



Universidade Estadual de Campinas
Instituto de Física Gleb Wataghin
Departamento de Física da Matéria Condensada

Tese de Doutorado

Estudos de Espalhamento de Raios X a Baixos Ângulos por Sistemas Biológicos: Teoria e Aplicações

*Este exemplar corresponde à redação final da tese
de Doutorado defendida pelo aluno Cristiano Luis
Pinto de Oliveira e aprovada pela Comissão Julgadora.*

22/06/2005
[Signature]

Cristiano Luis Pinto de Oliveira
IFGW - UNICAMP

Orientadora: Profa. Dra. Iris L. Torriani
IFGW – UNICAMP

Campinas – SP
Abril de 2005

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IFGW - UNICAMP

OL4e Oliveira, Cristiano Luis Pinto de
 Estudos de espalhamento de Raios-X a baixos
 ângulos : teoria e aplicações / Cristiano Luis Pinto de
 Oliveira. -- Campinas, SP : [s.n.], 2005.

 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. Modelagem ab initio. 3. Proteínas – Estrutura.
 I. Torriani, Iris Concepción Linares de. II. Universidade
 Estadual de Campinas. Instituto de Física "Gleb
 Wataghin". III. Título.

(vsv/ifgw)

- **Título em inglês:** Small angle X-ray scattering from biological systems : theory and applications
- **Palavras-chave em inglês (Keywords):**
 - 1. Small angle X-ray scattering
 - 2. Ab initio modeling
 - 3. Protein structure
- **Área de concentração:** Física da Matéria Condensada
- **Titulação:** Doutor em física
- **Banca examinadora:**
 - Prof^a Iris Concepción Linares de Torriani
 - Prof^a Maria Cristina dos Santos
 - Prof. Paulo Mascanello Bisch
 - Prof. José Antônio Brum
 - Prof. Fernando Cerdeira
- **Data da defesa:** 28.04.2005



MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE
CRISTIANO LUÍS PINTO DE OLIVEIRA – RA 950364 APRESENTADA E
APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE
ESTADUAL DE CAMPINAS, EM 28 / 04 / 2005.

COMISSÃO JULGADORA:

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

Profa. Dra. Maria Cristina dos Santos – IF/USP

Prof. Dr. Paulo Mascarello Bisch – IBCCF / UFRJ

Prof. Dr. José Antônio Brum – DFMC/IFGW/UNICAMP

Prof. Dr. Fernando Cerdeira – DFMC/IFGW/UNICAMP

Agradecimentos,

À Jeová, nosso Deus, pela dádiva da vida.

À minha amada esposa Leire, e a nosso amado filho Gudryan, pelo apoio e paciência na execução deste trabalho.

A minha sogra Mariza e meu cunhado Lezer pelo apoio e amizade.

À minha orientadora e amiga Iris Torriani que, com paciência e dedicação, me orientou no desenvolvimento desta tese.

Aos meus colegas do Laboratório de Cristalografia Aplicada e Raios X na UNICAMP, Dr. Carlos Leite, Ana Carolina e José Luis, muito obrigado pelo apoio ao longo de todos estes anos de convívio.

Aos meus colaboradores nos diversos experimentos mostrados nesta tese que com seu trabalho possibilitaram a realização de meu trabalho.

Ao Dr. Jan Skov Pedersen quando de meu estágio na Universidade de Aarhus, Dinamarca, me auxiliando no desenvolvimento de meu projeto.

Aos meus colegas do Laboratório Nacional de Luz Síncrotron que possibilitaram as condições para que eu pudesse realizar os meus experimentos.

À FAPESP por ter acreditado em mim, apoiando meu doutorado.

A todos vocês os meus mais sinceros e profundos agradecimentos.

Cristiano Luis Pinto de Oliveira

... a sabedoria está com os modestos,

Provérbios. 11:2

RESUMO

Apresentamos nesta tese estudos relacionados à obtenção de parâmetros estruturais e modelagem *ab initio* para macromoléculas biológicas e sistemas densos de partículas elipsoidais a partir de dados de espalhamento de raios X a baixos ângulos. Aspectos teóricos e experimentais são discutidos, complementados por simulações computacionais e pela apresentação de resultados experimentais que exemplificam a aplicação dos métodos de cálculo revelando assim os alcances e limites da técnica. Progressos recentes em biologia molecular estrutural revelam que a correlação entre a forma e a função de macromoléculas biológicas é crucial para a compreensão dos processos biológicos. Além disso, diversos estudos associam várias doenças a defeitos estruturais em proteínas, aumentando ainda mais o interesse na determinação das estruturas protéicas. A técnica de espalhamento de raios X a baixos ângulos permite a obtenção de parâmetros estruturais de proteínas, apresentando diversas vantagens sobre a cristalografia de proteínas e a ressonância magnética nuclear, mesmo não atingindo alta resolução.

Trataremos, portanto, dois assuntos principais: estudos de sistemas diluídos e estudos de sistemas densos. Para o estudo de sistemas diluídos o enfoque principal será na modelagem estrutural de proteínas em solução, onde são apresentados e discutidos diversos métodos de simulação tanto da estrutura das macromoléculas quanto de seus parâmetros hidrodinâmicos. Estes métodos foram aplicados ao estudo de diversos problemas, tais como: (i) as estruturas das hemoglobinas extracelulares de *Biomphalaria glabrata* e *Glossoscolex paulistus*, proteínas gigantes com elevada cooperatividade funcional entre suas subunidades; (ii) as diferenças estruturais entre as formas livres e ligadas a lipídeos da proteína básica de mielina, uma das proteínas principais do axônio nervoso e cujo mal funcionamento leva a doenças neuronais; (iii) a interação entre proteínas e polímeros, tema relevante para processos de cristalização e construção de sistemas bio-poliméricos; e (iv) a estrutura da proteína precursora amilóide, proteína transmembrana que, apesar de diversas funções conhecidas, é uma das precursoras principais do Mal de Alzheimer.

Já no estudo de sistemas densos é apresentada uma nova metodologia baseada em simulações de Monte Carlo para elipsóides de revolução. Como elipsóides podem servir de sistemas modelos para diversos sistemas reais, simulações de propriedades de soluções concentradas destes sistemas são de grande interesse. Com base nestas simulações, funções de espalhamento e funções de correlação puderam ser calculadas para um grande intervalo de anisotropias e concentrações, permitindo a obtenção de funções numéricas que podem ser utilizadas no estudo de sistemas reais uma vez que não existem expressões analíticas para este tipo de sistema. Como aplicação, este método foi utilizado no estudo de nanopartículas de ferro em matriz de silício, fornecendo parâmetros estruturais do sistema.

ABSTRACT

In this dissertation, theoretical and experimental aspects of the calculation of structural parameters and ab-initio modeling for biological macromolecules and condensed particle systems are being discussed. Two main applications will be presented, complemented by computational simulations and many experimental results. For dilute systems the focus will be on the structural modeling of several proteins in solutions, including the calculation of hydrodynamic parameters. Recent developments in structural molecular biology reveal that knowledge about the correlation existing between the structure and function of biological macromolecules is crucial for the understanding of biological processes. Many studies have proved that structural defects detected in certain proteins are responsible for serious diseases. Small angle X-ray scattering has proved to be very useful in the determination of the low resolution structure of macromolecules in solution. This technique offers clear advantages with respect to other techniques, like single crystal X-ray crystallography and nuclear magnetic resonance.

The low resolution structural studies of proteins included: (i) the 3D solution structure of the extracellular hemoglobin from the species *Biomphalaria glabrata* and *Glossoscolex paulistus*, giant proteins with high functional cooperativity between their subunits; (ii) the determination of the structural differences between the lipid free and lipid bound forms of myelin basic protein (MBP), one of the most important proteins of the axon sheath whose bad functioning leads to neural diseases; (iii) the study of protein-polymer interactions, relevant subject for crystallization processes and development of biopolymeric systems; (iv) the 3D solution structure of the amyloid precursor protein, APP, transmembrane protein which is related to the Alzheimer disease.

For the study of dense systems, a Monte Carlo based method for the simulation of a system of hard ellipsoids of revolution was developed. Ellipsoidal particles can be used as model systems for many real problems and simulations of the properties of concentrated solutions are of great interest. As a result of these simulations, scattering functions and correlation functions could be derived for a wide interval of anisotropies, permitting the calculation of numerical functions that can be used in real systems. As an example, this method was used for the study of iron nanoparticles in a silica matrix, leading to the determination of mean radius and size distribution of the particles.

Índice Geral

Capítulo I – Aspectos Teóricos

1.	Aspectos teóricos de Espalhamento a Baixos Ângulos	2
1.1.	Espalhamento por um elétron	2
1.2.	Espalhamento por um átomo	6
1.3.	Espalhamento de ondas planas pela matéria	7
1.4.	Espalhamento de raios X por partículas	10
1.4.1.	Equação de Debye	10
1.4.2.	Função de Correlação	11
1.4.3.	Espalhamento e Absorção de Raios X	15
1.5.	Espalhamento por um conjunto de partículas	16
1.5.1.	Monodispersidade e polidispersidade	20
1.5.2.	O conceito de contraste	21
1.5.3.	Sistemas concentrados e diluídos	22
1.6.	Sistemas Monodispersos	24
1.6.1.	Parâmetros Gerais	24
1.6.2.	Grau de Compacidade	28
1.6.3.	Função de distribuição de distâncias e anisometria de partículas	33
1.6.4.	Informação contida em curvas de espalhamento a baixos ângulos	36
1.6.5.	Metodologia <i>Ab initio</i> para modelagem da forma da partícula	37
1.6.5.1.	Método de esféricos Harmônicos	38
1.6.5.2.	Método de elementos finitos (<i>dummy atom models</i>)	40
1.6.5.3.	Método de pseudo resíduos (<i>chain like models</i>)	42
1.6.5.4.	Promediação de modelos e imposições de simetria e anisotropia	44
1.6.6.	Cálculo de Intensidades de Espalhamento a partir de Modelos atômicos ..	45
1.6.7.	Construindo modelos por rotação de corpo rígido	46
1.6.8.	Determinação do perfil de densidade de espalhamento por SAXS	47
2.	Cálculos de Parâmetros Hidrodinâmicos	50
2.1.	Introdução	50
2.2.	Propriedades Hidrodinâmicas – Descrição Geral	51
2.2.1.	Coeficiente de fricção	51
2.2.2.	Coeficiente de difusão translacional	52
2.2.3.	Coeficiente de sedimentação	53
2.2.4.	Viscosidade Intrínseca	55
2.3.	Simulação de Propriedades Hidrodinâmicas Para Modelos de Partículas	57

Capítulo II – Parte Experimental

1.	Aspectos Experimentais	60
1.1.	Introdução	60
1.2.	Linha de Espalhamento de Raios X a Baixos ângulos do LNLS	61
1.3.	Métodos para medida de intensidade absoluta	63
1.3.1.	Utilização da água como padrão secundário	65
1.3.2.	Utilização de uma proteína como padrão secundário	67
1.3.3.	Parâmetros Obtidos da Renormalização em Escala Absoluta	69

1.4.	Distorções da intensidade espalhada	71
1.4.1.	Influência das dimensões do feixe e detetor	72
1.4.2.	Policromaticidade do feixe	75
1.4.3.	Erros Estatísticos	75
1.4.4.	Expressão geral para a Intensidade Experimental	76
1.4.5.	Efeitos de colimação na linha de SAXS do LNLS	77
1.5.	Tratamento de Dados	81
1.5.1.	Programas Para Tratamento de Dados	86
2.	Cálculo de Parâmetros Hidrodinâmicos para Modelos ab initio	87
2.1.	Introdução	87
2.2.	Modelos Ab initio	89
2.3.	Estimativa do raio ideal para modelos chain model	91
2.4.	Modelagem tipo chain model assumindo grupos de resíduos	95
2.5.	Conclusões	100

Capítulo III – Estudos Estruturais de Proteínas

1.	Espalhamento a Baixos ângulos por proteínas com alto peso molecular	102
1.1.	Espalhamento de raios X a baixos ângulos para hemoglobinas de <i>B. glabrata</i> ..	103
a)	Introdução	103
b)	Experimental	103
c)	Resultados	104
d)	Cálculo de modelos	108
e)	Considerações finais	112
1.2.	Forma e estrutura da hemoglobina extracelular de <i>G. paulistus</i> por SAXS	115
a)	Introdução	115
b)	Parte experimental e métodos de cálculo utilizados	116
c)	Resultados	117
d)	Considerações sobre os resultados com detetor unidimensional	121
e)	Medidas em alta resolução para proteína <i>G. paulistus</i>	122
2.	Estudo da proteína Básica da membrana de Mielina por SAXS	125
a)	Introdução	125
b)	Experimental e Metodologia	126
c)	Resultados e Discussão	127
d)	Considerações finais	134
3.	Estudo da interação da lisozima com polietilenos e copolímeros	135
a)	Introdução	135
b)	Materiais e Métodos	137
c)	Aspectos Teóricos da Análise de SAXS	138
d)	Resultados Experimentais	143
e)	Considerações finais sobre o estudo	165
4.	Estudo da proteína precursora amilóide por SAXS	166
a)	Introdução	166
b)	Experimental	169
c)	Considerações Finais do Estudo com a Proteína Precursora Amilóide	201

Capítulo IV – Simulações para sistemas Densos

1. Cálculo de Intensidade de Espalhamento, Fator de Estrutura e Função de Correlação para Sistemas Densos	203
1.1. Introdução	203
1.2. Método Monte Carlo	204
1.2.1. Introdução	204
1.2.2. Cálculo de Funções de Espalhamento	206
1.3. Método Monte Carlo para Simulação de Elipsóides em Solução	209
1.3.1. Introdução	209
1.3.2. Simulações	210
1.3.3. Resultados	213
1.3.4. Conclusões Sobre os Resultados	234
1.4. Exemplo de Aplicação – Estudo de partículas de ferro em matriz de silício	235

Capítulo V – Considerações Finais e Bibliografia

1. Conclusões e Perspectivas	240
2. Referências e Bibliografia	243

Apêndice

A. Proteínas	252
A.1. Introdução	252
A.2. Aminoácidos e Cadeias Polipeptídicas	254
A.3. Estrutura Tridimensional de Proteínas	256
A.3.1. Estrutura Primária	257
A.3.2. Estrutura Secundária	258
A.3.2.1. A α -Hélice	260
A.3.2.2. Folhas β	261
A.3.2.3. Estruturas não repetitivas	263
A.3.2.4. Alinhamento de seqüências e predição de estrutura secundária	263
A.4. Proteínas Fibrosas	264
A.5. Proteínas Globulares	266
A.5.1. Estrutura Terciária	267
A.5.2. Estrutura Quaternária	268
A.6. Considerações Finais	269

Índice de Figuras

Figura 1 – Sistema de eixos utilizado no desenvolvimento teórico.	5
Figura 2 – Espalhamento de primeira ordem de uma onda plana por um potencial limitado no espaço.	7
Figura 3 – Definição do vetor de espalhamento q	8
Figura 4 – Perfil bidimensional de espalhamento para uma partícula esférica.	13
Figura 5 – Perfil bidimensional de espalhamento para uma partícula anisotrópica.	14
Figura 6 – Intensidade teórica de SAXS para random coils.	29
Figura 7 – Gráficos de Kratky para random coils.	30
Figura 8 – Intensidade de SAXS para partículas globulares compactas e com interface bem definida com o meio.	30
Figura 9 – Gráfico de Kratky para partículas globulares compactas com interface bem definida.	31
Figura 10 – Modelos esquemáticos para proteínas em diferentes graus de enovelamento.	31
Figura 11 – Variação do Gráfico de Kratky para uma proteína.	32
Figura 12 – Intensidades de espalhamento e funções de distribuição de distâncias para corpos simples.	33
Figura 13 – Intensidades de espalhamento e funções de distribuição de distâncias para corpos ocos.	34
Figura 14 – Modelagem ab initio usando esféricos harmônicos.	40
Figura 15 – Ajuste por modelos tipo "dummy atom".	42
Figura 16 – Ajuste por modelos tipo "chain atom".	43
Figura 17 – Diagrama de Transformadas de Fourier em espalhamento a baixos ângulos.	48
Figura 18 – Esquema de ultracentrífuga.	53
Figura 19 – Representação esquemática da construção do modelo hidrodinâmico em um análogo bidimensional.	57
Figura 20 – Representação esquemática de um experimento de SAXS e a transformação de Fourier do espaço real para o espaço recíproco.	60
Figura 21 – Esquema da estação experimental de SAXS do LNLS.	61
Figura 22 – Espalhamento da água para calibração.	67
Figura 23 – Efeitos de "smearing" causados por colimação de fendas (esquema).	73
Figura 24 – Espectro bidimensional teórico para o espalhamento de um sistema monodisperso de esferas.	78
Figura 25 – Simulação da distorção causada pela integração linear ao longo da fenda.	79
Figura 26 – Efeito de "smearing" causado pela integração na largura da fenda.	80
Figura 27 – Subtração do espalhamento parasita.	82
Figura 28 – Curva típica de espalhamento de uma solução de BSA, espalhamento do solvente e a curva de diferença.	83
Figura 29 – Modelos ab initio para espalhamento da lisozima.	90
Figura 30 – Modelos ab initio para espalhamento da albumina.	91
Figura 31 – Estimativa do raio ideal para cálculo de parâmetros hidrodinâmicos.	91
Figura 32 – Simulação ab initio para lisozima variando-se o raio das subunidades de modo a manter-se o mesmo volume.	97
Figura 33 – Purificação da hemoglobina de <i>Biomphalaria glabrata</i> à pH5.0 e pH7.5.	105
Figura 34 – Curva de espalhamento, ajuste teórico e função de distribuição de distâncias para hemoglobina de <i>B. glabrata</i>	106

Figura 35 – Extrapolação para intensidade na origem para diferentes concentrações de proteína.....	107
Figura 36 – Envelope externo para a hemoglobina de <i>B. glabrata</i> calculada utilizando expansão multipolar sem imposição de simetria.....	109
Figura 37 – Envelope externo para a hemoglobina de <i>B. glabrata</i> calculada utilizando expansão multipolar com simetria P222.....	110
Figura 38 – Modelo mais provável calculado para proteína <i>B. glabrata</i> por modelagem tipo dummy atom.....	112
Figura 39 – Modelo esquemático proposto para a hemoglobina de <i>B. glabrata</i>	114
Figura 40 – Curva experimental de espalhamento normalizada para seção de choque de uma partícula e ajuste teórico para hemoglobina de <i>G. paulistus</i>	118
Figura 41 – Modelo ab initio calculado com espaço inicial de busca esférico para <i>G. paulistus</i>	119
Figura 42 – Modelo ab initio calculado com espaço inicial de busca cilíndrico para proteína de <i>G. paulistus</i>	120
Figura 43 – Espectro Bidimensional de espalhamento para <i>G. paulistus</i> a 20mg/ml.....	122
Figura 44 – Resultados experimentais para <i>G. paulistus</i>	123
Figura 45 – Modelo ab initio tipo chain model para proteína de <i>G. paulistus</i>	124
Figura 46 – Curvas de SAXS para proteína básica de mielina nas formas LB-MBP e LF-MBP.....	128
Figura 47 – Modelagem da LB-MBP por perfil de densidade eletrônica.....	130
Figura 48 – Modelagem da LB-MBP por chain models.....	132
Figura 49 – Comparação do espalhamento da LB-MBP com modelo atômico teórico.....	133
Figura 50 – Superposição do modelo ab initio ao modelo teórico.....	133
Figura 51 – Superposição do modelo ab initio ao perfil de densidade eletrônica obtido pelo ajuste com modelo de esferas concêntricas.....	134
Figura 52 – Simulação de estruturas compactas e estruturas tipo “random coil”.....	143
Figura 53 – Experimentos de SAXS para lisozima com PEO3350.....	145
Figura 54 – Experimentos de SAXS para lisozima com F38.....	146
Figura 55 – Experimentos de SAXS para lisozima com P105.....	148
Figura 56 – Espalhamento da lisozima pura com variação de temperatura.....	149
Figura 57 – Cinética de agregação da lisozima à 80°C.....	150
Figura 58 – Espalhamento da lisozima+PEO3350 com variação de temperatura.....	151
Figura 59 – Cinética de agregação da lisozima + PEO3350 à 80°C.....	152
Figura 60 – Intensidades de SAXS para P105 e o sistema composto Lisozima + P105 para várias temperaturas.....	153
Figura 61 – Cinética de agregação da lisozima + P105 à 80°C.....	155
Figura 62 – Diferenças da formação de micelas na solução com e sem proteína para várias temperaturas.....	157
Figura 63 – Variação dos parâmetros dimensionais da lisozima com e sem polímeros em função da temperatura.....	158
Figura 64 – Variação do raio de giro do sistema lisozima+P105 em comparação com o esperado para um sistema com dois fatores de forma.....	159
Figura 65 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 60°C.....	161
Figura 66 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 70°C.....	162

Figura 67 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 80°C.....	163
Figura 68 – Estrutura de domínios e clivagem da APP..	167
Figura 69 – Espalhamento a baixos ângulos para sAPP α_{695}	170
Figura 70 – Cálculo ab initio para sAPP α_{695} utilizando esféricos harmônicos.....	171
Figura 71 – Cálculo ab initio para sAPP α_{695} utilizando modelagem tipo bead model..	172
Figura 72 – Experimentos de SAXS para sAPP α pura e com agentes denaturantes.	175
Figura 73 – Desenovelamento da sAPP α_{695} utilizando dicroísmo circular.	178
Figura 74 – Experimentos de SAXS em alta resolução para sAPP α ..	180
Figura 75 – Estrutura de domínios da APP e sAPP α ..	181
Figura 76 – Estrutura secundária predita para as isoformas sAPP α_{695} e sAPP α_{770} com base na seqüência primária.....	182
Figura 77 – Modelagem ab initio em alta resolução para as proteínas sAPP α_{695} e sAPP α_{770}	184
Figura 78 – Modelagem ab initio em alta resolução para as proteínas sAPP α_{695} e sAPP α_{770}	186
Figura 79 – Superposição dos modelos ab initio obtidos para a proteína sAPP α_{695} e sAPP α_{770}	187
Figura 80 – Proposta para adição de um domínio globular de 75 resíduos ao modelo ab initio da sAPP α_{695} de modo a representar o modelo ab initio da sAPP α_{770} ..	188
Figura 81 – Rotação de corpo rígido para dois possíveis posicionamentos do dummy_prot com o respectivo ajuste dos dados..	189
Figura 82 – Estrutura com resolução atômica para a parcela 28-123 da sAPP α	190
Figura 83 – Estrutura com resolução atômica para a parcela 124-189 da sAPP α	191
Figura 84 – Estrutura com resolução atômica para a parcela 287-344 da sAPP α_{770}	191
Figura 85 – Estrutura com resolução atômica para a parcela 299-490 da sAPP α_{770} e 374-565 da sAPP α_{770}	192
Figura 86 – Modelo molecular construído pelo docking do domínio N-terminal ligador de heparina (1MWP.PDB) e o domínio ligador de cobre (1OWT.PDB).....	194
Figura 87 – Sobreposição dos domínios com resolução atômica sobre o modelo ab initio obtido por experimentos de SAXS para a proteína sAPP α_{695}	196
Figura 88 – Sobreposição dos domínios com resolução atômica sobre o modelo ab initio obtido por experimentos de SAXS para a proteína sAPP α_{770}	197
Figura 89 – Ajuste dos dados experimentais utilizando o modelo construído por domínios atômicos.....	198
Figura 90 – Sistema de coordenadas e definição da anisotropia do elipsóide.	210
Figura 91 – Rede inicial para o cálculo Monte Carlo,.....	210
Figura 92 – Cella com partículas randomicamente orientadas.	211
Figura 93 – Cálculos teóricos para as frações de volume possíveis em uma rede FCC de elipsóides de revolução.	214
Figura 94 – Simulação para $\varepsilon=1$ e $\eta=0.06$..	215
Figura 95 – Funções calculadas pela simulação de Monte Carlo	216
Figura 96 – Simulações de Monte Carlo para esferas rígidas.	217
Figura 97 – Simulação para $\varepsilon=1/2$ e $\eta=0.06$..	218
Figura 98 – Funções calculadas pela simulação de Monte Carlo.	220

Figura 99 – Simulação para $\varepsilon = 2$ e $\eta=0.06$.	221
Figura 100 – Funções calculadas pela simulação de Monte Carlo.	222
Figura 101 – Simulação para $\varepsilon = 1/3$ e $\eta=0.06$.	223
Figura 102 – Funções calculadas pela simulação de Monte Carlo.	224
Figura 103 – Simulação para $\varepsilon = 3$ e $\eta=0.06$.	225
Figura 104 – Funções calculadas pela simulação de Monte Carlo.	226
Figura 105 – Simulação para $\varepsilon = 1/5$ e $\eta=0.06$.	227
Figura 106 – Funções calculadas pela simulação de Monte Carlo.	228
Figura 107 – Simulação para $\varepsilon = 5$ e $\eta=0.06$.	229
Figura 108 – Funções calculadas pela simulação de Monte Carlo.	230
Figura 109 – Variação de $r_{\min}$ e $r_{\max}$ com a anisotropia da partícula ε .	232
Figura 110 – Relação entre o raio de esfera rígida e a anisotropia das partículas ε .	233
Figura 111 – Simulação de Monte Carlo para sistema polidisperso de esferas.	236
Figura 112 – Foto de microscopia eletrônico para o sistema em estudo.	237
Figura 113 – Resultados da simulação MC para o sistema estudado.	238

Índice de Tabelas

Tabela 1 - <i>Raio de Giro para variados tipos de corpos homogêneos.</i>	26
Tabela 2 - <i>Comprimentos de espalhamento de raios X para alguns elementos.</i>	70
Tabela 3 - <i>Comprimentos de espalhamento para os aditivos* mais comuns</i>	71
Tabela 4 - <i>Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados pelos modelos tipo chain model e bead model.</i>	92
Tabela 5 - <i>Parâmetros hidrodinâmicos para Albumina. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados pelos modelos tipo chain model e bead model.</i>	93
Tabela 6 - <i>Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados por médias dos modelos tipo bead model.</i>	94
Tabela 7 - <i>Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados por médias dos modelos tipo chain model</i>	94
Tabela 8 - <i>Raios para as simulações.</i>	97
Tabela 9 - <i>Razões entre os parâmetros hidrodinâmicos calculados para cada um dos modelos e os parâmetros obtidos pela estrutura cristalográfica 6LYZ.</i>	99
Tabela 10 - <i>Parâmetros hidrodinâmicos experimentais e comparação com valores obtidos para modelos ab initio.</i>	111
Tabela 11 - <i>Parâmetros Hidrodinâmicos e peso molecular para hemoglobina de G. Paulistus obtida por SAXS e calculada pelos modelos ab initio. Modelo 1: Espaço de busca inicial esférico. Modelo 2: Espaço de busca inicial cilíndrico.</i>	121
Tabela 12 - <i>Parâmetros Hidrodinâmicos e peso molecular para hemoglobina de G. Paulistus obtida por SAXS e calculada pelos modelos ab initio. Modelo 3: Modelo tipo chain model.</i>	124
Tabela 13 - <i>Descrição dos polímeros utilizados neste estudo.</i>	137
Tabela 14 - <i>Mudanças no raio de giro da proteína na presença dos polímeros PEO, F38 e P105 em várias condições experimentais.</i>	145
Tabela 15 - <i>Amostras estudadas por SAXS, concentração e solução tampão utilizada.</i> ...	174
Tabela 16 - <i>Raios de giro para as duas isoformas de sAPPα na condição nativa e com aditivos.</i>	176
Tabela 17 - <i>Cálculos hidrodinâmicos para modelos ab initio das isoformas de sAPPα.</i> .	184
Tabela 18 - <i>Cálculos hidrodinâmicos para modelos ab initio das isoformas de sAPPα.</i> .	187

CAPÍTULO I

ASPECTOS

TEÓRICOS

1. Aspectos teóricos de Espalhamento a Baixos Ângulos

A teoria de espalhamento a baixos ângulos é desenvolvida em detalhe em diversos livros texto (Guinier e Fournet, 1955; Glatter e Kratly, 1982; Feigin e Svergun, 1987 entre outros), contendo notações e desenvolvimentos ligeiramente diferentes, detalhando aspectos e peculiaridades da teoria. Neste sentido, pretendi apresentar um desenvolvimento teórico uniforme e consistente, baseado nestes e em outros livros que serão citados ao longo do texto.

1.1. Espalhamento por um elétron

Os raios-X são ondas eletromagnéticas caracterizadas por um campo elétrico cuja intensidade varia senoidalmente com o tempo em qualquer ponto do feixe. O campo elétrico exerce força em uma partícula carregada e, devido a seu caráter oscilatório, esta carga (um elétron, por exemplo) oscilará em torno de sua posição média pela influência deste campo elétrico.

A aceleração ou desaceleração de uma carga gera a emissão de uma onda eletromagnética secundária (Jackson, 1963). Sendo assim, um elétron que é posto em oscilação por um feixe de raios-X é continuamente acelerado e desacelerado durante seu movimento, reemitindo uma onda eletromagnética. Neste sentido pode-se dizer que o elétron *espalha* raios X, sendo que esta onda espalhada é uma onda irradiada pelo elétron sob ação do feixe incidente.

O espalhamento é convenientemente caracterizado pela razão entre a quantidade de energia emitida pelo sistema espalhador em uma dada direção por unidade de tempo, e a densidade de fluxo de energia da radiação incidente. Seja dI é a energia radiada por um sistema em um ângulo sólido $d\Omega$ por segundo para uma onda incidente com vetor de Poynting $\vec{S} = (c/4\pi)\vec{E} \times \vec{H}$ (sendo $\vec{E}$ e $\vec{H}$ o campo elétrico e o campo magnético característicos da onda incidente e $\vec{S}$ possuindo dimensões de energia/área $\times$ tempo). A seção de choque de espalhamento $d\sigma$ (no ângulo sólido $d\Omega$) é definida como

$$d\sigma = \frac{\overline{dI}}{\overline{S}} \quad (1)$$

onde a barra significa média temporal. Desta formulação vemos que a seção de choque de espalhamento (do inglês *scattering cross section*) possui dimensão de área. Como, em geral, as velocidades de oscilação dos elétrons são muito menores do que a velocidade da luz, sempre teremos valores médios. Se assumimos uma dependência temporal harmônica para as ondas incidentes e espalhadas, as médias são feitas facilmente*. A integral σ de $d\sigma$ sobre todas as direções é denominada seção de choque total de espalhamento.

Vamos assumir o espalhamento produzido por uma carga em repouso na origem do sistema de coordenadas. Supondo que o feixe incidente seja uma onda plana linearmente polarizada, seu campo elétrico pode ser escrito na forma,

$$\vec{E} = \vec{E}_0 \cos(\vec{k}_0 \cdot \vec{r} - \omega t + \alpha) \quad (2)$$

sendo $\vec{E}_0$ a intensidade do campo elétrico, $\vec{k}_0$ a direção de propagação do feixe incidente, ω a frequência de oscilação do campo e α um fator de fase arbitrário. Como assumimos que as velocidades ($\vec{v}$) atingidas pela carga elétrica são pequenas quando comparadas à velocidade da luz, podemos considerar que a força agindo na carga será $e\vec{E}$, ao passo que a força $(e/c)\vec{v} \times \vec{H}$ gerada pelo campo magnético pode ser desprezada. Neste caso podemos também desprezar o efeito do deslocamento da carga durante sua vibração sobre a influência do campo. Se a carga oscila sempre ao redor de um ponto origem, podemos assumir que o campo que age na carga durante todo o tempo é o mesmo campo que age sobre ela na origem, ou seja, $\vec{E} = \vec{E}_0 \cos(\omega t - \alpha)$.

Como a equação do movimento de uma carga é

$$m\ddot{\vec{r}} = e\vec{E} \quad (3)$$

e seu momento de dipolo $\vec{d} = e\vec{r}$, teremos

$$\ddot{\vec{d}} = \frac{e^2}{m} \vec{E} \quad (4)$$

* Se $\vec{U}(\vec{r}, t)$ e $\vec{V}(\vec{r}, t)$ são vetores com dependência temporal harmônica (na forma $e^{i\omega t}$) as médias de operações vetoriais envolvendo estes vetores são facilmente feitas, a saber

$$\left. \begin{aligned} \vec{U}(\vec{r}, t) &= \vec{U}_0(\vec{r}) e^{i\omega t} \\ \vec{V}(\vec{r}, t) &= \vec{V}_0(\vec{r}) e^{i\omega t} \end{aligned} \right\} \Rightarrow \left\{ \begin{aligned} \overline{\vec{U}(\vec{r}, t) \cdot \vec{V}(\vec{r}, t)} &= \frac{\vec{U}_0 \cdot \vec{V}_0^*}{2} \\ \overline{\vec{U}(\vec{r}, t) \times \vec{V}(\vec{r}, t)} &= \frac{\vec{U}_0 \times \vec{V}_0^*}{2} \end{aligned} \right.$$

Para calcularmos a radiação espalhada, usamos a relação para radiação dipolar (que é justificável, pois a velocidade adquirida pela carga sobre a ação da onda incidente é pequena quando comparada à velocidade da luz). Assumimos também que a frequência da onda irradiada pela carga (isto é, espalhada por ela) seja a mesma da onda incidente (espalhamento elástico). Na expressão para radiação dipolar (Landau, 1951)

$$dI = \frac{1}{4\pi c^3} \left(\ddot{\vec{d}} \times \vec{n} \right)^2 d\Omega \quad (5)$$

se substituirmos a equação (4) obtemos,

$$dI = \frac{e^4}{4\pi m^2 c^3} \left(\vec{E} \times \vec{n} \right)^2 d\Omega \quad (6)$$

Por outro lado, o vetor de Poynting para a onda incidente é dado por,

$$S = \frac{c}{4\pi} E^2 \quad (7)$$

Com esta expressão para S obtemos para a seção de choque em um ângulo sólido $d\Omega$,

$$d\sigma = \frac{e^2}{4\pi c^3} \sin^2 \Theta d\Omega \quad (8)$$

onde Θ é o ângulo entre a direção de espalhamento (o vetor $\vec{n}$) e a direção do campo elétrico $\vec{E}$ da onda incidente. Sendo assim vemos que a seção de choque de espalhamento de uma carga livre é independente da frequência. Podemos determinar a seção de choque total σ . Para isso utilizamos coordenadas polares com origem na posição da carga e eixo polar ao longo de $\vec{E}$. Como $d\Omega = \sin \theta d\theta d\varphi$, integramos com respeito a θ de 0 a π , e sobre φ de 0 a 2π para obter

$$\sigma = \frac{8\pi}{3} \left(\frac{e^2}{mc^2} \right) \quad (9)$$

Agora podemos calcular a seção de choque $d\sigma$ para o caso onde a luz incidente é não polarizada. Para isso, devemos calcular a média do vetor $\vec{E}$ no plano perpendicular à direção de propagação da onda incidente (direção do vetor de onda $\vec{k}_0$). Introduzimos o sistema de coordenadas com o eixo z ao longo da direção $\vec{k}_0$ do feixe incidente e x ao longo de $\vec{E}$ (Figura 1).

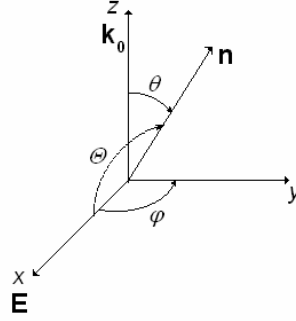


Figura 1 – Sistema de eixos utilizado no desenvolvimento teórico.

Assim o cosseno do ângulo Θ entre as direções de $\vec{n}$ e $\vec{E}$ será $\cos\Theta = \sin\theta\cos\varphi$, onde θ e φ são os ângulos polar e azimutal da direção $\vec{n}$. Fazer a média sobre todas as direções de $\vec{E}$ no plano perpendicular à $\vec{k}_0$ é equivalente a promediar sobre o ângulo azimutal φ . Assim,

$$\begin{aligned}\cos\Theta &= \sin\theta\cos\varphi \Rightarrow \sin^2\Theta = 1 - \sin^2\theta\cos^2\varphi \\ \overline{\sin^2\Theta} &= \frac{1}{2\pi} \int_0^{2\pi} (1 - \sin^2\theta\cos^2\varphi) d\varphi = \frac{1}{2\pi} \left[2\pi - \frac{\sin^2\theta}{2} \int_0^{2\pi} (1 + \cos 2\varphi) d\varphi \right] = 1 - \frac{\sin^2\theta}{2} = \frac{1 + \cos^2\theta}{2} \\ \therefore \overline{\sin^2\Theta} &= \frac{1 + \cos^2\theta}{2}\end{aligned}\quad (10)$$

Substituindo esta relação na equação (8) temos para a seção de choque efetiva de espalhamento de uma onda não-polarizada por uma carga livre,

$$d\sigma = \left(\frac{e^2}{mc^2} \right)^2 \left(\frac{1 + \cos^2\theta}{2} \right) d\Omega \quad (11)$$

sendo θ o ângulo entre as direções das ondas incidente e espalhada[†]. Esta é a fórmula de Thomson para espalhamento de radiação por uma carga livre. O valor $r_0 = e^2 / mc^2 = 2.82 \times 10^{-13} \text{ cm}$ é raio clássico de Thomsom para o elétron. Para ângulos pequenos, que é o que consideramos nesta tese, o fator angular pode ser tomado como sendo 1. Sendo assim a seção de choque de espalhamento de um elétron assume o valor constante $d\sigma / d\Omega \cong r_0^2 = 7.95 \times 10^{-26} \text{ cm}^2$. Já a seção de choque integral (equação 9) para um único elétron assume o valor $\sigma = (8\pi/3)r_0^2 = 6.66 \times 10^{-25} \text{ cm}^2$.

[†] Em cristalografia costuma-se também definir o ângulo entre as direções das ondas incidentes e espalhadas como sendo 2θ .

1.2. Espalhamento por um átomo

Quando um feixe de raios-X encontra um átomo, cada elétron em seu interior espalha parte da radiação coerentemente de acordo com a equação de Thomson. Podemos também esperar que o núcleo contribua em parte com o espalhamento coerente uma vez que possuindo carga, é capaz de oscilar sobre a influência do feixe incidente. No entanto, o núcleo possui uma massa extremamente grande quando comparada à massa do elétron e desta maneira sua oscilação acaba sendo desprezível. Na realidade, a equação de Thomson indica que a intensidade de espalhamento coerente é inversamente proporcional ao quadrado da massa da partícula espalhadora. Sendo assim, o efeito total do espalhamento coerente fornecido por um átomo é devido exclusivamente pelos elétrons em seu interior.

Para o espalhamento na mesma direção do feixe incidente ($\theta = 0$), as ondas espalhadas por todos os elétrons estarão em fase e as amplitudes destas ondas simplesmente se adicionam. Isto não é verdade para as outras direções de espalhamento. O fato de os elétrons em um átomo estarem situados em diferentes pontos no espaço introduz diferenças de fase entre as ondas espalhadas por diferentes elétrons. Neste sentido define-se o fator de espalhamento atômico, $f_{\text{atomo}}(\theta)$ como descrevendo a “eficiência” do espalhamento dado por um átomo em uma dada direção:

$$f_{\text{atomo}}(\theta) = r_0 \frac{A_{\text{atomo}}(\theta)}{A_{\text{eletron}}(\theta)}$$

sendo A_{atomo} e $A_{\text{elétron}}$ a amplitude da onda espalhada pelo átomo e pelo elétron respectivamente em um dado ângulo θ com relação ao feixe incidente. O fator $f_{\text{atomo}}(\theta)$ é comumente dado em unidades do raio do clássico do elétron, ou unidades de comprimento de espalhamento (*scattering length units*). Conforme θ aumenta as ondas espalhadas pelos elétrons individuais ficam mais e mais fora de fase, causando uma diminuição do valor de f_{atomo} . No entanto, para espalhamento de raios-X a baixos ângulos sempre teremos θ pequeno e assim poderemos considerar $f_{\text{atomo}}(\theta)$ como sendo constante. Sendo assim, para estudos em espalhamento de raios-X a baixos ângulos,

$$f_{\text{atomo}} = r_0 Z \quad (12)$$

, sendo Z o número atômico do átomo.

1.3. Espalhamento de ondas planas pela matéria

Podemos agora examinar a interação de ondas planas com a matéria. A onda interage com todos os núcleos e elétrons, os quais se tornam fontes de ondas secundárias. A superposição destas ondas fornece a primeira aproximação da onda espalhada. Também, estas ondas secundárias podem ser re-espalhadas pelos centros espalhadores, e a superposição de todas estas ondas fornece a segunda aproximação e assim sucessivamente para as outras aproximações.

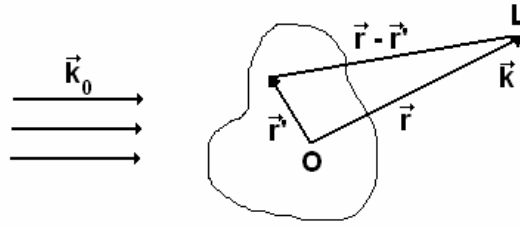


Figura 2 – Espalhamento de primeira ordem de uma onda plana por um potencial limitado no espaço.

Seja $\phi(\vec{r}) = A_0 \exp(i\vec{k}_0 \cdot \vec{r})$ uma onda plana, caracterizada por uma amplitude A_0 e um vetor de onda $\vec{k}_0$, que sofre espalhamento em um centro espalhador descrito por um potencial $V(\vec{r})$. Similarmente ao caso anterior iremos assumir variações lentas em comparação com a velocidade da luz de modo que as funções de onda podem ser consideradas como médias temporais. Vamos assumir um potencial local, isto é, um potencial definido em uma pequena região do espaço. A onda $\psi(\vec{r})$ espalhada pelo campo $V(\vec{r})$ é dada pela solução da equação de Schrödinger

$$[\nabla^2 + k_0^2 + \nu V(\vec{r})]\psi(\vec{r}) = 0 \quad (13)$$

onde $k_0 = |\vec{k}_0|$ representa o número de onda da onda incidente no espaço livre, ∇^2 é o operador laplaciano e ν é um parâmetro especificando intensidade da interação com o campo potencial.

A solução desta equação pode ser dada em séries de potência de ν . O termo livre desta série (ν^0) representa a onda incidente e o termo linear em ν é dado de acordo com a primeira aproximação da onda espalhada. Se nos restringirmos a esta primeira aproximação, a amplitude da onda espalhadora em um ponto de observação L é igual à soma de duas componentes (Sakurai, 1994),

$$\psi^{(0)}(\vec{r}) + \psi^{(1)}(\vec{r}) = A_0 \exp(i\vec{k}_0 \cdot \vec{r}) + \frac{A_0 V}{4\pi} \int \frac{\exp[i\vec{k} \cdot (\vec{r} - \vec{r}')] V(\vec{r}') \exp(i\vec{k}_0 \cdot \vec{r}') d\vec{r}'}{|\vec{r} - \vec{r}'|} \quad (14)$$

O ponto de observação L é usualmente escolhido de maneira que $|\vec{r} - \vec{r}'|$ é muito grande quando comparado às dimensões do centro espalhador. Neste caso temos $\vec{r} \gg \vec{r}'$ e assim $|\vec{r} - \vec{r}'| \cong r$. A onda espalhada assume então a forma assintótica:

$$\psi^{(0)}(\vec{r}) + \psi^{(1)}(\vec{r}) = A_0 \exp(i\vec{k}_0 \cdot \vec{r}) + \frac{A_0 V}{4\pi} \frac{\exp[i\vec{k}_0 \cdot \vec{r}]}{r} \int V(\vec{r}') \exp(i\vec{q} \cdot \vec{r}') d\vec{r}' \quad (15)$$

onde $\vec{q} = \vec{k} - \vec{k}_0$ representa o vetor de espalhamento com $|\vec{q}| = (4\pi \sin(\theta/2))/\lambda$, θ sendo o ângulo de espalhamento, como mostrado na Figura 3 abaixo

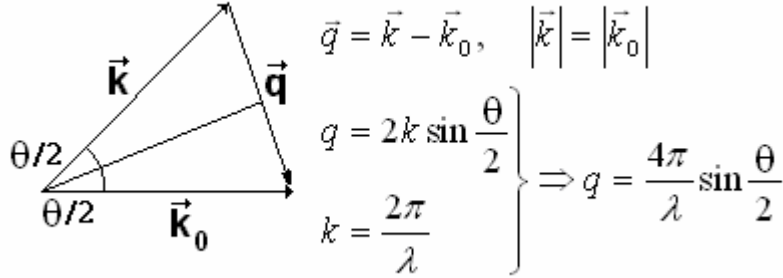


Figura 3 – Definição do vetor de espalhamento q .

No segundo termo da equação (15) podemos definir:

$$f(\vec{q}) = \frac{V}{4\pi} \int V(\vec{r}) \exp(i\vec{q} \cdot \vec{r}) d\vec{r} \quad (16)$$

que descreve a maneira com que a onda é espalhada pelo centro espalhador. Assim $f(\vec{q})$ é a amplitude do espalhamento elástico pelo campo $V(\vec{r})$. Esta expressão é também conhecida como a primeira aproximação de Born ou, em outras palavras, a aproximação por espalhamento único. Assim podemos escrever,

$$\begin{aligned} \psi(\vec{r}) &= A_0 \exp(i\vec{k}_0 \cdot \vec{r}) + A_0 \frac{\exp(i\vec{k}_0 \cdot \vec{r})}{r} f(\vec{q}) \\ \psi(\vec{r}) &= \phi(\vec{r}) + \frac{\phi(\vec{r})}{r} f(\vec{q}) \end{aligned} \quad (17)$$

Vemos claramente por esta expressão que a onda espalhada possui uma componente idêntica a onda incidente somada a uma onda esférica com amplitude $f(\vec{q})$.

A expressão (16) representa uma transformada de Fourier. A solução do problema inverso – cálculo de $V(\vec{r})$ usando a função $f(\vec{q})$ – é dado pela transformada inversa de Fourier

$$V(\vec{r}) = \frac{1}{2\pi^2\nu} \int f(\vec{q}) \exp(-i\vec{q} \cdot \vec{r}) d\vec{q} \quad (18)$$

Assim a aplicação da transformada de Fourier permite-nos calcular a amplitude de espalhamento de um sistema conhecido de centros espalhadores e o campo potencial para uma amplitude de espalhamento conhecida se as condições da aproximação de Born são satisfeitas (Sakurai, 1994).

O primeiro termo do lado direito da equação (17) representa a função de onda inicial deslocada sem mudança na forma, como se não tivesse ocorrido espalhamento, ou seja, como se não tivéssemos o potencial V no sistema. O segundo termo é uma função de onda de espalhamento esférica, indicando que, se colocarmos um detetor fora da direção do feixe incidente, ele detectará uma onda que se expande radialmente, sendo esta uma réplica da função “vista” pelo espalhador na origem, mas reduzida em amplitude pelo fator $1/r$ e modulada pela amplitude $f(\vec{q})$.

A probabilidade de se observar um fóton no detetor em um dado intervalo de tempo t e $t+dt$ é dada por (Merzbacher, 1970),

$$c \frac{|f(\vec{q})|^2}{r^2} |\psi(\vec{r})|^2 r^2 d\Omega dt$$

Desta forma, a probabilidade total de detecção será

$$|f(\vec{q})|^2 d\Omega \int_{-\infty}^{+\infty} |\psi(\vec{r})|^2 c dt \quad (19)$$

onde os limites de integração foram tomados como sendo $-\infty$ e $+\infty$ devido à extensão finita da função de onda. Por outro lado, a probabilidade de que uma partícula incidente passe por uma área localizada perpendicularmente ao feixe incidente imediatamente antes do objeto espalhador é dada por,

$$\int_{-\infty}^{+\infty} |\psi(\vec{r})|^2 c dt \quad (20)$$

Da definição de seção de choque diferencial (semelhante à equação (1)),

$$\frac{d\sigma}{d\Omega} = \frac{\text{número de partículas espalhadas em } d\Omega \text{ por unidade de tempo}}{\text{número de partículas incidentes atravessando uma unidade de área por unidade de tempo}}$$

Obtemos,

$$I(\vec{q}) = \frac{d\sigma(\vec{q})}{d\Omega} = |f(\vec{q})|^2 \quad (21)$$

$$\frac{dW(\vec{q})}{d\Omega} = \frac{A_0^2}{r^2} I(\vec{q}) \quad (22)$$

mostrando então que a seção de choque de espalhamento $d\sigma/d\Omega$ do potencial está diretamente relacionada à amplitude de espalhamento $f(\vec{q})$ do mesmo. Outra maneira de se definir a seção de choque de espalhamento é através da função $I(\vec{q})$, conhecida como a intensidade de espalhamento, que é um termo utilizado em análises de estruturas por raios X, uma vez que “seções de choque” são comumente utilizadas em espalhamento por nêutrons. Finalmente, a função $dW/d\Omega$ representa a intensidade experimental, sendo proporcional ao fluxo de energia espalhada ou o número de partículas espalhadas em uma dada direção. Posteriormente em *Aspectos Experimentais*, esta função será melhor descrita.

1.4. Espalhamento de raios X por partículas

1.4.1. Equação de Debye

Pode-se escrever uma equação explícita para a intensidade espalhada média para uma partícula quando as coordenadas dos átomos são conhecidas. Se existem n átomos com fatores de espalhamento $f_i(\vec{q})$ (que podem ser considerados como constantes na região a baixos ângulos: $f_i(\vec{q}) = f_i(0) = f_i$) e coordenadas $\vec{r}_i$ interiores à partícula teremos que,

$$I(q) = \sum_{i=1}^n \sum_{j=1}^n f_i f_j \frac{\sin(qr_{ij})}{qr_{ij}} \quad (23)$$

onde $r_{ij} = |\vec{r}_i - \vec{r}_j|$ enquanto que a soma é tomada sobre todos os átomos em uma partícula.

Esta equação foi derivada por Debye (Debye, 1915) e possui seu nome. Ela fornece uma expressão exata para o espalhamento a baixos ângulos por uma partícula, mas não é

possível resolver-se o problema inverso devido ao grande número de termos na soma dupla. Veremos posteriormente que esta expressão pode ser utilizada para descrever o espalhamento a baixos ângulos de partículas com coordenadas atômicas conhecidas ou modelos construídos com elementos finitos.

Se em vez de um conjunto de átomos utilizarmos a distribuição de densidades eletrônicas $\rho(\vec{r})$ para descrever a partícula, a equação de Debye toma a forma

$$I(q) = \iint_V \rho(\vec{r}_1) \rho(\vec{r}_2) \frac{\sin(qr_{12})}{qr_{12}} d\vec{r}_1 d\vec{r}_2 \quad (24)$$

1.4.2. Função de Correlação

Como mostramos na equação (21), a intensidade de espalhamento é dada pelo módulo quadrático da amplitude de espalhamento $f(\vec{q})$,

$$I(\vec{q}) = |f(\vec{q})|^2$$

No entanto, a amplitude $f(\vec{q})$ está relacionada ao potencial espalhador $V(\vec{r})$ através de uma transformada de Fourier (equações 16 e 18). Este potencial espalhador pode ser representado pela densidade eletrônica da partícula espalhadora $\rho(\vec{r})$ de modo que podemos escrever,

$$\begin{aligned} I(\vec{q}) &= |f(\vec{q})|^2 = f(\vec{q}) \bullet f(\vec{q})^* \\ &= F(\rho(\vec{r})) \bullet F(\rho(\vec{r}))^* \\ &= F(\rho(\vec{r}) * \rho(-\vec{r})) \\ &= \int_V \exp(i\vec{q} \cdot \vec{r}) d\vec{r} \int_V \rho(\vec{r}') \rho(\vec{r}' + \vec{r}) d\vec{r}' \\ &= \int_V \int_V \rho(\vec{r}') \rho(\vec{r}' + \vec{r}) \exp(i\vec{q} \cdot \vec{r}) d\vec{r} d\vec{r}' \\ \therefore I(\vec{q}) &= \int_V \int_V \rho(\vec{r}') \rho(\vec{r}' + \vec{r}) \exp(i\vec{q} \cdot \vec{r}) d\vec{r} d\vec{r}' \quad (25) \end{aligned}$$

Onde utilizamos propriedades de transformadas de Fourier e o teorema da convolução (Arfken e Weber, 1995; Butkov, 1968; Semenyuk e Svergun, 1987). O produto de convolução,

$$P(\vec{r}) = \int_V \rho(\vec{r}') \rho(\vec{r}' + \vec{r}) d\vec{r}'$$

é conhecido como auto-convolução da partícula, ou função de Patterson em análises estruturais.

O espalhamento a baixos ângulos é obtido por promediar-se $I(\vec{q})$ sobre o ângulo sólido no espaço recíproco. Colocando o espaço real em coordenadas polares (r, θ, φ ; $d\omega = \sin \theta d\theta d\varphi$ é um elemento de ângulo sólido), obtemos

$$\begin{aligned} I(q) &= (4\pi)^{-1} \int_{\Omega} I(\vec{q}) d\Omega \\ &= \frac{1}{4\pi} \int_0^\infty \int_0^{4\pi} r^2 dr d\omega \int_0^{4\pi} d\Omega P(\vec{r}) \exp(i\vec{q} \cdot \vec{r}) \\ &= \int_0^\infty r^2 dr \int_0^{4\pi} P(\vec{r}) d\omega \frac{1}{4\pi} \int_0^{4\pi} \exp(i\vec{q} \cdot \vec{r}) d\Omega \end{aligned} \quad (26)$$

Agora, a média de $\exp(i\vec{q} \cdot \vec{r})$ sobre todas as direções fornece (Guinier e Fournet, 1955)

$$\langle \exp(i\vec{q} \cdot \vec{r}) \rangle_{\Omega} = \frac{\sin(qr)}{qr} \quad (27)$$

E assim, a intensidade é escrita como,

$$I(q) = 4\pi \int_0^\infty \gamma(r) \frac{\sin(qr)}{qr} r^2 dr \quad (28)$$

onde

$$\gamma(r) = \frac{1}{4\pi} \int_0^{4\pi} P(\vec{r}) d\omega = \langle \rho(\vec{r}) * \rho(-\vec{r}) \rangle_{\omega} \quad (29)$$

(a auto-convolução média da distribuição de densidade de espalhamento) é a função de correlação da partícula, primeiro introduzida por Debye e Bueche (Guinier e Fournet, 1955). A transformada inversa da equação (28) pode ser escrita como

$$\gamma(r) = \frac{1}{4\pi^2} \int_0^\infty I(q) \frac{\sin(qr)}{qr} q^2 dq \quad (30)$$

As equações anteriores são de grande importância em análises em espalhamento a baixos ângulos. A intensidade de espalhamento é promediada em todas as orientações da partícula espalhadora e, devido a isso, é isotrópica, possuindo o mesmo perfil em todo o

cone de espalhamento. Para melhor compreensão deste efeito tomemos o espalhamento de uma esfera (Figura 4):

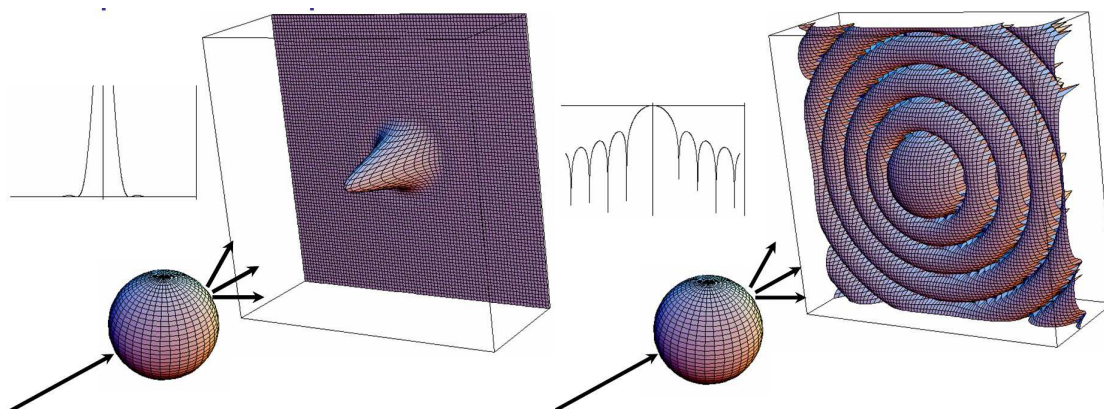


Figura 4 – Perfil bidimensional de espalhamento para uma partícula esférica. Esquerda: Intensidade espalhada em escala linear. Direita: Intensidade espalhada em escala logarítmica. Inserções: Cortes unidimensionais no plano de espalhamento. A simulação foi feita utilizando o programa Mathematica®.

Devido ao fato da esfera ter alta simetria, sua intensidade de espalhamento será naturalmente isotrópica de modo que qualquer corte linear no plano de espalhamento fornece o mesmo perfil de espalhamento. No entanto o mesmo não acontece para partículas não esféricas. Na Figura 5 (à esquerda) mostramos o espectro de espalhamento de uma partícula anisotrópica fixa no espaço. Vemos claramente que cortes no plano de espalhamento em diferentes direções fornecerão perfis da intensidade espalhada distintos. Já na Figura 5 (à direita) temos o espectro de espalhamento para uma média de todas as orientações da partícula. Neste caso o espectro de espalhamento é isotrópico e, qualquer corte linear fornecerá o mesmo perfil de intensidade. Esta propriedade é muito importante, pois permite que em experimentos reais tomemos apenas um corte (utilizando detectores lineares) da intensidade espalhada.

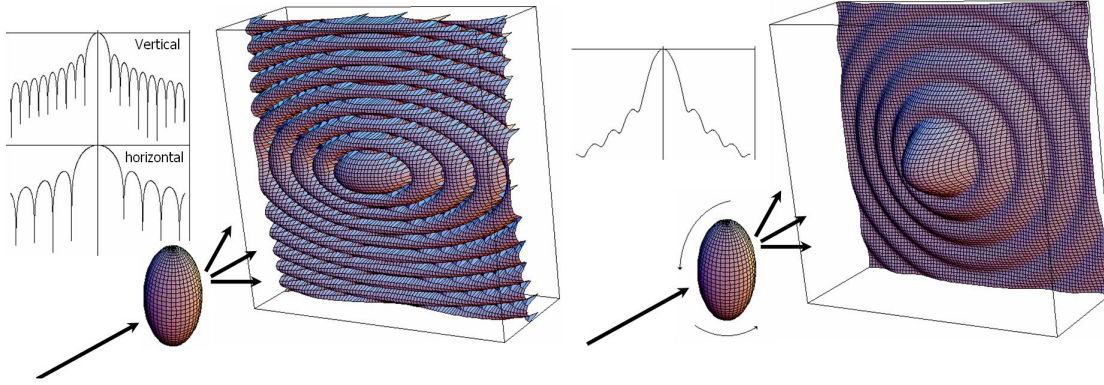


Figura 5 – Perfil bidimensional de espalhamento para uma partícula anisotrópica. Esquerda: Intensidade para a partícula fixa. Direita: Intensidade promediada para todas as orientações da partícula. Inserções: Cortes unidimensionais no plano de espalhamento. A simulação foi feita utilizando o programa Mathematica®.

A função de correlação, que pode ser calculada diretamente da curva de intensidade, permite tirar-se imediatamente uma grande quantidade de conclusões da estrutura da partícula. Em particular é claro da equação (30) que $\gamma(r) \equiv 0$ para $r > D$, onde D é a máxima dimensão da partícula. Assim a equação (28) pode ser reescrita como,

$$I(q) = 4\pi \int_0^D \gamma(r) \frac{\sin(qr)}{qr} r^2 dr \quad (31)$$

A função $\gamma(r)$ bem como a função

$$p(r) = r^2 \gamma(r) \quad (32)$$

(a função de distribuição de pares de distâncias da partícula) são muito utilizadas em análises de espalhamento a baixos ângulos. Estas funções representam a geometria da partícula, expressando numericamente o conjunto de distâncias unindo elementos de volume no interior da partícula. Particularmente no caso de partículas uniformes, nas quais as densidades de espalhamento podem ser expressas na forma,

$$\rho(\vec{r}) = \begin{cases} \rho & , \text{ se } \vec{r} \text{ está dentro da partícula} \\ 0 & , \text{ se } \vec{r} \text{ está fora da partícula} \end{cases} \quad (33)$$

as funções $\gamma(r)$ e $p(r)$ permitem tirar-se conclusões sobre a geometria e forma das partículas. Isso será mostrado posteriormente em *Sistemas Monodispersos*.

1.4.3. Espalhamento e Absorção de Raios X

Quando radiação eletromagnética interage com a matéria diversos processos podem ocorrer dependendo da energia da onda incidente. Em geral pode-se ter quatro processos: (1) Espalhamento elástico ou coerente (Rayleigh); (2) Espalhamento Compton; (3) Efeito Fotoelétrico e (4) Produção de Pares. Para a faixa de raios X, 7 - 12keV, os efeitos importantes são o efeito fotoelétrico e espalhamento elástico. O espalhamento Compton é importante para energia próximas de 1MeV e para energias maiores temos o domínio da produção de pares (Williams, 1991).

Quando um feixe de raios X incide em qualquer forma de matéria, ele é parcialmente transmitido e parcialmente absorvido. Experimentos mostram que a fração diminuída na intensidade I de um feixe de raios X quando este atravessa qualquer substância homogênea é proporcional à distância atravessada x . Na forma diferencial,

$$-\frac{dI}{I} = \mu \cdot dx \quad (34)$$

onde a constante de proporcionalidade μ é chamada *coeficiente de absorção linear* e depende da substância considerada, sua densidade e comprimento de onda dos raios X. Integrando-se a equação anterior temos

$$I = I_0 e^{-\mu \cdot x} \quad (35)$$

O coeficiente de absorção linear μ é proporcional à densidade d , o que significa que a quantidade μ/d é uma constante do material e independente de seu estado físico (Cullity, 1978). Esta última quantidade, chamada *coeficiente de absorção de massa* é usualmente tabelada e a equação (35) pode ser reescrita em uma forma mais usual,

$$I = I_0 e^{-(\mu/d) \cdot d \cdot x} \quad (36)$$

Para um dado elemento o coeficiente de absorção aumenta, em geral, com o comprimento de onda, mas este aumento não é contínuo. Existem *descontinuidades de absorção* para certos comprimentos de onda denominados, em ordem crescente, λ_K , λ_L , λ_M , ..., que correspondem a energias conhecidas como *bordas de absorção*. Nestas bordas o mecanismo de absorção muda e a nova camada eletrônica K, L, M... é ionizada por radiação de comprimento de onda menores do que o correspondente comprimento de onda

crítico. Ao mesmo tempo, são emitidos raios X das linhas K, L, M... do átomo excitado, sendo essa emissão conhecida como fluorescência (Guinier, 1963).

Como mencionado anteriormente, a matéria absorve raios X por dois processos distintos, por espalhamento e pelo efeito fotoelétrico, sendo que estes dois processos juntos correspondem à absorção total medida pela quantidade μ/d . Em geral o fenômeno de espalhamento, exceto para elementos muito leves, é responsável apenas por uma pequena fração da absorção total. Por este motivo muitas vezes denomina-se a absorção por efeito fotoelétrico como *absorção real* do material.

Assim, se os efeitos de absorção real e espalhada são independentes, a absorção total de um material pode ser escrita como a soma:

$$\mu = \mu_l = \sigma_l + \tau_l \quad (37)$$

onde σ_l e τ_l são as seções de choque de espalhamento e da absorção real, respectivamente.

Apresentamos apenas alguns resultados referentes a processos de absorção de raios X pela matéria. Aspectos mais detalhados podem ser obtidos em diversos livros texto (Cullity, 1978; Feigin e Svergun, 1987 e referências nestes textos).

1.5. **Espalhamento por um conjunto de partículas**

A amplitude do espalhamento elástico por um grupo de N átomos pode ser escrito como (Feigin e Svergun, 1987)

$$A(\vec{q}) = \sum_{i=1}^N f_i(\vec{q}) \exp(2\pi i \vec{q} \cdot \vec{r}_i) \quad (38)$$

onde $f_i(\vec{q})$ e $\vec{r}_i$ são os fatores de espalhamento e as coordenadas dos átomos (com respeito a uma dada origem) respectivamente. A maior dificuldade em estudos estruturais corresponde em recuperar-se o arranjo espacial dos átomos a partir dos padrões de espalhamento, ou seja, a função de intensidade espalhada $I(\vec{q}) = A(\vec{q}) \cdot A^*(\vec{q})$. Uma dificuldade que naturalmente aparece neste tipo de análise é o bem conhecido *problema da fase*, que decorre do fato de na medida da intensidade espalhada, perder-se informações sobre as fases $\exp(2\pi i \vec{q} \cdot \vec{r}_i)$ existentes na soma (38). Apesar de em cristalografia a determinação correta destas fases ser crucial para a determinação da estrutura, em análises

de espalhamento a baixos ângulos, em geral, não é preciso determinar-se estes termos, podendo-se utilizar apenas a intensidade $I(q)$.

Consideremos o sistema como sendo composto de N átomos idênticos (ou de modo mais geral, N partículas idênticas: moléculas, aglomerados, etc.) com fatores de espalhamento $f(\vec{q})$ distribuídos em um volume V_0 (volume iluminado pelo feixe incidente). Da equação (38) a intensidade de espalhamento do grupo completo é dado por

$$I(\vec{q}) = \sum_{i=1}^N \sum_{j=1}^N f_i(\vec{q}) f_j(\vec{q}) \exp[i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)] \quad (39)$$

onde $\vec{r}_i$ e $\vec{r}_j$ são as coordenadas dos átomos ou vetores descrevendo as posições das partículas. Separando os termos com $i = j$ e promediando no grupo de partículas, temos:

$$\begin{aligned} \langle I(\vec{q}) \rangle &= \left\langle \sum_{i=1}^N f_i^2(\vec{q}) \right\rangle + \left\langle \sum_{i \neq j} \sum f_i(\vec{q}) f_j(\vec{q}) \exp[i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)] \right\rangle \\ \langle I(\vec{q}) \rangle &= \langle N \rangle \langle f^2(\vec{q}) \rangle + \langle f(\vec{q}) \rangle^2 \left\langle \sum_{i \neq j} \sum \exp(i\vec{q} \cdot \vec{r}_{ij}) \right\rangle \end{aligned} \quad (40)$$

onde $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$. O valor médio $\langle N \rangle$ é escrito ao invés de N porque o número de átomos (partículas) pode flutuar no caso geral. Vamos supor que o sistema é isotrópico, isto é, que o arranjo atômico não depende da direção. Para se calcular o valor médio da soma dupla na equação (40) pode-se utilizar a função de probabilidade p_{ij} . Esta função fornece a probabilidade de se encontrar uma partícula no elemento de volume dv_i descrita pelo vetor $\vec{r}_i$ e outra simultaneamente no elemento de volume dv_j posicionada pelo vetor $\vec{r}_j$. Utilizando a média de $\exp(i\vec{q} \cdot \vec{r})$ ($\langle \exp(i\vec{q} \cdot \vec{r}) \rangle_\Omega = \sin(qr)/qr$) temos (Guinier e Fournet, 1955)

$$\left\langle \sum_{i \neq j} \sum \exp(i\vec{q} \cdot \vec{r}_{ij}) \right\rangle = \iint_{V_0} \frac{\sin(qr_{ij})}{qr_{ij}} p_{ij} dv_i dv_j \quad (41)$$

sendo V_0 o volume irradiado da amostra.

Vamos calcular a probabilidade $p_{ij} dv_i dv_j$. A probabilidade de ocorrerem dois eventos em um sistema é igual a probabilidade do primeiro evento multiplicada pela probabilidade do segundo, modificada pela existência do primeiro evento. A probabilidade

de ocorrência do primeiro evento, obter uma partícula no elemento de volume dv_i é $(\langle N \rangle / V_0) dv_i$. Quando isso ocorre:

- As restantes $\langle N \rangle - 1$ partículas se distribuem no volume $V_0 - dv_i$.
- Se os centros dos dois volumes dv_i e dv_j , cada um contendo uma partícula, estão separados pela distância r , os centros das partículas estão separados por uma distância aproximadamente igual à r . É obvio que as distâncias entre as partículas não são todas igualmente prováveis: se as partículas fossem esferas de raio R , a distância entre elas não poderia ser menor do que $2R$. Esta restrição é incorporada ao cálculo através da função $g(r)$ tal que, se $\vec{R}_i$ e $\vec{R}_j$ são os vetores posição dos centros das partículas e $r_{ij} = |\vec{R}_i - \vec{R}_j|$,

$$p_{ij} dv_i dv_j = \frac{\langle N \rangle}{V_0} dv_i \frac{\langle N \rangle - 1}{V_0 - dv_i} g(r_{ij}) dv_j \quad (42)$$

- A maneira como a função $g(r)$ foi introduzida indica que esta função tende à unidade conforme r aumenta, uma vez que não mais ocorrerão superposições entre as partículas e todas as posições passam a ser igualmente prováveis. Podemos desprezar dv_j com respeito à V_0 e usualmente “1” é desprezível com respeito à $\langle N \rangle$. Se utilizamos o volume médio oferecido para cada partícula, $v_1 = (V_0 / \langle N \rangle)$, uma quantidade característica do grupo de partículas, a equação anterior se reduz a,

$$p_{ij} dv_i dv_j = \frac{dv_i}{v_1} \frac{dv_j}{v_1} g(r_{ij}) dv_j \quad (43)$$

- Com isso a dupla soma na equação (41) se reduz a,

$$\left\langle \sum_{i \neq j} \exp(\vec{q} \cdot \vec{r}_{ij}) \right\rangle = \iint_{V_0} \frac{\sin(qr_{ij})}{qr_{ij}} g(r_{ij}) \frac{dv_i}{v_1} \frac{dv_j}{v_1} \quad (44)$$

Como a função $g(r)$ tende a unidade com o aumento de r , a estrutura do grupo de partículas é melhor caracterizada pela função $(1 - g(r))$, a qual é diferente de zero somente para pequenos valores de r . Sendo assim, fazendo-se a substituição $g(r) = 1 - (1 - g(r))$ na equação (44) e usando este resultado na equação (40) temos,

$$\begin{aligned}
\langle I(\vec{q}) \rangle = \langle N \rangle \langle f^2(\vec{q}) \rangle + \langle f(\vec{q}) \rangle^2 \iint_{V_0} \frac{\sin(qr_{ij})}{qr_{ij}} \frac{dv_i}{v_1} \frac{dv_j}{v_1} \\
+ \langle f(\vec{q}) \rangle^2 \iint_{V_0} [1 - g(r_{ij})] \frac{\sin(qr_{ij})}{qr_{ij}} \frac{dv_i}{v_1} \frac{dv_j}{v_1}
\end{aligned} \quad (45)$$

Vamos examinar separadamente cada um dos termos integrais:

- *Primeiro termo:* A contribuição deste termo para a intensidade total,

$$\langle I(\vec{q}) \rangle_A = \langle f(\vec{q}) \rangle^2 \iint_{V_0} \frac{\sin(qr_{ij})}{qr_{ij}} \frac{dv_i}{v_1} \frac{dv_j}{v_1}$$

pode (de acordo com a equação (24)) ser interpretado como representando a intensidade espalhada por uma “partícula” de volume V_0 com densidade de espalhamento constante uniforme $\rho = \langle f(\vec{q}) \rangle / v_1$. Dado o tamanho desta (pseudo) partícula, $\langle I(\vec{q}) \rangle_A$ é efetivamente zero para todos os ângulos observáveis. É importante notar a simples e geral interpretação geométrica do termo $\langle I(\vec{q}) \rangle_A$. Para obter o limite após o qual $\langle I(\vec{q}) \rangle_A$ é desprezível, deve-se considerar que para um dado $q=q_0$, a máxima diferença de fase entre dois raios espalhados por dois pontos na partícula deve ser da ordem de 2π radianos (Guinier e Fournet, 1955). Desta forma, se a dimensão média do volume irradiado V_0 pode ser descrita por $D_0 = V_0^{1/3}$, q_0 é determinado por $q_0 = 2\pi / D_0$.

- *Segundo termo:* Vamos considerar primeiro a integração com respeito à dv_j

$$\int_{V_0} \frac{\sin qr_{ij}}{qr_{ij}} (1 - g(r_{ij})) \frac{dv_j}{v_1}$$

a função $(1 - g(r_{ij}))$ tende a zero rapidamente com o aumento de r e, se desprezarmos efeitos de borda, este termo pode ser reescrito como,

$$\int_0^\infty \frac{\sin qr}{qr} (1 - g(r)) \frac{4\pi r^2}{v_1} dr$$

independente do índice j . A próxima integração com respeito a “ i ” se reduz a multiplicar o termo anterior pelo fator $\int_{V_0} dv_j / v_1$ o qual é igual à $\langle N \rangle$, o número médio de partículas no volume V_0 .

A relação final para a intensidade espalhada assume então a forma:

$$\langle I(\vec{q}) \rangle = \langle N \rangle \left\{ \langle f^2(\vec{q}) \rangle - \frac{\langle f(\vec{q}) \rangle^2}{v_1} \int_0^\infty [1 - g(r)] \frac{\sin(qr)}{qr} 4\pi r^2 dr \right\} \quad (46)$$

Se assumimos que a amplitude de espalhamento do átomo (ou partícula) é esféricamente simétrica (ou próxima disso), temos $\langle f^2(\vec{q}) \rangle = \langle f(\vec{q}) \rangle^2$ e assim podemos escrever,

$$I(q) = \langle N \rangle F^2(q) \left\{ 1 - \frac{1}{v_1} \int_0^\infty [1 - g(r)] \frac{\sin(qr)}{qr} 4\pi r^2 dr \right\} \quad (47)$$

onde $I(q)$ e $F^2(q)$ representam a intensidade de espalhamento média do sistema e do átomo (partícula) respectivamente. Esta equação é muito importante e expressa a intensidade de espalhamento como uma função do fator de forma de um átomo (partícula) bem como da função estatística que governa seu arranjo no sistema. Assim, por exemplo, podemos procurar os fatores $g(r)$ e $F^2(q)$ ao invés de tentar-se determinar as coordenadas atômicas do sistema.

1.5.1. Monodispersidade e polidispersidade

Na derivação das equações anteriores assumimos que todas as partículas eram idênticas (sistema monodisperso), para o qual pode-se obter uma conexão explícita entre o fator de forma da partícula $F^2(q)$ e a intensidade de espalhamento $I(q)$. Para sistemas polidispersos, onde as partículas podem possuir diferentes formas, tamanhos e densidades as expressões anteriormente obtidas não são válidas. Isso porque no processo de cálculo de médias no sistema devemos levar em conta os diferentes fatores de forma das partículas.

Para considerar-se sistemas polidispersos é conveniente descrever as partículas como possuindo um certo tamanho característico R . Uma função $D_N(R)$ pode ser

introduzida de modo que $D_N(R)dR$ corresponda ao número de partículas existentes no intervalo $(R, R+dR)$. A contribuição destas partículas a um fator de forma médio é então dada pelo produto $\langle F_0^2(q, R) \rangle m^2(R) D_N(R) dR$, onde $\langle F_0^2(q, R) \rangle$ é o fator de forma normalizado das partículas de tamanho R ($F_0(0, R)=1$); ao passo que $m(R)$ correlaciona cada tamanho R ao comprimento de espalhamento (*full scattering length*) da partícula correspondente. Em termos gerais $m(R)$ é proporcional ao volume da partícula espalhadora.

O fator de forma médio total será então,

$$\langle F^2(q) \rangle = \int_0^\infty \langle F_0^2(q, R) \rangle m^2(R) D_N(R) dR \quad (48)$$

Em um caso particular de um sistema polidisperso isotrópico, quando todas as partículas possuem tamanhos diferentes, mas forma idêntica, obtemos

$$F^2(q) = \int_0^\infty i_0(q, R) m^2(R) D_N(R) dR \quad (49)$$

onde $i_0(q, R)$ é a média sobre todas as orientações do fator de forma normalizado.

Vale a pena ressaltar que enquanto que para sistemas monodispersos é possível procurar-se a estrutura da partícula, para sistemas polidispersos isto somente pode ser feito quando se conhece a função $D_N(R)$. Existe outro enfoque que acaba sendo mais importante neste tipo de análise: obter-se $D_N(R)$ assumindo $i_0(q, R)$ conhecido. É evidente que sistemas monodispersos possibilitam uma análise mais detalhada da estrutura do objeto do que os sistemas polidispersos uma vez que não se faz suposições quanto a forma das partículas. No entanto, a polidispersidade é freqüentemente encontrada em sistemas reais que se apresentam para análise por espalhamento a baixos ângulos.

1.5.2. O conceito de contraste

Até agora o espalhamento do solvente foi desconsiderado (as partículas foram assumidas como estando no vácuo). Porém em casos reais as partículas estão em um solvente (ou matriz) com uma dada densidade eletrônica ρ_s . Se a densidade eletrônica de uma partícula puder ser escrita como uma função $\rho(\vec{r})$ (por simplicidade assumimos o caso

monodisperso), o espalhamento da radiação não será dado pelo valor particular de cada uma das densidades mas sim ao contraste entre as densidades eletrônicas da partícula e solvente dado por,

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \rho_s \quad (50)$$

também chamado de contraste de espalhamento. Para poder-se separar o espalhamento proveniente das partículas imersas no solvente, medidas experimentais do espalhamento da solução de partículas e do solvente puro devem ser feitas, seguidos da subtração do primeiro pelo segundo. Neste contexto, as expressões apresentadas anteriormente mantêm-se inalteradas, bastando substituir $\rho(\vec{r})$ por $\Delta\rho(\vec{r})$.

Para partículas aproximadamente homogêneas ($\rho(\vec{r})$ não varia muito ao longo da partícula) pode-se utilizar um *contraste eletrônico médio* $\Delta\rho$ definido por,

$$\Delta\rho = \langle \Delta\rho(\vec{r}) \rangle = V^{-1} \int_V [\rho(\vec{r}) - \rho_s] d\vec{r} = \bar{\rho} - \rho_s \quad (51)$$

sendo V o volume da partícula espalhadora. Desta expressão vemos que para altos valores de $\Delta\rho$ ($\bar{\rho} \gg \rho_s$ ou $\bar{\rho} \ll \rho_s$) as partículas (suas formas) são altamente contrastadas do solvente. Por outro lado, em contrastes baixos ($\bar{\rho} \approx \rho_s$, $\Delta\rho \approx 0$) a intensidade de espalhamento irá corresponder mais à flutuações da densidade eletrônica do meio do que a forma particular das partículas.

1.5.3. Sistemas concentrados e diluídos

As expressões gerais apresentadas (equação (46)) mostraram que a intensidade de espalhamento consiste de dois termos: o primeiro dependendo somente da estrutura da partícula, enquanto que o outro é função de sua distribuição espacial $g(r)$ e do volume disponível por partícula v_l por partícula. No entanto, a função $g(r)$ também depende de v_l : $g(r) = g(r, v_l)$. Se a equação (45) é expressa na forma

$$I(q) = \langle N \rangle \langle f^2(\vec{q}) \rangle - \langle f(\vec{q}) \rangle^2 \frac{\langle N \rangle^2}{V} \int_0^\infty [1 - g(r, v_l)] \frac{\sin(qr)}{qr} 4\pi r^2 dr \quad (52)$$

vemos que o primeiro termo corresponde ao espalhamento independente de $\langle N \rangle$ partículas, ao passo que o segundo termo, proporcional a $\langle N \rangle^2$, provém da interferência interpartículas.

Os efeitos de interferência são muito difíceis de serem tratados rigorosamente e a maior dificuldade vem justamente da dependência entre g e v_l . Uma maneira elegante de se reescrever a equação anterior seria,

$$I(q) = \langle N \rangle \langle f^2(\vec{q}) \rangle \langle S(\vec{q}) \rangle \quad (53)$$

onde

$$\langle S(\vec{q}) \rangle = 1 + \frac{\langle f(\vec{q}) \rangle^2}{\langle f^2(\vec{q}) \rangle} \frac{\langle N \rangle^2}{V} \int_0^\infty [g(r, v_l) - 1] \frac{\sin(qr)}{qr} 4\pi r^2 dr$$

sendo $S(\vec{q})$ denominado fator de estrutura do sistema. Cálculos analíticos para esta função são possíveis somente para partículas com simetrias simples (esferas, cilindros), não possuindo solução geral para qualquer outra forma. Posteriormente apresentaremos uma maneira de se calcular fatores de estrutura utilizando simulações de Monte Carlo.

O caso onde o volume médio disponível por partícula (v_l) é muito maior do que seu próprio volume (V) é de grande importância sendo muito favorável para o estudo de estruturas de partículas em solução uma vez que $g(r, v_l) \rightarrow 1$ para $v_l \gg V$. Nestes casos a interação interpartículas pode ser desprezado fornecendo,

$$I(q) = \langle N \rangle \langle f^2(\vec{q}) \rangle \quad (54)$$

Este resultado clássico informa que para sistemas diluídos de partículas idênticas a intensidade a baixos ângulos é proporcional ao espalhamento de uma única partícula promediado em todas as orientações. Sendo assim a forma bem como a estrutura interna da partícula podem ser investigados neste caso. Para partículas com mesma forma, porém tamanhos diferentes teremos da equação (49)

$$I(q) = \int_0^\infty i_0(q, R) D_N(R) m^2(R) dR \quad (55)$$

Com o aumento da concentração de partículas no meio, a curva de espalhamento sofre alteração em decorrência da interação interpartículas. Os efeitos de interferência (que serão relevantes posteriormente na análise de proteínas em solução) podem ser descritos qualitativamente da seguinte maneira:

- Eles aumentam com a concentração

- São significantes somente na parte inicial da curva uma vez que a função $\sin(qr)$ para grandes valores de q oscila rapidamente quando comparada com $g(r, v_1)$ e a integral na equação (52) tende a zero rapidamente
- Em geral diminuem a intensidade espalhada próxima a origem observada em experimentos.

1.6. Sistemas Monodispersos

Na descrição de sistemas de partículas foi mostrado que, para sistemas monodispersos altamente diluídos (partículas idênticas em baixa concentração, $v_1 \gg V$), as interações interpartículas podem ser desprezadas e a intensidade total correspondente ao sistema de N partículas é dado por

$$I(q) = \langle N \rangle \langle F^2(q) \rangle = \langle N \rangle I_1(q) \quad (56)$$

sendo $I_1(q)$ a intensidade de espalhamento, promediada em todas as orientações da partícula espalhadora. Desta forma, dados experimentais podem ser interpretados através do cálculo da função $I_1(q)$. Esta função pode ser calculada analiticamente para corpos geométricos simples ou simulada através da técnica de elementos finitos (Oliveira, 2001). A seguir mostraremos propriedades gerais desta função.

1.6.1. Parâmetros Gerais

Como mostramos anteriormente, se assumimos o solvente como uma matriz uniforme com densidade eletrônica constante, a intensidade espalhada por uma partícula randomicamente orientada presente nesta matriz pode ser escrita como,

$$I_1(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin qr}{qr} dr \quad (57)$$

e a transformada inversa

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I_1(q) \frac{\sin qr}{qr} dq \quad (58)$$

a distribuição de pares de distâncias no interior da partícula.

O comportamento da intensidade espalhada à valores muito pequenos ($q \rightarrow 0$) e valores muito grandes ($q \rightarrow \infty$) do vetor transferência de momento q está diretamente relacionado com os parâmetros dimensionais e estruturais da partícula espalhadora. Na realidade, próximo de $q = 0$ pode-se utilizar a expansão de McLaurin $\sin(qr)/qr \approx 1 - (qr)^2/3! + \dots$ na equação (57) obtendo-se

$$I(q) = 4\pi \left[\int_0^{D_{\max}} p(r) dr - \frac{1}{6} q^2 \int_0^{D_{\max}} p(r) r^2 dr + \dots \right]$$

$$I(q) = 4\pi \int_0^{D_{\max}} p(r) dr \left[1 - \frac{1}{3} \frac{\int_0^{D_{\max}} p(r) r^2 dr}{2 \int_0^{D_{\max}} p(r) dr} q^2 + \dots \right]$$

e, se definimos um raio médio quadrático R_g^2 ,

$$R_g^2 = \frac{\int_0^{D_{\max}} p(r) r^2 dr}{2 \int_0^{D_{\max}} p(r) dr} \quad (59)$$

e levando-se em conta que a intensidade transmitida $I_1(0)$ é proporcional ao quadrado da contraste de espalhamento da partícula

$$I_1(0) = \int_V \int_V \Delta\rho(\vec{r}) \Delta\rho(\vec{r}') d\vec{r} d\vec{r}' = 4\pi \int_0^{D_{\max}} p(r) dr = (\Delta\rho)^2 V^2 \quad (60)$$

obtemos finalmente,

$$I_1(q) = I_1(0) \left[1 - \frac{1}{3} R_g^2 q^2 + O(q^4) \right] \cong I_1(0) \exp\left(-\frac{1}{3} R_g^2 q^2\right) \quad (61)$$

O raio definido na equação (59), conhecido como raio de giro da partícula espalhadora, é o segundo momento normalizado da distribuição de distâncias na partícula ao redor do centro de sua densidade eletrônica. O conceito de raio de giro é similar ao raio de inércia em mecânica clássica, utilizando o contraste de densidade eletrônica como função peso ao invés da densidade de massa.

A equação (61), derivada por Guinier (Guinier e Fournet, 1955), durante muito tempo foi a ferramenta mais importante em análises de espalhamento por sistemas

isotrópicos e continua sendo muito útil nos primeiros estágios da análise de dados. Para sistemas monodispersos o Gráfico de Guinier ($\ln I_1(q)$ versus q^2) deve ter uma fração linear que intercepta $I_1(0)$ e cuja inclinação fornece o raio de giro R_g . A linearidade em um gráfico de Guinier pode ser considerada como um teste para a monodispersidade da amostra e desvios da linearidade podem indicar interações repulsivas ou atrativas entre as partículas decorrentes de efeitos de fator de estrutura ou agregação. No entanto, deve-se sempre ter em mente que a aproximação de Guinier é válida somente para ângulos muito pequenos. Para partículas esféricas o intervalo de validade deve ser $q < 1.3/R_g$, sendo ainda menor ($q < 0.7/R_g$) para partículas anisotrópicas. Um ajuste linear fora deste intervalo não tem sentido físico.

Em muitos casos pode-se propor modelos geométricos simples para as partículas espalhadoras e expressões analíticas para valores de raios de giro em função de tamanhos característicos destes corpos geométricos são muito úteis. Na Tabela 1 mostramos expressões analíticas de raios de giro para corpos geométricos simples (Feigin e Svergun, 1987).

Tabela 1 - Raio de Giro para variados tipos de corpos homogêneos.

Esfera de raio R	$R_g^2 = \frac{3}{5} R^2$
Casca esférica com raio $R_1 > R_2$	$R_g^2 = \frac{3}{5} \frac{R_1^5 - R_2^5}{R_1^3 - R_2^3}$
Elipse com semi-eixos a e b	$R_g^2 = \frac{a^2 + b^2}{4}$
Elipsóide com semi-eixos a , b e c	$R_g^2 = \frac{a^2 + b^2 + c^2}{5}$
Prisma com lados A , B e C	$R_g^2 = \frac{A^2 + B^2 + C^2}{12}$
Cilindro elíptico com semi-eixos a , b , e altura h	$R_g^2 = \frac{a^2 + b^2}{4} + \frac{h^2}{12}$
Cilindro circular oco com raio $R_1 > R_2$, e altura h	$R_g^2 = \frac{R_1^2 + R_2^2}{2} + \frac{h^2}{12}$

Enquanto que o raio de giro R_g caracteriza o tamanho da partícula, a intensidade transmitida $I_1(0)$ está relacionada com o peso molecular da partícula (M_w). Da equação

(60) vemos que o valor de $I_1(0)$ é proporcional ao quadrado do contraste de espalhamento da partícula multiplicado pelo volume ao quadrado da partícula espalhadora. Uma vez que o contraste pode ser calculado a partir da composição química e do volume parcial específico $\bar{v}$ da partícula (sendo M_1 a massa molecular da partícula espalhadora e V seu volume, $\bar{v} = V/M_1$), pode-se, em princípio obter o volume V da partícula espalhadora e conseqüentemente seu peso molecular. No entanto, para poder-se determinar este parâmetro deve-se ter a intensidade na chamada *escala absoluta*. Várias equações para determinação de peso molecular e renormalização à intensidade absoluta são apresentadas em diversos trabalhos (Glatter e Krakty, 1982) e serão mostradas posteriormente na parte *Experimental* desta tese. Tendo-se os dados em escala absoluta pode-se obter o peso molecular da partícula espalhadora adotando o desenvolvimento abaixo:

$$\begin{aligned} I_1(0) &= (\Delta\rho)^2 V^2 \\ I_1(0) &= (\Delta\rho)^2 M_1 \bar{v} \\ M_1 &= \frac{I_1(0)}{(\Delta\rho)^2 \bar{v}} \end{aligned} \quad (62)$$

sendo $\bar{v}$ o volume parcial específico da partícula espalhadora e M_1 a massa molecular (em gramas) da partícula espalhadora. Sabendo que o peso molecular M_w é definido como $M_w = M_1 \cdot N_A$, sendo N_A o número de Avogrado, teremos para o peso molecular da partícula espalhadora,

$$M_w = \frac{N_A \cdot I_1(0)}{(\Delta\rho)^2 \bar{v}} \quad (63)$$

Posteriormente apresentaremos dois métodos utilizados em nossos experimentos para a estimativa do peso molecular.

Para se analisar o comportamento de $I_1(q)$ a altos ângulos pode-se integrar a equação (28) duas vezes por partes e, tendo-se em conta que $\gamma(D_{\max})=0$, obtém-se (Ciccariello, 1999)

$$I_1(q) \cong 8\pi q^{-4} \gamma'(0) + O_1 q^{-3} + O_2 q^{-4} + o(q^{-5}) \quad (64)$$

onde O_1, O_2 são termos trigonométricos oscilantes na forma $\sin(qD_{\max})$. O termo principal responsável pela queda da intensidade a altos ângulos é então proporcional a q^{-4} , sendo conhecido como a lei de Porod (Glatter e Krakty, 1982). Mostra-se que para partículas

homogêneas, $\gamma'(0) = -(\Delta\rho)^2 S/4$, onde S é a área superficial da partícula espalhadora. Assim temos no limite de $q \rightarrow \infty$,

$$\lim_{q \rightarrow \infty} [q^4 \cdot I_1(q)] = -2\pi(\Delta\rho)^2 S \quad (65)$$

Para eliminar o termo referente ao contraste das partículas podemos utilizar o invariante de Porod (Guinier e Fournet, 1955)

$$Q = \int_0^\infty q^2 I_1(q) dq = 2\pi^2 \int_V (\Delta\rho(\vec{r}))^2 d\vec{r} \quad (66)$$

Para partículas homogêneas, $Q = 2\pi^2(\Delta\rho)^2 V$ e, tendo-se em conta que $I_1(0) = (\Delta\rho)^2 V^2$, o volume excluído por partícula (Porod) é $V = 2\pi^2 I_1(0) Q^{-1}$. Também o valor assintótico permite a determinação da superfície específica da partícula como $S/V = (\pi/Q) \lim_{q \rightarrow \infty} [q^4 \cdot I_1(q)]$. Vale ressaltar que, graças ao invariante de Porod, os dois parâmetros citados anteriormente podem ser obtidos de dados experimentais em escala relativa, sem a necessidade de renormalização.

Outro ponto importante deve ser mencionado. Na prática, inhomogeneidades internas causam variações na assíntota de Porod, as quais, para um sistema composto de macromoléculas, por exemplo, com contrastes suficientes, podem ser satisfatoriamente corrigidos por uma subtração de um valor constante da curva experimental (Wu and Schmidt, 1974). Pode-se demonstrar que o espalhamento a altos ângulos segue um comportamento linear em um gráfico $q^4 I_1(q)$ vs. q^4 : $q^4 I_1(q) \approx Bq^4 + A$, e a subtração da constante B de $I_1(q)$ fornece uma aproximação para o espalhamento proveniente de um corpo homogêneo.

1.6.2. Grau de Compacidade

Outra característica que pode ser obtida diretamente dos dados de espalhamento é o grau de compacidade da estrutura da partícula espalhadora. Para isso utilizamos os chamados *Gráficos de Kratky* (*Kratky Plots* - $I \cdot q^2$ vs. q).

Um modelo básico muito importante para diversos tipos de macromoléculas em solução é o chamado *wormlike chain model*. Em resumo, é um modelo no qual temos uma cadeia cujo eixo está sujeito a mudanças de direção, contínuas e completamente

randômicas, em qualquer ponto ao longo deste eixo. O grau de enrolamento da cadeia (*degree of coiling*) é caracterizado pelo chamado comprimento de persistência (*persistent length*). Se começamos em uma origem arbitrário nos movemos ao longo da cadeia e em um dado ponto calcularmos o cosseno entre a direção atual e a direção no ponto origem, esta quantidade decrescerá de um valor inicial de 1 e irá, em média, exponencialmente à zero. A distância medida ao longo da cadeia partindo da origem para o ponto com o valor médio $1/e$ é chamado comprimento persistente a (Glatter e Kratky, 1982). A curva de espalhamento de uma molécula com cadeias deste tipo, chamadas freqüentemente de cadeias gaussianas, consiste de 3 regiões.

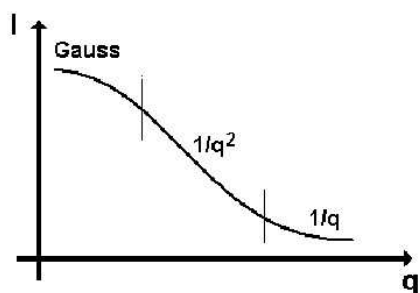


Figura 6 – Intensidade teórica de SAXS para random coils.

A parte inicial possui comportamento gaussiano. Em um gráfico de Guinier ($\ln I \times q^2$) esta região fornece o raio de giro da partícula como um todo. O formato da curva de espalhamento a ângulos maiores será dominado por elementos estruturais menores dentro da partícula. A parte final da curva irá, no entanto, ter a forma do espalhamento de uma agulha, ou seja, $I(q) \approx 1/q$ (Glatter e Kratky, 1982) uma vez que corresponde a regiões muito pequenas da molécula que podem ser aproximadas por retas. Logicamente que isso somente será válido se o comprimento destas retas for grande em comparação com o diâmetro da cadeia, o que é verdade para alguns tipos de polímeros e para proteínas completamente desnaturadas, por exemplo.

A parte intermediária das curvas de intensidade originada pelas subseções maiores da molécula pode ser visualizada como um arranjo randômico dos elementos da cadeia seguindo a relação $\sim 1/q^2$ (Glatter e Kratky, 1982). Em um gráfico $I.q^2$ vs. q esta região ficaria horizontal. Neste mesmo gráfico a parte $1/q$ segue uma tendência ascendente cujo prolongamento passa pela origem.

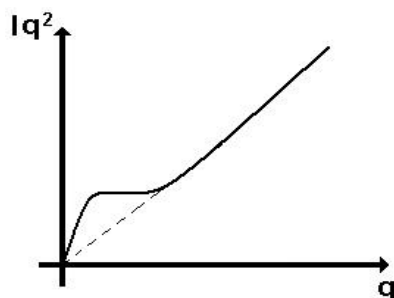


Figura 7 – Gráficos de Kratky para random coils.

Em seu estado nativo, podemos tomar uma proteína como sendo uma partícula homogênea, compacta e com uma interface bem definida[‡]. Assumindo uma conformação globular (que é uma forma comum para proteínas) também podemos dividir a curva de espalhamento em 3 regiões.

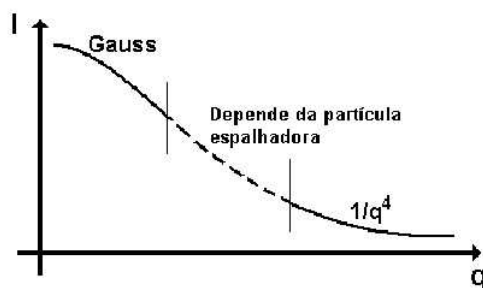


Figura 8 – Intensidade de SAXS para partículas globulares compactas e com interface bem definida com o meio.

Na parte inicial temos um comportamento exponencial de onde podemos extrair o raio de giro da partícula. A região central dependerá da forma da partícula espalhadora, podendo diferir bastante de uma partícula para outra. A parte final em geral seguirá a lei de Porod ($I \propto 1/q^4$) que corresponde a interfaces abruptas. Esta parte também pode conter modulações decorrentes da forma da partícula e/ou de sua estrutura interna.

Desta forma, para um sistema monodisperso de partículas em solução é possível prever qualitativamente a forma do gráfico de Kratky $I.q^2$ vs. q :

Parte inicial:

$$I(q).q^2 \propto e^{-q^2} . q^2 \cong \left(1 - q^2 - \frac{q^4}{3} + \dots\right) . q^2 = q^2 - q^4 - \frac{q^6}{3} + \dots$$

^{‡ ‡} Uma breve descrição das propriedades principais de proteínas é mostrada no apêndice A desta tese.

Parte central:

Depende da partícula espalhadora

Parte Final:

$$I(q) \cdot q^2 \propto \frac{1}{q^4} \cdot q^2 = \frac{1}{q^2}$$

Assim teremos:



Figura 9 – Gráfico de Kratky para partículas globulares compactas com interface bem definida.

A parte central para o caso de curvas suaves sem oscilações decorrentes de correlações intrapartícula fará com que a curva tenha a forma de um sino. No processo de desnaturação de uma proteína, independente do método de denaturação, partimos de um arranjo completamente compacto da proteína para uma situação onde suas cadeias estão soltas podendo este último ser imaginado como uma *random coil*:



Figura 10 – Modelos esquemáticos para proteínas em diferentes graus de enovelamento.

Com base no que foi apresentado anteriormente, podemos descrever qualitativamente a forma do *Gráfico de Kratky* em cada caso.

No primeiro caso temos a proteína no estado nativo, compacto e assim esperamos uma curva com formato tipo sino, contendo um máximo, indo a zero para grandes valores de q (Figura 11A).

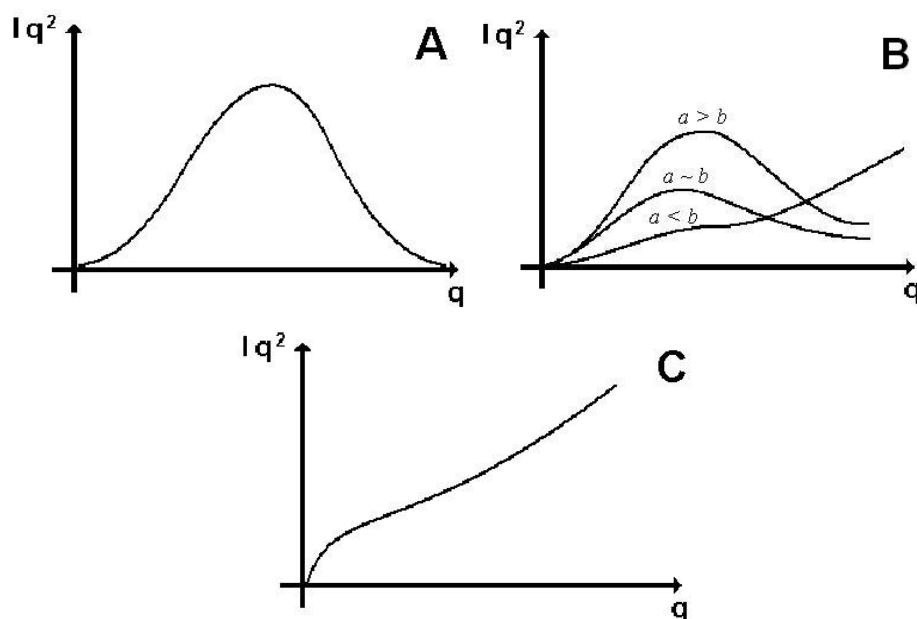


Figura 11 – Variação do Gráfico de Kratky para uma proteína. (A) Gráfico de Kratky para proteína globular compacta, (B) em estado crescente de denaturação e (C) completamente denaturada.

No estado intermediário (que pode constituir várias etapas) teremos uma mescla entre partes compactas da proteína e partes com comportamento de *random coil*. Assim de modo geral a curva de intensidade terá duas contribuições na parte final:

$$\lim_{q \rightarrow \infty} I = \frac{a}{q^4} + \frac{b}{q} \Rightarrow \lim_{q \rightarrow \infty} I \cdot q^2 = \frac{a}{q^2} + b \cdot q$$

dependendo da relação entre a parte compacta e a parte *random coil* da proteína (ou de modo simplificado, o valor das constantes a e b) teremos o comportamento da curva (Figura 11B). No estado de completa desnaturação teremos o comportamento típico de *random coil* (Figura 11C)

Com base nesta descrição qualitativa pode-se explicar o comportamento dos gráficos de Kratky no estudo de proteínas no estado nativo e em processos de desnaturação. Além disso, tem-se informações pertinentes ao estado de compactação da proteína em seu estado nativo.

1.6.3. Função de distribuição de distâncias e anisometria de partículas

Em princípio a função de distribuição de distâncias $p(r)$ contém a mesma informação da intensidade de espalhamento $I_1(q)$, mas a representação no espaço real fornecida pela função $p(r)$ é mais intuitiva e informações sobre a forma da partícula podem geralmente ser deduzidas por uma análise visual direta da forma da função $p(r)$ (Glatter e Kratky, 1982). Nas figuras abaixo mostramos cálculos teóricos da intensidade de espalhamento e da correspondente função de distribuição de distâncias.

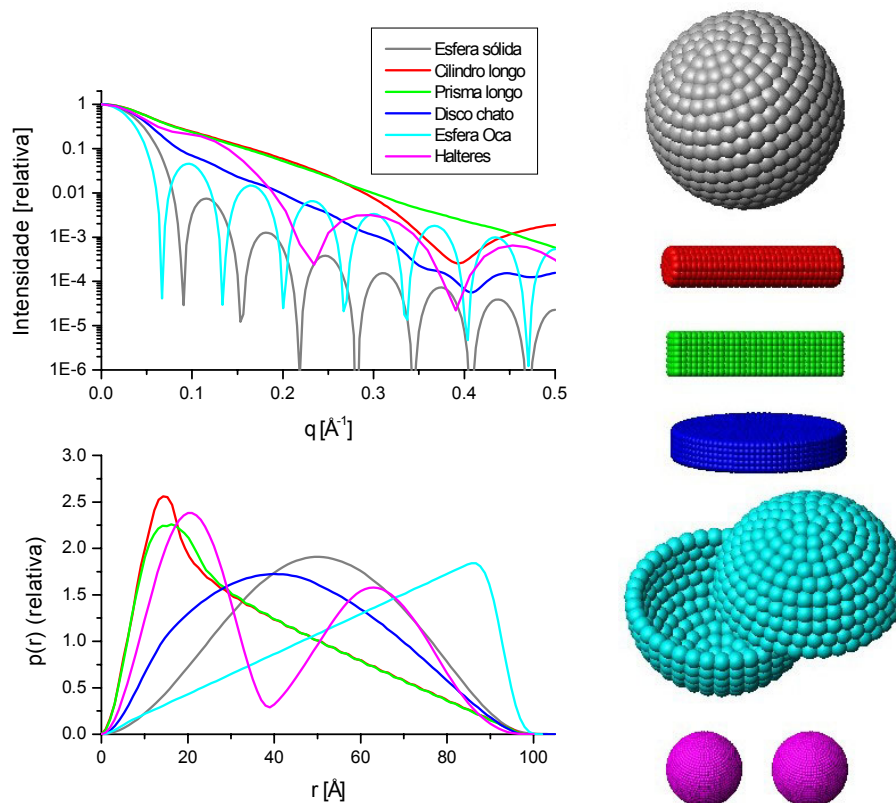


Figura 12 – Intensidades de espalhamento e funções de distribuição de distâncias para corpos simples.

Na Figura 12 mostramos curvas típicas de espalhamento e função de distribuição de distâncias para corpos geométricos com mesmo tamanho máximo. Partículas globulares (curva 1) fornecem uma função $p(r)$ com forma de “sino” (*bell-shaped*) com valores máximos em torno de $D_{\max}/2$. Partículas com algum tipo de anisotropia possuem um

deslocamento deste máximo para menores valores de r na curva $p(r)$ bem como uma diminuição das oscilações da intensidade espalhada. Partículas alongadas, por exemplo, possuem distribuições assimétricas com um máximo claro em distâncias pequenas, correspondendo ao raio médio de sua seção transversal. Partículas achatadas (*flattened*) possuem um máximo mais alargado e, como dissemos, deslocado a distâncias menores do que $D_{\max}/2$. Já um máximo deslocado a distâncias maiores do que $D_{\max}/2$ é usualmente uma indicação de que a partícula é oca (Glatter e Kratky, 1982). Partículas consistidas de subunidades bem separadas podem fornecer vários máximos, o primeiro correspondendo a distâncias intrasubunidade e a outra correspondendo à distância entre as subunidades. As diferenças nos próprios padrões de espalhamento permitem facilmente detectar objetos esféricamente simétricos como aqueles que possuem mínimos acentuados. Partículas com grande anisometria fornecem curvas que decaem mais lentamente do que partículas globulares, entre outras características. Porém as intensidades de espalhamento são em geral menos ilustrativas do que as funções $p(r)$.

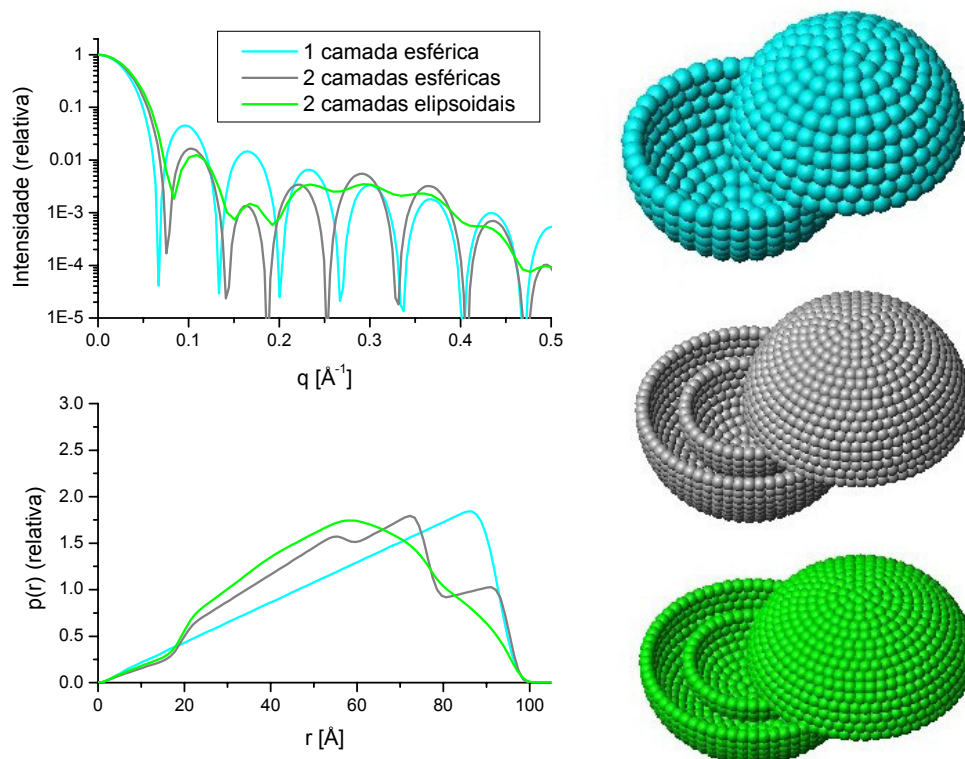


Figura 13 – Intensidades de espalhamento e funções de distribuição de distâncias para corpos ocos.

Para partículas com algum tipo de periodicidade intrínseca as funções de intensidade e $p(r)$ também fornecerão informações sobre sua forma. Isso é mostrado na Figura 13. No primeiro caso temos novamente a esfera oca (monocamada) e posteriormente uma bicamada esférica e uma bicamada elipsoidal. Das curvas vemos que a função $p(r)$ da bicamada esférica fornece 3 máximos bem definidos nos valores $r \sim 55\text{\AA}$, $\sim 75\text{\AA}$ e $\sim 95\text{\AA}$. É interessante notar que estes máximos se referem aos diâmetros das camadas esféricas, considerando o centro de massa das subunidades esféricas, isso é, como a camada esférica mais externa deve ter $100\text{\AA}$ de diâmetro e o raio das subunidades esféricas é $2.5\text{\AA}$ o diâmetro dos centros de massa destas subunidades seria $95\text{\AA}$. Sendo assim, a $p(r)$ fornece diretamente o número de camadas e o raio delas neste caso. O valor $55\text{\AA}$ corresponde à camada menor e os $75\text{\AA}$ correspondem à distância entre um lado da camada menor e o lado diametralmente oposto da camada maior. Na curva de intensidade a presença das multicamadas se reflete em uma modulação da intensidade. No entanto esta riqueza de informações se perde quando deixamos de ter a simetria esférica das camadas, passando a ter camadas elipsoidais.

Mesmo para corpos geométricos simples existem somente alguns casos em que é possível calcular-se analiticamente $I_1(q)$ e $p(r)$. O caso mais conhecido são as expressões para uma esfera sólida com raio R :

$$I_1(q) = \left[\frac{4\pi R^3}{3} \left(\frac{\sin qR - qR \cos(qR)}{(qR)^3} \right) \right]^2 \quad (67)$$

$$p(r) = \left(\frac{4\pi R^3}{3} \right) r^2 \left(1 - \frac{3t}{4} + \frac{t^3}{16} \right), \quad t = r/R \quad (68)$$

Expressões semi-analíticas para a intensidade de elipsóides, cilindros e prismas foram derivados por Mittelbach e Porod (Glatter e Krakty, 1982), e posteriormente expressões analíticas para a função $p(r)$ foram obtidas para partículas com as mesmas geometrias (Feigin e Svergun, 1987). Nos cálculos das intensidades anteriormente apresentadas a construção dos objetos e o cálculo das intensidades foi feita utilizando-se o programa BUILDER (Oliveira, 2001). Este programa utiliza a técnica de elementos finitos para construir as partículas espalhadoras e permite o cálculo de funções numéricas para corpos com variados tipos de complexidade bem como composição dos mesmos.

Na análise de dados experimentais um cálculo confiável da função de distribuição de distâncias $p(r)$ é um pré-requisito necessário para análises posteriores em termos de modelos tridimensionais. Transformações diretas de Fourier para dados experimentais usando a equação (58) não são possíveis uma vez que, em casos reais, os dados experimentais de $I(q)$ são medidos somente em um número finito de N pontos q_i em um intervalo finito $[q_{\min}, q_{\max}]$. Devido a isso, métodos indiretos mostram-se mais eficientes. Existem várias implementações do método de transformação indireta, diferindo no tipo de funções ortogonais que são utilizadas para representar $p(r)$ e também nos métodos numéricos utilizados. Para os cálculos desta utilizaremos dois programas: ITP desenvolvido por Glatter (Glatter, 1977) e GNOM desenvolvido por Semenyuk e Svergun (Semenyuk e Svergun, 1991).

Os métodos de transformação indireta mostraram-se muito eficientes devido ao fato de poderem impor vínculos fortes no cálculo, como condições de contorno para os valores da função e também a suavidade da $p(r)$. Na maioria dos casos requer-se uma estimativa para o valor de $D_{\max}$ (mesmo que tentativa) e posteriormente o ajuste é iterativamente refinado. Os valores da intensidade na origem e do raio de giro obtidos pela função $p(r)$ através das equações (58) e (59) com o uso do método de transformação indireta fornecem valores mais confiáveis do que aqueles obtidos pela aproximação de Guinier, uma vez que a função $p(r)$ é menos sensível a cortes dos dados experimentais a baixos ângulos (Glatter, 1988). Nos métodos indiretos o requerimento de ter-se um número suficiente de pontos para $q < 1.3/R_g$ no gráfico de Guinier é substituído pela condição menos restrigente $q_{\min} < \pi/D_{\max}$ (Glatter e Kratky, 1982).

1.6.4. Informação contida em curvas de espalhamento a baixos ângulos

Como vimos no item anterior, alguns parâmetros gerais (R_g , $D_{\max}$) podem ser obtidos diretamente da curva experimental sem assumir-se nenhum modelo e alguns outros podem ser obtidos quando assumimos que a partícula é aproximadamente homogênea (M_w , Volume V e área S). Isto faz com que venha a tona à questão sobre o número máximo de parâmetros independentes que podem, em princípio, ser extraídos dos dados de espalhamento. Uma medida da informação contida em uma curva de intensidade de

espalhamento $I(q)$ pode ser obtida pelo teorema de amostragem de Shannon (Shannon e Weaver, 1949; Moore, 1980). Deste teorema se deduz a equação (69), a qual indica que a função contínua $I(q)$ pode ser representada por seus valores em um conjunto discreto de pontos (*Shannon channels*) sendo $q_k = k\pi / D_{\max}$, o que faz com que $I(q)$ se torne uma função analítica em uma base que contenha um dado número de parâmetros.

$$qI(q) = \sum_{k=1}^{\infty} q_k I(q_k) \left[\frac{\sin D_{\max}(q - q_k)}{D_{\max}(q - q_k)} - \frac{\sin D_{\max}(q + q_k)}{D_{\max}(q + q_k)} \right] \quad (69)$$

O número mínimo de parâmetros (ou graus de liberdade) requeridos para representar a função analítica no intervalo $[q_{\min}, q_{\max}]$ é dado pelo número de *canais de Shannon* (*Shannon channels*) $N_S = D_{\max}(q_{\max} - q_{\min}) / \pi$ neste intervalo.

Sendo assim, o número de *Shannon channels* fornece um guia muito útil em uma medida experimental indicando, por exemplo, o valor de $q_{\min}$ que deve ser menor do que o *primeiro canal de Shannon* ($q_{\min} < \pi / D_{\max}$). Isto obviamente coloca alguns limites no uso de transformações indiretas que descrevemos anteriormente. Na prática, como as intensidades de espalhamento decaem rapidamente com q e são geralmente medidas com resoluções menores do que $10\text{\AA}$, temos que o número de *Shannon Channels* tipicamente não ultrapassa 10-15. Apesar de ser extremamente simplista afirmar-se que N_S limita o número de parâmetros que podem ser extraídos de dados de espalhamento, este número fornece um parâmetro inicial para a interpretação e modelagem de dados experimentais. O nível de detalhe dos modelos que pode ser deduzido dos padrões de espalhamento dependem não somente dos valores de N_S , mas também de outros fatores como a precisão dos dados e informações conhecidas *a priori*. Este tema será posteriormente abordado item de modelagem *ab initio* e nos casos experimentais este tema.

1.6.5. Metodologia *Ab initio* para modelagem da forma da partícula

Pela análise da quantidade de informação contida em curvas de espalhamento a baixos ângulos fica claro que a reconstrução de um modelo tridimensional de um objeto a partir da curva unidimensional de espalhamento é um problema mal definido (Svergun e Stuhrmann, 1991). Para simplificar a descrição dos modelos de baixa resolução que podem

ser obtidos, a interpretação dos dados é geralmente feita em termos de corpos homogêneos (a influência das inhomogeneidades internas pode ser grandemente eliminada pela subtração de uma constante como descrito anteriormente). No passado, a modelagem da forma era feita por tentativa e erro, calculando-se a intensidade de espalhamento para diferentes formas e comparando-as com dados experimentais. Os modelos podem ser calculados com expressões analíticas ou semi-analíticas disponíveis para corpos simples (esferas, prismas, elipsóides, etc) ou formas construídas utilizando um arranjo regular de subunidades (*beads*). Para estes *bead models* a intensidade espalhada é calculada utilizando-se a fórmula de Debye citada anteriormente (Glatter, 1972),

$$I(q) = \sum_{i=1}^K \sum_{j=1}^K f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \quad (70)$$

sendo K o número de esferas (*beads*), $f_j(q)$ a amplitude de espalhamento da j -ésima subunidade (usualmente uma esfera rígida) e $r_{ij} = |\vec{r}_i - \vec{r}_j|$ a distância entre um par de esferas. Este tipo de modelagem permite a construção de modelos complicados, mas deve ser vinculado a informações adicionais (por exemplo, microscopia eletrônica ou dados hidrodinâmicos, entre outras possíveis informações).

1.6.5.1. Método de esféricos Harmônicos

Historicamente o primeiro método para determinação *ab initio* de forma de partículas espalhadoras foi proposto por Stuhrmann (Stuhrmann, 1970). A forma da partícula era representada por uma função angular de envelope $F(\omega)$, descrevendo as fronteiras da partícula em coordenadas esféricas (r, ω) . Esta função é parametrizada como

$$F(r, \omega) \approx F_L(r, \omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega) \quad (71)$$

onde $Y_{lm}(\omega)$ são esféricos harmônicos e os coeficientes multipolares f_{lm} números complexos. Para uma partícula homogênea, a densidade dada por,

$$\rho_c = \begin{cases} 1, & 0 \leq r < F(\omega) \\ 0, & r \geq F(\omega) \end{cases} \quad (72)$$

e a intensidade de espalhamento da forma é expressa por (Stuhrmann, 1970)

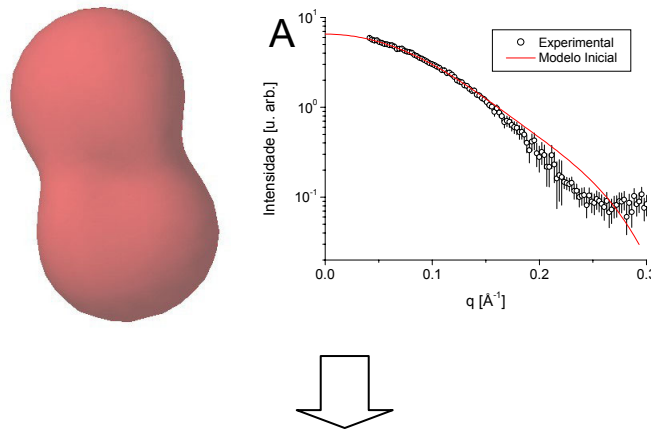
$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}(q)|^2 \quad (73)$$

sendo as amplitudes parciais $A_{lm}(q)$ calculadas a partir dos coeficientes de forma f_{lm} utilizando uma fórmula de recorrência baseada em coeficientes 3j-Wigner (Svergun e Stuhrmann, 1991). Estes coeficientes são determinados por uma otimização não linear, iniciando em um aproximação esférica à qual são adicionados outros componentes de modo a minimizar a discrepância χ entre dados experimentais e calculados,

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{\text{exp}}(q_j) - \eta I(q_j)}{\sigma(q_j)} \right]^2 \quad (74)$$

sendo η um fator de escala. O valor de truncamento L na equação (71) define o número de parâmetros independentes N_p , os quais, para envelopes de baixa resolução, são comparáveis com o número de canais de Shannon (*Shannon channels*) dos dados experimentais. Em geral, $N_p = (L+1)^2 - 6$, ou seja, necessita-se de 10-20 parâmetros para $L=3-4$, sendo que este número pode ser reduzido para corpos simétricos. Este método – implementado no programa SASHA (Svergun et al., 1996) – foi a primeira ferramenta automática para determinação ab initio a partir de dados de SAS.

Um exemplo de modelagem utilizando este método é mostrado abaixo, com dados de espalhamento para lisozima[§]:



[§] Uma breve descrição sobre as propriedades principais de proteínas é mostrada no apêndice A desta tese.

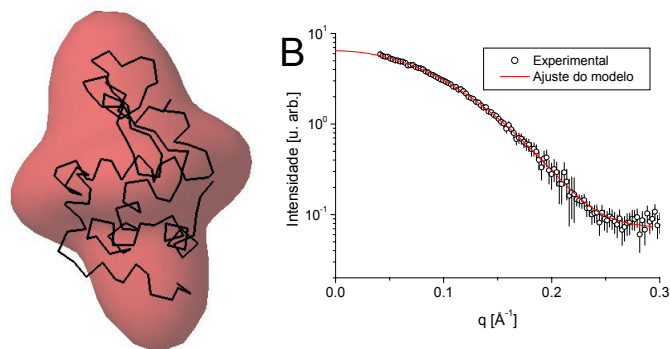


Figura 14 – Modelagem *ab initio* usando esféricos harmônicos. A) A otimização começa com poucos esféricos harmônicos. B) Ajuste final onde os melhores coeficientes f_{lm} foram obtidos pelo programa. O “esqueleto” da lisoizima foi sobreposto ao modelo para melhor visualização.

O formalismo de esféricos harmônicos mostrou-se ser extremamente útil em análises de dados de SAS sendo aplicado em vários outros métodos. Devido ao pequeno número de parâmetros, este método fornece solução única na maioria dos casos práticos e devido ao sucesso em várias aplicações, demonstrou-se que curvas de SAS *continham informações sobre a forma da partícula espalhadora*.

1.6.5.2. Método de elementos finitos (*dummy atom models*)

A utilização de funções angulares de envelope é limitada a partículas com formas relativamente simples (em particular, sem buracos em seu interior). Uma descrição melhor é alcançada em métodos baseados na idéia de elementos finitos (*bead methods*) (Chacón et al., 1998; Svergun, 1999), que utilizam o enorme poder de cálculo dos modernos computadores, revivendo as idéias da modelagem com a fórmula de Debye por tentativa e erro. Um volume (usualmente) esférico com diâmetro D_{max} é preenchido com M subunidades em empacotamento hexagonal (*close packed* – esferas com raio $r_0 \ll D_{max}$). Cada uma das subunidades pertence tanto à partícula (índice = 1) ou ao solvente (índice = 0), e a forma é então descrita por uma cadeia binária X de comprimento M . Iniciando em uma distribuição randômica de 0's e 1's, o modelo é randomicamente modificado utilizando métodos tipo Monte Carlo de modo a obter-se a cadeia X que ajusta os dados experimentais. Como os modelos de busca geralmente contem milhares de subunidades, a solução deve possuir vínculos. Por exemplo, no procedimento por *simulated annealing* (Ingber, 1993) implementado no programa DAMMIN (Svergun, 1999), um termo de

penalidade explícita $P(X)$ é adicionado à função de ajuste $f(X) = \chi^2 + P(X)$ de modo a garantir a compactação e conectividade da forma resultante. Também, ao invés de utilizar a fórmula de Debye, a intensidade é calculada utilizando esféricos harmônicos de modo a acelerar a computação. Outro ganho de velocidade é obtido pelo não recálculo da intensidade do modelo completo após cada modificação, mas somente atualizando a contribuição das subunidades alterando seu respectivo índice.

Além do programa DAMMIN diversos outros programas utilizando variados algoritmos e estratégias foram desenvolvidas. Por exemplo, o primeiro método de modelagem ab initio por *bead models* (DALAI_GA, Chacón et al., 1998) utilizando algoritmo genético não impunha vínculos explícitos, mas a solução é implicitamente vinculada pela gradual diminuição do raio das subunidades r_0 durante a minimização. Em sua última versão (Chacón et al., 2000) condições explícitas de conectividade foram adicionadas. Também, estratégias para modelagem ab initio baseados em Métodos Monte Carlo sem restringir o espaço de busca foram implementados. Um procedimento tipo '*given-take*' (Walther et al., 2000) implementado no programa SAXS3D coloca subunidades em uma rede hexagonal e, em cada passo, uma subunidade é adicionada, removida ou transladada de modo a melhorar a concordância com os dados experimentais. Por fim, o programa SASMODEL (Vigil et al., 2001) não utiliza uma rede fixa para as subunidades, mas representa o modelo pela superposição de elipsóides interconectados, utilizando uma busca de Monte Carlo (ou, em uma implementação posterior com otimização por algoritmo genético, Heller et al., 2002) de suas posições e tamanhos para ajustar os dados experimentais. De modo geral todos os métodos mencionados conseguem, com maior ou menor eficiência, dependendo do problema, restaurar satisfatoriamente a forma de baixa resolução a partir dos dados de espalhamento em solução (Takahashi et al., 2003).

Em diversos cálculos mostrados nesta tese utilizei o programa DAMMIN. Um exemplo de modelagem utilizando este método é mostrado abaixo:

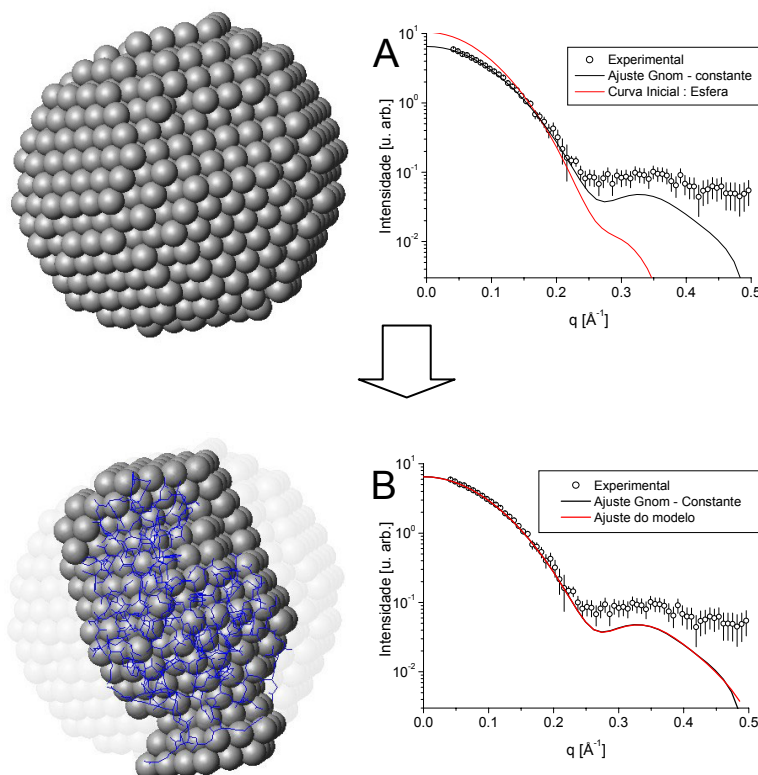


Figura 15 – Ajuste por modelos tipo "dummy atom". A) Início da otimização. Um espaço de busca esférico é criado e aleatoriamente o programa seleciona a composição de subunidades que melhor ajusta os dados experimentais. B) Ajuste final após otimização. A estrutura da lisozima é sobreposta ao modelo obtido.

1.6.5.3. Método de pseudo resíduos (*chain like models*)

Uma limitação fundamental dos métodos descritos anteriormente, o fato de ter-se de assumir uma densidade uniforme para a partícula, limita a resolução à 20-30 $\text{\AA}$ bem como a como a confiança nos modelos, de modo que somente porções restritas dos dados podem ser ajustados. Uma estratégia mais versátil para reconstruir modelos de proteínas a partir de dados de SAXS foi recentemente proposta (Petoukhov et al., 2001), sendo a proteína representada por um arranjo de *dummy residues* (DR). O número M de resíduos é usualmente conhecido pela seqüência primária da proteína ou pela seqüência do DNA transportador, sendo o objetivo da modelagem obter as coordenadas dos M DRs que melhor ajusta os dados experimentais, construindo então uma estrutura tipo proteína. Este método, implementado no programa GASBOR (Svergun et al, 2001; Petoukhov et al., 2002; Petoukhov et al., 2003), inicia com uma distribuição esférica dos DRrs em um volume de diâmetro D_{max} . Os DRs são relocados randomicamente no espaço de busca seguindo um

protocolo tipo *simulated annealing*, sendo que o critério de compactação utilizado nos métodos anteriores é substituído pelo requerimento de que o modelo possua um arranjo espacial dos DRs compatível com o que é sabido de cadeias protéicas (*'chain-compatible'*). Em particular, como os átomos C_{α} de resíduos de aminoácidos vizinhos na sequência primária são separados por $\approx 3.8\text{\AA}$, é requerido que cada DR tenha dois vizinhos à distância de $3.8\text{\AA}$.

Um exemplo de modelagem utilizando este método é mostrado abaixo:

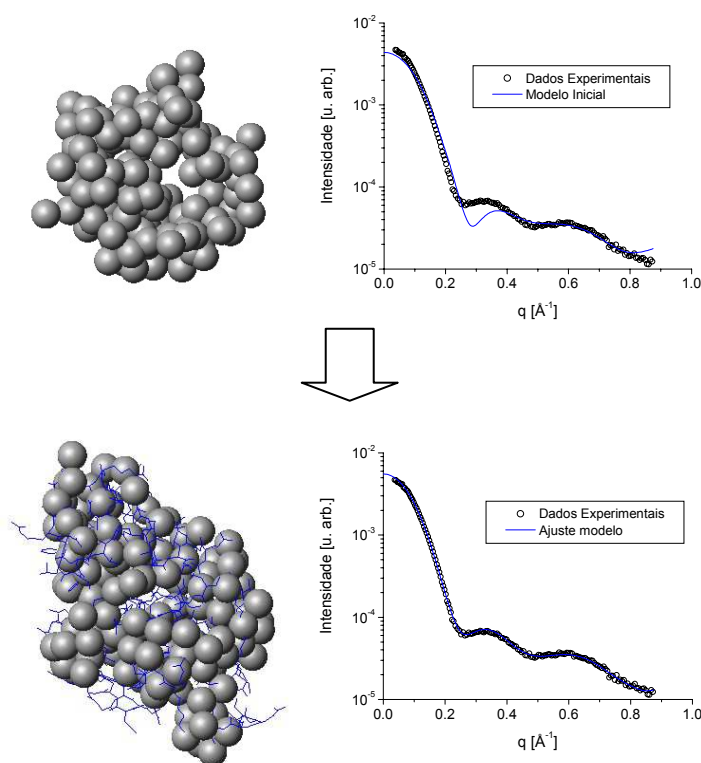


Figura 16 – Ajuste por modelos tipo "chain atom". A) Início da otimização. Uma configuração inicial esférica é proposta para o chain model. No processo o programa busca configuração da cadeia que melhor ajusta os dados experimentais. B) Ajuste final após otimização. A estrutura da lisozima é sobreposta ao modelo obtido.

Comparada com a determinação da forma realizada pelos métodos de esféricos harmônicos e *bead models*, a modelagem por DR aumenta substancialmente a resolução e confiança dos modelos, possuindo potencial para posteriores desenvolvimento. Em particular, este método pode ser aplicado para adicionar-se fragmentos faltantes a modelos incompletos de proteínas (Programa CREDO (Petoukhov et al, 2002)).

1.6.5.4. Promediação de modelos e imposições de simetria e anisotropia

Diferentemente do método de esféricos harmônicos onde, apesar das limitações inerentes ao método, consegue-se um resultado único para a forma da proteína, os métodos tipo elementos finitos (*bead models*) e pseudo cadeias (*chain models*) não fornecem resultados únicos. Como estes métodos iniciam em estruturas randômicas diferentes bem como utilizam métodos baseados em estratégias de Monte Carlo, com mudanças aleatórias em cada passo da otimização, teremos múltiplas soluções, com essencialmente o mesma qualidade de ajuste dos dados. Como proposto por Volkov e Svergun (Volkov e Svergun, 2003) a promediação de modelos independentes permite obter-se a configuração mais provável do respectivo modelo. Dada uma série de modelos calculados independentemente, a obtenção da estrutura mais provável pode ser feita com o pacote de programas DAMAVER (Volkov e Svergun, 2003). Este pacote utiliza um programa de alinhamento de modelos, SUPCOMB (Kozin e Svergun, 2001) que permite o alinhamento de dois modelos arbitrários, tanto de alta quanto de baixa resolução. Os modelos são representados por conjuntos de pontos e o alinhamento é feito minimizando a dissimilaridade medida, denominada discrepância espacial normalizada (*normalized spatial discrepancy - NSD*). Para cada ponto (subunidade ou átomo) no primeiro modelo, o valor mínimo correspondendo a distâncias entre este ponto e todos os outros no segundo modelo é obtido. Estas distâncias são adicionadas e normalizadas pelas distâncias entre pontos vizinhos para os dois modelos. Geralmente valores de NSD próximos da unidade indicam que dois modelos são similares. Todos os pares de modelos são alinhados pelo programa SUPCOMB e o modelo que fornece menor discrepância média com os outros modelos é tomado como referência (*reference model*). Todos os outros modelos, exceto os descartados pelos valores de NSD com relação à referência assumindo um valor limiar, são alinhados com o modelo de referência e o conjunto completo de subunidades ou DRs é remapeado em uma grade de subunidades densamente empacotada onde cada ponto da grade é caracterizado pelo seu fator de ocupação (o número de subunidades no conjunto completo que está na vizinhança do ponto da grade). Os pontos de rede com ocupância não nula formam a região total estendida (*total spread region - TSR*), e a porção de TSR com maior ocupância é selecionada (volume mais populado – *most populated volume MPV*) de modo a

ter-se um volume igual à média do volume excluído dos modelos de entrada no processo de média. O espalhamento calculado do modelo resultante MPV pode não ajustar os dados experimentais perfeitamente, mas este modelo preserva as características mais prováveis da solução. Utilizamos esta promediação em diversos experimentos realizados, como será mostrado posteriormente em *Estudos Estruturais de Proteínas*.

A confiança de modelos *ab initio* pode ser aumentada se informações adicionais sobre a partícula são conhecidas. Em particular, dados sobre anisotropia (prolata e oblata) e simetria permitem um aumento significativo da velocidade cálculo e reduzem o número de parâmetros no modelo. Nos programas mencionados acima SASHA, DAMMIN e GASBOR, restrições de anisotropia e de simetria associados com os grupos espaciais P2-P10 e P222-P62 podem ser impostos.

1.6.6. Cálculo de Intensidades de Espalhamento a partir de Modelos atômicos

Se o modelo de alta resolução da macromolécula inteira ou de seus fragmentos individuais é conhecida (por cristalografia ou NMR, por exemplo), uma interpretação mais detalhada de dados de SAXS é possível. Um pré-requisito necessário para a utilização de modelos atômicos é o cálculo preciso das intensidades de espalhamento em solução, o que não é uma tarefa trivial devido a influência do solvente, mais precisamente da camada de hidratação que se forma no entorno da macromolécula. Em geral a intensidade da partícula em solução pode ser escrita na forma,

$$I(q) = \left\langle \left| A_a(\vec{q}) - \rho_s A_s(\vec{q}) + \delta\rho_b A_b(\vec{q}) \right|^2 \right\rangle_{\Omega} \quad (75)$$

onde $A_a(\vec{q})$ é a amplitude de espalhamento da partícula no vácuo, $A_s(\vec{q})$ e $A_b(\vec{q})$ são respectivamente as amplitudes de espalhamento do volume excluído de solvente e da camada de hidratação. Esta equação leva em consideração o fato de que a densidade de espalhamento do solvente ligado à proteína ρ_b pode diferir daquela da matriz ρ_s , levando a um contraste não nulo da camada de hidratação $\delta\rho_b = \rho_b - \rho_s$. Na realidade, diversas análises mostram que a densidade desta camada é 5% - 20% maior que a da matriz. Esta consideração é crucial para a correta descrição de espalhamento por macromoléculas biológicas em solução (Svergun et al., 1995). Os programas CRY SOL para raios X e

CRYSON para nêutrons (Svergun et al., 1995), circundam a macromolécula por uma camada de hidratação com 3Å de espessura, com densidade ajustável ρ_b . Estes programas utilizam esféricos harmônicos para calcular as amplitudes parciais $A_{lm}(q)$ para todos os termos na equação anterior de tal modo que a média pode ser feita analiticamente (veja equação (73)). Dadas as coordenadas atômicas de uma estrutura, por exemplo obtida do banco de dados *Protein Data Bank* (Berman et al., 2000), estes programas podem tanto ajustar dados experimentais usando dois parâmetros livres, o volume excluído da partícula e o contraste da camada de hidratação $\delta\rho_b$, quanto predizer o padrão de espalhamento usando valores comuns para estes parâmetros. Posteriormente mostraremos exemplos de aplicação deste programa.

1.6.7. Construindo modelos por rotação de corpo rígido

Como vimos no item anterior, podem-se utilizar estruturas de alta resolução para verificar similaridades entre macromoléculas em cristais e em solução. No entanto, modelos estruturais de partículas complexas em solução podem ser construídos a partir de modelos de alta resolução de domínios individuais por refinamento de corpo rígido utilizando dados de espalhamento. Para ilustrar isso, consideremos uma macromolécula composta de dois domínios com estruturas atômicas conhecidas. Se fixarmos o domínio A enquanto transladamos e rotamos o domínio B , a intensidade espalhada da partícula será,

$$I(q, \alpha, \beta, \gamma, \vec{u}) = I_a(q) + I_b(q) + 4\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \text{Re}[A_{lm}(q)C_{lm}^*(q)] \quad (76)$$

onde $I_a(q)$ e $I_b(q)$ são as intensidades de espalhamento dos domínios A e B respectivamente. Os termos $A_{lm}(q)$ são as amplitudes parciais do domínio fixo A , e os termos $C_{lm}(q)$ as do domínio B rotado pelos ângulos de Euler α, β, γ e transladado pelo vetor $\vec{u}$. A estrutura da intensidade espalhada deste complexo depende então de seis parâmetros posicionais e rotacionais e estes podem ser refinados de modo a ajustar-se dados experimentais.

Cálculos com esféricos harmônicos associados a algoritmos eficientes (Svergun, 1991, 1994) permitem resultados suficientemente rápidos para permitir-se uma busca exhaustiva de parâmetros posicionais para o ajuste de dados experimentais a partir de um

complexo, minimizando a discrepância dada na equação (74). No entanto este método automatizado pode, apesar de poder-se ter um ótimo ajuste, falhar no que diz respeito à localização dos contatos entre as subunidades. Nestes casos, informações bioquímicas relevantes (por exemplo, contatos entre resíduos específicos) podem ser levados em conta utilizando um modo de busca interativo. Possibilidades para a combinação de estratégias automáticas e interativas são fornecidas pelos programas ASSA (plataformas UNIX – Kozin et al., 1997) e MASSHA (plataformas Windows – Konarev et al., 2001), disponibilizando ferramentas gráficas tridimensionais associadas com módulos de cálculo que implementam a equação (76). As subunidades podem ser transladadas e rotadas como corpos rígidos enquanto se observa as correspondentes mudanças no ajuste dos dados experimentais podendo-se, adicionalmente, utilizar um refinamento automatizado que faz uma busca exaustiva na vizinhança da configuração corrente. Exemplos da aplicação deste método serão posteriormente mostrados

1.6.8. Determinação do perfil de densidade de espalhamento por SAXS

Nos casos anteriores vimos métodos de determinação da forma de partículas espalhadoras diretamente de dados de espalhamento. Em todos os casos, assumiu-se (implicitamente) que a densidade de espalhamento $\Delta\rho(\vec{r})$ da partícula espalhadora era constante para valores de r dentro da partícula. Porém, existem casos onde $\Delta\rho(\vec{r})$ varia dentro da partícula espalhadora. Variações de densidade eletrônica dentro da partícula espalhadora influenciam a forma da intensidade espalhada $I(q)$ e da função $p(r)$ (Glatter e Kratky, 1982).

Não existe solução geral para a determinação de $\Delta\rho(\vec{r})$ para estruturas tridimensionais arbitrárias. No entanto, existem soluções para simetrias esféricas e cilíndricas com distribuição radial centro-simétrica (densidade sem dependência angular ou axial) e para lamelas centro-simétricas sem inomogeneidades ao longo de seu plano (Glatter, 1991).

É óbvio que não se perde informação fazendo-se a média espacial no caso de simetria esférica. Sendo assim, esperamos determinar a informação unidimensional $\Delta\rho(r)$ a partir da função unidimensional $I(q)$. As equações (29)-(31) mostram a correlação entre

a intensidade espalhada e o quadrado de convolução da densidade de espalhamento $\Delta\rho(\vec{r})$ ou a função $p(r)$. Para partículas com simetria esférica tem-se expressões similares para *amplitude de espalhamento* $F(q)$ e a *densidade* $\Delta\rho(r)$.

Podemos ilustrar isso com o *diagrama de Transformadas de Fourier* do espalhamento a baixos ângulos (Figura 17): as curvas de espalhamento no espaço recíproco, a amplitude $F(q)$ e a intensidade $I(q)$ estão relacionadas à função densidade $\Delta\rho(r)$ e a distribuição de pares de distâncias $p(r)$ no espaço real por uma transformada de Fourier, a intensidade é o quadrado das amplitudes e a função $p(r)$ se relaciona com a convolução quadrática da densidade. A inversão, isto é, o cálculo de $\Delta\rho(r)$ a partir de $I(q)$ é o que iremos discutir.

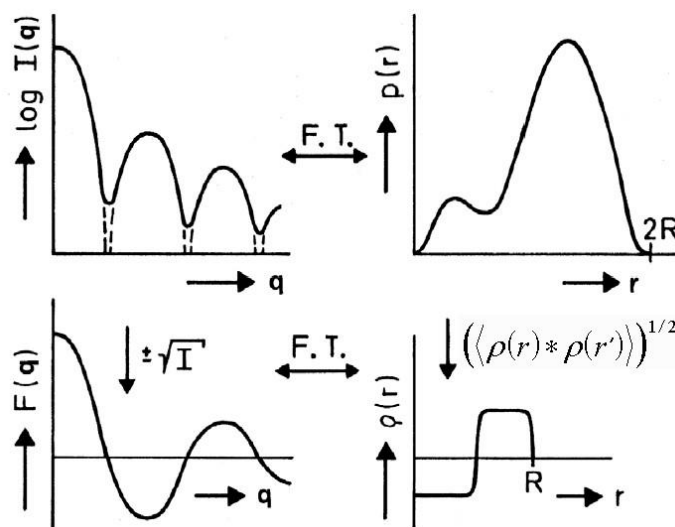


Figura 17 – Diagrama de Transformadas de Fourier em espalhamento a baixos ângulos. Correlações entre a densidade radial $\Delta\rho(r)$ e a função $p(r)$ e suas transformadas de Fourier, a amplitude de espalhamento $F(q)$ e a intensidade espalhada $I(q)$, assumindo-se simetria esférica. Figura baseada em Glatter (2002).

A maneira mais simples para se realizar a inversão é pelas amplitudes de espalhamento, mas o cálculo destas amplitudes a partir das intensidades espalhadas necessita, além de uma simples operação de raiz quadrada, os sinais corretos destas amplitudes. Este é o chamado *problema da fase* (*phase problem*) um problema central e bem conhecido em cristalografia, já mencionado anteriormente. Uma vez conhecidas as fases, o perfil de densidade radial pode ser calculado das amplitudes através de uma transformada de Fourier. Durante muito tempo esta estratégia utilizando amplitudes de

espalhamento foi a única maneira de se calcular perfis de densidade, mas não existe solução única para o problema da fase. Este problema da fase pode ser completamente evitado com outra metodologia, a técnica da raiz quadrada da convolução (*convolution square root technique*), que permite a deconvolução da função de distribuição de pares de distâncias $p(r)$ no perfil de densidade. Esta técnica, implementada no programa DECON (Glatter, 1981), elimina o problema da fase (exceto a arbitrariedade de um fator ± 1 para toda a função $\Delta\rho(r)$). Esta operação pode ser aplicada para funções limitadas em um intervalo, isto é, a função $p(r)$ deve ir para zero após um certo limite. Isto acontece para partículas com simetria esférica onde o limite é simplesmente o diâmetro. Esta condição não é satisfeita para estruturas com periodicidades a longo alcance e desta maneira não pode ser utilizada satisfatoriamente para lipossomas multilamelares (com muitas ordens de reflexão) e em cristalografia.

Técnicas similares de deconvolução existem para sistemas tipo bastão (*rod-like*) e estruturas lamelares. Uma vez que estas estruturas são facilmente reconhecidas pela forma da função $p(r)$ para a partícula toda, a função de distribuição de distâncias da seção transversal $p_c(r)$ e da espessura $p_t(r)$ podem ser calculadas. Estas funções podem ser deconvoluídas em perfis de densidade $\Delta\rho_c(r)$ e $\Delta\rho_t(r)$ pela técnica mencionada.

Os princípios matemáticos da técnica de deconvolução são descritos em diversos artigos (Glatter, 1981; Glatter e Hainisch, 1984; Bergmann et al, 2000) e vão além do escopo desta tese. No entanto, se as condições acima mencionadas são satisfeitas é possível obter-se da função $p(r)$ os perfis de densidade em cada caso. Diversas aplicações, mostradas posteriormente, utilizarão este método.

Na exposição de resultados experimentais os métodos mostrados anteriormente serão utilizados respectivamente à necessidade do problema a ser estudado. No entanto a presente explanação permite uma correta compreensão das metodologias de cálculo utilizadas.

2. Cálculos de Parâmetros Hidrodinâmicos

2.1. Introdução

As propriedades hidrodinâmicas de macromoléculas em solução foram usadas para a determinação da conformação ou forma de proteínas. Antes do desenvolvimento da cristalografia de raios X, as propriedades das macromoléculas em solução como coeficientes de difusão e sedimentação bem como viscosidades intrínsecas foram costumeiramente utilizadas para determinar o tamanho e forma de proteínas, em termos de modelos hidrodinâmicas simples como, por exemplo, elipsóides de revolução. De uma ou mais propriedades pode-se determinar a anisometria (razão axial) e o grau de hidratação de uma proteína.

Na década de 80 esta situação mudou, principalmente por dois desenvolvimentos. Por um lado, a cristalografia de raios X surgiu como uma ferramenta poderosa e comum na determinação da estrutura de proteínas a nível atômico. Mais recentemente, novas técnicas baseadas em ressonância nuclear magnética multidimensional (*multidimensional NMR*) foram desenvolvidas como ferramentas alternativas para determinação estrutural em solução. Isso foi importante para solucionar-se algumas dúvidas quanto a possíveis diferenças na estrutura determinada em um cristal com relação àquela obtida em solução, ou seja, em condições fisiológicas. Também, utilizando espalhamento a baixos ângulos, pode-se obter modelos de baixa resolução para a partícula espalhadora, diretamente da solução de partículas, para um grande intervalo de pesos moleculares. Como a cristalografia de proteínas requer-se cristais de boa qualidade, o que é difícil de se obter, e a técnica de NMR é restrita principalmente para proteínas pequenas, o espalhamento a baixos ângulos permite a obtenção de informações que nenhuma outra técnica poderia fornecer.

Em trabalhos recentes, Garcia de la Torre e colaboradores (Garcia de la Torre et al., 2000) desenvolveram métodos computacionais para determinar propriedades hidrodinâmicas de estruturas em nível atômico. No entanto, estes métodos também podem ser utilizados em modelos *ab initio* obtidos por SAXS. Face aos ótimos resultados deste desenvolvimento para proteínas reais quando comparado com resultados experimentais (de

la Torre et al., 2000), valores de grandezas hidrodinâmicas podem ser preditas com grande precisão a partir das coordenadas atômicas das estruturas protéicas.

Como descrito em detalhe em outra parte desta tese, a modelagem a partir de dados de espalhamento a baixos ângulos fornece estruturas de baixa resolução e muitas vezes diversas dúvidas quanto a confiabilidade dos modelos podem surgir. Neste sentido, a comparação de características destes modelos com resultados hidrodinâmicos experimentais correlacionados à proteína estudada podem servir de indicador de quais modelos podem ser utilizados bem como se a modelagem fornece resultados coerentes.

Esta estratégia foi utilizada em diversos estudos apresentados nesta tese e apresentaremos em linhas gerais a maneira como é feita. No entanto, visando um entendimento geral das propriedades hidrodinâmicas importantes que podem ser obtidas nestes cálculos bem como em experimentos, uma breve descrição destas propriedades será apresentada.

2.2. Propriedades Hidrodinâmicas – Descrição Geral

Para obter-se informações sobre dimensões, formas e pesos de macromoléculas, deve-se estudar sistemas nos quais os processos de transporte das partículas no meio estão ocorrendo, ou seja, os sistemas não estão em equilíbrio. Estes processos como sedimentação e difusão entre outros são irreversíveis no ponto de vista termodinâmico; o sistema em que tal processo ocorre é removido do estado de equilíbrio. A seguir apresentaremos as principais relações que descrevem estes processos (Van Holde, 1971).

2.2.1. Coeficiente de fricção

Imaginemos uma solução com uma molécula de soluto à qual se submete repentinamente uma força externa F . Independentemente da origem desta força, a molécula sofrerá uma aceleração na direção do campo de força. Porém esta aceleração terminará em um tempo muito pequeno (da ordem de nanosegundos), uma vez que, com o aumento de sua velocidade, a molécula experimenta um aumento da resistência para mover-se no meio. Esta força de fricção será dada por $f.v$, sendo v a velocidade e a constante f o *coeficiente de fricção* da molécula. A velocidade constante será atingida quando a força total na molécula é zero:

$$f.v + F = 0 \quad (77)$$

Assim, se medirmos a velocidade de movimento produzido por uma força conhecida, pode-se determinar o coeficiente de fricção. Por exemplo, para uma esfera de raio R , tem-se:

$$f = 6\pi\eta R \quad (78)$$

sendo η a viscosidade do meio. Esta equação é denominada *lei de Stokes* e equações análogas podem ser desenvolvidas para partículas com formas simples (Van Holde, 1971).

2.2.2. Coeficiente de difusão translacional

Dado um sistema contendo moléculas de solvente e soluto, imaginemos uma situação inicial onde existe uma interface abrupta entre a molécula e o soluto pela qual a concentração de soluto varia rapidamente de 0 a um valor C_0 . Das leis da termodinâmica sabe-se que este estado inicial não está em equilíbrio uma vez que a energia livre será mínima somente para uma mistura uniforme.

A mudança no potencial químico do meio com a distância na região de interface induz um fluxo J de soluto que irá levar ao equilíbrio. Pode-se demonstrar que este fluxo se relaciona com o gradiente de concentração no sistema na forma,

$$J_2 = -D_2 \frac{\partial C_2}{\partial x} \quad (79)$$

onde,

$$D_2 = \frac{RT}{N_A f_2} \left\{ 1 + C_2 \frac{\partial \ln y_2}{\partial C_2} \right\} \quad (80)$$

sendo N_A o número de Avogadro, f_2 o coeficiente de fricção da partícula, $R = kN_A$ é a constante universal dos gases com k a constante de Boltzmann, C_2 a concentração de soluto e y_2 um coeficiente de atividade (indica o afastamento do comportamento ideal do líquido).

A quantidade D_2 é definida *coeficiente de difusão* e é a medida da taxa com que as moléculas se deslocam através de gradiente de concentração. A equação (79) é chamada *primeira lei de Fick*, possuindo unidades (CGS) de $\text{cm}^2.\text{segundo}^{-1}$. Para soluções ideais o coeficiente de atividade pode ser desprezado e obtêm-se: $D_2 = RT / N_A f_2$. Além disso, como f_2 não varia com a concentração, D_2 se torna constante.

Utilizando a equação da continuidade,

$$\frac{\partial C_2}{\partial t} = \left(\frac{\partial J_2}{\partial x} \right) \quad (81)$$

tem-se com a equação (79)

$$\frac{\partial C_2}{\partial t} = D \left(\frac{\partial^2 C_2}{\partial x^2} \right) \quad (82)$$

sendo esta equação denominada *segunda lei de Fick*. Diversos métodos experimentais baseados nas equações anteriores foram desenvolvidos para obter-se os coeficientes de difusão para soluções de proteínas (Van Holde, 1971; Atkins, 1988).

2.2.3. Coeficiente de sedimentação

Num campo gravitacional, partículas pesadas se deslocam em uma coluna de solução por um processo denominado *sedimentação*. A taxa de sedimentação depende da intensidade do campo nas massas e das formas das partículas. Moléculas esféricas (e moléculas compactas em geral) sedimentam mais rápido do que partículas anisotrópicas. A sedimentação é normalmente muito lenta, mas este processo pode ser acelerado trocando o campo gravitacional por um campo centrífugo. Isso pode ser feito utilizando uma ultracentrífuga que é essencialmente um cilindro que pode ser rotado em uma alta velocidade ao redor de seu eixo, com a amostra em uma cela próxima de sua periferia (Figura 18). Ultracentrífugas modernas podem produzir acelerações equivalentes a 10^5 vezes a gravidade. Inicialmente a amostra é uniforme, mas a fronteira de soluto move-se para fora no processo de sedimentação.

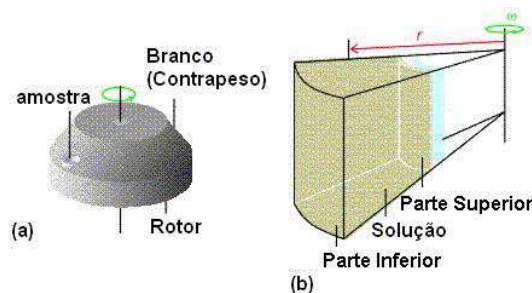


Figura 18 – Esquema de ultracentrífuga. Figura tirada de Atkins, 1988.

Uma partícula de soluto de massa m possui uma massa efetiva $m_{eff} = bm$ em relação ao fator de flutuação b do meio sendo,

$$b = 1 - \rho \cdot \bar{v} \quad (83)$$

sendo ρ a densidade do solvente, $\bar{v}$ o volume parcial específico do soluto ($\bar{v} = (\partial V / \partial m_B)_T$, com m_B a massa total do soluto) e $\rho \cdot \bar{v}$ a massa do solvente deslocada por grama do soluto. As partículas de soluto a uma distância r a partir do eixo de rotação em uma velocidade angular ω experimentam uma força centrífuga de magnitude $m_{eff} r \omega^2$. A aceleração centrífuga é contida pela força de fricção proporcional à velocidade, v , das partículas no meio. Esta força é escrita como $f \cdot v$ sendo f o coeficiente de fricção. As partículas adotam uma velocidade de deslocamento, estacionária através do meio, a qual é obtida equacionando as duas forças $m_{eff} r \omega^2$ e $f \cdot v$. As forças são iguais no processo estacionário e assim,

$$v = \frac{m_{eff} r \omega^2}{f} = \frac{b m r \omega^2}{f} \quad (84)$$

A velocidade de deslocamento depende da velocidade angular ω e do raio r ao centro de rotação, sendo conveniente definir o chamado *coeficiente de sedimentação* ou a *constante de sedimentação*, s , dada por,

$$s = \frac{v}{r \omega^2} \quad (85)$$

Sendo assim, obtemos para o coeficiente de sedimentação,

$$s = \frac{m(1-\bar{v})}{f} = \frac{M(1-\bar{v})}{N_A f} \quad (86)$$

sendo $M = N_A \cdot m$ o peso molecular da molécula de soluto. De acordo com esta equação s é proporcional ao peso molecular multiplicado pelo *fator de flutuação* $(1-\bar{v} \cdot \rho)$ e inversamente proporcional ao coeficiente de fricção. As unidades de s são segundos. Como valores da ordem de 10^{-13} seg são comumente encontrados, define-se 1×10^{-13} como 1 Svedberg. Unidades de Svedberg são convencionalmente denotadas pelo símbolo S .

Utilizando-se equações de estado e assumindo-se uma solução ideal (Van Holde, 1971) obtêm-se que o fluxo J de partículas de soluto em um processo em uma célula ultracentrífuga como sendo dado em função do coeficiente de sedimentação e do coeficiente de difusão:

$$J = s \omega^2 r C - D(\partial C / \partial r) \quad (87)$$

sendo $s = M(1-\bar{v} \cdot \rho) / N_A f$ e $D = RT / N_A f$ como mostrados anteriormente.

2.2.4. Viscosidade Intrínseca

A presença de um soluto macromolecular aumenta a viscosidade da solução. Este efeito é grande mesmo para baixas concentrações, porque moléculas grandes afetam o fluxo de fluido em uma grande região ao redor delas. A baixas concentrações a viscosidade, η , da solução se relaciona com a viscosidade de um solvente puro, η_0 , pela relação

$$\frac{\eta}{\eta_0} = 1 + \nu\phi \quad (88)$$

Sendo ν um fator numérico, com valor desconhecido e ϕ a fração de volume das partículas no meio. Em um tratamento rigoroso do problema, Einstein mostrou que para partículas esféricas $\nu = 5/2$. Para partículas com outras formas, em soluções diluídas, encontramos equações semelhantes a esta com diferentes valores de ν .

A equação (88) se aplica a soluções diluídas, onde se negligencia efeitos de interação soluto-soluto. Estes efeitos levam a termos quadráticos e de maior ordem na fração de volume, devendo-se escrever,

$$\eta_r = \frac{\eta}{\eta_0} = 1 + \nu\phi + \kappa\phi^2 + \dots \quad (89)$$

aqui introduzimos a viscosidade relativa (η_r) para a razão entre as viscosidades do solvente e da solução. Se subtrairmos a unidade deste valor obtemos a viscosidade específica η_{esp} , que é uma medida da mudança fracional na viscosidade produzida pela adição de soluto:

$$\eta_{esp} = \eta_r - 1 = \nu\phi + \kappa\phi^2 + \dots \quad (90)$$

como geralmente é difícil trabalhar na escala de concentrações em fração de volume, uma vez que no preparo de soluções se pesa soluto em um dado volume de solução, podemos reescrever a equação anterior notando que $\phi = C\bar{v}$, sendo C a concentração em g/ml e $\bar{v}$ o volume parcial específico do soluto. Assim, $\eta_{esp} = \bar{v}C + \kappa\bar{v}^2C^2 + \dots$ ou

$$\frac{\eta_{esp}}{C} = \bar{v} + \kappa\bar{v}^2C + \dots \quad (91)$$

Finalmente, no limite de η_{esp}/C com $C \rightarrow 0$ define-se a *viscosidade intrínseca*, $[\eta]$ do meio, que dependerá das propriedades das macromoléculas isoladas uma vez que efeitos de interação são eliminados pela extrapolação. Sendo que a viscosidade é dada em gramas

por centímetro-segundo (CGS), a unidade chamada poise (P), a viscosidade intrínseca $[\eta]$ será dada em mililitros por grama. De acordo com as equações anteriores para determinar $[\eta]$ mede-se a viscosidade relativa para uma série de concentrações de soluto e extrapola-se η_{esp} / C para diluição infinita, o que fornece,

$$[\eta] = \nu \cdot \bar{v} \quad (92)$$

que mostra que a viscosidade intrínseca de uma macromolécula depende de sua forma (através de ν) e de seu volume específico. O valor de ν como 2.5 para uma esfera é um valor mínimo: qualquer desvio de esfericidade ocasiona um valor maior. Para proteínas o volume específico parcial gira em torno de $0.75 \text{ cm}^3/\text{g}$ na maioria dos casos e, desta forma, pode-se estimar o mínimo valor para a viscosidade intrínseca (partículas esféricas, sem hidratação) como sendo $\sim 2 \text{ ml/g}$. A principal dificuldade na interpretação de dados de viscosidade intrínseca para proteínas globulares é o fato de que tanto a hidratação quanto a assimetria contribuem para a mesma quantidade. Se assumimos que a água da hidratação possui aproximadamente a mesma densidade do água solvente, a hidratação de δ g de água / g de proteínas “secas” leva a um volume parcial específico corrigido $\bar{v}' \equiv \bar{v}(1 + \delta \bar{v}_0 / \bar{v})$ (Van Holde, 1971), sendo $\bar{v}$ e $\bar{v}_0$ o volume parcial específico da macromolécula e da água respectivamente. Com esta consideração tem-se para a viscosidade intrínseca,

$$[\eta] = \nu(\bar{v} + \delta \bar{v}_0) \quad (93)$$

Se δ é conhecido (ou estimado), a razão axial da partícula pode ser calculada a partir de ν (imaginando-se modelos simples de elipsóides), ou tendo-se ν para uma partícula conhecida, estima-se δ .

Pelo apresentado anteriormente, todas as grandezas hidrodinâmicas apresentam limitações quanto à informação sobre tamanho e forma das partículas na solução. No entanto, a combinação destes parâmetros (viscosidade intrínseca, coeficiente de fricção e coeficiente de difusão, por exemplo) pode fornecer informações suficientes para poder tirar conclusões tanto sobre a forma como a hidratação das macromoléculas.

2.3. Simulação de Propriedades Hidrodinâmicas Para Modelos de Partículas

Em artigos recentes, Garcia de la Torre e colaboradores (de la Torre et al, 2000) desenvolveram várias metodologias para predição de propriedades hidrodinâmicas de modelos compostos tanto de estruturas cristalográficas quanto de *bead models*. A descrição completa das metodologias de cálculo envolvidas bem como das equações relacionadas está fora do escopo desta tese, podendo ser encontrados em diversos trabalhos (de la Torre et al, 2000 e referencias citadas nestes trabalho). No entanto, podemos descrevê-lo em linhas gerais. Dado um conjunto de coordenadas atômicas (ou das subunidades do modelo), os passos para a modelagem são esquematizados na Figura 19. Primeiramente todos os átomos (excetuando-se os hidrogênios) são representados por esferas idênticas de raio a ($\sim 1.8\text{\AA}$). As diferenças nos tamanhos entre carbonos, hidrogênios e oxigênios são promediadas por questão de simplicidade. O modelo resultante é praticamente livre de superposições entre as subunidades. Seguidamente, para compensar efeitos de hidratação da proteína em solução, assume-se que os valores de a , raio dos elementos atômicos (AER – *atomic elements radius*) são maiores do que os anteriormente assumidos, o que acaba sendo equivalente à adição de uma camada de hidratação, o que serve também para preencher pequenas falhas ou espaços no interior da proteína (Figura 19B). Neste estágio tem-se a partícula hidrodinâmica primária (PHP), sendo composta de um aglomerado compacto de esferas superpostas.

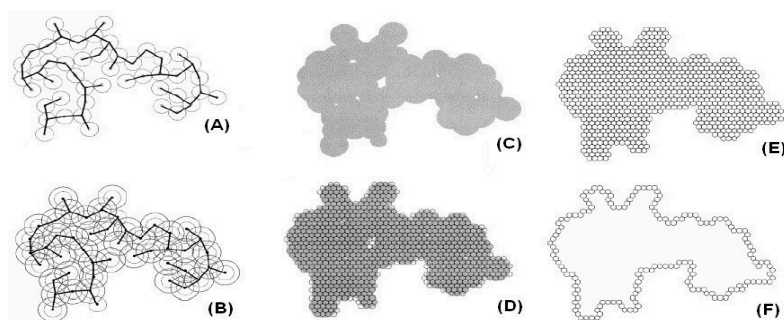


Figura 19 – Representação esquemática da construção do modelo hidrodinâmico em um análogo bidimensional. (A) O modelo de coordenadas atômicas, com esferas representando os átomos com seus raios covalentes. (B) As esferas são expandidas até um valor a de seu raio. (C) O resultante modelo hidrodinâmico primário (PHP). (D e E) O modelo preenchido, superposto ao PHP e sozinho. (F) Modelo de camada externa. Figura tirada de la Torre et al, 2000.

Neste ponto podemos procurar as propriedades hidrodinâmicas da PHP. Modelos construídos para a camada externa (*Shell models*) são adequados para a descrição hidrodinâmica de uma partícula com forma arbitrária. Desta forma aplicam-se ferramentas computacionais para desenvolver a modelagem da camada de uma partícula com forma geral para a PHP. Primeiramente se constrói o modelo preenchido (Figura 19D,E) no qual a partícula é preenchida por subunidades arranjadas em uma maneira mais compacta, uma rede hexagonal. Deste modelo pode-se obter o volume e o raio de giro, R_g , da macromolécula. Então, todas as subunidades que são internas, no sentido de estarem completamente rodeadas por um número de subunidades igual ao número de coordenação da rede (12 neste caso), são removidas. Com isso tem-se acesso ao assim chamado modelo de camada externa (*rough shell model*) (Figura 19F) que representa a superfície da partícula (que é onde as forças hidrodinâmicas agem) como uma camada de pequenas subunidades de raio σ . Para este modelo calculam-se as propriedades hidrodinâmicas (coeficiente de difusão translacional e rotacional e a viscosidade intrínseca). As descontinuidades deste modelo, causadas pelo tamanho discreto das subunidades e das limitações decorrentes de seu tamanho e arranjo de rede, são eliminados pela extrapolação de valores calculados com vários σ 's para o limite de camada correspondendo a $\sigma = 0$.

Este procedimento de cálculo fornece parâmetros hidrodinâmicos para estruturas cristalográficas de proteínas muito próximos dos valores experimentais. Em um estudo comparativo para uma série de proteínas dos mais variados parâmetros obtêm-se para o raio a dos elementos atômicos a ser considerados nos modelos (AER) o valor de médio $3.1\text{\AA}$, o qual fornece erros menores que 5 % no cálculo dos parâmetros. Como será apresentado nesta tese, esta metodologia de cálculo pode ser aplicada à modelos *ab initio* tipo *bead model* e *chain model*, utilizando outros valores para os raios médios das subunidades componentes do modelo.

CAPÍTULO II

ASPECTOS

EXPERIMENTAIS

1. Aspectos Experimentais

1.1. Introdução

O conceito de resolução permite a estimativa da região angular requerida para experimentos de espalhamento por soluções. Imaginemos que em um experimento utilizamos radiação monocromática com comprimento de onda $\lambda = 1\text{\AA}$ para estudar partículas de tamanho característico de $100\text{\AA}$ que requeira uma resolução, digamos de $d = 200$ à $10\text{\AA}$. Se tomamos a relação de resolução $2\pi/d = 4\pi \sin \theta / \lambda$ a correspondente região angular será em torno de 0.005 a 0.1 rad, ou seja, $0.3 - 6^\circ$. O padrão de espalhamento inteiro está então confinado a ângulos muito pequenos, o que leva ao nome da técnica: espalhamento a baixos ângulos (*Small Angle Scattering – SAS*; para raios X *SAXS – Small Angle X ray Scattering* e para nêutrons *SANS – Small Angle Neutron Scattering*)

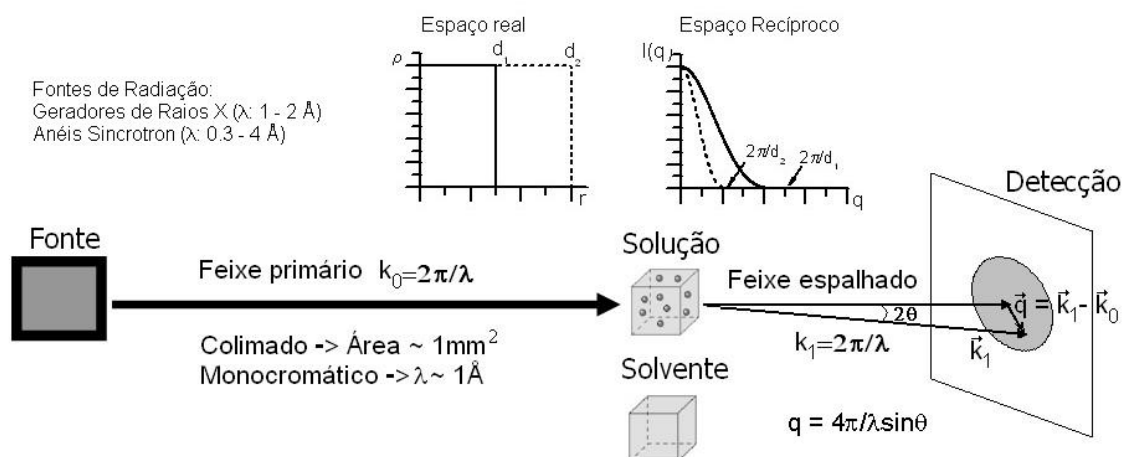


Figura 20 – Representação esquemática de um experimento de SAXS e a transformação de Fourier do espaço real para o espaço recíproco.

Conceitualmente experimentos de SAS são simples (Figura 20), mas o projeto de uma instrumentação eficiente para este fim requer desafios técnicos muito grandes principalmente para conseguir-se espalhamento parasítico (que definiremos melhor posteriormente) baixo nas proximidades do feixe incidente. Além disso, para sistemas biológicos o contraste das partículas é geralmente pequeno e o sinal útil pode ser fraco quando comparado com o espalhamento “branco”. Abaixo descrevemos sucintamente a linha de SAXS do Laboratório Nacional de Luz Síncrotron.

1.2. Linha de Espalhamento de Raios X a Baixos ângulos do LNLS

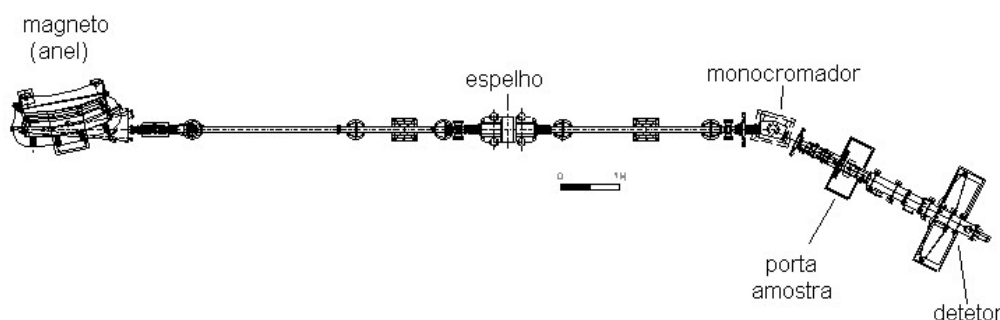


Figura 21 – Esquema da estação experimental de SAXS do LNLS.

A estação experimental de SAXS (linha DA11) do LNLS (Figura 21) opera numa faixa de 8 a 12 KeV e permite acesso a valores de espalhamento entre 0.005 e 1.4 Å⁻¹.

Em seu esquema, possui um espelho de Raios X coberto com uma película de ouro utilizado para a focalização vertical da radiação proveniente do anel sincrotron. O princípio de funcionamento dos espelhos de Raios X baseia-se no fenômeno de reflexão total da radiação por superfícies. Sua curvatura é obtida através da ação de seu próprio peso ao ser apoiado pela sua extremidade, o qual lhe confere um raio de alguns quilômetros.

Após o espelho um módulo de fendas de entrada limita o tamanho do feixe que chega até o monocromador. Ele possui quatro lâminas de Tântalo controladas por motores de passo com movimentos independentes.

Um monocromador de silício seleciona o comprimento de onda da fonte de radiação que deve ser utilizado durante o experimento. O princípio de funcionamento destes monocromadores é a difração de Raios X por monocristais, que segue a lei de Bragg,

$$n\lambda = 2d \sin \theta \quad (94)$$

sendo λ o comprimento de onda selecionado, d é o espaçamento dos planos cristalinos e θ o ângulo de incidência do feixe no cristal. Além de selecionar o comprimento de onda sua forma (curva) permite a focalização da radiação na horizontal.

A estação é montada dentro de uma blindagem que protege o usuário contra a radiação durante os experimentos. Os primeiros componentes, espelho e monocromador, têm sua performance prejudicada com qualquer contaminação de suas superfícies, assim a

parte inicial da linha antes do monocromador opera em condições de alto vácuo (10^{-8} mbar). O restante da estação não necessita de pressão tão baixa, podendo ser utilizadas pressões da ordem de 10^{-3} mbar, o suficiente para reduzir a radiação absorvida pelo ar, reduzindo o espalhamento parasita e protegendo os elementos óticos.

A seleção do comprimento de onda a ser utilizado no experimento é feita através da mudança do ângulo de incidência da radiação sobre o cristal do monocromador. Para cada novo comprimento de onda selecionado, o feixe é difratado em uma posição horizontal diferente. Dessa forma, a parte da estação depois do monocromador deve ser capaz de acompanhar o movimento do feixe monocromatizado. Para isso, toda a instrumentação é montada sobre uma mesa que possui como eixo de rotação o mesmo eixo do monocromador e pode girar no plano horizontal de forma adequada: para uma variação de θ no ângulo de incidência, ocorrerá uma mudança de 2θ da mesa.

Fendas anti-parasitas estão dispostas entre o monocromador e o porta amostras e tem a função de minimizar o espalhamento causado pelos elementos óticos da linha. Essas fendas não tocam o feixe e precisam ser posicionadas o mais próximo possível a este no plano vertical, onde as medidas de espalhamentos são usualmente realizadas.

Na região indicada porta amostra (Figura 21), podemos montar diversos tipos de porta amostras dependendo do tipo de sistemas a ser estudado, (sólidos, líquidos, gasosos etc.).

Ao final da linha estão o detetor e o “*beam stop*”, este ultimo tem a função de impedir que o feixe direto chegue ao detetor e danifique todo o sistema de detecção.

Para a detecção do feixe espalhado a estação experimental possui detetores lineares e bidimensionais. Como o espalhamento de raios X por sistemas diluídos de partículas randomicamente orientadas é isotrópico, basta tomarmos um corte (linear) na frente de onda espalhada para termos o perfil de espalhamento. Não descreverei o funcionamento dos detetores utilizados, mas posteriormente mostro efeitos de *smearing* em detetores lineares.

Informações mais detalhadas sobre as características da linha de SAXS do LNLS podem ser encontrados em vários artigos (Kellerman et al., 1997).

1.3. Métodos para medida de intensidade absoluta

Para espalhamento a baixos ângulos, tanto de Raios X (SAXS) quanto de nêutrons (SANS), a importância da medida de escala absoluta tem sido reconhecida desde a criação das técnicas. Inicialmente, somente o peso molecular ou espalhamento por partícula eram determinados, mas a calibração em escala absoluta pode implicar também a representação da curva de espalhamento em unidades de cm^{-1} , que permite, por exemplo, a determinação direta do contraste de densidade eletrônica em *scattering length units*. A conversão dos dados para a seção de choque de espalhamento por unidade de volume, $d\Sigma(q)/d\Omega$, tipicamente em unidades de cm^{-1} , deve ser realizada pela multiplicação de uma constante de calibração. O uso de escala absoluta não é necessário para a medida da função interferência (fator de forma; fator de estrutura). Por exemplo, o cálculo do raio de giro R_g de macromoléculas em solução, ou a determinação da distribuição de tamanhos de poros em matrizes vítreas podem ser obtidas utilizando dados em escalas relativas. No entanto, colocando os dados em escala absoluta, podemos determinar pesos moleculares além de muitos outros parâmetros já citados.

Apesar das inúmeras vantagens de colocar os dados em escala absoluta, o uso de técnicas padronizadas para calibração não é de modo algum universal no campo de espalhamento a baixos ângulos (Russel, 1988). Esta situação decorre em parte devido ao fato de que muitas das técnicas de calibração necessitam de equipamento especial, além de serem de difícil aplicação e requererem muito tempo de medida (Patel e Schmidt, 1971).

Antes de explicitarmos os métodos é interessante discutirmos a definição da seção de choque diferencial do ponto de vista experimental. De modo similar ao já mencionado anteriormente, a seção de choque de espalhamento $d\sigma(q)/d\Omega$ é definida (Turchin, 1965; Russel, 1988) como a razão entre o número de partículas (nêutrons ou fótons de Raios X) espalhadas por segundo dividido pelo fluxo incidente (partículas $\text{cm}^{-2} \text{s}^{-1}$) por unidade de ângulo sólido e assim possui dimensões de área (cm^2). Se renormalizamos pelo volume iluminado pelo feixe, a seção de choque $d\Sigma(q)/d\Omega$ assume unidades de cm^{-1} . A relação entre a seção de choque de espalhamento e a taxa de contagem medida $I(q)$ ($\equiv dW(q)/d\Omega$ na equação (22)) (contagens s^{-1}) em um elemento de área Δa do detedor e com eficiência de contagem ε , situado em ângulo reto com o feixe incidente à uma distância r da amostra é dada por

$$\left[\frac{d\Sigma}{d\Omega}(q) \right] = \frac{I(q)}{tT} \frac{r^2}{\epsilon I_0 \Delta a A} = \frac{I(q)r^2}{tTK_N} \quad (95)$$

sendo, I_0 a intensidade (contagens $s^{-1} \text{ cm}^{-2}$) na amostra de área A , espessura t e volume $V=at$. A transmissão medida T é dada por $T=e^{-\mu t}$ sendo μ a atenuação linear do feixe quando passa pela amostra. Para espalhamento a baixos ângulos, assumimos que o fator de atenuação é o mesmo para todos os Raios X (ou neutrons) espalhados e que o ângulo sólido subtendido pelo elemento do detetor é independente do ângulo θ (ângulo de espalhamento), pois como vimos $\cos\theta$ é muito próximo da unidade. Como as dimensões de tempo se cancelam tanto no numerador quando no denominador da equação (95), a calibração se reduz à comparação (obtenção do fator $K_N=\epsilon I_0 \Delta a$) com um padrão de seção de choque conhecido na mesma geometria e no mesmo período de tempo.

De modo geral, existem dois métodos diferentes para calibração em escala absoluta. O primeiro, denominado método direto, consiste de uma atenuação mecânica do feixe incidente por um disco rotatório (*Kratky's rotator method*) ou por uma fenda móvel (Kratky e Glatter, 1982). No entanto, esta atenuação mecânica não é bem sucedida quando utilizamos detetores sensíveis a posição (PSDs) (Russel, 1983). O segundo método é o método indireto, utilizando padrões calibrados como Lupolen, carbono vítreo, alumínio irradiado, partículas de sílica ou folhas calibradas (Hendricks, 1972; Shaffer e Hendricks, 1974; Russel et al., 1988). A calibração por folhas é difícil e atenuações muito fortes do feixe incidente podem distorcer a distribuição da radiação. Além disso, quando se utiliza radiação síncrotron, no processo de monocromatização, selecionamos uma família de comprimentos de onda, que satisfazem lei de Bragg (equação (94)). Com isso se escolhermos um comprimento de onda λ , não conseguiremos evitar os harmônicos $..3\lambda$, $2\lambda... \lambda/2$, $\lambda/3....$; os harmônicos de baixa energia podem ser desconsiderados pois são absorvidos pelo material; no entanto os harmônicos de alta energia, apesar de terem intensidade mais baixa, contribuirão para a intensidade transmitida por não serem absorvidos na mesma forma pelo material. Com isso a aplicação da calibração por folhas quando se utiliza radiação síncrotron torna-se bastante difícil, pois teríamos que levar em conta a distribuição espectral do feixe incidente, bem como a absorção dos harmônicos pelo material das fendas. Outro problema consiste no efeito de *smearing* presente no processo de colimação. Apesar de com a luz síncrotron, podermos utilizar um feixe aproximadamente

pontual, efeitos de *smearing* ainda podem ser notados, distorcendo, por exemplo, o início da curva espalhada. Assim, um padrão secundário ideal deveria, idealmente, possuir um espalhamento independente do ângulo. Esta propriedade é conseguida por elementos estruturais ou flutuações de densidade eletrônica presentes na amostra, pequenos o suficiente para não permitir uma dependência angular no regime de baixos ângulos. No entanto, tais amostras geralmente fornecem um espalhamento muito fraco.

Um representante deste tipo de elemento que fornece um espalhamento independente do ângulo devido a pequenas flutuações de densidade eletrônica é a **água**. Como citado por Orthaber et al (Orthaber et al, 2000) a água já tem sido utilizada como padrão secundário em SANS devido ao seu intenso espalhamento incoerente, bem como utilizada em técnicas de espalhamento de luz. Outra importante vantagem é que o espalhamento da água vem simplesmente do número de flutuações e assim depende somente das propriedades físicas da compressibilidade isotérmica. A compressibilidade de fluidos como fonte de espalhamento a baixos ângulos é discutida de maneira bastante detalhada por Guinier e Fournet (1955). A seção de choque de espalhamento foi medida por Hendricks *et al.* (1974), onde fornece uma equação semi-empírica para a variação da seção de choque diferencial de espalhamento para a água em função da temperatura.

Outros tipos de padrões secundários existem. Particularmente para proteínas pode-se utilizar uma proteína conhecida como padrão secundário da proteína de interesse. Este método será detalhado a seguir juntamente com a utilização da água como padrão secundário na linha de SAXS do LNLS. A escolha pela utilização de padrões secundários baseou-se principalmente por sua versatilidade, uma vez que podem ser aplicados a praticamente todos os tipos de porta-amostras disponíveis na estação experimental, além de não requerer a utilização de padrões calibrados.

1.3.1. Utilização da água como padrão secundário

Como vimos, existem vários padrões secundários que poderiam ser utilizados para calibração. No entanto, a utilização da água como padrão é facilitada, pois a intensidade espalhada independe do ângulo de espalhamento na faixa angular de medidas a baixos ângulos. Apesar de o espalhamento da água ser muito fraco, a utilização de radiação síncrotron permite a medição de um padrão razoável com boa estatística em tempo hábil.

A parte horizontal do espalhamento para o range de grandes valores de q se relaciona com a compressibilidade isotérmica da água conforme a expressão,

$$\frac{d\Sigma}{d\Omega} = \rho^2 k T \chi_T \quad (96)$$

onde k é a constante de Boltzmann, T é a temperatura em Kelvin, ρ é a densidade de comprimento de espalhamento e χ_T é a compressibilidade isotérmica. Como χ_T é uma função da temperatura, a intensidade de espalhamento constante da água também o será. Por exemplo para $T=293K$, $\chi_T = 4,58 \times 10^{-10} \text{ Pa}^{-1}$ e temos para a água uma intensidade constante de $1,6456 \times 10^{-2} \text{ cm}^{-1}$. Com estes valores é possível obter-se o espalhamento de uma amostra em relação ao espalhamento da água na mesma configuração e assim converter-se o espalhamento da amostra para escala absoluta. Para obter $d\Sigma/d\Omega$ para uma dada amostra basta utilizarmos a seguinte relação,

$$\left. \frac{d\Sigma}{d\Omega} \right|_{amostra} = \left[\frac{I(0)_{amostra}}{I(0)_{agua}} \right] \left. \frac{d\Sigma}{d\Omega} \right|_{agua} \quad (97)$$

sendo $I(0)_{amostra}$ e $I(0)_{agua}$ o valor da intensidade experimental a ângulo zero para a amostra desejada (obtido por extrapolação) e o valor da intensidade experimental constante da água e $d\Sigma/d\Omega_{agua}$ a seção de choque dada na equação (96).

Esta metodologia na forma que apresentamos faz necessário que, para cada experimento realizado, se meça o espalhamento da água nas mesmas condições. Como dissemos, isso pode ser realizado em um tempo hábil, podendo-se obter uma precisão razoável para o espalhamento da água.

A seguir mostramos uma curva típica de espalhamento da água (Figura 22). Para a medida da intensidade experimental realizamos uma série de 5 exposições com duração de 20 min cada uma:

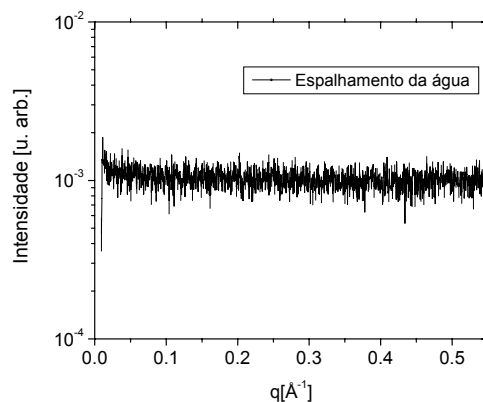


Figura 22 – *Espalhamento da água para calibração.*

1.3.2. Utilização de uma proteína como padrão secundário

Em medidas de SAXS por proteínas em solução um parâmetro muito importante que pode ser obtido dos experimentos é o valor do peso molecular. O método anterior, utilizando a água como padrão secundário, pode ser utilizado com sucesso para este fim, como mostraremos no próximo item. Outra possibilidade de se obter uma estimativa do peso molecular de uma amostra de proteínas em solução (e posteriormente transformar os dados em escala absoluta caso necessário) consiste em utilizar-se outra proteína, com peso molecular conhecido, como padrão secundário. A vantagem principal de utilizar-se outra proteína como padrão secundário é que o tempo requerido para ter-se uma boa curva experimental é muito menor. Da expressão para a intensidade de SAXS de um sistema monodisperso altamente diluído de partículas em solução (equações 56 e 60) temos:

$$I(q) = N \cdot (\Delta\rho)^2 \cdot V_p^2 \cdot P(q) \quad (98)$$

onde N é o número médio de partículas iluminado pelo feixe, $(\Delta\rho)$ é o contraste de espalhamento da partícula, V_p é o volume da partícula e $P(q)$ é o fator de forma normalizado da partícula. Da equação (98), para $q = 0$ temos:

$$I(0) = N \cdot (\Delta\rho)^2 \cdot V_p^2 \quad (99)$$

Em geral o valor do número de partículas geralmente não é conhecido, mas sim o valor da concentração (em mg/ml ou molar – mol/l) das partículas em solução. Se conhecemos a concentração da proteína em solução em mg/ml podemos escrever o número

de partículas N como sendo $N = M_T / M_1$, sendo M_T a massa total de partículas iluminada pelo feixe na cela experimental e M_1 a massa de uma proteína. A massa total pode ser escrita como a concentração de partículas na solução (c - mg/ml) multiplicado pelo volume iluminado da cela (V_c), a saber $M_T = c \cdot V_c$. Assim $N = V_c \cdot c / M_1$ e podemos escrever

$$I(0) = \frac{V_c \cdot c}{M_1} \cdot (\Delta\rho)^2 \cdot V_p^2 \quad (100)$$

da definição de volume específico parcial de uma proteína, sabemos que $V_p = \bar{v} \cdot M_1$ e assim

$$I(0) = \left(\frac{V_c \cdot \bar{v} \cdot \Delta\rho}{M_1} \right)^2 \cdot c \cdot M_1^2 \quad (101)$$

multiplicando-se e dividindo-se pelo número de Avogadro N_A ($N_A = 6.022 \times 10^{23}$ part./mol) e utilizando-se a definição de peso molecular em Daltons, $M_w = N_A M_1$ temos,

$$I(0) = \left(\frac{V_c \cdot \bar{v} \cdot \Delta\rho}{N_A} \right)^2 \cdot c \cdot M_w \quad (102)$$

Vamos assumir que realizamos experimentos com duas amostras de proteínas em solução. A amostra que denominaremos *STD* é nossa amostra padrão com peso molecular conhecido. Já a amostra *A* é a que desejamos obter o peso molecular. Nas mesmas condições experimentais, o volume iluminado V_c é o mesmo nos dois casos. Também para proteínas o volume parcial específico $\bar{v}$ é muito parecido bem como a densidade de espalhamento (mesmo conjunto de aminoácidos componentes das proteínas). Desta forma escrevendo a razão entre as intensidades na origem para as duas amostras temos:

$$\frac{I(0)_{STD}}{I(0)_A} = \frac{\left(\frac{V_c \cdot \bar{v} \cdot \Delta\rho}{N_A} \right)^2 \cdot c_{STD} \cdot M_{wSTD}}{\left(\frac{V_c \cdot \bar{v} \cdot \Delta\rho}{N_A} \right)^2 \cdot c_A \cdot M_{wA}} = \frac{c_{STD} \cdot M_{wSTD}}{c_A \cdot M_{wA}} \quad (103)$$

rearranjando-se os termos, temos:

$$M_{wA} = \left(\frac{I(0)_A / c_A}{I(0)_{STD} / c_{STD}} \right) M_{wSTD} \quad (104)$$

que fornece uma relação direta entre as massas moleculares da amostra padrão e da amostra desejada.

Se a concentração da amostra for conhecida em concentração molar (c_m - mol / l)

basta utilizar na equação anterior a relação entre concentração molar e concentração mássica (c - mg/ml) $c = c_m \cdot M_w$ e obter,

$$M_{wA} = \left(\frac{I(0)_A / c_{mA}}{I(0)_{STD} / c_{mSTD}} \right)^{1/2} M_{wSTD} \quad (105)$$

que permite o cálculo do peso molecular também para esta concentração molar. Se conhecemos a intensidade absoluta para a amostra padrão $I(0)_{STD}$ podemos utilizar a expressão anterior para converter os dados experimentais também em escala absoluta. Utilizando-se como proteína padrão uma que possua estrutura cristalográfica resolvida, a determinação da seção de choque diferencial desta proteína pode ser feita utilizando-se o programa CRY SOL, mostrado anteriormente.

1.3.3. Parâmetros Obtidos da Renormalização em Escala Absoluta

Das equações (56) e (102) temos que a intensidade espalhada por um sistema monodisperso de partículas em solução é dada em termos da seção de choque macroscópica $d\Sigma / d\Omega(q)$ (Orthaber et al., 2000) por,

$$\frac{d\Sigma(q)}{d\Omega} = n \cdot I_1(q) = \bar{v}^2 \cdot (\Delta\rho)^2 \cdot c \cdot M_1 \cdot P(q) \quad (106)$$

onde $n = N/V_c$, com V_c o volume iluminado pelo feixe de raios X, $\bar{v}$ o volume parcial específico da partícula ($\text{cm}^3 \text{g}^{-1}$), c a concentração de partículas (mg/ml), $\Delta\rho$ o contraste de densidade de comprimento de espalhamento (cm^{-2}) entre a partícula e a matriz, M_1 a massa molecular da partícula (em gramas) e $P(q)$ o fator de forma normalizado da partícula. O contraste de densidade de espalhamento depende do conjunto de átomos que compõem a partícula espalhadora e a matriz onde está imersa. Para partículas em solução em sistemas altamente diluídos, o contraste pode ser escrito como

$$\Delta\rho = \frac{1}{V_p} \left(\sum_j f_j \right) - \rho_{solv} = \frac{1}{M_1 \bar{v}} \left(\sum_j f_j \right) - \rho_{solv} \quad (107)$$

sendo V_p o volume da partícula espalhadora, f_j os fatores de espalhamento atômico dos átomos componentes da partícula (Tabela 6) e ρ_{solv} a densidade de espalhamento do solvente.

Tabela 2 - Comprimentos de espalhamento de raios X para alguns elementos.

Átomo	H	D	C	N	O	P	S	Au
Massa Atômica	1	2	12	14	16	30	32	197
N Elétrons	1	1	6	7	8	15	16	79
$f [10^{-12} \text{ cm}]$	0.282	0.282	1.69	1.97	2.16	3.23	4.51	22.3

Como vimos, proteínas possuem centenas ou milhares de átomos de modo que o cálculo explícito do primeiro termo da equação (107) requer que saibamos todos os átomos da proteína em questão bem como seu volume em solução. Como as proteínas possuem em geral a mesma população de resíduos em sua seqüência, costuma-se utilizar valores médios para a estimativa desta densidade de espalhamento. O valor $\bar{v}\Delta\rho = 3.22 \times 10^{23} \text{ e/g}$ é costumeiramente assumido para proteínas (Orthaber et al., 2000). Utilizando o comprimento de espalhamento de um elétron, $0.282 \times 10^{-12} \text{ cm}$, temos

$$\bar{v}\Delta\rho_{\text{proteínas}} = 9.082 \times 10^{10} \text{ cm/g} \quad (108)$$

dado o volume parcial específico $\bar{v} [g / \text{cm}^3]$ de uma dada proteína, obtêm-se o contraste $\Delta\rho [\text{cm}^{-2}]$.

Para o solvente é necessário considerar-se todos os elementos químicos que o compõem. Como em geral o solvente é constituído de certa substância diluída em água, à uma certa molaridade M , a densidade de espalhamento do solvente é dado por,

$$\rho_{\text{solvente}} = \frac{M_{H_2O} \times \rho_{H_2O} + M_{S_1} \times \rho_{S_1} + \dots + M_{S_K} \times \rho_{S_K}}{M_{H_2O} + M_{S_1} + \dots + M_{S_K}} \quad (109)$$

sendo, M_{H_2O} a molaridade da água (55.55 Mol/litro), ρ_{H_2O} a densidade de espalhamento da água à 20°C ($9.44 \times 10^{-10} \text{ cm}^{-2}$), M_{S_i} e ρ_{S_i} a molaridade e densidade de espalhamento de cada um dos componentes adicionais do solvente. Na Tabela 7 mostro as densidades de espalhamento para diversos aditivos utilizados para soluções protéicas.

Tabela 3 – Comprimentos de espalhamento para os aditivos* mais comuns.

Aditivo	$\rho_s [10^{11} \text{ cm}^{-2}]$
Acetato	1.3038E+11
NaCl	1.7642E+11
KCN	1.2928E+11
Acido Acético	2.2674E+11
HCl	1.7865E+11
TRIS	6.1121E+11

*Fórmulas químicas obtidas no *Handbook of Chemistry and Physics* (Lide, 1997)

Desta forma, uma solução tampão com TRIS 50mM diluída em água possui densidade de espalhamento $\rho_s = 9.48 \times 10^{10} \text{ cm}^{-2}$ (à 20°C) e, o contraste de densidade de espalhamento, para uma proteína com volume parcial específico $\bar{v} = 0.72 [\text{cm}^3 / \text{g}]$ seria $\Delta\rho = 3.13 \times 10^{10} \text{ cm}^{-2}$.

Conhecendo-se o contraste de densidade de espalhamento $\Delta\rho$ e a seção de choque macroscópica da partícula $d\Sigma/d\Omega(q)$, obtemos o peso molecular de uma partícula com volume parcial específico $\bar{v}$ usando a expressão,

$$M = \frac{d\Sigma(0)}{d\Omega} \left(\frac{N_A}{c\Delta\rho^2} \right) = M_1 N_A \quad (110)$$

sendo N_A o número de Avogadro (6.022×10^{23} moléculas/Mol). Da equação (102) obtemos o volume da partícula em solução fazendo $V_p = \bar{v} M_1$. Finalmente, a seção de choque diferencial por partícula $I_1(q)$ ($I_1(q) = n^{-1} d\Sigma/d\Omega(q) = d\sigma/d\Omega(q)$), será dada por,

$$I_1(q) = \frac{d\Sigma(q)}{d\Omega} \frac{1}{c^2} \frac{d\Sigma(0)}{d\Omega} \frac{1}{(\Delta\rho)^2 \bar{v}^2} \quad (111)$$

1.4. Distorções da intensidade espalhada

Em todo o desenvolvimento apresentado anteriormente, seja na obtenção dos parâmetros estruturais, ou na renormalização dos dados para escala absoluta, assumimos colimação pontual (foco ponto). Para aquilo que é definido como colimação pontual do feixe incidente e espalhado de Raios X em medidas de baixos ângulos, temos necessariamente que satisfazer as condições (Feigin, 1967):

$$\sqrt{\sigma} / R \ll \theta_{\min} \text{ e } \sqrt{\sigma_r} / R \ll \theta_{\min} \quad (112)$$

sendo σ e σ_r respectivamente as áreas de seção transversais do feixe incidente e do detector da radiação, R a distância amostra-detetor e θ o ângulo de espalhamento. Com isso teremos condições muito próximas do que é assumido na teoria, uma vez que o feixe incidente é praticamente paralelo e a energia espalhada é medida sobre um ângulo sólido muito pequeno.

Na realidade, na linha de SAXS do LNLS (bem como em muitos laboratórios síncrotrons) consegue-se um feixe muito intenso e de pequenas dimensões que satisfaz bem esta condição na amostra. Já no detetor como veremos, a utilização de fendas largas (visando aumentar o fluxo captado pelo detetor) pode ocasionar distorções à intensidade espalhada.

1.4.1. Influência das dimensões do feixe e detetor

Na maioria dos instrumentos de baixos ângulos, utiliza-se colimação de fendas para aumentar-se a intensidade primária e espalhada. Além disso, os detetores da radiação espalhada também possuem dimensões finitas. Isto causa distorções (*smearing*) na curva em certos intervalos angulares definidos pelos parâmetros geométricos do sistema de colimação, e pelo tamanho e sensibilidade do detetor. Como regra, estes efeitos de *smearing* resultam em sérias distorções do perfil real da curva experimental e, conseqüentemente, levam a necessidade de restaurarmos a curva espalhada à correspondente condição ideal (“ponto”) do experimento. Os métodos para a restauração da “curva ideal” são muito importantes, uma vez que sempre teremos efeitos experimentais ocorrendo.

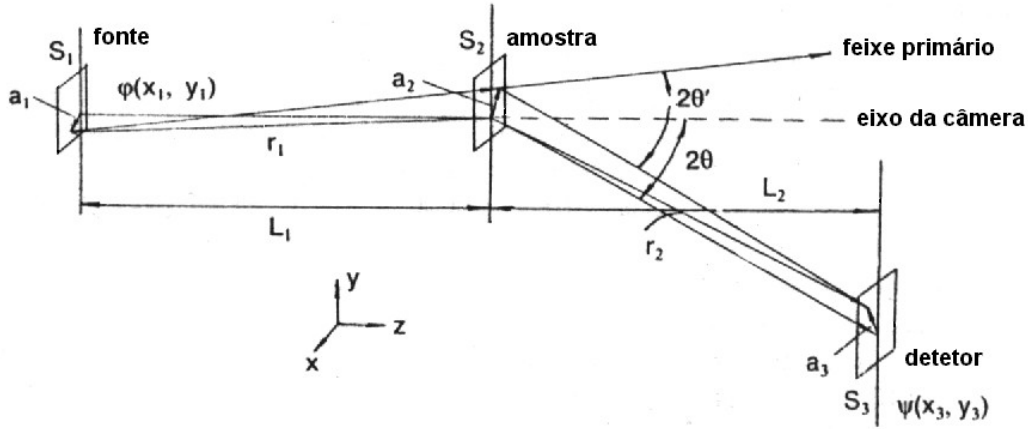


Figura 23 – Efeitos de “smearing” causados por colimação de fendas (esquema). Coordenadas dos vetores $a_1(x_1, y_1, -L_1)$, $a_1(x_2, y_3, 0)$, $a_1[x_3 + L_2 \tan(2\theta), y_3, L_2 \cos(2\theta)]$. Figura tirada de Semenyuk e Svergun, 1987.

Serão apresentadas, em linhas gerais, as causas das distorções e posteriormente a relação geral entre as curvas ideal e distorcida. A Figura 23 mostra um esquema de um experimento de difração no qual a fonte, amostra e detector são considerados objetos de dimensões finitas. O sistema de leitura está em uma posição para a qual o ângulo para colimação pontual seria igual a 2θ . Consideraremos um raio arbitrário detectado, o qual não passa pelo centro da fonte, amostra e detector. Se seu caminho é descrito pelos vetores $\mathbf{r}_1$, $\mathbf{r}_2$, e $\mathbf{a}_2$, então o ângulo de espalhamento é dado por (Feigin e Svergun, 1987)

$$2\theta' = \text{ArcSin} \left\{ \frac{\|(\vec{r}_1 - \vec{a}_2)(\vec{r}_2 - \vec{a}_2)\|}{\|\vec{r}_1 - \vec{a}_2\| \|\vec{r}_2 - \vec{a}_2\|} \right\} \quad (113)$$

As coordenadas dos vetores $\mathbf{r}_1$, $\mathbf{r}_2$, e $\mathbf{a}_2$ são mostrados na Figura 37 (a origem do sistema de coordenadas é o centro da amostra). Para espalhamento a baixos ângulos, as distâncias entre fonte, amostra e detector geralmente excedem as dimensões dos mesmos e o ângulo de espalhamento é pequeno o suficiente para assumirmos que $\tan(2\theta) \approx 2\theta$ e $\sin(2\theta) \approx 2\theta$. Neste caso temos,

$$2\theta' \approx [y'^2 + (2\theta - x')^2]^{1/2}$$

onde

$$y' = y_3 / L_2 + y_1 / L_1 - y_2 (L_1^{-1} + L_2^{-1})$$

e

$$x' = x_3 / L_2 + x_1 / L_1 - x_2 (L_1^{-1} + L_2^{-1})$$

Em um ângulo 2θ , o detetor irá ler também raios para os quais os vetores $\mathbf{a}_1$, $\mathbf{a}_2$, e $\mathbf{a}_3$ caem respectivamente dentro dos limites da fonte, amostra e detetor. A intensidade de espalhamento total será dada pela integração de todas as possíveis direções destes raios. Também deve-se levar em conta que em vários pontos da fonte e detetor a intensidade da radiação e a eficiência da leitura podem ser diferentes. Em termos da distribuição de intensidades do feixe incidente $\varphi(x_1, y_1)$ e da eficiência do detetor $\psi(x_3, y_3)$, a expressão geral para a intensidade lida é expressa na forma,

$$\tilde{I}(2\theta) = A \int_{S_1} \int_{S_2} \int_{S_3} \varphi(x_1, y_1) \psi(x_3, y_3) I \left[\left(y'^2 + (2\theta - x')^2 \right)^{1/2} \right] dx_1 dx_2 dx_3 dy_1 dy_2 dy_3 \quad (114)$$

sendo S_1 , S_2 , e S_3 as áreas correspondentes da fonte, amostra e detetor, e A uma constante de normalização.

Podemos reescrever a equação (114) de modo mais conveniente. Se assumirmos que $\varphi(x_1, y_1) = \varphi_x(x_1) \varphi_y(x_1)$ e $\psi(x_3, y_3) = \psi_x(x_3) \psi_y(x_3)$ (uma condição geralmente aceitável e que, na prática, é satisfeita com grau de precisão suficiente para todas as instalações experimentais) e introduzimos as variáveis q , $t = 4\pi y' / \lambda$ e $u = 4\pi x' / \lambda$ teremos

$$\tilde{I}(q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_w(u) W_l(t) I \left[\left(t^2 + (s - u)^2 \right)^{1/2} \right] dt du \quad (115)$$

Aqui as funções peso $W_w(u)$ e $W_l(t)$ são determinadas pela geometria instrumental bem como pelas funções φ e ψ , e obedecem as condições de normalização,

$$\int_{-\infty}^{\infty} W_w(u) du = \int_{-\infty}^{\infty} W_l(t) dt = 1 \quad (116)$$

A integração sobre t é feita na direção y (distorção por altura de fenda - “*slit-height smearing*”) e sobre u na direção x (distorção por largura de fenda - “*slit-width smearing*”). Apesar de em princípio, podermos medir experimentalmente as formas das funções peso, é mais conveniente as calcularmos analiticamente. Na realidade, existem vários trabalhos (Guinier e Fournet, 1955; Kratky e Glatter, 1982; Feigin e Svergun, 1987) nos quais os autores desenvolvem métodos gerais para o cálculo destas funções, para vários tipos de geometrias de sistemas de colimação. Por exemplo, para câmeras com fendas retangulares pode-se aplicar a equação simplificada,

$$W(x) = \begin{cases} 1/a, & |x| \leq (a-l)/2 \\ [(a+l)/2 - |x|]/al, & (a-l)/2 \leq |x| \leq (a+l)/2 \\ 0, & |x| > (a+l)/2 \end{cases} \quad (117)$$

para ambas $W_w(u)$ e $W_l(t)$. Na realidade a expressão anterior é uma função peso com perfil trapezoidal, possuindo bases $(a+l)$ e $(a-l)$. Aqui assumimos uma fonte pontual e uma sensibilidade uniforme do detetor. A quantidade a é a projeção da largura do feixe (ou altura) no plano de detecção e l é a largura (ou altura) da fenda do detetor.

1.4.2. Policromaticidade do feixe

Até aqui assumimos uma radiação incidente estritamente monocromática. Apesar de em anéis síncrotron termos uma boa resolução em energia (e em consequência uma polidispersidade muito pequena no valor de comprimento de onda) esse efeito pode ser levado em conta. Em um caso geral podemos ter em adição ao comprimento de onda principal λ_0 o feixe incidente contém outro comprimento de onda λ_1 . Assim, em algum ângulo de medição 2θ , o espalhamento com ambos $q=4\pi\theta/\lambda_0$ e $q_1=4\pi\theta/\lambda_1$ serão lidos. Neste caso,

$$\tilde{I}(q) = I(q) + \gamma I(q/\lambda) \quad (118)$$

sendo $\lambda=\lambda_1/\lambda_0$, e γ a intensidade relativa da radiação com comprimento de onda λ_1 .

Já se o feixe incidente contém um conjunto de comprimentos de onda caracterizados por uma função espectral $W_\lambda(\lambda)$, e assumindo também que a amostra não possui espalhamento anômalo, teremos para a radiação espalhada

$$\tilde{I}(q) = \int_0^\infty W_\lambda(\lambda) I(q/\lambda) d\lambda \quad (119)$$

onde a função peso é normalizada pela condição

$$\int_0^\infty W_\lambda(\lambda) d\lambda = 1 \quad (120)$$

Com isso, a policromaticidade da radiação, similar ao caso de sistemas com colimação finita, leva a distorções no perfil de espalhamento experimental.

1.4.3. Erros Estatísticos

A intensidade espalhada é medida para alguns valores discretos do ângulo de espalhamento. No entanto, cada resultado experimental não é uma função contínua $I(q)$, mas sim um conjunto de valores discretos $I(q_i)$ correspondendo a valores definidos q_i (erros nos valores de q_i são geralmente desprezíveis). Com isso, os valores de $\tilde{I}_i$ variam da intensidade $I(q)$, de modo geral, por números aleatórios, isto é,

$$\tilde{I}_i = \tilde{I}(q_i) = I(q_i) + \varepsilon_i \quad (121)$$

onde ε_i é o erro estatístico no ponto q_i . Estes erros são números aleatórios independentes e suas magnitudes dependem da precisão do experimento. Para checar a concordância entre experimentos sucessivos e para restaurar o perfil de espalhamento correto (fazendo-se a média de várias curvas experimentais, por exemplo) é necessário utilizar-se métodos estatísticos (cálculo de *chi-square*, etc).

1.4.4. Expressão geral para a Intensidade Experimental

As fontes de distorção consideradas anteriormente podem ser combinadas e generalizadas na forma,

$$J(q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_w(u) W_l(t) W_\lambda(\lambda) I \left[t^2 + (q-u)^2 \right]^{1/2} \lambda^{-1} d\lambda dt du \quad (122)$$

e

$$J_e(q_i) = \mu_1 [J(q_i) + \varepsilon_i] + \mu_2 [J_a(q_i) + \varepsilon_{ai}] \quad (123)$$

onde $J_e(q_i)$ é a intensidade de espalhamento experimental da amostra, $J_a(q_i)$ é a componente aditiva (*buffer*) distorcida pelos mesmos efeitos integrais, μ_1 e μ_2 as constantes de normalização para absorção, intensidade primária, etc enquanto que ε_i e ε_{ai} são os erros medidos.

Dentro destas relações, podemos formular os principais estágios para o processamento de dados de espalhamento a baixos ângulos:

- 1 Processamento preliminar das curvas (normalização);
- 2 Subtração da componente aditiva (*buffer*). Caso haja várias curvas para promediar, verificar a concordância entre as mesmas;
- 3 Correção por efeitos de colimação e policromaticidade;

Métodos computacionais para a resolução da equação geral (122) interpretam as distorções independentemente. Neste cálculo separa-se esta equação em três equações individuais e as correções devem ser introduzidas na seguinte ordem (Glatter e Kratky, 1982):

- 1 Correção dos efeitos de colimação para largura, ou seja, a solução da equação de convolução com respeito a $F(q)$,

$$J(q) = \int_{-\infty}^{\infty} W_w(u) F(q-u) du \quad (124)$$

- 2 Correção dos efeitos de colimação para altura, ou seja, a solução da equação de convolução com respeito a $I_n(q)$,

$$F(q) = \int_{-\infty}^{\infty} W_l(t) I_n(\sqrt{q^2 + t^2}) dt \quad (125)$$

- 3 Correções para policromaticidade do feixe, ou seja, solução para a equação com respeito a $I(s)$,

$$I_n(q) = \int_{-\infty}^{\infty} W_\lambda(\lambda) I(q/\lambda) d\lambda \quad (126)$$

1.4.5. Efeitos de colimação na linha de SAXS do LNLS

Anteriormente foram descritos todos os efeitos que causam distorções na intensidade espalhada, e que devem ser corrigidos para podermos extrair informações dos dados experimentais. No Laboratório Nacional de Luz Síncrotron, como utilizamos um monocromador, efeitos de policromaticidade não precisam ser levados em conta uma vez que temos uma ótima resolução em energia ($\Delta\lambda/\lambda < 1\%$ para o intervalo de comprimentos de onda utilizados na linha de SAXS). Além disso, os harmônicos de baixa energia são completamente absorvidos pelo material espalhador, ao passo que os de alta energia são ordens de grandeza menos intensos.

Por outro lado, como as dimensões do feixe incidente na amostra são muito pequenas (tipicamente 1x1mm) a condição proposta por Feigin (equação 112) é muito bem satisfeita para σ . Já para σ_r dependerá do detetor utilizado. Aqui descreverei apenas o caso onde utilizamos detetor linear sensível a posição (PSD). De modo muito simplificado, este detetor consiste essencialmente de uma câmara metálica (cátodo) contendo gás, onde está presente um fio condutor (ânodo). Os fótons incidentes produzem pares elétron-íon no gás. Os elétron-íons provocam uma avalanche de cargas gerando uma distribuição localizada de cargas negativas próximo ao ânodo. Esta distribuição de cargas induzirá cargas positivas no cátodo, produzindo um pulso elétrico no mesmo. Este sinal induz dois pulsos em um conjunto RLC que permite tanto a determinação da posição do mesmo ao longo do fio, quando sua intensidade. Dentre as características deste tipo de detetor, a que nos interessa é sua fenda de entrada e sua resolução. A fenda de entrada é colocada na frente do detetor selecionando a fatia do feixe espalhado que será coletado pelo mesmo. Já a resolução é

definida como a menor distância entre picos de um espectro que pode ser distinguido pelo sistema de detecção. Para esta classe de detetores (Gabriel, 1982), consegue-se resoluções da ordem de $300\mu\text{m}$. Sendo assim, podemos imaginar que cada canal do detetor é composto por uma fenda “imaginária” com largura igual à resolução do detetor e altura igual à fenda de entrada do mesmo. Como a largura desta fenda imaginária é pequena, podemos aproximá-la por uma função delta e assim na equação (122) deve-se levar em conta apenas o efeito da altura de fenda. As fendas de entrada disponíveis variam de 1 à 8mm, ao passo que as distâncias amostra detetor utilizadas variam de 300 a 2000mm. Neste contexto apresentarei agora um experimento hipotético de um sistema monodisperso de esferas em solução para mostrar quantitativamente a deformação da curva experimental causada pelo efeito de *smearing*.

O padrão de espalhamento (teórico) bidimensional para um sistema monodisperso de esferas é mostrado na Figura 24.

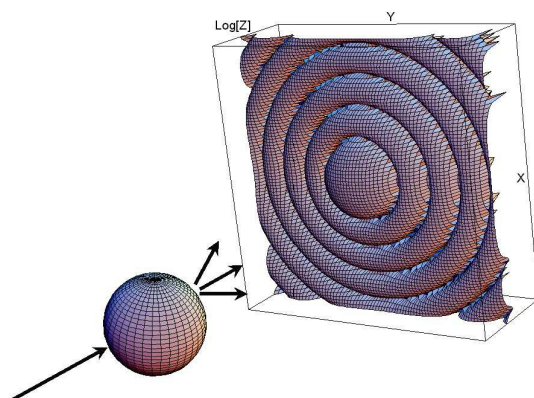


Figura 24 – Espectro bidimensional teórico para o espalhamento de uma esfera.

Como discutimos anteriormente na introdução teórica, o espectro de espalhamento em sistemas randomicamente orientados é isotrópico. Sendo assim, tomando-se uma fatia linear deste espectro, teremos a informação desejada do espalhamento do sistema. No entanto, como a fenda de entrada possui uma dimensão finita, na realidade não estaremos tomando uma fatia, mas sim integrando uma região, que nada mais é do que a largura da fenda de entrada do detetor. Como o espalhamento é isotrópico, a integração deveria ser *angular*. A integração na largura da fenda promoverá uma integração *linear* ao longo da fenda, e isto gerará distorções no espectro espalhado, como mostramos na simulação abaixo,

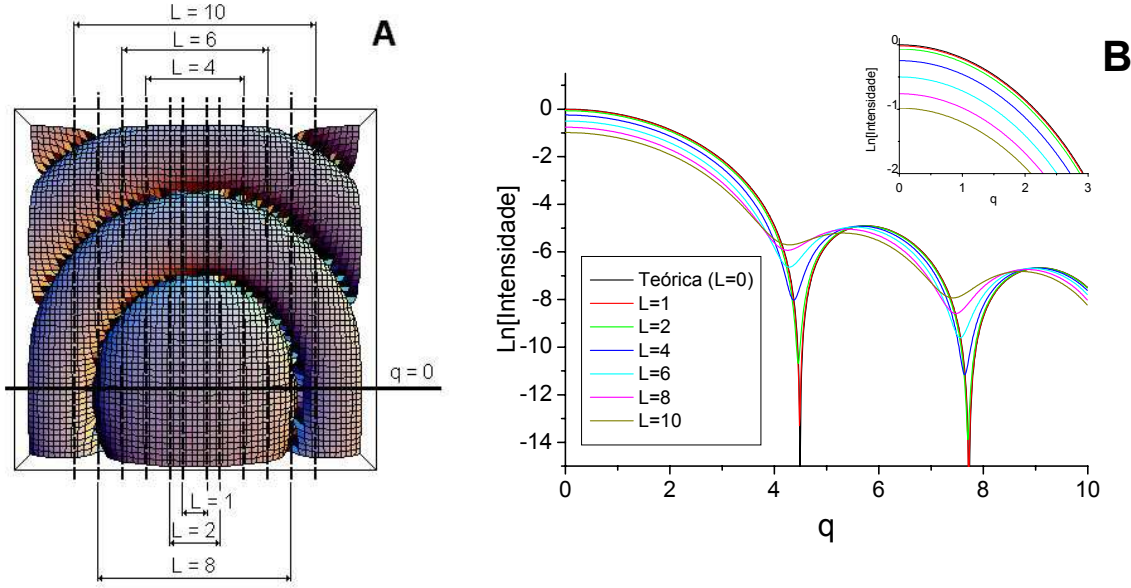


Figura 25 – Simulação da distorção causada pela integração linear ao longo da fenda. As unidades de “L” e “q”, são as mesmas. A) Representação da região de integração no espectro de espalhamento. B) Intensidade resultante da integração. Inserção: Região de Guinier. A simulação foi feita utilizando o programa Mathematica®.

Na Figura 25 anterior mostramos um caso extremo, onde a integração foi feita em uma região tão larga que distorce completamente a curva de espalhamento. No nosso caso não teremos distorções tão severas, pois a faixa a ser integrada geralmente será pequena. No entanto, como visto na Figura 25B, para todos os casos teremos a diminuição da intensidade na origem, o que é problemático, pois este valor é o responsável pela renormalização à escala absoluta e estimativa do peso molecular. Sendo um pouco mais específico, tomemos a configuração experimental da linha de SAXS do LNLS. Para a simulação anterior tomei a largura da fenda na mesma unidade do vetor de espalhamento q . Isto é necessário, pois a integração deverá ser feita no espaço recíproco. Para convertermos a largura da fenda para as unidades do vetor de espalhamento utilizamos a expressão do vetor q (Figura 3) na forma,

$$L = 2 \frac{4\pi}{\lambda} \sin\left(\frac{1}{2} \arctan\left(\frac{f}{2D}\right)\right) \approx \frac{2\pi f}{\lambda D} \quad (127)$$

sendo f a largura da fenda, D a distância amostra detector (f e D devem ter as mesmas unidades) e λ o comprimento de onda (assumirei que é dado em Å). Na última passagem usamos o fato de que para ângulos pequenos $\tan(\theta/2) \approx \sin(\theta/2) \approx \theta/2$. Com esta

relação vemos que a largura de integração na fenda variará dentro do intervalo $0.002 \text{ \AA}^{-1} < L < 0.1 \text{ \AA}^{-1}$. Assim é fácil perceber que a distorção será mais severa para distâncias pequenas. No entanto existe outro parâmetro a considerar: o tamanho das partículas espalhadoras. Partículas grandes ($300 \text{ \AA}$ por exemplo) fornecem um espectro de espalhamento restrito à ângulos mais baixos, ao passo que partículas pequenas ($30 \text{ \AA}$ por exemplo) fornecem um espectro de espalhamento que se estende por uma região muito maior (vide Figura 20). Assim, como não é possível mostrar uma expressão geral para o efeito de distorção, mostrarei uma simulação onde avalio a variação no valor de $I(0)$ integrando na região de L mostrada anteriormente, para 3 tamanhos diferentes de partícula: 10, 100 e $1000 \text{ \AA}$, e comprimento de onda de $1.608 \text{ \AA}$:

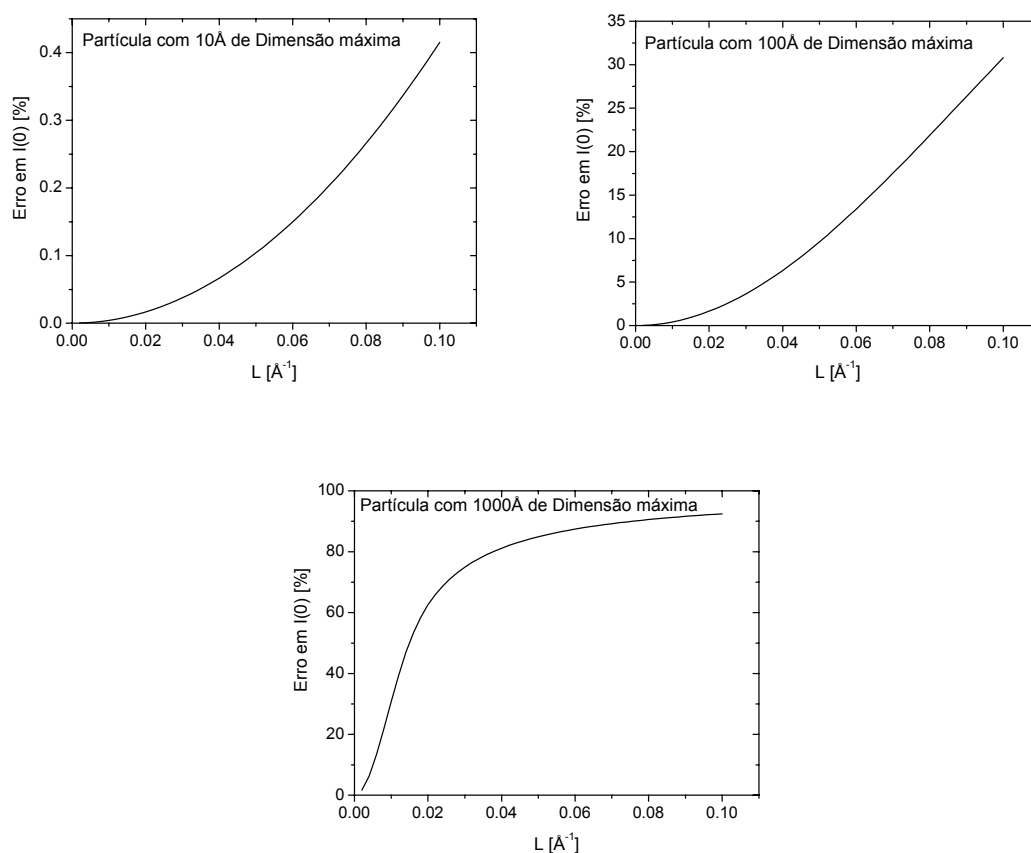


Figura 26 – Efeito de “smearing” causado pela integração na largura da fenda. Na figura mostramos cálculos para três partículas com tamanhos de 10, 100 e $1000 \text{ \AA}$.

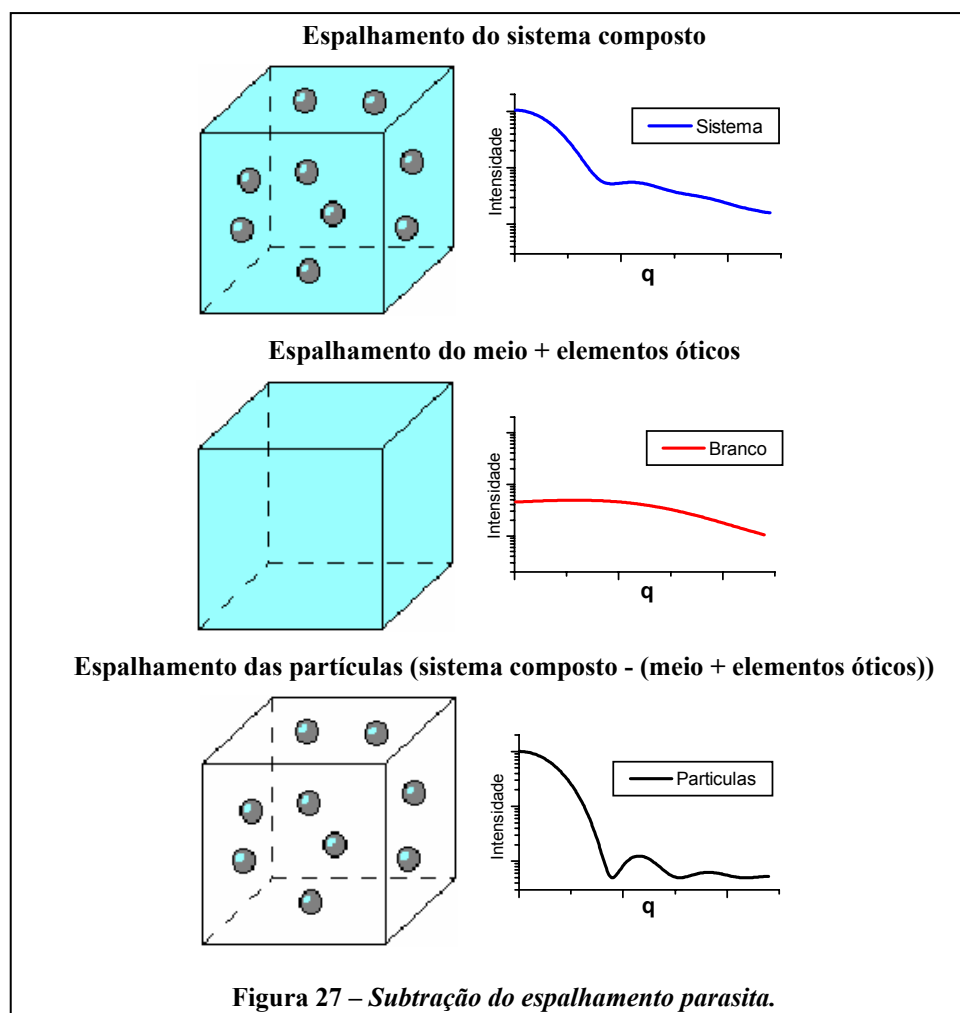
O cálculo anterior foi feito para espalhamento por esferas, mas se aplica a partículas globulares em geral. Vemos que para partículas pequenas, o erro é pequeno, mesmo para grandes aberturas de fenda. Já para partículas maiores, vemos que o erro aumenta consideravelmente.

É por isso que em experimentos de SAXS para partículas em solução, deve-se tomar medidas a várias distâncias amostra-detector. À longa distância, onde os efeitos de *smearing* são diminuídos, podemos determinar com maior precisão $I(0)$ experimental. Medidas à distâncias menores permitem uma melhor descrição do espalhamento a altos ângulos e, ao final, as curvas são unidas de modo a ter-se um grande intervalo experimental dos dados de espalhamento.

Procurei mostrar os efeitos que podem acontecer devido à colimação por fendas realizada na linha de SAXS do LNLS. Como vimos, de modo geral, para experimentos onde não uniremos curvas em distâncias diferentes e onde o arranjo experimental foi preparado adequadamente (de acordo com o tamanho da partícula estudada) o erro cometido é pequeno, ficando em torno de 10% nos piores casos. Este erro não afetará de modo proibitivo a determinação de parâmetros como o raio de giro, máxima dimensão etc, mas como afeta diretamente $I(0)$, qualquer utilização deste fator será prejudicada. Sendo assim, para aumentar a precisão e confiabilidade da renormalização para escala absoluta, devemos corrigir os efeitos de *smearing*. Existem vários programas que fazem a correção destes efeitos, por exemplo, os programas GNOM (Semenyuk e Svergun, 1991), ITP (Glatter, 1977), utilizando os métodos citados anteriormente. Experimentos com *detetores bidimensionais* em anéis síncrotron não requerem, em geral, nenhum tipo de correção de *smearing* pelo fato de não utilizar fendas de entrada.

1.5. Tratamento de Dados

Em medidas de espalhamento a baixos ângulos, sempre teremos o sistema de interesse (partículas, fases etc) imerso em um meio que na maioria das vezes também contribuirá para o espalhamento. Por exemplo, se estivermos interessados em espalhamento por partículas imersas em uma matriz, deveríamos levar em conta o espalhamento desta matriz bem como dos elementos óticos (fendas, etc).



No caso de partículas em solução, por exemplo (Figura 27), assumindo que a curva composta é uma superposição do espalhamento do solvente + partículas, devemos subtrair do espalhamento do sistema “partículas + meio + elementos óticos” (solução de partículas) a contribuição “meio + elementos óticos” (solução tampão). Isso é conseguido realizando um experimento “sem amostra” subtraindo esta curva de espalhamento da curva do experimento “com amostra”. Para sistemas em solução precisamos subtrair o espalhamento do solvente, ao passo que para amostras sólidas, em geral, subtraímos o espalhamento com o porta-amostras vazio. Sendo assim, experimentos de SAXS e o tratamento dos dados são compostos pelas seguintes etapas:

- 1 Medida do espalhamento do sistema composto (sistema de interesse + meio + elementos óticos);
- 2 Medida do espalhamento do "parasita" (meio + elementos óticos);

- 3 Normalização dos dados por efeitos experimentais (atenuação da amostra, intensidade integrada, eficiência do detetor etc.);
- 4 Obtemos a curva experimental desejada subtraindo-se a intensidade parasita da intensidade de espalhamento do sistema composto.

A correção por atenuação da amostra à radiação incidente é necessária, pois as amostras terão absorções distintas uma vez que os componentes serão diferentes em cada caso. A correção pela intensidade integrada normaliza as curvas quanto ao tempo de exposição tanto como a possíveis variações de intensidade da fonte de radiação (no caso de anéis síncrotron, por exemplo, a intensidade do feixe incidente diminui ao longo do tempo). Detetores lineares (ou de área) em geral não possuem a mesma eficiência de detecção em toda sua extensão. Assim normaliza-se a curva de espalhamento por uma curva de calibração do detetor que corrige estas variações de eficiência. A curva obtida será dada na maioria dos casos em *Intensidade x canais*, e desta maneira, é necessário calcular o vetor de espalhamento q ($q=4\pi \cdot \text{Sen}(\theta)/\lambda$, 2θ ângulo de espalhamento e λ o comprimento de onda da radiação usada) para que tenhamos a curva em Intensidade x q . Um exemplo de tratamento de dados é mostrado na Figura 28.

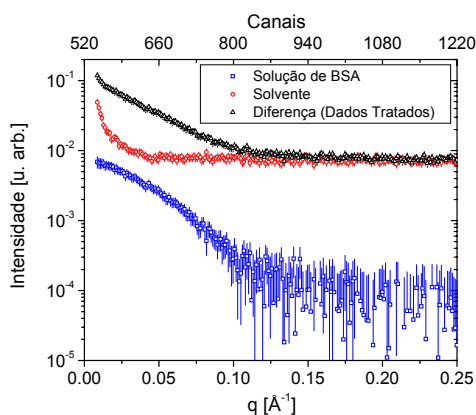


Figura 28 – Curva típica de espalhamento de uma solução de BSA, espalhamento do solvente e a curva de diferença. O solvente utilizado foi 50mM Tris-HCl em água mili Q, pH7.5. Os dados foram coletados na linha de SAXS do LNLS.

Se denominamos $I_{amostra}$ a intensidade do sistema composto, $I_{parasita}$ a intensidade do meio onde as partículas de interesse estão imersas, a intensidade corrigida $I_{corrigida}$ será:

$$I_{corrigida} = I_{Namostra} - I_{Nparasita} \quad (128)$$

com,

$$I_{Namostra} = \frac{I_{amostra} \cdot A_{Tamostra}}{Ef \cdot NIO_{amostra}} \quad (129)$$

$$I_{Nparasita} = \frac{I_{parasita} \cdot A_{Tparasita}}{Ef \cdot NIO_{parasita}} \quad (130)$$

sendo,

$I_{amostra}$ - a intensidade de espalhamento do sistema composto;

$I_{parasita}$ - a intensidade de espalhamento parasita (meio + elementos óticos)

$A_{Tamostra}$ - atenuação do feixe de Raios X pelo sistema composto;

$A_{Tparasita}$ - atenuação do feixe de Raios X pelo parasita;

$NIO_{amostra}$ - Intensidade integrada do sistema composto;

$NIO_{parasita}$ - Intensidade integrada do parasita;

Ef - Eficiência normalizada do detetor em cada canal.

$I_{Namostra}$ - intensidade normalizada do sistema composto;

$I_{Nparasita}$ - intensidade normalizada do parasita;

O cálculo do erro estatístico é feito utilizando-se a seguinte expressão (Knoll, 1989):

$$\sigma_{corr} = \sqrt{\sigma_{Namostra}^2 + \sigma_{Nparasita}^2} \quad (131)$$

com

$$\sigma_{Namostra} = I_{Namostra} \cdot \sqrt{\frac{1}{I_{amostra}} + \frac{1}{NIO_{amostra}} + \left(\frac{STD[Ef]}{Ef} \right)^2} \quad (132)$$

e

$$\sigma_{Nparasita} = I_{Nparasita} \cdot \sqrt{\frac{1}{I_{parasita}} + \frac{1}{NIO_{parasita}} + \left(\frac{STD[Ef]}{Ef} \right)^2} \quad (133)$$

sendo,

$\sigma_{Namostra}$ - Incerteza calculada para $I_{Namostra}$;

$\sigma_{Nparasita}$ - Incerteza calculada para $I_{Nparasita}$;

$STD[Ef]$ - Desvio padrão da eficiência normalizada do detetor na região útil;

O vetor de espalhamento correspondente a cada ponto experimental “j” é calculado fazendo-se,

$$q[j] = \frac{4\pi \cdot \sin(\theta[j]/2)}{\lambda} \quad (134)$$

com,

$$(\theta[j]) = \text{ArcTan} \left[\frac{(\text{Canal}[j] - \text{CanalZero}) \cdot K}{D} \right] \quad (135)$$

sendo,

θ - ângulo de espalhamento;

λ - comprimento de onda;

Canal[j] - Canal correspondente no detetor linear;

CanalZero - Canal de incidência do feixe direto;

K - Relação mm/canal (a distância em milímetros entre dois canais do detetor);

D - Distância amostra - Detetor

A curva obtida ainda poderá ser corrigida por efeitos de *smearing*, bem como ser renormalizada para escala absoluta, como discutido anteriormente. No entanto o tratamento preliminar dos dados estará concluído realizando-se os passos citados.

As expressões mostradas anteriormente são válidas para sistemas *diluídos*, isto é, casos onde a fração de volume das partículas espalhadoras é muito pequena. Para proteínas em solução a fração de volume para experimentos típicos (concentrações da ordem de 10mg/ml) é menor do que 1% e desta forma não é preciso levar-se em conta estes efeitos. No entanto para experimentos com amostras com frações de volume maiores requerem a consideração correta deste efeito. A expressão da intensidade corrigida para sistemas concentrados, com fração de volume ϕ é dada por (Glatter e Kratky, 1982),

$$I_{\text{corrigida}} = (I_{\text{Namostra}} - I_{\text{Ncap}}) - (1 - \phi)(I_{\text{NParasita}} - I_{\text{Ncap}}) \quad (136)$$

sendo I_{Namostra} e $I_{\text{NParasita}}$ como mencionados anteriormente e I_{Ncap} a intensidade normalizada de espalhamento obtida com o porta-amostras **vazio**. Para $\phi \sim 0$ recuperamos a expressão anteriormente mostrada.

1.5.1. Programas Para Tratamento de Dados

Tanto para detetores lineares quanto para detetores de área (bidimensionais) o procedimento descrito na seção anterior é o mesmo. Como já mencionado anteriormente, para sistemas com partículas randomicamente orientadas a intensidade de espalhamento é isotrópica de modo que basta tomarmos um corte desta intensidade de espalhamento. Por isso em muitos casos podem ser utilizados detetores unidimensionais. A vantagem destes detetores é que como resultado teremos uma curva unidimensional pronta para o tratamento posterior de dados. Além disso, os arquivos gerados são pequenos (~50kB em arquivos ASCII) o que facilita seu transporte e arquivamento. Por outro lado, quando tomamos apenas um corte do espalhamento deixamos de utilizar a maior parte da intensidade espalhada. Para sistemas que espalham bastante isso não é problema, pois mesmo tomando-se um corte ainda teremos uma boa estatística para a curva de espalhamento. Já para sistemas que espalham pouco (proteínas em solução, por exemplo, espalham $\sim 10^{-5}$ da intensidade incidente) tomar apenas um corte da intensidade espalhada acaba por limitar em muito as possibilidades do experimento, uma vez que a baixa estatística a altos ângulos (ocasionada pela queda quase exponencial da intensidade) impede que possamos ter informações a altos ângulos utilizáveis. Isso pode ser corrigido utilizando detetores bidimensionais que captem todo o ângulo sólido de espalhamento. Nestes espectros bidimensionais pode-se fazer uma integração angular, utilizando assim toda a intensidade espalhada, e assim melhorando em muito a estatística dos dados obtidos. No entanto para efetuar esta integração deve-se ter uma rotina capaz de integrar os dados e, além disso, os arquivos gerados são bem maiores do que no caso unidimensional (~16MB para arquivos ASCII e ~4MB para arquivos Binários). Com isso, deve-se avaliar pelas características do experimento, qual é o melhor detetor a utilizar. Outro ponto que deve ser ressaltado é o fato de em experimentos típicos de SAXS tomar-se dezenas (ou em alguns casos, centenas) de curvas. Desta forma o tratamento dos dados pode ser um processo demorado se for feito manualmente.

Durante meu Doutorado desenvolvi dois programas para tratamento de dados: tanto para dados unidimensionais, programa TRAT1D (Oliveira, 2003) e bidimensionais, programa TRAT2D (Oliveira, 2004). Os programas foram desenvolvidos para WindowsTM (95, 98, 2000, XP etc.) em linguagem DELPHITM (orientada ao objeto) com

ênfase na interface gráfica (janelas, menus etc) de modo a facilitar sua utilização. Com um pequeno número de passos o usuário consegue obter curvas experimentais tratadas e prontas para posteriores análises. Além disso, centenas de curvas (ou espectros bidimensionais) podem ser tratados simultaneamente, com o correto cálculo dos erros experimentais. O programa TRAT1D possui funções de regressão linear, corte de pontos, renormalização das curvas experimentais bem como rotinas de promediação dos dados experimentais, o que é muito útil quando se toma várias curvas da mesma amostra. Já no programa TRAT2D ferramentas de integração angular foram desenvolvidas, fornecendo como saída curvas unidimensionais que podem ser lidas pelo programa TRAT1D. Estes programas estão disponíveis para os usuários da linha de SAXS do LNLS, e vem sendo utilizados para o tratamento dos dados dos mais variados tipos de sistemas.

2. Cálculo de Parâmetros Hidrodinâmicos para Modelos *ab initio*

2.1. Introdução

Na introdução teórica apresentamos uma metodologia para cálculo de propriedades hidrodinâmicas para estruturas cristalográficas de proteínas. Pelo que foi apresentado vimos que para estruturas de alta resolução os resultados hidrodinâmicos eram muito semelhantes aos obtidos experimentalmente. Tentaremos agora aplicar esta metodologia para os modelos de baixa resolução obtidos por cálculos *ab initio* de SAXS.

Das três metodologias de cálculo *ab initio* apresentadas, as modelagens feitas com *bead models* e *chain models* podem ser aplicadas para cálculos hidrodinâmicos uma vez que fornecem como saída dos cálculos, arquivos com coordenadas das posições das subunidades componentes dos modelos. Em princípio, os modelos tipo *chain model* devem ter uma melhor correspondência de valores quanto às proteínas estudadas uma vez que, apesar de não restaurarem peculiaridades estruturais (alfa hélices, folhas beta), conseguem obter informações de forma e composição de domínios da proteína de interesse. Além disso, nesta modelagem assumimos um número fixo de “resíduos”, que é dado pela sequência primária da proteína de interesse. Já os modelos tipo *bead model*, onde podemos ter um número maior ou menor de subunidades na estrutura, dependendo do raio do *bead*

utilizado, são capazes de restaurar corretamente a forma externa geral da proteína estudada, mas não fornecem informações de enovelamento e composição de domínios.

Na modelagem tipo *bead model* já se leva em conta no ajuste dos dados experimentais a camada de hidratação presente na proteína, isto é, o arranjo final das subunidades deve ser imaginado como sendo da proteína com a camada de hidratação. Já nos modelos tipo *chain model* adiciona-se uma rede de átomos de água ao redor do modelo de modo a mimetizar esta camada de hidratação. Neste cálculo cada resíduo é representado por uma esfera e assim podemos esperar que os volumes de modelos deste tipo sejam menores do que a proteína real uma vez que esta aproximação não contempla a estrutura de cada resíduo.

O programa para o cálculo de parâmetros hidrodinâmicos (HIDROPRO) desconsidera átomos de hidrogênio e moléculas de água no cálculo, de modo que a camada de hidratação nos modelos tipo *chain model* não seria considerada no cálculo. Desta maneira podemos tentar, aumentando o raio das subunidades esféricas componentes do modelo, obter um raio ideal para os “*dummy residues*”, que restaure corretamente o volume bem como os parâmetros hidrodinâmicos da proteína estudada.

Esta análise não se aplica aos modelos tipo *dummy atom*. Como dissemos, estes modelos já levam em conta a camada de hidratação e, sendo assim, o raio destas subunidades não pode ser variado uma vez que levaria ou a perda de compacidade do modelo *ab initio* ou à uma superposição excessiva das subunidades componentes do modelo.

Para os cálculos *ab initio* utilizei dados experimentais para a proteína lisozima e albumina, com concentrações de 10mg/ml, diluídas em solução tampão Tris50mM, obtidos na linha de SAXS do LNLS com porta-amostras de capilar de 1.5mm de diâmetro. O cálculo dos modelos foi feito com os programas DAMMIN (*bead models*) e GASBOR (*chain model*), descritos em outra parte desta tese. Os parâmetros hidrodinâmicos foram obtidos com o programa HIDROPRO.

A estratégia para a determinação do raio ideal em modelos *chain model* foi:

- Faço uma série de modelos tipo *chain model* usando o programa GASBOR; (10 modelos);

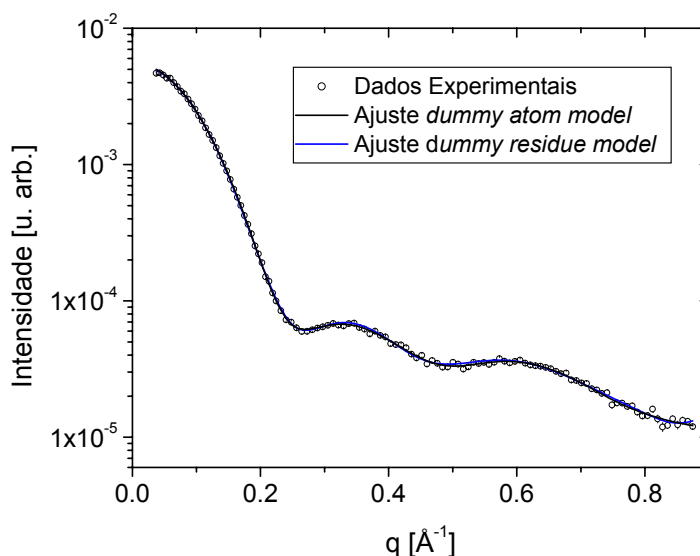
- Para estes modelos, utilizando o programa HYDROPRO, determino os parâmetros hidrodinâmicos, variando o raio dos “*dummy residues*”;
- Tabelaando os resultados vejo se existe um raio privilegiado;

Este procedimento é baseado no realizado pelo professor Garcia de la Torre quando em seu programa *HYDROPRO* (de la Torre et al., 2000) obteve um raio hidrodinâmico médio para os átomos de proteínas (estruturas cristalográficas reais) em solução como sendo de 3.1Å. Como descrito na introdução teórica, este raio para os átomos forneceu corretamente os parâmetros hidrodinâmicos experimentais para as proteínas estudadas naquele artigo.

2.2. Modelos *Ab initio*

Mostro abaixo a superposição dos modelos *ab initio* obtidos pelos métodos mencionados anteriormente com a estrutura cristalográfica da lisozima e albumina. Apenas um modelo é mostrado em cada caso. A semelhança entre os modelos e a estrutura cristalográfica é muito grande, o que indica que os parâmetros hidrodinâmicos devam ser semelhantes também.

- **Lisozima**



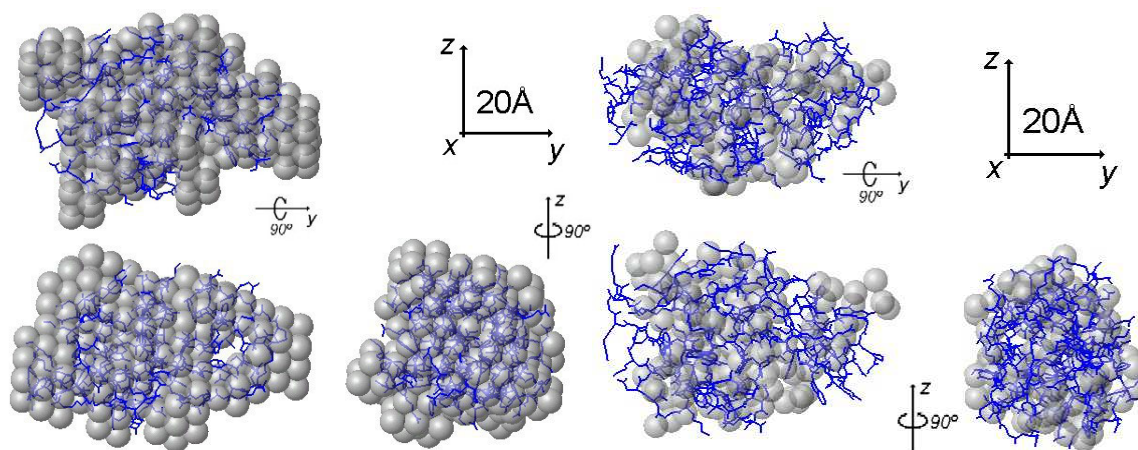
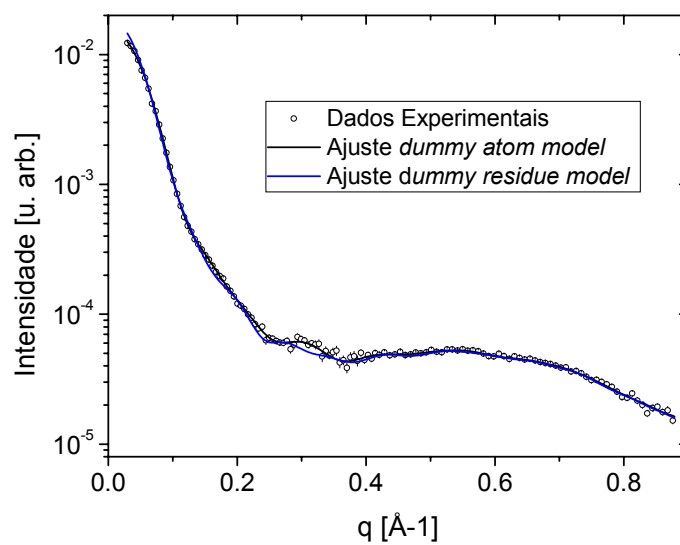


Figura 29 – Modelos *ab initio* para espalhamento da lisozima. *Topo: Ajuste dos dados experimentais pelos modelos. Abaixo Esquerda: Restauração com modelos tipo dummy atom. Abaixo Direita: Restauração com modelos tipo dummy residues. Esqueleto azul: estrutura cristalográfica da lisozima.*

- **Albumina**



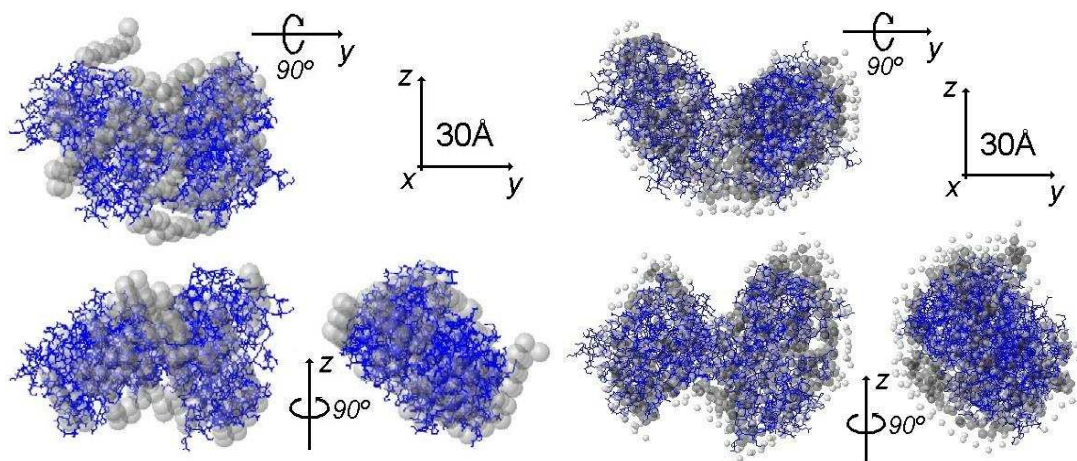


Figura 30 – Modelos *ab initio* para espalhamento da albumina. Topo: Ajuste dos dados experimentais pelos modelos. Abaixo Esquerda: Restauração com modelos tipo *dummy atom*. Abaixo Direita: Restauração com modelos tipo *dummy residues*. Esqueleto azul: estrutura cristalográfica da albumina.

2.3. Estimativa do raio ideal para modelos *chain model*

Tomamos inicialmente uma estrutura *ab initio* tipo *chain model* de lisozima e calculamos os parâmetros hidrodinâmicos, variando o raio dos *dummy chain atoms*. Calculamos estes mesmos parâmetros para a estrutura cristalográfica da lisozima (6LYZ.PDB) e graficamos a razão dos parâmetros hidrodinâmicos em função do raio dos *chain atoms*. O resultado é mostrado abaixo:

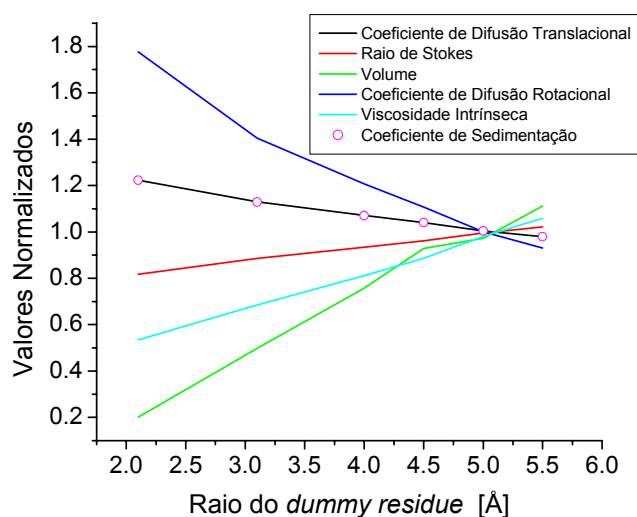


Figura 31 – Estimativa do raio ideal para cálculo de parâmetros hidrodinâmicos.

Temos a clara indicação de que existe um raio privilegiado onde todos os parâmetros hidrodinâmicos do modelo e da estrutura cristalográfica se igualam. Para uma determinação precisa do raio ideal para os *dummy residues*, tomamos os 10 modelos gerados independentemente pelo programa GASBOR e calculamos os parâmetros hidrodinâmicos para uma série de raios variando entre 4.5 e 6Å. As curvas obtidas são similares à mostrada acima. Fazendo a média dos raios ideais fornecidos neste caso obtemos:

$$Raio\ ideal = 5.36 \pm 0.14 \text{ \AA} \quad (137)$$

Sendo assim, este é o raio que devemos utilizar para cálculos hidrodinâmicos de modelos tipo *chain model*, assumindo as hipóteses apresentadas anteriormente.

Para averiguarmos o resultado, recalculamos os parâmetros hidrodinâmicos para um dos modelos da lisozima e também para outra proteína, a Albumina, usando este “raio ideal”, bem como para os modelos tipo *bead model*. O resultado é mostrado na tabela abaixo:

Tabela 4 – Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados pelos modelos tipo *chain model* e *bead model*.

Lisozima	Chain model			Bead model	
	6LYZ	Valores	Difer. [%]	Valores	Difer. [%]
Coef. Dif. Translacional [$10^{-6} \text{ cm}^2/\text{s}$]	1.08	1.09	0.74	1.08	0.19
Raio de Stokes [Å]	19.9	19.8	0.65	19.9	0.10
Raio de Giro [Å]	15.3	15.4	0.92	15.8	3.14
Volume [10^{-20} cm^3]	2.42	2.49	2.86	1.24	48.68
Coef. Dif. Rot. [10^7 s^{-1}]	1.95	1.98	1.38	1.92	1.59
Viscosidade Intrínseca [cm^3/g]	3.49	3.41	2.41	3.39	2.75
Coeficiente de Sedimentação [S]	1.839	1.851	0.65	1.84	0.16

Tabela 5 – Parâmetros hidrodinâmicos para Albumina. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados pelos modelos tipo *chain model* e *bead model*.

Albumina	Chain model			Bead model	
	1AIV	Valores	Difer. [%]	Valores	Difer. [%]
Coef. Dif. Translacional [$10^{-7} \text{ cm}^2/\text{s}$]	5.66	5.69	0.46	6.17	8.90
Raio de Stokes [Å]	37.9	37.7	0.45	34.8	8.18
Raio de Giro [Å]	30.7	30.5	0.65	28.5	7.19
Volume [10^{-19} cm^3]	1.36	1.37	0.88	0.533	60.77
Coef. Dif. Rot. [10^6 s^{-1}]	2.75	2.81	1.92	3.52	27.67
Viscosidade Intrínseca [cm^3/g]	4.533	4.465	1.50	3.43	24.35
Coeficiente de Sedimentação [S]	5.14	5.16	0.45	5.60	8.89

Vemos pelos valores obtidos que os parâmetros hidrodinâmicos são corretamente determinados pelos modelos tipo *chain model* com o raio ideal de 5.36Å.

Como destacado na parte de modelagem *ab initio* desta tese, em muitos casos tem-se uma melhor descrição da partícula espalhadora se fizermos a média espacial dos diferentes modelos *ab initio*, uma vez que essa média fornece a estrutura mais provável do modelo. Porém, como discutido naquele tópico, o modelo “médio” nem sempre ajusta os dados experimentais com a mesma qualidade dos modelos em cada cálculo independente. Sendo assim faremos dois cálculos distintos: dado um conjunto de 10 modelos independentes calculados, tanto para modelos *bead model* e *chain model*, calculamos os parâmetros hidrodinâmicos para cada um destes modelos e obtemos a média destes parâmetros. Posteriormente calculamos a média espacial dos modelos de SAXS e, da estrutura mais provável, obtemos os parâmetros hidrodinâmicos. Os resultados são apresentados a seguir na Tabela 6 :

Tabela 6 – Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados por médias dos modelos tipo bead model.

Lisozima	Média dos modelos			Modelo Médio	
	6LYZ	Valores	Difer. [%]	Valores	Difer. [%]
Coef. Dif. Translacional [$10^{-6}\text{cm}^2/\text{s}$]	1.08	1.08 ± 0.02	0.31	1.104	2.22
Raio de Stokes [Å]	19.9	19.82 ± 0.04	0.41	19.43	2.36
Raio de Giro [Å]	15.3	15.89 ± 0.02	3.88	15.25	0.33
Volume [10^{-20}cm^3]	2.42	1.27 ± 0.09	47.45	1.535	36.57
Coef. Dif. Rot. [10^7s^{-1}]	1.95	1.9 ± 0.1	0.54	2.076	6.46
Viscosidade Intrínseca [cm^3/g]	3.49	3.4 ± 0.2	2.87	3.257	6.68
Coeficiente de Sedimentação [S]	1.839	1.85 ± 0.04	0.36	1.882	2.34

Tabela 7 – Parâmetros hidrodinâmicos para Lisozima. Comparação entre o valor esperado (estrutura cristalográfica) e valores dados por médias dos modelos tipo chain model

Lisozima	Média dos modelos			Modelo Médio	
	6LYZ	Valores	Difer. [%]	Valores	Difer. [%]
Coef. Dif. Translacional [$10^{-6}\text{cm}^2/\text{s}$]	1.08	1.08 ± 0.09	0.44	1.037	3.98
Raio de Stokes [Å]	19.9	19.78 ± 0.01	0.60	20.69	3.97
Raio de Giro [Å]	15.3	15.47 ± 0.01	1.13	16.21	5.95
Volume [10^{-20}cm^3]	2.42	2.6 ± 0.1	8.13	3.489	44.17
Coef. Dif. Rot. [10^7s^{-1}]	1.95	1.98 ± 0.06	1.72	1.759	9.79
Viscosidade Intrínseca [cm^3/g]	3.49	3.44 ± 0.08	1.50	3.958	13.41
Coeficiente de Sedimentação [S]	1.839	1.85 ± 0.02	0.51	1.767	3.92

De modo geral, a média dos valores hidrodinâmicos possui a vantagem de informar o desvio padrão do resultado, ou seja, o grau de confiabilidade naquele valor. Já para os valores obtidos do modelo médio não temos essa informação. Para as médias dos valores hidrodinâmicos calculados para cada um dos modelos bead model vemos que a diferença entre valor médio e o da estrutura cristalográfica não apresentou mudanças significativas em relação ao obtido para apenas um modelo (Tabela 4,5), estando a principal diferença no valor do volume do modelo, 47%, ao passo que a diferença nos outros valores manteve-se menor do que 4%. Já para os resultados obtidos para o modelo médio vemos uma pequena melhora no valor do volume e raio de giro, 33% e 0.33%, mas um aumento sistemático da

variação com relação aos outros valores hidrodinâmicos. De qualquer maneira, a diferença neste caso ficou abaixo dos 7% em todos os valores (exceto o volume). Já a média dos valores hidrodinâmicos para cada um dos modelos *chain model* forneceu valores muito próximos aos obtidos para um modelo, estando a diferença principal no valor do volume médio (~8%). Olhando modelo a modelo vemos que em alguns casos a variação no volume chegou a 15% e, desta maneira, a promediação dos valores hidrodinâmicos permite uma melhor estimativa do volume da proteína com relação ao cálculo *ab initio*. Já a média dos modelos apresenta uma piora sistemática nos valores hidrodinâmicos, particularmente um grande aumento na diferença com relação ao volume (44%) e viscosidade intrínseca (13%). Isso já era esperado, pois no cálculo da estrutura mais provável constrói-se um modelo tipo *bead model*, perdendo desta maneira a característica *chain model* que, como vimos, fornece uma melhor descrição dos parâmetros hidrodinâmicos da partícula.

2.4. Modelagem tipo *chain model* assumindo grupos de resíduos

Como mostrado anteriormente, a modelagem tipo *chain model* permite uma melhor descrição da proteína espalhadora devido aos vínculos intrínsecos da modelagem de cadeia. No entanto, quando se tratam de proteínas grandes, com muitos resíduos, não é possível utilizar esta metodologia uma vez que o número de resíduos excede o valor máximo permitido no programa de modelagem. Como apresentamos, esta modelagem feita pelo programa GASBOR constitui-se da construção de uma seqüência de subunidades esféricas, formando uma cadeia linear, semelhante à uma estrutura primária de uma proteína. Nesta modelagem cada subunidade esférica corresponderia a cada resíduo individual. O programa busca a configuração desta cadeia que melhor descreve os dados experimentais. Como padrão, o programa assume distâncias entre resíduos de 3.8Å e raio dos resíduos de 1.9Å.

Uma alternativa para a utilização de modelagem tipo *chain model* em proteínas grandes consiste em fazer-se com que cada subunidade da seqüência do modelo corresponda à *mais de um resíduo* na seqüência primária da proteína. Para isso deveria ser possível aumentar o raio da subunidade de modo a corresponder ao mesmo volume do grupo de resíduos que se deseja mimetizar. Com o aumento do raio da subunidade diminui-

se o número total de resíduos a serem utilizados no cálculo, permitindo assim sua aplicação ao estudo de proteínas com um número elevado de subunidades. Se N_{DR} for o número total de resíduos da proteína original, assumindo o raio médio do resíduo como sendo $R_{DR}=1.9\text{\AA}$, o novo número de resíduos N_{new} e o novo raio das subunidades R_{new} são obtidos impondo-se que a massa total se mantenha constante:

$$N_{DR} \cdot R_{DR}^3 = N_{new} \cdot R_{new}^3 \quad (138)$$

No entanto, como o programa não permite que variemos o raio das subunidades devemos fazer um truque para poder-se realizar esta simulação. Da fórmula anterior obtemos a razão entre os dois raios (ou duas distâncias D quaisquer) em cada situação:

$$f = \frac{D_{new}}{D_{DR}} = \left(\frac{N_{DR}}{N_{new}} \right)^{1/3} \quad (139)$$

Sendo assim, a mudança do raio representa apenas uma mudança no fator de escala no espaço real. Já que não podemos alterar a distância entre as subunidades no modelo (espaço real) alteramos o vetor transferência de momento (espaço recíproco) de modo a corresponder à uma escala onde as distâncias entre as subunidades correspondam às dadas pelo programa de cálculo. Os valores de q que devem ser usados na simulação são dados por:

$$q_{simul} = q_{exp} \cdot f \quad (140)$$

Após a modelagem ser feita retorna-se à escala original utilizando as relações:

$$\begin{aligned} q_{new} &= \frac{q_{simul}}{f} \\ D_{new} &= D_{simul} \cdot f \end{aligned} \quad (141)$$

Para avaliar este método proposto, o aplicamos à descrição da lisozima. Realizaram-se simulações diminuindo N_{DR} desde o valor inicial (129 resíduos) até 13 resíduos. Os valores de N_{DR} e os respectivos raios são mostrados na Tabela 8.

Tabela 8 - Raios para as simulações.

N_{DR}	R [Å]	f
129	1.90	1.00
119	1.95	1.03
106	2.03	1.07
93	2.12	1.12
80	2.23	1.17
66	2.34	1.25
53	2.56	1.35
40	2.81	1.48
27	3.20	1.68
13	4.08	2.15

Renormalizando a intensidade espalhada e o modelo obtido pelas equações (140) e (141) com os valores de f dados pela tabela anterior, pude calcular modelos *ab initio* para cada um dos casos. Os resultados são mostrados abaixo:

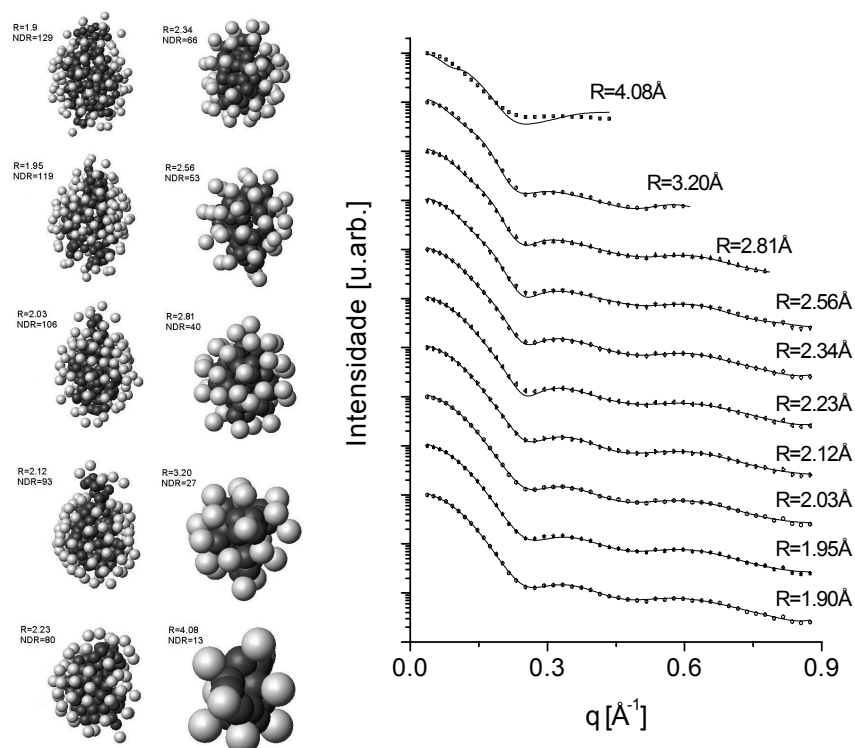


Figura 32 – Simulação *ab initio* para lisozima variando-se o raio das subunidades de modo a manter-se o mesmo volume. Esquerda: modelos obtidos em cada caso. Direita: ajuste dos dados experimentais

Como pode ser visto nos modelos acima, este método apresenta a desvantagem de "aumentar" o raio da camada de hidratação que fica no entorno da proteína. Esta camada, que idealmente possui $\sim 3\text{\AA}$ de espessura, acaba possuindo valores maiores em decorrência da correção de escala. Isto é crítico para casos onde a camada de hidratação seja comparável ao tamanho da partícula, como no caso com maior raio, $R=4.12\text{\AA}$ que corresponde a $\sim 9\%$ do tamanho da partícula espalhadora. Nosso intuito com esta metodologia alternativa de cálculo é possibilitar a modelagem tipo chain model em proteínas com elevado peso molecular que possuiriam milhares de resíduos. Para estes casos, uma pequena variação na espessura da camada de hidratação não fornece erros grandes, pois a influência desta camada é muito pequena.

Realizamos cálculos hidrodinâmicos para cada um dos modelos obtidos anteriormente. Nestes cálculos utilizamos para cada modelagem o raio ideal dado pela equação (137) determinado na seção anterior. Isso foi possível devido a todos os modelos estarem em uma escala onde os raios de suas subunidades eram $1.9\text{\AA}$. Além disso, como nos cálculos hidrodinâmicos deve-se informar o peso molecular da proteína, em cada caso devemos informar o peso molecular *efetivo* da partícula uma vez que alterando-se sua escala a mesma aparentará ser menos massiva. O peso molecular efetivo é dado pela relação:

$$Mw_{sim} = Mw / f^3 \quad (142)$$

Efetuada os cálculos hidrodinâmicos retornamos os valores das grandezas hidrodinâmicas à escala original utilizando as relações abaixo, determinadas com base nas fórmulas dos parâmetros hidrodinâmicos:

$$\begin{aligned} Mw &= Mw_{sim} \cdot f^3 \\ V &= V_{sim} \cdot f^3 \\ D_{new} &= D_{sim} / f \\ R_{stokes_{new}} &= R_{stokes_{sim}} \cdot f \\ R_{g_{new}} &= R_{g_{sim}} \cdot f \\ D_{R_{new}} &= D_{R_{sim}} / f^3 \\ \eta_{new} &= \eta_{sim} \\ S_{new} &= S_{sim} \cdot f^2 \\ Dmax_{new} &= Dmax_{sim} \cdot f \end{aligned} \quad (143)$$

sendo D_R o coeficiente de difusão rotacional (Van Holde, 1971) que é uma das saídas do programa de cálculo. Os valores obtidos da razão entre os parâmetros hidrodinâmicos calculados para cada modelo e o valor dado pela estrutura cristalográfica são mostrados na tabela abaixo:

Tabela 9 - Razões entre os parâmetros hidrodinâmicos calculados para cada um dos modelos e os parâmetros obtidos pela estrutura cristalográfica 6LYZ.

f	2.15	1.68	1.48	1.35	1.25	1.17	1.12	1.07	1.03
R_{new}	4.08	3.20	2.81	2.56	2.37	2.23	2.12	2.03	1.95
Mw ratio	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
D ratio	0.934	0.952	0.982	0.995	1.009	0.997	0.995	0.997	1.006
R_{Stokes} ratio	1.071	1.050	1.019	1.006	0.991	1.003	1.005	1.004	0.994
R_g ratio	1.092	1.059	1.025	1.010	1.000	1.009	1.021	1.012	1.021
V ratio	1.466	1.377	1.194	1.198	1.142	1.149	1.077	1.168	1.119
D_R ratio	0.823	0.873	0.959	1.000	1.038	0.995	0.984	0.983	1.017
η ratio	1.201	1.130	1.030	0.989	0.956	0.988	0.994	0.992	0.969
S ratio	0.952	0.971	1.001	1.014	1.029	1.017	1.015	1.016	1.026
D_{max} ratio	0.913	0.889	0.920	0.896	0.877	0.883	1.003	0.994	0.998

Como vemos anteriormente, mesmo utilizando-se um pequeno número de subunidades (cada uma correspondendo à vários resíduos) os modelos baseados na metodologia *chain model* conseguem restaurar valores hidrodinâmicos corretos mesmo nos casos extremos.

Uma aplicação direta desta metodologia é o estudo de eritrocruorinas, que apresentamos posteriormente na seção de Experimentos Realizados. A modelagem por SAXS deste tipo de proteína é feita através da utilização de modelos tipo *bead model* devido ao elevado peso molecular. No entanto, como mostramos nos itens anteriores, este tipo de modelos não fornecem valores hidrodinâmicos tão precisos quanto a modelagem por *chain models*. As limitações de cálculos hidrodinâmicos para eritrocruorinas partindo de modelos tipo *bead model* também são discutidas por Krebs et al. (Krebs et al., 2004) em um artigo recente. Desta forma, a metodologia apresentada aqui surge como alternativa para a correta determinação de parâmetros hidrodinâmicos para estes casos onde a metodologia por *chain models* não poderia ser aplicada.

2.5. Conclusões

Mostramos com as simulações anteriores que em modelagem *ab initio*, a melhor metodologia a ser aplicada para cálculo de parâmetros hidrodinâmicos consiste em, dada uma série de modelos *ab initio*, calcular os parâmetros hidrodinâmicos para cada um independentemente e após, calcular-se a média e o desvio padrão para o conjunto. Pelos nossos resultados os erros associados aos cálculos bem como a diferença dos modelos com relação ao valor real diminuem ao utilizar-se este enfoque. Excetuando-se o cálculo do volume do modelo, que somente pode ser calculado corretamente através de modelagem *chain model*, ambas as metodologias *ab initio* mostraram-se apropriadas para o cálculo de todos os outros parâmetros hidrodinâmicos (coeficiente de difusão translacional, Raio de Stokes, Raio de Giro, Coeficiente de difusão rotacional, viscosidade intrínseca e coeficiente de sedimentação). Esta análise foi crucial para a utilização correta do programa HIDROPRO no cálculo de parâmetros hidrodinâmicos para modelos *ab initio*. Ao final propomos uma nova aplicação da metodologia *ab initio* para modelagem *chain model* à proteínas que, inicialmente, não poderiam ser tratadas por esta metodologia. Pelos resultados obtidos vimos que aumentando-se o raio das subunidades correspondentes da “cadeia” do modelo *ab initio*, podemos aplicar esta metodologia de cálculo à modelagem de proteínas grandes com um número elevado de resíduos.

CAPÍTULO III

ESTUDOS

ESTRUTURAIS DE

PROTEÍNAS

1. Espalhamento a Baixos ângulos por proteínas com alto peso molecular^{**}

Diversos tipos de proteínas compõem-se de várias subunidades, formando uma estrutura quaternária extremamente grande. Em particular as hemoglobinas extracelulares encontradas em moluscos e anelídeos (também chamadas de eritrocruorinas) formam complexos cujas massas moleculares podem variar de ~1 a ~4 MDa, com tamanhos máximos variando entre ~150 e ~300 Å. Este tipo de proteína é praticamente impossível de ser cristalizado devido a sua elevada massa molecular. Diversos trabalhos (Weber e Vinogradov, 2001 e referencias citadas neste artigo) descrevem em detalhe as propriedades bioquímicas de diversas proteínas pertencentes a esta classe.

No passado, diversos experimentos de espalhamento de raios X a baixos ângulos foram realizados buscando obter informações conformacionais sobre componentes deste grupo de proteínas (Pilz et al., 1988, e ref. citadas). Porém as análises se limitavam à interpretação das funções de distribuição de pares de distâncias $p(r)$ correspondentes à intensidade espalhada bem como à construção de modelos simples. Em muitos casos esta modelagem era guiada por dados de microscopia eletrônica uma vez que o tamanho destas proteínas permitia sua visualização por microscópio eletrônico. Estas estruturas formam complexos altamente ordenados, com simetrias características (P222, P6, P62, P52 etc.). Com o desenvolvimento de técnicas para modelagem *ab initio* novas possibilidades para interpretação deste tipo de estruturas foram abertas. Mostraremos resultados para duas hemoglobinas extracelulares proveniente do molusco *Biomphalaria glabrata* e da minhoca da terra *Glossoscolex paulistus*.

^{**} Uma breve descrição das propriedades principais de proteínas é mostrada no apêndice A desta tese.

1.1. Espalhamento de raios X a baixos ângulos para hemoglobinas de *B. glabrata*^{††}

a) Introdução

No filo dos moluscos o principal transportador de oxigênio é a hemocianina (por conter cobre) enquanto que hemoglobinas são restritas a duas famílias de moluscos, Astartideos e Carditideos, e uma família de caracóis de água fresca, os Planorbideos. Nesta família, as hemoglobinas destes caracóis são proteínas com muitas subunidades e com pesos moleculares variando entre 1.54 a 2.25 MDa, reportados como sendo oligômeros compostos de um anel dipentamérico com subunidades com pesos moleculares de 350 a 400kDa. Alternativamente, foi proposto que a hemoglobina de *Helisoma trivolvis* possui um arranjo dihexamérico de subunidades, mas essa sugestão não foi suportada por experimentos posteriores (Herskovits e Hamilton, 1994). A hemoglobina de *Biomphalaria glabrata* (*B. glabrata*), também um caracol Planorbídeo, possui um peso molecular publicado de (1.75 ± 0.06) MDa, valor baseado em parâmetros hidrodinâmicos desta proteína (Almeida e Neves, 1974). Experimentos bioquímicos indicaram que, baseado no peso molecular desta proteína, poder-se-ia propor que a mesma contém 5 subunidades de 360kDa cada uma delas compostas por duas subunidades de 180kDa conectadas por pontes dissulfeto (Arndt e Santoro, 1998).

Neste trabalho realizamos experimentos de raios X a baixos ângulos com a proteína nas formas oxidada (*oxyhemoglobin*) à pH7.5 e de-oxidada (*cyanomethemoglobin*) à pH5.0. Como veremos, dos dados de SAXS foi possível obter-se informações estruturais para as duas formas bem como construir-se modelos *ab initio* de baixa resolução (Arndt e Oliveira et al., 2003). Destas informações pudemos mostrar que a hemoglobina de *B. glabrata* possui uma conformação octamérica e não decamérica como proposto para outros Planorbídeos.

b) Experimental

Os métodos de purificação e extração desta proteína dos organismos vivos é descrito em diversos trabalhos (Almeida e Neves, 1974, Arndt e Santoro, 1998). Por ser

^{††} Os resultados aqui apresentados formam parte da publicação Arndt MHL, **Oliveira CLP**, Régis WCB, Torriani IL and Santoro MM (2003), "Small Angle X-Ray Scattering of the Hemoglobin from *Biomphalaria glabrata*". *Biopolymers*, 69, 470-479.

abundante no sangue destes caracóis sua extração é muito eficiente, permitindo a purificação de uma grande quantidade de proteínas. A hemoglobina foi extraída no laboratório de Enzimologia e Físico-Química de Proteínas da Universidade Federal de Minas Gerais pelo grupo do professor Marcelo Santoro. As amostras preparadas foram trazidas para os experimentos de SAXS em Campinas.

As medidas de SAXS foram realizadas no Laboratório Nacional de Luz Síncrotron (LNLS) com concentrações que variaram de 0.5 a 10mg/ml. As intensidades de espalhamento foram coletadas utilizando um detector linear sensível à posição e o tratamento dos dados foi feito com o programa TRAT1D (Oliveira, 2003). Utilizamos duas distâncias amostra-detector (1530 e 630mm) de modo a ter-se um maior intervalo experimental. Com o comprimento de onda de 1.608Å o intervalo do módulo do vetor de espalhamento acessível aos experimentos foi de $0.0105 < q < 0.1814\text{\AA}^{-1}$. As medidas foram feitas tomando-se exposições de 900s, sendo as amostras mantidas em um porta amostras com 1mm de espessura e janelas de mica (Cavalcanti et al., 2004), mantidas a temperatura constante (20°C). Após os experimentos de SAXS testes para averiguar-se a monodispersidade da solução foram feitos bem como novas medidas dos valores das concentrações das amostras. Medidas de intensidades de espalhamento por água pura foram realizados nas mesmas condições experimentais de modo a permitir-se a renormalização à escala absoluta.

c) Resultados

De experimentos com cromatografia de filtração em gel^{††}, usada na purificação da proteína, vemos diferenças nos volumes de eluição da proteína em pH5.0 e pH7.5, o que reflete diferenças em sua massa molecular aparente. No entanto nenhum material dissociado foi observado nos experimentos (Figura 33).

^{††} Por este método de filtragem, também chamada de cromatografia de exclusão por tamanhos a solução de proteínas é passada por uma coluna de gel. Neste método as moléculas grandes atravessam a coluna mais rapidamente do que moléculas pequenas, que podem passar através dos poros. Sendo assim o volume de eluição (requerido para retirar a molécula da coluna) é maior para partículas menores (Voet et al., 2002).

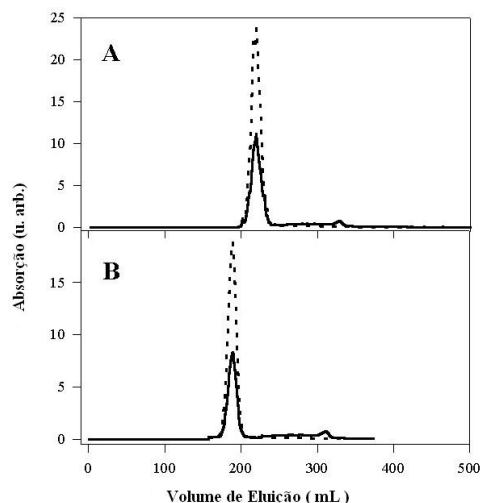
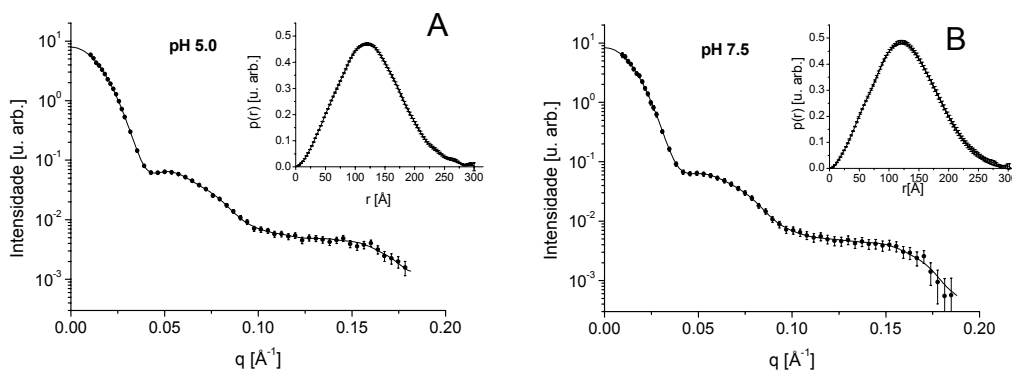


Figura 33 – Purificação da hemoglobina de *Biomphalaria glabrata* à pH5.0 e pH7.5. A – Cromatografia por filtração de gel com Sephacryl S400 em buffer acetato 0.16 mol/L pH 5.0 com 0.04 mol/L NaCl e 0.001 mol/L KCN. B - Cromatografia por filtração de gel com Sephacryl S400 em buffer 0.10 mol/L TRIS-HCl pH 7.5. A linha continua representa a absorção em 280nm e a linha tracejada à 419nm. Amostra: 3.5 mL de hemolimfa. Fluxo: 0.15 mL/min. Temperatura ambiente.

Das medidas de SAXS obtivemos os valores para o raio de giro (R_g) através do ajuste teórico da intensidade utilizando o programa GNOM (Semenyuk e Svergun, 1991). Do ajuste teórico tem-se também a extrapolação dos dados à $q=0$ o que permitiu o cálculo do peso molecular da proteína. Também deste ajuste, calcula-se a distribuição de pares de distâncias ($p(r)$) da qual pôde-se obter a máxima dimensão (D_{max}) da partícula bem como indicações sobre sua forma geral.

As intensidades de espalhamento e correspondentes $p(r)$'s são mostrados na figura abaixo.



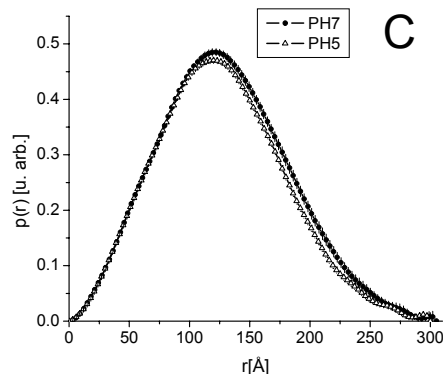


Figura 34 – Curva de espalhamento, ajuste teórico e função de distribuição de distâncias para hemoglobina de *B. glabrata*. A - Curva de espalhamento da proteína em pH5.0, buffer acetato 0.16 mol/L com 0.04 mol/L NaCl e 0.001 mol/L KCN à 20 °C. Inserção – Função $p(r)$. B – Curva de espalhamento para proteína em pH7.5, buffer TRIS-HCl 0.10 mol/L à 20 °C. Inserção - Função $p(r)$. C – Superposição das funções de distribuição de pares de distâncias.

Do ajuste obtemos que as proteínas a pH7.5 e 5.0 possuem raios de giro com valores $101.8 \pm 0.2 \text{ \AA}$ e $98.6 \pm 0.5 \text{ \AA}$ respectivamente e máximas dimensões $305 \pm 10 \text{ \AA}$ e $300 \pm 10 \text{ \AA}$. Nas curvas de espalhamento, vemos um mínimo em $q \sim 0.04 \text{ \AA}^{-1}$, indicando que a partícula possui uma forma globular. Pela forma da função $p(r)$ vemos que a partícula possui uma conformação ligeiramente prolata.

Utilizando a água como padrão secundário foi possível renormalizar-se a intensidade à intensidade absoluta de modo a calcularmos o peso molecular. A extrapolação da curva de espalhamento a concentração zero permitiu o cálculo do parâmetro $I(0)/c$ que se relaciona com a seção de choque da partícula através da equação (97)

$$\frac{1}{c} \frac{d\Sigma}{d\Omega} \Big|_{\text{sample}} = \left[\frac{I(0)_{\text{sample}} / c}{I(0)_{\text{water}}} \right] \frac{d\Sigma}{d\Omega} \Big|_{\text{water}}$$

sendo $I(0)_{\text{sample}}$ a intensidade a ângulo zero dos dados experimentais, $I(0)_{\text{water}}$ a intensidade do espalhamento da água a ângulo zero nas mesmas condições experimentais e $d\Sigma/d\Omega|_{\text{water}}$ a seção de choque macroscópica de espalhamento da água ($1.6456 \times 10^{-2} \text{ cm}^{-1}$ at 293K).

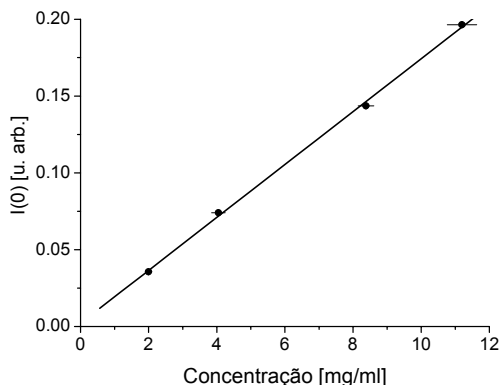


Figura 35 – Extrapolação para intensidade na origem para diferentes concentrações de proteína.

Utilizando esta relação obtemos a seção de choque de espalhamento para a proteína estudada bem como o peso molecular e seu volume usando as expressões já mostradas anteriormente (equação 110),

$$M = \frac{1}{c} \frac{d\Sigma(0)}{d\Omega} \left(\frac{N_A}{(\Delta\rho\bar{v})^2} \right)$$

$$V = \frac{\bar{v} \cdot M}{N_A}$$

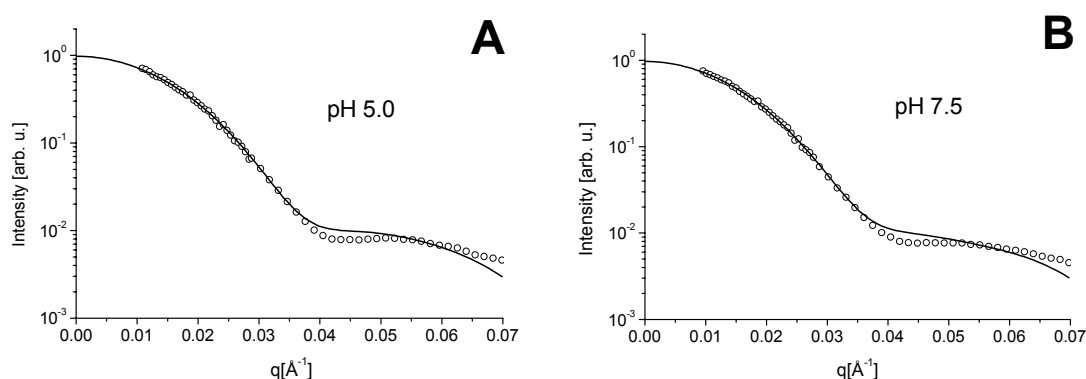
sendo $\Delta\rho$ o contraste de densidade de espalhamento da partícula e solvente e $\bar{v}$ o volume específico parcial da partícula. Calculamos com estas expressões o peso molecular e o volume da *B. glabrata*, obtendo os valores de (1.5 ± 0.1) MDa para o peso molecular e $(1.8 \pm 0.1) \times 10^{-18} \text{cm}^3$ para o volume da proteína. No entanto, como o volume parcial específico para a forma em pH 5.0 possui diferentes valores reportados na literatura, $0.682 \text{cm}^3/\text{g}$ e $0.73 \text{cm}^3/\text{g}$ (Almeida e Neves, 1974), e como o peso molecular depende crucialmente deste valor, o valor médio de 0.71 foi utilizado nos cálculos. Se compararmos o volume experimental com o de uma esfera com o mesmo raio de giro teremos a razão $V_{\text{partícula}}/V_{\text{esfera}} = 0.21$, indicando que a partícula não possui forma esférica, o que confirma nossa análise da forma da função $p(r)$.

As diferenças entre as curvas de espalhamento para as duas formas são pequenas. No entanto essas diferenças promovem valores distintos para os parâmetros dimensionais calculados. A curva $p(r)$ para as duas amostras é mostrada na Figura 34C, onde vemos

claramente a diferença na forma da curva (e conseqüentemente na forma da partícula) e também nos valores de máxima dimensão de cada proteína.

d) Cálculo de modelos

Esta proteína possui uma composição de vários domínios que devem se agrupar com uma determinada simetria. Como mencionado na introdução teórica, informações de simetria são muito úteis em modelagens a partir de curvas de espalhamento e por isso buscamos metodologias para podermos propor alguma simetria específica para esta proteína. Dos métodos existentes para modelagem *ab initio* somente o método de expansão multipolar (Programa SASHA - veja *Introdução Teórica – Método de Esféricos Harmônicos*) é capaz de fornecer resultados únicos (Svergun et al, 1997), dentro dos limites desta modelagem. Sendo assim, para tentarmos obter informações preliminares sobre a forma da partícula espalhadora, utilizamos a expansão multipolar sem nenhuma restrição de simetria para determinarmos o envelope externo de baixa resolução da proteína nas duas formas, para a região a baixos ângulos da intensidade espalhada ($q < 0.07 \text{ \AA}^{-1}$). O resultado é mostrado na Figura 36.



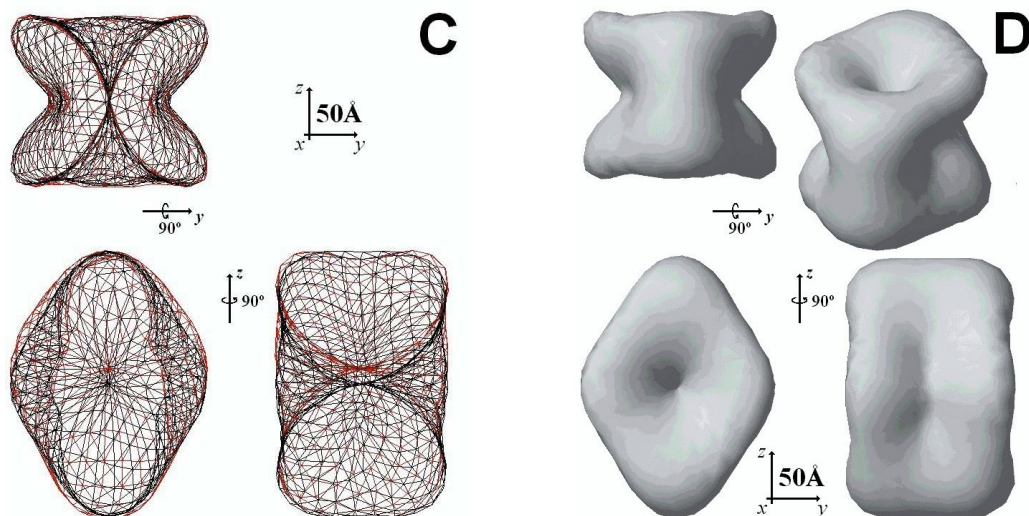


Figura 36 – Envelope externo para a hemoglobina de *B. glabrata* calculada utilizando expansão multipolar sem imposição de simetria. A – Ajuste da parte a baixos ângulos da curva de intensidade ($q_{max} \sim 0.07 \text{Å}^{-1}$) para proteína em pH5.0. B – Mesmo ajuste para proteína em pH7.5. C – Três vistas diferentes para envelope tridimensional das duas proteínas (pH5.0 (preto) e pH7.5 (vermelho)). A forma molecular é praticamente a mesma para os dois casos, indicando mesma simetria geral. Na figura os dois modelos estão superpostos. D – Superfície sólida para hemoglobina à pH5.0.

O ajuste das curvas é muito bom na parte inicial das curvas de SAXS, até $q \sim 0.035 \text{Å}^{-1}$. Isso nos dá a indicativa de que a forma geral da proteína é corretamente descrita pelos modelos obtidos. Nos modelos calculados vemos aspectos importantes sobre a forma destas proteínas. Primeiramente temos uma indicação do grau de anisotropia das partículas e, além disso, a presença de vínculos de simetria. Uma simetria P_{222} é claramente vista no envelope externo da proteína. Isto nos dá a indicativa de que poderíamos propor esta simetria para a proteína. Vemos uma pequena mudança na forma geral dos envelopes das proteínas indicando possíveis mudanças conformacionais entre os dois pH's. Para confirmar os resultados repetimos o ajuste acima impondo este vínculo de simetria. O resultado é mostrado na Figura 37.

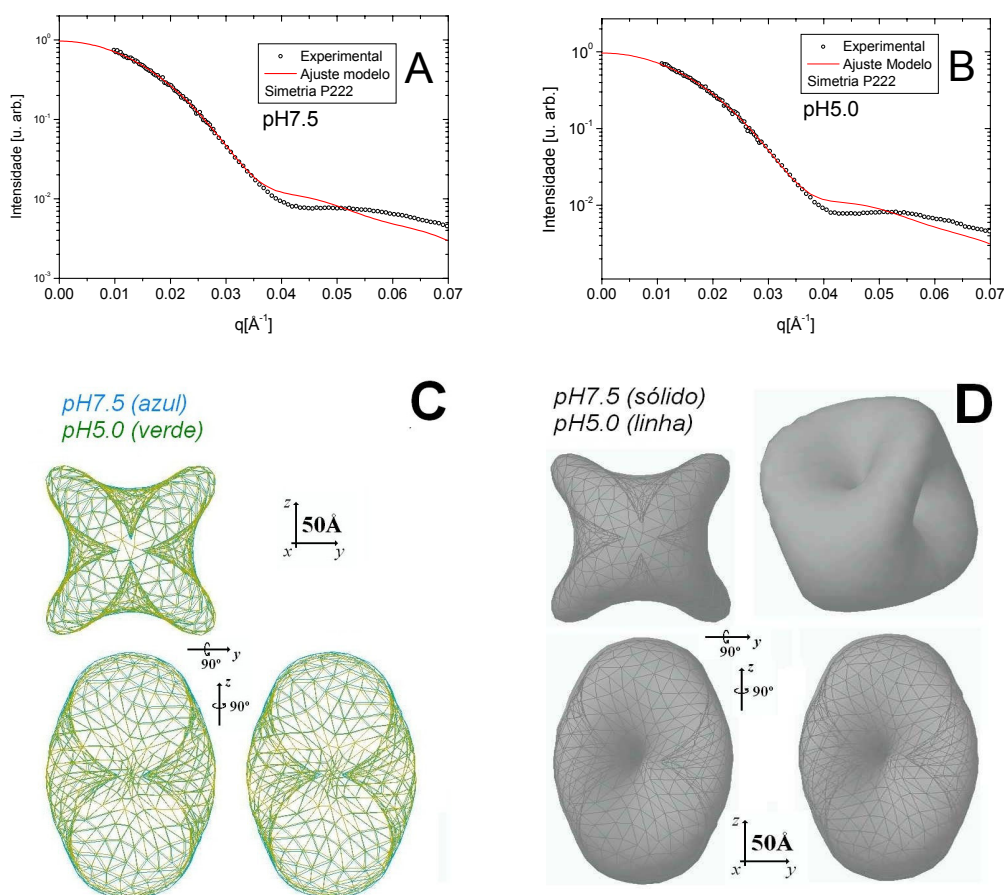


Figura 37 – Envelope externo para a hemoglobina de *B. glabrata* calculada utilizando expansão multipolar com simetria P222. A – Ajuste da parte a baixos ângulos da curva de intensidade ($q_{\text{max}} \sim 0.07 \text{\AA}^{-1}$) para proteína em pH7.5. B – Mesmo ajuste para proteína em pH5.0. C – Três vistas diferentes para envelope tridimensional das duas proteínas (pH5.0 (azul) e pH7.5 (verde)). Na figura os dois modelos estão superpostos. D – Superfície sólida para hemoglobina à pH7.5 superposto ao pH5.0 na forma de linhas.

Vemos que, por força da imposição de simetria, este segundo envelope é ligeiramente diferente do obtido sem imposição de simetria, mas ainda mantém suas características principais. Além disso, a qualidade do ajuste dos dados de SAXS é comparável, aumentando assim a confiabilidade dos resultados. As pequenas diferenças ainda podem ser vistas nos modelos, mas de uma maneira mais sutil devido à alta simetria imposta.

Esta informação de simetria obtida de forma conclusiva pelo método de expansão multipolar nos possibilitou a utilização de outra metodologia de cálculo que permitiu um ajuste melhor dos dados experimentais. Utilizamos este vínculo de simetria como

parâmetro de entrada do programa DAMMIN que constrói modelos tipo *bead model* (*Aspectos Teóricos – Método de elementos finitos*). Como vimos anteriormente, este tipo de ajuste é superior à expansão multipolar, pois permite a descrição de estruturas mais complexas do que as descritas por esféricos harmônicos. No entanto, como mostrado anteriormente, a solução não é única como na expansão multipolar e devido a isso realizamos a promediação dos modelos calculados. O intervalo completo do espaço recíproco utilizado nos cálculos forneceu uma quantidade de informação correspondente à 16 *Shannon Channels*, e uma resolução ($2\pi/q_{\max}$) de $\sim 30\text{\AA}$. A adição de vínculos de simetria aumenta a confiabilidade dos modelos uma vez que aumenta a quantidade de informação de sua estrutura, diminuindo a redundância.

Tabela 10 - Parâmetros hidrodinâmicos experimentais e comparação com valores obtidos para modelos *ab initio*.

Parâmetro Hidrodinâmico	Experimental	Model	Model
	pH 5.0	pH 5.0	pH 7.5
Coef. Dif. Trans. [D] ($10^{-7} \text{ cm}^2 \text{ s}^{-1}$)	1.75 ± 0.03^a	1.80	1.74
Raio de Stokes ($\text{\AA}$)	116 ± 1^b	117	120
Visc. Intr. $[\eta]$ (cm^3/g)	6.80 ± 0.42^c	7.0	7.4
Coef. Sedimentação [s] (Svedberg)	35 ± 1^b	34	33
f/f_0	1.59 ± 0.03^a	1.59^d	1.62^d

^a from Almeida and Neves, 1976.

^b data obtained by Arndt et al. 2000

^c Obtido por Arndt e Oliveira et al., 2003

^d Obtido a partir dos valores de D e s .

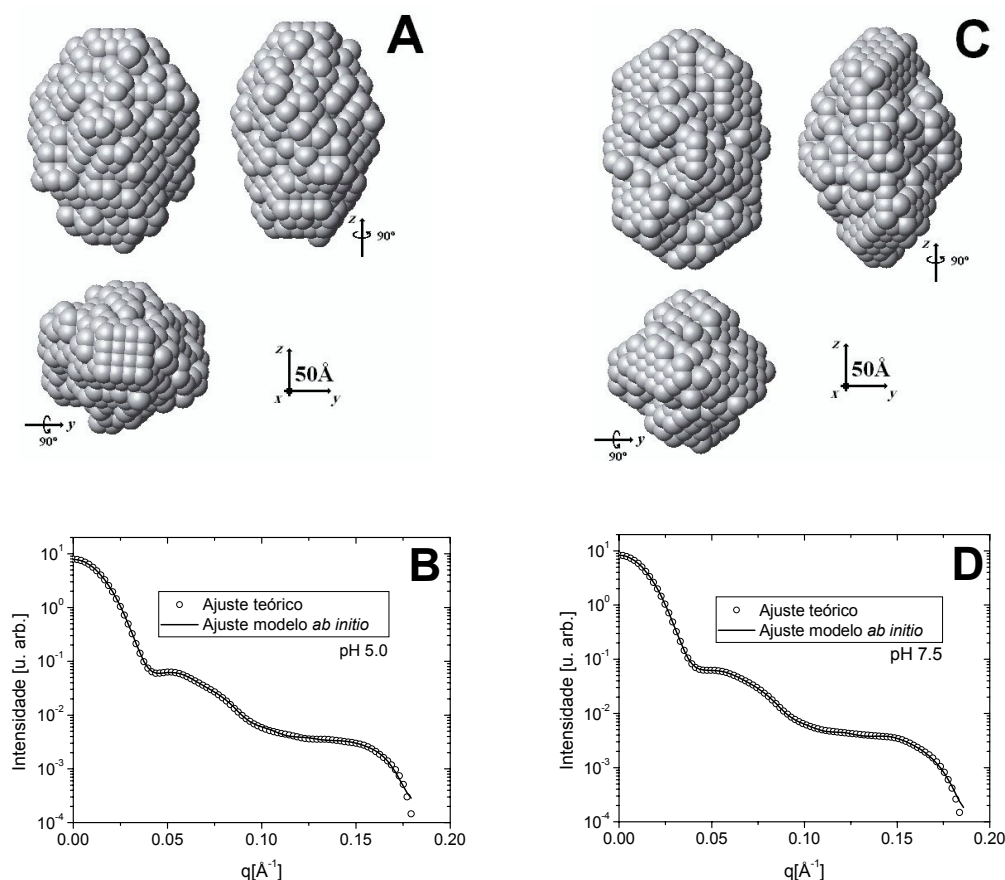


Figura 38 – Modelo mais provável calculado para proteína *B. glabrata* por modelagem tipo dummy atom. O modelo corresponde à média de 10 cálculos independentes com imposição de simetria P222. A – Modelo tridimensional para proteína à pH5.0. B – Ajuste dos dados à pH5.0 para um dos modelos. C – Modelo tridimensional para proteína à pH7.5. D – Ajuste dos dados à pH7.5 para um dos modelos.

Os modelos promediados são mostrados na Figura 38. Mostramos juntamente o ajuste da intensidade experimental pela correspondente dos modelos, sendo um ajuste praticamente perfeito. De modo a checarmos a confiabilidade dos modelos, cálculos hidrodinâmicos foram calculados para os modelos de SAXS e comparados com valores experimentais. Os resultados são mostrados na Tabela 10, com ótima concordância entre si.

e) Considerações finais

Observam-se diferenças significativas observadas nos volumes de eluição para a hemoglobina de *B. glabrata* em pH 5.0 e pH 7.5 (Figura 33). Estas diferenças podem vir de mudanças na forma da proteína ou de dissociação de suas subunidades. No entanto, nenhum material dissociado foi observado nos experimentos. Também, as duas soluções

tampão em que as proteínas estavam continham força iônica maior do que 0.1M, o que impede a ocorrência de interações iônicas das proteínas com o gel da matriz cromatográfica (Arndt e Oliveira et al., 2003). Sendo assim, este comportamento dos volumes de eluição foi interpretado como sendo consequência da expansão da proteína em pH 7.5 ocasionada pela repulsão de cargas negativas em excesso no interior da proteína, alterando sua estrutura mas preservando seu peso molecular absoluto.

Os parâmetros dimensionais R_g e D_{max} obtidos dos dados de SAXS são compatíveis com as dimensões obtidas utilizando esta mesma técnica por Pilz et al. (Pilz et al., 1988), para hemoglobina de *H. trivialis*. Além disso, os resultados dimensionais obtidos por microscopia eletrônica para a proteína *B. glabrata* (Arndt et al., 2000) e outras hemoglobinas de Planorbídeos (Herskovits e Hamilton, 1994) também apresentam ótima concordância com os parâmetros obtidos para as proteínas analisadas aqui.

O envelope obtido para a proteína, calculado a partir da região a baixos ângulos da curva experimental, é adequado para descrever-se superfícies suaves e pode fornecer informações de simetria à baixa resolução. Como dissemos anteriormente, o bom ajuste dos dados para $q_{max} \sim 0.07 \text{ \AA}^{-1}$ nos dá a indicativa de que a simetria P_{222} pode ser proposta para esta proteína além de indicar pequenas mudanças na forma geral das partículas com os valores de pH.

A utilização de vínculos de simetria nos modelos tridimensionais calculados para os dados de SAXS aumenta a informação pertinente à estrutura geral da proteína e, desta maneira, favorece a obtenção de um modelo mais confiável. Por isso, como segundo passo neste estudo utilizamos modelagem tipo *bead model* para curva completa dos dados experimentais, impondo vínculos de simetria. Vemos que a forma geral do modelo é similar à dada pelo método de elementos finitos, porém com maiores detalhes. Do conjunto de 10 ajustes independentes para cada pH, obtivemos a solução mais provável através do processo de média espacial dos modelos, o que aumenta em muito a confiabilidade do resultado. Dos modelos mostrados na Figura 38 vemos que a proteína em pH 7.5 é levemente mais expandida do que a em pH 5.0. Este fato se correlaciona muito bem com o que observamos na cromatografia de filtração por gel.

O modelo tetramérico para esta hemoglobina foi proposto, pela primeira vez, a partir de experimentos de *cross-linking* seguida de eletroforese em gel denaturante de

agarose, onde observou-se quatro bandas ao invés de 5, como esperado para uma estrutura pentamérica (Arndt et al., 2000). Neste mesmo trabalho, a partir de medidas de viscosidade e densidade, a viscosidade intrínseca ($[\eta]$) e o volume específico parcial ($\bar{v}$) foram calculados, fornecendo um valor de peso molecular de aproximadamente 1.5MDa, valor consistente com o obtido em nossos experimentos. Além disso, dados de microscopia eletrônica por transmissão não forneceram os anéis pentaméricos característicos para outros caramujos Planorbídeos (Ilan et al., 1986).

O peso molecular obtido pelos experimentos de SAXS é menor do que o obtido por Almeida e Neves (Almeida e Neves, 1974) e similar ao dado por Arndt et al. (Arndt et al., 2000). Conseqüentemente a relação oligômero/subunidade entre os pesos moleculares (1500/360) é igual a 4.2, o que está em concordância com a estrutura tetramérica para a molécula, como mostrado na Figura 39. A razão axial para este modelo simples é 1.6, a é bastante similar aos valores obtidos para os modelos de SAXS e também consistente com os valores de coeficiente friccionais relativos (f/f_0) obtidos para esta proteína (Tabela 10). Além disso, os valores hidrodinâmicos calculados para os modelos são bastante similares aos experimentais e com isso temos uma maior confiabilidade dos resultados de SAXS e da modelagem feita.

De acordo com os resultados bioquímicos citados aqui bem como as informações estruturais obtidas dos dados de espalhamento, a composição de subunidades da proteína é descrita como sendo de oito domínios idênticos, cada um representando uma cadeia de 180kDa, podendo ser também interpretado como contendo quatro subunidades de 360kDa.

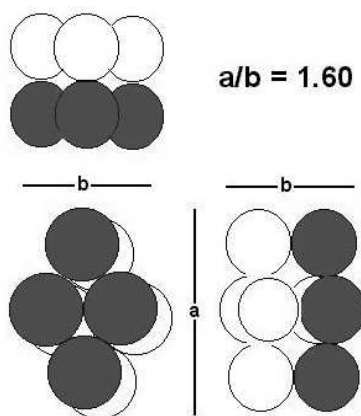


Figura 39 – Modelo esquemático proposto para a hemoglobina de *B. glabrata*. A razão a/b para esta figura geométrica é muito próxima à obtida para os modelos *ab initio*.

Do ponto de vista fisiológico, se compararmos as hemoglobinas hexaméricas em formato de anel, ou outras proteínas também em formato de anel de outros Planorbídeos, o arranjo de subunidades mostrado aqui possui uma área de contato com o solvente otimizada, fazendo com que esse arranjo seja mais eficiente para troca de oxigênio (O_2). Mesmo com as limitações da técnica de SAXS, algumas diferenças estruturais puderam ser detectadas para as duas condições de pH estudadas. Estas diferenças, juntamente com as mudanças nos parâmetros dimensionais observados, podem ser relacionadas com propriedades cooperativas já descritas para esta proteína (Bugge et al., 1999).

A principal conclusão deste trabalho foi que o arranjo de subunidades mais provável para a hemoglobina extracelular de *B. glabrata* não é um arranjo pentamérico oblato como proposto anteriormente, mas sim tetramérico, levemente prolato.

1.2. Forma e estrutura da hemoglobina extracelular de *G. paulistus* por SAXS

a) Introdução

As propriedades físico-químicas de proteínas extracelulares são conhecidas desde 1930. Em particular, diferentes afinidades com oxigênio e cooperatividades são reportadas para moléculas com quantidades similares de grupos heme, dimensões e pesos moleculares. Este fato leva muitos pesquisadores a focalizarem atenção nas possíveis diferenças estruturais que podem explicar este comportamento funcional diverso (Decker et al, 1990). Dois artigos muito detalhados explanam esta diferença funcional nas estruturas de hemoglobinas extracelulares (Chung, 1979; Weber, 2001). O desafio, em todos os casos, tem sido a elucidação da forma de interação para as mais de 200 subunidades destas proteínas respiratórias. Esta auto-organização espontânea e cooperativa na ligação de oxigênio, existente nestas proteínas não pôde até agora ser compreendida completamente. Neste estudo, apresentamos os resultados de experimentos com a hemoglobina extracelular de *Glossoscolex paulistus* (*G. paulistus*), com peso molecular de 3.1MDa. A partir de dados de espalhamento de raios X a baixos ângulos e as técnicas de modelagem *ab initio* mencionadas anteriormente, pudemos descrever o arranjo de subunidades destas moléculas bem como obter os parâmetros dimensionais que foram confirmados por cálculos hidrodinâmicos para os modelos propostos. É interessante notar que, no que diz respeito a

propriedades bioquímicas, apenas pequenas diferenças podem ser encontradas quando comparamos as hemoglobinas de *Lumbricus terrestris* (L. terrestris, Fushitani et al., 1991) e *G. paulistus*. A estrutura desta proteína foi estudada por dissociação induzida por pH e dissociação por alta pressão por Bonafe et al. (Bonafe et al., 1991) e Silva et al. (Silva et al., 1989), indicando uma similaridade marcante nestas duas proteínas apesar de diferenças em seu peso molecular. Por esta razão é interessante comparar o resultado da modelagem ab initio a partir dos dados de SAXS (Oliveira et al., 2004) com a estrutura cristalográfica proposta por Royer et al. (Royer et al., 2000) para a proteínas L. terrestris.

b) Parte experimental e métodos de cálculo utilizados

A preparação das amostras, compreendendo a coleta e purificação da hemoglobina é descrita em outros trabalhos (Silva et al., 1989). Similarmente à proteína previamente estudada, *B. glabrata*, a eficiência da purificação é muito grande, fornecendo amostras de boa qualidade e com alta concentração. (~40mg/ml). Os parâmetros moleculares (volume específico parcial $\bar{v}$ e coeficiente de Sedimentação s) para a proteína *G. paulistus* foram obtidos da literatura (Almeida e Neves 1976, Costa et al. 1988).

Os experimentos de espalhamento a baixos ângulos foram feitos na linha de SAXS do Laboratório Nacional de Luz Síncrotron, sendo a intensidade espalhada coletada em duas etapas de experimentos: na primeira etapa os dados foram coletados utilizando um detetor linear sensível à posição e em outro turno de medidas, utilizando um detetor bidimensional. Os resultados de modelagem apresentados aqui se referem aos dados obtidos com o detetor unidimensional, sendo que comentaremos os dados bidimensionais posteriormente no texto. Para as medidas com o detetor linear o intervalo de dados experimentais foi de $q = 0.005$ à $0.1882\text{\AA}^{-1}$, correspondendo aos dados unidos referentes a duas distâncias experimentais (1.74m e 0.84m). Várias concentrações de proteína foram utilizadas nos experimentos, variando as concentrações de 0.5-40mg/ml. O tampão utilizado nos experimentos foi 0.05M Tris-HCl em pH7.5 à temperatura constante de 20°C. Utilizando esta série de concentrações as curvas de intensidade foram extrapoladas à concentração zero. As medidas foram feitas utilizando porta-amostra de capilar com 1.5mm de diâmetro (Cavalcanti et al., 2004). Todo o tratamento dos dados foi feito utilizando o programa TRAT1D (Oliveira, 2003).

O ajuste teórico dos dados experimentais foi feito com o programa GNOM (Semenyuk e Svergun, 1991), permitindo a determinação do raio de giro da partícula espalhadora (R_g) bem como a curva de distribuição de distâncias da partícula $p(r)$. Com os valores extrapolados da intensidade na origem ($I(0)$) foi possível determinarmos o peso molecular para a proteína bem como renormalizar os dados à escala absoluta, através do espalhamento de uma amostra de água, nas mesmas condições experimentais. Além da determinação do peso molecular da proteína expressamos os dados em função da seção de choque de espalhamento por partícula $I_1(q)$ utilizando a expressão já mencionada anteriormente (equação 111)

$$I_1(q) = \frac{d\Sigma(q)}{d\Omega} \frac{1}{c^2} \frac{d\Sigma(0)}{d\Omega} \frac{1}{(\Delta\rho\bar{v})^2}$$

sendo os componentes desta equação os mesmos já foram descritos na análise dos dados da B. glabrata.

Na modelagem dos dados experimentais utilizamos informações baseadas em estudos por microscopia eletrônica de modo a poder propor vínculos de simetria para a proteína. Muitos autores (Weber e Vinogradov, 2001) sugerem estruturas em forma cilíndrica (oblata) compondo-se de bicamadas com subunidades grandes arranjadas em formas hexagonais ou pentagonais, subunidades estas compostas de pequenas partículas. Para a modelagem desta proteína utilizamos modelos tipo *bead model*, com o programa DAMMIN (Svergun, 1999) utilizando dois espaços de busca distintos: esférico e cilíndrico (vazado). Com os modelos ab initio realizamos cálculos hidrodinâmicos comparando com resultados experimentais para testar a validade dos modelos quanto às propriedades hidrodinâmicas da proteína.

c) Resultados

Os dados de SAXS renormalizados à intensidade absoluta são mostrados na Figura 40, onde mostramos também o ajuste com o programa GNOM e também a função e distribuição de distâncias (*inserção*), da qual pudemos obter os valores de R_g e D_{max} para a proteína.

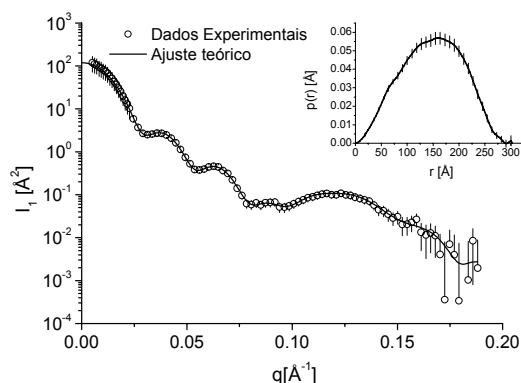


Figura 40 – Curva experimental de espalhamento normalizada para seção de choque de uma partícula e ajuste teórico para hemoglobina de *G. paulistus*. Inserção : função de distribuição de pares de distâncias ($p(r)$).

O raio de giro obtido foi $113.6 \pm 0.7 \text{ \AA}$ e a máxima dimensão $300 \pm 10 \text{ \AA}$. Estes valores são compatíveis às dimensões determinadas por experimentos de microscopia eletrônica (Bonafé et al., 1991) e medidas de SAXS previamente realizadas par esta proteína (Souza, 1990). Dos dados de microscopia é evidente a existência de uma simetria P62 na proteínas, informação esta que será útil no cálculo dos modelos. Dos cálculos em escala absoluta obtemos para esta proteína o peso molecular de $3.1 \pm 0.2 \text{ MDa}$ e $3.8 \pm 0.1 \times 10^6 \text{ \AA}^3$ para o volume hidratado. Se compararmos este volume com o de uma esfera de mesmo raio de giro da proteína obtemos a razão $V_{\text{partícula}}/V_{\text{esfera}} = 0.28$, indicando que a partícula possui uma forma não estritamente esférica. Na realidade, pela forma da função $p(r)$ vemos que a partícula possui uma conformação globular oblata.

Dos dados experimentais tem-se que a quantidade de informação para esta proteína é de ~ 18 *Shannon channels*. Visando aumentar a confiabilidade dos modelos ab initio obtidos vamos impor no cálculo vínculos de simetria de modo que o modelo quaternário seja melhor restaurado. Várias simetrias foram testadas nos cálculos (P3, P32, P6 e P62). O modelo obtido para a simetria P62 mostrou a melhor concordância com os dados de microscopia eletrônica. Na Figura 41B mostramos a média de 10 modelos independentes. O cálculo ab initio foi feito no modo lento (*slow mode*) onde utiliza-se um número maior de subunidades esféricas devido a seu raio ser menor. Além disso o processo de otimização por *simulated annealing* é mais lento, visando aproximar-se mais do mínimo global (Volkov e Svergun, 2003). Com este ajuste, os modelos *ab initio* provenientes dos cálculos

ficaram muito semelhantes ($NSD < 0.9$ para cada par de modelos). O ajuste dos dados experimentais para estes modelos é mostrado na figura 41C onde vemos uma ótima descrição dos dados para todo o intervalo medido. Para comparação mostramos a estrutura obtida por Royer et al. (2000) para a hemoglobina de *Lumbricus terrestris* (Figura 41A) utilizando cristalografia de proteínas. Dos dados de microscopia e também do modelo promediado, é notória a presença de uma região central da proteína vazada. Isso aliado à forma da proteína de *L. terrestris* nos dá a indicativa de que provavelmente a proteína *G. paulistus* possua uma região central vazada. Isso demonstra uma das vantagens de se utilizar modelos promediados. Dos dez modelos *ab initio* obtidos alguns possuíam região central vazada, outros apenas com menos densidade de subunidades na região central. No entanto o modelo promediado, que leva em conta a frequência de aparecimento de uma dada subunidade em uma região dos modelos, fornece que a estrutura mais provável possui realmente o interior vazado.

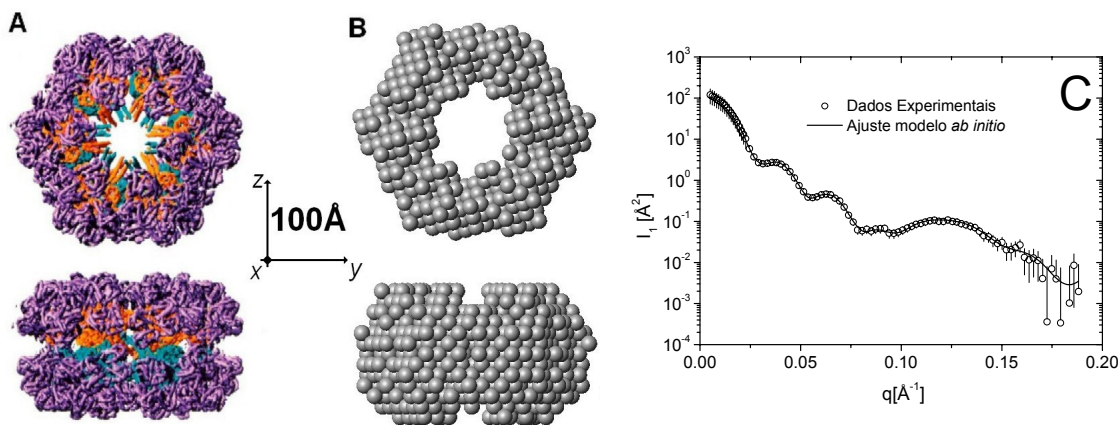


Figura 41 – *Modelo ab initio* calculado com espaço inicial de busca esférico para *G. paulistus*. A) Estrutura cristalográfica para *Lumbricus terrestris* – tirado de Royer et al.(2000). B) Modelo mais provável para proteína de *G. paulistus*. Este modelo corresponde à média de 10 modelos independentes calculados com espaço inicial de busca esférico e simetria P62. C) Ajuste para um dos modelos.

Esta informação adicional poderia ser utilizada como entrada nos cálculos *ab initio*. Uma região de diâmetro $\sim 50 \text{\AA}$ pode ser proposta como correspondendo a uma passagem central para a *L. terrestris* e podemos em princípio esperar um orifício desta mesma ordem de grandeza para a *G. paulistus* com base no modelo apresentado. Sendo assim, como cálculo complementar, ao invés de utilizar um espaço de busca inicial esférico, o qual

permite ao modelo assumir qualquer conformação (dentro do vínculo P62) utilizamos um espaço de busca mais restritivo, composto de um cilindro com diâmetro 320Å, altura 210Å e orifício central com diâmetro 40Å. Pelas dimensões do modelo médio obtido vemos que este espaço de busca é capaz de conter a estrutura da *G. paulistus*. Considerações semelhantes a estas foram propostas por Krebs et al (Krebs et al., 2004) em modelagens *ab initio*, levando-os a considerar naquele trabalho apenas as soluções que forneciam a região central vazada. Aqui, como constatamos um orifício central na proteína, podemos propor um novo espaço de busca.

O modelo mais provável para 10 cálculos independentes (novamente no modo lento) é mostrado na Figura 42 com o respectivo ajuste. Vemos a restrição do espaço de busca tornou o modelo um pouco mais simétrico, mas ainda muito semelhante ao obtido no caso de espaço de busca esférico (para os modelos promediados $NSD < 0.7$). A principal vantagem da restrição no espaço de busca foi o ganho de tempo no cálculo de cada modelo. Modelos com a restrição no espaço de busca foram calculados na metade do tempo dos modelos com espaço de busca esférico. No entanto, em ambos os casos tivemos resultados similares.

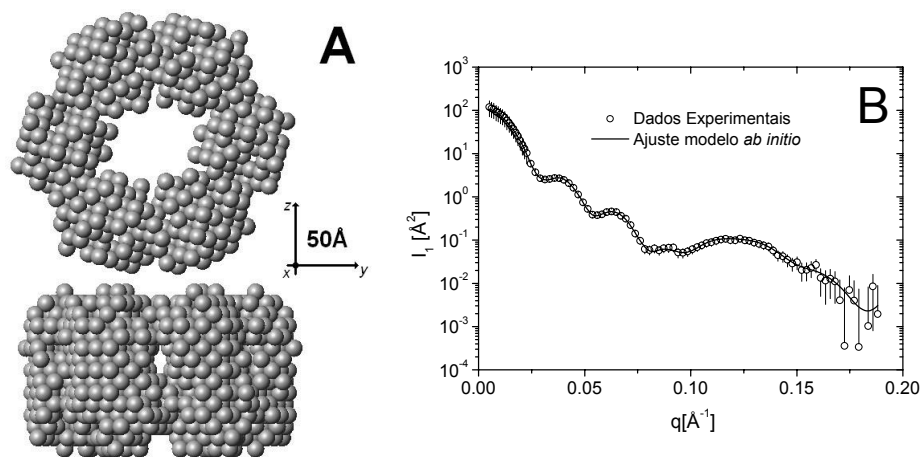


Figura 42 – *Modelo ab initio* calculado com espaço inicial de busca cilíndrico para proteína de *G. paulistus*. Este modelo corresponde à média de 10 modelos independentes calculados com espaço inicial de busca cilíndrico e simetria P62. B) Ajuste dos dados para um dos modelos.

Para comprovar a validade dos modelos realizamos cálculos hidrodinâmicos com os modelos *ab initio* obtidos de modo a comparar com valores experimentais. Por exemplo, o valor do coeficiente de fricção relativo da proteína (f/f_0) (Costa et al., 1988) é bastante parecido ao obtido para o modelo. Também, o peso molecular obtido pelos dados de SAXS

estão em ótima concordância com o esperado para esta proteína por cromatografia de filtração em gel (Costa et al. 1988). Todos os parâmetros hidrodinâmicos calculados para os modelos de SAXS e os obtidos por valores experimentais são mostrados na Tabela 11. Apesar de em alguns casos não termos valores idênticos, de maneira geral temos boa concordância de valores. Vale ressaltar que a medida de parâmetros hidrodinâmicos para proteínas com elevado peso molecular não é uma tarefa simples uma vez que a alta massa molecular pode acarretar erros de interpretação dos resultados (Krebs et al, 2004).

Tabela 11 – Parâmetros Hidrodinâmicos e peso molecular para hemoglobina de *G. Paulistus* obtida por SAXS e calculada pelos modelos *ab initio*. Modelo 1: Espaço de busca inicial esférico. Modelo 2: Espaço de busca inicial cilíndrico.

Parâmetro	Valor Experimental	Modelo 1	Modelo 2
Mw [MDa]	3.1 ± 0.2^a	---	---
Volume [10^6 Å]	3.8 ± 0.1^a	5.0 ± 0.8	4.7 ± 0.5
Coef. Sedimentação [S]	$\sim 58^b$	50 ± 1	50 ± 1
Raio de Stokes [Å]	138 ± 6^c	141 ± 4	143 ± 3
Coef. Difusão [10^{-7} cm ² /s]	1.56 ± 0.05^c	1.4 ± 0.2	1.5 ± 0.4
f/fmin	1.4 ± 0.2^c	1.35 ± 0.09	1.38 ± 0.08

^ados experimentos de SAXS.

^bde Costa et al. 1988.

^ccalculado a partir do coeficiente de sedimentação.

d) Considerações sobre os resultados com detetor unidimensional

Utilizamos a técnica de SAXS para obter modelos de baixa resolução da estrutura quaternária da hemoglobina extracelular *G. paulistus*. Devido a dificuldade inerente de crescer-se cristais de proteínas com elevado peso molecular como a *G. paulistus*, nossos resultados mostraram que a técnica de SAXS aliada a modelagem *ab initio* permite obter-se resultados confiáveis sobre a forma e arranjo quaternário de proteínas gigantes. Como visto no caso anterior para a proteína *B. glabrata*, cálculos hidrodinâmicos são um complemento importante para os resultados de SAXS e, como vemos na Tabela 11, temos uma boa concordância entre os valores experimentais e os preditos pelos modelos *ab initio*. A grande diferença no volume já foi mencionada antes para o caso de modelos tipo *dummy atom* e

refere-se ao fato de que neste tipo de modelos a superfície tende a ser mais suave do que nas proteínas reais, alterando o volume final.

Tomando-se por base o modelo de SAXS obtido bem como a medida do peso molecular também fornecido dos experimentos de espalhamento, pode-se propor para a proteína um conjunto de doze domínios idênticos em formato de dois anéis hexagonais sobrepostos. Cada domínio possuindo 260kDa, em concordância com a estrutura proposta por Silva et al. (1989) utilizando dissociação por pressão hidrostática. A comparação entre o modelo obtido e a estrutura cristalográfica para a *L. terrestris* aumenta a credibilidade do modelo final. O fator crucial para a correta modelagem foi o conhecimento prévio da simetria global (no caso P62) da proteína.

e) Medidas em alta resolução para proteína *G. paulistus*

Os resultados mencionados acima foram obtidos utilizando-se para a coleta dos dados um detector linear, compreendendo um intervalo de $0.005 < q < 0.1882 \text{ \AA}^{-1}$. Neste experimento atingimos uma resolução de $\sim 34 \text{ \AA}$ e um total de ~ 18 Shannon Channels. Repetimos estes experimentos utilizando-se o detector bidimensional para compararmos os resultados. Como mencionamos anteriormente, detectores bidimensionais permitem uma melhor descrição da região a altos ângulos. Utilizamos concentrações que variaram de 5 a 20mg/ml. Três distâncias foram utilizadas nos experimentos: 1590mm, 682mm e 395mm. As curvas foram tratadas pelo programa TRAT2D (Oliveira, 2004) bem como a integração angular foi feita usando este programa.

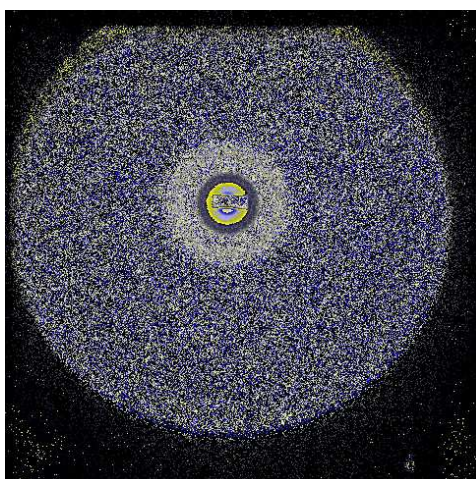


Figura 43 – Espectro Bidimensional de espalhamento para *G. paulistus* a 20mg/ml.

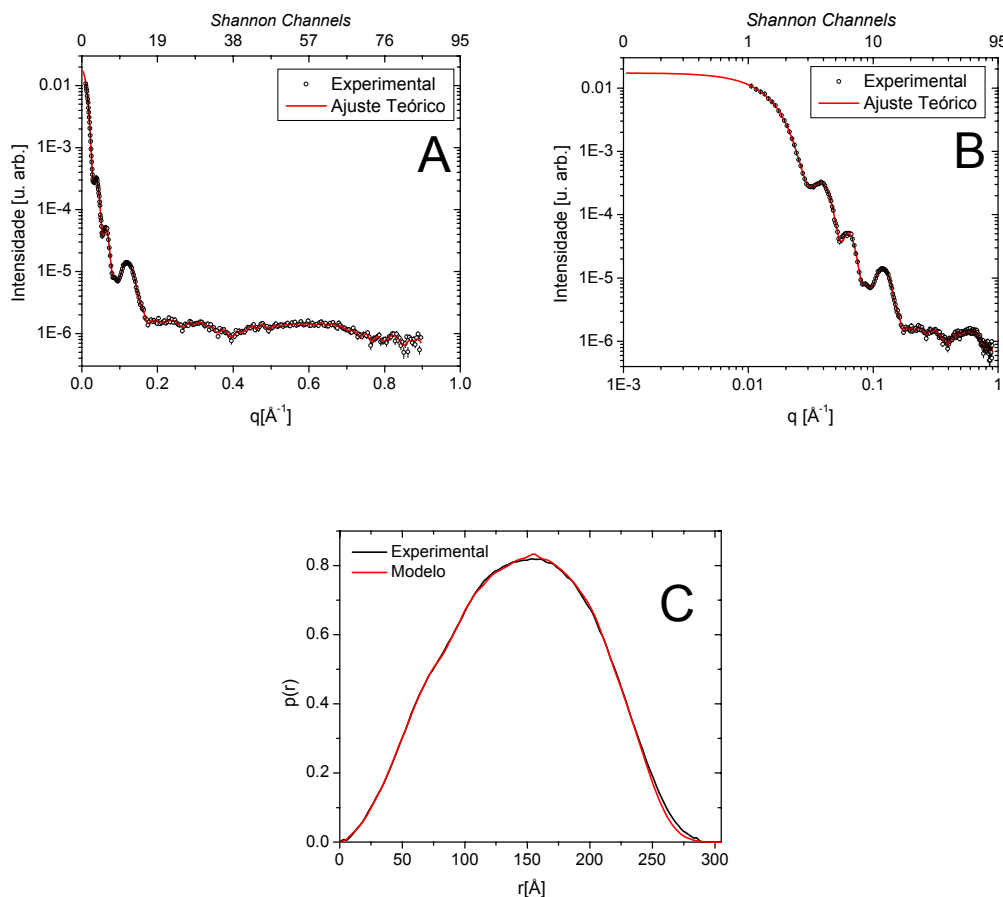


Figura 44 – Resultados experimentais para *G. paulistus*. A e B) Intensidade de espalhamento (em dois gráficos com escalas diferentes) e ajuste teórico. C) Função $p(r)$ calculada.

Como vemos na Figura 44, temos uma curva experimental muito rica, uma vez que a proteína é muito grande ($D_{\text{max}} \sim 300 \text{\AA}$) e chegamos a um $q_{\text{max}} \sim 0.9$, o que significa ~ 84 Shannon Channels. Os programas disponíveis para a construção de modelos não conseguem dar conta de uma curva deste tipo ainda uma vez que a proteína completa possuirá um total de ~ 28000 resíduos, impedindo uma modelagem por modelos tipo *chain model*, cada resíduo representando um aminoácido da proteína. Desta maneira, utilizamos a metodologia descrita na parte *Experimental*, representando um grupo de resíduos por um componente da *dummy chain*. Fazendo com que cada componente da cadeia represente um total de ~ 10 resíduos da proteína, o número de componentes da cadeia foi de 2820, impondo-se simetria P62. Realizando o ajuste dos dados experimentais obtemos o modelo mostrado abaixo (Figura 45). Vemos do ajuste dos dados experimentais (Figura 44C) que o

modelo descreve corretamente a forma da proteína mesmo com a utilização de pseudo resíduos com raios maiores. Além disso, as características principais dos modelos *ab initio* anteriormente mostrados são obtidas também neste modelo.

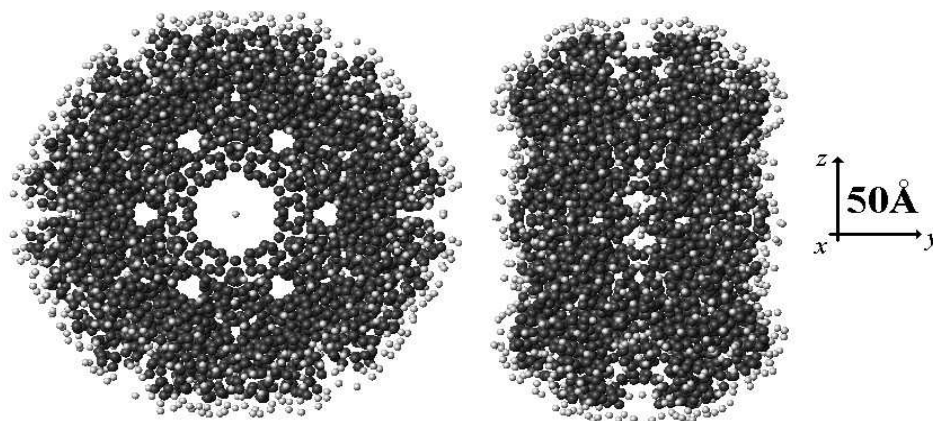


Figura 45 – Modelo *ab initio* tipo *chain model* para proteína de *G. paulistus*.

Para comparação calculamos as propriedades hidrodinâmicas (programa HYDROPRO) deste modelo e as apresentamos na Tabela 12. Comparando-se com os resultados dos modelos tipo *dummy atom* (Tabela 11) vemos que alguns parâmetros são melhor descritos por este modelo, o que é esperado pois como mostrado anteriormente nesta tese, modelos tipo *chain model*, mesmo que modificados para descreverem mais de um resíduo por subunidade, fornecem valores hidrodinâmicos mais precisos do que modelos tipo *dummy atom*.

Tabela 12 – Parâmetros Hidrodinâmicos e peso molecular para hemoglobina de *G. Paulistus* obtida por SAXS e calculada pelos modelos *ab initio*. Modelo 3: Modelo tipo *chain model*.

Parâmetro	Valor Experimental	Modelo 3
Mw [MDa]	3.1 ± 0.2^a	---
Volume [10^6 Å]	3.8 ± 0.1^a	6.5
Coef. Sedimentação [S]	$\sim 58^b$	56
Raio de Stokes [Å]	138 ± 6^c	135
Coef. Difusão [10^{-7} cm ² /s]	1.56 ± 0.05^c	1.59
f/fmin	1.4 ± 0.2^c	1.2

^ados experimentos de SAXS.

^bde Costa et al. 1988.

^ccalculado a partir do coeficiente de sedimentação.

Como a proteína é muito grande é de se esperar que o volume do modelo seja maior do que o valor real da proteína. Isso ocorre porque como cada subunidade da dummy chain representa ~10 resíduos da proteína real teremos o modelo ab initio inflado com relação ao volume real da proteína. Para proteínas menores este efeito diminui (veja em *Aspectos Experimentais*, o item 2.4).

2. Estudo da proteína Básica da membrana de Mielina por SAXS^{§§}

a) Introdução

A bainha de mielina no sistema nervoso central é uma membrana multilamelar rica em lipídeos fortemente ligada ao redor do axônio nervoso (Moscarello, 1996). A proteína básica de mielina (*myelin basic protein - MBP*) é a segunda proteína em maior número na mielina e talvez o seu componente mais estudado. A MBP é importante por duas razões principais: 1) acredita-se que seja a “cola” biológica pela qual se processa a formação, compactação e manutenção da estrutura multilamelar da mielina (Riccio et al., 1996); e 2), é candidato a autoantígeno no contexto de pesquisas em esclerose múltipla uma vez que pode induzir Encelafomielite Alérgica, um modelo animal de esclerose múltipla. Até o presente momento, além da sequência de aminoácidos e um modelo estrutural de uma isoforma com 18.5kDa (Ridsdale et al., 1997), muito pouco se sabe sobre sua estrutura e conformação tridimensional. Uma razão disso decorre de a MBP ser usualmente extraída sobre condições extremamente drásticas em uma conformação livre de lipídeos (*lipid free MBP – LF-MBP*), solúvel em água, e parcialmente desenhovelada. Na realidade, a maioria dos estudos da forma LF-MBP em soluções aquosas estão em concordância com conformações flexíveis para esta proteína. No entanto, a interação da MBP com detergentes, lipídeos e outras moléculas pode induzir uma estrutura mais ordenada (Haas et al., 1998, Polverini et al., 1999) além do fato de interações proteína-lipídeo serem críticas para a estabilidade da bainha de mielina.

^{§§} Os resultados aqui apresentados formam parte da publicação Haas H, **Oliveira CLP**, Torriani IL, Polverini E., Fasano A, Carlone G., Cavatorta P and Riccio P (2004) “Small Angle X-Ray Scattering from Lipid-Bound Myelin Basic Protein in Solution”, *Biophysical Journal*, Vol. 86, 1-6.

Uma outra maneira de se extrair MBP da mielina é conseguida utilizando-se detergentes moderados para manter a proteína em seu entorno lipídico natural durante o processo de purificação (Riccio et al., 1984, 1994). Esta forma é chamada MBP ligada à lipídeos (*lipid-bound* MBP – LB-MBP). A LB-MBP mostrou ser diferente em vários aspectos, principalmente imunológicos, da forma livre de lipídeos (Mazaanti et al., 1998). Por experimentos espectroscópicos mostrou-se que a LB-MBP possui uma proporção de estrutura secundária ordenada muito maior do que a forma LF-MBP, mesmo após a adição de detergentes e lipídeos a esta forma (Polverini et al., 1999). Tomando-se a LB-MBP e lipídeos, a auto-organização de membranas tipo mielina estáveis podem ser induzidas em condições nas quais os lipídeos sozinhos manteriam-se pouco organizados (Riccio et al., 1986, 2000).

Neste estudo, para obter informações relevantes sobre as estruturas das formas livre de lipídeos (LF-MBP) e ligada a lipídeos (LB-MBP), realizamos estudos de espalhamento de raios X a baixos ângulos (SAXS) em soluções contendo estas proteínas. Como veremos (Haas e Oliveira et al., 2004), os experimentos revelaram diferenças consideráveis entre as duas isoformas de proteínas. Complementarmente, a utilização do molelo cristalográfico teórico e cálculos *ab initio* para a forma LB-MBP permitiram a obtenção diversas informações sobre a estrutura tridimensional nativa desta proteína.

b) Experimental e Metodologia

As metodologias para obtenção das formas LF-MBP e LB-MBP do cérebro bovino são descritas em diversos trabalhos (LF: Deibler et al., 1972, 1984; LB: Riccio et al., 1984, 1994). O procedimento fornece soluções com baixa concentração tanto para as formas LF e LB.

Nossos colaboradores neste trabalho obtiveram por dinâmica molecular um modelo estrutural para a MBP bovina, utilizando como base o modelo estrutural obtido para a MBP humana por Ridsdale (Ridsdale et al., 1997), utilizando as coordenadas atômicas disponíveis no banco de dados *Protein Data Bank*, com código 1QCL. Os procedimentos de otimização e simulação do modelo teórico são descritos em outros trabalhos (Haas e Oliveira et al., 2004, Polverini et al., 1999), mas ao final um modelo teórico para a estrutura da MBP foi obtido e pôde ser comprovado através dos dados de SAXS.

Os experimentos de SAXS foram realizados no Laboratório Nacional de Luz Síncrotron (LNLS). O comprimento de onda utilizado foi de 1.608Å. O porta-amostras utilizado foi o de janelas de mica separadas à distância de 1mm para colocar-se a amostra. O sistema possuía temperatura controlada à 15°C. A detecção da intensidade espalhada foi feita utilizando um detetor linear sensível à posição. Duas distâncias amostra-detetor foram utilizadas, 1600 e 525mm. Amostras de proteína, contendo detergente e solução tampão foram preparadas pelo grupo do professor Paolo Riccio no *Dipartimento di Biochimica e Biologia Molecolare, University of Bari, Bari, Italy*. As amostras foram liofilizadas e transportadas à temperatura ambiente da Itália ao Brasil. Imediatamente antes dos experimentos de SAXS, água filtrada Milli-Q (Millipore, Bedford, MA) foi adicionada às amostras liofilizadas para fornecer as concentrações desejadas de proteína, e esta solução estoque foi utilizada nas medidas. Medidas do espalhamento parasita (buffer) foram feitas com soluções contendo todos os aditivos utilizados nas medidas com proteínas. Os experimentos com LB-MBP foram feitos à concentração de proteína de 1.28 mg/ml em um buffer HEPES 20mM, pH7.5, com 0.8% CHAPS. As medidas de LF-MBP foram feitas com concentrações de proteína de 3.2 mg/ml em buffer HEPES 10mM pH7.5 com 0.4% CHAPS.

Todas as soluções foram agitadas com um vortex diretamente antes dos experimentos. As amostras foram mantidas à temperatura de 15°C e exposições de 10min foram realizadas. Em alguns casos, várias exposições foram feitas para melhorar a qualidade dos dados. O tratamento dos dados foi feito utilizando-se o programa TRAT1D.

c) Resultados e Discussão

- Espalhamento de raios X a baixos ângulos para as formas LF-MBP e LB-MBP

As intensidades de espalhamento para as proteínas LB-MBP e LF-MBP são mostradas na Figura 46. Como pode ser visto por inspeção direta, as duas intensidades de espalhamento são claramente diferentes. Para a amostra LB-MBP obteve-se dados com boa relação sinal/ruído, apresentando uma modulação na curva de espalhamento típica de partículas globulares. O intervalo medido foi $0.02302 < q < 0.2549\text{Å}^{-1}$, sendo q o módulo do vetor transferência de momento já descrito anteriormente. Em contraste, as curvas para LF-

MBP não mostraram esta modulação. Além disso, o nível de ruído é maior, tornando a região útil menor do que para LB-MBP.

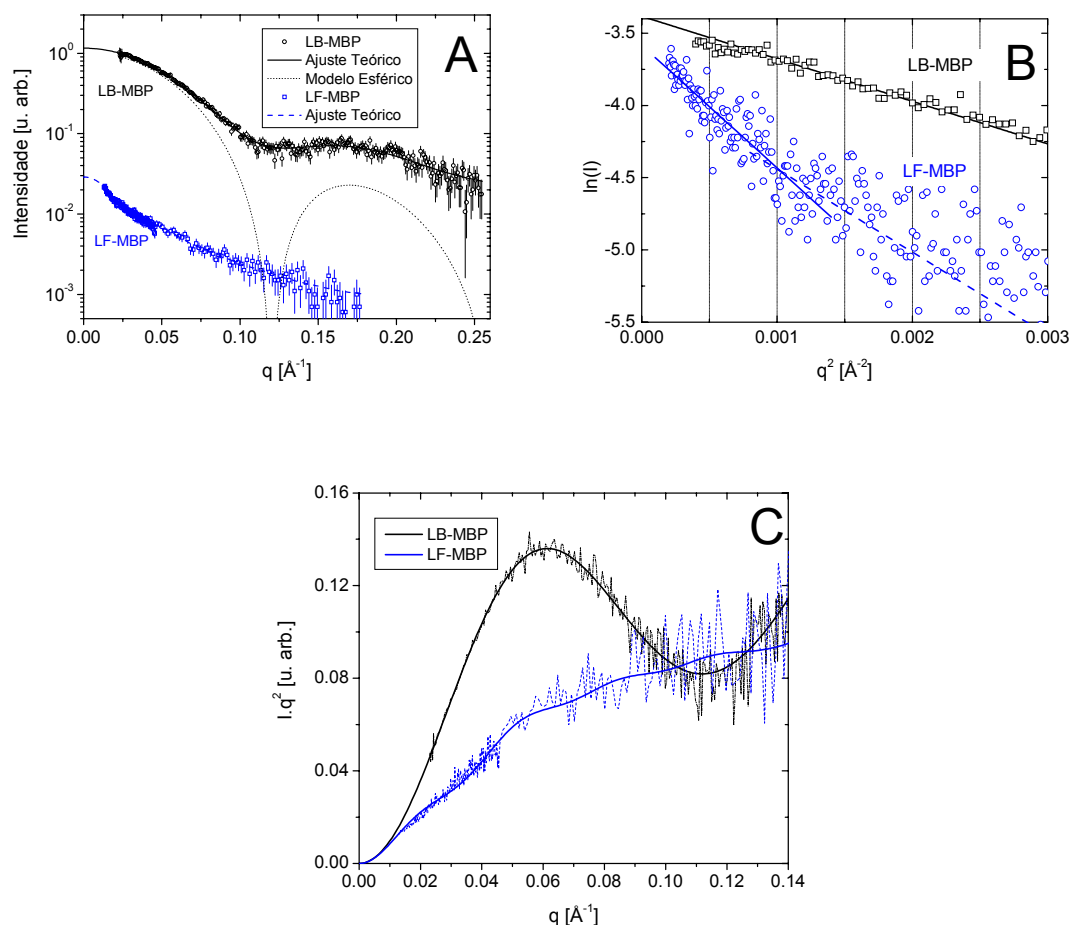


Figura 46 – Curvas de SAXS para proteína básica de mielina nas formas LB-MBP e LF-MBP. A) Intensidade de espalhamento para MBP na forma com lipídeos LB-MBP e livre de lipídeos LF-MBP. A intensidade LF-MBP foi dividida por um fator constante para facilitar a visualização. B) Gráfico de Guinier para LB-MBP e LF-MBP. C) Gráfico de Kratky para LB-MBP e LF-MBP.

As diferenças entre as curvas de espalhamento tornam-se evidentes quando analisamos a região a baixos valores de q , ou seja, a região de Guinier no gráfico de Guinier ($\ln(I) \times q^2$). Conforme mostrado anteriormente, para um sistema monodisperso de partículas o raio de giro R_g será determinado pela inclinação da curva de intensidade neste gráfico. A Figura 46B mostra o gráfico de Guinier para os dois tipos de proteínas. Para as medidas com LB-MBP (curva superior), obtém-se uma reta, indicando que, na realidade, o sistema em solução é monodisperso, com uma forma globular bem definida. Da regressão

linear da região de q^2 de 0.001 a 0.002 $\text{\AA}^{-2}$ obtém-se o valor de 29.7 $\text{\AA}$ para o raio de giro. A aproximação linear é mostrada na figura.

Para a LF-MBP (curva inferior), os dados são menos claros. Neste caso o ruído experimental é maior em decorrência da menor intensidade espalhada. A inclinação da curva é maior do que no caso anterior, não sendo constante na região analisada. Desta forma, não se pode associar uma estrutura globular a esta amostra. A região com melhor linearidade ocorre em baixos valores de q^2 , $\sim 0.0005 \text{\AA}^{-2}$. Da regressão desta região (linha sólida), um raio de giro da ordem de 50 $\text{\AA}$ é determinada. A valores maiores de q^2 ($\sim 0.015 \text{\AA}^{-2}$), a inclinação é menor, correspondendo ao valor de 42 $\text{\AA}$. Estes resultados em alguns aspectos são similares aos obtidos anteriormente em outros estudos da proteína LF-MBP por SAXS, com condições experimentais um pouco diferentes (sem detergente), com valores de 46 $\text{\AA}$ (Krigbaum and Iisu, 1975) e 39 $\text{\AA}$ (Erand et al., 1974) para o raio de giro.

Outra análise que pode ser realizada consiste no grau de compactação da proteína espalhada. Como mostramos anteriormente, através do *Kratky plot* ($I \cdot q^2 \times q$) pode-se ter informação da compactidade da proteína em solução. Na Figura 46C temos os *Gráficos de Kratky* para as duas proteínas e vemos claramente que, ao passo que a proteína LB-MBP apresenta uma conformação globular compacta (indicada pela forma de sino, porém com alguma flexibilidade uma vez que a curva não decai a zero) a proteína LF-MBP apresenta uma conformação típica de *random coil*, indicando novamente que os lipídeos ligados à forma LB-MBP mantém a compactidade do sistema protéico. Desta maneira, os dados de SAXS indicam diferenças claras entre as duas formas da proteína MBP.

- Modelagem com perfil de densidade eletrônica

Além de caracterizar as diferenças entre as duas formas, tivemos como objetivo obter informações sobre a estrutura da proteína LB-MBP. Devido ao fato desta proteína ser monodispersa e compacta, podemos utilizar o formalismo de sistemas monodispersos para prosseguir na análise dos dados experimentais. Como mostramos anteriormente na *Introdução Teórica* desta tese a transformada de Fourier da intensidade de espalhamento fornece a curva de distribuição de pares de distâncias $p(r)$, a qual provê indicações sobre a forma da partícula. Na Figura 47A temos a função $p(r)$ calculada a partir do ajuste dos dados experimentais com o programa GNOM. Vê-se a presença de um máximo em $\sim 40 \text{\AA}$ e

um ombro em $\sim 18\text{\AA}$, bem como a máxima dimensão da partícula sendo em torno de $80\text{\AA}$. A forma da função $p(r)$ indica a presença de regiões ocas no interior da proteína ou, melhor dizendo, uma variação da densidade eletrônica da partícula no sentido radial (compare com Figura 12). Por exemplo, um arranjo concêntrico de esferas com diferentes valores de densidade eletrônica fornece funções $p(r)$ deste tipo e, na realidade, este pode ser um modelo inicial para a LB-MBP uma vez que é esperada uma camada de lipídeos ou detergentes ao redor da proteína.

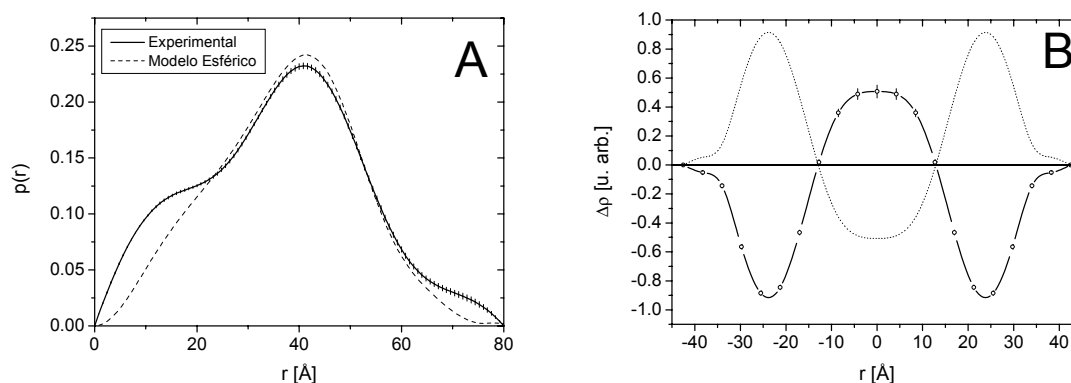


Figura 47 – Modelagem da LB-MBP por perfil de densidade eletrônica. A) Função de distribuição de pares de distâncias $p(r)$ (curva contínua) e ajuste com modelo esférico (curva tracejada). B) Perfil de densidade eletrônica obtido para modelo esférico.

Como primeiro teste utilizamos o programa DECON (Glatter, 1981) para tentar ajustar os dados experimentais com um modelo de partícula esférica com densidade eletrônica radial variável. Como resultado tem-se o perfil de densidade eletrônica que melhor ajusta os dados experimentais. Na Figura 47B mostramos o perfil de densidade eletrônica do ajuste e nas figuras Figura 46A e Figura 47A o ajuste da intensidade e da função $p(r)$ correspondente a este perfil. O perfil de densidade eletrônica é calculado utilizando 10 intervalos equidistantes (camadas concêntricas com mesma espessura) com densidade eletrônica constante em cada camada. Podemos interpretar esse resultado em termos de uma esfera interna com diâmetro da ordem de $30\text{\AA}$, onde a densidade eletrônica é maior do que a água, circundado por uma camada de $\sim 15\text{\AA}$ de espessura, com densidade eletrônica menor do que a da água (um perfil com sinais invertidos, como o mostrado em linha tracejada na Figura 47B, fornece o mesmo ajuste). Esta conformação poderia indicar um tipo de organização da proteína (que possui alta densidade eletrônica) circundada por

uma camada de lipídeos ou detergentes (com densidade eletrônica menor devido as cadeias de hidrocarboneto). Também, as dimensões são da mesma ordem dos valores esperados. Para a proteína MBP o volume, que pode ser estimado com base no volume parcial molar dos aminoácidos de sua cadeia (Perkins, 1988), assume o valor de $21000\text{\AA}^3$, o que corresponde a uma esfera compacta com raio $17\text{\AA}$. Similarmente, as dimensões de sua camada externa são esperadas para um envelope lipídico. Com este modelo, alguns aspectos da distribuição radial de densidade podem ser obtidos. No entanto, com este formalismo somente estruturas centro-simétricas podem ser corretamente representadas. As claras diferenças entre os ajustes para com os dados experimentais indicam que a estrutura da proteína deva diferir significativamente de um arranjo centro-simétrico. Desta forma utilizamos modelagem *ab initio* com o objetivo de obter novas informações.

- Modelagem *ab initio* e comparação com modelo teórico

Para os cálculos *ab initio* utilizamos modelos tipo *chain model*. Testes com modelos tipo *dummy atom*, forneceram resultados inconsistentes uma vez que em muitos casos domínios separados eram obtidos. Isso acontece devido à forma peculiar da proteína, como indicado pela função $p(r)$ e sendo assim, modelos com vínculo de contigüidade, como no caso de modelos baseados em *chain models*, impedirão este artefato de cálculo. Como os dados são de baixa resolução e informação, ($2\pi/q_{\text{max}} \sim 25\text{\AA}$, 6.5 *Shannon channels*) um número grande de configurações diferentes pode ajustar com a mesma qualidade os dados experimentais e, sendo assim, realizamos a média espacial dos modelos visando determinar o modelo mais provável.

Para a correta interpretação dos resultados deve-se ter em mente as limitações inerentes ao modelo. Como a proteína é estabilizada pela presença dos lipídeos (provavelmente posicionados em seu entorno) a aproximação da mesma por uma estrutura com densidade eletrônica constante (que é o caso de um modelo *chain model*) não fornecerá a estrutura da proteína em sim, mas sim, uma estrutura média que represente tanto a proteína quanto o entorno lipídico. Desta forma é de se esperar que o volume do modelo obtido seja *maior* do que a proteína real.

Na Figura 48 mostro o modelo mais provável obtido pela média dos modelos *ab initio* e também um dos ajustes. Vemos que a modelagem *ab initio* forneceu uma

conformação muito interessante, assumindo uma forma em “U” para a proteína estabilizada pelos lipídeos.

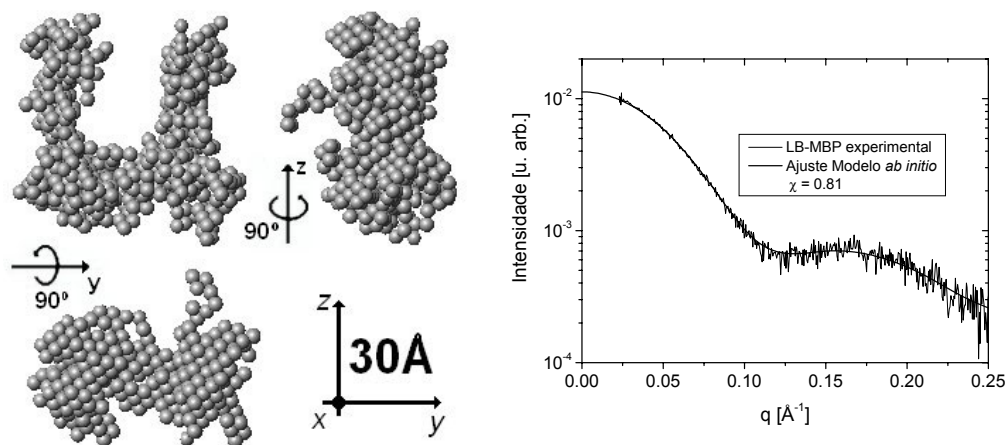


Figura 48 – Modelagem da LB-MBP por chain models. Modelo *ab initio* mais provável calculado a partir dos dados de SAXS e um dos ajustes.

Esse resultado está em ótima concordância com o modelo teórico construído para a proteína MBP bovina. Utilizando as coordenadas dos átomos do modelo obtido por dinâmica molecular, realizamos o ajuste dos dados experimentais com base na estrutura cristalográfica, utilizando o programa CRY SOL, já descrito anteriormente. O ajuste dos dados, considerando valores padrão para o volume excluído por resíduo, é mostrado na Figura 49 (curva tracejada). Vemos que, apesar do formato geral da curva de espalhamento ser o mesmo, o ajuste dos dados é ruim, indicando que o modelo se refere à uma estrutura *menor* do que àquela vista pelos dados de espalhamento. Para poder levar em conta a influência dos lipídeos, realizamos outro ajuste permitindo que o programa assumisse valores de volume excluído por resíduo, *maiores* do que os valores padrão. Isso teve o objetivo de mimetizar a presença dos lipídeos que, como argumentamos antes, poderiam estar inflando a proteína. O ajuste dos dados com esta nova consideração é mostrada Figura 49 (curva sólida).

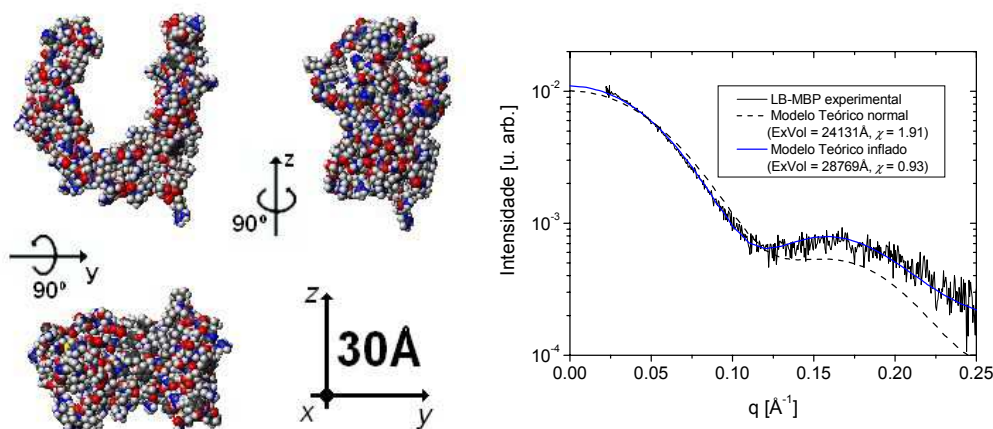


Figura 49 – Comparação do espalhamento da LB-MBP com modelo atômico teórico. Modelo teórico calculado por homologia com MBP humana (esquerda) e o correspondente ajuste dos dados experimentais com variação do volume excluído da proteína (direita). Veja o texto para detalhes.

Temos agora uma melhora significativa da qualidade do ajuste dos dados experimentais, indicando que para os dados de espalhamento a baixos ângulos, o efeito dos lipídeos no espalhamento foi o de aumentar o volume excluído total da proteína. Sendo assim, temos a confirmação de que o modelo teórico obtido por homologia com a MBP humana descreve muito bem os dados experimentais, ao menos ao nível de resolução que tivemos nestes experimentos. Outro fato interessante é o de que, pela modelagem *ab initio*, conseguimos recuperar corretamente esta conformação em “U” da proteína como mostrado pela superposição do modelo *ab initio* com o modelo teórico na Figura 50.

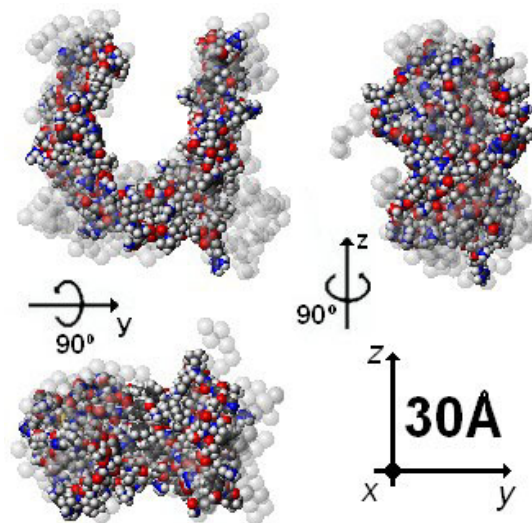


Figura 50 – Superposição do modelo *ab initio* ao modelo teórico.

Os modelos obtidos permitem-nos compreender a presença do ombro característico de estruturas ocas presente na função $p(r)$. Como o interior da proteína é “aberto” temos a presença de uma modulação (solvente – proteína – solvente – proteína – solvente) de modo a ocasionar esta característica na função $p(r)$. Desta forma, o resultado obtido pela aproximação por esferas concêntricas não estava inteiramente errôneo uma vez que os tamanhos característicos (Figura 51) foram corretamente estimados por aquela aproximação. No entanto, em decorrência da partícula não ser esférica, esta aproximação é muito grosseira para descrever corretamente os dados experimentais.

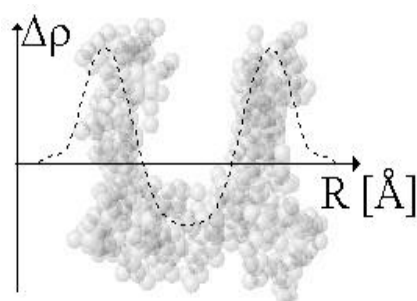


Figura 51 – *Superposição do modelo ab initio ao perfil de densidade eletrônica obtido pelo ajuste com modelo de esferas concêntricas.*

d) Considerações finais

O principal intuito deste trabalho foi o de demonstrar a importância de se estudar a proteína MBP extraída e purificada em um meio contendo lipídeos (LB-MBP) em comparação à forma livre de lipídeos (LF-MBP). Obtivemos também informações originais sobre a estrutura e conformação da proteína MBP em solução com seu entorno lipídico bem como modelos tridimensionais para esta estrutura através de medidas de espalhamento de raios X a baixos ângulos (Haas e Oliveira et al, 2004). Uma das principais conclusões que obtivemos aqui foi a de que a proteína MBP, extraída e purificada em seu entorno lipídico pelo procedimento proposto por Riccio (Riccio et al. 1984, 1994), mantém suas características nativas quando na membrana de mielina, permitindo assim estudar sua estrutura e função. Sendo assim, para esta proteína, o isolamento direto do complexo proteína-lipídeo é mais apropriado do que tentar-se renaturar a proteína na forma livre de lipídeos, denaturada, pela exposição da mesma a um meio lipídico.

Dos dados sobre a conformação da MBP em seu meio lipídico nativo tivemos a indicativa de uma estrutura com alta compacidade, não-esférica, de proteína+lipídeos. Da modelagem ab initio para a LB-MBP obtivemos uma conformação em “U”, reiterando o modelo teórico obtido através de dinâmica molecular. Além disso, foi possível mostrar que este modelo consegue descrever os dados experimentais se permitirmos que assuma, em solução juntamente com os lipídeos, um volume excluído maior do que o dado pela sua cadeia de aminoácidos, o que representaria a influência do entorno lipídico da proteína. Estes resultados demonstram também que a técnica de SAXS permite um estudo preciso deste tipo de sistema, isto é, sistemas de proteínas complexadas à lipídeos.

3. Estudo da interação da lisozima com polietilenos e copolímeros^{***}

a) Introdução

Óxidos Polietilenos (*poly(ethylene oxide)* - PEO) e copolímeros são muito utilizados em fórmulas biomédicas tanto pela sua alta solubilidade quanto pela sua biocompatibilidade. Uma das aplicações mais proeminentes se relaciona com a inibição da adsorção de proteínas em superfícies, aumentando sua biocompatibilidade. Razões similares levaram ao desenvolvimento do que é atualmente chamado “*pegylation*”, isto é, ligação química entre pequenos óxidos polietilenos, também conhecidos como polietileno glicols, PEG, à proteínas gerando bioconjugados ativos com vidas médias consideráveis (Abuchowski et al., 1977).

A natureza das interações físico-químicas entre PEO e proteínas ainda é uma questão em aberto. A propriedade de inibição da adsorção de proteínas sobre superfícies cobertas com PEO foi associada à um impedimento estérico, diretamente relacionado à conformação das cadeias de PEO (Harder et al., 1998), de modo que quanto maior for sua massa molar, menos efetivas são as forças de atração tipo van der Waals entre a proteína e a superfície e, assim, menor é a extensão da adsorção. No entanto, forças atrativas foram

^{***} Os resultados aqui apresentados formam parte da publicação Almeida NL, **Oliveira CLP**, Torriani IL and Loh W (2004). “Calorimetric and structural investigation of the interaction of lysozyme and bovine serum albumin with poly(ethylene oxide) and its copolymers”. Coll. And Surf. B: Biointerfaces, 38, 67-76.

detectadas entre PEO e proteínas em alguns casos particulares; especificamente foram detectadas forças de interação em superfícies cobertas com PEO de altas massas molares e submetidos à pressão; ou por PEO pouco hidratado à altas temperaturas, em aparelhos de medida para força em superfícies. Também existem trabalhos indicando a formação de complexos em soluções de PEO e albumina humana, baseados em evidências de cromatografia e medidas de espalhamento (Azegami et al., 2004).

Recentemente, artigos de Topchieva e colaboradores, apresentaram evidências da formação de complexos não covalentes entre alfa-chimotripsina e PEO (Topchieva et al., 2000), ou copolímeros óxido etileno/propileno (Topchieva et al., 1994). Em alguns casos, a formação dos complexos foi aumentada por centrifugação das misturas, o que foi associado à redução da solvatação do PEO quando submetido à pressão (Topchieva et al., 1994). Uma das evidências apresentadas para a formação do complexo incluiu um pequeno aumento da estabilidade térmica da proteína quando da adição do polímero, como indicado por medidas de calorimetria diferencial (DSC) e por cromatografia de afinidade com diferentes tempos de eluição. Por outro lado, um estudo anterior em misturas de proteínas (urease e interleuquina 2) com copolímeros F127EO/PO reportaram nenhuma mudança na estabilidade térmica da proteína (Wang et al., 1993).

Esta contradição e também a relevância do assunto nos motivou a executar este trabalho (Almeida e Oliveira et al., 2004). Utilizando o espalhamento de raios X a baixos ângulos procuramos detectar a formação dos complexos proteína-polímero para diferentes tipos de polímeros, correlacionando estes dados com informações de estabilidade térmica e medidas de cromatografia.

Neste estudo mostramos resultados de misturas de lisozima com PEO de diferentes massas molares e EO/PO copolímeros tribloco de variadas composições. Os polímeros utilizados neste estudo possuem sua estrutura química descritas na Tabela 13. O copolímero P105 representa um polímero com um grupo central PO hidrofóbico ligado a dois grupos EO hidrofílicos. O uso de copolímeros bloco EO-PO-EO se relaciona com a propriedade destes polímeros de promover auto-organização em solução (Glatter et. al., 1994). Esta agregação pode ser obtida pelo aumento de concentração, ocorrendo acima da concentração micelar crítica (c. m. c.) ou induzida pelo aumento da temperatura, com o limiar de micelação dado pela temperatura de micelação crítica (c. m. t.). Pelos resultados que

mostraremos não existe interação específica entre os polímeros e as proteínas. No entanto interações não específicas foram detectadas, bem como influências em processos de agregação e formação de micelas.

b) Materiais e Métodos

A amostra de lisozima foi obtida junto a Sigma, USA. Os polímeros e copolímeros utilizados são descritos na Tabela 13. Água de grau Mili Q foi utilizado nos experimentos. Soluções tampão a pH 7.0 foi preparada com *buffer* fosfato.

Tabela 13 – Descrição dos polímeros utilizados neste estudo.

Polímero	Fonte	M* / g.mol ⁻¹	Unidades de EO*	Unidades de PO*	c.m.t [†] (solução 2.5 %) / °C
PEO	Sigma (U.S.A.)	3350	76	-	-
F 38	ICI Surf., UK	4800	2 x 44	17	60
P 105	ICI Surf., UK	6500	2 x 37	58	18

* valores médios calculados pela massa molar nominal.

† temperatura crítica de micelação determinada.

A preparação das amostras foi realizada pelo grupo do professor Watson Loh no Instituto de Química da Universidade Estadual de Campinas, sendo transportadas à temperatura ambiente para o LNLS. As medidas de espalhamento de raios X a baixos ângulos foram realizadas na linha de SAXS do Laboratório Nacional de Luz Síncrotron (LNLS). O comprimento de onda utilizado nos experimentos foi de 1.488Å com distância amostra-detetor de 745mm. As medidas foram realizadas em um porta-amostra de capilar (Cavalcanti et al, 2004) com tubo de diâmetro 1.5mm. As amostras de lisozima estavam à concentração de 10mg/ml e pH 7.0 (*buffer* fosfato). Para amostras contendo polímeros, sua concentração foi de 2.5wt.%. Algumas amostras foram submetidas à alta pressão (25GPa) e centrifugação cerca de 30-60 minutos antes das medidas de SAXS, com metodologia descrita previamente (Almeida e Oliveira et al, 2004).

Inicialmente realizamos estudos à temperatura ambiente para amostras sem tratamento e também submetidas à pressão e centrifugação. Posteriormente para amostras sem tratamento fizemos um estudo variando a temperatura até o limiar de denaturação da proteína. O processo de denaturação seguido da formação de agregados protéicos foi acompanhado pelos experimentos de espalhamento.

- *Tratamento dos dados*

Os dados foram tratados com o programa TRAT1D. Como temos um sistema composto de três elementos (polímero, proteína e solvente) devemos considerar a melhor maneira de tratar os dados. Estamos procurando possíveis interações entre proteína e polímero utilizando espalhamento de raios X a baixos ângulos. Estas interações, em princípio, podem ser desde interações interpartículas até ligações proteína/polímero. Se não ocorrer nenhum tipo de interação ou complexação a intensidade total será simplesmente,

$$I(q) = \sum_i I_i(q) = I_{proteína}(q) + I_{polímero}(q) + I_{Solvente}(q)$$

ou seja, a soma das intensidades espalhadas por cada uma das possíveis fontes de espalhamento. Havendo algum tipo de interação, essa relação não será mais válida e isso nos dará a indicativa de algum tipo de complexação. Assim para detectar estas possíveis variações procederemos da seguinte maneira:

- Medimos o espalhamento da proteína padrão (lisozima) sem nenhum polímero, diluída no solvente;
- Medimos o espalhamento para os polímeros utilizados diluídos no solvente;
- Medimos o espalhamento do solvente;
- Realizamos experimentos com misturas proteína + polímero e deste espalhamento subtraímos o espalhamento do polímero. Caso não haja nenhum tipo de interação, o espalhamento restante deverá ser idêntico ao da lisozima inicial. Caso contrário, algum tipo de interação ocorreu. Nestes casos subtraímos somente o espalhamento do solvente e tentamos analisar o espalhamento do sistema binário proteína+polímero.

c) Aspectos Teóricos da Análise de SAXS

Como mostramos anteriormente a intensidade espalhada por um sistema monodisperso de partículas é dada por,

$$I(q) = N \cdot (\Delta\rho)^2 \cdot V^2 \cdot P(q) \cdot S(q)$$

sendo N o número de partículas, $\Delta\rho$ o contraste de densidade de espalhamento, V o volume da partícula, $P(q)$ o fator de forma e $S(q)$ o fator de estrutura. Se tivermos um sistema

composto por partículas de diferentes fatores de forma, porém monodisperso em cada tipo, a intensidade total espalhada por este sistema será dada por:

$$I(q) = \left[\sum_i N_i \cdot (\Delta \rho_i)^2 \cdot V_i^2 \cdot P_i(q) \right] S_{1,2,...,i,...}(q) \quad (144)$$

com o índice i designando cada tipo de partícula. A componente $S_{1,2,...,i,...}$ expressa a possibilidade de uma possível interação partícula-partícula que pode ser dada em várias formas (interação eletrostática, hidrofóbica etc.). No entanto, se considerarmos sistemas diluídos, podemos assumir como primeira aproximação que $S_{1,2,...,i,...}(q) \approx 1$ e assim temos

$$\begin{aligned} I(q) &= \sum_i N_i \cdot (\Delta \rho_i)^2 \cdot V_i^2 \cdot P_i(q) \\ I(q) &= N_1 \cdot (\Delta \rho_1)^2 \cdot V_1^2 \cdot P_1(q) + N_2 \cdot (\Delta \rho_2)^2 \cdot V_2^2 \cdot P_2(q) + \dots \\ I(q) &= I_1(q) + I_2(q) + \dots \\ I(q) &= \sum_i I_i(q) \end{aligned} \quad (145)$$

A função de distribuição de distâncias é definida como

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty r q I(q) \sin(rq) dq$$

Utilizando a forma de $I(q)$ na equação (145), obtemos para a $p(r)$ de um sistema com polidispersidade em fatores de forma,

$$\begin{aligned} p(r) &= \sum_i \left(\frac{1}{2\pi^2} \int_0^\infty r q I_i(q) \sin(rq) dq \right) \\ p(r) &= \sum_i p_i(r) \end{aligned} \quad (146)$$

Assim, a $p(r)$ final será a soma das funções $p(r)$ de cada tipo de partícula. Como esta função fornece informações sobre a forma da partícula e suas dimensões, podemos extrair informações para a proteína no sistema binário comparando com os parâmetros obtidos pela proteína pura em solução.

Outro parâmetro, já descrito anteriormente, que podemos obter é a dependência do raio de giro no caso da polidispersidade em fatores de forma. Para isso tomamos a equação (59) para o cálculo do raio de giro:

$$R_g^2 = \frac{\int_0^{D_{max}} p(r) r^2 dr}{2 \int_0^{D_{max}} p(r) dr}$$

sendo $p(r)$ a função de distribuição de distâncias da partícula e $D_{\max}$ a máxima dimensão da partícula. Se usarmos a expressão para a $p(r)$ mostrada na equação (146), temos:

$$R_g^2 = \frac{\sum_i \int_0^{D_{\max_i}} p_i(r) r^2 dr}{2 \sum_j \int_0^{D_{\max_j}} p_j(r) dr} \quad (147)$$

Multiplicando e dividindo o numerador por $\int_0^{D_{\max_i}} p_i(r) dr$,

$$R_g^2 = \frac{\sum_i \frac{\int_0^{D_{\max_i}} p_i(r) r^2 dr}{\int_0^{D_{\max_i}} p_i(r) dr} \cdot \int_0^{D_{\max_i}} p_i(r) dr}{2 \sum_j \int_0^{D_{\max_j}} p_j(r) dr} \quad (148)$$

Agora temos que $R_{g_i}^2 = \frac{\int_0^{D_{\max_i}} p_i(r) r^2 dr}{2 \int_0^{D_{\max_i}} p_i(r) dr}$. Assim

$$R_g^2 = \frac{\sum_i R_{g_i}^2 \cdot \int_0^{D_{\max_i}} p_i(r) dr}{\sum_j \int_0^{D_{\max_j}} p_j(r) dr} = \sum_i R_{g_i}^2 \cdot \left(\frac{\int_0^{D_{\max_i}} p_i(r) dr}{\sum_j \int_0^{D_{\max_j}} p_j(r) dr} \right)$$

$$R_g^2 = \sum_i R_{g_i}^2 \cdot \alpha_i, \text{ com } \alpha_i = \frac{\int_0^{D_{\max_i}} p_i(r) dr}{\sum_j \int_0^{D_{\max_j}} p_j(r) dr} \quad (149)$$

ou seja, teremos uma espécie de média quadrática ponderada dos raios de giro. Podemos simplificar ainda mais esta expressão. Para isso, tomemos a transformada de Fourier inversa da $p(r)$,

$$I(q) = 4\pi \int_0^\infty p(r) \frac{\sin qr}{qr} dr$$

se fizermos $q=0$, como $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ temos:

$$I(0) = 4\pi \int_0^\infty p(r) dr \quad (150)$$

No caso de partículas temos uma máxima dimensão bem definida, acima da qual $p(r)=0$. Assim a integral anterior pode ser escrita até um valor finito:

$$I(0) = 4\pi \int_0^{D_{max}} p(r) dr \quad (151)$$

Note que esta integral é a mesma usada no cálculo do raio de giro (equação 145). Escrevendo a expressão para o raio de giro médio em função de $I(0)$ temos:

$$R_g^2 = \sum_i R_{gi}^2 \cdot \alpha_i \quad \text{com} \quad \alpha_i = \frac{I_i(0)}{\sum_j I_j(0)} \quad (152)$$

Como já citamos anteriormente, o valor de $I(0)$ (para sistemas monodispersos) é dado por

$$I(0) = N(\Delta\rho)^2 V_p^2$$

assim, se assumimos que o contraste de densidade de espalhamento é o mesmo para todas as partículas na solução,

$$\alpha_i = \frac{N_i V_i^2}{\sum_j N_j V_j^2} \quad (153)$$

ou seja o peso da contribuição no raio de giro na média será dado pelo produto entre número de partículas e o volume ao quadrado da mesma.

- Modelamento para processo de agregação

Na equação (101) obtivemos que a intensidade de espalhamento na origem $I(0)$ se relaciona com o volume da partícula espalhadora V_p através da relação,

$$I(0) = [c \bar{v} V_c (\Delta\rho)^2] V_p$$

sendo c a concentração mássica de partículas (mg/ml), $\bar{v}$ o volume parcial específico da partícula [cm^3/g] e V_c o volume iluminado pelo feixe incidentes de raios X.

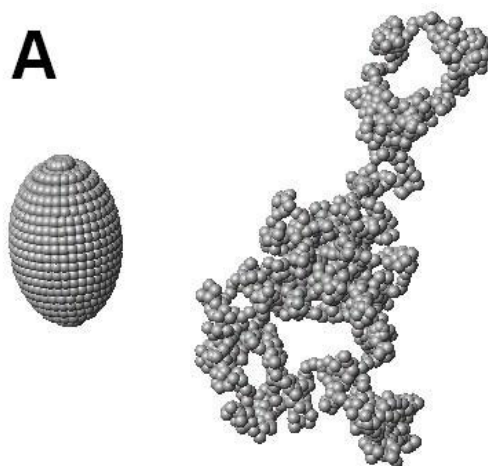
Se assumirmos que não ocorrem mudanças de fase e/ou reações químicas em um processo que ocorra na solução, a quantidade nos colchetes é constante e podemos escrever,

$$I(0) = C \cdot V_p \quad (154)$$

Assim, sabendo-se C podemos obter o volume (médio) das partículas espalhadoras. Para isso, voltando ao nosso experimento, tomamos uma medida padrão (por exemplo

lisozima à 20°C). Se as condições de monodispersidade são satisfeitas, obtemos a constante C para o padrão, já que sabemos seu volume pela sua estrutura cristalográfica. Com esta constante podemos seguir qualquer processo de agregação desta proteína.

Se considerarmos partículas no vácuo, quando a proteína se desnatura temos uma variação na curva de espalhamento, mas o valor de $I(0)$ não se altera. Isto porque o “volume espalhador” por partícula é o mesmo (seqüência de cadeias polipeptídicas). Já para partículas em solução isto é quase verdade. Como partículas em solução possuem uma camada de solvatação ao seu entorno, camada esta que contribui para o espalhamento, teremos uma intensidade ligeiramente maior para a proteína desenovelada do que para a proteína enovelada. Isto porque estruturas que na proteína enovelada estão internas, passam a ficar disponíveis na proteína desenovelada, aumentando assim a contribuição referente à camada de solvatação. No entanto a contribuição desta camada é bem menor do que a intensidade referente ao espalhamento da proteína. Logicamente que parâmetros como o raio de giro e máxima dimensão variarão, uma vez que estas quantidades estão relacionadas à forma da partícula. Abaixo mostro uma simulação exemplificando esta discussão: temos uma partícula globular compacta composta de 1785 esferas e outra com o mesmo número de esferas (mesma massa, portanto), mas em formato *random-coil* (representando proteína desenovelada)



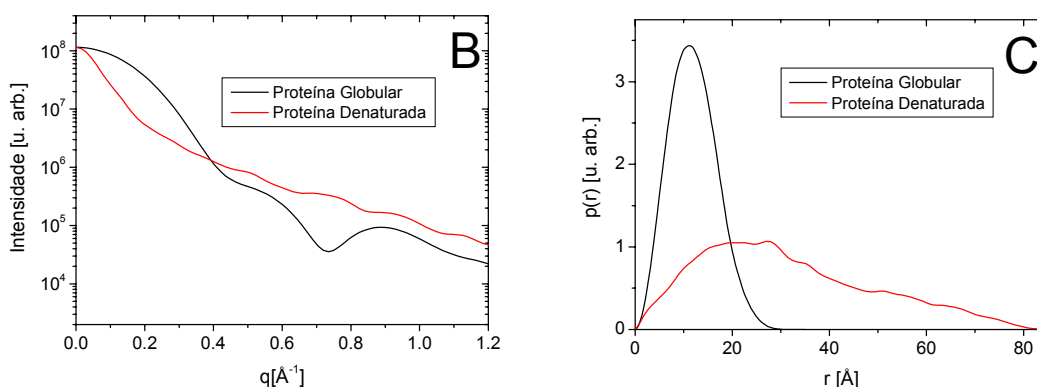


Figura 52 – Simulação de estruturas compactas e estruturas tipo “random coil”. A) Modelos tipo compacto e tipo “random coils” com o mesmo número de subunidades, B) Intensidades de espalhamento e C) funções $p(r)$ calculadas.

Na simulação consideramos que temos na composição dos modelos átomos de carbono, que em geral, são os componentes principais de proteínas. Os modelos e cálculos de intensidades foram feitos com o programa BUILDER (Oliveira, 2001).

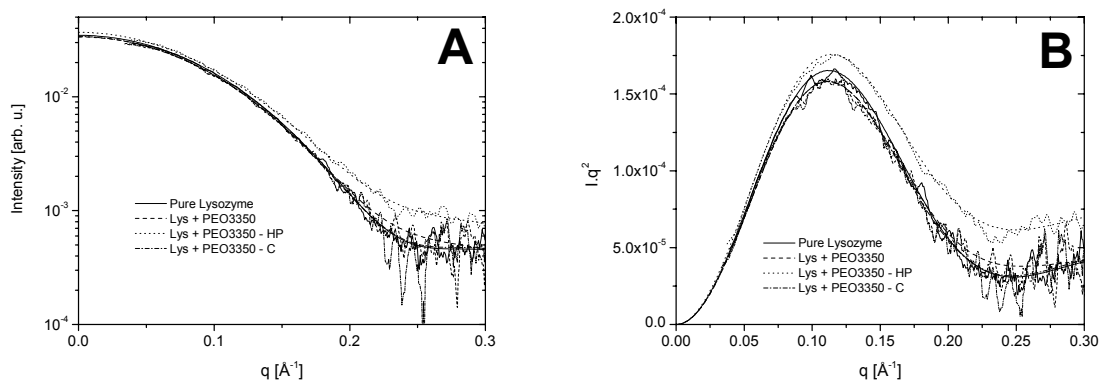
Outros tipos de análises podem ser feitas utilizando o gráfico de Kratky ($I \cdot q^2 \times q$), descrito anteriormente, cujo comportamento indica o grau de compactação da partícula espalhadora.

d) Resultados Experimentais

Como apresentado na modelagem teórica, os dados de SAXS permitem uma descrição estrutural detalhada de misturas de lisozima com PEO e copolímeros bloco. O estudo por SAXS compôs-se de várias etapas. Inicialmente investigamos a mistura proteína-polímero na temperatura ambiente, visando obter informações sobre a interação entre proteína e polímeros nesta situação. Posteriormente, um estudo com variação de temperatura foi realizado, com temperaturas de 50, 60, 70 e 80°C. Por volta de 80°C inicia-se a denaturação e a conseqüente agregação da lisozima. Este processo de agregação foi acompanhado de modo a determinar-se possíveis influências dos polímeros neste processo. Por fim, investigamos mudanças na conformação de micelas de polímero pela presença de proteínas.

- Medidas à temperatura ambiente

Na Figura 53 temos as intensidades de SAXS, os correspondentes gráficos de Kratky e função de distribuição de distâncias $p(r)$ para os três sistemas estudados. As intensidades de espalhamento para a lisozima+PEO3350 sobre condições normais, depois de sujeitas a alta pressão e depois de centrifugação (Figura 53), apresentam comportamentos similares. Para as três condições estudadas o raio de giro médio da partícula é praticamente o mesmo da proteína pura em solução, estando dentro do erro experimental (Tabela 14) valor este que está em ótima concordância com o valor conhecido para esta proteína. Uma pequena diferença é notada para a amostra proteína+polímero que foi submetida à pressão. Neste caso tem-se uma contribuição para o espalhamento total que pode ser proveniente de mudanças no grau de encadeamento (*degree of coiling*) do polímero sobre pressão na presença da proteína, ou da perda de parte da estrutura secundária e/ou terciária da proteína também induzida por pressão. O gráfico de Kratky para esta mesma amostra (Figura 53B) mostra um aumento na curva para altos valores de q , que é típico em sistemas contendo estruturas tipo *random coil*. As funções $p(r)$ mostradas na indicam que a forma geral da proteína não foi alterada pela presença do polímero na solução. Nota-se somente uma pequena mudança na forma desta curva quando submetida à alta pressão, com um aumento do número de distancias curtas em decorrência do aumento da intensidade de espalhamento a altos ângulos.



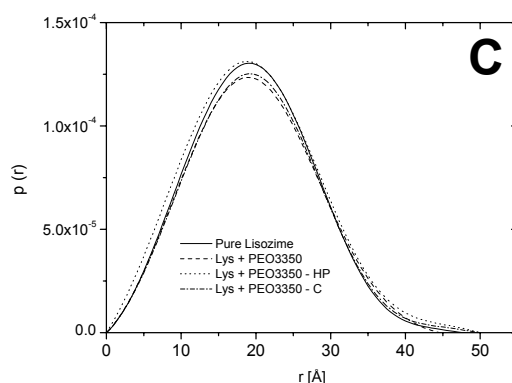


Figura 53 – Experimentos de SAXS para lisozima com PEO3350. A) Curvas de Intensidades de SAXS. B) Gráficos de Kratky e C) função de distribuição de pares de distâncias para o sistema lisozima+PEO3350 em condições normais, após ser submetido à alta pressão (HP) e após centrifugação (C). Em A) e B) cada curva experimental é superposta ao ajuste teórico correspondente (curva lisa).

Tabela 14 – Mudanças no raio de giro da proteína na presença dos polímeros PEO, F38 e P105 em várias condições experimentais.

Amostra	Rg/[Å]
Lisozima Pura	15.0 ± 0.3
(Lisozima + PEO)	15.0 ± 0.3
(Lisozima + PEO) – Sob alta pressão	15.2 ± 0.5
(Lisozima + PEO) – Centrifugada	15.2 ± 0.4
(Lisozima + F38)	14.6 ± 0.6
(Lisozima + F38) – Centrifugada	14.5 ± 0.4
(Lisozima + P105)	27 ± 10
(Lisozima + P105) – Sob alta pressão	33 ± 9
(Lisozima + P105) – Centrifugada	14.3 ± 0.7

Na Figura 54 temos os resultados para a mistura lisozima + F38. Para estas amostras novamente não obtivemos mudanças específicas na conformação da proteína. Como pode ser visto, a intensidade de espalhamento é praticamente coincidente com a curva da proteína pura e, como mostrado na Tabela 14, os valores de raio de giro calculados para as partículas espalhadoras também estão muito próximos aos valores da proteína nativa em solução. Os gráficos de Kratky (Figura 54B) indicam uma estrutura compacta e as funções $p(r)$ correspondem a partículas globulares sem diferenças significantes entre a proteína no

estado puro e para misturas proteína+polímero mesmo submetida a centrifugação ou pressão. É interessante notar que as diferenças vistas à alta pressão para o polímero PEO3350 não foram vistas aqui, indicando que a contribuição para as mudanças nos padrões de espalhamento naquela amostra referem-se a mudanças na conformação do polímero e não na proteína.

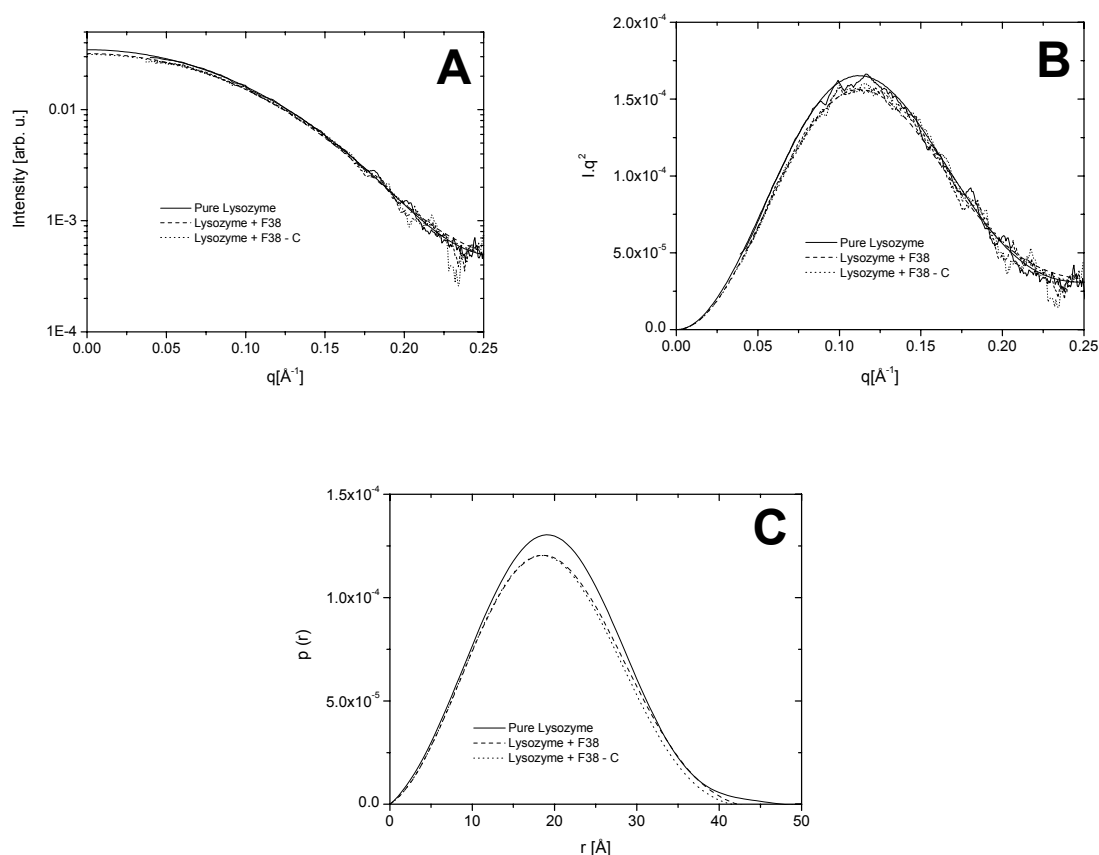
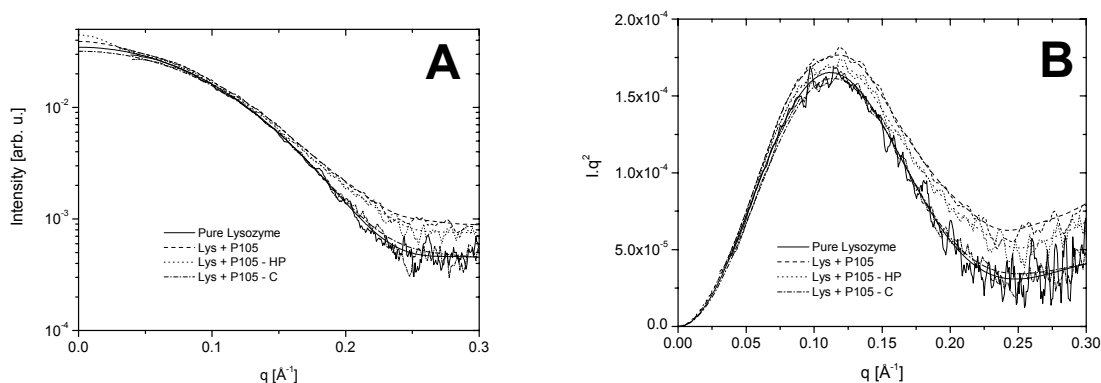


Figura 54 – Experimentos de SAXS para lisozima com F38. A) Curvas de Intensidades de SAXS. B) Gráficos de Kratky e C) função de distribuição de pares de distâncias para o sistema lisozima+F38 em condições normais, após ser submetido à alta pressão (HP) e após centrifugação (C). Em A) e B) cada curva experimental é superposta ao ajuste teórico correspondente (curva lisa).

Finalmente, na Figura 55 temos o resultado para o sistema lisozima + P105 (6500Da). As intensidades de espalhamento mostram diferenças a ângulos muito pequenos e também a maiores valores de q . Quando a amostra é submetida a pressão vê-se um aumento da intensidade a baixos ângulos (Figura 55A), que pode ser descrito como um espalhamento residual das micelas de polímero que não puderam ser eliminadas após a

subtração do espalhamento branco (polímero + solvente). Isto pode ser indicar uma população maior de micelas no sistema proteína + polímero + solvente sugerindo que a presença da proteína favorece, de alguma forma, a formação das micelas. O raio de giro para este sistema é muito maior do que a da proteína pura (Tabela 14). Isso pode ser atribuído à contribuição de micelas grandes presentes no sistema (quando partículas de vários tamanhos estão presentes, temos uma média ponderada dos valores como indicada na equação 152). Nota-se também que a centrifugação da amostra proteína-polímero produz um sistema cuja intensidade é muito parecida à da proteína pura. Isto é melhor observado na função $p(r)$ correspondente à lisozima pura e à amostra lisozima + P105 centrifugada (Figura 55C). Por outro lado, tanto o sistema lisozima + P105 em condições normais e após ser submetido à pressão, ambos possuem funções $p(r)$ nas quais a parte inicial coincide com a da proteína pura e uma parte posterior que se estende para valores muito maiores de r . Isso pode ser interpretado em termos da equação (146) como um resultado da contribuição de dois espalhadores independentes: proteínas e micelas de polímeros. De acordo com esta interpretação, as distâncias maiores decorrem de micelas de polímero (maiores do que a proteína) presentes no sistema. No entanto, pela coincidência entre as funções $p(r)$ para com a proteína pura, vemos que não ocorre a mudança da forma geral da proteína. A parcela da função $p(r)$ correspondente às micelas poliméricas apresenta algumas oscilações e uma descrição apenas satisfatória em decorrência da baixa concentração de micelas no sistema.



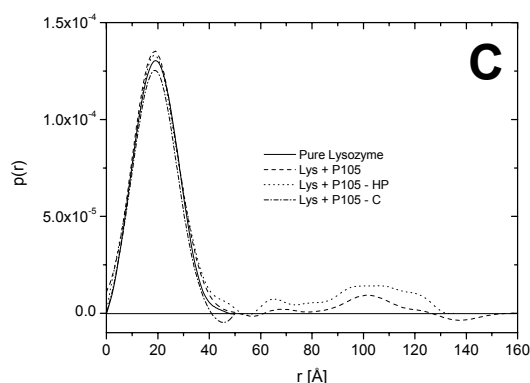


Figura 55 – Experimentos de SAXS para lisozima com P105. A) Curvas de Intensidades de SAXS. B) Gráficos de Kratky e C) função de distribuição de pares de distâncias para o sistema lisozima+P105 em condições normais, após ser submetido a alta pressão (HP) e após centrifugação (C). Em A) e B) cada curva experimental é superposta ao ajuste teórico correspondente (curva lisa).

Desta forma os dados de SAXS para o sistema à temperatura ambiente, nas várias condições mostradas não indicam a formação de nenhum tipo de complexo proteína-polímero estável como indicado por outros autores. Também, medidas de calorimetria diferencial e cromatografia de exclusão (Almeida e Oliveira et. al, 2004) suportam a não formação de complexos nestas condições.

- Medidas com variação de temperatura

- Lisozima pura

Visto que a lisozima no estado nativo será a base para comparações entre possíveis interações proteína-polímero, realizamos todos os experimentos para a lisozima pura e posteriormente para as misturas. Na Figura 56 mostramos a intensidade de espalhamento, função $p(r)$ e gráfico de Kratky para proteína pura submetida às temperaturas de 20, 50, 60, 70 e 80°C. Até 70°C o sistema manteve-se monodisperso e, em 80°C após o claro início da denaturação da proteína, observou-se um processo de agregação em consequência desta denaturação.

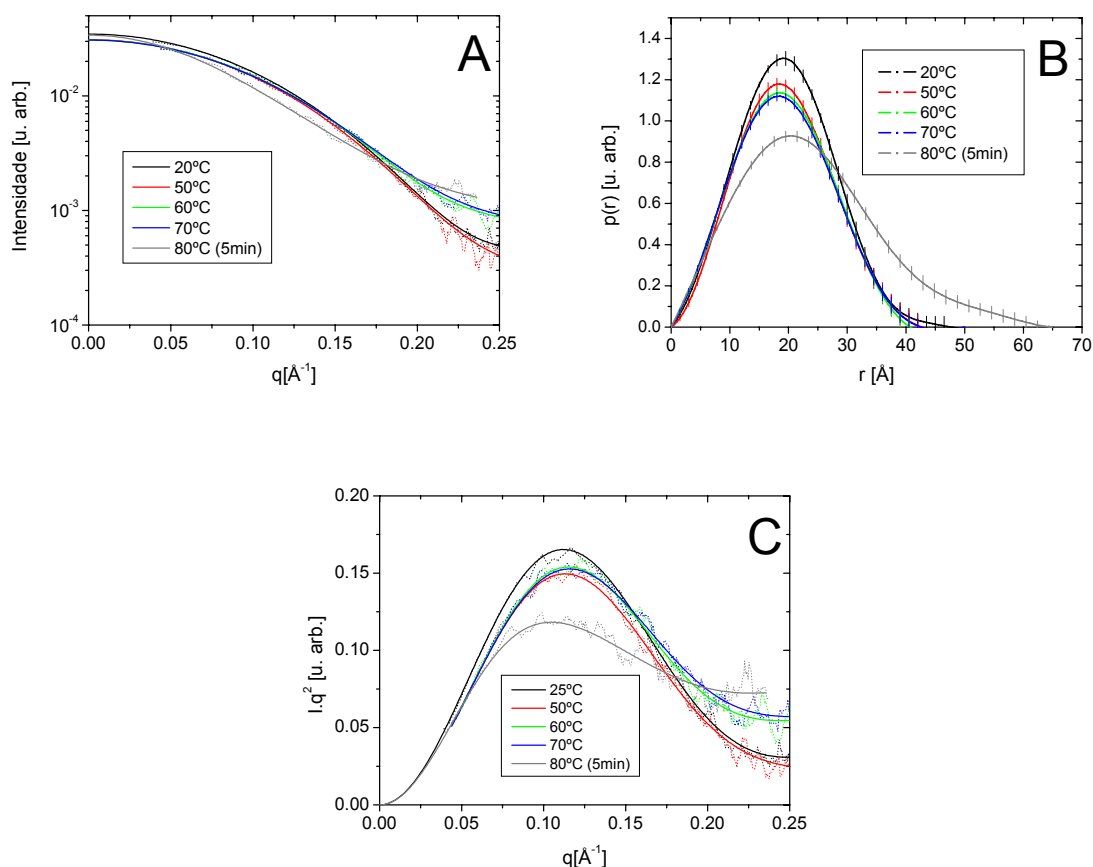


Figura 56 – Espalhamento da lisozima pura com variação de temperatura. A) Intensidades de espalhamento, B) funções $p(r)$ e C) Gráficos de Kratky para lisozima à 20, 50, 60, 70 e 80°C.

Como indicado em diversos outros trabalhos (Hirai et al, 1998, 1999) em 50, 60 e 70° temos uma leve mudança conformacional na lisozima em comparação com a estrutura à 20°C. A partir de 80°C vemos um grande aumento do raio de giro e da máxima dimensão, indicando a denaturação da proteína (compare com Figura 52). No gráfico de Kratky também vemos este efeito onde a curva perde sua forma simétrica indicando que a proteína não é mais globular compacta indicando a presença de cadeias flexíveis (*random coils*). Para a proteína a 80° fizemos uma cinética com exposições de 5 min de modo a acompanhar a dinâmica de agregação. Na Figura 57 mostramos estes resultados com 12 exposições de 5 minutos cada:

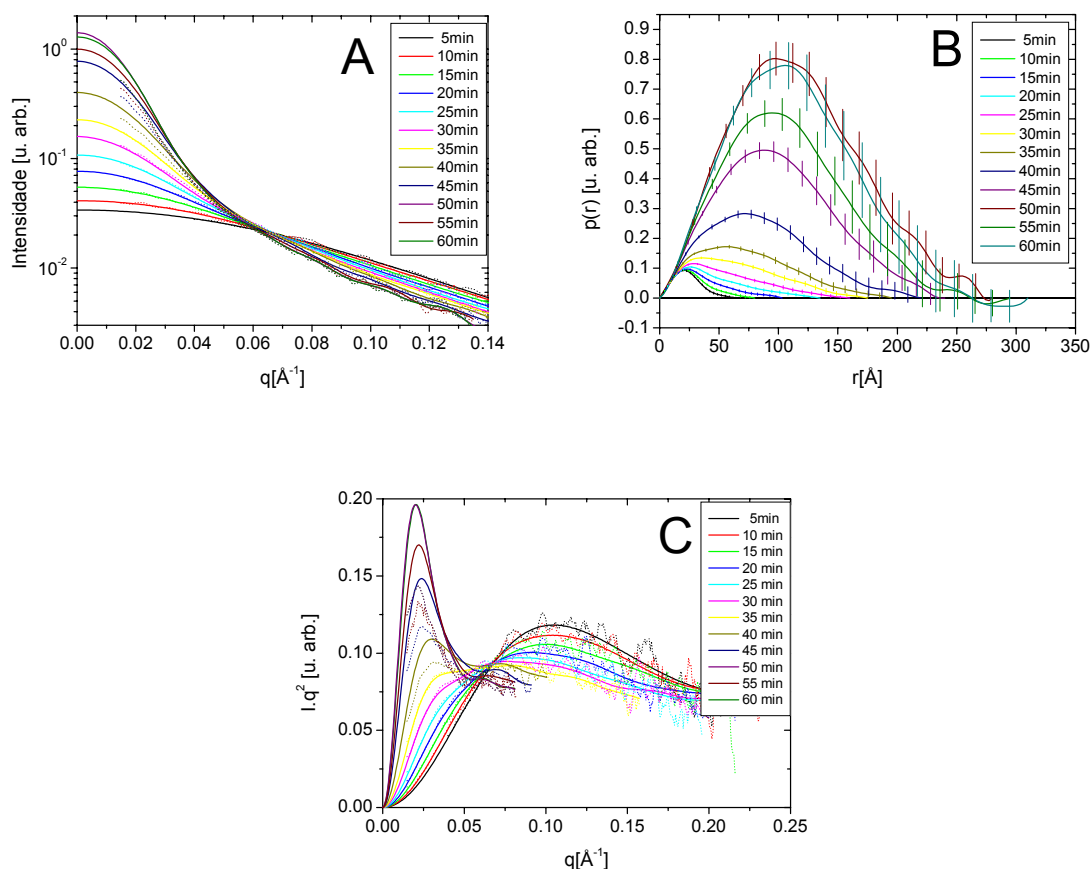


Figura 57 – Cinética de agregação da lisozima à 80°C. Cada curva corresponde a exposições de 5 minutos, tiradas sequencialmente. A) Intensidades de espalhamento, B) funções $p(r)$ e C) Gráficos de Kratky.

Na cinética de agregação acompanhamos a evolução do sistema utilizando o ajuste dos dados de espalhamento com o modelo de sistema monodisperso. Logicamente que isso é uma aproximação visto que no processo de agregação pode-se imaginar um sistema polidisperso. Entretanto, tendo em mente a equação (146), podemos interpretar a função $p(r)$ obtida como sendo uma média das contribuições dos centros espalhadores. Desta cinética vemos que a proteína parte de um estado globular parcialmente denaturado (como indicado pelas funções $p(r)$ e *Kratky plot*) nos primeiros 5 minutos em 80°C, para então iniciar o processo de agregação. Pela função $p(r)$ vemos que formam-se agregados globulares e pelo gráfico de Kratky vemos que os agregados maiores (após 40 minutos) são bastante compactos.

Estes resultados servirão como base de comparação para os sistemas lisozima+polímero.

- Lisozima + PEO3350

Similarmente ao mostrado anteriormente, utilizamos a intensidade de espalhamento do PEO3350 como “branco” do sistema Lisozima+PEO3350. Na Figura 58 mostramos a curva resultante. Vemos uma grande similaridade entre as curvas de espalhamento, função $p(r)$ e gráfico de Kratky para essa mistura lisozima-polímero em todas as temperaturas para com o caso da lisozima sozinha (Figura 56). Isso nos indica que nenhum tipo de interação (capaz de mudar o fator de forma da proteína) ocorreu. À 80° vemos uma forte mudança da intensidade de espalhamento, função $p(r)$ e compacidade da proteína, indicando o início de sua denaturação.

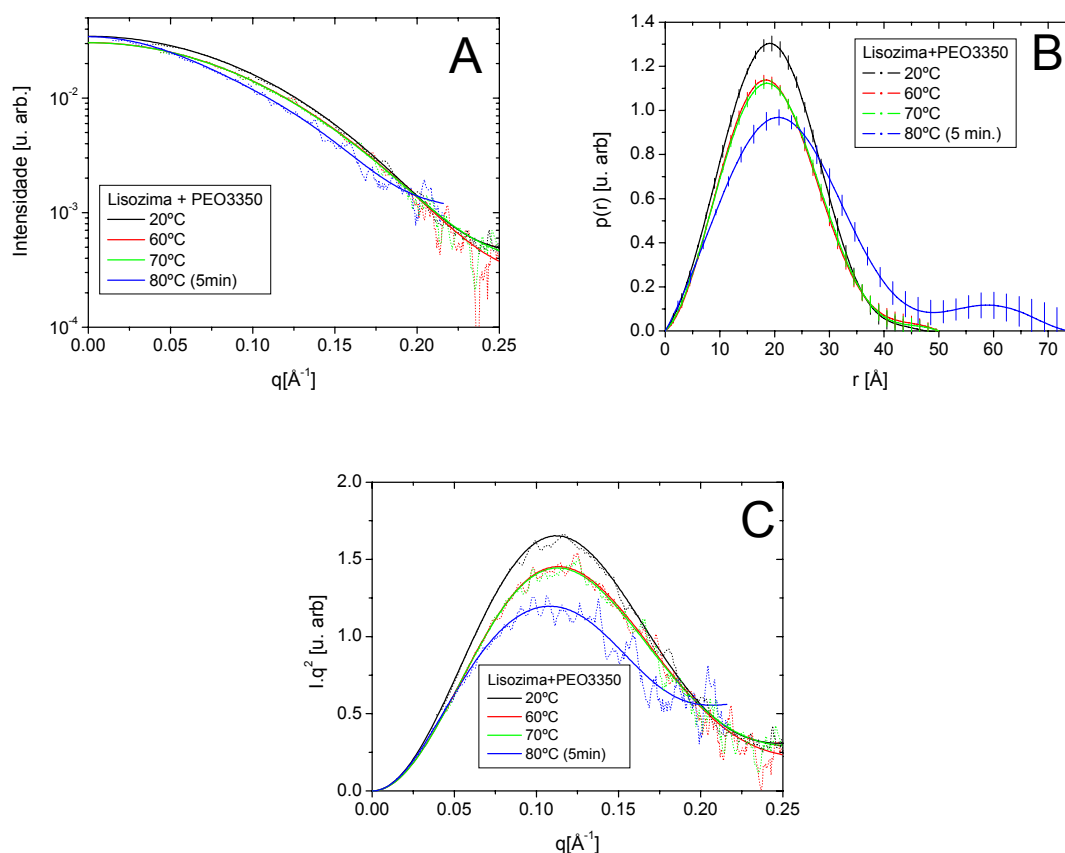


Figura 58 – Espalhamento da lisozima+PEO3350 com variação de temperatura. A) Intensidade de espalhamento, B) função $p(r)$ e C) Gráfico de Kratky para lisozima+ PEO3350 a 20, 60, 70 e 80°C.

Na cinética à 80°C notamos também uma semelhança muito grande ao espalhamento da lisozima pura indicando que, mesmo nestas condições (alta temperatura, agregação da proteína em meio a presença do polímero) nenhum tipo de efeito (capaz de alterar drasticamente a dinâmica do processo) ocorreu. Como veremos posteriormente na análise da mudança dos valores de raio de giro, a presença do polímero altera a forma dos agregados protéicos.

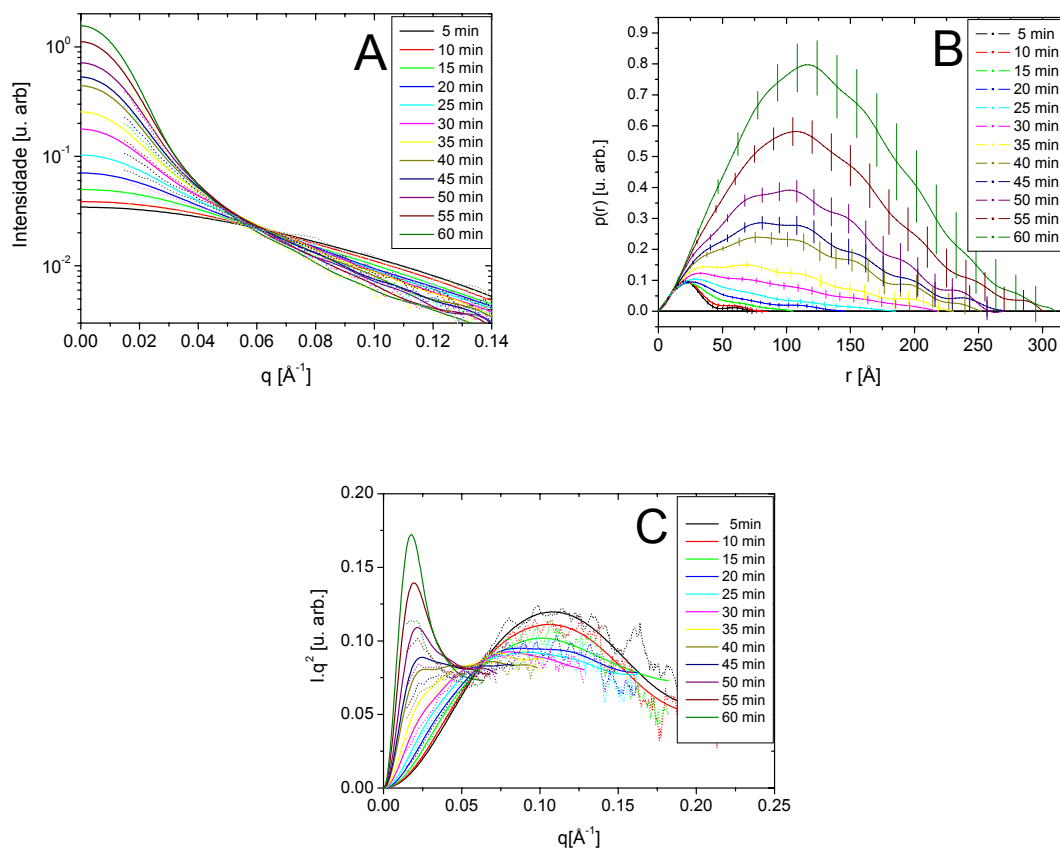


Figura 59 – Cinética de agregação da lisozima + PEO3350 à 80°C. Cada curva corresponde a exposições de 5 minutos, tiradas seqüencialmente. A) Intensidades de espalhamento, B) funções $p(r)$ e C) Gráficos de Kratky.

- Lisozima + P105

Como comentamos anteriormente, o copolímero P105 forma micelas em solução, sendo que sua formação é favorecida pelo aumento da temperatura. Como mostrado nos resultados à temperatura ambiente, se subtraímos do espalhamento da mistura lisozima + P105 o espalhamento somente do P105 percebemos uma influência de efeitos da presença

das micelas que não são completamente eliminados pela subtração do espalhamento “branco”. Desta maneira não é correto considerar como “branco” o espalhamento do copolímero, mas sim subtrair somente a contribuição do *solvente* e descrever o sistema binário resultante, tratando-o como possuindo polidispersidade em fatores de forma (equações (145) e (146)). De acordo com a introdução teórica, analisando a função $p(r)$ resultante poderemos detectar a contribuição de cada uma das partes.

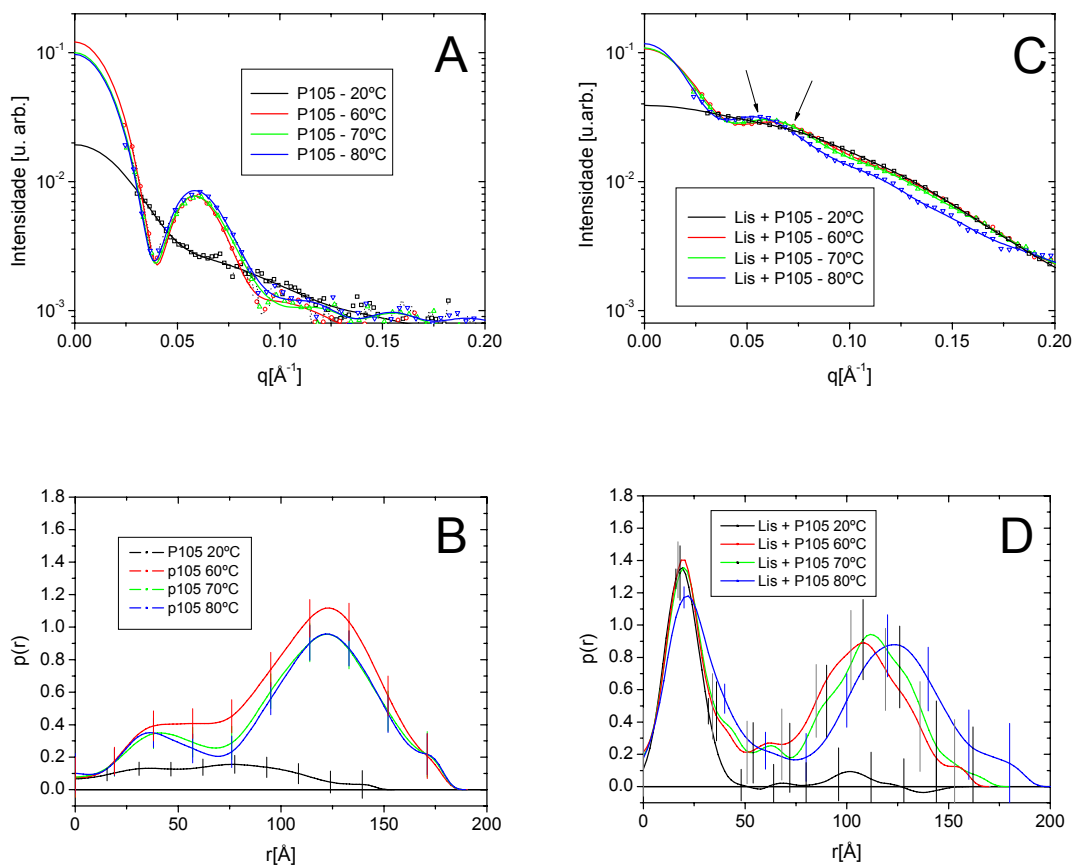


Figura 60 – Intensidades de SAXS para P105 e o sistema composto Lisozima + P105 para várias temperaturas. A) e B) Intensidade de espalhamento e função $p(r)$ para P105 à 20, 60, 70 e 80°C. C) e D) Intensidade de espalhamento e função $p(r)$ para P105+Lisozima à 20, 60, 70 e 80°C

Na Figura 60 mostro curvas de espalhamento correspondentes à lisozima+PEO e PEO puro para 20, 60, 70 e 80°C. À 20°C vemos na Figura 60A um pequeno ombro o que indica a presença de uma pequena população de micelas, que ainda não é suficiente para fornecer um máximo pronunciado de correlação. Em 60, 70 e 80°C vemos que o sistema assume a forma micelar, com características semelhantes para as três temperaturas Figura

60A,B. Isso nos indica que, apesar da temperatura micelar crítica para P105 ser próxima de 20°C (vide Tabela 13) somente acima de 60°C temos uma micelação pronunciada do copolímero tribloco P105 (Glatter et al, 1994). Já para as amostras de lisozima+PEO vemos características bastante interessantes. Na temperatura ambiente não temos uma população grande de micelas, e o espalhamento é dominado pelo da lisozima. Isto é visto na $p(r)$ (Figura 60D) pela forma da mesma: temos a curva similar à da lisozima pura, só que a máxima dimensão é forçosamente aumentada pela presença das micelas na solução. Com o aumento da temperatura para 60°C, o número de micelas e seu tamanho aumentam de modo que sua contribuição para a curva de intensidade passa a ser maior (Figura 60C setas indicativas), gerando o máximo de correlação mais pronunciado nas curvas de intensidade. Isso se reflete nas curvas $p(r)$ com o aparecimento do máximo característico de micelas na região a altos valores de r . Aumentando-se ainda mais a temperatura (70 e 80°C) inicia-se a desnaturação da lisozima (alargamento do máximo no início na $p(r)$) e, diferentemente do caso onde tínhamos as micelas sem a presença de proteína, também há mudança no tamanho das micelas (variação da posição do máximo no final da $p(r)$). Pelos resultados à 60 e 70°C as micelas formadas no sistema proteína+polímero são ligeiramente menores do que quando no sistema somente de polímero. Isto é visto na forma da função $p(r)$, onde o máximo da contribuição das micelas está a valores de r menores do que no caso de polímero sozinho, o que indica que a presença da proteína na solução influencia, de algum modo, a formação das micelas. Já à 80°C a posição do máximo fica na mesma posição do obtido para o polímero sozinho (compare Figura 60B e Figura 60D) o que indica que, a esta temperatura, a presença das proteínas não mais influencia o tamanho das micelas formadas.

Com base no que apresentamos na introdução teórica, vemos que a aproximação por um sistema de 2 fases pode ser aplicada a este sistema uma vez que é clara a contribuição da lisozima e das micelas na $p(r)$ da mistura. Desta forma, apesar de termos constatado uma influência da presença da proteína no tamanho das micelas, não pudemos detectar a presença de agregados proteína+polímero também para este sistema.

Olhemos agora para a variação à 80°C. Como vimos, nesta temperatura ocorre a desnaturação da lisozima seguida do processo de agregação. Na Figura 61 abaixo temos as intensidades de espalhamento, $p(r)$'s e Gráficos de Kratky para este processo.

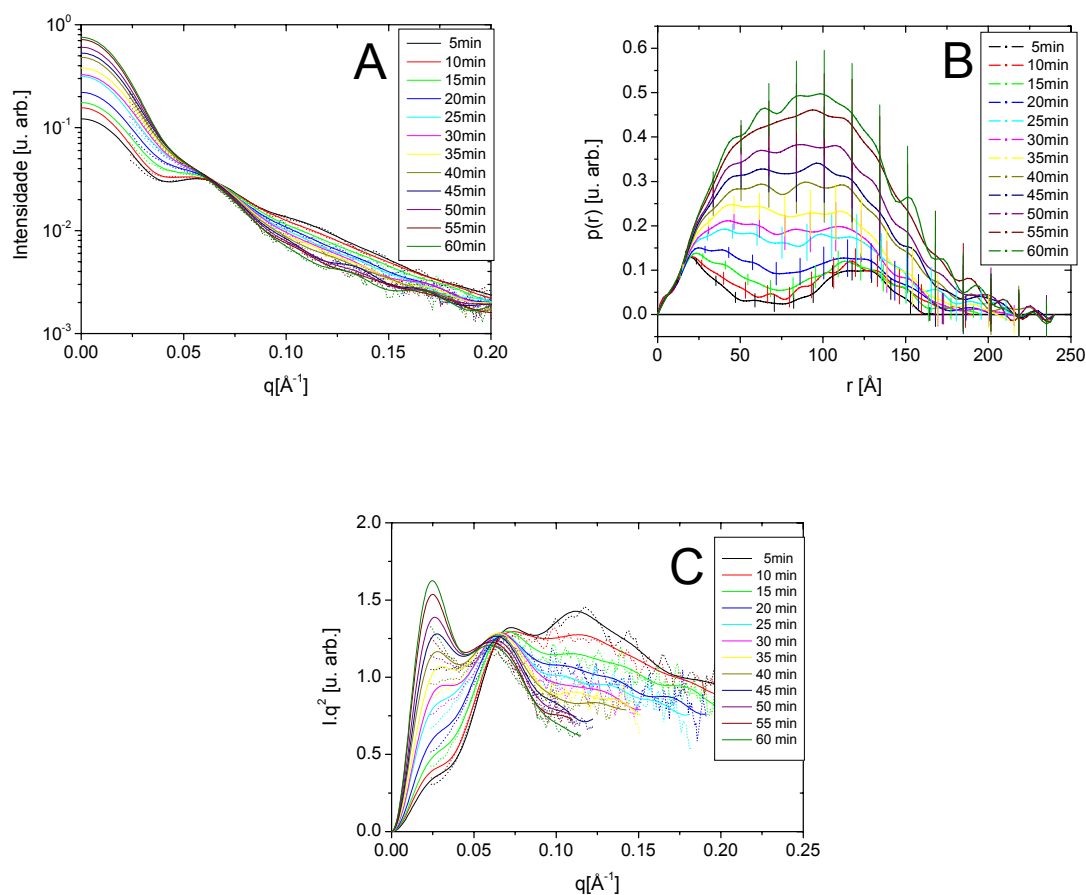


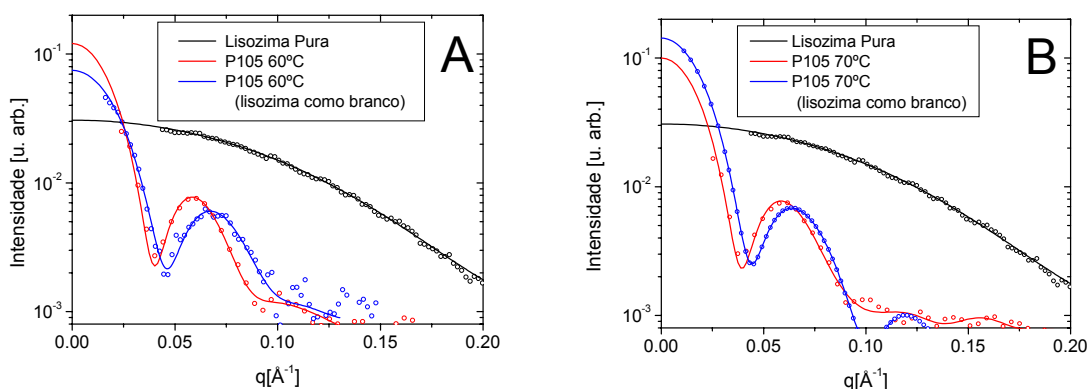
Figura 61 – Cinética de agregação da lisozima + P105 à 80°C. Cada curva corresponde a exposições de 5 minutos, tiradas sequencialmente. A) Intensidades de espalhamento, B) funções $p(r)$ e C) Gráficos de Kratky.

Pelas curvas da intensidade espalhada, vemos que a parcela referente às micelas permanece praticamente inalterada na curva de intensidade, enquanto que a contribuição decorrente da agregação da lisozima aumenta rapidamente. Isto é visto também na $p(r)$ (Figura 61B) onde inicialmente tínhamos a contribuição da lisozima e das micelas claramente indicadas para posteriormente vermos o aumento da contribuição dos agregados. A análise do gráfico de Kratky (Figura 61C) é dificultada pela presença do pico de correlação das micelas, mas podemos perceber que os agregados formados após 40 min assumem uma conformação compacta.

- Velocidade de agregação

Como mostramos na introdução teórica, podemos acompanhar a formação dos agregados utilizando a relação $I(0) = C.V_p$, sendo V_p o volume médio das partículas espalhadoras e C uma constante a determinar. Para a determinação de C utilizamos a lisozima como padrão. O volume desta proteína é determinado a partir de sua estrutura cristalográfica (*PDB entry 6LYZ*) e, com os dados à temperatura ambiente obtivemos C . Com este valor podemos obter o volume dos agregados formados diretamente de $I(0)$ obtido nas mesmas condições experimentais.

Para a amostra proteína+PEO3350 podemos utilizar diretamente esta relação e acompanhar o processo de agregação. Já para a amostra proteína+P105 alguns aspectos devem ser considerados. O polímero forma micelas na solução e, como vimos, a presença da proteína altera a conformação destas micelas. Isto foi indicado pela forma da função $p(r)$ e também é mostrado nas curvas a seguir (Figura 62), onde temos o espalhamento da lisozima, o do P105 em solução e da diferença [(lisozima + P105)-(lisozima)], medidos para 60, 70 e 80°C:



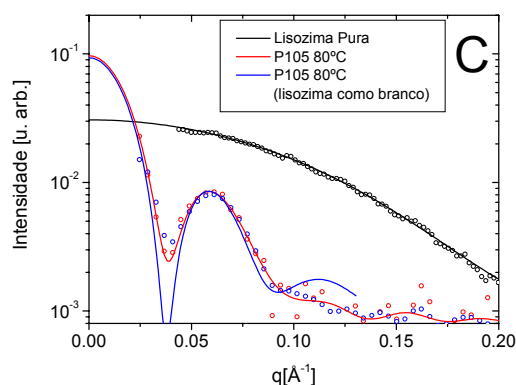


Figura 62 – Diferenças da formação de micelas na solução com e sem proteína para várias temperaturas.
 A) Intensidade de espalhamento para polímero P105 à 60°, B) Intensidade de espalhamento para polímero P105 à 70°, C) Intensidade de espalhamento para polímero P105 à 80°.

Com o aumento da temperatura, as micelas formadas na presença da proteína ficam cada vez mais parecidas às formadas com somente o polímero em solução, até que, em 80° as curvas para as micelas em ambos os casos ficam praticamente idênticas. Também o valor de $I(0)$ fica muito próximo, outra evidência de que as micelas formadas nos dois casos são muito parecidas. Desta forma, como discutido anteriormente pela análise das funções $p(r)$, o processo de agregação da lisozima não interfere de maneira agressiva na conformação das micelas nesta temperatura. Sendo assim, para as medidas à 80°C podemos subtrair a intensidade do polímero como sendo o espalhamento “branco”, de modo a termos somente a contribuição da agregação da lisozima e acompanhar o processo de agregação.

Os valores de $I(0)$ foram obtidos de um ajuste teórico das curvas de espalhamento com o programa GNOM, corrigindo também efeitos de *smearing* decorrentes da colimação por fendas. A variação do volume, raio de giro e máxima dimensão em função do tempo de exposição são mostrados abaixo:

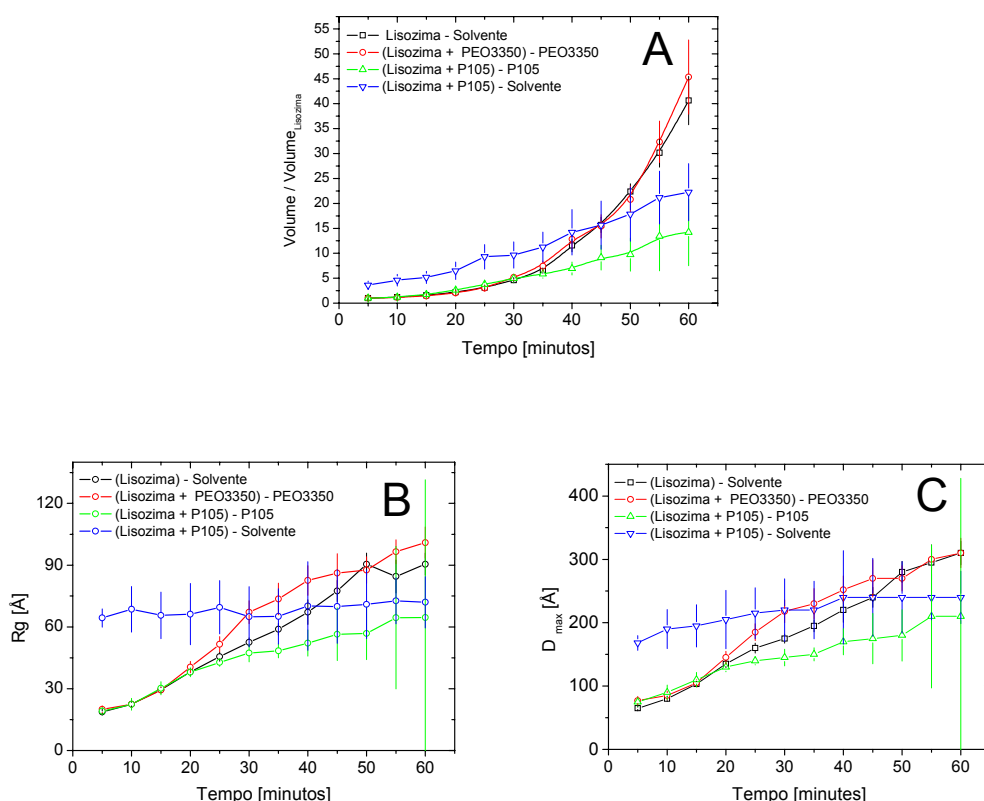


Figura 63 – Variação dos parâmetros dimensionais da lisozima com e sem polímeros em função da temperatura. A) Variação do volume médio, B) raio de giro e C) máxima dimensão dos agregados formados à 80°C em função do tempo para amostras de lisozima pura, lisozima+PEO3350 e lisozima+P105. Para a amostra com P105 dois casos foram considerados, subtraindo em um caso o espalhamento do polímero e no outro o espalhamento do solvente. Veja o texto para detalhes.

A presença do PEO3350 na solução não altera a velocidade de agregação da lisozima, fornecendo valores muito próximos dos obtidos para a lisozima pura Figura 63A. No entanto, vemos uma ligeira diferença na forma da curva da variação do raio de giro e também da máxima dimensão Figura 63B,C. Isto indica que a presença do PEO3350 não altera a velocidade de agregação, mas a forma dos agregados.

Para o P105 na solução, mostramos o comportamento assumindo o sistema composto somente por lisozima (subtraindo a contribuição do P105) e também para o sistema composto de lisozima+P105 (subtraindo somente o solvente). Pelos valores e forma das curvas, temos a confirmação de que as micelas praticamente não alteram sua forma durante a agregação. No entanto, tornam o processo de agregação da lisozima muito mais

lento do que nos casos de lisozima pura e lisozima+PEO3350. Isso nos indica que ocorre algum tipo de interação entre as micelas e as proteínas que inibe o processo de agregação. Dos dados experimentais, obtemos para as micelas à 80°C um raio de giro dado por $\approx 80\text{\AA}$ e máxima dimensão de $190\text{\AA}$ (em ótima concordância com valores já publicados por Nolan et al, 1997). Como mostrei na introdução teórica, o raio de giro para um sistema com polidispersidade em fatores de forma é dado por (equação (152)):

$$R_g^2 = \sum_i R_{g_i}^2 \cdot \alpha_i \text{ com } \alpha_i = \frac{I_i(0)}{\sum_j I_j(0)}$$

No nosso caso, se assumimos como sendo um sistema de 2 tipos de partículas (as micelas e os agregados) teremos

$$R_g = \sqrt{\frac{I_{micelas}(0)}{I_{micelas}(0) + I_{agregado}(0)} R_{g_{micelas}}^2 + \frac{I_{agregado}(0)}{I_{micelas}(0) + I_{agregado}(0)} R_{g_{agregado}}^2} \quad (155)$$

Para $R_{g_{agregado}}$ e $I(0)_{agregado}$ usamos os valores obtidos para o sistema [(lisozima+P105)-P105]. Já pra $R_{g_{micelas}}$ e $I(0)_{micelas}$ usamos os valores ($\sim$ constantes) para o espalhamento do P105 à 80°C. Graficando esta expressão em conjunto com o raio de giro experimental [(lisozima+p105)-solvente] temos,

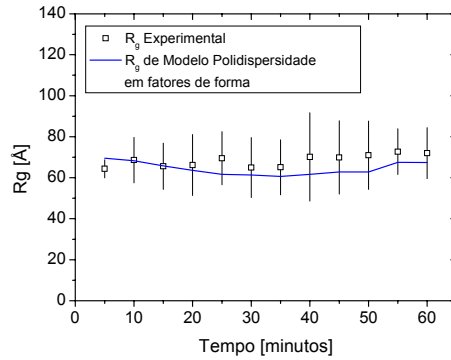


Figura 64 – Variação do raio de giro do sistema lisozima+P105 em comparação com o esperado para um sistema com dois fatores de forma.

Vemos uma ótima concordância, face a simplicidade do modelo adotado. Assim podemos dizer que à 80°C a forma das micelas não se altera com a presença dos agregados e que as micelas inibem a velocidade de agregação quando comparamos os resultados com a lisozima pura.

- Micelação do PEO105 em função da temperatura

Para tentar detectar diferenças nas micelas formadas na presença ou não da proteína, realizamos o cálculo do perfil de densidade eletrônica das mesmas. Devido as características hidrofóbicas, hidrofílicas das partes correspondente ao copolímero tribloco, temos a auto-organização das moléculas e a formação de micelas, com um centro PPO compacto, envolto por um arranjo disperso (*soft*) das cadeias de PEO em forma de estrela se estendendo dentro da fase aquosa (Glatter et al, 1994). Uma análise aprofundada deste tipo de sistema está fora do escopo desta tese, mas com base nos dados obtidos é possível obter-se alguns resultados. Cálculos baseados em métodos de Monte Carlo podem ser aplicados a estes sistemas, mimetizando a camada externa das micelas com um arranjo randômico de cadeias flexíveis (Svaneborg e Pedersen, 2002).

Nos cálculos utilizamos o espalhamento do P105 à temperatura de 60, 70 e 80°C e o comparamos ao espalhamento das micelas na presença da proteína. Como a presença do P105 praticamente não altera a forma da proteína (como argumentado nos parágrafos anteriores), utilizamos o *espalhamento da proteína como branco do sistema proteína+polímero* em cada temperatura, esperando com isso obter somente a contribuição do polímero na solução. Os ajustes são mostrados nas figuras seguintes:

P105 – 60°C

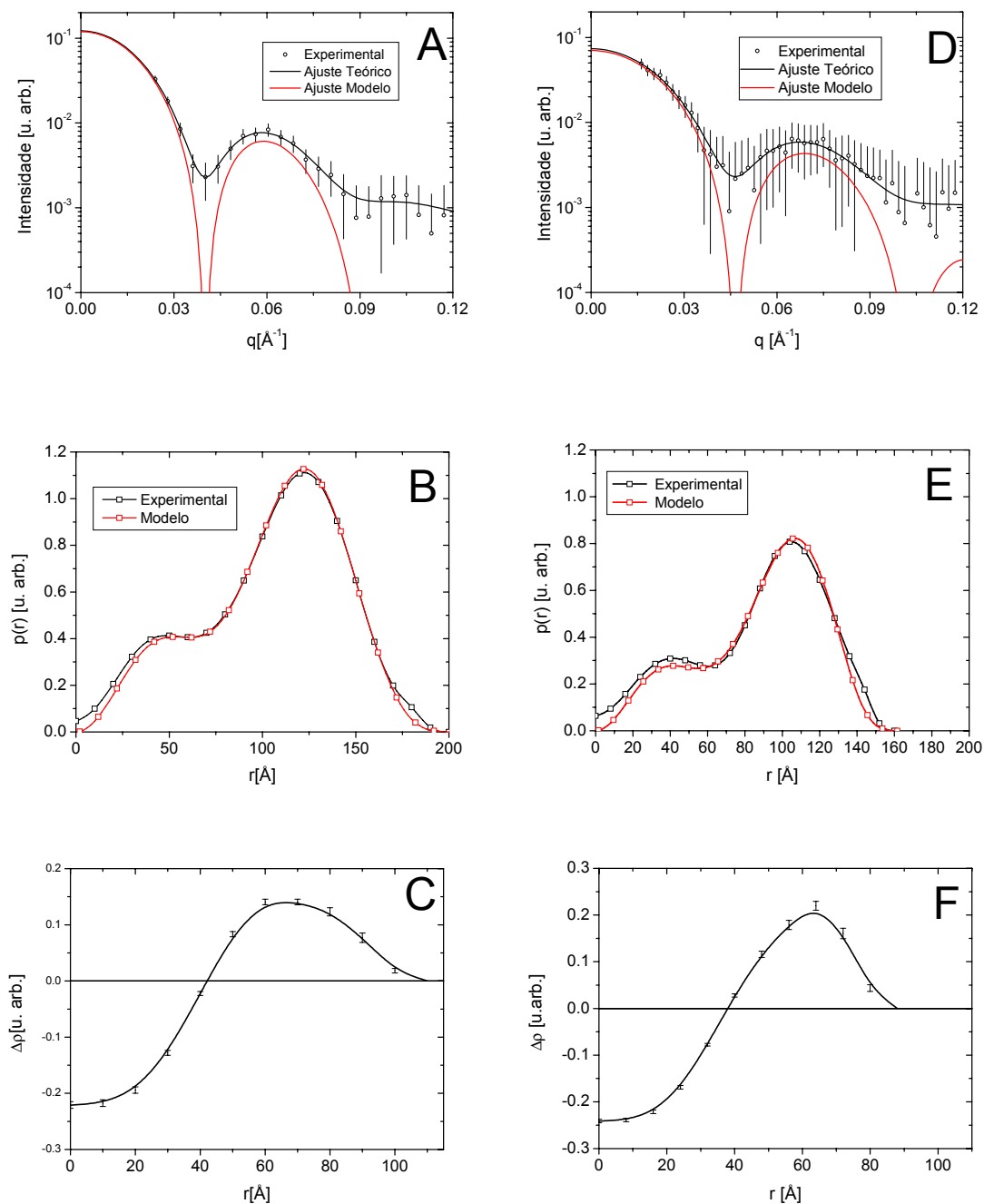


Figura 65 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 60°C. Dados para P105: A) Ajuste da intensidade de SAXS; B) Ajuste da função $p(r)$; C) Densidade eletrônica calculada (simetria esférica). Dados para P105 e lisozima: C) Ajuste da intensidade de SAXS; D) Ajuste da função $p(r)$; E) Densidade eletrônica calculada (simetria esférica).

P105 – 70°C

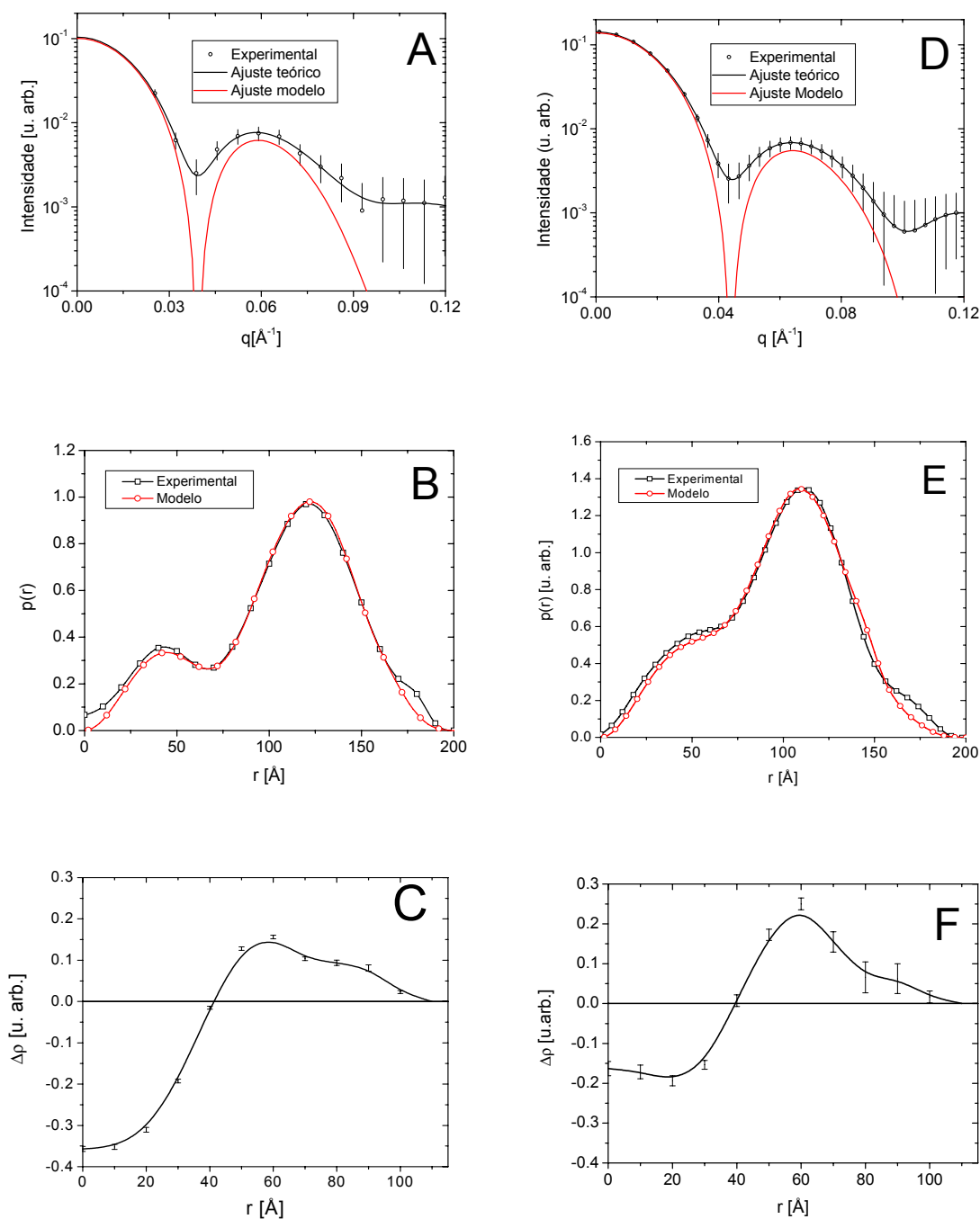


Figura 66 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 70°C. Dados para P105: A) Ajuste da intensidade de SAXS; B) Ajuste da função $p(r)$; C) Densidade eletrônica calculada (simetria esférica). Dados para P105 e lisozima: C) Ajuste da intensidade de SAXS; D) Ajuste da função $p(r)$; E) Densidade eletrônica calculada (simetria esférica).

P105 – 80°C

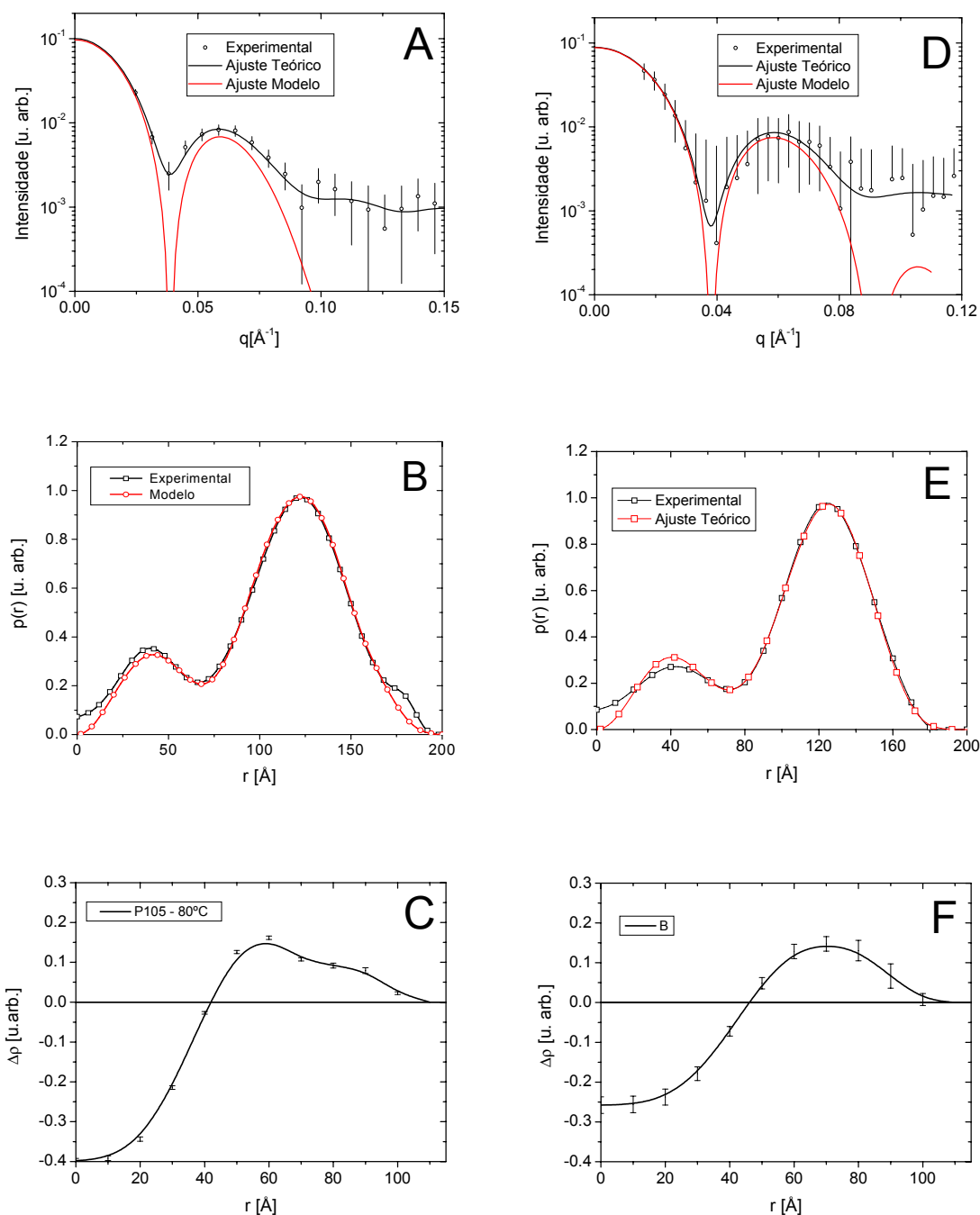


Figura 67 – Cálculo do perfil de densidade eletrônica através do ajuste dos dados experimentais à 80°C. Dados para P105: A) Ajuste da intensidade de SAXS; B) Ajuste da função $p(r)$, C) Densidade eletrônica calculada (simetria esférica). Dados para P105 e lisozima: D) Ajuste da intensidade de SAXS; E) Ajuste da função $p(r)$; F) Densidade eletrônica calculada (simetria esférica).

As curvas experimentais foram ajustadas usando-se os programas GIFT (Bergmann et al., 2000) e DECON (Glatter, 1984). O perfil de densidade eletrônica foi calculado assumindo-se simetria esférica. Como apresentado anteriormente, este programa calcula a variação do contraste na direção radial ($\Delta\rho(r)$) que melhor ajusta os dados experimentais. A $p(r)$ e o perfil de densidade eletrônica são característicos de sistemas micelares (Glatter, 1991). Em todos os casos o núcleo hidrofóbico apresentou um raio de 35-40 Å, o que é condizente com o esperado para este polímero (Glatter et. al, 1994, Nolan et al, 1997). Um fato interessante é que pelo fato de ser um núcleo compacto, não esperaria-se uma grande variação de seu valor em função da temperatura ou do meio, sendo isso confirmado pelos perfis de densidade eletrônica. O mesmo já não ocorre para com a camada externa, uma vez que a flexibilidade inerente às cadeias poliméricas permite um arranjo variável, susceptível à mudanças no meio e na temperatura.. Na realidade a descrição desta camada formada de cadeias flexíveis a por uma densidade eletrônica constante é apenas aproximada, não conseguindo levar em conta as flutuações de densidade eletrônica presentes nesta camada (Svaneborg e Pedersen, 2002). No entanto, ao nível de resolução alcançado nos experimentos, esta é uma boa aproximação. Em todos os casos a densidade eletrônica da camada externa decresceu suavemente, como esperado para este tipo de modelo.

Para 60° vemos que as micelas formadas para o polímero puro ($R_g=79\text{Å}$) são maiores dos que formados na presença da lisozima ($R_g=67\text{Å}$). Desta forma, a esta temperatura a presença da proteína no meio promove uma diminuição da camada externa de cadeias flexíveis, passando de $\sim 70\text{Å}$ para $\sim 50\text{Å}$. Para 70° temos o valor de 80Å para o raio de giro do polímero puro e 74Å para o polímero na presença da lisozima. No entanto a forma do perfil de densidade eletrônica é um pouco diferente nos dois casos, indicando uma pequena diminuição da camada externa. Finalmente para 80° temos valores de raios de giro de 80Å tanto para o polímero puro quanto para o polímero na presença da lisozima. Apesar de termos pequenas diferenças no perfil de densidade eletrônica a esta temperatura, a grande similaridade entre as curvas $p(r)$ e o valor de raio de giro nos indica que a esta temperatura as conformações assumidas pelas micelas são mais parecidas do que às outras temperaturas. Sendo assim, a camada flexível externa deve ser similar nos dois casos.

Pelos resultados mostrados, vemos que foi possível descrever-se de maneira simplificada a influência da proteína na formação das micelas de PEO-PPO-PEO. Principalmente para as temperaturas de 60 e 70°C, temos a indicação de algum tipo de influência da proteína no processo de micelação dos polímeros.

e) Considerações finais sobre o estudo

Tivemos como objetivo neste trabalho um estudo sistemático da interação entre a lisozima e diversos tipos de polímeros. Estudos anteriores (Topchieva et al., 1994, 2000) sugeriram a formação de complexos proteína-polímero estáveis à temperatura ambiente, os quais eram favorecidos se a amostra era centrifugada ou submetida à pressão. Repetimos estes experimentos, utilizando a técnica de SAXS para determinar-se a possível complexação entre a lisozima e os polímeros. Em todos os casos realizados à temperatura ambiente os dados de espalhamento não indicam a formação de complexos estáveis entre a lisozima e os polímeros estudados, sendo estes dados suportados também por outros resultados (Almeida e Oliveira et. al., 2004).

Nos estudos com variação de temperatura diversas conclusões puderam ser obtidas. Apesar de novamente não detectarmos a formação de complexos, interações não específicas foram constatadas entre a lisozima e os polímeros. Para o sistema lisozima+PEO3350 não constatou-se nenhum tipo de interação detectável à 60 e 70°C mas o caminho de agregação da lisozima (Figura 63) à 80° sofreu alterações indicando que a presença do polímero altera a forma dos agregados protéicos formados em decorrência da denaturação da proteína. Já para a lisozima+P105, a micelação decorrente de partes hidrofóbicas e hidrofílicas do polímero P105 permitiu determinar-se influência mútua das proteínas e dos copolímeros nos processos de formação das micelas e agregação da proteína. À 60 e 70° constatamos que a presença da proteína gerou alterações na forma e tamanho das micelas formadas quando em comparação com o polímero puro. Já à 80°C, apesar das micelas formadas em ambos os casos serem bastante parecidas, a velocidade de agregação da proteína foi fortemente diminuída, provavelmente em decorrência da presença das micelas no sistema. O modelo de polidispersidade de fatores de forma permitiu a interpretação dos dados de espalhamento para os sistemas binários estudados bem como descrever-se a variação do raio de giro total para o sistema lisozima+P105.

Sendo assim, nossos resultados sugerem que as indicações anteriores da formação de complexos entre polímeros hidrofílicos e proteínas devam ser revistas. No entanto, os resultados aqui apresentados sugerem a existência de diversos tipos de interações não específicas, ainda longe de serem completamente compreendidas, entre sistemas poliméricos e protéicos.

4. Estudo da Proteína Precursora Amilóide por SAXS^{†††}

a) Introdução

A doença de Alzheimer (AD) é a causa mais comum do declínio progressivo das funções cognitivas em pessoas idosas. Estima-se que AD afete entre 17-25 milhões de pessoas em todo o mundo. Nos países do ocidente, representa a quarta causa de morte após doenças do coração, câncer e derrame cerebral (Czeck et al., 2000). A combinação da crescente prevalência da doença em conjunto com longa sobrevivência (7 anos em média) após o diagnóstico da doença leva a cargas emocionais e econômicas muito pesadas para os membros da família e a sociedade. Nos EUA, por exemplo, os custos anuais diretos e indiretos associados à AD são estimados entre 80-100 bilhões de dólares (fonte: *National Institute of Aging, 1998*).

A proteína precursora amilóide (*amyloid precursor protein* – APP) está geneticamente e bioquimicamente ligada à gênese da AD. A APP humana é uma proteína bastante comum que transpassa a membrana plasmática somente uma vez, existindo em várias isoformas. As isoformas predominantes não neuronais, APP₇₅₁ e APP₇₇₀, contém o domínio inibidor de protease tipo Kunitz (Hynes et al., 1990), que participa na modulação de processos de coagulação (Botelho, Gralle e Oliveira et al., 2003). Já a forma

^{†††} Os resultados aqui apresentados formam parte das publicações:

- Gralle M, Botelho MM, **Oliveira CLP**, Torriani I and Ferreira ST (2002). “Solution studies and structural model of the extracellular domain of the human amyloid precursor protein”. *Biophys. J.*, 83, 3513-3524.
- Botelho MG, Gralle M, **Oliveira CLP**, Torriani I and Ferreira ST (2003). “Folding and stability of the extracellular domain of the human amyloid precursor protein”. *J. Biol. Chem.*, 278, 34259-34267.
- Gralle M, **Oliveira CLP**, Torriani IL and Ferreira ST. “Conformation of the extracellular domain of human amyloid precursor protein from solution X-ray scattering”, *in preparation*.

predominante em neurônios, APP₆₉₅, não possui este domínio inibidor de protease (De Strooper e Annaert, 2000).

A vida média da APP na membrana plasmática é <10min, sendo clivada extracelularmente ou reciclada em compartimentos intracelulares como vesículas endossômicas ou lisossomas. Quebras proteolíticas da APP podem também ocorrer intracelularmente, resultando em um conjunto de APP secretada (sAPP) em vesículas. A clivagem extracelular da APP ocorre predominantemente pela ação de uma α -secretase (Lammich et al., 1999), que quebra a APP em um sítio localizado à 13 aminoácidos a partir de sua inserção na membrana (Figura 68). A APP também pode ser clivada pela ação da β -secretase no sítio- β , localizado à 29 aminoácidos do ponto de inserção na membrana (Vassar et al., 1999). Estes fragmentos extracelulares solúveis gerados após a clivagem pela α -secretase (sAPP α) e β -secretases (sAPP β) podem ser detectados no meio extracelular. Outras clivagens podem ocorrer intra-membrana por γ -secretases (Durkin et al., 1999), liberando o peptídeo p3 ou o peptídeo β -amilóide para vesículas intracelulares ou o meio extracelular.

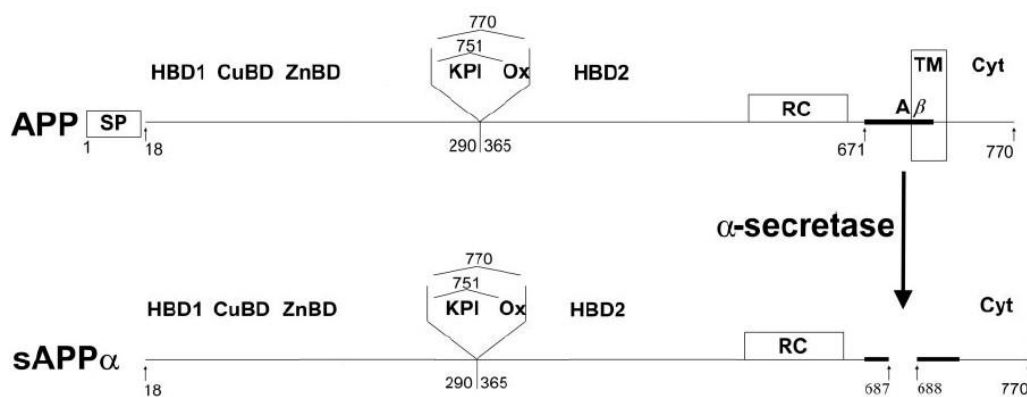


Figura 68 – Estrutura de domínios e clivagem da APP. APP: a proteína transmembrana inteira, incluindo o peptídeo sinal (SP), numerado como na APP₇₇₀. O domínio KPI com 56 resíduos e a sequência Ox2 com 19 resíduos (Ox). HBD1, HBD2: domínios ligadores de heparina. CuBD: domínio ligador de cobre. ZnBD: domínio ligador de zinco. RC: região predita como sendo tipo random coil. TM: domínio transmembrana. A β : sequência amiloidogênica que se insere parcialmente no domínio transmembrana. Cyt: domínio citoplasmático da APP. O diagrama mostra que a sAPP α ₇₇₀ é equivalente ao intervalo APP₁₁₈₋₆₈₇ (670 resíduos), enquanto que a sAPP α ₇₇₁ possui 651 resíduos e sAPP α ₆₉₅ possui 595 resíduos. Veja o texto para detalhes.

A AD é patologicamente caracterizada pela presença de um número muito grande de placas senis e fibras trançadas no cérebro, acompanhadas com uma perda maciça da atividade neuronal (De Felice e Ferreira, 2002). O peptídeo β -amilóide ($A\beta$), proveniente da proteína precursora amilóide é o maior componente destas placas e fibras. O domínio $A\beta$ forma agregados fibrilares não covalentes característicos tanto *in vitro* como *in vivo*, e a agregação e subsequente depósito no cérebro estão relacionadas à neurotoxicidade da AD. Além disso uma conexão causal entre a APP e a AD foi confirmada pela evidência genética de que mutações na sequência da APP causam hereditariedade levando a AD pré-senil (Goate et al., 1991). Além do interesse no que se refere à AD, muitas funções biológicas importantes foram descobertas para a APP. A APP transmembrana possui uma função importante no crescimento dos dendritos do axônio (Sabo et al., 2003), adesão neuronal, transporte axonal e metabolismo do cobre no organismo (Multhaup et al., 1999). A sAPP α , o principal fragmento da APP, pode ser neuroprotetor, modular a função sináptica e regular a proliferação de progenitores neuronais (Caille et al., 2004) e outros tipos de células.

A estrutura primária da APP já é conhecida à quase 20 anos e, apesar do grande interesse na correlação da APP com a AD e em muitos processos celulares importantes, sua estrutura tridimensional completa permanece desconhecida, além de seu mecanismo de ação molecular ainda ser um enigma. Recentemente, houve um substancial progresso na biologia estrutural de domínios isolados da APP (Barnham et al., 2003; Dulubova et al., 2004; Hynes et al., 1990; Kroenke et al., 1997; Rossjohn et al., 1999; Wang et al., 2004), mas estruturas em resolução atômica da sAPP α e APP ainda são desconhecidas. Na falta de estruturas em resolução atômica para a APP, apresentamos aqui uma caracterização da sAPP α por espalhamento de raios X a baixos ângulos. Conformação geral, estabilidade, processo de denaturação e modelos *ab initio* foram obtidos para a sAPP α em duas isoformas (a saber sAPP α_{695} e sAPP α_{770}) bem como a localização mais provável de fragmentos já determinados em resolução atômica no interior dos modelos de SAXS.

Mostraremos resultados pertinentes à vários experimentos. Inicialmente obtivemos uma estrutura em baixa resolução para a sAPP α_{695} (Gralle, Botelho e Oliveira et al., 2002). Posteriormente, utilizando medidas de SAXS em alta resolução foi possível obter-se informações sobre a estabilidade e enovelamento da sAPP α_{695} e sAPP α_{770} (Botelho, Gralle e Oliveira, 2003). Por fim, com os dados de alta resolução para as formas nativas modelos

ab initio de alta resolução puderam ser construídos, permitindo o posicionamento dos fragmentos com resolução atômica nos modelos de SAXS para a APP (Gralle e Oliveira et al., 2004).

b) Experimental

- ESTRUTURA EM BAIXA RESOLUÇÃO PARA O DOMÍNIO EXTRACELULAR DA PROTEÍNA PRECURSORA AMILÓIDE HUMANA

O processo de purificação desta proteína é descrito em diversos trabalhos (Gralle, Botelho e Oliveira et al., 2002). As amostras foram preparadas pelo grupo de pesquisas do professor Sérgio T. Ferreira, no Departamento de Bioquímica Médica da Universidade Federal do Rio de Janeiro e transportadas para Campinas para a realização dos experimentos.

Os experimentos de raios X a baixos ângulos foram realizados na linha de SAXS do laboratório Nacional de Luz Síncrotron (LNLS), com comprimento de onda de 1.488Å. Utilizou-se um porta amostras de mica com espaço entre janelas de 1mm (Cavalcanti et al., 2004), e a intensidade espalhada foi coletada com detetor linear sensível à posição. As soluções da sAPP α_{695} purificada, utilizada nos experimentos tiveram uma concentração máxima de 2.6mg/ml. A solução tampão utilizada foi 50mM Tris, pH7.4. As amostras foram mantidas à temperatura de 10°C durante as exposições. Uma segunda amostra de proteína mais diluída (~1.3mg/ml) também foi medida de modo a investigar possíveis efeitos de concentração e agregação. A aquisição dos dados foi feita tomando-se dez exposições de 15 minutos, o que permitiu um controle quanto a possíveis efeitos da radiação. Utilizando a distância de 445.1mm o intervalo do módulo do vetor transferência de momento acessível nos experimentos foi $0.02043\text{\AA}^{-1} < q < 0.2997\text{\AA}^{-1}$. O tratamento dos dados foi realizado com o programa TRAT1D (Oliveira, 2003). A análise dos dados foi feita utilizando os programas GNOM (Semenyuk e Svergun, 1991), SASHA (Svergun et al., 1996), DAMMIN (Svergun, 1999), DAMAVER (Volkov e Svergun, 2003) e HYDROPRO (de la Torre et al., 2000).

A curva de espalhamento com o ajuste teórico e cálculo da função distribuição de pares de distâncias $p(r)$ é mostrada na Figura 69. Pelo gráfico de Guinier ($\text{Log}(I) \times q^2$ - Figura 69A) vemos uma região linear o que indica uma monodispersidade satisfatória para

a amostra de proteína. O raio de giro estimado desta aproximação é de $R_g=35\text{\AA}$ e, além disso, não existe nenhum tipo de dependência do valor do raio de giro para com a concentração, como indicado pela na Figura 69A.

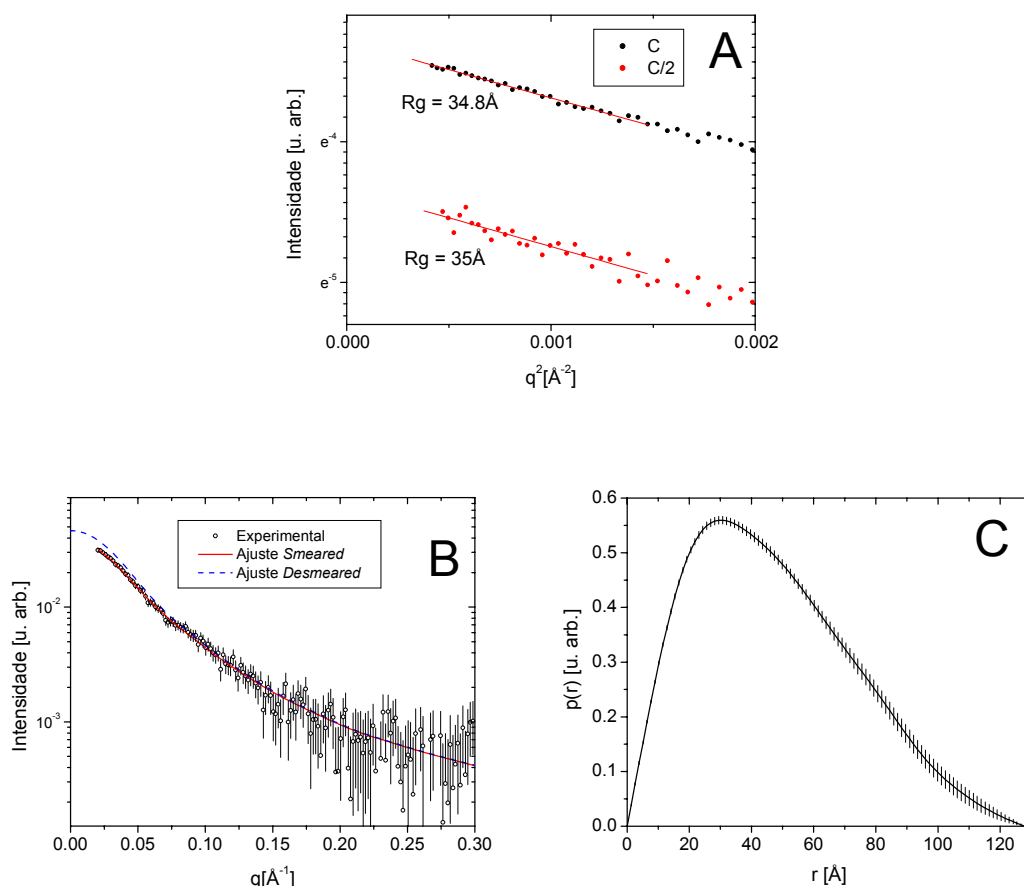


Figura 69 – Espalhamento a baixos ângulos para sAPP α_{95} . A) Gráfico de Guinier dos dados experimentais para determinação dos valores de raio de giro. Linhas sólidas: regressão linear. Linha superior: sAPP α_{95} à 2.6mg/ml. Linha inferior: diluição ~1:2. B) Intensidade de espalhamento dos dados experimentais e ajuste teórico. Linha sólidas: Ajuste teórico “distorcido” pelo efeito de colimação por fendas. Linha tracejada: Ajuste teórico corrigido por efeitos de colimação por fendas. C) Função de distribuição de pares de distâncias.

Do ajuste teórico e cálculo da função distribuição de distâncias $p(r)$ (Figura 69B,C) vemos que a proteína apresenta uma conformação prolata alongada, com raio de giro proveniente do ajuste dado por $R_g=37.9\pm0.8\text{\AA}$ e máxima dimensão de $130\text{\AA}$. O valor do raio de giro assim obtido está em boa concordância com o dado pela aproximação de Guinier. Obtivemos estimativas para o peso molecular desta proteína utilizando dois

padrões secundários, a proteína lisozima e água. A intensidade de espalhamento para estas amostras foram medidas nas mesmas condições experimentais. Utilizando a lisozima como padrão secundário obtemos o valor de $(62 \pm 3)\text{kDa}$ e com a água, $(65 \pm 4)\text{kDa}$. Considerando o caráter independente de cada um dos métodos utilizados para a estimativa do peso molecular, pode-se dizer que temos uma ótima concordância. Estes valores estão em boa concordância com o peso calculado a partir da sequência primária da sAPP α_{695} , que fornece o valor de 69kDa. A conclusão mais importante que obtemos do valor do peso molecular é que a sAPP α_{695} é indubitavelmente um monômero em solução mesmo na concentração utilizada (2.6mg/ml).

Como discutimos na introdução teórica desta tese, da curva de espalhamento podemos calcular o número de graus de liberdade (N_S - *Shannon channels*) associado aos dados experimentais. Para nossos experimentos obtemos $N_S = 12$. A baixa concentração de proteína utilizada nos experimentos não nos permitiu ter intensidades mensuráveis a altos valores de q . Como primeira indicação da forma da partícula em solução, utilizamos o método de expansão multipolar (programa SASHA). Uma aproximação inicial para a forma da partícula foi obtida ajustando-se o intervalo $0.02043\text{\AA}^{-1} < q < 0.17\text{\AA}^{-1}$ ($N_S \sim 6$).

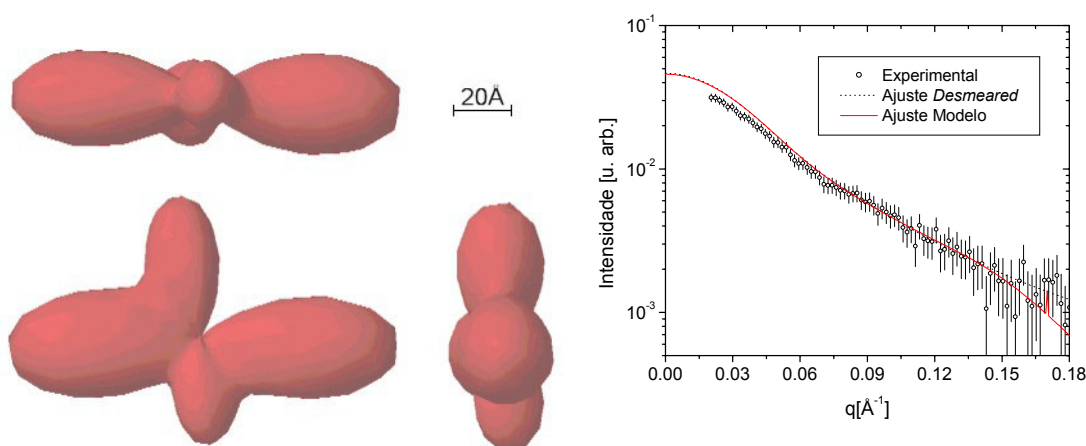


Figura 70 – Cálculo *ab initio* para sAPP α_{695} utilizando esféricos harmônicos. Esquerda) modelo *ab initio* obtido por expansão multipolar. Direita) Ajuste dos dados experimentais pelo modelo calculado.

Apesar das limitações do método, vemos que o modelo indica uma partícula alongada, confirmando o discutido com base na função $p(r)$. Para obtermos uma melhor aproximação para a forma da partícula, utilizamos modelagem tipo *bead model* (programa DAMMIN), no modo “lento” de modo a nos aproximarmos mais do mínimo global. O

ajuste foi feito em todo o intervalo medido ($N_s = 12$) e a média de dez modelos calculados independentemente foi obtida utilizando o programa DAMAVER.

O modelo obtido é mostrado na Figura 71 juntamente com o respectivo ajuste dos dados experimentais.

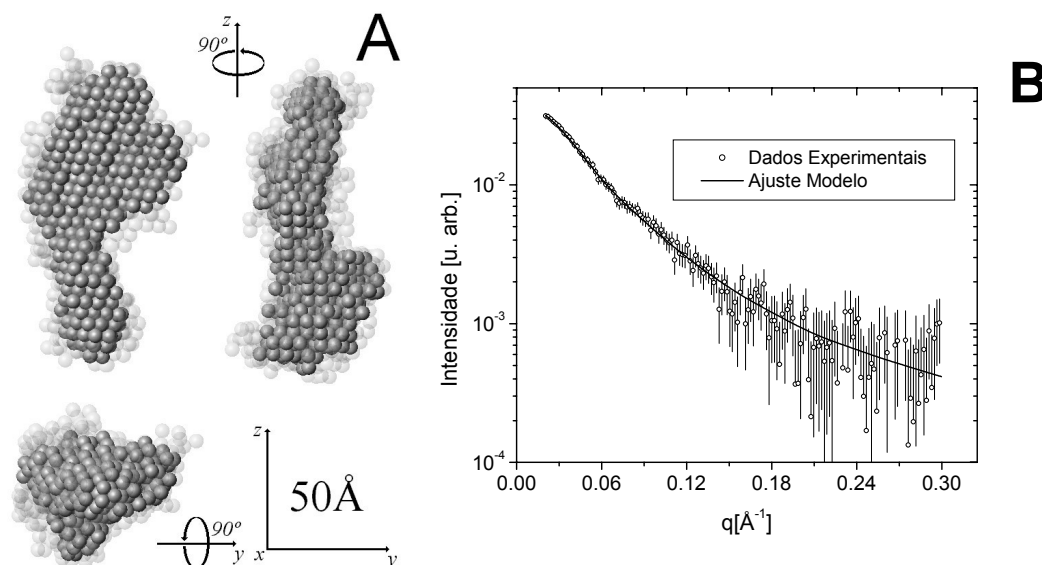


Figura 71 – Cálculo *ab initio* para sAPP α_{695} utilizando modelagem tipo *bead model*. A) Modelo mais provável proveniente da promediação de 10 modelos independentes. B) Ajuste para um dos modelos.

As características gerais deste modelo são semelhantes às obtidas para o método de esféricos harmônicos, mas agora temos indicações melhores sobre a forma da partícula, como por exemplo, a presença de, pelo menos, dois domínios bem definidos na estrutura da partícula. Como veremos posteriormente, os dados de SAXS em alta resolução confirmam estes resultados. Apesar das limitações inerentes à técnica, estes resultados de SAXS forneceram pela primeira vez informações estruturais e conformacionais da estrutura do fragmento solúvel sAPP α_{695} da proteína precursora amilóide humana.

Pelos resultados anteriores pudemos obter uma estimativa para a forma geral do fragmento solúvel da APP humana. Informações adicionais somente puderam ser obtidas devido a possibilidade de utilizarmos detectores bidimensionais, o que permitiu uma melhoria considerável da estatística dos dados a altos valores de q . Estes resultados são mostrados a seguir.

**- ENOVELAMENTO E ESTABILIDADE DO DOMÍNIO EXTRACELULAR DA
PROTEÍNA PRECURSORA AMILÓIDE HUMANA**

Similarmente ao fato de ter-se pouca informação estrutural para a proteína precursora amilóide APP, muito pouco se conhece de sua estabilidade em solução, quando comparado ao que já se obteve em aspectos patológicos relacionados à Doença de Alzheimer (AD). Desta maneira, utilizando espalhamento de raios X a baixos ângulos em conjunto com diversos outros resultados dados por outras técnicas, pudemos obter diversas informações concernentes à estabilidade e enovelamento da proteína precursora amilóide em duas isoformas, sAPP α_{695} e sAPP α_{770} .

O processo de expressão e purificação destas isoformas é descrito em diversos artigos. Os experimentos de raios X a baixos ângulos foram realizados na linha de SAXS do LNLS, com comprimento de onda de 1.488Å, utilizando porta-amostra de capilar com diâmetro de 1.5mm (Cavalcanti et al., 2004) com temperatura controlada à 10°C durante os experimentos. A coleta dos dados foi feita utilizando-se um detetor à gás bidimensional sensível à posição (Gabriel, 1982). A solução tampão utilizada foi 50mM Tris-HCl, pH7.4, com e sem denaturantes, como indicado na Tabela 15. As soluções de Tris-HCl, Hidrocloreto de Guanidina e Uréia foram obtidas da Sigma. Água Milli Q foi utilizada em todas as etapas do experimento. A aquisição dos dados experimentais foi feita tomando-se 3 exposições de 10min. Três distâncias amostra-detetor foram utilizadas, 1280mm, 691mm e 378mm o que permitiu-se obter um intervalo experimental de $0.0328\text{\AA}^{-1} < q < 0.8538\text{\AA}^{-1}$ para a sAPP α_{695} e $0.0241\text{\AA}^{-1} < q < 0.8609\text{\AA}^{-1}$ para a sAPP α_{770} (amostras sem aditivos). O tratamento dos dados foi feito utilizando-se o programa TRAT2D (Oliveira, 2004), que permite o tratamento de dados bidimensionais bem como integração angular, o que melhora bastante a estatística experimental. O ajuste dos dados tratados foi feito utilizando o programa GNOM.

Tabela 15 – Amostras estudadas por SAXS, concentração e solução tampão utilizada.

Amostra	Concentração	Solução Tampão
APP695 _{pura}	12.5	Buffer A
APP770 _{pura}	9.6	Buffer A
APP695 _{250mMGdn}	3.6	Buffer B
APP770 _{250mMGdn}	3.2	Buffer B
APP695 _{2MGdn}	3.6	Buffer C
APP770 _{2MGdn}	3.2	Buffer C
APP695 _{6MUreia}	3.6	Buffer D
APP770 _{6MUreia}	3.2	Buffer D

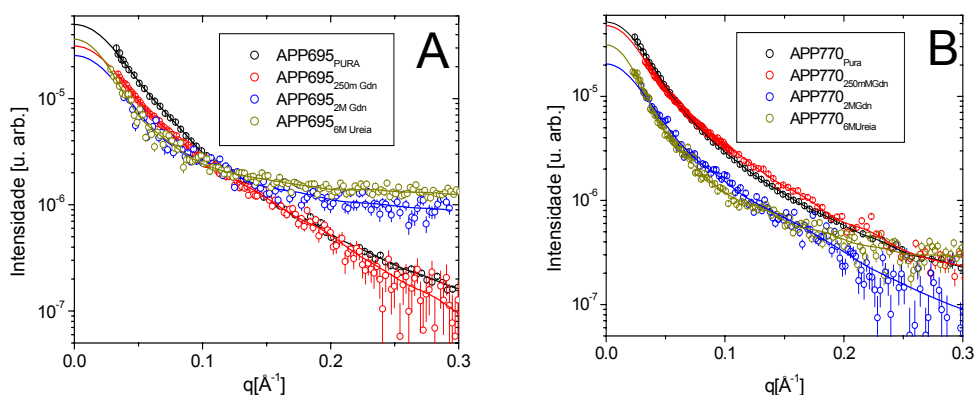
Buffer A: 50mM TrisHCl.

Buffer B: 50mM TrisHCl + 250mM Guanidina Hidroclorídrica (GdnHCl).

Buffer C: 50mM TrisHCl + 2M Guanidina Hidroclorídrica (GdnHCl).

Buffer D: 50mM TrisHCl + 6M Uréia.

Os dados experimentais e ajuste teórico são mostrados na Figura 72. Nas Figura 72A,B mostramos a intensidade de espalhamento para as proteínas nativas e com denaturantes. Nas Figura 72C,D mostramos a transformada de Fourier das intensidades de espalhamento, a função $p(r)$ e nas Figura 72E,F mostramos os respectivos gráficos de Kratky para as intensidades. Do valor da intensidade na origem, $I(0)$, obtemos as massas moleculares para a sAPP α_{695} e sAPP α_{770} como sendo (67 ± 3) e (74 ± 3) kDa respectivamente, utilizando uma amostra de albumina (66kDa) como padrão secundário. A diferença de ~ 7 kDa no valor da massa molecular está em ótima concordância com o fato da isoforma sAPP α_{770} conter 75 aminoácidos a mais do que a sAPP α_{695} .



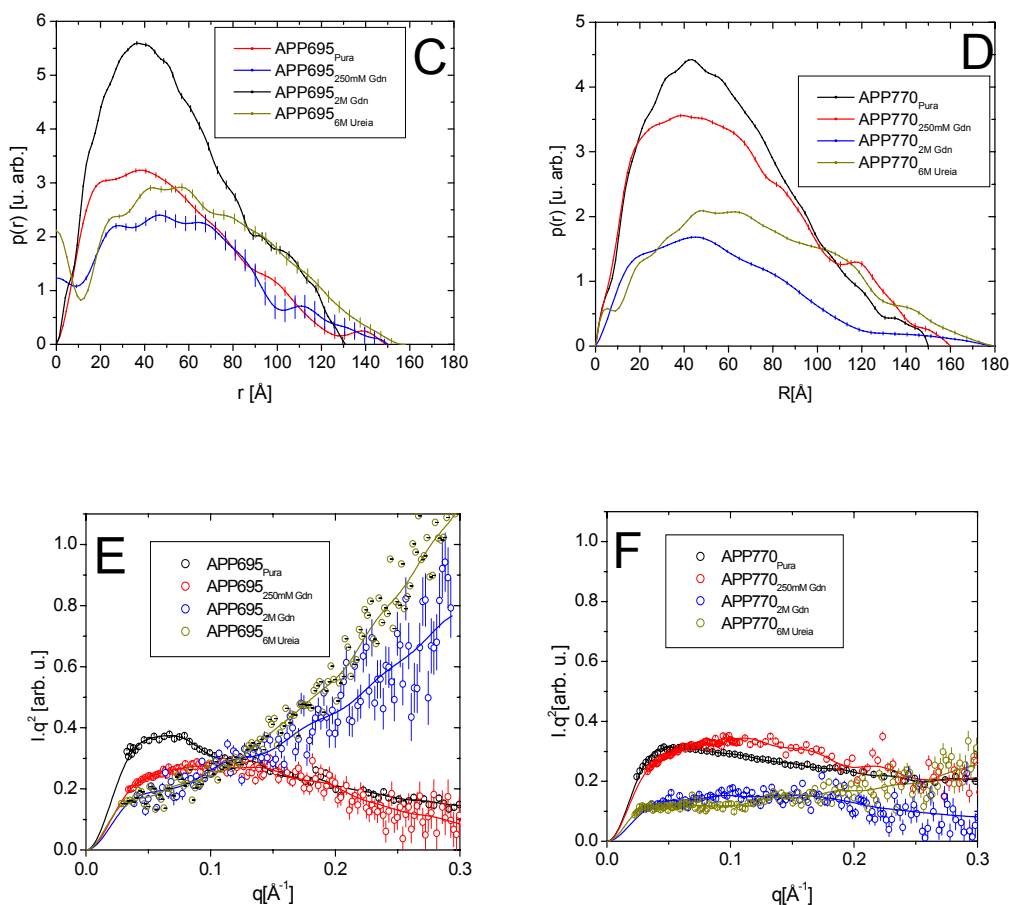


Figura 72 – Experimentos de SAXS para sAPP α pura e com agentes denaturantes. Curvas de intensidade de espalhamento (A,B), função de distribuição de distâncias (C,D) e gráficos de Kratky (E,F) para amostras de sAPP α_{695} e sAPP α_{770} puras e com aditivos conforme descrito na Tabela 15.

Para as duas isoformas o espalhamento proveniente da amostra com 250mM de GdnHCl é muito parecido ao fornecido pela proteína nativa (Figura 72A,B). Na presença de maiores concentrações de denaturante, as curvas de espalhamento para as duas isoformas de sAPP α mudam consideravelmente, indicando alterações significativas na forma e dimensões da partícula. Isso é melhor visto nas funções $p(r)$ (Figura 72C,D) onde temos uma grande mudança na forma da curva com a adição dos agentes denaturantes.

Os valores de raios de giro (R_g) para as duas isoformas em seu estado nativo e com a adição de denaturantes são mostrados na Tabela 16. O valor de R_g para a estrutura nativa da sAPP α_{695} obtida nestas medidas em alta resolução (42.5 $\text{\AA}$) com detetor bidimensional é ligeiramente maior do que o obtido anteriormente com o detetor unidimensional (~38 $\text{\AA}$).

Visto que os dados obtidos agora possuem maior resolução, correspondendo a uma maior região do espaço recíproco, devem fornecer valores mais confiáveis de Rg. Para a sAPP α_{770} nativa obtemos um valor de raio de giro 4Å maior do que a sAPP α_{695} nativa, o que está em concordância com o fato de, uma vez possuindo maior massa molecular, deva também possuir tamanho maior. Discutiremos detalhadamente os estados nativos destas isoformas, bem como a geração de modelos, no próximo item. A adição de 2M GdnHCl causa uma variação muito pequena no raio de giro da sAPP α_{695} e um aumento <2Å no raio de giro para sAPP α_{770} . Por outro lado, a adição de 6M uréia causa um aumento de 9.5 Å no raio de giro da sAPP α_{695} e 11.3Å para sAPP α_{770} .

Tabela 16 – Raios de giro para as duas isoformas de sAPP α na condição nativa e com aditivos.

Amostra	Rg [Å]	
	sAPP α_{695}	sAPP α_{770}
sAPP α Nativa	42.5 $\pm$ 0.3	46.5 $\pm$ 0.5
sAPP α + 250mM GdnHCl	43 $\pm$ 1	49.3 $\pm$ 0.6
sAPP α + 2M GdnHCl	44 $\pm$ 4	48 $\pm$ 1
sAPP α + 6M Uréia	52 $\pm$ 1	57.8 $\pm$ 0.5

Apesar de mudanças no raio de giro indicarem mudanças estruturais na proteína, para termos uma melhor descrição das alterações estruturais decorrentes dos denaturantes químicos podemos analisar os gráficos de Kratky conforme o apresentado na introdução teórica. As duas isoformas da sAPP α nativa apresentaram máximos bem definidos em baixos valores de q nos seus correspondentes gráficos de Kratky (Figura 72E,F), indicando que as partículas são relativamente compactas. No entanto, como as intensidades de espalhamento não decaem rapidamente à altos valores de q , como seria esperado para partículas estritamente compactas, temos uma indicação de que as isoformas de sAPP α possuem níveis de flexibilidade em sua estrutura mesmo no estado nativo. Como cerca de 30% da estrutura secundária da proteína não é composta de estruturas padrão (hélices α a ou folhas β) esta flexibilidade é esperada (Botelho, Gralle e Oliveira et al., 2003).

Na presença de 250mM de GdnHCl o máximo do gráfico de Kratky para ambas as isoformas desloca-se para maiores valores de q , mas as curvas ainda decaem com o aumento de q . Por outro lado, na presença de 2M de GdnHCl ou 6M de uréia as curvas não

apresentam nenhum máximo mas sim um platô inicial em torno de $q=0.09\text{\AA}^{-1}$. Este platô é típico de estruturas tipo cadeias gaussianas (*gaussian coils*, Glatter e Kratky, 1982). Além disso, as curvas aumentam com o aumento de q , indicando a presença de cadeias livres tipo *random coil* em sua estrutura. Desta forma, os gráficos de Kratky para a sAPP α na presença de 2M GdnHCl ou 6M de uréia indicam que as duas isoformas encontram-se desenoveladas, tendendo a um estado de completa denaturação. No entanto, nota-se que existem diferenças nas curvas para as duas isoformas. Para a sAPP α_{770} temos um aumento mais suave, indicando que esta isoforma não apresentou o mesmo nível de desenovelamento que a outra isoforma na mesma concentração de denaturantes. Como as maiores diferenças na curva de espalhamento ocorrem para $q > 0.1\text{\AA}^{-1}$, o que corresponde a detalhes da ordem de $\sim 30 - 60\text{\AA}$, temos a indicação de que para a isoforma sAPP α_{770} alguns domínios desta ordem de dimensão ainda não estavam completamente denaturados, mesmo com a perda da conformação geral da proteína. Por outro lado, para 6M de uréia o aumento da curva no gráfico de Kratky nos indica que a sAPP α_{770} assume uma conformação tipo *random coil* semelhantemente à outra isoforma, mas não ao mesmo nível de randomicidade, como indicado pela diferença nas inclinações.

Os resultados experimentais permitiram um estudo detalhado da estabilidade das isoformas da proteína precursora amilóide na presença de denaturantes. Dados de dicroísmo circular (Botelho, Gralle e Oliveira et al, 2003) indicam que a estrutura secundária destas isoformas apresenta uma quantidade substancial de estruturas em α -hélice (>35%) mas também uma quantidade significativa (>30%) de estruturas secundárias não padrão. O espectro de dicroísmo circular para a sAPP α_{695} é mostrado na Figura 73.

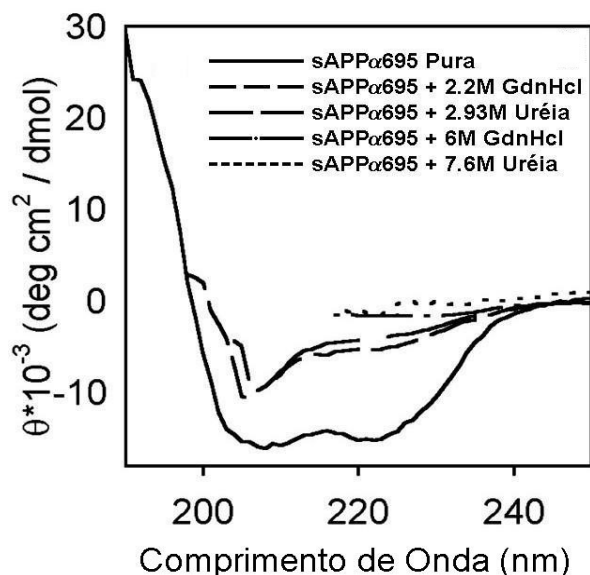


Figura 73 – Desenovelamento da sAPP α_{695} utilizando dicroísmo circular. Espectro para sAPP α_{695} nativa e com agentes denaturantes. Figura tirada de (Botelho, Gralle e Oliveira et al, 2003).

Vemos que a proteína apresenta estados intermediários parcialmente enovelados na presença de 2-3M GdnHCl. Isto está em ótima concordância com o indicado pelos gráficos de Kratky, onde vimos que a sAPP α em 2M de GdnHCl perde grande parte de sua estrutura compacta, como discutido acima, pela ausência de máximos a baixos ângulos (Figura 72E,F). No entanto, o raio de giro da sAPP α em 2M GdnHCl é similar ao obtido para a estrutura nativa (Tabela 16). Isto sugere que, apesar da proteína perder sua conformação terciária, como indicado pela forma da função $p(r)$ e gráfico de Kratky, a presença de domínios parcialmente desenovelados não promove um aumento significativo do raio de giro, como seria esperado para uma estrutura tipo *random coil* com este número de aminoácidos (Botelho, Gralle e Oliveira et al, 2003).

Por outro lado, um equilíbrio distinto é obtido nas curvas de desenovelamento na presença de uréia. Diferentemente do caso com GdnHCl, a uréia não estabiliza estados parcialmente desenovelados da APP, como indicado na Figura 73. Já em ~3M de uréia têm-se a proteína sem nenhuma estrutura secundária padrão, indicando seu completo desenovelamento nesta concentração. Isto nos indica que o desenovelamento da sAPP α não ocorre da mesma forma para diferentes agentes denaturantes. Enquanto que para a Uréia aparentemente temos somente dois estados (enovelado / desenovelado), para GdnHCl

temos estados intermediários parcialmente enovelados, com diferentes níveis de compacidade. O efeito da GdnHCl na estabilidade da sAPP α em estados intermediários parece dever-se a interações mais específicas de cátions guanidina com regiões da proteína (Botelho, Gralle e Oliveira et al, 2003).

Outro fator importante a mencionar é o fato de que a adição de guanidina aumentou em muito a absorção de raios X pelas amostras devido. Isso ocorre devido à absorção da radiação pelos ânions cloro neste solvente. Enquanto que para a proteína nativa tínhamos uma absorção de fator ~ 3 , com 2M de guanidina tivemos um fator ~ 7 e à 6 M de GdnHCl (dados não mostrados) um fator ~ 20 . Desta forma as medidas somente puderam ser efetuadas graças ao ganho estatístico decorrente da aquisição dos dados ter sido feita em detetores bidimensionais e sua posterior integração angular. Testes com detetor unidimensional nas mesmas condições experimentais não forneceram resultados mensuráveis mesmo para baixas concentrações de GdnHCl.

Finalmente, os dados de SAXS nos indicaram diferenças no processo de denaturação das duas isoformas. Além disso, para baixa concentração de GdnHCl tivemos a indução de uma conformação globular intermediária para esta proteína sem perda substancial de compacidade. Alternativamente, nossos resultados indicam a presença de estados intermediários de desenovelamento, bem como a presença de regiões flexíveis na sAPP α mesmo em condições nativas. Isso sugere que esta proteína possua plasticidade conformacional suficiente para permitir sua adaptação à diferentes ligantes, fazendo com que possa desempenhar diferentes funções fisiológicas no organismo.

- CONFORMAÇÃO DO DOMÍNIO EXTRACELULAR DA PROTEÍNA PRECURSORA AMILÓIDE A PARTIR DO ESPALHAMENTO A BAIXOS ÂNGULOS

No item anterior focamos nossa atenção na estabilidade e enovelamento da sAPP α baseando-nos nos dados de espalhamento da proteína na presença de denaturantes. No entanto, os dados de alta resolução para as formas nativas das isoformas da sAPP α nos permitem obter maiores informações estruturais sobre esta proteína bem como realizar uma modelagem mais apurada.

As intensidades de espalhamento e funções de distribuição de pares de distâncias para as proteínas nativas são mostradas na Figura 74:

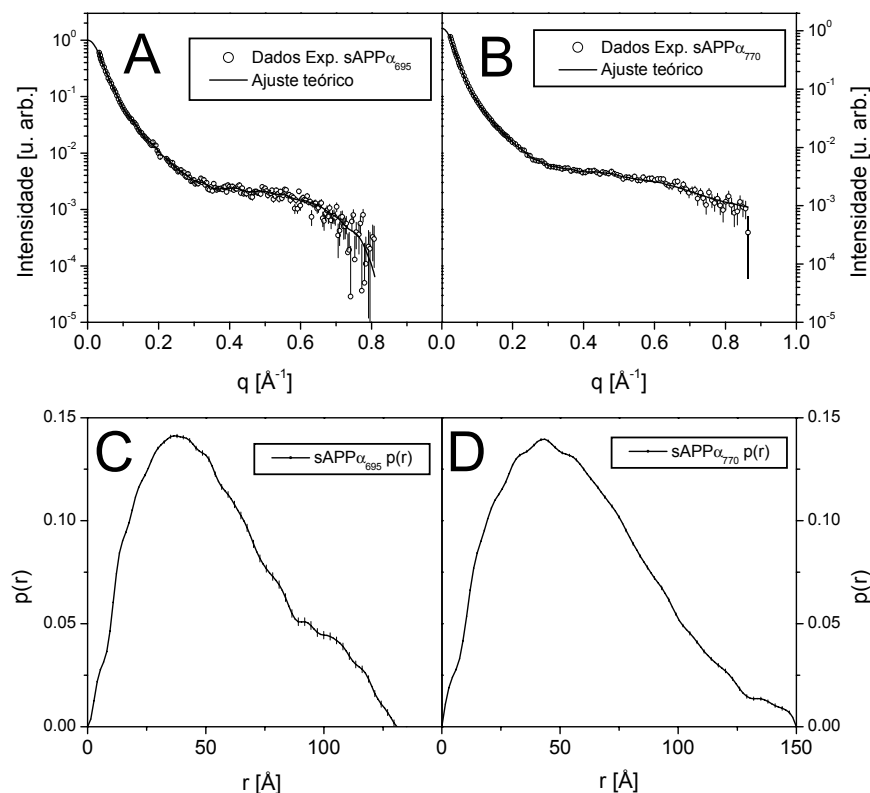


Figura 74 – Experimentos de SAXS em alta resolução para sAPP α . Linha superior, Intensidade espalhada para sAPP α_{695} (A) and sAPP α_{770} (B) em todo o intervalo angular medido. Círculos abertos: dados experimentais. Linhas sólidas: Ajuste teórico. Linha inferior, função distribuição de pares de distâncias calculadas a partir dos dados experimentais para sAPP α_{695} (C) e sAPP α_{770} (D).

Como mencionado no item anterior, para as proteínas nativas pudemos ter dados de espalhamento até $q \sim 0.8 \text{ \AA}^{-1}$, o que nos fornece dados com grande quantidade de informação bem como de alta resolução. Para a sAPP α_{695} temos que os dados de espalhamento correspondem à $N_s = 37$ e para a sAPP α_{770} $N_s = 41$, com resolução $(2\pi/q_{\max}) \sim 7 \text{ \AA}$ nos dois casos. Como discutido na introdução teórica, esta quantidade de informação e resolução permite a determinação de domínios protéicos bem como uma descrição precisa da estrutura terciária (ou quaternária) da proteína. Além disso, com base em diversos domínios

já resolvidos à nível atômico para a proteína precursora amilóide, pudemos propor o posicionamento destes domínios na estrutura ab initio obtida dos dados de SAXS.

- Estruturas primária e secundária para a sAPP α_{695} e sAPP α_{770}

A sequência primária das isoformas da sAPP α e o posicionamento dos domínios protéicos resolvidos a nível atômico na sequência primária são mostrados abaixo

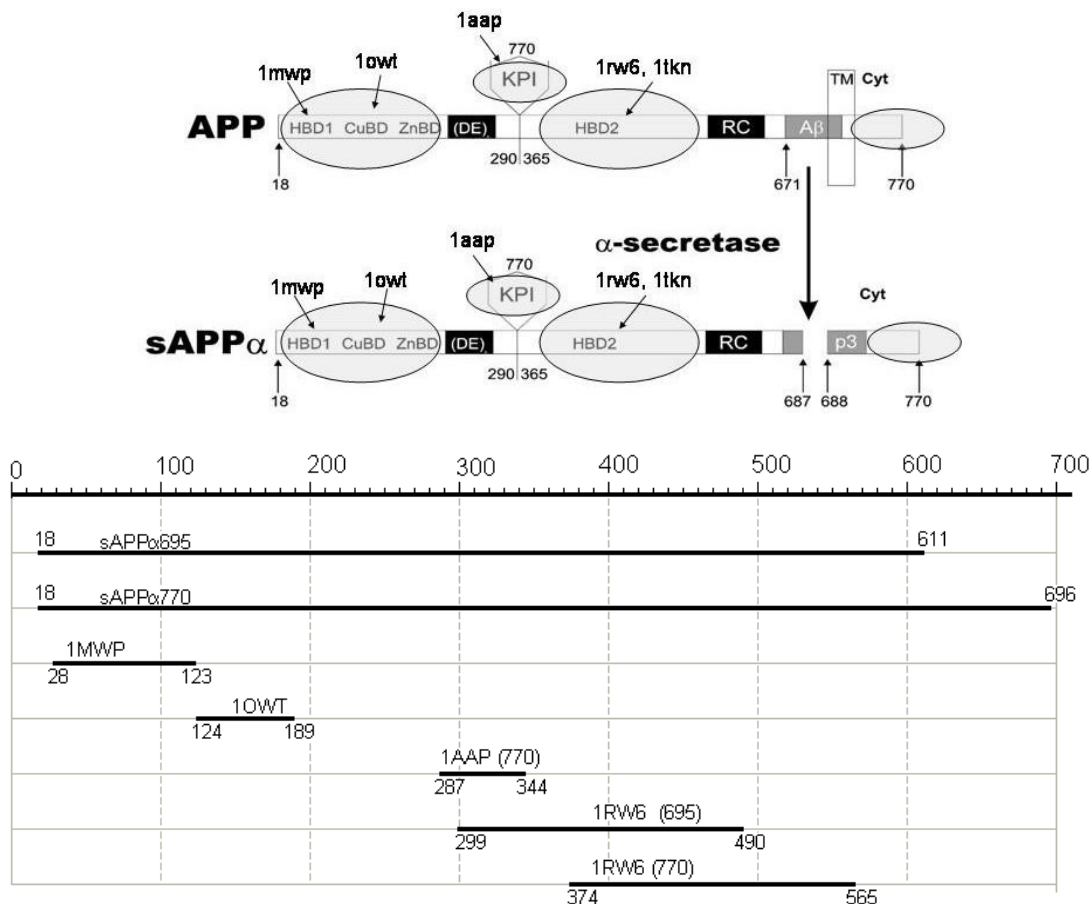


Figura 75 – Estrutura de domínios da APP e sAPP α Acima: Elipses cinza: Domínios conservados na família de isoformas de APP. Caixas pretas: parcelas não conservadas sugerindo possuir estruturas secundárias não padrão. A β : sequência amiloidogênica. TM: domínio transmembrana. 1mwp etc.: localização das estruturas com resolução atômica indicadas pelo número no PDB. Abaixo: Identificação similar à anterior em forma esquemática, identificando a posição na sequência de cada domínio.

As parcelas 1MWP e 1OWT são comuns às duas isoformas e ocorrem na mesma posição nas duas sequências. Já a parcela 1AAP ocorre somente na proteína APP770. A parcela 1RW6 ocorre nas duas proteínas, mas em posições distintas da sequência primária.

Utilizando o programa PSIPRED (McGuffin et al., 2000) obtemos a predição de estrutura secundária para as proteínas 695 e 770 (Figura 76). Isso é importante para avaliar a quantidade de partes não estruturadas (*coils*) na sequência.

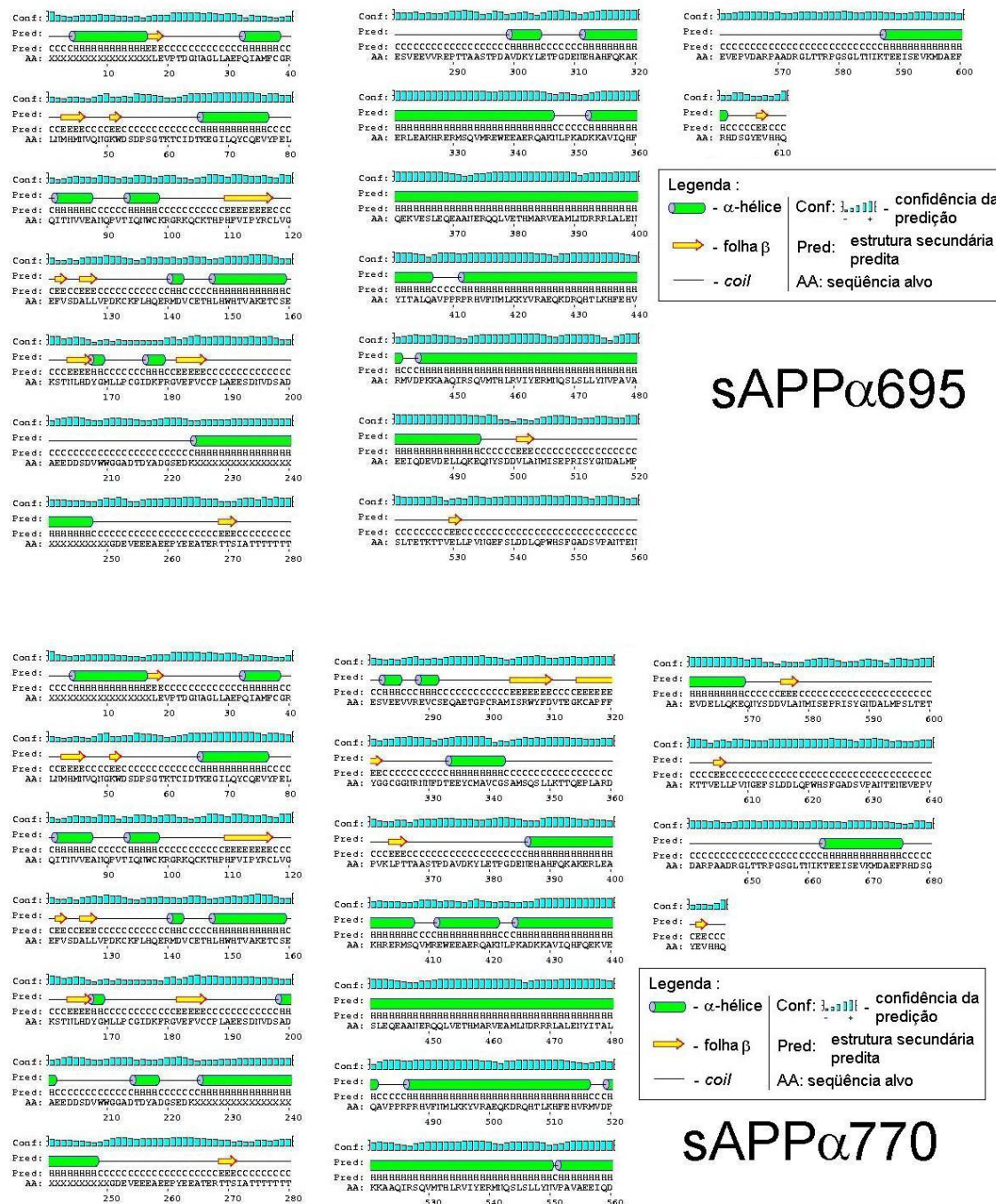


Figura 76 – Estrutura secundária predita para as isoformas sAPP α ₆₉₅ e sAPP α ₇₇₀ com base na sequência primária. Para a predição utilizou-se o programa PSIPRED.

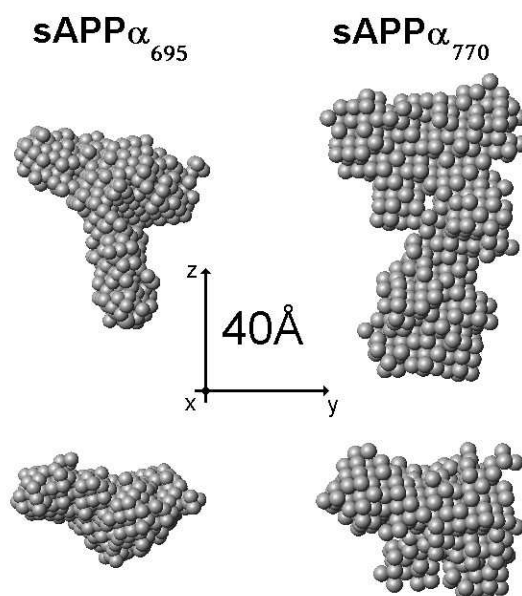
Como mencionado anteriormente, a grande quantidade de *coils* na estrutura (o que é confirmado por experimentos de CD) faz com que as estruturas da APP tenham uma flexibilidade intrínseca.

- Cálculos *Ab initio*

A este nível de resolução e informação dos dados experimentais podemos gerar modelos tipo *chain model* e tentar obter informações de sua estrutura de domínios. Como mencionado na introdução teórica, os modelos tipo *chain model* podem ser calculados ajustando-se tanto a intensidade de espalhamento quanto a curva de distribuição de distâncias $p(r)$. A escolha de uma ou outra forma de cálculo depende da qualidade dos modelos gerados. Para melhor compreensão, analisaremos as duas formas de cálculo:

- Modelos gerados com ajuste no espaço recíproco:

Nestes cálculos o programa GASBOR ajusta a curva de espalhamento experimental. A figura a seguir é o resultado de 10 ajustes independentes, com o modelo mais provável destacado:



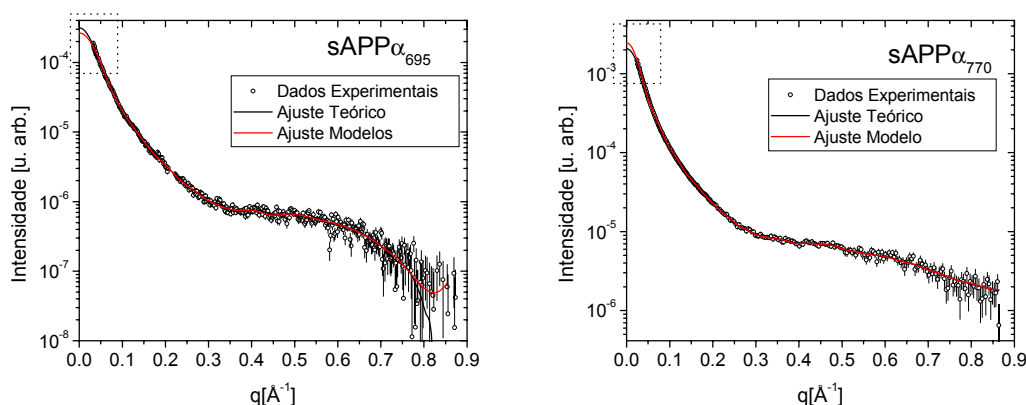


Figura 77 – Modelagem *ab initio* em alta resolução para as proteínas sAPP α_{695} e sAPP α_{770} . Acima: Modelos *ab initio* e ajuste dos dados experimentais de SAXS para amostras de sAPP α_{695} e sAPP α_{770} . Aqui a curva de intensidade de espalhamento é ajustada para o cálculo dos modelos. Os ajustes são mostrados abaixo dos modelos.

Apesar do ajuste da curva ser muito bom, a *parte inicial* não é corretamente ajustada, conforme indicado pelas caixas tracejadas. Como descrito na introdução teórica, a região inicial de uma curva de espalhamento descreve a dimensão (raio de giro) e forma geral da partícula espalhadora. Desta maneira um ajuste errado desta parte levará a aberrações nos modelos. Para avaliação dos modelos calculados utilizei o programa HYDROPPO para calcular o raio de giro dos modelos *ab initio* calculados. O resultado é mostrado na Tabela 17:

Tabela 17 – Cálculos hidrodinâmicos para modelos *ab initio* das isoformas de sAPP α

		Rg [Å]
sAPP α_{695}		
Valor Experimental		42.6 $\pm$ 0.3
Modelo <i>ab initio</i>		36.9 $\pm$ 0.7
sAPP α_{770}		
Valor Experimental		46.4 $\pm$ 0.1
Modelo <i>ab initio</i>		50.7 $\pm$ 0.5

Vemos que o ajuste errado da região inicial se refletiu em modelos *ab initio* com tamanhos diferentes dos valores experimentais para as proteínas. Isso se deve ao fato de, proporcionalmente, a região inicial é muito pequena em comparação com a curva toda de espalhamento. Desta maneira, este método não fornece resultados satisfatórios e outra metodologia deve ser aplicada à modelagem.

- *Modelos gerados com ajuste no espaço real:*

Teoricamente, a curva de espalhamento e sua transformada de Fourier, a função de distribuição de pares de distâncias $p(r)$ contém a mesma quantidade de informação. No entanto o enfoque dado em cada representação é diferente. No *espaço recíproco*, aumentando-se o vetor de espalhamento, teremos informações sobre detalhes cada vez menores da partícula espalhadora, com a resolução experimental dada pela relação $2\pi/q_{\max}$. Já no *espaço real* temos uma maior influência da região inicial da curva. A integral da função $p(r)$ está relacionada ao valor do raio de giro da partícula e ao valor da intensidade na origem $I(0)$ (equação (60)), além do fato de que a máxima dimensão da partícula ser dada principalmente pela região da intensidade espalhada próxima à origem. Sendo assim, um cálculo *ab initio* baseado no ajuste da função $p(r)$ deverá fazer com que os modelos calculados mantenham os valores de raio de giro e máxima dimensão corretos.

Devido aos fatores apresentados, os cálculos foram refeitos utilizando uma outra versão do programa GASBOR que ajusta a curva de distribuição de distâncias $p(r)$ e não a curva de espalhamento. O resultado é mostrado a seguir onde fiz um total de 20 modelos (ajustes independentes) e o cálculo do modelo mais provável com o programa DAMAVER. Os resultados são mostrados na Figura 78.

695

770

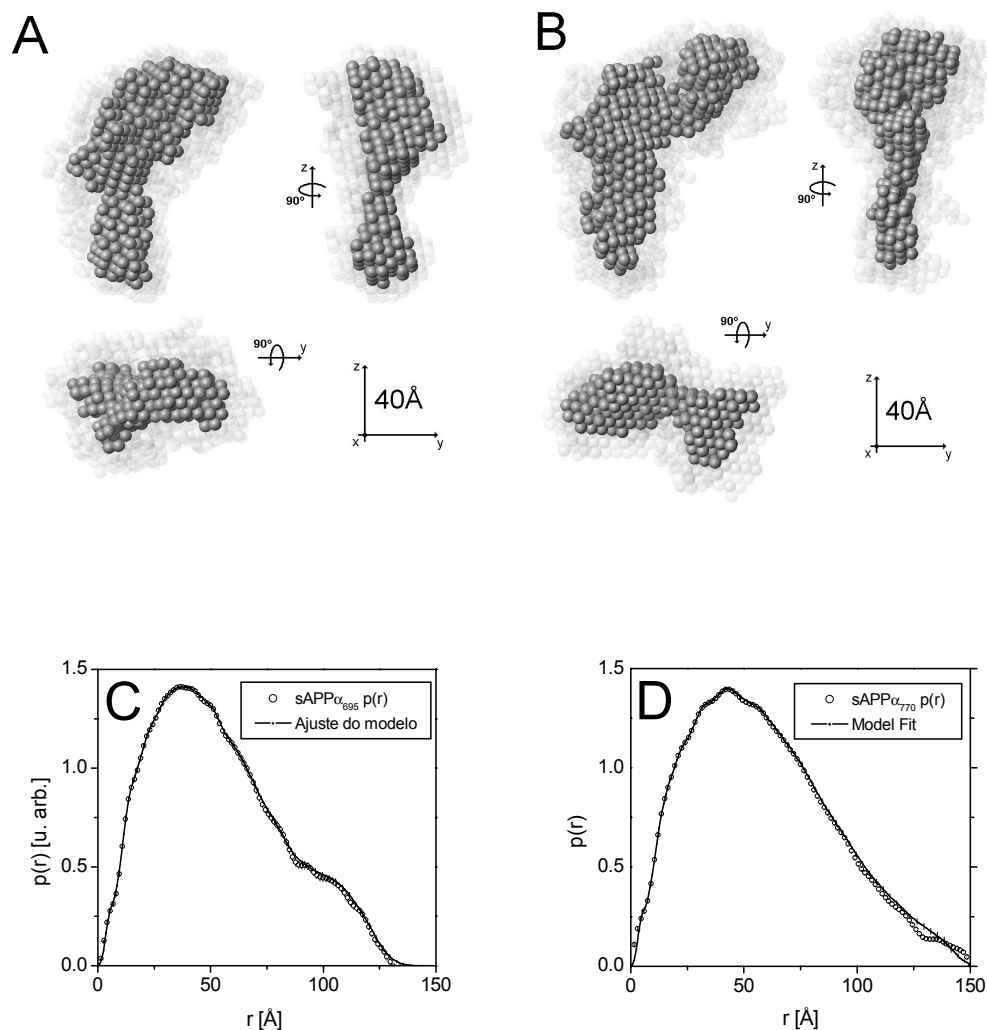


Figura 78 – Modelagem ab initio em alta resolução para as proteínas sAPP α_{695} e sAPP α_{770} . Modelos ab initio e ajuste dos dados experimentais de SAXS para amostras de sAPP α_{695} (A) e sAPP α_{770} (B). Aqui a curva de distribuição de distâncias é ajustada para o cálculo dos modelos. Os ajustes são mostrados em (C) e (D).

Vemos um bom ajuste da curva de distribuição de distâncias em ambos os casos. Calculando-se o raio de giro dos modelos ab initio com o programa HIDROPRO temos os resultados abaixo:

Tabela 18 – Cálculos hidrodinâmicos para modelos *ab initio* das isoformas de sAPP α

	Rg [Å]
sAPPa695	
Valor Experimental	42.6 ± 0.3
Modelo <i>ab initio</i>	42.9 ± 0.4
sAPPa770	
Valor Experimental	46.4 ± 0.1
Modelo <i>ab initio</i>	46.9 ± 0.6

Como esperado, os valores de raio de giro neste caso são praticamente idênticos aos valores experimentais, indicando que os erros de cálculo ocorridos no caso anterior foram agora corrigidos. Comparando-se os modelos *ab initio* vemos que existe uma grande similaridade entre as isoformas. A principal diferença ocorre na extremidade superior direita dos modelos onde, na isoforma sAPP α_{770} vemos uma saliência (Figura 79). Com base na seqüência primária das isoformas, vemos que esta diferença deve corresponder à influência da parcela correspondente à região de 75 aminoácidos que difere a seqüência das duas isoformas, região esta que contém a parcela 1AAP que possui estrutura resolvida a nível atômico.

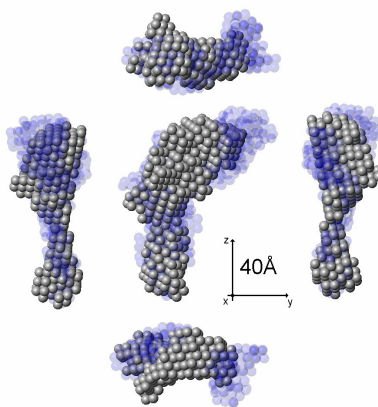


Figura 79 – Superposição dos modelos *ab initio* obtidos para a proteína sAPP α_{695} e sAPP α_{770} . sAPP α_{695} (esferas sólidas); sAPP α_{770} (esferas semitransparentes).

Para confirmar a presença destes 75 aminoácidos nesta região da proteína, podemos tentar adicionar um domínio com estes 75 resíduos ao modelo da sAPP α_{695} e compará-lo ao modelo da sAPP α_{770} (Figura 80). Para isso, gerei um domínio globular utilizando o

programa BUILDER (Oliveira,2001) composto de 75 resíduos podendo assim compor a diferença entre as proteínas. Como as proteínas possuem uma grande diferença na máxima dimensão, temos a clara indicação que esta parcela de 75 resíduos deve estar em uma das extremidades permitindo assim que se tenha um acréscimo da máxima dimensão.

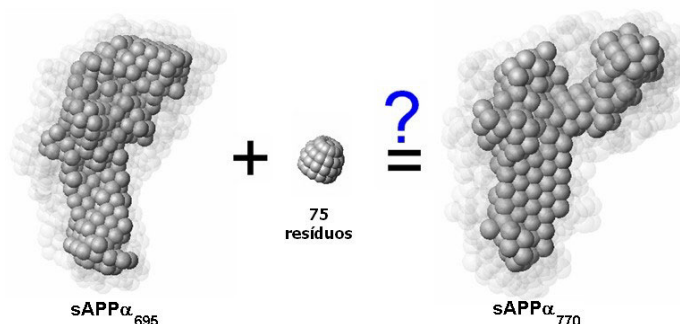


Figura 80 – Proposta para adição de um domínio globular de 75 resíduos ao modelo ab initio da $sAPP\alpha_{695}$ de modo a representar o modelo ab initio da $sAPP\alpha_{770}$.

Para realizar a comparação adicionamos o "domínio" (*dummy_prot*) em duas extremidades (superior e inferior) do modelo da $sAPP\alpha_{695}$ e o comparamos com o modelo da $sAPP\alpha_{770}$. Para podermos avaliar quantitativamente cada um dos casos ajustei a curva de espalhamento do modelo da $sAPP\alpha_{770}$ utilizando a intensidade calculada para o complexo $sAPP\alpha_{695} + dummy_prot$, utilizando o programa MASSHA (Konarev et al., 2002, veja item 1.6.7 do Cap.I desta tese) para rotação de corpo rígido. Neste procedimento o programa obtém a melhor orientação entre a $sAPP\alpha_{695}$ e o *dummy_prot* que fornecem o melhor ajuste da curva. Inicialmente o *dummy_prot* foi colocado em uma extremidade e posteriormente na outra extremidade. Isso é necessário porque o programa executa variações ao redor do ponto inicialmente proposto. Os ajustes são mostrados abaixo. Esta parcela adicionada é indicada com pontos vermelhos.

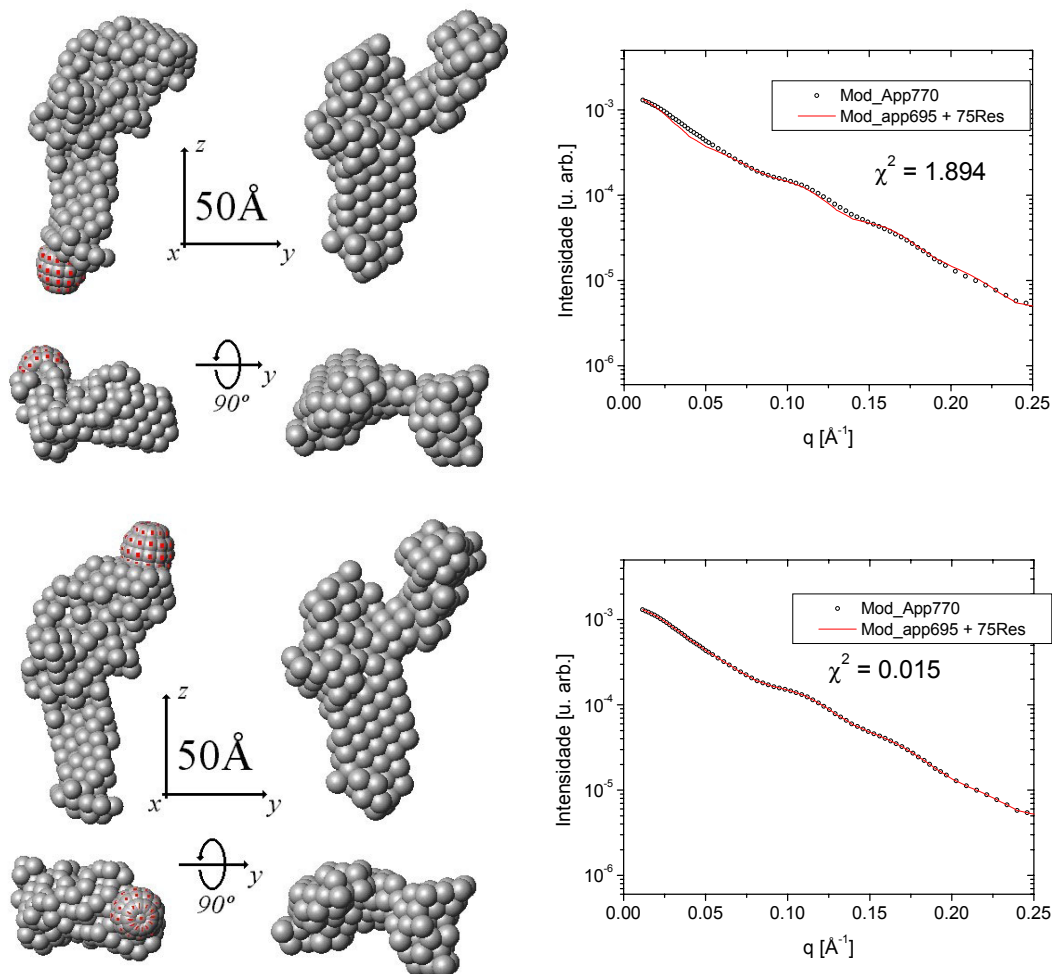


Figura 81 – Rotação de corpo rígido para dois possíveis posicionamentos do *dummy_prot* com o respectivo ajuste dos dados. Acima: *dummy_prot* posicionado na parte inferior da $sAPP\alpha_{695}$. Abaixo: *dummy_prot* posicionado na parte inferior da $sAPP\alpha_{695}$.

No primeiro caso a presença do domínio na parte inferior do modelo $sAPP\alpha_{695}$, forneceu um ajuste ruim da intensidade do modelo $sAPP\alpha_{770}$. Por outro lado a adição deste domínio na parte superior do modelo $sAPP\alpha_{695}$ fornece um ajuste perfeito da intensidade. Isto é evidente pelo valor de χ^2 em cada caso. Assim estes ajustes confirmam a hipótese de que a região referente à diferença entre as seqüências primárias deva estar posicionada naquela região do modelo da $sAPP\alpha_{695}$.

- Posicionamento dos domínios em resolução atômica nos modelos *ab initio*

Apesar do substancial progresso em biologia estrutural de domínios isolados da APP (Barnham et al., 2003; Dulubova et al., 2004; Hynes et al., 1990; Kroenke et al., 1997; Rossjohn et al., 1999; Wang et al., 2004), a estratégia de fragmentação por si mesma reflete a dificuldade de se obter bons cristais de sAPP completa. Entre os diversos motivos de insucesso no processo de cristalização, destaca-se a flexibilidade intrínseca da proteína APP, que discutimos anteriormente. Além disso, estes obstáculos são complementados pela ausência de experimentos de ressonância nuclear magnética (NMR) capazes de elucidar estruturas tridimensionais de proteínas do tamanho da sAPP.

Para estabelecer correlações topológicas entre as estruturas de resolução atômica e os modelos *ab initio* obtidos pelos dados de SAXS utilizamos informações disponíveis na literatura. Os domínios descritos em resolução atômica para a proteína APP são mostrados abaixo, juntamente com o mapa de carga superficial de cada domínio:

- 1MWP (28-123)

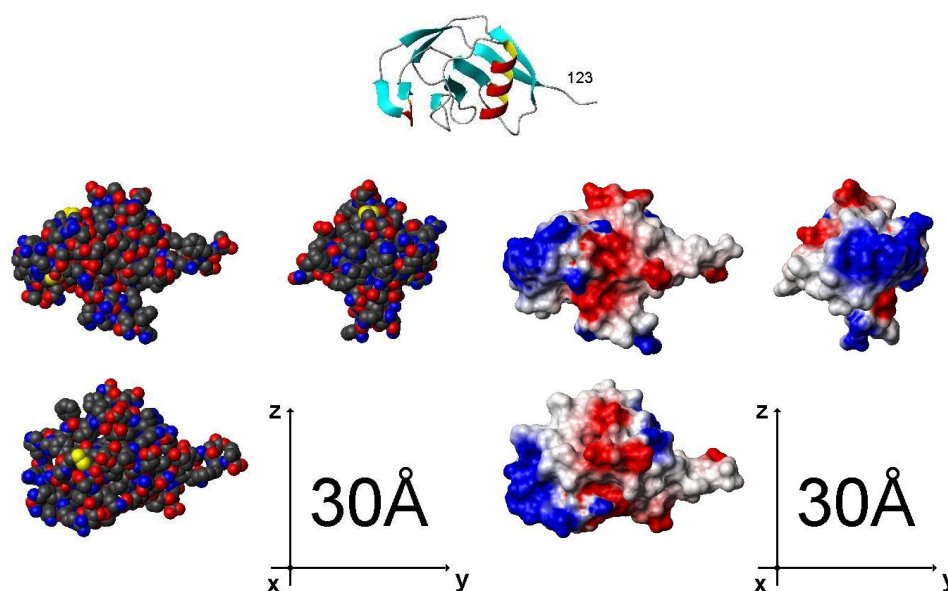


Figura 82 – Estrutura com resolução atômica para a parcela 28-123 da sAPP α Acima: Estrutura na forma de “ribbons”. Abaixo: esquerda) estrutura indicando os átomos, direita) estrutura indicando o potencial eletrostático na superfície da proteína. Entrada no Protein Data Bank: 1MWP.PDB (Rossjohn et al., 1999).

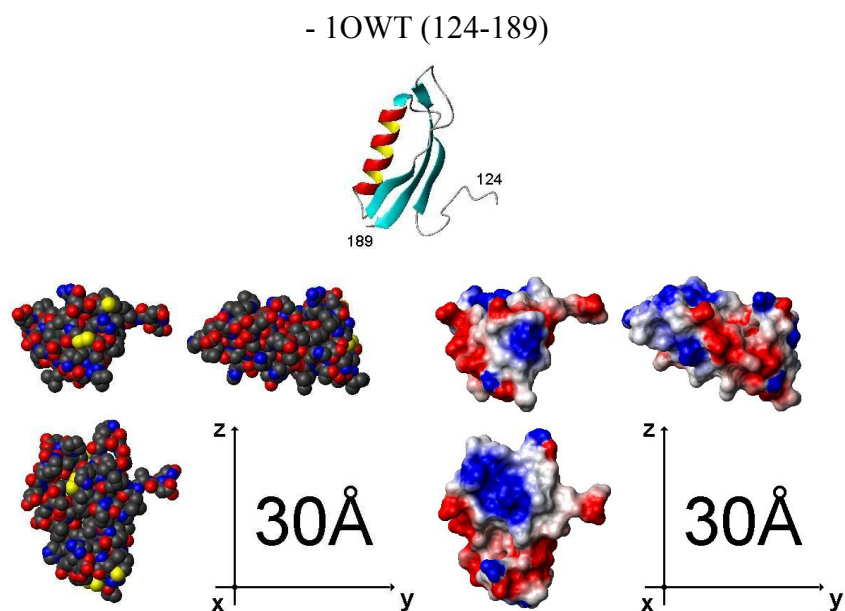


Figura 83 – Estrutura com resolução atômica para a parcela 124-189 da sAPP α . Acima: Estrutura na forma de “ribbons”. Abaixo: esquerda) estrutura indicando os átomos, direita) estrutura indicando o potencial eletrostático na superfície da proteína. Entrada no Protein Data Bank: 1OWT.PDB (Barnham et al., 2003).

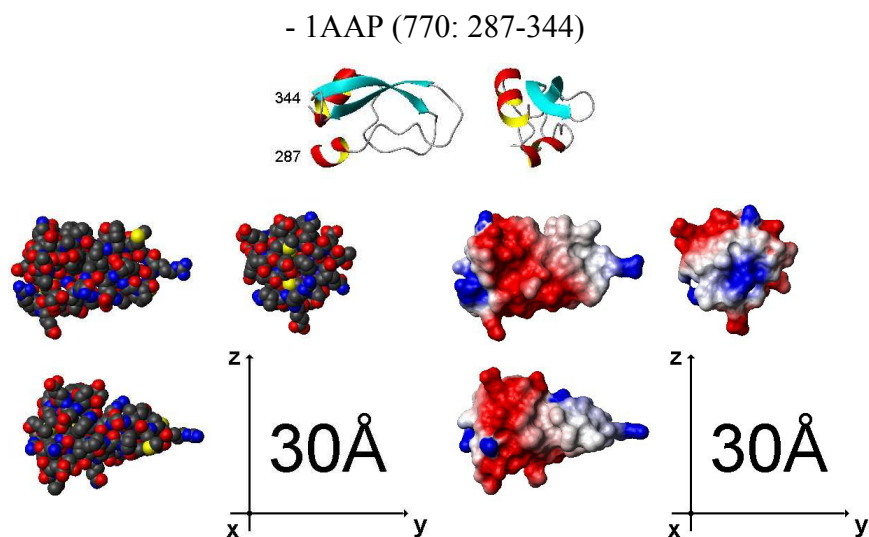


Figura 84 – Estrutura com resolução atômica para a parcela 287-344 da sAPP α_{770} . Acima: Estrutura na forma de “ribbons”. Abaixo: esquerda) estrutura indicando os átomos, direita) estrutura indicando o potencial eletrostático na superfície da proteína. Entrada no Protein Data Bank: 1AAP.PDB (Hynes et al., 1990).

- 1RW6 (695: 299-490, 770: 374-565)

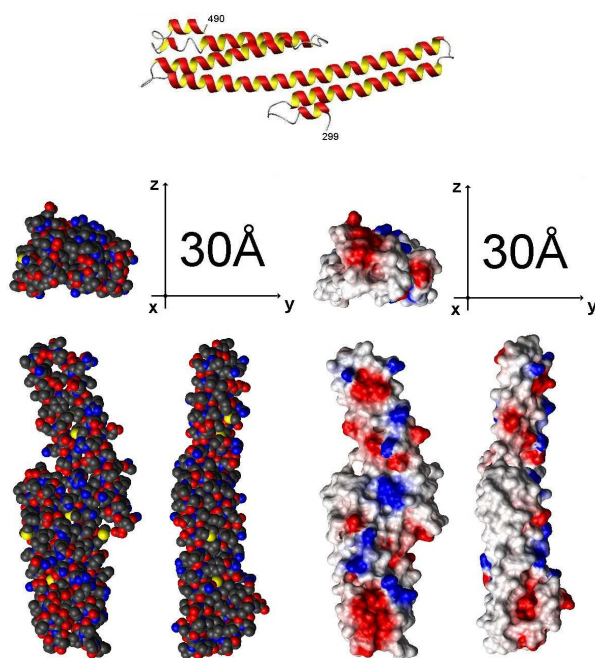


Figura 85 – Estrutura com resolução atômica para a parcela 299-490 da sAPP α_{770} e 374-565 da sAPP α_{770} . Acima: Estrutura na forma de “ribbons”. Abaixo: esquerda) estrutura indicando os átomos, direita) estrutura indicando o potencial eletrostático na superfície da proteína. Entrada no Protein Data Bank: 1RW6.PDB. (Wang et al., 2004).

O maior problema na localização correta dos domínios em resolução atômica nos modelos de SAXS decorre da existência de múltiplos *gaps* sem informação estrutural na estrutura da proteína. Uma tentativa inicial de descrever destes *gaps* correspondeu a testes de modelagem por homologia, utilizando rotinas que utilizam como *template* para a modelagem, estruturas conhecidas de proteínas homólogas (Fiser e Šali, 2003). Este procedimento não foi possível devido à baixa homologia entre a estrutura primária da APP e as estruturas protéicas conhecidas. Diversas rotinas de busca foram utilizadas, como as ferramentas BLAST (Altschul et al., 1990), GenTHREADER (McGuffin e Jones, 2003) e LOOP (Teodorescu et al., 2004). Mesmo para pedaços pequenos da sequência primária não obtivemos resultados que permitissem a utilização de uma determinada estrutura como modelo primário para a modelagem de partes da APP. Na ausência de modelos para estes *gaps*, o uso de ferramentas otimizadas de busca para posicionamento dos domínios nos modelos de SAXS, ferramentas estas guiadas pelo ajuste dos dados experimentais de SAXS, não é possível. Sendo assim, um posicionamento manual foi proposto, utilizando o

programa MOLMOL (Koradi et al., 1996) baseado em comparações dimensionais e conformacionais das estruturas atômicas conhecidas e dos modelos de SAXS.

Como primeiro passo buscamos a determinação mais provável das regiões contíguas 1MWP e 1OWT. Varias metodologias foram tentadas para obter o *docking* destas estruturas. Inicialmente tentamos obter este arranjo entre as subunidades por ferramentas de construção de proteínas por homologia (Fiser e Šali, 2003 - programa Modeller, Šali e Blundell, 1993). Estas ferramentas fornecem resultados muito confiáveis para o arranjo protéico quando temos homólogos sequenciais com homologia acima de 40%. O programa toma a estrutura cristalográfica do homólogo como *template* e constrói a nova proteína utilizando a correta seqüência primária. Infelizmente, fazendo uma busca no banco de dados não foi possível obter-se *templates* com alta homologia e os resultados obtidos acabaram por não ser satisfatórios.

A vantagem deste método de construção por homologia é que teríamos as proteínas naturalmente contíguas, isto é, unidas pelo ponto de ligação da seqüência primária. Como não foi possível aplicar este método, utilizou-se ferramentas de docking automático de proteínas (Programa GRAMM – Vakser, 1992). Utilizando o programa no modo de docking em baixa resolução (Vakser e Nikiforovich 1995, Vakser 1995, Vakser 1996) geramos um total de 1000 modelos. Agrupando-os em grupos de 50 modelos realizou-se uma comparação tridimensional entre os componentes deste grupo utilizando o programa SUPCOMB (Kozin e Svergun 2001). O modelo mais representativo do grupo foi selecionado para a promediação final. No cálculo do modelo final os 20 modelos selecionados foram promediados utilizando o programa DAMAVER (Volkov e Svergun, 2002). Mostra-se que este procedimento de promediação fornece resultados mais confiáveis para o *docking* (Li et al, 2003, Zeng e Doniach, 2002). Além de sabermos o ponto de ligação das proteínas, a proteína 1MWP possui um sítio de ligação de heparina e também um provável sítio para ligação da proteína 1OWT (Rossjohn et al 1999, Barnham et al 2003). Reunindo estas informações no modelo final obtido pela promediação, obtivemos o docking teórico mais provável para o complexo, mostrado na Figura 86:

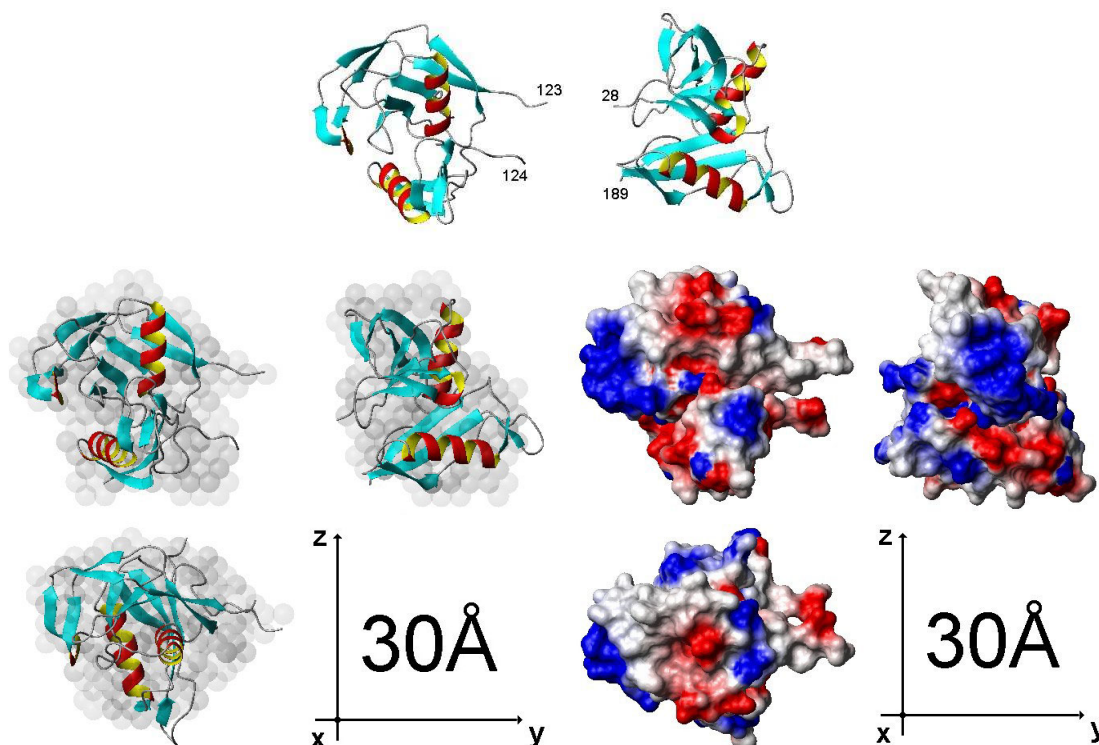


Figura 86 – *Modelo molecular construído pelo docking do domínio N-terminal ligador de heparina (1MWP.PDB) e o domínio ligador de cobre (1OWT.PDB). Para o docking utilizou-se o programa GRAMM e o processo de filtragem descrito no texto. Acima: Estrutura na forma de “ribbons”. Abaixo: esquerda) sobreposição da estrutura atômica com o modelo filtrado, direita) estrutura indicando o potencial eletrostático na superfície da proteína.*

É natural que os resíduos 123 e 124 fiquem "soltos" no modelo uma vez que, na estrutura completa estas regiões devam se curvar de modo a fazer a ligação entre as proteínas. De toda maneira o posicionamento próximo destes resíduos é uma indicativa de que o *docking* deve estar bastante próximo do real arranjo dos domínios.

Para a superposição dos domínios com resolução atômica nos modelos ab initio de SAXS assumimos duas hipóteses fundamentais:

1) *A forma das estruturas cristalográficas não muda muito na proteína completa.*

Esta hipótese é válida, pois diversos artigos mostram que parcelas cristalizadas de proteínas mantêm quase a mesma forma no complexo inteiro. As únicas diferenças que podem existir consistem em mudanças de arranjo de algumas partes móveis em decorrência da presença dos outros domínios protéicos;

2) *As isoformas da APP possuem a mesma conformação nas regiões onde há alinhamento seqüencial.* Similarmente ao mencionado acima esta hipótese é bastante boa uma vez que a estrutura terciária de uma proteína é ditada principalmente pela sua seqüência primária. No entanto, podem ocorrer pequenas mudanças estruturais nos domínios comuns às duas proteínas em decorrência da presença do domínio KPI (1AAP.PDB) na isoforma sAPP α_{770} .

Com base nas dimensões do domínio C-terminal (1RW6.PDB) mostrado anteriormente temos a indicação de que ele deve estar contido na região inferior esquerda, que se estende nas duas isoformas (Figura 78). Com esta informação e o domínio resultante do *docking* entre as parcelas 1MWP e 1OWT, pudemos superpor estas estruturas no modelo de SAXS da sAPP α_{695} . Também, o domínio KPI foi ajustado na protuberância do modelo da sAPP α_{770} com as posições dos outros domínios mantidos como na sAPP α_{695} . Os resultados são mostrados nas Figura 87 e Figura 88.

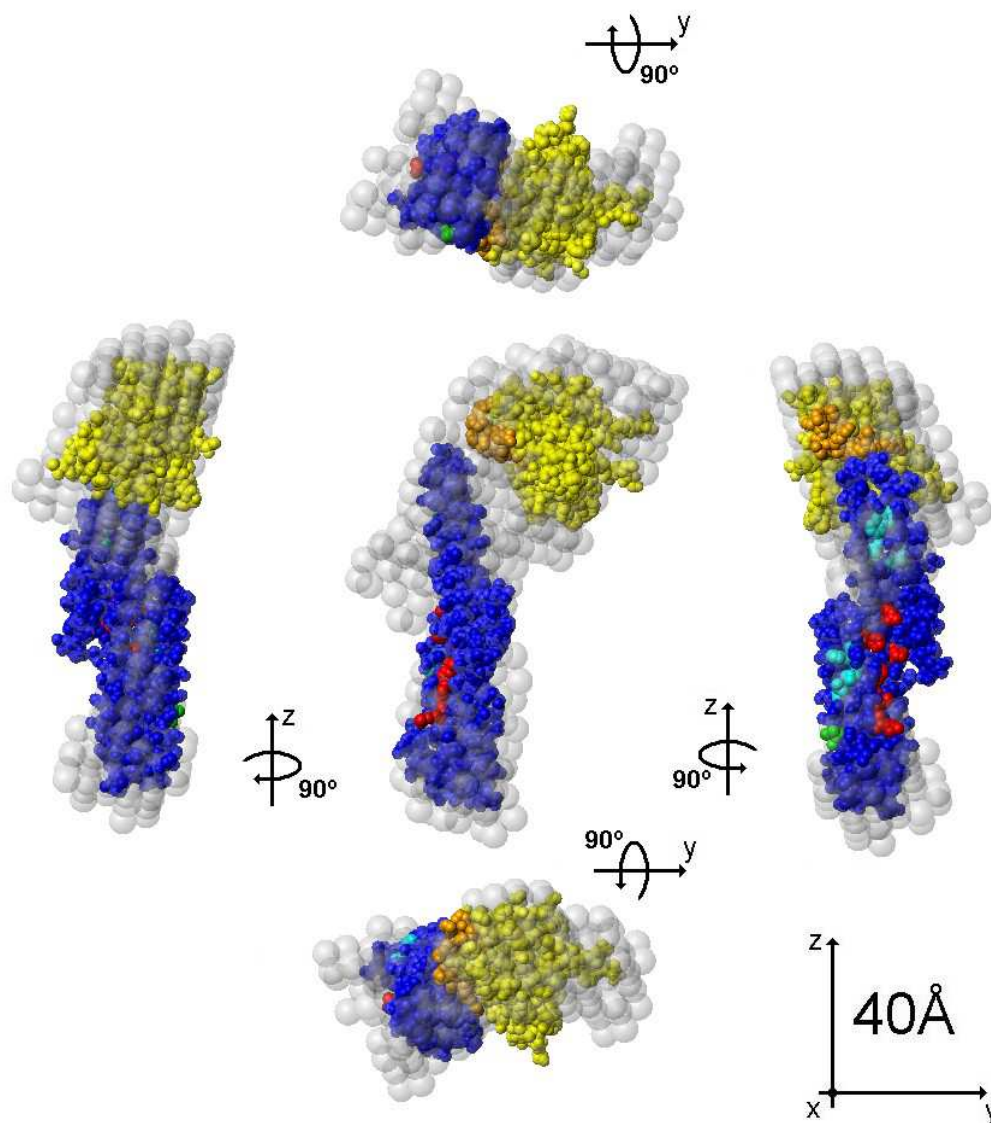


Figura 87 – Sobreposição dos domínios com resolução atômica sobre o modelo *ab initio* obtido por experimentos de SAXS para a proteína sAPP α_{695} . Domínio azul) Domínio C-terminal onde indicamos a região ligadora de heparina (vermelho), interface de dimerização do domínio (azul ciano) e sítio de glicosilação (verde). Domínio amarelo) Docking do domínio N-terminal 1MWP com o domínio ligador de cobre 1OWT. Em cor laranja indicamos a região ligadora de heparina. Esferas semitransparentes) modelo *ab initio* médio obtido pelos dados de SAXS para a isoforma sAPP α_{695} .

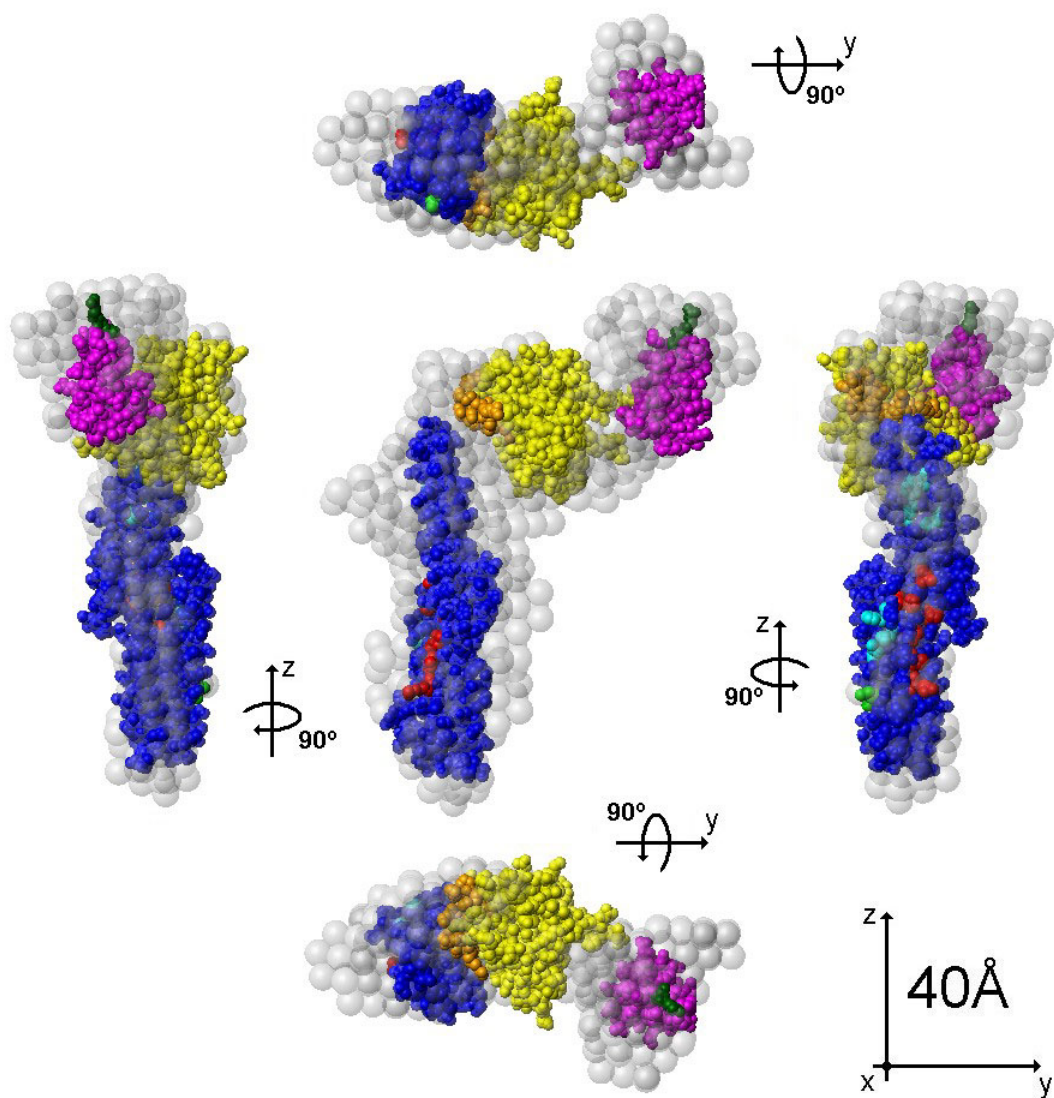


Figura 88 – Sobreposição dos domínios com resolução atômica sobre o modelo *ab initio* obtido por experimentos de SAXS para a proteína *sAPPα₇₇₀*. Domínio azul) Domínio C-terminal onde indicamos a região ligadora de heparina (vermelho), interface de dimerização do domínio (azul ciano) e sítio de glicosilação (verde). Domínio amarelo) Docking do domínio N-terminal 1MWP com o domínio ligador de cobre 1OWT. Em cor laranja indicamos a região ligadora de heparina. Domínio magenta) Domínio que contém o inibidor de protease 1AAP. Domínio verde) Sítio de ligação do domínio inibidor de protease. Esferas semitransparentes) modelo *ab initio* médio obtido pelos dados de SAXS para a isoforma *sAPPα₇₇₀*.

Com a superposição dos modelos podemos calcular a intensidade de espalhamento para o modelo construído pelos domínios atômicos e compará-la com os dados experimentais de SAXS. Utilizando o programa CRY SOL, realizamos este cálculo, mostrando os resultados na Figura 89 a seguir:

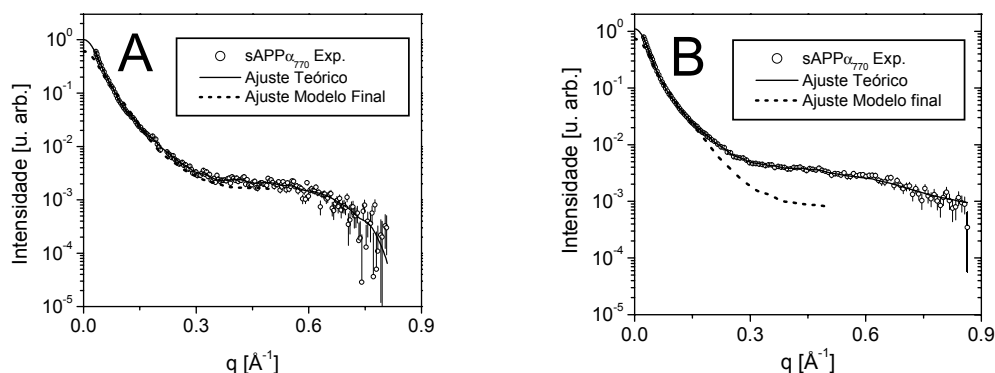


Figura 89 – Ajuste dos dados experimentais utilizando o modelo construído por domínios atômicos. A) ajuste para sAPP α_{695} . B) ajuste para sAPP α_{770} .

Como mencionado anteriormente, a existência de grandes *gaps* estruturais somente permitirá um ajuste parcial dos dados. Pequenas mudanças de posicionamento e orientação dos domínios cristalográficos utilizando o programa MASSHA não melhoraram a qualidade do ajuste. Apesar destas limitações, a parte inicial das curvas de SAXS (que dita a forma geral da partícula) é corretamente aproximada, fornecendo a indicação de que os domínios estão corretamente posicionados.

- Implicações biológicas dos resultados

- Função da sAPP α

A combinação das informações estruturais em vários níveis de resolução nos permitiram aqui obter um primeiro vislumbre da estrutura da sAPP α , nos trazendo ao limiar de compreensão as mudanças estruturais das quais dependem as funções da APP. A sAPP α é importante no fator de crescimento do cérebro e neurogênese em adultos (Sabo et al., 2003), paralelamente à sua importância em outros tecidos humanos (Quast et al., 2003). Apesar de os mecanismos moleculares de sua ação nos neurônios ainda não serem claros, possui efeitos de aumento de memória. Atividade neural pode levar a uma maior secreção

de sAPP α , sendo que esta APP mostra-se neuroprotetiva. Experimentos *in vivo* levaram a hipóteses de que AD é causada em parte pela ausência de sAPP α , em adição aos efeitos tóxicos da A β (Stein and Johnson, 2002). Ao passo que se mostrou que sAPP atua como fator de diferenciação celular e que liga-se para definir caminhos na superfície celular, nenhum receptor para a APP foi até agora descoberto ou sugerido. Nossos modelos para a sAPP α deste modo servem como ferramentas para investigar-se os possíveis ligantes e mecanismos de ligação deste importante fragmento da APP.

- Contexto Celular da APP

Nossos modelos fornecem informações da distância dos diferentes domínios da APP da membrana celular. Apesar da estrutura desordenada no domínio próximo à membrana, é provável que a região C-terminal da sAPP α , e desta forma a posição onde a APP é clivada pela α -secretase, ocorra na proximidade da região C-terminal do fragmento 1RW6, ou seja, na região inferior de nosso modelo (Figura 87 e Figura 88). Como para nenhuma isoforma de sAPP α sugere-se a ligação de lipídeos, é bastante provável que o eixo longitudinal da sAPP α_{695} e o eixo correspondente da sAPP α_{770} estendam-se verticalmente a partir da membrana. Neste caso, os sítios ligadores de heparina de laminine localizados no domínio helicoidal (Figura 87 e Figura 88 azul ciano), devem estar à uma distância de $\sim 50\text{\AA}$ da membrana, restringindo as possíveis situações de ligação fisiológica. Neste sentido, o domínio N-terminal e o domínio ligador de cobre (1MWP+1OWT) devem estar a uma distância de $\sim 100\text{\AA}$ da membrana, enquanto que o domínio KPI, o qual necessita interagir com o sítio ativo das proteases alvo da APP, está à uma distância de $\sim 120\text{\AA}$ da membrana.

Mesmo que nosso modelo não descreva a vizinhança dos sítios de clivagem da β -secretase, 6 e 16 aminoácidos a partir da região C-terminal da sAPP α , em alta resolução, a estrutura geral da sAPP α pode ser utilizada na modelagem da interação da APP com a β -secretase, a qual é uma proteína transmembrana de passagem única também.

- Sinalização pela APP

A proteína APP completa foi sugerida inicialmente como sendo um receptor transmembrana, transmitindo sinais extracelulares para o interior da célula. Apesar das funções celulares promovidas pela proteína transmembrana APP ficarem mais claras com o passar dos anos (Sabo et al., 2003), não se obteve conclusões sobre a forma de sinalização. Qualquer uma das diversas moléculas na matriz extracelular que ligam APP (heparina, laminina, fibulina e espondina) pode servir como sinal, mas as circunstâncias nas quais a ligação desta molécula à APP transmembrana leva à uma sinalização intracelular definida ainda deve ser estabelecida.

Um ponto central na transmissão de sinais pela APP é a questão de se a APP ou qualquer um de seus fragmentos é capaz de oligomerizar-se. Alguns estudos de superexpressão de APP em células mamárias mostraram uma pequena proporção de dímeros (Wang et al., 2004). Diversos experimentos com fragmentos da sAPP deram evidência de que podem existir na forma de dímeros. Scheuermann et al. (Scheuermann et al., 2001) obteve evidências cromatográficas de que o fragmento APP18-324 dimeriza em solução, enquanto que Wang e Ha (2004) (Wang and Ha, 2004) propuseram uma evidência similar de que o fragmento APP346-551 dimeriza em solução. Por outro lado, os únicos experimentos publicados para a sAPP α completa mostraram por SAXS, filtração de gel e análise de sedimentação de que ambas as isoformas sAPP α_{695} e sAPP α_{770} são monômeros mesmo em concentrações relativamente altas (Gralle, Botelho e Oliveira et al., 2002; Botelho, Gralle e Oliveira et al., 2003; Gralle e Oliveira et al., 2004). Nossos modelos oferecem neste sentido a solução a este paradoxo, uma vez que sugerem obstáculos históricos entre os diferentes domínios da sAPP α nativa, impedindo a formação de um dímero. A estabilidade do estado nativo da sAPP α é baixo o suficiente para permitir um possível rearranjo de seus domínios em certas condições fisiológicas e na presença de certos ligantes fisiológicos, conforme mostrados nos itens acima. Este rearranjo de domínios pode expor parcelas hidrofóbicas da superfície, facilitando a dimerização da sAPP α e possivelmente da APP transmembrana.

- Considerações finais

Destes resultados vemos que experimentos de SAXS combinados com modelagem ab initio e dinâmica molecular permitiram, pela primeira vez, propor um modelo completo para a estrutura e função da APP em suas duas maiores isoformas, sAPP α_{695} e sAPP α_{770} . Confirmando resultados obtidos por outros autores (Marquez et al., 2003), nosso trabalho mostra que estudos por SAXS com medidas em alta resolução são capazes de fornecer informações estruturais e funcionais em casos onde estruturas cristalográficas completas não estão disponíveis.

c) Considerações Finais do Estudo com a Proteína Precursora Amilóide

Mostramos um longo estudo da proteína APP utilizando a técnica de SAXS para obter informações estruturais e conformacionais. Diversos experimentos foram apresentados e discutidos, podendo-se ao final obter um modelo estrutural em alta resolução para a estrutura da sAPP α em suas duas maiores isoformas, além de resultados de sua estabilidade estrutural. Além disso, mostramos que o fragmento sAPP α ocorre na forma monomérica em solução, elucidando esta questão uma vez que seus fragmentos facilmente dimerizam. A maior compreensão dos mecanismos de ação da proteína APP aliados a informações estruturais permite um entendimento melhor de suas funções e, talvez, de inibição de suas propriedades patogênicas.

CAPÍTULO IV

SIMULAÇÕES PARA

SISTEMAS

DENSOS

1. Cálculo de Intensidade de Espalhamento, Fator de Estrutura e Função de Correlação para Sistemas Densos

1.1. Introdução

Apresento aqui resultados correspondentes a um estágio que realizei no mês de Agosto/2003 na universidade de Aarhus / Dinamarca, sobre a supervisão do Prof. Dr. Jan Skov Pedersen. Como mostrarei, o estágio foi muito frutífero e cumpri integralmente o projeto proposto na proposta de inicial.

Em resumo a proposta de trabalho consistia de simulações de Monte Carlo para partículas não esféricas. Nestas simulações utilizamos o método Monte Carlo no desenvolvimento de programas que permitissem o cálculo de intensidades de espalhamento (e outras funções) para sistemas com frações de volume variadas, onde partículas espalhadoras eram elipsóides de revolução com as mais variadas anisotropias. A motivação principal para este tipo de estudo consiste no fato de que, em baixa resolução, muitas partículas podem ser descritas por um elipsóide de revolução (oblato ou prolato). Em experimentos típicos de SAXS sempre tentamos utilizar concentrações baixas para tentar eliminar efeitos interpartículas (fator de estrutura). Porém, em sistemas vivos e interior de células, muitas vezes encontramos altas concentrações de proteínas, e assim um estudo apurado deste tipo de sistema requereria experimentos em alta concentração. Como expressões analíticas somente existem para o caso de partículas esféricas, simulações Monte Carlo para elipsóides de revolução podem ser uma alternativa para esta deficiência, possibilitando o estudo por SAXS de interações interpartículas de proteínas. Além disso, existem muitos outros sistemas onde as partículas geradas são elipsóides (micelas, vesículas, algumas ligas metálicas) em variadas concentrações, e o conhecimento da intensidade de espalhamento para estas condições enriquecerá em muito a análise destes dados.

1.2. Método Monte Carlo

1.2.1. Introdução

O método Monte Carlo (MC) é uma técnica de simulação muito útil no estudo de sistemas com contribuições randômicas decorrentes do elevado número de graus de liberdade. Por exemplo, para um sistema consistindo apenas de uma cadeia polimérica, é muito difícil obter-se expressões analíticas rigorosas que consigam descrever a estatística geral bem como as funções de espalhamento. Simulações Monte Carlo podem ser utilizadas na geração de séries de configurações possíveis e também na obtenção das séries de valores configuracionais médios de vários parâmetros, como raio de giro, máxima distância etc, que são amostrados durante a simulação. Neste sentido, indicações sobre o comportamento estatístico são conseguidas.

A maioria dos algoritmos MC possui um caráter dinâmico, que significa que devemos fazer uma operação (ou movimento) na configuração corrente para atingir a próxima (nova) configuração. Os tipos de movimentos possíveis dependem do sistema que está sendo estudado. Por exemplo, para um sistema composto de esferas, este movimento pode ser somente a translação da mesma. Para outros tipos de partículas poderia corresponder a uma translação e uma rotação, além de poder ser uma alteração na própria partícula (alteração de seu tamanho etc).

O método Monte Carlo mais simples é baseado nos seguintes princípios:

- Todas as configurações devem ser geradas com igual probabilidade;
- A probabilidade de ser ir de um estado A para um estado B deve ser idêntica à probabilidade de ir de um estado B para o estado A (princípio de balanço detalhado);
- Para dois estados, A e B, deve ser sempre possível ir de um estado A para um estado B (princípio ergódico);
- Se alguns estados são proibidos, por exemplo devido a algum tipo de superposição, o estado antigo é mantido e contado mais uma vez quando médias são calculadas;

Para sistema com interação de esfera rígida a decisão se o novo estado deve ser aceito é simplificada: as superposições entre todos os pares de esferas são checados e o

novo estado é aceito se não são detectadas superposições. Já para sistemas com potenciais de interação finitos, como potenciais de Lennard-Jones ou Coulomb, por exemplo, o assim chamado critério de Metropolis é usado (Metropolis et al, 1953). Por este critério a nova configuração é sempre aceita se está em um nível de energia menor que o antigo. Se possuir energia maior, é aceita com a probabilidade $\exp(-\Delta E/kT)$, sendo ΔE a diferença de energia entre o novo estado e o antigo estado e kT é a constante de Boltzmann multiplicada pela temperatura absoluta. Este enfoque fornece configurações que são pesadas pelo fator de Boltzmann $\exp(-E_A/kT)$, sendo E_A a energia do estado A, de tal modo que as médias termodinâmicas podem ser calculadas como médias simples sobre as configurações.

Como as simulações envolvem repetições muito longas de uma mesma operação e cálculos (10^6 vezes ou mais), é importante utilizar-se algoritmos eficientes. Usualmente a parte mais dispendiosa da simulação é o cálculo das energias (ou equivalentemente, a checagem das superposições) e isso, em princípio, requer N^2 operações para um sistema de N partículas. Em alguns tipos de sistemas alguns truques computacionais podem ser utilizados (Pedersen, 2002). Para simulações com muitas partículas o procedimento padrão é utilizar-se uma caixa com condições periódicas de contorno. Nestes casos a convenção de mínima imagem (*'minimum image convention'*) fornece uma maneira muito conveniente de seleccionar-se quais interações entre partículas devem ser consideradas (Allen and Tildesley, 1987). Como regra prática, uma simulação MC é mais eficiente no espaço de configurações se a fração de movimentos aceitos não é menor que 50%. Se a fração é significativamente menor, isto significa que uma configuração aceita é amostrada muitas vezes, o que é claramente ineficiente. Se a fração é significativamente maior, isto pode significar que a nova configuração difere apenas um pouco da anterior, o que também é ineficiente. Além disso, para todos os tipos de simulação é importante que um bom gerador de números aleatórios seja usado. Isto significa, em geral, que os geradores de números aleatórios internos dos computadores usuais não podem ser utilizados. A qualidade do gerador de números aleatórios é particularmente importante se somente uma pequena fração (tipicamente $< 1\%$) das tentativas de movimento são aceitas: A simulação acaba se tornando um teste do gerador ao invés de uma simulação real. No presente trabalho utilizou-se algoritmos de geração de números aleatórios descritos no livro *Numerical Recipes* (Press et al., 1992), que possuem qualidade aceitável para a maioria dos casos.

Na utilização da técnica Monte Carlo em um problema em particular é importante que o modelo seja escolhido com muito cuidado. O modelo deve incluir tantos detalhes quantos sejam necessários para descrever-se as medidas com a resolução que elas possuem. Ao mesmo tempo, ele deve ser o mais simples possível, de tal modo a permitir que as simulações possam ser realizadas com suficiente precisão estatística em um tempo razoável. Esta precisão deve ser melhor que a de dados experimentais que eventualmente venham a ser descritos por estes cálculos.

1.2.2. Cálculo de Funções de Espalhamento

O caso mais geral de espalhamento por sistemas monodispersos de partículas em solução pode ser descrito como sendo o caso em que os espalhadores possuem uma correlação espacial. Como mostrado na equação (53) esta correlação entre as partículas pode ser representada na intensidade de espalhamento através do fator de estrutura $S(\vec{q})$,

$$I(q) = \langle N \rangle \langle f^2(\vec{q}) \rangle \langle S(\vec{q}) \rangle$$

Para um sistema composto por partículas esféricamente simétricas a média indicada na equação anterior pode ser realizada individualmente para cada termo do produto e assim a intensidade de espalhamento pode ser escrita como

$$I(q) = N \Delta \rho^2 V^2 P(q) S(q) \quad (156)$$

sendo N o número de partículas, $\Delta \rho$ o contraste de densidade de espalhamento entre as partículas e o solvente/matriz, $P(q)$ é o fator de forma normalizado das partículas e $S(q)$ o fator de estrutura indicado na equação (53).

A expressão anterior assume simetria esférica para a forma da partícula. Para partículas idênticas com algum tipo de anisotropia a equação da intensidade assume a forma (Pedersen, 2002),

$$I(q) = \Delta \rho^2 V^2 \left[\sum_i F_i(q, \vec{e}_i)^2 + \frac{1}{N} \sum_{i,j} F_i(q, \vec{e}_i) F_j(q, \vec{e}_j) [S_{i,j}(q, \vec{e}_i, \vec{e}_j) - 1] \right] \quad (157)$$

onde as somas são feitas sobre todas as partículas na amostra e $F_i(q, \vec{e}_i)$ é a amplitude do fator de forma da i -ésima partícula com orientação dada pelo vetor unitário $\vec{e}_i$. As funções $S_{i,j}(q, \vec{e}_i, \vec{e}_j)$ são os fatores de estrutura parciais que dependem das respectivas orientações e N

é o número de partículas. Note que o primeiro termo na equação (157) é proporcional ao fator de forma médio: $P(q) \equiv \langle F^2(q) \rangle_0$.

Para partículas com uma pequena anisotropia podemos assumir que as interações são independentes das orientações e dado pelo tamanho médio das partículas. Isto leva a aproximação de desacoplamento (*decoupling approximation* – Kotlarchyk and Chen, 1984):

$$I(q) = \Delta \rho^2 V^2 P(q) [1 + \beta(q)(S(q) - 1)] \quad (158)$$

onde,

$$\beta(q) = \langle F(q) \rangle_0^2 / \langle F^2(q) \rangle_0 \quad (159)$$

e $S(q)$ é o fator de estrutura calculado para o tamanho médio da partícula definido como $R_{av} = [3V/(4\pi)]^{1/3}$, sendo V o volume da partícula.

Nos casos de partículas não esféricas simulações Monte Carlo podem ser utilizadas para obter-se os parâmetros desejados. Em uma caixa de tamanho determinado, colocam-se N partículas (com qualquer forma) randomicamente orientadas e, em cada passo da simulação translada-se e rota-se uma delas aleatoriamente, assumindo-se condições de contorno periódicas. Com isso pode-se calcular médias dentro deste grupo de partículas e obter-se diretamente a intensidade $I(q)$ e o fator de estrutura $S(q)$.

Para o cálculo destas funções em sistemas compostos de objetos independentes com orientação randômica, a intensidade de espalhamento e o fator de estrutura podem ser calculados utilizando as equações de Debye (Debye, 1915) na forma:

$$I(\vec{q}) = \left[\sum_{k=1}^N F_k(\vec{q}) \sin(\vec{q} \cdot \vec{R}_k) \right]^2 + \left[\sum_{k=1}^N F_k(\vec{q}) \cos(\vec{q} \cdot \vec{R}_k) \right]^2 \quad (160)$$

$$S(q) = \frac{1}{N^2} \sum_{i,j} \frac{\sin q r_{ij}}{q r_{ij}} \quad (161)$$

sendo N o número de centros espalhadores e r_{ij} a distância entre os pontos com índices “i” e “j”, F_k é o fator de forma da partícula espalhadora e R_k a posição da partícula “k” com relação à uma dada origem. Note que a normalização $S(q=0) = 1$ foi escolhida.

Para partículas interagentes, uma super cela com fronteiras periódicas é usada para reduzir os efeitos de borda e o tamanho finito do sistema. Para este tipo de cela a função de espalhamento deve ser calculada nos pontos da rede recíproca da cela unitária. Nesse

sentido, não podemos tentar obter informações do sistema para distâncias que são maiores que o comprimento da cela unitária. A 'densidade' dos pontos da rede recíproca se torna grande para grandes valores de q e assim é usual calcular-se a função somente nas direções representativas, como nas direções 001, 110 e 111 e suas equivalentes (Mulder e Frenkel, 1985; Frenkel e Mulder, 1985; Pedersen, 2002). Isto fornece um total de 13 direções. Para cada uma das direções (h,k,l) a funções de espalhamento são então calculadas como sendo:

$$I(\vec{q}) = \left\{ \left[\sum_{i=1}^N F_k(\vec{q}_p) \sin(\vec{q}_p \cdot \vec{R}_k) \right]^2 + \left[\sum_{i=1}^N F_k(\vec{q}_p) \cos(\vec{q}_p \cdot \vec{R}_k) \right]^2 \right\}_{\vec{q}_p \in [100], [110], [111]} \quad (162)$$

$$S(\vec{q}) = \left\{ \left| \sum_{i=1}^N \exp(-i\vec{q}_p \cdot \vec{R}_k) \right|^2 \right\}_{\vec{q}_p \in [100], [110], [111]} \quad (163)$$

onde L é o comprimento do lado da cela unitária, (x_j, y_j, z_j) é a posição do j-ésimo ponto e $q_p = p(2\pi/L)(h^2 + k^2 + l^2)^{1/2}$ é o módulo do vetor de espalhamento, com $p = 1, 2, \dots$. As funções de espalhamento das diferentes direções são promediadas e finalmente as funções das diferentes direções são sobrepostas. Veja que as equações (162) e (163) requerem somente $13N$ cálculos em oposição ao fator N^2 das equações (160) e (161). O uso da equação (163) pode ser combinado com rotinas de Transformadas de Fourier Rápidas (FFT) o que também acelera os cálculos (Pedersen, 2002). Devido a diversos erros numéricos envolvidos na discretização de operações FFT, é recomendável utilizar fórmulas de adição para seno e cosseno na equação (163) (Pedersen, 2002) o que também reduz significativamente o tempo de computação. Similarmente, o cálculo da intensidade de espalhamento somente será feito também nestas direções.

Algoritmos para simulações Monte Carlo dinâmico fornecem configurações que são correlacionadas, no sentido de que a nova configuração possui diversas partes e propriedades que são idênticas à anterior. Isto significa que algumas quantidades, como funções de espalhamento, por exemplo, não precisam ser calculadas em cada etapa. Isto é conhecido como 'amostragem estroboscópica' e em geral reduz o tempo de simulação significativamente. Quando se calcula os erros em uma quantidade amostrada, a possível presença de correlação deve ser levada em conta. A maneira mais simples consiste em "quebrar" a simulação em diversos blocos (10-20 tipicamente) cada um possuindo tamanhos significativamente maiores que o tempo de relaxação do sistema. Nestes casos os

blocos são individualmente independentes e o desvio padrão na média $\langle A \rangle$ da quantidade A é dado por

$$\sigma(\langle A \rangle)^2 = \frac{1}{M(M+1)} \sum_{m=1}^M [\langle A \rangle_m - \langle A \rangle]^2 \quad (164)$$

onde M é o número de blocos e $\langle A \rangle_m$ é a média da quantidade A no m -ésimo bloco. Esta equação também pode ser utilizada para as funções de espalhamento para cada um dos valores de q .

1.3. Método Monte Carlo para Simulação de Elipsóides em Solução

1.3.1. Introdução

O comportamento de líquidos formados por elipsóides duros (*hard ellipsoids*) foi muito estudado no passado (veja, por exemplo, Frenkel and Mulder (1985) e as referências citadas neste trabalho). Recentemente foram tratados sistemas de partículas elipsoidais (Sjoberg, (1999); Sjoberg and Mortensen (1997)). Exemplos destes sistemas podem ser soluções de proteínas e dispersões coloidais entre outros. Em seu trabalho combinado de simulação e experimentos em espalhamento a baixos ângulos, Sjöberg e colaboradores obtiveram resultados que não estão em concordância com comportamentos já conhecidos. Obtiveram resultados indicativos de que, mesmo para partículas estritamente esféricas e concentrações moderadas, os resultados analíticos de Percus-Yevick (Percus e Yevick, 1958) não estavam em concordância com as simulações. Isto está em discordância com várias outras comparações dos resultados de Percus-Yevick e dados de simulação. Foi posteriormente provado (Pedersen, 2002) que as discrepâncias tiveram sua origem em um tratamento errôneo por parte de Sjöberg et al das condições periódicas de contorno quando a intensidade era calculada.

Esta foi a principal motivação deste trabalho, onde refizemos estas simulações com o tratamento correto das condições periódicas de contorno. Como mostramos nas simulações a seguir, nossos resultados concordam inteiramente com a descrição teórica de Percus-Yevick, bem como mostramos que a “aproximação de desacoplamento” pode ser

aplicada para elipsóides de revolução mesmo para grandes anisotropias e frações de volume.

1.3.2. Simulações

Para a realização das simulações colocamos as partículas espalhadoras em uma caixa cúbica com condições periódicas de contorno. As partículas aqui são elipsóides de revolução com raio transversal unitário (por escolha) e anisotropia ε . A posição e orientação de cada elipsóide é dado por um vetor de 5 componentes $(x, y, z, \theta, \varphi)$, sendo (x, y, z) a posição em coordenadas cartesianas, e (θ, φ) a direção angular de seu eixo longitudinal em coordenadas esféricas (Figura 90).

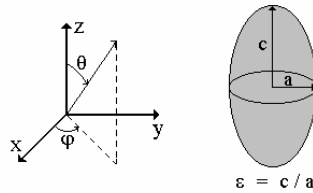


Figura 90 – Sistema de coordenadas e definição da anisotropia do elipsóide.

Dada a fração de volume η do sistema ($\eta = Nv/V$, onde N é o número de partículas, v é o volume de uma partícula e V o volume da cela), colocamos as partículas inicialmente em uma rede cúbica FCC onde o lado do cubo é calculado de acordo com a fração de volume desejada. Esta opção de colocar-se as partículas em uma rede FCC é aplicada para garantir-se que inicialmente não há superposição entre as partículas (Figura 91).

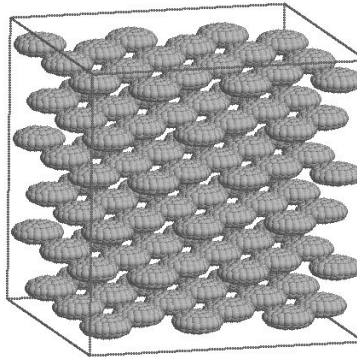


Figura 91 – Rede inicial para o cálculo Monte Carlo,

Nos passos Monte Carlo, variamos randomicamente em cada passo a posição e orientação de uma partícula. Após um período de equilíbrio (no qual a "ordem" existente na rede FCC é perdida) começamos os cálculos das funções de espalhamento desejadas (Figura 92).

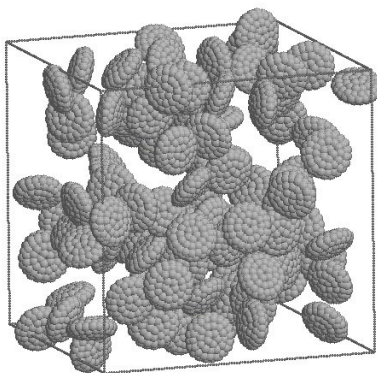


Figura 92 – Cella com partículas randomicamente orientadas.

A primeira dificuldade deste tipo de cálculo está no fato de que, em cada passo MC deve-se checar possíveis superposições. Para partículas esféricas este teste é simples, bastando apenas averiguar a distância entre os centros de duas partículas. Para elipsóides de revolução o teste é bem mais complexo. Nestes cálculos utilizou-se o algoritmo proposto por Perran e Wertheim (Perran e Wertheim, 1985), baseado no trabalho de Viellard-Baron (Viellard-Baron, 1972). Neste algoritmo dado um par de elipsóides A e B com mesma anisotropia calcula-se, com base em sua posição e orientação espacial, uma função definida positiva F_{AB} que fornecerá valor <1 para elipsóides com superposição, $=1$ para elipsóides cujas superfícies se tangenciam e >1 para elipsóides sem superposição. Este procedimento, que utiliza multiplicadores de Lagrange, foi completamente implementado e exaustivamente testado, pois a dinâmica correta da simulação Monte Carlo depende de um teste correto de superposição entre os elipsóides. Outra dificuldade na simulação consistiu na geração correta de direções randômicas. Como mostrado em diversos trabalhos, a geração de posições aleatórias na superfície de uma esfera não é um problema trivial, pois se não for feita corretamente não fornecerá uma distribuição uniforme (Cook, 1957). Para a geração das direções utilizamos o método de Marsaglia (Marsaglia, 1972) o qual apresenta os requisitos necessários para a correta geração destas direções.

Para o cálculo das funções de espalhamento utilizamos as equações (162) e (163) nas 13 direções de simetria da cela cúbica. A amplitude de espalhamento utilizada nos cálculos foi a de um elipsóide de revolução com semi-eixos $R, \epsilon R$:

$$F(\vec{q}) = F_1[q, r(q, \epsilon, \alpha)] \quad (165)$$

onde $r(q, \epsilon, \alpha) = R(\sin^2 \alpha + \epsilon^2 \cos^2 \alpha)^{1/2}$, com R o raio da secção transversal do elipsóide, α o ângulo entre o vetor de espalhamento e a direção de orientação do elipsóide, e F_1 o fator de forma de uma esfera (equação 67). O fator de forma para um sistema sem interação seria dados por:

$$P(q, R, \epsilon) = \int_0^{\pi/2} F_1[q, r(q, \epsilon, \alpha)]^2 \sin \alpha d\alpha \quad (166)$$

Em nosso cálculo utilizamos a equação (165), uma vez que a média em todas as orientações será dada pelo grupo de partículas. Neste sentido, a equação (166) será útil para calcularmos exatamente o fator de forma teórico para o dado elipsóide e compararmos com o resultado da simulação. Isto será um indicativo de que a orientação randômica dos elipsóides está sendo corretamente feita. A implementação do programa de simulação foi feita em Fortran 95 (arquivo ELLIPS.FOR). Os passos de cálculo do programa são:

- *Geração da cela unitária com as N partículas arranjadas em uma rede FCC.* A dimensão da cela cúbica depende da fração de volume desejada e da anisotropia da partícula. Na realidade o número N é dado por $N=4n^3$ onde n é o número de partículas em uma direção da rede FCC;
- *Equilibração do sistema.* O sistema é equilibrado fazendo-se $10^4 \cdot N$ passos Monte Carlo. Este número mostrou-se ser suficiente para todos os casos simulados. Neste período nenhuma função de espalhamento é calculada;
- *Cálculo das funções de espalhamento.* Valores típicos do número de passos de Monte Carlo utilizados nesta etapa foram $10^6 - 10^8$, dependendo do número de partículas na cela. O cálculo das funções de espalhamento é descrito no texto.
- *Gravação dos resultados.* Após o cálculo e geração dos valores médios, para todas as direções, os dados são exportados em arquivos texto.

Para possibilitar a visualização da dinâmica do sistema, ao longo da simulação algumas "fotos" são tiradas. Estas "fotos" são arquivos texto com as coordenadas e

orientação de cada elipsóide. Escrevi um programa que gera a partir destes arquivos texto, arquivos *protein Data Bank* (Berman et al, 2000), onde se constrói cada elipsóide de revolução utilizando subunidades esféricas (utilizando um algoritmo utilizado no programa BUILDER; Oliveira, 2001). Estes arquivos permitem a visualização tridimensional do sistema bem como de sua evolução ao longo da simulação.

Juntamente com as funções de espalhamento (intensidade espalhada, fator de estrutura e fator de forma) calculo também a distribuição de vizinhos do sistema. Para a computação desta função utilizamos a seguinte expressão:

$$g(r) = \sum_{i=1}^N \frac{L^3}{N \cdot 4\pi r^2 \cdot \Delta r} n(r, \Delta r, R_i) \quad (167)$$

sendo N o número de partículas, L o lado da cela unitária, R_i a posição de uma dada partícula, r o raio da casca esférica de correlação considerada com espessura Δr e n o número de partículas contida nesta casca esférica. O interesse de se calcular esta função consiste no fato de que ela se relaciona com o fator de estrutura através da expressão:

$$S(q) = 1 + \frac{N}{V} \int_0^\infty 4\pi r^2 [g(r) - 1] \frac{\sin(qr)}{qr} dr \quad (168)$$

Além disso, desta função de distribuição de vizinhos pode-se obter informações sobre a correlação de partículas no sistema, sendo por isso também chamada de função de correlação do sistema. Esta função somente possui expressão analítica exata para partículas esféricas.

1.3.3. Resultados

As simulações foram realizadas para variadas anisotropias das partículas espalhadoras. Para termos uma boa estatística utilizamos uma caixa com $N = 1372$ partículas. O número de passos de equilíbrio foi $10^4 \cdot N$ e para simulação utilizamos 10^8 passos. As simulações foram realizadas em computadores Pentium 4 de 1.6 e 2.5 GHz. Tipicamente, cada simulação completa requereu ~50 horas de máquina.

Como partimos de uma rede FCC temos que considerar as máximas frações de volume que este arranjo permite. Este limite dependerá da anisotropia (razão $\epsilon=c/a$) do elipsóide considerado. Na Figura 93 mostro a predição teórica das frações de volume máximas para um sistema FCC.

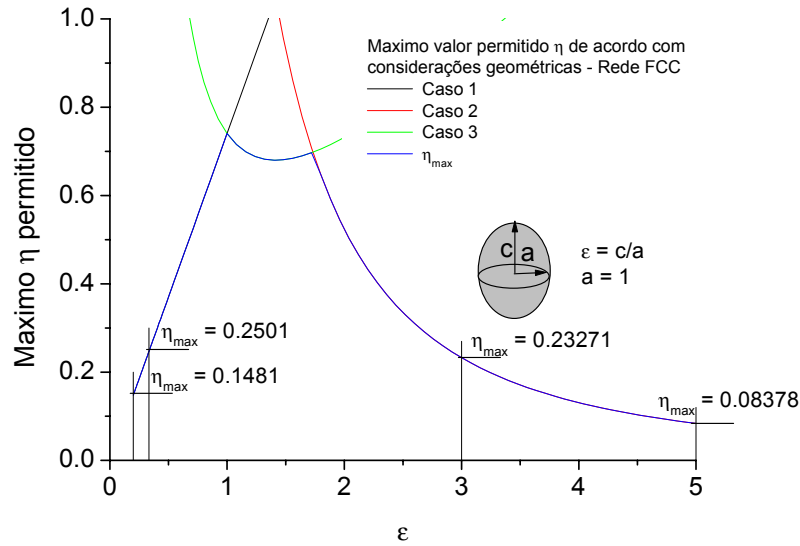
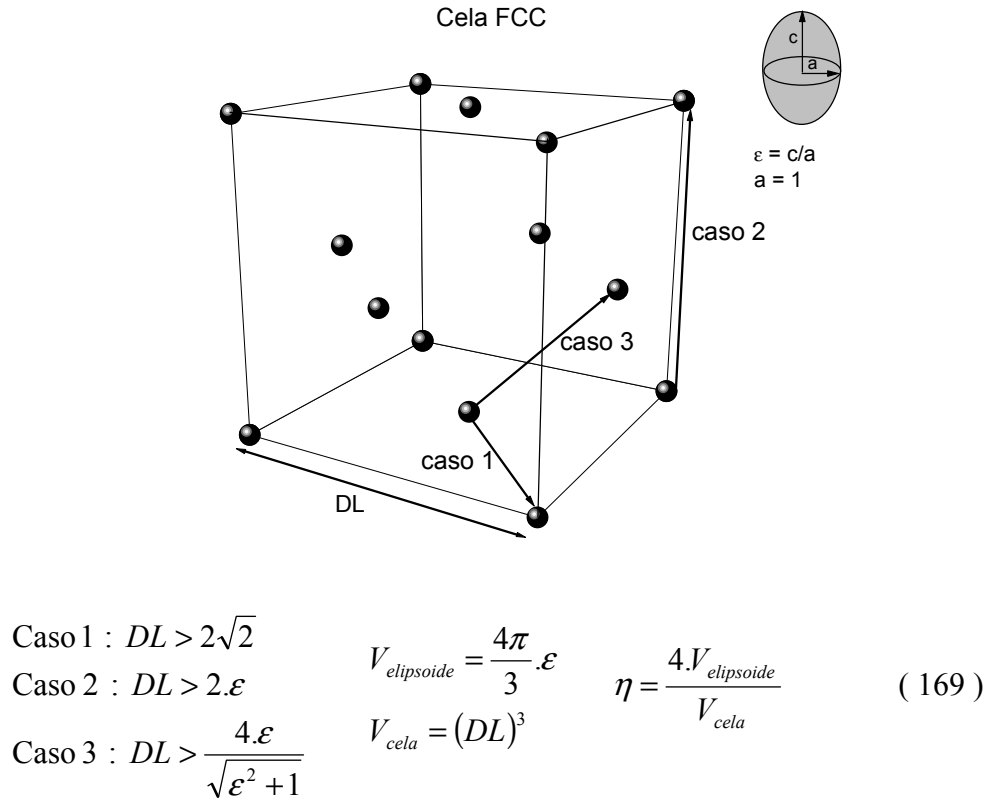


Figura 93 – Cálculos teóricos para as frações de volume possíveis em uma rede FCC de elipsóides de revolução. Acima) Definição dos vetores na rede FCC e condições de não superposição das partículas. Abaixo) Representação gráfica da máxima fração de volume em função da anisotropia ε .

Este cálculo nos permite saber a máxima fração volumétrica permitida para a configuração inicial em uma rede FCC composta de elipsóides de revolução. Aqui apresentarei cálculos para diversas anisotropias (1/5, 1/3, 1/2, 1, 2, 3, 5) e frações volumétricas (0.01, 0.03, 0.06, 0.1 (também 0.2, 0.3 e 0.4 em alguns casos)).

$\epsilon = 1$ (partículas esféricas)

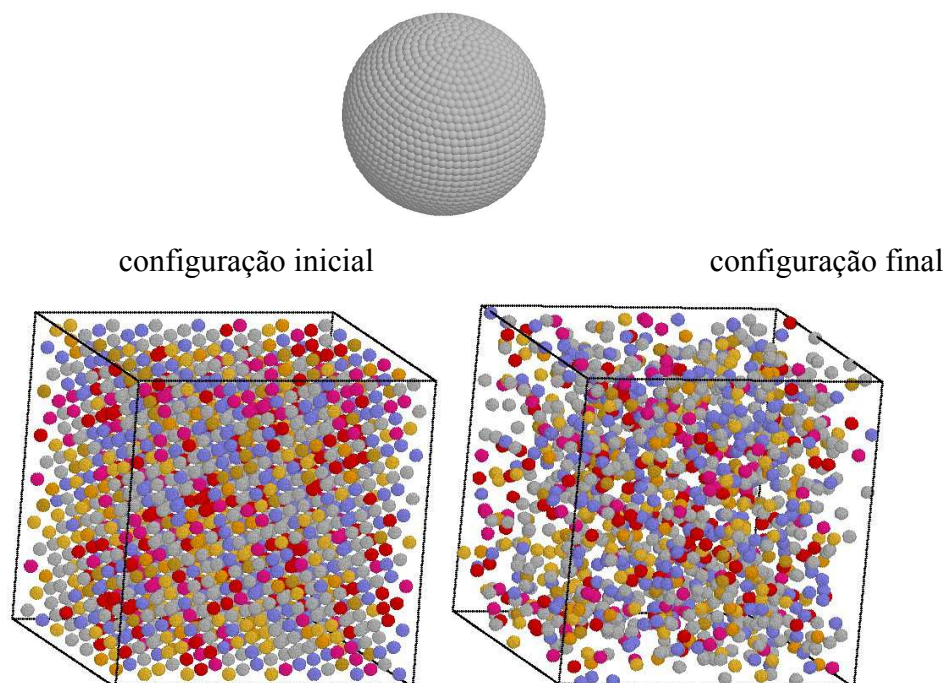


Figura 94 – Simulação para $\epsilon=1$ e $\eta=0.06$. Representação da partícula e configuração inicial e final do sistema^{***}.

Utilizando a metodologia e fórmulas descritas anteriormente realizamos as simulações Monte Carlo inicialmente para o caso de partículas esféricas. Estas simulações caracterizam um ótimo teste para o programa uma vez que o caso de partículas esféricas pode ser resolvido analiticamente (Percus e Yevick, 1958). Na Figura 95 apresento os gráficos de Intensidade, Fator de Estrutura e Função de correlação para este sistema para vários valores de fração de volume. Sobreposto aos dados simulados, mostro o cálculo teórico para este sistema. Na Figura 95C temos a aproximação de desacoplamento aplicada aos dados.

^{***} Nesta figura e em todas as figuras representativas das configurações do sistema, as cores nos elipsóides de revolução são apenas para facilitar a visualização. Todos os elipsóides envolvidos no cálculo são idênticos.

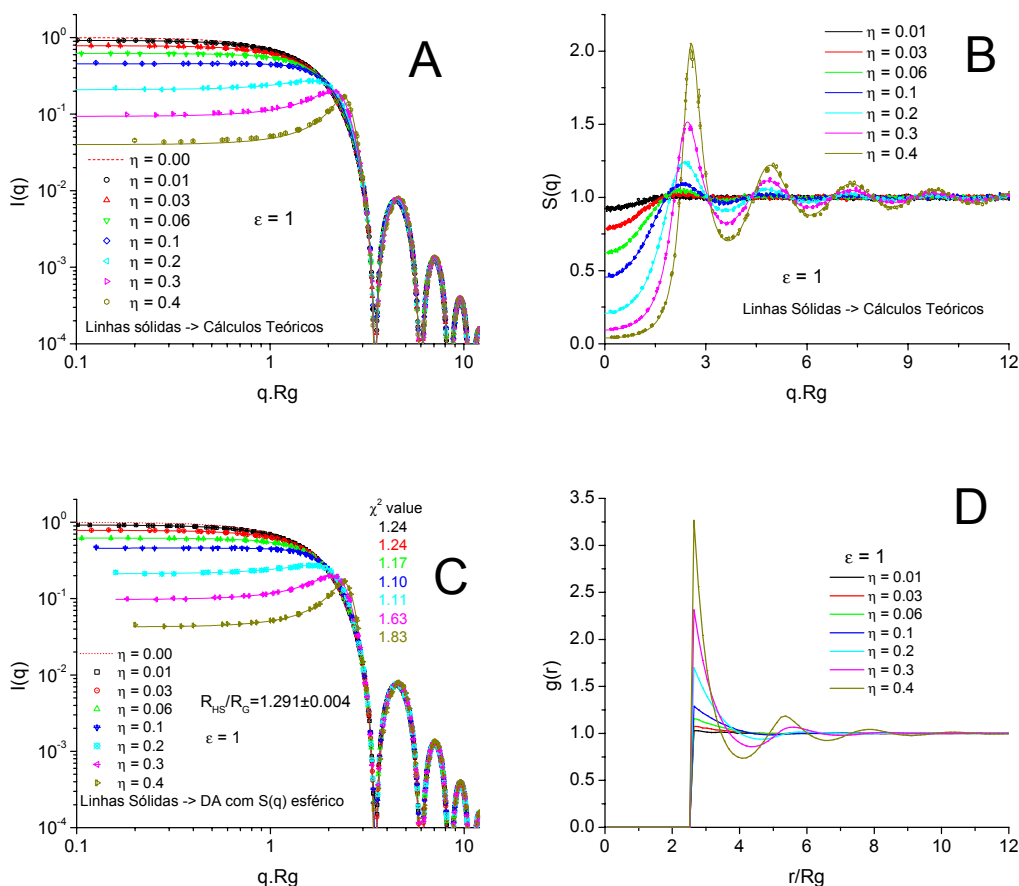


Figura 95 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas esféricas. Para a intensidade e o fator de estrutura o resultado teórico é plotado conjuntamente.

Vemos uma perfeita concordância com os dados teóricos mesmo para frações de volume elevadas. Isto nos deu a segurança de que a metodologia aplicada ao programa era válida para o caso de partículas esféricas ($\epsilon = 1$). Para que os dados fiquem adimensionais as escalas foram normalizadas pelo raio de giro da partícula espalhadora, tornando-os assim independentes do raio das esferas escolhidas na simulação (veja Tabela 1). É interessante notar que mesmo as oscilações provenientes das altas frações de volume, que aumentam a correlação espacial entre as partículas, são corretamente descritas pela simulação. Também, como esferas possuem um raio bem definido, é fácil compreender porque a função de correlação possui a queda abrupta para pequenos valores de r . Para baixas concentrações ($\eta=0.01$ ou menor) as partículas praticamente não possuem correlação espacial e assim podem ocupar quaisquer posições umas com relação às outras. Pelo fato de serem esferas

rígidas, sua distância sempre deverá ser maior do que $2R$, sendo R o raio da esfera. Como $R_g = \sqrt{3/5}R$ para esferas, isto significa que a mínima distância entre duas esferas seria $r/R_g \sim 2.58$, como indicado na Figura 95D. Para distâncias menores que esta não existirão distâncias entre pares de esferas e por isso a função de correlação possui valor zero. Para valores maiores do que este, quaisquer distâncias entre pares de esferas são possíveis e por isso a função de correlação tende a um valor unitário. No limite de baixas concentrações ($\eta \rightarrow 0$) esta função assume a forma de uma função degrau. Para altas concentrações de partículas a posição de uma partícula acabará influenciando no posicionamento (ou movimentação) das partículas vizinhas gerando a correlação entre suas posições o que se reflete em oscilações tanto na função de correlação quanto o fator de estrutura. Porém em todos os casos temos para a função de correlação uma queda abrupta para $r/R_g \sim 2.58$. A aproximação de desacoplamento (que é exata para o caso esférico) é mostrada na Figura 95C. Vemos um ótimo ajuste com baixos valores de χ^2 , indicando ótimo ajuste.

A título de comparação mostro na Figura 96 nossos dados e os dados obtidos por Sjöberg (Sjöberg, 1999) para simulações MC de esferas rígidas. Como mencionado antes, as simulações de Sjöberg não estão em concordância com os valores teóricos nem para pequenas frações de volume. Nossas simulações por outro lado estão em ótima concordância.

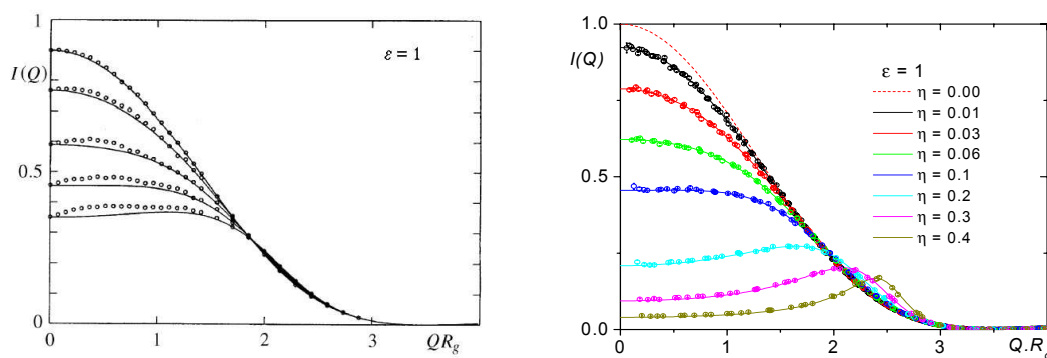


Figura 96 – Simulações de Monte Carlo para esferas rígidas. Esquerda- Resultado de Sjöberg (tirado de Sjöberg, 1999). Direita- Nossas simulações. Em ambos os casos as curvas sólidas representam o resultado teórico exato.

$$\epsilon = 1/2$$

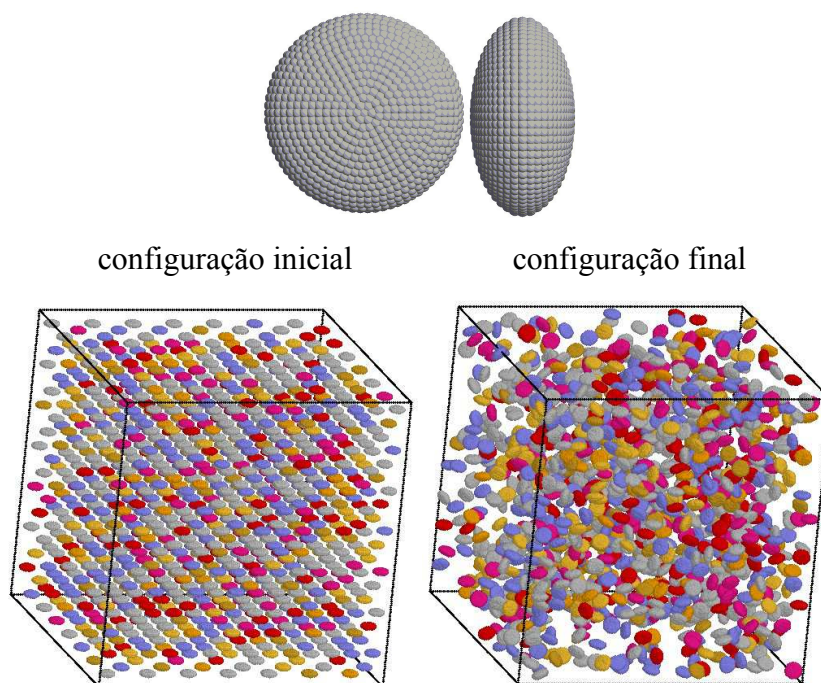


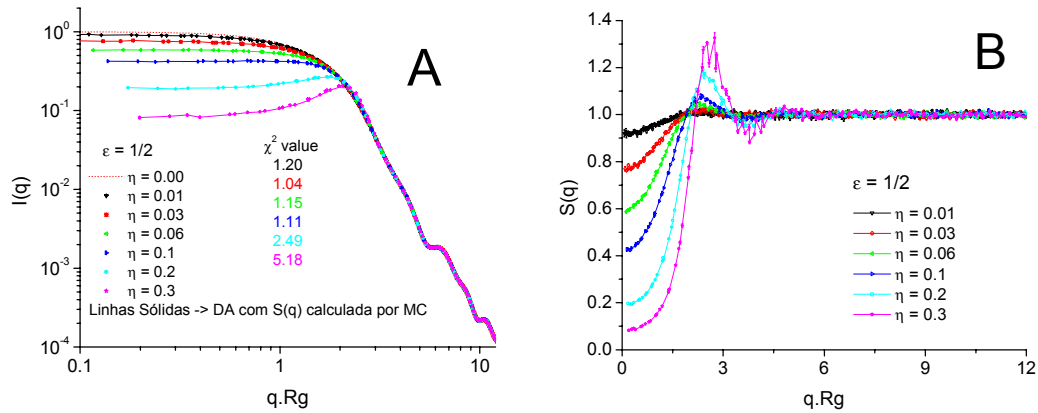
Figura 97 – Simulação para $\epsilon=1/2$ e $\eta=0.06$. Representação da partícula e configuração inicial e final do sistema.

Mostremos agora resultados para casos onde $\epsilon \neq 1$, ou seja, elipsóides de revolução prolatos e oblatos. Inicialmente consideraremos pequenas anisotropias. A seguir temos simulações para elipsóides com $\epsilon = 1/2$. Neste caso e nos seguintes teremos que, nas figuras “A” mostramos a intensidade de espalhamento calculada e a aproximação de desacoplamento em linhas sólidas utilizando o fator de estrutura calculado nos passos de MC. Nas figuras “B” e “D” mostramos o fator de estrutura e a função de correlação respectivamente. Nas figuras “C” mostramos a intensidade de espalhamento proveniente da simulação MC e a aproximação de desacoplamento calculada utilizando fator de estrutura de esferas.

Na Figura 98A vemos que a ótima superposição dos dados calculados e a aproximação de desacoplamento. Isso nos indica que ela é válida para este caso, para todas as frações de volume descritas. Na Figura 98C onde aplicamos a aproximação de desacoplamento com o fator de estrutura de esferas vemos que os ajustes até $\eta=0.2$ são muito bons, indicando que o cálculo utilizando o fator de estruturas de esferas, apesar de não estar correto, consegue descrever satisfatoriamente o comportamento do sistema. Neste

ajuste deixamos livres os valores de raio da esfera e fração de volume. Em todos os casos o valor de fração de volume ajustado correspondeu ao mesmo valor calculado na simulação de MC. Sendo assim, a variável importante neste ajuste é o valor do raio equivalente da esfera rígida que ajusta os dados simulados. Neste caso $R_{HS}=1.06$ ($R_{HS}/R_G=1.58$, sendo R_G o raio de giro do elipsóide).

No gráfico do fator de estrutura (Figura 98B) percebemos uma atenuação as oscilações desta função quando comparada com o caso de partículas esféricas. Para compreender porque isso ocorre olhemos para a função de correlação (Figura 98D). Diferentemente do caso anterior, a interface entre a região de não correlação e correlação total ($= 1$) não é abrupta, mas possui uma curvatura. Isto ocorre porque na dinâmica do sistema a mínima distância entre um par de elipsóides randomicamente orientados não será única, mas dependerá da orientação relativa de ambos. Aumentando-se a fração de volume esta curvatura muda suavemente e são geradas oscilações na função. Estas oscilações também possuem uma característica muito diferente do caso esférico, possuindo poucos máximos e mínimos.



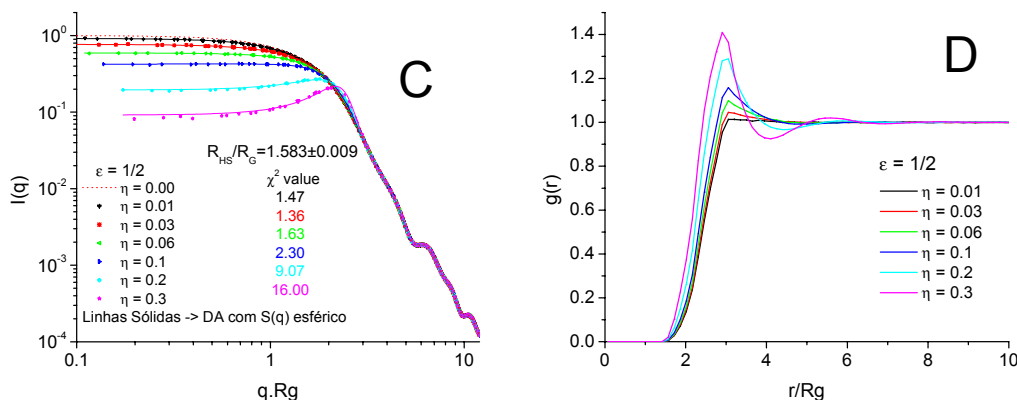


Figura 98 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas oblatas com $\epsilon = 1/2$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

Também na função de correlação, percebemos que a parte inicial da mesma cresce suavemente partindo do zero e, após atingir um máximo (mesmo para $\eta = 0.01$) tende à unidade. Como veremos, elipsóides prolatos possuirão outra característica para esta função. Nos gráficos “A” e “C” mostro em linhas tracejadas o fator de forma teórico para o respectivo elipsóide. Isto é importante porque indica se a média em todas as orientações foi corretamente feita para todas as frações de volume. Como dissemos anteriormente, essa média é feita no próprio grupo de elipsóides, onde assumimos que nos passos Monte Carlo os elipsóides assumem todas as orientações possíveis. Se essa média for corretamente feita, para altos valores de “ q ” onde os efeitos do fator de estrutura não são mais vistos, a intensidade calculada e deve ser igual ao fator de forma teórico.

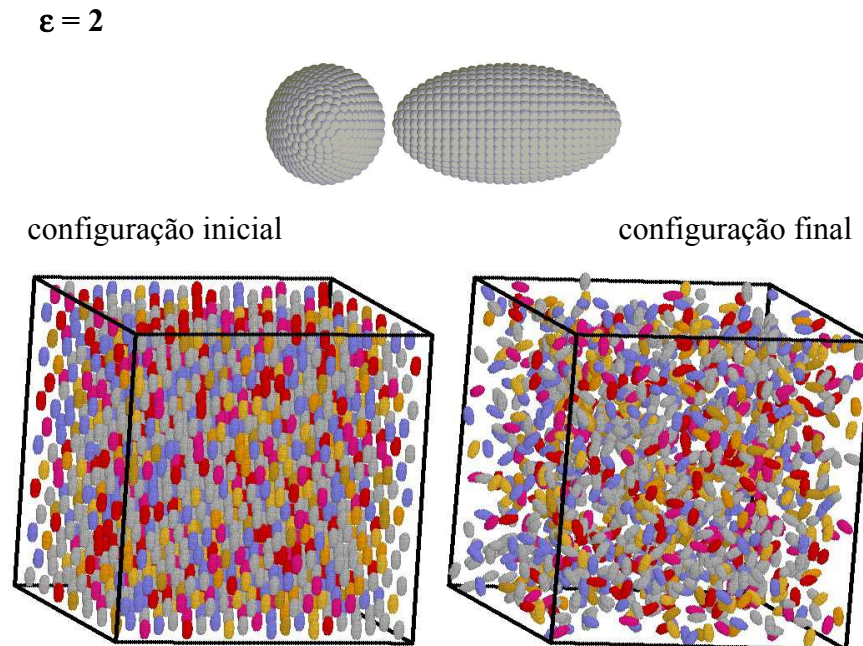


Figura 99 – Simulação para $\epsilon = 2$ e $\eta=0.06$. Representação da partícula e configuração inicial e final do sistema.

Vemos também neste caso que a aproximação de desacoplamento com o fator de estrutura calculado é válida na descrição destes dados uma vez que possui ótima concordância com os dados simulados. Já para o fator de estrutura esférico (Figura 100C) vemos que a descrição é razoável, ficando ruim para $\eta=0.2$. Isso nos indica que a utilização de fatores de estrutura esféricos fornecem ajustes satisfatórios para baixas anisotropias ($\epsilon=2$, $1/2$).

Como no caso anterior as oscilações no fator de estrutura e função de correlação são atenuadas. Isto indica que a presença de uma anisotropia seja prolata ou oblata, diminui a amplitude das oscilações nestas funções. O aspecto destas oscilações também é diferente nos dois tipos de anisotropia. Olhando para a função de correlação vemos que também deixa de possuir uma queda abrupta, para ter uma curvatura. No entanto, a característica desta curvatura é completamente diferente do caso oblato. O início da curva cresce de modo abrupto, e não de modo suave como no caso anterior. Por outro lado, após atingir o máximo valor, esta função decai de modo mais suave para o valor unitário. Logo, dependendo-se do tipo da anisotropia, teremos um diferente comportamento da função de correlação.

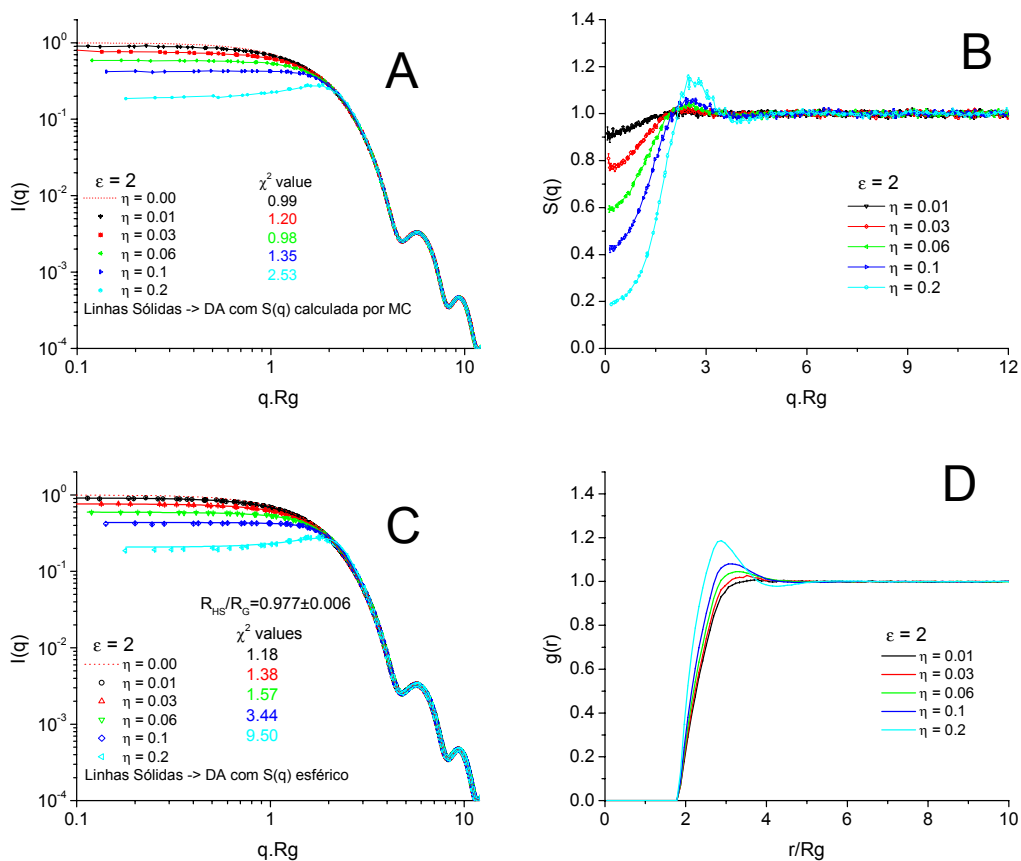


Figura 100 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas prolatas com $\epsilon = 2$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

$$\varepsilon = 1/3$$

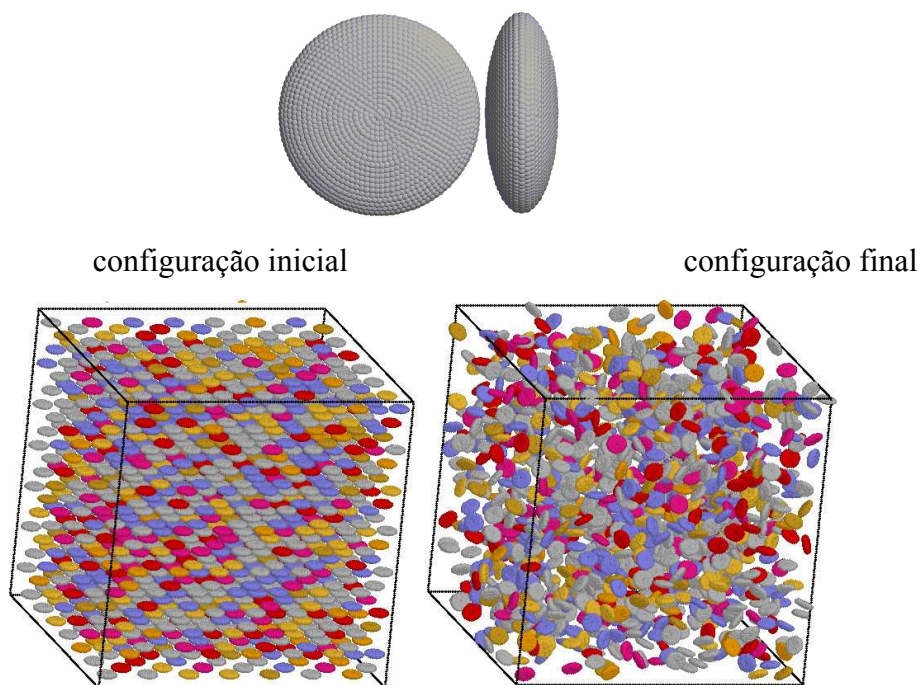


Figura 101 – *Simulação para $\varepsilon = 1/3$ e $\eta = 0.06$. Representação da partícula e configuração inicial e final do sistema.*

Aumentando-se a anisotropia percebemos que as observações tiradas nos casos anteriores se mantêm. As oscilações no fator de estrutura e função de correlação ficam mais atenuadas do que o caso $\varepsilon = 1/2$. Ao passo que a aproximação de desacoplamento continua perfeitamente válida para o caso onde utilizamos o fator de estrutura calculado nos passos MC (Figura 102A), a utilização do fator de estrutura esférico para esta anisotropia fornece ajustes piores (maiores valores de χ^2) do que os vistos para anisotropias menores (Figura 101C). Com isso temos a indicação de que a utilização de fatores de estrutura esféricos fica comprometida com o aumento da anisotropia da partícula. A forma da função de correlação, destacada anteriormente se mantém, possuindo uma variação suave no início da curva e uma queda abrupta após o máximo da função. Isto nos dá a indicativa de que este comportamento está ligado ao formato oblato, independente de sua anisotropia.

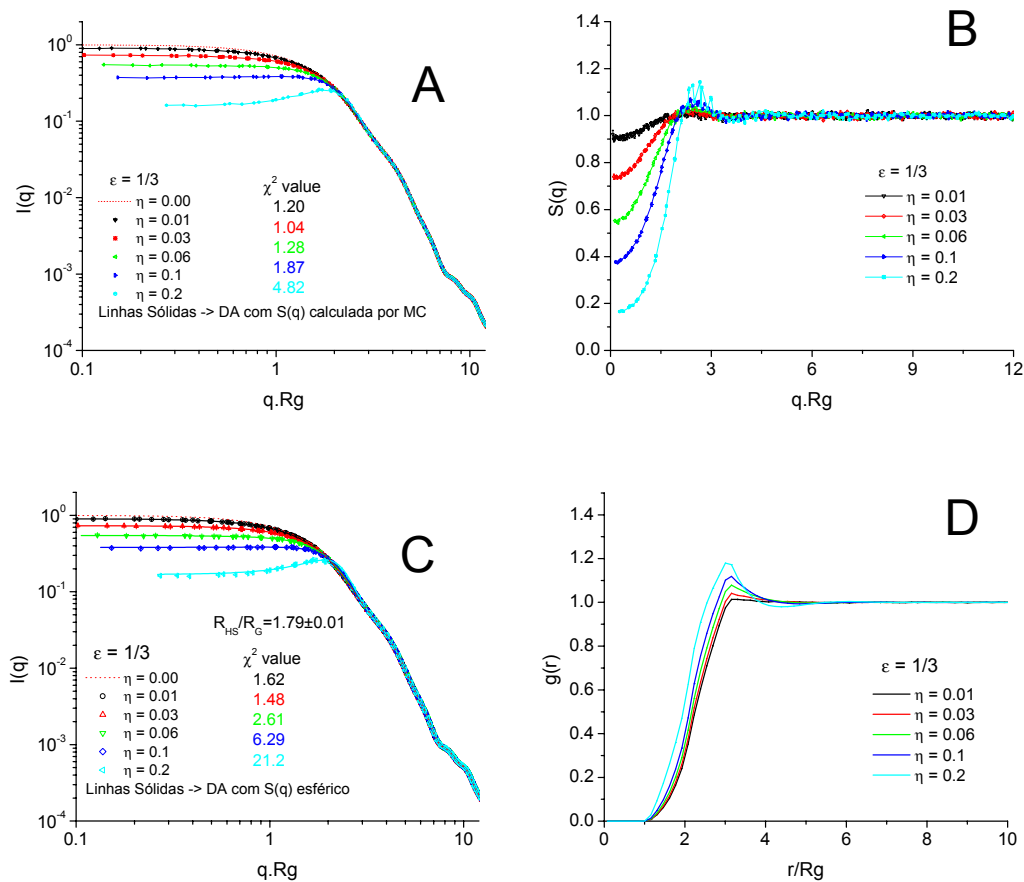


Figura 102 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas oblatas com $\varepsilon = 1/3$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

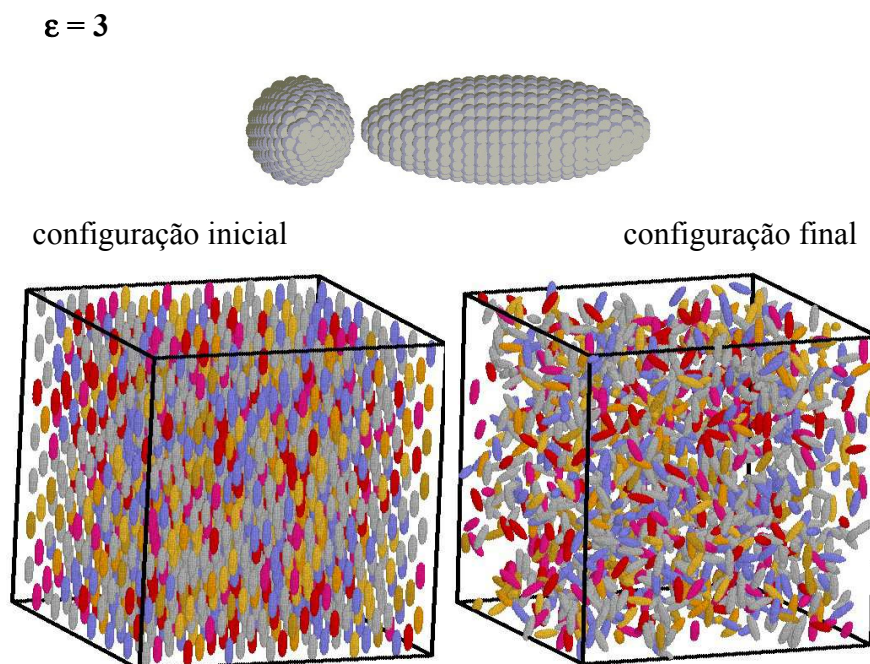


Figura 103 – Simulação para $\epsilon = 3$ e $\eta = 0.06$. Representação da partícula e configuração inicial e final do sistema

Similarmente ao caso anterior, a aproximação de desacoplamento é perfeitamente válida se utilizamos o fator de estrutura calculado (Figura 104A) e limitada se utilizamos fatores de estrutura esféricos (Figura 104C). A aproximação forneceu bom ajuste dos dados até concentrações de 0.06, ficando ruim para maiores valores de fração de volume.

A função de correlação mantém a mesma característica destacada anteriormente para o caso $\epsilon = 2$, com o início da curva apresentando um crescimento abrupto sendo seguido de um máximo. Após este máximo vê-se uma queda mais suave do que no caso com menor anisotropia.

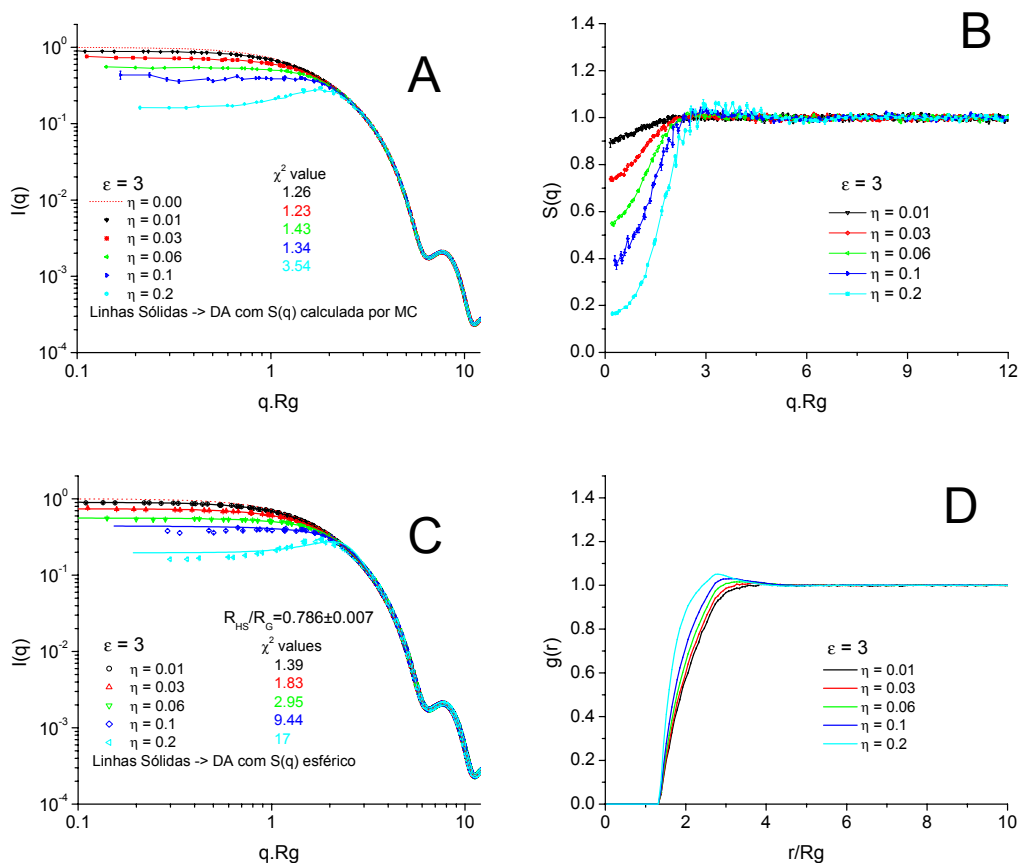


Figura 104 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas prolatas com $\varepsilon = 3$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

$$\varepsilon = 1/5$$

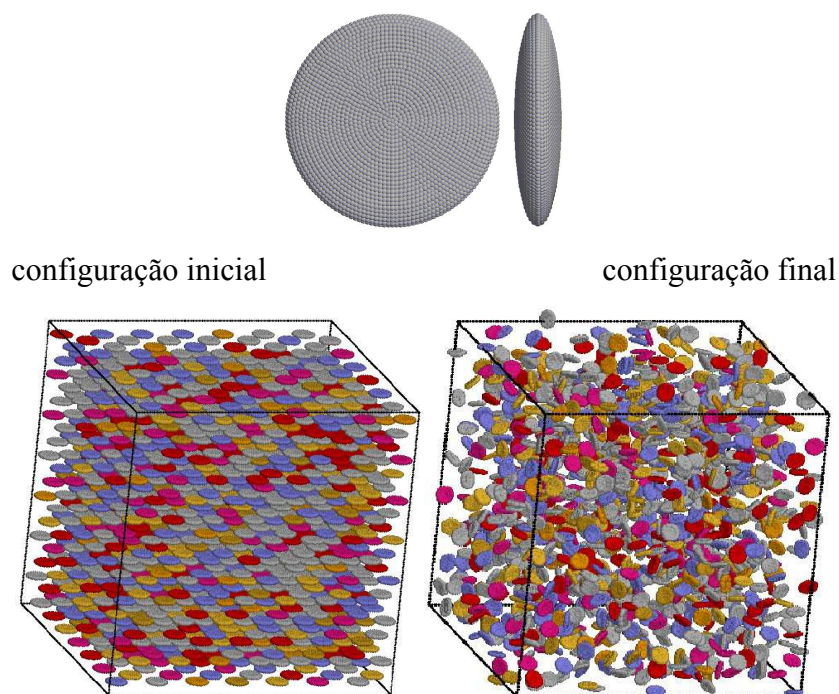


Figura 105 – Simulação para $\varepsilon = 1/5$ e $\eta = 0.06$. Representação da partícula e configuração inicial e final do sistema

Aqui temos um caso com anisotropia bastante acentuada. Como já notado nos casos anteriores, as oscilações diminuem bastante e praticamente somente são observadas em $\eta = 0.1$. A característica da função de correlação se mantém, possuindo máximos mais atenuados. Mesmo para esta anisotropia a aproximação de desacoplamento com o fator de estrutura calculado (Figura 106A) continua válida indicando que ainda poderia ser aplicada neste caso. Já para o caso onde utilizamos fatores de estrutura esféricos (Figura 106C) vemos que somente pode ser aplicada satisfatoriamente até $\eta = 0.06$, fornecendo ajustes ruins para maiores concentrações.

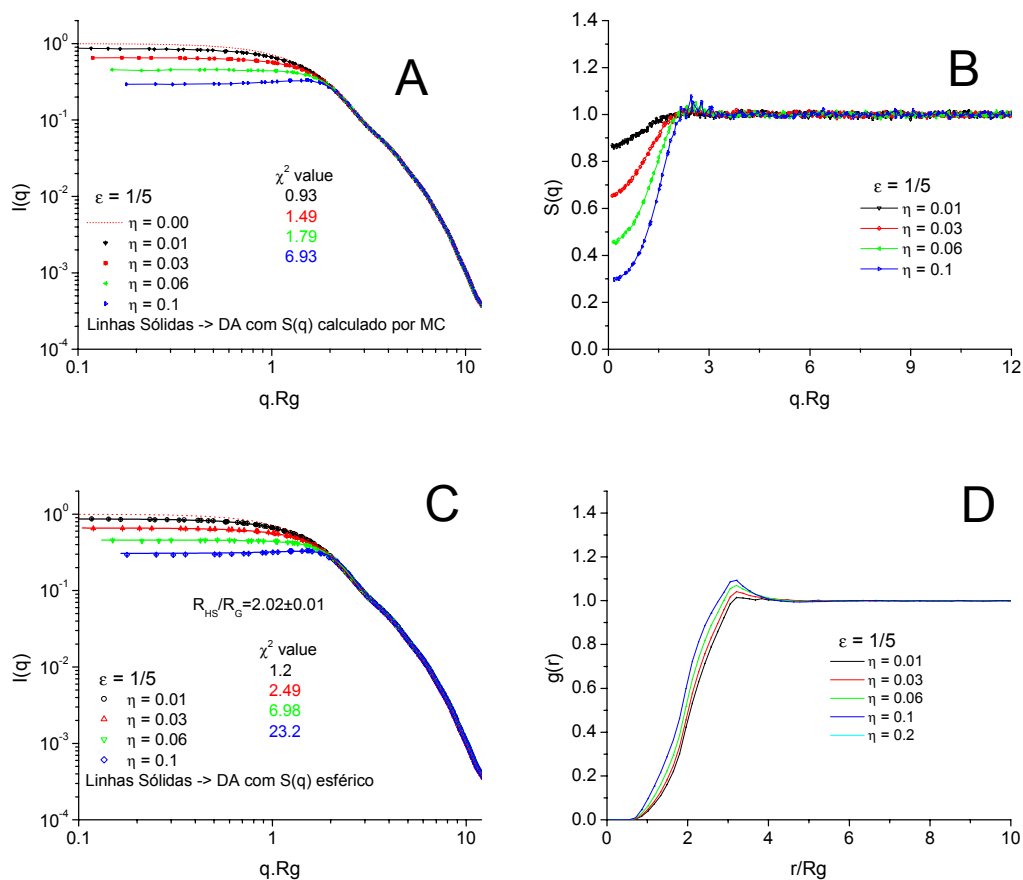


Figura 106 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas oblatas com $\epsilon = 1/5$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

$$\varepsilon = 5$$

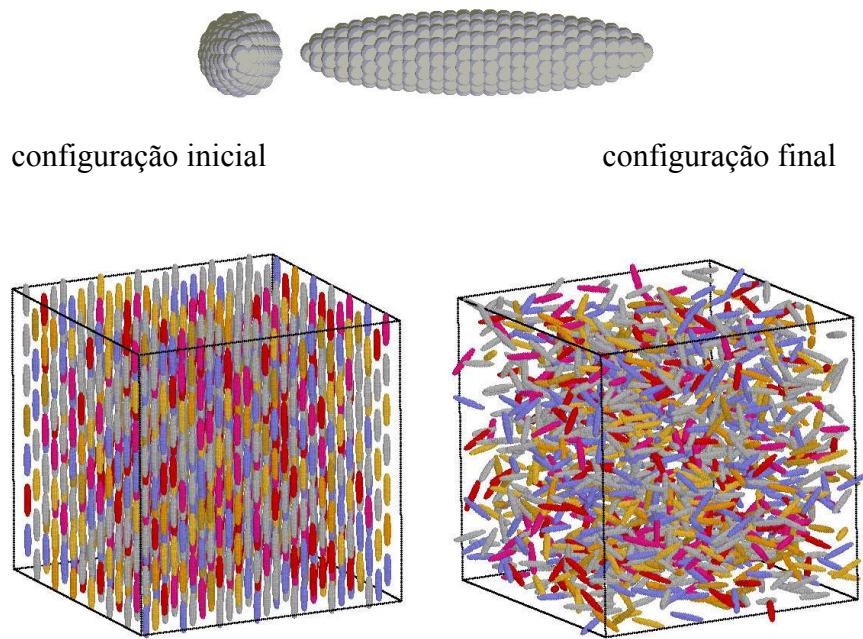


Figura 107 – *Simulação para $\varepsilon = 5$ e $\eta = 0.06$. Representação da partícula e configuração inicial e final do sistema.*

Neste caso teremos elipsóides extremamente alongados e, se observarmos a Figura 93 vemos que para $\eta = 0.1$ a configuração inicial apresentava superposição na rede FCC. Coloquei no algoritmo do programa uma rotina que busca obter uma configuração, com movimentos e rotações dos componentes da configuração inicial, sem superposição entre os elipsóides. Esta metodologia funcionou bem em vários casos, permitindo o cálculo de intensidades de espalhamento mesmo para casos de superposição na rede FCC inicial.

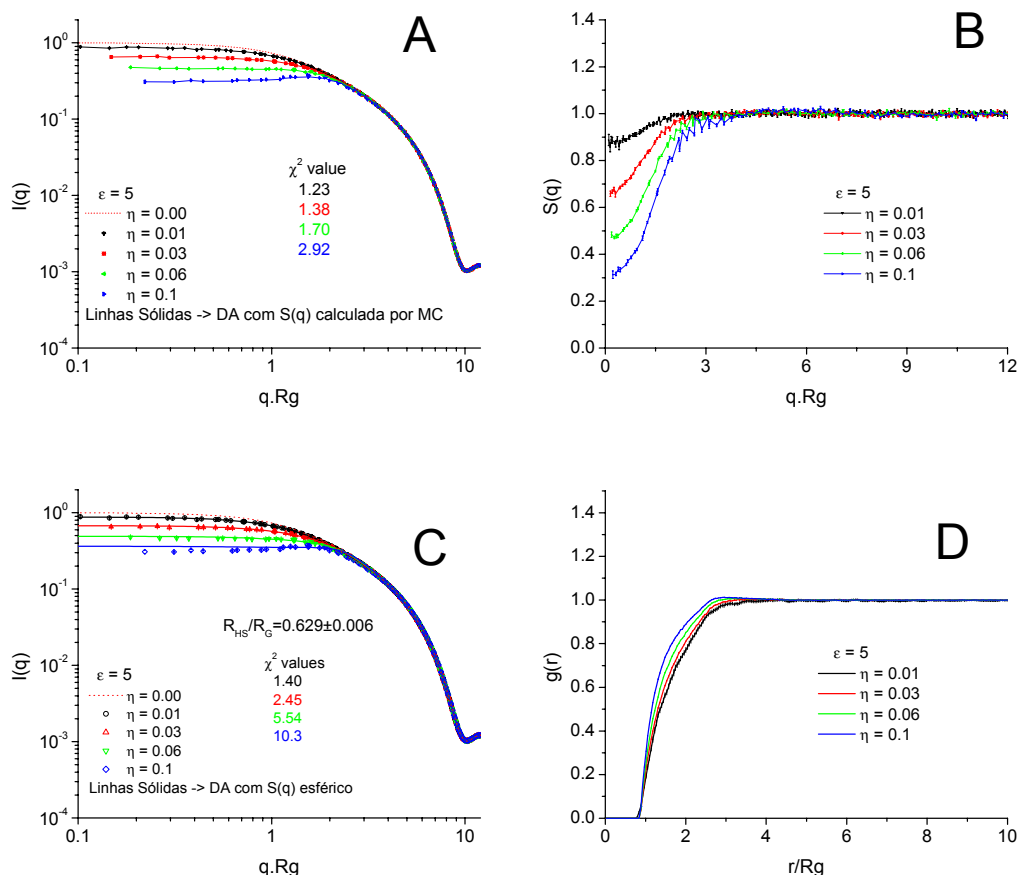


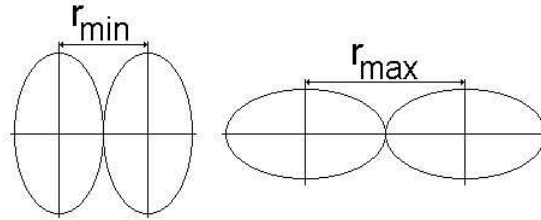
Figura 108 – Funções calculadas pela simulação de Monte Carlo. Intensidade de espalhamento $I(q)$, fator de estrutura $S(q)$ e função de correlação $g(r)$ para partículas oblatas com $\varepsilon = 5$. Para a intensidade a aproximação de desacoplamento é mostrada em linhas sólidas.

Como nos casos anteriores, vemos uma boa descrição dos dados pela aproximação de desacoplamento com fatores de estrutura calculados. Já a utilização de fatores de estrutura esféricos vemos que é válida principalmente até $\eta=0.06$.

As oscilações do fator de estrutura são ainda mais atenuados do que nos casos anteriores. A função de correlação possui um crescimento abrupto em seu início e, após o máximo, uma diminuição muito suave para o valor unitário, mesmo no caso de maior fração de volume.

Comportamento das funções de correlação e variação de R_{HS}

Como vimos, as funções $g(r)$ apresentaram um comportamento interessante, apresentando diferenças para o caso oblato e prolato. Diferentemente de uma esfera onde tínhamos um aumento abrupto do valor de $g(r)$, tivemos um aumento suave de seu valor. Para $\epsilon < 1$ (oblato) a função $g(r)$ aumenta com derivada zero em seu início e para $\epsilon > 1$ este aumento ocorre de forma abrupta. Nos dois casos o primeiro valor não nulo de $g(r)$ corresponde à mínima distância permitida entre dois elipsóides. Este valor ($r_{\min}$) pode ser predito teoricamente por geometria simples:



$$\left. \begin{array}{l} r_{\min} / R_{G \text{ oblate}} \\ r_{\max} / R_{G \text{ prolato}} \end{array} \right\} = \frac{2\epsilon\sqrt{5}}{\sqrt{2+\epsilon^2}} \quad (170)$$

$$\left. \begin{array}{l} r_{\min} / R_{G \text{ prolato}} \\ r_{\max} / R_{G \text{ oblate}} \end{array} \right\} = \frac{2\sqrt{5}}{\sqrt{2+\epsilon^2}}$$

Além disso, para partículas oblatas o primeiro máximo na função $g(r)$ está posicionado praticamente na mesma posição em todas as frações de volume. O valor de r neste caso corresponde à máxima distância entre centros para dois elipsóides tangentes ($r_{\max}$). Por outro lado, para partículas prolatas este primeiro máximo de $g(r)$ varia com os valores de η , fornecendo valores menores do que $r_{\max}$ dado na equação (170). Este comportamento é mostrado na Figura 109 onde mostramos o primeiro valor não nulo e a posição do primeiro máximo da função $g(r)$ em comparação aos valores de $r_{\min}$ e $r_{\max}$ esperados:

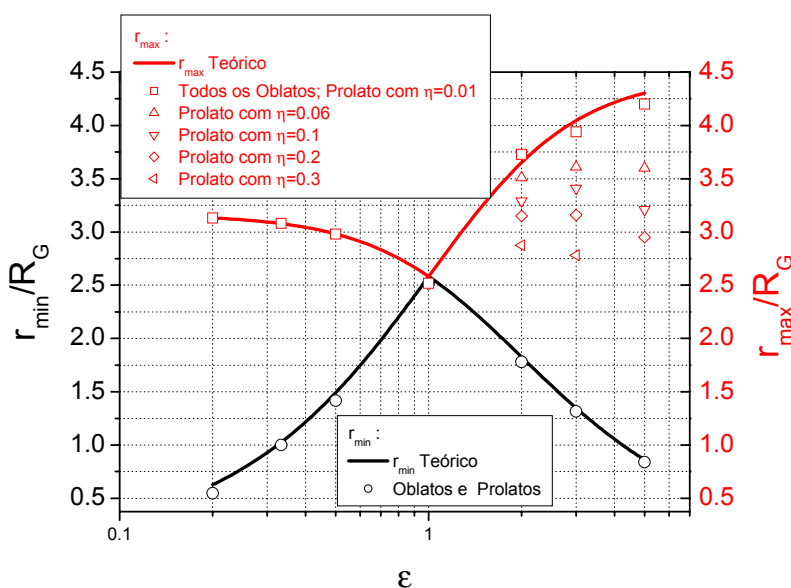


Figura 109 – Variação de r_{min} (preto) e r_{max} (vermelho) com a anisotropia da partícula ϵ . Detalhes no texto.

Deste resultado vemos que a mínima distância tangente é corretamente obtida pelas funções $g(r)$. O primeiro máximo da função $g(r)$ para partículas oblatas é diretamente relacionado à máxima distância entre centros para elipsóides tangentes. No entanto, para partículas prolatas isso não ocorre, fornecendo valores menores com o aumento da concentração do sistema. Esta notável diferença entre o comportamento das funções $g(r)$ indica também diferentes propriedades termodinâmicas para estes sistemas uma vez que esta função é relacionada à função de partição do sistema (Hansen e McDonald, 1990, Pathria, 1996). Este resultado é similar ao obtido por Frenkel e Mulder (Frenkel e Mulder, 1985) utilizando outros métodos de cálculo.

Aumentando-se a anisotropia, a aproximação de desacoplamento fornece valores maiores de χ^2 tanto para o fator de estrutura simulado quanto o de esferas. No entanto, comparativamente o fator de estrutura simulado fornece valores de χ^2 muito menores, indicando que mesmo para altas anisotropias e frações de volume essa aproximação é válida.

Porém, para fatores de estrutura esféricos o aumento da anisotropia ϵ diminui o intervalo de validade desta aproximação. Para $\epsilon=1/3$ (3) esta aproximação foi válida até $\eta=0.1$ e para $\epsilon=1/5$ (5) até $\eta=0.06$. Apesar disso, a possibilidade de se utilizar fatores de

estrutura esféricas mesmo para altas anisotropias e frações de volume moderadas é um resultado notável, justificando sua utilização como primeira aproximação. Do ajuste utilizando fatores de estrutura esféricos, obtínhamos o raio de esfera rígida R_{HS} equivalente em cada caso. Desta forma pudemos obter uma relação empírica entre o R_{HS} equivalente e a anisotropia da partícula ε (Figura 110):

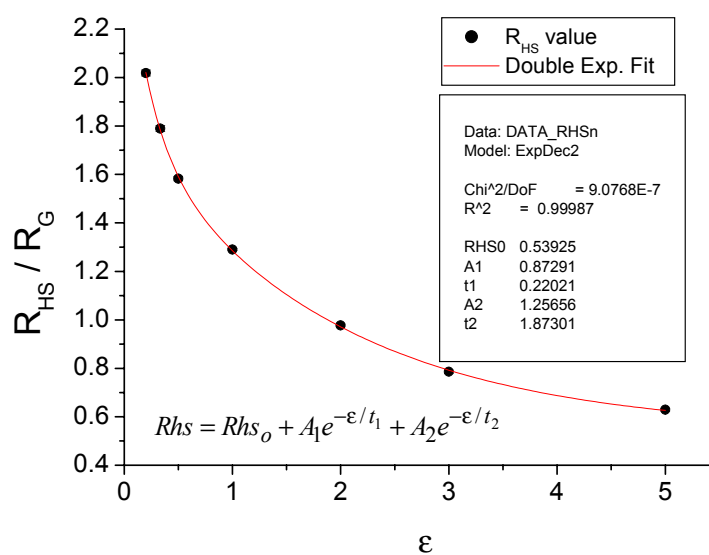


Figura 110 – Relação entre o raio de esfera rígida utilizado no cálculo de fator de estrutura R_{HS} e a anisotropia das partículas ε . A linha contínua corresponde à uma aproximação empírica por duas exponenciais.

Da ótima concordância ($R_{fator} \sim 1$) vemos que uma exponencial decrescente dupla descreve precisamente este comportamento. Uma relação (mesmo que empírica) é útil, pois permite a simulação da intensidade de espalhamento $I(q)$ para valores intermediários de ε ou, para um dado ajuste utilizando o fator de estrutura de esferas rígidas, obter o fator de estrutura $S(q)$ “correto” tomando-se o valor da anisotropia a partir do R_{HS} equivalente. Uma outra possibilidade que poderia ser testada é extrapolar este resultado para outros tipos de interação (esferas carregadas, esferas macias, etc) que são muito difíceis de se calcular (ou simular) utilizando elipsóides de revolução mas que, para esferas, possuem expressões analíticas fechadas.

1.3.4. Conclusões Sobre os Resultados

Apresentamos aqui resultados de simulações Monte Carlo para sistemas compostos de elipsóides de revolução em concentrações finitas. Como vimos pelas curvas apresentadas, os resultados foram muito bons, permitindo o cálculo das funções de espalhamento para este tipo de sistema. Diversos comportamentos e tendências puderam ser observados, permitindo a correlação entre a variação da anisotropia da partícula e a mudança no fator de estrutura e função de correlação. Mostramos claramente que a aproximação de desacoplamento pode ser aplicada para sistemas de elipsóides de revolução para as mais variadas anisotropias. Além disso, nossos dados corrigem as incoerências encontradas por Sjöberg em suas simulações Monte Carlo para elipsóides duros em solução.

Do comportamento da função de correlação $g(r)$ pudemos concluir que líquidos compostos de elipsóides prolatos e oblatos possuíam comportamentos termodinâmicos diferentes, o que decorre das claras diferenças entre a forma das funções.

A aproximação de desacoplamento foi testada para os fatores de estrutura calculados e para os fatores de estrutura de esfera rígida. Para os $S(q)$ calculados pudemos concluir que a aproximação é válida para todo o intervalo de anisotropias e concentrações simulado, indicando que as funções numéricas $S(q)$ (e conseqüentemente $g(r)$) podem ser utilizadas para descrever dados experimentais (supondo-se interação tipo esfera rígida entre as partículas) em termos da aproximação de desacoplamento sem perda de precisão. Além disso, para o uso de $S_{HS}(q)$ pudemos concluir que, ao menos a intensidade de espalhamento, pode ser satisfatoriamente descrita em termos deste fator de estrutura “errado” para o intervalo de anisotropias e fração de volume indicados nos resultados. Este é um resultado interessante, justificando sua utilização como primeira aproximação na descrição de dados experimentais. Por fim, obtivemos uma relação empírica entre os valores de R_{HS} e ϵ que permite obter a anisotropia em função do valor do raio de esfera rígida equivalente obtido em um ajuste.

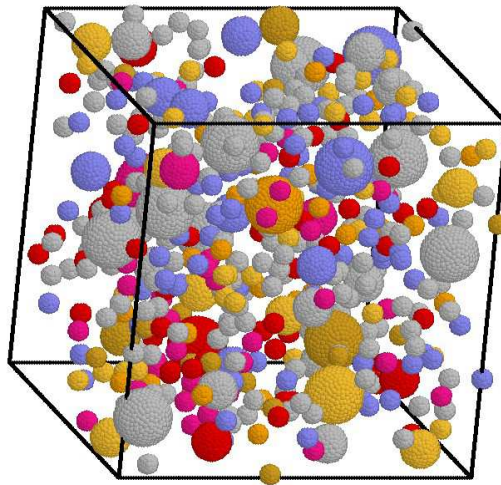
1.4. Exemplo de Aplicação – Estudo de partículas de ferro em matriz de silício

As simulações apresentadas anteriormente podem ser facilmente aplicadas a sistemas densos polidispersos de esferas. Para isso toma-se a equação de cálculo de intensidade mostrada anteriormente, utilizando agora a amplitude de espalhamento $F_k(\mathbf{q}, R_k)$ correspondente a uma com raio R_k que pode variar entre os componentes do sistema.

$$I(\vec{q}) = \left\{ \left[\sum_{k=1}^N F_k(\vec{q}_p, R_k) \sin(\vec{q}_p \cdot \vec{R}_k) \right]^2 + \left[\sum_{k=1}^N F_k(\vec{q}_p, R_k) \cos(\vec{q}_p \cdot \vec{R}_k) \right]^2 \right\}_{\vec{q}_p \in [100], [110], [111]} \quad (171)$$

sendo as outras funções calculadas da mesma forma que no caso anterior. Como mostrado na introdução teórica desta tese, para sistemas polidispersos trabalhamos com distribuições de tamanhos de partículas, isto é, um histograma do número de partículas para um dado raio. Assumindo-se uma forma para as partículas (em nosso caso, esferas), o objetivo deste tipo de cálculo passa a ser obter-se esta distribuição de tamanhos para um conjunto de dados experimentais.

Utilizando cálculos de Monte Carlo pode-se simular diretamente para um sistema proposto as intensidades de espalhamento e função de correlação. Como exemplo, imaginemos um caso onde temos um sistema composto por partículas esféricas sendo que 90% destas partículas possuem raio $R=1$, 7% $R=2$ e 3% $R=3$. Realizando-se simulações de Monte Carlo para este caso temos:



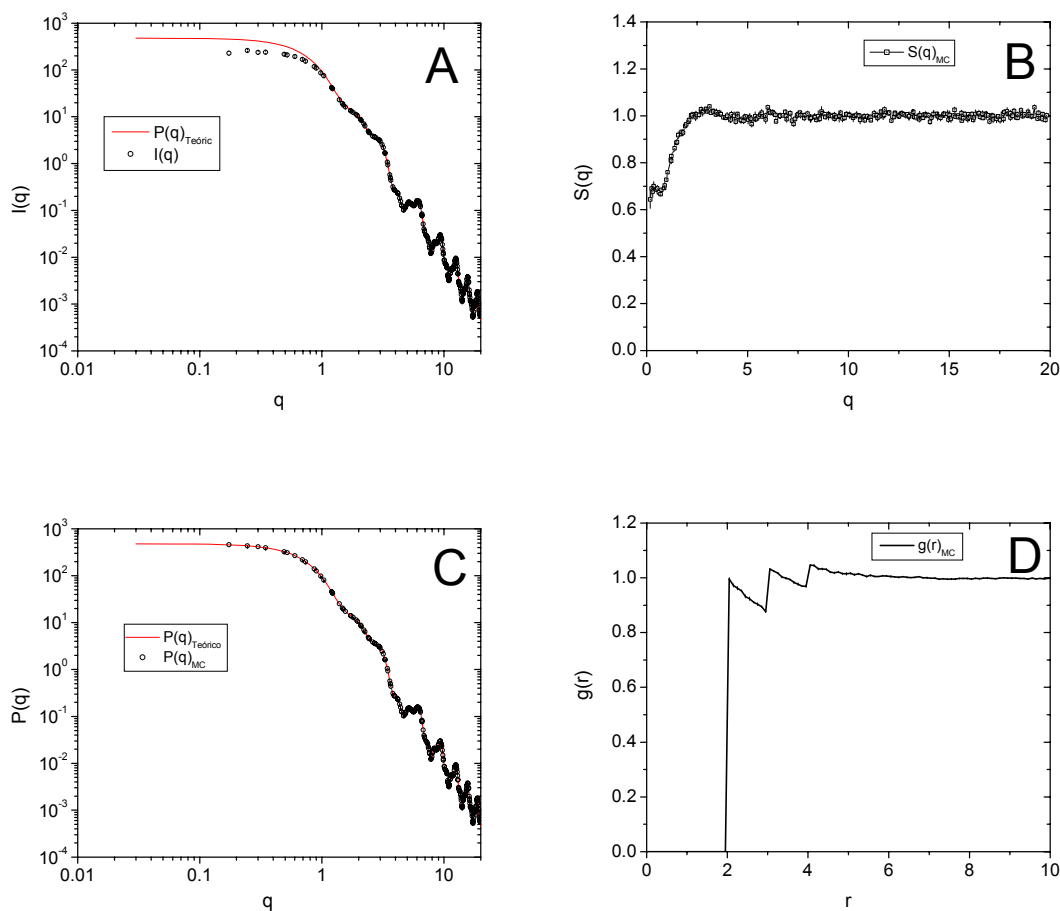


Figura 111 – Simulação de Monte Carlo para sistema polidisperso de esferas. *Topo:* Representação do sistema em um dos passos de cálculo. *Gráficos:* A) Intensidade de espalhamento calculada (pontos) e fator de forma teórico para o sistema (linha contínua); B) Fator de estrutura calculado; C) Fator de forma simulado (pontos) e teórico (linha contínua); D) Função de correlação do sistema.

Vemos algumas características interessantes neste cálculo. Do fator de forma teórico (Figura 111C) vemos que a média no *ensemble* é corretamente feita, pois temos uma perfeita superposição entre o fator de forma simulado e o teórico. O fator de estrutura apresenta oscilações atenuadas em decorrência de termos um sistema polidisperso. A função de correlação apresenta uma forma bastante peculiar. Temos variações abruptas para os valores de $r=2$, $r=3$ e $r=4$. Isso corresponde às distâncias mínimas entre a esfera de raio $R=1$ e as esferas de raio $R=1$, $R=2$ e $R=3$ respectivamente. Como o número de esferas com raio $R=1$ é muito superior ao número de esferas com raios $R=2$ e $R=3$, apenas as variações

com relação à esta esfera são vistas. As variações que se esperaria em $r=5$ e $r=6$ não são vistas nesta simulação devido ao à pequena fração de partículas no sistema.

Em sistemas reais não temos, em geral, valores discretos para os raios, mas sim distribuições de valores. Desta maneira, uma das rotinas do programa compõe-se de uma função que gera (ou lê de um arquivo externo) distribuições de tamanhos de partículas em função do raio. Em geral podemos assumir distribuições gaussianas com um determinado raio central r_0 e largura a meia altura σ . Isso foi utilizado no ajuste de dados experimentais de espalhamento a baixos ângulos de filmes de fitas metálicas de silício em que se formam partículas esféricas. O processo de produção destes filmes é descrito em diversos artigos (Socolovsky e Oliveira et al., 2005). Abaixo mostro uma figura de microscopia eletrônica de transmissão destes filmes, onde podemos identificar os domínios de Fe.

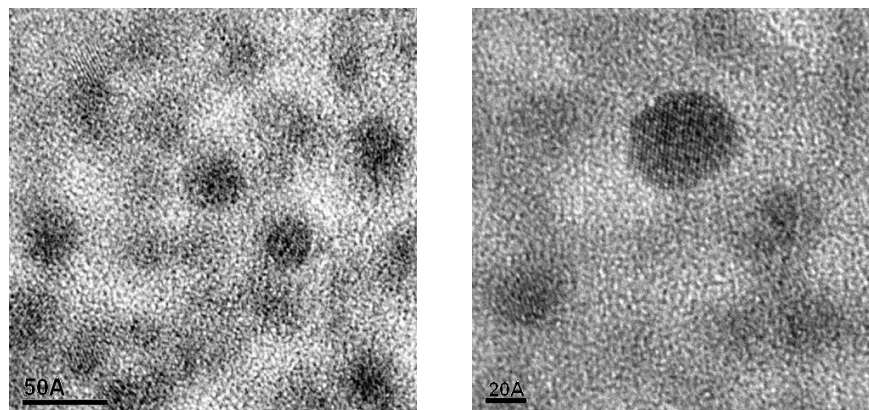
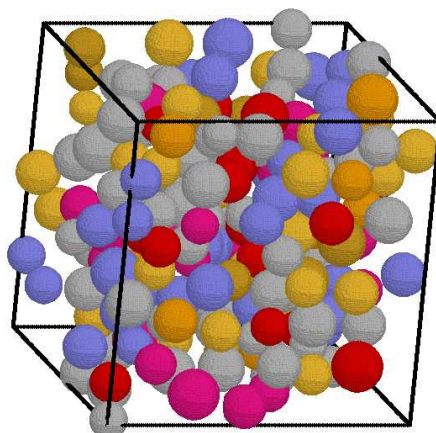


Figura 112 – Foto de microscopia eletrônico para o sistema em estudo. Tirado de Socolovsky e Oliveira et al, 2005.

Das simulações de MC, os parâmetros que fornecem os melhores ajustes dos dados são mostrados abaixo:



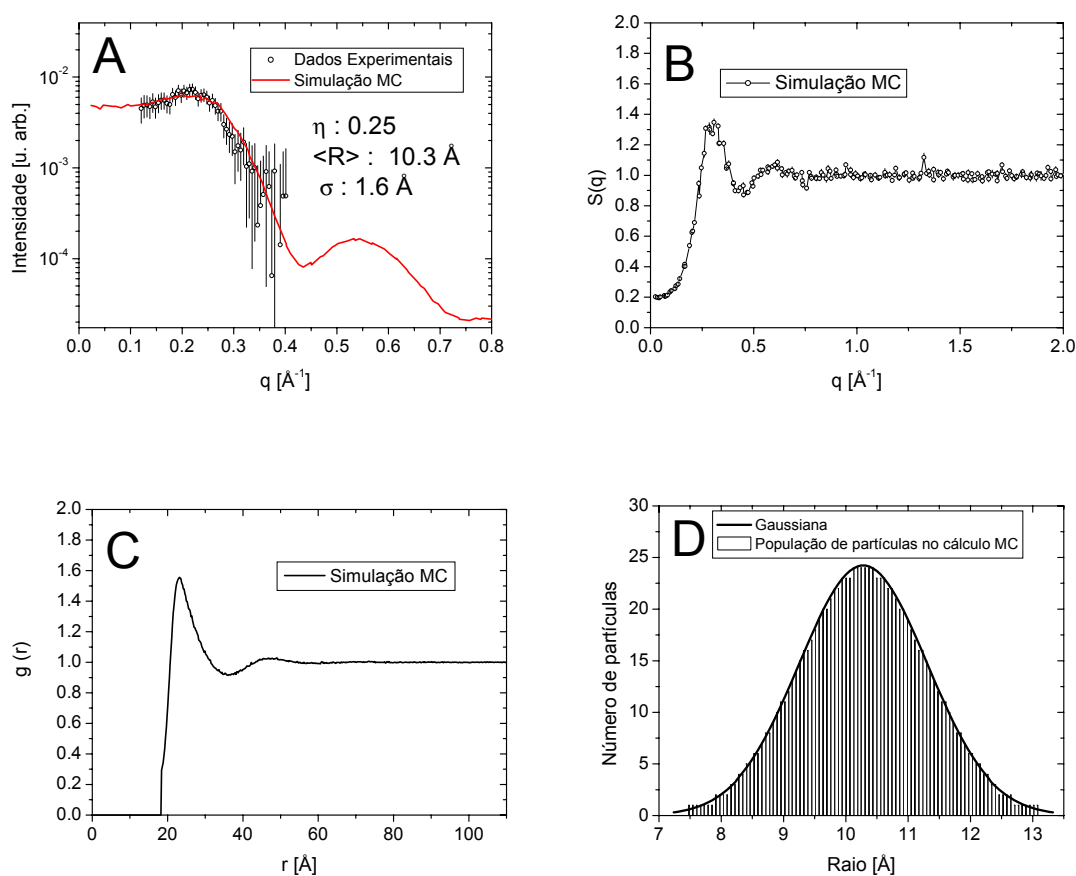


Figura 113 – Resultados da simulação MC para o sistema estudado. *Topo: Representação do sistema em um dos passos MC. Gráficos: A) Dados experimentais (pontos) e simulação de Monte Carlo (linha contínua) para o sistema; B) fator de estrutura do sistema; C) função de correlação do sistema; D) Distribuição gaussiana de partículas obtida do ajuste dos dados.*

Obtemos uma fração de volume de 25% ($\eta : 0.25$), raio médio das partículas $R = 10.3 \text{ \AA}$ e largura da distribuição $\sigma : 1.6 \text{ \AA}$. Sendo assim este método poder ser utilizado para a descrição de dados experimentais.

CAPÍTULO V

CONSIDERAÇÕES

FINAIS E

BIBLIOGRAFIA

1. Conclusões e Perspectivas

Nesta tese realizamos uma análise detalhada de vários aspectos teóricos e experimentais da técnica de espalhamento de raios X a baixos ângulos assim como de suas aplicações a sistemas diluídos (proteínas em solução) e sistemas densos. Detalhes teóricos e experimentais foram mostrados, sendo complementados por várias simulações que mostraram as limitações da técnica e exploraram novas possibilidades e protocolos para estender os limites de aplicação. Programas de cálculo de parâmetros hidrodinâmicos para estruturas cristalográficas foram adaptados para diversos tipos de modelos *ab-initio* sendo apresentadas as melhores metodologias para o cálculo e promediação dos valores obtidos. Foi proposta uma nova metodologia para cálculo dos parâmetros hidrodinâmicos para modelos *ab-initio* obtidos a partir de dados de espalhamento, usando modelagem por grupos de resíduos, aplicável principalmente para proteínas com elevado peso molecular.

Dos resultados experimentais informações estruturais inéditas sobre as proteínas estudadas foram obtidas, permitindo a correlação dos dados estruturais com aspectos biológicos e funcionais. Além disso, da interpretação dos dados, diversos protocolos de análise foram desenvolvidos e apresentados, permitindo a sua utilização no estudo de outros tipos de proteínas. Dado que a utilização de programas de modelagem deve estar atrelada à compreensão de seu funcionamento e limitações, tentou-se apresentar nesta tese diversos casos de resultados atípicos e peculiaridades dos programas de simulação.

No estudo de eritrocruorinas, a renormalização dos dados à escala absoluta foi fundamental para quantificar o número de subunidades da proteína bem como verificar a monodispersidade da amostra. Foi possível demonstrar que a hemoglobina de *B. glabrata* possui uma simetria tetramérica, com forma prolata, podendo ser imaginada como uma composição de 8 subunidades. Já para a hemoglobina de *G. paulistus* diversas metodologias de cálculo foram descritas, mostrando que a conformação desta proteína é muito semelhante à de *L. terrestris*, uma proteína da mesma família, porém de outra espécie. Para esta proteína a utilização de medidas de espalhamento a altos valores do vetor de espalhamento em conjunção com modelagens *chain model* para grupos de resíduos possibilitou a diminuição da ambigüidade dos modelos. Obteve-se também uma maior aproximação, quando comparado com dados experimentais, dos valores dos parâmetros

hidrodinâmicos calculados por esta metodologia do que os obtidos por outros métodos *ab initio*.

Dos resultados experimentais da proteína básica de mielina (MBP), vários pontos interessantes foram abordados. Os resultados de espalhamento de raios X a baixos ângulos permitiram, pela primeira vez, obter-se diferenças estruturais e conformacionais entre a proteína extraída com lipídeos (LB-MBP) e a proteína extraída sem lipídeos (LF-MBP). Com isso pode-se demonstrar que a estrutura nativa e funcional desta proteína somente é alcançada se a mesma estiver ligada a lipídeos. Da comparação entre modelos *ab initio* calculados para a LB-MBP e modelos teóricos calculados por métodos de dinâmica molecular, foi possível comprovar a conformação em “U” desta proteína bem como demonstrar a existência de uma camada lipídica no entorno da molécula.

Do estudo apresentado para interações entre proteínas e polímeros diversos resultados interessantes puderam ser obtidos. Os dados de espalhamento revelaram que, diferentemente do proposto em trabalhos anteriores, a lisozima e os polímeros analisados não se ligam de forma específica. Efeitos de interação entre as moléculas foram constatados devido a influências dos polímeros nos processos de agregação da lisozima bem como diferenças nas dimensões e perfil de densidade eletrônica das micelas formadas por polímeros blocos na presença de proteína em comparação com as formadas sem a presença de proteínas. Na análise dos resultados, modelos teóricos foram propostos os quais permitiram a correta interpretação dos resultados.

O longo estudo efetuado para a proteína precursora amilóide (APP) em duas isoformas permitiu a obtenção de resultados bastante interessantes. O estado de oligomerização, conformação tridimensional, estabilidade, enovelamento, etapas de denaturação e organização de domínios foram os principais aspectos estudados utilizando dados de espalhamento. Mostramos que em seu estado nativo as proteínas encontram-se em um estado monomérico e, ao final, a “montagem” da estrutura tridimensional da proteína através de suas parcelas conhecidas a nível atômico guiado pela modelagem *ab initio* de SAXS forneceu resultados inéditos sobre a conformação e funcionamento protéico. Diversas inferências puderam ser feitas sobre a atuação desta proteína nos processos biológicos, sendo suportadas por outros resultados obtidos na literatura.

Por fim, o estudo de sistemas densos através de simulações de Monte Carlo para elipsóides de revolução permitiu a descrição e compreensão de diversos comportamentos destes sistemas. Visando inicialmente corrigir erros apresentados em trabalhos na literatura, foi desenvolvido um programa que permitiu o cálculo de funções de espalhamento para sistemas de elipsóides de revolução com as mais variadas anisotropias e concentrações. Diversos comportamentos puderam ser analisados bem como um estudo quantitativo dos erros que se cometem utilizando fatores de estrutura de esferas para análises de dados experimentais de sistemas de partículas não esféricas. Aplicou-se este método ao estudo de nanopartículas de ferro em matriz de silício, obtendo-se parâmetros estruturais do sistema, como tamanho médio das partículas, dispersão de tamanhos, fator de estrutura e função de correlação.

Em face ao que foi apresentado, pode-se mostrar a grande versatilidade e abrangência da técnica de espalhamento de raios X a baixos ângulos no estudo de diversos tipos de sistemas. Isto, associado ao desenvolvimento de métodos teóricos e computacionais bem como a avanços de técnicas experimentais, favorecerão ainda mais a utilização desta técnica como uma das mais importantes para análise dos mais variados tipos de problemas.

2. Referências e Bibliografia

1. Abramowitz M and Stegun IA,(1972) "Handbook of mathematical functions: with formulas, graphs, and mathematical tables", Dover, New York.
2. Abuchowski AT Palczuk NC and. Davis FF (1977). "Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol". J. Biol. Chem. 252 3578.
3. Allen MP and Tildesley DJ (1987). "Computer Simulations of liquids". Clarendon, Oxford;
4. Almeida AP, Neves AGA (1974). "The hemoglobin of *Biomphalaria glabrata*: Chemical composition and some physicochemical properties". Biochimica et Biophysica Acta (BBA) – Prot. Struc., 371 (1), 140-146.
5. Almeida NL, Oliveira CLP, Torriani IL and Loh W (2004). "Calorimetric and structural investigation of the interaction of lysozyme and bovine serum albumin with poly(ethylene oxide) and its copolymers". Coll. And Surf. B: Biointerfaces, 38, 67-76.
6. Altschul SF, Madden TL, Schäffer AAJ, Zhang Z, Miller W, and Gapped LDJ (1997). "BLAST and PSI-BLAST: a new generation of protein database search programs". Nucleic Acids Res. 25:3389–3402.
7. Arfken GB, Weber HJ (1995). "Título Mathematical methods for physicists". Imprensa San Diego : Academic Press, 1995.
8. Arndt MHL, Santoro MM (1998), "Structure of the extracellular hemoglobin of *Biomphalaria glabrata*". Comp Biochem. Physiol. B, 119, 667–675.
9. Arndt, MHL, Régis WCB, Santoro MM (2000). "International Conference on Dioxxygen Binding Proteins", Roscoff, France, 2000.
10. Arndt MHL, Oliveira CLP, Régis WCB, Torriani IL and Santoro MM (2003), "Small Angle X-Ray Scattering of the Hemoglobin from *Biomphalaria glabrata*". Biopolymers, 69, 470-479.
11. Atkins PW (1988). "Physical Chemistry". Oxford University Press.
12. Azegami S, Tsuboy A, Izumi T, Hirata M, Dublin PL, Wang B and Kokufuta E (1999). "Formation of an Intrapolymer Complex from Human Serum Albumin and Poly(ethylene glycol)". Langmuir 15, 940-947.
13. Barnham KJ, McKinsty WJ, Multhaup G, Galatis D, Morton CJ, Curtain CC, Williamson NA, White AR, Hinds MG, Norton RS, Beyreuther K, Masters CL, Parker MW and Cappai R (2003). "Structure of the Alzheimer's disease amyloid precursor protein copper binding domain - A regulator of neuronal copper homeostasis". J. Biol. Chem., 278, 17401-17407.
14. Bergmann A, Fritz G and Glatter O (2000). "Solving the generalized indirect Fourier transform (GIFT) by Boltzmann simplex simulated annealing (BSSA)", J. Appl. Cryst. 33, 1212-1216;
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000). "The Protein Data Bank". Nucleic Acids Res. 28:235–242.
16. Bonafe CFS, Villas-Boas M, Suarez MC and Silva JL (1991). "Reassembly of a Large Multisubunit Protein Promoted by Nonprotein Factors: Effects of Calcium and Glycerol on the association of extracellular hemoglobin", J. Biol. Chem., 266 (20), 13210-13216.
17. Botelho MG, Gralle M, Oliveira CLP, Torriani I and Ferreira ST (2003). "Folding and stability of the extracellular domain of the human amyloid precursor protein". J. Biol. Chem., 278, 34259-34267.
18. Bracewell R (1965). "Fourier Transforms and Its Applications". McGraw-Hill, New York.
19. Bugge J, Weber RE (1999). "Oxygen binding and its allosteric control in hemoglobin of the pulmonate snail, *Biomphalaria glabrata*". Am J Physiol Regul. Integ. Comp Physiol, 276: R347– R356.
20. Butkov E (1968). "Mathematical physics". Imprensa Reading, Mass : Addison-Wesley.
21. Caille I, Allinquant B, Dupont E, Bouillot C, Langer A, Muller U and Prochiantz A (2004). "Soluble form of amyloid precursor protein regulates proliferation of progenitors in the adult subventricular zone". Development, 131, 2173-2181.

22. Cavalcanti LP, Torriani IL, Plivelic TS, Oliveira CLP, Kellermann G and 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. Inst.*, 75 (11), 4541-4546.
23. Chacon P, Morán F, Días JF, Pantos E and Andreu JM, (1998). "Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm". *Bioph. J.*, 74, 2760-2775.
24. Chacon P, Diaz JF, Moran F, Andreu JM (2000). "Reconstruction of protein form with X-ray solution scattering and a genetic algorithm". *J Mol Biol.*, 299:1289-1302.
25. Chung MCM and Ellerton HD (1979). "The physico-chemical and functional properties of extracellular respiratory haemoglobins and chlorocruorins". *Prog. Biophys. Molec. Biol.* 35, 33-102.
26. Ciccariello S and Sobry R (1999). "Small-angle scattering intensity behaviours of cylindrical, spherical and planar lamellae". *J. Appl. Cryst.* (1999). 32, 892-901.
27. Cohen-Tannoudji C, Diu B, Laloe F (1977). "Quantum Mechanics". New York : J. Wiley ; Paris : Hermann.
28. Cook JM (1957). "Rational formulae for the production of a spherically symmetric probability distribution". *Math. Tables Other Aids Comp.* 11, 81-82
29. Costa MCP, Bonafe CFS, Meirelles NC and Galembeck F, (1988). "Sedimentation Coefficient and Minimum molecular weight of extracellular hemoglobin of *Glossoscolex paulistus* (oligochaeta)". *Brazilian J. Med. Biol. Res.*, 21, 115-118.
30. Cullity BD (1978) "Elements of X-Ray Diffraction", Addison-Welsey Publ. Comp, Reading.
31. Czech C, Tremp G and Pradier L (2000). "Presenilins and Alzheimer's disease: Biological functions and pathogenic mechanisms". *Prog. Neurobiol.* 60:363-384.
32. De Felice FG and Ferreira ST (2002). "β-Amyloid Production, Aggregation and Clearance as Targets for Therapy in Alzheimer's Disease". *Cel. Mol. Neur.*, 22, 545-563.
33. De Strooper B and Annaert W (2000). "Proteolytic processing and cell biological functions of the amyloid precursor protein". *J. Cell Sci.* 113: 1857-1870.
34. Debye P (1915). "Zerstreuung von röntgenstrahlen. Scattering from non-crystalline substances". *Ann. Phys.* 46, 809-823;
35. Decker H and Sterner R (1990). "Hierarchien in der Struktur und Funktion von sauerstoffbindenden Proteinen". *Naturwissenschaften*, 77, 561-568.
36. Deibler GE, Martenson RE and Kies MW (1972). "Large-scale preparation of myelin basic protein from central nervous tissue of several mammalian species". *Prep. Biochem.* 2:139-165.
37. Deibler GE, Boyd LF and Kies MW (1984). "Proteolytic activity associated with purified myelin basic protein. In *Experimental Allergic Encephalomyelitis: A Useful Model for Multiple Sclerosis*". E. C. Alvord Jr., M. W. Kies, and A. J. Suckling, editors. Liss, New York. 249-256.
38. Dulubova I, Ho A, Huryeva I, Sudhof TC and Rizo J (2004). "Three-dimensional structure of an independently folded extracellular domain of human amyloid-beta precursor protein". *Biochemistry*, 43, 9583-9588.
39. Durkin JT, Murthy S, Husten EJ, Trusko SP, Savage MJ, Rotella DP, Greenberg BD and Siman R (1999). "Rank-order of potencies for inhibition of the secretion of Abeta40 and Abeta42 suggests that both are generated by a single gamma-secretase". *J. Biol. Chem.* 274: 20499-20504.
40. Epand RM, Moscarello MA, Zierenberg B, and Vail WJ (1974). "The folded conformation of the encephalogenic protein of the human brain". *Biochemistry*. 13:1264-1267.
41. Feigin LA, (1967) "The Absolute Intensity of Small Angle X-Ray Scattering in Apparatus Employing Slit Collimation, and the Calculation of the Molecular Weights of Macromolecules", *Sov. Phys. - Crystall.*, 12-2, 228;
42. Feigin LA, Svergun DI, (1987) "Structure Analysis by small-angle X-ray and neutron scattering", Plenum Press, New York;
43. Fiser A and Šali A (2003). "MODELLER: generation and refinement of homology-based protein structure models". *Methods in Enzymology*. Ed: Carter, C.W. and Sweet, R.M.. Academic Press, San Diego, 374, 463-493, 2003.

44. Frenkel D, Mulder BM (1985). "The Hard Ellipsoid-of-Revolution Fluid .1. Monte-Carlo Simulations", Mol. Phys. 55 (5): 1171-1192.
45. Frenkel D, Mulder BM, Mctague JP (1985). "Phase-Diagram Of Hard Ellipsoids Of Revolution". Mol. Cryst. And Liq. Cryst. 123 (1-4): 119-128.
46. Fushitani K and Riggs AF (1991). "The extracellular hemoglobin of earthworm *Lumbricus terrestris* - oxygenation properties of isolated chains, trimer and a reassociated product". J. Biol. Chem., 266 (16), 10275-10281.
47. Gabriel A (1982). "X-ray position sensitive detectors". Nuc.r Inst and Meth. in Phys. Research , 201 (1), 221-222;
48. García de La Torre J, Huertas ML and Carrasco B (2000). "Calculation of Hydrodynamic Properties of Globular Proteins from their Atomic-Level Structure". Biophys. J., 78, 719-730.
49. Glatter O, (1972). "X-Ray Small Angle Scattering of Molecules Composed of Subunits" ,Acta Phys. Austr., 36, 307-315.
50. Glatter O (1977). "New Method For Evaluation Of Small-Angle Scattering Data". J. Appl. Cryst. 10, 415-421.
51. Glatter O (1981). "Convolution square root of band-limited symmetrical functions and its applications to small-angle scattering data". J. Appl. Cryst. 14:101-108.
52. Glatter O, Kratky O (1982) "Small angle X-ray scattering", Academic Press, Londres;
53. Glatter O, Hainisch B (1984). "Improvements In Real-Space Deconvolution Of Small-Angle Scattering Data". J. Appl. Cryst., 17: 435-441.
54. Glatter O (1988). "Comparison Of 2 Different Methods For Direct Structure-Analysis From Small-Angle Scattering Data". J. Appl. Cryst., 21: 886-890.
55. Glatter O (1991). "Scattering studies on colloids of biological interest (Amphiphilic systems)". Progr. Colloid Polym. Sci 84, 46-54.
56. Glatter O, Scherf G, Schillen K, Brown W (1994) "Characterization Of A Poly(Ethylene Oxide) Poly(Propylene Oxide) Triblock Copolymer (Eo(27)-Po39-Eo(27)) In Aqueous-Solution".Macromol., 27 (21): 6046-6054 Oct 10 1994.
57. Glatter O (2002). "Fourier Transforms and Deconvolution (Chapter 5)". In Neutrons, X-Rays and Light, P. Lindner and Th. Zemb (Editors). Elsevier Science B.V.
58. Goate A, Chartier-Harlin,MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, James L and et al. (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature, 349, 704-706.
59. Gralle M, Botelho MM, Oliveira CLP, Torriani I and Ferreira ST (2002). "Solution studies and structural model of the extracellular domain of the human amyloid precursor protein". Biophys. J., 83, 3513-3524.
60. Gralle M, Oliveira CLP, Torriani IL and Ferreira ST. "Conformation of the extracellular domain of human amyloid precursor protein from solution X-ray scattering", *in preparation*.
61. Guinier A and Fournet G (1955). In "Small-angle Scattering of X-rays", Wiley, New York.
62. Guinier A (1963). "X-Ray Diffraction In Crystals, Imperfect Crystals and Amorphous Bodies", Dover, New York.
63. Haas H, Torrielli M, Steitz R, Cavatorta P, Sorbi R, Riccio P and Gliozzi A. (1998). "Myelin model membranes on solid substrate". Thin Sol. Films. 327-329:627-631.
64. Haas H, Oliveira CLP, Torriani IL, Polverini E., Fasano A, Carlone G., Cavatorta P and Riccio P (2004) "Small Angle X-Ray Scattering from Lipid-Bound Myelin Basic Protein in Solution", Bioph. J., Vol. 86, 1-6.
65. Hansen JP and McDonald IR (1990). "Theory of Simple Liquids". London: Academic Press.
66. Harder P, Grunze M, R. Dahint R, Whitesides GM, Laibinis PE (1998). "Molecular Conformation in Oligo(ethylene glycol)-Terminated Self-Assembled Monolayers on Gold and Silver Surfaces Determines Their Ability To Resist Protein Adsorption". J. Phys. Chem. B 102 (1998) 426.
67. Hendricks RW, (1972) "A Unified Theory of Absolute Intensity Measurements in Small-Angle X-ray Scattering", J. Appl. Cryst. 5, 315-324;
68. Herskovits TT, Hamilton MG (1994). "The molecular weight and subunit organization of *Helisoma trivolvis* (Say) hemoglobin: light-scattering and scanning transmission electron microscopic studies". Comp Biochem Physiol B 1994, 107(3), 433-441.

69. Hynes TR, Randal M, Kennedy LA, Eigenbrot C and Kossiakoff AA (1990). "X-ray crystal structure of the protease inhibitor domain of Alzheimer's amyloid beta-protein precursor". *Biochemistry*, 29, 10018-10022.
70. Ilan E, Hammel I, David MM and Daniel E (1986). "Erythrocrucorin from the aquatic snail *Helisoma trivolvis*. Quaternary structure and arrangement of subunits". *Biochemistry*, 25(21), 6551-6554.
71. Ingber L (1993). "Simulated Annealing: Practice versus Theory", *Math. Comp. Modelling*, 18(11), 29-57.
72. Jackson JDJ (1963). "Classical Electrodynamics". John Wiley & Sons.
73. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992), "Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques", *Proc. Natl. Acad. Sci. USA*, 89, 2195-2199.
74. Kellermann, G., Vicentin, F., Tamura, E., Rocha, M., Tolentino, H., Barbosa, A., Craievich, A. and Torriani, I.L. (1997) "The small angle X-ray scattering beamline of the brazilian synchrotron light laboratory", *J. Appl. Cryst.* 30: 880-883.
75. Knoll GF (1979). "Radiation detection and measurement". Imprenta New York: J. Wiley, c1979.
76. Konarev PV, Petoukhov MV and Svergun DI (2001). "MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data". *J. Appl. Crystallogr.* 34, 527-532.
77. Koradi R, Billeter M and Wüthrich K (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, 14, 51-55.
78. Kotlarchyk M, Chen SH (1983). Analysis of small angle neutron scattering spectra from polydisperse interacting colloids. *J. Chem. Phys* 79 (5), 2461-2469;
79. Kozin MB and Svergun DI (2001). "Automated matching of high- and low-resolution structural models". *J. Appl. Cryst.* 34, 33-41.
80. Krebs A, Durchschlag H and Zipper P (2004). "Small Angle X-Ray Scattering Studies and Modeling of *Eudistylia vancoverii* Chlorocruorin and *Macrobdella decora* Hemoglobin". *Bioph. J.*, 87, 1173-1185.
81. Krigbaum WR and Hsu TS (1975). "Molecular conformation of bovine A1 basic protein, a coiling macromolecule in aqueous solution". *Biochemistry*. 14: 2542-2546.
82. Kroenke CD, Ziemnicka-Kotula D, Xu J, Kotula L and Palmer AG (1997) "Solution conformations of a peptide containing the cytoplasmic domain sequence of the beta amyloid precursor protein". *Biochemistry*, 36, 8145-8152.
83. Lammich S, Kojro E, Postina R, Gilbert S, Pfeiffer R, Jasionowski M, Haass C and Fahrenholz F (1999). "Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease". *Proc. Natl. Acad. Sci. USA*, 96, 3922-3927.
84. Landau, L; Lifshitz, E, (1951) "The Classical Theory of Fields", Pergamon Press.
85. Lehninger AL (1990). "Principios de bioquímica". Trad. por Lodi WR, Simões AA. Imprenta São Paulo : Sarvier, 1990.
86. Li CH, Ma XH, Chen WZ and Wang CX (2003). "A protein-protein docking algorithm dependent on the type of complexes". *Prot. Eng.*, 16(4), 265-269.
87. Lide DR (1997) "Handbook of Chemistry and Physics", Boca Raton, New York.
88. Lindner P and Zemb T (Ed.) (2002). "In Neutrons, X-Rays and Light", Elsevier Science B.V.
89. Marquez JA, Smith CIE, Petoukhov MV, Lo Surdo P, Mattsson PT, Knekt M, Westlund A, Scheffzek K, Saraste M and Svergun DI (2003). "Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering". *EMBO J.*, 22, 4616-4624.
90. Marsaglia G (1972). "Choosing A Point From Surface Of A Sphere". *Annal. Math. Stat.* 43 (2): 645-646.
91. Mazzanti B, Vergelli M, Riccio P, Martin R, McFarland HF, Liuzzi MG, Amaducci L and Massacesi L (1998). "T-cell response to myelin basic protein and lipid-bound myelin basic protein in patients with Multiple Sclerosis and health donors". *J. Neuroimmunol.* 82:96-100.

92. McGuffin LJ, Bryson K, Jones DT (2000). "The PSIPRED protein structure prediction server". *Bioinformatics*, 16, 404-405.
93. McGuffin LJ, Jones DT (2003). "Improvement of the GenTHREADER method for genomic fold recognition". *Bioinformatics*, 19, 874-881.
94. Merzbacher E (1970). "Quantum Mechanics". Second Edition, John Wiley & Sons, Inc.
95. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, (1953). "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.*, 21 (5), 1087-1092.
96. Moore PB (1980). "Small-Angle Scattering. Information Content and Error Analysis, *J. Appl. Cryst.*, 13, 168-175.
97. Moscarello MA (1996). "Evolving biological concepts and therapeutic approaches". In *Cell Biology and Pathology of Myelin*. R. M. Devon, R. Doucette, B. H. J. Juurlink, A. J. Nazarali, D. J. Schreyer, and V. M. K. Verge, editors. Plenum Publishing, New York.
98. Mulder BM, Frenkel D (1985). "The Hard Ellipsoid-Of-Revolution Fluid .2. The Y-Expansion Equation Of State". *Mol. Phys.*, 55 (5): 1193-1215.
99. Multhaup G, Beyreuther K, Masters CL and Parker MW (1999). "Crystal structure of the N-terminal, growth factor-like domain of Alzheimer amyloid precursor protein". *Nat. Struct. Biol.*, 6, 327-331.
100. Nolan SL, Ronald JP, Cotts PM and Dungan SR. (1997), "Light Scattering Study on the Effect of Polymer Composition on the Structural Properties of PEO-PPO-PEO Micelles". *J. Coll. Interf. Sci.*, 191, 291-302;
101. Oliveira CLP (2001), "Simulação do Espalhamento de Raios X por Macromoléculas em Solução Através da Construção de Modelos Tridimensionais de Baixa Resolução", Tese de Mestrado – IFGW, TMO14s;
102. Oliveira CLP (2003). "TRAT1D - Computer Program for SAXS Data Treatment", LNLS Technical Manual MT01/2003.
103. Oliveira CLP (2004). "TRAT2D - Computer Program for Bidimensional SAXS Data Treatment", LNLS Technical Manual MT02/2004.
104. Oliveira CLP, Bonafe CFS, Torriani, IL. "SAXS of Glossoscolex Paulistus hemoglobin". *In preparation*.
105. Oliveira CLP, Torriani IL and Pedersen JS. "Monte Carlo Simulation Studies of Small-Angle Scattering from Hard Ellipsoids". *In preparation*.
106. Orthaber D, Bergmann A and Glatter O (2000). "SAXS experiments on absolute scale with Kratky systems using water as a secondary standard". *J. Appl. Cryst.*, 33:218–225.
107. Patel S, Schmidt PW (1971). "Small-angle X-ray Scattering Determination of Electron Density of the Particles in a Colloidal Silica Suspension". *J. App. Cryst.*, 4, 50.
108. Pathria RK (1995). "Statistical Mechanics". Oxford : Butterworth-Heinemann, 1996.
109. Pedersen JS (2002). "Monte Carlo Simulation Techniques Applied in the Analysis of Small-Angle Scattering Data from Colloids and Polymer Systems (Chapter 15)". In *Neutrons, X-Rays and Light*, P. Lindner and Th. Zemb (Editors). Elsevier Science B.V.
110. Percus JK, Yevick G (1958). "Analysis of Classical Statistical Mechanics by Means of Collective Coordinates". *Phys. Rev.* 110, 1-13
111. Perkins JP (1988). "X-ray and neutron solution scattering". In *Modern Physical Methods in Biochemistry*, Part B. A. Neuberger and L. L. M. van Deenen, editors. Elsevier, Amsterdam, The Netherlands. 134–265.
112. Perran JW, Wertheim MS (1985), "Statistical mechanics of hard ellipsoids. I. Overlap algorithm and the contact function". *J. Comp. Phys.* 58 (3), 409-416;
113. Petoukhov MV, Eady NAJ, Brown KA and DI Svergun DI (2002). "Addition of missing loops and domains to protein models using X-ray solution scattering". *Biophys J.* 2002 83(6) 3113-3125.
114. Petoukhov MV and Svergun, DI (2003). "New methods for domain structure determination of proteins from solution scattering data". *J. Appl. Cryst.* 36, 540-544.
115. Pilz I, Schwarz E and Vinogradov SN (1980), "Small-Angle X-Ray Studies of Lumbricus terrestris haemoglobin", *Int. J. Biol. Macromol.*, 2, 279-283;
116. Pilz I, Schwarz E, Tsfadia Y, Daniel E (1988). "Small-angle X-ray study on the quaternary structure of erythrocrucorin from *Helisoma trivolvis*". *E. Int J Biol Macromol* 1988, 10, 353–355.

117. Polverini E, Fasano A, Zito F, Riccio P and Cavatorta P (1999). "Conformation of bovine myelin basic protein purified with bound lipids". *Eur. Biophys. J.* 28:351–355.
118. Press WH, Flannery BP, Teukolsky SA and Vetterling WT, (1992). "Numerical Recipes in Pascal – The Art of Scientific Computing", Cambridge University Press.
119. Quast T, Wehner S, Kirfel G, Jaeger K, De Luca M and Herzog V (2003). "sAPP as a regulator of dendrite motility and melanin release in epidermal melanocytes and melanoma cells". *FASEB J.*, 17.
120. Riccio P, Rosenbusch JP and Quagliarello E (1984). "A new procedure to isolate brain myelin basic protein in a lipid-bound form". *FEBS Lett.* 177:236–240.
121. Riccio P, Masotti L, Cavatorta P, De Santis A, Juretic D, Bobba A, Pasquali-Ronchetti I and Quagliarello E (1986). "Myelin basic protein ability to organize lipid bilayers: structural transitions in bilayers of lysophosphatidylcholine micelles". *Biochem. Biophys. Res. Commun.* 134:313–319.
122. Riccio P and Quagliarello E (1993). "Lipid-bound, native-like, myelin basic protein: a well-known protein in a new guise, or an unlikely story?". *J. Neurochem.* 61:787–788.
123. Riccio P, Bobba A, Romito E, Minetola M and Quagliarello E (1994). "A new detergent to purify CNS myelin basic protein isoforms in lipidbound form". *Neuroreport.* 24:689–692.
124. Riccio P, Fasano A, Borenshtein N, Bleve-Zacheo T and Kirschner DA (2000). "Multilamellar packing of myelin modeled by lipidbound MBP". *J. Neurosci. Res.* 59:513–521.
125. Ridsdale RA, Beniac DR, Tompkins TA, Moscarello MA and Harauz G (1997). "Three-dimensional structure of myelin basic protein. II.Molecular modeling and considerations of predicted structures inmultiple sclerosis". *J. Biol. Chem.* 272:4269–4275.
126. Rossjohn J, Cappai R, Feil SC, Henry A, McKinstry WJ, Galatis D, Hesse L, Multhaup G, Beyreuther K, Masters CL and Parker MW (1999). "Crystal structure of the N-terminal, growth factor-like domain of Alzheimer amyloid precursor protein". *Nat. Struct. Biol.*, 6, 327-331.
127. Royer WE Jr, Strand K, van Heel M and Hendrickson WA. (2000). "Structural hierarchy in erythrocrurin, the giant respiratory assemblage of annelids", *PNAS*, 97 (13), 7107-7111.
128. Russel TP, (1988) "Intercalibration of Small-Angle X-Ray and Newtron Scattering Data", 21, 629-638;
129. Sabo SL, Ikin AF, Buxbaum JD and Greengard P (2003). "The amyloid precursor protein and its regulatory protein, FE65, in growth cones and synapses in vitro and in vivo". *J. Neurosci.*, 23, 5407-5415.
130. Sakurai JJ (1994). "Modern Quantum Mechanics". Imprenta Reading: Addison-Wesley, 1994.
131. Šali A and Blundell TL (1993). "Comparative Protein Modelling by Satisfaction of Spatial Restraints". *J. Mol. Biol.* 234, 779-815.
132. Semenyuk V and Svergun DI (1991) "GNOM - A Program Package for Small-Angle Scattering Data-Processing", *J. Appl. Cryst.*, 24, 537-540;
133. Shaffer LB and Hendricks RW (1974), "Calibration of Polyethylene (Lupolen) as a Wavelength-Independent Absolute Intensity Standard", *J. Appl. Cryst.* 7, 159;
134. Shannon C and Weaver W (1949). "The Mathematical Theory of Communication". The University of Illinois Press: Urbana.
135. Silva JL, Villas-Boas M, Bonafe CF and Meirelles,NC (1989). "Anomalous pressure dissociation of large protein agregates. Lack of concentration dependence and irreversibly at extreme degrees of dissociation of extracellular hemoglobin". *J Biol Chem.*, 25, 264(27): 15863-8.
136. Sjöberg B and Mortensen K (1997). "Structure and thermodynamics of nonideal solutions of colloidal particles: Investigation of salt-free solutions of human serum albumin by using small-angle neutron scattering and Monte Carlo simulation" *Biophys. Chem.* 65, 75-83.
137. Sjöberg B, (1999), "Small-angle scattering from collections of interacting hard ellipsoids of revolution studied by Monte Carlo simulations and other methods of statistical thermodynamics", *J. Appl. Cryst.* 32, 917-923;
138. Socolovsky LM, Oliveira CLP, Denardin JC, Knobel M and Torriani IL (2004). "Giant Hall effect in granular Co-SiO₂ thin films modified by thermal treatment". *In preparation.*
139. Souza CF (1990). "Espalhamento de Raios X a Baixo Ângulo: Estudo de Hemoglobinas Extracelulares em Solução, Tese de Doutorado – IFGW, TDS089e.

140. Stein TD and Johnson JA (2002). "Lack of neurodegeneration in transgenic mice overexpressing mutant amyloid precursor protein is associated with increased levels of transthyretin and the activation of cell survival pathways". *J. Neurosci.*, 22, 7380-7388.
141. Stuhrman HB (1970). "Interpretation Of Small-Angle Scattering Functions Of Dilute Solutions And Gases - A Representation Of Structures Related To A One-Particle-Scattering Function". *Acta Cryst. A*, 26: 297-307.
142. Svaneborg C and Pedersen JS (2002). "Form Factors of Block Copolymer Micelles with Excluded- Interactions of the Corona Chains Determined by Monte Carlo Simulations". *Macromolecules* 35(3), 1028-1037.
143. Svergun DI (1991). "Mathematical methods in small-angle scattering data analysis". *J. Appl. Cryst.* 24, 485-492.
144. Svergun DI, Stuhrmann HB, (1991). "New Developments In Direct Shape Determination From Small-Angle Scattering .1. Theory And Model-Calculations". *Acta Cryst. A*, 47: 736-744.
145. Svergun DI (1994). "Solution scattering from biopolymers: advanced contrast variation data analysis". *Acta Cryst. A*, 50, 391-402.
146. Svergun DI, Barberato C and Koch MHJ (1995). "CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates". *J. Appl. Cryst.* 28, 768-773.
147. Svergun DI, Volkov VV, Kozin MB and Stuhrmann HB (1996). "New developments in direct shape determination from small-angle scattering. 2. Uniqueness". *Acta Cryst. A*, 52, 419-426.
148. Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB, Barberato C and Koch MHJ (1997). "Shape determination from solution scattering of biopolymers". *J. Appl. Cryst.* 30, 798-802.
149. Svergun DI (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing". *Biophys. J.* 76, 2879-2886.
150. Svergun DI, Petoukhov MV and Koch MHJ (2001). "Determination of domain structure of proteins from X-ray solution scattering". *Biophys. J.* 80, 2946-2953.
151. Svergun DI and Koch MHJ (2003). "Small-angle scattering studies of biological macromolecules in solution". *Rep. Prog. Phys.* 66 (2003) 1735-1782.
152. Takahashi Y, Nishikawa Y and Fujisawa T (2003). "Evaluation of three algorithms for ab initio determination of three-dimensional shape from one-dimensional solution scattering profiles". *J. Appl. Cryst.* 36: 549-552.
153. Teodorescu O, Galor T, Pillardy J and Elber R (2004). "Enriching the sequence substitution matrix by structural information", *Proteins, Structure, Function and Genetics*, 54:41-48.
154. Topchieva IN, Efermova NV, Snitko YAE, Khvorov NV, *Doklady Chem.* 339 (1994) 257.
155. Topchieva IN, Sorokina EM, Efermova NV, Ksenofontov AL, Kurganov BI (2000), *Bioconjugate Chem.* 11, 22.
156. Turchin, VF (1965), "Slow Neutrons", p16, Israel Sci. Transl., Jerusalem: Sivan Press;
157. Vakser IA (1995), "Protein docking for low-resolution structures", *Protein Eng.*, 8, 371- 377.
158. Vakser IA (1996), "Low-resolution docking: Prediction of complexes for underdetermined structures", *Biopolymers*, 39, 455-464.
159. Van Holde KE (1971). "Physical biochemistry". Imprenta Englewood Cliffs, N. J.: Prentice-Hall, c1971.
160. Vassar R (2002). "Beta-secretase (BACE) as a drug target for Alzheimer's disease". *Adv Drug Deliv Rev*, 54, 1589-1602.
161. Viellard-Baron J (1972). "Phase Transitions of the Classical Hard-Ellipse System" *J. Chem. Phys.* 56, 4729-4744.
162. Vigil D, Gallagher SC, Trewthella J and García AE (2001). "Functional Dynamics of the Hydrophobic Cleft in the N-Domain of Calmodulin". *Biophys. J.* 2001 80: 2082-2092
163. Voet D, Voet JG and Pratt CW (2000). "Fundamentos de Bioquímica"; trad. Arthur Germano Fett Neto. Porto Alegre: Artmed.
164. Volkov VV and Svergun DI (2003). "Uniqueness of ab initio shape determination in small-angle scattering". *J. Appl. Cryst.* 36, 860-864.

165. Walther D, Cohen FE and Doniach S (2000). "Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules". *J. Appl. Cryst.* (2000). 33, 350-363.
166. Wang RCL, Kreuzer HJ, Grunze M (1997). "Molecular Conformation and Solvation of Oligo(ethylene glycol)-Terminated Self-Assembled Monolayers and Their Resistance to Protein Adsorption". *J. Phys. Chem. B* 101 (1997) 9767.
167. Wang Y, Ha Y, Dulubova I, Ho A, Huryeva I, Sudhof TC and Rizo J (2004). "The X-ray structure of an antiparallel dimer of the human amyloid precursor protein E2 domain". *Mol Cell*, 15, 343-353.
168. Weber RE and Vinogradov SN (2001). "Nonvertebrate Hemoglobins: Functions and Molecular Adaptations", *Physiological Reviews*, 81 (2), 569-628.
169. Williams WSC (1991). "Nuclear and particle physics". Imprenta Oxford: Clarendon, c1991.
170. Wu H and Schmidt PW (1974). "The relation between the particle shape and the outer part of the small-angle X-ray scattering curve". *J. Appl. Cryst.*, 7, 131-146.
171. Zeng W and Doniach S (2002). "Protein Structure Prediction Constrained by Solution X-ray Scattering Data and Structural Homology Identification", *JMB*, 316, 173-187.

APÊNDICE

A. Proteínas

Apresentamos aqui um breve resumo das propriedades principais das proteínas, tanto estruturais quanto funcionais. Estas informações foram obtidas em diversos livros textos (Voet et al, 2000; Lehninger, 1990), e uma descrição mais aprofundada pode ser obtida nestas fontes.

A.1. Introdução

Proteínas são as macromoléculas mais abundantes nas células vivas e constituem 50 por cento ou mais de seu peso seco. Elas se encontram em todas as células e em todas as partes das células. As proteínas também ocorrem em grande diversidade, centenas de diferentes tipos podem ser encontrados em uma única célula. Além disso, as proteínas têm diferentes papéis biológicos por serem instrumentos moleculares através dos quais se expressa a informação genética.

A chave da estrutura das milhares de proteínas diferentes entre si é o grupo de moléculas fundamentais relativamente simples com as quais as proteínas são constituídas. Todas as proteínas, seja das mais antigas linhagens de bactérias ou das formas de vidas mais complexas, são construídas com o mesmo conjunto de 20 aminoácidos, unidos covalentemente em seqüências características. Dependendo do agrupamento destes aminoácidos, a proteína possuirá uma característica particular. Os aminoácidos são o alfabeto da estrutura protéica, pois eles podem ser agrupados em um número quase infinito de seqüências para formar um número quase infinito de diferentes proteínas.

As proteínas possuem muitas funções biológicas diferentes. Podemos identificar várias grandes classes de proteínas de acordo com seus papéis biológicos (Tabela A1).

Tabela A1 – Classificação das Proteínas de acordo com a Função Biológica.

Classe	Exemplo
Enzimas	Ribonuclease, Tripsina
Proteínas Transportadoras	Hemoglobina, Albumina, Mioglobina, β_1 -lipoproteína
Proteínas nutritivas e de reserva	Gliadina (trigo), Ovoalbumina, Caseína (leite), Ferritina
Proteínas contráteis ou de movimento	Actina, Miosina, Tubulina, Dineína
Proteínas estruturais	Queratina, Fibroína, Colágeno, Elastina, Proteoglicanas
Proteínas de defesa	Anticorpos, Fibrinogênio, Trombina, Toxina botulínica, Toxina diftérica, Veneno de serpentes, Ricina
Proteínas reguladoras	Insulina, Hormônio de crescimento, Corticotrofina, Repressores

As proteínas são historicamente classificadas como fibrosas e globulares (Figura A1), dependendo de sua morfologia geral. Nas proteínas globulares as cadeias polipeptídicas estão fortemente enroladas em uma forma globular ou esférica. Nas proteínas fibrosas as cadeias estão estendidas ao longo de um eixo; estas moléculas são insolúveis em água.

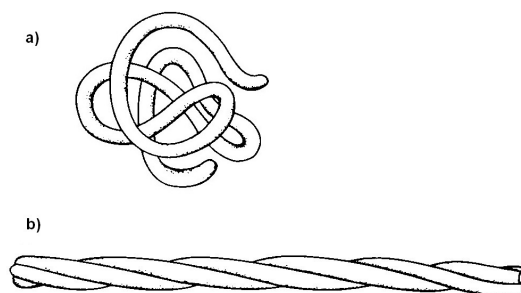


Figura A1– Proteínas globulares e proteínas fibrosas. (a) Nas proteínas globulares a cadeia polipeptídica enovela-se de forma compacta. Elas são quase sempre solúveis em meios aquosos. (b) Em proteínas fibrosas as cadeias polipeptídicas dispõem-se ao longo de um eixo. No desenho temos três cadeias enroladas uma ao redor da outra, compondo uma estrutura semelhante a uma corda.

Apresentaremos a seguir uma descrição sucinta dos componentes principais das proteínas. Isso será muito importante para a interpretação dos dados de SAXS que realizamos ao longo da tese.

A.2. Aminoácidos e Cadeias Polipeptídicas

A análise de um número muito grande de proteínas de praticamente todas as fontes conceituadas, mostra que *todas as proteínas são compostas de 20 aminoácidos “padrão”* (Tabela A2).

Tabela A2 – Abreviações dos Aminoácidos

Aminoácido	Abreviação de três letras	Abreviação de uma letra
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Ácido Aspártico	Asp	D
Ácido Glutâmico	Glu	E
Cisteína	Cys	C
Glicina	Gly	G
Glutamina	Gln	Q
Histidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Tirosina	Tyr	Y
Treonina	Thr	T
Triptofano	Trp	W
Valina	Val	V

Estas moléculas são conhecidas como *α -aminoácidos* porque, com exceção da prolina, todas possuem um grupo primário amina e um grupo substituinte de ácido carboxílico (Figura A2) no mesmo átomo de carbono.

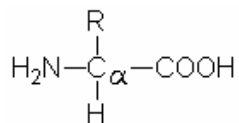


Figura A2 – Fórmula estrutural geral para α -aminoácidos. Existem 20 diferentes grupos R, gerando assim os 20 aminoácidos padrão.

Na faixa de pH fisiológico, tanto o ácido carboxílico e os grupos amina dos α -aminoácidos estão completamente ionizados. (Figura A3).

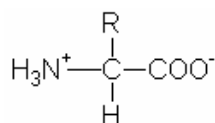


Figura A3 – A forma dos α -aminoácidos em valores de pH fisiológico.

Os α -aminoácidos polimerizam, ao menos conceitualmente, através da eliminação de uma molécula de água (Figura A4).

A ligação CO - NH é conhecida como *ligação peptídica*. Polímeros compostos de dois, três, poucos (3 – 10) ou muitos *resíduos aminoácido* (alternativamente chamados unidades polipeptídicas) são conhecidos, respectivamente, como *dipeptídica*, *tripeptídicas*, *oligopeptídicos* e *polipeptídios*. Estas subunidades, no entanto, são usualmente chamadas de “peptídeos”. *Proteínas são moléculas que se consistem de uma ou mais cadeias polipeptídica*. Esses polipeptídios variam no comprimento de 40 à 400 resíduos aminoácidos e, como a massa média de um resíduo aminoácido é $\sim 110\text{Da}^{\text{§§§}}$, possuem massas moleculares variando de ~ 4 até mais de $\sim 440\text{kDa}$.

^{§§§} O peso molecular de proteínas é expresso em Daltons (Da), definido como 1/12 das massas do átomo de ^{12}C ($1000\text{Da} = 1\text{kilodalton} = 1\text{kDa}$).

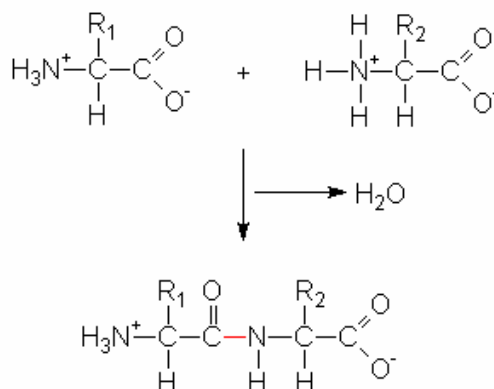


Figura A4 – A condensação de dois α -aminoácidos para formar um dipeptídeo. A ligação peptídica é mostrada em vermelho.

Polipeptídios são *polímeros lineares*, ou seja, cada amino ácido é ligado ao seu vizinho em uma configuração cabeça-com-final ao invés de formar cadeias ramificadas. Os resíduos das extremidades da cadeia fazem apenas uma ligação peptídica cada um. Os resíduos com o grupo amino livre (por convenção, o resíduo da extrema esquerda como mostrado na figura anterior) é chamado de **aminoterminal** ou **N-terminal**. O resíduo com um grupo carboxílico livre (o da direita) é chamado **carbóxi-terminal** ou **C-terminal**.

A forma mais útil de classificar os 20 aminoácidos-padrão é pela polaridade de suas cadeias laterais (*grupos R*). Isto porque as proteínas enovelam-se para sua conformação nativa em resposta a uma tendência de remover suas cadeias laterais hidrofóbicas do contato com água e fazer o contrário com suas cadeias hidrofílicas. De acordo com esse esquema de classificação, existem três tipos de aminoácidos: (1) aqueles com grupos R não polares, (2) aqueles com grupos R polares não carregados, (3) aqueles com grupos R carregados (Voet et al., 2002).

A.3. Estrutura Tridimensional de Proteínas

As propriedades funcionais de uma proteína são grandemente determinadas pela sua estrutura tridimensional. A estrutura tridimensional da proteína *nativa* (enovelada fisiologicamente) é especificada pela estrutura primária formando assim um conjunto único de características. Com isso a função de uma proteína pode ser entendida em termos de sua estrutura, ou seja, relações tridimensionais entre os átomos componentes. A descrição

estrutural das proteínas, bem como de outros materiais poliméricos, tem sido tradicionalmente descrito em termos de quatro níveis de organização:

- A *estrutura primária* é a sequência de aminoácidos de suas cadeias polipeptídicas.
- A *estrutura secundária* é o arranjo local e espacial de suas cadeias polipeptídicas sem nos preocuparmos com a conformação de suas cadeias laterais.
- A *estrutura terciária* se refere à estrutura tridimensional do polipeptídio como um todo. A distinção entre estrutura secundária e terciária pode parecer um pouco vaga: na prática o termo estrutura secundária indicará entidades estruturais como, por exemplo, hélices e folhas beta.
- Muitas proteínas são compostas de duas ou mais cadeias polipeptídicas, referidas como *subunidades*, associadas por interações não-covalentes e em alguns casos pontes dissulfeto. A *estrutura quaternária* se refere ao arranjo espacial destas subunidades.

A.3.1. Estrutura Primária

A estrutura primária de uma proteína consiste na sequência de aminoácidos da sua cadeia polipeptídica ou das suas cadeias polipeptídicas, no caso de ela ser constituída por mais de uma cadeia. Um exemplo de sequência de aminoácidos é mostrado abaixo para a proteína lisozima:

```
>6LYZ: _ LYSOZYME - CHAIN
KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINS
RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDV
QAWIRGCRL
```

Em geral as proteínas possuem pelo menos 40 resíduos, sendo os polipeptídios com menos de 40 resíduos denominados **peptídeos**. A maioria dos polipeptídios contém entre 100 e 1000 resíduos. Além destas limitações de tamanho, os polipeptídios estão sujeitos a limitações mais severas na composição de aminoácidos. Os 20 aminoácidos padrão não aparecem com a mesma frequência nas proteínas. Por exemplo, os aminoácidos mais abundantes são Leu, Ala, Gly, Ser, Val e Glu; os mais raros são Trp, Cys, Met e His. Pelo fato de cada resíduo possuir propriedades físicas e químicas características, sua presença em uma dada posição em uma proteína influencia suas propriedades. Além disso, as

características de uma proteína dependem mais da sequência dos aminoácidos do que da composição propriamente dita, pela mesma razão que “amor” e seu anagrama “roma” são palavras bem diferentes.

O conhecimento da sequência de aminoácidos de uma proteína é um pré-requisito para determinar sua estrutura tridimensional. Além disso, ferramentas computacionais poderosas disponíveis atualmente permitem a busca por sequências homólogas entre proteínas de espécies diferentes, o que fornece elementos para entender a sua função bem como relações entre as proteínas e os organismos que as produzem.

A.3.2. Estrutura Secundária

A estrutura secundária das proteínas inclui os padrões regulares de dobramento de polipeptídios, como as hélices, as folhas pregueadas e as voltas. Todas estas estruturas podem ser compreendidas com base nas propriedades geométricas dos grupos peptídicos.

Estudos realizados nas décadas de 30 e 40, utilizando raios X indicam que *os grupos peptídicos tem uma estrutura rígida planar*. Grupos peptídicos, com apenas algumas exceções, assumem a conformação *trans*: na qual átomos C_α sucessivos estão em lados opostos da ligação peptídica que os une (Figura A5). Isto é o resultado de uma interferência que ocorre na conformação *cis* e a torna $\sim 8 kJ.mol^{-1}$ menos estável que a conformação *trans*.

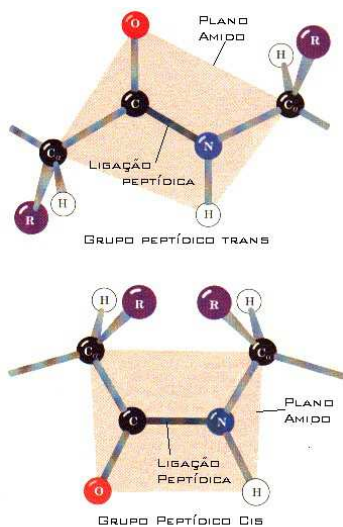


Figura A5 – Grupos peptídicos planares trans e cis. Figura tirada de Voet et al, 2002.

As considerações anteriores são importantes, pois indicam que a cadeia da proteína é uma seqüência de grupos planares unidos. Podemos no entanto especificar a conformação da cadeia pelos *ângulos de torção* (ângulos de rotação ou *ângulos diedrais*) entre a ligação $C_{\alpha} - N$ (ϕ), e a ligação $C_{\alpha} - C$ (ψ) de cada um dos resíduos aminoácido. Estes ângulos ϕ e ψ (Figura A6) são definidos como 180° quando a cadeia polipeptídica é planar, totalmente estendida e aumenta para uma rotação horária partindo de C_{α} .

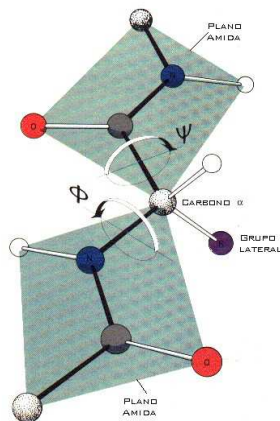


Figura A6 – *Uma porção da cadeia polipeptídica indicando os ângulos torcionais de cada unidade peptídica. Os únicos movimentos razoáveis ocorrem em torno da ligação $C_{\alpha} - N$ (ϕ) e na ligação $C_{\alpha} - C$ (ψ). Os ângulos de torção são ambos 180° na figura, e aumentam como indicado. Figura tirada de Voet et al, 2002.*

Os valores permitidos para ϕ e ψ podem ser determinados calculando-se as distâncias entre os átomos de um tripeptídio em todos os valores de ϕ e ψ para o peptídeo central. Conformações proibidas são aquelas onde as distâncias interatômicas são menores que a correspondente distância de *Van der Waals* de cada átomo. Tal informação é sumarizada no *mapa conformacional* ou *diagrama de Ramachandran* (Figura A7).

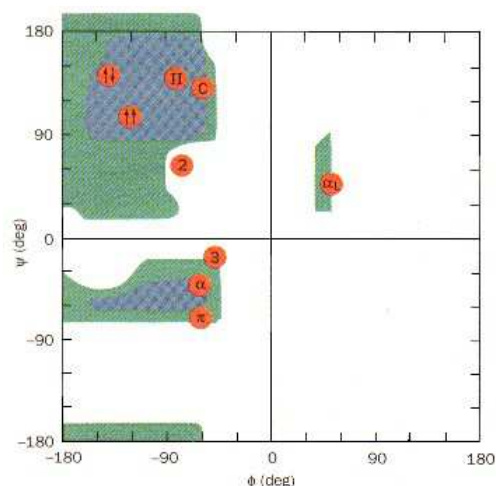


Figura A7 – Diagrama de Ramachandran. Este diagrama mostra os ângulos ϕ e ψ permitidos para a poli-L-alanina. O diagrama foi calculado utilizando-se distâncias de van der Waals entre ligações. Regiões normalmente possíveis de ϕ e ψ são indicadas em azul, enquanto que as regiões em verde correspondem a conformações que atingem o limite extremo das distâncias de van der Waals. Os ângulos conformacionais de diversas estruturas são indicados abaixo. Figura tirada de Voet et al, 2002

Tabela A3 – Ângulos ϕ e ψ para diversas estruturas.

Estrutura Secundária	ϕ (deg.)	ψ (deg.)
α -hélice à direita (α)	-57	-47
Conformação β Paralela ($\uparrow\uparrow$)	-119	113
Conformação β Antiparalela ($\uparrow\downarrow$)	-139	135
3_{10} helice (3)	-49	-26
π hélice (π)	-57	-70
2.2 ₇ ribbon (2)	-78	59
Hélices poliglicina II e poli-L-prolina II (II)	-79	150
Colágeno (C)	-51	153
α -hélice à esquerda (α_L)	57	47

Por este mapa vemos que a maior parte das áreas são inacessíveis para a cadeia polipeptídica.

A.3.2.1. A α -Hélice

Apenas uma hélice polipeptídica possui um padrão de pontes de hidrogênio e valores de ϕ e ψ que estão dentro da região totalmente permitida do diagrama de

Ramachandran: a α -hélice. Na Figura A8 temos uma α -hélice orientada para a direita, ou seja, é torcida na mesma direção em que os dedos da mão direita se fecham quando o polegar aponta na direção em que a hélice sobe. A α -hélice possui 3,6 resíduos por volta e um passo (a distância que a hélice aumenta ao longo do eixo por volta) de 5,4Å.

Na Figura A8 vemos que as pontes de hidrogênio de uma hélice são arranjadas de modo que a ligação peptídica C=O do n -ésimo resíduo aponta ao longo da hélice em direção ao grupo peptídico N-H do $(n+4)$ -ésimo resíduo. Isso resulta em uma ponte de hidrogênio muito forte, que possui uma distância ótima N...O da ordem de 2,8 Å.

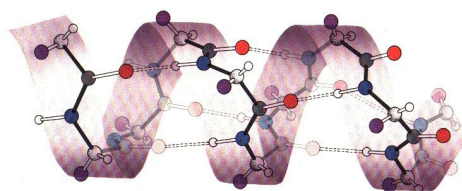


Figura A8 – Representação esquemática de uma α -hélice à direita. As pontes de hidrogênio entre os grupos N-H e C=O são indicados com linhas tracejadas. Figura tirada de Voet et al, 2002

A α -hélice é um elemento comum da estrutura secundária tanto de proteínas fibrosas quanto de proteínas globulares. Nas proteínas globulares, α -hélices tem um conjunto de 12 resíduos, o que corresponde a aproximadamente 3 passos de hélice e um comprimento de 18Å. No entanto, α -hélices com mais de 53 resíduos podem ser encontrados.

A.3.2.2. Folhas β

Da mesma forma que as α -hélices, a folha β utiliza todas as pontes de hidrogênio do esqueleto polipeptídico. Entretanto, na folha β , as pontes de hidrogênio ocorrem entre cadeias polipeptídicas vizinhas em vez de no interior da cadeia, como na α -hélice. As folhas apresentam duas variações

- A *conformação β antiparalela*, na qual cadeias polipeptídicas vizinhas unidas por pontes de hidrogênio ocorrem em direções opostas (Figura A9a).
- A *conformação β paralela*, na qual as cadeias unidas por pontes de hidrogênio estendem-se na mesma direção (Figura A9b).

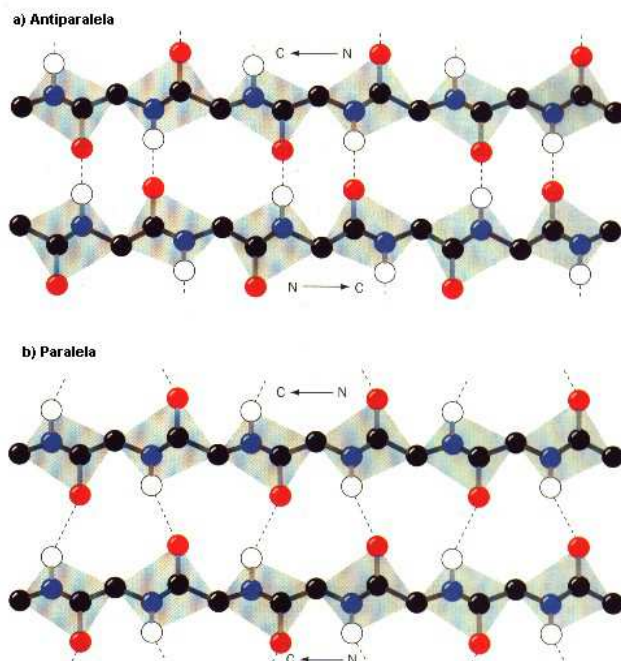


Figura A9 – Associações de pontes de hidrogênio na conformação β Cadeias laterais foram omitidas para clareza. (a) Conformação β antiparalela. (b) Conformação β paralela. Figura tirada de Voet et al, 2002.

As conformações nas quais as estruturas β apresentam quantidades ótimas de pontes de hidrogênio variam um pouco das dos polipeptídios estendidos mostrados na Figura A9. Portanto apresentam uma aparência enrugada ou pregueada nas bordas (Figura A10) e, por essa razão, são algumas vezes denominadas “folhas pregueadas”. Cadeias laterais sucessivas de uma cadeia polipeptídica em folha β estendem-se por lados opostos dessa folha, com repetições de dois resíduos em uma distância de $7,0\text{\AA}$.

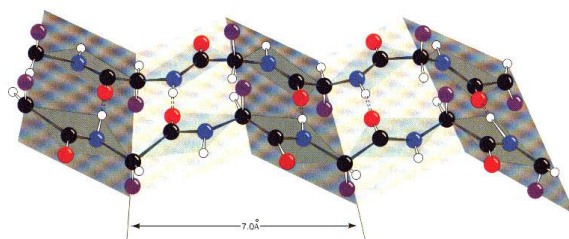


Figura A10 – Conformação β de cadeia dupla desenhada de modo a enfatizar seu caráter plano. Linhas tracejadas indicam pontes de hidrogênio. Figura tirada de Voet et al, 2002.

A conformação β é uma estrutura muito comum em proteínas. Em proteínas globulares, consistem de 2 a mais de 15 cadeias polipeptídicas, a média é de 6, formando um agregado com largura da ordem de $\sim 21\text{\AA}$.

Folhas β paralelas com menos de cinco cadeias são raras. Esta observação sugere que as folhas β paralelas são menos estáveis que conformação β antiparalelas, possivelmente porque as pontes de hidrogênio de conformações β paralelas estão distorcidas, em comparação com aquelas na conformação β antiparalela. Folhas β com misturas de cadeias paralelas e antiparalelas ocorrem com frequência. Além disso, quase invariavelmente as folhas β exibem um giro pronunciado para a direita quando vistas ao longo de suas cadeias polipeptídicas.

A.3.2.3. Estruturas não repetitivas

Estruturas secundárias regulares – hélices e estruturas β – correspondem aproximadamente à metade das cadeias polipeptídicas de uma proteína globular. As cadeias polipeptídicas restantes são ditas como possuindo *conformação em laço*. Isto não quer dizer que estas estruturas secundárias não-repetitivas são menos ordenadas que hélices ou conformação β ; elas são simplesmente irregulares e mais difíceis de descrever. Não devemos confundir o termo conformação em laço com o termo *laço aleatório* (ou *random coil*), que se refere a um conjunto de conformações totalmente desordenado e com flutuações rápidas que ocorrem quando temos proteínas denaturadas em solução.

Segmentos com estrutura secundária regular como as α -hélices e fitas de folhas β são tipicamente unidas por pedaços de polipeptídios que mudam abruptamente de direção. Tais **voltas reversas** ou **curvaturas β** (denominadas genericamente por *coils*) quase sempre ocorrem na superfície das proteínas.

A.3.2.4. Alinhamento de seqüências e predição de estrutura secundária

A estrutura de uma proteína, e conseqüentemente sua função protéica, está intimamente ligada à sua estrutura primária. Sendo assim é de se esperar que proteínas com seqüências (ou também estruturas) semelhantes possuam funções semelhantes. A determinação do grau de semelhança (homologia) entre uma dada seqüência protéica com

relação à um grupo de outras seqüências (banco de dados) denomina-se alinhamento de seqüência primária.

No alinhamento de seqüências protéicas as ferramentas de busca alinham não somente a seqüência de aminoácidos em si, mas também buscam correlações entre as estruturas secundárias das proteínas. Alguns programas predizem, através da comparação com um banco de dados, a estrutura secundária a partir da seqüência de aminoácidos (programas BLAST (Altschul et al., 1990), PSIPRED (McGuffin et al., 2000) entre outros). Neste sentido, a informação seqüencial completa de uma proteína pode ser mostrada na forma (fonte *Protein Data Bank* (<http://www.rcsb.org>):

```
Chain 6LYZ:_
Compound      Lysozyme (E.C. 3.2.1.17) - Chain _
Type           Protein
Molecular Weight 14296
Number of Residues 129
Number of Alpha 4          Content of Alpha 30.23
Number of Beta  3          Content of Beta  6.20
```

Sequence and secondary structure

```
1  KVFGRCLEAA AMKRHGLDNY RGYSLGNWVC AAKFESNFNT QATNRNTDGS
   B  HHHHHH HHHHTTTTTB TTBTHHHHHH HHHHHHTTBT T EEE TTS
51  TDYGILQINS RWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCAKKIVS
   EEETTTTEET TTT B SS S S   TT SBG GGGGSS  HH HHHHHHHHHT
101 DGNGMNAWVA WRNRCKGTDV QAWIRGCRL
   SSSGGGGTHH HHHHTTTTTG GGGTTT
```

E -> folhas β ; H -> α -hélices; S,T,B,G -> tipos de coils

A.4. Proteínas Fibrosas ^{****}

Proteínas fibrosas são moléculas muito alongadas cuja estrutura secundária é a parte dominante na estrutura. Do ponto de vista biológico as proteínas fibrosas desempenham importantes papéis na anatomia e fisiologia dos animais. Em grandes vertebrados elas podem constituir um terço ou mais do total de proteínas corporais. Dão proteção externa, pois são os componentes principais da camada mais externa da pele, cabelos, penas, unhas

**** O tema central desta tese serão as proteínas globulares. No entanto para melhor compreensão do contexto, apresentamos uma breve discussão sobre proteínas fibrosas, com suas características principais.

e chifres. As proteínas fibrosas também dão sustentação e forma já que são o maior componente dos tecidos conjuntivos. Podem também ter funções motoras como as proteínas nos músculos e dos cílios. Além disso, são estruturalmente mais simples que as proteínas globulares, o que facilita em muito a determinação de sua estrutura tridimensional utilizando raios X.

As informações relevantes que tiramos do estudo de proteínas fibrosas são basicamente três. Primeiro, vemos que as proteínas possuem não só uma estrutura primária, isto é, um esqueleto covalente, mas também uma estrutura secundária característica, a maneira pela qual resíduos sucessivos de aminoácidos estão arranjados no espaço.

Em segundo lugar vemos que a estrutura secundária, α -hélice e conformação β , é o resultado automático e espontâneo de seu conteúdo e seqüência de aminoácidos. A estrutura secundária característica de uma proteína é a sua forma mais estável em um dado conjunto de condições biológicas, e ocorre principalmente devido à existência de muitas pontes de hidrogênio tanto *intracadeias* (α -hélice) quanto *intercadeias* (conformação β).

Além disso, as conformações tridimensionais das proteínas fibrosas são adaptadas ao exercício de funções biológicas específicas. Por exemplo, a α -hélice da α -queratina é bem adaptada à função de proteção dos cabelos, penas, chifres e escamas em vertebrados; as folhas- β estão adaptadas a prover as características de flexibilidade e inextensibilidade das fibras da seda e das teias de aranha; a conformação do colágeno oferece alta resistência à tensão requerida pelos tendões, etc. Isso é sumarizado na **Tabela A4**.

Tabela A4 – Estrutura secundária e propriedades das Proteínas Fibrosas.

Estrutura	Características	Exemplos
α -hélice, interligada por cistina	Estruturas de proteção resistentes e insolúveis com dureza e flexibilidade variáveis	Cabelos, penas e unhas
Conformação β	Filamentos flexíveis e moles	Seda
Hélice tripla do colágeno	Alta força de tensão, sem esticar-se	Tendões e matriz óssea
Cadeias de elastina interligadas por desmosina	Elástica e podendo esticar-se em duas dimensões	Ligamentos

A.5. Proteínas Globulares

Nas proteínas globulares a cadeia polipeptídica dobra-se várias vezes, resultando em uma estrutura globular compacta. Como classe, elas têm conformação mais compacta que as proteínas fibrosas. Quase todas as duas mil enzimas já identificadas são proteínas globulares. Outras proteínas globulares funcionam como transportadoras de oxigênio, de nutrientes e de íons no sangue; algumas são anticorpos, outras são hormônios e, ainda outras, são componentes de membranas e de ribossomos.

Duas evidências importantes indicam que a cadeia polipeptídica das proteínas globulares é enovelada de forma compacta e que esta conformação é importante para a função biológica exercida por cada uma delas. A primeira evidência é que proteínas globulares sofrem *denaturação* quando aquecidas, expostas a valores extremos de pH, ou tratadas com uréia. Quando a proteína globular é denaturada, seu esqueleto covalente permanece intacto, mas a cadeia polipeptídica desdobra-se ao acaso em conformações espaciais variáveis e irregulares (Figura A11).

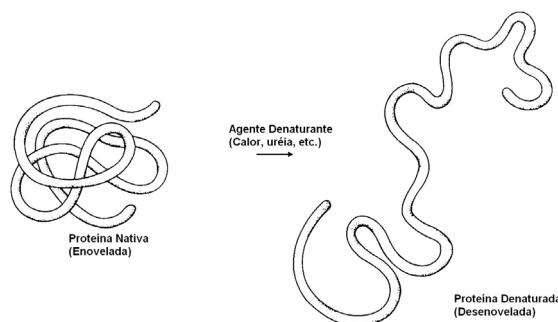


Figura A11 – Esquema de denaturação de uma proteína.

Uma vez denaturada, uma proteína globular geralmente torna-se insolúvel em sistemas aquosos de pH próximo a 7.0, e quase sempre perde sua atividade biológica.

A segunda evidência do dobramento múltiplo sobre si mesma, e conseqüente enovelamento da cadeia polipeptídica das proteínas globulares, procede das diferenças entre o tamanho do comprimento desta cadeia em comparação às dimensões típicas das proteínas. Por exemplo, a soroalbumina (PM. 64.500) tem uma cadeia única com 584 resíduos de aminoácidos. Se esta cadeia estiver em conformação β totalmente estendida ela terá comprimento de quase 2000Å e espessura perto de 5Å. Já se estiver em forma α -hélice, teria perto de 900Å de comprimento e 11Å de espessura. No entanto medidas físico-

químicas da soroalbumina mostram que seu comprimento é apenas de 130Å e sua largura próxima a 30Å. Temos aí a clara evidência de que a cadeia polipeptídica precisa estar enrolada de forma bem compacta.

A.5.1. Estrutura Terciária

A *estrutura terciária* de uma proteína é seu arranjo tridimensional, ou seja, o empacotamento de sua estrutura secundária, em conjunto com a disposição de suas cadeias laterais. Os tipos mais encontrados para a estrutura secundária em proteínas globulares são as α -hélices e as folhas β , variando em proporções e combinações. Algumas proteínas, como a lisozima por exemplo, constitui-se de partes em α -hélices e outras em folhas β (Figura A12). Existem casos onde a proteína somente possui α -hélices ou folhas β , bem como casos onde a maior proporção é de partes não estruturadas (*coils*). Devido ao enorme número de átomos nas proteínas a visualização de todas as posições atômicas em um modelo não fornece informações diretas sobre a sua estrutura secundária. Desta maneira utilizam-se construções simples para representar o grupo de átomos correspondente a esta ou aquela estrutura. O curso da cadeia peptídica pode ser seguido traçando-se as posições dos seus átomos de C_α (para a sequência completa ou somente nas regiões não estruturadas) ou representando-se as hélices como fitas helicoidais ou cilindros e as folhas β como conjuntos de flechas planas apontando da extremidade N- para a C-terminal.

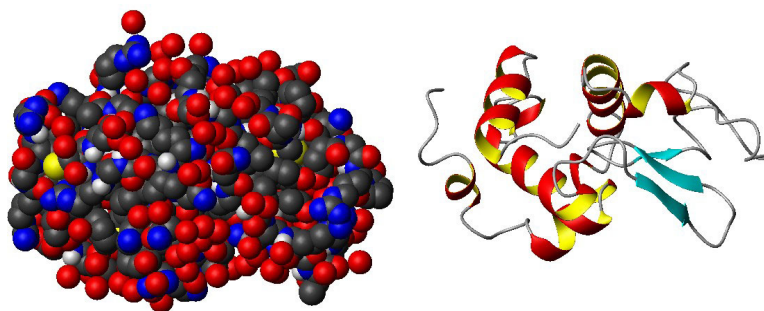


Figura A12 – Estrutura cristalográfica para a lisozima em duas representações diferentes.

As proteínas globulares são muito compactas: existe um espaço muito pequeno em seu interior, de modo que a água é grandemente excluída de seu interior devido a um arranjo tipo micelar de suas cadeias laterais (grupos polares fora, apolares dentro). A *densidade de empacotamento* (razão entre o volume ocupado pelos envelopes de Van der Waals dos átomos na região e o volume da região) da região interior das proteínas

globulares é em média ≈ 0.75 , o que é da mesma ordem de cristas moleculares de pequenas moléculas orgânicas. Em comparação, um empacotamento de esferas do mesmo tamanho possui uma densidade de empacotamento de 0.74 enquanto que líquidos orgânicos (gotas de óleo) possuem densidades de empacotamento entre 0.60 e 0.70. Desta forma, o *interior de uma proteína é mais semelhante a um cristal do que uma gota de óleo; ou seja, é empacotado eficientemente.*

A.5.2. Estrutura Quaternária

A maioria das proteínas, particularmente as com massas moleculares > 100 kDa, é constituída por mais de uma cadeia polipeptídica. Essas subunidades polipeptídicas associam-se com uma geometria específica. O arranjo dessas subunidades é conhecido como *estrutura quaternária* da proteína.

Existem várias razões para proteínas com várias subunidades serem tão comuns. Em grandes coleções de proteínas, como as fibrilas de colágeno, por exemplo, as vantagens da construção por subunidades sobre a síntese de uma cadeia robusta de polipeptídios são análogas àquelas existentes em componentes pré-fabricados na construção de prédios. Defeitos podem ser reparados pela simples troca de subunidades danificadas, e a forma da subunidade pode ser diferente da do produto final, além de a única informação necessária para a especificação de toda proteína são alguns tipos diferentes de subunidades. No caso de enzimas, por exemplo, o aumento do tamanho da proteína melhora o posicionamento tridimensional dos grupos formadores de seus sítios ativos. No entanto é mais eficiente associar-se várias subunidades idênticas do que se aumentar o comprimento da cadeia polipeptídica.

Uma proteína com muitas subunidades pode ser constituída por cadeias polipeptídicas idênticas ou diferentes. Proteínas com mais de uma subunidade são denominadas oligômeros e suas unidades são denominadas protômeros. Um protômero pode ser constituído por uma cadeia polipeptídica ou por várias cadeias polipeptídicas diferentes.

Na grande maioria das proteínas oligoméricas, os protômeros são arranjados simetricamente; isto é, cada protômero ocupa posições geométricas equivalentes no oligômero. Contudo, as proteínas não podem ter simetria de inversão ou reflexão-inversão,

porque isso requereria a conversão de resíduos L em resíduos D (Voet et al., 2002). Como os aminoácidos componentes de proteínas somente possuem configuração estereoquímica L, as proteínas podem ter apenas *simetria rotacional*.

A.6. Considerações Finais

A conformação tridimensional de uma proteína ou de um oligômero protéico é o resultado de várias interações como efeitos hidrofóbicos, interações eletrostáticas, pontes de hidrogênio e interações de Van der Waals. A estrutura final da proteína é o resultado do balanço delicado entre estas poderosas forças concorrentes. Como as proteínas compõem-se de muitos átomos, teremos um problema de muitos corpos o qual não possui solução simples. Mesmo para peptídeos pequenos, a determinação *ab initio*, a partir de sua estrutura primária, de sua estrutura terciária não é possível de ser feita uma vez que não se possui uma metodologia capaz de prever com certeza a conformação que será assumida pela sequência. Neste sentido, métodos que permitam a determinação tridimensional de proteínas são cruciais para a correta compreensão de suas funções. É neste ponto que a cristalografia de proteínas e ressonância nuclear magnética (RMN) possuem importância fundamental uma vez que permitem a determinação da estrutura tridimensional de proteínas. No entanto, como tratam-se de experimentos difíceis de serem executados (seja por dificuldades de cristalização, seja por limitações da técnica de RMN) somente uma pequena fração das proteínas conhecidas pode ser cristalizadas ou resolvidas por RMN. Devido a isso, a técnica de espalhamento a baixos ângulos é útil, pois permite a determinação de informações tridimensionais diretamente das partículas em solução.