



**UNESP UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
INSTITUTO DE QUÍMICA CAMPUS DE
ARARAQUARA-SP**



ANDERSON APARECIDO DO ESPIRITO SANTO

**Estudo sobre a variação da energia livre de
ligação em complexos lectina-peroxidase
através da dinâmica molecular**

Araraquara

2019

ANDERSON APARECIDO DO ESPIRITO SANTO

**Estudo sobre a variação da energia livre de ligação em
complexos lectina-peroxidase através da dinâmica
molecular**

Dissertação apresentada ao Instituto de
Química da Universidade Estadual Paulista,
como parte dos requisitos para obtenção do
título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Gustavo Troiano Feliciano

Araraquara

2019

FICHA CATALOGRÁFICA

S237e Santo, Anderson Aparecido do Espirito
Estudo sobre a variação da energia livre de ligação em complexos lectina-peroxidase através da dinâmica molecular / Anderson Aparecido do Espirito Santo. – Araraquara : [s.n.], 2019
77 f. : il.

Dissertação (mestrado) – Universidade Estadual Paulista, Instituto de Química
Orientador: Gustavo Troiano Feliciano

1. Biomoléculas. 2. Biossensores. 3. Dinâmica molecular. 4. Energia livre. 5. Reconhecimento molecular. I. Título.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara



CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: "Estudo sobre a variação da energia livre de ligação em complexos lectina-peroxidase através da dinâmica molecular"

AUTOR: ANDERSON APARECIDO DO ESPIRITO SANTO

ORIENTADOR: GUSTAVO TROIANO FELICIANO

Aprovado como parte das exigências para obtenção do Título de Mestre em BIOTECNOLOGIA, pela Comissão Examinadora:

Prof. Dr. GUSTAVO TROIANO FELICIANO

Departamento de Físico-Química / Instituto de Química - UNESP - Araraquara

Prof. Dr. EDUARDO MAFFUD CILLI

Departamento de Bioquímica e Tecnologia Química / Instituto de Química - UNESP - Araraquara

Prof. Dr. MAURÍCIO DOMINGUES COUTINHO NETO

Centro de Ciências Naturais e Humanas / Universidade Federal do ABC - UFABC - Santo André

Araraquara, 06 de agosto de 2019

Agradecimentos

Ao Prof. Dr. Gustavo Troiano Feliciano, por me orientar e apoiar por esses dois anos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa durante o período de realização deste mestrado.

Resumo

A capacidade de reconhecimento molecular da enzima peroxidase (HRP) tem sido muito estudada para o desenvolvimento de biossensores eletroquímicos através de sua ligação com receptores formados pela lectina (ArtinM) complexada com uma molécula de manotriose. Os biossensores eletroquímicos podem ser usados para estimar o número de sítios de ligação ocupados na superfície de um detector para medir uma constante de afinidade relacionada a energia livre de ligação entre a HRP e ArtinM. A incorporação de moléculas de água na estrutura durante o reconhecimento molecular entre as duas proteínas poderia influenciar o valor da constante de afinidade medida pelo biossensor, sendo um caso de estudo que pode ser esclarecido com uso de dinâmica molecular clássica em conjunto com o método de cálculo de energia livre de ligação MM-PBSA. Foram realizadas dinâmicas moleculares com possíveis conformações do complexo formado por HRP e ArtinM gerados por modelos de docking de corpo rígido. As simulações foram realizadas em diferentes condições iniciais onde os complexos se encontravam livres em solução ou imobilizados em um modelo de superfície representando, de forma simplificada, o ambiente existente no biossensor. Os cálculos para energia livre de ligação obtidos através das simulações de dinâmica molecular confirmam o valor da constante de afinidade medida experimentalmente pelos biossensores eletroquímicos e demonstram que a influência de possíveis moléculas de água estruturais é muito pequena em comparação com a massa total das proteínas do complexo. As análises das trajetórias das simulações demonstram que o reconhecimento molecular é quase inteiramente mediado pela manotriose, com contribuições pequenas de outros resíduos do sítio de reconhecimento.

Palavras-chaves: Biomoléculas. Biossensores. Dinâmica molecular. Energia livre. MM-PBSA.

Abstract

The molecular recognition capacity of the enzyme peroxidase (HRP) has been extensively studied for the electrochemical biosensors by its binding to receptors formed by lectin (ArtinM) complexed with a mannotriose molecule. Electrochemical biosensors can be used to estimate the number of binding sites occupied on the surface of a detector to measure an affinity constant related to free binding energy between HRP and ArtinM. The incorporation of structural water molecules during molecular recognition between the two proteins could influence the value of the affinity constant measured by the biosensor, being a case study that can be clarified by using classical molecular dynamics in conjunction with the method of MM-PBSA binding free energy calculation. Molecular dynamics were performed with possible conformations of the complex formed by HRP and ArtinM generated by rigid body docking models. The simulations were performed under different initial conditions where the complexes were free in solution or immobilized in a surface model representing, in a simplified way, the existing environment in the biosensor. The calculations for binding free energy obtained through molecular dynamics simulations confirm the value of the affinity constant experimentally measured by electrochemical biosensors and demonstrate that the influence of possible structural water molecules is very small compared to the total protein mass of complex. Analysis of the simulation trajectories show that molecular recognition is almost entirely mediated by the mannotriose, with small contributions from other recognition site residues.

Key-words: Biomolecules. Biosensors. Molecular dynamics. Free energy. MM-PBSA.

Lista de ilustrações

Figura 1	– Campo autoconsistente de Hartree-Fock.	21
Figura 2	– Metodologia.	31
Figura 3	– Estruturas em PDB das proteínas lectina (JEYAPRAKASH et al., 2004) e peroxidase (HENRIKSEN et al., 1998) visualizadas no VMD. Como podemos notar, a molécula de lectina possui uma mantriose ligada de forma não covalente. O Grupo heme da peroxidase é ligado de forma covalente a um resíduo de histidina no interior da enzima.	32
Figura 4	– Predições estruturais dos complexos lectina-peroxidase sobrepostos visualizados no VMD. As 10 melhores predições indicam que a mantriose tem alta seletividade pela região do sítio ativo da peroxidase.	33
Figura 5	– Potencial eletrostático superficial do complexo lectina-peroxidase. A distribuição não uniforme de cargas poderia ter alguma influência no reconhecimento molecular.	34
Figura 6	– Resíduo 105 da lectina ligado covalentemente ao ácido mercaptoundecanóico. A estrutura se encaixa no modelo de monocamada construído para a simulação.	35
Figura 7	– Modelo da monocamada: A : Bastão B : Raio de van der Waals. A falha no modelo de monocamada é o local onde se encaixa o resíduo da lisina 105 da lectina ligado covalentemente ao ácido da monocamada.	36
Figura 8	– Complexo lectina-peroxidase imobilizado. A lectina é imobilizada através de ligação covalente entre o resíduo de lisina 105 com um ácido localizado no centro da monocamada.	37
Figura 9	– Molécula de mantriose vista no VMD.	38
Figura 10	– Ácido 11-mercaptoundecanóico desprotonado. No campo de força gromos53a6 os grupos CH_2 são descritos como um único átomo.	38
Figura 11	– Grupo lateral do resíduo de lisina 105 da lectina ligada covalentemente ao ácido mercaptoundecanóico. O carbono beta é saturado com hidrogênios para a realização o cálculo quântico. A topologia da ligação covalente é descrita usando os mesmos parâmetros que o campo de força usa nos planos peptídicos de proteínas, onde há compartilhamento de dupla ligação entre o nitrogênio da lisina o oxigênio do ácido carboxílico.	39
Figura 12	– Topologia do ácido mercaptoundecanóico desprotonado criada para o campo de força. As constantes que descrevem as interações entre os átomos foram retiradas do próprio campo de força gromos53a6, usando grupos químicos semelhantes, como ácido carboxílico, tiol e amida.	39
Figura 13	– Cargas da mantriose calculadas pelo Orca. As cargas de Mulliken foram calculadas em conjunto com a base def2-SVP.	40

Figura 14 – Moléculas de água explícitas colocadas nas proximidades da região do grupo ácido carboxílico exposto ao solvente.	41
Figura 15 – Cargas do ácido mercaptoundecanóico calculadas pelo Orca.	42
Figura 16 – Confórmeros 1C_4 e 4C_1 da molécula de manose. Em solução aquosa, temperatura de 298 K e pressão de 1 bar, aproximadamente 100% dos confórmeros são do tipo 4C_1	43
Figura 17 – Ângulos diédricos das conformações da manose. Na maior parte do tempo de dinâmica molecular a conformação 4C_1 se mantém. Existem poucos instantes onde a conformação pode mudar para 1C_4 . A baixa probabilidade de haver a conformação 1C_4 é uma propriedade bem definida experimentalmente, demonstrando a boa qualidade da parametrização.	44
Figura 18 – Gráfico RMSD das estruturas nos sistemas livres durante a dinâmica.	48
Figura 19 – Termo entrópico da energia livre calculado pela aproximação quase-harmônica em função do tempo. O método falha antes de 60 ns e depois de 300 ns, indicando que a estrutura não é adequada para a análise de componentes principais.	49
Figura 20 – Sobreposição do deslocamento dos átomos do complexo com condição inicial 1. As cores representam diferentes posições da estrutura em função do tempo de dinâmica molecular. O grande deslocamento demonstra que os átomos não se comportam de forma quase-harmônica originando a falha da aproximação do método PCA.	50
Figura 21 – Conformações da molécula de manotriose A : Na condição inicial 1 a extremidade da manotriose é deslocada do sítio de ligação original na alça da lectina. B : Na condição inicial 3 a extremidade da manotriose se mantém no sítio original da lectina. Nas duas situações, o sítio foi identificado com a letra S.	52
Figura 22 – Gráfico RMSD das estruturas nos sistemas imobilizados durante a dinâmica.	55
Figura 23 – Termo entrópico do sistema desprotonado a partir de 2 ns de simulação.	55
Figura 24 – Termo entrópico do sistema desprotonado a partir de 4 ns de simulação.	56
Figura 25 – Conformações da molécula de manotriose A : No sistema totalmente protonado a extremidade da manotriose se mantém no sítio de ligação original na alça da lectina. B : No sistema totalmente desprotonado a extremidade da manotriose desloca do sítio original da lectina. Nas duas situações o sítio foi identificado com a letra S.	58
Figura 26 – Sobreposição dos gráficos de número de modos normais em função das frequências de todas as simulações.	59
Figura 27 – Número de modos normais em função das frequências sistema imobilizado semi-protonado e sistema com condição inicial 3.	60

Figura 28 – Contribuição individual de cada resíduo, íon ou grupo prostético na entalpia de todos os complexos. Os resíduos da lectina possuem uma baixa contribuição na energia livre. Íons cálcio da peroxidase têm uma grande contribuição, mesmo estando distantes da região de reconhecimento. A maior contribuição à energia livre favorável vem da manotriose.	62
Figura 29 – Contribuição individual através decomposição de energia livre. Resíduos de cor vermelha contribuem de forma favorável a ligação e de cor azul de forma desfavorável.	63
Figura 30 – Momento de dipolo da molécula de manotriose do sistema de condição inicial 3. .	64
Figura 31 – Momento de dipolo da molécula de manotriose do sistema semi-protonado. . . .	65
Figura 32 – Momento de dipolo da molécula de manotriose do sistema totalmente protonado.	65
Figura 33 – Número de ligações de hidrogênio em função do tempo. O sistema com condição inicial 2 é capaz de fazer o maior número de ligações de hidrogênio em função do tempo.	66
Figura 34 – Ligação de hidrogênio da manotriose na lectina com o resíduo de serina 73 da peroxidase.	67
Figura 35 – Gráfico do número de moléculas de água próximas em função do tempo.	68
Figura 36 – Área de superfície acessível ao solvente dos complexos.	69
Figura 37 – Script Tcl usado para contagem das moléculas de água.	70
Figura 38 – Número exato de moléculas de água a 2 angström de distância do complexo. . .	70
Figura 39 – Moléculas de água estruturais no complexo simulado no sistema com condição inicial 1 e no sistema totalmente protonado.	71

Lista de tabelas

Tabela 1 – Energia livre experimental.	44
Tabela 2 – Energia livre condição inicial 1.	46
Tabela 3 – Energia livre condição inicial 2.	47
Tabela 4 – Energia livre condição inicial 3.	47
Tabela 5 – Termo entrópico da energia nos complexos livres.	51
Tabela 6 – Energia livre total dos sistemas livres.	51
Tabela 7 – Energia livre do sistema imobilizado totalmente desprotonado.	53
Tabela 8 – Energia livre do sistema imobilizado semi-protonado.	53
Tabela 9 – Energia livre do sistema imobilizado totalmente protonado.	54
Tabela 10 – Termo entrópico da energia nos complexos imobilizados.	57
Tabela 11 – Energia livre total nos complexos imobilizados.	57

Lista de abreviaturas e siglas

ASAS	Área de Superfície Acessível ao Solvente
EDC	1-Etil-3-[3-Dimetilaminopropil]Carbodiimida
kDa	Kilodaltons
MM-PBSA	Molecular mechanics Poisson-Boltzmann surface area
NHS	N-Hidroxisuccinimida
nm	nanômetro
ns	nanosegundos
OSDD	Open source drug discovery consortium
PCA	Análise de componentes principais
PDB	Protein data bank
QCM	Microbalança de cristal de quartzo
RDF	Função de distribuição radial
RMSD	Desvio padrão médio quadrático
VMD	Visual Molecular Dynamics

Lista de símbolos

α	Letra grega minúscula Alfa
β	Letra grega minúscula Beta
γ	Letra grega minúscula GAMA
Δ	Letra grega maiúscula Delta
ν	Letra grega minúscula Ípsilon
λ	Letra grega minúscula Lambda
μ	Letra grega minúscula Mi
ρ	Letra grega minúscula Rho
ϕ	Letra grega minúscula Phi
χ	Letra grega minúscula Chi
Ψ	Letra grega maiúscula Psi
∇	Símbolo grego Nabla

Sumário

1	Introdução	14
1.1	Campo de força	16
1.2	Temperatura e pressão	17
1.3	Equações do movimento	17
1.4	Soma de Ewald e condições de contorno periódicas	18
1.5	Método variacional	18
1.6	Método de Hartree-Fock	19
1.6.1	Análise populacional de Mulliken	21
1.7	A energia livre de Gibbs	23
1.8	Cálculo da energia livre	24
1.8.1	Entropia quase-harmônica	26
1.8.2	Entropia de modos normais	27
1.9	Momento de dipolo	28
1.10	Docking de corpo rígido proteína-proteína	28
2	Objetivos	30
3	Metodologia	31
3.1	Programas de dinâmica molecular	31
3.2	Preparação das estruturas moleculares	32
3.3	Construção do modelo teórico em formato PDB	35
3.4	Preparação da topologia das moléculas	37
3.5	Cálculo das cargas dos átomos	40
3.6	Validação da parametrização	43
3.7	Valor experimental da energia livre	44
4	Resultados	45
4.1	Execução da dinâmica molecular	45
4.2	Complexos livres	45
4.2.1	Componente entálpica	46
4.2.2	Entropia quase-harmônica nos complexos livres	48
4.2.3	Entropia de modos normais nos complexos livres	51
4.2.4	Energia livre dos sistemas livres	51
4.3	Complexos imobilizados	52
4.3.1	Entropia quase-harmônica nos complexos imobilizados	54
4.3.2	Entropia de modos normais nos complexos imobilizados	56
4.3.3	Energia livre dos complexos imobilizados	57
4.4	Comparação com o valor experimental da energia livre	58

4.4.1	Número de modos normais	59
4.5	Contribuição individual de cada resíduo	61
4.6	Momento de dipolo	63
4.7	Tipos de interações intermoleculares	66
4.8	Número de moléculas de água ao redor do complexo	68
4.9	Moléculas de água estruturais	71
5	Conclusão	72
	 Referências	 74

1 Introdução

Uma das propriedades mais importantes das moléculas biológicas é a sua capacidade de reconhecimento molecular, que é intrinsecamente relacionada a sua estrutura e função. Moléculas biológicas, como as proteínas, podem enovelar formando estruturas tridimensionais que conseguem reconhecer alvos com alta seletividade e especificidade através de ligações não covalentes (NELSON; COX, 2018; CHATTERJI, 2016).

As proteínas como a lectina (ArtinM) e a peroxidase (HRP) têm sido muito estudadas na biotecnologia por sua capacidade de reconhecimento molecular de polissacarídeos relacionados a importantes funções biológicas. Elas são conhecidas por exercerem atividades antitumorais devido à maior facilidade de se ligarem às glicoproteínas presentes em membranas celulares de células cancerígenas (COELHO et al., 2017).

A ArtinM é uma proteína extraída da planta *Artocarpus heterophyllus*, com 16 kDa de massa, normalmente encontrada na forma tetramérica, com grande capacidade de ligar-se de forma reversível a carboidratos. Por isso, utilizada como marcadores naturais de patologias de grande interesse clínico, além de atuar como moduladora e estimuladora do sistema imunológico (CARVALHO et al., 2014).

A HRP é uma enzima com aproximadamente 44 kDa de massa muito estudada por sua capacidade de oxidar vários substratos orgânicos diferentes e conhecida por ter uma alta afinidade por manose, permitindo que seja usada em combinação com outras proteínas glicosiladas em muitas aplicações bioquímicas (HUMER; SPADIUT, 2019).

A capacidade de reconhecimento molecular entre a ArtinM e a HRP mediada por uma terceira molécula de carboidrato (PESQUERO et al., 2010) permite aplicar a potencialidade dessas moléculas, nos mais variados sistemas biológicos, possibilitando o desenvolvimento de estratégias bioquímicas inteiramente novas, como biossensores eletroquímicos baseados em técnicas de microbalança de cristal de quartzo (QCM) e na capacitância eletroquímica (GIMÉNEZ-ROMERO et al., 2013; SANTOS et al., 2014).

Biossensores QCM conseguem usar a proteína ArtinM imobilizada covalentemente em uma monocamada automontada de ácido mercaptoundecanóico. Nessa situação, a ArtinM funciona como um sítio de ligação na superfície. Essa monocamada se localiza sobre um transdutor físico-químico baseado em um cristal piezoelétrico (GIMÉNEZ-ROMERO et al., 2013).

Dessa forma, quando o biossensor é exposto a uma solução contendo moléculas de HRP, o reconhecimento molecular e consequente formação de um complexo HRP-

ArtinM provoca uma mudança na frequência de vibração do cristal piezoelétrico gerando um sinal detectável. O biossensor QCM consegue estimar o número de sítios de ligação ocupados na superfície, e, assim, medir uma constante de afinidade entre as duas proteínas.

Nos biossensores que usam a capacitância eletroquímica, a proteína ArtinM é imobilizada sobre uma monocamada automontada de ácido mercaptoundecanóico, que é utilizada como um capacitor de placas paralelas. Quando esse biossensor é exposto à solução contendo HRP, a formação do complexo HRP-ArtinM provoca uma mudança na capacitância da monocamada que é convertida em um sinal detectável. De maneira análoga ao QCM, o biossensor de capacitância eletroquímica estima o número de sítios ocupados, medindo a constante de afinidade entre as proteínas (SANTOS et al., 2014).

Um grande problema que pode existir no funcionamento desses biossensores eletroquímicos é o efeito que a formação de um complexo HRP-ArtinM tem sobre as moléculas de água do meio, que poderiam ficar aprisionadas entre as duas proteínas ou nas regiões próximas, convertendo-se em águas estruturais. A presença de moléculas de água estruturais poderia afetar a constante de afinidade medida pelos biossensores eletroquímicos.

A constante de afinidade medida pelos biossensores eletroquímicos está diretamente relacionada a energia livre de ligação (COURNIA; ALLEN; SHERMAN, 2017), uma das grandezas termodinâmicas mais críticas de obter com precisão através de cálculos computacionais. O estudo da energia livre nos sistemas biológicos é de fundamental importância, na medida em que desempenha um papel central na determinação das afinidades vinculativas entre as moléculas (KUKOL, 2015).

Nesse contexto, a realização de um estudo de dinâmica molecular clássica pode ajudar a esclarecer muitas das dúvidas existentes sobre o mecanismo de reconhecimento entre a HRP e a ArtinM no sistema existente no biossensor eletroquímico.

A dinâmica molecular é uma área de estudo baseada na simulação do movimento de partículas e moléculas em um sistema de muitos corpos. Na dinâmica molecular os movimentos dos núcleos dos átomos do sistema são calculados por meio da resolução numérica da Segunda lei de Newton após os cálculos das energias de interações clássicas (FRENKEL; SMIT, 1996; KUKOL, 2015).

As simulações de biomoléculas permitem explorar uma fração significativa do espectro conformacional das proteínas HRP e ArtinM, gerando arquivos de trajetórias que podem ser usados para calcular a energia livre de ligação através do método de mecânica molecular acoplada ao método de Poisson-Boltzmann (MM-PBSA) (UL-HAQ; MADURA, 2015, WANG et al., 2018).

A energia livre de ligação calculada pelo método MM-PBSA (CAMPANERA;

POUPLANA, 2010) permite comparar os valores medidos experimentalmente, ajudando a esclarecer os mecanismos envolvidos no sistema do biossensor, além de possibilitar a estimativa do número de moléculas de água estruturais que poderiam interferir no funcionamento do sistema experimental.

1.1 Campo de força

Para realizar a dinâmica molecular do sistema escolhido, devemos optar por um campo de força que melhor descreva as forças de ligação entre os diferentes tipos de átomos e moléculas presentes no sistema.

Um dos campos de força mais usados em simulações moleculares, devido ao seu alto desempenho, são os potenciais gromos53a6, que funcionam com o método de átomos unidos, o que aumenta muito a velocidade da simulação, que é importante em sistemas muito grandes como o sistema das proteínas imobilizadas.

Os parâmetros de interação não ligados do gromos53a6 são otimizados para reproduzir com grande qualidade as propriedades da fase líquida do sistema e os parâmetros torcionais são adequados para fornecer dados experimentais. O potencial pode ser dividido em várias componentes como podemos ver na equação a seguir (Tu et al., 2016):

$$V_{R_{i,j}} = V_{Ligação} + V_{\hat{Angulos}} + V_{Diédrico} + V_{N\grave{a}o-ligado} \quad (1)$$

Aqui a energia de ligação química entre dois átomos é representado pela equação clássica do potencial da mola representado a seguir:

$$V_{Ligação} = \sum_{Ligação} k_r (r - r_0)^2 \quad (2)$$

De maneira análoga, podemos obter a energia potencial angular através da equação a seguir:

$$V_{\hat{Angulos}} = \sum_{\hat{Angulos}} k_\theta (\theta - \theta_0)^2 \quad (3)$$

As energias relativas aos ângulos de torção entre os planos das ligações químicas são calculadas por meio das somas das energias representadas na equação seguinte (Bell et al., 2017):

$$V_{Diédrico} = \sum_{Diédrico} \sum_{i=1}^4 \frac{V_i}{2} (1 + \cos(i\phi - \phi_i)) \quad (4)$$

E as energias das interações entre átomos não ligados são calculadas através das equações para os potenciais de van der Waals e Coulomb:

$$V_{N\grave{a}o-ligado} = \sum_{i>j} \left[C_{i,j} \left(\frac{A_{i,j}}{r_{i,j}^{12}} + \frac{B_{i,j}}{r_{i,j}^6} \right) + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{i,j}} \right] \quad (5)$$

1.2 Temperatura e pressão

A temperatura do sistema é relacionada à energia cinética média das moléculas que se movem no sistema:

$$\langle E_c \rangle = \frac{nk_bT}{2} \quad (6)$$

Onde n é o número de graus de liberdade da molécula, k_b é a constante de Boltzman e T é a temperatura em Kelvin. Para manter as condições de temperatura e pressão próximas dos valores de referência escolhidos para a simulação, é necessário o uso de um banho térmico.

O efeito de banho térmico é simulado usando o acoplamento de Berendsen, fazendo com que todos os átomos e moléculas do sistema analisado sigam uma função de distribuição de Maxwell-Boltzmann. Dessa forma, se consegue uma equação que relaciona a grandeza termodinâmica de interesse $F(t)$ com a grandeza F_0 de referência (BERENDSEN et al., 1984):

$$\frac{\partial}{\partial t} F(t) = \tau^{-1} [F_0 - F(t)] \quad (7)$$

Onde τ é um parâmetro que determina a velocidade do acoplamento. Ao resolvermos essa equação, podemos controlar facilmente a temperatura e a pressão do sistema através de um fator de correção λ_t :

$$\lambda_t = \left[1 + \frac{\delta t}{\tau} \left\{ \frac{F_0}{F_t} - 1 \right\} \right]^{\frac{1}{n}} \quad (8)$$

Onde n vale 2 para o acoplamento da temperatura e 3 para a o acoplamento da pressão.

1.3 Equações do movimento

Uma vez calculado o valor do potencial, basta aplicar a Segunda lei de Newton, que relaciona diretamente a taxa de variação do momento dos átomos das moléculas com a soma das forças derivadas dos campos de força que atuam sobre os átomos.

$$\vec{F}_i = m_i \frac{\partial^2}{\partial t^2} \vec{r}_i = - \frac{\partial}{\partial r_i} V(r_1, \dots, r_n) \quad (9)$$

Como as equações não possuem solução analítica, são necessárias técnicas de integração numérica para resolução da equação diferencial originada da Segunda lei de Newton. Um dos métodos de integração numérica mais utilizados, implementado no software GROMACS, é o método de Verlet (VERLET, 1967):

$$x_{n+1} = 2x_n - x_{n-1} + a_{(t_n, v_n)} h^2 \quad (10)$$

Onde x_n são as posições unidimensionais do átomo e h é o passo de tempo usado na equação entre os intervalos de integração.

1.4 Soma de Ewald e condições de contorno periódicas

Quando realizamos uma simulação de dinâmica molecular estamos limitados a uma caixa com poucas centenas de milhares de átomos. Esse fato leva a problemas quando trabalhamos com interações de longo alcance, como o potencial de Coulomb, além da dificuldade de tratar a região de fronteira da caixa.

Realizar o corte desse potencial a uma determinada distância geraria problemas que comprometem muito a qualidade dos resultados obtidos. O problema das condições periódicas de contorno geralmente é tratado usando métodos baseados em somas de Ewald, que, basicamente, consiste em transformar um potencial de longa distância em dois potenciais, um de curta distância e outro de longa distância.

$$E = \frac{1}{2} \sum_l \sum_{i,j}^N \frac{B_{i,j}}{[r_{i,j} + l]^n} = \frac{1}{2} \sum_l \sum_{i,j}^N \frac{B_{i,j} \phi(r_{i,j})}{[r_{i,j} + l]^n} + \frac{1}{2} \sum_l \sum_{i,j}^N \frac{B_{i,j} (1 - \phi(r_{i,j}))}{[r_{i,j} + l]^n} \quad (11)$$

O primeiro termo à direita se refere ao potencial de Coulomb dentro do raio de corte adequado para a caixa de simulação. O segundo termo se refere ao termo de longa distância, com todas as imagens periódicas da caixa de simulação, que é otimizada para ser resolvida através de transformadas de Fourier.

Esse método faz com que todos os átomos da caixa façam interações de longa distância com suas imagens periódicas e, portanto, simula um sistema infinito com baixo custo computacional, resolvendo o problema das interações de longo alcance (WELLS; CHAFFEE, 2015).

1.5 Método variacional

Analogamente à mecânica clássica, onde, para entender o movimento dos corpos, devemos resolver a equação da Segunda lei de Newton, na mecânica quântica, para entender os sistemas físicos em escala subatômica, devemos resolver a equação de Schrödinger:

$$H\psi = E\psi \quad (12)$$

Ao resolver a equação devemos encontrar todas as funções de onda que são soluções dessa equação e todos os autovalores relacionados a essas funções. Infelizmente, quando se trata de átomos e moléculas com mais de um elétron, essa equação não possui solução analítica.

O método variacional é uma ferramenta matemática aplicada na mecânica quântica como uma forma de resolver a equação de Schrödinger para átomos ou moléculas com muitos elétrons.

A ideia básica por trás desse método é o fato de que qualquer função de onda tentativa χ que usarmos para representar um sistema físico irá retornar uma energia maior ou igual à energia do estado fundamental desse sistema. A única forma da função de onda retornar uma energia exata E_0 é se a função tentativa for exatamente a função de onda da molécula ψ_0 .

$$\langle \chi^* | H | \chi \rangle \geq \langle \psi_0^* | H | \psi_0 \rangle = E_0 \quad (13)$$

Isso significa que ao variarmos os parâmetros da função tentativa χ , de forma que a energia seja a menor possível, nos aproximamos da função de onda verdadeira. Encontrar uma solução aproximada para a função de onda é a base dos métodos variacionais na mecânica quântica (ALCÁCER, 2007, p. 175).

1.6 Método de Hartree-Fock

O método de Hartree Fock é um método aproximado, baseado no método variacional, para se obter a função de onda e energia resultantes da mecânica quântica, aplicada ao problema de moléculas com M prótons e N elétrons, que, usando a aproximação Born-Oppenheimer, é descrita pelo operador Hamiltoniano a seguir:

$$H = \sum_{i=1}^N \left[\hat{T}_{(c)} + V_{(r_i)} \right] + \frac{1}{2} \sum_{i,j=1}^N V_{(r_i, r_j)} \quad (14)$$

Onde $\hat{T}_{(c)}$ é o termo da energia cinética dos elétrons, $V_{(r_i)}$ o termo de interação elétron-núcleo e $V_{(r_i, r_j)}$ o termo de interação elétron-elétron. Se analisarmos o operador Hamiltoniano, notamos que ele pode ser dividido em dois tipos: um dependente de um elétron e, outro tipo, dependente de dois elétrons.

Para expressões de energia simétricas, podemos aplicar o teorema variacional na função de onda tentativa, de forma que a energia seja minimizada.

$$E_{ele} = \langle \Psi | \hat{H}_{ele} | \Psi \rangle \quad (15)$$

Para o caso de uma molécula com N elétrons e M núcleos, e levando em conta que o somatório deve se estender para todos os orbitais-spin ocupados (osoc), temos a seguinte expressão para o operador de Fock.

$$F = h + \sum_{i=1}^{osoc} (J_a - K_i) \quad (16)$$

Onde o termo h do operador representa a energia cinética e o potencial atrativo entre o elétron e o núcleo do átomo, exatamente como nos átomos hidrogenóides.

O componente J_a do operador representa a repulsão coulombiana entre os orbitais eletrônicos:

$$J_a \psi_2 = \int \frac{\psi_1^* \psi_1 d\tau \psi_2}{r_{12}} \quad (17)$$

O termo K_i representa o operador de troca, que não possui análogos clássicos, podendo ser interpretado como a correlação entre os movimentos dos elétrons nos orbitais com spins paralelos que baixam a energia ao se repelirem (ALCÁCER, 2007, p. 196):

$$K_i \psi_1 = \int \frac{\psi_2^* \psi_1 d\tau \psi_1}{r_{12}} \quad (18)$$

A ideia do método de Hartree-Fock é resolver a Equação 16, onde F é o operador de Fock e ψ_a são os orbitais, de forma parecida com a equação de Schrödinger:

$$F\psi_a = \epsilon_i \psi_a \quad (19)$$

Onde o orbital ψ_a é uma combinação linear de orbitais atômicos χ_a :

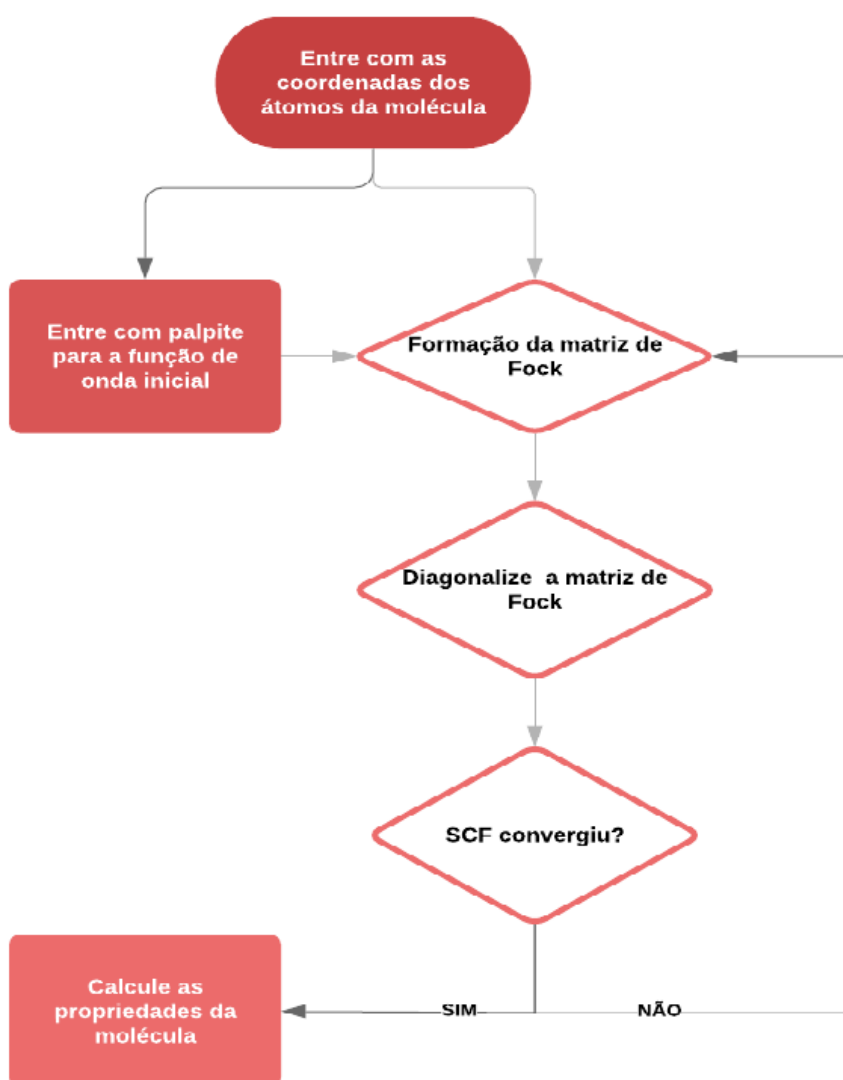
$$\psi_a = \sum_p^L C_p^a \chi_a \quad (20)$$

Resultando no sistema de equações seculares que têm soluções, não todas nulas, conhecidas como equações de Roothaan:

$$\sum_p^L (F - \epsilon_a S_p^q) C_p^a = 0 \quad (21)$$

Para encontrar as energias, basta resolver numericamente as equações de polinômios de grau L originadas da equação, através do método autoconsistente de Hartree-Fock:

$$\text{Det} \{ F_p^q - \epsilon_a S_p^q \} = 0 \quad (22)$$

Figura 1 – Campo autoconsistente de Hartree-Fock.

Fonte: Autor, 2019.

1.6.1 Análise populacional de Mulliken

A única forma de descobrirmos quais são as prováveis cargas dos átomos das moléculas novas usadas na dinâmica molecular é resolvendo a equação de Schrödinger para determinar qual a função de onda da molécula associada à densidade eletrônica.

$$\rho(r) = \sum_i^{oc} n_i |\psi_i(r)|^2 \quad (23)$$

Ao integrarmos a densidade eletrônica em todo espaço, obtemos o número total

de elétrons.

$$N = \int \rho_r dr \quad (24)$$

Usando o fato de a função de onda ser uma combinação linear de uma base de funções normalizadas ϕ_p , a seguinte igualdade é válida:

$$\rho(r) = \sum_i^{oc} n_i \psi_i(r) = \sum_p \sum_q D_p^q \phi_p^*(r) \phi_q(r) \quad (25)$$

Onde D_p^q é a matriz dos elementos de densidade dada pela seguinte expressão:

$$D_p^q = \sum_i^{oc} n C_p^i C_q^{i*} \quad (26)$$

Dessa forma, podemos relacionar os elementos de densidade com o traço do produto das matrizes D_p^q e S_p^q :

$$N = \int \rho_r dr = \sum_p \sum_q D_p^q S_p^q = tr(DS) = N \quad (27)$$

Onde a integral de recobrimento S_p^q é dada pela seguinte expressão:

$$S_p^q = \int \phi_p^*(r) \phi_q(r) dr \quad (28)$$

Ao conhecermos a função densidade eletrônica da molécula, é possível construir um mapa de população eletrônica e, conseqüentemente, calcular a carga dos átomos da molécula. Uma forma de calcular as cargas é usando a análise populacional da matriz de Mulliken (MULLIKEN, 1955):

$$M_p^q = (D_p^q S_p^q) \quad (29)$$

Em que M é a matriz onde os termos são produtos dos elementos correspondentes das matrizes S e D, sendo usada para selecionar os elétrons através de suas contribuições. Um elemento da diagonal dessa matriz M_p^p é o número de elétrons no orbital p. Os elementos M_p^q representam metade dos elétrons compartilhados pelos orbitais p e q. Ao somarmos as contribuições dos orbitais atômicos de um átomo A, obtemos a população N_A de elétrons associados a esse átomo.

$$N_A = \sum_{p \in A} \sum_q^{OA} D_p^q S_p^q \quad (30)$$

A análise populacional de Mulliken é um dos métodos mais tradicionais usados pelos químicos, mas importante salientar que esse método é muito questionado, devido a sua grande dependência das funções de base usadas, além de dividir igualmente a população eletrônica entre os átomos sem levar em conta a sua eletronegatividade (ALCÁCER, 2007, p. 217).

1.7 A energia livre de Gibbs

Para serem estudados de forma adequada, os sistemas termodinâmicos necessitam de uma descrição baseada em funções variáveis de estado, como a temperatura, pressão e volume, chamadas de potenciais termodinâmicos. A energia livre de Gibbs é um potencial termodinâmico (SALINAS, 1997, p. 72) e pode ser interpretada como a energia disponível para realização de trabalho mecânico, uma função da energia que descreve um problema de mínimo local na situação de equilíbrio termodinâmico.

$$G = U + PV - TS \quad (31)$$

A energia livre de Gibbs pode ser interpretada como sendo a soma da energia interna do sistema U com a energia necessária para criar uma cavidade do meio onde o sistema será introduzido PV menos o calor que o ambiente (Reservatório térmico) fornece TS . A variação da energia livre é uma grandeza termodinâmica relacionada à espontaneidade de um processo, e pode ser descrita pela seguinte equação (Sears; Salinger, 1979):

$$\Delta G = \Delta U + P\Delta V + \Delta PV - T\Delta S - \Delta TS \quad (32)$$

As reações dos processos biológicos geralmente acontecem em condições onde a temperatura e pressão são constantes e, portanto, ΔG se reduz à seguinte expressão:

$$\Delta G = U + P\Delta V - T\Delta S \quad (33)$$

Onde o termo $\Delta U + P\Delta V$ é a entalpia do sistema e a variação de energia livre pode ser descrita da seguinte forma (Nelson; Cox, 2014, p. 23):

$$\Delta G = \Delta H - T\Delta S \quad (34)$$

No caso da ligação entre biomoléculas, os valores de $\Delta G < 0$ indicam que essa ligação é espontânea, e, para valores maiores que zero, a ligação não é espontânea.

1.8 Cálculo da energia livre

Como as proteínas não são estáticas, os dados de evolução temporal dos sistemas moleculares são uma ferramenta preciosa no cálculo das energias livres, permitindo que se explore muitas conformações diferentes que o sistema possa assumir.

Essa função é feita pelo pacote `g_mmpbsa`, desenvolvido pelo *Open Source Drug Discovery Consortium* (OSDD), que faz uma interface entre GROMACS e o APBS, permitindo calcular as energias livres de ligação para cada frame da simulação e analisar a evolução temporal das energias livres e da energia livre média.

Outra função importante do pacote `g_mmpbsa` (KUMARI; KUMAR; LYNN, 2014) é permitir o cálculo das contribuições individuais de cada resíduo, ou ligante, na energia livre de ligação total do sistema. Isso é feito através do método de decomposição de energia, permitindo a fácil análise dos resultados por meio das ferramentas mais comuns usadas na visualização de moléculas. As abordagens para cálculo de energia livre podem ser divididas em dois tipos, as de caminho e as de ponto final. As medidas baseadas em caminho são computacionalmente muito caras e requerem cálculos infinitesimais da progressão do sistema entre o estado inicial no vácuo e o estado final no solvente.

A segunda forma de cálculo da energia livre é bem mais simples e usa o fato da energia livre ser uma função de estado. Dessa forma, para calcularmos sua variação, precisamos apenas calcular o estado inicial e estado final, sem nos preocupar com o caminho entre os dois estados ao levar a molécula do vácuo até o solvente, e, sendo assim, muito mais rápida computacionalmente (UL-HAQ; MADURA, 2015).

O cálculo da energia livre de um complexo usando método direto é bem simples e resume-se a subtrair da energia livre do complexo lectina-peroxidase ΔG_{LP} as energias livres da lectina ΔG_L e da peroxidase ΔG_P , calculadas separadamente como na equação (KOLLMAN et al., 2000) :

$$\Delta G_{Lig} = \Delta G_{LP} - \Delta G_L - \Delta G_P \quad (35)$$

Onde cada termo ΔG_i é a soma das energias mecânicas E_{mm} do campo de força usado da dinâmica molecular com a energia livre de solvatação ΔG_{solv} menos a contribuição entrópica $T\Delta S$:

$$\Delta G_i = E_{mm} + \Delta G_{solv} - T\Delta S \quad (36)$$

Quando combinadas, as equações resultam na equação:

$$\Delta G_{lig} = E_{mm} + (\Delta G_{solv,LP} - \Delta G_{Solv,L} - \Delta G_{Solv,P}) - TS \quad (37)$$

As energias livres de solvatação podem ser divididas em uma soma da energia de solvatação polar e solvatação apolar.

$$\Delta G_{Solv} = \Delta G_{solv,Polar} + \Delta G_{Solv,Apolar} \quad (38)$$

A energia de solvatação apolar é considerada proporcional à área de superfície das moléculas acessível ao solvente (*ASAS*):

$$\Delta G_{Solv,Apolar} \propto \gamma ASAS \quad (39)$$

Já a energia de solvatação polar é calculada através da resolução da equação de Poisson-Boltzmann Linearizada. Com ela, é possível obter o potencial eletrostático entre diferentes grupos químicos quando se conhece o potencial associado a determinadas distribuições de íons em volta da molécula. A equação de Poisson relaciona a variação do potencial ϕ_r com a função distribuição de densidade de carga ρ_r no meio com constante dielétrica ϵ_r (SOARES; FERREIRA, 2004):

$$\nabla \cdot \epsilon_r \nabla = -4\pi \rho_r \quad (40)$$

A constante dielétrica pode ter dois valores devido à diferença de polarizabilidade entre a água e o interior das proteínas. No sistema biológico, as cargas da biomolécula podem ser tratadas como sendo fixas, enquanto as cargas dos íons dissolvidos na solução podem ser descritas como sendo móveis e com densidade seguindo a distribuição de Boltzmann:

$$\frac{n_j(r)}{n_j^0(r)} = e^{\frac{-q_i \phi_r}{k_b T}} \quad (41)$$

Onde $n_j(r)$ é a densidade volumétrica de íons do tipo j em uma região do espaço; $n_j^0(r)$ é a densidade volumétrica de íons do tipo j em todo o espaço; q_i é a carga do íon; ϕ é o potencial eletrostático da região. Dessa forma, podemos modificar a Equação 25 obtendo a equação de Poisson-Boltzmann:

$$\nabla \cdot \epsilon_r \nabla = -4\pi \left[\rho_r + \lambda_r \sum_i q_i n_i e^{\frac{-q_i \phi}{k_b T}} \right] \quad (42)$$

Onde λ_r é igual a 0, nas regiões inacessíveis ao solvente, e igual a 1, nas regiões acessíveis. Podemos aplicar a teoria de Debye-Hückel à equação de Poisson-Boltzmann em sistemas formados por um conjunto de cargas em meio com baixa força iônica e constante dielétrica uniforme. Dessa forma, o termo exponencial da distribuição de cargas passa a ser tratado linearmente (SOARES; FERREIRA, 2004):

$$\nabla \cdot \epsilon_r \nabla = -4\pi \rho_r' + \lambda_r k^{-2} \varphi_r \quad (43)$$

Onde k é a constante de Debye-Hückel e representa a blindagem provocada pelos íons no meio. A equação é resolvida numericamente através da geração de malhas, decompondo o domínio físico em elementos menores.

Como o cálculo da energia livre do MM-PBSA utiliza uma média, o software `g_mmpbsa` usa a fórmula estatística para estimar o erro do padrão σ do método (KUMARI; KUMAR; LYNN, 2014), como podemos ver na seguinte equação:

$$\sigma = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta G - \langle \Delta G \rangle)^2} \quad (44)$$

Onde N é o número de *frames* usados no cálculo. O valor do erro padrão da energia livre representa o quanto a média da amostra medida aproxima-se da média população de energias calculadas (ALTMAN; BLAND, 2005).

1.8.1 Entropia quase-harmônica

Os arquivos de trajetória gerados na dinâmica molecular são convertidos em uma matriz, onde as coordenadas são sobrepostas a partir de um conjunto de coordenadas de referência. Assim, obtemos um vetor de deslocamento de cada resíduo ΔR_t em cada ponto de tempo t . O conjunto de coordenadas M é agrupado em uma matriz de flutuação $3N \times M$. Onde $3N$ é o número de deslocamentos de cada átomo nas coordenadas x , y e z , para cada um dos M *frames* da simulação (NUMATA; WAN; KNAPP, 2007).

$$\Delta R = \begin{bmatrix} R_1(t_1) & R_1(t_2) & R_1(t_3) & \dots & R_1(t_m) \\ R_2(t_1) & R_2(t_2) & R_2(t_3) & \dots & R_2(t_m) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ R_n(t_1) & R_n(t_2) & R_n(t_3) & \dots & R_n(t_m) \end{bmatrix} \quad (45)$$

A partir da matriz de flutuação, obtemos uma matriz de variância-covariância C (AMADEI; LINSSEN; BERENDSEN, 1993):

$$C = \Delta R_t \Delta R_t^t \quad (46)$$

A matriz C carrega as informações sobre todas as regiões atingidas durante a simulação. O PCA utiliza uma transformação ortogonal em C convertendo o conjunto de deslocamentos dos átomos durante a dinâmica molecular em um novo conjunto linearmente independente de componentes principais.

Com as componentes principais podemos obter autovalores λ_{ij} relacionados às frequências de vibração através do teorema da equipartição de energia (YANG et al.,

2009; SEEBER et al., 2007):

$$\alpha_i = \frac{\hbar\omega_i}{k_b T} = \frac{\hbar}{\sqrt{k_b T}} \frac{1}{\lambda_{ij}} \quad (47)$$

Com as frequências calculadas, podemos obter a entropia vibracional através da seguinte equação:

$$S = k_b \sum_{i=1}^{3N-6} \left\{ \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \right\} \quad (48)$$

1.8.2 Entropia de modos normais

Outra forma de obtenção da entropia vibracional das moléculas é usando o cálculo dos modos normais de vibração. Os modos normais são as flutuações ao redor do estado fundamental de energia das moléculas. Esses movimentos são aproximados pela Lei de Hooke, e todas as forças que atuam sobre o átomo nessa direção podem ser expressas como a derivada segunda da energia potencial (SKJAERVEN; MARTINEZ; REUTER, 2010).

Dessa forma, podemos sempre encontrar 3N-6 modos normais de vibração oscilando em frequências normais de vibração e linearmente independentes. O valor 6 subtraído refere-se aos 3 modos rotacionais e 3 modos translacionais da molécula.

Essas equações podem ser expressas na forma de uma matriz Hessiana no caso de muitos átomos, como uma proteína (CARRINGTON; MANCERA, 2004).

$$H(V) = \begin{bmatrix} \frac{d^2V}{dx_1 dx_1} & \frac{d^2V}{dx_1 dx_2} & \cdots & \frac{d^2V}{dx_1 dx_n} \\ \frac{d^2V}{dx_2 dx_1} & \frac{d^2V}{dx_2 dx_2} & \cdots & \frac{d^2V}{dx_2 dx_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2V}{dx_n dx_1} & \frac{d^2V}{dx_n dx_2} & \cdots & \frac{d^2V}{dx_n dx_n} \end{bmatrix} \quad (49)$$

No caso dos modos normais, as forças se alinham com o deslocamento, e podemos escrever a expressão para o sistema da seguinte forma:

$$Hx = \lambda x \quad (50)$$

Onde λ é um escalar e x é um autovetor que pode ser encontrado ao combinarmos os termos da seguinte forma usando a matriz identidade I.

$$[H - \lambda I] x = 0 \quad (51)$$

Dessa forma, podemos encontrar os autovalores calculando determinante da matriz e obtemos as 3N-6 soluções.

$$\text{Det} [H - \lambda I] = 0 \quad (52)$$

Cada solução terá um autovalor λ relacionado ao seu autovetor x , que nada mais é que o deslocamento de um átomo em uma determinada direção. Com os autovalores, podemos usar a Equação 48 para calcular a entropia vibracional (CUI; BAHAR, 2005).

1.9 Momento de dipolo

Um dipolo elétrico é caracterizado pela presença de duas cargas de mesmo módulo e sinais opostos, separadas por uma distância d . O vetor momento de dipolo \vec{P} é definido pela equação (NUSSENZVEIG, 1997, p. 51):

$$\vec{P} = q\vec{d} \quad (53)$$

Onde o vetor \vec{d} indica a direção da carga negativa para a positiva. No caso de uma distribuição discreta de várias cargas pontuais, o momento de dipolo nada mais é que a soma vetorial:

$$\vec{P} = \sum_i q_i \vec{d}_i \quad (54)$$

O momento de dipolo é um conceito útil na representação clássica de dinâmica molecular para descrever as interações entre átomos e moléculas.

1.10 Docking de corpo rígido proteína-proteína

Os algoritmos de *docking* proteína-proteína funcionam usando uma abordagem puramente geométrica, aplicando técnicas usadas no reconhecimento de padrões. O algoritmo visa discretizar a estrutura dessas proteínas no formato de uma grade. Dessa forma, procuram identificar a complementaridade da superfície molecular expressando a energia de interação em cada conformação como sendo a soma de P funções de correlação:

$$E(\alpha, \beta, \gamma, \lambda, \mu, \nu) = \sum_{p=1}^P \int \overline{R(n, l, m)_p} \hat{T}(\lambda, \mu, \nu) \hat{D}(\alpha, \beta, \gamma) L_p(x, y, z) dV \quad (55)$$

Onde R_p e L_p são definidos no receptor e ligante, e \hat{T} e \hat{D} são operadores translacionais e rotacionais. Com a introdução, por Katchalski-katzir et al. (1992), das

transformadas rápidas de Fourier tornou-se possível explorar milhões de conformações com baixo custo computacional, onde R_p e L_p são definidos em grades no espaço 3D, assumindo que a função energia pode ser expressa como uma transformada de Fourier (PADHORN et al., 2016):

$$E(\alpha, \beta, \gamma, \lambda, \mu, \nu) = \sum_{p=1}^P \sum_{n,l,m} \overline{R(n, l, m)} L_p(\alpha, \beta, \gamma, n, l, m) e^{\frac{-2\pi i}{N}(n\lambda + l\mu + m\nu)} \quad (56)$$

Dessa forma, o algoritmo fornece uma lista de valores de correlação onde se obtém a correspondência geométrica entre as superfícies das moléculas, encontrando as melhores predições estruturais.

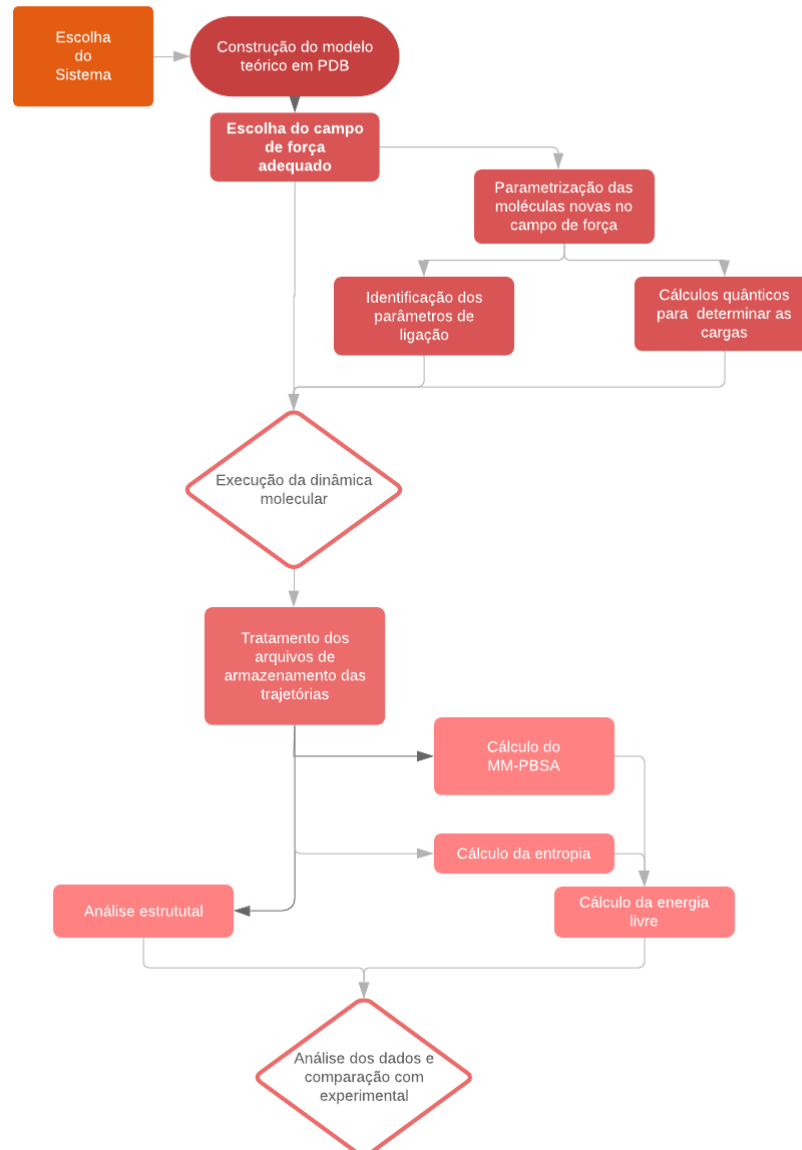
2 Objetivos

O objetivo principal deste projeto consiste na construção de um modelo teórico-computacional, com uso de técnicas de dinâmica molecular, sobre o reconhecimento molecular de lectinas e peroxidases em biossensores eletroquímicos. Dentro deste, encontram-se os objetivos específicos listados a seguir:

- Realizar os cálculos computacionais para energia livre de ligação entre complexos lectina-peroxidase immobilizadas em superfície de ouro coberta por uma monocamada de ácido 11- mercaptoundecanóico.
- Comparar os valores para energias livres de ligação calculadas através das técnicas computacionais e valores obtidos experimentalmente por grupos de pesquisa que utilizam biossensores eletroquímicos.
- Identificar os efeitos da ligação entre lectina e peroxidase nas moléculas de água estruturais das proteínas.
- Verificar se o processo de adsorção na monocamada é um fator importante, tendo contribuição na energia de ligação.
- Analisar a estabilidade da estrutura das proteínas immobilizadas, a ocorrência de possíveis mudanças conformacionais e estruturais nas proteínas individuais.

3 Metodologia

Figura 2 – Metodologia.



Fonte: Autor, 2019.

3.1 Programas de dinâmica molecular

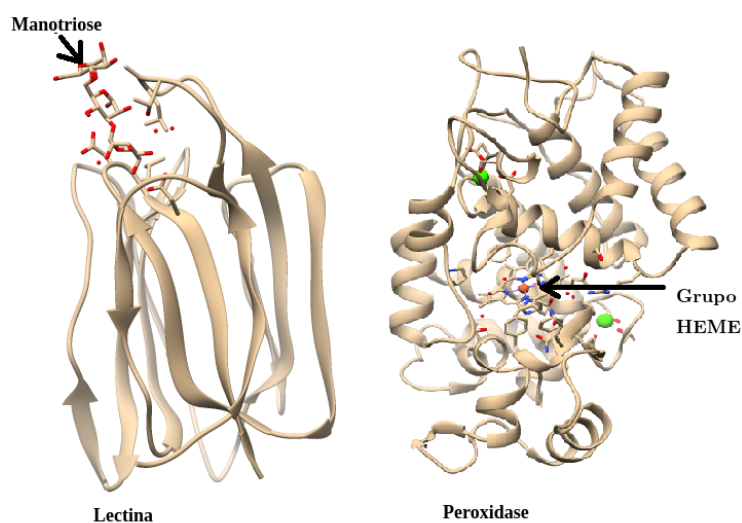
No projeto foram usadas ferramentas computacionais conhecidas, como o pacote de dinâmica molecular GROMACS 4.5.5 (*Groningen machine for chemical simulations*), o pacote de solução numérica da equação de Poisson-Boltzmann (APBS) 1.3, o

software de cálculos quânticos Orca 4.0.1.2 , o software para visualização de dinâmica molecular (VMD) 1.9.3. Além desses, foi utilizado o pacote (g_mmpbsa) da plataforma colaborativa internacional, que visa projetar e descobrir fármacos para doenças tropicais negligenciadas como a malária, leishmaniose e tuberculose. As dinâmicas moleculares foram realizadas em um computador com processador Intel Xeon E5-2670.

3.2 Preparação das estruturas moleculares

Os arquivos das estruturas cristalográficas das proteínas usadas no projeto foram escolhidas no Protein Data Bank (PDB), como não existe uma estrutura cristalográfica contendo um complexo formado por lectina e peroxidase, as estruturas da lectina artinM (JEYAPRAKASH et al., 2004) e *horseradish* peroxidase (HENRIKSEN et al., 1998) foram submetidas à ferramenta de ancoragem molecular de corpo rígido online, de uso livre acadêmico, ZDOCK (PIERCE et al., 2014). As estruturas das proteínas podem ser vistas na Figura 3.

Figura 3 – Estruturas em PDB das proteínas lectina (JEYAPRAKASH et al., 2004) e peroxidase (HENRIKSEN et al., 1998) visualizadas no VMD. Como podemos notar, a molécula de lectina possui uma manotriose ligada de forma não covalente. O Grupo heme da peroxidase é ligado de forma covalente a um resíduo de histidina no interior da enzima.



Fonte: Autor, 2019.

Como podemos ver na Figura 3, a lectina artinM é uma proteína pequena, composta por apenas 148 resíduos de aminoácidos, cuja estrutura secundária é formada principalmente por folhas beta antiparalelas. Na região de reconhecimento há um resíduo de manotriose.

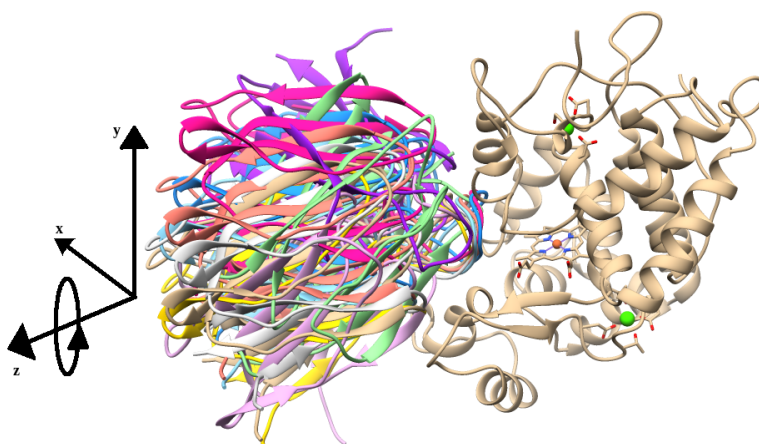
A molécula de peroxidase é uma enzima composta de 306 resíduos de aminoácidos, com estrutura secundária formada principalmente por alfa-hélices. Essa estrutura

ainda contém em seu interior dois íons cálcio e um grupo prostético HEME ligado covalentemente à região interna da enzima, no sítio catalítico.

A ancoragem molecular é necessária devido a limitação de tempo computacional. Seria inviável realizar uma dinâmica molecular por tempo suficiente para que a peroxidase e a lectina se ligassem naturalmente, como ocorre no experimento do biossensor eletroquímico.

O programa ZDOCK retorna 10 modelos possíveis de ligação entre as moléculas, um número estatisticamente não significativo comparado ao grande número de conformações que duas proteínas podem atingir. Os modelos podem ser vistos na Figura 4.

Figura 4 – Predições estruturais dos complexos lectina-peroxidase sobrepostos visualizados no VMD. As 10 melhores predições indicam que a mantriose tem alta seletividade pela região do sítio ativo da peroxidase.



Fonte: Autor, 2019.

Como relatado por Giménez-Romero et al. (2013), sabe-se através de experimentos que as moléculas de lectina têm forte capacidade de ligar-se a monômeros das moléculas de peroxidase, mas perdem essa capacidade quando, devido à elevada concentração, ocorre a formação de dímeros de peroxidase. Esse fato leva à suposição de que a região de ligação entre a lectina e a peroxidase é obstruída quando ocorre a formação de dímeros de peroxidase.

Segundo Ignatenko et al. (2013), as regiões dos sítios ativos das duas moléculas de peroxidase ficam muito próximas umas das outras. Dessa forma, é muito provável que a formação de dímeros de peroxidase gere uma restrição para determinados substratos no sítio ativo. O produto da oxidação de um sítio ativo poderia interferir na entrada de substrato no outro sítio.

Todos os modelos do complexo gerados pelo ZDOCK mostram que a região

preferencial de ligação entre as duas proteínas é a superfície onde se localiza o sítio ativo da peroxidase e a superfície onde se localiza o açúcar da lectina.

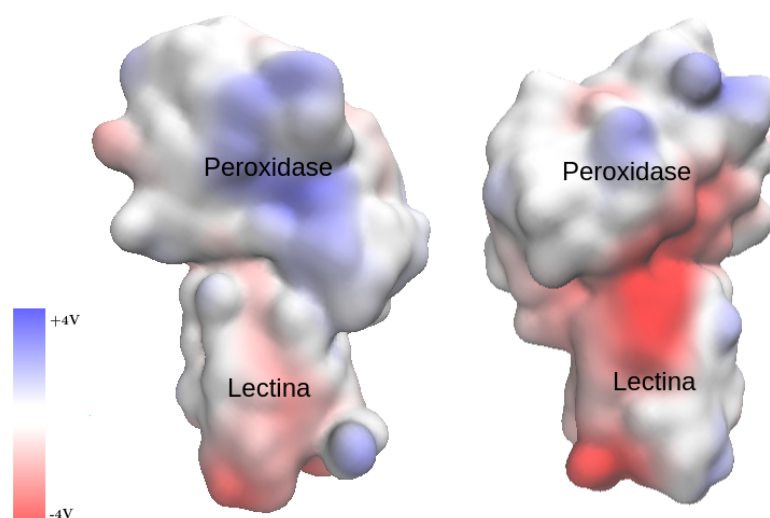
Os modelos gerados diferem apenas na rotação da lectina em torno do eixo de ligação, como podemos ver na Figura 4, onde são mostradas todas as conformações da lectina em torno da região do sítio ativo da peroxidase.

Essas evidências experimentais reforçam a qualidade do modelo gerado pelo *docking* entre as duas proteínas e explicariam o porquê dessa região ficar estericamente impedida na presença de altas concentrações de peroxidase.

Outro fator importante a ser enfatizado é que uma das formas refinar-se o modelo de *docking* é através da realização de dinâmica molecular. Sendo assim, não devemos ver o modelo como um resultado final, mas como uma “condição inicial” para começar uma simulação de dinâmica molecular.

Para entendermos uma das razões da região de reconhecimento molecular entre as proteínas ser favorecida, podemos analisar o potencial eletrostático superficial na Figura 5.

Figura 5 – Potencial eletrostático superficial do complexo lectina-peroxidase. A distribuição não uniforme de cargas poderia ter alguma influência no reconhecimento molecular.



Fonte: Autor, 2019.

O potencial eletrostático da peroxidase parece ter uma maior predominância de cargas positivas (azul) na superfície, enquanto a lectina parece ter uma superfície com densidade de carga mais negativas (vermelho). Esse fato pode ter alguma contribuição na afinidade entre as duas proteínas.

3.3 Construção do modelo teórico em formato PDB

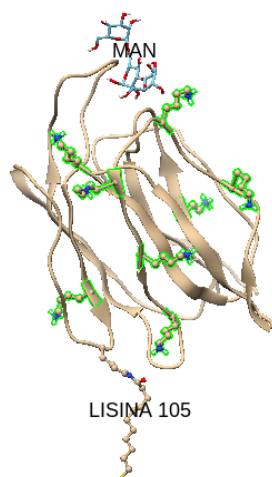
Devido à alta polarizabilidade dos átomos de ouro, esse elemento tem forte potencial de Lennard-Jones, como podemos ver nos parâmetros encontrados em Pu et al. (2007). Sendo assim, qualquer molécula orgânica poderia facilmente adsorver sobre a superfície de ouro do biossensor e não haveria especificidade na detecção. Dessa forma, utiliza-se a técnica de monocamada automontada formada por ácido 11-mercaptopundecanóico para cobrir superfície de ouro.

Embora, no sistema experimental, a ligação covalente entre o grupo tiol do ácido e o ouro ocorra devido à interação ácido-base da teoria de Pearson, no modelo teórico, aplicamos uma restrição ao movimento do átomo de enxofre para representar essa ligação. Essa restrição é feita por um potencial semelhante ao de ligação química, mas restringindo o movimento nas direções x, y e z com uma constante de mola de $10000 \text{ kJmol}^{-1} \text{ nm}^{-2}$.

Como detalhado por Giménez-Romero et al. (2013), a proteína de lectina é imobilizada na superfície do biossensor, formando uma ligação covalente entre um dos ácidos da monocamada e um resíduo da lectina. Essa reação é catalisada por 1-etil-3-[3-dimetilaminopropil]carbodiimida (EDC) e N-Hidroxisuccinimida (NHS), formando um grupo amida.

Existem vários resíduos contendo aminas primárias na proteína de lectina, havendo, assim, muitos pontos onde é possível ocorrer a ancoragem na monocamada. Na análise visual da lectina não se detectam esses resíduos (verde) na região da mantriose, onde ocorre a ligação com a peroxidase, como vemos na Figura 6.

Figura 6 – Resíduo 105 da lectina ligado covalentemente ao ácido mercaptopundecanóico. A estrutura se encaixa no modelo de monocamada construído para a simulação.



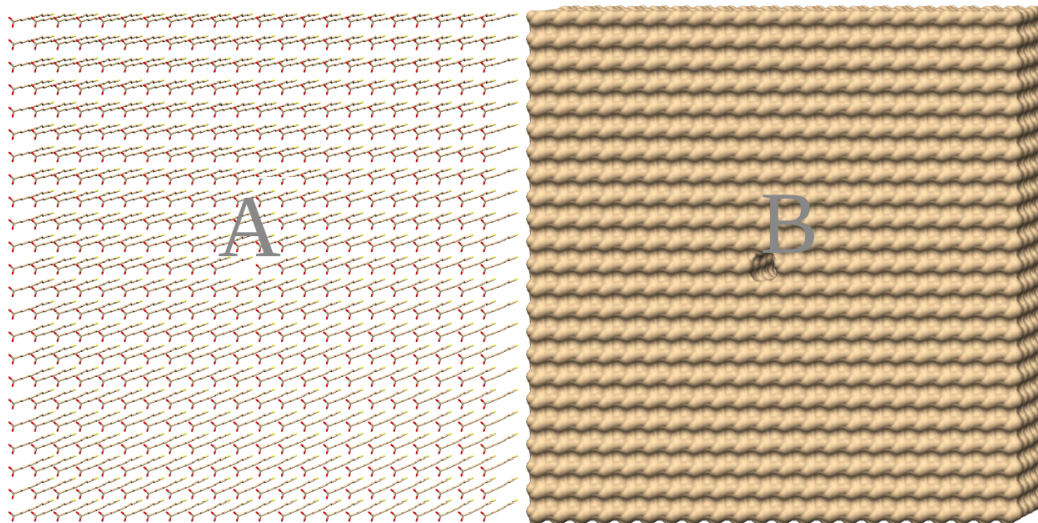
Fonte: Autor, 2019.

A ancoragem de resíduos mais próximos à manotriose poderia inclinar a proteína de lectina deixando a superfície de reconhecimento voltada para a monocamada durante algum tempo, mas não impediria que a ligação ocorresse em algum momento.

Dessa forma, embora não se detecte uma grande diferença com relação às demais possibilidades de ancoragem, optou-se por fazer a imobilização da lectina pelo resíduo de lisina 105. Isto porque, que em análise visual no VMD, ela demonstra ser menos internalizada, além de possuir uma densidade superficial de cargas positivas, tendo, assim, maior probabilidade de ancoragem.

Para decidir a melhor forma de montar a monocamada no modelo teórico, considerou-se o fato de que do raio de Lennard-Jones para os grupos CH_2 das moléculas de ácido possuem cerca de 0.5 nm no campo de força de dinâmica molecular. As moléculas foram colocadas sempre a 0.5 nm umas das outras, em formato de grade. Dessa forma, obteve-se uma monocamada que bloqueia totalmente a superfície de ouro sem problemas de minimização de energia provocados por impedimento estérico, como vemos na Figura 7:

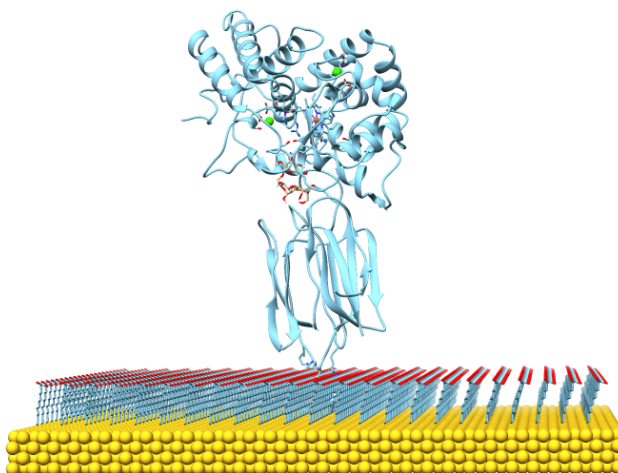
Figura 7 – Modelo da monocamada: **A**: Bastão **B**: Raio de van der Waals. A falha no modelo de monocamada é o local onde se encaixa o resíduo da lisina 105 da lectina ligado covalentemente ao ácido da monocamada.



Fonte: Autor, 2019.

Como podemos ver no modelo, uma falha é deixada entre os ácidos, onde posteriormente encaixa-se o resíduo imobilizado da lectina o mais distante possível das fronteiras da caixa periódica, como podemos ver na Figura 8.

Figura 8 – Complexo lectina-peroxidase imobilizado. A lectina é imobilizada através de ligação covalente entre o resíduo de lisina 105 com um ácido localizado no centro da monocamada.



Fonte: Autor, 2019.

Outra vantagem da escolha do resíduo de lisina 105 é o fato de evitar que o complexo extrapole as condições periódicas de contorno, permitindo o uso de uma caixa menor e reduzindo o tempo e custo computacional.

3.4 Preparação da topologia das moléculas

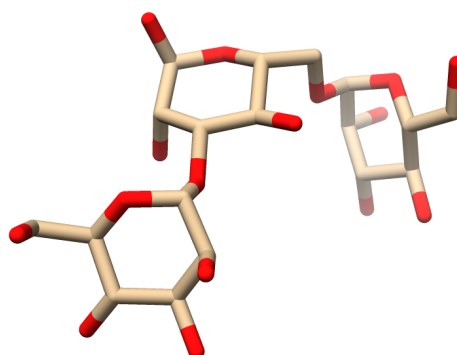
Como todos os softwares de dinâmica molecular, o GROMACS não contém as topologias de todas as moléculas existentes, e, em muitas situações, é necessário que o usuário obtenha essas informações para o programa realizar a simulação.

No complexo estudado, há muitas moléculas que não são pré-catalogadas no banco de dados de topologias do gromos53a6, como uma molécula de manotriose pertencente a proteína de lectina na Figura 9.

Como se trata de uma única molécula de açúcar de estrutura complexa, uma solução cabível é usar o servidor online do PRODRG (VAN AALTEN et al., 1996) para obter a topologia com parâmetros de ligação do próprio gromos53a6. Mas como visto em Lemkul, Allen e Bevan (2010), devemos evitar usar as cargas dos átomos fornecidas por este programa.

Existem muitos servidores *online* que fornecem automaticamente os parâmetros necessários para o campo de força, mas, dependendo da complexidade da molécula, construir a topologia manualmente é uma possibilidade.

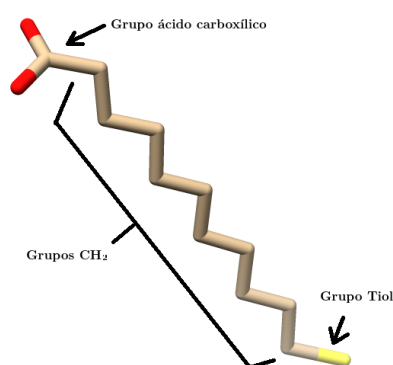
O banco de dados de moléculas do gromos53a6 é descrito de uma forma bem intuitiva, com poucos arquivos onde existem nomes de átomos e definições de

Figura 9 – Molécula de mantriose vista no VMD.

Fonte: Autor, 2019.

constantes dos potenciais usados, sendo simples criar-se topologias diretamente no campo de força, aumentando o banco de dados, evitando a necessidade de modificação de arquivos de topologias manualmente e minimizando a possibilidade de erros.

Outra molécula usada na simulação é a do ácido 11-mercaptoundecanóico, que forma uma monocamada sobre a superfície de ouro do detector do biossensor. Nesse caso, existem centenas de moléculas, mas elas são relativamente simples, como podemos ver na Figura 10.

Figura 10 – Ácido 11-mercaptoundecanóico desprotonado. No campo de força gromos53a6 os grupos CH_2 são descritos como um único átomo.

Fonte: Autor, 2019.

A molécula do ácido contém apenas os grupos químicos do tipo ácido carboxílico e tiol já existentes nos resíduos catalogados nos arquivos de topologias do GROMACS. Um modo de resolver esse problema é criar um resíduo de ácido 11-mercaptoundecanóico no banco de dados usando as informações pré-existent, mas, novamente, há o problema das cargas de cada átomo, que são desconhecidas.

onde podemos ver que os grupos CH_2 são fundidos em um único átomo por razão da eficiência; por último, na terceira coluna, as cargas individuais. No campo *bonds* é definido quais ligações químicas são feitas entre os átomos. Cada ligação é representada pela instrução *gb*, um padrão que especifica a informação sobre a constante de mola da ligação química e a distância de ligação. Analogamente, no campo *angles*, estão as informações sobre cada trio de átomos que formam um ângulo na molécula, bem como as instruções *ga* que especificam a informação sobre a constante de mola e o ângulo formado entre os átomos. Devido à simplicidade das moléculas de ácido, as topologias não contêm ângulos impróprios e o campo *impropers* fica vazio. No campo *dihedrals* estão as informações sobre quais átomos formam um potencial diédrico e, nas instruções *gd*, estão especificados os parâmetros desse potencial.

3.5 Cálculo das cargas dos átomos

Ao utilizarmos o software Orca (NEESE, 2011) devemos entrar com uma função de onda tentativa que, após ser otimizada pelos métodos variacionais, permite a obtenção das cargas individuais dos átomos. O Orca retorna as cargas os átomos da molécula usando o método de Mulliken, como vemos na Figura 13.

Figura 13 – Cargas da manotriose calculadas pelo Orca. As cargas de Mulliken foram calculadas em conjunto com a base def2-SVP.

```

-----
MULLIKEN ATOMIC CHARGES
-----

```

0 O : -0.327821	OH	33 C : 0.270680	
1 H : 0.196841		34 H : 0.004642	
2 C : 0.112036		35 O : -0.290813	OH
3 H : 0.001440		36 H : 0.212018	OH
4 C : 0.156582		37 O : -0.403642	
5 H : -0.006780		38 C : 0.120974	
6 O : -0.331279	OH	39 H : -0.001541	
7 H : 0.196003		40 C : 0.307592	
8 C : 0.121341		41 H : 0.004728	
9 H : -0.007058		42 H : 0.014152	
10 O : -0.326728	OH	43 O : -0.391033	
11 H : 0.199776		44 C : 0.327707	
12 C : 0.130911		45 H : -0.011818	
13 H : -0.018416		46 O : -0.406701	
14 C : 0.237995		47 C : 0.116620	
15 H : -0.002410		48 H : 0.005250	
16 H : 0.000951		49 C : 0.208117	
17 O : -0.375317	OH	50 H : 0.010858	
18 H : 0.208937		51 H : 0.020622	
19 O : -0.403760		52 O : -0.364884	OH
20 C : 0.404446		53 H : 0.200152	OH
21 H : -0.014052		54 C : 0.143791	
22 O : -0.400708		55 H : -0.004122	
23 C : 0.198488		56 O : -0.335338	OH
24 H : 0.002710		57 H : 0.199603	OH
25 C : 0.153046		58 C : 0.118608	
26 H : -0.007431		59 H : 0.014558	
27 O : -0.347581	OH	60 O : -0.314655	OH
28 H : 0.194970		61 H : 0.203497	OH
29 C : 0.101734		62 C : 0.190824	
30 H : 0.008595		63 H : -0.010774	
31 O : -0.301519	OH	64 O : -0.315824	OH
32 H : 0.188703	OH	65 H : 0.211510	OH

Sum of atomic charges: -0.0

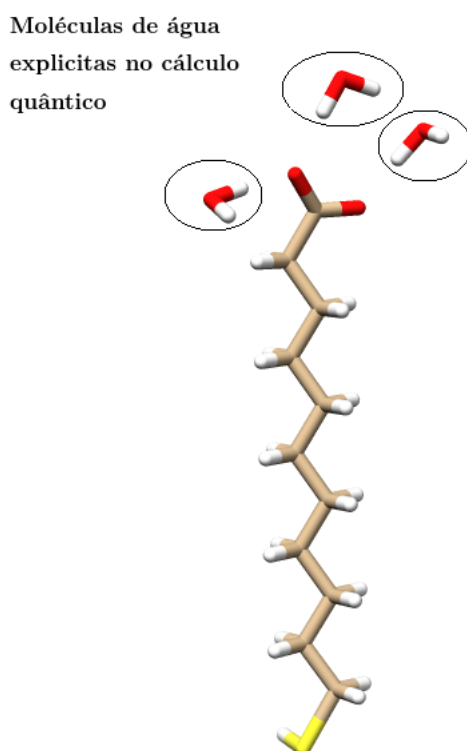
Fonte: Autor, 2019.

No campo de força gromos53a6 é usado o conceito de átomos unidos. Dessa forma, a inserção das cargas na topologia é feita somando as cargas dos hidrogênios de cada carbono com o respectivo átomo.

Uma limitação do método de Mulliken para determinar as cargas a partir da função de onda otimizada é o fato dele ser sensível a base utilizada. Dessa forma, optamos pela base de segunda geração def2-SVP (WITTE; NEATON; HEAD-GORDON, 2016), uma das mais usadas atualmente pelo seu desempenho.

No caso de moléculas expostas ao solvente, como a parte polar do ácido mercaptoundecanóico da monocamada, há uma grande interferência das moléculas de água próximas à nuvem eletrônica. Portanto, para se obter um resultado ainda mais realista, faz-se necessária a introdução de moléculas de água explícitas, para representar as interações específicas de curto alcance, antes da realização dos cálculos quânticos, como podemos ver na Figura 14.

Figura 14 – Moléculas de água explícitas colocadas nas proximidades da região do grupo ácido carboxílico exposto ao solvente.



Fonte: Autor, 2019.

Para determinar um número adequado de moléculas de água devemos lembrar que o pKa de uma molécula de ácido está diretamente relacionado à variação da energia livre de dissociação do próton desse ácido. A escolha de 3 moléculas de água próximas não é aleatória, sendo fartamente documentado na literatura. Quando se

calcula pKa teórico para moléculas usando primeiros princípios, os melhores resultados e mais próximos do sistema experimental, são alcançados com a adição de moléculas de água explícitas, como podemos ver em Pliego e Riveros (2019).

Quando submetemos a estrutura do ácido com moléculas de água, o software Orca retorna as cargas individuais de cada átomo do ácido mercaptoundecanóico na forma desprotonada e protonada, como vemos na Figura 15.

Figura 15 – Cargas do ácido mercaptoundecanóico calculadas pelo Orca.

MULLIKEN ATOMIC CHARGES	MULLIKEN ATOMIC CHARGES
0 C : 0.196397	0 C : 0.148011
1 C : -0.001516	1 C : 0.047944
2 C : 0.011475	2 C : 0.008619
3 C : 0.007108	3 C : 0.017807
4 C : 0.013620	4 C : 0.021067
5 C : 0.015992	5 C : 0.021155
6 O : -0.313759	6 O : -0.424669
7 S : -0.152125	7 S : -0.155727
8 O : -0.252344	8 O : -0.429613
9 C : 0.010783	9 C : 0.019042
10 C : 0.017841	10 C : 0.021793
11 C : -0.009089	11 C : -0.001292
12 C : 0.053959	12 C : 0.056528
13 C : -0.099024	13 C : -0.089855
14 H : 0.201764	14 H : -0.047587
15 H : 0.053279	15 H : 0.013649
16 H : -0.000225	16 H : -0.016124
17 H : 0.004704	17 H : -0.009879
18 H : 0.001227	18 H : -0.011105
19 H : -0.001462	19 H : -0.011740
20 H : -0.003530	20 H : -0.011463
21 H : -0.003205	21 H : -0.010410
22 H : -0.007090	22 H : -0.010416
23 H : -0.004304	23 H : -0.010075
24 H : -0.008046	24 H : -0.009629
25 H : -0.003703	25 H : -0.009282
26 H : -0.007755	26 H : -0.008045
27 H : -0.002830	27 H : -0.007730
28 H : -0.006751	28 H : -0.003068
29 H : 0.002178	29 H : -0.002822
30 H : -0.001904	30 H : -0.001336
31 H : 0.003178	31 H : 0.002265
32 H : 0.002760	32 H : 0.052162
33 H : 0.056515	33 H : 0.055781
34 H : 0.056627	34 H : 0.079518
35 H : 0.078586	35 O : -0.438259
36 O : -0.325863	36 H : 0.130003
37 H : 0.169551	37 H : 0.110627
38 H : 0.202050	38 O : -0.385527
39 O : -0.323667	39 H : 0.170797
40 H : 0.208659	40 H : 0.157708
41 H : 0.160254	41 O : -0.364760
42 O : -0.350906	42 H : 0.176211
43 H : 0.182880	43 H : 0.159724
44 H : 0.167714	Sum of atomic charges: -1.0000000
Sum of atomic charges: -0.0000000	

Fonte: Autor, 2019.

A primeira coisa a se perceber na Figura 15, é que a soma das cargas no ácido não é exatamente -1 ua (Unidade atômica) na forma desprotonada. O resultado é afetado porque a carga se distribui ao longo das nuvens eletrônicas das moléculas de água colocadas explicitamente no sistema.

O mesmo ocorre na molécula de ácido protonado, onde há uma leve carga

negativa, ao contrário do que se esperaria. Isso se deve a contribuição eletrônica das moléculas de água que doam densidade eletrônica ao ácido, mas, como no caso anterior, a carga total no sistema não muda.

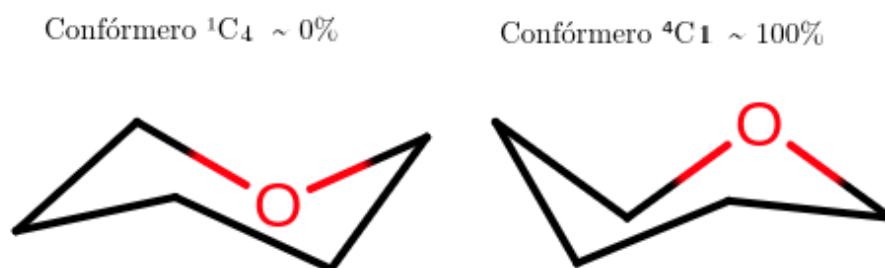
As cargas atômicas devem ser novamente somadas em grupos não polares para manter o conceito de átomos unidos do grupo e, então, podem ser adicionadas diretamente na topologia do campo de força, permitindo o início da simulação.

3.6 Validação da parametrização

Com as topologias desenvolvidas para a dinâmica molecular finalizadas, devemos usar alguma propriedade das moléculas que possa ser comparada a dados experimentais, como garantia da qualidade da parametrização.

Uma forma muito utilizada para testar a parametrização de moléculas de carboidratos é através da conformação química (PANCZYK; PLAZINSKI, 2018). A manose possui dois isômeros conformacionais 1C_4 e 4C_1 com proporção em solução bem definidas pela ressonância magnética nuclear (POL-FACHIN et al., 2012), como podemos ver na Figura 16.

Figura 16 – Confórmeros 1C_4 e 4C_1 da molécula de manose. Em solução aquosa, temperatura de 298 K e pressão de 1 bar, aproximadamente 100% dos confórmeros são do tipo 4C_1 .

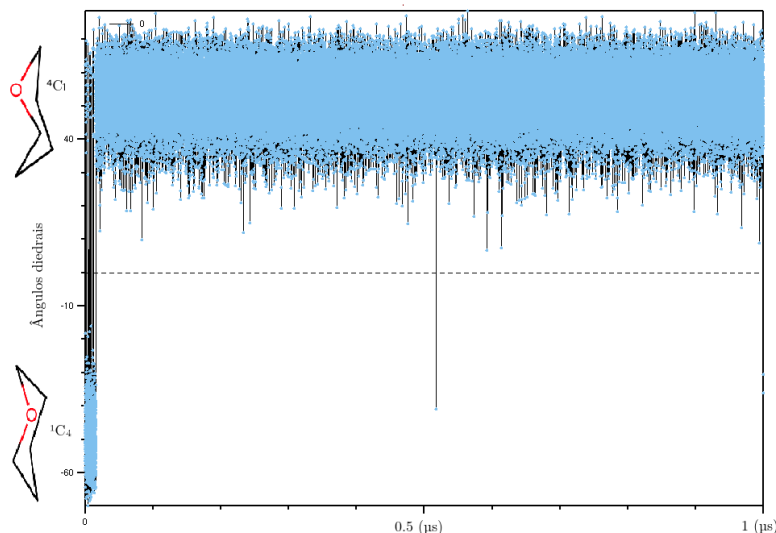


Fonte: Autor, 2019.

Dessa forma, podemos realizar uma dinâmica molecular com apenas um dos isômeros espaciais da manose, usando a conformação 1C_4 como condição inicial, para verificarmos se a parametrização usada consegue reproduzir as propriedades conformacionais. A dinâmica molecular da manose foi executada por um tempo de $1\mu s$, com temperatura a 300 K, pressão de 1 bar e caixa cúbica com arestas de 2.77 nm.

Como podemos verificar na Figura 17, a manose permanece na conformação 4C_1 durante quase todo o tempo de dinâmica molecular. A probabilidade de encontrarmos a molécula na conformação 1C_4 é próxima de zero (POL-FACHIN et al., 2012).

Figura 17 – Ângulos diédricos das conformações da manose. Na maior parte do tempo de dinâmica molecular a conformação 4C_1 se mantém. Existem poucos instantes onde a conformação pode mudar para 1C_4 . A baixa probabilidade de haver a conformação 1C_4 é uma propriedade bem definida experimentalmente, demonstrando a boa qualidade da parametrização.



Fonte: Autor, 2019.

O uso da condição inicial na forma 1C_4 demonstra que a parametrização consegue reproduzir bem essa propriedade experimental da molécula. Com essa validação, concluímos que a metodologia de parametrização das moléculas pode ser usada no campo de força com confiabilidade.

3.7 Valor experimental da energia livre

Ao analisarmos os dados experimentais obtidos por Santos et al. (2014), podemos usar a constante de afinidade $K_a = (1.6 \pm 0.6)10^8 \text{ Lmol}^{-1}$ para obter a energia livre através da equação:

$$\Delta G = -RT \ln(k_a \pm \Delta k_a) = -RT \left[\ln(k_a) \pm \frac{\Delta k_a}{k_a} \right] \quad (57)$$

Dessa forma, ao substituirmos os valores na equação 57, usando as constantes $R = 8.31 \text{ Jmol}^{-1} \text{ K}^{-1}$ e $T = 298 \text{ K}$, obtemos o valor para energia livre experimental e podemos comparar com a energia livre que calculamos teoricamente.

Tabela 1 – Energia livre experimental.

Energia livre	kJ.mol^{-1}
Experimental	46.780 ± 0.928

Fonte: Santos et al. (2014).

4 Resultados

4.1 Execução da dinâmica molecular

As estruturas dos complexos lectina-peroxidase foram submetidas à dinâmica molecular com o software GROMACS, onde, por meio do cálculo de um campo de força, é realizada a evolução temporal de biomoléculas solvatadas dentro de uma caixa de água, pressão de 1 bar e temperatura de 300 Kelvin, onde existem condições periódicas de contorno para evitar problemas com descontinuidade de energia (WELLS; CHAFFEE, 2015).

Uma vantagem das condições de contorno periódicas o fato de permitir a simulação de uma superfície 2D infinita, que se assemelha muito à situação das moléculas imobilizadas, com baixo custo computacional.

A simulação molecular deve ser executada por um tempo relativamente longo, em torno de 200 ns, para permitir que uma fração estatisticamente significativa das conformações sejam alcançadas pelo sistema em estudo. Além disso, está medida garante o cálculo correto do termo entrópico na energia livre, por meio da aproximação quase-harmônica e dos modos normais de vibração.

O uso de diferentes ambientes químicos nas simulações moleculares permite a análise do efeito desses fatores na capacidade de reconhecimento molecular. A variação das condições iniciais é uma forma de explorarmos um número maior de conformações do sistema, dentro do limite de tempo da simulação.

Outra vantagem de usar um tempo tão longo de simulação, é a possibilidade de garantir que o intervalo de tempo escolhido para se calcular a energia seja o mais estável possível, diminuindo a margem de erro nos cálculos da energia livre.

Com o GROMACS, foram gerados os arquivos contendo as informações da evolução temporal das posições das moléculas livres em solução dentro da caixa. Isso permite o cálculo da variação estrutural das moléculas durante todo o tempo de realização da dinâmica molecular, através do desvio padrão médio quadrático (RMSD). Com essa informação, é possível definir com mais precisão os períodos da simulação molecular em que o sistema está mais estável.

4.2 Complexos livres

Foram realizadas três dinâmicas moleculares com o complexo livre em solução, em condições iniciais diferentes. As duas primeiras foram realizadas utilizando as duas predições estruturais melhor ranqueadas no ZDOCK. A terceira usou como condição inicial a conformação do complexo em 10 ns de uma simulação molecular que teve

como condição inicial o *frame* melhor ranqueado.

As simulações foram executadas por um período aproximado de 200 ns. As dimensões das caixas cúbicas de dinâmica molecular usadas nos sistemas livres possuem arestas de, aproximadamente, 10 nm. Todas as caixas usaram 3 íons sódio para neutralizar a carga do complexo formado pelas proteínas.

Essas simulações em condições iniciais diferentes tiveram como finalidade a averiguação da estabilidade do complexo durante o tempo de simulação, além de estimar a quantidade ideal de *frames* necessários para se calcular a energia livre por um tempo razoável, sem perder a qualidade dos dados.

Cada uma das dinâmicas moleculares realizadas gerou aproximadamente cem mil *frames*. Calcular a média da energia livre de todos poderia levar anos e não geraria um resultado melhor do que calcular a média de um número estatisticamente significativo.

Um teste realizado com um modelo proteína-ligante, em menor escala, demonstrou que, a partir de 500 *frames*, já não havia uma grande mudança em comparação com 25 mil *frames*. Dessa forma, as trajetórias foram reduzidas a aproximadamente 2000 *frames*, onde o cálculo da energia livre leva 5 dias para ficar pronto, com a confiabilidade de uma amostra estatisticamente significativa.

A simulação das moléculas livres, além de ser mais rápida, pois exige uma caixa menor, permite a comparação dos efeitos da imobilização na afinidade entre as proteínas que poderiam ser favoráveis ou desfavoráveis à energia livre de ligação.

4.2.1 Componente entálpica

Nas Tabelas 2, 3 e 4 podemos observar o resultado do cálculo das componentes entálpicas da energia livre para as três diferentes condições iniciais. As condições iniciais 1 e 2 referem-se a primeira e segunda predições estruturais melhor ranqueadas pelo ZDOCK. A condição inicial 3 refere-se a uma conformação retirada de outra dinâmica molecular.

Tabela 2 – Energia livre condição inicial 1.

Contribuição entálpica	kJ.mol^{-1}
Energia de van der Waals	-461.059 ± 48.322
Energia eletrostática	-169.657 ± 56.089
Energia de solvatação polar	433.897 ± 83.777
Energia de solvatação apolar	-47.743 ± 6.316
Energia de ligação	-244.562 ± 92.700

Fonte: Autor, 2019.

Tabela 3 – Energia livre condição inicial 2.

Contribuição entálpica	$\text{kJ}\cdot\text{mol}^{-1}$
Energia de van der Waals	-387.237 ± 31.133
Energia eletrostática	-198.302 ± 81.346
Energia de solvatação polar	416.212 ± 120.734
Energia de solvatação apolar	-43.753 ± 5.543
Energia de ligação	-213.080 ± 73.979

Fonte: Autor, 2019.

Tabela 4 – Energia livre condição inicial 3.

Contribuição entálpica	$\text{kJ}\cdot\text{mol}^{-1}$
Energia de van der Waals	-405.231 ± 43.133
Energia eletrostática	-170.797 ± 60.686
Energia de solvatação polar	365.395 ± 103.098
Energia de solvatação apolar	-44.094 ± 6.314
Energia de ligação	-254.727 ± 73.726

Fonte: Autor, 2019.

O cálculo da parte entálpica da energia livre realizada nos mostra que todos os complexos possuem energia menor que -200 kJmol^{-1} . Além disso, os resultados dos três complexos estão dentro do erro padrão do método (KUMARI; KUMAR; LYNN, 2014), mas, como vemos na Equação 36, ainda há uma fração significativa da energia livre que não foi computada, pois a entropia é desconsiderada no método MM-PBSA.

Outra informação interessante que percebemos é o fato de a maior contribuição para a energia livre de ligação ser negativa, portanto, favorável, vem do termo de van der Waals. Como essa é uma das interações intermoleculares menos intensas, esse resultado indica que há uma grande área de contato entre as proteínas.

Em todos os complexos a energia eletrostática contribui favoravelmente à energia livre de ligação, mas essa contribuição é sempre inferior na energia de van der Waals, sendo um indicativo de que as proteínas possuem uma baixa complementaridade, algo já esperado pois elas não têm nenhuma correlação biológica e possuem funções distintas.

O termo da energia de solvatação apolar é quase idêntico nos três complexos, mostrando que a área de superfície acessível ao solvente é quase igual em ambos os casos. O termo de solvatação polar é o único que retorna valores positivos para a energia.

A maior fonte de imprecisão do cálculo da energia livre vem dessa componente. Esse fato pode ser explicado pelo processo de linearização da equação de Poisson-Boltzmann, o que reduz o custo computacional, mas parece ser muito sensível a flutuações.

A flutuação do método MM-PBSA nos três sistemas está acima de 70 kJmol^{-1} , indicando a complexidade de lidar com cálculo de energia média em sistemas dinâmicos.

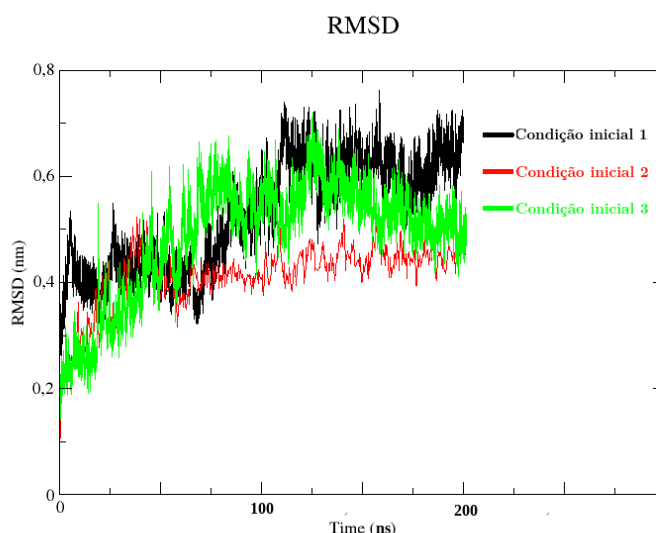
A maior imprecisão desse método ocorre devido a grupos muito carregados em sistemas com carga líquida, o que afetaria a decomposição em domínios físicos. No entanto, como vemos da análise estrutural, o complexo não contém tais grupos. A carga total das proteínas é de apenas -3 ua, onde a maior parte carga localizada resulta da presença de íons cálcio. Dessa forma, podemos considerar que a média da energia de solvatação polar deve ser mais importante do que a flutuação intrínseca ao método.

Para os cálculos da parte vibracional da entropia foram usados os métodos de modos normais e de aproximação quase-harmônica.

4.2.2 Entropia quase-harmônica nos complexos livres

A aproximação quase-harmônica exige que a estrutura seja extremamente estável durante toda sua trajetória e em dinâmicas extremamente longas para haver convergência. Essa análise explora toda a extensão global do espaço conformacional que o complexo pode atingir a uma temperatura T . Além disso, toda rotação e translação do complexo deve ser removida em um pós-tratamento dos arquivos de trajetória. Para averiguar a estabilidade das estruturas durante as dinâmicas, foi calculado o RMSD dos carbonos alfa nos três complexos.

Figura 18 – Gráfico RMSD das estruturas nos sistemas livres durante a dinâmica.



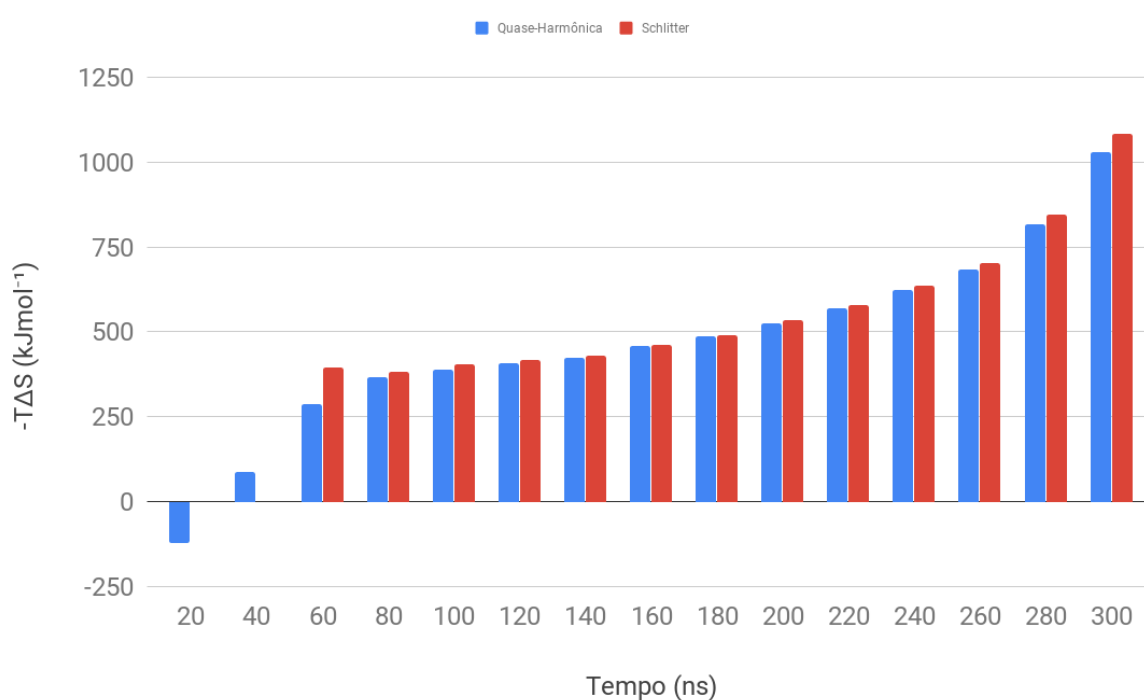
Fonte: Autor, 2019.

No gráfico da Figura 18, percebemos que as estruturas dos sistemas livres permanecem estáveis durante toda a dinâmica, com RMSD sempre abaixo de 1

nm. Isso indica que nenhuma delas sofre grandes mudanças conformacionais, uma condição necessária para o cálculo da aproximação quase-harmônica.

Devido à estabilidade, o sistema na condição inicial 1 parece ser o mais indicado para o cálculo da entropia quase-harmônica. Então, após tratamento das trajetórias com remoção de rotação e translação do complexo, foram obtidos os resultados que podem ser vistos no gráfico da Figura 19.

Figura 19 – Termo entrópico da energia livre calculado pela aproximação quase-harmônica em função do tempo. O método falha antes de 60 ns e depois de 300 ns, indicando que a estrutura não é adequada para a análise de componentes principais.



Fonte: Autor, 2019.

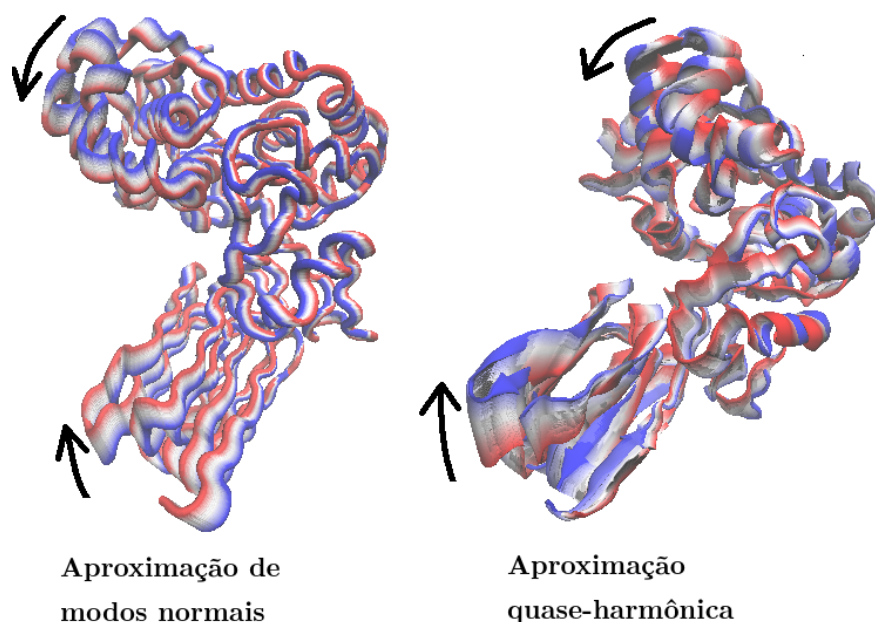
Com dinâmicas moleculares longas, seria esperado que o método PCA encontrasse os componentes principais para estimar as frequências. A partir desse ponto, o resultado sofreria pouca variação e a entropia deveria convergir. Ao analisarmos o gráfico da Figura 19, percebemos entropia quase-harmônica não converge, mesmo com dinâmicas moleculares estendidas, mais longas que 300 ns, além de retornar valores maiores que o módulo do termo entálpico até o tempo de 300 ns.

Antes de 60 ns e após 300 ns, o método falha ao calcular a entropia do sistema usando a fórmula de Schlitter (1993), como podemos ver no *output* do programa anaeig do GROMACS: “*The Entropy due to the Schlitter formula is nan J/mol K*”, indicando um valor não válido. Ou seja, o sistema não é estável o bastante para usar esse método de cálculo de entropia e o termo entrópico, $-T\Delta S$, retorna valores negativos.

O único valor obtido que explicaria a estabilidade do complexo seria a entropia de 89.46 kJmol^{-1} obtido em 40 ns, mas não há qualquer indicativo de que o sistema convergiu para esse valor. Além disso, o valor da entropia quase-harmônica usando a fórmula de Schlittler falha, e, portanto, não podemos usar esse valor.

Ao analisarmos a sobreposição dos movimentos dos átomos na Figura 20, no arquivo de movimento do primeiro modo normal de vibração do complexo e o deslocamento dos átomos do complexo na trajetória de dinâmica molecular, onde foram removidas a rotação e translação, obtemos imagens muito parecidas.

Figura 20 – Sobreposição do deslocamento dos átomos do complexo com condição inicial 1. As cores representam diferentes posições da estrutura em função do tempo de dinâmica molecular. O grande deslocamento demonstra que os átomos não se comportam de forma quase-harmônica originando a falha da aproximação do método PCA.



Fonte: Autor, 2019.

Essas imagens nos mostram a razão pela qual a entropia de aproximação quase-harmônica não funciona nesse sistema, uma vez que observamos um grande deslocamento dos átomos em comparação com sua posição inicial.

O sistema comporta-se como um joelho sendo flexionado, estando longe da idealidade necessária a este método, onde se supõe que os átomos oscilem ao redor do estado fundamental de forma quase-harmônica.

Com a inviabilidade de usar a aproximação quase-harmônica para estimar a entropia, devemos usar outro método.

4.2.3 Entropia de modos normais nos complexos livres

O cálculo de modos normais analisa o espaço de configuração local, com base em um único ponto, supondo que seja o mínimo da energia potencial local. Dessa forma, feita a retirada de um dos *frames* da trajetória, e, com o mesmo campo de força da dinâmica molecular, é realizada a minimização de energia para garantir que todos os átomos estejam no estado fundamental de energia.

Tabela 5 – Termo entrópico da energia nos complexos livres.

MM-PBSA	kJ.mol⁻¹
Condição inicial 1	128.820
Condição inicial 2	118.560
Condição inicial 3	202.650

Fonte: Autor, 2019.

Embora seja necessária a minimização de energia, o que modifica um pouco o estado original do complexo, como podemos observar na Tabela 5, a entropia de modos normais retorna valores mais condizentes com o esperado, uma energia positiva que contribui de modo desfavorável a energia livre de ligação.

4.2.4 Energia livre dos sistemas livres

Ao somarmos a entalpia calculada pelo método MM-PBSA com a contribuição entrópica calculada pelo método dos modos normais de vibração, obtemos o seguinte resultado:

Tabela 6 – Energia livre total dos sistemas livres.

Energia livre	kJ.mol⁻¹
Condição inicial 1	-115.742 ± 92.700
Condição inicial 2	-94.52 ± 73.979
Condição inicial 3	-52.077 ± 73.726

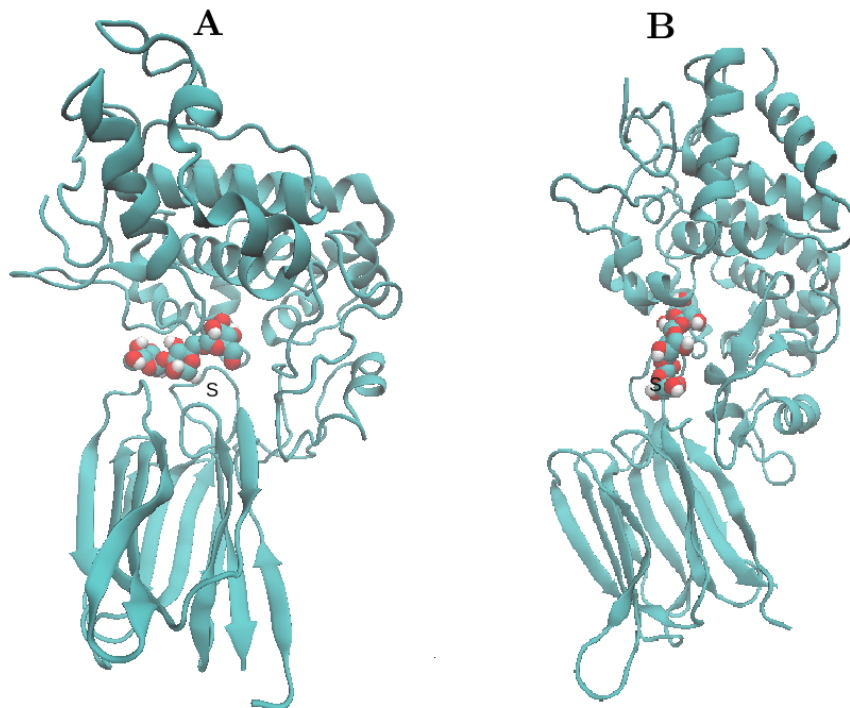
Fonte: Autor, 2019.

Como podemos ver na Tabela 6, os valores de energia livre final dos complexos varia de -115.742 kJmol⁻¹ a -52.077 kJmol⁻¹, sendo que esses valores se encontram dentro do erro padrão do método MM-PBSA calculado pela equação 44.

Os valores negativos são coerentes com o que seria esperado de um complexo, indicando que a energia livre é favorável à ligação. O complexo que usou a condição inicial 3 obteve o valor mais próximo do valor experimental de 46.780 kJmol⁻¹.

Os módulos das energias livres nas condições iniciais 1 e 2 possuem aproximadamente o dobro do módulo na condição inicial 3. Esse fato explicado pelo deslocamento da molécula de manotriose do sítio de ligação original na lectina, como podemos ver na Figura 21:

Figura 21 – Conformações da molécula de manotriose **A**: Na condição inicial 1 a extremidade da manotriose é deslocada do sítio de ligação original na alça da lectina. **B**: Na condição inicial 3 a extremidade da manotriose se mantém no sítio original da lectina. Nas duas situações, o sítio foi identificado com a letra S.



Fonte: Autor, 2019.

O resultado demonstra que a manotriose mais exposta ao contato com a peroxidase pode ser a responsável pela redução da energia livre calculada nos sistemas.

4.3 Complexos imobilizados

Uma dúvida levantada sobre o sistema é qual o real estado de protonação da monocamada, dado que, após a imobilização da lectina no biossensor, os ácidos carboxílicos remanescentes são recobertos com fragmentos de proteínas hidrolisadas, cuja composição é incerta. Dessa forma, optou-se por fazer três dinâmicas moleculares de 200 ns do sistema, com três estados de protonação distintos e a mesma condição inicial do complexo, usando a predição estrutural melhor ranqueada pelo ZDOCK.

Essa variação do estado de protonação da monocamada justifica-se devido a uma das limitações das dinâmicas moleculares clássicas que não permite a observação de fenômenos eletrônicos, como a mudança de estado de protonação de um ácido devido à presença de algum grupo químico. Portanto, ao variarmos o estado de protonação, é possível ver o efeito provocado pelo pH na afinidade entre as moléculas do complexo.

As caixas de simulação dos sistemas imobilizados possuem aproximadamente 12 nm em x, 12 nm em y e 17 nm em z, com um volume aproximado de 2541 nm³. A neutralidade no sistema é um fator essencial para que o algoritmo realize as condições periódicas de contorno. Dessa forma, quando se simula uma monocamada, é necessário adicionar íons sódio para neutralizar as cargas da monocamada.

Para a monocamada totalmente desprotonada foram adicionados 363 íons sódio e no sistema com metade dos ácidos protonados foram necessários 206 íons sódio. No sistema totalmente protonado, assim como nos sistemas livres, foram necessários apenas 3 íons sódio para neutralizar a carga dos complexos.

A diferença de valor entre o número de íons, que não é exatamente metade entre os sistemas totalmente protonado e semi-protonado deve-se ao fato de que as monocamadas foram construídas manualmente, fazendo com que o número ácidos na superfície variasse. Além disso, as cargas dos ácidos não são números inteiros, uma consequência do cálculo quântico com águas explícitas.

A primeira dinâmica foi realizada com grupos ácidos na monocamada totalmente desprotonados; a segunda com metade deles desprotonados, equivalendo ao estado de pK_a dos ácidos da monocamada; e a terceira, e última, com todos os ácidos protonados.

O uso de uma monocamada totalmente protonada justifica-se pois, apesar do pK_a dos ácidos carboxílicos ser inferior a 5, queremos simular uma monocamada totalmente neutralizada, situação que é melhor representada pelo ácido protonado de carga total nula.

Tabela 7 – Energia livre do sistema imobilizado totalmente desprotonado.

Contribuição entálpica	kJ.mol⁻¹
Energia de van der Waals	-400.072 ± 48.091
Energia eletrostática	-204.342 ± 68.957
Energia de solvatação polar	475.242 ± 125.723
Energia de solvatação apolar	-44.313 ± 7.002
Energia de ligação	-173.485 ± 76.473

Fonte: Autor, 2019.

Tabela 8 – Energia livre do sistema imobilizado semi-protonado.

Contribuição entálpica	kJ.mol⁻¹
Energia de van der Waals	-331.986 ± 48.953
Energia eletrostática	-69.749 ± 76.148
Energia de solvatação polar	390.703 ± 120.372
Energia de solvatação apolar	-37.004 ± 7.335
Energia de ligação	-48.036 ± 94.294

Fonte: Autor, 2019.

Tabela 9 – Energia livre do sistema imobilizado totalmente protonado.

Contribuição entálpica	kJ.mol^{-1}
Energia de van der Waals	-340.905 ± 54.174
Energia eletrostática	-107.624 ± 92.520
Energia de solvatação polar	357.481 ± 149.153
Energia de solvatação apolar	-38.742 ± 7.257
Energia de ligação	-129.792 ± 81.772

Fonte: Autor, 2019.

A primeira diferença que se pode notar, é como o módulo da contribuição entálpica da energia livre diminui nos sistemas com diferentes estados de protonação, quando comparada à energia dos sistemas livres em solução.

O valor mais notável é o que ocorre no complexo imobilizado com metade dos ácidos protonados, onde há uma contribuição entálpica de $-48.036 \text{ kJmol}^{-1}$. Esse valor baixo em módulo inicialmente levantou a suspeita de que houvesse algum erro de indexação, fazendo com que resíduos importantes a ligação não fossem contabilizados nos cálculos do MM-PBSA.

Para averiguar um possível erro, um dos frames do complexo durante a trajetória foi retirado e usado na dinâmica do sistema livre como condição inicial 3, que, como podemos ver na Tabela 4, originou uma contribuição entálpica de $-254.727 \text{ kJmol}^{-1}$.

No caso imobilizado percebemos que o termo de van de Waals, novamente, é o que mais contribuiu para a energia livre de ligação favorável, indicando que a grande área de contato entre as proteínas parece ter um papel mais importante do que a alta especificidade da ligação.

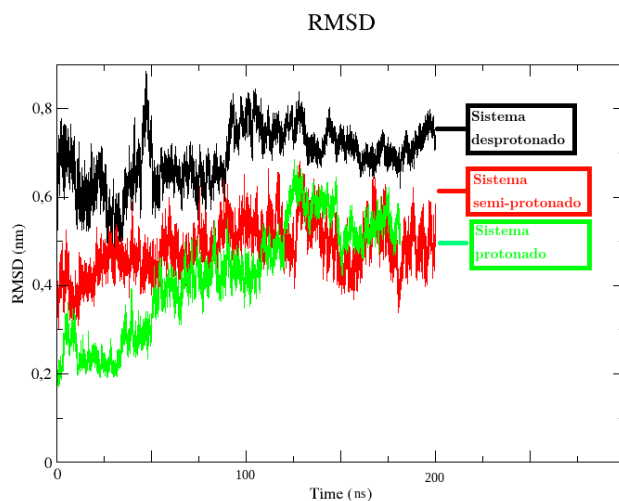
O termo energia de solvatação polar no caso imobilizado é maior em módulo que o termo de van de Waals em todos os complexos imobilizados, o que é uma mudança perceptível em relação ao complexo livre, no qual em apenas um caso esse termo foi maior em módulo.

O termo de energia eletrostática é o que mais sofre alterações nos complexos imobilizados, variando de $-204.342 \text{ kJmol}^{-1}$ no sistema totalmente desprotonado até $-69.749 \text{ kJmol}^{-1}$ no sistema semi-protonado. A variação da energia eletrostática pode ser explicada por uma mudança conformacional, onde a área de contato entre as proteínas foi menor, o que pode ser percebido no termo de energia de van der Waals menos favorável no sistema semi-protonado.

4.3.1 Entropia quase-harmônica nos complexos imobilizados

A estabilidade dos complexos imobilizados foi novamente avaliada pelo RMSD como podemos ver na Figura 22.

Figura 22 – Gráfico RMSD das estruturas nos sistemas imobilizados durante a dinâmica.

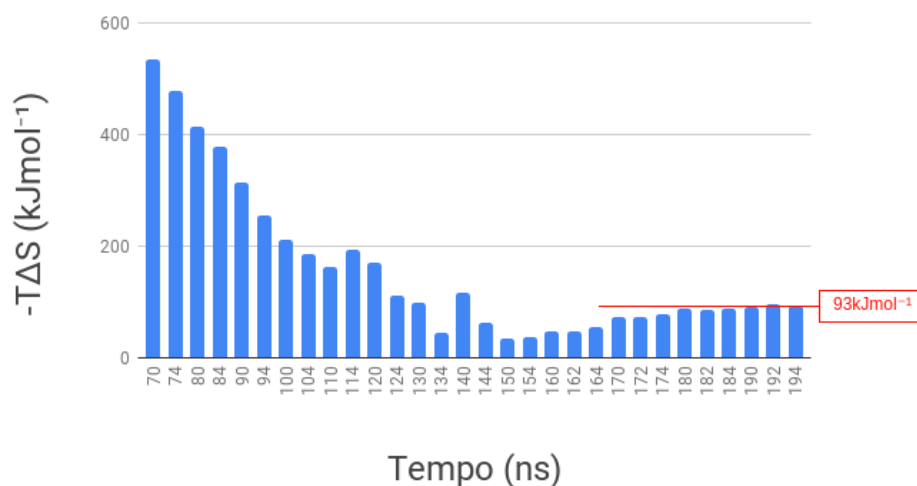


Fonte: Autor, 2019.

No gráfico, todos os complexos imobilizados possuem um RMSD menor que 0.8 nm durante as suas trajetórias, indicando que não sofrem grandes mudanças estruturais de enovelamento ou desnaturação. No entanto, em nenhum dos gráficos o complexo atinge um platô, que indicaria uma alta estabilidade estrutural.

Devido a sua maior estabilidade do RMSD, a metodologia de aproximação quase-harmônica foi testada no sistema totalmente desprotonado, usando todos os *frames* a partir de 2 ns, como vemos na Figura 23:

Figura 23 – Termo entrópico do sistema desprotonado a partir de 2 ns de simulação.



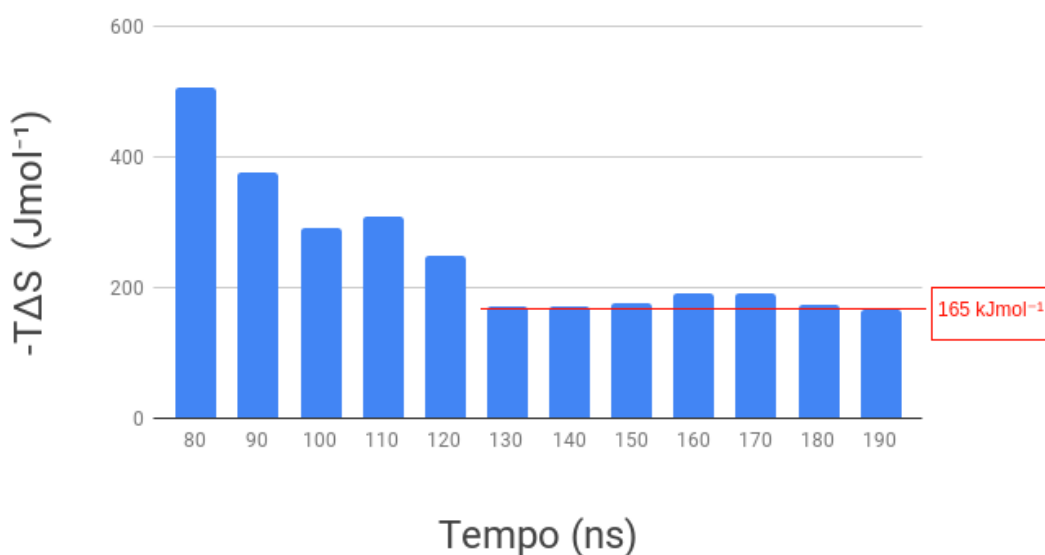
Fonte: Autor, 2019.

Quando analisamos o gráfico da Figura 23, percebemos o que parece ser uma convergência a partir de 170 ns, onde o termo entrópico da energia livre parece

estabilizar em torno 93 kJmol^{-1} .

Embora esse resultado seja coerente com o esperado, ele deve ser obtido começando de tempos de dinâmica diferentes, como uma garantia de que não seja apenas um valor aleatório. Dessa forma, o método da aproximação quase-harmônica foi aplicado novamente, usando todos os *frames* a partir de 4ns, como vemos na Figura 24:

Figura 24 – Termo entrópico do sistema desprotonado a partir de 4 ns de simulação.



Fonte: Autor, 2019.

No gráfico, a aproximação quase-harmônica gera um resultado que parece ser novamente uma convergência, mas o valor é consideravelmente mais alto, acima de 165 kJmol^{-1} , quase o dobro do valor anterior.

Esse resultado indica que realmente a aproximação quase-harmônica não é a abordagem mais adequada para se estimar a contribuição entrópica no complexo analisado, pois, mesmo quando parece haver uma convergência, ele pode levar a resultados enganosos.

O resultado divergente ocorre provavelmente devido às pequenas flutuações que vemos no gráfico de RMSD, onde, em nenhum momento, os complexos parecem atingir um platô estável, indicando que, mesmo pequenas modificações conformacionais, afetam muito o método de aproximação quase-harmônica.

4.3.2 Entropia de modos normais nos complexos imobilizados

Ao calcularmos a entropia de modos normais, a dúvida sobre o possível erro no cálculo do MM-PBSA do complexo imobilizado semi-protonado desaparece, como

podemos ver na tabela a seguir:

Tabela 10 – Termo entrópico da energia nos complexos imobilizados.

MM-PBSA	kJ.mol^{-1}
Totalmente desprotonado	77.010
Semi-protonado	3.390
Totalmente protonado	85.560

Fonte: Autor, 2019.

Quando analisamos os valores, notamos que a contribuição entrópica para a energia livre dos sistemas imobilizados diminui consideravelmente quando comparado à energia dos sistemas livres. Dessa forma, a diferença de energia acaba sendo compensada, não havendo uma variação muito grande na energia livre total.

4.3.3 Energia livre dos complexos imobilizados

Ao somarmos a entalpia com a contribuição entrópica podemos entender melhor o porquê das diferenças aparentes entre os sistemas livres e imobilizados.

Tabela 11 – Energia livre total nos complexos imobilizados.

Energia livre	kJ.mol^{-1}
Totalmente desprotonado	-96.475 ± 76.473
Semi-protonado	-44.646 ± 94.294
Totalmente protonado	-44.232 ± 81.722

Fonte: Autor, 2019.

Ao analisarmos a Tabela 11, concluímos que a imobilização provoca mudanças nos valores da entalpia e do termo entrópico, alterando a energia livre de ligação.

O caso mais interessante é o do sistema imobilizado com metade dos ácidos desprotonados, onde, apesar da grande diferença em módulo da entalpia com sistema na condição inicial 3, essa diferença é quase compensada pelo termo entrópico, mostrando que os dois complexos têm energias livres de ligação próximas.

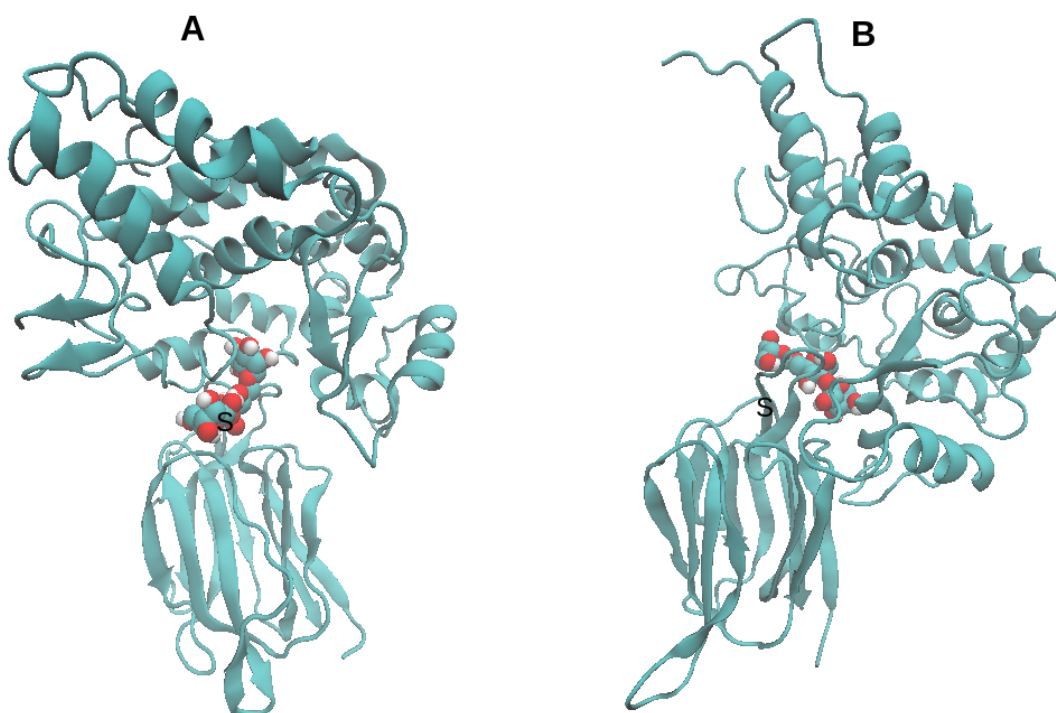
Esse resultado parecido não é coincidência, pois o sistema de condição inicial 3 surgiu de um *frame* retirado do sistema imobilizado semi-protonado para analisar o porquê dos valores de entalpia serem baixos. Como percebemos, esse valor aumentou em quase 5 vezes, apesar das conformações muito parecidas, diferença provocada apenas pela imobilização. Esse resultado confirma a eficiência do método de modos normais, pois o termo entrópico da energia livre parece acompanhar a entalpia em pequenas mudanças conformacionais.

No sistema totalmente protonado, que seria o mais próximo à condição experimental, com grupos ácidos neutralizados, obtemos uma energia livre total de -44.232

kJmol^{-1} , quase igual ao resultado do sistema semi-protonado. Esses valores estão muito próximos ao valor experimental de $46.780 \text{ kJmol}^{-1}$, mostrando que o estado de protonação parece afetar pouco a energia livre.

O resultado nos mostra que na situação onde o sistema estava totalmente desprotonado, o módulo da energia livre calculada é praticamente o dobro daquele nos sistemas semi-protonado e totalmente protonado. Ao analisarmos os arquivos de trajetória, localizamos novamente o deslocamento da molécula de mantriose para fora sítio original de ligação na lectina, como podemos verificar na Figura 25:

Figura 25 – Conformações da molécula de mantriose **A**: No sistema totalmente protonado a extremidade da mantriose se mantém no sítio de ligação original na alça da lectina. **B**: No sistema totalmente desprotonado a extremidade da mantriose desloca do sítio original da lectina. Nas duas situações o sítio foi identificado com a letra S.



Fonte: Autor, 2019.

Esse resultado demonstra que quanto maior exposição da molécula de mantriose maior o módulo da energia livre. Os valores mais próximos ao experimental são obtidos quando a mantriose se mantém no sítio original de ligação na lectina.

4.4 Comparação com o valor experimental da energia livre

O resultado da energia livre experimental publicado por Santos et al. (2014) é bem próximo dos valores obtidos nos sistemas semi-protonado e totalmente protonado, sendo um pouco menor que o valor do sistema de condição inicial 3. Além disso, está

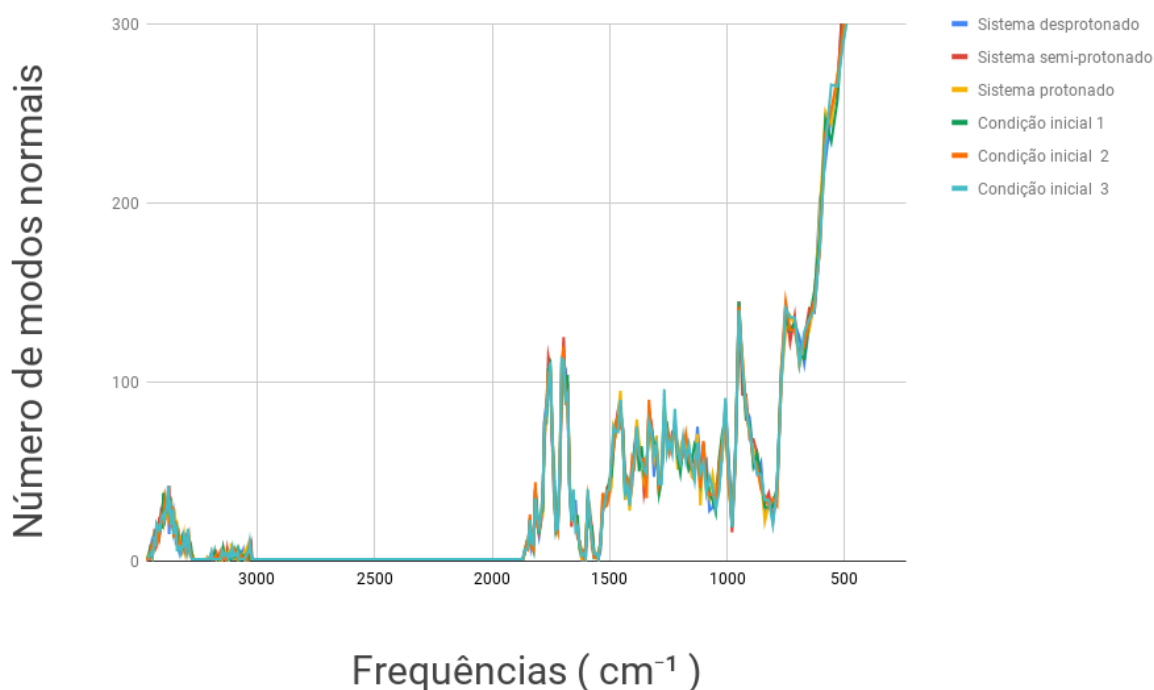
dentro da margem de erro padrão de todos os outros valores estimados para energia livre.

Devemos levar em consideração que no sistema experimental o reconhecimento entre as moléculas não é estático e o valor obtido no biossensor é uma média de todas as conformações atingidas durante o experimento. Dessa forma, resultado teórico indica que, possivelmente, a conformação nos três sistemas que mais se aproximaram da energia livre experimental podem ser parecidas com a média das conformações das proteínas no biossensor eletroquímico.

4.4.1 Número de modos normais

Uma outra análise que pode ser feita é a verificação da alteração do número de modos normais de vibração das proteínas em função da frequência como resultado do processo de imobilização na monocamada do biossensor. Uma alteração nesses modos normais poderia interferir no termo entrópico da energia e, conseqüentemente, na forma como ocorre o reconhecimento molecular. Na Figura 26, podemos ver o gráfico do número de modos normais em função das frequências.

Figura 26 – Sobreposição dos gráficos de número de modos normais em função das frequências de todas as simulações.

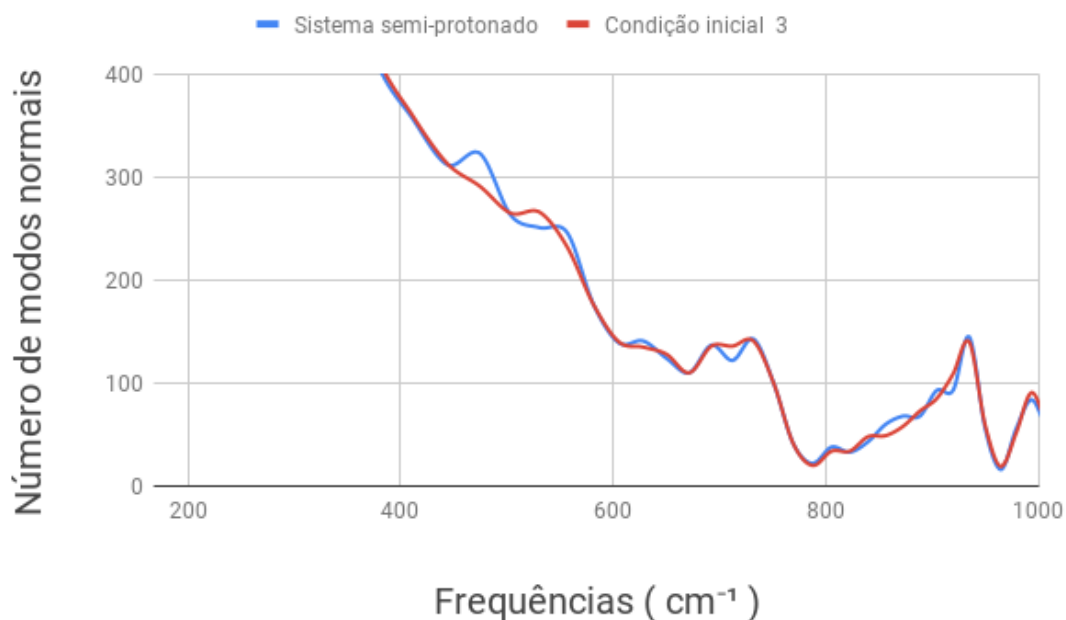


Fonte: Autor, 2019.

Quando analisamos o gráfico da Figura 26, podemos perceber que a imobilização das moléculas não parece alterar as frequências normais de vibração das proteínas.

Todos os gráficos praticamente se sobrepõem, indicando que a maioria dos modos normais são de baixas frequências, como seria esperado para proteínas.

Figura 27 – Número de modos normais em função das frequências sistema imobilizado semi-protonado e sistema com condição inicial 3.



Fonte: Autor, 2019.

Na Figura 27, ao aproximarmos a imagem do gráfico para a região de baixas frequências, percebemos pequenas mudanças no número de modos normais no sistema com condição inicial 3 e no complexo imobilizado semi-protonado. Essas alterações poderiam ser as responsáveis pelas mudanças no valor do termo entrópico da energia livre dos complexos.

O fato de haver uma imobilização pontual em apenas um único resíduo da lectina não parece ser suficiente para alterar significativamente os modos normais. Sendo assim, as variações nos valores de energia provavelmente são provocadas por mudanças conformacionais.

Essas pequenas mudanças no gráfico poderiam ser uma consequência da metodologia usada, pois a minimização de energia para o cálculo de modos normais utiliza apenas a estrutura do complexo retirado dos últimos *frames* das trajetórias de dinâmica molecular.

A ausência da monocamada poderia afetar o resultado final dos modos normais, mas, como podemos ver na comparação de energia entre o complexo na condição inicial 3 e o complexo imobilizado semi-protonado, essa diferença não parece afetar muito o resultado. A conformação dos dois complexos é bastante próxima e origina

energias livres finais muito parecidas, com a aproximação de modos normais retornando valores muito coerentes, mesmo sem a presença da monocamada.

Devemos destacar o fato de que, ao calcularmos da energia livre, estamos supondo que o complexo atingiu uma conformação estável, na qual as proteínas oscilam ao redor de um mínimo de energia local, que corresponde a média da energia livre de ligação nesse intervalo. Dessa forma, ao minimizarmos a energia com o campo de força do GROMACS, estamos estimando o termo entrópico correspondente à média da energia livre e não do *frame* usado.

4.5 Contribuição individual de cada resíduo

Com exceção das dinâmicas moleculares dos sistemas semi-protonado, totalmente protonado e do complexo com condição inicial 3, em todas as dinâmicas realizadas ocorre o deslocamento da manotriose para fora do sítio original na lectina. O deslocamento não desfaz o complexo, mas aumenta a superfície da molécula de manotriose em contato com a peroxidase. Em todas essas situações, ocorre um aumento do módulo da energia livre de ligação.

Esse aumento pode ser uma interpretação enganosa, pois a manotriose não está covalentemente ligada a lectina. Isso apenas indica que a peroxidase pode ter uma afinidade maior pela manotriose do que a própria lectina, e, em tempos de simulação maiores, poderia deslocar a manotriose do sítio de ligação da lectina e desfazer o complexo.

Dessa forma, a situação mais provável, validada pela energia calculada experimentalmente, é de uma divisão onde uma das extremidades da manotriose está em seu sítio de ligação na lectina e a outra extremidade no sítio de ligação da peroxidase.

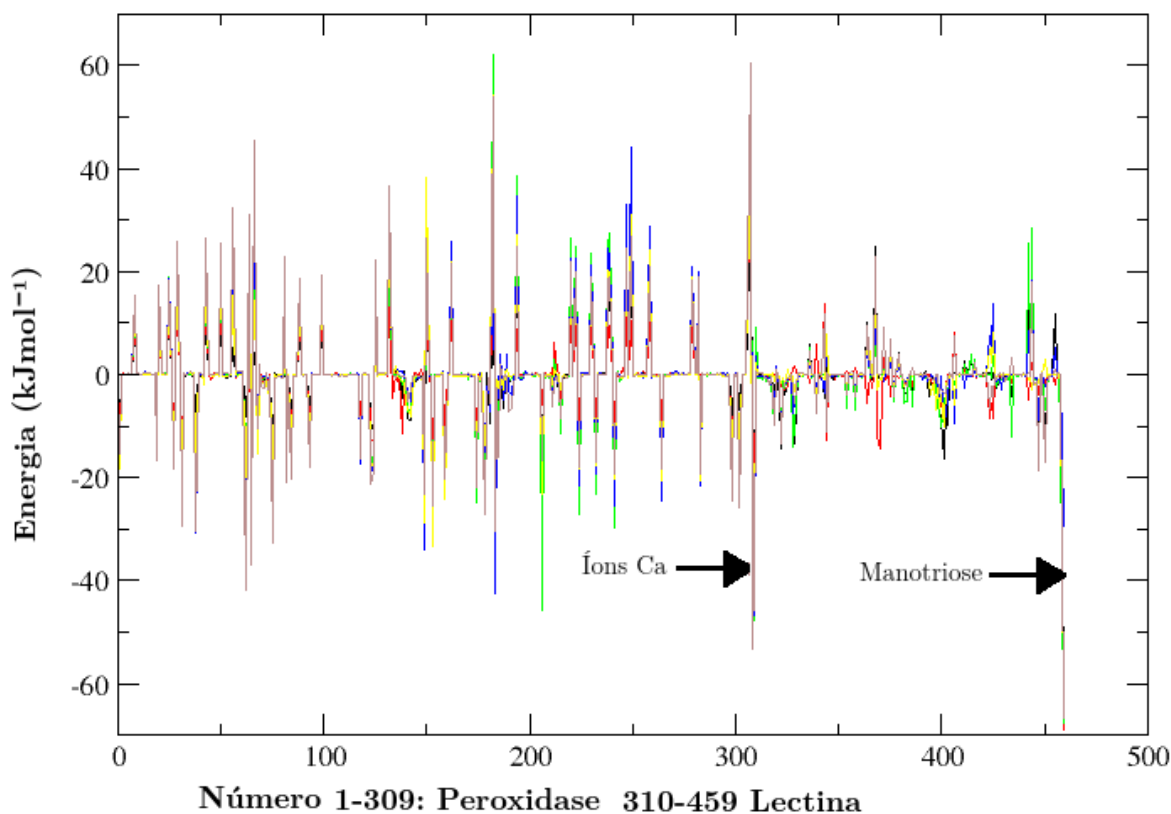
A ferramenta de decomposição de energia do pacote *g_mmpbsa* ajuda a entender as variações de energia encontradas nas diferentes simulações. Inicialmente, a ferramenta calcula a entalpia de cada átomo do complexo na forma ligada e na forma livre. Dessa forma, a contribuição, de cada resíduo, íon ou grupo prostético, é obtida através da seguinte equação (KUMARI; KUMAR; LYNN, 2014):

$$\Delta R_x^{BE} = \sum_{i=1}^n (A_i^{livre} - A_i^{ligado}) \quad (58)$$

A ferramenta do *g_mmpbsa* retorna uma coloração vermelha para os grupos que contribuem favoravelmente a ligação, e azul para aqueles que contribuem desfavoravelmente.

Na Figura 28, o gráfico da contribuição individual de todos os complexos so-

Figura 28 – Contribuição individual de cada resíduo, íon ou grupo prostético na entalpia de todos os complexos. Os resíduos da lectina possuem uma baixa contribuição na energia livre. Íons cálcio da peroxidase têm uma grande contribuição, mesmo estando distantes da região de reconhecimento. A maior contribuição à energia livre favorável vem da manotriose.



Fonte: Autor, 2019.

brepostos mostra claramente que o grupo que mais contribui para a energia livre de ligação é a manotriose da lectina. No entanto, como essa ferramenta não incorpora o termo entrópico, o resultado pode ser enganoso.

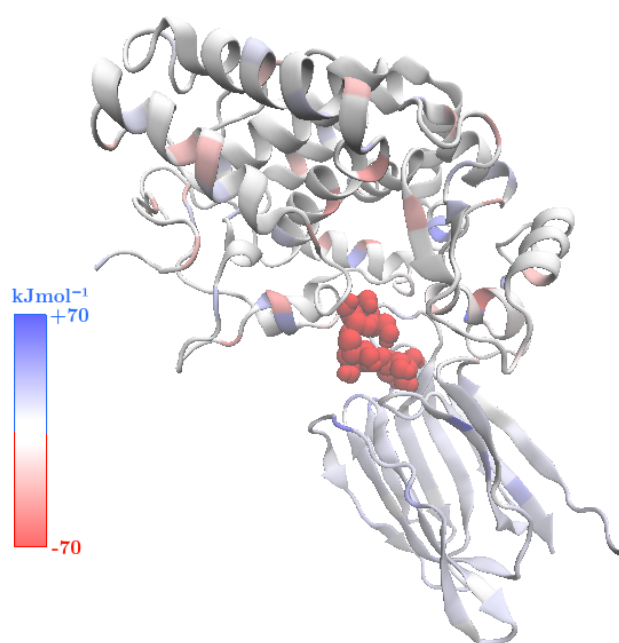
A maior contribuição na componente entálpica da energia livre é facilmente explicada porque a manotriose contém vários grupos hidroxila e cada um desses grupos é capaz de fazer 3 ligações de hidrogênio, sendo duas como acceptor e uma como doador. Como a peroxidase possui vários resíduos polares na região de ligação, pode-se concluir que a principal contribuição a energia livre de ligação se dá por ligações de hidrogênio e interações de dipolos permanentes dos grupos hidroxila.

As interações de dipolos e de van de Waals são muito mais fracas energeticamente, mas ocorrem em uma quantidade muito maior. Dessa forma, o reconhecimento molecular acaba sendo dominado por interações mais fracas que, quando somadas podem ter um peso, muito relevante.

Os grupos 308 e 309 aparentemente contribuem muito para a energia livre, mas são íons cálcio presentes na estrutura da peroxidase que se localizam a uma distância considerável da região de ligação. Portanto, é de se supor que essa contribuição é cancelada na hora que se soma o termo entrópico da energia livre. Sendo assim, a parte entálpica da energia livre não é suficiente para determinar, com confiabilidade, a contribuição de cada resíduo na energia livre total, servindo somente a uma análise qualitativa.

Como podemos ver na Figura 29, o resíduo em vermelho mais intenso e que mais contribui para a entalpia é a manotriose da lectina.

Figura 29 – Contribuição individual através decomposição de energia livre. Resíduos de cor vermelha contribuem de forma favorável a ligação e de cor azul de forma desfavorável.



Fonte: Autor, 2019.

Essa situação gera uma enorme flexibilidade das possíveis formas de ligação entre a lectina e a peroxidase, pois, como vemos, as diferenças de entalpia calculadas entre o complexo imobilizado semi-protonado e o sistema na condição inicial 3 devem-se ao fato de que a posição de ligação entre as duas proteínas mudou. No entanto, como o que realmente contribui para a energia livre é a manotriose, o módulo da energia livre de ligação altera-se muito pouco.

4.6 Momento de dipolo

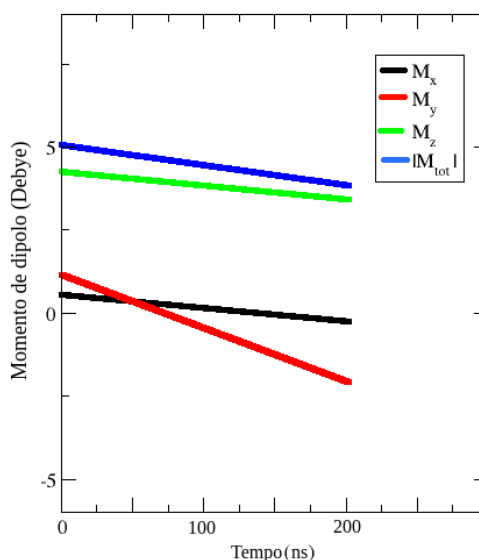
O software GROMACS possui a função chamada de `g_dipoles`. Essa ferramenta permite analisar os arquivos de trajetórias da dinâmica molecular e, a partir delas, os

momentos de dipolo das conformações de todos os átomos da molécula, retornando o valor total das componentes vetoriais e do módulo do vetor momento de dipolo.

Essa informação é bastante útil, pois, como sabemos, as interações de longa distância com dipolos são muito importantes para o reconhecimento molecular. Além disso, um momento de dipolo maior na molécula de manotriose pode explicar a razão do reconhecimento molecular ser favorecido nesta região.

No gráfico da Figura 30, podemos perceber que o módulo do momento de dipolo total (azul) da manotriose no sistema com condição inicial 3 está na faixa de 5 Debye - o que é mais que o dobro do momento de dipolo da molécula de água livre, que é de 1.85 Debye (BAKÓ; MAYER, 2016).

Figura 30 – Momento de dipolo da molécula de manotriose do sistema de condição inicial 3.

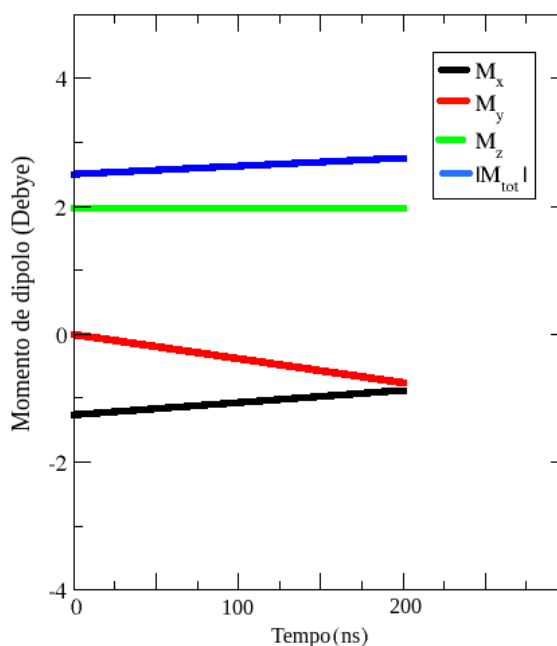


Fonte: Autor, 2019.

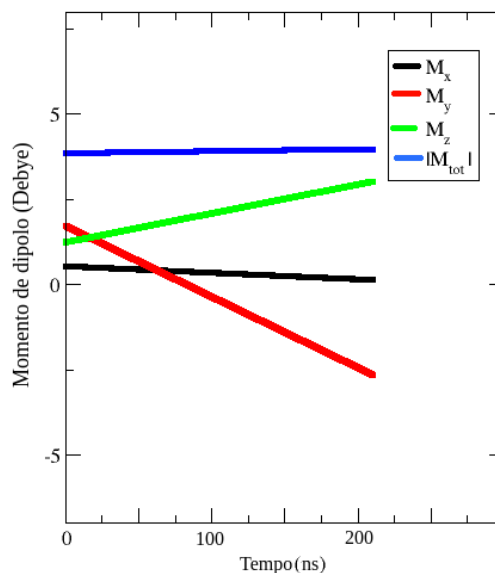
O momento de dipolo na direção z (verde) é dominante, tendo módulo praticamente igual ao momento total.

Quando analisamos o gráfico da Figura 31, percebemos uma redução considerável no módulo do momento de dipolo da molécula de manotriose, um valor aproximado de 3 Debye durante a dinâmica molecular. Essa queda no valor do momento de dipolo da manotriose é provocada por uma mudança na conformação da molécula e poderia ser a causa da diferença entre os valores calculados para a energia livre nos dois sistemas. Novamente, a componente z do momento de dipolo é maior em módulo.

Ao verificarmos o momento de dipolo da manotriose no sistema totalmente protonado, na Figura 32, percebemos que o valor retorna para a faixa de 5 Debye. No entanto, como vemos na Tabela 11, o valor em módulo da energia livre desse sistema

Figura 31 – Momento de dipolo da molécula de mantriose do sistema semi-protonado.

Fonte: Autor, 2019.

Figura 32 – Momento de dipolo da molécula de mantriose do sistema totalmente protonado.

Fonte: Autor, 2019.

foi menor do que o valor do sistema semi-protonado. Essa diferença de energia livre entre o sistema de condição inicial 3 e o sistema totalmente protonado, apesar de terem um momento de dipolo que oscila na faixa de 5 Debye, indica que ela não tem uma grande contribuição para a energia livre de ligação total do complexo estudado.

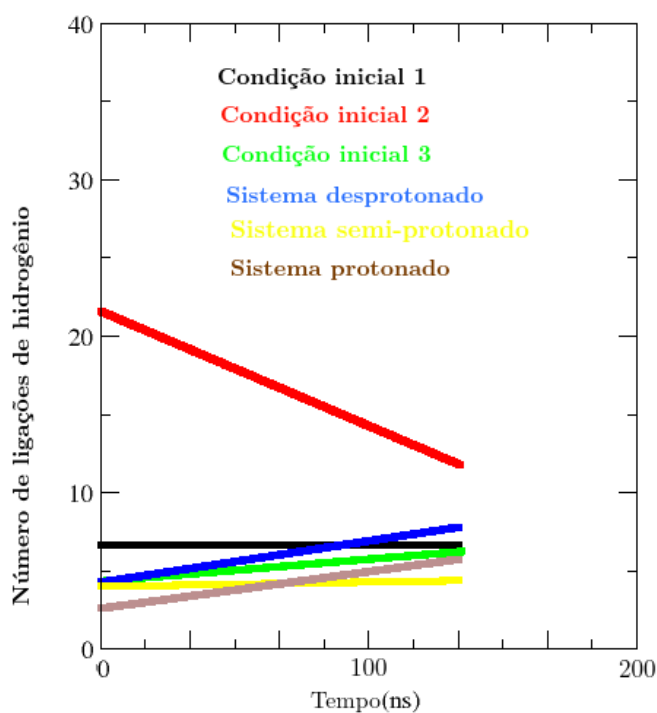
Em todos os casos analisados nos quais o complexo estava imobilizado na mo-

nocamada, a componente vetorial na direção z (verde) parece ter uma leve dominância sobre as demais. Este resultado pode ser explicado facilmente, dado que o sistema simula uma monocamada no plano xy e a mantriose é quase ortogonal a esse plano durante toda a dinâmica.

4.7 Tipos de interações intermoleculares

Ao usarmos a função hbond do GROMACS podemos calcular o número de ligações de hidrogênio entre lectina e peroxidase em função do tempo, como apresentado na Figura 33.

Figura 33 – Número de ligações de hidrogênio em função do tempo. O sistema com condição inicial 2 é capaz de fazer o maior número de ligações de hidrogênio em função do tempo.



Fonte: Autor, 2019.

Ao analisarmos o gráfico da Figura 33, percebemos que o sistema de condição inicial 2 (vermelho), parece fazer um maior número de ligações de hidrogênio entre a lectina e a peroxidase em função do tempo. O número, que começa na faixa de 20 ligações, decai até aproximar de 10 ligações. O número de ligações de hidrogênio é superior ao das dinâmicas moleculares dos outros complexos, mas existe uma tendência de que esse número se aproxime, em caso de simulações mais longas.

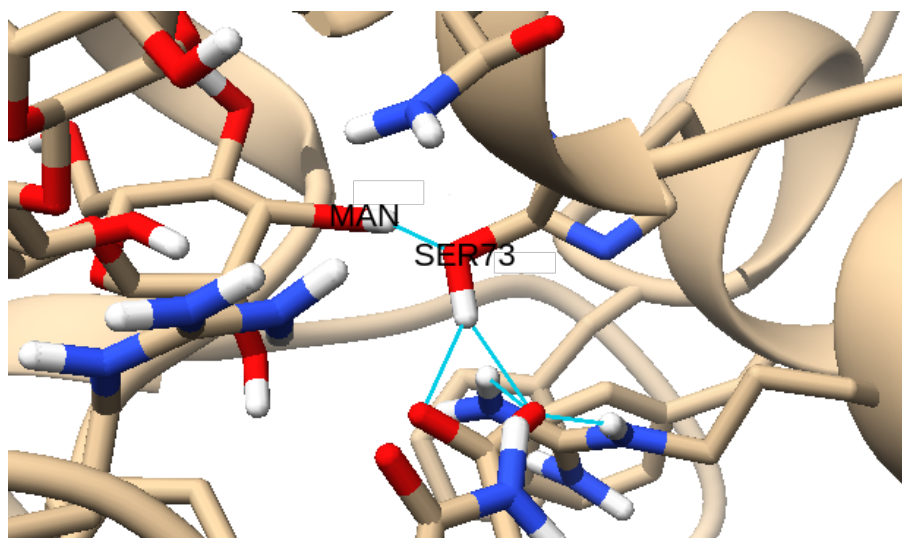
Esse maior número de ligações de hidrogênio não parece resultar em maior afinidade de ligação, pois a energia livre de ligação do sistema com condição inicial 2 é menor em módulo do que as energias do sistema com condição inicial 1 e do sistema

totalmente desprotonado. Esta diferença pode ser explicada no fato de que esse sistema utilizou um palpite inicial, baseado na segunda melhor predição do ZDOCK, enquanto as outras foram baseadas na primeira predição.

Essa estabilidade no número de ligações de hidrogênio em complexos, com diferenças consideráveis nos valores de energia, parece confirmar que a maior parte da contribuição para a energia livre de ligação entre a lectina e peroxidase ocorre por interações menos intensas, como as interações de dipolos e van der Waals, cujo grande número acaba favorecendo a interação entre as moléculas.

Ao visualizarmos o último *frame* do complexo imobilizado totalmente protonado, podemos ver uma dessas ligações de hidrogênio formadas entre a mantriose (MAN) e o resíduo de Serina 73 da peroxidase, conforme mostrado na Figura 34.

Figura 34 – Ligação de hidrogênio da mantriose na lectina com o resíduo de serina 73 da peroxidase.



Fonte: Autor, 2019.

Na mesma Figura, observa-se outras ligações de hidrogênio, algumas entre resíduos da peroxidase e resíduos da própria lectina, mostrando que, apesar da enorme contribuição, o reconhecimento molecular não ocorre apenas devido à mantriose da lectina.

As ligações de hidrogênio podem acontecer entre vários resíduos da região do sítio ativo da peroxidase e mudam a cada *frame* da dinâmica molecular. Portanto, a contribuição energética dessas ligações é mais uma média de todas as interações possíveis do que a contribuição da interação de dois resíduos.

As forças eletrostáticas mais intensas no sistema seriam as interações íon-íon, mas a análise estrutural do complexo nos mostra que apenas a molécula de peroxidase possui íons cálcio. Esses localizam-se muito distantes da região de ligação, e, dessa forma, o reconhecimento molecular só pode ocorrer através das interações entre

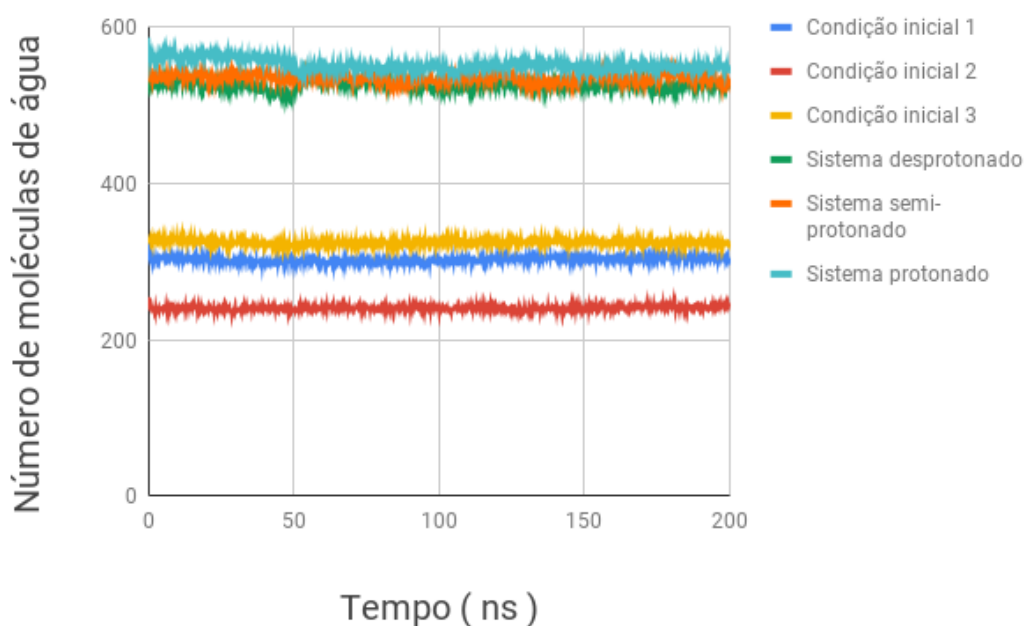
dipolos, ligações de hidrogênio e forças de van der Waals.

4.8 Número de moléculas de água ao redor do complexo

Uma dúvida existente nos biossensores eletroquímicos, e que pode ser esclarecida com a dinâmica molecular do sistema, o possível aumento da massa molecular aferida, provocado pela presença de moléculas de água próximas do complexo, atuando como águas estruturais.

Usando o programa RDF (Função de distribuição radial) do GROMACS, podemos fazer uma estimativa sobre o número de moléculas de água próximas do complexo durante a simulação, que poderia interferir no resultado do valor medido experimentalmente. O resultado do RDF nos mostra o número de moléculas de água próximas do complexo em função do tempo, conforme apresentado na Figura 35.

Figura 35 – Gráfico do número de moléculas de água próximas em função do tempo.



Fonte: Autor, 2019.

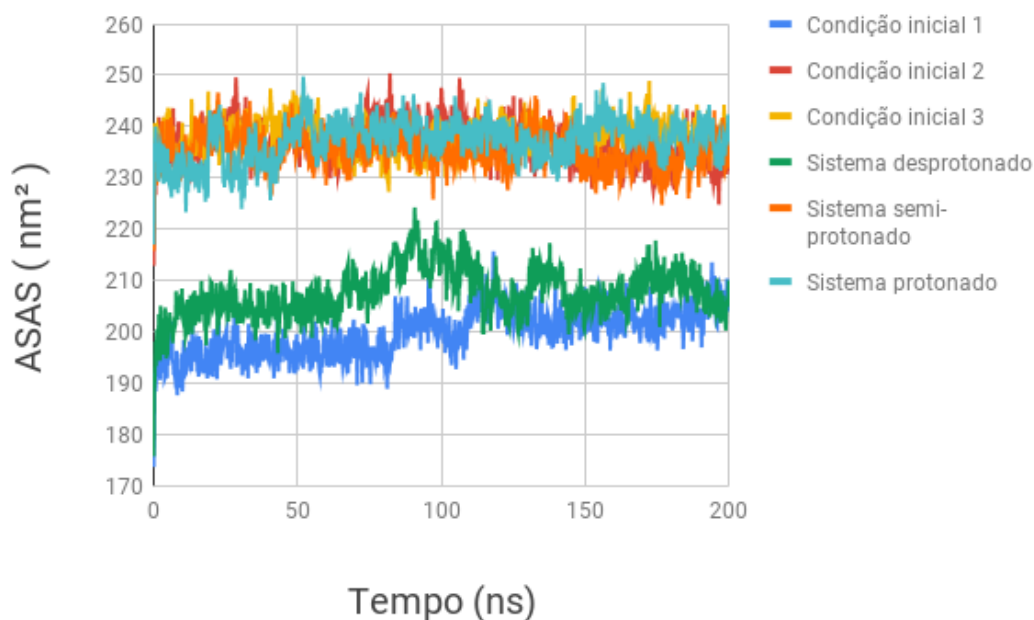
Ao analisarmos o gráfico, percebemos que nos complexos livres há aproximadamente 300 moléculas de água, e que, no caso no complexo imobilizado, esse número aumenta, superando 500 moléculas de água.

Como o biossensor eletroquímico detecta a adsorção, mais moléculas de água em volta do complexo poderiam explicar um módulo maior da energia livre, pois elas seriam contabilizadas.

Uma explicação para esse fato poderia ser o aumento da Área de Superfície Acessível ao Solvente (ASAS) provocada pela imobilização do complexo. Sendo assim,

uma análise o usando a ferramenta `g_sas` do GROMACS pode esclarecer essa dúvida, como podemos ver na Figura 36:

Figura 36 – Área de superfície acessível ao solvente dos complexos.



Fonte: Autor, 2019.

Ao verificarmos o gráfico na Figura 36, percebemos os complexos com condição inicial 1 e aqueles imobilizados no sistema totalmente desprotonado têm ASAS na faixa de 200 nm². Todos os outros complexos têm ASAS na faixa de 240 nm².

Dessa forma, podemos concluir que esse maior número de moléculas de água ao redor dos complexos imobilizados não resulta em maior acessibilidade ao solvente devido ao processo de imobilização.

Esse resultado do ASAS contradiz o resultado do RDF. Dessa forma, não podemos confiar totalmente nele e devemos buscar outros métodos que permitam verificar se realmente ocorre um aumento tão significativo de 200 moléculas por complexo (como indica o RDF), pois isso resultaria em um aumento considerável na constante de afinidade medida no biossensor.

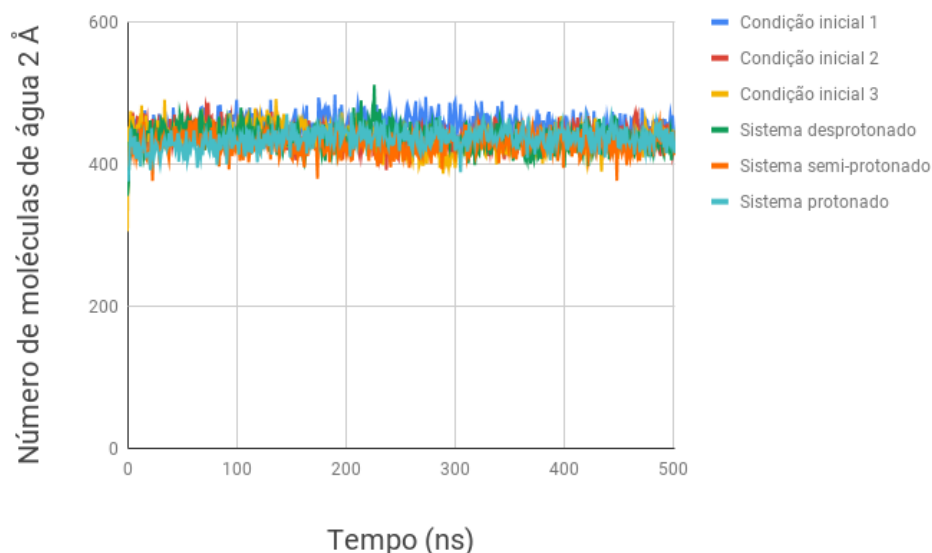
A análise da trajetória com a contagem manual das moléculas de água também contradiz o resultado do RDF. No entanto, analisar individualmente cada *frame* para contar o número de moléculas de água seria inviável. Dessa forma, foi desenvolvido um *script* em linguagem não compilada Tcl para ser executada no VMD, conforme apresentado na Figura 37.

Figura 37 – Script Tcl usado para contagem das moléculas de água.

```
set frames [molinfo top get numframes]
set outfile [open "numero.out" w]
set soma 0.0
set N 0
set numero_de_frames 0
for {set i $numero_de_frames} {$i < $frames} {incr i} {
  set a [atomelect top "(not resname MAN)&&(not resname HEM)&&not
protein &&(not resname UAK)&&(not resname UBK)&&same resid as within 2
of protein " frame $i]
  set num [$a num]
  set x [expr { $num/3 }]
  set soma [expr {$soma + $x}]
  puts "Frame: $i Numero: $x"
  puts $outfile "$i $x"
  $a delete
  set N [expr {$N + 1}]
}
set media [expr { $soma/$N }]
puts "Frames: $N Media: $media"
puts $outfile "Frames: $N Media: $media"
```

Fonte: Autor, 2019.

O *script* realiza a contagem exata do número de moléculas de água a 2 angström de distância do complexo em 500 *frames*, retornando um arquivo que permite gerar o gráfico da Figura 38.

Figura 38 – Número exato de moléculas de água a 2 angström de distância do complexo.

Fonte: Autor, 2019.

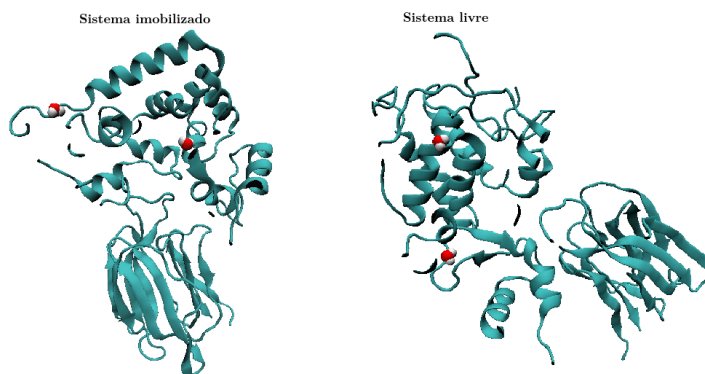
No algoritmo, qualquer átomo da molécula de água que estiver a uma distância menor que a distância mínima será contado. Dessa forma, o resultado não deve ser comparado diretamente ao método RDF, mas o método permite uma comparação muito precisa entre o complexo livre e imobilizado.

No gráfico da Figura 38, percebemos que o número exato de moléculas de água ao redor do complexo não muda significativamente com a imobilização, permanecendo sempre na faixa de 430 moléculas em todos os *frames*. Dessa forma, podemos concluir que a imobilização não teria como afetar a constante de afinidade medida pelo biossensor eletroquímico.

4.9 Moléculas de água estruturais

A maior interferência possível na massa da proteína detectada aconteceria no caso de moléculas de água estruturais presentes na proteína individual da peroxidase ou incorporadas na formação do complexo, ou seja, moléculas de água aprisionadas na cavidade formada entre as duas proteínas, como vemos na Figura 39.

Figura 39 – Moléculas de água estruturais no complexo simulado no sistema com condição inicial 1 e no sistema totalmente protonado.



Fonte: Autor, 2019.

Quando verificamos, na Figura 39, um plano de corte da peroxidase nos arquivos de trajetória das dinâmicas moleculares do sistema de condição inicial 1 e imobilizado totalmente protonado, podemos localizar moléculas de água que permanecem, uma boa parte do tempo da simulação, próximas a peroxidase, como se fizessem parte da estrutura da proteína. Mesmo essas moléculas não conseguem ficar durante todo o tempo de simulação ligados a estrutura.

Devemos lembrar que o modelo de molécula de água usado da dinâmica molecular é mais simplificado e não considera todas as interações que uma molécula de água faria com seus dois pares de elétrons não ligados.

Essas moléculas podem ser responsáveis por aumentar a massa da proteína medida experimentalmente no biossensor. No entanto, ainda assim, a quantidade de moléculas de água é desprezível em relação à massa total da proteína, não afetando significativamente a precisão da medida. Além disso a região de contato entre as duas proteínas no modelo teórico parece ser muito pequena, não favorecendo o aprisionamento de águas estruturais.

5 Conclusão

O projeto foi bem-sucedido em tentar reproduzir o meio experimental existente no biossensor eletroquímico, usando as condições periódicas de contorno da caixa de dinâmica molecular para simular uma monocamada infinita, com moléculas de lectina imobilizadas covalentemente, semelhante à situação experimental, guardadas as proporções entre o tamanho das moléculas e do transdutor físico-químico.

O modelo de *docking* molecular de corpo rígido usado para gerar as condições iniciais sobre a conformação entre as proteínas gerou bons resultados, o que reduz a necessidade de dinâmicas moleculares muito extensas. No entanto, quando somados os tempos de simulação do sistema nas mais variadas condições superamos 1 microssegundo, o que é um tempo de simulação alto.

O método de parametrização das moléculas com o uso das cargas de Mulliken em conjunto com base def2-SVP parece ser adequado. Usando as parametrizações de moléculas, foi possível reproduzir até mesmo as conformações da molécula de manotriose em solução. Devido à componente eletrostática ser apenas um dos componentes no cálculo da energia livre, o resultado não parece ser muito sensível ao método usado para obter as cargas da parametrização.

Todos os cálculos de energia livre realizados parecem confirmar valores experimentais, nos quais metade das simulações retornou valores muito próximos, mostrando a eficiência do método MM-PBSA em conjunto com o método de modos normais para o cálculo da componente entrópica da energia livre. O método de modos normais parece funcionar muito bem em proteínas globulares, onde o reconhecimento não altera muito sua estrutura.

Os cálculos de decomposição de energia demonstram que a principal responsável pela energia livre de ligação é a molécula de manotriose da lectina, que, apesar de não ser covalentemente ligada, acaba exercendo, quase, uma função de âncora que se liga à molécula de peroxidase.

O cálculo do MM-PBSA, como podemos imaginar, não é perfeito, estando sujeito a flutuações devido à limitação da amostragem em função do tempo. Segundo Salinas (1997, p. 24), na teoria dos erros, o desvio relativo deveria decair na ordem de $N^{-\frac{1}{2}}$, onde N é o número de *frames*. Neste caso, esbarramos nos limites computacionais para calcular a energia de um sistema extremamente dinâmico. Os melhores computadores atualmente conseguem simular apenas microssegundos. Dessa forma, devemos sempre enfatizar que o resultado é uma média da energia livre no intervalo simulado.

Os métodos empregados para estimar a componente entrópica da energia livre demonstraram ter muitas limitações. A estrutura desloca demais seus átomos para

que se use a aproximação quase-harmônica e estimar a entropia vibracional. Já o método de aproximação de modos normais demonstra funcionar de forma adequada na estrutura das duas proteínas, mas devemos lembrar que essa estimativa considera um único *frame*. Em dinâmicas moleculares mais longas haveria possibilidade de mudanças conformacionais que afetam a estimativa de modos normais.

A análise do tipo de interações de longa distância mostra que, apesar das ligações de hidrogênio serem mais intensas, a maior contribuição para a energia livre de ligação vem de interações de curto alcance. A maior contribuição para a energia livre vem de interações de dipolos e forças de van der Waals, com destaque para essa última, devido à maior quantidade dessas interações.

A simulação do sistema demonstra que o processo de imobilização das proteínas realmente afeta a entalpia e a entropia, diminuindo o módulo dessas grandezas. No entanto, parece afetar pouco a energia livre de ligação.

A simulação demonstra que o processo de imobilização não altera o número de moléculas de água nas proximidades do complexo e, como a região de ligação é pequena em comparação ao tamanho das duas proteínas, não ocorre aprisionamento de moléculas de água, que se tornariam moléculas de água estruturais e afetariam a constante de afinidade medida no biossensor.

O número de moléculas de água estruturais é pequeno, sendo desprezível em comparação com a massa total da proteína. A previsão teórica da energia livre de ligação de moléculas imobilizadas no biossensor eletroquímico mostra-se eficiente, e esse conhecimento poderia ser aplicado em outras moléculas de interesse para ajudar no desenvolvimento de outros biossensores baseados nessa tecnologia.

Referências

- ALCÁCER, L. **Introdução à química quântica computacional**. 2. ed. Lisboa: IST-Instituto Superior Técnico, 2007. 351 p.
- ALTMAN, D. G.; BLAND, J. M. Standard deviations and standard errors. **Bmj**, [s.l.], v. 331, n. 7521, p. 903-903, 13 oct. 2005.
- AMADEI, A.; LINSSEN, A. B. M.; BERENDSEN, H. J. C. Essential dynamics of proteins. **Proteins: Structure, Function, and Genetics**, [s.l.], v. 17, n. 4, p. 412-425, dec. 1993.
- BAKÓ, I.; MAYER, I. On dipole moments and hydrogen bond identification in water clusters. **The Journal of Physical Chemistry A**, [s.l.], v. 120, n. 25, p. 4408-4417, 17 june 2016.
- BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The journal of Chemical Physics**, [s.l.], v. 81, n. 8, p. 3684-3690, 15 oct. 1984.
- CAMPANERA, J. M.; POUPLANA, R. MMPBSA decomposition of the binding energy throughout a molecular dynamics simulation of amyloid-beta (A β 1035) aggregation. **Molecules**, [s.l.], v. 15, n. 4, p. 2730-2748, 15 apr. 2010.
- CARRINGTON, B. J.; MANCERA, R. L. Comparative estimation of vibrational entropy changes in proteins through normal modes analysis. **Journal of Molecular Graphics and Modelling**, [s.l.], v. 23, n. 2, p. 167-174, oct. 2004.
- CARVALHO, F. et al. Evaluating the equilibrium association constant between artinM lectin and myeloid leukemia cells by impedimetric and piezoelectric label Free Approaches. **Biosensors**, [s.l.], v. 4, n. 4, p. 358-369, 3 oct. 2014.
- CHATTERJI, D. **Basics of molecular recognition**. [s.l.]: CRC Press, 2016. 141 p.
- COELHO, L. C. B. B. et al. Lectins, interconnecting proteins with biotechnological/pharmacological and therapeutic applications. **Evidence-based Complementary and Alternative Medicine**, [s.l.], v. 2017, p. 1-22, 2017.
- COURNIA, Z.; ALLEN, B.; SHERMAN, W. Relative binding free energy calculations in drug discovery: recent advances and practical considerations. **Journal Of Chemical Information And Modeling**, [s.l.], v. 57, n. 12, p. 2911-2937, 15 dec. 2017.
- CUI, Q.; BAHAR, I. **Normal mode analysis: theory and applications to biological and chemical systems**. New York: Chapman and Hall/CRC, 2005. 432 p.

- GIMÉNEZ-ROMERO, D. et al. Elucidation of carbohydrate molecular interaction mechanism of recombinant and native artinM. **The Journal of Physical Chemistry B**, [s.l.], v. 117, n. 28, p. 8360-8369, 9 july 2013.
- FRENKEL, D.; SMIT, B. **Understanding molecular simulation**: from algorithms to applications. [s.i.]: Academic Press, 1996. 638 p.
- HENRIKSEN, A. et al. Structural interactions between horseradish peroxidase C and the substrate benzhydroxamic acid determined by X-ray crystallography. **Biochemistry**, [s.l.], v. 37, n. 22, p. 8054-8060, june 1998.
- HUMER, D.; SPADIUT, O. Improving the performance of horseradish peroxidase by site-directed mutagenesis. **International Journal Of Molecular Sciences**, [s.l.], v. 20, n. 4, p. 916-930, 20 feb. 2019.
- IGNATENKO, O.V. et al. Electrochemistry of chemically trapped dimeric and monomeric recombinant horseradish peroxidase. **Advances In Biosensors and Bioelectronics (ABB)**, [s.l.], v. 2, p. 25-34, sep. 2013.
- JEYAPRAKASH, A.A. et al. Structural basis for the carbohydrate specificities of artocarpin: variation in the length of a Loop as a strategy for generating ligand specificity. **Journal of Molecular Biology**, [s.l.], v. 338, n. 4, p. 757-770, may 2004.
- KATCHALSKI-KATZIR, E. et al. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques.. **Proceedings of the National Academy of Sciences**, [s.l.], v. 89, n. 6, p. 2195-2199, 15 mar. 1992.
- KUKOL, A. **Molecular modeling of proteins**. 2nd ed. Hatfield: Humana Press, 2015. 474 p.
- KUMARI, R.; KUMAR, R. ; LYNN, A. G_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. **Journal of Chemical Information and Modeling**, [s.l.], v. 54, n. 7, p. 1951-1962, 28 july 2014.
- LEMKUL, J. A.; ALLEN, W. J.; BEVAN, D. R. Practical considerations for building GROMOS-compatible small-molecule topologies. **Journal of Chemical Information and Modeling**, [s.l.], v. 50, n. 12, p. 2221-2235, 27 dec. 2010.
- NUSSENZVEIG, H. M. **Curso de Física Básica: Eletromagnetismo**. São Paulo: Edgard Blücher Ltda, 1997. 295 p.
- MULLIKEN, R. S. Electronic population analysis on LCAO–MO molecular wave functions. **The Journal Of Chemical Physics**, [s.l.], v. 23, n. 10, p. 1833-1840, oct. 1955.

- NELSON, D. L.; COX, M. M. **Princípios de bioquímica de Lehninger**. 7. ed. Porto Alegre: Artmed, 2018. 1312 p.
- NEESE, F. The ORCA program system. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, [s.l.], v. 2, n. 1, p. 73-78, 28 june 2011.
- NUMATA, J.; WAN, M. ; KNAPP, E.W. Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. **Genome Informatics 2007**, [s.l.], p. 192-205, dec. 2007.
- PADHORN, D. et al. Protein–protein docking by fast generalized Fourier transforms on 5D rotational manifolds. **Proceedings of the National Academy of Sciences**, [s.l.], v. 113, n. 30, p. 4286-4293, 13 july 2016.
- PANCZYK, K.; PLAZINSKI, W. Pyranose ring puckering in aldopentoses, ketohexoses and deoxyaldohexoses. A molecular dynamics study. **Carbohydrate Research**, [s.l.], v. 455, p. 62-70, jan. 2018.
- PESQUERO, N.C. et al. Real-time monitoring and kinetic parameter estimation of the affinity interaction of jArtinM and rArtinM with peroxidase glycoprotein by the electrogravimetric technique. **Biosensors And Bioelectronics**, [s.l.], v. 26, n. 1, p. 36-42, 15 sep. 2010.
- PIERCE, B. G. et al. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. **Bioinformatics**, [s.l.], v. 30, n. 12, p. 1771-1773, 14 feb. 2014.
- PLIEGO, J. R.; RIVEROS, J. M. Hybrid discrete-continuum solvation methods. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, [s.l.], p. 1-25, 4 ago. 2019.
- POL-FACHIN, L. et al. GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. **Journal of Chemical Theory and Computation**, [s.l.], v. 8, n. 11, p. 4681-4690, 18 sep. 2012.
- PU, Q. et al. Molecular simulations of stretching gold nanowires in solvents. **Nanotechnology**, [s.l.], v. 18, n. 42, p. 424007-424012, 13 sep. 2007.
- SALINAS, S. R. A. **Introdução à física estatística**. São Paulo: Edusp, 1997. 466 p.
- SANTOS, A. et al. Impedance-derived electrochemical capacitance spectroscopy for the evaluation of lectin–glycoprotein binding affinity. **Biosensors and Bioelectronics**, [s.l.], v. 62, p. 102-105, dec. 2014.
- SCHLITTER, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. **Chemical Physics Letters**, [s.l.], v. 215, n. 6, p. 617-621, dec. 1993.

- SEARS, F. W; SALINGER, G. L. **Termodinâmica**: teoria cinética e termodinâmica estatística. 3 ed. Rio de Janeiro: Guanabara Dois, 1979. 402 p.
- SEEBER, M. et al. Wordom: a program for efficient analysis of molecular dynamics simulations. **Bioinformatics**, [s.l.], v. 23, n. 19, p. 2625-2627, 23 ago. 2007.
- SOARES, T. A.; FERREIRA, R. Aplicação da equação de Poisson-Boltzmann ao cálculo de propriedades dependentes do pH em proteínas. **Química Nova**, [s.l.], v. 27, n. 4, p. 640-647, ago. 2004.
- SKJAERVEN, L. ; MARTINEZ, A. ; REUTER, N. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. **Proteins: Structure, Function, and Bioinformatics**, [s.l.], v. 79, n. 1, p. 232-243, 5 nov. 2010.
- TU, Q. et al. Rotating carbon nanotube membrane filter for water desalination. **Scientific Reports**, [s.l.], v. 6, n. 1, p. 1-11, may 2016.
- UL-HAQ, Z.; MADURA, J. D. **Frontiers in computational chemistry**. Shāriqah: Bentham Science Publishers, 2015. v.1, 362 p.
- VAN AALTEN, D. M. F. et al. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. **Journal Of Computer-Aided Molecular Design**, [s.l.], v. 10, n. 3, p. 255-262, june. 1996.
- VERLET, L. C. **Physical Review**, [s.l.], v. 159, n. 1, p. 98-103, 5 july 1967.
- WANG, C. et al. Recent developments and applications of the MMPBSA method. **Frontiers In Molecular Biosciences**, [s.l.], v. 4, p. 1-18, 10 jan. 2018.
- WELLS, B. A.; CHAFFEE, A. L. Ewald summation for molecular simulations. **Journal of Chemical Theory and Computation**, [s.l.], v. 11, n. 8, p. 3684-3695, 3 ago. 2015.
- WITTE, J.; NEATON, J. B.; HEAD-GORDON, M. Push it to the limit: characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory. **The Journal Of Chemical Physics**, [s.l.], v. 144, n. 19, p. 194306-194318, 18 may 2016.
- YANG, L. et al. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. **Bioinformatics**, [s.l.], v. 25, n. 5, p. 606-614, 15 jan. 2009.