UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

NOVOS MÉTODOS PARA ANÁLISE DE CURVAS
DE ESPALHAMENTO A BAIXO ÂNGULO APLICADOS
A UM INIBIDOR DE α -AMILASE, À HEXOCINASE E
À ASPARTATO TRANSCARBAMILASE

Claudio Barberato

Tese apresentada no Instituto de Física de
São Carlos, da Universidade de São
Paulo, para obtenção do título de Doutor
em Ciências na área de Física Aplicada.

Orientadora: Yvonne Primerano Mascarenhas

São Carlos
1996

IFSC-USP SERVIÇO DE BIBLIOTECA E
INFORMAÇÃO

Barberato, Claudio

Novos métodos para a análise de curvas de espalhamento a baixo ângulo aplicados a um inibidor de α -amilase, à hexocinase e a aspartato transcarbamilase.

Claudio Barberato - São Carlos, 1996.

100 p.

Tese (Doutorado) - Instituto de Física de São Carlos, 1996

Orientadora: Prof. Dra. Yvonne Primerano Mascarenhas.

1. Espalhamento a baixo ângulo 2. Biologia Molecular

**UNIVERSIDADE
DE SÃO PAULO**

Instituto de Física de São Carlos

Av. Dr. Carlos Botelho, 1465
CEP 13560-250 - São Carlos - SP
Brasil

Fone (016) 272-6222

Fax (016) 272-2218

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE **CLAUDIO BARBERATO** APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, UNIVERSIDADE DE SÃO PAULO, EM 01/08/1996.

COMISSÃO JULGADORA:

Prof.ª Dra. Yvonne Primerano Mascarenhas/IFSC-USP

Prof. Dr. Eduardo Ernesto Castellano/IFSC-USP

Prof.ª Dra. Lia Queiroz do Amaral/IF-USP

Prof.ª Dra. Iris C. Linares de Torriani/UNICAMP

Prof. Dr. Edgar Dutra Zanutto/UFSCar

Aos meus pais

Este trabalho recebeu o apoio financeiro da:

Fundação de Amparo a Pesquisa no Estado de São Paulo (FAPESP)

e Conselho Nacional de Pesquisa e Desenvolvimento (CNPq)

AGRADECIMENTOS

Ao Dr. Dmitri Svergun e Dr. Michel Koch, pela orientação durante o período (nov./93-dez./95) em que estagiei no European Molecular Biology Laboratory (EMBL).

Prof. Dra. Yvonne P. Mascarenhas, pelos ensinamentos e orientação e amizade.

Aos professores Glaucius Oliva, Richard Garrat, Eduardo E. Castellano e Júlio Zuckerman-Schpector pelas discussões, ensinamentos e amizade.

Ao Paulo Loureiro, pela inestimável amizade via e-mail.

À toda turma do grupo de cristalografia: Paulão, Cristina, Tereza, Portezani, Jorge, Beatriz, Bianca, Zac, Ignez, Marcão, Valma, Fernando, Wanda, Maria Helena... pela ajuda e bons momentos.

À toda turma do European Molecular Biology Laboratory: Tasos, Gwyndaf, Ivo, Thomas, Paola, Renaud, Patricia, Vangelis, Martin, Michael, Gabi, Niki, Adelia, Sandra, Howard, Sus, Susana, Alessia, Margret, Zhera, Michael Kozin, Vladimir, Bernd, ... pela ajuda e bons momentos.

Ao meu cunhado Luíz Antônio Genghini pela amizade, telefonemas e ajuda.

À minha família.

SUMÁRIO

Lista de figuras.....	viii
Lista de tabelas.....	x
Lista de Abreviaturas.....	xi
Resumo.....	xii
Abstract.....	xiii
CAPÍTULO 1 - INTRODUÇÃO.....	1
CAPÍTULO 2 - RAIOS-X.....	2
2.1 - Absorção de raios-X.....	3
CAPÍTULO 3 - PRINCÍPIOS GERAIS DO ESPALHAMENTO	
DE RAIOS-X A BAIXO ÂNGULO.....	5
3.1 - Introdução.....	5
3.2 - Espalhamento de uma onda plana por uma densidade eletrônica.....	7
3.3 - Intensidade de uma partícula fixa.....	9
3.4 - Intensidade de uma partícula em solução.....	11
3.5 - Partículas com densidade eletrônica uniforme em solução.....	13
3.6 - Comportamento assintótico da Intensidade (Lei de Porod).....	16
3.7 - Determinação dos parâmetros integrais de partículas em solução.....	17
3.7.1 - Volume e superfície.....	17
3.7.2 - Raio de giro.....	18
3.8 - Informação contida em uma curva de SAXS.....	20
CAPÍTULO 4 - DETERMINAÇÃO DO ELIPSÓIDE EQUIVALENTE.....	24
4.1 - <i>ELLFIT</i> - Um programa para a avaliação do elipsóide equivalente...	24

CAPÍTULO 5 - INTERPRETAÇÃO DO ESPALHAMENTO DE RAIOS-X

POR PARTÍCULAS INOMOGÊNEAS..... 30

5.1 - Espalhamento de raios-X por uma partícula inhomogênea..... 30

5.2 - Técnica da variação de contraste..... 35

CAPÍTULO 6 - DETERMINAÇÃO DO ENVELOPE MOLECULAR

ATRAVÉS DA TEORIA DE MULTIPÓLOS..... 39

6.1 - Harmônicos esféricos..... 39

6.2 - Expansão de multipolos..... 41

6.3 - Determinação do envelope molecular de uma
inibidora de α -amilase..... 47

6.3.1 - Resultados..... 49

6.3.2 - Discussão..... 50

CAPÍTULO 7 - CÁLCULO DA CURVA DE ESPALHAMENTO DE

PROTEÍNAS EM SOLUÇÃO BASEADO EM

COORDENADAS ATÔMICAS..... 55

7.1 - *CRY SOL*.....57

7.1.1 - Avaliação das amplitudes parciais.....60

7.1.2 - Implementação do programa..... 65

7.1.3 - Aplicações de *CRY SOL*..... 67

7.1.3.1 - Lisozima..... 67

7.1.3.2 - Hexocinase.....69

7.2 - Discussão..... 71

CAPÍTULO 8 - DETERMINAÇÃO DA POSIÇÃO RELATIVA ENTRE	
AS SUBUNIDADES DE UMA PROTEÍNA.....	73
8.1 - <i>HOMDIM</i>	80
8.2 - Aspartato Transcarbamilase.....	82
CAPÍTULO 9 - CONCLUSÕES E PERSPECTIVAS FUTURAS.....	92
REFERÊNCIAS BIBLIOGRÁFICAS.....	94
APÊNDICE A - Tratamento de dados experimentais.....	97
ANEXO 1 - Disquete com as estruturas da hexocinase, com os	
pacotes dos programas <i>CRY SOL</i>, <i>ELLFIT</i> e <i>HOMDIM</i>	
e com os dados experimentais para as proteínas	
estudadas nesta tese.	

LISTA DE FIGURAS

Capítulo 2

Figura 2.1 - Espectro típico de radiação de raios X emitido por um tubo de cobre	2
---	---

Capítulo 3

Figura 3.1 - Experimento típico de SAXS.	6
Figura 3.2 - Feixe de raios X incidente e espalhado por uma densidade de cargas.	7
Figura 3.3 - Geometria do espalhamento de raios X.	7
Figura 3.4 -função de Patterson...	11
Figura 3.5 - Conceito do volume comum (V_c).	15
Figura 3.6 - Representação da função $sI(s)$ segundo o teorema da amostragem.	21

Capítulo 4

Figura 4.1 - Ajustes entre curvas experimentais (pontos) e elipsóides equivalentes	26
---	----

Capítulo 5

Figura 5.1 - A densidade eletrônica de uma proteína...	30
Figura 5.2 - Densidade eletrônica da proteína no vácuo como superposição...	35
Figura 5.3 - A densidade eletrônica de uma proteína em solução como a superposição...	36

Capítulo 6

Figura 6.1 - Curva de SAXS para o modelo da inibidora de α -amilase...	49
--	----

Figura 6.2 - Modelo da inibidora de α -amilase obtida por determinação...	52
---	----

Figura 6.3 - Modelo da figura 6.2 rodado 90° ao redor do eixo x.	53
--	----

Figura 6.4 - Modelo da figura 6.3 rodado 45° ao redor do eixo z.	54
--	----

Capítulo 7

Figura 7.1 - Volume excluído e suas representações:	55
--	----

Figura 7.2 - Representação esquemática de uma proteína em solução...	57
---	----

Figura 7.3 - Densidade eletrônica de átomos ... esferas Gaussianas...	63
--	----

Figura 7.4 - Avaliação da função $F(\varpi)$...	64
---	----

Figura 7.5 - Curva de SAXS para a lisozima...	67
--	----

Figura 7.6 - Intensidade espalhada pela estrutura atômica no vácuo...	68
--	----

Figura 7.7 - Curva de experimental para a hexocinase...	70
--	----

Capítulo 8

Figura 8.1 - Exemplificação do procedimento para a obtenção da estrutura...	79
--	----

Figura 8.2 - Modelos para as estruturas quaternárias da ATCase...	84
--	----

Figura 8.3 - Comparação entre as curvas de SAXS para os estados...	86
---	----

Figura 8.4 - Ajuste entre a curva experimental ... <i>HOMDIM</i> ...	88
---	----

Figura 8.5 - Valor χ -quadrado <i>versus</i> a metade do deslocamento...	90
--	----

LISTA DE TABELAS

Capítulo 4

Tabela 4.1 - Tabela 4.1: Comparação entre as dimensões máximas... 27

Tabela 4.2 - Resultados obtidos utilizando o programa *ELLFIT*. 28

Capítulo 6

Tabela 6.1 - Primeiros termos dos esféricos harmônicos. 40

Tabela 6.2 - Coeficientes f_{lm} encontrados pela minimização do funcional 8.25. 48

Tabela 6.3 - Parâmetros geométricos da inibidora de α -amilase. 50

Capítulo 7

Tabela 7.1 - Parâmetros dos átomos e grupos atômicos. 62

Tabela 7.2 - Valores obtidos para as três conformações da hexocinase encontradas... 71

LISTA DE ABREVIATURAS

SAXS - Small Angle X-ray Scattering.

ATCase - Aspartato Transcarbamilase.

RESUMO

Este trabalho teve por finalidade a implementação e desenvolvimento de novos métodos para a análise de curvas de espalhamento de raios X a baixo ângulo por sistemas monodispersos. O resultado básico final deste trabalho foi a confecção de três programas de computador e suas aplicações em proteínas de interesse biológico. *ELLFIT* é um programa de computador que encontra o elipsóide cuja curva de SAXS melhor se ajusta a uma dada curva experimental. Para casos favoráveis este programa é capaz de determinar a dimensão máxima e algumas características básicas do formato da partícula. *CRY SOL* é um programa para a avaliação de curvas de espalhamento em solução para proteínas com estrutura atômica conhecida. O programa usa a expansão de multipolos para o cálculo rápido da promediação espacial e simula uma camada de hidratação ao redor da proteína. *CRY SOL* pode predizer a curva de SAXS de uma determinada proteína e compara-la com dados experimentais. *HOMDIM* é um programa para a determinação da posição das sub-unidades de um homodímero no caso de ser conhecida somente a estrutura da sub-unidade sozinha. Dada a curva experimental e a amplitude da sub-unidade, *HOMDIM* procura os parâmetros posicionais que descrevem o homodímero. Estes e outros programas foram aplicados a várias proteínas. O método da expansão de multipolos foi usado na determinação do envelope molecular de uma inibidora de α -amilase. O programa *CRY SOL* foi utilizado para resolver uma ambiguidade na estrutura quaternária cristalina da hexocinase e o programa *HOMDIM* para a proposição de um novo modelo para a estrutura quaternária da aspartato transcarbamilase no estado R em solução.

ABSTRACT

This work was aimed at the implementation and development of new methods for solution scattering analysis of monodisperse systems. The basic final result of this work was the development of three programs and their applications to proteins of biological interest. *ELLFIT* is a computer program which finds the ellipsoid whose SAXS curve has the best fit to a given experimental curve. In favourable cases, this program is able to determine the maximum dimension and some basic characteristics of the particle shape. *CRY SOL* is a program for evaluating the solution scattering from proteins of known structure. The program uses multipole expansion for fast calculation of the spherically averaged scattering pattern and takes into account the hydration shell. Given the atomic coordinates it can predict the solution scattering curve and compare it with the experimental scattering curve. *HOMDIM* is a program to determine the position of both subunits of a homodimer when only one sub-unit structure is known. Given the experimental curve of the homodimer and the subunit scattering amplitudes, *HOMDIM* searches for the positional parameters which describe the homodimer. These and other programs were used to study several proteins. The multipole expansion method was used in the shape determination of an α -amylase inhibitor. The program *CRY SOL* was used to solve the ambiguity in the hexokinase quaternary crystal structures and the program *HOMDIM* was utilized for the quaternary structure modeling of the R-state of the aspartate transcarbamylase in solution.

CAPÍTULO 1 - INTRODUÇÃO

Os últimos anos evidenciaram um enorme esforço da comunidade científica para o entendimento dos processos bioquímicos envolvendo proteínas. Para que as descrições desses processos sejam confiáveis, estas tem que compatibilizar os resultados obtidos por diferentes técnicas de investigação. Resultante das estruturas atômicas e supramoleculares, o espalhamento de raios X a baixo ângulo (Small Angle X-ray Scattering, SAXS) por soluções de proteínas pode conectar informações estruturais fornecidas por técnicas com diferentes níveis de resolução, como cristalografia de proteínas, ressonância magnética nuclear e microscopia eletrônica. Um outro atrativo da técnica de SAXS é que esses estudos são feitos com a proteína em condições mais próximas do seu estado nativo (soluções, emulsões, géis, etc.).

Novas abordagens teóricas tem sido desenvolvidas no tratamento e interpretação de dados de espalhamento a baixo ângulo permitindo a análise detalhada da estrutura global das proteínas, como determinação direta do envelope molecular (Svergun; Stuhrmann, 1991), predição de estruturas quaternárias em oligômeros (Svergun, 1994) e verificação de possíveis diferenças entre as estruturas moleculares no estado cristalino e em solução (Svergun; Barberato; Koch, 1995).

Os objetivos principais deste trabalho são a implementação das novas abordagens teóricas na forma de programas de computador e a aplicação destes no estudo estrutural de algumas proteínas em solução.

CAPÍTULO 2 - RAIOS X

Raios X são ondas eletromagnéticas cujo comprimento de onda varia entre 0,1 e 100 Å. Eles são usualmente produzidos pela rápida desaceleração de elétrons, que convertem parte de sua energia cinética em radiação eletromagnética. Em máquinas convencionais termo elétrons são gerados num filamento quente e acelerados em direção a um alvo de metal (cobre, molibdênio ou tungstênio). Um espectro contínuo de raios X é gerado no alvo pela rápida desaceleração que os elétrons sofrem ao colidir com os átomos do alvo. Certos elétrons entretanto, tem energia suficiente para arrancar elétrons que orbitam nas camadas mais internas desses átomos. A vacância nessas camadas são preenchidas pela transição de elétrons das camadas superiores e a diferença de energia potencial provocada por esta mudança de orbital é convertida em forma de radiação eletromagnética. Como as energias das camadas são bem definidas, essa radiação aparece na forma de picos no espectro de radiação (figura 1).

Figura 2.1: Espectro típico de radiação de raios X emitido por um tubo de cobre.

Devido às baixas intensidades obtidas por estes processos, ocorre atualmente um uso crescente de raios X gerados em aceleradores de partículas. Estas máquinas foram originalmente desenvolvidas para a pesquisa de física de altas energias. Nestas máquinas feixes de partículas subatômicas são aceleradas em anéis circulares a velocidades próximas a da luz. A geração de radiação eletromagnética de altíssima intensidade (radiação síncrotron) nos ímãs de deflexão do feixe era vista como um processo pernicioso de perda de energia pois impunha um limite para a velocidade das partículas. Portanto, nos primeiros estágios a radiação síncrotron era tida como um produto parasita dos anéis de altas energias. Entretanto, devido às suas características únicas, a radiação síncrotron se tornou nos dias atuais em uma importante ferramenta de pesquisa em física, química e biologia. Para este fim anéis de armazenamento tem sido especialmente desenhados com parâmetros otimizados para a geração de radiação síncrotron, isto é, feixes contínuos de longa vida, fontes estáveis e dispositivos de inserção para produzir radiação com propriedades específicas. Algumas características especiais deste tipo de radiação são alta intensidade, boa colimação intrínseca, 100% de polarização no plano de órbita, radiação pulsada e eficiência energética de 4-12%.

2.1 - Absorção de raios X pela matéria

Quando um feixe de raios X passa através da matéria sua intensidade diminui. O enfraquecimento é resultado de dois efeitos: A mudança de direção do feixe incidente, provocada pelo fenômeno de espalhamento, e a absorção de fótons associada

à ionização dos átomos pelo efeito fotoelétrico. A absorção da intensidade I por um material amorfo é dada pela equação:

$$I = I_0 e^{-\mu x} \quad 2.1$$

onde I_0 é a intensidade incidente, $\mu = \sigma + \tau$ é o coeficiente de absorção total, σ é o coeficiente de espalhamento que é praticamente uma constante para quase todos os materiais:

$$\sigma \approx C_s \quad 2.2$$

e τ é o coeficiente de absorção linear dado aproximadamente por (Feigin; Svergun, 1987):

$$\tau \approx CZ\lambda^3 \quad 2.3$$

onde C varia muito pouco com o comprimento de onda (λ) e Z é o número atômico. Esta equação mostra que a absorção cresce rapidamente com o aumento do número atômico e do comprimento da onda incidente. Estas equações impõem limites para o uso de raios X na pesquisa em materiais. A absorção do feixe incidente em estudos de raios X é determinada principalmente por τ que é significativamente maior que σ . Para efeitos práticos, os comprimentos de onda de raios X comumente utilizados variam de 0,5 a 2,0 Å.

CAPÍTULO 3 - PRINCÍPIOS GERAIS DO ESPALHAMENTO DE RAIOS X A BAIXO ÂNGULO

Este e os próximos capítulos irão introduzir os conceitos básicos para a análise de curvas de espalhamento de raios X a baixo ângulo por sistemas monodispersos. No começo do desenvolvimento desta técnica somente informações unidimensionais, como raio de giro, volume e superfície, eram extraídos de um experimento de SAXS (Guinier; Fournet, 1955). Como será mostrado na seção 3.8 uma curva de SAXS pode conter mais informações do que apenas estes parâmetros. Com o desenvolvimento das técnicas de coleta e análise de dados, modelos tridimensionais simples, como elipsóides, tubos e discos, foram sendo associados às partículas em solução (Glatter; Kratky, 1982). Recentes desenvolvimentos, como a teoria de multipolos (Stuhrmann, 1970ab, Svergun; Stuhrmann, 1991), permitem explorar ainda mais as informações obtidas num experimento de SAXS e obter diretamente a partir destas o envelope molecular de macromoléculas em solução.

3.1 - Introdução

Em um experimento típico de SAXS uma solução de proteínas é atravessada por um feixe de raios X. Devido a diferença de densidade eletrônica entre a proteína e o solvente parte do feixe incidente é espalhado. Por causa da orientação randômica das proteínas em solução a intensidade do feixe espalhado depende apenas do ângulo entre este e o feixe incidente. O produto final deste experimento é uma curva de intensidade

versus o ângulo de espalhamento (ou outra unidade equivalente) (figura 3.1). Devido à lei da reciprocidade, que apresenta uma relação inversa entre o ângulo de espalhamento e o tamanho da partícula, e o comprimento de onda dos raios X comumente usados ($0,5-2,0 \text{ \AA}$) o ângulo máximo de espalhamento de raios X observado por este sistema é bem pequeno ($\approx 5^\circ$). Decorre deste pequeno ângulo de espalhamento dois fatos interligados: 1) Podemos obter apenas informações dos aspectos gerais da partícula em solução, como forma, raio de giro e volume; 2) No entanto, podemos tratar esta partícula, para a maioria dos casos, como se ela tivesse uma distribuição de densidade eletrônica uniforme.

Figura 3.1: Experimento típico de SAXS.

3.2 - Espalhamento elástico de uma onda plana por uma densidade eletrônica

O processo de espalhamento pode ser visualizado da seguinte maneira: Um feixe de raios X atravessando uma partícula induz os elétrons desta a oscilar e emitir ondas secundárias com a mesma frequência do feixe incidente. A onda espalhada resultante será a soma de todas as ondas secundárias emitida pelos elétrons.

Figura 3.2: Feixe de raios X incidente e espalhado por uma densidade de cargas.

Figura 3.3: Geometria do espalhamento de raios X.

A figura 3.2 mostra um feixe de raios X incidindo e sendo espalhado por uma densidade de cargas $\rho(\vec{r})$, que é definida de forma que a sua integral sobre o volume ocupado por um átomo seja igual ao número total de elétrons (Ne) neste átomo:

$$Ne = \int \rho(\vec{r}) dv_{\text{atomo}} \quad 3.1$$

Na figura 3.2 o ponto O é tomado como origem, $\vec{r}$ é o vetor que vai da origem até o ponto P , α é o ângulo entre $\vec{r}$ e o feixe incidente e β é o ângulo entre $\vec{r}$ e o feixe espalhado. A diferença de fase entre a onda espalhada em P e a em O é dada por:

$$\frac{2\pi}{\lambda} r(\cos\beta - \cos\alpha) \quad 3.2$$

onde λ é o comprimento da onda da radiação incidente. A amplitude da onda espalhada em P por um elemento de volume dv relativamente a uma onda espalhada por um eletrão na origem é dada por:

$$dA = (\rho(\vec{r}) dv) \exp(i\vec{r} \cdot \vec{s}) \quad 3.3$$

onde dA é o diferencial da amplitude espalhada, $\vec{s} = \vec{S} - \vec{S}_0$ ($s = 4\pi\sin(\theta)/\lambda$) é o vetor de espalhamento, $\vec{S}$ e $\vec{S}_0$ são vetores de módulo $2\pi/\lambda$ e com a direção e o sentido do feixe espalhado e incidente respectivamente (figura 3.3).

A amplitude total, $A(\vec{s})$, da onda espalhada será então:

$$A(\vec{s}) = \int \rho(\vec{r}) \exp(i\vec{r} \cdot \vec{s}) dv \quad 3.4$$

que é a transformada de Fourier da densidade eletrônica $\rho(\vec{r})$. A solução do problema inverso, cálculo de $\rho(\vec{r})$ a partir de $A(\vec{s})$, é dado simplesmente pela transformada de Fourier inversa:

$$\rho(\vec{r}) = \frac{1}{2\pi^2} \int A(\vec{s}) \exp(-i\vec{r} \cdot \vec{s}) dV_s \quad 3.5$$

Onde dV_s denota o elemento de volume do espaço recíproco (o espaço recíproco é definido pelo vetor de espalhamento $\vec{s}$). No entanto, podemos medir em um experimento de espalhamento de raios X somente o fluxo de energia do feixe espalhado, ou seja, a intensidade $I(\vec{s})$ que é proporcional ao quadrado da amplitude $A(\vec{s})$:

$$I(\vec{s}) \propto A(\vec{s}) \cdot A^*(\vec{s}) = |A(\vec{s})|^2 \quad 3.6$$

A intensidade medida também é proporcional à intensidade do feixe incidente (I_0), porém omitiremos este parâmetro para efeito de simplificação das equações. A obtenção de $\rho(\vec{r})$ a partir somente do módulo da amplitude é o objetivo principal da cristalografia.

3.3 - Intensidade de uma partícula fixa

A intensidade de uma partícula com uma orientação fixa no espaço é dada pela combinação das equações 3.4 e 3.6:

$$I(\vec{s}) = |A(\vec{s})|^2 = \iint \rho(\vec{r}_1) \rho(\vec{r}_2) \exp(i\vec{s}(\vec{r}_1 - \vec{r}_2)) d\vec{r}_1 d\vec{r}_2 \quad 3.7$$

Esta é também uma transformada de Fourier envolvendo somente a distância relativa $(\vec{r}_1 - \vec{r}_2)$ para cada par de pontos $\vec{r}_1$ e $\vec{r}_2$. É conveniente portanto integrar a equação 3.7 em dois passos. O primeiro passo é integrar todos os pares $\vec{r}_1$ e $\vec{r}_2$ de forma que $\vec{r}_1 - \vec{r}_2 = \text{constante} = \vec{r}$, e depois integrar sobre $\vec{r}$.

A primeira operação é chamada de auto-correlação e definida por:

$$P(\vec{r}) = \int \rho(\vec{r}_1) \rho(\vec{r}_1 + \vec{r}) d\vec{r}_1 \quad 3.8$$

Esta equação é denominada de função de auto-correlação eletrônica e é também conhecida em cristalografia como função de Patterson (1935). Como definida pela equação 3.8 a função $P(\vec{r})$ corresponde a integração da multiplicação entre a densidade eletrônica e a mesma densidade deslocada pelo vetor $\vec{r}$. Seguem desta definição algumas propriedades. Como mostrado na figura 3.4a, a função $P(\vec{r})$ alcançará seu valor máximo para $\vec{r} = 0$, que é quando as duas densidades eletrônicas se sobrepõem completamente, e seu valor mínimo:

$$P(\vec{r}) = 0 \quad , \text{ para } \vec{r} \geq D \quad 3.9$$

onde D é a máxima dimensão da partícula (figura 3.4b). A figura 3.4c mostra também que $P(\vec{r}) = P(-\vec{r})$, ou seja, a função de auto-correlação é centrossimétrica.

O segundo passo para a avaliação de 3.7 é a integração sobre $\vec{r}$:

$$I(\vec{s}) = \int P(\vec{r}) \exp(i\vec{s} \cdot \vec{r}) d\vec{r} \quad 3.10$$

que é outra transformada de Fourier. Como $P(\vec{r})$ é uma função par podemos simplificar a equação acima e escrever:

$$I(\vec{s}) = \int P(\vec{r}) \cos(\vec{s} \cdot \vec{r}) d\vec{r} \quad 3.11$$

Figura 3.4: Nesta figura a densidade eletrônica deslocada é representada por linhas horizontais enquanto que a densidade fixa é representada por linhas verticais. a) Máximo valor para a função de Patterson ($\vec{r} = 0$). b) Mínimo valor ($\vec{r} > D$). c) A função de Patterson é centrossimétrica.

3.4 - Intensidade de uma partícula em solução

Dois fatos devem ser levados em conta no estudo de partículas em solução. O primeiro deles é que a densidade eletrônica $\rho(\vec{r})$ deve ser subtraída da densidade

eletrônica média do solvente, ρ_0 , e o segundo é que as partículas estão randomicamente orientadas em solução. Portanto a intensidade medida num experimento de espalhamento de raios X será a média das intensidades para todas as direções possíveis que a partícula pode assumir (promediação espacial):

$$\begin{aligned}
 I(s) &= \langle I(\vec{s}) \rangle_{\Omega} = \frac{1}{4\pi} \int_0^{4\pi} I(\vec{s}) d\Omega = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} I(\vec{s}) \sin\Phi d\Phi d\Psi \\
 &= \frac{1}{4\pi} \int_0^{4\pi} \left(\int_0^{4\pi} \int_0^{\infty} P(\vec{r}) \cos(\vec{s} \cdot \vec{r}) r^2 dr d\omega \right) d\Omega \\
 &= \int_0^{\infty} \left(\int_0^{4\pi} P(\vec{r}) d\omega \right) \left(\frac{1}{4\pi} \int_0^{4\pi} \cos(sr \cos\Phi) d\Omega \right) r^2 dr \\
 I(s) &= 4\pi \int_0^D \gamma(r) \frac{\sin(sr)}{sr} r^2 dr
 \end{aligned} \tag{3.12}$$

onde $\langle \rangle$ significa a promediação espacial, $\vec{s} = (s, \Omega) = (s, \Phi, \Psi)$ são as coordenadas do vetor $\vec{s}$ no espaço recíproco, $\vec{r} = (r, \omega) = (r, \theta, \phi)$ são as coordenadas do vetor $\vec{r}$ no espaço real, $d\Omega = \sin\Phi d\Phi d\Psi$ é o elemento infinitesimal do ângulo sólido no espaço recíproco, $d\omega = \sin\theta d\theta d\phi$ é o elemento infinitesimal de área no espaço direto, D é a máxima dimensão da partícula ($P(\vec{r}) = 0$, para r maior que D) e $\gamma(r)$ é a função de correlação definida como a promediação espacial da função de auto-correlação:

$$\gamma(r) = \frac{1}{4\pi} \int_0^{4\pi} P(\vec{r}) d\omega \tag{3.13}$$

Podemos obter $\gamma(r)$ diretamente da curva $I(s)$ através da transformada inversa da equação 3.12:

$$\gamma(r) = \frac{1}{2\pi^2} \int_0^\infty I(s) \frac{\sin(sr)}{sr} s^2 ds \quad 3.14$$

A equações 3.12-3.14 são de grande importância na análise de dados de SAXS. A função de correlação pode ser calculada diretamente das intensidades medidas, equação 3.14, permitindo desta forma obter um conhecimento direto da estrutura da partícula. A função $\gamma(r)$, bem como a função distribuição de distâncias $p(r)$:

$$p(r) = r^2 \gamma(r) \quad 3.15$$

são amplamente utilizadas em SAXS, pois ambas dependem da geometria da partícula e expressam numericamente o conjunto de distâncias entre os átomos da mesma. Estas funções são especialmente ilustrativas para o caso de partículas uniformes.

3.5 - Partículas com densidade eletrônica uniforme em solução

Devido a baixa resolução desta técnica, somente as características mais gerais de uma partícula em solução podem ser obtidas através da análise de curvas de SAXS. Por esta razão pequenas alterações no posicionamento dos átomos não alterarão de

forma significativa a curva de SAXS. Portanto, para a maioria dos casos, podemos tratar a densidade eletrônica $\rho(\vec{r})$ como constante através da partícula.

A densidade eletrônica uniforme $\rho_u(\vec{r})$ de uma partícula uniforme pode ser expressa na forma:

$$\rho_u(\vec{r}) = \begin{cases} \rho & \text{se } \vec{r} \text{ estiver dentro da partícula} \\ 0 & \text{se } \vec{r} \text{ estiver fora da partícula} \end{cases} \quad 3.16$$

Para este caso é facilmente visto, equação 3.13, que:

$$\gamma(0) = \rho^2 V \quad 3.17$$

onde V é o volume da partícula. A comparação da equação acima com a equação 3.14 permite chegar a seguinte conclusão:

$$2\pi^2 \gamma(0) = \int_0^\infty I(s) s^2 ds = 2\pi^2 \rho^2 V = Q \quad 3.18$$

Esta função, conhecida como invariante de Porod, é proporcional à energia total espalhada e é muito usada no cálculo de parâmetros estruturais. Outra característica importante desta equação é que ela independe da forma da partícula espalhadora.

A função característica $\gamma_o(r)$ (Porod, 1951) é definida por:

$$\gamma_o = \frac{\gamma(r)}{\gamma(0)} \quad 3.19$$

Esta função depende somente da forma externa da partícula e representa a probabilidade de encontrar um ponto dentro da partícula a uma distância r de um outro ponto. Esta função assume o valor unitário para $r = 0$ e zero para $r > D$. Wilson (1949) deu uma construção geométrica muito simples para a função $\gamma_o(r)$. Se uma partícula homogênea de volume V é deslocada pelo vetor $\vec{r}$, a função $\gamma_o(r)$ será dada pela promediação espacial do volume em comum, V_c , entre a partícula original e a deslocada (Figura 3.5):

$$\gamma_o(r) = \frac{\langle V_c(\vec{r}) \rangle_\omega}{V} \quad 3.20$$

Figura 3.5: Conceito do volume comum (V_c).

Substituindo 3.19 na equação 3.12 teremos a intensidade espalhada para uma partícula uniforme:

$$I(s) = 4\pi\rho^2V \int_0^D \gamma_o(r) \frac{\text{sen}(sr)}{sr} r^2 dr \quad 3.21$$

3.6 - Comportamento assintótico da curva de Intensidade (Lei de Porod)

Consideraremos agora o comportamento de $I(s)$ para altos valores de s . Observa-se da equação 3.21 que $I(s)$ e $\gamma(r)r$ estão relacionados por um transformada de Fourier, portanto, a região de altos valores de s estão relacionados com baixos valores de r . Para pequenos valores de r a função característica $\gamma_o(r)$ pode ser aproximada por (Glatter; Kratky, 1982):

$$\gamma_o(r) \approx 1 - \frac{S}{4V}r + \dots \quad 3.22$$

onde S é a superfície total da partícula em solução. Substituindo esta equação na equação 3.21 temos:

$$I(s) = 4\pi\rho^2V \int_0^D \left(1 - \frac{S}{4V}r + \dots\right) \frac{\text{sen}(sr)}{sr} r^2 dr \quad 3.22$$

A equação acima pode ser resolvida por partes (Abregó, 1987) resultando em:

$$I(s) \approx \frac{2\pi}{s^4} \rho^2 S \quad s \rightarrow \infty \quad 3.23$$

Dessa forma, o comportamento assintótico da intensidade para altos valores de s fica relacionado com a superfície da partícula.

3.7 - Determinação dos parâmetros integrais de partículas em solução

Nesta seção as equações para as invariantes das curvas de SAXS serão obtidas. O termo invariantes referem-se aos parâmetros estruturais como volume, raio de giro e superfície, que podem ser diretamente obtidos a partir da curva da intensidade $I(s)$ e independem da forma da partícula.

3.7.1 - Volume e superfície

A intensidade $I(s)$ de uma partícula homogênea para $s = 0$ pode ser calculada diretamente da equação 3.7:

$$I(0) = \int \int \rho_u(\vec{r}_1) \rho_u(\vec{r}_2) dv_1 dv_2 = \rho^2 \int \int dv_1 dv_2 = \rho^2 V^2 \quad 3.24$$

Podemos então calcular facilmente o volume V fazendo uso do invariante Q (equação 3.18):

$$V = \frac{2\pi^2 I(0)}{Q} \quad 3.25$$

O volume assim calculado é conhecido por volume de Porod. O conhecimento da intensidade de SAXS, mesmo quando em escala arbitrária, é então suficiente para determinar o volume de partículas homogêneas em solução.

A função Q também pode ser utilizada para o cálculo da superfície S . Multiplicando a equação 3.23 por s^4 obtemos:

$$s^4 I(s) \underset{s \rightarrow \infty}{\approx} 2\pi\rho^2 S = c_4 \quad 3.26$$

portanto a superfície S da partícula pode obtida da seguinte forma:

$$\frac{S}{V} = \frac{\pi}{Q} \lim_{s \rightarrow \infty} [s^4 I(s)] = \frac{\pi}{Q} c_4 \quad 3.27$$

As equações 3.25 e 3.27 podem também ser aplicadas para partículas não-homogêneas suficientemente grandes (> 50 kDa).

3.7.2 - Raio de giro

O comportamento de $I(s)$ será agora analisado para valores baixos de s . Expandindo o termo $\sin(sr)/sr$ em termos de uma série de McLaurin obtemos:

$$\frac{\text{sen}(sr)}{sr} = 1 - \frac{(sr)^2}{6} + \frac{(sr)^4}{120} - \dots \quad 3.28$$

Substituindo apenas os dois primeiros termos da equação acima na equação 3.12 teremos nas vizinhanças de $s=0$:

$$I(s) = I(0) \left(1 - \frac{s^2 R_g^2}{3} \right) \quad 3.29$$

onde:

$$I(0) = 4\pi \int_0^D \gamma(r) r^2 dr \quad 3.30$$

e R_g é o raio de giro da partícula definido por:

$$R_g^2 = \frac{1}{2} \frac{\int_0^D \gamma(r) r^4 dr}{\int_0^D \gamma(r) r^2 dr} \quad 3.31$$

Uma outra maneira de se definir o raio de giro é como a raiz quadrada da distância quadrática média entre os elétrons da partícula e o centro de massa eletrônico

$\bar{r}_e$:

$$R_g^2 = \frac{1}{N_e V} \int \rho(\vec{r}) |\vec{r} - \vec{r}_e|^2 dv \quad 3.32$$

onde N_e é o número total de elétrons. Para uma partícula uniforme com a origem de coordenadas coincidindo com $\vec{r}_e$ a equação anterior se simplifica para:

$$R_g^2 = \frac{1}{V} \int |\vec{r}|^2 dv \quad 3.33$$

O raio de giro é então uma medida da distribuição das cargas em relação ao centro de massa eletrônico de uma partícula. Este parâmetro pode ser muito útil na investigação estrutural de uma partícula em solução, pois um aumento ou diminuição do raio de giro de uma partícula provocado por qualquer variação nas condições iniciais da solução indicaria uma mudança nas posições relativas entre os átomos que compõem a partícula (mudança conformacional).

A expressão do lado direito da equação 3.29 pode ser considerado como os primeiros termos da série de McLaurin para a função $\exp(-s^2 R_g^2 / 3)$. Podemos então escrever para o começo da curva:

$$I(s) = I(0) \exp(-s^2 R_g^2 / 3) \quad 3.34$$

Esta é a equação de Guinier e foi derivada a quase 60 anos atrás (Guinier, 1939), e pode ser aplicada para a maioria dos casos precisando correções somente para o caso de partículas altamente anisométricas.

3.8 - Informação contida em uma curva de SAXS

As seções anteriores mostraram alguns dos parâmetros estruturais possíveis de serem obtidos a partir de curvas de SAXS. Esta seção fará uma estimativa da quantidade de informação contida nestas curvas, ou mais precisamente, o número de parâmetros estruturais independentes da partícula que podem ser obtidos de uma dada curva de SAXS.

Figura 3.6: Representação da função $sI(s)$ segundo o teorema da amostragem.

A equação 3.14 mostra que a função $sI(s)$ é a imagem seno-Fourier da função $r\gamma(r)$, a qual difere de zero dentro do intervalo finito $0 \leq r \leq D$. Então, devido ao teorema da amostragem (Shannon; Weaver, 1949), é válida para a função $sI(s)$ a seguinte expansão:

$$\begin{aligned}
 sI(s) &= \sum_{k=1}^{\infty} s_k I(s_k) \left\{ \frac{\text{sen}[D(s - s_k)]}{D(s - s_k)} - \frac{\text{sen}[D(s + s_k)]}{D(s + s_k)} \right\} \\
 &= \sum_{k=1}^{\infty} s_k I(s_k) \phi(s, k, D)
 \end{aligned} \tag{3.35}$$

onde $s_k = k\pi/D$. Isto significa que a intensidade pode ser inteiramente definida pelos valores de $I(s)$ nos pontos discretos s_k , $k = 1, 2, 3, \dots$. A figura 3.6 ilustra esquematicamente a função $sI(s)$ e vários termos da série.

A equação acima impõem um valor mínimo para o intervalo de amostragem (Δs), pois a medida de curvas de SAXS com a condição $\Delta s \geq \pi/D$ acarretaria em perdas de informação (Damaschun *et al.*, 1968). Na prática um menor intervalo Δs é recomendável devido aos erros experimentais.

A equação 3.35 não define intrinsicamente o número de parâmetros independentes que descrevem a função $I(s)$, desde que a série é infinita. Entretanto, devido que o comportamento de $sI(s)$ nas redondezas de s_k é determinado principalmente pelo k -ésimo termo da série (figura 3.6), podemos aproximar a equação 3.35 por:

$$sI(s) \approx \sum_{k=k_{\min}}^{k_{\max}} s_k I(s_k) \phi(s, k, D) \tag{3.36}$$

onde $k_{\min} = \text{inteiro}[s_{\min} D/\pi] + 1$, $k_{\max} = \text{inteiro}[s_{\max} D/\pi]$, e $s_{\min}$ e $s_{\max}$ determinam o intervalo de $I(s)$ medido experimentalmente.

O número suficiente de parâmetros independentes para descrever $I(s)$ no intervalo $[s_{min}, s_{max}]$ é, desprezando-se os efeitos terminais da truncagem, dado pelo número de “pontos básicos” s_k pertencentes ao intervalo.

É natural concluir então que este número, $n = k_{max} - k_{min}$, é igual ao número de parâmetros estruturais independentes da partícula que podem ser avaliados de uma curva de SAXS. O problema principal com esta afirmação está no problema inverso, isto é, descobrir até que ponto parâmetros estruturais como volume, raio de giro e superfície são independentes entre si.

CAPÍTULO 4 - ELIPSÓIDE EQUIVALENTE

Como foi visto na seção 3.8 uma curva de SAXS pode conter mais informações além daquelas dadas pelos parâmetros integrais (Volume, superfície e R_g). Por esta razão vários métodos tem sido desenvolvidos para extrair mais dados estruturais de partículas em solução a partir destas curvas. Dentre estes podemos citar *O Elipsóide Equivalente de Inércia* (Müller; Schrauber, 1992), *Corpos de Simetria Simples* (Mittelbach; Porod, 1961, 1962), *Agregados de Pequenas Esferas* (Rolbin; Feigin; Scedrin, 1971) e *Grupo de Átomos Normalizados* (Müller, 1983). Em geral, estes métodos baseiam-se em procedimentos de “tentativa-e-erro” e são utilizados para a determinação de modelos de baixa resolução para a partícula em solução.

Definimos como elipsóide equivalente o elipsóide tri-axial, descrito pelos semi-eixos a , b e c , cuja curva de SAXS melhor se ajusta a uma dada curva experimental.

4.1 - *ELLFIT* - Um programa para a avaliação do elipsóide equivalente

Elipsóides tri-axiais podem ser usados como uma primeira aproximação para modelos de baixa resolução para a forma de partículas em solução. Naturalmente a quantidade de informação estrutural obtida pela associação da forma da partícula com um elipsóide é bastante reduzida e, mesmo assim, essa associação só é válida para o caso de partículas com peso molecular maior que 50 kDa (quando a partícula pode ser considerada homogênea) e cuja forma seja aproximadamente globular. Nestes casos

pode se obter a dimensão máxima e algumas características básicas do formato da partícula, isto é, se a partícula apresenta uma forma mais alongada ou achatada.

O elipsóide equivalente pode ser obtido pela minimização do seguinte funcional:

$$\Phi^2(a, b, c, S_{fac}) = \frac{\sum_{i=1}^N [S_{fac} I_{exp}(s) - I_{elip}(a, b, c)]^2 W(s)}{\sum_{i=1}^N [S_{fac} I_{exp}(s)]^2 W(s)} \quad 4.1$$

onde a, b e c são os semi-eixos do elipsóide equivalente, S_{fac} é o fator de escala entre os dados experimentais e teóricos, N é número de pontos da curva, $I_{exp}(s)$ e $I_{elip}(s)$ são as curvas de SAXS experimental e teórica, respectivamente e $W(s)$ é uma função peso que é tomada como o inverso do quadrado dos erros experimentais $\sigma_{exp}(s)$:

$$W(s) = \left(\frac{1}{\sigma_{exp}(s)} \right)^2 \quad 4.2$$

A curva de SAXS de um elipsóide tri-axial (a, b, c) é dado por (Feigin; Svergun, 1987):

$$I_{elip}(s, a, b, c) = \int \int 9 \left(\frac{\sin(sR) - sR \cos(sR)}{(sR)^3} \right)^2 dx dy \quad 4.3$$

onde

$$R(a,b,c,x,y) = \sqrt{(a^2 \cos^2(1/2\pi x) + b^2 \sin^2(1/2\pi y))(1-y^2) + c^2 y^2} \quad 4.4$$

Figura 4.1: Ajustes entre curvas experimentais (pontos) e elipsóides equivalente (linhas contínuas). a) Lisozima, b) Inibidora da α -amilase, c) hexocinase e d) nitrose redutase. Descrições a respeito da obtenção das curvas para a lisozima, hexocinase e da inibidora da α -amilase serão dadas nas seções posteriores. A curva da nitrato redutase foi gentileza do Dr. Grossmann (Grossmann *et al.*, 1991).

*ELLFIT*¹ é um programa que encontra o elipsóide equivalente cuja curva de SAXS melhor se ajusta a uma dada curva de SAXS. O programa utiliza as equações acima descritas obtendo o elipsóide equivalente através dos seguintes passos:

- 1- A curva experimental de SAXS é lida de um arquivo.

¹ *ELLFIT* (Barberato, Svergun e Volkov, 1994, não publicado)

2- O raio de giro da partícula é avaliado usando a aproximação de Guinier.

Também usando esta aproximação o começo da curva experimental é extrapolada até $s = 0$.

3- Usando a curva extrapolada o volume é avaliado através da equação 3.25.

4- O intervalo de minimização ($s_{\min}$ - $s_{\max}$) é determinado de forma a conter pelo menos 3 parâmetros independentes segundo o critério desenvolvido na seção 3.8 ($s_{\max} = 3\pi/D$). A dimensão máxima é aproximada por $D = 3Rg$.

5- O funcional 4.1 é minimizado dentro do intervalo ($s_{\min}$ - $s_{\max}$) utilizando o algoritmo de Dennis, Gay e Welsch (1980).

Os resultados da aplicação deste programa em 4 proteínas encontram-se ilustrados nas figuras 4.1a,b,c e d e nas tabelas 4.1 e 4.2. Uma análise qualitativa dos gráficos apresentados na figura 4.1 confirma que o ajuste entre os pontos experimentais e curvas teóricas tende a melhorar com o aumento do peso molecular da proteína.

Tabela 4.1: Comparação entre as dimensões máximas de algumas proteínas (x,y,z) e os elipsóides equivalentes obtidos com o programa *ELLFIT* (a,b,c).

Proteína	Peso Molecular (kDa)	x (Å)	a(Å)	y (Å)	b (Å)	z (Å)	c (Å)
lisozima	14,3	45	40	30	40	30	20
ini. α -amilase	52,0	88*	97	44*	40	35*	37
hexocinase	104,4	110	100	60	100	60	36
nitrato red.	103,0	75	84	65	77	65	51

Obs: As dimensões máximas foram estimadas a partir das estruturas cristalográficas com exceção da inibidora de α -amilase cujas dimensões máximas foram estimadas a partir do envelope molecular (para detalhes veja seção 6.3).

A lisozima é um caso bem ilustrativo da limitação dessa técnica para o caso de proteínas de baixo peso molecular pois, apesar de apresentar uma forma relativamente elipsoidal, o ajuste entre as curvas experimental e do elipsóide é bastante pobre. Esta falta de concordância entre as curvas se reflete nas discrepâncias encontradas pela comparação entre as distâncias máximas da proteína e os semi-eixos do elipsóide e os raios de giro (tabelas 4.1 e 4.2). Neste caso a influência da falta de homogeneidade da densidade eletrônica da lisozima no espalhamento de raios-X não pode ser desprezada.

Os resultados preliminares da cristalização da inibidora da α -amilase (Oliva, Iulek e Eda, 1996) assim como estudos feitos em solução (seção 6.3) indicam que o oligômero formado é uma partícula com uma forma alongada. Os resultados obtidos com *ELLFIT* reforçam esta suposição e concordam razoavelmente com as dimensões máximas obtidos a partir da determinação do envelope molecular (seção 6.3).

Tabela 4.2: Resultados obtidos utilizando o programa *ELLFIT*. S_{exp} é o intervalo angular experimental, S_{min} é o intervalo utilizado na minimização, R_{fac} é o fator de concordância entre curva experimental e a curva teórica para o intervalo em questão, $R_{g_{exp}}$ é o raio de giro experimental, $R_{g_{eli}}$ é o raio de giro do elipsóide, V_{por} é o volume de Porod e V_{eli} é o volume do elipsóide.

Proteína	$S_{exp} (\text{\AA}^{-1}) - R_{fac}$	$S_{min} (\text{\AA}^{-1}) - R_{fac}$	$R_{g_{exp}} (\text{\AA})$	$R_{g_{eli}} (\text{\AA})$	$V_{por} (\text{\AA}^3)$	$V_{eli} (\text{\AA}^3)$
lisozima	0,04 - 0,50 (0,30)	0,04 - 0,30 (0,15)	15,2	13,5	15.000	16.100
inb. α -amil.	0,03 - 0,25 (0,04)	0,03 - 0,17 (0,01)	24,7	24,9	74.200	47.300
hexocinase	0,02 - 0,30 (0,10)	0,02 - 0,13 (0,03)	32,1	32,4	163.000	186.000
nit. red.	0,02 - 0,24 (0,06)	0,02 - 0,14 (0,02)	29,5	28,0	177.000	173.000

Apesar do alto peso molecular da hexocinase a concordância entre as curvas experimental e teórica, assim como a comparação entre os valores para dimensões máximas obtidos com *ELLFIT*, não foram satisfatórios. Isto se deve principalmente ao fato da hexocinase apresentar um envelope molecular extremamente anisotrópico

(Steitz, 1976), violando portanto a condição de globularidade necessária para esta técnica. A nitrato redutase (Grossmann *et. al.*, 1991) por sua vez satisfaz as duas condições, ela é suficientemente grande (103 kDa) e razoavelmente globular. A aplicação do programa a sua curva de SAXS rende portanto um excelente ajuste entre as curvas e parâmetros dimensionais bem próximos daqueles obtidos por cristalografia (tabelas 4.1 e 4.2).

A aplicação de *ELLFIT* a curvas de SAXS pode render portanto algumas informações estruturais a respeito de uma proteína em solução. Para os casos favoráveis, isto é, quando ocorre uma boa concordância entre as curvas, pode-se obter valores aproximados das dimensões máximas de uma proteína, e nos casos desfavoráveis, isto é, quando, apesar de um alto peso molecular, as curvas não apresentam uma boa concordância pode-se concluir que a partícula em questão não é globular.

CAPÍTULO 5 - INTERPRETAÇÃO DO ESPALHAMENTO DE RAIOS-X POR PARTÍCULAS INOMOGÊNEAS

5.1 - Espalhamento de raios-X por uma partícula inhomogênea¹

Num experimento comum de SAXS partículas com amplitudes de espalhamento $A_v(\vec{s})$ são dissolvidas num solvente com densidade eletrônica média ρ_0 e então irradiadas por um feixe incidente. Como mostrado na figura 5.1 a densidade eletrônica de uma partícula em solução pode ser considerada como a sobreposição da densidade eletrônica da partícula no vácuo, do volume de solvente deslocado pela partícula (volume excluído) homogeneamente preenchido com uma densidade eletrônica $-\rho_0$ e do elemento de volume V_0' preenchido inteiramente com o solvente.

Figura 5.1: A densidade eletrônica de uma proteína, representada aqui por um conjunto de círculos conectados, em solução pode ser representada como a superposição de sua densidade eletrônica no vácuo, do volume de solvente deslocado por ela (volume excluído) e do volume irradiado pelo feixe incidente.

O volume médio ocupado por uma partícula em solução, V_0' é dado por:

¹ Inhomogêneo é uma palavra que não consta no vocabulário português. Trata-se de uma adaptação da palavra inglesa *inhomogeneous* que significa não-homogêneo.

$$V_0' = V_0 / N \quad 5.1$$

onde V_0 é o volume de solução irradiado pelo feixe incidente e N é o número de partículas neste volume.

A intensidade espalhada total será dada pela soma das amplitudes ao quadrado dos N conjuntos partícula + volume excluído + V_0' :

$$\begin{aligned} I(\vec{s}) &= \left| \sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s}) + A_{V_0'_i}(\vec{s})) \right|^2 = \left| \left(\sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s})) \right) + A_{V_0}(\vec{s}) \right|^2 \\ &= \left| \sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s})) \right|^2 + \\ &\quad \left(\sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s})) \right) A_{V_0}^*(\vec{s}) + A_{V_0}(\vec{s}) \left(\sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s})) \right)^* + |A_{V_0}(\vec{s})|^2 \end{aligned} \quad 5.2$$

onde $A_v(\vec{s})$ é a amplitude espalhada pela partícula no vácuo:

$$A_v(\vec{s}) = \int \rho(\vec{r}) \exp(-i\vec{s} \cdot \vec{r}) dv \quad 5.3$$

$A_s(\vec{s})$ é a amplitude espalhada pelo volume excluído:

$$A_s(\vec{s}) = -\rho_0 \int \rho_s(\vec{r}) \exp(-i\vec{s} \cdot \vec{r}) dv \quad 5.4$$

$\rho_s(\vec{r})$ é a função que descreve o volume excluído:

$$\rho_s(\vec{r}) = \begin{cases} 1 & \text{para } \vec{r} \text{ dentro da partícula} \\ 0 & \text{para } \vec{r} \text{ fora da partícula} \end{cases} \quad 5.5$$

$A_{V_0}(\vec{s})$ é a amplitude espalhada pelo volume de solvente irradiado:

$$A_{V_0}(\vec{s}) = \sum_i^N A_{V_{0j}}, = \rho_0 \int_{V_0} \rho_V(\vec{r}) \exp(-i\vec{s} \cdot \vec{r}) dV \quad 5.6$$

e $\rho_V(\vec{r})$ é a função que descreve o volume irradiado:

$$\rho_V(\vec{r}) = \begin{cases} 1 & \text{para } \vec{r} \text{ dentro do volume irradiado} \\ 0 & \text{para } \vec{r} \text{ fora do volume irradiado} \end{cases} \quad 5.7$$

Devido à lei da reciprocidade e às dimensões macroscópicas do feixe incidente, a amplitude espalhada pelo volume irradiado V_0 estará totalmente concentrada a baixíssimos valores de s quando comparadas à amplitude espalhada pelas partículas + volume excluídos. Portanto os termos cruzados $\left(\sum_{i=1}^N (A_{V_i}(\vec{s}) + A_{S_i}(\vec{s})) \right) A_{V_0}(\vec{s})$ podem ser omitidos e a equação 5.2 torna-se:

$$I(\vec{s}) \approx I_p(\vec{s}) + I_{V_0}(\vec{s}) \quad 5.8$$

onde:

$$I_p(\vec{s}) = \left| \sum_i^N (A_{v_i}(\vec{s}) + A_{s_i}(\vec{s})) \right|^2 \quad 5.9$$

é a intensidade espalhada somente pelo conjunto proteína + volume excluído e

$$I_{v_0}(\vec{s}) = \left| A_{v_0}(\vec{s}) \right|^2 \quad 5.10$$

é a intensidade espalhada pelo volume de solvente irradiado.

A intensidade $I_p(\vec{s})$, equação 5.9, pode então ser obtido experimentalmente através da subtração da intensidade de espalhamento da solução (partícula + solvente), $I(\vec{s})$, pela intensidade de espalhamento do solvente puro, $I_{v_0}(\vec{s})$.

Se as partículas em solução estão randomicamente orientadas e suficientemente espaçadas² os termos cruzados, $A_{v_i}(\vec{s})A_{s_j}(\vec{s})^*$ para $i \neq j$, desaparecem e a equação 5.9 simplifica-se para:

$$I_p(\vec{s}) = I_p(s) = N \left\langle \left| A_v(\vec{s}) + A_s(\vec{s}) \right|^2 \right\rangle_\Omega \quad 5.11$$

onde $\langle \rangle_\Omega$ significa a promediação espacial no espaço recíproco.

Desenvolvendo o quadrado da equação acima chegamos a:

² Existe a possibilidade de se obter uma orientação líquida de proteínas em solução através do uso de capilares extremamente finos e campos magnéticos (Koch, dados não publicados).

$$\begin{aligned}
 I_p(s) &= N \left\langle A_v(\vec{s}) A_v(\vec{s})^* + A_s(\vec{s}) A_v(\vec{s})^* + A_v(\vec{s}) A_s(\vec{s})^* + A_s(\vec{s}) A_s(\vec{s})^* \right\rangle_{\Omega} \\
 &= I_v(s) + \rho_0 I_{vs}(s) + \rho_0^2 I_s(s)
 \end{aligned}
 \tag{5.12}$$

onde:

$$\begin{aligned}
 I_v(s) &= N \left\langle A_v(\vec{s}) A_v(\vec{s})^* \right\rangle_{\Omega} \\
 &= N \left\langle \int \int \rho(\vec{r}_1) \rho(\vec{r}_2) \exp(i\vec{s} \cdot (\vec{r}_1 - \vec{r}_2)) d\vec{r}_1 d\vec{r}_2 \right\rangle_{\Omega}
 \end{aligned}
 \tag{5.13}$$

é a intensidade espalhada para uma partícula no vácuo,

$$I_{vs}(s) = \frac{1}{\rho_0} N \left\langle A_v(\vec{s}) A_s(\vec{s})^* + A_s(\vec{s}) A_v(\vec{s})^* \right\rangle_{\Omega}
 \tag{5.14}$$

é o termo cruzado e

$$\begin{aligned}
 I_s(s) &= \frac{1}{\rho_0^2} N \left\langle A_s(\vec{s}) A_s(\vec{s})^* \right\rangle_{\Omega} \\
 &= N \left\langle \int \int \rho_s(\vec{r}_1) \rho_s(\vec{r}_2) \exp(i\vec{s} \cdot (\vec{r}_1 - \vec{r}_2)) d\vec{r}_1 d\vec{r}_2 \right\rangle_{\Omega}
 \end{aligned}
 \tag{5.15}$$

é o espalhamento do volume excluído da partícula com uma densidade uniforme ρ_0 .

5.2 - Técnica da variação de contraste

Stuhrmann e Kirste (1967) propuseram uma forma mais conveniente de representar a densidade eletrônica de uma partícula em solução. Neste novo formalismo a função $\rho(\vec{r})$ é descrita como a soma de duas partes (figura 5.2):

$$\rho(\vec{r}) = \rho_f(\vec{r}) + \bar{\rho} \rho_s(\vec{r}) \quad 5.16$$

onde a função $\rho_f(\vec{r})$ descreve as flutuações da densidade eletrônica da partícula em torno da sua densidade eletrônica média $\bar{\rho}$:

$$\bar{\rho} = \frac{1}{V} \int \rho(\vec{r}) dv \quad 5.17$$

e $\rho_s(\vec{r})$ descreve o envelope da partícula exatamente como na equação 5.5.

Figura 5.2: Densidade eletrônica da proteína no vácuo como superposição das flutuações da densidade ($\rho_s(\vec{r}) = \rho(\vec{r}) - \bar{\rho}$) e da densidade média ($\bar{\rho}$).

Utilizando o mesmo procedimento da seção 5.5, a densidade eletrônica da partícula em solução será representada pela sobreposição da função $\rho_f(\vec{r})$, do volume excluído com uma densidade:

$$\Delta\rho = \bar{\rho} - \rho_0 \quad 5.18$$

e pelo volume V_0 (figura 5.3).

Figura 5.3: A densidade eletrônica de uma proteína em solução como a superposição das flutuações de densidade eletrônica, do volume excluído com densidade $\delta\rho$ e do volume irradiado pelo feixe incidente.

A diferença entre a densidade média da partícula e a densidade média do solvente, $\Delta\rho$, é definida como o contraste da solução. A intensidade espalhada por este sistema será:

$$I(\vec{s}) = \left| \sum_i^N (A_{f_i}(\vec{s}) + A_{s_i}(\vec{s}) + A_{V_{0_i}}(\vec{s})) \right|^2 \quad 5.19$$

onde $A_f(\vec{s})$ é a amplitude espalhada pelas flutuações de densidade eletrônica da partícula:

$$A_f(\vec{s}) = \int \rho_f(\vec{r}) \exp(-i\vec{s} \cdot \vec{r}) dV \quad 5.20$$

$A_s(\vec{s})$ é a amplitude espalhada pelo volume excluído com uma densidade $\Delta\rho$:

$$A_s(\vec{s}) = \Delta\rho \int \rho_s(\vec{r}) \exp(-i\vec{s} \cdot \vec{r}) dV \quad 5.21$$

onde $\rho_s(\vec{r})$ é a função que descreve o volume excluído (equação 5.5). Utilizando os mesmos procedimentos usados na seção 5.1 chegamos facilmente a uma equação para a intensidade espalhada $I_p(s)$:

$$I_p(s) = I_f(s) + \Delta\rho I_{fs}(s) + \Delta\rho^2 I_s(s) \quad 5.22$$

A função $I_p(s)$ fica dessa forma decomposta em três funções básicas, sendo $I_f(s)$ a intensidade provocada pelas inhomogeneidades da partícula, $I_s(s)$ a intensidade do volume excluído com densidade $\Delta\rho$ e o termo cruzado $I_{fs}(s)$.

Se pelo menos três curvas forem medidas a diferentes contrastes $\Delta\rho_i$, as funções básicas $I_p(s)$, $I_f(s)$ e $I_{fs}(s)$ podem ser obtidas através da resolução do seguinte sistema linear:

$$\begin{aligned}
 I_p(s_k)_1 &= I_f(s_k) + \Delta\rho_1 I_{fs}(s_k) + \Delta\rho_1^2 I_s(s_k) \\
 I_p(s_k)_2 &= I_f(s_k) + \Delta\rho_2 I_{fs}(s_k) + \Delta\rho_2^2 I_s(s_k) \\
 I_p(s_k)_3 &= I_f(s_k) + \Delta\rho_3 I_{fs}(s_k) + \Delta\rho_3^2 I_s(s_k)
 \end{aligned}
 \tag{5.23}$$

As variações no contraste podem ser obtidas pela adição de sucrose na solução de proteína ou água deuterada nas experiências com difração de neutrons.

A rigor ambas equações 5.12 e 5.23 podem ser usadas para se obter as equações básicas, porém, a equação 5.23 apresenta a vantagem de que $I_f(s)$ e $I_{fs}(s)$ são ambas iguais a zero na origem (equação 3.24 para $\rho = 0$), e esta informação adicional pode ser usada para corrigir eventuais erros de avaliação do contraste $\Delta\rho_i$.

CAPÍTULO 6 - DETERMINAÇÃO DO ENVELOPE MOLECULAR ATRAVÉS DA TEORIA DE MULTIPOLos

Uma curva de SAXS geralmente contém mais informações estruturais além daquelas reveladas pelo elipsóide equivalente. Neste capítulo iremos descrever um método direto (Stuhrmann, 1970ab) que utiliza toda a curva de SAXS para a obtenção do envelope molecular de uma partícula em solução. Devido a limitada quantidade de informações contida nestas curvas este envelope não poderá conter detalhes a nível de resolução atômico, limitando-se a resoluções não melhores que aproximadamente 20 Å.

A seção seguinte fará uma pequena introdução aos harmônicos esféricos devido à extensa utilização destes na teoria de multipolos.

6.1 - Harmônicos esféricos

Harmônicos esféricos ocorrem numa grande variedade de problemas físicos sempre que uma equação de onda:

$$\nabla^2 \psi = \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} \quad 6.1$$

é resolvida por separação de variáveis em coordenadas esféricas. O harmônico esférico $Y_{lm}(\theta, \varphi)$, $-l \leq m \leq l$, é uma função de duas coordenadas θ e φ na superfície de uma

esfera. Os harmônicos esféricos são funções ortogonais para valores diferentes de l e m , são normalizados de tal forma que a integral sobre a esfera seja a unidade:

$$\left\langle Y_{lm}(\theta,\varphi) Y_{kl}^*(\theta,\varphi) \right\rangle_{\omega} = \int Y_{lm}(\theta,\varphi) Y_{kl}^*(\theta,\varphi) d\omega = \delta_{lk} \delta_{ml} \tag{6.2}$$

e satisfazem a seguinte relação:

$$Y_{l,-m}(\theta,\varphi) = (-1)^m Y_{lm}^*(\theta,\varphi) \tag{6.3}$$

O primeiros harmônicos esféricos normalizados estão tabelados abaixo:

Tabela 6.1: Primeiros termos dos esféricos harmônicos.

$Y_{00} = \sqrt{\frac{1}{4\pi}}$
$Y_{11} = \sqrt{\frac{3}{8\pi}} \operatorname{sen}\theta \exp(i\varphi)$ $Y_{10} = \sqrt{\frac{3}{4\pi}} \cos\theta$
$Y_{22} = \frac{1}{4} \sqrt{\frac{15}{2\pi}} \operatorname{sen}^2\theta \exp(2i\varphi)$ $Y_{21} = -\sqrt{\frac{15}{8\pi}} \operatorname{sen}\theta \cos\theta \exp(i\varphi)$ $Y_{20} = \sqrt{\frac{5}{4\pi}} \left(\frac{3}{4} \cos^2\theta - \frac{1}{2} \right)$

6.2 - Expansão de multipolos

Transformadas de Fourier convencionais não são muito apropriadas para o desenvolvimento de métodos diretos em SAXS. Somente para os casos de partículas muito simples (esferas, paralelepípedos, elipsóides...) o conceito de que a amplitude espalhada é a transformada de Fourier da densidade eletrônica pode ser usada. Este capítulo introduzirá uma nova abordagem para a descrição do espalhamento de raios X a baixo ângulo. Um método que usava harmônicos esféricos no problema de espalhamento de raios X foi primeiramente criado por Harrison (1969) e mais tarde desenvolvido compreensivelmente por Stuhrmann (1970a). Este método tem se mostrado de grande importância na teoria de SAXS e em particular na determinação dos envelopes moleculares de partículas em solução.

Qualquer densidade eletrônica tridimensional $\rho(\vec{r})$ pode ser representada por uma série da forma:

$$\rho(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \rho_{lm}(r) Y_{lm}(\omega) \quad 6.4$$

onde:

$$\rho_{lm}(r) = \int \rho(\vec{r}) Y_{lm}^*(\omega) d\omega \quad 6.5$$

são as funções radiais. No espaço recíproco a amplitude espalhada, $A(\vec{s})$, também pode ser expressa da mesma forma:

$$A(\vec{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}(s) Y_{lm}(\Omega) \quad 6.6$$

onde:

$$A_{lm}(s) = \int A(\vec{s}) Y_{lm}^*(\Omega) d\Omega \quad 6.7$$

são as amplitudes parciais. Por outro lado, a amplitude espalhada pode ser escrita como a transformada de Fourier da densidade eletrônica $\rho(\vec{r})$ (equação 3.4):

$$A(\vec{s}) = (2\pi)^{-3/2} \int \rho(\vec{r}) \exp(i\vec{s} \cdot \vec{r}) dv \quad 6.8$$

onde o termo $(2\pi)^{-3/2}$ foi empregado para efeito de simetrização entre o espaço recíproco e o real. Dessa forma a densidade eletrônica $\rho(\vec{r})$ é escrita como:

$$\rho(\vec{r}) = (2\pi)^{-3/2} \int A(\vec{s}) \exp(-i\vec{s} \cdot \vec{r}) dv_s \quad 6.9$$

Substituindo a seguinte relação (Edmonds, 1957):

$$\exp(i\vec{s} \cdot \vec{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l i^l j_l(sr) Y_{lm}(\Omega) Y_{lm}^*(\omega) \quad 6.10$$

onde $j_l(sr)$ são as funções esféricas de Bessel:

$$j_l(z) = z^l \left(-\frac{1}{z} \frac{d}{dz} \right)^l \frac{\sin(z)}{z} \quad 6.11$$

e a equação 6.4 na equação 6.6 temos:

$$\begin{aligned} A(\vec{s}) &= (2\pi)^{-3/2} \int_0^{\infty} \left(\sum_{l=0}^{\infty} \sum_{m=-l}^l \rho_{lm}(r) Y_{lm}(\omega) \right) \left(4\pi \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p j_p(sr) Y_{pq}(\Omega) Y_{pq}^*(\omega) \right) dv \\ &= \sqrt{\frac{2}{\pi}} \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p \int \rho_{lm}(r) j_p(sr) r^2 dr Y_{pq}(\Omega) \int Y_{lm}(\omega) Y_{pq}^*(\omega) d\omega \end{aligned} \quad 6.12$$

Usando a relação 6.2 chegamos a:

$$\begin{aligned} A(\vec{s}) &= \sqrt{\frac{2}{\pi}} \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p \int_0^{\infty} \rho_{lm}(r) j_p(sr) r^2 dr Y_{pq}(\Omega) \delta_{lp,mq} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l \sqrt{\frac{2}{\pi}} i^l \int \rho_{lm}(r) j_l(sr) r^2 dr Y_{lm}(\Omega) \end{aligned} \quad 6.13$$

Comparando a equação acima com a equação 6.6 concluímos que:

$$A_{lm}(s) = \sqrt{\frac{2}{\pi}} i^l \int \rho_{lm}(r) j_l(sr) r^2 dr \quad 6.14$$

isto é, as amplitudes parciais $A_{lm}(s)$ são as transformadas de Hankel das funções radiais $\rho_{lm}(r)$. A transformação inversa:

$$\rho_{lm}(r) = \sqrt{\frac{2}{\pi}} (-i)^l \int A_{lm}(s) j_l(sr) s^2 ds \quad 6.15$$

pode ser derivada da mesma forma.

Substituindo a equação 6.6 na equação 3.12 obtemos uma fórmula simples para o cálculo da intensidade $I(s)$:

$$\begin{aligned} I(s) &= \left\langle A(\vec{s}) A^*(\vec{s}) \right\rangle_{\Omega} = \left\langle \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}(s) Y_{lm}(\omega) \sum_{p=0}^{\infty} \sum_{q=-p}^p A_{pq}^*(s) Y_{pq}^*(\omega) \right\rangle_{\Omega} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{p=0}^{\infty} \sum_{q=-p}^p A_{lm}(s) A_{pq}^*(s) \left\langle Y_{lm}(\omega) Y_{pq}^*(\omega) \right\rangle_{\Omega} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{p=0}^{\infty} \sum_{q=-p}^p A_{lm}(s) A_{pq}^*(s) \delta_{lp,mq} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}(s)|^2 \end{aligned} \quad 6.16$$

O principal objetivo da interpretação dos dados de SAXS usando a expansão de multipolos é distinguir entre as diferentes contribuições dos multipolos para $I(s)$ as quais permitiriam restaurar a densidade eletrônica $\rho(\vec{r})$. Esta tarefa é obviamente impossível dado que o número de parâmetros obtidos a partir de uma curva de SAXS, seção 3.8, é muito inferior ao número de parâmetros necessários para representação de $\rho(\vec{r})$. Faz-se necessário então fornecer informações adicionais a respeito da partícula

de modo a diminuir o número dos parâmetros necessários para a representação de $\rho(\vec{r})$. Uma das mais importantes restrições é a atribuição de uma densidade eletrônica homogênea (ρ_0) a $\rho(\vec{r})$. Outra restrição também muito importante é a respeito da forma da partícula. Se esta for aproximadamente globular podemos representa-la usando a função angular de forma $F(\omega)$ que descreve a superfície da partícula. Com estas duas restrições a densidade eletrônica fica representada da seguinte forma:

$$\rho(\vec{r}) = \begin{cases} \rho_0, & \text{se } 0 \leq \vec{r} < F(\omega) \\ 0, & \text{se } \vec{r} \geq F(\omega) \end{cases} \quad 6.17$$

Esta função também pode ser representada por uma série de harmônicos esféricos:

$$F(\omega) \approx F_L(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega) \quad 6.18$$

onde L determina a resolução da representação da partícula e f_{lm} são os coeficientes dos multipolos dados por:

$$f_{lm} = \int F(\omega) Y_{lm}^*(\omega) d\omega \quad 6.19$$

Podemos representar as funções esféricas de Bessel como uma série de potências:

$$j_l(sr) = \sum_{p=0}^P d_{lp}(sr)^{l+2p} \quad 6.20$$

onde P é o máximo número de elementos na série e

$$d_{lp} = \frac{(-1)^p}{2^p p! [2(l+p) + 1]!!} \quad 6.21$$

Substituindo esta série e a equação 6.5 em 6.14, obtemos:

$$\begin{aligned} A_{lm}(s) &= \sqrt{\frac{2}{\pi}} i^l \int_0^{F(\omega)} \left(\int \rho(\vec{r}) Y_{lm}^*(\omega) d\omega \right) \left(\sum_{p=0}^P d_{lp} (sr)^{l+2p} \right) r^2 dr \\ &= \sqrt{\frac{2}{\pi}} i^l \rho_0 \sum_{p=0}^P d_{lp} s^{l+2p} \int_0^{F(\omega)} \left(\int r^{l+2p+2} dr \right) Y_{lm}^*(\omega) d\omega \\ &= \sqrt{\frac{2}{\pi}} i^l \rho_0 \sum_{p=0}^P d_{lp} s^{l+2p} \int \frac{F(\omega)^{l+2p+3}}{(l+2p+3)} Y_{lm}^*(\omega) d\omega \\ &= \sqrt{\frac{2}{\pi}} i^l \rho_0 \sum_{p=0}^P \frac{d_{lp}}{(l+2p+3)} f_{lm}^q s^{l+2p} \end{aligned} \quad 6.22$$

onde:

$$f_{lm}^q = \int [F(\omega)]^q Y_{lm}^*(\omega) d\omega \quad 6.23$$

Dessa forma a amplitude espalhada e consequentemente a intensidade $I(s)$ ficam representadas por uma combinação não linear dos coeficiente de multipolos:

$$I(s) = \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s)|^2 \propto \sum_{l=0}^L \sum_{m=-l}^l \left| \sum_{p=0}^P f_{lm}^p s^{l+2p} \right|^2 \quad 6.24$$

portanto o envelope molecular da partícula pode ser obtido através da minimização de um funcional do tipo:

$$\Phi = \Phi(s, f_{lm}) = \int_{s \min}^{s \max} [I_{\exp}(s) - I(s, f_{lm})]^2 W^2(s) ds \quad 6.25$$

onde $W(s)$ é uma função peso.

6.3 - Determinação do envelope molecular de uma inibidora de α -amilase

As α -amilases são encontradas em vários organismos vivos e são responsáveis pela degradação das ligações glicosídicas do tipo $\alpha(1-4)$. Encontram-se na literatura muitas proteínas capazes de inibir este tipo de proteína (Silano, 1987) e suspeita-se que estas substâncias estão envolvidas no mecanismo de defesa das plantas contra as α -amilases de insetos. Estudos da inibidora de α -amilase em solução foram feitos com o objetivo de determinar seu envelope molecular.

A proteína foi purificada de grãos de trigo (*Triticum aestivum*) de acordo com o método descrito por Granum e Whitaker (1977). Inicialmente os grãos de trigo foram triturados e misturados com água. Esta solução foi filtrada e o sobrenadante aquecido para inativar α -amilases endógenas. O inibidor precipita a 20-50% de saturação de $(\text{NH}_4)_2\text{SO}_4$. Esta fração foi filtrada e depois redissolvida em uma solução

de fosfato, concentrada e levada a uma coluna de CM-celulose. Este procedimento levou a obtenção de uma proteína altamente ativa contra a α -amilase da saliva humana (Oliva, Iulek e Ida, 1996).

Foram utilizados nos experimentos de SAXS 100 μ l de solução aquosa com uma concentração de aproximadamente 5 mg de proteína por ml. Os dados foram colhidos no feixe de luz do EMBL (European Molecular Biology Laboratory) no HASYLAB (Hamburg Synchrotron Laboratory) usando uma câmara com aproximadamente 2,0 m de distância entre a amostra e o detector que cobria o intervalo de angular entre 0-0,25 $\text{\AA}^{-1}$. A concentração de proteína na solução foi determinada experimentalmente através da comparação entre a curva de SAXS da inibidora da α -amilase com outra curva obtida com uma solução de lisozima com uma concentração conhecida e medida instantes antes. O valor obtido (5 mg/ml) foi significativamente menor do que o valor esperado (10 mg/ml). Esta baixa concentração acarretou numa baixa relação sinal/ruído para valores de $s > 0,15 \text{ \AA}^{-1}$. Os dados foram então tratados de acordo com os procedimentos do apêndice A.

Tabela 6.2: Coeficientes f_{lm} encontrados pela minimização do funcional 8.25. O valor de b_0 para estes coeficientes é 21,4.

l	m	f_{lm}	
0	0	3,193868	0,000000
2	0	0,385453	0,000000
2	2	0,080238	0,178488
4	0	-0,170421	0,000000
4	2	0,001744	0,203286
4	4	-0,011746	0,059439

A determinação do envelope foi feita impondo a simetria 222, a qual foi sugerida por estudos cristalográficos (Oliva, Iulek e Ida, 1996), e utilizando um programa que usa as equações descritas previamente assim como novos algoritmos (Svergun, Volkov, Kozin, Barberato e Stuhmann, 1996). Os coeficientes f_{lm} , tabela 6.2, são números imaginários cujo módulo não excede valor de 5 quando é usado um fator de escala $b_0 = (3V/4\pi)$ onde V é o volume da partícula.

6.3.1 - Resultados

Figura 6.1: Curva de SAXS para o modelo da inibidora de α -amilase (linha continua) ajustada à curva experimental (linha tracejada).

A curva experimental da inibidora de α -amilase assim como a curva calculada do modelo obtido estão ilustrados na figura 6.1. O fator de discordância entre elas é de 0,2%. A máxima dimensão da proteína encontrada utilizando o programa *GNOM* (Svergun *et al.*, 1988) foi de 85 Å. Esta distância juntamente com o valor de s_{max} indicam, segundo o teorema de Shannon (Seção 3.8), que a curva experimental pode ser descrita com 8 parâmetros independentes. A simetria 222 impõem fortes restrições

nos coeficientes f_{lm} e se a simetria coincide com os eixos cartesianos todos os coeficientes exceto f_{00} , f_{20} , f_{22} , f_{40} , f_{42} , f_{44} ... desaparecem. O número final dos parâmetros que descrevem o envelope fica então reduzido a 9, o qual está próximo da estimativa de parâmetros independentes segundo o teorema de Shannon. Estes coeficientes ainda guardam relações entre si pois parâmetros integrais como volume e raio de giro podem ser representados por uma combinação destes (Svergun e Stuhrmann, 1991). Desta forma a fixação do volume e do raio de giro automaticamente implica na redução do número de coeficientes f_{lm} necessários para representação do envelope molecular. Este recurso não foi utilizado entretanto com o intuito de se comparar os valores obtidos pela minimização do funcional 8.25 com aqueles obtidos através das equações 3.25 e 3.34. Esta comparação esta ilustrada na tabela 6.3 e mostra uma razoável concordância entre os valores obtidos pelos dois diferentes métodos.

Tabela 6.3: Parâmetros geométricos da inibidora de α -amilase.

	Rg (Å)	Volume (Å ³)	D _{max} (Å)
Experimental	24,7	74.200	85
Modelo	24,9	82.000	90

6.3.2 - Discussão

O envelope molecular obtido apresenta uma forma alongada com dimensões máximas de 88 x 44 x 35 Å e com um volume de aproximadamente 82.000 Å³ (figuras 6.2, 6.3 e 6.4). Estudos cristalográficos sugerem um tetrâmero com simetria 222 na unidade assimétrica. Nestas condições o valor calculado para um tetrâmero de 52,4 kDa

seria de aproximadamente $64.000 \text{ \AA}^3$. Levando-se em conta que o volume hidratado obtido por SAXS é geralmente de 20-30% maior que o volume da proteína seca podemos afirmar que o volume encontrado concorda razoavelmente com valor esperado. A forma alongada encontrada também é consistente com duas outras técnicas experimentais. Estudos de dicroísmo circular indicam que a estrutura secundária desta proteína é composta basicamente de hélices- α . Estudos de ressonância magnética nuclear com outra inibidora de α -amilase revelaram uma estrutura formada por um pacote (*bundle*) de hélices- α (Strobl *et. al.*, 1995). Um tetrâmero formado pela combinação de quatro “*bundles*” de hélices- α é compatível com o envelope encontrado.

O número de parâmetros utilizados (9) usados para descrever o modelo é compatível com o ponto de vista da quantidade de informação contida na curva experimental segundo o teorema de Shannon. É importante enfatizar porém que o envelope obtido não deve ser considerado como uma solução única. A interpretação dos dados de espalhamento é sempre ambígua e diferentes envelopes podem ser encontrados os quais produzem curvas quase idênticas de SAXS dentro um intervalo de s finito. Além disso a curva de SAXS utilizada é somente uma aproximação da curva provocada somente pelo volume excluído da proteína (Capítulo 5). A última só poderia ser determinada através do método de variação de contraste (seção 5.2). Não é possível portanto utilizar as partes da curva com valores de s superiores a $0.25 \text{ \AA}^{-1}$ porque a intensidade neste intervalo é influenciada pela flutuações da densidade eletrônica. Com estas restrições o modelo apresentado pode ainda ser considerado confiável pois além de ajustar os dados experimentais ele é compatível com os resultados obtidos por

experimentos independentes. Será muito interessante poder comparar o envelope obtido através de curvas de SAXS com a estrutura cristalográfica se esta vier a ser determinada (Oliva, Iulek e Ida, 1996).

Figura 6.2: Modelo da inibidora de α -amilase obtida por determinação do envelope molecular.

Figura 6.3: Modelo da figura 6.2 rodado 90° ao redor do eixo x.

Figura 6.4: Modelo da figura 6.3 rodado 45° ao redor do eixo z.

CAPÍTULO 7 - CÁLCULO DA CURVA DE ESPALHAMENTO DE PROTEÍNAS EM SOLUÇÃO BASEADO EM COORDENADAS ATÔMICAS

Figura 7.1: Volume excluído e suas representações: a) Volume excluído, b) Volume excluído representado por elementos cúbicos (Método do cubo) e c) Volume excluído representado por esferas com raios de van der Waals (Método do fator de estrutura efetivo). Abaixo destes encontram-se gráficos que mostram a densidade eletrônica (ρ) nos pontos em que o vetor $\vec{r}$ atravessa o volume excluído e suas representações. Fica claro por estes que o método do cubo faz uma representação mais fiel do volume excluído do que o método do fator de estrutura efetivo.

Comparações entre curvas experimentais de SAXS e aquelas calculadas a partir de estruturas cristalográficas tem sido amplamente utilizadas para validar modelos teóricos, verificar similaridades estruturais entre proteínas em cristais e em solução e prever estruturas quaternárias (Langridge *et al.*, 1960; Ninio; Luzzati; Yaniv, 1972; Fedorov; Ptisyn; Voronin, 1972; Fedorov; Denesyuk, 1978; Müller, 1983; Pavlov; Sinev; Timchenko; Ptitsyn, 1986; Grossmann *et al.*, 1993; Svergun; Barberato; Koch, 1995).

O principal problema na avaliação da curva de SAXS a partir de estruturas atômicas é a correta determinação do volume excluído (figura 7.1a) e os vários métodos que tem sido desenvolvidos para a avaliação da curva de SAXS diferem basicamente na representação deste. No método do fator de estrutura efetivo (Langridge *et al.*, 1960; Lattman, 1989) o volume excluído é construído colocando-se “átomos de solvente” na posição de cada átomo da estrutura. Os “átomos de solvente” são estruturas fictícias, geralmente representadas por esferas, de tal forma que a soma dos volumes de todos essas esferas seja igual ao volume excluído da proteína. Essa representação tem como principal inconveniente o preenchimento inomogêneo do volume excluído como ilustrado pela figura 7.1c. Esta aproximação é justificada para resoluções menores de 10-20 Å, isto é, para um intervalo $0 \leq s \leq 0,3 \text{ Å}^{-1}$. Para melhores resoluções o preenchimento inomogêneo do volume excluído pode introduzir desvios sistemáticos. O método do cubo (Ninio; Luzzati; Yaniv, 1972; Fedorov; Ptisyn; Voronin, 1972) e suas modificações (Pavlov; Fedorov, 1983; Müller, 1983) preenchem homogeneamente o volume excluído (figura 7.1b) providenciando então uma melhor acurácia para valores mais altos do vetor de espalhamento ($s > 0,3 \text{ Å}^{-1}$).

Os métodos acima descritos não levam em conta a camada de hidratação que envolve as macromoléculas em solução. A omissão dessa camada pode ocasionar erros sistemáticos significativos nos valores de $I(s)$ próximos da origem e, portanto, comprometer a interpretação dos resultados. Complexos procedimentos tem sido utilizados para a inclusão da camada de hidratação (Hubbard; Hodgson; Doniach, 1988; Grossmann *et al.*, 1993), porém nenhum desses procedimentos foi implantado na forma de um programa. As próximas seções irão introduzir o programa *CRY SOL* (Svergun; Barberato; Koch, 1995) que avalia curvas de SAXS levando em conta o espalhamento provocado pela camada de hidratação.

Figura 7.2: Representação esquemática de uma proteína em solução. Para maiores explicações veja o texto.

7.1 - *CRY SOL*

Macromoléculas em solução podem ser esquematicamente representadas como ilustrado na figura 7.2. A partícula com uma densidade eletrônica $\rho(\vec{r})$ esta imersa num solvente com uma densidade eletrônica média ρ_0 . A camada de hidratação é

aproximada por uma camada na superfície da proteína com espessura Δ e densidade ρ_b que pode diferir de ρ_0 . A intensidade espalhada por este sistema será:

$$I(s) = \left\langle \left| A_p(\vec{s}) - \rho_0 A_s(\vec{s}) + \delta \rho A_b(\vec{s}) \right|^2 \right\rangle_{\Omega} \quad 7.1$$

onde $A_p(\vec{s})$ é a amplitude espalhada pela proteína no vácuo, $A_s(\vec{s})$ e $A_b(\vec{s})$ são, respectivamente, a amplitude espalhada pelo volume excluído e pela camada de hidratação, ambas com densidade unitária e $\delta \rho = \rho_0 - \rho_b$.

A promediação espacial na equação 7.1 é grandemente facilitada pelo uso da expansão de multipolos (seção 6.2). A amplitude espalhada por uma proteína composta de N átomos com fatores de forma $f_i(s)$ é dada por:

$$A_p(\vec{s}) = \sum_{i=1}^N f_i(s) \exp(i\vec{s} \cdot \vec{r}_i) \quad 7.2$$

Substituindo a relação 6.10 na equação acima temos:

$$A_p(\vec{s}) = \sum_{l=0}^L \sum_{m=-l}^l A_{lm}(s) Y_{lm}(\Omega) \quad 7.3$$

onde $A_{lm}(s)$ são as amplitudes parciais:

$$A_{lm}(s) = 4\pi i^l \sum_{i=1}^N f_i(s) j_l(sr_i) Y_{lm}^*(\omega_i) \quad 7.4$$

A amplitude espalhada pelo volume excluído, quando este é representado por uma superposição de “átomos” de solvente com fatores de forma $g_i(s)$ e centrados nas mesmas coordenadas $\vec{r}_i$ dos átomos da proteína, é expressa na forma de 7.4, onde as amplitudes parciais $C_{lm}(s)$ são dadas por:

$$C_{lm}(s) = 4\pi i^l \sum_{i=1}^N g_i(s) j_l(sr_i) Y_{lm}^*(\omega_i) \quad 7.5$$

Usando os mesmos argumentos utilizados na seção 6.2 podemos descrever densidade eletrônica da camada de hidratação, $\rho_b(\vec{r})$ com a ajuda da função $F(\omega)$:

$$\rho_b(\vec{r}) = \begin{cases} \rho_0, & \text{se } F(\omega) \leq \vec{r} < F(\omega) + \Delta \\ 0, & \text{se } 0 \leq \vec{r} < F(\omega) \text{ ou } \vec{r} \geq F(\omega) + \Delta \end{cases} \quad 7.6$$

Como mostrado na seção 6.2 as amplitudes parciais são a transformada de Hankel da funções radiais $\rho_{lm}(r)$:

$$B_{lm}(s) = i^l (2/\pi)^{1/2} \int_0^\infty \rho_{lm}(r) j_l(sr) r^2 dr \quad 7.7$$

onde:

$$\rho_{lm}(r) = \int \rho_b(\vec{r}) Y_{lm}^*(\omega) d\omega \quad 7.8$$

Utilizando os mesmos procedimentos usados na seção 6.2 pode-se mostrar que:

$$B_{lm}(s) = i^l (2/\pi)^{1/2} \int_{F(\omega)}^{F(\omega)+\Delta} Y_{lm}^*(\omega) d\omega \int_{F(\omega)} j_l(sr) r^2 dr \quad 7.9$$

Substituindo 7.9, 7.5 e 7.4 em 7.1 obtemos uma fórmula na qual a amplitude espalhada total fica representada como uma somatória das amplitudes parciais $A_{lm}(s)$, $B_{lm}(s)$ e $C_{lm}(s)$. Devido as propriedades ortogonais dos harmônicos esféricos, todos os termos cruzados desaparecem após a promediação espacial resultando numa fórmula simples para a intensidade espalhada:

$$I(s) = \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s) - \rho_0 C_{lm}(s) + \delta \rho B_{lm}(s)|^2 \quad 7.10$$

onde o valor L define a resolução da representação da partícula.

7.1.1 - Avaliação das amplitudes parciais

Esta seção irá descrever como os fatores de forma $f_j(s)$ e $g_j(s)$ e a função $F(\omega)$, necessários para a avaliação das amplitudes parciais $A_{lm}(s)$, $C_{lm}(s)$ e $B_{lm}(s)$, respectivamente, são especificados.

É muito comum ignorar os átomos de hidrogênio nos cálculos das amplitudes espalhadas para os experimentos de difração de raios X em cristais. Porém, átomos de

hidrogênio representam em média 15% da quantidade de elétrons numa proteína exercendo portanto uma significativa influência na intensidade espalhada nas proximidades de $s = 0$. Entretanto, incluir os hidrogênios na mesma base que os outros átomos implicaria em primeiro desenvolver um complicado algoritmo para o correto posicionamento destes e segundo multiplicar o número de átomos na equação 7.4 por um fator de 2 ou 3. Uma alternativa a estes procedimentos é substituir os átomos ligados covalentemente a hidrogênios por grupos atômicos de acordo com a tabela 7.1. Os fatores de forma de cada grupo atômico ficam definidos como a raiz quadrada da intensidade espalhada promediada do respectivo grupo. A equação abaixo exemplifica o caso para a obtenção do fator de forma para o grupo metil ($f_{CH_3}(s)$):

$$f_{CH_3}(s) = \sqrt{\left\langle \left| f_C(s) + f_H(s) \exp(i\vec{s} \cdot \vec{r}_1) + f_H(s) \exp(i\vec{s} \cdot \vec{r}_2) + f_H(s) \exp(i\vec{s} \cdot \vec{r}_3) \right|^2 \right\rangle_\Omega} \quad 7.11$$

onde $f_C(s)$ e $f_H(s)$ são, respectivamente, os fatores de forma para o carbono e o hidrogênio, calculados a partir de uma aproximação de 5 Gaussianas (*International Tables for X-ray Crystallography*, 1974), e $\vec{r}_1$, $\vec{r}_2$ e $\vec{r}_3$ determinam a posição dos hidrogênios ao redor do átomo de carbono para um grupo metil (o átomo de carbono foi situado na origem do sistema de coordenadas para efeito de simplificação nos cálculos). Os fatores de forma para os átomos assim calculados são tabelados no intervalo $0 \leq s \leq 1,0\text{\AA}$.

Os fatores de forma para os “átomos de solvente” são expressos como sugerido por Frase, MacRae e Suzuki (1978):

$$g_j(s) = \exp(-\pi s^2 V_j^{2/3}) \quad 7.12$$

onde $V_j = (4\pi/3) r_{wj}^3$ é o volume de solvente deslocado pelo j -ésimo átomo/grupo atômico (tabela 7.1).

Tabela 7.1: Parâmetros dos átomos e grupos atômicos.

Átomos ou grupos atômicos	Volume de solvente deslocado ($\text{\AA}^3$)	Raio ($\text{\AA}$)
H	5,15	1,07
C*	16,44	1,58
CH*	21,59	1,73
CH ₂ **	26,74	1,85
CH ₃ **	31,89	1,97
N*	2,49	0,84
NH**	7,64	1,22
NH ₂ **	12,79	1,45
NH ₃ **	17,94	1,62
O*	9,13	1,30
OH**	14,28	1,50
S***	19,86	1,68
SH**	25,10	1,81
Mg***	17,16	1,60
P***	5,73	1,11
Ca***	31,89	1,97
Mn***	9,20	1,30
Fe***	7,99	1,24
Cu***	8,78	1,28
Zn***	9,85	1,33

* Volumes deslocados segundo Frase, MacRae e Suzuki (1978)

** Avaliado pela adição do volume dos hidrogêneos correspondentes

*** Raios extraídos da *International Tables for X-ray Crystallography* (1968)

Nesta representação as densidades eletrônicas dos “átomos de solvente” são representados por esferas Gaussianas:

$$\rho(r) = \rho_0 \exp(-(r/r_0)^2) \quad 7.13$$

que minimizam as inogeneidades da densidade eletrônica do volume excluído quando comparado com esferas (figura 7.3).

Figura 7.3: Densidade eletrônica de átomos de solvente representados por esferas Gaussianas (linhas tracejadas) e sua soma (linha contínua em negrito). Note que a densidade do volume excluído assim representado apresenta variações menores quando comparado com a densidade do volume excluído quando este é representado por esferas (figura 7.1b).

Para avaliar a função $F(\varpi)$ a partícula é primeiramente movida de tal forma que seu centro geométrico coincida com a origem. A função $F(\varpi)$ é avaliada somente para um conjunto discreto de direções ϖ_i . Este conjunto forma uma malha quase uniforme de direções e é avaliada usando os números de Fibonacci (Svergun, 1994). Cada átomo da estrutura atualiza o valor da função $F(\varpi)$ na direção ϖ_j se a distância mínima entre o átomo e esta direção é menor que a soma do raio atômico r_{gi} e o raio de uma molécula de água, $r_w = 1,5 \text{ \AA}$. O valor atualizado é:

$$F(\varpi_i) = \max \{ F'(\varpi_i), (r_{ji} + 0,5 r_{gi}) \} \quad 7.14$$

onde $F'(\varpi_j)$ é o valor corrente da função $F(\varpi_i)$ e r_{ji} é a projeção de r_j na direção de (figura 7.4).

Figura 7.4: Avaliação da função $F(\varpi)$. A coordenada atômica corrente referente a origem O é $\vec{r}_j$, a direção corrente é ϖ_i , $|OP| = r_{ji}$, $|PQ| = 0.5 r_{gj}$, $F(\varpi_i) = |OQ|$. Para detalhes veja o texto.

Depois de todos os átomos serem processados a função $F(\varpi)$ contém as distâncias entre a origem e a superfície da partícula para cada direção ϖ_i . As amplitudes $C_{lm}(s)$ podem ser então avaliadas através da integração numérica da equação 7.9.

A densidade do solvente ligado a proteína pode diferir significativamente da densidade do solvente não ligado somente para regiões que distam poucos ângströms da superfície da proteína (Cheng; Schoenborn, 1990), isto é, a condição $F(\varpi) \gg \Delta$ é usualmente satisfeita. Isto significa que a contribuição da camada de hidratação depende majoritariamente no produto $\delta\rho \Delta$, portanto, podemos fixar um dos parâmetros sem perda de generalidade. No programa *CRY SOL*, a largura efetiva da camada de hidratação é tomada como $3,0 \text{ \AA}$, que é aproximadamente a largura da primeira camada observada em cristais (Badger, 1993). Desta forma deixamos como variável no cálculo da intensidade de SAXS, equação 7.10, somente o contraste eletrônico $\delta\rho$ da camada

de hidratação. O limite superior desta variável é $\delta\rho_{max} \approx 0,09 \text{ elétrons}/\text{\AA}^3$ (Perkins, 1986).

7.1.2 - Implementação do programa

Os algoritmos acima descritos foram implementados num programa iterativo chamado *CRY SOL*. Este programa foi escrito em Fortran 77, o qual avalia curvas de SAXS executando os seguintes passos:

(1) As coordenadas atômicas são lidas de um arquivo no formato do *Protein Data Bank* (PDB) (Bernstein *et al.*, 1977). O centro geométrico é determinado e a proteína é deslocada para este ponto.

(2) Uma malha de direções angulares quase-uniforme ϖ_i , $i = 1, 2, \dots, N_g$ ($N_g \leq 4185$), os fatores de forma e as integrais das funções de Bessel (equação 7.9) são tabuladas no intervalo angular (s) e no nível de resolução L indicados pelo usuário.

(3) Os tipos dos átomos/grupo atômicos são lidos de novo do arquivo PDB em blocos de mil átomos e armazenados temporariamente numa área de memória. O artifício de ler as coordenadas em blocos de mil átomos permite evitar limites no número de átomos para o arquivo de entrada. Para cada átomo ou heteroátomo, o grupo atômico e o volume deslocado são identificados de acordo com a tabela 1 e as contribuições para as amplitude parciais $A_{lm}(s)$, $C_{lm}(s)$ e para a função $F(\varpi)$ são avaliadas. Os átomos pertencentes a moléculas de água são ignorados.

(4) As amplitudes $B_{lm}(s)$ são avaliadas através da equação 7.9 utilizando a função $F(\varpi)$ avaliada no passo anterior.

(5) A intensidade espalhada $I(s)$ é calculada usando a equação 7.10 para o valores padrões $\rho_0 = 0,334 \text{ elétrons/\AA}^3$ e $\delta\rho = 0,03 \text{ elétrons/\AA}^3$.

(6) Se a curva experimental $I_e(s)$ é dada, o contraste na camada de hidratação é variado de forma a ajustar a curva teórica com a experimental. O valor de $\delta\rho$ é variado dentro do intervalo $0 \leq \delta\rho \leq 0,09 \text{ elétrons/\AA}^3$ de modo a minimizar o funcional

$$\chi^2(\delta\rho) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(s_i) - cI(s_i, \delta\rho)}{\sigma(s_i)} \right]^2 \quad 7.15$$

onde N_p é o número de pontos experimentais, $\sigma(s_i)$ são os erros experimentais e

$$c = \left[\sum_{i=1}^{N_p} \frac{I_e(s_i)I(s_i, \delta\rho)}{\sigma(s_i)^2} \right] \left[\sum_{i=1}^{N_p} \frac{I(s_i, \delta\rho)^2}{\sigma(s_i)^2} \right]^{-1} \quad 7.16$$

é o fator de escala. Um gráfico da curva experimental e da curva calculada é mostrado na tela do computador.

(7) Os resultados (parâmetros integrais da partícula, a função $F(\varpi)$, as amplitudes parciais, intensidades e o ajuste entre as curvas experimental e a calculada) são armazenados em arquivos binários e ASCII. Os dados podem assim ser recuperados para cálculos adicionais com outros parâmetros e/ou dados experimentais.

O programa é compilado em computadores tipo IBM PC ou compatíveis usando o Microsoft Fortran PowerStation 1.0 com o Phar Lap MS-DOS extender e requer uma versão do DOS 3.3 ou mais recente, 2 Mbytes de memória livre (convencional + estendida) e um monitor EGA/VGA/SVGA. Versões para

computadores utilizando UNIX (Sun, Silicon Graphics, DEC Alpha) as quais utilizam o pacote gráfico *GNU-PLOT*, são também disponíveis.

7.1.3 - Aplicações de *CRY SOL*

As seções 7.1.3.1 e 7.1.3.2 irão mostrar a aplicação de *CRY SOL* para duas proteínas. A primeira proteína escolhida foi a lisozima porque esta tem sido utilizada para propósitos ilustrativos por diversos autores (Pickover; Engelman, 1982; Pavlov; Fedorov, 1983; Lattman, 1989). A segunda proteína é a hexocinase que é uma enzima composta de duas sub-unidades idênticas. Os resultados da resolução cristalográfica da estrutura desta proteína apresentaram três possibilidades de dimerização. Experimentos de espalhamento de raios X em solução mostraram que somente uma conformação é compatível com os dados obtidos.

7.1.3.1 - Lisozima

Figura 7.5: Curva de SAXS para a lisozima (Pontos) e o melhor ajuste usando *CRY SOL* (linha cheia). A curva tracejada indica a melhor curva para $\delta\rho = 0$.

A curva de espalhamento da lisozima (figura 7.5) foi obtida utilizando os procedimentos padrões da câmara X33 do European Molecular Biology Laboratory (EMBL) no HASYLAB em Hamburgo, Alemanha. As coordenadas da lisozima (Diamond, 1974) foram obtidas do arquivo PDB 6lyz.pdb. A figura 7.6 mostra as curvas de espalhamento $I_a(s) = \langle |A_a(\vec{s})|^2 \rangle$, $I_b(s) = \delta\rho^2 \langle |A_b(\vec{s})|^2 \rangle$ e $I_c(s) = \rho_0 \langle |A_c(\vec{s})|^2 \rangle$ avaliados para $L = 12$ e $N_g = 2585$. Na figura 7.5 o melhor ajuste entre dados teóricos e experimentais ($\chi^2 = 0,477$) ocorre para o valor de $\delta\rho = 0,025 \text{ e/\AA}^3$, correspondendo a uma hidratação de $0,4 \text{ g (H}_2\text{O) / g (Proteína)}$. O raio de giração obtido para a curva experimental (R_{ge}) foi $15,2 \text{ \AA}$, que é bem próximo do raio de giro teórico (R_{gt}), $14,8 \text{ \AA}$. Os raios de giro foram obtidos utilizando a aproximação de Guinier. É importante notar que quando a camada de hidratação não é incluída nos cálculos ocorre um pior ajuste entre as curvas ($\chi^2 = 0,765$, $R_{gt} = 14,3 \text{ \AA}$).

Figura 7.6: Intensidade espalhada pela estrutura atômica no vácuo $I_a(s)$ (1), a intensidade espalhada pelo volume excluído $I_b(s)$ (2) e pela camada de hidratação $I_b(s)$ (3). As linhas tracejadas são as curvas avaliadas pelo programa de Pavlov.

Os resultados de *CRY SOL* foram comparados com o programa de Pavlov e Fedorov (1993) (figura 7.6), o qual usa o método do cubo e métodos numéricos para a promediação espacial. As curvas $I_a(s)$ e $I_c(s)$ apresentam uma boa concordância até $s \approx 0,4 \text{ \AA}^{-1}$. Para maiores ângulos existem desvios devidos ao preenchimento inhomogêneo do volume excluído. A curva calculada pelo programa de Pavlov, o qual não inclui a camada de hidratação nos cálculos, apresenta desvios sistemáticos para valores baixos de s ($\chi^2 = 0,687$, $R_{gt} = 14,7 \text{ \AA}$). Esses desvios podem ser reduzidos mudando-se artificialmente a densidade do solvente para $\rho_0 = 0,310 \text{ e/\AA}^3$ resultando um melhor valor para χ^2 (0,581) e deixando praticamente inalterado o valor de R_{gt} (14,5 $\text{\AA}$). O tempo total de CPU exigido por *CRY SOL* em um IBM AT/486 DX50 foi de 310 s. Comparativamente o programa de Pavlov necessitou 660 s para executar a mesma tarefa.

7.1.3.2 - Hexocinase

A transferência de fosforilas é uma reação básica na bioquímica. Uma enzima que catalise a transferência de um grupamento fosforila do ATP para um aceptor é chamada de uma *cinase*. A hexocinase, portanto, catalisa a transferência de uma fosforila do ATP para uma variedade de oses com seis carbonos (*Hexoses*), como a glicose e a manose. A hexocinase de trigo tem um peso molecular de 104 kDa e é composta de duas sub-unidades idênticas. A primeira estrutura cristalográfica com resolução atômica (Steitz *et al.*, 1976, grupo espacial $P2_12_12_1$, dimensões da cela $a = 144,2$, $b = 87,4$ $c = 99,4 \text{ \AA}$ e resolução de $3,5 \text{ \AA}$) apresentava uma curiosa e inédita

característica: as duas sub-unidades não eram relacionadas por um elemento de simetria. Entretanto, outra estrutura cristalográfica (Reißner *et al.*, comunicação pessoal, grupo espacial $P2_12_12$, dimensões da cela $a = 215,2 \text{ \AA}$, $b = 91,3 \text{ \AA}$, $c = 58,3 \text{ \AA}$, resolução $3,2 \text{ \AA}$) mostrava que outras duas estruturas quaternárias eram possíveis para a hexocinase.

Figura 7.7: Curva de experimental para a hexocinase (pontos) comparada com as curvas calculadas de três possíveis dimeros. Dímeros 1 e 2 foram resolvidos por Reißner e o dímero 3 por Steitz (1976).

Soluções de hexocinase em pH's fisiológicos e com concentrações de 3-20 mg/ml foram utilizadas para obtenção de uma curva de SAXS. Os dados foram coletados na câmara X33 do EMBL no HASYLAB usando um detector linear. O programa *CRY SOL* foi então utilizado para comparar as curvas teóricas dos três dimeros com a curva experimental e os resultados estão mostrados na figura 7.7. Uma análise qualitativa desta figura mostra que a conformação que mais se aproxima da curva experimental é o dímero encontrado por Steitz (1976). Esta conclusão é

corroborada pelos valores obtidos para os desvios (χ^2) entre as curvas e os raios de giro mostrados na tabela 7.2. O menor valor encontrado para χ^2 assim como o raio de giro teórico que mais se aproximou do raio de giro experimental foram encontrados para o dímero 3.

Tabela 7.2: Valores obtidos para as três conformações da hexocinase encontradas em cristais. Os dímeros 1 e 2 foram obtidos por Reißner (comunicação pessoal) e o dímero 3 por Steitz *et al.* (1976). As coordenadas atômicas deste dímeros encontram-se disponíveis em um disquete anexo a esta tese.

Dímero	χ^2	R _g (Å)*	$\delta\rho$ (eletrons/Å ³)	Hidratação**
1	2,612	35,0	0,064	0,274
2	3,096	37,1	0,088	0,288
3	1,283	32,8	0,010	0,272

* O valor experimental para o raio de giro da hexocinase, R_{ge}, é 32,2 Å.

** A unidade de hidratação é dada por “g de H₂O por g de proteína”.

7.2 - Discussão

A importância da contribuição da camada de hidratação para o espalhamento de raios X tem sido discutida por diversos autores (Hubbard; Hodgson; Doniach, 1988; Schoenborn, 1988; Badger, 1993; Grossmann *et al.*, 1993). As estruturas disponíveis no *Protein Data Bank* usualmente contém menos que 50% das moléculas de água ligadas à proteína e, portanto, dificilmente podem representar a camada de hidratação em solução. Tentativas de incluir as moléculas de água com o programa de Pavlov somente pioraram os resultados.

A representação da camada de hidratação introduzida por *CRY SOL* é, naturalmente, um modelo simplificado. Para moléculas de formas complicadas o

envelope molecular pode não ter um valor simples e o uso de $F(\varpi)$ preencheria as cavidades internas. Embora $F(\varpi)$ não seja então apropriado para descrever o espalhamento do volume excluído para altos valores de s , podemos ainda utilizá-lo para simular a camada de hidratação. A justificativa para esta afirmação vem é que, estando na periferia da proteína, a camada de hidratação contribui significativamente para o espalhamento somente para pequenos valores de s , onde a resolução é baixa e, portanto, uma descrição detalhada da superfície da proteína faz-se desnecessária. Uma proteína pode conter até 3 camadas de moléculas de água ordenadas em sua superfície. A representação da hidratação em *CRY SOL*, com uma casca de espessura igual a 3,0 Å e densidade eletrônica uniforme ao redor da superfície da proteína procura simular a primeira camada de hidratação, a qual contém a maioria das moléculas organizadas de solvente e, portanto, domina o espalhamento provocado pela hidratação (Tanki; Thorton; Goodfellow, 1988; Cheng; Schoenborn, 1990; Badger, 1993). Para todas as proteínas estudadas até agora a contribuição da camada de hidratação tem significativamente melhorado os ajustes das curvas teóricas com as curvas experimentais, sendo que a razão de hidratação encontrada variou entre 0,2-0,3 g g⁻¹. O valor de 0,4 g g⁻¹ encontrado para a lisozima foi excepcionalmente alto.

O problema da dimerização da hexocinase exemplifica claramente a validade dos experimentos de SAXS para a determinação ou confirmação da estrutura quaternária de uma proteína em solução. Nestes casos *CRY SOL* provou ser uma ferramenta bastante útil devido a sua rapidez e acurácia na avaliação das curvas de SAXS.

CAPÍTULO 8 - DETERMINAÇÃO DA POSIÇÃO RELATIVA ENTRE AS SUB-UNIDADES DE UMA PROTEÍNA

Experimentos de SAXS podem ser utilizados para a determinação da posição relativa de sub-unidades em proteínas oligoméricas, isto é, dada a curva experimental $I_{exp}(s)$ do complexo e as amplitudes de espalhamento de cada sub-unidade é possível, em princípio, determinar a posição relativa das sub-unidades minimizando as diferenças entre as curvas experimental e teórica. Vamos considerar aqui um oligômero composto pelas sub-unidades A e B . A intensidade de raios X espalhada por este heterodímero é facilmente obtida quando as amplitudes espalhadas pelas sub-unidades, $A_0(\vec{s})$ e $B_0(\vec{s})$, são descritas usando a expansão de multipolos (seção 6.2):

$$A_0(\vec{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}^0(s) Y_{lm}(\Omega) \quad 8.1$$

e

$$B_0(\vec{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l B_{lm}^0(s) Y_{lm}(\Omega) \quad 8.2$$

onde as amplitudes parciais

$$A_{lm}^0(s) = \int A_0(\vec{s}) Y_{lm}^*(\Omega) d\Omega \quad 8.3$$

e

$$B_{lm}^0(s) = \int B_0(\vec{s}) Y_{lm}^*(\Omega) d\Omega \quad 8.4$$

são especificadas de tal maneira que os centros de massa de cada sub-unidade coincidam com a origem de coordenadas no espaço real. A posição da partícula B em relação a partícula A pode ser descrita através da rotação de B pelos ângulos Eulerianos α , β e γ seguido de uma translação do seu centro de massa por um vetor $\vec{r}$, $\vec{r} = (r, \omega) = (r, \theta, \varphi)$. A intensidade $I(s)$ de tal heterodímero é escrita como:

$$I(s) = \left\langle \left| A_0(\vec{s}) + B(\vec{s}) \right|^2 \right\rangle_{\Omega} \quad 8.5$$

onde $B(\vec{s})$ é a amplitude da partícula transladada:

$$B(\vec{s}) = \exp(i\vec{s} \cdot \vec{r}) B'(\vec{s}) \quad 8.6$$

e $B'(\vec{s})$ é a amplitude espalhada da partícula rotacionada:

$$B'(\vec{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l B'_{lm}(s) Y_{lm}(\Omega) \quad 8.7$$

As amplitudes parciais da partícula rotacionada relacionam-se com as amplitudes parciais da partícula na origem através da relação:

$$B'_{lm}(s) = \sum_{l'=-l}^l D_{m'l'}^l(\alpha \beta \gamma) B_{l'}^0(s) \quad 8.8$$

sendo $D_{m,t}^l(\alpha \beta \gamma)$ o elemento de matrix para rotações finitas (Edmonds, 1957):

$$D_{m,t}^l(\alpha \beta \gamma) = \exp(im\gamma) d_{m,t}^l(\beta) \exp(im\alpha) \quad 8.9$$

onde:

$$d_{m,t}^l(\beta) = \left[\frac{(l+m)!(l-m)!}{(l+t)!(l-t)!} \right]^{1/2} \sum_{k=k \min}^{k \max} \binom{l+t}{l-m-k} \binom{l-t}{k} (-1)^{l-m-k} \left(\cos \frac{\beta}{2}\right)^{2k+m+t} \left(\sin \frac{\beta}{2}\right)^{2l-2k-m-t} \quad 8.10$$

$$k \min = \max(0, -m-t) \quad 8.11$$

e

$$k \max = \min(l-m, l-t) \quad 8.12$$

As amplitudes parciais da partícula transladada são relacionadas com as amplitudes parciais da partícula rotacionada por:

$$\begin{aligned} B_{lm}(s) &= \int B(s) Y_{lm}^*(\Omega) d\Omega \\ &= \int \exp(i\vec{s} \cdot \vec{r}) B'(s) Y_{lm}^*(\Omega) d\Omega \\ &= \int \exp(i\vec{s} \cdot \vec{r}) \sum_{k=0}^{\infty} \sum_{t=-k}^k B'_{kt}(s) Y_{kt}(\Omega) Y_{lm}^*(\Omega) d\Omega \end{aligned} \quad 8.13$$

substituindo a relação (Edmonds, 1957):

$$\exp(i\vec{s} \cdot \vec{r}) = 4\pi \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p j_p(sr) Y_{pq}(\Omega) Y_{pq}^*(\omega) \quad 8.14$$

e a equação 8.7 em 8.13 obtemos:

$$\begin{aligned} B_{lm}(s) &= \int \left(4\pi \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p j_p(sr) Y_{pq}(\Omega) Y_{pq}^*(\omega) \right) \left(\sum_{k=0}^{\infty} \sum_{t=-k}^k B'_{kt}(s) Y_{kt}(\Omega) \right) Y_{lm}^*(\Omega) d\Omega \\ &= 4\pi \sum_{p=0}^{\infty} \sum_{q=-p}^p i^p j_p(sr) Y_{pq}^*(\omega) \sum_{k=0}^{\infty} \sum_{t=-k}^k B'_{kt}(s) \int Y_{lm}^*(\Omega) Y_{pq}(\Omega) Y_{kt}(\Omega) d\Omega \end{aligned} \quad 8.15$$

Esta equação pode ser simplificada utilizando a seguinte relação:

$$\int Y_{lm}(\Omega) Y_{pq}(\Omega) Y_{kt}(\Omega) d\Omega = [(2l+1)(2p+1)(2k+1)/4\pi]^{1/2} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ m & q & t \end{pmatrix} \quad 8.16$$

onde:

$$\begin{aligned} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} &= \\ (-1)^{j_1-j_2-m_3} (2j_3+1)^{-1/2} \delta(m_3, m_1+m_2) & \\ \left(\frac{(j_1+j_2-j_3)!(j_3+j_1-j_2)!(j_3+j_2-j_1)!(2j_3+1)}{(j_1+j_2+j_3+1)!} \right)^{1/2} & \\ \sum_{k=k_{\min}}^{k_{\max}} \frac{(-1)^k (j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!(j_3+m_3)!(j_3-m_3)!}{k!(j_1+j_2-j_3-k)!(j_1-m_1-k)!(j_2+m_2-k)!(j_3-j_2+m_1+k)!(j_3-j_1-m_2+k)!} & \end{aligned} \quad 8.19$$

são os coeficientes de Wigner. Os limites das somatórias são definidas por:

$$k \min = \max (0, j_2 - j_3 - m_1, j_1 - j_3 + m_2) \quad 8.20$$

e

$$k \max = \min(j_1 + j_2 - j_3, j_1 - m_1, l_2 + m_2) \quad 8.21$$

Os coeficientes de Wigner apresentam a seguinte propriedade:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = 0 \quad \text{se} \begin{cases} j_3 > j_1 + j_2 \\ j_3 < |j_1 - j_2| \\ m_3 \neq -m_1 - m_2 \end{cases} \quad 8.22$$

Substituindo a equação 8.16 na equação 8.15 obtemos:

$$\begin{aligned} B_{lm}(s) &= 4\pi(-1)^m \sum_{p=0}^{\infty} i^p j_p(sr) \sum_{q=-p}^p Y_{pq}^*(\omega) \\ &\sum_{k=|l-p|}^{l+p} [(2l+1)(2p+1)(2k+1) / 4\pi]^{1/2} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ -m & q & m-q \end{pmatrix} \sum_{t=-k}^k B'_{kt}(s) \end{aligned} \quad 8.23$$

As substituições das equações 8.5 e 8.8 na equação 8.23 leva-nos finalmente a uma fórmula a qual descreve a intensidade do heterodímero pelos parâmetros α , β , γ , r , θ e φ e pelas amplitudes parciais $A_{lm}^0(s)$ e $B_{lm}^0(s)$:

$$I(s) = \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}^0(s) + B_{lm}^0(s)|^2 = F(\alpha, \beta, \gamma, r, \theta, \varphi) \quad 8.24$$

Esta fórmula pode ser usada para a determinação dos seis parâmetros posicionais através da minimização das diferenças entre $I(s)$ e $I_{exp}(s)$.

A busca dos parâmetros posicionais α , β , γ , r , θ e φ requer várias milhares de avaliações da equação 8.20, portanto esforços foram feitos para reduzir o tempo de CPU utilizado no cálculo de $I(s)$. É evidente da análise das equações 8.8 e 8.23 que o maior esforço computacional na avaliação de $I(s)$ é dispensado na translação da partícula B . A mesma intensidade $I(s)$ pode ser obtida de maneira mais rápida se uma escolha diferente de rotação/translação é feita. No novo procedimento a partícula B é rotacionada por (α, β, γ) seguido por uma nova rotação de ambas partículas A e B por $(0, \theta, \varphi)$. Esta última rotação tem por objetivo fazer coincidir o vetor de translação com o eixo cartesiano z . A translação é agora feita ao longo do eixo z , $\vec{r} = (r, 0, 0)$, a qual significativamente simplifica a equação 8.23, isto é, tal translação reduz os harmônicos esféricos $Y_{lm}(\varpi)$ para:

$$Y_{pq}^*(\varpi) = \begin{cases} 0 & q \neq 0 \\ \left[\frac{2p+1}{4\pi} \right]^{1/2} & q = 0 \end{cases} \quad 8.25$$

Substituindo a equação 8.25 na equação 8.23 obtemos:

$$B_{lm}(s) = (-1)^m \sum_{p=0}^{\infty} i^p j_p(sr) (2p+1) \sum_{k=l-p}^{l+p} [(2l+1)(2k+1)/4\pi]^{1/2} \begin{pmatrix} l & p & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & k \\ -m & 0 & m \end{pmatrix} \sum_{t=-k}^k B'_{kt}(s) \quad 8.26$$

Este novo procedimento provou avaliar curvas teóricas de heterodímeros de maneira muito mais rápida.

Figura 8.1: Exemplificação do procedimento para a obtenção da estrutura de um homodímero a partir somente da estrutura de uma subunidade. O procedimento é composto de basicamente de quatro operações: a) A estrutura da subunidade, que é dada numa posição arbitrária e com o centro de massa coincidindo com a origem, é rotacionada pelos ângulos Eulerianos (α, β, γ) (Neste exemplo a subunidade tem seu eixo maior orientado ao longo do eixo x). b) A subunidade é transladada $r/2$ ao longo do eixo z . c) A outra subunidade é obtida através da aplicação do eixo de ordem 2 que é coincidente com o eixo y . d) A curva de SAXS do homodímero é calculada e comparada com a curva experimental. No alto da figura encontra-se a representação de um homodímero e a sua respectiva curva de SAXS "experimental". Os números 1) e 2) são duas tentativas para a determinação de sua estrutura partindo da estrutura de uma subunidade. Em 1) a subunidade não é rotacionada ($\alpha=0, \beta=0, \gamma=0$) e a curva de SAXS do homodímero obtido em c) não apresenta uma boa concordância com a curva "experimental". Em 2) a subunidade é primeiramente rotacionada por ($\alpha=\pi/2, \beta=0, \gamma=0$) e a comparação feita em d) indica que o modelo obtido tem uma boa probabilidade de ser aquele que originou a curva experimental.

Um caso particular de um oligômero com duas sub-unidades é o homodímero com um eixo de ordem 2 relacionando suas sub-unidades. Esta partícula pode ser descrita usando apenas 4 parâmetros: A orientação da partícula, dada pelos ângulos

Eulerianos (α, β, γ) , em relação ao eixo de ordem 2 e a distância r entre os centros de massa de cada sub-unidade. As equações envolvidas no cálculo da intensidade do homodímero são bastante simplificadas se uma escolha criteriosa das rotações e translações é feita. A figura 8.1 exemplifica o procedimento escolhido para a obtenção da posição relativa entre as subunidades de um dímero. A intensidade espalhada é obtida da seguinte forma: A sub-unidade B , cujo centro de massa esta na origem, é rotacionada por (α, β, γ) e então transladada por $r/2$ ao longo do eixo z . A segunda sub-unidade é então gerada pelo eixo de ordem 2, isto é, pela rotação desta sub-unidade por π ao redor do eixo y . A última rotação transforma as amplitudes parciais da seguinte maneira:

$$\begin{aligned} B'_{lm}(s) &= \sum_{l=-l}^l D'_{m,l}(0\gamma\ 0) B_l(s) \\ &= \sum_{l=-l}^l (-1)^{l-m} \delta_{l,-m} B_l(s) = (-1)^{l-m} B_{l-m}(s) \end{aligned} \quad 8.27$$

onde $B'_{lm}(s)$ são as amplitudes parciais da sub-unidade rotacionada+transladada, obtida através das equações 8.8 e 8.26. Usando a seguinte propriedade das amplitudes parciais:

$$A_{l-m} = (-1)^{l+m} A_{lm}^*(s) \quad 8.28$$

é facilmente mostrado que:

$$B'_{lm}(s) = (-1)^{l-m} (-1)^{l+m} B^*_{lm}(s) = (-1)^{2l} B^*_{lm}(s) = B^*_{lm}(s) \quad 8.29$$

ou seja, as novas amplitudes parciais da partícula rotacionada por ($\alpha = 0, \beta = \pi, \gamma = 0$) são as amplitudes parciais conjugadas da partícula não rotacionada. Substituindo as equações 8.8, 8.26 e 8.29 em 8.5 obtém-se uma equação para o cálculo da intensidade $I(s)$ para um homodímero:

$$I(s) = \sum_{l=0}^{\infty} \sum_{m=-l}^l |2(-1)^m| \quad 8.30$$

$$\sum_{p=0}^{\infty} \text{Re} \left[i^p j_p(sr) (2p+1) \sum_{k=|l-p|}^{l+p} [(2l+1)(2k+1)]^{1/2} \begin{pmatrix} l & p & q \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & q \\ -m & 0 & m \end{pmatrix} \sum_{t=-k}^k B_{kt}(s) \right]^2$$

A qual também pode ser utilizada na determinação da posição relativa entre as sub-unidades de um oligômero.

8.1 - HOMDIM

O algoritmo para a determinação de um homodímero foi implementado num programa denominado *HOMDIM*¹ o qual, dado a curva experimental e a amplitude da sub-unidade, procura os parâmetros posicionais α, β, γ, r que descrevem o homodímero executando basicamente os seguintes passos:

- (1) A intensidade experimental $I_{exp}(s)$ é lida de um arquivo ASCII.

¹ Este programa foi desenvolvido por (Barberato; Svergun, não publicado).

(2) As amplitudes parciais $A_{lm}(s)$ são lidas de um arquivo binário. Essas amplitudes tem que ser avaliadas previamente a partir de suas coordenadas atômicas usando o programa *CRY SOL*.

(3) Um conjunto de ângulos Eulerianos $(\alpha_i, \beta_i, \gamma_i)$ e distâncias r_i são determinadas pelo usuário.

(4) Para cada conjunto de valores $(\alpha_i, \beta_i, \gamma_i, r_i)$ o funcional:

$$\chi^2(\alpha_i, \beta_i, \gamma_i, r_i) = \frac{1}{N_p} \sum_{j=1}^{N_p} \left[\frac{I_{\text{exp}}(s_j) - cI(s_j, \alpha_i, \beta_i, \gamma_i, r_i)}{\sigma(s_j)} \right]^2 \quad 8.31$$

é calculado, onde N_p é o número de pontos experimentais, $\sigma(s_j)$ são os erros experimentais e

$$c = \left[\sum_{j=1}^{N_p} \frac{I_e(s_j) I(s_j, \alpha_i, \beta_i, \gamma_i, r_i)}{\sigma(s_j)^2} \right] \left[\sum_{j=1}^{N_p} \frac{I(s_j, \alpha_i, \beta_i, \gamma_i, r_i)^2}{\sigma(s_j)^2} \right]^{-1} \quad 8.32$$

é o fator de escala.

(5) Os cinquenta conjuntos de parâmetros $(\alpha_i, \beta_i, \gamma_i, r_i)$ cujos χ^2 associados são os menores valores encontrados são selecionados e mostrados na tela do computador em ordem crescente de χ^2 e armazenados num arquivo de saída. Um gráfico mostrando a curva experimental e a curva calculada com a melhor solução é mostrada na tela do computador.

O programa foi compilado exatamente como o programa *CRY SOL* e segue basicamente as mesmas especificações que este. Um arquivo com instruções detalhadas para o uso deste programa encontra-se no apêndice D desta tese. A próxima seção irá ilustrar o uso deste programa para determinação de um modelo para o estado R da aspartato transcarbamilase em solução.

8.2 - ASPARTATO TRANSCARBAMILASE

Aspartato transcarbamilase (ATCase) de *Escherichia coli* é uma enzima que catalisa a carbamilação do grupo amino L-aspartato pelo carbamil-fosfato, sendo este o primeiro passo na via de produção da pirimidina. Suas seis cadeias catalíticas (c) agrupadas em dois trímeros e suas seis cadeias regulatórias (r) agrupadas em três dímeros (figura 8.2A) formam um complexo de 306 kDa com uma quase-simetria D_3 (Kantrowitz; Lipscomb, 1988). As propriedades cooperativas desta proteína são bem compatíveis com o modelo teórico MWC (Monod; Wyman; Changeux, 1965) o qual é baseado numa transição reversível entre um estado de baixa afinidade T (Tensa) e um estado de alta afinidade R (Relaxada). A mudança conformacional associada com a transição $T \rightarrow R$ foi primeiramente detectada por experimentos de sedimentação (Gerhart; Schachman, 1968). As estruturas cristalinas dos dois estados foram determinadas e o modelo da transição $T \rightarrow R$ foi descrito essencialmente em termos de movimentos de corpo rígido das sub-unidades c e r (Stevens; Gouaux; Lipscomb, 1990).

Figura 8.2: Modelos para as estruturas quaternárias da ATCase. Os dois trímeros catalíticos estão coloridos em amarelo (cadeias c_1) e azul (cadeias c_6) e os três dímeros regulatórios em magenta. Na primeira coluna vertical (A) contém três vistas do modelo no estado T, a segunda coluna (B) três vistas para o modelo no estado R cristalino e a terceira (C) três vista para o modelo no estado R em solução. A linha de cima é a vista ao longo do eixo de ordem 3. A segunda linha é obtida através de uma rotação de 90° no sentido anti-horário ao redor do eixo x dos modelos na linha 1. A última é obtida através da rotação de 90° no sentido horário do modelo na linha 1.

Os movimentos mais proeminentes são um aumento de 11 Å na distância entre os trímeros catalíticos seguida de uma rotação de 6° ao redor do eixo de ordem 3 e uma rotação de 15° dos dímeros regulatórios ao redor dos eixos de ordem 2 (figura 8.2 A, B). Estes movimentos são acompanhados pelo fechamento de uma fenda entre os domínios catalíticos e por mudanças na estrutura terciária do *loop* 240s que forma contatos entre os trímeros catalíticos (Ke *et al.*, 1988).

A alongação da enzima ao longo do eixo de ordem três na transição T→R foi predita tanto pela comparação entre os eixos c das duas estruturas (o eixo c no cristal tipo R é 11 Å maior do que no cristal tipo T) e em experimentos de SAXS (Monaco *et al.*, 1978; Moody *et al.*, 1979). Os desvios sistemáticos observados em comparações anteriores entre curvas calculadas de SAXS baseadas nas estruturas cristalinas e dados experimentais foram atribuídos às contribuições da camada de hidratação que não tinham sido levadas em conta nos cálculos teóricos (Altman, Ladner e Lipscomb, 1982).

Usando as estruturas cristalinas e o programa *CRY SOL* foram calculadas as curvas de SAXS para os dois estados e comparadas com dados experimentais (Svergun; Barberato; Koch; Fetler; Vachette; 1996). O melhor ajuste ($\chi = 2,6$) entre as curvas teórica e experimental para o estado T (figura 8.3) é obtido para um valor de contraste na camada de hidratação $\delta\rho = 0,068 \text{ e/\AA}^3$, correspondendo a uma hidratação de 0,23 g g⁻¹. Se a camada de hidratação não é levada em conta o ajuste não é tão bom e os desvios observados (curva tracejada na figura 8.3) correspondem a aqueles desvios relatados anteriormente (Altman, Ladner e Lipscomb, 1982). Portanto, as estruturas

quaternárias do estado T em solução e em cristais não devem diferir, e as discrepâncias observadas são devidas somente a não ter sido levado em conta os efeitos do solvente.

Figura 8.3: Comparação entre as curvas de SAXS para os estados T e R (pontos). A Curva 1 é modelo para o estado R em solução (linha tracejada), e a curva 2 é o modelo para o estado R cristalino. A curva 3 é o estado T sem a camada de hidratação e a curva 4 é com a camada.

No entanto, o mesmo não ocorre com a enzima no estado R. Ainda que a camada de hidratação melhore o ajuste entre as curvas para o estado R (O valor de χ baixa de 8,2 para 4,8) o melhor ajuste para os dados experimentais ($\delta\rho = 0,09 \text{ e/\AA}^3$) permanece pobre (figura 8.3). Os grandes desvios próximos da origem da curva de espalhamento e ao redor do primeiro máximo subsidiário (resolução de 50-100 Å) apontam para diferenças significativas entre as estruturas cristalina e em solução.

A comparação entre as curvas na figura 8.3 indica que o primeiro máximo da curva de espalhamento para a curva do estado R cristalino reside entre o máximo da curva de espalhamento do estado T e o máximo da curva experimental para o estado R. Isto sugere que a estrutura para o estado R em solução é mais alongada que a observada em cristais. O programa *HOMDIM* foi então utilizado para a comprovação desta hipótese. A estrutura cristalográfica para o estado R foi inicialmente rotacionada de forma a coincidir o eixo de ordem 3 da partícula com o eixo z do sistema de coordenadas e um dos eixos de ordem 2 com o eixo cartesiano y . A estrutura foi então dividida em duas sub-unidades relacionadas pelo eixo de ordem 2. Cada uma destas sub-unidades maiores era composta de três cadeias catalíticas e três regulatórias. As amplitudes parciais $A_{lm}(s)$ de uma das sub-unidades foram avaliadas utilizando o programa *CRY SOL* e introduzidas juntamente com os dados experimentais no programa *TWODIM*. Devido que a orientação da proteína em relação ao eixo de ordem 2 já estava fixada os ângulos Eulerianos $(\alpha_i, \beta_i, \gamma_i)$, que determinam essa orientação, permaneceram todos em zero ($\alpha_i = 0, \beta_i = 0, \gamma_i = 0$). Desta forma, o único parâmetro variável do funcional 8.31 ficou sendo a meia-distância r entre as duas sub-unidades. Foi permitido ao programa variar esta distância no intervalo $18,1 < r < 28,1 \text{ \AA}$, com um passo de $0,1 \text{ \AA}$ ($r = 23,1 \text{ \AA}$ corresponde a estrutura cristalográfica). O melhor ajuste entre as curvas experimental e teórica foi obtido para $r = 26,2 \text{ \AA}$ e ($\chi = 0,90$) (figura 8.4) o que corresponde a aumentar a distância entre as sub-unidades maiores em $6,2 \text{ \AA}$. Ficou assim mostrado que o estado R apresenta uma forma mais alongada em solução quando comparado com a estrutura cristalográfica.

Figura 8.4: Ajuste entre a curva experimental (pontos) e teórica (linha contínua) obtido com o programa *HOMDIM*.

Porém, o aumento de 6,2 Å na distância entre as sub-unidades acarreta na perda de contato entre estas. De maneira a preservar os contatos entre as cadeias regulatórias e catalíticas o centro da interface r_1-c_1 , o qual é pouco afetado na transição $T \rightarrow R$, foi determinado ($r_c = (x_c, y_c, z_c) = (24,8; 10,9; -21,2)$). Deslocamentos deste ponto correlacionam as translações e rotações das duas cadeias de forma a manter as distâncias D_3 e D_2 entre este ponto e os eixos de ordem 3 e 2, respectivamente, constantes. Um aumento na distância entre os trómeros catalíticos (ΔD_c) implica em uma rotação deste ao redor do eixo de ordem 3 ($\Delta\phi$) e um deslocamento (ΔD_r) mais uma rotação ($\Delta\theta$) dos dímeros regulatórios ao redor dos respectivos eixos de ordem 2. As fórmulas que relacionam ΔD_c com $\Delta\phi$, $\Delta\theta$ e ΔD_r encontram-se a seguir:

$$\Delta\theta = \theta_c - \cos^{-1}(z_c / (D_3 + (\Delta D_c / 2))) \quad 8.33$$

$$\Delta\phi = \phi_c - \sin^{-1}(\sin(\theta_c - \Delta\theta) D_2 / D_3) \quad 8.34$$

$$\Delta D_r = D_3 (\cos(\varphi_c + \Delta\varphi) - \cos(\varphi_c)) \quad 8.35$$

$$D_3 = \sqrt{x_c^2 + y_c^2} \quad 8.36$$

$$D_2 = \sqrt{y_c^2 + z_c^2} \quad 8.37$$

A distância dos trímeros catalíticos foi variada entre 0,0 e 5,0 Å e a cada novo valor de ΔD_c um novo modelo para o estado R foi gerado através do deslocamento e rotação das sub-unidades segundo as fórmulas 8.33-8.37. A figura 8.5 apresenta as discrepâncias entre a curva experimental e aquelas dos modelos *versus* o aumento dos deslocamentos ΔD_c ($\Delta D_c = 0$ é a estrutura cristalina para o estado R). O gráfico mostra um claro mínimo ($\chi_{\min} = 1,0$) para $\Delta D_c = 3,5$ Å, o que implica numa rotação dos trímeros catalíticos de 11° em direções opostas ao redor do eixo de ordem 3. Simultaneamente, cada um dos dímeros regulatórios é rodado por 9° ao redor do eixo de ordem 2 correspondente e deslocado 1,4 Å ao longo deste eixo. O ajuste entre a curva obtida por este modelo e os dados experimentais é mostrado na figura 8.3.

O desvio quadrático médio entre as coordenadas dos C_α do estado cristalino R e T é 6,2 Å. O mesmo valor para o modelo encontrado e o estado T é de 0,99 Å. Estes valores indicam que as mudanças globais que acompanham a transição T→R em solução são aproximadamente 35% maiores que aquelas em cristais.

O espalhamento de raios X em soluções fornecem informações de baixa resolução e sua interpretação em termos de estruturas com alta resolução requer

cautela. A concordância observada para o estado T constitui então somente um forte argumento em favor da identidade entre as estruturas cristalinas e em solução, e não uma prova inequívoca. Para o estado R, no entanto, a significativa discordância com os dados experimentais prova uma diferença entre a estrutura quaternária entre as formas cristalinas e em solução. Nenhuma tentativa foi feita de modelar possíveis mudanças na estrutura terciária e a interpretação foi restrita a movimentos de corpo rígido das sub-unidades, de maneira similar ao primeiro modelamento do estado R em cristais (Ladner *et al.*, 1982).

Figura 8.5: Valor χ -quadrado *versus* a metade do deslocamento entre as sub-unidades catalíticas.

A ligação de substratos e mudanças conformacionais em enzimas alostéricas envolvem somente interações não covalentes de baixa energia. A energia de ativação da transição da estrutura quaternária é então também baixa e as forças de empacotamento originárias de ligações não covalentes entre moléculas vizinhas pode distorcer a arquitetura de proteínas selecionadas evolucionariamente para uma fácil reorganização. Esta pode ser a causa das distorções observadas na simetria molecular destes oligômeros nos cristais de ATCase. De fato, os três eixos de ordem 2 nestes

cristais são não-cristalográficos e apresentam somente uma simetria D_3 aproximada. Análises mostram que os desvios ocasionados pela simetria não cristalográfica na estrutura para o estado T são, entretanto, locais e correspondem a resíduos individuais, enquanto que as grandes discrepâncias no estado R afetam domínios inteiros, como apontado previamente (Krause; Volz; Lipscomb, 1985). Isto fortemente sugere que o estado R cristalino seja distorcido quando comparado com o meio isotrópico em solução.

No modelo para o estado R em solução os trímeros catalíticos estão praticamente em posição eclipsada. Por causa do aumento da separação, a preservação dos contatos entre os *loop* 240s implicariam na adoção de uma nova conformação mais estendida. Os contatos entre os átomos na interface entre as sub-unidades permanecem praticamente inalterados.

O presente estudo claramente ilustra a validade, especialmente para macromolécula com multi-domínios, de testar as estruturas cristalinas com dados de SAXS. A integração das informações obtidas por este trabalho em modelos teóricos deverá se consistir em mais uma ferramenta para a compreensão do mecanismo de cooperatividade e regulação em proteínas alostéricas.

CAPÍTULO 9 - CONCLUSÕES E PERSPECTIVAS FUTURAS

Apesar das vantagens do estudo de sistemas biológicos em solução é preciso sempre lembrar que o espalhamento de raios X a baixo ângulo é uma técnica com baixa resolução e, conseqüentemente, a quantidade de informações obtidas é limitada. Informações adicionais, obtidas por técnicas como cristalografia de proteínas, ressonância magnética nuclear ou microscopia eletrônica, ajudam a restringir o problema e dessa forma a chegar a resultados mais conclusivos. Foram os casos para a inibidora de α -amilase, onde a determinação do envelope com razoável resolução só foi possível através da imposição da simetria 222, para a hexocinase, onde somente os casos de dimerização cristalografica foram testados, e para a aspartato transcarbamilase, onde se explorou somente os movimentos de corpo rígido. Estes estudos exemplicam claramente as limitações e o valor dos estudos feitos em solução.

O trabalho desenvolvido nesta tese faz parte de um projeto mais ambicioso, que é o de formar um pacote completo de programas para a análise de dados de SAXS. Este projeto abrange a comparação de dados estruturais obtidos por diferentes técnicas, modelagem de estruturas quaternárias baseados em dados experimentais, determinação do envelope molecular de macromoléculas em solução e a criação de um banco de dados com envelopes moleculares. Uma boa parte destes programas já foram escritos, porém o projeto ainda continua em andamento.

Pretende-se no futuro desenvolver as seguintes atividades:

- Implementar os programas já desenvolvidos nos laboratórios de SAXS do estado de São Paulo e aplicar estes em proteínas que vem sendo estudadas no grupo de

cristalografia do Instituto de Física de São Carlos e outros lugares que venham a mostrar interesse pela técnica de SAXS.

- Implementar o método do cubo no programa *CRY SOL*.
- Adaptar o método de determinação do envelope molecular para processar os dados de difração de cristais de proteína.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abregó, J.R.B. *Estudo de Macromoléculas de interesse biológico em solução por espalhamento de raios X a baixo ângulo*. São Carlos 1987, 30p. Tese - Instituto de Física e Química de São Carlos, Universidade de São Paulo.
- Altman, R.R.; Ladner, J.E.; Lipscomb, W.N. *Biochem. Biophys. Res. Commun.*, **108**, 592-595, 1982.
- Badger, J. *Biophys. J.*, **65**, 1656-1659, 1993.
- Bernstein, F.C.; Koetzle T.F.; Williams, G.J.B.; Meyer, E.F.Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.*, **112**, 535-542, 1977.
- Cheng, X.; Schoenborn, B. P. *Acta Cryst.*, **B46**, 195-208, 1990.
- Damaschun, G.; Müller, J.J.; Pürschel, H.V. *Monatsh. Chem.*, **99**, 2343-2348, 1968.
- Dennis, J.E.Jr.; Gay, D.M.; Welsch, R.E. *An Adaptive nonlinear least-square algorithm, ACM Trans. Math. Software*. 1980.
- Diamond, R. *J. Mol. Biol.*, **82**, 371-391, 1974
- Edmonds, A.R. *Angular Momentum in Quantum Mechanics*. Princeton: Princeton Univ. Press, 1957.
- Fedorov, B.A.; Ptitsyn, O.B.; Voronin, L.A. *FEBS Lett.*, **28**, 188-190, 1972.
- Fedorov, B.A.; Denesyuk, A.I. *J. Appl. Cryst.*, **11**, 473-477, 1978.
- Feigin, L.A.; Svergun, D.I. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Plenum Press, New York. 1987. Cap. 1, p. 22.
- Fetler, L.; Tauc, P.; Hervé, G.; Moody, M.F.; Vachette, P. *J. Mol. Biol.*, **251**, 243-255, 1996.
- Fraser, R.D.; MacRae, T.P.; Suzuki, E. *J. Appl. Cryst.*, **11**, 693-694, 1978.
- Gerhart, J.C.; Schachman, H.K. *Biochemistry*, **7**, 538-552, 1968.
- Glatter, O.; Kratky, O. *Small-Angle X-Ray Scattering*, Academic Press, London.
- Granum, P.E.; Whitaker, J.R. *J. Food Biochem.*, **1**, 385-401, 1977.
- Grossmann, G.; Abraham, Z.; Adman, E.; Neu, M.; Eady, R.; Smith, B.; Hasnain, S. *Biochemistry*, **32**, 7360-7366, 1993.
- Guinier, A. *Ann. Phys.*, **12**, 161-237, 1935.

- Guinier, A.; Fournet, G. *Small-Angle Scattering of X-ray*, Wiley, New York, 1955.
- Harrison, S.C. *J. Mol. Biol.*, **42**, 457-483, 1969.
- Hubbard, S.R.; Hodgson, K.O.; Doniach, S. *J. Biol. Chem.*, **263**, 4151-4158, 1987.
- International Tables for X-ray Crystallography. Vol.III. Birmingham: Kynoch Press, 1968.
- International Tables for X-ray Crystallography. Vol.IV. Birmingham: Kynoch Press, 1974.
- Kantorowitz, E.R.; Lipscomb, W.N. *Science*, **241**, 669-674, 1988.
- Ke, H.M.; Lipscomb, W.N.; Cho, Y.; Honzatko, R.B. *J.Mol. Biol*, **204**, 725-748, 1988.
- Krause, K.L.; Volz, K.W.; Lipscomb, W.N. *Proc. Natl. Acad. Sci. USA*, **82**, 1643-1647, 1985.
- Ladner, J.E.; Kitchell, J.P.; Honzatko, R.B.; Ke, H.M.; Volz, K.W.; Kalb, A.J.; Ladner, R.C.; Lipscomb, W.N. *Proc. Natl. Acad. Sci. USA*, **79**, 3125-3128, 1982.
- Langridge, R.; Marvin, D.A.; Seeds, W.E.; Wilson, H.R.; Cooper, C.W.; Wilkins M.H.F.; Hamilton, L.D. *J. Mol.Biol.*, **2**, 38-64, 1960.
- Lattman, E.E. *Proteins*, **5**, 149-155, 1989.
- Mittelbach, P.; Porod, G. *Acta Phys. Austriaca*, **14**, 185-211, 405-439, 1961.
- Mittelbach, P.; Porod, G. *Acta Phys. Austriaca*, **15**, 122-147, 1962.
- Monaco, H.L.; Crawford, J.L.; Lipscomb, W.N. *Proc. Natl.Acad. Sci. USA*, **75**, 5276-5280, 1978.
- Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.*, **12**, 88-118, 1965.
- Moody, M.F.; Vachette, P.; Foote, A.M. *J. Mol. Biol.*, **133**, 517-532, 1979.
- Müller, J.J. *J. Appl. Cryst.*, **16**, 74-82, 1983.
- Müller, J.J.; Schrauber, H. *J. Appl. Cryst.*, **25**, 181-191, 1992.
- Ninio, J.; Luzzati, V.; Yaniv, M. *J. Mol. Biol.*, **71**, 217-229, 1972.
- Oliva, G.; Iulek, J.; Ida, E.I. *Acta Cryst.*, **D52**, imprimindo, 1996.
- Patterson, A. L. *Phys. Rev.*, **46**, 372, 1935.
- Pavlov, M.Yu.; Fedorov, B.A. *Biopolymers*, **22**, 1507-1522, 1983.
- Pavlov, M.Yu.; Sinev, M.A.; Timchenko, A.A.; Ptitsyn, O.B. *Biopolymers*, **25**, 1385-1397, 1986.

- Perkins, S.J. *Eur. J. Biochem.*, **157**, 169-180, 1986.
- Pickover, C.A.; Engelman, D.M. *Biopolymers*, **21**, 817-831, 1982.
- Rolbin, J.A.; Feigin, L.A.; Scedrin, B.M. *Appar. Metody Rentgenovskogo Anal.*, **9**, 46-50, 1971.
- Schoenborn, B.P. *J. Mol. Biol.*, **201**, 741-749, 1988.
- Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana. 1949.
- Silano, V. *Enzimes and their Role in Cereal Technology*, Am. Ass. Cer. Chem., 1987.
- Steitz, T.A.; Fletterick, R.J.; Anderson, W.F.; Anderson, C.M. *J. Mol. Biol.*, **104**, 197-222, 1976.
- Stevens, R.C.; Gouaux, J.E.; Lipscomb, W.N. *Biochemistry*, **29**, 7691-7701, 1990.
- Strobl, S.; Mühlhahn, P.; Berstein, R.; Wilschek, R.; Maskos, K.; Wunderlich, M. Huber, R.; Glockshuber, R.; Holak, T.A. *Biochemistry*, **34**, 8281-8293, 1995.
- Stuhrmann, H.B.; Kirste, R.G. *Z. Phys. Chem. Frankfurt*, **46**, 247-250, 1967.
- Stuhrmann, H.B. *Acta Cryst.*, **A26**, 297-306, 1970a.
- Stuhrmann, H.B. *Z. Phys. Chem. Frankfurt*, **72**, 177-184, 185-198, 1970b.
- Svergun, D.I.; Stuhrmann, H.B. *Acta Cryst.*, **A47**, 736-744, 1991.
- Svergun, D.I. *J. Appl. Cryst.*, **24**, 485-492, 1991.
- Svergun, D.I. *Acta Cryst.*, **A50**, 391-402, 1994.
- Svergun, D.I.; Barberato, C.; Koch, M.H.J. *J. Appl. Cryst.*, **28**, 768-773, 1995.
- Svergun, D.I.; Barberato, C.; Koch, M.H.J.; Fetler, L.; Vachette, P. *Proteins*, submetido, 1996.
- Svergun, D.I.; Volkov, V.; Kozin, M.; Barberato, C.; Stuhrmann, H.B. *J. Appl. Cryst.*, submetido, 1996.
- Svergun, D.I.; Semenyuk, A.V.; Feigin, L.A. *Acta Cryst.*, **A44**, 244-250, 1988.
- Tanki, N.; Thorton, J.M.; Goodfellow, J.M. *J. Mol. Biol.*, **202**, 637-657, 1988.
- Wilson, A.J.C. *X-ray Optics*, Wiley, New York, 1949.

APÊNDICE A - TRATAMENTO DE DADOS EXPERIMENTAIS

Esta apêndice irá descrever os procedimentos padrões de tratamento de dados experimentais na câmara X33 do European Molecular Biology Laboratory (EMBL) no HASYLAB em Hamburgo, Alemanha. Antes porém, algumas particularidades envolvidas na obtenção de uma curva de SAXS serão brevemente abordadas.

Devido a imperfeições na sua fabricação os detectores sensíveis a posição geralmente apresentam regiões em que a detecção de fótons de raios X é mais ou menos favorecida. As curvas medidas com estes instrumentos apresentam portanto distorções. Podemos superar este problema através da divisão ponto a ponto dessas curvas pela resposta do detector, sendo a resposta do detector a curva resultante da exposição do detector a uma fonte de raios X com intensidade constante ($I(s) = C$).

Devido as variações da intensidade do feixe incidente cada curva medida deve ser normalizada pela intensidade do feixe incidente, desta forma as curvas de uma mesma amostra colhida em diferentes instantes estarão numa mesma escala.

Uma curva de SAXS pode mudar durante uma coleta por causa da degradação da amostra devido aos danos por radiação e mesmo problemas no sistema de aquisição. Estas mudanças podem ser monitoradas através da coleta sucessiva de várias curvas e sua posterior comparação.

O espalhamento devido somente ao volume de solvente irradiado (Capítulo 5) é comumente obtido substituindo o volume de solução irradiado (proteína + solvente) por um igual volume preenchido somente com o solvente utilizado na solução de proteína.

Tendo em vista as informações dadas acima, as curvas de SAXS são obtidas no EMBL-Hamburg seguindo-se os seguintes passos:

a) São colhidas N curvas de resposta do detector (geralmente $N=10$). O tempo de exposição para cada curva é geralmente de 1 minuto. A intensidade do feixe incidente é avaliada através de uma câmara de ionização no término da coleta de cada uma das curvas e armazenada. Cada uma das curvas, $I_k(s_j)$, é normalizada pela intensidade incidente (I_{inc}) correspondente:

$$I_k(s_j) = \frac{I_k(s_j)}{I_{inc}} \quad \text{A.1}$$

onde o índice k representa a k -ésima curva colhida e o índice j representa o j -ésimo vetor de espalhamento. Os erros associados a cada curva, $\sigma_k(s_j)$, são avaliados através da seguinte fórmula:

$$\sigma_k(s_j) = I_k(s_j) I_{inc} \sqrt{\frac{1}{I_k(s_j)} + \frac{1}{I_{inc}}} \quad \text{A.2}$$

b) Todas as curvas são comparadas com a primeira curva coletada tomada como padrão. O critério utilizado na comparação é o teste do χ quadrado:

$$\chi_k = \sqrt{\frac{1}{(N_p - 1)} \sum_{j=1}^{N_p} \frac{(I_1(s_j) - I_k(s_j))^2}{(\sigma_1^2 + \sigma_k^2)}} \quad \text{A.3}$$

onde χ_k é o χ quadrado referente a comparação entre a primeira e a k -ésima curva e Np é o número de pontos de cada curva. Desta forma, se o χ_k for maior que o estipulado pelo usuário (o valor utilizado normalmente é 0.1) a k -ésima curva é descartada.

c) As N' curvas que passaram pela comparação feita em b) são promediadas:

$$I_{\text{det}}(s_i) = \frac{1}{N'} (I_1(s_j) + I_2(s_j) + \dots + I_{N'}(s_j)) \quad \text{A.4}$$

onde $I_{\text{det}}(s_j)$ é a resposta do detector e

$$\sigma_{\text{det}}(s_i) = \frac{1}{N'} \sqrt{\sigma_1(s_j)^2 + \sigma_2(s_j)^2 + \dots + \sigma_{N'}(s_j)^2} \quad \text{A.5}$$

é seu respectivo erro associado.

d) São colhidas N curvas da solução de proteínas e N curvas de solvente. Todas recebem o mesmo tratamento dado para a obtenção da curva de resposta do detector (a-c). A curva resultante para a solução de proteínas é designada $I_{\text{prot}}(s_j)$ e a curva do solvente é designada $I_{\text{sol}}(s_j)$.

e) A curva $I_{\text{prot}}(s_j)$ é corrigida pela resposta do detector, $I_{\text{det}}(s_j)$, utilizando as seguintes equações:

$$I'_{prot}(s_j) = \frac{I_{prot}(s_j)}{I_{det}(s_j)} \quad A.6$$

$$\sigma'_{prot}(s_j) = \frac{I_{prot}(s_j)}{I_{det}(s_j)} \sqrt{\frac{\sigma_{prot}(s_j)}{I_{prot}(s_j)} + \frac{\sigma_{det}(s_j)}{I_{det}(s_j)}} \quad A.7$$

f) A curva do solvente é também corrigida pela resposta do detector como em e).

g) A curva devida somente ao espalhamento da proteína, $I_{exp}(s_j)$, é obtida subtraindo da curva obtida em e) a curva obtida em f).

$$I_{exp}(s_j) = I'_{prot}(s_j) - I'_{sol}(s_j) \quad A.8$$

$$\sigma_{exp}(s_j) = \sqrt{\sigma_{prot}(s_j)^2 + \sigma_{sol}(s_j)^2} \quad A.9$$