

Universidade de São Paulo
Instituto de Física

Análise de interações proteicas via SAXS e simulações de Monte Carlo

Fernando Takeshi Tanouye

Orientador(a): Prof.(a) Dr.(a)



Rosângela Itri

Tese de doutorado apresentada ao Instituto de Física
da Universidade de São Paulo, como requisito parcial
para a obtenção do título de Doutor(a) em Ciências.

Banca Examinadora:

Prof(a). Dr(a). Rosângela Itri (Orientadora) - IFUSP

Prof. Dr. Leandro Ramos Souza Barbosa - IFUSP

Prof. Dr. Fernando Luis Barroso da Silva - FFCLRP USP

Prof. Dr. Lus Fernando Mercier Franco - UNICAMP

Prof. Dr. Vitor Barbanti Pereira Leite - UNESP

São Paulo
2022

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

Tanouye, Fernando Takeshi

Análise de interações proteicas via SAXS e simulações de Monte Carlo. São Paulo, 2022.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Física. Depto. de Física Aplicada.

Orientador(a): Profa. Dra. Rosangela Itri.

Área de Concentração: Biofísica.

Unitermos: 1. Física computacional; 2. Física experimental; 3. Biofísica.

USP/IF/SBI-65/2022

University of São Paulo
Physics Institute

Analysis of protein interactions via SAXS and Monte Carlo simulations

Fernando Takeshi Tanouye

Supervisor: Prof.(a) Dr.(a) Rosangela Itri

Thesis submitted to the Physics Institute of the University of So Paulo in partial fulfillment of the requirements for the degree of Doctor of Science.

Examining Committee:

Prof.(a) Dr.(a) Rosangela Itri (Supervisor) - IFUSP

Prof. Dr. Leandro Ramos Souza Barbosa - IFUSP

Prof. Dr. Fernando Luis Barroso da Silva - FFCLRP USP

Prof. Dr. Lus Fernando Mercier Franco - UNICAMP

Prof. Dr. Vitor Barbanti Pereira Leite - UNESP

São Paulo
2022

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Esta pesquisa utilizou recursos do Laboratório Nacional de Luz Síncrotron (LNLS), uma instalação nacional aberta do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) operada pelo Centro Nacional de Pesquisa em Energia e Materiais (CNPEM). Agradecemos a equipe da Linha de Luz SAXS1 pela assistência durante os experimentos.

Meus agradecimentos ao corpo de funcionários do IFUSP, pelo apoio burocrático e demais suportes institucionais durante a realização deste trabalho. Em especial, pelo acolhimento e convivência com as secretárias do Departamento de Física Aplicada, nossas vizinhas de sala, Lia e Elza.

Agradeço à profa. Dra. Rosangela Itri por me receber em seu grupo de pesquisa, pela orientação acadêmica e por todos os ensinamentos transmitidos ao longo dos últimos anos, através dos quais pude me especializar na técnica de espalhamento de raios-X a baixos ângulos. Sou grato também aos nossos colaboradores prof. Dr. Francesco Spinozzi e prof. Dr. Jozismar Rodrigues Alves. O primeiro por ceder gentilmente parte dos dados de espalhamento utilizados nessa pesquisa, além das discussões e aprendizagens na área de SAXS. O segundo, amigo e interlocutor de longa data, pelo trabalho em conjunto na elaboração dos códigos de Monte Carlo e pelas discussões sempre cheias de ideias.

Aos colegas de grupo, com os quais dividimos tantas conversas, conhecimentos, risadas e experiências: Maressa, Pedro, Gustavo, Ju, Tay, Rafaela, Angela e tantos outros que já passaram ou estão chegando, meus agradecimentos. E por fim, mas não menos importante, ao apoio constante e incondicional dos meus familiares.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

This research used resources of the Brazilian Synchrotron Light Laboratory (LNLS), an open national facility operated by the Brazilian Centre for Research in Energy and Materials (CNPEM) for the Brazilian Ministry for Science, Technology, Innovations and Communications (MCTIC). The SAXS1 beamline staff is acknowledged for the assistance during the experiments.

Resumo

Proteínas são polímeros de aminoácidos envolvidos em praticamente todas as funções metabólicas dos seres vivos. Em seu contexto natural de atuação, ou mesmo em aplicações biotecnológicas, elas frequentemente se encontram imersas em ambientes altamente concentrados - no meio intracelular, por exemplo, há cerca de 50-400 g/L de proteínas totais. Nessas condições, o estudo das interações proteína-proteína é fundamental para a compreensão dos fatores que determinam a sua estabilidade.

Neste projeto, propomos a realização de experimentos de espalhamento de raios-X a baixos ângulos (SAXS) para o estudo das interações proteicas em sistemas simples e misturas de lisozima e albumina de soro bovino (BSA), duas proteínas modelo bastante conhecidas na literatura. Esta técnica nos permitiu, juntamente com uma abordagem teórica para os potenciais de interação, avaliar os efeitos das mudanças de concentração, temperatura e força iônica nas amostras. A análise teve como foco a modelagem do fator de estrutura por meio do potencial Hard-Sphere Double-Yukawa (HSDY) sob a relação de fechamento RPA.

Outra abordagem utilizada foi a realização de simulações de Monte Carlo, nas quais representamos as proteínas por meio de esferas duras carregadas (modelo primitivo). Num primeiro momento, realizamos simulações para testar a validade e a coerência do modelo HSDY/RPA empregado nos ajustes, o qual apresentou diversas limitações. Em seguida, propomos um novo método para o ajuste das curvas experimentais por meio de simulações do fator de estrutura com Monte Carlo. Este método se mostrou bastante viável e capaz de modelar as curvas de espalhamento, tanto das soluções simples quanto das misturas binárias, além de evitar os erros de aproximação decorrentes das relações de fechamento. Assim, pudemos descrever o comportamento das proteínas e obter informações complementares que não seriam acessíveis somente pela técnica de SAXS.

Palavras-chave: proteínas, interação, SAXS, Monte Carlo

Abstract

Proteins are polymers of amino acids involved in practically all the metabolic functions of living beings. In their natural context of action, or even in biotechnological applications, they are often immersed in highly concentrated environments - in the intracellular environment, for example, there are about 50-400 g/L of total proteins. Under these conditions, the study of protein-protein interactions is essential to understand the factors that determine its stability.

In this project, we propose to carry out small-angle X-ray scattering (SAXS) experiments to study protein interactions in simple systems and mixtures of lysozyme and bovine serum albumin (BSA), two model proteins well-known in the literature. This technique allowed us, together with a theoretical approach to the interaction potential, to evaluate the effects of changes in concentration, temperature and ionic strength in the samples. The analysis focused on the modeling of the structure factor through the Hard-Sphere Double-Yukawa (HSDY) potential under the RPA closure relation.

Another approach used here was to perform Monte Carlo simulations, in which we represent proteins by means of charged hard-spheres (primitive model). At first, we performed simulations to test the validity and coherence of the HSDY/RPA model used in the adjustments, which presented several limitations. Then, we proposed a new method for fitting the experimental curves through the simulation of the structure factors with Monte Carlo. This method proved to be quite viable and capable of modeling the scattering curves, both for simple solutions and for binary mixtures, in addition to avoiding the approximation errors resulting from the closure relations. Thus, we were able to describe the behavior of the proteins and obtain complementary information that would not be accessible only by the SAXS technique.

Keywords: proteins, interaction, SAXS, Monte Carlo

Lista de Abreviações

β $1/k_bT$

BSA Bovine Serum Albumin

DLVO Derjaguin-Landau-Verwey-Overbeek

HS Hard-Sphere

HSDY Hard-Sphere Double Yukawa

LYS Lysozyme

MC Monte Carlo

PM Primitive Model

PME Particle Mesh Ewald

PY Percus-Yevick

RPA Random Phase Approximation

RPM Restrict Primitive Model

SAXS Small-Angle X-Ray Scattering

Sumário

1	Introdução	1
1.1	Proteínas	1
1.2	Lisozima	4
1.3	Albumina sérica bovina (BSA)	5
1.4	Interações proteicas	6
1.5	Objetivos	8
2	Experimentos de SAXS	9
2.1	Medidas no Laboratório Nacional de Luz Síncroton (LNLS)	9
2.2	Medidas no Diamond Light Source	10
2.3	Software de ajuste: GENFIT	11
3	Teoria de SAXS	13
3.1	Equação Geral	13
3.2	Fator de Forma	18
3.3	Fator de Estrutura	28
3.4	Relações de Fechamento	33
4	Metodologia: Monte Carlo	39
1	Modelagem computacional	39
1.1	O modelo primitivo	39
1.2	Simulação	41
2	Cadeia de Markov e o algoritmo de Metropolis	42
3	Soma de Ewald	44
4	Implementação do algoritmo de MC	47
4.1	Truncagem da soma de Ewald	48
4.2	Movimentos	50
4.3	Validação do código	52
5	Análise das Proteínas Puras: Resultados e Discussão	59
1	Lisozima	59

1.1	Efeitos do sal e da temperatura	59
1.2	Efeitos de concentração	70
1.3	Análises com Monte Carlo	74
2	BSA	87
2.1	Efeito da temperatura	87
2.2	Efeito da concentração	89
2.3	Análise com Monte Carlo	92
6	Análise das Misturas Lisozima-BSA: Resultados e Discussão	97
1	Metodologia	97
1.1	Intensidade de espalhamento de misturas	97
1.2	Monte Carlo das misturas	98
2	Resultados e Discussões	102
2.1	SAXS das misturas	102
2.2	Análise de Monte Carlo das misturas	103
7	Conclusão	125
	Bibliografia	127
	Apêndice A	135
	Apêndice B	139

Introdução

1.1 Proteínas

As proteínas são macromoléculas biológicas constituídas de uma ou algumas cadeias longas de aminoácidos chamadas polipeptídeos. Essas cadeias podem assumir certas conformações espaciais próprias para desempenhar diversas atividades nos organismos vivos, tais como conferir suporte estrutural às células, transportar substâncias, sinalizar eventos, catalizar reações químicas, atuar como anticorpos etc [1]. Muitas dessas funções envolvem a atuação conjunta de vários tipos de proteínas e a formação de grandes complexos moleculares.

Como se sabe, a sequência de uma cadeia polipeptídica é produzida a partir de uma molécula de RNA num processo conhecido como tradução, e que a informação contida nesse RNA está, por sua vez, codificada no DNA (conforme o dogma central da biologia molecular [2]). Já a relação entre a função de uma proteína e a sua sequência específica de aminoácidos é menos clara, mas é amplamente admitido que a função é uma consequência direta da estrutura tridimensional da proteína e que esta estrutura está de alguma forma, ainda obscura, pré-determinada pela sequência de aminoácidos [3].

Em termos estruturais as proteínas são descritas de uma maneira hierarquizada. Chama-se estrutura primária a própria sequência de aminoácidos de uma cadeia. Um aminoácido é uma molécula formada por um carbono central (C_α) ligado a um hidrogênio, um grupo amina, um grupo carboxila e um grupo variável que distingue o tipo do aminoácido, sendo que há 20 tipos que ocorrem naturalmente com maior frequência. Quando dois aminoácidos se ligam durante a formação da cadeia o grupo amina do primeiro interage com o grupo carboxílico do segundo resultando numa ligação amídica, ou peptídica (fig. 1.1). Uma cadeia completa pode então ser representada como uma sequência do tipo DTHKSEIAH..., onde D representa um aminoácido aspartato, T uma treonina, H uma histidina e assim por diante.

O próximo nível hierárquico engloba as estruturas ditas secundárias, que são arranjos espaciais locais de um ou mais fragmentos da sequência. As estruturas secundárias mais

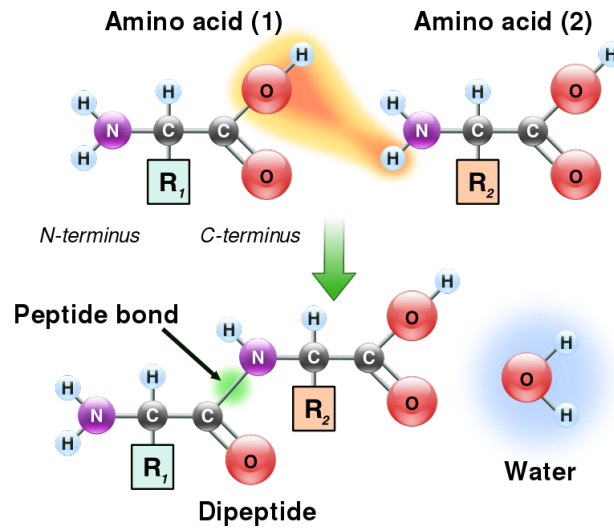


Figura 1.1: Formação de uma ligação peptídica. Imagem licenciada sob domínio público.

recorrentes são as hélices- α e as folhas- β (fig. 1.2). A primeira delas, como o nome sugere, é formada quando uma parte da sequência se dobra em formato helicoidal, de tal modo que cada amina da hélice faz uma ligação de hidrogênio com a carbonila do 4º aminoácido na sequência. Já a segunda se forma quando alguns fragmentos se alinham de maneira paralela ou antiparalela, dependendo do sentido de cada fragmento, constituindo uma rede de ligações de hidrogênio mais ou menos planar. A combinação de algumas estruturas secundárias também pode dar origem a arranjos regulares maiores, ditos *motivos*, como o barril beta, o β -hairpin, a chave-grega, o dedo de zinco, a hélice-alça-hélice, para citar alguns. São elementos intermediários entre os níveis secundário e terciário.

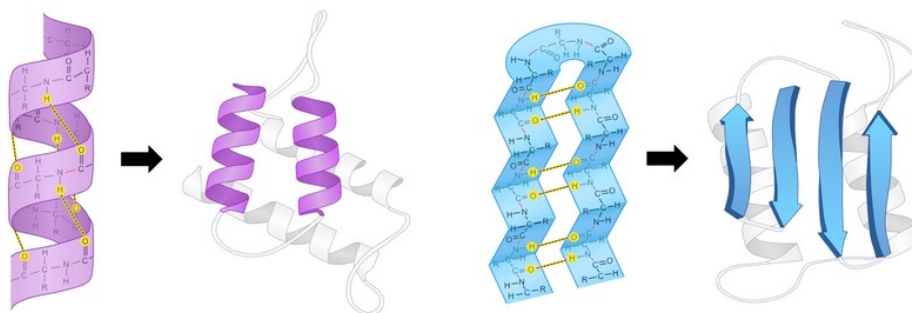


Figura 1.2: Estruturas secundárias de hélice- α e folha- β . Imagem licenciada sob domínio público.

A estrutura terciária diz respeito à conformação espacial global de uma cadeia, compreendendo os elementos de estrutura secundária e as porções mais desordenadas, geralmente resultando numa forma compacta e funcional. O núcleo de proteínas globulares é majoritariamente composto de aminoácidos hidrofóbicos, enquanto que a superfície contém resíduos polares e carregados em maior quantidade. Muitas interações intramo-

leculares e também envolvendo moléculas do solvente aquoso cooperam para estabilizar essa estrutura, como o efeito hidrofóbico, pontes salinas, ligações dissulfeto e de hidrogênio, as quais ajudam a manter o delicado balanceamento energético e entrópico estável (segundo a *hipótese termodinâmica* [4], $\Delta G < 0$ entre os estados enovelado e desenovelado). Quando dois ou mais polipeptídeos enovelados se juntam para formar um complexo funcional maior, temos então uma estrutura dita quaternária, as quais se diferenciam pela maneira como as subunidades proteicas estão conectadas entre si. O complexo de chaperonas GroEL-GroES-(ADP)₇, por exemplo, é uma estrutura quaternária formada por três anéis heptaméricos: dois deles alinhados lado a lado e compostos por um tipo de chaperona (GroEL), e o outro, composto por chaperonas menores (GroES), formando uma espécie de tampa para os primeiros [5], fig. 1.3.

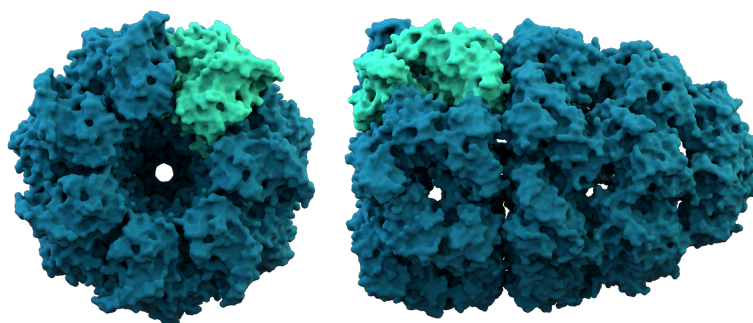


Figura 1.3: Estrutura quaternária do complexo de chaperonas GroEL-GroES-(ADP)₇, conforme a entrada do PDB 1AON. Uma subunidade de proteína GroEL está destacada em verde claro. By Thomas Splettstoesser - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=29509596>.

A tarefa de elucidar a estrutura terciária ou quaternária de proteínas constituiu um intenso campo de pesquisa desde o final da década de 50, quando Perutz e Kendrew resolveram as primeiras estruturas conhecidas da hemoglobina e da mioglobina, respectivamente, por difração de raios-X. Desde então várias técnicas novas surgiram ou foram aprimoradas com o objetivo de realizar estudos estruturais, como a espectroscopia de ressonância magnética (RMN) ou, mais recentemente, a criomicroscopia eletrônica (cryo-EM). Atualmente, em 2020, há cerca de 150.000 estruturas proteicas resolvidas depositadas na base de dados PDB, *Protein Data Bank*. Uma vez conhecida uma estrutura novas possibilidades de investigação se abrem, em particular nos campos de dinâmica molecular, produção de fármacos, bioinformática dentre outros.

Neste trabalho, faremos uso de duas proteínas modelo bastante estudadas na literatura: a lisozima e a albumina de soro bovino (BSA). Suas estruturas cristalográficas serão parte essencial na análise dos dados de espalhamento de raios-X a baixos ângulos (SAXS), como será visto mais adiante.

1.2 Lisozima

O nome lisozima foi cunhado em 1922 pelo microbiologista Alexander Fleming (1881-1955), para se referir a uma substância com propriedades antibacterianas que ele supôs estar presente em diversas secreções e tecidos humanos [6]. Segundo Fleming, ele notou pela primeira vez este efeito quando realizava um estudo onde analisava, em uma placa de cultivo, a secreção nasal de um paciente que sofria de coriza aguda e observou que ocorria uma inibição do crescimento de certas bactérias na placa. Antes de Fleming, porém, pelo menos outro pesquisador, Laschtschenko, já havia relatado em 1909 [7] o efeito bactericida da clara do ovo de galinha, que ocorre devido à presença de lisozima.

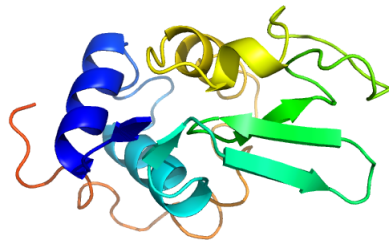


Figura 1.4: Estrutura cristalográfica da Lisozima de clara de ovo (PDB 6LYZ). Visualização tipo cartoon gerada pelo software PyMol.

Essa proteína, também conhecida como muramidase, é uma enzima da família das glicosídeo hidrolases e cataliza a hidrólise das ligações β -1,4 entre os resíduos de ácido N-acetilmurâmico e de N-acetil-D-glucosamina, que fazem parte dos peptidoglicanos que formam a parede celular de bactérias Gram-positivas. Assim, ela promove a lise da membrana desse tipo de bactéria protegendo o organismo de infecções, o que a torna parte do sistema de defesa dos animais. A lisozima pode ser encontrada em tecidos e secreções, como lágrima e saliva, mas uma grande fonte natural é a clara do ovo de galinha, compondo cerca de 3-4% das proteínas totais. Por ser tão abundante, muito fácil de purificar (a partir da clara do ovo) e de cristalizar, ela é uma das melhores proteínas para a realização de ensaios cristalográficos. Sua estrutura de difração foi determinada pela primeira vez em 1965 por David Chilton Phillips (1924-1999) e colaboradores, com uma resolução à época de 2 Å [8].

Uma lisozima típica contém aproximadamente 129 aminoácidos, seu peso molecular vale cerca de 14.4 kDa e possui um ponto isoelétrico $pI \approx 11$. Sua elevada estabilidade térmica pode chegar até uma temperatura de *melting* de 72 °C em pH 5 [9]. Uma representação da estrutura da lisozima pode ser conferida na fig. 1.4, acima, na qual se distingue quatro hélices- α e três fitas- β antiparalelas.

1.3 Albumina sérica bovina (BSA)

Albumina, em sentido amplo, é o nome genérico de proteínas solúveis em água presentes na clara do ovo (do latim, *albumen ovi*, e *albus*: branco), no leite, sangue e outros fluidos animais e vegetais. Em sentido restrito, diz respeito a uma família de proteínas presentes no plasma sanguíneo, as albuminas séricas. Essas proteínas, sintetizadas pelo fígado, são responsáveis pela manutenção da pressão osmótica coloidal do sangue e por transportar diversas substâncias como cátions (Ca^+ , K^+ , Na^+), ácidos graxos, bilirrubina, mas também alguns hormônios e fármacos. Em humanos, a HSA (Human Serum Albumin) é encontrada em concentrações de referência na faixa de 33-52 mg/ml [10].

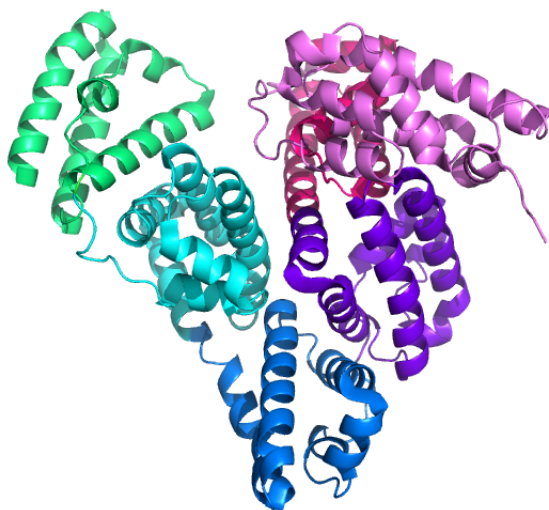


Figura 1.5: Estrutura cristalográfica da BSA (PDB 4F5S). Os subdomínios estão coloridos como segue: IA/IB, rosa claro/rosa escuro; IIA/IIB, roxo/azul; IIIA/IIIB, ciano/verde. Visualização tipo cartoon gerada pelo software PyMol.

A albumina de soro bovino (BSA) é composta de 583 aminoácidos e possui uma identidade de sequência de 75.8% com a HSA [11], além de estruturas muito parecidas. Seu peso molecular fica em torno de 66.4 kDa e o ponto isoelétrico (pI) entre 4.7-5.6, dependendo do meio. Em pH fisiológico, 7.4, ela apresenta uma elevada carga negativa, em razão da grande quantidade de aminoácidos ácidos, o que favorece a ligação de cátions em sua superfície.

A estrutura cristalográfica da HSA foi determinada em 1999 por Kobayashi et al. [12] e a da BSA (juntamente com as da ESA e LSA) por Anna Bujacz em 2012 [11], fig. 1.5. A BSA possui três domínios (I, II e III), cada um deles com dois subdomínios (A e B), que se juntam em um formato de coração. A maior parte dos seus elementos secundários é composta por 33 α -hélices. A presença de 17 pontes dissulfeto (ligações covalentes entre resíduos de cisteína), em posições altamente conservadas, tornam a BSA uma proteína bastante estável. O único resíduo de cisteína não pareado ocorre na

posição 34 e, eventualmente, promove a formação de um dímero ao se ligar à cisteína 34 de outra BSA [11].

Sua estrutura também pode variar dependendo das condições físico-químicas do meio. Conforme o pH diminui, por exemplo de 8.0 a 2.7, ela pode realizar transições entre três estados sucessivos: N (normal) \rightarrow F (fast) \rightarrow E (expanded), segundo estudos de dinâmica molecular [13, 14, 15]. No estado N a BSA está em sua conformação típica de coração, no F os domínios I e III se afastam deixando-a parcialmente estendida e no E a cadeia já se encontra totalmente estendida, embora preservando um certo conteúdo de estruturas secundárias. A exposição de resíduos hidrofóbicos provocada por essa abertura pode levar a eventos de agregação.

1.4 Interações proteicas

Quando falamos de interações entre proteínas, a depender do contexto, pode-se entender o acoplamento ou o compartilhamento de uma superfície de contato [16] onde duas ou mais proteínas ficam bastante próximas uma da outra por um certo tipo de encaixe. É o caso, por exemplo, da formação de estruturas quaternárias ou outros tipos de complexos proteicos, efêmeros ou mais duradouros. Em geral, essas interações estão relacionadas a alguma função biológica específica.

Aqui, porém, estaremos tratando de interações entre proteínas dispersas, ou seja, não acopladas. Em razão da presença de forças diversas (forças de Coulomb, dipolo-dipolo, van der Waals, volume-excluído, hidrofóbica etc.) atuando sobre proteínas em solução, ou colóides de maneira em geral, elas estarão sujeitas a uma soma de potenciais de interação atuando sobre várias escalas de distância. Nesse sentido, proteínas quaisquer podem interagir dependendo de suas cargas superficiais, da polaridade de seus resíduos de aminoácidos, assim como das condições físico-químicas do meio (temperatura, força iônica, pH) e outros fatores.

Uma das condições que afetam o modo como proteínas interagem diz respeito à própria concentração das mesmas, o chamado efeito de aglomeração, *crowding effect*. Tendo em vista que as proteínas de uma solução diluída podem se comportar diferentemente do que se estivessem em altas concentrações, como aquelas do ambiente intracelular, é interessante buscar, neste último caso, modelos de interação onde o efeito da alta aglomeração proteica seja um fator relevante. Alguns exemplos de sistemas altamente concentrados incluem o interior das células vermelhas do sangue, com uma estimativa bruta de 350 mg/ml de Hemoglobinas [17], e nossas lentes oculares com > 500 mg/ml de cristalinas [18].

Um dos meios para se estudar sistemas em altas concentrações é através da técnica de espalhamento de raios-X a baixos ângulos (SAXS). A técnica consiste em bombardear uma amostra com um feixe de raios-X, colimado e monocromático, e medir em

pequenos ângulos a luz espalhada pelas partículas num detector bidimensional (fig. 1.6). Por meio de uma integração radial das imagens assim coletadas, obtêm-se uma curva unidimensional, $I(q)$, da intensidade em função do módulo do vetor de espalhamento q . Em teoria, para certos sistemas simples, podemos escrever essa função como o produto de um fator, $P(q)$, que depende apenas da forma das partículas espalhadoras e de outro, $S(q)$, que contém informações sobre as interações entre elas: $I(q) = kn_p P(q) S(q)$, onde k é uma constante relacionada ao arranjo experimental e n_p a densidade numérica de partículas.

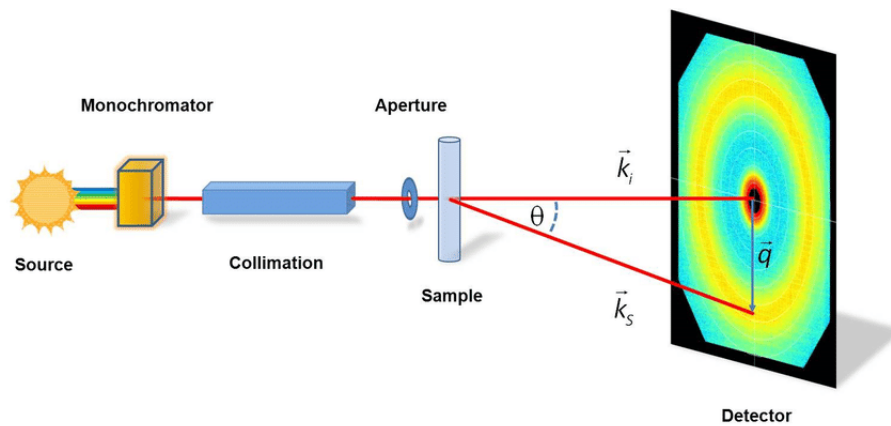


Figura 1.6: Esquema básico de um experimento de SAXS. Referência: *Structural investigations using small angle scattering techniques and contrast variation* - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Scheme-of-the-general-setup-of-a-small-angle-scattering-experiment_fig4_279829326 [accessed 7 May, 2020]

É possível relacionar o fator de estrutura $S(q)$, através de uma série de aproximações envolvendo a teoria dos fluidos, à energia potencial de pares efetiva $U(r)$ entre as partículas, o que nos permite obter detalhes desses potenciais de interação a partir do ajuste das curvas de SAXS. Um dos modelos mais bem estabelecidos dentro da teoria de colóides é o chamado potencial DLVO (Derjaguin-Landau-Verwey-Overbeek)¹, desenvolvido para explicar a estabilidade ou agregação de dispersões coloidais [19, 20]. Além da interação de volume excluído, o modelo propõe como principais forças atuantes nesses sistemas o potencial de Coulomb blindado (repulsivo para cargas de mesmo sinal) somado a uma interação atrativa de van der Waals. A teoria DLVO, de fato, é um paradigma nos estudos de colóides e o termo coulombiano com blindagem é quase uma peça indispensável em qualquer modelo com partículas carregadas. Já as contribuições atrativas para o potencial têm tido uma abordagem mais variada: enquanto alguns autores continuam a usar termos de van der Waals [21], outros preferem utilizar potenciais como o de Lennard-Jones [22], o poço quadrado [23], uma função tipo-Yukawa [24, 25] (nesse caso, chamaremos o potencial DLVO de HSDY, Hard-Sphere Double Yukawa), regulação de

¹Cf. Apêndice B para uma dedução da parte eletrostática deste potencial.

carga [26] etc., dependendo do tipo de modelagem e dos fatores físicos levados em consideração. Seja como for, a obtenção de uma teoria simples e robusta para a descrição das interações proteicas continua sendo grande um desafio.

Outra abordagem para se estudar as interações e a estrutura de partículas dispersas é por meio de simulações de Monte Carlo (MC), técnica que consiste num processo estocástico de amostragem massiva de dados. Diferentemente de uma simulação de dinâmica molecular que, a princípio, é totalmente determinística e possui uma capacidade limitada de acesso ao espaço de configurações, uma simulação de MC pode produzir uma amostragem aleatória, porém representativa do mesmo espaço. Por outro lado, por não envolver o tempo em sua dinâmica, o MC está limitado ao cálculo de grandezas estáticas, como as que dependem exclusivamente das posições.

Utilizando-se de modelos simplificados, é possível gerar uma amostragem de milhões de configurações representativas do sistema e, a partir delas, calcular propriedades estatísticas médias, como a energia, a distribuição radial e o fator de estrutura. Há muitas possibilidades de complementar os estudos de SAXS por meio de simulações deste tipo. A determinação rotineira do formato médio de uma partícula espalhadora (seu "envelope", ou, estrutura de baixa resolução), por exemplo, é feita por meio de uma simulação de MC que utiliza a função de distribuição de distâncias experimental como peso estatístico para a realização da busca. Também é possível testar modelos e aproximações teóricas empregados em análises de SAXS comparando suas previsões com os resultados "exatos" de simulação [27, 21]. No entanto, há pouquíssimos trabalhos na literatura que utilizam MC para analisar misturas binárias e nenhum que realize ajustes de misturas com simulações de MC. Assim, nossa proposta, aqui, vem no sentido de suprir esta lacuna na literatura de SAXS.

1.5 Objetivos

- A partir de dados de espalhamento de SAXS, investigar as interações de duas proteínas modelo: BSA e a lisozima
 - Tanto em sistemas puros quanto em misturas binárias;
 - Variando as condições de temperatura, força iônica e concentração proteica.
- A princípio, analisar as curvas de intensidade de espalhamento através de um modelo de interação teórico (HSDY/RPA).
- Depois, utilizar simulações de Monte Carlo:
 - Para testar a validade do modelo HSDY/RPA;
 - Como ferramenta de ajuste dos dados experimentais.

Experimentos de SAXS

2.1 Medidas no Laboratório Nacional de Luz Síncrotron (LNLS)

Preparação das amostras

No dia anterior às medições de SAXS, foram preparadas duas soluções estoque, uma só com lisozima e outra só com BSA, a 2.72 mM cada. Para isso, foram utilizadas as proteínas liofilizadas da marca Sigma-Aldrich, códigos L6876 (lisozima de clara de ovo) e A3059 (albumina sérica bovina). As proteínas foram solubilizadas à temperatura ambiente por agitação magnética durante uma hora em tampão PBS (NaCl 137 mM, Fosfato 10 mM, KCl 2.7 mM). O pH foi acertado em 7.4 com o auxílio de um pHmetro e soluções concentradas de NaOH e HCl. As soluções estoque foram então mantidas sob refrigeração até momentos antes de sua utilização.

Nas dependências da linha SAXS1 do Laboratório Nacional de Luz Síncrotron (LNLS, CNPEM), preparamos diversas amostras a partir da diluição ou da mistura das duas soluções estoque, ver tabelas 2.1 e 2.2. Para as soluções simples, foram preparadas dez concentrações diferentes de cada proteína, variando no intervalo de 0.14 a 1.36 mM. Já as misturas foram preparadas em diversas proporções de Lisozima:BSA, porém mantendo sempre a concentração molar total de proteínas fixa em 1.36 mM.

Coleta de dados

Os dados de espalhamento foram coletados na linha SAXS1 do LNLS, entre os dias 7 e 8 de maio de 2018. Utilizou-se uma distância entre o porta-amostras e o detector de cerca de um metro, resultando num intervalo de detecção entre $q = 0.013 - 0.5 \text{ \AA}^{-1}$. Os tempos de exposição aos raios-X foram de 10 segundos, para as amostras de BSA pura e as misturas, e de 40 segundos para as de lisozima pura. Todas as medidas foram realizadas em duplicata ou triplicata, em três temperaturas distintas: 10, 23 e 37 °C, sendo que as temperaturas foram mantidas relativamente estáveis no equipamento com o auxílio de um banho térmico.

Tabela 2.1: Diluições das soluções de Lisozima e BSA puras - Amostras medidas no LNLS.

Concentração (mM)	Concentração Lis. (mg/ml)	Concentração BSA (mg/ml)
1,36	19.5	90.3
1,22	17.6	81.3
1,17	16.7	77.4
1,02	14.6	67.7
0,82	11.7	54.2
0,68	9.8	45.2
0,54	7.8	36.1
0,34	4.9	22.6
0,19	2.8	12.9
0,14	2.0	9.0

Adicionalmente, também foram tomadas medidas de algumas amostras após 4 horas de preparação para verificarmos se poderia estar ocorrendo agregação das proteínas, mas nenhuma mudança nas curvas de espalhamento nesse intervalo de tempo foi observada. A integração radial das imagens foi realizada através de um software, CS-Studio, instalado nos computadores da própria linha de luz. O espalhamento devido somente às proteínas (e à camada de hidratação) foi isolado pela subtração da intensidade de espalhamento das amostras pela intensidade espalhada apenas pelo tampão. As incertezas na intensidade foram obtidas pela fórmula usual do desvio padrão:

$$\sigma_I^2(q) = \frac{1}{M-1} \sum_{j=1}^M [I_j(q) - \overline{I(q)}]^2 \quad (2.1)$$

sendo M o número de réplicas da medida.

2.2 Medidas no Diamond Light Source

Além dos dados obtidos no LNLS, também dispomos de um conjunto de medidas realizado e cedido pelo prof. Dr. Francesco Spinozzi (Università Politecnica delle Marche, Ancona, Itália), em colaboração com nosso grupo de pesquisa. As amostras, nesse caso, consistiram de misturas de lisozima e BSA, preparadas em tampão Tris e pH 8.0, em diversas proporções (ver tabela 2.3), todas com uma concentração proteica total fixa em 1.36 mM. Além disso, essas misturas foram preparadas com duas concentrações de NaCl: 0 e 50 mM.

As medidas foram tomadas na linha de luz B21 do síncrotron Diamond Light Source (Harwell Science & Innovation Campus, Didcot, UK), operando no comprimento de onda $\lambda = 1.00 \text{ \AA}$ e com uma distância entre a amostra e o detector de 4.014 m. O intervalo de

Tabela 2.2: Misturas de Lisozima e BSA - Amostras medidas no LNLS.

Proporção Li- sozima:BSA	Porcentagem de Lis. (%)	Porcentagem de BSA (%)	Concentração total (mM)	Concentração total (mg/ml)
0:1	0	100	1.36	90.3
1:9	10	90	1.36	83.2
1:6	14.3	85.7	1.36	80.2
1:3	25	75	1.36	72.6
2:3	40	60	1.36	62.0
1:1	50	50	1.36	54.9
3:2	60	40	1.36	47.8
3:1	75	25	1.36	37.2
6:1	85.7	14.3	1.36	29.6
9:1	90	10	1.36	26.6
1:0	100	0	1.36	19.5

q , neste caso, foi de 0.01 a 0.37 \AA^{-1} . As medidas foram replicadas em sete temperaturas distintas (10, 15, 20, 25, 30, 35 and 40° C). As intensidades de espalhamento calibradas foram obtidas considerando-se a intensidade do feixe incidente, a eficiência do detector, a transmissão das amostras e a subtração das contribuições do tampão.

2.3 Software de ajuste: GENFIT

Parte da análise das curvas de SAXS foi feita utilizando o software de ajuste GENFIT¹, desenvolvido pelo prof. Dr. Francesco Spinozzi et al. [28]. Este software disponibiliza uma série de modelos desenvolvidos para a análise e ajuste de curvas de SAXS, não só de proteínas, como também de membranas lipídicas, polímeros flexíveis, dentre outros. Em particular, utilizamos o modelo 15 (“All-Atoms Form Factor Of A PDB Structure With Solvation Shell Of Dummy Atoms And Multipole Expansion Average”) quando buscamos ajustar apenas os fatores de forma das proteínas, os quais foram calculados a partir dos arquivos PDB 6LYZ (lisozima) e 4F5S (BSA); mas principalmente o modelo 46 (Interacting PDB Structures: All-Atoms Form Factor Of A PDB Structure With Solvation Shell Of Dummy Atoms And Multipole Expansion Average And DLVO Potential Under RPA), o qual, além de utilizar os arquivos PDB para ajustar o fator de forma aos dados experimentais, também aplica a relação de fechamento *Random Phase Approximation* (RPA), juntamente com um modelo de interação tipo-DLVO (HSDY), para ajustar o chamado fator de estrutura. Este último modelo será devidamente explicado e discutido mais adiante.

¹Disponível para download em <https://sites.google.com/site/genfitweb/>.

Tabela 2.3: Misturas de Lisozima e BSA - Amostras medidas no Diamond Light Source

Proporção Li- sozima:BSA	Porcentagem de Lis. (%)	Porcentagem de BSA. (%)	Concentração de Lis. (mM)	Concentração de BSA (mM)
0:1	0	100	0	1.3636
1:19	5	95	0.0682	1.2955
1:9	10	90	0.1364	1.2273
3:17	15	85	0.2045	1.1591
1:4	20	80	0.2727	1.0909
1:3	25	75	0.3409	1.0227
3:7	30	70	0.4091	0.9545
7:13	35	65	0.4773	0.8864
2:3	40	60	0.5455	0.8182
9:11	45	55	0.6136	0.7500
1:1	50	50	0.6818	0.6818
11:9	55	45	0.7500	0.6136
3:2	60	40	0.8182	0.5455
13:7	65	35	0.8864	0.4773
7:3	70	30	0.9545	0.4091
3:1	75	25	1.0227	0.3409
4:1	80	20	1.0909	0.2727
17:3	85	15	1.1591	0.2045
9:1	90	10	1.2273	0.1364
19:1	95	5	1.2955	0.0682
1:0	100	0	1.3636	0

Teoria de SAXS

3.1 Equação Geral

Considere uma dada região V do espaço onde está contida uma distribuição eletrônica $\rho_e = -e\rho$, sendo e a carga elementar e ρ a densidade numérica de elétrons. A carga contida num pequeno volume arbitrário $\Delta\mathbf{r}_i$ dessa região pode ser expressa por:

$$q_i = -e\rho(\mathbf{r}_i)\Delta\mathbf{r}_i = -en_i \quad (3.1)$$

onde n_i é a fração numérica de elétrons neste elemento de volume. Imagine agora que por V passe uma onda eletromagnética plana, monocromática e linearmente polarizada. Se adotarmos um sistema de coordenadas em V de tal modo que o eixo \mathbf{y} coincida com a direção de propagação da onda e o eixo \mathbf{z} com sua direção de polarização, podemos escrever a equação da radiação incidente como:

$$\mathbf{E}_o(y,t) = (0,0,E_oe^{i\phi})e^{i(\omega t - ky)} = E_ze^{i(\omega t - ky)}\hat{\mathbf{z}} \quad (3.2)$$

Vamos supor que a onda provoque um movimento harmônico nos elétrons, de modo que eles possam ser considerados como dipolos oscilantes. A radiação emitida devido à aceleração da carga q_i é dada pela equação aproximada¹ para grandes distâncias de observação $r \gg r_i$:

$$\mathbf{E}_i(\mathbf{r},t) = \frac{\mu_0}{4\pi R_i} \left[(\hat{\mathbf{R}}_i \cdot \ddot{\mathbf{p}}_i) \hat{\mathbf{R}}_i - \ddot{\mathbf{p}}_i \right] \quad (3.3)$$

onde,

$$\mathbf{R}_i = \mathbf{r} - \mathbf{r}_i \quad (3.4)$$

$$\ddot{\mathbf{p}}_i(t_o) = q_i\ddot{\mathbf{r}}_i(t_o) \quad , \quad t_o = t - \frac{R_i}{c} \quad (3.5)$$

Repare que a derivada segunda em relação ao tempo do momento de dipolo da carga q_i é substituída por seu valor no tempo t_o para levar em conta o atraso na propagação do campo emitido. Se aplicarmos a segunda lei de Newton na equação (5), ignorando a

¹Cf. Griffiths, David J., *Eletrodinâmica*, 3 ed., Cap.11.

contribuição da componente magnética da força de Lorentz, teremos que:

$$\ddot{\mathbf{p}}_i(t_o) = \frac{q_i^2}{m_i} \mathbf{E}_o(y_i, t_o) = \frac{e^2}{m_e} n_i E_z e^{i(\omega t_o - ky_i)} \hat{\mathbf{z}} \quad (3.6)$$

onde usamos na última passagem que $m_i = n_i m_e$, sendo m_e a massa do elétron. Substituindo a equação (6) em (3), teremos:

$$\mathbf{E}_i(\mathbf{r}, t) = \frac{r_e}{R_i} n_i E_z e^{i(\omega t_o - ky_i)} \left[(\hat{\mathbf{R}}_i \cdot \hat{\mathbf{z}}) \hat{\mathbf{R}}_i - \hat{\mathbf{z}} \right] \quad (3.7)$$

sendo

$$r_e = \frac{\mu_0 e^2}{4\pi m_e} \quad (3.8)$$

o chamado raio clássico do elétron (2.8×10^{-13} cm). Vamos calcular a operação dos vetores separadamente, representando $\hat{\mathbf{R}}_i$ em coordenadas esféricas:

$$\begin{aligned} (\hat{\mathbf{R}}_i \cdot \hat{\mathbf{z}}) \hat{\mathbf{R}}_i - \hat{\mathbf{z}} &= \cos \theta_i \hat{\mathbf{R}}_i - \hat{\mathbf{z}} \\ &= (\cos \theta_i \sin \theta_i \cos \phi_i, \cos \theta_i \sin \theta_i \sin \phi_i, \cos^2 \theta_i - 1) \\ &= \sin \theta_i (\cos \theta_i \cos \phi_i, \cos \theta_i \sin \phi_i, -\sin \theta_i) \\ &= \sin \theta_i \hat{\boldsymbol{\theta}}_i \end{aligned} \quad (3.9)$$

Assim, a equação (7) fica:

$$\mathbf{E}_i(\mathbf{r}, t) = \frac{r_e}{R_i} n_i \sin \theta_i E_z e^{i(\omega t_o - ky_i)} \hat{\boldsymbol{\theta}}_i \quad (3.10)$$

Antes de prosseguirmos, vamos fazer algumas aproximações que levem em conta as condições de observação, isto é, detecção afastada dos centros espalhadores ($r \gg r_i$) e a baixos ângulos ($\cos 2\Theta = \hat{\mathbf{r}} \cdot \hat{\boldsymbol{\gamma}} \approx 1$). Nessas condições, podemos expandir R_i , usando a lei dos cossenos, em uma série de Taylor truncada até a primeira ordem:

$$R_i \approx r - \hat{\mathbf{r}} \cdot \mathbf{r}_i \approx R - \hat{\mathbf{r}} \cdot \mathbf{r}_i \quad (3.11)$$

onde $R = \mathbf{r} \cdot \hat{\boldsymbol{\gamma}} \approx r$ é a distância horizontal entre a origem e o detector. No denominador da equação (10), vamos aproximar ainda $R_i \approx R$, porém manteremos a aproximação até a primeira ordem na exponencial complexa, uma vez que seu argumento é mais sensível a pequenas variações. Nesse caso,

$$t_o = t - \frac{R_i}{c} \approx t - \frac{R}{c} + \frac{\hat{\mathbf{r}} \cdot \mathbf{r}_i}{c} \quad (3.12)$$

Outra aproximação que podemos fazer é em relação à direção de \mathbf{R}_i :

$$\hat{\mathbf{R}}_i = \frac{\mathbf{R}_i}{R_i} \approx \frac{\mathbf{r} - \mathbf{r}_i}{r} \approx \hat{\mathbf{r}} \quad (3.13)$$

$$\Rightarrow \sin \theta_i \hat{\boldsymbol{\theta}}_i \approx \sin \theta \hat{\boldsymbol{\theta}} \quad (3.14)$$

Assim, substituindo (11,12,14) em (10), ficamos com:

$$\mathbf{E}_i(\mathbf{r}, t) = \frac{r_e}{R} n_i \sin \theta E_z e^{i(\omega t - kR)} e^{ik(\hat{\mathbf{r}} \cdot \mathbf{r}_i - y_i)} \hat{\boldsymbol{\theta}} \quad (3.15)$$

Apenas para utilizarmos uma notação corrente na literatura, vamos definir os seguintes vetores: $\hat{\mathbf{s}} = \hat{\mathbf{r}}$ e $\hat{\mathbf{s}}_o = \hat{\mathbf{y}}$, ou seja, as direções de espalhamento e de incidência, de modo que possamos escrever:

$$\hat{\mathbf{r}} \cdot \mathbf{r}_i - y_i = \hat{\mathbf{s}} \cdot \mathbf{r}_i - \hat{\mathbf{s}}_o \cdot \mathbf{r}_i = (\hat{\mathbf{s}} - \hat{\mathbf{s}}_o) \cdot \mathbf{r}_i \quad (3.16)$$

Logo,

$$\mathbf{E}_i(\mathbf{q}, t) = \frac{r_e}{R} n_i \sin \theta E_o(R, t) e^{iq \cdot \mathbf{r}_i} \hat{\boldsymbol{\theta}} \quad (3.17)$$

onde definimos o chamado vetor de espalhamento \mathbf{q} :

$$\mathbf{q} = k(\hat{\mathbf{s}} - \hat{\mathbf{s}}_o) = \frac{2\pi}{\lambda} (\hat{\mathbf{s}} - \hat{\mathbf{s}}_o) \quad (3.18)$$

cujo módulo $q = 4\pi \sin \Theta / \lambda$ é facilmente deduzível através da lei dos cossenos. Agora, para obtermos o campo elétrico total emitido pela distribuição eletrônica em V , basta somarmos a contribuição de cada elemento de carga q_i :

$$\mathbf{E}(\mathbf{q}, t) = \sum_i \mathbf{E}_i = \frac{r_e}{R} \sin \theta E_o(R, t) \left(\sum_i n_i e^{iq \cdot \mathbf{r}_i} \right) \hat{\boldsymbol{\theta}} \quad (3.19)$$

O termo entre parênteses, por sua vez, pode ser interpretado como a transformada de Fourier da densidade numérica de elétrons:

$$\rho_{\mathbf{q}} = \sum_i n_i e^{iq \cdot \mathbf{r}_i} = \int \rho(\mathbf{r}) e^{iq \cdot \mathbf{r}} d\mathbf{r} \quad (3.20)$$

Por fim, a intensidade de espalhamento da distribuição “estática” $\rho(\mathbf{r})$, numa pequena região do detector definida pelo vetor de espalhamento, será dada pela média temporal do módulo do vetor de Poynting nessa região:

$$I(\mathbf{q}) = \langle S \rangle_t = \frac{c\epsilon_0}{2} \mathbf{E} \mathbf{E}^* = \frac{c\epsilon_0 E_o^2}{2} \frac{r_e^2}{R^2} \sin^2 \theta \rho_{\mathbf{q}} \rho_{-\mathbf{q}} = I_o \frac{r_e^2}{R^2} \sin^2 \theta \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \quad (3.21)$$

Definindo $\hat{r} = (\alpha, \beta, \gamma)$, também podemos escrever o termo angular como $\sin^2 \theta = 1 - \gamma^2$. Esse fator se deve exclusivamente ao estado de polarização linear da radiação incidente e à nossa escolha do eixo z estar na direção do campo elétrico. Se tivéssemos realizado a dedução com uma onda polarizada genérica (o que supõe apenas uma diferença de fase $\Delta\phi = \phi_x - \phi_z$ constante entre as componentes do campo), obteríamos, a baixos ângulos, uma média para cada componente:

$$\frac{1 - \gamma^2}{2} + \frac{1 - \alpha^2}{2} = \frac{1 + \beta^2}{2} \quad (3.22)$$

onde $\beta = \cos 2\Theta$ é o cosseno do ângulo de espalhamento. Assim, no caso mais geral de polarização, teríamos:

$$I(\mathbf{q}) = I_o \frac{r_e^2}{R^2} \frac{1 + \cos^2 2\Theta}{2} \rho_{\mathbf{q}} \rho_{-\mathbf{q}} = I_e(\mathbf{q}) \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \quad (3.23)$$

sendo que $I_e(\mathbf{q})$, conhecido como fórmula de Thomson, fornece o espalhamento de um único elétron nas mesmas condições experimentais e pode ser considerado praticamente uma constante a baixos ângulos. A equação acima diz respeito à intensidade espalhada por uma distribuição eletrônica “estática” (partículas espalhadoras em posições espaciais fixas), porém, se quisermos obter a intensidade de uma distribuição que varia no tempo, devemos realizar uma média de (23) no ensemble do sistema. Isso nos leva, finalmente, à expressão geral:

$$I(\mathbf{q}) = I_e(\mathbf{q}) \langle \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \rangle \quad (3.24)$$

O símbolo $\langle \rangle$ denota aqui a operação de obtenção do valor médio em um dado ensemble, o que envolve percorrer todo o espaço de fase do sistema. Em geral, supomos que a temperatura e o volume irradiado de amostra são mantidos fixos nas condições experimentais. Se, além disso, o sistema contido nesse volume não trocar partículas com o entorno, estaremos no chamado ensemble canônico; caso a troca de partículas for permitida, mas o potencial químico das espécies envolvidas se mantiver constante, a média será no ensemble grande canônico. Seja qual for o caso, estaremos supondo que o sistema envolvido esteja em equilíbrio termodinâmico, de modo que a densidade de probabilidade f dos microestados não mude com tempo, e que o intervalo de medição seja longo o bastante para obtermos uma boa amostragem do espaço conformacional. Devido à equivalência dos ensembles no limite termodinâmico e à maior simplicidade das expressões no canônico, adotaremos este ensemble para os cálculos seguintes.

Obs.1: É comum encontrarmos na literatura a equação 24 expressa da seguinte forma:

$$\frac{d\Sigma}{dq} \equiv V^{-1} \frac{I(\mathbf{q})}{I_e(\mathbf{q})} = V^{-1} \langle \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \rangle \quad (3.25)$$

sendo V o volume de amostra irradiada. Essa função é chamada de *seção-de-choque de espalhamento diferencial macroscópica* e, por simplicidade, também é referida como sendo a “intensidade” de espalhamento.

Obs.2: Se estamos interessados apenas no espalhamento das partículas (colóides, proteínas etc.) dispersas numa solução e não no espalhamento dos elétrons do meio, como frequentemente é o caso, o habitual é subtrair a intensidade total espalhada pela solução daquela espalhada somente pelo solvente. Desse modo, devemos substituir, na equação 24, a densidade eletrônica total, ρ , pela função *contraste de densidade eletrônica*: $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_o$, onde ρ_o é a densidade eletrônica média do solvente (que vale cerca de $0.33 e^-/\text{Å}^3$, no caso da água).

Sistemas monodispersos

Daqui pra frente, iremos tratar apenas de sistemas de partículas com estrutura interna fixa, de modo que possamos definir um centro (p.ex., o centro de massa, ou algum outro de alta simetria) e uma orientação para cada uma delas. As coordenadas espaciais dos centros das partículas ($\mathbf{r}^N = \mathbf{r}_1, \dots, \mathbf{r}_N$) e de suas orientações ($\boldsymbol{\omega}^N = \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N$) devem, então, ser suficientes para descrever todas as possíveis configurações de densidade e, portanto, de ρ_q . Assim, como a intensidade de uma dada conformação (23) não depende dos momentos, e, supondo que o Hamiltoniano do sistema possa ser decomposto na soma de um termo potencial $V(\mathbf{r}^N, \boldsymbol{\omega}^N)$ que depende apenas das variáveis espaciais e de um termo puramente cinético, a média estatística pode ser simplificada para integrais realizadas apenas sobre as coordenadas de posição e orientação. No ensemble canônico, teríamos:

$$\langle A \rangle = \int A(\mathbf{r}^N, \boldsymbol{\omega}^N) f(\mathbf{r}^N, \boldsymbol{\omega}^N) d\mathbf{r}^N d\boldsymbol{\omega}^N \quad (3.26)$$

onde $f = \exp(-\beta V)/Z_N$ é a função densidade de probabilidade do microestado ($\mathbf{r}^N, \boldsymbol{\omega}^N$). Ao todo, temos integrais sobre $6.N$ graus de liberdade: 3 para cada posição $r_i = (x_i, y_i, z_i)$ e 3 para cada orientação $w_i = (\theta_i, \phi_i, \chi_i)$, onde os ângulos de Euler representam a rotação dos eixos moleculares da i -ésima partícula com respeito ao sistema de coordenadas do laboratório.

Hipóteses adicionais: H_1 e H_2

Antes de prosseguirmos com o cálculo do valor médio, vamos ad admitir duas outras hipóteses, as quais chamaremos de H_1 e H_2 , mantendo a nomenclatura presente no livro-texto de Guinier e Fournet².

H_1 : As variáveis de orientação são independentes das de posição. Além disso, todas as orientações são igualmente prováveis.

H_2 As partículas estão distribuídas de maneira homogênea e isotrópica.

²Guinier, A., Fournet, G. *Small-Angle Scattering of X-Rays*, 1955, Cap. 2, p. 30.

A primeira hipótese é equivalente a dizer que o potencial de interação não depende das orientações, ou seja, $V = V(\mathbf{r}^N)$. O que implica que podemos escrever a função densidade de probabilidade como sendo o produto de duas outras funções de probabilidade:

$$f(\mathbf{r}^N, \boldsymbol{\omega}^N) = f_r(\mathbf{r}^N) f_\omega(\boldsymbol{\omega}^N) = f_r(\mathbf{r}^N) \Omega^N \quad (3.27)$$

sendo Ω uma constante que pode ser obtida impondo-se a condição de normalização

$$1 = \int f_\omega(\boldsymbol{\omega}^N) d\boldsymbol{\omega}^N = \Omega^N \left(\int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \sin \theta d\theta d\phi d\chi \right)^N = \Omega^N (8\pi^2)^N \quad (3.28)$$

$$\iff \Omega = \frac{1}{8\pi^2} \quad (3.29)$$

(Observação: moléculas lineares possuem um grau de liberdade a menos (θ, ϕ), de modo que a normalização implicaria $\Omega = 1/4\pi$). Essa hipótese é geralmente válida para partículas esféricas e funciona como uma primeira aproximação no caso de partículas globulares e não muito anisotrópicas. No caso geral, entretanto, devemos esperar que haja a curto alcance orientações preferenciais devido, por exemplo, a interações eletrostáticas. A segunda hipótese será utilizada mais adiante.

3.2 Fator de Forma

É conveniente, agora, reescrevermos a transformada da densidade eletrônica, $\rho_{\mathbf{q}}$, numa forma que nos permita separar posteriormente as contribuições do espalhamento devido às interferências intrapartícula e interpartículas. A primeira delas dará origem ao chamado fator de forma e a segunda ao fator de interferência. Para tanto, começamos por notar que a posição \mathbf{r}_i de um determinado centro espalhador, localizado na p -ésima partícula, pode ser escrita como:

$$\mathbf{r}_i = \mathbf{r}_p + \mathbf{r}_{p,i} \quad (3.30)$$

onde \mathbf{r}_p é a posição do centro da partícula p e $\mathbf{r}_{p,i}$ é a posição interna do elemento espalhador em relação a esse centro, ambos os vetores tomados no sistema de coordenadas do laboratório. Obviamente, a posição interna de um elemento espalhador não depende da posição do centro, mas tão somente da orientação da partícula p . Assim, a somatória da equação (20) pode ser dividida em duas: uma realizada sobre as partículas e outra sobre os elementos internos das mesmas, ou seja,

$$\rho_{\mathbf{q}} = \sum_p \sum_i n_{p,i} e^{i\mathbf{q} \cdot (\mathbf{r}_p + \mathbf{r}_{p,i})} = \sum_p F_p e^{i\mathbf{q} \cdot \mathbf{r}_p} \quad (3.31)$$

onde representamos a soma dos elementos internos de p pela função

$$F_p = F(\mathbf{q}, \boldsymbol{\omega}_p) = \sum_i n_{p,i} e^{i\mathbf{q} \cdot \mathbf{r}_{p,i}} = \int \rho_p(\mathbf{r}) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \quad (3.32)$$

sendo $\rho_p(\mathbf{r})$ a densidade de elétrons no i -ésimo elemento da partícula p (omitimos os índices na forma integral, mas não percamos de vista sua interpretação física), o qual dista \mathbf{r} do centro para uma dada orientação $\boldsymbol{\omega}_p$. No entanto, apesar da posição interna de um dado elemento depender da orientação, a densidade de elétrons contida nele não varia. Logo, a dependência explícita com a orientação deve vir unicamente do produto interno $\mathbf{q} \cdot \mathbf{r}$. De fato, o vetor posição \mathbf{r} , expresso no sistema de coordenadas do laboratório, e o mesmo vetor \mathbf{u} , expresso no sistema de coordenadas da partícula, estão relacionados pela transformação: $\mathbf{r} = \mathbf{u}\Phi(\boldsymbol{\omega}_p)$, onde a matriz de rotação [29] é dada por:

$$\Phi(\boldsymbol{\omega}) = \begin{pmatrix} c\phi c\theta c\chi - s\phi s\chi & s\phi c\theta c\chi + c\phi s\chi & -s\theta c\chi \\ -c\phi c\theta s\chi - s\phi c\chi & -s\phi c\theta s\chi + c\phi c\chi & s\theta s\chi \\ c\phi s\theta & s\phi s\theta & c\theta \end{pmatrix} \quad (3.33)$$

Na chamada *matriz de cossenos diretores* acima, usamos ‘c’ para abreviar os cossenos e ‘s’ os senos. Também assumimos que \mathbf{u} , assim como \mathbf{r} , estão sendo usados como vetores linha. Com essa convenção, o produto interno com o vetor de espalhamento, tomado como um vetor coluna, deve ser expresso na ordem adequada:

$$\mathbf{r} \cdot \mathbf{q} = \mathbf{u}\Phi(\boldsymbol{\omega}_p)\mathbf{q} \quad (3.34)$$

Uma vez que as dependências com as posições e orientações das partículas estão expressas numa notação minimamente limpa, estamos agora em condições de calcular a média:

$$\langle \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \rangle = \sum_p \sum_m \langle F_p F_m^* e^{i\mathbf{q} \cdot (\mathbf{r}_p - \mathbf{r}_m)} \rangle = \sum_p \langle |F_p|^2 \rangle + \sum_{p \neq m} \langle F_p F_m^* e^{i\mathbf{q} \cdot (\mathbf{r}_p - \mathbf{r}_m)} \rangle \quad (3.35)$$

Vamos definir o termo que aparece na primeira somatória do lado direito como sendo o *fator de forma* da partícula p :

$$\begin{aligned} P(\mathbf{q}) &= \langle |F_p|^2 \rangle = \int |F_p(\boldsymbol{\omega}_p)|^2 f_r(\mathbf{r}^N) \Omega^N d\mathbf{r}^N d\boldsymbol{\omega}^N \\ &= \left(\int f_r(\mathbf{r}^N) d\mathbf{r}^N \right) \left(\Omega^N \int d\boldsymbol{\omega}^{N-1} \right) \left(\int |F_p(\boldsymbol{\omega}_p)|^2 d\boldsymbol{\omega}_p \right) \\ &= \frac{1}{8\pi^2} \int |F_p(\boldsymbol{\omega}_p)|^2 d\boldsymbol{\omega}_p \end{aligned} \quad (3.36)$$

onde as simplificações acima foram feitas com o auxílio da hipótese H_1 e a suposição de

que f_r já está normalizada. Usando as equações (3.32) e (3.34), teremos:

$$\begin{aligned}
 &= \frac{1}{8\pi^2} \int \int \int \rho_p(\mathbf{r}) \rho_p(\mathbf{r}') e^{i(\mathbf{r}-\mathbf{r}') \cdot \mathbf{q}} d\mathbf{r} d\mathbf{r}' d\boldsymbol{\omega}_p \\
 &= \int \int \rho_p(\mathbf{r}) \rho_p(\mathbf{r}') \left(\frac{1}{8\pi^2} \int e^{i\mathbf{u}\Phi(\boldsymbol{\omega}_p)\mathbf{q}} d\boldsymbol{\omega}_p \right) d\mathbf{r} d\mathbf{r}' \quad (3.37)
 \end{aligned}$$

sendo que aqui \mathbf{u} representa o vetor $\mathbf{r} - \mathbf{r}'$, mas expresso no sistema de referência da partícula. Isso implica que $u = |\mathbf{r} - \mathbf{r}'|$ é igual à distância entre o par de elementos espalhadores da mesma molécula definidos pelos vetores \mathbf{r} e \mathbf{r}' . O termo entre parênteses na equação anterior é igual ao valor médio do fator exponencial sobre todas as orientações da partícula p . Sem perda de generalidade, podemos escolher a orientação dos eixos dos sistemas de coordenadas do laboratório e da partícula de tal modo que $\mathbf{q} = q\hat{\mathbf{z}}$ seja paralelo ao eixo z do primeiro e $\mathbf{u} = u\hat{\mathbf{z}}$ seja paralelo ao eixo z do segundo – o resultado, é claro, deve ser independente dessa escolha arbitrária, porém conveniente. Sendo assim, a média sobre as orientações de p será:

$$\begin{aligned}
 \frac{1}{8\pi^2} \int e^{iuq[\hat{\mathbf{z}}\Phi(\boldsymbol{\omega}_p)\hat{\mathbf{z}}]} d\boldsymbol{\omega}_p &= \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi e^{iuq \cos \theta_p} \sin \theta_p d\theta_p d\phi_p d\chi_p \\
 &= \frac{(2\pi)^2}{8\pi^2} \int_{-iuq}^{iuq} \frac{e^a}{iuq} da = \frac{e^{iuq} - e^{-iuq}}{2iuq} \\
 &= \frac{\sin(uq)}{uq} \quad (3.38)
 \end{aligned}$$

onde a integral em θ_p foi realizada através da mudança de variável $a = iuq \cos \theta_p$. Assim, substituindo a equação (3.38) em (3.37), obtemos

$$P(q) = \int \int \rho_p(\mathbf{r}) \rho_p(\mathbf{r}') \frac{\sin(uq)}{uq} d\mathbf{r} d\mathbf{r}' \quad (3.39)$$

um resultado análogo àquele obtido pela primeira vez por Debye (1915) [30], demonstrando que a aleatoriedade das orientações não anula os efeitos de interferência³. Uma vez que $\rho_p(\mathbf{r})d\mathbf{r}$ representa a quantidade de elétrons contida em $d\mathbf{r}$, e $\rho_p(\mathbf{r}')d\mathbf{r}'$ a quantidade de elétrons em $d\mathbf{r}'$, o produto dos dois termos equivale à quantidade de pares de elétrons entre esses elementos de volume. A equação (3.39), portanto, nada mais é do

³Por uma grande coincidência, no dia em que o artigo de Debye foi recebido pelos editores outro artigo de P. Ehrenfest sobre o mesmo tema é apresentado em um encontro da Amsterdam Academy por H. A. Lorentz e H. Kamerlingh Onnes. Este trabalho, apesar de tratar apenas de moléculas diatômicas (sendo cada átomo um centro espalhador pontual), chega à mesma conclusão, embora restrita a essa geometria. Cf. Ehrenfest, P. *About Interference Phenomena to be Expected when X-rays Pass Through a Diatomic Gas*. Amst. Akad., 23, 1132 (1915)

que uma soma de pares de elétrons ponderada por uma função que depende da distância u entre eles.

A partir da expressão acima, podemos definir algumas outras funções que são comumente empregadas na análise de curvas de SAXS, em particular a função de distribuição de pares $p(u)$. Primeiro, vamos reescrever (3.39) usando a mudança de variável $\mathbf{u} = \mathbf{r} - \mathbf{r}'$:

$$\begin{aligned} P(q) &= \int \rho_p(\mathbf{r}') \left(\int \frac{\sin(uq)}{uq} \rho_p(\mathbf{u} + \mathbf{r}') d\mathbf{u} \right) d\mathbf{r}' \\ &= \int \frac{\sin(uq)}{uq} \left(\int \rho_p(\mathbf{r}') \rho_p(\mathbf{u} + \mathbf{r}') d\mathbf{r}' \right) d\mathbf{u} \\ &= \int \frac{\sin(uq)}{uq} V(\mathbf{u}) d\mathbf{u} \end{aligned} \quad (3.40)$$

A função $V(\mathbf{u})$ definida na última passagem é, claramente, a autoconvolução da função de densidade. Para uma partícula com densidade uniforme ρ_o , essa função é simplesmente dada por $\rho_o^2 v(\mathbf{u})$, sendo $v(\mathbf{u})$ o volume da intersecção entre a partícula e uma cópia da mesma deslocada de \mathbf{u} . Ela também pode ser interpretada como a densidade de pares eletrônicos formados pelos elétrons contidos nessa intersecção. Passando a integral em (3.40) para coordenadas esféricas, teremos:

$$\begin{aligned} &= \int_0^\infty \frac{\sin(uq)}{uq} u^2 \left(\int_0^\pi \int_0^{2\pi} V(\mathbf{u}) \sin \theta_u d\theta_u d\phi_u \right) du \\ &= \int_0^\infty \frac{\sin(uq)}{uq} 4\pi u^2 \bar{V}(u) du \end{aligned} \quad (3.41)$$

onde $\bar{V}(u)$ é a média de $V(\mathbf{u})$ sobre todas as orientações. Podemos, enfim, definir a *função de distribuição de pares* como:

$$p(u) = 4\pi u^2 \bar{V}(u) \quad (3.42)$$

e, portanto,

$$P(q) = \int_0^\infty p(u) \frac{\sin(uq)}{uq} du \quad (3.43)$$

Para interpretarmos a função $p(u)$, é conveniente reescrevê-la da seguinte forma (vamos omitir aqui o índice da partícula):

$$p(u) du = \int d\mathbf{r}' \rho(\mathbf{r}') \left(\int_{\Omega_u} \rho(\mathbf{r}' + \mathbf{u}) d\mathbf{u} \right) \quad (3.44)$$

que é igual ao produto do número de elétrons em $d\mathbf{r}'$ pela quantidade de elétrons contidos na casca esférica de raio u centrada nesse elemento de volume, integrado sobre todo o volume da partícula, o que equivale ao número total de pares eletrônicos na partícula que distam u entre si (obs. o mesmo par é contado duas vezes). Assim, $p(u)$ é a densidade de pares por unidade de distância. Além disso, se integrarmos a equação (3.44) para todas as distâncias, obteremos:

$$P(q \rightarrow 0) = \int_0^\infty p(u)du = \int d\mathbf{r}' \rho(\mathbf{r}') \left(\int \rho(\mathbf{r}' + \mathbf{u})d\mathbf{u} \right) = n_e^2 \quad (3.45)$$

onde n_e é o número total de elétrons espalhadores da partícula. Outra propriedade da função $p(u)$, a qual decorre de sua definição (3.42), é que $p(0) = 0$ e $p(u \geq D_{\max}) = 0$, ou seja, se a distância u for maior do que a dimensão máxima da partícula, caso em que mais nenhum par pode ser encontrado. Assim, $p(u)$ está inteiramente concentrada na região das distâncias típicas da partícula ($0 < u < D_{\max}$) e fornece uma indicação direta da sua dimensão máxima no ponto em que vai definitivamente a zero.

Podemos obter uma expressão mais prática para $p(u)$ através da transformada inversa da equação (3.43):

$$\begin{aligned} \int_0^\infty q \sin(uq)P(q)dq &= \int_0^\infty q \sin(uq) \left(\int_0^\infty p(a) \frac{\sin(aq)}{aq} da \right) dq \\ &= \int_0^\infty \frac{p(a)}{2a} da \int_{-\infty}^\infty \sin(uq) \sin(aq) dq \end{aligned} \quad (3.46)$$

onde usamos o fato da segunda integral ser sobre uma função par para estendemos o intervalo de integração a menos infinito. Usando agora a representação imaginária das funções seno e algumas propriedades básicas do delta de Dirac, teremos:

$$\begin{aligned} &= \int_0^\infty \frac{p(a)}{2a} da \int_{-\infty}^\infty \frac{e^{iuq} - e^{-iuq}}{2i} \cdot \frac{e^{iaq} - e^{-iaq}}{2i} dq \\ &= \int_0^\infty \frac{p(a)}{8a} da \int_{-\infty}^\infty [e^{iq(u-a)} - e^{iq(u+a)} - e^{-iq(u+a)} + e^{-iq(u-a)}] dq \\ &= \int_0^\infty \frac{p(a)}{8a} 2\pi [\delta(u-a) - \delta(u+a) - \delta(-u-a) + \delta(-u+a)] da \\ &= \int_0^\infty \frac{p(a)}{8a} 2\pi [2\delta(u-a)] da = \frac{\pi}{2u} p(u) \end{aligned}$$

Acima, supomos que $u > 0$ para podermos eliminar os deltas intermediários e usamos a paridade dos deltas para somar o primeiro e o último termos nos colchetes. Assim,

isolando $p(u)$, obtemos:

$$p(u) = \frac{2u}{\pi} \int_0^\infty P(q)q \sin(uq) dq \quad (3.47)$$

A fórmula acima, diferentemente da definição (3.42), nos permite, a princípio, calcular a função $p(u)$ a partir de dados experimentais. Na prática, porém, como não é possível medir $P(q)$ para todos os ângulos de espalhamento, realiza-se uma truncagem na integral para a obtenção da $p(u)$.

Além da função de distribuição de pares, também é comum encontrarmos na literatura a chamada *função característica* $\gamma_o(u)$, introduzida por Porod (1951) [31] e dada por:

$$\gamma_o(u) = \frac{\bar{V}(u)}{V} \Rightarrow p(u) = 4\pi u^2 V \gamma_o(u) \quad (3.48)$$

sendo V o volume total da partícula. Essa função apresenta uma série de propriedades interessantes, as quais não abordaremos aqui – para uma discussão mais detalhada ver Guinier (1955) [32].

Exemplos de $P(\mathbf{q})$

Vamos utilizar as definições que foram apresentadas para calcular o fator de forma de uma esfera homogênea. Em seguida, apresentamos uma simplificação válida para partículas com simetria radial, dando como exemplos a esfera com casca e a casca esférica. Outros fatores de forma conhecidos na literatura (elipsóide, cilindro etc.) são também mencionados e comparados entre si.

(A) Esfera Homogênea (Rayleigh, 1914)

A função de autoconvolução de uma esfera homogênea de raio R e densidade ρ_o pode ser escrita como:

$$V(\mathbf{u}) = (\rho * \rho)(\mathbf{u}) = \rho_o^2 v(\mathbf{u}) \quad (3.49)$$

sendo que o volume da intersecção entre as esferas, $v(\mathbf{u})$, pode ser obtido por simples argumentos geométricos:

$$v(u) = \frac{\pi}{12}(4R + u)(2R - u)^2 \quad (3.50)$$

Como essa função não depende da orientação do deslocamento, afinal o sistema é idêntico em todas as direções, a função de distribuição de pares terá a seguinte forma:

$$p(u) = 4\pi u^2 \bar{V}(u) = \frac{u^2 \pi^2 \rho_o^2}{3}(4R + u)(2R - u)^2 \quad (3.51)$$

válida para $u \leq 2R$, sendo nula para $u > 2R$. O fator de forma, por sua vez, será dado pela equação (3.43):

$$\begin{aligned}
 P_{\text{esf}}(q) &= \int_0^{2R} \frac{u^2 \pi^2 \rho_o^2}{3} (4R + u)(2R - u)^2 \frac{\sin(uq)}{uq} du \\
 &= \frac{\pi^2 \rho_o^2}{3q} \int_0^{2R} [16R^3 u - 12R^2 u^2 + u^4] \sin(uq) du \\
 &= \frac{16\pi^2 \rho_o^2}{q^6} [\sin(qR) - qR \cos(qR)]^2
 \end{aligned} \tag{3.52}$$

ou, em termos do volume V da esfera:

$$P_{\text{esf}}(q) = [\rho_o V \Phi(qR)]^2, \quad \Phi(qR) = 3 \frac{\sin(qR) - qR \cos(qR)}{(qR)^3} \tag{3.53}$$

Para ângulos de espalhamento tendendo a zero, $\Phi(q \rightarrow 0) = 1$, de modo que:

$$P_{\text{esf}}(q \rightarrow 0) = [\rho_o V]^2 = n_e^2 \tag{3.54}$$

o que está de acordo com a equação (3.45).

É claro que o cálculo do fator de forma pela via das definições apresentadas anteriormente pode se tornar extremamente trabalhoso, se não inviável do ponto de vista analítico, para partículas com geometrias diversas ou mesmo esfericamente simétricas. Para este último caso, no entanto, podemos introduzir algumas simplificações que facilitam enormemente as contas.

(B) Partículas centrossimétricas e/ou com simetria radial

Se a partícula em questão possui um centro de simetria, então, para cada elemento espalhador, localizado numa posição r_i , haverá outro com a mesma fração de elétrons n_i localizado na direção oposta $-r_i$. Sendo assim, a somatória da equação (3.32) pode ser reescrita de modo a considerar a soma de pares simétricos (omitiremos aqui o índice p da partícula):

$$F(\mathbf{q}, \omega) = \sum_i n_i e^{i\mathbf{q} \cdot \mathbf{r}_i} = \sum_i n_i \frac{e^{i\mathbf{q} \cdot \mathbf{r}_i} + e^{-i\mathbf{q} \cdot \mathbf{r}_i}}{2} = \sum_i n_i \cos(\mathbf{q} \cdot \mathbf{r}_i) \tag{3.55}$$

sendo que o fator 2 foi introduzido para não contarmos o mesmo elemento duas vezes. A diferença óbvia, neste caso, é que F já não é mais uma função complexa. Logo, a média no ensemble dessa função será:

$$\begin{aligned}
\langle F \rangle(q) &= \sum_i n_i \langle \cos(\mathbf{q} \cdot \mathbf{r}_i) \rangle \\
&= \sum_i \frac{n_i}{8\pi^2} \int \cos[r_i q \cos \theta] \sin \theta \, d\theta \, d\phi \, d\chi \\
&= \sum_i n_i \frac{(2\pi)^2}{8\pi^2} \int_{-r_i q}^{r_i q} \frac{\cos a}{r_i q} \, da = \sum_i n_i \frac{\sin(r_i q)}{r_i q}
\end{aligned} \tag{3.56}$$

ou, na forma contínua:

$$\langle F \rangle(q) = \int \rho(\mathbf{r}) \frac{\sin(rq)}{rq} \, d\mathbf{r} \tag{3.57}$$

Este resultado também é válido para partículas desprovidas de um centro de simetria, pois as médias $\langle \exp(i\mathbf{q} \cdot \mathbf{r}_i) \rangle = \langle \cos(\mathbf{q} \cdot \mathbf{r}_i) \rangle = \sin(r_i q)/r_i q$ são idênticas nos dois casos. O fator de forma, por sua vez, será expresso por:

$$P(q) = \langle F^2 \rangle = \sum_i \sum_j n_i n_j \langle \cos(\mathbf{q} \cdot \mathbf{r}_i) \cos(\mathbf{q} \cdot \mathbf{r}_j) \rangle \tag{3.58}$$

Essa média, no entanto, não é tão simples de resolver quanto a anterior, uma vez que só podemos assumir ou \mathbf{r}_i ou \mathbf{r}_j como sendo paralelo ao eixo z do sistema de coordenadas da partícula. Se determinamos um o outro também estará determinado, de modo que esses vetores não são independentes e, portanto, em geral:

$$\begin{aligned}
\langle \cos(\mathbf{q} \cdot \mathbf{r}_i) \cos(\mathbf{q} \cdot \mathbf{r}_j) \rangle &\neq \langle \cos(\mathbf{q} \cdot \mathbf{r}_i) \rangle \langle \cos(\mathbf{q} \cdot \mathbf{r}_j) \rangle \\
\Rightarrow \langle F^2 \rangle &\neq \langle F \rangle^2
\end{aligned}$$

Uma exceção importante à desigualdade acima ocorre para partículas com simetria radial. Nesses casos, F não deve depender da orientação $\boldsymbol{\omega}$ da partícula, de modo que:

$$P(q) = \langle F^2 \rangle = F^2 = \langle F \rangle^2 = \left[\int \rho(r) \frac{\sin(rq)}{rq} \, d\mathbf{r} \right]^2 \tag{3.59}$$

ou, integrando em coordenadas esféricas:

$$P(q) = \left[\frac{4\pi}{q} \int_0^R r \rho(r) \sin(rq) \, dr \right]^2 \tag{3.60}$$

sendo R o raio máximo da partícula. Assim, o cálculo do fator de forma de uma esfera homogênea, por exemplo, se torna bem mais direto:

$$P_{\text{esf}}(q) = \left[\frac{4\pi}{q} \rho_o \int_0^R r \sin(rq) dr \right]^2 = \left[\frac{4\pi R^3}{3} \rho_o \Phi(qR) \right]^2$$

resultado idêntico àquele expresso pela equação (3.53).

(C) Esfera com Casca

O fator de forma de uma esfera com densidades ρ_1 , para $0 < r < R_1$, e ρ_2 , para $R_1 < r < R_2$, também segue diretamente da equação (3.60):

$$P_{\text{esf2}}(q) = \left[4\pi\rho_1 \int_0^{R_1} r \frac{\sin(rq)}{q} dr + 4\pi\rho_2 \int_0^{R_2} r \frac{\sin(rq)}{q} dr - 4\pi\rho_2 \int_0^{R_1} r \frac{\sin(rq)}{q} dr \right]^2$$

$$= [V_1\rho_1\Phi(qR_1) + V_2\rho_2\Phi(qR_2) - V_1\rho_2\Phi(qR_1)]^2 \quad (3.61)$$

(D) Casca Esférica

Fazendo $\rho_1 \rightarrow 0$ na equação anterior, obtemos o fator de forma de uma casca esférica de raios interno e externo iguais a R_1 e R_2 , respectivamente:

$$P_{\text{casca}}(q) = \rho_2^2 [V_2\Phi(qR_2) - V_1\Phi(qR_1)]^2 \quad (3.62)$$

Alguns outros exemplos de fatores de forma são mencionados, a seguir, para partículas com simetrias não-radiais.

(E) Elipsóide de revolução (eixos $2R$, $2R$ e $2vR$)[33]

$$P_{\text{elp}}(q) = (\rho_o V)^2 \int_0^{\pi/2} \Phi^2 \left(qR \sqrt{\cos^2 \theta + v^2 \sin^2 \theta} \right) \cos \theta d\theta \quad (3.63)$$

(F) Cilindro de revolução (altura $2H$ e diâmetro $2R$)[34]

$$P_{\text{cil}}(q) = (\rho_o V)^2 4 \int_0^{\pi/2} \frac{\sin^2(qH \cos \theta)}{q^2 H^2 \cos^2 \theta} \cdot \frac{J_1^2(qR \sin \theta)}{q^2 R^2 \sin \theta} d\theta \quad (3.64)$$

sendo J_1 a função de Bessel de primeira espécie e ordem 1.

(G) Disco (raio R e espessura infinitesimal)[35]

$$P_{\text{disco}}(q) = (\rho_o V)^2 \frac{2}{q^2 R^2} \left[1 - \frac{J_1(2qR)}{qR} \right] \quad (3.65)$$

(H) **Vara** (comprimento H e seção transversal infinitesimal. Neugebauer, 1943)

$$P_{\text{vara}}(q) = (\rho_o V)^2 \left[\frac{\text{Si}(2qH)}{qH} - \frac{\sin^2(qH)}{q^2 H^2} \right], \quad \text{Si}(x) = \int_0^x \frac{\sin u}{u} du \quad (3.66)$$

Graficamos abaixo, figs. 3.1 e 3.2, alguns desses fatores de forma e suas respectivas funções de distribuição de pares $p(r)$, ambas normalizadas pelo quadrado do número total de elétrons n_e :

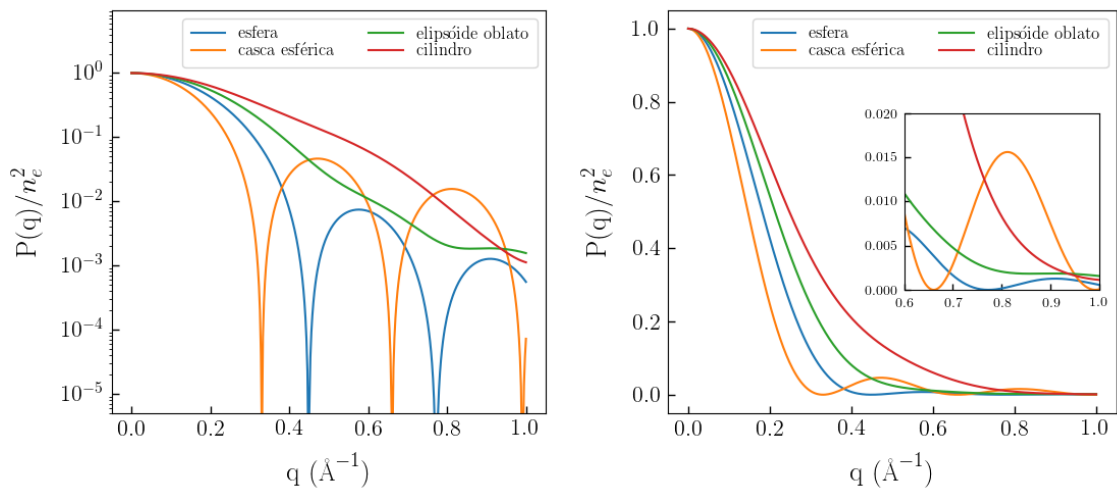


Figura 3.1: Fatores de forma de uma esfera (azul), casca esférica (amarelo), elipsóide oblato (verde) e cilindro (vermelho), todos com uma dimensão máxima de 20 Å. À direita, gráfico em escala logarítmica e, à esquerda, escala linear.

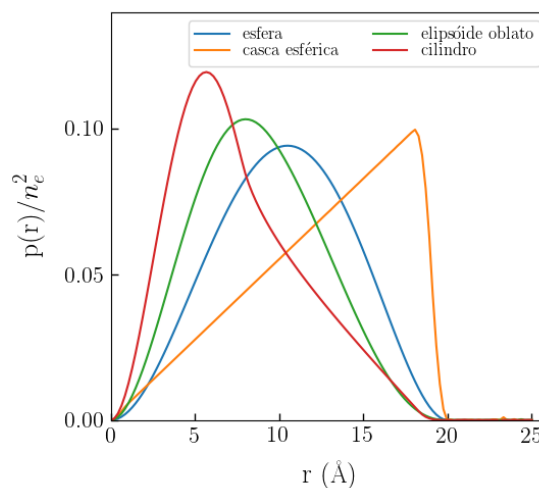


Figura 3.2: Funções de distribuição de pares das partículas descritas na figura 3.1.

3.3 Fator de Estrutura

Na seção anterior lidamos apenas com o primeiro termo da equação (3.35), ou seja, o fator de forma, o qual leva em consideração somente as interferências entre os elementos espalhadores de uma mesma partícula. Agora, vamos desenvolver o segundo termo daquela equação relativo às interferências cruzadas, a saber:

$$\langle F_p F_m^* e^{i\mathbf{q}\cdot(\mathbf{r}_p - \mathbf{r}_m)} \rangle = \int F_p F_m^* e^{i\mathbf{q}\cdot(\mathbf{r}_p - \mathbf{r}_m)} f_r(\mathbf{r}^N) \Omega^N d\mathbf{r}^N d\boldsymbol{\omega}^N \quad (3.67)$$

Na expressão acima usamos novamente a hipótese H_1 , o que nos permitirá separar as integrais como segue:

$$= \left(\frac{1}{8\pi^2} \int F_p d\boldsymbol{\omega}_p \right) \left(\frac{1}{8\pi^2} \int F_m^* d\boldsymbol{\omega}_m \right) \int e^{i\mathbf{q}\cdot(\mathbf{r}_p - \mathbf{r}_m)} f_r(\mathbf{r}^N) d\mathbf{r}^N \quad (3.68)$$

Usando a definição do delta de Dirac, podemos reescrever o fator exponencial como:

$$= \langle F_p \rangle \langle F_m^* \rangle \int \int \int e^{i\mathbf{q}\cdot(\mathbf{r} - \mathbf{r}')} \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) f_r(\mathbf{r}^N) d\mathbf{r}^N d\mathbf{r} d\mathbf{r}' \quad (3.69)$$

de modo que podemos definir o valor médio do produto dos deltas por:

$$\begin{aligned} \langle \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) \rangle &= \int \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) f_r(\mathbf{r}^N) d\mathbf{r}^N \\ &= \int f_r(\mathbf{r}, \mathbf{r}', \mathbf{r}^{N-2}) d\mathbf{r}^{N-2} \end{aligned} \quad (3.70)$$

Essa média corresponde à densidade de probabilidade de encontrarmos a partícula p na posição \mathbf{r} e a partícula m em \mathbf{r}' , simultaneamente. Assim, substituindo (3.69) e (3.70) na equação (3.35), obtemos

$$\langle \rho_q \rho_{-q} \rangle = \sum_p \langle |F_p|^2 \rangle + \sum_{p \neq m} \langle F_p \rangle \langle F_m^* \rangle \int \int e^{i\mathbf{q}\cdot(\mathbf{r} - \mathbf{r}')} \langle \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) \rangle d\mathbf{r} d\mathbf{r}' \quad (3.71)$$

Se o sistema em questão for composto de partículas idênticas, é possível simplificar a expressão anterior retirando-se as médias orientacionais das somatórias, ou seja,

$$= N \langle |F|^2 \rangle + \langle F \rangle \langle F^* \rangle \sum_{p \neq m} \int \int e^{i\mathbf{q}\cdot(\mathbf{r} - \mathbf{r}')} \langle \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) \rangle d\mathbf{r} d\mathbf{r}' \quad (3.72)$$

onde N é o número total de partículas irradiadas. Colocando o primeiro termo em evidência e permutando os símbolos de soma e integral, podemos reescrever a equação anterior na seguinte forma:

$$= N \langle |F|^2 \rangle \left(1 + \frac{|\langle F \rangle|^2}{N \langle |F|^2 \rangle} \int \int e^{i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} \left\langle \sum_{p \neq m} \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) \right\rangle d\mathbf{r} d\mathbf{r}' \right) \quad (3.73)$$

O termo entre parênteses em (3.73) é definido como o *fator de estrutura medido* $S_M(q)$ do sistema. Assim, usando também a definição do fator de forma (3.36), temos que:

$$\langle \rho_q \rho_{-q} \rangle = NP(q) S_M(q) \quad (3.74)$$

porém, precisamos desenvolver ainda mais a expressão do fator de estrutura medido para colocá-la numa forma que possa ser melhor trabalhada analiticamente. Para isso, faremos uso de algumas definições e hipóteses simplificadoras (H_2).

A média das funções delta que aparece na equação (3.73) é chamada de função *densidade de pares* (ou, *densidade de 2-partículas*) e representa a densidade de probabilidade de encontrarmos duas partículas quaisquer ocupando as posições \mathbf{r} e \mathbf{r}' , independente das posições das demais e de seus momentos. Como as partículas são idênticas, a média pode ser escrita como sendo o número total de pares multiplicado pela média sobre um único par qualquer:

$$\begin{aligned} \rho_N^{(2)}(\mathbf{r}, \mathbf{r}') &= \sum_{p \neq m} \langle \delta(\mathbf{r} - \mathbf{r}_p) \delta(\mathbf{r}' - \mathbf{r}_m) \rangle \\ &= N(N-1) \langle \delta(\mathbf{r} - \mathbf{r}_1) \delta(\mathbf{r}' - \mathbf{r}_2) \rangle \\ &= \frac{N!}{(N-2)!} \frac{1}{Z_N} \int e^{-\beta V(\mathbf{r}, \mathbf{r}', \mathbf{r}^{N-2})} d\mathbf{r}^{N-2} \end{aligned} \quad (3.75)$$

onde $\beta = 1/k_b T$ é o inverso da constante de Boltzmann multiplicada pela temperatura. Essa densidade de pares é apenas um caso particular das funções *densidade de n-partículas*:

$$\rho_N^{(n)}(\mathbf{r}^n) = \frac{N!}{(N-n)!} \frac{1}{Z_N} \int e^{-\beta V(\mathbf{r}^n, \mathbf{r}^{N-n})} d\mathbf{r}^{N-n} \quad (3.76)$$

as quais formam uma hierarquia⁴ de funções, de tal modo que uma densidade de ordem

⁴As funções de densidade estão relacionadas entre si, na teoria de líquidos, através da chamada hierarquia YBG (Yvon-Born-Green), no caso de um sistema em equilíbrio termodinâmico. Nessa hierarquia, podemos escrever uma função densidade em termos de outra de grau superior. A solução de tal equação, entretanto, ainda necessitaria de alguma outra relação (de fechamento) independente.

n pode ser escrita em termos da densidade de ordem $n + 1$ por meio da integração de uma variável de posição. Para os nossos propósitos, aqui, somente estaremos interessados nas densidades de primeira e de segunda ordem. De acordo com (3.76), a densidade de 1-partícula é dada por:

$$\rho_N^{(1)}(\mathbf{r}) = \frac{N}{Z_N} \int e^{-\beta V(\mathbf{r}, \mathbf{r}^{N-1})} d\mathbf{r}^{N-1} = N \langle \delta(\mathbf{r} - \mathbf{r}_1) \rangle \quad (3.77)$$

de maneira que sua integral sobre todas as posições será:

$$\int \rho_N^{(1)}(\mathbf{r}) d\mathbf{r} = \frac{N}{Z_N} \int e^{-\beta V(\mathbf{r}^N)} d\mathbf{r}^N = N \quad (3.78)$$

Além disso, se o sistema for homogêneo, no sentido de que a probabilidade de um determinado elemento de volume $d\mathbf{r}$ estar ocupado é independente de sua posição \mathbf{r} , então, a integral em (3.77) também pode ser resolvida como:

$$\int \rho_N^{(1)}(\mathbf{r}) d\mathbf{r} = \rho_N^{(1)} \int d\mathbf{r} = \rho_N^{(1)} V \quad \Rightarrow \quad \rho_N^{(1)} = \frac{N}{V} = n \quad (3.79)$$

ou seja, a densidade de 1-partícula para um sistema homogêneo no equilíbrio é simplesmente igual à densidade total do sistema; no caso heterogêneo ela fornece a densidade local numa dada região infinitesimal. Estamos agora em condições de definir a chamada *função de distribuição radial* (ou, *função distribuição de 2-partículas*):

$$g_N^{(2)}(\mathbf{r}, \mathbf{r}') = \frac{\rho_N^{(2)}(\mathbf{r}, \mathbf{r}')}{\rho_N^{(1)}(\mathbf{r}) \rho_N^{(1)}(\mathbf{r}')} \quad (3.80)$$

a qual, para um sistema homogêneo, implica que

$$\rho_N^{(2)}(\mathbf{r}, \mathbf{r}') = n^2 g_N^{(2)}(\mathbf{r}, \mathbf{r}') \quad (3.81)$$

Para interpretarmos fisicamente a função de distribuição radial vamos considerar a seguinte razão:

$$\begin{aligned} \frac{\rho_N^{(2)}(\mathbf{r}, \mathbf{r}')}{\rho_N^{(1)}(\mathbf{r}')} &= (N - 1) \frac{\int e^{-\beta V(\mathbf{r}, \mathbf{r}', \mathbf{r}^{N-2})} d\mathbf{r}^{N-2}}{\int e^{-\beta V(\mathbf{r}', \mathbf{r}^{N-1})} d\mathbf{r}^{N-1}} \\ &= (N - 1) \int f_r(\mathbf{r}, \mathbf{r}^{N-2} | \mathbf{r}') d\mathbf{r}^{N-2} = (N - 1) \langle \delta(\mathbf{r} - \mathbf{r}_1 | \mathbf{r}') \rangle \end{aligned} \quad (3.82)$$

onde $f_r(\mathbf{r}, \mathbf{r}^{N-2} | \mathbf{r}')$ é a densidade de probabilidade de uma determinada partícula estar em $d\mathbf{r}$ dado que há uma partícula em $d\mathbf{r}'$. Em comparação com a equação (3.77), a equação (3.82) representa a densidade de partículas no elemento de volume $d\mathbf{r}$ dado que há uma partícula em $d\mathbf{r}'$ (obs. estamos considerando elementos de volume pequenos o

suficiente para caber o centro de uma única molécula). Logo, a função $g_N^{(2)}(\mathbf{r}, \mathbf{r}')$ é igual à densidade relativa de partículas numa região que dista $\mathbf{u} = \mathbf{r} - \mathbf{r}'$ de uma partícula localizada em \mathbf{r}' :

$$g_N^{(2)}(\mathbf{r}, \mathbf{r}') = \frac{(N-1)\langle \delta(\mathbf{r} - \mathbf{r}_1 | \mathbf{r}') \rangle}{N\langle \delta(\mathbf{r} - \mathbf{r}_1) \rangle} = \frac{\rho(\mathbf{r} | \mathbf{r}')}{\rho(\mathbf{r})} \quad (3.83)$$

sendo assim, essa função deve ser nula para distâncias u muito curtas, nas quais ocorreria sobreposição entre as partículas, e tende a 1 para distâncias longas o bastante para que a correlação entre elas seja desprezível.

Como estamos interessados em sistemas homogêneos, podemos usar (3.81) para escrever o fator de estrutura medido como:

$$\begin{aligned} S_M(q) &= 1 + \frac{|\langle F \rangle|^2}{N\langle |F|^2 \rangle} \int \int e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} \rho_N^2(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}' \\ &= 1 + \frac{|\langle F \rangle|^2 n^2}{\langle |F|^2 \rangle N} \int \int e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')} g_N^2(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}' \end{aligned} \quad (3.84)$$

Se, além de homogêneo, o sistema também for isotrópico, no sentido de que as propriedades locais numa dada posição são as mesmas para todas as orientações do vetor posição, então a função de distribuição radial depende exclusivamente da distância entre os pontos considerados:

$$g_N^2(\mathbf{r}, \mathbf{r}') = g(|\mathbf{r} - \mathbf{r}'|) = g(u) \quad (3.85)$$

Assim, substituindo (3.85) em (3.84) e usando a mudança de variável $\mathbf{u} = \mathbf{r} - \mathbf{r}'$, ficamos com:

$$\begin{aligned} S_M(q) &= 1 + \frac{|\langle F \rangle|^2 n^2}{\langle |F|^2 \rangle N} \int \int e^{i\mathbf{q}\cdot\mathbf{u}} g(u) d\mathbf{u} d\mathbf{r}' \\ &= 1 + \frac{|\langle F \rangle|^2}{\langle |F|^2 \rangle} n \int g(u) e^{i\mathbf{q}\cdot\mathbf{u}} d\mathbf{u} = 1 + \frac{|\langle F \rangle|^2}{\langle |F|^2 \rangle} n \hat{g}(q) \end{aligned} \quad (3.86)$$

sendo $\hat{g}(q)$ a transformada de Fourier da função de distribuição radial, a qual pode ser escrita, após feitas as integrais nas orientações, como:

$$\hat{g}(q) = \int_0^\infty 4\pi u^2 g(u) \frac{\sin(uq)}{uq} du \quad (3.87)$$

Fisicamente, a quantidade $4\pi u^2 n g(u) du$ representa o número de partículas contidas na casca esférica de raio u e centrada numa dada partícula de referência. Assim, a

quantidade $n\hat{g}(q)$ pode ser entendida como uma soma ponderada sobre as partículas equidistantes de uma certa partícula de referência. Finalmente, se as partículas tiverem simetria radial e, portanto, $|\langle F \rangle|^2 = \langle |F|^2 \rangle$, a equação (3.86) se reduz a:

$$S(q) = 1 + n\hat{g}(q) \quad (3.88)$$

Essa última função é chamada simplesmente de *fator de estrutura* e equivale ao fator de estrutura medido apenas no caso simétrico (embora seja uma aproximação razoável para partículas globulares). Muitas vezes, é mais conveniente escrever (3.88) em termos da chamada *função de correlação de pares*, $h = g - 1$, por razões que ficarão mais claras na próxima seção:

$$S(q) = 1 + n(\hat{h} + \hat{1}) = 1 + n\hat{h}(q) + n(2\pi)^3\delta(\mathbf{q}) \quad (3.89)$$

onde o termo da função delta em geral pode ser ignorado, uma vez que ele só é relevante para ângulos de espalhamento quase nulos, inacessíveis do ponto de vista experimental.

Resumo da teoria básica

Os resultados das três últimas seções podem ser agora sintetizados. Se substituirmos a equação (3.74) na equação geral (3.25), temos que a intensidade de espalhamento de um sistema de partículas monodisperso e sujeito às hipóteses H_1 (independência das posições e orientações) e H_2 (sistema homogêneo e isotrópico) será dada por:

$$\frac{d\Sigma}{dq} = nP(q)S_M(q) \quad (3.90)$$

onde

$$P(q) = \int_0^\infty p(u) \frac{\sin(uq)}{uq} du \quad \text{e} \quad S_M(q) = 1 + \frac{|\langle F \rangle|^2}{\langle |F|^2 \rangle} [S(q) - 1]$$

são os fatores de forma e de estrutura medido, respectivamente, conforme apresentados nas equações (3.43) e (3.86), e $S(q) = 1 + n\hat{g}(q) \approx 1 + n\hat{h}(q)$ é o fator de estrutura.

Para sistemas suficientemente diluídos ($n \rightarrow 0$) o fator de estrutura, assim como o fator de estrutura medido, tende à unidade, $S(q) \rightarrow 1$, de modo que a função de espalhamento passa a depender basicamente da geometria das partículas, expressa pelo seu fator de forma. Já para sistemas densos os efeitos do fator de estrutura não podem ser desprezados e, portanto, precisamos de alguns métodos teóricos para lidar com esse termo, os quais serão apresentados na próxima seção.

3.4 Relações de Fechamento

Na teoria dos fluidos é conveniente definir as chamadas *funções de correlação direta* $c^{(n)}(\mathbf{r}^n)$. Isso pode ser feito de maneira recursiva pela definição da função de primeira ordem:

$$c^{(1)}(\mathbf{r}) = \ln[\Lambda^3 \rho^{(1)}(\mathbf{r})] - \beta\psi(\mathbf{r}) \quad (3.91)$$

onde Λ é o comprimento de onda térmico de de Broglie, $\rho^{(1)}(\mathbf{r}) = \langle \rho(\mathbf{r}) \rangle$ é a média grande canônica da densidade de partículas e $\psi(\mathbf{r}) = \mu - \phi(\mathbf{r})$ é o potencial químico subtraído de um eventual campo externo. As demais funções de correlação direta são, então, dadas pela derivada funcional de suas predecessoras:

$$c^{(n+1)}(\mathbf{r}^{(n+1)}) = \frac{\delta c^{(n)}(\mathbf{r}^{(n)})}{\delta \rho^{(1)}(\mathbf{r})} \quad (3.92)$$

O que nos interessa aqui é que a função de correlação de pares total $h^{(2)}(\mathbf{r}, \mathbf{r}')$ está relacionada com a função de correlação direta de pares através da equação de Ornstein-Zernike [36]:

$$h^{(2)}(\mathbf{r}, \mathbf{r}') = c^{(2)}(\mathbf{r}, \mathbf{r}') + \int c^{(2)}(\mathbf{r}, \mathbf{r}'') \rho^{(1)}(\mathbf{r}'') h^{(2)}(\mathbf{r}'', \mathbf{r}') d\mathbf{r}'' \quad (3.93)$$

Para um sistema homogêneo e isotrópico, essa equação se escreve de uma forma mais simples como:

$$h(r_{12}) = c(r_{12}) + n \int c(r_{13}) h(r_{23}) d\mathbf{r}_{13} \quad (3.94)$$

onde as variáveis estão representadas na fig. 3.3, abaixo:

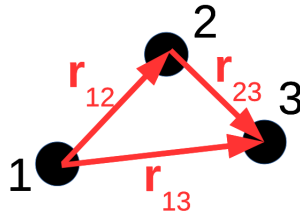


Figura 3.3: Variáveis presentes na equação de Ornstein-Zernike de um fluido homogêneo e isotrópico.

Ela expressa, portanto, a correlação de pares $h(r_{12})$ em termos da correlação direta $c(r_{12})$ entre duas partículas, acrescido de uma dependência com correlações indiretas que se propagam para as outras partículas do sistema, uma vez que $h(r_{23})$ se escreve em termos de $h(r_{34})$ pela mesma equação, $h(r_{34})$ em termos de $h(r_{45})$ e assim por diante. Chamando $r = r_{12}$, $r' = r_{13}$ e aplicando uma transformada de Fourier nos dois lados da equação 3.94, obtemos como resultado:

$$\hat{h}(q) = \hat{c}(q) + n \int c(r') \left[\int e^{i\mathbf{q}\cdot\mathbf{r}} h(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r} \right] d\mathbf{r}' \quad (3.95)$$

onde podemos usar a mudança de variável $\mathbf{r}'' = \mathbf{r} - \mathbf{r}'$ e chegar numa versão algébrica da equação de Ornstein-Zernike (OZ):

$$\begin{aligned} \hat{h}(q) &= \hat{c}(q) + n \int c(r') \left[\int e^{i\mathbf{q}\cdot(\mathbf{r}'+\mathbf{r}'')} h(r'') d\mathbf{r}'' \right] d\mathbf{r}' \\ &= \hat{c}(q) + n \left[\int e^{i\mathbf{q}\cdot\mathbf{r}'} c(r') d\mathbf{r}' \right] \left[\int e^{i\mathbf{q}\cdot\mathbf{r}''} h(r'') d\mathbf{r}'' \right] \\ &= \hat{c}(q) + n\hat{c}(q)\hat{h}(q) \end{aligned} \quad (3.96)$$

Assim, isolando a transformada da função de correlação, temos que:

$$\hat{h}(q) = \frac{\hat{c}(q)}{1 - n\hat{c}(q)} \quad (3.97)$$

Não há teoria até o momento que seja capaz de resolver a equação acima de forma exata, de modo que dependemos de aproximações adicionais que a tornem manejável. Chama-se uma *relação de fechamento* qualquer expressão independente de OZ que expresse $c(r)$ em termos de $h(r)$ para um dado potencial de pares $u(r)$. Junto com a equação de OZ elas levam a uma equação integral fechada (i.e., solúvel) para $h(r)$. Algumas dessas relações serão apresentadas a seguir.

Percus-Yevick (PY)

A aproximação de Percus-Yevick (1958) [37] é dada por:

$$c(r) = g(r) \left[1 - e^{\beta u(r)} \right] \quad (3.98)$$

Em geral, a PY é uma relação muito bem sucedida quando aplicada a partículas do tipo esfera-dura, inclusive acrescidas de interações de curto alcance. Para um potencial de esferas-duras (ED), $u(r < 2R) = \infty$, $u(r > 2R) = 0$ e $g(r < 2R) = 0$, logo, a função de correlação direta será:

$$c(r) = \begin{cases} g(r)e^{\beta u(r)}, & \text{se } r < 2R \\ 0, & \text{se } r > 2R \end{cases} \quad (3.99)$$

onde R é o raio das esferas. Em regimes de baixa densidade, é possível demonstrar⁵ que

⁵Para uma solução sucinta cf. Hansen e MacDonald (2013) [36], Apêndice D.

a solução da equação de OZ, com esse fechamento, vale exatamente:

$$c(r) = a + br + cr^3 \quad (3.100)$$

na região $r < 2R$, onde

$$a = -\frac{(1+2\eta)^2}{(1-\eta)^4} \quad b = \frac{3\eta(4+2\eta)^2}{(1-\eta)^4} \frac{1}{2R} \quad c = -\frac{\eta(1+2\eta)^2}{2(1-\eta)^4} \frac{1}{(2R)^3} \quad (3.101)$$

resultado que é válido com boa concordância, quando comparado com simulações de Monte Carlo, para frações volumétricas $\eta = n4\pi R^3/3 < 0.3$.

Repare que se substituirmos a expressão 3.97 em 3.89, podemos obter o fator de estrutura a partir da função de correlação direta:

$$S(q) \approx 1 + nh\hat{c}(q) = \frac{1}{1 - n\hat{c}(q)} \quad (3.102)$$

Assim, o fator de estrutura do fluido de esferas-duras relativamente diluído, $\eta < 0.3$, será dado pela transformada:

$$\begin{aligned} \hat{c}(q) &= \int_0^\infty 4\pi r^2 c(r) \frac{\sin(qr)}{qr} dr \\ &= \frac{4\pi}{q} \int_0^{2R} [ar + br^2 + cr^4] \sin(qr) dr \\ &= \frac{12\eta[\eta(3-\eta^2) - 2]}{n(1-\eta)^4} \frac{\Phi(2qR)}{2qR} \end{aligned} \quad (3.103)$$

onde $\Phi(x) = (\sin x - x \cos x)/x^2$. Logo,

$$[S(q)]^{-1} = 1 - \frac{12\eta[\eta(3-\eta^2) - 2]}{(1-\eta)^4} \frac{\Phi(2qR)}{2qR} \quad (3.104)$$

Mean-Spherical Approximation (MSA)

Introduzida na teoria do estado líquido primeiramente por Percus e Lebowitz (1966) [38], a aproximação MSA assume que:

$$\begin{cases} c(r) = -\beta u(r), & \text{se } r > 2R \\ g(r) = 0, & \text{se } r < 2R \end{cases} \quad (3.105)$$

Juntas, elas transformam OZ numa equação integral linear que permite determinar $c(r < 2R)$ e $g(r > 2R)$.

A relação de fechamento MSA fornece soluções analíticas para vários modelos de potencial de pares, como o poço quadrado, Coulombiano, Yukawa, esferas com momento de dipolo, dentre outros. Em geral, a MSA é mais apropriada para lidar com potenciais de curto alcance. Já para sistemas altamente repulsivos, ela pode levar a valores negativos de $g(r)$ próximos à superfície de contato, o que claramente não possui significado físico, pois $g(r) \geq 0$.

Para sistemas com baixa fração volumétrica η , a MSA prevê que:

$$g(r) \approx 1 + c(r) = 1 - \beta u(r) \quad \text{para } r > 2R \quad (3.106)$$

e, portanto, valores negativos de $g(r)$ quando $\beta u(r) > 1$.

Random Phase Approximation (RPA)

Suponha que possamos separar o potencial de pares de um fluido em dois termos:

$$u(r) = u_o(r) + \phi(r) \quad (3.107)$$

sendo o primeiro o potencial de um sistema de referência e o segundo um potencial perturbativo Fourier-integrável. Vamos assumir ainda que a função de correlação direta $c_o(r)$ do sistema de referência é conhecida de antemão. Assim, a relação de fechamento RPA estabelece para a correlação direta do sistema perturbado a seguinte expressão:

$$c(r) = c_o(r) - \beta \phi(r) \quad (3.108)$$

A vantagem desse tipo de abordagem é que podemos relacionar diretamente o fator de estrutura com o potencial perturbativo por meio da equação 3.102:

$$S(q) = \frac{1}{1 - n\hat{c}_o(q) + n\beta\hat{\phi}(q)} = \frac{1}{[S_o(q)]^{-1} + n\beta\hat{\phi}(q)} \quad (3.109)$$

Podemos utilizar como sistema de referência, por exemplo, o fluido de esferas duras, cujo fator de estrutura é dado em 3.104. Além disso, devido a uma ambiguidade na escolha do potencial perturbativo para $r < 2R$, isto é, na região onde o potencial de referência é infinito, a relação RPA não garante que teremos $g(r < 2R) = 0$ sempre.

Hypernetted-chain (HNC)

Desenvolvida por Leeuwen et al. (1959) [39], a relação HNC leva seu nome devido ao método de derivação por diagramas. Ela se baseia na seguinte aproximação de fechamento:

$$c(r) = -\beta u(r) + h(r) - \ln [1 + h(r)] \quad (3.110)$$

A equação de OZ fechada com a relação HNC só pode ser resolvida numericamente, mesmo no caso mais simples de esferas-duras. Assim como a Percus-Yevick, a HNC prevê com boa precisão a $g(r)$ em baixas densidades. Porém, enquanto a PY consegue prever o comportamento da $c(r)$ em distâncias longas somente em sistemas diluídos, a HNC consegue uma boa predição assintótica independentemente da densidade. Em contraste com a MSA, ela não consegue bons resultados para potenciais com uma cauda atrativa. A HNC também possui uma performance inferior à PY para lidar com o fluido de esferas-duras, mas consegue lidar bem com repulsões de longo alcance, como as de fluidos iônicos.

Outras relações de fechamento também podem ser encontradas na literatura, como a Rescaled MSA (RMSA) introduzida por Hansen e Hayter (1982) [40], a Rogers-Young (RY, 1984) [41] e outras aproximações perturbativas [42, 27].

Metodologia: Monte Carlo

Neste capítulo, vamos tratar as soluções proteicas como um fluido de partículas carregadas, dentro do escopo do chamado modelo primitivo. A estrutura líquida teórica das partículas será então calculada por meio de simulações de Monte Carlo (MC) no ensemble canônico. Para isso, faremos uso da abordagem convencional de Metropolis, para aceitar ou rejeitar estados durante as simulações, e utilizaremos a técnica da soma de Ewald para lidar com as interações eletrostáticas.

1 Modelagem computacional

1.1 O modelo primitivo

Uma das abordagens teóricas mais difundidas para modelar um sistema de partículas carregadas é feita através do chamado modelo primitivo (PM), no qual o potencial de pares pode ser expresso de maneira genérica como a soma de um termo de esferas-duras (hard-sphere, HS) e o potencial de Coulomb:

$$\phi_{ij}(r_{ij}) = \phi_{\text{HS}} + \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (4.1)$$

onde ϵ é a permissividade do meio, considerado um contínuo homogêneo. Quando as partículas possuem o mesmo diâmetro e cargas opostas, respeitando sempre a condição de neutralidade, dizemos que se trata do modelo primitivo restrito (RPM).

Uma versão do modelo primitivo foi aplicada com sucesso, por exemplo, na teoria de Debye-Hückel (1923) [43, 44], a qual representou um marco no estudo de soluções iônicas. Na abordagem dos autores, uma solução composta de íons esféricos, cujas cargas estão concentradas pontualmente nos seus centros, está sujeita à equação de Helmholtz (ou, Poisson-Boltzmann linearizada, cf. Apêndice A):

$$\nabla^2 \psi_j(r) = \kappa^2 \psi_j(r) \quad \text{com} \quad \kappa^2 = \frac{\beta e^2}{\epsilon} \sum_i n_i z_i^2 \quad (4.2)$$

sendo ψ_j o potencial eletrostático, tendo o íon do tipo j como origem do sistema de

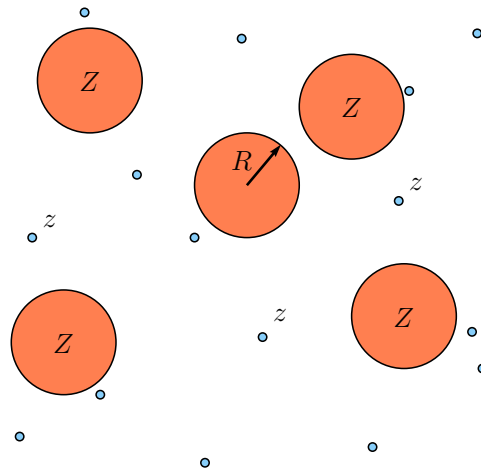


Figura 4.1: Modelo primitivo com macroíons (carga Z e raio R) e microíons pontuais de carga $z = \pm 1$.

coordenadas, e n_i a concentração numérica dos íons da espécie i . Aplicando as devidas condições de contorno para distâncias grandes ($r \rightarrow \infty$) e na distância de menor aproximação dos íons ($r = a$, igual a um diâmetro se os íons tiverem o mesmo raio), eles obtém a seguinte expressão para a energia potencial média de pares:

$$u_{ij}(r) = ez_i\psi_j(r) = \frac{e^2 z_i z_j}{4\pi\epsilon(1 + \kappa a)} \frac{e^{-\kappa(r-a)}}{r} \quad (r > a) \quad (4.3)$$

a qual inclui o efeito da blindagem eletrostática pela atmosfera iônica ao redor das partículas. Outra conquista dessa teoria foi a derivação de uma lei limite para o coeficiente de atividade em baixas concentrações, $\ln \gamma_j \rightarrow -z_j^2 e^2 \kappa / 2\epsilon k_b T$ ($\kappa \rightarrow 0$), alcançando na época um bom acordo com os experimentos [44].

Da mesma forma, a teoria DLVO [19, 20] também se baseia no modelo primitivo, no caso considera-se esferas carregadas de uma mesma espécie imersas em uma solução iônica de cargas pontuais. A resolução proposta leva em conta o efeito da aproximação entre duas esferas sobre a atmosfera iônica ao redor das mesmas, e deriva, como primeira aproximação, o famoso potencial (cf. Apêndice B):

$$u(r) = \frac{e^2 z_i z_j}{4\pi\epsilon(1 + 0.5\kappa a)^2} \frac{e^{-\kappa(r-a)}}{r} \quad (4.4)$$

válido para soluções diluídas e tais que o potencial elétrico seja pequeno $e\psi/k_b T \ll 1$. A expressão é muito parecida com aquela obtida pela teoria de Debye-Hückel, com exceção do fator $(1 + \kappa a)$ no denominador, que agora aparece ao quadrado e com metade do diâmetro.

À semelhança da abordagem DLVO, iremos tratar nossas soluções proteicas a partir

de um modelo primitivo que considera cada proteína (macroión) uma esfera com raio e carga definidos, enquanto que os microíons – compostos de contraíons (H^+ ou OH^-) e os íons dos sais – serão tratados como cargas pontuais (ver figura 4.1). Como o principal sal utilizado nos experimentos foi o NaCl, iremos considerar aqui somente microíons monovalentes, isto é, com cargas +1 ou -1. A carga total será sempre nula, repetindo a hipótese de neutralidade. Para soluções puras (ou seja, uma única espécie de macroión) e sem a adição de sal, os microíons consistirão somente de contraíons monovalentes com sinal oposto ao dos macroíons. Para soluções com sal, além dos contraíons serão incluídos microíons monovalentes positivos (Na^+) e negativos (Cl^-). Finalmente, as misturas binárias serão compostas por dois tipos de macroíons, com raios R_1 e R_2 e cargas Z_1 e Z_2 , seus respectivos contraíons e, dependendo do caso, os microíons de sal.

Para um sistema simples, os parâmetros que definem o modelo são: número de macroíons (N_M), carga dos macroíons (Z_M), carga dos microíons (Z_I), raio do macroión (R_M), concentração de macroíons (ρ_M), concentração de sal (ρ_s), temperatura (T) e a permissividade do meio (ϵ), sendo que o número de microíons (N_I) já fica determinado pela hipótese de neutralidade. Apesar do modelo possuir essas 8 variáveis independentes, pode-se descrevê-lo em termos de apenas três parâmetros reduzidos, suficientes para determinar toda a sua termodinâmica [45, 46]:

$$Z_r = -\frac{Z_M}{Z_I} \quad \Phi_M = \frac{4\pi R_M^3}{3} \rho_M \quad \Gamma_{II} = \frac{\lambda_B Z_I^2}{R_M} \quad (4.5)$$

onde o primeiro é a assimetria de cargas, o segundo a fração volumétrica dos macroíons e o terceiro um fator de acoplamento eletrostático. Essa escolha se deve ao fato de podermos reescrever a energia total de um sistema coulombiano como:

$$\begin{aligned} \beta E_C &= \frac{\beta e^2}{4\pi\epsilon} \sum_{(i,j)} \frac{z_i z_j}{r_{ij}} = \lambda_B \left[\sum_{MM} \frac{Z_M^2}{r_{ij}} + \sum_{MI} \frac{Z_M Z_I}{r_{ij}} + \sum_{II} \frac{Z_I^2}{r_{ij}} \right] \\ &= \Gamma_{II} \left[\sum_{MM} \frac{Z_r^2}{R_{ij}} + \sum_{MI} \frac{Z_r}{R_{ij}} + \sum_{II} \frac{1}{R_{ij}} \right] \end{aligned} \quad (4.6)$$

onde λ_B é o comprimento de Bjerrum, $R_{ij} = r_{ij}/R_M$ e a soma sobre pares (i,j) foi dividida em três somatórias sobre pares de macroíons (M) e microíons (I).

1.2 Simulação

Uma simulação de Monte Carlo do nosso modelo consiste em gerar uma sequência de configurações que representem de maneira ponderada o espaço de fase das posições. De maneira sucinta, as simulações partem de um estado inicial aleatório e certas condições de contorno, entrando em seguida num loop onde vários estados são gerados segundo

algum algoritmo de movimentação. As configurações são atualizadas em cada etapa, submetendo esses estados a um teste que garante a distribuição de probabilidades no ensemble. Aqui, faremos uso da abordagem convencional de Metropolis [47] para aceitar ou rejeitar estados no ensemble canônico. Em geral, o conjunto de etapas necessário para produzir duas configurações descorrelacionadas é chamado de passo de Monte Carlo.

Já para o cálculo da energia necessitaremos de um tratamento especial, por duas razões: (i) é a etapa mais custosa computacionalmente, pois envolve o cálculo repetitivo das interações de cada par de partículas e (ii) estaremos simulando sistemas carregados, o que implica em certos cuidados de convergência. Assim, para lidarmos com as interações eletrostáticas, utilizaremos a técnica da chamada soma de Ewald [48]. Descreveremos de maneira sucinta, nas próximas seções, os métodos citados e alguns detalhes de sua implementação.

2 Cadeia de Markov e o algoritmo de Metropolis

Uma amostragem de configurações de um dado sistema físico pode ser ordenada em termos de uma sequência, onde cada índice i representa um estado $\nu(i)$. Uma maneira interessante de gerar uma sequência como essa é através de um processo dito markoviano. A ideia consiste em tratar $\nu(i)$ como uma variável aleatória e associarmos a ela uma probabilidade $q_\nu(i)$ dela estar no estado ν na i -ésima posição, de tal modo que essa probabilidade dependa unicamente do estado anterior $\nu(i-1)$. Dizemos, então, que a sequência é uma cadeia de Markov [49, 50]. Definida desta maneira, teremos que:

$$q_\nu(i) = \sum_{\tau} p_{\nu \leftarrow \tau} q_\tau(i-1) \quad (4.7)$$

onde $p_{\nu \leftarrow \tau}$ é a taxa de transição para o estado ν dado que o estado anterior é τ . Obviamente,

$$\sum_{\tau} p_{\tau \leftarrow \nu} = 1 \quad (4.8)$$

Utilizando uma notação matricial, na qual a probabilidade $q_\nu(i)$ de cada estado é visto como um elemento do vetor coluna $\mathbf{q}(i)$ e $p_{\nu \leftarrow \tau}$ é um elemento de uma *matriz de transição* \mathbf{p} , podemos reescrever 4.7 de maneira geral como:

$$\mathbf{q}(i) = \mathbf{p} \cdot \mathbf{q}(i-1) \quad (4.9)$$

Assim, se aplicarmos 4.9 recursivamente até chegarmos à distribuição inicial de estados $\mathbf{q}(0)$, teremos que:

$$\mathbf{q}(i) = \mathbf{p}^i \cdot \mathbf{q}(0) \quad (4.10)$$

Para o caso em que a matriz de transição \mathbf{p} é *ergódica*, isto é, existe um natural n que

torna todos os elementos da matriz \mathbf{p}^n não-nulos (ou, em outras palavras, se qualquer estado pode ser atingido num "tempo" finito a partir de qualquer outro estado), então é possível provar que existe o limite:

$$\mathbf{Q} = \lim_{i \rightarrow \infty} \mathbf{p}^i \cdot \mathbf{q}(0) \quad (4.11)$$

de modo que, a partir de algum índice da cadeia, podemos considerar a probabilidade Q_ν de ocorrência do estado ν praticamente constante. Nessa condição estacionária, portanto, deve valer:

$$\mathbf{Q} = \mathbf{p} \cdot \mathbf{Q} \quad \Leftrightarrow \quad Q_\nu = \sum_{\tau} p_{\nu \leftarrow \tau} Q_\tau \quad \forall \nu \quad (4.12)$$

A partir da condição de normalização 4.8 e da condição de estado estacionário 4.12, podemos derivar a seguinte expressão:

$$\sum_{\tau} p_{\tau \leftarrow \nu} = 1$$

$$\Leftrightarrow Q_\nu \sum_{\tau} p_{\tau \leftarrow \nu} = Q_\nu \quad (4.13)$$

$$\Leftrightarrow \sum_{\tau} p_{\tau \leftarrow \nu} Q_\nu = \sum_{\tau} p_{\nu \leftarrow \tau} Q_\tau \quad (4.14)$$

Portanto, uma condição equivalente para a estacionaridade da cadeia é dada pela igualdade:

$$\sum_{\tau} [p_{\tau \leftarrow \nu} Q_\nu - p_{\nu \leftarrow \tau} Q_\tau] = 0 \quad (4.15)$$

Em geral, sabemos de antemão as probabilidades de equilíbrio. No ensemble canônico, por exemplo, elas devem seguir a distribuição de Boltzmann $Q_\nu = e^{-\beta E_\nu} / Z$. Assim, se formos capazes de encontrar $p_{\tau \leftarrow \nu}$ e $p_{\nu \leftarrow \tau}$ tais que satisfaçam a equação anterior, teremos garantido a condição de equilíbrio. Uma solução possível é aquela que satisfaz o chamado *balanceamento detalhado*, no qual impomos que cada termo da somatória seja nulo, ou seja:

$$p_{\tau \leftarrow \nu} Q_\nu = p_{\nu \leftarrow \tau} Q_\tau \quad \forall \tau, \nu \quad (4.16)$$

de modo que, no ensemble canônico, deve valer:

$$\frac{p_{\nu \leftarrow \tau}}{p_{\tau \leftarrow \nu}} = \frac{Q_\nu}{Q_\tau} = e^{-\beta(E_\nu - E_\tau)} \quad (4.17)$$

A equação anterior ainda não determina as taxas de transição de maneira unívoca. A proposta mais difundida e empregada nas rotinas de simulação é aquela que ficou conhecida como algoritmo de Metropolis (1953) [47], na qual se estabelece o seguinte

critério para as taxas:

$$p_{\nu \leftarrow \tau} = \begin{cases} e^{-\beta(E_\nu - E_\tau)} & \text{se } E_\nu > E_\tau \\ 1 & \text{se } E_\nu \leq E_\tau \end{cases} \quad (4.18)$$

assim, uma vez que o sistema esteja numa dada configuração τ e se produza uma nova configuração ν a partir dela, caso a energia do sistema diminua ou não varie ($E_\nu \leq E_\tau$) então a transição é sempre aceita; se a energia aumentar ($E_\nu > E_\tau$), então a transição ocorrerá com uma probabilidade $e^{-\beta(E_\nu - E_\tau)} < 1$. Em suma, $p_{\nu \leftarrow \tau} = \min(1, e^{-\beta\Delta E})$. O mesmo raciocínio pode ser facilmente estendido para outros tipos de ensemble, uma vez que se conheça as respectivas distribuições de equilíbrio.

O cálculo preciso da energia é, portanto, uma etapa fundamental em qualquer simulação de Monte Carlo que utilize o algoritmo de Metropolis. Além disso, como essa etapa é repetida numerosas vezes em cada ciclo do processo ela é, em geral, o principal limitante em termos do custo computacional. Na próxima seção discutimos como este cálculo pode ser feito em sistema carregados e como otimizá-lo.

3 Soma de Ewald

A energia total de um sistema de N partículas que interagem via potencial de Coulomb é dada pela seguinte equação:

$$E_C = \frac{e^2}{4\pi\epsilon} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{z_i z_j}{|\mathbf{r}_{ij}|} \quad (4.19)$$

sendo $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ e as somatórias percorrendo todos os pares (i, j) sem repetição.

No entanto, simulações de fluidos carregados são geralmente feitas com um número N limitado de partículas (algumas dezenas ou centenas) contidas em uma caixa com condições periódicas de contorno. Por um lado, isso se deve ao fato de que a quantidade de interações cresce com $\mathcal{O}(N^2)$, o que faz com que o tempo das simulações aumente muito rapidamente com o número de partículas. Por outro lado, simulações com poucas partículas podem produzir distorções significativas, como efeitos de borda, em volumes de paredes rígidas, de modo que a utilização de contornos periódicos é uma saída conveniente e que ajuda a aproximar os resultados do limite termodinâmico. Assim, para os propósitos de uma simulação, a energia de uma célula de N partículas deve levar em conta não apenas as interações entre elas, mas também delas com as suas infinitas réplicas vizinhas:

$$E_C = \frac{e^2}{4\pi\epsilon} \sum_{\mathbf{n}}' \left(\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{z_i z_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right) \quad (4.20)$$

onde $\mathbf{n} = (n_x, n_y, n_z)$ é um vetor deslocamento que indexa cada célula da rede e o

símbolo prima na somatória indica que quando $\mathbf{n} = (0,0,0)$, ou seja, quando indexa a célula principal, devemos desconsiderar os casos em que $i = j$ na soma dupla entre parênteses. O fator 2 dividindo aparece, agora, para não contarmos as mesmas interações duas vezes.

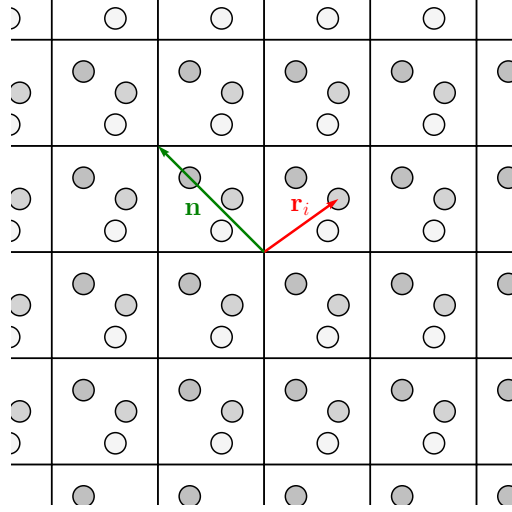


Figura 4.2: Rede usada no cálculo da energia de uma célula unitária.

A princípio, poderíamos ficar tentados a efetuar o cálculo da equação 4.20 por meio de algum tipo de truncamento, desprezando interações muito distantes umas das outras, mas o caso não se resolve assim tão facilmente. O uso de contornos periódicos em sistemas carregados coloca dificuldades adicionais que sistemas sem cargas ou com interações de curto alcance não enfrentam. Para o cálculo da energia coulombiana, por exemplo, não podemos simplesmente definir um raio de corte r_{cut} ao redor das partículas dentro do qual as energias serão contabilizadas, pois o erro cometido na truncagem diverge. De fato, para sistemas de um único componente cujo potencial de pares é $u(r)$, esse erro pode ser expresso como:

$$E_{\text{trunc.}} = \frac{N\rho}{2} \int_{r_{\text{cut}}}^{\infty} 4\pi r^2 u(r) dr \quad (4.21)$$

onde está implícito que a distribuição radial $g(r) = 1$ para $r > r_{\text{cut}}$. Pode ser demonstrado que a integral acima converge somente para potenciais que decaem mais rápido que r^{-3} no espaço tridimensional. Assim, não é permitido realizar o truncamento quando o potencial for coulombiano, ou mesmo do tipo carga-dipolo e dipolo-dipolo [51].

A soma expressa em 4.20, portanto, precisa ser efetuada de outra maneira. Obviamente que o cálculo bruto de todos os termos não é uma opção viável na prática, mesmo porque foi demonstrado que essa somatória é condicionalmente convergente [52], de sorte que o resultado final dependeria da ordem com que os termos são somados (é claro que a energia de sistemas físicos reais, como cristais, não diverge, devido ao tamanho limitado). O método criado por Paul Ewald (1921) [48] foi o primeiro que conseguiu lidar com esse problema, sendo aplicado na época para obter a energia eletrostática de cristais

iônicos. O método foi posteriormente aperfeiçoado por Leeuw e colaboradores (1980) [52], que incluíram um termo envolvendo os momentos de dipolo. Através de uma série de manipulações matemáticas, os autores conseguem transformar a somatória anterior em duas séries uniformemente convergentes que podem ser truncadas em outras duas somas limitadas:

$$\begin{aligned} \frac{4\pi\epsilon}{e^2} E_C = & \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-\frac{k^2}{4\alpha^2}}}{k^2} \left| \sum_{j=1}^N z_j e^{i\mathbf{k} \cdot \mathbf{r}_j} \right|^2 + \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i=1}^N \sum_{j=1}^N z_i z_j \frac{\text{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \\ & + \frac{2\pi}{3V} \left(\sum_{i=1}^N z_i \mathbf{r}_i \right)^2 - \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^N z_i^2 \end{aligned} \quad (4.22)$$

onde $\mathbf{k} = 2\pi V^{-1/3} \mathbf{u}$, com $\mathbf{u} = (l, m, n) \in \mathbb{Z}^3$, e α é um fator arbitrário utilizado para acelerar a convergência das séries. Uma revisão acessível da dedução desta expressão pode ser encontrada na dissertação defendida por Wagner Gomes no IFUSP (2011) [53].

Os termos acima possuem uma interpretação física relativamente simples. O método de acelerar a convergência da soma original consiste em blindar cada carga do sistema por uma distribuição gaussiana com o mesmo módulo da carga e sinal oposto. Isso é feito somando e subtraindo tais distribuições, de modo que a energia não seja alterada (ver fig. 4.3). A largura dessas gaussianas é inversamente proporcional ao fator α . O termo constante da soma de Ewald, $-\alpha/\sqrt{\pi} \sum z_i^2$, representa a energia de interação das cargas com as suas respectivas blindagens gaussianas. A série que contém a função erro (erfc) representa as interações das cargas com as demais cargas blindadas. E os demais termos representam as interações das cargas com as antblindagens gaussianas, sendo que o termo de dipolo contabiliza essas interações na célula principal ($\mathbf{k} = 0$) e a série de Fourier nas demais células, conferir fig. 4.4.

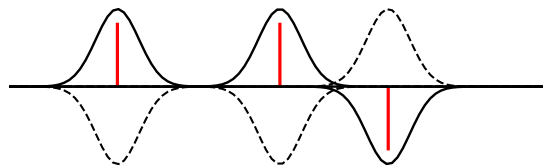


Figura 4.3: Gaussianas utilizadas na soma de Ewald. Somar e subtrair não altera a energia, mas auxilia na convergência. Cargas representadas como distribuições delta em vermelho e as gaussianas em preto.

Há algumas outras abordagens disponíveis na literatura para se calcular a energia de sistemas carregados. A eficiência delas depende bastante do número de partículas

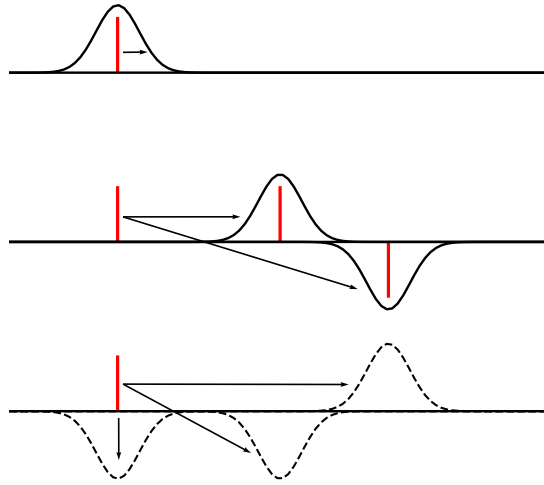


Figura 4.4: Cada carga interage (i) com a sua blindagem gaussiana, (ii) com as outras cargas blindadas e (iii) com as antiblindagens.

consideradas. A soma de Ewald escala com $\mathcal{O}(N^{3/2})$ [51], logo, sua aplicação ainda é limitada a sistemas com N pequeno (digamos, até a ordem de 10^3), mas continua sendo o método mais amplamente empregado. Os métodos PPPM (particle-particle/particle-mesh), desenvolvido por Eastwood e Hockney [54], e o PME (Particle Mesh Ewald) de Pedersen et al. [55] escalam com $\mathcal{O}(N \log(N))$, mas só começam a ser mais eficientes que Ewald pra N entre $10^3 - 10^4$. Já quando $N > 10^5$, o método FMM (Fast Multipole Method) de Greengard e Rokhlin [56], originalmente aplicado para calcular energias gravitacionais, é de longe o mais eficiente, escalando com $\mathcal{O}(N)$.

4 Implementação do algoritmo de MC

O algoritmo que implementamos para simular nosso modelo primitivo foi escrito em linguagem C++ e se encontra resumido no fluxograma abaixo:

- (1) Gera uma configuração aleatória inicial e calcula sua energia.
- (2) Começa um novo passo de MC.
- (3) Sorteia uma partícula da rede (sem repetição).
- (4) Se for um microíon, realize um movimento de translação, calcule a mudança na energia e aplique o teste de Metropolis.
- (5) Se for um macroíon, realize um movimento de cluster, calcule a mudança na energia e aplique o teste de Metropolis.
- (6) Se há partículas que não foram sorteadas, retorne ao passo (3). Se todas as partículas da rede já foram sorteadas, então, caso o número de passos de MC não tenha sido atingido, retorne ao passo (2), caso contrário termine a simulação.

Em nossas simulações utilizamos uma caixa cúbica de lado $L = V^{-1/3} = (N_M/\rho_M)^{-1/3}$, definido pelos parâmetros do modelo. Para trabalharmos com grandezas adimensionais, todas as distâncias foram normalizadas por L . Assim, a caixa terá lado unitário em coordenadas reduzidas e as posições das partículas serão dadas por $\mathbf{R}_i = \mathbf{r}_i/L$. Já a energia e a temperatura reduzidas serão expressas como:

$$E_C^* = \frac{E_C}{E_o} \quad T^* = \frac{k_B T}{E_o} = (\beta E_o)^{-1} \quad \text{onde} \quad E_o = \frac{e^2}{4\pi\epsilon L} \quad (4.23)$$

Além dos parâmetros do modelo, alguns outros precisam ser fornecidos previamente para que a simulação se inicie. Dentre eles estão os parâmetros usados na soma de Ewald: o fator de convergência ($\sigma = \alpha L$), o raio de corte (r_c) da série no espaço real e o comprimento de corte (u_c) da série no espaço recíproco. Mas também precisamos definir os parâmetros usados nos movimentos das partículas, como as distâncias máximas de deslocamento dos microíons (Δ_I) e dos macroíons (Δ_M), e o raio do cluster ($r_{\text{clust.}}$), que engloba as partículas que são movimentadas juntamente com um macroíon. Falaremos um pouco sobre eles nas próximas duas seções. Como gerador de números aleatórios utilizamos a função `ran2` descrita no livro *Numerical Recipes*[57].

4.1 Truncagem da soma de Ewald

Para trabalharmos com unidades adimensionais, reescrevemos a soma de Ewald (4.22) truncada da seguinte maneira:

$$E_C^* = \sum_{|\mathbf{u}| < u_c} \frac{e^{-\frac{\pi^2 u^2}{\sigma^2}}}{2\pi u^2} \left| \sum_{j=1}^N z_j e^{2\pi i \mathbf{u} \cdot \mathbf{R}_j} \right|^2 + \sum_{R_{ij} < r_c} z_i z_j \frac{\text{erfc}(\sigma R_{ij})}{R_{ij}} + \frac{2\pi}{3} \left(\sum_{i=1}^N z_i \mathbf{R}_i \right)^2 - \frac{\sigma}{\sqrt{\pi}} \sum_{i=1}^N z_i^2 \quad (4.24)$$

sendo que a série sobre o espaço recíproco se estende somente a vetores \mathbf{u} com módulo inferior a u_c e a série das funções erro, no espaço real, conta somente as interações das partículas na rede principal com todas as outras partículas dentro de um raio de corte $r_c < 0.5$, sem repetições. É claro que todo truncamento envolve um erro. Kolafa e Perram [58] demonstraram que os erros cometidos ao se truncarem a série real (δE_R) e a série de Fourier (δE_F) valem, em coordenadas reduzidas, aproximadamente:

$$\frac{\delta E_R}{E_o} \approx N Z_I Z_M \sqrt{\frac{r_c}{2}} \left(\frac{1}{\sigma r_c} \right)^2 \exp [-(\sigma r_c)^2] \quad (4.25)$$

$$\frac{\delta E_F}{E_o} \approx N Z_I Z_M u_c \left(\frac{\sigma}{\pi u_c} \right)^2 \exp \left[-\left(\frac{\pi u_c}{\sigma} \right)^2 \right] \left(1 + \frac{1}{\sigma \sqrt{u_c}} \right) \quad (4.26)$$

Se quisermos que os dois erros acima possuam ordens de grandeza semelhantes, podemos usar o truque de igualar os seguintes fatores:

$$\left(\frac{1}{\sigma r_c}\right)^2 \exp[-(\sigma r_c)^2] = \left(\frac{\sigma}{\pi u_c}\right)^2 \exp\left[-\left(\frac{\pi u_c}{\sigma}\right)^2\right] = \left(\frac{1}{s}\right)^2 \exp(-s^2) \quad (4.27)$$

onde $\delta = \exp(-s^2)/s^2$ é uma constante que fixamos para limitar o erro máximo que queremos cometer na truncagem. Impondo tal condição, é fácil ver que os parâmetros da soma de Ewald estarão vinculados por:

$$r_c = \frac{s}{\sigma} \quad u_c = \frac{\sigma s}{\pi} \quad (4.28)$$

de modo que os erros ficarão:

$$\frac{\delta E_R}{E_o} \approx N Z_I Z_M \sqrt{\frac{r_c}{2}} \delta \quad (4.29)$$

$$\frac{\delta E_F}{E_o} \approx N Z_I Z_M u_c \delta \left(1 + \frac{1}{\sigma \sqrt{u_c}}\right) \quad (4.30)$$

Se desprezarmos a contribuição do fator entre parênteses no erro da soma recíproca, o qual deve ser na prática próximo de 1, então a relação entre os dois erros será:

$$\delta E_R \approx \delta E_F \sqrt{\frac{r_c}{2u_c^2}} \quad (4.31)$$

É preciso garantir um certo equilíbrio na escolha dos parâmetros da soma de Ewald para que, por um lado, eles forneçam erros suficientemente pequenos e, por outro, não tornem a simulação muito custosa computacionalmente. Por exemplo, quanto maior for o fator de blindagem σ , menor precisa ser o raio de corte r_c e, portanto, menos pares de interação precisam ser calculados, o que reduz o tempo de computação. Porém, quanto maior for a blindagem mais termos na soma recíproca devem ser contados, o que aumenta o tempo de CPU, a menos que relaxemos mais o erro tolerado δ . Há várias maneiras de se determinar um conjunto satisfatório dos parâmetros σ , r_c e u_c . Aqui, usaremos simplesmente um método de "tentativa e erro". Primeiro, fixamos um valor de σ , digamos algo em torno de 8-14. Depois, testamos diversos valores de δ (para cada par (σ, δ) fica determinado o par (r_c, u_c) e os erros de truncagem, por meio das equações 4.28 e 4.29) e selecionamos os valores que forneçam os menores comprimentos de corte dentro da margem de erro desejada. Podemos repetir o procedimento para outros valores de σ e depois comparar todos os parâmetros selecionados ao final, escolhendo aqueles que melhor otimizem a simulação. Isso quer dizer, tomar os menores valores encontrados para (r_c, u_c) que mantenham um erro tolerável. No nosso caso, consideramos suficientes

erros até a terceira casa decimal.

4.2 Movimentos

Durante a simulação de Monte Carlo uma série de configurações ν_i é gerada a partir de certos movimentos realizados sobre as partículas. A escolha adequada desses movimentos, por sua vez, é fundamental para que o cálculo do valor médio de alguma grandeza de interesse $A(\nu)$ tenha uma boa precisão. De fato, se as configurações produzidas obedecem a distribuição de probabilidade p_ν do ensemble, então o valor esperado pode ser estimado por uma média simples:

$$\langle A \rangle = \sum_{\nu} p_{\nu} A(\nu) \approx \frac{\sum_{i=1}^M A(\nu_i)}{M} \quad (4.32)$$

sendo M o número total de configurações. Na prática, contudo, são usados movimentos que geram configurações sucessivas com algum grau de correlação, de modo que a amostragem final pode não seguir a distribuição de probabilidades desejada ou introduzir redundâncias que irão subestimar as incertezas. Para evitar a utilização de estados correlacionados, é comum definir-se uma função de autocorrelação:

$$f(k) = \frac{\langle A(\nu_i)A(\nu_{i+k}) \rangle - \langle A \rangle^2}{\langle A^2 \rangle - \langle A \rangle^2} \quad (4.33)$$

que avalia quantos movimentos são necessários para que sejam gerados dois estados descorrelacionados, $f = 0$. Esse intervalo de correlação pode ser reduzido dependendo dos tipos de movimento que serão adotados.

Para os nossos propósitos, apenas dois tipos de movimento serão suficientes. O primeiro deles consiste na translação de um único microíon: sorteia-se aleatoriamente um vetor de deslocamento \mathbf{d} , cujo módulo não ultrapasse um certo valor Δ_I , e transladamos a partícula para a nova posição (fig. 4.5). É desejável que o valor de Δ_I seja grande o bastante para que o microíon possa passear livremente por toda a rede e não se limite a deslocamentos locais.

Já o segundo movimento, envolvendo um macroíon, é um pouco mais complicado. Se quiséssemos transladar somente o macroíon teríamos enormes dificuldades, uma vez que ele se encontra, via de regra, rodeado por contraíons muito próximos de sua superfície, de maneira que facilmente ocorreriam sobreposições de partículas e o movimento seria rejeitado no teste de Metropolis. Ou então, os deslocamentos ficariam limitados a distâncias extremamente curtas, gerando grandes correlações. Uma saída possível consiste em definir um raio ($r_{\text{clust.}}$) no entorno do macroíon, formando um cluster com os microíons mais próximos, e deslocando simultaneamente o conjunto das partículas (fig. 4.6). O tamanho do raio do cluster obviamente deve levar em consideração o comprimento de cor-

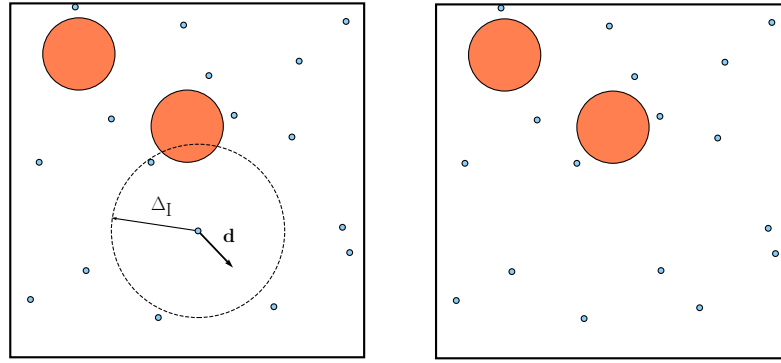


Figura 4.5: Movimento de translação de um microíon.

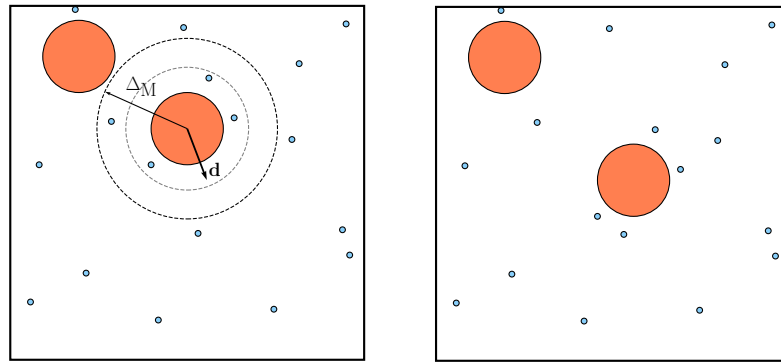


Figura 4.6: Movimento de cluster de um macroíon e seu entorno contido dentro de $r < r_{\text{clust.}}$ (círculo cinza tracejado).

relação dos microíons ao redor do macroíon, o que pode ser estimado pelo comprimento de blindagem de Debye κ^{-1} (ver equação 4.2). Porém, ele não deve ser maior do que um diâmetro do macroíon, sob o risco de englobar outro macroíon dentro do cluster, o que tornaria o movimento menos eficiente pelo mesmo motivo inicial. Já o deslocamento máximo do cluster Δ_M pode ser ajustado durante a simulação através da definição de uma taxa de aceite desejada. Por exemplo, se quisermos que 15% dos movimentos sejam aceitos no teste de Metropolis, adicionamos um condicional ao código que diminui um pouco Δ_M se a taxa de aceite estiver abaixo dos 15%, ou então o aumenta, caso contrário. É preciso simular algumas vezes o sistema para ganhar familiaridade com ele e ter noção de qual taxa de aceite produz melhores resultados para descorrelacionar as configurações. Taxas de aceite alta geralmente ocorrem quando os deslocamentos são pequenos, logo mais correlacionados, enquanto que uma taxa menor pode refletir em movimentos mais amplos e com menor correlação.

Apenas para mencionar, também existem outros movimentos de cluster possíveis [46, 51], em geral mais complexos do que este. Alguns, por exemplo, envolvem o movimento de dois ou mais macroíons ao mesmo tempo. Estes últimos são mais apropriados quando o sistema em questão passa por processos de agregação, implicando numa maior proximidade entre os macroíons.

4.3 Validação do código

Para assegurarmos que o nosso código esteja de fato realizando os cálculos da maneira correta realizamos alguns testes de validação, que consistiram em comparar alguns cálculos com resultados analíticos e com simulações de sistemas carregados da literatura. Seguindo a ordem de elaboração do programa, testamos primeiramente se ele é capaz de reproduzir a distribuição radial de um fluido de esferas-duras não carregado, cuja solução analítica se encontra em (2.103). Em seguida, as funções que calculam a energia através da soma de Ewald foram testadas com um modelo de cristal de NaCl. Nesse caso, a constante de Madelung foi obtida e comparada com o seu valor tabelado. Por fim, as distribuições radiais de alguns sistemas carregados foram calculadas e comparadas com os dados publicados por Per Linse (2000) [45]. A boa reprodutibilidade em todos esses casos nos dará mais confiança no código para as próximas simulações dos sistemas proteicos.

Fluido de Esferas-Duras

Simulamos um fluido de esferas-duras com $N = 200$ partículas de raio $R = 30 \text{ \AA}$ e densidade $\rho = 10^{-6} \text{ \AA}^{-3}$, o que corresponde a uma densidade reduzida (ou, fração volumétrica) $\rho^* = \rho 4\pi R^3/3 = 0.216$, dentro do intervalo no qual a aproximação de Percus-Yevick funciona bem. A função de distribuição radial foi, então, calculada da seguinte maneira:

$$g(r) = \frac{\rho(r)}{\rho} = \frac{\langle n(r) \rangle}{\rho V(r)} = \frac{\sum_{i=1}^M n_i(r)}{M\rho V(r)} \quad (4.34)$$

sendo $\langle n(r) \rangle$ o número médio de partículas distantes de $[r, r + dr]$ de uma dada esfera e $V(r)$ o volume da casca esférica contida entre $[r, r + dr]$. Como essa média deve ser a mesma para todas as N esferas, incluímos a contribuição total delas no cálculo:

$$g(r) = \frac{\sum_{j=1}^N \sum_{i=1}^M n_{ij}(r)}{NM\rho V(r)} \quad (4.35)$$

O número de passos de Monte Carlo foi de $M = 5 \times 10^5$, cada passo sendo composto de N movimentos, e um intervalo de termalização de 5×10^4 passos. O deslocamento máximo das partículas foi fixado como sendo igual à distância média entre elas, expressa pelo raio de Wigner-Seitz $r_o = [(4\pi/3)\rho]^{-1/3}$. Já a distribuição radial teórica foi obtida pela transformada de Fourier do fator de estrutura previsto pela relação de fechamento PY:

$$g(r) = 1 + \frac{1}{2\pi^2 r} \int_0^\infty q \left[\frac{S(q) - 1}{\rho} \right] \sin(rq) dq \quad (4.36)$$

Os gráficos das curvas teóricas e os resultados de simulação estão representados na fig. 4.7, a seguir. Dentro do intervalo de distâncias que pudemos observar (limitado pelo

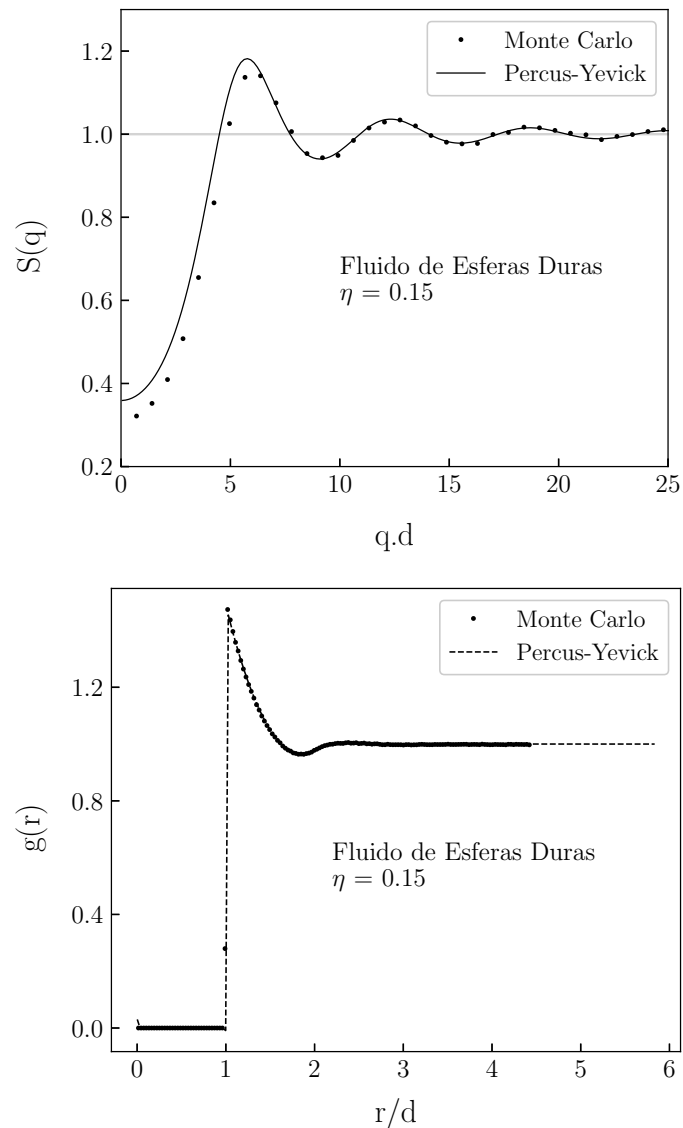


Figura 4.7: Acima, fator de estrutura teórico previsto pela relação de Percus-Yevick e, abaixo, as distribuições radiais teórica e simulada pelo método de Monte Carlo. O módulo do vetor de espalhamento e a distância estão em unidades adimensionais, sendo d o diâmetro das esferas.

tamanho da rede), há uma ótima concordância nas distribuições radiais entre teoria e simulação. Isso indica que o código desenvolvido até aqui deve estar realizando todas as operações prevista, incluindo o cálculo das interações de volume excluído, os movimentos na rede periódica, o cálculo das distribuições radiais etc.

Energia de um cristal de NaCl

Podemos usar um modelo de cristal de NaCl, cuja energia teórica é conhecida, para testar se o nosso cálculo da soma de Ewald fornece valores confiáveis [59]. Este teste, aliás, reproduz a motivação inicial por trás da criação do método. O modelo se baseia em uma célula unitária formada por 8 íons pontuais: 4 de Na^+ e 4 de Cl^- , dispostos

Tabela 4.1: Coordenadas dos íons na célula unitária de um cristal de NaCl.

x	y	z	carga (e)
0	0	0	-1
0.5	0.5	0	-1
0.5	0	0.5	-1
0	0.5	0.5	-1
0.5	0	0	1
0	0.5	0	1
0	0	0.5	1
0.5	0.5	0.5	1

segundo as coordenadas reduzidas da tabela 4.1.

Rearranjando os termos da equação 4.20, podemos escrever a energia eletrostática de uma célula unitária do cristal como:

$$E_C = \frac{1}{2} \frac{e^2}{4\pi\epsilon} \sum_{i=1}^N z_i \left(\sum_{\mathbf{n}}' \sum_{j=1}^N \frac{z_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right) \quad (4.37)$$

Em seguida, normalizando as distâncias pelo comprimento $l = L/2$ de menor separação entre dois íons do cristal, sendo L o lado da célula, obtemos:

$$E_C = \frac{1}{2} \frac{e^2}{4\pi\epsilon l} \sum_{i=1}^N z_i \left(\sum_{\mathbf{n}}' \sum_{j=1}^N \frac{z_j}{|\mathbf{r}_{ij} + \mathbf{n}|/l} \right) = E_o \sum_{i=1}^N z_i M_i \quad (4.38)$$

onde chamamos $E_o = e^2/4\pi\epsilon L$ e o termo entre parênteses é a própria definição da constante de Madelung M_i , associada ao íon do tipo i . Para um cristal de NaCl, as constantes de Madelung possuem o mesmo valor em módulo devido à simetria da rede, ou seja:

$$M_{\text{Na}^+} = -M_{\text{Cl}^-} \quad \text{e} \quad z_{\text{Na}^+} = -z_{\text{Cl}^-} \quad (4.39)$$

Portanto, chamando $M = |M_{\text{Na}^+}| = |M_{\text{Cl}^-}|$ e levando em consideração que uma célula unitária é composta de 4 cátions e 4 ânions, obtemos a seguinte expressão para M :

$$\left| \frac{E_C}{E_o} \right| = |4z_{\text{Na}^+} M_{\text{Na}^+} + 4z_{\text{Cl}^-} M_{\text{Cl}^-}| = |8z_{\text{Na}^+} M_{\text{Na}^+}| = 8M \quad (4.40)$$

$$\therefore M = \frac{1}{8} \left| \frac{E_C}{E_o} \right| \quad (4.41)$$

Assim, comparamos o valor tabelado de $M = 1.74756\dots$ com aquele obtido pelo cálculo da soma de Ewald, utilizando um raio de corte $r_c = 0.5$, para a soma real, e um comprimento de corte $u_c = 26$, para a soma de Fourier. Deixamos o fator de convergência σ variar dentro do intervalo 0 – 30 para observarmos como a precisão do método depende

de uma boa escolha desses três parâmetros. O resultado está apresentado na fig. 4.8. Com este conjunto de parâmetros, nosso código conseguiu reproduzir a constante de Madelung teórica até a sétima casa decimal para valores de σ entre 4-9, e até a segunda casa decimal com σ entre 2.4-15. Logo, a soma de Ewald parece ter sido implementada com sucesso.

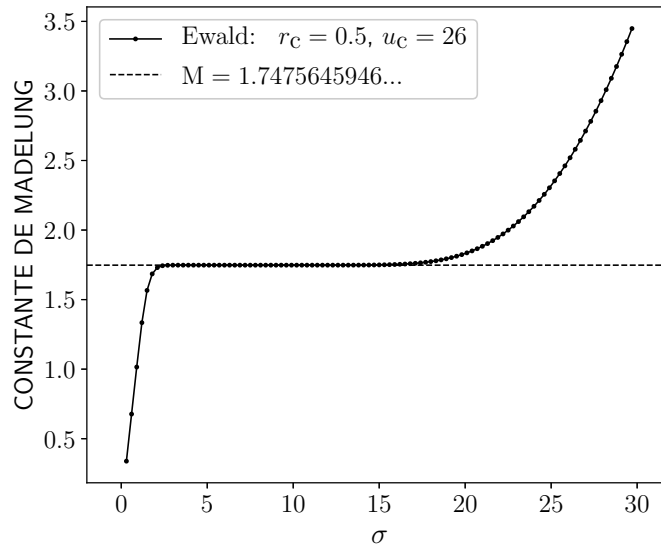


Figura 4.8: Constante de Madelung calculada em função do fator de convergência σ , com raios de corte fixos e iguais a $r_c = 0.5$ e $u_c = 26$.

Reprodução das simulações de Per Linse

No trabalho de Per Linse (2000) [45] o autor simula um modelo primitivo composto de macroíons e contraíons (nenhum sal incluído), explorando o diagrama de fase dos parâmetros reduzidos Z_r , Φ_M e Γ_{II} , descritos anteriormente. Nosso objetivo, aqui, foi reproduzir os gráficos de distribuição radial (referentes à fig. 3 do artigo citado) quando $\Phi_M = 0.01$ e $\Gamma_{II} = 0.3558$ são mantidos constantes e a carga Z_M do macroíon varia entre 10, 20, 40 e 80. Como o autor só especifica os parâmetros reduzidos e as cargas dos íons, temos certa liberdade para escolher os outros parâmetros do modelo (temperatura, raio, densidade, constante dielétrica), desde que os primeiros permaneçam constantes. Os parâmetros que utilizamos estão na tabela 4.2 a seguir. O número de macroíons é livre. Linse utiliza $N_M = 80$, mas afirma que estudos anteriores indicam que resultados praticamente idênticos para a distribuição radial podem ser obtidos com apenas 20 macroíons. Este foi o número que adotamos, pois deixa as simulações mais rápidas. Aqui, estamos definindo as funções de distribuição radial macroíon-macroíon (MM), macroíon-microíon (MI) e microíon-microíon (II) como:

$$g_{MM}(r) = \frac{\rho_{MM}(r)}{\rho_M} \quad g_{MI}(r) = \frac{\rho_{MI}(r)}{\rho_I} \quad g_{II}(r) = \frac{\rho_{II}(r)}{\rho_I} \quad (4.42)$$

Tabela 4.2: Parâmetros do modelo primitivo usados na reprodução dos dados de Linse (2000).

Z_I	Z_M	T(K)	ϵ_r	N_M	R_M (Å)	ρ (Å ⁻³)
-1	10	298.15	78	20	20.2	$2.9 \cdot 10^{-7}$
-1	20	298.15	78	20	20.2	$2.9 \cdot 10^{-7}$
-1	40	298.15	78	20	20.2	$2.9 \cdot 10^{-7}$
-1	80	298.15	78	20	20.2	$2.9 \cdot 10^{-7}$

onde, de maneira genérica, $\rho_{ij}(r)$ é a densidade média de partículas do tipo j a uma distância $[r, r + dr]$ de uma partícula de referência do tipo i .

O número de passos de Monte Carlo utilizado dependeu de qual distribuição radial estavam calculando e da carga do macroión. Para obtermos g_{MM} fixamos $M = 4 \cdot 10^5, 4 \cdot 10^5, 3 \cdot 10^5$ e 10^5 passos para $Z_M = 10, 20, 40$ e 80 , respectivamente. O intervalo de termalização foi sempre de 10^4 passos. Já para o cálculo das funções g_{MI} e g_{II} , as quais convergem bem mais rápido do que a primeira, uma vez que os microíons se movimentam com maior facilidade, utilizamos $M = 5 \cdot 10^4, 3 \cdot 10^4, 10^4$ e $3 \cdot 10^3$ passos para $Z_M = 10, 20, 40$ e 80 , respectivamente. Os resultados das simulações estão dispostos na fig. 4.9, onde usamos a mesma escala da fig. 3 do artigo do Linse (fig. 4.10) para facilitar a comparação.

De maneira geral, os gráficos que simulamos são muito parecidos com aqueles obtidos por Linse. Algumas diferenças, não muito significativas, podem ser comentadas. Em primeiro lugar, a curva de g_{MM} que calculamos se estende até aproximadamente 10 raios de distância. Isso é consequência direta de termos adotado somente $N_M = 20$ macroíons, o que limita o tamanho da caixa (as $g(r)$ são calculadas até $L/2$). Porém, até onde pudemos calcular, as distribuições seguem o mesmo formato e o primeiro pico ocorre na mesma posição, entre 7 e 8 raios. Nas curvas do Linse observamos uma pequena tendência de aumento dos picos quando a carga dos macroíons aumenta, já as nossas curvas com carga 20, 40 e 80 ficaram mais sobrepostas, sendo difícil ver essa tendência com a precisão que obtivemos. Talvez utilizando mais partículas em nossas simulações possamos distinguir melhor essas diferenças. As curvas de g_{MI} e g_{II} seguem também o padrão encontrado pelo autor, revelando o aumento das correlações entre um macroión e os microíons ao redor, como também entre microíons próximos da superfície de um macroión. Poderíamos refinar mais nossas g_{MI} escolhendo um intervalo menor de distâncias (dr), para determinar melhor o valor do máximo próximo de $r/R_M = 1$. Vale mencionar que Linse suaviza as curvas dele com funções analíticas, de modo que pequenas diferenças com os nossos resultados podem ser devidas a este pós-tratamento dos dados.

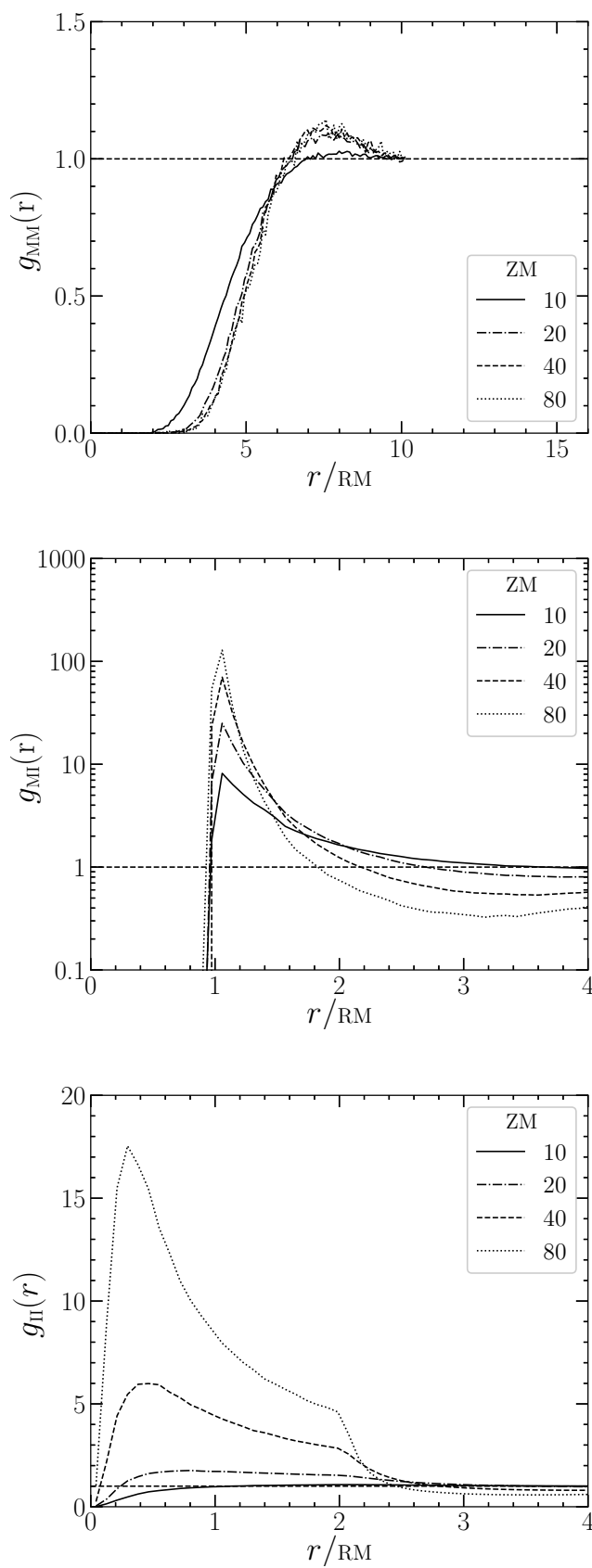


Figura 4.9: Distribuições radiais g_{MM} , g_{MI} e g_{II} de fluidos carregados com diferentes cargas do macroíon e contraíons monovalentes. As distâncias estão normalizadas pelo raio R_M do macroíon.

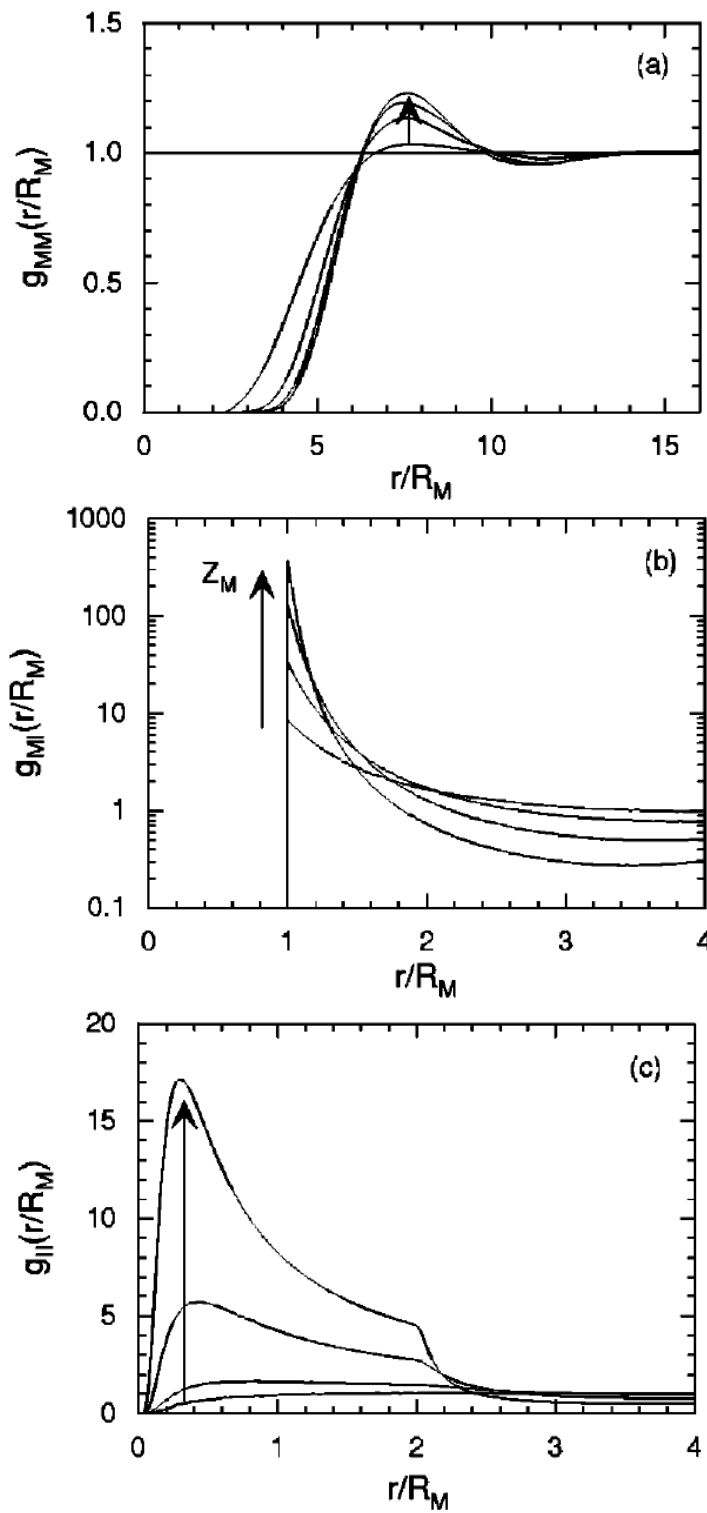


Figura 4.10: Distribuições radiais g_{MM} , g_{MI} e g_{II} de fluidos carregados com diferentes cargas do macroíon e contraíons monovalentes. Imagens retiradas do artigo de Linse [45], figura 3, para comparação com as nossas simulações.

Análise das Proteínas Puras: Resultados e Discussão

1 Liozima

1.1 Efeitos do sal e da temperatura

Nesta seção iremos analisar os efeitos provocados pela variação da temperatura e da concentração de sal monovalente (NaCl) sobre o espalhamento de raios-X a baixos ângulos das amostras de liozima (LYS). Para facilitar a exposição dividimos os dados experimentais em dois conjuntos, conforme mostra o diagrama abaixo: (i) o primeiro com concentração de proteína fixa em 1.36 mM e temperatura e sal variáveis; e (ii) o segundo com concentração de NaCl fixa em 150 mM, mas variando a temperatura e a concentração de liozima. Nos casos em que as concentrações de sal são iguais a 0 e 50 mM a temperatura variou em $T = 10, 15, 20, 25, 30, 35, 40$ °C, em $\text{pH} = 8$, enquanto as curvas a 150 mM de sal foram coletadas a 10, 23 e 37 °C e $\text{pH} = 7.4$.

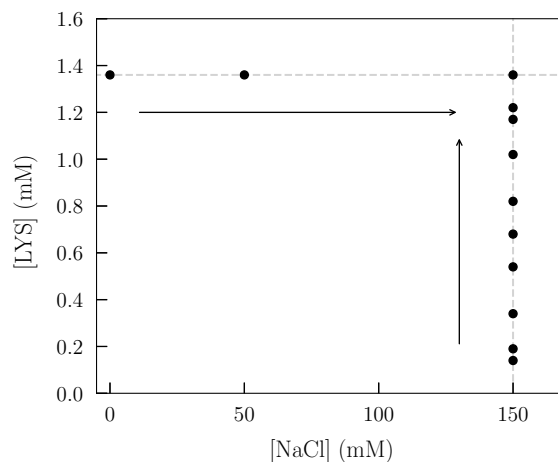


Figura 5.1: Diagrama densidade de liozima vs concentração de NaCl. Cada ponto representa uma amostra.

Uma primeira análise qualitativa consiste em graficar as curvas de espalhamento e

observar o seu comportamento conforme variamos a temperatura. Como as curvas, por razões experimentais, não estavam todas na mesma escala, foi necessário inicialmente acertar a escala para podermos compará-las. Isso foi feito ajustando as curvas com o modelo 46 do GENFIT, o qual utiliza o arquivo PDB da proteína, para ajustar o fator de forma, e um potencial de esferas duras somado a um potencial efetivo tipo-DLVO para ajustar o fator de estrutura através da relação de fechamento RPA (Random Phase Approximation). Este último potencial é composto de um termo repulsivo de Coulomb blindado somado a um termo atrativo, sendo que ambos possuem a forma de um potencial de Yukawa, ou seja, proporcionais a $\exp(-\alpha r)/r$:

$$\beta U(r) = \begin{cases} \infty & , \quad r < \sigma \\ \frac{\lambda_B Z^2}{(1+\kappa\sigma/2)^2} \frac{e^{-\kappa(r-\sigma)}}{r} - J\sigma \frac{e^{-(r-\sigma)/D}}{r} & , \quad r > \sigma \end{cases} \quad (5.1)$$

de modo que o chamaremos até o final da tese de potencial HSDY (Hard-Sphere Double-Yukawa). Temos aqui 5 parâmetros principais a serem ajustados: a carga numérica Z , o diâmetro σ das esferas, a profundidade J e a largura D do termo atrativo, além da força iônica I incorporada na constante de Debye. Além disso, os ajustes das curvas com as mesmas concentrações de proteína e de sal foram realizados de maneira simultânea, isto é, minimizando os parâmetros de todas as curvas com temperaturas distintas conjuntamente. Assim, assumimos que a carga efetiva, o diâmetro da proteína e a força iônica são parâmetros comuns às curvas, independentemente da temperatura, de modo que apenas D e J variaram caso a caso com a temperatura durante o procedimento de ajuste. Note-se que, ainda que mudanças de temperatura possam afetar o raio hidrodinâmico médio das proteínas ou o pKa dos grupos passíveis de protonação/desprotonação e, portanto, a carga, estamos supondo que esses efeitos são apenas secundários.

Com as curvas assim ajustadas, foi possível graficá-las de maneira que o fator de forma delas coincidisse, já que este deve ser o mesmo para todas elas (supondo que a estrutura da lisozima não se altere muito nesse intervalo de temperaturas). Por conveniência, também definimos $P(q_{\min}) = 1$. Com essa escala arbitrária, mas padronizada, a comparação se torna imediata, ver figs. 5.2 e 5.3.

Em primeiro lugar, o bom alinhamento das curvas na região de $q > 0.1 \text{ \AA}^{-1}$ indica uma boa coerência da estrutura real da lisozima em solução com a estrutura cristalográfica, e a preservação dessa forma estrutural com as mudanças de temperatura na faixa selecionada (10-40 °C, bem abaixo da temperatura de *melting*: $T_m = 72 \text{ °C}$). Uma vez que o fator de forma das proteínas está aparentemente sendo preservado, atribui-se o efeito da temperatura nas curvas a mudanças no fator de estrutura e, portanto, nas interações proteína-proteína. Segundo, o efeito do aumento da temperatura, em geral, consiste em diminuir a intensidade de espalhamento, sobretudo na região de $q < 0.1 \text{ \AA}^{-1}$. Isso significa que o potencial de pares tende a ser menos repulsivo, ou mais atrativo, em

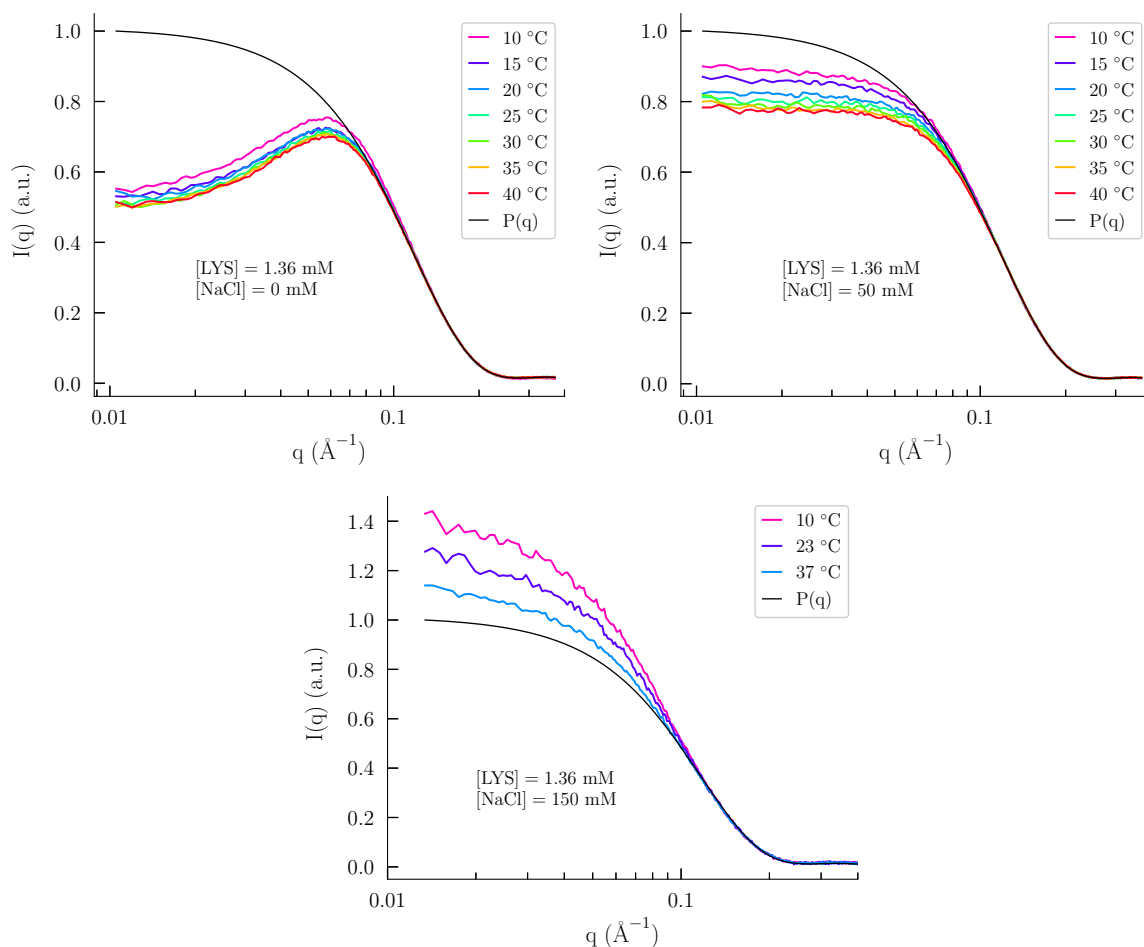


Figura 5.2: Efeito da temperatura no espalhamento da lisozima a 1.36 mM e em três concentrações distintas de NaCl: 0 mM, 50 mM e 150 mM. Os gráficos estão em escala linear-log. As curvas pretas representam o fator de forma da estrutura cristalográfica da lisozima 6LYZ (PDB).

temperaturas mais baixas. Este efeito já havia sido observado por Malfois et al. [60], para soluções de lisozima a 126 mg/ml (8.75 mM) em tampão fosfato 0,6 M e pH 7,5. Tardieu et al. [61] também observaram o mesmo efeito próximo ao ponto isoelétrico da lisozima, pH = 10,5, com $[LYS] = 74 \text{ mg/ml}$ (5.14 mM), o que significa que a temperatura deve estar afetando, pelo menos, a componente atrativa do potencial. Além disso, soluções mais diluídas e/ou com baixa força iônica são menos sensíveis às mudanças de temperatura, enquanto que o oposto é verdadeiro para sistemas mais densos e/ou mais blindados [62]. Finalmente, a adição de sal altera a interação de muito repulsiva para principalmente atrativa, indo ao encontro de resultados anteriores de Ducruix et al. [63], em pH = 4,5 e 18 °C.

A partir dos ajustes mencionados acima foi possível obter algumas funções características das nossas amostras, como o fator de estrutura, a energia potencial de pares e a função de distribuição radial, além dos parâmetros físico-químicos dos quais essas funções dependem. Cabe, agora, discutir mais detalhadamente esses parâmetros e funções

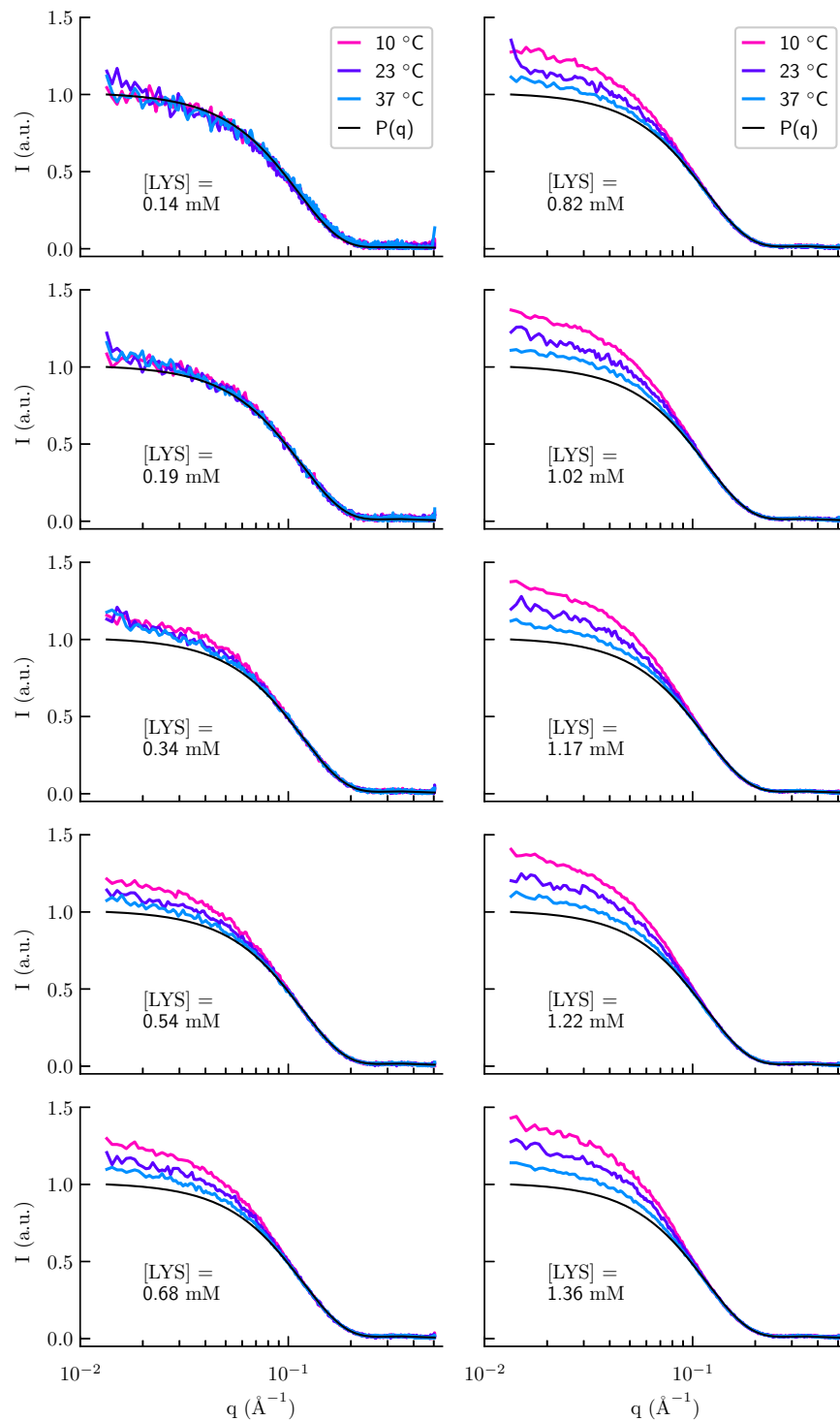


Figura 5.3: Efeito da temperatura no espalhamento da lisozima em diversas densidades variadas entre 0.14-1.36 mM (ou, cerca de 2-20 g/L) e concentração de NaCl fixa em 150 mM. As curvas pretas representam o fator de forma da estrutura cristalográfica de lisozima 6LYZ (PDB).

encontrados. Para evitar repetições, vamos analisar aqui somente as curvas com $[LYS] = 1.36$ mM, deixando para comentar a influência da temperatura em outras concentrações (5.3) numa seção posterior.

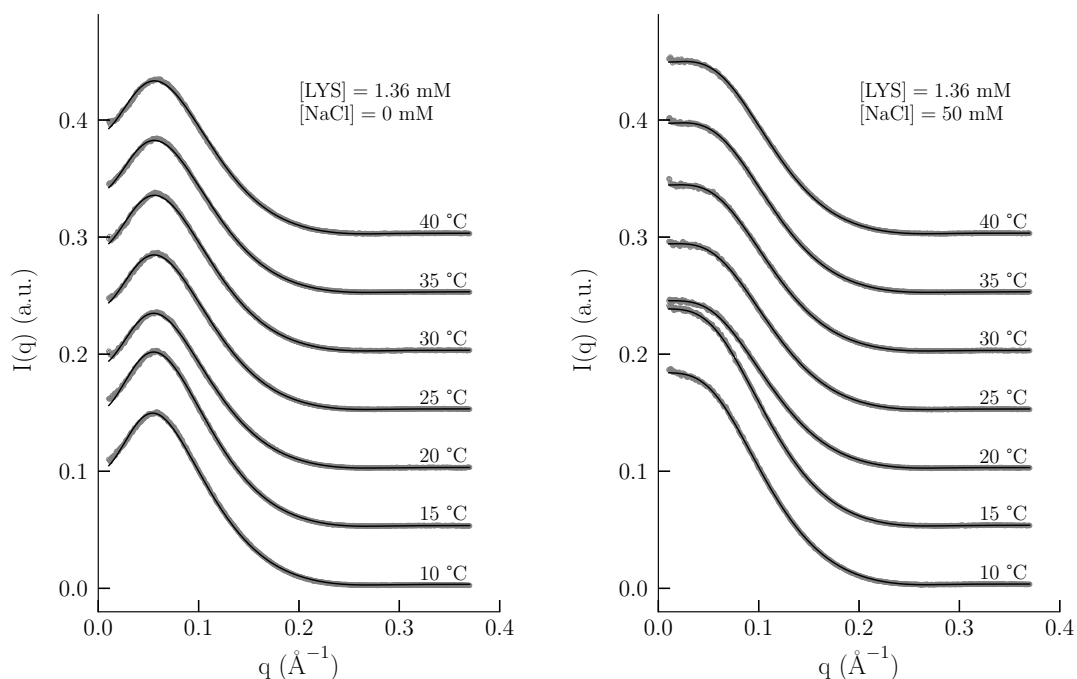


Figura 5.4: Ajuste global das soluções de lisozima com $[\text{NaCl}] = 0$ mM (direita) e 50 mM (esquerda). Os pontos cinzas são os dados experimentais e a linha preta contínua a função de ajuste, que usa um potencial tipo-DLVO, com carga e diâmetro fixos para todas as curvas, sob a aproximação RPA. As curvas estão deslocadas na vertical para melhor visualização.

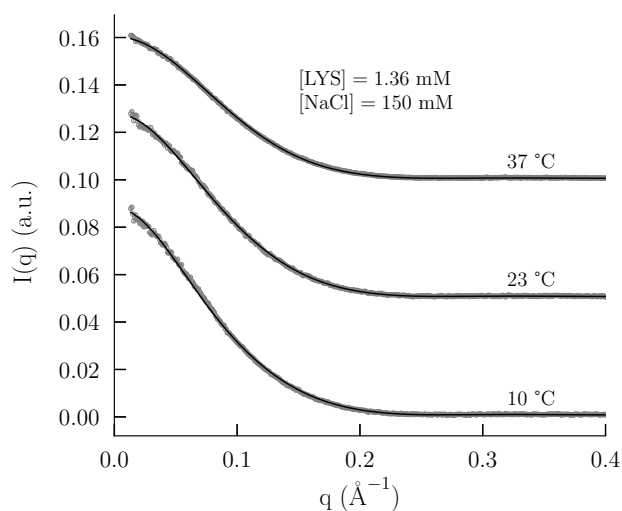


Figura 5.5: Ajuste global das soluções de lisozima a 150 mM de NaCl. A descrição do ajuste é a mesma da figura anterior. Curvas deslocadas na vertical para melhor visualização.

As figuras 5.4 e 5.5 mostram os dados experimentais, referentes às soluções com a maior concentração de lisozima (1.36 mM), sobrepostos às funções de ajuste (eq. 3.90). Para todas as concentrações de NaCl, o potencial de interação proteína-proteína representado por um potencial HSDY resulta em funções $S(q)$, dentro da relação de

fechamento de RPA, que levam a funções $I(q)$ modeladas que se ajustam muito bem aos dados de espalhamento. Os parâmetros ajustados são mostrados nas Tabelas 5.1, 5.2 e 5.3. No entanto, a interpretação física desses parâmetros, em especial para o caso sem sal, nos parece bastante problemática e será discutida mais adiante.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)
10	8.0	3.0	4.8	28
15	8.0	2.4	4.8	28
20	8.9	2.0	4.8	28
25	8.7	1.6	4.8	28
30	8.6	1.4	4.8	28
35	8.0	1.4	4.8	28
40	8.5	1.1	4.8	28

Tabela 5.1: Parâmetros do ajuste da lisozima a 1.36 mM, $[\text{NaCl}] = 0$ mM e $\text{pH} = 8$.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)
10	8.4	4.2	7.4	28.4
15	6.7	4.8	7.4	28.4
20	6.0	4.9	7.4	28.4
25	5.5	4.4	7.4	28.4
30	6.0	4.2	7.4	28.4
35	6.0	3.9	7.4	28.4
40	5.8	3.7	7.4	28.4

Tabela 5.2: Parâmetros do ajuste da lisozima a 1.36 mM, $[\text{NaCl}] = 50$ mM e $\text{pH} = 8$.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)
10	11	4.1	6.4	28
23	9	4.1	6.4	28
37	8	3.5	6.4	28

Tabela 5.3: Parâmetros do ajuste da lisozima a 1.36 mM, $[\text{NaCl}] = 150$ mM e $\text{pH} = 7.4$.

Como podemos ver, o diâmetro ajustado das proteínas (σ) praticamente não variou com a concentração de sal, permanecendo em torno de 28 Å, o que está de acordo com valores relatados na literatura [64, 65, 66]. Em relação à carga, porém, não era esperado que ela sofresse alterações tão grandes em função da concentração salina. De acordo com o software de predição de carga PROPKA [67, 68], deveríamos obter uma carga próxima de 7.4 em $\text{pH} 8^1$, exatamente o valor ajustado para as soluções com 50 mM de

¹Usamos como entrada do programa a mesma estrutura cristalográfica empregada nos ajustes, o arquivo PDB 6LYZ.

NaCl. Para 150 mM de sal a carga efetiva encontrada foi um pouco menor, 6.4, uma unidade a menos que a anterior. Já no caso sem sal, a carga efetiva encontrada, 4.8, ficou muito subestimada em relação ao valor de referência e, por isso, não deve ser considerada fisicamente adequada.

Em relação aos parâmetros do potencial atrativo temos que, para o caso sem sal, os valores de J mudaram apenas ligeiramente com a temperatura, enquanto que os valores de D oscilaram entre 1-3 Å, diminuindo com o aumento da temperatura. Já para as outras concentrações de sal, a largura D do potencial atrativo variou entre 3.5-4.9 Å. Esses valores são um pouco maiores do que aquele obtido por Malfois et al. [60] (3 Å) no mesmo pH, embora eles tenham utilizado uma força iônica quatro vezes maior do que a nossa, 600 mM. Da mesma forma, outros autores também propuseram uma largura atrativa para a lisozima em torno de 3 Å [61, 24]. A profundidade J , expressa em unidades de k_bT , assume valores maiores com 150 mM de sal (8-11) do que com 50 mM (5.5-8.4). Em ambos os casos ela atinge seu valor máximo na temperatura mais baixa, 10 °C. Nossa análise sugere, assim, que o termo atrativo deve mudar tanto com a temperatura quanto com a força iônica. Em contraste, Tardieu et al. [61] tentaram ajustar uma série de soluções de lisozima a 100 mg/ml (6.94 mM) e pH 4.5 dentro de uma faixa de concentrações de NaCl entre 0-350 mM, mantendo $J = 2.65$ e $D = 3$ Å constantes. Essa abordagem, que praticamente manteve o potencial atrativo invariante com a concentração de NaCl, exigiu que a carga protéica ajustada diminuísse progressivamente de $Z = 6$ para 2.1 à medida que a quantidade de sal aumentava, o que, segundo os próprios autores, não está de acordo com um processo realista de regulação de carga.

Os fatores de estrutura, os potenciais de pares e as distribuições radiais derivados dos ajustes podem ser conferidos, respectivamente, nas figuras 5.7, 5.8 e 5.6. Conforme mencionado anteriormente, a intensidade de espalhamento da solução sem NaCl tem uma fraca dependência com a temperatura, e isso se reflete nos fatores de estrutura da mesma (ver gráficos do topo da figura 5.7), enquanto que nas outras concentrações de sal o efeito é bem maior. Nos menores ângulos de espalhamento, em geral, o aumento da temperatura reduz o fator de estrutura. Apenas um efeito sutil é observado sobre as posições e as amplitudes dos picos e vales secundários. Em 50 mM de NaCl, a variação das curvas com a temperatura é bem mais perceptível do que no caso sem sal: o começo das curvas está mais próximo da unidade e varia cerca de 0.1 entre 10 e 40 °C; a altura do primeiro pico ora aumenta ora diminui, com uma tendência a se deslocar para a direita. Já a 150 mM de sal o formato inicial das $S(q)$ sofre uma inversão de concavidade e elas passam a ser maiores do que um, variando mais do que 0.2 quando a temperatura muda entre 10 e 37 °C. Quase não se observa o deslocamento dos picos, apenas uma redução de suas amplitudes conforme a temperatura cresce, resultando numa aparência geral mais "comportada" do que as curvas a 50 mM de sal.

Em relação aos potenciais de interação efetivos, observa-se um claro predomínio da

repulsão eletrostática na ausência de NaCl, com uma considerável barreira de potencial, cuja altura é maior do que $1 k_b T$ (topo da figura 5.8). A porção efetivamente atrativa do potencial é apenas de curtíssimo alcance, menor que um décimo do diâmetro da lisozima, enquanto que a extensão da porção repulsiva se prolonga para mais de quatro diâmetros. Aqui, como nas outras concentrações de sal, o efeito do aumento da temperatura se resume a diminuir a barreira de potencial - reduzindo, portanto, a estabilidade coloidal - e a deslocá-lo ligeiramente para a direita. A altura do pico também diminui conforme a concentração de sal aumenta, devido à maior blindagem das cargas das lisozimas pela atmosfera iônica. A 50 mM de NaCl essa altura não chega a $1 k_b T$ e a 150 mM nem mesmo a $0.1 k_b T$, quase desaparecendo nesse último caso em temperaturas abaixo de 23 °C, o que, do ponto de vista físico, é problemático, pois indica um sistema teoricamente instável e passível de agregação (sendo que, na prática, não se observou agregação nenhuma). A largura do pico também é bastante reduzida a 50 mM e mais ainda a 150 mM. Em suma, a parte atrativa se torna preponderante conforme a temperatura cresce e mais ainda com a adição de sal, chegando a dominar as interações repulsivas nas soluções de maior concentração salina.

As funções de distribuição radial foram derivadas dos fatores de estrutura por meio da seguinte transformada de Fourier:

$$g(r) = 1 + h(r) \approx 1 + \frac{1}{2\pi^2 \rho r} \int_0^\infty q [S(q) - 1] \sin(rq) dq \quad (5.2)$$

válida para soluções homogêneas e isotrópicas (ver eq. 3.88). Os resultados são mostrados na figura 5.6. Uma incoerência que logo é observada nas soluções sem NaCl é a violação da condição $g(r < d) = 0$, ou seja, para distâncias menores do que um diâmetro a distribuição radial deveria ser nula devido ao volume excluído da esfera, o que não ocorreu nas funções derivadas dos ajustes. Outra incoerência grave é que as curvas chegam a assumir valores negativos, o que é absurdo pois a função $g(r)$ deve ser sempre maior ou igual a zero. Essas violações são claros sinais de uma inadequação teórica do modelo e dos pressupostos de sua aplicabilidade, provavelmente ocasionada pelo forte potencial repulsivo (já que a blindagem dos íons é mais fraca na ausência de sal) combinado à densidade relativamente alta de proteínas (1.36 mM). Nessas condições, é possível que a relação de fechamento RPA não seja capaz de modelar adequadamente o fator de estrutura, uma vez que, nessa abordagem, o potencial de pares é tratado como uma pequena perturbação. As $g(r)$ das soluções a 50 e 150 mM de sal, pelo contrário, não apresentam os mesmos problemas de convergência na região de $r < d$, onde tendem a zero como deveriam, nem assumem valores negativos.

Ainda que a interpretação física das distribuições radiais na ausência de sal fique assim comprometida, podemos captar algumas tendências gerais nas curvas obtidas.

Como o potencial de força média é dado por $u_{fm}(r) = -k_b T \ln g(r)$, então à medida

que a concentração de sal aumenta observa-se uma mudança de um potencial médio predominantemente repulsivo (0 mM NaCl) para totalmente atrativo (150 mM NaCl) e de menor alcance, com cerca de meio diâmetro de largura. Do mesmo modo, o efeito da diminuição da temperatura é o de aumentar a atração média entre as lisozimas (embora este efeito seja mais sutil do que o do incremento de sal). Todas as curvas possuem também um pico estreito próximo à superfície das partículas, o qual assume valores maiores do que um (variando em torno de 3-10 unidades adimensionais), o que significa uma densidade local acima da média nessa região do pico, ou seja, indicando que podem se formar momentaneamente alguns pares de proteínas, mas sem muita estabilidade.

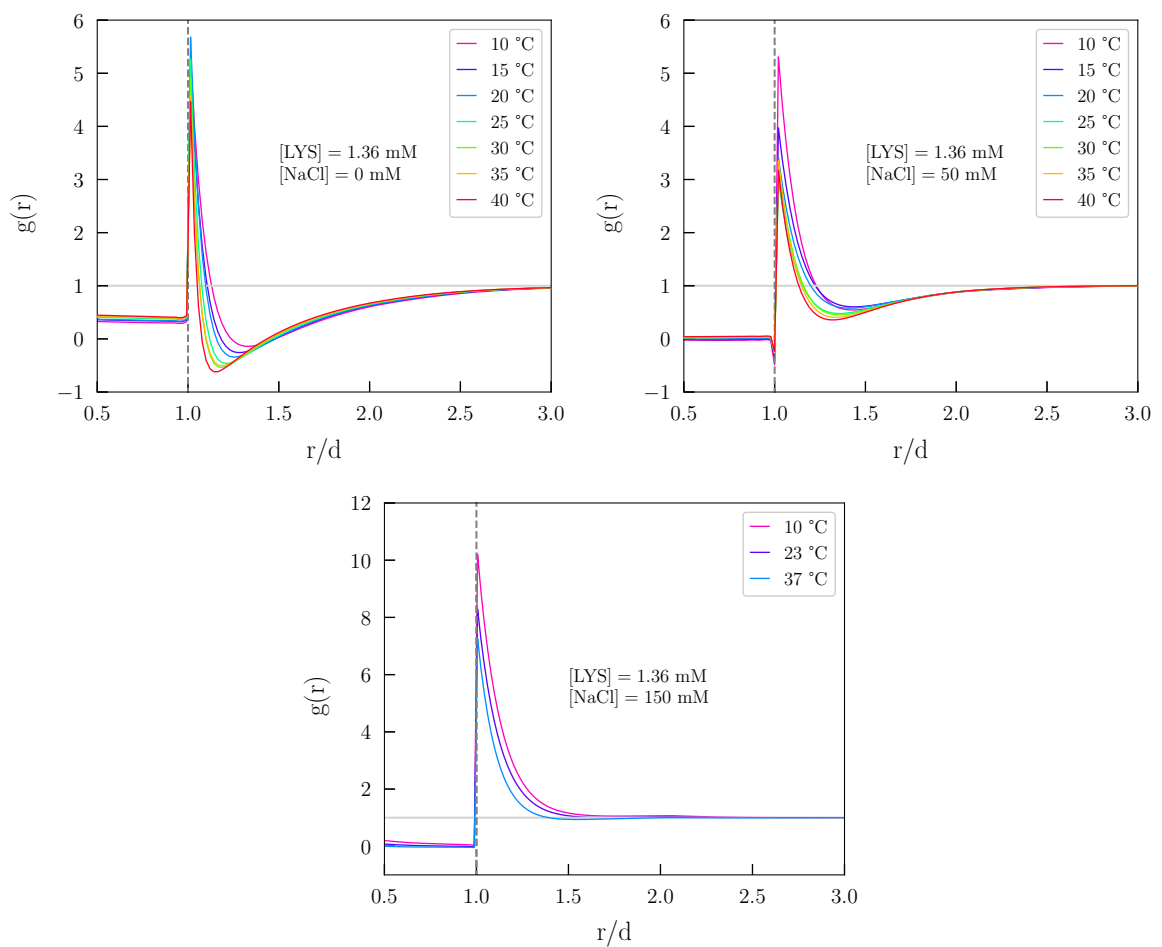


Figura 5.6: Funções de distribuição radial derivadas dos ajustes das curvas de espalhamento da lisozima, variando a temperatura e a concentração de sal.

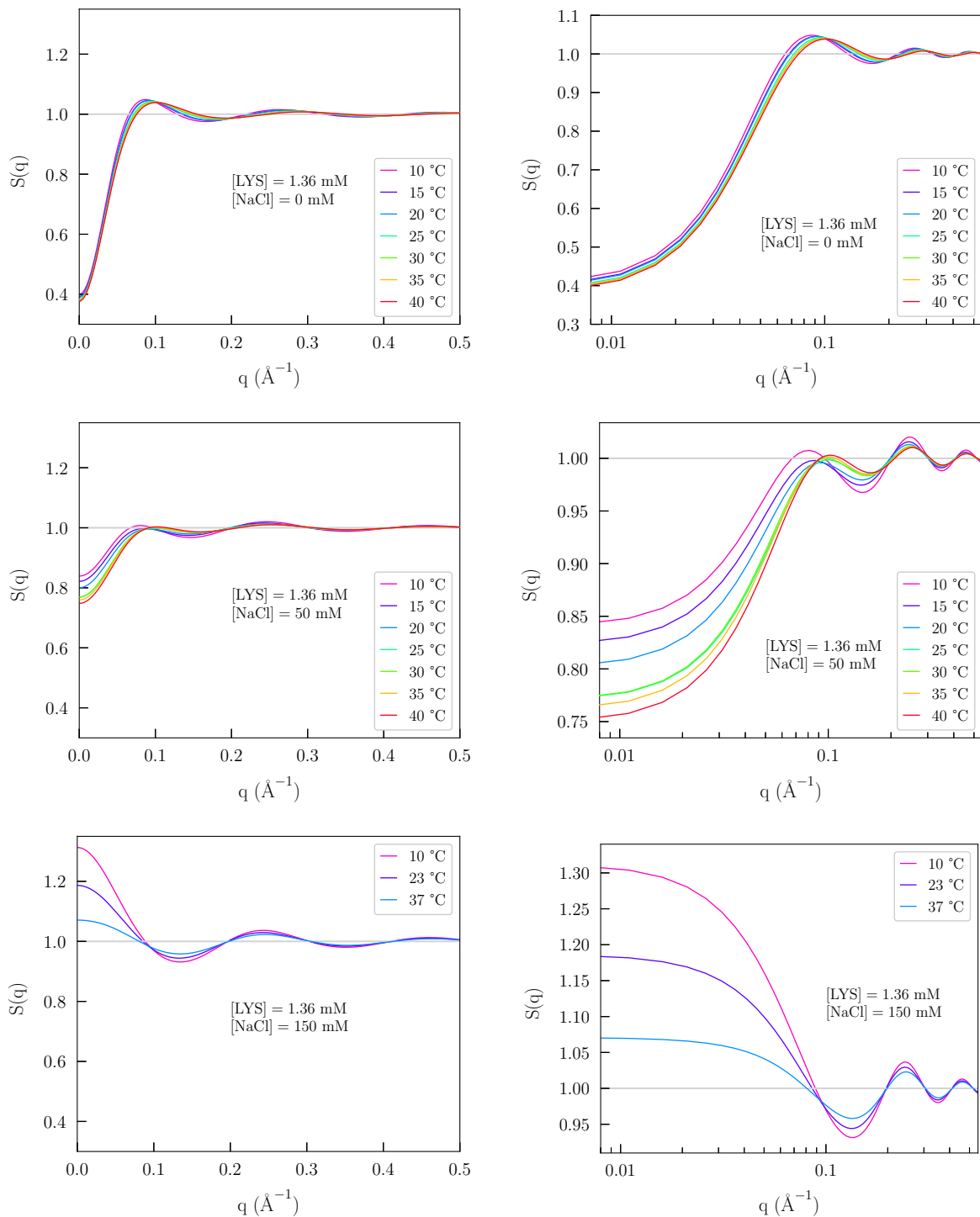


Figura 5.7: Esquerda: fatores de estrutura, $S(q)$, obtidos do ajuste das curvas de lisozima. Direita: os mesmos gráficos, mas em escala logarítmica no eixo horizontal para facilitar a visualização.

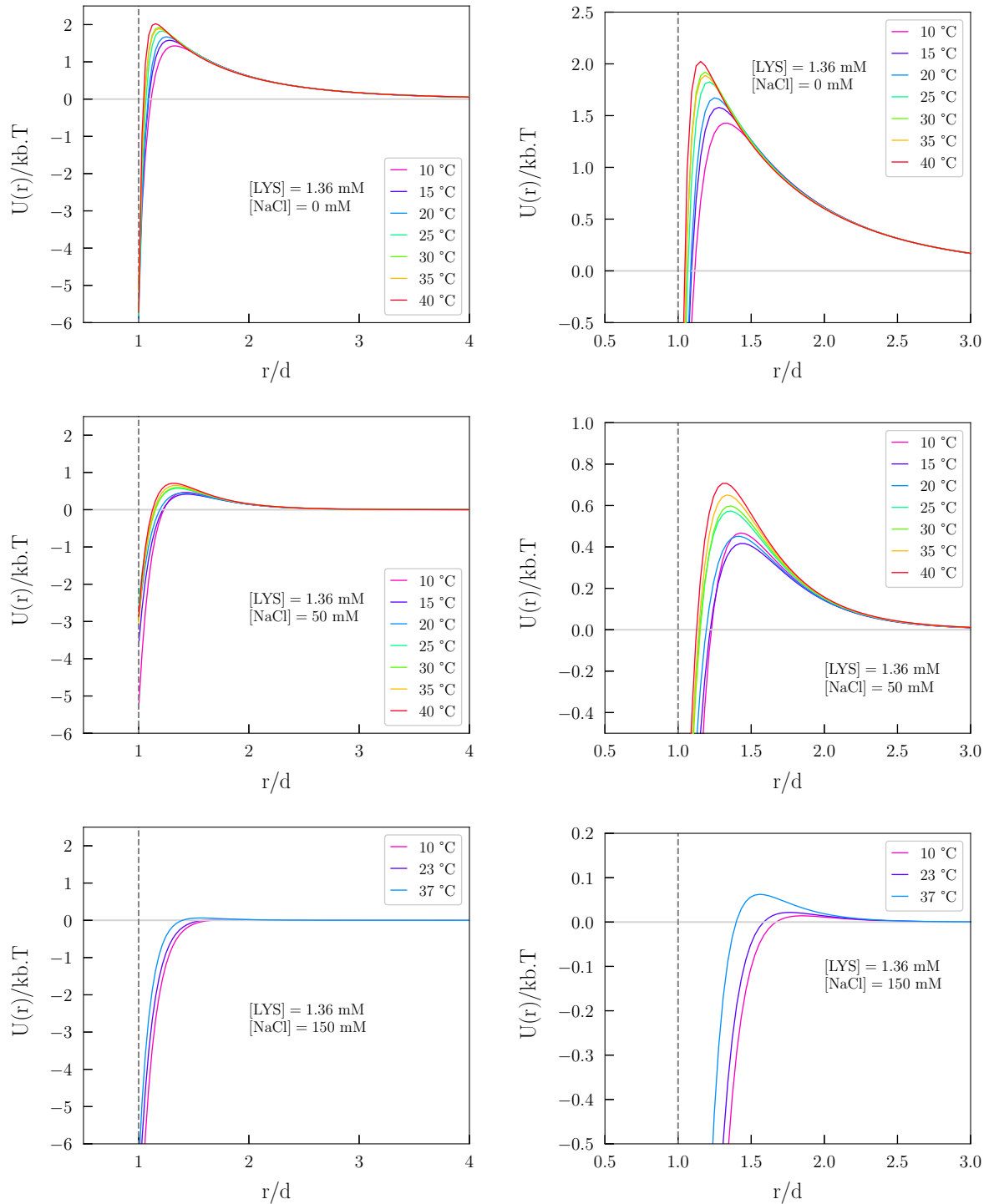


Figura 5.8: Esquerda: energia potencial efetiva, $U(r)$, em unidades de $k_b T$. Direita: as mesmas curvas, porém ampliadas na região do pico de potencial.

1.2 Efeitos de concentração

O efeito da concentração foi analisado em três condições de temperatura (10, 23 e 37 °C), mantendo fixa a concentração de NaCl (150 mM), conforme o diagrama abaixo. Em cada condição de temperatura variamos a concentração de lisozima entre dez valores distintos: 0.14, 0.19, 0.34, 0.54, 0.68, 0.82, 1.02, 1.17, 1.22 e 1.36 mM (o que equivale em peso a cerca de 2-20 g/L de proteína). Lembrando que todas essas soluções foram preparadas em tampão PBS e pH = 7.4.

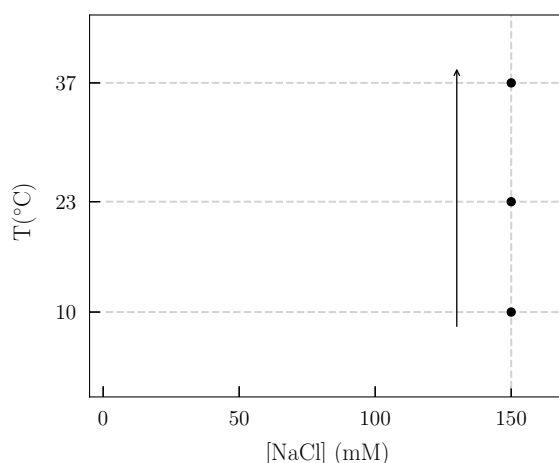


Figura 5.9: Diagrama temperatura vs concentração de NaCl. Para cada ponto variou-se a concentração de lisozima entre 10 valores.

O efeito da concentração já pôde ser parcialmente observado na figura 5.3, mas podemos graficar novamente as mesmas curvas agrupando-as segundo a temperatura, ver figura 5.10 a seguir. Em geral, observa-se um aumento na intensidade de espalhamento na região de $q < 0.1 \text{ \AA}^{-1}$ à medida em que a concentração aumenta. Tal efeito é tanto mais significativo quanto menor for a temperatura. A 10 °C, por exemplo, vemos uma clara separação entre as curvas, a qual se reduz sensivelmente a 23 °C até quase desaparecer a 37 °C. As interações interproteínas - ou, a correlação espacial entre as lisozimas -, portanto, são mais relevantes na menor temperatura observada, enquanto que na maior a solução está aparentemente mais próxima de um sistema aleatoriamente distribuído.

O ajuste das curvas experimentais foi feito utilizando-se o mesmo procedimento descrito na seção anterior, com a diferença de que, para uma mesma condição de temperatura e sal, as dez curvas com concentrações distintas foram ajustadas simultaneamente. Além disso, deixamos as concentrações variarem dentro de um pequeno intervalo de 0.5 g/L para levarmos em conta possíveis erros sistemáticos na preparação das amostras. O resultado se mostrou bastante satisfatório (figura 5.11) dentro do intervalo de concentrações e temperaturas testadas. Já havíamos notado que o potencial HSDY, junto com a relação de fechamento RPA, ajusta bem as curvas de lisozima com alta blindagem eletrostática, o que é o caso aqui (150 mM de NaCl).

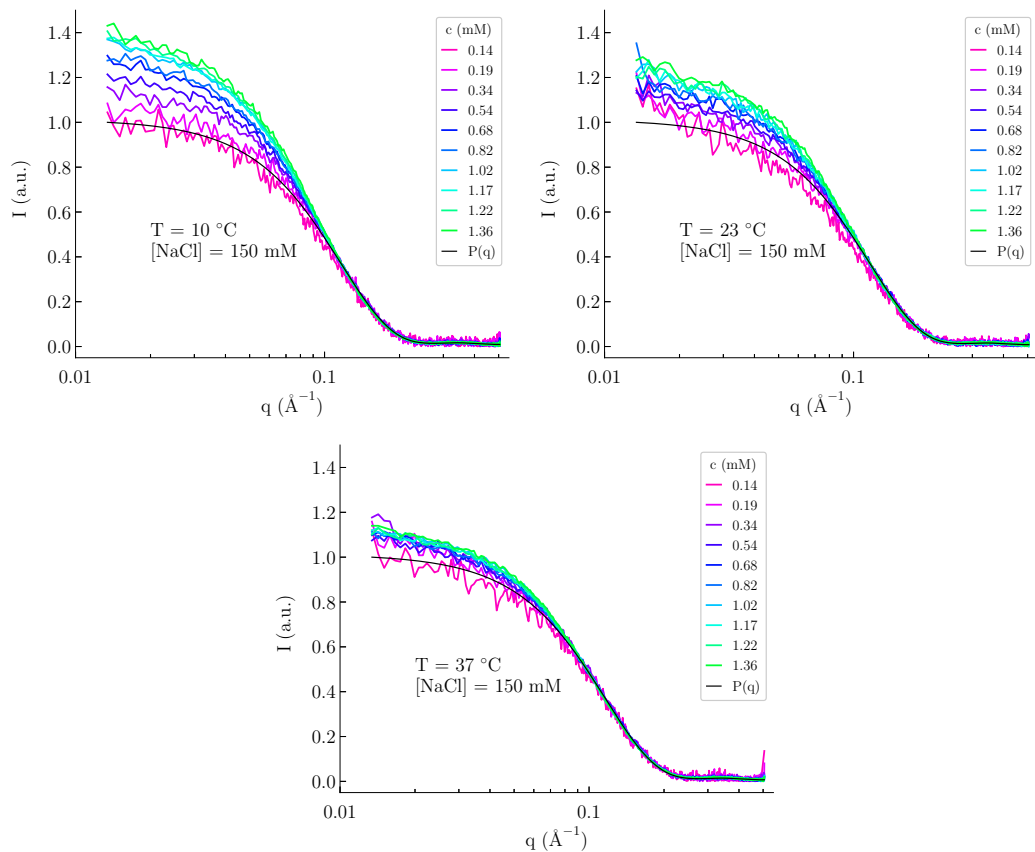


Figura 5.10: Efeito da concentração no espalhamento da lisozima em $T = 10, 23$ e 37 °C e 150 mM de NaCl. A escala está em linear-log. A curva preta representa o fator de forma da estrutura cristalográfica 6LYZ (PDB).

O pico principal das funções $S(q)$, figura 5.11, é sempre maior do que a unidade, indicando um potencial efetivo predominantemente atrativo. De maneira geral, os picos e vales dessas funções se tornam mais intensos conforme aumenta-se a concentração, refletindo uma maior correlação proteína-proteína, e diminuem à medida que cresce a temperatura e o sistema tende a se tornar mais desordenado. Nenhum deslocamento horizontal dos picos e vales foi verificado, como quando se variou a temperatura em baixas concentrações salinas.

Os parâmetros ajustados se encontram na tabela 5.4. Para uma mesma condição de temperatura, foi possível ajustar simultaneamente todas as curvas com diferentes concentrações utilizando um mesmo valor de carga Z , diâmetro σ e largura do poço atrativo D . A única variação entre elas, no caso, se deu com o parâmetro J , o qual apresentou um leve crescimento (ou seja, se tornou mais atrativo) com a redução da temperatura. Em comparação com os ajustes da seção anterior, feitos apenas com a concentração mais alta de lisozima (1.36 mM), os valores dos parâmetros obtidos neste ajuste não coincidem perfeitamente com aqueles ($J = 8 - 11$ ($k_b T$), $D = 3.5 - 4.1$ Å, $Z = 6.4$ (e) e $\sigma = 28$ Å, tabela 5.3), embora também não sejam muito discrepantes. As diferenças possivelmente se devem aos múltiplos mínimos locais presentes no processo

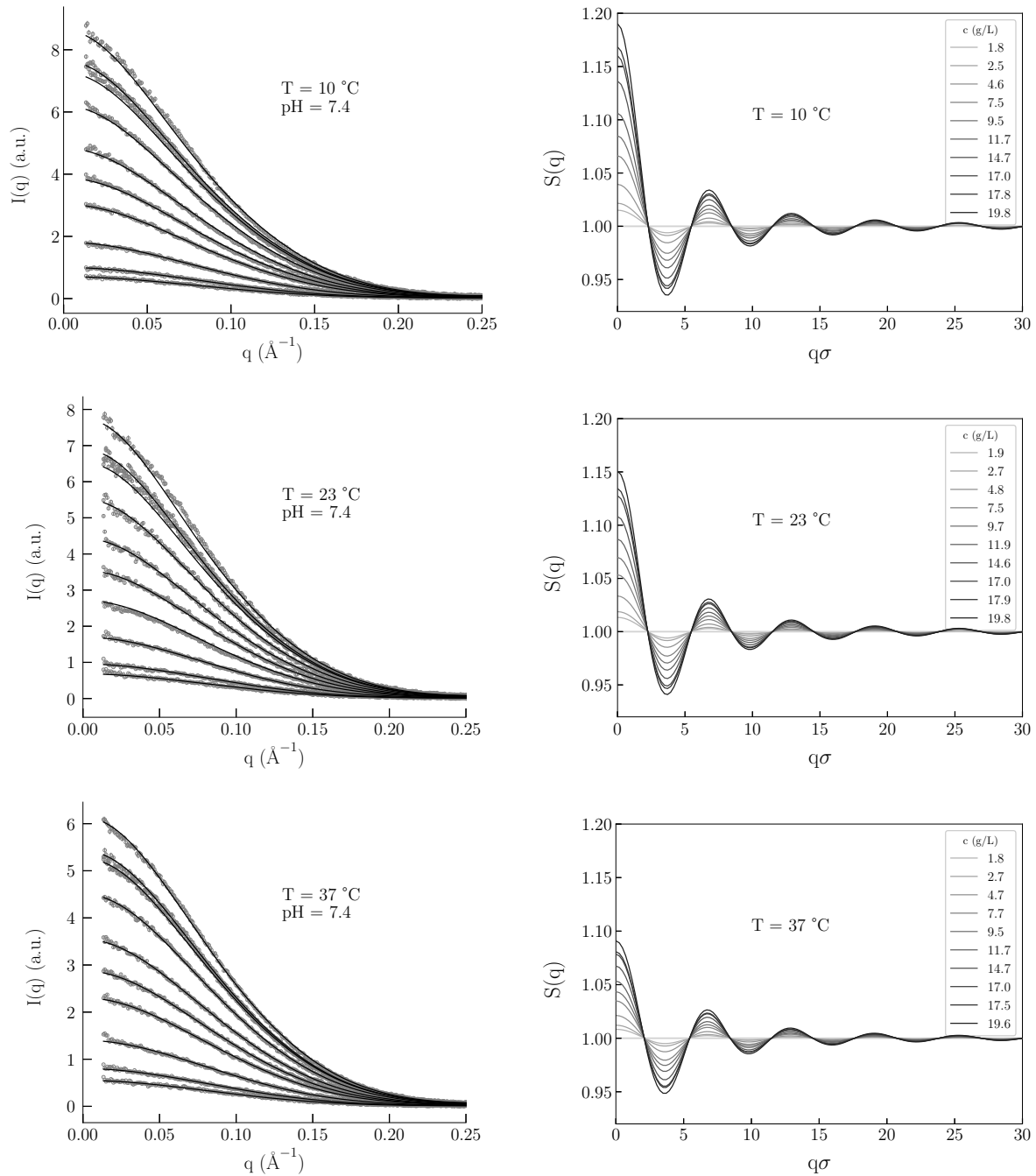


Figura 5.11: Coluna à esquerda: Ajustes das soluções de lisozima usando o GENFIT, com um potencial HSDY e a aproximação RPA. Coluna à direita: respectivos fatores de estrutura. O eixo horizontal, nesse caso, está multiplicado pelo diâmetro ajustado da lisozima, logo, está em unidades adimensionais.

de minimização do ajuste.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)
10	7,8	4,5	7,3	30,0
23	7,4	4,5	7,3	30,5
37	6,5	4,5	7,3	30,6

Tabela 5.4: Parâmetros do ajuste da lisozima com $[\text{NaCl}] = 150 \text{ mM}$, $\text{pH} = 7.4$ e dez concentrações entre 2-20 g/L.

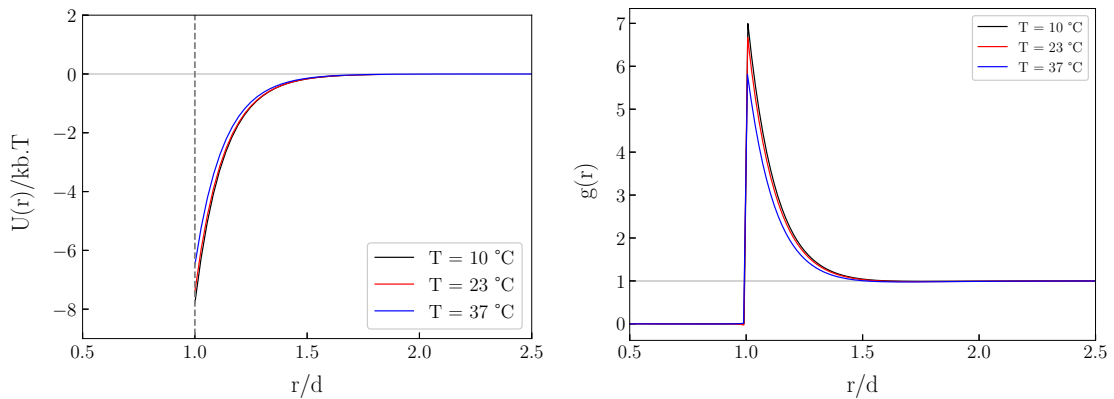


Figura 5.12: Esquerda: potencial efetivo ajustado (HSDY). Direita: funções de distribuição radial lisozima-lisozima derivadas dos ajustes. Distâncias em unidades adimensionais, normalizadas pelo diâmetro das proteínas ($d = \sigma$).

Na figura 5.12 podemos ver a energia potencial efetiva e as distribuições radiais. Notavelmente, pouquíssima diferença é percebida entre as curvas nas várias concentrações e temperaturas analisadas. Em relação à energia, há uma leve tendência de aumento da atração nas menores temperaturas, como observado anteriormente através dos fatores de estrutura, lá muito mais evidente do que aqui. Já em relação às funções $g(r)$ vemos um leve aumento no pico das distribuições em baixas temperaturas, indicando uma maior concentração relativa de proteínas ($\rho(r)/\rho$) ao redor de uma dada proteína. Tanto o potencial quanto as correlações de densidade têm um alcance de meio diâmetro. Comparadas às funções ajustadas anteriormente (figuras 5.8 e 5.6, lado inferior), elas possuem um comportamento substancialmente idêntico, com exceção dos pequenos picos (barreiras) de potencial que aparecem lá e que praticamente desaparecem aqui, o que agrava ainda mais a questão da estabilidade coloidal.

Em suma, apesar da variação na concentração de lisozima, na faixa entre 2-20 g/L (0.136-1.26 mM), alterar significativamente as curvas de espalhamento de raios-X (ao menos a 10 e 23 °C), quase nenhuma alteração é produzida sobre a distribuição radial e, portanto, sobre o potencial de força média. A pouca alteração é no sentido de um suave incremento na atração e na correlação espacial entre as lisozimas com a redução da temperatura.

1.3 Análises com Monte Carlo

As análises anteriores, utilizando o modelo de ajuste HSDY com fechamento RPA, mostraram algumas limitações no uso dessa abordagem, as quais passaremos a investigar com mais detalhes através da técnica de Monte Carlo. Em primeiro lugar, são feitos testes de aplicabilidade do modelo, tanto do potencial de interação quanto da aproximação RPA, e, em seguida, apresentaremos uma nova abordagem de ajuste das curvas de SAXS por meio de simulações de MC que permite evitar os erros de aproximação inerentes às relações de fechamento.

Teste do potencial de blindagem (SC)

Para verificarmos a validade do modelo de ajuste, vamos primeiro analisar se algumas hipóteses envolvidas na dedução do potencial de Coulomb blindado (Screened Coulomb, SC) estão de fato sendo satisfeitas. Como sabemos, este potencial foi deduzido por Verwey e Overbeek [20], que se basearam nos trabalhos anteriores de Derjaguin e Landau [19], tendo como ponto de partida a equação linearizada de Poisson-Boltzmann (PB)², também conhecida como equação de Helmholtz. As duas principais hipóteses que justificam a linearização da equação PB são as seguintes [69]: (1) o potencial de força média (pmf) é suposto igual à energia potencial média $U(r) = eZ\psi(r)$ (sendo ψ o potencial elétrico médio entre os macroíons), o que geralmente é o caso para sistemas suficientemente diluídos, onde os efeitos de muitos corpos se tornam insignificantes. Pode-se demonstrar também, por meio do equilíbrio mecânico, que essa hipótese equivale a assumir que os íons que formam as duplas-camadas não interagem entre si; e (2) $\psi(r)$ deve ser pequeno o suficiente para permitir a linearização da exponencial da equação PB, ou seja, $ve\psi/k_bT \ll 1$, onde v é a valência dos microíons e e a carga elementar.

A primeira hipótese pode ser testada pela obtenção do pmf "exato", calculado por simulações de Monte Carlo de um sistema de esferas duras carregadas interagindo através do potencial SC, nas mesmas condições físico-químicas dos nossos experimentos. Se estes dois potenciais forem equivalentes, então estaremos dentro dos limites dessa aproximação. As simulações foram realizadas com 30 macroíons de carga 7 e diâmetro 28 Å, além dos microíons de sal e contraíons de raio de 1 Å, todos monovalentes. O pmf foi calculado pela conversão das funções de distribuição radial. O que observamos, Figura 5.13, foi que a pmf simulada (que representa a função "exata") não corresponde ao potencial SC correspondente do sistema quando a concentração de sal é nula, onde a única contribuição para a força iônica provem dos contra-íons. Por outro lado, quando 50 ou 150 mM de sal foram adicionados à solução protéica, os dois potenciais de fato convergem, de modo que somente nestes casos a hipótese (1) está sendo satisfeita. No entanto, na prática, sabemos que as nossas soluções sem adição de NaCl também possuem uma força iônica

²Cf. Apêndice A para uma dedução dessa equação.

a mais devido ao tampão e aos ácidos e bases usados na correção do pH, o que tende a favorecer o cumprimento da hipótese.

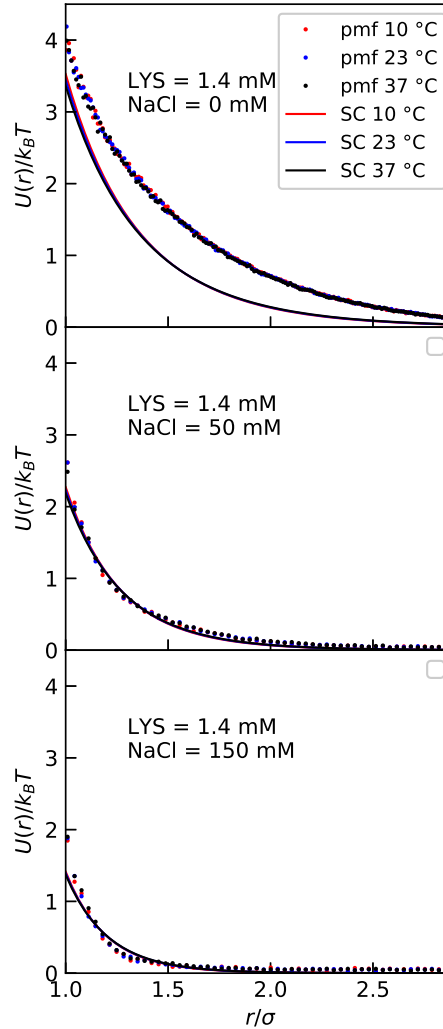


Figura 5.13: Comparação entre os potenciais de força média (pmf) e os potenciais de Coulomb blindado (SC) de um sistema de esferas duras carregadas, com $Z = 7$ e $\sigma = 28 \text{ \AA}$. A concentração de macroíons (nossas "proteínas") foi fixada em 1,4 mM e três condições de sal foram testadas: 0, 50 e 150 mM. As cores vermelha, azul e preta denotam temperaturas de 10, 23 e 37 °C, respectivamente.

Com relação à segunda hipótese, podemos apenas graficar a seguinte função usando a função energia potencial derivada dos nossos ajustes:

$$\frac{e\psi(r)}{k_b T} = \frac{U(r)}{Zk_b T} \quad (5.3)$$

já considerando a valência dos microíons (monovalentes) $v = 1$. Embora esta seja uma verificação *a posteriori*, é óbvio que o potencial ajustado precisa ser consistente com a hipótese inicial. A Figura 5.14 mostra os gráficos da função acima em três concentrações de sal, 1,36 mM de proteínas e $T = 10 \text{ °C}$ (as curvas para outras temperaturas são

praticamente as mesmas). Como se pode ver, o caso que menos satisfaz a condição $e\psi/k_bT \ll 1$ é novamente aquele que corresponde à menor força iônica. Como nenhuma das hipóteses (1) e (2) é totalmente verificada quando temos uma solução protéica não diluída e com baixa concentração de sal, decidimos investigar se isso poderia causar distorções nos fatores de estrutura e, conseqüentemente, nos parâmetros de ajuste .

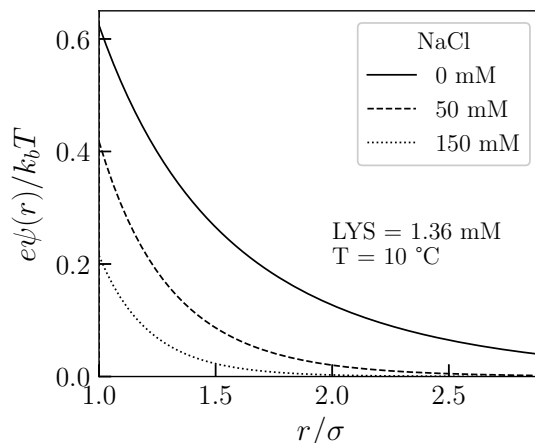


Figura 5.14: Potencial elétrico reduzido das soluções com 0 (sólido), 50 (tracejado) e 150 (pontilhado) mM NaCl, 1,36 mM lisozima e $T = 10^\circ \text{C}$. A aproximação DLVO requer que esse potencial assuma pequenos valores ($\ll 1$) para que possa ser aplicado, o que pode ser favorecido, por exemplo, pela adição de sal.

Assim, dando continuidade aos testes do potencial SC usando Monte Carlo, iremos agora simular e comparar os fatores de estrutura de um sistema composto por esferas carregadas interagindo através do potencial SC e de outro sistema composto pelas mesmas macroesferas, mas também pelos microíons (contraíons e íons de sal), todos interagindo através do potencial de Coulomb (C) usual. Espera-se que, se o potencial SC de fato é uma boa aproximação para o potencial médio entre os macroíons deste segundo sistema, conforme descrito por Verwey e Overbeek [20], então os fatores de estrutura dos dois devem ser muito próximos.

Realizamos essa comparação em diferentes valores de concentração proteica (0,2, 0,8 e 1,4 mM), temperatura (10, 23 e 37 °C) e sal (0, 50 e 150 mM), totalizando 27 condições distintas, algumas delas estão mostradas na figura (5.15). Como se pode observar, os fatores de estrutura SC e C são bastante semelhantes nos casos mais diluídos (0,2 mM) e/ou com maior força iônica (150 mM). Apenas pequenos desvios ocorrem à medida que a concentração aumenta e/ou o sal é removido, principalmente no início das curvas (menores ângulos q) e na região do pico principal. Essas diferenças, por si só, não nos parecem capazes de alterar significativamente os ajustes das curvas SAXS, mas podem ser uma grande fonte de erro para densidades maiores que 1,4 mM e baixo teor de sal. Assim, embora as hipóteses que testamos anteriormente não estejam sendo totalmente satisfeitas em alguns casos, o potencial SC parece uma aproximação razoável

nas condições estudadas, pelo menos em termos do fator de estrutura. O comportamento das curvas nas demais temperaturas, 23 e 37 °C, é completamente análogo.

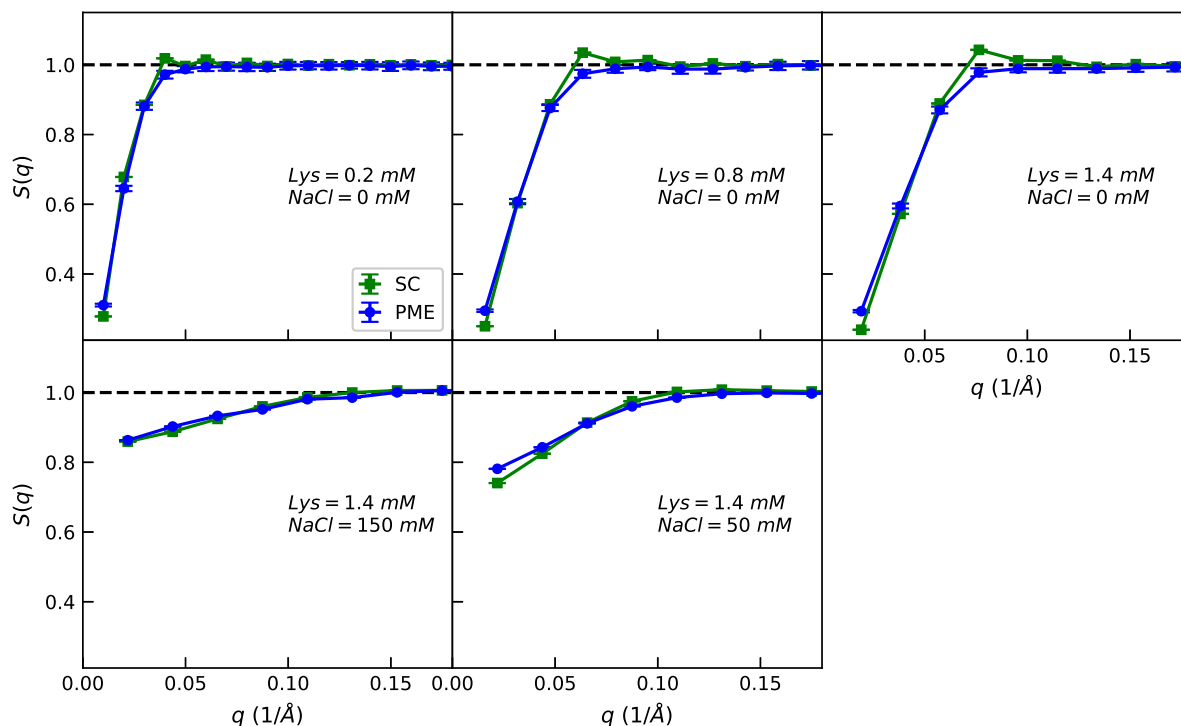


Figura 5.15: Comparação dos fatores de estrutura produzidos pelo potencial de Coulomb padrão, calculado pelo método PME, e o potencial de Coulomb blindado (SC). Os sistemas simulados foram compostos por 50 esferas rígidas carregadas ($Z = 7$, $\sigma = 28 \text{ \AA}$) com e sem microíons explícitos, respectivamente. Os gráficos mostram diferentes concentrações de macroíons e sal, todos a 10 °C.

Teste da relação de fechamento (RPA)

Como o potencial SC parece introduzir apenas pequenos desvios em nossos ajustes, decidimos então investigar se o mesmo ocorreria para a aproximação RPA. Para isso, foram feitos dois tipos de análise: (1) utilizando os mesmos parâmetros obtidos nos ajustes experimentais com o modelo HSDY/RPA (tabelas 5.1, 5.2 e 5.3) fizemos simulações de MC dos sistemas correspondentes e comparamos os fatores de estrutura obtidos, antes, pelo ajuste experimental e pelo agora simulado; na segunda análise fizemos o inverso: (2) usando os fatores de estrutura simulados no item anterior, derivamos as curvas teóricas de espalhamento (multiplicando as funções $S(q)$ pelo fator de forma da esfera) e então ajustamos essas curvas com o modelo HSDY/RPA para verificar se este era capaz de reproduz os parâmetros de entrada das simulações. Essas duas análises configuram testes de consistência importantes e que não encontramos em trabalhos do tipo na literatura.

Em relação ao primeiro teste (simulação com os parâmetros dos ajustes) observamos que há uma clara diferença, cf. figura 5.16, entre os fatores de estrutura ajustado e

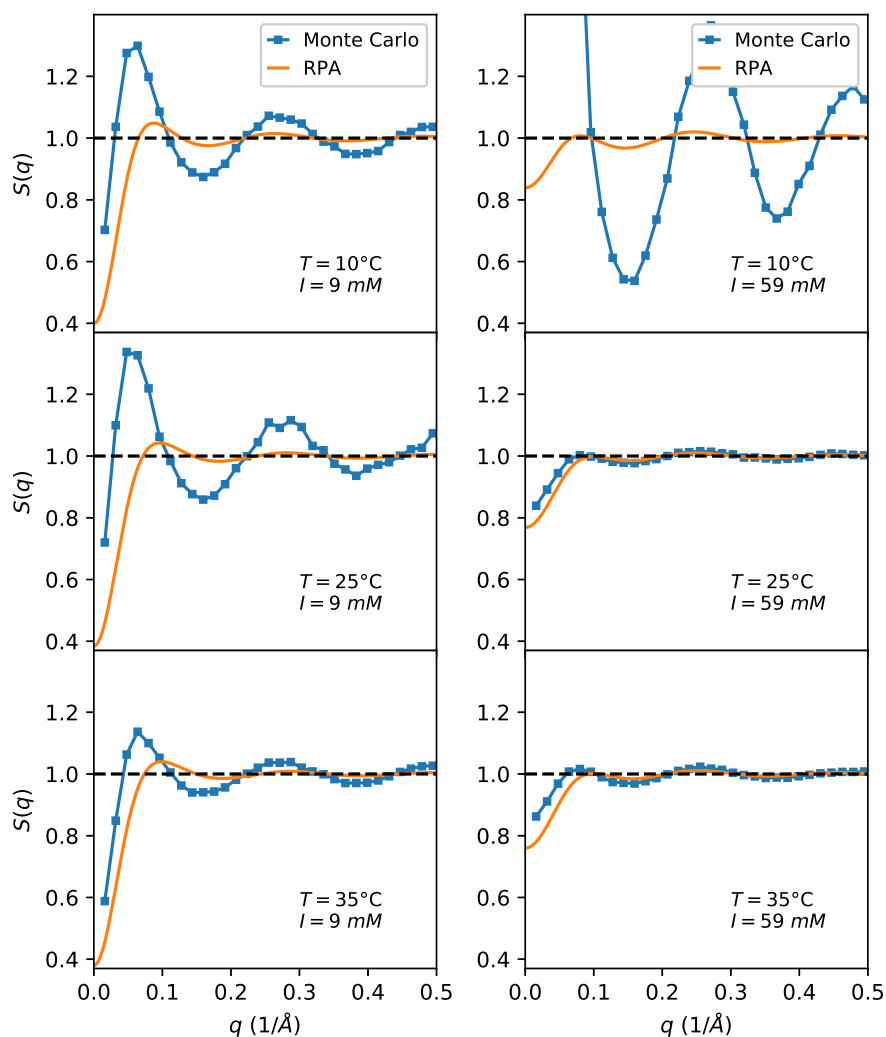


Figura 5.16: Comparação das funções $S(q)$ obtidas por ajuste experimental (linha sólida) e por simulação de MC (quadrados) em duas forças iônicas, 9 e 59 mM, e três temperaturas, 10, 25 e 35 °. A concentração de macroíons foi fixada em 1,36 mM. As simulações de MC foram realizadas com os mesmos parâmetros obtidos dos ajustes dos dados experimentais, nos quais foi aplicado o modelo HSDY sob a relação RPA.

simulado para as soluções sem adição de NaCl (cuja força iônica I fora ajustada em 9 mM). As funções $S(q)$ simuladas a 10 e 25 °C, por exemplo, mostram oscilações com amplitudes muito maiores do que as $S(q)$ derivadas do ajuste dos dados experimentais. Essa diferença só se torna menor a 35 °C, mas os picos continuam fora de fase. Aqui, evidentemente, a relação de fechamento RPA falha em termos de consistência com a simulação de MC, muito mais do que o mero potencial SC. Uma melhor consistência é observada quando elevamos a força iônica para 59 mM (veja a figura 5.16). Neste caso, há uma melhor coincidência das $S(q)$ a 25 e 35 °C, por exemplo. A única exceção foi a curva a 10 graus Celsius, que sofreu uma transição de fase nas simulações (agregação), ver figura 5.17. Da mesma forma, todas as simulações com 150 mM de sal também apresentaram agregação das esferas, provavelmente devido ao forte potencial atrativo

previsto pelo ajuste. Isso indica outra inconsistência dos parâmetros do ajuste, uma vez que as proteínas nas soluções reais não agregaram nessas condições. Talvez um estudo mais detalhado do diagrama de fases do modelo HSDY em questão possa revelar melhor as condições sob as quais ele leva a previsões não físicas como essa.

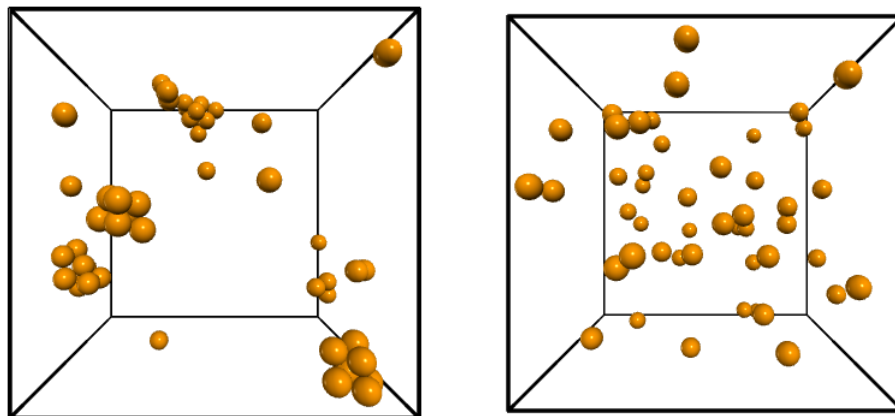


Figura 5.17: Configurações finais de duas simulações de MC, uma onde ocorreu agregação (figura à esquerda) e outra onde não houve (à direita). A primeira corresponde à temperatura de 10 °C e a segunda de 25 °C, ambas com 59 mM de força iônica. Os parâmetros usados nas simulações foram aqueles obtidos no ajuste com o modelo HSDY/RPA.

Por fim, tomando os fatores de estrutura simulados acima, derivamos as curvas de intensidade teóricas, $I(q) \propto P(q)S(q)$, correspondentes. Ao ajustarmos essas curvas com o modelo HSDY/RPA, obtivemos resultados que reforçam o que viemos observando até o momento. A saber, não conseguimos reproduzir, pelo novo ajuste, os parâmetros usados nas simulações das soluções densas (1,36 mM) e com baixa baixa força iônica (9 mM). A Figura 5.18, à esquerda, mostra três tentativas de ajustar esta condição a uma temperatura de 25 °C (resultados análogos são obtidos a 10 e 35 °C). O primeiro ajuste (fit 1) foi feito fixando os parâmetros iguais àqueles utilizados na simulação e os demais (ajuste 2 e 3) deixando esses parâmetros livres para variar ao redor dos valores exatos. Como podemos ver, cf. legenda da figura 5.18, embora seja possível reproduzir valores semelhantes para o diâmetro e a carga das esferas, os parâmetros do termo atrativo, J e D , apresentam desvios percentuais muito grandes. Esses desvios se tornam menores, mas não desprezíveis, à medida em que reduzimos a concentração de proteína para 0,8 e 0,2 mM (dados não mostrados). Por outro lado, para forças iônicas de 59 mM conseguimos ajustes muito mais robustos, veja a figura 5.18. Neste caso, podemos encontrar conjuntos de parâmetros com pequenos desvios percentuais em relação aos valores de referência e também valores de qui-quadrado menores, de onde podemos inferir uma clara consistência entre os ajustes e as simulações (considerando as duas análises que fizemos, um é capaz de reproduzir o outro com baixa margem de erro).

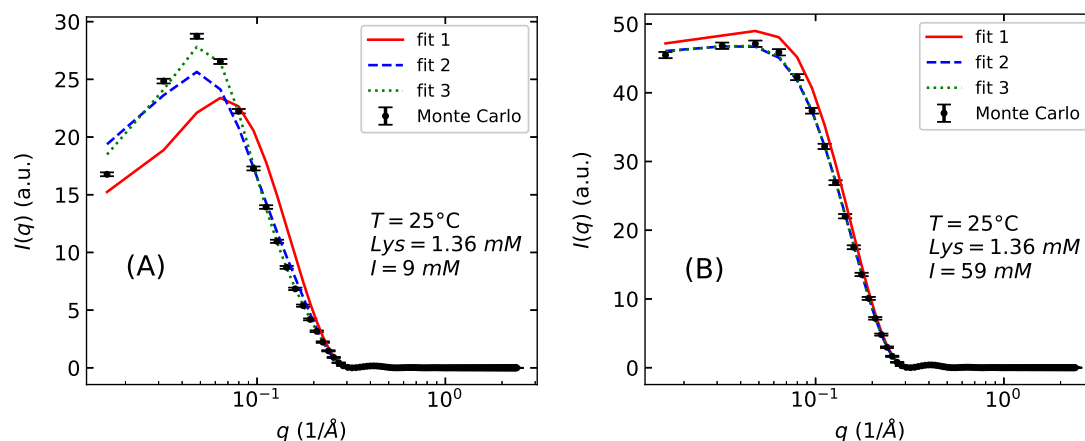


Figura 5.18: Ajustes (fit 1, 2 e 3) das curvas de intensidade derivadas das simulações (pontos pretos) para duas forças iônicas diferentes, (A) 9 e (B) 59 mM. A concentração de proteína e a temperatura estão fixadas em 1,36 mM e 25 °C, respectivamente. Os parâmetros ajustados e os valores de qui-quadrado, (Z, σ, J, D, χ^2), são: (A) fit 1 = (4.8, 28, 8.7, 1.6, 5.7), fit 2 = (4.2, 28.4, 12.2, 4.6, 31.4), fit 3 = (4.6, 27.6, 73, 1.2, 5.7); (B) fit 1 = (7.4, 28.4, 5.5, 4.4, 15.4), fit 2 = (7.1, 28.4, 5.1, 5.2, 0.3), fit 3 = (7.1, 28.4, 6.2, 4.3, 0.2). As unidades são as mesmas da tabela 5.1.

Apesar da carga e do diâmetro terem sido os parâmetros que apresentaram os menores desvios nesses últimos ajustes, é importante notar que isso também pode ser reflexo da maior quantidade de dados que a simulação produz em relação aos experimentos de SAXS. Enquanto que na simulação obtivemos valores de $S(q)$ até $q = 2.4 \text{ \AA}^{-1}$ com desvios padrão muito pequenos, no experimento medimos q no máximo até 0.5 \AA^{-1} com flutuações consideráveis nos maiores ângulos. Sem contar que o fator de forma das esferas apresenta oscilações, que dependem do diâmetro, muito mais bem definidas (e mais fáceis de ajustar) do que o comportamento do fator de forma de uma proteína. Assim, os desvios que observamos no ajuste dos dados simulados ainda subestimam os desvios que são produzidos no ajuste das curvas experimentais.

Além disso, existem vários trabalhos na literatura de SAXS que usam como modelo o potencial HSDY, em conjunto com alguma relação de fechamento, que relatam desvios na carga da lisozima em baixas concentrações de sal (tabela 5.5). Comentaremos eles a seguir.

No nosso caso, a carga ajustada para as soluções com baixo teor de sal, $Z = 4.8$ (Tabela 5.1), está bastante subestimada em comparação com os ajustes a 50 e 150 mM de NaCl, mas também com a previsão do PROPKA e com algumas curvas de titulação tabuladas [70, 71], as quais estimam uma carga mais próxima de $Z \approx 7 - 8$ nos pHs em que estamos trabalhando (7.4 e 8). Ortore et al. [72] também obtiveram uma carga de lisozima subestimada de aproximadamente 5.5 em pH 7, trabalhando com soluções densas (30-120 mg/ml) e nenhum sal além do tampão (cf. Tabela 5.5). Em um trabalho posterior [73], em contraste, utilizando soluções com maior força iônica devido à adição

de 30 mM NaOH, os autores relataram uma carga ajustada em torno de 8.5, em um pH semelhante. Já nos trabalhos de Winter et al. [74, 62, 75], embora eles trabalhem com soluções densas (100 mg/ml) e sem sal, em pH 7 e tampão Tris 25 mM, eles sempre consideram a carga de lisozima fixa a priori em seu valor esperado, $Z = 8$. Nesse caso, não sabemos se Z seria subestimado se fosse tratado como um parâmetro livre.

Alguns outros autores que trabalharam com soluções densas de lisozima e baixa concentração de sal também relataram uma subestimação da carga ao usar especificamente o modelo HSDY/RPA (cf. Tabela 5.5). Por exemplo, Tardieu et al. [61] estudando soluções de lisozima a 100 mg/ml em pH 4,5, obtiveram um valor de carga de $Z = 6$ para $[\text{NaCl}] = 0, 35$ e 70 mM, muito inferior ao o valor esperado, $Z \approx 10$, neste pH [70, 71]. Além disso, Narayanan e Liu [24] também encontraram um valor de carga de 6 quando ajustaram soluções de 40, 60 e 80 mg/ml de lisozima em pH 4,6 e concentração salina de 8,6 mM. É importante notar que esses autores trabalharam com densidades e cargas protéicas muito superiores às nossas, de modo que esses sistemas se desviam ainda mais das condições ideais de ajuste. Assim, ao contrário do que esses autores afirmam, o modelo HSDY sob RPA não é tão adequado para explicar as interações lisozima-lisozima nessas condições.

	LYS (mg/ml)	salt (mM)	pH	buffer (mM)	T (°C)	Z (e)	σ (Å)	J (k_bT)	D (Å)
Tardieu et al. (1999) [61]	100	0 – 350 (NaCl)	4.5	50 (acetate)	18	6 – 2.1	32.4	2.65	3
Narayanan et al. (2003) [24]	40 – 80	8.6 (NaCl)	4.6	H ₂ O	35	6	36	10 – 8	3.8 – 3.3
Ortore et al. (2008) [72]	30 – 120	0	7.0	100 (phosphate)	25	~ 5.5	*	*	2.5
Ortore et al. (2009) [73]	100	30 (NaOH)	7.5	D ₂ O	20	~ 8.5	*	~ 10	~ 3
Winter et al. (2011) [74]	100	0	7.0	25 (Bis-Tris)	5 – 25	8	29.9	7 – 4.4	2.7
Winter et al. (2012) [62]	~ 5 – 170	0	7.0	25 (Bis-Tris)	8 – 45	8	29.9	6 – 2	3
This paper (2021)	2 – 20	0	7.4	10 (PBS)	10 – 40	4.9	28	8 – 8.5	3 – 1

Tabela 5.5: Alguns trabalhos na literatura de SAXS de lisozima que utilizam o potencial HSDY para ajustar soluções com altas concentrações de proteína e baixas de sal. Os asteriscos representam dados que não são facilmente identificáveis nos artigos.

Ajuste das curvas de SAXS da lisozima com Monte Carlo

Do que foi discutido acima podemos afirmar que o ajuste das curvas de SAXS através do modelo HSDY/RPA está sujeito às limitações inerentes às aproximações envolvidas no modelo, em especial naquilo que se refere à relação de fechamento, mas também em menor escala ao potencial SC (e mesmo, possivelmente, à aproximação de Percus-Yevick para o fator de estrutura, já que estamos lidando com sistemas densos). A combinação dessas várias aproximações, ao que nos parece, pode levar à parâmetros ajustados sem sentido físico, ainda que as curvas sejam bem ajustadas, sobretudo em condições de baixa força iônica e alta densidade proteica.

Uma maneira de contornarmos o uso das relações de fechamento, como a RPA, PY ou outras, consiste em ajustar o fator de estrutura por meio de simulações de Monte Carlo. Neste caso, podemos manter como modelo apenas o potencial de interação HSDY,

tendo como entrada um certo conjunto de parâmetros iniciais³, e derivar o fator de estrutura $S(q)$ correspondente por meio de uma média termodinâmica no ensemble de configurações. A partir daí é possível obter, analogamente ao que fizemos antes, uma curva teórica para a intensidade, mas desta vez multiplicando a função $S_M(q)$ pelo fator de forma da proteína em questão:

$$I_{MC}(q) = nP(q)S_M(q) = n\langle|F|^2\rangle \left[1 + \frac{|\langle F \rangle|^2}{\langle|F|^2\rangle} (S(q) - 1) \right] \quad (5.4)$$

sendo n a densidade de proteínas. As funções $|\langle F \rangle|^2$ e $\langle|F|^2\rangle$, acima, podem ser calculadas através do arquivo PDB da proteína utilizando o próprio software GENFIT. A curva resultante pode então ser comparada com a curva experimental. Isso é feito visualmente graficando a curva de espalhamento medida, $I_{SAXS}(q)$, junto com a função de ajuste $I_{fit}(q) = \alpha I_{MC}(q)$, onde α é uma constante multiplicativa que podemos ajustar manualmente para acertar a escala das duas curvas de intensidade. Caso elas não se ajustem uma à outra, modificamos os parâmetros de entrada e repetimos todo o procedimento até obtermos um ajuste aceitável. Essa foi a estratégia geral que utilizamos aqui e, até onde pudemos constatar, ela representa uma novidade na literatura de SAXS de proteínas.

Os resultados desses ajustes podem ser conferidos na figura 5.19, onde, para cada concentração de sal, ajustamos três temperaturas distintas. Os parâmetros obtidos se encontram nas tabelas 5.6, 5.7 e 5.8.

Uma outra diferença com relação aos ajustes anteriores é que, ao invés de fixar a constante dielétrica igual a 78 (seu valor usual à temperatura ambiente), nós utilizamos agora uma equação empírica que leva em conta a sua dependência com a temperatura [76]:

$$\epsilon(T) \approx 10^{1.94315 - 0.00199720 * T} \quad (5.5)$$

A variação de ϵ entre 10-40 °C é pequena (cf. tabela 1.3) e não deve afetar os ajustes de forma significativa, mas decidimos considerar o efeito por uma questão de rigor.

T (°C)	$\epsilon(T)$
10	83.8359
23	79.0304
25	78.3159
35	74.8393
37	74.1628
40	73.1594

³Apenas como ponto de partida, tomamos como parâmetros iniciais dos ajustes com MC os mesmos parâmetros obtidos nos ajustes feitos com o GENFIT.

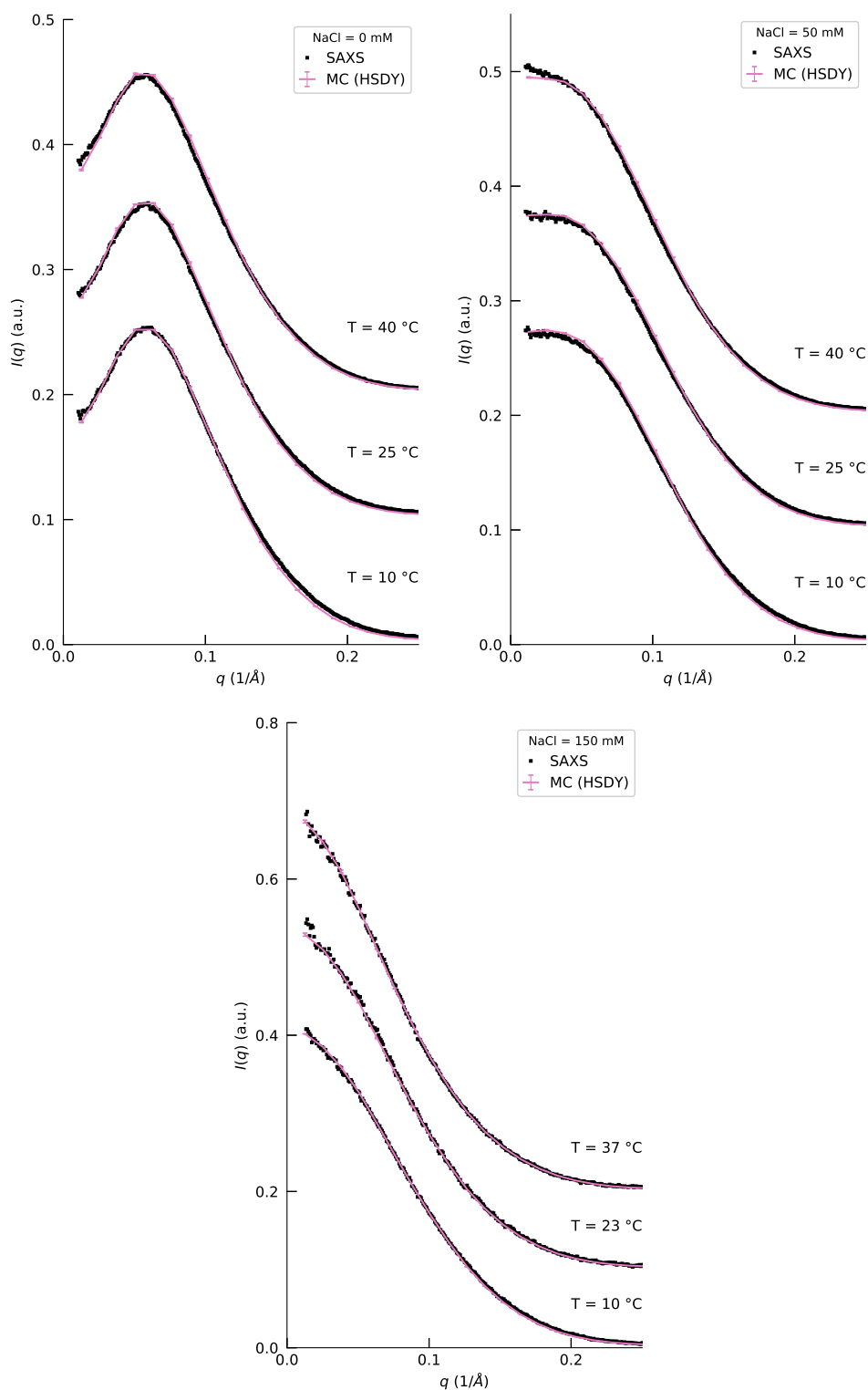


Figura 5.19: Ajustes das curvas experimentais de lisozima (ponto pretos) através de simulações de Monte Carlo (linha rosa). As curvas estão deslocadas na vertical apenas para facilitar a sua visualização.

Um resultado que chama a atenção foi que conseguimos obter bons ajustes usando os mesmos valores de D , Z e σ para todas as curvas acima. Os únicos parâmetros que tivemos que mudar, no caso, foram J e a força iônica I . É importante destacar que, com

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5	3	7	28	19
25	2	3	7	28	19
40	1	3	7	28	19

Tabela 5.6: Parâmetros ajustados com MC das amostras de lisozima a 1.36 mM, [NaCl] = 0 mM.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5	3	7	28	59.5
25	2	3	7	28	59.5
40	1	3	7	28	59.5

Tabela 5.7: Parâmetros ajustados com MC das amostras de lisozima a 1.36 mM, [NaCl] = 50 mM.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5.6	3	7	28	159.5
23	5.3	3	7	28	159.5
37	5.0	3	7	28	159.5

Tabela 5.8: Parâmetros ajustados com MC das amostras de lisozima a 1.36 mM, [NaCl] = 150 mM.

a técnica de Monte Carlo, foi possível ajustar as curvas com baixa força iônica usando uma carga bem próxima do seu valor esperado, o que não havíamos conseguido antes através do modelo HSDY/RPA. Isso reforça, novamente, a inadequação da relação de fechamento nessas condições. As inconsistências físicas que havíamos observado antes - distribuições radiais negativas e agregação nas simulações - também foram eliminadas nos novos ajustes.

Se compararmos as funções de distribuição radial e os fatores de estrutura ajustados com Monte Carlo (figs. 5.20 e 5.21) com aqueles obtidos através do ajuste com o modelo HSDY/RPA (figs. 5.6 e 5.7), notamos que o comportamento dessas curvas é bastante similar, porém com pequenas diferenças que valem ser apontadas. Enquanto todas as $g(r)$ ajustadas com a relação RPA possuem um pico localizado próximo à superfície das proteínas, nem todas as simuladas possuem. Neste caso, os picos vão se formando gradativamente conforme a concentração de sal aumenta ou a temperatura diminui. Além disso, na condição com 150 mM de NaCl as funções simuladas continuam tendo uma pequena região na qual $g(r) < 1$, ou seja, onde o potencial de força média é repulsivo, o que é essencial para que as proteínas não agreguem. Já em relação aos fatores de estrutura, o primeiro pico quase não aparece nas funções simuladas com 0 mM de NaCl, enquanto que este era bem visível nos ajustes com a RPA.

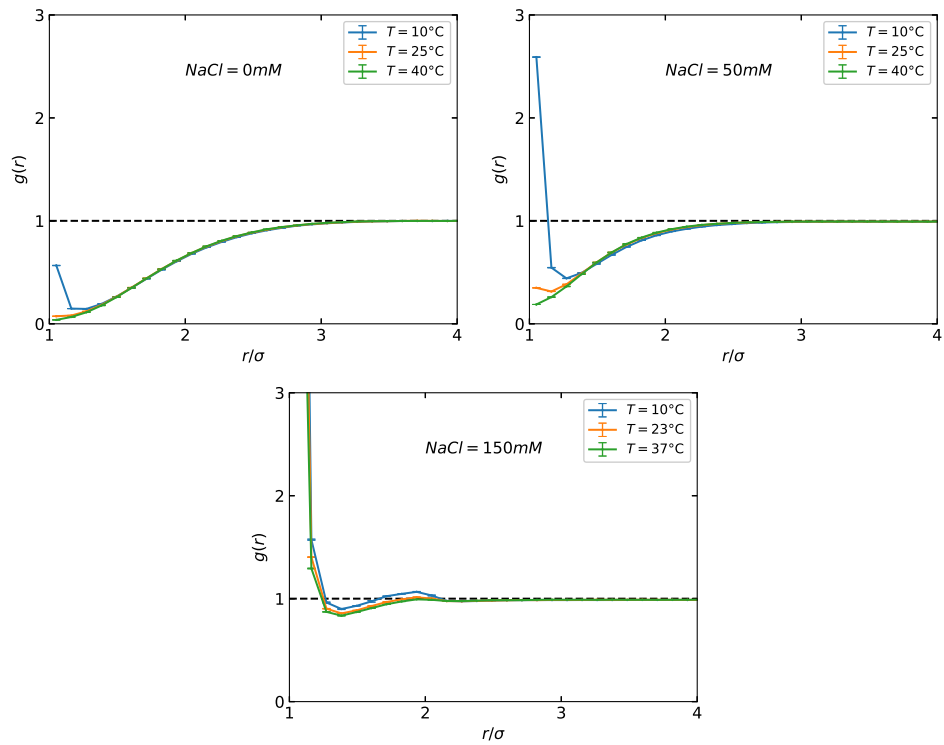


Figura 5.20: Distribuições radiais dos ajustes simulados com MC.

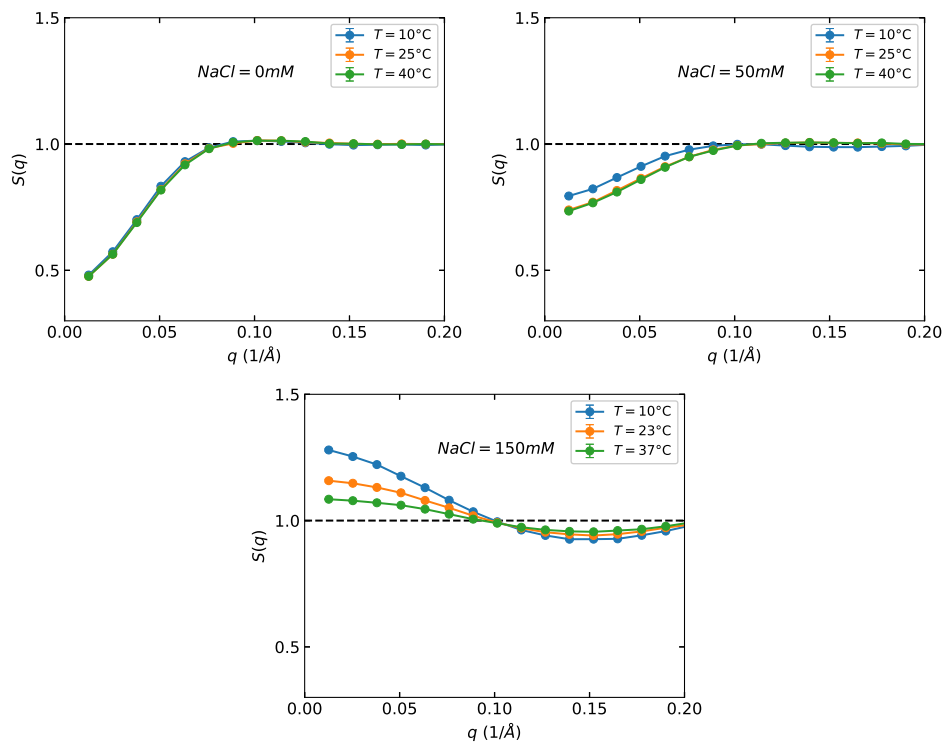


Figura 5.21: Fatores de estrutura dos ajustes simulados com MC.

Por fim, as interações proteína-proteína ajustadas com Monte Carlo (fig. 5.22) diferem dos potenciais obtidos com o modelo HSDY/RPA (fig. 5.8) nos seguintes pontos: estes últimos sempre apresentam uma pequena região atrativa próxima à superfície da proteína,

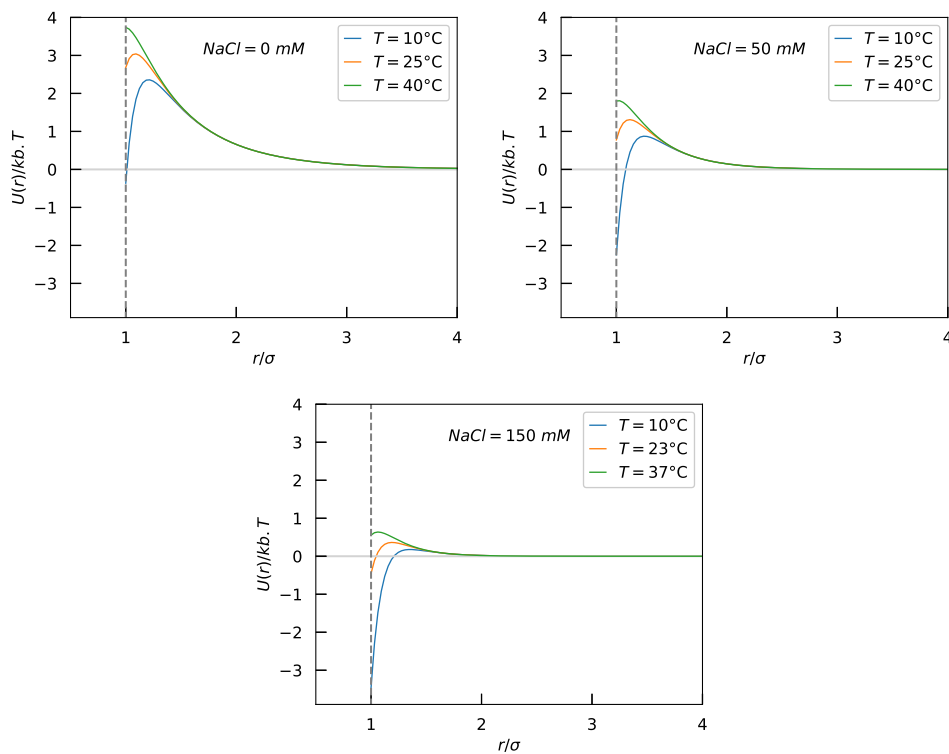


Figura 5.22: Potenciais médios de interação lisozima-lisozima dos ajustes simulados com MC.

que agora aparece apenas na temperatura de 10°C e sal igual ou superior a 50 mM , e um pouco em 23°C e 150 mM de sal. Em geral, os potenciais ajustados com MC são bem mais repulsivos que os anteriores. A maior barreira de potencial, neste caso, ajuda a explicar porque as soluções de lisozima permanecem estáveis e não sofrem agregação. Os valores de J encontrados aqui também são menores do que os anteriores, variando entre 1-5, para 0 e 50 mM de NaCl, e 5-5.6 para 150 mM de NaCl.

2 BSA

2.1 Efeito da temperatura

Assim como fizemos para a lisozima, graficamos juntas as curvas experimentais obtidas nas mesmas condições de densidade proteica e sal, mas variando a temperatura. Essas curvas foram alinhadas com base no fator de forma da estrutura cristalográfica do monômero de BSA (PDB 4F5S), como o objetivo de observarmos, qualitativamente, os desvios em relação a essa função e que geralmente podem ser atribuídos a mudanças no fator de estrutura, mas não somente. Diferentemente do que ocorre com a lisozima, não foi observado um efeito significativo da temperatura sobre a intensidade de espalhamento de raios-X da BSA. Os resultados estão representados nas figuras 5.23 (para uma mesma concentração estudada de 1.36 mM) e 5.24 (para todas as concentrações e 150 mM de sal), a seguir.

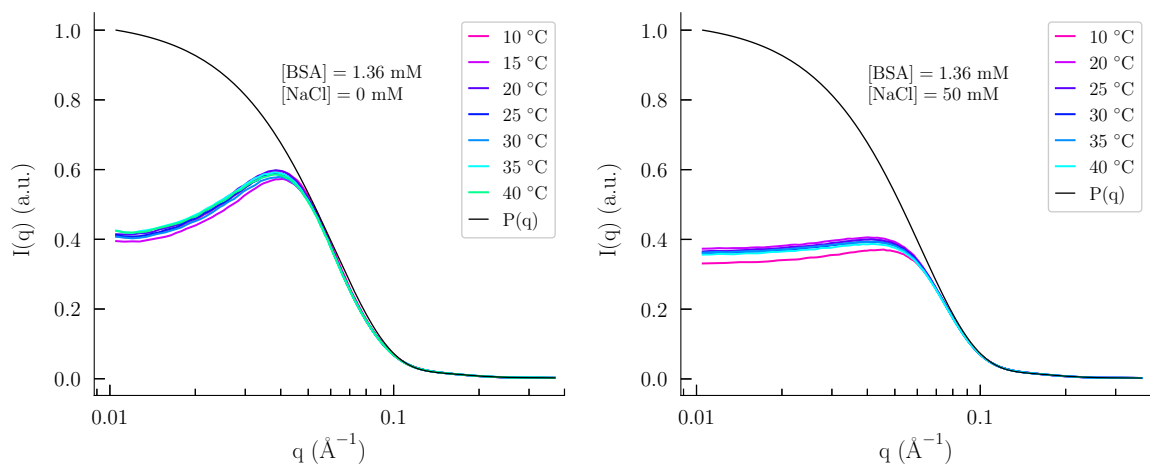


Figura 5.23: Efeito da temperatura no espalhamento da BSA a 1.36 mM e duas concentrações de NaCl: 0 e 50 mM. Os gráficos estão em escala linear-log. As curvas pretas representam o fator de forma da estrutura cristalográfica da BSA 4F5S (PDB).

No caso de 150 mM de sal e para a menor concentração de BSA (0.14 mM, ou, cerca de 9 mg/ml), ver figura 5.24 acima e à esquerda, notamos que as curvas não convergem para o fator de forma do monômero de BSA como seria o esperado no limite de baixas diluições. Esse desvio no início das curvas pode indicar que: (1) ou a solução não está diluída o suficiente para que possamos desprezar o fator de estrutura, ou (2) a amostra contém agregados oligoméricos [15]. Supondo que esta última hipótese seja verdadeira, realizamos um ajuste das curvas menos concentradas considerando que elas eram compostas de uma mistura de monômeros e dímeros. Neste caso, a intensidade de espalhamento seria dada por:

$$I(q) \propto xP_{\text{mon}}(q) + \frac{1-x}{2}P_{\text{dim}}(q) \quad (5.6)$$

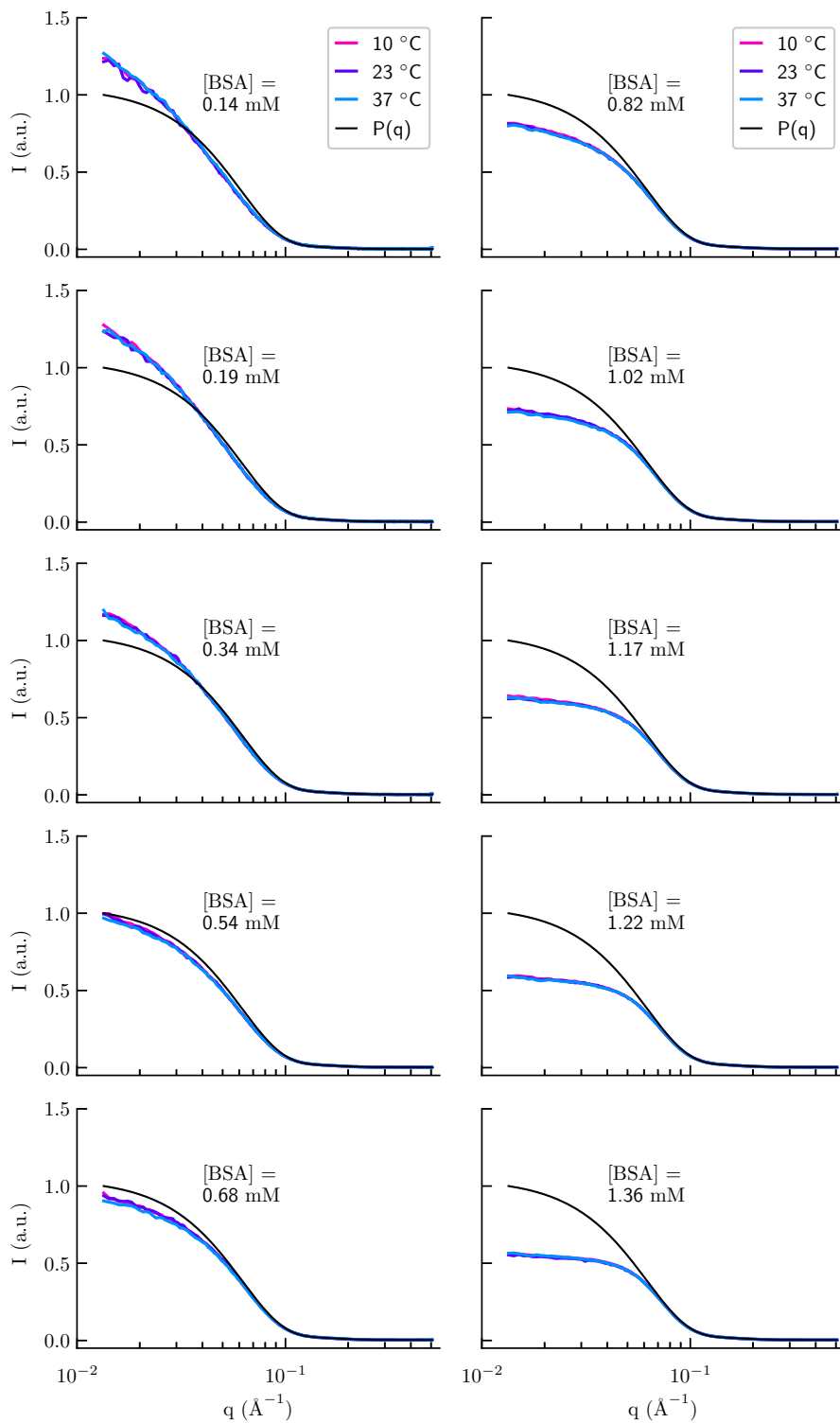


Figura 5.24: Efeito da temperatura no espalhamento da BSA em várias densidades da proteína e 150 mM de NaCl. Os gráficos estão em escala linear-log. As curvas pretas representam o fator de forma da estrutura cristalográfica da BSA 4F5S (PDB).

sendo x a fração de BSA na forma monomérica (número de monômeros dividido pelo número total de BSA) e $(1 - x)/2$ a fração de dímeros (número de dímeros dividido pelo número total de BSA).

Ajustando este modelo aos dados experimentais observamos uma melhor adequação teórica (Figura 5.25) em comparação com o modelo que considera apenas monômeros. A fração de monômeros encontrada aqui foi de $x = 0.54$, ou seja, cerca de metade das proteínas em solução estaria livre e a outra metade fazendo pares com outra proteína. Além disso, é possível perceber nos menores ângulos da figura que os pontos experimentais ainda estão um pouco acima da curva de ajuste, o que pode indicar a presença de estados oligoméricos maiores, como tetramêros, por exemplo, embora em quantidade bem inferior a dos monômeros e dímeros.

Ressaltamos, novamente, que essa análise por si só é inconclusiva quanto à presença ou não de oligômeros na solução e que a presença de um fator de estrutura nessas condições não pode ser descartada a priori. Neste caso, experimentos de DLS poderiam ter nos auxiliado a determinar o grau de polidispersividade das amostras, o que não fizemos à época das medidas.

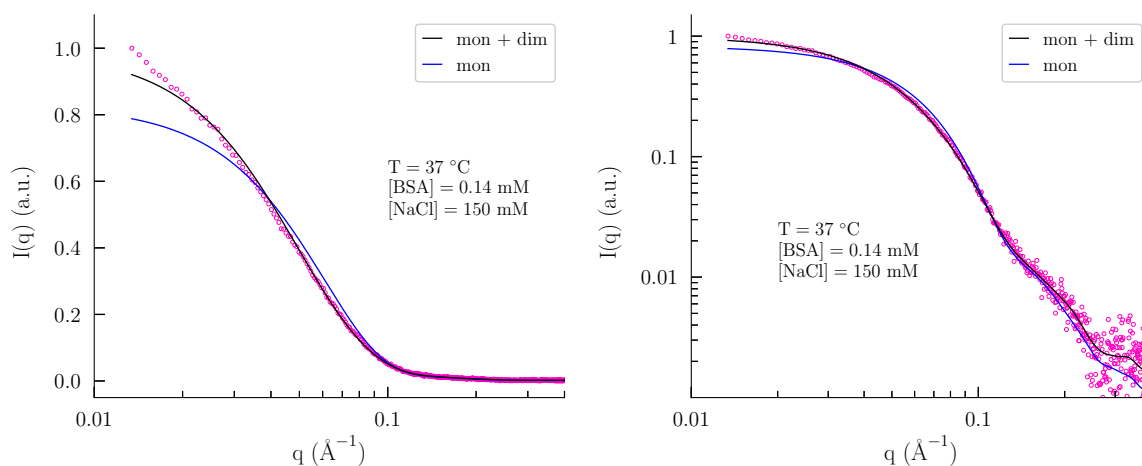


Figura 5.25: Ajuste da curva experimental da BSA a 0.14 mM (9 mg/ml), 150 mM de NaCl e 37 °C, representada pelos pontos rosas. A curva azul é o ajuste do modelo de monômeros e a curva preta do modelo de monômeros e dímeros. À esquerda, gráfico em escala linear-log e, à direita, o mesmo gráfico em escala log-log.

2.2 Efeito da concentração

A seguir, mesmo sabendo das limitações discutidas anteriormente a respeito do modelo HSDY/RPA, decidimos ajustar as curvas de espalhamento da BSA a 150 mM de sal em função da densidade utilizando este modelo. Aqui, portanto, todo desvio em relação ao fator de forma do monômero de BSA (como a oligomerização) será tratado como um efeito do fator de estrutura. O ajuste seguiu os mesmos procedimentos do caso da lisozima e, a partir dele, pudemos obter as funções dos fatores de estrutura, energias potenciais e as distribuições radiais (ver figuras 5.27 e 5.28, abaixo). No caso da BSA, não conseguimos realizar um ajuste global, ou seja, simultâneo, de todas as 10 concentrações medidas, pois

o modelo não era capaz de ajustar as curvas de maior concentração de BSA em solução. Assim, excluindo essas últimas do ajuste, pudemos ajustar simultaneamente somente as 6 curvas de menor densidade de partículas, ver figura 5.26. Quando a concentração é muito elevada (> 1 mM) o modelo começa a se mostrar inadequado, gerando parâmetros com pouco sentido físico ou inconsistências nas funções de distribuição radial. Muito provavelmente, nessas condições, as interações são fortes demais para serem tratadas dentro de uma teoria perturbativa, ou então o fator de forma de um monômero apenas em solução não é válido.

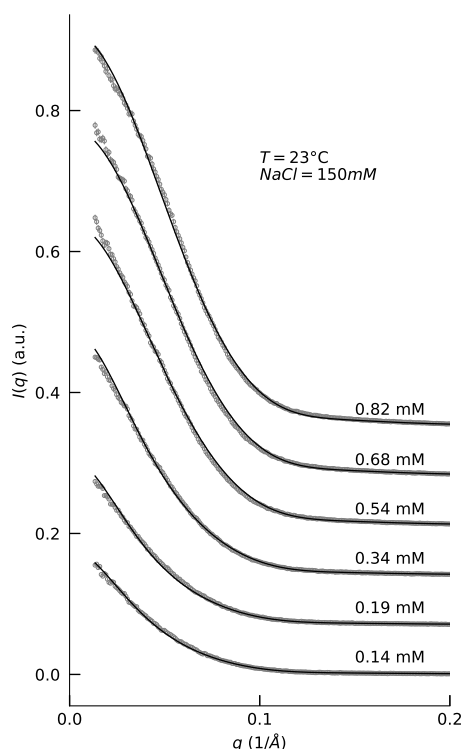


Figura 5.26: Ajuste global das 6 curvas experimentais menos concentradas de BSA, com 150 mM de NaCl e temperatura de 23 °C. O modelo utilizado foi o potencial HSDY com a relação de fechamento RPA.

Em relação aos fatores de estrutura, observamos a formação gradual de um pico nos menores ângulos de espalhamento a medida em que diminuimos a densidade de BSA, o que indica uma maior atração entre as proteínas quando estas estão mais diluídas. Isso se reflete de maneira evidente nas curvas de energia potencial e de distribuição radial. Os potenciais de interação resultaram predominantemente atrativos e de curto alcance, com uma barreira repulsiva insignificante (ver o zoom na figura 5.28). Já as $g(r)$ nos mostram um pico de densidade contido numa pequena região no entorno das proteínas (menor do que um raio), na qual a concentração proteica média é maior que a do bulk, e esse pico se torna mais acentuado conforme a concentração de BSA é reduzida. É interessante notar que, diferentemente do que observamos nos ajustes da lisozima, tanto a energia potencial quanto as distribuições radiais da BSA sofreram mudanças devido à

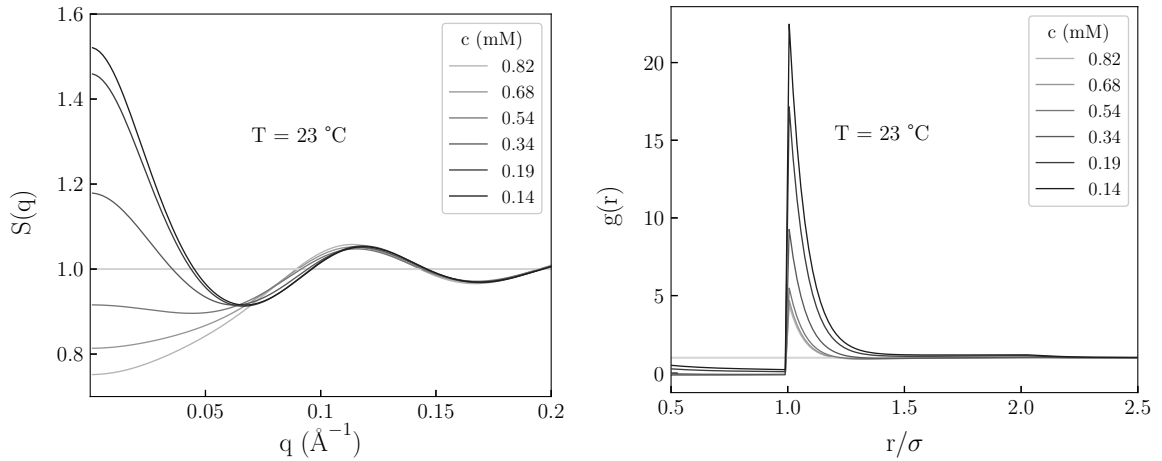


Figura 5.27: Fatores de estrutura ajustados e distribuições radiais das seis curvas de BSA menos concentradas entre 0.14 e 0.82 mM (9-54 g/L).

variação de densidade.

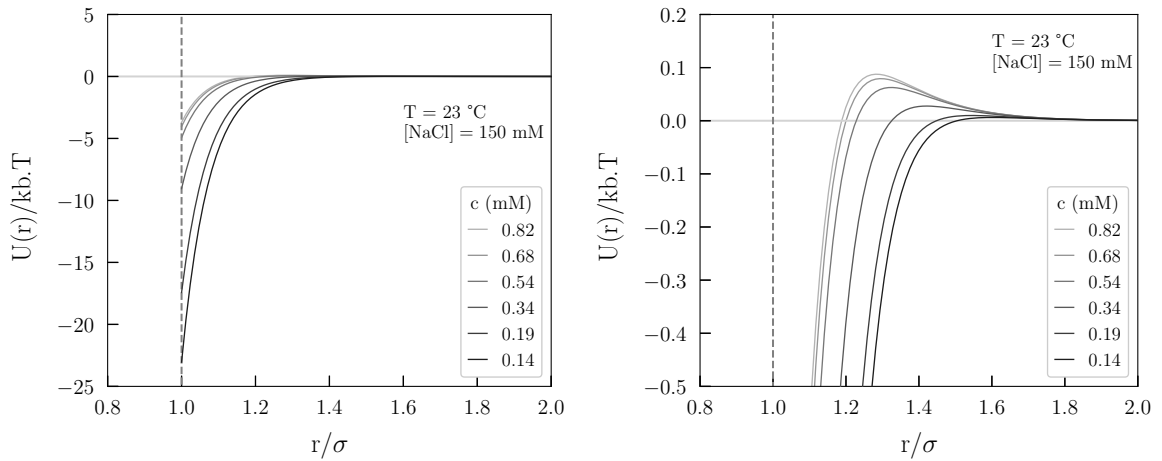


Figura 5.28: À esquerda, energia potencial ajustada das seis curvas de BSA menos concentradas (0.14-0.82 mM). À direita, um zoom das mesmas funções evidenciando o aumento de um pequeno pico repulsivo à medida que a concentração cresce.

BSA (mM)	BSA (g/L)	J (k_bT)	D (Å)	Z (e)	σ (Å)
0.14	8.9	26.0	5	-24	60
0.19	12.1	20.4	5	-24	60
0.34	21.5	12.0	5	-24	60
0.54	35.6	7.9	5	-24	60
0.68	46.0	7.0	5	-24	60
0.82	54.2	6.7	5	-24	60

Tabela 5.9: Parâmetros do ajuste da BSA com $[\text{NaCl}] = 150$ mM, $\text{pH} = 7.4$, $T = 23$ °C e seis concentrações entre 0.14-0.82 mM (ou, 8.9-54.2 g/L).

Os parâmetros do ajuste estão apresentado na tabela 5.9. Mantivemos D , Z e σ

como parâmetros comuns a todas as curvas e deixamos J livre para variar. A carga aqui encontrada, $Z = -24$, nos pareceu bastante sobrestimada em relação aos valores tabelados na literatura em condições semelhantes. Barbosa et al.[25], por exemplo, encontraram uma carga para a BSA de -13(2) em pH 7, assim como Tanford and Buzzel[77] em pH 7.3. Já Fologea et al.[78] obtiveram uma carga um pouco maior, $Z = -18$ em pH 7, mas ainda assim bem abaixo de -24. O software de predição de carga PROPKA, por sua vez, nos informa que a carga esperada para a BSA em pH 7.4, no qual as nossas medidas foram realizadas, ficaria em torno de -15 (utilizando como arquivo de entrada o monômero do PDB 4F5S).

2.3 Análise com Monte Carlo

Ajuste das curvas de SAXS da BSA com Monte Carlo

Tomando como referência o valor de carga previsto pelo PROPKA, $Z = -15$, tentamos ajustar com Monte Carlo as curvas de BSA na concentração máxima de 1.36 mM, em várias temperaturas e concentrações de NaCl. O procedimento foi o mesmo descrito anteriormente nos ajustes da lisozima.

Novamente, foi possível obter ajustes satisfatórios para as curvas experimentais mantendo a carga, o diâmetro (60 Å) e o parâmetro D (6 Å) constantes, ver figura 5.29 e as tabelas 5.10, 5.11 e 5.12. Os únicos parâmetros que precisaram ser variados caso a caso foram a força iônica I e o termo atrativo J . Note-se que mesmo as soluções sem NaCl só puderam ser ajustadas com uma força iônica não muito baixa de 26 mM, a qual se deve principalmente aos contraíons das proteínas. Já o termo J , ao contrário do que observamos com a lisozima, variou muito pouco com a temperatura e resultou em valores menores com a adição de sal.

As funções fator de estrutura, distribuição radial e energia potencial são mostradas nas figuras 5.30, 5.31 e 5.32, respectivamente. Assim como as curvas de espalhamento, essas funções têm uma dependência quase nula com a temperatura, sendo que apenas as curvas a 10 °C e 50 mM de sal destoaram um pouco das demais temperaturas. Nas amostras sem sal os fatores de estrutura exibem dois picos na região de $q < 0.15$ Å⁻¹, os quais somem e dão lugar a apenas um conforme NaCl é adicionado ao sistema. Nos menores ângulos, as funções sempre se encontram abaixo da unidade, não ocorrendo uma inversão para valores maiores do que um como vimos nas curvas da lisozima.

As distribuições radiais nos mostram que as proteínas tendem a ficar cada vez mais próximas umas das outras quando a concentração de sal aumenta, o que era algo esperado já que o sal reforça a blindagem eletrostática e reduz a repulsão entre elas. Isso se reflete, é claro, nos potenciais de interação, onde vemos a transição de um potencial basicamente repulsivo na ausência de sal, com uma barreira energética de $1 k_b T$, para um potencial quase nulo quando adicionamos 150 mM de NaCl. Nesse último caso, o sistema estaria se

comportando próximo a um fluido de partículas não-interagentes, de modo que apenas o volume excluído estaria ditando as suas propriedades.

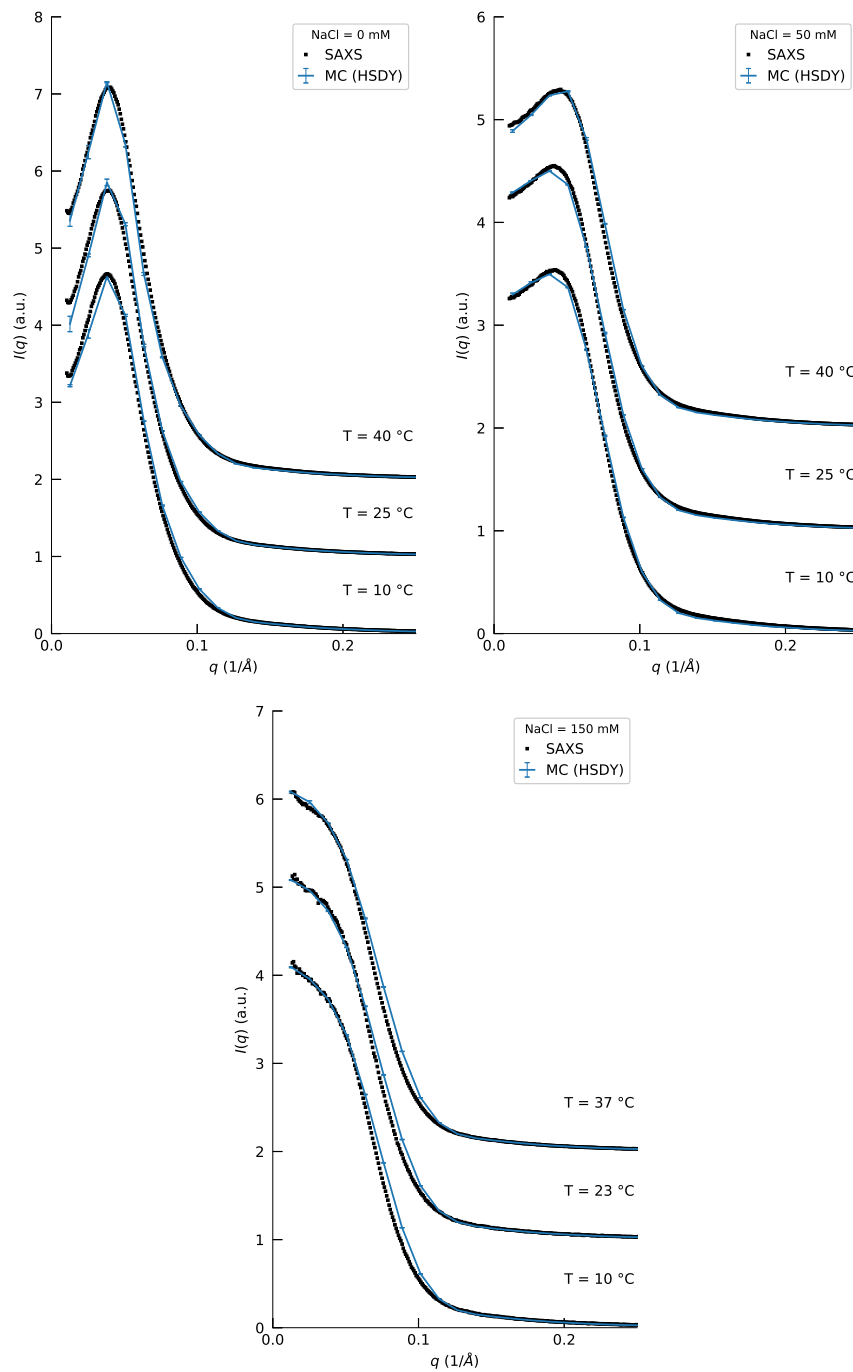


Figura 5.29: Ajustes das curvas experimentais de BSA (ponto pretos) através de simulações de Monte Carlo (linha azul). As curvas estão deslocadas na vertical para facilitar a visualização.

Talvez o fato da BSA possuir um tamanho maior que o da lisozima (tendo aproximadamente o dobro do diâmetro desta) e sendo quase duas vezes mais carregada explique a diferença no comportamento das duas proteínas em condições físico-químicas semelhantes. Ao que nos parece, proteínas menores e com carga não muito elevada são mais

suscetíveis a mudanças de temperatura e a variações na força iônica. Quando proteínas assim interagem há uma maior sobreposição das duplas camadas iônicas, cujos comprimentos de Debye dependem diretamente das grandezas mencionadas. No entanto, um estudo mais aprofundado seria necessário para podermos generalizar essas conclusões. Além disso, como as simulações de MC foram feitas com um único tamanho de esferas (60 Å), não conseguimos distinguir se o sistema experimental se encontra oligomerizado.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	6.8	6	-15	60	26
25	6.2	6	-15	60	26
40	6.2	6	-15	60	29

Tabela 5.10: Parâmetros ajustados com MC das amostras de BSA a 1.36 mM, [NaCl] = 0 mM.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	1	6	-15	60	72
25	2	6	-15	60	72
40	2	6	-15	60	72

Tabela 5.11: Parâmetros ajustados com MC das amostras de BSA a 1.36 mM, [NaCl] = 50 mM.

T (°C)	J (k_bT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	1	6	-15	60	170
23	1	6	-15	60	170
37	1	6	-15	60	170

Tabela 5.12: Parâmetros ajustados com MC das amostras de BSA a 1.36 mM, [NaCl] = 150 mM.

Com isso, terminamos as nossas análises envolvendo as proteínas (BSA e lisozima) puras, isto é, em solução simples. No próximo capítulo estudaremos as misturas binárias entre elas e, para isso, será fundamental os resultados que obtivemos aqui a partir dos ajustes com Monte Carlo.

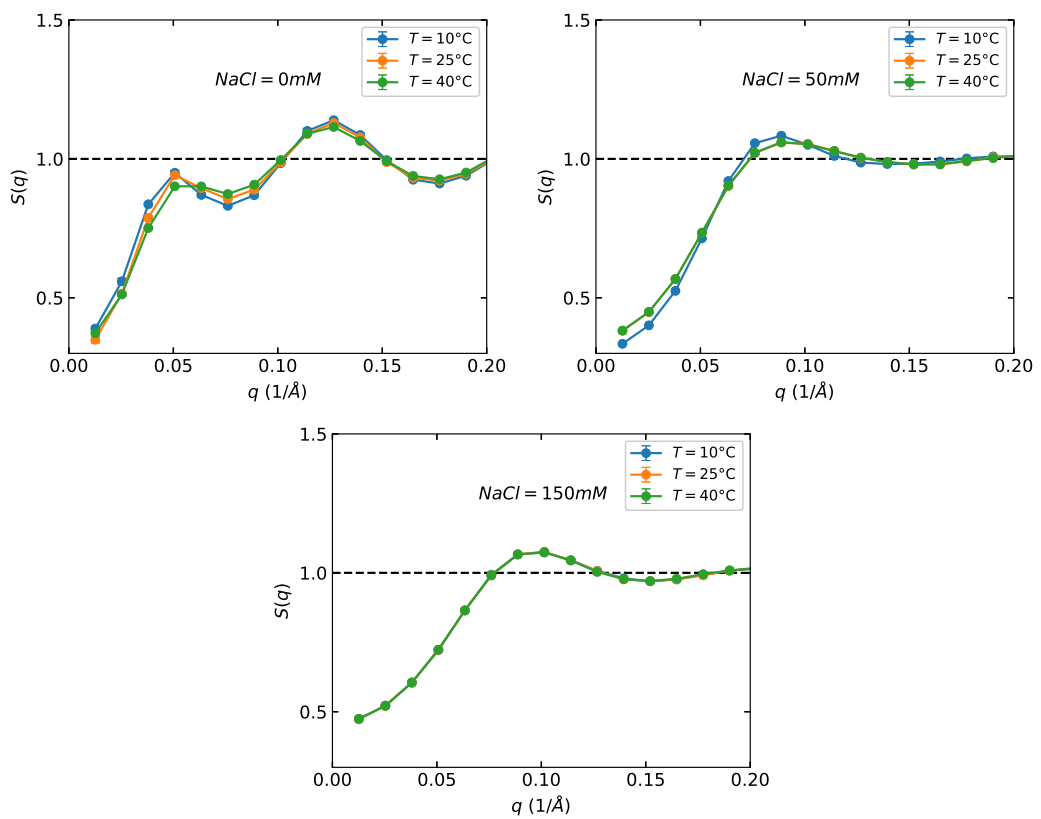


Figura 5.30: Fatores de estrutura da BSA, obtidas dos ajustes simulados com MC.

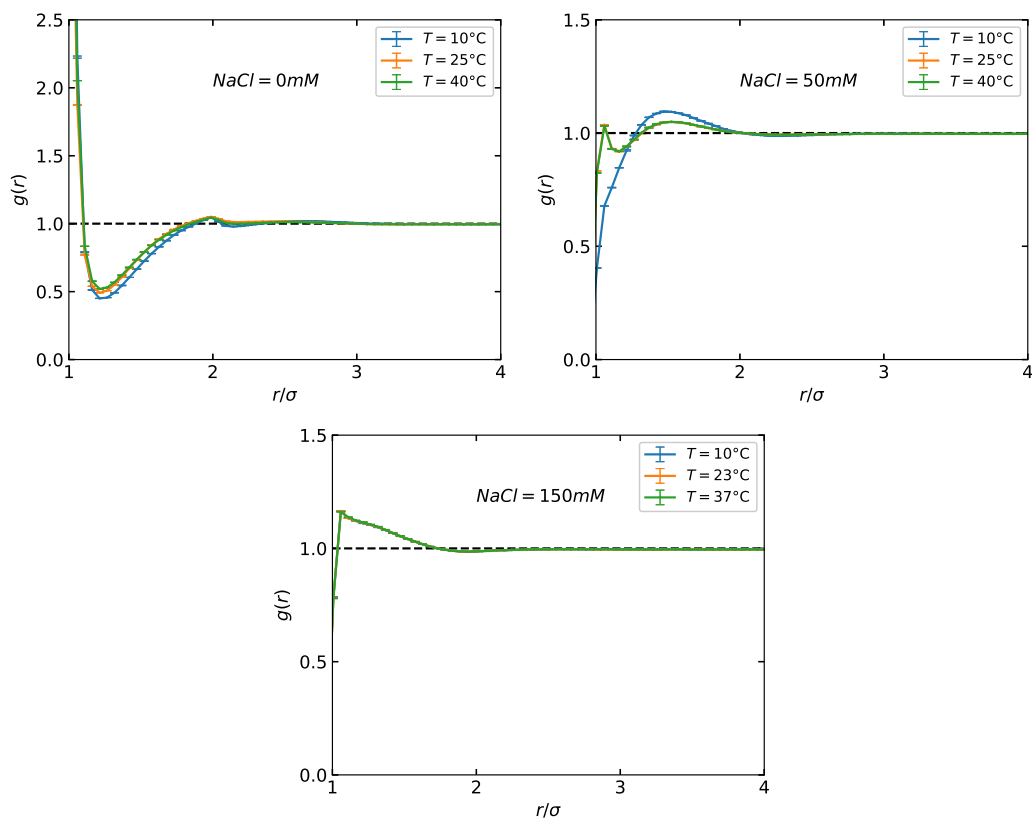


Figura 5.31: Distribuições radiais da BSA, obtidas dos ajustes simulados com MC.

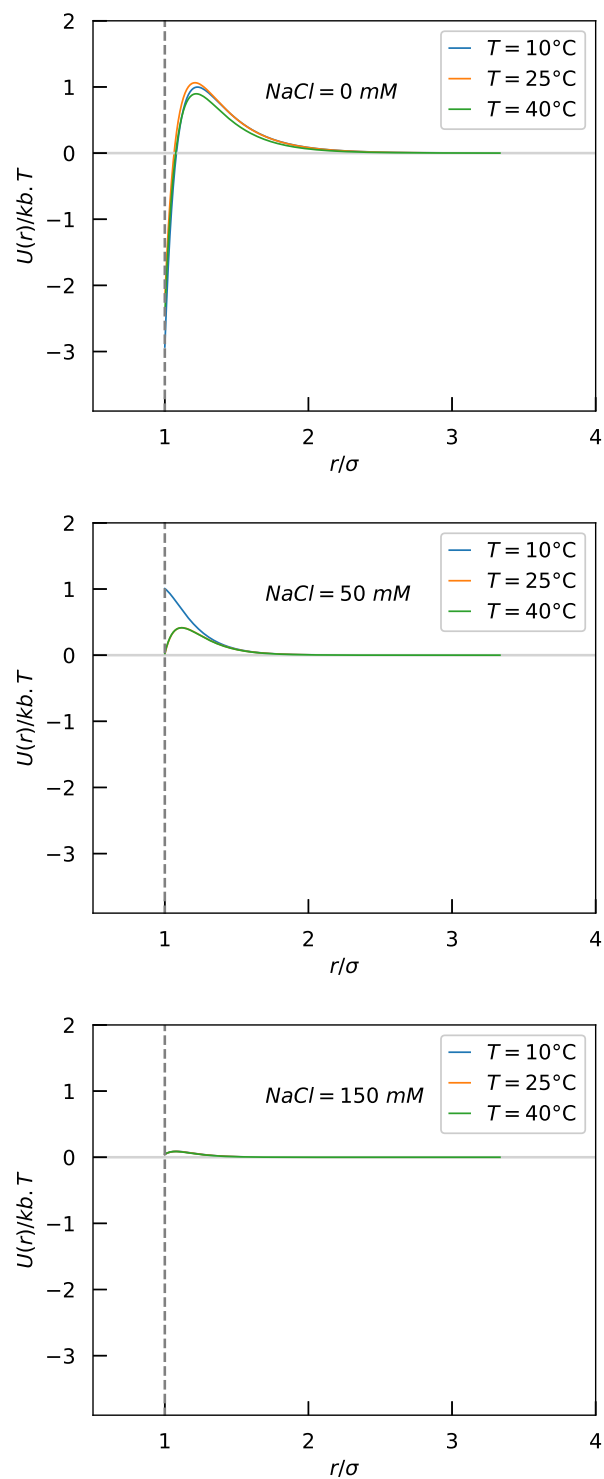


Figura 5.32: Potenciais médios de interação BSA-BSA obtidos dos ajustes simulados com MC.

Análise das Misturas Lisozima-BSA: Resultados e Discussão

1 Metodologia

1.1 Intensidade de espalhamento de misturas

No capítulo 3 derivamos uma expressão para a intensidade de espalhamento de um sistema monodisperso. No caso mais geral de sistemas (homogêneos e isotrópicos) polidispersos, contendo p espécies distintas de partículas, podemos escrever a intensidade¹ na forma de Fournet-Vrij[79]:

$$I(q) = \sum_{i=1}^p n_i \left[\langle |F_i(\mathbf{q})|^2 \rangle - |\langle F_i(\mathbf{q}) \rangle|^2 \right] + \sum_{i,j=1}^p (n_i n_j)^{1/2} \langle F_i(\mathbf{q}) \rangle \langle F_j^*(\mathbf{q}) \rangle S_{ij}(q) \quad (6.1)$$

onde n_i é a densidade numérica da i -ésima espécie, F_i é o seu fator de amplitude não-normalizado (cf. equação 2.31) e os brackets denotam a média orientacional sobre as coordenadas angulares da partícula. As funções $S_{ij}(q)$ são os fatores de estrutura parciais de Ashcroft-Langreth[80], definidos da seguinte maneira:

$$S_{ij}(q) = \delta_{ij} + (n_i n_j)^{1/2} h_{ij}(q) \quad (6.2)$$

sendo δ_{ij} o delta de Kronecker e $h_{ij}(q)$ a transformada de Fourier da função de correlação indireta $h_{ij}(r)$. Outra maneira de escrevermos a equação acima, muito mais útil do ponto de vista de simulações, é por meio de uma média nas posições $\{\mathbf{r}_k\}_{k=1,2,\dots}$ das partículas. Se o número $N_{1,2,\dots,p}$ de partículas de cada espécie for fixo, como no ensemble canônico, essa média pode ser representada como:

$$S_{ij}(q \neq 0) = \frac{1}{(N_i N_j)^{1/2}} \left\langle \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} e^{i\mathbf{q}(\mathbf{r}_k - \mathbf{r}_l)} \right\rangle \quad (6.3)$$

¹Seção-de-choque de espalhamento macroscópico diferencial

Para sistemas suficientemente diluídos ou ângulos de espalhamento suficientemente grandes, temos que $S_{ij} \approx \delta_{ij}$, de modo que a intensidade fica dependendo apenas das densidades e dos fatores de forma das partículas:

$$I(q) \approx \sum_{i=1}^p n_i \langle |F_i(\mathbf{q})|^2 \rangle \quad (6.4)$$

Em analogia com um sistema monodisperso diluído, essa última expressão é muitas vezes tratada como sendo um fator de forma efetivo, $P_{ef}(q)$, da mistura. Já no caso em que as interferências entre as partículas não podem ser desprezadas, define-se um fator de estrutura efetivo $S_{ef}(q)$, de tal modo que 6.1 pode ser escrita de maneira aparentemente simples como:

$$I(q) = P_{ef}(q)S_{ef}(q) \quad (6.5)$$

Trabalho precedente

Em sua tese de doutorado, defendida no IFUSP em 2018, Elisa Morandé[81] propôs um método de ajuste para a função $S_{ef}(q)$ de um sistema binário composto por lisozima e BSA em diversas proporções. Em sua abordagem, ela define:

$$S_{ef}(q) = 1 + \beta(q) [S(q) - 1] \quad (6.6)$$

onde $\beta(q)$ é uma função que envolve os fatores de forma e de amplitude de um cilindro hipotético e $S(q)$ é o fator de estrutura, sob a aproximação RPA, de um sistema de esferas duras monodisperso interagindo por um potencial HSDY. Assim, em termos de ajuste da S_{ef} , a mistura é tratada como se fosse um sistema simples. Como resultado foi possível ajustar as curvas de espalhamento e obter gráficos para os fatores de estrutura efetivos, além dos potenciais de interação desse suposto sistema simples equivalente. No entanto, essa abordagem (i) não é capaz de diferenciar as contribuições de cada proteína para a intensidade total, (ii) não fornece informações sobre os potenciais de pares reais e (iii) produz parâmetros de ajuste desprovidos de uma interpretação física imediata.

Essas limitações encontradas no trabalho da Dr.a Morandé nos motivou a desenvolver uma nova abordagem de ajuste para o caso das misturas binárias a partir de simulações de MC. O método que empregamos é descrito a seguir.

1.2 Monte Carlo das misturas

Para as nossas simulações consideramos um modelo de interação que consiste de uma adaptação do potencial HSDY para o caso de misturas:

$$u_{ij}(r) = \begin{cases} \infty, & \text{se } r < \sigma_{ij} \\ \frac{q_i q_j}{4\pi\epsilon(1+k\sigma_i/2)(1+k\sigma_j/2)} \frac{e^{-k(r-\sigma_{ij})}}{r} - \delta_{ij} J_i \sigma_i \frac{e^{-(r-\sigma_i)/D_i}}{r}, & \text{se } r \geq \sigma_{ij} \end{cases} \quad (6.7)$$

onde $u_{ij}(r)$ representa o potencial de pares entre as partículas do tipo i e j ($=$ L ou B), $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, δ_{ij} representa o delta de Kronecker e os outros parâmetros possuem o significado usual. Note que quando $i = j$ recaímos no potencial HSDY para sistemas monodispersos. Já o termo cruzado, u_{LB} , inclui apenas a contribuição de Coulomb bindado para partículas de tamanhos distintos. Não incluímos uma contribuição atrativa a mais ao potencial u_{LB} por dois motivos: (i) não temos como extrair empiricamente os parâmetros J e D da interação cruzada, diferentemente do que fizemos para as soluções simples de BSA e lisozima, e (ii) buscamos um modelo de ajuste com o menor número possível de parâmetros.

Os valores dos parâmetros q_i , σ_i , D_i e J_L foram fixados nos mesmos valores que encontramos nos ajustes das curvas experimentais das soluções simples (ver capítulo anterior). Assim, não consideramos possíveis variações nesses parâmetros ocasionadas pelo fato das proteínas agora estarem em uma mistura. O único parâmetro que modulamos caso a caso foi o termo atrativo J_B da BSA. Já a força iônica das misturas foi calculada pela média das forças iônicas das soluções simples ponderada pela proporção de cada proteína:

$$I_{LB} = \frac{n_L I_L + n_B I_B}{n_L + n_B} \quad (6.8)$$

Consideramos também a variação da constante dielétrica, $\epsilon_r = \epsilon/\epsilon_o$, com a temperatura, utilizando os valores já descritos no capítulo anterior. Desse modo, o comprimento de Debye $k_D = k^{-1}$ foi calculado caso a caso segundo a fórmula (no S.I.):

$$k^{-1} = \sqrt{\frac{\epsilon_r}{8\pi\lambda_B I_{LB}}} \quad (6.9)$$

sendo λ_B o comprimento de Bjerrum.

No total foram simuladas 59 condições físico-químicas nas quais variamos a temperatura em 10, 25 e 40 °C, quando as concentrações de sal eram iguais a 0 mM e 50 mM, e em 10, 23 e 37 °C quando tínhamos 150 mM de sal. As proporções de Lisozima:BSA utilizadas estão organizadas nas Tabelas 6.1 e 6.2. Mantivemos a concentração numérica total de proteínas sempre fixa em aproximadamente 1.36 mM e o número total de proteínas igual a $N = N_L + N_B = 100$.

Todas as simulações foram realizadas no ensemble canônico, em uma caixa cúbica com condições periódicas de contorno e utilizando o algoritmo de Metropolis. Foram utilizados $5 \cdot 10^4$ passos de MC para a termalização e 10^6 passos para a amostragem das

Tabela 6.1: Misturas de Lisozima e BSA. Proporções das amostras com 0 e 50 mM de sal monovalente

Proporção L:B	Porcentagem de L (%)	Porcentagem de B (%)	Concentração de L (mM)	Concentração de B (mM)
0:1	0	100	0	1.3636
3:17	15	85	0.2045	1.1591
3:7	30	70	0.4091	0.9545
1:1	50	50	0.6818	0.6818
7:3	70	30	0.9545	0.4091
17:3	85	15	1.1591	0.2045
1:0	100	0	1.3636	0

Tabela 6.2: Misturas de Lisozima e BSA. Proporções das amostras a 150 mM de sal monovalente.

Proporção L:B	Porcentagem de L (%)	Porcentagem de B (%)	Concentração de L (mM)	Concentração de B (mM)
0:1	0	100	0	1.3636
1:6	14.3	85.7	0.1948	1.1688
2:3	40	60	0.5454	0.8182
3:2	60	40	0.8182	0.5454
6:1	85.7	14.3	1.1688	0.1948
1:0	100	0	1.3636	0

configurações. Os fatores de estrutura foram obtidos por meio da fórmula 6.3, sendo o intervalo de decorrelação estimado em 15 passos de MC. Já os desvios padrão foram calculados através da fórmula usual:

$$\sigma^2(q) = \frac{1}{M-1} \sum_{i=1}^M [S_i(q) - \overline{S(q)}]^2 \quad (6.10)$$

sendo M o número de amostras usadas na média. No nosso caso, como as simulações foram realizadas em triplicata estamos tomando $M = 3$. Por fim, também calculamos as distribuições radiais das misturas como segue:

$$g_{ij}(r) = \frac{\langle n_j(r) \rangle}{V(r)\rho_j} \quad (6.11)$$

onde $\langle n_j(r) \rangle$ representa o número médio de partículas do tipo j dentro da casca esférica de volume $V(r) = (4\pi/3)[(r + dr)^3 - r^3]$, centrada em uma partícula de referência do tipo i , e $\rho_j = N_j/V$ é a densidade média total de j . As incertezas foram obtidas de maneira análoga ao fator de estrutura.

Aproximação para a intensidade teórica

Além dos fatores de estrutura e das distribuições radiais, gostaríamos de derivar uma curva "teórica" para a intensidade com o objetivo de realizar comparações diretas com as curvas experimentais. No caso de uma mistura homogênea e isotrópica composta por lisozima (L) e BSA (B) a equação 6.1 nos diz que:

$$I(q) = n_L \left[\langle |F_L|^2 \rangle - |\langle F_L \rangle|^2 \right] + n_B \left[\langle |F_B|^2 \rangle - |\langle F_B \rangle|^2 \right] \quad (6.12)$$

$$+ n_L |\langle F_L \rangle|^2 S_{LL} + n_B |\langle F_B \rangle|^2 S_{BB} + (n_L n_B)^{1/2} [\langle F_L \rangle \langle F_B^* \rangle + \langle F_L^* \rangle \langle F_B \rangle] S_{LB}$$

Assim, para podermos graficar uma intensidade de espalhamento teórica a partir dos resultados de simulação (S_{LL} , S_{LB} e S_{BB}) necessitamos conhecer também os fatores de amplitude e de forma das proteínas. Por um lado, o software GENFIT nos permite obter as funções $\langle |F_i|^2 \rangle$ e $|\langle F_i \rangle|$ a partir da estrutura cristalográfica das proteínas. É claro que as conformações destas podem se alterar ligeiramente quando em solução, porém iremos supor que essas funções não sofrem grandes desvios em relação às formas cristalográficas. Por outro lado, isso resolve parte do problema, pois ainda faltariam as funções $\langle F_L \rangle$ e $\langle F_B \rangle$, presentes no último termo da equação 6.16. Apesar dessas funções serem complexas a princípio, sabemos que para partículas centrossimétricas elas assumem valores reais, de modo que $\langle F_i \rangle = \langle F_i^* \rangle = |\langle F_i \rangle|$. Assim, considerando que nossas proteínas são partículas globulares, com simetria quase esféricas, parece razoável supor que:

$$\langle F_L \rangle \langle F_B^* \rangle + \langle F_L^* \rangle \langle F_B \rangle \approx 2 |\langle F_L \rangle| |\langle F_B \rangle| \quad (6.13)$$

Sendo válida essa suposição, temos toda informação necessária para calcular as curvas de intensidade teóricas e compará-las com as curvas experimentais.

Outra alternativa seria utilizar os fatores de amplitude das próprias esferas simuladas, cujas funções são conhecidas de antemão (cf. equação 3.53):

$$F_i = \rho_o V_i \Phi(q\sigma_i/2) \quad , \quad \Phi(x) = 3 \frac{\sin(x) - x \cos(x)}{x^3} \quad (6.14)$$

onde ρ_o é a densidade eletrônica da esfera (para proteínas globulares este valor é aproximadamente constante, em torno de $0.42 \text{ e}/\text{\AA}^3$) e V_i o volume da esfera de tipo i . Neste caso, a intensidade de espalhamento seria dada simplesmente por:

$$I(q) = n_L F_L^2 S_{LL} + n_B F_B^2 S_{BB} + 2(n_L n_B)^{1/2} F_L F_B S_{LB} \quad (6.15)$$

porém, preferimos seguir com a aproximação mais realista dada pela equação 6.13, fazendo a ressalva de que uma avaliação dessa hipótese ainda se faz necessária para julgarmos o quão boa ela é para os nossos ajustes.

2 Resultados e Discussões

2.1 SAXS das misturas

As medidas experimentais das misturas estão graficadas nas figuras 6.1 e 6.2, a seguir. Nas curvas a 150 mM de sal (fig. 6.2), as quais estão em escala absoluta, podemos observar a diminuição na intensidade de espalhamento conforme a proporção de lisozima aumenta nas soluções, o que já era esperado devido ao menor tamanho da proteína. As curvas a 0 e 50 mM de sal estão em escalas arbitrárias, o que dificulta identificar um efeito com a temperatura.

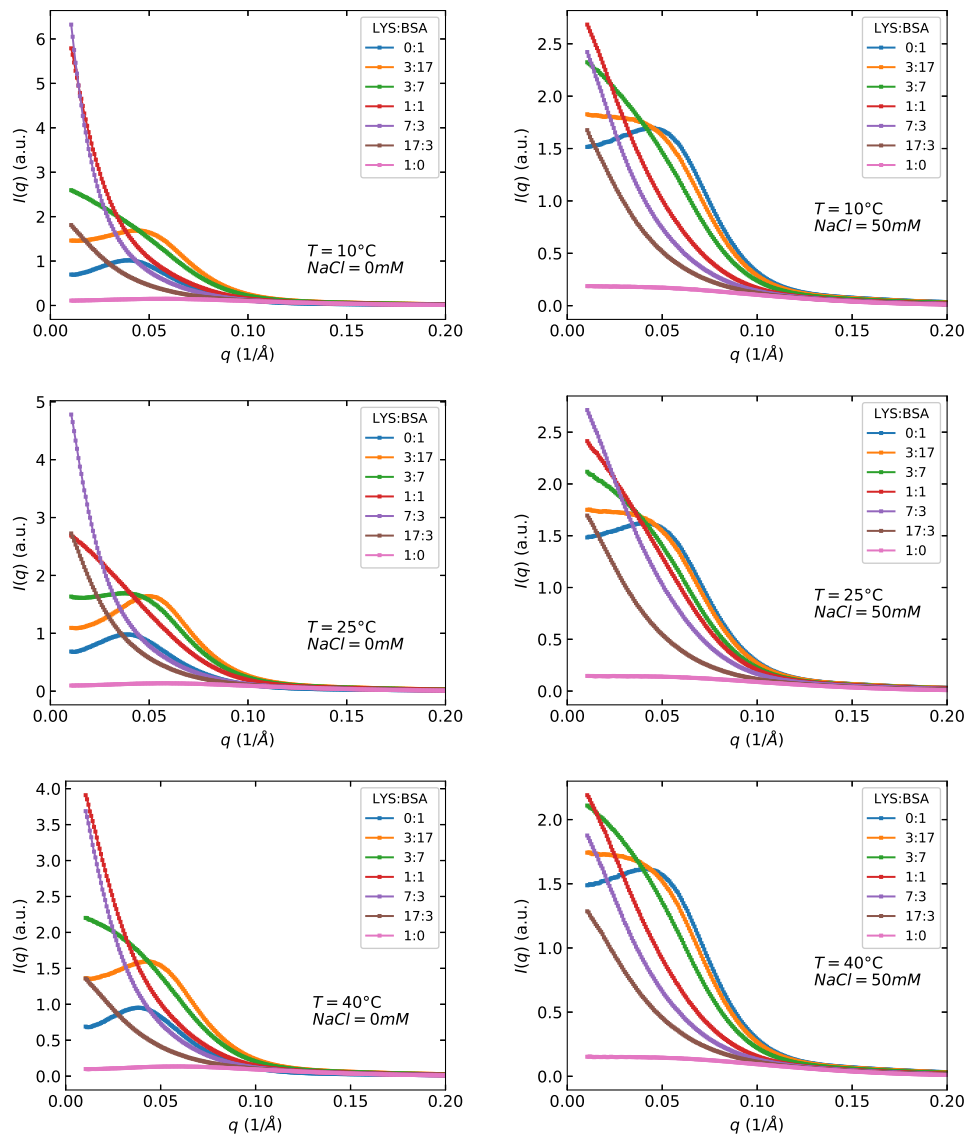


Figura 6.1: Intensidades de espalhamento das misturas de lisozima e BSA. Na primeira coluna, as medidas a 0 mM de sal e, na segunda, com 50 mM de sal. As temperaturas aumentam de cima para baixo.

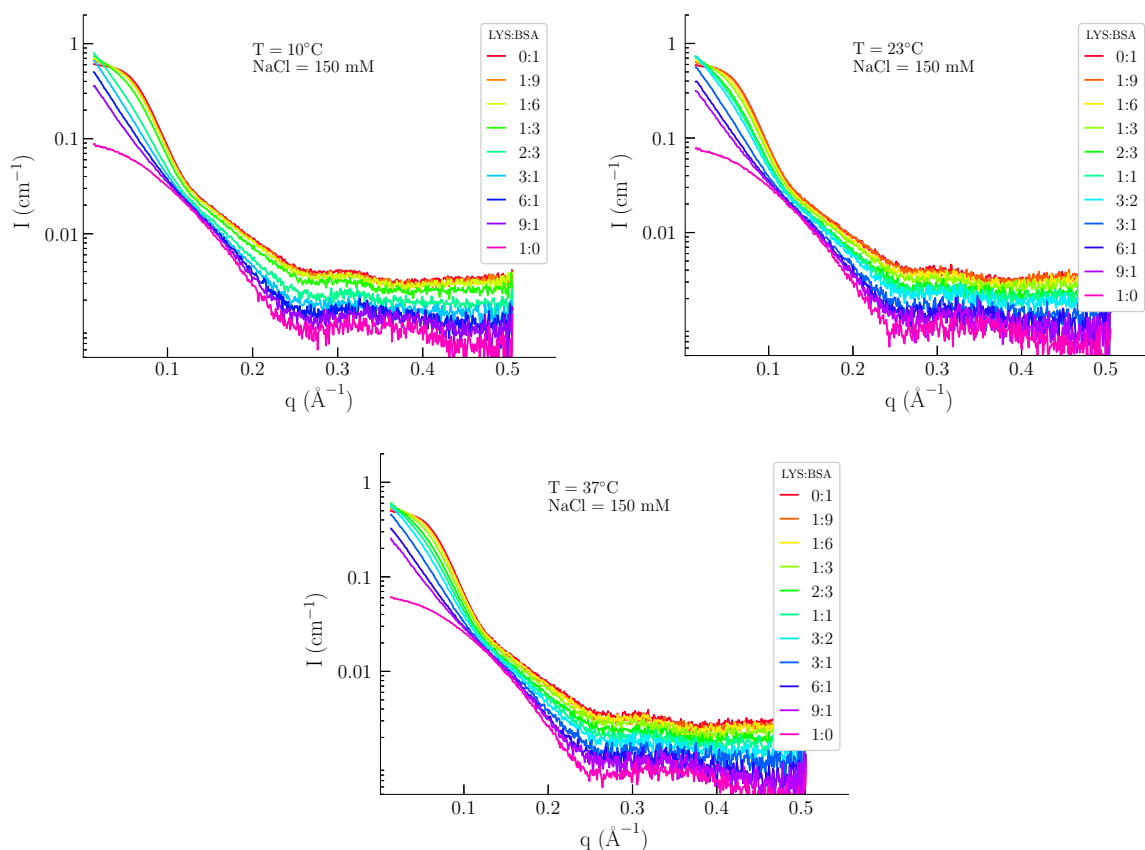


Figura 6.2: Intensidades de espalhamento das misturas de lisozima e BSA com 150 mM de sal.

Em todos os gráficos podemos ver que tanto as curvas de lisozima pura (1:0) quanto de BSA pura (0:1) possuem uma inclinação pequena nos menores ângulos e que ela vai aumentando à medida em que a proporção das duas proteínas vai ficando mais balanceada. Nas condições com 0 e 50 mM de sal, ora é a proporção 1:1 que possui maior inclinação ora a 7:3, o que poderia indicar uma maior aproximação das proteínas nesses casos. Vejamos que outras informações podemos extrair dos ajustes dessas curvas.

2.2 Análise de Monte Carlo das misturas

Nas figuras 6.3, 6.4 e 6.5, a seguir, graficamos as curvas experimentais novamente, mas agora junto com as funções de intensidade simuladas e ajustadas. As curvas estão deslocadas na vertical apenas para facilitar a sua visualização. Os ajustes das proporções 1:0 e 0:1 são os mesmos apresentados no capítulo anterior, pois, como dissemos, usamos os mesmos parâmetros derivados dos ajustes das soluções simples. A novidade se encontra nos ajustes das curvas nas proporções intermediárias. Como podemos ver, a metodologia proposta aqui foi capaz de produzir funções teóricas que se adequam muito bem aos dados medidos experimentalmente. Isso foi possível apenas permitindo com que o parâmetro atrativo J_B pudesse variar um pouco em cada caso. Os valores obtidos para esse termo

e os demais parâmetros se encontram sistematizados nas tabelas 6.3 e 6.4.

Os parâmetros que não estão listados nas tabelas acima foram aqueles mantidos constantes para todas as amostras, ou seja, a carga das proteínas ($Z_L = +7$ e $Z_B = -15$), seus diâmetros ($\sigma_L = 28 \text{ \AA}$ e $\sigma_B = 60 \text{ \AA}$) e as larguras dos potenciais atrativos ($D_L = 3 \text{ \AA}$ e $D_B = 6 \text{ \AA}$).

Para a maioria das curvas com 0 mM de sal foi possível realizar os ajustes variando J_B em menos de $1 k_bT$ em relação ao valor que obtivemos para BSA pura (6.8 e $6.2 k_bT$). A condição que precisou de uma maior variação foi na proporção 3:17, na qual tivemos que reduzir um pouco mais o valor de J_B para reproduzir o formato dos dados experimentais. Da mesma forma, as simulações com 50 e 150 mM de sal também exigiram oscilações maiores do parâmetro, no caso não maiores do que $5.5 k_bT$ e $4.3 k_bT$, respectivamente. Para essas condições, tivemos sempre que aumentar o valor de J_B conforme a proporção de lisozima crescia na mistura, o que indica que essa proteína induziu uma maior atração entre as BSA em função de sua densidade numérica.

No entanto, para entendermos como as interações e a estrutura líquida das proteínas está se comportando em cada condição precisamos olhar mais atentamente para os gráficos das energias potenciais (figuras de 6.6 até 6.14, fatores de estrutura (figuras de 6.15 até 6.23 e distribuições radiais (figuras de 6.24 até 6.32).

Em relação à interação lisozima-lisozima u_{LL} nas misturas, destacamos que ela é a única função potencial que sofre uma alteração significativa com a mudança de temperatura, e apenas nas concentrações de NaCl iguais a 0 e 50 mM. Basicamente ela se torna mais repulsiva em temperaturas maiores, repetindo o comportamento que havíamos observado para as soluções simples. A mudança se dá principalmente no início das curvas, próximo à superfície da proteína, onde o mínimo local em $r = \sigma_L$ passa de valores negativos (em $T = 10 \text{ }^\circ\text{C}$, figs. 6.6 e 6.9) para valores positivos ($T = 25 \text{ }^\circ\text{C}$, figs. 6.7 e 6.10), até se tornar um máximo local ($T = 40 \text{ }^\circ\text{C}$, figs. 6.8 e 6.11). Nas condições com 150 mM de sal não verificamos o mesmo efeito e, de fato, todos os potenciais se mostraram resilientes à temperatura quando a blindagem iônica se encontrou elevada.

Já quando olhamos as mudanças provocadas pela proporção das duas proteínas, vemos que é o potencial BSA-BSA, u_{BB} , o mais afetado. É claro que isso reflete o fato de termos variado apenas o parâmetro J_B em nossos ajustes, de modo que o cenário das interações que apresentamos aqui carrega a suposição de que os demais parâmetros atrativos (J_L , D_L e D_B) permaneceram inalterados. Sendo este o caso, o que observamos é que a adição de lisozima nas misturas diminui a repulsão entre as BSAs, quando olhamos as curvas a 50 e 150 mM de sal (figs. 6.9-6.11 e 6.12-6.14, respectivamente). Nas soluções sem NaCl (figs. 6.6-6.8), no entanto, a situação parece ser a inversa, com um aumento na repulsão das BSAs conforme a proporção de lisozima cresce. Em todos os

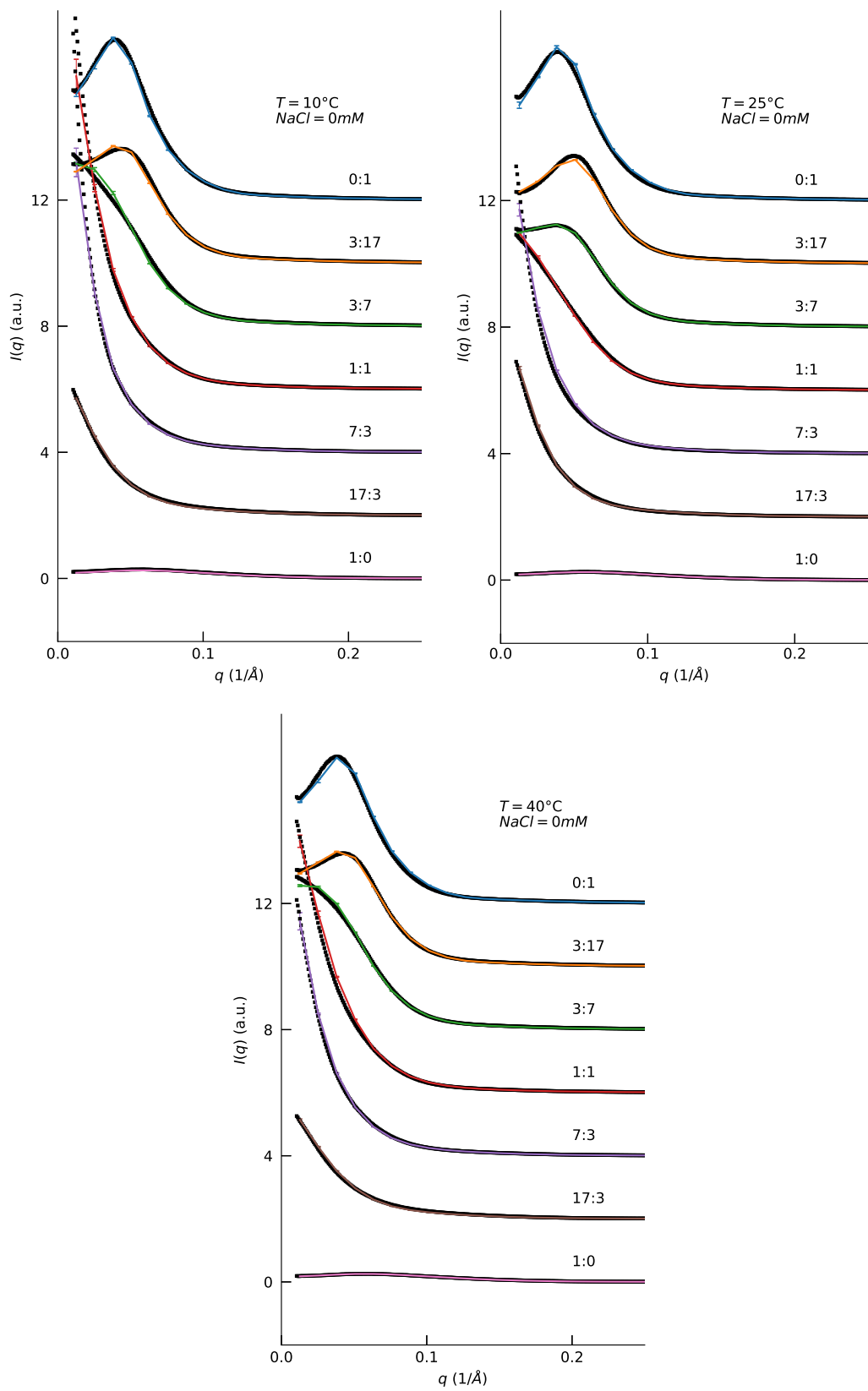


Figura 6.3: Ajustes com Monte Carlo das misturas a 0 mM de sal, em sete proporções distintas de LYS:BSA e temperaturas de 10, 25 e 40 °C.

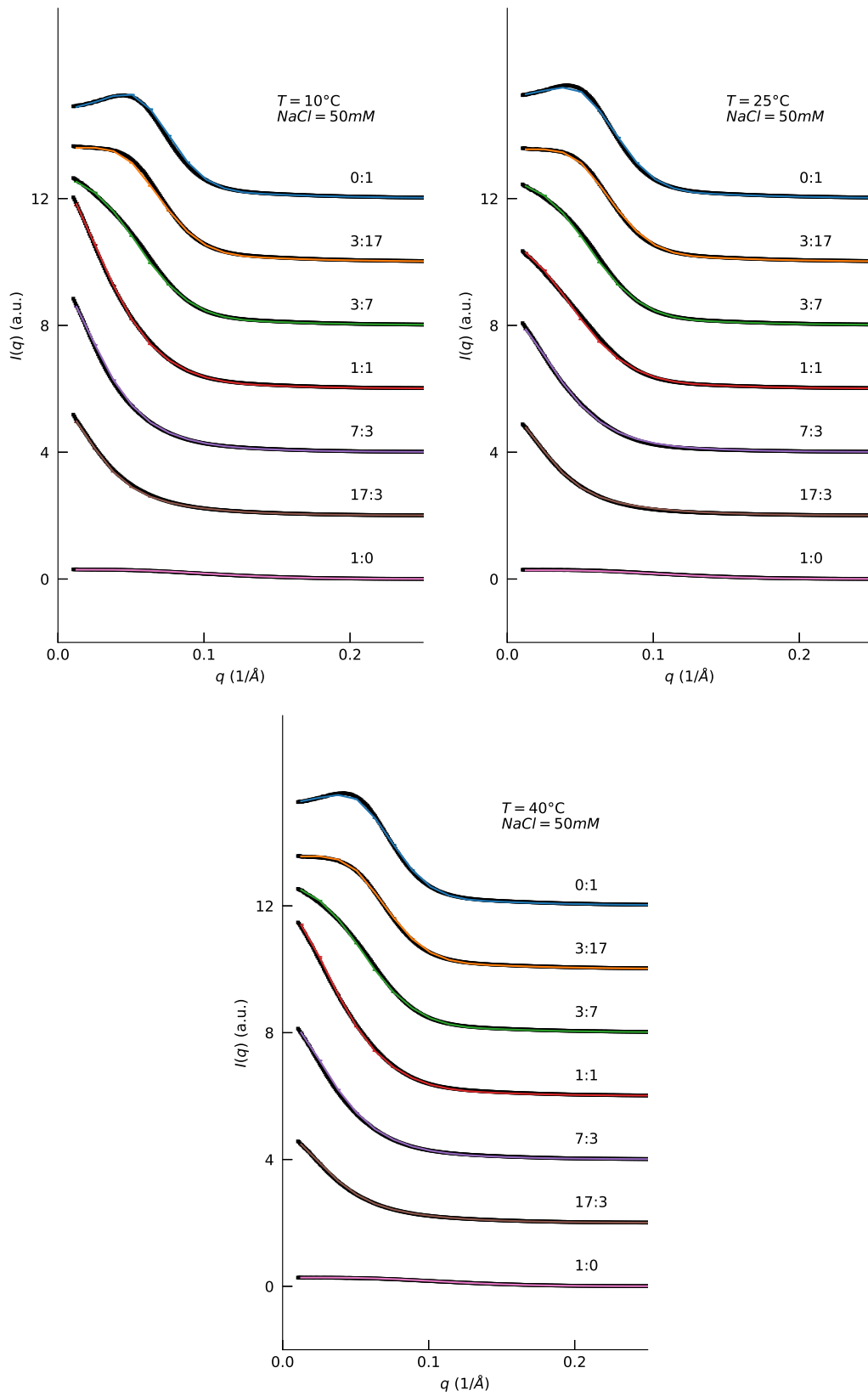


Figura 6.4: Ajustes com Monte Carlo das misturas a 50 mM de sal, em sete proporções distintas de LYS:BSA e temperaturas de 10, 25 e 40 °C.

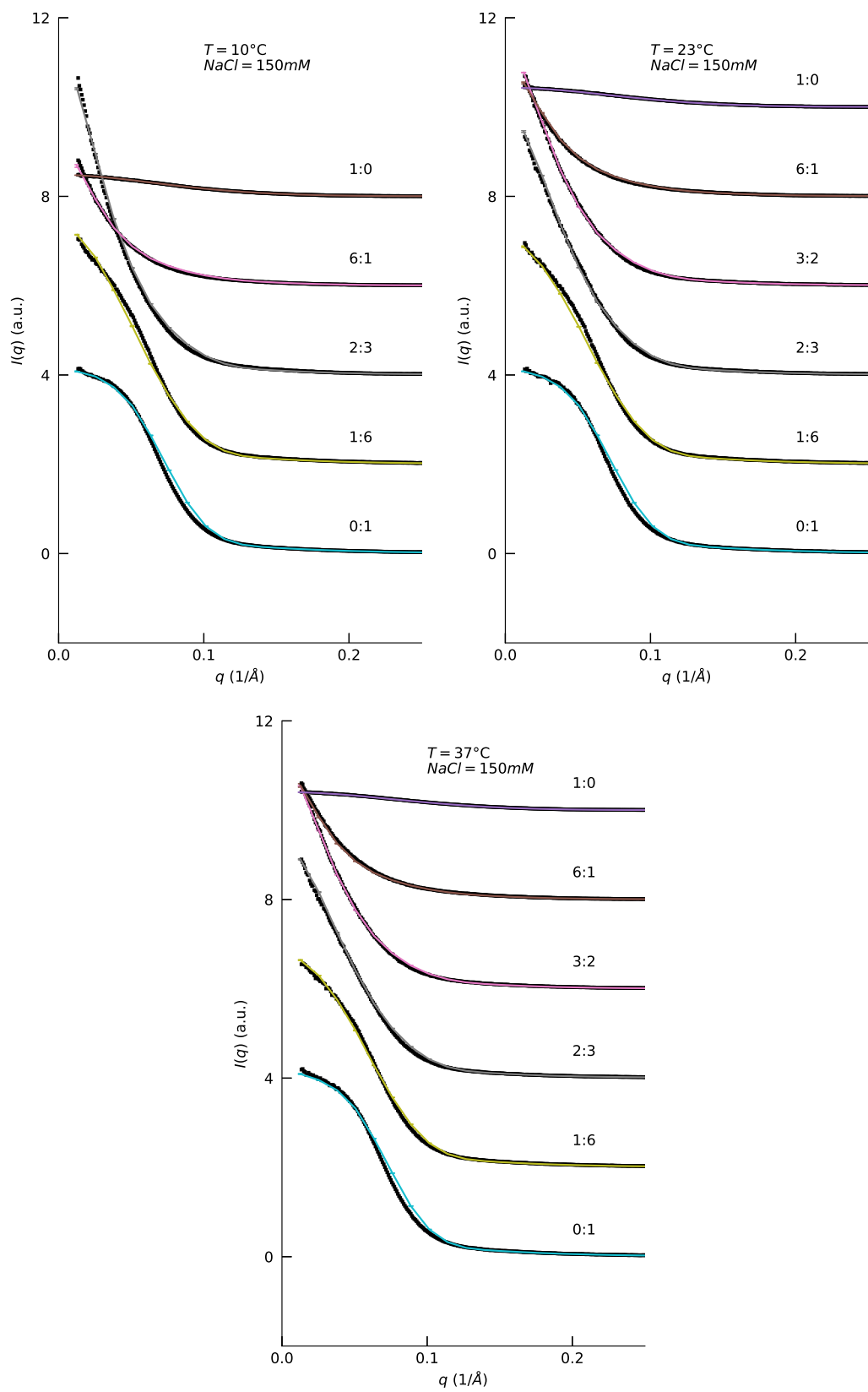


Figura 6.5: Ajustes com Monte Carlo das misturas a 0 mM de sal, em seis proporções distintas de LYS:BSA e temperaturas de 10, 23 e 37 °C.

Amostra	NaCl (mM)	T (°C)	LYS:BSA	J_L (k_bT)	J_B (k_bT)	I_{LB} (mM)
1	0	10	0:1	-	6.8	26
2	0	10	3:17	5	5.2	24.95
3	0	10	3:7	5	6.2	23.9
4	0	10	1:1	5	6.5	22.5
5	0	10	7:3	5	6.4	21.1
6	0	10	17:3	5	6.4	20.05
7	0	10	1:0	5	-	19
8	0	25	0:1	-	6.2	26
9	0	25	3:17	2	3.5	24.95
10	0	25	3:7	2	3.8	23.9
11	0	25	1:1	2	5.3	22.5
12	0	25	7:3	2	6.5	21.1
13	0	25	17:3	2	7.2	20.05
14	0	25	1:0	2	-	19
15	0	40	0:1	-	6.2	29
16	0	40	3:17	1	4.9	27.5
17	0	40	3:7	1	5.8	26
18	0	40	1:1	1	6.4	24
19	0	40	7:3	1	6.5	22
20	0	40	17:3	1	6.2	20.5
21	0	40	1:0	1	-	19
22	50	10	0:1	-	1	72
23	50	10	3:17	5	2.5	70.125
24	50	10	3:7	5	3.8	68.25
25	50	10	1:1	5	5.1	65.75
26	50	10	7:3	5	5.7	63.25
27	50	10	17:3	5	6.5	61.375
28	50	10	1:0	5	-	59.5
29	50	25	0:1	-	2	72
30	50	25	3:17	2	2.4	70.125
31	50	25	3:7	2	3.7	68.25
32	50	25	1:1	2	4.3	65.75
33	50	25	7:3	2	5.4	63.25
34	50	25	17:3	2	6.5	61.375
35	50	25	1:0	2	-	59.5
36	50	40	0:1	-	2	72
37	50	40	3:17	1	2.4	70.125
38	50	40	3:7	1	3.8	68.25
39	50	40	1:1	1	5	65.75
40	50	40	7:3	1	5.5	63.25
41	50	40	17:3	1	6.3	61.375
42	50	40	1:0	1	-	59.5

Tabela 6.3: Parâmetros dos ajustes das misturas a 0 e 50 mM de sal.

Amostra	NaCl (mM)	T (°C)	LYS:BSA	J_L (k_bT)	J_B (k_bT)	I_{LB} (mM)
43	150	10	0:1	-	1	170
44	150	10	1:6	5.6	2.3	168.5
45	150	10	2:3	5.6	3.65	165.8
46	150	10	6:1	5.6	5.3	161
47	150	10	1:0	5.6	-	159.5
48	150	23	0:1	-	1	170
49	150	23	1:6	5.3	2.1	168.5
50	150	23	2:3	5.3	3.3	165.8
51	150	23	3:2	5.3	4	163.7
52	150	23	6:1	5.3	5.3	161
53	150	23	1:0	5.3	-	159.5
54	150	37	0:1	-	1	170
55	150	37	1:6	5	1.9	168.5
56	150	37	2:3	5	3	165.8
57	150	37	3:2	5	3.9	163.7
58	150	37	6:1	5	5.3	161
59	150	37	1:0	5	-	159.5

Tabela 6.4: Parâmetros dos ajustes das misturas a 150 mM de sal.

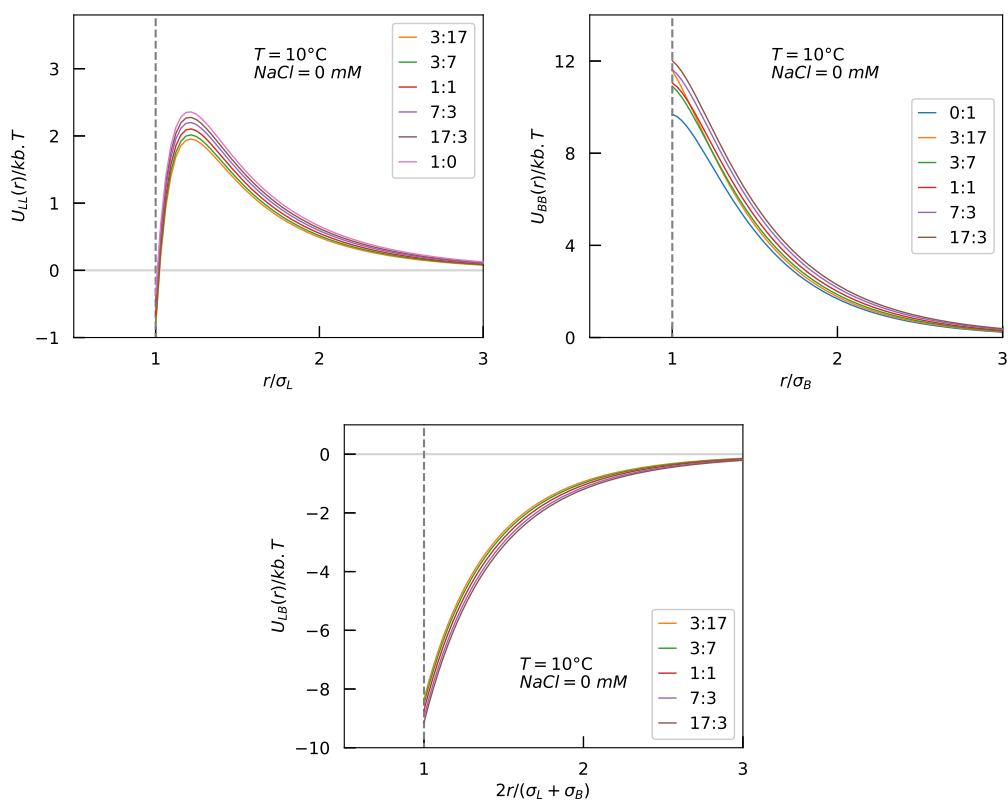


Figura 6.6: Energias potenciais das misturas a 0 mM de sal e temperatura de 10 °C.

casos, o aumento de BSA na mistura torna as interações lisozima-lisozima, u_{LL} , menos repulsivas. Estes pontos serão revistos na apresentação das funções de distribuição radial,

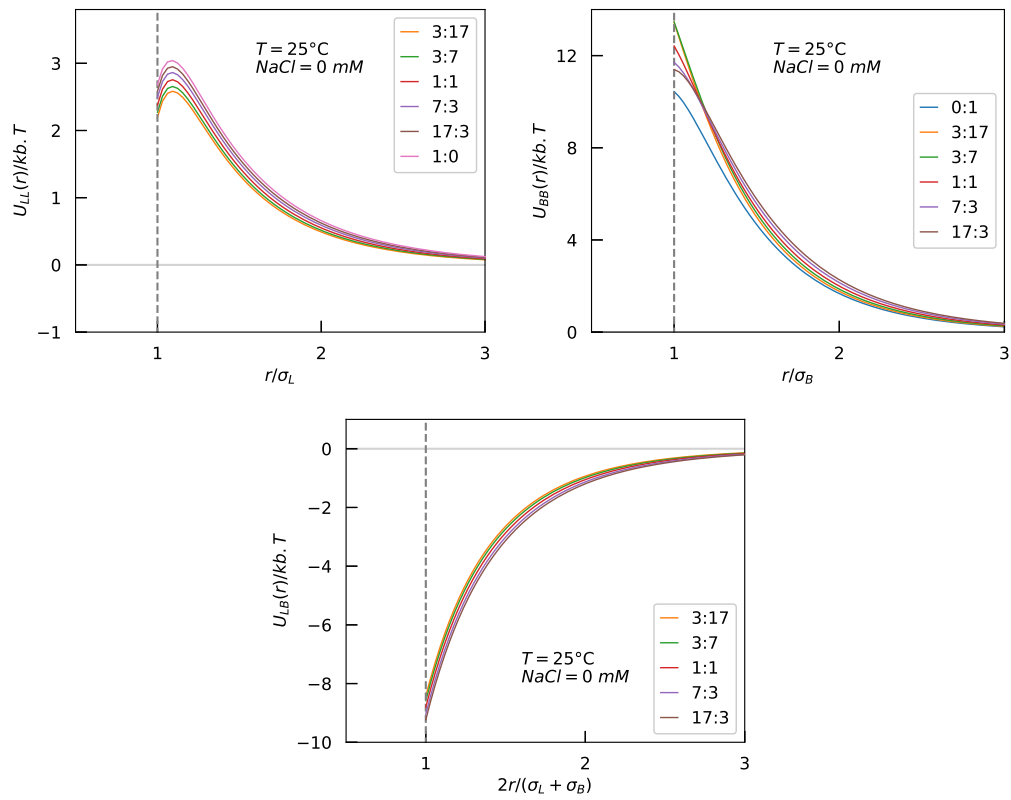


Figura 6.7: Energias potenciais das misturas a 0 mM de sal e temperatura de 25°C .

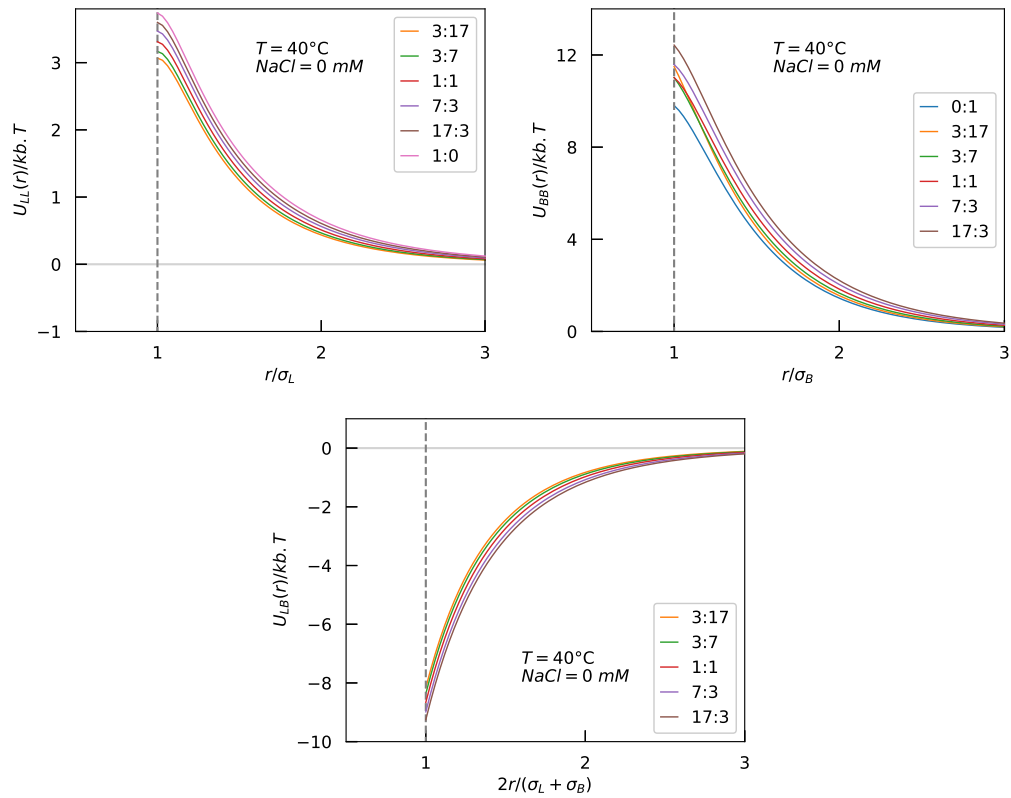


Figura 6.8: Energias potenciais das misturas a 0 mM de sal e temperatura de 40°C .

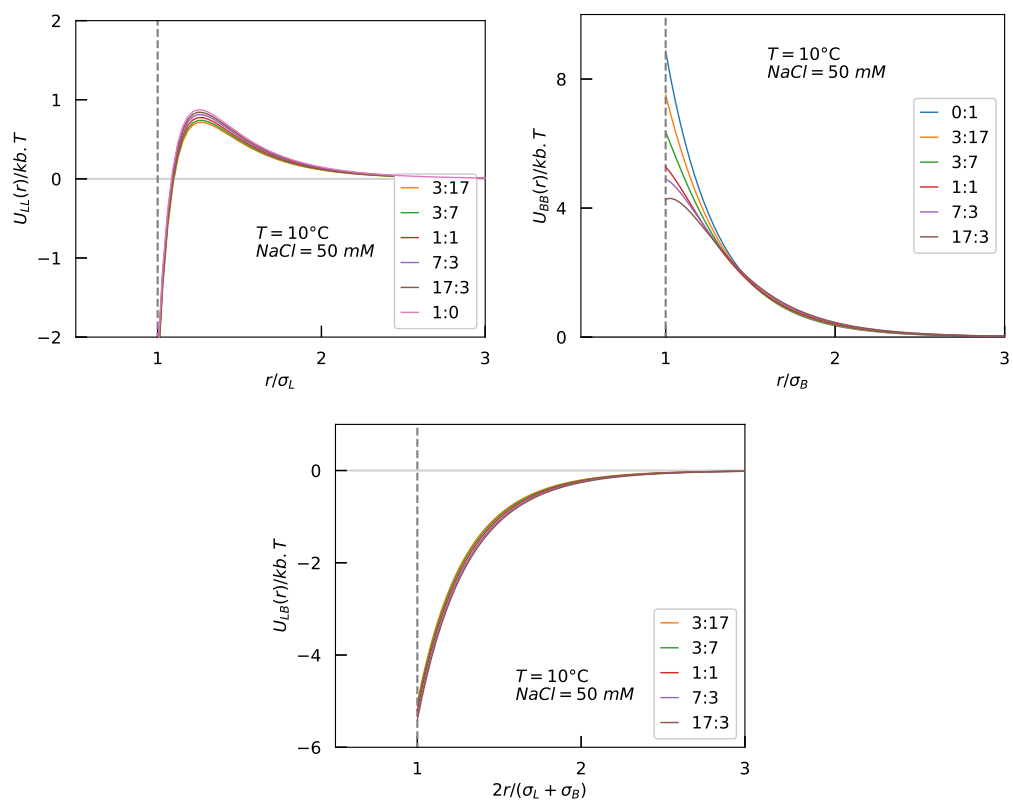


Figura 6.9: Energias potenciais das misturas a 50 mM de sal e temperatura de 10°C .

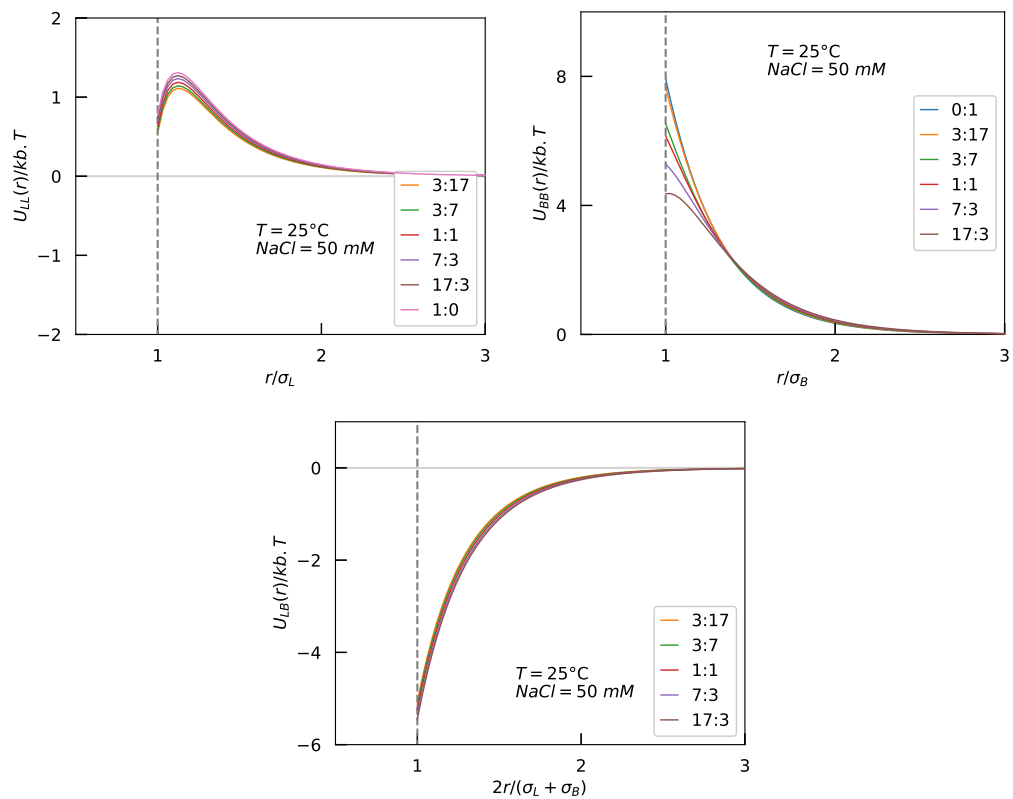


Figura 6.10: Energias potenciais das misturas a 50 mM de sal e temperatura de 25°C .

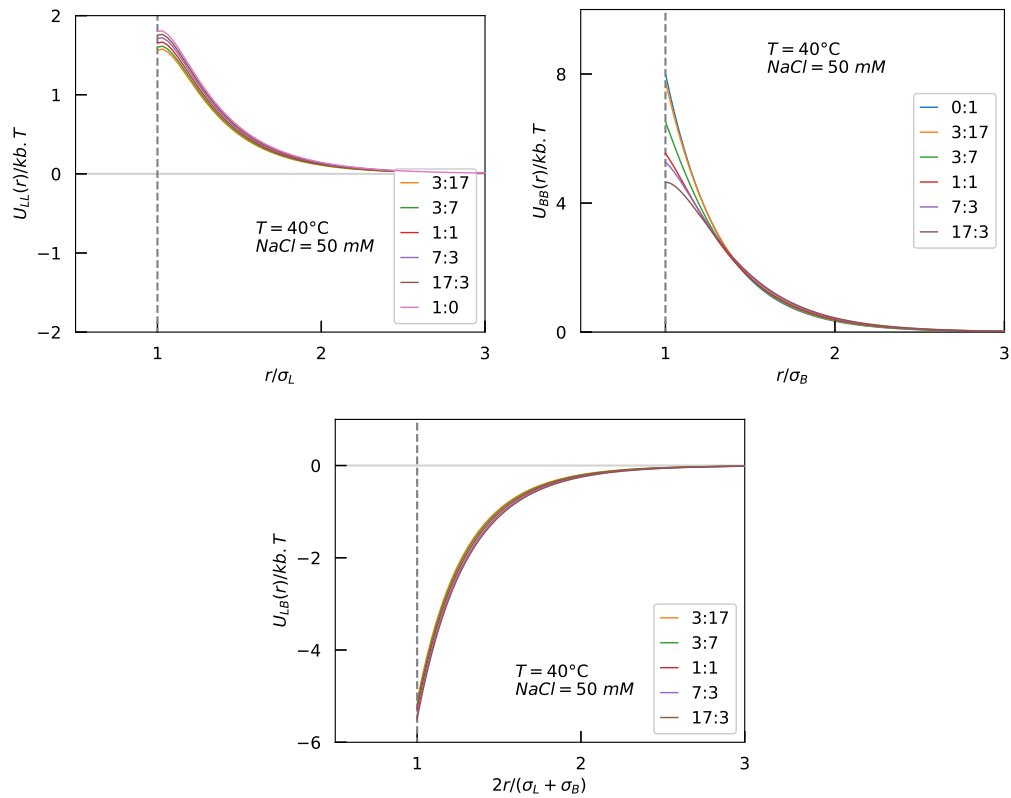


Figura 6.11: Energias potenciais das misturas a 50 mM de sal e temperatura de 40 °C.

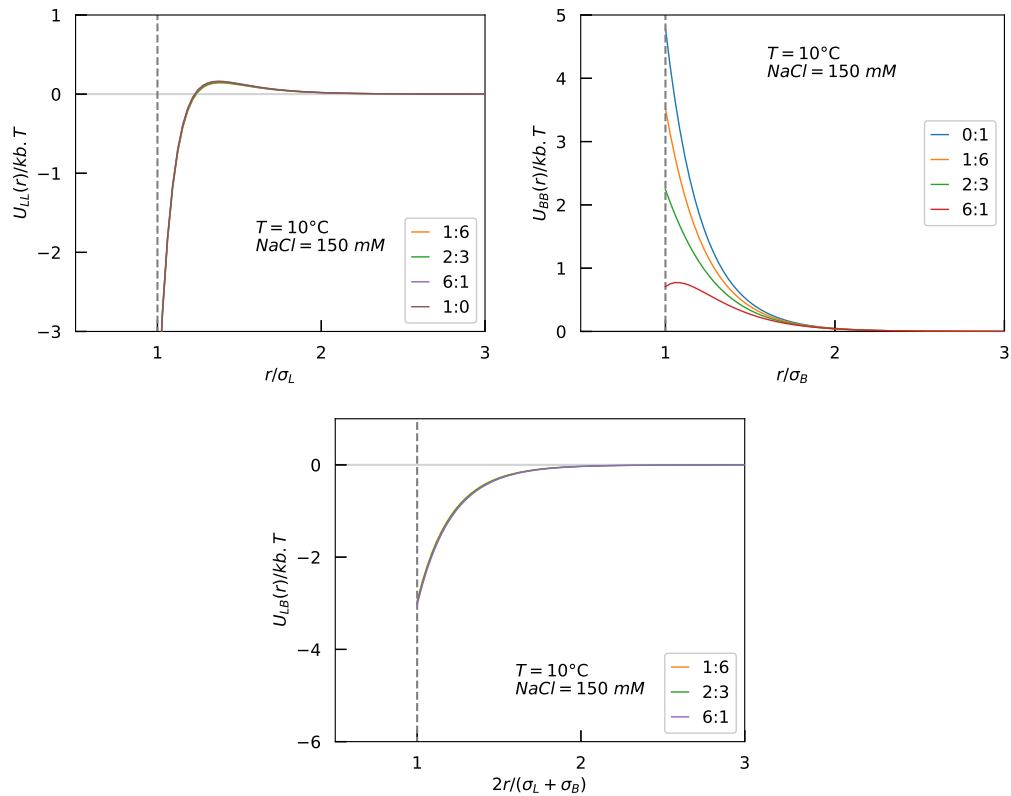


Figura 6.12: Energias potenciais das misturas a 150 mM de sal e temperatura de 10 °C.

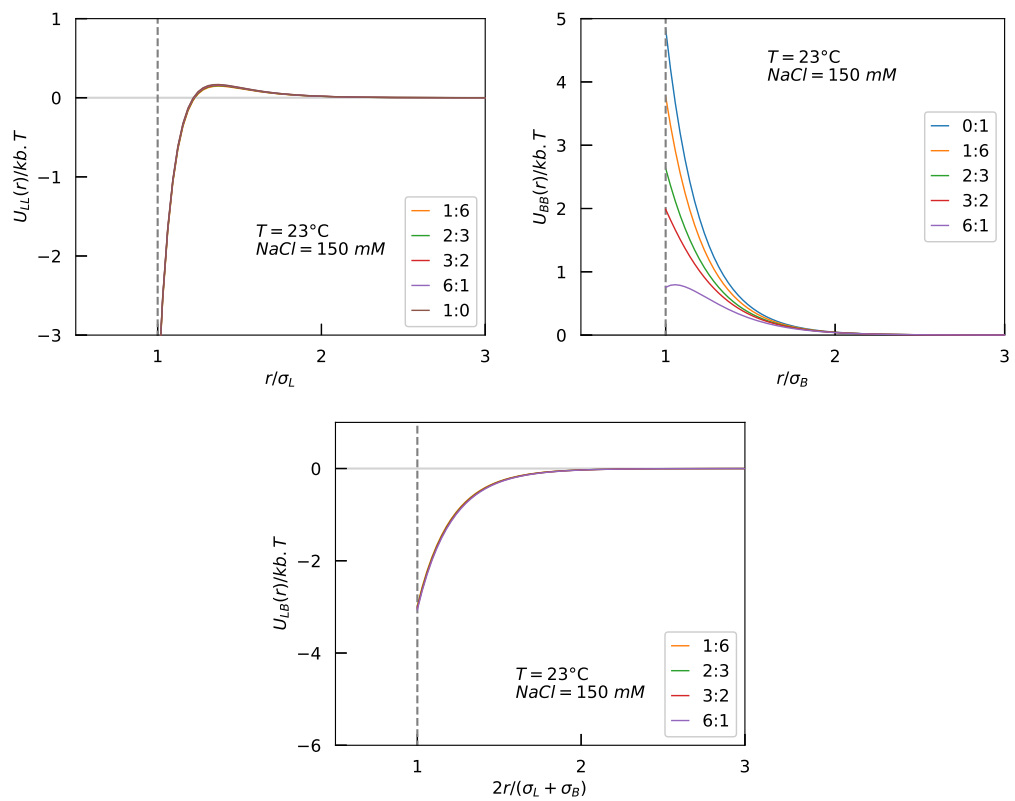


Figura 6.13: Energias potenciais das misturas a 150 mM de sal e temperatura de 25 °C.

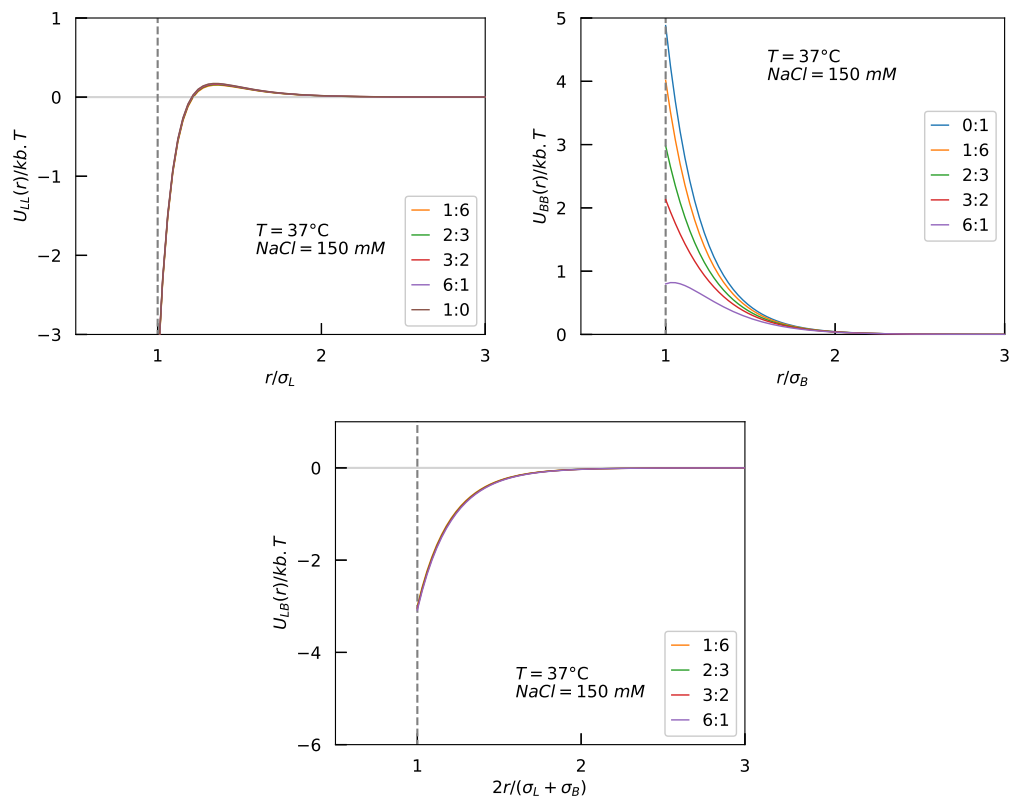


Figura 6.14: Energias potenciais das misturas a 150 mM de sal e temperatura de 40 °C.

mais adiante no texto.

Quanto às interações cruzadas, elas são muito pouco afetadas seja pela temperatura, seja pela proporção das duas proteínas. Essa característica também deriva do nosso modelo de ajuste, uma vez que u_{LB} só leva em conta o potencial de Coulomb blindado, o qual depende principalmente das cargas e dos diâmetros das partículas. Como este potencial também depende da força iônica vemos uma variação nas curvas no sentido de tornar menos atrativa a interação liozima-BSA conforme adicionamos sal na mistura. Essa tendência é fácil de ser compreendida, uma vez que quanto mais blindadas estão as proteínas menos elas irão interagir eletrostaticamente. Esse efeito de blindagem também pode ser verificado para os potenciais u_{LL} e u_{BB} . No caso das interações liozima-liozima, a adição de sal muda a tal ponto os potenciais que estes passam de puramente repulsivos para quase totalmente atrativos. Já as interações BSA-BSA, apesar de também sofrerem uma atenuação, continuam sempre repulsivas devido a sua maior carga elétrica total.

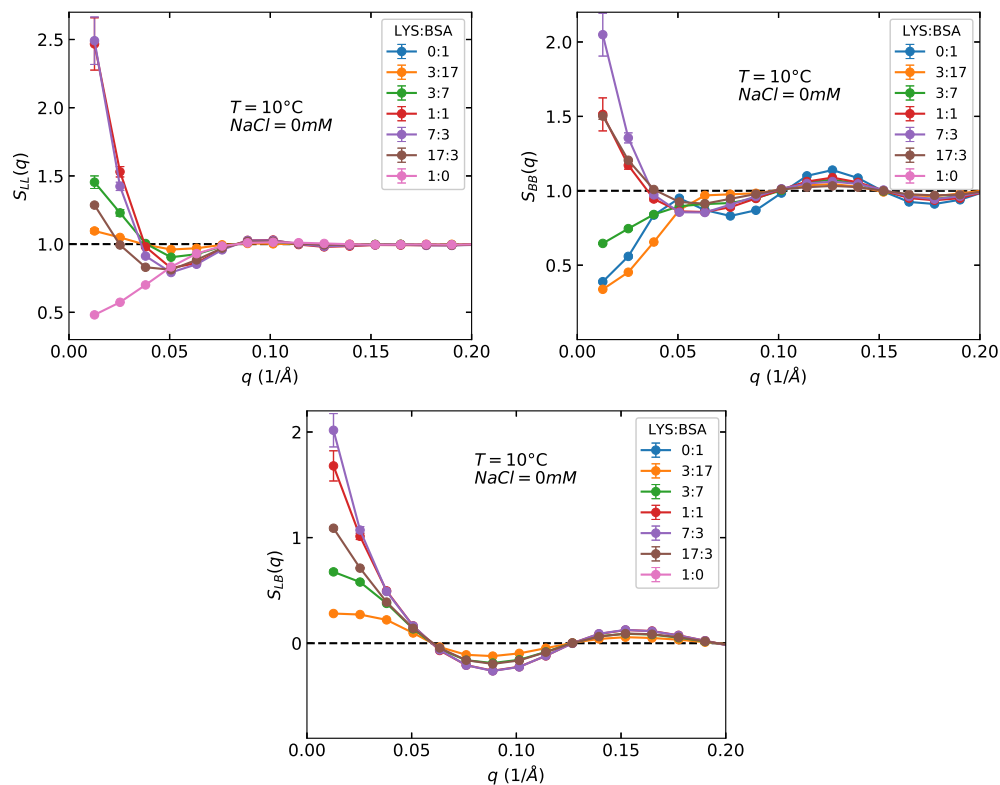


Figura 6.15: Fatores de estrutura das misturas a 0 mM de sal e temperatura de 10 °C.

Olhando agora para os fatores de estrutura, encontramos um comportamento bastante diverso em função da concentração de sal. A 0 mM de NaCl, por exemplo, as funções S_{LL} , S_{BB} e S_{LB} oscilam ora pra cima, ora para baixo a depender da proporção LYS:BSA. Partindo da solução simples de uma proteína (0:1 ou 1:0), vemos a princípio um aumento das funções, nos menores ângulos de espalhamento, à medida em que a proporção da

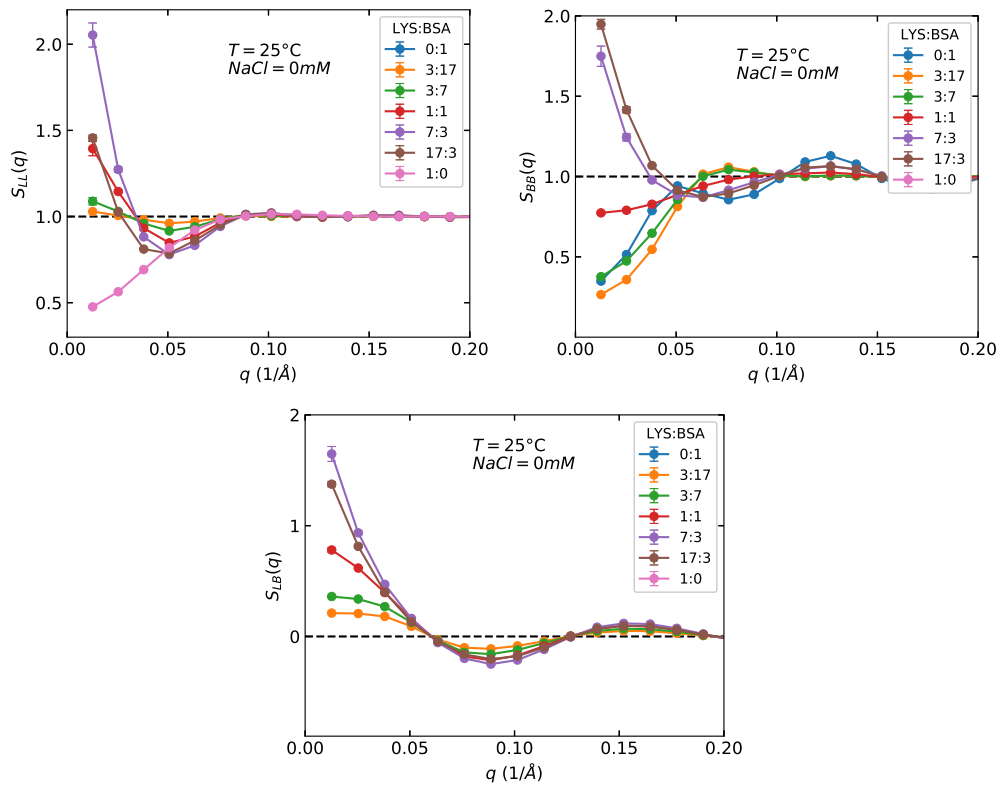


Figura 6.16: Fatores de estrutura das misturas a 0 mM de sal e temperatura de 25 °C.

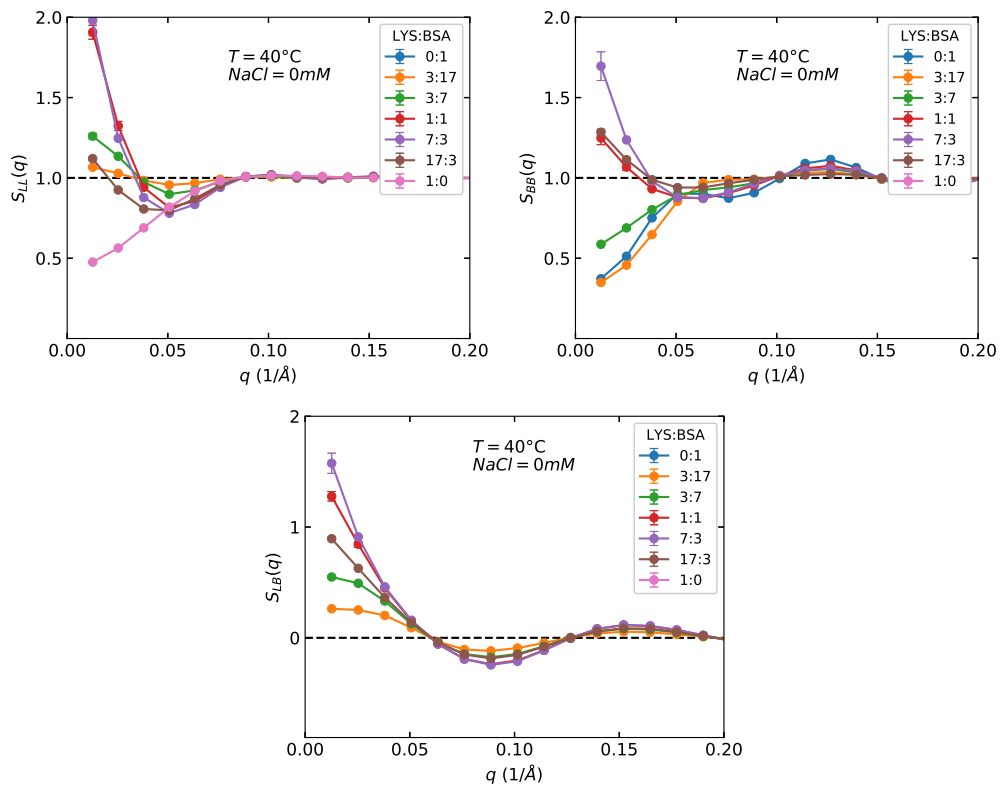


Figura 6.17: Fatores de estrutura das misturas a 0 mM de sal e temperatura de 40 °C.

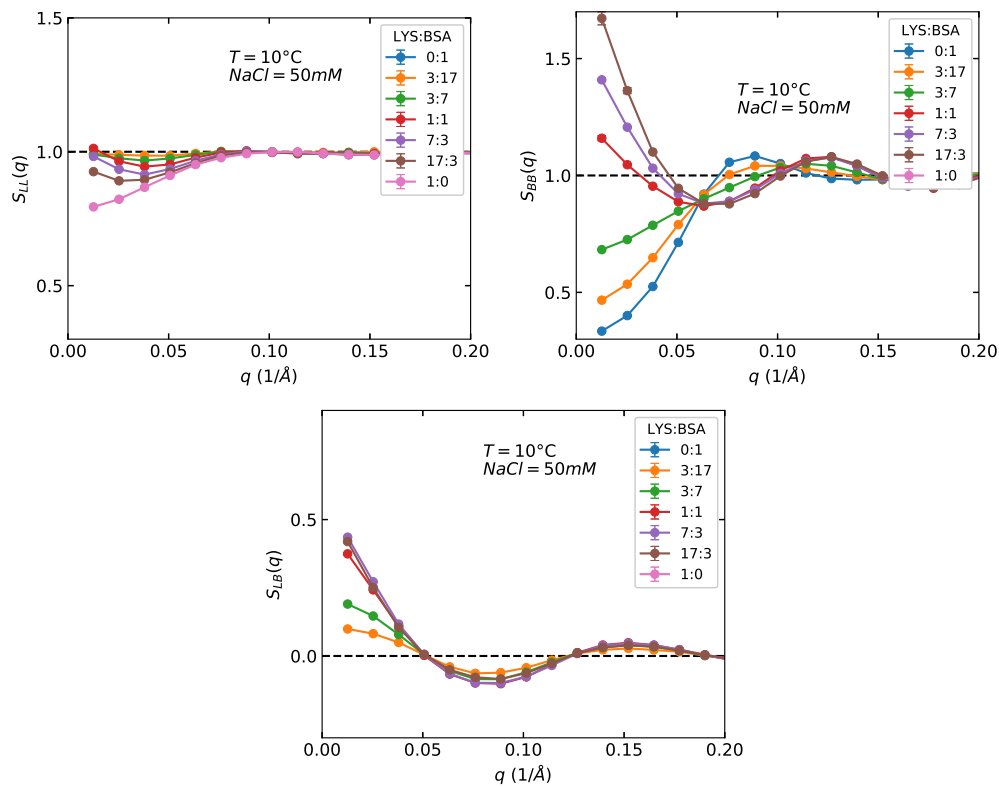


Figura 6.18: Fatores de estrutura das misturas a 50 mM de sal e temperatura de 10 °C.

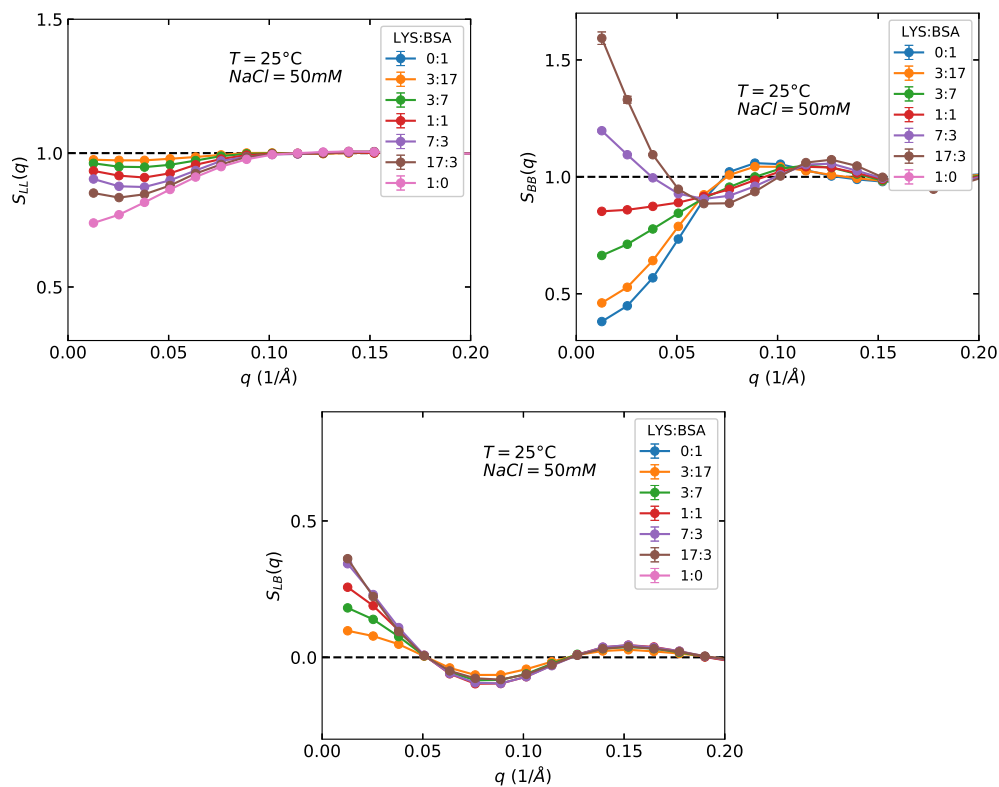


Figura 6.19: Fatores de estrutura das misturas a 50 mM de sal e temperatura de 25 °C.

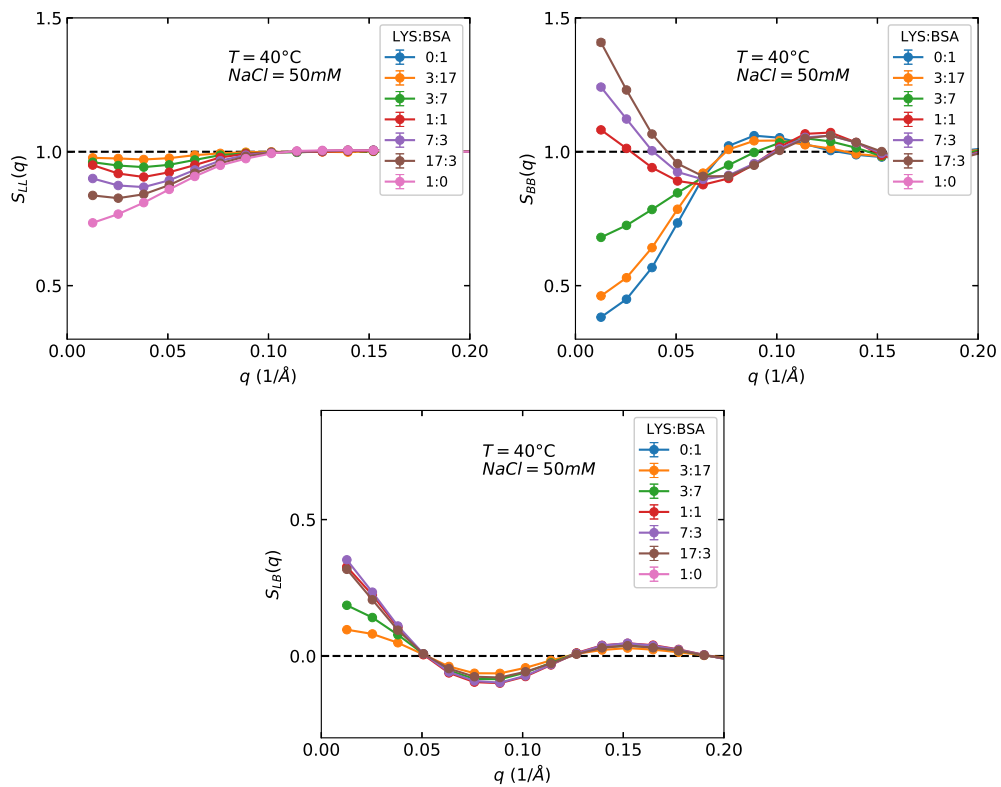


Figura 6.20: Fatores de estrutura das misturas a 50 mM de sal e temperatura de 40 °C.

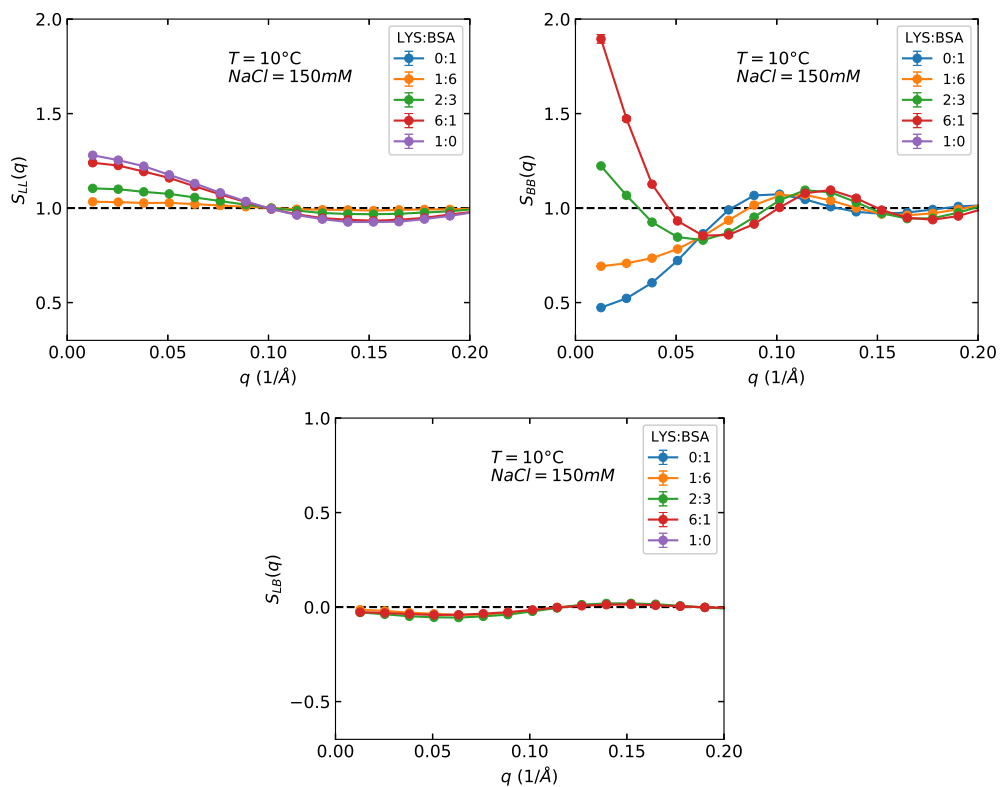


Figura 6.21: Fatores de estrutura das misturas a 150 mM de sal e temperatura de 10 °C.

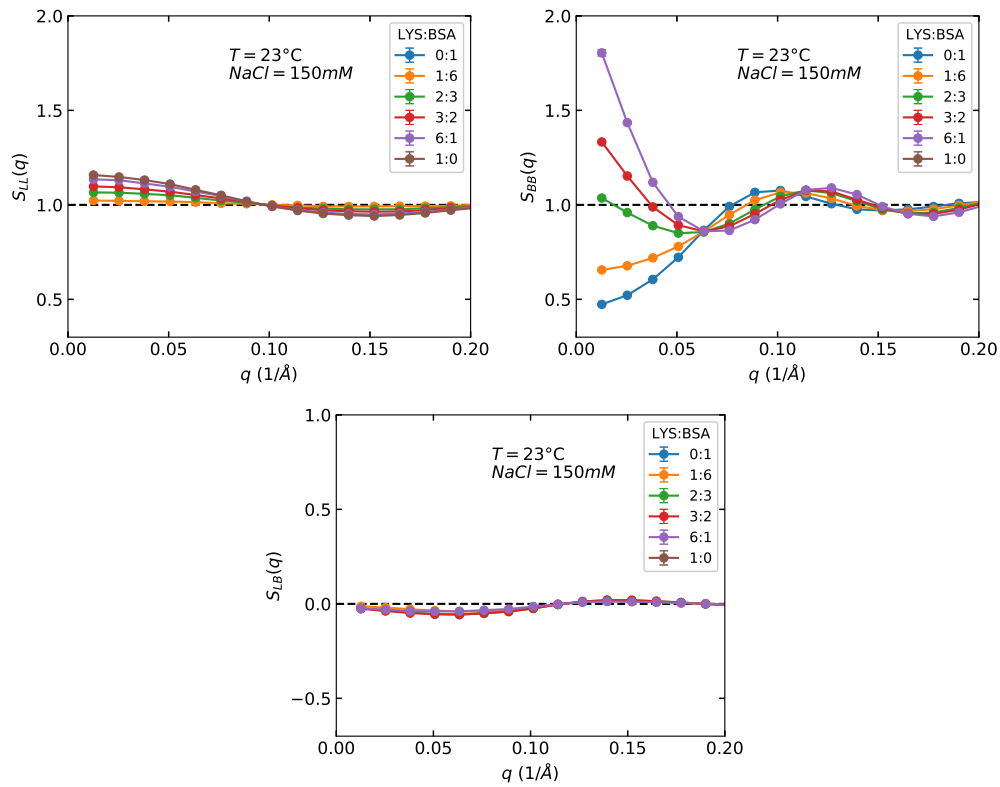


Figura 6.22: Fatores de estrutura das misturas a 150 mM de sal e temperatura de 23 °C.

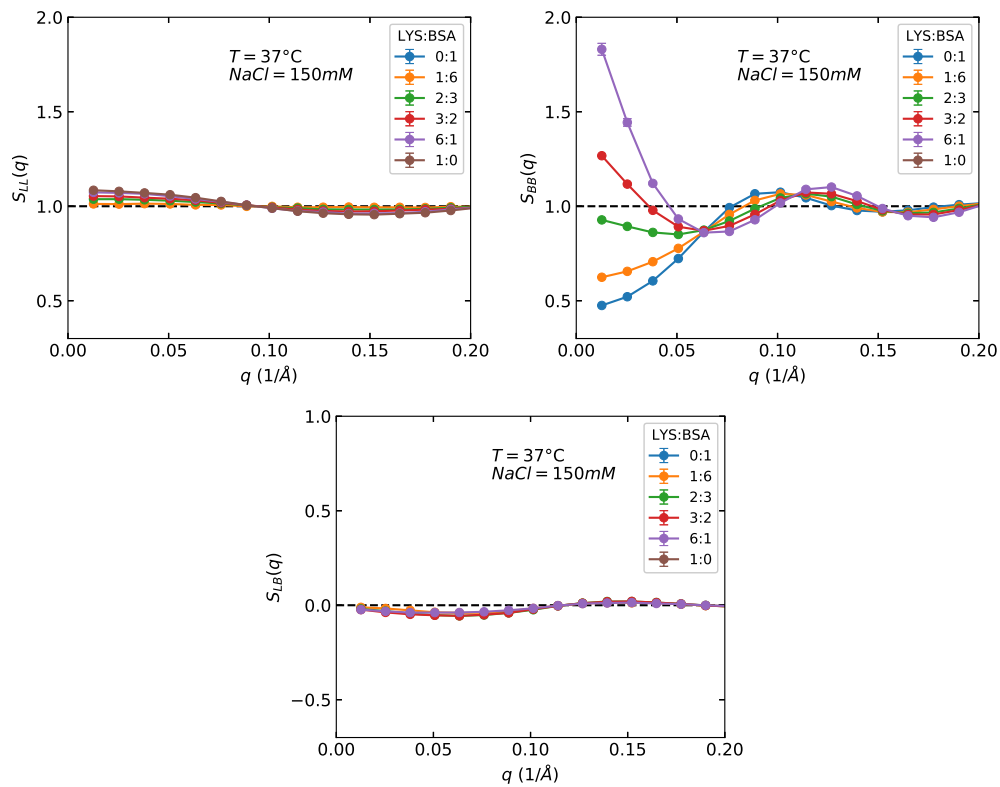


Figura 6.23: Fatores de estrutura das misturas a 150 mM de sal e temperatura de 37 °C.

outra proteína cresce na mistura, mas depois elas atingem um valor máximo e voltam a diminuir. As duas primeiras funções assumem valores tanto acima quanto abaixo da unidade em ângulos pequenos, enquanto que a S_{LB} fica sempre acima de 0 nessa região.

Com 50 mM de sal, não temos mais essas oscilações e as funções assumem um comportamento monotônico. Nesse caso, os fatores de estrutura S_{LL} ou S_{BB} só aumentam conforme a proporção de BSA ou lisozima, respectivamente, cresce em solução. Já o fator S_{LB} aumenta quanto mais lisozima há na mistura. Com 150 mM de sal, o comportamento da função S_{LL} se inverte: ela assume valores acima da unidade e diminui com o aumento da proporção de BSA. Já a função S_{BB} mantém um comportamento parecido com o que tinha a 50 mM de sal. O que é mais interessante, no entanto, é que a função S_{LB} praticamente se anula no regime de maior blindagem. Portanto, é como se pudéssemos descorrelacionar a contribuição de cada proteína para a intensidade total, ou seja, se $S_{LB} \approx 0$, então a equação 6.16 poderia ser escrita simplesmente como:

$$I(q) = n_L \langle |F_L|^2 \rangle S_{M,L} + n_B \langle |F_B|^2 \rangle S_{M,B} \quad (6.16)$$

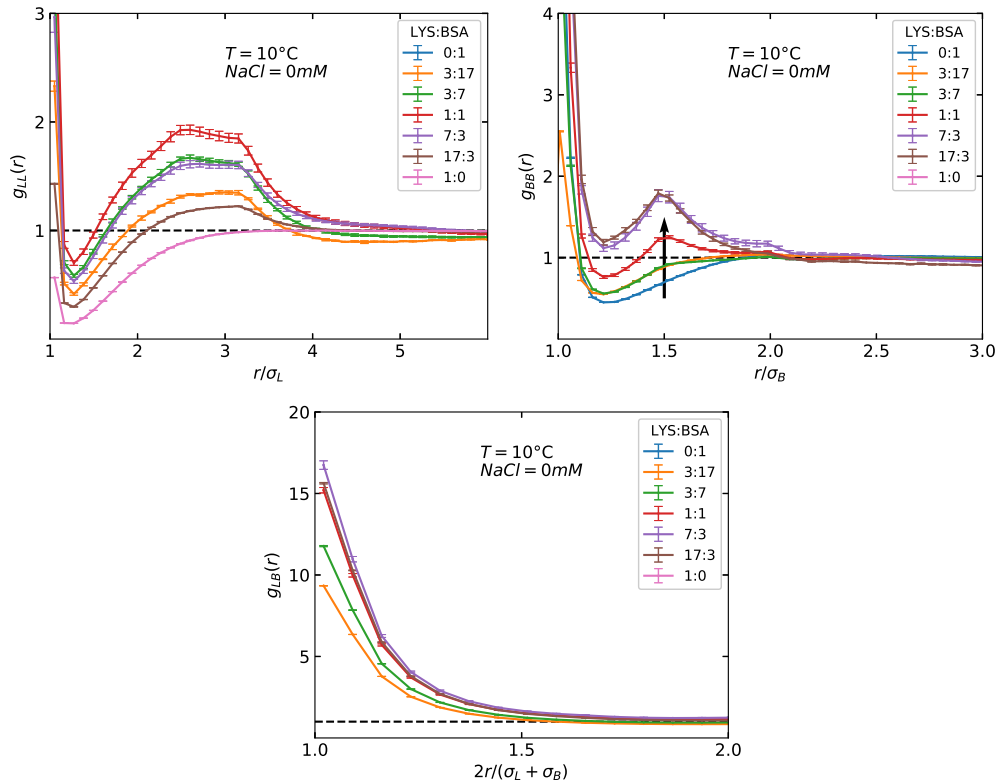


Figura 6.24: Distribuições radiais das misturas a 0 mM de sal e temperatura de 10 °C.

Por fim, as distribuições radiais (figs. 6.24-6.31) revelam um pouco melhor a organização das partículas em solução. Nas misturas sem sal (figs. 6.24-6.26), por exemplo, as funções g_{LL} nos mostram que as lisozimas começam a ficar mais próximas entre si

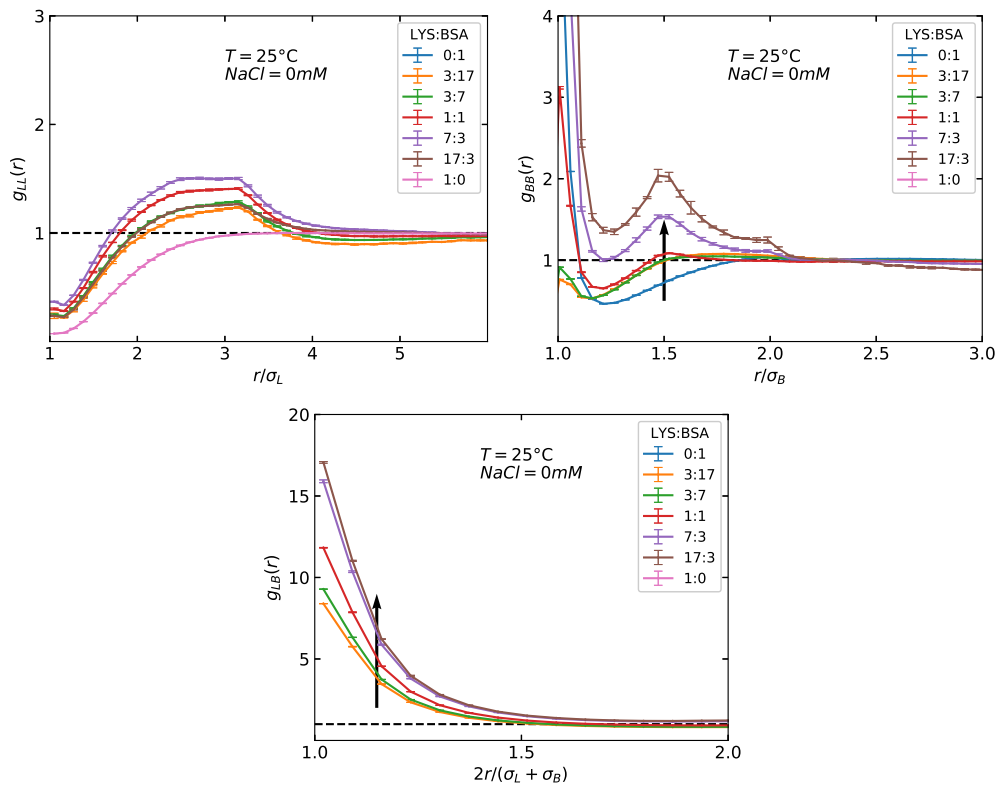


Figura 6.25: Distribuições radiais das misturas a 0 mM de sal e temperatura de 25 °C.

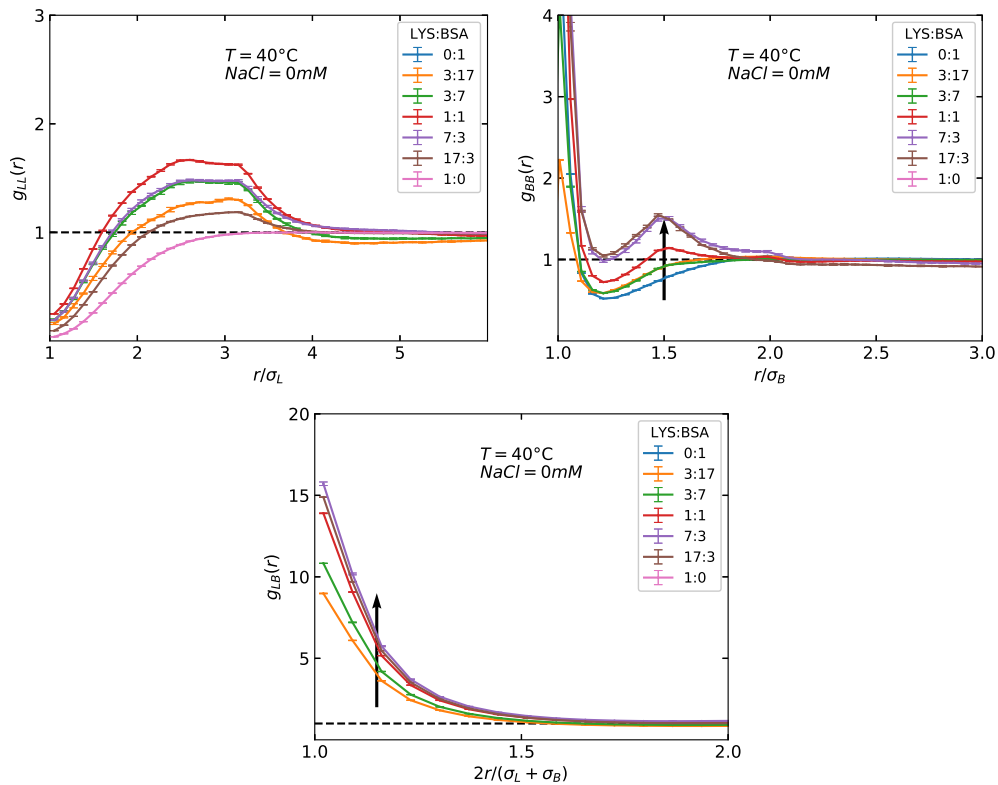


Figura 6.26: Distribuições radiais das misturas a 0 mM de sal e temperatura de 40 °C.

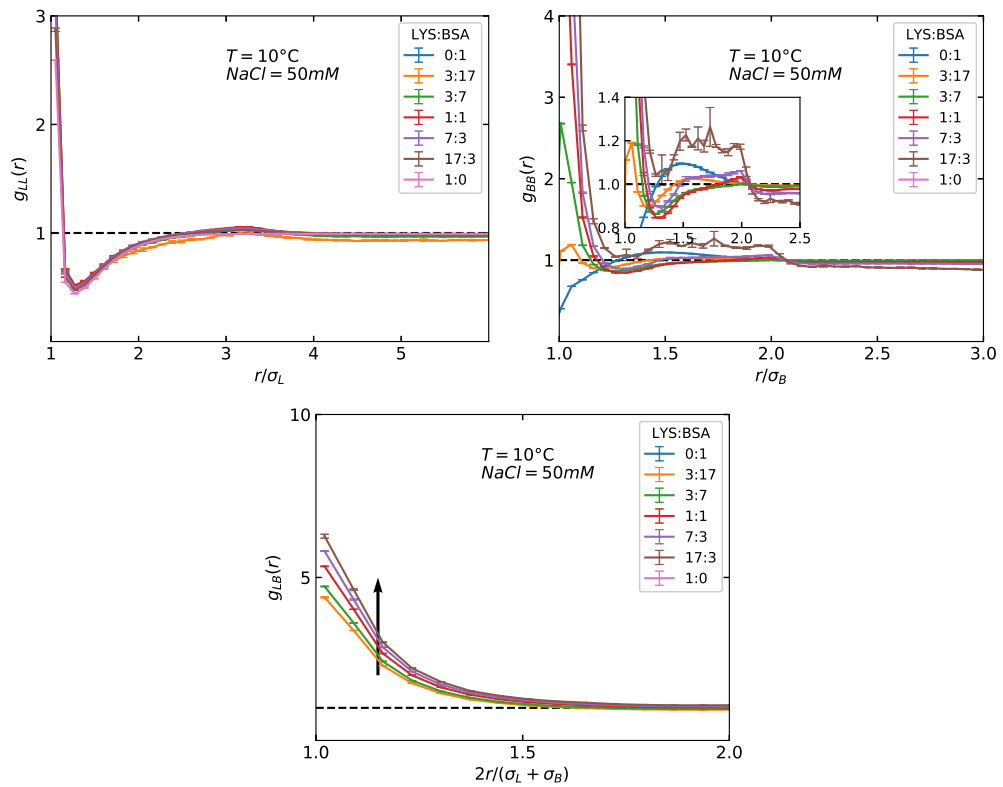


Figura 6.27: Distribuições radiais das misturas a 50 mM de sal e temperatura de 10 °C.

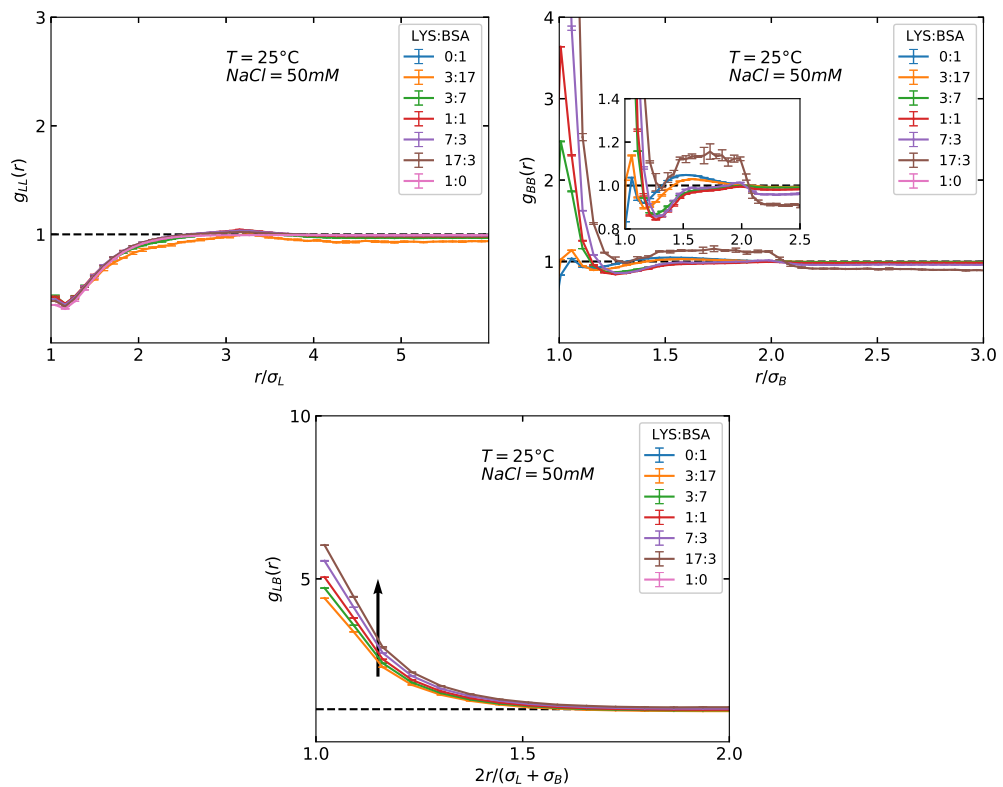


Figura 6.28: Distribuições radiais das misturas a 50 mM de sal e temperatura de 25 °C.

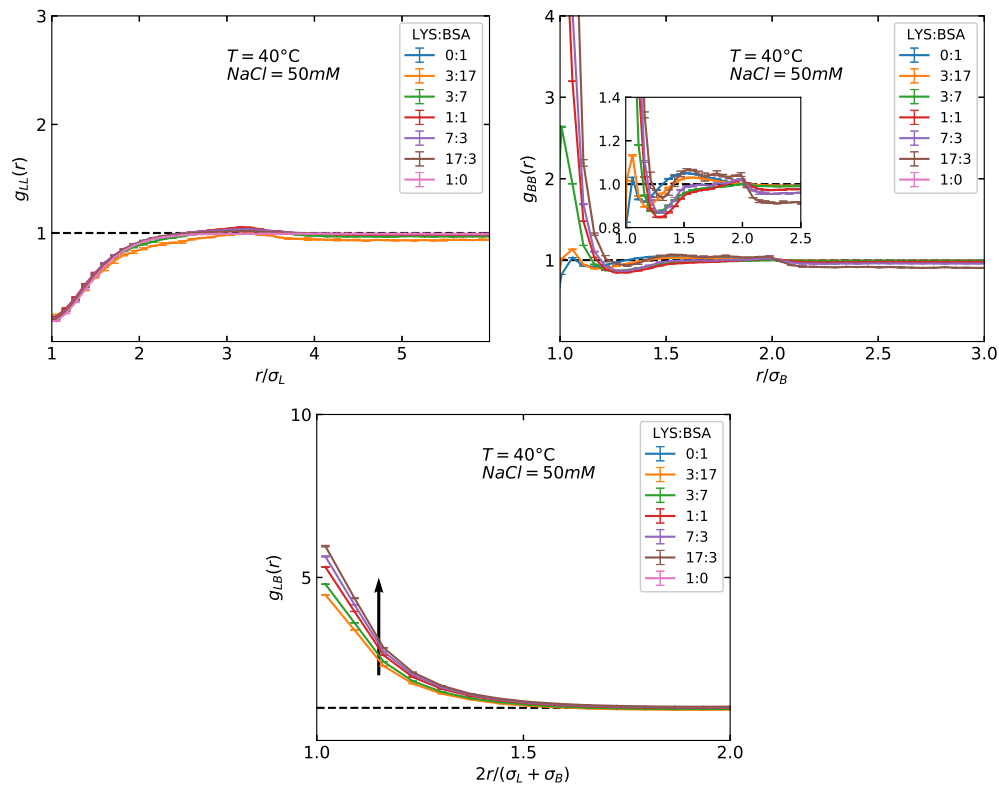


Figura 6.29: Distribuições radiais das misturas a 50 mM de sal e temperatura de 40 °C.

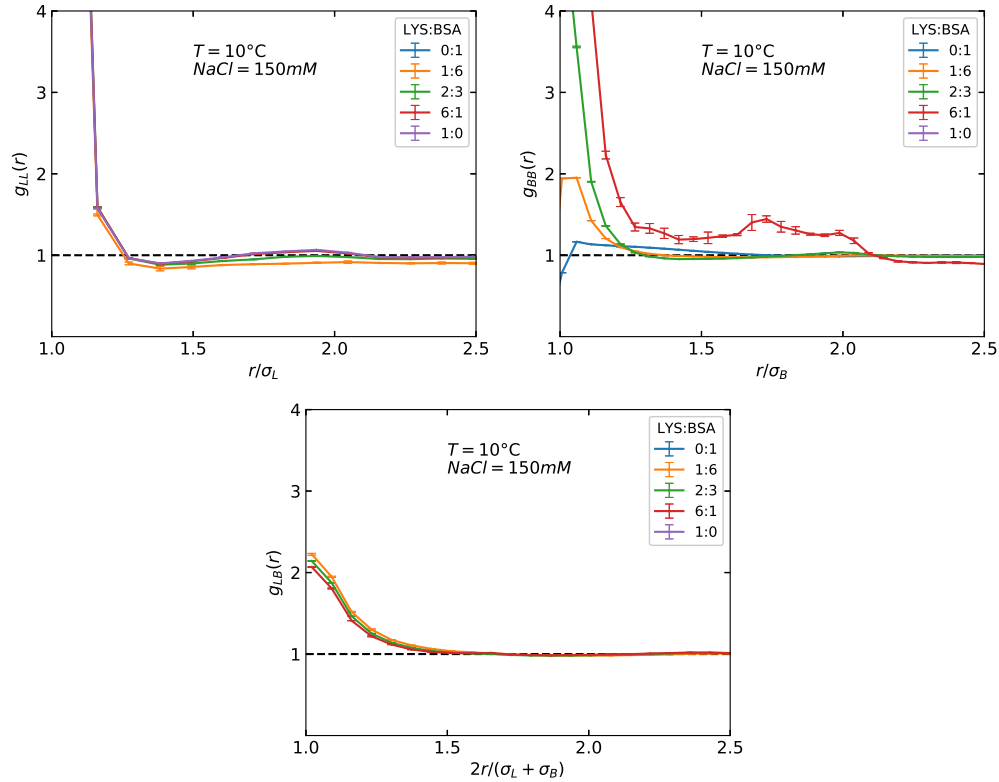


Figura 6.30: Distribuições radiais das misturas a 150 mM de sal e temperatura de 10 °C.

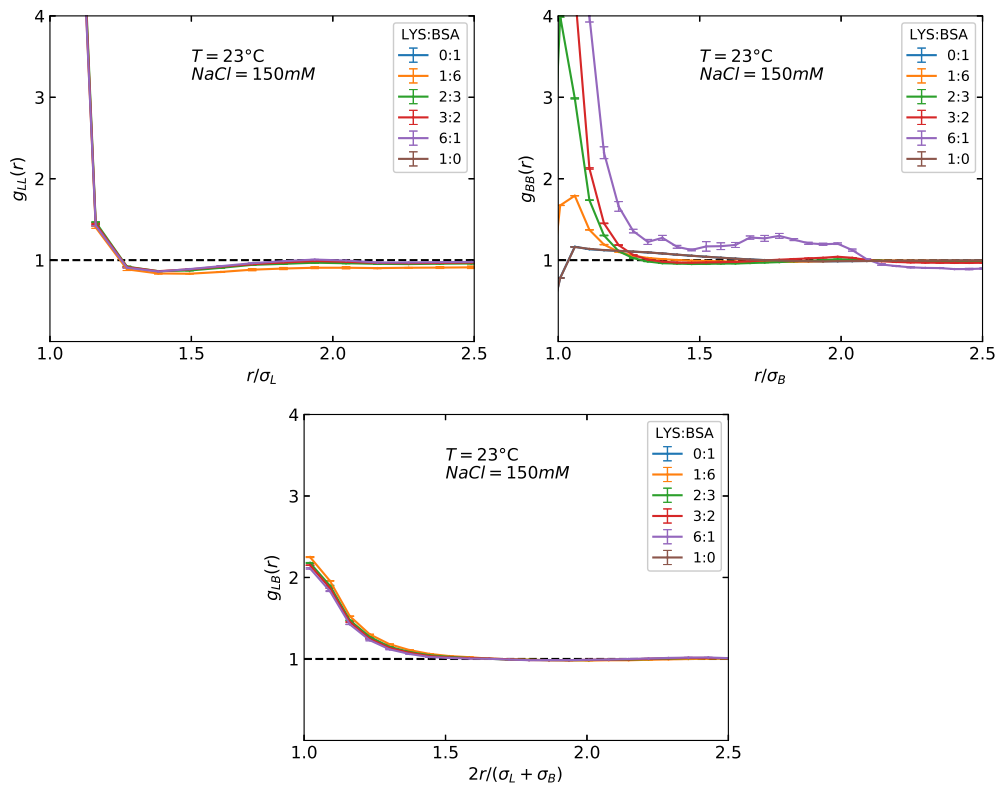


Figura 6.31: Distribuições radiais das misturas a 150 mM de sal e temperatura de 23 °C.

quando aumentamos a concentração de BSA até a proporção de 1:1 (para as temperaturas de 10 e 40 °C) ou 7:3 (para $T = 25$ °C), mas depois elas voltam a se afastar se mais BSA é adicionada. É interessante observar que essa oscilação na g_{LL} com a proporção LYS:BSA acontece apesar do potencial u_{LL} possuir um comportamento monotônico, ou seja, sempre fica menos repulsivo com o aumento de BSA na mistura. Isso mostra que, numa mistura, somente o potencial de interação não é suficiente para inferir a tendência de aproximação ou afastamento das proteínas. Além disso, é possível observar uma região entre 1 e $3.5 \sigma_L$ na qual essa organização é mais afetada. Já as BSAs sempre ficam mais próximas entre si, cf. g_{BB} , quando há mais lisozima na mistura, com uma distância preferencial de $1.5 \sigma_B$. De maneira análoga, quanto mais lisozima em solução maior é a aproximação entre uma lisozima e uma BSA, como podemos ver nos gráficos das funções g_{LB} , onde destacamos essas tendências com uma seta.

Nas concentrações de 50 e 150 mM de sal as curvas das distribuição radial lisozima-lisozima praticamente não mudam com a proporção de LYS:BSA. No primeiro caso, as lisozimas preferem ficar afastadas entre si e evitar contato (ver região onde $g_{LL} < 1$), enquanto no segundo elas tendem a ficar mais perto uma da outra ($g_{LL} > 1$) devido a maior blindagem das cargas. As funções (g_{BB}) apresentam um comportamento mais caótico com 50 mM de sal e não conseguimos enxergar um padrão distinto nas curvas. A julgar apenas pela região imediatamente próxima da superfície podemos perceber um

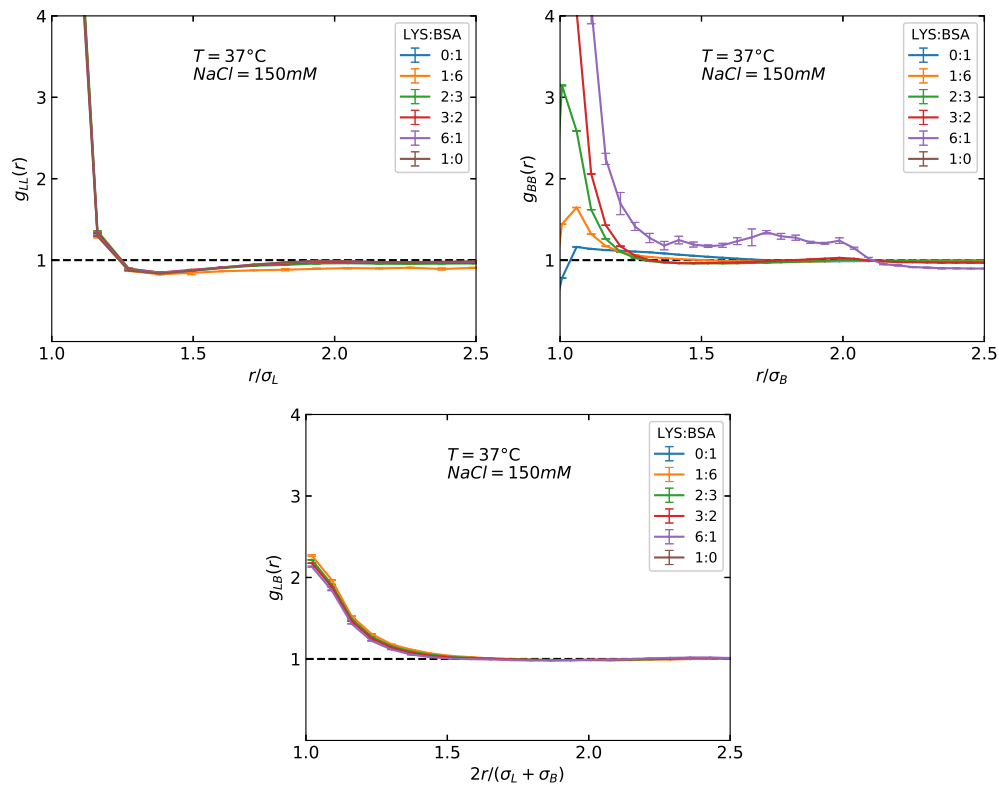


Figura 6.32: Distribuições radiais das misturas a 150 mM de sal e temperatura de 37 °C.

favorecimento da atração entre BSAs com a adição de lisozima. Já com 150 mM de sal esse efeito é claro para um intervalo maior de distâncias.

Em relação às funções g_{LB} , vemos que elas assumem valores sempre acima de um (isto é densidade local acima da média), mas cada vez menores conforme se acrescenta sal no sistema. Como é esperado, lisozima e BSA tendem a ficar próximas espacialmente devido às cargas opostas, mas essa tendência é obviamente atenuada quanto mais blindadas estão as proteínas.

Cabe mencionar também como a estrutura líquida das proteínas se apresenta mais organizada, ou seja, com distâncias espaciais mais definidas (como vimos a 0 mM de NaCl), quando predomina um potencial repulsivo entre as proteínas de um mesmo tipo. Essa estrutura, pelo contrário, vai sendo perdida com a blindagem das cargas, o que, no limite, sugere uma tendência de redução da estabilidade coloidal das soluções e possivelmente a gerar agregação devido ao predomínio das interações atrativas.

Conclusão

Neste trabalho, conseguimos utilizar de maneira integrada e complementar as técnicas de SAXS e Monte Carlo para o estudo de interações proteicas de soluções simples e misturas de BSA e lisozima. Para isso, além dos experimentos de espalhamento, implementamos, desenvolvemos e testamos algoritmos de simulação computacional de sistemas de esferas carregadas.

Num primeiro momento, estudando apenas as soluções de proteínas puras, realizamos uma análise crítica do modelo HSDY/RPA, mostrando que o mesmo apresenta sérias limitações para a modelagem de interações, sobretudo em condições de baixa força iônica e altas densidades. Dentre os problemas apresentados pelo modelo destacamos: (i) a subestimação da carga da lisozima, (ii) a produção de distribuições radiais sem sentido físico e (iii) a predição de potenciais que, quando simulados, geram agregação. Mostramos que essas dificuldades se devem principalmente à relação de fechamento RPA. Para contornar o uso dessa aproximação, propomos um novo método de ajuste da intensidade de espalhamento das proteínas, utilizando simulações de Monte Carlo para modelar o fator de estrutura. O método se mostrou capaz de ajustar as curvas das soluções simples, superando os problemas citados do método anterior.

Num segundo momento, nos debruçamos no problema do ajuste de curvas de SAXS de misturas binárias. Até então, houveram esforços do nosso grupo de pesquisa no sentido de ajustar essas curvas com modelos *ad hoc*, mas que resultavam em parâmetros de difícil interpretação física. Como uma extensão do método para sistemas simples, propomos uma abordagem de ajuste com Monte Carlo utilizando a versão do potencial HSDY para misturas, mas sem o termo de Yukawa para a interação cruzada. Tomando como base inicial os parâmetros das soluções simples, conseguimos obter bons ajustes variando apenas um parâmetro. A partir daí, foi possível obter as contribuições individuais de cada par de proteínas (lisozima-lisozima, lisozima-BSA e BSA-BSA) para a energia potencial, os fatores de estrutura e as funções de distribuição radial. Desse modo, pudemos extrair um conjunto mais robusto e facilmente interpretável de informações sobre as amostras.

Apesar de ser uma contribuição nova na literatura, os códigos e métodos desenvolvidos aqui ainda precisam ser otimizados, pois ainda dependem de um controle manual dos parâmetros a serem ajustados. Como perspectiva futura, pretendemos automatizar o

processo de ajuste para que o software busque, através da minimização de χ^2 , por exemplo, por conjuntos de parâmetros que se adequem melhor aos dados experimentais. Pretendemos também disponibilizar uma versão amigável para uso pela comunidade científica. Além disso, vemos como possibilidade para uso futuro do método o estudo de sistemas com polidispersão nas cargas e nos tamanhos das partículas.

Bibliografia

- [1] Donald Voet, Judith G Voet, and Charlotte W Pratt. *Fundamentals of biochemistry: Life at the molecular level*. John Wiley & Sons, Nashville, TN, 5 edition, 2016.
- [2] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, aug 1970.
- [3] Jonathan King. Deciphering the rules of protein folding. *Chemical and Engineering News*, 67(15):32–54, 1989.
- [4] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, jul 1973.
- [5] Zhaohui Xu, Arthur L. Horwich, and Paul B. Sigler. The crystal structure of the asymmetric GroEL-GroES-(ADP)₇ chaperonin complex. *Nature*, 388(6644):741–750, 1997.
- [6] Alexander Fleming. On a remarkable bacteriolytic element found in tissues and secretions. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 93(653):306–317, 1922.
- [7] P. Laschtschenko. Über die keimtötende und entwicklungshemmende Wirkung von Hühnereiweiß. *Zeitschrift für Hygiene und Infektionskrankheiten*, 64(1):419–427, 1909.
- [8] C. C.F. Blake, D. F. Koenig, G. A. Mair, A. C.T. North, D. C. Phillips, and V. R. Sarma. Structure of hen egg-white lysozyme: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature*, 206(4986):757–761, may 1965.
- [9] Sathyadevi Venkataramani, Jeremy Truntzer, and Denis R. Coleman. Thermal stability of high concentration lysozyme across varying pH: A Fourier Transform Infrared study. *Journal of Pharmacy and Bioallied Sciences*, 5(2):148–153, 2013.
- [10] J. D. Hoppe, P. C. Scriba, and H. Klüter. Transfusion Medicine and Hemotherapy - Chapter 5. Human Albumin. In *Transfusion medicine and hemotherapy : officielles*

Organ der Deutschen Gesellschaft für Transfusionsmedizin und Immunhamatologie, volume 36, chapter 5 Human AI, pages 399–407. 2009.

- [11] Anna Bujacz. Structures of bovine, equine and leporine serum albumin. *Acta Crystallographica Section D: Biological Crystallography*, 68(10):1278–1289, 2012.
- [12] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, and K. Kobayashi. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Engineering*, 12(6):439–446, 1999.
- [13] K. Baler, O. A. Martin, M. A. Carignano, G. A. Ameer, J. A. Vila, and I. Szleifer. Electrostatic unfolding and interactions of albumin driven by pH changes: A molecular dynamics study. *Journal of Physical Chemistry B*, 118(4):921–930, 2014.
- [14] Amit Kumar Srivastav, Sanjeev K. Gupta, and Umesh Kumar. A molecular simulation approach towards the development of universal nanocarriers by studying the pH- And electrostatic-driven changes in the dynamic structure of albumin. *RSC Advances*, 10(23):13451–13459, 2020.
- [15] G. Scanavachi, Y.R. Espinosa, J.S. Yoneda, R. Rial, J.M. Ruso, and R. Itri. Aggregation features of partially unfolded bovine serum albumin modulated by hydrogenated and fluorinated surfactants: Molecular dynamics insights and experimental approaches. *Journal of Colloid and Interface Science*, 572:9–21, jul 2020.
- [16] Ozlem Keskin, Attila Gursoy, Buyong Ma, and Ruth Nussinov. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chemical Reviews*, 108(4):1225–1244, 2008.
- [17] Roy Malka, Francisco Feijó Delgado, Scott R. Manalis, and John M. Higgins. In Vivo Volume and Hemoglobin Dynamics of Human Red Blood Cells. *PLoS Computational Biology*, 10(10), 2014.
- [18] Joseph Horwitz, Michael P. Bova, Lin Lin Ding, Dana A. Haley, and Phoebe L. Stewart. Lens α -crystallin: Function and structure. *Eye (Basingstoke)*, 13(3):403–408, may 1999.
- [19] B. Derjaguin and L. Landau. Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Progress in Surface Science*, 43(1-4):30–59, 1993.
- [20] E.J.W Verwey and J.Th.G Overbeek. Theory of the stability of lyophobic colloids. *Journal of Colloid Science*, 10(2):224–225, apr 1955.

- [21] Luís Fernando Mercier Franco, Cristiano Luis Pinto de Oliveira, and Pedro de Alcantara Pessoa Filho. Thermodynamics of protein aqueous solutions: From the structure factor to the osmotic pressure. *AIChE Journal*, 61(9):2871–2880, sep 2015.
- [22] Weimin Li, Björn A. Persson, Maxim Morin, Manja A. Behrens, Mikael Lund, and Malin Zackrisson Oskolkova. Charge-induced patchy attractions between proteins. *Journal of Physical Chemistry B*, 119(2):503–508, 2015.
- [23] Fajun Zhang, Maximilian W. A. Skoda, Robert M. J. Jacobs, Richard A. Martin, Christopher M. Martin, and Frank Schreiber. Protein Interactions Studied by SAXS: Effect of Ionic Strength and Protein Concentration for BSA in Aqueous Solutions. *The Journal of Physical Chemistry B*, 111(1):251–259, jan 2007.
- [24] Janaky Narayanan and X. Y. Liu. Protein interactions in undersaturated and supersaturated solutions: A study using light and X-ray scattering. *Biophysical Journal*, 84(1):523–532, 2003.
- [25] Leandro R.S. Barbosa, Maria Grazia Ortore, Francesco Spinozzi, Paolo Mariani, Sigrid Bernstorff, and Rosangela Itri. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: A small-angle x-ray scattering study. *Biophysical Journal*, 98(1):147–157, 2010.
- [26] Mikael Lund and Bo Jönsson. On the charge regulation of proteins. *Biochemistry*, 44(15):5722–5727, 2005.
- [27] Achille Giacometti, Domenico Gazzillo, Giorgio Pastore, and Tushar Kanti Das. Numerical study of a binary Yukawa model in regimes characteristic of globular proteins in solutions. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(3):1–10, 2005.
- [28] Francesco Spinozzi, Claudio Ferrero, Maria Grazia Ortore, Alejandro De Maria Antolinos, and Paolo Mariani. GENFIT: Software for the analysis of small-angle X-ray and neutron scattering data of macro-molecules in solution. *Journal of Applied Crystallography*, 47(3):1132–1139, 2014.
- [29] Steven S. Andrews. Using rotational averaging to calculate the bulk response of isotropic and anisotropic samples from molecular parameters. *Journal of Chemical Education*, 81(6):877–885, jun 2004.
- [30] P. Debye. Zerstreuung von Röntgenstrahlen. *Annalen der Physik*, 351(6):809–823, 1915.
- [31] G. Porod. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen - I. Teil. *Kolloid-Zeitschrift*, 124(2):83–114, 1951.

- [32] Andre Guinier. *Small-angle Scattering of X-rays*. John Wiley & Sons, New York, 1955.
- [33] André Guinier. La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Annales de Physique*, 11(12):161–237, apr 1939.
- [34] Gérard Fournet. Étude théorique et expérimentale de la diffusion des rayons X par les ensembles denses de particules. *Bulletin de la Société française de Minéralogie et de Cristallographie*, 74(1):37–172, 1951.
- [35] O Kratky and G Porod. Diffuse small-angle scattering of x-rays in colloid systems. *Journal of Colloid Science*, 4(1):35–70, feb 1949.
- [36] Jean Pierre Hansen and I. R. McDonald. *Theory of Simple Liquids: With Applications to Soft Matter: Fourth Edition*. Academic Press Inc. (London) Ltd., 4 edition, 2013.
- [37] Jerome K. Percus and George J. Yevick. Analysis of Classical Statistical Mechanics by Means of Collective Coordinates. *Physical Review*, 110(1):1–13, 1958.
- [38] J. L. Lebowitz and J. K. Percus. Mean spherical model for lattice gases with extended hard cores and continuum fluids. *Physical Review*, 144(1):251–258, 1966.
- [39] J. M.J. van Leeuwen, J. Groeneveld, and J. de Boer. New method for the calculation of the pair correlation function. I. *Physica*, 25(7-12):792–808, jan 1959.
- [40] Jean Pierre Hansen and John B. Hayter. A rescaled MSA structure factor for dilute charged colloidal dispersions. *Molecular Physics*, 46(3):651–656, jun 1982.
- [41] Forrest J. Rogers and David A. Young. New, thermodynamically consistent, integral equation for simple fluids. *Physical Review A*, 30(2):999–1007, 1984.
- [42] Francesco Spinozzi, Domenico Gazzillo, Achille Giacometti, Paolo Mariani, and Flavio Carsughi. Interaction of proteins in solution from small-angle scattering: A perturbative approach. *Biophysical Journal*, 82(4):2165–2175, 2002.
- [43] E. Hückel. Zur Theorie der Elektrolyte. *Ergebnisse der exakten naturwissenschaften*, 24:199–276, 2007.
- [44] Robert P H Gasser and W Graham Richards. *an Introduction To Statistical Thermodynamics*. Courier Corporation, New York, 2010.
- [45] Per Linse. Structure, phase stability, and thermodynamics in charged colloidal solutions. *The Journal of Chemical Physics*, 113(10):4359–4373, sep 2000.

- [46] Per Linse. Simulation of Charged Colloids in Solution. In *Advanced Computer Simulation Approaches for Soft Matter Sciences II*, volume 185, pages 111–162. Springer-Verlag, Berlin/Heidelberg, 2005.
- [47] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, jun 1953.
- [48] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.
- [49] A Markov and R Howard. Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain In Dynamic Probabilistic Systems. In *Markov Models*, volume 1, chapter Appendix B, pages 552–577. Wiley, 1 edition, 1971.
- [50] Tomé T. and M J D E Oliveira. *Dinâmica Estocástica e Irreversibilidade*. Edusp, São Paulo, 2 edition, 2001.
- [51] Daan Frenkel and Berend Smit. *Understanding molecular simulation: From algorithms to applications*. Academic Press Inc. (London) Ltd., 2002.
- [52] S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 373(1752):27–56, oct 1980.
- [53] Wagner Gomes Rodrigues Junior. “*Sistemas carregados: modelos de simulação*”. Dissertação de mestrado, Universidade de São Paulo, IFUSP, 2011.
- [54] J.W Eastwood and R.W Hockney. Shaping the force law in two-dimensional particle-mesh models. *Journal of Computational Physics*, 16(4):342–359, dec 1974.
- [55] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, jun 1993.
- [56] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73(2):325–348, dec 1987.
- [57] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C : the Art of Scientific Computing*. Cambridge University Press, New York, second edi edition, 1992.

- [58] Jiri Kolafa and John W. Perram. Cutoff Errors in the Ewald Summation Formulae for Point Charge Systems. *Molecular Simulation*, 9(5):351–368, jan 1992.
- [59] Ronald M. Pratt. Madelung constants for ionic crystals using the Ewald sum. *Jurnal Kejuruteraan*, 13:21–39, 2001.
- [60] Marc Malfois, Françoise Bonneté, Luc Belloni, and Annette Tardieu. A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. *Journal of Chemical Physics*, 105(8):3290–3300, 1996.
- [61] A. Tardieu, A. Le Verge, M. Malfois, F. Bonneté, S. Finet, M. Riès-Kautt, and L. Belloni. Proteins in solution : From X-ray scattering intensities to interaction potentials. *Journal of Crystal Growth*, 196(2-4):193–203, 1999.
- [62] Johannes Möller, Martin A. Schroer, Mirko Erkkamp, Sebastian Grobelny, Michael Paulus, Sebastian Tiemeyer, Florian J. Wirkert, Metin Tolan, and Roland Winter. The effect of ionic strength, temperature, and pressure on the interaction potential of dense protein solutions: From nonlinear pressure response to protein crystallization. *Biophysical Journal*, 102(11):2641–2648, 2012.
- [63] Arnaud Ducruix, Jean Pierre Guilloteau, Madeleine Riès-Kautt, and Annette Tardieu. Protein interactions as seen by solution X-ray scattering prior to crystallogenesis. *Journal of Crystal Growth*, 168(1-4):28–39, oct 1996.
- [64] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proceedings of the National Academy of Sciences*, 95(5):2267–2272, mar 1998.
- [65] José García de la Torre, María L. Huertas, and Beatriz Carrasco. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophysical Journal*, 78(2):719–730, feb 2000.
- [66] Thomas F. Kumosinski and Helmut Pessen. Estimation of sedimentation coefficients of globular proteins: An application of small-angle X-ray scattering. *Archives of Biochemistry and Biophysics*, 219(1):89–100, nov 1982.
- [67] Chresten R. Søndergaard, Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *Journal of Chemical Theory and Computation*, 7(7):2284–2295, jul 2011.
- [68] Mats H. M. Olsson, Chresten R. Søndergaard, Michał Rostkowski, and Jan H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p

- K a Predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, feb 2011.
- [69] Terrell L Hill. *Introduction to statistical thermodynamics*. Addison-Wesley Educational, Boston, MA, 1960.
- [70] Charles Tanford and Robert Roxby. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry*, 11(11):2192–2198, may 1972.
- [71] Daniel E. Kuehner, Jan Engmann, Florian Fergg, Meredith Wernick, Harvey W. Blanch, and John M. Prausnitz. Lysozyme Net Charge and Ion Binding in Concentrated Aqueous Electrolyte Solutions. *The Journal of Physical Chemistry B*, 103(8):1368–1374, feb 1999.
- [72] Maria Grazia Ortore, Raffaele Sinibaldi, Francesco Spinozzi, Flavio Carsughi, Daniel Clemens, Adalberto Bonincontro, and Paolo Mariani. New insights into urea action on proteins: A SANS study of the lysozyme case. *Journal of Physical Chemistry B*, 112(41):12881–12887, 2008.
- [73] Maria Grazia Ortore, Francesco Spinozzi, Paolo Mariani, Alessandro Paciaroni, Leandro R.S. Barbosa, Heinz Amenitsch, Milos Steinhart, Jacques Ollivier, and Daniela Russo. Combining structure and dynamics: Non-denaturing high-pressure effect on lysozyme in solution. *Journal of the Royal Society Interface*, 6(SUPPL. 5), 2009.
- [74] Martin A. Schroer, Jonas Markgraf, D. C. Florian Wieland, Christoph J. Sahle, Johannes Möller, Michael Paulus, Metin Tolan, and Roland Winter. Nonlinear pressure dependence of the interaction potential of dense protein solutions. *Physical Review Letters*, 106(17):2–5, 2011.
- [75] Johannes Möller, Sebastian Grobelny, Julian Schulze, Andre Steffen, Steffen Bieder, Michael Paulus, Metin Tolan, and Roland Winter. Specific anion effects on the pressure dependence of the protein-protein interaction potential. *Physical Chemistry Chemical Physics*, 16(16):7423–7429, 2014.
- [76] C.G. Malmberg and A.A. Maryott. Dielectric constant of water from 0 to 100 C. *Journal of Research of the National Bureau of Standards*, 56(1):1, jan 1956.
- [77] Charles Tanford and John G. Buzzell. The Viscosity of Aqueous Solutions of Bovine Serum Albumin between pH 4.3 and 10.5. *The Journal of Physical Chemistry*, 60(2):225–231, feb 1956.

- [78] Daniel Fologea, Bradley Ledden, David S. McNabb, and Jiali Li. Electrical characterization of protein molecules by a solid-state nanopore. *Applied Physics Letters*, 91(5):053901, jul 2007.
- [79] A. Vrij. Light scattering of a concentrated multicomponent system of hard spheres in the Percus–Yevick approximation. *The Journal of Chemical Physics*, 69(4):1742–1747, aug 1978.
- [80] N. W. Ashcroft and David C. Langreth. Structure of Binary Liquid Mixtures. I. *Physical Review*, 156(3):685–692, apr 1967.
- [81] Elisa Morandé Sales. *Estudo das interações proteína-proteína, proteína-membranas e proteína-agentes desnaturantes por espalhamento de raios-X a baixos ângulos*. PhD thesis, Universidade de São Paulo, 2018.

Apêndice A

Equação de Poisson-Boltzmann (PB)

Seja um sistema composto de N partículas carregadas, com coordenadas $\mathbf{r}^N = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ e imersas num meio homogêneo de constante dielétrica ϵ . A equação de Poisson, em unidades gaussianas, nos diz que:

$$\nabla^2 \psi(\mathbf{r}, \mathbf{r}^N) = -\frac{4\pi\rho(\mathbf{r}, \mathbf{r}^N)}{\epsilon} \quad (1)$$

Para simplificar, tomemos uma densidade de cargas pontuais:

$$\rho(\mathbf{r}, \mathbf{r}^N) = \sum_i e z_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (2)$$

onde e é a carga elementar e z_i é a carga numérica (adimensional). Fixando as coordenadas da partícula 1 em $\mathbf{r}_1 = \mathbf{r}'$ podemos tirar a média de ambos os lados da equação de Poisson sobre as coordenadas de todas as outras $N - 1$ partículas:

$$\nabla^2 \int \psi(\mathbf{r}, \mathbf{r}^N) f(\mathbf{r}^{N-1}|\mathbf{r}') d\mathbf{r}_2 \dots d\mathbf{r}_N = -\frac{4\pi e}{\epsilon} \sum_i z_i \int \delta(\mathbf{r} - \mathbf{r}_i) f(\mathbf{r}^{N-1}|\mathbf{r}') d\mathbf{r}_2 \dots d\mathbf{r}_N \quad (3)$$

sendo $f(\mathbf{r}^{N-1}|\mathbf{r}')$ a função densidade de probabilidade, condicionada pela posição da partícula 1 estar fixa em \mathbf{r}' . A integral do lado esquerdo é o potencial médio em \mathbf{r} devido à partícula 1:

$$\psi_1(\mathbf{r}) = \langle \psi(\mathbf{r}, \mathbf{r}^N) \rangle_{\mathbf{r}_2, \dots, \mathbf{r}_N} \quad (4)$$

Fazendo uma mudança de variável $\mathbf{u} = \mathbf{r} - \mathbf{r}'$, esse potencial dependerá apenas da distância u entre a partícula de referência e o ponto \mathbf{r} (se o sistema for homogêneo e isotrópico). Além disso, como as partículas de uma mesma espécie j são idênticas, o potencial será o mesmo para qualquer partícula dessa espécie escolhida como referência, de modo que escreveremos:

$$\psi_1(\mathbf{r}) = \psi_j(u) \quad (5)$$

As integrais do lado direito da Eq. (3) podem ser simplificadas como segue:

(A) Se $i = 1$:

$$\int \delta(\mathbf{r} - \mathbf{r}_1) f(\mathbf{r}^{N-1} | \mathbf{r}') d\mathbf{r}_2 \dots d\mathbf{r}_N = \delta(\mathbf{r} - \mathbf{r}_1) \int f(\mathbf{r}^{N-1} | \mathbf{r}') d\mathbf{r}_2 \dots d\mathbf{r}_N = \delta(\mathbf{r} - \mathbf{r}_1) \quad (6)$$

o qual será igual a zero nos pontos da solução onde $\mathbf{r} \neq \mathbf{r}_1$.

(B) Se $i > 1$:

$$\begin{aligned} \int \delta(\mathbf{r} - \mathbf{r}_i) f(\mathbf{r}^{N-1} | \mathbf{r}') d\mathbf{r}_2 \dots d\mathbf{r}_N &= \int \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}' - \mathbf{r}_1) f(\mathbf{r}^N) d\mathbf{r}_N \\ &= \langle \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}' - \mathbf{r}_1) \rangle_{\mathbf{r}^N} \end{aligned} \quad (7)$$

Nessa última passagem relaxamos a condição da partícula 1 estar fixa. A média acima representa a probabilidade das partículas 1 e i estarem nas posições \mathbf{r}' e \mathbf{r} , respectivamente. Substituindo as Eqs. (5-7) em (3), obtemos que:

$$\nabla^2 \psi_j(u) = -\frac{4\pi e}{\epsilon} \sum_i z_i \langle \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}' - \mathbf{r}_1) \rangle \quad (8)$$

onde o caso com $i = 1$ podemos considerar igual a zero. A soma em i sobre todas as partículas pode ser separada em uma soma em k , sobre todas as espécies carregadas, e outra em n_k , representando cada partícula de uma dada espécie k :

$$\nabla^2 \psi_j(u) = -\frac{4\pi e}{\epsilon} \sum_k z_k \sum_{n_k} \langle \delta(\mathbf{r} - \mathbf{r}_{n_k}) \delta(\mathbf{r}' - \mathbf{r}_1) \rangle \quad (9)$$

A segunda somatória representa a probabilidade de encontrarmos qualquer partícula da espécie k em \mathbf{r} , dado que a partícula 1 (da espécie j) está em \mathbf{r}' . Essa função pode ser escrita em termos da função de distribuição radial $g_k^j(\mathbf{r}, \mathbf{r}')$, segundo sua definição:

$$\rho_k^j(\mathbf{r}, \mathbf{r}') = \sum_{n_k} \langle \delta(\mathbf{r} - \mathbf{r}_{n_k}) \delta(\mathbf{r}' - \mathbf{r}_1) \rangle = \rho_k(\mathbf{r}) g_k^j(\mathbf{r}, \mathbf{r}') \quad (10)$$

onde

$$\rho_k(\mathbf{r}) = \sum_{n_k} \langle \delta(\mathbf{r} - \mathbf{r}_{n_k}) \rangle \quad (11)$$

é igual à densidade média $\rho_k = N_k/V$, se o sistema for homogêneo. Além disso, se o sistema também for isotrópico:

$$g_k^j(\mathbf{r}, \mathbf{r}') = g_k^j(|\mathbf{r} - \mathbf{r}'|) = g_k^j(u) \quad (12)$$

Assim, a Eq. (9) fica:

$$\nabla^2 \psi_j(u) = -\frac{4\pi e}{\epsilon} \sum_k z_k \rho_k g_k^j(u) \quad (13)$$

A função de distribuição radial, por sua vez, está relacionada com o potencial de força média $w_{jk}(u)$, entre partículas do tipo j e k , segundo a expressão:

$$g_k^j(u) = e^{-w_{jk}(u)/k_b T} \quad (14)$$

(i) Primeira aproximação: potencial de força média igual à energia potencial média

Para chegarmos à equação de Poisson-Boltzmann precisamos supor que valha a seguinte aproximação:

$$w_{jk}(u) \approx e z_k \psi_j(u) \quad (15)$$

isto é, que o potencial de força média à uma distância de u é igual ao trabalho médio para trazer uma carga $e z_k$ do infinito para uma posição que dista de u de uma carga da espécie j , trabalho devido somente ao potencial médio de j (ignorando as interações das outras partículas). Obtemos, com isso, a Eq. de Poisson-Boltzmann:

$$\nabla^2 \psi_j(u) = -\frac{4\pi e}{\epsilon} \sum_k z_k \rho_k e^{-e z_k \psi_j(u)/k_b T} \quad (16)$$

(ii) Segunda aproximação: linearização da equação PB

Linearizando a exponencial, $e^x \approx 1 + x$, $|x| \ll 1$, obtemos que:

$$\nabla^2 \psi_j(u) = -\frac{4\pi e}{\epsilon} \sum_k z_k \rho_k [1 - e z_k \beta \psi_j(u)] \quad (17)$$

valida para $|e z_k \beta \psi_j(u)| \ll 1$.

Equação de Helmholtz

Por simplicidade, vamos considerar agora um sistema composto por apenas uma espécie de macroíon, de diâmetro σ , e microíons de raio nulo. O potencial médio, a seguir, será relativo a um macroíon M . Assim, é preciso distinguir três casos:

(A) Para $r < \sigma/2$:

$$\nabla^2 \psi_M(u) = 0 \quad (18)$$

pois não há outros íons dentro do macroíon e sua carga está toda concentrada na origem.

(B) Para $\sigma/2 < r < \sigma$, não pode haver outros macroíons nessa região devido ao choque estérico. Logo, a somatória anterior só incluirá os microíons, o que indicaremos pelo símbolo $'$:

$$\nabla^2 \psi_M(u) = -\frac{4\pi e}{\epsilon} \sum_k' z_k \rho_k + \frac{4\pi e^2 \beta}{\epsilon} \psi_M(u) \sum_k' z_k^2 \rho_k \quad (19)$$

ou, de forma simplificada:

$$\nabla^2 \psi_M(u) = c + k_1^2 \psi_M(u) \quad (20)$$

onde,

$$c = -\frac{4\pi e}{\epsilon} \sum_k' z_k \rho_k \quad \text{e} \quad k_1^2 = \frac{4\pi e^2 \beta}{\epsilon} \sum_k' z_k^2 \rho_k \quad (21)$$

(C) Para $r > \sigma$ pode haver outros macroíons nessa região. Logo, a equação de Poisson-Boltzmann linearizada fica:

$$\nabla^2 \psi_M(u) = -\frac{4\pi e}{\epsilon} \sum_k z_k \rho_k + \frac{4\pi e^2 \beta}{\epsilon} \psi_M(u) \sum_k z_k^2 \rho_k \quad (22)$$

Devido à condição de neutralidade, o primeiro termo dessa equação é nulo. Assim, definindo a força iônica como $I = (1/2) \sum_i z_i^2 \rho_i$, teremos:

$$\nabla^2 \psi_M(u) = \frac{8\pi e^2 \beta I}{\epsilon} \psi_M(u) = k_2^2 \psi_M(u) \quad (23)$$

sendo k_2 o inverso do comprimento de Debye:

$$k_2^{-1} = k_D = \sqrt{\frac{\epsilon}{8\pi e^2 \beta I}} \quad (24)$$

A expressão (23) é a forma tridimensional da chamada equação de Helmholtz.

Apêndice B

Potencial de Coulomb blindado segundo Verwey e Overbeek

Vamos apresentar uma parte da derivação do potencial de Coulomb blindado segundo a abordagem de Verwey e Overbeek (VO), tirada do livro de 1948[20].

De acordo com a parte 1 do capítulo III do livro de VO, a energia-livre de um sistema de dupla-camada de geometria arbitrária pode ser escrita como:

$$F = - \int_S ds \int_0^{\psi_o} \sigma(\psi'_o) d\psi'_o \quad (25)$$

onde S é a superfície do sistema carregado, σ a densidade de carga superficial e ψ_o o potencial elétrico médio na superfície. Vamos supor que o sistema em questão seja um fluido de esferas carregadas e idênticas, imersas num meio com microíons (contraíons e íons de sal). As esferas interagem via um potencial elétrico de pares efetivo ψ .

(1) Hipótese: a densidade superficial é proporcional ao potencial médio na superfície

Suponha que a densidade superficial de carga é diretamente proporcional ao potencial médio na superfície ψ_o :

$$\sigma = \alpha\psi_o \quad (26)$$

Sendo assim, a segunda integral da equação (25) fica:

$$\int_0^{\psi_o} \alpha\psi'_o d\psi'_o = \frac{\alpha\psi_o^2}{2} = \frac{\sigma_o\psi_o}{2} \quad (27)$$

onde $\sigma_o = \alpha\psi_o$ é a densidade quando o potencial médio na superfície vale ψ_o . A energia-livre, portanto, será:

$$F = - \int_S \left(\frac{\sigma_o\psi_o}{2} \right) ds = - \frac{\psi_o}{2} \int_S \sigma_o ds = - \frac{\psi_o Q}{2} \quad (28)$$

sendo Q a carga total na superfície da esfera.

Segundo Verwey e Overbeek, a energia média de interação entre duas esferas separadas por uma distância centro-a-centro de r é igual à variação total de energia-livre ao

aproximá-las desde o infinito até essa separação r :

$$U(r) = 2\Delta F = 2[F(r) - F(\infty)] \quad (29)$$

Substituindo (28) em (29), obtemos:

$$U(r) = \psi_o(\infty)Q(\infty) - \psi_o(r)Q(r) \quad (30)$$

A partir daqui, uma de duas hipóteses é considerada: ou o potencial médio de superfície é considerado constante, independente da separação entre as esferas, ou é a carga superficial das mesmas que é fixa. O livro afirma que as duas hipóteses levam aos mesmos resultados para $U(r)$.

(2) Hipótese: Ou o potencial médio de superfície é constante, ou a carga superficial é constante

$$U(r) = \begin{cases} \psi_o [Q(\infty) - Q(r)] , & \text{se } \psi_o \text{ é constante} \\ Q [\psi_o(\infty) - \psi_o(r)] , & \text{se } Q \text{ é constante} \end{cases} \quad (31)$$

Enquanto Levine e Dube utilizaram a hipótese de carga constante em seus trabalhos, Verwey e Overbeek acham a hipótese do potencial de superfície ser constante mais razoável. Será esta, portanto, a que utilizaremos a seguir. Imagine então duas esferas carregadas, de raio a , separadas por uma distância r . O potencial médio na superfície de ambas é o mesmo e possui valor constante ψ_o , como havíamos suposto:

$$\psi_o = \frac{1}{4\pi a^2} \int_S \psi(\mathbf{r}) ds \quad (32)$$

Já a carga de uma esfera pode ser calculada aplicando-se o teorema de Gauss:

$$\frac{4\pi Q(r)}{\epsilon} = \int_S \vec{E} d\vec{s} = - \int_S \frac{\partial \psi(\mathbf{r})}{\partial r} ds \quad (33)$$

A equação (32), juntamente com a suposição de que o potencial efetivo total, $\psi(\mathbf{r})$, desse sistema de duas esferas, é nulo no infinito formam as condições de contorno do problema. Já a equação (33) nos permite, além de obter a carga superficial, escrever o potencial médio de superfície em termos de $Q(r)$. Assim, o problema se resume em resolver a equação de Helmholtz (ou de Poisson-Boltzmann linearizada):

$$\nabla^2 \psi(\mathbf{r}) = k^2 \psi(\mathbf{r}) \quad (34)$$

no espaço fora das duas esferas, aplicar as condições de contorno e deixar a solução em termos da carga superficial.

(3) Hipótese: Superposição dos potenciais individuais

Para resolver essa última equação outra aproximação é feita. Considera-se que o potencial efetivo total do sistema de duas esferas é dado pela soma dos potenciais efetivos individuais de cada esfera:

$$\psi(\mathbf{r}_1) = \psi_1(\mathbf{r}_1) + \psi_2(\mathbf{r}_2) \quad (35)$$

onde tomamos como origem do sistema de coordenadas o centro da esfera 1 (ver figura abaixo). Aqui, $\psi_1(\mathbf{r}_1)$ é a solução da equação de PB no caso de uma única esfera, ou seja, sem fixar uma segunda esfera a uma distância r da primeira. Idem para a função $\psi_2(\mathbf{r}_2)$. É claro que essa aproximação despreza o efeito que uma esfera produz na dupla camada da outra.

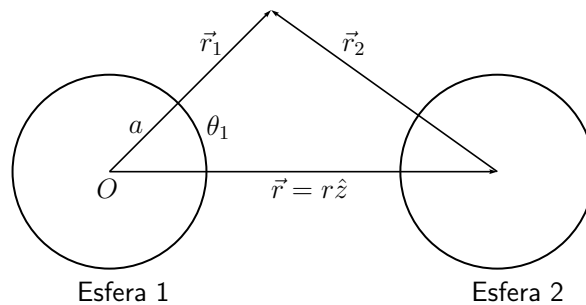


Figura 1: Geometria do sistema de duas esferas.

Ora, a solução do problema de uma única esfera fixa é fácil de ser resolvido, pois ele apresenta simetria radial. Nesse caso, para a esfera 1 teremos que:

$$\frac{1}{r_1^2} \frac{d}{dr_1} \left(r_1^2 \frac{d\psi_1}{dr_1} \right) = k^2 \psi_1 \implies \psi_1 = A \frac{e^{-kr_1}}{r_1} + B \frac{e^{+kr_1}}{r_1} \quad (36)$$

onde $B = 0$, uma vez que o potencial é suposto nulo no infinito. O mesmo vale para o potencial da esfera 2. Assim, o potencial do sistema duplo será:

$$\psi(r_1) = A \left(\frac{e^{-kr_1}}{r_1} + \frac{e^{-kr_2}}{r_2} \right) \quad (37)$$

Para determinar a constante A , aplicamos então a segunda condição de contorno, dada pela equação (32), na superfície da esfera 1:

$$\psi_o = \frac{1}{4\pi a^2} \int_0^{2\pi} \int_0^\pi \psi(\mathbf{r}_1) a^2 \sin \theta_1 d\theta_1 d\phi_1 \quad (38)$$

Se tomarmos o eixo z na direção que liga a esfera 1 até a esfera 2, então o sistema será simétrico na coordenada ϕ_1 , de modo que sobra apenas a integral na coordenada

polar:

$$\psi_o = \frac{A}{2} \int_0^\pi \left(\frac{e^{-ka}}{a} + \frac{e^{-kr_2}}{r_2} \right) \sin \theta_1 d\theta_1 \quad (39)$$

Como a integral é feita na superfície da esfera 1, temos que $r_1 = a$. Já r_2 pode ser expresso, pela lei dos cossenos, como:

$$r_2^2 = a^2 + r^2 - 2ar \cos \theta_1 \quad (40)$$

Diferenciando os dois lados da equação (40), podemos fazer uma substituição de variáveis na integral para deixá-la em termos de r_2 :

$$\psi_o = \frac{A}{2ar} \int_{r-a}^{r+a} \left(\frac{e^{-ka}}{a} + \frac{e^{-kr_2}}{r_2} \right) r_2 dr_2 \quad (41)$$

As integrais acima são resolvidas de maneira trivial e levam ao seguinte resultado:

$$\psi_o = A \frac{e^{-ka}}{a} \left[1 + \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka}) \right] \quad (42)$$

ou, isolando a constante A (que depende da separação r entre as esferas):

$$A = \frac{\psi_o a e^{ka}}{1 + \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka})} \quad (43)$$

Por fim, aplicaremos a lei de Gauss para relacionar o potencial médio de superfície, ψ_o , com a carga superficial Q . Tomando como superfície gaussiana o contorno da esfera 1, teremos que:

$$Q(r) = -\frac{\epsilon}{4\pi} \int_{S_1} \frac{\partial \psi(\mathbf{r}_1)}{\partial r_1} \Big|_{r_1=a} ds \quad (44)$$

Pela mesma razão de simetria já apontada, essa integral pode ser feita apenas sobre a coordenada polar:

$$Q(r) = -\frac{\epsilon a^2}{2} \int_0^\pi \frac{\partial \psi(\mathbf{r}_1)}{\partial r_1} \Big|_{r_1=a} \sin \theta_1 d\theta_1 \quad (45)$$

A derivada parcial no integrando pode ser calculada aplicando a regra da cadeia:

$$\frac{\partial \psi(\mathbf{r}_1)}{\partial r_1} = \frac{\partial}{\partial r_1} [\psi_1(r_1) + \psi_2(r_2)] = \frac{\partial \psi_1(r_1)}{\partial r_1} + \frac{\partial \psi_2(r_2)}{\partial r_2} \frac{\partial r_2}{\partial r_1} \quad (46)$$

sendo que a derivada parcial de r_2 em relação a r_1 pode ser obtida pela derivação da lei dos cossenos:

$$r_2^2 = r_1^2 + r^2 - 2r_1 r \cos \theta_1 \implies \frac{\partial r_2}{\partial r_1} = \frac{r_1 - r \cos \theta_1}{r_2} \quad (47)$$

e usando a lei dos cossenos novamente para deixar a expressão em termo só das distâncias r_1 e r_2 :

$$\frac{\partial r_2}{\partial r_1} = \frac{r_1 + \left(\frac{r_2^2 - r_1^2 - r^2}{2r_1}\right)}{r_2} = \frac{r_1^2 + r_2^2 - r^2}{2r_1 r_2} \quad (48)$$

Assim, substituindo (48) em (46) e fazendo a conta das outras derivadas, ficamos com:

$$\frac{\partial \psi(\mathbf{r}_1)}{\partial r_1} = -\frac{Ae^{-kr_1}}{r_1} \left(\frac{1}{r_1} + k\right) - \frac{Ae^{-kr_2}}{r_2} \left(\frac{1}{r_2} + k\right) \left(\frac{r_1^2 + r_2^2 - r^2}{2r_1 r_2}\right) \quad (49)$$

Finalmente, substituindo a equação (49) em (45), lembrando que $r_1 = a$ na superfície da esfera 1, e fazendo a mudança de variável de θ_1 para r_2 , chegamos na seguinte equação para a carga superficial:

$$Q(r) = -\frac{\epsilon a}{2r} \int_{r-a}^{r+a} \left[-\frac{Ae^{-ka}}{a} \left(\frac{1}{a} + k\right) - \frac{Ae^{-kr_2}}{r_2} \left(\frac{1}{r_2} + k\right) \left(\frac{a^2 + r_2^2 - r^2}{2ar_2}\right) \right] r_2 dr_2 \quad (50)$$

A integral acima pode ser facilmente separada em algumas integrais triviais, porém tediosas. Após esses cálculos, pode-se verificar que o resultado final será dado por:

$$Q(r) = A\epsilon e^{-ka} (1 + ka) \left[1 + \frac{e^{-k(r-2a)}}{2kr} \left(\frac{1 - ka}{1 + ka} - e^{-2ka} \right) \right] \quad (51)$$

Substituindo a constante A, equação (43), na expressão obtida, teremos a carga superficial em termos do potencial de superfície médio:

$$Q(r) = \psi_o \epsilon a (1 + ka) \frac{1 + \frac{e^{-k(r-2a)}}{2kr} \left(\frac{1 - ka}{1 + ka} - e^{-2ka} \right)}{1 + \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka})} \quad (52)$$

No limite em que as esferas estejam muito afastadas uma da outra, a carga superficial tenderá ao seguinte valor:

$$Q(\infty) = \psi_o \epsilon a (1 + ka) \quad (53)$$

É interessante notar que essa expressão para o potencial médio na superfície é análoga à expressão encontrada por Debye-Huckel para o potencial efetivo de uma única esfera:

$$\psi(r) = \frac{Q}{\epsilon(1 + ka)} \frac{e^{-k(r-a)}}{r} \implies \psi(a) = \frac{Q}{\epsilon a (1 + ka)} \quad (\text{Debye-Huckel}) \quad (54)$$

com a diferença de que, naquele caso, tanto a carga quanto o potencial de superfície são supostos constantes.

Temos, então, todas as peças para montar a energia de interação entre duas esferas.

Se substituirmos (52) e (53) na equação (31), ficaremos com:

$$U(r) = \psi_o^2 \epsilon a (1 + ka) \left[1 - \frac{1 + \frac{e^{-k(r-2a)}}{2kr} \left(\frac{1-ka}{1+ka} - e^{-2ka} \right)}{1 + \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka})} \right] \quad (55)$$

Com um pouco de álgebra, podemos reescrever essa expressão da seguinte maneira:

$$U(r) = \frac{\psi_o^2 \epsilon a^2 e^{-k(r-2a)}}{r} \frac{1}{1 + \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka})} \quad (56)$$

Se fizermos uma expansão em Taylor da segunda fração que aparece na equação acima, ou seja,

$$f(y) = \frac{1}{1+y} = \sum_{n=0}^{\infty} (-1)^n y^n, \quad |y| < 1 \quad (57)$$

com

$$y = \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka}) \quad (58)$$

obteremos a seguinte expressão:

$$U(r) = \frac{\psi_o^2 \epsilon a^2 e^{-k(r-2a)}}{r} \sum_{n=0}^{\infty} e^{-nk(r-2a)} \left(\frac{e^{-2ka} - 1}{2kr} \right)^n \quad (59)$$

Se aproximarmos a somatória até a ordem $n = 0$, ficamos com:

$$U(r) = \frac{\psi_o^2 \epsilon a^2 e^{-k(r-2a)}}{r} \quad (60)$$

ou então, escrevendo em termos da carga a grandes distâncias, equação (53), teremos a expressão mais conhecida da interação de Coulomb blindado segundo a teoria DLVO:

$$U(r) = \frac{Q(\infty)^2}{\epsilon(1+ka)^2} \frac{e^{-k(r-2a)}}{r} \quad (61)$$

É claro que essa primeira aproximação da série será válida na medida em que:

$$y(r) = \frac{e^{-k(r-2a)}}{2kr} (1 - e^{-2ka}) \ll 1 \quad (62)$$

Para ver quão boa é essa aproximação vamos analisar o seu valor máximo. É óbvio que $y(r)$ é uma função estritamente decrescente, logo seu valor máximo ocorre na menor separação entre as duas esferas, ou seja, quando $r = 2a$:

$$y(2a) = \frac{1 - e^{-2ka}}{4ka} \quad (63)$$

Essa função de ka , por sua vez, também é estritamente decrescente. Além disso, como o produto $ka > 0$, seu valor será sempre menor do que o limite de $ka \rightarrow 0$, o qual vale exatamente $1/2$. Logo,

$$y(r) \leq y(2a) < 1/2 \tag{64}$$

Assim, no pior dos casos, $y(r) < 0.5$. Mas é claro que em situações físicas nas quais o comprimento de Debye ($k_D = k^{-1}$) e o raio das esferas possuem a mesma ordem de grandeza, e a distância média entre elas for bem maior que $2a$, a condição (62) será facilmente cumprida.