

Carolina Tatiani Alves Ferreira

Desenvolvimento de ferramentas
computacionais
para o estudo de macromoléculas.

São José do Rio Preto

2016

Desenvolvimento de ferramentas computacionais para o estudo de macromoléculas.

Carolina Tatiani Alves Ferreira

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto-SP, para o exame de qualificação como parte das exigências para a obtenção do Título de Mestre em Biofísica Molecular.

Orientador: Prof. Dr. Leandro Cristante de Oliveira

São José do Rio Preto - SP

31 de agosto de 2016.

Ferreira, Carolina Tatiani Alves.

Desenvolvimento de ferramentas computacionais para o estudo de macromoléculas / Carolina Tatiani Alves Ferreira. -- São José do Rio Preto, 2016

80 f. : il., tabs.

Orientador: Leandro Cristante de Oliveira

Dissertação (mestrado) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Biofísica. 3. Macromoléculas. 4. Simulação (Computadores digitais) 5. Raios X - Espalhamento a baixo ângulo. 6. Algoritmos. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. II. Título.

CDU – 577.3

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Desenvolvimento de ferramentas computacionais para o estudo de macromoléculas.

Carolina Tatiani Alves Ferreira

Dissertação apresentada ao Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto-SP, para o exame de qualificação como parte das exigências para a obtenção do Título de Mestre em Biofísica Molecular.

Banca Examinadora:

Prof. Dr. Leandro Cristante de Oliveira
Professor assistente doutor
Unesp - São José do Rio Preto - SP
Orientador

Prof. Dr. José Ramon Abrego
Professor assistente doutor
Unesp - São José do Rio Preto - SP

Prof. Dr. Mario de Oliveira Neto
Professor assistente doutor
Unesp - Botucatu - SP

São José do Rio Preto - SP

31 de agosto de 2016.

*Por confiar no futuro acadêmico da criança teimosa e curiosa, te
dedico, Daniela Caprioglio (in memoriam).*

*“Se deres uma resposta a um homem, dás-lhe apenas um fato. Se
lhe deres uma pergunta, fá-lo-ás procurar as suas próprias
respostas.”*

Patrick Rothfuss, O Temor do Sábio

Agradecimentos

- Agradeço à minha família por toda confiança que demonstraram em relação às minhas escolhas nos últimos anos, especialmente aos meus pais por todo o amor e carinho, além de serem o melhor exemplo profissional que eu poderia ter.
- Agradeço ao meu orientador Leandro pela paciência e por me proporcionar uma excelente orientação através de uma nova linha de pesquisa. Aos meus colegas de grupo pelas contribuições, especialmente ao Victor por colaborar neste trabalho.
- Aos meus amigos Alex, Fernando, Ingrid e Janine, pela companhia e perseverança durante esses anos e por me permitirem fazer parte dessa família.
- Ao Guilherme pelo apoio inabalável.
- À CAPES pelo apoio financeiro.
- Aos meus professores por construírem meu conhecimento.
- À todos que foram parte da minha vida nesse período e, portanto, me ajudaram a construir esta conquista, eu agradeço e ofereço meu carinho.

Resumo

Estudos recentes têm explorado o comportamento de macromoléculas em solução e a implicação das mudanças conformacionais. Apesar da importância deste assunto, há um número limitado de técnicas para observar as alterações nestes sistemas. Isto se deve a complexidade e aos altos custos envolvidos. Uma técnica que apresenta vantagens com relação aos tamanhos das proteínas estudadas e baixos custos em comparação à outras similares é o espalhamento de raios-X à baixo ângulo, ou SAXS - do inglês *small angle X-ray scattering*. Entretanto, os resultados possuem uma baixa resolução, pois consiste apenas no "envelope" da partícula e não possui nenhuma informação sobre as estruturas secundárias, terciárias ou quaternárias. O objetivo deste trabalho é o desenvolvimento de uma ferramenta capaz de descrever teoricamente estruturas terciárias em solução, baseado no conhecimento das estruturas secundárias e do perfil de espalhamento de SAXS. As simulações geram novas conformações por um algoritmo de pivot. O critério de Metropolis minimiza a energia potencial formada pela energia de Lennard-Jones e um termo associado à similaridade entre os perfis de espalhamento teórico e experimental. Esta ferramenta obteve resultados iniciais satisfatórios e, após uma fase de calibração, estará disponível *online* à comunidade científica, no site: <http://oliveira.df.ibilce.unesp.br/>.

Palavras-chave: Espalhamento de raios-X à baixo ângulo. SAXS. Simulação computacional. Algoritmo de pivot.

Abstract

Recent studies have been exploring the dynamical behavior of macromolecules in solution and the implication of conformational changes. Despite all the importance revolving this subject, there are a limited number of techniques to observe these system variations. This is due to the complexity and costs involved. One interesting technique that presents no protein size limitations and has relatively low costs is the Small Angle X-ray Scattering (SAXS). However, it provides low-resolution results, consisting only of the particle's "envelope" that does not provide any accuracy on secondary, tertiary or quaternary structures. This work aims the development of a tool to provide theoretical structural description of representative conformations in solution based only on the secondary structure informations and the experimental scattering profile. The simulations generates new conformations by a pivot algorithm. The Metropolis criteria minimizes the potential energy composed by a Lennard-Jones term and an energetic term related to the similarity of the experimental and theoretical scattering profiles. The software achieved some satisfactory initial results and it will be available online to the scientific community, after a calibration phase, at <http://oliveira.df.ibilce.unesp.br/>.

Keywords: Small angle X-ray scattering. SAXS. Computational simulation. Pivot algorithm.

Lista de Figuras

| | | |
|------|---|----|
| 1.1 | Complexo do CNPEM e esquema da estrutura do Sirius no LNLS. | 14 |
| 2.1 | Ilustração geométrica da lei da reciprocidade. | 17 |
| 2.2 | Esboço de curvas de espalhamento qualitativas para a incidência de ondas com o mesmo comprimento de onda sobre partículas de diferentes tamanhos. | 17 |
| 2.3 | Esquema de montagem do experimento de SAXS. | 18 |
| 2.4 | Espalhamento de raios-X por dois elétrons. | 20 |
| 2.5 | Curvas de espalhamento devido à diferentes geometrias. | 28 |
| 2.6 | Gráfico de Guinier para a lisozima. | 31 |
| 2.7 | Representação de uma partícula e seu “fantasma”, formado pelo deslocamento. | 32 |
| 2.8 | Gráfico de Porod para a lisozima. | 32 |
| 2.9 | Diferentes curvas de espalhamento analisadas pelo gráfico de Kratky. | 33 |
| 2.10 | Macromolécula envolta por uma camada de hidratação de espessura Δ | 34 |
| 3.1 | Relação entre $f(q)$ e q para diferentes elementos químicos . . . | 39 |
| 3.2 | Fluxograma do algoritmo utilizado em SAXS-factor. | 41 |
| 3.3 | Perfil de espalhamento da Lisozima experimental em comparação com o teórico obtido pelo SAXS-factor. | 43 |

| | | |
|------|--|----|
| 3.4 | Análise comparativa entre os resultados do SAXS-factor e programas similares para a Lisozima. | 44 |
| 3.5 | Perfil de espalhamento de 28 pares de bases de DNA experimental em comparação com o teórico obtido pelo SAXS-factor. | 46 |
| 3.6 | Análise comparativa entre os resultados do SAXS-factor e programas similares para 28 pares de bases de DNA. | 47 |
| 3.7 | Perfil de espalhamento da Xilanase experimental em comparação com o teórico obtido pelo SAXS-factor. | 48 |
| 3.8 | Análise comparativa entre os resultados do SAXS-factor e programas similares para a Xilanase. | 49 |
| 3.9 | Análise comparativa entre os resultados do SAXS-factor e programas similares para a Vilina. | 50 |
| 3.10 | Análise comparativa entre os resultados do SAXS-factor e programas similares para a 1ENH. | 51 |
| 3.11 | Análise comparativa entre os resultados do SAXS-factor e programas similares para a 1YPC. | 51 |
| 3.12 | Representação do <i>backbone</i> de um aminoácido | 52 |
| 3.13 | Representação das cadeias laterais de todos os aminoácidos essenciais. | 53 |
| 3.14 | Representação dos ângulos em um aminoácido. | 54 |
| 3.15 | Conformação totalmente planar de uma ligação peptídica. Os ângulos ϕ e ψ estão em posição de zero. | 54 |
| 3.16 | Ramachandran <i>plot</i> observado para a Glicina. | 55 |
| 3.17 | Representação do angulos de euler com relação aos eixos cartesianos. | 56 |
| 3.18 | Representação do sistema de coordenadas 'linha' formado a partir da rotação de ω em torno de Z | 57 |
| 3.19 | Fluxograma do algoritmo utilizado em Ramatrix. | 59 |

| | | |
|------|--|----|
| 3.20 | Análise comparativa entre os resultados do Ramatrix e da literatura. | 61 |
| 3.21 | Análise comparativa entre os resultados do Ramatrix e da literatura. | 62 |
| 4.1 | Fluxograma do algoritmo utilizado em Pivot-SAXS. | 65 |
| 4.2 | Mudança estrutural da vilina causada pelo movimento de pivot. | 67 |
| 4.3 | Mudança estrutural da 1ENH causada pelo movimento de pivot. | 68 |
| 4.4 | Estruturação dos domínios α e β da proteína 1YPC causada pelo movimento de pivot. | 68 |
| A.1 | Fluxograma do critério de Metropolis. | 80 |

Lista de Tabelas

| | | |
|-----|---|----|
| 3.1 | Valores de a_i , b_i e c para diversos elementos químicos. | 40 |
| 4.1 | Valores de RMSD para cada uma das proteína na conformação aberta e final com relação à estrutura cristalográfica. | 68 |

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução. | 13 |
| 2 | Espalhamento de raios-X à baixo ângulo | 16 |
| 2.1 | Fundamentação teórica. | 16 |
| 2.1.1 | Lei da reciprocidade. | 16 |
| 2.1.2 | Espalhamento da Luz | 18 |
| 2.1.3 | A interferência e o vetor momento de transferência \vec{q} | 20 |
| 2.1.4 | Espalhamento por vários elétrons. | 21 |
| 2.1.5 | Função de auto-correlação. | 23 |
| 2.1.6 | Aproximação de Debye e função de correlação. | 24 |
| 2.1.7 | Sistema em solução. | 26 |
| 2.1.8 | Análises das curvas de SAXS. | 29 |
| 2.1.9 | Análise computacional dos dados de SAXS. | 33 |
| 3 | Resultados preliminares. | 37 |
| 3.1 | Desenvolvimento do SAXS-factor. | 37 |
| 3.2 | Calibração do SAXS-factor. | 42 |
| 3.2.1 | Moléculas modelo do banco de dados. | 42 |
| 3.2.2 | Moléculas retiradas do PDB. | 50 |
| 3.3 | Estudo dos ângulos diedrais de uma proteína. | 52 |
| 3.3.1 | Emprego de propriedades geométricas. | 52 |
| 3.3.2 | Ângulos de Euler. | 55 |
| 3.3.3 | Estrutura do programa Ramatrix. | 58 |

| | | |
|----------|--|-----------|
| 3.3.4 | Validação do programa Ramatrix. | 58 |
| 4 | Construção de potenciais para simulações de previsão conformacional com o Pivot-SAXS. | 63 |
| 4.1 | Motivações. | 63 |
| 4.2 | Desenvolvimento do programa. | 64 |
| 4.3 | Validação e resultados do Pivot-SAXS. | 66 |
| 5 | Discussão. | 69 |
| 6 | Trabalhos Futuros. | 73 |
| | Referências. | 74 |
| A | Critério de Metropolis. | 79 |

Capítulo 1

Introdução.

Fontes de luz de alto brilho são de grande importância para o desenvolvimento científico em diversas áreas do conhecimento. É uma das ferramentas com maior número de aplicações, auxiliando na solução de problemas relativos a energia, saúde e meio ambiente, por exemplo.

Neste aspecto, o Brasil está na vanguarda do desenvolvimento científico. Possuímos o único laboratório de luz síncrotron da América Latina, o LNLS - Laboratório Nacional de Luz Síncrotron. Há duas fontes de luz pertencentes a este laboratório: o UVX e uma segunda fonte, de maior potência, está em construção. Este é o projeto Sirius. Ambos localizados no campus do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) em Campinas-SP.

O Sirius vêm para manter o país em uma posição proeminente no cenário tecnológico mundial. Com um investimento orçado em 650 milhões de reais entre a construção do prédio e compra de equipamentos de ponta para o estudos de materiais. A estrutura abrigará um anel de cerca de 518 metros de circunferência e 165 metros de diâmetro, até 40 linhas de luz podendo alcançar energias de 3 GeV¹, um esquema da estrutura pode ser visto na

¹Dados obtidos do site do projeto: <http://lnls.cnpem.br/sirius/?la=br>

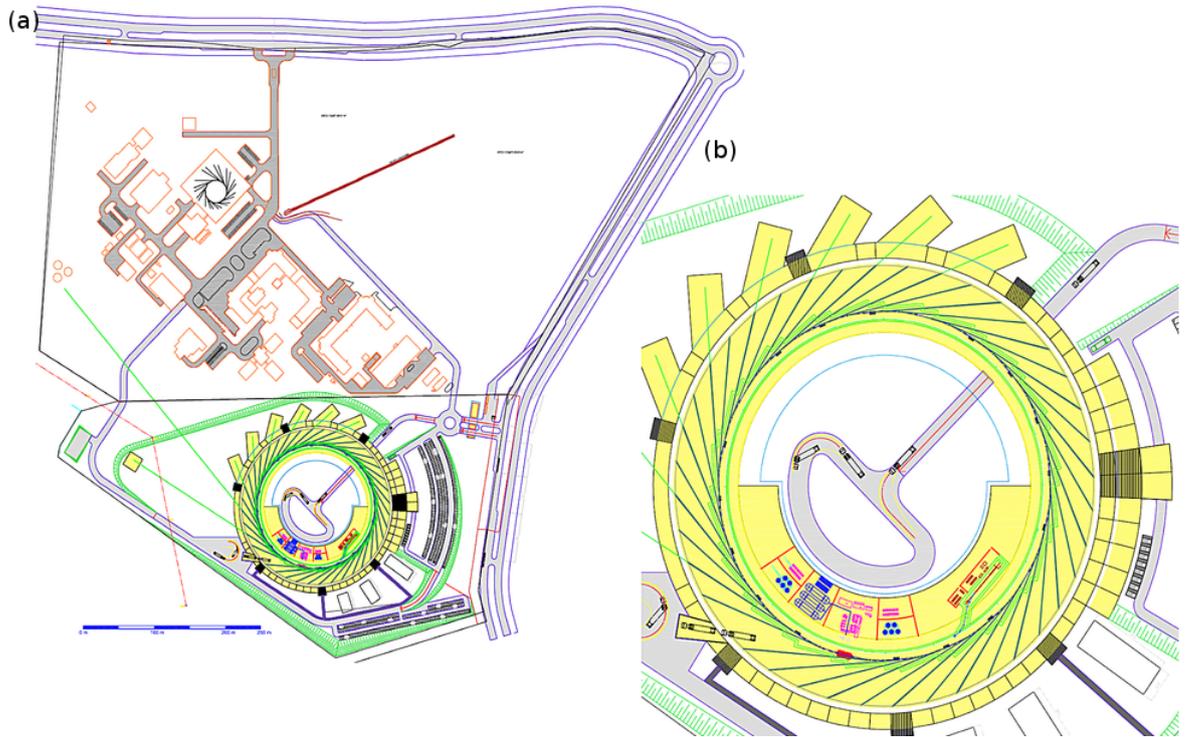


Figura 1.1: Complexo do CNPEM e esquema da estrutura do Sirius no LNL. Na figura (a) observa-se toda a área do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) e é possível comparar o tamanho da fonte UVX, o pequeno anel no canto superior esquerdo, com a Sirius, o anel na parte baixa da figura. A figura (b) mostra o anel do Sirius em detalhes, incluindo suas 40 linhas de luz. Figura retirada do site do LNL.²

figura 1.1.

O UVX possui uma estrutura mais modesta com um anel de cerca de 93 metros de circunferência e 30 metros de diâmetro. 15 linhas de luz com energia de 1.37 GeV estão em funcionamento e os experimentos são divididos em duas categorias: espectroscopia de materiais e estudos estruturais de materiais. Nesta última entram os experimentos de difração de raios-X, cristalografia de macromoléculas e espalhamento de raios-X à baixo ângulo (SAXS). Todas estas técnicas são muito utilizadas no estudo de proteínas, incluindo o SAXS, do nome em inglês *Small Angle X-ray Scattering*, que

²<http://lnls.cnpe.br/sirius/facility-layout/>

permite o estudo das moléculas em solução e possui uma linhas exclusivas no UVX, a SAXS 1. Esta é uma técnica que foi muito desenvolvida devido à luz síncrotron. Antes de seu surgimento, a medida de espalhamento era dificultada pela pouca intensidade e colimação das fontes disponíveis. O raio-X originado de uma fonte de luz síncrotron possui alta intensidade e pouca divergência, isto tornou SAXS uma das principais formas de análise de nanoestruturas atualmente. Por ter adquirido um bom status no meio acadêmico, SAXS é muito utilizada em pesquisas com macromoléculas, especificamente, proteínas. A técnica permite o estudo de aspectos estruturais da partícula, como raio de giro, estado de enovelamento e mudanças conformacionais[1]. Simulações computacionais são utilizadas para outros tipos de análises. A reconstrução do envelope das proteínas[2–5] possibilita os estudos sobre a superfície da molécula, a simulação também é usada para aprofundar os estudos sobre enovelamento e mudanças conformacionais[6].

Por ser uma ferramenta já muito utilizada no área da física biológica, desenvolvemos um programa que amplia sua aplicabilidade para além dos estudos de superfície. O Pivot-SAXS utiliza o perfil de espalhamento e movimentos de pivot para determinar a estrutura terciária da proteína em solução. Esta informação é muito importante para o estudos de mecanismos, verificação de mudanças conformacionais através de ensambles, além da possibilidade de usá-lo em conjunto com preditores de estrutura secundária para estruturas não-cristalizáveis, aumentando a quantidade de proteínas que podem ser estudadas.

Nos capítulos seguintes detalharemos mais sobre a teoria do espalhamento de raios-X à baixo ângulo, demonstraremos os passos e explicaremos todas as etapas de validação que foram necessárias durante a construção do programa final.

Capítulo 2

Espalhamento de raios-X à baixo ângulo

2.1 Fundamentação teórica.

O espalhamento de raios-X à baixo ângulo[1] é uma técnica de análise espectroscópica, amplamente utilizada para análise conformacional e estrutural de moléculas biológicas. As amostras não precisam ser cristalizadas, podendo ser analisadas em solução e permitindo o estudo de sistemas em diferentes condições de pH e concentração. Também é um recurso de baixo custo quando comparado com outras técnicas que visa o estudo dos aspectos estruturais de sistemas isotrópicos e não-periódicos de diversos tamanhos.

2.1.1 Lei da reciprocidade.

O SAXS, assim como várias outras técnicas de espalhamento, baseia-se na chamada *lei da reciprocidade*, responsável pela redução proporcional entre o ângulo de espalhamento observável e o tamanho da partícula. A técnica é assim eficiente na avaliação de estruturas de tamanhos coloidais, na faixa de dezenas e milhares de Å, consideradas gigantes quando comparadas com o comprimento de onda dos raios-X, cerca de 1,54Å para as linhas

comumente usadas em experimentos com proteínas. Tal propriedade pode ser observada pela comparação entre partículas de tamanhos diferentes.

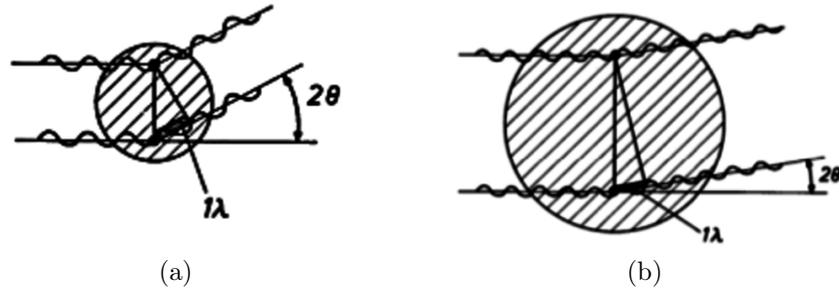


Figura 2.1: Ilustração geométrica da lei da reciprocidade. Figura extraída do livro de nome "Small Angle X-ray Scattering" de Glatter e Kratky[1]

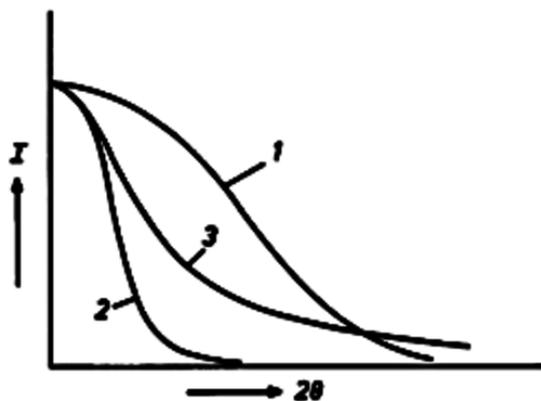


Figura 2.2: Esboço de curvas de espalhamento para a incidência de ondas com o mesmo comprimento de onda sobre partículas de diferentes tamanhos. Figura extraída de O.Glatter e O.Kratky, 1982[1]

As figuras 2.1 e 2.2 ilustram as ideias fundamentais do espalhamento pela lei da reciprocidade e por esboços de curvas de espalhamento. Na figura 2.1a duas ondas são espalhadas a um ângulo 2θ . Elas possuem uma diferença de caminho óptico de λ . Quando a contribuição de todos os elétrons é considerada, há uma interferência destrutiva na direção 2θ e, portanto, não há contribuição de espalhamento. Ao analisar ângulos menores a interferência se torna construtiva e a intensidade de espalhamento aumenta até atingir seu

máximo na direção em que o ângulo de espalhamento é zero e todas as ondas estão em fase. A curva de espalhamento gerada é representada pela curva 1 na figura 2.2. Considerando uma partícula maior, tal qual ilustrada na figura 2.1b, é possível perceber que a diferença de caminho ótico de 1λ ocorre para valores de 2θ relativamente menores. Neste caso, a interferência destrutiva se inicia em ângulos baixos o que torna a curva de espalhamento mais estreita e similar a curva 2 na figura 2.2.

Assim entende-se que o espalhamento de raio-X a baixo ângulo ocorre para partículas grandes em relação ao comprimento de onda do raio incidente, comprovando a lei da reciprocidade.

2.1.2 Espalhamento da Luz

Feixes de raios-X podem ser obtidos por luz síncrotron. Isso permite que a fonte seja contínua e de alta intensidade e colimação. Assim, é possível a realização de uma grande diversidade de experimentos com vários materiais e condições.

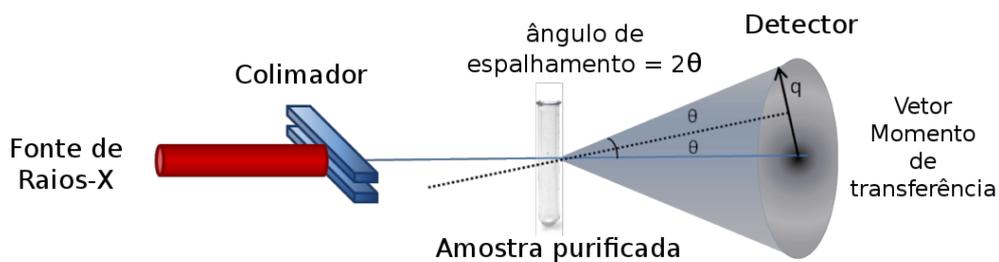


Figura 2.3: Esquema de montagem do experimento de SAXS. O sistema é mantido no vácuo. Figura extraída do artigo de Putnam, Lowe e Meiler[7].

O experimento está apresentado na figura 2.3 e consiste em uma fonte de raios-X paralelos ou colimados, um suporte para a amostra, um bloqueio na direção de incidência dos raios-X e um detector.

O feixe de raios-X deve incidir sobre a amostra fazendo com que os elétrons vibrem na mesma frequência. Devido a esta ressonância, os

elétrons emitem ondas secundária coerentes que interferem de forma destrutiva ou construtiva. A interferência destrutiva de todas as ondas secundárias espalhadas com ângulo 2θ leva a um espalhamento nulo nesta direção, e a máximo em $\theta = 0$. Na montagem o bloqueio, chamado *beamstop* é colocado neste ponto para evitar que o perfil seja distorcido pela alta intensidade, pois grande parte da radiação não interage com a amostra e, portanto, atingiria o coletor com ângulo $\theta = 0$. O *beamstop* impede que esta radiação afete os resultados, porém o valor de $I(0)$ também é perdido e este deve ser calculado computacionalmente¹.

É importante avaliar o balanço de energia do sistema. Nesta montagem, a energia de um fóton de raios-X é maior que a energia nuclear de ligação de um átomo, assim, todos os elétrons se comportarão como uma partícula espalhadora livre e as ondas espalhadas serão coerentes. A amplitude resultante, nomeada aqui por A , será dada pela soma das amplitudes A_n de todas as ondas espalhadas n como apresentado na equação 2.1.

$$A = \sum_{i=1}^n A_n \quad (2.1)$$

Este tipo de experimento envolve um onda eletromagnética interagindo com um elétron. Tal interação é chamada espalhamento elástico de Thompson. Ao ser espalhada, a onda perde uma quantidade de energia proporcional ao seu ângulo de espalhamento. E todas as ondas secundárias possuem a mesma intensidade dada por:

$$I_e(\theta) = I_p \cdot \frac{e^2}{mc^2} \cdot \frac{1}{r^2} \cdot \frac{1 + \cos^2(2\theta)}{2}, \quad (2.2)$$

onde I_p é a intensidade primária, ou seja, relativa à onda incidente, r se refere à distância entre a partícula espalhadora e o detetor. A relação (e^2/mc^2) é devido ao quadrado do raio do elétron clássico. A dependência

¹Cálculo explicado na seção 2.1.8.

da intensidade com o ângulo de espalhamento 2θ é bem sutil, podendo ser reduzida à 1 para os baixos valores de θ analisados no experimento.

2.1.3 A interferência e o vetor momento de transferência \vec{q} .

Por convenção, a amplitude e a intensidade das ondas espalhadas assumiram os valores de 1 por facilidade nos cálculos, fazendo com que somente exista diferença na fase, que está relacionada com a posição no espaço do elétron espalhador. Cada onda espalhada é representada pela função complexa $e^{i\varphi}$, onde φ indica a fase.

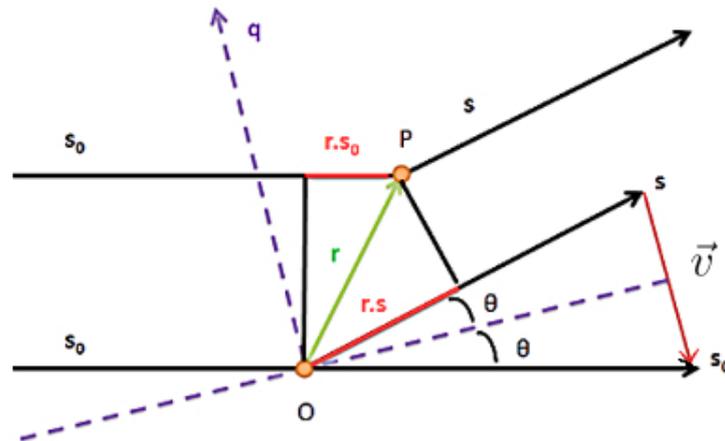


Figura 2.4: Espalhamento de raios-X por dois elétrons para uma partícula fixa. Figura modificada do artigo de Putnam, Lowe e Meiler[7].

A magnitude da fase é dada pelo fator $2\pi/\lambda$ multiplicado pela diferença de caminho entre as ondas incidente e espalhada, cujas direções são representadas na figura 2.4, por S_0 e S , respectivamente. O ponto P, na mesma figura, é dado pela distância \vec{r} até a origem O. Há uma diferença de caminho óptico entre estes pontos de $-\vec{r} \cdot \vec{v}$, em que \vec{v} é o vetor diferença entre a onda espalhada e a projeção da onda incidente. Assim, a fase será dada pela equação:

$$\varphi = -\frac{2\pi}{\lambda} \vec{r} \cdot \vec{v} \quad (2.3)$$

Um outro importante parâmetro que pode ser obtido é o vetor momento de transferência \vec{q} , também apresentado na figura 2.4, e que define a curva de espalhamento no espaço recíproco, possuindo unidade inversas de comprimento. Tal grandeza independe da distância até o detetor e do tipo de radiação incidente e é utilizada para comparação entre diferentes técnicas de espalhamento, tais quais o SAXS e a difração. O vetor \vec{q} possui módulo dado por:

$$|\vec{q}| = \frac{4\pi \sin \theta}{\lambda}. \quad (2.4)$$

Os vetores \vec{q} e \vec{v} possuem a mesma direção e, segundo o esboço da figura 2.4, a magnitude de \vec{v} é de $2 \sin \theta$, com θ sendo a metade do ângulo de espalhamento. Então ao substituir este valor na equação 2.4, obtêm-se o vetor \vec{q} :

$$\vec{q} = \frac{2\pi}{\lambda} \times \vec{v} \quad (2.5)$$

Analisando as equações 2.3 e 2.5, nota-se que a fase pode ser escrita como função do vetor momento de espalhamento e da distância \vec{r} .

$$\varphi = -\vec{q} \cdot \vec{r} \quad (2.6)$$

2.1.4 Espalhamento por vários elétrons.

Considerando que a amostra possui uma grande quantidade de elétrons, não precisamente localizados, o conceito de densidade eletrônica pode se fazer útil. A densidade eletrônica é definida como o número de elétrons por unidade de volume dV . No experimento de SAXS, a diferença de densidade eletrônica (ρ) entre a amostra e o solvente é observado. Se a $\rho_{amostra}$ é igual à $\rho_{solvente}$ não há espalhamento observável, pois as ondas espalhadas se

cancelam. Para obter a densidade eletrônica apenas da amostra, é subtraída a densidade eletrônica do solvente puro e a densidade eletrônica da solução (soluto + solvente). Em casos pela literatura científica[8], o termo densidade eletrônica pode ser usado para expressar a ideia de contraste ou diferença de densidade eletrônica entre soluto e solvente. Este conceito permite dizer que a amplitude A_n da onda espalhada por uma unidade de volume dV é dada por:

$$A_n = dV \cdot \rho(\vec{r})e^{-i\vec{q}\cdot\vec{r}} \quad (2.7)$$

Onde $e^{-i\vec{q}\cdot\vec{r}}$ é a representação complexa da onda espalhada pelo volume dV com densidade eletrônica $\rho(\vec{r})$, com \vec{q} dado pela equação 2.4. Ao transformar o somatório da equação 2.1 em uma integral sobre toda o volume V irradiado pelo feixe é possível calcular a amplitude A .

$$A(\vec{q}) = \iiint dV \cdot \rho(\vec{r})e^{-i\vec{q}\cdot\vec{r}} \quad (2.8)$$

Pode-se notar que a amplitude é dada, matematicamente, pela transformada de Fourier de $\rho(\vec{r})$ e, pela a propriedade da transformada inversa, $\rho(\vec{r})$ pode ser calculada a partir da amplitude $A(\vec{q})$ pela equação 2.9, tornando-a proporcional ao módulo $|A(\vec{q})|$ e a fase φ por meio da exponencial $e^{-i\vec{q}\cdot\vec{r}}$.

$$\rho(\vec{r}) = \iiint dV \cdot A(\vec{q})e^{-i\vec{q}\cdot\vec{r}} \quad (2.9)$$

O cálculo de $\rho(\vec{r})$ é relevante por ser uma variável real e positiva, porém a amplitude não é uma quantidade possível de ser medida. A intensidade de espalhamento é obtida experimentalmente e é proporcional ao quadrado da amplitude da forma explicitada na equação 2.10, onde A^* é o complexo conjugado da amplitude.

$$I(\vec{q}) \propto A^*A = |A(\vec{q})|^2 e^{-i\vec{q}\cdot\vec{r}} e^{i\vec{q}\cdot\vec{r}} \quad (2.10)$$

Dessa forma, a intensidade é proporcional apenas ao módulo da amplitude e a fase $\varphi = \vec{q} \cdot \vec{r}$ é perdida durante o cálculo, impossibilitando que $\rho(\vec{r})$ seja devidamente conhecido. Tal situação leva a um entrave chamado de *problema da fase*, discutido a seguir.

2.1.5 Função de auto-correlação.

O problema da fase impede o cálculo da densidade eletrônica diretamente, como exposto na seção anterior. Todavia, a relação entre a amplitude e a intensidade na equação 2.10 apresenta uma abordagem alternativa. Em seus estudos, Patterson[9] utilizou-se das relações de proporcionalidade entre a amplitude e a intensidade estabelecida pela equação 2.10, assim como da proporção entre a amplitude e a densidade eletrônica, demonstrada na equação 2.8.

Seguindo suas ideias é possível obter a intensidade como a transformada de Fourier de ρ .

$$I(\vec{q}) = \iiint \iiint dV_1 \cdot dV_2 \cdot \rho(\vec{r}_1)\rho(\vec{r}_2)\exp[-i\vec{q}(\vec{r}_1 - \vec{r}_2)] \quad (2.11)$$

Ao considerar uma partícula diluída a integral 2.11 torna-se relativa, apenas, às propriedades do elementos dV da própria partícula, ou seja, da distância intra partícula $\vec{r} = \vec{r}_2 - \vec{r}_1$ e das densidades eletrônicas ρ de cada um. Por ser uma integral de resolução complexa, a solução deve ser feita em dois passos:

- **Cálculo da função de auto-correlação $\tilde{\rho}(\vec{r})$:** Também conhecida como função de Patterson [1] é definida por uma relação entre o espaço real e o chamado espaço recíproco. Cada par disposto à uma distância relativa $\vec{r} = (\vec{r}_2 - \vec{r}_1)$ constante será representado por um ponto no espaço recíproco e a densidade desses pontos é dada por $\tilde{\rho}(\vec{r})$ na equação

2.12. Como cada par é contabilizado duas vezes, o espaço recíproco deverá possuir um centro de simetria.

$$\tilde{\rho}^2(\vec{r}) = \iiint dV_1 \rho(\vec{r}_1) \rho(\vec{r}_2). \quad (2.12)$$

- **Integração sobre todo o espaço recíproco:** Esse segundo passo leva á outra transformada de Fourier e prova que a intensidade depende apenas da estrutura do objeto de estudo.

$$I(\vec{q}) = \iiint dV \cdot \tilde{\rho}^2(\vec{r}) \cdot e^{i\vec{q}\cdot\vec{r}}. \quad (2.13)$$

Assim, a função de auto-correlação pode ser obtida a partir da transformada inversa da intensidade $I(\vec{q})$, dada por:

$$\tilde{\rho}^2 = \left(\frac{1}{2\pi}\right)^3 \iiint dq_x dq_y dq_z \cdot I(\vec{q}) \cdot e^{i\vec{q}\cdot\vec{r}}, \quad (2.14)$$

As equações acima demonstram uma relação de proporcionalidade inversa entre o espaço real e o espaço recíproco, pois são ligadas apenas pelo fator $e^{i\vec{q}\cdot\vec{r}}$. A equação demonstra matematicamente a lei da reciprocidade discutida na seção 2.1.1.

2.1.6 Aproximação de Debye e função de correlação.

É possível utilizar duas restrições, encontradas na maioria dos casos, para facilitar as análises das curvas.

- **Sistema estatisticamente isotrópico:** O espalhamento não é afetado pela direção de incidência, impondo que a distribuição de pares de distâncias depende apenas de $|\vec{r}|$ e confirma a centro-simetria do espaço recíproco. Devido a independência da distribuição de pares com relação

a direção, é possível substituir o fator de fase pela sua média sobre todas as direções.

$$\langle e^{i\vec{q}\cdot\vec{r}} \rangle_{\vec{r}} = \frac{\sin qr}{qr}. \quad (2.15)$$

Esta equação é chamada fórmula fundamental de Debye[1]. Com esta substituição a intensidade definida na equação 2.13 se torna:

$$I(q) = 4\pi \int_0^\infty r^2 \tilde{\rho}^2 \frac{\sin qr}{qr}. \quad (2.16)$$

- **Ausência de correlação entre pontos distantes:** No limite de grandes valores de \vec{r} , as densidades de elétrons $\rho_1(\vec{r}_1)$ e $\rho_2(\vec{r}_2)$ se tornam independentes e podem ser substituídas pela média $\langle \rho \rangle$, fazendo a função de Patterson tender para um valor constante $V \langle \rho \rangle$ e assumindo, na origem, seu valor máximo igual a $\tilde{\rho}^2(0) = V \overline{\rho^2}$.

A informação estrutural se concentra na região finita, em que $\langle \rho \rangle^2$ difere do valor final. Tal valor, por ser constante, não contribui para o padrão de difração, não tendo, portanto, relevância para as investigações sobre as estruturas estudadas.

Convém utilizar a flutuação de densidade eletrônica $\eta = \rho - \langle \rho \rangle$ por ser mais informativa que apenas a densidade eletrônica ρ . Deve-se, então redefinir a função de auto-correlação 2.12.

$$\tilde{\eta}^2 = \widetilde{(\rho - \langle \rho \rangle)^2} = V \cdot \gamma(r) \quad (2.17)$$

Na equação 2.17 foi introduzida a chamada função de correlação[10] $\gamma(r)$ que pode ser definida pela equação 2.18 com $r = |r_1 - r_2|$ e interpretada como a média do produto de duas flutuações à uma distância r uma da outra

$$\gamma(r) = \langle \eta(r_1)\eta(r_2) \rangle \quad (2.18)$$

Esta função também segue as restrições acima tendo $\gamma(0) = \langle \eta^2 \rangle$ e $\gamma \rightarrow 0$ para r muito grande.

2.1.7 Sistema em solução.

Em sistemas diluídos, como uma proteína em meio intracelular, o contraste entre a partícula (ρ) e o meio (ρ_0) é uma medida viável e informativa. É possível relacionar a função de correlação $\gamma(\vec{r})$ e $(\Delta\rho)^2 = (\rho - \rho_0)^2$ usando uma função γ_0 chamada *função intrínseca*[11], dada por:

$$\gamma(r) = (\Delta\rho)^2 \gamma_0(r) \text{ com } \begin{cases} \gamma_0(0) = 1 \\ \gamma_0(r \geq D_{m\acute{a}x}) = 0 \end{cases}, \quad (2.19)$$

onde $D_{m\acute{a}x}$ é a maior distância entre pontos internos da proteína e γ_0 depende apenas da geometria da partícula. Assim a intensidade de espalhamento, apresentada na equação 2.20, pode ser escrita de forma generalizada e intuitiva como uma transformada de Fourier da função intrínseca, possibilitando o cálculo da mesma pela transformada inversa na equação 2.21.

$$I(q) = 4\pi V(\Delta\rho)^2 \int_0^{D_{m\acute{a}x}} r^2 dr \cdot \gamma_0(r) \frac{\sin qr}{qr}, \quad (2.20)$$

$$V(\Delta\rho)^2 \gamma_0(r) = \int_0^\infty q^2 dq \cdot I(q) \frac{\sin qr}{qr} \quad (2.21)$$

A intensidade é uma variável medida experimentalmente, portanto é uma vantagem se ter uma quantidade possível de ser calculada através da mesma. A distribuição de distância de pares permite uma análise da geometria da partícula estudada e, também pode ser calculada a partir da intensidade, pois se relaciona com a função intrínseca da forma $p(r) = 4\pi r^2 \gamma_0(r)$. O espaço real e recíproco são, então, unidos pela equação 2.22.

$$I(q) = V(\Delta\rho)^2 \int_0^{D_{m\acute{a}x}} p(r) \cdot dr \frac{\sin qr}{qr} \quad (2.22)$$

A distribuição de pares $p(r)$ possui as mesmas informações que a intensidade, porém a $p(r)$, por estar representando o espaço real, possui uma visualização mais imediata sobre a geometria da molécula.

A figura 2.5(b) apresenta diversos objetos com a mesma distância interna máxima ($D_{m\acute{a}x}$) e as respectivas curvas de intensidade e distâncias de pares. As diferenças nas curvas de $p(r) \times r$ são bastante características de cada formato, estruturas globulares possuem uma curva gaussiana com o máximo, aproximadamente, em $D_{m\acute{a}x}/2$. Partículas achatadas ou alongadas causam um deslocamento do máximo para $r < D_{m\acute{a}x}/2$ enquanto para cascas esféricas o deslocamento é para $r > D_{m\acute{a}x}/2$. Estruturas compostas por subunidades separadas produzem dois máximos, um relativo às distâncias internas de cada subunidade e outro à distância entre as subunidades.

Para auxiliar os cálculos posteriores é possível estabelecer uma quantidade Q chamada de *invariante*. Em situações em que o fator de Debye $\frac{\sin qr}{qr}$ assume valor 1, ou seja, quando $r = 0$ ou $q = 0$, a intensidade $I(0)$ é dada pela equação 2.23 mas não pode ser medida, precisando ser extrapolada computacionalmente. Tornando $V\gamma(0)$, na equação 2.24, uma melhor opção de cálculo, pois evita a necessidade de calcular $I(0)$.

$$I(q) = 4\pi V(\Delta\rho)^2 \int_0^{D_{m\acute{a}x}} r^2 dr \cdot \gamma_0(r) \frac{\sin qr}{qr} \approx (\Delta\rho)^2 V^2 \quad (2.23)$$

$$V\gamma(0) = \frac{1}{2\pi^2} \int_0^\infty q^2 dq \cdot I(q) = V\overline{\eta^2}, \quad (2.24)$$

onde $\overline{\eta^2}$ é a média do quadrado da flutuação e, portanto $V\overline{\eta^2}$ é constante. Esta equação indica a conexão da intensidade com a flutuação de elétrons da partícula e não apenas com sua forma. Assim, se a partícula sofrer algum tipo de deformação estrutural sem que haja rupturas, mesmo que o padrão de espalhamento seja alterado, o valor Q na equação 2.25 deve se manter a

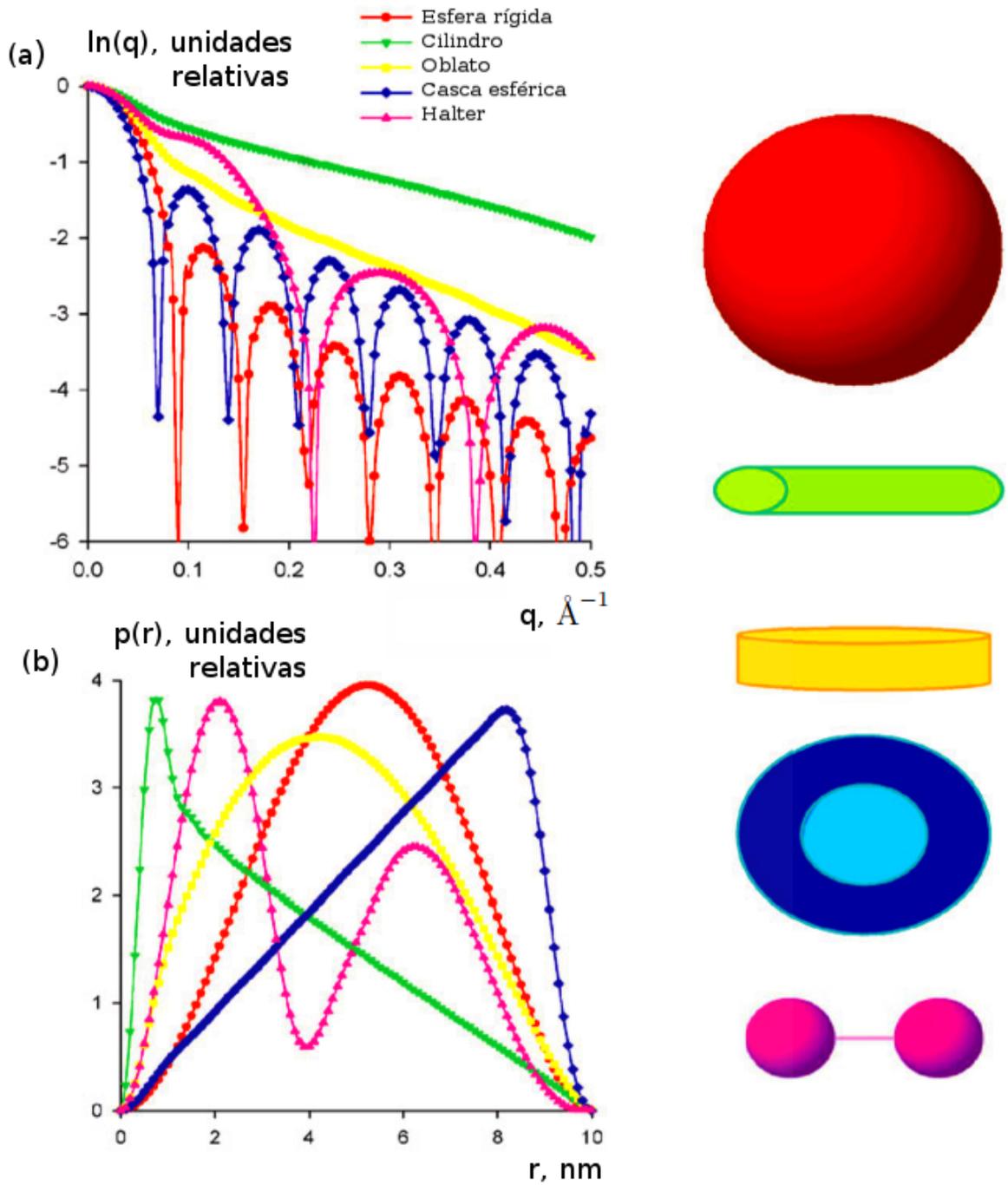


Figura 2.5: Na figura (a) observa-se o perfil de espalhamento devido à diferentes formas geométricas. E (b) está a distribuição de distância de pares para estas mesmas figuras. Figura modificada do artigo de Svergun e Koch[5].

mesma.

$$Q = \int_0^{\infty} q^2 dq \cdot I(q) \quad (2.25)$$

Por ser um valor constante pode ser escrita como na equação 2.25 ao considerar a intensidade calculada na por 2.23.

$$Q = 2\pi^2(\Delta\rho)^2V \quad (2.26)$$

2.1.8 Análises das curvas de SAXS.

A curva de espalhamento refere-se a estrutura externa, também chamada de *envelope* da molécula. É possível determinar algumas outras propriedades ao analisar as três regiões da curva: pequenos ângulos de espalhamento ($q \rightarrow 0$), para determinação do raio de giro (R_g), maiores ângulos de espalhamento ($q \rightarrow \infty$) e também convém estudar a região central para informações sobre estrutura interna, para a estrutura mais superficial.

Considerar a equação de Debye 2.15 para valores pequenos de q , ou seja, para **a região de Guinier**, possibilita o uso de uma expansão em série de potências[8] em série de McLaurin como apresentada na equação 2.27. Geralmente, usa-se os primeiros termos desta série na equação 2.16 para $q = 0$, obtendo assim a equação 2.28 que relaciona a intensidade e a distância quadrática média dos elétrons até o centro de massa da molécula, chamado de raio de giro, R_g .

$$\frac{\sin qr}{qr} = 1 - \frac{q^2 r^2}{6} + \frac{q^4 r^4}{120} + \dots \quad (2.27)$$

$$I(q) = I(0) \left[1 - \left(\frac{q^2 R_g^2}{6} \right) \right]. \quad (2.28)$$

A região de Guinier é delimitada pela regra:

$$q \cdot R_g < 1, 3. \quad (2.29)$$

Guinier e Fournet[12] observa que a intensidade pode ser escrita como uma função exponencial 2.30, pois o fator de expansão $1 - \frac{q^2 R_g^2}{6}$ também pode ser considerado como os primeiros termos de uma expansão da exponencial $\exp \frac{-q^2 R_g^2}{3}$.

$$I(q) = I(0)e^{-\frac{q^2 R_g^2}{3}} \quad (2.30)$$

Assim, o gráfico de guinier, dado por $\ln I(q) \times q^2$ e demonstrado na figura 2.6 e segue a equação 2.30 linearizada,

$$\ln I(q) = \ln I(0) - \frac{q^2 R_g^2}{6}. \quad (2.31)$$

Dessa forma, o gráfico de guinier será uma reta com coeficiente linear igual à $\ln I(0)$ e coeficiente angular proporcional ao raio de giro.

A **área central da curva** $I(q)$ está ligada com a estrutura e dimensões gerais da partícula. O volume excluído, também chamado de volume de Porod, é definido como o volume da intersecção entre uma partícula e o seu “fantasma” após um deslocamento de r , onde r é menor que o tamanho da partícula. A construção é representada na figura 2.7.

O volume de Porod é, portanto, uma dimensão importante que pode ser calculada pela equação 2.32 e está diretamente ligada ao valor γ_0 , de forma que os pontos contidos neste espaço são apenas àqueles que contribuem efetivamente para o cálculo da mesma.

$$V = 4\pi \int_0^D r^2 dr \cdot \gamma_0(r). \quad (2.32)$$

A intensidade em $q = 0$ dada pela equação 2.23 dividida pela inva-

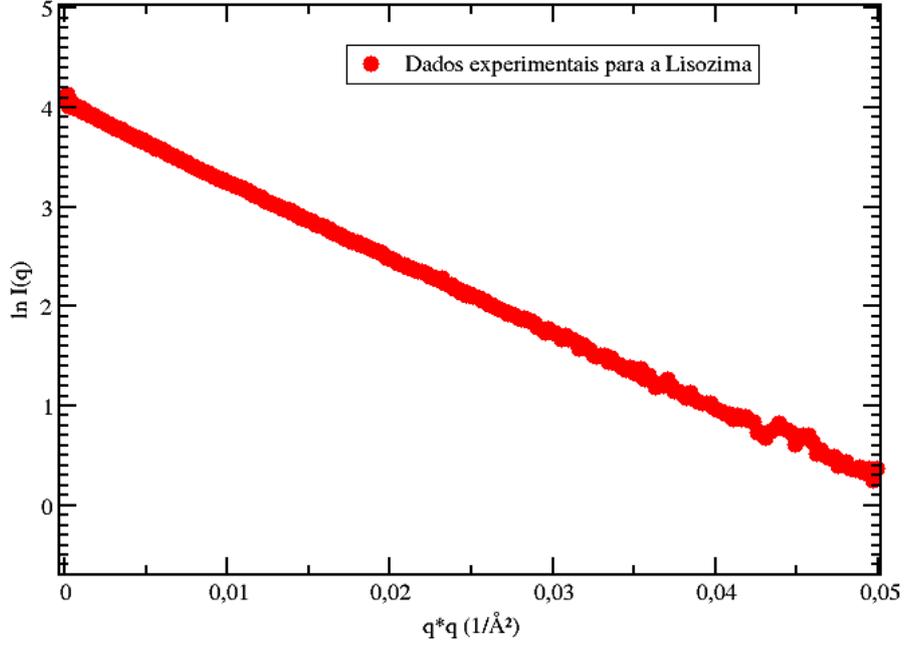


Figura 2.6: Gráfico de Guinier para a lisozima. Usado para o cálculo do raio de giro R_g da proteína e de $I(0)$ a partir da eq. 2.31. Gráfico plotado a partir de dados extraídos do BIOISIS[13]

riante de Porod da equação 2.26 leva à uma relação direta entre o volume excluído e a invariante de Porod. Essa relação permite que o volume não possua unidades destoantes dos resultados experimentais.

$$V = 2\pi^2 \frac{I(0)}{Q} \quad (2.33)$$

Sendo que $I(0)$ deve ser extrapolado usando a aproximação de guinier, discutida anteriormente.

A análise da **região de grandes q** , ou seja, com $q \rightarrow \infty$ é válida devido à sua dependência com a estrutura mais próxima da superfície da partícula e não deve ser influenciada por suas propriedades gerais. A relação entre q^{-4} e a intensidade, dada em 2.34 foi obtida por Debye[14] e Porod[11]

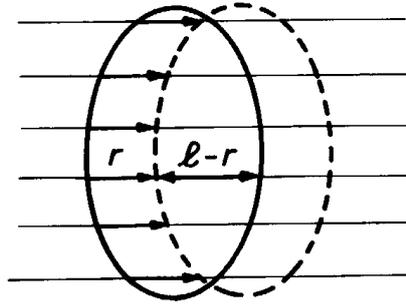


Figura 2.7: Representação de uma partícula e seu “fantasma”, formado pelo deslocamento. Figura retirada do livro *Small Angle X-ray Scattering* de Glatter e Kratky[1].

independentemente. Esta relação é bem representada no gráfico de Porod, mostrado na figura 2.8 dado por uma curva de $q^4 I(q) \times q$ do qual é possível obter informações de volume excluído V e deste com a superfície total da partícula S .

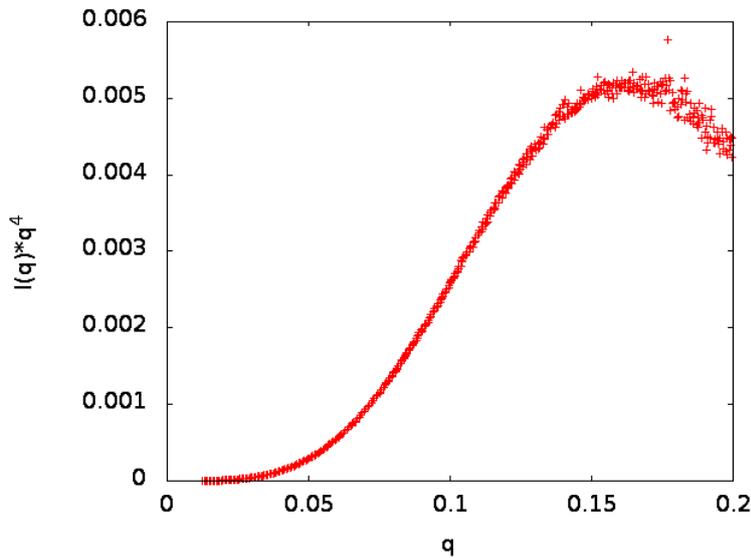


Figura 2.8: Gráfico de Porod para a lisozima. Usado para o cálculo do volume excluído da proteína e da superfície total a partir das relações 2.34 e 2.35. Gráfico plotado a partir de dados extraídos do BIOISIS[13]

$$I_1(q) \rightarrow (\Delta\rho)^2 \frac{2\pi}{q^4} \cdot S, \quad (2.34)$$

Considerando que as medidas experimentais exigem valores absolutos de intensidade, o uso da “invariante” se faz necessária de forma que relação entre S e V pode ser calculada por 2.35.

$$\frac{S}{V} = \frac{\pi}{Q} \lim_{q \rightarrow \infty} I(q)q^4. \quad (2.35)$$

Uma outra análise que pode ser feita é o chamada *gráfico de Kratky*, sendo importante para a análise de estados conformacionais da molécula. Ele é dado por um gráfico de $q^2 I(q) \times q$ é utilizada para a avaliação da flexibilidade de macromoléculas e do estado de enovelamento das proteínas. A figura 2.9 ilustra as diferentes formas que o gráfico pode assumir dependendo das condições de enovelamento da proteína.

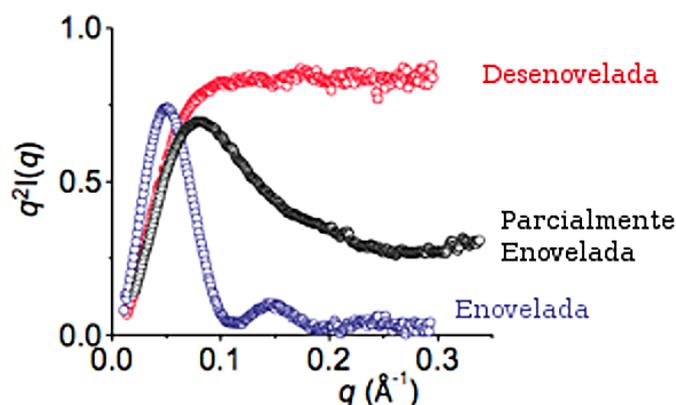


Figura 2.9: Diferentes curvas de espalhamento analisadas pelo gráfico de Kratky com diferentes estados de enovelamento da proteína. Figura retirada do site BIOISIS[13].

2.1.9 Análise computacional dos dados de SAXS.

O espalhamento de raios-X à baixo ângulo permite a caracterização estrutural de proteínas e a avaliação de mudanças conformacionais de proteínas em solução. A técnica possibilita a construção de um modelo tridimensional

por meio da curva de espalhamento. Porém, esta construção é dificultada pelo problema de fase discutido na seção 2.1.4 e seguintes.

O problema da fase provocou o meio científico ao desenvolvimento de modelos teóricos, obtidos à partir de formas geométricas simples, para o cálculo do envelope. Posteriormente, desenvolvem-se métodos para calcular a intensidade destes modelos e comparar com resultados experimentais.

Em 1970, Stuhrmann[2] usou esféricos harmônicos e polinômios de Laguerre para descrever matematicamente o envelope angular formado por uma proteína, a partir da curva de espalhamento. O *software* Crysol utiliza esta abordagem para fazer o processo inverso, ou seja, calcular uma curva de espalhamento teórica a partir do envelope. O sistema está suficientemente diluído para que seja considerada apenas uma partícula espalhadora. O conjunto está esquematizado na figura 2.10 e consiste em uma partícula com densidade de espalhamento $\rho_a(r)$ envolvida por uma camada de hidratação de espessura Δ e densidade ρ_b em um solvente de densidade ρ_0 .

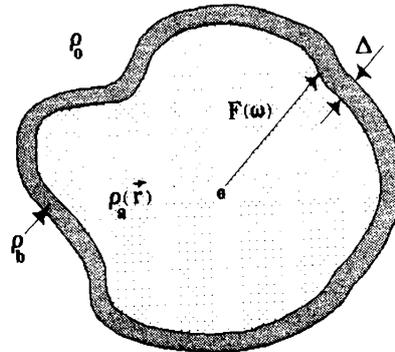


Figura 2.10: Macromolécula envolta por uma camada de hidratação de espessura Δ . Figura retirada do artigo de Svergun, Barberato e Koch[15].

Com esta conformação a intensidade de SAXS pode ser calculada pela equação 2.36, em que $\delta\rho = \rho_b - \rho_0$ e $A_a(\vec{q})$, $A_c(\vec{q})$ e $A_b(\vec{q})$ são a amplitude de espalhamento da partícula no vácuo, do volume excluído e dos átomos de borda, respectivamente.

$$I(q) = \langle |A_\alpha(q) - \rho_0 A_c(q) + \delta\rho(q)|^2 \rangle_\Omega \quad (2.36)$$

Ao ser fornecido um perfil de espalhamento como arquivo de entrada calcula-se um valor de discrepância (χ^2) pela equação 2.37

$$\chi^2(r_0, \delta\rho) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_{exp}(q_i) - cI(q_i, r_0, \delta\rho)}{\sigma(q_i)} \right]^2, \quad (2.37)$$

onde N_p e $\sigma(q_i)$ são a quantidade de pontos coletados e o erros da intensidade experimental $I_{exp}(q_i)$ de cada ponto e c é um fator de escala entre os dados de alta resolução e os dados experimentais.

Posteriormente, DAMMIN[3] e DAMMIF[16] e GASBOR[4] foram desenvolvidos para modelar os envelopes e analisar cavidades da moléculas. Estes programas utilizam pequenas esferas chamadas *dummy* gerando modelagens DAM². Estas esfera são utilizadas para preencher uma esfera de volume maior e podem ou não ser diferentes entre si, dependendo do sistema estudado. DAMMIN e DAMMIF usam *simulated annealing* para deformar a "esfera maior" e minimizar a função χ^2 na equação 2.38 para a análise da superfície da molécula. O programa GASBOR gera uma cadeia "proteica" formada por átomos *dummy*, que passa pelo processo de *simulated annealing* M vezes para minimizar χ^2 , possibilitando estudos sobre a camada de hidratação e cavidades da macromolécula.

$$\chi^2 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{N(i)} \left(\frac{I_{exp}^{(i)}(q_j) - I^{(i)}(q_j)}{\sigma(q_j)} \right)^2 \quad (2.38)$$

Na mesma época, Chacón et al.[17, 18] desenvolveram um *software* combinando algoritmos genéticos e simulações de Monte Carlo, aplicado à um modelo de átomos *dummy* para comparar as curvas geradas por estes sistemas com as curvas experimentais.

²Do inglês *dummy atoms model*

Mais recentemente, Schneidman-Duhovny, Hammel e Sali[19] utilizaram uma forma mais direta para o cálculo da intensidade baseado na aplicação da fórmula de Debye[20] para a intensidade de espalhamento apresentada na equação 2.39.

$$I(q) = \sum_{i=1}^N \sum_{j=1}^N f_i(q) f_j(q) \frac{\sin(qd_{ij})}{qd_{ij}}, \quad (2.39)$$

onde d_{ij} é a distância entre os átomos i e j e o fator de forma, ou fator de espalhamento atômico, f_i já considera a camada de hidratação e o solvente deslocado para o cálculo.

$$f_i(q) = f_v(q) - c_1 f_s(q) + c_2 s_i f_w(q), \quad (2.40)$$

onde $f_v(q)$ e $f_s(q)$ são o fator de espalhamento do átomo no vácuo e do átomo *dummy* que representa o solvente deslocado, respectivamente, s_i é a fração de superfície acessível ao solvente do átomo i , f_w é o fator de forma da água, c_1 é usado para ajustes no volume excluído total e c_2 é usado para ajustes na densidade de espalhamento da camada de hidratação, geralmente usa-se $c_1 = 1$ e $c_2 = 0$ como valores padrão. Quando um perfil de SAXS é fornecido ao software o cálculo de χ^2 é feito usando a equação 2.37.

O cálculo de um envelope diretamente da curva de espalhamento é um problema mal-posto que não possui uma solução única[21], entretanto esta abordagem permite o uso de vínculos impostos pela estrutura da molécula, proporcionando, especialmente pela geometria da ligações de carbonos e a formação de estruturas secundárias, uma redução do problema.

Capítulo 3

Resultados preliminares.

O projeto tem por objetivo desenvolver uma ferramenta, chamada Pivot-SAXS, capaz de prever a conformação de uma proteína em solução baseada na curva de espalhamento de SAXS e movimentos de pivot nas regiões de *coil* da estrutura. O Pivot-SAXS utiliza um potencial baseado na diferença entre a curva de espalhamento teórico da conformação atual e a curva fornecida pelo usuário, além de considerar os efeitos estéricos causados pelo tamanho dos átomos. Para isso foram necessários o desenvolvimento e validação das ferramentas computacionais SAXS-factor, para o cálculo da curva teórica, e Ramatrix, a qual efetua os movimentos de pivot e calcula a energia devido ao impedimento estérico. Este capítulo aborda o processo de validação destas ferramentas-base.

3.1 Desenvolvimento do SAXS-factor.

O programa explicado à seguir segue a mesma abordagem computacional do programa FoXS, utilizando a equação 2.39 para o cálculo da intensidade de espalhamento e 2.37 como parâmetro de ajuste. Em nossa implementação inicial, os fatores de forma utilizados na equação 2.40 são referentes, apenas, ao átomo no vácuo, não levando em consideração a camada

de hidratação ou o efeito do solvente. Isso faz com que os programas FoXS e Crysol, que utilizamos para comparação, sejam mais precisos. Todavia, isso também nos permite um maior controle das variáveis e de seus efeitos no problema.

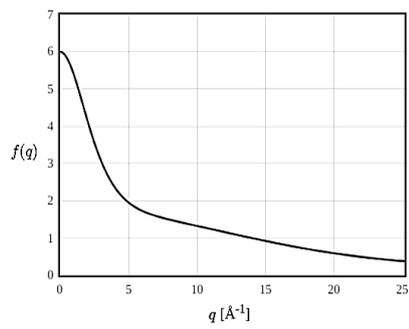
O programa, nomeado carinhosamente de *SAXS-factor*, foi desenvolvido em FORTRAN 77 e tem por objetivo ter uma elaboração simples e eficiente, com um algoritmo claro e de apresentação didática. SAXS-factor utiliza dois arquivos de entrada: um arquivo PDB retirado do bancos de dados de proteínas¹, conhecido como PDB[22] - *do inglês Protein Data Bank* - e um perfil de intensidade experimental. Neste trabalho não houve coleta de dados pelo grupo de pesquisa, sendo os dados utilizados retirados do banco de dados² BIOISIS[13]. Do arquivo PDB são obtidas as posições atômicas de cada elemento as quais são usadas para calcular a distância usada na equação 2.39 e os elementos químicos para o cálculo do fator de forma. Do perfil experimental obtêm-se os valores de $I(q)$, q e σ , onde q é substituído na equação 2.39 para calcular os valores de intensidade teórica e, posteriormente, os valores são usados para comparação e obtenção do valor de χ^2 , obtido na equação 2.37. O algoritmo do SAXS-factor é melhor demonstrado no fluxograma da figura 3.2, sendo que as linhas cheias representam a passagem de informação e as linhas pontilhadas a sequência do programa.

Para o cálculo dos fatores de espalhamento foi utilizada a equação 3.1, a qual possui uma dependência direta com o o vetor momento de espalhamento como mostra a figura 3.1. a_i , b_i e c são chamados coeficientes de Cromer-Mann e são diferentes para cada elemento químico e apresentados na tabela 3.1.

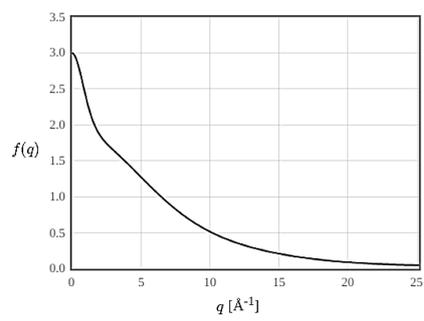
$$f(\vec{q}) = \sum_{i=1}^4 a_i \exp\left(-b_i \left(\frac{q}{4\pi}\right)^2\right) + c, \quad (3.1)$$

¹<http://www.rcsb.org/pdb/>

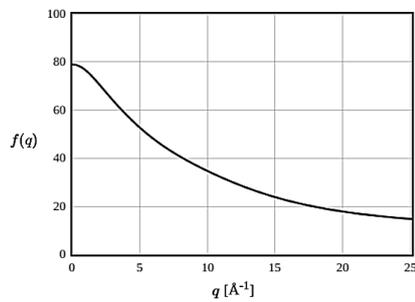
²<http://www.bioisis.net/>



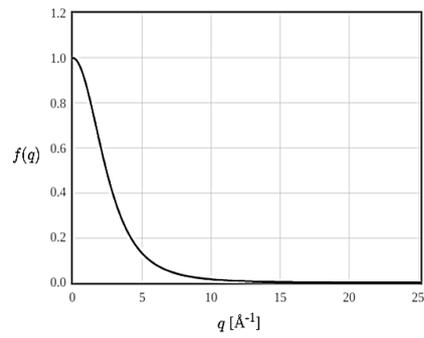
(a) Carbono



(b) Lítio



(c) Ouro



(d) Hidrogênio

Figura 3.1: Relação entre $f(q)$ e q para diferentes elementos químicos

Tabela 3.1: Valores de a_i , b_i e c para diversos elementos químicos.

| Elemento | a_1 | b_1 | a_2 | b_2 | a_3 | b_3 | a_4 | b_4 | c |
|----------|----------|---------|----------|---------|----------|---------|----------|---------|----------|
| H | 0.489918 | 20.6593 | 0.262003 | 7.74039 | 0.196767 | 49.5519 | 0.049879 | 2.20159 | 0.001305 |
| He | 0.8734 | 9.1037 | 0.6309 | 3.3568 | 0.3112 | 22.9276 | 0.178 | 0.9821 | 0.0064 |
| C | 2.31 | 20.8439 | 1.02 | 10.2075 | 1.5886 | 0.5687 | 0.865 | 51.6512 | 0.2156 |
| N | 12.2126 | 0.0057 | 3.1322 | 9.8933 | 2.0125 | 28.9975 | 1.1663 | 0.5826 | -11.529 |
| O | 3.0485 | 13.2771 | 2.2868 | 5.7011 | 1.5463 | 0.3239 | 0.867 | 32.9089 | 0.2508 |
| Ne | 3.9553 | 8.4042 | 3.1125 | 3.4262 | 1.4546 | 0.2306 | 1.1251 | 21.7184 | 0.3515 |
| Na | 4.7626 | 3.285 | 3.1736 | 8.8422 | 1.2674 | 0.3136 | 1.1128 | 129.424 | 0.676 |
| Mg | 5.4204 | 2.8275 | 2.1735 | 79.2611 | 1.2269 | 0.3808 | 2.3073 | 7.1937 | 0.8584 |
| P | 6.4345 | 1.9067 | 4.1791 | 27.157 | 1.78 | 0.526 | 1.4908 | 68.1645 | 1.1149 |
| S | 6.9053 | 1.4679 | 5.2034 | 22.2151 | 1.4379 | 0.2536 | 1.5863 | 56.172 | 0.8669 |
| Ca | 8.6266 | 10.4421 | 7.3873 | 0.6599 | 1.5899 | 85.7484 | 1.0211 | 178.437 | 1.3751 |
| Fe | 11.7695 | 4.7611 | 7.3573 | 0.3072 | 3.5222 | 15.3535 | 2.3045 | 76.8805 | 1.0369 |
| Zn | 14.0743 | 3.2655 | 7.0318 | 0.2333 | 5.1652 | 10.3163 | 2.41 | 58.7097 | 1.3041 |
| Se | 17.0006 | 2.4098 | 5.8196 | 0.2726 | 3.9731 | 15.2372 | 4.3543 | 43.8163 | 2.8409 |
| Au | 16.8819 | 0.4611 | 18.5913 | 8.6216 | 25.5582 | 1.4826 | 5.86 | 36.3956 | 12.0658 |

Fonte: *International Tables of Crystallography*. [23]

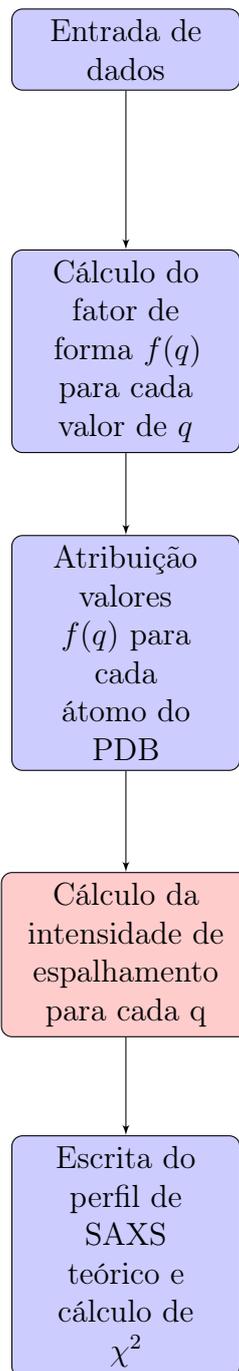


Figura 3.2: Fluxograma do algoritmo utilizado em SAXS-factor. O programa recebe os dados e calcula o fator de forma dos átomos da tabela 3.1 para cada valor de q . Os átomos do pdb recebem seus fatores de forma e a intensidade de espalhamento de toda a proteína é calculada para q . Por fim, obtêm-se o valor de χ^2 e um arquivof com o perfil de espalhamento teórico.

3.2 Calibração do SAXS-factor.

Durante o desenvolvimento do SAXS-factor o perfil de espalhamento modelo da Lisozima, retirado do banco de dados BIOISIS[24], foi utilizada como padrão de comparação. Posteriormente, os perfis de 28 pares de base de DNA e da Xilanase, também retiradas do BIOISIS, foram usadas para testes de semelhança.

Outra forma de validação do SAXS-factor foi a comparação com as versões *online* dos *softwares* similares, Crysol³ e FoXS⁴. Para esta análise, além das 3 moléculas citadas anteriormente, também foram analisadas, a vilina, a *engrailed homeodomain* de *Drosophila melanogaster*[25]⁵ e a a quimotripsina inibidora 2[26]⁶. Para as últimas não é calculado o valor de χ^2 pois estas proteínas são retiradas diretamente do PDB[22] e não possuem perfil de espalhamento experimental.

Em todos os casos as curvas são dadas em preto, vermelho, azul e verde para o perfil experimental e os perfis teóricos obtidos por SAXS-factor, FoXS e Crysol, respectivamente.

3.2.1 Moléculas modelo do banco de dados⁷.

Abaixo são apresentados os resultados gráficos calculados para 3 proteínas selecionadas para a calibração do modelo e as comparações entre os perfis experimentais e os teóricos obtidos com o SAXS-factor, FoXS e Crysol. Também são apresentados os valores de χ^2 entre a curva experimental e a calculada pelo SAXS-factor. Estas moléculas foram escolhidas para calibração do SAXS-factor e de análise dos efeitos das simplificações feitas no cálculo dos fatores de forma.

³Encontrada em <http://www.embl-hamburg.de/biosaxs/atsas-online/>

⁴Encontrada em <http://modbase.compbio.ucsf.edu/foxs/>

⁵PDB ID: 1ENH

⁶PDB ID: 1YPC

⁷BIOISIS[24]

A Lisozima⁸ foi utilizada nos primeiros testes do programa por ser uma proteína pequena e globular e, portanto, teria uma análise rápida e simples, além de já ter sido amplamente estudada e suas condições conformacionais são bem entendidas. Ainda é classificada como α/β envolvendo um bom grau de complexidade para estrutura secundária e organização. A figura 3.3 apresenta a comparação gráfica entre a curva experimental teórica e a curva teórica do SAXS-factor, o valor de $\chi^2 = 324.1$ também calculado pelo programa e uma figura representação da molécula.

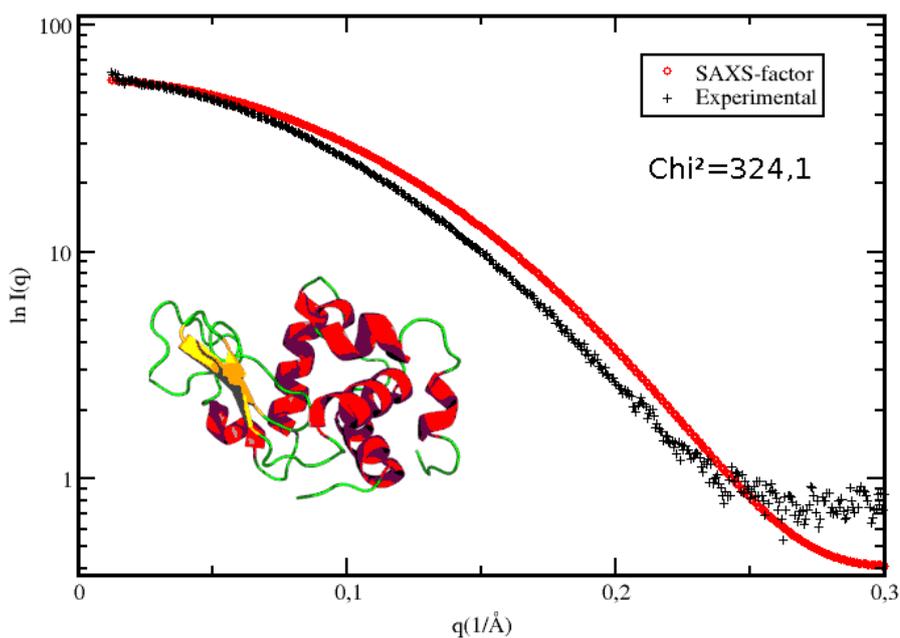
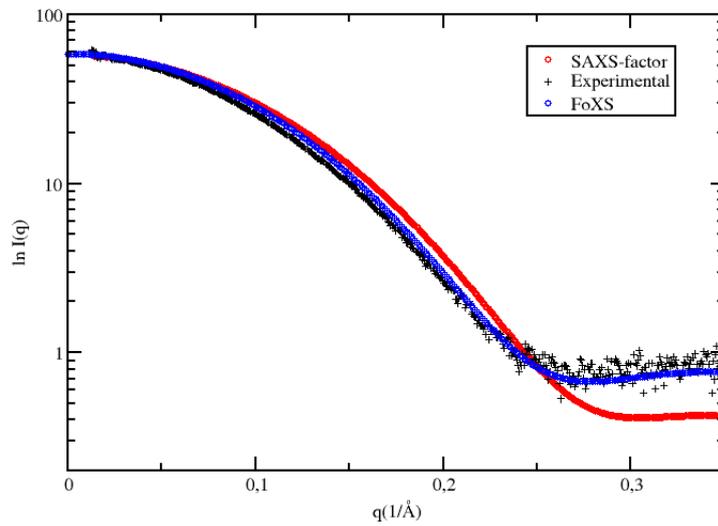


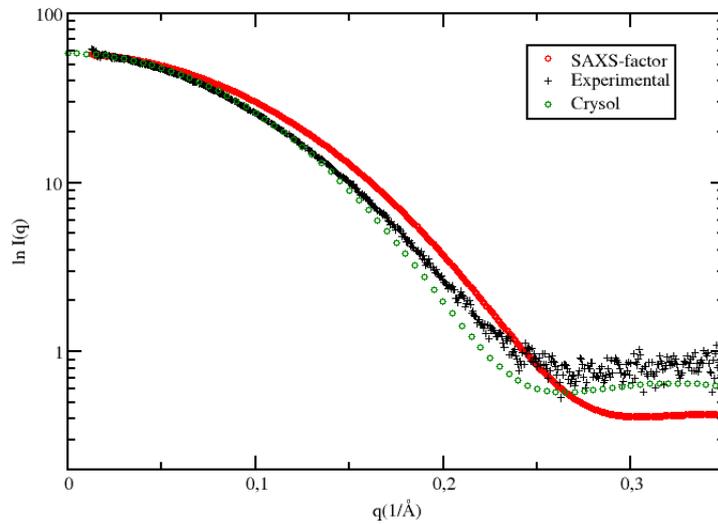
Figura 3.3: Perfil de espalhamento da Lisozima experimental, curva em preto, em comparação com o teórico obtido pelo SAXS-factor, curva em vermelho. O valor de χ^2 calculado para a Lisozima foi $\chi^2 = 324.1$. Uma representação estrutural também é apresentada ao lado da curva.

A análise comparativa entre o SAXS-factor e os *softwares* similares foi importante para verificar o como a abordagem feita neste trabalho afeta

⁸BIOISIS ID:LYSOZP



(a)



(b)

Figura 3.4: Análise comparativa dos resultados do SAXS-factor para a Lisozima. Os resultados obtidos pelo SAXS-factor, com curva vermelha, e o perfil experimental, com curva preta são comparados com os programas (a) FoXS, com curva azul, e (b) Crysol, curva verde.

a qualidade do resultado. A figura 3.4 mostra os resultados dados pelos *softwares online* FoXS e Crysol frente aos gerados no SAXS-factor e aos dados experimentais. Com estes gráficos é possível analisar os efeitos da camada de hidratação sobre o perfil de espalhamento, dado que os fatores de forma utilizados no programa SAXS-factor empregam apenas o átomo no vácuo.

Tanto o DNA quanto a Xilanase também foram escolhidas por terem dados experimentais disponíveis em bancos de dados[24], para provar a versatilidade do programa.

Utilizar um seguimento de DNA para comparação é interessante devido a possibilidade de utilização do software pra este tipo de molécula, permitindo estudos futuros de interação DNA-proteína ou RNA-proteína. Neste caso, 28 pares de base⁹ foram usados para o cálculo da intensidade teórica e comparados com a curva de intensidade experimental na figura 3.5, com a curva de intensidade do programa FoXS na figura 3.6a e com a curva obtida pelo Crysol na figura 3.6b.

O valor de χ^2 calculado para esta molécula foi de $\chi^2 = 13,3$. Pela equação 2.37 pode perceber que o valor de χ^2 é inversamente proporcional ao erro experimental. O erro para o DNA é cerca de 2 ordens de grandeza maior que o da lisozima. Esta é razão para lisozima possuir um valor de χ^2 maior enquanto, graficamente, parece ter um ajuste melhor que o do DNA.

Escolhida devido a sua importância industrial, a enzima xilanase¹⁰ possui uma grande relevância nos processos de degradação vegetal utilizado nos processos industriais de fábricas de papel e especialmente em pesquisas com biodiesel.

Formada, basicamente, por folhas β , a xilanase tem uma grande área com acesso ao solvente devido ao arranjo das cadeias laterais. A implementação de uma camada de hidratação deve melhorar o ajuste no final da curva

⁹BIOISIS ID:28BPDD

¹⁰BIOISIS ID:1XYNTP

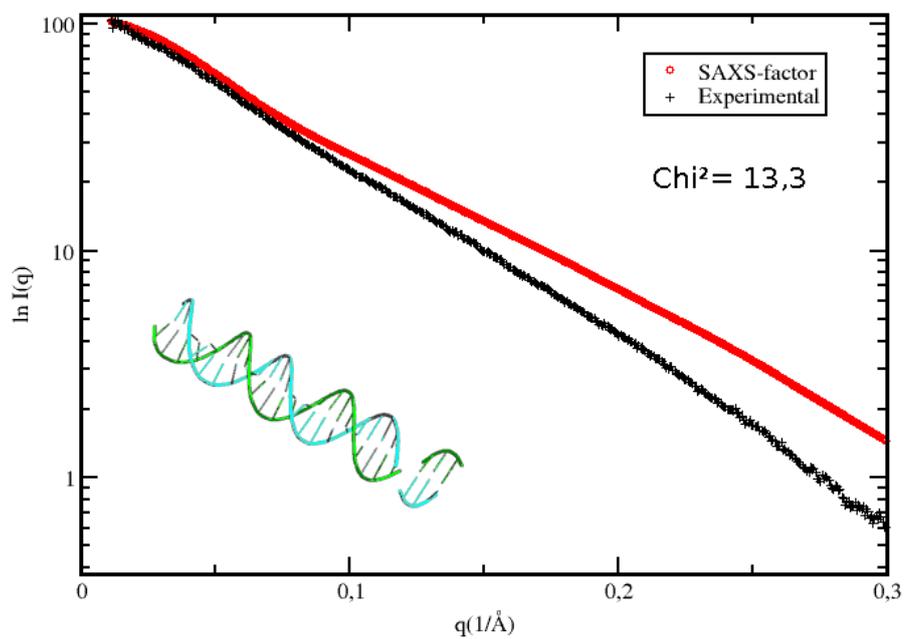
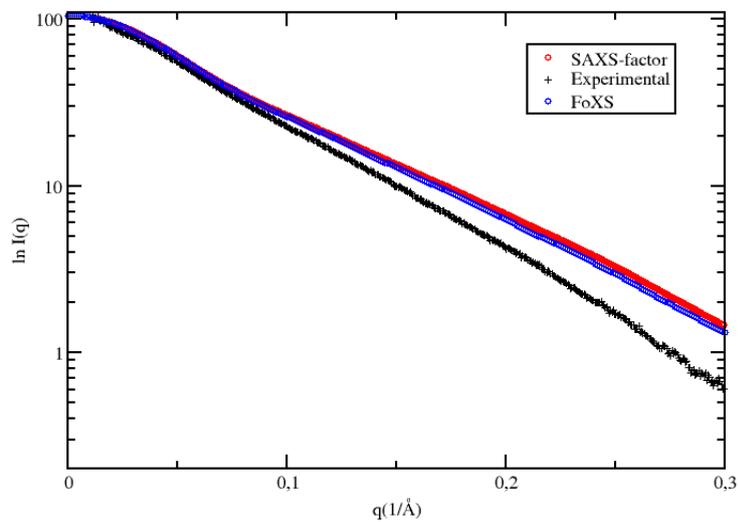
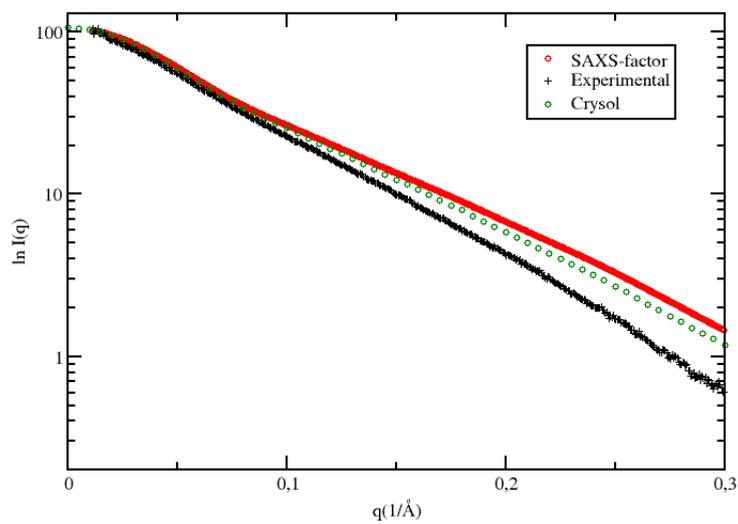


Figura 3.5: Perfil de espalhamento de 28 pares de base de DNA experimental, curva em preto, em comparação com o teórico obtido pelo SAXS-factor, curva em vermelho. O valor de χ^2 calculado para a o DNA foi $\chi^2 = 13.3$. Uma representação estrutural também é apresentada ao lado da curva.

de espalhamento.



(a)



(b)

Figura 3.6: Análise comparativa dos resultados do SAXS-factor para 28 pares de bases de DNA. Os resultados obtidos pelo SAXS-factor, com curva vermelha, e o perfil experimental, com curva preta são comparados com os programas (a) FoXS, com curva azul, e (b) Crysol, curva verde.

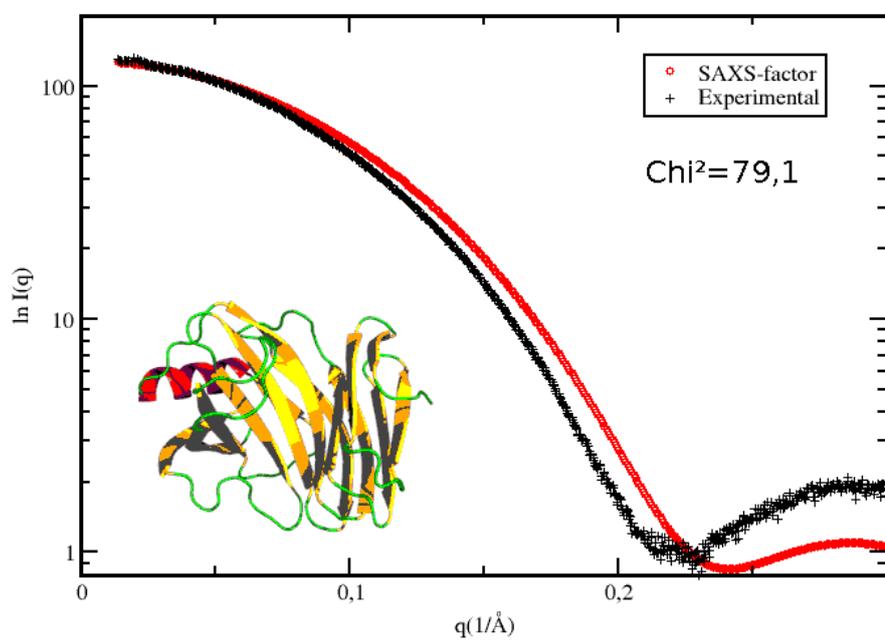
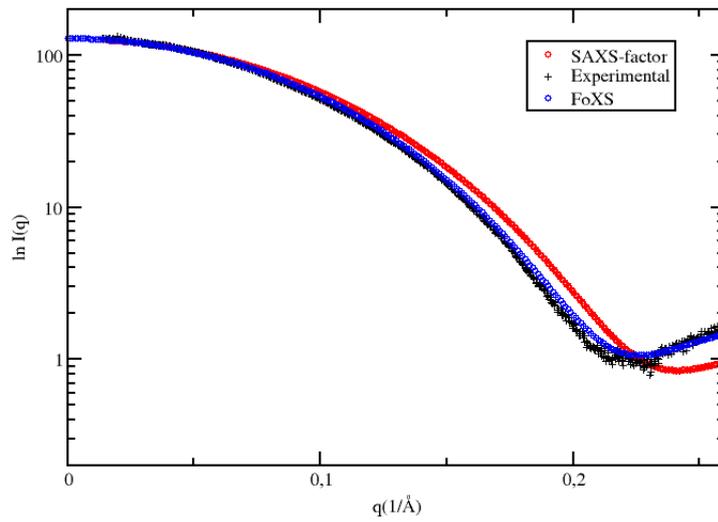
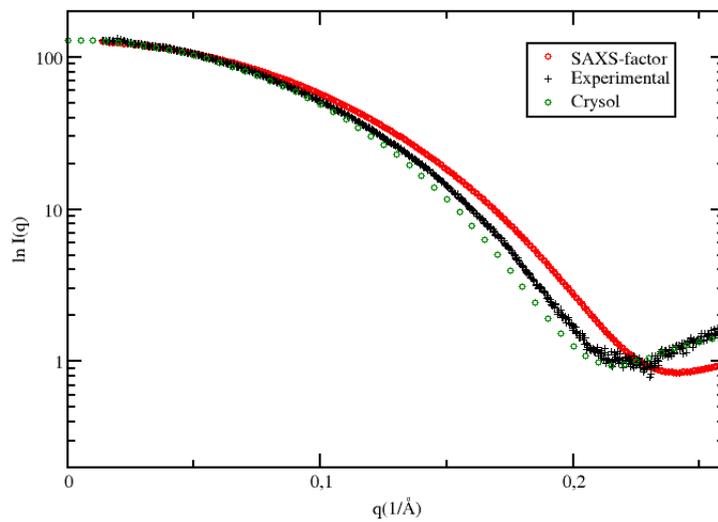


Figura 3.7: Perfil de espalhamento da Xilanase experimental, curva em preto, em comparação com o teórico obtido pelo SAXS-factor, curva em vermelho. O valor de χ^2 calculado para a Xilanase foi $\chi^2 = 79.1$. Uma representação estrutural também é apresentada ao lado da curva.



(a)



(b)

Figura 3.8: Análise comparativa dos resultados do SAXS-factor para a Xilanasase. Os resultados obtidos pelo SAXS-factor, com curva vermelha, e o perfil experimental, com curva preta são comparados com os programas (a) FoXS, com curva azul, e (b) Crysol, curva verde.

3.2.2 Moléculas retiradas do PDB[22].

As estruturas seguintes são comparadas apenas com os perfis de espalhamento teóricos calculados por FoXS e Crysol, pois não há disponibilidade dos dados experimentais. Tais estruturas são utilizadas no capítulo 3 para validação do programa Pivot-SAXS e foram escolhidas por seus tamanhos e estruturas secundárias, maiores detalhes serão discutidos posteriormente, no capítulo 3. As proteínas vilina¹¹, 1ENH[25] e 1YPC[26]¹² tiveram seus perfis de espalhamento teórico calculados pelos três *softwares* e comparados nos gráficos 3.9, 3.10 e 3.11.

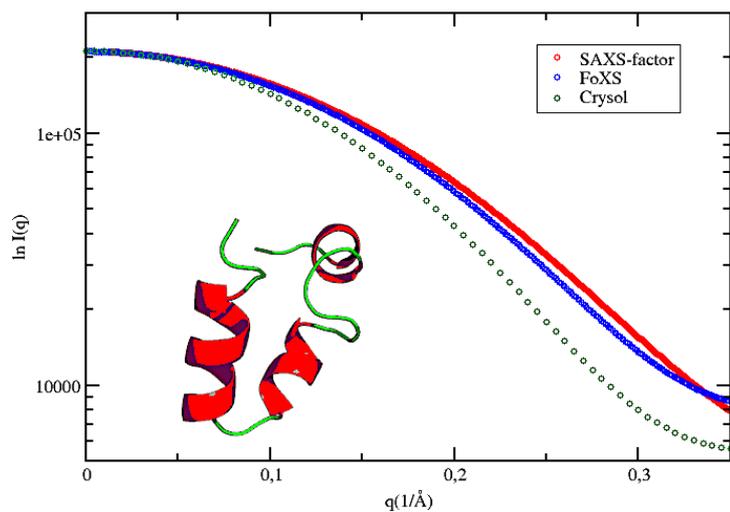


Figura 3.9: Análise comparativa dos resultados do SAXS-factor para a Vilina. Os resultados obtidos pelo SAXS-factor, com curva vermelha, são comparados com os programas FoXS, com curva azul, e Crysol, curva verde.

¹¹PDB ID:1VII

¹²As duas proteínas são referenciadas pelas suas próprias PDB ID.

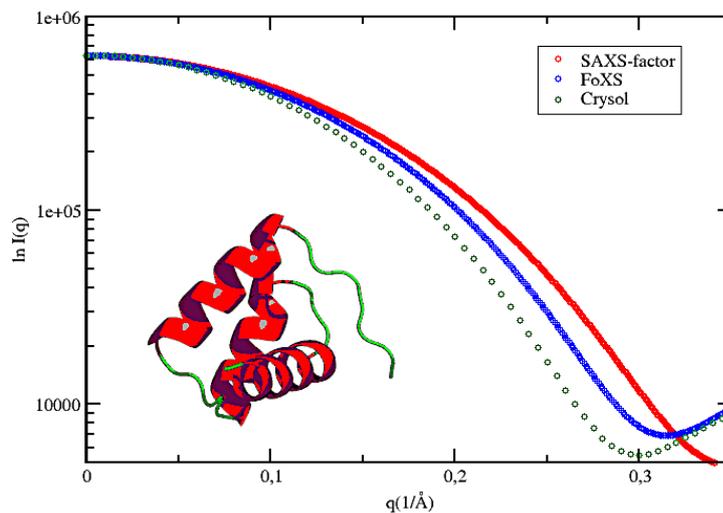


Figura 3.10: Análise comparativa dos resultados do SAXS-factor para a 1ENH. Os resultados obtidos pelo SAXS-factor, com curva vermelha, são comparados com os programas FoXS, com curva azul, e Crysol, curva verde.

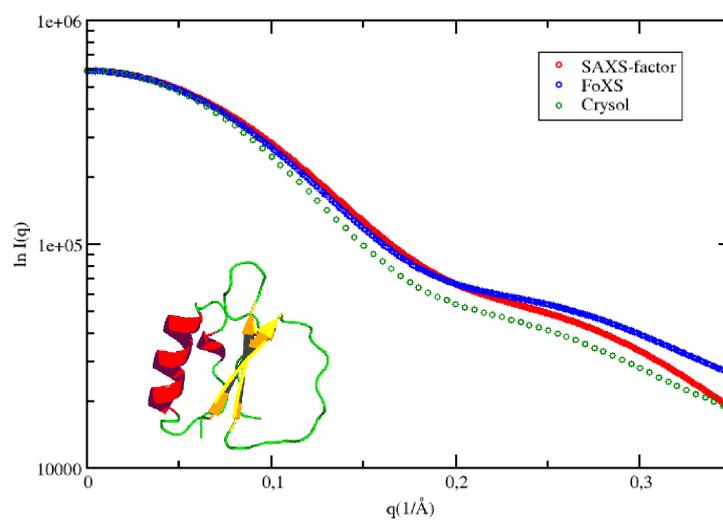


Figura 3.11: Análise comparativa dos resultados do SAXS-factor para a 1YPC. Os resultados obtidos pelo SAXS-factor, com curva vermelha, são comparados com os programas FoXS, com curva azul, e crysol, curva verde.

3.3 Estudo dos ângulos diedrais de uma proteína.

O grande interesse no estudo das proteínas e de suas estruturas se deve ao seu papel vital na natureza. Esta seção tem por objetivo abordar o conceito do uso de restrições estéricas na construção de proteínas. Reproduzir o gráfico de Ramachandran[27] permite avaliar quais conformações a proteína pode assumir sem sofrer efeitos estéricos.

3.3.1 Emprego de propriedades geométricas.

Os aminoácidos são bastante diferentes entre si mas possuem uma mesma estrutura básica, também chamada de *backbone*, formada por um carbono alfa (C_α) ligado à um radical amina e um radical carboxila da forma que mostra a figura 3.12. Cada um dos 20 aminoácidos se diferenciam por uma cadeia lateral (R) todas apresentadas na figura 3.13.

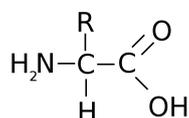


Figura 3.12: Representação do *backbone* de um aminoácido

Existem alguns ângulos de interesse no *backbone* dos aminoácidos, demonstrados na figura 3.14, entre eles os ângulos de ligação, os quais tendem a permanecer constantes, e os ângulos diedrais, que podem assumir diferentes valores e combinações, afetando a conformação das moléculas.

Os ângulos ϕ , ψ e ω são os ângulos diedrais. ω se refere à rotação em torno da ligação C_i-N_{i+1} e, diferentemente de ϕ e ψ , não assume uma gama alta de valores se restringindo às conformações *cis*, com $\omega = 0^\circ$, e *trans*, com $\omega = 180^\circ$ [28]. Portanto, os ângulos de interesse para o trabalho serão ϕ e ψ por serem diretamente responsáveis pelas mudanças conformacionais afetando a cadeia lateral e, quando se considera uma proteína, toda a estrutura. ϕ se

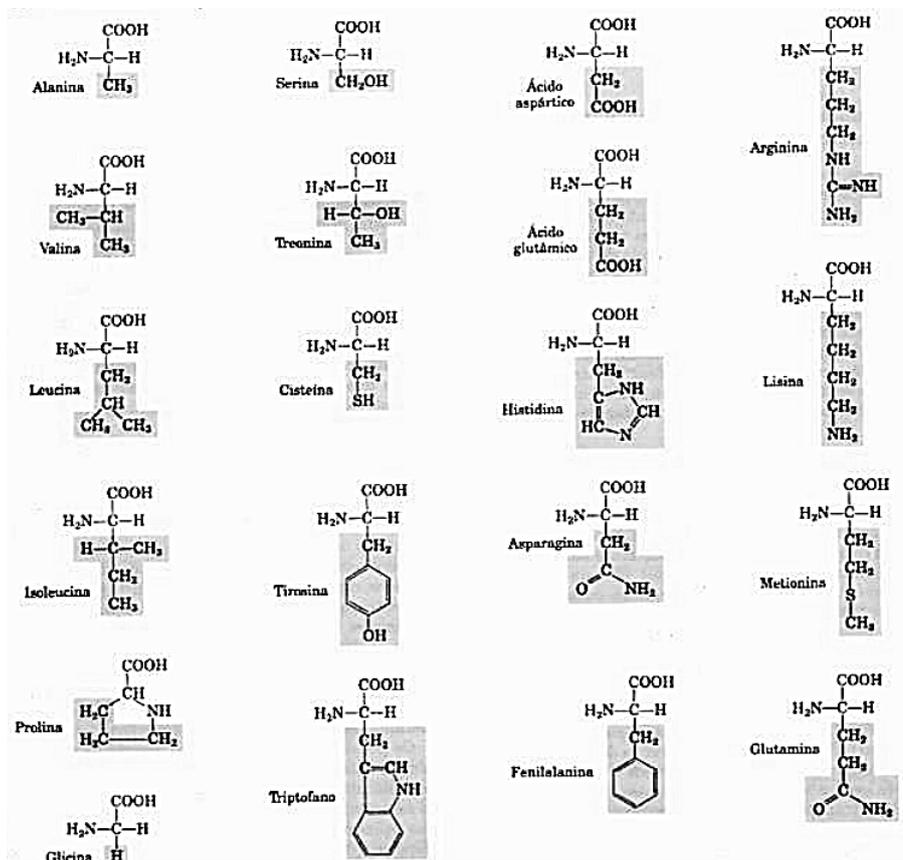


Figura 3.13: Representação das cadeias laterais de todos os aminoácidos essenciais.

refere ao ângulo em torno da ligação $N-C_{\alpha}$ assumindo o valor de 0° quando C_i e C_{i+1} estão *trans* com relação mesma enquanto ψ se refere à ligação $C_{\alpha}-C$ com valor 0° quando N_i e N_{i+1} estão em posição de *trans* em relação à ligação $C_{\alpha}-C$ [27], tais valores aumentam no sentido horários. A figura 3.15 mostra dois aminoácidos ligados e totalmente estendidos, com ambos os ângulos em posição de 0° .

Ao descrever peptídeos utilizando ϕ e ψ nota-se que não são todas as combinações que são fisicamente possíveis ou prováveis[27]. Existem impedimentos estéricos, ou seja, em certas combinações alguns átomos quase se sobrepõem, o gráfico da distribuição de ϕ e ψ é chamado Ramachandran *plot* e um exemplo deste gráfico está apresentado na fig 3.16. Ao considerar

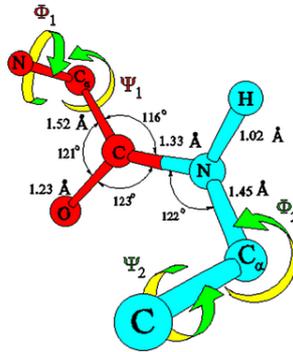


Figura 3.14: Representação dos ângulos em um aminoácido.

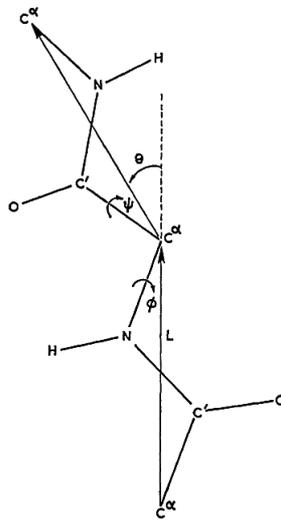


Figura 3.15: Conformação totalmente plana de uma ligação peptídica. Os ângulos ϕ e ψ estão em posição de zero.

que a distância interatômica mínima é dada pela soma dos raios de *van der Waals* de cada um dos átomos [29] é possível fazer uma análise da energia de repulsão de Lennard-Jones para cada combinação, usando a equação 3.2 considerando σ_{ij} a distância mínima e r_{ij} a distância real entre os átomos.

$$E_{lj} = \sum_i \sum_{j=i+3} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \quad (3.2)$$

Tal abordagem deve ser feita sobre toda a estrutura, desde que os átomos sejam ao menos terceiros vizinhos para evitar o efeito de sobreposição

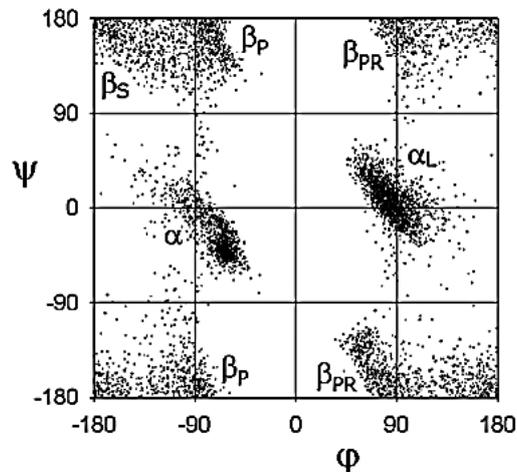


Figura 3.16: Ramachandran *plot* observado para a Glicina. Figura retirada de Ho e Brasseur[30]

de nuvem eletrônica devido à ligação covalente.

3.3.2 Ângulos de Euler.

Ângulos de Euler são definidos como o ângulo dado pela rotação em torno de um eixo[31], a figura 3.17 mostra os ângulos de euler em torno do eixos cartesianos. Desta forma os ângulos ϕ , ψ e ω são considerados ângulos de euler em relação ao eixo formado pelas ligações $N-C_\alpha$, $C_\alpha-C$ e C_i-N_{i+1} , respectivamente.

Tais ângulos são utilizados para mudar os sistema de coordenadas sem alterar a ortonormalidade do sistema cartesiano, ou seja, alterando apenas as coordenadas e mantendo o tamanho das partículas representadas. Aplicar isso em uma proteína permite a rotação dos ângulos ϕ e ψ , que podem assumir diversos valores, sem afetar as distâncias interatômicas, o tamanho dos átomos ou as proporções intramoleculares.

A rotação em torno de um eixo gera um novo sistema de eixos cartesianos, representado na figura 3.18, e, portanto, conjunto de coordenadas para cada ponto. Para determinar estas novas coordenadas utiliza-se a equação

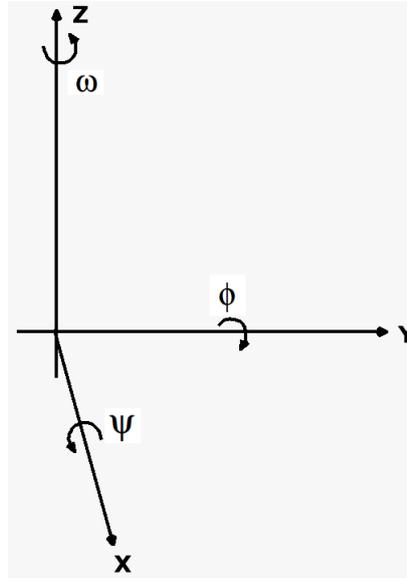


Figura 3.17: Representação dos ângulos de Euler em relação aos eixos cartesianos.

3.3, em que x, y e z são as coordenadas antigas, x', y' e z' são as novas posições e λ é uma matriz de rotação formada pelos cossenos dos ângulos formado entre o sistema de coordenadas antes e depois da rotação, dada pela equação 3.4.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (3.3)$$

$$\lambda = \begin{pmatrix} \cos(x', x) & \cos(x', y) & \cos(x', z) \\ \cos(y', x) & \cos(y', y) & \cos(y', z) \\ \cos(z', x) & \cos(z', y) & \cos(z', z) \end{pmatrix} \quad (3.4)$$

Os cossenos entre os dois sistemas de coordenadas são os próprios ângulos de Euler, assim a matriz de rotação λ pode ser simplificada para a rotação em torno de cada um dos eixos, sendo λ_x a matriz de rotação em torno do eixo x , λ_y em torno de y e λ_z de z , cada uma delas está apresentada na equação 3.5[31].

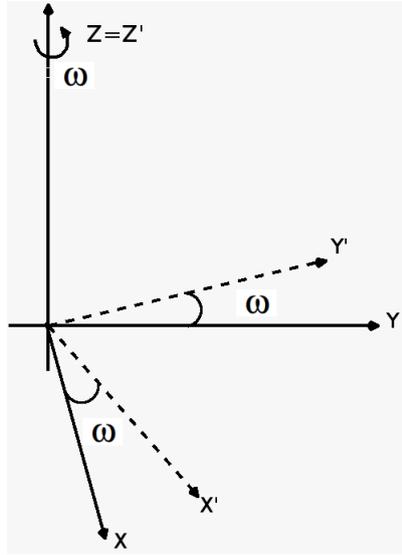


Figura 3.18: Representação do sistema de coordenadas 'linha' formado a partir da rotação de ω em torno de Z .

$$\lambda_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & \sin \psi \\ 0 & -\sin \psi & \cos \psi \end{pmatrix}$$

$$\lambda_y = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.5)$$

$$\lambda_z = \begin{pmatrix} \cos \omega & 0 & \sin \omega \\ 0 & 1 & 0 \\ -\sin \omega & 1 & \cos \omega \end{pmatrix}$$

Esta simplificação torna muito mais simples o cálculo da multiplicação de matrizes da equação 3.3.

3.3.3 Estrutura do programa Ramatrix.

Os conceitos de ângulos e matrizes de rotação dão a base para o desenvolvimento do program, daqui pra frente chamado Ramatrix, utiliza as matrizes de rotação 3.5 para rotacionar os ângulos ϕ e ψ . Para fazer uma análise de todas as combinações destes angulos utilizou-se destas rotações e do cálculo da energia de repulsão de Lennard-Jones pela equação 3.2 para determinar quais as conformações energeticamente mais favoráveis. Permitindo a construção de um potencial que limita a geometria à regiões permitidas.

Escrito em FORTRAN 77, o Ramatrix recebe um arquivo PDB sem hidrogênios, por facilidade de programação, para calcular a energia de Lennard-jones da conformação inicial, a partir de então iniciam-se as rotações em torno dos ângulos phi e psi, de forma que, para cada ângulo ψ , 360 novas conformações sejam geradas. Assim todas as 360×360 combinações de $\psi \times \psi$ são mapeadas e a energia de cada uma delas é calculada, permitindo que um espectro de Ramachandran seja plotado.

Para que os ângulos diedrais sejam utilizados como ângulos de Euler, a ligação em torno da qual ocorrerá a rotação deve coincidir com um dos eixos cartesianos, no caso, foi escolhido o eixo x . A proteína toda passa por uma translação e duas rotações a fim de ajustar esta posição para cada rotação. Inicialmente, é definido um vetor com o mesmo módulo, direção e sentido da ligação relativa à rotação ($N-C\alpha$ para psi e $C_\alpha-C$ para phi), a translação é feita para que o início do vetor esteja em $(0, 0, 0)$ e então uma rotação em torno do eixo y para zerar a coordenada z e outra em torno de z até zerar y . Um esquema simplificado do programa está na figura 3.19.

3.3.4 Validação do programa Ramatrix.

Os gráficos de *Ramachandran* da glicina e da alanina são os mais encontrados na literatura científica e funcionam como padrão. Portanto,

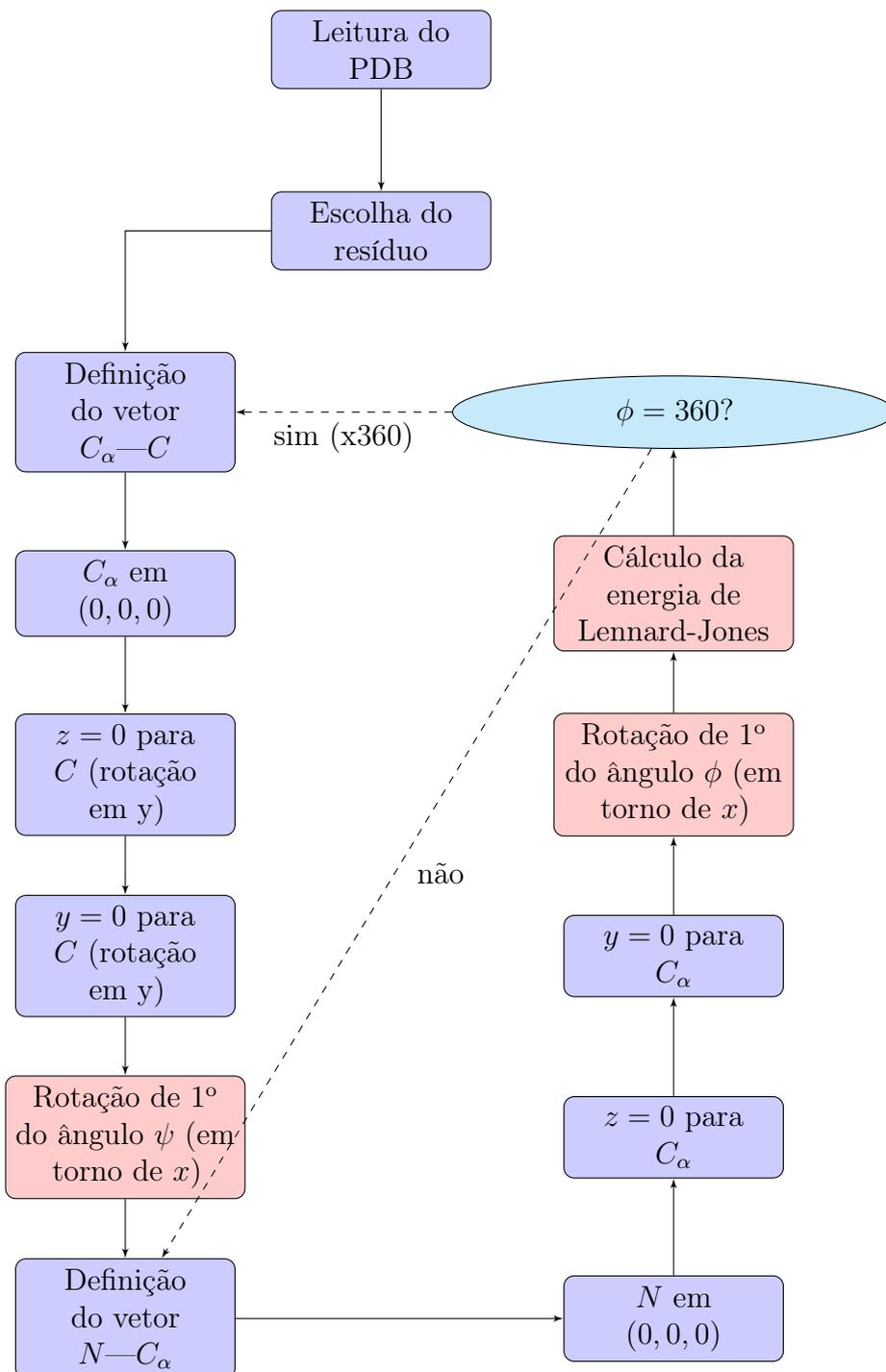


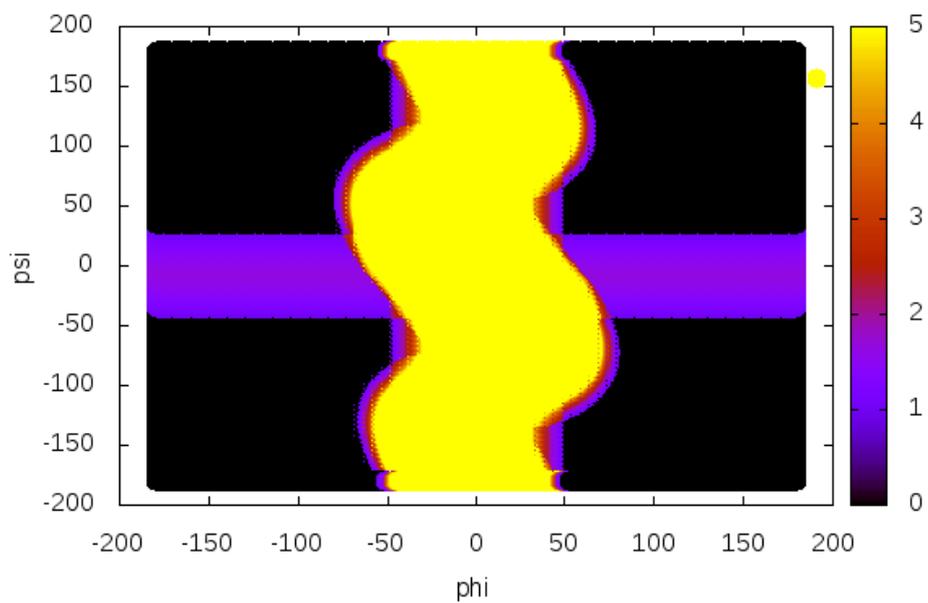
Figura 3.19: Fluxograma do algoritmo utilizado em Ramatrix.

estes foram os resíduos utilizados para a validação. Para isso foram gerados, computacionalmente através do programa *pymol*, 2 peptídeos de quatro

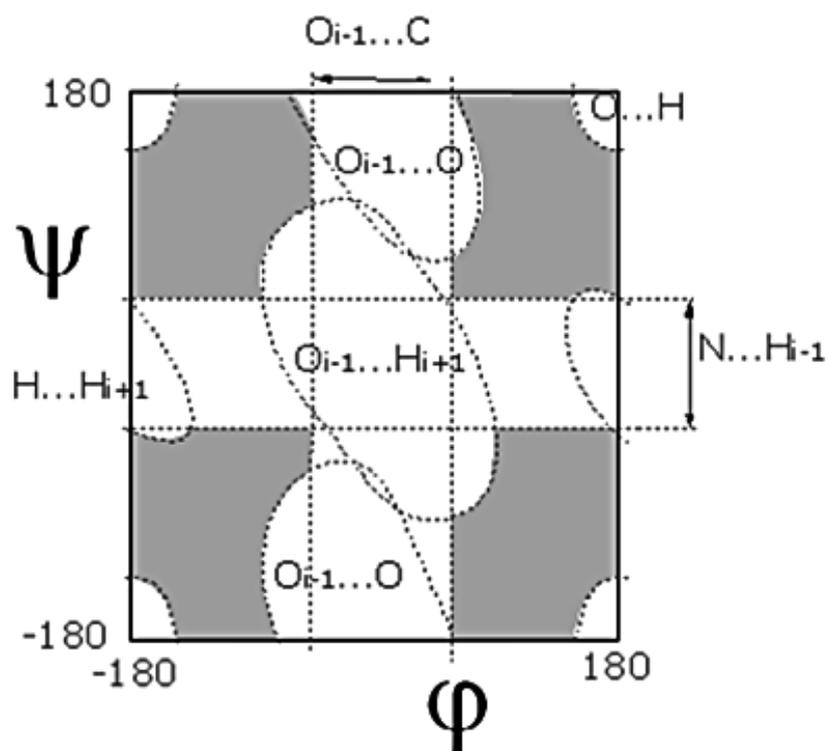
resíduos sem os hidrogênios: um formado apenas de glicina e outro de alanina. As rotações ocorreram em torno das ligações do resíduo dois, em ambos os casos.

O resultado para a glicina, apresentados na figura 3.20a, possuem o padrão simétrico característico devido à ausência de cadeia lateral. As regiões em preto possuem menor energia, ou seja, são mais permitidas, enquanto as regiões em amarelo são combinações de maior energia. Estes máximos são causados pela proximidade de átomos muito grandes que causam uma energia repulsiva muito alta. As análises mostram que os resultados estão coerentes com o mapa de impedimento estérico na figura 3.20b retirada do artigo de Ho e Brasseur[30] de 2005.

O gráfico de Ramachandran da alanina apresentado na figura 3.21a possui seus valores de máximo na região em amarelo. O gráfico apresenta o mesmo padrão da glicina adicionado de uma nova região proibida. Esta região existe devido a única diferença entre a estrutura da alanina e da glicina, a existência de um carbono na cadeia lateral. E esta região é a faixa de sobreposição entre o C_β e o oxigênio do resíduo anterior. A figura 3.21b mostra o mapa de impedimento estérico retirado do artigo Ho, Thomas e Brasseur[32] de 2003 e concorda com os resultados obtidos.

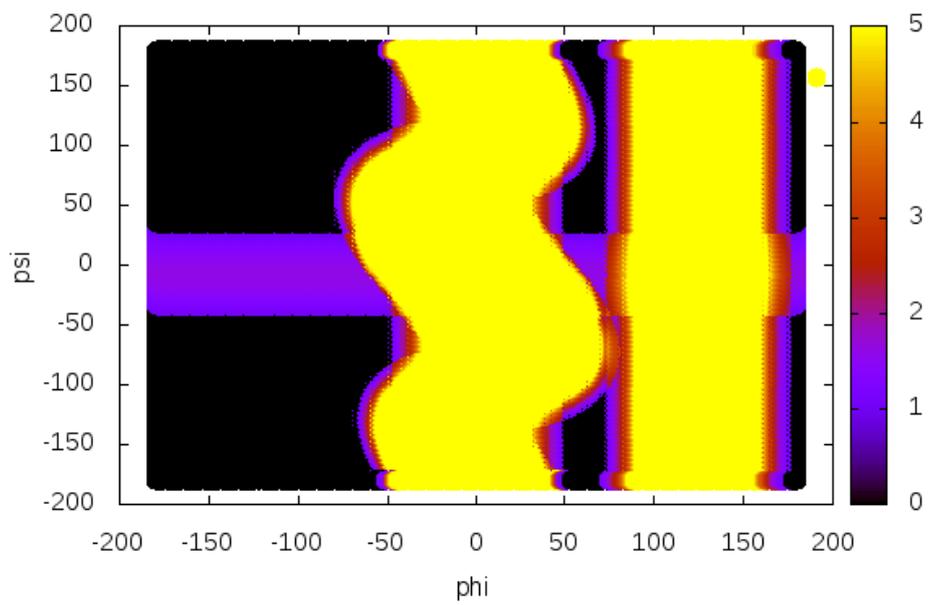


(a)

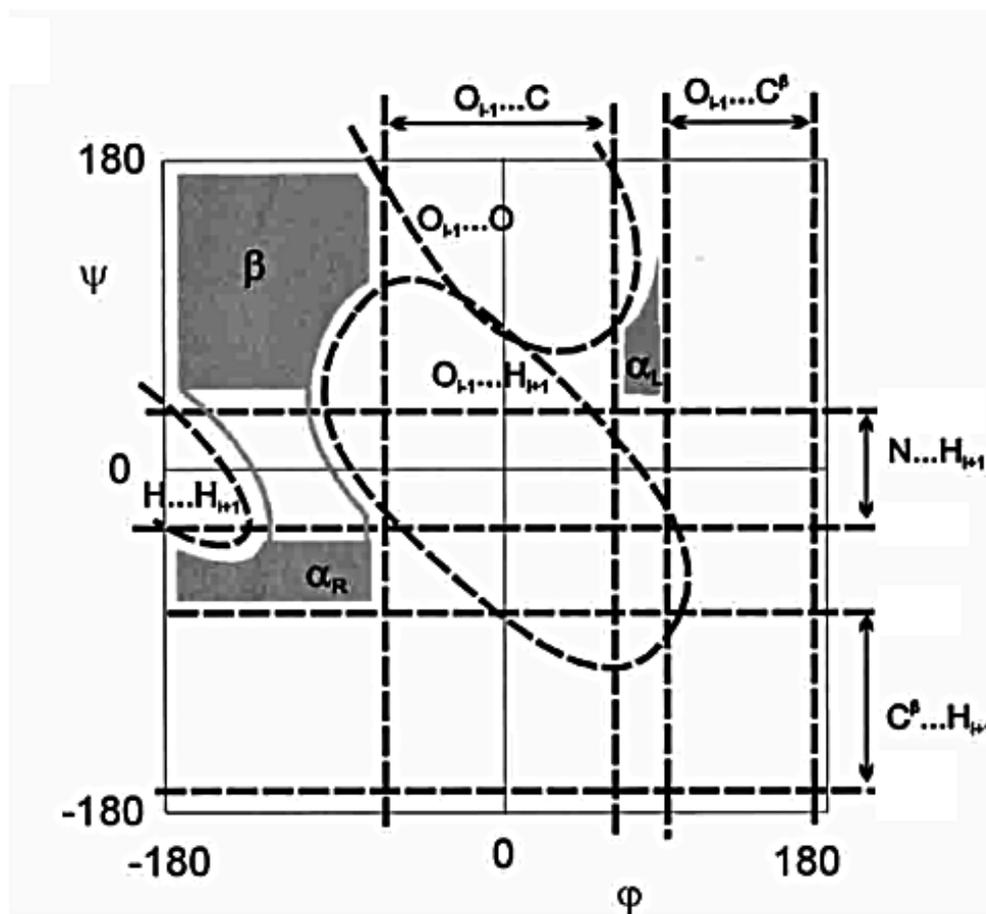


(b)

Figura 3.20: Análise comparativa dos resultados do SAXS-factor para o DNA.



(a)



(b)

Figura 3.21: Análise comparativa dos resultados do SAXS-factor para o DNA.

Capítulo 4

Construção de potenciais para simulações de previsão conformacional com o Pivot-SAXS.

4.1 Motivações.

Foram discutidas no capítulo 2, as vantagens do uso de SAXS para o estudo conformacional de moléculas biológicas em solução e como a visualização da estrutura interna da proteína é melhor analisada pela cristalografia. As técnicas espectroscópicas atuais, inclusive SAXS, permitem o estudo das superfícies da molécula porém não é possível saber o arranjo que as estruturas secundárias assumem em seu interior quando colocadas em solução. O conhecimento da estrutura interna permite estudos mais específicos às posições dos aminoácidos, como análise eletrostática dos resíduos de determinado sítio de interesse, por exemplo. Esta foi a busca que motivou este trabalho. Utilizando apenas o conhecimento de estruturas secundárias e o perfil de espalhamento da molécula, procuramos usar movimentos simples para que uma proteína em

estrutura aberta chegue à uma conformação similar à estrutura em solução.

Em casos que há mudanças conformacionais em solução o perfil de SAXS pode assumir uma curva intermediária causada pela combinação dos sinais de cada uma das estruturas[33]. A combinação ao Pivot-SAXS deve tornar possível a geração de um *ensemble* e associar quais conformações são representativas para reconstruir o sinal final.

Os movimentos de pivot usados no Ramatrix permite que novas conformações sejam geradas. O SAXS-factor é capaz de calcular o perfil de espalhamento desta nova configuração e calcular χ^2 quando comparado com a estrutura em solução. Assim, é possível definir uma energia formada por χ^2 e pela energia repulsiva de Lennard-Jones, dado na equação 4.1. Tal energia foi minimizada pelo método de Monte Carlo utilizando o critério de Metropolis¹[34]. Desta forma, as configuração aceitas evitam as sobreposições atômicas e possuem um perfil de espalhamento cada vez mais próximo ao comparativo, sendo fisicamente razoáveis.

Para evitar a desestruturação da proteína, as regiões de *coil* são as únicas que podem ser usados nos movimentos de pivot. Esta imposição reduz a quantidade de aminoácidos com ângulos que podem ser rotacionados pelo Ramatrix e número de configurações possíveis de serem acessadas pelo sistema, evitando parcialmente o paradoxo de Levinthal[35]².

4.2 Desenvolvimento do programa.

O Pivot-SAXS é formado à partir dos programas apresentados nos capítulos anteriores. O programa usa o Ramatrix como método para geração de novas conformações e o SAXS-factor para calcular o perfil de espalhamento e a discrepância entre estas e a estrutura de referência. Então, a energia

¹Detalhes no apêndice

²O paradoxo de Levinthal mostra que uma proteína deve ter restrições de movimento, além de um potencial para dirigi-los. Ela, portanto, não deveria conseguir acessar todas as configurações possíveis, pois assim o enovelamento levaria um tempo astronômico.

formada pela energia de Lennard-Jones e pelo fator de discrepância χ^2 , dada na equação 4.1 é minimizado através do critério de Metropolis.

$$E = \eta_j E_{lj} + \eta_\chi \chi^2, \quad (4.1)$$

onde η_j e η_χ são parâmetros de ajuste da energia de Lennard-Jones E_{lj} e de χ^2 , respectivamente.

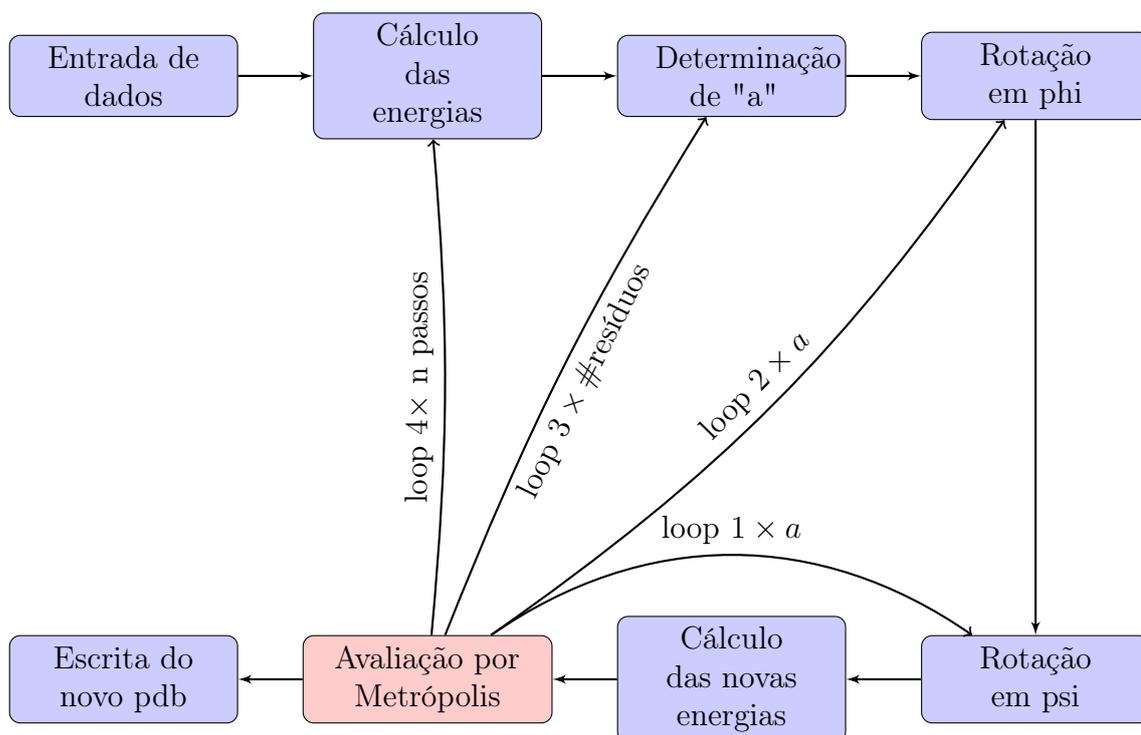


Figura 4.1: Fluxograma do algoritmo utilizado em Pivot-SAXS.

O movimento de pivot iniciado com grandes ângulos possibilita uma convergência mais rápida da estrutura aberta para uma conformação mais fechada pela falta de impedimentos estéricos. Neste ponto, a amplitude do movimento reduz a aceitação, e o ângulo de rotação precisa ser atenuado para possibilitar o ajuste das estruturas sob essas condições. Para manter o ciclos de 360° para cada um dos ângulos diedrais, uma variável "a" é introduzida. "a" representa a quantidade de estados que phi e psi podem acessar em 360 graus e é alterada dependendo da taxa de aceitação. Por exemplo, o

programa inicia com $a = 3$ e rotações de 120 graus em cada um dos ângulos, ao atingir uma aceitação menor que 60%, "a" recebe um valor maior e novas conformações são geradas à partir de rotações com ângulos menores.

No fluxograma 4.1 está representado as etapas executadas pelo programa Pivot-SAXS.

4.3 Validação e resultados do Pivot-SAXS.

A validação do Pivot-SAXS foi feita utilizando proteínas pequenas para reduzir o tempo computacional necessário para a execução do programa. Em todos os teste, a proteína passou primeiro pelo Pivot-SAXS modificado de forma que todas as conformações são aceitas. Uma conformação aberta é escolhida para iniciar o método descrito na sessão anterior.

A Vilina³ foi a primeira a ser testada. Foi, também, utilizada durante todo o processo de estruturação e adequação do programa final. A figura 4.2 mostra como a estrutura aberta converge para a conformação obtida pela cristalografia. No início da simulação o RMSD⁴ entre as estruturas era de 7,469Å. Após a execução do Pivot-SAXS, as estruturas, representadas na figura 4.2b, possuem um RMSD= 4,869Å.

Testes com a vilina permitiram validar o programa, ainda que um resultado exato não tinha sido obtido (figura 4.2). Todavia, se mostrou eficiente para seguir os testes com outras proteínas. Uma destas proteínas foi a *engrailed homeodomain* de *Drosophila melanogaster*[25]⁵, formada por três hélices e 53 resíduos. A simulação inicia em estruturas com um RMSD= 18,669Å, apresentada na figura 4.3a, e termina com RMSD= 5,122Å. A figura 4.3b compara a estrutura ao final da simulação com a estrutura cristalográfica.

³PDB ID: 1VII

⁴do inglês *root main square deviation* - desvio padrão entre as posições atômicas de duas estruturas proteicas.

⁵PDB ID: 1ENH

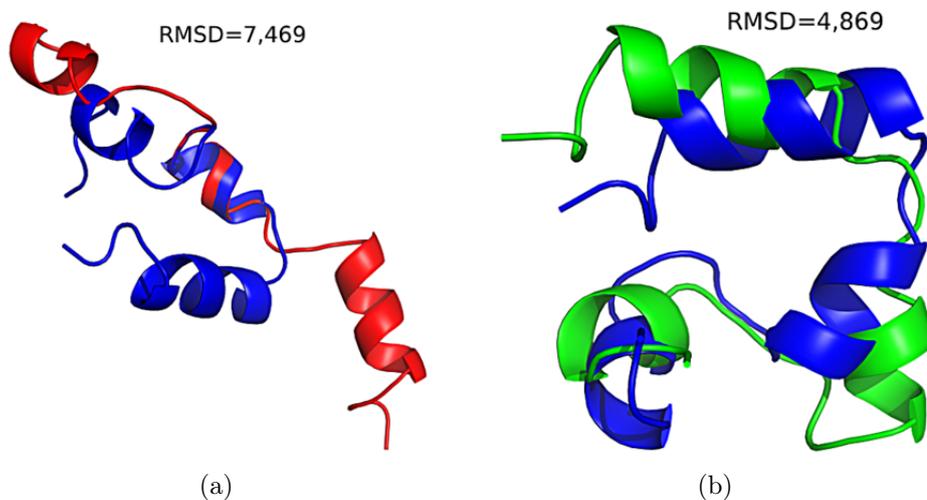


Figura 4.2: Mudança estrutural da vilina causada pelo movimento de pivot.

A quimotripsina inibidora 2 [26]⁶ tem 63 resíduos, formando 2 α -hélices e 2 folhas- β e na classificação SCOP é uma $\alpha+\beta$. Sua região N-terminal é em α -hélice enquanto a região em C-terminal é formada por folhas β . Esta característica a torna uma boa alternativa para testar a eficácia do potencial para a formação das folhas- β .

A quimotripsina é uma proteína não-globular e formada de domínios α e β e pela figura 4.4 podemos notar que tais domínios foram formados através do algoritmo de pivot.

Para proteínas em α -hélice o programa se mostrou eficiente na convergência das estruturas. O resultado para a proteína $\alpha + \beta$ é promissor enquanto resultado preliminar, porém mais testes precisam ser realizados para a validação da formação das folhas- β .

⁶PDB ID: 1YPC

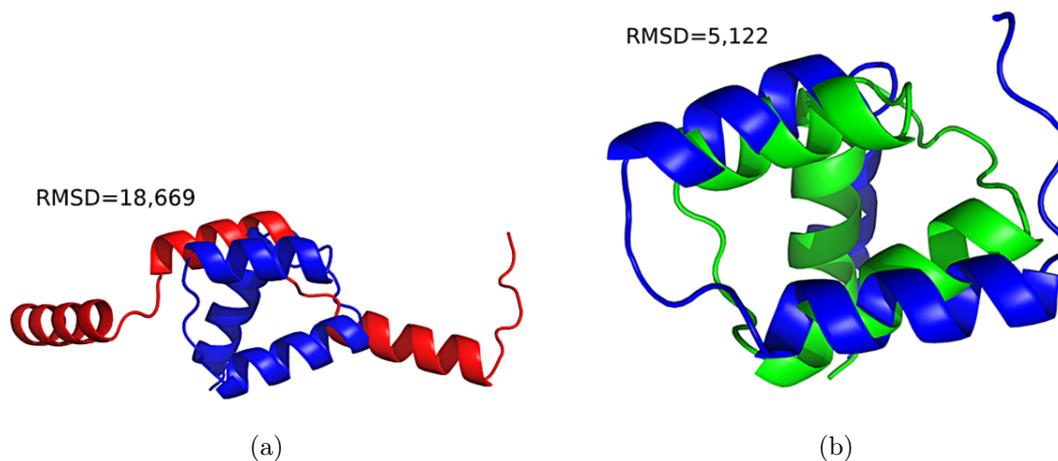


Figura 4.3: Mudança estrutural da 1ENH causada pelo movimento de pivot.

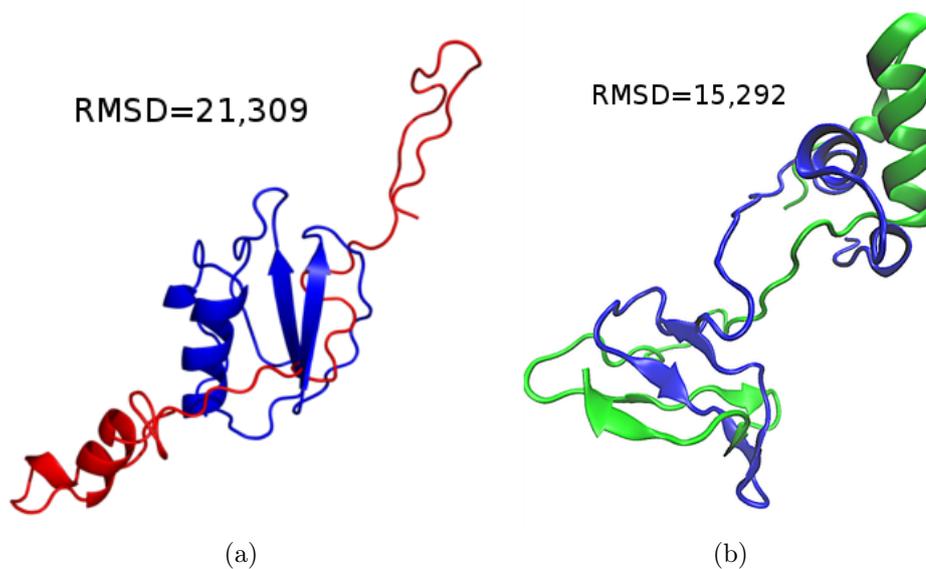


Figura 4.4: Estruturação dos domínios α e β da proteína 1YPC causada pelo movimento de pivot.

Tabela 4.1: Valores de RMSD para cada uma das proteína na conformação aberta e final com relação à estrutura cristalográfica.

| Proteínas | RMSD para cadeia aberta (Å) | RMSD para cadeia final (Å) |
|-----------|-----------------------------|----------------------------|
| 1VII | 7,469 | 4,869 |
| 1ENH | 18,669 | 5,122 |
| 1YPC | 21,309 | 15,269 |

Capítulo 5

Discussão.

O objetivo do trabalho é utilizar apenas movimentos de pivot para convergir um estrutura aberta em sua estrutura terciária representativa em solução. Para cumprir este objetivo, foi utilizado um potencial formado pela energia de Lennard-Jones e pelo fator de discrepância entre as curvas de espalhamento da estrutura cristalográfica¹ e da conformação gerada pelo programa. Além de minimizar o potencial, o programa precisava gerar uma nova conformação através da rotação de um dos ângulos diedrais e calcular a curva de espalhamento desta configuração, além de calcular o fator χ^2 . Os programas SAXS-factor e Ramatrix foram desenvolvido com estes fins. Seus algoritmos são parte do Pivot-SAXS, portanto, para que o objetivo seja alcançado cada etapa deve ser analisada e validada.

Desde o início do trabalho, aponta-se que uma das diferenças entre o SAXS-factor e programas similares disponíveis é a implementação do fator de espalhamento. O comportamento das curvas no capítulo 2 se referem à estrutura cristalográfica no vácuo. Isso implica que não há solvente para afetar o espalhamento ou alterar o estado de relaxamento da estrutura. Essa diferença é visível nas comparações com os programas FoXS e Crysol.

¹Essas curvas de espalhamento serão, futuramente, substituídas por curvas de SAXS.

Mesmo com essa alteração, o SAXS-factor se provou eficiente no cálculo de uma curva de espalhamento teórica equiparável à curva experimental. Isso é importante para garantir o entendimento do programa. Ainda, futuras versões do programa terão o termo do solvente implementado.

O desenvolvimento do Ramatrix foi bastante instrutivo sobre a forma de rotação dos ângulos diedrais. Esta etapa também foi usada para o ajuste do tamanho dos átomos e do cálculo da energia de Lennard-Jones. A comparação com o mapa de Ramachandran foi a forma encontrada para verificar tanto a rotação feita pelo programa quanto o cálculo da energia. O tamanho dos átomos de carbono precisam levar em conta a existência do hidrogênio, não presentes na nossa avaliação. Pela comparação dos gráficos gerados pelo Ramatrix com os mapas de Ramachandran, percebe-se que as regiões de impedimento estérico relativas ao hidrogênio também são incluídas.

Tendo ambos programas-base validados, iniciam-se os testes com Pivot-SAXS. Para isso, foi necessária uma versão modificada do programa para gerar uma estrutura aberta². Os resultados comparativos do capítulo 3 e os valores de RMSD da tabela 4.1 mostram que o potencial formado pela energia de Lennard-Jones e χ^2 é muito eficaz para o tratamento de proteínas formadas por hélices alpha.

O estudo para folhas beta foi iniciado com a 1YPC, uma proteína formada por duas hélices- α e duas folhas- β paralelas. Porém a estruturação de uma folha beta se mostrou um pouco complicada. Enquanto as hélices chegaram a estruturas similares, as folhas beta levaram um tempo computacional maior e não foram tão precisas. Esta diferença pode se referir ao posicionamento das cadeias laterais e a formação das ligações de hidrogênio. Ao criar uma estrutura aberta, as cadeias laterais assumem posições aleatórias e as ligações de hidrogênio são destruídas. O programa não conta com um potencial de pontes de hidrogênio, o que pode ser implementado. Mesmo

²Nesta versão a taxa de aceitação do critério de Metropolis foi mantida à 100%.

mantendo os resíduos das folhas beta rígidos, o coil entre elas é rotacionado. Durante o processo de convergência para a estrutura cristalográfica o número de combinações de ϕ e ψ é diretamente proporcional à quantidade de resíduos. Dentre estas, apenas algumas permitem o alinhamento necessário para a formação das ligações de hidrogênio.

O paradoxo de Levinthal aborda quanto tempo levaria para uma proteína acessar todos os estados conformacionais possíveis. Estes estados são contados considerando a quantidade de resíduos da proteína e quantas conformações aquele resíduo pode assumir. Caso uma proteína pudesse assumir qualquer valor de ϕ e ψ em todos os seus resíduos, seu espaço conformacional seria enorme e o tempo para acessá-lo por completo seria astronômico. Este é um conceito que pode ser aplicado ao tempo computacional do programa.

Uma das formas usadas para evitar que o processamento do programa fosse muito longo foi considerar os resíduos das α -hélices e folhas β como estruturas rígidas. Porém os resíduos de loop devem ser capazes de assumir diversos valores de phi e psi. Essa liberdade nos ângulos diedrais aumenta o número de conformações possíveis. Desta forma, o tempo de processamento do programa é extremamente alto. Este é um dos motivos para utilizar uma abordagem markoviana durante a simulação, ou seja as novas configurações não são geradas aleatoriamente em nenhum quesito. Esta abordagem não garante uma varredura completa do espaço conformacional, mas foi eficiente na convergência do sistema.

Outro fator analisado foi a amplitude do movimento de pivot. O ângulo de rotação é reduzido de acordo com a aceitação das novas conformações pelo critério de Metropolis. Quando a estrutura está aberta, a taxa de aceitação, assim como a convergência são altas e conforme a proteína se fecha, aumentando o empacotamento da estrutura, o número de choques estéricos aumenta. Isto ocorre devido a amplitude da rotação em pontos distantes de seu eixo.

Pode-se concluir que o movimento de pivot é eficiente para o objetivo proposto até que a estrutura se torne empacotada e a amplitude do movimento se torne um empecilho à convergência. Devido ao alto tempo de processamento, parece viável a associação de uma etapa de monte carlo tradicional para ampliar a varredura do espaço conformacional e outra de dinâmica molecular para os ajustes mais finos.

Capítulo 6

Trabalhos Futuros.

Nossos resultados visam a implementação de um *web-server*¹ com diversas complementações à serem feitas ao programa. Dentre elas, a inclusão da contribuição da camada de hidratação no cálculo do fator de espalhamento para melhorar o modelo atual, uma possível predição de estruturas terciárias empregando ferramentas de predição de estruturas secundárias e predição combinada com enterramento atômico[36].

Sabemos que a abordagem markoviana utilizada inicialmente não fornece uma boa varredura do espaço conformacional. Os resultados apresentados neste trabalho são válidos, porém precisam passar por uma etapa em que o Monte Carlo tradicional seja aplicada, ou seja, que as novas conformações sejam geradas de forma aleatória. Esta fase garante que a conformação obtida é o mínimo global do potencial minimizado. A taxa de aceitação do movimento de pivot decai conforme a energia do sistema diminui. Assim, o tempo computacional para acomodar as cadeias laterais é alto o suficiente para buscar outra abordagem para ajustes finos, o que abre a possibilidade de fazer um passo seguinte empregando Dinâmica Molecular. Então, o ideal seria iniciar a simulação em movimento de pivot e finalizá-la com uma dinâmica

¹O servidor será disponibilizado no site do grupo de pesquisa: <http://oliveira.df.ibilce.unesp.br/>.

molecular usando um modelo baseado em estrutura e o fator de discrepância χ^2 .

Uma adição interessante, é a possibilidade de usar preditores de estruturas secundárias. Com este recurso, permitimos que os dados de entrada sejam a sequência primária e o perfil de espalhamento e como resultado uma estrutura terciária compatível com a curva de espalhamento. Esta abordagem possibilita o uso do programa, inclusive, para proteínas que por diversos motivos não podem ser cristalizadas. Tornando-o uma ferramenta importante no estudo de proteínas em solução.

Referências.

- 1 GLATTER, O.; KRATKY, O. *Small Angle X-ray Scattering*. New York: Academic Press, 1982. 514 p. ISBN 0-12-286280-5.
- 2 STUHRMANN, H. B. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Crystallographica Section A*, International Union of Crystallography, v. 26, n. 3, p. 297–306, 1970. ISSN 0567-7394.
- 3 SVERGUN, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, v. 76, n. 6, p. 2879–2886, 1999. ISSN 0006-3495.
- 4 SVERGUN, D. I.; PETOUKHOV, M. V.; KOCH, M. H. J. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal*, v. 80, n. 6, p. 2946–2953, 2001. ISSN 00063495.
- 5 SVERGUN, D. I.; KOCH, M. H. J. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, v. 66, n. 10, p. 1735–1782, 2003. ISSN 0034-4885.
- 6 DONIACH, S. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews*, v. 101, n. 6, p. 1763–1778, 2001. ISSN 00092665.
- 7 PUTNAM, D. K.; LOWE, E. W.; MEILER, J. Reconstruction of SAXS Profiles from Protein Structures. *Computational and structural biotechnology journal*, v. 8, p. e201308006, 2013. ISSN 2001-0370.
- 8 FEIGN, L.; SVERGUN, D. I. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. New York: Plenum Press, 1989. v. 40. 224–224 p. ISSN 03237648. ISBN 0-306-42629-3.
- 9 PATTERSON, A. L. A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Kristallographie - Crystalline Materials*, v. 90, n. 1-6, 1 1935. ISSN 2196-7105.

- 10 DEBYE, P.; BUECHE, A. M. Scattering by an inhomogeneous solid. *Journal of Applied Physics*, v. 20, n. 6, p. 518–525, 1949. ISSN 00218979.
- 11 POROD, G. The X-ray small-angle scattering of close-packed colloid systems. *Kolloid Z*, v. 124, p. 83, 1951.
- 12 GUINIER, A.; FOURNET, G. *Small angle scattering of X-rays*. New York: John Wiley & Sons, Inc., 1956. v. 19. 594–594 p. ISSN 00223832.
- 13 RAMBO, R. P.; TAINER, J. A. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Current Opinion in Structural Biology*, Elsevier Ltd, v. 20, n. 1, p. 128–137, 2010. ISSN 0959440X.
- 14 DEBYE, P.; ANDERSON, H. R.; BRUMBERGER, H. Scattering by an Inhomogeneous Solid. II. The Correlation Function and Its Application. *Journal of Applied Physics*, v. 28, n. 6, p. 679, 1957. ISSN 00218979.
- 15 SVERGUN, D. I.; BARBERATO, C.; KOCH, M. H. CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, International Union of Crystallography, v. 28, n. 6, p. 768–773, 1995. ISSN 00218898.
- 16 FRANKE, D.; SVERGUN, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, v. 42, n. 2, p. 342–346, 2009. ISSN 00218898.
- 17 CHACÓN, P. et al. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophysical journal*, v. 74, n. 6, p. 2760–2775, 1998. ISSN 00063495.
- 18 CHACÓN, P. et al. Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *Journal of molecular biology*, v. 299, n. 5, p. 1289–1302, 2000. ISSN 0022-2836.
- 19 SCHNEIDMAN-DUHOVNY, D.; HAMMEL, M.; SALI, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research*, v. 38, n. SUPPL. 2, p. 540–544, 2010. ISSN 03051048.
- 20 DEBYE, P. Zerstreung von Röntgenstrahlen. *Annalen der Physik*, v. 351, n. 6, p. 809–823, 1915. ISSN 00033804.
- 21 VOLKOV, V. V.; SVERGUN, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *International Union of Crystallography*, p. 860–864, 2003.

- 22 BERMAN, H. M. et al. The Protein Data Bank. *Nucleic acids research*, v. 28, n. 1, p. 235–42, 1 2000. ISSN 0305-1048. Disponível em: <<http://www.rcsb.org/pdb/>>.
- 23 BROWN, P. J. et al. Intensity of diffracted intensities. In: *International Tables for Crystallography*. Chester, England: International Union of Crystallography, 2006. cap. 6.1, p. 554–595.
- 24 HURA, G. L. et al. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *PMC*, v. 6, n. 8, p. 606–612, 2011.
- 25 CLARKE, N. et al. Structural studies of the engrailed homeodomain. *Protein Sci.*, v. 3, p. 1779–1787, 1994.
- 26 HARPAZ, Y. et al. Direct observation of better hydration at the N terminus of an alpha-helix with glycine rather than alanine as the N-cap residue. *Proc. Natl. Acad. Sci. USA*, v. 91, p. 311–315, 1994.
- 27 RAMACHANDRAN, G. N.; SASISEKHARAN, V. Conformation of polypeptides and proteins. In: *Advances in protein chemistry*. [S.l.: s.n.], 1968. v. 23, n. September, p. 283–438. ISBN 0065-3233.
- 28 KLEYWEGT, G. J.; JONES, T. A. Phi/Psi-chology: Ramachandran revisited. *Structure*, v. 4, n. 12, p. 1395–1400, 1996. ISSN 09692126.
- 29 BRAGG, W. L. The arrangement of atoms in crystals. *Philosophical Magazine*, v. 40, n. 236, p. 169–189, 1920. ISSN 1941-5982.
- 30 HO, B. K.; BRASSEUR, R. The Ramachandran plots of glycine and pre-proline. *BMC structural biology*, v. 5, p. 14, 2005. ISSN 1472-6807.
- 31 THORNTON, S. T.; MARION, J. B. *Classical Dynamics of Particles and Systems*. 5^a. ed. [S.l.]: Thomson Learning, 2004. 672 p. ISBN 9780534408961.
- 32 HO, B. K.; THOMAS, A.; BRASSEUR, R. Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein science : a publication of the Protein Society*, v. 12, n. 11, p. 2508–2522, 2003. ISSN 0961-8368.
- 33 JAMROS, M. A. et al. A Combined small angle x-ray scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape. *Journal of Biological Chemistry*, v. 285, n. 46, p. 36121–36128, 2010. ISSN 00219258.
- 34 METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, v. 21, n. 6, p. 1087, 1953. ISSN 00219606.

35 ZWANZIG, R.; SZABO, A.; BAGCHI, B. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*, v. 89, n. 1, p. 20–2, 1992. ISSN 0027-8424.

36 ARAUJO, A. F. Pereira de; ONUCHIC, J. N. A sequence-compatible amount of native burial information is sufficient for determining the structure of small globular proteins. *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 45, p. 19001–4, 11 2009. ISSN 1091-6490. Disponível em: <<http://www.pnas.org/cgi/content/long/106/45/19001>>.

Apêndice A

Critério de Metropolis.

O critério de Metropolis é utilizado no programa como forma de minimização do potencial construído no capítulo 4. As estruturas geradas, de forma não aleatória, são comparadas com a conformação anterior, então, o sistema de menor energia, ou com flutuações térmicas razoáveis, é selecionado.

O critério de Metropolis foi implementado de acordo com o esquema A.1. A nova estrutura gerada no programa Pivot-SAXS tem sua energia calculada, assim como variação da energia com relação à estrutura anterior. Caso a variação seja negativa a nova conformação é aceita, ou seja, ela assume a posição de conformação antiga. Quando a variação é positiva, um número aleatório é comparado com o peso de Boltzmann P_b dado pela equação A.1. Se o número aleatório for maior a nova conformação é aceita, se não a antiga se mantém. O ciclo reinicia.

$$Pb = e^{\frac{-E_{nova}}{kT}} \quad (A.1)$$

As simulações são feitas fora do equilíbrio térmico e o método minimiza a energia do sistema aceita considerando pequenas flutuações térmicas da ordem de kT.

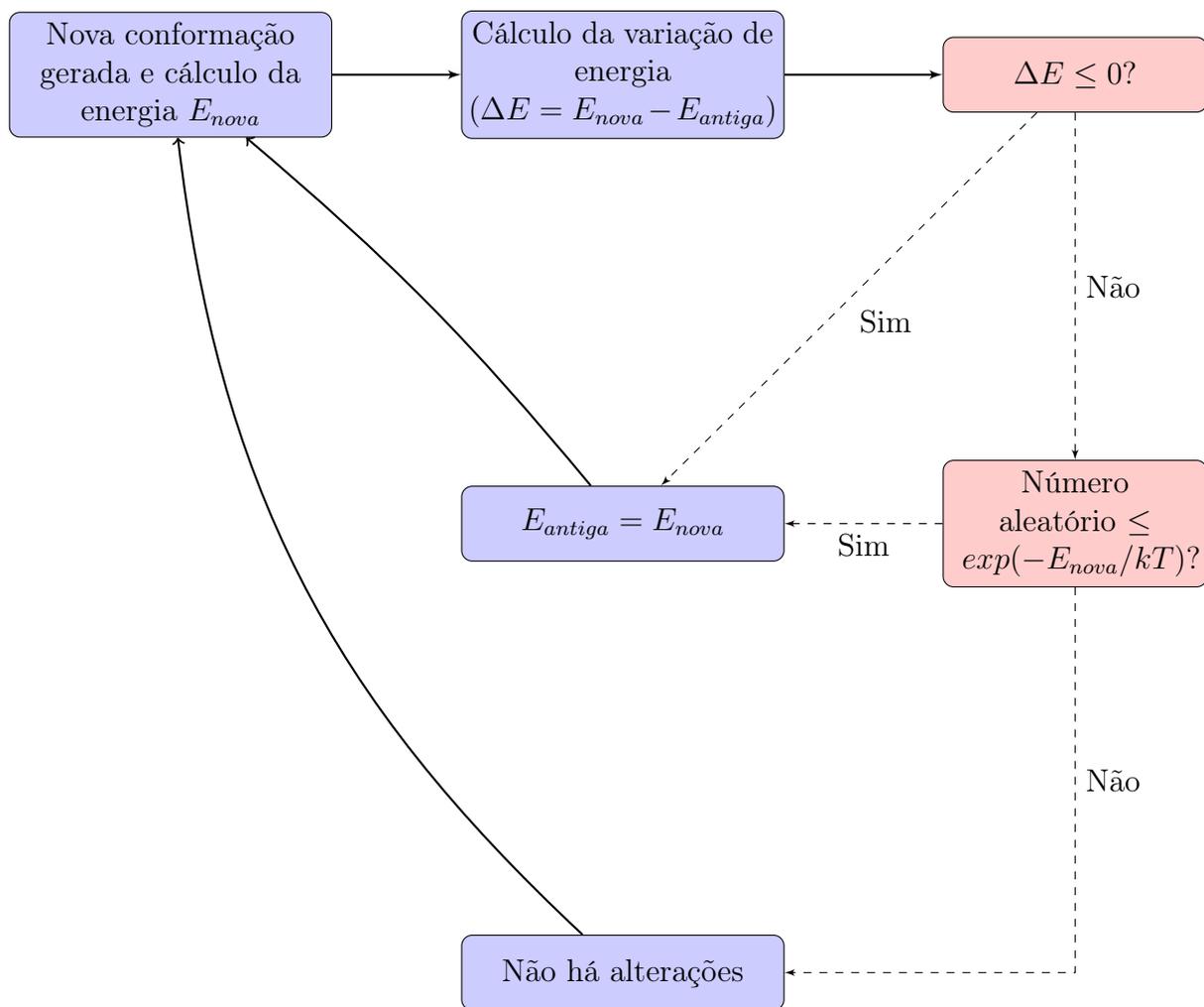


Figura A.1: Fluxograma do critério de Metropolis. A energia de uma conformação é calculada e comparada com a energia de uma conformação anterior. Se a energia diminuir, a conformação é aceita. Se a energia aumentar, um número aleatório é gerado e comparado com o peso de Boltzmann. Se o número for maior, a conformação é aceita, se não a simulação segue a partir da conformação anterior.