmente importante levando a mioglobina de um estado intermediário a outro pela superfície de energia.

Em 1987, Karplus calculou, com uma trajetória de dinâmica molecular de 300 ps, que as conformações visitadas nessa simulação, separadas a cada 0,05 ps e minimizadas, às vezes pertenciam ao mesmo poço de potencial energético e às vezes a outro, levando à confirmação da existência de múltiplos subestados conformacionais.¹⁴ Ele também pôde quantificar as flutuações na estrutura com cálculos de RMSD (*Root Mean Square Deviation*), e concluiu que entre 70 a 80 por cento delas surgem da transição entre pequenos poços de energia.

No mesmo ano, Karplus também considerou essas flutuações no cálculo da entropia configuracional de desnaturação da proteína BPTI (*bovine pancreatic trypsin inhibitor*).¹⁵

Figura 1 – Esquema de um movimento funcionalmente importante levando a mioglobina de um estado intermediário a outro, com diferentes valores de energia de Gibbs, pela superfície de energia representada apenas na direção de uma coordenada conformacional.

Fonte: Adaptada de FRAUENFELDER.¹²

Tradicionalmente estipulava-se, para este cálculo, que a entropia configuracional da proteína nativa fosse igual a zero, mas ela não deveria ser, considerando as flutuações entre subestados e mesmo dentro de um subestado. Com aproximações harmônicas, Karplus encontrou que a entropia vibracional interna era quase uma ordem de magnitude maior que a própria entropia configuracional de desnaturação encontrada experimentalmente. Ele resolveu essa inconsistência considerando haver duas contribuições no cálculo da entropia da proteína desnaturada, “uma devida às flutuações na vizinhança de uma estrutura bem definida, e a outra corresponde à existência de mais de uma estrutura bem definida”. Ou seja, este modelo considera que uma proteína possa ter inúmeras conformações de estrutura bem definidas, com flutuações ao redor de cada uma delas, correspondendo a um potencial com inúmeros poços harmônicos multidimensionais e desconexos entre si. O resultado experimental é válido somente quando a média ponderada da entropia vibracional da proteína desnaturada, derivada das flutuações em torno de cada estrutura bem definida, for semelhante à entropia vibracional da proteína enovelada, que supostamente possuía apenas uma conformação de estrutura bem definida. Dessa forma, no cálculo da diferença entre proteína enovelada e desnaturada, esse termo devido às flutuações pode ser negligenciado. No entanto, ele deve ser considerado explicitamente nos casos em que a proteína possua diferentes entropias de vibração em cada um de seus estados ligado ou desligado do ligante.¹⁵

O uso por Karplus dessas simulações, cálculos de mínimo de energia, RMSD e aproximações harmônicas deu suporte ao conceito proposto por Frauenfelder de *energy landscape* e dos diferentes tipos de movimento da proteína associados aos estados e subestados e às barreiras energéticas entre eles. Também deu suporte à ideia de que a dinâmica da proteína e sua entropia estejam intimamente relacionadas. Com o avanço das

técnicas de RMN (Ressonância Magnética Nuclear) para análise de dinâmica no final dos anos 90,¹⁶ os cálculos de Karplus receberam mais apoio experimental. Mais recentemente, em 2012, na associação de RMN com ITC (*Isothermal titration calorimetry*), Kalodimos chegou à conclusão de que a função da proteína é regulada pela combinação entre seus movimentos internos rápidos e lentos, que estão relacionados, respectivamente, com a entropia conformacional e com estados conformacionais pouco populosos.¹⁷ Testando 11 variações do regulador transcricional homodimérico CAP (*catabolite activator protein*) ele mediu tanto as componentes termodinâmicas ΔG , ΔH e $-T\Delta S$ por experimentos de ITC quanto a entropia conformacional por experimentos de RMN. Também por RMN, pela determinação de *chemical shifts*, ele encontrou a porcentagem de moléculas em estado ativo, distinguindo as populações em dois diferentes estados conformacionais dos domínios de ligação ao DNA (*DNA binding domain*, DBD). Essas porcentagens não foram proporcionais à afinidade da proteína, como era de se esperar (quanto maior a população de DBD em estado ativo, maior deveria ser a afinidade pelo DNA). Mas ele encontrou uma relação entre as componentes listadas e afinidade encontrada experimentalmente, e concluiu que “a afinidade nas interações entre proteína e ligante pode ser fortemente modulada pela resposta global de dinâmica interna, mesmo quando a interface de ligação não é perturbada”. Portanto ambos os movimentos, tanto aquele entre estados, quanto aquele relacionados a flutuações em equilíbrio, como denominados por Frauenfelder, participam para regular a interação da proteína.

Enfim, desde que Monod lançou o conceito de alosterismo, muito se avançou no que diz respeito ao movimento que as proteínas alostéricas podem fazer, seja entre estados ou dentro do mesmo estado. Compreender a dinâmica que remodela a estrutura de uma proteína a ponto de regular sua função, abre a oportunidade de intervir na atividade enzimática e na afinidade da proteína por seus ligantes sem de fato manipular seu sítio ativo, mas controlando seu movimento a partir de um ponto distante.

O avanço da nossa compreensão sobre a dinâmica da proteína foi acompanhada, ou possibilitada, pelo avanço também de técnicas experimentais, não só de RMN. Em um artigo de revisão feito por Kern e Henzler-Wildman em 2007,¹⁸ foram relacionados os movimentos de uma proteína com as barreiras energéticas que devem ser superadas, a escala de tempo envolvida em cada processo e que experimentos são capazes de detectá-los. Elas destacaram os métodos computacionais por terem a vantagem imbatível de prover a exata posição dos átomos a qualquer instante do movimento e a exata energia dessa conformação, cálculo possível devido ao enorme avanço na descrição dos campos de força. Desde 2007 a rapidez com que as equações de movimento são calculadas também evoluiu, e o então limite de poucos microssegundos já subiu para milissegundos desde 2009 com o supercomputador Anton, projetado especialmente para a simulação de dinâmicas moleculares.¹⁹ Apesar deste exemplo particular, o cálculo de longas dinâmicas também se tornou mais acessível com o desenvolvimento das unidades gráficas de processamento, as

GPUs. Nosso próprio grupo, por exemplo, possui um computador com dois processadores Xeon®, totalizando 32 cores, e mais duas GPUs Titan Z, cada uma com dois processadores gráficos. Nessa construção, podemos simular com o Gromacs 45 ns/dia de um sistema com pouco mais que 68 mil átomos. Assim, é possível observar nestas simulações movimentos biologicamente funcionais, associados a baixas frequências atingíveis apenas com o cálculo de longas trajetórias.

1.4 A análise computacional dos movimentos biologicamente funcionais de três proteínas

Poder simular longas dinâmicas moleculares em tempo plausível permite que analisemos os movimentos coletivos com importância para a função biológica que aparecem na dinâmica de proteínas alostéricas. Para estudá-los, escolhemos simular três proteínas diferentes, cada uma por uma abordagem diferente. Duas delas são reguladores transcricionais homodiméricos, cuja interface de interação com o DNA estudamos com minúcia e de forma generalizada, como comentado anteriormente. A terceira proteína que analisamos, cuja estrutura foi elucidada pelo professor Eduardo Horjales,²⁰ foi a primeira proteína a ter sua estrutura resolvida por cristalografia na América do Sul, a glicosamina 6-fosfato desaminase, ou DEA.

O primeiro regulador transcricional que descreveremos aqui é a proteína ElrR (*Enterococcus faecalis leucine-rich protein regulator*), cuja estrutura foi resolvida recentemente por Michel De Groote et al. em nosso grupo (códigos PDB 5v5t e 5v5u),²¹ por cristalografia, em sua forma apo. Seu ligante alostérico não é conhecido, nem sua conformação que o estabiliza juntamente com a ligação ao DNA. Portanto, neste projeto nos propomos a encontrar a conformação da ElrR em seu estado ligado ao DNA, que também deve ser uma conformação estável para a interação com o ligante alostérico. O ponto de partida foi o prognóstico de como deveriam se posicionar as hélices de reconhecimento nos sulcos de DNA, pela análise dos complexos TR-DNA. Calculamos então os movimentos coletivos, por diferentes métodos computacionais, necessários para levar as hélices de reconhecimento de sua posição original na estrutura apo para a posição final proposta. Esses movimentos também alteram coletivamente o sítio alostérico e com essas mudanças pudemos averiguar quais seriam os possíveis ligantes alostéricos que se encaixariam nessa conformação final.

O segundo regulador que analisaremos aqui é o inibidor transcricional MosR (*M. tuberculosis oxidation sense Regulator*). Ela é uma proteína pequena e possui duas conformações resolvidas por cristalografia, ligada e desligada do DNA, ambas em seu estado ativo. Seu estado inativo é atingido pela oxidação reversível de quatro cisteínas formando duas pontes dissulfeto intracatenárias, associadas à diminuição na afinidade pelo DNA.²² Como uma conformação nesse estado não é conhecida, buscamos simulá-la pela construção dessas duas pontes dissulfeto e averiguamos as diferenças dinâmicas e

estruturais que essa modificação causa na proteína quando comparadas com seu estado ativo, que também simulamos a partir da conformação desligada do DNA.

A terceira proteína alostérica é a DEA. Ela é a maior das três proteínas, sendo um hexâmero com 266 resíduos cada protômero, e por esse motivo ela nos permite avaliar outros dois aspectos do alosterismo: a simetria dos movimentos coletivos e a relação entre oligomerização e a coletividade dos movimentos de baixa frequência. Para isso simulamos a DEA em duas modalidades, tanto como hexâmero completo, quanto como monômero único, sendo o hexâmero estritamente simétrico durante a dinâmica pela imposição de que as interações com seus protômeros vizinhos fossem construídas a partir das operações de simetria de seu grupo pontual.

1.4.1 ElrR

A flexibilidade entre estados e a construção da conformação ligada ao DNA a partir de movimentos coletivos simulados na forma apo

A proteína ElrR é um regulador transcricional de *Enterococcus faecalis*. Ela ativa a expressão de ElrA (*enterococcal leucine-rich protein A*), que tem comprovada relação com a virulência de *E. faecalis*, bactéria comensal do trato intestinal e também um patógeno oportunista. O duplo estilo de vida dessa bactéria depende, dentre muitos outros fatores, da atividade de reguladores transcpcionais.²³ Em trabalho feito pelo grupo da professora Pascale, publicado em 2013,²³ identificou-se que a ElrR se liga especificamente à região promotora de ElrA, e as baixas expressões de ambas em laboratório ainda sugerem que sejam dependentes de condições ambientais. No entanto, não foi identificado nenhum sinal desencadeador, que ative ou expresse a ElrR e então propicie a expressão de ElrA.

Os componentes da superfamília de reguladores transcpcionais RNPP, a que foi creditada a participação da ElrR por este mesmo estudo mencionado, não possuem homologia na estrutura primária, mas grande semelhança estrutural e todos se ligam a peptídeos curtos para sua ativação. Três proteínas homólogas à ElrR são conhecidas com conformações estruturais resolvidas, e poderiam ajudar na questão de reconhecer um sinal desencadeador neste *quorum sensing* também associado a membros da família RNPP, um sistema de estímulo e resposta dependente da densidade das espécies envolvidas muito comum na regulação da expressão gênica de bactérias. Além das conformações apo conhecidas para as três, suas estruturas também foram resolvidas com algum ligante. No entanto, cada uma delas possui características únicas que impossibilitam que encontremos uma conformação por homologia para a ElrR com seu ligante alostérico. A Rgg2, cuja estrutura foi publicada em 2015, é a mais similar estruturalmente à ElrR, mas está ligada a um inibidor; a PlcR, a única das três resolvida com um peptídeo e o DNA concomitantemente, possui uma estrutura em sua forma apo muito diferente da ElrR, com

uma hélice que conecta ao domínio de ligação ao DNA muito maior; e a PrgX possui uma hélice a mais no domínio C-terminal que se desfaz para interagir com o peptídeo em seu sítio alostérico. Portanto, é provável que nenhuma das três possua um sítio alostérico como o da ElrR quando associado ao ligante.

O grupo da professora Pascale ainda identificou um possível sítio de DNA para a ElrR, e em colaboração com ela, Michel De Groote conseguiu em nosso grupo não somente resolver a estrutura da ElrR em sua forma apo por cristalografia, como também confirmou que ela se liga a esse proposto sítio de DNA por experimentos de ressonância plasmônica de superfície. Estes experimentos de SPR (*surface plasmon resonance*) mostraram uma interação de dissociação lenta, o que significa uma ligação específica como DNA, mas não foi observado um estado estacionário, o que significa que nem todos os DBDs se ligaram ao DNA. Isso indica que a ElrR seja capaz de se ligar especificamente ao sítio proposto, mas também que um ligante alostérico poderia estabilizar o complexo de forma mais eficiente. Coerentemente, esse mecanismo de ativação está de acordo com o *quorum sensing*.

Essa possibilidade é condizente com o modelo de Monod para o alosterismo,⁵ em que uma molécula livre de ligantes poderia transitar entre dois estados, e que diferentes ligantes estabilizariam um estado preferencialmente ao outro. E também se encaixa no modelo proposto por Nussinov de seleção conformacional e deslocamento populacional,²⁴ inspirado pelos conceitos desenvolvidos por Frauenfelder em 1991 sobre *energy landscapes*, em que a presença de um ligante deslocaria o equilíbrio do conjunto de proteínas para uma conformação não favorável e atingida por poucas moléculas por vez. Apesar de existirem algumas discrepâncias entre o dogma proposto por Monod em 1965 e a descrição mais atual de deslocamento populacional, ambas concordam que as diferenças estruturais verificadas entre estados diferentes só podem ser alcançadas através de movimentos coletivos.

Na conformação cristalográfica apo da ElrR, as duas unidades de DBD estão demasiadamente distantes entre si para se ligar ao DNA, mas, mantendo a coerência com os resultados experimentais e com o conceito de movimentos coletivos, é razoável esperar que a ElrR em sua forma apo possua flexibilidade para executar um movimento que leve as hélices de reconhecimento à posição de ligação ao DNA. E também que esse movimento ocorra naturalmente em solução mesmo na ausência de ligantes. Como as regiões que definem o sítio alostérico apresentam fortes contatos polares e hidrofóbicos,²¹ a movimentação coletiva dos DBDs deve ser acompanhada por essas regiões, de tal forma que a presença do ligante estabilize não só seu próprio sítio alostérico como também a conformação relativa dos DBDs ligada ao DNA. A Figura 2 indica, na estrutura da ElrR, o DBD e as duas regiões que compõem o sítio alostérico, assim como um movimento esperado das hélices de reconhecimento para que elas se liguem no DNA.

Em suma, os resultados experimentais sugerem duas conclusões que fundamentam os seguintes experimentos *in silico*. A primeira é que a ElrR em sua forma apo seja capaz

Figura 2 – a) Estrutura cristalográfica apo da ElrR representada em *cartoon* e com sua superfície. Em vermelho está indicado o DBD, em laranja uma região central e em amarelo uma região C-terminal. Prevê-se que o sítio alostérico se encontre entre essas duas últimas regiões. b) Esquema de um movimento que levaria as hélices de reconhecimento, indicadas em *cartoon*, a se encaixarem nos sulcos maiores do DNA.

Fonte: Elaborada pela autora.

de realizar um movimento que leve ao seu encaixe nos sulcos do DNA. E a segunda é que esse movimento seja coletivo e envolva tanto os DBDs quanto as regiões que formam os sítios alostéricos. Assim, encontrar esses movimentos, e a conformação que estavelmente se liga tanto ao DNA quanto a seu ligante alostérico, permitiria averiguar, pela conformação do sítio alostérico, qual ligante se encaixaria ali.

Existem alguns experimentos *in silico* que permitem calcular os possíveis movimentos coletivos de proteínas. Quatro metodologias muito populares são: a descrição por corpos rígidos; o modelo de redes elásticas (ENM); a análise de modos normais (NMA) e a análise de componentes principais (PCA). Descrever o movimento de uma proteína como sendo o movimento relativo entre algumas subunidades rígidas é o método mais simples. A dificuldade reside no reconhecimento dessas subunidades e a necessidade de pelo menos duas conformações conhecidas. Já os métodos de análise por redes elásticas ou modos normais propõem que a proteína esteja sujeita a um potencial harmônico, e para isso necessitam apenas uma conformação dela. A diferença reside na descrição das interações atômicas. Em redes elásticas os átomos, ou mesmo grupo de átomos, interagem por potenciais elásticos, como se estivessem conectados por molas aos seus vizinhos. Em modos normais essa descrição é mais detalhada, ela adota o potencial atomístico de um campo de forças, mas exige uma minimização *a priori*. O quarto método, por componentes principais, também descreve o movimento como uma transição harmônica. A vantagem

dele é não assumir que a proteína esteja presa a um único poço de potencial, mas que ela transite entre poços com conformações mais estáveis. A dificuldade está na amostragem dessas conformações, limitadas às encontradas durante uma ou mais dinâmicas moleculares. Com dinâmicas mais longas é possível não só aumentar o número de conformações visitadas, como também de estados visitados com menor frequência.

Os campos de força são atualmente muito próximos da realidade, mas o tempo de simulação ainda é um limitante, já que alguns movimentos de importância biológica estão previstos para ocorrer na faixa de centenas de nanossegundos a segundos.¹⁸ Para contornar o problema do tempo também existem simulações guiadas. É possível colocar uma conformação alvo para uma dinâmica ou puxar um grupo de átomos em determinada direção ou simplesmente dar preferência a uma direção no espaço conformacional. Em todos eles, existe a imposição de uma força externa que guie o movimento da proteína no sentido da transição que se deseja observar.²⁵ Uma diferença destes métodos em relação àqueles que descrevem uma transição harmônica, é que estes possibilitam a observação de movimentos locais ao longo do movimento coletivo, além de permitir encontrar estados de transição. Todos os métodos, no entanto, podem levar a distorções irreais da proteína, seja pela força da simplificação, ou literalmente pela imposição de uma força inexistente na realidade. No decorrer desta tese, enquanto é feita a análise de um modelo conformacional da ElrR ligada ao DNA, será discutido como os diferentes métodos abordados podem acarretar distorções conformacionais.

Escolhemos cinco métodos distintos para encontrar este modelo da ElrR. Como nosso interesse maior está em encontrar uma conformação final, e não o caminho da transição entre estados, os métodos de dinâmica guiada foram preteridos. Utilizamos, afinal, a descrição por corpos rígidos, por redes elásticas, por modos normais, pela simulação de uma dinâmica molecular não perturbada e pelo filtro de movimentos de baixa frequência dessa mesma dinâmica molecular.

Este último método, assim como a análise de modos normais, calcula uma frequência associada aos modos harmônicos obtidos. Mas, assim como na análise de componentes principais, os modos harmônicos são recuperados de uma dinâmica molecular. Ele é baseado na inversão de domínios pela transformada de Fourier, uma ferramenta matemática aplicável a inúmeras análises, como a reconstrução de imagens em criomicroscopia eletrônica de proteína. Como uma dinâmica molecular fornece uma trajetória da proteína como uma série de coordenadas em função do tempo, a transformada de Fourier decompõe essa trajetória em função de suas frequências. Sua aplicação para a recuperação de movimentos proteicos ocorrendo em diferentes janelas de tempo foi proposta por Osguthorpe em 1989.²⁶ No entanto, este método não se popularizou tanto quanto a análise por componentes principais, proposta por Almadei em 1993,²⁷ e nem tanto quanto a análise por modos normais proposta em 1983 por Go e Karplus separadamente.^{28,29}

Neste projeto de doutorado, adaptamos o método de filtro de frequências pela transformada de Fourier para analisar os movimentos coletivos das três diferentes proteínas que nos propomos estudar, mas cada análise recebeu um enfoque diferente, podendo ressaltar distintas qualidades deste método.

Para encontrar uma conformação da ElrR que fosse estável ligada ao DNA procuramos por movimentos, obtidos pelos métodos citados acima, que obrigatoriamente levassem os DBDs a uma posição relativa que permitisse o encaixe deles em dois sulcos consecutivos do DNA. Encontramos essa posição relativa pela comparação com outros complexos de regulador transcricional-DNA que analisamos em outra etapa deste projeto descrita anteriormente, considerando não só a estrutura da ElrR, mas também a sequência de bases do seu sítio de DNA. Já a qualidade dos cinco modelos obtidos foi primeiramente julgada pelas distorções nas estruturas provocadas por cada método escolhido e pela aproximação das hélices à posição que propomos, e depois pela estabilidade das novas interações encontradas em seus sítios alostéricos.

1.4.2 MosR

A regulação da afinidade pela flexibilidade conformacional dentro de cada estado

O bacilo de Koch, ou *Mycobacterium tuberculosis*, agente causador da tuberculose, vive dentro de fagossomos no pulmão de seu hospedeiro. Ele sobrevive nesse ambiente nocivo graças a inúmeros mecanismos de defesa que previnem a lise do fagossomo. Contanto que o sistema imune de seu hospedeiro não esteja prejudicado, as defesas do bacilo são suficientes apenas para equilibrar os ataques do fagossomo, e assim ele poderia coexistir em seu hospedeiro sem que os sintomas da tuberculose aparecessem nenhuma vez.³⁰

Esses mecanismos de defesa dependem de um sistema muito sensível a mudanças no ambiente, e capaz de uma pronta resposta. Neste cenário, a MosR (*M. tuberculosis oxidation sense Regulator*) possui um papel indispensável.²² Ela é um repressor transcricional sensível a oxidações no meio e regula a expressão de uma oxirredutase a ser secretada, que por sua vez afeta o pH no interior do fagossomo. Por esta função ela é caracterizada dentro da família MarR (*Multiple antibiotics resistance Regulator*), que inclui proteínas responsáveis pela resposta bacteriana a diferentes agentes como antibióticos e estresse oxidativo, e pelo controle da expressão de fatores de virulência e do catabolismo de componentes aromáticos presentes no ambiente.³¹ As proteínas dessa família também partilham de semelhanças estruturais, e são, assim como a MosR, homodímeros de forma triangular, com um domínio de dimerização que conecta os dois DBDs. Estes domínios de ligação ao DNA são do tipo *winged helix*, que apresentam, logo após o HTH, um par de fitas β , cuja volta de resíduos que as une é capaz de interagir especificamente com bases dos sulcos menores

no DNA. Cada um desses domínios se conecta diretamente ao domínio de dimerização por uma longa hélice α de sua própria cadeia, e o ponto de encontro entre essa hélice e o DBD funciona como uma dobradiça. Por outro lado, cada DBD também interage com o N-terminal da outra cadeia, como mostra a [Figura 3](#), caracterizando uma região conservada no espaço tridimensional das proteínas da família, onde se forma um bolsão para a ligação de pequenas moléculas, como o inibidor salicilato.³² Estes ligantes atuam alostericamente, promovendo ou estabilizando uma mudança conformacional relativa dos dois DBDs, que impede o encaixe de ambos ao sítio de DNA. Outra variedade de ligantes atua apenas transitoriamente, mas gera uma modificação covalente, como uma ponte dissulfeto, que também causa a diminuição da afinidade pelo DNA por mudanças conformacionais entre os DBDs.³³

Figura 3 – a) Estrutura da MosR em sua forma reduzida e ligada ao DNA representada em *cartoon*. As cisteínas que formam uma ponte de dissulfeto em ambiente oxidativo estão representadas em *stick*. O DBD está colorido em *rainbow*. b) superposição em outra perspectiva das duas estruturas cristalográficas da MosR, ligada e não ligada ao DNA, de códigos PDB 4fx4 e 4fx0, respectivamente. A estrutura ligada está colorida como em a), enquanto que a estrutura não ligada ao DNA está colorida em lilás.

Fonte: Elaborada pela autora.

A MosR possui ligantes dentro desta última variedade, e, assim como sua homóloga OhrR,³⁴ sofre a formação de duas pontes dissulfeto em ambiente oxidativo, situação que prejudica sua interação específica com o DNA e permite a expressão da oxirredutase. Para sua homóloga OhrR, as mudanças conformacionais necessárias para estabilizar as pontes de dissulfeto são grandes, já que as cisteínas reativas se encontram a 15,5Å em ambiente não oxidativo.³⁴ Essas duas cisteínas encontram-se em cadeias diferentes, uma está próxima a um N-terminal e a outra na longa hélice α que conecta um DBD. Com a formação da ponte essa hélice se quebra em duas, levando a um grande rearranjo do DBD, incluindo a rotação das hélices de reconhecimento, o que impede que se encaixem em dois sulcos maiores consecutivos no DNA.

Na MosR, as mudanças que a ponte dissulfeto causa não são tão óbvias, já que ambas as cisteínas se encontram na mesma cadeia, próximas ao N-terminal e separadas por apenas um resíduo, como pode ser observado na [Figura 3](#). Duas conformações da MosR foram resolvidas por cristalografia para elucidar esse mecanismo.²² No entanto, nenhuma das duas apresenta as pontes dissulfeto, e estão ambas em estado reduzido, uma ligada ao DNA e a outra em sua forma apo. Daqui em diante, no decorrer desta tese serão usadas as designações reduzida e oxidada, e estados reduzido e oxidado, para a ausência ou presença das pontes de dissulfeto na MosR respectivamente.

Ainda que a conformação apo esteja reduzida, nesse estado ativo a proteína pode estar dinamicamente transitando entre diferentes populações conformacionais, aptas ou não a se ligar no DNA. Assim, é provável que a forma apo não só apresente diferenças conformacionais em relação a forma holo, ligada ao DNA, como se assemelhe a uma conformação oxidada. As diferenças entre essas duas estruturas resolvidas devem, portanto, indicar como a oxidação das cisteínas afeta a afinidade da proteína pelo DNA.

Assim como em sua proteína homóloga, OhrR, as hélices de reconhecimento da MosR estão rotadas, aproximadamente 25°, quando comparadas as formas apo e holo. Essa rotação é acompanhada de diferenças também em todo o DBD, relativas ao domínio de dimerização, indicando que todo o DBD se mova como um corpo rígido. Outra diferença interessante está no comprimento do N-terminal observado na cristalografia. Na forma apo estão atribuídas posições atômicas a partir do resíduo 6, enquanto que na forma holo somente a partir do resíduo 9. Não é incomum que as extremidades da proteína não estejam inteiramente resolvidas, mas é aparentemente contraditório que na forma apo, mais livre para se movimentar, o N-terminal apresente menor flexibilidade que na forma holo, que possui seus movimentos restritos pelo DNA. Isso pode indicar que a flexibilidade do DBD e do N-terminal estejam de alguma forma relacionadas.

Também foi observado que os diferentes estados possuem distintas afinidades pelo sítio de DNA, como comprovam os experimentos de eletroferose da MosR com DNA, em que se detecta a dissociação do complexo proteína-DNA conforme a concentração de oxidantes se eleva. Também com experimentos de RT-PCR in vivo, em que se monitora a quantidade de RNA mensageiro da oxirredutase, e os maiores níveis de RNAm foram encontrados na adição de H₂O₂.²² As estruturas resolvidas por cristalografia também indicam que a oxidação esteja relacionada a mudanças conformacionais, o deslocamento dos DBDs, que restringem a interação com o DNA, ainda que não se verifiquem as pontes dissulfeto na conformação apo. Assim, nosso intuito ao analisar computacionalmente a MosR dentro deste projeto de doutorado é verificar se a forma apo reduzida é de fato capaz de passar por grandes mudanças conformacionais que possibilitem sua interação específica com o DNA, e que restrições as pontes dissulfeto provocam na estrutura e na flexibilidade a ponto de inibir a ligação ao DNA.

Para isso, simulamos duas dinâmicas moleculares de 400 ns para cada estado da proteína, reduzido ou oxidado. Como o N-terminal poderia ser peça chave na movimentação da proteína, incluímos os cinco resíduos faltantes na conformação apo cristalográfica. Testamos variadas posições iniciais desses resíduos em 10 conformações diferentes, modeladas manualmente considerando a superfície eletrostática utilizando o programa Coot.³⁵ Depois de minimizar e equilibrar essas estruturas, permitindo somente a flexibilidade no N-terminal, escolhemos duas conformações iniciais mais estáveis para dar início às dinâmicas da proteína inteira. O estado oxidado para cada uma dessas conformações foi obtido apenas substituindo um rotâmero da cisteína 12 por outro mais adequado para a formação da ponte dissulfeto.

Analizamos nessas dinâmicas a aproximação das hélices de reconhecimento à posição em que se encontram na conformação cristalográfica holo, além de acompanhar as orientações relativas dessas hélices, que descrevemos com três parâmetros de distância e ângulo. Também medimos as distâncias entre alguns resíduos mais significativos, como aqueles da hélice de reconhecimento que interagem especificamente com bases do DNA, ou aqueles descritos por Brugarolas que distinguem as conformações apo e holo. E a influência do N-terminal foi monitorada procurando os contatos intermonoméricos polares e apolares que apareciam com mais frequência em cada dinâmica.

A verificação quanto à posição das hélices de reconhecimento é especialmente importante porque dela depende a especificidade da ligação com o DNA. Interações inespecíficas são imprescindíveis para a aproximação da proteína ao DNA e são sempre possíveis, independente da conformação da proteína, pela existência de muitos resíduos carregados positivamente na interface de interação. Entretanto, para que haja o reconhecimento do sítio de DNA é preciso que as hélices de reconhecimento, ou seja, a segunda hélice do HTH de cada DBD, se posicionem favoravelmente ao encaixe em dois sulcos maiores consecutivos do DNA. Essa posição depende não somente das cargas dos resíduos, mas de uma geometria que seja compatível com a da hélice dupla fita do DNA, como discutimos neste projeto ao analisar uma variedade de complexos TR-DNA. De fato, as hélices de reconhecimento são as primeiras a interagir especificamente com o DNA como demonstra o experimento feito por Kalodimos et al.⁸ Nesse experimento de RMN de troca entre N e H, é possível encontrar um fator de proteção para cada resíduo, que indica quão rapidamente um resíduo se torna protegido pelo solvente. Kalodimos usou esse método para analisar a interação entre o repressor Lac e seu sítio de DNA, considerando resíduos que se tornam protegidos somente no estado ligado ao DNA. Ele encontrou que um resíduo crucial no N-terminal da segunda hélice do HTH, a hélice de reconhecimento, se torna protegido duas vezes mais rapidamente que outro resíduo também no N-terminal, mas na primeira hélice do HTH. E a proteção desse resíduo crucial ainda é dez vezes mais rápida que a de um resíduo fora do HTH que interage especificamente com o sulco menor do DNA.

Além disso, Kalodimos também ressalta a importância da combinação entre movimentos internos rápidos e lentos na regulação da proteína, em outro trabalho que discutimos anteriormente. Assim, em nosso projeto também analisamos a flexibilidade da proteína em cada uma das simulações, medindo o parâmetro de deslocamento atômico isotrópico e anisotrópico, além de verificar o comportamento somente dos movimentos lentos sob distintas condições encontradas nas diferentes trajetórias simuladas, usando, mais uma vez, o filtro de frequências pela transformada de Fourier.

1.4.3 DEA

Um estudo sobre a simetria dos movimentos coletivos

A glicosamina 6-fosfato desaminase (DEA) é uma enzima que catalisa uma reação química única e reversível de isomerização e desaminação de D-glicosamina 6-fosfato, resultando em D-frutose 6-fosfato e amônia. Esses açúcares aminados são importantes fontes de carbono e nitrogênio para a bactéria *E. coli*, mas também constituem a parede celular bacteriana. Assim, o catabolismo deles feito pela DEA nem sempre é desejável, e é necessária uma fina regulação de sua atividade. Não surpreendentemente, a DEA é uma enzima alostérica, e possui como ativador o açúcar aminado N-acetilglicosamina 6-fosfato. Ele se liga exclusivamente à forma R, ativa, da DEA, e na ausência dele existem 10 mil moléculas de proteína na forma T, inativa, para cada uma na forma R. E ainda que seu substrato não se ligue exclusivamente à forma R, sua afinidade por esta é 50 vezes maior que pela forma T, como indicam os ensaios de ativação feito por Altamirano et al..³⁶ Ademais, as curvas encontradas nesses experimentos se ajustam ao modelo MWC de alosterismo (Monod-Wyman-Changeux), tanto na ausência quanto na presença do ativador. Isso indica que tanto uma molécula de ativador quanto uma de substrato podem alostericamente ativar a conformação R da proteína, hetero ou homotropicamente, respectivamente.

O ajuste ao modelo MWC indica que a afinidade da proteína pelo ligante, não varia com o número de ligantes interagindo com ela, mas depende somente do estado quaternário em que ela se encontra, T ou R.³⁷ Assim, a fração de proteína em solução no estado R é necessariamente proporcional à fração de proteína com o sítio ocupado por um ligante, exceto no caso de ligações não exclusivas.³⁸ O modelo MWC também é chamado de modelo concertado, pois estas características se adequam a uma transição concertada da proteína entre seus estados T e R. Ou seja, os protômeros da proteína se movem simetricamente de forma que estejam todos sempre em estado R ou T, e basta apenas a interação de um ligante com um protômero para que todos os outros se estabilizem na conformação R.

Em oposição a esse modelo, existe o modelo sequencial, baseado na hipótese de encaixe induzido, que explica a cooperatividade pelas mudanças conformacionais induzidas por um protômero já ligado sobre um protômero vizinho.³⁷ No limite em que as interações

entre protômeros sejam muito acopladas, basta que um protômero mude do estado para que todos os outros também mudem. Mas este modelo não prevê que as mudanças conformacionais possam ocorrer mesmo na ausência do ligante, como descrito no modelo de Monod.

Atualmente, tanto movimentos locais quanto coletivos são compreendidos como intrínsecos às proteínas. Essa flexibilidade é fundamental para que elas exerçam suas funções. Monod ainda supôs que o eixo de simetria de proteínas oligoméricas também fosse um eixo que coordenasse a movimentação da proteína, principalmente descrita por rotações em torno dele mesmo. Assim, neste modelo, a simetria da estrutura está intimamente relacionada com a dinâmica da proteína, e Monod lista uma série de declarações para caracterizar proteínas alostéricas nesse âmbito.³⁹ Elas definem que: proteínas alostéricas são oligoméricas com ao menos um eixo de simetria; cada molécula ligante corresponde somente a um sítio em cada protômetro, que se dispõem na proteína com a mesma simetria; a conformação de cada protômero está constrita pelas interações com seus vizinhos; essas proteínas possuem dois estados reversivelmente acessíveis, e as interações entre protômeros são diferentes em cada um dos estados; a afinidade por ligantes também é diferente em cada estado; e, por fim, a simetria é conservada durante a transição entre estes estados.

A DEA não só possui atividade e ativação que se encaixam no modelo MWC³⁶ como também possui as características estruturais listadas acima. Ela é um hexâmero, com pelo menos um eixo de simetria que se mantém em seus dois estados. Como ambos foram resolvidos por cristalografia,^{20,40} pôde-se constatar que as interações entre protômeros também são diferentes em cada estado. A Figura 4 mostra tanto as conformações T e R da DEA, e ilustra a transição alostérica que ocorre entre elas. Ademais, ela possui um sítio ativo e um sítio alostérico por protômero que também se dispõem sob a mesma simetria molecular que os protômeros. Contudo a última declaração, sobre a simetria da transição, ainda não pôde ser comprovada para a DEA, pelo contrário, nossos estudos prévios mostraram que a simetria entre hexâmeros não se mantém em sua dinâmica.⁴¹

Será que esta definição de Monod é de fato razoável? Será que a simetria poderia se manter mesmo para movimentos locais de alta frequência que não estão imediatamente constritos pela associação dos protômeros? Ou será que essa afirmação só se mantém sob a luz do teorema ergódico, como uma conclusão sobre a média das conformações num ensemble? E se a simetria fosse rigorosamente mantida, a flexibilidade de cada protômero se alteraria? Ou será ainda que a simetria não se sustenta durante as transições, mas é válida somente para os estados inicial e final da transição? E como será para uma transição quase estática?

Estas questões são o mote para as análises com a DEA neste projeto de doutorado, e para abordá-las iniciamos duas dinâmicas moleculares distintas. Uma simulação *all-atom* com solvente explícito do hexâmero livre. E a segunda também *all-atom* com

Figura 4 – Esquema da transição alostérica entre os estados T e R da DEA, cujas estruturas cristalográficas possuem código de PDB 1fsf e 1dea, respectivamente, e estão desenhadas em *cartoon* com cada cadeia colorida de uma cor. Duas perspectivas são apresentadas: uma em cima e outra embaixo. Entre as conformações T e R estão representadas as mudanças conformacionais observadas indicadas por vetores para cada carbono α substituindo uma das cadeias, enquanto as demais estão representadas em *trace*.

Fonte: Elaborada pela autora.

solvente explícito, mas de apenas um protômero enquanto os outros cinco são calculados simetricamente, com as cinco operações de simetria de seu grupo espacial cristalográfico, como imagens do primeiro protômero.

2 OBJETIVOS

Este projeto procura por meios computacionais que auxiliem na compreensão dos movimentos coletivos que as proteínas são aptas a fazer e como eles contribuem para seu funcionamento. Para isso selecionamos alguns sistemas e buscamos por parâmetros que possam descrevê-los no espaço conformacional e também por meios de detectar as transições entre seus estados conformacionais.

- Encontrar parâmetros geométricos e de sequência primária que auxiliem na classificação e organização de grupos proteicos de DBDs contendo o motivo HTH;
- Desenvolver método de extração de movimentos coletivos de dinâmica molecular que sejam diretamente relacionados com a frequência com que ocorrem, para então aplicá-lo no estudo da dinâmica de três proteínas diferentes;
- Construir modelos razoáveis da ElrR ligada ao seu sítio de DNA, usando somente movimentos coletivos e a combinação deles, para então inferir sobre possíveis ligantes alostéricos;
- Encontrar diferenças estruturais e dinâmicas entre os dois estados, reduzido ou oxidado, da MosR, que justifiquem suas distintas afinidades pelo sítio de DNA;
- Avaliar os efeitos da imposição de simetria sobre a dinâmica da DEA.

3 BASE TEÓRICA E METODOLOGIA

3.1 O motivo HTH em Domínios de Ligação ao DNA

O motivo HTH (*helix-turn-helix*) foi reconhecido pioneiramente em 1982 por Sauer et al.⁴² em algumas das primeiras estruturas de fatores transcricionais resolvidas, quando também identificaram seu papel crítico na ligação ao DNA. Nestas proteínas, o HTH se encontrava atrelado a outras estruturas secundárias que em conjunto formavam um domínio de ligação ao DNA, denominado de domínio HTH. Ele recebeu esse nome, apesar da presença das outras estruturas secundárias, porque algumas características do motivo HTH eram muito marcantes e pouco variáveis, como: a direta e específica interação da segunda hélice do HTH com o DNA, também denominada de hélice de reconhecimento; o ângulo agudo que ela forma com a hélice anterior; e o comprimento da volta que as conecta.

Em termos de sequência primária, os HTH apresentam grande variabilidade, o que é justamente a causa para sua especificidade. Até mesmo o tamanho das hélices é variável, mas com apenas os 20 resíduos centrais, como historicamente definido, já se representa bem sua estrutura tridimensional bastante conservada.⁶ Apesar dessa geometria específica para o encaixe da segunda hélice no sulco maior do DNA, acompanhada do N-terminal da primeira hélice apontando para os fosfatos do DNA em oposição a sua carga, não existe nenhuma regra simples sobre como se dá este encaixe.⁴³ Inclusive o ângulo entre a hélice da proteína e a dupla hélice de DNA também é muito variado, assim como as interações entre resíduos e bases, sejam por ligações de hidrogênio ou interações de van der Waals, que tampouco seguem algum código.⁴⁴ Ademais, a segunda hélice do HTH não é a única capaz de interações específicas, em alguns casos outras estruturas dentro do domínio HTH também reconhecem bases específicas principalmente interagindo com os sulcos menores vizinhos ao sulco maior em que se encaixa a hélice de reconhecimento,⁴⁵ aumentando a especificidade de cada regulador.

Essas variadas estruturas secundárias que aparecem atreladas ao HTH dão origem a diversas versões do domínio HTH, compostas de famílias distintas. A mais simples das versões é constituída por somente três hélices, sendo as duas últimas constituintes do motivo HTH. Como explicado por Aravind,⁴⁶ essa terceira hélice forma uma fenda com o HTH, numa conformação aberta, que tende a ser compensada por outras estruturas. Uma outra versão do domínio é chamada de *winged helix*, ou wHTH, que apresenta uma folha beta interagindo hidrofobicamente com esta fenda e ainda é capaz de interagir especificamente com o DNA.

Dentro de cada família existe homologia e também alguns resíduos conservados que

interagem com certos trechos curtos de bases de DNA, como é o caso das homeoproteínas com o homeodomínio, uma região de 60 resíduos muito conservados e que compreende um HTH. Todas estas proteínas apresentam uma asparagina na posição 51, importante para reconhecer trechos do DNA ricos em AT. Logo, esta asparagina não é responsável pela especificidade de cada proteína do grupo.⁴⁷ Já na posição 50 pode estar presente ora uma glutamina, que reconhecerá as sequências TAATTA ou TAATTG, ora uma lisina que reconhecerá a sequência TAATCC. Ainda dentre aquelas com uma glutamina estão as proteínas Hox, que também conseguem interagir com o sulco menor. Portanto, é a composição de um conjunto de pequenas características que garante a especificidade entre DBD e DNA. Outro exemplo de estudo de especificidade foi a construção de quimeras de HNF-3 β .⁴⁸ Trocando os 20 resíduos que precedem a hélice de reconhecimento pelos mesmos da HFH-1 foi possível que a proteína reconhecesse o sítio da HFH-1. Portanto, algumas regras de reconhecimento podem ser identificadas, mas elas estão restritas não só a uma mesma versão do domínio HTH, como a uma mesma família, e tampouco existem poucas famílias. Estima-se que o último ancestral comum universal já possuísse de 6 a 11 versões distintas do domínio HTH.⁴⁶

Apesar de alguns exemplos de interações conservadas entre resíduos e bases, não existe nenhum conjunto de regras simples que as reja. Ainda assim, algumas tentativas foram feitas nesse sentido e foram encontradas predominâncias de interações, por exemplo entre argininas e guaninas,⁴⁹ independente da família da proteína ou do DBD. Essa predominância acontece por vários motivos: a guanina é a base que mais apresenta átomos susceptíveis a ligações de hidrogênio expostos tanto no sulco maior do DNA quanto no sulco menor; a arginina pode participar em mais de uma ligação de hidrogênio; e por último, é muito frequente que um mesmo resíduo possa interagir com mais de uma base ou par de bases e vice-versa, denominadas de interações complexas.⁴⁹ Esta última característica não só dificulta a generalização dessas interações, como também aumenta muito a especificidade de cada proteína. Além das ligações de hidrogênio, foram encontradas muitas interações de van der Waals, que de fato contribuem com dois terços de todas as interações consideradas no estudo em questão.⁴⁹ Mas para as interações de van der Waals existe uma variedade de combinações de resíduos e bases mais frequentes, ao contrário do que acontece com as ligações de hidrogênio.

Variabilidade também é encontrada na deformação da dupla hélice de DNA. Essa deformação pode ser justificada principalmente pela sequência de bases, considerando como elas se arranjam ao longo do eixo do DNA e em que concentração se encontram em determinada região, num sulco maior ou menor.⁴⁷ Mas uma associação também pode ser feita quanto à função da proteína. Aquelas relacionadas à catálise geralmente envolvem o DNA com uma rede de estruturas secundárias e *loops*, causando grandes distorções,⁴⁵ enquanto as proteínas que apenas se ligam ao DNA se detêm em apenas uma das faces do DNA e se encaixam predominantemente nos sulcos maiores.⁴⁵ Proteínas com o domínio

HTH, geralmente caem nesta última classificação, e normalmente não são enzimas, mas reguladores transcricionais. Portanto, as deformações no DNA não são grandes, mas também variam principalmente de acordo com as bases centrais do sítio de ligação, que se encontram em um sulco menor, já que a grande maioria forma homodímeros, como já explicado.

Essa homodimerização implica não só em uma susceptibilidade maior do centro do sítio de DNA a diferentes curvaturas, como também em certa palindromia em toda a extensão do sítio. A palindromia aparece porque as interações base-resíduo são, com poucas exceções, as mesmas para os dois domínios HTH que se encaixam no DNA, sejam para ligações de hidrogênio ou para interações de van der Waals, as deste último tipo com menor rigorosidade. Mas como a direção de encaixe de cada hélice de reconhecimento é distinta, uma fita de DNA diferente se aproxima da hélice de reconhecimento, como ilustra a [Figura 5](#).⁵⁰

Figura 5 – Exemplo de duas hélices de reconhecimento, de mesma sequência, interagindo simetricamente com dois sulcos maiores vizinhos no DNA. Cada hélice de proteína faz as mesmas interações, mas com uma fita diferente do DNA. O exemplo mostra o complexo de código PDB 4fx4.

Fonte: Elaborada pela autora.

Mesmo que a palindromia apresente certa flexibilidade, ela ainda pode ser detectada e por vezes é denominada de pseudo-palindromia, por aparecer com algumas falhas. Isso acontece justamente pela natureza de cada interação, que varia entre as famílias. Como já

mencionado é um conjunto de características que determina a especificidade do complexo regulador transcricional-DNA (TR-DNA). Kalodimos identificou em seu trabalho com o *Lac headpiece*⁸ que algumas interações acontecem de forma sequencial, e concluiu que a hélice de reconhecimento deveria se ligar ao sulco maior antes que outros resíduos interagissem com o sulco menor. Em especial, ele encontrou que a glicina 18 do *Lac headpiece* é crucial para a interação com o DNA e sua ligação acontece muito mais rapidamente que outras. A posição dessa glicina é especial. Ela se encontra no N-terminal da hélice de reconhecimento e, quando ligada ao DNA, ela é a que mais se aproxima do fundo do sulco maior, como se estivesse “encarando” essas bases do fundo.

Neste trabalho consideramos que essa posição seja de fato particular e crucial para a interação dos reguladores transcricionais estudados e seus respectivos sítios de DNA, e ao resíduo nesta posição denominamos de resíduo principal. Assim, levantamos uma série de parâmetros relacionados a ele e à orientação de sua hélice, considerando sempre a existência de dois HTHs simétricos, para fazer uma classificação geométrica que pudesse auxiliar, por exemplo, a encontrar uma posição de encaixe mais provável para os HTHs de outros reguladores transcricionais. Alguns parâmetros mostraram-se mais informativos e relevantes que outros, mas a lista completa deles e sua definição matemática encontra-se a seguir, assim como a definição de objetos necessários ao cálculo. A [Figura 6](#) auxilia na definição desses parâmetros.

- eixo da hélice de reconhecimento: o eixo o autovetor que compõe o tensor de inércia da hélice, com menor autovalor associado. O eixo aponta no sentido do C-terminal da hélice. O tensor de inércia é calculado considerando-se somente os átomos pesados da cadeia principal;
- ponto principal da hélice de reconhecimento: a projeção das coordenadas do carbono α do resíduo principal sobre o eixo de sua hélice.
- eixo do DNA: a composição de centros de massa ao longo da hélice dupla fita. Cada centro de massa é calculado considerando todos os átomos de uma volta de uma fita da hélice de DNA, compondo 10 bases. Quando é necessário considerá-lo como vetor, está definido como o vetor diferença entre dois desses centros de massa, escolhidos convenientemente.
- ângulo de simetria, ϵ . Verificamos se as hélices de reconhecimento de fato obedece a simetria de ordem 2 pela superposição de ambas. Da matriz de superposição obtemos o eixo de simetria, equivalente ao eixo de rotação, e o ângulo de simetria, equivalente ao ângulo de rotação.⁵¹
- β , ângulo entre as hélices de reconhecimento. Calculado diretamente como o produto escalar entre os eixos das duas hélices de reconhecimento. É definido pelo ângulo

que se forma voltado para o interior da proteína.

- D_p , distância entre os pontos principais das duas hélices de reconhecimento.
- D_h , distância entre as hélices. Calculada como a projeção do vetor diferença dos pontos principais sobre o vetor ortogonal às duas hélices de reconhecimento, resultante do produto vetorial entre os vetores que definem seus eixos.
- δ , menor ângulo formado entre o vetor ortogonal às duas hélices (de módulo D_h) e o vetor diferença entre os dois pontos principais (de módulo D_p). Calculado diretamente pelo produto escalar entre esses dois vetores.
- D_{ph} , distância de cada ponto principal ao eixo do DNA. Aqui o eixo de DNA é considerado como o vetor diferença entre os dois centros de massa de uma volta mais próximos do ponto principal.
- λ , menor ângulo formado entre o eixo do DNA e o vetor diferença entre os dois pontos principais (de módulo D_p). Aqui o eixo do DNA é considerado como o vetor diferença entre os centros de massa de uma volta mais distantes.
- α , menor ângulo formado entre o eixo de uma hélice de reconhecimento e o vetor diferença entre os dois pontos principais (de módulo D_p).
- γ , menor ângulo formado entre o eixo de uma hélice de reconhecimento e o eixo do DNA. Aqui o eixo do DNA é considerado como o vetor diferença entre os centros de massa de uma volta mais distantes.
- tamanho da hélice de reconhecimento em número de resíduos.
- tamanho do sítio de DNA em número de bases. Definido pelo número de bases entre as bases que interagem com o resíduo principal.
- estatísticas de resíduos e bases interagentes. Somente interações polares foram consideradas. Foram quantificadas as interações entre cada resíduo e bases específicas, além da quantidade de interações por resíduo e por sua posição na hélice de reconhecimento, e por base no sulco maior ou menor.
- curvatura do DNA. Avaliada qualitativamente pela observação da distribuição dos centros de massas de uma volta calculados para o DNA.

Cálculos semelhantes já foram feitos, tanto para a estatística de interações quanto para a geometria do encaixe das hélices de reconhecimento no sulco maior.^{43,49} Inclusive a amostragem de complexos abordados nesses estudos é maior que a considerada neste projeto de doutorado, mas tampouco é específica para o grupo de proteínas com o domínio HTH, assim como a avaliação geométrica do encaixe da hélice não considerava a simetria

Figura 6 – Duas perspectivas das hélices de reconhecimento e ligadas ao DNA de um complexo TR-DNA (código PDB 4fx4) para ilustrar os parâmetros levantados para análise. O resíduo principal está representado em CPK em branco. O ponto principal de cada hélice de reconhecimento está representado em preto. O cilindro em preto une os dois pontos principais. Os vetores em preto representam os eixos das hélices de reconhecimento. O vetor em laranja representa o vetor ortogonal aos eixos das duas hélices de reconhecimento. Os pontos em vermelho representam os centros de massa de cada volta do DNA. O vetor em vermelho representa o eixo do DNA usado nos cálculos de λ e γ . O vetor em azul representa o eixo de DNA usado no cálculo de D_{ph} , e o ponto em azul é a projeção do ponto principal sobre este vetor.

Fonte: Elaborada pela autora.

dessas proteínas, como fazemos aqui. Portanto a abordagem que fazemos é diferenciada e aplicada a um conjunto diferente de proteínas, mas ficou limitada a somente 27 complexos selecionados dentre as 59 estruturas depositadas no PDB até 2014 correspondentes à busca por reguladores transcricionais ligados a DNA. Dessas, 39 são homodiméricas com o domínio HTH e apenas 27 realmente possuem um sítio de ligação no DNA que apresenta certa simetria, ou palindromia, além de excluídas estruturas que representassem o mesmo regulador transcricional. Atualmente, muitas outras proteínas poderiam ser incluídas neste estudo. O crescente número de estruturas resolvidas só tende a melhorar as conclusões que podemos retirar, mas dificulta a análise detalhada de cada uma delas. Outros estudos têm sido feitos com o intuito de auxiliar nesta etapa de classificação de tais complexos, como o de AlQuraishi et al.,⁵² que levanta 64 DBDs com HTH ligados a DNA e provê informações estatísticas quanto às interações atômicas e como elas se distribuem.

3.2 O estudo de movimentos coletivos

No mesmo estudo em que Kalodimos identifica a associação sequencial de alguns resíduos com o DNA, ele também discute que uma região desordenada do *Lac headpiece* se torna ordenada após a associação do HTH no sulco maior do DNA.⁸ Ainda que alguns reguladores transcricionais dependam do DNA para se ordenar e estabilizar sua estrutura quaternária, muitas outras proteínas alostéricas, com estrutura apo ordenada, podem passar por essa transição de conformação na ausência dele como já discutimos na introdução.

A forma mais simples de abordar e representar quantitativamente esta transição é considerando que algumas regiões da proteína se comportem como corpos rígidos, executando somente translações e rotações relativas um ao outro. Essa representação não está longe da realidade, visto que alguns motivos estruturais são reconhecidos justamente por sua estabilidade tridimensional, como o HTH. E muitas proteínas têm sido representadas assim para analisar suas mudanças conformacionais biologicamente funcionais. Como é o caso da chaperona GroEl, quando passa de sua conformação aberta para fechada, formando a jaula de Anfinsen.⁵³ Ou como a DEA, que estudamos neste projeto, cuja transição entre os estados T e R pode ser definida por uma rotação da região mais externa.⁴⁰

Para dividir a estrutura de uma proteína em corpos rígidos é necessário que se conheça pelo menos duas conformações dela, para reconhecer que regiões apresentam menor desvio entre ambas. Aqui usamos duas estratégias para reconhecer estas regiões.

A primeira é por sucessivas superposições, ou alinhamentos estruturais, de diferentes regiões até que se encontrem regiões com os menores valores de RMSD (*root mean square deviation*). Calculamos essas superposições com um rotina escrita em Fortran90 usando o método de diagonalização da matriz quaternion,⁵⁴ acompanhado de um método iterativo, baseado na rotina Lsq do programa O,⁵⁵ que reescolhe resíduos da região base para o alinhamento de acordo com as distâncias dos carbonos α , de forma a convergir a uma lista de átomos pertencentes a um núcleo rígido, com menor RMSD.

O alinhamento, ou superposição, estrutural não é útil somente para o reconhecimento de regiões menos flexíveis, mas também para análise das regiões mais flexíveis durante simulações de dinâmica molecular. Esse alinhamento ressalta o movimento da região que se deseja avaliar, em detrimento de outra, cuja dinâmica de fato não passa de pequenas flutuações. Assim, a superposição estrutural local é preferida nas análises desenvolvidas neste projeto.

A segunda estratégia é pela determinação de domínios dinâmicos pelo programa *on-line* DynDom.⁵⁶ Neste método uma simples superposição estrutural é feita com as duas conformações conhecidas. Como desta superposição uma matriz de rotação é calculada, também a cada átomo designa-se um vetor de rotação. E com um algoritmo de clusterização é possível determinar grupos de átomos que passem pela mesma rotação. Esses grupos são

definidos como domínios dinâmicos.

Assim como neste último método, também é possível definir esses grupos de átomos a partir da Análise de Modos Normais (NMA). Neste método a divisão em corpos rígidos não é um resultado direto. Dele se obtém uma base de vetores que definem os movimentos coletivos intrínsecos a estrutura da proteína, e não é necessário duas conformações conhecidas para calcular essas oscilações naturais.

Todo corpo possui modos únicos de vibração, que dependem da sua distribuição de massa e das interações entre suas partes. Assim também pode ser compreendida a proteína, como um corpo composto de átomos que interagem entre si de acordo com as leis da mecânica clássica, de forma que ela está sujeita a energia potencial determinada pelas forças dessas interações. NMA aproxima a energia potencial a um poço harmônico de tantas dimensões quantos graus de liberdade da proteína (em coordenadas cartesianas, são três para cada átomo, totalizando $3N$ dimensões, onde N é o número total de átomos). Essa simplificação pode ser obtida pela expansão da energia potencial em uma série de Taylor até segunda ordem, como na [Equação 3.1](#),

$$V = V_0 + \sum_{i=1}^{3N} \frac{\partial V}{\partial q_i} \bigg|_0 q_i + \frac{1}{2} \sum_{i,j=1}^{3N} \frac{\partial^2 V}{\partial q_i \partial q_j} \bigg|_0 q_i q_j \quad (3.1)$$

onde V é o potencial a que está sujeita a proteína e q_i é uma coordenada generalizada, equivalente ao deslocamento da coordenada i de seu mínimo de energia, Δx_i , multiplicado pela raiz quadrada da massa do átomo correspondente, $\sqrt{m_i}$.

Assim, uma minimização de energia deve garantir que o termo $\frac{\partial V}{\partial q_i}$ seja pequeno o suficiente para ser desprezado frente ao termo de derivada em segunda ordem, ainda que multiplicado por q_i , e também permite anular o primeiro termo assumindo que a minimização de fato leve ao mínimo do potencial.

Encontrar os modos de vibração significa obter os autovetores e autovalores da matriz hessiana, ou seja, diagonalizar a matriz construída com as derivadas segundas do potencial V , como na [Equação 3.2](#),

$$\begin{bmatrix} \frac{\partial^2 V}{\partial q_1 \partial q_1} & \frac{\partial^2 V}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 V}{\partial q_1 \partial q_{3N}} \\ \frac{\partial^2 V}{\partial q_2 \partial q_1} & \frac{\partial^2 V}{\partial q_2 \partial q_2} & \cdots & \frac{\partial^2 V}{\partial q_2 \partial q_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 V}{\partial q_{3N} \partial q_1} & \frac{\partial^2 V}{\partial q_{3N} \partial q_2} & \cdots & \frac{\partial^2 V}{\partial q_{3N} \partial q_{3N}} \end{bmatrix} \begin{bmatrix} w_{1n} \\ w_{2n} \\ \vdots \\ w_{3Nn} \end{bmatrix} = \omega_n^2 \begin{bmatrix} w_{1n} \\ w_{2n} \\ \vdots \\ w_{3Nn} \end{bmatrix} \quad (3.2)$$

onde n é o índice para cada modo normal, a matriz explicitada é a matriz hessiana, w_n são os autovetores, que indicam a direção do movimento num espaço de $3N$ coordenadas, e ω_n são os autovalores.⁵⁷

A [Equação 3.3](#) é solução da [Equação 3.2](#), uma função harmônica no tempo para cada coordenada i de cada modo normal n .

$$q_{in} = \Delta x_{in} \sqrt{m_i} = w_{in} A_j \cos(\omega_n t + \phi_n) \quad (3.3)$$

Na [Equação 3.3](#), a amplitude, A_n , a frequência, ω_n , e a fase, ϕ_n , são independentes da coordenada i , mas diferentes para cada modo normal n . Ou seja, num mesmo modo normal, existe um deslocamento em cada coordenada de cada átomo, que ocorre com mesma frequência e fase. As amplitudes estão associadas a dada temperatura à proteína para que ela oscile em torno de seu mínimo. As oscilações de mais baixa frequência correspondem a movimentos de grande amplitude, que por sua vez correspondem a movimentos de importância biológica, relacionados a funcionalidade da proteína e coincidentes com dados experimentais, como exposto pioneiramente por Go e Karplus, separadamente em 1983.^{28,29}

NMA possibilitou grandes avanços na compreensão sobre os movimentos coletivos das proteínas e sobre a forma do potencial a que estão sujeitas. No entanto, pode ser muito custoso computacionalmente cumprir a exigência de se iniciar em um mínimo do potencial, principalmente para grandes proteínas e oligômeros. Não partir de uma conformação minimizada pode acarretar inconsistências como autovalores negativos. Ademais, a minimização tampouco é garantia de se alcançar o mínimo global, além de inconvenientemente afastar a proteína de suas coordenadas cristalográficas iniciais.

Em 1996 Tirion descreveu uma alternativa simples para esse problema, e que também gera um conjunto de movimentos coletivos bastante consistentes com os calculados por NMA.⁵⁸ Ela substituiu o potencial complexo, usado em NMA e também em simulações de dinâmica molecular, com constantes de força variadas para cada tipo de interação, por um potencial elástico cuja constante de força é a mesma para todos os pares considerados. O potencial clássico está descrito na [Equação 3.4](#). Ele é uma função da distância entre átomos ligados, b ; da distância entre átomos não ligados, r_{ij} ; do ângulo entre três átomos ligados, θ ; e do ângulo de diedro entre quatro átomos ligados ϕ ; além de depender das variadas constantes de força e de termos como a multiplicidade ou periodicidade, n , que contribui com mais uma somatória dentro da considerada para ângulos diedrais.⁵⁹

$$E_p = \frac{1}{2} \sum_{\text{ligações}} k_b (b - b_0)^2 + \frac{1}{2} \sum_{\text{ângulos}} k_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{\text{diedros}} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{pares não ligados}(ij)} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{D r_{ij}} \right] \quad (3.4)$$

O potencial elástico está descrito na [Equação 3.5](#). Nesta simplificação de redes elásticas a proteína é interpretada como um conjunto de nós conectados por molas. Na [Equação 3.5](#), o vetor conectando cada par de nós é denotado por $\vec{r}_{a,b}$ e o índice 0 indica

a conformação inicial. Tirion designou um átomo a cada um desses nós, mas outras designações podem ser feitas, como somente carbonos α ou agrupamento de átomos, dando um caráter de escalabilidade ao método. Além disso, a derivada primeira desse potencial elástico (que aparece no segundo termo da expansão de Taylor) tende a zero no equilíbrio, e como as coordenadas cristalográficas já são resultado de um equilíbrio entre as muitas moléculas do cristal, a minimização torna-se desnecessária.⁶⁰

$$E_p = \frac{C}{2}(|\vec{r}_{a,b}| - |\vec{r}_{a,b}^0|)^2 \quad (3.5)$$

Denominado de ENM (*Elastic Network Models*), este tipo de método permitiu a análise rápida de muitas proteínas e levou à compreensão de que os movimentos coletivos são em grande parte dominados pela forma da proteína, e não muito por detalhes em sua estrutura e energia.⁶⁰ No entanto, alguns exemplos de mudanças alostéricas, como as da DEA, que discutiremos adiante, ou as da Groel, dependem também de interações locais.⁶¹ Existem outros tipos de ENM inspirados pelo trabalho de Tirion, em que se variam as designações do nó, como no ANM (*Anisotropic Network Models*), ou o como no GNM (*Gaussian Network Models*), em que o quadrado da diferença das distâncias entre dois nós passa a ser o produto escalar da diferença dos vetores entre dois nós.⁶⁰

Entretanto, uma limitação que já existia em NMA e persiste nos ENMs, pela aproximação harmônica é a negligência sobre múltiplos mínimos e barreiras energéticas na superfície do potencial.⁶² Como explica Hinsen,⁶³ o mínimo da energia potencial é sempre rodeado por pequenas barreiras energéticas, relacionadas a poços estreitos, enquanto que à temperatura ambiente a proteína está sujeita a um potencial mais amplo e é pouco afetada pelas pequenas barreiras em seu fundo. São essas diferenças na largura do potencial que provocam incoerências nas frequências encontradas por NMA ou ENMs, geralmente muito mais altas que as esperadas para movimentos relacionados a função biológica.

Portanto, é atraente estudar os movimentos coletivos derivados de simulações de dinâmica molecular. Apesar de serem custosas no tempo de cálculo e dependentes da amostragem, as conformações calculadas são provenientes da passagem livre da proteína por barreiras e mínimos no potencial.

A simulação de Dinâmica Molecular consiste na resolução de forma numérica das equações de movimento de Newton para o sistema em questão, cuja função de energia é descrita pela [Equação 3.4](#). Assim, ela fornece o movimento das partículas individualmente em função do tempo, o que permite calcular muitos parâmetros com facilidade, como a distância entre grupos de átomos, ou a conservação de estruturas secundárias. Essas trajetórias permitem a análise de interações moleculares com um nível de detalhe e com a passagem do tempo impossível a outros experimentos. A escolha do passo de integração das equações de movimento ainda permite uma maleabilidade de acordo com os tempos

dos eventos que se deseja observar. No entanto, a aproximação clássica representa átomos como esferas com carga e massa exatas e não permite a observação da formação ou quebra de ligações covalentes. Também existe uma limitação temporal a depender do tamanho do sistema e da quantidade de graus de liberdade.

PCA (*Principal Component Analysis*) ou *Essential Dynamics*, proposta em 1993,²⁷ é a mais conhecida das técnicas que analisam simulações de dinâmica molecular em busca de movimentos coletivos. Como em qualquer uma de suas inúmeras aplicações, a análise de componentes principais retorna vetores que descrevem a distribuição dos objetos de estudo num espaço cujo número de dimensões é justamente a quantidade de parâmetros que descrevem esses objetos. Na análise de uma dinâmica molecular os objetos são as conformações e os parâmetros são todas as suas $3N$ coordenadas, onde N é o número de átomos considerados. Nessa distribuição, as conformações podem se agrupar de acordo com o estado que visitam durante a simulação, justamente por serem conformacionalmente distintos. Assim, os novos vetores podem fornecer uma direção e distância entre esses grupos, e uma transição entre eles pode ser representada harmonicamente.

Assim como para NMA, os vetores resultantes de PCA são autovetores provenientes de uma diagonalização da matriz de covariância, o que significa que esses vetores são ortogonais. Por este motivo movimentos relacionados não linearmente estão fracamente representados, como explicado por David e Jacobs.⁶⁴ Ademais há um alto custo de memória computacional envolvido na diagonalização, o que torna quase impossível usar estes métodos para grandes proteínas. Existem alternativas como escolher um subgrupo de coordenadas, ou encontrar outras descrições como os ângulos ϕ e ψ da cadeia principal, ou mesmo abordar de forma diferente como no *kernel PCA*.⁶⁴

O método de filtro de movimentos coletivos pela transformada de Fourier aparece como uma alternativa tanto ao problema da harmonização quanto ao da ortogonalidade. Ele interpreta diretamente as conformações calculadas por dinâmica molecular, como PCA, mas os vetores obtidos não são ortogonais. Podem-se extrair tantos movimentos harmônicos quanto o número de conformações proporcionadas à análise, e não o número de parâmetros.

O filtro de frequências como proposto por Osguthorpe em 1989²⁶ segue o desenvolvimento em série de Fourier de uma função dependente do tempo, como em análise de sinais.^{65,66} Considera-se que qualquer função do tempo discreta e periódica, $f(t)$, pode ser expandida em uma série de senos e cossenos, como na [Equação 3.6](#).

$$f(t) = \sum_{n=1}^{\infty} A_n \cos\left(\frac{\pi n t}{T}\right) + B_n \sin\left(\frac{\pi n t}{T}\right) \quad , \quad \begin{cases} t = k\delta t \\ k = 0, 1, \dots, K \end{cases} \quad (3.6)$$

A_n e B_n são determinados por suas transformadas de Fourier discretas, considerando

que nossa função $f(t)$ é a trajetória da proteína representada por um número limitado de conformações separadas pelo mesmo intervalo, δt , (Equação 3.7 e Equação 3.8), assim k é um número inteiro, entre 0 e K , o tempo t é igual a $k\delta t$, e K multiplicado por este intervalo, δt , equivale ao tempo total da trajetória, T . n também é um número inteiro relativo a um modo harmônico incluso na somatório, que podem ser infinitos. Entretanto, no tratamento numérico, com um número finito de amostras, ou conformações, não é possível obter mais que esta mesma quantidade de modos harmônicos independentes.⁶⁶

$$A_n = \frac{1}{T} \int_{-T}^T f(t) \cos\left(\frac{nt\pi}{T}\right) dt = \frac{1}{K} \sum_{k=0}^{K-1} f(t) \cos\left(\frac{\pi n \delta t k}{T}\right) \quad (3.7)$$

$$B_n = \frac{1}{T} \int_{-T}^T f(t) \sin\left(\frac{nt\pi}{T}\right) dt = \frac{1}{K} \sum_{k=1}^{K-1} f(t) \sin\left(\frac{\pi n \delta t k}{T}\right) \quad (3.8)$$

Em análise de sinais, um filtro para $f(t)$ é construído selecionando uma faixa de frequências em A_n e B_n , ou seja, multiplicando-as por uma função moduladora, e transformando de volta para o domínio de t . Este procedimento é comum em processamento de imagens e sons, por exemplo, para eliminar ruído.

Como mostrado em,^{26,67} este método pode ser usado para filtrar frequências de qualquer parâmetro dependente do tempo que possa ser extraído de uma dinâmica molecular, incluindo as $3N$ coordenadas que descrevem a trajetória da proteína. Aplicado nestas coordenadas, o filtro resulta em uma nova trajetória, em que movimentos de alta ou de baixa frequência podem ser negligenciados e as análises seguintes podem ser feitas somente numa faixa de frequências de interesse.

Em⁶⁸ este método de filtragem é usado para separar movimentos harmônicos, selecionando apenas uma frequência, e assim compará-los com os obtidos por NMA, ainda que as dinâmicas calculadas simulassem cerca de apenas 200 ps. O filtro é aplicado separadamente para cada uma das $3N$ coordenadas, gerando novas trajetórias harmônicas, como na Equação 3.9, constituída por apenas uma componente da soma na Equação 3.6, onde Δx_{in} é o deslocamento na coordenada i no modo harmônico n .

$$\Delta x_{in} = A_{in} \cos\left(\frac{\pi n k \delta t}{T}\right) + B_{in} \sin\left(\frac{\pi n k \delta t}{T}\right) = F_n w_{in} \cos\left(\frac{\pi n k \delta t}{T} + \phi_{in}\right) \quad (3.9)$$

Diferentemente da Equação 3.3, as fases encontradas, $\phi_i n$, na Equação 3.9 não são as mesmas para cada coordenada i num mesmo modo. De fato, a Equação 3.9 descreve uma elipse, com duas amplitudes relacionadas às componentes em seno e cosseno com mesma frequência. Essa diferença dificulta a comparação com o movimento descrito por modos normais. Em⁶⁸ é proposto que a direção relativa do movimento seja escolhida de

acordo com o maior valor da parte real ou imaginária. Neste trabalho propomos uma variação nesse sentido.

A divergência nas fases surge das condições de contorno da trajetória, ou da falta de simetria inerente. Ao impor que seus limites inicial e final sejam iguais a zero, e também uma simetria ímpar em torno de $t = 0$, as componentes em cosseno se anulam, e a transformada resulta num desenvolvimento somente em termos de seno, todos com a mesma fase. Isso não implica em perda de generalidade, já que os conjuntos das transformadas tanto em seno, quanto em cosseno ou nos dois, possuem o mesmo número de funções base. É importante ressaltar que essas imposições não provocam nenhuma descontinuidade, pois as equações de movimento de Newton são simétricas em zero, com respeito ao intervalo de tempo. Isto é, se temos uma solução entre 0 e T , ampliando com a simetria no intervalo de $-T$ a 0, ela ainda é solução no intervalo de $-T$ a T , e isso é válido para todas as soluções da equação do movimento.

Para que este desenvolvimento seja aplicável, $f(t)$ deve obedecer:

$$f(0) = f(T) = 0 \quad (3.10)$$

Para garantir isso, é preciso remover uma linha reta de $f(0)$ a $f(T)$. Isto simula meio período. A simetria ímpar é construída com $f(-t) = -f(t)$ de $t = 0$ a $t = -T$.

A transformada em seno, para uma função discreta, é então calculada como na [Equação 3.11](#).

$$F_n = \frac{1}{K} \sum_{k=-K}^{K-1} f(t) \sin\left(\frac{\pi n k \delta t}{T}\right) \quad (3.11)$$

Como $f(t)$ é uma função ímpar, periódica em $[-T, T]$, a [Equação 3.11](#) se reduz à [Equação 3.12](#):

$$F_n = \frac{2}{K} \sum_{k=0}^{K-1} f(t) \sin\left(\frac{\pi n k \delta t}{T}\right) \quad (3.12)$$

De forma que somente os pontos de $t = 0$ a $t = T$ são necessários. Como no filtro de Fourier, depois da transformação para o domínio das frequências, uma faixa de frequências, ou somente uma frequência, é selecionada através da multiplicação da função F_n por uma função quadrada $H(n)$ ([Equação 3.13](#)).

$$H(n) = \begin{cases} 1 & , \quad n_{min} < n < n_{max} \\ 0 & , \quad \text{caso contrário} \end{cases} \quad (3.13)$$

A transformada em senos de volta para o domínio do tempo resulta numa função $f(t)$ filtrada e periódica. A aplicação desse método a cada coordenada de uma dinâmica molecular, filtrando somente uma frequência, resulta numa nova trajetória, descrita pela [Equação 3.14](#), em que F_{in} pode ser normalizada para $w_{in}S_n$, sendo S_n a amplitude para cada componente de seno filtrada do movimento de todas as coordenadas em conjunto.

$$\Delta x_{in} = F_{in} \sin(2\pi\omega_n t) = S_n w_{in} \sin(2\pi\omega_n t) \quad (3.14)$$

Esta última equação agora é semelhante à [Equação 3.3](#), e modos normais podem agora ser comparados por produto escalar, por exemplo, como será explicado adiante.

Um resumo para o método de filtro por senos proposto acima consiste de 5 tópicos:

1. subtrair uma reta do início ao fim da função original, tornando-a periódica com simetria ímpar e nula em seus contornos;
2. tomar a transformada em senos;
3. filtrar uma frequência multiplicando a transformada em seno por $H(n)$;
4. tomar a transformada em senos de volta para o domínio do tempo;
5. adicionar o coeficiente linear da reta calculada em 1. A adição do coeficiente angular remove a periodicidade da trajetória filtrada resultante.

Esse método pode ser aplicado a qualquer propriedade de uma dinâmica molecular que seja uma função do tempo. Mas quando aplicada à trajetória de uma molécula, esses passos devem ser feitos para todas as $3N$ coordenadas separadamente.

As frequências dos modos harmônicos obtidos por este método, estão diretamente relacionadas ao comprimento da trajetória simulada, e não com as frequências básicas intrínsecas a estrutura e suas interações, como teoricamente se obtém por NMA e ENM, que ademais perdem totalmente a relação com essas frequências devido às aproximações. Ainda assim, é uma vantagem deste método poder relacionar frequências aos movimentos coletivos encontrados, diferente de PCA que desassocia as conformações da função de tempo. De fato, conformações de diferentes simulações podem ser analisadas em conjunto por PCA.

As frequências neste método de filtro são calculadas como na [Equação 3.15](#).

$$\omega_n = \frac{n}{2\delta t K} \quad , \quad n = 0, 1, \dots, K \quad (3.15)$$

Elas estão limitadas justamente pelo comprimento da dinâmica e pelo intervalo entre cada conformação. A frequência mais alta que se pode verificar é o inverso deste

intervalo, enquanto que a mais baixa é o inverso do dobro do comprimento da dinâmica. As frequências intermediárias são todas múltiplos da mais baixa.

Como já mencionado, os vetores que descrevem os modos harmônicos assim obtidos, tampouco são ortogonais. É possível que o modo obtido da frequência mais baixa na realidade seja muito similar ao segundo modo de frequência mais baixa. Assim, por mais que uma transição entre estados funcionais da proteína não seja observada durante a dinâmica molecular, é possível que um modo harmônico nesta direção possa ser obtido como combinação de outros de mais alta frequência e já detectáveis em uma dinâmica. Desta análise, é natural supor que um conjunto de modos é mais representativo de uma transição entre conformações relacionadas à função biológica da proteína, que um modo por si só.

Esta expectativa já existia em relação aos modos harmônicos calculados pelos outros métodos descritos.^{69,70} De fato, qualquer direção no movimento pode ser reconstruída por uma combinação de modos normais, visto que eles formam uma base ortogonal. A dúvida reside em saber se é possível reconstruir mudanças conformacionais relacionadas à função a partir de um conjunto limitado de modos.⁶⁰ Para responder essa questão muitos estudos envolvem a comparação entre os modos harmônicos calculados e a diferença de duas conformações cristalográficas que representem estados distintos, quando disponíveis para uma mesma proteína. Como esclarece Bahar,⁶⁰ matematicamente essa comparação se faz procurando uma amplitude para o vetor do modo harmônico em questão, que somado às coordenadas de uma das estruturas resolvidas, minimize o desvio desta em relação à segunda estrutura. Isso pode ser obtido pela derivada do quadrado da diferença na amplitude que buscamos, como na [Equação 3.16](#),

$$\frac{\partial}{\partial s_n} [(\vec{r}_a - s_n \vec{u}_n) - \vec{r}_b]^2 = 0 \quad (3.16)$$

onde s_n é amplitude a ser calculada que multiplica cada modo harmônico n , u_n , corresponde ao vetor do modo harmônico, enquanto $\vec{r}_a$ ou $\vec{r}_b$ correspondem ao vetor que define as conformações **A** ou **B**, respectivamente.

Desta definição é preciso que a amplitude que multiplica o vetor do modo harmônico seja igual ao produto escalar do vetor diferença entre as conformações conhecidas e este vetor, ambos normalizados. Por isso, este produto escalar tem sido amplamente usado como o parâmetro de sobreposição (*overlap*), um indicativo da similaridade entre as mudanças conformacionais verificadas nas conformações resolvidas e as observadas no decorrer de um modo harmônico. E este modo harmônico pode ser proveniente de qualquer um dos métodos citados, sendo parte de uma base ortonormal ou não. Bahar ainda exemplifica para 6 proteínas o quanto que o RMSD entre as duas conformações pode diminuir quando é incluído o deslocamento por cada um dos 20 modos normais de mais baixa frequência, e

conclui que sempre há pelo menos um que abaixe o RMSD significativamente.⁶⁰

Em 2001, num dos primeiros estudos do tipo, Tama e Sanejouand ainda levantam outros parâmetros para comparar com mudanças conformacionais observadas experimentalmente em 20 proteínas.⁷⁰ São eles a correlação e o grau de coletividade, além de comparar também os deslocamentos por resíduo. Já em 2005, Lindahl propõe que um conjunto de modos normais possa ser usado para refinar estruturas, dadas as coordenadas de um modelo e mais alguns dados experimentais da estrutura que se deseja obter ao final, sejam provenientes da difração de raios-X ou de criomicroscopia eletrônica.⁶⁹ Ele demonstrou que é possível diminuir o RMSD entre as conformações apo e holo de algumas proteínas, ao somar a contribuição de alguns modos normais à conformação apo. Para isso ele multiplica cada um dos autovetores dos modos considerados pelo valor do produto escalar entre este autovetor e o vetor diferença entre as conformações apo e holo. Outra possibilidade de uso que Lindahl propõe para um conjunto de modos normais é a geração de *decoys*, conformações distantes um dado valor de RMSD de uma conformação original, geradas pela contribuição de alguns modos com amplitudes atribuídas randomicamente.⁷¹

Neste projeto de doutorado tentamos também construir uma nova conformação da proteína ElrR, considerando as contribuições de alguns modos harmônicos somadas a sua conformação apo. Para escolher qual conjunto de modos seria usado consideramos todas as combinações possíveis com os 20 modos de mais baixa frequência. Assim, consideramos tanto a possibilidade de usar somente um desses 20 modos, como todos eles, ou qualquer subconjunto.

Para encontrar as amplitudes a multiplicar cada um dos modos, exigimos, assim como Bahar, que o desvio quadrático entre duas estruturas conhecidas devesse ser mínimo. Nesta primeira condição consideramos que vamos somar um conjunto de modos, M modos, e não somente um. Nestes termos o MSD pode ser escrito como na [Equação 3.17](#), em que as variáveis estão definidas como na [Equação 3.16](#).

$$MSD = \left[\sum_{n=1}^M s_n \vec{u}_n - (\vec{r}_a - \vec{r}_b) \right]^2 \quad (3.17)$$

O mínimo do MSD é obtido através das derivadas com respeito a cada uma das amplitudes desconhecidas, formando o conjunto de equações a seguir ([Equação 3.18](#)), como

num ajuste por mínimos quadrados, $\vec{u}_0$ é o vetor diferença entre as conformações **A** e **B**:

$$\left\{ \begin{array}{l} \vec{u}_1 \sum_{n=1}^M s_n \vec{u}_n - \vec{u}_1 \vec{u}_0 = 0 \\ \vec{u}_2 \sum_{n=1}^M s_n \vec{u}_n - \vec{u}_2 \vec{u}_0 = 0 \\ \qquad \qquad \qquad \vdots \qquad = \vdots \\ \vec{u}_M \sum_{n=1}^M s_n \vec{u}_n - \vec{u}_M \vec{u}_0 = 0 \end{array} \right. \quad (3.18)$$

Para resolver este conjunto, reescrevemo-lo na forma matricial, como indica a [Equação 3.19](#).

$$\begin{bmatrix} \vec{u}_1 \vec{u}_1 + \vec{u}_1 \vec{u}_2 + \cdots + \vec{u}_1 \vec{u}_M \\ \vec{u}_2 \vec{u}_1 + \vec{u}_2 \vec{u}_2 + \cdots + \vec{u}_2 \vec{u}_M \\ \vdots \qquad \qquad \qquad \vdots \\ \vec{u}_M \vec{u}_1 + \vec{u}_M \vec{u}_2 + \cdots + \vec{u}_M \vec{u}_M \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{bmatrix} = \begin{bmatrix} \vec{u}_1 \vec{u}_0 \\ \vec{u}_2 \vec{u}_0 \\ \vdots \\ \vec{u}_M \vec{u}_0 \end{bmatrix} \quad (3.19)$$

Esta pode ser resolvida pela decomposição de valores singulares, que retorna o vetor contendo cada uma das amplitudes que desejamos encontrar.

Assim, é possível construir um modelo estrutural de uma proteína combinando um conjunto de modos calculados a partir de uma conformação inicial e somados a ela, para que se aproxime de uma segunda conformação de interesse. É claro que desta conformação de interesse não é necessário conhecer todas as coordenadas exatas, do contrário a construção do modelo seria redundante. Na proposta de refinamento de Lindahl⁶⁹ apenas alguns dados experimentais são necessários. Aqui propomos que se use apenas um trecho da estrutura. Neste projeto usamos justamente as hélices de reconhecimento da ElrR, cujas coordenadas dos carbonos α propomos que sejam iguais às de um dos complexos TR-DNA que analisamos. Para ajustar as hélices de reconhecimento a esta nova posição, pelo método dos mínimos quadrados e combinação linear de modos harmônicos, obtemos as amplitudes s_n para cada modo harmônico n considerado. Como cada modo harmônico também descreve uma direção para os outros átomos da estrutura, e não só os das hélices de reconhecimento, com essas mesmas amplitudes também podemos obter novas posições para o restante da estrutura, e assim construir um novo modelo.

4 MÉTODOS

4.1 Parâmetros de análise estrutural e de flexibilidade

Além dos métodos de simulação acima descritos, também buscamos por parâmetros de caracterização das conformações estudadas no decorrer deste projeto. Eles estão descritos a seguir.

4.1.1 RMSD

Para calcular o quanto uma conformação se desvia de outra em relação a suas coordenadas ou subconjunto delas usamos o RMSD, como definido na [Equação 4.1](#), que pode ser escrito também como uma função do tempo.

$$RMSD = \sqrt{\frac{1}{3N} \sum_{i=1}^{3N} (r_i - \bar{r}_i)^2} \quad (4.1)$$

4.1.2 Conservação de hélices α

Para construir um modelo da ElrR a partir de um conjunto de modos harmônicos, consideramos todos os possíveis conjuntos dos 20 modos de mais baixa frequência, obtidos tanto por NMA, quanto pelo filtro de frequências em uma dinâmica molecular. Assim, muitos modelos foram construídos, mas alguns apresentavam grandes distorções na estrutura. Então um parâmetro de seleção dos modelos foi a conservação de hélices α , visto que a ElrR não apresenta fitas β . O cálculo de estrutura secundária foi feito assim como descrito por Kabsch e Sander.⁷² Uma energia é atribuída a ligação de hidrogênio a depender das distâncias entre a carbonila de um resíduo e a amida de um resíduo na quarta posição à frente. Se esta energia for menor que -0,5 kcal/mol, então esses dois resíduos formam uma ligação de hidrogênio característica de hélice α e formam uma volta de hélice junto com os três resíduos entre eles.

Nos modelos da ElrR ligada ao DNA, aqueles resíduos que na conformação apo já eram considerados como pertencentes a uma hélice α e continuam sendo contam na porcentagem de hélices α conservadas.

4.1.3 Distorções nas distâncias entre carbonos α vizinhos

Ainda para avaliar as distorções presentes nos modelos gerados para ElrR ligada ao DNA, contamos a quantidade de carbonos α que não estivessem entre 3,6 e 4,1 Å distantes de um outro carbono α vizinho.

4.1.4 Mapa de contatos

Para avaliar diferenças conformacionais entre os modelos da ElrR gerados e também para caracterizar a contribuição do N-terminal nas dinâmicas da MosR, contamos quantos carbonos α se aproximavam um do outro a menos de 6 Å. Também contamos como contatos polares todo os pares de átomos de nitrogênio ou oxigênio distantes um do outro por menos de 3,5 Å, além de contarmos como contatos apolares quaisquer outros pares de átomos, diferentes também de hidrogênio, distantes entre si menos de 4 Å.

4.1.5 Parâmetro de deslocamento atômico isotrópico e anisotrópico

Os parâmetros de deslocamento atômico (ADP) refletem as flutuações de um átomo em torno de um ponto médio num ensemble de conformações. Eles podem ser determinados experimentalmente por difração de raios-X, se a resolução for alta o suficiente. Assim como também possuem uma definição teórica que pode ser aplicada para quantificar flutuações em uma simulação computacional.⁷³ Os ADPs anisotrópicos são seis parâmetros que compõem a seguinte matriz simétrica (Equação 4.2),

$$\begin{bmatrix} \langle \Delta x \Delta x \rangle & \langle \Delta x \Delta y \rangle & \langle \Delta x \Delta z \rangle \\ \langle \Delta y \Delta x \rangle & \langle \Delta y \Delta y \rangle & \langle \Delta y \Delta z \rangle \\ \langle \Delta z \Delta x \rangle & \langle \Delta z \Delta y \rangle & \langle \Delta z \Delta z \rangle \end{bmatrix} \quad (4.2)$$

onde o Δ significa o deslocamento de um átomo de sua posição média em cada uma das três coordenadas e os parênteses angulares denotam a média temporal.

Os seis parâmetros podem aparecer em arquivos de estruturas, no formato pdb, multiplicados por 10.000, e são facilmente visualizados como elipsoides, que demarcam uma região onde seria possível encontrar o respectivo átomo com uma probabilidade de 50%.

O traço da matriz equivale ao quadrado do deslocamento médio, que se assemelha a definição de RMSF (*Root Mean Square Fluctuation*) a menos da raiz quadrada. Ele também está na definição do fator de temperatura, ou fator B, ou fator de Debye-Waller, ou ADP isotrópico, multiplicado por $8\pi^2/3$.

4.1.6 Produto escalar, ou *overlap*

Um produto escalar, $P_{nn'}$, entre dois vetores sempre pode ser calculado (Equação 4.3), indicando quão semelhantes eles são, ou quão semelhantes são as direções dos movimentos que eles representam.

$$P_{nn'} = \vec{u}_n \vec{u}_{n'} \quad (4.3)$$

Aqui, ambos os vetores w_n e w'_n são normalizados para que $P_{nn'}$ igual a 1 indique mesma direção e sentido dos vetores, -1 indica mesma direção, mas em sentidos opostos, enquanto que 0 significa que os vetores são ortogonais, ou representam movimentos ortogonais.

4.1.7 Amplitude do modo harmônico

Como definidos pela [Equação 3.14](#), qualquer modo harmônico pode ser descrito com um versor, w_{in} , e uma amplitude, S_n , onde i corresponde à coordenada. Esse versor pode ser de dimensão $3N$, onde N é o número total de átomos, ou qualquer subconjunto deles. Para comparar os modos harmônicos filtrados de diferentes dinâmicas da MosR, consideramos como amplitude global o valor de S_n , associado ao versor composto apenas das coordenadas do carbonos α , e consideramos como amplitude das hélices de reconhecimento a associada a um versor composto apenas das coordenadas dos carbonos α das hélices de reconhecimento.

4.1.8 Grau de coletividade

Também comparamos os modos harmônicos, provenientes dos diferentes métodos que adotamos, por seu grau de coletividade, como definido por Tama e Sanejouand⁷⁰ calculado como indica a [Equação 4.4](#),

$$\kappa_n = \frac{1}{N} e^{-\sum_{i=1}^N S_n(\Delta r_{in})^2 \log[S_n(\Delta r_{in})^2]} \quad (4.4)$$

onde S_n é definidos como na [Equação 3.14](#), Δr_{in} é o deslocamento de cada átomo i no modo n e N é o número total de átomos considerados, que no caso correspondem a apenas os carbonos α .

4.1.9 Alinhamento, ou superposição, estrutural

Muitas análises que fazemos aqui requerem um alinhamento, ou superposição, prévio das conformações. Para isto escrevemos uma subrotina em Fortran90 usando o método de diagonalização da matriz quaternion,⁵⁴ acompanhado de um método iterativo, baseado na rotina Lsq do programa O,⁵⁵ que reescolhe resíduos da região base para o alinhamento de acordo com as distâncias dos carbonos α , de forma a convergir a uma lista de átomos pertencentes a um núcleo rígido, com menor RMSD.

Apenas para o cálculo de RMSD entre conformações da MosR, usamos a rotina *align* do Pymol,⁷⁴ e também para a superposição dos complexos TR-DNA especificamente para a construção da [Figura 8](#).

4.2 Parâmetros de simulação

Cada uma das três proteínas consideradas neste estudo serviu para análises diferentes e foi abordada por métodos distintos. Assim, estes métodos estão descritos abaixo de acordo com a proteína.

4.2.1 ElrR

Apenas uma simulação de dinâmica molecular foi feita com a ElrR. Uma simulação *all-atom* de 100 ns usando o campo de força charmm36⁷⁵ com o programa NAMD.⁷⁶ O sistema foi construído com 6 contraíons de sódio e solvatado explicitamente com o modelo de água TIP3P formando uma camada de água de no mínimo 10 Å entre a proteína e as paredes da caixa. A conformação inicial, em sua forma apo, foi cedida por Michel De Groote em fase de refinamento, atualmente depositada no PDB sob o código 5v5t. Como as duas histidinas presentes em cada cadeia apareciam expostas ao solvente elas foram completamente protonadas. Esse sistema foi previamente minimizado e equilibrado por 10.000 passos e 200 ps, respectivamente, a pressão constante de 1 atm controlada pelo pistão de Langevin,⁷⁷ temperatura constante de 310 K e com um coeficiente de amortecimento de 1 ps^{-1} para simular uma dinâmica de Langevin.⁷⁸ Condições periódicas de contorno foram usadas com PME (*particle mesh* Ewald)⁷⁶ como método de cálculo eletrostático e passos de 2 fs. A dinâmica foi então realizada por 100 ns, guardando conformações a cada 0,5 ps, nas mesmas condições que nas etapas anteriores exceto pelo amortecimento que foi desligado e a temperatura controlada por acoplamento ao banho térmico.

O filtro de frequências foi aplicado utilizando as últimas 131.072 conformações desta dinâmica, o que corresponde a um período de 65,536 ns.

Análise de Modos Normais (NMA) foi realizada para a ElrR com a mesma conformação inicial que a da dinâmica descrita acima, mas com o potencial charmm27⁷⁹ e o programa Gromacs⁸⁰ e solvente implícito pelo método GBSA (*Generalized Born/Solvent-Accessible Surface Area*).⁸¹ Neste caso as histidinas foram protonadas apenas no carbono ϵ . Duas etapas de minimização foram feitas. A primeira com o método de gradientes conjugados que atingiu uma força máxima de $40,9 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. A segunda etapa foi feita com o método L-BFGS (*limited-memory* Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian),⁸² que alcançou uma força máxima de $2,0 \cdot 10^{-9} \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Após a diagonalização da matriz hessiana, os 6 primeiros autovalores resultaram muito próximos de zero, da ordem de 10^{-5} ou 10^{-6} , enquanto que nenhum dos seguintes resultou em valores negativos.

Alguns modelos da ElrR ligada ao DNA foram construídos pela geração de decoys com o método proposto por Lindahl,⁷¹ como comentado anteriormente. Calculados diretamente do site do NOMAD-REF, foram usados os primeiros 20 modos normais calculados pelo modelo de redes elásticas de Tirion.⁵⁸ Apenas 20 modos foram usados dada a limitação

do tempo de cálculo para cada modelo gerado, cujo de número dobra com cada modo considerado. Foi usado um parâmetro de peso para as distâncias de 5 Å, e um parâmetro de corte de 10 Å para gerar 10 conjuntos com 100 estruturas cada um. Cada conjunto possui um RMSD em relação a estrutura original de 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5,0, 5,5, ou 6,0 Å. A estrutura original foi a mesma usada nas simulações anteriores.

O modelo construído por rotação e translação de corpos rígidos foi construído dividindo a ElrR em 3 corpos rígidos diferentes, duas compostas pela região N-terminal e a respectiva região central mais próxima, e a terceira composta pelas duas regiões C-terminais de cada cadeia. As duas primeiras subunidades foram posicionadas de acordo com a superposição das hélices de reconhecimento com as mesmas da estrutura de referência. A terceira subunidade foi posicionada de forma intermediária entre as duas primeiras, superpondo as hélices de transição entre as regiões central e C-terminal. O algoritmo de superposição utilizado foi o mesmo descrito anteriormente usando somente a cadeia principal dos exatos trechos mencionados para alinhamento, sem permitir as iterações de otimização.

Todos os modelos construídos para representar a ElrR em sua conformação ligada ao DNA foram minimizados com o programa CHARMM⁸³ e o campo de força charmm27,⁸⁴ fixando os átomos da cadeia principal das hélices de reconhecimento. Duas etapas de minimização foram feitas, a primeira de 100 passos usando o algoritmo de *steepest descent*, e a segunda foi feita com o método ABNR (*Adopted Basis* Newton-Raphson) até que a força média fosse de 0,01 kcal mol⁻¹ Å⁻¹.

4.2.2 MosR

Cinco simulações de dinâmica molecular *united-atom* (hidrogênios não polares estão implícitos) foram realizadas com a MosR. Todas foram iniciadas a partir da conformação cristalográfica apo (código PDB: 4fx0). Como alguns trechos de ambas as cadeias estavam faltando, e com o propósito de avaliar o papel do N-terminal não detectável na cristalografia, com o programa Coot³⁵ e em especial utilizando a rotina de regularização de geometria, reconstruímos o restante da sequência da proteína, completando dos resíduos 1 a 148, para uma das cadeias, e pela simetria do grupo espacial também a segunda cadeia. 12 conformações distintas foram construídas alternando as posições dos 7 primeiros resíduos, incluindo os 5 adicionados manualmente.

Para cada uma das 12 conformações construímos um sistema com 16 contraíons de cloro e solvente explícito com o modelo de água SPC/E.⁸⁵ Com o programa Gromacs⁸⁰ e o campo de força Gromos54a7,⁸⁶ cada sistema foi minimizado até que a força máxima fosse menor que 103 kJ mol⁻¹ nm⁻¹. Mantendo a proteína restrita, o solvente foi então equilibrado por 100 ns mantendo constantes o volume e a temperatura, a 310 K; e depois por mais 100 ns, mas mantendo constantes a temperatura e a pressão, a 1 bar. Mais

um passo de equilibração foi feito permitindo a flexibilidade somente do N-terminal, os 7 primeiros resíduos, e do solvente. Este passo foi de 1 ns, à pressão constante de 1 bar, e com a temperatura iniciando em 0 K, e lentamente incrementada durante 500 ps para que alcançasse 310 K. Todos estes passos de equilibração foram calculados usando condições periódicas de contorno, LINCS como algoritmo de constrição das ligações de hidrogênio,⁸⁷ PME para cálculos eletrostáticos,⁸⁸ o algoritmo *velocity rescaling* para controle de temperatura,⁸⁹ o de Parrinello-Rahman para controle da pressão⁹⁰ e passos de 2 fs.

Com duas dessas conformações prosseguimos com uma equilibração de 1 ns com as mesmas condições anteriores, mas permitindo a flexibilidade de toda a proteína com o mesmo protocolo de aquecimento gradual, e depois a dinâmica molecular por 400 ns, guardando conformações a cada 10 ps.

Com essas mesmas duas conformações construímos conformações oxidadas. A única mudança nas coordenadas foi a escolha de outro rotâmero para a cisteína 12 de cada cadeia, aproximando o átomo de enxofre ao da cisteína 10. Uma nova topologia foi gerada usando a opção -ss na rotina pdb2gmx do Gromacs. A solvatação, a minimização e as equilibrações seguiram nas mesmas condições detalhadas para as outras simulações, exceto pelo passo em que somente o N-terminal da proteína é livre para se mexer. E por fim também foram simuladas duas dinâmicas de 400 ns para as conformações oxidadas.

Com uma terceira conformação foi feita uma simulação de 100 ns usando uma restrição na distância entre cada par dos átomos de nitrogênio da alanina 6 e de oxigênio da arginina 65 da outra cadeia. Essa restrição foi imposta com os valores de 0,24 nm, 0,4 nm e 0,5 nm para os parâmetros *low*, *up1* e *up2*, respectivamente. Ela foi mantida durante todas as etapas simuladas, desde a minimização até a dinâmica molecular, calculadas com as mesmas condições descritas para os outros sistemas.

4.2.3 DEA

Duas simulações de dinâmica molecular foram iniciadas para a DEA, ambas partindo da conformação apo (código PDB: 1fsf) e ambas com o programa CHARMM⁸³ e o campo de força charmm27. Uma é a simulação de todo o hexâmero livre, e a outra é de apenas um monômero interagindo com suas próprias imagens dispostas como na simetria do grupo espacial observado. Ambas estruturas tiveram suas histidinas protonadas de acordo com nossos estudos anteriores.⁴¹

A primeira foi solvatada explicitamente com o modelo de água TIP3P formando uma camada de água de no mínimo 10 Å entre a proteína e as paredes da caixa cúbica, simulada com condições periódicas de contorno. Após uma minimização de 100 passos com o método de *steepest descent*, foram adicionados contraíons, 45 de cloro e 21 de potássio para neutralizar o sistema a uma concentração aproximada de 0,15 M de sal. Essa neutralização

foi acompanhada de alguns passos de minimização, e mais passos de minimização foram feitos intercalando os métodos de *steepest descent* e ABNR e gradualmente liberando a proteína para se mexer, primeiro somente os hidrogênios, depois só a cadeia lateral e por fim a proteína inteira, até que o sistema atingisse uma força média de $0,1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. A equilibração foi feita por 400 ps em passos de 2 fs com o método de Andersen *heatbath*⁹¹ para controle da temperatura, que foi gradualmente incrementada até atingir 310 K, e o barostato MC para controle da pressão a 1 atm.⁹² A dinâmica molecular continuou a partir daí, guardando conformações a cada 10 ps, e já remonta 340 ns.

A segunda também foi solvatada explicitamente com o modelo de água TIP3P, mas compondo apenas 1/6 de uma caixa hexagonal. A conformação de partida foi resolvida com o grupo espacial *P6322*, mas como na cela unitária deste grupo existe muito espaço vazio, ou com água no cristal, optamos por simular o hexâmero como um monômero como unidade assimétrica em um grupo espacial *P321* para aproveitar melhor o espaço, mas mantendo a simetria em cada hexágono. Para isso foi preciso uma rotação dele de 90° em torno do eixo z. As dimensões da caixa hexagonal foram ajustadas para resultar em não muito mais que 70% de átomos de solvente. Definimos 5 operações de simetria para posicionar os outros 5 protômeros, assim como no hexâmero formado pela simetria cristalográfica. Elas aparecem na figura 4.1d e são as mesmas que definem o grupo espacial *P321*. As possíveis transformações foram geradas com um *cut-off* de $141 \text{ \AA}$. Elas definem as imagens a serem construídas, obedecendo as condições periódicas de contorno, e incluem tanto as translações quanto as rotações determinadas pela simetria do grupo espacial. Elas devem ser no mínimo equivalentes ao número de caixas vizinhas à simulada, quando não são aplicadas operações de simetria. Como aqui definimos 5 operações, para remontar as fronteiras das caixas vizinhas são necessárias no mínimo mais cinco transformações para cada.

Para solvatar esse sistema partimos das coordenadas de moléculas de água que preencheriam a mesma caixa hexagonal por completo e então excluímos as águas de 5/6 da caixa, como na região indicada na Figura 7a, com as restrições geométricas indicadas na Figura 7c. Procedemos com as mesmas etapas de minimização e neutralização como no sistema com o hexágono completo, adicionando 29 átomos de cloro e 25 de potássio de contraíons. A equilibração foi simulada então por 200 ps como num ensemble microcanônico, com volume e energia constantes, enquanto a temperatura foi incrementada gradualmente até alcançar 310 K. A dinâmica molecular prosseguiu a partir daí com as mesmas condições guardando conformações a cada 10 ps e já remonta 50 ns.

NMA também foi realizada para a DEA, partindo do mesmo hexâmero que na primeira simulação. Para este método usamos o programa Gromacs⁸⁰ com o campo de força ENCADV.⁹³ Três métodos de minimização foram utilizados intercaladamente, gradientes conjugados, *steepest descent* e L-BFGS, para alcançar uma força máxima da ordem de

Figura 7 – a) monômero solvatado compondo um sexto de uma caixa hexagonal. b) hexagonal construído a partir do monômero simulado e suas imagens. c) as cinco operações de simetria cristalográfica que recompõem o hexâmero a partir do monômero. d) inequações geométricas usadas para excluir as moléculas de água cuja posição do oxigênio cumprisse essas características, restringindo o solvente a apenas um sexto da caixa hexagonal.

Fonte: Elaborada pela autora.

$10^{-6} \text{ kJ mol}^{-1} \text{ nm}^{-1}$.

Para todas as análises todas as conformações foram superpostas pelos resíduos de 207 a 233 e 244 a 250, sendo que para o hexâmero as 6 cadeias foram superpostas concomitantemente.

5 RESULTADOS E DISCUSSÃO

Os resultados estão divididos em quatro seções. A primeira delas trata dos complexos TR-DNA analisados, quando discutimos as variadas posições que as hélices de reconhecimento podem adotar, como reflexo da sequência de resíduos da hélice de reconhecimento, da sequência de bases no sítio de DNA, da simetria das hélices de reconhecimento e do encaixe concomitante delas em dois sulcos maiores do DNA, também simétricos. A segunda seção descreve os resultados obtidos no processo de construção do modelo da ElrR ligada ao DNA. Ela inicia justamente discutindo a escolha da posição alvo das hélices de reconhecimento, e finaliza com a análise do sítio alostérico nas conformações fechadas obtidas. A terceira seção retrata a MosR simulada em seus dois estados de oxidação diferentes, e discute as diferenças conformacionais e dinâmicas encontradas nas amostragens obtidas. A quarta mostra os primeiros resultados para discutir a influência da simetria estrutural na dinâmica da DEA.

5.1 Sobre os complexos TR-DNA

Estudamos 27 complexos TR-DNA cuja proteína fosse homodimérica com o motivo HTH presente no DBD, e cujo sítio de DNA apresentasse certa simetria na sequência das bases. Deste grupo, a primeira característica geral que encontramos é a existência de um resíduo que mais se aproxima do fundo do sulco maior do DNA, que denominamos de resíduo principal. AlQuraishi et al.⁵² definiu como resíduo central aquele cujo carbono α estivesse mais próximo do eixo do DNA. Essas definições parecem ser idênticas, mas no grupo de complexos estudados por AlQuraishi et al., o resíduo central também aparece no meio da hélice de reconhecimento. Enquanto que, no conjunto que analisamos, o resíduo principal sempre se encontra na primeira volta da hélice de reconhecimento, já que observamos o N-terminal dela sempre apontando para a dupla hélice de DNA. Assim, o resíduo principal é aquele voltado para o fundo do sulco do DNA na primeira volta da hélice de reconhecimento. Seguindo a definição de Harrison,⁹⁴ o resíduo principal ocuparia a posição de número 13.

Como o resíduo principal é o que mais se aproxima do fundo do sulco, a inserção da hélice de reconhecimento no sulco maior é em grande parte regida por este resíduo. A variedade de aminoácidos ocupando esta posição é pequena. Dos 27 complexos que analisamos, 7 possuem uma serina como resíduo principal, 5 possuem glicinas, 3 possuem treoninas, 3 possuem lisinas, 3 possuem prolinas, 2 possuem argininas, 2 possuem glutaminas, um possui uma asparagina e um possui uma alanina. Na grande maioria dos casos o resíduo principal faz pontes de hidrogênio específicas com bases de DNA, exceto no caso das prolinas, da alanina e algumas glicinas – outras glicinas conseguem fazer pontes de

hidrogênio através de seus átomos da cadeia principal. Esse resíduo parece influenciar na posição da hélice de reconhecimento pela distância do ponto principal ao eixo da hélice, Dph (ver a seção métodos para definição dessa distância). Dphs menores que 7 Å só foram encontradas quando o resíduo principal é uma glicina, enquanto que Dphs maiores que 10 Å só foram encontradas quando o resíduo principal é uma arginina ou uma lisina.

Encontrado o resíduo principal em cada complexo, superpusemos todos estes por apenas um dos dois HTHs de cada complexo. A [Figura 8](#) mostra estes HTHs superpostos ao centro, além das hélices de reconhecimento do segundo HTH de cada complexo, e uma dupla hélice de DNA de um dos complexos. A superposição foi feita considerando 10 resíduos antes do principal e 9 depois dele. Os HTHs superpostos se alinham muito bem devido a conservada geometria desse motivo, mas as hélices de reconhecimento da outra cadeia adotam uma posição distinta em cada complexo. Dessa figura é possível observar três grupos mais discrepantes quanto à posição dessa outra hélice de reconhecimento. Do ponto de vista do observador da [Figura 8](#), no primeiro grupo essa hélice se posiciona a direita da central; no segundo grupo ela se posiciona a esquerda da central e seu C-terminal se encontra acima do eixo do DNA; e no terceiro grupo ela também se encontra a esquerda da central, mas o C-terminal se encontra abaixo do eixo do DNA.

Figura 8 – Superposição de complexos TR-DNA por apenas um dos motivos HTH. A superposição foi feita considerando sempre o resíduo principal, os 10 resíduos anteriores a ele e os 9 posteriores a ele. No centro está representada essa região superposta de todos os complexos e também as respectivas hélices de reconhecimento completas. A hélice de reconhecimento da outra cadeia também está representada em cartoon para todos os complexos. A dupla hélice de DNA desenhada corresponde ao complexo de código PDB 4aik. Além deste estão desenhados os complexos de código PDB 4ivz, 2vz4, 3q5f, 3s8q, 3u3w, 3zkc, 3zpl, 4fth, 4fx4, 4ihs, 1z9c, 4jl3, 4l62, 2e1c, 1jt0, 2yvh, 3vok, 4i6z, 1zlk, 3a12 e 1h0m.

Fonte: Elaborada pela autora.

Quanto ao posicionamento das hélices de reconhecimento à esquerda ou à direita da central, definimos esses dois arranjos como do tipo I ou do tipo II, respectivamente.

Eles refletem dois distintos arranjos da estrutura quaternária, de forma que é impossível superpor simultaneamente os dois DBDs de TRs de tipos diferentes. As do tipo I apresentam o N-terminal da hélice de reconhecimento mais afastado do eixo de simetria da proteína, enquanto que as de tipo II apresentam o N-terminal mais próximo ao eixo de simetria, como exemplificado na [Figura 9](#). A [Figura 9](#) também apresenta outro ponto de vista. Com o eixo do DNA normal ao plano da figura, no tipo I a hélice de reconhecimento mais próxima do observador apresenta seu C-terminal à direita do eixo de simetria da proteína, enquanto que no tipo II o C-terminal dessa hélice aparece à esquerda do eixo de simetria.

Figura 9 – Dois complexos TR-DNA exemplificando o tipo I acima (a e b) e o tipo II abaixo (c e d), sob dois pontos de vista rotados em 90° ao redor do eixo de simetria mostrado em preto. Coloridos em rainbow estão os motivos HTH. O tipo I é exemplificado pelo complexo de código PDB 4ivz, enquanto o de tipo II pelo de código PDB 2e1c.

Fonte: Elaborada pela autora.

Algumas restrições nos arranjos dos complexos podem ser associadas aos tipos I ou II. A primeira que observamos é possibilidade das proteínas do tipo I apresentarem um DBD com a subfamília wHTH. Nenhuma proteína do tipo II, nesse conjunto que estudamos, possui essa extensão ao HTH em forma de folha β com um longo *loop* que interage com o sulco menor do DNA. Ademais, o *winged helix* parece proporcionar maior variabilidade a forma de encaixe DBD, pois somente nessa subfamília é que as hélices de reconhecimento apresentam um ângulo entre si, β (ver a seção de métodos), maior que 180°. Essa diferença é aparente já na [Figura 8](#), em que esse ângulo β se percebe pelo posicionamento do C-terminal das hélices de reconhecimento abaixo ou acima do eixo do

DNA. Coerentemente, as de tipo II não possuem exemplos de hélices de reconhecimento com o C-terminal abaixo do eixo do DNA na [Figura 8](#), já que o ângulo β para esse tipo é sempre menor que 180° . Os menores ângulos β são encontrados no tipo II, podendo até ser agudos, enquanto que o menor encontrado para o tipo I foi de 107° .

No tipo II também são encontrados somente ângulos λ (definido na seção de métodos) positivos, enquanto que no tipo I valores negativos de λ são encontrados para todos os complexos com λ maior que 180° e somente para estes. O ângulo λ reflete o quanto as hélices de reconhecimento se inserem no sulco maior. Assim, o fato de valores negativos estarem atrelados a β maior que 180° mostra que este ângulo β relativo entre as hélices impede que elas se insiram mais no sulco. Em contrapartida, valores de β maiores que 180° sempre são acompanhados da extensão de folha β da subfamília wHTH, possibilitando mais contatos com o DNA. Conforme as hélices de reconhecimento se inserem mais ou menos no sulco maior, diferentes voltas delas se aproximam dos grupos fosfato do DNA, e nestas observamos sempre algum resíduo carregado positivamente para equilibrar as cargas. Portanto, a distribuição desses resíduos está diretamente associada a posição da hélice no sulco do DNA. A [Figura 10](#) mostra dois complexos do tipo I com diferentes valores de λ , em um esse valor é de $-17,2^\circ$ enquanto no outro é de $15,1^\circ$, e fica evidente como os resíduos carregados se distribuem de forma distinta.

Figura 10 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos do tipo I, evidenciando a distribuição de cargas dependente do valor do ângulo *lambda*. Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. À esquerda está representado o complexo de código PDB 3q5f, enquanto que à direita está representado o de código PDB 4hf1.

Fonte: Elaborada pela autora.

Também observamos que os complexos do tipo II apresentam valores de Dp e Dh em média maiores que os do tipo I, comportamento mais pronunciado para o parâmetro Dp. Para o tipo I, Dp e Dh valem respectivamente $29,0 \pm 2,9$ Å e $27,0 \pm 2,9$ Å. Enquanto para o tipo II eles valem respectivamente $34,7 \pm 2,8$ Å e $32,4 \pm 4,3$ Å. Essa pequena diferença de 5 Å pode estar associada a uma preferência das hélices de reconhecimento em se acomodarem mais próximas a uma fita de DNA que à outra, para que a primeira hélice do HTH também se acomode com seu N-terminal apontando para uma fita de DNA no tipo I e para a outra no tipo II. Essa distância das hélices de reconhecimento também é acompanhada de um ligeiro aumento na quantidade de bases entre aquelas que interagem especificamente com resíduos iguais nas duas hélices, que é em média 7 bases para o tipo I e 9 para o tipo II. A [Figura 11](#) mostra os dois complexos com o menor e o maior valor de tamanhos de sítio (como definido na seção de métodos), que correspondem respectivamente a complexos pertencentes ao tipo I e ao tipo II.

Figura 11 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos, evidenciando o tamanho do sítio de DNA, definido pelo número de bases entre as bases interagentes com os resíduos principais, e a distância pd, definida como a distância entre os pontos principais. Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. Os resíduos principais e as bases com que eles fazem pontes de hidrogênio estão coloridos em preto. À esquerda está representado o complexo de código PDB 1z9c, enquanto que à direita está representado o de código PDB 2yvh.

Fonte: Elaborada pela autora.

Na [Figura 11](#) é evidente como o tamanho do sítio afeta o posicionamento das hélices de reconhecimento. Essa distância entre elas também é acompanhada de interações distintas dos resíduos carregados positivamente nas hélices de reconhecimento. No tipo I

eles se aproximam de uma fita do DNA, enquanto no tipo II eles se aproximam da outra, como pode ser visto na [Figura 8](#) e na [Figura 11](#). No entanto, a distância pd parece ser sensível a outros fatores. Para os 4 complexos que apresentam um tamanho de sítio de 6 bases, pd varia de 23,9 Å a 27 Å. É provável que interações do DNA com outros resíduos da hélice de reconhecimento provoquem um deslocamento dessa hélice, ou mesmo alguma distorção no DNA.

Essas outras interações de resíduos diferentes do principal com as bases ou mesmo os grupos fosfato também afetam o ângulo com que as hélices de reconhecimento se posicionam, que medimos como α e γ (ver seção de métodos). Em média eles são semelhantes para os tipos I e II, sendo por volta de 76° para ambos. Esse valor é um meio-termo entre 90° , o que corresponderia à hélice de reconhecimento perpendicular ao eixo do DNA, e 62° , que equivale à inclinação de cada fita do DNA em torno do seu eixo. A [Figura 12](#) mostra dois complexos, com diferentes inserções no sulco maior e diferentes inclinações em relação ao eixo do DNA.

Figura 12 – Hélices de reconhecimento e seus respectivos sítios de DNA de dois complexos, evidenciando a inclinação das hélices de reconhecimento em relação ao eixo de DNA, definida por γ . Resíduos carregados positivamente estão coloridos em azul, e resíduos carregados negativamente estão coloridos em vermelho. À esquerda está representado o complexo de código PDB 4fx4 pertencente ao tipo I, enquanto que à direita está o representado o de código PDB 2e1c pertencente ao tipo II.

Fonte: Elaborada pela autora.

A [Figura 12](#) exemplifica diferentes inclinações da hélice de reconhecimento em relação ao DNA e fica evidente como as interações com os resíduos da hélice também são distintas. No complexo à esquerda na [Figura 12](#) as hélices não estão muito inseridas no sulco do DNA e a inclinação delas acompanha uma das fitas de DNA, que interage com os resíduos carregados positivamente localizados apenas na face da hélice próxima a esta fita de DNA. Assim o valor de γ neste complexo é próximo ao ângulo de inclinação das próprias

fitas de DNA em torno de seu eixo. Já no complexo à direita na [Figura 12](#), as hélices de reconhecimento estão bastante inseridas nos sulcos e apresentam resíduos carregados positivamente interagindo com ambas as fitas de DNA, o que leva γ a apresentar um valor próximo a 90° .

O ângulo δ também reflete em parte essa inclinação da hélice de reconhecimento em relação ao eixo do DNA, mas por outra parte também está associado à inserção delas no sulco do DNA. De fato, existem muitos fatores que interferem na posição das hélices de reconhecimento e a combinação deles em muitos casos dificulta a interpretação dos parâmetros que levantamos, assim como a curvatura do DNA, cuja relação com os demais parâmetros não foi analisada em detalhe (por não apresentar uma relação simples). As medidas que encontramos para todos os 27 complexos analisados estão mostrados na [Tabela 1](#) para os complexos de tipo I e na [Tabela 2](#) para os de tipo II. Certamente, uma maior amostragem esclareceria melhor a influência de cada interação base-resíduo no posicionamento das hélices de reconhecimento.

Quanto à distribuição de bases no sítio de DNA ainda podemos levantar mais algumas características além do comprimento. Todos os sítios analisados possuem no mínimo cinco bases simétricas, obedecendo uma palindromia completa ou pseudo-palindromia (ver seção de métodos). Como a posição das hélices de reconhecimento não depende apenas das interações que o resíduo principal faz, podem existir mais bases fazendo interações específicas com a proteína e estas também devem ser compatíveis com a simetria da proteína. Ademais, algumas bases estão presentes para conferir ao DNA maior flexibilidade ou rigidez. Para este conjunto de 27 complexos encontramos uma preponderância de guaninas e citosinas no sulco maior do DNA fazendo interações específicas com a proteína, enquanto que no sulco menor observamos mais timinas e adeninas, como mostra a [Tabela 3](#). Essa preferência é devida provavelmente às diferentes acessibilidades aos grupos de doadores e receptores de hidrogênio dispostos nas bases, quando presentes num sulco maior ou menor. A [Tabela 3](#) ainda discrimina as interações encontradas não só por base, mas também por resíduo. É interessante notar como a arginina faz o maior número de interações, em concordância com o observado em⁴⁹ e, no entanto, a serina é o resíduo que mais aparece como principal. Provavelmente, a arginina ainda consegue interagir especificamente com as bases de DNA mesmo não pertencendo à primeira volta da hélice de reconhecimento por possuir uma cadeia lateral mais longa. Já na situação inversa, se uma arginina fosse o resíduo principal, uma serina na segunda volta da hélice não teria alcance para interagir com as bases. É válido comentar também que o contato feito pela histidina foi encontrado em apenas um complexo e envolve sua protonação do N_ϵ , como discutido em.⁹⁵

Por fim, fica claro que a posição das hélices de reconhecimento é bastante influenciada pela interação resíduo-base, e pela distribuição tanto das bases quanto dos resíduos. Com essa informação nossa amostragem de complexos, ainda que pequena, já foi suficiente

Tabela 1 – Valores observados nos complexos de tipo I estudados para os parâmetros que levantamos (ver seção de métodos). A coluna hélice mostra os resíduos que definem a hélice de reconhecimento na sequência primária da proteína, a coluna rp remete ao resíduo principal, a coluna outros mostra outros resíduos interagentes especificamente com bases, na coluna sítio está o tamanho do sítio de DNA e WHTH remete à participação na subfamília *winged helix*

complexo	hélice	rp	resíduos	ϵ	β	Dph1	Dph2	Dp	δ	Dh	sítio	$\alpha 1$	$\alpha 2$	$\gamma 1$	$\gamma 2$	λ	WHTH
4ht1	38 - 52	40 S	EQ	179,8	106,7	7,5	8,1	36,2	9,9	35,6	9	81,6	82,6	86,6	85	15,1	sim
4ivz	34 - 45	36 T	R	179,1	133	8	8,6	29,5	23,2	27,1	8	70	67,9	76,8	79,2	10	não
1r8e	18 - 28	20 K	-	180	216,6	10,6	10,9	36,4	20,6	34	7	70,5	70,5	54,1	54,7	-17,2	sim
2vz4	15 - 25	17 R	-	180	221,5	10,7	10,6	31,4	12,8	30,6	5	78,1	78,1	65,4	65,8	-13,4	sim
3q5f	59 - 73	61 P	R	178	213,9	7,3	6,9	27	3,7	26,9	6	89,4	87,5	73,4	75,6	-17,9	sim
3s8q	34 - 43	36 T	R	174,7	121,9	7,9	8,4	30,7	21,5	28,6	9	68,8	76	86,1	79,7	12,1	não
3u3w	30 - 39	32 S	R	179,9	148	8,3	8,3	29,1	19	27,5	8	71,7	71,8	77,3	77,5	5,8	não
3zkc	27 - 36	29 S	RQ	178,5	132,7	7,3	7	28,9	11	28,3	8	80,2	79,7	82,5	83,3	3,5	não
3zpl	71 - 85	73 S	RH	178,9	222,7	7,8	7,6	26,4	9,9	26	6	80,4	81,3	67,2	68,1	-14,3	sim
4fth	61 - 72	63 S	RN	179,9	133,3	8,8	8,1	30,5	28	16,9	8	64,4	64,5	73,7	70,2	9,3	não
4fx4	69 - 83	71 T	RR	179,4	228,2	6,7	7,4	24,6	11,4	24,1	6	79,4	79,8	64,9	64,7	-16,4	sim
4ihs	28 - 44	30 P	RQ	178,2	121,8	7,6	7,8	30,5	30,3	26,3	5	65,5	62,4	73,3	71,3	9,8	sim
1z9c	67 - 80	69 G	S	179,7	218	6,3	6,6	24,3	7,3	24,1	4	82,6	84,2	71	69,1	-14,2	sim
4aik	59 - 73	61 P	Q	179	213,5	8,2	7,7	23,9	3,4	23,9	6	86,6	87,3	78,3	78,6	-15,3	sim
3gz6	189 - 201	191 K	R	179,6	140,3	7,4	7,2	30	22,5	27,7	9	68,8	69	70,2	70,9	1,9	sim
3o9x	97 - 104	97 N	R	179,1	164,9	8	7,7	24,9	3,5	24,9	7	89,8	89,3	89,3	89,2	2,2	não
média	-	-	-	-	-	8 $\pm$ 1		29 $\pm$ 3	15 $\pm$ 8	27 $\pm$ 3	7 $\pm$ 1	77 $\pm$ 7		74 $\pm$ 7		-	-

Fonte – Elaborada pela autora.

Tabela 2 – Valores observados nos complexos de tipo II estudados para os parâmetros que levantamos (ver seção de métodos). A coluna hélice mostra os resíduos que definem a hélice de reconhecimento na sequência primária da proteína, a coluna rp remete ao resíduo principal, a coluna outros mostra outros resíduos interagentes especificamente com bases, na coluna sítio está o tamanho do sítio de DNA e wHTH remete à participação na subfamília *winged helix*

complexo	hélice	rp	resíduos	ϵ	β	Dph1	Dph2	Dp	δ	Dh	sítio	$\alpha 1$	$\alpha 2$	$\gamma 1$	$\gamma 2$	λ	wHTH
4jl3	46 - 54	48 Q	K	170,9	87,5	8,7	8,7	37,4	4,3	37,3	9	86,4	87,6	82,4	88,4	7,5	não
4l62	41 - 47	42 G	K	174,6	95	7,2	7,7	35,6	31,3	30,4	10	65,1	70,2	59,6	64,1	8,7	não
2e1c	54 - 68	56 S	E	180	75	7,7	8	29,2	15,1	28,2	9	80,9	80,9	89,5	88,4	16,8	não
1jt0	35 - 43	37 G	K	174,3	104,8	7,2	6,7	38,4	9,4	37,9	10	83,9	81,6	75,2	74,1	10,4	não
2yvh	37 - 45	39 S	K	177,7	113,9	8,3	7,6	37,6	8,6	37,2	11	81,7	88,6	69,4	77,8	13,9	não
3vok	45 - 52	47 G	R	180	80,6	8,2	9,7	34,7	35,3	28,3	9	68	68	60,3	53,7	22,8	não
4i6z	37 - 44	38 G	-	178	100,9	6,3	7,1	37,4	11,4	36,6	10	82,3	80,3	77,7	74,9	6,6	não
1rzz	15 - 24	17 A	R	175,1	34,5	8	8,5	33,9	21,2	31,6	8	85,5	82,3	81,9	72,4	+27.7	não
1zlk	177 - 193	179 K	K	179,6	110,5	7,6	9	30,6	31,2	26,1	8	64,71	64,7	82,3	85,3	23,9	não
4a12	39 - 50	41 Q	-	174,6	97,2	9,1	9,4	37	7,3	36,7	9	87,3	82,9	80,7	88,3	12,6	não
1h0m	201 - 218	203 R	R	179,8	124,1	9,7	10,6	29,5	28,8	25,9	6	64,5	65,1	85,6	85,3	23,8	não
média	-	-	-	-	-	8 $\pm$ 1		35 $\pm$ 3	18 $\pm$ 10	32 $\pm$ 4	9 $\pm$ 1	77 $\pm$ 8		77 $\pm$ 8		-	-

Fonte – Elaborada pela autora.

Tabela 3 – Número de interações específicas encontradas nos complexos estudados distribuídas em função dos resíduos e bases interagentes.

Interações resíduo-base no sulco maior				
	adenina	timina	citossina	guanina
arginina	3	5	0	34
serina	8	6	6	6
lisina	2	0	0	21
glutamato	3	4	3	0
glutamina	6	0	2	1
glicina	1	0	2	5
treonina	0	0	6	0
asparagina	0	2	0	4
histidina	0	2	2	0
Interações resíduo-base no sulco menor				
arginina	13	18	0	1
tirosina	0	0	0	2
alanina	0	0	0	2

Fonte – Elaborada pela autora.

para encontrar um complexo semelhante a ElrR nesses quesitos, o que consiste no primeiro passo para a construção de um modelo da ElrR ligada ao DNA, como será explicado na próxima seção de resultados.

5.2 Sobre a construção de um modelo para a ElrR ligada ao DNA

5.2.1 As hélices de reconhecimento alvo

Com base na análise dos complexos TR-DNA, como descrito na seção de resultados anterior, nós encontramos que as posições das hélices de reconhecimento, uma em relação a outra e ao DNA, são afetadas significativamente pela sequência tanto de resíduos das hélices quanto de bases do sítio de DNA. Assim, nós selecionamos um complexo, dentre os estudados, com ambas as sequências muito similares às da ElrR e de seu sítio de DNA, apesar de as proteínas não serem da mesma família. Este complexo é identificado pelo código de PDB 3gz6, e apresenta a estrutura cristalográfica da NrtR de *Shewanella oneidensis* ligada a um sítio de DNA de 27 pares de bases. As sequências das hélices de reconhecimento da NrtR e da ElrR, e de seus respectivos sítios de DNA estão apresentadas na [Figura 13a](#), com as bases e resíduos semelhantes ressaltados.

O resíduo principal, como definimos na seção de resultados anterior – por sua

Figura 13 – a) seqüências de resíduos das hélices de reconhecimento das proteínas ElrR e NrtR, e as seqüências de seus respectivos sítios de DNA. Em vermelho estão destacados resíduos que fazem ligação de hidrogênio com bases do sítio, os resíduos em verde são uma mutação conservativa e aqueles em amarelo são idênticos. As bases que obedecem uma simetria palindrômica no sítio de DNA da ElrR e da NrtR estão coloridas em amarelo à esquerda do centro de simetria, e em verde à direita do centro. b) estrutura resolvida do sítio de DNA da NrtR, colorida em verde e amarelo de acordo com as posições das bases que preservam a simetria das bases no sítio da ElrR, como indicado na tabela em a). O par de bases central está colorido em vermelho.

Fonte: Elaborada pela autora.

inserção no sulco maior e grande probabilidade de formar interações específicas com as bases de DNA –, ocupa a segunda posição, destacada em vermelho, em ambas as seqüências mostradas na [Figura 13a](#) para a ElrR e a NrtR, sendo ambos resíduos básicos com uma longa cadeia lateral. O complexo com a NrtR indica que apenas mais um resíduo da hélice de reconhecimento faz uma ligação de hidrogênio com uma base no sítio de DNA: é uma arginina que ocupa a sexta posição, destacada em vermelho, na seqüência mostrada na [Figura 13a](#). Ela está localizada exatamente uma volta atrás do resíduo principal, e na seqüência da ElrR também existe uma arginina. Vizinha ao resíduo principal existe uma serina em ambas as proteínas. Nesta posição só caberia um resíduo polar de cadeia curta, pois o complexo com a NrtR indica que existe um fosfato distante apenas 6,0 Å de seu carbono α . A Tyr ou a Phe logo após estas serinas são ambas aromáticas e estão mais próximas do centro do DBD em cada proteína, estabilizando seus respectivos domínios estruturais. Na quinta posição, existem resíduos polares de cadeia lateral longa, e poderiam interagir igualmente bem com a cadeia de açúcar e fosfato. Na sétima posição da hélice da NrtR, como mostrado na seqüência da [Figura 13a](#), existe mais uma arginina, que interage com um fosfato e com uma região desordenada que interage especificamente com bases do DNA pelo sulco menor. No entanto, a ElrR não possui essa região e na sétima posição existe uma Phe, também apontando para o centro do DBD. A oitava posição em ambas as proteínas é ocupada por valinas, que estão novamente orientadas para o centro do DBD. Na nona posição aparecem resíduos polares de longa cadeia lateral, uma glutamina ou um glutamato. A NrtR ainda apresenta mais dois resíduos no C-terminal da hélice de

reconhecimento, e se conecta a uma folha β curta com uma longa volta, que interage especificamente com bases do DNA pelo sulco menor, mas essa região mais uma vez não existe na ElrR.

Os sítios de DNA de ambas foram identificados com um número ímpar de bases e o par de bases AT no centro. Este par central também é o centro da simetria palindrômica, que corresponde diretamente à simetria entre o DBD de cada protômero (as duas proteínas são diméricas). A sequência do sítio de DNA da NrtR é inteiramente palindrômica, enquanto a do sítio de ElrR não é, como indica a [Figura 13a](#). Isso pode estar relacionado com as regiões extras da NrtR que interagem especificamente com as bases do seu sítio, e que a ElrR não possui, mas também pode ter outras causas, por exemplo, evolutivas. Considerando que o par de bases central esteja na posição 0, como mostra a [Figura 13a](#), os pares de base nas posições de 2 a 5 e de 8 a 11 são os que preservam a palindromia na ElrR. Os pares de bases nas posições de 2 a 5 estão localizados no sulco maior, como mostra a [Figura 13b](#), interagindo diretamente com resíduos da hélice de reconhecimento da NrtR, e em ambas as sequências os pares de bases AT e CG estão igualmente intercalados. Já as bases nas posições de 8 a 11 estão localizadas no sulco menor, e apresentam somente pares AT nas duas sequências. Este arranjo similar de bases pode indicar não apenas interações semelhantes com a proteína, como também uma torção parecida da dupla hélice de DNA.

Esta grande quantidade de características semelhantes sugere que as hélices de reconhecimento da NrtR e da ElrR tenham a mesma posição e orientação em relação ao DNA. Portanto, nós usamos as hélices de reconhecimento da NrtR, cuja estrutura resolvida é identificada pelo código PDB 3gz6, como alvo para a movimentação dos corpos rígidos.

5.2.2 A decomposição e a movimentação em corpos rígidos

Três análises diferentes, além da observação de algumas fortes interações presentes na estrutura apo resolvida, nos levaram a dividir a ElrR em três corpos rígidos. Estas análises são: a observação qualitativa dos parâmetros de deslocamento atômico anisotrópicos (ADP); a observação de alguns modos normais de mais baixa frequência, com exceção dos 6 primeiros; e os resultados obtidos pela divisão de domínios dinâmicos feita pelo DynDom na comparação de duas estruturas cristalográficas da ElrR resolvidas em distintos grupos espaciais. A [Figura 14](#) mostra o sétimo modo normal (de mais baixa frequência), a representação elipsoidal dos ADPs e a decomposição proposta pelo DynDom.

Da análise feita pelo DynDom, que considera apenas uma cadeia por vez, apareceram duas regiões de “dobradiça”, do resíduo 71 ao 75 e do 180 ao 200. Elas definem um ponto de apoio para a rotação relativa entre os domínios. A segunda região compreende a hélice 11, que selecionamos como região de referência para a superposição do terceiro corpo rígido, como explicado na seção de métodos. Em torno da primeira região existe uma rotação de $16,9^\circ$ do domínio 1 em relação ao 3, e em torno da segunda região ocorre uma

Figura 14 – a) representação elipsoidal dos ADPs para todos os átomos pesados da ElrR. b) Representação do sétimo modo normal com vetores de deslocamento atômico para os carbonos α . Nas duas figuras em a) e b) os domínios 1, 2 e 3 estão coloridos respectivamente em vermelho, laranja, e amarelo, exceto pelas regiões de conexão dos domínios, que compreendem os resíduos de 63 a 77 e 175 a 198, coloridas em branco. c) tabela indicando a divisão de domínios dinâmicos que resulta da análise feita pelo DynDom, mostrando os resíduos que compreendem cada domínio, e o RMSD pela superposição de cada domínio de duas distintas conformações da ElrR.

Fonte:Elaborada pela autora.

rotação de $18,9^\circ$ do domínio 2 em relação ao 3. Nesses dois movimentos o domínio 3 é considerado fixo e resultam numa aproximação do domínio 2 em direção ao 3. Entre estes dois domínios se forma uma região exposta ao solvente que se fecha nos dois movimentos em 26,2% ou 78,9%, respectivamente.

Interessantemente, os domínios 1 e 2 participam de um *domain swapping* na dimerização da ElrR. Existem fortes interações intercatenárias entre o domínio 1 de uma cadeia e o domínio 2 da outra, e esses dois domínios parecem se mover coletivamente pelo que mostram tanto os modos normais quanto os ADPs. Assim, dividimos a ElrR em três corpos rígidos, e cada domínio descrito pelo DynDom foi designado a um corpo rígido diferente, como indicado na [Figura 15](#), e explicado na seção de métodos. A [Figura 15](#) ainda mostra um passo intermediário para a construção do modelo da ElrR ligada ao DNA, em que dois dos corpos rígidos são superpostos às hélices de reconhecimento da NrtR, além do modelo final ainda não minimizado. Originalmente, o RMSD entre as hélices de reconhecimento da estrutura cristalográfica e das hélices alvo era de 2,49 Å, enquanto que o RMSD do modelo completo ligado ao DNA em relação à estrutura original é de 4,55 Å. As hélices de reconhecimento deste modelo foram então usadas como alvo para construir outros modelos com o método de mínimos quadrados.

Figura 15 – a) estrutura cristalográfica da ElrR dividida em três corpos rígidos, circulados em preto. b) superposição de dois corpos rígidos da ElrR com a estrutura cristalográfica da NrtR ligada ao seu sítio de DNA (código PDB 3gz6), alinhando estruturalmente somente os carbonos α das hélices de reconhecimento. c) reconstrução dos três corpos rígidos para que as hélices de reconhecimento coincidam com as hélices alvo da NrtR e o terceiro domínio se posicione com o menor desvio possível das α -hélices 11, das duas cadeias concomitantemente. A estrutura do sítio de DNA da NrtR está representada em complexo com este modelo. Em toda a figura, as estruturas do complexo de código 3gz6 (NrtR e seu sítio de DNA) estão coloridas em azul. A ElrR está colorida seguindo a divisão de regiões descrita na legenda da [Figura 14](#).

Fonte: Elaborada pela autora.

5.2.3 Os *decoys*

Outra abordagem que adotamos para construir um modelo da ElrR ligada a seu sítio, foi a procura por conformações numa amostragem próxima à estrutura cristalográfica apo. A geração de *decoys* é um método rápido e mais permissivo em relação a mudanças conformacionais que o movimento de corpos rígidos, já que as mudanças não são baseadas somente na translação e rotação de grandes grupos de átomos, mas também em movimentos como os semelhantes à respiração, por exemplo. Eles são calculados como uma combinação linear de modos elásticos, com suas amplitudes escolhidas randomicamente e multiplicadas por um fator de escala para que a estrutura final atinja um certo RMSD em relação à estrutura original, no nosso caso a ElrR cristalográfica apo. Escolhemos 10 valores distintos de RMSD a ser atingido por esse método, como explicado na seção de métodos, variando de 1,5 a 6,0 Å, e para cada valor de RMSD geramos 100 estruturas. Por esta abordagem a amostragem é bastante ampla, mas apenas alguns *decoys* puderam ser selecionados como aptos a ligar ao sítio de DNA. Dentre 1000 conformações geradas, escolhemos 4 com um valor de RMSD em relação às hélices alvo menor que 1,5 Å. Estas quatro conformações

apresentam um valor de RMSD em relação a estrutura apo de 5,0, 5,5 e 6,0 Å. No entanto, quanto maior o desvio da conformação inicial, mais distorcidas são as conformações finais, com ligações e ângulos irreais, como mostra a [Figura 16a](#). Por outro lado, dentre os conjuntos de conformações calculados, conforme o RMSD em relação à estrutura apo aumenta, as hélices de reconhecimento aproximam-se mais das hélices alvo, como pode ser observado na [Figura 16b](#), pelo desvio absoluto médio dos valores de RMSD em relação às hélices alvo. O gráfico na [Figura 16b](#) mostra o RMSD médio em relação as hélices alvo e o número médio de distorções entre cada carbono α , para cada conjunto de *decoys* distribuídos segundo o RMSD em relação à conformação original, cristalográfica apo.

Figura 16 – a) um decoy do conjunto com RMSD em relação à estrutura original de 6,0 Å, com 35% de distorções nas distâncias entre carbonos α vizinhos. b) gráfico de distribuição dos *decoys* de acordo com o RMSD em relação à estrutura de origem, com o RMSD em relação às hélices alvo e ao número de distorções entre cada carbono α vizinho. Distorções são consideradas apenas quando a distância entre dois carbonos α vizinhos ultrapassa a faixa entre 3,6 e 4,1 Å.

Fonte: Elaborada pela autora.

Esta abordagem pode ser melhorada em dois aspectos diferentes. Um é a complexidade dos modos normais, já que os *decoys* foram gerados a partir de modos elásticos. Outro é a objetividade da amostragem. Para o cálculo dos *decoys* as amplitudes de cada modo dentro da combinação linear foram escolhidas ao acaso para explorar o espaço conformacional ao redor da estrutura original, sem uma direção específica. Mas podemos especificar a direção ao estipular as hélices de reconhecimento da Nrtr como alvo, e também podemos utilizar modos harmônicos provenientes de métodos mais detalhados. Na próxima seção mostramos os modos com as menores frequências obtidos por NMA e pelo método de filtro da dinâmica molecular.

5.2.4 Os modos normais e modos harmônicos filtrados de dinâmica molecular

Os modos normais têm sido usados com sucesso para descrever mudanças conformacionais com relevância biológica.⁶³ No entanto, eles são obtidos pela aproximação da superfície de energia potencial a um poço harmônico. Esta aproximação pode ser evitada se considerarmos as mudanças conformacionais que realmente ocorrem durante uma dinâmica molecular. Assim, nós também analisamos os modos harmônicos filtrados de uma trajetória simulada. A [Figura 17](#) mostra o RMSD em relação às hélices alvo de cada conformação amostradas no decorrer de uma dinâmica molecular de 100 ns. Fica claro deste gráfico, [Figura 17](#), que, durante a trajetória simulada, a ElrR em sua forma apo de fato se aproxima de uma conformação apta a se ligar ao DNA, considerando que uma orientação das hélices de reconhecimento compatível com a disposição dos sulcos maiores é um passo primário para a ligação ao DNA.⁸

Figura 17 – Gráfico do RMSD em relação às hélices alvo em função do tempo. As conformações foram guardadas a cada 0,5 ps durante uma simulação de dinâmica molecular de 100 ns.

Fonte: Elaborada pela autora.

Também consideramos para maiores análises a conformação cujas hélices de reconhecimento mostraram o menor valor de RMSD em relação às hélices alvo. Entretanto, o RMSD muda rapidamente no período em que esta conformação aparece. Então, é provável que ela seja proveniente de mudanças mais locais e transitivas. Estas mudanças

podem não afetar coletivamente o sítio alostérico de modo a interagir otimamente com o ligante alostérico. Assim, nós recuperamos desta dinâmica os modos harmônicos de mais baixa frequência, para então combiná-los e gerar outros modelos, que provavelmente compreendam mudanças também no sítio alostérico. A [Tabela 4](#) compara os 20 modos harmônicos filtrados e os 20 modos normais com as menores frequências obtidas. Eles estão listados e ordenados de forma crescente quanto à frequência calculada, desconsiderando os 6 primeiros autovalores obtidos por NMA, que correspondem a movimentos de translação e rotação da proteína inteira. A [Tabela 4](#) ainda relaciona o grau de coletividade e o produto escalar com respeito ao vetor diferença entre as hélices de reconhecimento alvo, posicionadas como na estrutura da NrtR, e as hélices em sua conformação original, como resolvida cristalograficamente. Este vetor diferença denominamos de vetor da transição esperada. Também estão mostradas as amplitudes obtidas para os modos harmônicos filtrados, considerando tanto todos os carbonos α da proteína, quanto aqueles somente das hélices de reconhecimento.

A [Tabela 4](#) mostra que o primeiro modo harmônico filtrado corresponde ao movimento que ocorre na dinâmica com a maior amplitude, mas ele não teve o tempo para se repetir durante a simulação, ou mesmo de ser completado. Assim, nesta simulação, o modo básico da proteína, com a frequência natural mais baixa, é pobremente representado. Entretanto, existem modos com alto valor de produto escalar com o vetor da transição esperada. O modo harmônico com o vetor que mais se aproxima do vetor da transição esperada é o modo de número 15, com uma frequência de 114,4 MHz, o que significa que este modo se repetiu durante a trajetória simulada a cada 8,7 ns. O segundo modo harmônico filtrado mais similar à transição esperada ocorre com uma frequência mais baixa, repetindo-se a cada 32,8 ns, e também é o modo com o segundo maior valor de amplitude para o movimento das hélices de reconhecimento.

As frequências obtidas por NMA, ao contrário, não dependem da amostragem em uma dinâmica molecular. Teoricamente, a frequência mais baixa obtida por NMA deveria ser a frequência natural mais baixa da proteína, mas o valor obtido é muito maior que o observado para modos recuperados da simulação de dinâmica molecular. De fato, a aproximação harmônica para a energia potencial afeta profundamente o valor calculado para as frequências.⁶³ De qualquer forma, alguns dos modos normais apresentam os maiores valores para o grau de coletividade, corroborando com as observações de que eles se assemelham a movimentos com relevância funcional biológica, apesar das frequências errôneas. Contudo, eles não possuem valores para o produto escalar com o vetor da transição esperada tão altos quanto os obtidos para os modos harmônicos filtrados.

Tabela 4 – Relação dos 20 modos normais e modos harmônicos filtrados com os menores valores de frequência (Freq.). Eles foram enumerados em ordem crescente de acordo com o aumento na frequência. Os primeiros 6 modos normais foram desconsiderados. Estão relacionados também o grau de coletividade (Col.) e o produto escalar (PE) com o vetor diferença entre as hélices de reconhecimento alvo, e sua posição original na conformação cristalográfica. Para os modos harmônicos filtrados também estão relacionadas as amplitudes do movimento, considerando tanto todos os carbonos α (amplitude global - Amp.global), quanto somente aqueles das hélices de reconhecimento (amplitude local - Amp.local).

modos normais				modos harmônicos filtrados				
modo	Freq.(GHz)	PE	Col.	Freq.(MHz)	Amp.global(Å)	Amp.local(Å)	PE	Col.
1	39,69	0,34	0,44	7,6	34,17	4,58	-0,41	0,59
2	41,97	0,17	0,50	15,3	16,60	1,67	-0,01	0,52
3	52,56	0,42	0,39	22,9	15,54	2,43	-0,08	0,57
4	69,21	0,24	0,33	30,5	13,31	2,50	0,83	0,53
5	76,56	-0,20	0,46	38,1	8,96	1,17	-0,62	0,49
6	82,86	-0,01	0,63	45,8	7,73	0,95	-0,08	0,53
7	86,97	-0,32	0,78	53,4	8,31	1,44	-0,77	0,61
8	90,48	0,31	0,68	61,0	6,40	1,35	-0,48	0,65
9	97,50	-0,21	0,45	68,7	11,54	2,11	-0,63	0,48
10	99,72	0,27	0,73	76,3	8,57	1,77	-0,44	0,66
11	108,48	-0,29	0,60	83,9	7,48	0,35	0,04	0,41
12	117,24	0,25	0,67	91,5	5,96	0,85	0,79	0,49
13	122,61	0,31	0,62	99,2	4,65	0,57	0,16	0,47
14	130,32	-0,15	0,70	106,8	5,30	0,69	0,33	0,31
15	137,22	0,53	0,71	114,4	5,73	0,53	0,92	0,55
16	146,07	-0,07	0,47	122,1	3,94	0,61	0,21	0,68
17	148,35	-0,17	0,32	129,7	4,92	0,56	-0,63	0,47
18	151,41	0,43	0,24	137,3	4,12	0,44	0,28	0,57
19	156,39	-0,64	0,62	145,0	5,48	0,61	-0,51	0,49
20	160,32	0,54	0,52	152,6	4,60	0,37	-0,47	0,59

Fonte – Elaborada pela autora.

5.2.5 A combinação linear dos modos no método de mínimos quadrados

Como explicado na seção de métodos, com os 20 modos provenientes ou de NMA ou do método de filtro, nós consideremos 220 modelos possíveis. Eles são compostos pela combinação de diferentes modos, considerando no mínimo um modo e no máximo

20, sem repetir nenhum. Alguns modelos mostraram uma boa aproximação às hélices de reconhecimento, enquanto outros mostraram graves distorções na estrutura final da proteína. O melhor modelo deve apresentar as hélices de reconhecimento tão próximas quanto possível do alvo, e a menor quantidade de distorções possível, considerando a estrutura inteira, e também partes dela. A [Figura 18](#) mostra a distribuição de modelos de acordo com o RMSD em relação às hélices alvo e um segundo parâmetro que quantifica as distorções geradas, como o RMSD em relação à estrutura inicial, o número de distorções nas distâncias entre carbonos α vizinhos, ou a porcentagem de estrutura secundária preservada.

Fica claro pelos gráficos na [Figura 18](#) que os modelos obtidos pela combinação de modos harmônicos filtrados fornecem os menores valores de RMSD em relação às hélices de reconhecimento alvo, enquanto também não desviam demasiadamente da conformação inicial. Os modos harmônicos filtrados também resultaram em menores porcentagens de distorções na distância entre carbonos α vizinho, ao mesmo tempo em que também aproximam as hélices de reconhecimento ao alvo o máximo possível, assim como também preservam mais de sua estrutura secundária, quando comparados com os modelos obtidos pela combinação de modos normais.

Um modelo da combinação de modos de cada método foi selecionado para ser minimizado e então servir de apoio para o desenho de possíveis ligantes alostéricos, mostrados na [Figura 18d](#) e na [Figura 18e](#). Eles foram escolhidos pela consideração somente dos valores de RMSD em relação às hélices de reconhecimento alvo, e à conformação inicial, mas os outros parâmetros também revelam valores razoáveis.

O modelo construído com modos normais apresenta um RMSD das hélices de reconhecimento alvo de 1,19 Å, RMSD da conformação inicial de 5,9 Å, porcentagem resíduos em hélices α conservadas de 96,8% e porcentagem de distorções encontradas nas distâncias entre carbonos α vizinhos de 52,2%. Ele foi obtido usando 11 modos (de números 1, 2, 3, 6, 8, 9, 10, 12, 15, 16, e 20). Este conjunto de modos normais compreende os três com menor frequência, dois com alto valor de produto escalar com o vetor de transição esperada para as hélices de reconhecimento, e dois com alto grau de coletividade. E outros modos com nenhuma dessas qualidades também participam deste conjunto.

O modelo construído a partir de modos harmônicos filtrados apresenta um RMSD das hélices de reconhecimento de 0,81 Å, RMSD da conformação inicial de 4,9 Å, porcentagem das hélices α preservadas de 95,5%, e porcentagem de distorções na distância entre carbonos α vizinhos de 42,3%. Ele foi construído usando apenas 9 modos (de números 3, 4, 6, 7, 8, 9, 10, 12, e 15). Fazem parte deste grupo os modos com o maior valor de produto escalar com o vetor transição das hélices de reconhecimento, e também alguns modos com alto grau de coletividade, enquanto a frequência dos modos ficou principalmente na faixa de 22 a 92 MHz. O modo com o maior valor para o produto escalar é o único cuja frequência fica fora desta faixa, sendo de 114,4MHz. Isso sugere que o movimento que mais contribui

Figura 18 – Os gráficos em a), b) e c) mostram a distribuição dos modelos de acordo com o RMSD em relação às hélices alvo e um segundo parâmetro que quantifica as distorções geradas, como o RMSD em relação à estrutura inicial, o número de distorções nas distâncias entre carbonos α vizinhos, ou a porcentagem de estrutura secundária preservada, respectivamente. Os modelos construídos pela combinação linear de modos normais estão representados em cinza, enquanto que os construídos por modos harmônicos filtrados estão representados em preto. d) modelo construído a partir de modos harmônicos filtrados para seguir nas análises. e) modelo construído a partir de modos normais para continuar com análises. Os modelos mostrados estão representados tanto em cartoon quanto em linhas, que também representa as cadeias laterais.

Fonte: Elaborada pela autora.

para a transição esperada das hélices de reconhecimento ocorra com alta frequência, mas dependa de outros modos com frequência menor para atingir a conformação que melhor se encaixa nos sulcos maiores do DNA. Este comportamento é consistente com a rápida mudança no RMSD em torno da conformação ligante, como indica a [Figura 17](#), e o fato de que raramente essa conformação ligante apareça durante a trajetória simulada.

5.2.6 A análise dos modelos selecionados em cada método

Analizamos nesta seção os modelos, que selecionamos pelos cinco métodos que usamos, um pela superposição de corpos rígidos, quatro *decoys*, uma conformação recuperada da dinâmica, um modelo gerado pela combinação linear de modos harmônicos filtrados, e outro pela combinação de modos normais. Cada método difere principalmente na complexidade e grau de detalhes considerado nos cálculos. Apesar da simplicidade ser preferida, algumas características dependem de maior acurácia. Por exemplo, o modelo obtido por superposição de corpos rígidos incorre em muitos espaços vazios no interior da proteína, que provavelmente causariam grande instabilidade, com mostra a [Figura 19](#). Este e os próximos modelos que discutiremos passaram por uma minimização de energia, como explicado na seção de métodos.

Figura 19 – a) modelo gerado pela superposição de corpos rígidos. A cadeia principal está representada em cartoon e colorida como descrito na [Figura 14](#). A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza. b) outro ponto de vista indicando o sítio alostérico com uma seta vermelha.

Fonte: Elaborada pela autora.

Os quatro modelos de *decoys* resultaram em grandes desvios na região C-terminal. Esta região mostrou os menores ADPs experimentais, tanto isotrópicos quanto anisotrópicos, assim como baixo deslocamento por NMA e o método de filtro, e um RMSD de 0,49 Å na decomposição de domínios feita pelo DynDom. No entanto, o RMSD de *decoy* para esta região quando comparados com a conformação original resultou em 1,6 Å, 2,1 Å, 2,3 Å, e 2,4 Å. O modelo com o menor RMSD ainda poderia ser considerado, mas ele mostrou sítios alostéricos demasiadamente abertos.

A conformação retirada da MD também mostrou sítios alostéricos amplos, como pode ser observado na [Figura 20](#), e não é possível prever um ligante através dela.

Figura 20 – Conformação retirada da MD quando as hélices de reconhecimento mais se aproximaram do alvo. As duas perspectivas mostram os dois sítios alostéricos, indicados por uma flecha vermelha. A cadeia principal está representada em cartoon e colorida como descrito na [Figura 14](#). A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza.

Fonte: Elaborada pela autora.

Por fim, os dois modelos obtidos da combinação linear de modos normais ou modos harmônicos filtrados resultaram em conformações do sítio alostérico mais fechadas. Pela superposição da proteína PrgX, homóloga à ElrR, é possível observar que o peptídeo, cuja estrutura foi resolvida em complexo com a PrgX, se encaixa no sítio alostérico destes modelos, como mostrado na [Figura 21](#).

Os dois modelos poderiam acomodar um peptídeo ligante no sítio alostérico, mas eles o envolvem de diferentes maneiras, impedindo o acesso do solvente com diferentes contatos entre os domínios 2 e 3. A [Figura 22](#) mostra mapas de contato desses dois modelos e da estrutura cristalográfica apo.

A [Figura 22](#) mostra mapas de contato indicando os carbonos α com até 12 Å entre eles. O padrão na diagonal dos três mapas é consistente com a sequência de hélices α . As curtas linhas em paralelo ao lado da linha central indicam contatos entre hélices em paralelo, enquanto as linhas perpendiculares indicam contatos entre hélices α em antiparalelo. O padrão central é preservado nos dois modelos, o que indica que as estruturas secundárias e terciárias foram mantidas. Os quadrantes em que os padrões de hélices estão ausentes correspondem a contatos intercatenários. Na conformação apo, esses contatos intercatenários são principalmente responsáveis pela dimerização e estabilização do *domain swapping*, mas outros contatos aparecem além destes nos modelos construídos por combinação linear dos modos. A [Figura 22](#) resalta esses contatos circulados em branco. As diferenças encontradas residem na região entre os resíduos 80 e 90, na maioria. Essa região corresponde à primeira hélice α do domínio 2, a α -hélice 6. No modelo construído pela combinação de modos normais, esta hélice aparece próxima das α -hélices 10 e 11 da

Figura 21 – a) e b) modelo obtido pela combinação de modos harmônicos filtrados. c) e d) modelo obtido pela combinação de modos normais. As duas perspectivas mostram os dois sítios alostéricos. A cadeia principal está representada em cartoon e colorida como descrito na [Figura 14](#). A superfície eletrônica também está representada com certa transparência e colorida de acordo com a cadeia em branco ou cinza. O peptídeo resolvido em complexo com a proteína PrgX (código PDB 2axz), superposta com cada modelo, aparece em cartoon colorido em azul.

Fonte: Elaborada pela autora.

sua própria cadeia, fechando o topo do sítio alostérico. Já no modelo construído com modos harmônicos filtrados, a α -hélice 6 se encontra próxima do *loop* que conecta a α -hélice 11 do domínio 3 da outra cadeia, fechando o sítio alostérico de outra forma. Apesar dessas diferenças, alguns resíduos que compõem o sítio alostérico são comuns nos dois modelos, e eles também estão presentes nas proteínas homólogas PrgX e PlcR. Ambas interagem com peptídeos curtos, mais ou menos inseridos dentro do sítio alostérico. Nos modelos para a ElrR, o sítio alostérico parece formar um bolsão mais fundo, mas o peptídeo não deve ter mais de 7 resíduos. Agora, uma biblioteca pode ser escrita considerando as similaridades entre os modelos e as proteínas homólogas, mas também suas divergências.

5.3 Sobre a afinidade da MosR pelo DNA

Abaixo mostramos os resultados de quatro simulações, cada uma com 400 ns. Elas diferem na posição inicial dos N-terminais ou no estado de oxidação da proteína. Por estado de oxidação nos referimos à existência de duas pontes dissulfeto intracatenárias, construídas como explicado na seção de métodos, entre os resíduos Cys10 e Cys12. No estado oxidado as duas pontes estão formadas, enquanto que no estado reduzido nenhuma está. As simulações 1 e 2 estão no estado reduzido, enquanto as simulações 3 e 4 estão no estado oxidado. Por outro lado, as simulações 1 e 3 partem da mesma conformação,

Figura 22 – Mapas de contato a) da conformação cristalográfica apo, b) do modelo construído pela combinação de modos harmônicos filtrados, e c) do modelo construído pela combinação de modos normais. d) Legenda de cores para as distâncias entre carbonos α . e) Conformação cristalográfica apo ressaltando os resíduos de número 45, 75, 180 e 200, também indicados nos eixos dos mapas de contato.

Fonte: Elaborada pela autora.

enquanto que as simulações 2 e 4 partilham uma outra conformação inicial.

Os resultados a seguir são uma tentativa de caracterizar o que é comum e o que é distinto na forma apo da MosR em ambos os estados, e, ademais, como as características diferentes afetam a afinidade por seu sítio de DNA.

A [Figura 23](#) mostra ambos os estados simulados, calculados como as conformações médias das simulações 1 e 2 para o estado reduzido, e das simulações 3 e 4 para o estado oxidado, superpostos com as estruturas cristalográficas da MosR em sua forma apo e holo, ligada ao DNA. Complementarmente, o [Quadro 1](#) indica o RMSD entre as conformações mostradas na [Figura 23](#), e também o RMSD entre suas hélices de reconhecimento. Apesar de simulados em estados de oxidação diferentes, ambas as médias simuladas apresentam um valor mais baixo de RMSD quando comparadas à estrutura cristalográfica apo que quando comparadas à estrutura holo, o que fica ainda mais evidente considerando apenas suas hélices de reconhecimento. Ademais, a conformação média das quatro simulações juntas apresenta um RMSD em relação à estrutura apo de apenas 1,0 Å. Não obstante, algumas

Figura 23 – Quatro conformações da MosR superpostas, ressaltando o posicionamento das hélices de reconhecimento, indicadas por uma espessura maior do *ribbons*, em duas orientações diferentes. Em verde está representada a conformação média calculada para o estado reduzido, considerando duas simulações de 400 ns, em azul está representada a conformação média calculada para o estado oxidado, considerando duas simulações de 400 ns, em rosa está representada a estrutura cristalográfica apo (código PDB 4fx0) e em amarelo está representado a estrutura cristalográfica holo (código PDB 4fx4), com seu DNA representado em laranja.

Fonte: Elaborada pela autora.

diferenças podem ser observadas entre os estados oxidado e reduzido, em concordância com as diferenças também observadas entre as estruturas cristalográficas.²² A formação da ponte dissulfeto no início da α -hélice 1 causa uma tensão extra na volta que a sucede, adotando uma estrutura ligeiramente diferente, como mostra a [Tabela 5](#), que indica a distância entre os carbonos α dos resíduos Cys10 e Thr14 nas conformações cristalográficas e nas médias calculadas para as simulações. De fato, hélices 3/10 são mais propensas a formarem pontes dissulfeto intracatenárias que α -hélices,⁹⁶ e outras mudanças conformacionais envolvendo uma ponte dissulfeto com cisteínas em posição similar já foram observadas, por exemplo, para a tiorredoxina redutase.⁹⁷ Além disso, uma região apolar aparece em torno da ponte dissulfeto, que repele resíduos carregados próximos às cisteínas. De fato, a Asn37 se encontra mais distante da Cys12 da outra cadeia no estado oxidado simulado e na estrutura cristalográfica apo, que no estado reduzido simulado e na estrutura cristalográfica holo, como também indicado na [Tabela 5](#).

Contudo, para as hélices de reconhecimento, observamos uma rotação máxima de apenas 5,9° (ver [Figura 23](#)) entre as médias calculadas para os estados reduzido e oxidado. Assim, o próximo conjunto de resultados avalia se as simulações podem reproduzir os dados experimentais em relação à afinidade da MosR pelo DNA, cujo complexo não se forma no estado oxidado. Poderíamos avaliar também outras diferenças, além da orientação das hélices de reconhecimento, que foram observadas entre as duas estruturas cristalográficas, mas algumas delas podem ter sido causadas pela interação com o DNA, sendo difícil atribuí-las ao estado da proteína, reduzido ou oxidado. De qualquer forma, tem-se encontrado

Quadro 1 – RMSDs entre as quatro conformações mostradas na [Figura 23](#), estrutura cristalográfica apo (código PDB 4fx0), estrutura cristalográfica holo (código PDB 4fx4), conformação média calculada para o estado reduzido simulado e conformação média calculada para o estado oxidado simulado. Dois valores de RMSD estão mostrados para cada par de conformações comparadas, primeiro considerando todas os carbonos α , e depois somente aqueles pertencentes às hélices de reconhecimento.

RMSD (Å)	estrutura apo	estrutura holo	conformação média oxidada
conformação média reduzida	1,112 – 0,621	1,887 – 3,076	1,665 – 0,873
conformação média oxidada	1,369 – 0,693	2,684 – 3,214	
estrutura holo	1,520 – 3,092		

Fonte: Elaborada pela autora.

Tabela 5 – Distâncias atômicas observadas nas quatro conformações da MosR mostradas na [Figura 23](#), estrutura cristalográfica apo (código PDB 4fx0), estrutura cristalográfica holo (código PDB 4fx4), conformação média calculada para o estado reduzido simulado e conformação média calculada para o estado oxidado simulado. Duas distâncias são mostradas para cada conformação em relação à cadeia A e outras duas em relação à cadeia B. A primeira é a distância intracatenária da cisteína 10 à treonina 14, e a segunda é a distância intercatenária entre a cisteína 12 e asparagina 37 da cadeia oposta.

	estrutura apo	estrutura holo	média oxidada	média reduzida
cadeia A	6,2 – 7,7	6,5 – 7,4	7,3 – 8,6	5,9 – 6,9
cadeia B	6,1 – 8,0	6,1 – 7,6	7,5 – 8,3	6,0 – 7,1

Fonte – Elaborada pela autora.

que as hélices de reconhecimento são as primeiras a interagir com o DNA,⁸ portanto, as mudanças na posição e orientação delas devem ocorrer previamente a outras interações com o DNA. Assim, detemos nossas análises sobre as hélices de reconhecimento para caracterizar a capacidade da proteína a se ligar ao DNA.

5.3.1 As hélices de reconhecimento

A [Figura 24](#) mostra o RMSD das hélices de reconhecimento em função do tempo simulado para cada MD em relação à estrutura cristalográfica ligada ao DNA (código PDB 4fx4). Fica claro que a simulação 1, no estado reduzido, apresenta o conjunto de conformações que mais se aproxima da estrutura cristalográfica holo (o menor RMSD observado foi de 0,9 Å). No entanto, a simulação 3 também apresenta algumas conformações que consideravelmente se aproximam da estrutura holo (o menor RMSD observado foi de 1,3 Å). A simulação 2 também apresenta um menor RMSD de 1,3 Å, enquanto que a

simulação 4 nunca apresenta um RMSD menor que 1,5 Å. Estes resultados levam à certa ambiguidade nas conclusões pois a maioria das simulações apresenta uma aproximação das hélices de reconhecimento à posição de encaixe nos sulcos do DNA.

Figura 24 – RMSD das hélices de reconhecimento comparadas às da estrutura cristalográfica holo. As quatro trajetórias simuladas estão representadas. Em verde escuro a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4.

Fonte: Elaborada pela autora.

Apesar de valores baixos de RMSD certamente representarem conformações semelhantes, quanto maiores os valores, menos informação temos sobre a origem dessas diferenças. Será que elas são originadas de mudanças conformacionais coletivas, ou apenas de deformações locais? Por exemplo, a conformação de menor RMSD para as hélices de reconhecimento que aparece na simulação 1 possui um RMSD global para a proteína de 1,993 Å em relação à estrutura holo, já a que aparece na simulação 2 possui um RMSD global de 1,98 Å, enquanto que a conformação que aparece na simulação 3 possui um RMSD global de 2,882 Å. Com o objetivo de apresentar uma descrição mais informativa, calculamos a orientação relativa das hélices de reconhecimento através de três parâmetros (ver a seção de métodos). O produto escalar direto entre as duas hélices de reconhecimento não separou bem as conformações retiradas das simulações 1 e 3, assim, a [Figura 25](#) mostra somente os outros dois parâmetros: a distância entre as duas hélices, cuja direção do vetor que a representa é normal aos eixos das duas hélices; e a distância entre os centros

de massa, como a projeção do vetor diferença entre os centros de massa numa direção ortogonal à primeira e paralela às hélices. A estrutura cristalográfica holo também está representada na [Figura 25](#) por um círculo laranja. Com esta análise fica claro como as conformações da simulação 1 são as que mais se aproximam da cristalográfica. Também, a probabilidade de atingir uma conformação ligante do sítio de DNA é muito maior nas simulações de estado reduzido que nas de estado oxidado. Ainda assim, as probabilidades de atingir a conformação ligante são baixas mesmo para o estado reduzido, o que pode indicar a existência de outros efeitos reguladores complementares, por exemplo os de um ligante alostérico, como mencionado anteriormente.²²

Figura 25 – Distribuição das conformações retiradas das simulações de acordo com a distância relativa entre as hélices de reconhecimento durante as trajetórias simuladas em comparação com a estrutura cristalográfica holo. No eixo das abscissas está a distância entre as hélices, cujo vetor possui uma direção normal às hélices. No eixo das ordenadas está a distância entre os centros de massa (CM), definida pela projeção do vetor diferença entre os CMs num plano paralelo às hélices. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3, em azul claro a simulação 4, e em laranja a estrutura cristalográfica holo.

Fonte: Elaborada pela autora.

Outra descrição pode ser feita para confirmar este resultado. Na [Figura 26](#) as conformações simuladas estão distribuídas de acordo com o produto escalar entre cada hélice de reconhecimento e sua respectiva hélice na estrutura cristalográfica holo. Assim,

apesar do RMSD entre as hélices ser baixo para a simulação 3 (Figura 24), fica claro mais uma vez como as conformações da simulação 1 são as que mais se assemelham à da estrutura cristalográfica holo.

Figura 26 – Distribuição das conformações retiradas das simulações de acordo com os ângulos entre cada hélice de reconhecimento e sua respectiva na estrutura cristalográfica holo. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4. o ponto de coordenadas (0,0) corresponderia à própria estrutura cristalográfica holo.

Fonte: Elaborada pela autora.

Interessantemente, as conformações na simulação 1 também estão distribuídas de forma mais ampla que nas outras simulações (ver Figura 25 e Figura 26) no espaço definido pelos parâmetros selecionados, indicando uma maior variedade de orientação para as hélices de reconhecimento. A Figura 27 compara a flexibilidade da MosR em cada simulação representada pelo ADP isotrópico calculado para cada janela de tempo de 1 ns e cada 50 ns. Para 1 ns, a MosR apresenta quase o mesmo padrão de flexibilidade para todas as simulações. Já para a janela de 50 ns, fica claro que a MosR nas simulações 1 e 3, que partilham da mesma conformação inicial, é mais flexível que nas outras duas simulações. A flexibilidade das hélices do DBD é ainda maior para a cadeia B na simulação 1. As diferentes janelas de tempo permitem que se distingam movimentos locais e coletivos. Assim, os resultados indicam que a variedade de orientações para as hélices de reconhecimento observada na simulação 1 são provenientes de movimentos coletivos do DBD da cadeia B.

Figura 27 – ADPs isotrópicos calculados para cada simulação e separadamente para cada cadeia em diferentes janelas de tempo: a) média dos valores obtidos as janelas de 1 ns; b) média dos valores obtidos para as janelas de 50 ns. Em verde escuro está representada a simulação 1, verde claro a simulação 2, azul escuro a simulação 3 e em azul claro a simulação 4.

Fonte: Elaborada pela autora.

Este conjunto de resultados mostra, por fim, que a MosR é mais provável de atingir uma conformação ligante de DNA no estado reduzido, já a simulação 1 neste estado apresenta maior flexibilidade (Figura 27 e também maior proximidade à conformação cristalográfica holo (Figura 25 e Figura 26). Entretanto, a MosR na simulação 2 também estava em estado reduzido, mas não mostrou capacidade equivalente de adotar uma conformação ligante de DNA. Assim, também podemos concluir que a conformação do N-terminal esteja afetando o espaço acessível das hélices de reconhecimento. O próximo conjunto de resultados trata desta possibilidade.

5.3.2 Sobre a participação do N-terminal

As coordenadas do N-terminal não puderam ser resolvidas nas estruturas cristalográficas, indicando que não há uma conformação preferencial. Assim as conformações selecionadas para início das MDs são apenas casos particulares de conformações possíveis de se visitar. A Figura 28 mostra o mapa de contatos dos dois N-terminais da MosR em cada simulação. Somente os contatos entre os primeiros 18 resíduos de uma cadeia e todos os resíduos da cadeia oposta são considerados. Um contato é contado para cada vez que dois carbonos α ficam a menos de 6 Å um do outro. Das estruturas cristalográficas que não apresentam os N-terminais completos, espera-se que eles sejam bastante flexíveis e possam interagir de forma transitiva com um grande número de resíduos em sua vizinhança. Assim, a contagem de contatos pode variar de uma simulação para outra. Se estas interações mudassem com rapidez e as simulações fossem longas o suficiente, então o padrão de

contatos seria semelhante para todas as simulações. Por exemplo, em todas as simulações o resíduo 4 aparece perto dos resíduos da α -hélice 2' e do loop que conecta à α -hélice 3' da outra cadeia, mas com taxas ligeiramente diferentes. O resíduo 3 também aparece sempre com alguma taxa de interação com resíduos da α -hélice 5', exceto o da cadeia A na simulação 1, em estado reduzido. Por outro lado, este resíduo 3 da cadeia A na simulação 1 é o único que faz contato com o resíduo 64' na α -hélice 3', a primeira do HTH. Na [Figura 29](#), estão ressaltadas as interações do N-terminal da cadeia A nas simulações 1 e 3, que partilham da mesma conformação inicial, mas correspondem a estados de oxidação diferentes, reduzido e oxidado, respectivamente. Apesar de ambos mostrarem contatos semelhantes com a α -hélice 3', no estado reduzido observamos contatos que parecem impossíveis no estado oxidado. As conformações retiradas das simulações 1 e 3 mostradas na [Figura 29c](#) e na [Figura 29d](#) são de fato distintas. No estado reduzido (simulação 1), o N-terminal consegue se interpor entre a α -hélice 1 e α -hélice 3', o que não acontece no estado oxidado (simulação 3). Em concordância com os dados experimentais, a conformação cristalográfica holo mostra uma distância maior entre essas duas hélices que a conformação apo, como representado na [Figura 29e](#). Procurando especificamente por contatos polares entre os resíduos do N-terminal e a α -hélice 3', encontramos que a simulação 1 foi a única a apresentar a ligação de hidrogênio entre os átomos da cadeia principal da Ala6 da cadeia A e da Arg65' da cadeia B, como indicado na [Figura 29c](#).

Para verificar o comportamento desta interação específica encontrada na simulação 1, entre a Ala6 e a Arg65', monitoramos a distância entre o nitrogênio da Ala6 e o oxigênio da Arg65'. A [Figura 30](#) mostra essas distâncias em função do tempo para os dois N-terminais da MosR na simulação 1 e na simulação 3. Somente na simulação 1 essa distância é tão curta quanto uma ligação de hidrogênio. Essa interação não é permanente, mas é estável durante metade da simulação. Coincidentemente, o menor valor de RMSD encontrado para as hélices de reconhecimento em relação à posição delas na estrutura cristalográfica holo é encontrado também durante a primeira metade da simulação 1. Já para a simulação 3, essa ligação não chega a se formar, mas os dois átomos mantêm uma distância estável, cerca de 5 Å, durante mais da metade da simulação, que parece indicar uma ligação de hidrogênio mediada por uma molécula de água.

Em seguida verificamos se esta interação poderia afetar a flexibilidade do DBD, visto que a cadeia B da simulação 1 já se mostrou a mais flexível (ver [Figura 27](#)). Para isso dividimos a simulação 1 em duas metades, e a [Figura 31](#) mostra os ADPs isotrópicos calculados para cada parte. Desta análise fica evidente que a cadeia B na primeira metade da simulação é mais flexível que na segunda. A [Figura 32](#) ainda mostra os ADPs anisotrópicos calculados para cada metade da simulação, confirmando que a flexibilidade do DBD da cadeia B, com movimentos coletivos, é fundamental para que seja alcançada a posição das hélices de reconhecimento que se encaixe nos sulcos maiores do DNA. Em suma, observamos que a ligação de hidrogênio direta entre Ala6 do N-terminal e Arg65' do HTH

Figura 28 – Mapas de contato entre o N-terminal e a cadeia oposta. Um contato é considerado quando dois carbonos α ficam a menos de 6 Å entre si. A escala de cores indica quantas vezes um contato foi encontrado durante as simulações, sendo no máximo 40.000 vezes, já que foram guardadas 40 mil conformações para cada trajetória de 400 ns. No eixo das abscissas estão indicados os números dos resíduos do N-terminal, enquanto no eixo das ordenadas estão os números dos resíduos da cadeia oposta. Cada mapa indica os contatos encontrados em cada N-terminal simulado. Em a) o N-terminal da cadeia A da simulação 1, b) o N-terminal da cadeia B da simulação 1, c) o N-terminal da cadeia A da simulação 2, d) o N-terminal da cadeia B da simulação 2, e) o N-terminal da cadeia A da simulação 3, f) o N-terminal da cadeia B da simulação 3, g) o N-terminal da cadeia A da simulação 4, e em h) o N-terminal da cadeia B da simulação 4.

Fonte: Elaborada pela autora.

Figura 29 – Interações do N-terminal com a α -hélice 3' da cadeia oposta. a) detalhe do mapa de contato na [Figura 28e](#), mostrando os contatos entre os resíduos de 1 a 7 do N-terminal da cadeia A com os resíduos de 62 a 70 da cadeia oposta na simulação 3, em estado oxidado. b) detalhe do mapa de contato na [Figura 28a](#), mostrando os contatos entre os resíduos de 1 a 7 do N-terminal da cadeia A com os resíduos de 62 a 70 da cadeia oposta na simulação 1, em estado reduzido. c) conformação retirada da simulação 3 aos 50 ns, mostrando a posição do N-terminal da cadeia A. d) conformação retirada da simulação 1 aos 50 ns, mostrando a posição do N-terminal da cadeia A. e) estruturas cristalográficas apo e holo superpostas e mostrando a posição relativa das α -hélices 1 e 3', e as diferentes ligações de hidrogênio feitas pela I66'. A estrutura holo está em amarelo, com o DNA em complexo colorido em laranja, e a estrutura apo está em rosa.

Fonte: Elaborada pela autora.

Figura 30 – Distância atômica intercatenária entre o nitrogênio da Ala6 e o oxigênio da Arg65 da cadeia oposta durante a) a simulação 1 e b) a simulação 3.

Fonte: Elaborada pela autora.

aumenta a flexibilidade deste domínio de ligação ao DNA.

Para avaliar em maior profundidade os movimentos da MosR em cada metade da simulação 1, recuperamos os modos harmônicos associados a baixas frequências. A [Figura 33](#) representa, com vetores de deslocamento atômico para os carbonos α , o movimento com a menor das frequências em cada metade, enquanto a [Tabela 6](#) indica a frequência de cada modo, as amplitudes global e local (somente das hélices de reconhecimento) e o grau de coletividade associado. Interessantemente, a primeira parte da simulação 1 apresenta maiores amplitudes e grau de coletividade que na segunda parte. É possível que este aumento no grau de coletividade surja do rearranjo da estrutura proveniente da interação entre Ala6 da cadeia A e Arg65' da cadeia B. De fato, a conformação média desta primeira parte da simulação 1 apresenta um deslocamento considerável da conformação média da segunda parte (ver [Figura 32](#)).

5.3.3 Impondo uma restrição

Para nos certificar de que esta interação específica de fato está associada a um aumento na flexibilidade e na afinidade pelo sítio de DNA, simulamos uma MD curta, de 100 ns, em que restringimos a distância entre o nitrogênio da Ala6 de uma cadeia e o oxigênio da Arg65' da outra cadeia. Apesar desta restrição ter sido aplicada simetricamente para cada par de átomos, apenas um par manteve-se perto o suficiente por tempo o bastante para ser considerada uma ligação de hidrogênio estável, como mostra a [Figura 34a](#). A [Figura 34](#) ainda mostra os ADPs isotrópicos calculados para esta nova trajetória, assim como os parâmetros que indicam a aproximação das hélices de reconhecimento a uma conformação similar a da estrutura cristalográfica holo, comparados com os mesmos resultados obtidos para a primeira parte da simulação 1. Este conjunto de resultados mostra que a cadeia A desta simulação restringida se comporta como a cadeia B na primeira parte da simulação 1.

Figura 31 – Parâmetros de deslocamento atômico isotrópico calculados somente para os carbonos α como médias de janelas de 50 ns. Os valores calculados para a primeira metade da simulação 1 estão em rosa e os da segunda metade estão em laranja.

Fonte: Elaborada pela autora.

Figura 32 – Parâmetros de deslocamento atômico anisotrópico representados como elipsoides somente para os carbonos α , calculados considerando os 200 ns de cada metade da simulação 1. As conformações médias de cada metade da simulação estão mostradas em a) rosa para a primeira metade e b) laranja para a segunda metade. As hélices de reconhecimento de cada conformação média foram superpostas com as hélices de reconhecimento da estrutura cristalográfica holo, e estão representadas em preto, assim como o sítio de DNA ligado a elas.

Fonte: Elaborada pela autora.

Figura 33 – Vetores de deslocamento atômico para cada carbono α representando o modo de mais baixa frequência recuperado de a) a primeira parte da simulação 1, e b) a segunda parte da simulação 1. As conformações representadas são as primeiras do período considerado para esta análise. A da primeira parte está colorida em rosa e a da segunda parte em laranja. Cada cadeia está colorida com um tom diferente.

Fonte: Elaborada pela autora.

Tabela 6 – Frequência (Freq.), grau de coletividade (Col.), amplitude global (Amp.global) e amplitude somente das hélices de reconhecimento (Amp.local) dos 5 primeiros modos recuperados de cada metade da simulação 1 (primeira parte e segunda parte).

primeira parte				segunda parte		
Freq.(MHz)	Amp.global(Å)	Amp.local(Å)	Col.	Amp.global(Å)	Amp.local(Å)	Col.
3,05	66,6	24,8	0,38	60,3	13,6	0,22
6,10	25,3	10,0	0,49	29,4	9,9	0,41
9,15	14,7	5,8	0,48	17,9	3,8	0,22
12,21	10,7	4,8	0,43	17,1	3,3	0,22
15,26	11,1	3,7	0,41	12,9	3,2	0,30

Fonte – Elaborada pela autora.

Assim, é possível associar a interação encontrada com as diferenças observadas na dinâmica da MosR, e que podem aumentar a afinidade pelo sítio de DNA. Além disso, as mudanças conformacionais também são observadas quando esta interação está presente. Por fim, a [Figura 34](#) também mostra a conformação média desta simulação restringida superposta à conformação média da primeira parte da simulação 1, e às estruturas cristalográficas. Para esta última simulação restringida, a hélice de reconhecimento da cadeia A apresenta uma rotação de $24,6^\circ$ em relação à estrutura cristalográfica apo, o que é compatível com a rotação detectada para a ligação com o DNA. Pela [Figura 34e](#) e pela [Figura 34f](#) observamos que tanto na conformação média da primeira parte da simulação 1 quanto na conformação média da simulação restringida, uma das hélices de reconhecimento permanece mais próxima da forma apo enquanto a outra passa por uma rotação no sentido de se aproximar

da forma holo. Para a simulação restringida a hélice mais próxima da forma holo é a cadeia A, que de fato passa por uma mudança conformacional de todo o DBD coletivamente, resultando num RMSD, da cadeia inteira, de apenas 1,181 Å em relação a forma holo e de 2,301 Å em relação a forma apo. Já para a primeira parte da simulação a hélice que mais se aproxima da forma holo é a da cadeia B, que apresenta um RMSD de 0,942 Å em relação a forma holo e de 1,388 Å em relação a forma apo.

5.4 Sobre os modos coletivos da DEA

Uma das vantagens do método de filtro de modos harmônicos reside nas frequências obtidas. Em NMA as frequências estão relacionadas ao poço harmônico que se assume como aproximação do potencial. Enquanto que no método de filtro de dinâmica molecular as frequências correspondem à visita da proteína em seu espaço conformacional simulado, de forma que a amostragem é um fator limitante.

Por essa dependência, pode ser difícil associar os modos obtidos de mais baixa frequência a movimentos com importância biológica. Mas se o espaço conformacional for abrangente o suficiente, oscilações naturais serão observadas e também a repetição delas. Se um modo harmônico natural da proteína for suficientemente repetido dentro da simulação, então a frequência obtida para este modo já não será mais limitada pelo tempo de simulação. Por exemplo, no trabalho inicial de Osguthorpe, em que ele apresenta o método de filtro para o estudo de movimentos coletivos,²⁶ as maiores flutuações na distância entre o carbono β e o carbono γ de uma histidina são encontradas numa faixa de frequências entre 30 e 39 THz (1000 a 1300 cm^{-1}). E ainda que o período analisado (8 ps de simulação) permitisse encontrar movimentos em frequências tão baixas quanto 0,1 THz, não foram observadas amplitudes maiores que nesta faixa de frequências. A expectativa é que da mesma forma possamos encontrar uma faixa de frequências que delimite as maiores amplitudes para as oscilações da proteína inteira, contanto que o tempo simulado seja longo o suficiente.

O tempo de simulação deve ser muito maior que aqueles 8 ps justamente porque as oscilações de mais baixa frequência da proteína inteira devem corresponder a movimentos coletivos de grande amplitude, muito maiores que as flutuações na distância entre dois carbonos. Assim, também espera-se que a frequência seja dependente da massa do sistema analisado. Lembrando que num sistema massa-mola as frequências das oscilações naturais são proporcionais ao inverso da raiz quadrada da massa.

A DEA, um homohexâmero, possui uma massa total de cerca de 179 kDa e um total de 25080 átomos. Cada um de seus monômeros possui 4180 átomos e cerca de 30 kDa. No modelo de Monod a transição alostérica consiste num rearranjo concertado de cada protômero, e a análise das duas conformações da DEA, T e R, mostra que além das mudanças quaternárias também há mudanças na estrutura terciária de cada protômero,

Figura 34 – Análise da simulação restringida em comparação com a primeira parte da simulação 1, ambas em estado reduzido. As conformações da simulação restringida estão representadas em marrom, as conformações da simulação 1 em verde, a estrutura cristalográfica holo em laranja e a estrutura cristalográfica apo em rosa. a) distâncias atômicas entre o nitrogênio da Ala6 e o oxigênio da Arg65 da cadeia oposta durante a simulação restringida. b) ADPs isotrópicos para as duas cadeias da MosR calculados como médias das janelas de 50 ns tanto para a simulação restringida, quanto para a primeira parte da simulação 1. c) distâncias relativas entre as duas hélices de reconhecimento. d) ângulo que cada hélice de reconhecimento da conformação simulada faz com sua respectiva hélice da estrutura holo. e) e f) duas perspectivas da conformação média da simulação restringida superposta à conformação média da primeira parte da simulação 1 e à estrutura cristalográfica apo. As hélices de reconhecimento estão em destaque.

Fonte: Elaborada pela autora.

proveniente da rotação da região externa em torno de um core rígido.⁴⁰ Assim, surge a pergunta de qual seria a dependência da frequência de oscilação da transição alostérica com a oligomerização e a simetria.

Pela aproximação de redes elásticas – que reproduz oscilações harmônicas mais dependentes da forma do sistema que de detalhes interatômicos, como discutido anteriormente – observamos oscilações simétricas de movimentos relativos entre os protômeros. Para o modo de mais baixa frequência observamos uma rotação de todo o protômero em torno do eixo de ordem 3, mas não a mudança conformacional na estrutura terciária, como indicado na [Figura 35](#). A região circulada em vermelho na [Figura 35](#) indica o sítio alostérico, que é mais fechado na conformação R que na T pela aproximação do *loop* de resíduos de 144 a 154. Apenas quando o sítio se fecha é que a ligação a um fosfato pode se estabelecer e assim regular a ativação do sítio ativo.

A [Figura 35](#) mostra as duas conformações T e R superpostas ([Figura 35a](#)), indicando a transição alostérica observada experimentalmente por vetores atômicos para os carbonos α ([Figura 35b](#), [Figura 35c](#) e [Figura 35d](#)) e a mudança na estrutura terciária que distorce o sítio alostérico e regula o sítio ativo, região circulada em vermelho na [Figura 35](#). Também está na [Figura 35](#) o modo de mais baixa frequência obtido por redes elásticas ([Figura 35e](#), [Figura 35f](#) e [Figura 35g](#)). A rotação da região externa do monômero parece bem representada por esse modo, o que indica uma relação desse movimento com a forma hexamérica da proteína. No entanto, faltaram detalhes interatômicos para separar o core interno, que permanece estático, e representar a mudança na região reguladora como a deformação e a aproximação do *loop* com resíduos de 144 a 154 ao core interno, como descrito em⁴⁰ e destacado na [Figura 35h](#). É possível verificar na [Figura 35b](#) e na [Figura 35h](#), como as direções dos vetores atômicos nesta região, circulada em vermelho, são diferentes na transição observada experimentalmente e no modo elástico 7. Assim se justifica que analisemos os movimentos coletivos da DEA por NMA.

A [Tabela 7](#) relaciona cada um dos 10 modos de mais baixa frequência obtidos por NMA e ENM, comparando-os pela frequência calculada, pelo grau de coletividade e pela correlação com a transição alostérica observada experimentalmente para o hexâmero. As frequências calculadas por ENM não proveêm valores com unidades, e são sempre relativas ao primeiro modo. Já NMA possibilita que a cada modo seja designada uma frequência exata. A frequência obtida para o modo 7 por NMA é menos de mil vezes mais baixa que a obtida por Osguthorpe para o par de carbonos, o que é compatível com a razão da raiz quadrada de suas massas, $\sqrt{425}$ (considerando a massa de todo o hexâmero). Mas no período correspondente a essa frequência para o modo 7, cerca de 43 ps, não é possível observar a transição entre os estados numa simulação de dinâmica molecular.⁴¹ Ademais, outros trabalhos já indicam que esse tipo de movimento coletivo ocorre na faixa de microssegundos a segundos.¹⁸

Figura 35 – Representação em *trace* de diferentes conformações da DEA. a) superposição das conformações T e R cristalográficas, T em preto e R em branco. b) A conformação cristalográfica T com os vetores de deslocamento atômico para os carbonos α representando a transição para a conformação R. c e d) hexâmero da conformação T cristalográfica com uma das cadeias representada apenas pelos vetores de deslocamento atômico dos carbonos α da conformação T para a R. Em c) o plano da figura está perpendicular ao eixo de ordem 3 da DEA, e em d) o plano da figura está paralelo a ele. e) e f) o hexâmero na conformação T cristalográfica e os vetores de deslocamento dos carbonos α no modo elástico 7. Cada cadeia está colorida por uma cor diferente. O hexâmero em e) e f) estão orientados assim como em c) e d), respectivamente. g) conformação T cristalográfica representando o modo elástico 7 por vetores de deslocamento atômico para os carbonos α . A região reguladora está circutada em vermelho, destacando o *loop* com os resíduos de 144 a 154, e ampliada em h).

Fonte: Elaborada pela autora.

Tabela 7 – Relação dos 10 modos de mais baixa frequência obtidos por redes elásticas e NMA. Os modos estão relacionados por número, frequência (Freq.), grau de coletividade (Col.) e produto escalar (PE) com o vetor que descreve a transição entre as conformações cristalográficas T e R.

modos elásticos				modos normais		
modo	Freq.	Col.	PE	Freq.(GHz)	Col.	PE
modo 7	1	0,73	0,9	23,42	0,7	0,78
modo 8	1,22	0,49	0	33,93	0,55	-0,06
modo 9	1,22	0,53	0	38,31	0,49	-0,29
modo 10	1,68	0,63	0	40,34	0,53	-0,22
modo 11	1,68	0,63	0	44,72	0,58	0,08
modo 12	1,76	0,62	0,08	49,67	0,59	0,03
modo 13	1,93	0,7	0	51,95	0,62	0,1
modo 14	1,97	0,45	0	54,57	0,65	0,01
modo 15	1,97	0,51	0	57,8	0,59	-0,13
modo 16	2,12	0,51	0	60,13	0,5	-0,03

Fonte – Elaborada pela autora.

Então realizamos duas simulações da DEA, em sua conformação T, na tentativa de obter a frequência da aproximação harmônica que melhor reproduza a transição alostérica. Uma delas simula o hexâmero explicitamente. A outra simula apenas um monômero, enquanto os outros cinco são imagens do primeiro obtidas pelas operações de simetria do cristal (ver a seção de métodos). O tempo simulado de cada uma permitiu que analisássemos pelo método do filtro os períodos de 327,68 ns e 40,96 ns, respectivamente. Entretanto, para ambas as simulações observamos um aumento na amplitude dos modos conforme o decréscimo da frequência, de forma que não foi possível determinar uma faixa de frequência para uma oscilação natural da DEA. Ademais, o produto escalar calculado entre os 10 primeiros modos filtrados e a transição alostérica experimental são todos menores que 0,5, em ambas as simulações, como mostra a [Tabela 8](#).

Quanto ao produto escalar com a transição alostérica, observamos o maior valor para o modo elástico 7, ainda que ele não represente corretamente a mudança conformacional na região do sítio alostérico, como descrito. Os outros modos calculados por ENM resultam em um produto escalar ínfimo, indicando modos muito discrepantes da transição experimental. Já os modos obtidos por NMA não resultam em valores de produto escalar menos divergentes. O modo normal 7, de mais baixa frequência, é o que mais se assemelha à transição experimental, apresentando a rotação da região externa em cada

Tabela 8 – Relação dos 20 primeiros modos obtidos pelo método de filtro de frequências da simulação do hexâmero e da do monômero. Os modos estão relacionados por número, frequência (Freq.), grau de coletividade (Col.) e produto escalar (PE) com o vetor que descreve a transição entre as conformações cristalográficas T e R.

	hexâmero			monômero		
modo	Freq.(MHz)	Col.	PE	Freq.(MHz)	Col.	PE
modo 1	1,53	0,5	0,31	12,21	0,4	-0,02
modo 2	3,05	0,58	0,29	24,41	0,52	-0,09
modo 3	4,58	0,63	0,23	36,62	0,49	-0,4
modo 4	6,1	0,51	-0,09	48,8	0,38	-0,12
modo 5	7,63	0,62	-0,08	61,03	0,45	-0,02
modo 6	9,15	0,48	-0,09	73,24	0,47	0,01
modo 7	10,68	0,48	-0,42	85,45	0,34	-0,07
modo 8	12,21	0,54	0,14	97,66	0,53	-0,26
modo 9	13,73	0,59	-0,05	109,86	0,61	-0,27
modo 10	15,26	0,59	-0,17	122,07	0,53	-0,21
modo 11	16,79	0,65	0,05	134,31	0,5	0,54
modo 12	18,31	0,59	0,01	146,52	0,38	0,46
modo 13	19,84	0,52	-0,06	158,73	0,61	-0,52
modo 14	21,36	0,54	0,14	170,94	0,31	0,29
modo 15	22,89	0,5	0,34	183,15	0,47	0,16
modo 16	24,42	0,58	0,27	195,36	0,55	0,4
modo 17	25,94	0,63	0,01	207,57	0,44	-0,28
modo 18	27,47	0,62	0,36	219,78	0,49	-0,6
modo 19	28,99	0,36	-0,21	231,99	0,37	0,05
modo 20	30,52	0,55	-0,04	244,14	0,53	0,22

Fonte – Elaborada pela autora.

protômero, mas a mudança conformacional na região reguladora ainda pouco se assemelha à experimental. Já o modo normal 9, com o segundo maior produto escalar, apresenta uma mudança conformacional na região reguladora de uma das cadeias mais parecida com a experimental, como mostra a [Figura 36](#). Assim, é possível que uma combinação desses dois modos resultasse num movimento até mais semelhante ao experimental que o modo elástico 7. A [Figura 36](#) mostra o modo normal 7 ([Figura 36a](#) e [Figura 36b](#)), destacando a cadeia B para mostrar a região reguladora circulada em vermelho ([Figura 36c](#)). A [Figura 36d](#)

mostra a cadeia C no modo normal 9, também indicando a região reguladora circulada em vermelho.

A [Figura 36](#) também mostra o modo de mais baixa frequência calculado pelo método de filtro das dinâmicas, tanto do hexâmero ([Figura 36e](#) e [Figura 36f](#)) quanto do monômero ([Figura 36g](#) e [Figura 36h](#)). Nenhum dos modos filtrados apresenta movimentos semelhantes aos de rotação da região externa em cada cadeia, e tampouco cada cadeia se move de forma simétrica em relação às outras. A falta de simetria fica evidente tanto na [Figura 36](#) quanto na [Tabela 9](#), que mostra o produto escalar de cada cadeia separadamente para os 20 modos filtrados de mais baixa frequência, para o modo normal 7 e para o modo elástico 7. Enquanto esses valores são muito próximos para as 6 cadeias no modo elástico, eles variam um pouco mais no modo normal e são totalmente discrepantes nos modos filtrados. Como mostrado na [Figura 36](#), cada cadeia do hexâmero do modo filtrado 1 se move de forma desconcertada do restante, apresentando um grau de coletividade de apenas 0,5.

Observamos que prevalece nos modos filtrados o movimento da tampa do sítio ativo de cada cadeia, e não a rotação da região externa. É possível que esta rotação seja dependente de um movimento em fase do hexâmero, refletindo sua forma, como evidenciado pelo cálculo dos modos elásticos, mais sensíveis à forma da proteína que a detalhes interatômicos, como discutido anteriormente. Já as mudanças na região reguladora podem ser observadas em conjunto com movimentos da tampa do sítio ativo. Considerando os 100 primeiros modos filtrados, tanto para cada cadeia do hexâmero simulado explicitamente, quanto para o monômero, o maior valor de produto escalar com o vetor de transição alostérica encontrado é igual a 0,67. Este está mostrado na [Figura 36](#), e representa o modo 41 filtrado da simulação do monômero, equivalente a uma frequência de 0,5 GHz (mais de 40 vezes menor que a menor das frequências obtidas por NMA). A amplitude do movimento é muito baixa, mas já mostra uma deformação na região reguladora semelhante à deformação experimental, como indica [Figura 36](#), o que justifica o valor encontrado para o produto escalar, indicando uma aproximação do *loop* com resíduos de 144 a 154 em relação ao core interno, como descrito em,⁴⁰ e a região mais acima também se movimenta no sentido de se aproximar ao core interno, apesar do movimento da tampa do sítio ativo aparecer em direção quase perpendicular.

Portanto, o período simulado para a DEA, tanto do hexâmero quanto do monômero, não foi suficiente para apresentar a rotação da região externa predominante na transição alostérica observada experimentalmente, mas já mostra detalhes na mudança conformacional que os modos elásticos não representam, e também indicam que as frequências associadas a esses movimentos são muito menores que o calculado por NMA. Além disso, a simulação do monômero de fato mostrou valores de produto escalar mais altos, com destaque para a faixa de 100 a 170 MHz, que para qualquer cadeia na simulação do

Figura 36 – Representação em *truce* de diferentes conformações da DEA. a), b), c) e d) representam a conformação T após a minimização necessária para ANM, com os vetores de deslocamento atômico para os carbonos α representando um modo normal. Em a) e b) está representado o hexâmero com cada cadeia colorida por uma cor diferente e os vetores indicando o modo normal 7, os planos das figuras estão orientados perpendicularmente e paralelamente ao eixo de ordem 3, respectivamente. Em c) está representada a cadeia B com os vetores indicando o deslocamento dos carbonos α no modo normal 7. Em d) está representada a cadeia C com os vetores indicando o deslocamento dos carbonos α no modo normal 9. e) e f) representam a primeira conformação da dinâmica do hexâmero considerada para o cálculo do filtro de frequências, cada cadeia está colorida por uma cor diferente e os vetores indicam o modo harmônico filtrado número 1. Os planos das figuras estão orientados como em a) e b), respectivamente. g) e h) representam a primeira conformação da dinâmica do monômero e os vetores indicam os modos harmônicos filtrados números 1 e 41, respectivamente, sendo que para o modo 41 os módulos estão multiplicados por 20. A região reguladora está circulada em vermelho, destacando o *loop* com os resíduos de 144 a 154.

Fonte: Elaborada pela autora.

Tabela 9 – Valores do produto escalar entre o vetor que representa cada cadeia do hexâmero em cada modo listado e o vetor que representa a transição alostérica entre as conformações cristalográficas T e R. Estão listados os 20 primeiros modos harmônicos filtrados da simulação do hexâmero, além do modo normal 7 e o modo elástico 7

modo	cadeia A	cadeia B	cadeia C	cadeia D	cadeia E	cadeia F
modo filtrado 1	0,47	0,06	-0,18	0,16	-0,02	-0,48
filtrado 2	0,22	0,23	-0,47	0,03	0,26	-0,29
filtrado 3	0,46	0,03	-0,33	0,22	0,09	-0,45
filtrado 4	0,11	-0,05	-0,55	0,16	0,47	-0,16
filtrado 5	0,13	-0,18	-0,42	0,34	0,24	-0,15
filtrado 6	-0,09	0,2	-0,23	-0,09	0,36	-0,16
filtrado 7	-0,11	0,04	0,04	-0,13	0,17	0
filtrado 8	-0,05	0,01	-0,14	-0,08	0,47	-0,2
filtrado 9	-0,2	0,25	0,2	-0,4	0,16	0,01
filtrado 10	0,23	-0	-0,27	0,27	0,16	-0,38
filtrado 11	-0,15	0,35	0,12	-0,36	0,02	0,03
filtrado 12	0,34	0,06	-0,12	0,27	-0,2	-0,36
filtrado 13	0,1	0,23	-0,44	0,01	0,09	-0
filtrado 14	0,26	-0,11	-0,38	0,28	0,13	-0,19
filtrado 15	0	0,33	-0,12	-0,18	0	-0,02
filtrado 16	-0,02	-0,37	-0,34	0,38	0,29	0,01
filtrado 17	-0,11	0,11	-0,33	-0,1	0,54	-0,12
filtrado 18	-0,12	-0,22	-0,24	0,24	0,43	-0,12
filtrado 19	-0,16	-0,3	-0,44	0,04	0,51	0,05
filtrado 20	-0,11	0,15	-0,02	-0,05	0,18	-0,15
normal 7	0,82	0,83	0,75	0,79	0,77	0,79
elástico 7	0,89	0,9	0,9	0,91	0,91	0,91

Fonte – Elaborada pela autora.

hexâmero, o que sugere que a simetria proveniente do hexâmero afete a frequência dos movimentos coletivos.

6 CONCLUSÕES

Assim como dividimos os objetivos em cinco tópicos específicos, as conclusões também estão divididas de acordo. Elas estão expostas a seguir na mesma ordem: sobre os complexos TR-DNA, sobre o método desenvolvido para obtenção de modos harmônicos filtrados das dinâmicas, sobre a construção do modelo da ElrR ligada ao DNA, sobre a afinidade da MosR pelo DNA, e sobre os modos coletivos da DEA. Todavia, algumas conclusões parecem ser recorrentes, visto que as temáticas como ligação ao DNA e movimentos coletivos também estão difundidas entre vários destes objetos de estudo.

6.1 Sobre os complexos TR-DNA

Como discutimos anteriormente, a geometria e distribuição de cargas específicas para a dupla hélice de DNA remete a existência de condições equivalentes para os DBDs, mas os muitos intentos realizados neste sentido mostraram que não há nenhuma relação ou regra simples entre os dois. Assim, restringir o grupo de DBDs estudados tem-se mostrado mais frutífero, como analisar apenas complexos que interajam através de uma hélice α ,⁴³ ou considerar somente DBDs que possuam o motivo HTH.⁵² Neste projeto restringimos ainda mais ao considerar somente reguladores transcricionais homodiméricos, cuja simetria de ordem 2 fosse acompanhada por certa palindromia do sítio de DNA. Essa restrição permitiu que definíssemos novos parâmetros, alguns mais superficiais, como os tipos I ou II, e outros mais detalhados, como γ , λ e Dp. De qualquer forma, todos os parâmetros conduziram à mesma conclusão principal, de que o encaixe das hélices α reflete as interações entre proteína e DNA, tanto específicas, com as bases de DNA, quanto inespecíficas, com os grupos de açúcar e fosfato.

Este estudo mostrou que conhecer algumas características da proteína ou do sítio de DNA permite prever com maior ou menor precisão como será o encaixe dos dois. Por exemplo, identificar um TR do tipo II já indica que, quando ligado ao DNA, o ângulo entre suas duas hélices de reconhecimento, β , será menor que 180° , que a distância entre seus resíduos principais será por volta de $35 \text{ \AA}$, e que eles provavelmente interagirão com uma base do par CG, cuja base simétrica se posicionará num sulco maior vizinho, separada por cerca de 9 pares de base. Mais informações quanto aos resíduos que compõem a hélice de reconhecimento e as bases que constituem o sítio de DNA devem auxiliar na precisão desse prognóstico. Por enquanto, o conjunto de complexos que levantamos é pequeno demais para avaliar a contribuição específica do posicionamento de cada resíduo ou base. Mas a conclusão geral, de que é possível caracterizar o tipo de ligação apenas a partir da sequência do sítio de DNA, é de grande valor para futuros estudos.

Além disso, este conjunto já foi suficiente para encontrar um complexo com muitas

características semelhantes à ElrR e seu sítio de DNA, suficientes para fazer uma previsão bem embasada de como devem se encaixar as hélices de reconhecimento da ElrR nos sulcos maiores do DNA. Uma base mais ampla de dados daria maior precisão às médias calculadas para alguns parâmetros, e também forneceria complexos para fazer mais hipóteses como a que fizemos para a ElrR. Assim, acreditamos que isto sirva de motivação para expandir essas análises a outros complexos de TR-DNA.

6.2 Sobre o método de filtro de modos harmônicos

O algoritmo escrito pôde ser aplicado na análise das três proteínas sem dificuldades e se mostrou muito satisfatório quanto ao tempo de cálculo e custo da memória. O cálculo de transformada de Fourier requer que se armazene um vetor de tamanho igual ao número de conformações fornecidas para cada coordenada. Um vetor representa o custo mínimo de memória. Para poupar tempo no processo de leitura da trajetória, armazenamos numa matriz uma coluna para cada coordenada que vamos avaliar, podendo ser as coordenadas de todos os átomos da proteína ou somente alguns. Assim, existe um balanço nos custos de tempo e memória que pode ser otimizado a depender do computador. De forma geral, ele não nos tomou mais que alguns minutos para gerar um arquivo contendo as amplitudes de cada coordenada para cada frequência em estudo, sendo que dos sistemas que estudamos consideramos sempre todos os átomos e escolhemos até as 100 menores frequências para analisar. Como perspectiva, esperamos disponibilizar o código fonte junto com um artigo sobre sua aplicação. Mas suas possíveis aplicações ainda estão em fase inicial, e seria ideal poder aplicá-lo a uma dinâmica muito mais longa que as que pudemos simular neste projeto. Com trajetórias mais longas caberia uma discussão mais ampla sobre os modos harmônicos mais básicos que as proteínas apresentam.

6.3 Sobre a construção do modelo da ElrR ligada ao DNA

Uma vez feita a hipótese de como as hélices de reconhecimento da ElrR se encaixariam em seu sítio de DNA, pudemos construir modelos por diferentes aproximações. Os métodos mais simples resultaram em modelos não suficientemente detalhados, o que justifica o uso de métodos mais complexos. O modelo por superposição de corpos rígidos apresentou buracos no interior da proteína que desestabilizariam sua estrutura. Os *decoys* levaram a mudanças conformacionais na região C-terminal que não eram coerentes com outros experimentos. E a conformação retirada diretamente da MD não mostrou um sítio alostérico fechado o suficiente para permitir a previsão do ligante.

A abordagem mais complexa por combinação linear de modos, usando o método dos mínimos quadrados para aproximar as hélices de reconhecimento à posição alvo, gerou uma enorme quantidade de modelos. A maioria deles tinha graves distorções na estrutura, mas alguns puderam ser selecionados, minimizados e considerados para análise do sítio

alostérico. Essa seleção foi baseada em características geométricas, mas vários modelos eram similares nesses quesitos, e um método mais apurado poderia ser desenvolvido.

Tanto o modelo construído com modos harmônicos filtrados quanto o com modos normais contribuíram para a análise da conformação fechada do sítio alostérico. Entretanto, a combinação de modos harmônicos filtrados resultou em modelos com valores menores de RMSD para as hélices de reconhecimento e também com um menor grau de distorções, possivelmente porque os modos harmônicos filtrados mostraram valores menores para o produto escalar com o vetor de transição das hélices de reconhecimento. Já os modos normais possuíam maiores graus de coletividade, mas valores menores para o produto escalar. Assim, modos não tão similares à transição das hélices de reconhecimento podem ter sido considerados com grande amplitude, que por sua vez levou a um grande deslocamento também em outras regiões da proteína.

O método de filtro também proporcionou uma análise interessante das frequências relacionadas aos modos que compunham o modelo selecionado. O modo com maior valor de produto escalar, 0,92, corresponde a uma frequência de 114 MHz, mas as hélices de reconhecimento não alcançaram a conformação alvo com essa frequência na trajetória simulada. De fato, o modelo foi composto por mais outros 8 modos, todos numa faixa menor de frequências. Isso sugere que apesar do movimento na direção correta ocorrer rapidamente, ele dependa de outros movimentos para levar as hélices de reconhecimento à uma posição ideal para o encaixe no DNA. É então a combinação dos modos que justifica a raridade deste evento.

Portanto, o uso de métodos mais complexos produz modelos mais realistas, e também fornece uma descrição mais acurada, além de relacionada ao tempo, para a transição entre dois estados biológicos.

6.4 Sobre a afinidade da MosR pelo DNA

As simulações da MosR em sua forma apo mostraram-se em boa concordância com resultados experimentais. O total de 1,6 μ s de trajetórias simuladas, metade em estado reduzido e metade em estado oxidado, resultou numa conformação média com apenas 1 Å de RMSD em relação à estrutura cristalográfica apo. A formação da ponte dissulfeto também promoveu a tensão na hélice e a redistribuição de cargas como esperado. A conformação média no estado oxidado de fato mostrou-se a mais distante da conformação cristalográfica holo, mas a média no estado reduzido tampouco mostrou a rotação das hélices de reconhecimento para encaixe no DNA como na conformação holo.

Pela análise do espaço conformacional definido por variáveis coletivas (semelhantes as levantadas para o estudo dos complexos de TR-DNA), observamos poucas conformações das hélices de reconhecimento compatíveis com o encaixe no DNA e em apenas uma das

simulações em estado reduzido. Isto sugere que a interação específica com o DNA é rara e possível somente no estado reduzido. De fato, encontramos que ela depende das interações que o N-terminal faz, e também está associada a mudanças na flexibilidade.

Identificamos que quando os resíduos Ala6 e Arg65' de cadeias opostas conseguem fazer uma ligação de hidrogênio através de seus átomos da cadeia principal, o DBD a que pertence a Arg65' apresenta maior flexibilidade. Situação que perdurou pelos primeiros 200 ns de uma simulação em estado reduzido. Nessa situação, a conformação média desta cadeia com DBD mais flexível possui um RMSD de apenas 1,08 Å em relação à conformação cristalográfica holo.

Essa ligação de hidrogênio entre Ala6 e Arg65' promove o posicionamento do N-terminal entre a α -hélice 1 e α -hélice 3', afastando o DBD do domínio de dimerização e aumentando sua flexibilidade. No estado oxidado essa conformação do N-terminal parece impossível, já que a ponte dissulfeto provoca uma tensão no N-terminal da α -hélice 1 e também um afastamento da Asn37' na α -hélice 2, acompanhado de todo o DBD em direção ao domínio de dimerização. Assim, no estado oxidado observamos uma interação entre Ala6 e Arg65' apenas mediada por uma molécula de água, sendo incapaz de promover as mudanças conformacionais necessárias para a ligação com o DNA.

Quando a posição do N-terminal é favorável, os modos harmônicos filtrados resultam muito mais coletivos, sendo que o maior grau de coletividade encontrado está associado a uma frequência de 60 MHz, e também observamos um aumento na flexibilidade do DBD já na frequência de 20 MHz, pela análise de APDs.

É intrigante, no entanto, que a restrição no N-terminal seja acompanhada de maior flexibilidade, quando esta é geralmente associada a estruturas mais relaxadas e desordenadas. Para a MosR, o N-terminal é em sua maior parte do tempo mais livre e flexível, mas num subconjunto de conformações possíveis ele interage de forma estável com o DBD, e eles passam a se movimentar coletivamente, N-terminal e DBD.

Por fim, a afinidade da MosR pelo DNA parece depender de vários fatores, associados a diferentes níveis de mudança conformacional, mas as ferramentas que apresentamos parecem apropriadas para discerni-los.

6.5 Sobre os modos coletivos da DEA

As análises que fizemos para a DEA também variaram em grau de complexidade, assim como os diferentes métodos usados para construir os modelos da ElrR ligada ao DNA. Usamos três métodos distintos para obter os modos coletivos da DEA, ENM, NMA e o método de filtro da MD. O modo elástico 7 mostrou grande concordância com a transição esperada para a DEA, entre suas duas conformações cristalográficas nos estados T e R, com um produto escalar de 0,9. Apesar desse valor bastante alto, esse modo não

foi capaz de reproduzir as mudanças no sítio alostérico, que provêm de uma rotação da região externa em torno de um eixo deslocado do eixo de simetria do hexâmero. Além disso, todos os outros modos elásticos resultaram num produto escalar menor que 0,1.

O modo normal 7, por outro lado mostrou um produto escalar de 0,78 e em outros dois modos ele foi maior que 0,2. Enquanto o produto escalar obtido com o modo elástico pode ser superestimado por partir da conformação cristalográfica, com as mesmas exatas coordenadas, os modos normais podem ser subestimados por partirem de uma conformação minimizada, e que também resultou em movimentos ligeiramente assimétricos em torno do eixo de simetria do hexâmero. Apesar do valor menor, o modo normal 9 mostrou boa concordância com as mudanças conformacionais esperadas para o sítio alostérico. Assim, os modos normais podem se combinar para resultar num movimento com maior semelhança com a transição experimental de uma forma que os modos elásticos não poderiam.

Os modos filtrados da simulação com o hexâmero podem resultar num produto escalar ainda mais subestimado por partir de uma conformação mais afastada da cristalográfica e assimétrica, mas mesmo analisando cada cadeia separadamente, todos os produtos escalares calculados foram menores que 0,6. O maior produto escalar obtido para um modo harmônico filtrado foi justamente com a simulação do monômero, o modo 41, que corresponde a um produto escalar de 0,67 com a transição esperada e está associado a uma frequência de 0,5 GHz. Isso sugere que a simetria imposta, de fato promova movimentos coletivos mais semelhantes com a transição experimental.

Como o modo elástico mostrou um alto produto escalar, e a simetria imposta para a simulação do monômero também elevou o produto escalar, fica claro que a forma da proteína, neste caso a forma hexamérica, de fato influencia nos movimentos coletivos, como discutido por Tama e Sanejouand,⁷⁰ e a transição experimental é em grande parte constituída por esse movimento coletivo dependente da forma. Mas a forma não é a única que contribui para a direção dos movimentos coletivos, as interações atômicas também, como provam os maiores detalhes considerados em NMA e no método de filtros. Cada contribuição, seja a forma ou sejam as interações atômicas, estão associadas a um frequência diferente. Assim, a frequência relacionada à transição experimental também deve ser uma combinação de outras.

Por fim, ainda que as simulações não tenham sido longas o suficiente para observarmos um movimento de indubitável frequência natural, a transição experimental parece ser composta por modos harmônicos que já estamos encontrando.

6.6 Conclusões finais

Para as três proteínas que estudamos os movimentos com relevância funcional parecem ser melhor descritos por uma combinação de modos harmônicos com distintas

frequências, e estes modos são influenciados tanto pela forma da proteína quanto por interações atômicas, como as estabelecidas pelo N-terminal da MosR, ou as presentes na região reguladora da DEA.

Assim, a nossa compreensão dos movimentos funcionais das proteínas pode ser muito desenvolvida não só pelo cálculo de modos harmônicos, mas também pela combinação deles. Os detalhes atômicos que os modos normais consideram de fato melhoram a descrição dos movimentos feita somente por modos elásticos, e os modos harmônicos filtrados, talvez por não serem ortogonais, por sua vez também indicam serem bons descritores, quando combinados. Ademais, mesmo que as simulações não sejam longas o suficiente para encontrar os modos naturais da proteína, a combinação de modos filtrados pode indicar também como as frequências se combinam para resultar numa frequência natural.

Uma vez encontrada essa frequência natural, poderíamos elaborar experimentos que interfiram na ressonância da proteína para controlar sua função. Ou talvez percebamos que já estamos sujeitos no nosso dia a dia a campos eletromagnéticos cujas frequências interferem no bom funcionamento das nossas proteínas. Por hora, a frequência dos movimentos funcionais das proteínas é uma área ainda pouco explorada, e nós esperamos que os métodos abordados aqui sejam um caminho para compreendê-la.

REFERÊNCIAS

- 1 LEVINE, M.; TJIAN, R. Transcription regulation and animal diversity. **Nature**, v. 424, n. 6945, p. 147–151, 2003.
- 2 JAROSZ, D. F.; TAIPALE, M.; LINDQUIST, S. Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. **Annual Review of Genetics**, v. 44, n. 1, p. 189–216, 2010.
- 3 SCHRECK, R.; ALBERMANN, K.; BAEUERLE, P. A. Nuclear factor K κ B: an oxidative stress-responsive transcription factor of eukaryotic cells (a review). **Free Radical Research Communications**, v. 17, n. 4, p. 221–237, 2009.
- 4 PABO, C. O.; SAUER, R. T. Transcription factors: structural families and principles of DNA recognition. **Annual Review of Biochemistry**, v. 61, 1992. doi=10.1146/annurev.bi.61.070192.005201.
- 5 MONOD, J. From enzymatic adaptation to allosteric transitions. **Science**, v. 154, n. 748, p. 475–483, 1966.
- 6 HARRISON, S. C. A structural taxonomy of DNA-binding domains. **Nature**, v. 353, 1991. doi:10.1038/353715a0.
- 7 BERMAN, H. M. et al. The Protein Data Bank. **Nucleic Acids Research**, v. 28, 2000. doi:10.1093/nar/28.1.235.
- 8 KALODIMOS, C. G.; BOELEN, R.; KAPTEIN, R. A residue-specific view of the association and dissociation pathway in protein–DNA recognition. **Nature Structural Biology**, v. 52, p. 193–197, 2002. doi:10.1038/nsb763.
- 9 MONOD, J.; CHANGEUX, J.-P.; JACOB, F. Allosteric proteins and cellular control systems. **Journal of Molecular Biology**, v. 6, n. 4, p. 306–329, 1963.
- 10 WEBER, G. Energetics of ligand binding to proteins. **Advances in Protein Chemistry**, v. 29, p. 1–83, 1975. doi:10.1016/S0065-3233(08)60410-6.
- 11 FRAUENFELDER, H.; SLIGAR, S. G.; WOLYNES, P. G. The energy landscapes and motions of proteins. **Science**, v. 254, n. 5038, p. 1598–1603, 1991.
- 12 ANSARI, A. et al. Protein states and proteinquakes. **Proceedings of the National Academy of Sciences**, v. 82, n. 15, p. 5000–5004, 1985.
- 13 SWAMINATHAN, S. et al. Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor. **Biochemistry**, v. 21, n. 21, p. 5230–5241, 1982.
- 14 ELBER, R.; KARPLUS, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. **Science**, v. 235, n. 4786, p. 318–321, 1987.
- 15 KARPLUS, M.; ICHIYE, T.; PETTITT, B. Configurational entropy of native proteins. **Biophysical Journal**, v. 52, n. 6, p. 1083–1085, 1987.

- 16 ISHIMA, R.; TORCHIA, D. Protein dynamics from NMR. **Nature structural & molecular biology**, v. 76, n. 2-3, p. 145–152, 2000.
- 17 TZENG, S.-R.; KALODIMOS, C. G. Protein activity regulation by conformational entropy. **Nature**, v. 488, n. 7410, p. 236–240, 2012.
- 18 HENZLER-WILDMAN, K.; KERN, D. Dynamic personalities of proteins. **Nature**, v. 450, n. 7172, p. 964–972, 2007.
- 19 SHAW, D. E. et al. Millisecond-scale molecular dynamics simulations on Anton. In: . **CONFERENCE ON HIGH PERFORMANCE COMPUTING NETWORKING, STORAGE AND ANALYSIS**, 9., 2009, Portland. Proceedings...Portland: IEEE, 2009. doi:10.1145/1654059.1654126.
- 20 OLIVA, G. et al. Structure and catalytic mechanism of glucosamine 6-phosphate deaminase from *Escherichia coli* at 2.1 Å resolution. **Structure**, v. 3, n. 12, p. 1323–1332, 1995.
- 21 DE GROOTE, M. C. R. **Estudos cristalográficos do ElrR, regulador transicional do fator de virulência ElrA de *Enterococcus faecalis*, e indícios de interação com a região de ligação ao DNA**. 2017. Tese (Doutorado em Ciências) — Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.
- 22 BRUGAROLAS, P. et al. The oxidation-sensing regulator (MosR) is a new redox-dependent transcription factor in *Mycobacterium tuberculosis*. **Journal of Biological Chemistry**, v. 287, n. 45, p. 37703–12, Nov 2012.
- 23 DUMOULIN, R. et al. Enterococcal Rgg-like regulator ElrR activates expression of the *elrA* operon. **Journal of Bacteriology**, v. 195, n. 13, p. 3073–83, 2013.
- 24 TSAI, C. J. et al. Folding funnels, binding funnels, and protein function. **Protein Science**, v. 8, n. 6, p. 1181–1190, 1999.
- 25 HUANG, H.; OZKIRIMLI, E.; POST, C. B. A comparison of three perturbation molecular dynamics methods for modeling conformational transitions. **Journal of Chemical Theory and Computation**, v. 5, n. 5, p. 1301–1314, 2009.
- 26 SESSIONS, R. B.; DAUBER-OSGUTHORPE, P.; OSGUTHORPE, D. J. Filtering molecular dynamics trajectories to reveal low-frequency collective motions: Phospholipase A2. **Journal of Molecular Biology**, v. 210, n. 3, p. 617–633, 1989.
- 27 AMADEI, A.; LINSEN, A. B. M.; BERENDSEN, H. J. C. Essential dynamics of proteins. **Proteins**, v. 17, n. 4, p. 412–425, 1993.
- 28 BROOKS, B.; KARPLUS, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. **Proceedings of the National Academy of Science of the United States of America**, v. 80, n. 21, p. 6571–6575, 1983.
- 29 GO, N.; NOGUTI, T.; NISHIKAWA, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. **Proceedings of the National Academy of Sciences**, v. 80, n. 12, p. 3696–3700, 1983.

-
- 30 PIETERS, J. Mycobacterium tuberculosis and the macrophage: maintaining a balance. **Cell Host & Microbe**, v. 3, n. 6, p. 399–407, 2008.
- 31 WILKINSON, S. P.; GROVE, A. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. **Current Issues in Molecular Biology**, v. 8, n. 1, p. 51–62, 2006.
- 32 SARIDAKIS, V. et al. Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. **Journal of Molecular Biology**, v. 377, n. 3, p. 655–667, 2008.
- 33 PERERA, I. C.; GROVE, A. Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. **Journal of Molecular Cell Biology**, v. 2, n. 5, p. 243–254, 2010.
- 34 NEWBERRY, K. J. et al. Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. **Molecular Cell**, v. 28, n. 4, p. 652–664, 2007.
- 35 EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D**, v. 66, n. 4, p. 486–501, 2010.
- 36 ALTAMIRANO, M. M. et al. Asymmetric allosteric activation of *Escherichia coli* glucosamine-6-phosphate deaminase produced by replacements of Tyr 121. **Biochemistry**, v. 34, n. 18, p. 6074–6082, 1995.
- 37 VOET, D.; VOET, J. G. Regulação Alostérica. In: FRAGA, S. (Ed.). **Bioquímica**. 4. ed. Porto Alegre: Artmed, 2013. cap. 10, p. 352.
- 38 RUBIN, M. M.; CHANGEUX, J.-P. On the nature of allosteric transitions: implications of non-exclusive ligand binding. **Journal of Molecular Biology**, v. 21, n. 2, p. 265–274, 1966.
- 39 MONOD, J.; WYMAN, J.; CHANGEUX, J.-P. On the nature of allosteric transitions: a plausible model. **Journal of Molecular Biology**, v. 12, n. 1, p. 88–118, 1965.
- 40 HORJALES, E. et al. The allosteric transition of glucosamine-6-phosphate deaminase: the structure of the T state at 2.3 Å resolution. **Structure**, v. 7, n. 5, p. 527–537, 1999.
- 41 CÂMARA, A. S. **Uma transição assimétrica entre estados simétricos: o alosterismo da Glucosamina 6-fosfato Desaminase**. 2013. 118 p. Dissertação (Mestrado em Ciências) — Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2013.
- 42 SAUER, R. T. et al. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. **Nature**, v. 298, n. 5873, p. 447–451, 1982.
- 43 SUZUKI, G. E.; GERSTEIN, M. Binding geometry of alpha helices that recognize DNA. **Proteins**, v. 23, 1995. doi:10.1002/prot.340230407.
- 44 GARVIE, C. W.; WOLBERGER, C. Recognition of specific DNA sequences. **Molecular Cell**, v. 8, n. 5, p. 937–946, 2001.

- 45 LUSCOMBE, N. M. et al. An overview of the structures of protein-DNA complexes. **Genome Biology**, v. 1, n. 1, 2000. doi:10.1186/gb-2000-1-1-reviews001.
- 46 ARAVIND, L. et al. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. **FEMS Microbiology Reviews**, v. 29, n. 2, p. 231–262, 2005.
- 47 ROHS, R. et al. Origins of specificity in protein-DNA recognition. **Annual Review of Biochemistry**, v. 79, p. 233–269, 2010. doi:10.1146/annurev-biochem-060408-091030.
- 48 GAJIWALA, K. S.; BURLEY, S. K. Winged helix proteins. **Current Opinion in Structural Biology**, v. 10, n. 1, p. 110–6, 2000.
- 49 LUSCOMBE, N. M.; LASKOWSKI, R. A.; THORNTON, J. M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. **Nucleic Acids Research**, v. 29, n. 13, p. 2860–2874, 2001.
- 50 ALBERTS, B. et al. DNA-binding motifs in gene regulatory proteins, New York: Garland Science, 2002.
- 51 FILLMORE, J. **A note on rotation matrices**. 1984. Disponível em: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4055639.html>. Acesso em: 22 jun. 2017.
- 52 ALQURAISHI, M.; TANG, S.; XIA, X. An affinity-structure database of helix-turn-helix: DNA complexes with a universal coordinate system. **BMC Bioinformatics**, v. 16, n. 1, p. 390, 2015.
- 53 MA, J.; KARPLUS, M. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 15, p. 8502–8507, 1998.
- 54 KEARSLEY, S. K. On the orthogonal transformation used for structural comparisons. **Acta Crystallographica Section A**, v. 45, n. 2, p. 208–210, 1989.
- 55 JONES, A. **A to Z of O**. Disponível em: <http://xray.bmc.uu.se/alwyn/TAJ/Home.html>. Acesso em: 11 jun. 2017.
- 56 HAYWARD, S.; BERENDSEN, H. J. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. **Proteins**, v. 30, n. 2, p. 144–154, 1998.
- 57 GÖ, N. A theorem on amplitudes of thermal atomic fluctuations in large molecules assuming specific conformations calculated by normal mode analysis. **Biophysical Chemistry**, v. 35, n. 1, p. 105–112, 1990.
- 58 TIRION, M. M. Large amplitude elastic motions in proteins from a single-parameter, tomic analysis. **Physical Review Letters**, v. 77, n. 9, p. 1905–1908, 1996.
- 59 BECKER, O. M. et al. **Computational Biochemistry and Biophysics**, New York: Marcel Dekker, 2001.
- 60 BAHAR, I. et al. Global dynamics of proteins: bridging between structure and function. **Annual Review of Biophysics**, v. 39, n. 1, p. 23–42, 2010.

-
- 61 ZHENG, W.; BROOKS, B. R.; THIRUMALAI, D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. **Biophysical Journal**, v. 93, n. 7, p. 2289–2299, 2007.
- 62 ALEXANDROV, V. Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. **Protein Science**, v. 14, n. 3, p. 633–643, 2005.
- 63 HINSEN, K. **Normal mode theory and harmonic potential approximations**. 2005. Disponível em: <<http://www.crcnetbase.com/doi/pdfplus/10.1201/9781420035070.ch1>>. Acesso em: 22 jun. 2017.
- 64 DAVID, C. C.; JACOBS, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. **Methods in Molecular Biology**, p. 193–226, 2014. doi:10.1007/978-1-62703-658-011.
- 65 BLOOMFIELD, P. **Fourier analysis of time series: an introduction**, New York: John Wiley & Sons, 2004.
- 66 PRESS, W. H. et al. **Numerical recipes in Fortran 90: the art of parallel scientific computing**. Cambridge: Cambridge University Press, 1996.
- 67 DAUBER-OSGUTHORPE, P.; OSGUTHORPE, D. J. Extraction of the energetics of selected types of motion from molecular dynamics trajectories by filtering. **Biochemistry**, v. 29, n. 36, p. 8223–8228, 1990.
- 68 DAUBER-OSGUTHORPE, P. et al. Low frequency motion in proteins. **Journal of Computational Physics**, v. 151, n. 1, p. 169–189, 1999.
- 69 LINDAHL, E.; DELARUE, M. Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization. **Nucleic Acids Research**, v. 33, n. 14, p. 4496–4506, 2005.
- 70 TAMA, F.; SANEJOUAND, Y.-H. Conformational change of proteins arising from normal mode calculations. **Protein Engineering Design and Selection**, v. 14, n. 1, p. 1–6, 2001.
- 71 LINDAHL, E. et al. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. **Nucleic Acids Research**, v. 34, n. Supplement 2, p. 52–56, 2006.
- 72 KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577–2637, 1983.
- 73 NORTHRUP, S. H. et al. Molecular dynamics of ferrocycytochrome c. **Journal of Molecular Biology**, v. 153, n. 4, p. 1087–1109, 1981.
- 74 Schrodinger LLC. **The PyMOL Molecular Graphics System, Version 1.8**. 2015.
- 75 BEST, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ψ and side-chain χ_1 and χ_2 dihedral angles. **Journal of Chemical Theory and Computation**, v. 8, n. 9, p. 3257–3273, 2012.

- 76 PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1781–1802, 2005.
- 77 MARTYNA, G. J.; TOBIAS, D. J.; KLEIN, M. L. Constant pressure molecular dynamics algorithms. **Journal of Chemical Physics**, v. 101, n. 5, p. 4177–4189, 1994.
- 78 BRÜNGER, A. T. **X-PLOR**: a system for X-ray crystallography and NMR, version 3.1, New Haven: Yale University, 1992.
- 79 BJELKMAR, P. et al. Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. **Journal of Chemical Theory and Computation**, v. 6, n. 2, p. 459–466, 2010.
- 80 ABRAHAM, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 2015. doi:10.1016/j.softx.2015.06.001.
- 81 QIU, D. et al. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. **Journal of Physical Chemistry A**, v. 101, n. 16, p. 3005–3014, 1997.
- 82 ZHU, C. et al. Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization. **ACM Transactions Mathematical Software**, v. 23, n. 4, p. 550–560, 1997.
- 83 BROOKS, B. R. et al. CHARMM: The biomolecular simulation program. **Journal of Computational Chemistry**, v. 30, n. 10, p. 1545–1614, 2009.
- 84 MACKERELL, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. **Journal of Physical Chemistry B**, v. 102, n. 18, p. 3586–3616, 1998.
- 85 BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P. The missing term in effective pair potentials. **Journal of Physical Chemistry**, v. 91, n. 24, p. 6269–6271, 1987.
- 86 SCHMID, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. **European Biophysics Journal**, v. 40, n. 7, p. 843, 2011.
- 87 HESS, B. et al. LINCS: a linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, 1997.
- 88 DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. **Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, 1993.
- 89 BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **Journal of Chemical Physics**, v. 126, n. 1, p. 14101, 2007.
- 90 PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: a new molecular dynamics method. **Journal of Applied Physics**, v. 52, n. 12, p. 7182–7190, 1981.

-
- 91 ANDERSEN, H. C. Molecular dynamics simulations at constant pressure and/or temperature. **Journal of Chemical Physics**, v. 72, n. 4, p. 2384–2393, 1980.
- 92 CHOW, K.-H.; FERGUSON, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. **Computer Physics Communications**, v. 91, n. 1, p. 283–289, 1995.
- 93 LEVITT, M. et al. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. **Computer Physics Communications**, v. 91, n. 1, p. 215–231, 1995.
- 94 HARRISON, S. C.; AGGARWAL, a. K. DNA recognition by proteins with the helix-turn-helix motif. **Annual Review of Biochemistry**, v. 59, p. 933–69, 1990. doi:10.1146/annurev.bi.59.070190.004441.
- 95 STEVENSON, C. E. M. et al. Investigation of dna sequence recognition by a streptomycete marr family transcriptional regulator through surface plasmon resonance and x-ray crystallography. **Nucleic Acids Research**, v. 41, n. 14, p. 7009–7022, 2013.
- 96 DE SIMONE, A. et al. Limited tendency of α -helical residues to form disulfide bridges: a structural explanation. **Journal of Peptide Science**, v. 12, n. 12, p. 740–747, 2006.
- 97 WAKSMAN, G. et al. Crystal structure of Escherichia coli thioredoxin reductase refined at 2 Å resolution: implication for a large conformational change during catalysis. **Journal of Molecular Biology**, v. 236, n. 3, p. 800–816, 1994.

Apêndices

**APÊNDICE A – ARTIGO EM FASE DE CORREÇÃO PARA SER SUBMETIDO
EM BREVE NA PLOS ONE**

Computer simulations reveal changes in collective motions that regulate DNA-binding affinity in the transcriptional regulator MosR.

Short title: Collective motions in DNA-binding regulation.

Key words: Low-frequency motions, allostery, long molecular dynamics, conformational change, conformational space, frequency filter.

Authors:

Amanda Souza Câmara

Instituto de Física de São Carlos, Universidade de São Paulo.

Eduardo Horjales Reboredo

Instituto de Física de São Carlos, Universidade de São Paulo.

Work was performed at Instituto de Física de São Carlos, Universidade de São Paulo.

Correspondence author:

Eduardo Horjales Reboredo

Instituto de Física de São Carlos, Universidade de São Paulo.

Avenida João Dagnone, nº 1100 - Jardim Santa Angelina.

e-mail: horjales@ifsc.usp.br

Abstract

MosR is a transcriptional regulator from *Mycobacterium tuberculosis*. It senses the environment oxidation and regulates the expression of a secreted oxidoreductase, thus defending the bacilli against oxidative stress from the phagosome. While most of the members of MarR family are ligand-responsive, MosR may dissociate from its DNA site upon formation of an intrachain disulphide bond. However, the structure of MosR in its oxidized state is not known, and it is not clear how the formation of this disulphide bond would lead to conformational changes, required for dissociation of the DNA. Nonetheless, MosR presents two crystallographically resolved conformations in its reduced state, bound and unbound to DNA. We managed to simulate MosR unbound to the DNA, in the presence or in the absence of the disulphide bond. We propose a set of analysis to describe collective motions, retrieving low frequency motions from the simulated trajectories, and to verify global parameters that distinguish different conformational population, regarding its affinity for the DNA site. Our results indicate that this disulphide bond stabilizes a salt bridge and thus affects the collective motions of the DNA-binding domain. Comparing to other members of the family, one can see that local interactions near this conserved hinge region also holds these two helices closer and interfere in the collective motions in a similar way. And though the residues in this region might change from one protein to another, the proposed mechanism with which the collective motions are regulated is probably conserved in the family.

Introduction

Mycobacterium tuberculosis is the causative agent for tuberculosis. Once it reaches the lungs of its host, it is phagocytized by macrophage cells, as a defence mechanism of our body response to the exogenous agent. Inside the phagosome the bacilli thrives in a very harmful environment. While the macrophage is trying to lysate it, it releases several defences, managing to prevent the maturation process needed to the phagosome to fuse with the lysosome. As long as the immune system of its hostage is not injured, the bacilli may live there for several years, without ever causing tuberculosis symptoms to appear [1].

To prosper inside the phagosome, *M. tuberculosis* must have evolved a very sensitive perception of its surroundings, and also a prompt response to any disturb. In this scenario,

MosR plays an indispensable role. It is a transcriptional repressor that senses the environment oxidation and regulates the expression of a secreted oxidoreductase, thus affecting the pH in the phagosome interior. Hence its name, *M. tuberculosis* oxidation sense Regulator [2]. With this functional description, MosR is characterized into the MarR family (Multiple antibiotics resistance Regulator), which includes critical proteins for controlling of bacterial response to oxidative stresses and antibiotics, as well as control of virulence factor production and catabolism of environmental aromatic compounds [3]. Proteins of this family share structural similarities that are also present in MosR. They are homodimers of triangular shape, presenting a dimerization domain, which connects to two DNA-binding domains (DBD) (see figure 1). The dimerization is stabilized mainly by hydrophobic contacts between intercalating N- and C-terminal α -helices. In particular, the long helices $\alpha 1$ and $\alpha 5$ are the connection to the DBD, described as a winged helix motif [3, 4]. Proteins of this family may also contain ligand binding sites that reside in the DBD, where the ligand appears as a competitive effector; or in between the dimerization domain and the DBDn, in which case the ligand participates of an allosteric mechanism that promotes large conformational changes to the DNA binding site [3, 5].

The crystallographic structure of MosR has already been elucidated in its reduced conformation, bound and unbound to DNA [2]. An effector site is described [2] to lie in the dimerization domain, with two cysteine residues susceptible to the formation of a reversible disulphide bond. These two cysteines are important for the environment oxidation sensing mechanism, as concluded by experiments performed by Brugarolas and collaborators [2]. From electrophoretic mobility shift assays of MosR with DNA upon increasing concentrations of oxidants, they have detected the dissociation of the protein upon the addition of 0.5-1 mM of oxidant. Using *in vivo* RT-PCR assays, they were able to monitor the levels of the oxidoreductase mRNA, and concluded that it was increased upon addition of H₂O₂. Moreover, punctual mutation of one of the cysteine residues, combined with an increased concentration of oxidants, they have identified partial dissociation of the MosR from DNA, and lower levels of expression of the oxidoreductase mRNA. This could indicate the presence of additional oxidation-sensing mechanisms, but the overall conclusion is that MosR has greater affinity for its specific DNA site in its reduced form, while in its oxidized form it is not likely for binding, allowing the oxidoreductase to be expressed [2]. However, the structure of the oxidized state remains unknown and it is not

still clear how the disulphide bond formation would lead a protein molecule to dissociate from its DNA site.

OhrR, a homologous protein of MosR (28% of amino acids sequence identity when compared to OhrR from *B. subtilis* and 25% identical compared to OhrR from *X. campestris*), also sensitive to oxidation, had its molecular structure elucidated in its oxidized state (from *Xanthomonas campestris*) [6]. Both proteins may display a reversible disulphide bond, where one cysteine residue is in a conserved position, surrounded by aromatic amino acids (typically found in oxidation-sensing regulators [7]) and experimental results for both indicate that this specific bond may alter the affinity of the protein for its DNA site. Precisely, the oxidized structure of OhrR indicates a rotation of the recognition helix $\alpha 4$, compared to the reduced conformation, which prevents the protein from binding to its DNA site when oxidized [6]. This rotation arises when Cys22, in the N-terminal of helix $\alpha 1$, forms a disulphide bond with Cys127', in long helix $\alpha 5'$, which connects the dimerization domain to the DBD. This strong interaction causes helix $\alpha 5$ to break in half, and leads to a different orientation of the DBD.

A striking difference between MosR and OhrR lies in which residue is capable of forming the disulphide bond. In OhrR the residues are the Cys22 from one chain and the Cys127' from the other, holding close two different helices, $\alpha 1$ and $\alpha 5'$, and symmetrically, Cys22' from $\alpha 1'$ and Cys127 from $\alpha 5$. While for MosR, the bond is formed between Cys10 and Cys12 from the same chain and same helix $\alpha 1$, and symmetrically, Cys10' and Cys12' in helix $\alpha 1'$. Therefore, it is not likely that the disulphide bonds in MosR would promote a break in the long helices $\alpha 5$ and $\alpha 5'$, as in OhrR. Nevertheless, a rotation of helices $\alpha 4$ and $\alpha 4'$ is also observed in MosR by comparison of its two resolved structures with two different states: bound and unbound to DNA.

Comparing the two crystallographic structures, it is possible to observe major differences on the DBD, and that helix $\alpha 4$, which fits into the major groove of the DNA helix, is rotated for about 25° from one conformation to the other, similarly to OhrR upon the formation of the disulphide bond. These differences are observed, albeit the crystallographic structures are both at the reduced state. Nevertheless, the reduced state of MosR is also the active state, and unbound to DNA, it may be dynamically changing between different conformational populations. Thus, it is likely that the structure of MosR reduced and unbound to DNA resembles its non-binding oxidized structure, and to analyse the different

features between bound and unbound conformations may bring insights on the oxidized conformation and how the changes caused by the disulphide bond may lead to dissociation from DNA.

Another difference is that the conformation unbound to DNA presents three more residues at the N-terminus loop than the bound conformation, indicating greater flexibility of this region when the protein is bound to DNA. Interestingly, the N-terminus loop resides in a key position: it is not only closely followed in sequence by the two cysteines that may form the disulphide bond, but it is also located in a region responsible for great conformational changes in the protein, in-between helix $\alpha 5$ and the DBD. However, neither of the resolved structures present the coordinates for the entire N-terminus. Indeed, five residues are missing in the unbound conformation, and the bound conformation lacks eight residues. This short but rather flexible disordered region may interact with helix $\alpha 5$ as well as with helices $\alpha 2$ and $\alpha 3$ from the DBD, and it maybe affecting the motions of this domain.

This greater flexibility of the DNA binding domain, which moves as a rigid body with respect to the dimerization domain [3], is essential for MosR to adopt different conformations as a mechanism to regulate its affinity for DNA. Because the geometry of two consecutive major grooves on the DNA double helix structure requires the DNA recognition α -helix of the protein to adopt a particular orientation [8], some conformations would be more likely of binding to DNA than others. Therefore, the disulphide bond on MosR should prevent it from reaching conformations of greater affinity to its DNA site. This may be done by decreasing the flexibility of the DNA binding domain and holding the protein molecule in a conformation where the recognition α -helices do not present a most favourable orientation [9].

A protein conformation is favourable for binding when the α -helices responsible for the DNA recognition have the same orientation as the two consecutive major grooves. Thus, the orientation of the recognition α -helices determines the ability of the protein to bind to DNA. Once the recognition residues could interact with specific bases in the DNA grooves a fine tune of the DNA binding can be done by the rest of the residues and the α -helix main chain can adopt a better fit orientation [10]. Thus, the binding of the protein to its DNA site requires both collective and local conformational changes: collective changes to approximate both α -helices to two consecutive major grooves, and local deviations to fine tune the binding.

These collective and local conformational changes are related to the number of atoms they involve, and the total inertial mass of these group of atoms, subjected to the same kinetic energy, corresponds to low or high frequency motions, respectively. Low frequency motions involve residues from the same region moving as a rigid body, while high frequency motions are characteristic from side chain rotations or even bond vibrations [11, 12]. Thus, it takes longer times to approximate the α -helices than to specifically fit their orientation.

Because the disulphide bond in MosR is located away from the DBDn, for the oxidation to promote a change in the DNA-binding-affinity, it must affect a low frequency motion that connects both sites in the same group of atoms. Thus, it probably rather affects the low frequency motions than the high frequency ones, preventing the ligand α -helices to approximate to an adequate orientation. In fact, many works have been emphasizing how important it is, for the biological functioning of the protein, to visit a broad conformational ensemble. A recent work highlights the importance of the internal motions of proteins on regulating activity [13]. It demonstrates, using NMR spectroscopy and thermodynamic data, how much is the internal dynamics of the catabolite activator protein (fast and slow motions) responsible for its binding to DNA.

Few experiments can reveal about low frequency motions of proteins. NMR spectroscopy, laser Raman spectroscopy, and TLS parameters and anisotropic displacements from crystallography are among them. In this work we make a computational approach, simulating MosR for long molecular dynamics, to uncover how the formation of a disulphide bond could affect the accessible conformations during thermal motions of the protein and its affinity for DNA.

Because the N-terminal region maybe crucially affecting the flexibility of the DBD, we have started our simulations with a set of possible positions for this unresolved region and present here a detailed analysis of the interactions it may form in the presence or in the absence of the disulfide bonds.

Moreover, because the biological conformational changes are related to low frequency motions, we focused on retrieving these motions from molecular dynamics simulations of both states of MosR: reduced and oxidized. Therefore we use an analysis method, based on the Fourier filter transform proposed by Oguthorpe [14, 15]. With this tool it is possible to

observe the collective motions that appear in a simulated trajectory, while relating them to the frequency they occur with, and to non-permanent structural features.

The application of this method on the simulations of MosR showed distinct dynamics of the DBDs depending on the oxidation state of the protein and on the N-terminal interactions. These results not only endow the known regulatory function of MosR, but also improve the discussions on how to identify and analyse collective conformational changes, and moreover, on the role of dynamics in allosteric regulation.

Materials and Methods

The molecular dynamics simulations

All simulations were carried out using Gromacs program [X] and the united-atom force field Charmm27 [X]. Solvation of the system was explicit considered using SPCE model for water molecules and 16 chlorine counterions in a box whose sides are minimum 1 nm away from a protein atom.

The original structure of MosR used to perform the simulations was the crystallographic conformation unbound to DNA resolved by Brugarolas and collaborators (PDB entry 4fx0) [2]. This structure presents missing residues mainly in chain A, hence a starting conformation for the simulations was built using a copy of chain B superposed to the original chain A. Residues 1 to 5, 51 to 53, 92 to 95 and 148 were previously added to chain B using Coot program and the regularize zone routine [X]. Additionally, 12 different positions of the N-terminal residues were tested. Each conformation was minimized for 1,000 steps using steepest descent algorithm. An equilibration was performed for 100 ps restraining the position of all protein atoms in a canonical *ensemble* at 310 K, followed by another 100 ps long step of equilibration still restraining the protein but in an isobaric-isothermal *ensemble* at 1 bar. A third equilibration step was taken for 1 ns in this same *ensemble* allowing the N-termini (residues from 1 to 7) and the solvent to move, and the system was gradually heated from 0 to 310 K during 500 ps. All these equilibration steps were calculated using periodic boundary conditions, LINCS algorithm to constraint the hydrogen bonds, PME algorithm to calculate electrostatic forces, V-rescale algorithm on the temperature control, Parrinello-Rahman on the pressure control when necessary and the steps were 2 fs long.

Two of these conformations were chosen, for their stabilities, to proceed with a 1 ns long step of equilibration allowing the entire protein to move in an isobaric-isothermal *ensemble* gradually heated from 0 to 310 K during 500 ps. A 400 ns long molecular dynamics was then simulated for each one, saving conformations each 10 ps.

Still with these two conformations, after the N-terminal equilibration, we manually chose another conformer for each Cys12, bringing the sulfur atom closer to the one in Cys10, using the regularize zone routine from Coot program. A new topology for these structures was then built, which constitute the oxidized state of MosR. These two oxidized conformations were minimized and equilibrated following the same protocol used with the original reduced conformations, except for the N-terminal equilibration, and a 400 ns long molecular dynamics was also simulated for each one, again saving conformations each 10 ps.

A third conformation was chosen to simulate a 100 ns long molecular dynamics using a distance restraint for the two pair of atoms consisting in the nitrogen atom from Ala6 and the oxygen atom from Arg65 of the opposite chain. This restriction was imposed by setting the parameters low, up1 and up2 with the values of 0.24 nm, 0.4 nm and 0.5 nm, respectively. Minimization and equilibration were also under these restraints, besides the above mentioned default conditions, as well as the molecular dynamics simulation, from which a conformation was saved each 10 ps.

Methods of analysis

The results showed in this work came from three different methods of analysis. We wrote our own algorithms in Fortran 90 to perform them. They are based on the subroutines from [21] to read and write trajectory files in dcd format and recipes from [22] to calculate the Fourier transform of the simulated trajectories and the principal inertia axes of the α helices.

The first method is simply for measuring geometrical quantities of the protein's structure during the simulated trajectory, such as distances between chosen atoms or centres of mass, principal axes of α helices, and the angle between them. They allow us to characterize the orientation of the recognition helices in the DNA-binding domain, and are described with more details in the supplementary material.

The second method, also described in the supplementary material, is for calculating isotropic and anisotropic displacements as defined in [23, 24]. They give us information on the flexibility of the protein and on the direction of the motion it performs during the simulations.

The third is the method already mentioned for retrieving low-frequency motions from molecular dynamics, which we adapted from the filtering method of Osguthorpe, and it is also explained below.

For visual analysis we used PyMOL [20] to represent the ellipses of the anisotropic atomic displacements and VMD [25] for the representation of the atomic displacement vectors by arrows.

The filtering analysis

In signal analysis it is common the usage of Fourier transform to filter frequencies [26, 22]. This is usually applied to remove noisy frequencies, as in image processing. As showed in [14, 27] this method can be performed to filter any time dependent parameter extracted from a molecular dynamics, including the 3N coordinates from the protein trajectory (where N is the number of atoms of the protein). This last application results in a filtered trajectory, where high or low frequency motions can be neglected, allowing the analysis of the protein motion in certain frequencies of interest.

In [15] this filtering method is used to extract harmonic motions, selecting one specific frequency, and thus compare to the modes obtained by Normal Mode Analysis. However, this process maintains the asymmetry present in the simulated trajectories and it results on different phases for each coordinate. We preferred here to generate, from the simulated trajectory, a new function of time, symmetric around the origin, with twice the time interval and that still obeys the equations of motion of Newton. Then, the Fourier transform allows us to select only one frequency and designate to it a single phase and amplitude in a harmonic motion for the entire protein, as in equation (1),

$$\overrightarrow{\Delta x_f} = \widehat{w}_f S_f \sin(\omega_f t) \quad (1)$$

where the subindex f stands for each calculated mode (the total number of modes is equal to the total number of considered conformations), $\widehat{w}_f$ is a versor that represents the direction

of each harmonic motion of the protein in the 3N space, S_f is the squared amplitude of this motion, ω_f its frequency, and $\overrightarrow{\Delta x}_f$ is the displacement vector of the entire protein in each harmonic mode. A more detailed explanation lies on the supplementary material, and a summary for the method, consisting of six steps, is listed below:

1. Take the boundary values of the function and equal them to zero, by removing an end-to-end straight line over all time steps;
2. Generate an odd symmetry;
3. Take the sine transform;
4. Filter one or a range of frequencies;
5. Transform back to the time domain;
6. Add the linear coefficient from the straight line calculated in 1.

This filtering can be made over any property of the molecular dynamics that is a function of time. But when filtering the trajectory of a molecule, these steps should be taken over all of its 3N coordinates separately, resulting in unidirectional motions of the entire protein in the 3N space, which may be represented by atomic displacement vectors for each atom of the protein, in a similar way as Normal Mode Analysis (NMA) or Principal Component Analysis (PCA).

An advantage of this method is that it directly relates a frequency to an amplitude of a harmonic motion retrieved from the molecular dynamics trajectory, which is simulated in specific conditions of temperature, pressure and constraints. This fills a blank left by PCA, which cannot associate a frequency to the determined modes, and also gets around a misleading result from NMA, which incorrectly approximates the broad potential energy wells to the narrow wells present at the bottom of the potential energy landscape [28, 29].

However, the harmonic modes calculated by the Sine Filtering Analysis we propose does not form an orthonormal basis. The method actually provides as many modes as the number of given conformations. And the range of frequencies it may analyse is restricted to the time window of the trajectory, as in equation 2,

$$\omega_f = \frac{f}{2T\delta} \quad , \quad f = 1, \dots, T \quad (2)$$

where ω_f is the frequency, f is the mode number, T is the total number of conformations and δ is the time step between each conformation.

Nevertheless, they may be quantitatively analysed and compared not only by the associated frequency and amplitude, but also by dot products to other harmonic modes, that indicates the similarity on the direction of the motion. Thus providing complementary information to the other mentioned methods.

In this work we only compare modes obtained by this filtering method. We compare them by their frequencies, amplitudes and degree of collectivity. The amplitude is obtained from equation 1, and it can be calculated for any set of atoms. We consider all the α carbon atoms for the global amplitude, and only the α carbon atoms from the recognition helices for the amplitude of the recognition helices. The degree of collectivity κ_f is considered as suggested by Tama and Sanejouand (X) and it is calculated as in equation 3, where Δx_{if} is the displacement of each coordinate i of the mode f , as calculated in equation 1; S_f is the squared amplitude, also from equation 1; and N is the total number of coordinates, which we considered to be all α carbon atoms.

$$\kappa_f = \frac{1}{N} \exp \left(- \sum_{i=1}^N S_f \Delta x_{if}^2 \log S_f \Delta x_{if}^2 \right) \quad (3)$$

Results and discussion

The results of four simulations are showed below, each one is 400 ns long. These four simulations differ from each other in the initial position of the N-termini or the oxidation state of the protein. By oxidation state of the protein we refer to the existence of two intrachain disulphide bonds, built as explained earlier, between residues Cys10 and Cys12. In the oxidized state both disulphide bonds are present, while in the reduced state they are not. Simulations 1 and 2 are in the reduced state, while 3 and 4 are in the oxidized state. Also, simulations 1 and 3 share the same initial conformation, whereas 2 and 4 share another initial conformation.

The following results are an attempt to characterize the common and the distinct features of the apo MosR in its oxidized or reduced state, and moreover, how this distinct features affect the affinity for the DNA site.

Figure 1 shows both simulated states of MosR superposed, calculated as the average conformation from simulations 1 and 2, for the reduced state, and 3 and 4 for the oxidized state. Table 1 indicates the RMSD of each conformation to the other and to the crystallographic structures. Though simulated in different states, both conformations have a lower RMSD compared to the apo structure than to the holo structure, and moreover, an average conformation of all four simulations presents a RMSD from the apo crystallographic structure of only 1.0 Å. Nevertheless, a few differences between the oxidized state and the reduced state can be observed, in accordance to the differences also observed between the crystallographic structures (brugarolas). The formation of the disulfide bond at the beginning of α helix 1 causes an extra tension at the next turn, which adopts a slightly different structure (see table 2), besides, an apolar region appears around the disulfide bond, which repels charged residues close to it. Indeed, Asn37 appears to be farther from Cys12' in the simulated oxidized state and in the apo crystallographic structure, than in the reduced state and the holo crystallographic structure (see table 2). However, a maximum rotation of only 5.9° (see figure 1) is observed between the recognition helices of the oxidized and the reduced state. Therefore, the next set of results evaluate whether the simulations can reproduce the experimentally observed difference between the two states regarding the affinity for the DNA site, which we treat here as a consequence of the positioning and orientation of the recognition helices. Other differences are observed between the two crystallographic structures, but some of them maybe caused by the interaction with the DNA and its difficult to attribute them to a state of the protein, reduced or oxidized. Nevertheless, the recognition helices are known to be the first to interact with DNA (X), and thereby the overall changes on their position and orientation must occur prior to the interaction with DNA. Therefore, we have detained our analysis on the recognition helices to characterize the ability of the protein to bind to DNA.

	apo structure (pdb entry 4fx0)	Holo structure (pdb entry 4fx4)	Oxidized average conformation
Reduced average conformation	1.112	1.887	1.665
Oxidized average conformation	1.369	2.684	
Holo structure (pdb entry 4fx4)	1.520		

	Apo structure (pdb entry 4fx0)	Holo structure (pdb entry 4fx4)	Oxidized average conformation	Reduced average conformation
Chain A	6.2	6.5	7.3	5.9
Chain B	6.1	6.1	7.5	6.0

The recognition helices

Figure 2 shows the RMSD of the recognition helices in each simulation compared to the recognition helices of the crystallographic structure bound to DNA (PDB entry 4fx4). It is clear that simulation 1, in the reduced state, presents the conformations which most approximate to the crystallographic structure (the least RMSD observed is 0.9 Å). Nevertheless, simulation 3, in the oxidized state, also presents conformations that considerably approaches the holo structure (the least RMSD observed is 1.3 Å). Simulation 2 also present a least RMSD of 1.3 Å, while simulation 4 never shows a RMSD lower than 1.5 Å.

Though low RMSD values are certain to represent very similar conformations, the higher the values are, the more poorly represented are the similarities. We then describe the relative orientation of the recognition helices, related to each other, in a less ambiguous way, by three parameters (see the methods section). The direct dot product between the recognition helices axes do not separate well conformations from simulations 1 and 3, and thereby figure 3 shows only the other two parameters: the distance between the two helices, which is normal to both axes; and the distance between their centres of mass, which is orthogonal to the direction of the latter. The crystallographic holo conformation is also represented in figure 3 by an orange circle. In this analysis, it is clear how the conformations of simulation 1 are the ones which most approaches the crystallographic conformation.

Relative distance of the two recognition helices

Another description may be done to confirm this result. In figure 4 the simulated conformations are plotted according to the dot product of each recognition helices and its respective helices in the crystallographic apo structure. Thus, though the RMSD between them may be low in simulation 3 (see figure 1), it is clear once again that the conformations from simulation 1 are the ones which most resembles the crystallographic holo structure.

angle between each helix of the holo structure

Interesting, conformations in simulation 1 also seem to be more spread out than in the other simulations (see figures 3 and 4) in the space defined by these selected parameters, presenting a variety of orientations for the recognition helices. Figure 5 compares the flexibility of MosR in each simulation, represented by the isotropic atomic displacement parameter calculated for time windows of 1 ns and 50 ns. For 1 ns, MosR presents nearly the same flexibility in all simulations. While for a 50 ns time window, it is clear that MosR in simulations 1 and 3, which share the same initial conformation, is more flexible than in the other two simulations. The flexibility of the DBD helices is even higher for the chain B in simulation 1. The different time windows allow one to distinct local and collective motions. Therefore, this results indicate that the variety of orientations of the recognition helices observed in simulation 1 may arise from a collective motion of the DBD in chain B.

Isotropic atomic displacement parameter

isotropic atomic displacement parameter

This set of results also shows that MosR is more likely to achieve a conformation able to bind to DNA in a reduced state, since it presents higher flexibility (see figure 5). However, MosR in simulation 2 was also in a reduced state, but did not show such capability to approach the DNA-binding orientation of the recognition helices. Thereby, the initial conformation of the N-terminus is probably affecting the accessible space of the recognition helices. The next set of results are in regard of this possibility.

On the N-terminus contribution

Figure 6 shows the contact maps for both N-termini of MosR in each simulation. Only contacts between the first 18 residues of one chain and the entire other chain are considered. A contact is accounted for each time two α carbons are not more than 6 Å apart from each other. From the crystallographic structures, which present missing residues in the N-termini, one may expect that the N-termini are very flexible and can transiently interact with a broad number of residues surrounding them in the neighbourhood. Thus, the

counting of contacts may vary from one simulation to another. If these interactions are rapidly changing and the simulations are long enough, then the pattern of contacts would be similar in all simulations. For example, in all simulations residue 4 appears close to residues from α helix 2 and the loop that connects to α helix 3 of the other chain, but with slightly different rates. In the other hand, residue 4 is also frequently observed to be close to residues from α helix 5 of the other chain. Particularly, however, the N-terminus of chain A in simulation 1 hardly contacts the residues from α helix 5, and contacts are rather observed with residues from α helix 3 of chain B, which correspond to the first helix of the HTH. Both N-termini in simulation 3 also appear to be making contact with α helix 3, thus this interaction was probably favoured by the initial conformation. However, the N-terminal residues participating are slightly different in each simulation. Looking specifically into the polar contacts made by these N-terminal residues and α helix 3 (supplementary material) it is clear that the N-terminus from chain A in simulation 1 is the only one to present an interaction between the nitrogen atom of Ala6 and the oxygen atom from Arg65 of chain B.

To verify the behaviour of this specific interaction found in simulation 1, we have monitored the distance between the two atoms involved. Figure 7 shows this distance as a function of time for both N-termini of MosR in simulation 1. It is clear that this polar interaction is not permanent, but it endures for half of the simulation. Coincidentally, the orientation of the recognition helices which best fit to the holo crystallographic conformation also appears during the first half of this simulation, while this interaction between Ala6 of chain A and Arg65 from chain B is present.

We next verified whether this interaction could be affecting the flexibility of the DBD, as chain B in simulation 1 was already observed to be the most flexible (see figure 5). Hence we have divided simulation 1 into two halves, and figure 8 shows the calculated isotropic atomic displacement for each part. Evidently, from this analysis, chain B in the first half is more flexible than in the second half. Moreover, figure 9 shows the anisotropic atomic displacements calculated for each half of the simulation, confirming that this higher flexibility of the DBD from chain B is key to allow the best positioning of the recognition helices into the major grooves of the DNA.

To further evaluate the differences in the motions that MosR performed in both halves of simulation 1, we have retrieved the motions associated to the lowest frequencies. Figure 10 represents the motion with the lowest frequency of both halves with arrows for each α carbon, and table 3 indicates the frequency, amplitude and the degree of collectivity for the five motions with the lowest frequencies (see supplementary material for animated motion videos). Interesting, the harmonic motions retrieved from the first half have higher amplitudes and higher degrees of collectivity than the ones from the second half. It is possible that the increased degree of collectivity arises from a rearrangement of the

structure provided by the interaction between Ala6 from chain A and Arg65 from chain B. Indeed, the average conformation from this half of the simulation considerably deviates from the average conformation from the second half (see figure 9).

	First half			Second half		
Frequency (MHz)	Global Amplitude (Å)	Recognition helices amplitude (Å)	Degree of collectivity	Global amplitude (Å)	Recognition helices amplitude (Å)	Degree of collectivity
30.5	66.6	24.8	0.38	60.3	13.6	0.22
61.0	25.3	10.0	0.49	29.4	9.9	0.41
91.5	14.7	5.8	0.48	17.9	3.8	0.22
122.1	10.7	4.8	0.43	17.1	3.3	0.22
152.6	11.1	3.7	0.41	12.9	3.2	0.30

Imposing a restriction

To assure that this specific interaction is in fact associated to an increase of the flexibility and of the affinity for the DNA site, we have simulated a short molecular dynamics in which the distance between the nitrogen atom from Ala6 of one chain and the oxygen atom from Arg65' of the other chain is restrained.

isotropic atomic displacement parameter

relative distance of the two recognition helices

angle between each recognition helix of the holo structure

Though the restriction was applied symmetrically for both pair of atoms, only one pair remains close enough to be considered as an enduring interaction (see supplementary material). Figure 11 shows the isotropic atomic displacements calculated for this new trajectory, as well as the described parameters that indicate the recognition helices approach a conformation similar to the crystallographic holo structure, always comparing with the results obtained for the first half of simulation 1. This set of results shows that chain A from this restrained simulation behaves similarly to chain B in the first half of simulation 1. Thereby, it is possible to associate the detected interaction to the observed differences in the dynamics of MosR that may increase its affinity for the DNA site. Moreover, large conformational changes are also observed when this interaction is present. At last, figure 12 shows the average conformation from this restrained dynamics, and the one from the first half of simulation 1, superposed to the apo crystallographic structure. In this case the recognition helix of chain A is rotated 24.6° from its respective helix in the apo structure, which is compatible with the rotation detected between the apo and holo crystallographic structures.

Conclusions

A first conclusion withdrawn from the simulations is the agreement with some experimental results. The formation of the disulphide bond does promote local changes that mimics real biological characteristics, such as the tension at the beginning of the α helix in which the cysteines are localized (X), and the redistribution of net charges surrounding the disulphide bridge, which is replaced by two protons and two electrons, in the reduced state. Moreover, the average conformation calculated from the four simulations of MosR in the absence of

DNA, which totalize 1.6 μ s, reproduces well the crystallographic apo structure, presenting a RMSD value of only 1.0 Å.

It is a more difficult task to determine if a simulated apo conformation is able to bind to DNA by the direct comparison with the holo structure. Thereunto we have selected three collective variables to describe the position and orientation of the two recognition helices relative to one another. In a space characterized by these variables, conformations of simulation 1, in the reduced state, are more spread out than in the other simulations, closely approaching the crystallographic holo structure. This broader *ensemble* is also related to higher flexibility. By calculating the atomic displacements with different time windows, we could attribute this higher flexibility to low frequency motions, or collective motions. Moreover we could specifically identify that only chain B of MosR in simulation 1 presented this higher flexibility.

The contradictorial dynamical behaviour of MosR in simulations 1 and 2 indicates that the N-terminal residues is probably affecting the flexibility of the DBD. In fact, counting the contacts made between N-terminal residues and the opposite DBD showed an interaction unique to chain B in simulation 1. This interaction, made up by the mainchain atoms of Ala6 from chain A and Arg65 from chain B, directly connects the N-terminus to α helix 3, the first one in the HTH motif of chain B. Forcing the proximity of this two residues in a restrained simulation only reinforced the conclusions that this interaction is indeed associated to a higher flexibility of the DBD and its capability to bind to DNA, but intriguing questions arise from these results. How can an interaction, a restriction in structure, increase the flexibility while it is usually associated to loose and disordered structures? And why is this interaction not observed in the oxidized state?

The studies of Tama and Sanejouand suggest that collective motions are highly subjected to the shape of the protein, rather than to more complex atomic properties (X). Interesting, the low frequency harmonic modes filtered from the trajectory period where that interaction appears in simulation 1 have a higher degree of collectivity than in any other period of the four simulations. The average conformation of chain B from this particular period deviates 1.44 Å from the apo crystallographic structure (RMSD value), while it deviates only 1.08 Å from the holo crystallographic structure. Allaying these RMSD values to the degree of collectivity, one may conclude this period reveals a different shape for MosR, which is closer to the holo structure.

This different shape probably arises from the specific interactions of the N-terminal residues with α helix 3, but it is primarily originated on the surroundings of the cysteines. Upon the formation of the disulfide bond, not only the beginning of α helix 1 assumes a slightly different conformation, as the Asn37 in α helix 2 is pushed away from the apolar disulfide bridge. The distancing of α helix 2 is accompanied by the entire DBD, making it impossible to form a hydrogen bond between the mainchain atoms of Ala6 and Arg65'. Therefore, the intrachain oxidation of cysteines 10 and 12 is able to preclude the functional conformational change and the associated high flexibility of the neighbouring DBD.

A final remark must be made on the analysis method of harmonic mode filtering from the molecular dynamics. It is likely that this method was not thoroughly explored in the early years of 2,000 due to computing limitations, but it does bring a complementary view on the collective motion analysis. Unlike NMA, it does not make approximations on the potential energy landscape, and differently from essential dynamics, it considers the time information provided by the simulations. Moreover, motions with the same frequency maybe retrieved from different periods of the trajectory, as we have showed here. We think this method would be of great value to analyse even longer simulations and to obtain a more realistic clue on the frequency of the motions with biological relevance.

The different harmonic modes we have retrieved from our simulations, are already distinct on the frequency of $1/50 \text{ ns}^{-1}$. However, the different shape observed for MosR maybe associated to a rarer event, since it could not be observed in simulation 2. We conclude that for MosR to bind to DNA it is not only required to be in a reduced state, but also to achieve a subset of conformations with different shape and higher flexibility, perhaps defined by a specific substate.

There still much to learn on the flexibility, frequency and biological functionality of collective motions, but we expect this work may show a way to further develop these issues.

Acknowledgements

We thank Gilmar Bertolote Jr for his help on setting up powerful personal computers. We are also grateful to the USP-Rice BG collaboration for enabling the access to the Blue Gene

supercomputer. And we gratefully acknowledge CNPq for financially supporting Amanda on her PhD and Fapesp for the financial support to the research project 2013/26313-5.

References

- [1] Pieters, J. *Mycobacterium tuberculosis* and the macrophage: maintaining a balance. *Cell host & microbe* 2008;3:399-407.
- [2] Brugarolas P, Movahedzadeh F, Wang Y, Zhang N, Bartek IL, Gao YN, Voskuil MI, Franzblau SG, He C. The oxidation-sensing regulator (MosR) is a new redox-dependent transcription factor in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* 2012;287:37703-12.
- [3] Wilkinson SP, Grove A. Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Current issues in molecular biology* 2006;8:51.
- [4] Gajiwala KS, Burley SK. Winged helix proteins. *Current opinion in structural biology* 2000;10:110-6.
- [5] Lim D, Poole K, Strynadka NC. Crystal structure of the MexR repressor of the mexRAB-oprM multidrug efflux operon of *Pseudomonas aeruginosa*. *Journal of Biological Chemistry* 2002; 277:29253-9.
- [6] Newberry KJ, Fuangthong M, Panmanee W, Mongkolsuk S, Brennan RG. Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Molecular cell* 2007;28:652-64.
- [7] Poole LB, Nelson KJ. Discovering mechanisms of signaling-mediated cysteine oxidation. *Current opinion in chemical biology* 2008;12:18-24.
- [8] Branden C, Tooze J. *Introduction to protein structure*. New York: Garland publisher inc; 1999. 302 p.
- [9] Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. *Nature structural & molecular biology* 2006;13:831-8.
- [10] Kalodimos CG, Boelens R, Kaptein R. A residue-specific view of the association and dissociation pathway in protein–DNA recognition. *Nature Structural & Molecular*

- Biology 2002;9:193-7.
- [11] Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature* 2007;450:964-72.
- [12] Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben IE, Sauke TB, Shyamsunder E, Young RD. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences* 1985;82:5000-4.
- [13] Tzeng SR, Kalodimos CG. Protein activity regulation by conformational entropy. *Nature* 2012;488:236-40.
- [14] Sessions RB, Dauber-Osguthorpe P, Osguthorpe DJ. Filtering molecular dynamics trajectories to reveal low-frequency collective motions: phospholipase A 2. *Journal of molecular biology* 1989;210:617-33.
- [15] Dauber-Osguthorpe P, Osguthorpe DJ, Stern PS, Moult J. Low frequency motion in proteins: comparison of normal mode and molecular dynamics of streptomyces griseus protease A. *Journal of computational physics* 1999;151:169-89.
- [16] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 2005;26:1781-802.
- [17] Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, MacKerell Jr AD. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of chemical theory and computation* 2012;8:3257-73.
- [18] Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of chemical physics* 1995;103:4613-21.
- [19] Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* 1994;101:4177-89.
- [20] The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
- [21] L. Martínez, MDAnalysis. Version 1.1.1.299, Institute of Chemistry - University of Campinas, Campinas, SP, Brazil, 2014.
- [22] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical recipes in fortran 77: the art of scientific computing*. New York: Press Syndicate of the University of

Cambridge, 1996.

- [23] Dunitz JD, Schomaker V, Trueblood KN. Interpretation of atomic displacement parameters from diffraction studies of crystals. *The Journal of Physical Chemistry* 1988;92:856-67.
- [24] <http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>. [Accessed 4 april 2016].
- [25] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics* 1996;14:33-8.
- [26] P. Bloomfield. *Fourier analysis of time series: an introduction*. New York: John Wiley and sons, 2013. 288 p.
- [27] Dauber-Osguthorpe P, Osguthorpe DJ. Extraction of the energetics of selected types of motion from molecular dynamics trajectories by filtering. *Biochemistry* 1990;29:8223-8.
- [28] K. Hinsen, "Normal mode theory and harmonic potential approximations," in *Normal mode analysis theory and applications to biological and chemical systems*, Chapman and Hall/CRC, 2005.
- [29] Hinsen K. *Normal mode theory and harmonic potential approximations*. Boca Raton: Chapman & Hall/CRC. 2006:1-6.
- [30] Frauenfelder H, Chen G, Berendzen J, Fenimore PW, Jansson H, McMahon BH, Stroe IR, Swenson J, Young RD. A unified model of protein dynamics. *Proceedings of the National Academy of Sciences* 2009;106:5129-34.
- [31] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. *Science* 1991;254:1598-603.
- [32] Perera IC, Grove A. Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *Journal of molecular cell biology* 2010:mjq021.
- [33] Saridakis V, Shahinas D, Xu X, Christendat D. Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. *Journal of molecular biology* 2008;377:655-67.
- [34] Perera IC, Lee YH, Wilkinson SP, Grove A. Mechanism for attenuation of DNA

- binding by MarR family transcriptional regulators by small molecule ligands. *Journal of molecular biology* 2009;390:1019-29.
- [35] Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology* 1965;12:88-118.
- [36] Goldstein H, Poole CP, Safko JL. *Classical Mechanics*. Mass: Addison-Wesley, 2001. 683 p.
- [37] Peraire J, Widnall S, "Lecture L26 - 3D rigid body dynamics: the inertia tensor," 16.07 Dynamics MIT course, 2008.
- [38] <https://www.newton.ac.uk/files/seminar/20050804090010001-149165.pdf> [Accessed 6 June 2016]
- [39] Allen TW, Andersen OS, Roux B. On the importance of atomic fluctuations, protein flexibility, and solvent in ion permeation. *The Journal of general physiology* 2004;124:679-90.
- [40] <http://xray.bmc.uu.se/~alwyn/TAJ/Home.html>. [Accessed 6 april 2016].
- [41] Kleywegt GJ, Jones TA. Detecting folding motifs and similarities in protein structures. *Methods in enzymology* 1997;277:525.
- [42] http://xray.bmc.uu.se/alwyn/Essential_O/lsq_frameset.html. [Accessed 6 april 2016].
- [43] Kearsley SK. On the orthogonal transformation used for structural comparisons. *Acta Crystallographica Section A: Foundations of Crystallography* 1989;45:208-10.