



**UNICAMP**

***Tese de Mestrado***

***“SIMULAÇÃO DO ESPALHAMENTO DE RAIOS X POR  
MACROMOLÉCULAS EM SOLUÇÃO ATRAVÉS DA  
CONSTRUÇÃO DE MODELOS TRIDIMENSIONAIS DE  
BAIXA RESOLUÇÃO”***

**Cristiano Luis Pinto de Oliveira  
IFGW - UNICAMP**

***Comissão Julgadora***

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

Prof. Dr. Igor Polikarpov  
LNLS/Campinas-SP

Prof. Dr. Antonio Rubens Brito de Castro  
IFGW/UNICAMP

## FICHA CATALOGRÁFICA ELABORADA PELA

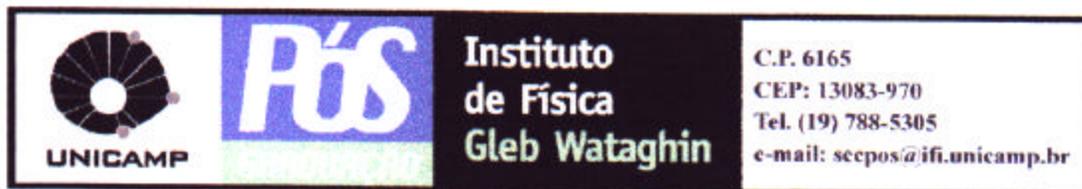
### 1. BIBLIOTECA DO IFGW - UNICAMP

OL4s

Oliveira, Cristiano Luis Pinto de  
Simulação do espalhamento de raios X por macromoléculas em solução através da construção de modelos tridimensionais de baixa resolução / Cristiano Luis Pinto de Oliveira. -- Campinas, SP : [s.n.], 2001.

Orientador: Iris C. L. de Torriani.  
Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Física "Gleb Wataghin".

1. Raios X - Espalhamento de baixo ângulo. 2. Raios X - Métodos de simulação. 3. Modelos e construção de modelos. 4. Macromoléculas - Métodos de simulação. 5. Moléculas - Modelos. I. Torriani, Iris C. L. de. II. Universidade Estadual de Campinas. Instituto de Física "Gleb Wataghin". III. Título.



MEMBROS DA COMISSÃO JULGADORA DA TESE DE MESTRADO DO SR. **CRISTIANO LUIS PINTO DE OLIVEIRA – RA 950364**, APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 16 DE MARÇO DE 2001.

**COMISSÃO JULGADORA:**

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

Prof. Dr. Igor Polikarpov  
LNLS/Campinas-SP

Prof. Dr. Antonio Rubens Brito de Castro  
IFGW/UNICAMP

*Agradecimentos,*

*Seria impossível querer exprimir em poucas palavras os agradecimentos a todas as pessoas que, direta e indiretamente me auxiliaram na conclusão deste trabalho. No entanto existem sempre aquelas que se destacam por terem sido de fundamental importância. A minha orientadora, Dra.Íris Torriani muito obrigado pelo apoio e confiança em meu trabalho. A meus colegas do laboratório que no dia-a-dia tornaram mais fácil e agradável a caminhadas. A meus pais e minha irmã pelo apoio e amor que tornaram possível qualquer passo meu e, sobretudo a Deus por me conceder o dom da vida...*

*Sinceramente, o meu muito obrigado,*

*Cristiano Oliveira*

*16/03/2001*

## RESUMO

Neste trabalho apresentamos um estudo sobre a potencialidade do método de espalhamento a baixos ângulos para a reconstrução da estrutura tridimensional de partículas em solução. Neste sentido muitas simulações foram realizadas para testar e avaliar os métodos utilizados. Inicialmente apresento a teoria geral de espalhamento a baixos ângulos por partículas. Posteriormente enfatizo a construção de modelos o que permite relacionarmos a forma da partícula a seu espalhamento e a sua curva de distribuição de distâncias ( $P(r)$ ). Ao final utilizo estes modelos para avaliar métodos *ab initio* de reconstrução tridimensional de partículas. Com base nesta avaliação pudemos analisar os resultados de sistemas reais compostos por proteínas em solução. Como veremos pelos resultados, a técnica de SAXS é capaz de fornecer informações conclusivas sobre a forma, anisotropia, simetria e mudanças conformacionais das partículas espalhadoras. Dos métodos utilizados, a otimização por *simulated annealing* somada a utilização de vínculos adequados no algoritmo, forneceu resultados melhores na aproximação da forma das partículas estudadas do que os obtidos por otimização com algoritmo genético.

## ABSTRACT

In this work we present a careful study about the potentiality of the small angle scattering technique and the inverse scattering problem ie., the reconstruction of the particle shape directly from the scattering profile. In this way, we made many simulations to test and analyzed various methods. First I show the general scattering theory, emphasizing the scattering from particles. Next I present a model building computing routine that shows the relationship between the scattering profile and the particle shape. Finally, I apply the three-dimensional reconstruction methods to several models in order to evaluate the power of these methods. This calculation served as a basis to analyze real problems like proteins in solution. As we show in the results, the SAXS technique furnishes interesting results like shape, anisotropy, symmetry and conformational changes of the scattering particles. From the *ab initio* reconstruction methods used, optimization by simulated annealing plus the introduction of adequate constraints gave the best shape approximation when compared to genetic algorithm results.

# Índice Geral

1. TEORIA GERAL.....	7
1.1. INTRODUÇÃO .....	7
1.2. PRINCÍPIOS GERAIS .....	8
1.3. SISTEMAS MONODISPERSOS.....	12
1.3.1. PARÂMETROS ESTRUTURAIS GLOBAIS .....	12
1.3.2. FORMA E ESTRUTURA DAS PARTÍCULAS.....	15
1.3.3. INTERFERÊNCIA INTERPARTÍCULAS E EFEITOS DE CONCENTRAÇÃO.....	22
2. SIMULAÇÕES E CÁLCULO DE MODELOS .....	27
2.1. SIMULAÇÕES .....	27
2.2. CÁLCULO DE MODELOS.....	27
2.2.1. CÁLCULO DE INTENSIDADES DE ESPALHAMENTO.....	27
2.2.2. CONSTRUÇÃO DE MODELOS .....	29
3. BUILDER – PROGRAMA PARA CONSTRUÇÃO DE MODELOS .....	33
3.1. METODOLOGIA PARA CONSTRUÇÃO DE MODELOS.....	33
3.1.1. CONSTRUÇÃO DE PRISMAS.....	34
3.1.2. CONSTRUÇÃO DE CILINDROS E CASCAS CILÍNDRICAS .....	35
3.1.3. CONSTRUÇÃO DE ESFERAS, CASCAS ESFÉRICAS, ELIPSÓIDES DE REVOLUÇÃO E CASCAS ELIPSOIDAIIS DE REVOLUÇÃO .....	35
3.2. EXEMPLOS DE INTENSIDADES DE ESPALHAMENTO PARA MODELOS DIVERSOS .....	39
4. RECUPERAÇÃO DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS.....	51
4.1. INFORMAÇÃO CONTIDA EM ESPALHAMENTO A BAIXO ÂNGULO .....	51
4.2. DETERMINAÇÃO DA FORMA DE MACROMOLÉCULAS UTILIZANDO ESFÉRICOS HARMONICOS – EXPANSÃO MULTIPOLAR .....	52
4.3. DETERMINAÇÃO DA FORMA DE MACROMOLÉCULAS UTILIZANDO ALGORITMOS ESTOCÁSTICOS – MÉTODO DE ELEMENTOS FINITOS.....	63
4.3.1. ALGORITMO GENÉTICO ( GA ) .....	64
4.3.2. SIMULATED ANNEALING ( SA ) .....	66
4.3.3. SIMULAÇÕES .....	69
5. APLICAÇÃO DOS MÉTODOS DA RECUPERAÇÃO DE ESTRUTURAS USANDO DADOS EXPERIMENTAIS .....	99
5.1. HEMOGLOBINAS EXTRACELULARES .....	99
5.2. RESTAURAÇÃO TRIDIMENSIONAL DE PROTEÍNAS EXTRACELULARES UTILIZANDO MÉTODOS ESTOCÁSTICOS.....	102
6. CONSIDERAÇÕES FINAIS.....	109
7. REFERÊNCIAS E BIBLIOGRAFIA .....	111



# Índice de Figuras

<i>Figura 1-1: Diferença de fase entre raios incidentes e espalhados.....</i>	<i>9</i>
<i>Figura 1-2: O problema inverso de espalhamento ( figura baseada de ref. [41]).....</i>	<i>10</i>
<i>Figura 1-3: A altura da função <math>p(r)</math> para um certo valor de <math>r</math> é proporcional ao número de todas as linhas com comprimento entre <math>r</math> e <math>r+dr</math> no interior da partícula ( figura baseada de ref. [41]).....</i>	<i>11</i>
<i>Figura 1-4: Comparação entre funções de espalhamento de uma esfera (---) e um cubo (—) com o mesmo raio de giro ( figura baseada de ref. [22]).....</i>	<i>16</i>
<i>Figura 1-5: Funções de distribuição de distâncias para uma esfera (---) e um cubo (—) com o mesmo raio de giro e mesma intensidade na origem. ( figura baseada de ref. [22]).....</i>	<i>17</i>
<i>Figura 1-6: Comparação entre funções <math>p(r)</math> de uma esfera (—), um elipsóide prolato de revolução com relação entre eixos dada por 1:1:3 (---), e um elipsóide oblato de revolução 1:1:0.2 (----) com o mesmo raio de giro ( figura baseada de ref. [22]).....</i>	<i>18</i>
<i>Figura 1-7: Comparação da função de espalhamento de uma esfera, elipsóide prolato e oblato. (veja legenda da Figura 1-6) ( figura baseada de ref. [22]).....</i>	<i>18</i>
<i>Figura 1-8: Partícula tipo bastão.....</i>	<i>19</i>
<i>Figura 1-9: Funções de distribuição de distâncias de paralelepípedos homogêneos com lados: (a)50x50x500Å; (b)50x50x250Å; (c)50x50x150Å. ( figura baseada de ref. [22]).....</i>	<i>20</i>
<i>Figura 1-10: Três paralelepípedos com comprimento constante <math>L</math> (400Å) e seção transversal constante, variando-se as dimensões: — 40x40Å; --- 80x20Å; ---- 160x10Å ( figura baseada de ref. [22]).....</i>	<i>20</i>
<i>Figura 1-11: Função <math>f(r)</math> para três paralelepípedos com comprimento constante <math>L</math> (400Å) e seção transversal constante, variando-se as dimensões: — 40x40Å; --- 80x20Å; ---- 160x10Å. ( figura baseada de ref. [22]).....</i>	<i>21</i>
<i>Figura 1-12: Funções de distribuição de distâncias para dímeros construídos a partir de elipsóides prolatos. — monômeros, --- dímeros, — diferença entre dímeros e monômeros. ( figura baseada de ref. [41]).....</i>	<i>22</i>
<i>Figura 1-13: Tipos característicos das funções de espalhamento: (a) tipo-gás; (b) espalhamento por partículas; (c) tipo líquido. ( figura baseada de ref. [41]).....</i>	<i>23</i>
<i>Figura 1-14: Função de distribuição de distâncias para modelo de esfera rígida. Funções <math>p(r)</math> teóricas: — <math>\Theta = 0</math>, --- <math>\Theta = 0.25</math>, ---- <math>\Theta = 0.5</math>, ————— <math>\Theta = 1.0</math>, <math>h_1.R = 2.0</math>, 2% de erro estatístico, <math>D_{max}=300\text{Å}</math>, <math>\Delta R_g=0.5\%</math>, <math>\Delta I_0=1.2\%</math> ( figura baseada de ref. [41]);.....</i>	<i>25</i>
<i>Figura 2-1: Exemplo representativo da dependência entre o raio da subunidade e o máximo valor da função de espalhamento no espaço recíproco.....</i>	<i>30</i>
<i>Figura 3-1: Prisma reto com lados <math>a, b, c</math> centrado na origem. ....</i>	<i>34</i>
<i>Figura 3-2A: Seccionamento do prisma da figura anterior em <math>N</math> camadas.....</i>	<i>34</i>
<i>Figura 3-3: Prisma construído pelo programa BUILDER. Na figura mostro uma camada em detalhe.....</i>	<i>34</i>
<i>Figura 3-4A: Esquema de preenchimento de um círculo com esferas. ....</i>	<i>35</i>
<i>Figura 3-5A: Cilindro e tubo construídos com o programa BUILDER. Na figura detalho uma das camadas em cada sólido.....</i>	<i>35</i>
<i>Figura 3-6: Esquema do preenchimento de um elipsóide de revolução com esferas.....</i>	<i>36</i>
<i>Figura 3-7: Esfera e elipsóides prolato e oblato construídos com o programa BUILDER. ....</i>	<i>36</i>
<i>Figura 3-8: Inaplicabilidade do método anterior para construir-se cascas elipsoidais. ....</i>	<i>37</i>
<i>Figura 3-9A: Parametrização da elipse em função do ângulo <math>\Theta</math> e do raio <math>\rho</math>.....</i>	<i>37</i>
<i>Figura 3-10A: Cascas esférica e elipsoidais construídas pelo programa BUILDER. ....</i>	<i>38</i>
<i>Figura 3-11A: Seccionamento das cascas da figura anterior para representação da estrutura interna. ..</i>	<i>38</i>
<i>Figura 3-12: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para anel circular de raio 10nm. ....</i>	<i>39</i>

<i>Figura 3-13: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para composição de anéis circulares de raios 10nm.</i>	39
<i>Figura 3-14: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para prisma de lados 10x15x30nm.</i>	40
<i>Figura 3-15: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para esfera de raio 20.</i>	40
<i>Figura 3-16: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para elipsóide oblato com semi-eixos (50,50,10).</i>	41
<i>Figura 3-17: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para elisóide prolato com semi-eixos (10,10,45).</i>	41
<i>Figura 3-18: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para dímeros em L. Os monômeros são elipsóides prolatos com semi-eixos (10,10,20).</i>	42
<i>Figura 3-19: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para cilindro de raio 20nm e altura 50nm.</i>	42
<i>Figura 3-20: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para tubo de raio 10nm e altura 40nm.</i>	43
<i>Figura 3-21: Modelo tridimensional, função de espalhamento e <math>p(r)</math> para arranjo de tubos de raio 10nm e altura 40nm.</i>	43
<i>Figura 3-22: Função de espalhamento para um sistema de multicamadas esféricas. —desenvolvimento analítico, —método de elementos finitos, —gráfico das diferenças entre os cálculos.</i>	44
<i>Figura 3-23: Modelo de Multicamadas esféricas.</i>	45
<i>Figura 3-24: Modelo de Multicamadas elipsoidais.</i>	45
<i>Figura 3-25: Função de espalhamento para o Modelo de Multicamadas elipsoidais.</i>	46
<i>Figura 3-26: Modelo de Multicamadas elipsoidais com pesos diferentes para cascas alternadas.</i>	46
<i>Figura 3-27: Função de espalhamento para o Modelo de Multicamadas elipsoidais com diferentes pesos. Note o aparecimento de mais picos.</i>	47
<i>Figura 3-28: Modelo de Multicamadas Cilíndricas.</i>	47
<i>Figura 3-29: Intensidade de espalhamento para o modelo de multicamadas cilíndricas.</i>	48
<i>Figura 3-30: Modelo de Multicamadas planas.</i>	48
<i>Figura 3-31: Intensidade de espalhamento para o modelo de multicamadas planas.</i>	49
<i>Figura 4-1: Envelope ab initio restaurado e ajuste obtido para uma esfera utilizando esféricos harmônicos.</i>	55
<i>Figura 4-2: Envelope ab initio restaurado e ajuste obtido para um elipsóide prolato utilizando esféricos harmônicos.</i>	56
<i>Figura 4-3: Envelope ab initio restaurado e ajuste obtido para um elipsóide oblato utilizando esféricos harmônicos.</i>	57
<i>Figura 4-4: Envelope ab initio restaurado e ajuste obtido para um dímero de elipsóides prolatos utilizando esféricos harmônicos.</i>	58
<i>Figura 4-5: Envelope ab initio restaurado e ajuste obtido para dímero de elipsóides oblatos utilizando esféricos harmônicos.</i>	59
<i>Figura 4-6: Envelope ab initio restaurado e ajuste obtido para cilindro utilizando esféricos harmônicos.</i>	60
<i>Figura 4-7: Envelope ab initio restaurado e ajuste obtido albumina utilizando esféricos harmônicos.</i>	61
<i>Figura 4-8: Esquema da implementação do método do algoritmo genético em simulações de SAXS para resolver numericamente o problema inverso de espalhamento. ( figura tirada de ref. [62]).</i>	65
<i>Figura 4-9: Um exemplo da estratégia de mascaramento. O resultado de aumentos sucessivos de resolução pela criação de novos espaços de busca com subunidades menores. <math>N</math> é o número de subunidades do espaço de busca inicial para cada um dos passos do algoritmo e <math>R</math> o raio da subunidade. . (figura tirada de ref. [62])</i>	66
<i>Figura 4-10: Retauração ab initio e ajuste obtido para cilindro utilizando simulated annealing.</i>	70
<i>Figura 4-11: Retauração ab initio e ajuste obtido para cilindro utilizando algoritmo genético.</i>	71

<i>Figura 4-12: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>72</i>
<i>Figura 4-13: Retauração ab initio e ajuste obtido para elipsóide prolato utilizando simulated annealing</i>	<i>73</i>
<i>Figura 4-14: Retauração ab initio e ajuste obtido para cilindro utilizando algoritmo genético.....</i>	<i>74</i>
<i>Figura 4-15: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>74</i>
<i>Figura 4-16: Retauração ab initio e ajuste obtido para elipsóide oblato utilizando simulated annealing .</i>	<i>75</i>
<i>Figura 4-17: Retauração ab initio e ajuste obtido para elipsóide oblato utilizando algoritmo genético..</i>	<i>76</i>
<i>Figura 4-18: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>77</i>
<i>Figura 4-19: Retauração ab initio e ajuste obtido para dímero de elipsóides prolatos utilizando simulated annealing.....</i>	<i>78</i>
<i>Figura 4-20: Retauração ab initio e ajuste obtido para dímero de elipsóides prolatos utilizando algoritmo genético.....</i>	<i>79</i>
<i>Figura 4-21: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>79</i>
<i>Figura 4-22: Retauração ab initio e ajuste obtido para dímero de elipsóides oblatos utilizando simulated annealing .....</i>	<i>80</i>
<i>Figura 4-23: Retauração ab initio e ajuste obtido para dímero de elipsóides oblatos utilizando algoritmo genético.....</i>	<i>81</i>
<i>Figura 4-24: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>82</i>
<i>Figura 4-25: Retauração ab initio e ajuste obtido para esfera utilizando simulated annealing.....</i>	<i>83</i>
<i>Figura 4-26: Retauração ab initio e ajuste obtido para esfera utilizando algoritmo genético.....</i>	<i>84</i>
<i>Figura 4-27: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>84</i>
<i>Figura 4-28: Retauração ab initio e ajuste obtido para albumina utilizando simulated annealing.....</i>	<i>85</i>
<i>Figura 4-29: Retauração ab initio e ajuste obtido para albumina utilizando algoritmo genético.....</i>	<i>86</i>
<i>Figura 4-30: Sobreposição dos resultado obtidos pelos métodos. ....</i>	<i>87</i>
<i>Figura 4-31: Esquema da variação angular entre monômeros componentes de um dímero.....</i>	<i>90</i>
<i>Figura 4-32: Intensidade espalhada e curva de distribuição de distâncias para dímeros em diferentes ângulos.....</i>	<i>90</i>
<i>Figura 4-33: Restauração ab initio e ajuste obtido para dímero em 60° utilizando simulated annealing. Na simulação utilizamos simetria P2.....</i>	<i>91</i>
<i>Figura 4-34: Restauração ab initio e ajuste obtido para dímero em 60° utilizando algoritmo genético..</i>	<i>92</i>
<i>Figura 4-35: Restauração ab initio e ajuste obtido para dímero em 90° utilizando simulated annealing. Na simulação utilizamos simetria P2.....</i>	<i>93</i>
<i>Figura 4-36: Retauração ab initio e ajuste obtido para dímero em 60° utilizando algoritmo genético....</i>	<i>94</i>
<i>Figura 4-37: Restauração ab initio e ajuste obtido para dímero em 120° utilizando simulated annealing. Na simulação utilizamos simetria P2 .....</i>	<i>95</i>
<i>Figura 4-38: Restauração ab initio e ajuste obtido para dímero em 90° utilizando algoritmo genético..</i>	<i>96</i>
<i>Figura 5-1: Grupo Heme (Fé – protoporfirina IX).....</i>	<i>99</i>
<i>Figura 5-2: Representação de um oligômero composto de proteínas.....</i>	<i>99</i>
<i>Figura 5-3: Modelo tridimensional, ajuste das curvas de espalhamento e distribuição de distâncias obtidos por Cristina F. Souza. [ 36 ].....</i>	<i>101</i>
<i>Figura 5-4: Restauração ab initio e ajuste obtido para glossoscolex paulistus, utilizando simulated annealing. Na simulação utilizamos simetria P3 .....</i>	<i>103</i>
<i>Figura 5-5: Restauração ab initio e ajuste obtido para glossoscolex paulistus, utilizando algoritmo genético.....</i>	<i>104</i>
<i>Figura 5-6: Restauração ab initio e ajuste obtido para arenicula marina, utilizando simulated annealing. Na simulação utilizamos simetria P3 .....</i>	<i>105</i>
<i>Figura 5-7: Restauração ab initio e ajuste obtido para arenicula marina, utilizando algoritmo genético. ....</i>	<i>106</i>



# 1. TEORIA GERAL

## 1.1. INTRODUÇÃO

Os princípios físicos de espalhamento são os mesmos para difração a alto ângulo e espalhamento de raios-X a baixo ângulo. O campo elétrico das ondas incidentes induz uma oscilação dipolar nos átomos. A energia dos raios X é tão alta que todos os elétrons são excitados. As cargas assim aceleradas geram ondas secundárias que se adicionam a longas distâncias (enfoque de campo distante) formando a amplitude espalhada total. Todas as ondas secundárias possuem a mesma frequência mas podem ter diferentes fases devido a diferenças de caminho. Devido à sua alta frequência é somente possível detectar a intensidade espalhada – o quadrado da amplitude espalhada – e sua dependência com o ângulo de espalhamento.

A dependência angular da amplitude de espalhamento está relacionada com a densidade eletrônica do centro espalhador por uma transformada de Fourier. Tudo isso é verdadeiro tanto para difração a alto ângulo quanto para espalhamento de raios X a baixo ângulo. A principal diferença é que em uma estrutura cristalina, temos um arranjo periódico de centros espalhadores idênticos (partículas), isto é, o meio espalhador é periódico nas três dimensões com um grande número de repetições, enquanto que em espalhamento de raios X a baixo ângulo as partículas, por exemplo, proteínas, não estão ordenadas periodicamente. Elas estão imersas com orientação arbitrária e distâncias irregulares entre si, em uma matriz aquosa. O número de partículas é grande podendo ser assumidos idênticos como na cristalografia.

A transformada de Fourier de um sistema periódico em cristalografia (difração por cristais) corresponde a Séries de Fourier, isto é, a estrutura periódica é expandida em um sistema periódico de funções.

A transformada de Fourier de uma estrutura limitada não periódica (espalhamento a baixo ângulo) corresponde a uma Integral de Fourier. Matematicamente, significa expandirmos uma função não periódica utilizando um sistema de funções periódicas.

Assim as diferenças entre cristalografia e espalhamento a baixo ângulo são equivalentes às diferenças entre séries de Fourier e integrais de Fourier. Pode parecer uma atitude tola expandir-se uma função não periódica com funções periódicas, mas este é o princípio que o espalhamento utiliza e não existem outros processos físicos tão poderosos quanto este para estudar este tipo de estruturas.

Em adição a isto existe uma perda de informação em experimentos espalhamento a baixo ângulo devido à média realizada sobre todas as orientações no espaço. A estrutura tridimensional é representada por uma função unidimensional (a dependência da curva de espalhamento com o ângulo de espalhamento). O cálculo da função de espalhamento para uma estrutura conhecida é chamado solução do problema de espalhamento. Este problema pode ser resolvido exatamente para diversos tipos de estruturas. O inverso, isto é, a estimativa da estrutura tridimensional partindo da função de espalhamento não possui solução única.

Utilizando esta técnica de espalhamento a baixos ângulos, é possível estudar-se sistemas monodispersos e polidispersos. No caso de sistemas monodispersos é possível determinar-se o tamanho, forma e, sobre certas circunstâncias, a estrutura interna. Monodispersidade não pode ser deduzida por esta técnica devendo ser assumida ou confirmada por métodos independentes.

Para sistemas polidispersos uma função de distribuição de tamanhos pode ser obtida devendo-se assumir alguma forma da partícula.

Todas as afirmações anteriores são estritamente verdadeiras para sistemas altamente diluídos, onde a distância interpartícula é muito maior que o tamanho das partículas. No caso de sistemas semidiluídos o resultado de experimentos de espalhamento a baixos ângulos são influenciados pela estrutura da partícula e seu arranjo espacial. Em sistemas densos, isto é, sistemas cuja fração de volume das partículas (centros espalhadores) é da mesma ordem de magnitude da fração de volume da matriz, é possível determinar-se estas frações de volume e o comprimento característico das fases. No entanto, nos concentraremos em sistemas diluídos neste trabalho.

Como os experimentos a baixo ângulo estão relacionados com os outros experimentos de espalhamento? Espalhamento a baixo ângulo, utiliza radiação com comprimento de onda na faixa de 1 a 10 Å, dependendo do problema a se estudar e da fonte utilizada. Esta faixa é similar para raios X e nêutrons, mas os nêutrons interagem com o núcleo, enquanto que os raios X interagem com os elétrons. A eficiência do espalhamento aumenta linearmente com o número atômico para raios X. Já para nêutrons depende do fator de espalhamento (*scattering length*) de cada átomo.

## 1.2. PRINCÍPIOS GERAIS

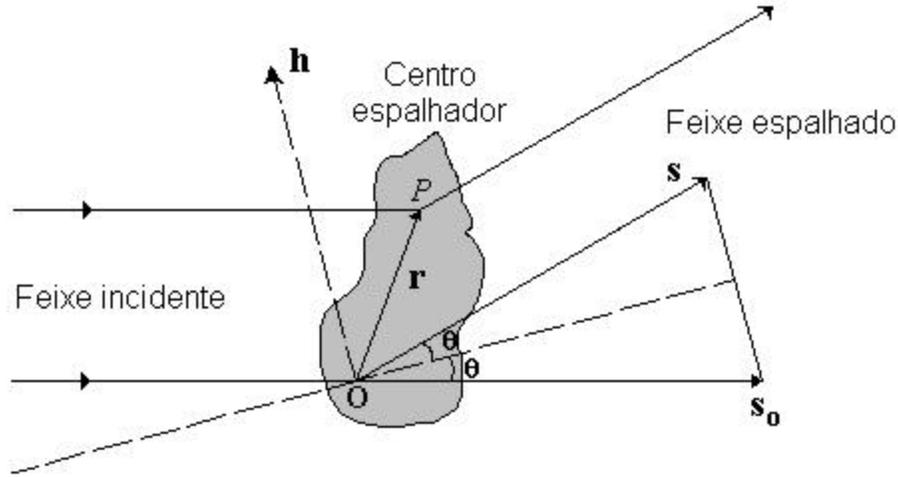
Nos concentraremos em raios X somente, mas todas as equações que serão apresentadas podem ser aplicadas, com pequenas modificações, a nêutrons ou difração de elétrons. Quando uma onda de raios X incide em um objeto, cada elétron se torna uma fonte de uma onda espalhada. Todas estas ondas possuem a mesma intensidade dada pela fórmula de Thomson

$$I_e(\theta) = I_p \cdot T_f \cdot \frac{1}{a^2} \cdot \frac{1 + \cos^2 2\theta}{2} \quad (1-1)$$

sendo  $I_p$  é a intensidade primária,  $a$  a distância do objeto e detector. O fator  $T_f$  é o quadrado do raio clássico do elétron ( $e^2/mc^2 = 7.90 \times 10^{-26} [cm^2]$ ). O ângulo de espalhamento  $2\theta$  é o ângulo entre a onda primária e a onda espalhada. O último termo em (1-1) é o fator de polarização que é praticamente igual a 1 para todos os problemas que trataremos.  $I_e$  deveria aparecer em todas as equações seqüentes mas iremos omiti-lo, ou seja, a amplitude da onda espalhada por um elétron será tida como 1.  $I_e$  somente é necessário nos casos em que temos interesse na intensidade absoluta.

As amplitudes diferem somente pela fase  $\phi$ , que depende da posição dos elétrons no espaço. Espalhamento incoerente (Compton) pode ser desprezado para espalhamento de raios X a baixo ângulo. A fase  $\phi$  é  $2\pi/\lambda$  vezes a diferença entre o caminho ótico percorrido pela onda espalhada e uma onda arbitrária de referência. A direção do feixe incidente é definida pelo vetor unitário  $s_0$  e o feixe espalhado por  $s$ . O ângulo entre estes dois vetores unitários (ângulo de espalhamento) é  $2\theta$ . A diferença de fase entre os raios em um ponto P e uma origem arbitrária O é  $-\vec{r}(\vec{s} - \vec{s}_0)$ . A fase é dada por  $\phi = -\vec{h} \cdot \vec{r}$  se definimos o vetor de espalhamento  $\vec{h}$  como

$$\vec{h} = \frac{2\pi}{\lambda}(\vec{s} - \vec{s}_0) \quad (1-2)$$



**Figura 1-1:** Diferença de fase entre raios incidentes e espalhados

Este vetor bissecciona o ângulo entre o feixe espalhado e o incidente e tem como módulo  $h = \frac{4\pi}{\lambda} \sin \theta$ . Para espalhamento de raios X a baixos ângulos, podemos substituir  $\sin \theta$  por  $\theta$ . Agora introduzimos a densidade eletrônica  $\rho(\vec{r})$ . Esta função informa o número de elétrons por unidade de volume na posição  $\vec{r}$ . Um elemento de volume  $dV$  em  $\vec{r}$  contém  $\rho(\vec{r})dV$  elétrons. A amplitude de espalhamento de todo o volume irradiado é dada por

$$A(\vec{h}) = \iiint \rho(\vec{r}) e^{-i\vec{h} \cdot \vec{r}} dV \quad (1-3)$$

Vemos que a amplitude  $A$  é a transformada de Fourier da distribuição de densidade eletrônica  $\rho$ . A intensidade  $I(\vec{h})$  da amplitude complexa  $A(\vec{h})$  é o quadrado absoluto dado pelo produto da amplitude pelo seu complexo conjugado  $A^*$

$$I(\vec{h}) = A(\vec{h}) A(\vec{h})^* = \iiint \tilde{\rho}^2(\vec{r}) e^{-i\vec{h} \cdot \vec{r}} dV \quad (1-4)$$

onde  $\tilde{\rho}^2(\vec{r})$  é a função de convolução quadrática (Bracewell, 1965).

$$\tilde{\rho}^2(\vec{r}) = \iiint \rho(\vec{r}_1) \rho(\vec{r}_1 - \vec{r}) dV_1 \quad (1-5)$$

A distribuição de intensidades em  $\vec{h}$  (espaço recíproco) é unicamente determinada pela estrutura no espaço real.

Até agora discutimos o processo de espalhamento de uma partícula em uma orientação fixa no vácuo. Na maioria dos casos de espalhamento a baixos ângulos, as seguintes situações ocorrem:

- Os espalhadores (partículas ou inhomogeneidades) são estatisticamente isotrópicos e não existe ordenação a longo alcance, isto é, não existe correlação entre pontos em grandes distâncias espaciais.

- Os espalhadores estão imersos em uma matriz. A matriz é considerada como um meio homogêneo com densidade eletrônica  $\rho_0$ . Esta situação ocorre para partículas em solução ou para inhomogeneidades em um sólido. A densidade eletrônica nas equações (1-3 – 1-5) deve ser substituída pela diferença em densidade eletrônica  $\Delta\rho = \rho - \rho_0$ , o qual pode ter valores positivos e negativos.

A média sobre todas as orientações  $\langle \rangle$  fornece

$$\langle e^{-i\vec{h}\cdot\vec{r}} \rangle = \frac{\text{sen}(hr)}{hr} \quad (1-6)$$

e a equação (1-4) se reduz a forma

$$I(h) = 4\pi \int_0^\infty r^2 \Delta\tilde{\rho}^2(r) \frac{\text{sen}(hr)}{hr} dr \quad (1-7)$$

ou, com

$$p(r) = r^2 \Delta\tilde{\rho}^2(r) = r^2 V\gamma(r) \quad (1-8)$$

teremos

$$I(h) = 4\pi \int_0^\infty p(r) \frac{\text{sen}(hr)}{hr} dr \quad (1-9)$$

$\gamma(r)$  é a chamada função de correlação, ou função característica. A função  $p(r)$  é chamada função distribuição de pares de distâncias PDDF (Guinier & Fournet, 1955; Glatter, 1979). A transformação inversa para (1-9) é dada por

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty I(h) hr \text{sen}(hr) dh \quad (1-10)$$

ou por

$$V\gamma(r) = \frac{1}{2\pi^2} \int_0^\infty I(h) h^2 \frac{\text{sen}(hr)}{hr} dh \quad (1-11)$$

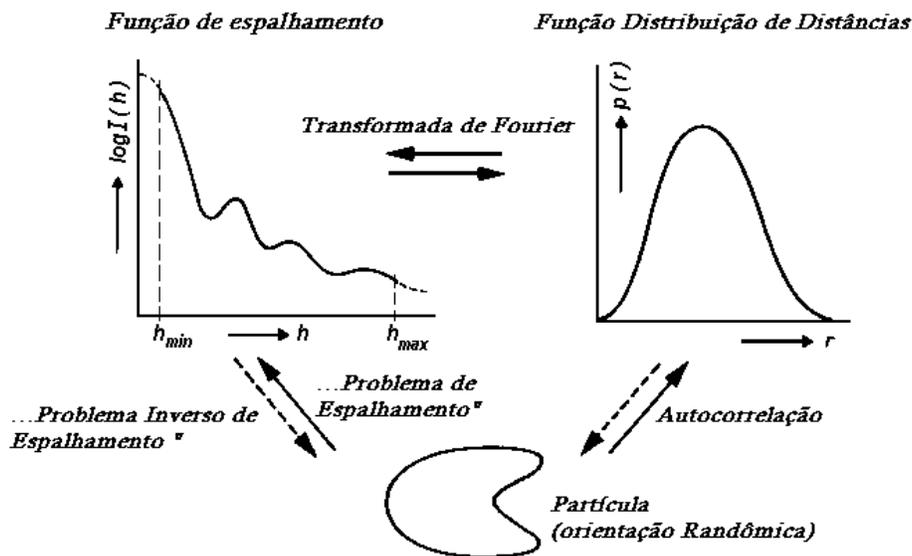
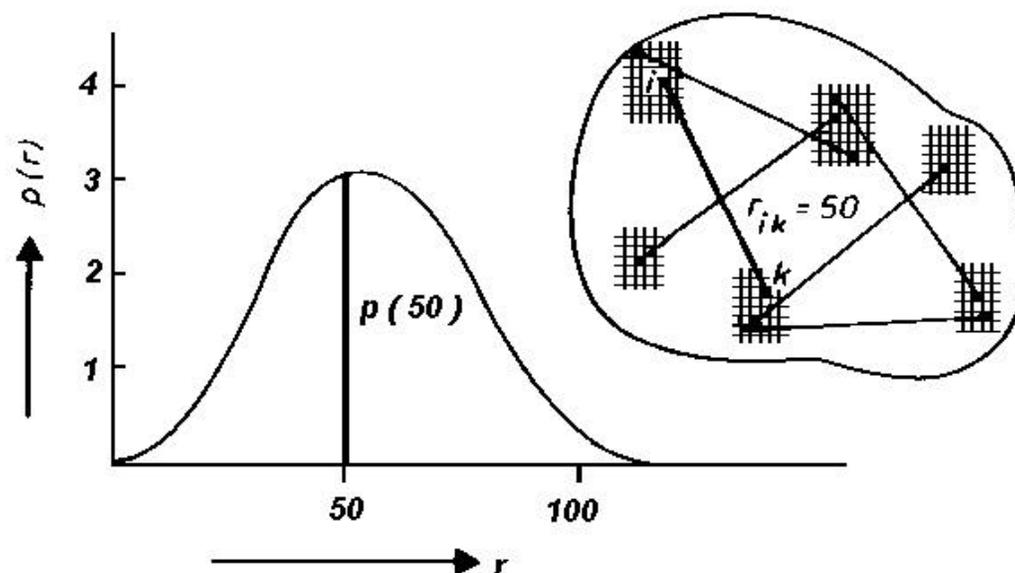


Figura 1-2: O problema inverso de espalhamento ( figura baseada de ref. [41])

A função  $p(r)$  é diretamente conectada com a intensidade de espalhamento mensurável e é muito importante para a solução do problema inverso de espalhamento. Este problema é ilustrado na Figura 1-2. Partindo da partícula com estrutura tridimensional conhecida podemos calcular a correspondente PDDF e a intensidade espalhada. Este problema é o *problema de espalhamento* e pode ser solucionado sem dificuldades essenciais, analiticamente ou numericamente. O *problema inverso de espalhamento* é a tentativa para obtermos o tamanho, forma e estrutura interna da partícula a partir da intensidade de espalhamento medida. Este problema não possui solução única mas veremos na seqüência que a PDDF é de muita ajuda neste contexto. Além disso, nos capítulos seguintes apresentaremos métodos numéricos de recuperação tridimensional de estruturas partindo da curva de espalhamento.

Antes de entrar em detalhes devemos discutir as equações (1-9) e (1-10). A PDDF pode ser definida como: a função  $p(r)$  fornece a distribuição de distâncias entre pares de elétrons dentro da partícula.

Vamos subdividir a partícula em um número muito grande de elementos idênticos de volume. A função  $p(r)$  é proporcional ao número de linhas com comprimento  $r$  e  $r + dr$  que são encontradas na combinação de qualquer elemento de volume  $i$  e qualquer outro elemento de volume  $k$  da partícula (Figura 1-3). Para  $r=0$  não existe *outro* elemento de volume, assim  $p(r)$  deve ser zero, e aumenta com  $r^2$  pois o número de elementos de volume vizinhos é proporcional a superfície de uma esfera com raio  $r$ . Partindo de um ponto arbitrário na partícula existe a probabilidade de atingir-se a superfície na distância  $r$ . Isto irá fazer com que a função  $p(r)$  decresça como uma parábola e finalmente a PDDF vai a zero para todo  $r > D$  onde  $D$  é a dimensão máxima da partícula. Neste sentido a função  $p(r)$  é um *histograma das distâncias na partícula*. Não há nenhuma informação sobre a orientação destas linhas na  $p(r)$ , devido á média espacial de orientações.



**Figura 1-3:** A altura da função  $p(r)$  para um certo valor de  $r$  é proporcional ao número de todas as linhas com comprimento entre  $r$  e  $r+dr$  no interior da partícula ( figura baseada de ref. [41])

No caso de partículas inomogêneas devemos pesar cada linha com o produto do número da diferença de elétrons  $\Delta\rho.dV$  do elemento de volume. Isso pode fornecer *contribuições negativas* a PDDF.

Podemos ver da equação ( 1-9 ) que cada distância  $r$  fornece a contribuição  $\frac{\text{sen}(hr)}{hr}$  com o peso  $p(r)$  para a intensidade espalhada total.  $I(h)$  e  $p(r)$  contém a mesma informação, mas na maioria dos casos é mais fácil fazer análises em termos das distâncias que em termos das contribuições  $\frac{\text{sen}(hr)}{hr}$ . A PDDF pode ser computada exatamente com a equação ( 1-10 ) se  $I(h)$  for conhecido em *todo o espaço recíproco*.

Para  $h = 0$  temos da equação ( 1-9 )

$$I(0) = (\overline{\Delta\rho})^2 V^2 = 4\pi \int_0^{\infty} p(r) dr \quad (1-12)$$

isto é, a intensidade espalhada em  $\theta$  igual a zero é proporcional à área sobre a PDDF. Da equação (1-11) encontramos

$$V\gamma(0) = \frac{1}{2\pi^2} \int_0^{\infty} I(h)h^2 dh = V \overline{\Delta\rho^2} \quad (1-13)$$

(Porod, 1982), isto é, a integral da intensidade multiplicada por  $h^2$  está relacionada com a média quadrática das flutuações de densidade eletrônica independente da estrutura.. Podemos modificar a forma da partícula, a função de espalhamento  $I(h)$  deverá mudar consideravelmente mas a integral ( 1-13 ) deverá continuar invariante (Porod, 1951)

$$\text{Invariante } Q = \int_0^{\infty} I(h)h^2 dh \quad (1-14)$$

### 1.3. SISTEMAS MONODISPERSOS

Discutiremos o espalhamento proveniente de sistemas monodispersos, isto é, todas as partículas possuem a mesma forma e estrutura interna. Estas condições são usualmente satisfeitas por macromoléculas biológicas em solução. Ademais assumimos que estas soluções possuem diluição infinita, o que experimentalmente pode ser levado em consideração medindo-se uma série de funções de espalhamento em diferentes concentrações e extrapolando estes dados à concentração zero.

A solução é monodispersa. Assim existe somente um tipo de partícula na solução com uma distribuição eletrônica associada  $\rho(\vec{r})$  e intensidade espalhada  $i_1(h)$ . A intensidade espalhada experimental  $I(h)$  está relacionada com  $i_1(h)$  pela relação

$$I(h) = N i_1(h) \quad (1-15)$$

onde N é o número de partículas da amostra.

#### 1.3.1. PARÂMETROS ESTRUTURAIS GLOBAIS

Existem alguns parâmetros que podem ser obtidos diretamente da curva de intensidade espalhada sem a necessidade de um tratamento apurado dos dados, bastando apenas uma análise do regime assintótico.

### 1.3.1.1. INTENSIDADE NA ORIGEM

Para  $r = 0$ , utilizando as equações ( 1-5 ), ( 1-7 ) e ( 1-15 ), temos

$$i_1(0) = \int_{V_r} \int_{V_{r'}} \Delta\rho(\vec{r})\Delta\rho(\vec{r}')dV_r dV_{r'} = \Delta m^2 = m^2 - m_0^2 \quad (1-16)$$

sendo  $m$  o número de elétrons da partícula e  $m_0$  o número de elétrons correspondente ao volume de água deslocado pela molécula.

Podemos escrever a intensidade total na origem como

$$I(0) = Ni_1(0) = Nm^2(1 - \rho_0\Psi)^2 \quad (1-17)$$

sendo  $\Psi$  o volume específico (eletrônico) parcial da partícula. Geralmente utiliza-se a concentração em massa por volume (g/l)  $c = N\mu m / N_A$ , sendo  $N_A$  o número de Avogadro e  $\mu$  a razão  $M/m$  do peso molecular para o correspondente número de elétrons, o que depende da composição química da partícula (para proteínas  $M/m = 1.87$  é uma boa aproximação (Vachette & Svergun, 1998). Com isso,

$$\frac{I(0)}{c} = \frac{N_A M}{\mu} (1 - \rho_0\Psi)^2 \quad (1-17a)$$

A intensidade a ângulo zero dividida pela concentração da solução é proporcional ao peso molecular  $M$ . Se  $\Psi$ ,  $\rho_0$  e  $c$  são conhecidos, e a intensidade do feixe incidente é conhecido em escala absoluta (número de fótons durante o experimento) a intensidade na origem pode determinar o peso molecular.

### 1.3.1.2. RAI0 DE GIRO: LEI DE GUINIER

Próximo a origem, a intensidade espalhada pode ser expandida em potências de  $h^2$ ,

$$I(h) = I(0) \left[ 1 - \frac{1}{3} R_g^2 h^2 + k.h^4 + \dots \right] \quad (1-18)$$

$$I(h) \cong I(0).e^{-h^2 R_g^2 / 3} \quad (1-19)$$

com o raio de giro  $R_g$  sendo somente um parâmetro. Essa aproximação é conhecida como Lei de Guinier, e o raio de giro é definido pela expressão

$$R_g = \frac{\sqrt{\int \Delta\rho(\vec{r})r^2 dV}}{\sqrt{\int \Delta\rho(\vec{r})dV}} \quad (1-20)$$

onde a origem dos vetores é tomada no centro de massa da partícula.

Como dito anteriormente podemos calcular a intensidade espalhada (analítica ou numericamente) para qualquer sólido. As curvas de espalhamento de duas partículas possuindo o mesmo raio de giro *coincidirão* para ângulos muito pequenos, e somente no final da curva, devido aos termos de ordem  $h^4$ ,  $h^6$ , etc. teremos diferenças marcantes devido a influência da forma das partículas. Com isso, podemos suspeitar que possa existir uma aproximação universal para todos os tipos de partículas. O raio de giro de um objeto mede seu grau de (não) esfericidade. Para um dado volume, o *menor raio de giro* é o de uma esfera

$$R_g = \sqrt{\frac{3}{5}} \cdot R \quad (1-21)$$

sendo  $R$  o raio da esfera. Na prática utiliza-se uma representação linearizada, plotando-se  $\ln[I(h)]$  vs.  $h^2$  (Gráfico de Guinier). Uma simples regressão linear fornece o raio de giro da inclinação, bem como a intensidade na origem. Para o caso de partículas esféricas a aproximação de Guinier é válida até  $h \cdot R_g = 0.13$  (Guinier & Fournet, 1955).

- CILINDROS E LAMELAS

No caso de partículas muito alongadas, o raio de giro da *seção transversal* pode ser determinado, usando uma representação similar (Glatter and Kratky, 1982), plotando agora  $h \cdot I(h)$  vs.  $h^2$ ,

$$h \cdot I(h) \propto \exp\left(\frac{1}{2} R_c^2 h^2\right) \quad (1-22)$$

Para o caso de lamelas, a espessura pode ser obtida de um gráfico de  $h^2 \cdot I(h)$  vs.  $h^2$ ,

$$h^2 \cdot I(h) \propto \exp(-R_l^2 h^2) \quad \text{e } R_l = T / \sqrt{12} \quad (\text{T espessura}) \quad (1-23)$$

### 1.3.1.3. INVARIANTE DE POROD

Como derivado anteriormente para  $r = 0$

$$V\gamma(0) = \frac{1}{2\pi^2} \int_0^\infty I(h) h^2 dh = V \overline{\Delta\rho^2} \quad (1-13a)$$

$$\text{Invariante } Q = \int_0^\infty I(h) h^2 dh = 2\pi^2 V \overline{\Delta\rho^2} \quad (1-14a)$$

A integral (1-14a) é chamada de invariante de Porod, cujo valor não depende da forma da partícula (na realidade, este resultado também vale para sistemas não particulados) e é igual à média quadrática do contraste eletrônico do sistema.

### 1.3.1.4. VOLUME HIDRATADO

A intensidade  $i_1(0)$  espalhada por uma partícula pode ser escrita como

$$i_1(0) = (V \overline{\Delta\rho})^2 \quad (1-24)$$

onde  $V$  é o volume da partícula hidratada. Fazemos esta diferenciação entre *volume da partícula* e *volume da partícula hidratada* pois, particularmente para o caso de macromoléculas em solução, forma-se uma fina camada de água na superfície da partícula, o que, conseqüentemente, altera seu volume real. Se a partícula possui densidade homogênea,  $(\overline{\Delta\rho})^2 = \overline{\Delta\rho^2}$  e assim,

$$V = \sqrt{\frac{i_1(0)}{\gamma(0)}} = \frac{i_1(0)}{Q} \quad (1-25)$$

### 1.3.1.5. REGIME ASSINTÓPTICO

Se assumirmos que a partícula possui densidade eletrônica uniforme e uma interface bem definida com o solvente, pode-se demonstrar (Porod, 1951) que neste caso, o regime assintótico da intensidade espalhada é dado por

$$\lim_{h \rightarrow \infty} [i_1(h)] = (\overline{\Delta\rho})^2 \cdot \frac{2\pi}{h^4} \cdot S + B \cdot h^4 + \dots \quad (1-26)$$

Onde S é a área da interface entre soluto e solvente, enquanto que B é um termo constante que surge devido as flutuações na densidade eletrônica (*background*), bem como erros experimentais nos valores de  $i_1(h)$  para ângulos grandes. Na realidade, a determinação de S pela relação anterior, necessita que façamos experimentos utilizando intensidade absoluta, o que é muito difícil de ser feito. Para contornar esta limitação podemos utilizar o *invariante Q*

$$I(h)/Q \rightarrow \frac{1}{\pi} \cdot \frac{S}{V} \cdot \frac{1}{h^4} \Rightarrow \frac{S}{V} = \pi \cdot \lim_{h \rightarrow \infty} [I(h) \cdot h^4 / Q] \quad (1-27)$$

Assim a superfície específica pode ser determinada pelo padrão de espalhamento, sem necessitarmos de informações adicionais. A dificuldade reside no fato de que as flutuações de densidade eletrônica, bem como os erros experimentais nesta faixa angulares podem ser muito grandes. Isto geralmente limita a utilização desta lei no estudo de proteínas.

- CILINDROS E LAMELAS

Semelhantemente para o caso da aproximação de Guinier, para partículas alongadas o regime para  $h \rightarrow \infty$  não seguirá a tendência  $\approx h^{-4}$ . Como vemos na equação ( 1-22 ) surge um fator  $h$  multiplicando a intensidade espalhada, para que a mesma apresente um regime exponencial. Este fator acarretará uma inclinação final de  $\approx h^{-3}$ . A determinação da superfície específica dada pelas equações ( 1-26 ) e ( 1-27 ) permanece válida, com a diferença de que o significado de  $S/V$  neste caso pode ser substituído por *circunferência / área*.

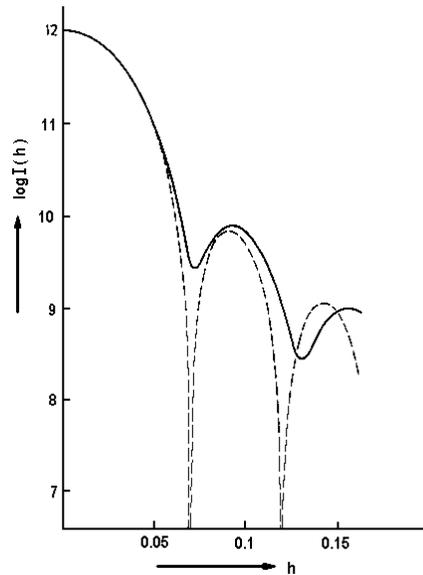
Para lamelas como vemos na equação ( 1-23 ) surge um termo  $h^2$  multiplicando a intensidade e ocasionará uma tendência dada por  $\approx h^{-2}$ . A superfície específica, como obtida por esta lei, se reduzirá simplesmente à  $1/T$ . Todas as considerações sobre flutuações na densidade eletrônica continuam valendo para estes casos.

### 1.3.2. FORMA E ESTRUTURA DAS PARTÍCULAS

Veremos agora como a forma, tamanho e estrutura das partículas espalhadoras são refletidas na função de intensidade  $I(h)$  e na função distribuição de distâncias (PDDF)  $p(r)$ . Em geral é mais simples discutir mudanças na PDDF, mas algumas características como simetria, produzem efeitos mais pronunciados no espaço recíproco. Posteriormente veremos que estas análises, são a base de um dos métodos para recuperação tridimensional de estruturas.

### 1.3.2.1. PARTÍCULAS GLOBULARES

Somente um pequeno número de problemas de espalhamento pode ser resolvido analiticamente. A partícula mais trivial é a esfera.



**Figura 1-4:** Comparação entre funções de espalhamento de uma esfera (---) e um cubo (—) com o mesmo raio de giro ( figura baseada de ref. [22])

### 1.3.2.2.

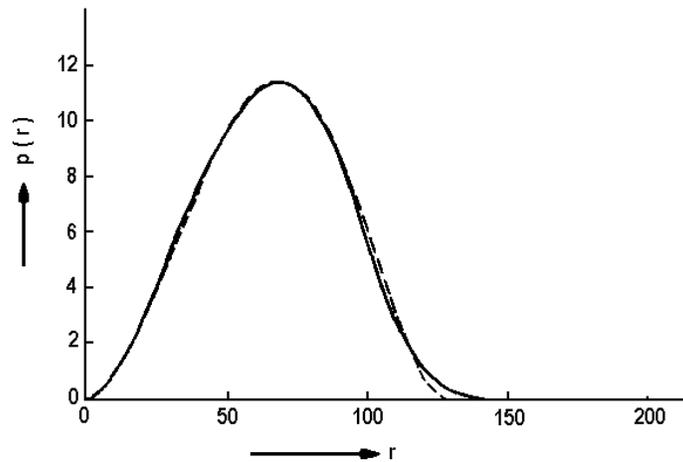
A expressão analítica normalizada da intensidade espalhada é dada por

$$I(h) = \left( 3 \cdot \frac{\text{sen}(hR) - hR \cdot \cos(hR)}{(hR)^3} \right)^2 \quad (1-28)$$

e para a função PDDF (Porod, 1946)

$$p(r) = 12x^2(2 - 3x + x^2) \quad x = r/(2R) \quad (1-29)$$

sendo  $R$  o raio da esfera. A representação gráfica da função de espalhamento é usualmente feita com um gráfico semi-log ( $\log I(h)$  vs.  $h$ ) ou com um gráfico log-log ( $\log I(h)$  vs.  $\log h$ ), a função PDDF é mostrada como um gráfico linear. A título de comparação de duas diferentes estruturas, é preferível manter-se a intensidade a ângulo zero constante (área sobre PDDF) e o raio de giro constantes.



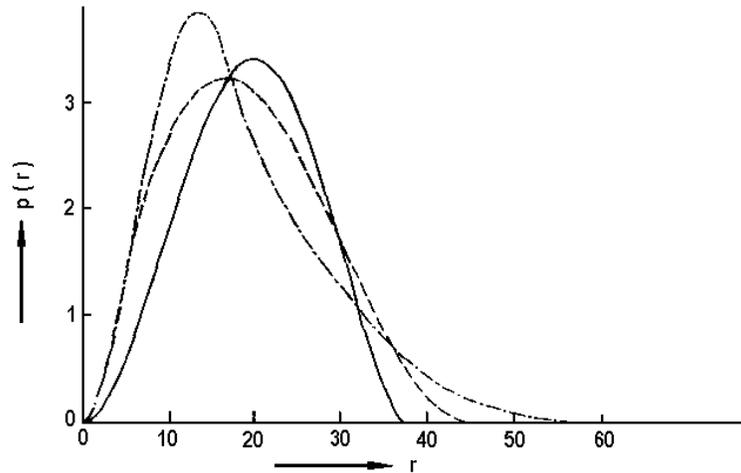
**Figura 1-5:** Funções de distribuição de distâncias para uma esfera (---) e um cubo (—) com o mesmo raio de giro e mesma intensidade na origem. (figura baseada de ref. [22])

A função de espalhamento de uma esfera com raio  $R = 65$  é mostrada na Figura 1-4 (linha tracejada,  $\log I(0)$  normalizado para 12). Vemos mínimos distintos o que são típicos para partículas com alta simetria. Podemos determinar o tamanho da esfera diretamente da posição dos zeros  $h_{01}$  e  $h_{02}$  (Glatter, 1972)

$$R \cong \frac{4.493}{h_{01}} \quad \text{ou} \quad R \cong \frac{7.725}{h_{02}} \quad (1-30)$$

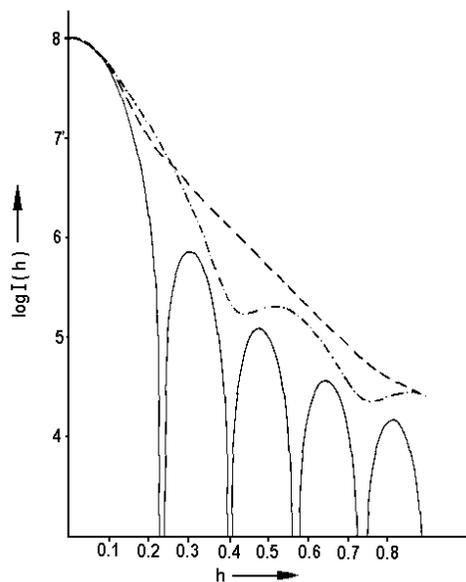
ou da posição do primeiro máximo ( $R_g \approx 4.5/h_1$ ). Os mínimos são consideravelmente atenuados no caso de cubos (linha cheia na Figura 1-4). As diferenças correspondentes no espaço real ( $p(r)$ ), não são tão evidentes (Figura 1-5).

A função  $p(r)$  para uma esfera tem um máximo próximo à  $r = R = D/2$  ( $x \cong 0.525$ ) e vai a zero, como toda PDDF, em  $r = D$ , sendo  $D$  a máxima dimensão da partícula (aqui, seu diâmetro). A função  $p(r)$  para o cubo com o mesmo  $R_g$  é zero em  $r \cong 175$ . A função é muito prolongada nesta região. Este fato demonstra a dificuldade de precisão na determinação de  $D$  quando levamos em conta erros experimentais. Qualquer variação da simetria esférica, fará com que o máximo se desloque para valores menores de  $r$  e o valor de  $D$  irá aumentar (considerando  $I(0)$  e  $R_g$  constantes). Uma comparação das funções  $p(r)$  para esfera, elipsóide oblato de revolução (razão entre eixos 1:1:0.2) e elipsóide prolato de revolução (1:1:3) é mostrado na Figura 1-6.



**Figura 1-6:** Comparação entre funções  $p(r)$  de uma esfera (—), um elipsóide prolato de revolução com relação entre eixos dada por 1:1:3 (---), e um elipsóide oblato de revolução 1:1:0.2 (----) com o mesmo raio de giro (figura baseada de ref. [22])

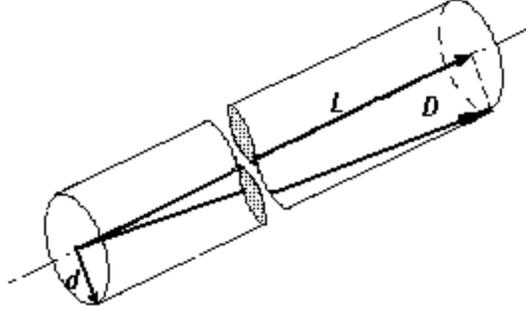
Quanto mais alteramos da forma esférica para estruturas alongadas uni ou bidimensionalmente, mais o máximo da função se desloca para valores menores de  $r$  ao mesmo tempo que temos um aumento em  $D$ . Neste sentido vemos que a função  $p(r)$  é muito informativa. Na Figura 1-7 temos as curvas teóricas de espalhamento para a esfera e os elipsóides, onde vemos uma grande alteração na forma das curvas. Uma discussão sistemática das propriedades da função  $p(r)$  não será apresentada aqui, mas pode ser obtida em vários trabalhos (Glatter, 1979; Glatter, 1982b).



**Figura 1-7:** Comparação da função de espalhamento de uma esfera, elipsóide prolato e oblato. (veja legenda da Figura 1-6) (figura baseada de ref. [22])

### 1.3.2.3. PARTÍCULAS TIPO BASTÃO

Algumas propriedades assintóticas deste tipo de partículas foram apresentadas anteriormente, mas agora analisaremos o comportamento da função  $p(r)$  para este tipo de partículas. O primeiro exemplo que vimos de partículas alongadas em uma direção (elipsóide prolato) foi mostrado nas Figuras 1-6 e 1-7. Uma classe importante são partículas alongadas em uma direção com seção transversal constante de forma arbitrária, porém constante ao longo da direção longitudinal (cilindros longos, prismas, etc). A seção transversal  $A$  (com máxima dimensão  $d$ ) pode ser pequena em comparação com o comprimento de toda a partícula  $L$ :



**Figura 1-8:** Partícula tipo bastão.

$$d \ll L \quad L = \left( D^2 - d^2 \right)^{\frac{1}{2}} \cong D \quad (1-31)$$

A curva de espalhamento de tais partículas pode ser escrita como

$$I(h) = L(\pi/h) I_c(h) \quad (1-32)$$

onde a função  $I_c(h)$  está relacionada à seção transversal e o fator  $1/h$  é característico de partículas tipo bastão (Kratky & Porod, 1948; Porod, 1982). A função de seção transversal  $I_c(h)$  é

$$I_c(h) = (L\pi)^{-1} \cdot I(h) \cdot h = \text{const} \cdot I(h) \cdot h \quad (1-33)$$

Adicionalmente temos,

$$I_c(h) = 2\pi \int_0^{\infty} p_c(r) \cdot J_0(hr) \cdot dr \quad (1-34)$$

onde  $J_0(hr)$  é uma função de Bessel de ordem zero e

$$p_c(r) = \frac{1}{2\pi} \int_0^{\infty} I_c(h) \cdot (hr) \cdot J_0(hr) \cdot dh \quad (1-35)$$

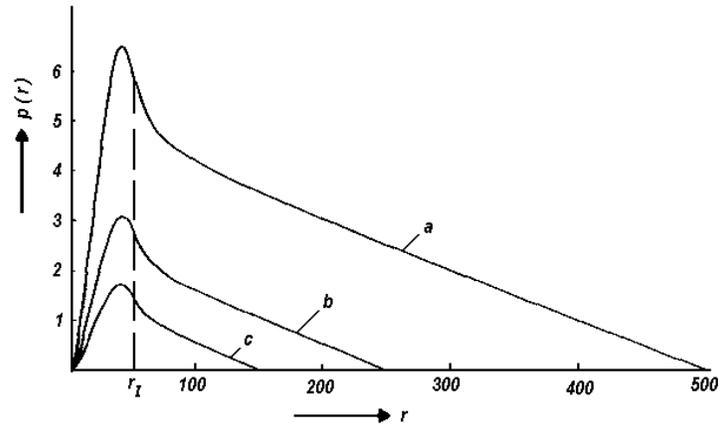
(Glatter, 1982a). A função  $p_c(r)$  é a PDDF da seção transversal com

$$p_c(r) = r \cdot \gamma_c(r) = \langle \Delta\rho(r_c) * \Delta\rho(-r_c) \rangle \quad (1-36)$$

O símbolo \* indica a operação matemática denominada convolução e  $\langle \rangle$  indica uma média espacial sobre todas as direções no *plano da seção transversal*. Partículas tipo bastão com seção transversal constante apresentam uma dependência linear decrescente na  $p(r)$  para  $r \gg d$  se  $D > 2.5d$ . A inclinação desta parte linear é proporcional ao quadrado da área da seção transversal

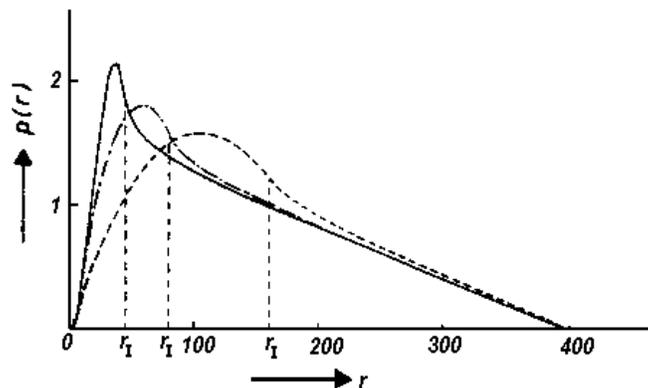
$$\frac{dp}{dr} = \frac{A^2 \Delta p^2}{2} \quad (1-37)$$

Funções PDDF de prismas com a mesma seção transversal mas diferentes comprimentos L são mostrados na Figura 1-9.



**Figura 1-9:** Funções de distribuição de distâncias de paralelepípedos homogêneos com lados: (a) 50x50x500Å; (b) 50x50x250Å; (c) 50x50x150Å. (figura baseada de ref. [22])

O máximo corresponde à seção transversal e o ponto de inflexão  $r_I$  fornece uma aproximação da máxima dimensão da seção transversal. Isto é mostrado mais claramente na Figura 1-10 onde três prismas com área de seção transversal iguais mas seções transversais de diferentes dimensões são mostrados.



**Figura 1-10:** Três paralelepípedos com comprimento constante L (400Å) e seção transversal constante, variando-se as dimensões: — 40x40Å; - - - 80x20Å; ---- 160x10Å (figura baseada de ref. [22]).

### 1.3.2.4. PARTÍCULAS TIPO DISCO

Partículas chatas, isto é, partículas alongadas em duas dimensões (discos, prismas chatos) com uma espessura constante  $T$  muito menor que a dimensão geral  $D$ , pode ser tratada em modo similar. A função de espalhamento pode ser escrita como

$$I(h) = A \cdot \frac{2\pi}{h^2} \cdot I_t(h) \quad (1-38)$$

onde  $I_t(h)$  é denominado fator de espessura (Kratky & Porod, 1948) ou

$$I_t(h) = (A \cdot 2\pi)^{-1} \cdot I(h) \cdot h^2 = \text{const} \cdot I(h) \cdot h^2 \quad (1-39)$$

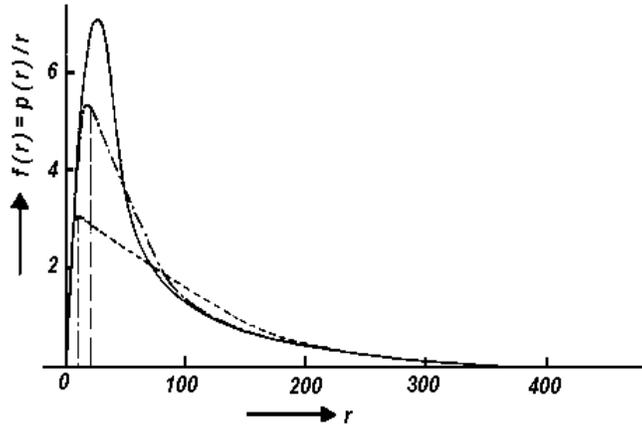
que pode ser utilizada para a determinação de  $R_t$ ,  $T$  e  $M_t$  (Raio da seção transversal, espessura e massa por unidade de espessura respectivamente). Além disso temos também

$$I_t(h) = 2 \int_0^{\infty} p_t(r) \cdot \cos(hr) \cdot dr \quad (1-40)$$

e

$$p_t(r) = \gamma_t(r) = \frac{1}{\pi} \int_0^{\infty} I_t(h) \cdot \cos(hr) \cdot dh = \Delta\rho_t(r) * \Delta\rho_t(-r) \quad (1-41)$$

Funções PDDF para partículas chatas não demonstram características claras e por isso é melhor estudar a função  $f(r) = \frac{p(r)}{r}$  (Glatter, 1979). As funções correspondentes para partículas lamelares com o mesmo plano basal, mas com diferentes espessuras são mostradas na Figura 1-11.

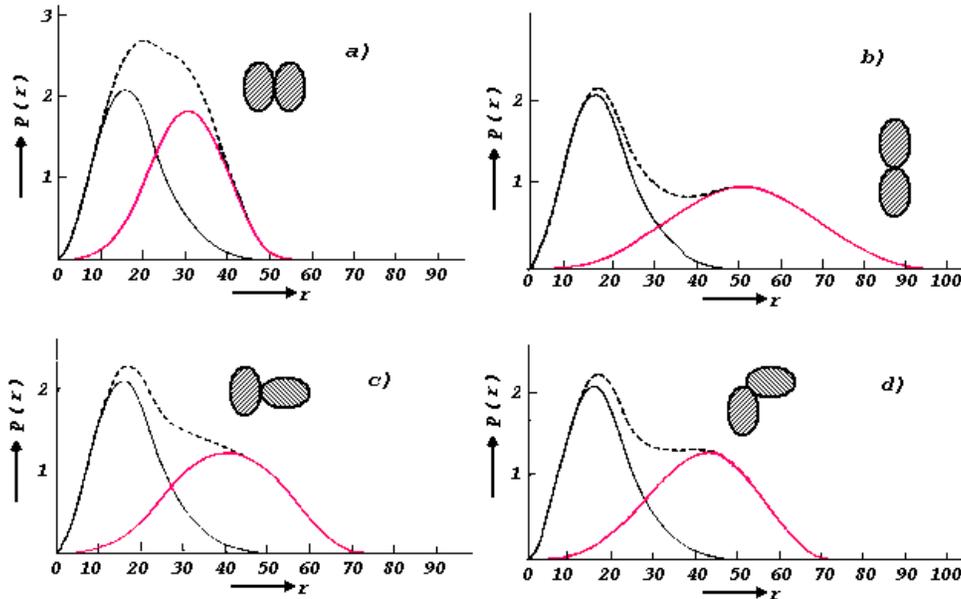


**Figura 1-11:** Função  $f(r)$  para três paralelepípedos com comprimento constante  $L$  ( $400\text{Å}$ ) e seção transversal constante, variando-se as dimensões: —  $40 \times 40\text{Å}$ ; - - - -  $80 \times 20\text{Å}$ ; ----  $160 \times 10\text{Å}$ . (figura baseada de ref. [22])

Os pontos de transição assinalados podem ser utilizados para determinar a espessura. A função PDDF da espessura  $p_t(r)$  pode fornecer mais informação em alguns casos, especialmente para partículas inhomogêneas (Glatter & Kratky, 1982).

### 1.3.2.5. PARTÍCULAS COMPOSTAS – AGREGADOS, SUBUNIDADES

Uma análise qualitativa da formação de dímeros pode ser analisado qualitativamente utilizando a função  $p(r)$  (Glatter, 1979). Porém, como apresentaremos posteriormente, a informação da existência de dímeros, possibilitará impormos restrições de simetria em nossas partículas, o que auxiliará em muito a busca pela estrutura tridimensional da partícula.



**Figura 1-12:** Funções de distribuição de distâncias para dímeros construídos a partir de elipsóides prolatos. —monômeros, --- dímeros, — diferença entre dímeros e monômeros. (figura baseada de ref. [41])

Para uma análise inicial é somente necessário conhecer-se a PDDF do monômero. Tipos diferentes de agregados fornecerão diferenças na PDDF. Agregados maiores geralmente não podem ser classificados sem ambigüidade. Informações adicionais sobre o sistema podem simplificar o problema.

Partículas constituídas de agregados com um número muito grande de subunidades idênticas mostram, em baixa resolução a estrutura geral de uma única partícula. Em ângulos maiores (alta resolução) a influência das subunidades individuais poderá ser vista. No caso especial de subunidades globulares, pode ser possível determinar-se o tamanho da subunidade a partir da posição do mínimo na função de espalhamento e a equação ( 1-30 ).

### 1.3.3. INTERFERÊNCIA INTERPARTÍCULAS E EFEITOS DE CONCENTRAÇÃO

Até agora apenas o espalhamento de partículas tem sido tratado imaginando-se, é claro, que um grande número de partículas sempre está presente. Assumimos que as intensidades somente se adicionam para fornecer o padrão de difração. Isto é verdade para sistemas bastante diluídos, mas aumentado-se a concentração aparecerão efeitos de interferência. Amostras biológicas sempre requerem altas concentrações para um sinal suficientemente mensurável da intensidade espalhada. Podemos tratar o problema de duas maneiras diferentes:

- Usamos o termo de interferência como uma informação adicional sobre o sistema em estudo analisando assim o arranjo espacial das partículas;
- Tratamos o efeito de interferência como uma perturbação de nosso conceito de partícula única e tentamos removê-lo.

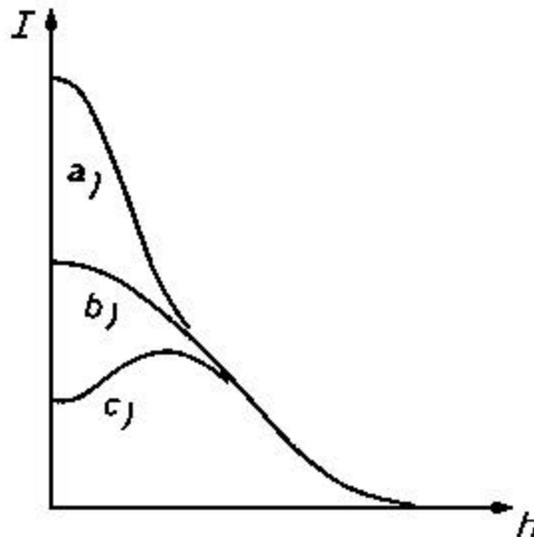
O primeiro ponto de vista é o caso mais geral porém requer muita informação sobre o sistema (tipo de agregação, mecanismos etc.). Para aplicações práticas o segundo ponto de vista é o mais utilizado.

### 1.3.3.1. A FUNÇÃO DE DISTRIBUIÇÃO RADIAL

Para obter uma descrição geral iremos nos restringir a um arranjo isotrópico e monodisperso de esferas. Isso torna possível descrever a situação introduzindo uma função de distribuição radial interpartículas  $P(r)$ . Cada partícula possui a mesma vizinhança. Nos fixaremos em uma partícula central e procuramos a probabilidade de que outra partícula seja encontrada em um elemento de volume  $dV$  a uma distância  $r$ . O valor médio é  $\left(\frac{N}{V}\right)dV$ ; qualquer variação deste valor será levado em conta pelo fator  $P(r)$ . Na faixa de impenetrabilidade ( $r < D$ ), temos  $P(r) = 0$  e em longo alcance ( $r \gg D$ )  $P(r) = 1$ . Assim a equação correspondente toma a forma,

$$I(h) = N \cdot I_1(h) \cdot \left[ 1 + \frac{N}{V} \int_0^\infty 4\pi r^2 (P(r) - 1) \frac{\text{sen}(hr)}{hr} dr \right] \quad (1-42)$$

O segundo termo contém todas as interferências interpartículas. A parte predominante é o “buraco” de raio  $D$  onde  $(P(r) - 1) = -1$ .



**Figura 1-13:** Tipos característicos das funções de espalhamento: (a) tipo-gás; (b) espalhamento por partículas; (c) tipo líquido. ( figura baseada de ref. [41])

Isto causa uma diminuição da intensidade espalhada principalmente na parte central, o que resulta em um padrão tipo líquido (Figura 1-13). Isso pode ser explicado pela redução do contraste causado pelo grande número de partículas vizinhas. Mesmo se uma distribuição de tamanhos é assumida os efeitos permanecem os mesmos. Até o presente

momento nenhuma expressão analítica para a função  $P(r)$  existe. A situação se torna mais complicada se levamos em conta interações atrativas ou repulsivas ou partículas não esféricas (efeitos de orientação).

Se tivermos um sistema de esferas com tamanho conhecido  $D$  podemos usar a equação (1-28) para  $I_1(h)$  na equação (1-42), dividir por esta função e calcular  $P(r)$  de dados experimentais por inversão de Fourier.

Se existirem forças atrativas haverá uma tendência para agregação. Esta tendência pode, por exemplo, ser introduzida em alguns passos do processo de preparação de amostras biológicas. Tal agregação tende a aumentar a intensidade na parte central (tipo gás). Neste caso teremos um sistema polidisperso de monômeros e oligômeros. Novamente não existem métodos para analisar tais sistemas de modo único.

### 1.3.3.2. ELIMINAÇÃO DOS EFEITOS DE CONCENTRAÇÃO – TIPO LÍQUIDO

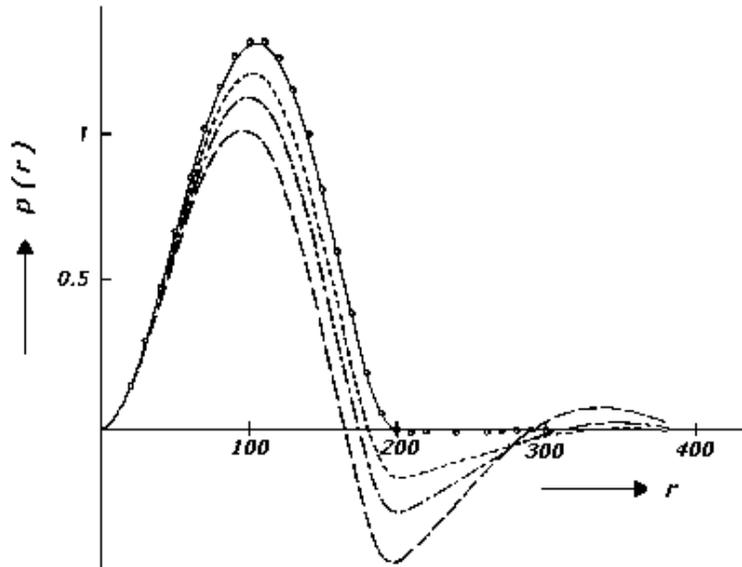
Na maioria dos casos o efeito de interferência é uma perturbação de nosso experimento, onde estamos somente interessados na função de espalhamento da partícula. Qualquer efeito de concentração persistente causará erros nos parâmetros resultantes. Como vemos na Figura 1-13 o efeito ocorre principalmente para valores de  $h$  pequenos, influenciando assim  $I(0)$ ,  $R_g$  e a função PDDF em grandes valores de  $r$ .

Podemos tratar deste problema nos sistemas tipo líquido da seguinte forma: medimos a função de espalhamento  $I(h)$  em diferentes concentrações (tipicamente de poucos  $mg/ml$  até algo em torno de  $50mg/ml$ ). A influência da concentração pode ser vista em um gráfico comum destas curvas, dividindo-se pelas respectivas concentrações. Para grandes valores de  $h$  estas curvas são idênticas. Na região de pequenos valores de  $h$  as curvas devem ser extrapoladas para concentração zero. Dependendo do problema, um ajuste linear pode ser suficiente, e em outros um polinômio de segunda ordem pode ser utilizado. A extrapolação pode ser feita em um gráfico padrão  $I(h)/c$  vs.  $h$  ou em um gráfico tipo Zimm  $(I(h)/c)^{-1}$  vs.  $h$ .

Como mencionado anteriormente, a região inicial da função de espalhamento é abaixada e com isso o raio de giro aparente diminui com o aumento da concentração. O comprimento da região linear do gráfico de Guinier pode ser aumentado pelo efeito de interferência para partículas não esféricas. Assim uma prolongada região linear no gráfico de Guinier não é uma garantia da completa eliminação dos efeitos de concentração. Interferências interpartícula restantes não podem ser reconhecidas no espaço recíproco.

A função PDDF é afetada consideravelmente por interferências interpartículas (Glatter, 1979). Ela é atenuada com o aumento de  $r$ , passando por um mínimo negativo na região da máxima dimensão  $D$  da partícula, apresentando oscilações que desaparecem para valores grandes de  $r$ . Isto é mostrado para o modelo de esferas rígidas Figura 1-14. As oscilações desaparecem quando a concentração vai a zero.

O mesmo comportamento é observado em dados experimentais mesmo no caso de partículas não esféricas.



**Figura 1-14:** Função de distribuição de distâncias para modelo de esfera rígida. Funções  $p(r)$  teóricas: —  $\theta = 0$ , - - - -  $\theta = 0.25$ , ·····  $\theta = 0.5$ , ————  $\theta = 1.0$ ,  $h_1 \cdot R = 2.0$ , 2% de erro estatístico,  $D_{max} = 300 \text{ \AA}$ ,  $\Delta R_g = 0.5\%$ ,  $\Delta \theta = 1.2\%$  (figura baseada de ref. [41]);

Em alguns casos é impossível fazer experimentos com diferentes concentrações. Isto pode ocorrer, por exemplo, se a estrutura da partícula depender da concentração. Sobre certas circunstâncias é possível obter-se os parâmetros da partícula desprezando-se a parte inicial da função de espalhamento influenciada pelo efeito de concentração (Müller & Glatter, 1982).

### 1.3.3.3. AGREGADOS – TIPO GÁS

Quando as partículas possuem tendência a se agregarem com o aumento da concentração podemos procurar os mesmos procedimentos discutidos para o tipo líquido, isto é, fazendo-se uma série de concentrações e extrapolando-se as curvas  $I(h)/c$  à concentração zero.

No entanto, na maioria dos casos a tendência de agregação existe em qualquer concentração, isto é, tanto para um sistema bastante diluído temos um certo número de oligômeros coexistindo com monômeros. Novamente não existe uma maneira única de se obter os parâmetros reais das partículas nestes casos. Não é suficiente desprezar-se a parte inicial da função de espalhamento pois isso provoca uma perda crescente da informação essencial sobre a partícula (monômero).



## 2. SIMULAÇÕES E CÁLCULO DE MODELOS

### 2.1. SIMULAÇÕES

Simulações podem ajudar na busca dos limites do método e também estimar erros sistemáticos que podem aparecer no processo de tratamento de dados. Simulações são realizadas com sistemas modelo exatamente conhecidos. Estes sistemas devem ser similares à estrutura de interesse. Os dados provenientes do modelo devem ser transformados (se necessário) de acordo com as características experimentais (perfis de colimação e distribuição de comprimentos de onda), partindo da curva PDDF teórica. “Pontos experimentais” podem ser gerados em uma região de  $h$  e também se adicionando ruído estatístico. Pode-se também, se necessário, adicionar-se *background* à função gerada.

### 2.2. CALCULO DE MODELOS

Veremos posteriormente métodos de processamento de dados experimentais. Tais métodos fornecem uma estimativa parcial da estrutura da partícula em estudo, diretamente dos dados experimentais. Uma maneira de refinar a análise consiste em comparar os resultados com funções de espalhamento e PDDF obtidas de modelos. Veremos também, que existem “algoritmos inteligentes” que podem fornecer uma estimativa da forma da partícula em questão.

#### 2.2.1. CALCULO DE INTENSIDADES DE ESPALHAMENTO

##### 2.2.1.1. MODELOS TRIAXIAIS

A amplitude de espalhamento para corpos simples triaxiais como elipsóides, prismas e cilindros podem ser calculados analiticamente para qualquer direção. A média espacial deve ser feita numericamente. O erro na integração pode ser tomado abaixo de um limite arbitrário, diminuindo-se o passo da integração. Tais cálculos semi-analíticos foram feitos em um grande número de diferentes modelos por Porod e Mittelbach, cujos trabalhos são citados em [ 22 ].

##### 2.2.1.2. MODELOS DE MUITOS CORPOS

Muitas moléculas não podem ser aproximadas satisfatoriamente por modelos simples triaxiais. No entanto, métodos semi-analíticos podem ser estendidos para construir-se corpos compostos de muitos corpos triaxiais. As amplitudes de espalhamento das subunidades individuais constituintes são calculadas analiticamente e adicionadas de acordo com sua orientação particular no modelo. A amplitude total é quadrada e sua média é feita em todas as direções.

Novamente a integração é feita numericamente. Modelos compostos somente de subunidades esféricas podem ser calculados utilizando-se a fórmula de Debye (Debye,1915; Glatter,1980b):

$$I(h) = \sum_{i=1}^n g_i^2 \phi_i^2(h) + 2 \cdot \sum_{i=1}^{n-1} \sum_{k=i+1}^n g_i g_k \phi_i(h) \phi_k(h) \frac{\text{sen}(d_{ik} \cdot h)}{d_{ik} \cdot h} \quad (2-1)$$

onde  $h = \frac{4\pi \sin \theta}{\lambda}$ ,  $\lambda$  o comprimento de onda utilizado e  $2\theta$  o ângulo de espalhamento.  $g_i$  é o peso da  $i$ -ésima esfera,  $\phi_i(h)$  é o fator de forma da  $i$ -ésima esfera e  $d_{ik}$  é a distância entre os centros da  $i$ -ésima e da  $k$ -ésima esfera.

$$g_i = \rho_i V_i = \rho_i \frac{4\pi}{3} R_i^3 \quad (2-2)$$

sendo  $\rho_i$  a densidade eletrônica da  $i$ -ésima esfera (de raio  $R_i$ ) e  $V_i$  o volume desta esfera. O fator de forma  $\phi_i(h)$  é dado por,

$$\phi_i(h) = 3 \frac{\text{sen}(R_i h) - (R_i h) \cos(R_i h)}{(R_i h)^3} \quad (2-3)$$

### 2.2.1.3. MÉTODO DOS ELEMENTOS FINITOS

Qualquer modelo pode ser aproximado por um número finito de elementos homogêneos muito pequenos, de densidade eletrônica variável. Os elementos devem ser menores que o menor detalhe estrutural de interesse.

Existem alguns métodos utilizados para o cálculo. Citaremos aqui apenas dois:

- (a) *Método de esferas*: Neste método, os elementos consistem de esferas de tamanho igual. O diâmetro destas esferas deve ser escolhido independentemente da distância entre primeiros vizinhos, de modo que o volume do modelo seja representado exatamente pela soma dos volumes de todos os elementos (isso corresponde a um pequeno *overlap* entre esferas adjacentes). A intensidade de espalhamento é calculada utilizando-se a fórmula de Debye (2-1), com  $\phi_i(h) = \phi_k(h) = \phi(h)$ .

O modelo de esferas pode ser utilizado na computação de curvas de espalhamento para macromoléculas em solução de um cristal cuja estrutura é conhecida. Os pesos dos átomos são dados por:

$$Z_{eff} = Z - \rho_0 V_{eff} \quad (2-4)$$

onde  $V_{eff}$  é o volume aparente de um átomo, dado por Langridge et al. (1960) [22, p160]. Consiste em uma aproximação grosseira da influência do solvente.

- (b) *Método de cubos*: Este método desenvolvido independentemente por Fedorov et al. (1972, 1974a, b) e por Ninio e Luzzati (1972), faz a computação de intensidades espalhadas por macromoléculas em solução cuja estrutura eletrônica é conhecida. Neste método a macromolécula é colocada mentalmente em um paralelepípedo, o qual é subdividido em pequenos cubos (com lados de 1-1.5Å). Cada cubo é examinado para decidir se está dentro da macromolécula, ou na região de solvente. Após a consideração de todos os cubos, é possível representar o volume da partícula dissolvida com grande precisão. O conteúdo matemático deste método é descrito em detalhe por Fedorov et al. (1974a,b) [22, p161].

#### 2.2.1.4. CÁLCULO DA FUNÇÃO DISTRIBUIÇÃO DE DISTÂNCIAS PARA MODELOS

A função distribuição de distâncias  $p(r)$  pode ser calculada analiticamente somente para um pequeno número de modelos simples, e em todos os outros casos devemos utilizar o método de elementos finitos com esferas. É possível definir uma equação análoga à fórmula de Debye (equação 2-1) no espaço real (Glatter, 1980c). A função distribuição de distâncias  $p(r)$  pode ser expressa por,

$$p(r) = \sum_{i=1}^N \rho_i^2 p_0(r, R_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N \rho_i \rho_k \bar{p}(r, d_{ik}, R_i, R_k) \quad (2-5)$$

$p_0(r, R_i)$  é a função distribuição de distâncias de uma esfera de raio  $R_i$  e densidade eletrônica igual a unidade,  $\bar{p}(r, d_{ik}, R_i, R_k)$  é a distribuição de distâncias cruzada entre a  $i$  ésima e a  $k$ -ésima esfera (raios  $R_i$  e  $R_k$ ) com distância mutua  $d_{ik}$ .

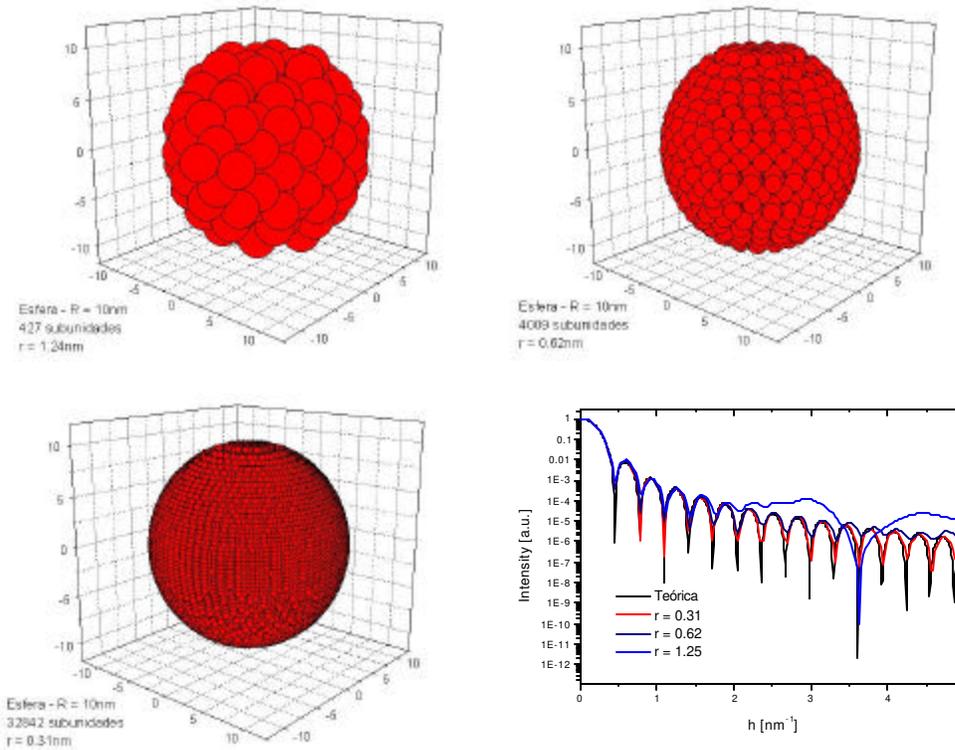
As equações ( 2-5 ) (e ( 2-1 )) podem ser utilizadas de duas maneiras diferentes para o cálculo funções modelo. Algumas vezes é possível aproximar a macromolécula a um agregado de algumas esferas de tamanho bem definido representando *diferentes* subunidades globulares. O fator de forma das subunidades é nestes casos parâmetro real do modelo. No entanto, na maioria dos casos não temos essa possibilidade e necessitamos utilizar o método de elementos finitos, ou seja, ajustamos nosso modelo com um grande número de esferas suficientemente pequenas e de igual tamanho e, se necessário, pesos diferentes. O fator de forma destas pequenas esferas não é agora um parâmetro, introduzindo assim um limite na resolução.

Podemos utilizar a Transformada de Fourier (equação (1-10)) pode ser utilizada no cálculo da função distribuição de distâncias de um modelo arbitrário se a função de espalhamento deste modelo é conhecida em uma faixa suficientemente grande do espaço recíproco.

### 2.2.2. CONSTRUÇÃO DE MODELOS

Para o cálculo numérico das curvas de intensidade de modelos usando elementos finitos, existem diferentes rotinas e métodos. Dentre várias, utilizarei o pacote denominado *MULTIBODY* (Glatter 1980b). Fornecendo para o software as coordenadas das subunidades esféricas, este software resolve as somas de ( 2-5 ) e ( 2-1 ) e com isso fornece a curva de espalhamento e  $p(r)$  para um sistema *monodisperso* da partícula construída usando as subunidades esféricas. Maiores informações sobre limitações deste programa podem ser obtidas na referência citada. No entanto um dos principais fatores que temos de ter conta é a resolução que desejamos ter em nosso modelo. O diâmetro da subunidade que iremos utilizar no modelo, deverá ser menor do que o menor detalhe que queremos representar. Sendo assim para termos modelos muito precisos, pagamos o preço de utilizar muitas subunidades.

A figura abaixo demonstra essa situação. Nosso objetivo era simular o espalhamento de uma esfera. Para isso construímos 3 esferas com o mesmo raio só que utilizando diferentes diâmetros para as subunidades.



**Figura 2-1:** Exemplo representativo da dependência entre o raio da subunidade e o máximo valor da função de espalhamento no espaço recíproco. Os modelos usados aqui foram construídos com o programa *BUILDER*

Para a região de baixos ângulos todas as três esferas apresentam a mesma curva de espalhamento. A diferença surge para ângulos maiores. Como veremos posteriormente, para modelos deste tipo, nossa resolução no espaço recíproco está restrita ao valor  $2\rho / d$  sendo  $d$  a distância entre os centros das subunidades. É interessante notar que diminuindo-se o raio da subunidade esférica de 4 vezes, aumentamos o número de esferas a utilizar em aproximadamente 76 vezes. Isso ocorre pois o número de esferas aumenta com  $1/r^3$ .

Outro fator importante diz respeito ao *empacotamento* do modelo. Como estamos construindo um modelo utilizando *subunidades esféricas*, se simplesmente *tangenciarmos* as esferinhas umas nas outras não estaremos mantendo o *volume* do modelo idêntico ao que queremos modelar. Uma solução para isso consiste em fazer com que as esferas estejam levemente superpostas. Para isso fazemos o diâmetro das subunidades um pouco maior que a distância  $d$  entre seus centros. Se tivermos um arranjo cúbico das esferas, teremos uma relação para o raio das esferas ( $R$ ) e a distância entre os centros das subunidades ( $d$ ):

$$R = \left( \frac{3 \cdot d^3}{4 \cdot \pi} \right)^{\frac{1}{3}} \cong 0,62035 \cdot d \quad (2-6)$$

Podemos compreender este último resultado como um *vínculo de homogeneidade*. Esferas arranjadas segundo esta distância característica ( $d$ ) denotarão um meio único, ao menos até o limite da resolução. A resolução do método está ligada ao fato de que as subunidades esféricas apresentam zeros no fator de forma. A curva de espalhamento do

modelo como um todo estará errada naquela região, isto é, a relação ( 1-30 ) pode ser aplicada denotando o máximo valor de h,

$$h_{\max} < \frac{4.5}{R} \quad ( 2-7 )$$

Outro problema que pode ocorrer surge de um arranjo alongado das subunidades. Tal arranjo pode ser visto como uma quasi-rede para os centros das subunidades. Isto causará erros nas curvas de espalhamento para os correspondentes valores de Bragg. Podemos impedir estes erros se a relação abaixo é satisfeita,

$$h_{\max} < \frac{2.\pi}{d} \quad ( 2-8 )$$

sendo d o arranjo regular entre os centros das subunidades. Tal arranjo pode gerar variações agudas com periodicidade d na função distribuição de distâncias. Na prática se as relações ( 2-6 ) e ( 2-8 ) são satisfeitas, ( 2-7 ) será automaticamente satisfeita também.

Deste modo vemos que a construção de um modelo requer alguns vínculos importantes para que a técnica de elementos finitos não distorça os resultados. Sendo assim, quanto maior o número de esferas, mais difícil será a construção do modelo, pois devemos satisfazer os vínculos necessários. Neste trabalho de tese foi desenvolvido um programa para possibilitar a construção destes modelos.



### 3. BUILDER – PROGRAMA PARA CONSTRUÇÃO DE MODELOS

Como discutido anteriormente, a construção de um modelo tridimensional composto de subunidades pode ser uma tarefa tediosa quando desejamos construir modelos sofisticados e compostos de subunidades pequenas. Modelos complexos requerem uma visão espacial muito apurada do usuário. Sendo assim, construí o programa **BUILDER**<sup>1</sup> em Pascal 7.0<sup>TM</sup> visando facilitar a construção do modelo e, além disso, permitir uma visão tridimensional do mesmo.

Trata-se de um programa interativo, que pergunta ao usuário cada passo seguinte. Possui várias ferramentas de construção sendo elas pontos, retas, círculos, prismas, elipsóides de revolução, cascas elipsoidais de revolução, cilindros de revolução e cascas elipsoidais de revolução, cabendo ao usuário apenas fornecer informações simples como do coordenadas posicionais, raios, dimensões laterais, semi-eixos etc, e o programa calcula todas as posições das subunidades, utilizando o empacotamento dado em ( 2-6 ). Depois de construído o modelo, o programa constrói o arquivo de entrada do MULTIBODY e chama este programa para que calcule a curva de espalhamento e a curva de distribuição de distância  $p(r)$ . Como saída, o programa BUILDER fornece um roteiro das atividades (nomearq.log) com todos os comandos utilizados na simulação. Um arquivo padrão *protein data Bank* (nomearq.pdb) formatado de modo que permita a visualização do modelo construído de maneira tridimensional em diversos softwares de visualização de proteínas (Rasmol, Raswin, Protein Explorer etc). O arquivo de entrada do MULTIBODY (nomearq.mul) e os arquivos de saída deste último contendo a curva de intensidade e a  $p(r)$  (nomearq.imu e nomearq.pmu) e por fim um arquivo tipo texto com todas as coordenadas das esferas utilizadas na forma (X, Y, Z) podendo ser lido em qualquer planilha eletrônica. Uma descrição sucinta do algoritmo utilizado no programa é descrita a seguir

#### 3.1. METODOLOGIA PARA CONSTRUÇÃO DE MODELOS

Apresentarei aqui em linhas gerais a rotina efetuada pelo programa BUILDER na construção dos modelos. Uma descrição detalhada do algoritmo bem como uma versão *freeware* do programa estão em fase final de elaboração [ 68 ].

Como salientado no texto, um vínculo importante na construção de modelos é a distância entre subunidades adjacentes. Se quisermos que o meio seja *contínuo* devemos manter esta distância constante e assim não surgirão aberrações na curva de espalhamento.

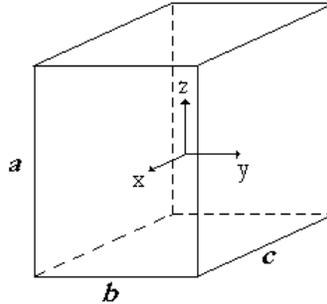
Para objetos simples como retas, círculos, etc, este vínculo pode facilmente ser satisfeito. No entanto para objetos mais complexos como esferas, elipsóides, cilindros, placas, prismas, cascas elipsoidais e cilíndricas etc, este vínculo não é simples de ser mantido, e deve ser aplicado de diferentes formas dependendo do objeto a construir.

---

<sup>1</sup> Para maiores informações sobre o programa e rotinas, contate o autor da tese.

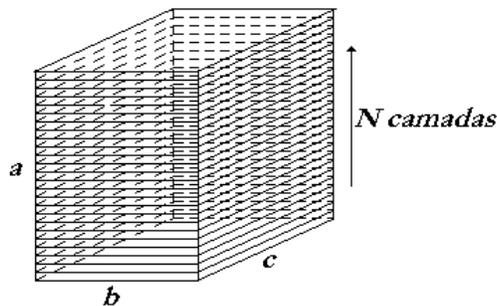
### 3.1.1. CONSTRUÇÃO DE PRISMAS

Dos sólidos citados anteriormente este é um dos mais simples de construir pelo método de elementos finitos. Consideraremos apenas prismas regulares retos. Seja um prisma reto com lados  $a$ ,  $b$ ,  $c$ ,



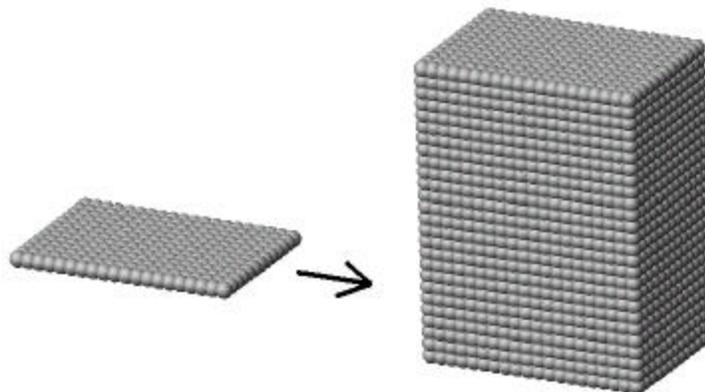
**Figura 3-1:** Prisma reto com lados  $a, b, c$  centrado na origem.

Podemos imaginar este sólido como a composição de muitas placas retangulares de lados  $b$  e  $c$ ,



**Figura 3-2A:** Seccionamento do prisma da figura anterior em  $N$  camadas.

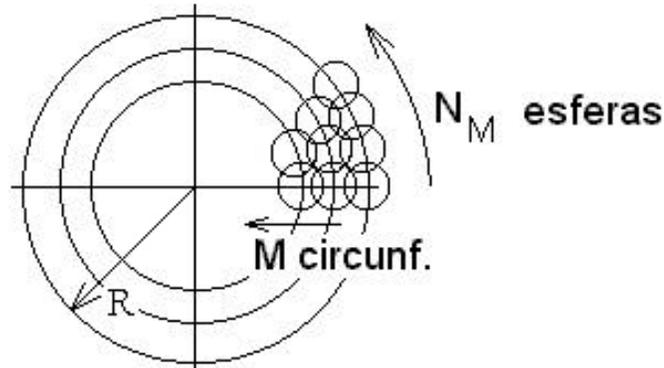
Assim tudo se resume a realizarmos  $N$  cópias de uma placa retangular. Dada a distância  $d$  entre duas subunidades adjacentes, basta discretizarmos o plano em uma matriz com o numero de linhas e colunas correspondentes a distância escolhida para as subunidades. Esta matriz é copiada  $N$  vezes, formando o sólido desejado. O processo é esquematizado abaixo,



**Figura 3-3:** Prisma construído pelo programa BUILDER. Na figura mostro uma camada em detalhe.

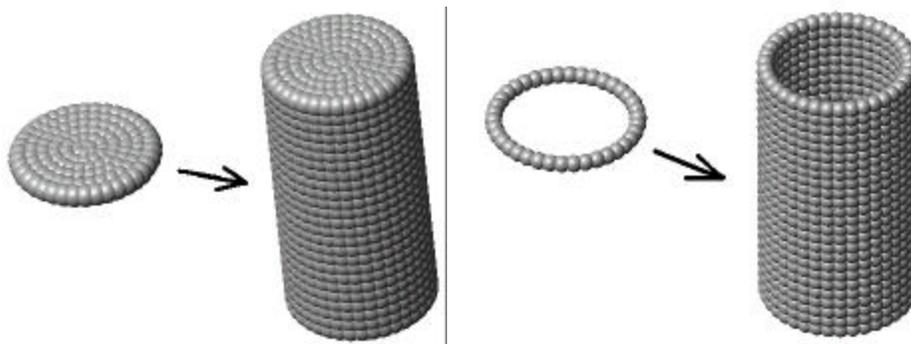
### 3.1.2. CONSTRUÇÃO DE CILINDROS E CASCAS CILÍNDRICAS

O processo descrito para a construção de prismas pode ser aplicado diretamente na construção de cilindros e cascas cilíndricas. Tudo o que precisamos é de uma base circular. Para preencher um círculo com esferas resolvi preencher o círculo de *fora para dentro*. Com isso, garanto que a dimensão radial externa seja exatamente a desejada. Dado um círculo de raio  $R$ , teremos um total de  $M$  circunferências, cada uma composta de  $N_M$  esferas.



**Figura 3-4A:** Esquema de preenchimento de um círculo com esferas.

Podendo-se construir um círculo, é direta a construção do cilindro. Basta copiarmos o número necessário para fazer a altura  $H$  desejada para o cilindro. Para construirmos uma casca cilíndrica, basta construirmos primeiramente um anel de esferas e posteriormente copiá-lo o número desejado de vezes.

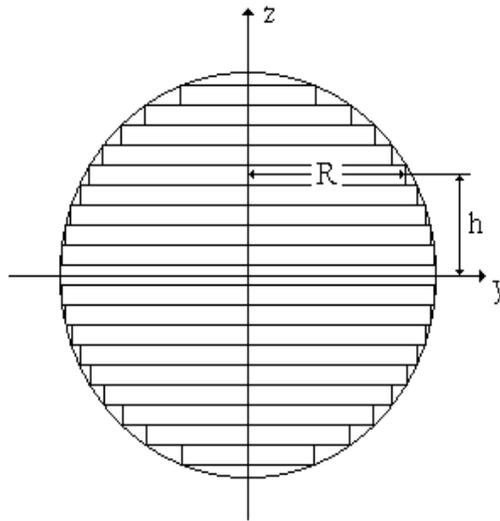


**Figura 3-5A:** Cilindro e tubo construídos com o programa BUILDER. Na figura detalho uma das camadas em cada sólido.

### 3.1.3. CONSTRUÇÃO DE ESFERAS, CASCAS ESFÉRICAS, ELIPSÓIDES DE REVOLUÇÃO E CASCAS ELIPSOIDAIIS DE REVOLUÇÃO

Temos duas situações distintas: no caso de sólidos, devemos ter a preocupação de manter o vínculo de distância entre as subunidades esféricas *tridimensionalmente*, e no caso de corpos ocos o vínculo deverá ser satisfeito *bidimensionalmente* o que faz uma diferença fundamental na construção do modelo. Para construir elipsóides de revolução (onde

também se inclui a esfera, com semi-eixos iguais), podemos proceder da mesma forma que na construção dos cilindros, isto é, como a superposição de camadas. A diferença agora, é que o raio das camadas circulares irá variar de acordo com a altura da mesma,



**Figura 3-6:** Esquema do preenchimento de um elipsóide de revolução com esferas

Dada a equação da elipse de semi-eixos  $a$  e  $b$  no plano  $YZ$ ,

$$\frac{y^2}{a^2} + \frac{z^2}{b^2} = 1 \quad (3-1)$$

Para uma altura  $h$ , o raio do prato circular será,

$$R = \sqrt{a^2 - \frac{h^2}{b^2}} \quad (3-2)$$

Quando menor o raio da subunidade esférica, mais precisa será a aproximação.

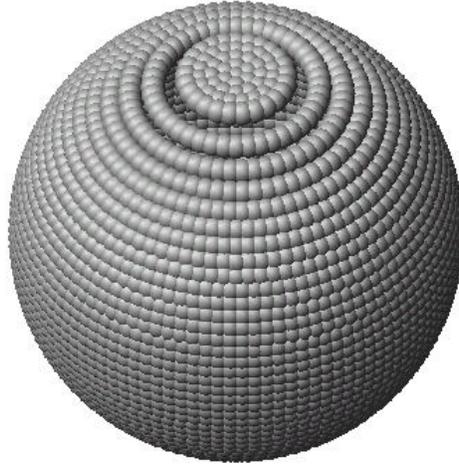


**Figura 3-7:** Esfera e elipsóides prolato e oblato construídos com o programa BUILDER.

Para assegurar que tridimensionalmente as subunidades mantenham-se a uma distância  $d$ , pagamos o preço de não ter uma ótima aproximação na superfície.

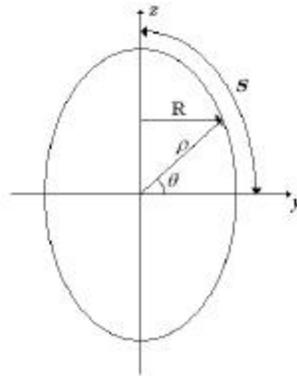
Já para o caso de cascas elipsoidais de revolução a situação é diferente. Se procedêssemos como no caso do cilindro oco, isto é, tomando apenas o círculo mais externo de cada camada, não teríamos uma uniformidade, pois, como vemos na figura

Figura 3-8, ao longo do eixo Z a aproximação fica precária para os valores próximos ao máximo valor em Z.



**Figura 3-8:** Inaplicabilidade do método anterior para construir-se cascas elipsoidais.

Dada esta dificuldade, devemos partir para outra abordagem. O problema anterior aconteceu, pois estávamos discretizando na altura, de modo que a distância entre duas placas sucessivas fosse  $d$ , e assim o vínculo de homogeneidade fosse mantido. Porém, como estamos utilizando agora apenas anéis, a distância entre anéis sucessivos deve ser  $d$ . Isso é satisfeito para valores próximos a  $z=0$ , mas não para valores próximos a  $z=b$ . A solução para essa dificuldade consiste em discretizarmos no comprimento lateral:



**Figura 3-9:** Parametrização da elipse em função do ângulo  $\theta$  e do raio  $\rho$

Como construiremos cascas de revolução, somente precisamos nos preocupar com a secção longitudinal da mesma. As secções transversais serão círculos. O total de anéis dependerá do comprimento lateral da elipse. Cada um destes anéis terá raio  $R$ , sendo  $R = \rho \cdot \cos(\theta)$ . Porém,  $\rho$  não é constante e varia conforme o ângulo  $\theta$ . Usando a forma paramétrica da elipse teremos,

$$\begin{aligned} z &= b \cdot \text{sen}(\theta) \\ y &= a \cdot \text{cos}(\theta) \end{aligned} \quad (3-3)$$

assim  $\rho$  será dado por,

$$\rho = \sqrt{z^2 + y^2} = \sqrt{b^2 \cdot \text{sen}(\theta)^2 + a^2 \cdot \text{cos}(\theta)^2} \quad (3-4)$$

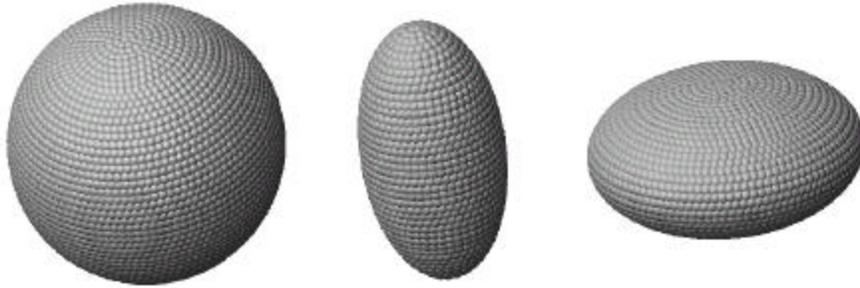
com isso, para um dado ângulo  $\theta$ , o raio do anel componente da casca será dado por,

$$R = \text{cos}(\theta) \cdot \sqrt{b^2 \cdot \text{sen}(\theta)^2 + a^2 \cdot \text{cos}(\theta)^2} \quad (3-5)$$

Para obtermos S devemos resolver a integral,

$$S = \int_0^{\pi/2} \rho \cdot d\theta = \int_0^{\pi/2} \sqrt{b^2 \cdot \text{sen}(\theta)^2 + a^2 \cdot \text{cos}(\theta)^2} \cdot d\theta \quad (3-6)$$

Utilizando as expressões anteriores, construímos os objetos a seguir,



**Figura 3-10:** Cascas esférica e elipsoidais construídas pelo programa BUILDER.

Vale lembrar, que estes objetos são “ocos”. Na realidade, se realizássemos um corte no plano XY teríamos,



**Figura 3-11:** Seccionamento das cascas da figura anterior para representação da estrutura interna.

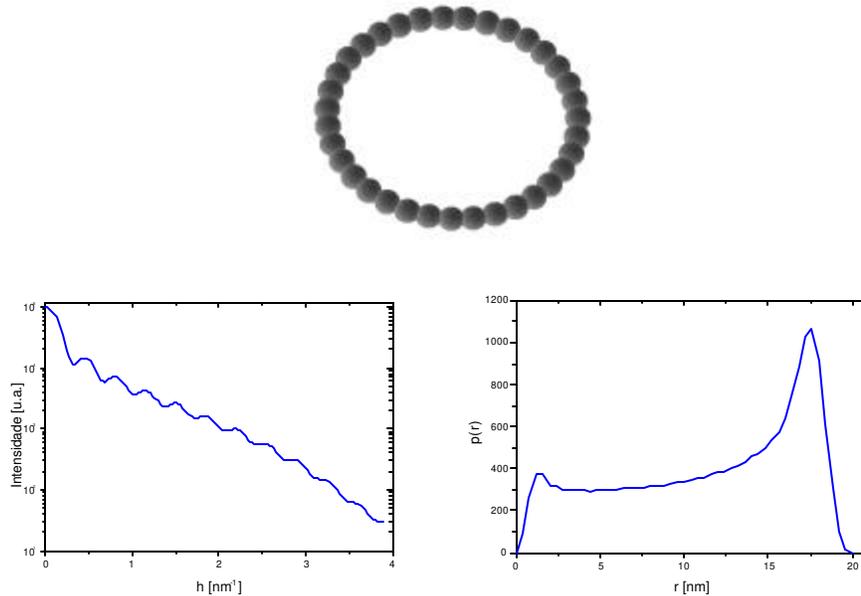
, onde vemos que a espessura da casca é exatamente igual ao diâmetro da subunidade escolhida.

Este método pode ser aplicado para a construção de *qualquer* superfície parametrizada suave, bastando apenas discretizarmos a mesma de modo a mantermos a distância entre as subunidades.

### 3.2. CÁLCULO DE INTENSIDADES DE ESPALHAMENTO PARA MODELOS DIVERSOS

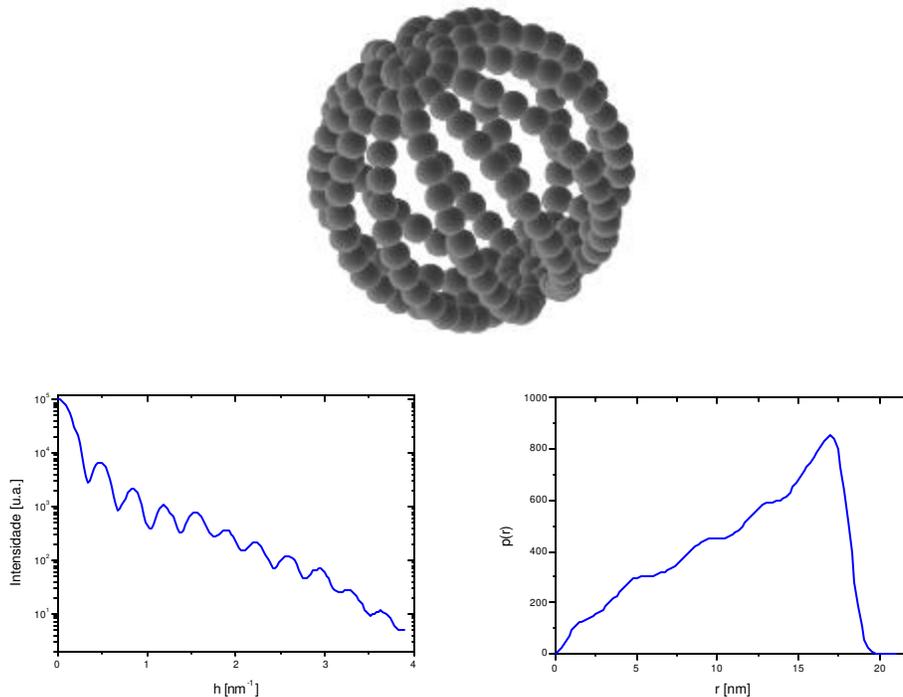
Alguns modelos construídos pelo BUILDER são mostrados abaixo, juntamente com a curva de espalhamento e  $p(r)$ :

- **Exemplo 1: Anel Circular:**



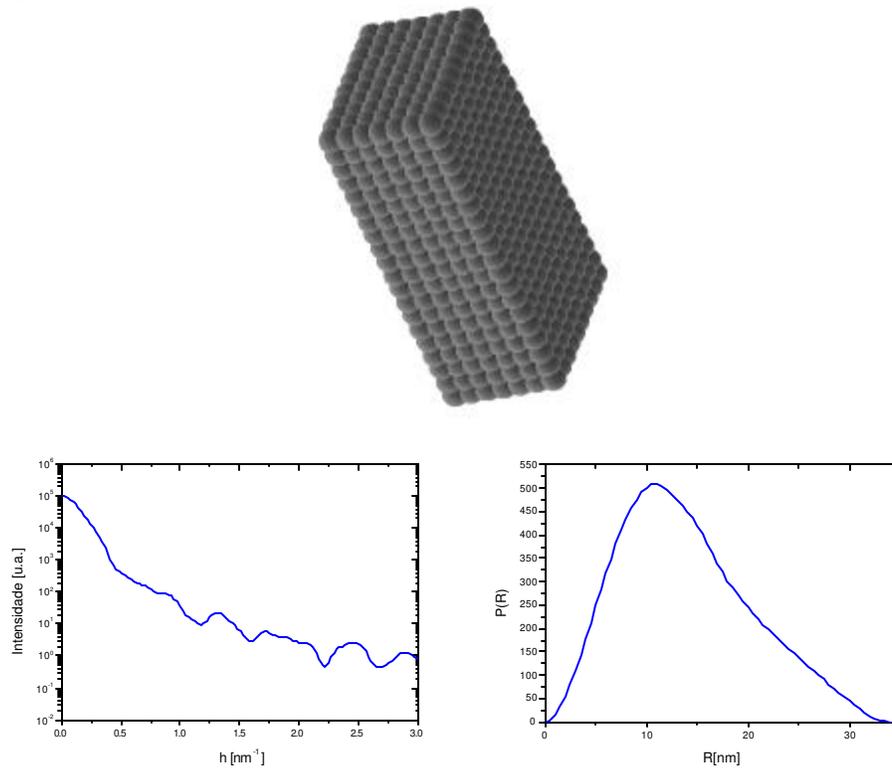
**Figura 3-12:** Modelo tridimensional, função de espalhamento e  $p(r)$  para anel circular de raio 10nm.

- **Exemplo 2: Composição de Anéis Circulares**



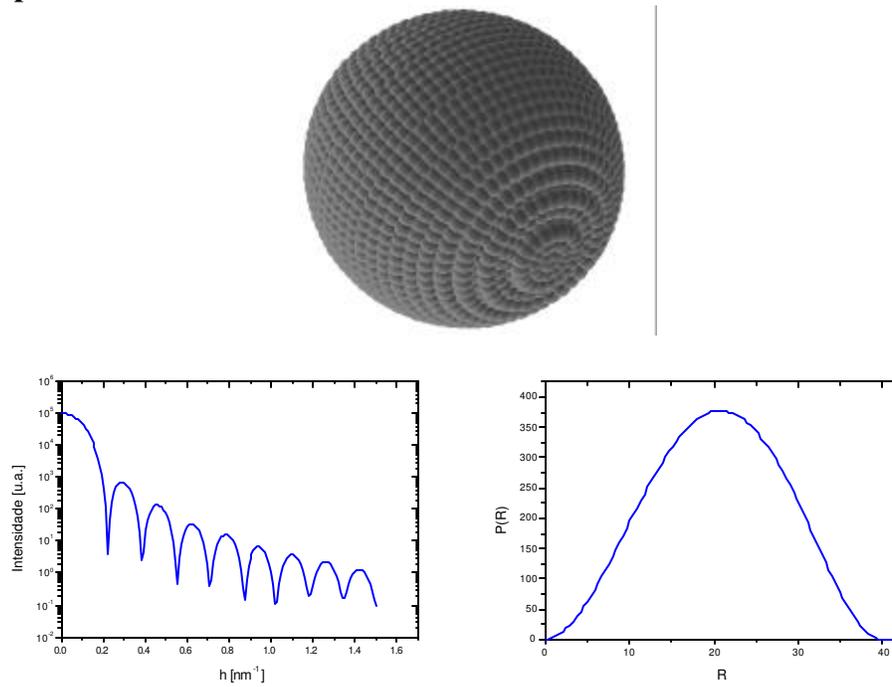
**Figura 3-13:** Modelo tridimensional, função de espalhamento e  $p(r)$  para composição de anéis circulares de raios 10nm.

- Exemplo 3: Prisma



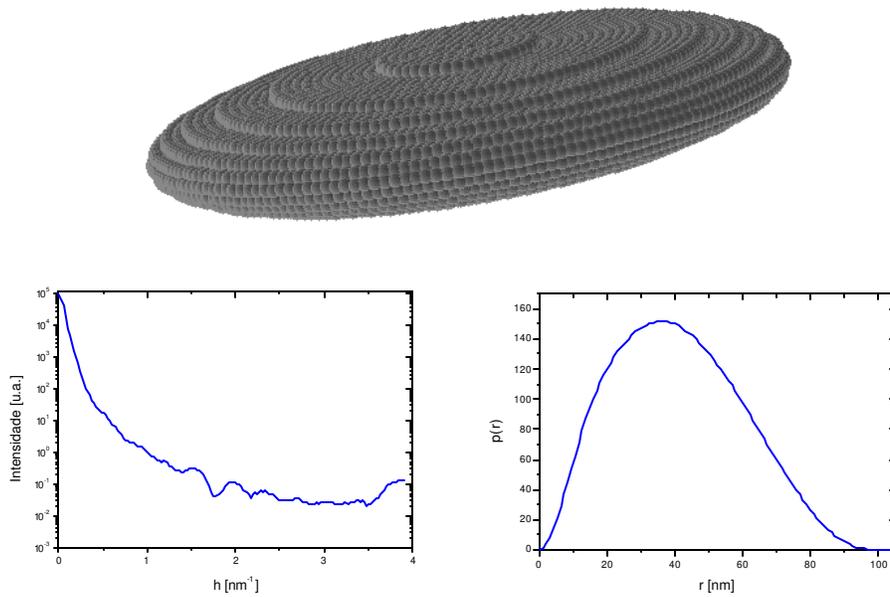
**Figura 3-14:** Modelo tridimensional, função de espalhamento e  $p(r)$  para prisma de lados  $10 \times 15 \times 30 \text{nm}$ .

- Exemplo 4: Esfera



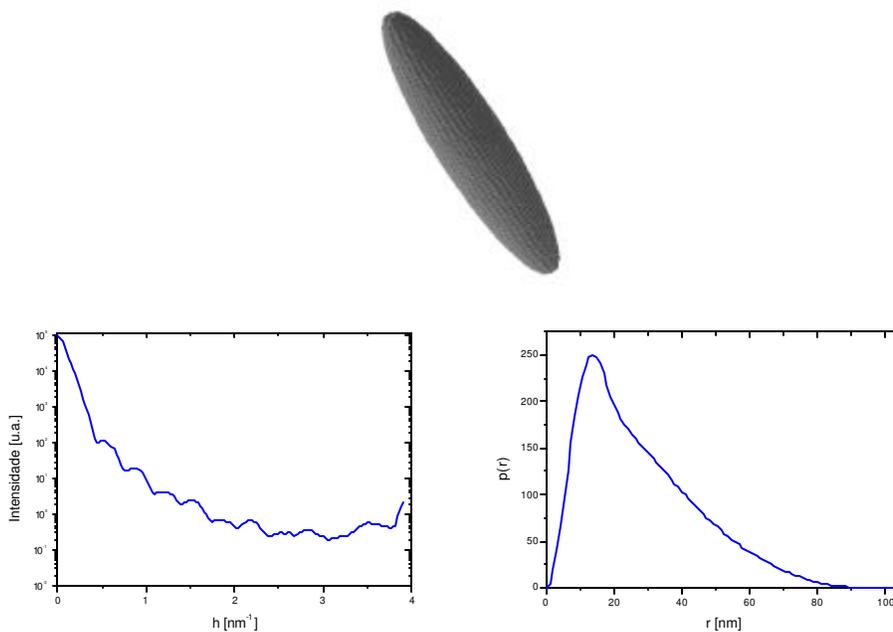
**Figura 3-15:** Modelo tridimensional, função de espalhamento e  $p(r)$  para esfera de raio 20.

- **Exemplo 5: Elipsóide Oblato**



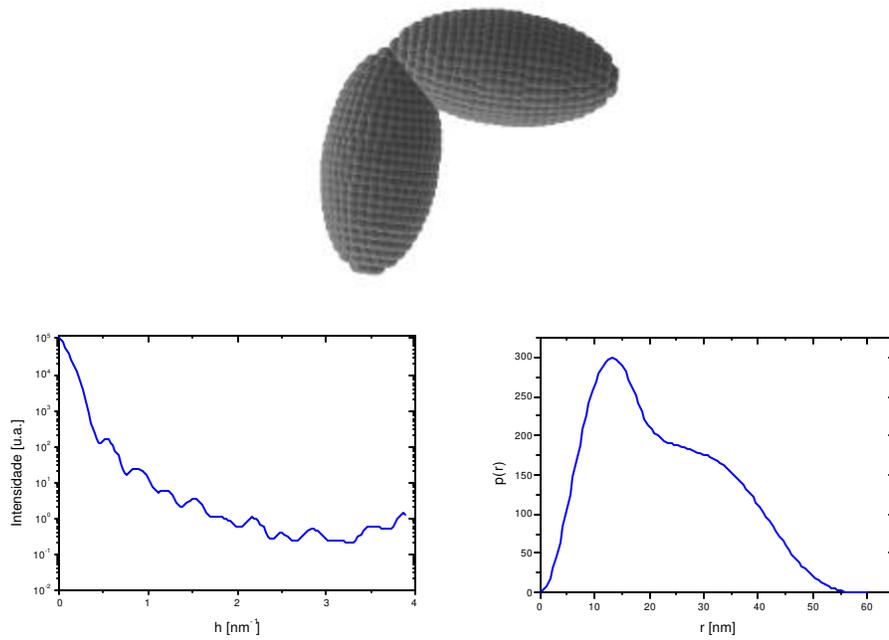
**Figura 3-16:** Modelo tridimensional, função de espalhamento e  $p(r)$  para elipsóide oblato com semi-eixos (50,50,10).

- **Exemplo 6: Elipsóide Prolato**



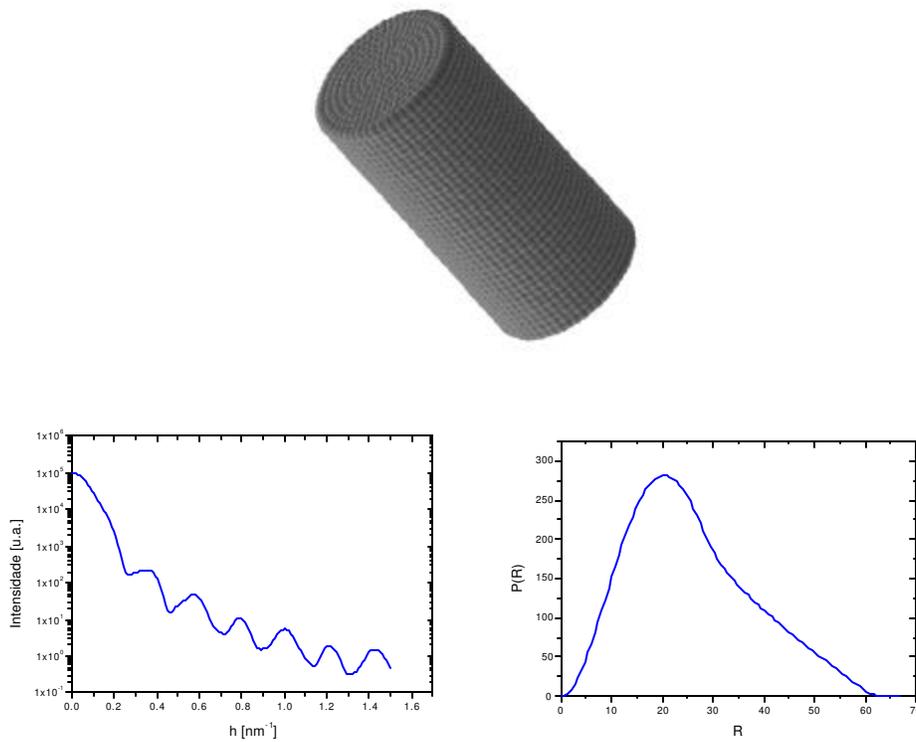
**Figura 3-17:** Modelo tridimensional, função de espalhamento e  $p(r)$  para elipsóide prolato com semi-eixos (10,10,45).

- Exemplo 7: Dímero em L



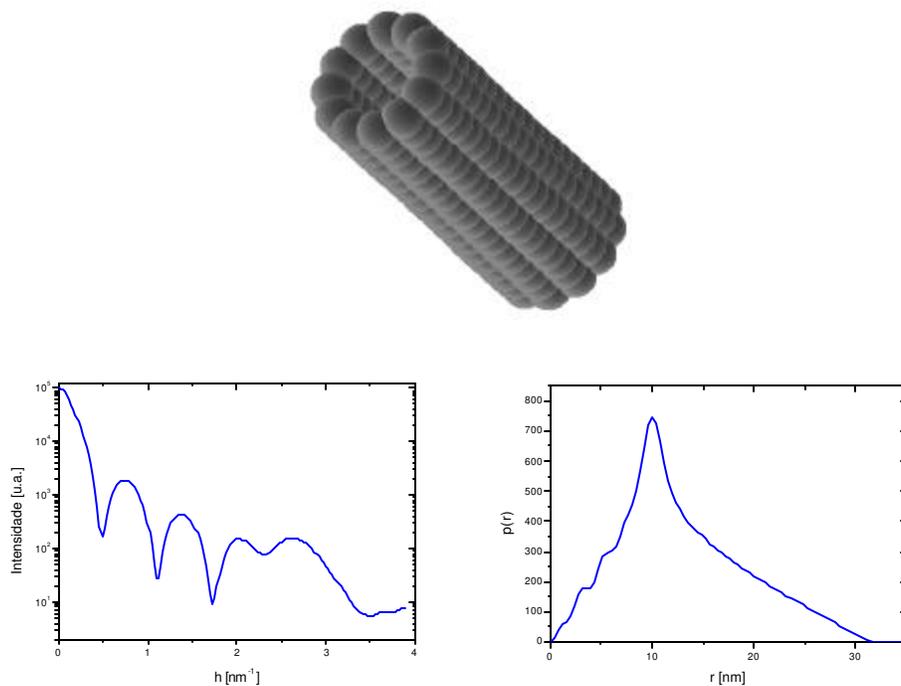
**Figura 3-18:** Modelo tridimensional, função de espalhamento e  $p(r)$  para dímeros em L. Os monômeros são elipsóides prolatos com semi-eixos (10,10,20),

- Exemplo 8: Cilindro



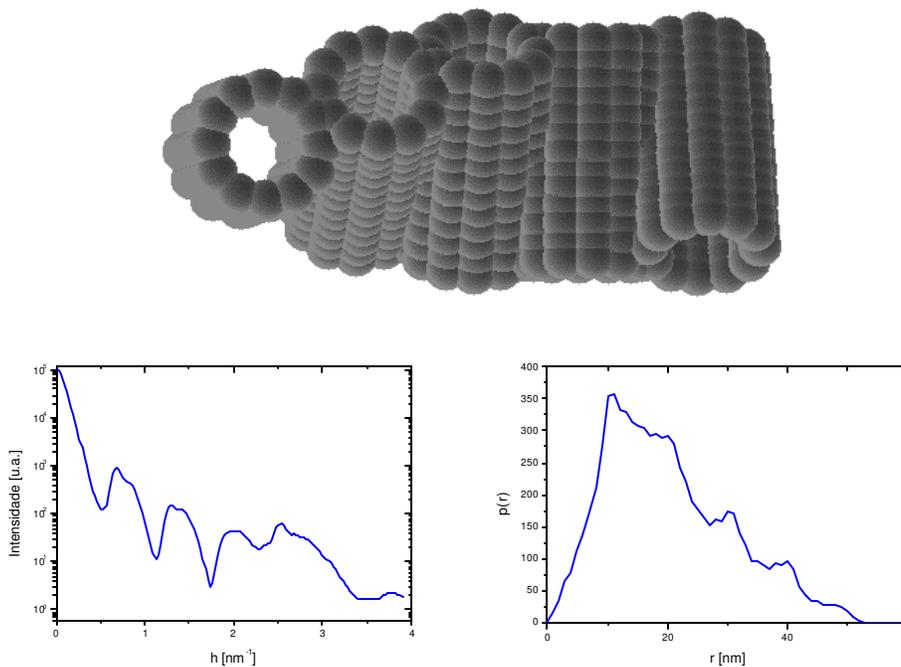
**Figura 3-19:** Modelo tridimensional, função de espalhamento e  $p(r)$  para cilindro de raio 20nm e altura 50nm.

- **Exemplo 9: Tubo**



**Figura 3-20:** Modelo tridimensional, função de espalhamento e  $p(r)$  para tubo de raio 10nm e altura 40nm.

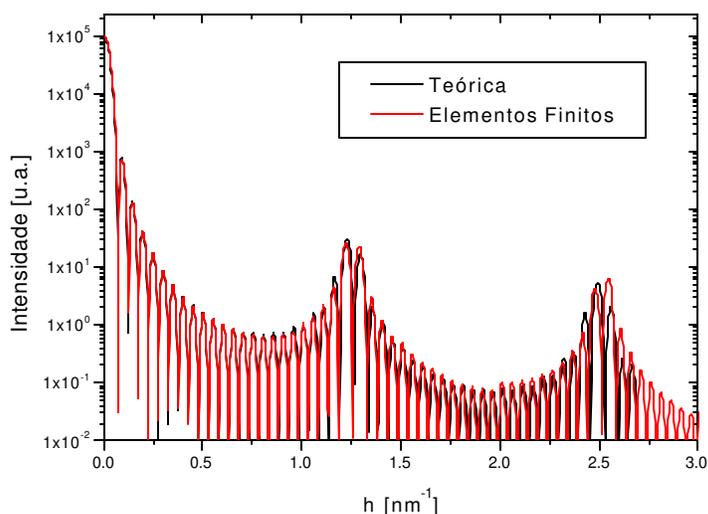
- **Exemplo 10: Arranjo de Tubos**



**Figura 3-21:** Modelo tridimensional, função de espalhamento e  $p(r)$  para arranjo de tubos de raio 10nm e altura 40nm.

Pelos exemplos anteriores vemos a utilidade da construção de modelos no estudo de espalhamento a baixos ângulos: analisando a forma do modelo e a curva de espalhamento e  $p(r)$  simuladas, criamos uma *percepção* sobre estas curvas que pode ser muito importante na análise de dados experimentais.

Para demonstrar a potencialidade do método de elementos finitos para o cálculo das curvas de espalhamento e  $p(r)$  para modelos, e do programa **BUILDER** para a construção de modelos sofisticados, podemos tentar descrever estruturas *periódicas*, como por exemplo, *multicamadas planas, cilíndricas, esféricas e elipsoidais*. Exemplos reais destas estruturas (particularmente para o caso esférico e elipsoidal) são moléculas biológicas que apresentam características hidrofóbicas e hidrofílicas. Algumas destas estruturas, quando colocadas em solução aquosa, formam *micelas* arranjadas de tal modo que a parte *hidrofóbica* não entre em contato com a água e a *hidrofílica* fique em contato. Sem entrar muito em detalhe sobre o processo de formação das micelas, formam-se cascas globulares compostas pelas partes hidrofóbicas e hidrofílicas, com uma *periodicidade intrínseca*. Esta periodicidade gerará picos semelhantes aos picos de Bragg na curva de espalhamento. A descrição teórica destas estruturas é muito complexa e, exceto para cascas esféricas, somente valores assintóticos são obtidos (Cicariello/1999). Apresento a seguir o resultado da simulação de um sistema composto por 12 cascas esféricas concêntricas espaçadas por uma distância de 5nm, com um raio central de 5 e um final de 60nm. No modelo foram utilizadas 78596 subunidades, requerendo um tempo de simulação da ordem de 30 min em um computador Pentium III - 700 MHz. Para a comparação do resultado, utilizei um programa gentilmente fornecido pelo Dr. Salvino Cicariello, que simula o espalhamento por cascas esféricas, utilizando desenvolvimento analítico.



**Figura 3-22:** Função de espalhamento para um sistema de multicamadas esféricas. —desenvolvimento analítico, —método de elementos finitos.

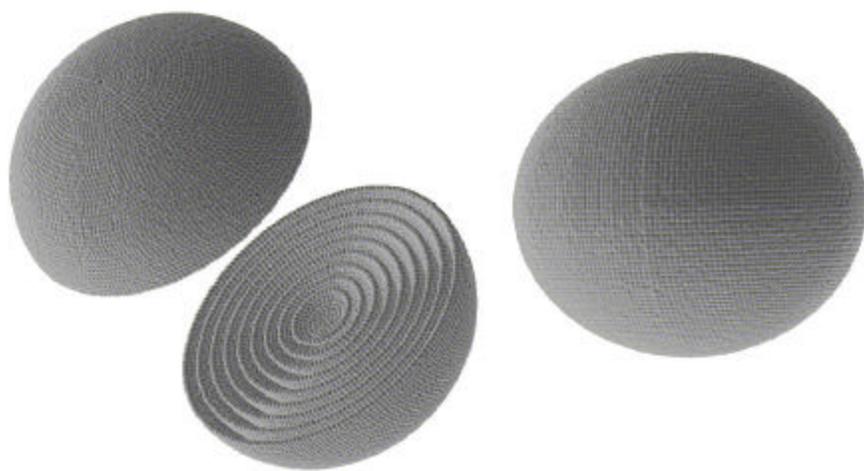
O cálculo analítico assume *meios contínuos*, ou seja, sem a composição de subunidades. No entanto vemos uma ótima concordância entre as curvas de intensidade, havendo uma variação para ângulos maiores. Isto deve-se ao mesmo fenômeno visto na

simulação de esferas (Figura 2-1) e ocorre pois aumentando-se o vetor  $h$ , aumentamos nossa resolução e com isso, a aproximação de um meio contínuo por elementos finitos se torna menos adequada.

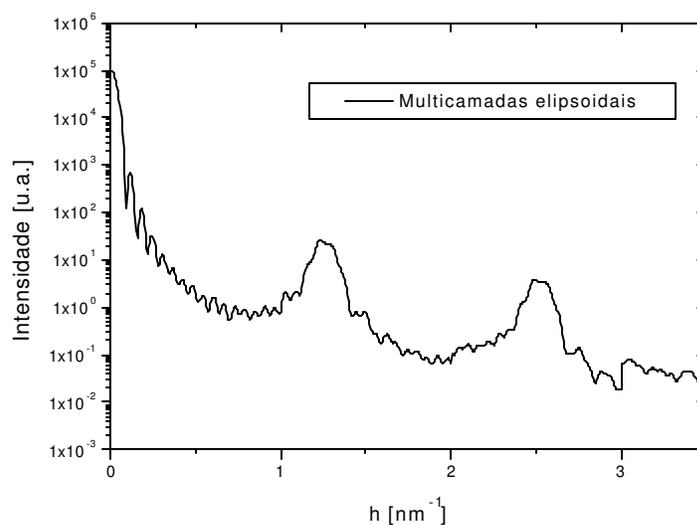


**Figura 3-23:** *Modelo de Multicamadas esféricas.*

Utilizamos multicamadas esféricas simplesmente para poder realizar uma comparação com os resultados apontados em (Cicariello/1999). Com o **BUILDER** é possível construir-se *qualquer* tipo de casca elipsoidal de revolução. Como exemplo tomamos o seguinte caso onde temos cascas elipsoidais iniciando em semi-eixos (10,10,3) para em seguida manter uma periodicidade 5 entre as cascas, totalizando 9 camadas:



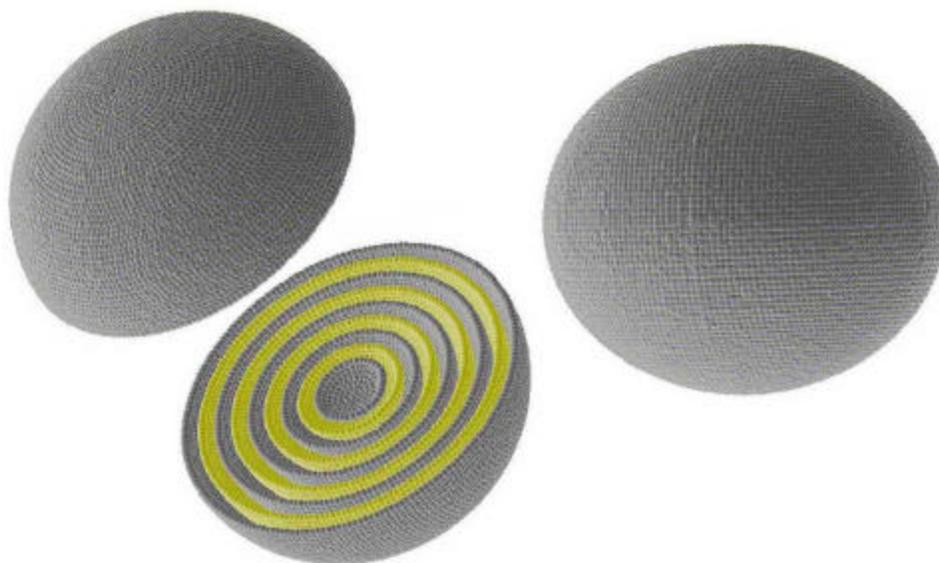
**Figura 3-24:** *Modelo de Multicamadas elipsoidais.*



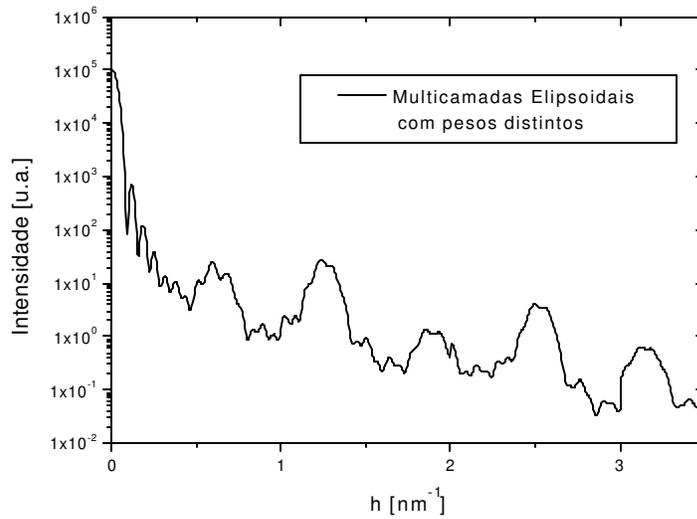
**Figura 3-25:** Função de espalhamento para o Modelo de Multicamadas elipsoidais.

Vemos que ocorre um alargamento dos picos. Isso ocorre devido a quebra de simetria esférica (Cicariello/1999).

Até agora, estamos assumindo todas as subunidades com o mesmo *peso*. Tomemos este mesmo exemplo só que atribuindo pesos 1 e 3 para as subunidades alternadamente,



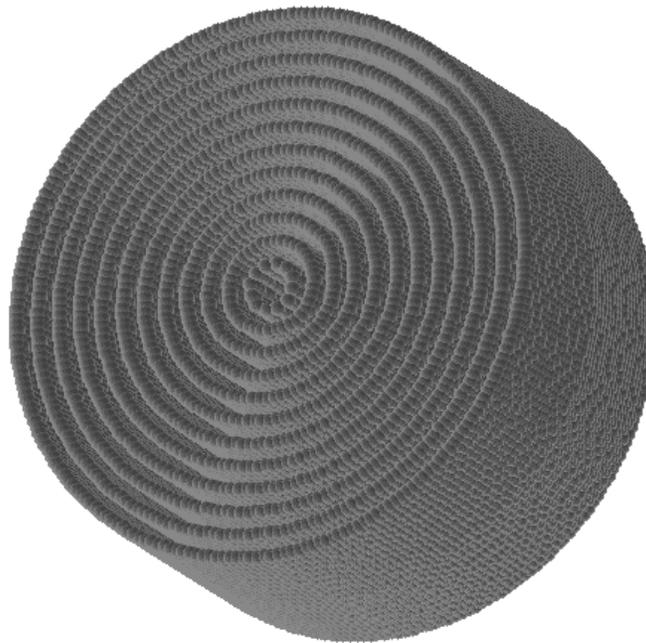
**Figura 3-26:** Modelo de Multicamadas elipsoidais com pesos diferentes para cascas alternadas.



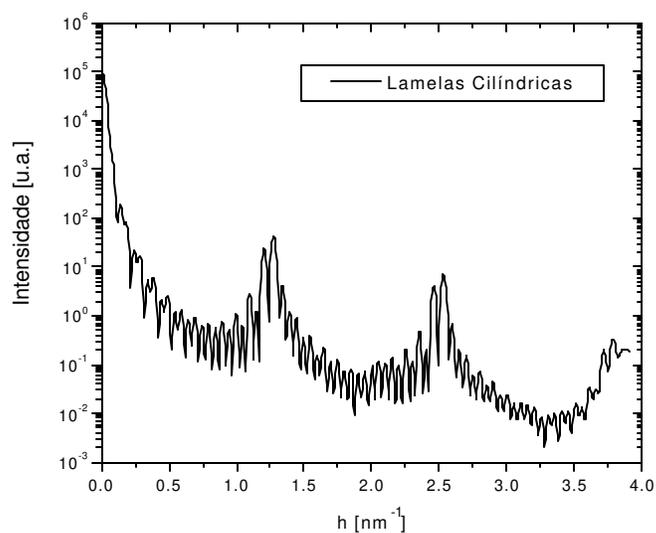
**Figura 3-27:** Função de espalhamento para o Modelo de Multicamadas elipsoidais com diferentes pesos. Note o aparecimento de mais picos.

Comparando-se com a Figura 3-25 vemos o aparecimento de mais picos. Estes picos se devem à superestrutura formada pela variação dos pesos, que confere ao sistema outra periodicidade característica.

Para monocamadas cilíndricas utilizamos o modelo a seguir (Figura 3-28). Note a diferença na conformação dos máximos (Figura 3-29) quando comparados com o caso esférico.

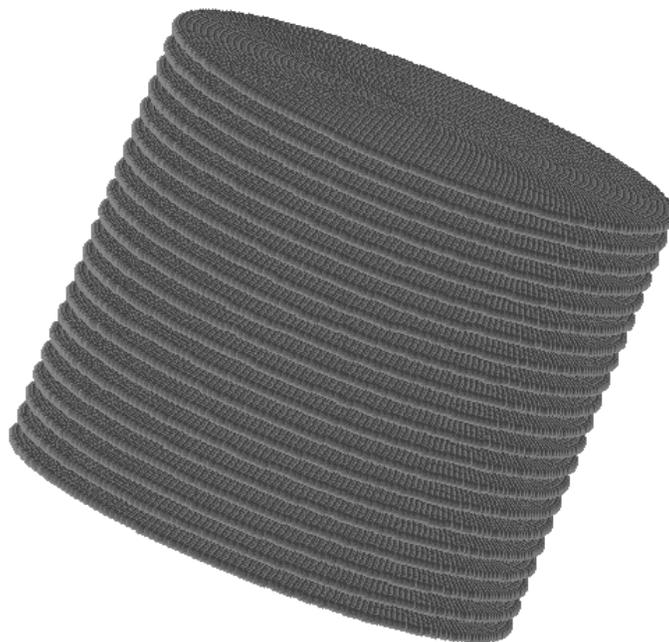


**Figura 3-28:** Modelo de Multicamadas Cilíndricas.

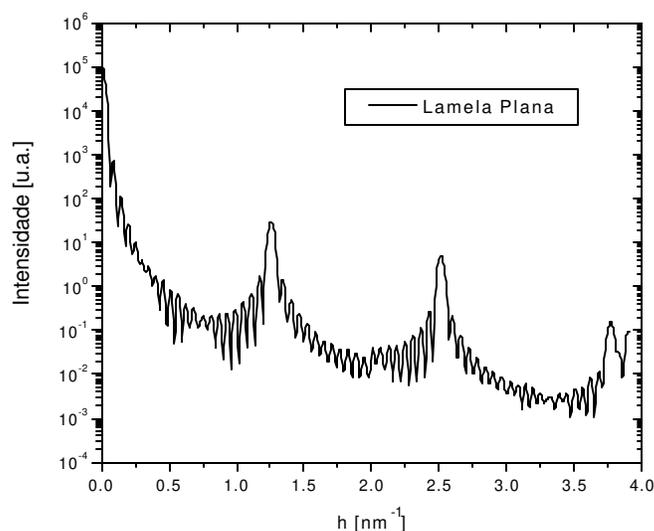


**Figura 3-29:** Intensidade de espalhamento para o modelo de multicamadas cilíndricas.

Para monocamadas planas utilizamos o modelo abaixo. Note a diferença máximos componentes da Figura 3-30 para com o caso esférico e cilíndrico. Quanto maior a área dos planos componentes das multicamadas, mais estreitos seriam os picos de difração.



**Figura 3-30:** Modelo de Multicamadas planas.



**Figura 3-31:** Intensidade de espalhamento para o modelo de multicamadas planas.

Os exemplos anteriores demonstram a potencialidade do método de elementos finitos no estudo de sistemas monodispersos, podendo-se até descrever possíveis periodicidades intrínsecas as partículas componentes do sistema. Como já havíamos salientado, no caso de multicamadas elipsoidais, cilíndricas e planas, somente expressões assintóticas podem ser desenvolvidas em cálculos analíticos. No entanto, com o método de elementos finitos, temos acesso a *toda* curva de espalhamento, partindo desde a origem do espaço recíproco.

Nos casos apresentados geramos um modelo e partindo deste obtemos a curva de espalhamento. Como citado anteriormente o processo inverso, ou seja, a restauração da estrutura tridimensional a partir da curva de espalhamento, não possui solução única. Os argumentos para esta afirmação, bem como a apresentação de métodos para a reconstrução tridimensional serão discutidos a seguir.



## 4. RECUPERAÇÃO DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

O objetivo central de estudos estruturais em biologia molecular é estabelecer uma relação entre a estrutura (ou precisamente mudanças estruturais) e a função de macromoléculas biológicas. Nos últimos anos uma quantidade enorme de informações estruturais tem sido obtida utilizando cristalografia de proteínas e ressonância nuclear magnética (NMR). Estes métodos de alta resolução somente podem ser aplicados sobre condições muito específicas: geralmente é muito difícil crescer-se cristais com alto peso molecular (MW), cristais estes que são necessários para difração. A aplicação de NMR é limitada a pequenas proteínas (MW < 30 kd). A maioria das funções celulares executadas pelas proteínas está intimamente ligada ao seu enovelamento e, deste modo, métodos alternativos de obter-se informações estruturais devem ser explorados.

Espalhamento a baixos ângulos seja de raios X ou de nêutrons apenas fornece informações em baixa resolução (de 1 a 100 nm). No entanto é uma técnica aplicável a uma grande variedade de condições e tamanhos das partículas e por isso, permite a análise de macromoléculas biológicas e seus complexos em condições próximas às fisiológicas e também permite um estudo direto das mudanças estruturais que podem ocorrer quando alteramos as condições externas.

### 4.1. INFORMAÇÃO CONTIDA EM ESPALHAMENTO A BAIXO ÂNGULO

O espalhamento fornecido por macromoléculas em solução está diretamente relacionado com a sua estrutura tridimensional. Se a estrutura atômica da molécula é conhecida, o perfil de espalhamento dado por um “ensemble” delas pode ser calculado com considerável precisão. Não é possível, no entanto reverter o processo e obter a estrutura molecular a partir da curva de espalhamento. A desordem dos espalhadores produz uma média rotacional de seu espalhamento o que gera uma perda de informação sobre a direção relativa dos vetores interatômicos, sem os quais não podemos restaurar a estrutura tridimensional.

Agora que sabemos que a informação contida em um perfil de espalhamento é pequena uma questão permanece: como quantificar a quantidade de informação. Além disso, dado o erro em perfis de espalhamento, com qual precisão podemos especificar a informação? Respostas para estas perguntas são necessárias quando desejamos restaurar a estrutura tridimensional.

A análise do problema de espalhamento a baixos ângulos baseando-se no teorema da amostragem de Shannon (Shannon & Weaver, 1949) pode servir como guia para responder as questões apresentadas anteriormente. Detalhes matemáticos deste desenvolvimento pode ser obtido em diversos artigos (Moore, 1980; Taupin and Luzzati, 1982), no entanto, utilizando o teorema da amostragem é possível determinar-se o número de graus de liberdade (Shannon channels) associados com o perfil de espalhamento  $I(h)$  em um intervalo  $h_{\min} < h < h_{\max}$  como sendo

$$N_s = D_{\max} (h_{\max} - h_{\min}) / \pi \quad (4-1)$$

, sendo  $D_{\max}$  a dimensão máxima da partícula. Devido a curva de espalhamento decair muito rapidamente com  $h$  ela é registrada em baixa resolução e na prática  $N_s$  não excede 20 – 25. Segundo esta estimativa, SAS não é considerado somente uma técnica de baixa resolução mas também de baixa informação. Informações adicionais sobre a estrutura da partícula podem ser obtidas por variação de contraste (Stuhrmann and Kirste, 1965). Como vimos anteriormente, o contraste de uma partícula ou uma densidade eletrônica  $\rho(r)$  em um solvente de densidade  $\rho_s$  é a densidade média efetiva  $\Delta\rho = \langle \rho(\vec{r}) \rangle - \rho_s$ . Para proteínas, experimentos em diferentes  $\rho_s$  (em princípio) permitem a obtenção de mais informações sobre a forma da partícula.

#### 4.2. DETERMINAÇÃO DA FORMA DE MACROMOLÉCULAS UTILIZANDO ESFÉRICOS HARMÔNICOS – EXPANSÃO MULTIPOLAR

Para o cálculo da intensidade de espalhamento, integramos a equação ( 1-7 ) utilizando coordenadas cartesianas até agora. No entanto este sistema de coordenadas não fornece muito “insight” sobre a relação existente entre  $\rho(\vec{r})$  e  $I(h)$ . O fato da equação ( 1-7 ) assumir orientação randômica de um sistema diluído de partículas e uma integração sobre um ângulo sólido favorece a utilização do sistema de coordenadas polares para  $\rho(\vec{r})$ :

$$\rho(\vec{r}) = \rho(r, \theta, \varphi) \equiv \rho(r, \omega) \quad (4-2)$$

Esta função pode ser representada como uma série

$$\rho(\vec{r}) \equiv \rho_L(\vec{r}) = \sum_{l=0}^L \sum_{m=-l}^l \rho_{lm}(\vec{r}) Y_{lm}(\omega) \quad (4-3)$$

sendo

$$\rho_{lm}(\vec{r}) = \int \rho(\vec{r}) Y_{lm}^*(\omega) d\omega \quad (4-4)$$

a função radial e  $Y_{lm}(\omega)$  esféricos harmônicos. O valor  $L$  (máxima ordem da expansão) é a precisão na representação da estrutura da partícula :  $\rho_L(\vec{r}) \rightarrow \rho(\vec{r})$  quando  $L \rightarrow \infty$ . Com esta expansão, a intensidade espalhada da partícula é expressa como (Stuhrmann, 1970a)

$$I(h) = \sum_{l=0}^{\infty} \sum_{m=-l}^l 2\pi^2 |A_{lm}(h)|^2 \quad (4-5)$$

sendo  $h$  o módulo do vetor de espalhamento  $\vec{h}$  ( $h = (4\pi/\lambda)\text{sen } \theta$ ),  $\lambda$  o comprimento de onda e  $2\theta$  o ângulo de espalhamento) e

$$A_{lm}(h) = i^l (2/\pi)^{1/2} \int_0^{\infty} \rho_{lm}(\vec{r}) j_l(h\vec{r}) r^2 dr \quad (4-6)$$

é a transformada de Hankel da função radial, sendo  $j_l(h\vec{r})$  funções de Bessel esféricas.

A estrutura de uma grande variedade de partículas homogêneas pode ser descrita com o auxílio de uma função de forma angular  $F(\omega)$  sendo, (Svergun et al, 1996, 1997)

$$\rho(\vec{r}) = \begin{cases} 1 & 0 \leq r < F(\omega) \\ 0 & r \geq F(\omega) \end{cases} \quad (4-7)$$

Esta função também pode ser desenvolvida em séries,

$$F(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega) \quad (4-8)$$

e os coeficientes multipolares são números complexos,

$$f_{lm} = \sum_{l=0}^L \sum_{m=-l}^l F(\omega) Y_{lm}(\omega) \quad (4-9)$$

Em termos destes coeficientes podemos reescrever as amplitudes parciais  $A_{lm}(h)$  como a série de potências,

$$A_{lm}(h) = i^l (2/\pi)^{1/2} \sum_{p=0}^{p_{\max}} \left( (-1)^p f_{lm}^{(l+2p+3)} \cdot [2^p p!(l+2p+3)[2(l+p)+1]! \right]^{-1} s^{2p} \quad (4-10)$$

e os coeficientes da  $q$ -ésima potência são dados por

$$f_{lm}^{(q)} = \int_{\omega} [F(\omega)]^q Y_{lm}^*(\omega) d\omega \quad (4-11)$$

Apesar do fato de que a série na equação (4-5) é infinita, a contribuição das amplitudes parciais para  $l > L$  é relativamente pequena.

Dada uma curva de espalhamento experimental  $I_{\text{exp}}(h)$  especificada em  $N$  pontos  $h_i$ , a determinação da forma é feita minimizando-se a função,

$$\Phi(I_{\text{exp}}, f_{lm}) = R_l^2(I_{\text{exp}}, f_{lm}) + \mu \vec{r}_0^2 + \nu H^2 \quad (4-12)$$

sendo,

$$R_l^2(I_{\text{exp}}, f_{lm}) = \sum_{i=1}^N \left\{ [I_{\text{exp}}(h_i) - I(h_i)] W(h_i) \right\}^2 \times \left\{ \sum_{i=1}^N [I_{\text{exp}}(h_i) W(h_i)]^2 \right\}^{-1} \quad (4-13)$$

a variação em quadrados mínimos entre  $I_{\text{exp}}(h)$  e  $I(h)$  com a função peso  $W(h_i) = h_i^2 / [\sigma(h_i) / I_{\text{exp}}(h_i)]$ , sendo  $\sigma(h_i)$  o desvio padrão no  $i$ -ésimo ponto;  $\vec{r}_0$  é a posição do centro de massa e  $H(F)$  é a medida relativa da negatividade de  $F(\omega)$ :

$$H(F) = \frac{\int F^2(\omega) d\omega}{\int_{\omega} F^2(\omega) d\omega} \quad (4-14)$$

sendo que  $\omega^-$  indica a integração sobre valores negativos de  $F(\omega)$ . O primeiro vínculo na equação (4-12),  $\mu \vec{r}_0^2$ , prende a partícula próxima à origem; o segundo assegura que  $F_L(\omega)$  é positivo definido ( $\mu$  e  $\nu$  são multiplicadores auto-ajustáveis). Partindo de uma forma inicial descrito por um conjunto de coeficientes  $f_{lm}^{(\text{inicial})}$  (por exemplo uma esfera), os coeficientes que minimizam (4-12) podem ser obtidos por um algoritmo de minimização não linear (Svergun, 1996).

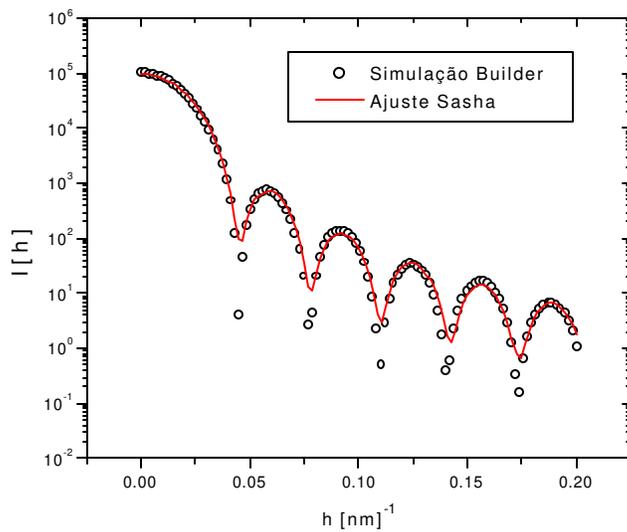
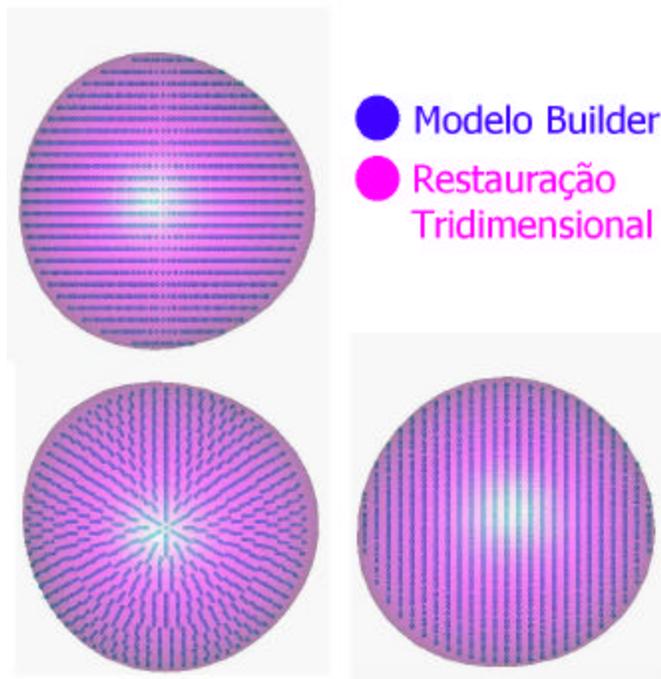
Os procedimentos práticos para a minimização e ajuste com dados experimentais podem ser obtidos em diversos artigos (Svergun et al., 1997; Kozin, Volkov & Svergun, 1997). No entanto a discussão anterior sobre a quantificação do número de graus de liberdade (Shannon channels) do sistema aplica-se diretamente nesta aplicação pois auxilia a determinação do máximo valor de  $L$ , que como vimos, trunca as somas.

A restauração tridimensional utilizando esféricos harmônicos retringe-se a partículas globulares sem cantos ou arestas. Isso ocorre devido ao fato de que a informação contida na curva de espalhamento não permite a utilização de grandes valores de  $L$  (tipicamente  $L = 4$ ) e conseqüentemente somente superfícies com contornos suaves podem ser restauradas. Mais que isso, mesmo que tenhamos uma grande região espalhada, isto é, um intervalo grande de  $h$ , e conseqüentemente um grande número de Shannon channels, esse método poderá não ajustar os dados experimentais. Isso ocorrer devido a vários fatores. Se a forma da partícula espalhadora for simples e puder ser descrita por poucos esféricos harmônicos, o ajuste poderá ser feito para qualquer range de  $h$ . No entanto, em proteínas reais não é isso que acontece. Indo para valores maiores de  $h$ , estaremos aumentando a resolução e observando pequenos detalhes estruturais da partícula em questão. Com isso a expansão multipolar não se aplicará neste caso. Sendo assim quando realizamos ajustes em dados experimentais utilizando este método, devemos nos concentrar em regiões tais que o número de Shannon channels não seja muito grande, pois o método não se aplicará. Informações adicionais sobre características da partícula (partícula esférica, prolata, oblata, com algum tipo de simetria, etc), permite o aumento do valor de  $L$  uma vez que podem anular várias componentes de (4-9).

O que foi discorrido anteriormente será melhor explicado através dos exemplos a seguir. Mostrarei algumas simulações efetuadas utilizando o programa BUILDER e a correspondente restauração tridimensional utilizando-se o programa Sasha (Kozin, Volkov & Svergun, 1997). Posteriormente utilizando o programa Crysol (Svergun D.I., Barberato C. & Koch M.H.J. (1995)) simulei a curva de espalhamento de uma proteína cuja estrutura já foi resolvida através de cristalografia de proteínas. Esta curva pode representar uma curva experimental e, como veremos a restauração tridimensional utilizando a expansão multipolar limita-se somente ao envelope externo da proteína.

- **ESFERA**

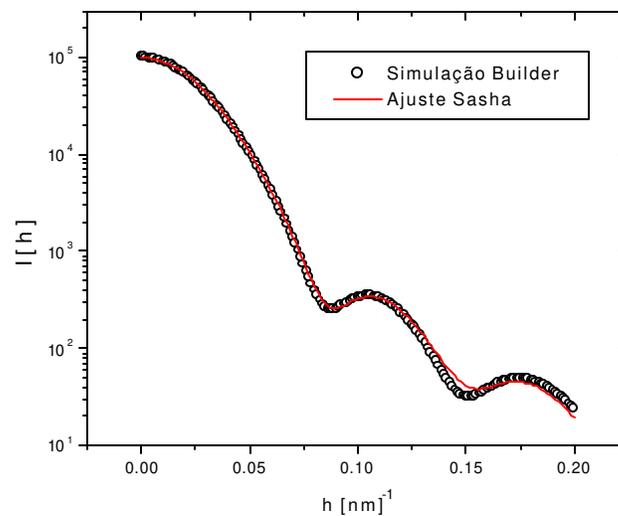
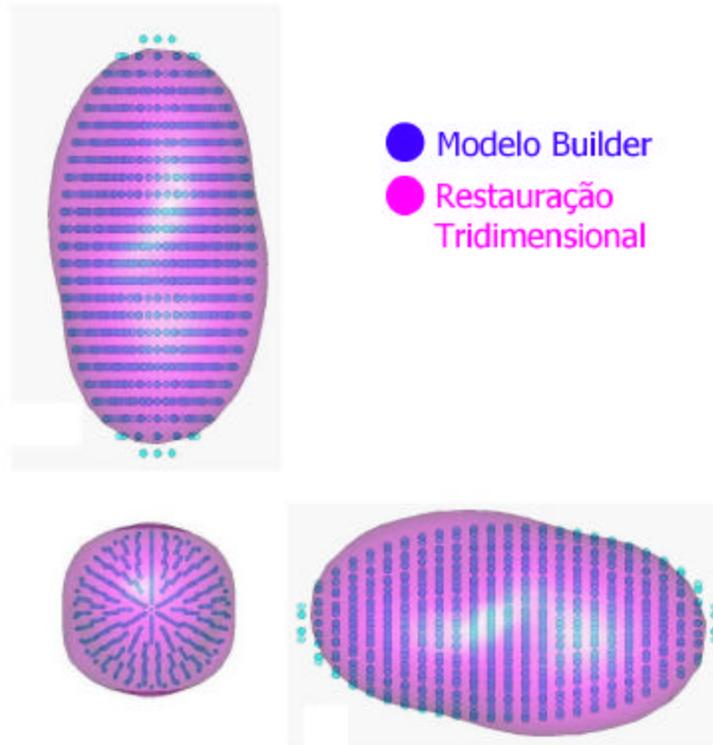
⇒ Temos um ótimo ajuste do espalhamento da esfera. A leve anisotropia do modelo gerado decorre do fato de que o modelo construído por elementos finitos não é uma esfera perfeita.



**Figura 4-1:** Envelope ab initio restaurado e ajuste obtido para uma esfera utilizando esféricos harmônicos.

- **ELIPSÓIDE PROLATO**

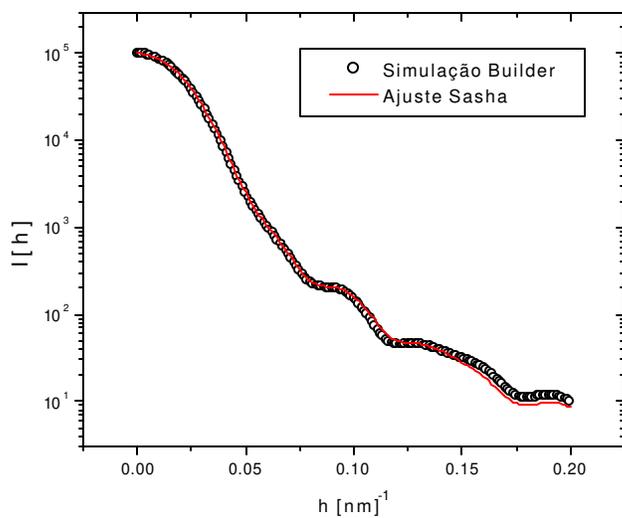
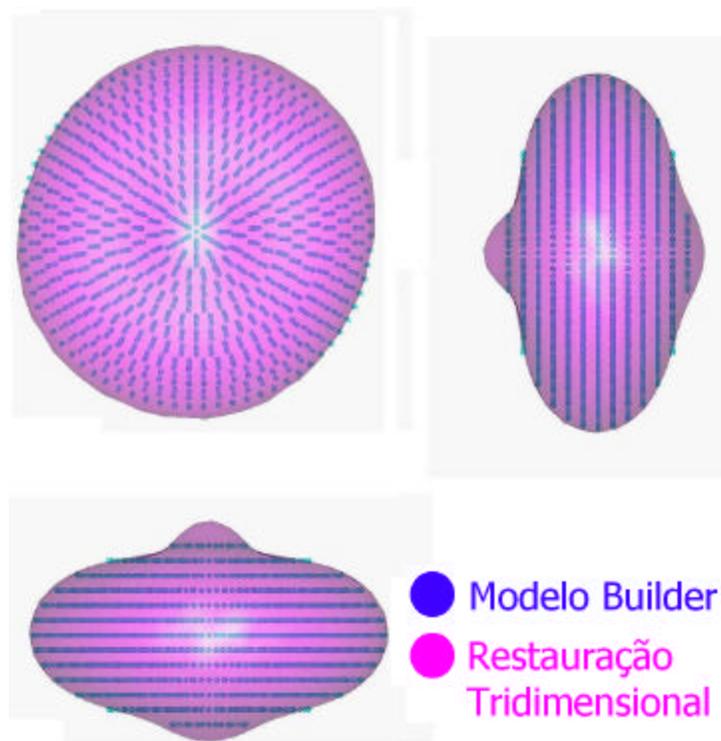
⇒ Temos um ótimo ajuste do espalhamento do elipsóide. A leve anisotropia do modelo gerado decorre do fato de que o modelo construído por elementos finitos não é um elipsóide perfeito.



**Figura 4-2:** Envelope ab initio restaurado e ajuste obtido para um elipsóide prolato utilizando esféricos harmônicos .

- **ELIPSÓIDE OBLATO**

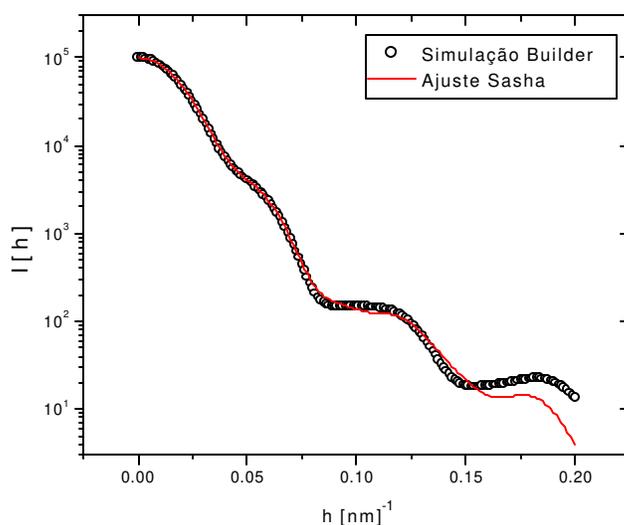
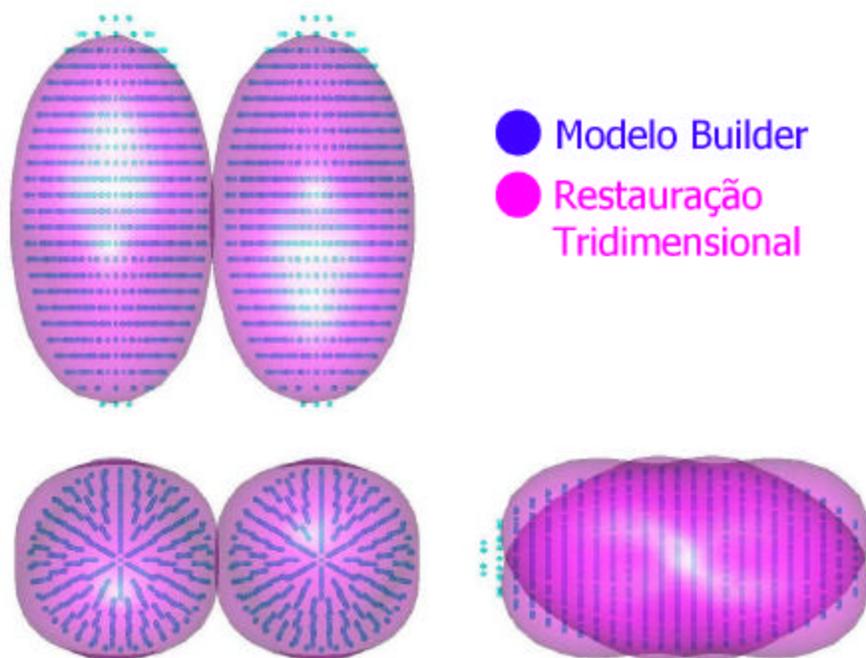
⇒ Temos um ótimo ajuste do espalhamento do elipsóide. A leve anisotropia do modelo gerado decorre do fato de que o modelo construído por elementos finitos não é um elipsóide perfeito.



**Figura 4-3:** Envelope ab initio restaurado e ajuste obtido para um elipsóide oblato utilizando esféricos harmônicos .

- **DÍMERO DE ELIPSÓIDES PROLATOS**

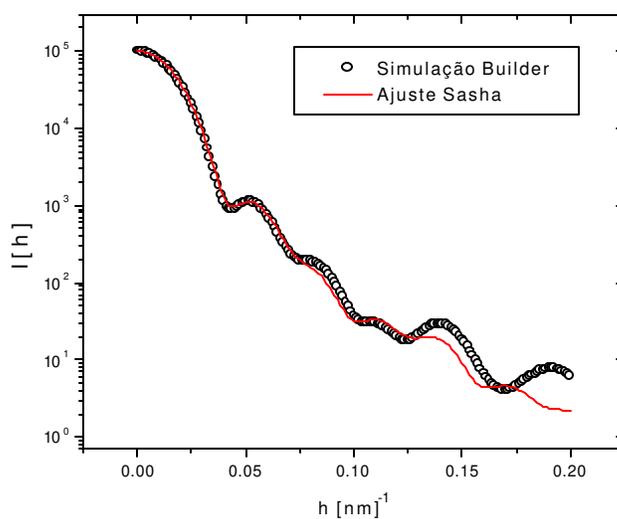
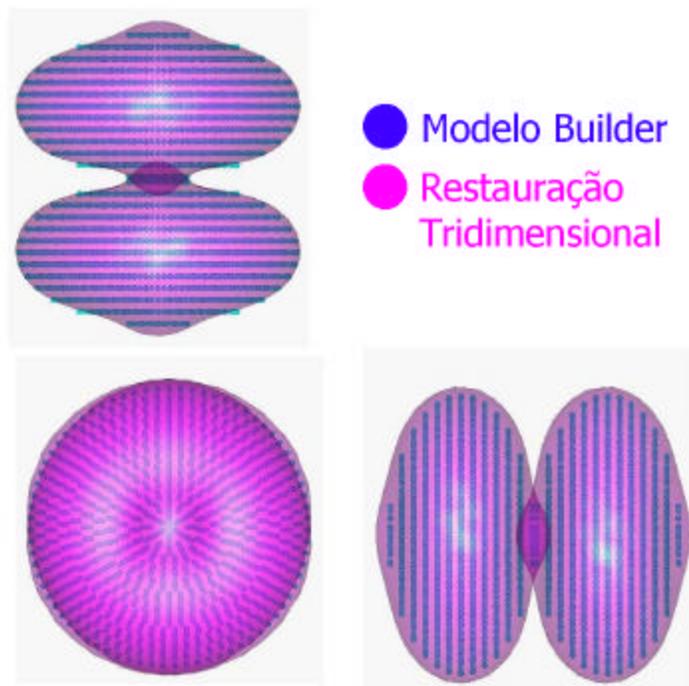
⇒ O programa Sasha permite que informemos a curva de espalhamento do monômero (quando conhecido) para o ajuste de espalhamento por dímeros. O ajuste seguinte somente foi possível utilizando este artifício. Sem a informação do monômero o software restauraria a forma precariamente (pouca informação).



**Figura 4-4:** Envelope ab initio restaurado e ajuste obtido para um dímerno de elipsóides prolatos utilizando esféricos harmônicos.

- **DÍMERO DE ELIPSÓIDES OBLATOS**

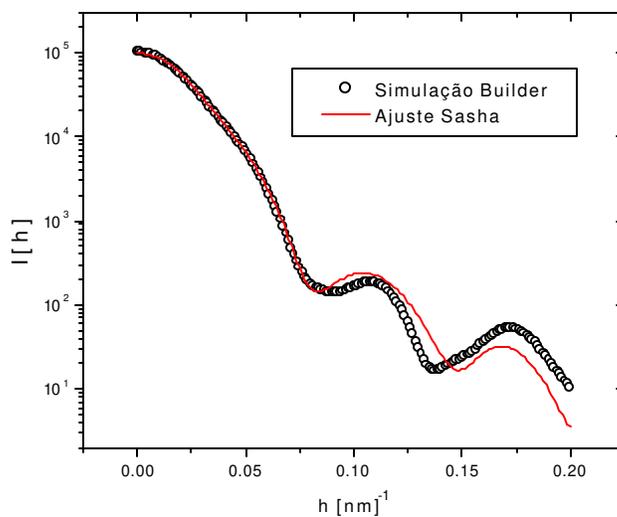
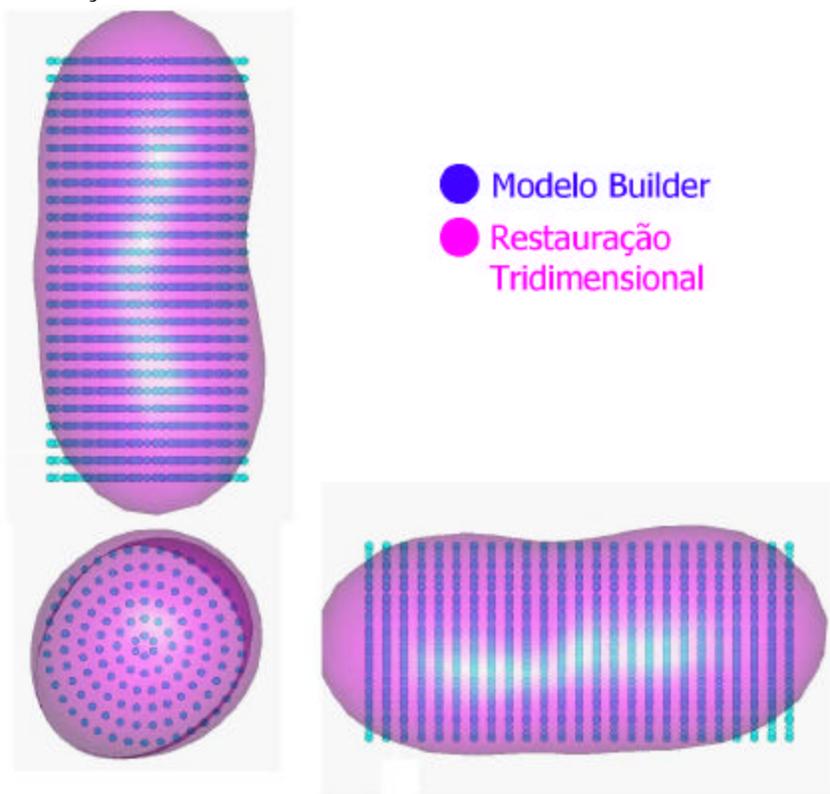
⇒ O programa Sasha permite que informemos a curva de espalhamento do monômero (quando conhecido) para o ajuste de espalhamento por dímeros. O ajuste seguinte somente foi possível utilizando este artifício. Sem a informação do monômero o software restauraria a forma precariamente (pouca informação).



**Figura 4-5:** Envelope ab initio restaurado e ajuste obtido para dímero de elipsóides oblatos utilizando esféricos harmônicos.

- **CILINDRO**

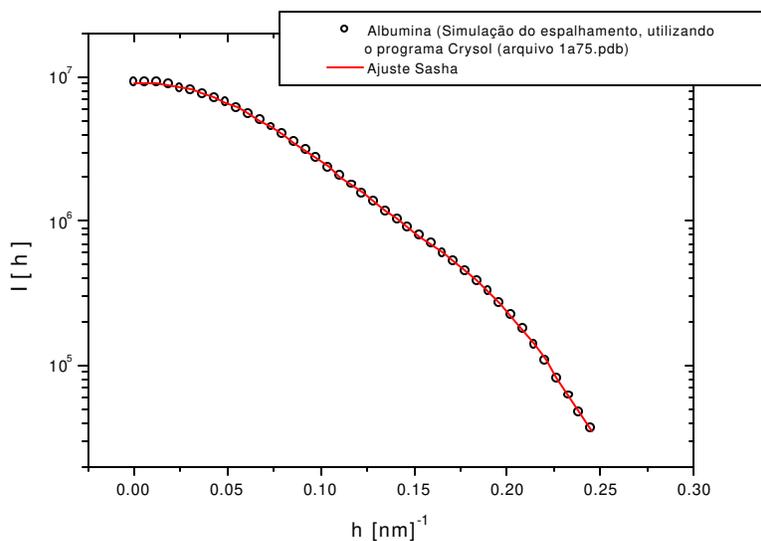
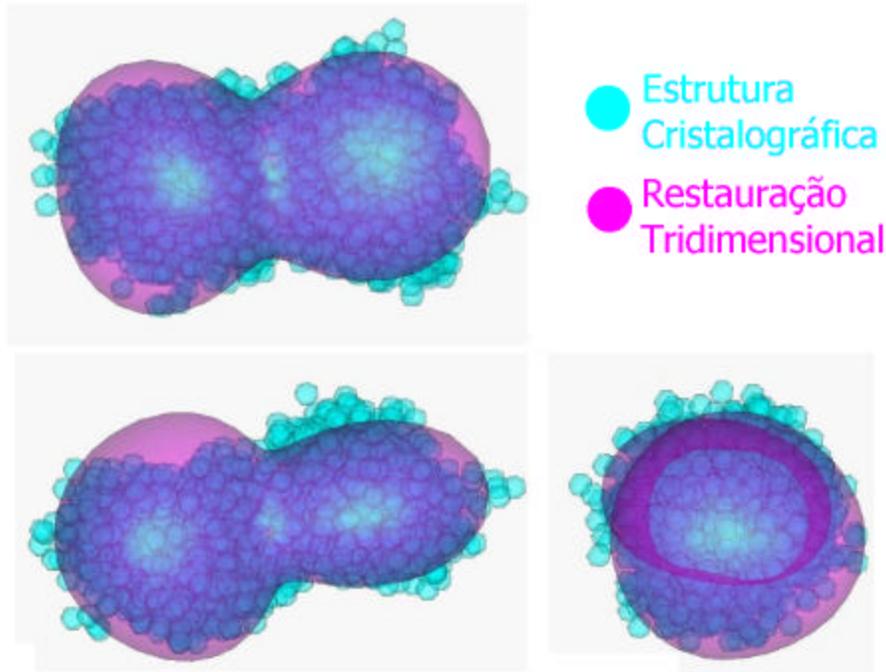
⇒ Aqui fica clara a limitação da expansão multipolar somente a contornos suaves. Como o cilindro possui um ângulo reto necessitaria-se de muitos esféricos harmônicos para descreve-lo. Com isso o ajuste da curva de espalhamento atingiu baixa resolução.



**Figura 4-6:** Envelope ab initio restaurado e ajuste obtido para cilindro utilizando esféricos harmônicos.

- **ALBUMINA**

⇒ Como a estrutura cristalográfica possui muitos detalhes pequenos não temos informação deles tomando apenas a região inicial da curva de espalhamento. No entanto, a expansão multipolar consegue descrever a anisotropia da proteína.



**Figura 4-7:** Envelope ab initio restaurado e ajuste obtido albumina utilizando esféricos harmônicos.

Nas simulações com o BUILDER o empacotamento utilizado é o mesmo citado anteriormente em CONSTRUÇÃO DE MODELOS. No entanto para efeito de visualização nas figuras anteriores indiquei as subunidades por pequenos pontos.

Para sólidos simples (esfera, elipsóides, dímeros) vemos que a restauração tridimensional por expansão multipolar se aplica satisfatoriamente. Algumas distorções são percebidas, mas vale lembrar que esta é uma determinação ab-initio, isto é, baseada somente na curva de espalhamento e também em indicações de simetria. Para o caso dos dímeros, o ajuste somente pôde ser feito fornecendo ao programa as curvas de monômeros e dímeros conjuntamente e o mesmo ajustava a distância entre eles. Se fornecêssemos apenas a curva de dímero, o resultado seria completamente errôneo.

Na realidade, como demonstrado por Svergun (Svergun et al.,1996), se o número de parâmetros independentes descrevendo a forma da partícula espalhadora for da ordem de 1 a 1.5 vezes o número de Shannon channels dado pelo conjunto de dados, a forma restaurada pelo método de esféricos harmônicos é única, caso os dados não possuam erros. A existência de erros pode gerar pequenas distorções na forma restaurada. Na descrição de uma esfera por exemplo, o número de parâmetros necessários para descreve-la é apenas 1 uma vez que na base de esféricos harmônicos ela é representada por  $Y_0^0(\omega)$ . Para os elipsóides, precisamos de mais parâmetros, mas ainda constitui um conjunto pequeno.

Já para o cilindro vemos que a aproximação é muito grosseira. Isso era esperado uma vez que o cilindro apresenta uma aresta em ângulo reto (topo e lateral) o que não pode ser descrito através de poucos esféricos harmônicos, porém, as características principais da forma são restauradas (prolaticidade, raio de giro etc.).

Para a albumina temos o mesmo efeito. A simulação da curva de espalhamento através da estrutura cristalográfica resolvida pode ser feita para, em princípio, um range muito extenso do espaço recíproco. No entanto, como salientado anteriormente, o procedimento por esféricos harmônicos não consegue ajustar a curva nesta região, além de gerar aberrações de forma. De qualquer modo o ajuste da curva é perfeito e sobrepondo a estrutura cristalográfica com o envelope gerado vemos que a expansão multipolar consegue identificar as duas regiões mais densas da proteína bem como a parte flexível (central) da mesma.

Pelos exemplos anteriores vemos claramente as potencialidades e limitações da expansão multipolar. Esta discussão é amplamente apresentada em Svergun/1996, com uma grande variedade de testes. Para estruturas simples e aproximadamente globulares a recuperação tridimensional é feita facilmente e condiz com a estrutura espalhadora. Já para estruturas mais complexas a aproximação é grosseira e o ajuste da curva espalhada também é precário. Este método é aplicável somente a um valor de 15 a 17 Shannon channels; para valores acima deste intervalo os erros numéricos acabam destruindo o processo de minimização. Isto ocorre devido ao fato das amplitudes  $A_{lm}(h)$  serem calculadas pela série de potências (4-10) que acelera o processo de cálculo, porém aumenta significativamente o erro numérico para valores acima de 15 Shannon channels. Segundo Svergun, uma melhor precisão seria obtida se utilizássemos a forma original de  $A_{lm}(h)$  equação (4-6), porém a integração numérica necessitaria de um tempo de computação extremamente longo. Com isso, mesmo que tenhamos dados experimentais com um grande número de Shannon channels, não poderemos ajustar toda a curva, pois causaria aberrações na forma restaurada ocasionadas por erros numéricos.

A determinação do envelope tridimensional de partículas em solução pode ser feita através da curva de espalhamento deste sistema através deste método, e apesar das limitações discutidas no texto, a restauração tridimensional obtida do nível de resolução ajustado é *única*, pois utiliza-se um processo de minimização global. Um passo posterior seria a utilização do método de elementos finitos para a recuperação tridimensional de partículas. No entanto o procedimento é muito mais complexo uma vez que o número de graus de liberdade do sistema aumenta enormemente. Por exemplo, se para a descrição utilizando a expansão multipolar usarmos  $L = 4$ , teremos que obter 25 fatores  $f_{lm}$  ( $(L+1)^2$  fatores) um número considerável de parâmetros mas ainda assim manejável de modo direto se tivermos um algoritmo eficiente. O mesmo já não ocorre se quisermos utilizar modelos baseados em elementos finitos. Como veremos a determinação direta da forma da estrutura é impossível, pois necessitaria de um tempo computacional extremamente longo. Neste caso se faz necessário métodos estocásticos que obtêm a estrutura mais provável por “tentativa e erro”. Com isso, diferentemente do caso de esféricos harmônicos, não poderemos garantir unicidade no resultado, devido ao processo de minimização. A seguir apresentarei dois procedimentos para a determinação da estrutura tridimensional através do método de elementos finitos: um utilizando o algoritmo genético e outro utilizando *simulated annealing*.

#### **4.3. DETERMINAÇÃO DA FORMA DE MACROMOLÉCULAS UTILIZANDO ALGORITMOS ESTOCÁSTICOS – MÉTODO DE ELEMENTOS FINITOS.**

A otimização combinatória consiste em uma classe de problemas centrais em disciplinas de ciência da computação e engenharia. Pesquisas nestas áreas possuem como objetivo o desenvolvimento de técnicas eficientes para a obtenção de máximos e mínimos para funções de muitas variáveis independentes. Estas funções usualmente denominadas *cost function* ou *objective function* representam uma medida quantitativa de ajuste para alguns sistemas complexos. Estas funções dependem da configuração detalhada de muitas partes do sistema.

Um problema clássico que envolve esta análise é o do caixeiro viajante (*traveling salesman problem*). Dada uma lista de  $N$  cidades e maneiras de se calcular o custo de viagem entre quaisquer duas cidades, deve-se planejar a rota do viajante de modo que passe em todas as cidades e retorne a cidade inicial, minimizando o custo total. Todos os métodos exatos para resolução deste problema necessitam de esforços computacionais que aumentam exponencialmente com  $N$ , e na prática, soluções exatas somente podem ser obtidas em problemas envolvendo algumas centenas de cidades. Deste modo, a abordagem exata deste tipo de problemas se torna impossível de ser realizada quando o número de parâmetros é grande. Por isso, métodos heurísticos vêm sendo desenvolvidos com esforços computacionais proporcionais a pequenas potências de  $N$ . No entanto estes métodos recaem em um problema específico: não existe garantia de que um procedimento heurístico para se obter a solução de um dado problema seja eficiente em um outro. No entanto, existem poucos trabalhos na literatura sobre performances de algoritmos heurísticos, pois a análise é usualmente muito difícil e a natureza das médias apropriadas para o estudo quase sempre não é clara.

O problema de se determinar a estrutura tridimensional em baixa resolução de partículas em solução, utilizando o método de elementos finitos se encaixa nesta classe de problemas. Independente do método que utilizarmos, a busca pela estrutura ideal será realizada em um *espaço de busca* (search space) que contenha a estrutura que desejamos. Este espaço, que comumente pode ser uma esfera, é construído por subunidades esféricas (semelhante ao construído na Figura 3-15 ). O método deverá “escolher” dentre as subunidades componentes do espaço de busca, aquelas que constroem melhor a estrutura desejada. Isso faz com que tenhamos um sistema binário, no sentido de que ou a subunidade está na estrutura que queremos (1) ou não está (0). No entanto o número de configurações possíveis é dado por  $2^N - 1$ , sendo  $N$  o número de subunidades utilizadas. Para um modelo composto de  $\mathcal{T}$  subunidades por exemplo, um laço simples contando de 1 até este valor ( $2^{\mathcal{T}}$ ), levaria cerca de 200000 anos em um processador típico de 200 MHz de velocidade (Chacón et al./1998) . Vemos claramente a necessidade da aplicação de um método heurístico para este problema.

#### 4.3.1. ALGORITMO GENÉTICO ( GA )

Na natureza, os organismos melhor sucedidos em competições ocasionadas pela escassez de recursos (por exemplo: comida, espaço) sobrevivem e se reproduzem. Geram sua prole, permitindo a transmissão de suas hereditariedade por meio dos genes contidos em seus cromossomos. A adaptação a um meio em continua mudança é essencial para a perpetuação dos indivíduos de cada espécie. Sendo assim, a seleção natural implica na sobrevivência dos indivíduos mais adaptados, mas isto também implica na sobrevivência dos *genes* mais adaptados. O processo de reprodução permite a diversificação de grupos genes de espécies. A evolução se inicia quando cromossomos de dois pais recombinam durante a reprodução. Novas combinações de genes são geradas partindo dos anteriores e um novo grupo genético é gerado. Segmentos dos dois cromossomos pais são trocados durante o cruzamento, criando a possibilidade de combinações “corretas” de genes, gerando indivíduos melhores. Mutações introduzem mudanças esporádicas e randômicas nos cromossomos. A seleção, cruzamento e mutação, causam uma evolução contínua de um grupo genético de uma dada espécie, e a geração de indivíduos melhor adaptados para o meio competitivo. Os algoritmos genéticos se baseiam nos princípios Darwinianos de seleção natural e evolução apresentados anteriormente. Eles manipulam uma população de soluções potenciais para um problema de otimização. Especificamente, operam representações codificadas de modo a equivalerem a cromossomos de indivíduos na natureza. Cada solução é associada com um valor de ajuste que reflete a qualidade daquela solução em comparação com as outras da população. A seleção então assegura que sobrevivam apenas os indivíduos melhor *ajustados*. As manipulações do *material genético* são feitas através de cruzamento e mutações.

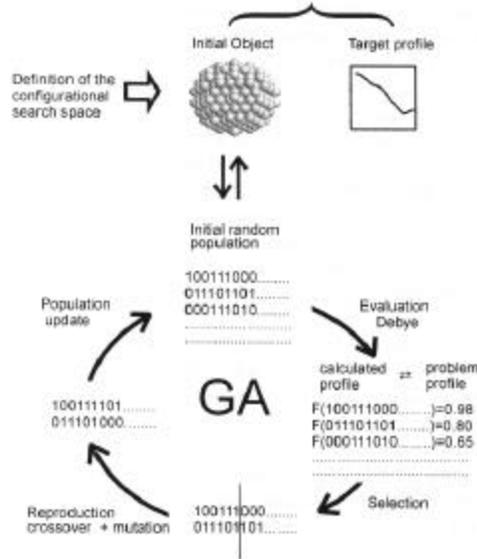
Existem várias formas de se implementar o GA. Apresentarei a seguir, em linhas gerais, a escolhida por Chacón, para a solução do problema inverso de espalhamento.

A codificação é feita em uma cadeia binária. A  $i$ ésima posição na cadeia é descrita pela presença (valor 1) ou falta (valor 0) de uma subunidade nesta posição. Cada posição corresponde à coordenada tridimensional de uma subunidade no espaço de busca.

Os passos executados pelo algoritmo são:

- **Geração das condições iniciais**

A população inicial é gerada aleatoriamente, baseando-se na dimensão máxima da partícula e em uma forma particular que pode ser estabelecida, por exemplo uma esfera, um elipsóide etc. O ponto fundamental é que a estrutura inicial seja capaz de “conter” a estrutura que desejamos construir. O tamanho das subunidades deve ser tal que satisfaça a resolução desejada;



**Figura 4-8:** Esquema da implementação do método do algoritmo genético em simulações de SAXS para resolver numericamente o problema inverso de espalhamento. ( figura tirada de ref. [62])

- **Cálculo do critério de ajuste (F)**

Alterações aleatórias do seqüenciamento são realizadas e estas seqüências, bem como os valores de F são guardados. Para o critério de ajuste é utilizada a função:

$$F = \left( \sum_i (\log(I_{\text{exp}}(S_i)) - \log(I_{\text{model}}(S_i)))^2 \right)^{\frac{1}{2}} \quad (4-15)$$

outro critério que também poderia ser utilizado é:

$$Rf = \sum_i \left\| \frac{\log(I_{\text{exp}}(S_i)) - \log(I_{\text{model}}(S_i))}{\log(I_{\text{exp}}(S_i))} \right\| \quad (4-16)$$

- **Reprodução**

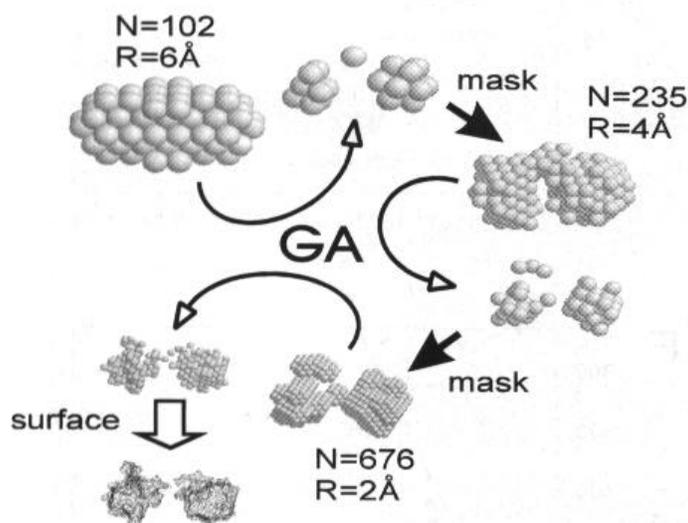
Uma fração ( $P_{\text{elite}}$ ) da população total ( $P_{\text{total}}$ ), logicamente a que possui maiores fatores  $F$  é preservado para a próxima geração. Os membros restantes são apagados de modo a termos novos cromossomos. A regeneração da população é feita utilizando dois operadores genéticos: mutação, que introduz nova informação na população, e *crossover*, que combina a informações dos melhores cromossomos. Além disso todos os cromossomos duplicados são descartados.

- **Update da população**

Finalmente, após a mutação, crossover e rejeição de duplicadas, os novos cromossomos gerados substituem a fração  $(1-P_{elite})$  da população inicial e o processo se repete.

- **Aumento da resolução**

Quanto maior a resolução, maior o tempo gasto na computação. Deste modo, um procedimento foi feito para aumentar a resolução ao longo do processo. Este processo constitui-se de diminuir o raio das subunidades gradativamente a partir de uma estrutura obtida. Isso se faz reconstruindo um modelo obtido com subunidades de Raio  $R_1$  por um empacotamento hexagonal de esferas com raio  $R_2$  ( $R_2 < R_1$ ). O valor mínimo de  $R_2$  está associado à máxima resolução permitida pela curva de espalhamento (equação ( 1-30 )). Esta técnica diminui o tempo de computação e é muito útil quando as dimensões do objeto inicial são desconhecidas.



**Figura 4-9:** Um exemplo da estratégia de mascaramento. O resultado de aumentos sucessivos de resolução pela criação de novos espaços de busca com subunidades menores.  $N$  é o número de subunidades do espaço de busca inicial para cada um dos passos do algoritmo e  $R$  o raio da subunidade. (figura tirada de ref. [62])

Simulações utilizando este método serão apresentadas posteriormente em **SIMULAÇÕES**.

#### 4.3.2. SIMULATED ANNEALING ( SA )

Este método se utiliza da profunda e útil conexão entre mecânica estatística (o comportamento de sistemas com muitos graus de liberdade em equilíbrio térmico a uma temperatura finita) e o processo de otimização combinatório ou multivariável (obter o mínimo de uma dada função que depende de muitos parâmetros). A analogia detalhada com o resfriamento de sólidos fornece um método para a otimização de propriedades de sistemas grandes e complexos.

Não é meu intuito aqui apresentar em profundidade este método. Restringirei-me a uma explanação em linhas gerais, como feito no caso do GA.

A mecânica estatística é a disciplina central em física da matéria condensada, e consiste de um conjunto de métodos para análise de propriedades de agregados de grandes números de átomos encontrados em amostras líquidas ou sólidas. Devido o número elevado de átomos por centímetro cúbico ( $\sim 10^{23}$ ) somente o comportamento mais provável do sistema no equilíbrio térmico a uma dada temperatura será observado em experimentos. Isso pode ser caracterizado por uma média e pequenas flutuações em torno do comportamento médio do sistema, quando a média é tomada em um *ensemble* de sistemas idênticos, introduzidos por Gibbs. Neste *ensemble* cada configuração, definida por um conjunto de posições atômicas  $\{r_{ij}\}$ , do sistema é pesado pelo seu fator de probabilidade de Boltzmann,  $\exp(-E(\{r_{ij}\})/k_B T)$ , sendo  $E(\{r_{ij}\})$  a energia da configuração,  $k_B$  a constante de Boltzmann e  $T$  a temperatura.

Uma questão fundamental em mecânica estatística diz respeito ao que ocorre em um sistema no limite de baixas temperaturas - por exemplo, quantos átomos permanecem fluidos ou solidificam, e se solidificarem, se formam uma estrutura cristalina ou um vidro. Estados fundamentais e configurações próximas a esta em energia são extremamente raras visto todas as configurações possíveis do meio macroscópico e, suas características predominam somente em baixas temperaturas uma vez que como  $T$  é diminuído, a distribuição de Boltzmann colapsa no estado de mínima energia.

Encontrar o estado de mínima temperatura do sistema quando a expressão de cálculo para sua energia é dada, é um problema de otimização combinatória. No entanto o conceito de temperatura de um sistema pode não encontrar um equivalente óbvio nos sistemas que queremos otimizar. Introduzida uma temperatura efetiva para otimização, procedemos para um resfriamento simulado (*simulated annealing*) e esperamos que este processo forneça a melhor solução heurística possível para o problema de otimização combinatória. Utilizamos no lugar da energia do sistema a *cost function* e definimos um conjunto de parâmetros  $\{x_{ij}\}$  relacionados ao nosso problema em particular.

A aplicação desta metodologia na procura da estrutura tridimensional de partículas em solução utilizando a curva de espalhamento foi realizada por Svergun (Svergun/1999), e a descrição do método é sumariada a seguir.

- **Critérios de ajuste**

Para um conjunto de  $M \geq 1$  curvas com variação de contraste  $I_{\text{exp}}^{(i)}(h)$ ,  $i = 1, \dots, M$ , é natural procurarmos uma configuração  $X$  que minimize a discrepância,

$$\chi^2 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{N(i)} \left[ \left( I_{\text{exp}}^{(i)}(h_j) - I^{(i)}(h_j) \right) / \sigma(h_j) \right]^2 \quad (4-17)$$

sendo  $N(i)$  o número de pontos da  $i$ -ésima curva e  $\sigma(h_j)$  denota os erros experimentais.

Esta relação é genérica e pode ser aplicada para experimentos com variação de contraste de densidade eletrônica. Na prática esta classe de experimentos é realizada com *nêutrons*. Para experimentos com raios X a variação de contraste não é utilizada e deste modo teremos  $M = 1$  e assim a forma da expressão (4-17) assemelha-se às utilizadas pelo GA (4-15 e 4-16).

Para uma descrição adequada da estrutura o número de átomos do modelo deve ser grande ( $N \approx 10^3$ ). Além disso, pelos mesmos argumentos descritos no método de expansão multipolar, a unicidade dos modelos não pode ser garantida, mesmo que os dados experimentais sejam otimamente ajustados.

Além disso neste método é introduzida em conjunto com a discrepância um critério de empacotamento (*looseness*). Por este critério são favorecidas as configurações que possuem o maior número de contatos, isto é, tenta-se ao máximo impedir a existência de átomos “soltos”. Para isso, uma lista de contatos é definida para cada átomo do modelo. O número de contatos para um empacotamento hexagonal é  $N_c = 12$  (ou menos para átomos próximos a borda) A conectividade individual de um átomo não solvente é caracterizada pela contagem de seus contatos para o número de átomos da mesma fase  $N_e$ . Uma forma exponencial,

$$C(N_e) = 1 - P(N_e) = 1 - [\exp(-0.5N_e) - \exp(-0.5N_c)]$$

pode ser utilizada para enfatizar átomos fracamente conectados. Esta função decai lentamente de  $C(12) = 1$  (conectividade ideal) para  $C(6) = 0.943$  (metade dos contatos pode indicar por exemplo átomos próximos a borda), seguido de uma queda brusca para átomos com poucos contatos chegando à  $C(0) = 0.002$  (um átomo do modelo isolado, não deveria nunca aparecer). A compactação de uma dada configuração  $X$  pode ser considerada através da média  $P(X) = 1 - \langle C(N_e) \rangle$ . Este valor depende bastante da conectividade de átomos individuais, mas também da anisometria da partícula representada pelos átomos. Por exemplo, para um sistema de uma fase composto de  $N \approx 2000$ ,  $P \approx 0.007$  para uma esfera sólida e  $0.012$  para um elipsóide prolato de revolução com relação axial de 1:10. Preenchendo os dois volumes randomicamente com átomos de fase 0 (solvente) e fase 1 (partícula), teremos  $P \approx 0.1$  nos dois casos.

- **Procedimento de Minimização**

O método para restaurarmos o modelo tridimensional de baixa resolução partindo da curva de espalhamento pode ser formulado da seguinte maneira: dado um espaço de busca, descobrimos configurações  $X$  que minimizam a função (*cost function*)

$$f(X) = \chi^2 + \alpha P(X) \quad (4-18)$$

, sendo  $\alpha > 0$  o peso do vínculo de compactação. Quando utilizamos vínculos, é usual (Svergun/1999) escolher-se o peso de modo que o segundo termo seja uma significativa (por exemplo  $\sim 10 - 50\%$ ) contribuição para a função no final da minimização. Como  $\chi^2$  é esperado como sendo em torno de 1 para uma solução correta e  $P(X)$  da ordem de  $10^{-2}$  para corpos compactos,  $\alpha \approx 10^1$ .

A utilização do SA neste caso consiste em fazer-se modificações randômicas no sistema (isto é, no vetor  $X$ ) escolhendo modificações que diminuam a energia  $f(X)$ , porém algumas vezes também aceitando algumas que aumentam  $f(X)$ . A probabilidade de aceitar-se este tipo de soluções diminui no curso do processo de minimização (o sistema é esfriado). No início, a temperatura é alta e as mudanças quase que randômicas, enquanto que no final a configuração com mínima energia é obtida. O algoritmo foi implementado utilizando *faster simulated quenching* (Press et al., 1992; Ingber, 1993) :

1. Inicie partindo de uma configuração  $X_0$  em uma temperatura alta  $T_0$ .

2. Selecione um átomo aleatoriamente, altere randomicamente sua fase (passando para a configuração  $X'$ ), e calcule  $\Delta = f(X') - f(X_0)$
3. Se  $\Delta < 0$ , use  $X'$ , se  $\Delta > 0$ , faça isso com uma probabilidade  $\exp(-\Delta/T)$ . Repita o passo 2 partindo de  $X'$  (se aceito) ou de  $X_0$  novamente.
4. Mantenha a temperatura constante para 100N reconfigurações ou para 10N reconfigurações bem sucedidas, dependendo da qual ocorrer primeiro, e a seguir resfrie o sistema ( $T = 0.9T$ ). Continue esfriando até que nenhuma melhora em  $f(X)$  seja observada.

Somente um átomo é modificado por movimento e somente sua contribuição é modificada na expressão de espalhamento. Esta é a etapa de maior gasto computacional, o cálculo da curva de espalhamento, e este método acelera em muito o processo. Esta aceleração possibilita a utilização do SA, o qual é muito robusto (Ingber, 1993) mas pode também ser muito lento. Alguns milhões de funções devem ser calculados para um refinamento típico.

### 4.3.3. SIMULAÇÕES

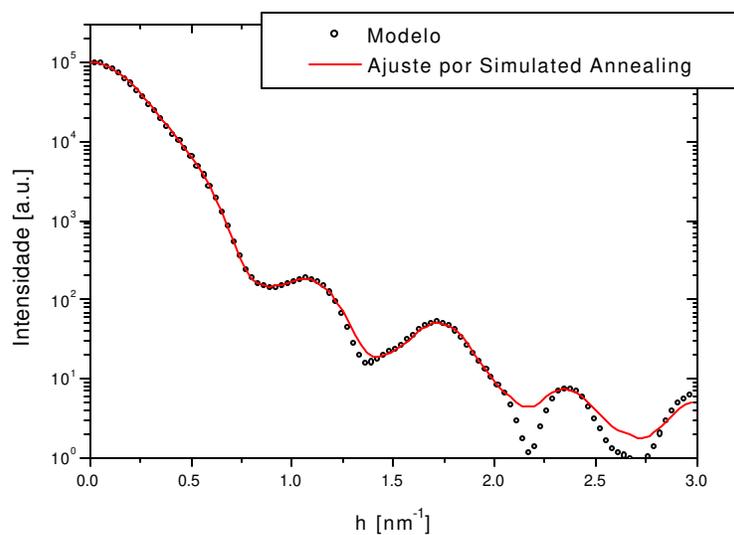
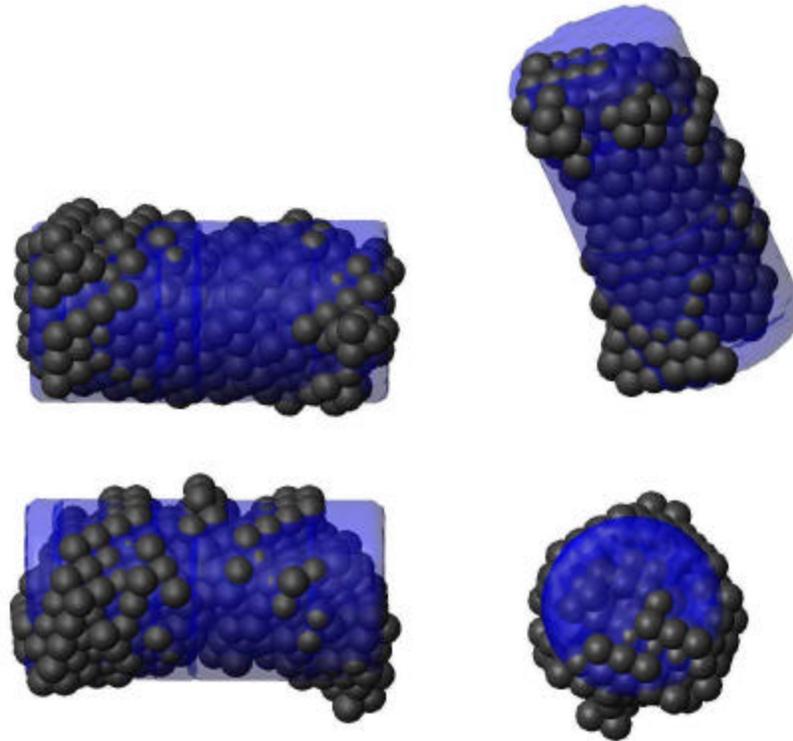
Pelo apresentado anteriormente, ambos os métodos mostram-se capazes de, em princípio, solucionar o problema proposto de reconstrução tridimensional partindo da curva de espalhamento. No entanto, como salientado exaustivamente em toda a discussão, a técnica de SAXS é de baixa resolução e baixa informação. As implicações disso ficaram evidentes quando utilizamos a expansão multipolar para restauração tridimensional. Estávamos restritos naquela análise, a uma pequena região do espaço recíproco (número baixo de Shannon channels). Com o método de elementos finitos podemos, em princípio, alcançar qualquer resolução, porém o número de Shannon channels, mesmo para medidas de alta resolução, é pequeno e novamente teremos o problema de não unicidade na solução. Isto ficará evidenciado nos exemplos a seguir. Poderemos comparar os métodos apresentados anteriormente e também avaliar a importância ou não do vínculo de empacotamento imposto por Svergun no método de *Simulated Annealing*. Tomei os mesmos modelos utilizados anteriormente, só que abrangendo uma região maior do espaço recíproco. Na realidade tomei valores típicos, alcançados em experimentos de alta resolução. Além disso, em experimentos reais, temos um erro muito grande para altos valores de  $h$  o que logicamente dificultaria ainda mais a restauração.

Nos exemplos seguintes, o envelope externo transparente foi obtido através de modelos construídos com o programa **BUILDER**. Ambos os modelos (tanto do Builder quanto o obtido pelos softwares) são compostos por subunidades esféricas, mas optamos por representar os modelos do BUILDER por envelopes transparentes para facilitar a comparação.

#### 4.3.3.1. CILINDRO

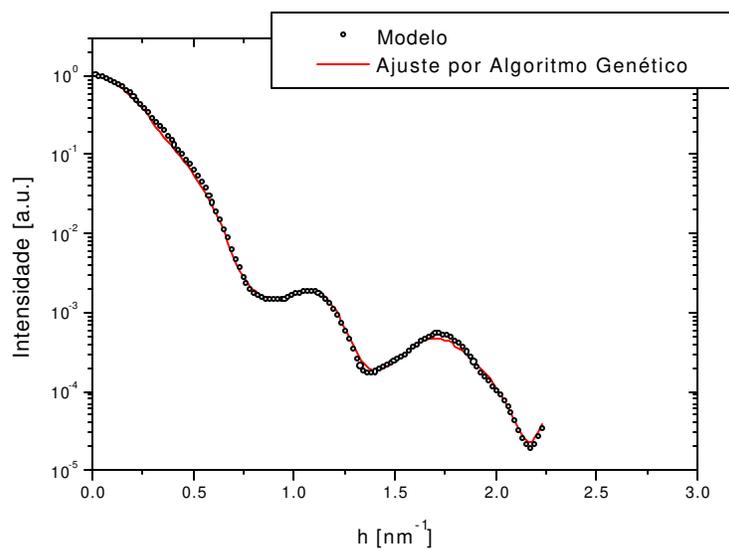
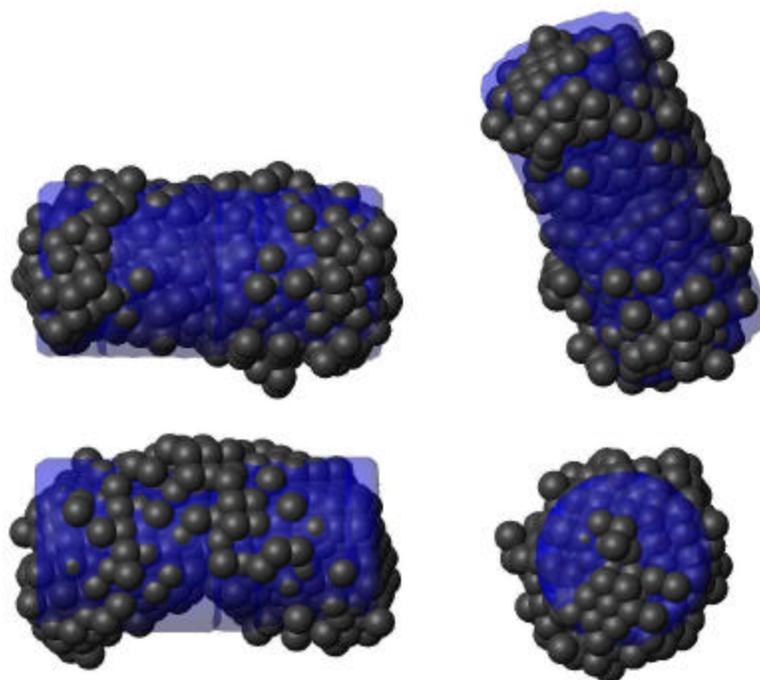
⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma distorcida obtida sobreposta ao envelope do modelo original. As diferenças no intervalo de ajuste decorrem de peculiaridades dos algoritmos dos métodos e serão discutidas posteriormente

- Simulated Annealing



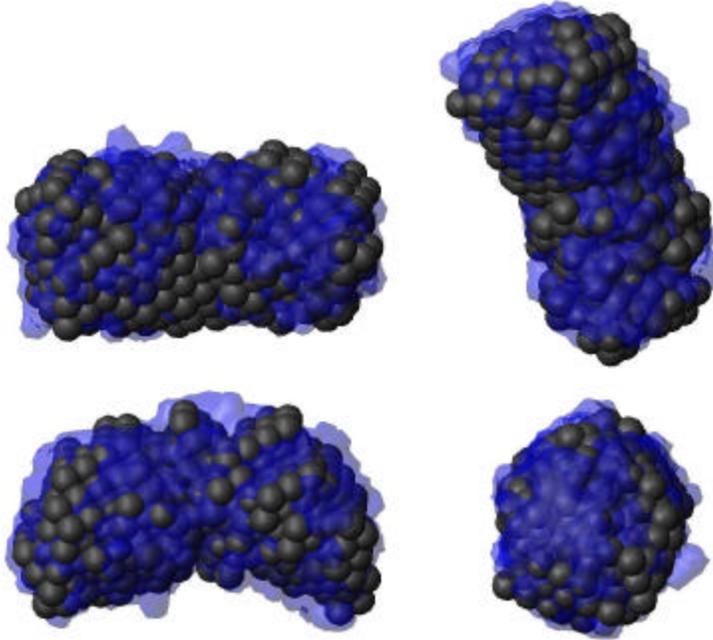
**Figura 4-10:** Restauração ab initio e ajuste obtido para cilindro utilizando simulated annealing .

- Algoritmo Genético



**Figura 4-11:** Restauração ab initio e ajuste obtido para cilindro utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso do cilindro são muito parecidas salvo pequenas variações. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

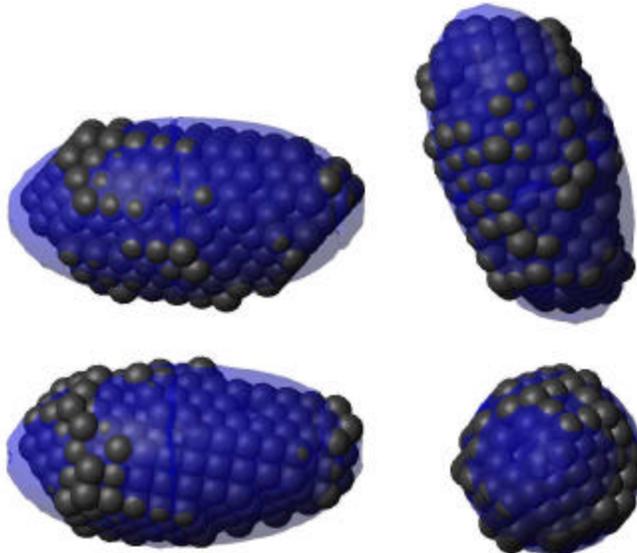


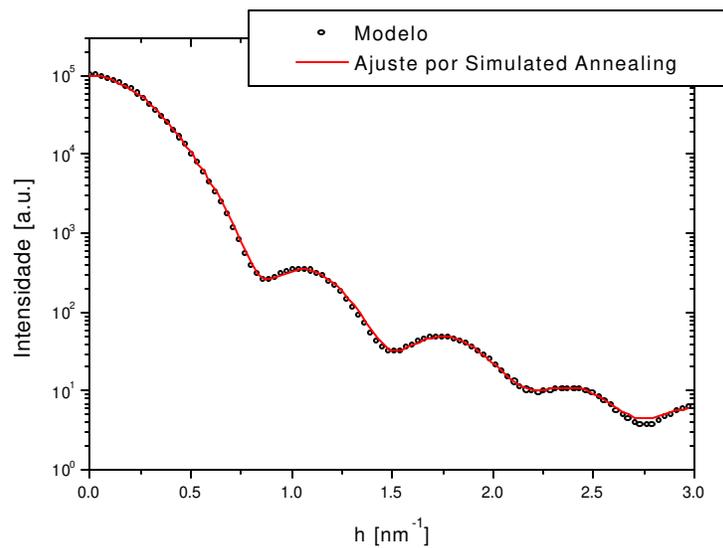
**Figura 4-12:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.2. ELIPSÓIDE PROLATO

⇒ A forma restaurada para este caso é mais próxima da estrutura real do que no caso do cilindro, apesar de em ambos os casos o ajuste da intensidade espalhada ser perfeito.

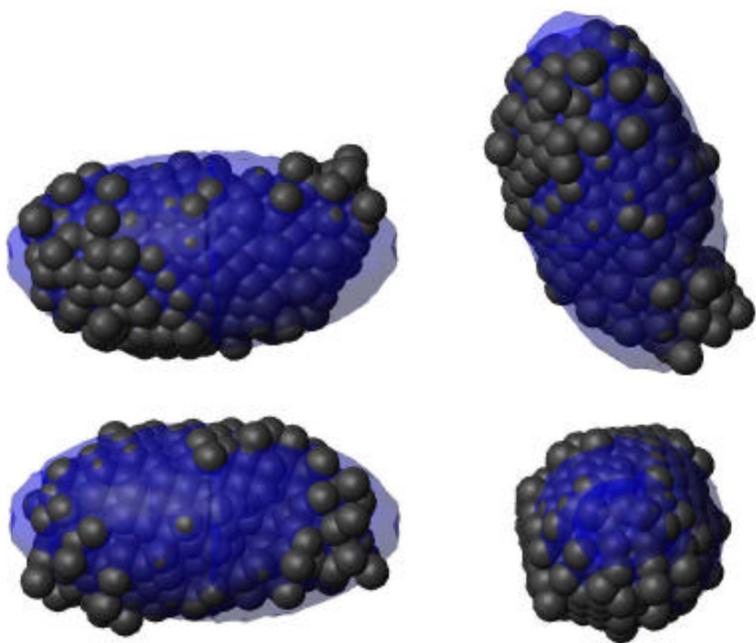
- Simulated Annealing

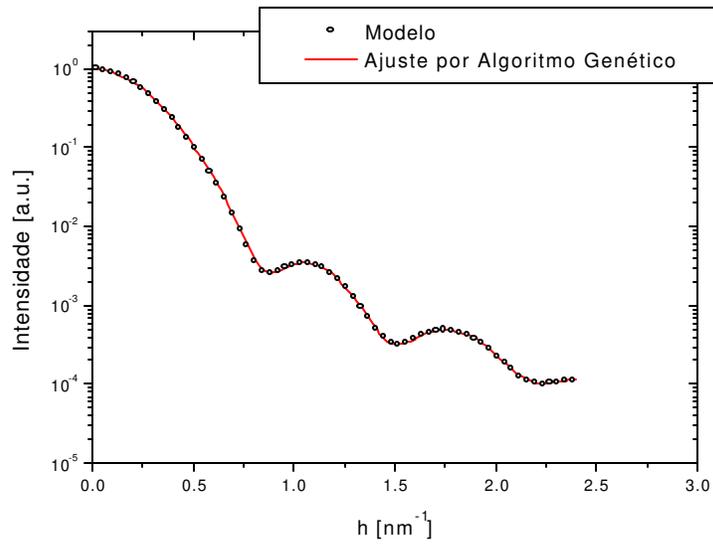




**Figura 4-13:** Restauração *ab initio* e ajuste obtido para elipsóide prolato utilizando *simulated annealing*

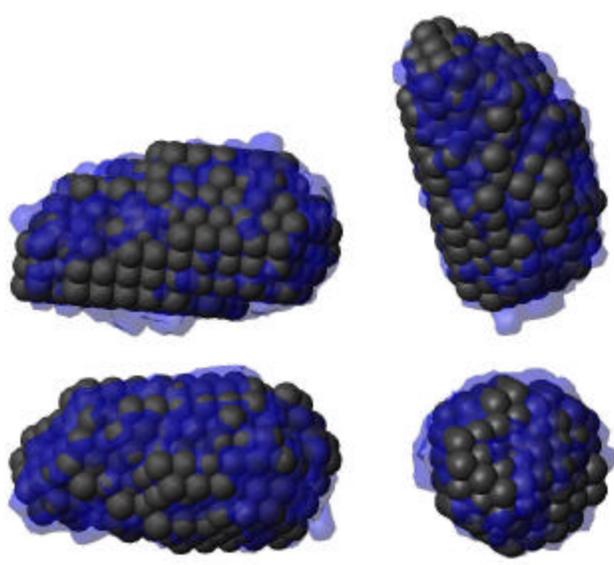
- Algoritmo Genético





**Figura 4-14:** Restauração *ab initio* e ajuste obtido para cilindro utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso do elipsóide prolato são muito parecidas salvo pequenas variações. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

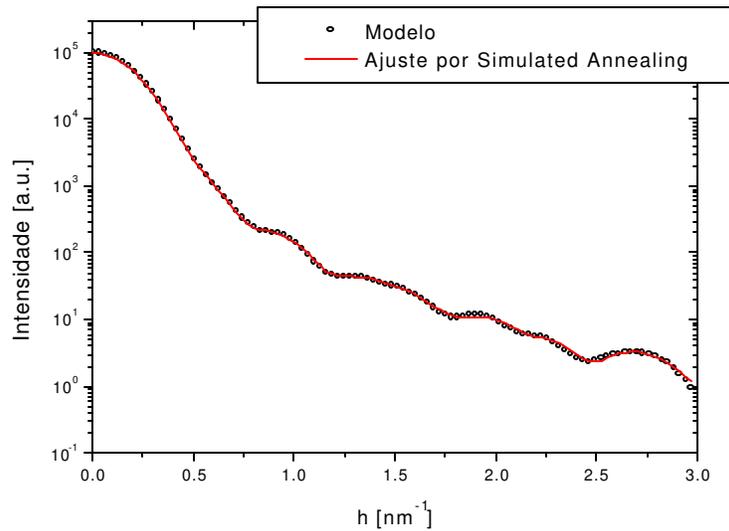
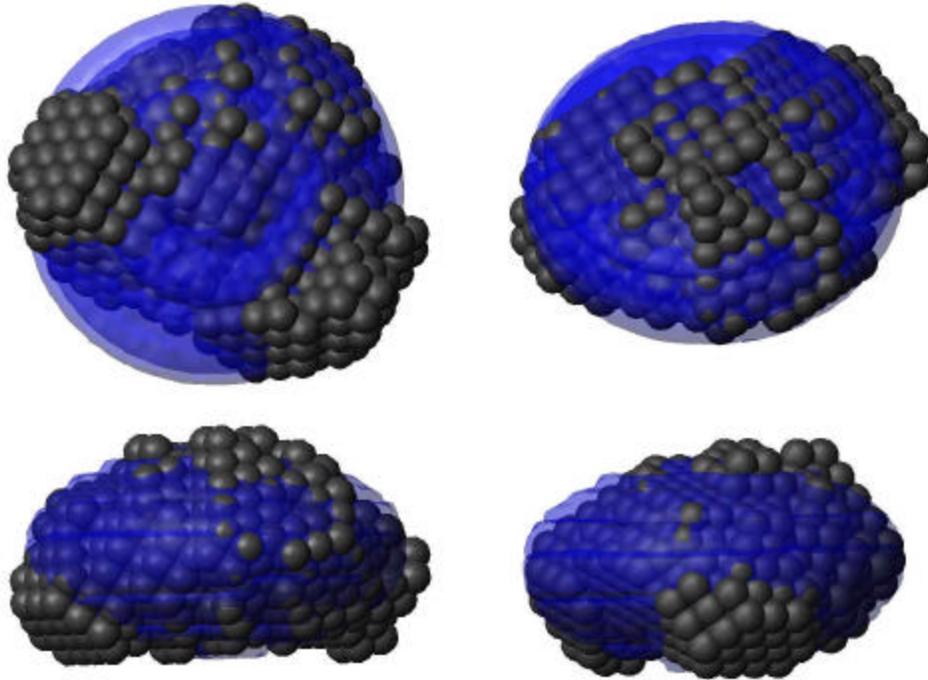


**Figura 4-15:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.3. ELIPSÓIDE OBLATO

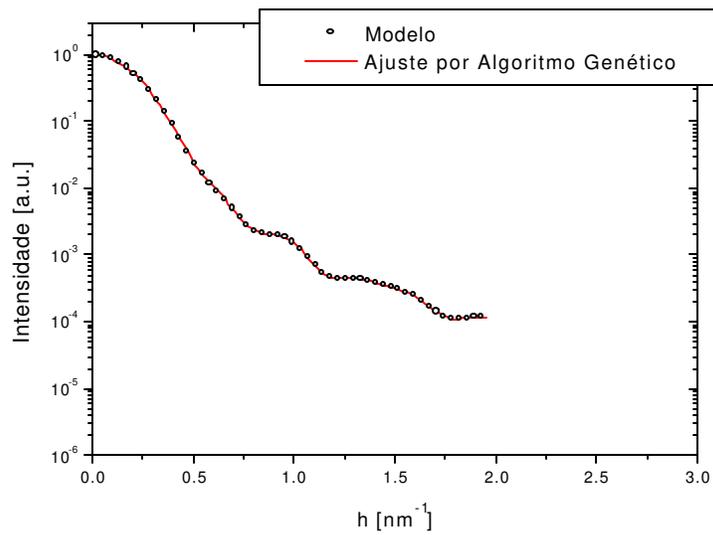
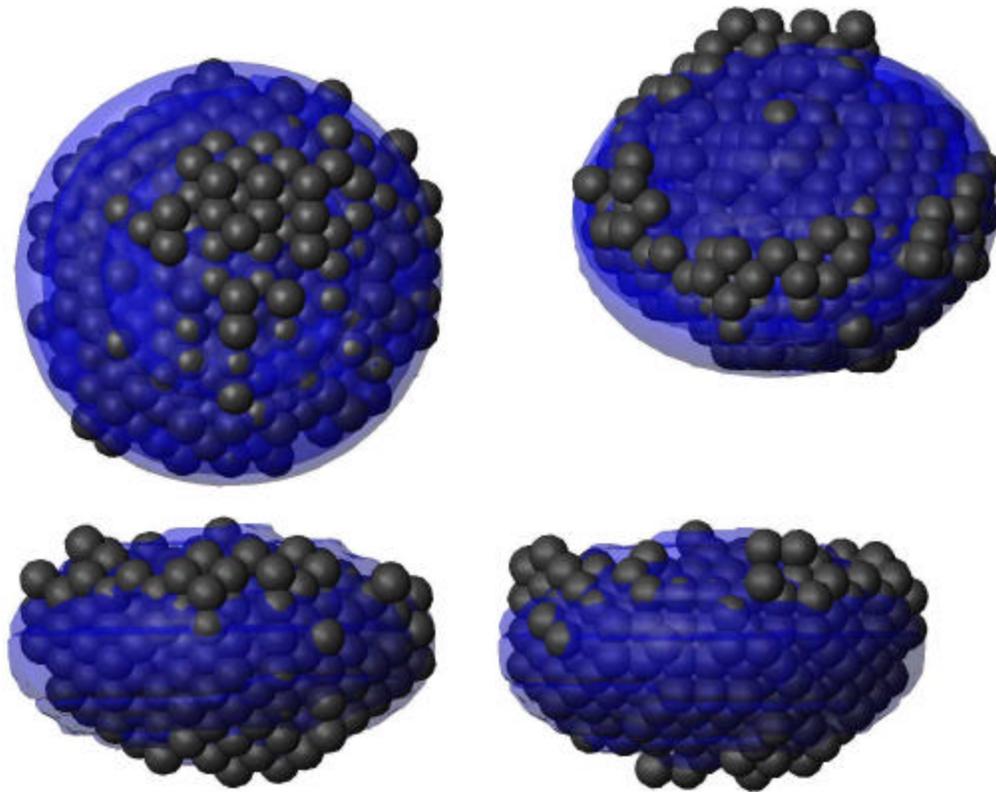
⇒ O ajuste das curvas de espalhamento tanto para o SA quanto para o GA. é perfeito enquanto que a forma obtida é bastante distorcida do modelo original.

- Simulated Annealing



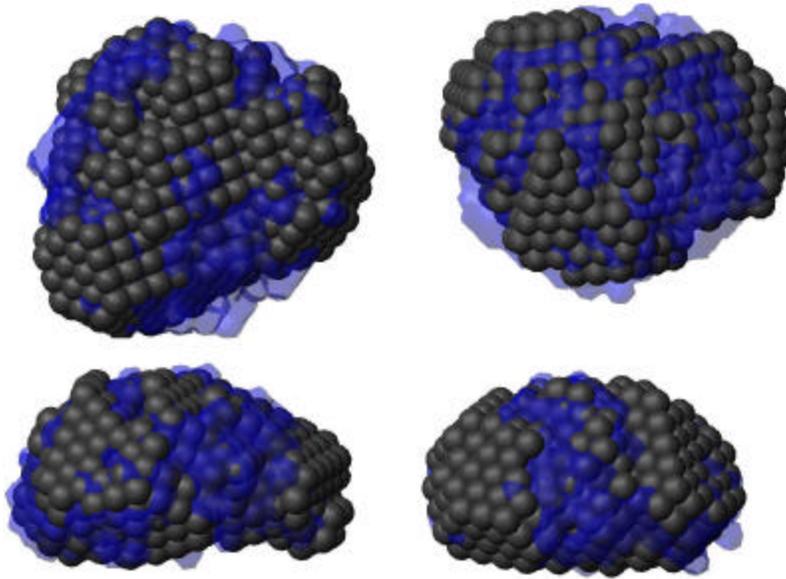
**Figura 4-16:** Restauração ab initio e ajuste obtido para elipsóide oblato utilizando simulated annealing

- Algoritmo Genético



**Figura 4-17:** Restauração ab initio e ajuste obtido para elipsóide oblato utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso do elipsóide oblato são bastante distorcidas do modelo original e entre si também diferem como vemos a seguir. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

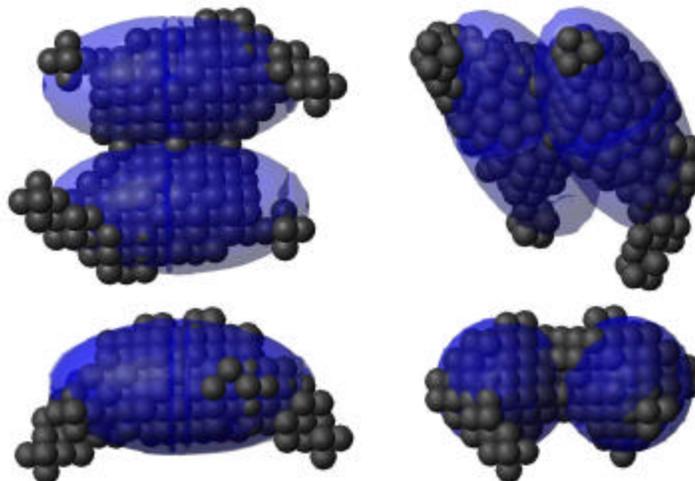


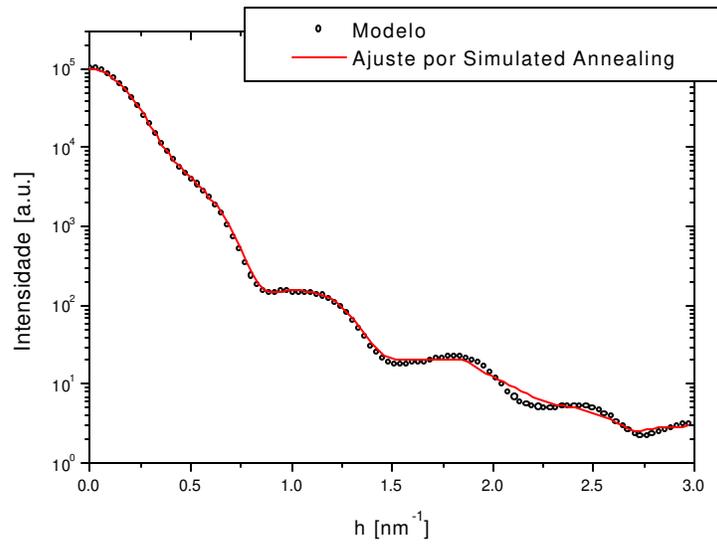
**Figura 4-18:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.4. DIMERO DE ELIPSOIDES PROLATOS

⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma distorcida obtida sobreposta ao envelope do modelo original. Para o SA utilizamos um vínculo de simetria (P2) (peculiaridade do algoritmo de otimização), enquanto que o GA não permite a inclusão de vínculos de simetria.

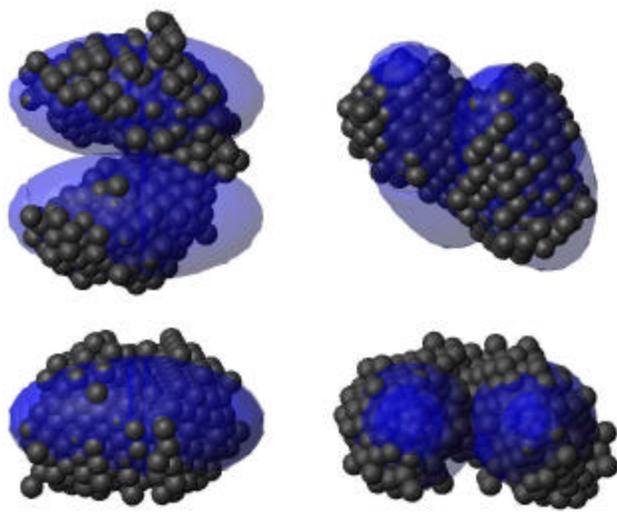
- Simulated Annealing

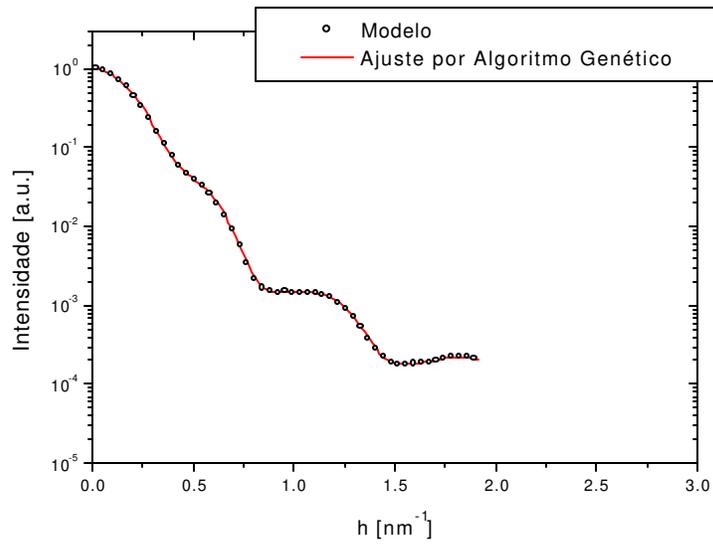




**Figura 4-19:** Restauração *ab initio* e ajuste obtido para dímero de elipsóides prolatos utilizando *simulated annealing*.

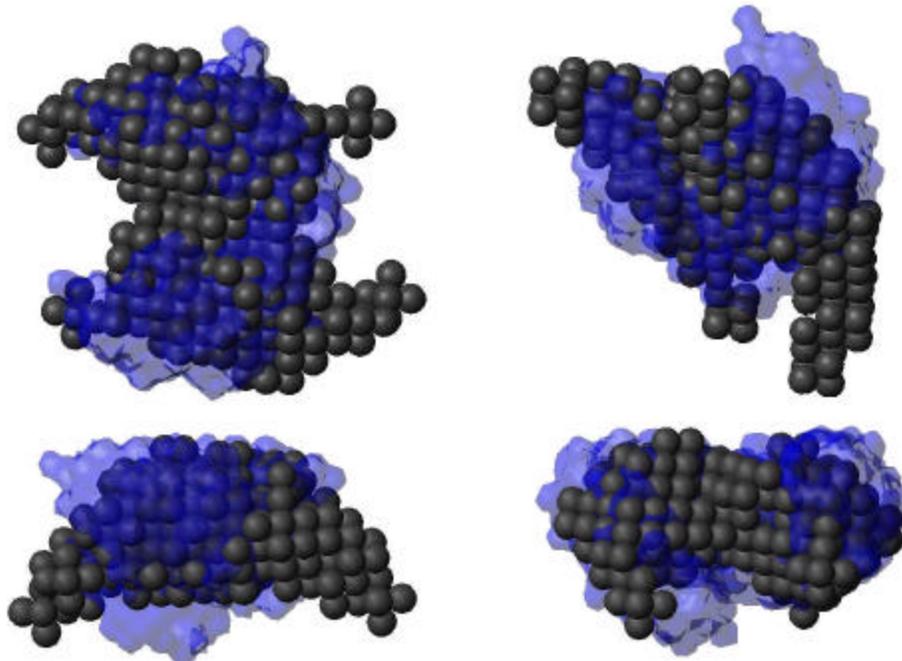
- Algoritmo Genético
  - Mesmo não incluindo o vínculo de simetria, a estrutura final apresenta um eixo de simetria 2.





**Figura 4-20:** Restauração *ab initio* e ajuste obtido para dímero de elipsóides prolatos utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso de dímero de elipsóides prolatos são bastante distorcidas do modelo original e entre si também diferem como vemos a seguir. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

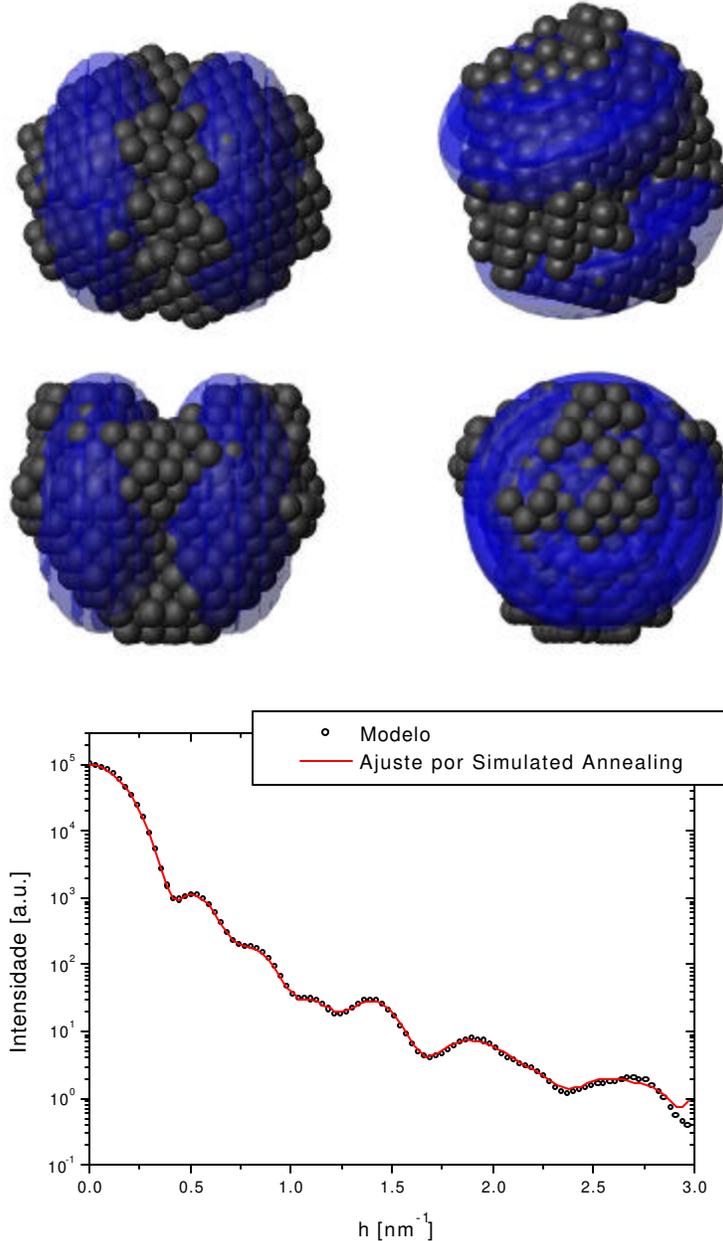


**Figura 4-21:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.5. DIMERO DE ELIPSOIDES OBLATOS

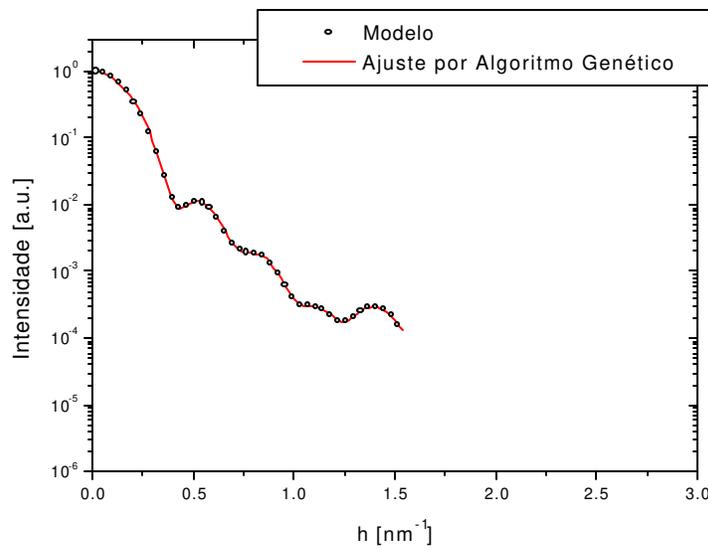
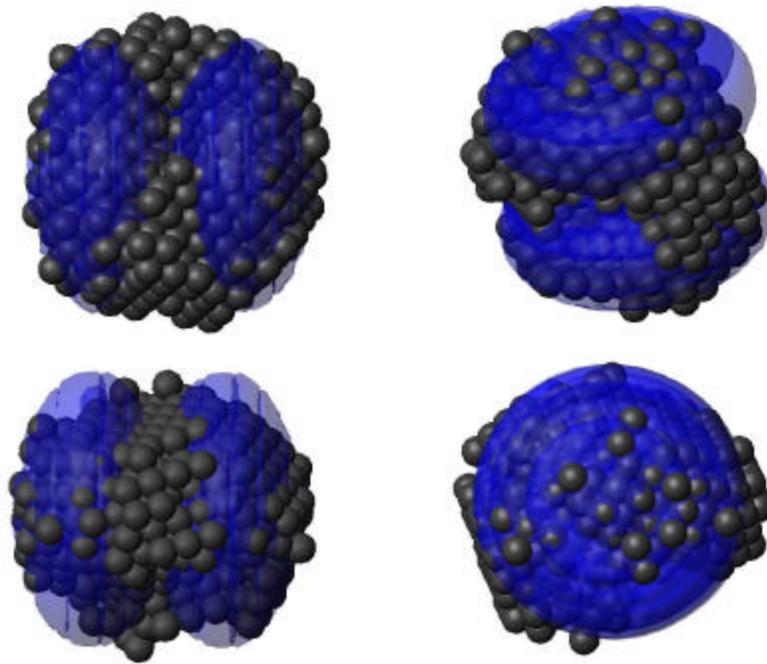
⇒ Também neste caso utilizamos um vínculo de simetria (P2) no SA (peculiaridade do algoritmo de otimização), enquanto que o GA não permite a inclusão de vínculos de simetria. Mesmo com o vínculo de simetria, a forma obtida é consideravelmente diferente do modelo, apesar do ajuste da curva de espalhamento ser perfeito.

- Simulated Annealing



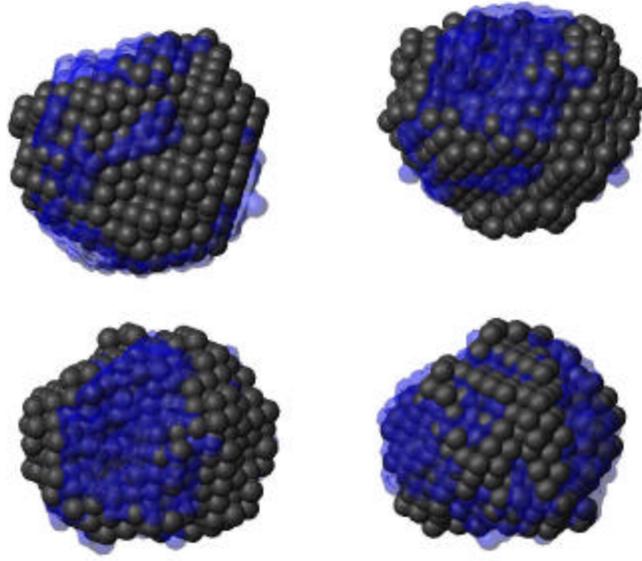
**Figura 4-22:** Restauração *ab initio* e ajuste obtido para d mero de elips ides oblato utilizando simulated annealing

- Algoritmo Genético



**Figura 4-23:** Restauração *ab initio* e ajuste obtido para dímero de elipsóides oblatos utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso de dímero de elipsóides oblatos são bastante distorcidas do modelo original e entre si também diferem como vemos a seguir. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

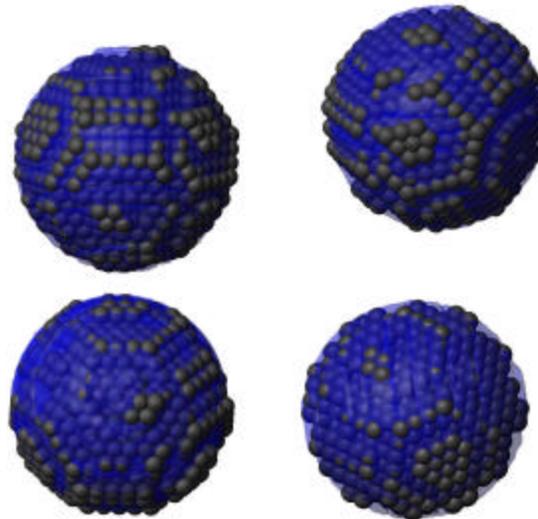


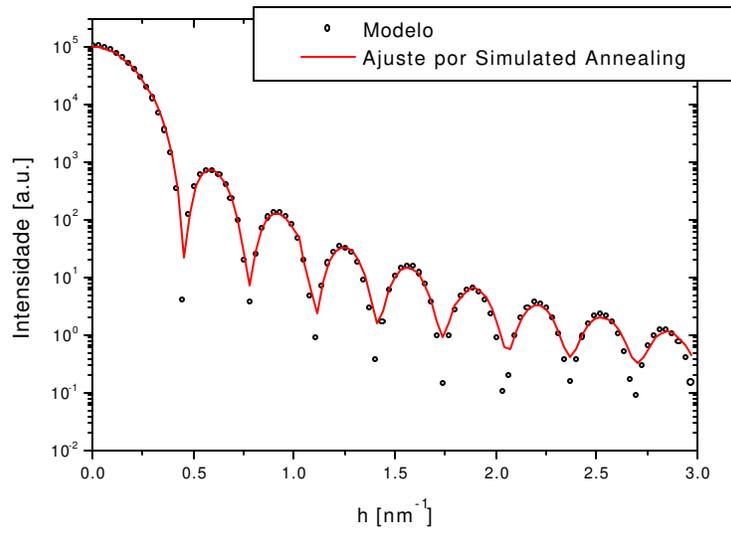
**Figura 4-24:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.6. ESFERA

⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma levemente distorcida obtida sobreposta ao envelope do modelo original. Como o espaço de busca inicial é uma esfera, a otimização neste caso é muito facilitada.

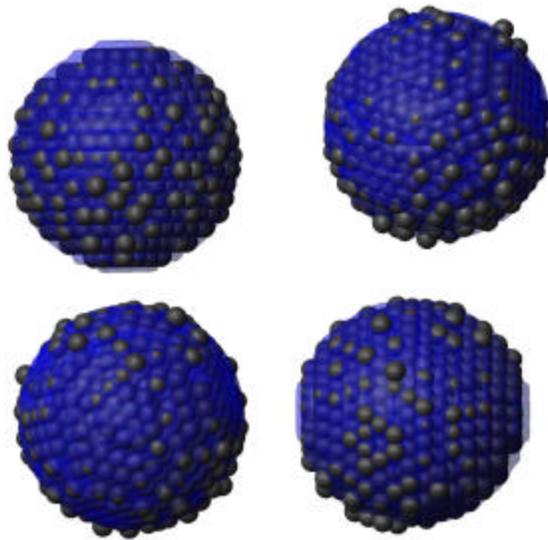
- Simulated Annealing

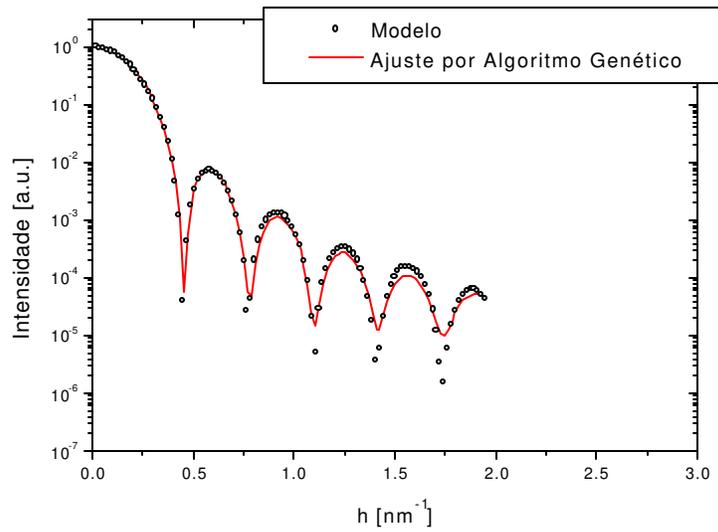




**Figura 4-25:** Restauração *ab initio* e ajuste obtido para esfera utilizando *simulated annealing*

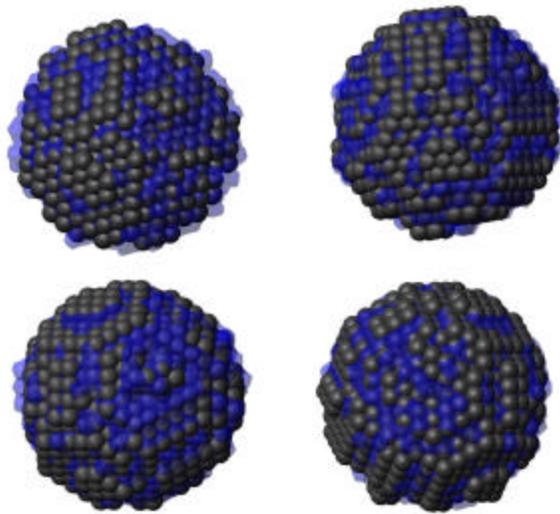
- Algoritmo Genético





**Figura 4-26:** Restauração *ab initio* e ajuste obtido para esfera utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais para o caso de esferas são muito semelhantes ao modelo original e entre si também diferem como vemos a seguir. Na figura abaixo o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA.

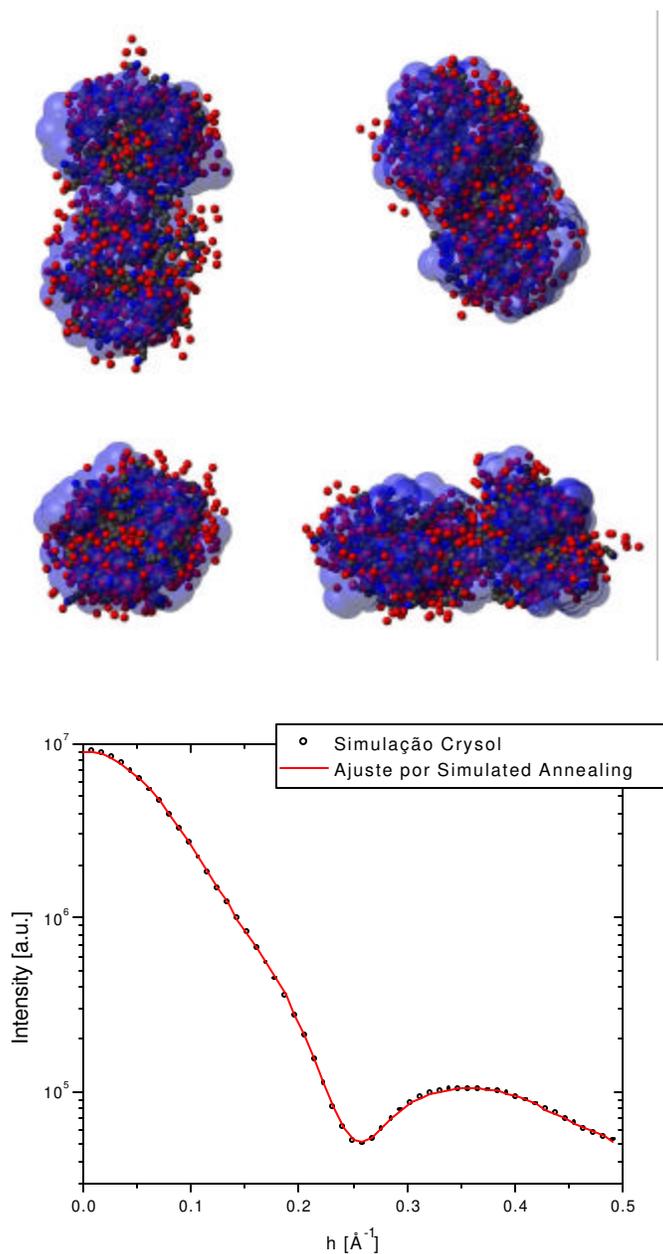


**Figura 4-27:** Sobreposição dos resultado obtidos pelos métodos.

#### 4.3.3.7. ALBUMINA

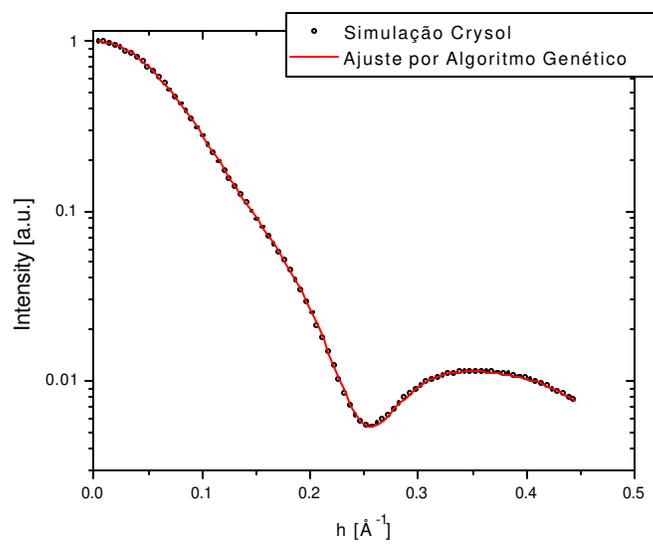
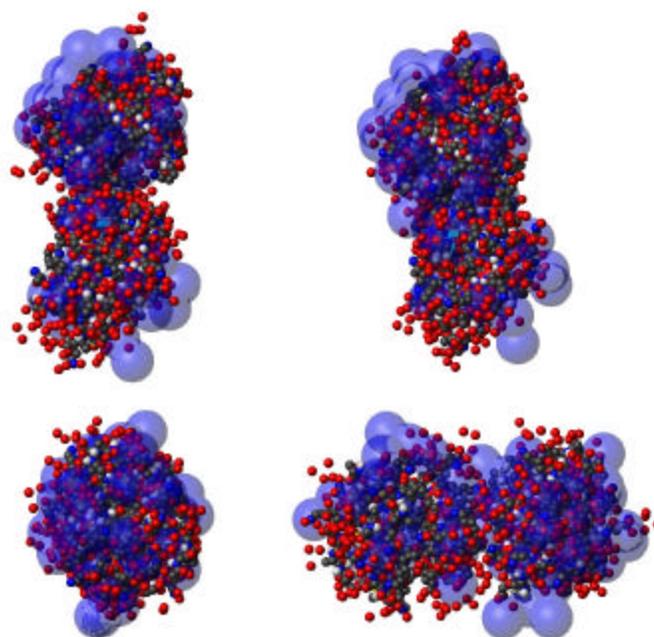
⇒ Neste exemplo, outra peculiaridade do algoritmo do SA se torna evidente, sendo esta o vínculo de compactação, que não permite que as subunidades fiquem muito separadas umas das outras. No entanto o ajuste da curva de espalhamento em ambos os casos é perfeito.

- Simulated Annealing



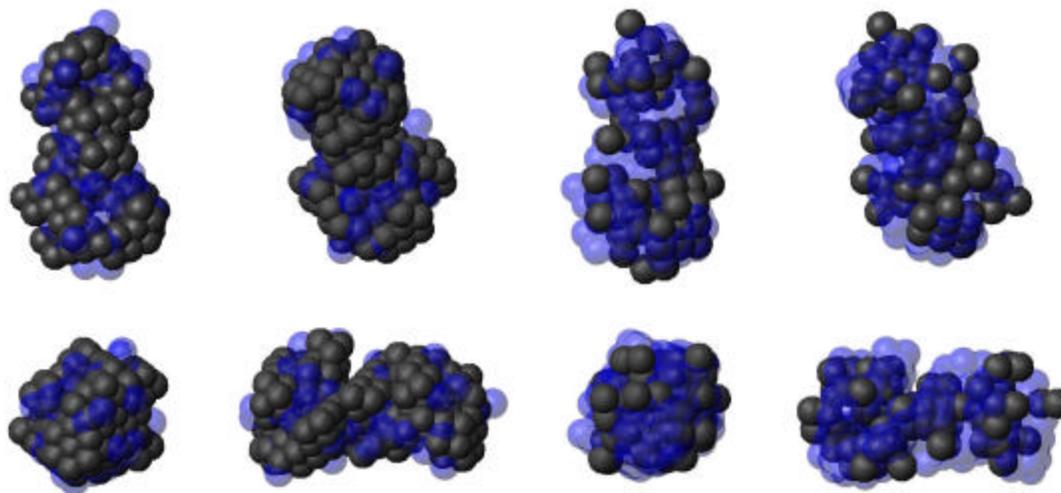
**Figura 4-28:** Restauração *ab initio* e ajuste obtido para albumina utilizando simulated annealing

- Algoritmo Genético



**Figura 4-29:** Restauração *ab initio* e ajuste obtido para albumina utilizando algoritmo genético.

- Sobreposição dos Resultados
  - As estruturas finais são bastante parecidas porém o vínculo de compactação força o modelo obtido por SA a ser mais compacto que o obtido por GA. Na figura abaixo, a esquerda, o envelope é dado pelo ajuste por GA e o modelo de esferas pelo SA. Na figura da direita as posições se invertem.



**Figura 4-30:** Sobreposição dos resultado obtidos pelos métodos.

Os exemplos anteriores demonstram claramente o que havíamos salientado em nossa discussão. Mesmo aumentando o range no espaço recíproco (aumentando o número de Shannon channels) , e obtendo um ajuste muito bom das curvas, os resultados, isto é, as formas tridimensionais restauradas representam aproximações “bizarras” das formas simuladas. Comparando-se com o obtido pelo método de expansão multipolar, podemos ter a falsa impressão de que aquele método é superior a este. Na realidade, a expansão multipolar foi favorecida pela forma simples de alguns modelos (esfera, elipsóides) que necessitam de poucos esféricos para serem descritos o que facilita em muito os cálculos. Já para os dímeros utilizamos uma propriedade do programa de ajustar simultaneamente a curva de monômero e dímero. Olhando pelo ponto de vista de informação, estamos aumentando em muito o número de Shannon channels o que permite ao software restaurar quase que perfeitamente a forma. No entanto, se não tivéssemos utilizado deste expediente, digo, inserir as duas curvas, o resultado fornecido seria completamente errôneo, o que ficou evidente para o cilindro. Voltemos agora para as simulações anteriores. Antes de explicitamente comentar os resultados, quero salientar algumas características dos programas utilizados. No *simulated annealing*, podemos inserir condições de simetria (simetria 2, 3, tetramérica, etc), além de condições de forma (desconhecido, prolato, oblato); tudo isso representa mais *informação*, que pode em muitos casos contribuir para a restauração da forma. Já o *algoritmo genético* não permite a inserção destas informações. Com isso os cálculos anteriores são particularmente interessantes para avaliar a relevância destas informações. Outro ponto que gostaria de salientar é o fato de que os ajustes para o algoritmo genético em todos os casos não atingem a mesma resolução alcançada pelo *simulated annealing*. Isto ocorre pois existe uma restrição de 2000 subunidades esféricas no algoritmo genético, para o espaço de busca. Como vimos, na discussão do algoritmo, ocorre

um processo de mascaramento onde a cada passo o diâmetro da subunidade é diminuído e com isso aumentada a resolução. Se em um destes passos o número de subunidades for maior que 2000 o programa é finalizado. Com isso, o máximo valor de  $h$  alcançado é variável para este software. Isto não ocorre no caso do *simulated annealing* pois o programa insere *todas* as subunidades de uma só vez, e o raio destas é ditado pelo máximo valor de  $h$  a ser ajustado.

Para alguns modelos, repeti o ajuste varias vezes para ver se os resultados eram reproduzíveis. Algumas subunidades se alteravam, mas a forma sempre se mantinha. Olhando todas as simulações, vemos que a forma obtida é muito semelhante para os dois métodos. Com isso ambos os métodos chegam a resultados muito próximos. Para o cilindro e o elipsóide prolato, vemos que o caráter alongado é restaurado. Como o ajuste das curvas é perfeito para a região próxima a origem, temos a certeza de que o volume e raio de giro do objeto restaurado são idênticos ao modelo simulado. Isto é de grande utilidade para a análise de macromoléculas em solução. A técnica de SAXS por si só, é incapaz de restaurar a forma exata da partícula espalhadora, no entanto fornece indicações importantíssimas sobre a forma da mesma e, a possibilidade de partindo da curva de espalhamento, independente de termos pouca informação e resolução, obtermos uma *aproximação* para a forma é algo notável. Logicamente que é uma aproximação, e por isso, muitos cuidados devem ser tomados na análise dos resultados.

Para o cilindro e os elipsóides, os resultados são muito parecidos como vemos na sobreposição dos resultados. Os ajustes são praticamente perfeitos e o fato do método do algoritmo genético não alcançar a máxima resolução possível não influencia na forma obtida. Isso é esperado, pois na curva de espalhamento a informação sobre a *forma* geral da partícula está confinada próxima a origem. O aumento da resolução acrescentará mais e mais peculiaridades que auxiliam na “lapidação” desta forma.

Para os dímeros de elipsóides (onde utilizamos simetria 2 no *simulated annealing*), vemos que a restrição de simetria melhora a restauração, mas não de forma significativa. Na realidade o algoritmo genético consegue evidenciar esta simetria para dímero de elipsóides prolatos (Figura 4-20), porém a forma restaurada pelo outro método é muito melhor. Para o caso de elipsóides oblatos ambos os resultados são bastante similares e o vínculo de simetria não é relevante. Isso é compreensível uma vez que este dímero pode facilmente se transformar em, digamos, um cilindro por exemplo, o que torna muito difícil sua restauração pelos algoritmos.

Para a esfera, ambos os algoritmos restauram quase que perfeitamente a forma. Isso de certo modo é esperado, pois em todos os casos o espaço de busca inicial é uma esfera de diâmetro igual ou superior à obtida para a máxima dimensão da partícula. Deste modo o processo é facilitado e como vemos na sobreposição, as esferas obtidas são praticamente idênticas.

Em último temos a restauração da albumina. Se retomarmos a restauração feita pela expansão multipolar, vemos que pelo fato da proteína possuir uma forma muito complexa tornava o resultado muito grosseiro, visto que a expansão multipolar não alcança alta resolução. Esperaríamos que estes métodos estocásticos apresentados até agora pudessem restaurar a forma de modo mais preciso que a anterior e, como vemos nas Figura 4-28 e 4-29 isto ocorre. A forma alongada da proteína é obtida em ambos os casos e o ajuste da curva é praticamente perfeito. Como se trata de uma proteína pequena não poderíamos simular em uma região maior do espaço recíproco, pois não representaria um caso real. Neste exemplo fica claro a importância da condição de compactação imposta por Svergun

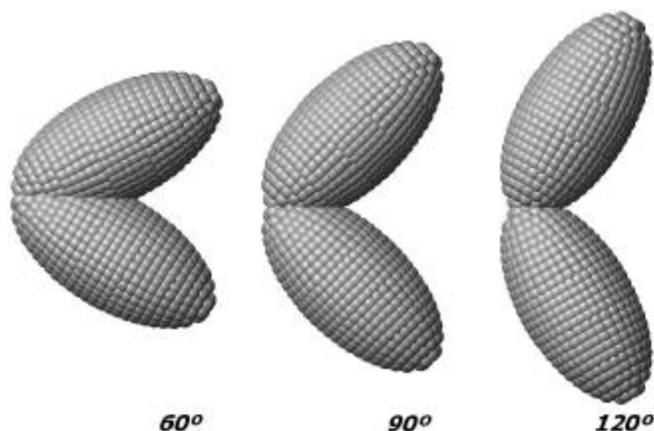
no *simulated annealing* apresentado anteriormente. O objeto restaurado por este método é muito mais compacto do que o caso do algoritmo genético. Nas simulações anteriores, isso não ficou evidente pois as subunidades eram pequenas (alta resolução) e esta compactação era automaticamente feita. Já neste caso, temos poucas subunidades e isso torna mais importante esta restrição. Na realidade, em proteínas biológicas globulares, as estruturas são bastante compactas e é válida esta imposição. Podemos encarar esta imposição como a adição de *informação* ao sistema, o que é de suma importância para o resultado final. O ajuste é perfeito independente desta restrição, no entanto, visto a não unicidade da forma obtida, a imposição da compactação torna a forma mais próxima da real, no caso de proteínas em solução.

De tudo que apresentamos até agora, podemos tirar algumas conclusões sobre a perspectiva de reconstruir-se a forma de objetos partindo da curva de espalhamento. Dos resultados anteriores vemos que a forma restaurada é, na maioria dos casos, uma aproximação grosseira da partícula espalhadora. No entanto é notável perceber que as características relevantes a uma análise conformacional do objeto espalhador são fornecidas, raio de giro, volume, características gerais da forma, etc; e ambos os métodos conseguem resgatar estas informações. Com relação aos algoritmos em si, não existe um método direto para a comparação. Porém se assumirmos que o *simulated annealing* e o algoritmo genético possuem a mesma velocidade de convergência, a implementação realizada por Svergun possui uma vantagem fundamental, ressaltada anteriormente, porém sem comentários: em seu método, no processo de ajuste, ele altera a posição de uma subunidade (na realidade, muda seu peso de 0 para 1) e recalcula a curva de espalhamento alterando *apenas* a contribuição *daquela* subunidade; já o algoritmo genético, a cada passo, calcula a contribuição de *todas* as subunidades novamente. Para espaços de busca com poucas subunidades (digamos  $N < 200$ ) isto não confere grandes diferenças entre os métodos, isto é, ambos convergem aproximadamente no mesmo tempo (no caso da albumina por exemplo). Porém para maiores valores de  $N$ , a diferença começa a ficar relevante chegando a ponto de que, nos exemplos anteriores, o método de Svergun mostrasse de 4 a 5 vezes mais rápido que o outro, e essa diferença cresce com o número de subunidades. Utilizando um Pentium III, 700 MHz e 256MB de RAM, o *simulated annealing* realizou os cálculos anteriores (no caso dos modelos construídos pelo Builder) em um intervalo de 8 horas, em média, enquanto que o *algoritmo genético* o fez em um intervalo de aproximadamente 35 horas. Isto se deve ao fato de termos utilizado um grande range do espaço recíproco, o que forneceu um número de *Shannon Channels* em torno de 20 - 30 (muito difícil de se obter em experimentos reais) e que por isso, exigia um grande número de subunidades.

Agora que sabemos as limitações destes métodos na reconstrução tridimensional, realizaremos um teste interessante, simulando mudanças conformacionais de uma proteína hipotética. Este exemplo é mostrado a seguir.

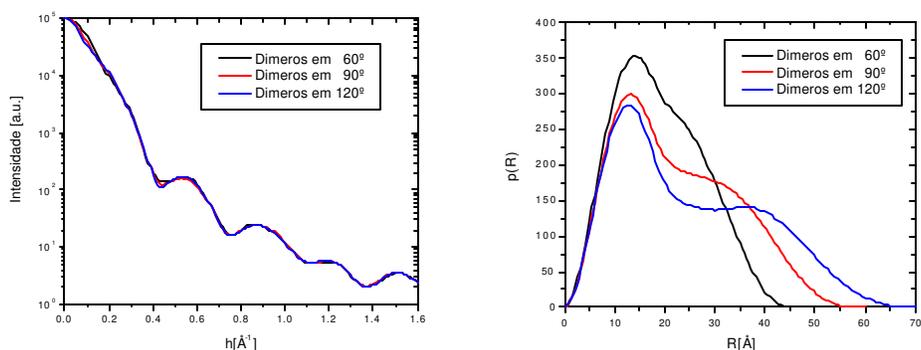
#### 4.3.3.8. MUDANÇA DE ÂNGULO ENTRE DÍMEROS

Imaginemos que uma proteína qualquer possua uma forma dimérica que varia a angularidade entre as subunidades de acordo com o pH, por exemplo. O processo é esquematizado abaixo.



**Figura 4-31:** Esquema da variação angular entre monômeros componentes de um dímero.

Como veremos, a variação do ângulo gera pequenas alterações na curva de espalhamento principalmente próxima a origem. Isso é esperado, pois a mudança no raio de giro é percebida nesta região. Para ângulos maiores as curvas são praticamente idênticas uma vez que os monômeros são os mesmos. Já havíamos citado anteriormente como uma conformação dimérica pode ser visualizada na curva de distribuição de distâncias,  $p(R)$ . A seguir temos a curva de espalhamento e a  $p(R)$  para o sistema anterior,



**Figura 4-32:** Intensidade espalhada e curva de distribuição de distâncias para dímeros em diferentes ângulos.

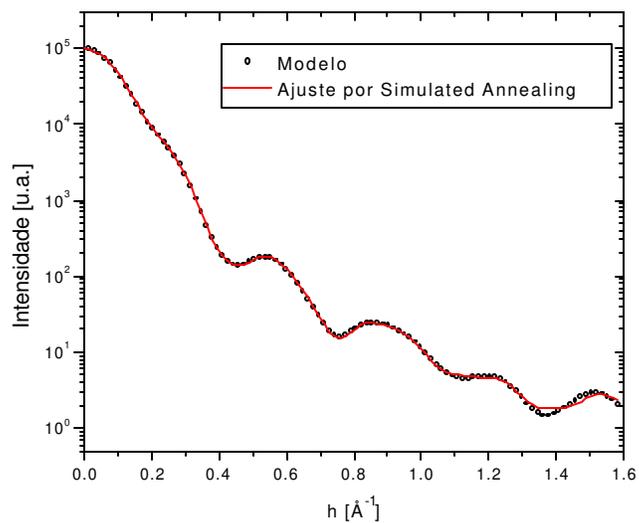
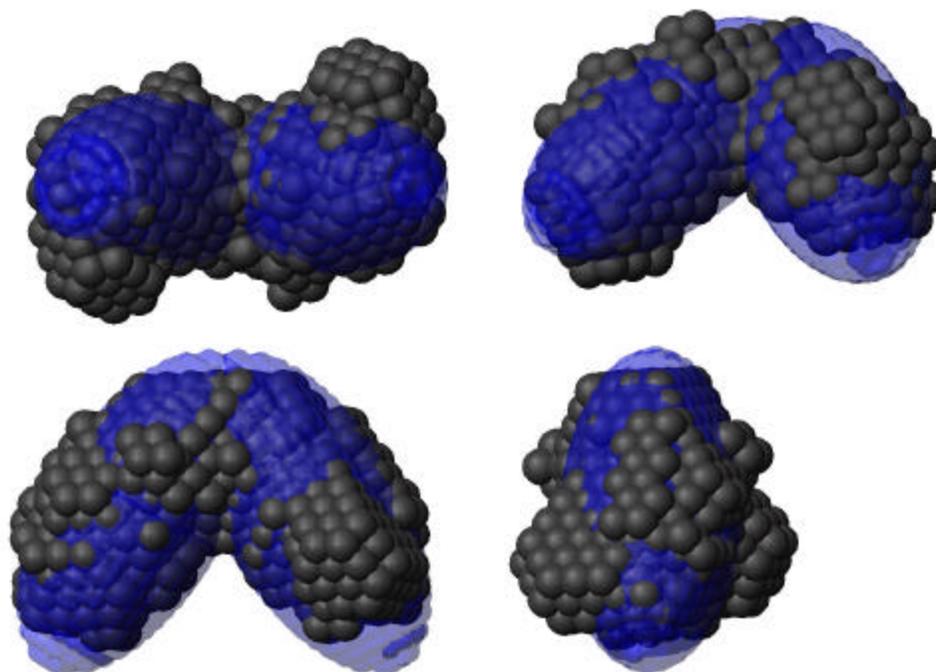
As pequenas alterações no início da curva de espalhamento conferem grandes alterações na  $p(R)$ . Estas alterações que nos permitem vislumbrar as mudanças conformacionais no sistema em questão. Um fato interessante é que as mudanças fundamentais concentram-se no início da curva de espalhamento. Sendo assim, em experimentos reais, o aparato experimental deveria permitir atingir-se ângulos bem pequenos de espalhamento. Caso isso não ocorra, não se faz possível a detecção destas variações conformacionais.

Inserindo estas curvas de espalhamento como entrada para os programas de restauração tridimensional, obtemos os resultados a seguir,

### A) DÍMERO EM 60°

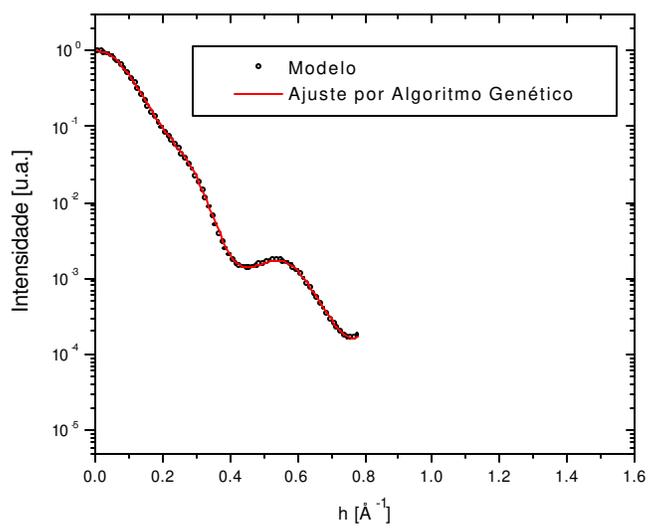
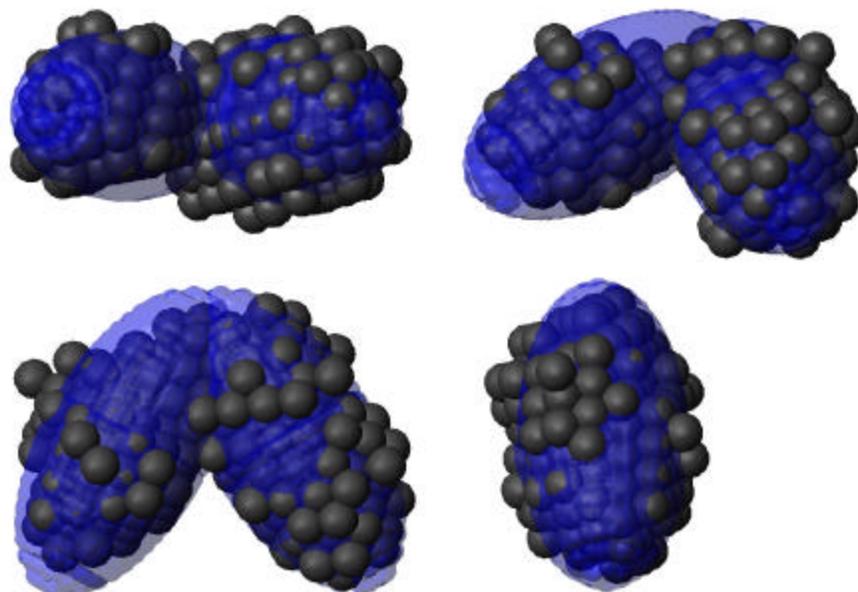
⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma levemente distorcida obtida sobreposta a estrutura. No entanto a angularidade é detectada por ambos os métodos. No SA utilizamos simetria P2.

- Simulated Annealing



**Figura 4-33:** Restauração *ab initio* e ajuste obtido para dímerno em 60° utilizando simulated annealing. Na simulação utilizamos simetria P2

- Algoritmo Genético

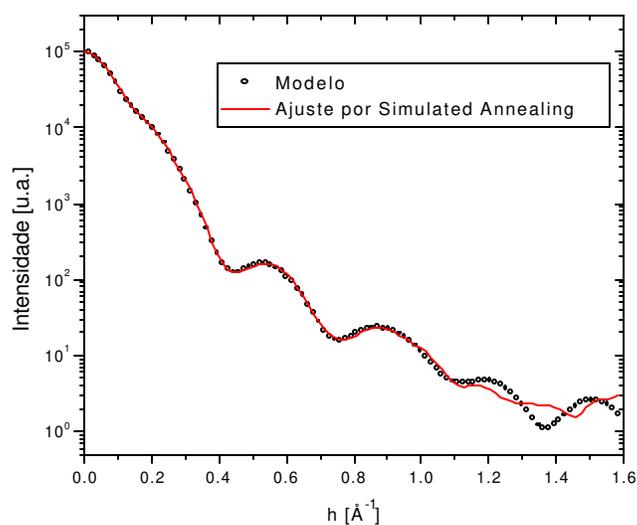
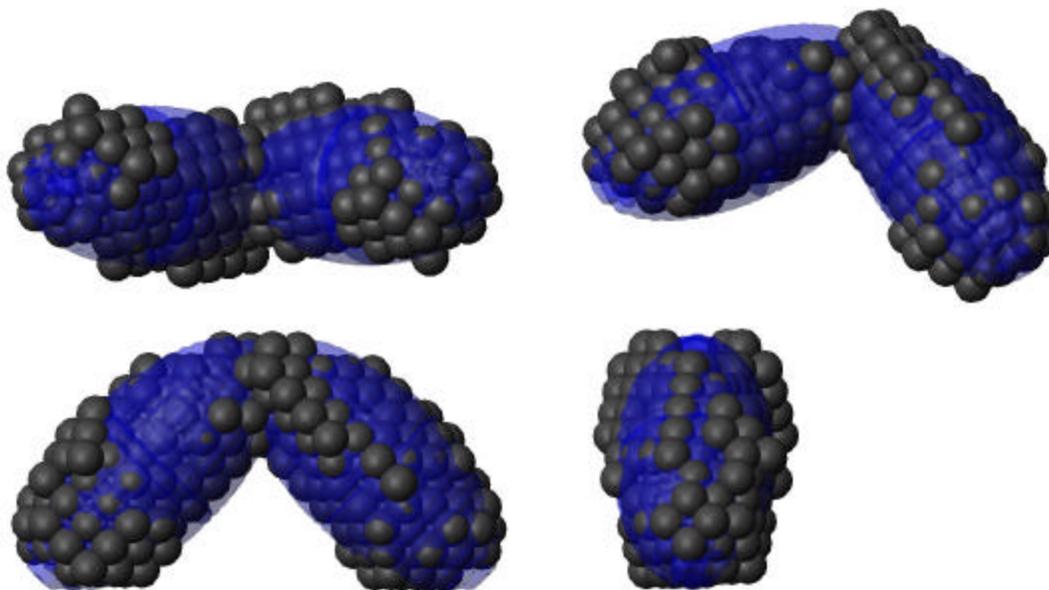


**Figura 4-34:** Restauração *ab initio* e ajuste obtido para d mero em  $60^\circ$  utilizando algoritmo gen tico.

## B) DÍMERO EM 90°

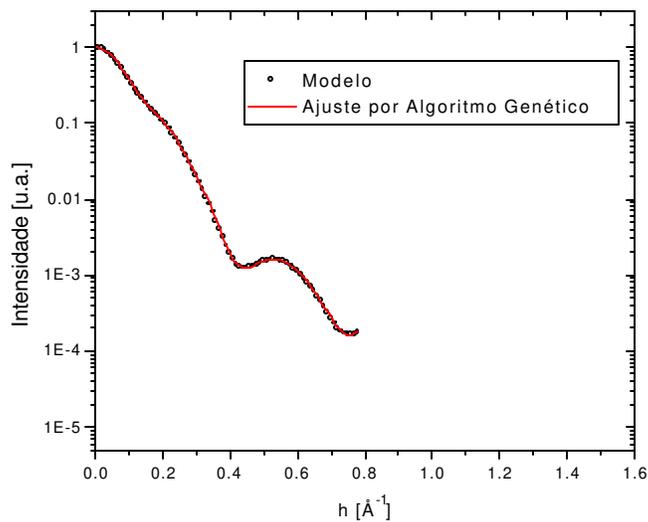
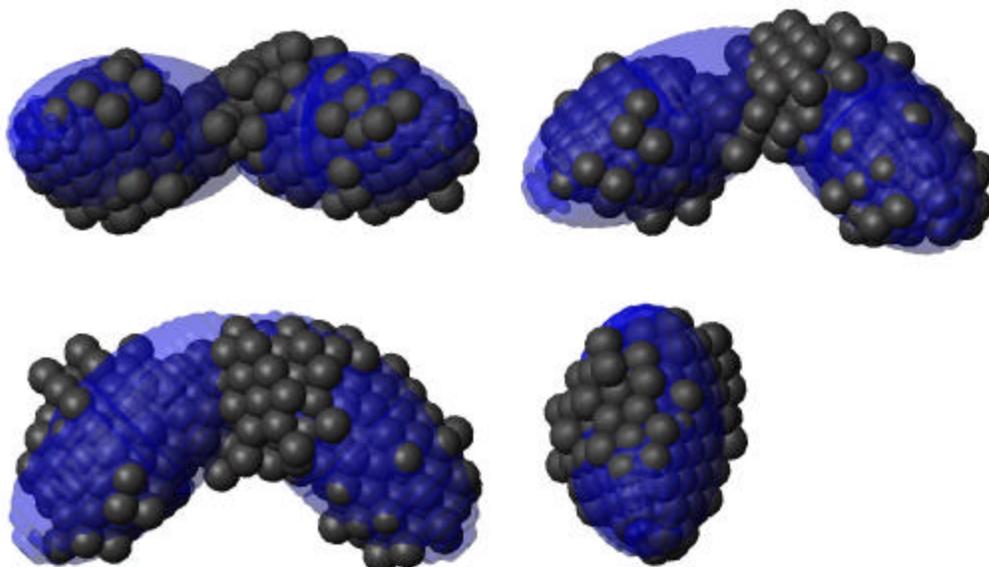
⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma levemente distorcida obtida sobreposta a estrutura. No entanto a angularidade é detectada por ambos os métodos. No SA utilizamos simetria P2.

- Simulated Annealing



**Figura 4-35:** Restauração *ab initio* e ajuste obtido para dímero em 90° utilizando simulated annealing. Na simulação utilizamos simetria P2

- Algoritmo Genético

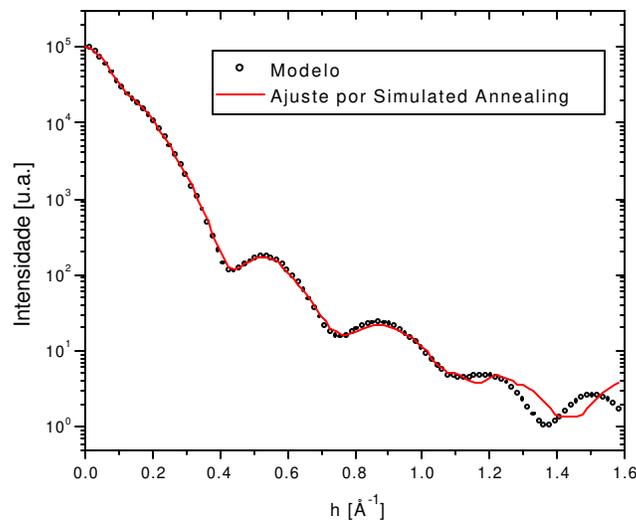
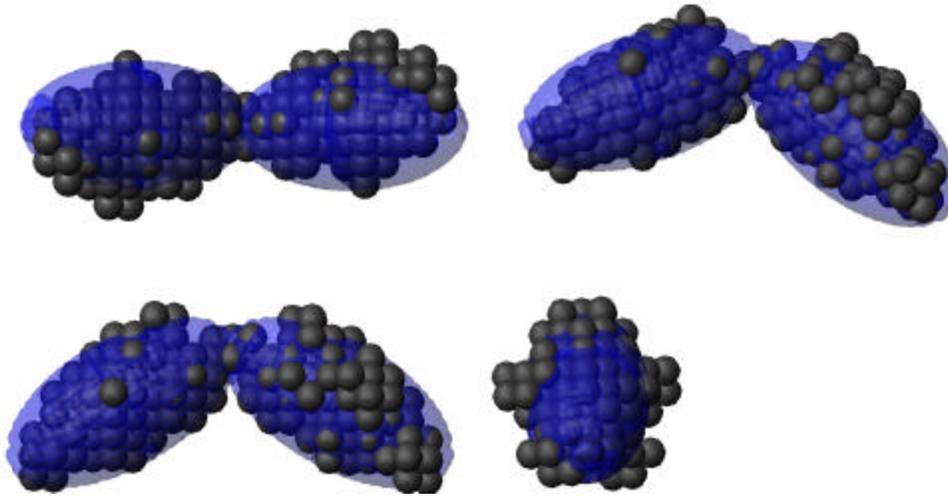


**Figura 4-36:** Restauração ab initio e ajuste obtido para dímero em  $90^\circ$  utilizando algoritmo genético.

### C) DÍMERO EM 120°

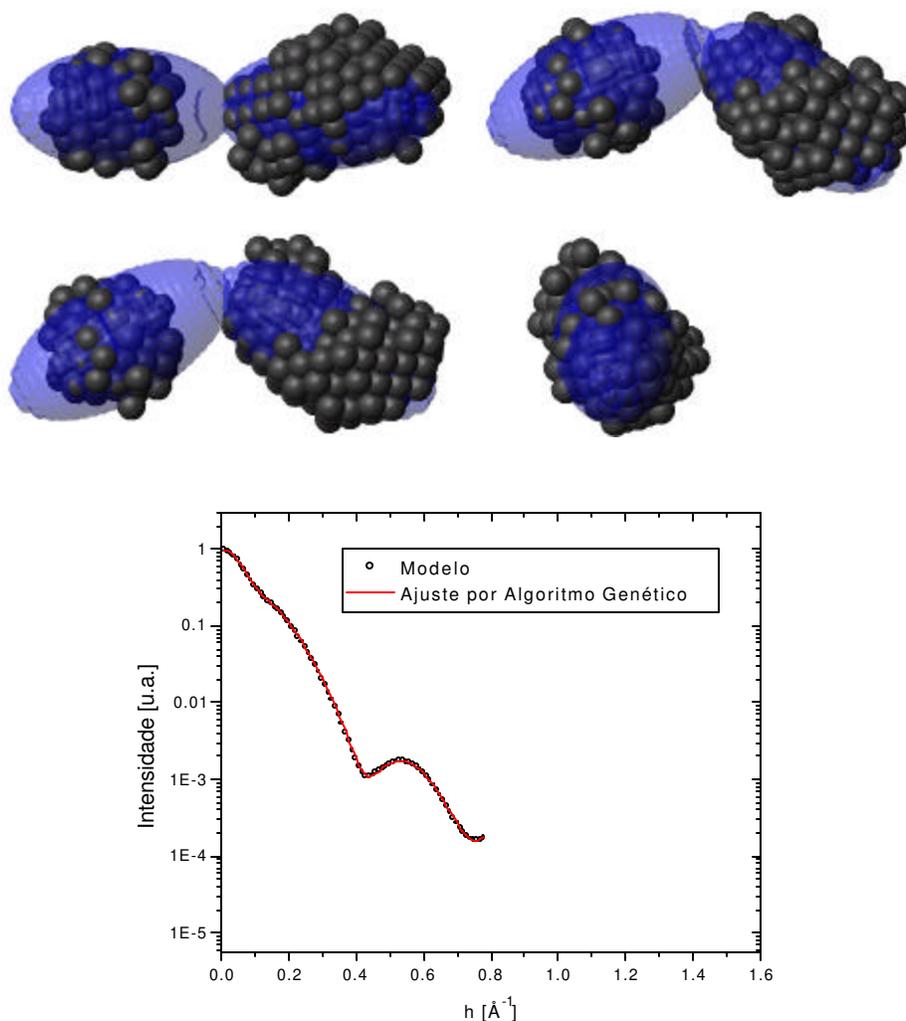
⇒ Note o perfeito ajuste das curvas de espalhamento tanto para o SA quanto para o GA., em comparação com a forma levemente distorcida obtida sobreposta a estrutura. No entanto a angularidade é detectada por ambos os métodos. Um fato interessante é notar que o modelo obtido pelo GA apareceu seccionado. Isso decorre diretamente do fato de não ter-se um vínculo de compactação. No SA utilizamos simetria P2.

- Simulated Annealing



**Figura 4-37:** Restauração *ab initio* e ajuste obtido para dímero em 120° utilizando simulated annealing. Na simulação utilizamos simetria P2

- Algoritmo Genético



**Figura 4-38:** Restauração *ab initio* e ajuste obtido para dímero em 90° utilizando algoritmo genético.

Novamente o método do algoritmo genético não conseguiu ajustar toda a curva de espalhamento, devido aos mesmos motivos citados anteriormente. Neste exemplo, fica clara a importância da condição de simetria e vínculo de compactação no resultado obtido. É surpreendente como ambos os métodos conseguem resgatar o ângulo entre os dímeros. Logicamente que a forma elipsoidal dos monômeros é apenas aproximada, mas a variação angular dos mesmos é reproduzida. No entanto, o *simulated annealing* apresenta vantagens claras, uma vez que permitindo a inserção de simetria dois e o vínculo de compactação, restaura de modo muito melhor os dímeros, independente do ângulo entre eles. Já o

algoritmo genético, apesar de o ajuste da curva de espalhamento ser perfeito ao nível de resolução conseguido, não fornece resultados tão pronunciados. Para 60 e 90 graus a forma restaurada, apesar de não ser tão boa quanto do outro método, consegue evidenciar os dímeros e os ângulos entre eles. No entanto para 120 graus, a ausência do vínculo de compactação inseriu aberrações enormes ao resultado. Como apresentado na apresentação dos métodos, este algoritmo vai gradativamente diminuindo o raio das subunidades e assim, aumentando a resolução. Porém, a ausência do vínculo de compactação permitiu que, em um dos passos, as subunidades ficassem desconectadas e, no processo seguinte de mascaramento, gerou-se dois núcleos separados. Se acompanharmos os eixos de simetria dos núcleos, o ângulo entre os dímeros é restaurado, mas em experimentos reais não temos essa informação, isto é, não saberemos que se trata de uma aberração do programa e isso pode levar a análises errôneas dos resultados. Estas discussões em conjunto com as anteriormente apresentadas nos permitem saber as limitações destes métodos estocásticos e as diferenças fundamentais entre eles na restauração tridimensional do objeto espalhador.

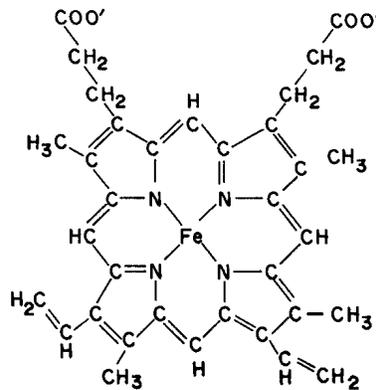


## 5. APLICAÇÃO DOS MÉTODOS DA RECUPERAÇÃO DE ESTRUTURAS USANDO DADOS EXPERIMENTAIS

Aplicaremos os métodos estocásticos apresentados anteriormente em um problema real: a obtenção da forma de hemoglobinas extracelulares gigantes

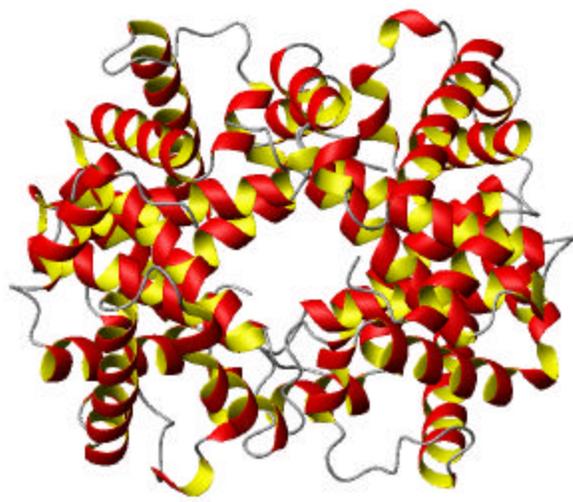
### 5.1. HEMOGLOBINAS EXTRACELULARES

Hemoglobinas são proteínas respiratórias que contém grupos heme. Tais grupos são constituídos por uma parte orgânica (protoporfirina IX) e um átomo de ferro. Conforme a figura abaixo, vemos que o Fe fica no centro do anel tetrapirrótico da protoporfirina, fazendo ligações com os 4 átomos de nitrogênio. É ele que vai se ligar de forma reversível ao oxigênio.



**Figura 5-1:** Grupo Heme (Fé – protoporfirina IX)

As hemeoproteínas podem ser intra ou extracelulares. As primeiras são normalmente moléculas pequenas, com massa molecular menores que 70000. São encontradas, por exemplo, nos vertebrados, onde se apresentam invariavelmente na forma tetramérica, com 4 grupos heme e massa molecular da ordem de 64500.



**Figura 5-2:** Hemoglobina tetramérica humana.

As extracelulares, também chamadas de eritrocruorinas, são encontradas nos invertebrados e podem ser classificadas em 4 grupos distintos de acordo com suas estruturas de subunidades. Em particular as eritrocruorinas, que tem várias subunidades muitas das quais unidas por pontes dissulfeto e algumas sem grupo heme, caracterizam-se por uma aparência de 2 camadas hexagonalmente simétricas e distinguem-se pela elevada massa molecular que varia entre 3500 e 4000 kDa.

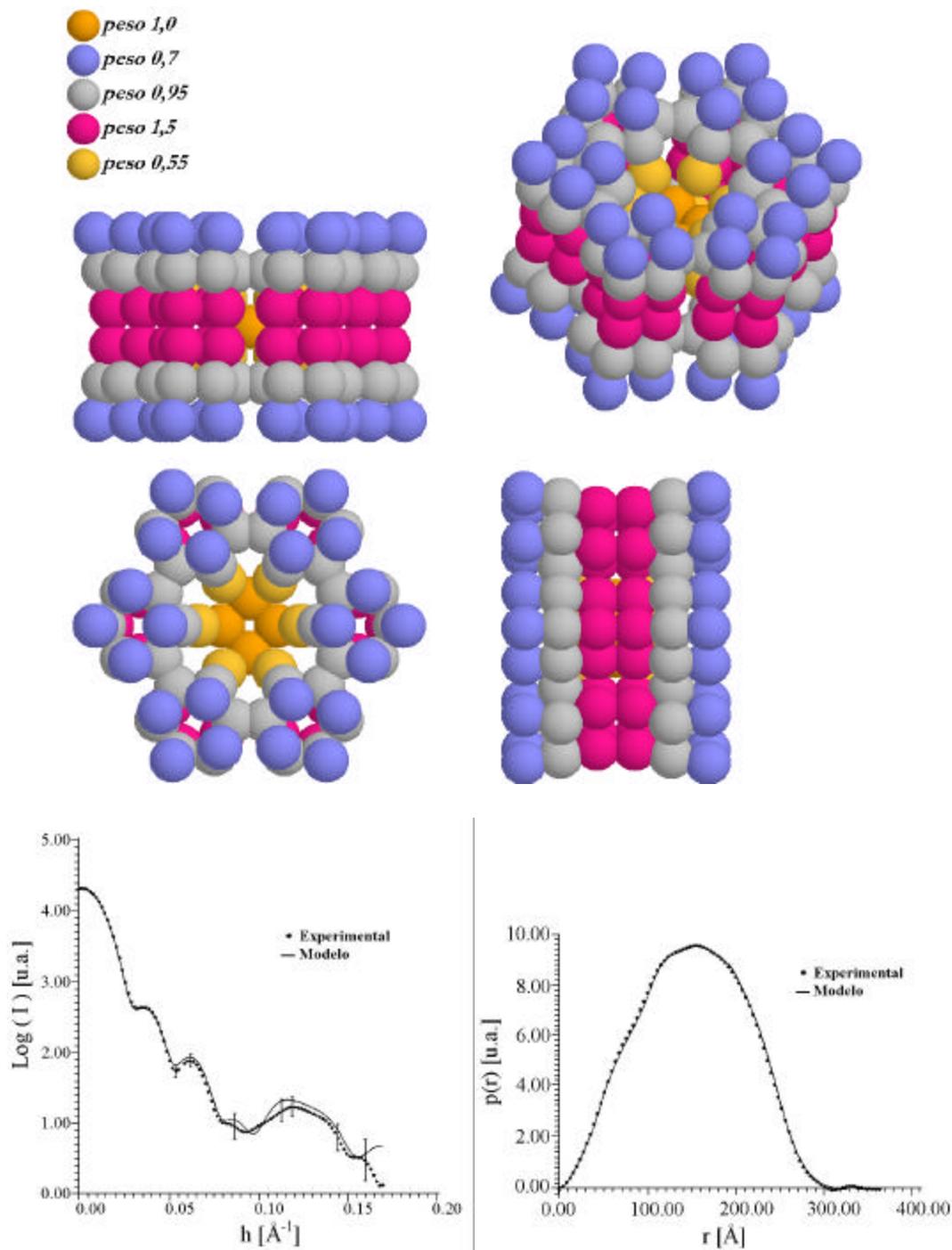
Devido às dimensões destas proteínas, sua cristalização é praticamente impossível de ser efetuada e assim, sua estrutura não pôde ser resolvida a nível atômico utilizando cristalografia de proteínas. Neste contexto, a técnica de espalhamento de raios X a baixo ângulo pode fornecer informações sobre a forma destas proteínas. Na realidade, isto foi objeto de estudo em muitos trabalhos (Stöckel et al, 1973; Pilz et al, 1979, etc) onde modelos tentativos foram propostos para estas proteínas extracelulares.

Usaremos dados obtidos com desta classe de proteínas provenientes da espécie *Glossoscolex paulistus* previamente estudadas por Cristina Franco de Souza na sua tese de doutoramento (TD So89e – IFGW) defendida em maio de 1990. A massa molecular desta eritrocruorina é de  $3.23 \times 10^6$  e um peso molecular mínimo de 25250 Da, o que constitui então, algo em torno de 128 grupos heme.

O trabalho de tese constituiu-se em um estudo sistemático da forma destas proteínas extracelulares baseando-se na curva de espalhamento a baixo ângulo, culminando em um modelo construído pela técnica de elementos finitos, apresentada anteriormente, sendo o ajuste do modelo efetuado manualmente.

Neste trabalho foram utilizadas informações de microscopia eletrônica e dados da literatura para construir e refinar um modelo para esta proteína. As medidas foram realizadas utilizando radiação síncrotron, sendo os dados coletados na estação experimental do anel DCI no LURE (Laboratoire pour l'Utilisation du Rayonnement Electromagnétique), Orsay, França. Detalhes sobre o aparato experimental e preparação das amostras podem ser encontrados em seu trabalho de tese. No entanto, consistiu-se de medidas em alta resolução em com baixo nível de ruído.

A análise dos dados foi realizada utilizando-se o formalismo dos programas ITP e MULTIBODY (Glatter, 1979, 1980). O resultado da construção de 570 modelos com alterações feitas manualmente são mostrados na Figura 5-3. O melhor ajuste é apresentado na figura citada, juntamente com o modelo tridimensional reconstruído a partir de dados da tese usando o programa BUILDER.



**Figura 5-3:** Modelo tridimensional, ajuste das curvas de espalhamento e distribuição de distâncias obtidos por Cristina F. Souza. [ 36 ]

Vemos um ótimo ajuste entre a curva experimental e o modelo proposto. Também a curva de distribuição de distâncias apresenta uma ótima concordância. Devemos compreender o significado de alterarmos os pesos das subunidades. Não significa diretamente que esta ou aquela parte da macromolécula possua um peso maior. Na realidade, se utilizássemos subunidades menores, aquelas regiões mais “pesadas” poderiam

representar locais onde temos uma população maior de subunidades que outra. Deste modo, a proteína poderia ser restaurada pelos métodos anteriormente apresentados, mesmo que aqueles métodos adotem o mesmo peso para todas as subunidades. Como veremos, os métodos estocásticos anteriores, usarão raio de 10Å para as subunidades componentes do modelo, conferindo mais detalhes ao modelo (como padrão, ambos os métodos usam uma resolução maior que a fornecida pela curva de espalhamento).

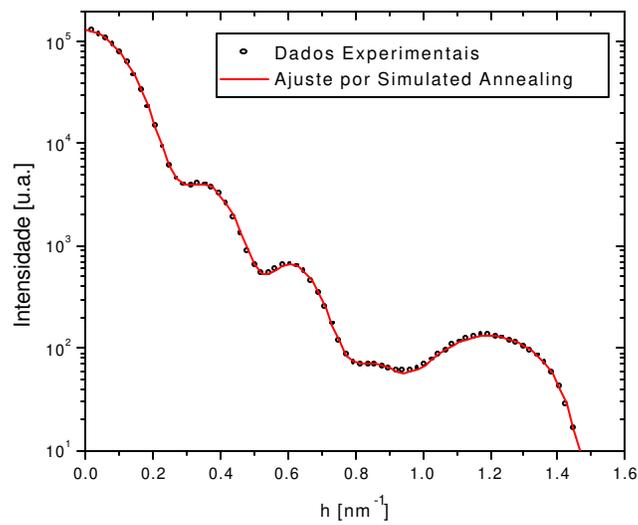
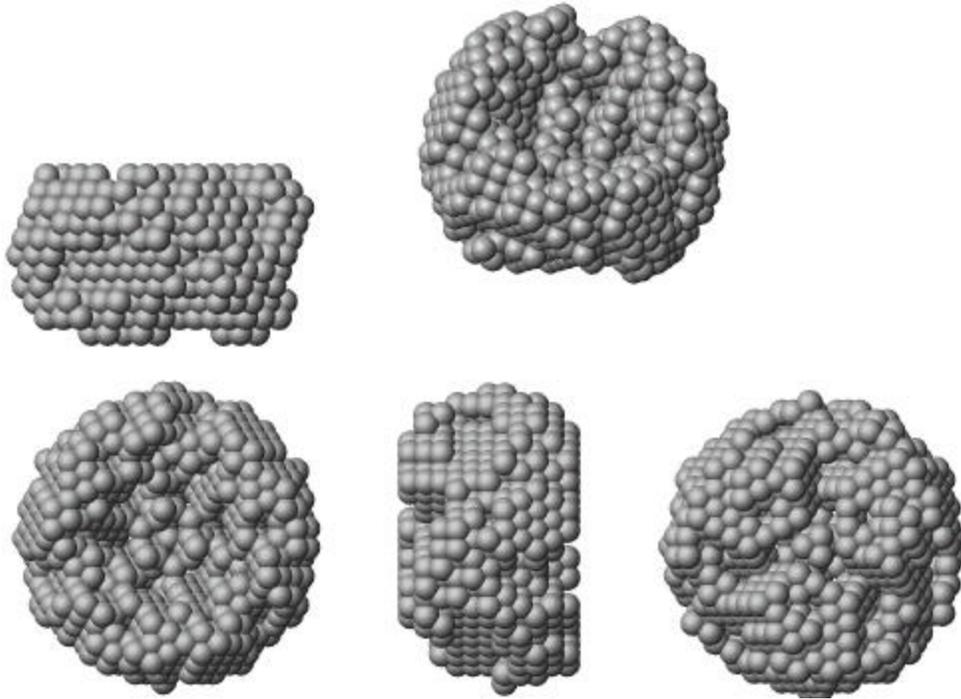
Não havíamos mencionado anteriormente, mas existem diferenças quanto a forma dos dados de entrada no método de *simulated annealing* e no algoritmo genético. No primeiro, o procedimento é exatamente o mesmo feito por Cristina Franco, os dados de entrada do programa não são exatamente os dados experimentais, mas sim o ajuste feito pelo programa GNOM (Svergun et al., 1992). Isto é necessário para que o programa tenha a informação da intensidade a ângulo zero e assim possa estimar o volume da partícula, bem como calcular com precisão o raio de giro além de a curva ser muito mais suave que a experimental. Outra peculiaridade do *simulated annealing* é tentar subtrair uma constante da curva de espalhamento, de modo que o comportamento a alto ângulo siga a lei de Porod (a intensidade diminui com  $h^{-4}$ ) mencionada no desenvolvimento da teoria. Quando apresentamos as simulações feitas com os modelos, este fato não era importante, pois a constante subtraída era nula. No algoritmo genético entramos com os dados experimentais, que devem ser normalizados de modo que a ângulo zero tenhamos intensidade unitária. Os autores deste programa sugerem que utilizemos programas como o ITP ou o GNOM para estimarmos o valor de  $I_0$ . Além disso, o algoritmo genético não tenta subtrair uma constante da curva de espalhamento, e caso queiramos corrigir este efeito, devemos subtraí-la manualmente.

Novamente o método de *simulated annealing* forneceu resultados mais compactos e simétricos, pelo fato de podermos inserir estes vínculos na rotina (Svergun,2000)] (utilizamos simetria 3, que pode ser evidenciada nas fotos de microscopia eletrônica). No entanto ambos os métodos demonstram uma região com pouca densidade eletrônica na parte central da proteína, o que está em acordo com o modelo proposto por Cristina Franco em sua tese, bem como diversos autores, que até mesmo propunham um modelo muito simplificado, composto de um cilindro oco (Stöckel,1973). Os resultados são apresentados a seguir.

## **5.2. RESTAURAÇÃO TRIDIMENSIONAL DE PROTEÍNAS EXTRACELULARES UTILIZANDO MÉTODOS ESTOCÁSTICOS**

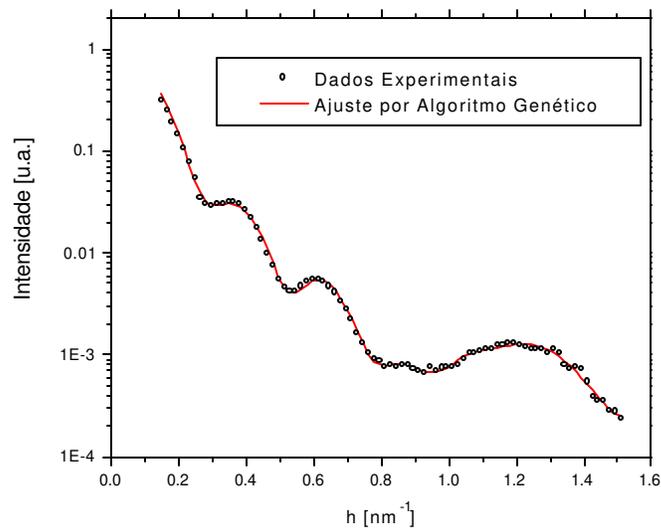
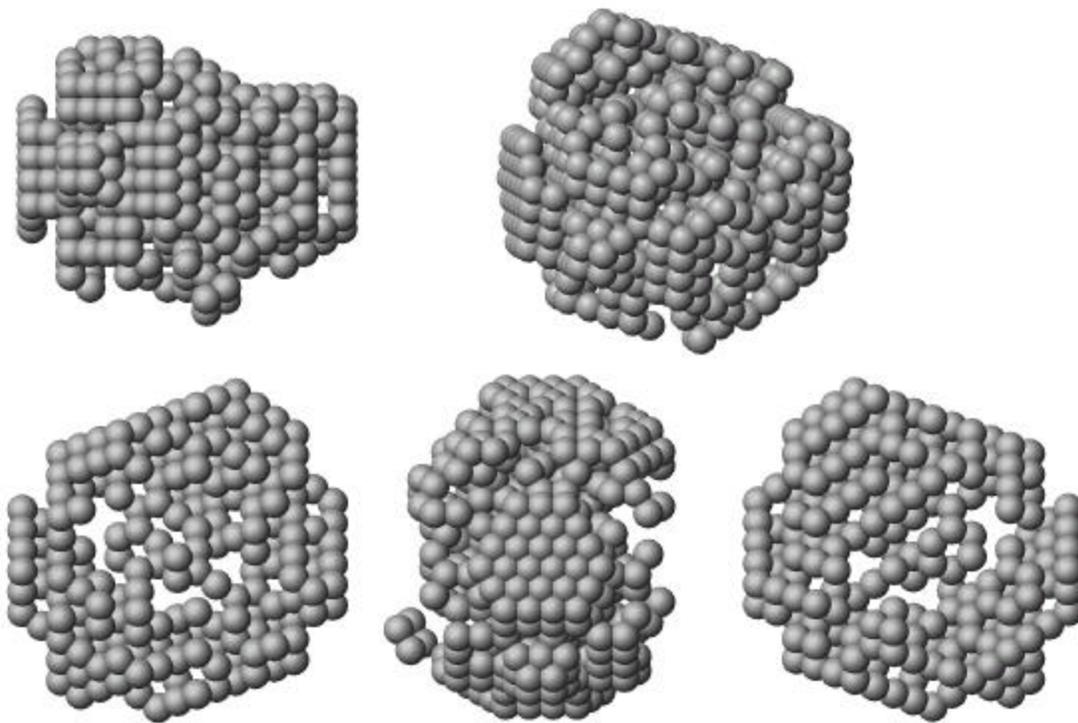
Utilizando os dados experimentais da tese de Cristina Franco, aplicamos os métodos estocásticos discutidos anteriormente.

- Simulated Annealing



**Figura 5-4:** Restauração *ab initio* e ajuste obtido para *glossoscolex paulistus*, utilizando *simulated annealing*. Na simulação utilizamos simetria P3

- Algoritmo Genético

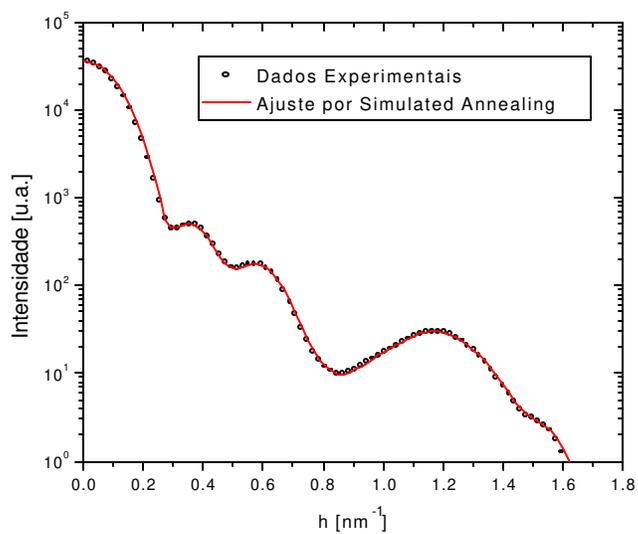
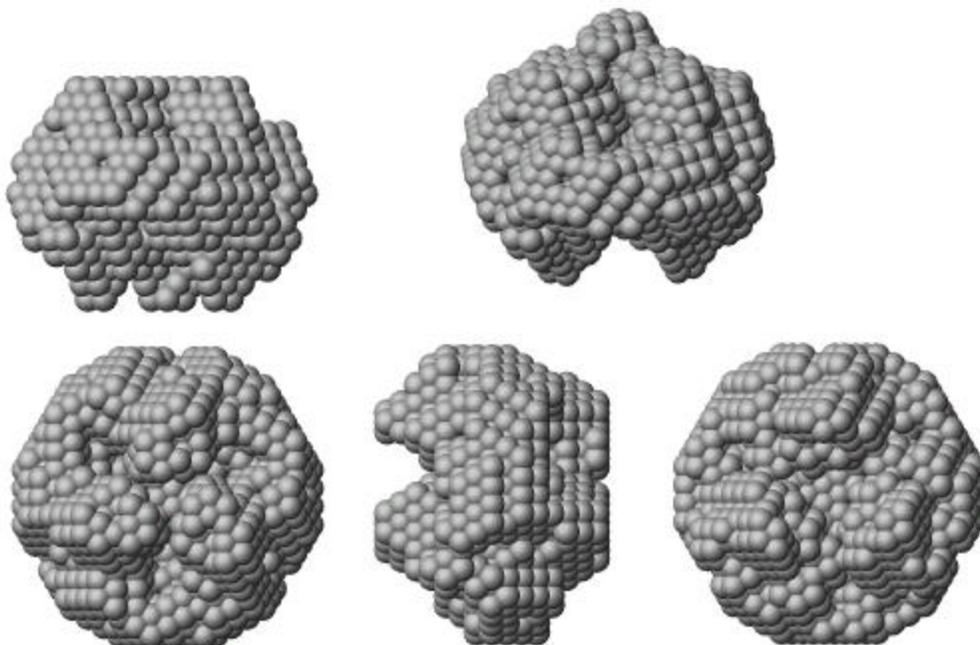


**Figura 5-5:** Restauração *ab initio* e ajuste obtido para *glossoscolex paulistus*, utilizando algoritmo genético.

Na mesma seqüência de medidas realizadas no LURE, foi medido a eritrocruorina de outra espécie de anelídeo (*Arenicola Marina*) que, de acordo com dados da literatura possui uma estrutura diferente. O estudo conformacional e comparativo entre eles podem

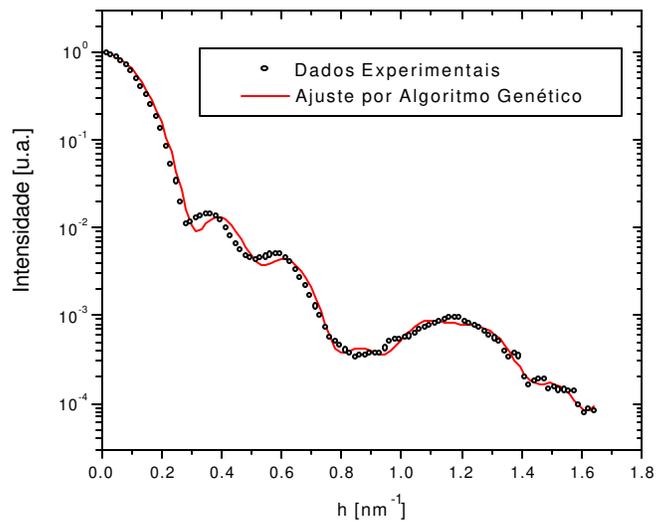
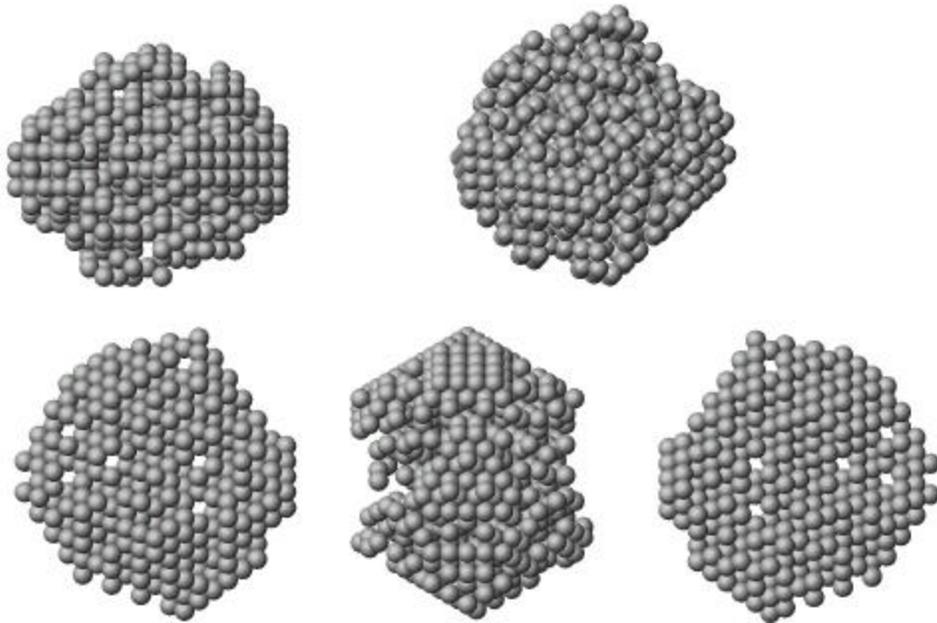
fornecer informações funcionais destas proteínas. Os resultados são apresentados a seguir para ambos os métodos.

- Simulated Annealing



**Figura 5-6:** Restauração *ab initio* e ajuste obtido para *arenicola marina*, utilizando simulated annealing. Na simulação utilizamos simetria P3

- Algoritmo Genético



**Figura 5-7:** Restauração *ab initio* e ajuste obtido para *arenicola marina*, utilizando algoritmo genético.

Os valores de raio de giro e máxima dimensão obtidos pelos dados anteriores condizem com os citados em muitos outros trabalhos, e são apresentados na tabela a seguir,

Anelídeo	Raio de Giro [Å]	Máxima Dimensão [Å]
<i>Glossoscolex paulistus</i>	$116 \pm 1$	$292 \pm 2$
<i>Arenicola Marina</i>	$119 \pm 1$	$310 \pm 2$

Obtemos estes valores utilizando o programa GNOM (Svergun,1990), e temos claramente a indicação de que a eritrocruorina da espécie marinha é maior que aquela da espécie terrestre tanto pelo valor do raio de giro, quanto pela máxima dimensão. Outro fator interessante é que a eritrocruorina da espécie *A. marina*. não apresenta a mesma característica da outra na região central. A *Glossoscolex* possui uma região com pouca densidade eletrônica na região central, enquanto que na *Arenicula* não observamos esta propriedade. Isto significa que a proteína marinha possui uma densidade eletrônica muito maior na região central do que a outra. Vale ressaltar que estas características que estamos discutindo são fornecidas pela restauração tridimensional de ambos os métodos e constituem características inerentes as proteínas as quais o SAXS e os métodos que estamos utilizando conseguiram evidenciar mesmo tendo em mente que a forma restaurada, está longe de ser a real (como vimos nos modelos anteriormente apresentados).

A capacidade do SAXS, juntamente com os métodos estocásticos que apresentamos, de restaurar a forma tridimensional e discernir partes com mais ou menos densidade eletrônica da proteína, bem como a possibilidade de evidenciar mudanças conformacionais (como vimos nos exemplos dos dímeros), fica evidenciado nos exemplos do trabalho aqui apresentado.

Comparando-se especificamente os resultados do *simulated annealing* e do algoritmo genético, vemos que este último não conseguiu ajustar precisamente a curva de espalhamento da eritrocruorina de *A. marina*. Isto pode ter várias causas, mas o fato de o método do *simulated annealing* utilizar não a curva experimental, mas sim um ajuste prévio que permite a eliminação de efeitos experimentais (concentração, ruído etc) pode ter sido decisivo no ajuste dos dados desta proteína.

Com isso, após a apresentação de todos os resultados, podemos avaliar que apesar de ambos os métodos consistirem de algoritmos estocásticos, o método baseado na otimização por *simulated annealing* apresenta ingredientes que possibilitam a obtenção de formas e ajustes muito melhores que o método de otimização por algoritmo genético.



## 6. CONSIDERAÇÕES FINAIS

Ao longo de todo o texto, os resultados apresentados eram comentados e analisados, cabendo agora uma análise geral de todos os tópicos discutidos.

Podemos encarar a teoria de espalhamento de várias maneiras, mas tudo recairá em quanta *informação* temos na curva de espalhamento. Quanto mais informação, mais conclusões podemos tirar de nossos dados. Como vimos, para a curva de espalhamento em si, a quantidade de informação pode ser dada pelo número de *Shannon channels* que temos na mesma, mas a análise completa de curvas de espalhamento vai muito além disso. Informações complementares, por outras técnicas ou mesmo a experiência do experimentador são cruciais na interpretação dos resultados. Essa análise pode ser feita em vários níveis, desde uma simples medida do raio de Guinier, a uma tentativa de restauração tridimensional, mas tudo deve ser alicerçado pelo pleno conhecimento do sistema em estudo.

Enfatizo isso pelo simples fato de que, uma suposição errônea, leva a resultados catastróficos. Um dos exemplos mais simples é a existência de polidispersidade no sistema investigado. Nosso estudo foi centrado em sistemas *monodispersos*, mas em experimentos reais de proteínas em solução, por exemplo, sempre haverá algum tipo de agregação ou coisa do tipo que acabem gerando a presença de polidispersidade no sistema. No entanto, existem maneiras de se minimizar estes efeitos procurando deixá-los ao menos ao nível do erro da medida. Outro efeito que é impossível de ser completamente eliminado é o de concentração. Como mostramos, esse efeito mascara o raio de Guinier, provocando uma diminuição da intensidade próxima a origem do espaço recíproco. A eliminação completa deste efeito requer a medida de uma série de concentrações e a posterior extrapolação para concentração zero.

Apresentadas estas ressalvas, podemos voltar à linha inicialmente tomada. A técnica de espalhamento a baixo ângulo mostra-se uma técnica impar, no sentido de que em muitos casos, nenhuma outra técnica poderia fornecer os resultados que ela fornece, ao menos com a mesma facilidade. Como vimos, a restauração tridimensional da forma é possível, logicamente com todos os cuidados interpretativos exaustivamente levantados em toda a discussão, e a forma obtida mesmo não sendo a real, fornece informações muito importantes como aspectos gerais da anisotropia (se a partícula é esférica, prolata, oblata), a existência ou não de regiões com menos densidade eletrônica (como evidenciamos no caso das eritrocruorinas). Além disso, os resultados das simulações com os dímeros, propiciaram a certeza de que informações como variação da angularidade entre partes de uma proteína, ou mesmo a variação conformacional da mesma pode ser acompanhada de modo direto pelo espalhamento de raios X a baixo ângulo.

Somente foi possível obter-se a informação sobre as possibilidades e limitações dos métodos estocásticos através de sua aplicação em modelos construídos pelo programa BUILDER, desenvolvido nesta tese. Sendo assim a utilidade da técnica de elementos finitos guiada é claro pelos programas BUILDER e MULTIBODY, mostrou-se crucial para a realização de testes e também para a compreensão da relação entre a forma da partícula e seu espalhamento.

A utilidade da técnica de espalhamento a baixo ângulo é indiscutível. No entanto os avanços computacionais na interpretação e apresentação dos resultados são recentes e

necessitam ser aplicados amplamente no estudo dos mais variados tipos de sistemas, pois representam novos caminhos para a restauração de estruturas moleculares a partir de dados de espalhamento.

## 7. REFERÊNCIAS E BIBLIOGRAFIA

- [ 1 ] C. SHANNON, W. WEAVER, (1949), *The Mathematical Theory of Communication*, The University of Illinois Press: Urbana;
- [ 2 ] P. DEBYE, A. M. BUECHE, (1949). *Scattering by an Inhomogeneous Solid*, *J. App. Phys.*, **20** , 518-525;
- [ 3 ] A. GUINIER, AND G. FOURNET,(1955). In “*Small-angle Scattering of X-rays*”, Wiley, New York;
- [ 4 ] R. BRACEWELL ,(1965). “*Fourier Transform and Its Applications*”, McGraw-Hill New York;
- [ 5 ] H. B. STURHMANN, (1970). *Interpretation of Small-Angle Scattering Functions of Dilute Solutions and Gases. A Representation of the Structures Related to a One-Particle Function* , *Acta Cryst.*, **A26** , 297;
- [ 6 ] S. PATEL, P. W. SCHIMIDT, (1971). *Small-Angle X-ray Scattering Determination of the Electron Density of the Particles in a Colloidal Silica Suspension*, *J. App. Cryst.*, **4** , 50;
- [ 7 ] O. GLATTER, (1972). *X-Ray Small Angle Scattering of Molecules Composed of Subunits* ,*Acta Physica Austriaca*, **36**, 307-315;
- [ 8 ] H. B. STURHMANN, (1973). *Comparison of the Three Basic Scattering Functions of Myoglobin in Solution with those from the Known Structure in Crystalline State*, *J. Mol. Biol.*, **77** , 363-369;
- [ 9 ] P. STÖCKEL, A. MAYER, R. KELLER, (1973), *X-Ray Small-Angle Scattering Investigation of a Giant Respiratory Protein: Hemoglobin Tubifex tubifex*, *Eur. J. Biochem.*, **37**, 193-200;
- [ 10 ] H. WU, P. W. SCHIMIDT, (1974). *The Relation Between the article and the Outer Part of the Small-Angle X-Ray Scattering Curve* , *J. App. Cryst.*, **7** , 131;
- [ 11 ] O. GLATTER, (1977). *A New Method for Evaluation of Small-Angle Scattering Data*, *J. Appl. Cryst.* **10**, 415-421;
- [ 12 ] H. B. STURHMANN, A. MILLER. (1978). *Small-Angle Scattering of Biological Structures*, *J. Appl. Cryst.*, **11** , 325-345;
- [ 13 ] O. GLATTER, (1979), *The Interpretation of Real-Space Information from Small-Angle Scattering Experiments*. *J. Appl. Cryst.* , **12**, 166-175;

- [ 14 ] *J. MÜLLER, G. DAMASHUN, (1979). Comparison and Optimization of Smoothing Procedures for Small-Angle X-Ray Scattering Curves. Polynomial Fitting and Modified Frequency Filtering, J. Appl. Cryst,12, 267-274;*
- [ 15 ] *O. GLATTER, (1980), Determination of Particle-Size Distribution Functions from Small-Angle Scattering Data by Means of the Indirect Transformation Method, J. Appl. Cryst. 13, 7-11;*
- [ 16 ] *O. GLATTER, (1980), Evaluation of Small-Angle Scattering Data from Lamellar and Cylindrical Particles by Indirect Transformation Method. J. Appl. Cryst. (1980), 10, 7-11;*
- [ 17 ] *O. GLATTER, (1980), Computation of Distance Distribution Functions and Scattering functions of Models for Small Angle Scattering Experiments Acta Physica Austriaca , 52, 243-256;*
- [ 18 ] *P. B. MOORE, (1980). Small-Angle Scattering. Information Content and Error Analysis, J. Appl. Cryst, 13, 168-175;*
- [ 19 ] *I. PILZ, E. SCHWARZ, S. N. VINOGRADOV, (1980). Small-Angle X-Ray Studies of Lumbricus terrestris haemoglobin, Int. J. Biol. Macromol., 2, 279-283;*
- [ 20 ] *P. WILHELM, I. PILZ, S. N. VINOGRADOV, (1980). Comparison of Arenicola marina and Lumbricus terrestris haemoglobins by Small-Angle X-Ray Scattering, Int. J. Biol. Macromol., 2, 383-384;*
- [ 21 ] *O. GLATTER, (1981), Convolution Square Root of Band-Limited Symmetrical Functions and its Application to Small Angle Scattering Data, J. Appl. Cryst. 14, 101-108;*
- [ 22 ] *O. GLATTER, O. KRATKY, (1982) "Small Angle X-ray Scattering", Academic Press, New York;*
- [ 23 ] *D. I. SVERGUN, L. A. FEIGIN, B. M. SCHEDRIN (1982), Small-Angle Scattering: Direct Structure Analysis, Acta Cryst., A38, 827-835;*
- [ 24 ] *D. TAUPIN, V. LUZZATI, (1982). Information Content and Retrieval in Solution Scattering Studies. 1. Degrees of Freedom and Data Reduction,, J. Appl. Cryst ,15, 289-300;*
- [ 25 ] *T. GERBER, P. W. SCHIMIDT, (1983). The Sampling Theorem and Small-Angle Scattering, J. Appl. Cryst, 16, 581-589;*
- [ 26 ] *S.KIRKPATRICK, C. D. GELATT JR. M. P. VECCHI,, (1983). Optimization by Simulated Annealing, Science, 220, 671-680;*

- [ 27 ] O. GLATTER, (1984), *Improvements in Real-Space Deconvolution of Small-Angle Scattering Data*, *J. Appl. Cryst.* **17**,435-441;
- [ 28 ] S. CICCARIELLO, (1985), *Deviations from the Porod Law due to Parallel Equidistant Interfaces*, *Acta Cryst.*, **A41** , 560-568;
- [ 29 ] S. N. VINOGRADOV, S. D. LUGO, M. G. MAINWARING, O. H. KAPP (1986). *Bracelet protein: A quaternary structure proposed for a giant extracellular haemoglobin of Lumbricus terrestris*, *Proc. Natl. Acad. Sci. USA.*, **83**, 8034-8038;
- [ 30 ] LEHNINGER, A. L., "Princípios de Bioquímica", (1986) Sarvier, São Paulo;
- [ 31 ] A. FEIGIN, D. I. SVERGUN, (1987), *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum Publishing Corporation – Plenum Press, New York;
- [ 32 ] S. CICCARIELLO, (1988), *On the Porod Law*, *J. App. Cryst.*, **21** , 117-128;
- [ 33 ] O. GLATTER, (1988), *Comparison of Two Different Methods for Direct Structure Analysis from Small-Angle Scattering Data*, *J. Appl. Cryst.* **21**,886-890;
- [ 34 ] D. I. SVERGUN, A. V. SEMENYUK L. A. FEIGIN,(1988), *Small-Angle Scattering-Data Treatment by the Regularization Method*, *Acta Cryst.*, **A44**, 244-250;
- [ 35 ] S. CICCARIELLO, (1989), *Particle Shapes and Oscillatory Deviations from the Porod Law: the Hyperbolic Contact Point Case*, *Acta Cryst.*, **A45** , 86-99;
- [ 36 ] C. F. SOUZA, (1990), *Espalhamento de Raios X a Baixo Ângulo: Estudo de Hemoglobinas Extracelulares em Solução*, Tese de Doutorado – IFGW, TDS089e;
- [ 37 ] O. GLATTER, (1991) *Scattering Studies on colloids of biological interest (Amphiphilic Systems)* *Prog. Colloid. Polym Sci.*, **84**, 46-54;
- [ 38 ] D. I. SVERGUN, (1991) *Mathematical-Methods in Small-Angle Scattering Data-Analysis*. *J. Appl. Cryst.* **24**, 485-492;
- [ 39 ] V. SEMENYUK AND D. I. SVERGUN,(1991) *Gnom - A Program Package for Small-Angle Scattering Data-Processing*. *J. Appl. Cryst.*, **24**, 537-540;
- [ 40 ] D. I. SVERGUN, H. B. STUHRMANN,(1991) *New Developments in Direct Shape Determination from Small-angle Scattering. 1. Theory and Model Calculations*, *Acta Cryst.*, **A47**, 736-744;
- [ 41 ] P. LINDNER, TH. ZEMB (Editors), (1991) *Newtron, X-Ray and Light Scattering*, Elsevier Schience Publishers B.V., 33-78;

- [ 42 ] W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, W. T. VETTERLING, (1992), *Numerical Recipes in Pascal – The Art of Scientific Computing* , Cambridge University Press;
- [ 43 ] D. I. SVERGUN, (1992), *Determination of the Regularization Parameter In Indirect-Transform Methods Using Perceptual Criteria*, *J. Appl. Cryst.*, **25**, 495-503;
- [ 44 ] D. I. SVERGUN, (1993), *A Direct Indirect Method of Small-Angle Scattering Data Treatment*, *J. Appl. Cryst.*, **26**, 258-267;
- [ 45 ] O. GLATTER, K. GRUBER (1993), *Indirect Transformation in Reciprocal Space: Desmearing of Small-Angle Scattering Data from Partially Ordered Systems*, *J. Appl. Cryst.* **26**,512-518;
- [ 46 ] L. INGBER,(1993). *Simulated Annealing: Practice versus Theory*, *Mathl. Comput. Modelling*, vol **18**, n° 11, 29-57;
- [ 47 ] D. I. SVERGUN, J. SKOV PEDERSEN (1994), *Propagating Errors in Small-Angle Scattering Data Treatment*, *J. Appl. Cryst.*, **27**, 241-248;
- [ 48 ] E. PANTOS, J. BORDAS, (1994), *Supercumputer Simulation of Small Angle X-Ray Scattering, Electron Micrographs and X-Ray Diffraction Patterns of Macromolecular Structures*, *Pure & Appl. Chem.*,vol **66**,n° 1, 77-82;
- [ 49 ] R. ZHANG, R. M. SUTER, AND J. F. NAGLE (1994), *Theory of the structure factor of lipid bylayers*, *Phys. Rev. E* **50**, 5047-5060.
- [ 50 ] D. VOET, J. G. VOET(1995), *Biochemistry*, New York , J. Wiley;
- [ 51 ] D. SVERGUN,C. BARBERATO, M. H. J. KOCH, (1995), *CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates*, *J. Appl. Cryst.*, **28**, 768-773;
- [ 52 ] Y. ZENG, P. C. DOERSCHUK, J. E. JOHNSON, (1995). *Determination of Three-Dimensional Low-Resolution Viral Structure from Solution X-Ray Scattering Data*, *Biophysical Journal.*, **69** , 619-639;
- [ 53 ] R.. A. SAYLE AND E. J. MILNER-WHITE, *RasMol: Biomolecular graphics for all Trends in Biochemical Sciences*, **20**:374-376, 1995.
- [ 54 ] D. I. SVERGUN, V. V. VOLKOV, M. B. KOZIN, H. B. STUHRMANN, (1996). *New Developments in Direct Shape Determination from Small-Angle Scattering. 2. Uniqueness*, *Acta Cryst.*, **A52** , 419-426;
- [ 55 ] KORADI, R., BILLETER, M., AND WÜTHRICH, K. (1996), *MOLMOL: a program for display and analysis of macromolecular structures*.*J Mol Graphics* **14**, 51-55.

- [ 56 ] S. CICCARIELLO, (1997), *Behavior of Small-Angle Scattering Intensity Tails*, *J. Appl. Cryst.*, **21** , 50;
- [ 57 ] D. I. SVERGUN, V.V.VOLKOV, M.B. KOZIN, H. B. STUHRMANN, C. BARBERATO, M. H. J. KOCH (1997), *Shape Determination from Solutions Scattering of Biopolymers*, *J Appl. Cryst.*, **30**, 798-802;
- [ 58 ] D. I. SVERGUN, (1997) *Restoring Three-Dimensional Structure of Biopolymers from Solution Scattering*, *J. Appl. Cryst.*, **30**, 792-797;
- [ 59 ] B. KOZIN, V. V. VOLKOV, D. I. SVERGUN (1997), *ASSA, a Program for Three-Dimensional Rendering in Solution Scattering from Biopolymers*, *J. Appl. Cryst.*, **30**, 811-815;
- [ 60 ] A. G. RUGGIERO, V. L. R. LOPES, (1997), *Cálculo Numérico – Aspectos Teóricos e Computacionais*, 2ª Edição, Makron Books;
- [ 61 ] P. VACHETTE AND D. SVERGUN,(1998) “*Small-Angle X-ray Scattering by solutions of Biological Macromolecules*”, *minicurso dado no HERCULES 98 – Glenoble* ;
- [ 62 ] P. CHACÓN,, F. MORÁN, J. F. DÍAS, E. PANTOS, AND J. M ANDREU, (1998). *Low-resolution structures of proteins in solution retrieved from X-ray scattering whit a genetic algorithm*. *Biophysical Journal* **74**,2760-2775;
- [ 63 ] R. MITTELBACH, O. GLATTER, (1998) *Direct Structure Analysis of Small-Angle Scattering Data from Polydisperse Colloidal Particles*, *J. App. Cryst.*, **31**, 600-608;
- [ 64 ] S. CICCARIELLO, R. SOBRY (1999), *Small-Angle Scattering Intensity Behaviors of Cylindrical, Spherical and Planar lamellae*, *J. Appl. Cryst.*, **32** , 892-901;
- [ 65 ] D. I. SVERGUN, (1999), *Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing*, *Biophysical Journal*, **76**, 2879-2886
- [ 66 ] D. I. SVERGUN, (2000), *Advanced solution scattering data analysis methods and their applications*, *J. Appl. Cryst.*, **33**, 530-534.
- [ 67 ] P. CHACÓN,, F. MORÁN, J. F. DÍAS, E. PANTOS, AND J. M ANDREU, (2000), *Reconstruction of Protein Form with X-ray Solution Scattering and a Genetic Alrorthm*, *J. Mol. Biol.*, **299**, 1289-1302.
- [ 68 ] C. OLIVEIRA, I. TORRIANI, *in preparation*