

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA - CURSO DE ENGENHARIA MECÂNICA
TRABALHO DE CONCLUSÃO DE CURSO

ANÁLISE ESTRUTURAL DE PROTEÍNAS UTILIZANDO ELEMENTOS FINITOS

por

João Marcos Regauer

Monografia apresentada ao Departamento de Engenharia Mecânica da Escola de Engenharia da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do diploma de Engenheiro Mecânico.

Porto Alegre, maio de 2021

CIP - Catalogação na Publicação

Regauer, João Marcos
ANÁLISE ESTRUTURAL DE PROTEÍNAS UTILIZANDO
ELEMENTOS FINITOS / João Marcos Regauer. -- 2021.
28 f.
Orientador: Ignacio Iturrioz.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Escola de
Engenharia, Curso de Engenharia Mecânica, Porto
Alegre, BR-RS, 2021.

1. estrutura de proteínas. 2. análise modal. 3.
flexibilidade. 4. lisozima. I. Iturrioz, Ignacio,
orient. II. Título.

João Marcos Regauer

ANÁLISE ESTRUTURAL DE PROTEÍNAS UTILIZANDO ELEMENTOS FINITOS

ESTA MONOGRAFIA FOI JULGADA ADEQUADA COMO PARTE DOS
REQUISITOS PARA A OBTENÇÃO DO TÍTULO DE
ENGENHEIRO MECÂNICO
APROVADA EM SUA FORMA FINAL PELA BANCA EXAMINADORA DO
CURSO DE ENGENHARIA MECÂNICA

Prof. Mário Roland Sobczyk Sobrinho
Coordenador(a) do Curso de Engenharia Mecânica

Área de concentração: Mecânica dos Sólidos

Orientador: Prof. Ignacio Iturrioz

Comissão de Avaliação:

Prof. Ignacio Iturrioz

Prof. Jakson Manfredini Vassoler

Prof. Rodrigo Rossi

Porto Alegre, maio de 2021

*everything that living things do
can be understood in terms of the jiggling and wiggling of atoms.
Feynman, Richard (1963). "Lectures on Physics"*

ANÁLISE ESTRUTURAL DE PROTEÍNAS UTILIZANDO ELEMENTOS FINITOS

Regauer, João Marcos
joaoregauer@gmail.com

Resumo. Quando a mecânica dos sólidos se depara com estruturas orgânicas e com dimensões próximas de ligações químicas, as hipóteses da mecânica do contínuo não se aplicam e a interação com diferentes áreas de conhecimento é inevitável para compreender seu comportamento estrutural. As proteínas podem ser compreendidas como estruturas dinâmicas, sendo sua flexibilidade e sua vibração fundamentais para compreender seus complexos mecanismos de ação e atividade biológica. Este trabalho busca avaliar a estrutura de uma proteína, construída através de suas ligações químicas de forma semelhante a uma estrutura unifilar, utilizando simulações de elementos finitos e rotinas de otimização para verificar a viabilidade das técnicas de mecânica dos sólidos. São realizadas simulações na proteína Lisozima em três diferentes estados de conformação (códigos do PDB: 1DPX, 1DPW e 4YM8) onde a análise modal é realizada para investigar a dinâmica de cada estado. Os resultados são comparados utilizando a correlação de Pearson entre os resultados da simulação e o fator de temperatura, obtido experimentalmente para cada átomo. A aplicação de ferramentas de otimização utilizando o fator de temperatura para definir a rigidez dos elementos aumentou a correlação com dados experimentais. O método trás aceitável correlação com baixo custo computacional, havendo margem para melhorias e trabalhos futuros, indicando a viabilidade desta técnica.

Palavras-chave: estrutura de proteínas, análise modal, flexibilidade, lisozima

Structural analysis in proteins using finite elements.

Abstract. When the mechanics of solids are faced with organic structures and with dimensions close to chemical bonds, the hypotheses of the mechanics of the continuum do not apply and interaction with different areas of knowledge is inevitable to understand their structural behavior. Proteins can be understood as dynamic structures, their flexibility and vibration being fundamental to understand their complex mechanisms of action and biological activity. This work seeks to evaluate the structure of a protein, built through its chemical bonds in a similar way to a truss structure, using finite element simulations and optimization routines to verify the feasibility of using solid mechanics techniques. Simulations are carried out on the lysozyme protein in three different conformation states (PDB codes: 1DPX, 1DPW and 4YM8) where the modal analysis is performed to investigate the dynamics of each state. The results are compared using Pearson's correlation between the results of the simulation and the temperature factor, obtained experimentally for each atom. The application of optimization tools using the temperature factor to define the stiffness of the elements increased the correlation with experimental data. The method brings an acceptable correlation with low computational cost, with room for improvement and future work, indicating the feasibility of this technique.

Keywords: protein structure, modal analysis, flexibility, lysozyme

SUMÁRIO

1. INTRODUÇÃO	1
Revisão bibliográfica.	1
Objetivo Geral.	2
Objetivo Específico.	2
2. FUNDAMENTAÇÃO	2
Formação e estruturas de proteínas.	2
Características experimentais e bancos de dados de proteínas.	4
Método de busca de raízes	5
Parâmetros Calculados	5
3. METODOLOGIA	6
4. RESULTADOS	9
4.1 Análise dos procedimentos de otimização.	9
4.2 Fatores de temperatura experimental e simulados.	10
4.2.1 Estruturas primárias.	10
4.2.2 Estruturas primárias com adição de ligações secundárias.	12
4.3 Resultados do coeficiente de Pearson entre experimental e resultados	13
4.4 Avaliação dos modos de vibração.	14
5. CONCLUSÃO	15
REFERÊNCIAS BIBLIOGRÁFICAS	16
APÊNDICES	
A: Parâmetros utilizados nos átomos e elementos.	18
B: Figuras dos procedimentos de otimização.	18
C: Figuras dos fatores de temperatura experimental e simulados para ligações primárias e primárias otimizadas para 1DPX e 1DPW.	20
D: Figuras dos fatores de temperatura experimental e simulados utilizando ligações de hidrogênio e otimizadas para 1DPX e 1DPW.	21
E: Figuras dos Modos de vibração.	22

1. INTRODUÇÃO

Existem evidências de que o comportamento mecânico das proteínas desempenha um papel importante nos processos biológicos dessa estrutura molecular, desde o enovelamento que ocorre em todas as proteínas até processos degenerativos e na capacidade de transmissão de um vírus. O enovelamento da proteína, que é crucial para sua função biológica, geralmente está associado às frequências mais baixas de vibrações, da ordem de 1 TeraHertz. Os estados de conformação das proteínas não são estáticos, variando com parâmetros como o ambiente externo, pH e temperatura.

A atividade biológica das proteínas ocorre de forma dinâmica, sendo sua flexibilidade e sua vibração fundamentais para compreender seus complexos mecanismos de ação. Assim, a dinâmica das estruturas geralmente são investigadas tanto em forma experimental como por técnicas numéricas. Geralmente, a técnica numérica mais utilizada é a chamada dinâmica molecular (DM) (ALBERTS B. et al, 2009), que estuda o movimento dos átomos e moléculas através das equações que resultam em considerar todas as forças de interação atômicas existentes no sistema. Porém, este método exige grande capacidade computacional, sendo ainda impraticável em alguns cenários. Os movimentos deste tipo de estrutura associados aos modos de vibração mais altos são os mais relevantes do ponto de vista biológico e eles podem ser simulados com esquemas simplificados como o *Coarse-Grained* (CG) (SCARAMOZZINO, D., et al, 2019) e *All-Atom* (AA) (CARPINTERI, A, et al, 2018) que são modelos semelhantes a uma estrutura formada por elementos finitos unifilares que une os átomos em interação. Diversas alternativas existem para determinar as rigidezes das vinculações entre átomos de forma a representar o comportamento do conjunto. Métodos simplificados como estes mostram grande potencial no estudo de estruturas biológicas complexas, principalmente quando combinados com dados experimentais utilizados para sua calibração como comentado em Ma J. (2005). Muitas técnicas experimentais podem se aliar aos modelos em elementos finitos e Dinâmica Molecular (MD), como as técnicas de Espectroscopia Raman (LACIDOGNA, G., et al 2017), Ressonância magnética nuclear (MITTERMAIER, A; KAY, L.E., 2006), *time-resolved crystallography* (HAJDU, J. et al, 2000) e *fluorescence resonance energy transfer* (BAE W. et al, 2013).

A primeira proteína a ser revelada pela cristalografia de raio-x foi a lisozima, por Blake et al. (1965). Segundo Goodsell, D. (2000), por ser uma enzima pequena e estável, ela é ideal para pesquisas sobre a estrutura e função das proteínas. Sendo a proteína mais comum no RCSB Protein Data Bank (PDB) (BERMAN H.M. et al, 2000). Ela é usada no corpo humano para defesa contra infecções, a enzima quebra as cadeias de açúcar nas paredes das células bacterianas, destruindo a integridade estrutural da bactéria, o que causa o rompimento das bactérias. É usado e pode ser encontrado naturalmente na clara do ovo, nas lágrimas humanas e no muco, nesses locais permite que os humanos resistam à infecção em superfícies expostas ao ambiente externo.

Revisão bibliográfica: Existem diversas alternativas na bibliografia que procuram representar as estruturas das proteínas em forma simplificada, dentre elas pode-se citar: Os modelos Carpinteri, et al. (2017) que utilizou uma metodologia semelhante ao deste estudo, também aplicada a proteína 4YM8, porém neste trabalho foram adotadas simplificações na área de seção dos elementos e na definição da massa dos átomos, sendo estas iguais à média de todos os átomos. Este procedimento reduz a complexidade e o custo computacional em detrimento da perda de detalhes locais, sendo adequado para avaliar o comportamento global da proteína.

Giordani, G et al. (2021) realizou uma análise modal para investigar as propriedades dinâmicas da proteína lisozima (código PDB: 4YM8) utilizando os métodos *Coarse-Grained* (CG) e *All-Atom* (AA), realizando uma avaliação do parâmetro de *cut-off*, distribuição de

frequências, modos de vibração e flexibilidade em ambos modelos. Os resultados indicaram uma correlação aceitável para flexibilidade global do método.

Davide Bochicchio et al. (2017) sugere uma modelagem para estruturas poliméricas longas que se comunicam com estruturas moleculares mais pequenas chamadas monômeros. As ligações entre as partes neste caso são mais fracas. Comparação entre os valores simulados com esta proposta e com parâmetros experimentais medidos não só permite avaliar a consistência do modelo como entender de que dependem estas interações e este último é fundamental para avançar na nossa compreensão de como projetar materiais bioinspirados com propriedades dinâmicas controláveis.

Hu & Buehler (2020) analisa os modos de vibração de diferentes linhagem e mutações das proteínas spike de *coronavirus-2* causadores de síndrome respiratória aguda grave (SARS-CoV-2). O modelo para simulação representa átomos como massas pontuais e ligações químicas como molas em um modelo CG. São realizadas análises dos modos e da flexibilidade, principalmente na região de ligação desta proteína com o receptor para infecção. Os resultados da flexibilidade desta região mostram uma correlação positiva com a taxa de letalidade e relação inversa com a infectividade do vírus.

Escobedo, A. Et al. (2019) estuda uma hélice de poliglutamina, que causa a doença atrofia muscular espinobulbar. Utiliza MD em uma simulação de movimento, ao longo de 5 μ s da movimentação da estrutura em passos de 4 femtosegundos (10^{-15} segundos), para compreender a dinâmica e a influência de uma ligação secundária na estabilização de uma α hélice que está correlacionada a doença de atrofia muscular espinobulbar. Também são utilizados outros métodos experimentais e numéricos para determinação das ligações e da estrutura. Os resultados sugerem explicações para mudanças que causam a doença e a relevância das ligações secundárias na estabilização de estruturas.

Nesta breve revisão, que pode ser considerada incomum no contexto de um TCC na Engenharia Mecânica, se pretende ilustrar o tipo de problemas que podem ser resolvidos relacionados com a biologia desde a perspectiva da Mecânica. Em todos os trabalhos vistos, é uma constante que a fisiologia dos sistemas biológicos vistos está intimamente relacionado com a mecânica de suas estruturas e isso nos leva a compreender que nosso papel como futuros engenheiros não está só em otimizar o processo para construção de máscaras ou na criação de válvulas mais efetivas em equipamentos médicos, problemas estes que são de singular importância. Mas também dar-se conta que conhecimentos sobre estruturas podem ser de grande utilidade aplicados por exemplo no conhecimento da estrutura de proteínas que formam parte de um vírus para estudar como inibi-lo ou desativá-lo é uma área na qual também podemos atuar. A continuação definimos os objetivos do presente trabalho.

Objetivo Geral: Explorar a simulação da estrutura de proteínas empregando modelos numéricos de diferentes níveis de complexidade.

Objetivos Específicos: (a) Criar modelos empregando sistemas de elementos finitos unifilares para avaliar 3 estados de conformação da Proteína Lisozima, utilizando sistemas estruturais familiares na simulação de macroestruturas. (b) Ajustar os parâmetros destes modelos empregando parâmetros medidos experimentalmente (c) Discutir a influência das ligações secundárias na calibração dos modelos propostos.

2. FUNDAMENTAÇÃO

Formação e estrutura de proteínas: proteínas são obtidas a partir da decodificação do DNA celular. São formadas a partir de uma sequência de aminoácidos, que são moléculas orgânicas compostas de um átomo de carbono central (carbono alfa) ligado a um grupo amino, um grupo carboxila, um átomo de hidrogênio e um componente que varia para cada aminoácido chamado de cadeia lateral (Ou grupo R)(ALBERTS B. et al, 2009). Os aminoácidos são

formados por ligações covalentes características que vinculam carbono e nitrogênio, as chamadas ligações peptídicas. As longas cadeias de aminoácidos que formam a estrutura básica de todas as proteínas são chamadas de estrutura primária. Na figura 2.1 está representado a estrutura básica de um aminoácido na parte superior à esquerda, embaixo uma representação de como os aminoácidos se conectam por ligações peptídicas e como tipicamente se posicionam as cadeias laterais, denotadas por R, indicadas por círculos coloridos.

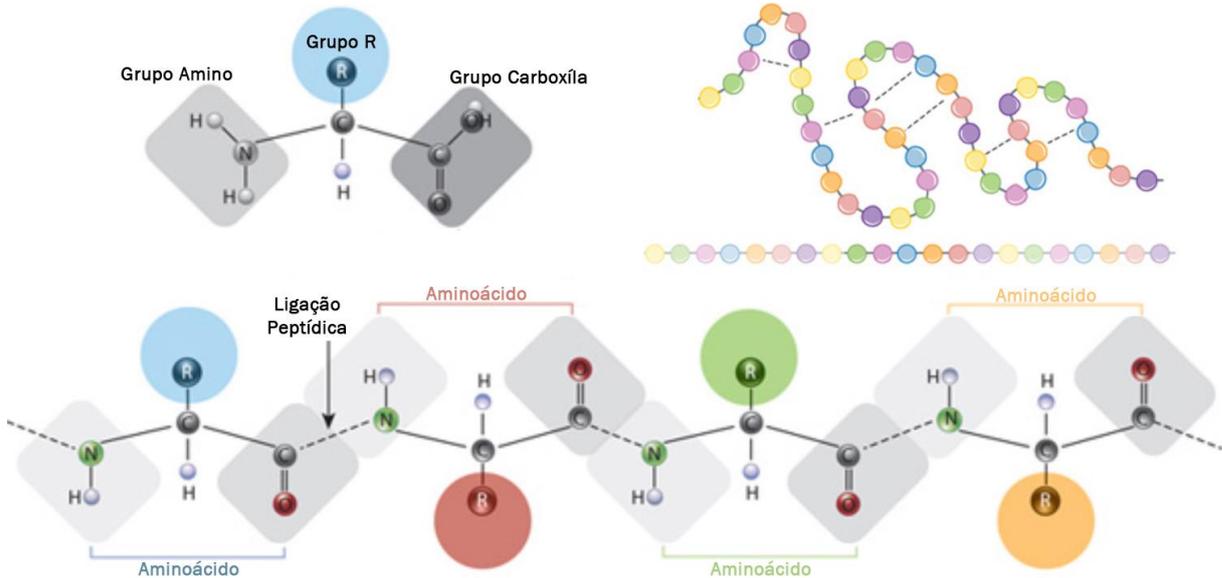


Figura 2.1 – Esquema de formação e estrutura da cadeia de aminoácidos. Adaptado de O'Connor; Adams (2010)

Todas as proteínas são constituídas a partir de vinte aminoácidos que possuem uma cadeia lateral única. Cada cadeia possui diferentes características químicas, como polaridade e eletronegatividade, podendo gerar diferentes ligações químicas entre a estrutura primária. Uma representação dessas ligações pode ser vista na figura 2.1 na parte superior direita, através de linhas tracejadas entre a sequência de diferentes aminoácidos. Estas ligações são formadas entre os átomos da cadeia lateral, são conhecidas como estruturas terciárias e são fundamentais para determinar o formato e a conformação em que a proteína será encontrada em um determinado meio.

Existe uma diferença na forma dos aminoácidos quando teoricamente isolados e quando ligados à proteína. Segundo o *RCSB Protein Data Bank* (Berman H.M. et al (2000)), os aminoácidos e proteínas nas suas formas ideais não possuem nenhuma ligação ou campo externo que interfira na sua estrutura. Já os aminoácidos quando ligados a uma macromolécula, possuem influência do campo elétrico, interações de polaridade, interações hidrofóbicas, ligações de Van der Waals entre outros fenômenos complexos que interagem com os átomos da vizinhança e alteram suas posições e suas propriedades. As ligações secundárias e terciárias são estimadas com base nessa configuração e no posicionamento dos átomos na estrutura conformada, sendo estas ligações mais fracas fundamentais para definir os estados de conformação da estrutura.

Um caso importante para compreender a estrutura de proteínas são as estruturas secundárias, que são definidas pelas ligações químicas entre a cadeia principal de aminoácidos (cadeia principal se refere a cadeia de aminoácidos sem incluir a cadeia lateral). Nessas ligações, dois casos particulares ocorrem com frequência nas proteínas, conhecidos como alfa-hélice e a folha-beta, que são representados na figura 2.2. Esses padrões se formam a partir de ligações de hidrogênio, no caso da alfa-hélice as ligações torcem a estrutura primária formando uma hélice com 3,6 voltas aminoácidos por volta e com os grupos R

posicionados para fora da hélice. No caso da folha-beta, dois ou mais seguimentos se alinham e se conectam por ligações de hidrogênio, com os seus grupos R posicionados para dentro. Certos aminoácidos têm uma propensão para formar uma alfa-hélice, enquanto outros têm uma propensão para formar uma folha-beta.

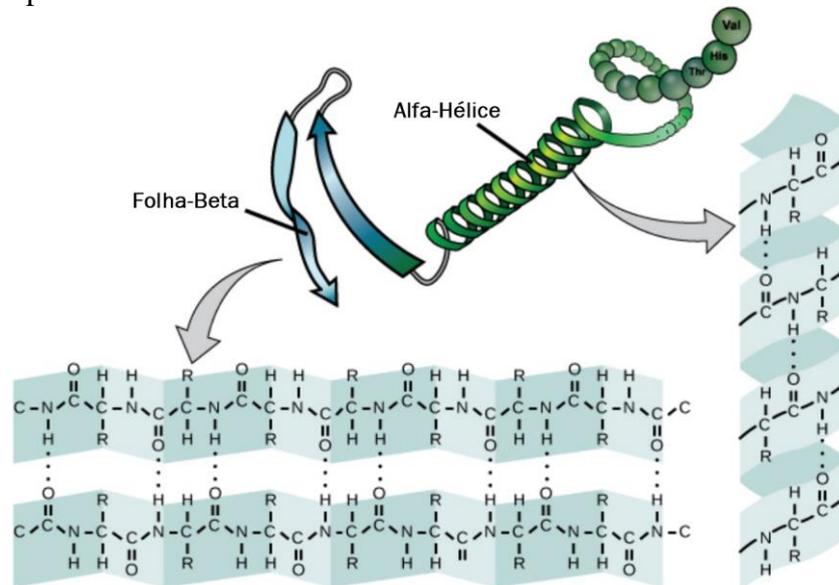


Figura 2.2 – Representação simplificada e ligações químicas secundárias de folha-beta (β -pleated sheet) e alfa-hélice (α -helix). Fonte: Clark, M (2018)

Características experimentais e bancos de dados de proteínas: As proteínas possuem tamanho entre 1-100 nm, apenas grandes agregados de proteínas podem ser vistos a partir de um microscópio. Para visualização da estrutura, são necessários métodos indiretos, como Cristalografia de raio-x (HAJDU, J. et al, 2000). Neste método, cristais de proteína purificados são dispostos em um feixe de raio-x e o padrão de deflexão dos raios é utilizado para prever as posições dos átomos presentes no cristal da proteína, podendo chegar a uma resolução de 1 Angstrom.

Quando observado experimentalmente a distribuição da densidade de elétrons de um átomo, esta varia a partir de sua posição ideal. Isto pode ser ocasionado por vibrações dentro da estrutura atômica ou diferenças entre as muitas moléculas presentes na estrutura cristalina. Assim, a observação do elétron possui uma “mancha” em torno do ponto ideal deste elétron, estes movimentos e a variação dos elétrons em torno deste ponto são incorporados ao modelo como um parâmetro que é medido experimentalmente chamado de fator de temperatura. Este fator é uma medida de confiança quanto a posição do átomo em torno de suas coordenadas, átomos na superfície das proteínas tendem a flutuar mais livremente, tendo assim um fator de temperatura mais elevados.

O *Protein Data Bank* (BERMAN H.M. et al, 2000) é um banco de dados para informações estruturais de proteínas e ácidos nucleicos, que podem ser usadas livremente, geralmente obtidos por cristalografia de raio-x. As informações podem ser exportadas em formato *.pdb* que é amplamente utilizada pelos *softwares* da área. Com as informações disponíveis no PDB, é possível definir as coordenadas de cada átomo da proteína, seu elemento químico, a que aminoácido pertence e seu fator de temperatura. No PDB é utilizada a nomenclatura proposta pela *IUPAC-IUB* (1970) que padroniza a numeração e o código para cada átomo de cada aminoácido. Na figura 2.3 segue exemplo para nomenclatura da cadeia principal e de alguns aminoácidos, a numeração indica a ordem em que é listado no arquivo exportado pelo PDB, a primeira letra indica o elemento químico do átomo e o restante identifica a posição do átomo dentro do aminoácido.

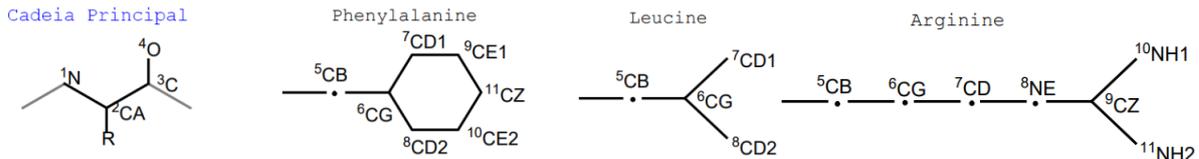


Figura 2.3 – Formatação padrão utilizada no PDB para cadeia principal e três cadeias laterais como exemplo. fonte: (IUPAC-IUB 1969)

O *software Chimera* (PETTERSEN E.F., et al, 2004), permite a leitura do arquivo exportado pelo PDB, além de diversas funções, como interface gráfica e diferentes representações tridimensionais da proteína, é possível selecionar e exportar as ligações químicas existentes na estrutura primária da proteína.

O *Protein Contact Atlas (PCA)* (KAYIKCI et al, 2018) é um banco de dados de ligações não-covalentes, que disponibiliza milhares de estruturas existentes no PDB com detalhes sobre suas ligações secundárias e terciárias. O post-processador do PCA permite avaliar proteínas através de um arquivo *.pdb* carregado a partir do site e fazer o *download* de um arquivo com todas ligações não covalentes obtidas. A Fig. 2.4 representa entre quais estruturas secundárias existem ligações secundárias e terciárias, para as proteínas avaliadas neste trabalho, as cores utilizadas para cada estrutura são as mesmas da representação tridimensional simplificada na Fig. 3.1.

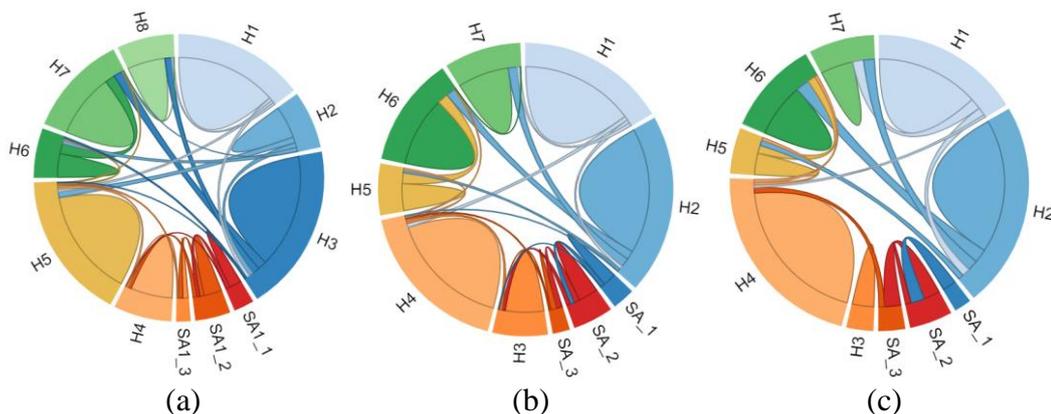


Figura 2.4 - Representação das ligações secundárias e terciárias, das três configurações da Lisozima estudadas 4YM8 (a), 1DPX (b) e 1DPW (c), entre as alfa-hélices (H1 até H8) e folhas-beta (SA_1 até SA_3). Fonte: (KAYIKCI et al, 2018)

Método de busca de raízes para otimização: O método da bisseção é utilizado para busca de raízes de uma função contínua. Utiliza o teorema do valor intermediário, que garante que se existe um máximo local entre um determinado intervalo de uma função, existe um ponto com derivada igual a zero e ele é um ponto de máximo local. A partir de um intervalo inicial, o método testa o ponto intermediário e define um novo subintervalo que contenha os valores de interesse, reduzindo o intervalo de busca e convergindo linearmente para a solução. O método é lento comparado com outros métodos, como o de Newton ou das Secantes, sendo necessário maior número de iterações para obter precisão. São métodos robustos quando lidam com funções que possuem maior suavidade próximo a solução.

Parâmetros calculados: O fator de temperatura calculado a partir de uma simulação de análise modal, pode ser obtido pela Eq. 1, sendo B_i o fator de temperatura calculado para cada átomo i como o resultado do somatório dos quadrados dos deslocamentos em x , y e z , dividido pelo inverso da frequência angular ω_n , para cada modo de vibração e multiplicados por uma constante. O termo kb na Eq. 1 indica a constante de Boltzmann ($\sim 1,38 \times 10^{-23}$ J / K) e T é a temperatura absoluta em Kelvin. A normalização dos parâmetros é uma forma melhor de avaliar a distribuição dos valores em torno da média e do desvio da população. Utilizando

a Eq. 2, é subtraído de B_i (valor tirado da amostra) a média da amostra $\mu(B)$ e dividido pelo desvio padrão da amostra $\sigma(B)$, obtendo B'_i que é o valor normalizado.

O coeficiente de correlação de Pearson para uma população, é calculado conforme a Eq. 3, sendo X e Y duas amostras independentes a serem avaliadas, onde $E(X)$ a esperança de X , $E(Y)$ a esperança de Y , $E(XY)$ a esperança de XY , $Var(X)$ a Variância de X e $Var(Y)$ a Variância de Y . Os valores obtidos para a correlação são entre 1 e -1, sendo 1 uma perfeita correlação entre os valores de X e Y , sendo -1 uma correlação perfeita inversa onde o aumento de X leva a diminuição de Y e 0 indicando nenhuma correlação entre os valores. As Eq. 1, 2 e 3 também são utilizadas por Giordani, G et al. (2021) e Scaramozzino, D. et al (2019) para comparação dos resultados experimentais e simulados.

$$B_i = \frac{8 \pi^2}{3} k_b T \sum_{n=7}^{3N} \frac{(a_{ix}^2 + a_{iy}^2 + a_{iz}^2)_n}{\omega_n^2} \quad (1), \quad B'_i = \frac{B_i - \mu(B)}{\sigma(B)} \quad (2), \quad \rho_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}} \quad (3).$$

3. METODOLOGIA

São escolhidos para este estudo três diferentes configurações da proteína lisozima, obtidos por cristalografia de raio-x e com dados disponíveis no *RCSB Protein Data Bank (PDB)* com os códigos de acesso 4YM8 (SUGAHARA, M. et al, 2015), 1DPX e 1DPW (WEISS, M.S. et al, 2000). As duas últimas foram obtidas pelos autores em meio levemente alcalino e a 1DPW utilizando 2-metil-2,4-pentanodiol (MPD) como precipitante, que se aloja entre as proteínas, causa uma diferença na formação dos cristais, e adiciona ligações terciárias na região e altera seu comportamento em relação a 1DPX. As proteínas citadas podem ser observadas na figura 3.1, que mostra a estrutura tridimensional simplificada, evidenciando a estrutura secundária com suas alfa-hélices e folhas-beta coloridas e a leve diferença entre os estados de conformação, as cores utilizadas nas estruturas são as mesmas da Fig 2.4.

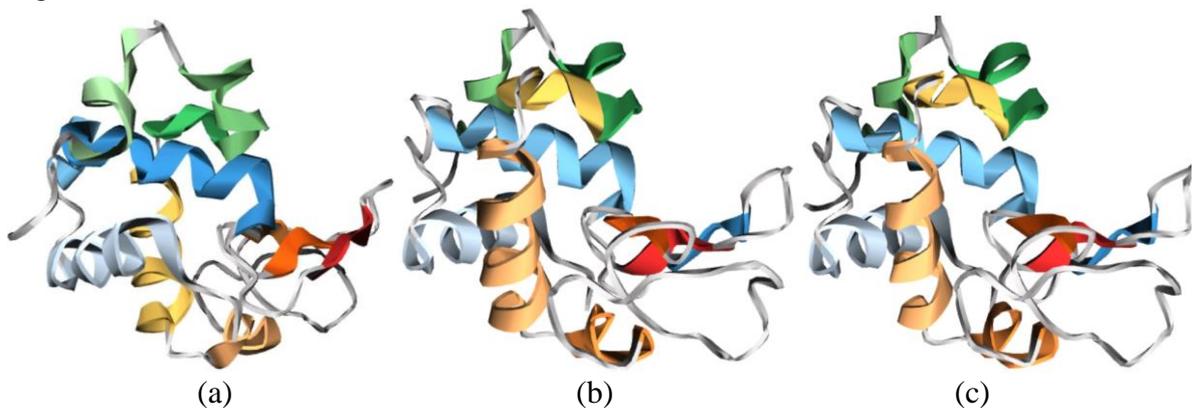


Figura 3.1 – Representação simplificada das proteínas 4YM8 (a), 1DPX (b) e 1DPW (c).
fonte : (KAYIKCI et al, 2018)

Na tabela A.1 no apêndice encontram-se os valores de massa, rigidez da ligação covalente e tipo de ligação usados no modelo para definir as vinculações primárias. Os valores de rigidez para estas ligações químicas covalentes, são estimados a partir da energia de dissociação da ligação, dada em kcal/mol, que é convertida para N/m de forma linear utilizando a ligação simples entre carbonos como referência. Sendo uma simplificação e aproximação das forças da ligação química para que um elemento estrutural unifilar com tal rigidez possua dinâmica semelhante.

As ligações de hidrogênio forte e hidrogênio fraca são estimadas como a média entre as ligações, utilizando como referência Jeffrey, G. A e Saenger, W (1991) para ligações de

hidrogênio que tipicamente ocorrem dentro de proteínas. Como exemplo da força das ligações covalentes, as ligações metálicas possuem uma rigidez entre 15 e 75 N/m.

Neste estudo, o *software* utilizado para as simulações foi o sistema de elementos finitos comercial *ANSYS-APDL* (ANSYS, INC (2006)). São definidos elementos de massa pontual para cada átomo da proteína e elementos de viga são definidos para representar as interações entre os átomos

Utilizando o *software Chimera* para obter as ligações primárias dos estados de conformação da lisozima, obtém-se mais de 1000 ligações a serem adicionadas como elementos. Com isso, o número de elementos a serem introduzidos no programa e os cálculos a serem realizados em cada elemento, tornam inviável sua manipulação de forma manual. Buscando uma forma de automatizar o procedimento, e proposto o roteiro apresentado na Figura 3.2 que se descreve a seguir: . O *software Chimera* é utilizado para leitura do arquivo *.pdb* e obtenção das ligações químicas da estrutura primária da proteína, sendo estes dados e o próprio arquivo *.pdb* inseridos no *Excel* na forma de uma planilha. Dentro do *Excel* é utilizado o *Virtual Basic Applications (VBA)* – que é uma linguagem de programação do Excel – que realiza os cálculos necessários, as verificações de cada tipo de ligações químicas, exporta os dados necessários para a simulação e para comparação de resultados, constrói o algoritmo em *APDL* com seus nós e elementos e exporta essas informações a serem lidas pela interface do programa *ANSYS APDL*.

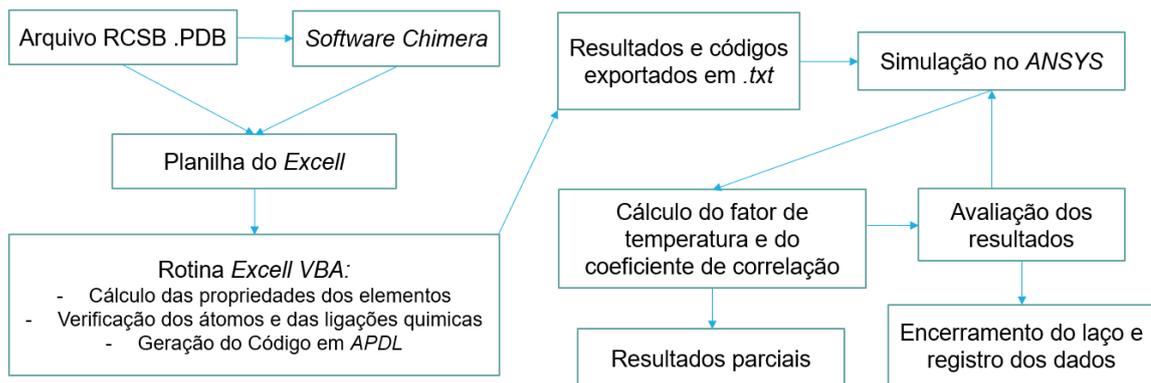


Figura 3.2 – Roteiro de algoritmos para simulação e obtenção de dados.

Utilizando a expressão $k_i = EA_i/l_i$, em cada ligação i , sendo o módulo de elasticidade (E) constante e igual a 100000 [N/m²], l_i a distância entre os átomos de cada ligação, e k_i a rigidez de cada ligação fornecidas na Tabela A.1 do apêndice A, é calculado para cada ligação a área de seção A_i para que cada elemento tenha rigidez equivalente a sua ligação.

É utilizada uma escala entre os valores reais e os numéricos, para evitar erros numéricos ao realizar os cálculos, por estar lidando com valores de massa e comprimentos muito pequenos, detalhes sobre esta conversão se podem ser encontrados em Carpinteri et al 2017. Na tabela 3.2, segue escala utilizada para realizar os modelos numéricos.

Tabela 3.2 – Escala para valores reais e numéricos.

Grandeza física	Escala
Massa, M (kg)	Numérico = real x10 ²⁶
Distância, L (m)	Numérico = real x10 ¹⁰
Rigidez, FL ⁻¹ (N/m)	Numérico = real x10 ⁰
Frequência, f (Hz)	Numérico = real x10 ¹³

Para as constantes de momento de inércia e módulo torsional são utilizados valores de 10⁶ [m⁴], que são muito altos em relação ao módulo de elasticidade, permitindo apenas a

flexibilidade axial das vigas e que elas atuem como molas com rigidez axial de acordo com a ligação química, sem interferências de efeitos de torção ou de flexão nos elementos.

Para a simulação, é utilizada uma rotina em linguagem *ANSYS APDL* para os três estados de conformação escolhidos. Esta rotina lê os parâmetros definidos e calculados pelo Excel, gerando os nós (átomos) com massas pontuais e elementos unifilares (ligações químicas). É realizado na estrutura uma análise modal, obtendo as frequências e deslocamentos para cada átomo em cada um dos 1000 primeiros modos de vibração, que serão utilizados para os cálculos das Eq. 1 e 2. O fator de temperatura experimental normalizado também é calculado pela rotina que é comparado aos resultados com o coeficiente de Pearson. Todos estes resultados são salvos externamente para serem avaliados. Como as estruturas são definidas livres e sem restrições ou engastes, os primeiros 6 modos de vibração são descartados dos cálculos, por conterem apenas movimentações de corpo livre.

Nas simulações para estrutura primária e para a adição de ligações de hidrogênio, o módulo de elasticidade dos elementos é constante para todos elementos unifilares. Para tentar melhorar os resultados, é definido um processo de otimização onde o módulo de elasticidade é definido pela Eq. 5, calculada para cada ligação, onde C é uma constante a ser definida, E é o módulo de elasticidade usado para as demais simulações e β_m é a **média do fator de temperatura experimental normalizado entre os átomos de cada ligação**. Quando C é igual a 1, o valor de E é igual para todas ligações; quando C é igual a 2, uma ligação que tenha 1 desvio em relação a média dos valores experimentais terá metade do módulo de elasticidade E . Ou seja, quanto maior a constante C , mais as ligações serão afetados pelo fator de temperatura experimental, assim, os átomos que tenham fatores elevados, terão elementos associados com menor rigidez.

$$E(\beta_m) = E * C^{-\beta_m} \quad (5)$$

A partir da Eq.5, através dos resultados da simulação, é buscado o valor de C que resulte no melhor coeficiente de correlação de Pearson entre os fatores de temperatura simulados e experimental possível. Para encontrar este valor, são realizadas simulações iterativamente, alterando o valor de C até este convergir para um ponto ótimo. Na primeira iteração é utilizado $C = 1$, são calculados os resultados até obter o coeficiente de Pearson, no fim da iteração é decidido qual será o novo C a ser testado pelo procedimento. Como eu não sei o intervalo onde está o melhor valor de C , o critério inicial é adicionar ao valor de C um valor chamado de passo a cada iteração, sendo $C = C + \text{passo}$. O passo utilizado é de 0,5. Quando os resultados retornam um valor de Pearson menor que o resultado anterior, o código deixa de adicionar valores a C e define o intervalo onde está o máximo local. Após saber este intervalo, é utilizada a aproximação deste ponto pelo método da bisseção, que testa valores de C ao redor do melhor valor de C encontrado até então. O método não possui rápida convergência, comparado com outros métodos, porém avalia um maior número de valores em torno do valor ótimo, permitindo uma melhor avaliação do comportamento da função. É utilizado um critério de parada pelo número de iterações, são realizadas 20 iterações para , para testar o método, buscar pelo melhor valor de correlação e avaliar o comportamento da alteração do módulo de elasticidade dos elementos utilizando a Eq. 5.

A partir do *Protein Contact Atlas*, são extraídas as ligações secundárias e terciárias. O número de ligações deste tipo para Lisozima passa de 4000 ligações de diversos tipos. Para adicionar ao modelo e verificar sua influência na dinâmica da estrutura, são selecionadas apenas as ligações de hidrogênio e hidrogênio fracas, excluídas as ligadas as moléculas de água ao redor da proteína, estas ligações são as mais rígidas presentes depois das ligações da estrutura primária. Contabilizando assim, em torno de 100 ligações de hidrogênio fracas e 140 ligações de hidrogênio. As ligações são exportadas do *PCA* e seguem o mesmo procedimento das ligações do *Chimera* para inserir os respectivos elementos no *software ANSYS*. Estas

ligações, quando presentes, possuem variação do módulo de elasticidade em função do fator de temperatura experimental, conforme Eq. 5. A representação tridimensional das ligações da estrutura primária podem ser visualizadas na Fig. 3.3, retirada do *ANSYS APDL*, representadas pelas linhas em azul. As ligações utilizadas do *PCA* estão representadas por linhas vermelhas e os asteriscos em azul representam as massas pontuais dos átomos.

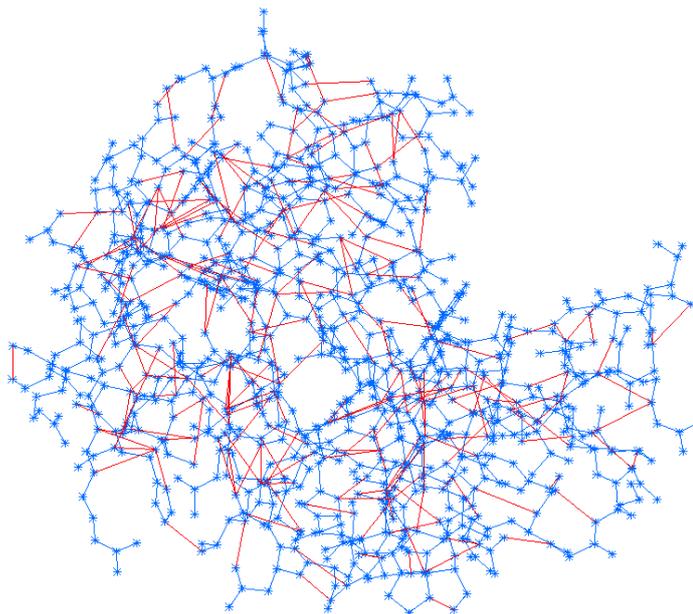
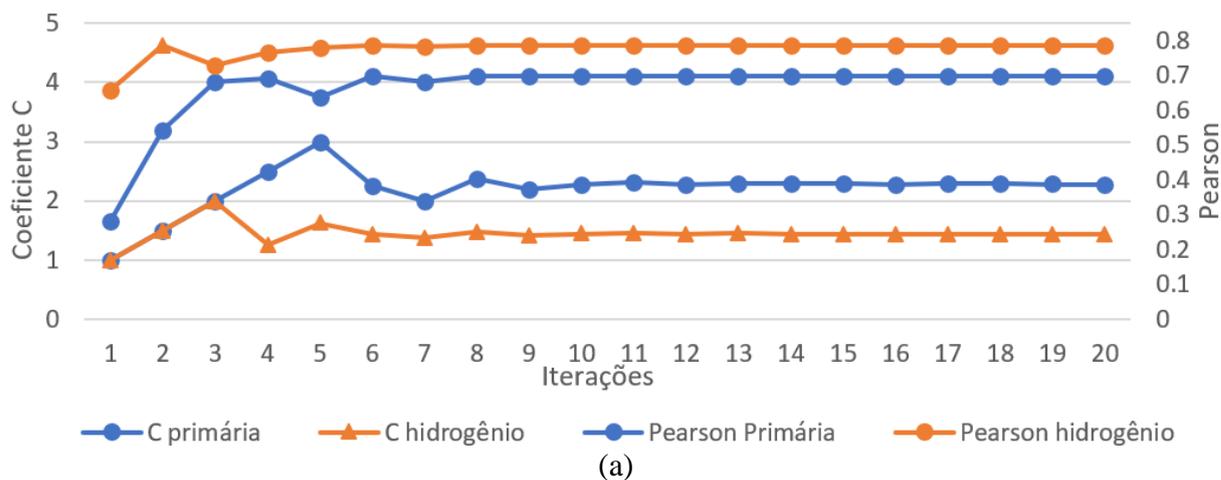


Figura 3.3 – Elementos da estrutura da 4YM8 inseridos no *software ANSYS*, com linhas representando ligações e asteriscos representando as massas pontuais dos átomos. Em vermelho ligações retiradas do *PCA*, em azul ligações retiradas do *Software Chimera*.

4. RESULTADOS

4.1 Análise dos procedimentos de otimização

Para cada estado de conformação são realizados dois processos de otimização, para a estrutura primária e para estrutura primária com a adição das ligações de hidrogênio. Na Figura 4.1(a), temos os resultados do coeficiente C (utilizado na Eq. 5 para variar o módulo de elasticidade dos elementos) e o coeficiente de *Pearson* obtido em relação ao fator de temperatura experimental, para as simulações realizadas para o estado de conformação 4YM8. Na Figura 4.1 (b) temos para cada simulação realizada os valores dos coeficientes de *Pearson* no eixo vertical e o coeficiente C no eixo horizontal.



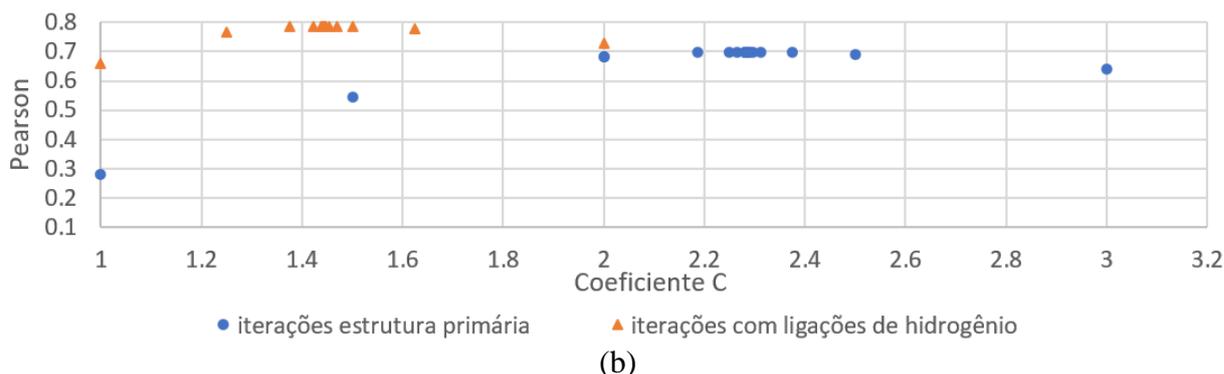


Figura 4.1 (a) Gráfico do coeficiente C simulado (eixo à esquerda) e Coeficiente de *Pearson* em relação ao experimental (eixo à direita) para cada iteração, utilizando o estado de conformação 4YM8. (b) Gráfico do coeficiente de *Pearson* em relação ao Coeficiente Simulado, utilizando a configuração 4YM8.

Na Fig. 4.1 (a) vê-se que na iteração 3 para ligações de hidrogênio, utilizando um $C = 2$, há uma redução no resultado de *Pearson*, indicando que o melhor valor de C se encontra entre 1 e 2. Para ligações de hidrogênio, isto ocorre na iteração 5, utilizando um valor de $C = 3$, indicando um valor ótimo entre 2 e 3. As demais iterações reduzem o intervalo de busca, convergindo para os valores da Tabela 4.1. Para os estados de conformação 1DPX e 1DPW, os gráficos do comportamento das iterações e da convergência são muito semelhantes, os resultados para estas são apresentados no Apêndice B.

O método da bisseção, apesar de sua simplicidade, em sua oitava iteração obteve um erro relativo ao melhor valor encontrado de 0,2%. Os resultados apresentados na Fig. 4.1 (b) mostram em ambos os casos um aumento no valor de *Pearson* com o aumento de C até um ponto de máximo global, onde a inclinação e sensibilidade ao parâmetro C é baixa. A adição de ligações de hidrogênio resultou em um coeficiente C menor mas com um perfil semelhante. Este comportamento também ocorreu nos outros estados de conformação, indicando a possibilidade do uso de um método com maior grau de convergência, ou investigar a existência de um comportamento padrão em análises de estruturas primárias utilizando uma regressão não linear utilizando uma amostra maior de estruturas.

Os valores de C que retornam o melhor resultado para o coeficiente de *Pearson* são exibidos na Tabela 4.1. Os coeficientes encontrados para os diferentes estados de conformação são semelhantes para estruturas primárias e para estruturas com ligações de hidrogênio, tendo a maior diferença entre 4YM8 e 1DPW, de 15%, para a estrutura primária.

Tabela 4.1 – Valores ótimos para coeficiente C .

	4YM8	1DPX	1DPW
Primária	2.282	1.997	1.979
Hidrogênio	1.443	1.507	1.416

4.2 Fatores de temperatura experimental e simulados

4.2.1 Estruturas primárias

Na Figura 4.2 se apresentam os fatores de temperatura experimentais para 4YM8, normalizados pela Eq. 2 (em preto). No eixo vertical temos os desvios em relação a média e no eixo horizontal o número de cada átomo. São sobrepostos ao experimental os resultados do fator de temperatura normalizados obtidos pela Eq. 1 das simulações com ligações primárias (em verde) e para o melhor resultado das ligações primárias otimizado (em amarelo).

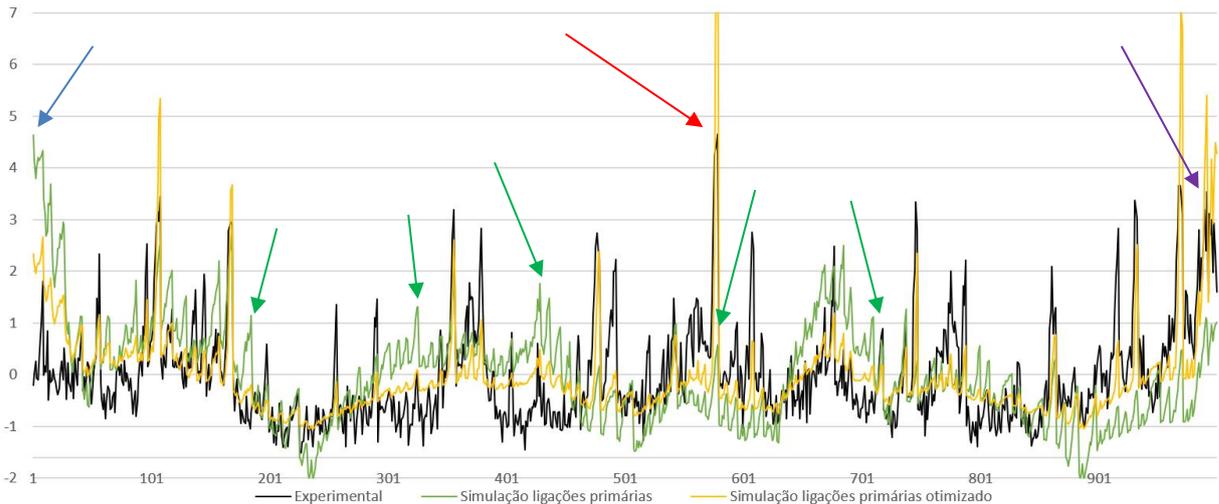


Figura 4.2 – Fatores de temperatura experimental e simulados para cada átomo utilizando ligações primárias e primárias otimizado para o estado de conformação 4YM8.

Como a numeração dos átomos segue um padrão para cada aminoácido, iniciando pelos átomos da cadeia principal seguidos da cadeia secundária. A cadeia principal esta sempre conectada a outros dois aminoácidos, com exceção das extremidades da proteína. Na simulação com ligações primárias, em verde na Fig. 4.2, é possível observar que há diversos picos ao longo da proteína, na grande maioria estes se encontram na cadeia lateral do aminoácido, que fica a vibrar como uma viga em balanço conectada ao átomo de carbono da cadeia principal, como é o caso do átomo 578 indicado pela seta vermelha que faz parte do aminoácido arginina, a cadeia lateral é apresentada na Fig. 2.3 e o pico corresponde aos átomos NE, CZ, NH1 e NH2. Muitos picos coincidem com o experimental, mesmo que haja uma certa distância nos valores causada pela vibração da cadeia principal, porém, no experimental nem sempre estes ocorrem ou ocorrem de maneira mais acentuada, como indicam as setas verdes na Fig. 4.2. Também podemos observar na Fig. 4.2 que ela possui um aumento de flexibilidade nas extremidades, principalmente próximo ao átomo 1 indicado pela seta azul e próximas ao átomo 1000 indicado pela seta roxa. Este aumento não ocorre no experimental para o átomo 1, indicando que existem restrições na região que a simulação é limitada em compreender.

Ainda na Fig. 4.2, temos os resultados da simulação com ligações primárias otimizado, em amarelo. Nas extremidades, onde o modelo era distante em reproduzir a flexibilidade, o resultado se aproxima do experimental, indicado pelas setas azul e roxa. Em alguns picos que já coincidiam, o valor dos picos se aproxima, porém ocorrem alguns exageros como é o pico em torno do átomo 578, indicado pela seta vermelha, que chega a quase 10 desvios e é omitido no gráfico para facilitar a visualização.

Apêndice C, A Fig. C.1, apresenta os resultados para o estado de conformação 1DPX, da mesma forma que na Figura 4.2. Podemos perceber que o resultado da simulação para ligações primárias é praticamente idêntico ao obtido para 4YM8, isto também pode ser percebido comparando os modos apresentados no Apêndice E, Fig. E.1, que são muito semelhantes. Os resultados experimentais são semelhantes aos da 4YM8, com mudanças em alguns picos. O resultado otimizado novamente traz melhorias para as extremidades e aumento de flexibilidade em picos do experimental. Os picos nos átomos em 970 e 991, indicados pelas setas vermelhas, são de 12,6 e 10,3 desvios, respectivamente, extrapolando as dimensões do gráfico, esta região pode ser visualizada no Apêndice E, na Fig. E.2 no modo 4 da simulação H indicado pela seta, onde ocorre o maior deslocamento no modo. Na região próxima ao átomo 240, indicado pela seta azul no Apêndice C, na Fig. C.1, o resultado

otimizado teve um aumento de flexibilidade aproximando do experimental, apesar do fator de temperatura experimental ser negativo e causar um aumento de rigidez nos elementos.

No Apêndice C, na Fig. C.2, se apresentam os resultados para a configuração 1DPW, apresentados da mesma forma que na figura 4.2. Novamente o resultado da simulação para ligações primárias é praticamente idêntico ao obtido para 4YM8 e 1DPX. O pico no átomo 970 é de 14,3 desvios, indicado pela seta vermelha.

Quando otimizada a alteração do módulo de elasticidade utilizando o fator de temperatura experimental, de maneira geral, atenua variações locais em regiões de baixa flexibilidade e evidencia alguns picos em regiões de alta flexibilidade. O resultado ótimo nas três avaliações acima causaram pelo menos um pico exagerado, mas que coincide com o maior valor experimental.

4.2.2 Estruturas primárias com adição de ligações secundárias

Com a adição das ligações secundárias, novamente são obtidos e comparados os fatores de temperatura experimentais e os calculados, os resultados podem ser visualizados na Fig. 4.3 para a configuração 4YM8. São adicionados ao gráfico triângulos em vermelho representando as ligações de hidrogênio adicionadas e em triângulos verdes as ligações de hidrogênio fracas, ambas abaixo do gráfico. Também abaixo do gráfico, está indicado por uma linha em roxo os dos átomos que formam alfa-hélices (H1 até H8) e folhas-beta (SA1 até SA3). Com a adição de ligações de hidrogênio há uma visível melhora em relação ao resultado experimental, principalmente nas extremidades e na região que contém SA1, SA2 e SA3, indicados na Fig. 4.3 pelas setas verdes. Porém os picos que coincidem às cadeias laterais se acentuam em regiões sem ligações de hidrogênio, alguns coincidindo com valores experimentais e outros não, como indicam as setas em amarelo. De forma semelhante à estrutura primária nas Figuras 4.2, no Apêndice C, na Fig. C.1 e Fig. C.2, a otimização aproximou o resultado, intensificando alguns picos que coincidem e reduzindo os demais.

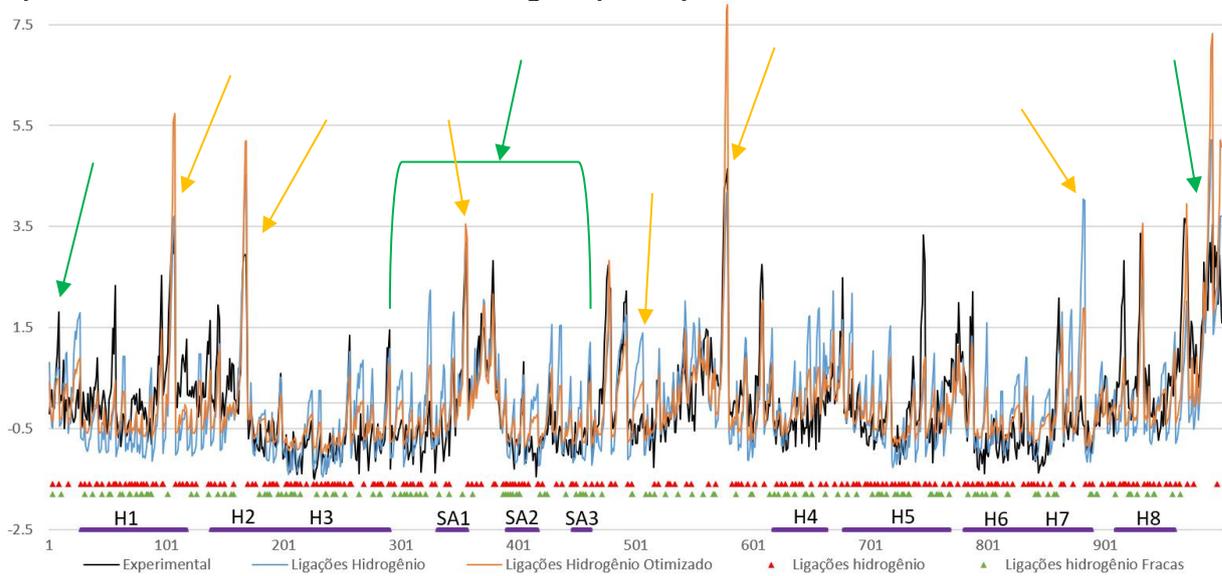


Figura 4.3 – Fatores de temperatura experimental e simulados para cada átomo utilizando ligações de hidrogênio e otimizado para a configuração 4YM8.

No Apêndice D, na Fig. D.1 e Fig. D.2 são apresentados os resultados para 1DPX e 1DPW, exibidos da mesma forma que na Fig. 4.3. No Apêndice D, na Fig. D.1 o pico em torno do átomo 992 é de 10,9 desvios, indicado pela seta vermelha. No Apêndice D, na Fig. D.2 o pico em torno do átomo 970 é de 10,6 desvios, indicado pela seta vermelha.

Comparando as Fig. 4.3, e Apêndice D, na Fig. D.1 e Fig. D.2, podemos perceber que há pouca variação nas ligações de hidrogênio e hidrogênio fracas entre os diferentes estados de conformação. Os resultados para Ligações de Hidrogênio são muito semelhantes, com pequenas variações entre alguns picos onde há mudança nas ligações de hidrogênio. Os resultados da otimização para ligações de hidrogênio da 1DPX e 1DPW é muito semelhante do resultado para 4YM8, porém é possível visualizar ajustes onde existem variações nos picos de valores experimentais, como indicam as setas verdes nas figuras do Apêndice D, Fig. D.1 e Fig. D.2. A seta amarela indica um ponto onde há ligações de hidrogênio na 1DPX, diferente das demais, nesse ponto não há um bom ajuste pela otimização, que pode indicar uma falha do método ou um erro na determinação das ligações.

4.3 Resultados do coeficiente de Pearson entre experimental e resultados

Todos os resultados para o fator de temperatura calculados são comparados na Figura 4.4, onde é calculado o coeficiente de correlação de Pearson entre as amostras de todos resultados simulados. O gradiente de cores está de acordo com o resultado, indo de azul para correlação igual a 1 até vermelho para correlação igual a 0, os nomes são abreviados para facilitar a visualização como segue: fator de temperatura experimental (FTE), ligações primárias (P), ligações primárias otimizadas (PO), ligações primárias e hidrogênio (H) e ligações primárias e hidrogênio otimizadas (HO).

	4ym8	4ym8	4ym8	4ym8	4ym8	1dpx	1dpx	1dpx	1dpx	1dpx	1dpw	1dpw	1dpw	1dpw	1dpw
	FTE	P	PO	H	HO	FTE	P	PO	H	HO	FTE	P	PO	H	HO
FTE	1	0.28	0.70	0.66	0.79	0.80	0.28	0.56	0.66	0.70	0.83	0.28	0.58	0.69	0.74
4ym8	P	1	0.58	0.40	0.33	0.18	1	0.49	0.43	0.30	0.10	1	0.47	0.41	0.25
	4ym8	PO	1	0.63	0.82	0.62	0.58	0.84	0.66	0.77	0.56	0.58	0.81	0.66	0.77
		4ym8	H	1	0.90	0.63	0.40	0.56	0.91	0.80	0.55	0.40	0.48	0.95	0.75
			4ym8	HO	1	0.73	0.33	0.70	0.86	0.91	0.64	0.33	0.59	0.87	0.83
				1dpx	FTE	1	0.18	0.69	0.60	0.78	0.85	0.18	0.64	0.67	0.80
					1dpx	P	1	0.49	0.43	0.30	0.11	1	0.47	0.41	0.25
						1dpx	PO	1	0.56	0.79	0.53	0.49	0.88	0.63	0.82
							1dpx	H	1	0.85	0.54	0.43	0.49	0.92	0.73
								1dpx	HO	1	0.61	0.29	0.61	0.81	0.81
									1dpw	FTE	1	0.10	0.66	0.61	0.78
										1dpw	P	1	0.47	0.41	0.25
											1dpw	PO	1	0.58	0.85
												1dpw	H	1	0.85
													1dpw	HO	1

Figura 4.4 – Fatores de correlação obtidos pelo cruzamento com experimental e com todos os resultados.

Na Fig. 4.9, podemos observar que os resultados entre FTE para as três conformações possuem uma boa correlação, de 0.8 entre 1dpx e 4ym8, 0.83 entre 1dpw e 4ym8 e 0.85 entre 1dpw e 1dpx. Apesar da boa correlação, o resultado indica o grau de diferença entre a dinâmica dessas estruturas, sendo a 1dpx mais semelhante a 1dpw do que a 4ym8 entre as demais. Os resultados de P em relação a FTE são de 0.28 para 4ym8, 0.18 para 1dpx e 0.1 para 1dpw, indicando alguma correlação apenas para 4ym8. O resultado melhor para 4ym8 e pior para 1dpw pode ser um indicativo de que o comportamento dinâmico de 4ym8 depende

mais de sua estrutura primária em relação a 1dpw. O cruzamento dos resultados (comparando P de 4ym8 em relação a P de 1dpx, por exemplo) para as simulações primárias (P), apresentou correlação muito próxima de 1 para os 3 casos, todas maiores que 0.999 e arredondadas para 1 para melhor visualização. Isso ressalta a semelhança observada nas figuras da subseção 4.2, indicando que a simulação da estrutura primária da proteína lisozima é incapaz de trazer informações sobre seu estado de conformação.

Os resultados para PO em relação à FTE são de 0.7 para 4ym8, 0.69 para 1dpx e 0.66 para 1dpw, mostrando melhora e seguindo proporção semelhante em relação aos resultados de P. O cruzamento dos resultados para PO mostra um resultado semelhante a diferença encontrada pelo cruzamento dos valores experimentais, próximo a 0.84, novamente tendo um melhor resultado comparando 1dpx e 1dpw de 0.88.

Para as simulações H os resultados em relação à FTE são de 0.66 para 4ym8, 0.6 para 1dpx e 0.61 para 1dpw, indicando boa melhora em relação aos obtidos para P, novamente com melhor valor para 4ym8. O cruzamento entre as simulações H obteve correlação de 0.91 entre 1dpx e 4ym8, 0.95 entre 1dpw e 4ym8 e 0.92 entre 1dpw e 1dpx, o que indica a pequena mudança entre os resultados, conforme comentado na subseção 4.2. Quando comparado os resultados da simulação com o fator experimental de outro estado de conformação, os resultados para H para 1dpw geram melhor correlação comparados com o FTE de 4ym8 e 1dpx do que com 1dpw, o que pode indicar que a diferença do comportamento dinâmico da 1dpw em relação as demais não é capturado pela adição das ligações de hidrogênio propostas, apesar da melhora do coeficiente.

Os resultados para HO em relação à FTE são de 0.79 para 4ym8, 0.78 para 1dpx e 0.78 para 1dpw, sendo os melhores resultados obtidos, mostrando boa melhora em relação H. Em comparação com os resultados para PO, houve um aumento de aproximadamente 0.1. O resultado para simulação HO para 1dpw, obteve melhor correlação quando comparado com 1dpx do que com 1dpw, indicando que mesmo o melhor resultado não distingue a diferença dinâmica entre os estados mais parecidos avaliados. Os resultados para HO possuem boa correlação com os resultados de PO e H.

4.4 Avaliação dos modos de vibração

As simulações realizadas resultaram em cerca de 1000 modos de vibração. De acordo com a literatura, os primeiros modos trazem mais informações sobre o comportamento biológico da proteína, sendo estes caminhos de menor energia para o deslocamento e vibração. Como forma de discutir o método proposto, são analisados apenas os modos iniciais de cada configuração. Na Fig. E.2 do Apêndice E, são exibidos os quatro primeiros modos de cada simulação para o estado de conformação 1DPX, a figura está com a deformação causada pela movimentação, em escala representativa, e os vetores indicam o máximo deslocamento durante o modo, o gradiente de cores vai do vermelho, para o ponto de máximo deslocamento no modo, para o azul no ponto de mínimo. No primeiro modo, para todos os casos, pode ser observado a movimentação longitudinal ao longo do maior comprimento da proteína. A forma dos quatro primeiros modos de P e PO são muito semelhantes, com pequenas variações de direção e intensidade em alguns átomos. No terceiro modo para P e PO, a movimentação dos átomos entra e sai do plano da folha, com ponto de máximo indicado pelas setas.

Com exceção do primeiro modo, os modos de 2 até 4 para H e HO, são distintos dos obtidos para P e PO, como mostra a Fig. E.2 do Apêndice E. As setas que são adicionadas indicam o átomo com maior deslocamento no modo. Os modos H e HO são semelhantes entre si, porém com alguns pontos com deslocamentos exagerados pela otimização em HO. No modo 4 para H e no modo 1 e 4 para HO, as setas indicam o átomo número 992, que possui um pico para o fator experimental e para as simulações (mostrado no Apêndice D, na Fig. D.1

e indicado por uma seta vermelha). Nas simulações P e PO, o átomo 992 tem menor deslocamento nos 4 primeiros modos, a adição de ligações de hidrogênio aumentou o seu deslocamento nos modos iniciais, com seu fator de temperatura se aproximando do experimental, com a otimização esse deslocamento aumentou para os 4 primeiros modos, causando um erro para este ponto. Nos modos 2 e 3 temos semelhanças entre si, com as setas apontando para a mesma região. No modo 2 apontam para o átomo 677 que no Apêndice D, na Fig. D.1 é uma região com média flexibilidade, indicada pela seta azul, sem grandes picos. No modo 3 apontam para o átomo 168, que possui alto fator de temperatura experimental e também nas simulações de H e HO. Diferente do ponto 992, estes pontos não tiveram um aumento significativo pelo procedimento de otimização.

5. CONCLUSÃO

O presente trabalho teve como objetivo explorar modelos numéricos de diferentes níveis de complexidade, utilizando elementos finitos, aplicado a dinâmica de proteínas. É utilizado para simulação das estruturas o *software ANSYS* utilizando rotinas na linguagem *APDL* para inclusão dos dados, otimização e coleta dos resultados. Os dados experimentais são obtidos do *Protein Data Bank*, as ligações químicas primárias do *software Chimera* e as não-covalentes do *Protein Contact Atlas*. Os dados são processados e o código gerado em *APDL* utilizando o *Software Excel*. São realizadas simulações na proteína Lisozima em três configurações diferentes, com códigos do PDB 4YM8, 1DPX e 1DPW, a fim de testar a capacidade dos métodos propostos em identificar a diferença entre estes estados. No primeiro modelo são realizadas simulação utilizando apenas a estrutura primária, no segundo modelo são adicionadas algumas ligações de hidrogênio e é proposto uma forma de otimizar as ligações utilizando o parâmetro experimental.

Os resultados para estrutura primária com otimização do estado de conformação 4YM8 obteve resultados próximos aos de Giordani, G et al. (2021), que utilizou o método CG para 4YM8. O melhor resultado obtido foi com a adição de algumas ligações de hidrogênio e com a alteração da rigidez das ligações utilizando otimização. O uso apenas da estrutura primária se demonstrou inútil em diferenciar a dinâmica entre os estados de conformação. Apesar das ligações de hidrogênio terem de 5 a 12% da rigidez das ligações carbono-carbono, a adição resultou em melhorias significativas para os resultados, indicando a possibilidade de diferenciar os estados de conformação. A utilização de um maior número de ligações não-covalentes se mostra uma possibilidade de melhoria do método.

O método de otimização, apesar de simples, teve boa convergência. O comportamento da correlação de Pearson em relação à função proposta se mostrou suave e indicando um ponto de máximo global, mostrando a possibilidade de utilizar outros métodos melhores de otimização. O aumento da correlação nas estruturas otimizadas indicou uma melhora na dinâmica global da estrutura, podendo haver picos e distorções locais, bem como alterações na direção do deslocamento dos átomos nos modos de vibração avaliados. Podem ser sugeridas alterações na função otimizada que minimizem os erros locais.

A análise deste trabalho possui espaço para melhoria dos algoritmos e adição de mais informações nos modelos, como ligações não-covalentes, átomos de hidrogênio da estrutura e moléculas de água. Estruturas biológicas conhecidas podem ser testadas para validação do método. Estruturas de vírus, bactérias e mutações em estruturas podem ser avaliadas investigando correlações de sua dinâmica com suas funções biológicas. Por fim, melhorar a comunicação com áreas como a bioquímica e biologia, direcionando os resultados para possíveis aplicações.

REFERÊNCIAS BIBLIOGRÁFICAS

Alberts, B.; Bray D.; Hopkin, K.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; and Walter, P.. **Essential Cell Biology. Third Edition. (2009)** New York: Garland Science (Taylor & Francis Group). ISBN: 978-0-8153-4129-1.

Alberto Carpinteri, Gianfranco Piana, Andrea Bassani & Giuseppe Lacidogna (2018): **Terahertz vibration modes in Na/K-ATPase, Journal of Biomolecular Structure and Dynamics**, DOI: 10.1080/07391102.2018.1425638

ANSYS, INC. **ANSYS Verification Manual**. Ansys Release 9.0. Canonsburg, 2004

Bae W, Choi MG, Hyeon C, Shin YK, Yoon TY. **Real-Time Observation of Multiple-Protein Complex Formation with Single-Molecule FRET**. *J Am Chem Soc*. 2013;135(28):10254–10257. pmid:23808880

Berman H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) **The Protein Data Bank Nucleic Acids Research**, 28: 235-242.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965). **Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution**. *Nature*, 206, 757-761.

Bochicchio, D., Salvalaglio M., Pavan M., “**Into the Dynamics of a Supramolecular Polymer at submolecular Resolution - Nat Commun**. 2017; 8: 147. Published online 2017 Jul 27. doi: 10.1038/s41467-017-00189-0

Carpinteri A, Lacidogna G , Piana G, Bassani A; **Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis**
<http://dx.doi.org/10.1016/j.molstruc.2017.02.099> 0022-2860/© 2017 Elsevier

Clark, M; Douglas, M; Choi, J. - **Biology 2e (2018)** – OpenStax, Rice University - Digital: ISBN-10: 1-947172-52-2 - Disponível em:
<https://openstax.org/details/books/biology-2e>

Escobedo, A., Topal, B., Kunze, M.B.A. *et al.* **Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor**. *Nat Commun* 10, 2034 (2019).
<https://doi.org/10.1038/s41467-019-09923-2>

Giordani, G.; Scaramozzino, D.; Iturrioz, I.; Lacidogna, G.; Carpinteri, A. **Modal Analysis of the Lysozyme Protein Considering All-Atom and Coarse-Grained Finite Element Models**. *Appl. Sci.* 2021, 11, 547. <https://doi.org/10.3390/app11020547>

Goodsell, D. (2000). **Molecule of the Month: Lysozyme**. DOI: 10.2210/rcsb_pdb/mom_2000_9 RCSB Protein Data Bank. Available online:
<http://www.rcsb.org/pdb>.

Hajdu, J; Neutze, R; Sjögren, T; Edman, K; Szöke, A; Wilmouth, RC; Wilmot, CM (2000). "Analyzing protein functions in four dimensions". *Nature Structural Biology*. 7 (11): 1006–12. doi:10.1038/80911

Hu X., Buehler X., “**Comparative Analysis of Nanomechanical Features of Coronavirus Spike Proteins and Correlation with Lethality and Infection Rate**”, *Matter* 4, 265–275 January 6, 2021 ^a 2020 Elsevier Inc. <https://doi.org/10.1016/j.matt.2020.10.032>

IUPAC-IUB Commission on Biochemical Nomenclature. "**Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969)**", *J. Biol. Chem.* 245, 6489 (1970).

Jeffrey, G. A.; Saenger, W. **Hydrogen Bonding in Biological Structures**, 1st ed.; Springer-Verlag: Berlin, Heidelberg, 1991

Kayikci, M. Venkatakrishnan A, Scott-Brown J., Ravarani C. N. J, Flock T, Babu M.-
Protein contacts atlas: visualization and analysis of non-covalent contacts in biomolecules - *Nat Struct Mol Biol.* **2018** Feb; 25(2): 185–194 doi: 10.1038/s41594-017-0019-z

Lacidogna, G., Piana, G., Bassani, A., & Carpinteri, A. (2017). **Raman spectroscopy of Na/K-ATPase with special focus on low-frequency vibrations.** *Vibrational Spectroscopy*, 92, 298–301.

Ma, Jianpeng. **Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes Structure**, Vol. 13, 373–380, March, 2005, Elsevier Ltd. DOI 10.1016/j.str.2005.02.002

Mittermaier A, Kay LE. **New tools provide new insights in NMR studies of protein dynamics.** *Science.* **2006** Apr;312(5771):224–228. pmid:16614210

O'Connor, C. M. & Adams, J. U. **Essentials of Cell Biology.** Cambridge, MA: NPG Education, 2010.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. **UCSF Chimera - a visualization system for exploratory research and analysis.** *J Comput Chem.* **2004** Oct;25(13):1605-12.

Scaramozzino, D., Lacidogna, G., Piana, G. *et al.* **A finite-element-based coarse-grained model for global protein vibration.** *Meccanica* **54**, 1927–1940 (2019).
<https://doi.org/10.1007/s11012-019-01037-9>

Sugahara, M., Nakane, T., Suzuki, M., Nango, E. (2015). **Crystal structure of hen egg-white lysozyme** DOI: 10.2210/pdb4ym8/pdb – Disponível em: <
<https://www.rcsb.org/structure/4YM8> >

Weiss, M.S., Palm, G.J., Hilgenfeld, R. (2000) **Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD** *Acta Crystallogr D Biol Crystallogr* **56**: 952-958

Weiss, M.S., Palm, G.J., Hilgenfeld, R. **STRUCTURE OF HEN EGG-WHITE LYSOZYME**, DOI: 10.2210/pdb1DPX/pdb - Disponível em: <
<https://www.rcsb.org/structure/1DPX> >

Apêndice A: Parâmetros utilizados nos átomos e elementos.

Tabela A.1 – Parâmetros utilizados nos átomos e elementos.

Átomo	Massa (g x10 ⁻²³)	Ligação	Rigidez (N/m)
Carbono (C)	1.994475	C – C	180
		C = C	320
Nitrogênio (N)	2.325918	C – N	160
		C = N	322
Oxigênio (O)	2.656698	C – O	190
		N – O	105
Enxofre (S)	5.324655	C – S	135
		S – S	118
Hidrogênio (H)	–	Forte	21
		Fraca	8

Apêndice B: Figuras dos procedimentos de otimização.

1DPX:

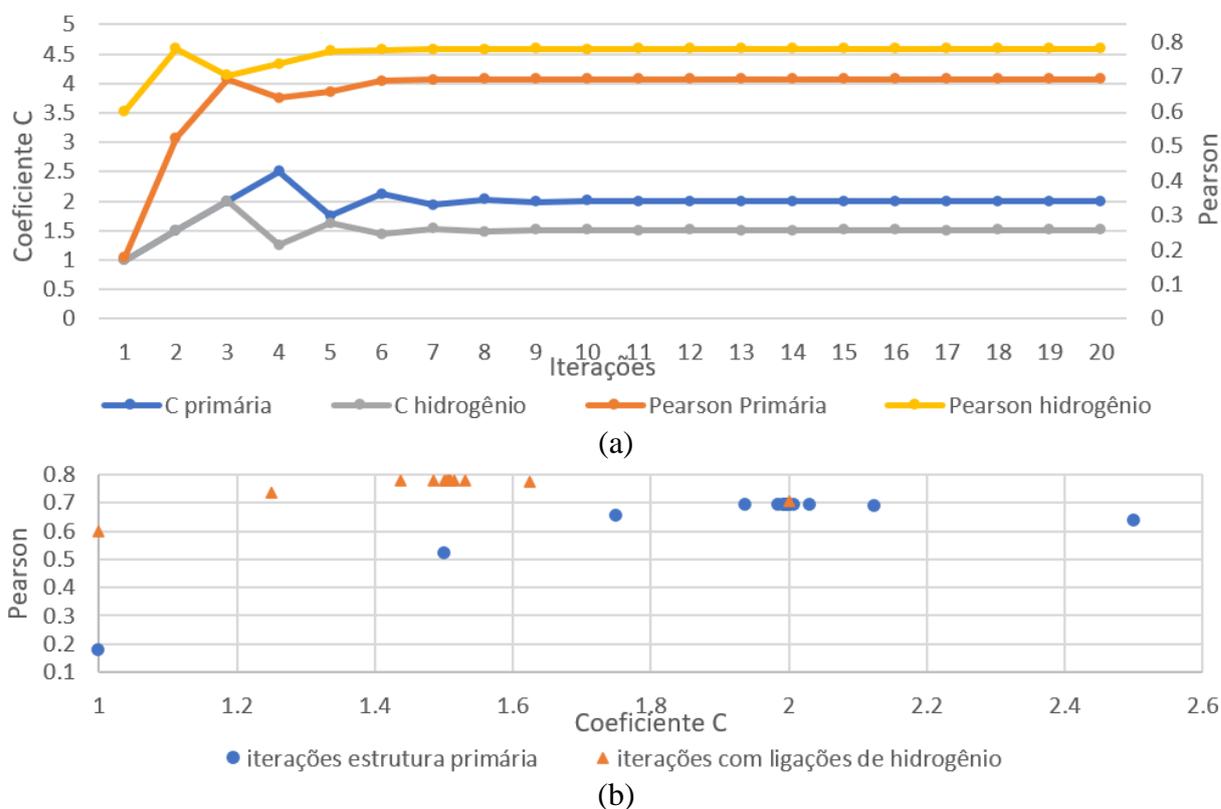


Figura B.1 (a) Gráfico do coeficiente C simulado (eixo à esquerda) e Coeficiente de $Pearson$ em relação ao experimental (eixo à direita) para cada iteração, utilizando o estado de conformação 1DPX. (b) Gráfico do coeficiente de $Pearson$ em relação ao Coeficiente Simulado, utilizando a configuração 1DPX.

1DPW:

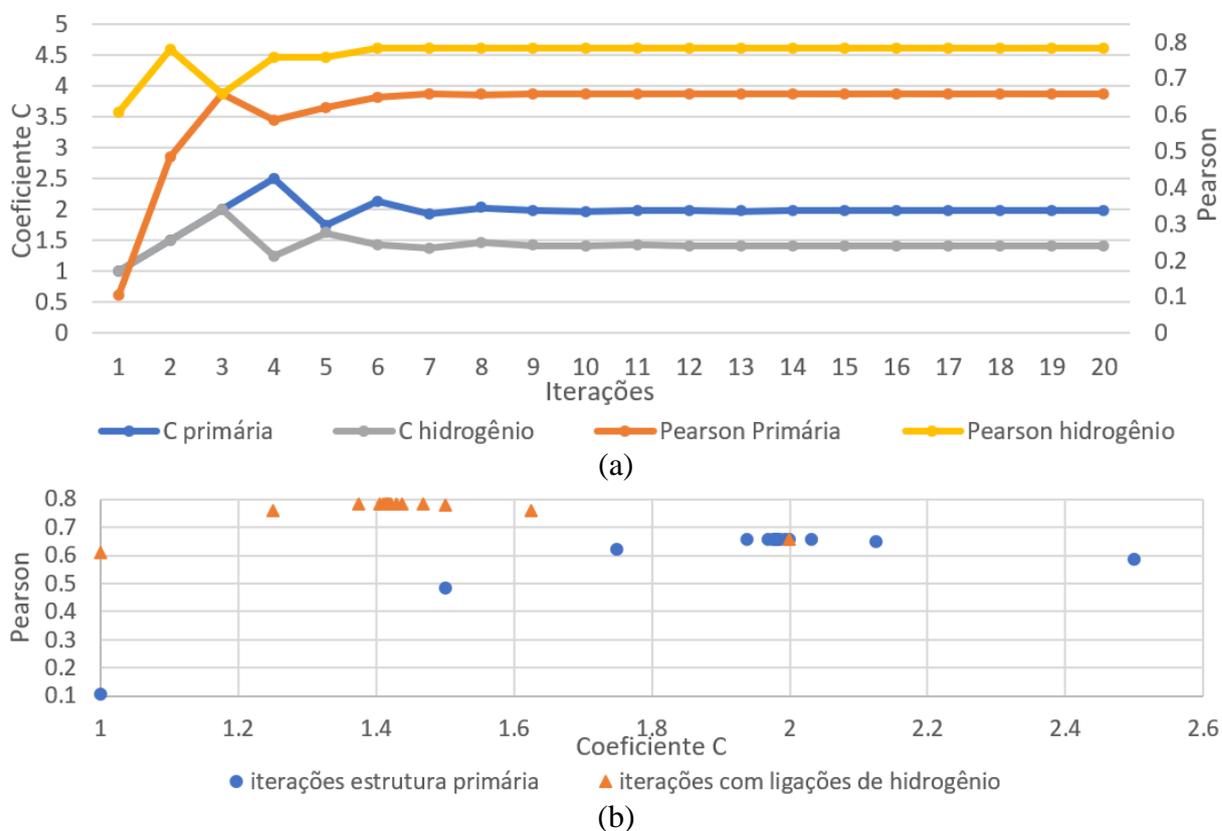


Figura B.2 (a) Gráfico do coeficiente C simulado (eixo à esquerda) e Coeficiente de *Pearson* em relação ao experimental (eixo à direita) para cada iteração, utilizando o estado de conformação 1DPW. (b) Gráfico do coeficiente de *Pearson* em relação ao Coeficiente Simulado, utilizando a configuração 1DPW.

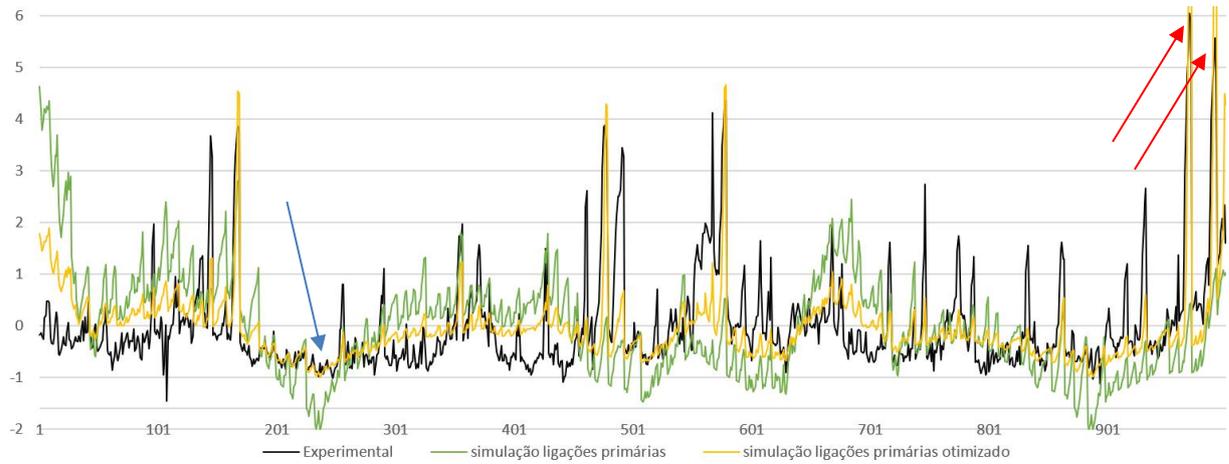
Apêndice C:

Figura C.1 - Fatores de temperatura experimental e simulados para cada átomo utilizando ligações primárias e primárias otimizado para o estado de conformação 1DPX.

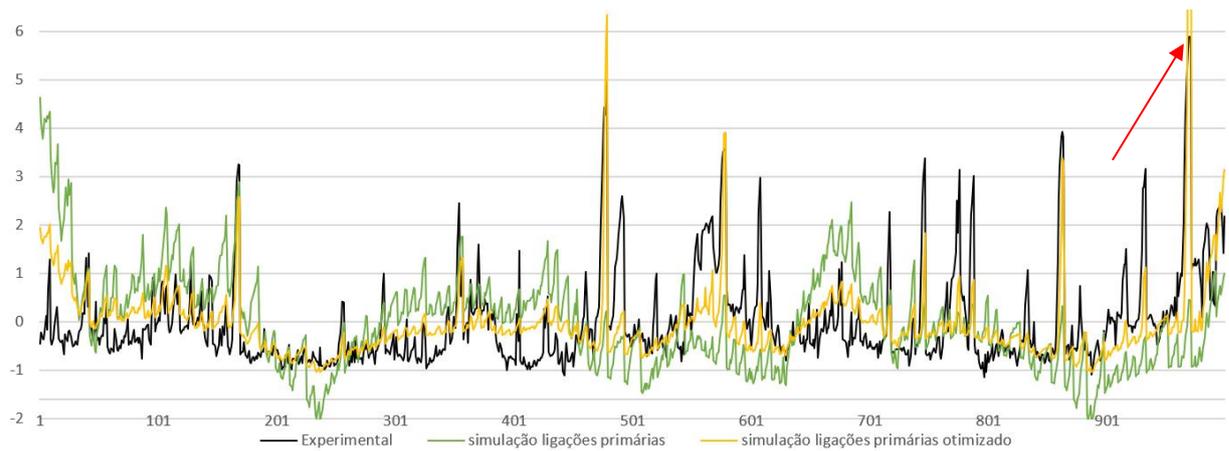


Figura C.2 - Fatores de temperatura experimental e simulados para cada átomo utilizando ligações primárias e primárias otimizado para o estado de conformação 1DPW.

Apêndice D:

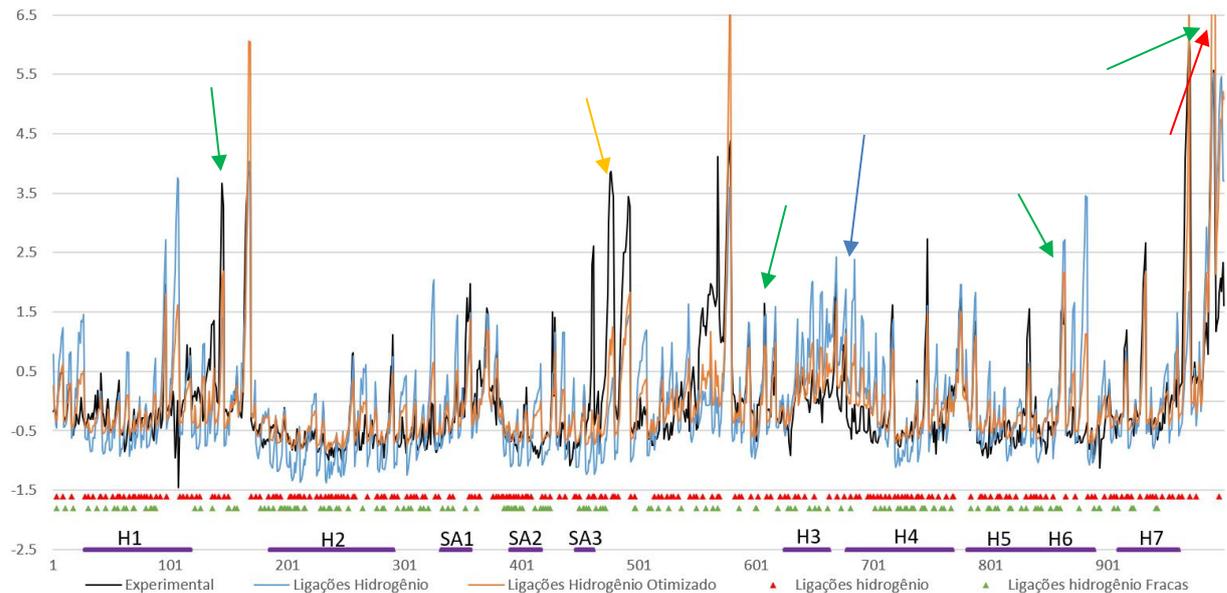


Figura D.1 – Fatores de temperatura experimental e simulados para cada átomo utilizando ligações de hidrogênio e otimizado para o estado de conformação 1DPX.

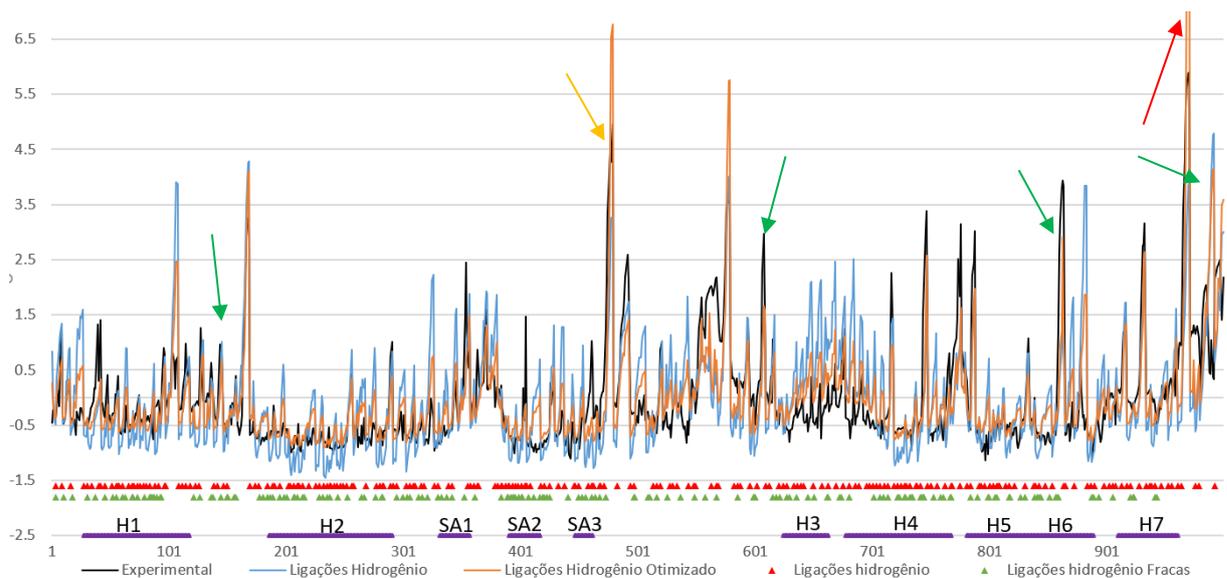
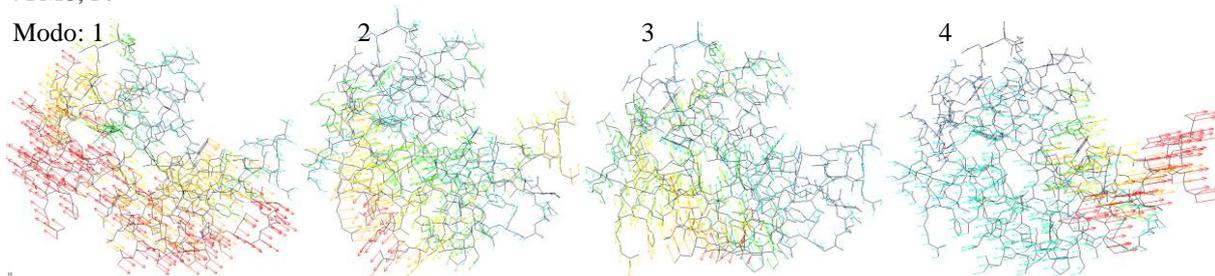


Figura D.2 – Fatores de temperatura experimental e simulados para cada átomo utilizando ligações de hidrogênio e otimizado para o estado de conformação 1DPW.

Apêndice E:

4YM8, P:

Modo: 1



1DPW, P:

Modo: 1

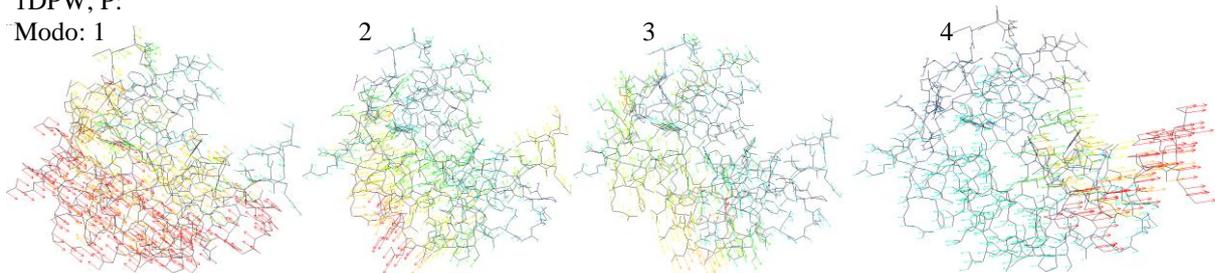
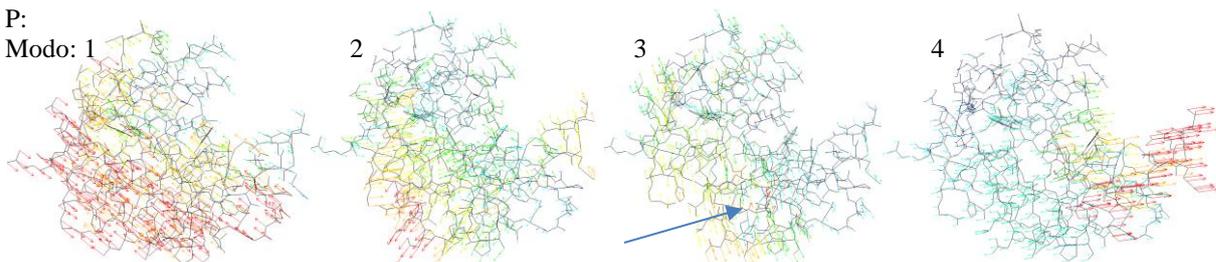


Figura E.1 – Modos de vibração para as simulações realizadas no estado de conformação 1DPW e 4YM8 utilizando a estrutura primária.

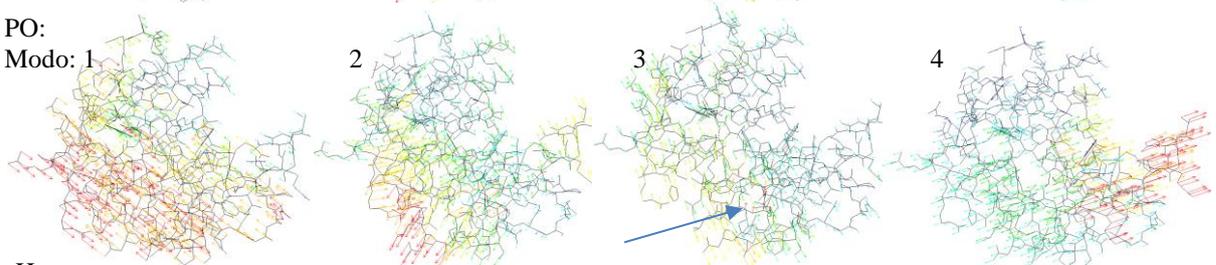
P:

Modo: 1



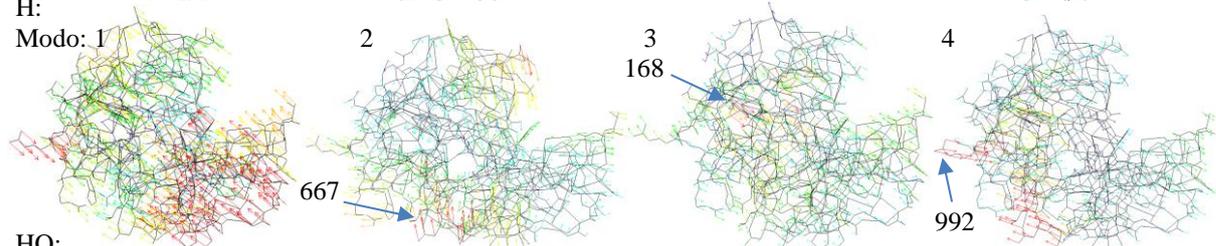
PO:

Modo: 1



H:

Modo: 1



HO:

Modo: 1

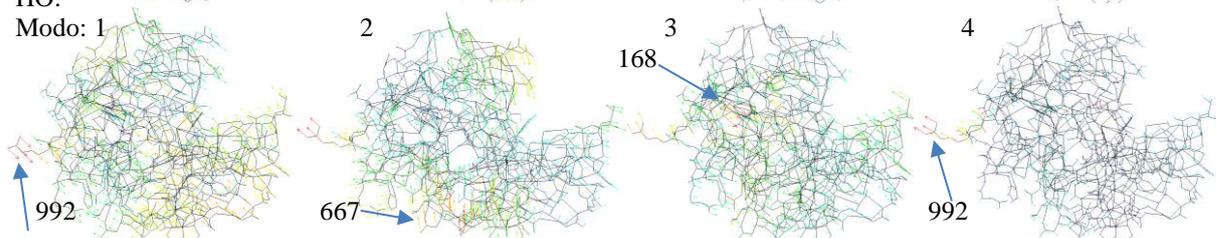


Figura E.2 – Modos de vibração para as simulações realizadas no estado de conformação 1DPX.