



**VANESSA CADAN SCHEFFER**

**CORRELAÇÃO ENTRE MÉTRICAS ESTRUTURAIS PARA  
PROTEÍNAS E AJUSTES DE CURVAS DE SAXS POR  
MODOS NORMAIS**

**INCONFIDENTES-MG  
2014**

**VANESSA CADAN SCHEFFER**

**CORRELAÇÃO ENTRE MÉTRICAS ESTRUTURAIS PARA  
PROTEÍNAS E AJUSTES DE CURVAS DE SAXS POR  
MODOS NORMAIS**

Trabalho de Conclusão de Curso apresentado como pré-requisito de conclusão do curso de graduação Licenciatura em Matemática no Instituto Federal de Educação Ciência e Tecnologia do Sul de Minas Gerais - Câmpus Inconfidentes, para obtenção do título de Licenciado em Matemática.

Orientador: Prof. Dr. Marcelo Augusto dos Reis

**INCONFIDENTES-MG  
2014**

VANESSA CADAN SCHEFFER

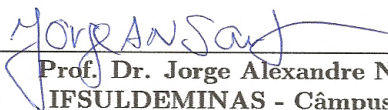
**CORRELAÇÃO ENTRE MÉTRICAS ESTRUTURAIS PARA  
PROTEÍNAS E AJUSTES DE CURVAS DE SAXS POR  
MODOS NORMAIS**

Data de Aprovação: 12 de Novembro de 2014



---

**Prof. Dr. Marcelo Augusto dos Reis (Orientador)**  
IFSULDEMINAS - Câmpus Inconfidentes



---

**Prof. Dr. Jorge Alexandre Nogueira Santos**  
IFSULDEMINAS - Câmpus Inconfidentes



---

**Prof. MSc. Flávio Fernandes Barbosa Silva**  
IFSULDEMINAS - Câmpus Inconfidentes

## DEDICATÓRIA

Dedico esse trabalho à minha querida mãe que nunca mediu esforços, muitas vezes abdicando de seus interesses, a fim de me proporcionar uma formação de qualidade.



## AGRADECIMENTOS

- Agradeço à minha família e amigos pelo apoio e compreensão nos momentos difíceis. Não poderia deixar de citar alguns nomes especiais em minha vida: Ana Maria, Hudson, Marcos, Gisele, Bernadete e Rafael, à todos vocês minha gratidão.
- À todos professores que fizeram diferença nessa caminhada, principalmente à Lidiane Teixeira, quem incentivou e acompanhou os primeiros trabalhos desenvolvidos.
- Ao orientador desse trabalho pela atenção, dedicação e paciência. Sinto grata pela oportunidade de desenvolver um projeto que vai além de cumprir com a obrigação burocrática. Agradeço além da aprendizagem que me proporcionou, a preocupação em mostrar novos caminhos.
- Aos colegas das turmas em que tive oportunidade de estudar. Patrícia, Leonardo e Kátia muito obrigada pelo apoio, sentirei saudades dos momentos ímpares que vivemos.
- Ao IFSULDEMINAS pelas grandes oportunidades vivenciadas durante essa formação.
- Ao IFSULDEMINAS - Câmpus Inconfidentes, por meio do Núcleo Institucional de Pesquisa e Extensão - NIPE, pelo recebimento de bolsa de iniciação científica, através da qual foi possível realizar parte desse trabalho.

# Resumo

Proteínas são macromoléculas biológicas que desempenham funções vitais em um organismo, tais como: catalisar reações, estruturar partes do corpo, regular o metabolismo e atuar no sistema imunológico. A função biológica geralmente está ligada a forma tridimensional da molécula. Para um grupo específico de proteínas a função biológica pode estar ligada aos movimentos estruturais. Portanto, estudar métodos computacionais inerentes ao esclarecimento dessa relação função-estrutura se faz relevante.

Embora métodos computacionais estejam disponíveis atualmente, a falta de tutoriais e literatura adequada para iniciantes do campo da matemática aplicada não foram encontrados.

Neste trabalho, duas abordagens foram levadas em conta a fim de contribuir com essa literatura básica. Em primeiro lugar, concentramos nossos esforços para tentar introduzir o campo da bioinformática para todos aqueles cuja base se encontra no campo da matemática. Por outro lado, utilizamos ferramentas computacionais para caracterizar estruturas proteicas utilizando duas estratégias diferentes.

O primeiro método é conhecido por espalhamento de raios X a baixos ângulos (SAXS). Essa é uma técnica experimental cujo resultado sobre as estruturas proteicas é de baixa resolução. Neste trabalho realizamos simulações computacionais baseadas em estruturas de proteínas a fim de obter seus respectivos perfis de SAXS.

*Template Modeling Score* (TM-score) foi o segundo método computacional usado. Através dessa técnica duas estruturas de proteínas foram comparadas. Essa técnica apresenta resultados de maior resolução que os obtidos por SAXS.

Esses métodos foram simulados para um grupo específico de proteínas que exibem conformação em estado aberto e fechado. Através do coeficiente de correlação de Pearson verificamos que ambas técnicas apresentam alto grau de correlação, sendo seu valor na ordem de -0,71.

A fim de prever possíveis conformações entre estado aberto e fechado foram realizados um conjunto de simulações computacionais baseadas em modos normais de proteínas comparando-as através de perfis de SAXS. Encontramos resultados em duas proteínas, que os melhoramentos dos perfis de SAXS foram da ordem de 42% e 64% respectivamente.

Palavras-chave: Proteínas, modelagem matemática, estruturas tridimensionais.

# Abstract

Proteins are macromolecules involved in several functions in live organisms, for example: catalysis, structural support, metabolism and immune system. Usually, biological activity is related to their three-dimensional structure. There are a group of proteins which may exhibit large structural motions for performing biological function. Therefore, it is imperative dealing with methods and computational approaches in order to shed light on the function-structure relationships.

Although some computational methods are currently available, the lack of tutorials and proper literature designed for beginners from applied mathematics field have not been found.

In this work, one addresses all these issues by using both approaches. In one hand, our efforts were driven by trying to introduce the bioinformatic field for all those whose background lies in the mathematics field. On the other hand, we have used computational tools in order to characterize proteins structures using two diferents strategies.

The first method is known as Small Angle X-ray Scattering (SAXS). In fact, this is an experimental technique which provides low resolution information about any protein structure. In our work, we have performed computer simulations based on proteins structures in order to get their correspondent SAXS profiles. Template Modeling Score (TM-score) is the second computational approach we have used to compare two proteins structures each other. It may be considered as higher resolution technique than SAXS.

These methods were performed to a selected group of proteins which are known for exhibiting open and closed conformations. From Pearson's correlation coefficient (PCC) between these two computational methods, our results show these two techniques are high correlated with  $PCC = - 0.71$ .

In order to predict possibles conformations between open and closed states we have launched a set of computer simulations based on normal modes of a protein structure against SAXS profile. As a result, we have been found two proteins where there were improvement in the SAXS agreement bewteen open and closed states about 42% and 64%, respectively.

Key words: Proteins, mathematical modeling, three-dimensional structures.

# Conteúdo

Resumo . . . . .	i
Abstract . . . . .	ii
Lista de figuras . . . . .	v
Lista de tabelas . . . . .	vii
<b>1 Introdução</b>	<b>1</b>
1.1 Modelagem matemática . . . . .	2
1.2 Proteínas . . . . .	3
1.3 Bioinformática . . . . .	4
1.4 Objetivos e organização do trabalho . . . . .	5
1.4.1 Objetivos Gerais . . . . .	5
1.4.2 Objetivos específicos . . . . .	5
1.4.3 Organização dos capítulos . . . . .	5
<b>2 Revisão da literatura</b>	<b>7</b>
2.1 Fundamentos biológicos . . . . .	7
2.1.1 Ligações Peptídicas . . . . .	9
2.1.2 Formas estruturais das proteínas . . . . .	11
2.2 Fundamentos físicos e matemáticos . . . . .	12
2.2.1 Determinação experimental de estruturas proteicas . . . . .	12
2.2.2 Luz síncrotron . . . . .	13
2.2.3 SAXS ( <i>Small Angle X-ray Scattering</i> ) . . . . .	15
2.2.4 Modelagem computacional por modos normais . . . . .	16
2.2.5 Oscilador Harmônico Simples . . . . .	16
2.2.6 Origem dos modos normais . . . . .	19
2.2.7 Rede elástica para proteínas . . . . .	21
2.2.8 Modos normais de vibração . . . . .	22
<b>3 Materiais e métodos</b>	<b>24</b>
3.1 Programas e dados utilizados . . . . .	24
3.2 Perfil simulado de SAXS . . . . .	25
3.2.1 Caracterização de estruturas utilizando SAXS . . . . .	26
3.3 <i>Template Modeling Score</i> (TM-score) . . . . .	26
3.3.1 Análise qualitativa da pontuação do TM-score . . . . .	27
3.4 Coeficiente de correlação de Pearson . . . . .	28
3.5 Modos normais de vibração e perfis de SAXS . . . . .	29
<b>4 Resultados e discussões</b>	<b>31</b>
4.1 TM-score <i>versus</i> SAXS-score . . . . .	31
4.2 Ajuste dos perfis de SAXS com auxílio dos modos normais . . . . .	35

<b>5</b>	<b>Conclusões</b>	<b>40</b>
<b>A</b>	<b>Proteínas, TM-score e SAXS-score</b>	<b>42</b>
	Tabela: proteínas, TM-score e SAXS-score . . . . .	42
<b>B</b>	<b>Modos normais de vibração e perfis de SAXS</b>	<b>45</b>
B.1	<i>LAO binding protein</i> . . . . .	45
B.2	<i>Lactoferrin</i> . . . . .	46
B.3	<i>Maltodextrin binding protein</i> . . . . .	47
B.4	<i>Dihydrofolate reductase</i> . . . . .	47
B.5	<i>Triglyceride lipase</i> . . . . .	48
B.6	<i>Citrate synthase</i> . . . . .	49

# Lista de Figuras

1.1	Esquema geral para modelagem matemática . . . . .	2
2.1	Estrutura geral de um aminoácido. . . . .	8
2.2	Aminoácidos . . . . .	8
2.3	Esquema geral de uma ligação peptídica. . . . .	10
2.4	Esquema de um pentapeptídio. . . . .	10
2.5	Representação dos quatro níveis estruturais de proteínas. . . . .	12
2.6	Laboratório Nacional de Luz Síncrotron (LNLS). . . . .	14
2.7	Esquema geral de um experimento de SAXS. . . . .	15
2.8	Perfil de um experimento de SAXS. . . . .	16
2.11	Sistema harmônico acoplado. . . . .	20
3.1	Comparação entre dois perfis simulados de SAXS. . . . .	26
3.2	Rede elástica da proteína <i>Triose phosphate isomerase</i> . . . . .	29
3.3	Simulação de um modo normal de vibração. . . . .	30
4.1	Resultados obtidos pelo TM-score e SAXS-score . . . . .	32
4.2	Perfil simulado de SAXS e estrutura tridimensional da proteína <i>Triose phosphate isomerase</i> . . . . .	33
4.3	Perfil simulado de SAXS e estrutura tridimensional da proteína <i>Adenylate kinase</i> . . . . .	34
4.4	Perfil simulado de SAXS e estrutura tridimensional da proteína <i>Diphtheria toxin</i> . . . . .	34
4.5	Perfil simulado de SAXS e estrutura tridimensional da proteína <i>Calmodulin</i> . . . . .	35
4.6	Perfil de SAXS para os 4 primeiros modos normais de vibração da proteína <i>Diphtheria toxin</i> . . . . .	36
4.7	Perfil de SAXS para o modo 8 da proteína <i>Diphtheria toxin</i> . . . . .	37
4.8	Perfil de SAXS para o modelo -21 do modo 8 da proteína <i>Diphtheria toxin</i> . . . . .	38
4.9	Perfil de SAXS para os 4 primeiros modos normais de vibração da proteína <i>Calmodulin</i> . . . . .	38
4.10	Perfil de SAXS para o modo 8 da proteína <i>Calmodulin</i> . . . . .	39
4.11	Perfil de SAXS para o modelo 19 do modo 8 da proteína <i>Calmodulin</i> . . . . .	39
B.1	Perfis de SAXS para a proteína <i>LAO binding</i> . . . . .	46
B.2	Perfis de SAXS para a proteína <i>Lactoferrin</i> . . . . .	46
B.3	Perfis de SAXS para a proteína <i>Maltodextrin binding</i> . . . . .	47
B.4	Perfis de SAXS para a proteína <i>Difhydrofolate reductase</i> . . . . .	48
B.5	Perfis de SAXS para a proteína <i>Triglyceride lipase</i> . . . . .	48
B.6	Perfis de SAXS para a proteína <i>Adenylate kinase</i> . . . . .	49



# Lista de Tabelas

2.1	Aminoácidos essenciais e não essenciais . . . . .	9
A.1	Tabela de proteínas com seus respectivos valores de SAXS-score e TM-score	44



# Capítulo 1

## Introdução

Como consequência do advento das tecnologias da informação têm-se a automação de uma grande quantidade de processos que antes eram possíveis de se realizarem somente manualmente. Como ponto positivo, destacam-se as pesquisas e descobertas em diversas áreas que vão desde à qualidade de vida das pessoas até avanços tecnológicos que permitem compreender cada vez mais o universo. Porém, as mudanças não se restringem aos processos, elas atingem o cotidiano das pessoas criando um novo perfil de homem.

Esse novo homem tem acesso à calculadora e computador que são capazes de realizar os mais diversos e complexos cálculos que ele necessita, causando uma acomodação desse indivíduo no sentido dele não necessariamente ter que aprender matemática para conseguir cumprir uma determinada tarefa. Nesse contexto, faz-se o questionamento sobre qual é o papel da matemática na formação do aluno. Na linguagem dos alunos: *“Para que serve essa matéria professor?”*.

Segundo o dicionário HOUAISS, licenciatura é *“Grau universitário que dá o direito de exercer o magistério do segundo segmento do ensino fundamental e do ensino médio”*. A licenciatura deve proporcionar que o futuro docente amplie sua bagagem de conhecimento e se sinta motivado a aperfeiçoar-se em determinada área de seu interesse, pois um professor motivado e com conhecimentos sólidos possui o potencial para entrar numa sala de aula e lecionar de uma forma responsável. Esse profissional será capaz de fazer ligações do conteúdo a ser ensinado com o cotidiano de seus alunos mostrando que matemática está além das fórmulas, matemática faz parte do cotidiano até mesmo em situações que aparentemente não possuem nenhuma ligação direta com a matemática.

A utilização da matemática para tratar problemas reais é o que deu origem a matemática que conhecemos hoje. Foram os problemas encontrados pelos primeiros homens que motivaram a busca pelo conhecimento, dando origem ao que conhecemos por modelagem matemática.

## 1.1 Modelagem matemática

Algumas situações do mundo real são intrincadas e difíceis de serem estudadas, seja por limitação de equipamentos, seja pela localização, pelo custo, dentre outros fatores. Para realizar estudos complexos utilizam-se modelos e realizam-se simulações sobre seu comportamento a fim de entender os mecanismos e prever futuras ações desses sistemas. Diversas áreas utilizam modelos tais como física, química, biologia, economia, engenharias, seja qual for a área a modelagem do problema se dará através de modelos matemáticos. Segundo STEWART, 2013, p.22, [1] “Um modelo matemático é a descrição matemática de um fenômeno do mundo real. O propósito desses modelos é entender o fenômeno e talvez fazer previsões sobre seu comportamento futuro”.

Pode-se dizer então que modelagem matemática é a atividade da descrição de matemática de um problema a fim de encontrar padrões e soluções que condizem com o mundo real. A Figura abaixo mostra um esquema geral de procedimentos para uma modelagem matemática.

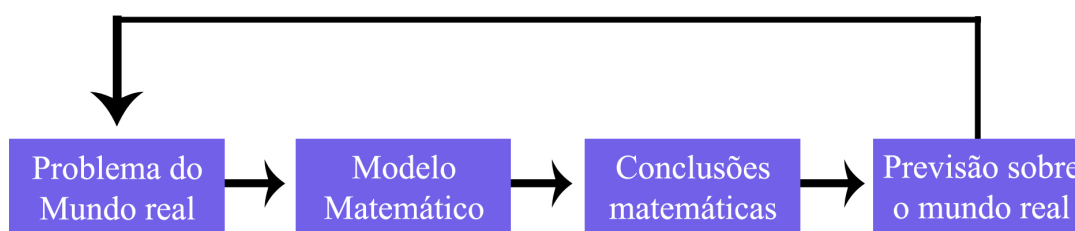


Figura 1.1: Esquema geral para modelagem matemática. A partir de um problema qualquer do mundo real, modelos matemáticos são estudados como uma simplificação do problema a fim de inferir possíveis resultados. Os resultados são comparados com a situação real e caso não sejam suficientes, o problema é reelaborado.

A modelagem matemática é um ambiente rico para exploração, e não se limita a simples realização de cálculos. Sua essência consiste em utilizar os cálculos para resolver um problema do mundo real, em outras palavras, pode-se dizer que a modelagem matemática traz sentido real aos algoritmos que muitas vezes são simplesmente decorados durante a formação do aluno.

Com o progresso da tecnologia, computadores cada vez mais poderosos estão à disposição das pessoas, permitindo que modelos matemáticos sejam implementados. Como resultado, tem-se *softwares e scripts* capazes de realizarem cálculos e simulações do fenômeno estudado em uma velocidade e precisão antes inimaginável.

## 1.2 Proteínas

O Dogma Central da Biologia Molecular foi postulado por Francis Crick em 1958. Ele explica como ocorre o fluxo de informações do código genético, e a origem das proteínas. Uma proteína é formada a partir dos ácidos nucleicos, mas um ácido nucleico não pode ser formado a partir de uma proteína.

Segundo esse dogma, o fluxo da informação genética segue o seguinte sentido: DNA  $\rightarrow$  RNA  $\rightarrow$  PROTEÍNAS. Porém sabe-se que algumas enzimas são capazes de utilizar o RNA para produzir DNA.

No DNA está contida a informação genética, que pode ser transcrito em moléculas de RNA. No processo de transcrição, uma molécula de DNA serve como molde para a criação de uma molécula de RNA. É nessa molécula de RNA que é encontrado o código usado para organizar a sequência de aminoácidos e formar as proteínas no processo de tradução. Esse processo consiste na união de aminoácidos, obedecendo à ordem de códons apresentados em um RNA mensageiro.

Segundo LEHNINGUER, 2011, [2] as proteínas são as mais abundantes macromoléculas biológicas encontradas nas células e desempenham funções vitais em um organismo. Estudar o funcionamento de um organismo vivo implica explorar as cadeias proteicas. As proteínas podem ser estudadas sob os enfoques biológico, químico e físico.

Desenvolver técnicas para análise de estruturas de proteínas é um desafio em ascensão, uma vez que estudos revelam a ligação entre a função biológica que essa desempenha com sua estrutura física e seu enovelamento

Existe a necessidade do desenvolvimento de técnicas robustas para a predição da estrutura de uma proteína a partir da sequência de aminoácidos. Isso representa um grande desafio, mas promete grandes recompensas em termos de entendimento da função proteica, identificação de doenças relacionadas com sequências proteicas anormais e desenho de drogas para alterar a estrutura ou a função de proteínas. (VOET, et al., p. 164 [3])

DIMA, THIRUMALAI, 2004, p. 2345 [4], citam a ascensão da pesquisa voltada à determinação das propriedades estruturais de macromoléculas, relacionando doenças às formas estruturais. Além disso, ressaltam a importância de identificar as proteínas em famílias, a fim de compreender as funções biológicas e aprimorar as técnicas utilizadas para predição estrutural.

Como as proteínas estão presentes em todos os organismos vivos, desde a menor e mais simples das bactérias até os seres humanos, estudar sua estrutura visando o esclarecimento do funcionamento biológico significa contribuir tanto para a saúde humana, como também para o combate de bactérias que degradam os organismos e o meio ambiente.

## 1.3 Bioinformática

A Bioinformática surgiu do avanço tecnológico que proporciona o surgimento de áreas cada vez mais específicas a fim de estudar certos fenômenos. Essa área se caracteriza na utilização de recursos e técnicas computacionais a fim de analisar fenômenos biológicos, combinando conhecimentos de química, física, biologia, ciência da computação e matemática. (BENTON, 1996, p. 262, [5]).

Os fundamentos teóricos, a relação da informática com as ciências biológicas são temas de debates. OUZOUNIS, 2002, p. 377 [6] discute a posição da computação com a ciência experimental. Para o autor, o termo *bioinformática* reflete a teoria e a prática da computação que embasam todos os aspectos da ciência biológica, desde os algoritmos até a aplicação destes em experimentos.

Esse trabalho aborda uma das áreas da bioinformática, a estrutural. São objetos de estudo da bioinformática estrutural moléculas que possuem forma tridimensional característica como, por exemplo, as proteínas. A fim de caracterizar as estruturas e compreender o comportamento e a interação entre as moléculas, *softwares* específicos são desenvolvidos para simular e prever tais situações.

As maiores esperanças no campo das ciências médicas nas próximas décadas estão, sem dúvida, centradas no Projeto Genoma. Os pares de bases do genoma humano que foram sequenciados recentemente constituem um “rascunho” importante para a elucidação do genoma completo. Desafio maior, entretanto, será a elucidação estrutural de novos alvos moleculares, principalmente proteínas, enzimas e receptores, e novas drogas que certamente emergirão dos dados provenientes do genoma humano e de outros genomas em estudo. Estas estruturas serão a base da revolução da “medicina do futuro”, em que a compreensão dos fenômenos biológicos a nível molecular terá papel cada vez mais relevante. Esta nova medicina molecular certamente implicará em avanços significativos das técnicas de diagnóstico e tratamento. Com isso, espera-se que seja possível iniciar os tratamentos clínicos no feto ou na primeira infância, muito antes do surgimento dos primeiros sintomas. (FILHO, ALENCASTRO, 2003, p.253 [7])

Sob esse contexto, esforços têm sido feito em todo o mundo a fim de elucidar estruturas proteicas. A modelagem de proteínas por homologia tem sido utilizada nesse desafio. O termo homólogo refere-se a proteínas que derivam de um mesmo ancestral, e conseqüentemente possuem semelhanças estruturais. (FILHO, ALENCASTRO, 2003, p. 254 [7]). Partindo desse princípio, os métodos baseados em homologia utilizam como molde estruturas tridimensionais de proteínas já conhecidas e depositadas em bancos de dados de proteínas, e as comparam com novas estruturas obtidas experimentalmente ou por simulações computacionais.

## 1.4 Objetivos e organização do trabalho

### 1.4.1 Objetivos Gerais

Sendo este um trabalho de conclusão de um curso de licenciatura em matemática, busca-se através das noções de modelagem ligar a teoria aprendida em sala de aula com fenômenos reais, a fim de compreender com mais clareza e profundidade determinados conteúdos. A natureza do objeto de estudo remete ao ramo da bioinformática estrutural, na qual técnicas e conceitos inerentes à essa pesquisa são discutidas.

Sob o contexto da importância das proteínas para a vida humana surge a proposta do desenvolvimento desse trabalho, que utiliza informações disponíveis em bancos de dados de proteínas e recursos computacionais para inferir características sobre estruturas proteicas.

### 1.4.2 Objetivos específicos

- Aferir a correlação entre duas métricas distintas - TM-score e SAXS-score, a fim de caracterizar estruturas proteicas a partir dos resultados obtidos;
- Empregar modos normais de vibração em ajustes de perfis de SAXS, inferindo sobre possíveis conformações estruturais proteicas com características de estado aberto e fechado.

### 1.4.3 Organização dos capítulos

O trabalho está dividido em 5 capítulos, os quais tratam as informações consideradas relevantes dentro do escopo do desenvolvimento.

No Capítulo 2 é feita a revisão de literatura sobre os principais temas biofísicos pertinentes a esse trabalho. Para melhor apresentação o capítulo divide-se em duas partes, na primeira são tratados os principais fundamentos biológicos referentes ao objeto de estudo, as proteínas. A segunda parte trata dos fundamentos físicos inerentes às técnicas utilizadas, e os fundamentos matemáticos aplicados na solução do sistema.

No Capítulo 3 é apresentado os materiais e métodos utilizados para a caracterização das estruturas proteicas. Duas métricas foram empregadas para quantificar a qualidade da sobreposição de estruturas. O algoritmo TM-score considerado uma métrica de alta resolução e o SAXS-score, considerado de baixa resolução. Além disso, simulações baseadas nos modos normais de vibração foram realizadas.

O Capítulo 4 trata da discussão dos resultados obtidos entre as duas métricas. Utiliza-se o coeficiente de correlação de Pearson para determinar o grau de correlação entre ambas. É discutido também o emprego dos modos normais de vibração às duas

estruturas que apresentaram maiores diferenças estruturais segundo resultado obtido pelo SAXS-score.

O Capítulo 5 fecha o ciclo de análises ressaltando os principais pontos de cada técnica, evidenciando seus limites e potencialidades.

# Capítulo 2

## Revisão da literatura

Devido a natureza multidisciplinar do trabalho, a revisão de literatura se divide em duas partes. Na primeira seção são abordados os aspectos biológicos envolvidos com o objeto de estudo, enquanto a segunda seção trata dos aspectos físicos e matemáticos que embasam as simulações computacionais.

### 2.1 Fundamentos biológicos

Milhares de diferentes tipos de proteínas podem ser encontrados na natureza. Esses polímeros variam amplamente de tamanho, desde relativamente pequenos peptídeos a enormes polímeros com massas moleculares na ordem de milhares de Daltons<sup>1</sup>, e que podem ser encontradas em uma única célula.

Além disso, acredita-se que as proteínas sejam as biomoléculas mais versáteis nos organismos vivos, pois desempenham uma variedade de funções: catalíticas, enzimáticas, estrutural, receptor ou transporte de certas substâncias para as células. *Proteínas são instrumentos moleculares através dos quais a informação genética é reproduzida.* (LEHNINGHER, et al., 2011, p.75 [2])

Cada proteína, desde a mais simples até a mais complexa, é formada pela combinação de 20 aminoácidos-padrão. Pode-se dizer que essas 20 pequenas moléculas formam o alfabeto que escreve cada estrutura de proteína. Os aminoácidos são ligados covalentemente numa sequência linear específica<sup>2</sup>. Nem todas as proteínas contêm os 20 aminoácidos, eles podem ocorrer com frequências diferentes em proteínas distintas.

Os aminoácidos podem ser classificados pelo seu radical, em particular, pela sua

---

<sup>1</sup>Dalton é uma unidade de medida de massa utilizada para expressar a massa de partículas atômicas. É definida em função da massa de um átomo de carbono. Seu valor é 1/12 da massa do carbono, que equivale à  $1,67 \times 10^{-27}$  kg.

<sup>2</sup>Cada molécula que compõe a proteína é chamada de  $\alpha$  aminoácido. Essa convenção foi adotada devido a todos os aminoácidos possuírem a mesma estrutura básica, a qual é formada por um grupo carboxílico (COOH), um grupo amina (H<sub>2</sub>N), um átomo de hidrogênio e uma cadeia lateral também chamada de Radical ou simplesmente grupo R, todos ligados ao carbono alfa (C <sub>$\alpha$</sub> ).

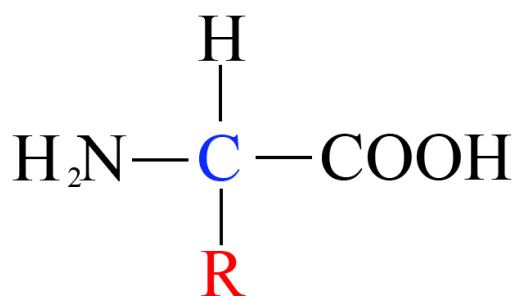


Figura 2.1: Estrutura geral de um aminoácido. Essa estrutura é comum a todos aminoácidos. O grupo R (em vermelho) ligado ao C<sub>α</sub> (em azul) é diferente em cada aminoácido.

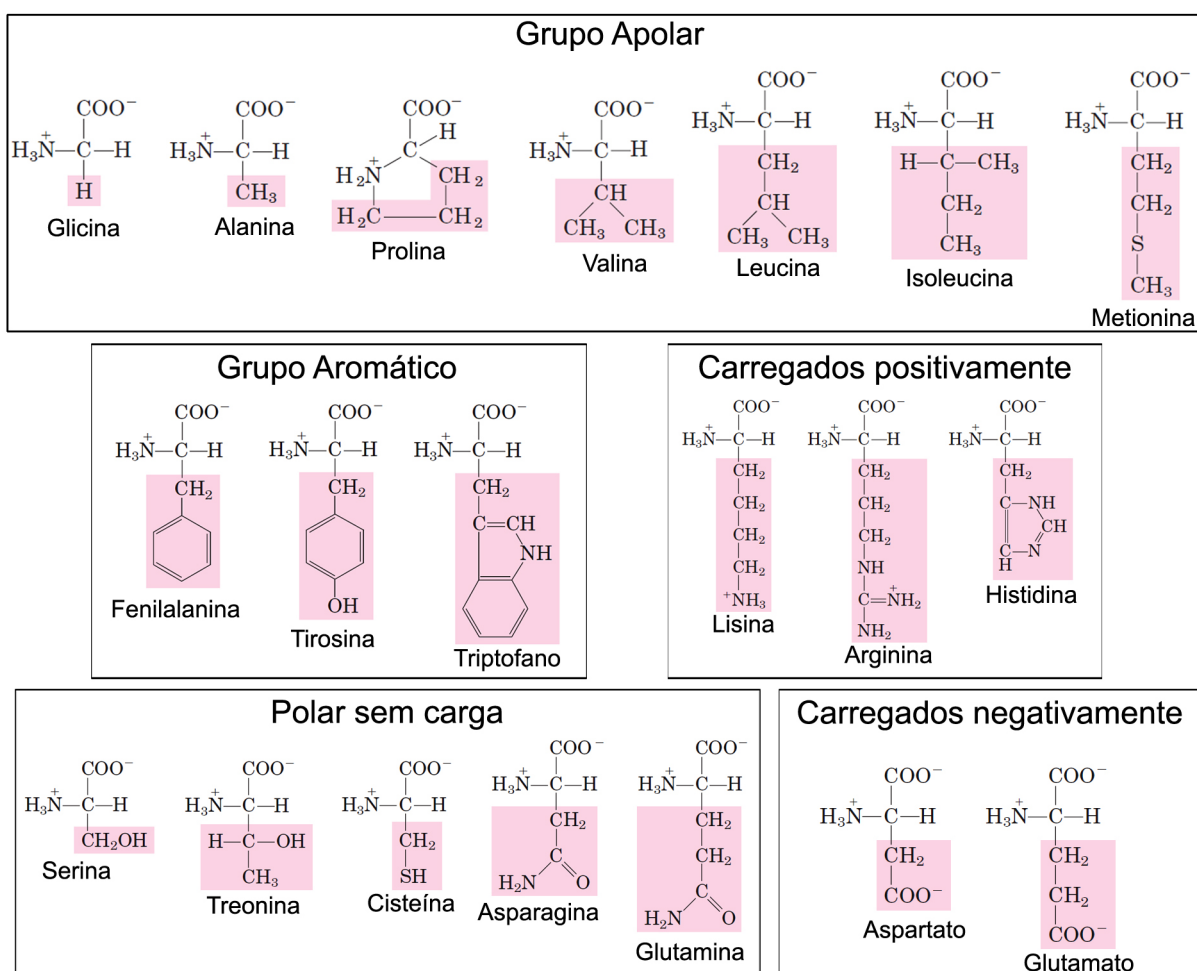


Figura 2.2: Forma estrutural dos 20 aminoácidos mais comuns encontrados na natureza. Estão divididos em grupos característicos.

polaridade ou tendência em interagir com a água em pH biológico (próximo AP pH 7,0). A polaridade do grupo R varia amplamente desde apolares e hidrofóbicos até altamente polares e hidrofílicos (solúveis em água), como mostra a Figura 2.2.

- Grupo apolar: Os radicais dos aminoácidos desse grupo são apolares e hidrofóbicos.



- Grupo aromático: Possuem cadeias laterais aromáticas, são relativamente não polares.
- Polar sem carga: Esses aminoácidos possuem em seus radicais moléculas mais solúveis em água que os do grupo não polar.
- Carregado positivamente: Os radicais dos grupos carregados positivamente ou negativamente são os mais hidrofílicos. Pertencem a esse grupo os radicais que apresentam carga positiva no pH 7.0.
- Carregado negativamente: Os radicais apresentam carga negativa no pH 7.0.

Embora o corpo humano necessite de todos os aminoácidos citados acima para produzir uma proteína, ele não é capaz de produzir todos. Aqueles que o organismo é capaz de sintetizar são chamados de não-essenciais, e os essenciais são os que necessitam ser ingeridos através dos alimentos.

Não essenciais	Essenciais
Alanina	Fenilalanina
Asparagina	Isoleucina
Cisteína	Leucina
Glicina	Lisina
Glutamina	Metionina
Serina	Treonina
Arginina	Triptofano
Histidina	Valina
Prolina	
Tirosina	
Aspartato	
Glutamato	

Tabela 2.1: Aminoácidos não essenciais são aqueles cujo organismo é capaz de sintetizar, e aminoácidos essenciais são obtidos através da ingestão de alimentos específicos.

### 2.1.1 Ligações Peptídicas

Duas moléculas de aminoácidos podem ser conectadas de forma covalente através de uma ligação peptídica, produzindo assim um dipeptídeo. Uma ligação peptídica é a união do grupo amino ( $H_2N$ ) de um aminoácido com o grupo carboxila ( $COOH$ ) de outro aminoácido. A síntese é feita retirando do grupo carboxila um agrupamento  $-OH$ , e do grupo amina retira-se um átomo de hidrogênio  $-H$ . Essa síntese é conhecida como desidratação, pois perde-se uma molécula de  $H_2O$ .

Centenas de aminoácidos podem ser conectados sequencialmente através de ligações peptídicas formando os polipeptídeos. Sendo assim os termos polipeptídeo e proteína são muitas vezes usados como sinônimo.

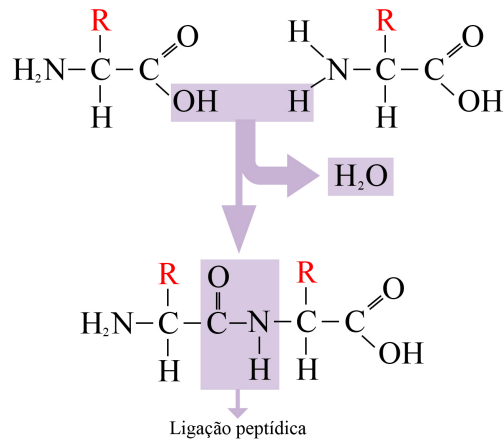


Figura 2.3: Esquema geral de uma ligação peptídica. O grupo R (em vermelho) ligado ao carbono  $\alpha$  representa o radical que varia em cada aminoácido. Em destaque a síntese da desidratação

Cada aminoácido num polipeptídeo é frequentemente chamado de resíduo (O termo resíduo reflete a perda do elemento da água quando ocorre uma ligação peptídica). O resíduo que possui um grupo amino livre é chamado aminoterminal ou N-terminal, e o resíduo que possui um grupo carboxílico livre é chamado de carboxiterminal ou C-terminal e esses dois resíduos indicam o início e o final da sequência de aminoácidos.

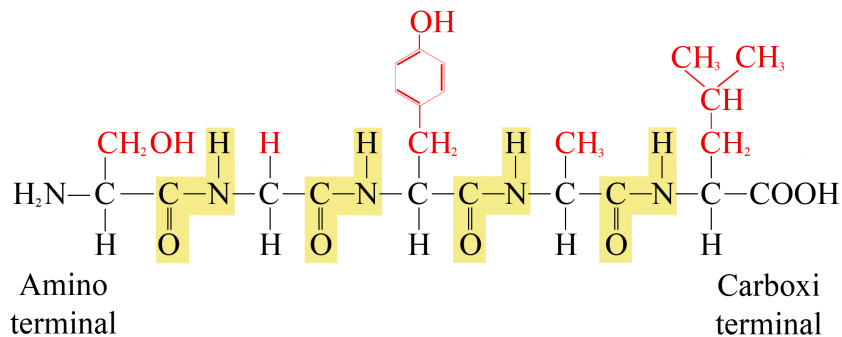


Figura 2.4: Esquema de um pentapeptídeo formado pela sequência de aminoácidos Ser-Gly-Tyr-Ala-Leu. O peptídeo inicia com o resíduo que representa o amino terminal que convencionalmente localiza-se do lado esquerdo e finaliza com o resíduo do carboxi terminal, localizado na outra extremidade. As ligações peptídicas estão sombreada em amarela. (Figura adaptada de LEHNINGUER, 2011, p. 86 [2]).

Cada átomo ocupa um lugar no espaço tridimensional e a disposição desses átomos em uma proteína é chamado de conformação. Esse termo é utilizado como sinônimo para estado estrutural de uma proteína, pois a conformação se refere a qualquer estado estrutural que a proteína pode assumir sem a quebra de suas ligações covalentes. Mudanças conformacionais podem ser induzidas em uma proteína devido à influência de seu ambiente químico ou de moléculas nas quais ela possui afinidades.

A função biológica de cada proteína está intrinsecamente ligada à disposição tridimensional dos aminoácidos.

Pesquisar e compreender as estruturas proteicas é colaborar com o avanço nas pesquisas que visam, por exemplo, detectar doenças genéticas, pois estas podem estar relacionadas às proteínas defeituosas do ponto de vista conformacional. *O defeito pode variar de uma única mudança na sequência de aminoácidos a deleções de uma grande parte da cadeia polipeptídica.* (LEHNINGHER, A. et al., 2011, p. 93 [2])

### 2.1.2 Formas estruturais das proteínas

Para decifrar a função biológica que uma proteína desempenha, é crucial conhecer sua estrutura. Seus níveis de organização se dividem em estrutura: primária, secundária, terciária e quaternária.

- **Estrutura primária:** É dada pela sequência de resíduos de aminoácidos ao longo da cadeia polipeptídica. É o nível estrutural mais simples, porém conhecer a sequência é um pré-requisito para determinar sua estrutura tridimensional. A estrutura primária da proteína resulta em uma longa cadeia de aminoácidos, com uma extremidade aminoterminal e uma extremidade carboxiterminal.
- **Estrutura secundária:** É dada pelo arranjo espacial de aminoácidos próximos entre si na sequência primária da proteína. Os arranjos secundários de uma cadeia polipeptídica são particularmente estáveis ocorrendo de forma regular através de ligações de hidrogênio graças à possibilidade de rotação das ligações entre os carbonos  $\alpha$  dos aminoácidos e os seus grupos amina e carboxílico. Cada tipo de estrutura secundária pode ser completamente descrita pelos ângulos diedros  $\phi$  e  $\psi$ , associados a cada resíduo. Dois tipos de estruturas secundárias mais conhecidas devido sua regularidade são as hélices  $\alpha$  e as conformações  $\beta$ . Quando nenhum desses padrões é observado diz-se que a estrutura é indefinida ou espiral aleatória.

As hélices  $\alpha$  e as conformações  $\beta$ , são as principais estruturas secundárias, elas compõem a estrutura de um grande número de proteínas. Sua regularidade se dá devido seus ângulos diedros característicos conhecidos por  $\phi$  e  $\psi$ . A maioria dos valores dos ângulos obtidos através de proteínas conhecidas cai em regiões esperadas, com alta concentração próximo aos valores preditos para as hélices  $\alpha$  e as conformações  $\beta$ .(LEHNINGUER, 2011, p. 122). A nomenclatura da  $\alpha$  hélice foi atribuída por Pauling e Corey, que determinaram ainda os valores dos ângulos  $\phi = -57^\circ$  e  $\psi = -47^\circ$ , e cada volta da hélice é formada por 3,6 resíduos de aminoácidos.

- **Estrutura terciária:** A estrutura terciária de uma proteína é o arranjo tridimensional de todos os átomos da molécula. Diferentemente da estrutura secundária na qual o arranjo espacial dos resíduos se dá entre os que são adjacentes em um segmento polipeptídico, a estrutura terciária inclui as cadeias laterais e as posições de quaisquer grupos, em outras palavras, a estrutura terciária possui uma abrangência

mais longa. Essa abrangência inclui tanto os resíduos que fazem parte da mesma cadeia polipeptídica, como os que estão em outra cadeia, isso ocorre devido ao dobramento que a proteína é capaz de adquirir na estrutura secundária.

- **Estrutura quaternária:** Algumas proteínas podem ter duas ou mais cadeias polipeptídicas distintas, ou subunidades que podem ser idênticas ou diferentes. O arranjo destas subunidades proteicas em complexos tridimensionais constitui a estrutura quaternária.

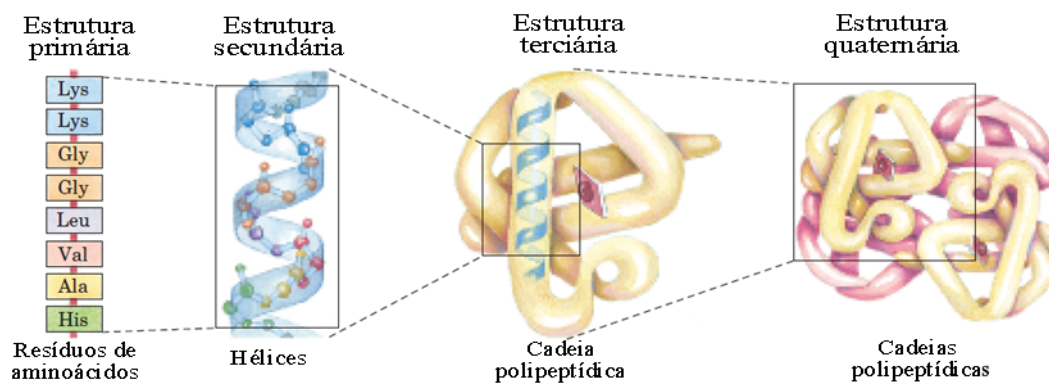


Figura 2.5: Representação dos quatro níveis estruturais de proteínas adaptada de LEHNINGUER. A estrutura primária formada pela sequência dos aminoácidos; Secundária sendo representada por uma  $\alpha$  hélice; Terciária é a continuação do enovelamento das estruturas secundárias; Quaternária representa a união de estruturas terciárias.

## 2.2 Fundamentos físicos e matemáticos

Para a descrição e estudo de fenômenos do mundo real usa-se teorias físicas a fim de compreender tais fenômenos e fazer previsões sobre comportamentos futuros. As teorias físicas dão sentido ao estudo de um fenômeno através de resultados obtidos por modelos matemáticos. Embora sabe-se que “*um modelo matemático nunca é uma representação completamente precisa de uma situação física - é uma idealização*” (STEWART, 2013, p. 13 [1]), através de modelos bem elaborados e resultados analisados cautelosamente é possível inferir afirmações e previsões sobre o problema estudado. Esses modelos permitem a simulação do fenômeno, e conseqüentemente seu estudo sob baixo investimento.

Tomando as estruturas proteicas como objeto de estudo, os conceitos físicos utilizados para caracterização estrutural desse trabalho são apresentados nesta seção.

### 2.2.1 Determinação experimental de estruturas proteicas

O panorama atual de pesquisa na área de biofísica encontra-se em amplo desenvolvimento, isso se deve ao aumento significativo de pessoas oriundas de diversas

áreas como biologia, física, matemática e computação, que juntas vêm aprimorando e desenvolvendo novas técnicas para a resolução da estrutura proteica.

O reflexo desse esforço conjunto resulta no aumento do número de proteínas já resolvidas e disponíveis nos bancos de dados de proteínas. Um dos mais conhecidos e utilizado nesse trabalho pode ser acessado no endereço <http://www.rcsb.org/pdb/home/home.do>. Atualmente esse banco de dados conta com um acervo de cerca de 95 mil estruturas de proteínas. Para o estudo e resolução estrutural das proteínas, três técnicas têm sido amplamente empregadas tais como, a Cristalografia, a Ressonância Magnética Nuclear (NMR), e a Microscopia Eletrônica (EM). [8]. No banco de dados consultado, das 95 mil estruturas 89% foram resolvidas utilizando a Cristalografia, 10% usando a NMR e 1% usando a EM.

Como pode-se observar a Cristalografia é uma das técnicas mais utilizadas. Em VOET, 2008, [3] encontra-se uma breve descrição do processo.

A proteína passa por um processo de cristalização, e é exposta a um feixe colimado de raios X, e o padrão de difração resultante, que se forma pelas posições regularmente repetidas dos átomos no cristal, é registrado em um contador de radiação ou, atualmente com menos frequência, em um filme fotográfico. Os raios X usados nos estudos estruturais são produzidos por geradores de raios X de laboratório ou, com cada vez mais frequência, por um síncroton, um tipo de acelerador de partículas que produz raios X de intensidade muito maior. As intensidades de difração máximas, são então, utilizadas para construir matematicamente uma imagem tridimensional da estrutura do cristal por meio de métodos complexos. Os raios X interagem quase exclusivamente com os elétrons na matéria, não com o núcleo atômico, o que resulta, portanto, uma imagem da densidade eletrônica do objeto em estudo. Tais mapas de densidade eletrônica são usualmente apresentados com a ajuda de um computador gráfico como um ou mais conjuntos de contorno. (VOET, et al., 2008, p. 145 [3])

O crescimento significativo dos bancos de dados permite que novas técnicas possam ser desenvolvidas utilizando como ponto de partida essas informações já disponíveis.

## 2.2.2 Luz síncroton

A visão humana ocorre pela captação de luz visível - que é um tipo de radiação eletromagnética - espalhada pelos objetos. Da mesma forma para observar objetos nanométricos pode-se usar radiação em determinado comprimento de onda. Uma das opções utilizadas é a luz síncroton, produzida em grandes máquinas que aceleram elétrons.

A luz produzida por um acelerador de partículas é capaz de “enxergar” o interior da matéria. Segundo o *site*:

<http://www.cnpem.br/blog/tag/laboratorio-de-luz-sincroton/> (acessado em

02/09/2014), existem 60 fontes de luz síncrotron no mundo espalhadas em 17 países. O Brasil é o único país da América Latina que possui um laboratório capaz de produzir este tipo de radiação. Esse laboratório localiza-se na cidade de Campinas, interior de São Paulo, e faz parte do Centro Brasileiro de Pesquisa em Energia e Materiais (CNPEM). O laboratório surgiu com incentivo do Ministério da Ciência, Tecnologia e Inovação, porém não está vinculada a nenhuma universidade específica.

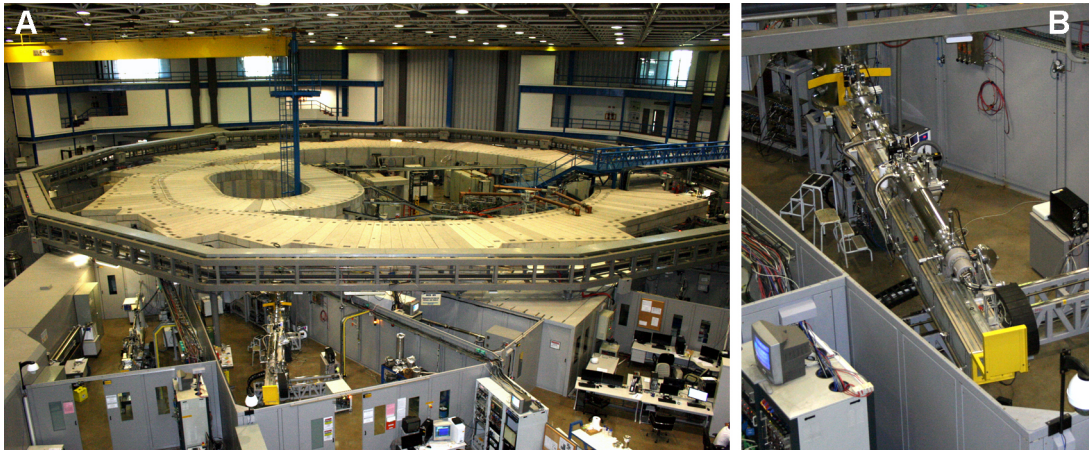


Figura 2.6: Laboratório Nacional de Luz Síncrotron (LNSL), localizado na cidade de Campinas, interior de SP. Do lado esquerdo, a imagem do anel de armazenamento onde os elétrons circulam a uma velocidade próxima à da luz. Do lado direito, a imagem de uma linha experimental SAXS. Fonte: a autora.

Um síncrotron é uma fonte de luz brilhante que os cientistas podem usar para coletar informações sobre as propriedades estruturais e químicas dos materiais em nível molecular. A luz é produzida dentro de um anel usando eletroímãs e ondas de radiofrequência que aceleram os elétrons a uma velocidade próxima à da luz. O anel é um polígono feito de sessões retas e ângulos de curvatura específicos a partir dos quais a radiação é emitida. A medida que os elétrons passam por esses ângulos, os ímãs fazem com que os elétrons percam energia gerando uma luz altamente brilhante e focalizada. Essa “luz”, que pode estar na região do infravermelho, ultravioleta e raios X, é então, canalizada para fora do anel de armazenamento e vai para as estações experimentais (linhas de luz), onde os pesquisadores escolhem o comprimento de onda desejado para estudar suas amostras.

A utilização de luz síncrotron está diretamente associada ao estudo da matéria e a análise de um grande número de processos biológicos, físicos, químicos, geológicos, entre outros. As informações obtidas podem ser usadas para variados fins como projetar novos medicamentos, desenvolvimento de óleos de motores mais eficazes, construção de pequenos *chips* de computadores mais potentes, desenvolvimento de novos materiais para implantes médicos mais seguros, limpeza de resíduos de mineração, além de outras utilidades.

### 2.2.3 SAXS (*Small Angle X-ray Scattering*)

O espalhamento de raios X a baixos ângulos (SAXS) é uma técnica experimental, utilizada para análise de estruturas de matérias. (SVERGUN, KOCH, 2003, p. 1737 [9]). SAXS está baseada nos mesmos princípios físicos básicos da Cristalografia. (BLUNDELL, JOHNSON, 1976 *apud* REIS, 2013, p. 29 [10, 11]).

Nessa técnica, uma proteína em solução é exposta a um feixe de raios X. Os elétrons dos átomos das proteínas interagem com a radiação incidente espalham-a até um detector. Esse esquema é demonstrado na Figura 2.7.

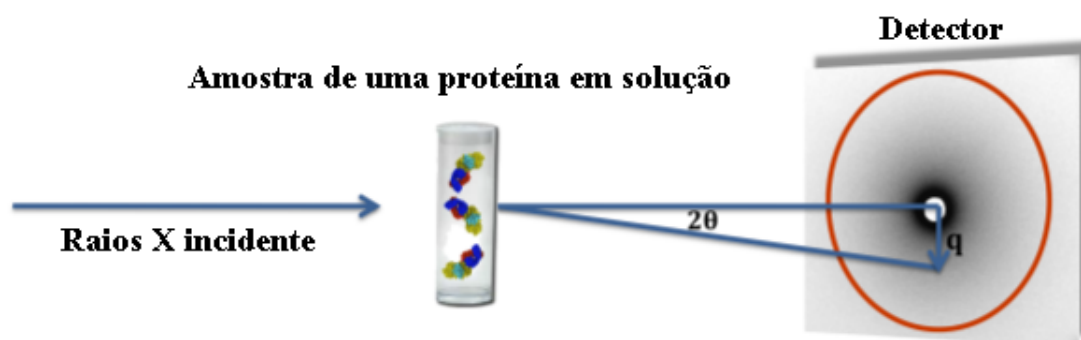


Figura 2.7: Esquema geral de um experimento de SAXS. O feixe de raios X incidente interage com a proteína espalhando-as por um ângulo de  $2\theta$  até atingir o detector.

O resultado é um perfil relacionado à intensidade do espalhamento em função do ângulo. (REIS, et al., 2011. e SVERGUN, KOCH, 2003. [12, 9])

Através de modelos físicos é possível obter uma análise das características geométricas globais da estrutura proteica. Esse resultado é de baixa resolução, pois SAXS não é capaz de fornecer as coordenadas atômicas, diferentemente da cristalografia. O perfil decorre do contraste do sinal entre a densidade eletrônica média da partícula e do seu ambiente químico. (SVERGUN, KOCH, 2003, p. 1737 [9])

A intensidade do espalhamento dos raios X é isotrópica, devido a proteína estar em solução, e caracteriza-se como mais intensa próximo ao ângulo do raio incidente, e diminui à medida que o ângulo vai aumentando, justificando o nome de espalhamento “a baixos ângulos”. A Figura 2.8 exemplifica um perfil de SAXS, no qual é possível observar que quanto maior o ângulo ( $q$ ), menor a intensidade.

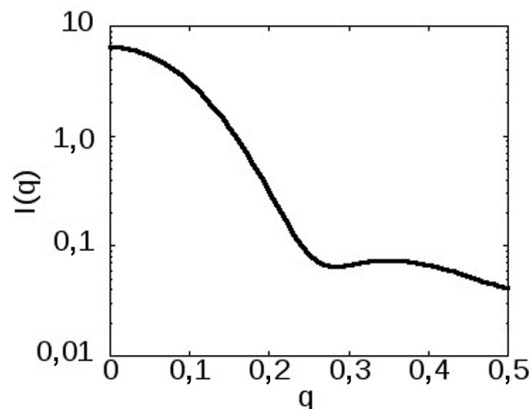


Figura 2.8: Perfil de um experimento de SAXS. A intensidade  $I(q)$  é maior para os ângulos  $q$  próximos ao feixe incidente.

## 2.2.4 Modelagem computacional por modos normais

A dinâmica molecular é a técnica tradicionalmente utilizada no estudo de estruturas proteicas, devido a seu alto grau de precisão, porém a dinâmica de uma proteína abrange uma ampla gama de escalas temporais: de femtossegundos a segundos, e escalas espaciais: de flutuações atômicas e movimentos coletivos de domínios, o que torna as simulações de dinâmica molecular (DM) inviáveis para o estudo de movimentos de larga escala, já que esta técnica trabalha com uma escala temporal tipicamente compreendida no intervalo de dezenas a centenas de nanossegundos, principalmente por limitações dos recursos computacionais disponíveis.

Diante dessas particularidades, surge a necessidade da elaboração de modelos simplificados, normalmente conhecidos por *coarse-grained*<sup>3</sup>. Esses modelos vêm ganhando espaço devido a facilidade de implementação, quando comparado aos modelos mais complexos, e ao baixo tempo que suas simulações têm despendido, sem no entanto, prejudicar a qualidade dos resultados.

## 2.2.5 Oscilador Harmônico Simples

A fim de compreender as técnicas inerentes às simulações computacionais de conformações proteicas o estudo do comportamento de um sistema harmônico se faz necessário.

O mundo está repleto de oscilações seja nos fenômenos naturais, como o vento por exemplo, sejam nos objetos e materiais. Todas as oscilações possuem um padrão normal de vibração que está associado às características físicas do objeto que vibra. Com

<sup>3</sup>Esses modelos trabalham com algum tipo de simplificação na estrutura da proteína. O modelo *coarse-grained* escolhido para esse trabalho, utiliza somente os  $C_\alpha$ . O carbono alfa em química orgânica refere-se ao carbono adjacente ao grupo funcional (o carbono é ligado a primeira posição, ou posição alfa, logo após o carbono que possui a função), este carbono é assimétrico, o que define o  $C_\alpha$  como um centro quiral e confere propriedades ópticas às moléculas.



isso torna-se possível estudar e conhecer esses padrões a fim de prever as ações que estão ligadas às vibrações dos objetos.

Considere o sistema da Figura 2.9. Uma partícula de massa  $m$  está presa a uma mola e está conectada a uma superfície fixa sobre o plano  $x$ , que deve ser considerado sem atrito. Nesse sistema a partícula se desloca proporcionalmente em relação à origem no sentido oposto à força aplicada, configurando assim um movimento harmônico simples. “Movimento harmônico simples é o movimento executado por uma partícula sujeita a uma força de módulo proporcional ao deslocamento da partícula e orientada no sentido oposto” (HALLIDAY, RESNICK, 2013, p. 92 [13]).

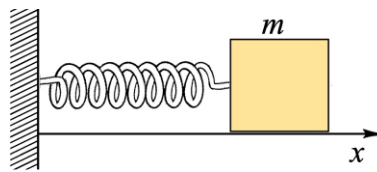


Figura 2.9: Ilustração de um sistema massa-mola. Uma massa  $m$  está presa a uma mola e esta conectada a uma superfície fixa sobre o plano  $x$  que deve ser considerado sem atrito. Esse sistema gera um movimento harmônico simples.

O movimento harmônico simples é representado pela expressão matemática da lei de Hooke:

$$F = -kx, \quad (2.1)$$

que depende de duas variáveis, a constante de força da mola  $k$  e o deslocamento da partícula  $x$ . O sinal negativo indica que o movimento é no sentido contrário à força aplicada. A oscilação harmônica, isto é, a função  $x(t)$  pode ser representada por um gráfico seno ou cosseno, pois todo movimento harmônico se repete num determinado intervalo de tempo. A Figura 2.10 mostra a função do deslocamento pelo tempo, onde  $x_m$  e  $-x_m$  é o máximo de deslocamento possível e nesses pontos a velocidade é zero.

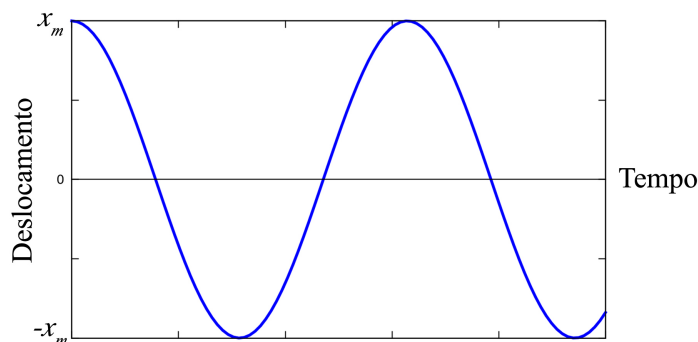


Figura 2.10: Função cosseno. O movimento harmônico simples pode ser representado por uma função cosseno do deslocamento da partícula pelo tempo. No ponto onde  $x = 0$ , a velocidade é máxima. O deslocamento se repete num determinado período que está relacionado com a frequência angular.

A segunda lei de Newton diz que:

$$F = ma, \quad (2.2)$$

onde  $F$  é a força,  $m$  a massa que pretende-se deslocar e  $a$  a aceleração.

Mas a aceleração é a derivada segunda do deslocamento em função do tempo:

$$a = \frac{d^2x}{dt^2}, \quad (2.3)$$

Igualando as Equações 2.1 e 2.3:

$$-kx = m \frac{d^2x}{dt^2}. \text{ Reescrevendo, } \frac{d^2x}{dt^2} + \frac{k}{m}x = 0, \text{ ou ainda:}$$

$$\ddot{x} + \frac{k}{m}x = 0. \quad (2.4)$$

A modelagem matemática de um sistema harmônico simples é dada por uma equação diferencial de segunda ordem, que relaciona a lei de Hooke com a segunda lei de Newton.

A resolução dessa equação diferencial se transforma na equação do deslocamento da partícula em função do tempo:

$$x(t) = x_m \cos(\omega t + \phi), \quad (2.5)$$

onde  $x_m, \omega$  e  $\phi$  são constantes.

A grandeza  $x_m$  denominada amplitude do movimento, é uma constante que representa o máximo de deslocamento possível em um dos sentidos do movimento. O deslocamento  $x(t)$  varia entre os limites  $\pm x_m$  (Figura 2.10). Esse é um parâmetro livre do sistema, ou seja, não é condição inicial fixa.

A grandeza dependente do tempo ( $\omega t + \phi$ ) é chamada de fase do movimento. A constante  $\phi$  é chamada de constante da fase ou ângulo da fase, e depende do deslocamento e da velocidade no instante  $t = 0$ , sendo também um parâmetro livre <sup>4</sup>.

A constante  $\omega$  é denominada frequência angular do movimento. Como o MHS é uma função senoidal, seu deslocamento se repete numa frequência de  $2\pi$  rad, pois  $x(t)$  deve ser igual a  $x(t + T)$ . Assim  $\omega = 2\pi f$ , e a unidade de frequência angular no SI<sup>5</sup> é o radiano por segundo.

---

<sup>4</sup>Toda equação diferencial de segunda ordem possui dois parâmetros livres.

<sup>5</sup>Sistema Internacional de Medidas

## Velocidade e aceleração do Oscilador Harmônico Simples

A velocidade de uma partícula em movimento harmônico simples é dada derivando a Equação 2.5 em relação ao tempo:

$$v(t) = -\omega x_m \text{sen}(\omega t + \phi). \quad (2.6)$$

A aceleração é calculada derivando a Equação da velocidade 2.6:

$$a(t) = -\omega^2 x_m \text{cos}(\omega t + \phi). \quad (2.7)$$

A aceleração é proporcional ao negativo do deslocamento e as duas grandezas estão relacionadas pelo quadrado da frequência angular. Combinando as Equações 2.5 e 2.7, obtêm-se:

$$a(t) = -\omega^2 x(t). \quad (2.8)$$

Conhecida a aceleração do Movimento Harmônico Simples é possível determinar qual é a força que atua no sistema. Usando a segunda lei de Newton 2.2 e a lei de Hooke 2.1 que são as forças conhecidas do sistema elástico, têm-se a constante da mola  $k$ :

$$\begin{aligned} F &= ma = -(m\omega^2)x. \\ F &= -kx. \\ -(m\omega^2)x &= -kx. \\ k &= m\omega^2. \end{aligned} \quad (2.9)$$

A frequência angular  $\omega$  do movimento harmônico simples está relacionada à constante elástica  $k$  e à massa  $m$  do bloco, sendo representada pela equação:

$$\omega = \sqrt{\frac{k}{m}}. \quad (2.10)$$

Combinando ainda a Equação 2.10 com a frequência angular  $\omega = 2\pi f$  têm-se o período:

$$T = 2\pi \sqrt{\frac{m}{k}}. \quad (2.11)$$

Todo sistema oscilatório possui uma certa “elasticidade” e uma certa “inércia” e, portanto, se assemelha a um oscilador linear tal como descrito nessa seção.

### 2.2.6 Origem dos modos normais

Um sistema massa-mola simples pode ser expandindo para inúmeras massas e molas, criando assim um grande sistema que vibra. Quanto maior o número de elementos, maior o número de equações necessárias para resolver o sistema.

A Figura 2.11, mostra um sistema harmônico acoplado no qual duas massas estão conectadas a três molas. Esse sistema é usado afim de exemplificar os modos normais de vibração.

Esse sistema descreve de forma análoga o comportamento de moléculas que possuem dinâmica linear, como por exemplo o dióxido de carbono  $\text{CO}_2$ .

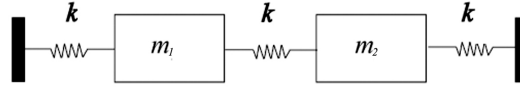


Figura 2.11: Sistema harmônico com duas massas e três molas. Após uma força ser aplicada no sistema as molas vibrarão dentro de padrões pré-estabelecidos. Esses padrões são os modos normais de vibração.

A energia potencial desse sistema é representada pela soma da energia potencial de cada mola.

$$V = \frac{1}{2} kx_1^2 + \frac{1}{2} kx_2^2 + \frac{1}{2} k(x_1 - x_2)^2, \quad (2.12)$$

onde  $k$  é a constante da mola, e  $x_1$ ,  $x_2$  representam os deslocamentos das massas  $m_1$ ,  $m_2$ .

A força desse sistema é dado pelo negativo do gradiente aplicado na energia potencial:

$$F_{x_1} = -\frac{\partial V}{\partial x_1} = -kx_1 - k(x_1 - x_2) = m \frac{\partial^2 x_1}{\partial t^2}.$$

$$F_{x_2} = -\frac{\partial V}{\partial x_2} = -kx_2 - k(x_2 - x_1) = m \frac{\partial^2 x_2}{\partial t^2}.$$

Reescrevendo as equações de força, obtêm-se um sistema de equações diferenciais, onde  $x_1$  e  $x_2$  dependem-se mutuamente:

$$\begin{cases} -2kx_1 + kx_2 = m \frac{\partial^2 x_1}{\partial t^2}, \\ kx_1 - 2kx_2 = m \frac{\partial^2 x_2}{\partial t^2}. \end{cases}$$

A resolução do movimento harmônico simples é dado pela Equação 2.5. O deslocamento é uma função cosseno. Resolvendo então o sistema de equações diferenciais, escolhe-se apropriadamente possíveis funções solução cuja, derivadas de segunda ordem continuam sendo as mesmas funções:  $x_1 = A_1 e^{i\omega t}$  e  $x_2 = A_2 e^{i\omega t}$ . A exponencial complexa escolhida como possível solução representa uma função cosseno, segundo a identidade de Euler ( $e^{ix} = \cos x + i \sin x$ ).

Reescrevendo então o sistema de equações diferenciais:

$$\begin{cases} -2kA_1 e^{i\omega t} + kA_2 e^{i\omega t} = -m\omega^2 A_1 e^{i\omega t}, \\ kA_1 e^{i\omega t} - 2kA_2 e^{i\omega t} = -m\omega^2 A_2 e^{i\omega t}. \end{cases}$$

Reescrevendo na forma matricial a fim de encontrar os valores de  $A_1$  e  $A_2$ :

$$\begin{bmatrix} -2k + m\omega^2 & k \\ k & -2k + m\omega^2 \end{bmatrix} \cdot \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Calculando o determinante dessa matriz em função de  $\omega$  obtêm-se os autovalores:

$$(-2k + m\omega^2)^2 = k^2$$

Os autovalores  $\omega$  representam as frequências e os autovetores os modos normais de vibração.

$$\begin{array}{ll} \omega_1 = \sqrt{\frac{k}{m}}, & \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \omega_2 = \sqrt{\frac{3k}{m}}, & \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ x_1 = A_1 e^{i\omega_1 t} & x_1 = e^{i\omega_1 t} \\ x_2 = A_2 e^{i\omega_1 t} & x_2 = e^{i\omega_1 t} \\ x_1 = A_1 e^{i\omega_2 t} & x_1 = e^{i\omega_2 t} \\ x_2 = A_2 e^{i\omega_2 t} & x_2 = -e^{i\omega_2 t} \end{array}$$

O primeiro autovalor e seu respectivo autovetor encontrado indica que ambas as massas se deslocam na mesma direção ( $A_1 = A_2$ ). Enquanto o segundo autovalor e seu respectivo autovetor indica que as massas se movimentam em sentidos opostos ( $A_1 = -A_2$ ). O movimento desse sistema pode ser descrito por qualquer combinação linear desses autovetores, ou seja, qualquer múltiplo da solução é solução também.

## 2.2.7 Rede elástica para proteínas

Expandindo a ideia inicial do oscilador harmônico simples, um sistema pode conter inúmeras massas, e inúmeras ligações criando uma grande rede elástica. Transpondo essa ideia para o escopo desse trabalho, os aminoácidos das proteínas são considerados “nós” de uma grande rede elástica, onde cada aminoácido é considerado uma “massa”, e as ligações entre eles são as “molas” que fazem com que o sistema vibre.

Existem diversos modelos de rede elástica possíveis de se implementar, porém cada um apresenta um grau de dificuldade e recursos computacionais específicos.

O modelo de rede elástica que foi utilizado nesse trabalho é o descrito por TIRION, 1996. Nesse modelo, a autora assume a conformação cristalográfica, obtida do *Protein Data Bank* (PDB), como sendo a conformação de equilíbrio já minimizada, eliminando assim uma das etapas mais críticas encontradas em outros modelos.

TIRION propõe que as interações entre os átomos sejam descritas por um potencial harmônico Hookiano, controlado por um simples parâmetro para todos os pares de átomos, assim a rede elástica proposta se transforma em um sistema harmônico.

O valor da constante de elasticidade entre cada aminoácido pode inclusive ser

igual para todas as interações. “*Esse modelo é suficiente para descrever os movimentos de baixa frequência que correspondem aos movimentos de grande amplitude de vibração.*” (TIRION, 1996, p. 4 [14])

## 2.2.8 Modos normais de vibração

Um modelo *coarse-grained* que vêm sendo utilizado é conhecido como Modos Normais de Vibração (NMA), do inglês *Normal Model Analysis*. Nesse modelo a simulação computacional da conformação proteica é feita calculando os modos normais de vibração do sistema, que dependem do modelo de rede elástica escolhida para implementação.

Segundo MENDONÇA, 2013, p. 17 [15], os primeiros estudos de NMA para proteínas foram reportados na década de 80. Os estudos foram realizados por meio de simulações de dinâmica molecular as quais utilizavam campos de forças altamente complexos. As coordenadas internas eram tomadas como as variáveis dinâmicas. Os modos normais e as frequências normais de vibração eram calculados via diagonalização da matriz Hessiana<sup>6</sup> (matriz das constantes de força). Pelo fato de que os elementos da matriz Hessiana são calculados na conformação de equilíbrio, era necessário efetuar a minimização da energia.

As flutuações conformacionais de proteínas são descritas pelo fator Debye-Waller (fator-B). Este fator corresponde à dinâmica local dos resíduos em nível atômico, e está associado à temperatura. Quanto maior for a temperatura maior será a dinâmica. Através das simulações computacionais dos modelos *coarse-grained* os fatores-B são reproduzidos em tempos computacionais curtos.

Expandindo o sistema acoplado descrito na Seção 2.2.6 para um sistema contendo  $N$  massas, a modelagem será dada por  $N$  equações diferenciais de segunda ordem. A solução desse sistema se dará transpondo os dados para uma matriz Hessiana. Dada uma função real de várias variáveis,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , a matriz Jacobiana que é a derivada do gradiente é denominado matriz Hessiana de  $f$ . A diagonalização da matriz Hessiana gera os autovetores, que são chamados de modos normais, e os respectivos autovalores estão associados às frequências de vibrações.

Disto decorre que uma vibração qualquer da estrutura de uma proteína pode ser descrita por uma combinação linear de todos os modos normais que formam uma base vetorial. A amplitude de movimento no modelo dos modos normais é arbitrária, embora todos os resíduos oscilem na mesma frequência em um dado modo.

Uma molécula contendo  $N$  átomos possui  $3N$  graus de liberdade, onde os 6 primeiros estão associados aos movimentos de translação e rotação de estruturas, portanto não predizem alterações das distâncias relativas entre aminoácidos. Grandes amplitudes

---

<sup>6</sup>Em matemática, a matriz Hessiana de uma função  $f$  de  $n$  variáveis é a matriz quadrada das derivadas parciais de segunda ordem da função.

vibracionais estão relacionadas às mais baixas frequências e portanto, aos primeiros modos normais que predizem alterações mais expressivas nas distâncias. O primeiro modo normal que descreve movimentos de distâncias entre átomos é o modo 7.

# Capítulo 3

## Materiais e métodos

Um grupo específico de proteínas possui característica estrutural de estado aberto e estado fechado. Esses estados, caracterizados pela estrutura tridimensional, estão associados à função biológica que a proteína desempenha. As reações químicas que acarretam essas conformações não fazem parte do escopo deste trabalho. Toda a discussão é feita sob o ponto de vista exclusivamente estrutural.

Pretende-se nesse trabalho de conclusão de curso exemplificar alguns casos de proteínas que exibem um comportamento de estado aberto e fechado procurando correlacioná-los sobre duas métricas e inferir possibilidades de conversão de uma estrutura na outra utilizando os modos normais de vibração. Isto se deu utilizando algoritmos específicos que realizam simulações computacionais.

### 3.1 Programas e dados utilizados

Para realização desse trabalho somente *softwares* livres foram usados. O sistema operacional escolhido foi o Linux Ubuntu versão 14.04. Os programas utilizados para gerar e diagonalizar a matriz Hessiana estão disponíveis para download no endereço <http://ecole.modelisation.free.fr/modes.html>. O algoritmo TM-score utilizado para comparar quantitativamente duas estruturas de proteínas pode ser encontrados no site <http://zhanglab.ccmr.med.umich.edu/TM-score/>. Os algoritmos SAXS-score e todos os *scripts* necessários para a automação dos cálculos foram fornecidos pelo orientador do projeto.

As estruturas proteicas utilizadas foram retiradas do banco de dados PDB disponível para acesso e *download* no endereço <http://www.rcsb.org/pdb/home/home.do>. Os arquivos que armazenam essas informações são simples arquivos de texto chamados de PDB, do inglês *Protein Data Bank*.

A partir da literatura TAMA, SANEJOUAND, 2001 [16], um grupo de 18 proteínas com características de estado aberto e fechado foram selecionados totalizando 36 estruturas.



Em posse dos 36 arquivos PDB, o primeiro passo foi remover todas as informações irrelevantes, deixando somente as informações sobre os  $C_\alpha$ . O segundo passo consistiu em aplicar as métricas a fim de comparar seus resultados. A primeira métrica utilizada basea-se na técnica de SAXS.

## 3.2 Perfil simulado de SAXS

A proposta desse trabalho é a caracterização de estruturas utilizando simulações computacionais. Sendo assim nenhum experimento de SAXS foi realizado em laboratório.

Tal estudo é possível porque essa técnica permite obter resultados através de simulações de perfis de SAXS, sendo assim estudos sobre proteínas podem ser realizados com recursos mínimos de pesquisa.

O perfil de SAXS é simulado através da Equação de Debye (DEBYE, 1915, *apud* REIS, 2013) [17]:

$$I(q) = \sum_{i=1}^{N+W} \sum_{j=1}^{N+W} f_i(q) f_j(q) \frac{\text{sen}(qd_{ij})}{qd_{ij}}, \quad (3.1)$$

onde,

$q = (4\pi \sin\theta)/\lambda$  é o vetor de espalhamento e sua unidade é o  $\text{\AA}^{-1}$ ;  $\lambda$  é o comprimento da onda de raios X;  $2\theta$  é o ângulo de espalhamento;

$N$  é o número de átomos do modelo;

$W$  é o número de moléculas de água que representam a camada de hidratação;

$d_{ij}$  é a distância entre os átomos  $i$  e  $j$ ;

$f(q)$  é o fator de espalhamento atômico, que depende das características de cada átomo;

Os elétrons são sensíveis aos campos, elétrico e magnético oscilando senoidalmente, o que justifica o seno na Equação 3.1.

Os perfis de SAXS obtidos, relacionam a intensidade do espalhamento  $I(q)$  com o ângulo  $q$ , e suas unidades de medidas são respectivamente unidades arbitrárias [u.a], e  $[\text{\AA}^{-1}]$ .

Para realizar a simulação computacional do perfil de SAXS é necessário apenas um arquivo PDB. Nesse trabalho de conclusão de curso, utilizou-se o algoritmo módulo SAXS-score do *pipeline* SAXTER (REIS, M. A.; APARICIO, R.; ZHANG, Y., 2011 [12]), para gerar os perfis simulados de SAXS. Como parâmetros de entrada para o algoritmo utilizou-se os 36 arquivos PDB das proteínas nos estados aberto e fechados, obtendo assim os perfis das 18 proteínas escolhidas.

### 3.2.1 Caracterização de estruturas utilizando SAXS

A partir da curva que caracteriza o perfil de SAXS é possível inferir sobre os modelos tridimensionais de uma proteína através da comparação de perfis.

Dois perfis de SAXS podem ser comparados numericamente através da equação:

$$\chi = \sqrt{\sum (I_{aberta}(q) - c I_{fechada}(q))^2}. \quad (3.2)$$

O resultado  $\chi$  retorna a raiz quadrada do somatório das diferenças entre as duas curvas elevado ao quadrado, onde  $I_{aberta}(q)$  representa o perfil simulado para a estrutura aberta, e  $I_{fechada}(q)$  representa o perfil simulado para a estrutura fechada. A constante  $c$  é um fator de escala para ajustar as duas curvas. O  $\chi$  resultante é a quantificação da diferença estrutural entre as estruturas aberta e fechada.

A Figura 3.1, exemplifica a comparação entre duas curvas simuladas. Cada curva representa uma estrutura de proteína, e o gráfico resultante é apresentado em escala logarítmica. Se tratando de uma análise comparativa de estruturas através dos perfis, as partes em que as curvas não coincidem implicam em diferenças estruturais da proteína entre seu estado aberto e fechado.

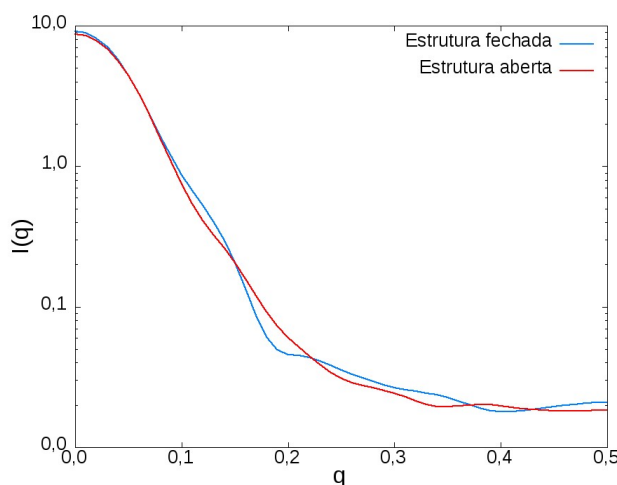


Figura 3.1: Comparação entre dois perfis simulados de SAXS. Cada curva representa uma estrutura, a diferença entre as curvas pode ser quantificada pela equação 3.2.

A comparação entre perfis simulados de SAXS e perfis teóricos obtidos, por exemplo, por Cristalografia, tem contribuído para validar modelos estruturais verificando diferenças entre proteínas e prevendo estruturas quaternárias.

### 3.3 *Template Modeling Score* (TM-score)

Após obter os resultados de SAXS, usou-se o algoritmo TM-score desenvolvido para quantificar a sobreposição de duas estruturas de proteínas, a partir dos arquivos

PDB.

Disponível para *download* no endereço <http://zhanglab.ccmb.med.umich.edu/TM-score/>, esse método utiliza informações de estruturas conhecidas. Esses arquivos PDBs foram obtidos através de técnicas experimentais, possuindo informações precisas sobre as coordenadas atômicas, conseqüentemente o TM-score se caracteriza como uma técnica de alta resolução.

Dada duas estruturas a quantificação da sobreposição é feita tomando pares de resíduos das duas estruturas e medindo suas distâncias. A função de pontuação proposta por ZHANG, SKOLNICK, 2004 [18] é dada pela fórmula:

$$TM - score = Max \left[ \frac{1}{L_N} \sum_{i=1}^{L_T} \frac{1}{1 + \left(\frac{d_i}{d_o}\right)^2} \right], \quad (3.3)$$

onde  $L_N$  é o tamanho da estrutura nativa, tomada como alvo;  $L_T$  é a quantidade de resíduos da estrutura experimental,  $d_i$  é a distância entre o  $i_{th}$  par dos resíduos alinhados e  $d_o$  é a escala para normalizar a diferença. “Max” quer dizer que a função retornará o valor máximo da sobreposição.

Segundo os autores, o raio de giro médio da estrutura de uma proteína globular depende de uma lei associada ao comprimento  $\langle R \rangle \approx L_N^{0,39}$ , o que permite estimar uma distância média dos pares de resíduos de estruturas escolhidas aleatoriamente. A partir da distância média é possível determinar o valor de  $d_o$ , o qual após minucioso estudo os autores chegaram ao valor de  $\approx 0,17$ .

O que difere o TM-score de outros métodos é o fator de normalização  $d_o$  de sua fórmula, o qual favorece a pontuação dos pares de átomos mais próximos em detrimento das combinações mais distantes, caracterizando essa técnica como uma análise quantitativa na qual domínios com alto grau de sobreposição são mais relevantes para a pontuação que a visão geral da sobreposição. Além disso, a técnica foi desenvolvida a fim de tornar irrelevante para o cálculo o tamanho das proteínas, já que essas podem ser comparadas de forma aleatória.

O procedimento padrão para a predição de estruturas de proteínas pode ser descrito por dois passos:

- (i) Encontrar uma estrutura já resolvida para ser utilizada como alvo;
- (ii) Construir uma estrutura baseada em modelos.

Feito isso, é possível comparar as duas estruturas utilizando o TM-score.

### 3.3.1 Análise qualitativa da pontuação do TM-score

O TM-score é um algoritmo que fornece respostas quantitativas sobre a sobreposição de duas proteínas.

Motivados em entender o quão significativo é esse resultado, XU e ZHANG [19] desenvolveram uma pesquisa na qual o significado do valor resultante pode ser analisado qualitativamente.

A pontuação do TM-score está compreendida entre [0,1], onde o valor 1 significa a perfeita sobreposição de duas estruturas.

Após realizarem testes com proteínas selecionadas aleatoriamente concluíram que:

- Para o TM-score acima de 0,5 as proteínas pertencem a mesma família;
- Quando o TM-score é inferior a 0,5 significa que as estruturas comparadas não pertencem a mesma família;
- Para o TM-score inferior a 0,17, as proteínas não possuem nenhuma semelhança estrutural.

### 3.4 Coeficiente de correlação de Pearson

Duas técnicas foram utilizadas para comparar a estrutura do estado aberto com a do estado fechado de 18 proteínas. A primeira técnica (SAXS) é de baixa resolução e a segunda TM-score é de alta resolução. Interessa-nos saber se existe e qual o grau de correlação das respostas obtidas por ambas técnicas.

Para obter essa resposta utilizou-se o coeficiente de correlação de Pearson dado pela equação:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.4)$$

Esse coeficiente mede o grau de correlação linear de duas variáveis, além de indicar a direção dessa correlação.

O resultado  $\rho$  pode variar entre -1 e 1, onde:

- -1 significa uma correlação negativa perfeita. Isso significa que enquanto uma variável aumenta o valor, a outra diminui;
- 0 significa que as duas variáveis não dependem linearmente uma da outra, mas isso não quer dizer que não exista correlação. Caso a resposta seja zero, é necessário utilizar outros métodos não lineares para apurar se existe ou não correlação;
- 1 significa uma correlação perfeita positiva entre as duas variáveis.

### 3.5 Modos normais de vibração e perfis de SAXS

Após obter os perfis de SAXS para as 18 proteínas identificou-se quais os maiores resultados de  $\chi$ . O objetivo agora é aplicar os 4 primeiros modos normais de vibração às estruturas abertas dessas proteínas que obtiveram os maiores  $\chi$ , afim de aferir se algum desses modos promove melhoramentos no perfil de SAXS, o que implicaria um possível enovelamento do estado aberto para o fechado.

O modo normal de vibração investigado nesse projeto basea-se na rede elástica proposta por TIRION. Nesse modelo a rede elástica é formada somente pelos  $C_\alpha$  dos resíduos de aminoácidos, os quais estão conectados um ao outro até uma certa distância que é determinada explicitamente pelo raio de corte ( $R_c$ ). Para o cálculo dos modos normais nesse trabalho usou-se um raio de corte de 10 Å na rede elástica. A Figura 3.2 (A) mostra a estrutura tridimensional da proteína *Enolase*, e a Figura (B) ilustra o formato da rede com raio de corte de 10 Å.

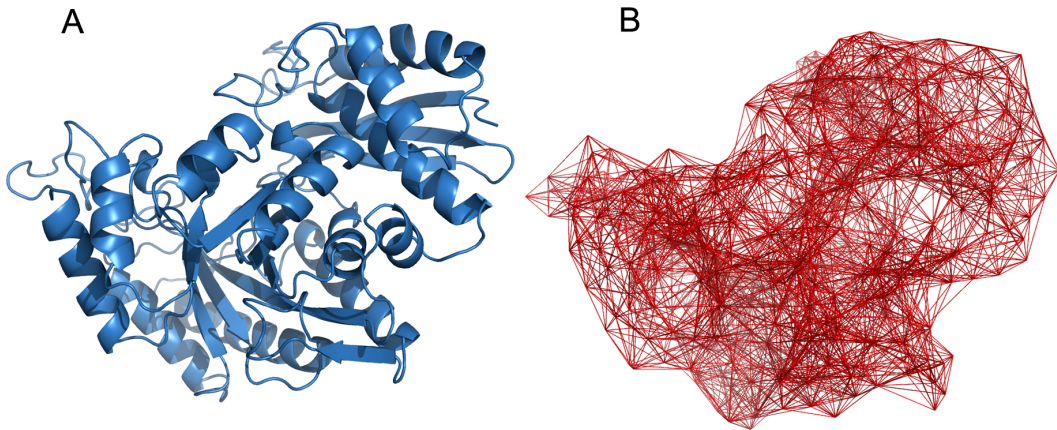


Figura 3.2: (A) Estrutura tridimensional da proteína *Enolase*. (B) Rede elástica da mesma proteína construída com raio de corte de 10 Å.

Neste método, as interações entre os átomos são consideradas harmônicas e as vibrações são descritas por coordenadas  $x, y, z$ .

De acordo com o modelo de rede elástica proposto por TIRION, a energia potencial é dada por:

$$E_{elastica} = \frac{1}{2} \sum_{d_{ij}^0 < R_c} k_{ij} (d_{ij} - d_{ij}^0)^2, \quad (3.5)$$

onde  $k$  é a constante de mola Hookiana entre os átomos  $i$  e  $j$ ,  $d_{ij}$  são as distâncias euclidianas entre dois  $C_\alpha$  e  $d_{ij}^0$  representa as distâncias na estrutura cristalográfica.

Segundo TIRION, 1996, [14], esse modelo foi desenvolvido e testado para baixas e altas frequências, e pode ser visto que a combinação linear dos modos vibracionais predizem os fatores de Debye-Waller inerentes à estrutura, que possuem dependência com

a temperatura.

Baseado no modelo de rede elástica escolhido, um algoritmo foi utilizado para gerar a matriz Hessiana. Considerando que todas as massas do sistema, ( $C_\alpha$ ), possuem o mesmo peso e que todos os átomos vizinhos, dentro de um raio de corte, estão conectados por uma mola sujeitos a mesma constante de força é calculado a derivada da força em relação às massas gerando assim a matriz Hessiana. Através da diagonalização da matriz realizada por outro algoritmo obtêm-se os autovetores, que são os modos normais de vibração. O resultado é armazenado em um arquivo no formato PDB.

A fim de realizar simulações conformacionais do estado aberto para o fechado calculou-se os modos normais de vibração para as estruturas abertas. A partir das coordenadas  $x, y$  e  $z$  dos  $C_\alpha$  obtidas pela diagonalização da matriz Hessiana, 50 modelos com variações das coordenadas foram geradas, 25 deslocando-se para o lado direito, e 25 para o lado esquerdo da estrutura geral tomada como sendo o modelo zero. A distância máxima de deslocamento<sup>1</sup> foi determinada para 20 Å da posição original. Para cada um desses 50 modelos foi calculado o  $\chi$  entre o modelo obtido pela simulação e a estrutura fechada da proteína. Assim foi possível verificar se o perfil de SAXS entre a estrutura aberta e fechada de uma determinada proteína foi melhorado aplicando determinado modo normal.

Modos normais de vibração é uma técnica de simulação computacional caracterizada como *coarse-grained*. Essa técnica indica possíveis direções de deslocamento de uma estrutura. A Figura 3.3 ilustra essa ideia. A estrutura tridimensional representada é da proteína *Lactoferrin*, e as setas indicando a direção que os domínios dessa proteína se deslocariam de acordo com o modo normal de vibração número 9.

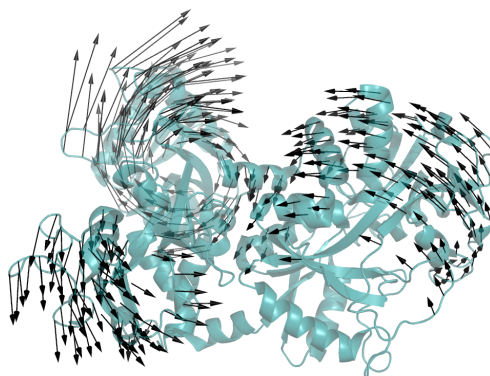


Figura 3.3: Estrutura tridimensional da proteína *Lactoferrin*. As setas indicam as direções de deslocamento dos domínios dessa proteína de no máximo de 20 Å, para o modo normal de vibração número 9.

---

<sup>1</sup>O deslocamento baseado nos modos normais é linear. O efeito visual da pseudorotação é causado pela variação de amplitude no movimento de cada átomo.

# Capítulo 4

## Resultados e discussões

Nesse capítulo são apresentados e discutidos os principais resultados obtidos após aplicar as técnicas descritas no Capítulo 3.

A primeira discussão se dará pela comparação dos resultados obtidos pelas duas métricas utilizadas para caracterização das estruturas proteicas.

### 4.1 TM-score *versus* SAXS-score

Após utilizar as duas técnicas para caracterizar a sobreposição das estruturas proteicas no estado aberto e fechado, obteve-se o gráfico da Figura 4.1, onde no eixo  $x$  estão representados os valores obtidos pelo SAXS-score e no eixo  $y$  os valores obtidos pelo TM-score. Cada ponto do gráfico representa a interseção dos resultados para cada estrutura de proteína, totalizando 18 pontos, os quais foram divididos em 3 grupos:

- O grupo **verdadeiro positivo** composto por 15 pontos representa as proteínas cujas estruturas apresentam grandes similaridades entre seus estados aberto e fechado. O resultado obtido pelo TM-score nesse grupo apresenta valores somente acima da reta que representa a constante 0,5 no eixo  $y$ . Como já mencionado na Seção 3.3.1, com esse resultado pode-se afirmar que as proteínas pertencem a mesma família com significativas semelhanças nos domínios estruturais. Ainda nesse grupo são observados os valores obtidos pelo SAXS-score todos abaixo do valor dois, com grande concentração dos pontos abaixo do valor unitário, ou seja, próximos de zero, o que também representa significativas semelhanças entre as estruturas comparadas, mas nesse caso considerando a estrutura global, pois diferentemente do TM-score o SAXS-score não possui um fator de normalização que privilegia a pontuação sobre as partes similares. A comparação é feita utilizando as curvas de perfis de SAXS que relacionam o ângulo com a intensidade do espalhamento.
- O caso **verdadeiro negativo** indica que essa proteína possui significativas diferenças estruturais entre seu estado aberto e fechado. O único caso encontrado nesse estudo

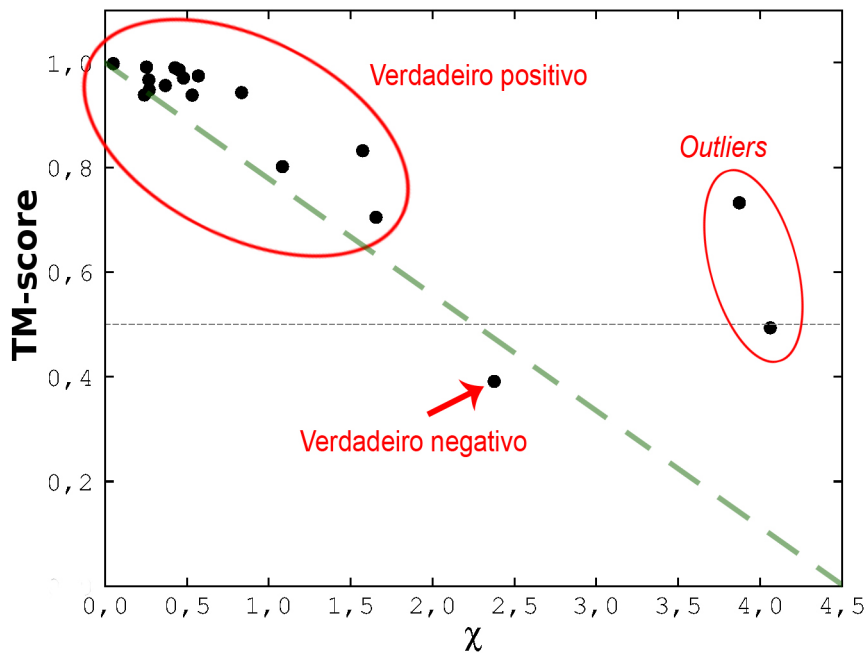


Figura 4.1: Resultados obtidos pelo TM-score (eixo  $y$ ) e SAXS-score (eixo  $x$ ), separados em 3 grupos. O grupo verdadeiro positivo representa as estruturas com grandes similaridades entre estado aberto e fechado segundo as duas métricas utilizadas. O caso verdadeiro negativo indica que essa proteína possui significativas diferenças estruturais entre seu estado aberto e fechado e esse resultado é indicado também pelas duas métricas. O grupo dos *outliers* mostra os casos nos quais as respostas das duas técnicas divergem entre si, e por isso estão fora da linha de tendência dos pontos (linha verde).

foi o da proteína *Adenylate kinase*, cujo  $\chi$  calculado foi de 2,38, e o TM-score foi de 0,39. Embora não exista um estudo quantitativo padronizado sobre o resultado do SAXS-score, apenas sabe-se que o valor zero implica a perfeita similaridade estrutural. A partir do resultado do TM-score é possível afirmar que estas estruturas possuem de fato significativas diferenças.

- O grupo dos *outliers* apresenta os casos emblemáticos de dissimilaridades estruturais. São casos nos quais as respostas obtidas pelas métricas utilizadas se divergem. Essas ocorrências se caracterizam pela essência das métricas, SAXS-score baseia-se na quantificação da estrutura global, enquanto TM-score baseia-se na pontuação sobre domínios. Nesse trabalho identificou-se dois casos de *outliers*, os quais influenciam diretamente no coeficiente de correlação dos pontos.

A partir dos resultados observar-se uma correlação entre os 18 pontos na ordem de  $\rho = -0.71$  calculado através do coeficiente de correlação de *Pearson* conforme a Equação 3.4. O maior valor na direção decrescente que o coeficiente de correlação de *Pearson* pode atingir é o valor unitário negativo (-1). Sendo assim concluí-se que o TM-score e o SAXS-score apresentam uma razoável correlação, embora *outliers* sempre existam, pois, TM-score foi projetado para pontuar sobre domínios, enquanto SAXS-score pontua sobre a globalidade das estruturas, resultado da baixa resolução da informação obtida pelo



experimento de SAXS.

## Exemplo do grupo verdadeiro positivo

A Figura 4.2 (A) apresenta os perfis simulados de SAXS para a proteína *Triose phosphate isomerase*, cujo  $\chi$  e o TM-score são respectivamente 0,05 e 0,99, o primeiro valor extremamente próximo a zero, e o segundo extremamente próximo ao valor unitário. Ainda na Figura 4.2 (B), pode ser conferida a sobreposição tridimensional da mesma proteína. Visualmente não é possível identificar diferenças tanto nas curvas quanto na estrutura tridimensional, pois as estruturas após serem sobrepostas de uma maneira ótima se equipararam perfeitamente.

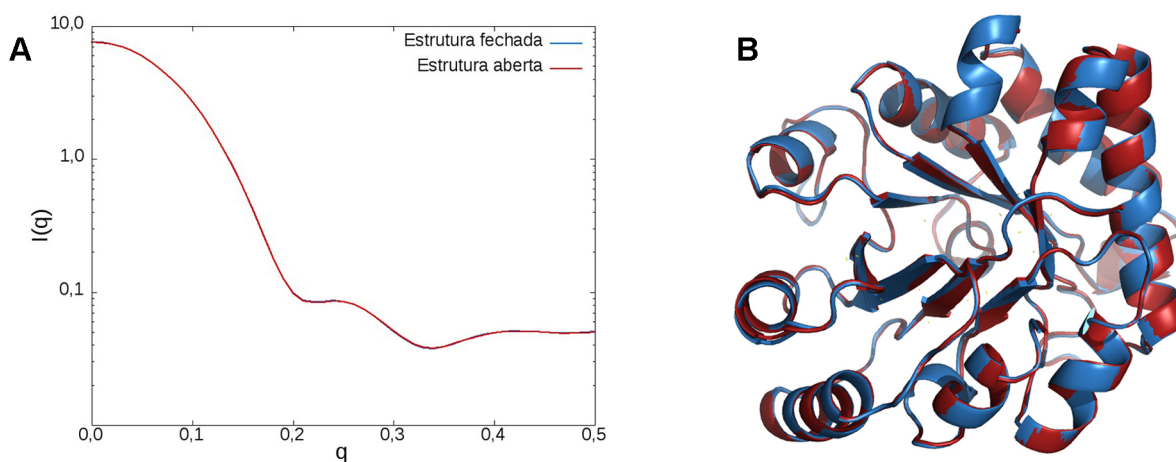


Figura 4.2: (A) Perfil simulado de SAXS calculado entre a estrutura aberta e fechada da proteína *Triose phosphate isomerase*. (B) Visualização tridimensional da mesma proteína calculada pelo TM-score.

## Exemplo do grupo verdadeiro negativo

Os perfis simulados de SAXS e a sobreposição tridimensional da única proteína pertencente ao grupo verdadeiro negativo pode ser conferida na Figura 4.3. Nos perfis de SAXS observa-se uma grande diferença entre as curvas do estado fechado e aberto e na representação tridimensional observa-se poucos domínios similares.

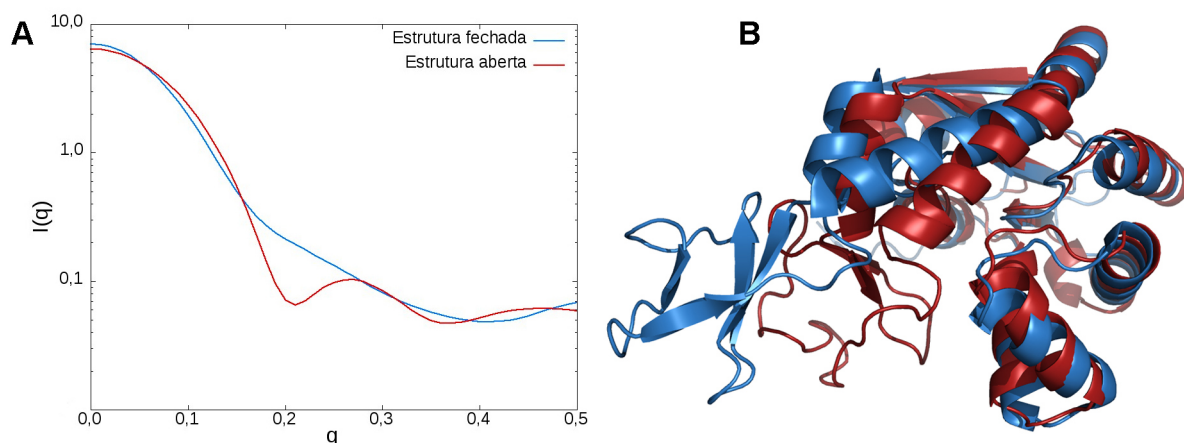


Figura 4.3: (A) Perfil simulado de SAXS calculado entre a estrutura aberta e fechada da proteína *Adenylate kinase*. (B) Visualização tridimensional da mesma proteína calculada pelo TM-score.

## Exemplos de *outliers*

Uma das proteínas desse grupo é a *Diphtheria toxin*, que possui TM-score de 0,73. Essa proteína tem como resultado do  $\chi$  o valor de 3,87, que está muito longe do valor de zero indicando assim que essa estrutura possui significativas diferenças globais estruturais. Porém, o resultado do TM-score está acima de  $y = 0,5$  e próximo ao valor unitário, o que segundo a análise qualitativa do TM-score nos permite concluir que essa proteína possui significativas similaridades de domínios.

Os perfis simulados de SAXS e a estrutura tridimensional podem ser observados na Figura 4.4.

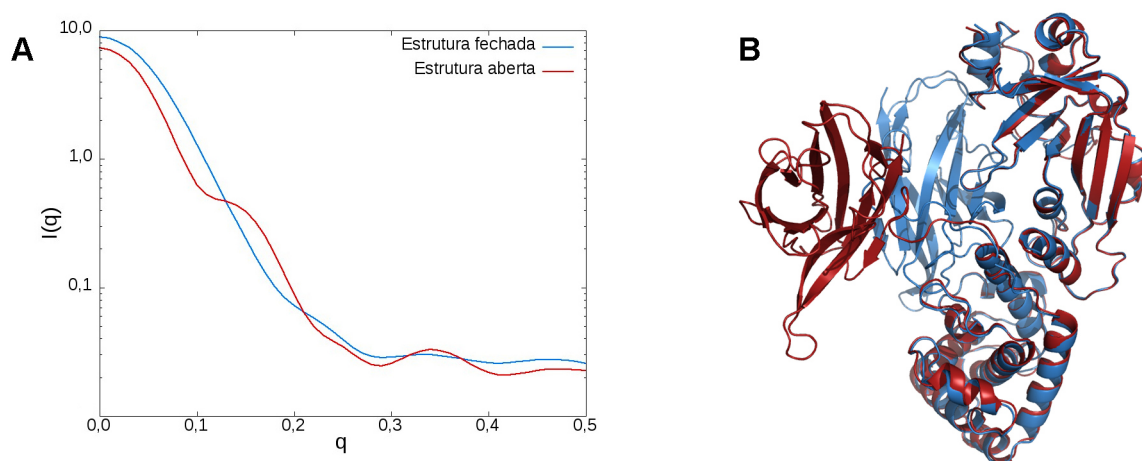


Figura 4.4: (A) Perfil simulado de SAXS calculado entre a estrutura aberta e fechada da proteína *Diphtheria toxin*. (B) Visualização tridimensional da mesma proteína calculada pelo TM-score.

Utilizando esse exemplo ressalta-se a discussão em torno das diferenças entre as métricas utilizadas. A proteína *Diphtheria toxin* cujo TM-score e SAXS-score apresentam

grande discrepância devido à essência das técnicas. Ao analisar a Figura 4.4, é possível verificar que grandes domínios estão equiparados com grandes similaridades, e que um domínio menor não foi possível sobrepor devido às dissimilaridades estruturais. O auto resultado do TM-score (0,73) se justifica pelo fator de normalização  $d_0$  na Equação 3.3, tornando essa métrica uma das mais importantes utilizadas para quantificar sobreposição de estruturas proteicas.

O segundo caso de *outlier* é a proteína *Calmodulin*, cujo resultado do SAXS-score da ordem de 4,06 é o maior dentre todas proteínas analisadas, e o TM-score é 0,49. Embora o TM-score esteja abaixo de  $y = 0,5$ , a discrepância entre os resultados das duas técnicas acarretam na classificação dessa proteína como também sendo um caso que está fora da linha de tendência do conjunto de pontos. Os perfis de SAXS e a estrutura tridimensional podem ser observados na Figura 4.5.

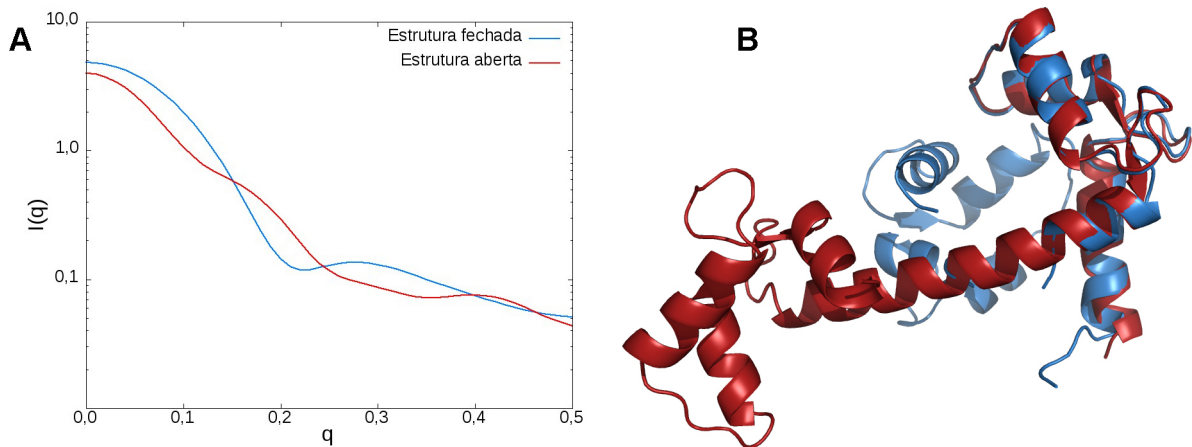


Figura 4.5: (A) Perfil simulado de SAXS calculado entre a estrutura aberta e fechada da proteína *Calmodulin*. (B) Visualização tridimensional da mesma proteína calculada pelo TM-score.

É possível aferir através dos exemplos as características inerentes à cada métrica. TM-score, uma técnica de alta resolução, que pontua sobre domínios, e SAXS-score uma técnica experimental de baixa resolução, que pontua através da totalidade das estruturas, apresentaram um auto grau de correlação embora ocorram *outliers*, que são consequências inerentes às particularidades das técnicas.

## 4.2 Ajuste dos perfis de SAXS com auxílio dos modos normais

Depois de conhecer os resultados obtidos pelas duas métricas que quantificam a qualidade da sobreposição de estruturas, agora é de interesse saber se existe um modo normal de vibração que melhore os perfis de SAXS calculados entre estrutura fechada e aberta, ou seja, se existe uma distorção que seja capaz de aproximar as duas estruturas.

Nesse trabalho, como estamos comparando a estrutura aberta com a fechada de uma determinada proteína, encontrar um modo normal de vibração que melhore o perfil de SAXS, implica encontrar uma possível conformação estrutural do estado aberto para o estado fechado.

A análise dos modos normais de vibração foi feita para as proteínas que apresentaram valores de  $\chi$  maiores que 0,5. Perfis de SAXS calculados para simulações de estruturas que apresentam  $\chi$  inferior a 0,5 se tornam inviáveis devido a grande similaridade que as estruturas já possuem. No Apêndice B Seções B.4 e B.5, são mostrados dois casos em que o  $\chi$  é da ordem de 0,83 e 0,57. É possível verificar através das Figuras B.4 e B.5 que a maioria dos valores de  $\chi$  calculados entre as simulações e a estrutura fechada são maiores que o  $\chi$  entre estrutura aberta e fechada original (Representado pelo modelo zero). Diante disso, torna-se interessante a discussão utilizando proteínas que apresentam significativas diferenças estruturais.

Nessa seção serão discutidos os dois maiores valores de  $\chi$  ocorrentes no grupo dos *outliers*. Os demais casos podem ser conferidos no apêndice B.

A Figura 4.6 apresenta o gráfico dos 4 primeiros modos normais de vibração da proteína *Diphtheria toxin* que tem como resultado do  $\chi$  o valor de 3,87, e o do TM-score 0,73. Cada modo normal de vibração foi calculado utilizando 50 variações do modelo original, sendo 25 deslocamentos em um determinado sentido a partir da estrutura original, e 25 para outro sentido, o que justifica o domínio do eixo  $x$  estar definido para o intervalo  $[-25,25]$ , onde a estrutura original é representado pelo modelo zero. Para cada um desses modelos foi calculado o  $\chi$  entre a estrutura fechada e o modelo simulado, o resultado encontra-se no eixo  $y$ .

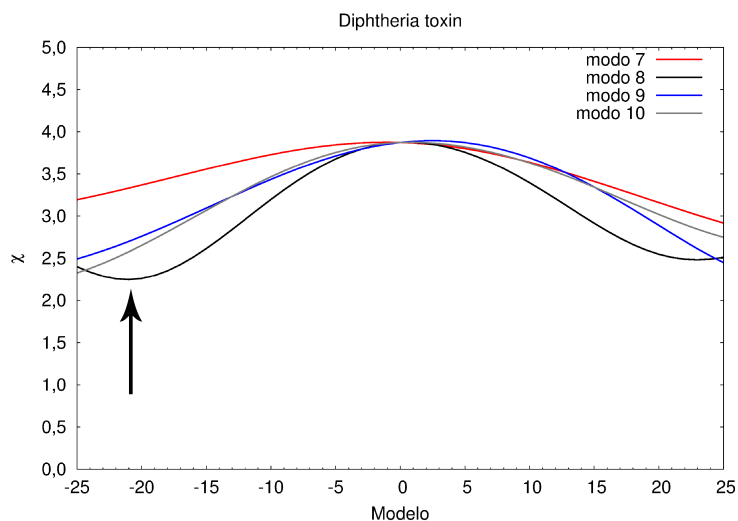


Figura 4.6: Perfil de SAXS para os 4 primeiros modos normais de vibração da proteína *Diphtheria toxin*. O eixo  $x$  representa cada estrutura do modo normal de vibração, e o eixo  $y$  apresenta o valor obtido pela simulação do SAXS-score entre a estrutura fechada o modelo.

O modo normal 8, foi o que apresentou o maior mínimo global para o valor de  $\chi$ , conseqüentemente ele é o modo cujo perfil de SAXS melhor se ajusta a estrutura da proteína no estado fechado. A Figura 4.7 (A) mostra com detalhes a curva do modo 8, e a Figura (B) mostra o sentido do deslocamento causado por esse modo.

O menor valor obtido do  $\chi$  é da ordem de 2,25 e refere-se a estrutura do modelo -21. Esse valor é referente a vigésima primeira posição do deslocamento do modo 8 em relação à estrutura original.

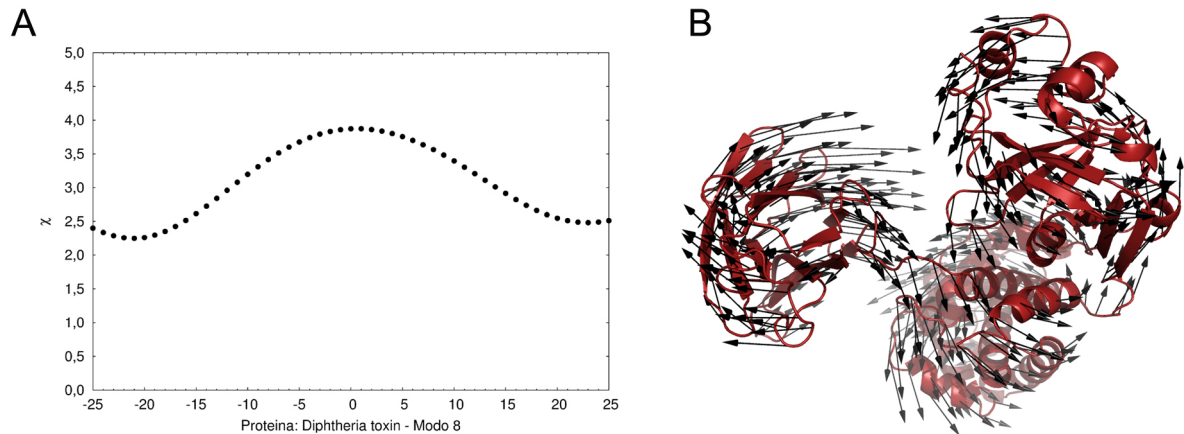


Figura 4.7: Perfil de SAXS para o modo 8 da proteína *Diphtheria toxin*. O eixo  $x$  representa cada estrutura do modo normal de vibração, e o eixo  $y$  o valor obtido pela simulação do SAXS-score entre a estrutura fechada o modelo.

O resultado dos perfis de SAXS e a sobreposição tridimensional entre o modelo -21 e a estrutura fechada podem ser conferido na Figura 4.8. A curva em verde na Figura 4.8 (A) representa a simulação do perfil de SAXS para o modo 8. Observa-se que essa curva está mais próxima da estrutura fechada (curva azul). E na Figura 4.8 (B), representado por esferas verdes a estrutura do modelo -21, o qual apresentou o menor  $\chi$  na simulação com a estrutura fechada.

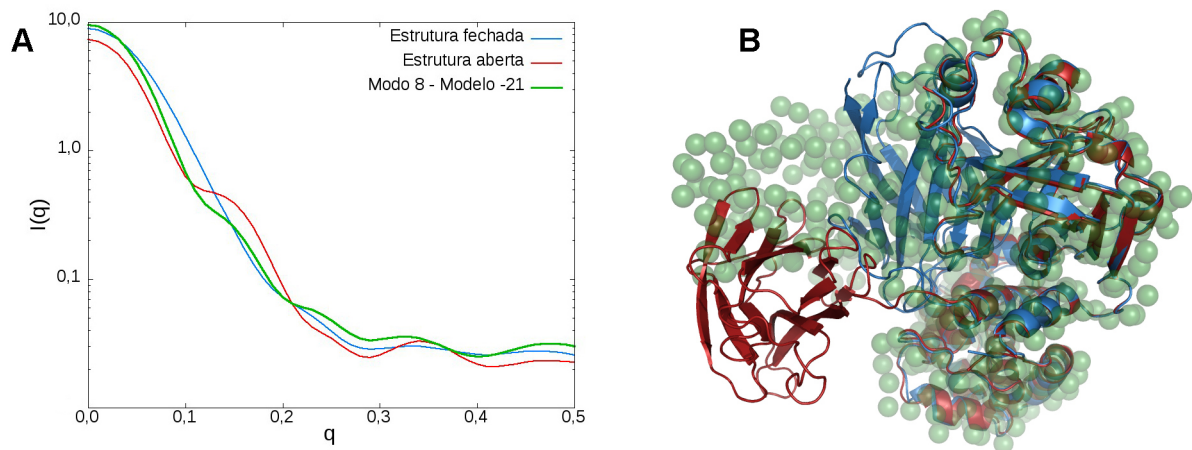


Figura 4.8: (A) Perfil de SAXS para o modelo -21 do modo 8 da proteína *Diphtheria toxin*. (B) Comparação entre as sobreposições do modelo -21 (esferas verdes) com a estrutura aberta (vermelho) e a estrutura fechada (azul).

A Figura 4.9 apresenta o gráfico dos 4 primeiros modos normais de vibração da proteína *Calmodulin* que tem como resultado do  $\chi$  o valor de 3,87, e o do TM-score 0,73.

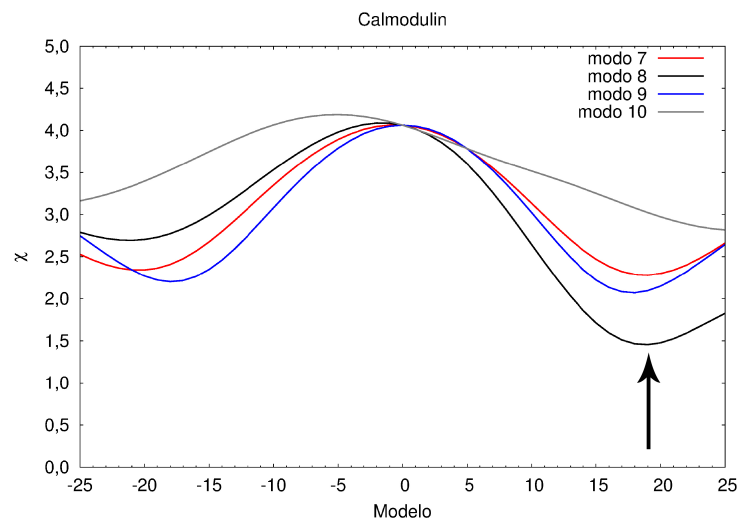


Figura 4.9: Perfil de SAXS para os 4 primeiros modos normais de vibração da proteína *Calmodulin*. O eixo  $x$  representa cada estrutura do modo normal de vibração, e o eixo  $y$  apresenta o valor obtido pela simulação do SAXS-score entre a estrutura aberta o modelo.

Da mesma forma que o caso anterior, cada modo normal de vibração foi calculado utilizando 50 variações do modelo original. Para cada um desses 50 modelos foi calculado o  $\chi$  entre a estrutura fechada e o modelo simulado, o resultado é representado no eixo  $y$ . O modo normal 8, foi o que apresentou mínimo global para o valor de  $\chi$ , conseqüentemente ele é o modo cujo perfil de SAXS melhor se ajusta a estrutura da proteína no estado fechado. A Figura 4.10 (A) mostra com detalhes a curva do modo 8, e Figura 4.10 (B) mostra um dos sentidos do deslocamento desse modo.



O menor valor obtido do  $\chi$  é da ordem de 1,45 e refere-se a estrutura do modelo 19. Esse valor é referente a décima nona posição do deslocamento do modo 8, na direção direita da estrutura original.

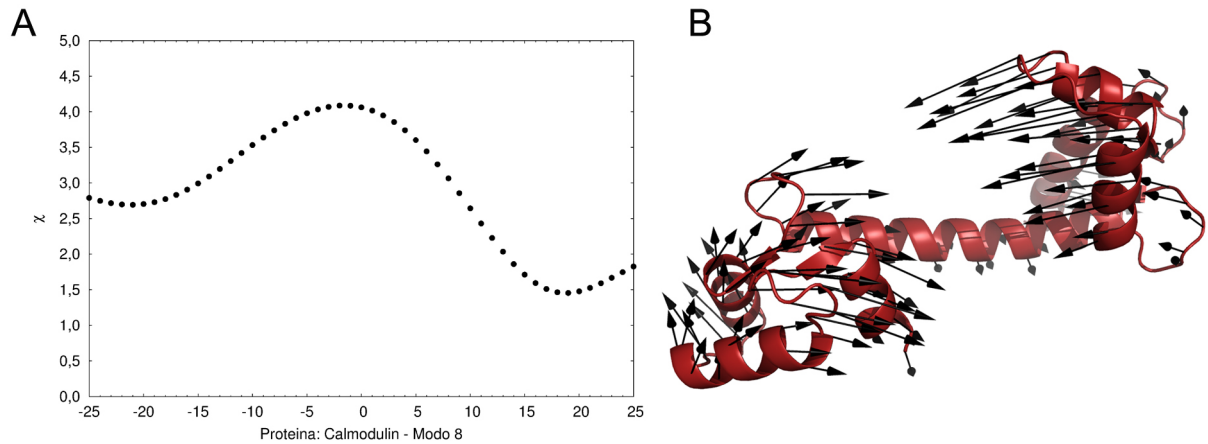


Figura 4.10: Perfil de SAXS para o modo 8 da proteína *Calmodulin*. O eixo  $x$  representa cada estrutura do modo normal de vibração, e o eixo  $y$  apresenta o valor obtido pela simulação do SAXS-score entre a estrutura aberta o modelo.

O resultado do perfil de SAXS e a sobreposição tridimensional entre o modelo 19 e a estrutura fechada podem ser conferido na Figura 4.11. A linha verde na Figura 4.11 (A) representa a simulação dos perfis de SAXS para o modo 8. É possível observar que essa curva está mais próxima da estrutura fechada (curva azul). E na Figura 4.11 (B), observa-se em esferas verdes a estrutura do modelo 19, o qual apresentou o menor  $\chi$  na simulação com a estrutura fechada.

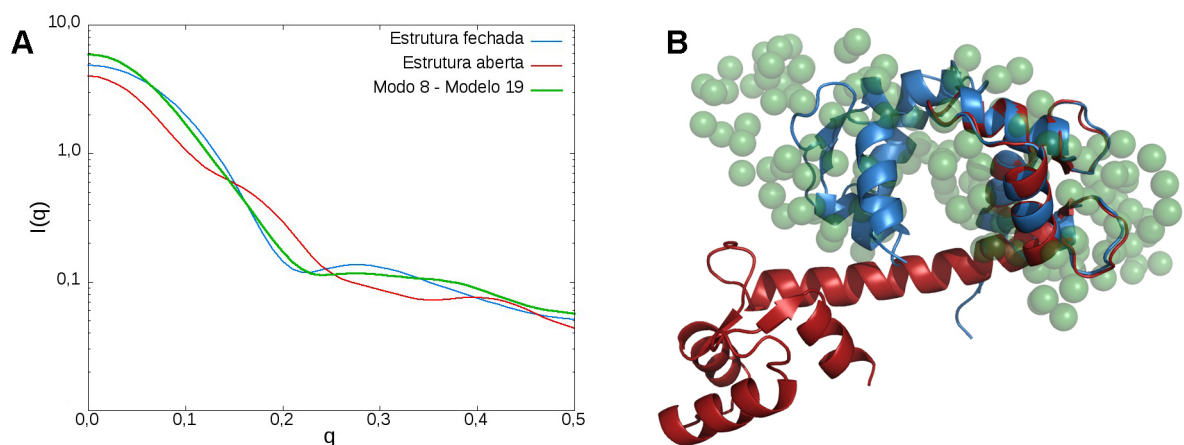


Figura 4.11: (A) Perfil de SAXS para o modelo 19 do modo 8 da proteína *Calmodulin*. (B) Comparação entre as sobreposições do modelo 45(esferas verdes) com a estrutura aberta (vermelho) e a estrutura fechada (azul).

# Capítulo 5

## Conclusões

Duas métricas distintas foram utilizadas para caracterizar estruturas tridimensionais de um grupo específico de proteínas que apresentam características de estado aberto e fechado. Confrontando os resultados obtidos por ambas técnicas, a primeira experimental, e a segunda teórica, foi possível verificar que, mesmo utilizando informação de baixa resolução, a técnica de SAXS obteve a maioria dos resultados compatíveis com o TM-score. Isso ocorreu nos grupos verdadeiro positivo e verdadeiro negativo, totalizando 16 pontos. A fim de verificar a existência de correlação entre as métricas utilizou-se o coeficiente de correlação de Pearson que obteve resultado na ordem de  $\rho = -0,71$ . Esse resultado implica numa forte correlação entre as técnicas, pois o valor máximo, em módulo, que se poderia obter desse coeficiente é da ordem de 1.

Simulações de modos normais de vibração foram utilizados a fim de obter melhoramentos nos perfis de SAXS. Foram realizadas simulações para os modos 7, 8, 9 e 10, pois como descrito na Seção 2.2.8, grandes amplitudes vibracionais estão relacionadas às mais baixas frequências e portanto, aos primeiros modos normais a partir do modo 7. Calculou-se o  $\chi$  entre cada modelo de deslocamento e a estrutura fechada. O modo 8 foi o que apresentou o menor  $\chi$  para as duas proteínas que apresentavam maiores diferenças estruturais (*Diphtheria toxin* e *Calmodulin*).

O perfil de SAXS entre estrutura aberta e fechada para a proteína *Diphtheria toxin* é da ordem de 3,87. Após aplicar os 4 primeiros modos normais, verificou-se que o modo 8 fez com que o  $\chi$  diminuísse para 2,25, obtendo um melhoramento do perfil de SAXS da ordem de 42%. O  $\chi$  da proteína *Calmodulin* também calculado entre estado aberto e fechado é 4,06. Através do modo normal de vibração número 8, esse  $\chi$  abaixou para 1,45. O melhoramento do perfil foi da ordem de 64%.

A primeira técnica, espalhamento de raios X a baixos ângulos (SAXS), é uma técnica experimental de baixa resolução. Embora utilize os mesmos princípios físicos da Cristalografia, esse experimento não fornece informações detalhadas como, por exemplo, as coordenadas atômicas da estrutura da proteína. O resultado obtido é um perfil que relaciona o ângulo de espalhamento com a intensidade coletada no detector. Nesse método



a caracterização de estruturas pode ser feita tanto comparando perfis teórico – obtidos através de técnicas mais precisas – com perfis experimentais, como utilizando perfis simulados de SAXS, como os utilizado nesse trabalho descrito na Seção 3.2. A comparação entre dois perfis de SAXS permite inferir resultados sobre a semelhanças entre estruturas proteicas.

O segundo método utilizado - TM-score - caracteriza estruturas proteicas comparando modelos já existentes disponíveis em bancos de dados. Duas estruturas com coordenadas atômicas definidas são sobrepostas de uma maneira ótima, e então é feito a quantificação da semelhança entre as duas estruturas. A utilização das coordenadas atômicas como parâmetros garante a essa técnica resultados de alta resolução.

Além de caracterizar estruturas tridimensionais, utilizar modos normais de vibração em estruturas proteicas significa simular possíveis enovelamentos dessas proteínas. Nesse trabalho, a simulação e os cálculos realizados, implicam em possíveis conformações estruturais do estado aberto para o fechado das proteínas analisadas. Sendo o modo 8 o que melhor prediz a conformação das duas estruturas com maiores valores de  $\chi$ , esses resultados acordam com a literatura descrita na Seção 2.2.8.

A fim de se obter resultados mais acurados utilizando SAXS novas pesquisas vêm sendo desenvolvidas. O algoritmo SAXSTER, descrito no artigo *Improving Protein Template Recognition by Using Small-Angle X-Ray Scattering Profiles* escrito por REIS, et al., 2011 [12], combina os resultados de SAXS com outra técnica de caracterização de estruturas proteicas. Esse algoritmo pode ser utilizado no endereço: <http://zhanglab.ccmb.med.umich.edu/SAXSTER/>.

Por fim, através desse trabalho conclui-se que é possível realizar interessantes estudos sob o ponto de vista científico à baixos investimentos. Esses estudos se realizados no meio acadêmico, durante a formação dos alunos, pode incentivar o contato com a pesquisa e com o conhecimento científico, contruindo para a formação e posterior qualificação de pessoas para o meio acadêmico como futuros docentes bem preparados, ou para o meio científico desenvolvendo pesquisas que visam contribuir de alguma forma com a sociedade.

# Apêndice A

## Proteínas, TM-score e SAXS-score

A tabela A.1 mostra os valores obtidos pelo TM-score e SAXS-score para as 18 proteínas estudadas. Também encontra-se os respectivos códigos de identificação no banco de dados PDB das proteínas no estado aberto e fechado, e uma breve descrição da função biológica de cada uma dessas macromoléculas.

As 18 proteínas foram divididas em 3 grupos conforme descrito no Capítulo 4.

Fazem parte do grupo verdadeiro positivo as proteínas: Triose phosphate isomerase; Annexin V; Thymidylate synthase; Triglyceride lipase; Seryl-tRNA synthetase; Alcohol dehydrogenase; Aspartate aminotransferase; HIV-1 protease; Dihydrofolate reductase; Che Y protein; Citrate synthase; Lactoferrin; Maltodextrin binding; LAO binding.

Somente uma proteína compôs o grupo verdadeiro negativo: Adenylate kinase.

E as duas proteínas que apresentaram os maiores valores obtidos pelo SAXS-score fazem parte do grupo dos *outliers*: Calmodulin e Diphtheria toxin.

<b>Proteína</b>	<b>PDB ID Aberto - Fechado</b>	<b>TM score</b>	<b>SAXS score</b>	<b>Função biológica</b>
Adenylate kinase	1aky - 2ak3	0,39	2,38	É uma enzima responsável pela regulação do glicogênico, açúcar e o metabolismo lipídico. As proteínas quinases são a chave central da comunicação no controle intracelular, regulação e transdução de sinais.
Alcohol dehydrogenase	8adh - 6adh	0,97	0,27	Os membros desta família de enzimas são responsáveis por quebrar os grupos de álcool e tóxicos durante o metabolismo humano e em alguns animais.
Annexin V	1avr - 1avh	0,99	0,25	É uma proteína com função anticoagulante. Durante a apoptose celular é translocada para a camada externa da membrana celular.
Aspartate aminotransferase	9aat - 1ama	0,96	0,37	Encontrada no miocárdio, fígado, rins, cérebro, pulmões, pâncreas, baço, células vermelhas e brancas. Sua elevação indica comprometimento celular mais profundo podendo revelar hepatite.
Calmodulin	1cll - 1ctr	0,49	4,06	Proteína ativadora termo-estável, de baixo peso molecular encontrada principalmente no encéfalo e coração. É responsável pelo transporte de cálcio.
Che Y protein	3chy - 1chn	0,94	0,24	Responsável pela direção do movimento de bactérias que se dará de acordo com a quantidade de nutrientes no ambiente.
Citrate synthase	5csc - 6csc	0,94	0,53	É uma enzima transferase que controla o primeiro passo do ciclo do ácido cítrico. O ciclo tem papel central nos mecanismos metabólicos de obtenção de energia.
Dihydrofolate reductase	4dfr - 5dfr	0,94	0,83	Catalisa o processo de redução do ácido fólico a ácido tetraidrofólico, um transporte para o grupo metilo necessária para a síntese de purinas, ácido timidílico, e certos aminoácidos.
Diphtheria toxin	1ddt - 1mdt	0,73	3,87	Essa toxina entra no citoplasma da célula e inibe a síntese de proteínas, causando inflamação e lesão em partes das vias respiratórias.
Enolase	3enl - 7enl	0,99	0,43	A enolase é uma enzima que atua na cascata glicolítica, catalizando a nona reação da via. O 2-fosfoglicerato, formado na reação anterior, é agora desidratado e forma 1 molécula de água e uma de fosfoenolpiruvato, um composto altamente energético.

*Continua na próxima página*

Proteína	PDB ID Aberto - Fechado	TM score	SAXS score	Função biológica
HIV-1 protease	1hhp – 1ajx	0,95	0,27	Enzima fundamental para que o vírus da AIDS se replique. As proteases são enzimas que clivam as cadeias peptídicas em fragmentos de aminoácidos menores. Os peptídios são responsáveis pela patogenicidade do vírus.
Lactoferrin	1lfh – 1lfg	0,83	1,57	Proteína encontrada no leite que cumpre uma valiosa função de antibiótico natural. A lactoferrina tem a capacidade de reter átomos de ferro impedindo que eles sejam roubados por bactérias.
LAO binding protein	2lao – 1lst	0,70	1,65	Grupo de ligantes responsável pelo transporte de solutos através da membrana citoplasmática.
Maltodextrin binding protein	1omp – 1anf	0,80	1,08	É responsável pela absorção eficiente e catabolismo de maltodextrinas. É um sistema de regulação e transporte complexo que envolve muitas proteínas e complexos de proteínas.
Seryl-tRNA synthetase	1ses – 1set	0,97	0,48	Família de enzimas que asseguram a correta ligação entre um aminoácido e seu tRNA, compondo fatores essenciais para o processo de síntese proteica.
Thymidylate synthase	3tms – 2tsc	0,99	0,45	Atua na conversão do grupo metileno a grupo metila obtendo como resultado substrato para a síntese de DNA. Como as células cancerígenas sofrem rápida divisão celular, essa enzima é altamente ativa.
Triglyceride lipase	3tgl – 4tgl	0,97	0,57	Essa proteína é responsável pelo controle dos triglicerídeos que são armazenados em adipócitos para serem usados depois como fonte de energia.
Triose phosphate isomerase	3tim – 6tim	0,99	0,05	Está relacionada com importantes vias metabólicas, como glicólise, gliconeogênese e síntese de ácidos graxos, responsáveis principalmente pela produção de energia.

Tabela A.1: Tabela contendo as 18 proteínas, seus respectivos códigos PDB (o primeiro refere-se a estrutura aberta e o segundo a fechada), os valores obtidos pelos algoritmos TM-score e SAXS-score, e uma breve descrição da função biológica de cada proteína.

# Apêndice B

## Modos normais de vibração e perfis de SAXS

Esse Apêndice foi criado a fim de melhorar a organização e compreensão do trabalho. Foram gerados 18 perfis simulados de SAXS a fim de comparar as estruturas das proteínas em seu estado aberto e fechado. Das 18 proteínas estudadas, 10 apresentaram perfis de SAXS com valores superiores a 0,5. Os dois maiores valores foram discutidos no Capítulo 4. Os demais podem ser conferidos nas seções abaixo:

### **B.1** *LAO binding protein*

De forma análoga ao caso anterior o perfil simulado de SAXS calculado entre a estrutura aberta e fechada foi da ordem de 1,65. Através da Figura B.1 (A), é possível observar diferenças significativas entre as curvas que representam a comparação entre as estruturas. Devido a essa diferença significativa foi aplicado os quatro primeiros modos normais de vibração a fim de aferir se haveria melhora nos perfis de SAXS. A Figura B.1 (B), mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 9 é o que apresenta o menor mínimo global entre as curvas na ordem de 0,68 para o modelo 12, significando que a simulação referente ao modo 9 pode estar associada ao enovelamento dessa proteína.

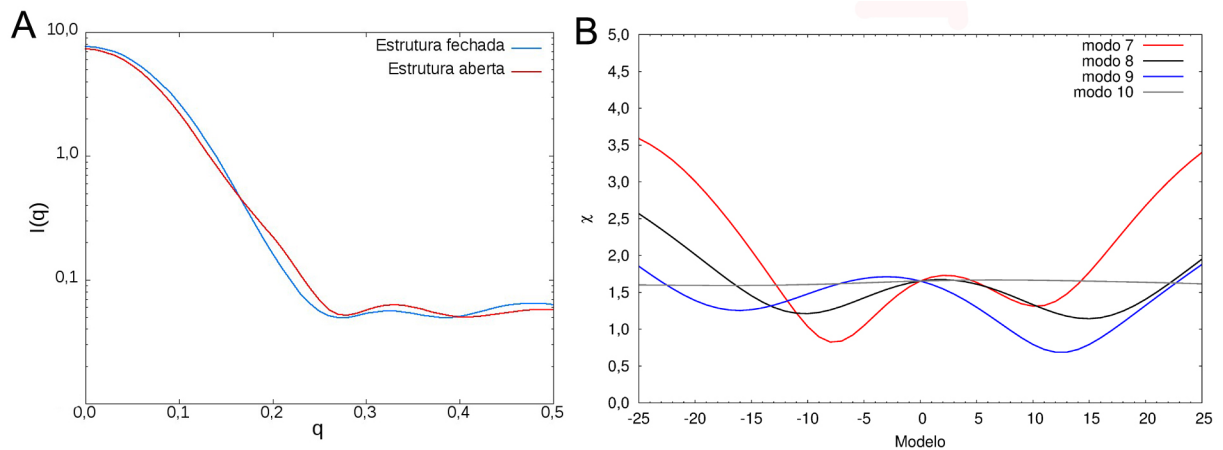


Figura B.1: (A) Perfil de SAXS para a proteína *LAO binding protein* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.

## B.2 *Lactoferrin*

A Figura B.2 (A) mostra perfil simulado de SAXS calculado entre a estrutura aberta e fechada da ordem de 1,57. A Figura B.2 (B), mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 9 é o que apresenta o menor mínimo global entre as curvas na ordem de 0,68 para o modelo -14, significando que a simulação referente ao modo 9 pode estar associada ao enovelamento dessa proteína.

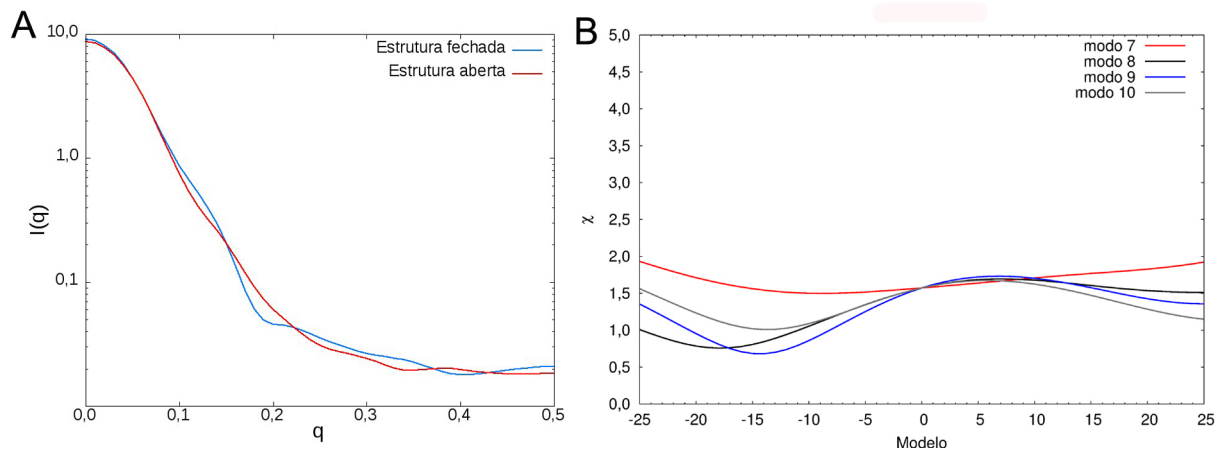


Figura B.2: (A) Perfil de SAXS para a proteína *Lactoferrin* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.

### B.3 *Maltodextrin binding protein*

A Figura B.3 (A) mostra perfil simulado de SAXS calculado entre a estrutura aberta e fechada da ordem de 1,08. A Figura B.3 (B), mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 7 é o que apresenta o menor mínimo global entre as curvas na ordem de 0,37 para o modelo -8, significando que a simulação referente ao modo 7 pode estar associada ao enovelamento dessa proteína.

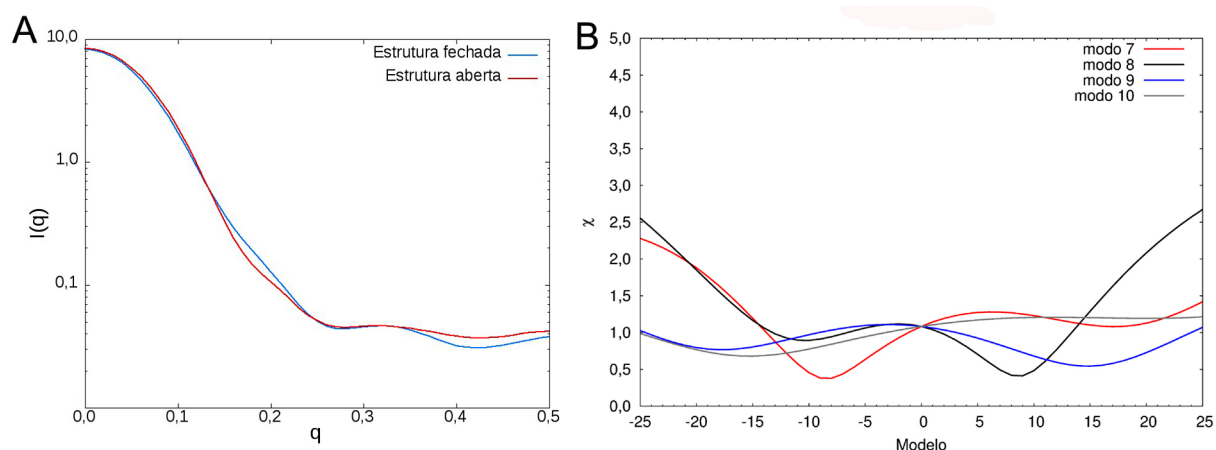


Figura B.3: (A) Perfil de SAXS para a proteína *Maltodextrin binding protein* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.

### B.4 *Dihydrofolate reductase*

A Figura B.4 (A) mostra perfil simulado de SAXS calculado entre a estrutura aberta e fechada da ordem de 0,83. Observa-se que esse  $\chi$  abaixo de 1,0 as curvas possuem poucas diferenças. A Figura B.4 (B), mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 7 é o que apresenta o menor mínimo global entre as curvas. Entretanto o valor obtido pelo modelo 5 é da ordem de 0,72, isso significa uma pequena modificação estrutural. Quanto menor o valor resultante menos pode-se inferir sobre os modos normais.

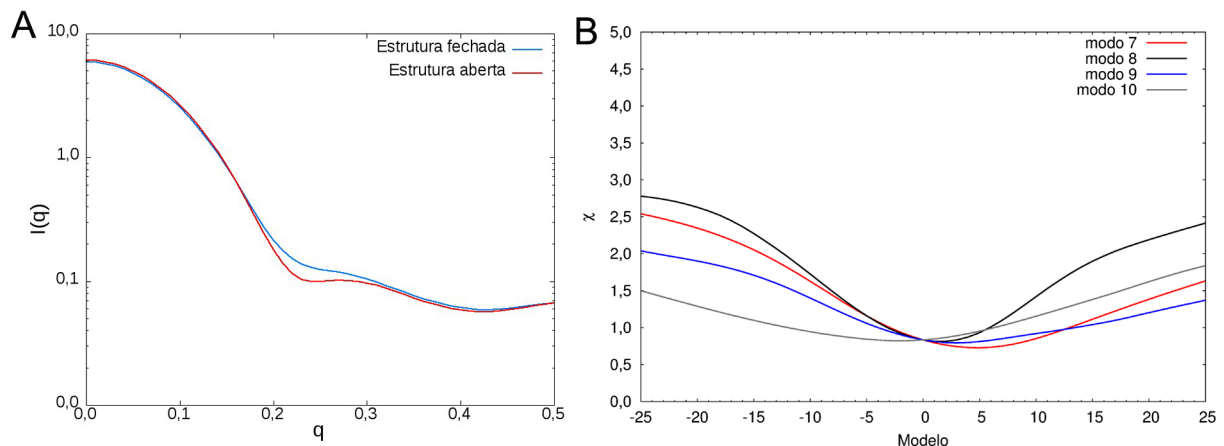


Figura B.4: (A) Perfil de SAXS para a proteína *Dihydrofolate reductase* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.

## B.5 *Triglyceride lipase*

A Figura B.5 (A) mostra perfil simulado de SAXS calculado entre a estrutura aberta e fechada da ordem de 0,57. A Figura B.5 B, mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 10 é o que apresenta o menor mínimo global entre as curvas, e o modelo 5 foi o que obteve o menor  $\chi$  na ordem de 0,45. Assim como no caso estudado acima, quanto menor o valor do  $\chi$  menos se pode inferir sobre possíveis conformações proteicas utilizando modos normais de vibração, pois a mudança estrutural é mínima.

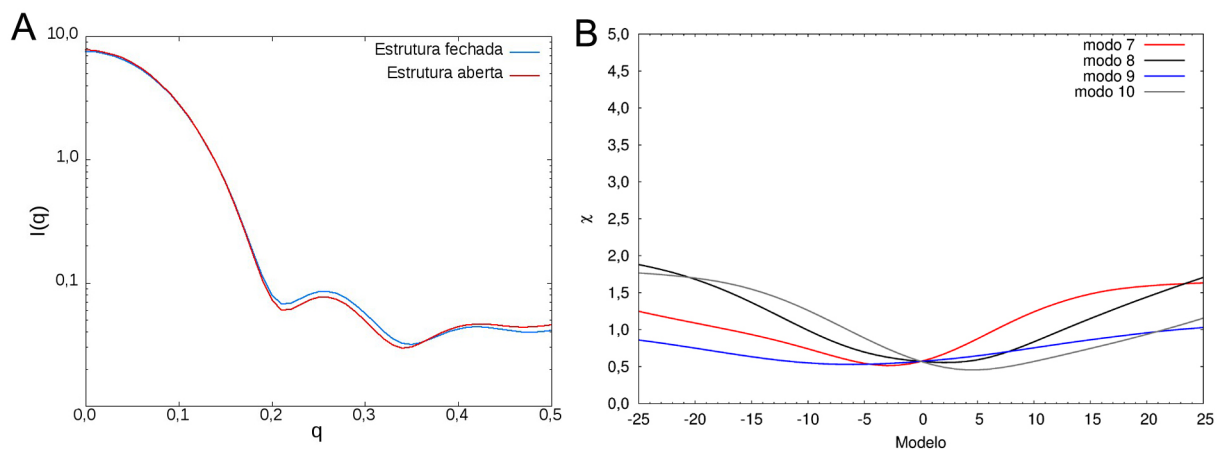


Figura B.5: (A) Perfil de SAXS para a proteína *Triglyceride lipase* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.



## B.6 Citrate synthase

Para essa proteína o perfil simulado de SAXS calculado entre a estrutura aberta e fechada foi da ordem de 0,53. Através da Figura B.6 (A), é possível observar que não há diferenças significativas entre as curvas que representam a comparação entre as estruturas. A Figura B.6 (B), mostra os perfis de SAXS dos modos normais de vibração. O perfil do modo 8 é o que apresenta o menor mínimo global entre as curvas na ordem de 0,35 para o modelo 11, significando que a simulação referente ao modo 8 seria o possível envelhecimento dessa proteína.

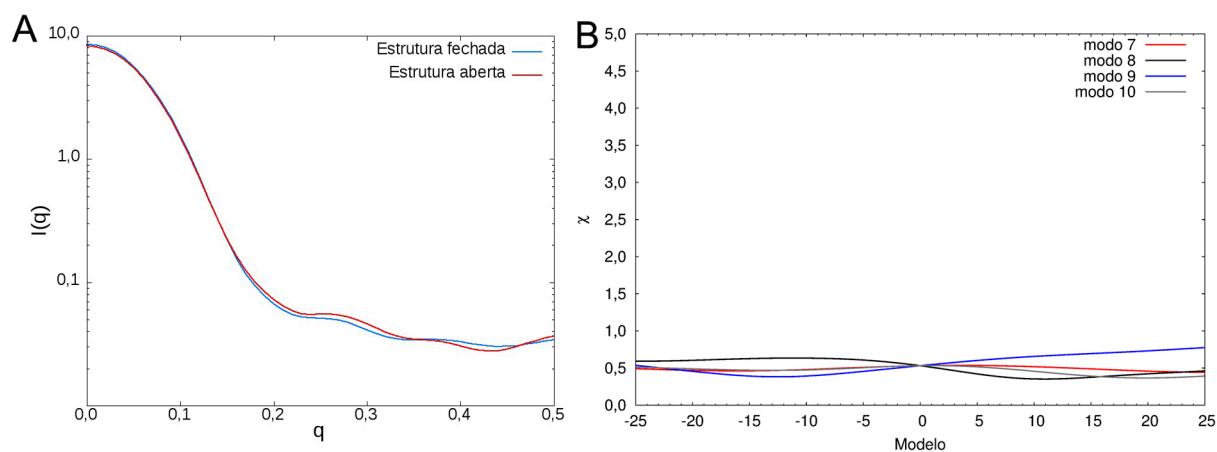


Figura B.6: (A) Perfil de SAXS para a proteína *Adenylate kinase* entre sua estrutura no estado aberto e fechado. (B) Comparação entre os 4 primeiros modos normais de vibração para a mesma proteína, utilizando perfis de SAXS.

# Bibliografia

- [1] STEWART, J. *Cálculo 1*. 7. ed. Brasil: Cengage Learning, 2013.
- [2] LEHNINGER, A.; NELSON, D. L.; COX, M. M. *Princípios de bioquímica de Lehninger*. 5. ed. Porto Alegre: Artmed, 2011.
- [3] VOET, D.; VOET, J. G.; PRATT, C. W. *Fundamentos de bioquímica: a vida em nível molecular*. 2. ed. Porto Alegre: Artmed, 2008.
- [4] DIMA, R. I.; THIRUMALAI, D. Proteins associated with diseases show enhanced sequence correlation between charged residues. *Bioinformatics*, v. 20, n. 15, p. 2345 – 2354, April - 2004.
- [5] BENTON, D. Bioinformatics - principles and potential of a new multidisciplinary tool. *Trends in Biotechnology*, v. 14, n. 8, p. 261 – 272, August - 1996.
- [6] OUZOUNIS, C. Bioinformatics and the theoretical foundations of molecular biology. *Bioinformatics*, v. 18, n. 3, p. 377 – 378, 2002.
- [7] FILHO, O. A. S.; ALENCASTRO, R. B. de. Modelagem de proteínas por homologia. *Química Nova*, v. 26, n. 2, p. 253 – 259, Setembro - 2002.
- [8] BERMAN, H. A. et al. The Protein Data Bank. *J. Mol. Biol.*, v. 28, n. 11, p. 235 – 242, December - 2000.
- [9] SVERGUN, D.; KOCH, M. H. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.*, v. 66, p. 1735 – 1782, 2003.
- [10] BLUNDELL, T. L.; JOHNSON, L. Protein Crystallography. *Academic Press*, p. 2770 – 2781, 1976.
- [11] REIS, M. A. *Aplicação de técnicas de espalhamento de raios X na caracterização estrutural de proteínas e modelagem computacional utilizando vínculos experimentais obtidos por SAXS*. 2013. 150 f. Tese (Doutorado em Ciências) - Departamento de Química, Universidade Estadual de Campinas. Dezembro 2013.

- [12] REIS, M. A.; APARICIO, R.; ZHANG, Y. Improving protein template recognition by using Small-Angle X-Ray Scattering Profiles. *Biophysical Journal*, v. 101, n. 11, p. 2270 – 2781, December - 2011.
- [13] HALLIDAY, D. *Fundamentos de física: gravitação, ondas e termodinâmica*. 9. ed. Rio de Janeiro: Gen LTC, 2013.
- [14] TIRION, M. M. Large amplitude elastic motions in proteins from a single-parameter, Atomic Analysis. *Phys. Rev. Lett.*, v. 77, n. 9, p. 1905 – 1908, August - 1996.
- [15] MENDONÇA, M. R. *Análise de modos normais em proteínas*. 2010. 71 f. Dissertação (Mestrado em Ciências) - Departamento de Física e Matemática, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo. 2010.
- [16] TAMA, F.; SANEJOUAND, Y. H. Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, v. 14, n. 1, p. 1 – 6, 2001.
- [17] DEBYE, P. A. *Physik*, v. 46, p. 809 – 823, 1915.
- [18] ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, v. 57, n. 4, p. 702 – 710, 2004.
- [19] XU, J.; ZHANG, Y. How significant is a protein structure similarity with TM-score= 0.5? *Bioinformatics*, v. 26, n. 7, p. 889 – 895, 2010.