Universidade de São Paulo
Instituto de Física

SIMULAÇÃO E MODELAGEM COMPUTACIONAL DE
DADOS DE ESPALHAMENTO À BAIXOS ÂNGULOS -
ENFOQUE EM ESTRUTURAS DE ALTA SIMETRIA

Cássio Alves

Orientador: Prof. Dr. Cristiano Luis Pinto de Oliveira

Tese de doutorado apresentada ao
Instituto de Física para a obtenção do
título de Doutor em Ciências

Banca Examinadora:

Prof. Dr. Cristiano Luis Pinto de Oliveira (IF-USP)

Profa. Dra. Kaline Rabelo Coutinho (IF-USP)

Prof. Dr. Luis Maurício Trambaioli da Rocha e Lima (UFRJ)

Prof. Dr. Roberto Kopke Salinas (IQ-USP)

Prof. Dr. Sérgio Luis Morelhão (IF-USP)

São Paulo

2015

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Alves, Cássio

Simulação e modelagem computacional de dados de
espalhamento à baixos ângulos - enfoque em estruturas de
alta simetria. São Paulo, 2015.

Tese (Doutorado) - Universidade de São Paulo.
Instituto de Física, Departamento de Física Experimental.

Orientador: Prof. Dr. Cristiano Luis Pinto de Oliveira
Área de Concentração: Física da Matéria Condensada

Unitermos: 1. Física da matéria condensada; 2. Biofísica molecular;
3. Física molecular; 4. Física experimental; 5. Física computacional.

USP/IF/SBI-044/2015

AGRADECIMENTOS

Gostaria de deixar registrado minha imensa gratidão e meu reconhecimento a todos que contribuíram, direta ou indiretamente, para o êxito deste trabalho.

Primeiramente ao orientador deste trabalho, o professor Doutor Cristiano L P Oliveira, que com muita competência mostrou o caminho a ser traçado, sempre com muitas idéias, atenciosa e sólida orientação, apontando o possível no meio de impossíveis.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela bolsa de Doutorado, ao INCT de Fluidos Complexos, pelos demais financiamentos e apoio e a comissão de pós graduação do Instituto de Física pelo apoio.

Ao Programa Ciências Sem Fronteiras (processo: 237230/2012-7/SWE) pelo financiamento completo do estágio realizado no grupo de Biologia Estrutural, chefiado pelo professor Alessandro Desideri na Universidade de Roma Tor Vergata.

Aos professores Alessandro Desideri, Mattia Falaconi e Blasco, aos estudantes Iacovelli, Oteri e Andrea e a todos do grupo de Biologia Estrutural por todo apoio, orientação e a calorosa recepção em Roma.

A todos os professores, técnicos e estudantes do grupo de Fluídos Complexos pelo apoio.

Ao Custódio, Cecília, Antônio, Terezinha, Lorena, Cassiano, Claudio pela agradável companhia, carinho, imenso e inumerável apoio recebido, além da paciência nos altos e baixos momentos nos quais sopram as nossas vidas.

Ao Amarildo, Jussara e Norberto pelas inúmeras e agradáveis discussões sobre quase tudo.

RESUMO

Esta tese apresenta uma abordagem para a criação rápida de modelos em diferentes geometrias (complexas ou de alta simetria) com objetivo de calcular a correspondente intensidade espalhada, podendo esta ser utilizada na descrição de experimentos de espalhamento à baixos ângulos. A modelagem pode ser realizada com mais de 100 geometrias catalogadas em um Banco de Dados, além da possibilidade de construir estruturas a partir de posições aleatórias distribuídas na superfície de uma esfera. Em todos os casos os modelos são gerados por meio do método de elementos finitos compondo uma única geometria, ou ainda, compondo diferentes geometrias, combinadas entre si a partir de um número baixo de parâmetros.

Para realizar essa tarefa foi desenvolvido um programa em FortranTM, chamado de *Polygen*, que permite modelar geometrias convexas em diferentes formas, como sólidos, cascas, ou ainda com esferas ou estruturas do tipo DNA nas arestas, além de usar esses modelos para simular a curva de intensidade espalhada para sistemas orientados e aleatoriamente orientados. A curva de intensidade de espalhamento é calculada por meio da equação de Debye e os parâmetros que compõe cada um dos modelos, podem ser otimizados pelo ajuste contra dados experimentais, por meio de métodos de minimização baseados em *simulated annealing*, Levenberg-Marquardt e algorítmicos genéticos. A minimização permite ajustar os parâmetros do modelo (ou composição de modelos) como tamanho, densidade eletrônica, raio das subunidades, entre outros, contribuindo para fornecer uma nova ferramenta para modelagem e análise de dados de espalhamento.

Em outra etapa desta tese, é apresentado o *design* de modelos atomísticos e a sua respectiva simulação por Dinâmica Molecular. A geometria de dois sistemas auto-organizados

de DNA na forma de octaedro truncado, um com *linkers* de 7 Adeninas e outro com *linkers* de ATATATA, foram escolhidas para realizar a modelagem atomística e a simulação por Dinâmica Molecular. Para este sistema são apresentados os resultados de *Root Mean Square Deviations* (RMSD), *Root Mean Square Fluctuations* (RMSF), raio de giro, torção das hélices duplas de DNA além da avaliação das ligações de Hidrogênio, todos obtidos por meio da análise de uma trajetória de 50 ns.

Palavras-chave: Modelagem, Simulação, Espalhamento a Baixos Ângulos, Dinâmica Molecular

ABSTRACT

This thesis presents an approach to the fast creation of models in different geometries (complex or high symmetry) in order to calculate the scattering intensity, which can be used for the description of small angles scattering experiments. The modeling can be performed using more than 100 geometries cataloged in a database, besides the possibility to build structures from random positions distributed on the surface of a sphere. In all cases the models are generated using the finite element method composing a single geometry, or composing different geometries combined with each other, using a small number of parameters.

To accomplish this task it was developed a program called Polygen, written in FortranTM language, which allows the modeling of convex polyhedrons in different geometries, as solids, shells, with aligned beads or DNA-like structures at the edges. To simulate the scattering intensity curve, these models are used and is possible simulate oriented and randomly oriented systems. The scattering intensity curve is calculated using the Debye equation and the main parameters describing the models, can be optimized by the fitting of the calculated curves against experimental data. The optimization is performed by the use minimization methods based on simulated annealing, Levenberg-Marquardt and genetic algorithmic. The minimization procedures allows the optimization of the model parameters (or models of composition) as size, electron density, gyration radius, among others, contributing to provide a new tool for modeling and scattering data analysis.

In a further step of this thesis, the design of atomistic models is presented and the respective simulation by Molecular Dynamics. Two geometries for DNA self-assembly

systems, where chosen: on octahedron truncated, using linkers with 7 adenine and another using linkers of ATATATA. For these systems it is presented the results of Root Mean Square Deviations (RMSD), Root Mean Square Fluctuations (RMSF), gyration radius, bend of double helix of DNA besides the evaluation of Hydrogen bonds, all obtained through the analysis of a trajectory of 50 ns.

Keywords: Modeling, Simulation, Small Angle Scattering, Molecular Dynamics

SUMÁRIO

Lista de Abreviaturas	x
Lista de Figuras	xi
Lista de Tabelas	xx
1 Introdução	1
2 Fundamentos Teóricos	9
2.1 Descrição do Fenômeno de Espalhamento de Radiação	10
2.1.1 Espalhamento Rayleigh ou o Espalhamento por um Potencial . .	10
2.1.2 Espalhamento Discreto por Potenciais em um Meio	24
2.2 Espalhamento à Baixos Ângulos Aplicado a Diferentes Sistemas	27
2.3 Sistemas Auto-Organizados	40
2.3.1 Sistemas Auto-Organizados de DNA	43
2.4 Dinâmica Molecular	49
3 SAS, MODELAGEM E SIMULAÇÃO	53
3.1 Modelagem	54
3.2 Simulação	61
3.2.1 Intensidade Média - Sistemas Aleatoriamente Orientados	62
3.2.2 Intensidade Em Sistemas Orientados	80
3.2.3 Verificação dos Graus de Precisão e Custos Computacionais . .	102
3.3 SAXS: Conclusões Parciais e Perspectivas	109

4	Dinâmica Molecular (DM)	112
4.1	<i>Design</i> e Modelagem	114
4.2	Protocolos de Equilíbrio, Termalização e DM	115
4.3	Análise da DM	116
4.3.1	Análise da Estabilidade e de parâmetros geométricos das <i>Nanocages</i>	116
4.3.2	<i>Linkers</i>	121
4.4	DM: Conclusões Parciais e Perspectivas	125
5	Conclusões Gerais e Perspectivas	129
A	Banco de Dados das estruturas de alta simetria	134
B	Síntese das Propriedades Estruturais do DNA	137
C	Sequências Nucleotídicas	139
	Referências Bibliográficas	140

LISTA DE ABREVIATURAS

AFM	<i>Atomic Force Microscopy</i> (Microscopia de Força Atômica)
bp	pares de bases
Cryo-EM	<i>Cryo-Electron Microscopy</i> (Crio Microscopia Eletrônica)
GFCx	Grupo de Fluídos Complexos (Grupo de Fluídos Complexos)
CSF	Ciência Sem Fronteiras
Da	dalton ou Unidade de Massa Atômica
DLS	<i>Dynamic Light Scattering</i> (Espalhamento de Luz Dinâmico)
DM	<i>Molecular Dynamics</i> (Dinâmica Molecular)
DNA	Ácido Desoxirribonucleico (ADN)
ds	<i>double strand</i> (dupla hélice)
EM	<i>Electron Microscopy</i> (Microscopia Eletrônica)
IFT	Transformada Indireta de Fourier
PAGE	Polyacrylamida Gel Electroforese
pdb	<i>protein data bank</i> (banco de dados de proteínas)
RMSD	<i>Root Mean Square Deviations</i>
RMSF	<i>Root Mean Square Fluctuations</i>
SAS	<i>Small Angle Scattering</i> (Espalhamento à Baixos Ângulos)
SAXS	<i>X-Ray Small Angle Scattering</i> (Espalhamento de Raios X à Baixos Ângulos)
ss	<i>single strand</i> (fita simples)
TIP3P	<i>Transferable Intermolecular Potential 3 Point-charge</i>
XFEL	<i>X-Ray free-electron lasers</i> (Laser de Raio-X de Elétrons-livres)
TF	Transformada de Fourier

LISTA DE FIGURAS

2.1	<i>Feixe incidente é paralelo ao eixo Oz. A zona de influência do potencial $V(\mathbf{r})$ está centrada em O. O detector mede o número de partículas dn espalhadas por unidade de ângulo sólido $d\Omega$. O coeficiente $d\sigma(\theta, \phi)$ é a seção de choque diferencial de espalhamento na direção (θ, ϕ) [1]. Figura adaptada de [1].</i>	12
2.2	<i>Representação das distâncias de influência do potencial, em que $\mathbf{r} - \mathbf{r}'$ é a distância entre o ponto O e M. O ponto P está situado dentro da zona de influência do Potencial que tem uma zona de influência da ordem de L. A distância entre O e P é muito pequena comparada com a distância entre O e M [1]. Figura adaptada de [1].</i>	14
2.3	<i>Diagrama da representação entre os vetores de onda incidente $\mathbf{k}_i$ e espalhado $\mathbf{k}_s$ e vetor de onda transferido $\mathbf{q}$ [1]. Figura adaptada de [1].</i>	18
2.4	<i>Representação esquemática: a) onda incidente e onda espalhada pela interação com o potencial, aproximação de Born; b) ondas espalhadas duas vezes pelo potencial, termo de segunda ordem em U na expansão de Born [1]. Figura adaptada de [1].</i>	23

2.5	<i>Desenho esquemático do processo de aplicação de polidispersidade diretamente no histograma de pares de distâncias. A) Função de polidispersão, do tipo gaussiana, com valor médio $\langle x \rangle$. B) Distribuição de pares de distâncias, de um cilindro do tipo bastão, em diferentes valores da curva de dispersão de tamanho. C) Distribuição de pares de distâncias resultante da dispersão de tamanho de um conjunto de espalhadores.</i>	41
3.1	<i>Desenho esquemático do processo de modelagem das geometrias. A) Posição dos vértices de um cubo. B) União dos vértices com segmentos de retas. C) Modelo do tipo sólido. D) União dos vértices com subunidades esféricas. E) União dos vértices com DNA. F) Modelo do tipo casca. Figura adaptada [2].</i>	58
3.2	<i>Desenho esquemático do processo de modelagem das geometrias com uso das coordenadas espaciais e das referências aos planos. O vetor $\vec{n}$ é normal ao plano definido pelos vetores v_1 e v_2. O vetor $\vec{v}_3$ conecta o ponto de referencia p_{ref} no plano ao ponto p_0. A) Se $\vec{v}_3 \times \vec{n} = 0$, p_0 está no plano. Se $\vec{v}_3 \times \vec{n} < 0$, p_0 está no semi-espaco negativo do plano. Se $\vec{v}_3 \times \vec{n} > 0$, p_0 está no semi-espaco positivo do plano. B) p_0 está no interior do tetraedro quando $\vec{v}_3 \times \vec{n} < 0$ para todos os vetores $\vec{n} = \vec{v}_1 \times \vec{v}_2$, desenhados com v_1 e v_2 orientados no sentido anti-horário das faces. Figura adaptada [2].</i>	60
3.3	<i>(a) Histograma de pares de distâncias entre as subunidades. b) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen da esfera e suas variantes. Figura adaptada [2].</i>	64
3.4	<i>(a) Histograma de pares de distâncias entre as subunidades. b) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen para o cilindro do tipo Prato (plate), bastão (rod) e cubo (cube). Figura adaptada [2].</i>	66
3.5	<i>(Esquerda) Histograma de pares de distâncias entre as subunidades, as geometrias apresentam polidispersão gaussiana em número com variâncias de 5%, 10% e 20% para: (A) esfera, (B) casca esférica, (C) cilindro do tipo bastão (rod) e (D) cilindro do tipo Prato (plate). (Direita) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen para: (A) esfera, (B) casca esférica, (C) cilindro do tipo bastão (rod) e (D) cilindro do tipo Prato (plate). Figura adaptada [2].</i>	68

3.6	<i>Resultados para modelos com arestas preenchidas com subunidades alinhadas. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].</i>	70
3.7	<i>Resultados para modelos de alta simetria em que as arestas são preenchidas com estruturas do tipo DNA. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].</i>	71
3.8	<i>Resultados para modelos do tipo cascas. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].</i>	72
3.9	<i>Resultados para modelos do tipo sólidos. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].</i>	73
3.10	<i>Efeito da polidispersidade nas nanocages de DNA. (Esquerda) Histograma dos pares de distâncias. (Direita) Intensidade de espalhamento simulada. Figura adaptada [2].</i>	74
3.11	<i>Composição de uma casca icosaédrica com uma nanocage de DNA em seu interior (A) Modelos. (B) histograma de pares de distâncias para cada modelo. (C) Curva de intensidade simulada com Polygen. O tamanho e as densidades eletrônicas para cada modelo estão detalhados na própria figura. Figura adaptada [2].</i>	77
3.12	<i>Exemplo do modelo de cages de DNA "aleatórias". Estes modelos foram gerados a partir de um conjunto aleatório de coordenadas espaciais, distribuídos em uma superfície esférica de raio igual a 100. (A) Histogramas de pares de distâncias para cada modelo. (B) Modelo obtido por 5 e 9 de coordenadas espaciais aleatórias. Nos modelos: (esquerda) as arestas são preenchidas com subunidades alinhadas (centro) as arestas são preenchidas com estruturas do tipo DNA e para visualizar a semelhança dos modelos à (direita) estão as composições dos modelos de subunidades alinhadas e do tipo DNA. (C) Curva de intensidade espalhada dos modelos com subunidades alinhadas e com DNA nas arestas. Figura adaptada [2].</i>	78

- 3.13 *Espalhamento de um feixe incidente k_i em uma determinada direção k_s , medido em um plano com raio q . O espalhamento é mensurado em um plano bidimensional. 82*
- 3.14 *(A) Imagem bidimensional da intensidade espalhada de um cilindro. Os cortes transversais são um guia para os olhos e mostram como uma imagem bidimensional pode ser composta com curvas unidimensionais. (B) Intensidade espalhada. (abaixo) o plano bidimensional. (acima) o gráfico tridimensional da imagem de espalhamento. 82*
- 3.15 *Histograma $h(\mathbf{qr})$ das projeções dos pares de distâncias $\mathbf{qr}$, em função do raio $\mathbf{q}$ e do ângulo do plano da imagem θ , variando de 0 a 2π 83*
- 3.16 *Simulação das imagens de espalhamento do cilindro do tipo bastão. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0 , 30 , 45 , 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0 , 30 , 45 , 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções. 88*
- 3.17 *Simulação das imagens de espalhamento do cilindro do tipo prato. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0 , 30 , 45 , 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0 , 30 , 45 , 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções. 89*

- 3.18 *Simulação das imagens de espalhamento do paralelepípedo. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.* 90
- 3.19 *Simulação das imagens de espalhamento do elipsoide de revolução. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.* 91
- 3.20 *A à C) imagens de espalhamento, do cilindro do tipo bastão em três diferentes orientações em relação ao feixe de incidência. D à F) Imagens de espalhamento, do cilindro do tipo prato em três diferentes orientações em relação ao feixe de incidência.* 92
- 3.21 *A à C) Imagens de espalhamento, do elipsoide de revolução em três diferentes orientações em relação ao feixe de incidência. D à F) imagens de espalhamento, do paralelepípedo em três diferentes orientações em relação ao feixe de incidência.* 93

3.22	<i>Simulação das imagens de espalhamento dos poliedros de Johnson: J14, J15 e J16. A) Histograma de pares de distâncias. B) (acima) Composição dos modelos dos sólidos de Johnson: J14, J15 e J16, na orientação de incidência do feixe. (abaixo) modelos rotacionados para visualização. C) Intensidade média simulada. D à F) Imagens de espalhamento, de acordo com os modelos (acima) na orientação de incidência do feixe; (abaixo) modelos rotacionados para visualização. Da esquerda para direita: J14, J15 e J16.</i>	95
3.23	<i>Simulação das imagens de espalhamento dos poliedros de Johnson: J53 e J54. A) (cinza escuro) Orientação do modelo do sólido de Johnson J53 na orientação de incidência do feixe e abaixo o mesmo modelo rotacionado. (cinza claro) Orientação do modelo do sólido de Johnson J54 na orientação de incidência do feixe e abaixo o mesmo modelo rotacionado. Os modelos são compostos em uma mesma figura mais abaixo. B) Histograma de pares de distâncias. C) Intensidade média simulada. D) Imagem da intensidade de espalhamento do sólido J53, de acordo com a orientação do modelo em A. E) Imagem da intensidade de espalhamento do sólido J54, de acordo com a orientação do modelo em A.</i>	96
3.24	<i>Simulação das imagens de espalhamento de diferentes poliedros com geometrias similares. A) Histograma de pares de distâncias. B) Da esquerda para direita: (acima) Orientação do modelo do tetraedro, octaedro, J12, J13, J49 e J50, na orientação de incidência do feixe. (abaixo) modelos rotacionados para visualização. O modelo mais a direita é uma composição dos modelos. C) Intensidade média simulada. Imagens de espalhamento, de acordo com os modelos em B: D) tetraedro, E) octaedro, F) J12 G) J13, H) J49 e I) J50.</i>	98
3.25	<i>Simulação da intensidade espalhada para pirâmides pentagonais. A) Histograma de pares de distâncias. B) Modelos da pirâmide pentagonal em três diferentes Perspectivas. C) Intensidade média simulada. D à F) Imagens da intensidade espalhada, de acordo com os modelos acima de cada plano. Da esquerda para direita três diferentes orientações da pirâmide pentagonal de diâmetro de 20 nm.</i>	99

3.26	<i>Simulação das imagens de espalhamento para modelos de proteínas obtidos em Protein Data Bank (RCSB PDB, www.rcsb.org) [3]. A) Histograma de pares de distâncias. B (acima) Orientação dos modelos de lysozyme pdb id: 4NY5 [4], colágeno de peptídeo: pdb id: 2CUO [5], e albumin (somente a cadeia A) pdb id: 3V03 [6], na orientação de incidência do feixe de incidência. (abaixo) modelos rotacionados para visualização. C) Intensidade média simulada. D à I) Modelos na orientação simulada e imagens de espalhamento, de acordo com os modelos em B: D) colágeno de peptídeo; E e F) duas orientações da albumina; G, H e I) três orientações da lysozyme.</i>	101
3.27	<i>Avaliação da influência de raio da subunidade, r, na precisão dos cálculos de intensidade, junto a densidade, d, de subunidades. A) Modelos obtidos com MEF, para cada um dos diferentes tamanhos de subunidades. B) Curva de intensidade. Dentro do painel B, um zoom realizado para altos valores de q. C) Soma acumulada da diferença normalizada, entre a curva de intensidade simulada e a calculada semi-analiticamente. . . .</i>	105
3.28	<i>Avaliação da influência do passo do histograma para a precisão dos cálculos de intensidade. A) Os vários histogramas normalizados e realizados com diferentes passos. B) Curva de intensidade. C) Soma acumulada da diferença normalizada, entre a curva de intensidade simulada e a calculada semi-analiticamente.</i>	106
4.1	<i>Estruturas de uma nanocage, auto-organizada de DNA, na geometria de um octaedro truncado. (a) Modelo atomístico na representação de átomos como esferas. (b) Modelo atomístico na representação da superfície acessível ao solvente.</i>	114
4.2	<i>RMSD das nanocages auto-organizadas de DNA com γA e γT e γAT nos linkers. Os dados são apresentados separados para comparar somente os linkers, somente as duplas hélices de cada nanocage e a composição de ambos (linkers + duplas hélices = nanocage). Dados da Nanocage com linkers de γ Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	117

4.3	<i>Raio de giro das nanocages de DNA com com 7A e 7T e 7AT nos linkers. Os dados são apresentados separados para comparar somente os linkers, somente as duplas hélices de cada nanocage e a composição de ambos (linkers + duplas hélices = nanocage). Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	118
4.4	<i>RMSF Local de cada nucleotídeo das duplas hélices de DNA presentes nas nanocages 7A (preta), 7T(vermelha) e 7AT (azul). Dados da nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	119
4.5	<i>RMSF Global de cada nucleotídeo das duplas hélices de DNA presentes nas nanocages 7A(preta), 7T(vermelha) e 7AT (azul). Dados da nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	120
4.6	<i>Torção (Bend) média sofrida em cada dupla hélice das nanocages de 7A, 7AT e 7T e as médias (Avg) para cada nanocage comparadas aos valores B-DNA. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	121
4.7	<i>Número médio de ligações de Hidrogênio entre: (A) as extremidades das duplas hélices, (B) os linkers ou hélices simples, (C) entre os pares de base das duplas hélices ou edges. (D) Síntese dos valores médios dos conjuntos das extremidades das duplas hélices e linkers. Dados da nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	123
4.8	<i>Número ou frequência de interações do tipo stacking entre os anéis aromáticos das bases de Adenina ou Timinas nos linkers para cada nanocage. (A) Número de interações do tipo stacking sandwich e T-shape somadas em cada passo da trajetória. (B) Número de interações do tipo stacking sandwich (C) Número de interações do tipo stacking T-shape. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].</i>	124
4.9	<i>Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com 7 Adeninas. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm.</i>	125

- 4.10 *Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com 7 Timinas. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].* 126
- 4.11 *Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com ATATATA. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm.* 127

LISTA DE TABELAS

C.1 Sequência de bases em cada par da dupla hélice de DNA	139
---	-----

CAPÍTULO 1

INTRODUÇÃO

5

Nas últimas décadas as áreas de atuação dos laureados com o Prêmio Nobel, têm refletido a fascinação pelas estruturas de biomoléculas, de 1956 a 2006 são pelo menos 12 prêmios nobéis na área de medicina, fisiologia ou química (1 a cada 4) [8]. Na década de 1996 a 2006 cerca de um terço dos trabalhos publicados e um em cada dois prêmios nobéis em química está relacionado a estruturas macromoleculares [8] e devido a essa intensa investigação, em estruturas biomoleculares, existe uma demandada crescente no desenvolvimento das técnicas e métodos de caracterização estrutural. Dentre as mais populares, destacam-se as técnicas de microscopia, entre elas a microscopia de força atômica (AFM) e a microscopia eletrônica (EM).

15 As técnicas de microscopias têm sido muito empregadas para caracterizar a estrutura das inúmeras geometrias de sistemas moleculares, principalmente porque oferecem a possibilidade de visualizar direta e individualmente as micro ou nanoestruturas [9]. Uma das grandes vantagens da utilização de EM é poder observar a forma geral, a simetria e

a geometria das estruturas, o que permite obter informações com valores dimensionais, além de, em muitos casos permitir reconstruir uma imagem tridimensional, a partir da reconstrução das imagens bidimensionais, de acordo com a simetria escolhida. Ao mesmo tempo, essas técnicas apresentam algumas limitações. No caso da microscopia
5 eletrônica (EM), por exemplo, as amostras podem ser visualizadas somente depois de sofrerem um determinado processo de preparação, que depende da amostra e do tipo de observação e análise pretendida. Em muitos casos, estas etapas de preparação acabam por alterar o equilíbrio do sistema, ou seja, alterando a medida e os resultados.

Uma alternativa as técnicas microscópicas, são as técnicas de espalhamento de radiação.
10 Nestas técnicas, um feixe de radiação incide sobre um sistema que espalha a radiação em uma determinada direção, permitindo investigar as formas, os tamanhos, estruturas e arranjos internos. Uma das principais vantagens destas aplicações é poder obter informações dos sistemas *in natura*, diretamente em solução. Por outro lado, uma das principais desvantagens, é apresentar informações do espaço recíproco (espaço de
15 Fourier) [10–14], ao contrário das técnicas microscópicas, que fornecem informações do espaço real.

As técnicas de espalhamento, diferentemente das técnicas de AFM e EM, apresentam facilidades na preparação das amostras, já que as estruturas podem ser observadas no seu estado nativo¹ além de permitir quantificar: a geometria média das estruturas ²,
20 modificações termodinâmicas e interações das estruturas em solução. Dentre as técnicas de espalhamento destacam-se, neste trabalho, a técnica de espalhamento à baixos ângulos (SAS)³.

¹Para este trabalho entende-se que o estado nativo é aquele estado em que o objeto molecular em estudo está em condições próximas da nativa, que em muitos casos significa estar disperso em uma solução.

²Cálculo realizado para uma experiência de espalhamento de radiação de uma amostra de moléculas de cerca de 500 kDa em uma concentração de cerca de $1 \mu g \setminus \mu l$ exposta a um feixe de radiação de $1 mm^2$, fornece em média informações sobre 10^{12} moléculas.

³*Small Angle Scattering*

Os experimentos de SAS são relativamente simples e aplicados, em geral, à investigação de sistemas diluídos. Para esses casos, a medida de intensidade espalhada representa uma média sobre todas as orientações do sistema, não dependendo de posições e/ou orientações específicas [15]. Se as partículas de um sistema em solução têm uma geometria simples, como esferas ou cilindros, há um conjunto de equações analíticas ou

5 ometria simples, como esferas ou cilindros, há um conjunto de equações analíticas ou semi-analíticas que permitem calcular a intensidade espalhada. No entanto, se as partículas apresentam geometrias ou formas um pouco mais complexas, torna-se necessário aplicar um conjunto de métodos numéricos para realizar o cálculo da intensidade de espalhamento [2,10,15–20]. Para os casos, nos quais os modelos são de partículas únicas

10 ou unitárias, existem abordagens baseadas em métodos de Monte Carlo (MC), Transformada Rápida de Fourier (FFT) e Transformada Indireta de Fourier (IFT) [21]. Por outro lado, para calcular a intensidade espalhada de sistemas compostos de partículas espacialmente correlacionadas ou orientadas, os métodos analíticos e numéricos são mais restritos. Em 1998 e 2002 McAlister & Grady [19,20] propuseram simular as in-

15 tensidades de espalhamento de sistemas orientados de partículas por meio do método de MC e recentemente, Fritz-Popovski [14] propôs analisar sistemas orientados em duas dimensões por meio da extensão do método de IFT.

Em recentes aplicações, técnicas de baixa resolução como as técnicas de espalhamentos a baixos ângulos SAS, têm sido empregadas no estudo de sistemas auto-organizados de

20 DNA “projetados” por bioinformática e biogenética [22]. Nestes casos, essas técnicas têm contribuído na obtenção de parâmetros físicos das estruturas, tais como: Raio de giro (R_g), polidispersão e ainda na identificação da forma geométrica média. Embora as técnicas de SAS tenham sido usadas para caracterizar sistemas auto-organizados, a grande maioria destes sistemas é ainda caracterizada com uso das técnicas de EM. Um

25 dos motivos que tornam as técnicas de microscopias, muito utilizadas, é apresentarem informações no espaço real de objetos unitários. Contudo, essas informações são obtidas, em geral, a partir de um número pequeno de estruturas comparadas ao número de

estruturas medidas por experimentos de SAS. No caso de Crio Microscopia Eletrônica (*Cryo*-EM) esse valor pode chegar a cerca 10^4 moléculas ⁴ enquanto num experimento de SAS esse número pode chegar a 10^{13} moléculas ⁵. Apesar das limitações de cada tipo de técnica, juntas elas podem ser complementares. Isso é possível porque, de um lado as técnicas de EM apresentam informações no espaço real de partículas individuais, e do outro, as técnicas de SAS apresentam informações no espaço recíproco que representam uma média do comportamento do sistema [22]. Quando unidas, contribuem para obter os parâmetros físicos de partículas individuais e do sistema como um todo [23–25].

Por outro lado, para realizar a caracterização estrutural de sistemas auto-organizados com uso de técnicas de alta resolução, como Ressonância Magnética Nuclear (NMR) e Cristalografia, existem uma série de dificuldades. Entre as dificuldades, destaca-se, para o caso da Cristalografia, a dificuldade de obter cristais com boa qualidade de difração; e para o caso de NMR, limitações no tamanho ou peso molecular que não pode ultrapassar 100 kDa.

Um outro experimento, baseado no espalhamento de radiação, que começa a ser aplicada ao estudo de sistemas moleculares, é o de Laser de Raio-X de Elétrons-livres (XFEL) ⁶. Neste experimento, é possível irradiar, com um feixe de raios-x intenso e de pulsos curtos, um pequeno número de objetos idênticos [26]. A análise dos dados destes experimentos poder ser realizada por um conjunto de ferramentas [26–41], que têm como base principal, o método proposto por Kam [42]. Basicamente, nestas abordagens é possível realizar a recuperação da imagem do espaço real, de um objeto espalhador, por meio da análise de funções de autocorrelação angulares de padrões de difração. Nestes casos, o espaço real é recuperado a partir de um conjunto de dados instantâneos de padrões de difração (imagens) orientados em volume tridimensional no espaço recíproco [43]. Em outras palavras, um conjunto de imagens orientadas são

⁴Esse número pode ser maior ou menor, mas é sempre em torno de algumas centenas ou milhares

⁵A média dessas estruturas é realizada com cerca 10^{11} - 10^{13} espalhadores.

⁶X-ray free-electron lasers

correlacionadas no espaço recíproco, e a análise destas correlações permite obter, no espaço real, a estrutura individual do objeto espalhador. Estas novas aplicações são possíveis porque, nestas experiências, o comprimento de pulso de radiação incidente (feixe coerente de raios X), ou o tempo necessário para obter uma imagem de cada
5 objeto, é menor do que a sua difusão de rotação [26].

Uma das grandes dificuldades intrínsecas ao emprego das técnicas de espalhamento, na determinação de nanoestruturas de sistemas auto-organizados, é que em muitos casos, para um mesmo conjunto de dados, existem um conjunto de diferentes modelos que podem representar igualmente os dados. Para os sistemas auto-organizados ainda existe
10 um outro problema, que é o fato dos métodos analíticos estarem restritos a um conjunto limitado de geometrias. Para o caso de SAS, por exemplo, existem métodos do tipo *ab initio* que permitem obter modelos tridimensionais diretamente de uma curva unidimensional [17, 18, 44, 45], mas mesmo nestes casos, é possível obter diferentes resultados para os mesmos conjuntos de dados. Essa dificuldade surge principalmente porque,
15 durante o processo de medida, ocorre uma grande perda de informações. Em um experimento de espalhamento à baixos ângulos, por exemplo, um objeto espalha radiação que é mensurada em detectores unidimensionais ou bidimensionais e já nessa operação existe a perda das informações relativas às fases da radiação da onda espalhada. Além disso, o acesso experimental, acontece somente em uma pequena região de espaço re-
20 cíproco, o que automaticamente limita a quantidade de informação disponível. Desta forma, a caracterização de sistemas auto-organizados, baseados em geometrias de alta simetria, pode ser bastante restrita. Um dos sistemas auto-organizados que inspiraram essa tese, foram os sistemas de DNA [23, 25] que são, em sua maioria, sintetizados em geometrias de alta simetria, compondo vários tipos de poliedros.

25 Além das técnicas experimentais de caracterização estrutural, as técnicas de modelagem e simulação da dinâmica destes sistemas vêm ganhando espaço e contribuindo para completar a caracterização de nanoestruturas. Em um trabalho pioneiro de aplicação

de Dinâmica Molecular (DM) em sistemas auto-organizados e tridimensionais de DNA, Falconi et al. [24] aplicaram a técnica baseados nos resultados experimentais obtidos com técnicas de SAXS e Microscopias. A DM neste caso, contribuiu para realizar as simulações de sistemas reais em nível atomístico colaborando para compreensão do comportamento de parâmetros físicos e químicos. Muitos destes parâmetros são, em muitos casos, complicados ou ainda impossíveis de serem obtidos com as atuais técnicas experimentais, dentre estes destacam-se, as interações do tipo $\pi - \textit{stacking}$, superfície *buried*, número de *clusters* mais ocupados (em um trajetória), ligações de H, entre outros. O trabalho de Falconi et al. [24] é um exemplo de que é possível aumentar a resolução na identificação das variáveis responsáveis pela estabilidade de nanoestruturas além de obter um conjunto de parâmetros físicos e químicos que são difíceis de obter experimentalmente para esses tipos de sistemas, o que contribuiu para aumentar a compreensão sobre a eficiência de montagem, geometrias, estabilidade e funcionalidades de um sistema específico [7, 24, 25].

Assim, em busca de métodos e ferramentas que permitam desvendar as estruturas de sistemas moleculares auto organizados, que em geral apresentam alta simetria, mas também aqueles que são não centro-simétricos, em suas diversas composições, é que o presente trabalho se justifica. Desta forma, optou-se por escolher a técnica de SAS buscando contribuir para o avanço da mesma, aplicada à investigação de estruturas de sistemas nanométricos. Observada uma lacuna científica, na qual sistemas biomoleculares de alta simetria não podiam ser analisados a partir de um baixo número de parâmetros, o presente trabalho desenvolveu abordagens, aplicadas à modelagem de sistemas de alta simetria e a simulação de dados de espalhamento, para o caso de curvas unidimensionais e imagens de espalhamento bidimensionais. Para realizar esse objetivo, foi desenvolvido um software em FortranTM que modela sistemas de alta simetria, e simula curvas de intensidade de espalhamento em uma e duas dimensões.

Inspirados em trabalhos prévios de sistemas auto organizados de DNA [7, 24, 25] e

motivados pela compreensão dos parâmetros que conferem estabilidade a este tipo de sistema, foram realizados também estudos com uso de DM. Neste caso, foram avaliados os parâmetros de *Root Mean Square Deviations* (RMSD), *Root Mean Square Fluctuations* (RMSF), Rg, ligações de H, número de interações do tipo $\pi - \text{stacking}$. Nesta etapa, foram avaliadas duas estruturas auto-organizadas de DNA na forma de octaedro truncado, no qual as arestas eram constituídas de duplas hélices de DNA com 18 pares de bases cada. A ligação entre cada hélice, das arestas dos octaedro truncado, foi realizada por meio de *linkers* de hélices simples de DNA. As sequências das duplas hélices podem ser vistas no apêndice C e no caso do *linker* todas as sequências para uma *nanocage* era de 7 Adeninas e da outra era de ATATATA.

Breve Resumo das Atividades Realizadas

Durante o período de realização deste trabalho foram realizadas e publicadas as seguintes atividades:

- Entre 2012 e 2013 foi realizado um projeto com a ajuda do programa Ciência Sem Fronteiras (CSF) ⁷, e o autor deste trabalho realizou um estágio de 6 meses no Grupo de Biologia Estrutural na Universidade de Roma - Tor Vergata sob supervisão do professor Alessandro Desideri. O estágio foi realizado com o objetivo de realizar o *design* e modelagem de nanoestruturas atômicas, simulação e análises de DM.
- Cassio Alves, Jan Skov Pedersen, and Cristiano Luis Pinto Oliveira. Modelling of High Symmetry Nanoscale Particles by Small-Angle Scattering. *Journal of Applied Crystallography*, 47:84-94, 2014. doi:10.1107/S1600576713028549
- Federico Iacovelli, Cassio Alves, Mattia Falconi, Francesco Oteri, Cristiano L. P. de Oliveira, and Alessandro Desideri. Influence of the single-strand linker com-

⁷ Processo: 237230/2012-7/SWE.

position on the structural/dynamical properties of a truncated octahedral DNA nano-cage family. *Biopolymers*, 101(10):992-999, 2014. doi:10.1002/bip.22475

- Cassio Alves and Cristiano Luis Pinto Oliveira. Simulation of two-dimensional image by scattered intensities of Nanoscale Particles. Em preparação.

CAPÍTULO 2

FUNDAMENTOS TEÓRICOS

5

O principal objetivo deste trabalho é desenvolver métodos para a análise e estudo da estrutura tridimensional, de sistemas nanométricos, a partir do ajuste de dados simulados, com uso de modelos obtidos *a priori*, contra dados experimentais da intensidade de espalhamento de curvas unidimensionais ou bidimensionais, do espaço recíproco. Neste caso, a técnica de caracterização estrutural abordada é a técnica de SAS, e neste capítulo será descrito o fenômeno de espalhamento a baixos ângulos e algumas abordagens disponíveis para análise de dados da intensidade espalhada. Além do objetivo principal, os itens finais descrevem brevemente, o estado da arte de sistemas auto-organizados compostos de DNA, sistemas que motivaram grande parte dos desenvolvimentos deste trabalho, e que foram estudados com uso da técnica de Dinâmica Molecular (DM), descrita sucintamente ao final do capítulo.

10

15

2.1 Descrição do Fenômeno de Espalhamento de Radiação

Existem diferentes descrições dos fenômenos de interação de radiação com a matéria. Neste trabalho o interesse está no espalhamento de um feixe incidente por um espalhador, no qual o espalhamento é elástico ou quase elástico e não afeta as estruturas internas do espalhador, de modo que, o foco é o do espalhamento de radiação por um Potencial $V(\mathbf{r})$.

2.1.1 Espalhamento Rayleigh ou o Espalhamento por um Potencial

A abordagem realizada, nesta seção, é a descrição do fenômeno de espalhamento de um feixe de radiação incidente por um potencial espalhador¹. A descrição será realizada com foco no espalhamento de fótons de raios X, mas também pode ser usada para descrever, com pequenas alterações, o espalhamento por nêutrons e por luz [15]. O potencial é de um elétron ligado a um núcleo e nesta abordagem, o conceito de localização, do elétron, será substituído pelo de densidade de probabilidade, de encontrar o elétron em uma dada posição [46]. Essa densidade de probabilidade, pode ser vista como uma nuvem eletrônica que interage com a radiação incidente, que é espalhada com o mesmo ou quase o mesmo comprimento de onda incidente (espalhamento elástico ou quase elástico). Nesta abordagem, os tempos de interações são pequenos e consequentemente será aplicada a aproximação de Born no caso não-relativístico [46].

Para realizar esse estudo, algumas hipóteses simplificadoras serão admitidas [1].

1. As partículas não tem spins;
2. O espalhamento é elástico ou quase elástico e não afetará as estruturas internas das partículas envolvidas;
3. O alvo é fino o suficiente, o que permite negligenciar múltiplos processos de esp-

¹ O estudo desta seção foi baseado na referência [1].

lhamento.

4. A possibilidade de coerência entre as ondas espalhadas será negligenciada.
5. A interação entre as partículas 1 e 2 podem ser descritas pela energia potencial $V(\mathbf{r}_2 - \mathbf{r}_1)$ que depende exclusivamente de suas posições relativas, $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. O potencial $V(\mathbf{r})$ cai mais rapidamente que $\frac{1}{r}$ quando $r \rightarrow \infty$, excluindo neste caso o potencial de Coulomb. O centro de massa será $\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$.

Tratando-se de um espalhamento elástico ou quase elástico, de uma partícula incidente por um potencial em uma escala atômica ou nuclear, a evolução da função de onda da partícula incidente, usando a Mecânica Quântica, estará associada com as interações entre a partícula incidente e o alvo.

Assim, para caracterizar o fenômeno de espalhamento assume-se que o potencial espalhador, $V(\mathbf{r})$, está localizado em torno da origem O do sistema de coordenadas, figura(2.1), e que Oz é a direção de incidência de partículas de massa μ . O fluxo das partículas incidentes é dado por F_i , que nada mais é do que o fluxo de partículas por unidade de tempo que atravessam a superfície perpendicular a Oz. O número de partículas n , espalhadas por unidade de tempo dn , no ângulo sólido Ω na direção $\mathbf{r} = \Omega(\theta, \phi)$ será dado por,

$$dn = F_i \sigma(\theta, \phi) d\Omega \quad (2.1)$$

Neste caso, a equação de Schrödinger, equação (2.4), descreve a evolução de uma partícula que interage com o potencial $V(\mathbf{r})$, e tem como solução a equação (2.5) com os estados estacionários da energia E associados.

$$E = \frac{\hbar^2 k^2}{2\mu}, \quad (2.2)$$

Figura 2.1: Feixe incidente é paralelo ao eixo Oz . A zona de influência do potencial $V(\mathbf{r})$ está centrada em O . O detector mede o número de partículas dn espalhadas por unidade de ângulo sólido $d\Omega$. O coeficiente $d\sigma(\theta, \phi)$ é a seção de choque diferencial de espalhamento na direção (θ, ϕ) [1]. Figura adaptada de [1].

$$V(\mathbf{r}) = \frac{\hbar^2}{2\mu} U(\mathbf{r}). \quad (2.3)$$

$$\left[-\frac{\hbar^2 \nabla^2}{2\mu} + V(\mathbf{r}, t) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r}, t) \quad (2.4)$$

Assumindo somente as soluções, da equação (2.5) de Schrödinger, com energia E positiva, que é igual a energia cinética da partícula incidente antes da interação com o potencial, e sabendo que $\psi(\mathbf{r})$ é uma solução da equação de autovalores,

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-iEt/\hbar}, \quad (2.5)$$

é possível escrever a função de onda estacionária como:

$$(\nabla^2 + k^2) \psi(\mathbf{r}) = U(\mathbf{r}) \psi(\mathbf{r}), \quad (2.6)$$

- 5 Antes de chegar na zona de influência do potencial $V(\mathbf{r})$, a partícula incidente está livre e seu estado é representado por um pacote de onda plana com termo o e^{ikz} . Assim,

quando o pacote de onda incidente aproxima-se da região de influência de $V(\mathbf{r})$, sua estrutura é profundamente modificada e a sua evolução é complicada. Contudo, para regiões distantes da influência do potencial o pacote de onda está dividido em um pacote, transmitido, que se propaga ao longo da direção positiva de Oz com termo e^{ikz} ,
 5 e um pacote de onda espalhado. Para valores de $\mathbf{r}$ muito grandes, o valor do potencial $V(\mathbf{r})$ é praticamente zero e o pacote de onda incidente não sofre influência do potencial, de modo que para $r \rightarrow \infty$, o pacote de onda espalhado não é modificado pelo potencial. Assim, a função de onda $\Psi(\mathbf{r})$ representa o estado de espalhamento estacionário com uma dada energia $E = \hbar^2 k^2 / 2\mu$ que pode ser obtida por meio da superposição da
 10 onda plana transmitida (e^{ikz}) e da onda espalhada. Como a onda espalhada depende da interação com o potencial $V(\mathbf{r})$, mas a sua descrição é realizada para uma região livre da influência do potencial, a função de onda espalhada deve apresentar algumas características para valores grandes de $\mathbf{r}$ [1]:

- a amplitude de espalhamento depende da direção (θ, ϕ) ;
- em uma dada direção (θ, ϕ) a dependência é radial da forma e^{ikr}/r e

$$(\nabla^2 + k^2) \frac{e^{ikr}}{r} = 0, \quad (2.7)$$

15 para $r \geq r_0$, $r_0 > 0$;

Desta forma, a solução² da função de onda, equação (2.6), associada com o estado de

²A solução da equação (2.6) é obtida para cada k , que satisfaz a equação (2.11), de forma que, para obter a evolução temporal, do pacote de onda que representa a partícula incidente, é necessário expandir em termos dos autoestados do Hamiltoniano total. Assim a função de onda é:

$$\Psi(\mathbf{r}, t)_{r \rightarrow \infty} \sim \int_0^x \Psi_k(\mathbf{r}) e^{-iE_k t / \hbar} g(k) dk \quad (2.8)$$

espalhamento estacionário é dada por:

$$\Psi(\mathbf{r}) \sim \Psi_0(\mathbf{r}) + \int \Psi(\mathbf{r}') G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') d^3 r', \quad (2.9)$$

em que, $\Psi_0(\mathbf{r})$ é a solução da equação homogênea $(\nabla^2 + k^2) \Psi_0(\mathbf{r}) = 0$ e $G(\mathbf{r})$ é a função de Green do operador $(\nabla^2 + k^2)$, que satisfaz $(\nabla^2 + k^2) G(\mathbf{r}) = \delta(\mathbf{r})$ e é igual a:

$$G(\mathbf{r}) = -\frac{1}{4\pi} \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \sim_{r \rightarrow \infty} -\frac{1}{4\pi} \frac{e^{ikr}}{r} e^{-ik\mathbf{u}\mathbf{r}'} \quad (2.10)$$

de modo que, $\mathbf{u}$ é um vetor unitário, conforme figura (2.2).

Figura 2.2: Representação das distâncias de influência do potencial, em que $|\mathbf{r} - \mathbf{r}'|$ é a distância entre o ponto O e M. O ponto P está situado dentro da zona de influência do Potencial que tem uma zona de influência da ordem de L. A distância entre O e P é muito pequena comparada com a distância entre O e M [1]. Figura adaptada de [1].

- 5 A onda plana incidente apresenta a forma $\Psi_0(\mathbf{r}) = e^{ikz}$ e como o ponto de observação é muito maior que a distância de influência do centro espalhador, $|\mathbf{r} - \mathbf{r}'| \cong r - \mathbf{u}\mathbf{r}'$, pode-se assumir uma solução assintótica, com uso da equação (2.10), de modo a reescrever

equação (2.9) e obter a função de onda espalhada no estado estacionário,

$$\Psi(\mathbf{r})_{r \rightarrow \infty} \sim e^{ikz} - \int \Psi(\mathbf{r}') \left(\frac{1}{4\pi} \frac{e^{ikr}}{r} e^{-ik\mathbf{u}\mathbf{r}'} \right) U(\mathbf{r}') d^3r' \quad (2.11)$$

Desta forma, é possível definir a amplitude de espalhamento $A_k(\theta, \phi)$, que depende do potencial espalhador $V(\mathbf{r})$, como:

$$A_k(\theta, \phi) = \frac{1}{4\pi} \int \Psi(\mathbf{r}') e^{-ik\mathbf{u}\mathbf{r}'} U(\mathbf{r}') d^3r', \quad (2.12)$$

A função de onda, no estado estacionário, equação (2.13), tem dois termos de acordo com:

$$\Psi(\mathbf{r})_{r \rightarrow \infty} \sim e^{ikz} - A_k(\theta, \phi) \frac{e^{ikr}}{r}. \quad (2.13)$$

O primeiro termo, descreve a onda incidente que não foi espalhada e o segundo termo
 5 representa a onda espalhada por um potencial. Como a equação (2.11) apresenta um fator desconhecido, $\Psi(\mathbf{r}')$, será realizada a expansão de Born, no estado estacionário da função de onda espalhada, para poder resolver essa equação.

Aproximação de Born

Para que a equação (2.13) seja útil, é necessário determinar a amplitude de espalha-
 10 mento $A_k(\theta, \phi)$. Quando a energia de colisão é grande ou quando o potencial espalhador é pequeno, ou ainda, ambas as condições, é possível usar a aproximação de Born para determinar esta amplitude. Neste sentido, e na falta de informações do $\Psi(\mathbf{r}')$, será realizada uma mudança de notação, $\mathbf{r} \Rightarrow \mathbf{r}'$ e $\mathbf{r}' \Rightarrow \mathbf{r}''$ de modo a obter:

$$\Psi_k(\mathbf{r}) = e^{ik\mathbf{i}\mathbf{r}'} + \int \Psi(\mathbf{r}'') G(\mathbf{r}' - \mathbf{r}'') U(\mathbf{r}'') d^3r''. \quad (2.14)$$

A inserção da equação (2.14) na equação (2.9), permite obter:

$$\begin{aligned}\Psi_k(\mathbf{r}) &= e^{i\mathbf{k}_i\mathbf{r}'} + \int G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') e^{i\mathbf{k}_i\mathbf{r}'} d^3r' \\ &\quad + \int d^3r' \int G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') G(\mathbf{r}' - \mathbf{r}'') \Psi(\mathbf{r}'') U(\mathbf{r}'') d^3r''.\end{aligned}\quad (2.15)$$

O procedimento anterior de mudança, $\mathbf{r} \Rightarrow \mathbf{r}'$ e $\mathbf{r}' \Rightarrow \mathbf{r}''$, pode ser repetido, da seguinte forma:

$$\begin{aligned}\Psi_k(\mathbf{r}) &= e^{i\mathbf{k}_i\mathbf{r}'} + \int G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') e^{i\mathbf{k}_i\mathbf{r}'} d^3r' \\ &\quad + \int d^3r' \int G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') G(\mathbf{r}' - \mathbf{r}'') \Psi(\mathbf{r}'') U(\mathbf{r}'') e^{i\mathbf{k}_i\mathbf{r}'} d^3r'' \\ &\quad + \int d^3r' \int d^3r'' \int G(\mathbf{r} - \mathbf{r}') U(\mathbf{r}') G(\mathbf{r}' - \mathbf{r}'') \Psi(\mathbf{r}'') U(\mathbf{r}'') \\ &\quad \times G(\mathbf{r}'' - \mathbf{r}''') U(\mathbf{r}''') \Psi_k(\mathbf{r}''') d^3r'''.\end{aligned}\quad (2.16)$$

Na equação (2.16) os três primeiros termos são conhecidos e o termo desconhecido $\Psi(\mathbf{r}''')$ é o quarto termo. Desta forma, cada termo apresenta uma potência maior que o termo seguinte, assim se o potencial é fraco a cada termo sucessivo e é menor que o precedente, pode-se negligenciar o termo mais a direita, que apresenta $\Psi(\mathbf{r}''')$, em relação aos termos mais a esquerda, obtendo a expansão de Born. O uso da primeira ordem da expansão de Born em $U(\mathbf{r})$ permite reescrever a equação (2.12) substituindo $\Psi(\mathbf{r}')$ por $e^{i\mathbf{k}_i\mathbf{r}'}$ e obter a amplitude de espalhamento:

$$A_k^{Born}(\theta, \phi) = -\frac{1}{4\pi} \int e^{i\mathbf{k}_i \mathbf{r}'} e^{-ik\mathbf{u} \mathbf{r}'} U(\mathbf{r}') d^3 r' \quad (2.17)$$

$$A_k^{Born}(\theta, \phi) = -\frac{1}{4\pi} \int e^{-i(\mathbf{k}_s - \mathbf{k}_i) \mathbf{r}'} U(\mathbf{r}') d^3 r' \quad (2.18)$$

$$A_k(\mathbf{q}) = -\frac{1}{4\pi} \int e^{-i\mathbf{q} \mathbf{r}'} U(\mathbf{r}') d^3 r' \quad (2.19)$$

Assim a amplitude de espalhamento é:

$$A_k(\mathbf{q}) = -\frac{\mu}{2\pi\hbar^2} \int e^{-i\mathbf{q} \mathbf{r}} V(\mathbf{r}) d^3 r. \quad (2.20)$$

A amplitude de espalhamento usou o vetor de espalhamento $\mathbf{q}$ obtido com uso da análise da geometria de espalhamento. Nesta análise, o vetor de onda incidente $\mathbf{k}_i$, $|\mathbf{k}_i| = k = \frac{2\pi}{\lambda}$, está na direção do feixe incidente Oz de modo que

$$e^{ikz} = e^{i\mathbf{k}_i \mathbf{r}},$$

e vetor de onda espalhado $\mathbf{k}_s$, $|\mathbf{k}_s| = k$ tem direção dos ângulos θ e ϕ , de modo que

$$\mathbf{k}_s = k \cdot \mathbf{u}.$$

O que permite concluir que o vetor de onda transferido $\mathbf{q}$ na direção (θ, ϕ) é a diferença entre $\mathbf{k}_s$ e $\mathbf{k}_i$, de acordo com a figura (2.3), [47].

$$\mathbf{q} = \mathbf{k}_s - \mathbf{k}_i \Rightarrow q = 2k \sin(\theta) \Rightarrow q = 4 \frac{\pi}{\lambda} \sin\theta \quad (2.21)$$

A ligação entre a descrição teórica, do fenômeno de espalhamento de um pacote de
5 onda incidente por um potencial espalhador, e o respectivo experimento, é realizada

Figura 2.3: Diagrama da representação entre os vetores de onda incidente $\mathbf{k}_i$ e espalhado $\mathbf{k}_s$ e vetor de onda transferido $\mathbf{q}$ [1]. Figura adaptada de [1].

pela medida experimental do espalhamento. Em um experimento de espalhamento não se mede a amplitude de espalhamento, $A_k(\theta, \phi)$, mas o fluxo de energia espalhada, ou ainda, o número de partículas espalhadas (fótons ou nêutrons) que é proporcional ao quadrado da amplitude de espalhamento. Para entender o fenômeno de espalhamento
 5 do ponto de vista experimental é necessário descrever o fluxo de energia ou partículas que atravessam uma determinada região, que neste caso, chamaremos de seção de choque diferencial, porque depende de θ e ϕ .

Espalhamento Medido na Seção de Choque

Em um experimento de espalhamento de radiação é possível medir o fluxo de energia
 10 espalhado ou número de partículas espalhadas (fótons ou nêutrons), que atravessam uma determinada seção de choque diferencial. Essa medida é proporcional ao quadrado da amplitude de espalhamento e é conhecida como intensidade espalhada [48]. Neste caso, a seção de choque total é dada por:

$$\sigma(\theta, \phi) = \int_{\Omega} \frac{d\sigma}{d\Omega} d\Omega = \int_{\Omega} |A_k(\theta, \phi)|^2 d\Omega = \int_{\Omega} I(\mathbf{q}) d\Omega. \quad (2.22)$$

A seção de choque de espalhamento, $\frac{d\sigma}{d\Omega}$, é igual ao número de partículas espalhadas,

por unidade de tempo e por unidade de fluxo incidente, dentro do ângulo sólido $d\Omega$ na direção $\mathbf{r} = \Omega(\theta, \phi)$, ou seja, é igual ao fluxo de partículas espalhadas J_s , pelo potencial espalhador, que atravessam a superfície $d\Omega$ dividido pelo fluxo incidente J_i ,

$$\frac{d\sigma}{d\Omega}(\theta, \phi) = \frac{J_s}{J_i} r^2 d\Omega. \quad (2.23)$$

Para determinar a densidade de probabilidade de encontrar a partícula em um estado estacionário, em que $\varrho(\mathbf{r}, t)$ é a densidade de probabilidade e $\mathbf{J}(\mathbf{r}, t)$ é a densidade de corrente de probabilidade, será utilizada a equação da continuidade,

$$\frac{\partial \varrho}{\partial t}(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) = 0. \quad (2.24)$$

A densidade de probabilidade de encontrar a partícula em um estado estacionário pode ser obtida por meio da densidade de corrente de probabilidade ($\mathbf{J}$) que é dada segundo a equação (2.25).

$$\mathbf{J}(\mathbf{r}) = \frac{1}{r} \text{Re} \left[\psi^*(\mathbf{r}) \frac{\hbar}{i} \nabla \psi(\mathbf{r}) \right]. \quad (2.25)$$

Para o caso de uma onda plana incidente, $\psi(\mathbf{r}) = e^{ikz}$, a densidade de corrente de probabilidade $\mathbf{J}_i$ é:

$$|\mathbf{J}_i| = \frac{\hbar k}{\mu}. \quad (2.26)$$

O fluxo incidente F_i (número de partículas do feixe incidente que atravessa uma unidade de superfície perpendicular a Oz em uma unidade de tempo), que é proporcional a

densidade de corrente de probabilidade $\mathbf{J}_i$, é dado por:

$$F_i = C|\mathbf{J}_i| = C \frac{\hbar k}{\mu}, \quad (2.27)$$

onde C é uma constante.

De modo similar, o número de partículas dn (fótons ou nêutrons) que atravessam a superfície $d\Omega$ e chegam ao detector em uma unidade de tempo é proporcional a densidade
5 de probabilidade J_s .

A densidade de corrente de probabilidade da onda espalhada $\mathbf{J}_s$ ³ é obtida por meio da equação (2.11) com a substituição de $\psi(\mathbf{r})$ por $A_k(\theta, \phi)e^{ikr}/r$:

$$(\mathbf{J}_s)_r = \frac{\hbar k}{\mu} \frac{1}{r^2} |A_k(\theta, \phi)|^2, \quad (2.28)$$

o que permite obter o número de partículas espalhadas:

$$dn = C\mathbf{J}_s \cdot d\Omega = C \frac{\hbar k}{\mu} |A_k(\theta, \phi)|^2 d\Omega. \quad (2.29)$$

Como discutido no subitens anteriores, assume-se que r é grande tornando $\mathbf{J}_{s\theta}$ e $\mathbf{J}_{s\phi}$
10 negligenciáveis, de forma que a densidade de corrente de probabilidade $\mathbf{J}_s$ é praticamente radial.

Assim, a seção de choque diferencial ou seção de choque de espalhamento, equação (2.30), pode ser finalmente encontrada com a substituição de F_i e dn na primeira equação desta seção, equação (2.1).

$$\frac{d\sigma}{d\Omega}(\theta, \phi) = |A_k(\theta, \phi)|^2 \quad (2.30)$$

³em coordenadas esféricas

Uma outra forma de escrever a seção de choque de espalhamento, equação (2.30), é com uso das equações (2.19 e 2.3) [1]:

$$\frac{d\sigma^{Born}}{d\Omega_k}(\theta, \phi) = \left(\frac{\mu}{2\pi\hbar}\right)^2 \left| \int e^{-i\mathbf{q}\mathbf{r}} V(\mathbf{r}) d^3r \right|^2. \quad (2.31)$$

A seção de espalhamento, equação (2.31), foi calculada considerando que o potencial espalhador $V(\mathbf{r})$ era pontual, contudo, fontes importantes de interação possuem uma
 5 estrutura interna mais ou menos extensa que faz com que o potencial tenha uma distribuição espacial significativa [49].

Um exemplo deste tipo de interação pode ser assumido quando um elétron interage com um átomo, de núcleo positivo e pontual, que possui elétrons orbitais produzindo uma densidade de carga que estende-se por alguns raios de Born. Considerando, como exem-
 10 plo um caso eletrostático, em que a interação do potencial é dada com uma distribuição de cargas de densidade $\rho(\mathbf{r})$, de acordo com:

$$V(r) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r'. \quad (2.32)$$

A substituição do potencial de interação, equação (2.32), na amplitude de espalhamento, equação (2.20), permite obter:

$$A_k(\mathbf{q}) = \frac{\mu}{2\pi\hbar} \int e^{-i\mathbf{q}\mathbf{r}} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d^3r d^3r' \quad (2.33)$$

$$A_k(\mathbf{q}) = \frac{\mu}{2\pi\hbar} \int \frac{e^{-i\mathbf{q}\mathbf{r}}}{|\mathbf{r} - \mathbf{r}'|} d^3r \int \rho(\mathbf{r}') e^{-i\mathbf{q}\mathbf{r}'} d^3r' \quad (2.34)$$

Desta forma, a segunda integral da equação (2.34), é o fator de forma do átomo $F(\mathbf{q})$:

$$F(\mathbf{q}) = \int \rho(\mathbf{r}') e^{-i\mathbf{q}\mathbf{r}'} d^3r', \quad (2.35)$$

que composta com a amplitude de espalhamento pontual, da ordem de $|\mathbf{r} - \mathbf{r}'|$, fornece:

$$A_k(\mathbf{q}) = \frac{\mu}{2\pi\hbar} F(\mathbf{q}) \int \frac{e^{-i\mathbf{q}\mathbf{r}}}{|\mathbf{r} - \mathbf{r}'|} d^3r \quad (2.36)$$

De maneira simplificada, a amplitude de espalhamento de um potencial espalhador extenso é dada pelo espalhamento pontual e pelo fator de forma, uma espécie de raio
5 de influência da amplitude de espalhamento pontual, de acordo com:

$$A_k(\mathbf{q}) = F(\mathbf{q}) A_k(\mathbf{q})^{pontual} \quad (2.37)$$

ou ainda,

$$A_k(\mathbf{q}) = F(\mathbf{q}) A_e(\mathbf{q})^{eletron}. \quad (2.38)$$

Desta forma, a equação (2.37), aplicada a seção de choque diferencial permite ver que é possível determinar o fator de forma do átomo $F(\mathbf{q})$ (ou os fatores de forma) da distribuição do potencial, a partir do estudo do espalhamento produzido por um objeto
10 espalhador com estrutura, $\frac{d\sigma}{d\Omega}|_{extensa}$, e da seção de choque teórica, $\frac{d\sigma}{d\Omega}|_{pontual}$, conhecendo previamente o potencial de interação [49], de acordo com:

$$\left. \frac{d\sigma}{d\Omega} \right|_{extensa} = F(\mathbf{q}) \left. \frac{d\sigma}{d\Omega} \right|_{pontual}. \quad (2.39)$$

Interpretação das fórmulas

Na seção 2.1.1, foi assumido que a zona de influência do meio, no qual atua o potencial espalhador, apresenta densidade proporcional a $U(\mathbf{r})$ e que a função $G(\mathbf{r} - \mathbf{r}')$ na equação (2.10) representa a amplitude em um ponto $\mathbf{r}$ de uma onda irradiada por uma fonte pontual situada em $\mathbf{r}'$. Por meio destas restrições, pode-se afirmar que os primeiros dois termos da equação (2.16) descrevem o total da onda espalhada em um ponto $\mathbf{r}$ como resultado da superposição da onda incidente, $e^{i\mathbf{k}_i\mathbf{r}}$, e um número infinito de ondas vindas da fonte secundária, que são as ondas incidentes que foram induzidas pelo meio espalhador. Dessa forma, a amplitude de cada onda espalhada é proporcional a onda incidente, $e^{i\mathbf{k}_i\mathbf{r}}$, e a densidade de material espalhado $U(\mathbf{r}')$, calculado no ponto $\mathbf{r}'$, figura (2.4). Essa interpretação remete ao *princípio de Huygens*.

Como o meio espalhador estende-se em uma determinada área, a fonte secundária é excitada não somente pela onda incidente, mas também pela onda espalhada vindas de outras fontes secundárias, que podem ser desconsideradas se $U(\mathbf{r}')$ é muito pequeno.

Figura 2.4: Representação esquemática: a) onda incidente e onda espalhada pela interação com o potencial, aproximação de Born; b) ondas espalhadas duas vezes pelo potencial, termo de segunda ordem em U na expansão de Born [1]. Figura adaptada de [1].

A partir dos resultados obtidos é possível formular a solução da equação de Schrodinger,

equação(2.4), que descreve o fenômeno do espalhamento elástico, no limite assintótico de uma onda por um potencial [1], como:

$$\Psi(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} - \left(\frac{1}{4\pi} \frac{e^{ikr}}{r} \right) A_k(\mathbf{q}) \quad (2.40)$$

Por último, para o caso de experimentos de espalhamento com ângulos pequenos, a seguinte inequação delimita o tamanho da unidade espalhadora:

$$qd = \frac{4\pi \sin(\theta)}{\lambda} d \ll 1. \quad (2.41)$$

5 Deste modo, o vetor de espalhamento máximo e mínimo são dados de acordo com os raios de interação de cada espalhador:

$$q_{min} = \frac{\pi}{R_{max}} \quad (2.42)$$

$$q_{max} = \frac{\pi}{R_{min}} \quad (2.43)$$

2.1.2 Espalhamento Discreto por Potenciais em um Meio

Na seção 2.1.1 foi abordado o fenômeno de espalhamento de uma onda por um potencial, nesta seção será feita uma generalização desse fenômeno para um conjunto de potenciais
 10 j , suspensos em um meio com constante dielétrica diferente de zero. Neste caso, além da solução obtida para o espalhamento por um potencial, equação (2.40), é necessário levar em conta, no termo da onda espalhada, o espalhamento provocado pelo potencial

do meio [15]. Desta forma, a equação que descreve a onda espalhada $\Psi(\mathbf{r})_s$ será:

$$\Psi_s(\mathbf{r}) = - \left(\frac{1}{4\pi} \frac{e^{ikr}}{r} \right) \left[\int \Delta\rho(\mathbf{r}'_j) e^{-i\mathbf{q}\mathbf{r}'_j} d^3r'_j \right] e^{-i\mathbf{q}\mathbf{r}_j}, \quad (2.44)$$

$$\Psi_s(\mathbf{r}) = - \left(\frac{1}{4\pi} \frac{e^{ikr}}{r} \right) b(\mathbf{q}) e^{-i\mathbf{q}\mathbf{r}_j} \quad (2.45)$$

em que $b(\mathbf{q})$ é comprimento de espalhamento,

$$b(\mathbf{q}) = \int \Delta\rho(\mathbf{r}'_j) e^{-i\mathbf{q}\mathbf{r}'_j} d^3r'_j. \quad (2.46)$$

Neste caso, r'_j é a posição do elemento de volume, de um potencial espalhador, relativa ao seu centro de massa; e r_j é a posição relativa a um origem arbitrária, onde está esse mesmo potencial espalhador. O termo de propagação da onda espalhada pelo conjunto de potenciais é dado por $e^{-i\mathbf{q}\mathbf{r}_j}$, de modo que, $\Delta\rho(\mathbf{r}'_j) = \rho(\mathbf{r}'_j) - \rho_{meio}$ é a composição entre o potencial dos objetos espalhadores e o potencial do meio, no qual os mesmos objetos estão suspensos [50].

A equação (2.45) permite ver que o espalhamento total é a soma dos espalhamentos dos objetos individuais, ou seja, é o produto da amplitude de espalhamento, da distribuição do potencial no interior do objeto, e o fator de fase, determinado pela posição do objeto no meio [15]. Para permitir a diferença de fase, a amplitude de espalhamento é expressa como uma quantidade complexa, de forma que, a intensidade espalhada por um sistema, é obtida pelo produto da amplitude de espalhamento com seu conjugado $A(\mathbf{q})A^*(\mathbf{q})$ [12]. Em um experimento de espalhamento de radiação mede-se a intensidade da radiação espalhada,

$$I = A(\mathbf{q})A^*(\mathbf{q}) = |A(\mathbf{q})|^2 = I_e F(\mathbf{q})F^*(\mathbf{q}) = I_e |F(\mathbf{q})|^2, \quad (2.47)$$

que é proporcional ao quadrado da amplitude de espalhamento $|A(\mathbf{q})|^2$, em que $I_e = |A_e(\mathbf{q})|^2$.

Para o caso de n espalhadores, com comprimento de espalhamento $b_j(\mathbf{q})$, a amplitude resultante para o sistema é obtida pela adição das amplitudes de onda espalhada, de
 5 cada espalhador, e o seu fator de fase:

$$A(\mathbf{q}) = A_e F(\mathbf{q}) = A_e \sum_i^n b_j(\mathbf{q}) e^{i\mathbf{q}\mathbf{r}}. \quad (2.48)$$

Se a distribuição de cargas $\Delta\rho(\mathbf{r})$ é contínua as somas da equação (2.48) são trocadas por integrais de volume:

$$A(\mathbf{q}) = A_e F(\mathbf{q}) = A_e \int_V \Delta\rho(\mathbf{r}) e^{i\mathbf{q}\mathbf{r}} dV. \quad (2.49)$$

A partir destes últimos resultados, é possível obter a intensidade da radiação espalhada, por n objetos de iguais elementos de volume, em posições relativas $\mathbf{q}(\mathbf{r}_j - \mathbf{r}_1)$:

$$I(\mathbf{q}) = I_e \sum_j^n \sum_l^n b_j(\mathbf{q}) b_l(\mathbf{q}) e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_1)} \quad (2.50)$$

10 ou assumindo somente a parte real

$$I(\mathbf{q}) = I_e \sum_j^n \sum_l^n b_j(\mathbf{q}) b_l(\mathbf{q}) \cos [\mathbf{q}(\mathbf{r}_j - \mathbf{r}_1)]. \quad (2.51)$$

No caso de distribuição de cargas contínuas, $\Delta\rho(\mathbf{r})$, as somas do comprimento de espalhamento $b_j(\mathbf{q})$ são trocadas por integrais de volume [12]:

$$I(\mathbf{q}) = I_e \int_V \int_V \Delta\rho(\mathbf{r}_j) \Delta\rho(\mathbf{r}_1) \cos [\mathbf{q}(\mathbf{r}_j - \mathbf{r}_1)] dV_j dV_l. \quad (2.52)$$

Em um experimento, no qual o feixe de radiação incidente, atinge um conjunto de n espalhadores aleatoriamente orientados, a intensidade total será uma média $\langle I(\mathbf{q}) \rangle_{orientacoes}$ sobre todas as possíveis orientações dos espalhadores, de acordo com:

$$\langle I(\mathbf{q}) \rangle = \left\langle I_e \sum_j^n \sum_l^n f_j(\mathbf{q}) f_l^*(\mathbf{q}) e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)} \right\rangle_{orientacoes}, \quad (2.53)$$

Em outras palavras, a intensidade total espalhada, é uma média em um espaço tridimensional das possíveis orientações de todos os objetos.

2.2 Espalhamento à Baixos Ângulos Aplicado a Diferentes Sistemas

No item anterior, foi discutido o fenômeno de espalhamento para um e vários potenciais espalhadores. Neste item, serão discutidas algumas aplicações à diversos sistemas usando a teoria descrita e as equações (2.50, 2.51, 2.53). Em um experimento típico de espalhamento, um feixe incide em uma amostra e é espalhado, a radiação espalhada é coletada por um detector, e em experimentos de SAS o detector pode ser unidimensional ou bidimensional. Neste último caso, a intensidade espalhada forma uma imagem bidimensional que pode ser isotrópica ou anisotrópica, dependendo do tipo de amostra espalhadora. Assim, em um experimento de espalhamento, o espaço real, ou o espaço dos vetores $\mathbf{r}$ s, espalha radiação do feixe incidente que é medida em um detector. O detector por sua vez, revela uma imagem do espaço recíproco do real, ou o espaço dos vetores $\mathbf{q}$.

Em síntese, os experimentos de SAS constituem uma forma de analisar sistemas na escala nanométrica, a mesma escala do comprimento de onda do feixe incidente, da ordem de 1 Å, a partir da transformação de dados e informações do espaço recíproco pode-se tentar obter informações correspondentes ao espaço real. O espaço recíproco,

assim como o real, é tridimensional, mas o processo de medida permite o acesso somente a uma projeção ou corte bidimensional. Desta forma, a perda de informação decorre do acesso limitado à apenas uma parte do espaço recíproco além da média orientacional realizada pelo espalhador. Como a média orientacional é realizada sobre
5 todas as orientações do espalhador a intensidade de espalhamento bidimensional é independente do ângulo azimutal, e neste caso a imagem pode ser integrada gerando um curva unidimensional. Desta forma, o trabalho de análise de dados de espalhamento consiste sempre em tentar obter as informações do espaço real, a partir de dados no espaço recíproco. Existem muitas formas de obter as informações do espaço real, e
10 em geral, é realizado um processo de minimização de dados calculados ou simulados contra dados experimentais, obtidos a partir de uma curva unidimensional de dados de espalhamento.

O cálculo da intensidade espalhada pode ser realizado de forma analítica, semi-analítica ou ainda com uso de diferentes métodos. Dentre esses métodos, destacam-se os métodos
15 elementos finitos (MEF) [2, 20, 51, 52] e os métodos *ab initio* [17, 18, 44, 45].

Para o caso de análise de sistemas auto-organizados, os métodos *ab initio* podem apresentar modelos razoáveis no espaço real a partir de curvas unidimensionais do espaço recíproco. Mas mesmo com a possibilidade de poderem ser combinados com modelos de alta resolução, o resultado não são modelos únicos e em alguns casos eles são muito
20 diferentes daqueles esperados. Além disso, é comum, diferentes modelos ajustarem, com uma mesma precisão, os mesmos dados experimentais [53–55]. Por outro lado, os métodos de elementos finitos se combinados com informações *a priori* dos sistemas auto-organizados podem fornecer importantes resultados [23, 25].

Assim, os próximos subitens tratarão de algumas das diferentes abordagens teóricas realizadas para descrever a intensidade espalhada de sistemas: orientados, aleatoriamente
25 orientados, centro-simétricos, não centro-simétricos, monodispersos e polidispersos. O

centro da abordagem será expor o método de EF, aplicado à SAS, e discutir algumas das limitações dos métodos analíticos e semi-analíticos. A escolha pelo método de EF deve-se ao fato deste trabalho priorizar aqueles métodos que podem ser aplicados a sistemas de alta simetria, que incluem os sistemas auto-organizados de DNA, além da
5 versatilidade e facilidade de poder simular qualquer tipo de sistema com um baixo número de parâmetros, mesmo sabendo do custo computacional.

Sistemas Orientados e Aleatoriamente Orientados

Novamente, em um experimento de SAXS, o sistema em estudo pode estar orientado, ou aleatoriamente orientado. Em um sistema aleatoriamente orientado, ou diluído,
10 o feixe incidente atinge um número grande de espalhadores que estão em orientações aleatórias e neste caso, não existe correlação espacial entre os espalhadores, de modo que, a equação que descreve a intensidade espalhada é a equação (2.53), que pode ser reescrita, de acordo com o tipo de potenciais (partículas, moléculas, objetos, etc) [50]:

- potenciais não idênticos,

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = \sum_{j=1}^n \langle |b_j(\mathbf{q})|^2 \rangle + \left\langle \sum_{j \neq l}^n \sum_{l=1}^n b_j(\mathbf{q}) e^{-i\mathbf{q}(\mathbf{r}_j)} b_l^*(\mathbf{q}) e^{i\mathbf{q}(\mathbf{r}_l)} \right\rangle, \quad (2.54)$$

- 15 • ou potenciais idênticos,

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = \sum_{j=1}^n \langle |b_j(\mathbf{q})|^2 \rangle = \langle n \rangle \langle |b(\mathbf{q})|^2 \rangle \quad (2.55)$$

Para o caso de potenciais espalhadores idênticos, equação (2.55), a intensidade espalhada é a soma das intensidades de cada espalhador e os fatores exponenciais não contribuem para a intensidade total porque estão distribuídos aleatoriamente. Uma forma mais breve de escrever equação (2.55) é com o uso do conceito de fator de forma,
20 $P(q)$, equação (2.56). O fator de forma descreve a diferença entre a distribuição de

densidade eletrônica das partículas e o meio, e contribui para descrever a forma do potencial eletrônico dos espalhadores [50].

$$P(q) = \frac{\langle |b(\mathbf{q})|^2 \rangle}{\langle |b(0)|^2 \rangle} \quad (2.56)$$

Neste caso $P(q) \rightarrow 1$ quando $q \rightarrow 0$ [50].

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n \langle |b(0)|^2 \rangle P(q) \quad (2.57)$$

Se os potenciais espalhadores apresentam simetria esférica, $\langle |b_j(\mathbf{q})|^2 \rangle = \langle b_j(\mathbf{q}) \rangle^2 = b^2(q)$, eles podem ser considerados como constantes na região à baixos ângulos [47], de modo que, realizando a integração de $\langle e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)} \rangle$ em todas as orientações ⁴,

$$\langle e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)} \rangle_{orientacoes} = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi e^{-iq\langle |\mathbf{r}_{jl}| \rangle \cos\theta} \sin\theta d\theta = \frac{\text{sen}(q|\mathbf{r}_{jl}|)}{q|\mathbf{r}_{jl}|} \quad (2.58)$$

é possível reescrever a equação (2.54) e obter a equação de Debye:

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n|b(q)|^2 + 2 \sum_{j \neq l}^n \sum_{l=1}^n |b(q)|^2 \frac{\text{sen}(qr_{jl})}{qr_{jl}}. \quad (2.59)$$

ou de forma simplificada:

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = \sum_{j=1}^n \sum_{l=1}^n |b(q)|^2 \frac{\text{sen}(qr_{jl})}{qr_{jl}} \quad (2.60)$$

No caso de um sistema concentrado, ou ainda, em um sistema orientado, as posições espaciais do sistema, no espaço real, estão correlacionadas. Desta forma, o cálculo da intensidade média ou a intensidade medida sobre todas as possíveis orientações do

⁴Para facilitar a notação $|\mathbf{r}_{jl}| = r$.

sistema, equação (2.54), deve levar em consideração que os potenciais espalhadores estão espacialmente correlacionados. Neste caso, é conveniente usar uma função de probabilidade (para n partículas incidentes e normalizada por V_0 ⁵) de encontrar o centro de um potencial espalhador, a uma distância r , dado um outro centro de potencial

5 espalhador, equação (2.61) [48, 50].

$$p_{jl} = \int_0^{V_0} \langle n \rangle \frac{g(r)}{V_0} dv_j = n \frac{4\pi}{V_0} \int_0^\infty g(r) r^2 dr, \quad (2.61)$$

Para calcular a intensidade espalhada sobre todas as possíveis orientações é necessário levar em conta a função de probabilidade, $g(r)$, de encontrar um segundo potencial espalhador dado o primeiro, e neste caso, a integração do segundo termo da equação (2.54) pode ser realizada de acordo com:

$$\begin{aligned} \left\langle \sum_{j \neq l}^n \sum_{l=1}^n e^{-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_l)} \right\rangle &= p_{jl} \frac{\text{sen}(q|\mathbf{r}_{jl}|)}{q|\mathbf{r}_{jl}|} \int_0^{V_0} \frac{dv_l}{V_0} \\ &= n \frac{4\pi}{V_0} \int_0^\infty g(r) \frac{\text{sen}(qr)}{qr} r^2 dr \int_0^{V_0} \frac{dv_l}{V_0} \end{aligned} \quad (2.62)$$

10 de modo que:

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n \langle |b(\mathbf{q})|^2 \rangle + \langle |b(\mathbf{q})|^2 \rangle \frac{4\pi}{V_0} \int_0^\infty [1 - (1 - g(r))] \frac{\text{sen}(qr)}{qr} r^2 dr \quad (2.63)$$

Assumindo que $g(r) = 1 - (1 - g(r))$, a equação (2.63) pode ser reescrita, levando em conta a probabilidade $g(r)$:

⁵Volume da amostra iluminada pelo feixe.

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n \langle |b(\mathbf{q})|^2 \rangle + n^2 \langle |b(\mathbf{q})|^2 \rangle \frac{4\pi}{V_0} \int_0^\infty \frac{\text{sen}(qr)}{qr} r^2 dr - n^2 \langle |b(\mathbf{q})|^2 \rangle \frac{4\pi}{V_0} \int_0^\infty [1 - g(r)] \frac{\text{sen}(qr)}{qr} r^2 dr \quad (2.64)$$

Do ponto de vista experimental, o segundo termo da equação (2.64) pode ser desconsiderado, no caso em que, o volume do feixe incidente é maior que inverso do volume do potencial espalhador $V_0 > \frac{1}{r_1 r_2 r_3}$. Ou seja, o segundo termo pode ser desconsiderado no caso em que o modulo do vetor de espalhamento q , em uma direção, é maior que o
 5 inverso de um eixos: $\frac{q}{2\pi} > \frac{1}{r}$ [48]. A equação (2.64) pode então ser novamente reescrita:

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n \langle |b(\mathbf{q})|^2 \rangle \left[1 - n \frac{4\pi}{V_0} \int_0^\infty (1 - g(r)) \frac{\text{sen}(qr)}{qr} r^2 dr \right] \quad (2.65)$$

Na equação (2.65) o termo entre colchetes pode ser chamado de fator de estrutura, $S(q)$, de acordo com :

$$S(q) = 1 - n \frac{4\pi}{V_0} \int_0^\infty (1 - g(r)) \frac{\text{sen}(qr)}{qr} r^2 dr, \quad (2.66)$$

de modo que, a intensidade média de sistemas concentrados pode ser entendida como uma composição entre o fator de forma $P(q)$ e fator de estrutura $S(q)$,

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = n \langle |b(0)|^2 \rangle P(q) S(q). \quad (2.67)$$

10 No caso de espalhadores idênticos e simétricos, em que $\Delta\rho$ é número de elétrons em excesso, ou em outras palavras, a diferença de densidade eletrônica entre os potenciais espalhadores do soluto e solvente, e $\langle |b(0)|^2 \rangle = \Delta\rho^2 V^2$, a equação de intensidade média

normalizada será:

$$\frac{\langle I(q) \rangle}{I_e} = n\Delta\rho^2 V^2 P(q) S(q). \quad (2.68)$$

Métodos Analíticos e Semi-Analíticos

Um vez que o experimento de SAS foi realizado, é possível investigar os dados de intensidade espalhada por meio da equação (2.67) e uso de métodos analíticos, mas se o sistema espalhador apresenta fator de estrutura, $S(q)$, (a posição dos espalhadores está correlacionada) os casos em que a intensidade pode ser calculada analiticamente são bem mais restritos [56]. A maioria dos casos em que é possível realizar o cálculo do fator de estrutura, analiticamente ou semi-analiticamente, são aqueles em que os espalhadores têm simetria esférica e interagem com um potencial também de simetria esférica [56], o que torna esse método restrito a poucas geometrias e potenciais de interação. Por outro lado, se os potenciais espalhadores estão aleatoriamente orientados sem qualquer correlação espacial, a intensidade total deve ser calculada sobre todas as possíveis orientações, de acordo com a equação (2.53 e 2.57). Neste caso, o fator de forma será obtido pela integração do comprimento de espalhamento sobre todas as possíveis orientações. Assim, diferentes geometrias de espalhadores apresentam diferentes simetrias e consequentemente, diferentes formas de realizar a integral sobre todas as orientações. Abaixo, estão descritos os fatores de forma para as geometrias usadas neste trabalho.

Nas equações seguintes $b(\mathbf{q})$ representa a amplitude do fator de forma, ou o comprimento de espalhamento, calculado analiticamente; e o fator de forma é dado por $P(q) = \langle b^2(q) \rangle_{\text{orientações}}$. O raio da geometria é R e o volume é representado por V . As equações foram extraídas de Pedersen [56].

- Esfera:

$$b_e(q, R) = \frac{3 [\text{sen}(qR) - qR \cos(qR)]}{(qR)^3}. \quad (2.69)$$

- Casca esférica:

$$b_{ce}(q) = \frac{V(R_{out})b_e(q, R_{out}) - V(R_{in})b_e(q, R_{in})}{V(R_{out}) - V(R_{in})}, \quad (2.70)$$

em que:

$$V(R) = \frac{4\pi R^3}{3}. \quad (2.71)$$

- Cascas esféricas concêntricas:

$$b_{cec}(q) = \frac{1}{M} \left[\rho_1 V(R_1) b_e(q, R_1) + \sum_{i=2}^N (\rho_i - \rho_{i-1}) V(R_i) b_e(q, R_i) \right], \quad (2.72)$$

5 onde: ρ é a densidade de espalhamento e

$$M = \rho_1 V(R_1) + \sum_{i=2}^N (\rho_i - \rho_{i-1}) V(R_i). \quad (2.73)$$

- Elipsoide de Revolução:

$$b_{eli}(q) = b_e(q, R(R_1, R_2, R_3, \alpha, \beta)) \quad (2.74)$$

em que $R(R_1, R_2, R_3, \alpha, \beta) = [(R_1^2 \sin^2 \beta + R_2^2 \cos^2 \beta) \sin^2 \alpha + R_3^2 \cos^2 \alpha]^{1/2}$

$$P_{eli}(q) = [b_{eli}(q, R(R_1, R_2, R_3, \alpha, \beta))]^2 \quad (2.75)$$

- Cilindro:

$$b_c(q) = \frac{2J_1(qR\sin\alpha)}{qR\sin\alpha} \frac{\sin(qL\cos\alpha/2)}{qL\cos\alpha/2}, \quad (2.76)$$

em que J_1 é a função de Bessel de primeira ordem e L é o comprimento do cilindro.

$$P(q) = \int_0^{\pi/2} [b_c(q)]^2 \sin\alpha \, d\alpha. \quad (2.77)$$

- Cubo ou paralelepípedo retangular:

$$b_p(q) = \frac{\sin(qa \sin\alpha \cos\beta)}{qa \sin\alpha \cos\beta} \frac{\sin(qb \sin\alpha \sin\beta)}{qb \sin\alpha \sin\beta} \frac{\sin(qc \cos\alpha)}{qc \cos\alpha}, \quad (2.78)$$

em que a, b, c representam o tamanho dos eixos.

$$P(q) = \frac{2}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} [b_p(q)]^2 \sin\alpha \, d\alpha \, d\beta. \quad (2.79)$$

5 Método de Elementos Finitos

O item 2.2 mostrou algumas possibilidades de aplicações dos métodos analíticos e semi-analíticos para o cálculo do fator de forma e do fator de estrutura, porém estes cálculos podem ser aplicados em um conjunto muito restrito de sistemas. Em outras palavras, pode-se afirmar que o cálculo analítico do fator de forma é possível, somente para formas ou geometrias simples. Por outro lado, o método de elementos finitos (MEF) permite a construção de modelos com formas ou geometrias complexas, além calcular rapidamente a respectiva intensidade de espalhamento. Desta forma, é possível modelar uma forma ou geometria complexa definida por um conjunto pequeno de parâmetros e a respectiva intensidade espalhada pode ser computada ou ajustada à dados experimentais. Em

diversas aplicações [21] o MEF é usado para o cálculo da intensidade, e para estes casos a forma ou geometria do modelo é obtida por meio do uso de algoritmos heurísticos, dentre estes o de Monte Carlo é comumente usado.

Uma das grandes utilidades do MEF é quando são conhecidas indicações sobre a forma da partícula e neste caso para a aplicação do MEF é necessário conhecimento da amplitude de espalhamento. Assim, durante todo este trabalho, o termo MEF indicará a construção de modelos com forma conhecida. Em outras palavras, é possível calcular a intensidade espalhada a partir de um modelo de espalhador construído com MEF.

Uma das formas de modelar um espalhador é representá-lo como um conjunto de subunidades de amplitude de espalhamento conhecida, como por exemplo uma esfera. Desta maneira, um modelo de cilindro poderia ser representado como um conjunto de subunidades esféricas, distribuídas ao longo do volume cilíndrico. Se o modelo possui uma distribuição aleatória de subunidades, usa-se o método de Monte Carlo (MMC) para posicioná-las. Mas se a distribuição de posições das subunidades é correlacionada, esta tese denominará este modelo de tipo *coarse grain*. Para o caso de modelos do tipo *coarse grain* é usual substituir um aminoácido, ou o conjunto de átomos deste aminoácido, por um grão. Desta forma, para posicionar o grão é também usual utilizar ou a posição do Potácio (P) ou de um Carbono (C), de acordo com a notação do *Research Collaboratory for Structural Bioinformatics, Protein Data Bank* (RCSB PDB, www.rcsb.org) [3].

Em resumo, o MEF é um método para modelagem de espalhadores, que propõe modelar um espalhador a partir de elementos finitos ou subunidades conhecidas, de modo que, as subunidades devem ser menores que o menor detalhe estrutural em interesse [57]. A partir do modelo do potencial espalhador, é possível calcular a intensidade espalhada com uso das equações (2.51 e 2.53). Essas duas equações permitem calcular a intensidade dos espalhadores, por que nestes modelos, a forma das subunidades e as respectivas po-

sições são conhecidas, o que permite realizar a simulação de dados de espalhamento. A equação (2.53) calcula a intensidade média sobre todas as orientações dos espalhadores e a equação (2.51) calcula a intensidade espalhada em uma determinada direção, relacionada a orientação do espalhador.

5 Uma das grandes vantagens, da utilização do método de elementos finitos, é que ele permite modelar corpos geométricos simples e complexos por meio da construção dos mesmos com pequenas subunidades. Essa vantagem também carrega uma desvantagem intrínseca, já que as equações (2.51 e 2.53) possuem dupla somatória que tem como limite o número de subunidades (n), de modo que, a complexidade computacional é da
10 ordem de $O(n^2)$. Em muitos destes modelos, quando o número de subunidades é muito alto, o custo computacional pode ser inviável. Assim, uma alternativa, é otimizar esses cálculos utilizando a adaptação, proposta por Glatter [58, 59], Bordas e Pantos [52].

A otimização do cálculo da intensidade, para modelos de potenciais espalhadores discretos e com subunidades iguais, propõe substituir a dupla somatória (kl), por uma
15 somatória simples baseada em um histograma, $h(r_k)$. Para o caso da equação (2.53), a substituição acontece nas distâncias e conseqüentemente as somas duplas de distâncias tornam-se um histograma de distâncias de $nbins$, equação (2.80). Quase da mesma forma, na equação (2.51), a dupla soma de vetores de distâncias, tornam-se um histograma de projeções de distâncias de $nbins$, tomados sempre entre os ângulos -90° e
20 90° , de acordo com a equação (2.81).

$$\frac{\langle I(\mathbf{q}) \rangle}{I_e} = \sum_{k=1}^n \langle I_k(q) \rangle + 2b_{subunidade}^2 \sum_{k=1}^{nbins} h(r_k) \frac{\text{sen}(qr_k)}{qr_k} \quad (2.80)$$

$$\frac{I(\mathbf{q})}{I_e} = \sum_{k=1}^n I_k(\mathbf{q}) + 2b_{subunidade}^2 \sum_{k=1}^{nbins} h_p(\mathbf{qr}_k) \cos(\mathbf{qr}_k) \quad (2.81)$$

Aqui, $h(r_k)$ é o peso do histograma de distâncias, r_k , e $h_p(\mathbf{qr}_k)$ é o peso do histograma

de projeções de distâncias, $\mathbf{r}_k$. Salienta-se que ambos histogramas são ponderados pelas densidades eletrônicas das subunidades, o que poderá gerar partes negativas, no histograma, quando os modelos em análise são compostos com densidades eletrônicas negativas e positivas. Para o conjunto deste trabalho, a expressão histograma de dis-

5 tâncias e histogramas de projeções de distâncias deve ser entendida, respectivamente como, histograma de pares de distâncias ponderadas pelas densidades eletrônicas e histogramas de pares de projeções de distâncias ponderadas pelas densidades eletrônicas.

As simulações com modelos que usam distribuição uniforme de subunidades, empregando o método de MC e MEF, podem, em muitos casos apresentar um número da

10 ordem de 10^6 subunidades. Esse número grande, torna as simulações de intensidades lentas (para a capacidade computacional de hoje), mesmo com uso das equações adaptadas (2.80 e 2.81). Para otimizar ainda mais as simulações com grandes quantidades de subunidades, Hansen [51] e Pedersen et al. [60] sugerem dividir a construção do histograma. Em um exemplo, Pedersen et al. [60] realizam a modelagem de cascas

15 esféricas, construídas com cerca de $40 \cdot 10^3$ subunidades, de forma que para calcular a curva de intensidade, dividem a construção do histograma, de pares de distâncias, em 10 histogramas de 4000 subunidades cada, da seguinte forma:

$$\frac{1}{2} \times N \times N \Rightarrow 10 \times \frac{1}{2} \times \frac{N}{10} \times \left(\frac{N}{10} - 1 \right). \quad (2.82)$$

Uma outra otimização da equação (2.53) pode ser realizada se o espalhador é centro simétrico [61] e neste caso a dupla somatória que relaciona as subunidades kl , pode

20 ser substituída por somas simples que relacionam cada subunidade com um centro específico (uma espécie de posição de origem) do modelo espalhador. Neste caso, o vetor distância entre subunidades, $\mathbf{r}_j - \mathbf{r}_l$, é substituído pelo vetor distância entre cada

subunidade e o centro do modelo $\mathbf{r}_1$, de modo que, a intensidade espalhada é dada por:

$$\frac{I(\mathbf{q})}{I_e} = \left[\sum_{l=1}^n b_l(\mathbf{q}) \sin(\mathbf{q}\mathbf{r}_1) \right]^2 + \left[\sum_{l=1}^n b_l(\mathbf{q}) \cos(\mathbf{q}\mathbf{r}_1) \right]^2 \quad (2.83)$$

Sistemas Polidispersos

No caso de sistemas nos quais as partículas espalhadoras são não interagentes e apresentam polidispersidade, as equações (2.51 e 2.53) devem levar em consideração a dispersão do sistema, que pode ser de forma, tamanho, densidade eletrônica, etc. Este trabalho optou por enfatizar nos casos de polidispersidade de tamanho, assumindo partículas de mesma forma, assim a variação de tamanho ocorre isotropicamente em todas as direções. Neste caso, a intensidade de radiação espalhada de um sistema de partículas espalhadoras, polidisperso em tamanho, pode ser avaliada por meio de três tipos: de número, volume e intensidade. Deste modo, a expressão para a intensidade de radiação espalhada (sistema polidisperso) é dada por [62, 63]:

$$I(q) = \Delta\rho^2 \int_0^\infty D(R)W(q, R)dR \quad (2.84)$$

em que $\int_0^\infty D(R)dR = 1.0$ é a distribuição, $V_{volume}(R)$ é o volume da partícula com raio R e $W(q, R)$ é

- $W(q, R) = P(q, R)V_{volume}^2(R)$, em que $D(R)$ é a função para distribuição em número; ou
- $W(q, R) = P(q, R)V_{volume}(R)$, em que $D(R)$ é a função para distribuição em volume; ou
- $W(q, R) = P(q, R)$, em que $D(R)$ é a função para distribuição em intensidade;

e $P(q, R)$ é o fator de forma do espalhador.

Quando a polidispersão de tamanho é afim, ou em outras palavras, se o tamanho do modelo polidisperso varia uniformemente em todas as direções do modelo, é possível aplicar a polidispersão diretamente no histograma de pares de distâncias com uso da equação (2.80), otimizando o custo computacional, quando a criação de modelos é
5 muito mais lenta. Esta estratégia é possível porque ao invés de realizar os diferentes modelos polidispersos e encontrar um histograma de pares de distâncias para cada modelo, figura 2.5-B, que depois será composto, pode-se aplicar diretamente a polidispersão no histograma, de acordo com a figura 2.5-C.

A figura 2.5, permite visualizar no painel A, que se a distribuição de tamanho de um
10 espalhador é do tipo gaussiana, então a distribuição de pares de distâncias de cada espalhador em suas diferentes dimensões seguirá a distribuição gaussiana, painel B. Desta forma, ao invés de realizar o histograma da distribuição de pares de distâncias para cada diferente tamanho de espalhador, basta aplicar a polidispersão gaussiana, diretamente no histograma da distribuição de pares de distâncias do espalhador, conforme painel C,
15 otimizando o tempo de cálculo. Esta aplicação é válida somente para polidispersão do tipo afim com baixos valores de dispersão.

Assim a seção de espalhamento é fechada com a equação (2.84), que permite calcular ou realizar a simulação da intensidade de espalhamento de um sistema de espalhadores, não interagentes e polidisperso.

20 2.3 Sistemas Auto-Organizados

Sistemas auto-organizados⁶ podem ser entendidos como uma organização autônoma de componentes em padrões ou estruturas sem intervenção humana [64]. O processo de criação (auto-organização), em várias escalas e tipos de interações, pode acontecer naturalmente por meio da “própria natureza” ou ainda com uso de tecnologia. O estudo

⁶O termo auto-organizados(as), aplicado a sistemas moleculares, é empregado durante em todo este trabalho como tradução do termo *self-assembly*, do inglês.

Figura 2.5: Desenho esquemático do processo de aplicação de polidispersidade diretamente no histograma de pares de distâncias. A) Função de polidispersão, do tipo gaussiana, com valor médio $\langle x \rangle$. B) Distribuição de pares de distâncias, de um cilindro do tipo bastão, em diferentes valores da curva de dispersão de tamanho. C) Distribuição de pares de distâncias resultante da dispersão de tamanho de um conjunto de espalhadores.

desses sistemas é motivado por uma série de fatores, entre eles destacam-se [64]:

- células de sistemas vivos são auto-organizadas;

- constitui uma das poucas estratégias práticas para construção de *ensembles* de nanoestruturas, essenciais para estudos nanotecnológicos;
 - suas aplicações abrangem desde a manufatura até a robótica;
 - são componentes dinâmicos e podem formar materiais que podem ser usados como
- 5 sensores, ou ainda, em rede de computadores.

Atualmente os sistemas auto-organizados, são empregados em muitas áreas da ciência. Químicos, e bioquímicos tendem a resolver os problemas de *design*, síntese e em alguns fazendo modificações, misturando partes preexistentes. Engenheiros e físicos acabam assumindo o papel de observadores dessas estruturas enquanto, bioinformáti-

10 cos e cientistas computacionais, geralmente realizam a modelagem e simulações desses sistemas [64].

Whitesides e Grzybowski predizem algumas áreas, nas quais, a estratégia de síntese e fabricação desses sistemas poderão ser aplicadas [64]:

1. Cristalização, em todas as escalas, o que permitiria desenvolver novos meios para
- 15 criação de dispositivos microeletrônicos.
2. Robótica e manufatura. Seguindo a tendência de miniaturização, podem servir na micro ou nano fabricação, tanto como componentes quanto como métodos.
 3. Nanociência e tecnologia. Fabricação de nano sistemas e o desenvolvimento de novos métodos de fabricação.
- 20 4. Microeletrônica. Fabricação de dispositivos.
5. Sistemas reticulados.
 6. Estruturas auto-organizadas de DNA [64] em que é possível programar a montagem de DNA compondo inúmeras geometrias.

Com o objetivo de decifrar as propriedades que conferem estabilidade aos sistemas moleculares auto-organizados compostos por DNA, por meio da técnica de DM, e com o objetivo de desenvolver aplicações para esses sistemas com uso de experimentos de SAS, as próximas seções apresentarão um breve panorama geral deste tipo de sistema.

5 2.3.1 Sistemas Auto-Organizados de DNA

O DNA é uma biomolécula que carrega a informação genética disposta em um conjunto muito grande de possibilidades de sequências entre seus pares de bases. Esses pares podem ser formados por quatro tipos de nucleotídeos e suas combinações formam uma estrutura de dupla hélice que são antiparalelas, de cerca de 2 nm de diâmetro com um
10 passo de repetição de 3.5 nm [65]. A área de biologia molecular claramente explora as possíveis interações entre os pares de base do DNA na forma de replicação de informação genética e suas expressões. Além da biologia outras áreas começaram a explorar a complementaridade dos pares de base de DNA para controlar a estrutura da matéria, de forma que, agora é possível construir sistemas auto-organizados de DNA na escala
15 de nanômetros.

No início dos anos 90 o cristalógrafo Nadrian Seeman publicou as primeiras nanoestruturas auto-organizadas de DNA com conexões de um cubo e um octaedro [66]. Entre a década de 90 e primeira década deste século, o número dos trabalhos publicados com esses sistemas cresceu de forma quase exponencial [67], inaugurando uma nova área nas
20 ciências aplicadas, apelidada pelo próprio Seeman de: DNA *nanotechnology* [65,68].

Esse campo de estudos cresceu enormemente, principalmente a partir de 2001, ano no qual uma série de avanços foram alcançados [68], como o desenvolvimento da técnica de origamis de DNA [68] que contribuiu para o aumento na produção e sofisticação das estruturas bidimensionais (2D) e tridimensionais (3D) [67]. Entre esses avanços,
25 pode-se destacar, os sistemas auto-organizados, cristalinos, bidimensionais e tridimensionais e a realização de estruturas algorítmicas bidimensionais. A partir desse de-

envolvimento um número vasto de aplicações em nanoescala pode ser desenvolvida, entre elas destacam-se: os dispositivos nanomecânicos, que foram incorporados em estruturas origami e em matrizes de DNA 2D, as nanoestruturas que têm servido como armações, aprisionando outras moléculas, e as nanoestruturas usadas para a replicação
5 simples [68].

As principais estruturas, tridimensionais, montadas, após às idealizadas por Seeman na década de 90 (cubo [66] e octaedro truncado [69]), foram: octaedro [23, 70, 71], tetraedro [72–75], bipirâmide trigonal [76], dodecaedro [73], icosaedro [77, 78] e icosaedro truncado [73]. Similares as estruturas anteriores, outras estruturas auto-organizadas
10 foram elaboradas a partir de componentes híbridos, usando ligações sintéticas nos vértices [79–81], grandes estruturas [22, 82–85] baseadas no desenvolvimento da técnica de origamis de DNA [86] e ainda estruturas capazes de encapsular proteínas [87] ou nanopartículas [76, 80].

No atual desenvolvimento da área de DNA *nanotechnology* [67] é possível encontrar
15 trabalhos sobre containers, “Legos” de DNA em 2D e 3D e dispositivos “inteligentes”, com altas curvaturas e torções [88, 89].

Além das inúmeras formas ou geometrias construídas, realizadas por algumas dezenas de grupos em todo globo [65], Goodman e colaboradores, sintetizaram em 2008, um sistema auto-organizado de DNA na forma de tetraedro, que pode apresentar “alterações” que modificam parâmetros da geometria inicial do sistema, o que faz com ele
20 possa estar no estado aberto ou fechado [90]. Nesses sistemas auto-organizados, o DNA pode ser usado para ligar outras estruturas, tornando possível a criação de sistemas que formam diferentes geometrias.

Quase todo esse desenvolvimento recente, na elaboração das nanoestruturas tridimen-
25 sionais de sistemas auto-organizados de DNA [65] foi baseado em estruturas de alta simetria, concentradas principalmente em formas poliédricas convexas. Em linhas ge-

rais, foi possível elaborar nanoestruturas poliédricas preenchidas com DNA nas arestas e nas faces [65] e a maioria dos trabalhos, referentes a essas estruturas tridimensionais convexas, baseiam-se em geometrias poliédricas, com exceção dos recentes trabalhos de Dongran et al. [88, 89].

5 Do ponto de vista da síntese experimental destes sistemas, existem três grandes propriedades [65]:

i hibridização;

ii estabilidade na ramificação de DNA;

iii síntese do *design* de sequências.

10 Neste sentido, para realizar um projeto na área de nanotecnologia de sistemas auto-organizados de DNA existem, basicamente, dois passos necessários.

i Realização de padrões de *design* (*Motif Design*). O padrão de *design* é uma operação de troca recíproca que calcula as conectividades entre as ligações das duplas hélices de DNA com fins de obter uma nova conectividade entre as hélices.

15 ii *Design* de sequência (*sequence design*). O *design* de sequência é o método responsável pela otimização da síntese das sequência das fitas de DNA [65].

Já do ponto de vista da caracterização, as geometrias destes sistemas podem ser desvendadas por meio da verificação das topologias das ligações e por meio de métodos de observação direta ou indireta, com as técnicas microscópicas e de espalhamento de radiação. Os primeiros sistemas auto-organizados de DNA em 3D, foram melhor des-
critos como espécies topológicas, já que, suas estruturas não eram robustas o suficiente, podendo ser muito flexíveis, de modo que, durante a primeira década deste século um número, relativamente grande, de sistemas auto-organizados de DNA pode ser caracterizado, por meio de eletroforese (PAGE) e somente recentemente por microscopia
25 eletrônica [65].

As topologias das ligações e ramificações das fitas de DNA podem ser bem definidas pelas características obtidas por meio de experiências de desnaturação em gel de eletroforese, PAGE (*Polyacrylamida Gel Electroforese*) [65]. Bastante popular nas caracterizações em bioquímica, essa técnica pode fornecer informações indiretas sobre a topologia das ligações, das estruturas presentes nesses sistemas, ou ainda informações do peso molecular das mesmas. Essas informações podem ser obtidas baseadas na distância percorrida por uma amostra em um gel, ambos mergulhados em uma solução e sofrendo a atuação de uma diferença de potencial. Nestes casos, se as estruturas foram montadas ou estão auto-organizadas, devido a ligações químicas que modificam uma topologia inicial, elas percorrem diferentes distâncias em relação aquelas não montadas ou que não estão auto-organizadas [65] pois permitem a confirmação da montagem da estrutura.

No entanto, as técnicas microscópicas têm apresentado importantes contribuições quando aplicadas à caracterização de sistemas auto-organizados de DNA, mas essas técnicas apresentam algumas limitações, já que as amostras podem ser visualizadas somente depois de sofrerem um determinado processo de preparação. Dentre estes processos, as mais conhecidas para o caso de microscopia eletrônica (EM) estão as preparações de: congelamento por mergulho⁷, crio-microscopia eletrônica (Cryo-EM), coloração negativa, crio-coloração negativa e para obter as imagens de nanoestruturas alguns dos métodos utilizados incluem cristalografia eletrônica, tomografia eletrônica e EM de partícula-unitária [9].

Uma outra técnica de microscopia usada para caracterizar a geometria de sistemas na escala micro e nanométrica, é a técnica de AFM. Essa técnica permite que estruturas bidimensionais e tridimensionais possam ser observadas individualmente com precisão de nanômetros [22, 86]. Contudo, para o caso da visualização de estruturas tridimensio-

⁷essa técnica permite que a amostra possa ser observada em condições próximas ao seu estado *in natura*

nais, complicações podem acontecer quando a ponteira que percorre a amostra interage com ela e a distorce. Um exemplo desse problema, pode ser visto no trabalho de Andersen e colaboradores, que observaram que a geometria de uma caixa cúbica de DNA sofria deformações com a interação da ponteira, responsável pela medida [22]. Embora as técnicas microscópicas tenham apresentado grande utilidade para desvendar as estruturas tridimensionais, principalmente porque permitem a visualização direta e no espaço real de sistemas auto-organizados de DNA, elas também apresentam algumas desvantagens e entre elas destacam-se:

- complicações nas preparações das amostras;
- amostras precisam “resistir” ao alto vácuo (EM);
- a observação pode alterar a amostra;
- a reconstrução da estrutura 3D requer dezenas ou até centenas de imagens das projeções (EM);
- as amostras não podem ser estudadas em solução, no estado nativo, o que dificulta obter informações sobre as interações do sistema. Uma alternativa, é uso de Cryo-EM, mas neste caso, é necessário o congelamento da amostra;
- difícil de realizar modificações termodinâmicas na amostra.
- quanto maior a resolução do aparelho menor será a área visualizada na amostra, e conseqüentemente, menor será a quantidade de estruturas visualizadas no conjunto geral.

Cristalografia de raios X [65], técnicas de fluorescência [90] e técnicas de espalhamento de radiação (DLS [73] e SAXS [22]) também têm sido usadas, mas em um número inferior quando comparado as técnicas microscópicas.

Por último, as nanoestruturas de sistemas auto-organizados de DNA ainda apresentam

custos muito elevados para realização da síntese, de modo que, simulações de Dinâmica Molecular (DM) oferecem uma alternativa de poder prever a estabilidade de um determinado sistema, evitando desperdícios na síntese e nos posteriores experimentos de caracterização. Em muitos casos, somente a técnica de DM consegue descrever com
5 detalhes, as interações atomísticas que conferem estabilidade a esses sistemas. Desta forma, a técnica de DM têm sido empregada para descrever as propriedades estruturais que conferem estabilidade a essas nanoestruturas [24], mas é importante destacar que, na maioria dos casos as simulações de DM, sempre partem de restrições ou parâmetros que foram previamente obtidos por dados experimentais.

10 ***Nanocages* de DNA: Aplicações**

O desenvolvimento das várias geometrias de *nanocages* de DNA fez surgir o interesse para possíveis aplicações dessas estruturas nas áreas de biologia, ciências e engenharia de materiais. Dentre essas aplicações, destacam-se: i) a construção de sistemas híbridos de ácidos nucleicos e metais que podem ser usados para redox, fotoquímica, controle
15 catalítico e magnético; e ainda ii) encapsulação e transporte de partículas/moléculas, realizadas no interior das *nanocages*.

Em uma destas aplicações Yang e colaboradores [80] construíram *nanocages* de DNA-metal que incorporaram metais de transição, nas faces triangulares de uma bipirâmide de DNA com cerca de $25\text{-}30\text{ nm}^3$ de volume interno. Neste caso, a caracterização da
20 montagem das estruturas foi realizada com uso de PAGE, técnicas de microscopia e de fluorescência. Essa nova habilidade de construção de estruturas híbridas poderá contribuir para novas aplicações como encapsulamento, sensores, modificação e liberação de biomoléculas e nanomateriais.

No campo de encapsulamento de moléculas por meio de estruturas auto-organizadas,
25 Bathia e colaboradores [77], desenvolveram um icosaedro de DNA, de cerca de 20 nm de raio, no qual encapsularam nanopartículas. Nesta aplicação, os autores observaram

a montagem das estruturas por meio de PAGE e EM. Em uma outra aplicação de encapsulamento, Erben e colaboradores encaixaram uma proteína, citocromo *c*, em uma estrutura tetraédrica de DNA que podia acomodar uma esfera de cerca 2.6 nm, com peso molecular de 60 kDa. Nesta aplicação, a proteína podia variar sua posição
5 ao longo do eixo do tetraedro [87].

As estruturas auto-organizadas de DNA, na forma de poliedros convexos, podem ser construídas com DNA nas arestas e com DNAs nas faces dos poliedros. Ke e colaboradores construíram uma dessas estruturas, um tetraedro com DNAs nas faces. Cada eixo, das faces triangulares do tetraedro, apresentava dimensão de cerca de 54 nm, com-
10 pondo um volume (de uma pirâmide triangular) interno e externo estimado em cerca de $1.8 \cdot 10^{-23} m^3$ e $1.5 \cdot 10^{-23} m^3$, respectivamente. Em um outro trabalho, Andersen e colaboradores construíram uma caixa cúbica de DNA de dimensões de $42 \times 36 \times 36 nm^3$ que pode ser aberta externamente por "chaves" de DNA [22].

Por último, em uma das aplicações mais interessantes, Walsh e colaboradores [91]
15 reportaram a entrada de tetraedros de DNAs no interior de células de mamíferos e segundo os autores, os tetraedros permaneceram intactos até cerca de 48h após a transferência. A observação da entrada das *nanocages*, no interior das células, foi realizada com técnicas de microscopia de fluorescência, fornecendo o primeiro passo para o desenvolvimento da engenharia de sistemas auto-organizados de DNA, com foco na entrega
20 e controle da atividade de cargas dentro de células [91].

2.4 Dinâmica Molecular

A técnica de Dinâmica Molecular (DM) é empregada, neste trabalho, para investigar as propriedades que conferem estabilidade a estruturas auto-organizadas de DNA na forma de octaedro truncado. Assim a presente seção discute alguns conceitos muito
25 básicos e genéricos empregados pela técnica.

Pode-se entender a DM como um estudo de forças moleculares clássicas, envolvidas na descrição de sistemas e usada para estudar o movimento molecular. Basicamente, a DM permite avaliar sucessivas configurações geradas pela integração das leis de movimento de Newton [92], tornando possível acompanhar a trajetória de moléculas ou átomos. Essa trajetória pode ser observada num espaço tridimensional ao longo do tempo. A força que atua sobre cada átomo é dada pelas interações modeladas, de acordo com a forma e precisão do potencial clássico escolhido. Existem inúmeras maneiras de realizar a DM, e entre os elementos essenciais destacam-se [93]:

- a descrição do sistema (topologia das moléculas, ligações entre átomos, detalhes sobre as moléculas do solventes, as coordenadas espaciais, massa, carga, raios entre outros, etc.);
- a descrição do sistema de condições iniciais e periódicas (a temperatura, pressão etc) e do ensemble estatístico;
- a definição do Campo de Forças.

O Campo de força é um dos principais elementos na realização da DM porque descreve as forças das quais os constituintes do sistema estão sujeitos. Basicamente, os campos de força representam um conjunto de funções potencial e um conjunto de parâmetros numéricos. No caso das funções potencial ou dos potenciais é possível dividi-los basicamente em dois tipos: aqueles que contribuem com interações ligantes e aqueles que contribuem com interações não ligantes. Os mais básicos são descritos abaixo [93, 94].

Interações ligantes:

- Potencial de ligação, descreve a energia do movimento de vibração entre os pares i e j ligados por meio de uma ligação covalente de constante elástica k_{ij}^r ,

$$U_{ligacao} = \frac{1}{2} k_{ij}^r (r_{ij} - r_{eq})^2 \quad (2.85)$$

em que $r_{ij} = |r_i - r_j|$ é a distância entre os pares ligados i e j , e r_{eq} é a distância de equilíbrio.

- Potencial de ângulo de ligação, descreve a energia do movimento angular de vibração entre o trio de i , j e k ligados por meio de ligações covalentes,

$$U_{ang. ligacao} = \frac{1}{2} k_{ijk}^{\theta} (\theta_{ijk} - \theta_{eq})^2 \quad (2.86)$$

5 em que θ_{ijk} é o ângulo entre sucessivas ligações e k_{ijk}^{θ} é um ângulo constante.

- Potencial de ângulo de torção, descreve a energia do movimento entre dois planos, formados pelo quarteto i , j , k e l ligados por meio de ligações covalentes. O primeiro desses planos é formado entre o trio i , j e k ligados por meio de duas ligações covalentes e o segundo plano formado entre o trio j , k e l ligados por meio de duas ligações covalentes,

$$U_{ang. torcao} = \frac{1}{2} V_{ijkl}^{\phi} (1 + \cos(\phi_{ijkl} - \gamma)) \quad (2.87)$$

em que ϕ_{ijkl} é o ângulo de torção entre três sucessivas ligações, γ é uma fator de fase [93, 94] e V_{ijkl}^{ϕ} é um potencial de barreira.

Interações não ligantes:

- Potencial de Coulomb, é o potencial que modela a interação eletrostática entre duas cargas q_i e q_j ,

$$U_{Coulomb}(r_{ij}) = \frac{q_i q_j}{r_{ij}^2} \quad (2.88)$$

em que r_{ij} é a distância entre as cargas q_i e q_j .

- Lennard-Jones, é o potencial que modela a interação entre um par de átomos

ou moléculas neutras. O primeiro termo deste potencial, equação(2.89), descreve a atuação de uma uma força repulsiva (força ligada ao princípio de exclusão de Pauli) até o limite da menor distância e o segundo termo descreve a força atrativa (força de van der Waals) até o limite da maior distância,

$$U_{Lennard-Jones}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (2.89)$$

5 em que ϵ é a profundidade de campo e σ é a distância em que o potencial é zero [93, 94].

A análise da trajetória de uma simulação de DM molecular pode revelar inúmeros comportamentos dos constituintes envolvidos em um sistema, que em muitas vezes são muitos difíceis ou impossíveis de se obter por meio de uma técnica experimental. Neste
10 sentido, uma boa escolha das configurações das condições iniciais e dos campos de força envolvidos, são fatores essenciais, mas que podem ser obtidos pela análise de algumas das condições teóricas e experimentais previamente avaliadas.

CAPÍTULO 3

5 SAS, MODELAGEM E SIMULAÇÃO

Nos capítulos anteriores foi discutido o fenômeno de espalhamento junto a uma abordagem da teoria de espalhamento à baixos ângulos (SAS). Essa discussão tinha como objetivo abordar os conceitos que serão novamente discutidos neste capítulo para aplicação em sistemas de alta simetria. Como um dos principais objetivos deste trabalho é
10 contribuir na análise de dados de espalhamento, que neste caso traduz-se em propor novas abordagens para o ajuste de curvas simuladas (de modelos de alta simetria, obtidos com informações *a priori*) contra dados experimentais, o presente capítulo apresenta alguns dos resultados obtidos durante o período de doutorado do autor.

Como discutido no capítulo 2, um dos grandes desafios de utilizar as técnicas de espalhamento à baixos ângulos é trabalhar com dados do espaço recíproco. Neste sentido,
15 uma das estratégias possíveis, é partir de um modelo, do espaço real, para simular a curva de intensidade de espalhamento, do espaço recíproco. Essa curva pode então ser ajustada com a curva obtida experimentalmente, de modo a obter um modelo final,

ajustado aos dados experimentais.

Para mostrar os resultados obtidos dessa estratégia, esse capítulo será dividido em duas partes: modelagem e simulação. A parte de modelagem apresenta o método de elementos finitos (MEF) utilizado para modelar geometrias de altas simetrias, em diversas
5 aplicações. A segunda parte, de simulação, apresenta a utilização dos mesmos modelos para obter as curvas da intensidade espalhada e neste caso, diferentes abordagens são aplicadas dependendo do modelo ou do sistema em estudo. Antes de cada abordagem, testes de validação são realizados, de modo que, as curvas simuladas são apresentadas e comparadas com formulações obtidas analiticamente ou semi-analiticamente, para
10 verificação de confiabilidade.

Todo o desenvolvimento computacional, apresentado neste capítulo, foi implementado em um programa chamado *Polygen*, escrito em FortranTM [2] e está disponível por meio do contato com o autor (alves.cassio arroba gmail.com) ou orientador (crilspo arroba if.usp.br) desta tese.

15 3.1 Modelagem

Um número muito grande de estruturas auto-organizadas de DNA têm sido montadas, principalmente a partir de 2001, ano no qual uma série de avanços foram alcançados [68]. O desenvolvimento recente de nanoestruturas auto-organizadas de DNA [65] foi baseado em estruturas de alta simetria, concentrando em formas poliédricas¹. No
20 atual desenvolvimento da área de nanotecnologia de estruturas de DNA 3D [65], já é possível desenvolver nanoestruturas preenchidas com DNA nas arestas e também nas faces. Praticamente quase todos esses trabalhos, com exceção do trabalho de Dongran et al. [88], baseiam-se em poliedros convexos.

Entre as várias técnicas utilizadas para realizar a caracterização de sistemas molecula-

¹Para o caso de estruturas 3D.

res, as técnicas de espalhamento não são as mais utilizadas. Talvez um dos principais motivos que dificultem o uso, seja a dificuldade de análise e modelagem dos dados de espalhamento. Uma abordagem que unifica a modelagem ao cálculo da intensidade espalhada, já foi apresentada previamente por Zhou et al. [95], mas com algumas li-
mitações, entre elas estavam um pequeno e restrito conjunto de modelos, e formas de
modelagem, além de que, alguns destes modelos já possuem expressões analíticas ou
semi-analíticas para o cálculo da intensidade espalhada. Diante dessa lacuna científica,
o presente trabalho pretende contribuir para tornar acessível os métodos de análise de
dados de SAS para o conjunto de aplicações descritas nas seções anteriores. Para tanto,
sugere a modelagem de estruturas usando o método de elementos finitos (MEF) para
uso em simulações e ajuste à dados experimentais. Para uma melhor compreensão,
as subseções seguintes também farão um breve resumo de estruturas de alta simetria
baseadas em poliedros convexos e formas de modelagem usando MEF.

Poliedros Convexos - Breve Contexto

Historicamente esses corpos foram objetos de estudo e teoria dos filósofos da antiguidade
grega, chegando até mesmo a comporem partes de teorias sobre o universo. Os políedros
podem ser entendidos, brevemente, a partir do significado dado pelos gregos que diz que
poly significa muitos e *edro* significa face. Do ponto de vista formal e contemporâneo,
poliedros convexos são poliedros com um número finito n ($n \geq 4$) de polígonos, planos
convexos (ou regiões poligonais convexas) tais que [96]:

- duas faces/polígonos não estão num mesmo plano;
- uma aresta divide somente duas faces/polígonos;
- o semi-espaco das faces/polígonos é dividido pelo plano de cada face/polígono,
de modo que, ficam determinados n semi-espacos, cada um com origem em seu
plano e que contém os restantes;

Para relacionar o número de vértices v , o número de arestas a e o número de faces f , de um poliedro convexo é possível usar a relação de Euler [97]:

$$v + f = a + 2. \quad (3.1)$$

Os principais poliedros convexos podem ser basicamente divididos em regulares e não regulares. Os regulares são também conhecidos como sólidos platônicos e são caracterizados por apresentarem o mesmo número de arestas em todas as faces. Já o número de arestas que encontram um vértice é o mesmo para todos os vértices. Os poliedros regulares são cinco: tetraedro, cubo, octaedro, dodecaedro e icosaedro.

Os poliedros não regulares podem ser sólidos arquimedianos, sólidos de Catalan, prismas, antiprismas, e sólidos de Johnson (construção baseada em pirâmides, cúpulas, rotundas, sólidos platônicos e sólidos arquimedianos). Os sólidos arquimedianos são aqueles obtidos por operações geométricas sobre os sólidos platônicos e apresentam faces de polígonos regulares, de modo que, para cada vértice existe ao menos dois arranjos de polígonos semelhantes. No total são treze: tetraedro truncado (tetraedro triakis), cuboctaedro (dodecaedro rômico), cubo truncado (octaedro triakis), octaedro truncado (hexaedro tetrakis), rombicuboctaedro (icosaedro deitodal), cuboctaedro truncado (dodecaedro disdiakis), icosidodecaedro (triacontaedro rômico), dodecaedro truncado (icosaedro triakis), icosaedro truncado (dodecaedro pentakis), rombicoidodecaedro (hexecontaedro deltoidal), icosidodecaedro truncado (triacontaedro disdiakis), cubo snub (icositetraedro pentagonal), icosidodecaedro snub (hexecontaedro pentagonal).

Os sólidos duais de Arquimedes são conhecidos como sólidos de Catalan e apresentam faces poligonais irregulares que são congruentes entre si. Assim como os sólidos arquimedianos, são treze: tetraedro triakis, dodecaedro rômico, octaedro triakis, hexaedro tetrakis, icositetraedro deltoidal, dodecaedro disdiakis, triacontaedro rômico,

icosaedro triakis, dodecaedro pentakis, hexecontaedro deltoidal, triacontaedro disdiakis, icositetraedro pentagonal e hexecontaedro pentagona.

Os prismas, são poliedros de faces poligonais de n lados (uma face superior e uma face inferior que são paralelas e congruentes) formados com outras n faces que são paralelogramos e que encontram as arestas dos polígonos base. Seus nomes são baseados na
5 forma de suas bases poligonais. Antiprismas são similares aos prismas exceto pelo fato que suas bases são relativamente giradas, uma em relação a outra e suas faces laterais são triângulos. A nomenclatura dos antiprismas é derivada do nome dos polígonos bases.

10 Por último, os sólidos de Johnson apresentam faces poligonais regulares, com todas as arestas de comprimentos iguais. Ao todo, existem pelo menos 92 sólidos de Johnson, e por facilidades de nomenclatura são também conhecidos como J1, J2, ... e J92. No apêndice [A](#) é apresentada uma lista com os nomes de cada um dos sólidos de Johnson.

Construção de um Banco de Dados de Poliedros Convexos

15 Com o objetivo de modelar estruturas de alta simetria, baseadas em sistemas auto-organizados já publicados, o presente trabalho buscou na literatura um banco de dados de formas poliédricas que pudesse servir de base para a modelagem dessas nanoestruturas. Nesta busca, foi encontrado um banco de dados, relativamente completo, com cerca de 100 poliedros convexos [\[98\]](#). O banco de dados (BD) utilizado, apêndice [A](#),
20 reúne um conjunto de posições de vértices e indicações dos vértices de cada face. Ele foi usado para modelar diferentes estruturas a partir das coordenadas espaciais dos vértices com uso do algoritmo *winged edge* [\[97\]](#).

O uso do algoritmo *winged edge* [\[97\]](#), somado a estrutura de dados, com as posições dos vértices e suas respectivas arestas, permitiu iniciar a modelagem de estruturas de
25 alta simetria, figura [3.1](#). Inicialmente, a modelagem inicia-se com a determinação da

posição dos vértices, figura 3.1-A. Em seguida a união ou identificação do vértice à aresta, figura 3.1-B e com as informações de vértices, arestas e faces é possível realizar diferentes conjuntos de modelagens, figuras 3.1 - C à F.

Figura 3.1: *Desenho esquemático do processo de modelagem das geometrias. A) Posição dos vertices de um cubo. B) União dos vértices com segmentos de retas. C) Modelo do tipo sólido. D) União dos vértices com subunidades esféricas. E) União dos vértices com DNA. F) Modelo do tipo casca. Figura adaptada [2].*

Na figura 3.1 - E é possível ver um exemplo de modelagem de um cubo contendo arestas
 5 elaboradas a partir de um modelo *coarse grain* de DNA, que pode ser construído com a
 informação das coordenadas espaciais dos átomos de carbono C_α ou dos fosfatos do P .
 Neste tipo de modelo, as coordenadas espaciais dos átomos escolhidos são usadas para
 posicionar as subunidades do tipo *coarse grain*, que por sua vez, podem representar
 um conjunto de átomos². No caso do DNA, o nucleotídeo e o *backbone* correspondente,
 10 podem ser representados por somente uma subunidade, que pode estar centralizada na
 mesma coordenada espacial do C_α ou do P . Neste trabalho, a modelagem na qual, as
 coordenadas espaciais das subunidades são pré definidas e representam um conjunto de
 centro espalhadores, será denominada de modelagem do tipo *coarse grain*.

²Nesta tese, denomina-se modelagem do tipo *coarse grain*, aquelas construídas com baixa resolução em que os grãos ou subunidades possuem coordenadas espaciais correlacionadas

Além da modelagem *coarse grain* em que as coordenadas das subunidades são correlacionadas, uma outra forma de estrutura pode ser representada quando as coordenadas espaciais são distribuídas ao longo de um volume, como no caso de cascas e estruturas coloidais. Nestes últimos casos, é possível realizar a modelagem com subunidades po-
 5 sicionadas em coordenadas espaciais aleatórias e para tal, usa-se o método de Monte Carlo (MMC), que é usado para gerar um conjunto de coordenadas com distribuição uniforme.

A aplicação do MMC na geração de corpos, com geometrias poliédricas, inicia-se delimitando um volume no interior de um espaço de busca que pode ser preenchido inter-
 10 namente e/ou externamente pelas coordenadas espaciais das subunidades, figura 3.1-C e F. Assim, para construir o modelo com MMC, o *Polygen* cria um volume preenchido com subunidades, ou em outras palavras, cria um espaço de busca de subunidades. Os modelos pretendidos são construídos realizando-se recortes neste espaço de busca, que possui densidade uniforme de subunidades de acordo com:

$$d_{densidade} = \frac{n_{subunidades}}{V_{volume\ espaço\ de\ busca}}. \quad (3.2)$$

15 Para preencher o volume interno ou externo de um poliedro convexo é necessário obter informações espaciais sobre limites, internos e externos. Assim, um método vetorial para saber se uma determinada coordenada espacial está no interior ou exterior de um poliedro pode ser visualizado na figura 3.2. Neste método, a coordenada espacial p_0 é avaliada por meio de produtos vetoriais entre ela e todos os vetores normais as
 20 faces/polígonos do poliedro. Assumindo que $\vec{v}_3 = p_0 - p_{ref}$ é o vetor entre a coordenada avaliada p_0 , e um ponto escolhido na face do poliedro p_{ref} . Se o vetor normal desta face é dado por $\vec{n} = \vec{v}_1 \times \vec{v}_2$, onde v_1 e v_2 são vetores coplanares à face (orientados no sentido anti-horário das faces), então se, $\vec{v}_3 \times \vec{n} < 0$, para todos os vetores normais a todas as faces, p_0 está no interior deste poliedro.

Figura 3.2: Desenho esquemático do processo de modelagem das geometrias com uso das coordenadas espaciais e das referências aos planos. O vetor $\vec{n}$ é normal ao plano definido pelos vetores v_1 e v_2 . O vetor $\vec{v}_3$ conecta o ponto de referencia p_{ref} no plano ao ponto p_0 . A) Se $\vec{v}_3 \times \vec{n} = 0$, p_0 está no plano. Se $\vec{v}_3 \times \vec{n} < 0$, p_0 está no semi-espaco negativo do plano. Se $\vec{v}_3 \times \vec{n} > 0$, p_0 está no semi-espaco positivo do plano. B) p_0 está no interior do tetraedro quando $\vec{v}_3 \times \vec{n} < 0$ para todos os vetores $\vec{n} = \vec{v}_1 \times \vec{v}_2$, desenhados com v_1 e v_2 orientados no sentido anti-horário das faces. Figura adaptada [2].

Se a geometria pretendida não corresponder a nenhum poliedro disponível no banco de dados é ainda possível, realizar a modelagem de novos poliedros convexos, a partir de coordenadas espaciais distribuídas em uma superfície esférica. Neste caso, o algoritmo *Gift Wrapping* é usado para criar as conectividades de faces triangulares e modelar um poliedro convexo a partir de pontos regulares ou aleatórios distribuídos em uma superfície esférica [97, 99, 100].

Esse conjunto de estratégias para realizar a modelagem permite definir de maneira bastante simples uma forma geométrica, como por exemplo um icosaedro. O icosaedro é uma forma geométrica com relativa complexidade e pode ser definido de várias formas, mas no caso deste trabalho, a modelagem necessita de um número muito baixo de parâmetros como o comprimento o raio de inscrição dos vértices e a forma de preenchimento, como por exemplo a do tipo sólido. Para este caso, o icosaedro sólido é modelado com somente dois parâmetros, e o mesmo vale para qualquer outra geometria disponível no programa *Polygen* que exporta ao final da modelagem um modelo no formato pdb, *Research Collaboratory for Structural Bioinformatics, Protein Data Bank* (RCSB PDB, www.rcsb.org).

Uma vez realizado a modelagem do sistema, é possível realizar a simulação da curva de intensidade de espalhamento, como poderá ser visto na próxima seção.

3.2 Simulação

A partir das informações prévias de um sistema espalhador é possível realizar um modelo
5 no espaço real e simular a intensidade espalhada. Nestes casos, os modelos podem ser compostos com subunidades ou coordenadas espaciais dos centros espalhadores, em posições pré-definidas (modelos *coarse grain* e atomísticos, vide item 2.2), figura 3.1 - D e E, ou ainda distribuídas aleatoriamente (MMC)³, figuras 3.1 - C e F. Assim, a partir de um modelo ou uma composição de modelos, conhecidos previamente — como
10 por exemplo, aqueles obtidas por experiências de EM, ou ainda, pelo *design* prévio do sistema em estudo — é possível minimizar os parâmetros relativos ao processo de modelagem do sistemas, por meio do ajuste da curva de intensidade simulada contra a experimental. Esse processo, permite identificar a conformação mais provável dos sistemas, em suas diferentes formas geométricas, além de obter outras informações
15 importantes como distribuição de tamanho, densidade eletrônica, polidispersão, entre outros.

Para realizar esta estratégia, são apresentados, nos itens seguintes, alguns resultados para a simulação da curva de intensidade de SAXS de vários modelos em diferentes geometrias. Inicialmente, são apresentados modelos que podem ser considerados modelos
20 testes (esfera, cilindros, elipsóide de revolução e paralelepípedo) porque a intensidade pode ser calculada analiticamente ou semi-analiticamente. Em uma segunda parte, são apresentadas as simulações da curva de intensidade realizadas para modelos com

³Se o modelo possui uma distribuição aleatória de subunidades, usa-se o método de Monte Carlo (MMC) para posicioná-las. Mas se a distribuição de posições das subunidades é correlacionada, esta tese denominará este modelo de tipo *coarse grain*. Para o caso de modelos do tipo *coarse grain* é usual substituir um aminoácido, ou o conjunto de átomos deste aminoácido, por um grão. Desta forma, para posicionar o grão é também usual utilizar ou a posição do Potácio (P) ou de um Carbono (C), de acordo com a notação do *Research Collaboratory for Structural Bioinformatics, Protein Data Bank* (RCSB PDB, www.rcsb.org) [3].

geometrias complexas em várias aplicações. Desta forma, a apresentação desta seção divide-se em simulações da curva de intensidade de: sistemas aleatoriamente orientados e orientados, neste último caso a orientação é realizada em uma direção específica do vetor de espalhamento $\mathbf{q}$.

5 3.2.1 Intensidade Média - Sistemas Aleatoriamente Orientados

Neste subitem os sistemas estudados são de espalhadores sem orientação preferencial, ou seja, aleatoriamente orientados. Este tipo de sistema representa com muita precisão o caso de partículas em solução, de modo que, o feixe incidente atinge um número muito grande de espalhadores, aleatoriamente orientados. A expressão utilizada para simular
 10 a intensidade espalhada destes modelos é a equação (2.80) normalizada [52, 58, 59]. O cálculo usando esta equação, foi realizada com a aplicação da divisão do histograma de pares distâncias, segundo a relação de Pedersen et al. [60], equação (2.82). Neste item não são apresentados ajustes contra dados experimentais, mas o programa *Polygen* realizou ajustes de curvas simuladas contra curvas de sistemas que possuem equações
 15 analíticas ou semi-analíticas obtendo valores de chi-quadrado próximos de 1. Para implementar os ajustes de dados, foram aplicados os métodos de minimização baseados em *simulated annealing*, Levenberg-Marquardt e algorítmicos genéticos.

Todos os resultados das simulações, foram realizados a partir de modelos elaborados com o MEF do tipo *coarse-grain* (coordenadas espaciais correlacionadas) ou do tipo
 20 MMC (coordenadas espaciais distribuídas aleatoriamente em um volume).

Com o objetivo de verificar se o processo de modelagem e cálculo da intensidade média estava correto, foram realizados testes de verificação. Para realizar estes testes, foram escolhidos modelos geométricos simples que possuem equação analítica ou semi-analítica para o cálculo de intensidades. As formas geométricas escolhidas foram: a esfera e suas
 25 variantes, o cilindro e o cubo. A intensidade espalhada, dos modelos de referência, foram calculadas analiticamente ou semi-analiticamente com uso da equação (2.68), assumindo

sistemas não interagentes, $S(q) = 1$, e com o fatores de forma, $P(q)$, correspondentes a cada geometria escolhida, vide item 2.2 sobre fatores de forma.

Sistemas Monodispersos - Sistemas Testes

Esfera e suas variantes

Os modelos elaborados foram três: (vermelho) esfera com raio de $R = 90$ e densidade eletrônica de $\rho = 1.0$, (azul) casca esférica de raio interno de $R_{in} = 90$, raio externo de $R_{out} = 100$ e densidade eletrônica de $\rho = -1.0$ e (verde) núcleo-casca esférica ou *core-shell esférica* com raio interno de $R_{in} = 90$ e densidade eletrônica de $\rho_{in} = 1.0$, composto com uma casca esférica de raio interno $R_{in} = 90$, raio externo $R_{out} = 100$ e densidade eletrônica de $\rho_{out} = -1.0$ ⁴.

A comparação teste do modelo da esfera, foi realizada com a curva da intensidade simulada, com *Polygen*, contra a curva de intensidade analítica, obtida para o fator de forma esférico, equação (2.69), ambas podem ser vistas na figura 3.3. A comparação entre essas curvas permite ver a boa precisão do cálculo de intensidades a partir do modelo realizado. Nesta mesma figura, é possível visualizar também que o histograma de pares de distâncias, das subunidades do modelo, apresenta uma distribuição simétrica em torno do raio da esfera, o que era esperado.

A casca esférica, diferentemente da esfera, apresenta um histograma de pares de distâncias com uma distribuição assimétrica em torno do valor do raio, como esperado. Já o núcleo-casca esférica foi modelado com núcleo esférico de densidade eletrônica positiva e uma casca esférica, concêntrica, com densidade eletrônica negativa. Neste caso, o histograma de pares distâncias apresenta uma distribuição assimétrica no eixo das abscissas, devido a diferença entre os valores de densidade eletrônicas, como esperado,

⁴Para estes casos usa-se a densidade eletrônica ρ por tratar-se de partículas em vácuo, mas para um sistema em solução a densidade eletrônica seria a diferença entre as densidades do soluto em relação ao solvente, $\Delta\rho = \rho_{soluto} - \rho_{solvente}$.

figura 3.3.

Figura 3.3: (a) Histograma de pares de distâncias entre as subunidades. b) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen da esfera e suas variantes. Figura adaptada [2].

As curvas de intensidades simuladas para ambos os modelos foram comparadas com as curvas de intensidade analítica, compostas pelos respectivos fatores de formas de acordo com as equações (2.69, 2.70 e 2.72). Em todos os casos, as comparações, entre as intensidades simuladas com *Polygen* e aquelas calculadas analiticamente, apresentam compatibilidade com valores de intensidade muito próximas entre si, figura 3.3. Existem algumas diferenças para valores altos de q , devido a precisão limitada do método de Debye para altos ângulos. No entanto, dados experimentais reais contém a presença de diversos ruídos de fundo entre outros efeitos, e a diferença apresentada, em muitos dos casos, não é relevante. Os resultados com o uso do *Polygen* (usando o método de EF) ficam mais próximos do teórico a medida que a densidade de subunidades por volume é maior. Quanto maior é a densidade de subunidades do modelo, maior a precisão da curva de intensidade, uma vez que, o volume da estrutura modelada fica mais preenchido. Em outras palavras, quanto mais próximo o modelo utilizado está em relação ao sistema estudado, maior será a precisão da curva de intensidade. Por outro lado, o aumento da densidade de subunidades tende a deixar a simulação computacional mais lenta, o que em casos de minimização, pode não ser indicado. Nestes casos, a

divisão em blocos proposta por Pedersen et al., [60] pode contribuir para otimizar o tempo de processamento.

Cilindros e Cubo

O modelo de cilindro do tipo bastão, ou *rod*, foi contruído de forma que o raio $R = 30$ seja muito menor que a altura $L = 100$. Na figura 3.4, o histograma de pares distâncias do bastão apresenta uma distribuição assimétrica, apresentando uma concentração maior em valores em torno do raio do cilindro, o que era esperado. A curva de intensidade simulada do bastão, foi comparada com a curva de intensidade, que usa o fator de forma do cilindro, equação (2.77) e mostra que ambas apresentam valores muito próximos entre si.

Já o modelo de cilindro tipo prato ou, *plate*, é quase o contrário do modelo de cilindro tipo bastão. No caso do prato o raio $R = 100$ é relativamente maior que a altura do cilindro $L = 30$. Neste caso, o histograma de pares distâncias apresenta uma distribuição com concentração maior nos valores em torno do raio do cilindro, figura 3.4, como era esperado. Para este caso, a curva de intensidade simulada foi comparada com a curva de intensidade, que usa do fator de forma do cilindro dado pela equação (2.77), e novamente os dados de simulação e aqueles calculados são muito similares à baixos ângulos.

A curva de intensidade do cubo, figura 3.4, inscrito numa esfera de raio igual a 100, foi simulada e comparada com a curva de intensidade com uso do fator de forma do cubo, equação (2.79), e assim como os resultados anteriores, apresenta uma baixa diferença entre a intensidade simulada e a analiticamente calculada.

Assim como para o caso esferas, as comparações das geometrias cilíndricas e cúbicas, entre a intensidade simulada com *Polygen* e aquelas obtidas por meio do cálculo com as respectivas equações analíticas, são muito próximas. Também como observado para esferas, existem algumas diferenças para valores altos de q devido as limitações de

Figura 3.4: (a) Histograma de pares de distâncias entre as subunidades. b) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen para o cilindro do tipo Prato (plate), bastão (rod) e cubo (cube). Figura adaptada [2].

precisão do cálculo. A consideração, realizada no subitem anterior, sobre os valores de densidade de subunidades também vale para esses casos.

Em síntese, os resultados obtidos com *Polygen* indicam que a simulação obtida para a curva de intensidade é muito próxima dos valores teóricos esperados. Assim, os testes aplicados a sistemas conhecidos, contribuem para realizar a verificação das simulações e dos métodos de modelagem, além de contribuir para dar o suporte necessário para o cálculo da intensidade de espalhamento, de sistemas monodispersos de qualquer forma ou geometria.

Sistemas Polidispersos - Sistemas Testes

Muitos sistemas realizados experimentalmente na escala micro e nanométrica podem apresentar polidispersão. O estudo desses sistemas não interagentes, com uso de espalhamento de radiação à baixos ângulos é baseado na informação média da intensidade de radiação espalhada. Como a informação da radiação espalhada é influenciada pela forma, tamanho e principalmente pelo contraste óptico/eletrônico/nuclear do sistema em solução [62, 63] ela pode ser avaliada de por meio de três tipos de polidispersidade⁵:

⁵Para este caso, os espalhadores do sistema, apresentam a mesma forma ou geometria.

em número, volume e intensidade, segundo equação (2.84).

Com o objetivo de verificar a qualidade e precisão das simulações da curva de intensidade espalhada, obtidas pelo programa *Polygen*, aplicado a sistemas polidispersos, foram feitas comparações contra as curvas calculadas analiticamente e semi-analiticamente, 5 que contém a contribuição de polidispersão, de acordo com a equação (2.84). No caso das simulações, a polidispersidade dos sistemas foi aplicada ao histograma dos pares de distâncias das geometrias modeladas, e o cálculo da intensidade foi realizado com uso da equação (2.80), para detalhes vide subitem 2.2. Os sistemas testes escolhidos foram a esfera ($R = 100$), a casca esférica ($R_{in} = 90$ e $R_{out} = 100$), e os cilindros do tipo prato 10 ($R = 100$ e $L = 30$) e bastão ($R = 30$ e $L = 100$). A polidispersão utilizada foi do tipo gaussiana com variâncias de 5%, 10% e 20% e todas as polidispersões foram aplicadas em até 5 desvios padrões para ambos os lados do centro da distribuição. Os resultados obtidos para esses sistemas podem ser vistos na figura 3.5.

Na figura 3.5, é possível observar que os valores obtidos pela curva simulada, com *Poly-* 15 *gen*, contra as curvas analíticas são muito próximos e os vales característicos de sistemas monodispersos são menos pronunciados, já que a curva de intensidades é uma média ponderada do sistema. Da mesma forma que foi avaliado para sistemas monodispersos, os sistemas polidispersos apresentaram resultados compatíveis com o esperado para os sistemas padrões, de modo que, esse último caso de polidispersidade, finaliza os testes 20 sobre a precisão do programa desenvolvido. Esses resultados indicam que os resultados das simulações podem ser avaliados em outras geometrias, nas quais as equações analíticas não estão disponíveis tanto para os sistemas monodispersos como polidispersos.

Figura 3.5: (Esquerda) Histograma de pares de distâncias entre as subunidades, as geometrias apresentam polidispersão gaussiana em número com variâncias de 5%, 10% e 20% para: (A) esfera, (B) casca esférica, (C) cilindro do tipo bastão (rod) e (D) cilindro do tipo Prato (plate). (Direita) Comparação da curva de intensidade obtida com equação analítica e a obtida por simulação com Polygen para: (A) esfera, (B) casca esférica, (C) cilindro do tipo bastão (rod) e (D) cilindro do tipo Prato (plate). Figura adaptada [2].

Sistemas ou Modelos *coarse grain*

Neste item são apresentados os resultados para modelos de alta simetria com geometrias simples e complexas. Dentre os diferentes tipos de modelagem dois procedimentos foram usados. Um deles usou coordenadas espaciais pré-definidas, figuras 3.6 e 3.7, e o
5 outro usou coordenadas espaciais distribuídas aleatoriamente (MMC), figuras 3.8 e 3.9. A avaliação dos resultados deste item é baseada em modelos de alta simetria, inscritos em uma esfera de raio $100 \text{ \AA}$, e o cálculo da intensidade de espalhamento foi realizado com uso da equação (2.80). Nestas figuras, todos os modelos possuem parâmetros de modelagem idênticos, embora os tipos de preenchimento e as geometrias sejam diferen-
10 tes. O objetivo das comparações entre diferentes modelos, diferentes geometrias e suas intensidades revelam o quanto o método de modelagem e simulação poderá permitir distinguir sistemas realizados experimentalmente.

A comparação de uma geometria individual em diferentes tipos de modelagem, visualizada ao longo da observação das figuras 3.6, 3.7, 3.8 e 3.9, permite ver que a abordagem
15 proposta, contribui para realizar a distinção e evidenciar a diferença entre os histogramas de pares de distâncias e as curvas de intensidade espalhada. Do mesmo modo, quando se observa as diferentes geometrias e o mesmo processo de modelagem em cada uma das figuras anteriores, é possível observar que abordagem contribui para permitir distinguir a grande maioria dos sistemas por meio da comparação dos histogramas de
20 pares de distâncias e das curvas de intensidades, o que em um processo de minimização contra dados experimentais, pode validar um modelo em detrimento de outro.

Esse conjunto de resultados apresentados, nas figuras citadas anteriormente, evidenciam a aplicabilidade da abordagem proposta neste trabalho, para espalhadores individuais. Contudo, nem sempre um sistema experimental apresenta espalhadores monodispersos,
25 de modo que, em muitos dos casos, os espalhadores podem apresentar polidispersão em tamanho, número, intensidade etc. Assim, para esses sistemas, é necessário introduzir

Figura 3.6: Resultados para modelos com arestas preenchidas com subunidades alinhadas. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].

a polidispersão dos espalhadores.

Felizmente, para o caso de sistemas de alta simetria também é possível aplicar poli-

Figura 3.7: Resultados para modelos de alta simetria em que as arestas são preenchidas com estruturas do tipo DNA. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].

dispersidade. Nestes casos, a polidispersidade pode ser aplicada a qualquer geometria, além de também poder ser incluída nos modelos com DNA nas arestas. Porém, para

Figura 3.8: Resultados para modelos do tipo cascas. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].

Figura 3.9: Resultados para modelos do tipo sólidos. (acima) Histograma de pares de distâncias entre as subunidades e à direita os respectivos modelos geométricos. (abaixo) Exemplos de curvas de intensidade obtidas por simulação com Polygen para diferentes modelos geométricos. Figura adaptada [2].

os casos de sistemas com DNA nas arestas, a polidispersão nem sempre corresponde a simplesmente reescalar as estruturas de DNA, o que significa modificação do tamanho do DNA ou do número de nucleotídeos do modelo pretendido. Isso acontece porque em muitos destes casos, o número de nucleotídeos é o mesmo, mas a flexibilidade das estruturas é alterada, o que acaba limitando a polidispersidade de tamanho em valores não muito grandes.

Um exemplo de polidispersão em tamanho, de sistemas auto-organizados de DNA, pode ser visto na figura 3.10. Nesta figura, é possível observar o efeito, no histograma de pares de distâncias e na curva de intensidade de espalhamento, de 20% na polidispersidade de tamanho, em um modelo dodecaédrico com DNA nas arestas, inscrito em uma esfera de raio 120.0. Neste caso, a polidispersidade contribui para suavização das oscilações da curva de intensidade, uma vez que, a curva de intensidade espalhada representa uma média sobre os tamanhos dispersos.

Figura 3.10: Efeito da polidispersidade nas nanocages de DNA. (Esquerda) Histograma dos pares de distâncias. (Direita) Intensidade de espalhamento simulada. Figura adaptada [2].

Os resultados dos modelos, simples e complexos, apresentados neste item mostraram que quando usados para obter as curvas simuladas de espalhamento, podem contribuir

para o ajuste de dados experimentais. Em síntese, a abordagem proposta ao longo deste trabalho, pretende partir de um modelo de alta simetria, no espaço real, que é avaliado no espaço recíproco, com dados experimentais, para poder voltar novamente ao espaço real “ajustado”. Essa abordagem configura-se como um procedimento útil
 5 de análise de dados de experimentos de SAXS, aplicados à sistemas de alta simetria. Trata-se de uma abordagem inovadora porque, para realizar o cálculo da intensidade de sistemas de alta simetria não existe qualquer tipo de equação analítica ou semi-analítica, para a grande maioria dos casos. Os procedimentos desenvolvidos com o programa *Polygen* permitem a construção rápida de modelos muito complexos além
 10 do cálculo da intensidade espalhada. Acredita-se que essa poderá ser uma poderosa ferramenta de modelagem e simulação que pode ser usada tanto para análise de dados de espalhamento quanto para diversas outras aplicações.

Modelos *coarse grain* de Nanomateriais

Existem inúmeras maneiras de realizar o *design* e síntese de sistemas auto-organizados e
 15 até este item, este trabalho apresentou diversos tipos de modelagem de sistemas, dentre eles, alguns baseados em DNA, mas além destes, existem inúmeros *designs* possíveis. Em alguns destes casos, nanoestruturas sintetizadas experimentalmente são capazes de encapsular proteínas [87] ou nanopartículas [76, 77, 80, 91, 101], em outros, diferentes materiais podem ser compostos para sintetizar nanomateriais [65, 101–107]. Inspirados
 20 nestas diversas aplicações, um modelo composto de diferentes geometrias foi modelado e simulado com *Polygen*, figura 3.11. Este modelo, é uma combinação de um modelo de nanoestrutura de DNA na forma de um octaedro truncado, inscrito em um esfera de raio $R_{in} = 60$ e densidade eletrônica $\rho = 30$, composto junto a um modelo de casca icosaédrica, circunscrito em uma esfera de raio interno $R_{in} = 99$ e inscrito em uma
 25 esfera de raio externo de $R_{out} = 100$ com densidade eletrônica $\rho = 1$.

A modelagem da estrutura de DNA foi elaborada com uso do MEF, com as coorde-

nadas espaciais de subunidades pré-definidas. Em um momento seguinte, as mesmas coordenadas espaciais foram substituídas por coordenadas espaciais distribuídas aleatoriamente, definidas dentro de um pequeno volume centrado nas posições pré-definidas⁶.

A modelagem do icosaedro foi realizada com coordenadas espaciais distribuídas aleatoriamente com uso do MMC. A densidade de subunidades, para o modelo composto
5 final, é igual a densidade de subunidades de cada geometria.

Destaca-se nesta aplicação o fato de que, a composição de modelos diferentes não significa que a curva de intensidade será uma soma simples de cada intensidade. Como mostra a figura 3.11-C, a intensidade espalhada de cada modelo têm características
10 particulares que quando compostas não representam uma soma simples. A mesma conclusão vale para o histograma de pares de distâncias, figura 3.11-B. Este é um exemplo no qual seria, praticamente impossível, realizar o cálculo analítico ou semi-analítico para o sistema proposto, mas com ajuda deste tipo de modelagem, que pode ser realizada com número infinito de composições de modelos e geometrias, utilizando um
15 baixo número de parâmetros, a simulação do cálculo da intensidade espalhada pode ser facilmente computada. A partir do cálculo da curva de intensidade espalhada do modelo é então possível ajustá-la contra dados experimentais, com fins de encontrar os parâmetros que melhor ajustam o modelo ao sistema experimental. Espera-se que essa abordagem permita contribuir para facilitar o uso da técnica de SAXS em sistemas
20 auto-organizados e de nanomateriais com alta simetria.

Aplicações a Modelos Desconhecidos

Uma última aplicação para o caso do cálculo da intensidade espalhada de sistemas aleatoriamente orientados é apresentada. A aplicação consiste em modelar, com *Polygen*, qualquer sistema a partir de um conjunto aleatório de coordenadas espaciais inscritas
25 em uma superfície esférica. Basicamente, com esse conjunto aleatório de coordenadas

⁶Este procedimento é necessário para assegurar que os valores relativos da densidade eletrônica estão corretos, assim o volume do espaço de busca, é o volume da densidade eletrônica

Figura 3.11: Composição de uma casca icosaédrica com uma nanocage de DNA em seu interior (A) Modelos. (B) histograma de pares de distâncias para cada modelo. (C) Curva de intensidade simulada com Polygen. O tamanho e as densidades eletrônicas para cada modelo estão detalhados na própria figura. Figura adaptada [2].

espaciais é possível modelar qualquer poliedro convexo e inserir diferentes composições tais como DNA nas arestas. Para gerar esses modelos é necessário apenas escolher um número de vértices, o raio da superfície da esfera e o tipo de modelo, como por exemplo:

um tipo casca, sólido, DNA nas arestas entre outros.

Um exemplo desta aplicação pode ser visualizada na figura 3.12, em que um conjunto de 5 e 9 coordenadas espaciais, distribuídas aleatoriamente na superfície de uma esfera, foram estruturados para modelar dois diferentes poliedros. Neste caso, foram simulados o histograma de pares de distâncias e a curva de intensidade espalhada. Na mesma figura, são apresentadas duas geometrias e a partir de cada geometria foram gerados dois diferentes modelos: um com subunidades alinhadas nas arestas e outro com DNA nas arestas. Os resultados são avaliados para cada geometria em diferentes modelos e para cada modelo em diferentes geometrias.

Figura 3.12: Exemplo do modelo de cages de DNA "aleatórias". Estes modelos foram gerados a partir de um conjunto aleatório de coordenadas espaciais, distribuídos em uma superfície esférica de raio igual a 100. (A) Histogramas de pares de distâncias para cada modelo. (B) Modelo obtido por 5 e 9 de coordenadas espaciais aleatórias. Nos modelos: (esquerda) as arestas são preenchidas com subunidades alinhadas (centro) as arestas são preenchidas com estruturas do tipo DNA e para visualizar a semelhança dos modelos à (direita) estão as composições dos modelos de subunidades alinhadas e do tipo DNA. (C) Curva de intensidade espalhada dos modelos com subunidades alinhadas e com DNA nas arestas. Figura adaptada [2].

No caso de geometrias diferentes, mas comparando separadamente aquelas modeladas com DNA e aquelas modelas com subunidades alinhadas, o histograma de pares de distâncias apresenta diferenças mais pronunciadas. Já no caso de geometrias semelhantes, mas comparando as diferentes maneiras de construir o modelo, com DNA ou subunidades alinhadas, as curvas de intensidades apresentam diferenças mais pronunciadas.

Em síntese, esse conjunto de resultados somados ao uso de técnicas de minimização contra dados experimentais, apontam que a abordagem apresentada pode ser usada para realizar a diferenciação das geometrias em suas diferentes formas de composição em um sistema.

- 5 Espera-se que esse procedimento de modelagem possa apresentar um grande potencial, para ser desenvolvido em um momento seguinte, porque poderá ser aplicado como um método de modelagem *ab initio* para dados experimentais de SAXS, que combinado com um processo de heurística poderá permitir a otimização de modelos diretamente contra um conjunto de dados experimentais, variando-se neste processo o número de
- 10 vértices além de suas coordenadas espaciais.

3.2.2 Intensidade Em Sistemas Orientados

Do ponto de vista experimental, a técnica de espalhamento à baixos ângulos apresenta relativa simplicidade, e se o sistema investigado está diluído a intensidade espalhada é uma média sobre todas as possíveis orientações do espalhador, uma vez que, as posições e orientações dos espalhadores não estão correlacionadas. Ainda para esses sistemas, se os espalhadores apresentam uma geometria simples como esferas, cilindros e paralelepípedos existem um conjunto de equações analíticas ou semi-analíticas para realizar o cálculo da intensidade de espalhamento. Para geometrias mais complexas e de espalhadores únicos existem um conjunto de abordagens baseados em: transformada rápida de Fourier (FFT), transformada Indireta de Fourier (IFT) e métodos de elementos finitos (MEF) e Monte Carlo (MMC) [21]. Por outro lado, quando o experimento de espalhamento à baixos ângulos é aplicado a sistemas de espalhadores que estão orientados ou correlacionados espacialmente, os métodos analíticos e numéricos são muito mais restritos, como mostrado no item 2.2.

Uma alternativa para realizar a análise dos dados de intensidade espalhada, de sistemas com espalhadores orientados ou espacialmente correlacionados, foi proposta por Fritz-Popovski [14] e consiste em analisar a imagem bidimensional do espalhamento. Este tipo de procedimento é diferente do que em geral se realiza, que é analisar a curva unidimensional de espalhamento, em outras palavras, analisar a curva unidimensional que representa a integração angular da imagem bidimensional. De maneira geral, o método proposto por Fritz-Popovski [14], consiste em uma extensão do método de IFT, em que realiza o cálculo de uma função de distribuição de pares de distâncias em duas dimensões, e é aplicado diretamente à análise das imagens de espalhamento, permitindo realizar a análise de sistemas orientados com simetria cilíndrica, em que o eixo cilíndrico está em uma orientação bem definida em relação ao feixe de radiação incidente. Já McAlister & Grady [19, 20] propõe uma outra alternativa para realizar a

análise de sistemas orientados e espacialmente correlacionados. Nesta última proposta, a intensidade espalhada é analisada em uma direção específica do vetor de espalhamento $\mathbf{q}$, usando modelos obtidos com o MMC e simulando a intensidade com a equação de Debye otimizada, equação (2.81).

5 Baseados na proposta de McAlister & Grady [19, 20], de analisar sistemas orientados e espacialmente correlacionados, a partir da intensidade espalhada em uma determinada direção, o presente item apresenta uma alternativa para realizar a análise destes sistemas, a partir da composição de projeções de curvas de espalhamento.

A equação (2.81) permite calcular a intensidade espalhada de um sistema que está ori-
 10 entado em uma determinada direção, em relação ao feixe de incidência. Assim, um conjunto de curvas de projeções de intensidade espalhada, quando compostas em uma distribuição angular no plano, podem formar uma imagem bidimensional de espalhamento, figuras 3.13 e 3.14. Ou seja, se as equações de espalhamento permitem obter uma curva de intensidade, em uma determinada direção do vetor de espalhamento $\mathbf{q}$, é
 15 possível compor todas essas curvas para formar uma imagem de espalhamento, em que o vetor $\mathbf{q}$ é o raio da imagem, distribuído ao longo dos 360 graus de um plano, conforme figuras 3.13 e 3.14.

Nesta aplicação, a intensidade é calculada com uso do histograma de projeções de pares de distâncias, uma vez que, o interesse está justamente em obter o espalhamento em uma
 20 determinada direção. Desta forma, o histograma deixa de ser bidimensional e torna-se tridimensional, figura 3.15, porque para cada direção de $\mathbf{q}$ é obtida uma curva de intensidade. Neste caso, utiliza-se o cálculo do histograma de projeções de distâncias transformando o cálculo da soma dupla em soma simples [52, 58, 59] e o sistema de projeções de distâncias está vinculado a orientação do espalhador, em relação ao feixe
 25 incidente. Assim, o plano de vetores de $\mathbf{q}$ é um plano do espaço recíproco e representa um corte bidimensional de um espaço tridimensional, de acordo com uma determinada

Figura 3.13: Espalhamento de um feixe incidente k_i em uma determinada direção k_s , medido em um plano com raio q . O espalhamento é mensurado em um plano bidimensional.

Figura 3.14: (A) Imagem bidimensional da intensidade espalhada de um cilindro. Os cortes transversais são um guia para os olhos e mostram como uma imagem bidimensional pode ser composta com curvas unidimensionais. (B) Intensidade espalhada. (abaixo) o plano bidimensional. (acima) o gráfico tridimensional da imagem de espalhamento.

orientação.

Figura 3.15: *Histograma $h(\mathbf{qr})$ das projeções dos pares de distâncias $\mathbf{qr}$, em função do raio q e do ângulo do plano da imagem θ , variando de 0 a 2π .*

Para avaliar a validade desta proposta, foram realizados casos testes, assim como os realizados na seção anterior. Para esta avaliação, foram usados os modelos obtidos com *Polygen*, em que a imagem de espalhamento foi formada com a composição espacial de curvas unidimensionais, obtidas com equação (2.81). As imagens simuladas, foram
 5 então comparadas às imagens obtidas por meio da composição de curvas unidimensionais de espalhamento, calculadas analiticamente ou semi-analiticamente, com uso das equações (2.52 e 2.46), usando o comprimento de espalhamento $b(\mathbf{q})$ relativo a cada geometria, vide item 2.2. Nestes casos, o comprimento de espalhamento é calculado para uma determinada projeção do vetor de espalhamento de $\mathbf{q}$. Os modelos testes
 10 escolhidos (sistemas orientados e de espalhadores unitários), foram aqueles de geometria simples e que possuem equação analítica para o cálculo de intensidades: cilindro, paralelepípedo e elipsoide de revolução.

Após o subitem de sistemas testes, são apresentadas aplicações desta abordagem à sistemas de alta simetria, em diversas orientações relativas, modelados e simulados
 15 com *Polygen*. Por último, são apresentados resultados obtidos com modelos extraídos do *Research Collaboratory for Structural Bioinformatics, Protein Data Bank* (RCSB PDB, www.rcsb.org), e simulados com *Polygen*.

Sistemas Testes

Os modelos testes elaborados para validação do procedimento proposto foram quatro:

- 20 • cilindro do tipo bastão com raio de $R = 20$, altura de $L = 80$ e densidade eletrônica de $\rho = 1.0$,
- cilindro do tipo prato com raio de $R = 40$, altura de $L = 20$ e densidade eletrônica de $\rho = 1.0$,
- paralelepípedo de lados $a_x = 80$, $b_y = 20$, $c_z = 40$ e densidade eletrônica de
 25 $\rho = 1.0$

- elipsoide de revolução com raios de $R_x = 20$, $R_y = 40$, $R_z = 80$.

As figuras dos modelos testes 3.16, 3.17, 3.18 e 3.19 são compostas de 7 painéis. Nos painéis B estão os modelos gerados. Acima, ao lado dos eixos, estão os modelos orientados em relação à direção de incidência do feixe, que entra no plano da página, no sentido negativo do eixo Z. Abaixo, estão os modelos rotacionados para visualização. Os painéis A e C refazem o cálculo do histograma de pares de distâncias e o cálculo da intensidade média, respectivamente. Nos painéis C a intensidade representa uma média do modelo calculado sobre todas as orientações possíveis, e são comparadas a intensidade simulada com *Polygen* contra a intensidade teórica — equações (2.68) com os respectivos fatores de forma para cada geometria: cilindros equação (2.77), paralelepípedo equação (2.79) e elipsoide de revolução equação (2.75). Os painéis D mostram a imagem da intensidade de espalhamento simulada com *Polygen* a partir dos modelos orientados segundo painel B. Os painéis E mostram a imagem calculada analiticamente ou semi-analiticamente para as respectivas geometrias — com uso da equação (2.52), e do comprimento de espalhamento $b(\mathbf{q})$ para: cilindros equação (2.76), paralelepípedo equação (2.78) e elipsoide de revolução equação (2.74). Cortes nos ângulos de 0, 30, 45, 60 e 90 graus realizados nos painéis D e E são apresentados nos painéis G, para comparação. Os painéis F, apresentam os cortes nos mesmos ângulos dos painéis G, mas neste caso, os pontos representam o histograma de projeções de pares de distâncias, em cada uma das direções, como mostrado na figura 3.15.

Para poder avaliar os resultados obtidos com uso dos sistemas testes, foram realizadas comparações entre os resultados esperados e aqueles obtidos para a imagem de intensidade espalhada (painéis D e E). Os modelos cilíndricos, paralelepípedo e elipsoide de revolução apresentam boa semelhança entre a imagem simulada e a calculada. Essa semelhança é mais notável quando se observa as projeções das intensidades em um direção específicas nos painéis G. Evidentemente, é também menos pronunciada para altos valores de q e principalmente para ângulos intermediários aos eixos perpendiculares da

imagem da curva simulada. A baixa precisão nestes ângulos, acontece porque, durante o processo de construção do histograma de projeções de distâncias, equação (2.81), são computadas diferentes distâncias que acabam apresentando semelhantes projeções. Em outras palavras, dependendo da distribuição numérica de subunidades em uma determinada orientação, relativa ao feixe de incidência, a precisão do cálculo pode ser maior nos eixos perpendiculares entre o feixe e a parte do modelo com maior distribuição ou concentração espacial de subunidades. Por outro lado, o histograma de projeções de distâncias, pode sobrepor projeções que são relativas à valores de pares de distâncias muito diferentes. Assim, dependendo da distribuição das subunidades do modelo, as partes que formam ângulos intermediários aos perpendiculares ao feixe, podem apresentar uma precisão menor.

Para os mesmos modelos usados para os testes de verificação, do procedimento de cálculo das imagem de espalhamento, foram simuladas também as imagens em três diferentes orientações, em relação ao feixe de incidência. Na figura 3.20, é possível inclusive comparar dois diferentes modelos de cilindros em três orientações, e na figura 3.21 a comparação é entre modelos do elipsoide de revolução e o paralelepípedo. As imagens de espalhamento evidenciam que, para cada uma das diferentes orientações, o feixe incidente atinge o espalhador com um determinado ângulo formando diferentes padrões de espalhamento. Desta forma, seria possível obter mais informações do espalhador a partir da imagem de espalhamento, em comparação com a curva unidimensional.

A comparação realizada entre as simulações das imagens com *Polygen*, contra aquelas calculadas com equações analíticas, evidenciam que é possível realizar a simulação da imagem de intensidade espalhada a partir da composição de curvas unidimensionais de projeções das intensidades. Além desta estratégia de calcular a imagem espalhada, é possível calcular a imagem de espalhamento pixel a pixel, o que tornaria a precisão da imagem mais refinada ao mesmo tempo que aumenta o custo computacional e espaço necessário para armazenar os dados dos cálculos intermediários, que pode

chegar facilmente a 1 giba bytes.

Os próximos itens apresentam possíveis aplicações que poderiam ser realizadas com a abordagem apresentada, contribuindo para resolver sistemas orientados, centro-simétricos ou não centro-simétricos e também aqueles que possuem correlação espacial. Essa abordagem, para sistemas com correlação espacial, poderá representar um importante passo, para resolver sistemas que não apresentam equações analíticas nem estratégias numéricas para o cálculo de intensidade.

Figura 3.16: Simulação das imagens de espalhamento do cilindro do tipo bastão. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.

Figura 3.17: Simulação das imagens de espalhamento do cilindro do tipo prato. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.

Figura 3.18: Simulação das imagens de espalhamento do paralelepípedo. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.

Figura 3.19: Simulação das imagens de espalhamento do elipsoide de revolução. A) Histograma de pares de distâncias. B) (acima) modelo usado para o cálculo da intensidade projetada na orientação de incidência do feixe, que atravessa perpendicularmente a folha. (abaixo) modelo rotacionado para visualização. C) Intensidade média simulada e a calculada analiticamente. D) Imagem de espalhamento simulada com Polygen de acordo com o modelo em B. E) Imagem calculada. F) Histograma de projeções de pares de distâncias em 5 projeções: 0, 30, 45, 60 e 90 graus. G) Intensidade projetada nas mesmas projeções de (F), projeções: 0, 30, 45, 60 e 90 graus. As intensidades correspondem a cortes nas figuras (E) e (F), nas respectivas projeções.

Figura 3.20: A à C) imagens de espalhamento, do cilindro do tipo bastão em três diferentes orientações em relação ao feixe de incidência. D à F) Imagens de espalhamento, do cilindro do tipo prato em três diferentes orientações em relação ao feixe de incidência.

Figura 3.21: A à C) Imagens de espalhamento, do elipsoide de revolução em três diferentes orientações em relação ao feixe de incidência. D à F) imagens de espalhamento, do paralelepípedo em três diferentes orientações em relação ao feixe de incidência.

Modelos de Alta Simetria

Os primeiros modelos escolhidos para esta aplicação, foram modelos que apresentam geometrias muito semelhantes entre si, quando orientados na mesma direção. Assim, além de poder visualizar os diferentes padrões da intensidade espalhada é possível avaliar a capacidade de distinguir as diferentes geometrias a partir da imagem de intensidade espalhada. Na figura 3.22 é possível ver os modelos escolhidos, três sólidos de Johnson: J14, J15 e J16. Nesta figura, a painel B apresenta (acima) os modelos na orientação de realização da simulação — o feixe incide perpendicular ao plano da página — e (abaixo) os mesmos modelos rotacionados para visualização. Da esquerda para direita estão os modelos: J14, J15, J16 e o último modelo, é uma sobreposição dos modelos anteriores. Os modelos J14 e J15 foram construídos inscritos em uma esfera de raio $R = 80$ e o modelo J16 em uma esfera de raio $R = 60$, todos eles com densidades eletrônicas iguais de $\rho = 1.0$. Os painéis A e C mostram respectivamente o histograma de pares de distâncias e a curva de intensidade média. As imagens da intensidade espalhada, para cada geometria, são apresentadas nos painéis: D para o J14, E para o J15 e F para o J16. Para as geometrias J15 e J16 é possível ver uma pequena diferença nas curvas de intensidade média que é amplificada nas comparações entre as imagens bidimensionais de espalhamento.

Em um outro conjunto de geometrias, foram usados modelos ainda mais semelhantes que os anteriores, os sólidos de Johnson J53 e J54. Esses modelos foram gerados inscritos em uma esfera de raio de $R = 80$ e densidades eletrônicas iguais de $\rho = 1.0$. Os resultados para simulações das intensidades espalhadas destes modelos podem ser vistas na figura 3.23. Nesta figura, o painel A mostra os modelos (preto) J53 (acima) orientado em relação à direção de incidência do feixe e (abaixo) o mesmo modelo rotacionado. Também no mesmo painel está o modelo (cinza) J54 (acima) orientado em relação à direção de incidência do feixe e (abaixo) o mesmo modelo rotacionado. E

Figura 3.22: Simulação das imagens de espalhamento dos poliedros de Johnson: J14, J15 e J16. A) Histograma de pares de distâncias. B) (acima) Composição dos modelos dos sólidos de Johnson: J14, J15 e J16, na orientação de incidência do feixe. (abaixo) modelos rotacionados para visualização. C) Intensidade média simulada. D à F) Imagens de espalhamento, de acordo com os modelos (acima) na orientação de incidência do feixe; (abaixo) modelos rotacionados para visualização. Da esquerda para direita: J14, J15 e J16.

por último, para comparação visual, é mostrada a sobreposição espacial dos modelos. Nesta orientação, o feixe de incidência está entrando no plano da página. Os painéis B e C representam respectivamente o histograma de pares de distâncias e a curva de intensidade média espalhada para cada modelo. A curva de intensidade média de ambos

os modelos são muito próximas entre si. As imagens, D do modelo J53 e E do modelo J54, mostram a imagem da intensidade espalhada de cada modelo. A comparação entre as imagens permite ver padrões de espalhamento bem diferentes entre os modelos, amplificando novamente a pequena diferença visualizada nas curvas unidimensionais de espalhamento.

Figura 3.23: Simulação das imagens de espalhamento dos poliedros de Johnson: J53 e J54. A) (cinza escuro) Orientação do modelo do sólido de Johnson J53 na orientação de incidência do feixe e abaixo o mesmo modelo rotacionado. (cinza claro) Orientação do modelo do sólido de Johnson J54 na orientação de incidência do feixe e abaixo o mesmo modelo rotacionado. Os modelos são compostos em uma mesma figura mais abaixo. B) Histograma de pares de distâncias. C) Intensidade média simulada. D) Imagem da intensidade de espalhamento do sólido J53, de acordo com a orientação do modelo em A. E) Imagem da intensidade de espalhamento do sólido J54, de acordo com a orientação do modelo em A.

Uma outra aplicação de modelos de alta simetria é apresentada. Nesta aplicação um

outro conjunto de diferentes modelos que apresentam características geométricas próximas entre si, são avaliados. Os modelos escolhidos foram: tetraedro, octaedro, J12, J13, J49 e J50, todos com densidade eletrônica de $\rho = 1.0$ e inscritos em uma esfera de raio de $R = 80$. Na figura 3.24 os modelos podem ser visualizados junto aos histogramas de distribuição de distâncias, a curva de intensidade média e as respectivas imagens de espalhamento dos modelos orientados. Assim como apresentado para os modelos de alta simetria dos itens anteriores, neste caso é possível ver novamente que, a imagem bidimensional de espalhamento apresenta mais informações que a curva unidimensional, o que contribui para evidenciar a diferença nas imagens de espalhamento entre os modelos.

Com último exemplo para sólidos geométricos, a intensidade espalhada para sistemas orientados e aleatoriamente orientados é apresentada baseada em um sistema sintetizado [108]. Neste sistema, uma pirâmide de base pentagonal pode alcançar diferentes diâmetros durante um específico processo de síntese. Baseado neste trabalho, foram escolhidos quatro diferentes diâmetros que inscrevem a pirâmide pentagonal, reportados por Banerjee et al. [108]. Os quatro modelos foram inscritos em esferas de diâmetros iguais a: 20 nm, 35 nm, 48 nm e 65 nm e para cada um destes modelos foi gerado um histograma de distâncias e uma curva de intensidade média. Por fim o modelo de 20 nm foi construído em três diferentes orientações e para cada um deles, foi simulada uma imagem de espalhamento. Os resultados obtidos podem ser visualizados na figura 3.25. O histograma de distâncias mostra as diferentes dimensões de cada modelo e a intensidade média apresenta diferentes perfis, assim como, as imagens de intensidade espalhada apresentam diferentes padrões para o mesmo modelo em três diferentes orientações. Este último resultado tem como objetivo apresentar uma das possíveis abordagens disponíveis no programa *Polygen*, aplicada em um sistema experimental.

Figura 3.24: Simulação das imagens de espalhamento de diferentes poliedros com geometrias similares. A) Histograma de pares de distâncias. B) Da esquerda para direita: (acima) Orientação do modelo do tetraedro, octaedro, J12, J13, J49 e J50, na orientação de incidência do feixe. (abaixo) modelos rotacionados para visualização. O modelo mais a direita é uma composição dos modelos. C) Intensidade média simulada. Imagens de espalhamento, de acordo com os modelos em B: D) tetraedro, E) octaedro, F) J12 G) J13, H) J49 e I) J50.

Figura 3.25: Simulação da intensidade espalhada para pirâmides pentagonais. A) Histograma de pares de distâncias. B) Modelos da pirâmide pentagonal em três diferentes Perspectivas. C) Intensidade média simulada. D à F) Imagens da intensidade espalhada, de acordo com os modelos acima de cada plano. Da esquerda para direita três diferentes orientações da pirâmide pentagonal de diâmetro de 20 nm.

Proteínas

As imagens de espalhamento geradas nos itens anteriores foram baseadas em modelos construídos com o programa *Polygen*, neste subitem os modelos usados foram extraídos do *Research Collaboratory for Structural Bioinformatics, Protein Data Bank* (RCSB PDB, www.rcsb.org) [3], e simulados com *Polygen*. Três modelos de proteínas foram escolhidos: lisozima - *pdb id*: 4NY5 [4], colágeno de peptídeo - *pdb id*: 2CUO [5], e albumina (somente a cadeia A) - *pdb id*: 3V03 [6]. As experiências de espalhamento são conhecidas como experiências de baixa resolução porque fornecem informações médias sobre um conjunto muito grande de espalhadores e por esse motivo os modelos do tipo *coarse grain*, usando para esse caso as coordenadas do átomos C_α para posicionar as subunidades de densidade eletrônica $\rho = 1$.

A figura 3.26 mostra os resultados obtidos para a simulação com *Polygen* das proteínas. As imagens da simulação dos histograma de pares de distâncias e da curva de intensidade média podem ser vistos respectivamente nos painéis A e C. Os modelos das proteínas simuladas podem ser vistos no painel B. As imagens da intensidade de espalhamento e os respectivos modelos, em sua orientação de simulação, podem ser vistas nos painéis D à I. Para os modelos da proteína *albumin* são simuladas duas orientações perpendiculares e as informações dos padrões obtidos em cada imagem simulada são distintas. O mesmo acontece, para a simulação das imagens da proteína *lysozyme* simulada em três orientações perpendiculares diferentes. Estes últimos exemplos, mostram uma nova abordagem para realizar a análise de dados de espalhamento a partir da imagem da intensidade espalhada, de sistemas orientados, centro-simétricos ou não centro-simétricos e pode servir de base para novas metodologias de análise de dados.

Figura 3.26: Simulação das imagens de espalhamento para modelos de proteínas obtidos em Protein Data Bank (RCSB PDB, www.rcsb.org) [3]. A) Histograma de pares de distâncias. B (acima) Orientação dos modelos de lysozyme pdb id: 4NY5 [4], colágeno de peptídeo: pdb id: 2CUO [5], e albumin (somente a cadeia A) pdb id: 3V03 [6], na orientação de incidência do feixe de incidência. (abaixo) modelos rotacionados para visualização. C) Intensidade média simulada. D à I) Modelos na orientação simulada e imagens de espalhamento, de acordo com os modelos em B: D) colágeno de peptídeo; E e F) duas orientações da albumina; G, H e I) três orientações da lysozyme.

3.2.3 Verificação dos Graus de Precisão e Custos Computacionais

Os métodos de modelagem e simulação foram apresentados, e neste item será realizada a verificação dos graus de precisão usados nos cálculos, das curvas de intensidade espalhada, e a correspondente discussão junto aos custos computacionais.

5 A exata determinação dos graus de precisão e custos computacionais demanda um esforço e um grau de refinamento superior ao que será apresentado neste item. Uma vez que esta tarefa não é o escopo deste trabalho, o objetivo principal nesta etapa é quantificar possíveis limitações, além de, desenhar estratégias para otimizar o tempo e o nível de precisão dos cálculos da curva de intensidade.

10 Assim, para realizar esta atividade, as verificações foram realizadas para o modelo de um cilindro, do tipo bastão, com raio de $R = 30$ e comprimento de $L = 100$, de modo que, todas as simulações das intensidades foram comparadas com o cálculo semi-analítico da intensidade, obtido segundo as equações (2.68 e 2.69).

A equação (2.60) de Debye, é por definição uma equação analítica, e neste trabalho é aplicada no cálculo da intensidade espalhada de modelos compostos por um conjunto de subunidades. Deste modo, a verificação do grau de precisão dos cálculos da intensidade, consiste em verificar a capacidade do uso do MEF, aplicado à modelagem, junto ao uso da equação de Debye otimizada. Assim, a principal verificação da precisão consiste em avaliar a capacidade de modelagem de sistemas espalhadores, representando-os como
15 um conjunto de espalhadores conhecidos, no caso, subunidades esféricas, e comparando-os com os resultados analíticos conhecidos.

Desta forma, a primeira verificação será realizada avaliando o tamanho das subunidades, que está relacionado ao número e a densidade de subunidades, que compõe um determinado modelo. A segunda verificação será realizada sobre o histograma de pares
25 de distâncias, na busca pela melhor relação entre a otimização do custo computacional

e a precisão da descrição do modelo.

A modelagem com uso do MEF, pressupõe que o tamanho das subunidades deve ser sempre menor que o menor detalhe da estrutura a ser modelada [20, 51, 56, 58, 59, 109] de acordo com:

$$(vn)_{subunidades} = (V)_{modelo} . \quad (3.3)$$

5 Em que v e V são os volumes das subunidades e do modelo, respectivamente.

Para a primeira avaliação, o passo dos canais do histograma de pares de distâncias esteve fixo em 0.01, de forma que, a variação dos raios das subunidades foi realizada de acordo com a relação (3.3), levando em consideração o tamanho do modelo. Os resultados desta avaliação podem ser vistos na figura 3.27, observando-se a variação
10 do próprio raio da subunidade, ou ainda, a densidade de subunidades que compõe o modelo. Para este caso, quanto maior o raio da subunidade, menor será a densidade em número, afim de preencher o volume do modelo, e vice-versa.

Basicamente para este caso teste, assume-se que raio mínimo do modelo é metade da menor dimensão do cilindro, neste caso $R_{min} = 7.5$, de forma que $q_{max} = 0.42$ de
15 acordo com a relação 2.43. Assim, a representação do modelo com subunidades deve usar o valor da metade do menor detalhe estrutural do cilindro, de acordo com o limite de Nyquist. Ou seja, o diâmetro da subunidade deve ser de 3.5, o que fornece um raio de 1.75. Essa relação pode ser verificada na figura 3.27, em que é possível ver que valores da intensidade, para os modelos gerados com raio de subunidade maiores
20 que 1, começam a divergir dos valores esperados. O painel C da figura 3.27 mostra com mais detalhes o efeito do tamanho da subunidades na comparação entre a curva da intensidade simulada e a calculada semi-analiticamente, conforme mostra a soma

acumulada destas diferenças, ξ . Em que:

$$\xi(q) = \frac{1}{n} \sum_{i=1}^n \frac{|I_i^{\text{simulado}}(q) - I^{\text{teorico}}(q)|}{I^{\text{teorico}}(q)} \quad (3.4)$$

Para o modelo de cilindro usado, em que o cálculo da intensidade de espalhamento vai no máximo até $q_{\max} = 0.4$, e o raio das subunidade menor é que 2, não existem grandes divergências, entre a intensidade calculada usando o modelo construído com MEF e a
 5 calculada semi-analiticamente. Por outro lado, para fins de testes, assumindo-se um $q_{\max} = 1.0$ ou um modelo com precisão de $R_{\min} = 3.0$, o valor do raio da subunidade deverá ser menor que $r < 0.7$, para uma boa representação. Assim, quanto maior a precisão do modelo a ser representado, menor deverá ser o tamanho da subunidade. Este resultado leva a conclusão que, os valores do raio de giro, vetor de espalhamento
 10 e raio da subunidade devem sempre ser avaliados de forma correlacionada, antes da representação de qualquer modelo.

Como discutido nos itens anteriores, os experimentos de espalhamento à baixos ângulos são experimentos de baixa resolução, de maneira que, a intensidade espalhada é sempre o resultado de um conjunto muito grande de espalhadores. Assim, quanto mais homo-
 15 gêneo possível é o volume do conjunto de subunidades esféricas, em relação ao volume total do modelo pretendido, maior será o grau de precisão do cálculo [57].

De modo geral, a figura 3.27 evidência que os raios de subunidades menores que 2 são os mais adequados para a representação de modelos nanométricos, porque preenchem, com um grau de uniformidade e homogeneidade o modelo pretendido.

20 Uma vez obtido o tamanho de subunidade ou a correspondente densidade de subunidades, necessária para a uniformidade e homogeneidade do modelo, será realizada a avaliação do tamanho do passo do histograma, ou em outras palavras, o número de intervalos (que está relacionado ao tamanho de cada intervalo). Para esta avaliação, o

A **B** **C**

Figura 3.27: Avaliação da influência de raio da subunidade, r , na precisão dos cálculos de intensidade, junto a densidade, d , de subunidades. A) Modelos obtidos com MEF, para cada um dos diferentes tamanhos de subunidades. B) Curva de intensidade. Dentro do painel B, um zoom realizado para altos valores de q . C) Soma acumulada da diferença normalizada, entre a curva de intensidade simulada e a calculada semi-analiticamente.

raio da subunidade escolhido foi de $r = 1$. O teste do histograma levou em consideração o limite de Nyquist, que diz que, o tamanho do passo do histograma não deve exceder

uma largura maior que a metade da menor dimensão do modelo [21]. Neste caso, o passo do histograma, dr deve seguir a seguinte relação $dr < \pi/q_{max}$ ou $dr < R_{min}/2$ [21], que para o caso em estudo, significa que $dr < 0.5$, em que o raio mínimo é o raio da subunidade. Seguindo esta relação, a figura 3.28 apresenta os resultados para estes testes, com diferentes valores de passo para o histograma.

Figura 3.28: Avaliação da influência do passo do histograma para a precisão dos cálculos de intensidade. A) Os vários histogramas normalizados e realizados com diferentes passos. B) Curva de intensidade. C) Soma acumulada da diferença normalizada, entre a curva de intensidade simulada e a calculada semi-analiticamente.

Os resultados da figura (3.28) apontam que valores de histograma com passo 10× menor que o valor limite de Nyquist podem ser uma boa escolha para construir o histograma e consequentemente realizar o cálculo da intensidade de espalhamento, sem grandes perdas de precisão, como mostra a figura (3.28)-C. Neste caso, o teste do histograma foi realizado para o cálculo da intensidade média (sistemas aleatoriamente orientados), já no caso da intensidade de sistemas orientados, a avaliação realizada não pode ser extrapolada, uma vez que, calcula-se o histograma de projeções de distâncias e não o módulo de distâncias. Desta forma, como no caso de sistemas aleatoriamente orien-

tados, o valor do passo do histograma deve seguir a seguinte relação $dr < R_{subunidade}$. No entanto, para o caso de sistemas orientados o histograma é realizado sobre a projeção de distâncias, de modo que, $R_{min}^{projção} = R_{min} \cdot \hat{\mathbf{k}}_s \propto \cos(R_{min})$ (podendo assumir qualquer valor inclusive zero), podem haver casos em que seja necessário utilizar passos
 5 muito pequenos nos histogramas. Isso pode acontecer, por exemplo, para partículas com grandes anisotropias (partículas muito alongadas, achatadas, etc). Nas aplicações apresentadas nesta tese o passo de 0.01 (i.e., 1/100 do raio da subunidade) mostrou-se suficientemente preciso para os cálculos da intensidade, em relação ao valor obtido analiticamente. Ainda que seja necessário um passo menor, o *Polygen* permite a alteração
 10 destes valores.

Nas imagens de espalhamento apresentadas, é possível ver que existe um padrão que se repete em cada um dos quadrantes, de forma que, para otimizar o tempo computacional para o caso do cálculo da imagem de espalhamento, é possível fazer o cálculo para o primeiro quadrante do plano e completar os demais quadrantes copiando e refletindo
 15 o quadrante calculado, sobre cada um dos semi-eixos, reduzindo o cálculo em até $4\times$. Ainda para compor a imagem de espalhamento, uma série de curvas de projeções de intensidade unidimensionais são calculadas separadamente, de forma que, um procedimento de paralelização de partes do algoritmos do programa, poderia reduzir o tempo de cálculo, da imagem, em até $5\times$.

20 Por último, para otimizar o cálculo do histograma de pares de distâncias usando a proposta de Hansen [51] e Pedersen et al. [60], neste caso dividindo o cálculo do histograma de pares de distâncias em blocos, é necessário definir uma amostragem mínima, para cada bloco de distâncias. Assumindo um tamanho de espalhadores de cerca de 150Å, que é modelado com uso do MMC, as subunidades possuem distribuição
 25 uniforme de posição ao longo do volume do modelo, o tamanho mínimo da amostra escolhida, para cada bloco, deve ser da ordem de $(R_{max} - R_{min})^2/12$ [110], ou cerca de 2000 subunidades, o que para o mesmo modelo poderia reduzir o cálculo $20\times$. Em

síntese, os processos de otimizações podem reduzir até $400\times$ o número de iterações, o que permitiria elaborar estratégias para realizar o ajuste de dados simulados contra dados experimentais, diretamente na imagem de espalhamento.

3.3 SAXS: Conclusões Parciais e Perspectivas

Um método para a modelagem rápida e a simulação da curva de intensidade foi apresentado e pode ser utilizado com um número baixo de parâmetros. Para unificar o processo de modelagem e simulação, um programa, inicialmente apelidado de *Polygen*,
5 foi elaborado. O programa permite construir um conjunto de modelos do tipo sólidos, cascas, com esferas ou DNA nas arestas, que podem ser unitários ou compostos em diferentes geometrias convexas e com diferentes densidades eletrônicas. A partir destes modelos é possível simular a curva de intensidade média que pode ser usada junto à dados experimentais e consequentemente ajustada pelos métodos de minimização de
10 dados. Esse processo de minimização pode revelar os melhores parâmetros estruturais ou eletrônicos que ajustam um determinado modelo.

Os modelos ou geometrias disponíveis são apresentados no Apêndice A, além daqueles que podem ser modelados a partir de posições aleatórias distribuídas na superfície de uma esfera. A composição de modelos também pode ser realizada de diversas formas
15 e com diferentes polidispersões. Trabalhos anteriores [22, 23, 25], mostraram que o uso de métodos de espalhamento permitem o acesso direto à estrutura média de sistemas auto-organizados contribuindo para fornecer informações sobre o rendimento de determinada forma ou geometria. Neste sentido, acredita-se que a possibilidade de combinar vários tipos de modelos, usando um baixo número de parâmetros, com o objetivo de
20 descrever um determinado sistema molecular, poderá fornecer uma poderosa estratégia para a análise de dados de espalhamento, principalmente quando as atuais ferramentas apresentadas podem contribuir para analisar sistemas com formas complexas, que em geral não possuem métodos analíticos para análise.

Além das simulações apresentadas, da intensidade média para sistemas aleatoriamente
25 orientados, uma alternativa para sistemas orientados e correlacionados espacialmente foi apresentada. Nesta aplicação, o cálculo da intensidade, projetada em uma determinada

direção do vetor de espalhamento $\mathbf{q}$, é usado para compor imagens de espalhamento. Esta abordagem, permite simular a intensidade de sistemas orientados, reduzindo as perdas de informações com os processos de integração de imagens.

Recentes investigações de sistemas nanométricos estão usando uma fonte conhecida
5 como *X-ray free-electron lasers* (XFEL) capaz de irradiar um espalhador separadamente e obter uma imagem de espalhamento unitária para cada espalhador. Nestes casos, a intensidade espalhada é coerente e corresponde a uma determinada orientação do espalhador, quando é atingindo pelo feixe incidente, de modo que, o pulso de incidência é mais rápido que o movimento de rotação do próprio espalhador. As análises das
10 imagens de espalhamento, obtidas pela experiência de XFEL, são baseadas no método proposto por Kam [42] e consistem, basicamente, em correlacionar as várias imagens de espalhamento, do mesmo espalhador em várias orientações. Deste modo, um volume de imagens de espalhamento é correlacionado no espaço recíproco, em uma esfera de Ewald, de modo a obter o volume no espaço real. Inspirados por este método e baseados nos
15 resultados apresentados acredita-se que, os procedimentos para simular as imagens de espalhamento, apresentados neste trabalho, podem contribuir para fornecer informações de sistemas orientados e correlacionados espacialmente. Assim acredita-se que, esta última abordagem apresentada junto a procedimentos de minimização de imagens e um banco de dados de imagens de espalhamento, dispostas em uma esfera de Ewald,
20 poderiam contribuir para obter, além dos parâmetros estruturais do sistema em estudo, a orientação do sistema em relação ao feixe de incidência. Neste caso, as imagens do banco de dados contribuiriam para facilitar a determinação da orientação de um modelo, sem a necessidade de realizar todos os cálculos de intensidades.

Os resultados apresentados neste capítulo, para intensidade média, fazem parte do ar-
25 tigo [2]: Cassio Alves, Jan Skov Pedersen, and Cristiano Luis Pinto Oliveira. Modelling of High Symmetry Nanoscale Particles by Small-Angle Scattering. *Journal of Applied Crystallography*, 47:84-94, 2014.

Os resultados para a intensidade de sistemas orientados fazem parte de um texto em vias de finalização para publicação.

CAPÍTULO 4

DINÂMICA MOLECULAR (DM)

Neste capítulo são apresentados os resultados obtidos no estágio realizado na Universidade de Roma Tor Vergata (15-09-2012 a 15-03-2013), sob orientação do professor Alessandro Desideri e Mattia Falconi, na área de Dinâmica Molecular ¹ além da orientação do supervisor deste trabalho.

Uma das vantagens que a DM oferece, neste caso, é contribuir para encontrar os melhores parâmetros do *design* experimental, além de, prever a estabilidade do sistema auto-organizado, antes de qualquer gasto com síntese de materiais. Num momento *a posteriori*, a mesma trajetória de DM pode ser usada para realizar as análises, em conjunto com os resultados experimentais, aumentando a capacidade de análise e previsão do comportamento do sistema.

O primeiro passo para obter a trajetória foi realizar a modelagem e o *design* das nanoestruturas do ponto de vista atomístico, e uma vez realizada, a simulação com DM, as

¹Programa ciências sem fronteiras (processo: 237230/2012-7/SWE).

nanoestruturas são avaliadas por meio da análise da trajetória.

Os dois sistemas, auto-organizados de DNA, estudados neste trabalho, tinham a geometria de um octaedro truncado em que as arestas eram modeladas com duplas hélices (ds) de DNA de 18 bp, e as ligações entre elas feitas de fitas simples de DNA (ss). As fitas simples serão denominadas de *linkers*, porque ligam as duplas hélices de DNA. As sequências das duplas hélices podem ser vistas no apêndice C e para o caso dos *linkers*, uma das nanocages tinha todos os links compostos de 7 adeninas (7A) e outra tinha a seguinte sequência ATATATA (7AT).

O estudo de DM com esses sistemas representa uma extensão de um conjunto de trabalhos publicados [7,24,25] nos quais são avaliados a importância do comprimento dos *linkers* (2, 3, 4, 5 ou 7 Timinas) para estabilidade de *nanocages* na forma de um octaedro truncado.

Inspirados na avaliação da estabilidade da *nanocage* octaédrica [7,24,25], que investigou a importância do comprimento dos *linkers* de Timinas [7,24,25], este trabalho optou em realizar a avaliação da estabilidade investigando diferentes composições de um mesmo comprimento de *linker*. Assim, outras duas *nanocages* idênticas, na mesma forma de um octaedro truncado, foram estudadas com *linkers* de 7A e 7AT. Como o modelo de *nanocage*, com *linkers* de 7T, já havia sido elaborado em trabalhos anteriores, assim como trajetória de DM já havia sido simulada [7,24,25], esses resultados foram usados para realizar a comparação com os resultados obtidos com os modelos de *nanocages*, de *linkers* de 7A e 7AT, simulados neste trabalho com DM. Este tipo de comparação, buscava verificar a estabilidade dessas novas nanoestruturas, avaliando parâmetros que podem ser mensurados experimentalmente, bem como, avaliando parâmetros que não são possíveis de obter por meio da realização de experiências. Os parâmetros calculados foram: *Root Mean Square Deviations* (RMSD), *Root Mean Square Fluctuations* (RMSF), Raio de Giro (Rg), número de ligações de hidrogênio H, número de interações

do tipo π – *stacking*.

4.1 *Design* e Modelagem

O *design* das *nanocages* foi baseado nos trabalhos já publicados [7,24,25] e basicamente constitui em posicionar as duplas hélices de DNA, cada uma com 18 bp, nas arestas de um octaedro truncado. Cada dupla hélice neste caso será identificada como *DH* para um total de 12 duplas hélices, apêndice C. Os *linkers* foram posicionados entre as duplas hélices, unindo as extremidades de cada uma, no total de 24 *linkers DH*- por *nanocage*. O número total de nucleotídeos, usados nos modelos, são de 600, figura 4.1.

Figura 4.1: Estruturas de uma *nanocage*, auto-organizada de DNA, na geometria de um octaedro truncado. (a) Modelo atômico na representação de átomos como esferas. (b) Modelo atômico na representação da superfície acessível ao solvente.

A modelagem atômica, do octaedro truncado, foi realizada com uso do programa SYBYL 6.0 (TRIPOS, <http://www.tripos.com/>). A topologia do sistema foi realizada por meio do programa Tleap AMBER12 [111] e a *nanocage* foi parametrizada com o campo de força *parmbsc0* [112], um refinamento do campo de força AMBER99 [113].

Para realizar a simulação de DM a *nanocage* foi imersa em uma caixa octaédrica com água. O modelo das moléculas do solvente utilizado foi o TIP3 [114] e a distância mínima, entre a *nanocage* e a borda da caixa, foi de 14 Å. A neutralização das cargas foi realizada com 600 contraíons de Na^+ .

5 4.2 Protocolos de Equilíbrio, Termalização e DM

As simulações de DM foram realizadas com o programa AMBER. Foram realizadas três minimizações de energia. Nas duas primeiras minimizações foram impostas restrições energéticas de 500 kcal/mol e nenhuma restrição energética para a última. Na primeira minimização de energia, as restrições foram aplicadas para todos os átomos da *nanocage*,
10 e na segunda foram impostas somente nas duplas hélices para minimizar os *linkers*.

O sistema foi simulado com condições periódicas de contorno com valor de *cutoff* de 12 Å para o cálculo de interações de curto alcance e o método de *Particle Mesh Ewald* (PME) foi utilizado para interações eletrostáticas de longo alcance. A temperatura da DM foi fixada em 300 K, com uso da dinâmica de Langevin com pressão constante de
15 1 atm. O processo de termalização, do sistema, foi dividido em três etapas de 30 ps cada, com passos de 1 fs. A primeira foi de 0 K até 100 K, a segunda foi até 200 K e a última até chegar a 300 K. A DM foi realizada em passos de 2 fs durante 50 ns. O tempo máximo usado, foi de 50 ns, porque após esse valor o sistema não demonstra nenhuma grande mudança, como mostram também os trabalhos anteriores [7, 24, 25].
20 O intervalo de tempo, entre 15 ns e 50 ns, foi utilizado para realizar as análises da trajetória da simulação porque tanto o R_g como os valores de RMSD da *nanocage* são estáveis somente após 15 ns.

4.3 Análise da DM

Nesta seção é apresentada uma síntese das análises da trajetória de DM da *nanocage* com *linkers* de 7A e 7AT estudada e simulada durante o estágio do programa CSF (programa Ciências Sem Fronteiras, processo: 237230/2012-7/SWE). Os resultados destas

5 *nanocage* foram comparados com a *nanocage* de *linkers* de 7T (modelada, medida experimentalmente e analisada também com DM em trabalhos anteriores [7, 24, 25]). Para realizar as comparações entre as diferentes *nanocages* (7A, 7AT e 7T) foi realizada o mesmo tipo de análise para ambas. A trajetória da *nanocage* de 7T, simulada pelo estudante Francesco Oteri [7], foi obtida com o professor Alessandro Desideri da Uni-

10 versidade de Roma Tor Vergata e toda a análise da trajetória foi refeita neste trabalho e é apresenta de modo comparativo nos itens seguintes.

4.3.1 Análise da Estabilidade e de parâmetros geométricos das *Nanocages*

O RMSD é um parâmetro de descreve o desvio médio quadrático do movimento dos átomos da *nanocage* por meio da comparação entre as posições dos átomos, em cada

15 passo da trajetória, com a sua estrutura topológica inicial, figura 4.1. Na figura 4.2 é possível ver a evolução do parâmetro RMSD para as *nanocages* de DNA 7A, 7AT e 7T [7, 24, 25].

A figura 4.3 mostra a evolução do raio de giro (R_g) de ambas as *nanocages*, 7A, 7AT e 7T, ao longo da trajetória. O valor do $R_g = 6.25(07) \text{ nm}$, da *nanocage* de 7T obtido

20 experimentalmente [25], é um valor próximo aos valores finais obtidos com trajetória das *nanocages* simuladas que é de $R_g = 6.4 \text{ nm}$ (50 ns). Nesta mesma figura é possível observar que, para tempos maiores que 15 ns (15000 ps), os raios de giros de ambas as *nanocages* não apresentam grandes oscilações, o que pode indicar uma maior estabilidade dessas estruturas, e a variação relativa entre cada *nanocage* é pequena. A

25 estabilidade ou a pequena oscilação dos valores também é vista, na figura 4.2, para a

Figura 4.2: *RMSD das nanocages auto-organizadas de DNA com 7A e 7T e 7AT nos linkers. Os dados são apresentados separados para comparar somente os linkers, somente as duplas hélices de cada nanocage e a composição de ambos (linkers + duplas hélices = nanocage). Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].*

evolução do RMSD.

As análises de RMSD e Rg indicam que as *nanocages* apresentam maior estabilidade a partir de 15 ns e por esse motivo os demais parâmetros analisados neste trabalho serão avaliados entre 15 e 50 ns.

- 5 Similarmente ao RMSD o RMSF descreve a flutuação média quadrática em relação a uma referência. O RMSF Local e Global, foram obtidos para as duplas hélices de DNA, descontados os seus movimentos de rotação e translação. No caso do RMSF Local, figura 4.4, a flutuação foi calculada sem a referência da conformação espacial da *nanocage*, e no caso do RMSF Global, figura 4.5, a flutuação foi calculada usando como
- 10 referência a conformação espacial de toda a *nanocage*. Neste último caso, os valores de

Figura 4.3: *Raio de giro das nanocages de DNA com com 7A e 7T e 7AT nos linkers. Os dados são apresentados separados para comparar somente os linkers, somente as duplas hélices de cada nanocage e a composição de ambos (linkers + duplas hélices = nanocage). Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].*

RMSF são influenciados pelos movimentos de toda a *nanocage* [7].

Os resultados de RMSF Local, figura 4.4, mostram poucas diferenças entre as flutuações das *nanocages* 7A, 7AT e 7T. Nesta mesma figura, é possível observar que nas extremidades das hélices a flutuação é maior, em torno de 0.3 nm, para ambas as *nanocages*.

5 Nas partes mais centrais das hélices, de ambas as *nanocages*, a flutuação é de cerca de 0.2 nm. Essa diferença nos valores de RMSF para cada parte das hélices indica que, os nucleotídeos mais próximos aos *linkers* apresentam maiores flutuações do que os mais afastados, que flutuam com amplitudes menores [7].

Para o caso dos valores de RMSF Global, figura 4.5, é possível verificar uma pequena,
10 e pouco significativa, diferença entre as flutuações para as *nanocages*. Neste caso,

as flutuações com formas senoidais indicam que as hélices apresentam um movimento rotacional que ocorre em torno do eixo tangencial as bordas externas da hélice, mas que nem sempre é paralelo ao eixo principal da hélice e as vezes é deslocado, ou apresenta algum tipo de inclinação [7]. Grandes flutuações dos nucleotídeos podem indicar que o movimento global das *nanocages* é caracterizado pelo movimento rotacional das duplas hélices [7, 24, 25].

O raio de Giro, RMSD (*Root Mean Square Deviations*) e RMSF (*Root Mean Square Fluctuations*) foram obtidos com uso do programa GROMACS DM *package version* 4.5.3 [115].

Figura 4.4: *RMSF Local* de cada nucleotídeo das duplas hélices de DNA presentes nas *nanocages* 7A (preta), 7T(vermelha) e 7AT (azul). Dados da *nanocage* com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

Figura 4.5: *RMSF Global* de cada nucleotídeo das duplas hélices de DNA presentes nas nanocages 7A (preta), 7T (vermelha) e 7AT (azul). Dados da nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

O grau de torção média sofrida em cada dupla hélice, no próprio eixo, durante toda a trajetória (15-50 ns), foi avaliado com uso do programa Curves+ [116,117] para ambas as nanocages e comparados aos valores obtidos para o valor padrão de B-DNA. Esses valores são mostrados na figura 4.6. Na mesma figura o valor médio para a nanocage de

5 7A está em torno de 22 graus, 26 graus para 7T e 21 graus para 7AT. Esses valores são diferentes do B-DNA, que é próximo de zero. Esse resultado, indica uma maior torção ou curvatura para as duplas hélices na nanocage 7T.

Figura 4.6: Torção (*Bend*) média sofrida em cada dupla hélice das nanocages de 7A, 7AT e 7T e as médias (*Avg*) para cada nanocage comparadas aos valores B-DNA. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

4.3.2 Linkers

Oteri et al. [7] mostraram que a estabilidade de *nanocages*, com diferentes *linkers*, é fortemente influenciada pelo tamanho dos *linkers*. Essa estabilidade, segundo Oteri et al. [7], é em parte justificada pela estabilidade das ligações de hidrogênio nas pontas 5 (*tips*) das duplas hélices de DNA.

Com o objetivo de verificar a estabilidade da *nanocage* com *linkers* de 7A e 7AT foram calculados o número médio de ligações de hidrogênio, figura 4.7, nas extremidades (*tips*) das duplas hélices. Para este caso, o programa VMD foi usado para calcular as ligações de hidrogênio, com um *cutoff* de até 0.35 nm e de até 30 graus.

10 A figura 4.7-A permite observar que na maior parte do tempo ambas as *nanocages*

apresentam o número médio de 3 ligações de hidrogênio nas extremidades das duplas hélices. Esse resultado indica uma certa estabilidade no movimento dos *linkers*, ou ainda, que o movimento dos *linkers* não apresenta uma amplitude grande suficiente para desestabilizar as ligações nas extremidades, das duplas hélices. Da mesma forma,

5 o número médio de ligações de Hidrogênio foi calculado para todos os nucleotídeos das duplas hélices e na figura 4.7-C é possível observar que existe uma certa estabilidade no número de ligações, para todas as hélices em ambas as *nanocages*. A figura 4.7-B mostra uma certa diferença no número de ligações de H, para os *linkers* individualmente em cada tipo *nanocage*, mas na média, o número de ligações de Hidrogênio é quase o

10 mesmo. No caso dos *linkers*, o número de ligações de Hidrogênio pode revelar uma determinada conformação da estrutura octaédrica, durante a trajetória. Um número grande de ligações de H, somados a um número grande de interações do tipo *stacking* poderiam indicar, por exemplo, a evidência de uma redução do valor do Rg inicial da *nanocage*.

15 Inicialmente os nucleotídeos dos *linkers* não estão ligados por ligações de H a uma outra hélice vizinha, e por esse motivo, podem apresentar movimentos espaciais com uma amplitude maior. Esse movimento pode facilitar ligações de H com nucleotídeos mais distantes e/ou ainda permitir interações entre os anéis aromáticos (dos nucleotídeos), por meio de interações do tipo *stacking*. A interação do tipo *stacking*, entre anéis

20 aromáticos, pode acontecer de duas formas distintas: *sandwich* ou *T-shape* [7]. No caso dos nucleotídeos de Adenina, como existem dois anéis aromáticos é esperado que a interação *stacking* do tipo *sandwich* seja igual ou maior se comparada aquelas dos nucleotídeos de Timina. Na figura 4.8-C é possível ver que a interação *stacking* do tipo *sandwich* é praticamente a mesma para ambas as *nanocages*, somente após 30 ns é que

25 aumenta um pouco para a *nanocage* com *linkers* de 7T, figura 4.8-D. A soma do número de interações do tipo *stacking sandwich* e *stacking T-shape* para cada *nanocage*, figura 4.8-A, não evidencia grandes diferenças para o comportamento dos diferentes *linkers*.

Figura 4.7: Número médio de ligações de Hidrogênio entre: (A) as extremidades das duplas hélices, (B) os linkers ou hélices simples, (C) entre os pares de base das duplas hélices ou edges. (D) Síntese dos valores médios dos conjuntos das extremidades das duplas hélices e linkers. Dados da nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

As interações do tipo *stacking* foram calculadas com uso de um programa desenvolvido por Francesco Oteri e Andrea Colleta.

O número de *clusters* ou a porcentagem de agrupamentos estruturais mais recorrentes em cada uma das trajetórias (7A, 7AT e 7T) podem ser visualizados na figuras 4.9, 4.10 e 4.11. Nessas três figuras é possível notar pequenas diferenças entre a porcentagem de *clusters* mais ocupados em cada *nanocage*, mas nenhuma grande diferença que indique mudanças significativas entre as estruturas de cada uma. Assim, nas três figuras, 4.9 e

Figura 4.8: Número ou frequência de interações do tipo stacking entre os anéis aromáticos das bases de Adenina ou Timinas nos linkers para cada nanocage. (A) Número de interações do tipo stacking sandwich e T-shape somadas em cada passo da trajetória. (B) Número de interações do tipo stacking sandwich (C) Número de interações do tipo stacking T-shape. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

4.10, um baixo porcentual de ocupação, por estado conformacional, indica uma grande variabilidade de conformações dos *linkers*, por período de tempo.

Figura 4.9: Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com 7 Adeninas. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm.

4.4 DM: Conclusões Parciais e Perspectivas

As análises da trajetória de DM da *nanocage* com *linkers* de 7A e 7AT não mostraram diferenças significativas comparadas aquela da *nanocage* de *linkers* de 7T avaliada experimentalmente e por DM [7, 24, 25]. O R_g e RMSD avaliados para a *nanocage* de 7A e 7AT não apresentaram grandes flutuações, o que pode indicar que as montagens das *nanocages* com 7A e 7AT podem apresentar estabilidade, indicando a realização experimental. Desta forma, a baixa variação dos parâmetros estudados, evidenciam que a conformação geométrica do sistema, é mais importante para a estabilidade, que as diferentes sequência dos *linkers*.

Figura 4.10: Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com 7 Timinas. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm. Dados da Nanocage com linkers de 7 Timinas foram obtidos por meio da trajetória publicada por Oteri et al. [7].

O parâmetro que apresentou diferenças mais evidentes entre as *nanocages* foi o *Bend* ou o grau de Torção das duplas hélices de DNA. Este parâmetro pode apresentar uma importância maior em trabalhos futuros de encapsulação de partículas. Neste caso, as paredes ou superfícies das *nanocages* não podem ser muito flexíveis a ponto de

5 dobrarem-se permitindo a saída ou entrada de partículas sem o devido controle.

Durante a realização deste trabalho novas dificuldades foram encontradas. Entre elas destaca-se, a necessidade da criação de um programa para a transformação dos modelos

Figura 4.11: *Porcentagem de clusters mais ocupados na trajetória analisada entre 15 ns e 50 ns para o octaedro truncado com ATATATA. Os três clusters mais ocupados são apresentados. Em vermelho o cluster mais ocupado, seguido do verde e o menos ocupado em azul. O método utilizado foi gromos com cutoff de 0.2 nm.*

coarse grain para modelos atomísticos, já que, atualmente utiliza-se o programa Sybil para realizar esse tipo de modelagem. O Sybil tem um custo aproximado de 20 mil euros, por licença, e não tem geometrias previamente definidas, o que torna a modelagem um processo longo e trabalhoso. Com o desenvolvimento do trabalho iniciado no Brasil e

5 na Itália pretende-se modelar além dos modelos *coarse grain*, já disponíveis, também modelos atomísticos por meio de programas desenvolvidos com esta colaboração.

A integração das ferramentas computacionais, realizadas neste trabalho, somada a outras contribuições dos pesquisadores envolvidos, poderão permitir a criação automa-

tizada dos modelos de DNA em nível atômico possibilitando a independência do uso do programa Sybil e uma maior facilidade na criação de modelos de sistemas auto-organizados. O desenvolvimento das técnicas de DM, para sistemas auto-organizados de DNA, poderá contribuir para avaliações prévias dos sistemas planejados, antes mesmo
5 da síntese destes materiais. Por outro lado, também poderão contribuir para auxiliar na compreensão das análises destes nanomateriais.

O estágio realizado no grupo de Biologia Estrutural da Universidade de Roma Tor Vergata, permitiu ao autor deste trabalho um primeiro contato com as técnicas de modelagem, simulação e análise de MD, aplicados à sistemas atomísticos. Este trabalho
10 inicial permite agora avançar na compreensão e utilização dos métodos de MD em outras aplicações.

Os resultados apresentados neste capítulo para intensidade média fazem parte do artigo [118]: Federico Iacovelli, Cassio Alves, Mattia Falconi, Francesco Oteri, Cristiano L. P. de Oliveira, and Alessandro Desideri. Influence of the single-strand linker composition on the structural/dynamical properties of a truncated octahedral DNA nano-cage
15 family. Biopolymers, 101(10):992-999, 2014. doi:10.1002/bip.22475

CAPÍTULO 5

CONCLUSÕES GERAIS E PERSPECTIVAS

A técnica de SAS apresenta a possibilidade de investigar sistemas na escala nanométrica sem muitas dificuldades experimentais, uma vez que é uma técnica relativamente simples. Por outro lado, é uma técnica de baixa resolução, que fornece valores médios,
10 no espaço recíproco, sobre um sistema espalhador. Justamente por apresentar dados do espaço recíproco, o processo de análise pode ser bastante complexo, e desta forma, se existem informações prévias sobre a geometria do sistema espalhador as chances de êxito na compreensão destes sistemas aumentam.

Ao longo de vários anos, inúmeras abordagens têm sido apresentadas para tentar sim-
15 plificar as análises de dados de espalhamento, e dentre todas, destacam-se aquelas que partem diretamente da intensidade espalhada (métodos *ab initio*), para se gerar um modelo de espalhador, e aquelas que calculam a curva de espalhamento com uso de

informações prévias sobre a forma geométrica [10, 15–18, 20, 21]. Uma das grandes dificuldades para realizar a análise de dados de SAS, quando se possui informações prévias dos espalhadores, é realizar o cálculo da intensidade de espalhamento para um dado modelo. No capítulo item 2.2 foi mencionado que a obtenção de formulas analíticas ou
5 semi-analíticas só é possível para geometrias simples. Assim uma das particularidades deste projeto foi apresentar uma abordagem para realizar o processo de modelagem de variadas geometrias, com uso do MEF, a partir de um número muito baixo de parâmetros. Dependendo das especificidades, a construção dos modelos pode ser bastante elaborada, assim para realizar a verificação do modelo criado, é gerado um arquivo no
10 formato pdb, de acordo com *Protein Data Bank* (RCSB PDB, www.rcsb.org) [3], que permite a visualização tridimensional da estrutura em *softwares* gráficos [119, 120].

Uma vez criado o modelo, existe um conjunto de *softwares* e programas (Multidody [121], Crysol [122], Fox [123], Aquasaxs [124], entre outros) que calculam a intensidade espalhada, em uma dimensão. No entanto, de modo a permitir a completa integração
15 entre a modelagem, simulação e ajuste à dados experimentais, tanto para dados unidimensionais quanto bidimensionais, desenvolveu-se no período de doutorado do autor, módulos computacionais para o cálculo de intensidades. Esse conjunto de ferramentas desenvolvidas, a partir das abordagens apresentadas, está unificado em um *software* de nome *Polygen* e constitui um pacote computacional inédito na literatura. Em resumo,
20 adotada uma geometria, o programa gera o modelo utilizando MEF, e permite o cálculo de curvas unidimensionais da intensidade espalhada (modelo aleatoriamente orientado) além de realizar imagens bidimensionais, para um modelo orientado em um direção específica em relação ao feixe incidente.

O Polygen contém um banco de dados (expansível) de cerca 100 diferentes poliedros
25 convexos, além de, permitir realizar a modelagem de um poliedro, a partir de vértices dispostos em uma superfície esférica. Os modelos podem ser compostos com diferentes geometrias e com diferentes formas de preenchimento, podendo ser modelos sólidos, ou

cascas, ou ainda modelos do tipo *coarse grain*, que podem ser construídos com subunidades: alinhadas nas arestas, com DNAs nas arestas ou outro modelo qualquer nas arestas. A metodologia para elaborar os modelos é baseada nos métodos de elementos finitos e de Monte Carlo e utilizam os algoritmos *winged edge* e *gift wrapping* [97,99,100].

5 A simulação das curvas de intensidade, baseadas nos modelos elaborados com MEF foram realizadas com uso da equação de Debye, em suas várias formas e otimizações [20, 51, 52, 58, 59, 109] e pode ser realizada para sistemas orientados ou aleatoriamente orientados, monodispersos e polidispersos. Para o caso de sistemas aleatoriamente orientados, obtêm-se uma curva unidimensional e no caso de sistemas orientados
10 têm-se uma imagem de espalhamento. A imagem de espalhamento pode apresentar um conjunto de informações maiores que as curvas unidimensionais, individuais, e neste caso, espera-se que esta abordagem possa contribuir para a investigação de quaisquer tipos de sistemas experimentais, sejam eles orientados, com correlações espaciais, centro-simétricos e não centro-simétricos. Em ambos os casos, a intensidade espalhada pode
15 ser ajustada contra dados experimentais usando métodos de minimização de dados.

Para o caso de sistemas aleatoriamente orientados é possível vislumbrar que o procedimento de modelagem baseado nos algorítmicos *gift wrapping* [97,99,100] possa ser desenvolvido para torna-se um método do tipo *ab initio* para análise de dados experimentais de SAXS, permitindo a otimização de modelos diretamente contra um conjunto de
20 dados experimentais. No caso de sistemas orientados, a simulação das imagens de espalhamento poderia ser combinada com procedimentos de minimização contra imagens experimentais e em conjunto com um banco de dados, de imagens em diferentes orientações, poderia revelar um modelo com parâmetros estruturais ajustados em uma determinada orientação em relação ao feixe de incidência.

25 O presente trabalho contribui para a caracterização de sistemas moleculares com uso das técnicas de espalhamento, fornecendo ferramentas para modelagem, simulação e análise

destes dados, com foco em estruturas de alta simetria, por meio de uma abordagem rápida e interativa. Assim, dado um sistema do qual se conhece as informações prévias da geometria, é possível calcular modelos, com diferentes estratégias, e a partir deste modelo simular a imagem ou a curva da intensidade espalhada. O ajuste das curvas unidimensionais contra dados experimentais pode revelar os parâmetros do sistema que melhor ajustam os dados. Em outras palavras, este trabalho contribuiu para que a partir de um modelo no espaço real seja possível obter, de um modo bastante simples, dados no espaço recíproco, que quando ajustados a dados experimentais, podem confirmar a forma de uma dada partícula.

A grande totalidade dos trabalhos publicados até a presente data, sobre nanoestruturas de sistemas auto-organizados de DNA, têm aplicado um conjunto restrito de técnicas experimentais para caracterização dos materiais sintetizados. No sentido de unificar um conjunto abrangente de técnicas experimentais, que contribuam na compreensão de parâmetros estruturais de sistemas auto-organizados, alguns trabalhos ganharam destaque [7, 22–25]. Nestes últimos trabalhos, são apresentadas abordagens nas quais os resultados experimentais são conectados aos simulados para uma maior abrangência na descrição físico e bioquímica desses sistemas [7, 22–25].

Neste sentido, o presente trabalho inspirou-se nos desenvolvimentos de sistemas auto-organizados que possuem alta simetria e descobriu que existia uma lacuna científica referente a métodos utilizados na caracterização por experimentos de espalhamento. Todo o desenvolvimento, realizado na tese, foi inspirado em sistemas auto-organizados de DNA, previamente estudados pelo orientador desta tese [7, 23–25], e durante o desenvolvimento do doutorado do autor percebeu-se a possibilidade de unir os desenvolvimentos obtidos para as técnicas de espalhamento com as técnicas de modelagem atomísticas, para fins de simulação com uso de DM. Diante da possibilidade de poder compreender melhor sistemas auto-organizados de DNA, foi realizado um estágio no grupo de Biologia Estrutural, chefiado pelo professor Alessandro Desideri na Universidade de Roma

Tor Vergata. Neste estágio, foram comparadas as trajetórias obtidas com DM de três nanoestruturas de DNA, na forma de um octaedro truncado. O objetivo desta última parte do trabalho, foi compreender os parâmetros que conferem estabilidade a essas nanoestruturas, focados na investigação dos ligantes (*linkers* de DNA ss) entre duplas
5 hélices de DNA, que segundo trabalhos prévios eram os elementos que contribuíam de forma mais significativa. Os resultados obtidos, a partir das análises das trajetórias de DM, para cada nanoestrutura, revelaram que a conformação geométrica do sistema é mais importante para a estabilidade, que as diferentes sequências dos *linkers*, e que essas diferentes sequências não contribuem de forma significativa para a estabilidade.

10 A possibilidade de trabalho com DM apontou novas perspectivas para a continuidade do trabalho proposto desta tese, focadas na modelagem de sistemas e na simulação com DM. Como discutido previamente, os sistemas auto-organizados ainda apresentam um custo de síntese relativamente alto, de maneira que, a simulação de DM pode antecipar perdas de materiais ou pode apontar sucessos de síntese, por meio da escolha
15 de modelos com maior chance de estabilidade. Assim, na oportunidade de simular novos modelos em diferentes geometrias é necessário realizar a modelagem, que não é um processo simples. Diante destas limitações, este trabalho iniciou a extensão do programa *Polygen* para além de, realizar a modelagem *coarse grain*, permitir também a modelagem com precisão atomística. Resultados iniciais já foram obtidos e poderão
20 ser temas de investigação do autor e orientador desta tese no futuro. A modelagem de diferentes modelos atomísticos poderá contribuir para facilitar o uso e planejamento da síntese de sistemas auto-organizados, permitindo a elaboração e testes de sistemas em geometrias mais complexas.

APÊNDICE A

5 BANCO DE DADOS DAS ESTRUTURAS DE ALTA SIMETRIA

Vide referência [98].

0 - tetrahedron 1 - cube 2 - octahedron 3 - dodecahedron 4 - icosahedron 22 - triangular
prism 23 - pentagonal prism 24 - hexagonal prism 25 - octagonal prism 26 - decagonal
10 prism 27 - square antiprism 28 - pentagonal antiprism 29 - hexagonal antiprism 30 -
octagonal antiprism 31 - decagonal antiprism 33 - rhombic dodecahedron 34 - triakis
octahedron 35 - tetrakis hexahedron 36 - trapezoidal icositetrahedron 37 - hexakis oc-
tahedron 38 - pentagonal icositetrahedron (dextro) 39 - rhombic triacontahedron 40 -
triakis icosahedron 41 - pentakis dodecahedron 42 - trapezoidal hexecontahedron 43
15 - hexakis icosahedron 44 - pentagonal hexecontahedron (dextro) 45 - square pyramid
(J1) 46 - pentagonal pyramid (J2) 47 - triangular cupola (J3) 48 - square cupola (J4)
49 - pentagonal cupola (J5) 50 - pentagonal rotunda (J6) 51 - elongated triangular

pyramid (J7) 52 - elongated square pyramid (J8) 53 - elongated pentagonal pyramid
 (J9) 54 - gyroelongated square pyramid (J10) 55 - gyroelongated pentagonal pyramid
 (J11) 56 - triangular dipyramid (J12) 57 - pentagonal dipyramid (J13) 58 - elonga-
 ted triangular dipyramid (J14) 59 - elongated square dipyramid (J15) 60 - elongated
 5 pentagonal dipyramid (J16) 61 - gyroelongated square dipyramid (J17) 62 - elongated
 triangular cupola (J18) 63 - elongated square cupola (J19) 64 - elongated pentagonal
 cupola (J20) 65 - elongated pentagonal rotunds (J21) 66 - gyroelongated triangular
 cupola (J22) 67 - gyroelongated square cupola (J23) 68 - gyroelongated pentagonal
 cupola (J24) 69 - gyroelongated pentagonal rotunda (J25) 70 - gyrobifastigium (J26)
 10 71 - triangular orthobicupola (J27) 72 - square orthobicupola (J28) 73 - square gyrobi-
 cupola (J29) 74 - pentagonal orthobicupola (J30) 75 - pentagonal gyrobicupola (J31)
 76 - pentagonal orthocupolarotunda (J32) 77 - pentagonal gyrocupularotunda (J33)
 78 - pentagonal orthobirotunda (J34) 79 - elongated triangular orthobicupola (J35)
 80 - elongated triangular gyrobicupola (J36) 81 - elongated square gyrobicupola (J37)
 15 82 - elongated pentagonal orthobicupola (J38) 83 - elongated pentagonal gyrobicupola
 (J39) 84 - elongated pentagonal orthocupularotunda (J40) 85 - elongated pentagonal
 gyrocupularotunda (J41) 86 - elongated pentagonal orthobirotunda (J42) 87 - elonga-
 ted pentagonal gyrobirotunda (J43) 88 - gyroelongated triangular bicupola (J44) 89 -
 gyroelongated square bicupola (J45) 90 - gyroelongated pentagonal bicupola (J46) 91
 20 - gyroelongated pentagonal cupularotunda (J47) 92 - gyroelongated pentagonal biro-
 tunda (J48) 93 - augmented triangular prism (J49) 94 - biaugmented triangular prism
 (J50) 95 - triaugmented triangular prism (J51) 96 - augmented pentagonal prism (J52)
 97 - biaugmented pentagonal prism (J53) 98 - augmented hexagonal prism (J54) 99 -
 parabiaugmented hexagonal prism (J55) 100 - metabiaugmented hexagonal prism (J56)
 25 101 - triaugmented hexagonal prism (J57) 102 - augmented dodecahedron (J58) 103 -
 parabiaugmented dodecahedron (J59) 104 - metabiaugmented dodecahedron (J60) 105
 - triaugmented dodecahedron (J61) 106 - metabidiminished icosahedron (J62) 107 -

tridiminished icosahedron (J63) 108 - augmented tridiminished icosahedron (J64) 109
 - augmented truncated tetrahedron (J65) 110 - augmented truncated cube (J66) 111
 - biaugmented truncated cube (J67) 112 - augmented truncated dodecahedron (J68)
 113 - parabiaugmented truncated dodecahedron (J69) 114 - metabiaugmented trunca-
 5 ted dodecahedron (J70) 115 - triaugmented truncated dodecahedron (J71) 116 - gyrate
 rhombicosidodecahedron (J72) 117 - parabigyrate rhombicosidodecahedron (J73) 118 -
 metabigyrate rhombicosidodecahedron (J74) 119 - trigyrate rhombicosidodecahedron
 (J75) 120 - diminished rhombicosidodecahedron (J76) 121 - paragyrate diminished
 rhombicosidodecahedron (J77) 122 - metagyrate diminished rhombicosidodecahedron
 10 (J78) 123 - bigyrate diminished rhombicosidodecahedron (J79) 124 - parabidiminished
 rhombicosidodecahedron (J80) 125 - metabidiminished rhombicosidodecahedron (J81)
 126 - gyrate bidiminished rhombicosidodecahedron (J82) 127 - tridiminished rhombi-
 cosidodecahedron (J83) 128 - snub disphenoid (J84) 129 - snub square antiprism (J85)
 130 - sphenocorona (J86) 131 - augmented sphenocorona (J87) 132 - sphenomegacorona
 15 (J88) 133 - hebesphenomegacorona (J89) 134 - disphenocingulum (J90) 135 - biluna-
 birotunda (J91) 136 - triangular hebesphenorotunda (J92) 137 - tetrahemihexahedron
 138 - octahemioctahedron 139 - small ditrigonal icosidodecahedron 140 - dodecadode-
 cahedron 141 - echidnahedron

APÊNDICE B

SÍNTESE DAS PROPRIEDADES ESTRUTURAIS DO DNA

O DNA, ácido desoxirribonucleico, é um dos dois tipos de ácidos nucleicos. Esses ácidos nucleicos são polímeros de nucleotídeos em que um grupo fosfato liga, por ligações fosfodiester, os açúcares das posições 3' para 5' (C3'para C5') sucessivamente. Em cada molécula de açúcar está ligado uma base nitrogenada, que são por sua vez planares e aromáticas. Essas bases podem ser (A) Adenina, (T) Timina, (C) Citosina ou (G) Guanina e ocupam o núcleo da hélice. A orientação das bases é quase perpendicular à cadeia (*backbone*) de açúcar e fosfato, que envolve as bases perifericamente. Cada uma das bases, no núcleo da hélice, pode *parear-se*, com sua *complementar* por ligações de Hidrogênio. Esse fenômeno de complementaridade acontece para aquelas bases que realizam a mesma quantidade de ligações de Hidrogênio entre si, ou seja, A com T (duas ligações de Hidrogênio) e G com C (três ligações de Hidrogênio) [125].

A conformação estrutural do DNA encontrado na sua forma nativa, ou biologicamente funcional, mais comum, é conhecida como B-DNA ou ainda estrutura de Watson-Crick [125]. Esta estrutura é composta de duas cadeias antiparalelas orientadas no sentido dextrógiro que formam uma dupla hélice de cerca de 20 Å de diâmetro. O

5 passo de repetição desse B-DNA apresenta cerca de 34Å ou 10 pb com ângulo de torção de 36°. As bases por suas vez apresentam-se empilhadas (*stacking*) uma em relação a outra e a modificação das posições no *backbone* não altera a geometria do mesmo. As hélices possuem uma geometria na qual existem dois sulcos expondo diferentes eixos dos pares de base. No sulco menor o eixo exposto é aquele com os átomos C1' e o sulco

10 maior é o eixo oposto [125].

A estabilidade da dupla hélice de DNA é afetada pela quantidade de ligações de Hidrogênio entre as bases, ou seja, a dupla hélice será mais estável se apresentar mais pares de bases do tipo GC ou vice-versa. Dependendo da sequência de pares de bases o B-DNA pode denaturar ou renaturar em uma temperatura maior ou menor.

15 Os fosfatos do *backbone* apresentam uma carga negativa cada e tendem a repelir-se. As bases no centro do DNA são mais hidrofóbicas enquanto que o *backbone* é mais hidrofílico, característica que promove o empilhamento ou *stacking* das bases o centro da dupla hélice.

APÊNDICE C

5

SEQUÊNCIAS NUCLEOTÍDICAS

DH	Nucleotídeos																	
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
1	C	G	A	T	G	T	C	T	A	A	G	C	T	G	A	C	C	G
2	G	G	A	C	C	G	T	G	A	T	T	C	C	A	T	G	A	C
3	C	T	T	A	G	A	G	T	T	G	C	C	A	C	C	A	G	G
4	G	A	A	T	C	C	T	A	T	G	C	T	C	G	G	A	C	G
5	G	G	C	T	C	A	C	A	T	T	G	G	C	T	A	C	A	G
6	C	T	A	T	C	C	G	A	T	C	G	A	G	G	C	A	T	G
7	C	A	T	A	C	T	G	A	G	A	G	C	G	T	T	C	C	G
8	G	T	C	G	C	A	G	T	T	C	A	G	A	T	A	C	G	C
9	C	G	G	T	T	A	C	G	G	T	A	C	A	A	T	G	C	C
10	C	G	C	A	A	G	A	C	G	T	T	A	G	T	G	T	C	C
11	C	C	A	C	C	G	A	A	T	G	G	T	G	T	A	T	C	G
12	G	T	A	T	G	A	C	G	C	A	G	C	A	C	T	T	G	C

Tabela C.1: Sequência de bases em cada par da dupla hélice de DNA

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Cohen-Tannoudji. *Quantum mechanics*, volume II. Wiley-Interscience, 2006. [xi](#), [10](#), [12](#), [13](#), [14](#), [18](#), [21](#), [23](#), [24](#)
- 5 [2] Cassio Alves, Jan Skov Pedersen, and Cristiano Luis Pinto Oliveira. Modelling of high-symmetry nanoscale particles by small-angle scattering. *Journal of Applied Crystallography*, 47:84–94, 2014. [xii](#), [xiii](#), [3](#), [28](#), [54](#), [58](#), [60](#), [64](#), [66](#), [68](#), [70](#), [71](#), [72](#), [73](#), [74](#), [77](#), [78](#), [110](#)
- 10 [3] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, January 2000. 18310. [xvii](#), [36](#), [61](#), [100](#), [101](#), [130](#)
- [4] L. Messori and A. Merlino. Ruthenium metalation of proteins: the X-ray structure of the complex formed between NAMI-A and hen egg white lysozyme. *Dalton Trans*, 43:6128–6131, December 2013. 00000. [xvii](#), [100](#), [101](#)
- 15 [5] C. Hongo, K. Noguchi, K. Okuyama, Y. Tanaka, and N. Nishino. Repetitive interactions observed in the crystal structure of a collagen-model peptide, [(Pro-Pro-Gly)9]3. *J.Biochem.(Tokyo)*, 138:135–144, May 2005. 00030. [xvii](#), [100](#), [101](#)
- [6] K. A. Majorek, P. J. Porebski, A. Dayal, M. D. Zimmerman, K. Jablonska, A. J. Stewart, M. Chruszcz, and W. Minor. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol.Immunol.*, 52:174–182, December 2011. 00000. [xvii](#), [100](#), [101](#)
- 20 [7] Francesco Oteri, Mattia Falconi, Giovanni Chillemi, Felicie F. Andersen, Cristiano L. P. Oliveira, Jan S. Pedersen, Birgitta R. Knudsen, and Alessandro Desideri. Simulative Analysis of a Truncated Octahedral DNA Nanocage Family Indicates the Single-Stranded Thymidine Linkers as the Major Player for the Conformational Variability. *The Journal of Physical Chemistry C*, 115(34):16819–16827, 2011. doi: 10.1021/jp203999x <http://pubs.acs.org/doi/abs/10.1021/jp203999xlink>. [xvii](#),
- 25

- xviii, xix, 6, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 132
- [8] MICHAEL SERINGHAUS and MARK GERSTEIN. Chemistry Nobel Rich in Structure. *Science*, 315(5808):40–41, January 2007. Doi: 10.1126/science.315.5808.40 <http://www.sciencemag.org/content/315/5808/40.full.pdf>link. 1
- [9] Bjoern Sander and Monika M. Golas. Visualization of bionanostructures using transmission electron microscopical techniques. *Microscopy Research and Technique*, 74(7):642–663, 2011. doi: 10.1002/jemt.20963 <http://dx.doi.org/10.1002/jemt.20963link>. 1, 46
- [10] A. Guinier and G. Fournet. *Small angle scattering of X-rays*. Wiley, 1955. 2, 3, 130
- [11] O. Glatter and O. Kratky. *Small-angle X-ray Scattering*. London: Academic Press, 1982. 2
- [12] Paul W Schmidt. Some Fundamental Concepts and Techniques Useful in Small-Angle Scattering Studies of Disordered Solids. In *Modern Aspects of Small Angle Scattering*, volume 451 of *Nato Science Series C*. 1995. 2, 25, 26
- [13] Jan Skov Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares. 1997. 00781. 2
- [14] Gerhard Fritz-Popovski. Two-dimensional indirect Fourier transformation for evaluation of small-angle scattering data of oriented samples. *Journal of Applied Crystallography*, 46(5):1447–1454, 2013. 00001. 2, 3, 80
- [15] P. Lindner and Th Zemb, editors. *Neutrons, X-Rays and Light - Scattering Methods Applied to Soft Condensed Matter*. Elsevier, 2002. 3, 10, 25, 130
- [16] H. B. Stuhrmann. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Crystallographica Section A*, 26(3):297–306, May 1970. 00182. 3, 130
- [17] D. I. Svergun and H. B. Stuhrmann. New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations. *Acta Crystallographica Section A Foundations of Crystallography*, 47(6):736–744, November 1991. 00141. 3, 5, 28, 130
- [18] D. I. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal*, 76(6):2879–2886, 1999. 01166. 3, 5, 28, 130
- [19] B. C. McAlister and B. P. Grady. Simulation of Small-Angle X-ray Scattering from Single-Particle Systems. *Journal of Applied Crystallography*, 31(4):594–599, August 1998. 00013. 3, 80, 81

- [20] Bryan C. McAlister and Brian P. Grady. The use of Monte-Carlo simulations to calculate small-angle scattering patterns. *Macromolecular Symposia*, 190(1):117–130, 2002. 00010. 3, 28, 80, 81, 103, 130, 131
- [21] Daniel P. Olds and Phillip M. Duxbury. Efficient algorithms for calculating small-angle scattering from large model structures. *Journal of Applied Crystallography*, 47(3):1077–1086, 2014. 00000. 3, 36, 80, 106, 130
- [22] Ebbe S. Andersen, Mingdong Dong, Morten M. Nielsen, Kasper Jahn, Ramesh Subramani, Wael Mamdouh, Monika M. Golas, Bjoern Sander, Holger Stark, Cristiano L. P. Oliveira, Jan Skov Pedersen, Victoria Birkedal, Flemming Besenbacher, Kurt V. Gothelf, and Jorgen Kjems. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, 459(7243):73–76, 2009. doi: 10.1038/nature07971 <http://dx.doi.org/10.1038/nature07971link>. 3, 4, 44, 46, 47, 49, 109, 132
- [23] Felicie F. Andersen, Bjarne Knudsen, Cristiano Luis Pinto Oliveira, Rikke F. Frøhlich, Dinna Krüger, Jörg Bungert, Mavis Agbandje-McKenna, Robert McKenna, Sissel Juul, Christopher Veigaard, Jørn Koch, John L. Rubinstein, Bernt Guldbbrandtsen, Marianne S. Hede, Göran Karlsson, Anni H. Andersen, Jan Skov Pedersen, and Birgitta R. Knudsen. Assembly and structural analysis of a covalently closed nano-scale DNA cage. *Nucleic Acids Research*, 36(4):1113–1119, 2008. \doi: 10.1093/nar/gkm1124 <http://nar.oxfordjournals.org/content/36/4/1113.abstractlink>. 4, 5, 28, 44, 109, 132
- [24] Mattia Falconi, Francesco Oteri, Giovanni Chillemi, Felicie F. Andersen, David Tordrup, Cristiano L. P. Oliveira, Jan S. Pedersen, Birgitta R. Knudsen, and Alessandro Desideri. Deciphering the Structural Properties That Confer Stability to a DNA Nanocage. *ACS Nano*, 3(7):1813–1822, 2009. doi: 10.1021/nn900468y <http://pubs.acs.org/doi/abs/10.1021/nn900468ylink>. 4, 6, 48, 113, 114, 115, 116, 119, 125, 132
- [25] Cristiano Luis Pinto Oliveira, Sissel Juul, Hanne L\’a erke Jørgensen, Bjarne Knudsen, David Tordrup, Francesco Oteri, Mattia Falconi, Jørn Koch, Alessandro Desideri, Jan Skov Pedersen, Felicie Faucon Andersen, and Birgitta Ruth Knudsen. Structure of Nanoscale Truncated Octahedral DNA Cages: Variation of Single-Stranded Linker Regions and Influence on Assembly Yields. *ACS Nano*, 4(3):1367–1376, 2010. doi: 10.1021/nn901510v <http://pubs.acs.org/doi/abs/10.1021/nn901510vlink>. 4, 5, 6, 28, 109, 113, 114, 115, 116, 119, 125, 132
- [26] D. K. Saldin, V. L. Shneerson, M. R. Howells, S. Marchesini, H. N. Chapman, M. Bogan, D. Shapiro, R. A. Kirian, U. Weierstall, K. E. Schmidt, and J. C. H. Spence. Structure of a single particle from scattering by many particles randomly oriented about an axis: toward structure solution without crystallization? *New Journal of Physics*, 12(3):035014, March 2010. 00035. 4, 5

- [27] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, July 1999. 01106. [4](#)
- [28] Henry N. Chapman, Anton Barty, Michael J. Bogan, Sébastien Boutet, Matthias Frank, Stefan P. Hau-Riege, Stefano Marchesini, Bruce W. Woods, Saša Bajt, W. Henry Benner, Richard A. London, Elke Plönjes, Marion Kuhlmann, Rolf Treusch, Stefan Düsterer, Thomas Tschentscher, Jochen R. Schneider, Eberhard Spiller, Thomas Möller, Christoph Bostedt, Matthias Hoener, David A. Shapiro, Keith O. Hodgson, David van der Spoel, Florian Burmeister, Magnus Bergh, Carl Coleman, Gösta Huldt, M. Marvin Seibert, Filipe R. N. C. Maia, Richard W. Lee, Abraham Szöke, Nicusor Timneanu, and Janos Hajdu. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics*, 2(12):839–843, December 2006. 00615. [4](#)
- [29] A. Barty, S. Marchesini, H. N. Chapman, C. Cui, M. R. Howells, D. A. Shapiro, A. M. Minor, J. C. H. Spence, U. Weierstall, J. Ilavsky, A. Noy, S. P. Hau-Riege, A. B. Artyukhin, T. Baumann, T. Willey, J. Stolken, T. van Buuren, and J. H. Kinney. Three-Dimensional Coherent X-Ray Diffraction Imaging of a Ceramic Nanofoam: Determination of Structural Deformation Mechanisms. *Physical Review Letters*, 101(5):055501, July 2008. 00066. [4](#)
- [30] Magnus Bergh, Gösta Huldt, Nicusor Timneanu, Filipe R. N. C. Maia, and Janos Hajdu. Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction. *Quarterly Reviews of Biophysics*, 41(3-4):181–204, November 2008. 00057. [4](#)
- [31] K. Schmidt, J. Spence, U. Weierstall, R. Kirian, X. Wang, D. Starodub, H. Chapman, M. Howells, and R. Doak. Tomographic Femtosecond X-Ray Diffractive Imaging. *Physical Review Letters*, 101(11), September 2008. 00020. [4](#)
- [32] Russell Fung, Valentin Shneerson, Dilano K. Saldin, and Abbas Ourmazd. Structure from fleeting illumination of faint spinning objects in flight. *Nature Physics*, 5(1):64–67, January 2009. 000000. [4](#)
- [33] Ne-Te Duane Loh and Veit Elser. Reconstruction algorithm for single-particle diffraction imaging experiments. *Physical Review E*, 80(2):026705, August 2009. 00093. [4](#)
- [34] D K Saldin, V L Shneerson, R Fung, and A Ourmazd. Structure of isolated biomolecules obtained from ultrashort x-ray pulses: exploiting the symmetry of random orientations. *Journal of Physics: Condensed Matter*, 21(13):134014, April 2009. 00042. [4](#)
- [35] M. J. Bogan, S. Boutet, A. Barty, W. H. Benner, M. Frank, L. Lomb, R. Shoeman, D. Starodub, M. M. Seibert, S. P. Hau-Riege, B. Woods, P. Decorwin-Martin, S. Bajt, J. Schulz, U. Rohner, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, J. Hajdu, and H. N. Chapman. Single-shot femtosecond x-ray diffraction from

- randomly oriented ellipsoidal nanoparticles. *Physical Review Special Topics - Accelerators and Beams*, 13(9):094701, September 2010. 00014. [4](#)
- [36] N. D. Loh, M. J. Bogan, V. Elser, A. Barty, S. Boutet, S. Bajt, J. Hajdu, T. Ekeberg, F. R. N. C. Maia, J. Schulz, M. M. Seibert, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, R. L. Shoeman, L. Lomb, M. Frank, M. Liang, and H. N. Chapman. Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns. *Physical Review Letters*, 104(22):225501, June 2010. 00054. [4](#)
- [37] Filipe R. N. C. Maia, Tomas Ekeberg, David van der Spoel, and Janos Hajdu. *Hawk* : the image reconstruction package for coherent X-ray diffractive imaging. *Journal of Applied Crystallography*, 43(6):1535–1539, December 2010. 00000. [4](#)
- [38] Veit Elser. Three-dimensional structure from intensity correlations. *New Journal of Physics*, 13(12):123014, December 2011. 00010. [4](#)
- [39] Henry N. Chapman, Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, Joachim Schulz, Daniel P. DePonte, Uwe Weierstall, R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L. Shoeman, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Peter Holl, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sébastien Boutet, Michael J. Bogan, Jacek Krzywinski, Christoph Bostedt, Saša Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, André Hömke, Christian Reich, Daniel Pietschner, Lothar Strüder, Günter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Kühnel, Marc Messerschmidt, John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu, Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea Rocker, Olof Jönsson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schröter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M. Holton, Thomas R. M. Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Björn Nilsson, and John C. H. Spence. Femtosecond X-ray protein nanocrystallography. *Nature*, 470(7332):73–77, February 2011. 00607. [4](#)
- [40] D. K. Saldin, H. C. Poon, M. J. Bogan, S. Marchesini, D. A. Shapiro, R. A. Kirian, U. Weierstall, and J. C. H. Spence. New Light on Disordered Ensembles: Ab Initio Structure Determination of One Particle from Scattering Fluctuations of Many Copies. *Physical Review Letters*, 106(11), March 2011. 00049. [4](#)
- [41] B. Pedrini, A. Menzel, M. Guizar-Sicairos, V. A. Guzenko, S. Gorelick, C. David, B. D. Patterson, and R. Abela. Two-dimensional structure from random multiparticle X-ray scattering images using cross-correlations. *Nature Communications*, 4:1647, April 2013. 00009.

- [42] Zvi Kam. Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations. *Macromolecules*, 10(5):927–934, 1977. 00073. 4, 110
- [43] J C H Spence, U Weierstall, and H N Chapman. X-ray lasers for structural and dynamic biology. *Reports on Progress in Physics*, 75(10):102601, October 2012. 00068. 4
- [44] P. Chacón, F. Morán, J. F. Díaz, E. Pantos, and J. M. Andreu. Low-Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm. *Biophysical Journal*, 74(6):2760–2775, June 1998. 00000. 5, 28
- [45] D. Walther, F. E. Cohen, and S. Doniach. Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *Journal of Applied Crystallography*, 33(2):350–363, April 2000. 5, 28
- [46] C. R. F. Castro. *Tomografia por Difração de Raios X em Tecidos Biológicos Utilizando Radiação Síncrotron*. PhD thesis, Univerdidade Federal do Rio de Janeiro, 2006. http://www.con.ufrj.br/DScTeses/Carlos/Tese_DSc_Castro.pdflink. 10
- [47] C. L. P. Oliveira. *Estudos de espalhamento de Raios-X a baixos ângulos: teoria e aplicações*. PhD thesis, Universidade Estadual de Campinas, 2005. <http://webbif.ifi.unicamp.br/teses/apresentacao.php?filename=IF385link>. 17, 30
- [48] Feigin L. A and Svergun D. I. *Structure Analysis by small-angle X-ray and neutron scattering*. Plenum Press,, New York, 1987. 18, 31, 32
- [49] Peña Luis de la. *Introducción a la mecánica cuántica*. Fondo de Cultura Económica, México, D.F., 1st. edition edition, December 2006. 00000. 21, 22
- [50] Peter N. Pusey. Introduction to Scattering Experiments. In *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matte*, North-Holland Delta Series, pages 1–21. Elsevier Science, 2002. 00021. 25, 29, 30, 31
- [51] S. Hansen. Calculation of small-angle scattering profiles using Monte Carlo simulation. *Journal of Applied Crystallography*, 23(4):344–346, August 1990. doi:10.1107/S0021889890002801 <http://dx.doi.org/10.1107/S0021889890002801link>. 28, 38, 103, 107, 131
- [52] E. Pantos and J. Bordas. Supercomputer simulation of small angle X-ray scattering, electron micrographs and X-ray diffraction patterns of macromolecular structures. *Pure and Applied Chemistry*, 66(1):77–82, 1994. doi: 10.1351/pac199466010077 <http://www.iupac.org/publications/pac/66/1/0077/pdf>link. 28, 37, 62, 81, 131
- [53] D. I. Svergun, V. V. Volkov, M. B. Kozin, and H. B. Stuhrmann. New Developments in Direct Shape Determination from Small-Angle Scattering. 2. Uniqueness.

- Acta Crystallographica Section A Foundations of Crystallography*, 52(3):419–426, May 1996. 00140. 28
- [54] Vladimir V. Volkov and Dmitri I. Svergun. Uniqueness of *ab initio* shape determination in small-angle scattering. *Journal of Applied Crystallography*, 36(3):860–864, June 2003. 00000. 28
- [55] Maxim V. Petoukhov and Dmitri I. Svergun. Applications of small-angle X-ray scattering to biomacromolecular solutions. *The International Journal of Biochemistry & Cell Biology*, 45(2):429–437, February 2013. 00027. 28
- [56] Pedersen Jan Skov. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Advances in Colloid and Interface Science*, 70(0):171–210, July 1997. doi: 10.1016/S0001-8686(97)00312-6 <http://www.sciencedirect.com/science/article/pii/S0001868697003126link>. 33, 103
- [57] C. L. P. de Oliveira. *SIMULAÇÃO DO ESPALHAMENTO DE RAIOS X POR MACROMOLÉCULAS EM SOLUÇÃO ATRAVÉS DA CONSTRUÇÃO DE MODELOS TRIDIMENSIONAIS DE BAIXA RESOLUÇÃO*. PhD thesis, Universidade Estadual de Campinas, 2001. <http://webbif.ifi.unicamp.br/teses/apresentacao.php?filename=IF1272link>. 36, 104
- [58] O. Glatter. X-Ray Small Angle Scattering of Molecules Composed of Subunits. *Acta Physica Austriaca*, 36:307–315, 1972. 37, 62, 81, 103, 131
- [59] G. Porod. General Theory. In O. Glatter and O. Kratky, editors, *Small-angle X-ray Scattering*, pages 17–52. London: Academic Press, 1982. 37, 62, 81, 103, 131
- [60] Jan Skov Pedersen, Cristiano L.P. Oliveira, Henriette Baun Hübschmann, Lise Arleth, Søren Manniche, Nicolai Kirkby, and Hanne Mørck Nielsen. Structure of Immune Stimulating Complex Matrices and Immune Stimulating Complexes in Suspension Determined by Small-Angle X-Ray Scattering. *Biophysical Journal*, 102(10):2372–2380, May 2012. 00003. 38, 62, 65, 107
- [61] Bo Sjöberg. Small-angle scattering from collections of interacting hard ellipsoids of revolution studied by Monte Carlo simulations and other methods of statistical thermodynamics. *Journal of Applied Crystallography*, 32(5):917–923, 1999. 00000. 38
- [62] Otto Glatter. Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter. North-Holland Delta Series, pages 391–420. Elsevier Science, 2002. 39, 66
- [63] Jan S. Pedersen. Monte Carlo Techniques Applied in the Analysis of Small Angle scattering Data from Colloids and Polymer Systems. In *Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter*, North-Holland Delta

- Series, pages 391–420. Elsevier Science, zemb, th. and lindner, p. edition, 2002. 00000. [39](#), [66](#)
- [64] George M. Whitesides and Bartosz Grzybowski. Self-Assembly at All Scales. *SCIENCE*, 295:2418–2421, 2002. doi: 10.1126/science.1070821
5 <http://www.sciencemag.org/content/295/5564/2418.fulllink>. [40](#), [41](#), [42](#)
- [65] Nadrian C. Seeman. Nanomaterials Based on DNA. *Annu. Rev. Biochem*, 79:65–87, November 2010. doi: 10.1146/annurev-biochem-060308-102244
<http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060308-102244link>. [43](#), [44](#), [45](#), [46](#), [47](#), [54](#), [75](#)
- 10 [66] Junghuei Chen and Nadrian C. Seeman. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature*, 350(6319):631–633, 1991. doi: 10.1038/350631a0 <http://dx.doi.org/10.1038/350631a0link>. [43](#), [44](#)
- [67] Veikko Linko and Hendrik Dietz. The enabled state of DNA nanotechnology. *Current Opinion in Biotechnology*, pages –, 2013. \doi: 10.1016/j.copbio.2013.02.001
15 <http://www.sciencedirect.com/science/article/pii/S0958166913000153>link. [43](#), [44](#)
- [68] Nadrian C. Seeman. Structural DNA Nanotechnology: Growing Along with Nano Letters. *Nano Letters*, 10(6):1971–1978, 2010. doi: 10.1021/nl101262u
<http://pubs.acs.org/doi/abs/10.1021/nl101262ulink>. [43](#), [44](#), [54](#)
- 20 [69] Yuwen Zhang and Nadrian C. Seeman. Construction of a DNA-Truncated Octahedron. *Journal of the American Chemical Society*, 116(5):1661–1669, 1994. doi: 10.1021/ja00084a006 <http://pubs.acs.org/doi/abs/10.1021/ja00084a006link>. [44](#)
- [70] William M. Shih, Joel D. Quispe, and Gerald F. Joyce. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature*, 427(6975):618–621, 2004. doi: 10.1038/nature02307 <http://dx.doi.org/10.1038/nature02307link>.
25 [44](#)
- [71] Yu He, Min Su, Ping-an Fang, Chuan Zhang, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. On the Chirality of Self-Assembled DNA Octahedra. *Angewandte Chemie International Edition*, 49(4):748–751, 2010. doi: 10.1002/anie.200904513
30 <http://dx.doi.org/10.1002/anie.200904513link>. [44](#)
- [72] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, and A. J. Turberfield. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science*, 310(5754):1661–1665, 2005. doi: 10.1126/science.1120367
35 <http://www.sciencemag.org/content/310/5754/1661.abstractlink>. [44](#)
- [73] Yu He, Tao Ye, Min Su, Chuan Zhang, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature*, 452(7184):198–201, 2008. doi: 10.1038/nature06597
40 <http://dx.doi.org/10.1038/nature06597link>. [44](#), [47](#)

- [74] Zhe Li, Bryan Wei, Jeanette Nangreave, Chenxiang Lin, Yan Liu, Yongli Mi, and Hao Yan. A Replicable Tetrahedral Nanostructure Self-Assembled from a Single DNA Strand. *Journal of the American Chemical Society*, 131(36):13093–13098, 2009. doi: 10.1021/ja903768f <http://pubs.acs.org/doi/abs/10.1021/ja903768f>. 44
- [75] Takayuki Kato, Russell P. Goodman, Christoph M. Erben, Andrew J. Turberfield, and Keiichi Namba. High-Resolution Structural Analysis of a DNA Nanostructure by cryoEM. *Nano Letters*, 9(7):2747–2750, 2009. doi: 10.1021/nl901265n <http://pubs.acs.org/doi/abs/10.1021/nl901265n>. 44
- [76] Christoph M. Erben, Russell P. Goodman, and Andrew J. Turberfield. A Self-Assembled DNA Bipyramid. *Journal of the American Chemical Society*, 129(22):6992–6993, 2007. doi: 10.1021/ja071493b <http://pubs.acs.org/doi/abs/10.1021/ja071493b>. 44, 75
- [77] Dhiraj Bhatia, Shabana Mehtab, Ramya Krishnan, Shantinath S. Indi, Atanu Basu, and Yamuna Krishnan. Icosahedral DNA Nanocapsules by Modular Assembly. *Angewandte Chemie International Edition*, 48(23):4134–4137, 2009. doi: 10.1002/anie.200806000 <http://dx.doi.org/10.1002/anie.200806000>. 44, 48, 75
- [78] Chuan Zhang, Min Su, Yu He, Xin Zhao, Ping-an Fang, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. Conformational flexibility facilitates self-assembly of complex DNA nanostructures. *Proceedings of the National Academy of Sciences*, 105(31):10665–10669, 2008. doi: 10.1073/pnas.0803841105 <http://www.pnas.org/content/105/31/10665.abstract>. 44
- [79] Faisal A. Aldaye and Hanadi F. Sleiman. Modular Access to Structurally Switchable 3D Discrete DNA Assemblies. *Journal of the American Chemical Society*, 129(44):13376–13377, 2007. doi: 10.1021/ja075966q <http://pubs.acs.org/doi/abs/10.1021/ja075966q>. 44
- [80] Hua Yang, Christopher K. McLaughlin, Faisal A. Aldaye, Graham D. Hamblin, Andrzej Z. Rys, Isabelle Rouiller, and Hanadi F. Sleiman. Metal–nucleic acid cages. *Nat Chem*, 1(5):390–396, 2009. doi: 10.1038/nchem.290 <http://dx.doi.org/10.1038/nchem.290>. 44, 48, 75
- [81] Jan Zimmermann, Martin P. J. Cebulla, Sven Mönninghoff, and Günter von Kiedrowski. Self-Assembly of a DNA Dodecahedron from 20 Trisigonucleotides with C3h Linkers. *Angewandte Chemie International Edition*, 47(19):3626–3630, 2008. doi: 10.1002/anie.200702682 <http://dx.doi.org/10.1002/anie.200702682>. 44
- [82] Hendrik Dietz, Shawn M. Douglas, and William M. Shih. Folding DNA into Twisted and Curved Nanoscale Shapes. *Science*, 325(5941):725–730, 2009. doi: 10.1126/science.1174251 <http://www.sciencemag.org/content/325/5941/725.abstract>. 44
- [83] Shawn M. Douglas, Hendrik Dietz, Tim Liedl, Bjorn Hogberg, Franziska Graf, and William M. Shih. Self-assembly of DNA into nanoscale three-

- dimensional shapes. *Nature*, 459(7245):414–418, 2009. doi: 10.1038/nature08016 <http://dx.doi.org/10.1038/nature08016link>. 44
- [84] Masayuki Endo, Kumi Hidaka, Takayuki Kato, Keiichi Namba, and Hiroshi Sugiyama. DNA Prism Structures Constructed by Folding of Multiple Rectangular Arms. *Journal of the American Chemical Society*, 131(43):15570–15571, 2009. doi: 10.1021/ja904252e <http://pubs.acs.org/doi/abs/10.1021/ja904252elink>. 44
- [85] Yonggang Ke, Jaswinder Sharma, Minghui Liu, Kasper Jahn, Yan Liu, and Hao Yan. Scaffolded DNA Origami of a DNA Tetrahedron Molecular Container. *Nano Letters*, 9(6):2445–2447, 2009. doi: 10.1021/nl901165f <http://pubs.acs.org/doi/abs/10.1021/nl901165flink>. 44
- [86] Paul W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006. doi: 10.1038/nature04586 <http://dx.doi.org/10.1038/nature04586link>. 44, 46
- [87] Christoph M. Erben, Russell P. Goodman, and Andrew J. Turberfield. Single-Molecule Protein Encapsulation in a Rigid DNA Cage. *Angewandte Chemie International Edition*, 45(44):7414–7417, 2006. doi:10.1002/anie.200603392 <http://dx.doi.org/10.1002/anie.200603392link>. 44, 49, 75
- [88] Dongran Han, Suchetan Pal, Jeanette Nangreave, Zhengtao Deng, Yan Liu, and Hao Yan. DNA Origami with Complex Curvatures in Three-Dimensional Space. *Science*, 332(6027):342–346, 2011. doi: 10.1126/science.1202998 <http://www.sciencemag.org/content/332/6027/342.abstractlink>. 44, 45, 54
- [89] Dongran Han, Suchetan Pal, Yang Yang, Shuoxing Jiang, Jeanette Nangreave, Yan Liu, and Hao Yan. DNA Gridiron Nanostructures Based on Four-Arm Junctions. *Science*, 339(6126):1412–1415, 2013. doi: 10.1126/science.1202998 <http://www.sciencemag.org/content/339/6126/1412.abstractlink>. 44, 45
- [90] Russell P. Goodman, Mike Heilemann, Soren Doose, Christoph M. Erben, Achillefs N. Kapanidis, and Andrew J. Turberfield. Reconfigurable, braced, three dimensional DNA nanostructures. *Nat Nano*, 3(2):93–96, 2008. doi:10.1038/nnano.2008.3 <http://dx.doi.org/10.1038/nnano.2008.3link>. 44, 47
- [91] Anthony S. Walsh, HaiFang Yin, Christoph M. Erben, Matthew J. A. Wood, and Andrew J. Turberfield. DNA Cage Delivery to Mammalian Cells. *ACS Nano*, 5(7):5427–5432, 2011. doi: 10.1021/nn2005574 <http://pubs.acs.org/doi/abs/10.1021/nn2005574link>. 49, 75
- [92] A. R. Leach. *Molecular Modelling Principles and Applications*. Pearson Education Limited, 2nd edition, 2001. 50
- [93] M. Bitar and M. Offman. Molecular Dynamics. Technical report, Technische Universität München, Department for Bioinformatics and Computational Biology, München. 50, 51, 52

- [94] Michael P. Allen. *Introduction to Molecular Dynamics Simulation*, volume 23. Norbert Attig, Kurt Binder, Helmut Grubmüller, Kurt Kremer, Jülich, NIC series edition, 2004. 50, 51, 52
- 5 [95] J. Zhou, A. Deyhim, S. Krueger, and S.K. Gregurick. LORES: Low resolution shape program for the calculation of small angle scattering profiles for biological macromolecules in solution. *Computer Physics Communications*, 170(2):186–204, August 2005. 55
- [96] Osvaldo Dolce and José Nicolau Pompeo. *Geometria Espacial Posição e Métrica*, volume 10 of *Fundamentos de Matemática Elementar*. Atual, 5 edition, 2005. 55
- 10 [97] Franco P. Preparata; and Michael Ian Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 2 edition, 1988. 56, 57, 60, 131
- [98] *Netlib - Polyhedra*. November 2011. Published: 12h dia 15 <http://www.netlib.org/polyhedra/http://www.netlib.org/polyhedra/>. 57, 134
- 15 [99] Donald R. Chand and Sham S. Kapur. An Algorithm for Convex Polytopes. *J. ACM*, 17(1):78–86, January 1970. 00317. 60, 131
- [100] R. A. Jarvis. On the identification of the convex hull of a finite set of points in the plane. *Information Processing Letters*, 2(1):18–21, March 1973. 00448. 60, 131
- 20 [101] M. Uchida, M. T. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. O. Liepold, M. Young, and T. Douglas. Biological Containers: Protein Cages as Multifunctional Nanoplatfoms. *Advanced Materials*, 19(8):1025–1042, 2007. DOI: 10.1002/adma.200601168 <http://onlinelibrary.wiley.com/doi/10.1002/adma.200601168/abstractlink>. 75
- 25 [102] Seidel S. Russell and Stang Peter J. High-Symmetry Coordination Cages via Self-Assembly. *Accounts of Chemical Research*, 35(11):972–983, 2002. doi: 10.1021/ar010142d <http://pubs.acs.org/doi/abs/10.1021/ar010142dlink>. 75
- 30 [103] John J. Perry, Jason A. Perman, and Michael J. Zaworotko. Design and synthesis of metal–organic frameworks using metal–organic polyhedra as supermolecular building blocks. *Chemical Society Reviews*, 38:1400–1417, 2009. doi: 10.1039/b807086p <http://pubs.rsc.org/en/content/articlelanding/2009/cs/b807086p#!divAbstractlink>. 75
- [104] Trevor Douglas and Mark Young. Viruses: Making Friends with Old Foes. *Science*, 312(5775):873–875, 2006. 75
- 35 [105] Tamara D. Hamilton and Leonard R. MacGillivray. Enclosed Chiral Environments from Self-Assembled Metal–Organic Polyhedra. *Crystal Growth & Design*, 4(3):419–430, 2004. 75

- [106] Andrew Stephenson, Stephen P. Argent, Thomas Riis-Johannessen, Ian S. Tidmarsh, and Michael D. Ward. Structures and Dynamic Behavior of Large Polyhedral Coordination Cages: An Unusual Cage-to-Cage Interconversion. *Journal of the American Chemical Society*, 133(4):858–870, 2011. 75
- 5 [107] Amy Szuchmacher Blum, Carissa M. Soto, Charmaine D. Wilson, John D. Cole, Moon Kim, Bruce Gnade, Anju Chatterji, Wendy F. Ochoa, Tianwei Lin, John E. Johnson, and B. R. Ratna. *Nano Letters*, 4(5):867–870, 2004. doi: 10.1021/nl0497474. <http://pubs.acs.org/doi/abs/10.1021/nl0497474link>. 75
- 10 [108] S. Banerjee, K. Loza, W. Meyer-Zaika, O. Prymak, and M. Epple. Structural Evolution of Silver Nanoparticles during Wet-Chemical Synthesis. *Chemistry of Materials*, 26(2):951–957, 2014. 97
- [109] S. J. Henderson. Monte Carlo modeling of small-angle scattering data from non-interacting homogeneous and heterogeneous particles in solution. *Biophysical Journal*, 70(4):1618–1627, April 1996. 00037. 103, 131
- 15 [110] Sharon L. Lohr. *Sampling: Design and Analysis*. Cengage Learning, Boston, Mass, 2 edition edition, December 2009. 107
- [111] T. A. Darden Case, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, Luo R. C., Walker, W. Zhang K. M., Merz B., Roberts S., and Hayik A. *AMBER 12*. University of California, San Francisco, 2012. 114
- 20 [112] Alberto Perez, Ivan Marchan, Daniel Svozil, Jiri Sponer, Thomas E. Cheatham III, Charles A. Laughton, and Modesto Orozco. Refinement of the {AMBER} Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical Journal*, 92(11):3817–3829, 2007. doi: 10.1529/biophysj.106.097782 <http://www.sciencedirect.com/science/article/pii/S0006349507711827link>. 114
- 25 [113] Thomas E. Cheatham, Piotr Cieplak, and Peter A. Kollman. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *Journal of Biomolecular Structure and Dynamics*, 16(4):845–862, 1999. doi: 10.1080/07391102.1999.10508297 <http://www.tandfonline.com/doi/abs/10.1080/07391102.1999.10508297link>. 114
- 30 [114] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983. doi:10.1063/1.445869 <http://link.aip.org/link/?JCP/79/926/1link>. 115
- 35 [115] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008. doi:10.1021/ct700301q <http://pubs.acs.org/doi/abs/10.1021/ct700301qlink>. 119

- [116] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research*, 37(17):5917–5929, September 2009. 00213. [120](#)
- 5 [117] Christophe Blanchet, Marco Pasi, Krystyna Zakrzewska, and Richard Lavery. CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Research*, 39(suppl 2):W68–W73, July 2011. 00023. [120](#)
- 10 [118] Federico Iacovelli, Cassio Alves, Mattia Falconi, Francesco Oteri, Cristiano L. P. de Oliveira, and Alessandro Desideri. Influence of the single-strand linker composition on the structural/dynamical properties of a truncated octahedral DNA nano-cage family: Single-Strand Linkers on Three Truncated Octahedral DNA Nano-Cages. *Biopolymers*, 101(10):992–999, October 2014. 00000. [128](#)
- 15 [119] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004. [130](#)
- [120] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996. [130](#)
- 20 [121] O. Glatter. Computation of distance distribution functions and scattering functions of models for small-angle scattering experiments. *Acta Phys. Austriaca*, 52:243–256, 1980. [130](#)
- [122] D. Svergun, C. Barberato, and M. H. J. Koch. *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, 28(6):768–773, December 1995. [130](#)
- 25 [123] Dina Schneidman-Duhovny, Michal Hammel, and Andrej Sali. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research*, 38(suppl 2):W540–W544, July 2010. [130](#)
- [124] Frédéric Poitevin, Henri Orland, Sebastian Doniach, Patrice Koehl, and Marc Delarue. AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Research*, 39(suppl 2):W184–W189, July 2011. [130](#)
- 30 [125] Donald Voet and Judith Voet. *Biochemistry*. JOHN WILEY & SONS , INC., 4 edition, 2010. [137](#), [138](#)