

**Universidade Estadual de Campinas**

**Instituto de Física “Gleb Wataghin”**

**Estudos de macromoléculas biológicas  
parcialmente desestruturadas usando  
Espalhamento de Raios-X**

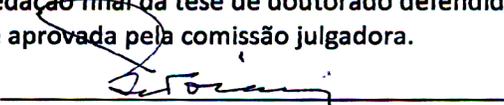
**Júlio César da Silva**

Orientadora:

Professora Dra. Íris L. Torriani

Tese de Doutorado apresentada no Instituto de Física “Gleb Wataghin” da  
Universidade Estadual de Campinas como partes dos requisitos exigidos para  
a obtenção do título de Doutor em Física.

Este exemplar corresponde à redação final da tese de doutorado defendida  
pelo aluno Júlio César da Silva e aprovada pela comissão julgadora.

  
\_\_\_\_\_  
Prof. Dra. Iris L. Torriani

**Campinas – São Paulo**

**2010**

A presente pesquisa foi desenvolvida no Instituto de Física “Gleb Wataghin” da UNICAMP e no Laboratório Nacional de Luz Síncrotron, com recursos próprios, do CNPq e da FAPESP.

### FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IFGW - UNICAMP

Si38e Silva, Júlio César da  
Estudos de macromoléculas biológicas parcialmente  
desestruturadas usando espalhamento de raios-X / Júlio César da  
Silva. -- Campinas, SP : [s.n.], 2010.

Orientador: Iris Concepción Linares de Torriani.  
Tese (doutorado) - Universidade Estadual de Campinas,  
Instituto de Física “Gleb Wataghin”.

1. Raios X – Espalhamento a baixo ângulo. 2. Proteínas  
intrinsecamente desestruturadas. 3. Proteínas multidomínios.  
4. Flexibilidade intramolecular. I. Torriani, Iris Concepción Linares  
de. II. Universidade Estadual de Campinas. Instituto de Física “Gleb  
Wataghin”. III. Título.

(vsv/ifgw)

- **Título em inglês:** Study of partially unstructured macromolecules using X-ray scattering
- **Palavras-chave em inglês (Keywords):**
  1. SAXS (Small angle X-ray scattering)
  2. Intrinsically unstructured proteins
  3. Multidomain proteins
  4. Intramolecular flexibility
- **Área de Concentração:** Física
- **Titulação:** Doutor em Ciências
- **Banca Examinadora:**

Prof<sup>a</sup> Iris Concepción Linares de Torriani  
Prof. Antonio Rubens Britto de Castro  
Prof. Marcus Aloizio Martinez de Aguiar  
Prof. Luis Mauricio Trambaioli da Rocha e Lima  
Prof. Marcelo Matos Santoro
- **Data da Defesa:** 25-02-2010
- **Programa de Pós-Graduação em:** Física



Secretaria de Pós-Graduação - Tel: (19) 3521-5305 FAX: (19) 3521-4142  
MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE **JÚLIO CÉSAR DA SILVA - RA 009027**, APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA "GLEB WATAGHIN" DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 25/02/2010

**COMISSÃO JULGADORA:**

Prof. Dra. Iris Concepcion Linares de Torriani – DFMC/IFGW/UNICAMP  
(Orientadora do Candidato)

Prof. Dr. Antonio Rubens Britto de Castro – DFMC/IFGW/UNICAMP

Prof. Dr. Marcus Aloizio Martinez de Aguiar – DFMC/IFGW/UNICAMP

Prof. Dr. Luis Mauricio Trambaioli da Rocha e Lima – Faculdade de Farmácia/UFRJ

Prof. Dr. Marcelo Matos Santoro – Instituto de Ciências Biológicas/UFGM



# Agradecimentos

- À Professora Dra. Íris Torriani pela orientação, ensino, muita paciência e dedicação com os quais a realização deste trabalho foi possível
- Ao Dr. Tomás Sigfrido Plivelic, ao Dr. Fernando Queiruga Rey e ao Rodrigo Martinez pelo apoio, amizade e ajuda na montagem da instrumentação da linha de luz durante as experiências deste trabalho.
- Ao professor Dr. Jörg Kobarg pela colaboração em diversos trabalhos desta tese, pelo fornecimento das amostras das proteínas FEZ1, Ki-1/57 e hnRNP-Q, e também por permitir usar o laboratório de biologia molecular do Laboratório Nacional de Biociências (LNBio) para a expressão e purificação de proteínas.
- Ao Dr. Cristiano Luis Pinto de Oliveira pelas importantes dicas computacionais e de trabalho, e também pela amizade.
- Ao Marcos Alborghetti, doutorando, e à Ariane Furlan, aluna de iniciação científica, ambos alunos do Dr. Jörg Kobarg no LNBio, por me ensinarem e auxiliarem na expressão e purificação de proteínas.
- Ao Dr. Daniel Lanza, pelas importantes discussões e colaboração no trabalho com a proteína FEZ1.
- Ao Dr. Gustavo Bressan, pelas importantes discussões e colaboração no trabalho com a proteína Ki-1/57.
- Ao Dr. Alex Quaresma, pelas importantes discussões e colaboração no trabalho com a proteína hnRNP-Q.
- Ao professor Dr. Carlos Ramos (Instituto de Química da UNICAMP) e ao professor Dr. Júlio Borges (Instituto de Química da USP de São Carlos) pelo fornecimento das amostras das proteínas HSP40 e pela colaboração no trabalho.
- Ao Dr. Esftraios Mylonas (Spring-8, Japão) pelas importantes dicas sobre proteínas nativamente desenhadas.
- À minha família que sempre me apoiou em cada passo deste trabalho.
- Aos meus amigos, cujos nomes não estão aqui porque eu precisaria de muitas folhas para caber todos eles, porém saibam que sou extremamente grato a todos.

- Ao professor Dr. Alex Antonelli e ao professor Dr. Fernando Cerdeira, ambos do IFGW/UNICAMP, pela participação na banca examinadora de meu Exame de Qualificação de Doutorado.
- Ao professor Dr. Douglas Soares Galvão e ao professor Dr. Mario Noboru Tamashiro, ambos do IFGW/UNICAMP, pela participação na banca examinadora de meu Seminário de Pré-requisito de Defesa de Tese de Doutorado.
- Aos funcionários do Laboratório de Cristalografia Aplicada e Raios-X (LCARX) da Unicamp: Ana Carolina, Carlos Leite e José Luis.
- Aos funcionários da secretaria de pós-graduação e da biblioteca do Instituto de física “Gleb Wataghin” da Unicamp (IFGW/UNICAMP). Também aos funcionários do Laboratório Nacional de Luz Síncrotron (LNLS) e do Laboratório Nacional de Biociências (LNBio).
- Ao CNPq e a FAPESP pelo apoio financeiro através de bolsas de doutorado.
- Ao Laboratório Nacional de Luz Síncrotron (LNLS) e ao Laboratório Nacional de Biociências (LNBio) pelo uso das instalações experimentais.
- A todas as pessoas que colaboraram de maneira direta ou indireta para a realização deste trabalho e confecção desta tese de doutorado.

A Deus.

À minha família pelo apoio,  
em especial à minha mãe e ao  
meu pai (*in memoriam*).

Dedico.



“Nunca deixe que alguém lhe diga que você não pode fazer algo. Se você tem um sonho tem que protegê-lo. As pessoas que não podem fazer por si mesmas, dirão que você não consegue”

(Christopher Gardner no filme “*A procura da felicidade*”)

“A vida vai ficando mais dura perto do topo”

(Friedrich Nietzsche)

“Quando me faltam palavras, ainda me resta a música”

(Júlio César da Silva)



# Resumo

As técnicas de caracterização estrutural de macromoléculas tradicionais se baseiam no fato de uma macromolécula possuir uma conformação compacta e estruturada. Partes flexíveis ou regiões desordenadas têm sido sempre consideradas como grandes obstáculos para técnicas como a cristalografia de raios-X e a ressonância magnética nuclear (RMN). A necessidade de entender a atividade funcional de proteínas nativamente desenhadas e de proteínas flexíveis com múltiplos domínios tem adquirido grande importância recentemente, mesmo porque essas proteínas desafiam o paradigma de que uma proteína precisa de uma estrutura bem definida para ser funcional. É bem nesse ponto que a técnica de espalhamento de raios-X a baixos ângulos (SAXS) surge oferecendo ferramentas únicas para realizar estudos de macromoléculas flexíveis ou parcialmente desestruturadas, com aplicações muito bem sucedidas em polímeros, matéria mole e macromoléculas em solução.

Neste trabalho de tese decidimos enfrentar o desafio de caracterizar proteínas que não possuem uma estrutura bem definida. A teoria do espalhamento mereceu especial cuidado para se adequar tanto aos métodos experimentais da técnica quanto aos tratamentos matemáticos em cálculos usados para estudar esse tipo de proteínas. Apresentamos aqui o estudo de duas proteínas pertencentes à classe das proteínas nativamente desenhadas: (1) a proteína FEZ1, que é necessária para o crescimento de axônios; (2) a proteína Ki-1/57, que é encontrada em diversas células com câncer principalmente em tumores do sistema linfático. Estudamos também algumas proteínas com múltiplos domínios conectados por regiões flexíveis e que são: (1) duas chaperonas da classe das HSP40 (proteínas Sis1 e Ydj1) juntamente com construções onde alguns domínios dessas proteínas foram cortados; (2) a proteína ribonucléica heterogênea hnRNP-Q que está envolvida em importantes funções do RNA. Experiências de SAXS foram realizadas, fornecendo parâmetros dimensionais e informações de forma dessas proteínas em solução. Modelos de baixa resolução das possíveis conformações foram calculados a partir das curvas de SAXS usando métodos de modelagem *ab initio* combinados com modelagem de corpos rígidos. Os resultados forneceram informações importantes para elucidar as funções biológicas dessas proteínas.

É importante ressaltar que, para realizar os estudos com proteínas em solução, é necessário contar com uma instrumentação adequada e devidamente montada para a aplicação da técnica de SAXS. Para isso, durante o período de desenvolvimento deste doutorado houve um grande investimento na montagem, teste e caracterização de instrumentos, junto à equipe de profissionais do Laboratório Nacional de Luz Síncrotron (LNLS), completando o comissionamento da estação experimental SAXS2 do LNLS.



# Abstract

The traditional techniques for structural characterization of macromolecules are based on a compact and structured conformation of the macromolecule. Flexible or disordered regions have usually been regarded as a great hindrance to techniques like X-ray protein crystallography and nuclear magnetic resonance (NMR). The need to study functional activity of natively unfolded proteins and flexible multidomain proteins came to the light rather recently, defying the classical structure–function paradigm where a protein must have a well-defined 3-D structure to be functional. In this type of situation, the small-angle X-ray scattering (SAXS) technique appears as a unique tool to deal with this problem. Indeed, the application of SAXS methods to the characterization of soft matter (e.g. polymers) and macromolecules in solution has already succeeded during the last years.

In this work we decided to face the challenge of characterizing proteins that do not have a well defined structure. The SAXS experimental technique as well as the mathematical methods and calculations needed special attention in order to be correctly applied to study the specific problem of unstructured proteins in solution. Thus, it was possible to find evidence of the structural details of these proteins and obtain a low resolution 3-D average structure. Here we present the study of two proteins that belong to the group of natively unfolded proteins: (1) The FEZ1 protein, which is necessary for axon growth, and (2) the proteins indentified as Ki-1/57, which is found in diverse cancer cells mainly in lymphatic systems tumors. We also studied some flexible multidomain proteins: (1) two chaperones from the groups of HSP40 (the proteínas Sis1 e Ydj1), and two mutant constructions where some domains were deleted; (2) the heterogeneous ribonucleoprotein hnRNP-Q which is related to an array of important functions of RNA. Several SAXS experiments were performed providing overall parameters and important shape information about those proteins in solution. Low resolution models for the possible conformations of these proteins were restored from the SAXS curves using *ab initio* modeling methods combined with rigid body modeling. The SAXS results provided a unique structural background for the biologists to deal with the function of these proteins.

SAXS experiments with proteins in solution demand the use of a specific instrumentation properly developed for those studies. So, it is important to mention that, throughout the duration of this doctorate, specific instrumentation development and testing was done together with the technical staff of the Brazilian Synchrotron Light Laboratory (LNLS, Campinas, SP, Brazil), collaborating with the commissioning of the new SAXS2 workstation, completed in 2008.



# Índice

<b>Resumo</b> .....	<b>xi</b>
<b>Abstract</b> .....	<b>xiii</b>
<b>Índice de figuras</b> .....	<b>xix</b>
<b>Índice de tabelas</b> .....	<b>xxi</b>
<b>Lista de abreviações</b> .....	<b>xxiii</b>
<b>Introdução</b> .....	<b>1</b>
<b>Capítulo 1 – Breve descrição da teoria e métodos utilizados nesta tese</b> .....	<b>7</b>
1.1. Escopo .....	7
1.2. Princípios de espalhamento de raios-X a baixos ângulos.....	8
1.3. Espalhamento de raios-X por partículas em solução.....	13
1.4. Parâmetros estruturais globais.....	14
1.4.1. Intensidade na origem.....	14
1.4.2. Raio de giro e $I(0)$ : Aproximação de Guinier.....	15
1.4.3. Raio de giro, máxima dimensão e $I(0)$ : função $p(r)$ .....	16
1.4.4. Partículas flexíveis: Aproximação de Debye .....	17
1.4.5. Aproximação de Porod.....	18
1.4.6. Partículas flexíveis: aproximação assintótica para grandes valores de $q$ .....	19
1.4.7. Estimativa de massa molecular da partícula.....	19
1.4.7.1. Utilizando água como padrão para determinação de massa molecular .....	20
1.4.7.2. Utilizando uma proteína secundária como padrão .....	21
1.5. Representação de Kratky da intensidade: $q^2I(q)$ vs. $q$ .....	22
1.6. Tratamento dos dados de SAXS.....	23
1.7. Processamento dos dados: Transformada de Fourier da intensidade.....	25
1.8. Quantidade de informação: canais de Shannon.....	26
1.9. Determinação de forma: modelagem tridimensional a partir dos dados de SAXS .....	27
1.9.1. Estratégia geral de modelagem.....	28
1.9.2. Método de minimização: algoritmo de recozimento simulado.....	29
1.9.3. Modelagem <i>ab initio</i> com átomos fictícios .....	30
1.9.4. Modelagem <i>ab initio</i> com resíduos fictícios .....	32
1.9.5. Modelagem de corpo rígido combinada com cálculos <i>ab initio</i> .....	33
1.9.6. Método de otimização de conjuntos de conformações .....	35

1.10. Analisando diferentes soluções no processo de modelagem.....	36
1.11. Breve revisão sobre proteínas.....	38
1.12. Diferentes níveis estruturais das proteínas.....	38
<b>Capítulo 2 – Instrumentação dedicada à realização de experiências de espalhamento de raios-X no Laboratório Nacional de Luz Síncrotron.....</b>	<b>41</b>
2.1. Introdução.....	41
2.2. Esquema Geral de uma linha de luz dedica à realização de experiências de SAXS.....	42
2.3. Características da linha D02A-SAXS2 do LNLS.....	43
2.4. Elementos ópticos da linha D02A-SAXS2 do LNLS.....	44
2.4.1. Espelho de raios-X da linha de luz.....	44
2.4.2. Monocromador de raios-X da linha de luz D02A-SAXS2.....	45
2.4.3. Conjuntos de fendas.....	47
2.4.4. Monitores de intensidade, filtros e absorvedor do feixe direto.....	48
2.5. Características dos detectores disponíveis na linha para experiências de SAXS.....	49
2.5.1. Detector linear a gás sensível a posição.....	50
2.5.2. Detector CCD.....	50
2.5.3. Detector de área com a tecnologia de pixels híbridos de silício.....	51
2.6. Porta-amostras para soluções de macromoléculas biológicas.....	52
2.7. Rotinas computacionais de tratamento de dados.....	54
2.8. Alguns exemplos de dados obtidos na linha de luz D02-SAXS2 e comparação com a antiga linha de SAXS do LNLS.....	54
<b>Capítulo 3 – Proteínas nativamente desenvoladas.....</b>	<b>59</b>
3.1. Introdução.....	59
3.2. Proteína de enfeixamento e prolongamento Zeta-1 (FEZ1).....	62
3.2.1. Descrição da proteína FEZ1.....	62
3.2.2. Materiais e métodos.....	64
3.2.2.1. Preparação das amostras.....	64
3.2.2.2. Descrição das experiências de SAXS com as amostras de FEZ1 em condição nativa.....	64
3.2.2.3. Mudanças conformacionais da FEZ1 em função da concentração de desnaturante (uréia).....	65
3.2.2.4. Região de dimerização da FEZ1: o fragmento FEZ1(92-194).....	66
3.2.2.5. Modelagem tridimensional <i>ab initio</i> da conformação da proteína FEZ1 em solução.....	67

3.2.2.6. Modelagem da FEZ1(1-392) considerando a topologia da região com propensão à formação de <i>coiled-coils</i> .....	67
3.2.2.7. As conformações de FEZ1 em solução: método de otimização de conjuntos de conformações.....	68
3.2.3. Resultados .....	68
3.2.3.1. Resultados das experiências com a proteína na condição nativa.....	68
3.2.3.2. Resultados das mudanças conformacionais da proteína FEZ1 com a adição de uréia .....	71
3.2.3.3. Resultados das análises da região de dimerização de FEZ1: o fragmento FEZ1(92-194) .....	73
3.2.3.4. Resultados da modelagem da conformação da proteína FEZ1 .....	74
3.2.3.5. Modelagem da topologia de FEZ1 considerando a região com propensão à formação de motivos “ <i>coiled-coils</i> ” .....	77
3.2.3.6. Resultados obtidos com o método de otimização de conjuntos de conformações.....	79
3.2.4. Discussões sobre os resultados obtidos com a proteína FEZ1 .....	83
3.3. Proteína regulatória humana Ki-1/57 tem características de proteína intrinsecamente desestruturada.....	85
3.3.1. Descrição da proteína Ki-1/57.....	85
3.3.2. Materiais e métodos.....	86
3.3.2.1. Preparação das amostras.....	86
3.3.2.2. Descrição das experiências de SAXS com a proteína Ki-1/57 .....	87
3.3.2.3. Análises dos dados de SAXS e modelagem da conformação da proteína Ki-1/57(122-413) .....	87
3.3.2.4. As conformações de Ki-1/57 em solução: método de otimização de conformações.....	88
3.3.3. Resultados .....	88
3.3.3.1. Principais resultados de SAXS da proteína Ki-1/57 .....	88
3.3.3.2. Modelagem da conformação da proteína humana Ki-1/57(122-413)...	89
3.3.3.3. Método de otimização de conjunto de conformações.....	91
3.3.4. Discussões sobre os resultados obtidos com a proteína ki-1/57 .....	93
<b>Capítulo 4 – Proteínas flexíveis com múltiplos domínios .....</b>	<b>95</b>
4.1. Introdução.....	95
4.2. Proteínas de choque térmico (HSP40) Sis1 e Ydj1: chaperonas moleculares.....	97
4.2.1. Descrição das proteínas de choque térmico de 40 kDa (HSP40) .....	97

4.2.2. Materiais e métodos.....	100
4.2.2.1. Preparação das amostras.....	100
4.2.2.2. Descrição das experiências de SAXS .....	100
4.2.2.3. Modelagem <i>ab initio</i> baseada nos dados de SAXS .....	101
4.2.2.4. Modelagem da topologia da organização de domínios na conformação das proteínas .....	101
4.2.3. Resultados .....	102
4.2.3.1. Resultados das experiências de SAXS .....	102
4.2.3.2. Resultados da modelagem <i>ab initio</i> das conformações das proteínas	104
4.2.3.3. Resultados da modelagem de corpos rígidos do posicionamento de domínios nas conformações das proteínas.....	106
4.2.4. Discussões sobre os resultados obtidos com as proteínas .....	107
4.3. A proteína ribonucléica heterogênea tipo Q (hnRNP-Q): uma proteína com múltiplos domínios .....	108
4.3.1. Descrição da proteína hnRNP-Q .....	108
4.3.2. Materiais e métodos.....	109
4.3.2.1. Preparação das amostras.....	109
4.3.2.2. Descrição das experiências de SAXS com a GST-hnRNP-Q(1-443). 109	
4.3.2.3. Modelagem <i>ab initio</i> baseada nos dados de SAXS da GST-hnRNP-Q(1-443).....	110
4.3.2.4. Modelagem da topologia do arranjo de domínios da proteína.....	110
4.3.3. Resultados .....	111
4.3.3.1. Resultados das experiências de SAXS com a proteína hnRNP-Q.....	111
4.3.3.2. Modelagem <i>ab initio</i> baseada da GST-hnRNP-Q(1-443).....	113
4.3.3.3. Modelagem da topologia do arranjo de domínios da proteína GST-hnRNP-Q(1-443) .....	114
4.3.4. Discussões sobre os resultados obtidos com a proteína hnRNP-Q .....	116
<b>Considerações finais.....</b>	<b>119</b>
<b>Lista de publicações científicas .....</b>	<b>123</b>
<b>Outras publicações científicas.....</b>	<b>124</b>
<b>Referências.....</b>	<b>125</b>

# Índice de Figuras

<b>Figura 1.1</b> – Representação esquemática do espalhamento por uma partícula fixa no espaço.....	9
<b>Figura 1.2</b> – Representação da função de distribuição de pares de distâncias $p(r)$ .....	12
<b>Figura 1.3</b> – Exemplos de representações de Kratky ( $q^2I(q)$ vs. $q$ ) da curva de intensidade para três tipos de estruturas: globular compacta, parcialmente desestruturada e desestruturada .....	23
<b>Figura 1.4</b> – Configurações de esferas densamente empacotadas do processo de modelagem <i>ab initio</i> com átomos fictícios. ....	32
<b>Figura 1.5</b> – Representação esquemática de uma proteína como um arranjo de resíduos fictícios para a modelagem <i>ab initio</i> .....	33
<b>Figura 1.6</b> – Representação típica de um modelo de uma proteína com múltiplos domínios obtido dos dados de SAXS usando a combinação de cálculos <i>ab initio</i> e modelagem de corpos rígidos .....	34
<b>Figura 1.7</b> – Esquemática do método de otimização de conjuntos de conformações de proteínas em solução .....	36
<b>Figura 1.8</b> – Organização hierárquica dos quatro diferentes aspectos das estruturas de proteínas .....	39
<b>Figura 2.1</b> – Fotografia mostrando a nova linha de luz (D02-SAXS2) dedicada à realização de experiências de SAXS do LNLS .....	42
<b>Figura 2.2</b> – Esquema geral do arranjo da instrumentação necessária para a realização de experiências de SAXS em uma linha de luz síncrotron.....	43
<b>Figura 2.3</b> – Desenho esquemático da linha D02A-SAXS2 .....	44
<b>Figura 2.4</b> – Fotografia do espelho de raios-X instalado na linha de luz D02A-SAXS2 dentro da câmara de vácuo .....	45
<b>Figura 2.5</b> – Monocromador de cristal curvado da linha de luz D02A-SAXS2.....	47
<b>Figura 2.6</b> – Desenho do módulo de fendas com os motores de passo e transladores .....	48
<b>Figura 2.7</b> – Porta-amostra para amostras líquidas com um cela de $1\text{ mm}$ de espessura e vedada por duas janelas de mica .....	53
<b>Figura 2.8</b> – Porta-amostra para amostras líquidas usando tubo capilar cilíndrico com $1\text{ mm}$ de diâmetro na seção transversal .....	53
<b>Figura 2.9</b> – Curvas experimentais de espalhamento da proteína lisozima medidas nas linhas D01A-SAXS1 e D02A-SAXS2.....	55
<b>Figura 2.10</b> – Comparação entre a curva experimental de espalhamento da lisozima medida na linha D02A-SAXS2 e a curva teórica calculada da estrutura cristalográfica da lisozima.....	56
<b>Figura 2.11</b> – Comparação entre a curva experimental de espalhamento da melalina em DMSO medida na linha D02A-SAXS2 usando um comprimento de onda de $\lambda = 1,045\text{ \AA}$ .....	57
<b>Figura 3.1</b> – Desenho esquemático do comprimento das seqüências de aminoácidos das amostras estudadas: FEZ1(1-392), FEZ1(1-227) e FEZ1(92-194).....	63
<b>Figura 3.2</b> – Curvas experimentais de SAXS para 6xHis-FEZ1(1-392), 6xHis-FEZ1(1-227) e FEZ1(92-194) 70	
<b>Figura 3.3</b> – Dados experimentais de SAXS mostrando mudanças conformacionais em FEZ1(1-392) e em FEZ1(1-227).....	73

<b>Figura 3.4</b> – Resultados de SAXS das experiências com o fragmento FEZ1(92-194) da região N-terminal da proteína FEZ1 .....	74
<b>Figura 3.5</b> – Modelos <i>ab initio</i> de baixa resolução obtidos dos dados de SAXS da FEZ1 calculados sem a imposição de simetria .....	76
<b>Figura 3.6</b> – Modelos de resíduos fictícios de baixa resolução obtidos dos dados de SAXS da FEZ1(1-392) através de cálculo <i>ab initio</i> e impondo simetria <i>P2</i> nos cálculos .....	76
<b>Figura 3.7</b> – Modelos da proteína FEZ1(1-392) levando em conta a topologia da região estruturada com propensão à formação de motivos “coiled-coils” obtidos dos dados de SAXS através da combinação de cálculos <i>ab initio</i> e modelagem global de corpos rígidos .....	78
<b>Figura 3.8</b> – Superposição do envelope molécula de baixa resolução calculado para FEZ1(1-392) através da combinação de métodos <i>ab initio</i> e modelagem de corpos rígidos sobre o modelo de resíduos fictícios .....	79
<b>Figura 3.9</b> – Resultados da aplicação do método de otimização de conjuntos de conformações aos dados de SAXS de FEZ1(1-392) em soluções com 0, 3 e 6 M de uréia .....	81
<b>Figura 3.10</b> – Resultados da aplicação do método de otimização de conjuntos de conformações aos dados de SAXS de FEZ1(1-227) em soluções com 0, 3 e 6 M de uréia .....	82
<b>Figura 3.11</b> – Resultados experimentais de SAXS obtidos com a proteína Ki-1/57(122-413) .....	89
<b>Figura 3.12</b> – Resultados da modelagem <i>ab initio</i> da proteína Ki-1/57(122-413) em solução a partir dos dados de .....	90
<b>Figura 3.13</b> – Resultados da aplicação do método de otimização de conjuntos de conformações aos dados de SAXS da proteína Ki-1/57(122-413) em solução .....	92
<b>Figura 4.1</b> – Representação esquemática das seqüências de aminoácidos e organização dos domínios das construções das HSP40s e mutantes .....	99
<b>Figura 4.2</b> – Resultados de SAXS para as proteínas Sis1, Sis1_Δ <sub>124-174</sub> , Sis1_Δ <sub>121-257</sub> e Ydj1_Δ <sub>106-255</sub> .....	103
<b>Figura 4.3</b> – Resultados dos cálculos <i>ab initio</i> dos modelos de átomos fictícios, resíduos fictícios e combinando cálculos <i>ab initio</i> com modelagem de corpos rígidos para Sis1 e suas mutantes .....	105
<b>Figura 4.4</b> – Resultados dos cálculos <i>ab initio</i> dos modelos de átomos fictícios, resíduos fictícios e combinando cálculos <i>ab initio</i> com modelagem de corpos rígidos para Ydj1_Δ <sub>106-255</sub> .....	106
<b>Figura 4.5</b> – Representação esquemática da organização dos domínios da proteína na construção GST-hnRNP-Q(1-443) .....	109
<b>Figura 4.6</b> – Dados experimentais de SAXS para a proteína GST-hnRNP-Q(1-443) em solução .....	112
<b>Figura 4.7</b> – Modelos <i>ab initio</i> para a conformação de GST-hnRNP-Q(1-443) em solução .....	114
<b>Figura 4.8</b> – Modelos de baixa resolução da organização dos domínios na conformação de GST-hnRNP-Q(1-443) em solução .....	115
<b>Figura 4.9</b> – Representação pictórica da estrutura proposta para a GST-hnRNP-Q(1-443) inspirada nos modelos de baixa resolução obtidos por SAXS .....	117

# Índice de Tabelas

<b>Tabela 3.1</b> – Parâmetros dimensionais das amostras de FEZ1(1-392), FEZ1(1-227) e FEZ1(92-194) obtidos a partir dos dados de SAXS.....	70
<b>Tabela 3.2</b> – Resultados das mudanças de raio de giro ( $R_g$ ) de FEZ1(1-392) e FEZ1(1-227) em soluções com uréia.....	72
<b>Tabela 3.3</b> – Parâmetros dimensionais do fragmento FEZ1(92-194) obtidos por SAXS antes e após a redução da ponte dissulfeto com TCEP .....	74
<b>Tabela 4.1</b> – Parâmetros dimensionais obtidos dos dados de SAXS para as proteínas e mutantes de HSP40: Sis1, Sis1_Δ <sub>124-174</sub> , Sis1_Δ <sub>121-257</sub> e Ydj1_Δ <sub>106-255</sub> .....	104



# Lista de abreviações

AcD: *Acidic Domain* (domínio ácido).

CD: Circular Dichroism (dicroísmo circular, em português).

Dmax: dimensão máxima da molécula.

DNA: ácido desoxirribonucléico.

DTT: Ditioneitol (agente redutor).

FEZ1: *Fasciculation and Elongation protein Zeta 1*.

GST: *Glutathione S-transferase* (proteína).

HCl: Ácido clorídrico.

His-tag: Cauda de histidina.

HSP: *Heat Shock Protein* (proteína).

hnRNP-Q: proteína humana regulatória que liga ao RNA.

kDa: kilo Daltons (unidade de massa molecular).

Da: Dalton – unidade de massa molecular onde 1 Da é equivalente a massa do átomo de H.

Ki-1/57: antígeno (proteína) humano de 57 kDa.

p53: proteína supressora de tumor.

PDB: Protein Data Bank (Base de estruturas de alta resolução de proteínas).

MDM2: *Mouse Double-Minute 2 protein* (proteína).

NaCl: Cloreto de Sódio.

Rg: raio de giro da molécula.

RMN: ressonância magnética nuclear.

RNA: ácido ribonucléico.

RRM: Domínio de ligação de RNA.

SAXS: *Small Angle X-ray Scattering* (Espalhamento de raios-X a baixos ângulos).

Sis1: Proteína de levedura (Chaperona).

SDS: *sodium dodecyl sulfate*.

STC1: proteína Staniocalcina 1.

TCEP: Tris(2-carboxietil)fosfina (agente redutor de ponte de sulfeto).

Tris: Tris(hidroximetil)aminometano.



# Introdução

O estudo estrutural de macromoléculas em solução é algo extremamente importante para o conhecimento das propriedades e funções de diversas moléculas que constituem os seres vivos. Compreendendo a sua estrutura, é possível entender a relação dessas macromoléculas com a causa ou cura de diversas doenças crônicas, como por exemplo, o câncer. Porém, algumas dessas macromoléculas possuem regiões flexíveis ou regiões desestruturadas, tais como, as proteínas intrinsecamente desestruturadas ou as proteínas flexíveis como múltiplos domínios. Em vista disso, realizar análises estruturais de proteínas desses tipos de proteína é uma difícil tarefa para as técnicas de alta-resolução devido às partes móveis presentes nas suas estruturas. Muitas técnicas de caracterização estrutural não permitem obter informações estruturais dessas macromoléculas. Isso se aplica também para o estudo de mudanças conformacionais, pois muitas técnicas não permitem acompanhar as alterações estruturais das moléculas. Nesse cenário, o espalhamento de raios-X a baixos ângulos (SAXS<sup>i</sup>) aparece como uma técnica essencial para esses tipos de estudos. A aplicação da técnica de SAXS à caracterização estrutural de proteínas flexíveis com mais de um domínio ou proteínas parcialmente desestruturadas permite identificar possíveis conformações para a proteína em solução, além de fornecer informações estruturais importantes sobre elas. Por essa razão, SAXS vem sendo aplicada com êxito ao estudo de desenovelamento de proteínas [Doniach, 2001; Hirai *et al.*, 2004; Hirai *et al.*, 1999; Chen *et al.*, 1996] e também ao estudo de proteínas flexíveis ou desestruturadas [Lanza *et al.*, 2009; Bressan *et al.*, 2008; Tidow *et al.*, 2007; Moncoq *et al.*, 2004], e a demanda nas estações experimentais de SAXS por estudos estruturais de proteínas tem aumentado significativamente nos últimos anos.

Sendo assim, vamos primeiro tentar compreender a importância das proteínas em nossas vidas. Quando observamos uma célula usando um microscópio ou analisando sua atividade elétrica e/ou bioquímica, nós estamos essencialmente observando proteínas. A célula é constituída, em sua maior parte, por proteínas. Elas não são somente os constituintes básicos dos quais as células são feitas, mas elas também executam quase todas as funções celulares. Por exemplo, as enzimas são proteínas que catalisam (i.e., aceleram) as reações químicas dentro da célula [Lehninger, 1995]. Proteínas na membrana plasmática formam canais e bombas que controlam a passagem de moléculas para dentro e para fora da célula [Cho & Stahelin, 2005]. Outras proteínas carregam mensagens de uma célula à outra, ou atuam como integradores de sinal que retransmitem conjuntos

---

<sup>i</sup> SAXS é uma abreviação que vem do nome da técnica em inglês: *Small-Angle X-ray Scattering*.

de sinais da membrana plasmática para o núcleo da célula [Cho & Stahelin, 2005, Cho, 2006]. Há ainda outras proteínas que servem como pequenas máquinas moleculares com partes móveis: a quinesina [Gross *et al.*, 2007; Kull *et al.*, 1996], por exemplo, impulsiona organelas através do citoplasma; já a topoisomerase [Champoux, 2001] pode desemaranhar moléculas de DNA embaraçadas. Outras proteínas especializadas atuam como anticorpos, toxinas, hormônios, fibras elásticas, moléculas que controlam a temperatura na célula, fontes de luminescência e outras funções. Antes de podermos entender como os genes funcionam, como os músculos contraem, como os nervos conduzem eletricidade, como os embriões se desenvolvem ou como nossos corpos funcionam, nós devemos primeiro nos ater ao entendimento das estruturas e funções das proteínas. Do ponto de vista químico, as proteínas são de longe as macromoléculas mais sofisticadas, estruturalmente e funcionalmente. Isso não é uma surpresa se nos dermos conta de que a estrutura e a química de cada uma delas vêm sendo desenvolvida e ajustada durante os bilhões de anos de história evolucionária.

Para entender esse complexo funcionamento das proteínas, é preciso ter em mente que as propriedades biológicas de uma proteína dependem de suas interações físicas com outras moléculas (proteínas com outras funções e características, hormônios, inibidores entre outras). Portanto, a caracterização estrutural das proteínas é muito importante para o entendimento dos processos celulares e como eles afetam os seres vivos. As técnicas experimentais de destaque para a caracterização estrutural são a cristalografia de proteínas e a ressonância magnética nuclear (RMN), devido à possibilidade de fornecer detalhes estruturais em nível atômico. Porém, ainda existem muitos paradigmas nesse cenário que precisam ser resolvidos. Para se ter uma idéia, em um momento em que o problema fundamental do enovelamento de proteínas ainda não foi completamente resolvido, já surge um grande novo desafio para o estudo de proteínas na era pós-genômica: as proteínas intrinsecamente desestruturadas (algumas vezes também chamadas de proteínas nativamente desenoveladas). Há cada vez mais evidências experimentais que comprovam a existência desse tipo de proteínas [Dyson & Wright, 2005; Fink, 2005; Tompa, 2002; Dyson & Wright, 1999]. Elas são macromoléculas com grandes regiões desordenadas (ou desestruturadas), podendo até haver mais desordem que ordem na sua estrutura. Elas realizam, porém, importantes funções celulares. Elas desafiam o princípio clássico de ter sua função ditada por sua estrutura tridimensional bem definida. Pelo contrário, muitas proteínas apresentam funções que requerem essa desordem conformacional, como por exemplo, funções de transporte, reconhecimento molecular, modificações pós-traducionais, sinalização celular e mecanismos regulatórios [Dyson & Wright, 2005; Tompa, 2002].

Nos últimos anos, o entendimento da organização conformacional de proteínas intrinsecamente desestruturadas não tem proporcionado somente um desafio intelectual nas áreas que vão desde a bioquímica física à mecânica estatística, mas também proporciona um imenso potencial de aplicações. Em particular, o estudo estrutural dessas proteínas é visto como um importante passo na compreensão de algumas doenças crônicas ainda sem cura, como o câncer [Tidow *et al.*, 2007], na engenharia de proteínas e no desenvolvimento de drogas para tratar essas doenças. Por essa razão, há uma clara necessidade para o desenvolvimento de abordagens não somente qualitativas, mas também quantitativas para a caracterização da estrutura de proteínas flexível e com regiões desordenadas em sua estrutura.

Atualmente, as análises estruturais de proteínas flexíveis, tais como as proteínas intrinsecamente desestruturadas, ainda apresentam grandes dificuldades para os pesquisadores, quase excluindo as técnicas de alta-resolução, como por exemplo, a cristalografia de raios-X (essas proteínas dificilmente cristalizam) e a ressonância magnética nuclear (muitas dessas proteínas possuem massa molecular superior a 50 kDa<sup>ii</sup>, que o limite superior para essa técnica atualmente). Nesse caso, a técnica que se destaca no estudo desse tipo de macromolécula é a de espalhamento de raios-X a baixos ângulos (SAXS). As experiências de SAXS podem eficientemente revelar a conformação espacial dessas proteínas ou a organização de domínios ligados por regiões desordenadas em proteínas com domínios múltiplos. Além disso, ainda permite o estudo de mudanças conformacionais de proteínas em solução. SAXS fornece parâmetros dimensionais importantes dessas proteínas, além de uma análise quantitativa do grau de compactidade (que pode ser correlacionado com o grau de enovelamento) de uma proteína flexível e/ou com grandes regiões desestruturadas.

SAXS é uma técnica que permite estudar características estruturais de tamanho coloidal (de 10 Å a 1000 Å) de moléculas em solução e suas organizações dentro de um material. O mais surpreendente da técnica de SAXS é a possibilidade de realizar uma modelagem de baixa resolução, usando cálculos *ab initio*, para se obter um modelo tridimensional para o envelope molecular de uma proteína. Essa reconstituição tridimensional fornece um modelo de baixa resolução, já que os dados de SAXS fornecem informações de dimensões coloidais sobre a forma da partícula. Isso é particularmente útil em biologia estrutural para o estudo estrutural e conformacional de proteínas e

---

<sup>ii</sup> A noção de massa atômica foi referida pela primeira vez por Dalton, em sequência da teoria atômica que desenvolveu. Dalton não tinha forma de calcular a massa de um átomo, mas precisava relacionar a massa de átomos de diferentes elementos entre si. Para tal foi necessário escolher um elemento específico como referência. Sendo assim, Dalton escolheu como elemento de referência o hidrogênio. Portanto, 1 Dalton (Da) é a unidade de massa equivalente a massa de um átomo de hidrogênio.

seus complexos em solução. Porém, um dos mais importantes aspectos da técnica de SAXS é que ela permite o estudo de proteínas em soluções muito próximas das condições fisiológicas, além de permitir a análise de mudanças estruturais em resposta a variações das condições externas.

Apesar de dados de espalhamento a baixos ângulos só conterem informações sobre o envelope molecular da molécula em estudo, muitas vezes esse envelope molecular pode auxiliar na solução dos problemas de ambigüidades na determinação de estruturas por outras técnicas de alta resolução, como por exemplo, o problema de sincronização de fases de reflexões de baixa resolução na cristalografia de proteínas [Hao *et al.*, 1999; Hao, 2001; Aparício *et al.*, 2002; Liu *et al.*, 2003]. Outras vezes, esse modelo de baixa resolução é tudo que se pode obter sobre a estrutura de determinadas proteínas, como acontece com as proteínas com regiões flexíveis ou desordenadas, e até mesmo para proteínas que não cristalizam ou são grandes demais (> 50kDa) para serem estudadas pela técnica de ressonância magnética nuclear (RMN). Esses modelos fornecem uma boa idéia do espaço conformacional adotado por uma proteína intrinsecamente desestruturada, assim como, as suas possíveis conformações adotadas em solução.

É importante ressaltar que a técnica de SAXS deve ser empregada em combinação com outros métodos bioquímicos e estruturais para uma completa caracterização estrutural de proteínas. Essa complementação é essencial para gerar modelos tridimensionais consistentes para as proteínas estudadas. A importância da técnica de SAXS no estudo de macromoléculas biológicas tem aumentado devido à acessibilidade aos instrumentos de pesquisa em instalações de larga escala, além dos avanços computacionais que têm permitido a criação de poderosos programas para análise de dados experimentais [Svergun & Koch, 2003]. Naturalmente, o requerimento básico para a uma correta caracterização estrutural de partículas é a correta realização de uma experiência de espalhamento, utilizando uma boa e adequada instrumentação (ver Capítulo 2). Monodispersidade da amostra, por exemplo, é um requerimento essencial para o estudo bem sucedido de proteínas em solução. Hoje em dia, a técnica de SAXS pode ser considerada como um método de precisão com um alto padrão experimental. Instrumentos bem desenvolvidos e disponíveis são confiáveis e podem, até certo ponto, ser automatizados, permitindo que a técnica de espalhamento a baixos ângulos seja desde uma ferramenta útil para trabalhos de rotina até uma técnica para caracterizações avançadas, como por exemplo, cinéticas de transformação da matéria e o estudo de grandes complexos macromoleculares (como são os ribossomos).

Para a demanda mais extrema, existem instalações experimentais com instrumentos poderosos e altamente especializados para experiências de espalhamento a baixos ângulos ao redor do mundo. Essas instalações são laboratórios que contêm anéis de armazenamento e fazem uso da radiação síncrotron para as experiências de espalhamento, onde a intensidade da radiação excede

aquela produzida por geradores convencionais de raios-X por várias ordens de grandeza. Com essas ferramentas, é possível realizar estudos resolvidos no tempo (por exemplo, análises de estados transientes durante mudanças conformacionais de macromoléculas), e também estudos de moléculas com diferença de densidade eletrônica muito baixa, cujo espalhamento é muito fraco (por exemplo, proteínas que não podem ser produzidas em alta concentração devido à agregação ou degradação). Além disso, o feixe de raios-X produzido pela radiação síncrotron possui uma baixa emitância<sup>iii</sup> o que é importante para que possamos distinguir a radiação espalhada, que é fraca, daquela do feixe incidente.

Por fim, os detectores usados para coletar a radiação espalhada merecem uma atenção especial, pois eles são peças importantes para se obter dados com qualidade boa o suficiente para as análises. Além disso, seu tamanho físico, muitas vezes, limita o intervalo angular observado da curva de espalhamento, o que implica realizar as experiências em mais de uma distância amostra-detector para alcançar o intervalo angular necessário. Hoje em dia, os detectores sensíveis a posição disponíveis contem dispositivos de detecção através de fios (conhecidos como detectores a gás), placa de imagem, matriz CCD e com a nova tecnologia de pixels híbridos de silício. Cada um possui suas peculiaridades, sensibilidade e ruído. Seus detalhes serão discutidos no capítulo 2 dessa tese, quando discutiremos a instrumentação de SAXS que montamos no Laboratório Nacional de Luz Síncrotron no Brasil.

Nesta tese de doutorado, o objetivo é a caracterização de proteínas flexíveis e/ou com desordem estrutural utilizando a técnica de SAXS, mostrando o imenso leque de possibilidades oferecidas por essa técnica no estudo de macromoléculas com essas características. No capítulo 3, apresentaremos nossos estudos com proteínas intrinsecamente desestruturadas. Tivemos a possibilidade de caracterizar duas importantes proteínas dessa classe que são: a proteína FEZ1 (abreviação do nome em inglês *Fasciculation and Elongation Zeta-1 protein*) [Lanza *et al.*, 2009; Assmann *et al.*, 2006], que está envolvida no crescimento e desenvolvimento de axônios; e a proteína regulatória humana denominada Ki-1/57 (nome dado devido ao gene que codifica essa proteína) [Bressan *et al.*, 2008], cujos dados funcionais disponíveis sugerem seu complexo papel na regulação da expressão gênica. Tivemos a possibilidade, nesse caso, de caracterizar duas proteínas com múltiplos domínios, onde existem regiões desordenadas estão conectando esses domínios. Uma delas foi a chaperona molecular (Hsp40) Sis1, juntamente com alguns mutantes construídos com a deleção de alguns domínios. A segunda foi proteína ribonucléica hnRNP-Q, que está relacionada

---

<sup>iii</sup> Feixe de baixa emitância significa que o feixe possui uma pequena divergência e uma pequena seção transversal. Isso permite focar o feixe de raios-X de tal maneira que suas dimensões são pequenas o suficiente no foco, sem que haja uma grande perda de intensidade.

com um amplo leque de importantes funções do RNA. Os estudos realizados sobre as estruturas de proteínas com múltiplos domínios serão apresentados no capítulo 4.

# 1

## Breve descrição da teoria e métodos utilizados nesta tese

### 1.1. ESCOPO

Para a caracterização estrutural de macromoléculas, técnicas como a cristalografia de raios-X e SAXS utilizam radiação numa faixa de comprimentos de onda (entre 1 e 2Å; energias entre 8 e 12 keV) especialmente úteis para esse tipo de caracterização por terem energia suficientemente alta para penetrar a matéria e interagir com os elétrons dos átomos de acordo com o fenômeno de espalhamento e difração. Em vista disso, para descrever o conteúdo dos trabalhos desenvolvidos nesta tese de doutorado, é preciso apresentar alguns conceitos fundamentais da interação dessa radiação com a matéria. Em particular, neste capítulo será apresentada uma breve revisão das principais características da técnica de espalhamento de raios-X a baixos ângulos (SAXS) no estudo de proteínas em solução. Obviamente, além de conhecer a técnica para a caracterização estrutural de uma determinada amostra, também é necessário conhecer a amostra em estudo. Como estaremos

trabalhando com proteínas, logo após a revisão de SAXS, uma pequena e breve revisão sobre estruturas de proteínas será apresentada nesse capítulo.

## 1.2. PRINCÍPIOS DE ESPALHAMENTO DE RAIOS-X A BAIXOS ÂNGULOS

A análise estrutural da matéria utilizando técnicas de espalhamento a baixos ângulos pode ser feita utilizando tanto espalhamento de raios-X (SAXS<sup>iv</sup>) quanto espalhamento de nêutrons (SANS<sup>v</sup>). Embora aqui somente iremos nos preocupar com o espalhamento de raios-X, todos os resultados podem ser aplicados também ao espalhamento de nêutrons, bastando somente pequenas adaptações, já que os nêutrons possuem massa e interagem com a matéria de maneira diferente dos raios-X. Uma boa introdução sobre a técnica de SANS pode ser encontrada em [http://www.ncnr.nist.gov/staff/hammouda/the\\_SANS\\_toolbox.pdf](http://www.ncnr.nist.gov/staff/hammouda/the_SANS_toolbox.pdf). Já sobre a técnica de SAXS, há vários livros [Guinier & Fournet, 1955; Glatter & Kratky, 1982; Feigin & Svergun, 1987] e revisões [Putman *et al.*, 2007; Lipfert & Doniach, 2007; Svergun & Koch, 2003; Koch *et al.*, 2003] explicando os detalhes da técnica.

Espalhamento de luz pode ser estudado como um problema de eletrodinâmica clássica ou como um fenômeno quântico. Nessa seção, vamos nos ater ao tratamento de eletrodinâmica clássica, mas uma prévia do tratamento quântico pode ser encontrada em algumas referências [Castro, 2004; Farina, 1973; Tang & Karplus, 1971; Wu & Ohmura, 1962]. No caso de raios-X incidindo sobre os átomos de um objeto, todos os elétrons atingidos pelas ondas tornam-se fonte de ondas espalhadas. Quando a frequência (e conseqüentemente a energia) de um fóton de raios-X é muito maior que a frequência de ressonância dos elétrons, todos os elétrons irão se comportar como se eles estivessem livres. Conseqüentemente, todas as ondas espalhadas têm a mesma intensidade, dada pela bem conhecida formula de Thomson [Cowley, 1995; Glatter & Kratky, 1982; Cullity, 1978]:

$$I_e(\theta) = I_p \left( \frac{e^2}{mc^2} \right)^2 \frac{1}{a^2} \frac{1 + \cos^2(2\theta)}{2} \quad (1.1)$$

onde  $I_p$  é a intensidade primária,  $a$  é a distância entre o objeto e o ponto de detecção,  $e$  é carga elementar do elétron,  $m$  é a massa do elétron,  $c$  é a velocidade da luz e  $2\theta$  é o ângulo de espalhamento. O fator numérico  $(e^2/mc^2)$  é conhecido como raio clássico do elétron. A

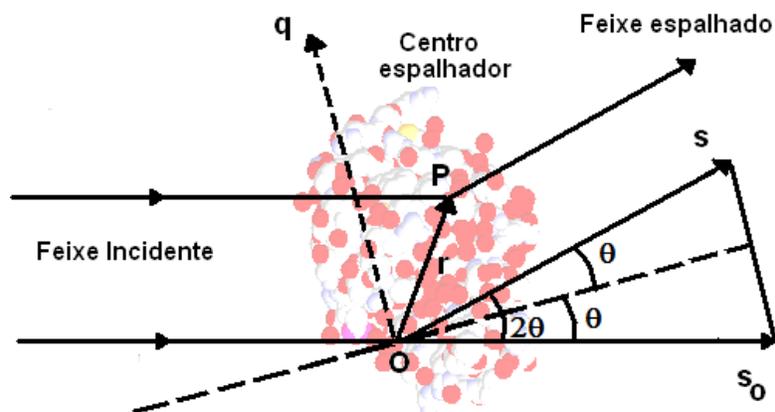
---

<sup>iv</sup> SAXS: abreviação do termo em inglês *Small Angle X-ray Scattering*

<sup>v</sup> SANS: abreviação do termo em inglês *Small Angle Neutron Scattering*

intensidade de espalhamento de um elétron depende somente do ângulo de espalhamento  $2\theta$  através do fator de polarização  $([1 + \cos^2(2\theta)]/2)$ , que é praticamente igual a  $I(u)$  para ângulos pequenos, que é o caso em todos os problemas de interesse aqui. Como a intensidade de espalhamento  $I_e(\theta)$  se aplica a todas as fórmulas de intensidade de espalhamento que se seguem, ela será omitida por brevidade.

O processo de espalhamento é caracterizado por uma lei de reciprocidade, que fornece uma relação inversa entre o tamanho da partícula e o ângulo de espalhamento. Dimensões coloidais (entre dezenas e vários milhares de angstroms) são enormes comparadas ao comprimento de onda dos raios-X (da ordem de  $1\text{\AA}$ ), o que faz o intervalo angular de espalhamento observável ser correspondentemente pequeno. Como os raios-X são espalhados pelos elétrons, o espalhamento de raios-X a baixos ângulos é sempre e somente observado quando as inhomogeneidades de densidade eletrônica de tamanhos coloidais existem na amostra [Glatter & Kratky, 1982]. Essas considerações se aplicam ao espalhamento coerente, pois as ondas espalhadas são coerentes. Mesmo que também ocorra espalhamento incoerente, esse pode ser desprezado quando somente ângulos pequenos estão envolvidos na experiência. Coerência aqui significa que as amplitudes são adicionadas, e a intensidade é então dada pelo quadrado do valor absoluto da amplitude de espalhamento resultante [Cowley, 1995; Glatter & Kratky, 1982]. As amplitudes são de igual magnitude, e diferem somente por suas fases  $\varphi$ , que por sua vez dependem das posições dos elétrons no espaço.



**Figura 1.1** – Representação esquemática do espalhamento por uma partícula fixa no espaço. A direção do feixe incidente está indicada pelo vetor unitário  $\hat{s}_0$ , e do feixe espalhado por  $\hat{s}$ . O ângulo de espalhamento é  $2\theta$  e o vetor de espalhamento  $\vec{q}$  tem a mesma direção de  $(\vec{s} - \vec{s}_0)$ .

É conveniente representar cada onda espalhada pela sua forma complexa:  $e^{i\varphi}$ . A fase  $\varphi$  é dada por  $(2\pi/\lambda)$  multiplicada pela diferença de caminho óptico em alguma referência arbitrária. O cálculo de  $\varphi$  é ilustrado na figura 1.1. Vamos denominar a direção do feixe incidente pelo vetor unitário  $\hat{s}_0$ , e do feixe espalhado por  $\hat{s}$ . A diferença de caminho entre um ponto  $P$ , especificado pelo vetor  $\vec{r}$ , e a origem  $\theta$  é então dado por  $-\vec{r} \cdot (\hat{s} - \hat{s}_0)$ . Portanto, a fase é dada por  $\varphi = -(2\pi/\lambda) \vec{r} \cdot (\hat{s} - \hat{s}_0)$ . Isso pode ser também escrito na forma  $\varphi = -\vec{q} \cdot \vec{r}$ , onde  $\vec{q} = (2\pi/\lambda)(\hat{s} - \hat{s}_0)$  é o chamado vetor de espalhamento ou vetor de transferência de momento. Da figura 1.1, podemos observar que o vetor  $(\hat{s} - \hat{s}_0)$  tem magnitude  $(2\text{sen}\theta)$ . Aqui,  $\theta$  é um ângulo cujo valor é metade do ângulo de espalhamento  $2\theta$ , conforme indicado na figura 1.1. Conseqüentemente, o vetor  $\vec{q}$  tem a mesma direção de  $(\hat{s} - \hat{s}_0)$  e módulo  $q = (4\pi/\lambda)\text{sen}\theta$ .

É possível obter a amplitude de espalhamento resultante pela soma das ondas espalhadas, sendo cada onda representada pelo termo complexo  $e^{-i\vec{q} \cdot \vec{r}}$ , onde já reescrevemos a fase como  $\varphi = -\vec{q} \cdot \vec{r}$ . Entretanto, considerando o enorme número de elétrons na partícula espalhadora e o fato de que um único elétron não pode ser exatamente localizado, será conveniente introduzir o conceito de densidade eletrônica. Essa densidade, definida como o número de elétrons por unidade de volume, é denotada aqui por  $\rho(\vec{r})$ . Um elemento de volume  $dV$  na posição  $\vec{r}$  irá conter  $\rho(\vec{r})dV$  elétrons. Assim a soma das ondas espalhadas pode ser repassada pela integral sobre o volume inteiro  $V$  irradiado pelo feixe incidente da seguinte forma:

$$F(\vec{q}) = \iiint \rho(\vec{r}) e^{-i\vec{q} \cdot \vec{r}} dV \quad (1.2)$$

Seguindo a definição feita na equação (1.2), do ponto de vista matemático, a amplitude de espalhamento  $F(\vec{q})$  em uma determinada direção especificada por  $\vec{q}$  é a transformada de Fourier da distribuição de densidade eletrônica  $\rho(\vec{r})$ . A intensidade de espalhamento é então dada por  $I(\vec{q}) = F(\vec{q})F^*(\vec{q})$ , onde  $F^*(\vec{q})$  é o complexo conjugado de  $F(\vec{q})$ :

$$I(\vec{q}) = F(\vec{q})F^*(\vec{q}) = \iiint \iiint \rho(r_1)\rho(r_2)e^{-iq \cdot (\vec{r}_1 - \vec{r}_2)} dV_1 dV_2 \quad (1.3)$$

Essa expressão é novamente uma integral de Fourier, envolvendo somente a distância relativa  $|\vec{r}_1 - \vec{r}_2|$  para todos os pares de pontos dentro da partícula. Dessa forma, é conveniente calcular essa integral em dois passos: primeiramente integrando sobre os pares com a mesma distância relativa e depois integrando sobre todas as distâncias relativas (incluindo o fator de fase).

O primeiro passo pode ser realizado definindo a função de autocorrelação de densidade eletrônica dada por [Glatter & Kratky, 1982]:

$$\gamma(\vec{r}) = \iiint \rho(\vec{r}_1) \rho(\vec{r}_2) dV_1 \quad (1.4)$$

onde  $\vec{r} = \vec{r}_1 - \vec{r}_2$  é mantido constante, logo, os vetores  $\vec{r}$ ,  $\vec{r}_1$  e  $\vec{r}_2$  estão correlacionados. Sendo assim, o segundo passo resulta na seguinte expressão:

$$I(\vec{q}) = \iiint \gamma(\vec{r}) e^{-i\vec{q}\cdot\vec{r}} dV \quad (1.5)$$

Esse resultado nos fornece uma equação para a intensidade de espalhamento de uma partícula fixa no espaço. Entretanto, no estudo de partículas em solução, elas estão aleatoriamente orientadas. Dessa forma, podemos reescrever a intensidade de espalhamento de uma partícula aleatoriamente orientada no espaço calculando a média da intensidade na equação (1.5) em um ângulo sólido  $d\Omega = \text{sen}\theta' d\theta' d\varphi'$  no espaço recíproco [Feigin & Svergun, 1987]:

$$I(q) = \langle I(\vec{q}) \rangle_{\Omega} = 4\pi \int_0^{\infty} r^2 \gamma(r) \frac{\text{sen}(qr)}{qr} dr \quad (1.6)$$

onde  $\gamma(r)$  é a média espacial da função de correlação (definida na equação 1.4) calculada em um ângulo sólido  $d\varpi = \text{sen}\theta d\theta d\varphi$  no espaço real:

$$\gamma(r) = \langle \gamma(\vec{r}) \rangle_{\varpi} = \frac{1}{4\pi} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} \gamma(\vec{r}) \text{sen}\theta d\theta d\varphi \quad (1.7)$$

E que pode ser encontrada também pela transformada de Fourier inversa da intensidade:

$$\gamma(r) = \frac{1}{2\pi^2} \int_0^{\infty} q^2 I(q) \frac{\text{sen}(qr)}{qr} dq \quad (1.7a)$$

A função  $\gamma(r)$  é comumente chamada de função característica da partícula e representa a probabilidade de encontrar um dado ponto dentro da partícula em uma distância  $r$  de um outro dado ponto. Dessa forma, multiplicando a função  $\gamma(r)$  da equação (1.7a) por  $r^2$  podemos definir uma nova função chamada função distribuição de pares de distâncias,  $p(r)$ , dentro da partícula [Glatter & Kratky, 1982; Feigin & Svergun, 1987]:

$$p(r) = r^2 \gamma(r) = \frac{1}{2\pi^2} \int_0^{\infty} r^2 q^2 I(q) \frac{\text{sen}(qr)}{qr} dq. \quad (1.8)$$

Essa função  $p(r)$  representa um histograma de pares de distâncias dentro de uma partícula. Se dividirmos a partícula em um número muito grande de elementos infinitesimais de volume idênticos, a função  $p(r)$  será proporcional ao número de linhas com comprimento entre  $r$  e  $r+dr$  que

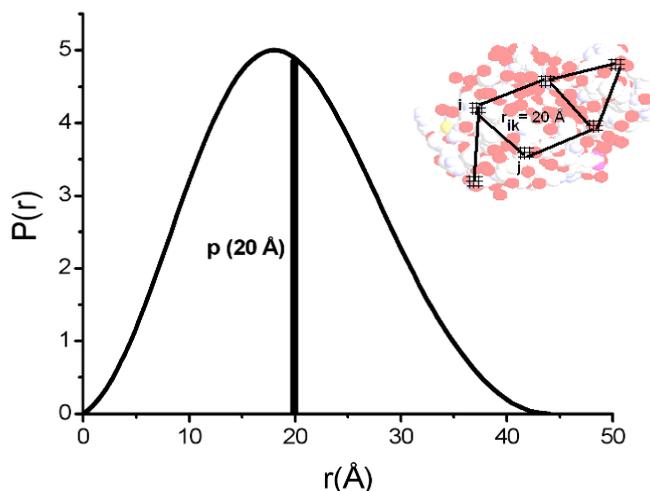
serão encontradas na combinação de qualquer elemento de volume arbitrário  $i$  e outro elemento de volume  $j$ , também arbitrário, da partícula, como está exemplificada na figura 1.2 [Glatter & Kratky, 1982].

A importância da função  $p(r)$  é que ela está no espaço real, o que nos fornece uma noção mais intuitiva sobre a forma da partícula do que a função  $I(q)$ , que por sua vez está no espaço recíproco. Entretanto, o que obtemos experimentalmente é a função  $I(q)$ . Sendo assim, gostaríamos de obter a função  $p(r)$  a partir dos dados de  $I(q)$ . Isso é possível reescrevendo as equações (1.6) e (1.8) para  $I(q)$  e  $p(r)$ , respectivamente, da seguinte forma:

$$I(q) = 4\pi \int_0^{\infty} p(r) \frac{\text{sen}(qr)}{qr} dr \quad (1.9)$$

$$p(r) = \frac{1}{2\pi^2} \int_0^{\infty} r q I(q) \text{sen}(qr) dq \quad (1.10)$$

Postas dessa forma, essas equações nos revelam uma importante relação matemática entre essas duas funções. Notamos que a função  $p(r)$  é a transformada de Fourier inversa da função intensidade  $I(q)$  e, obviamente,  $I(q)$  é a transformada de Fourier direta de  $p(r)$ . A importância dessa relação está no fato de que uma pode ser obtida a partir da outra através de uma transformada de Fourier (direta ou inversa, conforme for o caso).



**Figura 1.2** – Representação da função de distribuição de pares de distância,  $p(r)$ , de uma partícula, mostrando a probabilidade de encontrar pares de distâncias com  $r = 20\text{Å}$  entre elementos de volume arbitrários  $i$  e  $j$  dentro da molécula. No destaque está uma representação esquemática bidimensional de pares distâncias de  $20\text{Å}$  de comprimento dentro da partícula.

Seguindo nessa linha de pensamento, é possível calcular a curva de espalhamento esperada para uma partícula de qualquer forma geométrica. Obviamente, algumas poderão ser calculadas analiticamente e outras só numericamente. Como o espalhamento de raios-X a baixos ângulos permanecerá essencialmente inalterado se o deslocamento dos elétrons dentro da partícula ocorrer por distâncias menores que a dimensão global dessa partícula, então partículas com inhomogeneidades pequenas podem ser tratadas como se tivessem uma distribuição de densidade eletrônica uniforme.

### 1.3. ESPALHAMENTO DE RAIOS-X POR PARTÍCULAS EM SOLUÇÃO

Até agora foi assumido implicitamente que as partículas estivessem no vácuo, tal que a amplitude de espalhamento era proporcional ao número de moles de elétrons por unidade de volume, isto é, a densidade eletrônica  $\rho$ . Para partículas dissolvidas em um solvente (proteína em um tampão, por exemplo), estamos tratando de um problema de duas fases. Nesse caso, somente a diferença de densidade eletrônica  $\Delta\rho = \rho_2 - \rho_1$  ( $\rho_2$  = densidade eletrônica do soluto e  $\rho_1$  = densidade eletrônica do solvente) é efetiva. Portanto, a amplitude de espalhamento é proporcional a  $\rho_2 - \rho_1$  e, conseqüentemente, a intensidade de espalhamento será proporcional a  $(\Delta\rho)^2 = (\rho_2 - \rho_1)^2$ . Se essa diferença é nula, ou seja,  $\rho_2 = \rho_1$ , então o feixe de raios-X não “enxergará” as partículas, desde que a solução então representará um espaço contínuo de densidade eletrônica homogênea, e as ondas espalhadas em qualquer direção serão conseqüentemente extintas. O termo “densidade eletrônica” é freqüentemente usado para a diferença de densidade eletrônica, que por sua vez é também conhecida como “contraste”.

Além disso, já sabemos como a técnica funciona para uma partícula isolada. Sendo assim, agora é importante entender o espalhamento de um grupo disperso de partículas, isto é, uma solução. Em solução, cada partícula possui orientação aleatória e vai espalhar os raios-X de tal maneira que a intensidade de espalhamento seja dada pela equação (1.9). Para soluções suficientemente diluídas (não interagentes) e homogeneamente dispersas (monodispersas), as intensidades de espalhamento das partículas individuais simplesmente se adicionam. Por monodispersas se entende que haja somente um tipo de partícula em solução, cada uma com uma distribuição de densidade eletrônica  $\rho(\vec{r})$  e uma intensidade de espalhamento  $I_1(q)$ . A intensidade de espalhamento de uma solução de  $N$  partículas irradiadas pelo feixe de raios-X será então [Vachette & Svergun, 2000]:

$$I(q) = N I_1(q) \quad (1.11)$$

Dessa forma, o problema das análises de espalhamento a baixos ângulos consiste em deduzir tamanho, forma, massa e até mesmo a distribuição de densidade eletrônica de uma partícula a partir da curva de espalhamento das partículas em solução. Isso é possível devido ao fato que a intensidade de espalhamento conjunto de partículas será proporcional a intensidade de espalhamento de uma só partícula (equação (1.11)), desde que a solução seja monodispersa.

## 1.4. PARÂMETROS ESTRUTURAIS GLOBAIS

A análise das curvas de espalhamento a baixos ângulos permite determinar parâmetros estruturais da partícula espalhadora. Tradicionalmente, os primeiros parâmetros que podem ser diretamente determinados da curva de espalhamento são o raio de giro ( $R_g$ ), volume excluído da partícula e o valor da intensidade na origem ( $I(0)$ ), que por sua vez pode ser usada para estimar a massa molecular da partícula espalhadora. Esses parâmetros são obtidos através de aproximações assintóticas da curva de intensidade para pequenos ou grandes valores de  $q$ . Da função  $p(r)$  também é possível obter o valor de  $R_g$  e de  $I(0)$ . Mais ainda, ela fornece o valor da máxima dimensão da partícula ( $D_{máx}$ ) e dá uma idéia intuitiva da forma da molécula. Obviamente, ao trabalhar com proteínas, isso tudo só é possível quando a solução estudada é suficientemente diluída e monodispersa. A seguir vamos discutir como todos esses parâmetros podem ser obtidos.

### 1.4.1. INTENSIDADE NA ORIGEM

A intensidade de espalhamento em  $q=0$ , que vamos chamar de  $I(0)$ , é muito útil nas análises de SAXS para estimar a massa molecular de uma proteína em solução. Para isso, precisamos entender como  $I(0)$  se relaciona com a massa molecular de uma partícula espalhadora. Primeiramente, da equação (1.11),  $I(0)=N I_1(0)$ , onde  $I_1(0)$  é dada por [Feigin & Svergun, 1987; Glatter & Kratky, 1982]:

$$I_1(0) = \iiint_{V_1} \iiint_{V_2} \Delta\rho(\vec{r}_1)\Delta\rho(\vec{r}_2)dV_1dV_2 = \Delta n_e^2 = n_e^2 - n_0^2 \quad (1.12)$$

onde  $n_e$  é o número de elétrons da partícula e  $n_0$  é o número de elétrons do volume excluído do solvente pela partícula. Então, da equação (1.11), para  $N$  partículas irradiadas temos:

$$I(0) = N I_1(0) = N n_e^2 (1 - \rho_0 v_p)^2 \quad (1.15)$$

onde  $v_p$  é o volume específico parcial da partícula e  $\rho_0$  é a densidade eletrônica do solvente. Geralmente, para indicar as concentrações de soluções de proteínas, costuma-se usar a unidade de

g/L (gramas por litro) ou mg/mL (miligramas por mililitro). Sendo assim, podemos escrever  $c = N\mu n_e / N_A$ , onde  $c$  é a concentração das partículas (massa/volume) em solução,  $N_A$  é o número de Avogrado, e  $\mu$  é a razão  $M/n_e$  da massa molecular ( $M$ ) pelo número de elétrons ( $n_e$ ), razão essa que depende da composição química da partícula (para proteínas, uma boa aproximação é  $M/n_e=1,87$ ). Daí, segue que [Vachette & Svergun, 2000]:

$$\frac{I(0)}{c} = \frac{N_A M}{\mu^2} (1 - \rho_0 v_p)^2 \quad (1.16)$$

Da equação (1.16) vemos que a intensidade em  $q=0$  normalizada pela concentração,  $I(0)/c$ , é proporcional à massa molecular  $M$  da partícula espalhadora. Dessa forma, se  $v_p$ ,  $\rho_0$  e  $c$  são conhecidos, e se a intensidade do feixe incidente é também conhecida em escala absoluta (número real de fótons incidentes durante a medida), então o valor de  $I(0)$  pode ser usado para estimar o valor da massa molecular de uma partícula. Posteriormente, estaremos discutindo como obter o valor da massa molecular usando um padrão secundário (proteína-padrão ou água), para obter o valor absoluto da intensidade incidente, o que não é trivial, ou eliminar a necessidade de obter esse valor.

#### 1.4.2. RAIOS DE GIRO E $I(0)$ : APROXIMAÇÃO DE GUINIER

A primeira análise geralmente feita após coletar e tratar a curva experimental de espalhamento é a análise de Guinier, utilizada para obter o valor de raio de giro da partícula ( $R_g$ ) e intensidade na origem ( $I(0)$ ). Isso é feito através de uma análise assintótica para pequenos valores de  $q$ , expandindo a intensidade da equação (1.9) em uma série de potências e mantendo até termos de potência  $q^2$ . Para pequenos valores de  $q$  (próximo a origem), essa expansão da equação (1.9) fica:

$$I(q) = I(0) \left[ 1 - \frac{1}{3} R_g^2 q^2 + O(q^4) \right] \quad (1.17)$$

Rearranjando os termos e restringindo o intervalo de valores de  $q$  tal que  $qR_g < 1$ , isso dá origem à conhecida aproximação de Guinier [Guinier & Fournet, 1955]:

$$I(q) = I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right) \quad (1.18)$$

onde o raio de giro  $R_g$  é expresso por (ou, analogamente, como mostraremos a seguir na equação (1.20)):

$$R_g^2 = \frac{\int_V r^2 \Delta\rho(\vec{r}) d^3r}{\int_V \Delta\rho(\vec{r}) d^3r} \quad (1.19)$$

O Raio de Giro ( $R_g$ ) é a distância quadrática média de uma dada extremidade da partícula em relação ao seu centro de gravidade (em um sistema onde as forças gravitacionais agindo sobre o corpo são uniformes, o centro de gravidade é igual ao centro de massa). Experimentalmente, o  $R_g$  é o parâmetro que fornece uma medida média do tamanho de uma macromolécula independente da forma geométrica. Matematicamente, seu valor quadrático,  $R_g^2$ , é o segundo momento da distribuição de distâncias em torno do centro de massa de uma partícula. Por analogia com a mecânica clássica, o  $R_g$  pode ser considerado como o raio de giro eletrônico da partícula em torno de seu centro de massa (Teraoka, 2002; Guinier & Fournet, 1955).

A aproximação de Guinier na equação (1.18) é válida somente quando a contribuição do próximo termo (termo de ordem  $q^4$ ) na expansão de potências de  $I(q)$  é desprezível. Na prática, podemos linearizar a equação (1.18) aplicando logaritmo natural dos dois lados da equação da seguinte forma:

$$\ln I(q) = \ln I(0) - q^2 \left( \frac{R_g^2}{3} \right) \quad (1.18a)$$

Sendo assim, a curva  $\ln(I(q))$  vs.  $q^2$  pode ser ajustada por uma reta através de uma regressão linear dentro da região de validade  $qR_g < 1$ , rendendo os valores de  $I(0)$  e  $R_g$  de maneira direta utilizando a equação 1.18a. A linearidade da parte inicial das curvas experimentais de espalhamento (sem desvios dessa linearidade) é um indicativo de monodispersidade das partículas espalhadoras. Logo, essa análise é aplicada a todas as curvas de espalhamento para verificar a monodispersidade da solução de partículas que está sendo estudada.

### 1.4.3. RAIOS DE GIRO, MÁXIMA DIMENSÃO E $I(0)$ : FUNÇÃO $P(r)$

Parâmetros como o raio de giro ( $R_g$ ), máxima dimensão da partícula ( $D_{\max}$ ) e  $I(0)$  também podem ser extraídos da função  $p(r)$  [Guinier & Fournet, 1955; Glatter & Kratky, 1982; Feigin & Svergun, 1987]. Como o quadrado do raio de giro de uma partícula é metade da média quadrática dos pares de distância dentro da partícula, cuja distribuição é dada pela função  $p(r)$ , podemos obter o valor de  $R_g$  através da seguinte relação:

$$R_g^2 = \frac{\int_0^{D_{m\acute{a}x}} r^2 p(r) dr}{2 \int_0^{D_{m\acute{a}x}} p(r) dr} \quad (1.20)$$

onde  $D_{m\acute{a}x}$  é a máxima dimensão da partícula e que pode ser obtida para o valor de  $r$  tal que a função  $p(r)$  vá a zero, já que  $p(r \geq D_{m\acute{a}x}) = 0$ .

E para obter o valor de  $I(0)$ , substituímos o  $q$  por  $q=0$  na equação (1.9) e obtemos a seguinte expressão:

$$I(0) = 4\pi \int_0^{D_{m\acute{a}x}} p(r) dr \quad (1.21)$$

Esses cálculos de  $R_g$  e  $I(0)$  são de interesse porque são independentes de qualquer modelo e levam em conta toda a curva de espalhamento experimental, não somente a região inicial, como na aproximação de Guinier. A comparação entre estas duas formas de cálculo fornece uma validação dos resultados.

#### 1.4.4. PARTÍCULAS FLEXÍVEIS: APROXIMAÇÃO DE DEBYE

Para cadeias flexíveis (desestruturadas), como ácidos nucléicos ou proteínas desenoveladas (desestruturadas), a aproximação de Guinier tem uma região bastante estreita de validade, sobrando poucos pontos para a regressão linear. Isso porque as partículas espalhadoras irão conter distâncias internas grandes, tal que os termos com potências maiores na expansão em série de potências de  $q$  da equação (1.17) se tornam significativos para valores pequenos de  $q$ , o que é suficiente para transformar a aproximação de Guinier numa aproximação sem uso prático. Foi mostrado que nesse caso, que a função de espalhamento pode ser aproximada no domínio  $qR_g < 1.4$  pela equação de Debye [Debye, 1947]:

$$\frac{I(q)}{I(0)} = \frac{2}{(q^2 R_g^2)^2} (q^2 R_g^2 - 1 + e^{-q^2 R_g^2}) \quad (1.22)$$

Essa relação é a expressão analítica correta para uma cadeia Gaussiana infinitamente fina, sem comprimento de persistência (isto é, sem rigidez devido às interações de curto alcance entre os monômeros). Entretanto, essa relação permanece válida em uma região restrita de valores de  $q$  no caso de uma cadeia de volume excluído ou com comprimento de persistência. Equações com adição desses últimos efeitos são apresentadas em diversas referências [Guinier & Fournet, 1955; Glatter & Kratky, 1982, Feigin & Svergun, 1987, Koch *et al.*, 2003]. O mais interessante da equação (1.22) é que ela depende de um único parâmetro somente: o raio de giro,  $R_g$ .

### 1.4.5. APROXIMAÇÃO DE POROD

Vamos agora considerar o comportamento de  $I(q)$  quando  $q$  tende a grandes valores. Integrando duas vezes a equação (1.9) por partes e lembrando que  $\gamma(r) = 0$  para  $r \geq D_{\text{máx}}$  (e conseqüentemente  $p(r) = r^2 \gamma(r) = 0$  para  $r \geq D_{\text{máx}}$ ), temos:

$$I(q) = -\frac{8\pi}{q^4} \gamma'(0) + \frac{O_1}{q^3} + \frac{O_2}{q^4} + O(q^{-5}) \quad (1.23)$$

onde  $\gamma'(0)$  é a primeira derivada de  $\gamma(r)$  em  $r = 0$  e  $O_1$  e  $O_2$  são termos trigonométricos oscilantes da forma  $\text{sen}(qD_{\text{máx}})$  [Svergun & Koch, 2003]. Todos os termos proporcionais a  $q^{-3}$  e  $q^{-4}$ , exceto o primeiro, são oscilatórios de tal forma que o termo principal responsável pelo decaimento de intensidade em altos valores de  $q$  é dado por:

$$I(q) \approx -\frac{8\pi}{q^4} \gamma'(0). \quad (1.24)$$

Além disso, para partículas homogêneas e compactas (rígidas),  $\gamma'(0)$  é igual a  $-(\Delta\rho)^2 \frac{S}{4}$  onde  $S$  é a área superficial da partícula [Feigin & Svergun, 1987]. Sendo assim, substituindo  $\gamma'(0)$  na equação (1.24), temos:

$$I(q) \approx -\frac{2\pi}{q^4} (\Delta\rho)^2 S. \quad (1.25)$$

Essa relação é conhecida como Lei de Porod e foi obtida independentemente por [Debye & Bueche, 1949; Porod, 1951]. Ela indica que, para partículas homogêneas e compactas (rígidas), o decaimento da curva de intensidade de espalhamento para valores grande de  $q$  segue uma lei de potência e seu comportamento vai com  $q^{-4}$ . Outra função importante também obtida por Porod é o invariante de Porod, definido como [Svergun & Koch, 2003]:

$$Q = \int_0^{\infty} q^2 I(q) dq = 2\pi^2 \int_V (\Delta\rho(\vec{r}))^2 d^3 r \quad (1.26)$$

onde as integrais do espaço real e recíproco são iguais devido ao Teorema de Parseval [Arfken & Weber, 2001]. Para partículas homogêneas,

$$Q = 2\pi^2 (\Delta\rho)^2 V \quad (1.27)$$

e levando em conta que também  $I(0) = (\Delta\rho)^2 V^2$ , o volume excluído da partícula (conhecido como Volume de Porod) é:

$$V = 2\pi^2 \frac{I(0)}{Q}. \quad (1.28)$$

Conseqüentemente, o comportamento assintótico da curva de intensidade normalizada também permite estimar a superfície específica da partícula como:

$$\frac{S}{V} = \left( -\frac{\pi}{Q} \right) \lim_{q \rightarrow \infty} (q^4 I(q)) \quad (1.29)$$

Note que, graças ao invariante de Porod, ambos os parâmetros podem ser obtidos dos dados em escala relativa. Porém, essa estratégia só funciona bem para partículas compactas, para os quais a integral da equação (1.26) pode ser aproximada com base no comportamento assintótico da intensidade para  $q \rightarrow 0$  e  $q \rightarrow \infty$ . Sendo assim, notamos que a teoria apresentada nessa seção está baseada na hipótese de decaimento da intensidade com  $q^{-4}$ .

#### 1.4.6. PARTÍCULAS FLEXÍVEIS: APROXIMAÇÃO ASSINTÓTICA PARA GRANDES VALORES DE $Q$

Nós já vimos a aproximação assintótica para grandes valores de  $q$  da curva de intensidade de espalhamento para partículas compactas: a aproximação de Porod (equação (1.25)). Porém, para partículas não-compactas e flexíveis, a intensidade de espalhamento tem o seguinte limite para grandes valores de  $q$  [Glatter & Kratky, 1982]:

$$\lim_{q \rightarrow \infty} I(q) = \frac{2}{q^2 R_g^2} \left[ 1 - \frac{1}{q^2 R_g^2} \right] \quad (1.30)$$

Dessa forma, para grandes valores de  $q$ ,  $I(q)$  decai com  $q^{-2}$  para partículas flexíveis e desestruturadas em vez do decaimento com  $q^{-4}$  da Lei de Porod para partículas globulares e compactas. Essa diferença no comportamento assintótico fornece uma maneira prática de monitorar o grau de compacidade de uma partícula (que pode ser correlacionado ao grau de estrutura ou grau de envelhecimento de uma proteína, por exemplo). Isso pode ser convenientemente visto usando a representação de Kratky que discutiremos na seção seguinte.

#### 1.4.7. ESTIMATIVA DE MASSA MOLECULAR DA PARTÍCULA

Outro importante parâmetro para o estudo de partículas em solução que pode ser obtido dos dados de SAXS é a massa molecular ( $M$ ) da partícula. Ela é obtida utilizando o valor da intensidade de espalhamento para  $q = 0$ . Entretanto, o valor de  $I(0)$  não pode ser experimentalmente medido de

uma maneira direta, tal que uma extrapolação dos dados experimentais para o valor nulo de  $q$  deve ser feita. Isso pode ser feito utilizando a aproximação de Guinier (equação 1.18) ou a função  $p(r)$  (equação 1.21).

Como foi mencionado anteriormente, a medida da intensidade de espalhamento em escala absoluta é a principal tarefa para a determinação do valor da massa molecular [Orthaber *et al.*, 2000]. Todavia, a medida da intensidade absoluta, que é a razão da intensidade de espalhamento da molécula pela intensidade espalhada por um elétron (equação (1.1)), não é tarefa fácil. Felizmente, existem dois métodos diferentes para obter medida em escala absoluta. O primeiro deles é um método mecânico direto, que utiliza um mecanismo de atenuação do feixe primário por discos rotatórios [Kratky & Wawra, 1963] ou uma fenda móvel [Stabinger & Kratky, 1978]. Entretanto, essa atenuação mecânica não é muito útil quando a intensidade é registrada com detectores sensíveis à posição [Russell, 1983].

A segunda maneira é o método indireto, usando padrões secundários calibrados como Lupolen, carvão vítreo, água e outros. A água é especialmente útil como padrão de calibração porque seu espalhamento é independente do ângulo de espalhamento. Ainda podemos utilizar uma proteína estável e já bem caracterizada, cuja massa molecular seja conhecida, como padrão secundário para a determinação direta do valor de  $M$ , sem antes precisar calibrar a intensidade para uma escala absoluta. Detalhes serão discutidos nas próximas seções.

#### 1.4.7.1. UTILIZANDO ÁGUA COMO PADRÃO PARA DETERMINAÇÃO DA MASSA MOLECULAR

A vantagem da água como padrão é que seu espalhamento, além de ser independente do ângulo, depende somente da propriedade física de compressibilidade isotérmica. Isso é muito útil na calibração da escala de intensidade. Entretanto, o espalhamento da água é muito fraco e, é necessário levar em conta a sua absorção e o espalhamento do branco da câmara (espalhamento da câmara vazia) em que a água está [Orthaber *et al.*, 2000] durante a medida experimental.

A dependência da seção de choque da água  $d\Sigma/d\Omega$ , que é proporcional a intensidade, com sua compressibilidade isotérmica  $\beta_T$  aparece para valores maiores de ângulos de espalhamento como:

$$\frac{d\Sigma}{d\Omega} = \rho^2 k_B T \beta_T \quad (1.30)$$

onde  $\rho$  é a densidade eletrônica da água,  $T$  é temperatura e  $k_B$  ( $1,38 \times 10^{-23}$  J/K) é a constante de Boltzmann. A intensidade de espalhamento constante da água é  $1,632 \times 10^{-12}$   $cm^{-1}$  em 20°C (293 K),

com  $\beta_T = 4,58 \times 10^{-10} \text{ Pa}^{-1}$ . É importante notar que a compressibilidade da água é uma função da temperatura  $T$  [Bosio *et al.*, 1989; Fritz *et al.*, 2000]. Tabelas com os valores para a compressibilidade da água e  $[d\Sigma/d\Omega]_{\text{água}}$  podem ser encontradas em <http://physchem.kfunigraz.ac.at/sm/>. Com esses valores é possível obter o espalhamento da amostra em relação ao espalhamento da água e transformar a escala dos dados de espalhamento da amostra para a escala absoluta. Sendo assim, para obter  $d\Sigma/d\Omega$  da amostra em escala absoluta fazemos [Orthaber *et al.*, 2000]:

$$\left[ \frac{d\Sigma}{d\Omega} \right]_{\text{amostra}} = \frac{I(0)_{\text{amostra}}}{I(0)_{\text{água}}} \left[ \frac{d\Sigma}{d\Omega} \right]_{\text{água}} . \quad (1.31)$$

Tendo estimado o valor da intensidade em escala absoluta e determinado  $I(0)$  nessa escala pela equação (1.31), é possível estimar a massa molecular da amostra usando a seguinte fórmula [Orthaber *et al.*, 2000]:

$$M = \frac{N_A I(0)/c}{\Delta\rho_M^2} \quad (1.32)$$

onde  $I(0)/c$  é a intensidade absoluta em  $q=0$  normalizada pela concentração da amostra,  $\Delta\rho_M = [\rho_{M,prot} - (\rho_{solv} \bar{v})]r_0$  é o contraste por unidade de massa,  $N_A = 6,023 \times 10^{23} \text{ mol}^{-1}$  é o Número de Avogadro,  $\rho_{M,prot} = 3,22 \times 10^{23} \text{ e/g}$  é o número de elétrons por massa da proteína seca,  $\rho_{solv} = 3,34 \times 10^{23} \text{ e/cm}^3$  é o número de elétrons por volume de um solvente aquoso,  $\bar{v}$  é o volume específico parcial da proteína, e  $r_0 = 2,8179 \times 10^{-13} \text{ cm}$  é o comprimento de espalhamento de um elétron.

#### 1.4.7.2. UTILIZANDO UMA PROTEÍNA SECUNDÁRIA COMO PADRÃO

Outra maneira de obter a massa molecular ( $M$ ) de uma partícula através dos dados de espalhamento a baixo ângulo é utilizando os dados de espalhamento de uma proteína conhecida obtidos nas mesmas condições experimentais ou em condições bem próximas. Isso é muito usado para estimar o peso molecular de proteínas, pois a água espalha muito pouco e uma curva de espalhamento da água leva mais tempo para ser obtida, ao passo que a curva de espalhamento de uma proteína pode ser obtida mais rapidamente. Essa proteína-padrão precisa ser estável o suficiente para não alterar suas propriedades estruturais durante a experiência e estar bem caracterizada, inclusive com o peso molecular determinado. Duas proteínas muito utilizadas são a lisozima e a albumina, cujas estruturas já estão bem caracterizadas.

Dessa forma, vemos na equação (1.16) que  $M$  está diretamente relacionado com a intensidade de espalhamento para  $q = 0$ , escrita como  $I(0)$ , sendo que  $I(0) \propto cM$ , onde  $c$  é a concentração (g/L) das partículas em solução e  $M$  é a massa molecular de cada partícula. Sendo assim, para estimar a massa molecular, a amostra de proteína-padrão é usada para a calibração dos dados de espalhamento da molécula em estudo. Obviamente, as concentrações tanto da solução da proteína estudada quanto da solução da proteína padrão precisam ser determinadas com a máxima precisão possível utilizando outros métodos (técnicas espectroscópicas, por exemplo, usando a lei de Beer [Layne, 1957; Stoscheck, 1990]). Fazendo isso podemos relacionar os valores de  $I(0)$ ,  $c$  e  $M$  das duas proteínas da seguinte forma [Mylonas & Svergun, 2007]:

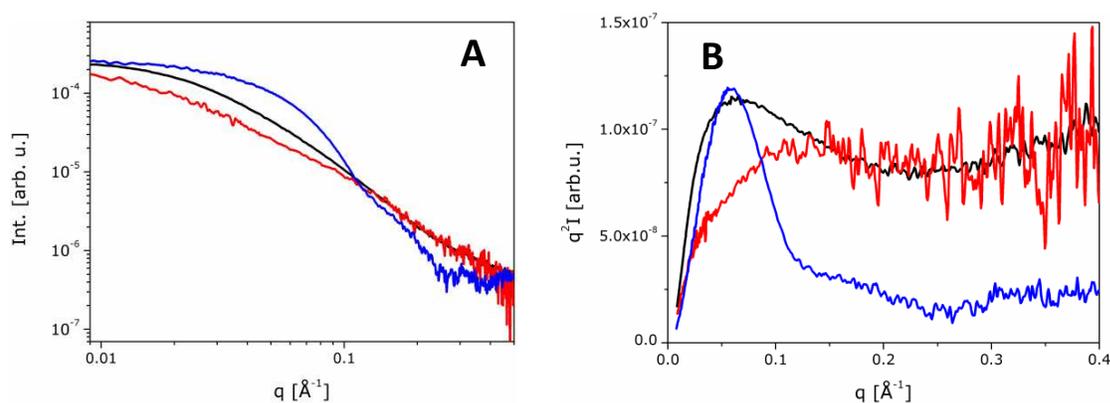
$$\frac{(I(0)/c)_{proteína}}{(I(0)/c)_{padrão}} = \frac{M_{proteína}}{M_{padrão}} \quad (1.33)$$

onde  $(I(0)/c)_{proteína}$  e  $(I(0)/c)_{padrão}$  são as intensidades de espalhamento em  $q=0$ , normalizadas pela concentração, da proteína e do padrão, respectivamente.

## 1.5. REPRESENTAÇÃO DE KRATKY DA INTENSIDADE: $q^2I(q)$ vs. $q$

Uma análise útil e importante da compacidade das proteínas pode ser feita a partir dos dados de SAXS através de uma representação especial da curva de intensidade de espalhamento:  $q^2I(q)$  vs.  $q$ . Essa representação é conhecida como gráfico de Kratky ou representação de Kratky (conhecido em inglês como *Kratky Plot*). Essa representação vem da teoria de espalhamento por polímeros lineares em solução e é muito útil no estudo do enovelamento de proteínas. De fato, cada tipo de conformação (cadeia Gaussiana ideal, cadeia enovelada compacta, cadeias flexíveis com comprimento de persistência etc.) pode ser identificado pelo perfil característico do gráfico de Kratky associado. Assim, a intensidade espalhada por uma proteína globular compacta comporta-se aproximadamente como  $1/q^4$  no domínio de altos valores de  $q$ , dando lugar a um “*Kratky Plot*” na “forma de sino” e com um máximo bem definido (Curva azul da figura 1.3). Por outro lado, a representação de Kratky para uma cadeia Gaussiana ideal apresenta um *plateau* em altos valores de  $q$ , devido ao fato de que  $I(q)$  varia com  $1/q^2$  num intervalo específico de  $q$  seguido de um crescimento monotônico característico do espalhamento de um cilindro em valores maiores de  $q$  [Glatter & Kratky, 1982; Feigin & Svergun, 1987; Perez *et al.*, 2001]. A figura 1.3 exemplifica esses comportamentos.

Com base nessas observações, podem ser feitas afirmações de que a intensidade espalhada por uma proteína completamente desenovelada deve exibir esse último tipo de comportamento (curva vermelha da figura 1.3). Uma proteína parcialmente desestruturada deve exibir o comportamento representado pela curva preta da figura 1.3. Em particular, quando há a ausência de um aumento na intensidade em altos valores de  $q$  em uma representação de Kratky é uma prova de um desenovelamento incompleto. Na verdade, a relação entre a forma geral do gráfico de Kratky e a natureza de uma cadeia desenovelada é complexa. Dentro dessa estrutura do modelo de comportamento da intensidade para grandes valores de  $q$ , uma variedade de formas pode ser observada: nenhum aumento de intensidade em altos valores de  $q$ , ausência de um *plateau*, ou até mesmo uma diminuição de intensidade em altos valores de  $q$ , todas elas descrevendo uma cadeia completamente desenovelada. Em suma, a aparência do gráfico de Kratky de uma cadeia polipeptídica de seção finita depende das características específicas dessa cadeia que descreve a proteína, do raio de giro da seção transversal  $R_c$  e do intervalo de  $q$  explorado [Glatter & Kratky, 1982; Feigin & Svergun, 1987; Perez *et al.*, 2001].



**Figura 1.3** – Exemplos de representações de Kratky ( $q^2 I(q)$  vs.  $q$ ) da curva de intensidade para três tipos diferentes de estruturas: globular compacta (curva azul), parcialmente estruturada (curva preta) e desestruturada (curva vermelha). (A) Curvas experimentais de intensidade. (B) Representação de Kratky,  $q^2 I(q)$  vs.  $q$ , das curvas de intensidade.

## 1.6. TRATAMENTO DOS DADOS DE SAXS

Após a aquisição dos dados de espalhamento a baixo ângulo, algumas correções devem ser feitas antes que a curva de intensidade possa ser usada para o estudo de interesse. Aqui será apresentada uma breve introdução de como isso é feito. Uma explicação mais detalhada e apurada

do tratamento de dados de SAXS pode ser encontrada algumas referências [Glatter & Kratky, 1982; Koch *et al.*, 2003].

Para as experiências de SAXS de soluções, é preciso coletar o espalhamento da solução contendo as proteínas e o espalhamento do solvente em que a proteína está imersa, para descontar a contribuição do solvente. Assim, espera-se que a curva resultante contenha somente informações das moléculas de proteína. Além disso, ainda que a medida tenha sido feita numa câmara na qual o caminho que o feixe espalhado percorre esteja sob vácuo, existe sempre um espalhamento adicional causado pelos elementos ópticos inseridos no caminho do feixe após a amostra (fendas, janela etc.) entre a amostra e o detector. Esse espalhamento adicional é conhecido como espalhamento parasita. Assim, a subtração do espalhamento do solvente também estará eliminando esses efeitos da curva de espalhamento da amostra de interesse.

Os dados experimentais devem também ser corrigidos por absorção de raios-X tanto pela solução com a proteína quanto solvente. Para isso, durante a experiência temos que levar em conta a atenuação dessas soluções medindo o valor de  $A_t$  (= número de fótons incidentes dividido pelo número de fótons transmitidos) e usando  $A_t$  como fator multiplicativo da intensidade. Além disso, em experiências realizadas com radiação síncrotron, a intensidade incidente na amostra não é constante no tempo e, para corrigir pela variação de intensidade no tempo, os dados são divididos pela contagem de fótons incidentes na amostra e integrada no tempo de medida,  $NI_o$ . Por outra parte, se a resposta do detector sensível à posição não é homogênea em toda a região de detecção, será também preciso corrigir os dados por essa inhomogeneidade de detecção. Isso é feito dividindo os dados coletados pela curva de resposta do detector. Dessa forma, temos:

$$I_{corrigida}(q) = \left[ \frac{I_a(q) \cdot A_{t_a}}{NI_{o,a}} - \frac{I_s(q) \cdot A_{t_s}}{NI_{o,s}} \right] \frac{1}{RD} \quad (1.34)$$

onde  $I_a(q)$  e  $I_b(q)$  são as intensidades de espalhamento da amostra e do solvente, respectivamente,  $A_{t_a}$  é a atenuação da amostra (sempre um fator maior que  $I(um)$ ),  $A_{t_s}$  é a atenuação do solvente (tampão),  $NI_{o,a}$  é a intensidade integrada no tempo total de medida da amostra (número de fótons incidente durante a medida da amostra),  $NI_{o,s}$  é a intensidade integrada no tempo total de medida do solvente (número de fótons incidente durante a medida do solvente) e  $RD$  é curva de resposta do detector (inhomogeneidades). Para as análises dos dados de SAXS nós usamos a curva  $I_{corrigida}(q)$ .

## 1.7. PROCESSAMENTO DOS DADOS: TRANSFORMADA DE FOURIER DA INTENSIDADE

A função característica (função  $\mathcal{Y}(r)$ ) e a função de distribuição de pares de distâncias (função  $p(r)$ ) poderiam, em princípio, serem calculadas através de uma transformada de Fourier dos dados experimentais de intensidade. Porém, isso não é assim tão simples, porque a curva experimental  $I(q)$  que temos é medida com um número finito de pontos ( $q_i$ ) no intervalo  $[q_{\min}, q_{\max}]$  em vez de  $[0, \infty]$ . A precisão dessas medidas é determinada pelos erros estatísticos ( $\sigma_i$ ) correspondentes, mas sempre há também os erros sistemáticos.

Portanto, é necessário calcular a função  $p(r)$  indiretamente pela transformada inversa de Fourier, dentro do intervalo  $[0, D_{\max}]$ , da curva de intensidade experimental,  $I(q)$ , conforme equações (9) e (10). Para esse propósito, a função  $p(r)$  é representada por uma combinação linear de funções ortogonais  $\varphi_k$  (uma base de funções) no intervalo  $[0, D_{\max}]$  da seguinte maneira [Glatter, 1977]:

$$p(r) = \sum_{k=1}^K c_k \varphi_k(r) \quad (1.35)$$

A estimativa *a priori* de  $D_{\max}$  pode ser refinada no decorrer do processo por cálculos iterativos de  $p(r)$  com diferentes valores de  $D_{\max}$ . Os coeficientes  $c_k$  na equação (1.35) são determinados ajustando os dados experimentais e minimizando o seguinte funcional:

$$\Phi_{\alpha} = \sum_{i=1}^N \left[ \frac{I_{\text{exp}}(q_i) - \sum_{k=1}^K c_k \psi_k(q_i)}{\sigma(q_i)} \right]^2 + \alpha \int_0^{D_{\max}} \left( \frac{dp(r)}{dr} \right)^2 dr \quad (1.36)$$

onde as funções  $\psi_k(q)$  são as transformadas de Fourier das respectivas funções  $\varphi_k(q)$ . O multiplicador de regularização  $\alpha \geq 0$  é usado para fazer um balanço entre a qualidade do ajuste (primeiro termo do funcional da equação (1.36)) e a suavidade (“*Smoothing*”) da função  $p(r)$  (segundo termo do funcional da equação (1.36)). Esse método é o chamado método da transformada de Fourier indireta [Glatter, 1977].

O principal problema ao usar o método da transformada indireta é selecionar o valor apropriado do multiplicador de regularização  $\alpha$ . Com valores pequenos demais desse multiplicador, as soluções são instáveis com os erros experimentais, enquanto que valores grandes demais apresentam desvios sistemáticos dos dados experimentais. Em um algoritmo desenvolvido por

*Semenyk & Svergun* [Semenyk & Svergun, 1991; Svergun, 1992; Svergun, 1991], um conjunto de critérios percentuais descrevem a qualidade da solução para guiar a escolha do melhor  $\alpha$ . Esse algoritmo pode automaticamente selecionar a “solução ótima” ou, então, detectar soluções duvidosas (por exemplo, com o valor de  $D_{m\acute{a}x}$  muito distante do verdadeiro). Infelizmente, essa escolha nem sempre é totalmente confiável, necessitando da avaliação humana a fim de selecionar a melhor solução. A função  $p(r)$  calculada é então usada para obter o valor de  $I(0)$  usando a equação (1.21) e o valor de  $R_g$  usando a equação (1.20).

Como a curva de espalhamento inteira é usada no cálculo desses dois parâmetros, em vez do limitado intervalo angular em baixos ângulos da aproximação de Guinier (equação (1.18)) ou da aproximação de Debye (equação (1.22)), os valores obtidos são mais confiáveis e menos sensíveis a efeitos de interações residuais ou pequenos níveis de agregação. O uso dessas aproximações requer um número suficiente de pontos experimentais no intervalo  $q_{min} < q < 1/R_g$  para aproximação de Guinier, e  $q_{min} < q < 1,4/R_g$  para a aproximação de Debye, com  $q_{min} < \pi/D_{m\acute{a}x}$  enquanto que o método de transformada indireta permite o cálculo de funções  $p(r)$  sob condições menos exigentes. Uma vez que a função  $p(r)$  está disponível, tudo está pronto para o próximo passo, que é a modelagem tridimensional a partir dos dados de SAXS, como será brevemente explicado nas seções seguintes. Entretanto, para entender o processo de modelagem, é importante entender a conteúdo de informações contidas na curva de espalhamento a fim de ter curvas com informações suficientes para cada algoritmo de modelagem, como veremos na próxima seção.

## 1.8. QUANTIDADE DE INFORMAÇÃO: CANAIS DE SHANNON

A quantidade de informação de uma curva de intensidade de espalhamento é geralmente estimada com base no teorema de amostragem de Shannon [Shannon & Weaver, 1949]. Como indicado na equação (1.9), a curva de intensidade  $I(q)$  é a transformada de Fourier da função de distribuição de pares de distâncias, que por sua vez é igual a zero para distâncias intrapartícula excedendo  $D_{m\acute{a}x}$ . Após integrar duas vezes por partes a equação (1.9), nós encontramos uma relação da seguinte forma recorrente [Koch *et al.*, 2003]:

$$I(q) = \sum_{k=1}^{\infty} q_k I(q_k) \left\{ \frac{\text{sen}[(q - q_k)D_{m\acute{a}x}]}{(q - q_k)D_{m\acute{a}x}} - \frac{\text{sen}[(q + q_k)D_{m\acute{a}x}]}{(q + q_k)D_{m\acute{a}x}} \right\} \quad (1.37)$$

Isso significa que a função contínua  $I(q)$  pode ser representada por seus valores em um conjunto discreto de pontos (Canaís de Shannon) onde  $q_k = k\pi/D_{m\acute{a}x}$ . O número de parâmetros (ou graus de liberdade) requeridos para representar uma função analítica no intervalo  $[q_{min}, q_{m\acute{a}x}]$  é,

portanto, dada pelo número de canais de Shannon neste intervalo. Esse número pode ser calculado através da seguinte expressão:

$$N_s = (q_{m\acute{a}x} - q_{m\acute{i}n}) \frac{D_{m\acute{a}x}}{\pi} \quad (1.38)$$

Na prática, as curvas de intensidade de espalhamento decaem rapidamente com  $q$  e elas são normalmente detectadas somente em resoluções menores que  $1nm$ , tal que o número de canais de Shannon ( $N_s$ ) tipicamente não excede *10 (dez)* ou *15(quinze)* canais. A questão aqui é se  $N_s$  representa o número máximo de parâmetros independentes que podem ser extraídos dos dados de espalhamento. Isso tem sido muito discutido durante os anos [Moore, 1980; Taupin & Luzzati, 1982]. Grandes valores de  $N_s$  claramente correspondem a um maior conteúdo de informações nos dados, mas ao mesmo tempo a série na equação (1.37) deve conter um número infinito de termos para a intensidade de espalhamento ter sentido físico.

Extração de parâmetros estruturais de dados de espalhamento de uma solução é geralmente um problema não-linear, de tal forma que o número de parâmetros não pode ser realmente predito considerando o intervalo angular experimental somente. O número de canais de Shannon realmente fornece um guia muito útil para realizar uma boa medida. Em particular, o valor de  $q_{m\acute{i}n}$  a ser utilizado nos cálculos não deve ser menos que aquele do primeiro canal de Shannon ( $q_{m\acute{i}n} < \pi/D_{m\acute{a}x}$ ). Todavia, o nível de detalhes de modelos de baixa resolução, que podem ser obtidos dos dados de SAXS, também depende de outros fatores como a precisão dos dados ou das informações disponíveis *a priori*.

## **1.9. DETERMINAÇÃO DE FORMA: MODELAGEM TRIDIMENSIONAL A PARTIR DOS DADOS DE SAXS**

Curvas de intensidade de espalhamento de raios-X de soluções monodispersas de proteínas contem informação sobre a estrutura tridimensional dessa proteína. Dessa forma, é possível construir um modelo tridimensional da proteína graças ao fato de que, para soluções monodispersas, a intensidade de espalhamento é proporcional ao espalhamento de uma única partícula. Porém, a quantidade de informação nos dados de espalhamento é drasticamente reduzida pelas orientações aleatórias das partículas em solução. Por isso, o modelo que conseguimos calcular é de baixa resolução, fornecendo uma representação tridimensional somente do envelope molecular. Em vista disso, notamos que a reconstrução de um modelo tridimensional de um objeto a partir dos dados unidimensionais de espalhamento é complicada e sem solução única. Informações *a priori*

(seqüência de aminoácidos da proteína, domínios com estruturas conhecidas, estruturas de moléculas homólogas) sobre a proteína em estudo são requeridas para obter um bom resultado.

A primeira idéia para uma modelagem *ab initio* a partir de dados de SAXS foi a abordagem padrão de tentativa-e-erro, onde as curvas de espalhamento eram calculadas de modelo com diferentes formas e comparadas com as curvas experimentais. Porém, isso mudou após Stuhrmann [Stuhrmann, 1970] introduzir a expansão multipolar, representando a amplitude de espalhamento da seguinte forma:

$$A(\vec{q}) = \sum_{l=0}^L \sum_{m=-1}^1 A_{lm}(q) Y_{lm}(\Omega) \quad (1.39)$$

onde  $Y_{lm}(\Omega)$  são harmônicos esféricos e o truncamento da série no valor  $L$  determina a precisão dessa expansão. Os coeficientes  $A_{lm}(q)$  são denominados amplitudes parciais de espalhamento, dados pela relação:

$$A_{lm}(q) = 4\pi^l \int_0^\infty j_l(qr) \rho_{lm}(r) r^2 dr \quad (1.40)$$

onde  $j_l(qr)$  são funções de Bessel esféricas e  $\rho_{lm}(r)$  é a expansão multipolar da densidade eletrônica dada por:

$$\rho_{lm}(\vec{r}) = \int_{\omega} \rho(\vec{r}) Y_{lm}^*(\omega) d\omega. \quad (1.41)$$

Dessa forma, devido a relação de ortogonalidade dos harmônicos esféricos, podemos escrever uma expressão simples para a intensidade de espalhamento de uma partícula:

$$I(q) = 2\pi^2 \sum_{l=0}^L \sum_{m=-1}^1 |A_{lm}(q)|^2. \quad (1.42)$$

Com base nisso, algumas abordagens *ab initio* foram desenvolvidas para dar conta dessa modelagem e iremos discuti-las nas próximas seções.

### 1.9.1. ESTRATÉGIA GERAL DE MODELAGEM

A tarefa de restaurar o modelo de baixa resolução de uma proteína a partir dos dados de SAXS foi inicialmente idealizada por Glatter [Glatter, 1988; Pilz *et al.*, 1972; Glatter, 1972] e depois automatizada por Svergun [Svergun, 1999]. Ela pode ser formulada como segue [Svergun, 1999]: dada uma certa configuração espacial tridimensional composta por elementos menores (digamos pequenas esferas) e representada por um vetor de posições  $X_0$  (com  $N$  elementos), temos que encontrar a configuração  $X_f$  (também com  $N$  elementos) que minimiza a função alvo:

$$E(X) = \chi^2[I(q), I_{\text{exp}}(q)] + \sum_i \alpha_i P_i \quad (1.43)$$

onde  $\alpha > 0$  é o peso das penalidades, que por sua vez consideram informações adicionais sobre a proteína, como por exemplo, informações de contato entre os elementos menores da configuração, evitando superposição entre eles, além de evitar que o modelo fique com partes desconectadas (frouxo). O termo  $\chi^2[I(q), I_{\text{exp}}(q)]$  é a discrepância entre os dados experimentais de intensidade  $I_{\text{exp}}(q)$  e os valores de intensidade calculados  $I(q)$  do modelo:

$$\chi^2[I(q), I_{\text{exp}}(q)] = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(q_j) - cI(q_j)}{\sigma(q_j)} \right]^2 \quad (1.44)$$

onde  $\sigma(q_j)$  é o erro experimental na curva experimental de espalhamento e  $c$  é uma constante de ajuste. Geralmente, o peso de cada penalidade é selecionado tal que o segundo termo na equação (1.43) tenha uma contribuição significativa (digamos, entre 10 e 50 %) para a função alvo no final do processo de minimização.

## 1.9.2. MÉTODO DE MINIMIZAÇÃO: ALGORITMO DE RECOZIMENTO SIMULADO

Dado o grande número de variáveis e a natureza combinatória do problema de modelagem, é preciso um protocolo para a minimização da equação (1.43) e construção de um modelo da proteína a partir dos dados de SAXS. O algoritmo de recozimento simulado (tradução livre do termo em inglês *Simulated annealing*) [Kirkpatrick *et al.*, 1983] parece ser o mais apropriado para a implementação dessa tarefa, já que é um método de minimização global. A idéia principal do método é realizar modificações aleatórias no sistema, isto é, no vetor de posições  $X$ , movendo o sistema para uma configuração que diminui o valor de  $E(X)$ , mas algumas vezes pode até aumentar seu valor. A probabilidade de aceitação desse último movimento (passo aleatório que aumenta  $E(X)$ ) diminui no decorrer do processo dado que o sistema é resfriado (a temperatura é reduzida). No começo, a temperatura<sup>vi</sup> é alta e as mudanças são quase todas aleatórias, enquanto que próximo do final, essas mudanças aleatórias já estão reduzidas e um mínimo da função  $E(X)$  é alcançado na configuração final do sistema. Esquemmatizando o método de minimização utilizando o protocolo de recozimento simulado, ele fica da seguinte forma [Svergun, 1999]:

---

<sup>vi</sup> O termo temperatura aqui não tem o significado físico conhecido de temperatura. O nome só é usado em razão do algoritmo que usa a idéia de probabilidade de ocupação de um estado da função de Boltzmann.

1. Começa com uma configuração inicial aleatória  $X=X_0$  (vetor com  $N$  elementos) em uma alta temperatura  $T_0$ .
2. Seleciona aleatoriamente um elemento do vetor (digamos, uma pequena esfera) e aleatoriamente troca sua posição, ficando com o vetor  $X'$ , e calcula a diferença  $\Delta E = E(X') - E(X)$ .
3. Se  $\Delta E < 0$ , aceita o passo e move o sistema para  $X'$ . Se  $\Delta E > 0$ , testa a probabilidade de aceitação com  $\exp(-\Delta E/T)$ . Daí, repete o passo 2 fazendo  $X=X'$  (se aceito) ou mantendo  $X$  (caso contrário).
4. Mantém a temperatura  $T$  constante até que haja  $(100 \times N)$  reconfigurações ou  $(10 \times N)$  sucessos, o que vier primeiro, então resfria o sistema:  $T' = 0,9 T$ .
5. Continua repetindo o processo e resfriando até que nenhuma melhora em  $E(X)$  seja mais observada no sistema.

### 1.9.3. MODELAGEM *AB INITIO* COM ÁTOMOS FICTÍCIOS

O método de modelagem *ab initio* com átomos fictícios (do inglês *Dummy Atoms*) é o método mais conhecido e usado para modelar proteínas em solução, porém, é o método de menor resolução dentre os outros métodos. Esse método é muito útil para restaurar o envelope molecular de baixa resolução de uma proteína até resolução de até cerca de 20 Å. Vamos à descrição do método. Primeiro, um volume esférico de raio  $R=D_{máx}/2$  é preenchido com  $N$  átomos fictícios, isto é, pequenas esferas de raio  $r_0 \ll R$  densamente empacotadas (figura 1.4). Para cada átomo fictício (pequena esfera) é associado um índice  $X_j$  indicando a fase a qual ele pertence dentro de um vetor de fases  $X$ . As fases podem ser igual a  $0$  (zero), significando que ele pertence ao solvente, ou  $1$  (um), se pertence a partícula. É possível usar mais fases nesse método se quisermos designar diferentes domínios dentro da partícula, mas isso só em caso especiais e com muito cuidado, sendo extremamente necessário informações sobre a localização desses domínios na partícula e suas regiões de contato. Continuando, dada a posição fixa dos átomos fictícios, a forma e estrutura do modelo são completamente descritas pelo vetor (configuração) de associação de fases  $X$  com  $N \approx (R/r_0)^3$  componentes.

Assumindo que os átomos fictícios da  $k$ -ésima fase ( $0$  ou  $1$ , para solvente ou partícula) tenha contraste  $\Delta\rho_k$ , a intensidade de espalhamento do modelo será:

$$I(q) = \left\langle \sum_{k=1}^N \Delta \rho_k A_k(\vec{q}) \right\rangle_{\Omega} \quad (1.45)$$

onde  $A_k(\vec{q})$  é a amplitude de espalhamento do volume ocupado pelo átomo fictício na  $k$ -ésima posição. Representando as amplitudes por harmônicos esféricos  $Y_{lm}(\Omega)$ :

$$A_k(\vec{q}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}^k(q) Y_{lm}(\Omega) \quad (1.46)$$

obtemos [Stuhrmann, 1970; Svergun, 1994; Svergun, 1999]:

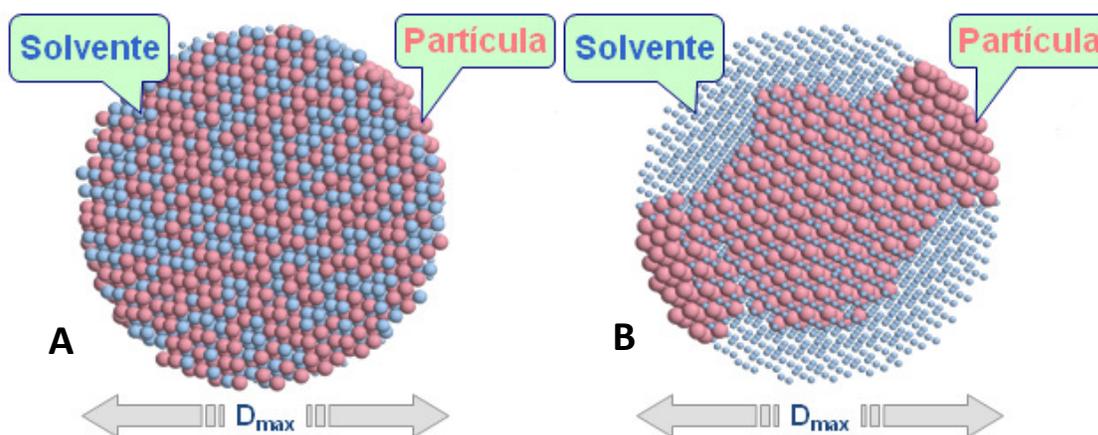
$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \left\{ \sum_{k=1}^N [\Delta \rho_k A_{lm}^k(\vec{q})]^2 + 2 \sum_{n>k} \Delta \rho_k A_{lm}^k(\vec{q}) \Delta \rho_n [A_{lm}^n(\vec{q})]^* \right\} \quad (1.47)$$

As amplitudes parciais do volume ocupado pela  $k$ -ésima fase do modelo são dadas pela relação [Svergun, 1999]:

$$A_{lm}^k(q) = 4\pi i^l f(q) \sum_{j=1}^{N_k} j_l(qr_j) Y_{lm}^*(\omega_j) \quad (1.48)$$

onde a soma corre sobre os átomos fictícios da  $k$ -ésima fase,  $r_j$  e  $\omega_j = (\theta_j, \varphi_j)$  são suas coordenadas polares,  $j_l(qr_j)$  são funções de Bessel esféricas, e  $f(q)$  é o fator de espalhamento de cada átomo fictício (fator de espalhamento atômico). As equações (1.47) e (1.48) permitem o cálculo da curva de intensidade de espalhamento de um modelo para uma configuração arbitrária  $X$  e contrastes arbitrários  $\Delta \rho_k$ .

Começando da configuração arbitrária inicial, o protocolo de recozimento simulado é empregado para minimizar a função alvo  $E(X)$  (equação (1.43)), onde  $I(q)$  é calculada pela equação (1.47) e  $I_{exp}(q)$  é a curva experimental. Somente um átomo fictício é trocado por movimento tal que somente uma somatória na equação (1.48) deve ser atualizada por vez para calcular as amplitudes parciais. Como esse passo é a operação que mais consome tempo, isso acelera o processo, o que torna possível o uso do protocolo de recozimento simulado, que é muito robusto [Ingber, 1993], mas seria proibidamente devagar demais, considerando que milhões de cálculos de funções são requeridos para um processo típico de modelagem.



**Figura 1.4** – Configurações de esferas densamente empacotadas do processo de modelagem *ab initio* com átomos fictícios. (A) Configuração inicial dos átomos fictícios da partícula (em rosa) e do solvente (em azul). (B) Resultado final da modelagem. A forma do envelope molecular da partícula é mostrada pela organização dos átomos fictícios em rosa.

#### 1.9.4. MODELAGEM *AB INITIO* COM RESÍDUOS FICTÍCIOS

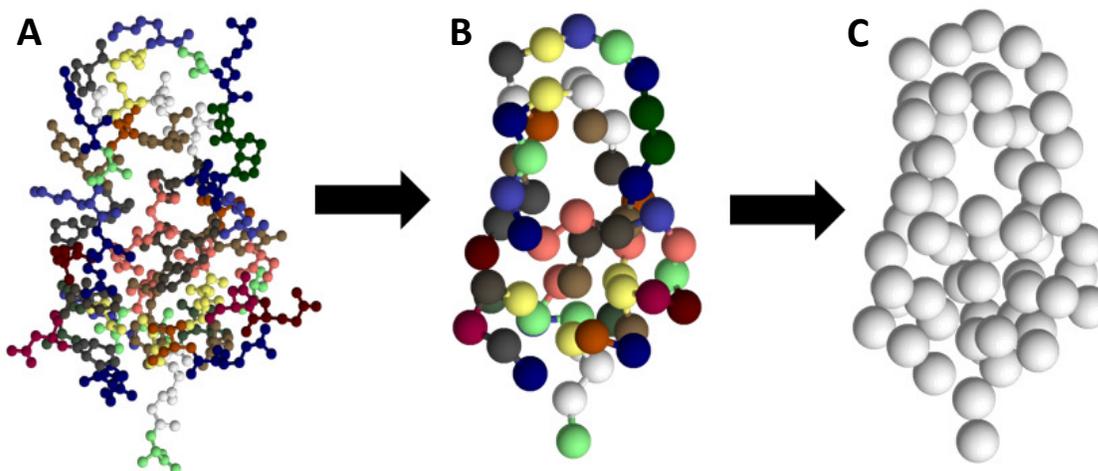
Nesta abordagem, a proteína é representada como uma cadeia de resíduos simulados (do inglês *Dummy Residues* - DRs), que são pequenas esferas conectadas em forma de cadeia polipeptídica. Esse método leva em conta a curva inteira de espalhamento e não somente a porção inicial, que corresponde à forma somente [Svergun *et al.*, 2001]. Na referência Svergun *et al.*, 2001 há uma figura que mostra as diferentes porções de uma curva de intensidade e que tipo de informação cada uma fornece.

Proteínas tipicamente consistem de cadeias polipeptídicas compostas de resíduos de aminoácidos separados por uma distância de aproximadamente  $3,8 \text{ \AA}$  medida entre átomos do  $C_{\alpha}$  adjacentes na estrutura. Em uma resolução de  $5 \text{ \AA}$ , uma estrutura de proteína pode ser considerada como um aglomerado de resíduos fictícios centrados nas posições dos carbonos-alfa ( $C_{\alpha}$ ) [Svegun *et al.*, 2001], como apresentado na figura 1.5. O número de resíduos fictícios é geralmente conhecido *a priori* a partir da seqüência de aminoácidos da proteína. Dessa forma, considerando um conjunto de resíduos fictícios distribuídos aleatoriamente dentro de um volume esférico de diâmetro  $D_{m\acute{a}x}$ , o protocolo de recozimento simulado é utilizado para minimizar a função  $E(X)$  da equação (1.43). Dessa forma, o método procura um arranjo espacial tipo cadeia polipeptídica dos resíduos fictícios cujo espalhamento calculado ajuste os dados experimentais de espalhamento. Para representar a camada de solvatação, o modelo da proteína é envolvido por uma camada de hidratação de espessura  $\Delta r = 3 \text{ \AA}$  representada por pequenas esferas. A intensidade de espalhamento

do arranjo de  $K$  resíduos fictícios com coordenadas  $r_i$  é calculada usando a fórmula de Debye [Debye, 1915]:

$$I(q) = \sum_{i=1}^K \sum_{j=1}^K g_i(q)g_j(q) \frac{\text{sen}(qr_{ij})}{qr_{ij}} \quad (1.49)$$

onde  $g_i(q)=f(q)$  é o fator de forma atômico para um resíduo fictício e  $g_i(q) = (4\pi r_i^2)\Delta r \delta\rho_b$  [Svergun *et al.*, 2001] é o equivalente do fator de forma atômico para um átomo fictício do solvente, e  $r_{ij} = |r_i - r_j|$  é a distância entre o  $i$ -ésimo e o  $j$ -ésimo ponto.



**Figura 1.5** – Representação esquemática de uma proteína como um arranjo de resíduos fictícios para a modelagem *ab initio*. (A) Estrutura de uma proteína mostrando seus resíduos de aminoácidos em cores diferentes. (B) Modelo para a estrutura da proteína com os resíduos sendo representados ainda por cores, mas cada resíduo está sendo mostrado na forma de uma esfera com um tamanho médio de um aminoácido. (C) Modelo típico para uma proteína calculado com o método de resíduos fictícios, onde os resíduos também são representados por pequenas esferas, sem distinção entre eles, porém mantendo o arranjo tipo cadeia polipeptídica.

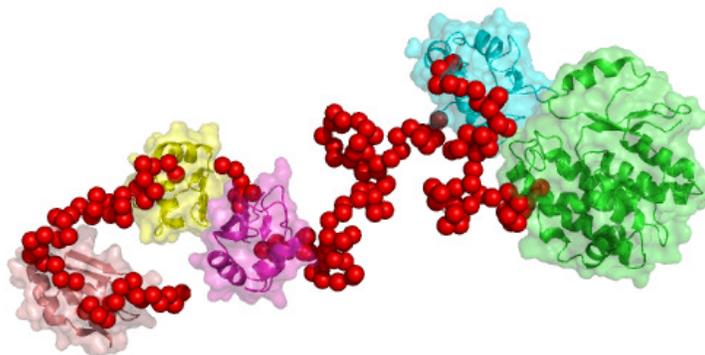
### 1.9.5. MODELAGEM DE CORPO RÍGIDO COMBINADA COM CÁLCULOS *AB INITIO*

Muitas vezes, a proteína estudada possui estruturas de alta resolução de domínios ou partes da sua estrutura já resolvidas por outras técnicas, mas falta conhecer o arranjo estrutural da proteína inteira. Para resolver isso usando a modelagem a partir dos dados de SAXS, há um método que

combina cálculos *ab initio* com modelagem de corpos rígidos<sup>vii</sup>. As estruturas das regiões disponíveis são tratadas como corpos rígidos conectados por regiões flexíveis de estrutura desconhecida e criadas com modelagem *ab initio* de resíduos fictícios [Petoukhov *et al.*, 2002; Petoukhov *et al.*, 2005]. O protocolo de recozimento simulado é empregado para encontrar as posições e orientações ótimas desses corpos rígidos e as prováveis conformações das regiões flexíveis conectadas no apropriado resíduo do domínio. Aqui, a seqüência de aminoácidos da proteína é essencial para montar o modelo e deve ser levada em conta no decorrer do processo. A intensidade de espalhamento do modelo inteiro é expressa por [Petoukhov *et al.*, 2005]:

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \left| \sum_k A_{lm}^k(q) + \sum_i D_{lm}^i(q) \right|^2 \quad (1.50)$$

Aqui,  $A_{lm}^k(q)$  são as amplitudes parciais de espalhamento dos domínios rígidos calculados de suas coordenadas atômicas como descrito por Svergun *et al.*, 1995 [Svergun *et al.*, 1995]. As amplitudes parciais de espalhamento das regiões flexíveis que conectam esses domínios são calculadas a partir das posições dos resíduos fictícios [Petoukhov *et al.*, 2002; Svergun *et al.*, 2001]. Começando de uma configuração arbitrária de domínios e regiões conectoras, modificações aleatórias são feitas no modelo a fim de minimizar a função alvo  $E(X)$  da equação (1.43). Esse método é muito útil para modelar proteínas flexíveis ou com regiões desestruturadas como são as proteínas com múltiplos domínios e as proteínas intrinsecamente desestruturadas, desde que haja ao menos uma estrutura já resolvida de alguma porção da proteína.



**Figura 1.6** – Representação típica de um modelo de uma proteína com múltiplos domínios obtido dos dados de SAXS usando a combinação de cálculos *ab initio* e modelagem de corpos rígidos. As regiões flexíveis criadas por cadeias de resíduos fictícios são mostradas vermelho.

<sup>vii</sup> Um sistema material  $S$  comporta-se como um corpo rígido se a distância entre qualquer par de pontos  $P_i$  e  $P_j$  pertencentes a esse sistema permanece constante durante sua evolução, ou seja, o corpo não sofre nenhuma deformação durante o cálculo, somente é transladado ou rotacionado.

### 1.9.6. MÉTODO DE OTIMIZAÇÃO DE CONJUNTOS DE CONFORMAÇÕES

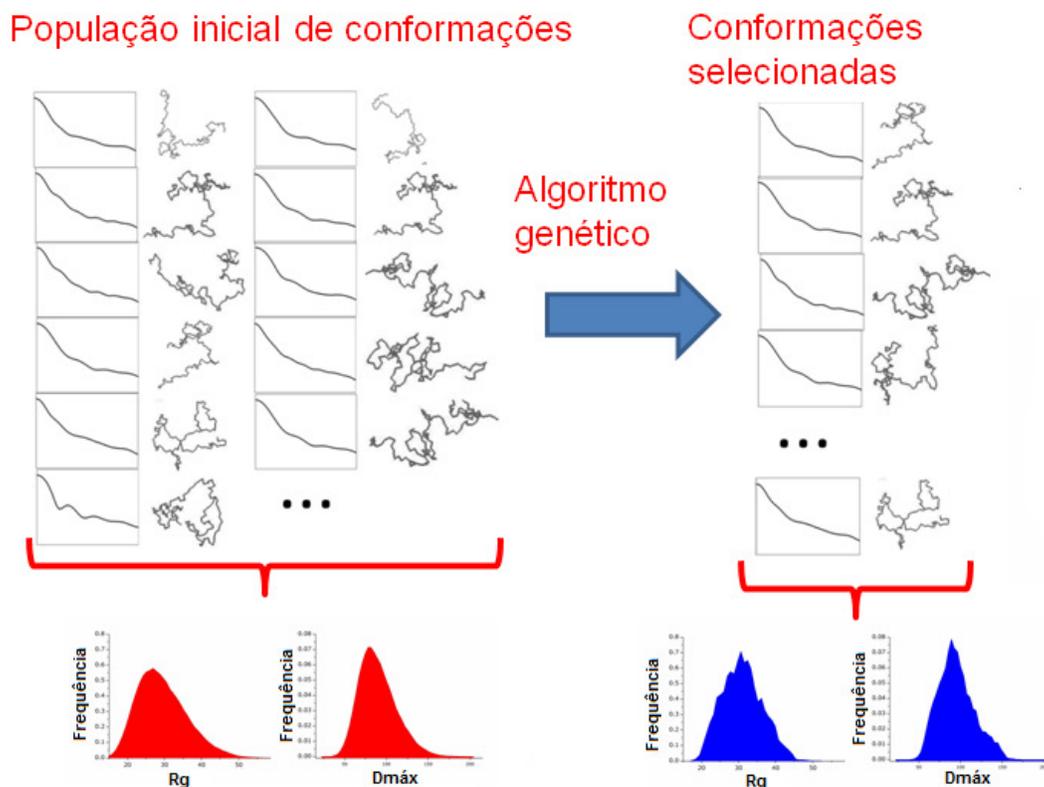
O método de otimização de conjuntos de conformações foi desenvolvido para a análise de sistemas flexíveis, em particular, proteínas com múltiplos domínios conectados por regiões flexíveis e proteínas desenoveladas ou parcialmente desestruturadas. Neste método, é assumido (podendo ser mostrado matematicamente) que os dados experimentais de intensidade de espalhamento possam ser representados por uma média do espalhamento das diferentes conformações da molécula [Bernadó *et al.*, 2007]. Dessa forma, para  $N$  diferentes conformações:

$$I(q) = \frac{1}{N} \sum_{j=1}^N I_j(q) \quad (1.51)$$

onde  $I_j(q)$  é a intensidade de espalhamento da  $j$ -ésima conformação.

Uma população com um grande número ( $M \gg N$ ) de possíveis conformações é gerada usando o método de Monte-Carlo para cobrir todo o espaço de configurações. Essas conformações são construídas assumindo que os planos dos peptídeos são entidades rígidas conectadas através dos átomos de carbono-alfa ( $C_\alpha$ ) da cadeia polipeptídica. As conformações de aminoácidos são aleatoriamente selecionadas de uma biblioteca de uma base de dados de conformações de estruturas de alta resolução [Lovell *et al.*, 2003]. Os ângulos diédricos nas ligações  $C_\alpha-C_\alpha-C_\alpha$  são selecionados para concordar com o histograma de ângulos de ligação de cadeias de proteínas conhecidas e os ângulos diédricos (ou ângulo de torção nas cadeias  $C_\alpha-C_\alpha-C_\alpha-C_\alpha$ ) são selecionados de tal forma a concordar com o diagrama quase-Ramachandran [Bernadó *et al.*, 2007; Kleywegt, 1997], além evitar contatos da cadeia com ela mesma. Esse histograma de ângulos de ligação e o diagrama quase-Ramachandran estão apresentados nas referências Svergun *et al.* [2001] e Petoukhov *et al.* [2002]. Uma lista de 500 proteínas tipicamente usada para a construção de diagramas de Ramachandran pode ser encontrada no endereço eletrônico: <http://kinemage.biochem.duke.edu/databases/top500.php>. Isso faz com que a cadeia polipeptídica do modelo seja bastante próxima de uma cadeia real de proteína. Detalhes podem ser encontrados no trabalho de Bernadó *et al.* [Bernadó *et al.*, 2005]. Depois disso, o algoritmo genético é aplicado para selecionar os conjuntos mais apropriados de conformações que ajustam os dados experimentais de SAXS [Bernadó *et al.*, 2007]. Os modelos resultantes dentro do conjunto que gera o melhor ajuste mostram possíveis conformações adotadas pela proteína em solução. A figura 1.7 apresenta um esquema do método. Esses cálculos resultam também em um conjunto de conformações cujos valores médios de raio de giro ( $R_g$ ) e dimensão máxima ( $D_{máx}$ ) concordam com aqueles obtidos de nossas análises prévias e convencionais de SAXS (a lei de Guinier e a transformada de Fourier

indireta), mas agora os cálculos fornecem as distribuições desses valores para o conjunto todo obtido por esse método.



**Figura 1.7** – Esquemática do método de otimização de conjuntos de conformações de proteínas em solução. À esquerda estão mostradas as curvas calculadas das respectivas conformações da população inicial e a direita estão mostradas aquelas do grupo selecionado de conformações. Abaixo de cada grupo estão mostradas representações das distribuições de raios de giro ( $R_g$ ) e dimensões máximas ( $D_{máx}$ ) de cada grupo.

## 1.10. ANALISANDO DIFERENTES SOLUÇÕES NO PROCESSO DE MODELAGEM

Devido ao fato de não existir uma solução única nos métodos *ab initio* de modelagem a partir dos dados de SAXS, é necessário encontrar alguma maneira de analisar essas soluções e selecionar as melhores e mais prováveis. Para cumprir essa tarefa, há um procedimento que é descrito para analisar essas múltiplas soluções e estimar a estabilidade e confiabilidade dessas reconstruções tridimensionais da forma da proteína. Nesse procedimento, os modelos obtidos são alinhados entre eles, aos pares, e comparadas suas diferenças e semelhanças [Kozin & Svergun,

2001; Volkov & Svergun, 2003]. Para isso há um algoritmo de alinhamento e comparação de modelos. Dentro desses métodos, cada modelo é representado por um conjunto de pontos (as posições dos átomos fictícios ou resíduos fictícios, por exemplo). Começando de um alinhamento dos eixos de inércia de cada modelo, tratado como um objeto topológico, o algoritmo minimiza o valor da discrepância espacial normalizada (abreviada aqui como NSD devido à expressão em inglês: *Normalized Spatial Discrepancy*), que é usada como uma estimativa quantitativa de similaridade entres os modelos comparados. Dessa forma, esse método também pode ser usado para uma análise comparativa de modelos estruturais obtidos por diferentes métodos, por exemplo, estruturas atômicas de alta resolução obtidas por cristalografia e modelos de baixa resolução obtidos dos dados de espalhamento ou de microscopia eletrônica.

Portanto, vamos introduzir a discrepância espacial normalizada (NSD) de acordo com a definição de Kozin & Svergun [Kozin & Svergun, 2001]. Para todos os pontos  $s_{1i}$  do primeiro conjunto  $S_1 = \{s_{1i}, i = 1, \dots, N_1\}$ , representando um modelo, o valor mínimo das distâncias entre  $s_{1i}$  e todos os pontos no segundo conjunto  $S_2 = \{s_{2i}, i = 1, \dots, N_2\}$ , representando outro modelo, é denotado por  $\rho(s_{1i}, S_2)$ . O valor de NSD entre os conjuntos  $S_1$  e  $S_2$  é definido como uma média normalizada dos valores de  $\rho(s_{1i}, S_2)$  dada por [Kozin & Svergun, 2001]:

$$\rho(S_1, S_2) = \left\{ \frac{1}{2} \left[ \frac{1}{N_1 d_2^2} \sum_{i=1}^{N_1} \rho^2(s_{1i}, S_2) + \frac{1}{N_2 d_1^2} \sum_{i=1}^{N_2} \rho^2(s_{2i}, S_1) \right] \right\}^{1/2} \quad (1.52)$$

onde  $N_i$  é o número de pontos em  $S_i$  e  $d_i$  é a distância média entre os pontos vizinhos em  $S_i$ . Quando há uma superposição ideal de objetos, NSD fica próxima entre 0 e 1, indicando que os objetos são bastante similares, mas excede 1 se os objetos são sistematicamente diferentes um do outro.

Tendo feito o alinhamento e calculado os valores de NSD entre cada par dentro do grupo de modelos, um valor médio sobre todos os pares  $\langle NSD \rangle$  e a dispersão  $\Delta(NSD)$  são então calculados. Para cada modelo, o valor de sua NSD com respeito ao resto do conjunto é calculado. A reconstrução com menor NSD é então selecionada como referência. Aqueles modelos cujos valores de NSD excedem  $\langle NSD \rangle + 2\Delta(NSD)$  são desconsiderados. Todos os modelos, exceto os descartados, são superpostos ao modelo de referência. Feito isso, cria-se um modelo médio, porém esse modelo pode ter um volume excluído (Volume de Prod) maior que o dos modelos individuais. Por isso, esse modelo médio é então filtrado para assegurar que o seu volume seja igual ao volume excluído médio do grupo de modelos. Esse último procedimento é descrito por Volkov & Svergun [Volkov & Svergun, 2003].

## 1.11. BREVE REVISÃO SOBRE PROTEÍNAS

As proteínas são constituintes básicos dos organismos vivos e fazem parte de uma classe fundamental de moléculas em biologia. As proteínas são formadas por um arranjo linear de resíduos de aminoácidos<sup>viii</sup> ligados através de ligações peptídicas (ligações covalentes). A seqüência de aminoácidos em uma proteína é definida pelo código genético. Elas são essenciais para a estrutura e funcionamento de todas as células dos seres vivos. Cada uma é única do ponto de vista de sua estrutura primária. Sua forma, que é o arranjo tridimensional das cadeias polipeptídicas de aminoácidos no seu estado nativo, está intimamente relacionada à sua função [Lehninger *et al.*, 1995; Voet & Voet, 1995].

Muitas proteínas são enzimas que catalisam as reações bioquímicas e são vitais para o metabolismo. Outras proteínas possuem funções estruturais ou mecânicas, tais como as proteínas que formam o citoesqueleto das células [Voet & Voet, 1995]. As proteínas são também importantes na marcação de células, respostas de imunidade, adesão da célula e no ciclo celular<sup>ix</sup>. Elas também formam um componente necessário em nossa dieta, desde que animais e seres humanos não podem sintetizar todos os aminoácidos essenciais e devem obtê-los através dos alimentos. Através do processo de digestão, os seres vivos quebram a proteína ingerida em aminoácidos livres que podem ser usados para a síntese de outras proteínas.

## 1.12. DIFERENTES NÍVEIS ESTRUTURAIS DAS PROTEÍNAS

As proteínas são macromoléculas com quatro aspectos distintos em suas estruturas [Lehninger *et al.*, 1995; Voet & Voet, 1995], que são classificados da seguinte forma:

Estrutura primária: consiste da seqüência linear de resíduos de aminoácidos formando uma estrutura covalente. Ela contém as informações e instruções que guiam a organização estrutural da proteína.

Estrutura secundária: consiste de estruturas locais regularmente repetidas estabilizadas por pontes de hidrogênio. Essas estruturas locais podem ser as alfa-hélices ou as folhas-beta como mostradas na figura 1.8(B).

Estrutura Terciária: corresponde à forma global de uma proteína. É o arranjo tridimensional que a proteína assume para realizar suas funções. Essa estrutura exhibe a relação espacial entre as

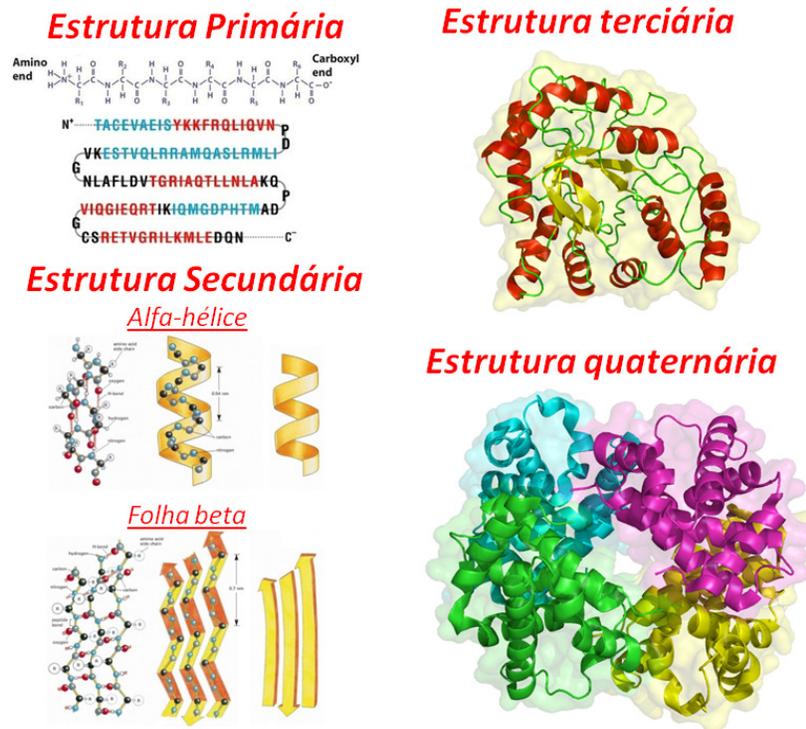
---

<sup>viii</sup> Quando dois ou mais aminoácidos perdem uma molécula de água ao formar uma cadeia polipeptídica, o que permanece de cada aminoácido é chamado resíduo de aminoácido.

<sup>ix</sup> O ciclo celular, ou ciclo de divisão da célula, é a série de eventos que acontecem na célula induzindo a sua divisão e duplicação (replicação).

estruturas secundárias. Estruturas terciárias são geralmente estabilizadas por interações não locais, mais comumente a formação de um núcleo hidrofóbico, mas também por ligações de hidrogênio, ligações de dissulfeto, entre outras. O termo estrutura terciária é frequentemente usado para se referir à proteína enovelada.

Estrutura Quaternária: forma ou estrutura que resulta da interação de mais de uma subunidade, iguais ou diferentes, de proteína para complementar sua função. É uma associação estável de múltiplas cadeias polipeptídicas que resulta em uma unidade ativa.



**Figura 1.8** – Organização hierárquica dos quatro diferentes aspectos das estruturas de proteínas.

A figura 1.8 exemplifica bem esses quatro aspectos distintos da estrutura de uma proteína. No contexto desses arranjos estruturais, as estruturas terciárias e quaternárias são normalmente referidas como conformações. Além dos níveis estruturais, as proteínas podem ser divididas em três classes principais, que estão correlacionadas com as estruturas terciárias típicas, e que são: proteínas globulares, proteínas fibrosas e proteínas de membrana. Porém, atualmente existem evidências experimentais que sugerem mais duas importantes classes de proteínas: as proteínas intrinsecamente desestruturadas e as proteínas flexíveis com múltiplos domínios.

Quase todas as proteínas globulares são solúveis em água e muitas são enzimas. Já as proteínas fibrosas são frequentemente estruturais e as proteínas de membrana geralmente servem

como receptores ou como canais para moléculas carregadas ou polares passarem através da membrana celular. Já as proteínas intrinsecamente desestruturadas possuem poucos elementos de estrutura secundária e funcionam como proteínas transportadoras, fazem reconhecimento molecular, modificações pós-traducionais, sinalização celular, além de estarem envolvidas em mecanismos regulatórios da célula. Quanto às proteínas flexíveis com múltiplos domínios, essas possuem domínios estruturados conectados por regiões flexíveis e tipicamente pertencem a essa classe as chaperonas (proteínas que auxiliam outras proteínas a se enovelarem) e proteínas que interagem com o DNA ou o RNA. Maiores detalhes dessas proteínas serão discutidas no capítulo 3 para proteínas intrinsecamente desestruturadas e no capítulo 4 para proteínas flexíveis com múltiplos domínios.

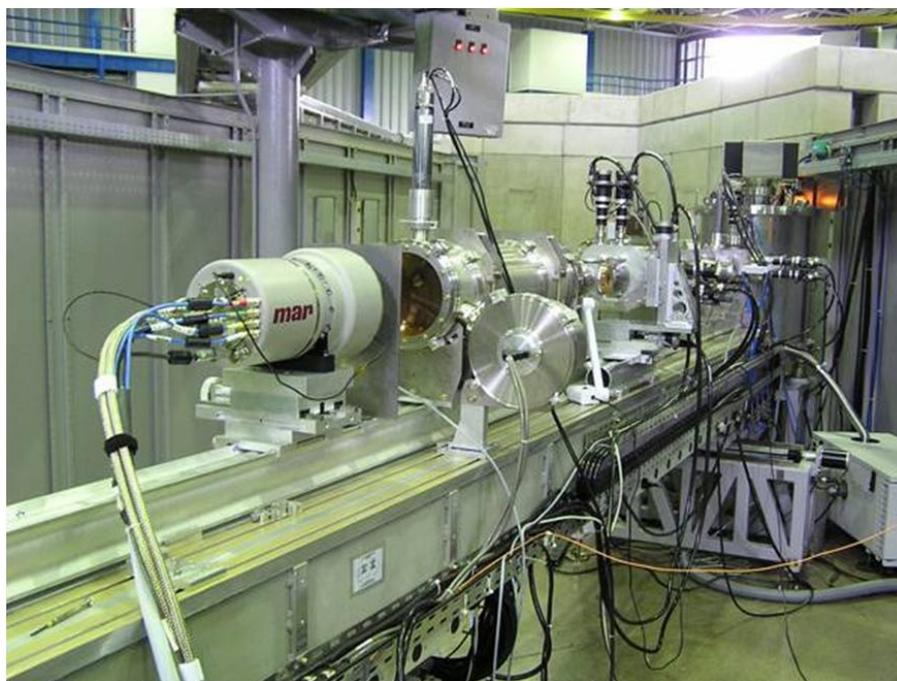
# 2

## **Instrumentação dedicada à realização de experiências de espalhamento de raios-X no Laboratório Nacional de Luz Síncrotron**

### **2.1. INTRODUÇÃO**

A primeira linha de luz dedicada à técnica de SAXS do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, SP, Brasil) foi aberta aos usuários no ano de 1997. Devido à grande demanda de usuários para a linha SAXS1, uma nova linha de luz, denominada SAXS2, com grandes melhorias na óptica de enfoque e a mecânica (banco óptico) da estação experimental começou a ser construída em 2005. Essa nova linha de luz foi aberta à comunidade de usuários em 2007 e uma foto dela está apresentada na figura 2.1. Juntamente com a equipe do LNLS, eu pude participar da sua construção e, durante o doutorado, também participei do comissionamento e dos melhoramentos técnico-científicos dessa nova linha de luz. Por isso, neste capítulo descreveremos brevemente a instrumentação da linha de luz D02A-SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, SP, Brasil). Essa estação experimental é dedicada a experiências de

SAXS para estudos de matéria condensada, materiais poliméricos e macromoléculas biológicas. Apresentaremos também resultados que mostram as vantagens que esta nova linha oferece atualmente fazendo comparações com experiências realizadas na linha SAXS1. Todas as experiências de SAXS desta tese de doutorado foram realizadas nessa linha de luz, portanto, são exemplos dos dados que podem ser obtidos nessa estação experimental.

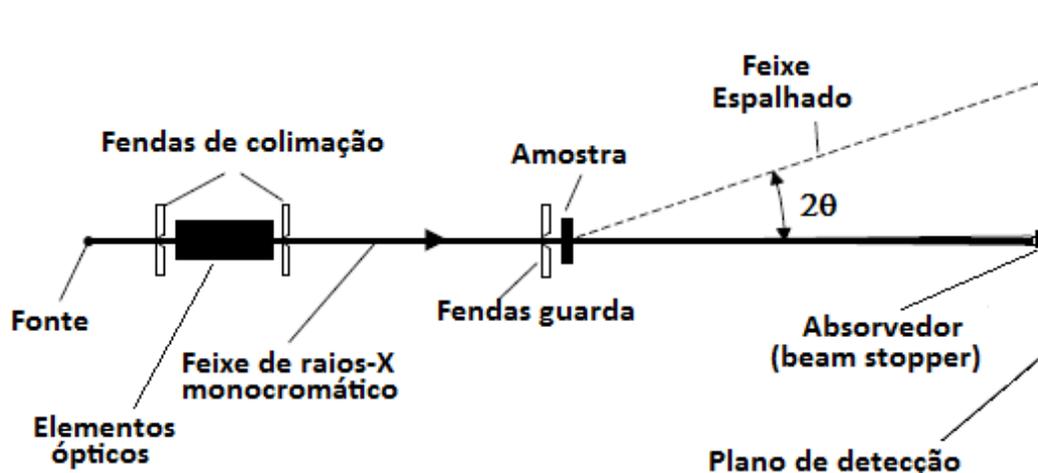


**Figura 2.1** – Fotografia mostrando a nova linha de luz (D02A-SAXS2) dedicada a realização de experiências de espalhamento de raios-X a baixos ângulos (SAXS) do Laboratório Nacional de Luz Síncrotron (LNL) [Foto tirada por Júlio César da Silva].

## **2.2. ESQUEMA GERAL DE UMA LINHA DE LUZ DEDICADA À REALIZAÇÃO DE EXPERIÊNCIAS DE SAXS**

Um esquema geral do arranjo instrumental necessário para a realização de experiências de SAXS é apresentado na figura 2.2. Basicamente necessitamos de um feixe monocromático colimado incidindo sobre a amostra e um sistema de detecção, mas na prática muitos mais detalhes devem ser considerados. Conjuntos de fendas são utilizados para a colimação antes de chegar à amostra passando por alguns elementos ópticos como o espelho e o monocromador. Além dessas fendas, há também as fendas parasitas, ou fendas-guarda, posicionadas antes da amostra. A

finalidade dessas fendas parasitas é absorver o espalhamento originado pela interação do feixe primário com as fendas de colimação e outras janelas no caminho do feixe até a amostra. Esse espalhamento extra é chamado de espalhamento parasita e não deve influenciar os dados espalhados pela amostra. Além disso, ao atravessar a amostra, somente parte do feixe incidente é espalhada. Como esse feixe não espalhado (feixe direto) é muito intenso em comparação ao espalhado, é essencial evitar que ele atinja o detector, pois pode atrapalhar o registro dos dados a muito baixo ângulo ou até mesmo danificar o detector. Para isso, utilizamos um absorvedor (*beam stopper*) posicionado imediatamente antes do detector com a função de bloquear o feixe direto, mas sem mascarar a região útil da curva de espalhamento a baixos ângulos. Maiores detalhes sobre instrumentação de SAXS podem ser encontrados em diversas referências [Koch *et al.*, 2003; Kellermann, 1997; Glatter & Kratky, 1982; Guinier & Fournet, 1995; Feigin & Svergun, 1987]. A instrumentação da linha D02A-SAXS2 do LNLS foi apresentada em congresso internacional [Torriani *et al.*, 2008] e um artigo sobre a linha está pronto para ser submetido para publicação.

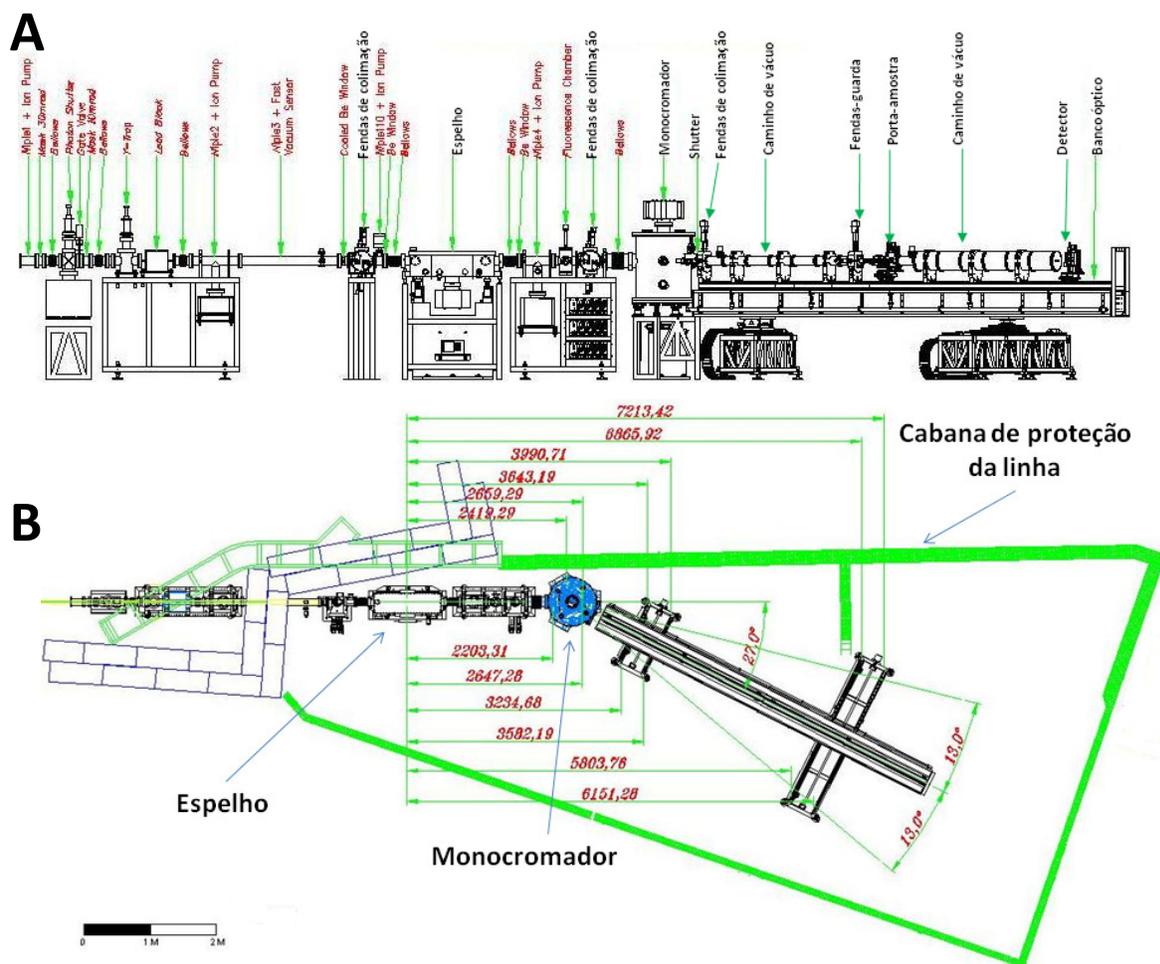


**Figura 2.2** – Esquema geral do arranjo da instrumentação necessária para a realização de experiências de SAXS em uma linha de luz síncrotron.

### 2.3. CARACTERÍSTICAS DA LINHA D02A-SAXS2 DO LNLS

A nova linha de luz dedicada à técnica de SAXS do LNLS está instalada no dipolo magnético D02, saída A, do anel de armazenamento do LNLS. Por isso, essa linha é identificada como D02A-SAXS2. Um esquema da configuração dessa linha está apresentado na figura 2.3. Seus elementos ópticos principais são o cristal monocromador e o espelho de raios-X. Além desses, há

também três módulos de fenda, dois monitores de intensidade (cintiladores), um antes e outro depois da amostra, filtros absorvedores, caminhos de vácuo, porta-amostra, o absorvedor do feixe direto e os detectores para a aquisição dos dados. A distância amostra-detector pode ser variada de 400 mm até 2500 mm e o intervalo de módulo do vetor de espalhamento é de  $\sim 0,004\text{\AA}^{-1}$  a  $\sim 2,000\text{\AA}^{-1}$ .



**Figura 2.3** – Desenho esquemático da linha D02A-SAXS2. (A) Vista lateral. (B) Vista superior. Os valores indicados no desenho significam as distâncias em milímetros (mm) [Desenho feito pelo grupo de projetos do LNLS].

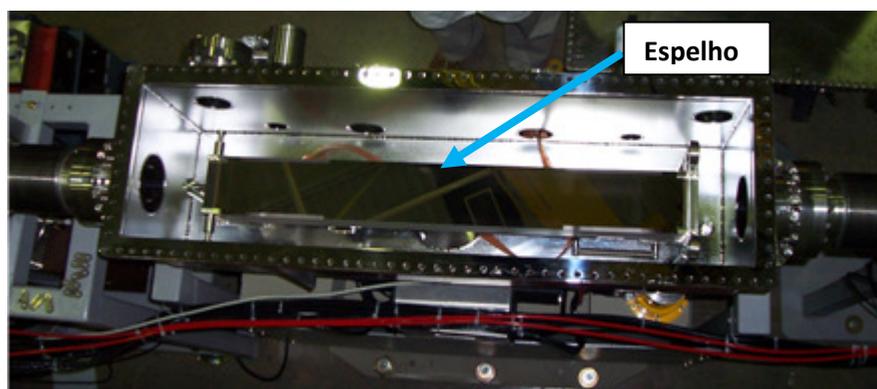
## 2.4. ELEMENTOS ÓPTICOS DA LINHA D02A-SAXS2 DO LNLS

### 2.4.1. ESPELHO DE RAIOS-X DA LINHA DE LUZ

O espelho de raios-X instalado na linha tem duas funções principais: eliminação de harmônicos (ordens de frequências mais altas que a fundamental) do feixe e focalização vertical.

O espelho utilizado na linha D02A-SAXS do LNLS é feito de sílica (quartzo) com baixo coeficiente de expansão térmica (*Ultra Low Expansion – ULE*) e recoberto com ródio (elemento altamente refletor) [Neuenschwander & Tavares, 2001]. Ele é colocado dentro uma câmara refrigerada que é mantida em vácuo (pressão base de operação da ordem de  $1 \times 10^{-8}$  mbar, usando bombas iônicas). Uma fotografia do espelho instalado na linha dentro dessa câmara está apresentada na figura 2.4. Dentro dessa câmara estão instalados motores para o posicionamento e alinhamento do feixe, tudo controlado por computadores. Esses motores também podem curvar levemente o espelho de tal modo a fazer a focalização na vertical do feixe. Essa câmara com o espelho está posicionada a uma distância de 6 m do dipolo magnético do anel de armazenamento (figura 2.3).

A relação entre o ângulo de incidência rasante crítico ( $\theta_c$ ) e o comprimento de onda para quartzo fundido é [Hendrix *et al.*, 1979]:  $\theta_c(\text{mrad}) = 26,3\lambda(\text{nm})$ , tal que  $\theta_c$  é tipicamente da ordem de 3 a 4 mrad. Note que somente comprimentos de onda maiores que aqueles dados por essa relação podem ser refletidos. Portanto, dessa maneira o espelho atua como um filtro de alta frequência e essa característica é particularmente útil para eliminar altos harmônicos do feixe de raios-X. Para o espelho utilizado na linha D02A-SAXS2, cálculos teóricos indicam que a refletividade é de 90% para um ângulo de incidência de 4,2 mrad na faixa de energias que vai de 6 a 12 keV.



**Figura 2.4** – Fotografia do espelho de raios-X instalado na linha de luz D02A-SAXS2 dentro da câmara de vácuo. O espelho é feito de quartzo e recoberto com ródio, que é um elemento altamente refletor [Foto tirada por Tomas S. Plivelic].

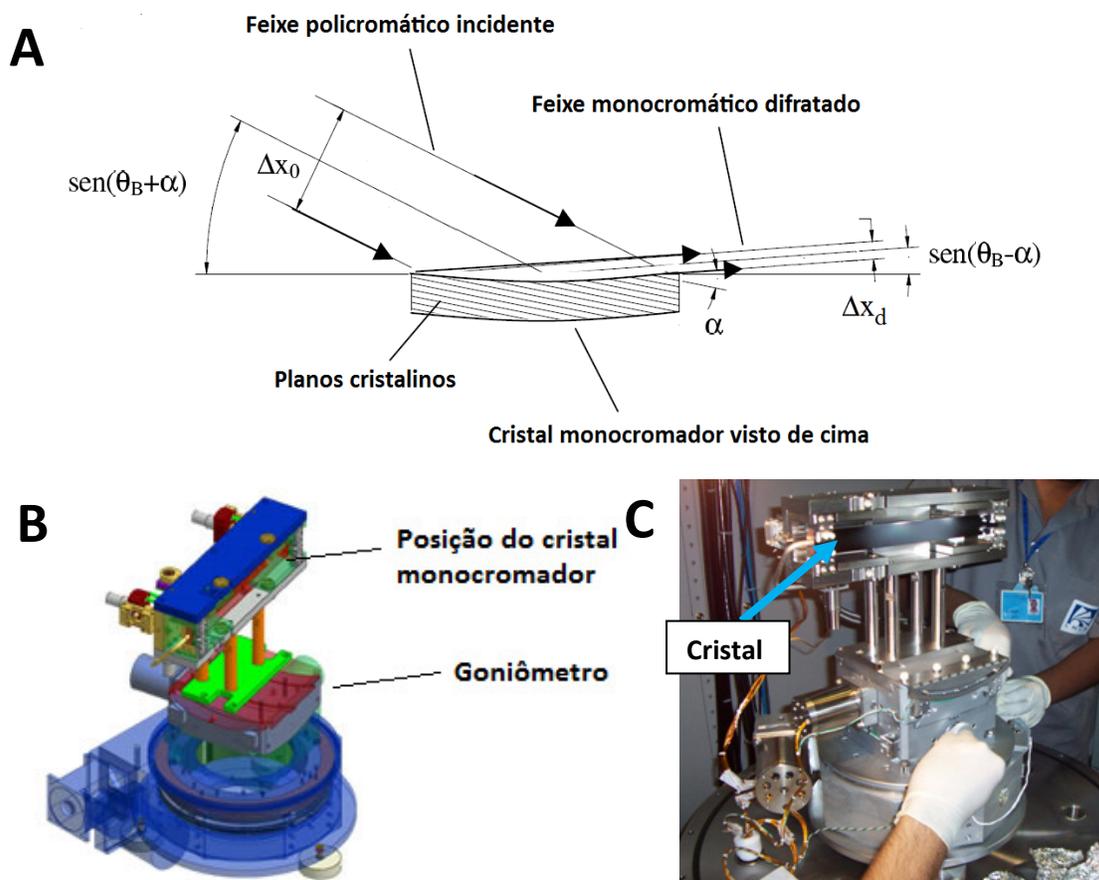
#### **2.4.2. MONOCROMADOR DE RAIOS-X DA LINHA DE LUZ D02A-SAXS2**

O monocromador é formado por um monocristal de silício com formato retangular que pode ser curvado mecanicamente na forma cilíndrica, permitindo ao mesmo tempo focalizar e selecionar

o comprimento de onda do feixe a ser utilizado na experiência. A monocromatização é feita utilizando os planos (111) do silício, de acordo com lei de Bragg:  $\lambda = 2d \text{sen}\theta_B$ , onde  $d$  é a distância entre os planos cristalinos,  $\lambda$  é o comprimento de onda da radiação difratada e  $\theta_B$  é o ângulo de Bragg [Cullity, 1978], ou seja, o ângulo entre o feixe incidente (e por consequência o difratado também) e os planos cristalinos do monocromador. O cristal do monocromador foi cortado de tal forma que esses planos cristalinos façam um ângulo de  $8^\circ$  com a superfície do cristal como esquematizado na figura 2.5. Esse tipo de corte do cristal foi proposto por Fankuchen [Fankuchen, 1937] com o intuito de concentrar a potência do feixe difratado de maneira a diminuir a seção transversal desse feixe (Figura 2.5(A)). Se o ângulo entre a superfície do cristal e seus planos cristalinos é  $\alpha$  e o ângulo de Bragg é  $\theta_B$ , então os ângulos feitos pelo feixe incidente e pelo difratado com a superfícies são, respectivamente,  $\theta_B + \alpha$  e  $\theta_B - \alpha$ . Dessa forma, a razão entre a largura da seção transversal do feixe incidente ( $\Delta X_0$ ) e do feixe difratado ( $\Delta X_d$ ) será  $\Delta X_0/\Delta X_d = (\text{sen}(\theta_B + \alpha)/\text{sen}(\theta_B - \alpha))$  [Guinier & Fournet, 1955; Glatter & Kratky, 1982].

Esse monocromador foi desenhado para cobrir um intervalo de energias que vão de 6 keV a 12 keV (comprimentos de onda no intervalo  $1\text{Å} < \lambda < 2\text{Å}$ ), sendo que ele está otimizado para o maior fluxo em torno da energia de 8 keV ( $\lambda = 1,488\text{Å}$ ). A sua resolução de energia é de  $5 \times 10^3$  em 8 keV e o fluxo de fótons, nessa mesma energia, é da ordem de  $10^{10}$  fótons/segundo para um foco de com FWHM de 1 mm na vertical por 3 mm na horizontal. Esse monocromador é posicionado dentro de uma câmara de vácuo (da ordem de  $5 \times 10^{-5}$  Pa, usando bombas iônicas) na linha de luz, logo após o espelho (como mostra a Figura 2.3), e está isolado da câmara do espelho e do restante da linha por janelas de berílio. Dentro dessa câmara de vácuo, há um sistema que curva o cristal monocromador em cima de três círculos goniométricos (Figura 2.5). As características detalhadas desse monocromador e desse sistema podem ser encontradas em Neuenschwander & Tolentino [Neuenschwander & Tolentino, 2004] e em Torriani *et al.* [Torriani *et al.*, 2008].

Para dar uma idéia das dimensões do feixe na posição focal (no detector) e na posição da amostras, vamos dar o exemplo de valores para uma condição típica de medida na linha de luz D02-SAXS2. Para esse exemplo, a linha está configurada com uma distância amostra-detector de aproximadamente 1 m, funcionando com um feixe monocromático com energia de 8 keV ( $\lambda = 1,488\text{Å}$ ). Nessas condições, o feixe que chega à amostra tem dimensões típicas da seção transversal de 4,0 mm na horizontal e 0,6 m na vertical (FWHM) e o feixe na posição focal tem dimensões típicas de 0,9 mm na horizontal por 0,5 m na vertical (FWHM).

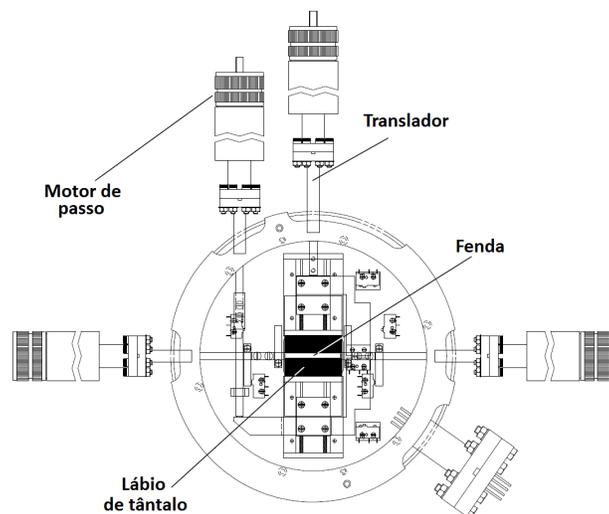


**Figura 2.5** – Monocromador de cristal curvado da linha de luz D02A-SAXS2. (A) Esquema mostrando a configuração utilizada para reduzir o tamanho horizontal do feixe difratado.  $\theta_B$  é o ângulo de Bragg;  $\alpha$  é o ângulo entre os planos cristalinos e a superfície do cristal;  $\Delta x$  é o tamanho da seção reta do feixe, onde os índices  $0$  e  $d$  se referem aos feixes incidindo no cristal e difratado pelo cristal, respectivamente. (B) Desenho do sistema que curva o cristal monocromador em cima de três círculos goniométricos [Desenho feito pelo grupo de projetos do LNLS]. (C) Fotografia do monocromador sendo instalado na linha de luz [Foto tirada por Tomas S. Plivelic].

### 2.4.3. CONJUNTOS DE FENDAS

Na linha de luz existem dois tipos de fendas: aquelas para a colimação do feixe e outras para a eliminação do espalhamento parasita (fendas-guarda). Fendas de colimação estão posicionadas em três lugares diferente da linha: um conjunto antes do espelho, outro antes do monocromador e mais outro conjunto depois do monocromador (Figura 2.3). Já o conjunto de fendas de guarda (parasitas) está posicionado antes do porta-amostra (Figura 2.3). Cada conjunto de fenda é constituído por quatro lábios como representado na figura 2.6. Os lábios superiores e

inferiores são responsáveis pela definição do tamanho vertical do feixe e o par de lábios laterais define seu tamanho horizontal. O movimento dos lábios é controlado por transladores lineares acionados por motores de passo.



**Figura 2.6** – Desenho do módulo de fendas com os motores de passo e transladores visto na direção perpendicular a de propagação do feixe [Desenho feito pelo grupo de projetos do LNLS].

#### **2.4.4. MONITORES DE INTENSIDADE, FILTROS E ABSORVEDOR DO FEIXE DIRETO**

Os monitores de intensidade são detectores de cintilação (cintiladores) posicionados antes e depois da posição do porta-amostra para monitorar o feixe incidente e o feixe transmitido através da amostra, respectivamente. A intensidade do feixe incidente, isto é, os fótons chegando à amostra são monitorados pelo cintilador colocado antes da amostra e que mede a intensidade espalhada por uma janela de um material semitransparente (kapton). Esse material permite a passagem da maior parte do feixe, mas reflete uma pequena fração que pode ser detectada pelo cintilador e que é proporcional à intensidade do feixe incidente. A intensidade do feixe transmitido é monitorada pelo cintilador que mede o espalhamento quando ele incide no absorvedor do feixe direto (*beam-stopper*). As contagens acumuladas nesses monitores durante a medida permitem determinar a atenuação do feixe pela amostra e normalizar as curvas de SAXS conforme descrito na seção 1.6 do capítulo 1.

O absorvedor (*beam-stopper*), que está posicionado na frente do detector na posição onde o feixe iria atingir a superfície ativa, é constituído de uma lâmina de aço inox com aproximadamente  $0,7\text{ mm}$  de espessura que atenua o feixe direto. Esse absorvedor não atenua totalmente o feixe,

deixando passar uma porção da radiação incidente somente para marcar a posição do ângulo zero (origem do eixo de valor de  $q$ ) do espalhamento. Essa origem também pode ser encontrada usando uma amostra cristalina em pó (Behenato de prata, por exemplo), que apresente anéis de difração isotrópicos a baixos ângulos, encontrando o centro desses anéis. As reflexões da amostra padrão também permitem medir com exatidão a distância amostra-detector, cujo conhecimento é necessário para a correta calibração do módulo do vetor de espalhamento (vetor  $\vec{q}$ ) da intensidade espalhada pela amostra.

Além desses dispositivos, na linha existem alguns filtros constituídos de materiais com bordas de absorção dentro de um intervalo de comprimentos de onda que vão de 1,35 a 1,80 Å (Cu, Ni, Co e Fe). Esses filtros são utilizados para alinhar o monocromador de tal forma a oferecer o feixe com a energia desejada. A calibração de energia é feita medindo a intensidade do feixe transmitido através do filtro desejado em função da posição angular  $\theta$  do cristal monocromador em unidades relativas. Nessa operação o banco óptico (Figura 2.3) deve efetuar uma varredura angular para registrar valores de intensidade do feixe difratado em diferentes valores de energia maiores e menores que a borda de absorção. A energia da borda de absorção do elemento escolhido é determinada selecionando a posição angular onde a derivada dessa curva tem seu valor máximo (ponto de inflexão da borda de absorção). Além dessa função, os filtros também podem ser usados para atenuar o feixe incidente sobre a amostra quando ela produz espalhamento muito intenso (geralmente para alguns materiais sólidos, mas nunca para proteínas) para evitar saturação do detector. Os filtros permitem também observar detalhes menos intensos acumulando durante um maior tempo de exposição na região de mais alto ângulo.

Todos os motores necessários para o posicionamento dos dispositivos e os cilindros pneumáticos que controlam a inserção dos filtros são controlados remotamente a partir de um painel de controle ou por computador. Monitores de posição acoplados aos motores de passo permitem controlar os dispositivos de forma precisa e reproduzir configurações anteriores.

## **2.5. CARACTERÍSTICAS DOS DETECTORES DISPONÍVEIS NA LINHA PARA EXPERIÊNCIAS DE SAXS**

Os detectores utilizados em medidas de SAXS ficam localizados na extremidade final da linha, na posição correspondente ao ponto de focalização do feixe monocromático. Eles podem ser movidos ao longo do banco óptico para se acomodarem a diferentes valores de distância amostra-detector. Entres os detectores disponíveis na linha de SAXS do LNLS estão os detectores lineares

(unidimensionais) a gás sensíveis a posição e os detectores de área (bidimensionais), que utilizam matrizes CCD ou com a nova tecnologia de pixels híbridos de silício. O surgimento dos detectores de área melhorou muito a relação sinal-ruído dos dados, o que é interessante para amostras que tipicamente espalham muito pouco os raios-X. Mais detalhes são dados abaixo.

### **2.5.1.DETECTOR LINEAR A GÁS SENSÍVEL A POSIÇÃO**

Os detectores a gás sensíveis a posição contam com a ionização de átomos de um gás para produzir um sinal eletrônico e codificar as coordenadas dos fótons [Gabriel, 1977]. Uma mistura de gases nobres é usada para minimizar as excitações de baixo nível molecular. Na linha de luz se utilizam estes para a montagem de experimentos especiais de tipo SAXS-WAXS [Torriani *et al.*, 2008]. Esses detectores têm janelas de berílio (Be) com comprimento efetivo de  $98\text{ mm}$  ao longo da direção de detecção e largura de  $8\text{ mm}$ . Eles têm resolução espacial entre  $100$  e  $350\ \mu\text{m}$  e podem contar até cerca de 200 fótons por segundo por canal (cada canal é uma região de  $100\ \mu\text{m}$  ao longo da direção sensível) sem perda de linearidade na eficiência de detecção. Embora haja uma variação na resposta do detector de aproximadamente 8% na eficiência relativa ao longo dos canais de detecção, esses detectores são muito rápidos, o que é bom para coletas cinéticas como acompanhamento em tempo real de mudanças na amostra. Obviamente, a correção pela resposta do detector, conforme apresentada na seção 1.6 do capítulo 1, é essencial e necessária para o correto tratamento dos dados antes das análises. Além disso, esses detectores apresentam uma limitação de alcance dinâmico (*dynamic range*), o que pode implicar em dificuldade nas coletas de dados de amostras que espalham pouco, como as proteínas em solução. Quanto mais alto o alcance dinâmico, melhor é a relação sinal-ruído. Mais discussões sobre esses detectores podem ser encontradas em algumas outras referências [Gabriel, 1977; Gursky, 1975; Russel, 1991].

### **2.5.2.DETECTOR CCD**

Os detectores CCD são tipicamente detectores de área, que utilizam uma matriz de CCDs para coletar a intensidade de espalhamento experimental. Esses detectores são extremamente versáteis e estão disponíveis com até milhares de *pixels* (quantidade essa que cresce cada vez mais). Por outro lado, o tamanho desses *pixels* é extremamente reduzido, com cerca de  $10$  a  $50\ \mu\text{m}$ , em geral. O tempo de leitura também é menor que 1 segundo, o que é importante para experiências em tempo real. Na linha de luz D02A-SAXS2, há um detector bidimensional da marca *Rayonix*<sup>TM</sup> (antiga *MARCCD*<sup>TM</sup>), modelo *SX-165* ([http://www.mar-usa.com/products/sx\\_165.htm](http://www.mar-usa.com/products/sx_165.htm)). Esse

equipamento possui uma área de detecção efetiva de  $165\text{ mm}$ , um alto range dinâmico e resoluções de espaciais  $79$ ,  $158$  e  $316\ \mu\text{m}$  em uma matriz com  $2048$ ,  $1024$  e  $516\ \text{pixels}$ , respectivamente. Além disso, esse detector possui um sistema de refrigeração que mantém sua temperatura de funcionamento em  $-80^\circ\text{C}$ , o que reduz bastante seu ruído de fundo. Na prática, essa temperatura é tão baixa que, se mantida assim, o sinal devido à atividade térmica do semicondutor do CCD (sinal esse chamado de corrente escura ou, em inglês, *dark current*) pode ser desprezado. Para termos uma idéia, o efeito desse sinal é bem menor (menos de  $0,5\%$ ) que os erros estatísticos da medida. Outra função interessante e útil oferecida por esse detector é o sistema de coletas no modo *frame-shift*, que permite coletas de dados em tempos da ordem de milissegundos em estudos resolvidos no tempo. Outras especificações desse detector MARCCD<sup>TM</sup> podem ser encontradas em sua ficha técnica que está em: [http://www.mar-usa.com/support/downloads/sx\\_series.pdf](http://www.mar-usa.com/support/downloads/sx_series.pdf).

A principal dificuldade para interpretar dados coletados com detectores CCD é o sinal que já está contido no chip CCD antes da medida ser realizada. Devido ao fato de que ruídos sempre estarão presentes na medida, e ruído é um processo estatístico, um valor fixo de sinal é adicionado aos CCDs antes da medida para garantir que a voltagem presente no conversor A/D (*Analog-to-digital*) de seus circuitos eletrônicos nunca seja negativa, já que ele não trabalha com valores negativos. Para o detector que utilizamos na linha, esse valor fixo é  $10$  (dez). Logo, a primeira coisa a ser feita após coletar os dados de espalhamento com o detector MARCCD é a subtração dessa constante dos dados de espalhamento para depois tratar cada curva experimental de espalhamento como apresentado na seção 1.6 do capítulo 1.

### **2.5.3.DETECTOR DE ÁREA COM A TECNOLOGIA DE PIXELS HÍBRIDOS DE SILÍCIO**

Os detectores de área (há lineares também) que utilizam a tecnologia de pixels híbridos de silícios operam no modo de “contagem de fóton único”. A principal diferença para os outros detectores existentes é que os fótons de raios-X que incidem no detector são diretamente transformados em carga elétrica, usando sensores de silício, e processados por chips de leitura do tipo CMOS (*complementary metal-oxide-semiconductor*, isto é, semicondutor metal-óxido complementar). São fabricados pela DECTRIS<sup>TM</sup> (<http://www.dectris.com/sites/dectris.html>) e há diversos modelos e tamanhos de detectores. Na linha de luz D02A-SAXS2, há disponível o detector PILATUS 300K (<http://www.dectris.com/sites/pilatus300k.html>), com área de  $83,8 \times 106,5\ \text{mm}^2$  e resolução de  $172\ \mu\text{m}$ .

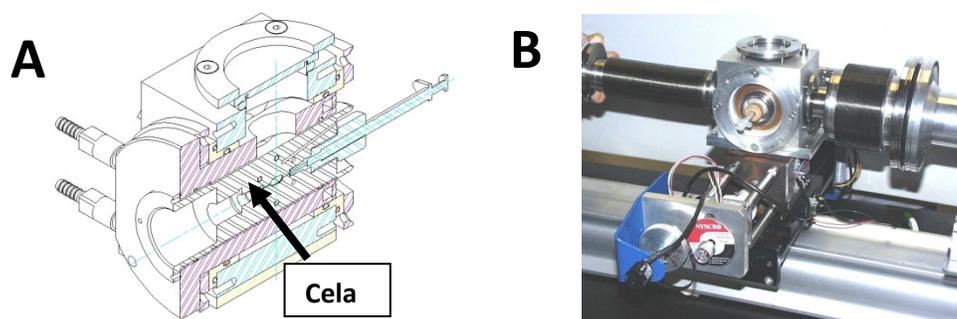
Essa tecnologia nova de detecção faz com que esse tipo de detector não apresente o efeito de corrente escura (o sinal devido à atividade térmica do semicondutor do CCD) e permite um alto range dinâmico (20 bits). Além disso, também permite coletas rápidas, com tempo de leitura menor que 3 milissegundos, com uma alta taxa de 200 imagens por segundo. A eficiência quântica do sensor de silício de 0,32 mm de espessura é ótima para experiências no intervalo de energia que vai de 3 a 15 keV, embora possam ser usados para energias de até 40 keV, bastando alguns cuidados. A taxa de leitura de contagem é acima de  $2 \times 10^6$  pixels/segundo para lidar com os altos fluxos das fontes de luz síncrotron. A desvantagem desses detectores é que eles são montados em módulos de tal forma que há um intervalo entre esses módulos onde não há detecção. Todavia, nos dados esses intervalos podem ser mascarados através de rotinas computacionais como, por exemplo, aquela que fazem parte da suíte computacional desenvolvida por Hammersley [Hammersley, 2004]. Esses detectores são o que há de mais moderno atualmente para coletar dados de SAXS.

## **2.6. PORTA-AMOSTRAS PARA EXPERIÊNCIAS DE SAXS COM SOLUÇÕES DE MACROMOLÉCULAS BIOLÓGICAS**

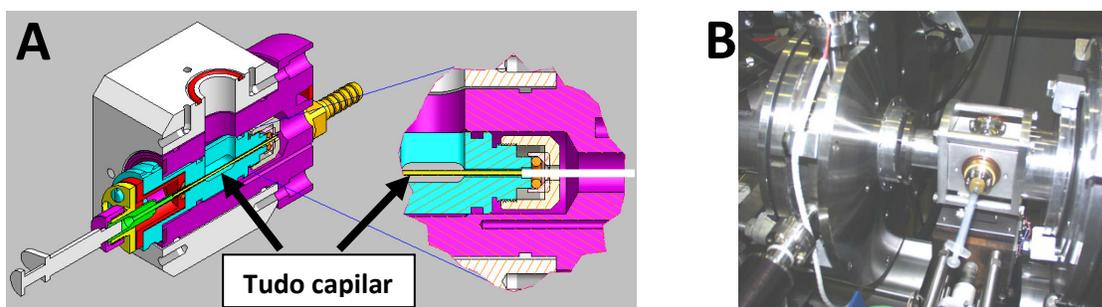
Para o estudo de macromoléculas biológicas em solução é necessário que haja porta-amostras para líquidos que possam ser usados na linha de SAXS. Como realizamos experiências onde o feixe de raios-X atravessa a amostra, é necessário também que o porta-amostra permita a transmissão dos raios-X com o mínimo de influência na intensidade do feixe além da absorção das amostras. Outro problema é que, geralmente, há um volume pequeno de proteínas em solução devido às dificuldades de obtenção de grandes quantidades monodispersas desse tipo de amostra. Pensando nisso, na linha de luz D02A-SAXS2 do LNLS há dois porta-amostras especialmente desenvolvidos para esse tipo de experiências e construídos pela equipe de SAXS [Cavalcanti *et al.*, 2004]. No primeiro, a amostra preenche uma cela de 1 mm de espessura e vedada por duas janelas de mica, uma de cada lado (ver figura 2.7). O volume a ser preenchido pela amostra é de aproximadamente 200  $\mu\text{L}$ . No segundo porta-amostras, o líquido é injetado em um tubo capilar cilíndrico com 1 mm de diâmetro (ver figura 2.8). A principal vantagem desse segundo porta-amostras é a quantidade menor de amostra necessária, sendo que basta um volume de aproximadamente 100  $\mu\text{L}$  para preencher o capilar. Ambos os porta-amostras possuem sistema de circulação de água para o controle de temperatura, que é feito através de uma conexão com um banho térmico, e estão conectados ao caminho de vácuo da linha de luz.

Comparando os dois porta-amostras, o primeiro é mais prático que o segundo, pois é mais robusto e simples de montar. O tubo capilar do segundo é extremamente frágil e tem que ser vedado

com cuidado para não ter vazamento, além de que o feixe tem que atravessar exatamente o centro do tubo (perpendicular ao eixo transversal principal) devido à sua geometria cilíndrica. O porta-amostra com janela de mica é bastante versátil, pois pode ser usado para diversos tipos de amostras, como por exemplo, géis, cristal líquido e dispersões de solventes orgânicos, além de macromoléculas em solução. Já o porta-amostra com capilar permite o registro da curva de intensidade até ângulos mais altos e o capilar pode ser eficientemente trocado após cada rodada de experiências, se necessário. A razão sinal-ruído e reprodutibilidade dos dados têm sido testadas através dos anos com diversas amostras e têm sido bastante satisfatórios.



**Figura 2.7** – Porta-amostra para amostras líquidas com um cela de  $1\text{ mm}$  de espessura e vedada por duas janelas de mica. (A) Desenho do projeto do porta-amostra. (B) Fotografia do porta-amostra montado na linha de luz.



**Figura 2.8** – Porta-amostra para amostras líquidas usando tubo capilar cilíndrico com  $1\text{ mm}$  de diâmetro na seção transversal. (A) Desenho do projeto do porta-amostra. (B) Fotografia do porta-amostra montado na linha de luz.

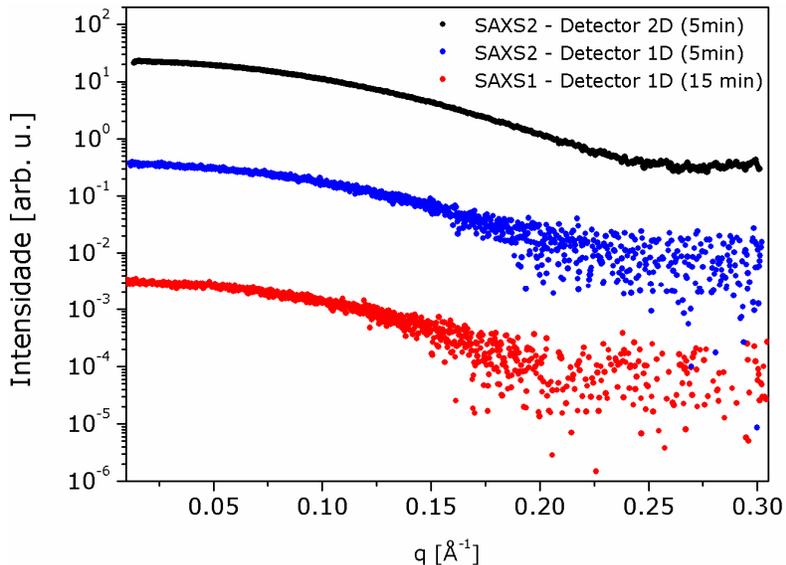
## **2.7. ROTINAS COMPUTACIONAIS PARA O TRATAMENTO DE DADOS**

Com a diminuição dos tempos de coleta de dados de SAXS, mais dados podem ser obtidos e maior é o volume de informação que deve ser processada. Isso torna necessária a aplicação reiterada das rotinas de tratamento das curvas experimentais de SAXS conforme descritos na seção 1.6 do capítulo 1. Para facilitar essa tarefa tediosa durante as experiências na linha D02A-SAXS2, eu desenvolvi durante meu doutorado uma rotina computacional em *linguagem C* para automatizar essa tarefa e dei o nome de PROTAS (Programa de Tratamento de dados de SAXS). Além disso, essa rotina permite também realizar o processo de transformação de uma imagem bidimensional para uma curva unidimensional, além de já calibrar o eixo de valores  $q$  de cada curva de espalhamento. Para essas últimas duas tarefas, nós inserimos em nosso algoritmo algumas das rotinas que fazem parte da suíte computacional FIT2D desenvolvida por Dr. Andy Hammersley, do European Synchrotron Radiation Facility (ESRF) [Hammersley, 2004].

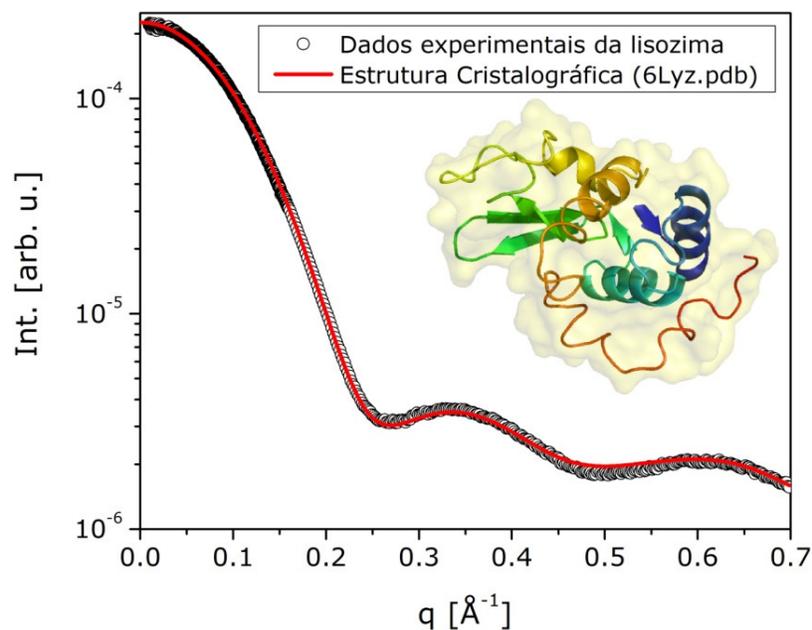
## **2.8. ALGUNS EXEMPLOS DE DADOS OBTIDOS NA LINHA DE LUZ D02A-SAXS2 E COMPARAÇÃO COM A ANTIGA LINHA DE SAXS DO LNLS**

A seguir mostraremos alguns exemplos de medidas experimentais realizadas na nova linha de luz D02A-SAXS2. É claro que os dados apresentados nos capítulos desta tese de doutorado também são grandes exemplos da potencialidade dessa linha. Nos exemplos que serão apresentados a seguir, nós poderemos notar a visível melhora de qualidade nos dados de SAXS que foi obtida com a construção da nova linha quando comparados com os dados da antiga linha, D01A-SAXS1. Aqui, nós realizamos experiências com a lisozima, uma proteína muito conhecida que já tem estrutura resolvida. Essa proteína é bastante usada como padrão de estimativa de massa molecular devido a sua estabilidade. Na figura 2.9, nós apresentamos três curvas de espalhamento experimentais obtidas das amostras de lisozima em uma concentração de  $5 \text{ mg/mL}$ . A curva em vermelho foi registrada na linha D01A-SAXS1 (antiga linha) durante um período de 15 minutos usando um detector linear a gás sensível a posição (ver seção 2.5.1 deste capítulo). Já a curva em verde foi registrada também com esse mesmo detector, porém durante 5 minutos e na linha D02A-SAXS2 (nova linha). Comparando as duas curvas, já podemos observar uma melhoria nos dados para valores de  $q > 0,20 \text{ \AA}^{-1}$ . Note que na curva obtida em menos tempo na SAXS2 que na

SAXS1, nós podemos observar uma melhor estatística dos dados para esses valores  $q$ . Isso mostra o ganho de fluxo de fótons que temos na nova linha. Continuando ainda mais com as experiências, obtivemos também mais uma curva de espalhamento da lisozima, também durante 5 minutos na SAXS2, porém com o detector bidimensional MARCCD<sup>TM</sup> (ver seção 2.5.2). O resultado é a curva preta da figura 2.9. Vemos claramente uma substancial melhora que nem se compara com as outras duas curvas. Para concluir, nós obtivemos uma última curva de espalhamento da lisozima na SAXS2 com o detector bidimensional numa região de  $q$  bem maior e comparamos com o espalhamento calculado da estrutura cristalográfica da lisozima. O excelente resultado que obtivemos está mostrado na figura 2.10. O espalhamento calculado a partir da estrutura de alta resolução da lisozima (código 6lyz no Protein Data Bank) foi obtido usando a metodologia descrita em Svergun *et al.* [Svergun *et al.*, 1995].

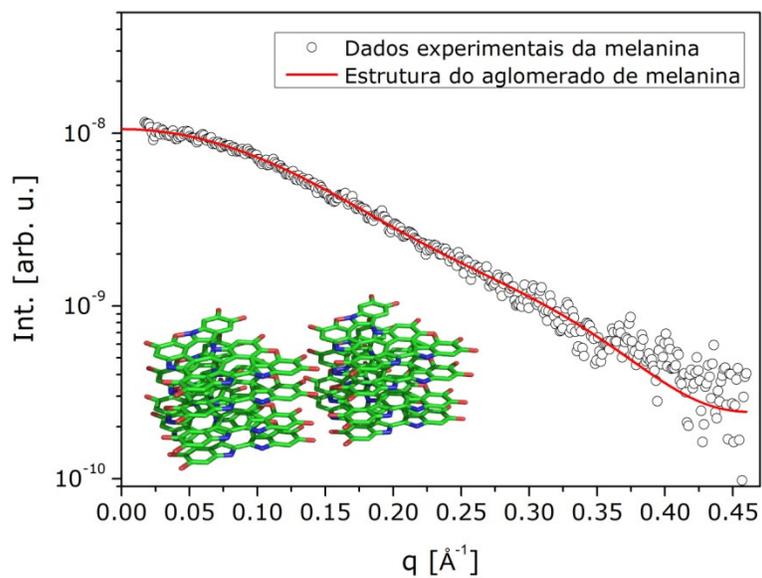


**Figura 2.9** – Curvas experimentais de espalhamento da proteína lisozima medidas nas linhas D01A-SAXS1 e D02A-SAXS2. A curva em vermelho foi registrada na linha D01A-SAXS1 com um detector unidimensional a gás sensível a posição durante 15 minutos. A curva em azul foi registrada na linha D02A-SAXS2 também com o detector unidimensional por 5 minutos e a curva em preto foi medida com um detector bidimensional durante 5 minutos. As curvas foram deslocadas por um fator constante para uma melhor visualização.



**Figura 2.10** – Comparação entre a curva experimental de espalhamento da lisozima medida na linha D02A-SAXS2 e a curva teórica calculada da estrutura cristalográfica da lisozima. No destaque da figura está mostrada a estrutura cristalográfica dessa proteína depositada no PDB com o código 6lyz.

Também realizamos uma experiência com melanina em solvente DMSO. O DMSO contém enxofre, que por sua vez absorve os raios-X no comprimento de onda típico da linha ( $\lambda = 1,488 \text{ \AA}$ ). Logo, graças ao maior fluxo da nova linha em relação à antiga (em torno de quatro vezes maior) e outros melhoramentos técnicos da linha, foi possível configurar a linha para um comprimento de onda de  $\lambda = 1,045 \text{ \AA}$ . Nesse comprimento de onda foi possível realizar experiências com melanina em DMSO e os resultados estão mostrados na figura 2.11. Com esses dados foi possível obter os primeiros resultados sobre o estado de aglomeração e empilhamento de moléculas de melanina em solução. Uma estrutura desse aglomerado foi calculada por dinâmica molecular. Daí, calculamos o espalhamento teórico dessa estrutura e comparamos com os dados experimentais. O resultado está mostrado na figura 2.11.



**Figura 2.11** – Comparação entre a curva experimental de espalhamento da melanina em DMSO medida na linha D02A-SAXS2 usando um comprimento de onda  $\lambda = 1,045 \text{ \AA}$  e a curva teórica calculada da estrutura teórica do aglomerado e empilhamento das moléculas de melanina. Essa estrutura está mostrada no destaque da figura.



# 3

## Proteínas Nativamente Desenoveladas

### 3.1. INTRODUÇÃO

O grande paradigma dos estudos estruturais de macromoléculas, principalmente em biologia molecular, é que o conhecimento da estrutura tridimensional é um pré-requisito para o entendimento de como essas moléculas funcionam. Essa determinação de estrutura tem sido feita principalmente através de técnicas como cristalografia de raios-X e ressonância magnética nuclear (RMN). Isso funciona bem para macromoléculas com estrutura terciária bem definida e rígida. Porém, grandes dificuldades aparecem no estudo de moléculas flexíveis. Macromoléculas possuindo flexibilidade inerente são difíceis de cristalizar. Além disso, as conformações de regiões desordenadas (*loops*) ou domínios móveis não podem ser capturados por técnicas cristalográficas e, também, são muito difíceis de ser univocamente determinadas por RMN. Entretanto, importantes funções de regulação de expressão gênica e no ciclo celular são freqüentemente executadas por proteínas intrinsecamente desestruturadas, que não possuem uma estrutura terciária fixa [Dunker *et al.*, 2002; Uversky, 2002a; Uversky, 2002b; Dyson & Wright, 2005]. Essas proteínas obviamente não podem ser cristalizadas e, embora RMN seja a única técnica de alta resolução aplicável, ela somente fornece informações locais sobre a estrutura [Uversky, 2002a; Shortle & Acherman, 2001; Louhivouri *et al.*, 2003; Mohana-Borges *et al.*, 2004; Jha *et al.*, 2005]. Portanto, há uma

necessidade clara de outros métodos de abordagem para uma caracterização quantitativa da estrutura dessas proteínas intrinsecamente desordenadas em solução. Uma técnica particularmente útil que aparece em destaque para o estudo de sistemas menos estruturados é a técnica de espalhamento de raios-X a baixo ângulo. Essa técnica vem sendo aplicada com sucesso a várias proteínas desestruturadas [Lanza *et al.*, 2009; Bressan *et al.*, 2008; Tidow *et al.*, 2007; Uversky & Fink, 2002; Uversky *et al.*, 1999].

As proteínas intrinsecamente desestruturadas são algumas vezes também denominadas como proteínas nativamente desenoveladas ou até mesmo proteínas parcialmente desenoveladas. Essa classe de proteínas difere de outras pelo fato de as estruturas dessas proteínas terem um grande volume hidratado, pouco ou quase nada de elementos de estrutura secundária, flexibilidade intramolecular acentuada e respostas não usuais às mudanças do meio [Uversky, 2002a; Uversky, 2002b]. De forma um pouco não intuitiva, elas são biologicamente ativas, em contraste com a noção clássica de que as proteínas necessitam de uma estrutura globular bem definida para serem ativas. Essa idéia pode ser melhor compreendida quando se pensa no fato de que a falta de uma estrutura globular rígida pode representar uma vantagem funcional para as proteínas nativamente desenoveladas, pois dessa forma essa flexibilidade permite que elas interajam com vários alvos diferentes o qual geralmente não ocorre com as proteínas enoveladas [Wright & Dyson, 1999; Dyson & Wright, 2002]. Embora a especificidade de interações proteína-proteína esteja relacionada com as pequenas partes estruturadas da molécula, as regiões desordenadas permitem que essas proteínas se adaptem a vários parceiros de ligação, aumentando a probabilidade de interações. Além disso, regiões desordenadas são sítios preferidos de regulação de proteínas por modificações pós-traducionais, principalmente fosforilação [Fuxreiter *et al.*, 2007]. O reconhecimento molecular, também realizado por proteínas desordenadas, pode ser usado para sinalização, dado que as regiões desordenadas podem ligar a múltiplos alvos com baixa afinidade [Dunker *et al.*, 2002]. De fato, mais e mais evidências vêm se acumulando na literatura de que essas proteínas formam uma grande fração (algumas predições dizem que são mais de 30%) de todas as proteínas eucarióticas [Dunker *et al.*, 2002; Tompa, 2002].

Do ponto de vista estrutural, as proteínas nativamente desenoveladas podem ainda ser classificadas em pelo menos três grupos conformacionais [Uversky, 2002a; Uversky, 2002b]: glóbulo pré-fundido (*pre-molten globule*), glóbulo fundido (*molten globule*), e completamente desestruturadas (*random coils*). O estado de glóbulo pré-fundido é caracterizado por uma estrutura terciária não rígida e uma considerável quantidade de estrutura secundária, porém menos pronunciada que aquela do estado de glóbulo fundido. Em geral, proteínas no estado de glóbulo pré-fundido contêm aproximadamente 50% de estrutura secundária em sua conformação, enquanto no

estado de glóbulo fundido as proteínas têm quase 100% de estrutura secundária [Uversky, 2002a]. A molécula de proteína no estado de glóbulo pré-fundido é consideravelmente menos compacta que aquela no estado de glóbulo fundido, mas ainda é mais compacta que aquela no estado sem estrutura (*random coils*, ou seja, totalmente desestruturada), estado este que é conhecido como estado desenovelado de uma proteína.

O mais interessante é que intuitivamente poderíamos pensar que uma proteína nativamente desenovelada não pudesse mais perder estrutura. Entretanto, é possível observar mudanças conformacionais nessas moléculas com a adição de desnaturantes (uréia, por exemplo) [Silva *et al.*, 2010a]. Isso mostra que a estrutura das proteínas intrinsecamente desestruturadas não é aleatória. Pequenas regiões formando núcleos hidrofóbicos e, em geral, algumas pequenas regiões com elementos de estrutura secundária (em maior parte, alfa-hélices) estão presentes na sua estrutura.

Para classificar essas proteínas com base na quantidade de desordem em suas estruturas, vários métodos bioquímicos são aplicados, como por exemplo, experiências de velocidade de sedimentação usando ultracentrifugação analítica, experiências de viscometria, espalhamento dinâmico de luz, algumas vezes até gel-filtração. Todos esses métodos fornecem parâmetros hidrodinâmicos, como por exemplo, raio hidrodinâmico, estimativas de massa molecular, volume e grau de monodispersidade, que auxiliam nessa classificação [Uversky, 2002a; Uversky, 2002b]. Além desse método, há também método de predição teórica de estrutura com base de aminoácidos e, também, a técnica de dicroísmo circular que fornece informações experimentais sobre a estrutura secundária (mesmo que pouca) da proteína. Das técnicas físicas disponíveis, a técnica experimental por excelência para estudar proteínas intrinsecamente desestruturadas é a técnica de SAXS. Aliás, há uma característica dessas moléculas que é necessária para uma análise mais apurada que é o grau de compactidade da molécula. Esse importante parâmetro reflete a presença ou ausência de domínios fortemente empacotados na molécula da proteína. Essa informação pode ser extraída exclusivamente das análises de dados de SAXS na chamada representação de Kratky da curva de intensidade ( $q^2 I(q)$  vs.  $q$ ), como discutido no capítulo 1.

Sendo assim, neste trabalho nós mostramos como a técnica de SAXS pode ser usada para essa caracterização de proteínas intrinsecamente desestruturadas, demonstrando todo seu potencial. Além disso, duas proteínas com características de proteínas intrinsecamente desestruturadas foram estudadas: a proteína humana identificada como FEZ1 (*Fasciculation and Elongation protein Zeta-1*) [Silva *et al.*, 2010a; Lanza *et al.*, 2009] e a proteína também humana denominada Ki-1/57 [Bressan *et al.*, 2008]. Estas proteínas foram expressas e fornecidas pelo grupo de biologia molecular do Dr. Jörg Kobarg do Laboratório Nacional de Biociências (LNBio) de Campinas, SP. Ambas as proteínas foram estruturalmente caracterizadas utilizando a técnica de SAXS, usando

diferentes abordagens matemáticas e computacionais, com o objetivo de se obter informações conformacionais dessas proteínas em solução e, até mesmo, modelos de baixa resolução que melhor as descrevam. Esses trabalhos renderam algumas publicações em revistas internacionais indexadas, sendo duas já publicadas [Bressan *et al.*, 2008; Lanza *et al.*, 2009] e mais uma em vias de submissão [Silva *et al.*, 2010a].

## **3.2. PROTEÍNA DE ENFEIXAMENTO E PROLONGAMENTO ZETA-1 (FEZ1)**

### **3.2.1. DESCRIÇÃO DA PROTEÍNA FEZ1**

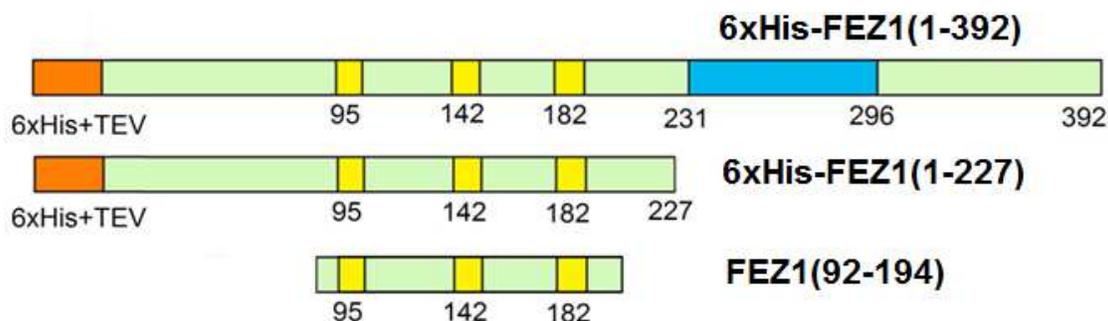
A proteína humana FEZ1 (*Fasciculation and Elongation protein Zeta 1*) é uma proteína, composta por 392 aminoácidos, com importantes funções centrossomais<sup>x</sup> dentro das células [Lanza *et al.*, 2008]. Essa proteína está relacionada com importantes doenças crônicas, entre as quais estão a leucemia [Lanza *et al.*, 2008; Assmann *et al.*, 2006] e esquizofrenia [Sakae *et al.*, 2008; Hodgkinson *et al.*, 2007; Koga *et al.*, 2007; Yamada *et al.*, 2004]. Estudos funcionais da FEZ1 sugerem o seu envolvimento em processos de transporte dentro da célula, o que tem claras implicações para seu papel na diferenciação neuronal e no crescimento de neurônios [Assmann *et al.*, 2006]. FEZ1 interage com uma grande quantidade de diferentes proteínas, que participam em diferentes funções celulares tais como a dinâmica do citoesqueleto celular, controle da transcrição, organização da cromatina, ciclo celular e transporte mitocondrial [Assmann *et al.*, 2006]. Essas interações sugerem o envolvimento da FEZ1 nesses processos. De fato, FEZ1 pode interagir com mais de 40 outras proteínas, o que sugere o fato de que ela seja uma proteína intrinsecamente desestruturada no seu estado nativo [Lanza *et al.*, 2008]. Em razão disso, foram realizadas predições teóricas de estrutura, experiências de dicroísmo circular, espectroscopia de fluorescência e proteólise limitada, cujos resultados indicam que FEZ1 pertença à classe de proteínas intrinsecamente desestruturadas (ou nativamente desoveladas) [Lanza *et al.*, 2008]. Outra coisa que é bem conhecida da literatura, e que reforça essa idéia, é que proteínas que possuem múltiplas funções, como é o caso da FEZ1, freqüentemente têm uma estrutura intrínseca desordenada [Singh & Dash, 2007; Dosztanyi *et al.*, 2006; Dunker *et al.*, 2005; Tompa, 2005].

---

<sup>x</sup> Centrossomal: relativo ao centrossoma. Centrossoma é uma região específica da célula, situada próxima ao núcleo, onde são organizados os microtúbulos, que por sua vez são estruturas protéicas que fazem parte do citoesqueleto nas células.

Como já foi falado, essa desordem estrutural e o baixo grau de compacidade dificultam a obtenção de estruturas de alta resolução dessas proteínas. Em vista disso, nós realizamos experiências de espalhamento de raios-X a baixo ângulo (SAXS) para obter parâmetros dimensionais e modelos conformacionais tridimensionais de baixa resolução para a proteína FEZ1, além de caracterizar seu grau de compacidade. As análises de SAXS foram realizadas com amostras da FEZ1 inteira, identificada como 6xHis-FEZ1(1-392), um fragmento do N-terminal, identificado como 6xHis-FEZ1(1-227), e um outro fragmento menor do N-terminal, identificado como FEZ1(92-194). Os valores entre parênteses representam a região de aminoácidos da construção e o prefixo 6xHis indica que a proteína estava fusionada a um N-terminal de 6xHis-tag (cauda de histidina), que é necessário para a expressão dessas proteínas. Por questões práticas, estaremos omitindo o sufixo 6xHis ao nos referirmos as amostras.

Os resultados mostraram uma dimerização da FEZ1, ao que tudo indica na região N-terminal, e uma forma estendida e flexível, compatível com uma proteína nativamente desenovelada. Maiores detalhes das experiências bioquímicas e predições teóricas deste trabalho estão descritos nos trabalhos de *Lanza et al.* [Lanza et al., 2009] e em *Silva et al.* [Silva et al., 2010a]. Nas seções a seguir, será dada atenção às experiências de SAXS, que foram da minha responsabilidade, e seus principais resultados.



**Figura 3.1** – Desenho esquemático do comprimento das seqüências de aminoácidos das amostras estudadas: FEZ1(1-392), FEZ1(1-227) e FEZ1(92-194). As regiões em laranja indicam a cauda de histidina, as regiões em amarelo indicam aminoácidos com alta probabilidade de formação de alfa-hélices e a região em azul indica a porção que foi predita com alta probabilidade de formação de “coiled-coils” em interações com outras proteínas.

## 3.2.2. MATERIAIS E MÉTODOS

### 3.2.2.1. PREPARAÇÃO DAS AMOSTRAS

As amostras da proteína FEZ1 inteira, FEZ1(1-392), e a região N-terminal, FEZ1(1-227), foram preparados conforme descrito em Lanza *et al.* [Lanza *et al.*, 2009]. O fragmento FEZ1(92-194) também foi preparado de maneira similar conforme descrito em Furlan *et al.* [Furlan *et al.*, 2010]. Brevemente, essas amostras de proteínas foram expressas em bactérias e purificadas por cromatografia de afinidade, seguindo protocolos convencionais de purificação de proteínas com cauda de histidina (His-tag) em colunas *HiTrap<sup>TM</sup> chelating* (GE Healthcare®, [www.gelifesciences.com](http://www.gelifesciences.com)). Essa cauda de histidina, fusionada na região N-terminal da proteína, possui afinidade com níquel. Logo, ao passar a solução com as bactérias (após expressarem a proteína e terem suas membranas rompidas) por uma resina porosa, equilibrada com níquel, somente as moléculas de proteína que possuem essa cauda irão se prender a essa resina. Posteriormente, as proteínas são liberadas dessa resina por eluição com um composto orgânico chamado imidazol, que compete com a histidina pela ligação com o níquel através de coordenação química. A estabilidade e pureza das amostras foram testadas e verificadas por gel de eletroforese (gel SDS) [Lanza *et al.*, 2009]. A figura 3.1 mostra um desenho esquemático do comprimento das seqüências das três amostras.

### 3.2.2.2. DESCRIÇÃO DAS EXPERIÊNCIAS DE SAXS COM AS AMOSTRAS DE FEZ1 EM CONDIÇÃO NATIVA

As três amostras da proteína FEZ1, aqui identificadas como FEZ1(1-392), FEZ1(1-227) e FEZ1 (92-194), foram submetidas à análise de DLS (espalhamento dinâmico de luz) para testar a monodispersidade das mesmas antes das experiências de SAXS. Feito isso, as amostras foram levadas para a linha de luz D02A-SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, SP, Brasil), onde as experiências de SAXS foram feitas. Foi utilizado um feixe de raios-X monocromático (comprimento de onda de  $\lambda = 1,488\text{\AA}$ ) e o espalhamento de raios-X foi registrado usando um detector bidimensional MARCCD sensível a posição. Duas distâncias amostra-detector foram utilizadas para detectar a região de espalhamento de interesse:  $606,2\text{ mm}$  e  $1812,6\text{ mm}$ , resultando em um intervalo total do módulo de vetor de espalhamento de  $0,009\text{\AA}^{-1} < q < 0,550\text{\AA}^{-1}$ , onde  $q$  é o módulo do vetor  $q$  definido por  $q = (4\pi / \lambda)\text{sen}\theta$  (onde  $2\theta$  é o ângulo de espalhamento como na seção 1.2 do capítulo 1). Todas as amostras de proteínas

estavam em tampão PBS (137 mM NaCl; 2,7 mM KCl; 10 mM Na<sub>2</sub>HPO<sub>4</sub>; 1,8 mM KH<sub>2</sub>PO<sub>4</sub>; pH 7,4) e foram centrifugadas por 30 min em uma ultracentrífuga, à 356.000 g, em 4°C para remover qualquer agregado e, posteriormente, colocadas no gelo. Para realizar as experiências de SAXS, as amostras foram inseridas, em um porta-amostra de 1 mm de espessura com janelas de mica [Cavalcanti *et al.*, 2004]. A temperatura das amostras foi controlada através de circulação de água no porta-amostra com a ajuda de um banho térmico monitorado por um termopar. Foram feitas três coletas sucessivas de 300 s cada para cada amostra. O espalhamento do tampão foi detectado separadamente antes e depois do espalhamento das amostras durante 300 s. Todas as curvas de espalhamento foram individualmente corrigidas pela resposta do detector e normalizadas pela intensidade integrada do feixe incidente e pela absorção das amostras como descrito na seção 1.6 do capítulo 1. A curva média do espalhamento do tampão foi então subtraída do espalhamento da amostra correspondente. As curvas resultantes foram cuidadosamente inspecionadas para verificar a formação de bolhas na solução e a existência de possíveis danos induzidos por radiação nas proteínas, mas tais efeitos não foram observados. Normalmente, a verificação da existência de dano induzido por radiação é realizada expondo a mesma amostra (sem retirá-la ou trocá-la do porta amostra) duas vezes ou mais com o mesmo tempo de exposição. Se assim as curvas tratadas (após subtração do espalhamento do tampão) forem sistematicamente diferentes, o dano por radiação está presente. As amostras foram medidas em mais de uma concentração para cada amostra: 10,05 mg/mL, 4,99 mg/mL e 2,57 mg/mL para FEZ1(1-392); 9,17 mg/mL, 5,6 mg/mL e 2,57 mg/mL para FEZ1(1-227) e, por último, 1,76 mg/mL e 0,97 mg/mL para FEZ1(92-194). Após normalizar as curvas pelas respectivas concentrações, as curvas de intensidade de espalhamento dessas amostras foram analisadas para investigar se havia dependência com a concentração, porém esse efeito não foi observado. Além disso, uma solução de albumina bovina (BSA, massa molecular de 66 kDa) no mesmo tampão das amostras, em uma concentração de 4,58 mg/mL, foi preparada e usada como amostra padrão para estimar as massas moleculares da proteína e seus fragmentos da maneira descrita na seção 1.4.7 do capítulo 1 (equação 1.33) dessa tese. As análises dos dados de SAXS e modelagem tridimensional foram feitas conforme descritas no capítulo 1.

### **3.2.2.3. MUDANÇAS CONFORMACIONAIS DA FEZ1 EM FUNÇÃO DA CONCENTRAÇÃO DE DESNATURANTE (URÉIA)**

A proteína FEZ1 possui uma região na porção C-terminal cujas predições teóricas e evolutivas indicam ser uma região estruturada com várias alfa-hélices juntas (entre os resíduos Glu231 e Gly296 da proteína) e que têm uma alta probabilidade de formarem motivos estruturais

chamados *coiled-coils*<sup>xi</sup> em suas interações. Note que esses motivos estão numa região da proteína que não existe no fragmento FEZ1(1-227). Dessa forma, para tentar comprovar indiretamente a existência dessa porção estruturada da FEZ1 através da técnica de SAXS, nós acompanhamos as mudanças estruturais das amostras FEZ1(1-392) e FEZ1(1-227) adicionando um agente desnaturante (uréia) na solução. A uréia vai modificar o ambiente de tal forma a desfazer os poucos elementos de estrutura secundária da proteína (irá “desnaturar” a proteína). Sendo assim, novas experiências de SAXS foram realizadas com essas amostras nas mesmas condições experimentais anteriores. Para estudar as mudanças de conformação da proteína, uma solução estoque de uréia a 12M foi previamente preparada no mesmo buffer da amostra (137 mM NaCl; 2,7 mM KCl; 10 mM Na<sub>2</sub>HPO<sub>4</sub>; 1,8 mM KH<sub>2</sub>PO<sub>4</sub>; pH 7,4). As amostras de proteínas em duas concentrações de uréia (3 e 6 M de uréia) foram volumetricamente preparadas dessa solução estoque e das soluções individuais de proteína. Depois disso, todas as amostras foram deixadas em temperatura ambiente por pelo menos duas horas antes dos experimentos.

#### **3.2.2.4. REGIÃO DE DIMERIZAÇÃO DA FEZ1: O FRAGMENTO FEZ1(92-194)**

As análises com o fragmento FEZ1(92-194) ocorreram após todas as análises com as amostras FEZ1(1-392) e FEZ1(1-227). Logo, para entender as análises com esse fragmento, alguns resultados serão adiantados aqui. Com base em resultados prévios de outras técnicas (e que nós confirmamos por SAXS) a proteína FEZ1 dimeriza na região N-terminal. Sabendo disso e conhecendo a seqüência da proteína, foi possível obter um fragmento estável da FEZ1 que engloba a região onde possivelmente ocorreria a dimerização da FEZ1, que identificamos aqui como FEZ1(92-194). Esse fragmento foi estudado com o objetivo de conhecer mais detalhes sobre essa dimerização. Analisando os aminoácidos dessa região, encontramos uma cisteína, que é um aminoácido que possui um átomo de enxofre e, por isso, pode formar uma ponte dissulfeto com outra cisteína. Em razão disso, acreditamos que essa ponte dissulfeto formada entre duas moléculas iguais do fragmento FEZ1(92-194) esteja contribuindo majoritariamente para a dimerização, embora possa haver outras razões, como por exemplo, equilíbrio de cargas, já que é uma região carregada da proteína. Sabendo disso, 2 mM de um agente redutor de pontes dissulfeto (TCEP, que é a abreviação para Tris(2-carboxyethyl)phosphine) foram adicionados à solução. Após isso,

---

<sup>xi</sup> *Coiled-coils* são motivos estruturais formados por elementos de alfa-hélices que se enrolam como se fossem o entrelaçado de uma corda.

também realizamos experiências de SAXS com a solução reduzida e não reduzida. Se a ponte dissulfeto estiver realmente contribuindo, estaremos esperando encontrar somente monômeros do fragmento na solução reduzida.

### **3.2.2.5. MODELAGEM TRIDIMENSIONAL *AB INITIO* DA CONFORMAÇÃO DA PROTEÍNA FEZ1 EM SOLUÇÃO.**

Embora a forma de proteínas nativamente desoveladas seja, por si só, algo difícil de definir, ainda é possível obter modelos da conformação média (ou daquela que aparece com maior frequência) da sua estrutura. Em vista disso, as conformações médias de FEZ1(1-392) e FEZ1(1-227) foram modeladas a partir das curvas experimentais de SAXS usando dois métodos de modelagem *ab initio*: o método de átomos fictícios (seção 1.9.3 do capítulo 1) e o método de resíduos fictícios (seção 1.9.4 do capítulo 1). É importante salientar, no entanto, que uma solução única não pode ser obtida a partir destes cálculos somente. Por esse motivo, dez cálculos diferentes foram realizados para cada amostra e cada método. Os modelos resultantes, separados por método e amostra, foram então comparados e as discrepâncias espaciais normalizadas (NSD) foram calculadas conforme descrito na seção 1.10 do capítulo 1. O parâmetro de discrepância espacial normalizada avalia de forma quantitativa as diferenças entre os modelos obtidos dos cálculos independentes. Dessa forma, o modelo mais típico (com o menor valor de NSD) foi considerado como a descrição mais próxima de uma conformação média de cada molécula. Para os modelos de átomos fictícios, é possível obter um modelo médio com o volume excluído corrigido pelo volume médio dos modelos individuais. Já no modelo de resíduos fictícios, essa média de modelos não pode ser feita para não perder o arranjo tipo cadeia do modelo. Sendo assim, as discrepâncias espaciais normalizadas (NSD) foram avaliadas, e os cinco modelos mais típicos (com menores valores de NSD) foram considerados representantes típicos das conformações possíveis da molécula em solução.

### **3.2.2.6. MODELAGEM DA FEZ1(1-392) CONSIDERANDO A TOPOLOGIA DA REGIÃO COM PROPENSÃO À FORMAÇÃO DE *COILED-COILS***

A topologia da proteína inteira FEZ1(1-392), incluindo região com propensão a formação de *coiled-coils*, foi examinada por modelagem de corpo rígido a partir dos dados de SAXS. Essa abordagem emprega o protocolo de recozimento simulado (*simulated annealing*, seção 1.9.2 do capítulo 1) para encontrar a posição e orientação ótimas de uma estrutura de alta resolução (no caso,

a região com propensão à formação de *coiled-coils*) e a conformação mais provável das regiões desconhecidas de tal modo a ajustar a curva de intensidade de espalhamento experimental da proteína usando uma cadeia de resíduos fictícios (ver seção 1.9.5 do capítulo 1). Para construir o modelo, a predição teórica do conteúdo de alfa hélices, juntamente com a predição de formação de *coiled-coils*, ambas apresentadas no artigo Lanza *et al.* [Lanza *et al.*, 2009], foram usadas em nossos cálculos. Uma estrutura de alta resolução, com várias alfa-hélices juntas e com propensão a formação de *coiled-coil*, foi encontrada no banco de dados “Protein Data Bank” (<http://www.rcsb.org/>). Essa estrutura está registrada com o código PDB *1P9I* e foi selecionada pelo seu tamanho e características. Para garantir que os resultados estivessem em concordância entre eles e pudessem representar a proteína, dez cálculos independentes foram feito com esse método. Os resultados foram então comparados e analisados conforme descrito na seção 1.10 do capítulo 1.

### **3.2.2.7. AS CONFORMAÇÕES DE FEZ1 EM SOLUÇÃO: MÉTODO DE OTIMIZAÇÃO DE CONJUNTOS DE CONFORMAÇÕES**

A fim de caracterizar o conjunto de conformações adotadas pela proteína FEZ1 em solução, o método de otimização de conjuntos de conformações (seção 1.9.6 do capítulo 1) foi aplicado aos dados de SAXS das proteínas FEZ1(1-392) e FEZ1(1-227). Neste método, é assumido que os dados experimentais de intensidade de espalhamento possam ser representados por uma média do espalhamento das diferentes conformações da molécula. No final, um conjunto de possíveis conformações de cada construção de FEZ1 analisada foi obtido. Também, como já mencionado, obtivemos as distribuições de  $R_g$  e  $D_{máx}$  das conformações do conjunto que melhor ajustam os dados experimentais.

## **3.2.3. RESULTADOS**

### **3.2.3.1. RESULTADOS DAS EXPERIÊNCIAS COM A PROTEÍNA NA CONDIÇÃO NATIVA**

Os dados de DLS das amostras mostraram um pico único e estreito para cada uma das três amostras, o que indica que há somente uma espécie em solução, satisfazendo a condição necessária de monodispersidade da solução para as experiências de SAXS. Sendo assim, as curvas experimentais de intensidade de espalhamento das amostras FEZ1(1-392), FEZ1(1-227) e FEZ1(92-

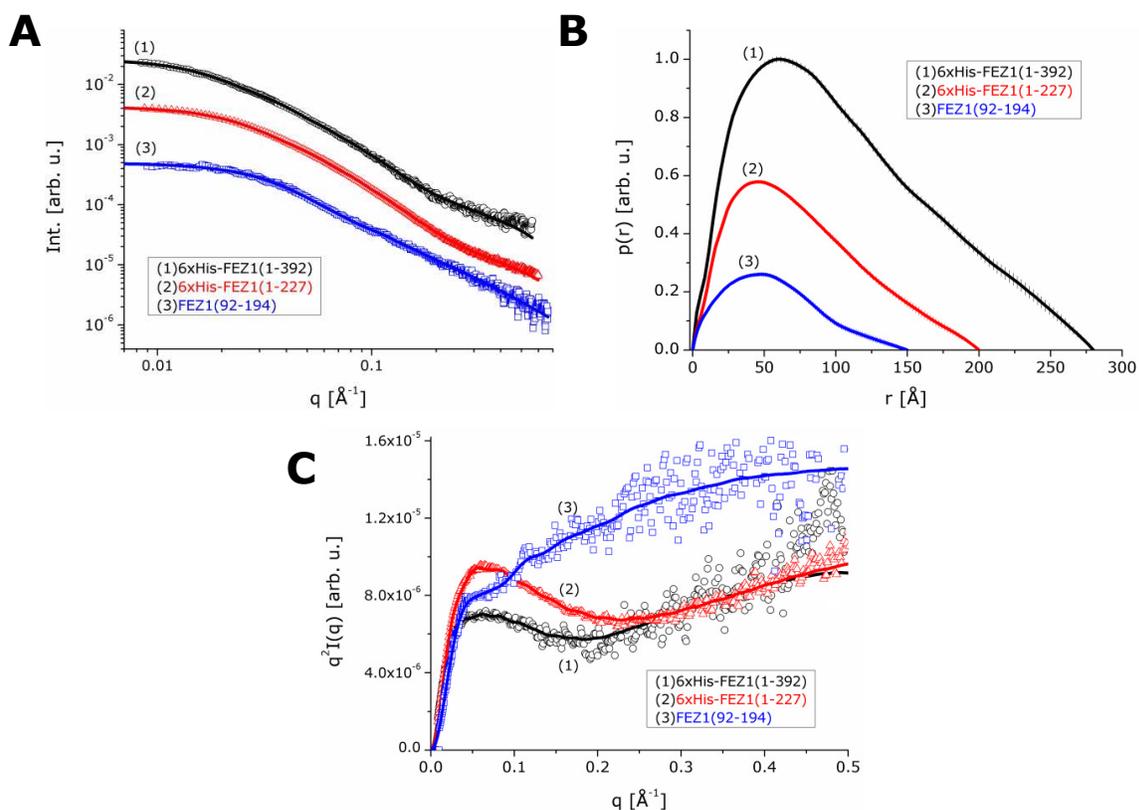
194) em solução são mostradas na figura 3.2(A), juntamente com os ajustes individuais devidamente calculados com o método da transformada de Fourier indireta (seção 1.7 do capítulo 1). As funções  $p(r)$  resultantes desses cálculos estão mostradas na figura 3.2(B).

Os valores de raio de giro ( $R_g$ ) obtidos da aproximação de Guinier (seção 1.4.2 do capítulo 1), da aproximação de Debye (veja equação 1.22 da seção 1.4.4 do capítulo 1) e da função  $p(r)$  (equação 1.20 da seção 1.4.3 do capítulo 1) são mostrados na tabela 3.1. Os valores de dimensão máxima ( $D_{max}$ ) obtidos para as moléculas também estão mostrados na tabela 3.1. Usando a solução de BSA como amostra de referência, as massas moleculares para as construções de FEZ1 obtidas dos resultados de SAXS também estão na tabela 3.1. Os valores de  $\sim 60kDa$  para 6xHis-FEZ1(1-227) e  $\sim 95kDa$  para 6xHis-FEZ1(1-392) indicam que ambas as construções de FEZ1 existem como dímeros em solução, desde que os valores obtidos foram aproximadamente duas vezes os valores calculados a partir da seqüência de aminoácidos (utilizamos a ferramenta ProtParam [Gasteiger *et al.*, 2005]):  $29,7 kDa$  para 6xHis-FEZ1(1-227) e  $48,6 kDa$  para 6xHis-FEZ1(1-392). Por razão dessa dimerização, como mencionado, decidimos também estudar o fragmento FEZ1(92-194), que engloba justamente a região em que acreditamos ser onde ocorre a dimerização. Como esperado, o valor de  $\sim 22 kDa$  também é aproximadamente duas vezes maior que o valor calculado para o monômero:  $12,2 kDa$ . Portanto, FEZ1 dimeriza pela sua região N-terminal e temos uma idéia de quais resíduos colaboram nessa dimerização.

Todos esses parâmetros dimensionais juntos sugerem que as moléculas tenham formas relativamente alongadas, o que é também evidenciado pelas formas assimétricas e características das funções  $p(r)$ , cujo máximo está deslocado para menores valores de  $q$ . As representações de Kratky das três construções da proteína são mostradas na figura 1(C). Todas essas representações apresentam características típicas de cadeias flexíveis (veja seção 1.5 e figura 1.3 do capítulo 1). As construções FEZ1(1-227) e FEZ1(1-392) ainda parecem apresentar uma oscilação no começo (uma espécie de máximo local), podendo indicar a existência de algumas porções de estrutura, sendo em quantidade ligeiramente maior na FEZ1(1-392). Isso pode indicar a presença da região estruturada com propensão à formação de motivos *coiled-coil*. Agora, a representação de Kratky do fragmento FEZ1(92-194) indica uma conformação bastante flexível e nada compacta, sugerindo a ausência quase total de regiões estruturadas. Obviamente, isso não exclui a existência de regiões com elementos de estrutura secundária pequenas demais para serem sensíveis à técnica de SAXS. Esses dados sugerem que a proteína FEZ1 forma dímeros através de sua região N-terminal e possui conformações amplamente abertas e flexíveis em solução.

**Tabela 3.1** – Parâmetros dimensionais das amostras de FEZ1(1-392), FEZ1(1-227) e FEZ1(92-194) obtidos a partir dos dados de SAXS.

Amostra	$R_g$ (Å) (Guinier)	$R_g$ (Å) (Debye)	$R_g$ (Å) ( $\rho(r)$ )	$D_{m\acute{a}x}$ (Å)	M(kDa) (SAXS)	M(kDa) (Calc.)	Estado Oligomérico
6xHis-FEZ1(1-392)	80±2	84,6±0,5	83±1	280	~95	48,6	Dímero
6xHis-FEZ1(1-227)	62±2	62,9±0,5	61,6±0,6	200	~60	29,7	Dímero
FEZ1(92-194)	44±4	45±1	45±1	150	~22	12,2	Dímero



**Figura 3.2** - Curvas experimentais de SAXS para 6xHis-FEZ1(1-392), 6xHis-FEZ1(1-227) e FEZ1(92-194) em solução e os resultados dos procedimentos de ajuste. (A) Curva de espalhamento experimental de 6xHis-FEZ1(1-392) [círculos abertos pretos], 6xHis-FEZ1(1-227) [triângulos abertos vermelhos] e FEZ1(92-194) [quadrados abertos azuis] e os ajustes teóricos [linhas sólidas] dos dados usando a rotina de transformada de Fourier indireta. As curvas foram deslocadas por um fator constante para permitir uma melhor visualização. (B) Respectivas funções de distribuição de pares de distâncias  $p(r)$ . (C) Respectivas representações de Kratky das curvas de intensidade.

### 3.2.3.2. RESULTADOS DAS MUDANÇAS CONFORMACIONAIS DA PROTEÍNA FEZ1 COM A ADIÇÃO DE URÉIA

Devido a sua característica desestruturada, a interação da proteína FEZ1 com outras proteínas deve ser baixa afinidade, porém espera-se que essas interações sejam específicas. Essa especificidade é possível se ao menos pequenas porções estruturadas existirem na molécula da FEZ1. Além disso, já sabemos que ela dimeriza, característica não esperada em moléculas completamente desenoveladas. Os resultados mostram que sua estrutura é desordenada, porém não é uma estrutura aleatória, sem qualquer organização de estrutura secundária. Predições teóricas de estrutura com base na seqüência de aminoácidos já mostraram que há pequenas regiões contendo alfa hélices e uma possível região com probabilidade de formação de *coiled-coils*, como já mostramos antes [Lanza *et al.*, 2009]. Logo, com a adição de uréia na solução da proteína, mudanças conformacionais da proteína FEZ1 podem ser observadas.

Após a adição de 3 e 6 M de uréia na solução, nós notamos com base nos dados de SAXS que mudanças conformacionais ocorreram tanto na FEZ1(1-392) quanto na FEZ1(1-227), porém, não da mesma maneira. As curvas de intensidade de espalhamento experimentais que mostram essas mudanças estão apresentadas na figura 3.3(A) e 3.3(C) para FEZ1(1-392) e FEZ1(1-227), respectivamente. Além disso, as representações de Kratky mostram aumento da flexibilidade de ambas moléculas, como visto nas figuras 3.3(B) e 3.3(D), respectivamente. Isso pode ser correlacionado com a perda das poucas regiões estruturadas da FEZ1 após a adição de uréia. Como uréia é um composto químico que pode reduzir interações hidrofóbicas internas de uma proteína, nós acreditamos que as mudanças em nossas amostras ocorreram principalmente na região com probabilidade de formação de *coiled-coils*, que seria a região mais estruturada para a qual a técnica de SAXS seria bastante sensível. Entretanto, o que nos intrigou foi o fato que mudanças conformacionais também foram observadas na FEZ1(1-227) (figura 3.3(D)), o que indica a existência de pequenas regiões com elementos de estrutura secundária (possivelmente, pequenas hélices) na região N-terminal da FEZ1. Provavelmente, aquelas regiões em amarelo na figura 3.1 contêm aminoácidos que devam estar formando pequenas regiões com estrutura na porção N-terminal da FEZ1. Outra coisa bastante interessante foi que mudanças nas conformações também foram observadas em FEZ1(1-227) com a adição de 3 M de uréia na solução, porém não houve mais mudanças com adição de 6 M de uréia, como mostrado nas figuras 3.3(B) e 3.3(D), onde as duas curvas correspondentes se sobrepõem. Entretanto, com a FEZ1(1-392) aconteceu algo diferente. Primeiramente, mudanças foram observadas com a adição de 3 M de uréia e mais mudanças também foram observadas com 6 M de uréia na solução. Como mostrado nas figuras 3.3(A) e

3.3(B), vemos curvas diferentes para 3 M e 6 M de uréia, indicando uma conformação mais flexível e ainda menos compacta para FEZ1(1-392) com 6 M de uréia que sua conformação com 3 M de uréia. Essa diferença indica a presença de uma região estruturada na FEZ1(1-392) que não existe na FEZ1(1-227). Logo, de maneira indireta, a existência de uma região com alfa-hélices juntas, com probabilidade de formação de “*coiled-coil*”, na FEZ1(1-392) explica esses resultados. Porém, para comprovar essa idéia, seria necessário obter a estrutura de alta resolução da FEZ1, o que atualmente não é possível com as técnicas experimentais disponíveis.

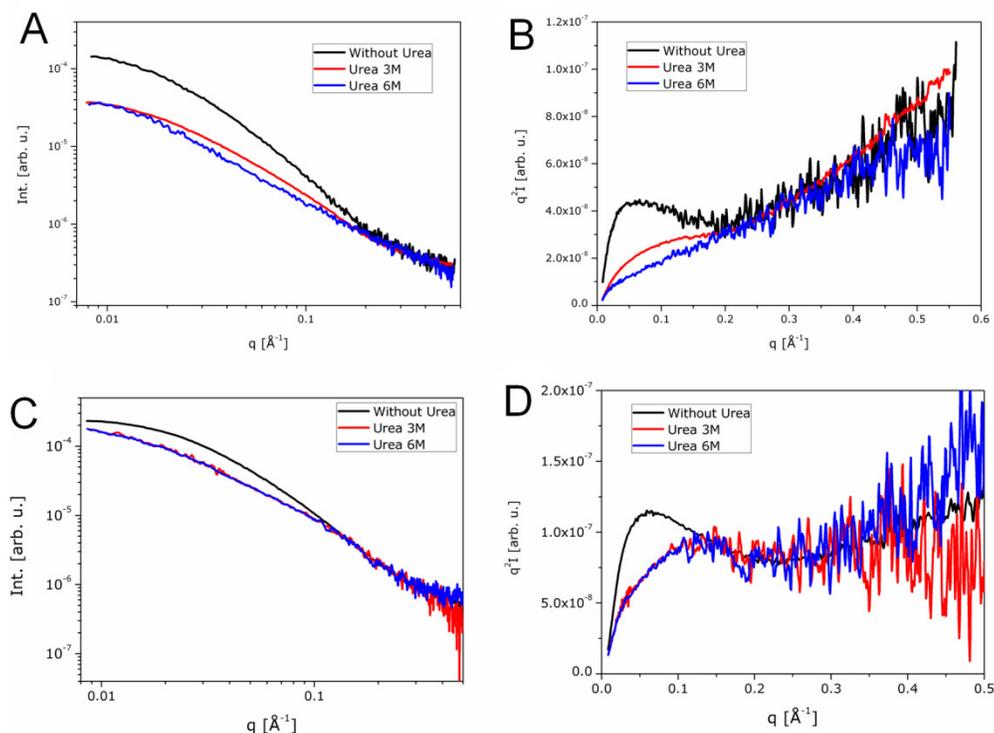
**Tabela 3.2** – Resultados das mudanças de raio de giro ( $R_g$ ) de FEZ1(1-392) e FEZ1(1-227) em solução com uréia.

Amostra	FEZ1(1-392)		FEZ1(1-227)	
	$R_g$ (Å) (Debye)	$R_g$ (Å) (Guinier)	$R_g$ (Å) (Debye)	$R_g$ (Å) (Guinier)
Sem uréia	$84,6 \pm 0,5$	$80 \pm 2$	$62,9 \pm 0,5$	$62 \pm 2$
3 M de uréia	$85 \pm 1$	$87 \pm 16$	$84 \pm 8$	$83 \pm 15$
6 M de uréia	$101 \pm 2$	$92 \pm 22$	$84 \pm 9$	$83 \pm 16$

Essas primeiras análises indicaram mudanças nas conformações de nossas amostras. Agora vamos analisar as mudanças nos parâmetros dimensionais delas com a adição de uréia. Os resultados estão mostrados na tabela 3.2. Os raios de giro ( $R_g$ ) da FEZ1(1-392) obtidos pela aproximação de Debye foram  $85 \pm 1 \text{ Å}$  e  $101 \pm 2 \text{ Å}$  para 3 e 6 M de uréia, respectivamente. Usando a aproximação de Guinier, esses valores ficam menos precisos, resultando em  $87 \pm 16 \text{ Å}$  para 3 M de uréia e  $92 \pm 22 \text{ Å}$  para 6 M de uréia. A determinação de  $R_g$  com a aproximação de Guinier apresentou altos valores de erros estatísticos devido ao fato de que a região de validade dessa aproximação para conformações altamente flexíveis é muito estreita, havendo poucos pontos experimentais. De qualquer forma, analisando os valores obtidos com aproximação de Debye, nós notamos que o valor de  $R_g$  quase não se altera com 3 M uréia em relação a conformação nativa ( $\sim 83 \text{ Å}$ ) e só observamos alteração com 6 M uréia, em que o valor de  $R_g$  ficou maior (tabela 3.2). Baseado nisso, nós poderíamos esperar que 3 M uréia não alteraria a conformação de FEZ1(1-392), porém não foi isso que observamos através das representações de Kratky das figura 3.3(B) e 3.3(D). Isso mostra que, mesmo com alterações no conjunto de conformações de FEZ1(1-392) em solução, o valor médio de  $R_g$  deve permanecer próximo daquele na condição nativa. Portanto, as representações de Kratky são essenciais para as análises aqui apresentadas.

De maneira similar, os raios de giro ( $R_g$ ) de FEZ1(1-227) obtidos pelas aproximações de Debye e Guinier são, respectivamente,  $84 \pm 8 \text{ Å}$  e  $83 \pm 15 \text{ Å}$  para a solução com 3 M de uréia e

valores idêntico para a solução com 6 M de uréia. Essa semelhança concorda com a semelhança das curvas de intensidade de espalhamento e vemos que o valor em torno de  $84 \text{ \AA}$  é maior que o valor de aproximadamente  $62 \text{ \AA}$  da conformação nativa, indicando que tanto as conformações quanto o valor médio de  $R_g$  se alteraram.



**Figura 3.3** – Dados experimentais de SAXS mostrando as mudanças conformacionais em FEZ1(1-392) e em FEZ1(1-227) com adição de uréia na solução. (A) Curvas experimentais de intensidade de espalhamento de FEZ1(1-392) em solução com 0M, 3M e 6M uréia. (B) Respectivas representações de Kratky mostrando a perda de estrutura de FEZ1(1-392). (C) Curvas experimentais de intensidade de espalhamento de FEZ1(1-227) em solução com 0, 3M e 6M uréia. (D) Respectivas representações de Kratky mostrando a perda de pequenos elementos estruturados de FEZ1(1-227).

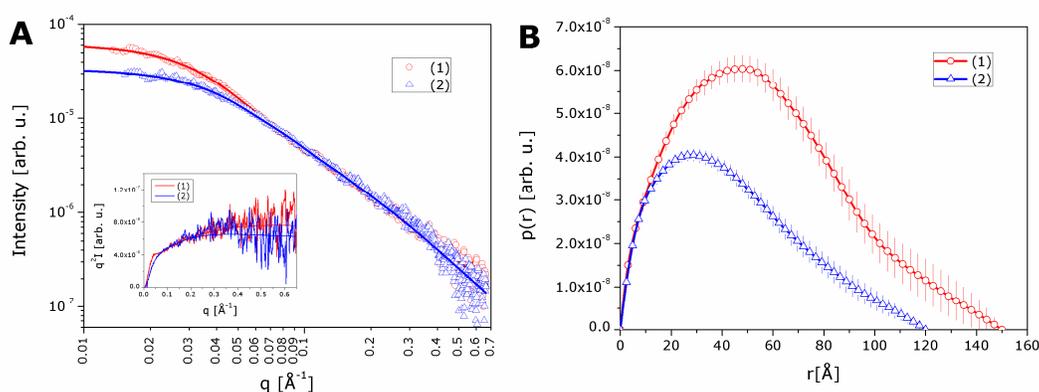
### 3.2.3.3. RESULTADOS DAS ANÁLISES DA REGIÃO DE DIMERIZAÇÃO DE FEZ1: O FRAGMENTO FEZ1(92-194)

Os resultados das experiências de SAXS com o fragmento FEZ1(92-194) em solução reduzida com 2 mM de TCEP e não reduzida são mostrados na figura 3.4. As curvas de intensidade de espalhamento são apresentadas na figura 3.4(A). As respectivas funções de distribuição de pares de distância foram calculadas através do método de transformada de Fourier indireta das intensidades (seção 1.7 do capítulo 1) e são apresentadas na figura 3.4(B). Os parâmetros

dimensionais obtidos estão mostrados na tabela 3.3. Com base na estimativa de massa molecular a partir dos dados de SAXS, confirmamos nossa hipótese de que com a adição de TCEP haveria somente monômeros em solução. Essa informação é importante, pois diz que a ponte dissulfeto tem uma grande colaboração na dimerização da FEZ1. É claro que não podemos excluir outros fatores que influenciam essa dimerização, como por exemplo, o equilíbrio de cargas dessa região no tampão utilizado.

**Tabela 3.3** – Parâmetros dimensionais do fragmento FEZ1 (92-194) obtidos por SAXS antes e após a redução da ponte dissulfeto com TCEP.

Amostra	$R_g$ (Å) (Guinier)	$R_g$ (Å) (Debye)	$R_g$ (Å) ( $p(r)$ )	$D_{max}$ (Å)	M(kDa) (SAXS)	M(kDa) (calc.)	Estado oligomérico
FEZ1(92-194)	44 ± 4	45 ± 1	45 ± 1	~150	~22	12,2	Dímero
FEZ1(92-194) reduzido com TCEP	34 ± 4	37 ± 1	36 ± 1	~120	~11	12,2	Monômero



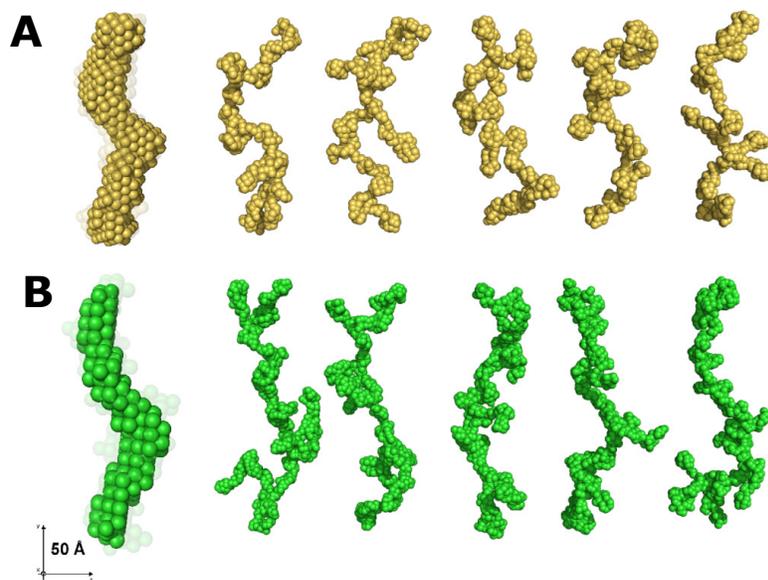
**Figura 3.4** – Resultados de SAXS das experiências com o fragmento FEZ1(92-194) da região N-terminal da proteína FEZ1. (A) Curva de intensidade de espalhamento do dímero (vermelha) e do monômero (azul), após a redução de pontes de dissulfeto com a adição de 2 mM de TCEP à solução. (B) Respectivas funções de distribuição de pares de distância,  $p(r)$ .

### 3.2.3.4. RESULTADOS DA MODELAGEM DA CONFORMAÇÃO DA PROTEÍNA FEZ1

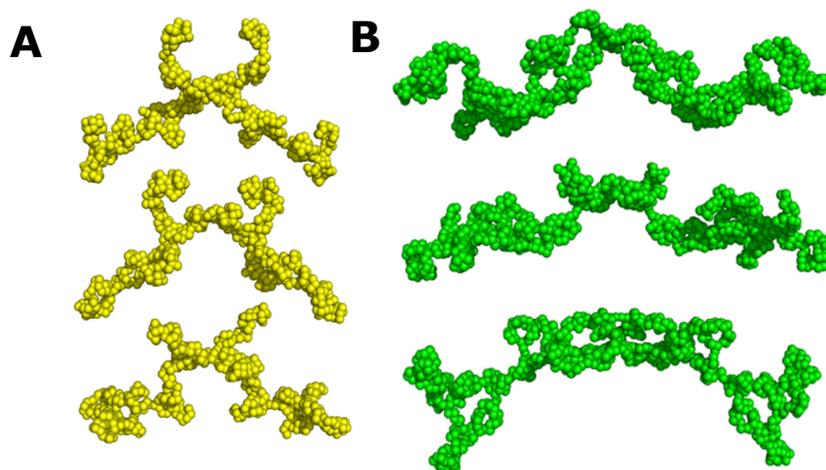
As conformações de baixa-resolução de FEZ1(1-227) e FEZ1(1-392) foram obtidas dos dados experimentais de SAXS usando os dois métodos descritos anteriormente. Os modelos de baixa resolução foram calculados sem a imposição de simetrias para os dímeros. Como já foi

mencionado antes, na ausência de uma solução única para os cálculos *ab initio*, vários cálculos foram realizados para cada proteína, obtendo resultados com certas similaridades. Para o caso de cada proteína, 10(*dez*) cálculos dos modelos foram realizados sem a imposição de restrições de simetria espacial (simetria *P1*) e, depois, mais 10(*dez*) cálculos foram realizados impondo uma simetria *P2* (simetria de rotação de 180° em torno de um eixo de simetria da molécula), por causa da conformação dimérica da proteína FEZ1. Os modelos obtidos sem a imposição de simetria (*P1*) para 6xHis-FEZ1(1-227) e 6xHis-FEZ1(1-392) são mostrados na figura 3.5(A) e 3.5(B), respectivamente. Já alguns modelos calculados para as mesmas proteínas, mas impondo a simetria *P2* nos cálculos estão mostrados na figura 3.6(A) e 3.6(B). Os valores de NSD para os conjuntos de modelos obtidos pelo método de átomos fictícios estiveram no intervalo de 0,87 a 1,00. A baixa resolução dessas representações somente mostra que as conformações dessas construções da proteína FEZ1 são estendidas. Inspeção das formas obtidas para 6xHis-FEZ1(1-227) e 6xHis-FEZ1(1-392) revela uma série de conformações similares para ambas as construções. A razão entre os valores de volume excluído ( $Vol_{1-392}/Vol_{1-227}$ ) dos modelos obtidos para FEZ1(1-392) e FEZ1(1-227) é igual a 1,52. Esse valor está bem próximo à razão 1,65 entre os números de aminoácidos (392/227) das duas construções da proteína, indicando que os modelos obtidos são coerentes.

Ainda na figura 3.5, cinco modelos *P1* representativos calculados com o método de resíduos fictícios são apresentados em linha para cada proteína, à direita do modelo médio dos modelos de átomos fictícios. E na figura 3.6, três modelos *P2* representativos são mostrados. Comparando os conjuntos de modelos, aqueles calculados com simetria *P2* parecem fornecer uma melhor visualização, de maneira mais intuitiva, da possível organização do dímero. Deve ser mencionado que, para esses modelos com resíduos fictícios, os valores de NSD são altos: os valores vão de 2,62 a 2,82 para 6xHis-FEZ1(1-227) e de 3,22 a 3,36 para 6xHis-FEZ1(1-392). Esses altos valores de NSD eram esperados, desde que eles refletem as diferenças conformacionais existentes nessas proteínas parcialmente desoveladas, portanto, os resultados são aceitáveis. Embora um modelo médio não possa ser obtido, a abordagem de mais alta resolução dos modelos com resíduos fictícios fornece uma melhor idéia do espaço conformacional ocupado pela proteína inteira, FEZ1(1-392) e pela construção do seu N-terminal, FEZ1(1-227).



**Figura 3.5** - Modelos *ab initio* de baixa resolução obtidos dos dados de SAXS da FEZ1 calculados sem a imposição de simetria. (A) Vistas selecionadas dos modelos do dímero de FEZ1(1-227) [amarelo]. (B) Vistas selecionadas dos modelos do dímero de FEZ1(1-392) [verde]. Os modelos são mostrados utilizando o programa PyMOL [DeLano, 2002]. Os modelos da primeira coluna em cada linha são os modelos médios dentre os modelos obtidos pelo método de átomos fictícios. Os restantes são cinco modelos selecionados para cada proteína dentre os modelos obtidos com método de resíduos fictícios.

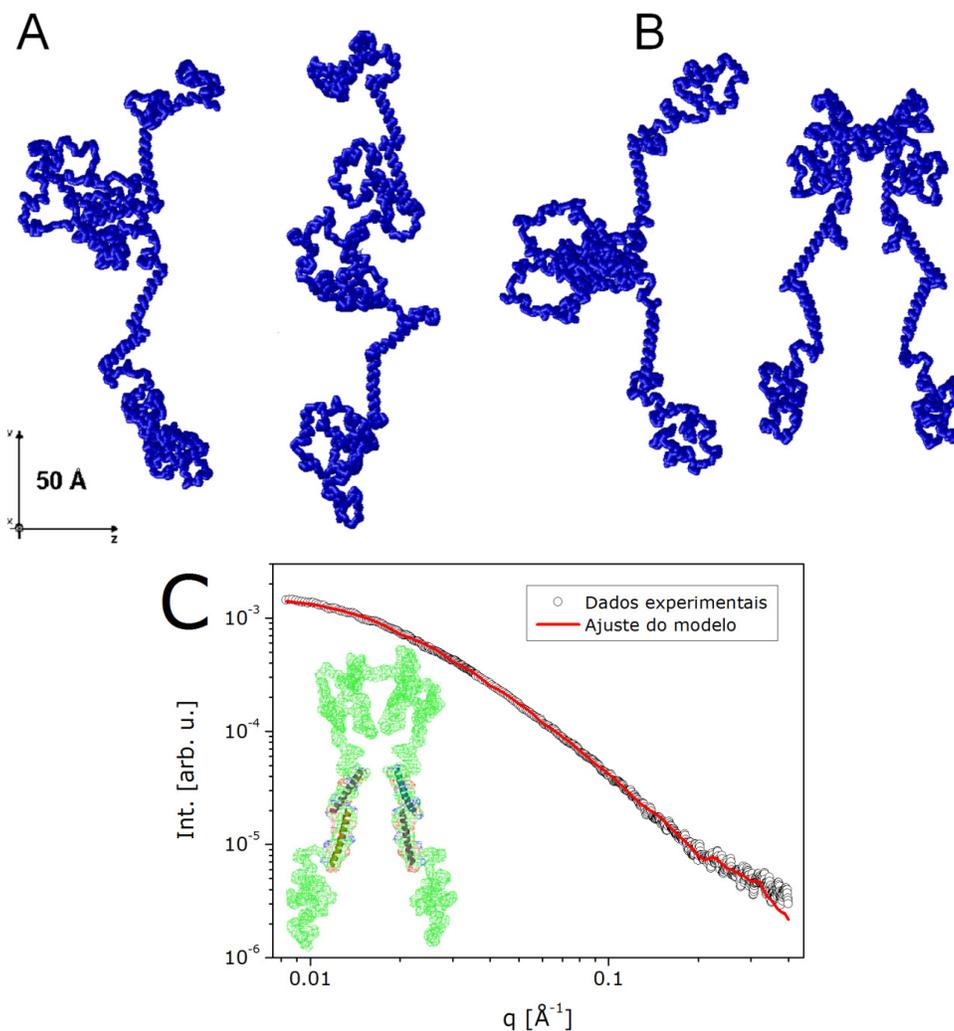


**Figura 3.6** - Modelos de resíduos fictícios de baixa resolução obtidos dos dados de SAXS da FEZ1(1-392) através de cálculos *ab initio* e impondo simetria P2 nos cálculos. (A) Vistas selecionadas dos modelos do dímero de FEZ1(1-227) [amarelo]. (B) Vistas selecionadas dos modelos do dímero de FEZ1(1-392) [verde]. Os modelos são mostrados utilizando o programa PyMOL [DeLano, 2002].

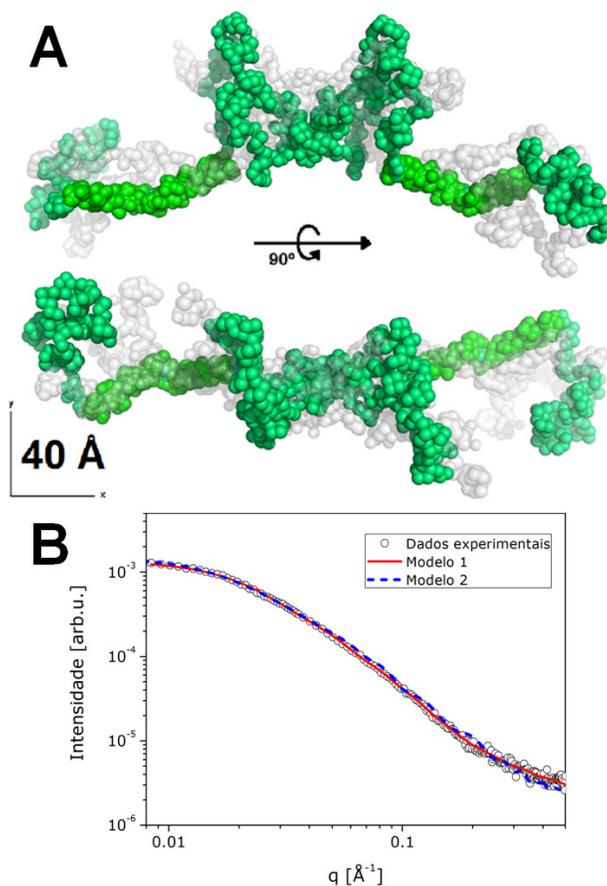
### 3.2.3.5. MODELAGEM DA TOPOLOGIA DE FEZ1 CONSIDERANDO A REGIÃO COM PROPENSÃO À FORMAÇÃO DE MOTIVOS “COILED-COILS”

Na tentativa de obter informações sobre a posição e topologia da região estruturada com propensão a formação de *coiled-coils* da proteína inteira FEZ1(1-392), nós usamos uma estrutura de alta resolução, com várias alfa-hélices juntas e com propensão a formação de *coiled-coils* (ver seção 3.2.2.6). Essa estrutura foi considerada como um corpo rígido durante o cálculo, sendo somente transladada e rotacionada. O restante da proteína foi construído com modelagem *ab initio* do tipo resíduos fictícios, permitindo flexibilidade dessas regiões. Dessa maneira, nós tentamos modelar a proteína realizando o cálculo sem a imposição de simetria espacial e com a imposição de simetria espacial simetria *P2* (simetria de rotação de 180° em torno de um eixo de simetria da molécula), por causa da conformação dimérica da proteína FEZ1. Dez cálculos foram realizados em cada conjunto. Dois modelos resultados típicos obtidos sem a imposição de simetria espacial são mostrados na figura 3.7(A) e dois modelos calculados com imposição de simetria espacial *P2* são apresentados na figura 3.7(B). A figura 3.7(C) mostra o ajuste do espalhamento calculado de um modelo típico aos dados experimentais. No destaque dessa figura está uma representação tipo *mesh* de um dos modelos, mostrando a posição e orientação da região estruturada.

Podemos agora, comparar o espaço ocupado por um modelo *ab initio*, por exemplo, um modelo de resíduos fictícios da FEZ1(1-392) com um modelo obtido com o método apresentado nessa equação. Para isso, os modelos de menor NSD em cada caso tiveram seus eixos de inércia alinhados e a discrepância espacial normalizada (NSD) entre os dois somente foi minimizada. O resultado está apresentado na figura 3.8. O modelo de resíduos fictícios, apresentado como superfície semi-transparente, apresentou uma ótima superposição com um dos modelos apresentados aqui, indicando consistências nos cálculos. O procedimento de superposição e comparação está descrito na seção 1.10 do capítulo 1.



**Figura 3.7** – Modelos da proteína FEZ1(1-392) levando em conta a topologia da região estruturada com propensão à formação de motivos “*coiled-coil*” obtidos dos dados de SAXS através da combinação de cálculos *ab initio* e modelagem global de corpos rígidos. (A) Dois modelos selecionados obtidos com cálculos sem a imposição de simetria espacial. (B) Dois modelos selecionados obtidos através de cálculos com a imposição de simetria espacial  $P2$  (simetria de rotação de  $180^\circ$  em torno de um eixo de simetria da molécula). (C) Ajuste do espalhamento calculado de um modelo típico aos dados experimentais. Destaque: uma representação do tipo *mesh* do modelo mostrando a região estruturada.



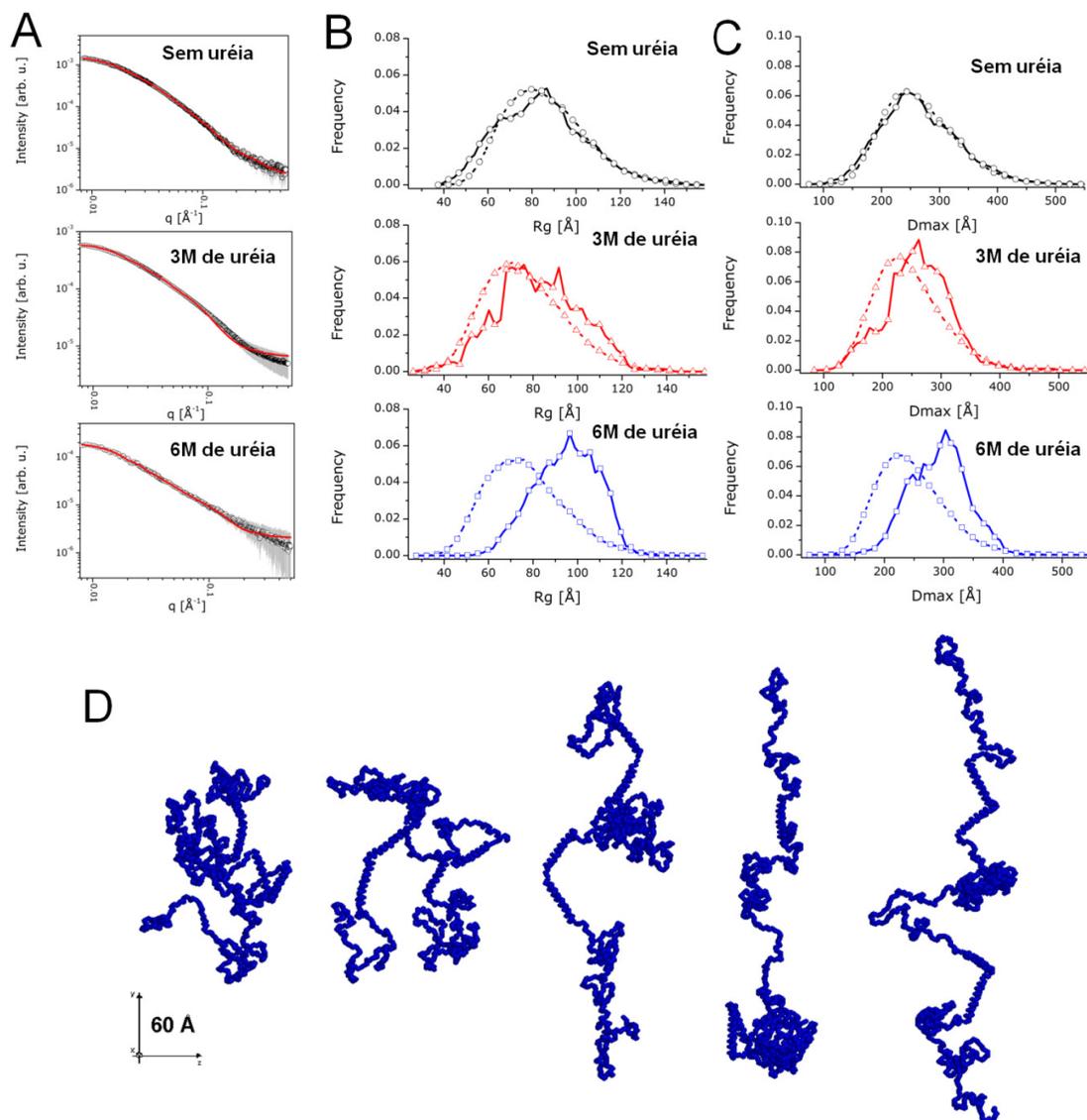
**Figura 3.8** – (A) Superposição do envelope molecular de baixa resolução calculado para FEZ1(1-392) através da combinação de métodos *ab initio* e modelagem de corpos rígidos (esfera verdes) sobre o modelo de resíduos fictícios (esferas cinzas semi-transparente) obtido com cálculo *ab initio* somente. Os cálculos foram realizados assumindo uma simetria P2 para o dímero de FEZ1 e encontrando a posição ótima dos motivos estruturados de *coiled-coil* (esferas verdes mais claras) na região onde foram preditos. (B) Ajuste dos modelos aos dados experimentais de SAXS (Modelo 1: modelo de resíduos fictícios; Modelo 2: envelope molecular levando em conta os motivos *coiled-coil*).

### 3.2.3.6. RESULTADOS OBTIDOS COM O MÉTODO DE OTIMIZAÇÃO DE CONJUNTOS DE CONFORMAÇÕES

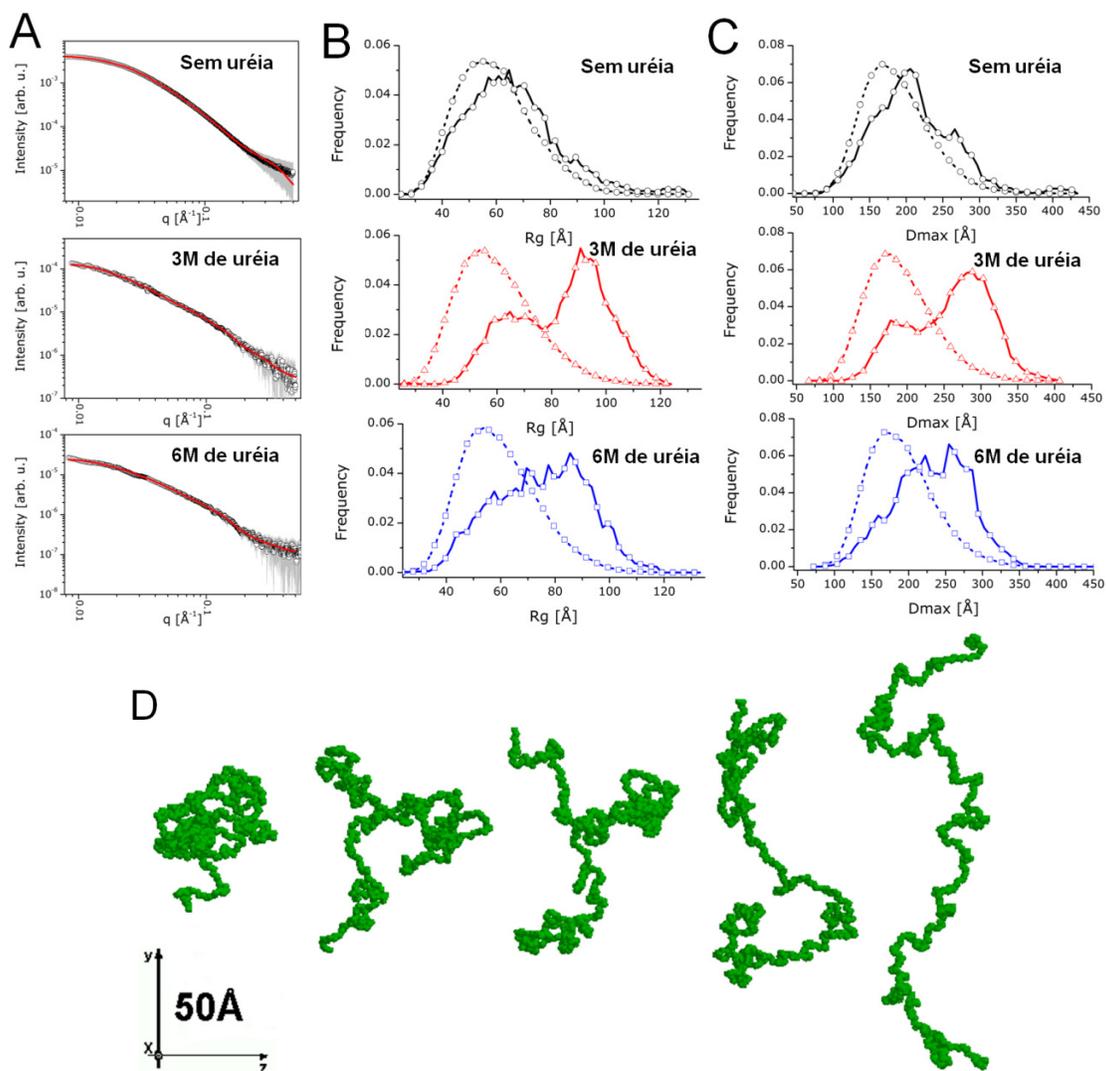
Os resultados obtidos dos cálculos com o método de otimização de conjuntos de conformações são apresentados nas figuras 3.9 e 3.10 para FEZ1(1-392) e FEZ1(1-227), respectivamente. Nós apresentamos os cálculos com as amostras na condição nativa e após a adição de 3 e 6 M de uréia. As curvas de intensidade (juntamente com os correspondentes ajustes do

método) e as distribuições de  $R_g$  e  $D_{máx}$  são apresentadas nas figuras 3.9(A), 3.9(B) e 3.9(C), respectivamente, para a FEZ1(1-392), assim como nas figuras 3.10(A), 3.10(B) e 3.10(C) para a FEZ1(1-227). Em cada uma das distribuições obtidas, há a distribuição inicial (linha tracejada + símbolo) calculada para o conjunto inicial de conformações e a distribuição final resultante da aplicação do algoritmo genético (linha sólida + símbolo), como já discutido na seção 1.9.6 do capítulo 1. As distribuições finais descrevem as possíveis conformações para ambas FEZ1(1-392) e FEZ1(1-227) cujo espalhamento médio ajusta a curva experimental de espalhamento em cada caso (figura 3.9(A) e 3.10(A)). Como esperado, os valores médios de  $R_g$  e  $D_{máx}$  dessas distribuições, para ambas as amostras, concordam com aqueles obtidos das análises típicas de SAXS já mostradas anteriormente. As distribuições de  $R_g$  e  $D_{máx}$  obtidas para a condição nativa de nossas amostras apresentaram um comportamento característico para proteínas desenoveladas e flexíveis, porque cada distribuição resultante não está muito deslocada da distribuição inicial aleatória de conformações. Além disso, o fato de haver somente um pico assegura que somente um tipo de espécie de molécula estava presente em solução em cada caso, não obstante o fato de que haja várias conformações em solução conforme largura do pico.

Após a adição de uréia, vimos que, para a condição com 3 M de uréia ( $\langle R_g \rangle = 86 \text{ \AA}$  e  $\langle D_{máx} \rangle = 268 \text{ \AA}$ ), as distribuições de  $R_g$  e  $D_{máx}$  para FEZ1 (1-392) que os valores médios não afastaram dos valores obtidos para a condição nativa ( $\langle R_g \rangle = 87 \text{ \AA}$  e  $\langle D_{máx} \rangle = 270 \text{ \AA}$ ), embora as distribuições tenham se tornado levemente mais alargadas. No entanto, uma leve mudança foi observada com a adição de 6 M de uréia. Nessa última, os valores médios de  $R_g$  e  $D_{máx}$  se alteraram de  $87 \text{ \AA}$  e  $270 \text{ \AA}$ , respectivamente, na condição nativa para  $97 \text{ \AA}$  e  $296 \text{ \AA}$  em 6 M de uréia. A forma das distribuições também mudou e, ainda, elas se deslocaram para valores maiores desses parâmetros em relação às distribuições iniciais de conformações. Com respeito à FEZ1(1-227), os valores médios de  $R_g$  e  $D_{máx}$  mudaram de  $62 \text{ \AA}$  e  $200 \text{ \AA}$ , respectivamente, na condição nativa para  $86 \text{ \AA}$  e  $235 \text{ \AA}$  em 3 M de uréia, e para  $78 \text{ \AA}$  e  $261 \text{ \AA}$  em 6 M de uréia. A distribuição final em 3M de uréia apresentou dois picos como numa distribuição bimodal, possivelmente indicando duas populações conformacionais coexistindo em solução. Em 6 M de uréia, as distribuições de  $R_g$  e  $D_{máx}$  apresentaram somente um pico, porém mais largas e levemente deslocadas para valores maiores desses parâmetros que da distribuição na condição nativa.



**Figura 3.9** – Resultados da aplicação do método de otimização de conjuntos de conformações aos dados de SAXS de FEZ1(1-392) em soluções com 0, 3 e 6 M de uréia. (A) Curvas experimentais de espalhamento (círculos aberto) com os ajustes correspondentes (linha sólida) obtidos espalhamento médio calculado para as conformações do conjunto otimizado pelo método. (B) Distribuições de  $R_g$  das conformações: inicial (linha tracejada + símbolo) e final (linha sólida + símbolo). (C) Distribuições de  $D_{máx}$  das conformações: inicial (linha tracejada + símbolo) e final (linha sólida + símbolo). (D) Cinco modelos representativos do conjunto de conformações otimizado pelo método.



**Figura 3.10** – Resultados da aplicação do método de otimização de conjuntos de conformações aos dados de SAXS de FEZ1(1-227) em soluções com 0, 3 e 6 M de uréia. (A) Curvas experimentais de espalhamento (círculos aberto) com os ajustes correspondentes (linha sólida) obtidos do espalhamento médio calculado para as conformações do conjunto otimizado pelo método. (B) Distribuições de  $R_g$  das conformações: inicial (linha tracejada + símbolo) e final (linha sólida + símbolo). (C) Distribuições de  $D_{máx}$  das conformações: inicial (linha tracejada + símbolo) e final (linha sólida + símbolo). (D) Cinco modelos representativos do conjunto de conformações otimizado pelo método.

Os cálculos com o método de otimização de conjunto de conformações também fornece modelos de baixa resolução para algumas conformações da proteína em solução. Com respeito a esses modelos, mostrar todos seria de pouca valia já que são em torno de vinte. Sendo assim, nós

decidimos apresentar ao menos cinco conformações representativas nas figuras 3.9(D) e 3.10(D) para FEZ1(1-392) e FEZ1(1-227), respectivamente, na condição nativa. Essas conformações foram selecionadas com base nos valores de  $R_g$  das distribuições finais, selecionado dois das extremidades (menor e maior valor de  $R_g$ ), uma em torno do valor médio e duas com valores intermediários de  $R_g$ . Os modelos das conformações com uréia também não são mostrados aqui, porque a visualização delas não acrescenta muito mais informação às nossas análises.

### 3.2.4.DISSCUSSÕES SOBRE OS RESULTADOS OBTIDOS COM A PROTEÍNA FEZ1

No trabalho feito aqui, nós pudemos verificar que FEZ1 realmente pertence à classe de proteínas intrinsecamente desestruturas (ou algumas vezes chamada nativamente desenoveladas) [Lanza *et al.*, 2009; Silva, *et al.*, 2010a]. Além disso, pudemos observar que essa proteína dimeriza na região N-terminal, através estudos da proteína inteira e da sua porção N-terminal sozinha. Também investigamos os detalhes da região de dimerização estudando o fragmento FEZ1(92-194), que contém os aminoácidos dessa região. Os resultados de SAXS sugerem que a conformação dimérica da FEZ1 permite que as duas regiões C-terminal do dímero estejam expostas para interagir com outras proteínas. A vasta maioria das interações de FEZ1 com suas parceiras são dependentes da região C-terminal e a maioria dos sítios de fosforilação<sup>xii</sup>, induzida pela Proteína quinase C (PKC), estão localizadas nessa região [Kuroda *et al.*, 1999].

Neste trabalho, modelos tridimensionais de baixa resolução do dímero da proteína FEZ1 foram propostos através das medidas de SAXS e modelagem *ab initio* e de corpos rígidos. Os modelos tanto da proteína inteira, quanto do N-terminal, são importante por fornecerem uma idéia de onde possivelmente está a região C-terminal da molécula. Essa região da FEZ1 é um importante domínio de interação proteína-proteína onde todas as proteínas testadas até agora interagem [Assmann *et al.*, 2006]. Se for assumido que as duas regiões C-terminais do dímero da FEZ1 inteira apontam para fora, distante do centro de dimerização (que envolve o N-terminal), isso sugere que as proteínas que interagem poderiam se ligar ao C-terminal para funções de transporte e que a FEZ1 possa funcionar como uma proteína adaptadora dimérica e bivalente.

Sendo isso verdade, é possível começar a entender porque a FEZ1 interage com tantas proteínas diferentes. A maioria das mais de 42 proteínas virais e celulares que interagem com a

---

<sup>xii</sup> Fosforilação é a adição de um grupo fosfato ( $PO_4^{3-}$ ) à molécula da proteína. Esse processo é um importante nos mecanismos de regulação das proteínas.

FEZ1 podem representar moléculas de carga [Assmann *et al.*, 2006]. A região “*coiled-coil*” na posição do C-terminal da FEZ1 pode servir como um domínio promíscuo de interação para um grande número de proteínas, pelo menos 27 das quais realmente mostram também alta probabilidade de possuir regiões com propensão à formação de “*coiled-coil*” nas suas seqüências, incluindo aquelas regiões que foram identificadas como possíveis candidatas para interagir com a FEZ1 [Assmann *et al.*, 2006].

Além da estrutura dimérica obtida por SAXS, a proteína FEZ1 mostra uma conformação longa e estendida, o que está de acordo com a previsão teórica, bem como com as análises espectroscópicas e de proteólise limitada [Lanza *et al.*, 2009]. De modo geral, esses dados sugerem que a FEZ1 possa pertencer ao grupo de proteínas nativamente desenhadas. A representação de Kratky das curvas de SAXS é típica de proteínas com um elevado grau de flexibilidade e alto conteúdo de conformações aleatórias (tipicamente chamados: *random coils*). Cabe ressaltar que a principal função do domínio N-terminal da FEZ1 pode ser o de fornecer uma plataforma para a dimerização. Sabendo da dimerização e com base no arranjo observado no modelo de SAXS, principalmente os da figura 3.6, as duas partes C-terminais do dímero permanecem separados um do outro e são livres para a interação com outras proteínas, que pode ou não representar proteínas de carga nos processos de transporte. Sua conformação dimérica em solução sugere que ela poderia atuar como uma proteína adaptadora<sup>xiii</sup> de transporte bivalente. As duas regiões C-terminais do dímero apontando para fora podem servir como domínios de interação para transporte e/ou fazer contatos com microtúbulos via outras proteínas já identificadas [Lanza *et al.*, 2009]. Além disso, agora sabemos que uma ponte dissulfeto na porção N-terminal tem um papel importante na dimerização da FEZ1, logo, poderíamos especular dizendo que isso pode dar indícios para o desenvolvimento de drogas que desfaçam esse dímero em casos de doenças causadas por super expressão dessa proteína na célula.

Por fim, o fato da proteína FEZ1 sofrer mudanças conformacionais indica que há pequenas regiões estruturadas que, apesar de possivelmente elas interagirem com baixa afinidade, são responsáveis pelas interações específicas de FEZ1 com outros parceiros de interação. De maneira resumida, podemos dizer que a estrutura da proteína FEZ1 é desestruturada, porém não é aleatória como a de uma proteína completamente desenhada (*Random Coil*).

---

<sup>xiii</sup> Proteínas adaptadoras são proteínas que são acessórias para as proteínas principais no trajeto de transformação de sinais dentro da célula.

### **3.3.PROTEÍNA REGULATÓRIA HUMANA KI-1/57 TEM CARACTERÍSTICAS DE PROTEÍNA INTRINSECAMENTE DESESTRUTURADA.**

#### **3.3.1. DESCRIÇÃO DA PROTEÍNA KI-1/57**

A proteína regulatória humana Ki-1/57 (esse nome vem do gene Ki-1/57 que expressa essa proteína) é uma proteína sobre o qual ainda se sabe pouco. Ela foi identificada pela primeira vez em células de linfoma de Hodgkin<sup>xiv</sup> [Hansen *et al.*, 1989; Froese *et al.*, 1987]. Dentro da célula, ela está localizada no citoplasma e no núcleo, onde é frequentemente encontrada em associação com diversas subestruturas nucleares [Rohde *et al.*, 1992]. Há alguns estudos que sugerem que a proteína Ki-1/57 possa estar relacionada ao desenvolvimento de tumores [Bressan *et al.*, 2008], porém, alguns estudos ainda precisam ser realizados para provar isso. Embora o seu exato papel celular ainda seja desconhecido, o perfil de interação proteína-proteína de Ki-1/57 indica que ela possa estar envolvida na regulação da expressão de genes tanto no processo de transcrição quanto por sua influência no metabolismo do RNA [Nery *et al.*, 2006a; Nery *et al.*, 2004]. Dessa forma, análises das propriedades estruturais de Ki-1/57 é um importante passo para esclarecer sua atividade funcional. A predição de baixa quantidade de estrutura secundária regular sugere que Ki-1/57 pertença à classe de proteínas intrinsecamente desestruturadas e experiências espectroscópicas e de análises bioquímicas reforçam essa idéia [Bressan *et al.*, 2008]. Porém, uma análise estrutural mais detalhada ainda é necessária, sendo a técnica de SAXS devidamente apropriada para isso.

Neste estudo, a técnica de SAXS foi utilizada para caracterizar as propriedades estruturais de uma construção do C-terminal da proteína humana Ki-1/57, que se estende na região de aminoácidos de 122-413 dessa proteína. Essa amostra é aqui identificada como Ki-1/57(122-413). Juntamente com previsões teóricas, todas as experiências que foram realizadas suportam a hipótese de que a proteína Ki-1/57 pertence à classe das proteínas nativamente desenoveladas. Os resultados de SAXS revelam uma conformação estendida e flexível para Ki-1/57(122-413), sugerindo a presença de grandes regiões desenoveladas. Uma forma alongada foi também observada nas experiências de cromatografia por exclusão de tamanho e experiências de velocidade de sedimentação por ultracentrifugação analítica, ambas realizadas antes das experiências de SAXS

---

<sup>xiv</sup> A Doença, ou Linfoma de Hodgkin, é uma forma de câncer que se origina nos linfonodos (gânglios) do sistema linfático, que é um conjunto composto por órgãos, tecidos que produzem células responsáveis pela imunidade e vasos que conduzem estas células através do corpo.

[Bressan *et al.*, 2008]. Em resumo, todos os dados que obtivemos apontaram para o fato de que Ki-1/57 possui características de uma proteína intrinsecamente desestruturada (proteína nativamente desenovelada), fornecendo novas direções para explicar e estudar a plasticidade de seu perfil de interação com outras proteínas. Maiores detalhes das experiências bioquímicas e predições teóricas que corroboram com esse trabalho estão descritas em Bressan *et al.* [Bressan *et al.*, 2008]. Nas seções a seguir, será dada especial atenção às experiências de SAXS, que foram da minha responsabilidade, e seus principais resultados.

### 3.3.2. MATERIAIS DE MÉTODOS

#### 3.3.2.1. PREPARAÇÃO DAS AMOSTRAS

A preparação da proteína 6xHis-Ki-1/57(122-413) foi realizada como descrito em Nery *et al.* [Nery *et al.*, 2006] e em [Bressan *et al.*, 2008]. Brevemente, essas produzidas em bactérias e purificadas em dois passos. Primeiro as amostras foram purificadas por cromatografia de afinidade com níquel, seguindo os protocolos convencionais de purificação de proteínas com cauda de histidina em colunas *HiTrap<sup>TM</sup> chelating* (GE Healthcare®, [www.gelifescience.com](http://www.gelifescience.com)). As frações purificadas pela afinidade com Ni<sup>2+</sup> foram acumuladas e dializadas no tampão: 50 mM de Tris-HCL (pH 7,5), 100 mM de NaCl, 10% de glicerol, 5 mM de EDTA. Feito isso, as amostras foram carregadas em uma coluna de troca iônica *Q-Sepharose* (GE Healthcare®). A proteína foi então eluída por um gradiente de 0 a 1 M de NaCl [Bressan *et al.*, 2008]. Essa troca iônica separa as proteínas pelo ponto isoelétrico<sup>xv</sup> de cada uma. Todos os procedimentos foram realizados na presença de inibidores de protease (*Calbiochem*®) a 4°C para prevenir degradação. Essa construção da proteína contém 29 aminoácidos adicionais no N-terminal de Ki-1/57(122-413) que incluem uma cauda de 6xHis (histidina) e outros aminoácidos ligando essa cauda à proteína [Bressan *et al.*, 2008]. Por questão práticas vamos abreviá-la por Ki-1/57(122-413).

---

<sup>xv</sup> O ponto isoelétrico (pI) é o valor pH da solução na qual há equilíbrio entre as cargas negativas e positivas dos grupamentos iônicos de uma proteína, isto é, apresenta carga elétrica líquida igual a zero.

### **3.3.2.2. DESCRIÇÃO DAS EXPERIÊNCIAS DE SAXS COM A PROTEÍNA KI-1/57**

As experiências de espalhamento de raios-X a baixo ângulo foram realizadas na linha de luz D02A-SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas-SP, Brasil). Antes da análise, as amostras de 6xHis-Ki-1/57 (122-413) em tampão 50 mM de Tris-HCl (pH 7,5), 400 mM de NaCl, 15% de glicerol, foram centrifugadas a 356.000 g durante 30 minutos a 4°C para remover eventuais partículas ou agregados. As medidas foram realizadas em um porta-amostra contendo uma cela de 1 mm de espessura com duas janelas de mica [Cavalcanti *et al.*, 2004] e a temperatura foi controlada através da circulação da água para a amostra se manter em  $T = 20^{\circ}\text{C}$ . As experiências foram feitas utilizando um feixe monocromático de raios-X (comprimento de onda de  $\lambda = 1,488 \text{ \AA}$ ) e o espalhamento foi registrado usando um detector MARCCD bidimensional sensível a posição. A distância amostra-detector usada foi de 1.374,4 mm, correspondente ao vetor espalhamento gama de  $0,01 < q < 0,25 \text{ \AA}^{-1}$ , onde  $q$  é o módulo do vetor  $q$  definido por  $q = (4\pi/\lambda)\text{sen}\theta$  (onde  $2\theta$  é o ângulo de espalhamento definido na seção 1.2 do capítulo 1). As medidas foram feitas com duas concentrações diferentes das amostras: 1,53 mg/mL e 2,59 mg/mL. Foram feitas cinco coletas sucessivas de 300 segundos para cada concentração. As curvas resultantes foram cuidadosamente inspecionadas para detectar possíveis danos induzidos por radiação, mas esses efeitos não foram observados. As curvas de espalhamento foram corrigidas pela resposta do detector e normalizadas pela absorção da amostra e pela intensidade integrada do feixe incidente. O espalhamento do tampão foi subtraído da curva de intensidade de cada amostra. Depois de normalizar pela concentração, os dados foram comparados para verificar a existência de efeitos de dependência com a concentração. Este efeito não foi observado. Uma solução com 4,09 mg/mL de albumina bovina (BSA, massa molecular de 66 kDa), no mesmo tampão das amostras, foi utilizada como padrão para determinar a massa molecular de Ki-1/57(122-413) da maneira descrita na seção 1.4.1.1 do capítulo 1 (equação 1.33) dessa tese.

### **3.3.2.3. ANÁLISES DOS DADOS DE SAXS E MODELAGEM DA CONFORMAÇÃO DA PROTEÍNA KI-1/57(122-413)**

Em vista do fato de a proteína Ki-1/57(122-413) apresentar estrutura desordenada, as análises dos dados de SAXS são bastante semelhantes àquelas feitas para a proteína FEZ1 (seção 3.1 do capítulo 3) e seguindo o que foi apresentado no capítulo 1 dessa tese. Os modelos de baixa

resolução foram obtidos usando os métodos de átomos fictícios (seção 1.9.3 do capítulo 1) e o método de resíduos fictícios (seção 1.9.4 do capítulo 1). Dez cálculos foram realizados com cada método e os resultados foram analisados conforme descrito na seção 1.10 do capítulo 1, calculando os valores de discrepância espacial normalizada (NSD). Novamente, calculamos o modelo médio dentro os modelos calculados com o método de átomos fictícios. Também, para os modelos de resíduos fictícios, essa média não pode ser obtida, após os valores de NSD serem avaliados, cinco modelos mais típicos (com menores valores de NSD) foram considerados representativos do conjunto das conformações possíveis da molécula em solução.

#### **3.3.2.4. AS CONFORMAÇÕES DE KI-1/57 EM SOLUÇÃO: MÉTODO DE OTIMIZAÇÃO DE CONJUNTOS DE CONFORMAÇÕES**

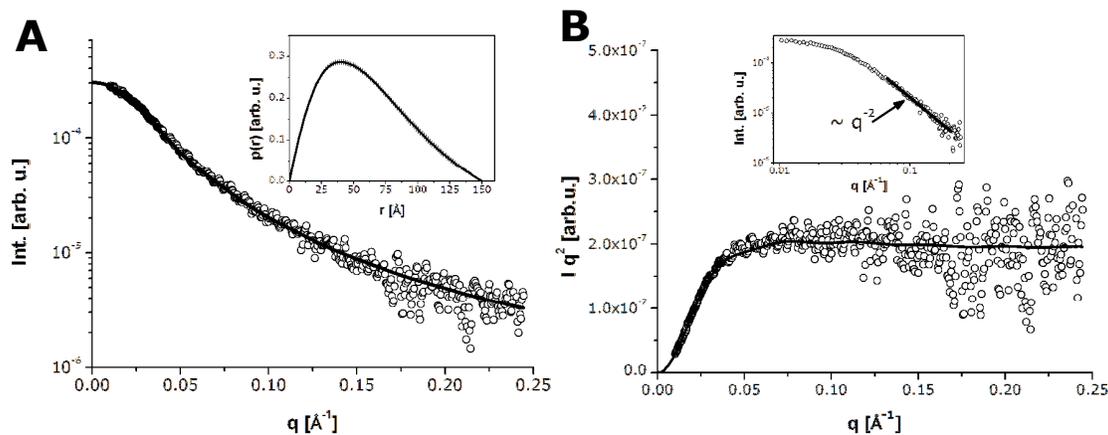
Devido a natureza desestruturada de Ki-1/57(122-413), o conjunto de conformações adotadas pela proteína em solução foi analisado usando o método de otimização de conjuntos de conformações foi aplicado aos dados de SAXS conforme está descrito na seção 1.9.6 do capítulo 1 dessa tese. Neste método, é assumido que os dados experimentais de intensidade de espalhamento possam ser representados por uma média do espalhamento de diferentes conformações da molécula. Como resultado da aplicação desse método, é obtido um conjunto de possíveis conformações da Ki-1/57(122-413) em solução e as distribuições de  $R_g$  e  $D_{máx}$  dessas conformações.

### **3.3.3. RESULTADOS**

#### **3.3.3.1. PRINCIPAIS RESULTADOS DE SAXS DA PROTEÍNA KI-1/57**

A figura 3.11(A) apresenta as curvas experimentais de SAXS corrigidas e normalizadas para Ki-1/57(122-413), juntamente com a curva calculada utilizando o método de transformada de Fourier indireta (seção 1.7 do capítulo 1) que melhor se ajusta à curva experimental. A função  $p(r)$  resultante deste cálculo é mostrada no destaque da figura 3.11(A). O valor obtido para a dimensão máxima ( $D_{máx}$ ) da proteína foi de  $150 \text{ \AA}$ . Os valores de  $R_g$  provenientes das aproximações de Guinier e de Debye, e através da função  $p(r)$  foram, respectivamente,  $47 \pm 2 \text{ \AA}$ ,  $48,1 \pm 0,4 \text{ \AA}$  e  $47,0 \pm 0,6 \text{ \AA}$ . Como se pode notar, todos esses valores estão bastante próximos como esperado. Utilizando os dados de espalhamento da amostra-padrão de BSA, a massa molecular para Ki-1/57 (122-413) foi estimada como sendo de  $\sim 40 \text{ kDa}$ . Este valor está de acordo com o calculado a partir da seqüência de aminoácidos da proteína com a ferramenta ProtParam [Gasteiger *et al.*, 2005] que

foi de 37 kDa, o que indica que a partícula é monomérica em solução. Estes parâmetros globais juntos sugerem uma forma alongada para Ki-1/57(122-413), que é também evidenciada pela correspondente forma assimétrica da função  $p(r)$  (destaque da fig. 3.11(A)). A representação de Kratky ( $q^2 I(q)$  vs.  $q$ ) não mostra um máximo como normalmente ocorre com moléculas compactas e globulares. Em vez disso, aparece um *plateau* para  $q > 0,07 \text{ \AA}^{-1}$  como mostrado na figura 3.9(B). Esta ausência de um máximo no gráfico sugere que Ki-1/57(122-413) possui uma cadeia flexível, sem um núcleo compacto, indicando Ki-1/57(122-413) apresenta uma conformação com grandes regiões desestruturadas em solução. No destaque mostrado na figura 3.9(B), é possível observar que a curva de intensidade na região  $0,07 < q < 0,25 \text{ \AA}^{-1}$  segue um comportamento proporcional a  $q^{-2}$ , como esperado para uma cadeia desenovelada típica de acordo com o modelo Gaussiano (ver capítulo 4.6 do capítulo 1).

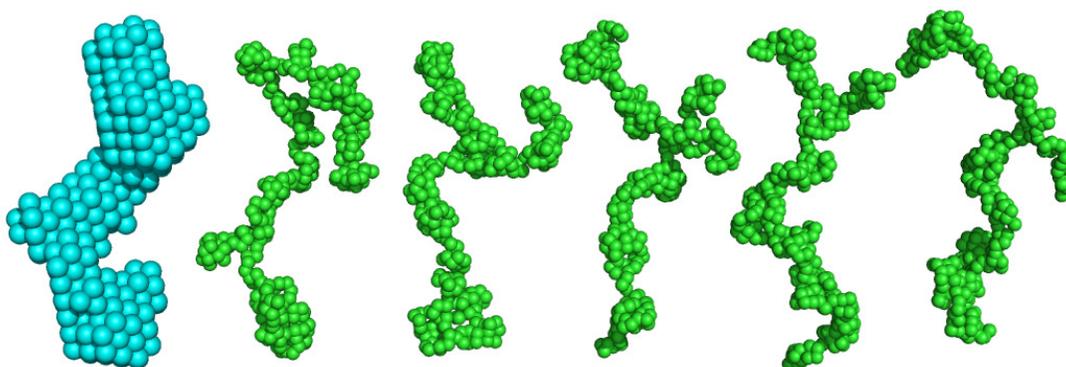


**Figura 3.11** – Resultados experimentais de SAXS obtidos com a proteína Ki-1/57(122-413) em solução. (A) Curva experimental de espalhamento (círculos aberto) e ajuste calculado com o método de transformada de Fourier indireta da intensidade (linha sólida). No destaque está a função de distribuição de pares de distância  $p(r)$  resultado desse ajuste e que mostra uma forma alongada para a proteína com dimensão máxima de  $150 \text{ \AA}$ . (B) Representação de Kratky da curva de intensidade. No destaque está a representação da curva de intensidade em escala *log-log* para mostrar o comportamento com  $1/q^2$  para valores grandes de  $q$ .

### 3.3.3.2. MODELAGEM DA CONFORMAÇÃO DA PROTEÍNA HUMANA KI-1/57(122-413)

Ao utilizar duas abordagens *ab initio* de modelagem tridimensional (átomos fictícios e resíduos fictícios) descritos anteriormente, envelopes moleculares foram obtidos para a proteína Ki-

1/57(122-413) a partir dados de espalhamento a baixos ângulos. Tendo em mente que o procedimento de modelagem não produz uma solução única, dez cálculos independentes foram realizados para cada abordagem. Embora tenham sido observados modelos conformacionais com pequenas diferenças em cada cálculo, todos os modelos apresentaram uma forma alongada típica e recorrente. A Fig. 3.12 mostra o modelo médio obtido com o método de modelagem com átomos fictícios para Ki-1/57(122-413). Cinco modelos representativos obtidos com o método de modelagem com resíduos fictícios são exibidos em linha a direita desse modelo de átomos fictícios médio. As discrepâncias espaciais normalizadas (NSD) para o conjunto de modelos obtidos com o método de átomos fictícios atingiram valores que vão de 0,88 a 1,01 (considerados como aceitáveis para esta modelagem, lembrando que valores de NSD até 1 indicam que modelos são bastantes similares).



**Figura 3.12** – Resultados da modelagem *ab initio* da proteína Ki-1/57(122-413) em solução a partir dos dados de SAXS. Na primeira posição está o modelo médio de baixa resolução obtido com o método de átomos fictícios [azul] seguido por cinco modelos representativos obtidos com o método de resíduos fictícios [verde]. Os modelos são mostrados usando o programa PyMOL [DeLano, 2002].

Além disso, uma vez que a proteína é desenovelada e altamente hidratada, o volume excluído (ver seção 1.4.5 do capítulo 1) calculado para este modelo é superestimado ( $1,38 \times 10^5 \text{ \AA}^3$ ). A baixa resolução dos modelos dessa representação só permite dizer que a conformação dessa proteína é alongada em solução, dando alguns poucos detalhes estruturais. Ao utilizar a abordagem do método de modelagem com resíduos fictícios, o conjunto de modelos obtidos teve valores de NSD variando de 2,32 a 2,43 (considerado bastante elevado, mas aceitáveis devido à abordagem tipo cadeia do método). Todavia, este tipo de reconstrução dá uma imagem mais realista das conformações possíveis para esse tipo de proteína que é desestruturada em solução. Uma

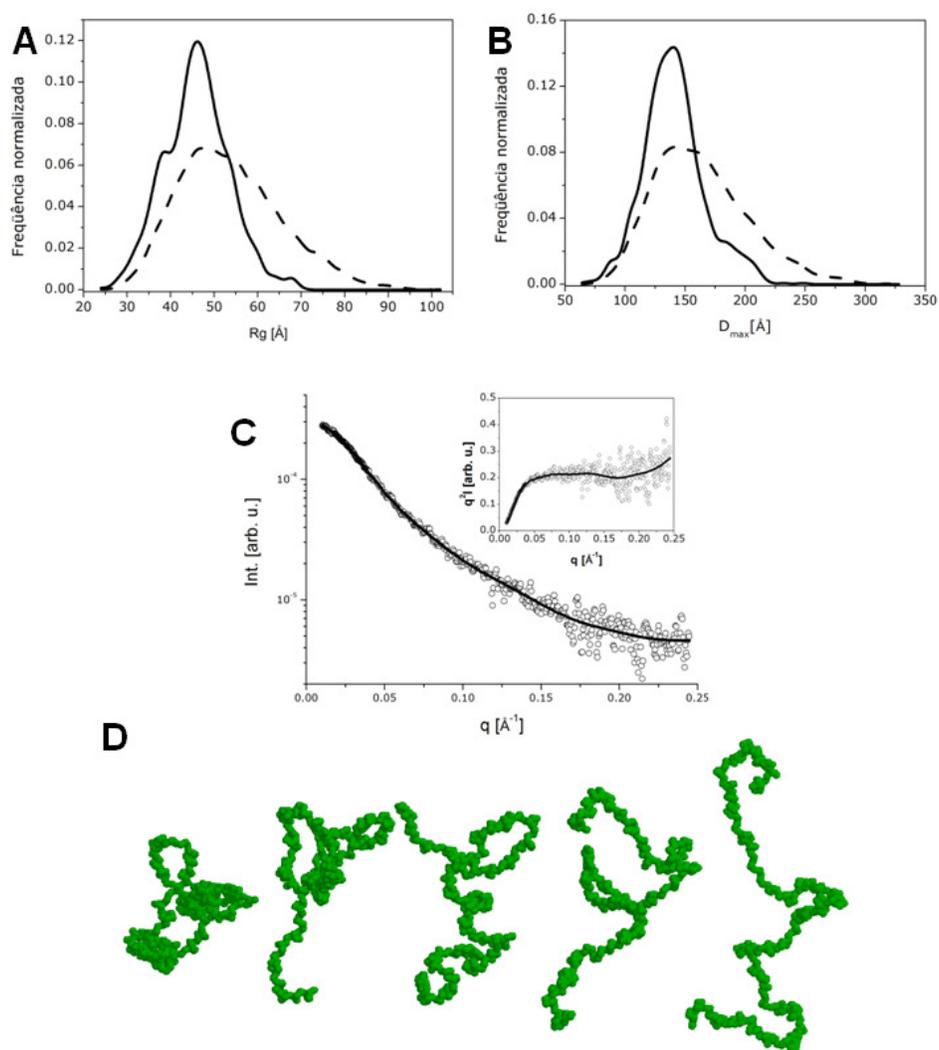
conformação média não pode ser calculada para esses modelos de resíduos fictícios devido ao seu arranjo tipo cadeia, mas pode mapear o espaço conformacional ocupado pela proteína de forma mais adequada. A resolução dos modelos não permite uma determinação unívoca da posição espacial dos elementos da estrutura secundária, mas retrata a forma global das conformações mais freqüentes adotadas pelas moléculas em solução. Os modelos obtidos para Ki-1/57(122-413) parecem explorar um grande espaço conformacional em solução. Estes resultados seguem a linha das características gerais obtidas para outras proteínas parcialmente desenoveladas [Silva *et al.*, 2010a; Lanza *et al.*, 2009; Bressan *et al.*, 2008; Tidow *et al.*, 2007; Bernadó *et al.*, 2007; Moncoq *et al.*, 2004].

### 3.3.3.3. MÉTODO DE OTIMIZAÇÃO DE CONJUNTO DE CONFORMAÇÕES

Os resultados obtidos com o método de otimização de conjuntos de conformações são mostrados na Figura 3.13. As distribuições de raio de giro ( $R_g$ ) e de máxima dimensão ( $D_{máx}$ ) da das diferentes conformações de Ki-1/57 em solução calculadas com o método estão apresentadas na Figura 3.13(A) e 3.13(B), respectivamente. Nessas figuras, estão mostradas juntas, em cada caso, a distribuição inicial das conformações criadas cobrindo o espaço de conformações e a distribuição final resultante após a otimização realizada para selecionar o melhor conjunto de conformações. Essa distribuição final de conformações para Ki-1/57(122-413) em solução representa o conjunto de modelos cujo espalhamento médio melhor ajusta os dados experimentais de espalhamento como mostra a figura 3.13(C). Os valores médios de  $R_g$  e  $D_{máx}$  obtidos a partir dessas distribuições foram  $48\text{Å}$  e  $145\text{Å}$ , respectivamente. Como esperado, os valores médios desses parâmetros calculados aqui estão em ótima concordância com aqueles obtidos das análises prévias e típicas das curvas de SAXS (ver seção 3.2.3.1). Essas duas distribuições mostraram um comportamento característico para proteínas desenoveladas e flexíveis, porque somente um máximo é observado ao redor do valor médio em cada caso e as distribuições finais de  $R_g$  e  $D_{máx}$  das diferentes conformações da proteínas não estão muito deslocadas das distribuições aleatórias iniciais.

Com respeito às conformações resultados, mostrar todas que foram obtidas no conjunto otimizado de conformações seria complicado, dado o grande número de conformações do conjunto. Por isso, nós decidimos apresentar ao menos cinco delas na Figura 3.11(D). Essas conformações mostradas foram selecionadas com base na distribuição de  $R_g$ , selecionando duas das extremidades (mais baixo e mais alto valor de  $R_g$ ), uma do valor médio e mais duas com valores intermediários de  $R_g$ . Como observamos nessas conformações, aquela com  $R_g$  mais próximo do valor médio de  $R_g$

(mais próxima do máximo da distribuição) apresenta bastante semelhança com os modelos obtidos com o método de resíduos fictícios, indicando que são modelos coerentes.



**Figura 3.13** – Resultados da aplicação do método de otimização de conjunto de conformações aos dados de SAXS da proteína Ki-1/57(122-413) em solução. (A) Distribuição de raios de giro  $R_g$  inicial (linha pontilhada) e final (linha sólida) do conjunto de conformações. (B) Distribuição de dimensões máximas  $D_{max}$  inicial (linha pontilhada) e final (linha sólida) do conjunto de conformações. (C) Ajuste calculado para a curva de intensidade de espalhamento experimental. *Destaque:* Representação de Kratky. (D) Cinco modelos para conformações representativas do conjunto que apresentou melhor ajuste aos dados experimentais de espalhamento.

### 3.3.4. DISCUSSÃO SOBRE OS RESULTADOS OBTIDOS COM A PROTEÍNA KI-1/57

Neste trabalho nós fornecemos novas informações referentes às propriedades conformacionais da região C-terminal de Ki-1/57 em solução. Os resultados das análises dos dados de SAXS obtidos para a construção C-terminal de Ki-1/57, englobando os aminoácidos 122-413, indicaram que Ki-1/57(122-413) ocupa um espaço conformacional estendido e possui uma alta flexibilidade intramolecular em solução. A técnica de SAXS também forneceu parâmetros dimensionais e de forma, além de um baixo grau de compacidade, que são típicos de uma molécula altamente flexível e alongada. Logo, Ki-1/57(122-413) não adota uma estrutura globular em solução e possui características de proteínas intrinsecamente desestruturadas. Os vários métodos de modelagem tridimensional *ab initio* e o método de otimização de conjunto de conformações contribuem para o entendimento das possíveis conformações adotadas pela proteína em solução. Essas características estruturais estão em concordância com os dados funcionais, disponíveis até o momento, dessa proteína.

Em vista disso, os dados funcionais disponíveis até o momento para Ki-1/57 sugerem um papel complexo para essa proteína na expressão e regulação gênica. Da mesma forma que as proteínas que são conhecidamente centros de conexão (*hubs*), Ki-1/57 tem a habilidade de se ligar a vários parceiros envolvidos na sinalização celular, regulação transcricional e metabolismo do RNA [Bressan *et al.*, 2009; Bressan *et al.*, 2008]. Recentemente, interações específicas de Ki-1/57 com seqüências de RNA ricos em uracila (U) foram descobertas [Bressan *et al.*, 2009]. Portanto, essas observações encaixam-se muito bem com as características de proteínas intrinsecamente desestruturadas e sugerem a necessidade de uma estrutura bastante flexível para Ki-1/57 que permita essa plasticidade funcional.

Longas regiões desordenadas têm sido encontradas em várias proteínas humanas associadas a doenças. Isso inclui diabetes, doenças auto-imunes, doenças neurodegenerativas, doenças cardiovasculares e câncer [Cheng *et al.*, 2006; Iakoucheva *et al.*, 2002] e reflete a importância funcional da classe de proteínas nativamente desoveladas. Recentemente, pesquisadores demonstraram a associação funcional e interação de Ki-1/57 com proteína de supressão tumoral p53 [Nery *et al.*, 2006b], sugerindo o envolvimento dessa interação no controle do crescimento celular. Isso aponta para a importância de um melhor entendimento das características estruturais de Ki-1/57, assim como apresentamos aqui, e também os detalhes da sua interação e relação funcional com essa proteína de supressão tumoral.

Juntando os resultados de SAXS, o perfil de interações proteína-proteína de Ki-1/57, as predições teóricas a partir da seqüência de aminoácidos que predizem várias regiões desordenadas na proteína e os outros resultados bioquímicos e biofísicos publicados em Bressan *et al.* [Bressan *et al.*, 2008], tudo nos induz a concluir que ki-1/57 é uma proteína intrinsecamente desestruturada.

# 4

## Proteínas flexíveis com múltiplos domínios

### 4.1. INTRODUÇÃO

A visão clássica do enovelamento de uma proteína considera que ela se enovela rapidamente até chegar a somente um estado nativo cuja estrutura é única e rígida. Porém, com o avanço da biologia molecular, principalmente com o avanço das técnicas de genômica e proteômica, foram encontradas proteínas com pouca estrutura ou com domínios estruturados, porém com partes flexíveis no estado nativo, o que fez com que essa antiga visão de enovelamento de proteínas fosse revista. Já discutimos no capítulo anterior sobre as proteínas intrinsecamente desestruturadas. Agora é a vez de falarmos de proteínas com múltiplos domínios estruturados ligados por regiões flexíveis e desestruturadas, o que fornece uma certa flexibilidade a esse tipo de proteína (em grau menor que no caso das proteínas nativamente desenoveladas). Essas proteínas são classificadas como proteínas flexíveis com múltiplos domínios.

A flexibilidade proporcionada por essas regiões flexíveis, tipicamente desestruturadas, está relacionada com as funções dessas proteínas. Essa característica é geralmente observada em proteínas que interagem com DNA ou RNA [Petoukhov *et al.*, 2006] e em algumas chaperonas, que são proteínas que auxiliam o enovelamento de outras proteínas [Krukenberg *et al.*, 2008]. Além

disso, cada domínio desse tipo de proteína pode realizar sua função independentemente ou com o auxílio de um domínio vizinho, sendo que a flexibilidade da região conectando esses domínios vizinhos tem um papel importante nesse caso.

Do ponto de vista da determinação de estruturas de alta resolução, essas proteínas constituem um verdadeiro desafio quando as técnicas de ressonância magnética nuclear (RMN) ou métodos cristalográficos são empregados. A existência de regiões flexíveis nas moléculas é um grande obstáculo. Sendo assim, mais uma vez a técnica de Espalhamento de Raios-X a Baixos Ângulos (SAXS) tem um papel importante nesse estudo. A técnica de SAXS fornece dados de baixa resolução que, em muitos casos, permitem a investigação de detalhes importantes no problema de estrutura-função das proteínas. SAXS ainda pode revelar de maneira eficiente a conformação espacial e a organização dos domínios dessas proteínas em solução, incluindo suas partes desestruturadas e/ou desconhecidas. Além disso, a característica mais potencialmente útil dessa técnica é a possibilidade de realizar modelagem de baixa resolução da proteína. Métodos de cálculos *ab initio*, que usam as rotinas bem conhecidas de modelagem com átomos fictícios ou resíduos fictícios (ver seções 9.3 e 9.4 do capítulo 1), permitem o cálculo de modelos de baixa resolução para o envelope molecular dessas proteínas em solução. Ainda há um método que permite combinar esses cálculos *ab initio* com modelagem de corpo rígido, no qual o modelo é restaurado dos dados de SAXS usando estruturas de alta resolução dos domínios individuais (resolvidas por outras técnicas) e simulando as regiões conectoras flexíveis como uma cadeia de resíduos fictícios (seção 9.5 do capítulo 1).

Algumas proteínas que se encaixam nessa classe são: duas proteínas da família das HSP40 (chaperonas), chamadas Sis1 (devido ao gene *S. cerevisiae* *Sis1* que a expressa) e Ydj1 (devido ao gene *Yeast DnaJ* que a expressa) [Ramos *et al.*, 2008; Sha *et al.*, 2000; Fan *et al.*, 2004], e a proteína ribonucléica heterogênea tipo Q (hnRNP-Q, que é a abreviação para o termo em inglês *heterogeneous Ribonucleoprotein Q*) [Quaresma *et al.*, 2009]. É possível encontrar vários trabalhos sobre as HSPs (abreviação para o termo em inglês *Heat Shock Proteins*) na literatura [Ramos *et al.*, 2008; Sha *et al.*, 2000; Fan *et al.*, 2004], porém pouco se conhece sobre suas estruturas. As HSPs são chaperonas que auxiliam outras proteínas no processo de enovelamento. Já sobre a hnRNP-Q, ainda não há trabalhos estruturais sobre ela; só funcionais [Quaresma *et al.*, 2009; Passos *et al.*, 2006]. A característica mais importante da hnRNP-Q é sua interação com RNA. Conseqüentemente ela está envolvida num amplo leque de funções do RNA. Sendo assim, para entender melhor as funções tanto da hnRNP-Q quanto das chaperonas Sis1 e Ydj1, é importante caracterizá-la estruturalmente, e isso é possível através da técnica de SAXS.

Neste trabalho descrevemos nossos estudos estruturais de alguns mutantes por deleção de domínios da Hsp40 Sis1 e da Hsp40 Ydj1. A atividade funcional desses mutantes ainda está sendo estudada, mas há indícios de que são funcionais [Silva, *et al.*, 2010b]. Além disso, também descrevemos nossos estudos estruturais com a proteína hnRNP-Q\_ΔRGGbox (o sufixo indica a deleção de uma região chamada *RGG Box*), contribuindo para o entendimento da sua interação com o RNA. Essas caracterizações estruturais foram feitas utilizando a técnica de SAXS com o objetivo de obter informações conformacionais sobre essas proteínas em solução e de entender a organização dos múltiplos domínios em sua estrutura, para isso, utilizando diferentes métodos de modelagem para construir a melhor representação tridimensional de seus envelopes moleculares.

## **4.2. PROTEÍNAS DE CHOQUE TÉRMICO (HSP40) SIS1 E YDJ1: CHAPERONAS MOLECULARES**

### **4.2.1. DESCRIÇÃO DAS PROTEÍNAS DE CHOQUE TÉRMICO DE 40KDA (HSP40)**

As chaperonas são, em biologia, proteínas cuja função é auxiliar outras proteínas a alcançar o enovelamento adequado. Muitas chaperonas são do tipo proteínas de choque térmico ou mais conhecidas pelo termo em inglês *Heat Shock Protein* (HSP), isto é, proteínas que atuam em resposta às temperaturas elevadas ou outros estresses celulares [Fink, 1999; Mayer & Bukau, 2005]. A razão para esse comportamento é que o enovelamento de proteínas é severamente afetado pelas altas temperaturas da célula e, portanto, as chaperonas da família das HSPs ajudam a reparar os possíveis danos causados pelo mau-enovelamento (“*misfolding*”) das proteínas.

Há várias famílias de HSPs, cada uma atuando de maneira diferente no auxílio ao processo de enovelamento. Uma dessas famílias é a HSP40, constituída pelas proteínas de choque térmico com massa molecular em torno de 40 kDa. Chaperonas da família HSP40 desempenham importantes papéis nas células trabalhando junto com as chaperonas moleculares da família HSP70 (proteínas de choque térmico com massa molecular em torno de 70 kDa) para promover o enovelamento, organização, deslocamento e degradação de proteínas [Hartl & Hayer-Hartl, 2002; Bukau, 1998; Hartl, 1996]. Devido ao alto peso molecular e flexibilidade intramolecular das HSP40, a técnica de SAXS se torna uma valiosa ferramenta para o estudo das conformações e da estrutura dessas proteínas. Inclusive, é uma técnica singular para estudar os complexos formados entre as próprias HSPs ou delas com outras proteínas que necessitam de sua ajuda para atingir o

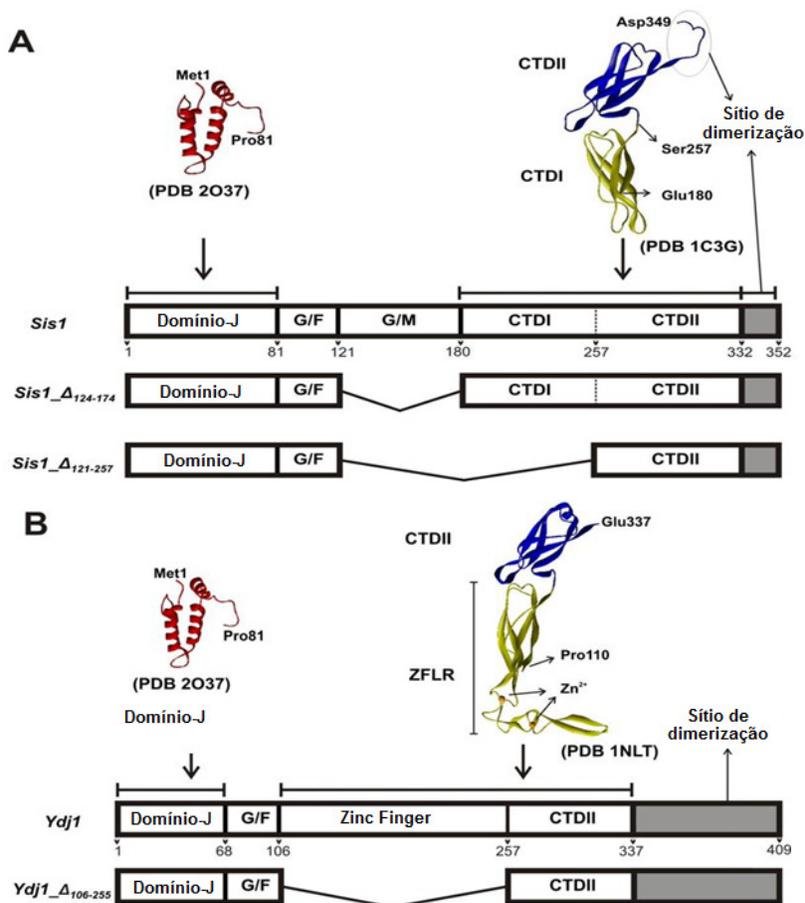
enovelamento adequado. Muitos trabalhos estão sendo feitos nos estudos das chaperonas e co-chaperonas utilizando várias técnicas bioquímicas e biofísicas [Fan *et al.*, 2004; Song & Masison, 2005; Carrigan *et al.*, 2006] e alguns deles inclusive usando a técnica de SAXS [Silva, *et al.*, 2010b; Borges *et al.*, 2005].

Os membros da família HSP40 tipicamente contêm um domínio J, que regula a ligação dependente de ATP dos peptídeos à HSP70 [Bukau, 1998; Wall *et al.*, 1994]. O nome domínio J vem do fato de que os 73 aminoácidos desse domínio conservam uma similaridade com a os 73 aminoácidos iniciais da proteína de *E. coli* DnaJ. Esses domínios J são conectados a um fragmento de ligação a peptídeo na região C-terminal via uma região rica em glicina e fenilalanina (região G/F). Entretanto, há dois subtipos principais de HSP40: o tipo I que contém um motivo estrutural de ligação a zinco, ou dedo-de-zinco<sup>xvi</sup> (do inglês *Zinc-finger*) entre o domínio J e o domínio de ligação a peptídeo, enquanto que o tipo II não apresenta essa região [Banecki *et al.*, 1996; Caplan & Douglas, 1991]. As diferenças estruturais entre as HSP40s tipo I e Tipo II podem ser consideradas para explicar as diferenças nas suas atividades de chaperona molecular. Para entender essas diferenças entre o tipo I e o tipo II, mais estudos estruturais precisam ser feitos com essas proteínas. Alguns estudos estruturais já realizados, junto com as análises termodinâmicas, mostraram que ambas HSP40 tipo I e tipo II forma dímeros em solução [Langer *et al.*, 1992]. A formação do dímero faz um papel crítico na atividade dessas proteínas, porque o rompimento dos motivos de dimerização das HSP40s de ambos os tipos I e II resulta em defeitos severos nas funções de chaperona [Sha *et al.*, 2000; Li *et al.*, 2003].

Para melhor compreender o papel dos múltiplos domínios nós realizamos experiências de SAXS com algumas construções das proteínas Sis1 (HSP40 do tipo II) e Ydj1 (HSP40 do tipo I), duas proteínas da família das HSP40s. *Sis1* é uma chaperona modelo dentre as HSP40s tipo II. Essa proteína contém 352 resíduos arranjados em um domínio-J, uma região central desordenada (dividida em duas regiões: uma rica em glicina e fenilalanina (G/F) e outra rica em glicina e metionina (G/M)) e dois subdomínios na região C-terminal, que são o domínio I da região C-terminal (CTDI) e o domínio II da região C-terminal (CTDII). Já a Ydj1, essa é uma importante HSP40 tipo I que contém 409 resíduos arranjados em um domínio-J, uma região central desordenada (rica em glicina e fenilalanina (G/F)), um domínio dedo-de-zinco (*zinc-finger*) e um subdomínio (CTDII) na região C-terminal.

---

<sup>xvi</sup> Dedo-de-zinco ou Zinc fingers pequenos motivos estruturais de uma proteína que podem coordenar um ou mais íons de zinco para ajudar no seu enovelamento da proteína.



**Figura 4.1** – Representação esquemática das seqüências de aminoácidos e organização dos domínios das construções das HSP40s e mutantes. (A) Organização dos domínios da proteína Sis1 e suas mutantes: Sis1\_Δ124-174 e Sis1\_Δ121-257. (B) Organização dos domínios da mutante Ydj1\_Δ106-255. Na parte de cima das representações estão mostradas as estruturas de alta resolução disponíveis para os domínios. *G/F*: região rica em glicina e fenilalanina; *G/M*: região rica em glicina e metionina; *Zinc Finger*: domínio dedo-de-zinco; *CTDI*: Domínio C-terminal I; *CTDII*: Domínio C-terminal II. O sítio de dimerização é representação por uma caixa cinza.

Para entender a relação entre suas estruturas e funções, foram cortados alguns domínios dessas proteínas. Essas construções dão origem ao que chamaremos aqui de mutantes por deleção de domínios. Com o intuito de entender o papel dos domínios de Sis1 na sua estrutura e função, duas construções foram produzidas. A primeira foi construída retirando a região *G/M* da Sis1 e identificaremos esse mutante como Sis1\_Δ124-174. A segunda foi construída retirando tanto a região *G/M* quanto o subdomínio CTDI, dando origem à Sis1\_Δ121-257. Essas construções estão apresentadas na figura 4.1(A). Novamente, para entender o papel de cada região na função e

estrutura da proteína, nós também estudamos uma construção que chamamos de Ydj1 $_{\Delta 106-255}$ , na qual o domínio dedo-de-zinco foi retirado da proteína Ydj1 como esquematizado na Figura 4.1(B).

## 4.2.2. MATERIAIS E MÉTODOS

### 4.2.2.1. PREPARAÇÃO DAS AMOSTRAS

As amostras de proteínas foram todas expressas em bactérias *E. coli* e purificadas como descrito em Fan *et al.* [Fan *et al.*, 2004] e em Silva *et al.* [Silva *et al.*, 2010]. De maneira breve, a purificações de Sis1, Sis1 $_{\Delta 124-174}$ , e Sis1 $_{\Delta 121-257}$  foram realizadas por cromatografia de troca catiônica e Ydj1 $_{\Delta 106-255}$  foi purificada por cromatografia de troca aniônica e por cromatografia de exclusão de tamanhos [Silva *et al.*, 2010]. Todos esses processos foram realizados seguindo protocolos padronizados de purificação. A estabilidade e pureza das amostras foram testadas e verificadas por gel de eletroforese (gel SDS) [Silva *et al.*, 2010]. As figuras 4.1(A) e 4.1(B) mostram representações esquemáticas das amostras analisadas neste trabalho.

### 4.2.2.2. DESCRIÇÃO DAS EXPERIÊNCIAS DE SAXS

Todas as amostras aqui analisadas (Sis1, Sis1 $_{\Delta 124-174}$ , Sis1 $_{\Delta 121-257}$  e Ydj1 $_{\Delta 106-255}$ ) foram previamente submetidas a análises de DLS e experiências de ultracentrifugação analítica para testar a monodispersidade das soluções antes das experiências de SAXS [Silva *et al.*, 2010]. Feito isso, as amostras foram levadas para a linha de luz D02A-SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, SP, Brasil), onde as experiências de SAXS foram feitas. O espalhamento de raios-X foi registrado usando um detector bidimensional MARCCD sensível a posição. As medidas foram realizadas com um feixe de raios-X monocromático (comprimento de onda de  $\lambda = 1,488\text{\AA}$ ) e um distância amostra-detector de  $1374,4\text{ mm}$ , resultando em um intervalo do módulo de vetor de espalhamento de  $0,01\text{\AA}^{-1} < q < 0,25\text{\AA}^{-1}$ , onde  $q$  é o módulo do vetor  $q$  definido por  $q = (4\pi/\lambda)\text{sen}\theta$  (onde  $2\theta$  é o ângulo de espalhamento como na seção 1.2 do capítulo 1). As amostras foram inseridas em um porta-amostra de  $1\text{ mm}$  de espessura com janelas de mica [Cavalcanti *et al.*, 2004] e temperatura constante ( $T=24^{\circ}\text{C}$ ) controlada através de circulação de água no porta-amostra com a ajuda de um banho térmico monitorado por um termopar. As curvas de espalhamento foram coletadas em duas concentrações para cada amostra:  $7,1\text{ mg/mL}$  e  $3,1\text{ mg/mL}$  para Sis1;  $6,8\text{ mg/mL}$  e  $4,2\text{ mg/mL}$  para Sis1 $_{\Delta 124-174}$ ;  $5,2\text{ mg/mL}$  e  $3,9\text{ mg/mL}$  para Sis1 $_{\Delta 121-}$

<sup>257</sup>; e 5,5 mg/mL e 2,75 mg/mL para Ydj1\_Δ<sub>106-255</sub>. Todas as amostras de proteínas estavam em tampão composto por 25 mM de Tris-HCl (pH 7,5) e 500 mM de NaCl. Foram feitas três coletas sucessivas de 300 s cada para cada amostra e uma mais, também de 300 s, para o tampão. Todas as curvas de espalhamento foram individualmente corrigidas pela resposta do detector e normalizadas pela intensidade integrada do feixe incidente e pela absorção das amostras como descrito na seção 1.6 do capítulo 1. Feito isso, a curva de espalhamento do tampão foi então subtraída do espalhamento da amostra correspondente. As curvas resultantes foram cuidadosamente inspecionadas para verificar a formação de bolhas na solução e a existência de possíveis danos induzidos por radiação nas proteínas, mas tais efeitos não foram observados. Após normalizar as curvas pelas respectivas concentrações, as curvas de intensidade de espalhamento dessas amostras foram analisadas para investigar se havia dependência com a concentração, porém esse efeito não foi observado. Por fim, uma solução de albumina bovina (BSA, massa molecular de 66 kDa) no mesmo tampão das amostras, em uma concentração de 5,6 mg/mL, foi preparada e usada como amostra padrão para estimar as massas moleculares da proteína e seus fragmentos da maneira descrita na seção 1.4.7 do capítulo 1 (equação 1.33) dessa tese. As análises dos dados de SAXS (transformada de Fourier, grau de compacidade e obtenção de parâmetros dimensionais) foram feitas conforme descritas no capítulo 1.

#### **4.2.2.3. MODELAGEM AB INITIO BASEADA NOS DADOS DE SAXS**

Cálculos *ab initio* baseados nos dados de SAXS foram realizados para obter modelos de baixa resolução para as conformações das amostras: Sis1, Sis1\_Δ<sub>124-174</sub>, Sis1\_Δ<sub>121-257</sub> e Ydj1\_Δ<sub>106-255</sub>. Nós utilizamos duas abordagens diferentes de modelagem: o método de átomos fictícios (seção 1.9.3 do capítulo 1) e o método de resíduos fictícios (seção 1.9.4 do capítulo 1). Os resultados de dez cálculos independentes com cada método e cada amostra foram analisados e comparados usando a metodologia descrita na seção 1.10 do capítulo 1. Seleccionamos o modelo mais típico (menor NSD) de cada conjunto e cada método para representar a proteína.

#### **4.2.2.4. MODELAGEM DA TOPOLOGIA DA ORGANIZAÇÃO DE DOMÍNIOS NA CONFORMAÇÃO DAS PROTEÍNAS**

A organização dos múltiplos domínios na conformação de cada amostra aqui analisada foi examinada por modelagem de corpo rígido a partir dos dados de SAXS. Essa abordagem emprega o protocolo de recozimento simulado (*simulated annealing*, seção 1.9.2 do capítulo 1) para encontrar

a posição e orientação ótimas de uma estrutura de alta resolução (domínios com estruturas resolvidas) e a conformação mais provável das regiões desconhecidas de tal modo a ajustar a curva de intensidade de espalhamento experimental da proteína usando uma cadeia de resíduos fictícios (ver seção 1.9.5 do capítulo 1). Para construir o modelo da proteína Sis1 e suas mutantes Sis1\_Δ<sub>124-174</sub>, Sis1\_Δ<sub>121-257</sub>, nós usamos a estrutura de alta resolução dos domínios de ligação a peptídeo (que inclui o sítio de dimerização e domínios CTDI e CTDII como mostrados na figura 4.1(A)) da região C-terminal de Sis1 e a estrutura de seu domínio-J, identificadas no PDB sob os códigos *IC3G* e *2O37*, respectivamente. Já para compor o modelo de Ydj1\_Δ<sub>106-255</sub>, nós usamos a estrutura de alta resolução dos domínios CTDII, dedo-de-zinco e sítio de dimerização, todos cristalizados juntos e depositados no PDB sob o código *1XAO*. Para o modelo de Ydj1\_Δ<sub>106-255</sub> ainda usamos a estrutura do domínio-J da Sis1 (código PDB *2O37*), porém cortamos os últimos 13 resíduos dessa estrutura, pois apresentam pequenas diferenças nessa região em relação à sequência de aminoácidos da Ydj1. Novamente, vários cálculos independentes foram realizados para cada amostra. Nós aplicamos uma restrição de simetria tipo *P2* (simetria em torno de um eixo de rotação de 180°) devido ao fato de que sabíamos *a priori* que todas as amostras formavam dímeros em solução. Após analisar e comparar os resultados conforme a metodologia descrita na seção 1.10 do capítulo 1, selecionamos o modelo mais típico para cada proteína.

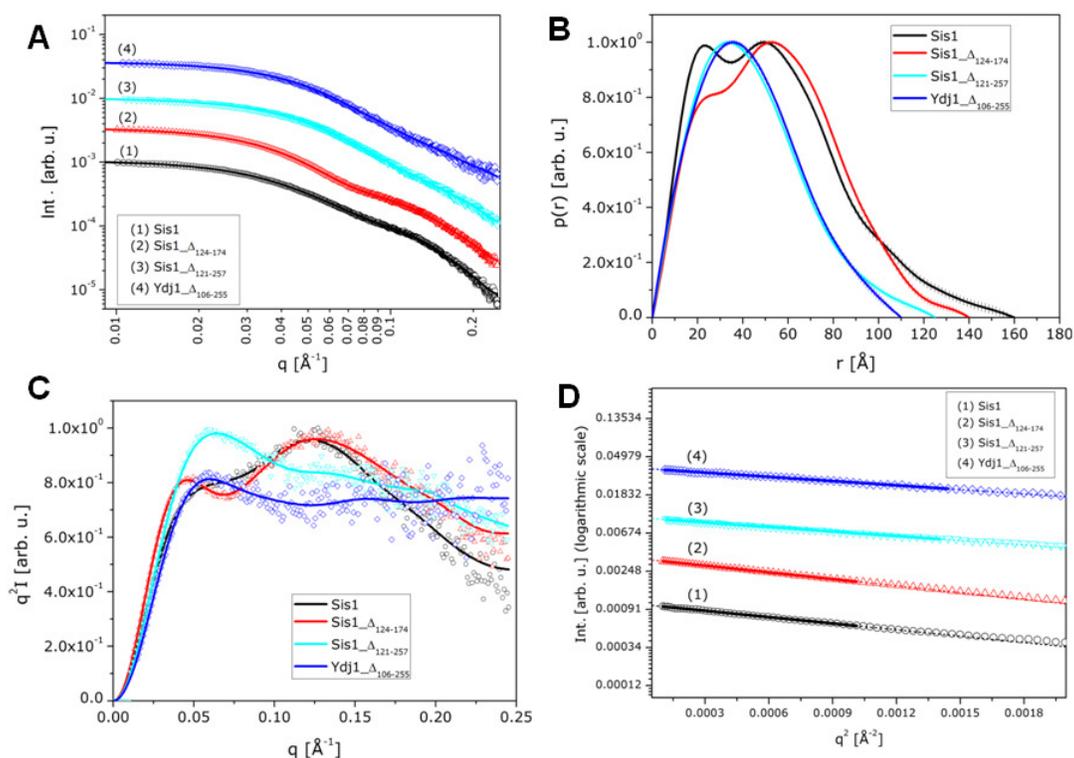
### 4.2.3. RESULTADOS

#### 4.2.3.1. RESULTADOS DAS EXPERIÊNCIAS DE SAXS

As curvas experimentais de SAXS corrigidas e normalizadas para as proteínas Sis1, Sis1\_Δ<sub>124-174</sub>, Sis1\_Δ<sub>121-257</sub> e Ydj1\_Δ<sub>106-255</sub> estão mostradas na figura 4.2(A), junto com o ajuste (linha sólida) calculado pelo método da transformada de Fourier indireta (seção 1.7 do capítulo 1). As correspondentes funções  $p(r)$  são mostradas na figura 4.2(B), todas indicando uma forma alongada para todas as proteínas aqui analisadas em solução, além de indicar características de proteínas com múltiplos domínios, realçadas nas funções  $p(r)$  das proteínas Sis1\_Δ<sub>124-174</sub> e Ydj1\_Δ<sub>106-255</sub>. Além disso, as representações de Kratky apresentadas na figura 4.2(C) mostraram que essas proteínas são bastante flexíveis em solução (seção 1.5 do capítulo 1), o que fornece um indício para a dificuldade de cristalização dessas proteínas e mutantes. Na figura 4.2(D) nós mostramos a linearidade da região de validade da aproximação de Guinier juntamente com as retas obtidas por regressão linear. Dessa vez, mesmo as proteínas sendo bastante flexíveis, o fato de

terem múltiplos domínios estruturados faz com a região de validade de aproximação de Guinier não seja tão estreita quanto aquela para as proteínas intrinsecamente desestruturadas. Sendo assim, a linearidade dessa região para cada amostra confirma a monodispersidade das soluções.

Os parâmetros dimensionais como a máxima dimensão ( $D_{máx}$ ) da molécula, os valores de raio de giro ( $R_g$ ) obtidos pela aproximação de Guinier (seção 1.4.2 do capítulo 1) e pela função  $p(r)$  (equação 1.20 da seção 1.4.3 do capítulo 1), assim como os valores de massa molecular estimados para as proteínas em solução estão apresentados na tabela 4.1. Como podemos notar, os valores de  $R_g$  obtidos pelas duas abordagens estão em muito boa concordância. Para conhecer o estado oligomérico das amostras, nós comparamos os valores de massa molecular estimados dos dados de SAXS com aqueles calculados (com a ferramenta ProtParam [Gasteiger *et al.*, 2005]) a partir das seqüências de aminoácidos das proteínas e os resultados também estão na tabela 4.1.



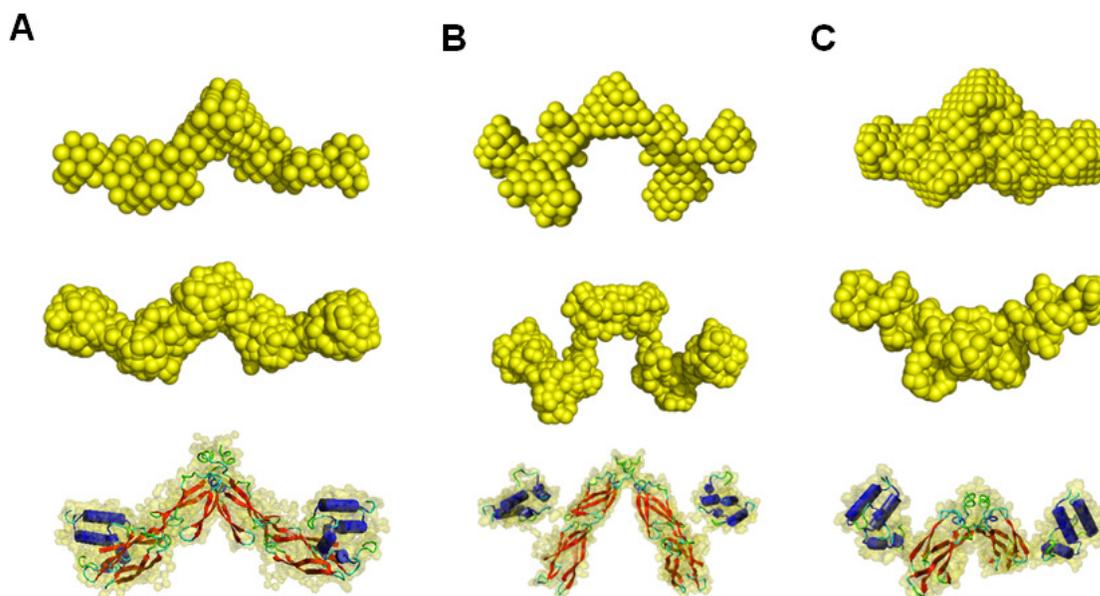
**Figura 4.2** – Resultados de SAXS para as proteínas Sis1, Sis1\_Δ<sub>124-174</sub>, Sis1\_Δ<sub>121-257</sub> e Ydj1\_Δ<sub>106-255</sub>. (A) Curvas de intensidade de espalhamento de todas as proteínas. (B) Funções  $p(r)$  correspondentes. (C) Respectivas representações de Kratky mostrando a alta flexibilidade das proteínas. (D) Região de validade da aproximação de Guinier das curvas de intensidade de espalhamento e as respectivas retas obtidas por regressão linear.

**Tabela 4.1** – Parâmetros dimensionais obtidos dos dados de SAXS para as proteínas e mutantes de HSP40: *Sis1*, *Sis1* $_{\Delta 124-174}$ , *Sis1* $_{\Delta 121-257}$ , *Ydj1* $_{\Delta 106-255}$ .

Proteína	C (mg/mL)	R <sub>g</sub> (Å) (Guinier)	R <sub>g</sub> (Å) $\langle p(r) \rangle$	D <sub>máx</sub> (Å)	M (kDa) (SAXS)	M (kDa) (calc.)	Estado oligom.
<i>Sis1</i>	7,1 e 3,1	42 ± 1	43,3 ± 0,5	160	~71	37,59	Dímero
<i>Sis1</i> $_{\Delta 124-174}$	6,8 e 4,2	42 ± 1	42,6 ± 0,3	140	~69	33,06	Dímero
<i>Sis1</i> $_{\Delta 121-257}$	5,2 e 3,9	35 ± 1	35,2 ± 0,3	125	~42	23,40	Dímero
<i>Ydj1</i> $_{\Delta 106-255}$	5,5 e 2,75	33 ± 1	34,2 ± 0,3	110	~47	28,06	Dímero

#### 4.2.3.2. RESULTADOS DA MODELAGEM *AB INITIO* DAS CONFORMAÇÕES DAS PROTEÍNAS

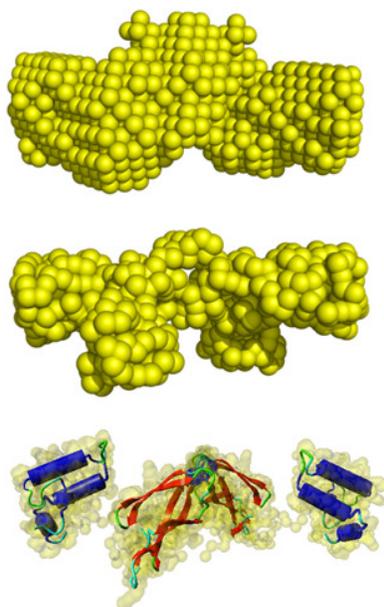
A Figura 5 apresenta os modelos de baixa resolução calculados para a proteína *Sis1* e suas mutantes *Sis1* $_{\Delta 124-174}$  e *Sis1* $_{\Delta 121-257}$ . Esses modelos foram obtidos dos dados de SAXS usando somente cálculos *ab initio* (métodos de átomos fictícios e métodos de resíduos fictícios) e os modelos obtidos combinando cálculos *ab initio* e modelagem de corpos rígidos. Em cada coluna, primeiro está mostrado o modelo de átomos fictícios, depois o de resíduos fictícios e, na última linha, o modelo de organização de domínios com estruturas resolvidas de cada proteína. Os modelos foram calculados impondo simetria espacial *P2* (simetria de rotação de 180° em torno de um eixo de simetria da molécula) devido às conformações diméricas das proteínas e também devido ao fato de que as estruturas cristalográficas da região C-terminal de ambas as proteínas *Sis* e de *Ydj1* cristalizam como dímeros e apresentando simetria *P2* [Sha *et al.*, 2000]. A forma dos modelos está de acordo com a forma predita pelas funções  $p(r)$  dessas proteínas. Eles possuem forma alongada e com características típicas de proteínas com múltiplos domínios, sendo que nos modelos para *Sis1* $_{\Delta 124-174}$  e *Ydj1* $_{\Delta 106-255}$  essa última característica é realçada como já indicavam as oscilações em torno do máximo das funções  $p(r)$  das respectivas proteínas. Nesta seção, vamos primeiro analisar os modelos *ab initio*, deixando o modelo de organização dos domínios para a próxima seção.



**Figura 4.3** – Resultados dos cálculos *ab initio* dos modelos de átomos fictícios (primeira linha), resíduos fictícios (segunda linha) e combinando cálculos *ab initio* com modelagem de corpos rígidos (terceira linha) para Sis1 e suas mutantes. (A) Modelos para Sis1. (B) Modelos para Sis1\_Δ<sub>124-174</sub>. (C) Modelos para Sis1\_Δ<sub>121-257</sub>. Todos os modelos foram visualizados pelo programa PyMOL [DeLano, 2002].

Comparando e analisando cada conjunto de modelos calculados para cada proteína (tipicamente 10 para cada método), calculamos e avaliamos os valores de NSD. Para o conjunto de modelos de átomos fictícios, os valores de NSD ficaram entre 1,00 e 1,22 para Sis1, entre 1,17 e 1,41 para Sis1\_Δ<sub>124-174</sub>, e entre 0,75 e 1,01 para Sis1\_Δ<sub>121-257</sub>. Todos esses valores são considerados muito bons para essa abordagem, indicando que os resultados foram bastante similares em cada conjunto. Com respeito aos modelos de resíduos fictícios, os valores de NSD foram de 1,41 a 1,97 para Sis1, de 1,84 a 2,68 para Sis1\_Δ<sub>124-174</sub>, e de 1,47 e 1,76 para Sis1\_Δ<sub>121-257</sub>, valores muito bons também para essa abordagem. Os modelos apresentados na figura 4.3 são os mais típicos de cada conjunto de modelos, ou seja, com menor NSD.

Com relação à construção Ydj1\_Δ<sub>106-255</sub>, os modelos resultantes são apresentados na figura 4.4 seguindo a mesma ordem daqueles apresentados na figura 4.3 para Sis1 e suas mutantes. Os valores de NSD para o conjunto de modelos de átomos fictícios dessa proteína variaram de 0,88 a 1,21, indicando similaridades nos resultados. E os valores de NSD para o conjunto de modelos de resíduos fictícios ficaram entre 1,38 e 1,56, o que também são muito bons para essa abordagem. Novamente, os modelos apresentados são os mais típicos de cada conjunto, ou seja, com menor NSD.



**Figura 4.4** – Resultados dos cálculos *ab initio* dos modelos de átomos fictícios (primeira linha), resíduos fictícios (segunda linha) e combinando cálculos *ab initio* com modelagem de corpos rígidos (terceira linha) para Ydj1\_Δ<sub>106-255</sub>. Todos os modelos foram visualizados pelo programa PyMOL [DeLano, 2002].

#### 4.2.3.3. RESULTADOS DA MODELAGEM DE CORPOS RÍGIDOS DO POSICIONAMENTO DOS DOMÍNIOS NAS CONFORMAÇÕES DAS PROTEÍNAS

Com o intuito de obter informações sobre as posições e orientações dos múltiplos domínios das proteínas em suas conformações, nós combinamos cálculos *ab initio* e modelagem de corpos rígidos a partir dos dados de SAXS, considerando as estruturas de alta resolução disponíveis dos domínios como corpos rígidos que podem ser transladados e rotacionados. Os modelos resultantes para Sis1 e suas mutantes estão apresentados na terceira linha de cada coluna da figura 4.3. Os valores de NSD médio para cada conjunto foi de 2,24 para Sis1, 2,26 para Sis1\_Δ<sub>124-174</sub>, e de 1,69 para Sis1\_Δ<sub>121-257</sub>, sendo esses valores razoavelmente bons para essa abordagem. Para a mutante Ydj1\_Δ<sub>106-255</sub>, os modelos estão apresentados na terceira linha da figura 4.4 e o valor médio de NSD foi de 2,01, que também é razoavelmente bom para proteínas flexíveis como esta. O mais interessante é que esses modelos têm formas muito semelhantes àqueles modelos obtidos somente com cálculos *ab initio*, mas agora o posicionamento dos múltiplos domínios foi refinado pela modelagem de corpos rígidos. Em todos os casos, aplicamos simetria espacial do tipo *P2* pelos

motivos já explicados na seção 4.2.3.2 deste capítulo. Os sítios de dimerização de ambas as proteínas, Sis1 e Ydj1, assim como de suas mutantes são conhecidos *a priori*, foram identificadas nas estruturas cristalográficas de seus domínios com estrutura resolvida e essa informação foi usada como informação adicional para os cálculos, pois assim sabemos quais aminoácidos deverão estar em contato.

#### **4.2.4. DISCUSSÕES SOBRE OS RESULTADOS OBTIDOS COM AS PROTEÍNAS**

Os resultados das experiências de SAXS mostraram que a HSP40 tipo I Sis1 e suas mutantes são altamente flexíveis em solução, assim como também é a mutante Ydj1\_Δ<sub>106-255</sub> da HSP40 tipo II. Apesar de observamos essa alta flexibilidade intramolecular nas construções aqui analisadas, nós não podemos descartar a possibilidade de que as alterações feitas na estrutura primária de cada proteína (como as deleções de domínios) tenham um efeito nos seus ambientes entrópicos e energéticos em solução, causando essa flexibilidade. Através dos resultados mostrados, pudemos confirmar a dimerização dessas proteínas (possivelmente na sua região C-terminal) em solução. A região C-terminal na Sis1 e suas mutantes têm o mesmo sítio de dimerização, assim como Ydj1\_Δ<sub>106-255</sub> também manteve o sítio de dimerização da Ydj1 inteira. Como observamos a dimerização em todas as amostras aqui analisada, podemos dizer que essa região C-terminal é muito importante para a função da HSP40, já que sabemos que a dimerização auxilia nas atividades funcionais dessas proteínas.

Usando três diferentes métodos de modelagem com base nos dados de SAXS, obtivemos envelopes moleculares similares para cada proteína considerando os resultados de diferentes métodos. A abordagem combinando cálculos *ab initio* com modelagem de corpos rígidos mostrou ser a mais apropriada para estudar essas proteínas flexíveis, pois elucidaram as possíveis posições e orientações dos domínios com estrutura resolvida nas conformações das proteínas. Isso é importante para entender melhor a relação função-estrutura das proteínas HSP40, principalmente dos mutantes, cujas atividades funcionais estão em estudo [Silva *et al.*, 2010]. Interessantemente, mesmo retirando alguns domínios conhecidamente flexíveis das proteínas Sis1 e Ydj1, as construções mutantes mantiveram a alta flexibilidade intramolecular. Além dessa importante característica, os mutantes mantiveram a arquitetura tipo grampo das proteínas inteiras e também mantiveram os domínios-J apontados para fora em direções opostas. Essa arquitetura é bastante favorável para as interações das HSP40 com a HSP70. Dessa forma, os modelos de baixa resolução aqui construídos concordam com o modelo funcional proposto por Qian *et al.* [Qian *et al.*, 2002], chamado “*Anchoring and*

*docking model*”, de como as HSP40s facilitam a entrega de polipeptídios desenovelados para a HSP70.

### **4.3. A PROTEÍNA RIBONUCLÉICA HETEROGÊNEA TIPO Q (HNRNP-Q): UMA PROTEÍNA COM MÚLTIPLOS DOMÍNIOS**

#### **4.3.1. DESCRIÇÃO DA PROTEÍNA HNRNP-Q**

As proteínas ribonucléicas heterogêneas (hnRNP, abreviação o termo em inglês *heterogeneous nuclear ribonucleoproteins*) são proteínas que desempenham importantes papéis na expressão gênica e metabolismo de mRNA [Choi *et al.*, 1986]. Essas proteínas estão relacionadas com uma ampla variedade de funções do RNA, incluindo: transcrição, processamento de pré-mRNA, transporte de mRNA [Choi *et al.*, 1986; Singh *et al.*, 2005]. As hnRNPs formam grandes complexos heterogênicos na célula e, embora elas tenham papéis no metabolismo do RNA, esses papéis não estão completamente caracterizados [Dreyfuss *et al.*, 1993]. A família das hnRNPs ainda contém várias subfamílias e uma delas é a chamada hnRNP tipo Q (hnRNPQ), que possui três isoformas que vão de Q1 a Q3. Sua menor isoforma Q1 tem uma massa molecular de 62 kDa, enquanto Q2 e Q3 têm 65 kDa e 70 kDa, respectivamente. Estruturalmente, todas as isoformas da hnRNP-Q têm uma organização modular com um domínio ácido (AcD) na porção N-terminal, seguido por três domínios de ligação a RNA (RRM) e uma região rica em Arginina e Glicina (RGGbox) na porção C-terminal.

Agora, no que diz respeito às estruturas das hnRNPs, essas são escassas na literatura. Com referência à proteína hnRNP-Q, somente alguns domínios RRM s tiveram suas estruturas resolvidas por RMN (código PDB 2DGU) e a estrutura do domínio ácido (AcD) foi calculada através de modelagem teórica por homologia e dinâmica molecular [Quaresma *et al.*, 2006]. Dessa forma, para caracterizar a proteína hnRNP-Q por completo, sua estrutura precisa ser elucidada. Aqui, experiências de SAXS foram realizadas para caracterizar estruturalmente a isoforma Q1 essa proteína fusionada a uma cauda de GST, porém sem o domínio RGGbox por questões de estabilidade e monodispersidade da amostra. Ela será identificada aqui como GST-hnRNP-Q(1-443). Além disso, métodos de modelagem *ab initio* de baixa resolução com base nas informações de SAXS foram usados para restaurar o envelope molecular de GST-hnRNP-Q(1-443). Também utilizamos a combinação de cálculos *ab initio* com modelagem de corpos rígidos para obter a organização espacial dos domínios na conformação de GST-hnRNP-Q(1-443). Maiores detalhes dos nossos estudos com a hnRNP-Q estão descritos em Silva *et al.* [Silva *et al.*, 2010c]

## 4.3.2. MATERIAIS DE MÉTODOS

### 4.3.2.1. PREPARAÇÃO DAS AMOSTRAS

Todas as experiências foram realizadas com a proteína hnRNP-Q(1-443) com uma cauda de GST fusionada na sua região N-terminal. Essa construção apresenta um domínio ácido (AcD) no N-terminal, seguida por três domínios de ligação a RNA (RRM), sem a região RGGbox. A razão para usar essa construção foi à dificuldade em obter a proteína inteira usando bactérias *E. coli* para expressar a proteína. A figura 4.5 mostra uma representação esquemática do arranjo dos domínios nessa construção. A proteína foi expressa e purificada como descrito em Quaresma *et al.* [Quaresma *et al.*, 2009; Quaresma *et al.*, 2006]. O tampão em que a proteína GST-hnRNP-Q(1-443) foi estudada continha 10 mM de Na<sub>2</sub>HPO<sub>4</sub> e 1,8 mM de KH<sub>2</sub>PO<sub>4</sub> (pH 7,4), 137 mM de NaCl, 2,7 mM KCl e 2 mM de DTT.



**Figura 4.5** – Representação esquemática da organização dos domínios da proteína na construção GST-hnRNP-Q(1-443). As regiões separando os domínios são preditas como flexíveis.

### 4.3.2.2. DESCRIÇÃO DAS EXPERIÊNCIAS DE SAXS COM GST-HNRNP-Q(1-443)

Os dados de SAXS foram coletados na linha D02A-SAXS2 do Laboratório Nacional de Luz Síncrotron (LNLS, Campinas, SP, Brasil) usando um detector bidimensional MARCCD. Curvas de espalhamento de GST-hnRNP-Q(1-443) foram medidas em duas concentrações: 3,4 mg/mL e 1,7 mg/mL. Todos os dados foram coletados com a solução em uma temperatura constante de 20°C, via circulação de água e banho térmico, usando um porta-amostra com uma cela de 1 mm de espessura e duas janelas de mica por onde passa o feixe de raios-X [Cavalcanti *et al.*, 2004]. As experiências foram realizadas com um feixe monocromático de comprimento de onda de 1,488 Å e uma distância amostra-detector de 1374,4 mm, resultando em um intervalo do módulo de vetor de espalhamento de  $0,01\text{Å}^{-1} < q < 0,25\text{Å}^{-1}$ , onde  $q$  é o módulo do vetor  $q$  definido por  $q = (4\pi/\lambda)\text{sen}\theta$  (onde  $2\theta$  é o ângulo de espalhamento como na seção 1.2 do capítulo 1). Foram

feitas três coletas sucessivas de 300 s cada amostra e mais uma, também de 300 s, para o tampão. Todas as curvas de espalhamento foram individualmente corrigidas pela resposta do detector e normalizadas pela intensidade integrada do feixe incidente e pela absorção das amostras como descrito na seção 1.6 do capítulo 1. As curvas resultantes foram cuidadosamente inspecionadas para verificar a formação de bolhas na solução e a existência de possíveis danos induzidos por radiação nas proteínas, mas tais efeitos não foram observados. A massa molecular da proteína foi estimada dos dados de SAXS conforme descrito na seção 1.4.7 do capítulo 1 e usando como padrão uma solução de albumina bovina (BSA) com uma concentração de 7,0 mg/mL. Todos os dados foram tratados e analisados seguindo os procedimentos padrões conforme descritos no capítulo 1.

#### **4.3.2.3. MODELAGEM *AB INITIO* BASEADA NOS DADOS DE SAXS DA GST-HNRNP-Q(1-443)**

Modelos de baixa resolução do envelope molecular de GST-hnRNP-Q(1-443) foram obtidos a partir dos dados de SAXS usando os dois métodos *ab initio* descritos no capítulo 1: o método de átomos fictícios (seção 1.9.3 do capítulo 1) e o método de resíduos fictícios (seção 1.9.4 do capítulo 1). Os resultados de dez cálculos independentes com cada método foram analisados e comparado usando a metodologia descrita na seção 1.10 do capítulo 1. Selecionamos o modelo mais típico (menor NSD) de cada conjunto e cada método para representar a proteína.

#### **4.3.2.4. MODELAGEM DA TOPOLOGIA DO ARRANJO DE DOMÍNIOS DA PROTEÍNA**

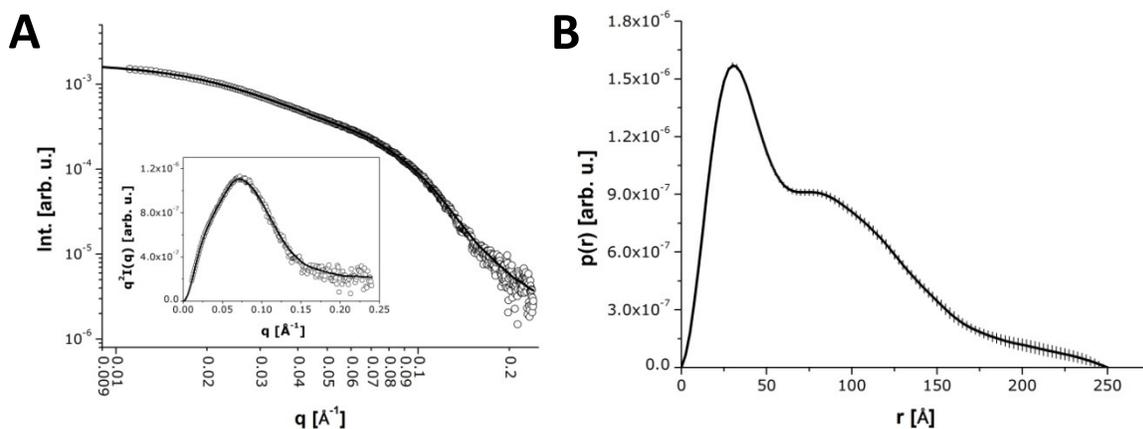
Usamos o método que combina cálculos *ab initio* e modelagem de corpos rígidos (seção 1.9.5 do capítulo 1) para encontrar um modelo para as possíveis posições e orientações dos domínios de GST-hnRNP-Q(1-443) em solução. As estruturas de alta resolução do domínio RRM3 (código PDB 2DGU) e da molécula de GST (código PDB 1GTA) estão disponíveis no banco de dados *Protein Data Bank*, abreviado por PDB ([www.rcsb.org](http://www.rcsb.org)). Já a estrutura do domínio ácido (AcD) foi obtida por homologia e dinâmica molecular por Quaresma *et al.* [Quaresma *et al.*, 2006] e nos foi gentilmente fornecida pelos autores. Infelizmente, o único domínio RRM da proteína hnRNP-Q com estrutura de alta resolução já resolvida é o RRM3. Devido a isso, nós usamos o fato de que os envelopes moleculares dos RRM3 são tipicamente similares em outras famílias de hnRNPs e têm aproximadamente o mesmo tamanho. Para se ter uma idéia dessa similaridade, esses domínios têm em média ~11kDa e tipicamente contém quatro elementos de folhas-beta e dois de

alfa-hélices [Petoukhov *et al.*, 2006]. Como não podemos lidar com a estrutura interna desses domínios RRM por SAXS, nós modelamos a proteína usando três estratégias como segue: (1) Usamos somente as estruturas do GST e do RRM3 como corpos rígidos e simulamos as partes restantes como cadeias de resíduos fictícios; (2) A estrutura do RRM3 foi adaptada, com base na seqüência de aminoácidos desses domínios, primeiro para ter o tamanho típico do domínio RRM1 e depois para ter o tamanho típico do domínio RRM2. Dessa forma, usamos quatro estruturas como corpos-rígidos: GST, RRM1, RRM2 e RRM3; (3) Subseqüentemente, combinamos todas as estruturas que temos, incluindo a estrutura calculada para o domínio ácido por dinâmica molecular, para construir o modelo, simulando as partes remanescentes por cadeias de resíduos fictícios. Cuidados especiais foram tomados para evitar problemas com essas estratégias. Somente os modelos cujos envelopes moleculares foram similares aos modelos de átomos fictícios e resíduos fictícios, obtidos anteriormente, foram considerados para nossos estudos. De certa forma, os modelos *ab initio* funcionaram como controle para os modelos aqui obtidos. Sendo assim, os resultados finais podem contribuir para a compreensão da relação estrutura-função dessa proteína e fornecer uma idéia de como possivelmente ela deva se ligar ao RNA.

### 4.3.3. RESULTADOS

#### 4.3.3.1. RESULTADOS DAS EXPERIÊNCIAS DE SAXS COM A PROTEÍNA HNRNP-Q

A curva de intensidade de espalhamento de GST-hnRNP-Q(1-443) está mostrada na figura 4.6(A) e representação de Kratky é mostrada no destaque dessa figura. A função de distribuição de pares de distância,  $p(r)$ , calculada com o método de transformada de Fourier indireta (seção 1.7 do capítulo 1) está mostrada na figura 4.6(B). A função  $p(r)$  indica que a proteína apresenta uma conformação alongada com dimensão máxima ( $D_{máx}$ ) de 250 Å. Os valores de raio de giro ( $R_g$ ) da proteína obtido pela aproximação de Guinier e pela função  $p(r)$  foram  $65 \pm 1$  Å e  $64 \pm 1$  Å, respectivamente. Usando a solução de BSA como padrão, a massa molecular estimada dos dados de SAXS para a GST-hnRNP-Q(1-443) foi de  $\sim 81$  kDa. Esse valor está em concordância com o valor calculado com a ferramenta ProtParam [Gasteiger *et al.*, 2005] para a seqüência de aminoácidos da proteína. O valor calculado foi de 76 kDa, indicando que a proteína é um monômero em solução.



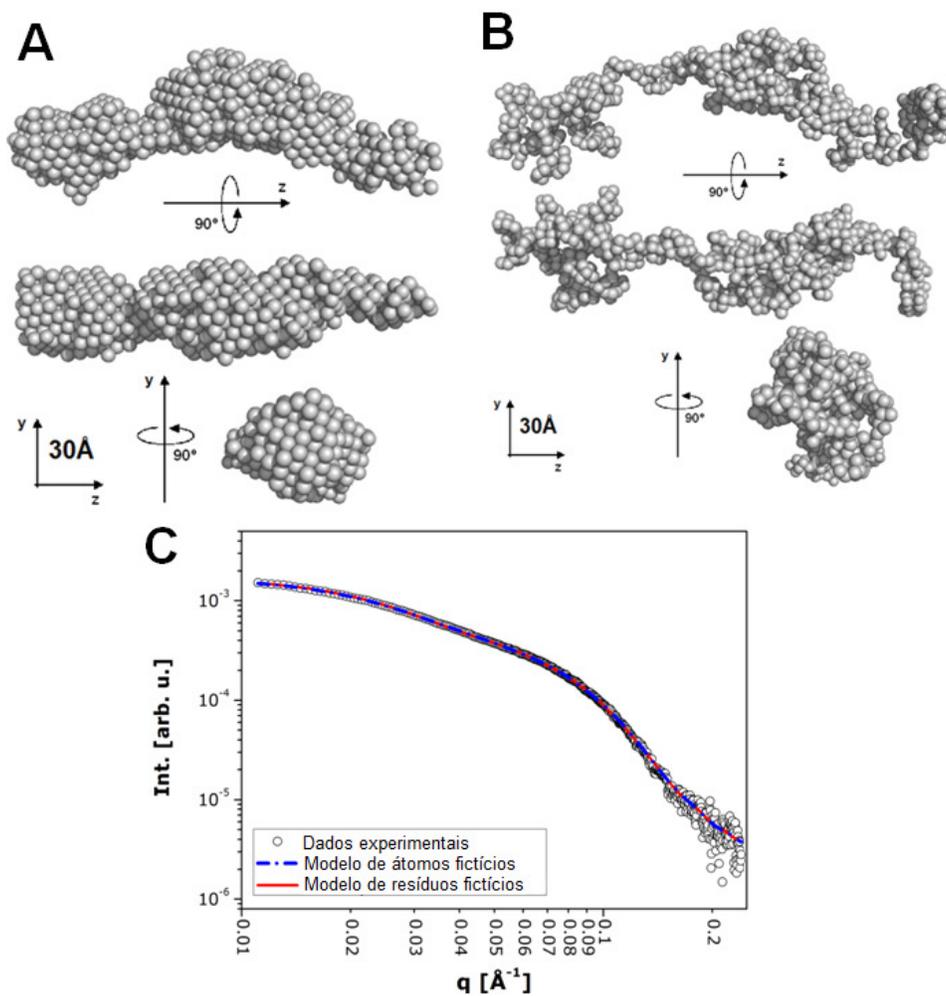
**Figura 4.6** – Dados experimentais de SAXS para proteína GST-hnRNP-Q(1-443) em solução. (A) Curva de intensidade experimental (círculos abertos) e o ajuste calculado com o método de transformada de Fourier indireta (linha sólida). *Destaque:* Representação de Kratky ( $q^2I(q)$  vs.  $q$ ). (B) Função de distribuição de pares de distâncias,  $p(r)$ , indicando uma forma alongada para a proteína com dimensão máxima de  $250\text{\AA}$ .

É interessante notar que a representação de Kratky da GST-hnRNP-Q(1-443) no destaque da figura 4.6(A) apresenta uma “forma de sino” com um máximo bem definido. Isso mostra que essa proteína apresenta uma certa compacidade, mais do que esperada, embora não possamos descartar a existência de flexibilidade na molécula. Muitas vezes quando a flexibilidade conformacional da proteína não resulta em variações de larga escala na estrutura, o efeito dessa flexibilidade pode ser bastante sutil e é facilmente mascarado por ruído nos dados de SAXS. Quando isso ocorre, fontes adicionais de informação, tais como estruturas de alta resolução das subunidades ou domínios e dados pertinentes a flexibilidade, são requeridos a fim de extrair e entender as informações dos dados de SAXS sobre proteínas flexíveis em solução. A estimativa de volume excluído da proteína usando a teoria de Porod também pode ser bastante útil (seção 1.4.5 do capítulo 1) nesse caso, pois um volume de Porod maior que o esperado pode ser um sinal que alguma flexibilidade existe entre os domínios rígidos. Além disso, existe uma relação empírica que diz que, para proteínas globulares, o volume de Porod (isto é, volume hidratado) em escala de  $\text{nm}^3$  é aproximadamente duas vezes maior que a massa molecular da proteína em escala de kDa [Petoukhov *et al.*, 2006; Heller, 2005]. Logo, se usarmos essa relação e obtivermos uma massa bem maior, teremos mais um indício de flexibilidade da proteína. Vale lembrar que, conforme discutido na seção 1.4.5 do capítulo 1, o procedimento de cálculo do volume de Porod é correto para decaimento da curva de intensidade com  $q^{-4}$  ou bem próximo desse comportamento. No caso da GST-hnRNP-Q(1-443), vemos da representação de Kratky que essa condição é satisfeita. Feito isso,

o volume de Porod estimado para GST-hnRNP-Q(1-443) foi de  $\sim 200\text{nm}^3$ , que fornece uma estimativa de massa de  $\sim 100\text{ kDa}$  para uma proteína globular, ou seja, um pouco maior que os  $\sim 81\text{ kDa}$  estimados anteriormente. Portanto, podemos inferir que a flexibilidade da GST-hnRNP-Q(1-443) existe, mas não é tão alta para aparecer na representação de Kratky ou talvez a molécula de GST seja tão maior que os outros domínios da proteína que contribuem bastante para a não flexibilidade da proteína.

#### **4.3.3.2. MODELAGEM *AB INITIO* DA GST-HNRNP-Q(1-443)**

Usando os métodos de átomos fictícios e de resíduos fictícios, nós restauramos o envelope molecular global da GST-hnRNP-Q(1-443) a partir dos dados de SAXS. Lembrando que esse processo não produz uma solução única, dez cálculos independentes foram realizados para cada método. Conformações similares foram observadas para cada grupo e até entre os grupos, onde todos apresentaram a mesma forma alongada recorrente. Após analisá-los conforme descrito na seção 1.10 do capítulo 1, o modelo médio (com volume corrigido) obtido do conjunto de 10 modelos obtidos com o método de átomos fictícios está mostrado na figura 4.7(A). Os valores de NSD desse conjunto ficaram entre 0,64 e 0,70, comprovando a alta similaridade entre eles (valores de NSD até 1 indicam similaridade). Quando usamos o método de modelagem com resíduos fictícios, o conjunto de 10 modelos teve valores de NSD mais altos (diferença típica observada entre os dois métodos de modelagem), que variaram de 1,92 até 2,13. Esses valores ainda são considerados bons para essa abordagem, indicando pequenas diferenças entre modelos, mas não alterando suas formas recorrentes. O modelo de resíduos fictícios mais típicos (com menor valor de NSD) está apresentado na figura 4.7(B). Os correspondentes ajustes dos modelos aos dados experimentais estão mostrados na figura 4.7(C). A resolução desses modelos não permite uma determinação unívoca das posições espaciais dos elementos de estrutura secundária da proteína, mas mesmo assim os modelos fornecem uma visão tridimensional da forma global da conformação alongada da GST-hnRNP-Q(1-443) em solução.

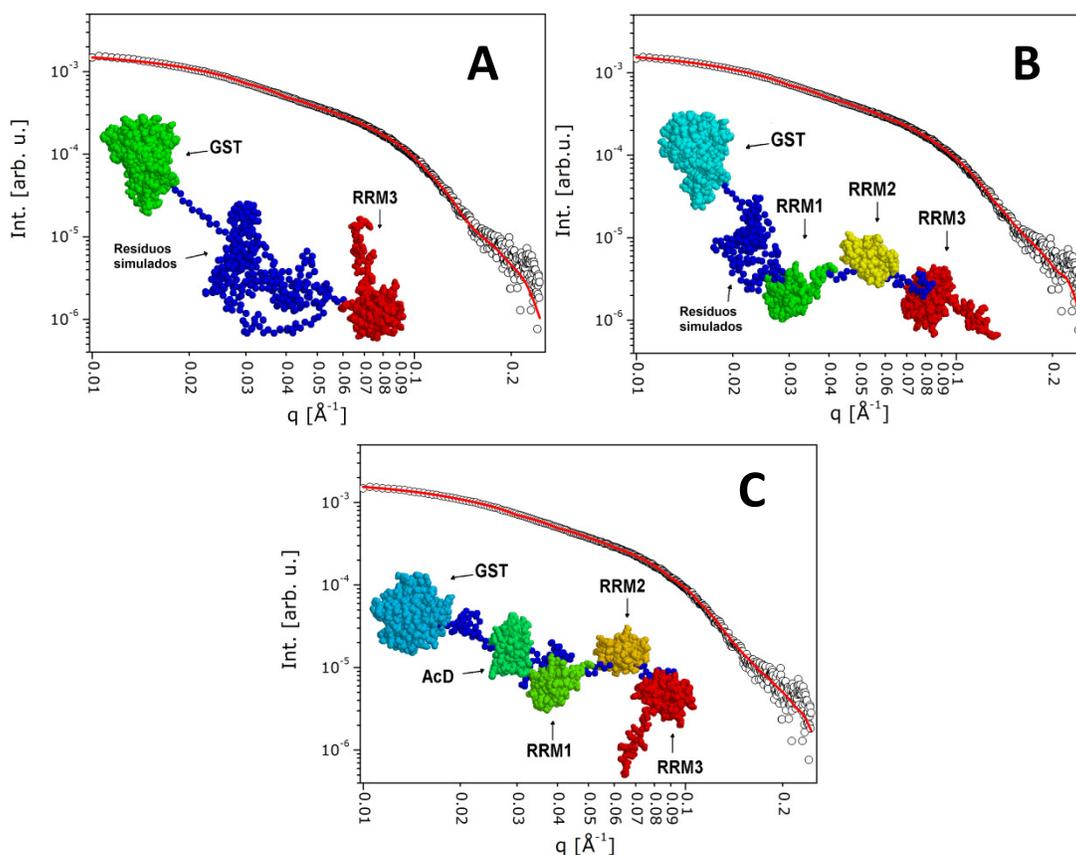


**Figura 4.7** – Modelos *ab initio* para a conformação de GST-hnRNP-Q(1-443) em solução. (A) Modelo médio (com volume excluído corrigido) obtido do conjunto de modelos de átomos fictícios calculados. (B) Modelo mais típico do conjunto de modelos de resíduos fictícios (com menor NSD). (C) Curva de intensidade de espalhamento experimental (círculos abertos) e as curvas de espalhamento calculadas dos modelos: de átomos fictícios (linha tracejada) e de resíduos fictícios (linha sólida). Os modelos foram apresentados usando o visualizador molecular PyMOL [DeLano, 2002].

#### 4.3.3.3. MODELAGEM DA TOPOLOGIA DO ARRANJO DE DOMÍNIOS DA PROTEÍNA GST-HNRNP-Q(1-443)

Para modelar o possível arranjo de múltiplos domínios da GST-hnRNP-Q(1-443) em solução, combinamos cálculos *ab initio* e modelagem de corpos rígidos. Nós adotamos as três

estratégias previamente descritas na seção de materiais e métodos (seção 4.3.2.4 deste capítulo). Todos os cálculos foram realizados sem a imposição de simetria espacial. Nós então dividimos os cálculos em três grupos, cada grupo seguindo uma estratégia. Começando com a primeira estratégia, ao usarmos somente as estruturas da molécula de GST e do domínio RRM3 como corpos rígidos e simulamos o restante como resíduos fictícios, o resultado está apresentado na figura 4.8(A), juntamente com ajuste do espalhamento calculado para o modelo à curva de intensidade experimental.



**Figura 4.8** – Modelos de baixa resolução da organização dos domínios na conformação de GST-hnRNP-Q(1-443) em solução. (A) Ajuste à curva experimental e modelo (no destaque) obtido usando somente as estruturas do GST e do RRM como corpos rígidos e simulando as partes restantes como resíduos fictícios. (B) Ajuste à curva experimental e modelo (no destaque) obtido usando somente as estruturas de RRM1, RRM2 e RRM3 (detalhes no texto) e a estrutura do GST. (C) Ajuste à curva experimental e modelo (no destaque) calculado usando todas as estruturas dos domínios (GST, RRM1, RRM2, RRM3 e AcD) e simulando as partes remanescentes como cadeias de resíduos fictícios.

Continuando, na segunda estratégia utilizamos a estrutura do GST mais as estruturas dos RRM's (conforme descrito na seção 4.3.2.4 deste capítulo) e o resultado está mostrado na figura 4.8(B). Por fim, incluindo agora as cinco estruturas dos cinco domínios (GST, domínios RRM's e a estrutura calculada de AcD) de GST-hnRNP-Q(1-443) e simulando as partes remanescentes como resíduos fictícios, nós obtivemos o resultado apresentado na figura 4.8(C). Como podemos observar, todos os modelos das diferentes estratégias apresentaram similaridade de forma, sempre com um formato alongado. Além disso, vemos que o modelo obtido com a terceira estratégia é bastante similar aos modelos obtidos por métodos *ab initio*, o que indica que todos são modelos coerentes com a conformação da GST-hnRNP-Q(1-443) em solução.

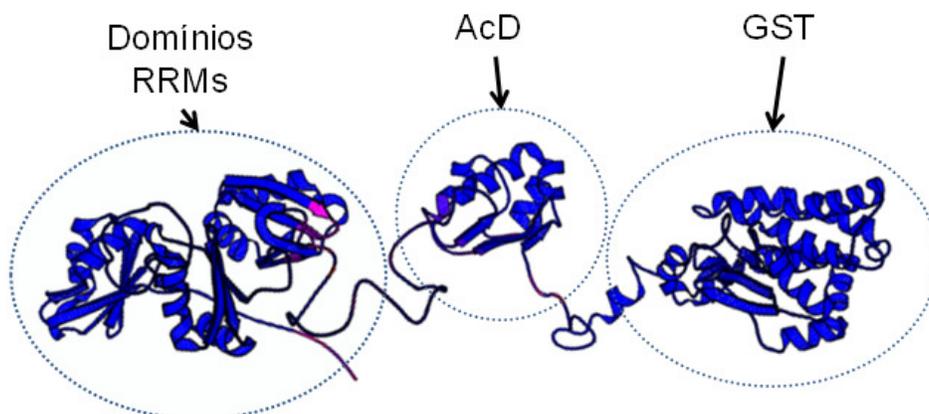
#### **4.3.4. DISCUSSÃO SOBRE OS RESULTADOS OBTIDOS COM A PROTEÍNA HNRNP-Q**

Neste trabalho, os resultados de SAXS de uma construção da proteína hnRNP-Q fusionada a um cauda de GST foi analisada. A região RGGbox foi cortada devido a dificuldades em obter a proteína inteira em bactérias *E. coli*. Esses resultados indicam que a GST-hnRNP-Q(1-443) possui uma forma alongada e uma conformação levemente compacta. Além disso, os resultados de SAXS também confirmaram que essa proteína é um monômero em solução. Nós ainda usamos três diferentes abordagens para modelar o envelope molecular de baixa resolução da GST-hnRNP-Q(1-443) a partir das curvas de SAXS dessa proteína. Os resultados de dois métodos de modelagem *ab initio* (átomos fictícios e resíduos fictícios) forneceram modelos com um formato recorrente para todos os modelos de cada método e similares quando comparados os modelos entre os métodos. Também foi possível obter um modelo com o possível arranjo espacial dos múltiplos domínios da GST-hnRNP-Q(1-443) em solução, obtido através de método que combina de cálculos *ab initio* e de modelagem de corpos rígidos. Esse último método é baseado nos dados de SAXS, onde o espalhamento calculado do modelo resultante ajustou bem os dados experimentais.

Baseado nesses resultados de SAXS, nós propomos que a estrutura a construção GST-hnRNP-Q(1-443) é composta por cinco domínios globulares: a molécula de GST, o domínio ácido (AcD) e os três domínios RRMS, conectados por regiões flexíveis, em um arranjo linear. Uma moderada flexibilidade da conformação dessa proteína foi observada dos dados de SAXS a partir da estimada de volume hidratado da proteína. Além disso, na figura 4.9 nós apresentamos uma representação pictórica da estrutura proposta inspirada nos modelos de baixa resolução da proteína obtidos por SAXS. Obviamente, os detalhes dessa estrutura proposta não podem ser levados em conta porque isso não é uma estrutura de alta resolução. No entanto, esses modelos estão em acordo

com as características estruturais desse tipo de proteína, onde os domínios de ligação a RNA estão bem próximos.

Tendo em mente a possível ambigüidade de interpretação dos dados de SAXS em termos de modelos tridimensionais, análises extensivas de validação dos repetidos cálculos devem ser feitas por comparação entre as diferentes abordagens. Essa ambigüidade nas orientações dos domínios entre diferentes soluções do método que trata esses domínios como corpos rígidos pode em parte ser atribuída à variação nas conformações da proteína. Entretanto, a ambigüidade e variabilidade das conformações das regiões flexíveis que conectam os domínios são devidas até certo ponto à baixa resolução dos dados de SAXS. Quando análises de corpos rígidos foram usadas, o método permitiu restaurar as posições relativas dos múltiplos domínios da GST-hnRNP-Q(1-443). Entretanto, por serem domínios individuais globulares (que é o caso dos domínios RRM, AcD e da molécula de GST), SAXS é menos sensível às suas orientações. Informações adicionais, de medidas de acoplamento dipolar e/ou medidas de relaxação usando a técnica de RMN, são requeridas para posteriores refinamentos do presente modelo.



**Figura 4.9** – Representação pictórica da estrutura proposta para a GST-hnRNP-Q(1-443) inspirada nos modelos de baixa resolução obtidos por SAXS. Os detalhes não devem ser levados em conta porque esta não é uma estrutura de alta resolução experimental. As possíveis posição e orientação dos domínios são sugeridas.



# Considerações finais

O estudo de proteínas com regiões desestruturadas continua ainda sendo um grande desafio para as ciências envolvidas na biologia estrutural. A maior parte das técnicas disponíveis hoje é baseada na hipótese de uma macromolécula compacta e rígida. Outras só fornecem informações hidrodinâmicas. Nesse cenário, a técnica de SAXS se apresenta como uma técnica que consegue fornecer informações estruturais quando há flexibilidade intramolecular, além de também permitir a construção de modelos de baixa resolução para o envelope molecular dessas macromoléculas. Para proteínas intrinsecamente desestruturadas, os modelos obtidos por SAXS são o mais próximo que se pode chegar da visualização tridimensional de suas estruturas. Graças a isso, importantes avanços estão sendo feitos no entendimento das funções de proteínas intrinsecamente desestruturadas ou flexíveis com múltiplos domínios. Obviamente, para esses estudos, não podemos usar somente uma técnica. A complementaridade de diversas técnicas e abordagens é necessária para a devida caracterização dessas macromoléculas em solução.

Neste trabalho de tese de doutorado, nós apresentamos o atual cenário dos estudos de proteínas com regiões desestruturadas utilizando a técnica de SAXS. As abordagens utilizadas e os diversos métodos de modelagem com base nos dados de SAXS são os mais avançados que temos disponíveis hoje para esse tipo de estudo. Ao estudar as proteínas intrinsecamente desestruturadas e as proteínas flexíveis com múltiplos domínios usando a técnica de SAXS, nós mostramos como é possível obter informações dessas macromoléculas que se correlacionam com suas funções. Isso é a razão central dos estudos estruturais de proteínas. Isso demonstra claramente o potencial da técnica de SAXS e o porquê ela é, até o momento, a técnica por excelência para os estudos estruturais de macromoléculas desestruturadas e flexíveis em solução.

Nos estudos com FEZ1 e Ki-1/57, nós apresentamos resultados que permitiram entender também o perfil de interações dessas moléculas com outras proteínas, já que proteínas nunca trabalham sozinhas nas células. Os resultados também indicaram características estruturais de FEZ1 que a caracterizam com uma proteína transportadora e de Ki-1/57 que sugerem sua função de proteína adaptadora. Já nas análises dos dados de SAXS das mutantes por deleção de domínios das HSP40s Sis1 e Ydj1, nós mostramos características estruturais dessas macromoléculas que indicam como essas proteínas conseguem carregar cadeias desenoveladas de proteínas recém-sintetizadas para as proteínas HSP70, que por sua vez irá auxiliar o enovelamento dessas novas proteínas juntamente com outras chaperonas (HSP60, HSP90 e outras mais). Por fim, os estudos estruturais

da proteína hnRNP-Q utilizando a técnica de SAXS revelaram características da organização de seus domínios que permitem compreender melhor sua interação com o RNA.

Analisando o conjunto atual de técnicas de caracterização estrutural disponíveis, ainda há muito para se fazer, tanto do ponto de vista experimental quanto do teórico, para que um dia possamos obter todas as respostas que esperamos com respeito à ligação dessas proteínas desestruturadas com doenças crônicas como o câncer. Há novas teorias e idéias para o estudo de macromoléculas desestruturadas, porém muitas delas ainda não funcionam muito bem ou são refutadas rapidamente. Inclusive novas abordagens para as análises de dados de SAXS estão sendo desenvolvidas, porém muitas delas estão baseadas nos estudos em polímeros em solução. As diferenças entre os vários níveis estruturais de uma proteína e de um polímero fazem com que essas abordagens não se apliquem muito bem para os dados de proteínas.

Quando se trata do estudo da estrutura e função de proteínas com regiões desestruturadas e flexíveis, as dificuldades não se encontram somente na física. As proteínas nativamente desoveladas, por exemplo, são difíceis de serem expressas, pois elas degradam facilmente ou então expressam numa pequena quantidade, muitas vezes sendo necessário expressá-las fusionada a outras moléculas como o GST e a cauda de Histidina. Além disso, essas proteínas têm uma alta propensão a se agregarem, dificultando a realização de qualquer experiência de caracterização estrutural cuja premissa básica é a monodispersidade da proteína. Isso exige uma padronização de tampões para proteína e rotas alternativas de purificação. Sendo assim, há a necessidade de criar procedimentos alternativos para a expressão e purificações de proteínas desse tipo, utilizando, por exemplo, a maquinaria de organismos mais complexos que bactérias.

Do ponto de vista instrumental, as linhas de luz dedicadas às experiências de SAXS recebem cada vez mais pesquisadores trazendo amostras de proteínas desestruturadas em solução. Muitos tentaram sem sucesso por anos cristalizá-las sem saber de suas características desestruturadas. Em vista disso, as linhas de luz são melhoradas cada dia mais, com novos porta-amostras, novos detectores, além de dispositivos de inserção de fontes síncrotron (Wigglers e Onduladores) que fornecem um feixe de luz mais brilhante, tudo isso colaborando muito para os estudos de proteínas em solução por SAXS. A construção de uma nova linha de SAXS no Laboratório Nacional de Luz Síncrotron permitiu um melhora significativa na qualidade de dados de SAXS coletados para amostras em solução.

As perspectivas para futuros trabalhos levam em conta os avanços na tecnologia de produção de radiação coerente, como os lasers de elétrons livres, que podem impulsionar os avanços das técnicas de caracterização estrutural de macromoléculas ou até o surgimento de novas técnicas. Por enquanto ainda não há lasers desse tipo produzindo radiação coerente com

comprimento de onda de raios-X, mas investimentos pesados estão sendo feitos para chegar a esse ponto. Só citando alguns projetos que já estão em andamento temos: o *European X-ray Laser Project XFEL* na Alemanha (<http://xfel.desy.de/>), o *SLAC Linac Coherent Light Source* nos Estados Unidos da América (<http://lcls.slac.stanford.edu/>) e *X-ray FEL SCSS Project* que estão sendo desenvolvido no *RIKEN Spring-8 Center* no Japão (<http://www.riken.jp/engn/r-world/research/lab/xfel/index.html>). Isso irá permitir a realização das chamadas experiências com molécula única, onde uma única molécula por vez é atingida por um pulso de radiação. As idéias das experiências com a radiação de laser de elétrons livres são bastante promissoras para fornecer informações de alta resolução sobre as estruturas de proteínas intrinsecamente desestruturadas.



# Lista de publicações científicas

- [1]. **Silva, J. C.**; Borges, J. C.; Cyr, D. M.; Ramos, C. H. I.; Torriani, I. L. (2010), “SAXS studies of heat shock proteins (*Hsp40*) *Sis1*, *Ydj1* and mutants: dimerization, flexibility and functionality”. **Manuscrito pronto para submissão.**
- [2]. **Silva, J. C.**; Quaresma, A. J. C.; Kobarg, J.; Torriani, I. L. (2010), “SAXS study of the human regulatory protein *hnRNP-Q* in solution”. **Manuscrito pronto para submissão.**
- [3]. **Silva, J. C.**; Lanza, D. C. F.; Kobarg, J.; Torriani, I. L. (2010), “SAXS analysis reveals structural details of the natively unfolded protein *FEZ1*”. **Manuscrito submetido para “Biophysical Journal”.**
- [4]. Alborghetti, M. R.; Furlan, A. S.; **Silva, J. C.**; Paes Leme A. F.; Torriani, I. C. L.; Kobarg, J. (2010). “The human *FEZ1* protein forms a disulfide bond mediated dimer: implications for cargo transport”. **Manuscrito submetido para “Journal of Proteome Research”.**
- [5]. Gonçalves, K. A.; Borges, J. C.; **Silva, J. C.**; Ferreira, P. P.; Bressan, G. C.; Torriani, I. L.; Kobarg, J. (2010). “Solution structure of the human signaling protein *RACK1*”. **Manuscrito submetido para “BMC Structural Biology”.**
- [6]. Breitreitz, M. C.; Tinti, F.; Barboza, F. D.; **Silva, J. C.**; Rost, M.; Bruns, R. E.; Torriani, I. L.; Matha V. (2009), “Development and characterization of a lipid-free delivery system for ciclosporin”. **Manuscrito submetido para “International Journal of Pharmaceutics”.**
- [7]. **Silva, J.C.**, Plivelic, T.S., Herrera, M.L., Rucheinsky, N., Kieckbusch, T.G., Luccas, V., Torriani, I. L. (2009), “Polymorphic phases of natural fat from *Cupuassu* (*Theobroma grandiflorum*) beans: a WAXS/SAXS/DSC study”. **Crystal Growth & Design**, v. **9**(12), p. **5155–5163**. [[doi:10.1021/cg901081j](https://doi.org/10.1021/cg901081j)].
- [8]. Lanza, D.C.F., **Silva, J.C.**, Assmann, E.M., Quaresma, A.J.C., Bressan, G.C., Torriani, I.L., Kobarg, J. (2009), “Human *FEZ1* has characteristics of a natively unfolded protein and dimerizes in solution”. **Proteins**, v.**74**, p.**104–121**. [[doi:10.1002/prot.22135](https://doi.org/10.1002/prot.22135)].
- [9]. Trindade, D.M, **Silva, J.C.**, Navarro, M.S., Torriani, I.C.L., Kobarg, J. (2009), “Low-resolution structural studies of human *Stanniocalcin-1*”. **BMC Structural Biology**, v.**9**, p. **57–68**. [[doi:10.1186/1472-6807-9-57](https://doi.org/10.1186/1472-6807-9-57)]
- [10]. Sousa, S.M., Rosselli, L.K., Kiyota, E., **Silva, J.C.**, Souza, G.H.M.F., Peroni, L.A., Stach-Machado, D.R., Eberlin, M.N., Souza, A.P., Koch, K.E., Arruda, P., Torriani, I.L., Yunes, J.A. (2009), “Structural and kinetic characterization of a maize aldose reductase”. **Plant Physiology and Biochemistry (Paris)**, v.**47**, p.**98-104**. [[doi:10.1016/j.plaphy.2008.10.009](https://doi.org/10.1016/j.plaphy.2008.10.009)]

- [11]. Bressan, G.C., **Silva, J.C.**, Borges, J.C., Passos, D.O., Ramos, C.H.I., Torriani, I.L., Kobarg, J. (2008), “*Human Regulatory Protein Ki-1/57 Has Characteristics of an Intrinsically Unstructured Protein*”. **Journal of Proteome Research**, v.7, p.4465 - 4474. [[doi:10.1021/pr8005342](https://doi.org/10.1021/pr8005342)]
- [12]. Oliveira, C.L.P., de la Hoz, L., **Silva, J.C.**, Torriani, I.L., Netto, F.M. (2007), “*Effects of gamma radiation on Beta-lactoglobulin: Oligomerization and aggregation*”. **Biopolymers**, v.85, p.284 - 294. [[doi:10.1002/bip.20610](https://doi.org/10.1002/bip.20610)]

## Outras publicações científicas

[1]. Bressan, G.C., **Silva, J.C.**, Borges, J.C., Passos, D.O., Ramos, C.H.I., Torriani, I.L., Quaresma, A.J.C., Moraes, E.C., Manfiolli, A.O., Gomes, M.D., Kobarg, J. “**Structural and functional insights to human Ki-1/57: an intrinsically unstructured protein that localizes to nuclear bodies related with RNA metabolism**”. *Science Highlight published in the 2008 Activity report from Brazilian Synchrotron Light Laboratory (LNLS), Brazil.*

[<http://www.lnls.br/ar2008/PDF/sh03.pdf>]

[2]. Lanza, D.C.F., **Silva, J.C.**, Assmann, E.M., Trindade, D.M., Quaresma, A.J.C., Bressan, G.C., Torriani, I.L., Kobarg, J. “**Low resolution structural X-ray studies of human FEZ1: a natively unfolded protein associated to the flower-like nuclei phenotype**”. *Science Highlight published in the 2007 Activity report from Brazilian Synchrotron Light Laboratory (LNLS), Brazil.*

[<http://www.lnls.br/ar2007/PDF/sh01.pdf>]

# Referências

Arfken, G. B.; & Weber, H. J. (2001). "Mathematical Methods for Physicists". 5<sup>th</sup> ed., Harcourt/Academic Press: San Diego, CA, United States of America.

Assmann, E. M.; Alborghetti, M. R.; Camargo, M. E.; & Kobarg, J. (2006). "FEZ1 dimerization and interaction with transcription regulatory proteins involves its coiled-coil region". *J. Biol. Chem.* **281**: 9869-9881.

Banecki, B.; Liberek, K.; Wall, D.; Wawrzynow, A.; Georgopoulos, C.; & Bertoli, E. (1996). "Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone". *J. Biol. Chem.* **271**, 14840-14848.

Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; & Svergun, D. I. (2007). "Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering". *J. Am. Chem. Soc.* **129**: 5656-5664.

Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. (2005). "A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering". *Proc. Natl. Acad. Sci. USA* **102**: 17002-17007.

Bosio, L., Teixeira, J.; & Bellissent-Funel, M.-C., (1989). "Enhanced density fluctuations in water analyzed by neutron scattering". *Phys. Rev. A* **39**: 6612-6613.

Bressan, G. C.; Silva, J. C.; Borges, J. C.; Passos, D. O.; Ramos, C. H.; Torriani, I. L.; & Kobarg, J. (2008). "Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein". *J. Proteome Res.* **7**(10): 4465-4474.

Bressan, G. C.; Quaresma, A. J.; Moraes, E. C.; Manfiolli, A. O.; Passos D. O.; Gomes, M. D.; & Kobarg, J. (2009). "Functional association of human Ki-1/57 with pre-mRNA splicing events". *FEBS J.*, **276**(14): 3770-3783.

Bukau, B.; & Horwich, A. L. (1998). "The Hsp70 and Hsp60 chaperone machines". *Cells* **92**: 351-366.

Caplan, A. J.; & Douglas, M. G. (1991). "Characterization of Ydj1: a yeast homologue of the bacteria DnaJ protein". *J. Cell. Biol.* **114**: 609-621.

Carrigan, P. E.; Sikkink, L. A.; Smith, D. F.; & Ramirez-Alvarado, M. (2006). "Domain:domain interactions within Hop, the Hsp70/Hsp90 organizing protein, are required for protein stability and structure". *Protein Sci.* **15**(3): 522-532.

Castro, A. R. B. (2004). "Synchrotron radiation: production and applications". *Couse notes (Fall 2004 – Summer 2005)*. Institut für Experimentalphysik, Universität Hamburg. Disponível na biblioteca do Instituto de Física "Gleb Wataghin" da Universidade Estadual de Campinas (UNICAMP).

Cavalcanti, L. P.; Torriani, I. L.; Plivelic, T. S.; Oliveira, C. L. P.; Kellermann, G.; & Neuenschwander, R. (2004). "Two new sealed sample cells for small-angle X-ray scattering from macromolecules in solution and complex fluids using synchrotron radiation". *Rev. Sci. Instr.* **75**: 4541-4546.

Champoux, J. J. (2001). "DNA topoisomerases: structure, function, and mechanism". *Annu. Rev. Biochem.* **70**: 369-413.

Chen, L.; Hodgson, K. O.; & Doniach S. (1996). "A Lysozyme Folding Intermediate Revealed by Solution X-ray Scattering". *J. Mol. Biol.* **261**: 658-671.

Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y. Y. J.; Romero, P.; Cortese, M. S.; Uversky, V. N.; & Dunker, A. K. (2006). "Rational drug design via intrinsically disordered protein". *Trends. Biotechnol.* **24**: 435-442.

Cho, W.; & Stahelin, R.V. (2005). "Membrane-protein interactions in cell signaling and membrane trafficking". *Annu. Rev. Biophys. Biomol. Struct.* **34**: 119-151.

Cho, W. (2006). "Building Signaling Complexes at the Membrane". *Sci. STKE* **321**, pe7.

Cowley, J. M. (1995). "Diffraction Physics". 3<sup>rd</sup> rev. ed., Elsevier Science B.V., Amsterdam, The Netherlands.

Cullity, B. D. (1978). "Elements of X-ray Diffraction. 2<sup>nd</sup> ed., Addison-Wesley Publishing Company, Inc., California, United States of America.

Debye, P. (1915). "Zerstreuung von röntgenstrahlen". (*Scattering from noncrystalline substances*). *Ann. Phys.* **46**: 809–823.

Debye, P. (1947). "Molecular-weight determination by light scattering". *J. Phys. Colloid. Chem.* **51**: 18-32.

Debye, P.; & Bueche, A. M., (1949). "Scattering by an Inhomogeneous Solid". *Appl. Phys. Lett.* **20**: 518-525.

Doniach, S. (2001). "Changes in biomolecular conformation seen by small angle X-ray scattering". *Chem. Rev.* **101**: 1763-78.

Dosztanyi, Z.; Chen, J.; Dunker, A. K.; Simon, I.; & Tompa, P. (2006). "Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution". *J. Proteome Res.* **5**, 2985-2995.

Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; & Uversky V. N. (2005). "Flexible nets. The roles of intrinsic disorder in protein interaction networks". *FEBS J.* **272**(20), 5129-5148.

Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; & Obradovic, Z. (2002). "Intrinsic disorder and protein function". *Biochemistry* **41**: 6573-6582.

Dyson H. J.; & Wright P. E. (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm". *J. Mol. Biol.* **293**: 321-331.

Dyson, H. J.; & Wright, P. E. (2002). "Insights into the structure and dynamics of unfolded proteins from NMR". *Adv. Protein Chem.* **62**: 311-340.

Dyson H. J.; & Wright P. E. (2005). "Intrinsically unstructured proteins and their functions". *Nat. Rev. Mol. Cell Biol.* **6**(3): 197-208.

Fan, C. Y.; Lee, S.; Ren, H. Y.; & Cyr, D. M. (2004). "Exchangeable Chaperone Modules Contribute to Specification of Type I and Type II Hsp40 Cellular Function". *Mol. Biol. Cell* **15**(2): 761-773.

Fankuchen, I. (1937). "A condensing Monochromator for X-Rays". *Nature* **139**: 193-194.

Feigin, L. A.; & Svergun, D. I. (1987). "Structure analysis by small-angle X-ray and neutron scattering", Plenum Press: New York.

Farina, J. E. G. (1973). "Quantum theory of scattering processes". 1<sup>st</sup> ed., Pergamon Press, Oxford, New York

Fink, A. L. (2005). "Natively Unfolded Proteins". *Curr. Opin. Struct. Biol.* **15**(1): 35-41.

Fink, A. L. (1999). "Chaperone-Mediated Protein Folding". *Physiol. Rev.* **79**: 425-449.

Fritz, G.; Scherf, G.; & Glatter, O. (2000). "Application of Densitometry, Ultrasonic Speed Measurements and Ultra Low Shear Viscosimetry to Aqueous Fluids". *J. Phys. Chem. B* **104**(15): 3463-3470.

Froese, P.; Lemke, H.; Gerdes, J.; Havsteen, B.; Schwarting, R.; Hansen, H.; & Stein, H. (1987). "Biochemical Characterization and biosynthesis of the Ki-1 antigen in Hodgkin-derived and virus-transformed human B and T lymphoid cell lines". *J. Immunol.* **139**: 2081-2087.

Fuxreiter, M.; Tompa, P.; & Simon, I. (2007). "Local structural disorder imparts plasticity on linear motifs". *Bioinformatics* **23**: 950-956.

Gabriel, A. (1977). "Position sensitive X-ray detector". *Rev. Sci. Instrum.* **48**: 1303-1305.

Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server". In: Walker, J. M., editor. *The proteomics protocols handbook*. Totowa, N. J., Humana Press, pp. 571-607.

Glatter, O.; & Kratky, O., (1982). "Small Angle X-Ray Scattering", Academic Press Inc. (London) Ltd.: London.

Glatter, O. (1977). "A New Method for the Evaluation of Small-Angle Scattering Data". *J. Appl. Cryst.* **10**, 415-421.

Glatter, O. (1988). "Comparison of two different methods for direct structure analysis from small-angle scattering data". *J. Appl. Cryst.* **21**, 886-890.

Glatter, O. (1972). "X-Ray Small Angle Scattering of Molecules Composed of Subunits" *Acta Phys. Austriaca* **36**: 307-315.

Gross, S. P.; Vershinin, M.; & Shubeita, G. T. (2007). "Cargo Transport: Two Motors Are Sometimes Better Than One". *Curr. Biol.* **17**: R478-R486.

Guinier, A.; & Fournet, G., (1955). "Small angle scattering of X-rays". Translated by Walker, C.B. & Yudowitch, K.L., New York: John Wiley & Sons.

Gursky, H.; & Schwartz, D. (1975). "In: X-ray Astronomy", Giacconi, R. and Gursky, H., eds., Reidel, Boston, pp. 44-51.

Hammersley, A. P. (2004). "FIT2D V12.012 Reference Manual V6.0 ESRF98HA01T", ESRF 1987-2004, [http://www.esrf.eu/computing/scientific/FIT2D/FIT2D\\_REF/fit2d\\_r.html](http://www.esrf.eu/computing/scientific/FIT2D/FIT2D_REF/fit2d_r.html).

Hansen, H.; Lemke, H.; Bredfeldt, G.; Könnecke, I.; & Havsteen, B. (1989). "The Hodgkin-associated Ki-1 antigen exists in a intracellular and a membrane-bound form". *Biol. Chem. Hoppe-Seyler*. **370**(5): 409-416.

Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding". *Nature* **381**: 571-580.

Hartl, F. U.; & Hayer-Hartl, M. (2002). "Molecular chaperones in the cytosol: from nascent chain to folded protein". *Science* **295**: 1852-1858.

Hennessy, F.; Nicoll, W. S.; Zimmermann, R.; Cheetham, M. E.; & Blatch, G. L. (2005). "Not all J domains are created equal: Implications for the specificity of Hsp40-Hsp70 interactions". *Protein Sci.* **14**(7): 1697-1709.

Heller, W. T. (2005). "Influence of multiple well defined conformations on small-angle scattering of protein in solution". *Acta Crystallogr. D Biol. Crystallogr.* **61**: 33-44.

Hendrix, J.; Koch, M. H. J.; & Bordas, J. (1979). "A Double Focusing X-ray Camera for Use with Synchrotron Radiation". *J. Appl. Cryst.* **12**: 467-472.

Hirai, M.; Arai, S.; & Iwase, H. (1999). "Complementary Analysis of Thermal Transition Multiplicity of Hen Egg-White Lysozyme". *J. Phys. Chem. B* **103**: 549-556.

Hirai, M.; Koizumi, M.; Hayakawa, T.; Takahashi, H.; Abe, S.; Hirai, H.; Miura, K.; & Inoue, K. (2004). "Hierarchical Map of Protein Unfolding and Refolding at Thermal Equilibrium Revealed by Wide-Angle X-ray Scattering". *Biochemistry* **43**: 9036-9049.

Hodgkinson, C. A.; Goldman, D.; Ducci, F.; DeRosse, P.; Caycedo, D. A.; Newman, E. R., Kane, J. M.; Roy, A.; & Malhotra, A. K. (2007). "The FEZ1 Gene Shows No Association to Schizophrenia in Caucasian or African American Populations". *Neuropsychopharmacology* **32**: 190-196.

Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; & Dunker, A. K. (2002). "Intrinsic disorder in cell-signalling and cancer-associated proteins". *J. Mol. Biol.* **323**:573-584.

Ingber, L. (1993). "Simulated annealing: practice versus theory". *Math. Computer Modelling.* **18**: 29-57.

Jha, A. K.; Colubri, A.; Freed, K. F.; & Sosnick, T. R. (2005). "Statistical coil model of the unfolded state: Resolving the reconciliation problem". *Proc. Natl. Acad. Sci. U.S.A.* **102**: 13099-13204.

Kellermann, G.; Vicenting, F.; Tamura, E.; Rocha, M.; Tolentino, H.; Barbosa, A.; Craievich, A.; & Torriani, I. L. (1997). "The Small-Angle X-ray Scattering Beamline of the Brazilian Synchrotron Light Laboratory". *J. Appl. Cryst.* **30**, 880-883.

Kirkpatrick, S.; Gelatt, C. D.; & Vecchi, M. P. (1983). "Optimization by simulated annealing". *Science* **220**: 671-680.

Kleywegt, G. J. J.(1997). "Validation of protein models from C-alpha coordinates alone". *Mol. Biol.* **273**(2): 371-376.

Koch, M. J.; Vachette, P.; & Svergun, D. I. (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution". *Q. Rev. Biophys.* **36**(2): 147-227.

Koga, M.; Ishiguro, H.; Horiuchi, Y.; Albalushi, T.; Inada, T.; Iwata, N.; Ozaki, N.; Ujike, H.; Muratake, T.; Someya, T.; & Arinami, T. (2007). "Failure to confirm the association between the FEZ1 gene and Schizophrenia in a Japanese population". *Neurosci. Lett.* **417**: 326-329.

Kozin, M. B.; & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**: 33-41.

Kratky, O.; & Wawra, H. (1963). "Messungen der Absolutintensität der diffusen Röntgenkleinwinkelstreuung durch mechanische Schwächung des Primärstrahls bei Anwendung der Impulszähltechnik". (Correção: Kratky, O.; & Wawra, H. (1963). "Errata". *Monatsh. Chem.* **94**(6): 1434-4475) *Monatsh. Chem.* **94**(5): 981-987.

Krukenberg, K. A.; Förster, F.; Rice, L. M.; Sali, A.; & Agard, D. A. (2008). "Multiple Conformation of *E. coli* Hsp90 in Solution: Insights into the Conformation Dynamics of Hsp90". *Structure* **16**: 755-765.

Kull, F. J.; Sablin, E. P.; Lau R.; Fletterick, R. J.; & Vale, R. D. (1996). "Crystal structure of the kinesin motor domain reveals a structural similarity to myosin". *Nature* **380**(6574): 550–555.

Kuroda, S.; Nakagawa, N.; Tokunaga, C.; Tatematsu, K.; & Tanizawa, K. (1999). "Mammalian homologue of the *Caenorhabditis elegans* UNC-76 protein involved in axonal outgrowth is a protein kinase C zeta-interacting protein". *J. Cell. Biol.* **114**: 403-411.

Langer, T.; Lu, C.; Echols, H.; Flanagan, J.; Hayer, M.K.; & Hartl, F.U. (1992). "Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding". *Nature* **356**: 683-689.

Lanza, D. C.; Silva, J. C.; Assmann, E. M.; Quaresma, A. J.; Bressan, G. C.; Torriani, I. L.; & Kobarg, J. (2009). "Human FEZ1 has characteristics of a natively unfolded protein and dimerizes in solution". *Proteins* **74**: 104-121.

Layne, E. (1957). "Spectrophotometric and Turbidimetric Methods for Measuring Proteins". *Methods in Enzymology* **3**: 447-455.

Lehninger, A. L.; Nelson, D. L.; & Cox, M. M. (1995), "Princípios de Bioquímica", 4ª. Ed., Sarvier, São Paulo.

Li, J.; Qian, X.; & Sha, B.D. (2003). "The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate". *Structure* **11**: 1475-1483.

Lipfert, J.; & Doniach, S., (2007). "Small-Angle X-Ray Scattering from RNA, Proteins, and Protein Complexes". *Ann. Rev. Biophys. Biomol. Struct.* **36**: 307-327.

Louhivouri, M.; Pääkkonen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; & Annala, A. (2003). "On the origin of residual dipolar couplings from denatured proteins". *J. Am. Chem. Soc.* **125**: 15647-15650.

Lovell, S. C.; Davis, I. W.; Arendall III, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C.(2003). "Structure validation by Calpha geometry: phi, psi and Cbeta deviation". *Proteins* **50**: 437-450.

Mayer, M.P.; & Bukau, B. (2005). "Hsp70 chaperones: Cellular functions and molecular Mechanism". *Cell. Mol. Life Sci.* **62**(6):670-684.

Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; & Wright P. E. (2004). "Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings". *J. Mol. Biol.* **34**: 1131-1142.

Moncoq, K.; Broutin, I.; Craescu, C. T.; Vachette, P.; Ducruix, A.; & Durand, D. (2004). "SAXS Study of the PIR Domain from the Grb14 Molecular Adaptor: A Natively Unfolded Protein with a Transient Structure Primer?". *Biophys. J.* **87**: 4056-4064.

Moore, P. B. (1980). "Small-angle scattering: information content and error analysis". *J. Appl. Cryst.* **13**: 168-175.

Mylonas, E.; & Svegun, D. I. (2007). "Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering". *J. Appl. Cryst.* **40**: s245-s249.

Nery, F. C.; Bressan, G. C.; Alborghetti, M. R.; Passos, D. O.; Kuniyoshi, T. M.; Ramos, C. H., Oyama, S. Jr.; & Kobarg, J. (2006a). "A spectroscopic analysis of the interaction between the human regulatory proteins RACK1 and Ki-1/57". *Biol. Chem.* **387**(5): 577-582.

Nery, F. C.; Rui, E.; Kuniyoshi, T. M.; & Kobarg, J. (2006b). "Evidence for the interaction of the regulatory protein Ki-1/57 with p53 and its interacting protein". *Biochem. Biophys. Res. Commun.* **341**: 847-855.

Nery, F. C.; Passos, D. O.; Garcia, V. S.; & Kobarg, J. (2004). "Ki-1/57 interacts with RACK1 and is a substrate for the phosphorylation by phorbol 12-myristate 13-acetate-activated protein kinase C". *J. Biol. Chem.* **279**(12): 11444-11455.

Neuenschwander, R. T.; & Tolentino, H. C. N. (2004). "Bent Crystal Monochromator with Constant Crystal Center Position and 2-theta Arm for a Dipersive Beamline". *AIP Conf. Proc.* **705**: 647-650.

Neuenschwander, R. T.; & Tavares, W. S. (2001). *Internal Technical Communication. Internal Technical Note, LNLS, Campinas, SP, Brazil.*

Orthaber, D.; Bergmann, A.; & Glatter, O. (2000). "SAXS experiments on absolute scale with Kratky systems using water as a secondary standard". *J. Appl. Cryst.* **33**: 218-225.

Passos, D. O.; Quaresma, A. J. C.; & Kobarg, J. (2006). "The methylation of the C-terminal region of hnRNPQ (NSAP1) is important for its nuclear localization". *Biochem. Biophys. Res. Commun.* **346**(2): 517-525.

Petoukhov, M. V.; Monie, T. P.; Allain, F. H.-T.; Matthews, S.; Curry, S.; & Svergun, D. I. (2006). "Conformation of Polypyrimidine Tract Binding Protein in Solution". *Structure* **14**: 1021-1027.

Petoukhov, M. V.; & Svergun, D. I. (2005). "Global rigid body modeling of macromolecular complexes against small-angle scattering data". *Biophys. J.* **89**: 1237-1250.

Petoukhov, M. V.; Eady, N. A.; Brown, K. A.; & Svergun, D. I. (2002). "Addition of missing loops and domains to protein models by x-ray solution scattering". *Biophys. J.* **83**: 3113-3125.

Pilz, I.; Glatter, O.; & Kratky, O. (1972). "Röntgenkleinwinkelstudien über die Substruktur von *Helixpomatia* Hämocyanin". *Zeitschrift für Naturforschung* **27b**: 518-524.

Porod, G. (1951). "Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. I. Teil. (X-ray low angle scattering of dense colloid systems. Part I.). *Kolloid-Zeitschrift* **124**: 83-114.

Putnam, C. D.; Hammel, M.; Hura, G. L.; & Tainer, J. A. (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution". *Q. Rev. Biophys.* **40**(3):191-285.

Quaresma, A. J. C.; Bressan, G. C.; Gava, L. M.; Lanza, D. C.; Ramos, C. H.; & Kobarg, J. (2009). "Human hnRNP Q re-localizes to cytoplasmic granules upon PMA, thapsigargin, arsenite and heat-shock treatments". *Exp. Cell Res.* **315**: 968-980.

Quaresma, A. J. C.; Oyama Jr., S.; Barbosa, J. A.; & Kobarg J. (2006). "The acidic domain of hnRNPQ (NSAP1) has structural similarity to Barstar and binds to Apobec1." *Biochem. Biophys. Res. Commun.* **350**: 288-297.

Ramos, C. H. I.; Oliveira, C. L. P.; Fan, C. Y.; Torriani, I. L.; & Cyr, D. M. (2008). "Conserved Central Domains Control the Quaternary Structure of Type I and Type II Hsp40 Molecular Chaperones". *J. Mol. Biol.* **383**: 155-166.

Rohde, D.; Hansen, H.; Hafner, M.; Lange, H.; Mielke, V.; Hansmann, M. L.; & Lemke, H. (1992). Cellular localizations and processing of the two molecular forms of the Hodgkin-associated Ki-1 (CD30) antigen. The protein kinase Ki-1/57 occurs in the nucleus". *Am. J. Pathol.* **140**, 473-482.

Russel, T. P. (1991). "In: *Handbook of Synchrotron Radiation*", Brown, G. S. and Moncton, D. E., eds., North-Holland, Amsterdam, **1**: 379-469.

Russel, T. P., (1983). "An absolute intensity standard for small-angle X-ray scattering measured with positron-sensitive detectors". *J. Appl. Cryst.* **16**: 473-478.

Sakae, N.; Yamasaki, M.; Kitaichi, K.; Fukuda, T.; Yamada, M.; Yoshikawa, H.; Hiranita, T.; Tatsumi, Y.; Kira, J.; Yamamoto, T.; Miyakawa, T.; & Nakayama, K. I. (2008). "Mice lacking the schizophrenia-associated protein FEZ1 manifest hyperactivity and enhanced responsiveness to psychostimulants". *Hum. Mol. Genet.* **17**(20): 3191-3203.

Semenyuk, A. V.; & Svergun, D. I. (1991). *GNOM – a program package for small-angle scattering data processing*. *J. Appl. Cryst.* **24**: 537-540.

Sha, B.; Lee, S; & Cyr, D. M. (2000) "The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1". *Structure* **8**: 799-807.

Shannon, C. E.; & Weaver, W. (1949). "The Mathematical Theory of Communication". Urbana, IL: University of Illinois Press.

Shortle, D.; & Acherman, M. S. (2001). "Persistence of Native-like Topology in a Denatured Protein in 8M urea". *Science* **293**: 487-489.

Silva, J. C.; Lanza, D. C. F.; Kobarg, J.; & Torriani, I. L. (2010a). SAXS structural analysis reveals conformational changes of the natively unfolded protein FEZ1. *Manuscripto submetido*.

Silva, J. C.; Borges, J. C.; Cyr, D. M.; Ramos, C. H. I.; & Torriani, I. L. (2010b), "SAXS studies of heat shock proteins (Hsp40) Sis1, Ydj1 and mutants: dimerization, flexibility and functionality". *Manuscripto submetido*.

Silva, J. C.; Quaresma, A. J. C.; Kobarg, J.; & Torriani, I. L. (2010c), "SAXS study of the human regulatory protein hnRNP-Q in solution". *Manuscripto submetido*.

Singh, G.P.; & Dash, D. (2007). "Intrinsic disorder in yeast transcriptional regulatory network". *Proteins* **68**, 602-605.

Song, Y.; & Masison, D.C. (2005). "Independent regulation of Hsp70 and Hsp90 chaperones by Hsp70/Hsp90-organizing protein Sti1 (Hop1)". *J. Biol. Chem.* **280**: 34178–34185.

Stabinger, H.; & Kratky, O. (1978). "A new technique for the measurement of the absolute intensity of x-ray small angle scattering. The moving slit method". *Makromol. Chem.* **179**: 1655-1659.

Stoscheck, C. M. (1990). "Quantitation of Protein". *Methods in Enzymology* **182**: 50-69.

Stuhrmann, H. B. (1970). "Interpretation of small-angle scattering of dilute solutions and gases. A representation of the structures related to a one-particle scattering functions". *Acta Crystallogr.* **A26**: 297-306.

Svergun, D. I. (1991). "Mathematical methods in small-angle scattering data analysis". *J. Appl. Crystallogr.* **24**: 485-492.

Svergun, D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria". *J. Appl. Cryst.* **25**: 495-503.

Svergun, D. I. (1994). "Solution scattering from biopolymers: advanced contrast variation data analysis". *Acta Crystallogr.* **A50**: 391-402.

Svergun, D. I.; Barberato, C.; & Koch, M. H. J. (1995). CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**: 768–773.

Svergun, D. I. (1999). "Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing". *Biophysical Journal* **76**: 2879-2886.

Svergun, D. I.; & Koch, M. H. J. (2003). "Small-Angle scattering studies of biological macromolecules in solution". *Rep. Prog. Phys.* **66**: 1735-1782.

Tang, K. T.; & Karplus, M. (1971). "Quantum Theory of (H, H<sub>2</sub>) Scattering: Approximate Treatments of Reactive Scattering". *Physical Review A* **4**(5): 1844-1858.

Taupin, D.; & Luzzati, V. (1982). "Informational content and retrieval in solution scattering studies. I. Degrees of freedom and data reduction". *J. Appl. Cryst.* **15**: 289-300.

Teraoka, I. (2002). *Polymer Solutions: An Introduction to Physical Properties*. John Wiley & Sons, Inc., New York, NY, USA.

Tidow, H.; Melero, R.; Mylonas, E.; Freund, S. M. V.; Grossmann, J. G.; Carazo, J. M.; Svergun, D. I.; Valle, M.; & Fersht, A. R. (2007). "Quaternary structures of tumor suppressor p53 and a specific p53-DNA complex". *Proc. Natl. Acad. Sci. USA* **104**: 12324-12329.

Tompa, P. (2005). "The interplay between structure and function in intrinsically unstructured proteins". *FEBS Lett.* **579**, 3346-3354.

Tompa, P. (2002). "Intrinsically unstructured proteins". *Trends in Biochemical Sciences* **27**: 527-533.

Torriani, I. L.; Plivelic, T. S.; Silva, J. C.; & Neuenschwander, R. T. (2008). "SAXS-WAXS Experiments at Small-Angle X-ray Scattering Beamline of the LCLS". *Sixth International Conference on Synchrotron Radiation in Materials Science - SRMS-6, Livro de resumos.*

Uversky, V. N. (2002a). "Natively unfolded proteins: a point where biology waits for physics". *Protein Sci.* **11**: 739-756.

Uversky, V. N. (2002b). "What does it mean to be natively unfolded". *Eur. J. Biochem.* **269**: 2-12.

Uversky, V.N.; & Fink, A.L. (2002). "Conformational behavior of human alpha-synuclein is modulated by familial Parkinson's disease point mutations A30P and A53T". *Neurotoxicology* **23**: 553-567.

Uversky, V.N.; Gillespie, J.R.; Millett, I.S.; Khodyakova, A.V.; Vasiliev A.M.; Chernovskaya T.V.; Vasilenko R.N.; Kozlovskaya G.D.; Dolgikh D.A.; Fink A.L.; Doniach S.; & Abramov, V.M. (1999). "Natively Unfolded" Human Prothymosin a Adopts Partially-Folded Conformation At Acidic pH. *Biochemistry*". *Biochemistry* **38**: 15009-15016.

Vachette, P.; & Svergun, D. I. (2000). "Small-Angle X-ray Scattering by Solutions of Biological Macromolecules". In: "Structure and Dynamics of Biomolecules". Edited by E. Fanchon, G. Geissler, J.-L. Hodeaus, J.-R. Regnard & P. A. Timmins. New York: Oxford University Press, 199-237.

Voet, D.; & Voet, J. G. (1995), "Biochemistry", 2<sup>nd</sup>. ed., New York: J. Wiley.

Volkov, V. V.; & Svergun, D. I. (2003). "Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**: 860-864.

Wall, D.; Zylicz, M.; & Georgopoulos, C. (1994). "The NH<sub>2</sub>-terminal 108 amino acids of the *Escherichia coli* DnaJ protein stimulate the ATPase activity of DnaK and are sufficient for  $\lambda$  replication". *J. Biol. Chem.* **269**(7): 5446-5451.

Wright, P. E.; & Dyson, H. J. (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm". *J. Mol. Biol.* **293**(2): 321-331.

Wu, Y.-Y. & Ohmura, T. (1962). "Quantum Theory of Scattering", *Prentice Hall International Series in Physics, Englewood Cliffs, Prentice Hall, New York*

Yamada, K.; Nakamura, K.; Minabe, Y.; Iwayama-Shigeno, Y.; Takao, H.; Toyota, T.; Hattori, E.; Takei, N.; Sekine, Y.; Suzuki, K.; Iwata, Y.; Miyoshi, K.; Honda, A.; Baba, K.; Katayama, T.; Tohyama, M.; Mori, N.; & Yoshikawa, T. (2004). "Association analysis of FEZ1 variants with schizophrenia in Japanese cohorts". *Biol. Psychiatry* **56**: 683-690.