

## RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta dissertação será disponibilizado somente a partir de 06/08/2020.



**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: <b>A</b> : Bastão <b>B</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 $\text{CH}_2$ 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 <b>A</b> : Na condição inicial 1 a extremidade da manotriose é deslocada do sítio de ligação original na alça da lectina. <b>B</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 <b>A</b> : No sistema totalmente protonado a extremidade da manotriose se mantém no sítio de ligação original na alça da lectina. <b>B</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

$\alpha$	Letra grega minúscula Alfa
$\beta$	Letra grega minúscula Beta
$\gamma$	Letra grega minúscula GAMA
$\Delta$	Letra grega maiúscula Delta
$\nu$	Letra grega minúscula Ípsilon
$\lambda$	Letra grega minúscula Lambda
$\mu$	Letra grega minúscula Mi
$\rho$	Letra grega minúscula Rho
$\phi$	Letra grega minúscula Phi
$\chi$	Letra grega minúscula Chi
$\Psi$	Letra grega maiúscula Psi
$\nabla$	Símbolo grego Nabla

# Sumário

<b>1</b>	<b>Introdução</b>	<b>14</b>
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
<b>2</b>	<b>Objetivos</b>	<b>30</b>
<b>3</b>	<b>Metodologia</b>	<b>31</b>
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
<b>4</b>	<b>Resultados</b>	<b>45</b>
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
<b>5</b>	<b>Conclusão . . . . .</b>	<b>72</b>
	 <b>Referências . . . . .</b>	 <b>74</b>



# 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):

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

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:

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

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:

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

## 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  $\tau$  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  $\lambda_t$ :

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

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  $\chi$  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  $\psi_0$ .

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

Isso significa que ao variarmos os parâmetros da função tentativa  $\chi$ , 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  $\psi_a$  são os orbitais, de forma parecida com a equação de Schrödinger:

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

Onde o orbital  $\psi_a$  é uma combinação linear de orbitais atômicos  $\chi_a$ :

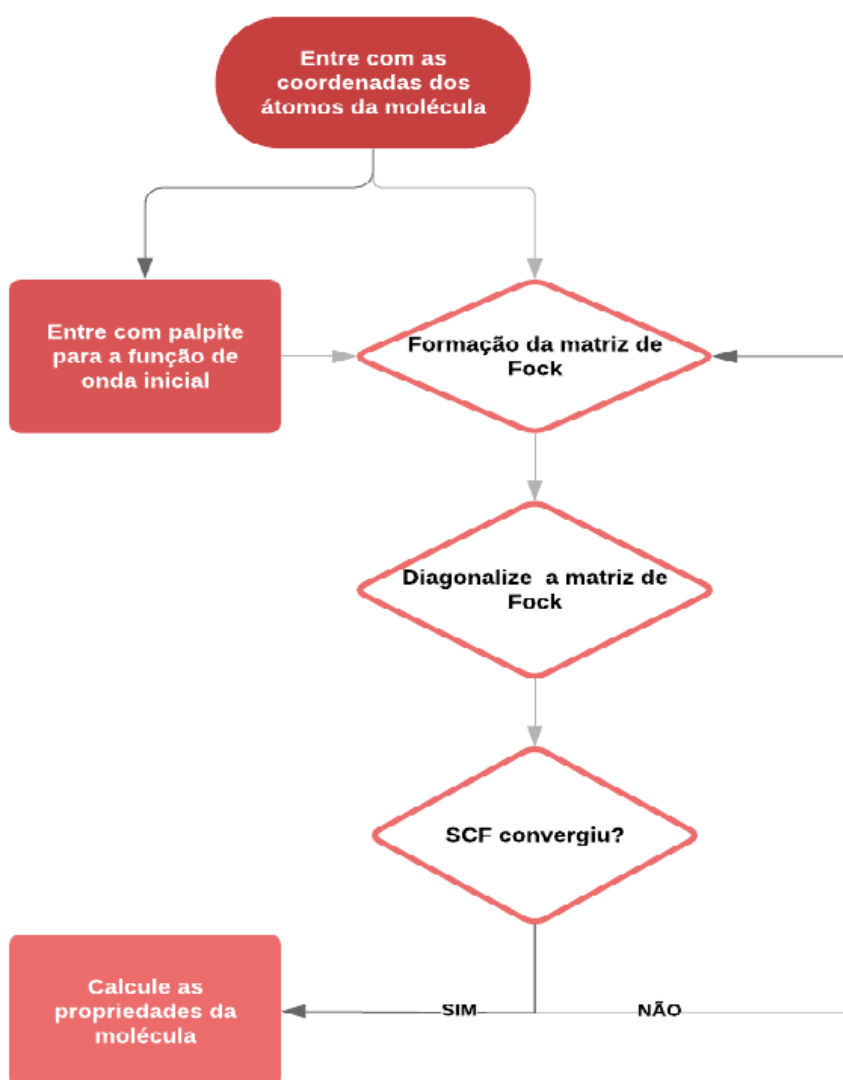
$$\psi_a = \sum_p^L C_p^a \chi_a \quad (20)$$

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  $\phi_p$ , 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 partilhados 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,  $\Delta G$  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  $\Delta G_{LP}$  as energias livres da lectina  $\Delta G_L$  e da peroxidase  $\Delta G_P$ , 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  $\Delta G_i$  é a soma das energias mecânicas  $E_{mm}$  do campo de força usado da dinâmica molecular com a energia livre de solvatação  $\Delta G_{solv}$  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  $\phi_r$  com a função distribuição de densidade de carga  $\rho_r$  no meio com constante dielétrica  $\epsilon_r$  (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;  $\phi$  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  $\lambda_r$  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  $\sigma$  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  $\Delta R_t$  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  $\lambda_{ij}$  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).

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

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  $\lambda$  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  $\lambda$  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.



## 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 $\beta$ 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.