



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Guilherme Henrique Martos

**Estudo dos Efeitos da Regulação de Carga em Modelos
Teóricos de Proteínas Intrinsecamente Desordenadas**

São José do Rio Preto

2025

Guilherme Henrique Martos

**Estudo dos Efeitos da Regulação de Carga em Modelos
Teóricos de Proteínas Intrinsecamente Desordenadas**

Dissertação apresentada à Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, para obtenção do título de Grau acadêmico Mestre em Biofísica Molecular.

Área de Concentração: Ciências Biomoleculares e Farmacológicas

Orientador: Prof. Dr. Sidney J. de Carvalho

**São José do Rio Preto
2025**

M387e Martos, Guilherme Henrique
Estudo dos efeitos da regulação de carga em modelos
teóricos de proteínas intrinsecamente desordenadas /
Guilherme Henrique Martos. -- São José do Rio Preto, 2024
59 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista
(UNESP), Instituto de Biociências Letras e Ciências Exatas,
São José do Rio Preto
Orientador: Sidney Jurado de Carvalho

1. Física. 2. Física Biológica. 3. Polieletrólitos. I. Título.

Guilherme Henrique Martos

Estudo dos Efeitos da Regulação de Carga em Modelos Teóricos de Proteínas Intrinsecamente Desordenadas

Dissertação apresentada à Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, para obtenção do título de Grau acadêmico Mestre em Biofísica Molecular.

Área de Concentração: Ciências Biomoleculares e Farmacológicas

Data da defesa: 13/09/2024

Banca Examinadora:

Prof. Dr. Sidney Jurado de Carvalho

UNESP - Instituto de Biociências, Letras e Ciências Exatas - Campus de São José do Rio Preto

Prof. Dr. João Ruggiero Neto

UNESP/Câmpus de São José do Rio Preto

Prof. Dr. Luiz Felipe Santos Mendes

Instituto de Física de São Carlos / Universidade de São Paulo

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Quero agradecer a todos os meus amigos pelos momentos de descontração e incentivo. Em especial ao Dante, Ricardo, Henry, Felipe e Ana Júlia, tanto pelas aventuras nas mesas de RPG como pela força em programação, que nunca foi o meu forte.

Agradeço demais à minha namorada Marjorie, por todo amor, carinho, inspiração e paz que vem me dando.

Agradeço a todos os professores da Unesp por todo o conhecimento que adquiri e sei que ainda vou aprender muito com eles. Em especial, aos professores João Ruggiero, Fátima e, principalmente, ao meu orientador Sidney, por toda a paciência, dedicação e inspiração. Espero, em minha carreira, ser ao menos uma fração dos excelentes cientistas que vocês são.

Por último, e não menos importante, agradeço ao IBILC e Fapesp por todo financiamento e infraestrutura, sem os quais não poderia ter feito este trabalho. Sinto que as palavras são poucas para expressar o quão grato sou a todos vocês, mesmo aqueles que não consegui mencionar aqui.

”No fim tudo dá certo, e se não deu certo é porque ainda não chegou ao fim”

– Fernando Sabino

Resumo

Proteínas intrinsecamente desordenadas, ou IDPs (*Intrinsically Disordered Proteins*), são proteínas que não possuem uma estrutura bem definida como no caso das proteínas globulares. Apesar disso, essa classe de proteínas participa de diversos processos biológicos importantes. Devido ao seu comportamento, que desafia o paradigma estrutura-função, as IDPs foram por muito tempo consideradas apenas exceções a essa regra, sendo relativamente pouco estudadas em relação às proteínas globulares. Verifica-se que as interações eletrostáticas são as interações predominantes nos processos envolvendo sua ligação com uma molécula alvo e nos mecanismos de reconhecimento molecular que promovem a estruturação da IDP, mas muitos detalhes sobre esses mecanismos ainda são desconhecidos. Nesse sentido, podemos especular que o mecanismo de regulação de carga pode ser importante ao reconhecimento molecular, uma vez que, nesse processo, a distribuição de carga da IDP sofre alterações quando na presença de uma molécula alvo, promovendo uma complementaridade eletrostática e conformacional. Por esse motivo, este trabalho tem como objetivo estudar os efeitos da regulação de carga em modelos teóricos de IDPs. Partimos da função de partição de um modelo de proteína caracterizado por uma linha contínua cuja densidade linear de carga é dada pela posição específica de seus resíduos. A energia livre foi obtida considerando as interações eletrostáticas e de volume excluído de maneira efetiva em um comprimento de Kunh renormalizado. Partindo desse modelo teórico, foi possível obter a distribuição de carga com o efeito da regulação e as distâncias *end-to-end* para os três modelos de IDP em função do pH. Foi possível comparar a carga líquida regulada com a situação onde a regulação de carga não é considerada, mostrando que, a carga líquida regulada é menor, causando uma diminuição também na distância *end-to-end*. Os próximos passos consistem em adicionar as contribuições referentes à adição do sal, e a interação com outras macromoléculas.

Palavras-Chave: regulação de carga; IDP; energia livre; interações eletrostáticas.

Abstract

Intrinsically Disordered Proteins (IDPs) are proteins that lack a stable conformation, unlike globular proteins. Despite this, this class of proteins plays a crucial role in various essential biological processes. Due to their behavior that defies the structure-function paradigm, IDPs have been considered exceptions to the rule, receiving relatively less attention compared to globular proteins. It was observed that electrostatic interactions are predominant in processes involving their binding to a target molecule and in the molecular recognition mechanisms that promote IDP structuring. However, many details about these mechanisms are still unknown. In this regard, it can be speculated that the charge regulation mechanism may be important in molecular recognition. In this process, the charge distribution of the IDP changes in the presence of a target molecule, promoting electrostatic and conformational complementarity. For this reason, this work aims to study the effects of charge regulation in theoretical models of IDPs. We start from the partition function of a protein model characterized by a continuous line whose linear charge density is given by the specific position of its residues. The free energy was obtained by considering electrostatic interactions and excluded volume interactions effectively over a renormalized Kuhn length. Based on this theoretical model, it was possible to obtain the charge distribution with the effect of regulation and the end-to-end distance for three IDP models as a function of pH. It was possible to compare the regulated net charge with the situation where charge regulation is not considered, showing that the regulated net charge is lower, causing a decrease in the end-to-end distance. The next steps involve adding contributions related to the addition of salt and the interaction with other macromolecules.

Keywords: charge regulation; IDP; free energy; electrostatic interactions.

Lista de Figuras

1.1	Estrutura molecular de um aminoácido.	12
1.2	Ligação peptídica entre dois aminoácidos.	12
1.3	Os 20 aminoácidos encontrados em proteínas, classificados de acordo com a polaridade de suas cadeias laterais.	13
4.1	Fluxograma representando os algoritmos computacionais utilizados para determinar a carga de cada resíduo das proteínas e para cada pH utilizado.	42
5.1	Três modelos de IDPs, onde E são ácidos glutâmicos e K são as lisinas. (a) Monômeros dispostos de forma alternada entre E e K, (b) monômeros dispostos de forma aleatória, ou seja, sua distribuição não segue um padrão definido, (c) monômeros dispostos em dois grandes blocos, o primeiro bloco contendo apenas E e o outro bloco composto por K.	44
5.2	Gráficos de carga líquida por pH para os três modelos de IDPs propostos: distribuição alternada dos resíduos (a); distribuição em trechos alternados (b); e distribuição em blocos (c). A curva preta representa a carga total sem a regulação de carga (SRC); A curva vermelha é a carga total considerando a regulação de carga com a aproximação de Debye-Hückel (DH); A curva em azul consideramos a regulação de carga mas sem essa aproximação.	46
5.3	Gráficos mostrando a carga de cada resíduo em pH=10,79 para os três modelos de IDPs propostos. Sem regulação de carga (SRC), com a regulação de carga e utilizando a aproximação de Debye-Hückel (DH), e com a regulação de carga, porém sem utilizar essa aproximação (RC). Os gráficos (a), (b) e (c) representam a distribuição alternada dos resíduos, os gráficos (d), (e) e (f) a distribuição em trechos alternados e (g), (h) e (i) a distribuição em blocos ao longo da proteína.	47

- 5.4 Gráficos da distância *end-to-end* renormalizada em unidades dessa grandeza para uma cadeia livre de interações (x), pelo pH para os três modelos de IDPs propostos: distribuição alternada dos resíduos (a); distribuição em trechos alternados (b); e distribuição em blocos (c). A curva preta representa a carga total sem a regulação de carga (SRC); A curva vermelha é a carga total considerando a regulação de carga com a aproximação de Debye-Hückel (DH); A curva em azul consideramos a regulação de carga mas sem essa aproximação. 48
- 5.5 Matrizes cujos elementos representam a raiz da distância quadrática média entre o i -ésimo resíduo com o j -ésimo resíduo para os três modelos de IDPs: (a) com resíduos ácidos e básicos distribuídos de forma alternada; (c) em trechos alternados; (e) em dois blocos. As imagens (b), (d) e (f) são meramente ilustrativas, representando espacialmente as IDPs. A escala ao lado está medida em Å e as distância foram calculadas para o pH=4,25. 50
- 5.6 Gráficos de carga líquida por pH para os dois mastoparans: (a) MK578; (b) MK4578911. A curva preta representa o peptídeo sem regulação de carga (SRC) e a curva azul com a regulação de carga (RC). 52
- 5.7 Gráficos mostrando a carga média por resíduo de dois mastoparans: (a)-(b) MK578; (c)-(d) MK4578911 obtidas em pH=9,60. Os pontos pretos representam o peptídeo sem regulação de carga (SRC) e em azul, com a regulação de carga (RC). 53
- 5.8 Matrizes cujos elementos representam a raiz da distância quadrática média entre o i -ésimo resíduo com o j -ésimo resíduo para os dois peptídeos: (a) MK578; (b) MK4578911. A escala ao lado está em Å e as distância foram calculadas para o pH=9,60. 54

Sumário

1	Introdução	11
1.1	Proteínas e Aminoácidos	11
1.2	Proteínas Intrinsecamente Desordenadas	15
1.3	Interações Eletrostáticas em Proteínas	16
2	Objetivos	18
3	Regulação de Carga	19
3.1	Descrição Fenomenológica	19
3.1.1	Descrição Estatística	20
3.1.2	Regulação de Carga Via Energia Livre	21
3.1.3	Aproximação Linear para a Regulação de Carga	24
4	Modelo Físico da Proteína	26
4.1	Determinação das Cargas da Proteína	39
5	Aplicação Para Modelos de IDPs	44
5.1	Aplicação em Peptídeos Antimicrobianos	49
6	Conclusão	55
	Referências	60

Capítulo 1

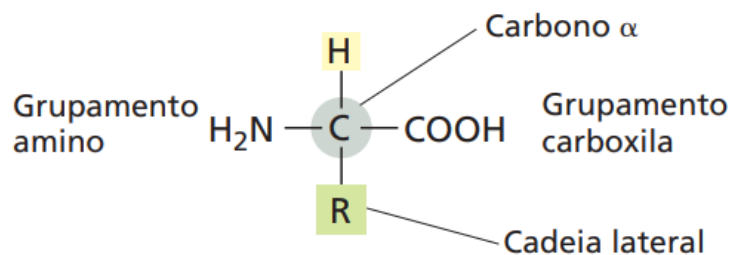
Introdução

1.1 Proteínas e Aminoácidos

As proteínas são macromoléculas biológicas que compõem a maior parte da matéria seca de uma célula (1). Além de participarem da composição celular, as proteínas desempenham funções biológicas muito importantes para a manutenção da vida. Uma classe importante de proteínas são as enzimas, que possuem conformações complexas para atuarem como catalisadores de diversas reações químicas importantes (1, 2). Proteínas imersas na membrana plasmática atuam controlando a passagem de pequenas moléculas que entram ou saem do citoplasma (1, 3). Essas macromoléculas também desempenham o papel de sinalizadores, permitindo a transmissão de informações de uma célula para a outra ou enviam sinais para que uma cascata ocorra no interior da célula, da membrana plasmática para o núcleo (4). Outras proteínas agem como máquinas moleculares com partes móveis, como por exemplo, a *cinesina*, que atua na propulsão de organelas no citoplasma e a *topoisomerase* que tem a função de separar moléculas de DNA emaranhadas (5, 6). Existem inúmeros outros exemplos das funções desempenhadas pelas proteínas além dos citados anteriormente, mostrando o quão fundamental é o conhecimento profundo sobre essas biomoléculas.

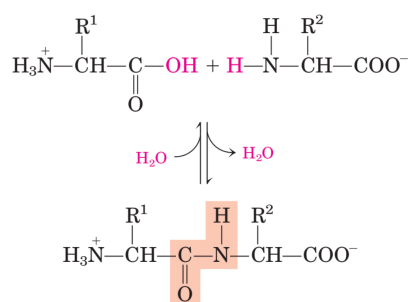
O que vai determinar a estrutura e a função de uma proteína é a sua composição química. Uma molécula de proteína consiste de uma longa cadeia não ramificada de resíduos de aminoácidos, que são suas unidades fundamentais. Estes, por sua vez, são moléculas compostas por um grupamento amina ($-NH_2$), um grupo carboxila ($-COOH$) e uma cadeia lateral ($-R$) ligados a um carbono central, chamado carbono α , como mostra a Figura 1.1. Embora tenham sido descritos mais de 300 aminoácidos na natureza, apenas 20 são encontrados em proteínas (1, 7, 8). Para formar a proteína, os aminoácidos se ligam uns com os outros através de uma ligação covalente, chamada de ligação peptídica e por esse motivo, muitas vezes as proteínas são referidas como cadeias polipeptídicas (1). Na ligação peptídica o grupo carboxila se liga ao grupo amina do aminoácido adjacente com a liberação de uma molécula de água, como representado na Figura 1.2.

Figura 1.1: Estrutura molecular de um aminoácido.



Fonte: Imagem obtida da referência (1).

Figura 1.2: Ligação peptídica entre dois aminoácidos.

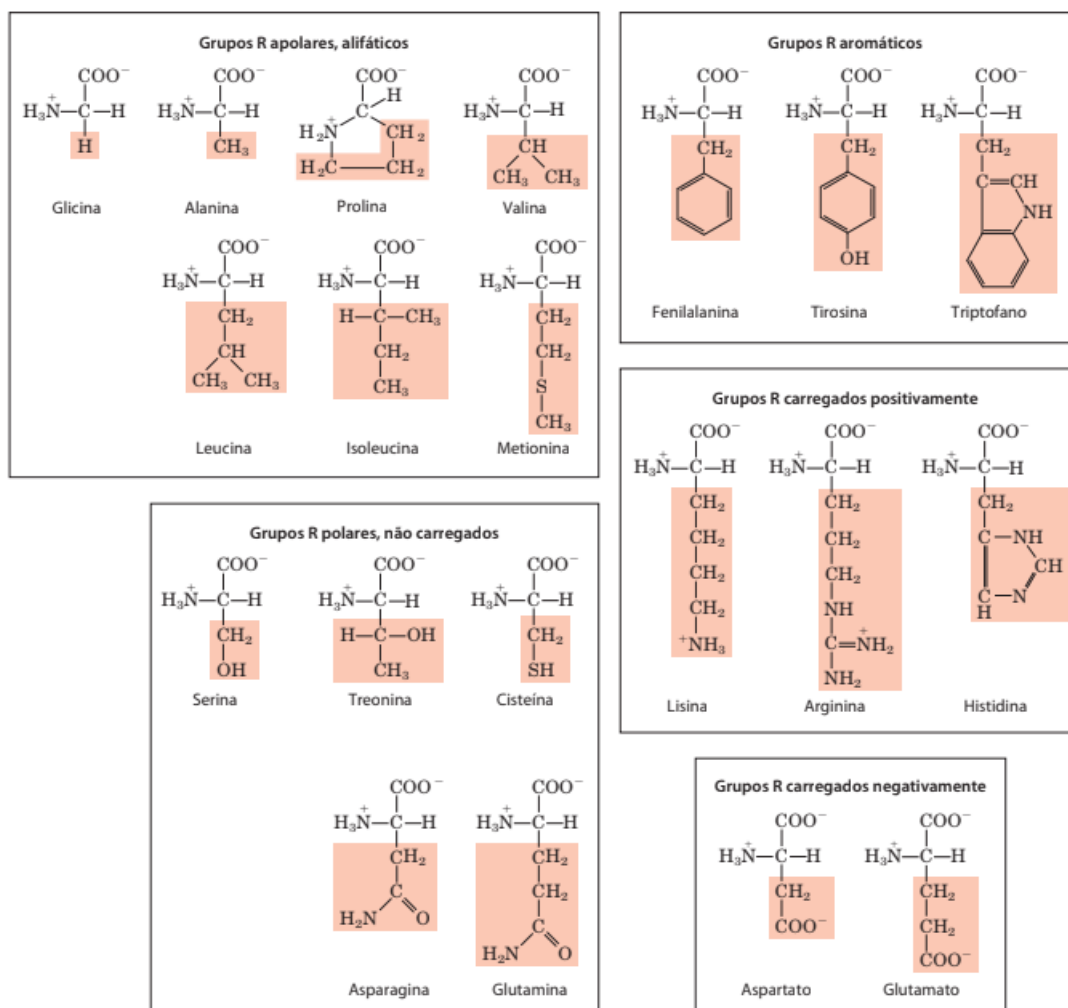


Fonte: Imagem obtida da referência (9)

O que diferencia um aminoácido de outro é a sua cadeia lateral, sendo responsável também pelas propriedades químicas distintas de cada um. Compreender essas propriedades é essencial para que seja possível entender o funcionamento das proteínas. Podemos, resumidamente, classificar os aminoácidos em cinco grupos com base em suas polaridades e tendência em interagir com a água em pH fisiológico (cerca de pH=7,0). A polaridade das cadeias varia amplamente, desde cadeias laterais apolares e hidrofóbicas (não hidrossolúvel) até cadeias polares e hidrofílicas (hidrossolúvel). Alguns aminoácidos são mais difíceis de caracterizar, não se encaixando perfeitamente em um desses grupos, é o caso da glicina, histidina e cisteína. Suas atribuições a alguns grupos vem de avaliações ponderadas e não absolutas, isto é, é feita uma análise comparativa com as propriedades de outros aminoácidos, ao invés de suas propriedades individuais (1, 9). Os 20 tipos de aminoácidos e suas classificações estão representados pela Figura 1.3.

- **Cadeias laterais apolares alifáticas:** As cadeias laterais neste grupo de aminoácidos são apolares e hidrofóbicas. Os aminoácidos alanina, valina, leucina e isoleucina tendem a ficar mais no interior da proteína, estabilizando sua conformação por interações hidrofóbicas. A glicina entra nesse grupo de aminoácidos, porém, por ser uma cadeia lateral relativamente simples, ela não contribui com as interações hidrofóbicas. A metionina, um dos dois aminoácidos que contém enxofre, possui um grupo tioéter, que é ligeiramente apolar. A prolina possui uma cadeia lateral alifática, sendo responsável por reduzir a flexibilidade de regiões peptídicas contendo esse aminoácido;

Figura 1.3: Os 20 aminoácidos encontrados em proteínas, classificados de acordo com a polaridade de suas cadeias laterais.



Fonte: Imagem obtida da referência (9).

- Cadeias laterais aromáticas:** como o nome sugere, são aminoácidos que possuem cadeias laterais compostas de anéis aromáticos, é o caso dos aminoácidos fenilalanina, tirosina e triptofano. Todas são relativamente apolares podendo participar de interações hidrofóbicas. A tirosina possui um grupo hidroxila na cadeia lateral e o triptofano possui um nitrogênio no anel aromático, o que os tornam mais polares do que a fenilalanina;
- Cadeias laterais polares não carregadas:** são aminoácidos relativamente mais polares e hidrofílicos em relação aos aminoácidos apolares. Isso se deve ao fato dos grupos funcionais das cadeias laterais desses aminoácidos formarem ligações de hidrogênio com água. Nessa classe de aminoácidos estão presentes a serina, treonina, cisteína, asparagina e glutamina. O que contribui para a polaridade dos aminoácidos serina e treonina são as hidroxilas de suas cadeias laterais. Para a asparagina e glutamina, o que vai contribuir para a suas polaridades é o grupo funcional amida. A cisteína aqui é um caso isolado, sua polaridade deriva do grupo sulfidrila, que é bastante modesta. A asparagina e a glutamina

tem sua polaridade derivada dos grupos aminos presentes em suas cadeias laterais;

- **Cadeias laterais polares carregadas positivamente (básicos):** a classe de aminoácidos mais hidrofílicos são aqueles cujas cadeias laterais podem doar ou receber prótons, dependendo do pH da solução e do pKa da cadeia lateral. Aminoácidos básicos, dependendo do pH podem protonar, adquirindo uma carga positiva. Essas propriedades decorrem dos grupos amino, guanidínio e imidazol aromático nas cadeias laterais da lisina, arginina e histidina, respectivamente;
- **Cadeias laterais carregadas negativamente (ácidos):** assim como no caso dos básicos, os aminoácidos ácidos possuem grupos capazes de desprotonar dependendo de seu pKa e do pH da solução, liberando prótons para a solução e adquirindo carga negativa. Isso ocorre com aspartato e o glutamato que possuem um grupo carboxila adicional em sua cadeia lateral.

Pelo fato das proteínas serem cadeias polipeptídicas muito extensas, existe um número muito grande de conformações tridimensionais possíveis. Dentre essas estruturas, poucas vão ser biologicamente ativas, as quais são chamadas de conformações nativas. Partindo de sua sequência específica de aminoácidos, essas biomoléculas formam estruturas tridimensionais estáveis para desempenhar uma função biológica específica (10). Devido a complexidade dessas conformações, tornou-se usual a definição de quatro níveis de estruturação para descrever como a proteína chega em sua conformação final (1, 9, 10):

- **Estrutura primária:** a chamada estrutura primária de uma proteína corresponde a ordem que os resíduos estão ligados, pois podemos ter duas proteínas com a mesma quantidade de aminoácidos e mesma composição, porém, distintas em relação a este ordenamento, sendo proteínas completamente diferentes em relação as suas funções biológicas. Essa é a primeira etapa unidimensional na especificação da estrutura tridimensional de uma proteína (10);
- **Estrutura secundária:** a estrutura secundária da proteína corresponde ao arranjo dos átomos do esqueleto da cadeia polipeptídica, isto é, o arranjo dos aminoácidos sem envolver suas cadeias laterais. Os tipos mais comuns de estrutura secundária são a α -hélice e a folha- β . As formas dessas estruturas são mantidas por ligações de hidrogênio entre o oxigênio do grupo carbonila de um aminoácido com o hidrogênio do grupo amino de outro aminoácido. O que muda entre essas duas estruturas são as posições dos átomos de cada aminoácido que participam das ligações de hidrogênio (10);
- **Estrutura terciária:** a estrutura terciária inclui o arranjo tridimensional de todos os átomos da proteína, incluindo os átomos das cadeias laterais ou grupos prostéticos (grupos de átomos não pertencentes a aminoácidos) (10);

- **Estrutura quartenária:** uma proteína pode consistir de múltiplas cadeias polipeptídicas, as quais cada uma dessas cadeias são chamadas de subunidades. O arranjo espacial dessas subunidades uma em relação a outra é a estrutura quartenária. Essas subunidades interagem por ligações não-covalentes, sendo por pontes de hidrogênio ou interações hidrofóbicas (10).

1.2 Proteínas Intrinsecamente Desordenadas

A pesquisa sobre proteínas se concentrou inicialmente em proteínas ordenadas, isto é, proteínas que adquirem uma conformação estável, seguindo os níveis de estruturação comentados anteriormente. Por consequência disso, os fundamentos para explicar a estruturação de uma proteína se baseou em proteínas ordenadas (11). Esses fundamentos tiveram como base a grande quantidade de estruturas obtidas experimentalmente no século 20 que foram armazenados em repositórios como PDB (*Protein Data Bank*) (12). Dessa forma não restavam dúvidas da relação entre a estrutura e a função das proteínas, o que levou ao paradigma da estrutura-função, em que uma proteína será biologicamente ativa se possuir uma estrutura tridimensional estável (13). As proteínas inicialmente eram consideradas moléculas biológicas rígidas, semelhantes a cristais, sem flexibilidade (13). Essa visão "estática" das proteínas era sustentada por uma das funções proteicas muito bem compreendida, que é a catálise (9, 13, 14). A visão de que as proteínas eram moléculas rígidas levou-se a supor que a complementaridade conformacional na interface das ligações e as forças não covalentes que surgem entre as moléculas são os principais fatores que promovem as interações envolvendo proteínas (15, 16). Apesar da riqueza de dados experimentais sobre as estruturas e funções das proteínas que contribuíram para o paradigma estrutura-função, houve um aumento de dados experimentais que desviavam desse paradigma, surgindo a necessidade de reavaliar esse modelo já bem estabelecido, o que por consequência fez surgir o paradigma desordem-função (13, 16).

Apesar do grande número de evidências de proteínas que não seguiam os níveis de estruturação comentados anteriormente, tudo era visto sob a luz do paradigma estrutura-função, de modo que os relatos de proteínas que apresentavam um comportamento diferente eram considerados exceções à regra (13, 16). Com a constante evidência de proteínas não tradicionais e uma intensa reavaliação do modelo clássico, proteínas não tradicionais foram consideradas como uma classe de proteínas, não sendo apenas uma exceção (13). Essas proteínas flexíveis desempenham seus papéis biológicos mesmo estando desordenadas (17). A própria flexibilidade conformacional dessas proteínas foi proposta como o fator principal para desempenhar suas funções biológicas (18). Ao longo dos anos, diversos termos foram usados para descrever essa classe de proteínas. Foi apenas recentemente que o termo "Proteínas Intrinsecamente Desordenadas" (do inglês *Intrinsically Disordered Protein* ou IDP) se destacou em relação aos outros termos utilizados (19, 20).

A propensão de uma proteína se ordenar ou permanecer desordenada depende da

sequência de aminoácidos que a compõem. Em outras palavras, o que vai definir se a proteína se enovela adquirindo uma conformação estável ou seja uma IDP, vai ser a sua composição de aminoácidos. Comparando a composição de proteínas ordenadas com proteínas desordenadas, verifica-se que as proteínas desordenadas são pobres em aminoácidos alifáticos (valina, leucina, isoleucina) e aromáticos (tirosina, fenilalanina, triptofano), e ricas em aminoácidos polares e carregados (metionina, lisina, arginina, serina, aspartato, glutamina) (18). Devido a abundância de aminoácidos polares e carregados em comparação com proteínas ordenadas, nas IDPs temos forças de repulsão mais intensas, além da carência de interações hidrofóbicas. Como as forças repulsivas são maiores do que as forças que promovem a compactação, a proteína tende a não ter uma conformação bem definida em relação às proteínas ordenadas (13).

Pelo fato das informações sobre IDPs terem sido muito escassas por muito tempo, parecia improvável que essas proteínas formassem uma classe de proteína tão abundante (13). Com o tempo, o acúmulo de dados experimentais permitiram que sofisticados algoritmos de bioinformática fossem feitos para prever a desordem das proteínas partindo de sua sequência de resíduos (13). Essas previsões mostraram que essas desordens estruturais estão presentes em todas as espécies (21). Pelo fato dessas proteínas serem mais flexíveis, elas participam de funções biológicas que envolvem o reconhecimento molecular, sendo relativamente mais abundantes em eucariotos em comparação com procariotos (22).

1.3 Interações Eletrostáticas em Proteínas

As interações eletrostáticas determinam as forças coesivas que unem átomos e moléculas em sólidos e líquidos. Desempenham um papel essencial em sistemas proteicos, envolvendo sua estrutura, mudanças conformacionais e interações com outras biomoléculas (23, 24). Verifica-se experimentalmente que essas interações são predominantes devido a forte dependência da variação da energia livre do sistema com o pH e a concentração de sal do meio em que estão inseridas (25–27).

As proteínas não possuem suas cargas fixas, e sua distribuição de carga depende de diversos fatores, sendo um deles, como comentado anteriormente, diz respeito a sua composição. Já que proteínas podem conter aminoácidos polares carregados, essas cargas vão depender do pH da solução, da concentração de sal do meio e interação com outras moléculas (23). Quando alguns fatores como mudança de concentração de sal, pHs próximos aos pKas dos resíduos ou proximidade com outras moléculas, as proteínas respondem modificando sua distribuição de carga, o que é conhecido como regulação de carga (*charge regulation* ou CR) (28). A origem desse mecanismo surge quando prótons migram de e para locais tituláveis da macromolécula, possibilitando que sua distribuição de carga se altere em resposta a alguma mudança no ambiente químico (28). Dessa forma ela se adapta a alguma perturbação no ambiente em que está inserida, sendo um processo muito importante ao se estudar interações entre duas proteínas (29) ou interações de proteínas com polieletrólitos (polímeros que adquirem cargas elétricas quando em

solução aquosa) (30). Um bom exemplo disso é a coacervação complexa, que corresponde a uma mudança de fase líquido-líquido de uma solução contendo polieletrólitos de cargas opostas, onde o CR influencia diretamente essa separação (23, 31).

A regulação de carga foi estudada inicialmente por Linderstrøm-Lang na década de 1920 (32–34) e posteriormente Kirkwood e Shumaker usando teoria de perturbação da mecânica estatística, mostraram que as correlações entre as distribuições de cargas entre duas proteínas favorecia interações atrativas entre elas (27), o que foi comprovado experimentalmente (35). Desde então diversos trabalhos foram desenvolvidos envolvendo mecanismo de regulação de carga nas interações proteína-proteína (36), proteína-superfície (30, 37, 38) e polieletrólitos-macroíons (39).

Pelo fato das IDPs serem ricas em resíduos polares carregados, elas geralmente são polianfóteros(1, 13). Embora essa classe de proteínas não possua uma conformação estável, elas podem adquirir certas conformações quando na presença de uma molécula alvo, o que vai definir essas conformações é a forma linear como seus resíduos estão distribuídos e a concentração de sal do meio, que é responsável por blindar as cargas dos aminoácidos da IDP (40–42). Em situações onde a distribuição de carga não é equilibrada, de modo que a carga líquida é diferente de zero, a proteína tem uma conformação estendida em baixa concentração de sal, devido a repulsão entre as cargas que predominam. Em concentrações altas de sal, essas interações são atenuadas devido à blindagem eletrostática e verifica-se uma forma mais compacta (43). Quando a carga líquida da proteína é zero, quanto mais separadas as cargas negativas das positivas, mais compacta a IDP se encontra (44). Ainda para carga líquida zero, quando temos distribuições em blocos com cargas positivas e blocos com cargas negativas, a proteína se compacta ganhando uma conformação semelhante a um "gancho" em baixas concentrações de sal, e para concentrações altas, essas interações são atenuadas e consequentemente a cadeia fica mais estendida (40).

As interações eletrostáticas possuem também um papel importante no processo de reconhecimento molecular das IDPs com uma molécula alvo, o que é verificado pelo fato dos aminoácidos participantes da ligação serem em sua maioria aminoácidos carregados, de modo que essas interações regulam a conformação da proteína quando na presença da molécula alvo (45). Embora seja conhecida a participação das interações eletrostáticas envolvida no processo de reconhecimento molecular, os detalhes desses mecanismos são em sua maioria desconhecidos (46). A regulação de carga levanta uma questão importante em relação a esses mecanismos, uma vez que tem o papel de alterar a distribuição de carga quando nas proximidades da molécula alvo, o que leva a uma determinada conformação que favorece a ligação com a molécula alvo (47).

Tendo em vista a importância das IDPs e o papel da regulação de carga nos processos envolvendo interações com uma molécula alvo, este trabalho tem como objetivo estudar os efeitos das interações eletrostáticas e particularmente da regulação de carga nas IDPs.

Capítulo 2

Objetivos

Este trabalho tem como objetivo principal combinar duas abordagens teóricas já existentes, sendo possível, por meio delas, observar os efeitos da regulação de carga em modelos representando IDPs. Vamos nos concentrar, inicialmente, em uma situação simplificada, onde a regulação de carga será estudada para um sistema contendo apenas a IDP em solução, sem considerar os efeitos do sal e de interações com outras macromoléculas carregadas. Dessa forma o efeito da regulação será observado somente pelas interações entre os resíduos tituláveis da proteína. Para alcançar esse objetivo, temos os seguintes desenvolvimentos:

- Fazer uma descrição fenomenológica e estatística da regulação de carga, introduzindo, de forma quantitativa, como esse fenômeno ocorre, de acordo com a referência (28);
- Estabelecer um formalismo geral que permite calcular as cargas, a distância quadrática média entre cada resíduo da proteína, e a distância *end-to-end* em unidades dessa grandeza para uma cadeia ideal para modelos de IDP já existentes via energia livre eletrostática do sistema;
- Utilizar o modelo proposto nas referências (48, 49) para calcular a energia livre eletrostática;

Capítulo 3

Regulação de Carga

3.1 Descrição Fenomenológica

Proteínas e uma variedade de outras macromoléculas contém resíduos ácidos e básicos que, dependendo das condições da solução podem protonar ou desprotonar (9), afetando sua distribuição de carga e interações com outras moléculas (28). A regulação de carga diz respeito a como o equilíbrio de protonação/desprotonação dos sítios tituláveis são afetados devido às interações entre seus próprios resíduos e/ou outras moléculas. Em outras palavras, a regulação de carga é uma resposta da proteína a alguma perturbação no ambiente químico (28).

Podemos descrever fenomenologicamente a regulação de carga analisando apenas um único sítio de ligação, que em condições de equilíbrio, pode ter a protonação/desprotonação representada pela seguinte reação química:



A constante de dissociação ácida é dada por:

$$K_d = \frac{\rho_{H^+} \rho_{A^-}}{\rho_{AH}} \frac{\gamma_{H^+} \gamma_{A^-}}{\gamma_{AH}}; \quad (3.2)$$

onde ρ representa as concentrações e γ são os coeficientes de atividade de cada participante da reação. Esse último leva em consideração as interações entre os solutos e solvente, introduzindo uma correção para o comportamento de soluções não ideais (50).

Tomando o logaritmo dos dois lados da equação 3.2 obtemos seguinte expressão:

$$\log K_d = \log (\rho_{H^+} \gamma_{H^+}) + \log \left(\frac{\rho_{A^-}}{\rho_{AH}} \right) + \log \left(\frac{\gamma_{A^-}}{\gamma_{AH}} \right),$$

identificando que $-\log K_d = pK_a$ e $-\log \rho_{H^+} = pH$, podemos reescrever a expressão anterior como:

$$-\log \left(\frac{\rho_{AH}}{\rho_{A^-}} \right) = pH - pK_a + \log \left(\frac{\gamma_{A^-}}{\gamma_{AH}} \right). \quad (3.3)$$

Para analisar o efeito da regulação de carga, vamos considerar o caso em que o $pH = pK_a$. Dessa forma, como a fração ρ_{AH}/ρ_{A^-} define a carga média do sítio, nessa situação ele vai depender apenas dos coeficientes de atividade, ou seja, vai ser afetado devido às interações entre seus resíduos ou com as outras moléculas do meio (28).

Considerando que nosso sistema seja composto apenas por uma proteína em solução, as interações ocorrem apenas entre seus resíduos. No equilíbrio térmico, o potencial químico $a_i = \rho_i \gamma_i$ será constante em toda a solução (51). Para que isso ocorra, quando o coeficiente de atividade muda, a concentração também muda, e como isso ocorre devido a proximidade entre os resíduos da proteína, podemos reescrever $\rho = \rho(R)$ e $\gamma = \gamma(R)$ onde R é a distância entre o sítio e outro resíduo da proteína (28).

3.1.1 Descrição Estatística

Anteriormente mostramos como a regulação de carga pode ser demonstrada utilizando argumentos termodinâmicos básicos. Considerando novamente um único sítio titulável, exposto a um potencial elétrico ϕ dado por todas as outras cargas do sistema e utilizando o formalismo do *ensemble* canônico (51), podemos expressar a valência média z do sítio por:

$$\langle z \rangle = \frac{\sum z_i \exp(-\beta e_0 z_i \phi)}{\sum \exp(-\beta e_0 z_i \phi)}, \quad (3.4)$$

onde $\beta = 1/k_B T$ é o inverso da energia térmica, com T sendo a temperatura absoluta, K_B a constante de Boltzmann e e_0 é a carga elementar.

Como vimos, qualquer alteração no ambiente químico afeta o potencial químico, e consequentemente a carga média do sítio. Dessa maneira, perturbando o potencial elétrico, a carga média também será alterada. Quantitativamente, podemos derivar a carga média em relação ao potencial ϕ , onde obtemos:

$$-\frac{1}{\beta e_0} \frac{\partial \langle z \rangle}{\partial \phi} = \langle z^2 \rangle - \langle z \rangle^2 \equiv c. \quad (3.5)$$

Aqui introduzimos a grandeza c , denominada capacitância (28, 52), que é simplesmente uma medida de quanta carga é induzida quando o sítio é exposto a um potencial externo (28).

Considerando uma solução ideal, onde os coeficientes de atividade são unitários, a variação da energia livre entre o estado protonado e desprotonado $\beta\epsilon$ pode ser escrita como (28, 53, 54):

$$\beta\epsilon = -\ln \left(\frac{\rho_{HA}}{\rho_{A^-}} \right), \quad (3.6)$$

a qual pode ser reescrita através da equação 3.3, considerando coeficientes de atividade unitários ($\gamma_{HA} = \gamma_{A^-} = 1$):

$$-\ln \left(\frac{\rho_{HA}}{\rho_{A^-}} \right) = \ln 10 (pH - pK_a), \quad (3.7)$$

resultando em

$$\beta\epsilon = \ln 10 (pH - pK_a). \quad (3.8)$$

Iremos utilizar a equação acima para descrever os dois estados possíveis para um único sítio titulável. Considerando como estado de referência o estado desprotonado, a energia será $\epsilon_0 = 0$. Assim a função de partição deste sítio pode ser escrita como:

$$Z = \sum_i e^{-\beta\epsilon_i} = 1 + e^{-(pH-pK_a)\ln 10}. \quad (3.9)$$

Para resíduos básicos, sendo a valência $z_0 = 0$ no estado desprotonado e $z_H = +1$ no estado protonado, a valência média do sítio é:

$$\langle z \rangle = \frac{z_0 + (z_0 + z_H) e^{-(pH-pK_a)\ln 10}}{Z}. \quad (3.10)$$

Dessa forma, vemos que

$$\langle z^2 \rangle = \frac{z_0^2 + (z_0 + z_H)^2 e^{-(pH-pK_a)\ln 10}}{Z}. \quad (3.11)$$

Assim, a variância é dada por

$$\langle z^2 \rangle - \langle z \rangle^2 = \frac{z_H^2 e^{-(pH-pK_a)\ln 10}}{Z^2}. \quad (3.12)$$

Variando a carga média com o pH, obtemos:

$$\frac{1}{\ln 10} \frac{\partial \langle z \rangle}{\partial pH} = \frac{z_H^2 e^{-(pH-pK_a)\ln 10}}{Z^2} = c. \quad (3.13)$$

Esse último resultado pode ser generalizado para a carga total de uma macromolécula, onde $Z = \sum z_i$. Assim, a capacitância total da macromolécula é:

$$C = \frac{1}{\ln 10} \frac{\partial \langle Z \rangle}{\partial pH} = \sum_i c_i. \quad (3.14)$$

3.1.2 Regulação de Carga Via Energia Livre

Vamos agora usar uma outra abordagem para formalizar a regulação de carga, partindo da energia livre envolvida nesse processo (55). Seja uma IDP contendo N resíduos io-

nizáveis com constante de dissociação $pKa^{(i)}$, com $i = 1, 2, \dots, N$. Os resíduos básicos, que pela protonação, adquirem uma carga positiva, que iremos escrever como $q_k^+ = e_0 \eta_k$ onde $\eta_k \in [0, 1]$, os resíduos ácidos, pela desprotonação, adquirem uma carga $q_k^- = e_0 (\eta_k - 1)$ e novamente $\eta_k \in [0, 1]$, e_0 é a carga elementar e η_k corresponde a fração de resíduos protonados (52, 56), ou seja, o número de resíduos protonados dividido pelo número total de resíduos. Consequentemente $(1 - \eta_k)$ é a probabilidade de encontrarmos o sítio desprotonado. A energia livre total para esse sistema pode ser expressa por:

$$F_{tot} = F_{CR} + F, \quad (3.15)$$

onde F_{CR} corresponde a energia livre de Helmholtz envolvida na regulação de carga e F a energia livre do polímero considerando as interações eletrostáticas, os graus de liberdade do polímero e as interações de volume excluído (48, 49).

Vamos inicialmente determinar a energia livre envolvida na regulação de carga, isto é, a energia livre envolvida nos processos de protonação/desprotonação da proteína. Essa energia pode ser escrita como (51):

$$F_{CR} = U - TS. \quad (3.16)$$

O primeiro termo da equação 3.16 corresponde a energia envolvida na associação/dissociação de prótons, o segundo termo é a entropia desse processo (51). Para entropia vamos usar a equação de Boltzmann (51):

$$S = k_B \ln \Omega, \quad (3.17)$$

onde k_B corresponde a constante de Boltzmann e Ω o número de maneiras de distribuir os sítios protonados em N resíduos.

Ao longo do nosso modelo de IDP, vamos considerar que dentre os N sítios tituláveis, temos N_p protonados e N_d desprotonados, de forma que $N = N_p + N_d$. Dessa forma, podemos escrever Ω como:

$$\Omega = \frac{N!}{N_p! (N - N_p)!}. \quad (3.18)$$

Substituindo a equação 3.18 na 3.17 a entropia é escrita como:

$$S = k_B \ln \frac{N!}{N_p! (N - N_p)!}. \quad (3.19)$$

Podemos eliminar os fatoriais utilizando a aproximação de Stirling (51, 57). Embora seja válida para números pequenos (da ordem da dezena), tendo um erro pequeno (na ordem de 0,01 %) essa aproximação se torna mais precisa para valores grandes de n e exata quando

$n \rightarrow \infty$. Sendo expressa como:

$$\ln n! \approx n \ln n - n.$$

Assim, usando essa aproximação na equação 3.19, obtemos:

$$S = -k_B \left[N_p \ln \frac{N_p}{N} + N_d \ln \left(1 - \frac{N_p}{N} \right) \right]. \quad (3.20)$$

Dividindo a equação 3.20 por N , identificando que $\eta_k = N_p/N$ e $1 - \eta_k = N_d/N$, obtemos a entropia por resíduo:

$$S_k = -k_B [\eta_k \ln \eta_k + (1 - \eta_k) \ln (1 - \eta_k)]. \quad (3.21)$$

Sendo a energia do estado protonado de um único sítio igual a ϵ , considerando o estado desprotonado com energia nula. Como temos N_p sítios protonados, a energia total é:

$$U = N_p \epsilon. \quad (3.22)$$

Dividindo essa última equação por N , obtemos a energia por monômero, expressa por:

$$U_k = \eta_k \epsilon. \quad (3.23)$$

Substituindo as equações 3.23 e 3.21 na equação 3.16, obtemos a contribuição da energia livre referente à regulação de carga:

$$F_{CR} = \sum_k \epsilon \eta_k + \frac{1}{\beta} \sum_k [\eta_k \ln \eta_k + (1 - \eta_k) \ln (1 - \eta_k)]. \quad (3.24)$$

A situação de equilíbrio é atingida quando a energia livre é minimizada em relação a fração de sítios carregados. Sendo assim, vamos minimizar a energia livre total com relação ao η_k utilizando a equação de Euler-Lagrange (58, 59). Como F_{tot} não tem dependência com as derivadas de η_k , obtemos que

$$\frac{\partial F}{\partial \eta_k} + \epsilon + \frac{1}{\beta} \ln \frac{\eta_k}{1 - \eta_k} = 0. \quad (3.25)$$

Partindo da equação 3.25 e identificando que $\beta \epsilon = \ln 10 (pH - pK_a^{(k)})$, onde aqui introduzimos um índice k para nos referirmos ao k -ésimo resíduo, podemos expressar η_k como:

$$\eta_k = \left[1 + \exp \left(\ln 10 (pH - pK_a^{(k)}) + \beta \frac{\partial F}{\partial \eta_k} \right) \right]^{-1}. \quad (3.26)$$

Com esses resultados conseguimos obter uma expressão para a carga de cada resíduo da proteína:

$$q_k^\pm = \frac{\pm e_0}{1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(k)} \right) \pm \beta \frac{\partial F}{\partial \eta_k} \right)}. \quad (3.27)$$

A equação 3.27 depende da derivada de F que será dada pelo modelo da proteína e contém as interações eletrostáticas entre os resíduos. Então, para determinar o valor das cargas é necessário inicialmente determinarmos a energia livre da proteína.

Determinando $\frac{\partial F}{\partial \eta_k}$ e inserindo na equação 3.27 obtemos um sistema de equações não lineares para as cargas q_k . Uma alternativa a isto, é utilizar uma aproximação linear (60) que nos permite obter um sistema de equações lineares.

3.1.3 Aproximação Linear para a Regulação de Carga

A aproximação de Debye-Hückel é amplamente utilizada na literatura, pois possibilita a obtenção de equações diferenciais analiticamente solúveis. Isso a torna especialmente atraente nos estudos de polieletrólitos (61, 62). Neste trabalho utilizamos essa aproximação e comparamos com a situação onde ela não é empregada, com o objetivo de estudarmos suas limitações. Nesta aproximação podemos simplificar a equação 3.27 (60), onde consideramos que o termo $\partial F / \partial \eta_k$ é muito pequeno quando comparado com a energia térmica $k_B T = \beta^{-1}$. Isso nos permite expandir a expressão 3.27 em uma série de potências em $\beta \partial F / \partial \eta_k$ e desprezar os termos não lineares. Assim, expandindo $q_i^\pm$ obtemos:

$$q_i^\pm = \frac{\pm e_0}{1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right)} - \beta e_0 \frac{\exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right)}{\left[1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right) \right]^2} \frac{\partial F}{\partial \eta_i}. \quad (3.28)$$

Para simplificar a última equação vamos definir as seguintes igualdades:

$$q_i^{(0)} = \frac{\pm e_0}{1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right)}, \quad (3.29)$$

$$c_i^{(0)} = \frac{1}{e_0 \ln 10} \frac{\partial q_i^{(0)}}{\partial pH} = \frac{\exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right)}{\left[1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right) \right]^2}, \quad (3.30)$$

onde a equação 3.29 corresponde a equação 3.27 na ausência de interações eletrostáticas entre os resíduos, e a equação 3.30 é a capacitância também na ausência dessas interações (52). Dessa forma, reescrevendo a equação 3.28 obtemos:

$$q_i = q_i^{(0)} - \beta e_0 c_i^{(0)} \frac{\partial F}{\partial \eta_i}, \quad (3.31)$$

Partindo das equações 3.31 e 3.27, podemos determinar as cargas de um polímero

considerando o efeito da regulação de carga, o que pode ser feito se determinarmos a energia livre da proteína.

Capítulo 4

Modelo Físico da Proteína

Vamos representar a IDP com base na sua estrutura linear, considerando que seu comprimento total seja muito maior do que as distâncias entre seus resíduos. Isso nos permite tratar a proteína como uma curva contínua (63, 64), na qual s corresponde a uma variável de integração ao longo do contorno dessa curva. O modelo segue uma abordagem *coarse grained*, no qual iremos considerar as interações entre cada resíduo sem levar em consideração seus detalhes interatômicos, isto é, não vamos considerar as interações entre cada átomo que compõe cada resíduo, de modo que cada resíduo seja representado apenas pelo seu pKa (49, 65).

Em relação as interações, vamos considerar que os resíduos interagem com um potencial harmônico; interações que surgem devido ao volume que os resíduos ocupam no espaço e as interações eletrostáticas entre resíduos carregados (63, 64). Podemos representar essas interações com o seguinte hamiltoniano (48, 49, 63):

$$\begin{aligned} \beta H_t = & \frac{3}{2l} \int_0^L ds \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 + l \int_0^L ds \int_0^s ds' \omega(s, s') \delta[\mathbf{R}(s) - \mathbf{R}(s')] \\ & + \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') \frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|}, \end{aligned} \quad (4.1)$$

onde s é uma variável que estamos integrando ao longo do contorno da cadeia, L corresponde ao comprimento total da cadeia, l denota o comprimento de Kunh (63), sendo um comprimento limite no qual a partir dele as orientações das ligações se tornam decorrelacionadas (64).

O primeiro termo da equação 4.1 representa as interações elásticas de cada ligação com $\mathbf{R}(s)$ sendo o vetor de posição do segmento s da cadeia. O segundo termo é a interação de volume excluído, com $\omega(s, s')$ sendo uma grandeza adimensional chamada parâmetro de interação de volume excluído entre os resíduos representados pelos segmentos s e s' , correspondendo as interações não eletrostáticas entre os monômeros e solvente (63, 64). Por fim, temos o terceiro termo que corresponde às interações eletrostáticas, seguindo a teoria de Debye-Huckel, onde cada resíduo possui uma carga $q(s)$ com $q(s) = \eta_s e_0$ para cargas positivas e

$q(s) = (\eta_s - 1) e_0$ para cargas negativas, em que η_s representa a probabilidade de encontrarmos o resíduo positivamente carregado, portanto $\eta_s \in [0, 1]$, onde e_0 sendo a carga elementar. Ainda no terceiro termo

$$l_b = e_0^2 / (4\pi\epsilon_0\epsilon k_B T)$$

é o comprimento de Bjerrum, ϵ_0 é a constante dielétrica no vácuo e ϵ a constante dielétrica do meio em que a proteína está inserida, κ é o parâmetro de Debye, e seu inverso $\lambda = \kappa^{-1} = (8\pi l_b c_b)^{-1/2}$ é o comprimento de Debye, no qual c_b é a concentração de sal do meio (66). Para a água temos que $l_b = 7,2 \text{ \AA}$.

Para determinar o tamanho médio da cadeia, uma abordagem variacional foi utilizada para simplificarmos a equação 4.1, utilizando o que é chamado de comprimento de Kunh renormalizado ou efetivo l_r (48, 67). Nessa simplificação, como mostraremos a seguir, as interações eletrostáticas e de volume excluído são consideradas fracas. Dessa forma, o polímero se comporta como uma cadeia ideal com comprimento de Kunh l_r e o hamiltoniano renormalizado é expresso de uma forma simplificada:

$$\beta H_r = \frac{3}{2l_r} \int_0^L ds \left(\frac{d\mathbf{R}(s)}{ds} \right)^2. \quad (4.2)$$

A média de algum observável físico A , pode ser escrita em termos do hamiltoniano renormalizado utilizando um método perturbativo. Nesse método vamos considerar que o hamiltoniano original (H_t) seja uma perturbação do hamiltoniano renormalizado (H_r), de maneira que podemos reescreve-lo como:

$$H_t = H_r + (H_t - H_r). \quad (4.3)$$

Como a média é feita sobre todas as configurações, ela é expressa em termos de integrais funcionais (67):

$$\langle A \rangle = \frac{\int \delta\mathbf{R} A e^{-\beta H_r - v}}{\int \delta\mathbf{R} e^{-\beta H_r - v}}, \quad (4.4)$$

na qual $\delta\mathbf{R}$ indica que a integral está sendo feito no espaço de funções, ou seja, a integração está sendo feita sobre todas as configurações acessíveis pela IDP. Para simplificar foi denotamos $v = \beta (H_r - H_t)$. Podemos reescrever $\langle A \rangle$ em relação ao hamiltoniano renormalizado, a qual vamos denotar por $\langle \dots \rangle_r$. Para fazer isso, inicialmente vamos expandir e^{-v} em uma série de potencias, obtêm-se:

$$\langle A \rangle = \frac{\int \delta\mathbf{R} A e^{-\beta H_r} [1 - v + \frac{1}{2}v^2 - \dots]}{\int \delta\mathbf{R} e^{-\beta H_r} [1 - v + \frac{1}{2}v^2 - \dots]}, \quad (4.5)$$

multiplicando e dividindo numerador e denominador pela função de partição renormalizada Z_r e identificando as médias de cada grandeza, obtemos:

$$\langle A \rangle = \frac{Z_r [\langle A \rangle_r - \langle Av \rangle_r + \frac{1}{2} \langle Av^2 \rangle_r - \dots]}{Z_r [1 - \langle v \rangle_r + \frac{1}{2} \langle v^2 \rangle_r - \dots]}. \quad (4.6)$$

Podemos inverter o denominador da equação 4.6 identificando uma série geométrica $1/(1-r)$, onde aqui $r = \langle v \rangle_r - \frac{1}{2} \langle v^2 \rangle_r + \dots$. Dessa forma obtém-se:

$$\begin{aligned} \langle A \rangle &= \left[\langle A \rangle_r - \langle Av \rangle_r + \frac{1}{2} \langle Av^2 \rangle_r - \dots \right] \left[1 + \langle v \rangle_r + \langle v \rangle_r^2 - \frac{1}{2} \langle v^2 \rangle_r - \dots \right] \\ &= \langle A \rangle_r - (\langle Av \rangle_r - \langle A \rangle_r \langle v \rangle_r) + \frac{1}{2} (\langle Av^2 \rangle_r - 2 \langle Av \rangle_r \langle v \rangle_r \\ &\quad + 2 \langle A \rangle_r \langle v \rangle_r^2 - \langle A \rangle_r \langle v^2 \rangle_r) + \dots, \end{aligned} \quad (4.7)$$

como estamos considerando as interações como fracas, os termos de ordem superior na expansão são desprezíveis, restando apenas os dois primeiros termos ao lado direito da equação 4.7. Vamos considerar também que $\langle A \rangle_r \approx \langle A \rangle$ (49), e dessa maneira obtém-se:

$$\langle A \rangle_r \langle (H_t - H_r) \rangle = \langle A (H_t - H_r) \rangle. \quad (4.8)$$

A escolha de A depende da quantidade de interesse, no nosso caso vamos determinar a distância quadrática média entre dois segmentos ao longo do contorno da proteína, nas posições $\mathbf{R}(s_1)$ e $\mathbf{R}(s_2)$, o que irá nos permitir calcular a distância quadrática média entre dois resíduos.

$$A = (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \quad (4.9)$$

Partindo da equação 4.8, reescrevendo essa igualdade de forma mais explicita com $A = [\mathbf{R}(s_2) - \mathbf{R}(s_1)]^2$, obtemos:

$$\begin{aligned} &\left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int_0^L ds \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 \right\rangle_r \\ &- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int_0^L ds \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 \right\rangle_r \\ &= \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 l \int_0^L ds \int_0^s ds' \omega(s, s') \delta[\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\ &- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle l \int_0^L ds \int_0^s ds' \omega(s, s') \delta[\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\ &+ \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') \frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|} \right\rangle_r \\ &- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') \frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|} \right\rangle_r. \end{aligned} \quad (4.10)$$

Podemos simplificar o lado esquerdo da igualdade em 4.10 através da definição de $\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r$, onde a média é feita sobre todas as configurações possíveis da proteína, no qual é expressa por

$$\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r = \frac{\int \delta \mathbf{R} \left((\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \exp \left(-\frac{3}{2l_r} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right) \right)}{\int \delta \mathbf{R} \exp \left(-\frac{3}{2l_r} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}, \quad (4.11)$$

Afim de se obter uma expressão simplificada, vamos fazer $u = \frac{3}{2}l_r$ em 4.11, de modo que obtemos:

$$\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r = \frac{\int \delta \mathbf{R} (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}{\int \delta \mathbf{R} \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}, \quad (4.12)$$

e então derivando em u , obtemos:

$$\begin{aligned} \frac{d}{du} \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r &= \frac{\int \delta \mathbf{R} (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \left(\frac{9}{4u^2} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right) \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}{\int \delta \mathbf{R} \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)} \\ &- \frac{\int \delta \mathbf{R} \left(\frac{9}{4u^2} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right) \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}{\int \delta \mathbf{R} \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)} \\ &\times \frac{\int \delta \mathbf{R} (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}{\int \delta \mathbf{R} \exp \left(-\frac{9}{4u} \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right)}. \end{aligned} \quad (4.13)$$

Multiplicando os dois lados da equação 4.13 por $\frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right)$ e identificando a média de cada grandeza, podemos reescrever a equação 4.13 como

$$\begin{aligned}
& \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \frac{d}{du} \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \\
&= \frac{1}{l_r^2} \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right\rangle_r \\
&- \frac{1}{l_r^2} \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle \left\langle \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right\rangle_r, \tag{4.14}
\end{aligned}$$

na qual, das propriedades de uma cadeia gaussiana, temos que $\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r = (s_2 - s_1) l_r$ (63, 64). Substituindo na equação 4.14, obtemos:

$$\begin{aligned}
& \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right\rangle_r \\
&- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle \frac{3}{2} \left(\frac{1}{l_r} - \frac{1}{l} \right) \int \left(\frac{d\mathbf{R}(s)}{ds} \right)^2 ds \right\rangle_r \tag{4.15}
\end{aligned}$$

$$= \left(\frac{1}{l_r} - \frac{1}{l} \right) (s_2 - s_1) l_r^2. \tag{4.16}$$

Para as interações de volume excluído, o que corresponde ao primeiro termo do lado direito da equação 4.10 podemos simplificá-la usando a transformada de Fourier para a função delta de Dirac (68, 69). Dessa forma, teremos:

$$\delta^3(\mathbf{r}) = \int \frac{d^3k}{(2\pi)^3} \exp i\mathbf{k} \cdot \mathbf{r}. \tag{4.17}$$

Substituindo a equação 4.17 no segundo termo da equação 4.10, obtemos:

$$\begin{aligned}
& \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 l \int_0^L ds \int_0^s ds' \omega(s, s') \delta[\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
&- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle l \int_0^L ds \int_0^s ds' \omega(s, s') \delta[\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
&= \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3k}{(2\pi)^3} \exp(i\mathbf{k} \cdot [\mathbf{R}(s) - \mathbf{R}(s')]) \right\rangle_r \\
&- \langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \rangle_r \left\langle l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3k}{(2\pi)^3} \exp(i\mathbf{k} \cdot [\mathbf{R}(s) - \mathbf{R}(s')]) \right\rangle_r. \tag{4.18}
\end{aligned}$$

Essa última igualdade pode ser reescrita como (67):

$$\begin{aligned}
& \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3 k}{(2\pi)^3} \exp i \mathbf{k} [\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
& - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3 k}{(2\pi)^3} \exp i \mathbf{k} [\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
& = -\frac{l_r^2}{9} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' \omega(s, s') (s - s_1)^2 \int \frac{d^3 k k^2}{(2\pi)^3} \exp \left(-\frac{k^2 l_r |s - s'|}{6} \right) \right. \\
& + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' \omega(s, s') (s - s')^2 \int \frac{d^3 k k^2}{(2\pi)^3} \exp \left(-\frac{k^2 l_r |s - s'|}{6} \right) \\
& + \int_{s_2}^L ds \int_0^{s_1} ds' \omega(s, s') (s_2 - s_1)^2 \int \frac{d^3 k k^2}{(2\pi)^3} \exp \left(-\frac{k^2 l_r |s - s'|}{6} \right) \\
& \left. + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' \omega(s, s') (s_2 - s')^2 \int \frac{d^3 k k^2}{(2\pi)^3} \exp \left(-\frac{k^2 l_r |s - s'|}{6} \right) \right], \tag{4.19}
\end{aligned}$$

em que a integral em k na equação 4.19 pode ser resolvida analiticamente. Fazendo $A = \frac{l_r |s - s'|}{6}$, temos:

$$\begin{aligned}
\int \frac{d^3 k k^2}{(2\pi)^3} \exp(-A k^2) &= \frac{1}{(2\pi)^3} \left[\int \int \int dk_x dk_y dk_z k_x^2 \exp(-(k_x^2 + k_y^2 + k_z^2) A) \right. \\
&+ \int \int \int dk_x dk_y dk_z k_y^2 \exp(-(k_x^2 + k_y^2 + k_z^2) A) \\
&\left. + \int \int \int dk_x dk_y dk_z k_z^2 \exp(-(k_x^2 + k_y^2 + k_z^2) A) \right], \tag{4.20}
\end{aligned}$$

onde podemos notar que as integrais em k_x , k_y e k_z são integrais gaussianas, de modo que resolvendo essas integrais para cada componente, obtemos:

$$\int \frac{d^3 k k^2}{(2\pi)^3} \exp \left(-\frac{k^2 l_r |s - s'|}{6} \right) = 3 \left(\frac{1}{2\pi} \right)^{3/2} \left(\frac{3}{l_r |s - s'|} \right)^{5/2}. \tag{4.21}$$

Substituindo esse último resultado na equação 4.19, temos:

$$\begin{aligned}
& \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3 k}{(2\pi)^3} \exp i\mathbf{k} [\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
& - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle l \int_0^L ds \int_0^s ds' \omega(s, s') \int \frac{d^3 k}{(2\pi)^3} \exp i\mathbf{k} [\mathbf{R}(s) - \mathbf{R}(s')] \right\rangle_r \\
& = -\frac{l_r^2}{3} \left(\frac{1}{2\pi} \right)^{3/2} \left(\frac{3}{l_r} \right)^{5/2} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' \omega(s, s') \frac{(s - s_1)^2}{(s - s')^{5/2}} \right. \\
& + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' \omega(s, s') (s - s')^{-1/2} \\
& + \int_{s_2}^L ds \int_0^{s_1} ds' \omega(s, s') \frac{(s_2 - s_1)^2}{(s - s')^{5/2}} \\
& \left. + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' \omega(s, s') \frac{(s_2 - s')^2}{(s - s')^{5/2}} \right]. \tag{4.22}
\end{aligned}$$

Vamos agora resolver para os termos contendo as interações eletrostáticas entre os resíduos, o que corresponde ao segundo termo do lado direito da equação 4.10. Inicialmente, vamos fazer

$$\frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|} \equiv \frac{\exp(-\kappa r)}{r}, \tag{4.23}$$

em que definimos r como a distância entre os resíduos. Podemos demonstrar que essa última igualdade corresponde a transformada inversa de Fourier da função (69):

$$G(k) = \frac{4\pi}{(k^2 + \kappa^2)}.$$

Sendo assim, fazendo a transformada inversa de $G(k)$ temos:

$$g(r) = 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2}. \tag{4.24}$$

Reescrevendo em coordenadas esféricas temos:

$$g(r) = 4\pi \int_0^\infty \int_0^\pi \int_0^{2\pi} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} k^2 \sin \theta dk d\phi, \tag{4.25}$$

que ao fazermos uma mudança de variável $u = \cos \theta$, obtemos:

$$g(r) = \frac{1}{2\pi^2} \left(4\pi \int_0^\infty \frac{k}{k^2 + \kappa^2} \left(\frac{\sin kr}{r} \right) dk \right), \tag{4.26}$$

onde a integral entre parenteses na última equação pode ser calculada por uma integração de contorno (69), pelo qual obtemos:

$$g(r) = \frac{\exp(-\kappa r)}{r}. \quad (4.27)$$

Dessa maneira, podemos reescrever o terceiro termo do lado direito da equação 4.10 como

$$\begin{aligned} & \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') \frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|} \right\rangle_r \\ & - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') \frac{\exp(-\kappa |\mathbf{R}(s) - \mathbf{R}(s')|)}{|\mathbf{R}(s) - \mathbf{R}(s')|} \right\rangle_r \\ & = \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r \\ & - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r. \end{aligned} \quad (4.28)$$

Podemos agora utilizar um método semelhante ao que foi feito entre o primeiro termo do lado direito da equação 4.18 até a 4.22 para mostrar que

$$\begin{aligned} & \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r \\ & - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r \\ & = -\frac{4\pi l_b l_r^2}{l^2} \frac{1}{9} \\ & \times \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' q(s) q(s') (s - s_1)^2 \int \frac{d^3 k k^2}{(k^2 + \kappa^2)} \exp\left(-\frac{k^2 l_r |s - s'|}{6}\right) \right. \\ & + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' q(s) q(s') (s - s')^2 \int \frac{d^3 k k^2}{(k^2 + \kappa^2)} \exp\left(-\frac{k^2 l_r |s - s'|}{6}\right) \\ & + \int_{s_2}^L ds \int_0^{s_1} ds' q(s) q(s') (s_2 - s_1)^2 \int \frac{d^3 k k^2}{(k^2 + \kappa^2)} \exp\left(-\frac{k^2 l_r |s - s'|}{6}\right) \\ & \left. + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' q(s) q(s') (s_2 - s')^2 \int \frac{d^3 k k^2}{(k^2 + \kappa^2)} \exp\left(-\frac{k^2 l_r |s - s'|}{6}\right) \right], \end{aligned} \quad (4.29)$$

para a última integral podemos usar a seguinte igualdade (70):

$$\begin{aligned} \int \frac{d^3 k k^2}{(k^2 + \kappa^2)} \exp\left(-\frac{k^2 l_r |s - s'|}{6}\right) & = 4\pi \left[\frac{\pi^{1/2}}{4} \left(\frac{6}{l_r |s - s'|}\right)^{3/2} \right. \\ & \quad - \frac{\pi^{1/2}}{2} \kappa^2 \left(\frac{6}{l_r |s - s'|}\right)^{1/2} \\ & \quad \left. + \frac{\pi}{2} \kappa^3 \exp\left(\frac{\kappa^2 l_r |s - s'|}{6}\right) \operatorname{erfc}\left(\sqrt{\frac{\kappa^2 l_r |s - s'|}{6}}\right) \right]. \end{aligned} \quad (4.30)$$

Com a finalidade de escrever uma expressão mais compacta, vamos reescrever o termo entre colchetes da equação 4.30 como:

$$C(s, s') = \left[\frac{\pi^{1/2}}{4} \left(\frac{6}{l_r |s - s'|} \right)^{3/2} - \frac{\pi^{1/2}}{2} \kappa^2 \left(\frac{6}{l_r |s - s'|} \right)^{1/2} + \frac{\pi}{2} \kappa^3 \exp \left(\frac{\kappa^2 l_r |s - s'|}{6} \right) \operatorname{erfc} \left(\sqrt{\frac{\kappa^2 l_r |s - s'|}{6}} \right) \right] \quad (4.31)$$

Dessa forma podemos reescrever a equação 4.29 como:

$$\begin{aligned} & \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r \\ & - \left\langle (\mathbf{R}(s_2) - \mathbf{R}(s_1))^2 \right\rangle_r \left\langle \frac{l_b}{l^2} \int_0^L ds \int_0^s ds' q(s) q(s') 4\pi \int \frac{d^3 k}{(2\pi)^3} \frac{\exp(i\mathbf{k} \cdot \mathbf{r})}{k^2 + \kappa^2} \right\rangle_r \\ & = -\frac{(4\pi)^2 l_b l_r^2}{l^2 9} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' q(s) q(s') (s - s_1)^2 C(s, s') \right. \\ & + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' q(s) q(s') (s - s')^2 C(s, s') \\ & + \int_{s_2}^L ds \int_0^{s_1} ds' q(s) q(s') (s_2 - s_1)^2 C(s, s') \\ & \left. + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' q(s) q(s') (s_2 - s')^2 C(s, s') \right], \quad (4.32) \end{aligned}$$

Substituindo 4.32, 4.22 e 4.15 na equação 4.10, obtemos uma igualdade que nos permite determinar a distância quadrática média entre cada par de resíduos ao longo da IDP:

$$\begin{aligned}
& \left(\frac{1}{l} - \frac{1}{l_r} \right) (s_2 - s_1) l_r^2 \\
&= -\frac{l_r^2}{3} \left(\frac{1}{2\pi} \right)^{3/2} \left(\frac{3}{l_r} \right)^{5/2} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' \omega(s, s') \frac{(s - s_1)^2}{(s - s')^{5/2}} \right. \\
&+ \int_{s_1}^{s_2} ds \int_{s_1}^s ds' \omega(s, s') (s - s')^{-1/2} \\
&+ \int_{s_2}^L ds \int_0^{s_1} ds' \omega(s, s') \frac{(s_2 - s_1)^2}{(s - s')^{5/2}} \\
&+ \left. \int_{s_2}^L ds \int_{s_1}^{s_2} ds' \omega(s, s') \frac{(s_2 - s')^2}{(s - s')^{5/2}} \right] \\
&- \frac{(4\pi)^2 l_b l_r^2}{l^2 9} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' q(s) q(s') (s - s_1)^2 C(s, s') \right. \\
&+ \int_{s_1}^{s_2} ds \int_{s_1}^s ds' q(s) q(s') (s - s')^2 C(s, s') \\
&+ \int_{s_2}^L ds \int_0^{s_1} ds' q(s) q(s') (s_2 - s_1)^2 C(s, s') \\
&+ \left. \int_{s_2}^L ds \int_{s_1}^{s_2} ds' q(s) q(s') (s_2 - s')^2 C(s, s') \right]. \tag{4.33}
\end{aligned}$$

Vamos agora mudar para uma variável discreta, onde as integrais se tornam somatórios de modo que $ds \rightarrow \Delta s$. Fazendo $s = ml$, $s' = nl$, $s_1 = jl$, $s_2 = il$ e $x_{i,j} = l_{r,ij}/l$, na qual m, n, i e $j \in N$. Dessa maneira podemos reescrever a equação 4.33 como:

$$\begin{aligned}
& -\frac{l_r^2}{3} \left(\frac{1}{2\pi}\right)^{3/2} \left(\frac{3}{l_r}\right)^{5/2} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' \omega(s, s') \frac{(s-s_1)^2}{(s-s')^{5/2}} \right. \\
& + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' \omega(s, s') (s-s')^{-1/2} \\
& + \int_{s_2}^L ds \int_0^{s_1} ds' \omega(s, s') \frac{(s_2-s_1)^2}{(s-s')^{5/2}} \\
& \left. + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' \omega(s, s') \frac{(s_2-s')^2}{(s-s')^{5/2}} \right] \\
& \rightarrow -\frac{l_r^2 l^{-1/2}}{3} \left(\frac{1}{2\pi}\right)^{3/2} \left(\frac{3}{l_r}\right)^{5/2} \left[\sum_{m=j}^i \sum_{n=1}^{j-1} \omega_{m,n} \frac{(m-j)^2}{(m-n)^{5/2}} \right. \\
& + \sum_{m=j+1}^i \sum_{n=j}^{m-1} \omega_{m,n} (m-n)^{-1/2} \\
& + \sum_{m=i+1}^N \sum_{n=1}^{j-1} \omega_{m,n} \frac{(i-j)^2}{(m-n)^{5/2}} \\
& \left. + \sum_{m=i+1}^N \sum_{n=j}^i \omega_{m,n} \frac{(i-n)^2}{(m-n)^{5/2}} \right] \tag{4.34}
\end{aligned}$$

e

$$\begin{aligned}
& -\frac{(4\pi)^2 l_b l_r^2}{l^2 9} \left[\int_{s_1}^{s_2} ds \int_0^{s_1} ds' q(s) q(s') (s-s_1)^2 C(s, s') \right. \\
& + \int_{s_1}^{s_2} ds \int_{s_1}^s ds' q(s) q(s') (s-s')^2 C(s, s') \\
& + \int_{s_2}^L ds \int_0^{s_1} ds' q(s) q(s') (s_2-s_1)^2 C(s, s') \\
& + \int_{s_2}^L ds \int_{s_1}^{s_2} ds' q(s) q(s') (s_2-s')^2 C(s, s') \\
& \rightarrow -\frac{(4\pi)^2 l_b l_r^2}{l^2 9} \left[\sum_{m=j}^i \sum_{n=1}^{j-1} q_m q_n C(m, n, x_{i,j}, \kappa l) (m-j)^2 \right. \\
& + \sum_{m=j+1}^i \sum_{n=j}^{m-1} q_m q_n C(m, n, x_{i,j}, \kappa l) (m-n)^2 \\
& + \sum_{m=i+1}^N \sum_{n=1}^{j-1} q_m q_n C(m, n, x_{i,j}, \kappa l) (i-j)^2 \\
& \left. + \sum_{m=i+1}^N \sum_{n=j}^i q_m q_n C(m, n, x_{i,j}, \kappa l) (i-n)^2 \right]. \tag{4.35}
\end{aligned}$$

Substituindo as equações 4.34 e 4.35 na equação 4.33 e rearranjando os termos, obtemos a seguinte expressão:

$$x_{i,j}^{3/2} \left(1 - \frac{1}{x_{i,j}}\right) = \left(\frac{3}{2\pi}\right)^{3/2} \frac{\Omega_{i,j}}{x_{i,j}} + \frac{1}{9} \frac{4\pi l_b}{l} \frac{2}{(2\pi)^2} Q_{i,j}, \quad (4.36)$$

no qual

$$\begin{aligned} \Omega_{i,j} = & \frac{1}{(i-j)} \left[\sum_{m=j}^i \sum_{n=1}^{j-1} \omega_{m,n} \frac{(m-j)^2}{(m-n)^{5/2}} \right. \\ & + \sum_{m=j+1}^i \sum_{n=j}^{m-1} \omega_{m,n} (m-n)^{-1/2} \\ & + \sum_{m=i+1}^N \sum_{n=1}^{j-1} \omega_{m,n} \frac{(i-j)^2}{(m-n)^{5/2}} \\ & \left. + \sum_{m=i+1}^N \sum_{n=j}^i \omega_{m,n} \frac{(i-n)^2}{(m-n)^{5/2}} \right], \end{aligned} \quad (4.37)$$

$$\begin{aligned} Q_{i,j} = & \frac{1}{(i-j)} \left[\sum_{m=j}^i \sum_{n=1}^{j-1} q_m q_n B(m, n, x_{i,j}, \kappa l) (m-j)^2 \right. \\ & + \sum_{m=j+1}^i \sum_{n=j}^{m-1} q_m q_n B(m, n, x_{i,j}, \kappa l) (m-n)^2 \\ & + \sum_{m=i+1}^N \sum_{n=1}^{j-1} q_m q_n B(m, n, x_{i,j}, \kappa l) (i-j)^2 \\ & \left. + \sum_{m=i+1}^N \sum_{n=j}^i q_m q_n B(m, n, x_{i,j}, \kappa l) (i-n)^2 \right], \end{aligned} \quad (4.38)$$

e

$$\begin{aligned} B(x_{i,j}, m, n, \kappa l) = & x_{i,j}^{3/2} \left[\frac{\pi^{1/2}}{4} \left(\frac{6}{x_{i,j}}\right)^{3/2} \frac{1}{(m-n)^{3/2}} \right. \\ & - \frac{\pi^{1/2}}{2} (l\kappa)^2 \left(\frac{6}{x_{i,j}}\right)^{1/2} \frac{1}{(m-n)^{1/2}} \\ & \left. + \frac{\pi}{2} (l\kappa)^3 \exp\left(\frac{(l\kappa)^2 (m-n)}{6}\right) \operatorname{erfc}\left(\sqrt{\frac{(l\kappa)^2 (m-n)}{6}}\right) \right]. \end{aligned} \quad (4.39)$$

Para condições em que não consideramos íons na solução ($\kappa = 0$), $Q_{i,j}$ é reescrito como:

$$\begin{aligned}
Q_{i,j} = & \frac{(6\pi)^{3/2}}{4\pi} \frac{1}{(i-j)} \left[\sum_{m=j}^i \sum_{n=1}^{j-1} q_m q_n \frac{(m-j)^2}{(m-n)^{3/2}} \right. \\
& + \sum_{m=j+1}^i \sum_{n=j}^{m-1} q_m q_n (m-n)^{1/2} \\
& + \sum_{m=i+1}^N \sum_{n=1}^{j-1} q_m q_n \frac{(i-j)^2}{(m-n)^{3/2}} \\
& \left. + \sum_{m=i+1}^N \sum_{n=j}^i q_m q_n \frac{(i-n)^2}{(m-n)^{3/2}} \right]. \tag{4.40}
\end{aligned}$$

Dessa forma a relação para $x_{i,j}$, dada pela equação 4.36, é reescrita como:

$$x_{i,j}^{3/2} \left(1 - \frac{1}{x_{i,j}} \right) = \left(\frac{3}{2\pi} \right)^{3/2} \frac{\Omega_{i,j}}{x_{i,j}} + \frac{l_b}{l} \sqrt{\frac{2}{3\pi}} Q_{i,j}. \tag{4.41}$$

Determinando os valores de $x_{i,j}$ para cada par de resíduos com a equação 4.41, podemos determinar a distância média quadrática entre quaisquer pares de resíduos ao longo da proteína por meio da seguinte equação, retiradas das referências (49, 65):

$$\langle R_{ij}^2 \rangle = |i-j| l^2 x_{i,j}. \tag{4.42}$$

Uma outra métrica importante para a conformação espacial da IDP é a distância média quadrática entre as extremidades da proteína, conhecida como distância *end-to-end* (63). Um caso especial da equação 4.41 nos permite determinar essa grandeza em unidades da distância *end-to-end* para uma cadeia livre de interações, que é obtida fazendo $i = N$, $j = 1$ e considerando $N - 1 \approx N$ (válido para cadeias grandes). Dessa forma $x = \langle R_{ee}^2 \rangle / \langle R_{ee,f}^2 \rangle$, em que R_{ee} é a distância *end-to-end* para a proteína de interesse e $R_{ee,f}$ essa grandeza para uma cadeia livre de interações. Com essas alterações a equação 4.41 é expressa por:

$$x^{3/2} \left(1 - \frac{1}{x} \right) = \left(\frac{3}{2\pi} \right)^{3/2} \frac{\Omega}{x} + \frac{l_b}{l} \sqrt{\frac{2}{3\pi}} Q, \tag{4.43}$$

e

$$\Omega = \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} \omega(m, n) \frac{1}{(m-n)^{1/2}}, \tag{4.44}$$

$$Q = \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} q(m) q(n) (m-n)^{1/2}. \tag{4.45}$$

A energia livre para a IDP em termos de x pode ser expressa como (49, 65):

$$\beta F(x) = \frac{3}{2}(x - \ln x) + \left(\frac{3}{2\pi}\right)^{3/2} \frac{\Omega}{x^{3/2}} + \frac{l_b}{l} \frac{Q}{x^{1/2}} \sqrt{\frac{6}{\pi}}. \quad (4.46)$$

Através dessa energia livre podemos determinar as cargas de cada resíduo considerando o mecanismo da regulação de carga, utilizando a equação 3.31.

4.1 Determinação das Cargas da Proteína

Para determinarmos as cargas de cada resíduo considerando o efeito da regulação de carga, de acordo com as equações 3.27 e 3.31 devemos determinar a derivada de 4.46 em relação a η_i . Partindo do caso mais simples, isto é, no limite de concentração nula de sal ($\kappa = 0$), temos:

$$\beta \frac{\partial F(x, \Omega, Q)}{\partial \eta_i} = \beta \left(\frac{\partial F}{\partial x} \frac{\partial x}{\partial \eta_i} + \frac{\partial F}{\partial Q} \frac{\partial Q}{\partial \eta_i} \right). \quad (4.47)$$

Podemos determinar $\beta \partial F / \partial x$ a partir da equação 4.46, pelo qual obtemos:

$$\beta \frac{\partial F}{\partial x} = \frac{3}{2}(1 - x^{-1}) - \frac{3}{2} \left(\frac{3}{2\pi}\right)^{3/2} \Omega x^{-5/2} + \frac{l_b}{l} \sqrt{\frac{6}{\pi}} \left(\frac{\partial Q}{\partial x} x^{-1/2} - \frac{1}{2} x^{-3/2} Q \right), \quad (4.48)$$

e partindo da equação equação 4.43 podemos determinar $\partial x / \partial \eta_i$ e $\partial Q / \partial x$:

$$\frac{\partial x}{\partial \eta_i} = \frac{\frac{e_0 l_b}{l} \sqrt{\frac{2}{3\pi}}}{\frac{3}{2} x^{1/2} - \frac{1}{2} x^{-1/2} + \left(\frac{3}{2\pi}\right)^{3/2} \Omega x^{-2}} \left(\frac{1}{N} \sum_{k=0}^N \sqrt{|k-i|} q_k \right), \quad (4.49)$$

$$\frac{l_b}{l} \sqrt{\frac{2}{3\pi}} \frac{\partial Q}{\partial x} = \frac{3}{2} x^{1/2} - \frac{1}{2} x^{-1/2} + \left(\frac{3}{2\pi}\right)^{3/2} \Omega x^{-2}. \quad (4.50)$$

Fazendo as devidas substituições, obtemos o primeiro termo ao lado direito da equação 4.47, expressa por:

$$\begin{aligned} \beta \frac{\partial F}{\partial x} \frac{\partial x}{\partial \eta_i} &= \left[6 - 3x^{-1} + \frac{3}{2} \left(\frac{3}{2\pi}\right)^{3/2} \Omega x^{-5/2} - \frac{l_b}{2l} \sqrt{\frac{6}{\pi}} x^{-3/2} Q \right] \\ &\times \left[\frac{3}{2} x^{1/2} - \frac{1}{2} x^{-1/2} + \left(\frac{3}{2\pi}\right)^{3/2} \Omega x^{-2} \right]^{-1} \frac{l_b e_0}{Nl} \sqrt{\frac{2}{3\pi}} \sum_{k=1}^N q_k \sqrt{|i-k|}. \end{aligned} \quad (4.51)$$

Procedendo de forma análoga, determinamos o segundo termo do lado direito da equação 4.47, onde:

$$\begin{aligned}
\beta \frac{\partial F}{\partial Q} \frac{\partial Q}{\partial \eta_i} &= \left[\frac{3}{2} - \frac{3}{2}x^{-1} - \frac{3}{2} \left(\frac{3}{2\pi} \right)^{3/2} \Omega x^{-5/2} - \frac{l_b}{2l} \sqrt{\frac{6}{\pi}} x^{-3/2} Q \right] \\
&\times \left[\frac{3}{2}x^{1/2} - \frac{1}{2}x^{-1/2} + \left(\frac{3}{2\pi} \right)^{3/2} \Omega x^{-2} \right]^{-1} \frac{l_b e_0}{Nl} \sqrt{\frac{2}{3\pi}} \sum_{k=1}^N q_k \sqrt{|i-k|} \\
&+ \frac{l_b e_0}{Nl} \sqrt{\frac{6}{\pi}} x^{-1/2} \sum_{k=1}^N q_k \sqrt{|i-k|}.
\end{aligned} \tag{4.52}$$

Substituindo as duas últimas equações na equação 4.47 e multiplicando os dois lados da equação por e_0 obtemos a seguinte igualdade:

$$\begin{aligned}
e_0 \beta \frac{\partial F}{\partial \eta_i} &= \left[\left(\frac{15}{2} - \frac{9}{2}x^{-1} - \frac{l_b}{l} \sqrt{\frac{6}{\pi}} x^{-3/2} Q \right) \left(\frac{3}{2}x^{1/2} - \frac{1}{2}x^{-1/2} + \left(\frac{3}{2\pi} \right)^{3/2} \Omega x^{-2} \right)^{-1} + 3x^{-1/2} \right] \\
&\times \frac{l_b e_0^2}{Nl} \sqrt{\frac{6}{\pi}} \sum_{k=1}^N q_k \sqrt{|i-k|}.
\end{aligned} \tag{4.53}$$

Para obtermos uma expressão mais compacta, vamos definir

$$\begin{aligned}
\xi_{ik} &= \left[\left(\frac{15}{2} - \frac{9}{2}x^{-1} - \frac{l_b}{l} \sqrt{\frac{6}{\pi}} x^{-3/2} Q \right) \left(\frac{3}{2}x^{1/2} - \frac{1}{2}x^{-1/2} + \left(\frac{3}{2\pi} \right)^{3/2} \Omega x^{-2} \right)^{-1} + 3x^{-1/2} \right] \\
&\times \frac{l_b e_0^2}{Nl} \sqrt{\frac{6}{\pi}} \sqrt{|i-k|}.
\end{aligned} \tag{4.54}$$

Dessa forma podemos reescrever a equação 4.47 como

$$e_0 \beta \frac{\partial F}{\partial \eta_i} = \sum_{k=1}^N \xi_{ik} q_k. \tag{4.55}$$

Com esses resultados podemos encontrar as cargas de cada resíduo da proteína para a situação de concentração nula de sal de duas formas distintas:

- substituindo 4.55 na equação 3.31 que surge de uma aproximação linear, e resolver o sistema de equações lineares resultando:

$$q_i^{(0)} = \sum_{k=1}^N \left(\delta_{ik} + c_i^{(0)} \xi_{ik} \right) q_k; \tag{4.56}$$

- substituindo 4.55 na equação 3.27 e resolver o sistema de equações não lineares:

$$q_i^\pm = \frac{\pm e_0}{1 + \exp\left(\pm \ln 10 \left(pH - pK_a^{(i)}\right)\right) \pm \sum_{k=1}^N \xi_{ik} q_k}. \quad (4.57)$$

Para resolver esses sistemas, foi desenvolvido um algoritmo computacional que consiste em um método iterativo. Nesse método inicialmente substituímos as cargas q_k pelas cargas q_k^0 dada pela equação 3.29 que depende apenas do pH da solução e do pK_a de cada resíduo. Com isso obtemos um valor para q_i . Esse processo se repete até que ocorra uma convergência, isto é, até que as diferenças entre as cargas de duas iterações consecutivas seja menor do que 10^{-6} . No próximo capítulo esses processos serão apresentados com mais detalhes.

Com os parâmetros obtidos, podemos determinar os valores das cargas médias de cada resíduo das IDPs. Para calcular as cargas utilizamos a equação 4.56 para o caso onde foi usada a aproximação DH (Debye Hückel) e a equação 4.57 para o na situação onde essa aproximação não é utilizada. O primeiro problema que encontramos ao tentar resolver essas equações (ou sistema de equações) é que as cargas estão correlacionadas, ou seja, para determinar uma única carga é necessário ter o valores de todas as demais cargas da IDP. Uma forma numérica de lidar com esse problema é utilizando um método autoconsistente (71). Nesse método a equação é resolvida assumindo uma solução inicial, que no caso correspondem as cargas na ausência da regulação de carga, em seguida, as novas soluções obtidas são inseridas novamente na equação, fazendo esse processo iterativo até que ocorre uma convergência. O fluxograma representado pela Figura 4.1 a seguir representa os algoritmos utilizados para obter as cargas:

- **Inserir parâmetros:** esses valores correspondem ao pK_a^i de cada resíduo da IDP, valor do pH utilizado, razão entre o comprimento de Bjerrum e o de Kuhn e o valor médio do parâmetro de volume excluído;
- **Calcular as cargas e capacitâncias na ausência de regulação de carga:** para a carga foi utilizada a equação:

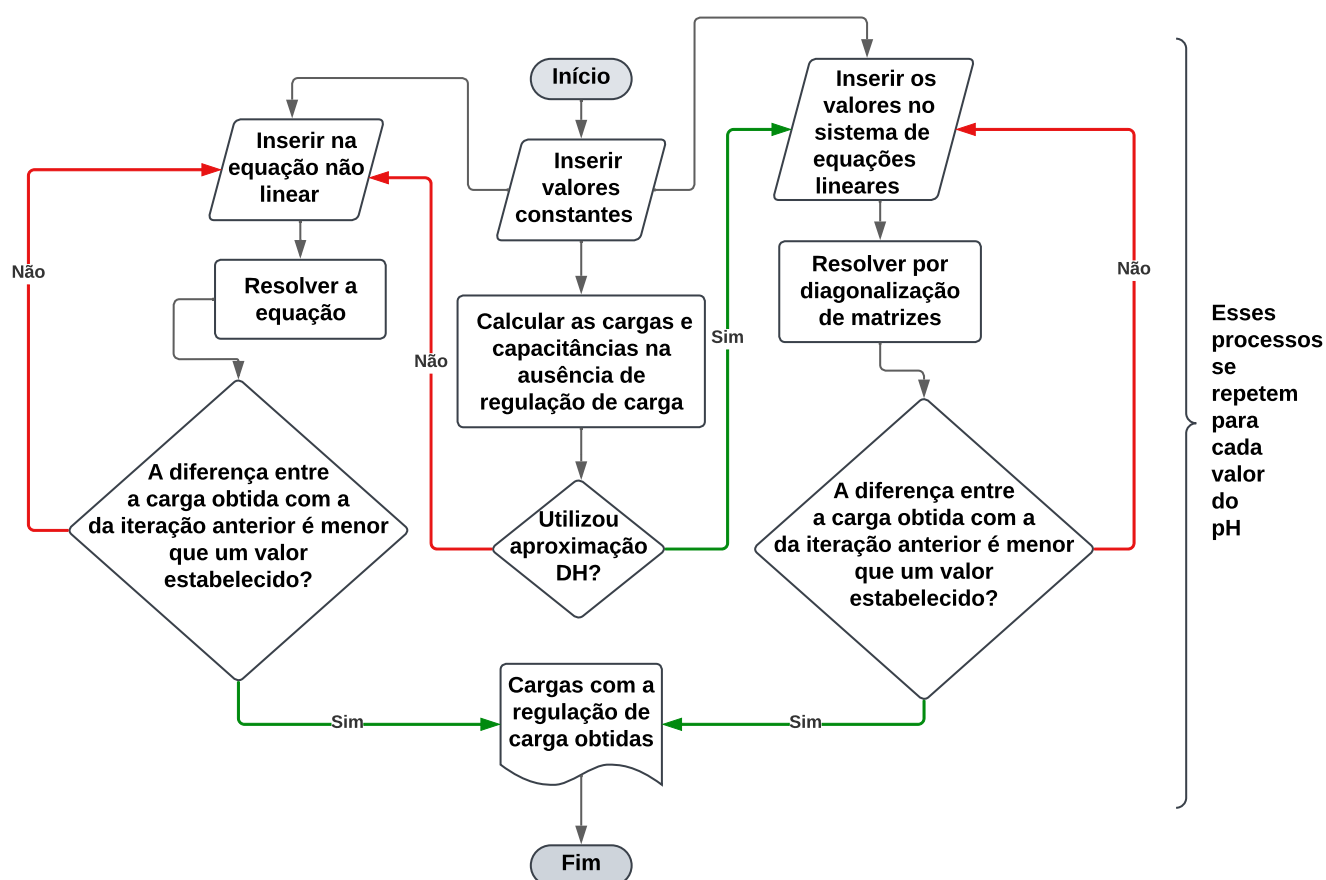
$$q_i^{(0)} = \frac{\pm e_0}{1 + \exp\left(\pm \ln 10 \left(pH - pK_a^{(i)}\right)\right)},$$

e para a capacitância:

$$c_i^{(0)} = \frac{\exp\left(\pm \ln 10 \left(pH - pK_a^{(i)}\right)\right)}{\left[1 + \exp\left(\pm \ln 10 \left(pH - pK_a^{(i)}\right)\right)\right]^2}.$$

- **Utilizar a equação correspondente:** como visto anteriormente, foram usadas duas abordagens: utilizando a aproximação DH e sem essa aproximação. A aproximação de DH resulta no seguinte sistema de equações:

Figura 4.1: Fluxograma representando os algoritmos computacionais utilizados para determinar a carga de cada resíduo das proteínas e para cada pH utilizado.



Fonte: Elaborado pelo autor.

$$q_i^{(0)} = \sum_{k=1}^N \left(\delta_{ik} + c_i^{(0)} \xi_{ik} \right) q_k,$$

em que esse sistema é resolvido através da diagonalização de matriz. Para o caso sem a aproximação, a seguinte equação foi utilizada e resolvida:

$$q_i^{\pm} = \frac{\pm e_0}{1 + \exp \left(\pm \ln 10 \left(pH - pK_a^{(i)} \right) \right) \pm \sum_{k=1}^N \xi_{ik} q_k}.$$

Em ambos os casos temos:

$$\begin{aligned} \xi_{ik} = & \left[\left(\frac{15}{2} - \frac{9}{2}x^{-1} - \frac{l_b}{l} \sqrt{\frac{6}{\pi}} x^{-3/2} Q \right) \left(\frac{3}{2}x^{1/2} - \frac{1}{2}x^{-1/2} + \left(\frac{3}{2\pi} \right)^{3/2} \Omega x^{-2} \right)^{-1} + 3x^{-1/2} \right] \\ & \times \frac{l_b e_0^2}{Nl} \sqrt{\frac{6}{\pi}} \sqrt{|i-k|}, \end{aligned}$$

onde podemos observar explicitamente onde os valores constantes são inseridos;

- **Processo iterativo para a convergência das cargas:** a partir da solução inicial, as novas cargas são comparadas com o valor da iteração anterior, de modo que se a diferença entre elas for menor do que um valor estabelecido (10^{-6}) significa que as cargas convergiram, e caso contrário ela é reinserida na equação corresponde e o processo se repete até que a condição de parada seja obtida.

Esses passos foram seguidos para todos os valores de pH utilizados. As cargas obtidas no final do processo representam a carga média de cada sítio com a regulação de carga, e nesse trabalho é o ponto de partida para determinar as demais grandezas de interesse.

Aplicação Para Modelos de IDPs

Com a finalidade de tornar mais evidente o mecanismo da regulação de carga, foram idealizados 3 modelos distintos de IDPs, cada um contendo 50 resíduos no total e mesma constituição, isto é, todos contendo 25 resíduos ácidos e o restante por resíduos básicos. Os resíduos ácidos foram considerados ácidos glutâmicos (Glu), cujo pK_a desse resíduo isolado é $pK_a = 4,25$ (1), enquanto os básicos, as lisinas (Lys), no qual seu pK_a isolado vale $pK_a = 10,79$ (1). A única diferença entre esses três modelos foi a forma como os resíduos estão dispostos ao longo da cadeia. Os três modelos estão representados na Figura 5.1.

Além disso, vamos considerar um sistema simplificado, no qual as IDPs vão estar em uma solução aquosa, na ausência de sal ($\kappa = 0$) e o comprimento de Bjerrum neste caso é $l_b = 7.2 \text{ \AA}$.

Os cálculos desenvolvidos nos permitem fazer análises comparativas entre os modelos, e verificar o efeito da regulação de carga. Para fazer essas comparações, determinamos a distribuição de carga e, conseqüentemente, a carga total das proteínas. Isso foi feito considerando a aproximação de Debye-Hückel, através da equação 4.56 e sem essa aproximação, através da equação 4.57. Ambas as equações 4.56 e 4.57, têm dependência com os parâmetros de volume excluído ($\omega_{m,n}$), que neste modelo foram considerados independente dos pares de resíduos (m,n), sendo substituído por um valor médio $\langle\omega\rangle$, apresentados na Tabela 5.1, retirados da referência

Figura 5.1: Três modelos de IDPs, onde E são ácidos glutâmicos e K são as lisinas. (a) Monômeros dispostos de forma alternada entre E e K, (b) monômeros dispostos de forma aleatória, ou seja, sua distribuição não segue um padrão definido, (c) monômeros dispostos em dois grandes blocos, o primeiro bloco contendo apenas E e o outro bloco composto por K.



Fonte: Imagem retirada da referência (49).

Tabela 5.1: Tabela contendo os valores médios do parâmetro de volume excluído de cada sequência de aminoácidos utilizados.

$\langle\omega\rangle$	Sequência
3.83	Alternados
3.4	Trechos Alternados
3.07	Blocos

Fonte: Elaborada pelo autor, com os dados fornecidos pelo artigo da referência (48).

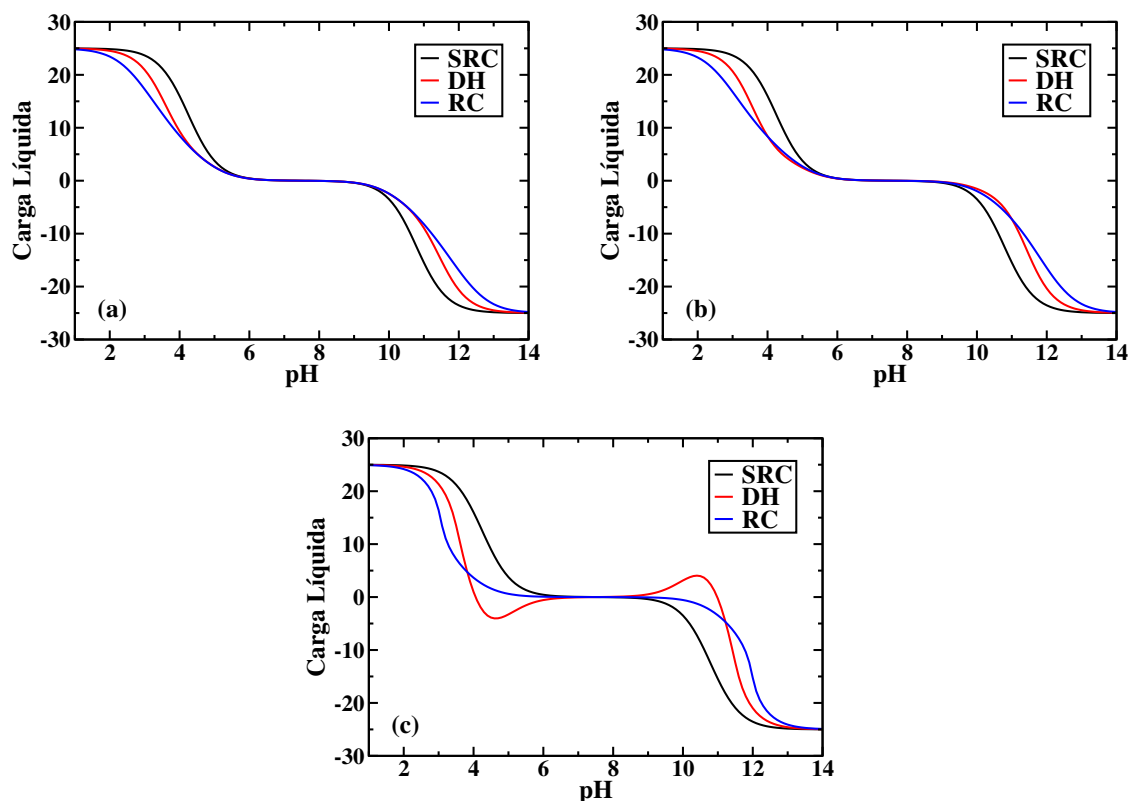
(48). Além disso, Das e Pappu estimaram o valor de $l = 5.8 \text{ \AA}$ para o comprimento de Kunh através de simulações computacionais (40, 48).

A Figura 5.2 (a) apresenta a carga líquida da IDP em função do pH cujos resíduos estão dispostos de forma alternada, isto é, cada resíduo de lisina terá um resíduo de ácido glutâmico ao lado, como mostra a Figura 5.1. Podemos notar uma diferença entre as curvas da carga líquida com a regulação de carga e a curva na ausência desse efeito. Essa diferença se torna mais evidente em pontos em que o pH é muito próximo do valor do pKa de cada resíduo. Para valores baixos de pH (entre 1 a 5), menores do que o pKa da lisina ($pKa = 10,79$), a carga total da proteína é positiva em virtude da protonação dos resíduos básicos. Pela forma como os monômeros estão dispostos, a carga positiva dos resíduos básicos adjacentes vão diminuir a probabilidade de protonação dos aminoácidos ácidos, fazendo com que esses sítios tenham maior probabilidade de estarem desprotonados, resultando em uma carga líquida menor para a proteína. Quando o pH fica próximo ao pKa da lisina, uma situação análoga ocorre, só que agora a carga líquida da proteína é negativa, devido a maioria dos sítios ácidos estarem desprotonados e negativamente carregados. Dessa forma, é gerada uma região negativamente carregada em torno do sítio básico devido a os resíduos ácidos adjacentes, favorecendo sua protonação. O efeito geral disso é uma carga líquida menos negativa. Além disso, observa-se uma diferença sutil entre as curvas que representam os resultados obtidos com a aproximação de Debye-Hückel 4.56 e sem essa aproximação 4.57, mostrando a sua validade para essa distribuição.

Na Figura 5.2 (b) apresenta o efeito da regulação de carga para o modelo representando uma IDP com seus resíduos dispostos em trechos alternados. Nesse modelo os monômeros estão dispostos como apresentado na Figura 5.1(b), não seguindo um padrão específico como no caso anterior. Nessa distribuição, os resíduos ácidos e básicos alternam em pequenos trechos com tamanhos distintos ao longo da cadeia. Desse gráfico, podemos notar uma semelhança muito grande com o caso anterior, na qual consideramos uma distribuição alternada. Essa semelhança se deve ao fato de que, apesar da distribuição não seguir uma ordem específica, ela possui apenas pequenos trechos compostos de resíduos do mesmo tipo. Além disso, verificamos que a aproximação de Debye-Hückel é adequada.

A última distribuição que analisamos o efeito da regulação de carga está representada na Figura 5.1(c). Nessa distribuição, os resíduos estão dispostos em dois grandes blocos, com os 25 primeiros resíduos compostos por ácidos glutâmicos e o restante por lisinas. Quando o

Figura 5.2: Gráficos de carga líquida por pH para os três modelos de IDPs propostos: distribuição alternada dos resíduos (a); distribuição em trechos alternados (b); e distribuição em blocos (c). A curva preta representa a carga total sem a regulação de carga (SRC); A curva vermelha é a carga total considerando a regulação de carga com a aproximação de Debye-Hückel (DH); A curva em azul consideramos a regulação de carga mas sem essa aproximação.

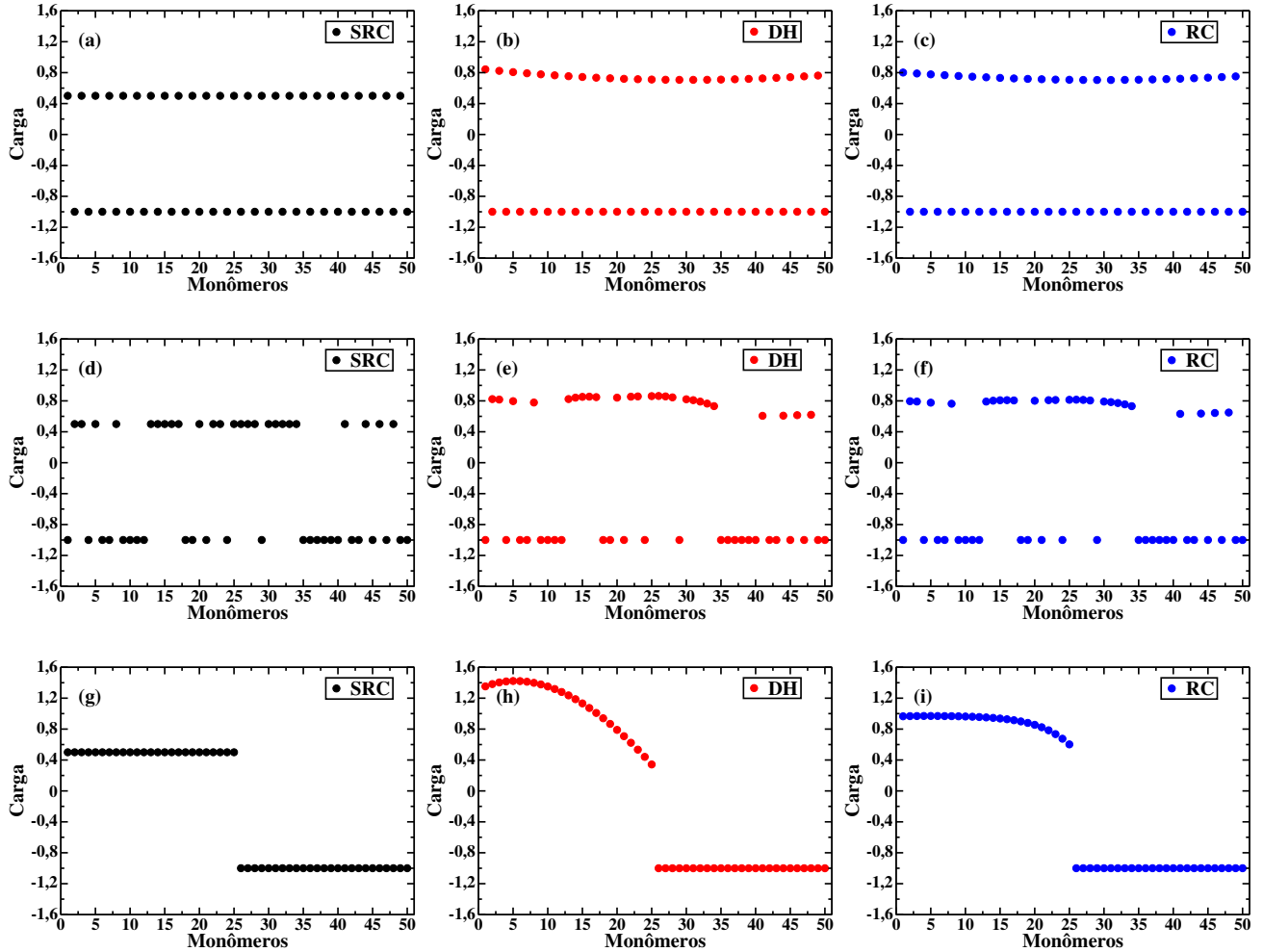


Fonte: Elaborado pelo autor.

pH está próximo do pK_a do ácido glutâmico ($pK_a = 4,25$), observa-se uma variação muito maior entre as curvas da carga total sem a regulação e as curvas onde a regulação de carga é considerada, a ponto da carga total ser quase nula na faixa onde o pH é próximo ao pK_a do ácido glutâmico ($pK_a = 4,75$). Algo semelhante é observado quando o pH é próximo ao pK_a da lisina ($pK_a = 10,79$). O que pode explicar essa diferença é a própria conformação da IDP, pois, geralmente, para uma distribuição de carga desse tipo a proteína tende a se dobrar, uma vez que o bloco carregado positivamente atrai o bloco com carga negativa, diminuindo muito a probabilidade de protonação dos sítios da cadeia. Como consequência disso observa-se uma carga total, em módulo, muito menor quando comparado com as outras distribuições estudadas. Nota-se também que a aproximação linear não é válida para essa distribuição pelo fato da proximidade entre os monômeros ser maior, resultando em interações eletrostáticas mais intensas.

A Figura 5.3 mostra o efeito da regulação para a carga média de cada resíduo nos três modelos de IDPs propostos em $pH = 10,79$, que corresponde a um pH em que a regulação de carga é mais evidente, como visto anteriormente. Aqui também foram utilizadas as duas abordagens desenvolvidas, isto é, a aproximação de Debye-Hückel (DH) (pontos vermelhos) e

Figura 5.3: Gráficos mostrando a carga de cada resíduo em pH=10,79 para os três modelos de IDPs propostos. Sem regulação de carga (SRC), com a regulação de carga e utilizando a aproximação de Debye-Hückel (DH), e com a regulação de carga, porém sem utilizar essa aproximação (RC). Os gráficos (a), (b) e (c) representam a distribuição alternada dos resíduos, os gráficos (d), (e) e (f) a distribuição em trechos alternados e (g), (h) e (i) a distribuição em blocos ao longo da proteína.

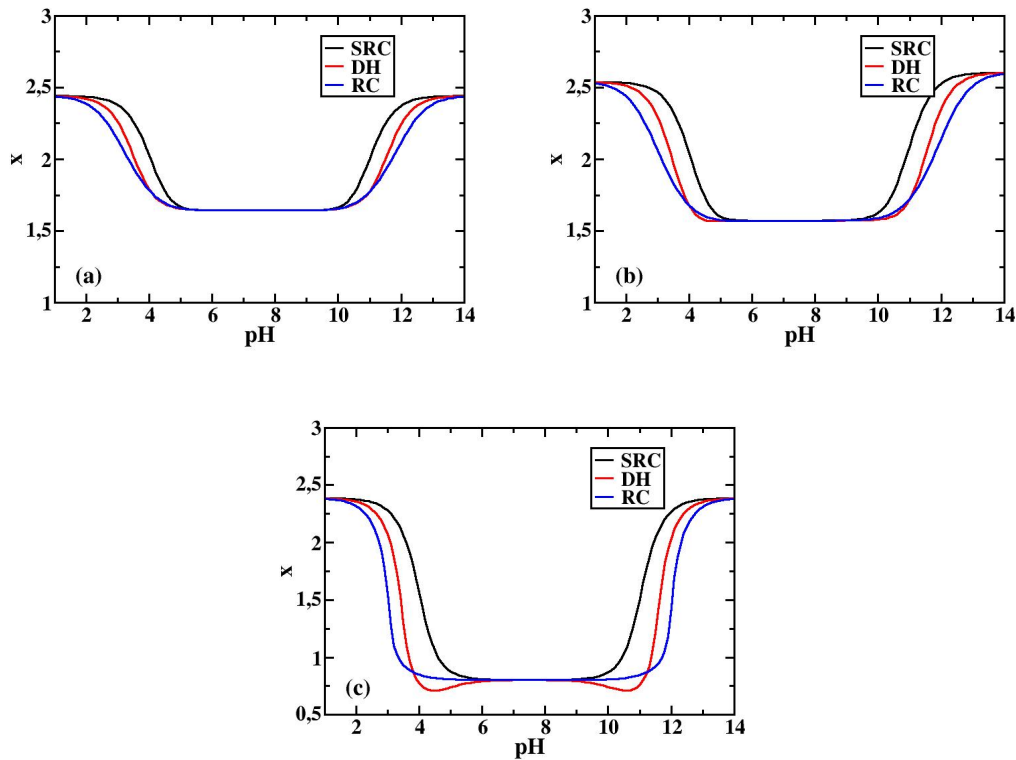


Fonte: Elaborado pelo autor.

sem essa aproximação (pontos azuis), nos permitindo analisar com mais detalhes quantitativos o quanto a probabilidade de protonação/desprotonação é afetada devido a regulação de carga. Podemos avaliar também a validade da aproximação de DH, vemos que para os gráficos representando IDPs com distribuição alternada e alternada por trechos, as duas abordagens não tiveram diferenças muito grandes entre as curvas representando a aproximação DH e em sua ausência, mostrando que essa aproximação é válida nesses casos. No caso da distribuição em dois blocos, vemos uma diferença muito grande entre as curvas representando a aproximação DH e sem ela, uma vez que, nesse caso os blocos carregados estão muito próximos um do outro, resultando em um potencial elétrico maior, não sendo possível a aproximação de DH. Além disso, observa-se valores médios das cargas maiores do que (+1), mostrando que essa aproximação não é válida, apresentando um resultado sem significado físico.

Obtendo os valores das cargas considerando as interações entre os resíduos, foi possível resolver numericamente a equação 4.43, obtendo o valor de $x = \langle R_{ee}^2 \rangle / \langle R_{ee,f}^2 \rangle$, que corresponde a uma grandeza que mede o quão expandida ou compactada a proteína está em relação a um estado de referência, que neste caso é uma cadeia livre de interações. Os resultados obtidos para cada modelo de IDP utilizado estão apresentados na Figura 5.4. De modo geral, nota-se que em pH s próximos a os pK as dos resíduos, na qual a proteína está eletricamente carregada, temos valores de $x > 1$, indicando que a proteína está mais estendida em relação a uma cadeia ideal. Isso ocorre devido as repulsões eletrostáticas entre os resíduos. A medida que o pH fica entre os valores de 5 a 10, a proteína se encontra eletrostaticamente neutra (exceto no caso da aproximação de DH para a distribuição em blocos). Dessa forma o valor de x é menor, pois se aproxima do caso de uma cadeia ideal. Podemos notar também que o valor de x é menor com a regulação de carga, o que é uma consequência do módulo da carga líquida da proteína ser menor. Dessa forma, as repulsões eletrostáticas são menos intensas, fazendo com que a proteína fique mais compacta. Com relação aos modelos de IDPs utilizados, vemos que os resultados das distribuições alternada (Figura 5.4 (a)) e alternada em trechos (Figura 5.4 (b))

Figura 5.4: Gráficos da distância *end-to-end* renormalizada em unidades dessa grandeza para uma cadeia livre de interações (x), pelo pH para os três modelos de IDPs propostos: distribuição alternada dos resíduos (a); distribuição em trechos alternados (b); e distribuição em blocos (c). A curva preta representa a carga total sem a regulação de carga (SRC); A curva vermelha é a carga total considerando a regulação de carga com a aproximação de Debye-Hückel (DH); A curva em azul consideramos a regulação de carga mas sem essa aproximação.



Fonte: Elaborado pelo autor.

são muito semelhantes, e ainda é válida a aproximação DH. Já para a distribuição em blocos (Figura 5.4 (c)), vemos valores de x muito menores com a regulação de carga se comparado com os gráficos das outras distribuições, uma vez que o bloco positivamente carregado atrai o bloco negativamente carregado, sendo uma cadeia mais compacta com relação a uma cadeia livre.

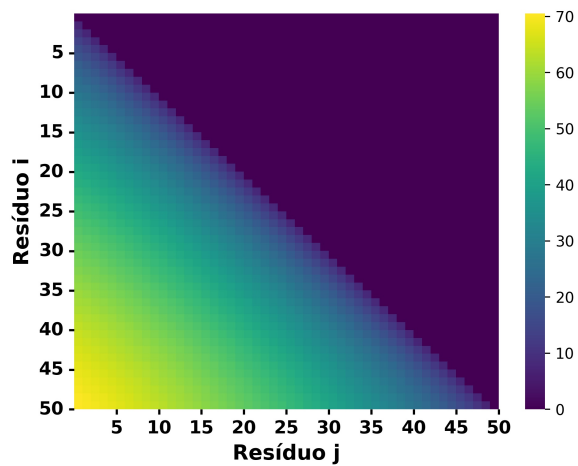
Através da equação 4.42 foi possível determinarmos a distância quadrática média entre cada par de resíduos ao longo da proteína para os três modelos considerando a regulação de carga sem a aproximação de DH. Vemos que essa equação são os elementos de uma matriz no qual suas linhas e colunas correspondem cada par de resíduos das IDPs. Essas matrizes estão representadas como mapas de cores na Figura 5.5. Para facilitar a visualização foi representada a raiz quadrada da distância quadrática média, onde a escala ao lado está em Å e os valores foram calculados para o pH=4,25. Além disso, foram feitas representações meramente ilustrativas, uma vez que nosso modelo não permite determinar uma conformação média tridimensional para representar a proteína. Nosso modelo nos permite obter dados de distância média quadrática apenas, nos permitindo fazer comparações entre os modelos descritos. Para o modelo de blocos, em especial, nota-se uma forma mais compacta com relação aos outros modelos, apesar de ser uma IDP. Estudos mostram que devido a forma como os resíduos estão distribuídos e dependendo das condições da solução (pH e concentração de sal) tem maior tendência de formarem agregados para soluções desse tipo de proteína (72).

Para o modelo com os resíduos dispostos de forma alternada, representado pela Figura 5.5 (a), podemos notar que a proteína é mais estendida devido as repulsões eletrostáticas entre os resíduos básicos, uma vez que são carregados em pHs baixos (em torno de pH=4,25) como mostra a Figura 5.5 (b). O modelo de IDP com trechos alternados segue um comportamento análogo, como podemos verificar na Figura 5.5 (c), exceto por algumas regiões, como as distâncias entre os resíduos 31 e 36, que são mais curtas devido a atração eletrostática entre os trechos positivos e negativos, como ilustrado pela Figura 5.5 (d). Por fim, para a distribuição em dois blocos, representada na Figura 5.5 (e), vemos que a distância entre as extremidades da proteína estão muito próximas, indicando que ela está mais compacta (Figura 5.5 (f)), o que corrobora com os resultados anteriores.

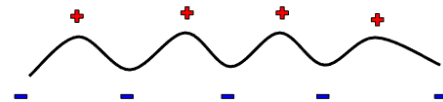
5.1 Aplicação em Peptídeos Antimicrobianos

O veneno de vespas contém diversas substâncias importantes para a sobrevivência do animal, tanto na caça quanto em sua defesa (73, 74). Além disso, vem chamando a atenção algumas substâncias encontradas nesses venenos que são de interesse farmacológico (73). Dentre diversos peptídeos derivados desses venenos, os mastoparans se destacam devido suas propriedades antimicrobianas, sendo muito eficazes contra bactérias gram-positivas, gram-negativas e leveduras (73, 75). Em relação a sua composição, os mastoparans são compostos por 14 resíduos de aminoácidos, amidado no C-terminal. Geralmente, são encontrados com três resíduos de lisinas nas posições 4/5 e/ou 11/12. Esses peptídeos se encontram sem uma conformação bem

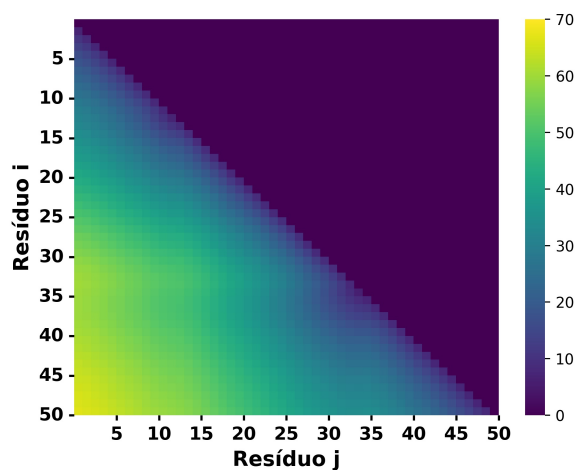
Figura 5.5: Matrizes cujos elementos representam a raiz da distância quadrática média entre o i -ésimo resíduo com o j -ésimo resíduo para os três modelos de IDPs: (a) com resíduos ácidos e básicos distribuídos de forma alternada; (c) em trechos alternados; (e) em dois blocos. As imagens (b), (d) e (f) são meramente ilustrativas, representando espacialmente as IDPs. A escala ao lado está medida em Å e as distâncias foram calculadas para o pH=4,25.



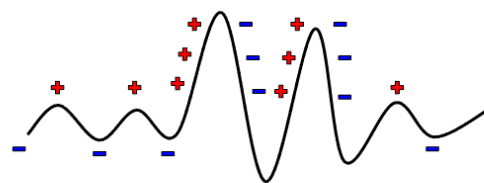
(a)



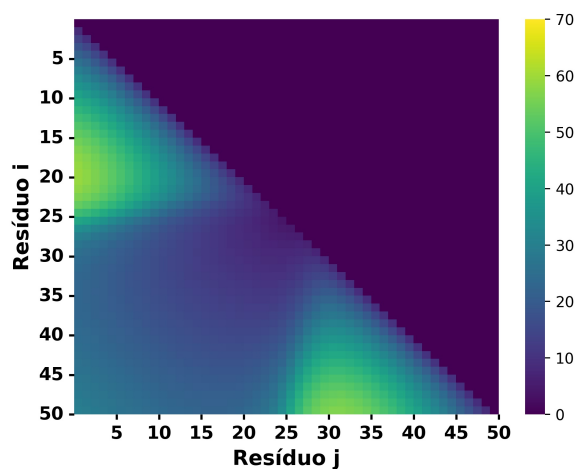
(b)



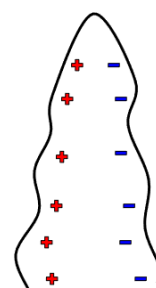
(c)



(d)



(e)



(f)

Fonte: Elaborado pelo autor.

definida em solução aquosa, adquirindo uma conformação helicoidal estável em ambientes como a membrana plasmática (75, 76). Sua atividade antimicrobiana se deve a essa conformação que auxilia na adesão de vários desses peptídeos na membrana plasmática, consequentemente, provocando poros e induzindo a exocitose (74, 75).

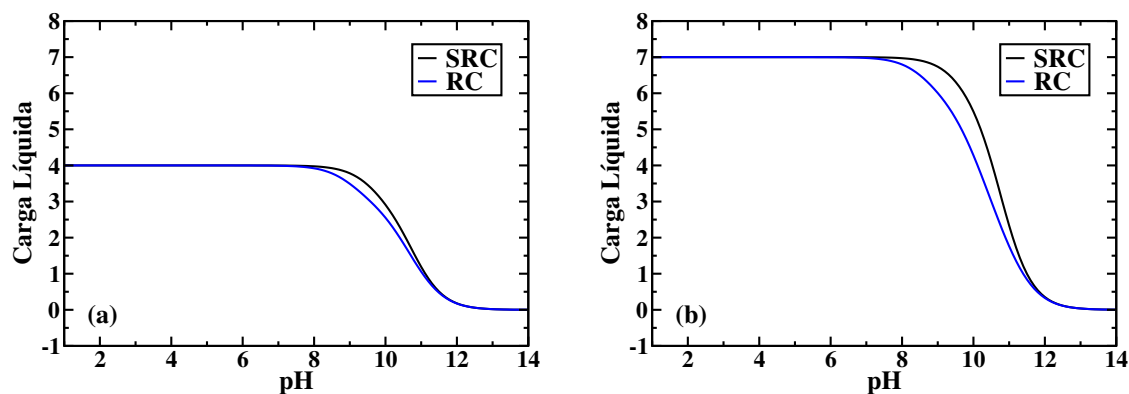
Estudos com dodecapeptídeos mostraram que certas regiões são mais suscetíveis a substituições do que outras e a adição de resíduos carregados aumenta a sua eficiência antimicrobiana (75, 77). Além disso, estudos com modelos de peptídeos antimicrobianos compostos apenas de lisinas e leucinas apresentam diferentes distribuições catiônicas e que a dispersão desses resíduos de aminoácidos na camada hidrofílica é um fator que contribui para a atividade antimicrobiana (75, 78). Devido a essas propriedades e a dependência com a distribuição de cargas do peptídeo, a regulação de carga pode desempenhar um papel importante para a sua formação helicoidal, sendo um ótimo modelo para aplicar o formalismo previamente desenvolvido. Para analisar os efeitos da regulação de carga foram utilizados dois mastoparans com distribuições e quantidades distintas de cargas, estudadas da referência (75). Os mastoparans utilizados foram as sequências MK578 (INWL**kAkk**VAGMIL- NH_2) e MK4578911 (INW**kkIk**kk**Vk**GML- NH_2), na qual na nomenclatura aqui utilizada, a letra M representa o mastoparan, K representa os resíduos de lisinas e os números se referem as suas posições ao longo da cadeia.

Assim como desenvolvido anteriormente, vamos analisar os efeitos da regulação de carga nos peptídeos antimicrobianos, calculando as cargas médias com a regulação de carga e comparando com o sistema na ausência desse efeito. Para determinar as cargas são necessários os valores médios do parâmetro de volume excluído e o comprimento de Kunh. Como os mastoparans se assemelham ao modelo com distribuição de trechos alternados, estabelecemos o valor de $\langle\omega\rangle = 3,5$ e mantivemos $l = 5,8$ Å.

A Figura 5.6 mostra a carga líquida para os dois peptídeos: (a) MK578 e (b) K4578911. Nessa figura podemos notar que a carga total passa a diminuir a medida que o pH está próximo dos valores do pKa da lisina (pKa=10,79) e do terminal amidado (pKa=9,60) (9). Nesses gráficos observa-se uma semelhança com o modelo de IDP no qual os resíduos estão distribuídos em trechos alternados, porém vemos que quanto maior o número de resíduos ionizáveis, maior o efeito da regulação de carga, uma vez que, nesses casos, as repulsões entre as lisinas são mais intensas, diminuindo a probabilidade de protonação desses resíduos e consequentemente suas cargas médias. Vemos isso com maior detalhe nos gráficos da Figura 5.7, no qual temos a carga de cada resíduo com a regulação e sem a regulação para os dois peptídeos calculados no pH=9,60, por ser um pH onde tanto as lisinas quanto o C-terminal amidado vão estar carregados, no caso do C-terminal, como esse pH corresponde é igual ao seu pKa as probabilidades de protonação e desprotonação são iguais.

Com as cargas obtidas, podemos determinar a distância quadrática média entre cada par de resíduos, como feito anteriormente. Os resultados obtidos estão apresentados na Figura 5.8, calculados para o $pH = 9,60$. No geral, podemos notar que o peptídeo com mais cargas positivas (MK4578911) está mais estendido com relação ao de menor carga líquida (MK578)

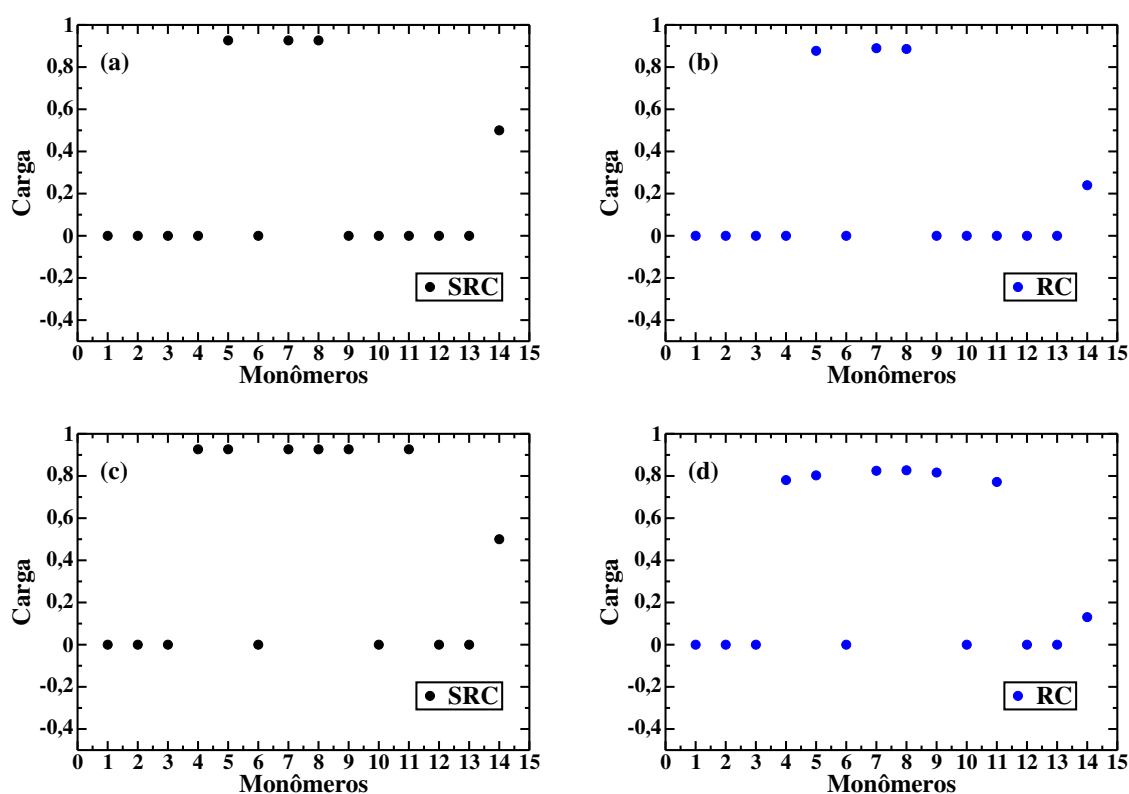
Figura 5.6: Gráficos de carga líquida por pH para os dois mastoparans: (a) MK578; (b) MK4578911. A curva preta representa o peptídeo sem regulação de carga (SRC) e a curva azul com a regulação de carga (RC).



Fonte: Elaborado pelo autor com os dados fornecidos da referência (75).

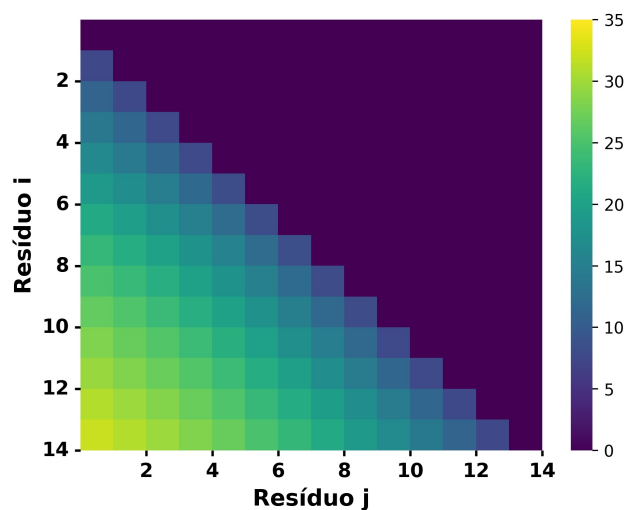
devido as repulsões eletrostáticas entre os resíduos positivos, como mostram as figuras (a) e (c), sendo de fato mais intensas para o MK4578911. Vemos que o papel da regulação de carga aqui é atenuar as repulsões, o que pode contribuir para a sua formação helicoidal.

Figura 5.7: Gráficos mostrando a carga média por resíduo de dois mastoparans: (a)-(b) MK578; (c)-(d) MK4578911 obtidas em pH=9,60. Os pontos pretos representam o peptídeo sem regulação de carga (SRC) e em azul, com a regulação de carga (RC).

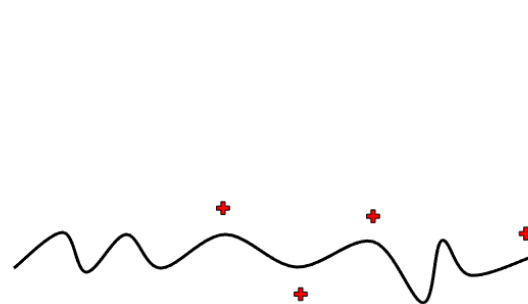


Fonte: Elaborado pelo autor com os dados fornecidos da referência (75).

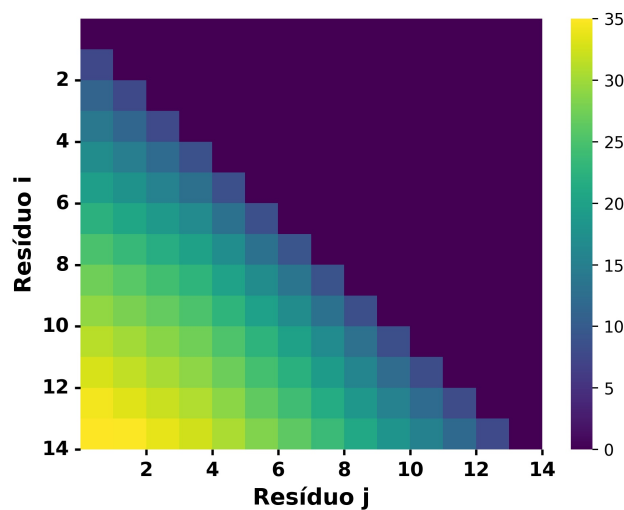
Figura 5.8: Matrizes cujos elementos representam a raiz da distância quadrática média entre o i -ésimo resíduo com o j -ésimo resíduo para os dois peptídeos: (a) MK578; (b) MK4578911. A escala ao lado está em Å e as distâncias foram calculadas para o pH=9,60.



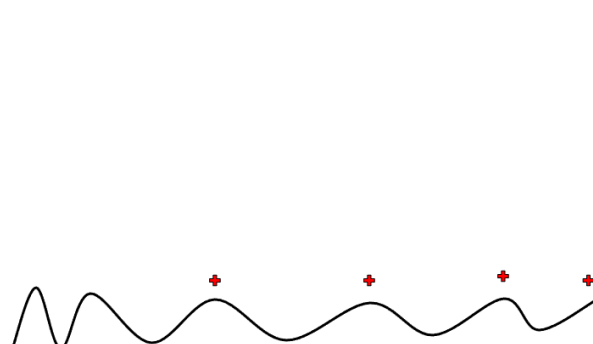
(a)



(b)



(c)



(d)

Fonte: Elaborado pelo autor.

Capítulo 6

Conclusão

Este trabalho teve como principal objetivo estudar os efeitos da regulação de carga em proteínas intrinsecamente desordenadas. Para isso, inicialmente foi feita uma descrição desse mecanismo, mostrando que ocorre em macromoléculas biológicas que possuam sítios tituláveis, permitindo que essas moléculas modifiquem suas distribuições de carga quando alguma variação no ambiente químico ocorra, como variação no pH, variação do potencial eletrostático e/ou concentração de sal. Como IDPs são ricas em resíduos ionizáveis, a regulação de carga pode desempenhar um papel central em interações envolvendo essa classe de proteínas, o que motivou o desenvolvimento deste trabalho.

Para estudar o mecanismo da regulação de carga em uma IDP foi utilizado o desenvolvimento da referência (52). Nessa referência, obtemos as expressões 4.56 e 4.57 através da energia livre do sistema, o que nos permitiu determinar as cargas de cada resíduo com o efeito da regulação de carga. Para a energia livre utilizou-se um modelo teórico simplificado para descrever a proteína de acordo com a referência (49), na qual não foram considerados os efeitos da concentração de sal, a constante dielétrica do meio foi utilizado um valor usual para modelos como esses. Não foram considerados os detalhes atômicos de cada resíduo, apenas seus valores de pK_a . Além disso, o hamiltoniano foi expresso com um comprimento de Kuhn renormalizado, contendo as interações de volume excluído, interações eletrostáticas e energia elástica das ligações entre os resíduos da cadeia, o que nos permitiu representar o hamiltoniano semelhante ao de uma cadeia livre de interações conhecida como cadeia gaussiana.

Os três modelos de IDPs utilizados foram idealizados como contendo dois tipos de aminoácidos apenas: o ácido glutâmico e a lisina. Afim de investigar o papel das distribuições de cargas, foi considerado que todas possuíam a mesma quantidade de resíduos, de modo que única diferença entre os três modelos foi a distribuição com que os resíduos das proteínas estavam dispostos: o primeiro foi considerado com uma disposição alternada; o segundo não seguia um padrão definido, com trechos de tamanhos distintos alternados ao longo da cadeia e o terceiro foi considerado como dois grandes blocos de resíduos básicos e ácidos. Para cada um desses modelos foram analisados os efeitos da regulação de carga em cada um de seus sítios tituláveis

em pH s iguais aos pK as de cada resíduo da proteína. Analisamos também como esse efeito afeta a conformação da proteína, através da distância *end-to-end* renormalizada e as distância entre os pares de resíduos.

De modo geral, nos três modelos foi possível observar que o efeito da regulação de carga causou a diminuição da carga líquida média da proteína quando o pH é próximo do valor do pK a dos resíduos. Isso ocorre porque nesses pH s a flutuação da carga dos resíduos atinge seu valor máximo, deixando-os mais suscetíveis a variar a probabilidade de protonação quando nas proximidades de um outro corpo carregado.

Como uma aplicação, exploramos o efeito da regulação de carga em peptídeos antimicrobianos, considerando que o sistema foi composto apenas pelos peptídeos, de forma análoga ao que foi feito para os modelos de IDPs. Foi possível analisar não só distribuições diferentes mas também quantidades distintas de cargas, pelo qual foi possível observar que quanto mais cargas no peptídeo, maior o efeito da regulação de carga, uma vez que as repulsões eletrostáticas são mais intensas, diminuindo a probabilidade de protonação dos resíduos básicos, portanto, diminuindo suas cargas médias. Como as repulsões são menos intensas, isso pode auxiliar na sua compactação, ajudando a estabelecer as ligações de hidrogênio para formar a hélice, consequentemente, na sua atividade antimicrobiana (79).

O modelo proposto neste trabalho, embora simplificado, isto é, considerando apenas a proteína isolada em solução, foi possível a partir dele determinar informações úteis em relação às cargas e conformações de IDPs, mostrando também sua extensão, ou seja, até onde foi válida a aproximação linear.

Referências

- 1 ALBERTS, B. et al. Biologia molecular da célula. [S.l.]: Artmed Editora, 2017.
- 2 PATOČKA, J.; KUČA, K.; JUN, D. Acetylcholinesterase and butyrylcholinesterase—important enzymes of human body. Acta Medica (Hradec Kralove), v. 47, n. 4, p. 215–228, 2004.
- 3 GUIDOTTI, G. Membrane proteins. Annual review of biochemistry, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 41, n. 1, p. 731–752, 1972.
- 4 ANANTHARAMAN, V.; BALAJI, S.; ARAVIND, L. The signaling helix: a common functional theme in diverse signaling proteins. Biology direct, Springer, v. 1, n. 1, p. 1–16, 2006.
- 5 CARTER, N. J.; CROSS, R. Mechanics of the kinesin step. Nature, Nature Publishing Group UK London, v. 435, n. 7040, p. 308–312, 2005.
- 6 POMMIER, Y. Topoisomerase i inhibitors: camptothecins and beyond. Nature Reviews Cancer, Nature Publishing Group UK London, v. 6, n. 10, p. 789–802, 2006.
- 7 LIEU, E. L. et al. Amino acids in cancer. Experimental & molecular medicine, Nature Publishing Group UK London, v. 52, n. 1, p. 15–30, 2020.
- 8 AKRAM, M. et al. Amino acids: A review article. Journal of Medicinal Plants Research, v. 5, n. 17, p. 3997–4000, 2011.
- 9 LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. Lehninger principles of biochemistry. [S.l.]: Macmillan, 2005.
- 10 CAMPBELL, M. K.; FARRELL, S. O.; MCDOUGAL, O. M. Biochemistry. [S.l.]: Cengage Learning, 2016.
- 11 UVERSKY, V. N. Intrinsically disordered proteins from a to z. The international journal of biochemistry & cell biology, Elsevier, v. 43, n. 8, p. 1090–1103, 2011.
- 12 HUANG, Y.-H.; ROSE, P. W.; HSU, C.-N. Citing a data repository: a case study of the protein data bank. PloS one, Public Library of Science San Francisco, CA USA, v. 10, n. 8, p. e0136631, 2015.
- 13 TRIVEDI, R.; NAGARAJARAM, H. A. Intrinsically disordered proteins: an overview. International Journal of Molecular Sciences, MDPI, v. 23, n. 22, p. 14050, 2022.
- 14 FISCHER, E. Einfluss der configuration auf die wirkung der enzyme. ii. Berichte der deutschen chemischen Gesellschaft, Wiley Online Library, v. 27, n. 3, p. 3479–3483, 1894.
- 15 BORGIA, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature, Nature Publishing Group UK London, v. 555, n. 7694, p. 61–66, 2018.
- 16 DUNKER, A. K. et al. Intrinsically disordered protein. Journal of molecular graphics and modelling, Elsevier, v. 19, n. 1, p. 26–59, 2001.
- 17 LERMYTE, F. Roles, characteristics, and analysis of intrinsically disordered proteins: A minireview. Life, MDPI, v. 10, n. 12, p. 320, 2020.
- 18 WALLIN, S. Intrinsically disordered proteins: structural and functional dynamics. Research and Reports in Biology, Taylor & Francis, p. 7–16, 2017.
- 19 OLDFIELD, C. J.; DUNKER, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry, Annual Reviews, v. 83, p. 553–584, 2014.
- 20 TOMPA, P. Intrinsically unstructured proteins. Trends in biochemical sciences, Elsevier, v. 27, n. 10, p. 527–533, 2002.

- 21 TOMPA, P. Intrinsically disordered proteins: a 10-year recap. Trends in biochemical sciences, Elsevier, v. 37, n. 12, p. 509–516, 2012.
- 22 PANCSA, R.; TOMPA, P. Structural disorder in eukaryotes. PloS one, Public Library of Science San Francisco, USA, v. 7, n. 4, p. e34687, 2012.
- 23 BOZIC, A.; PODGORNIK, R. Site correlations, capacitance, and polarizability from protein protonation fluctuations. The Journal of Physical Chemistry B, ACS Publications, v. 125, n. 46, p. 12902–12908, 2021.
- 24 ZHOU, H.-X.; PANG, X. Electrostatic interactions in protein structure, folding, binding, and condensation. Chemical reviews, ACS Publications, v. 118, n. 4, p. 1691–1741, 2018.
- 25 SEYREK, E. et al. Ionic strength dependence of protein-polyelectrolyte interactions. Biomacromolecules, ACS Publications, v. 4, n. 2, p. 273–282, 2003.
- 26 SEYREK, E.; DUBIN, P. L.; HENRIKSEN, J. Nonspecific electrostatic binding characteristics of the heparin-antithrombin interaction. Biopolymers: Original Research on Biomolecules, Wiley Online Library, v. 86, n. 3, p. 249–259, 2007.
- 27 KIRKWOOD, J. G.; SHUMAKER, J. B. Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 38, n. 10, p. 863–871, 1952.
- 28 LUND, M.; JÖNSSON, B. Charge regulation in biomolecular solution. Quarterly reviews of biophysics, Cambridge University Press, v. 46, n. 3, p. 265–281, 2013.
- 29 KIRKWOOD, J. G.; SHUMAKER, J. B. The influence of dipole moment fluctuations on the dielectric increment of proteins in solution. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 38, n. 10, p. 855–862, 1952.
- 30 BIESHEUVEL, P. M.; WITTEMANN, A. A modified box model including charge regulation for protein adsorption in a spherical polyelectrolyte brush. The Journal of Physical Chemistry B, ACS Publications, v. 109, n. 9, p. 4209–4214, 2005.
- 31 ZHENG, B. et al. Phase separation of polyelectrolytes: The effect of charge regulation. The Journal of Physical Chemistry B, ACS Publications, v. 125, n. 28, p. 7863–7870, 2021.
- 32 LINDERSTRØM-LANG, K. Om proteinstoffernes ionisation. Comptes Rendus des Travaux du Laboratoire Carlsberg, v. 15, p. 1–29, 1924.
- 33 GITLIN, I.; CARBECK, J. D.; WHITESIDES, G. M. Why are proteins charged? networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis. Angewandte Chemie International Edition, Wiley Online Library, v. 45, n. 19, p. 3022–3060, 2006.
- 34 JAENICKE, R.; SEMENZA, G. Selected Topics in the History of Biochemistry, Personal Recollections: VI. [S.l.]: Elsevier, 2000.
- 35 TIMASHEFF, S. N. et al. Studies of molecular interaction in isoionic protein solutions by light-scattering. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 41, n. 10, p. 710–714, 1955.
- 36 AGUILAR, B. et al. Statistics and physical origins of pK and ionization state changes upon protein-ligand binding. Biophysical journal, Elsevier, v. 98, n. 5, p. 872–880, 2010.
- 37 QUADRE, A. B. Efeitos da regulação de carga na interação entre macroíons e superfícies carregadas. Universidade Estadual Paulista (Unesp), 2022.
- 38 CAETANO, D. L. et al. Adsorption of lysozyme into a charged confining pore. Physical Chemistry Chemical Physics, Royal Society of Chemistry, v. 23, n. 48, p. 27195–27206, 2021.
- 39 SILVA, F. L. Barroso da et al. On the complexation of proteins and polyelectrolytes. The Journal of Physical Chemistry B, ACS Publications, v. 110, n. 9, p. 4459–4464, 2006.
- 40 DAS, R. K.; PAPPU, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences, National Acad Sciences, v. 110, n. 33, p. 13392–13397, 2013.
- 41 BAUL, U. et al. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. The Journal of Physical Chemistry B, ACS Publications, v. 123, n. 16, p. 3462–3474, 2019.
- 42 MAITY, H.; BAIDYA, L.; REDDY, G. Salt-induced transitions in the conformational ensembles of intrinsically disordered proteins. The Journal of Physical Chemistry B, ACS Publications, v. 126, n. 32, p. 5959–5971, 2022.

- 43 WOHL, S.; JAKUBOWSKI, M.; ZHENG, W. Salt-dependent conformational changes of intrinsically disordered proteins. The Journal of Physical Chemistry Letters, ACS Publications, v. 12, n. 28, p. 6684–6691, 2021.
- 44 DOBRYNIN, A. V.; COLBY, R. H.; RUBINSTEIN, M. Polyampholytes. Journal of Polymer Science Part B: Polymer Physics, Wiley Online Library, v. 42, n. 19, p. 3513–3538, 2004.
- 45 YANG, J. et al. Electrostatic interactions in molecular recognition of intrinsically disordered proteins. Journal of Biomolecular Structure and Dynamics, Taylor & Francis, v. 38, n. 16, p. 4883–4894, 2020.
- 46 DYSON, H. J.; WRIGHT, P. E. Intrinsically unstructured proteins and their functions. Nature reviews Molecular cell biology, Nature Publishing Group UK London, v. 6, n. 3, p. 197–208, 2005.
- 47 CHU, W.-T. et al. Role of non-native electrostatic interactions in the coupled folding and binding of puma with mcl-1. PLoS computational biology, Public Library of Science San Francisco, CA USA, v. 13, n. 4, p. e1005468, 2017.
- 48 SAWLE, L.; GHOSH, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. The Journal of chemical physics, AIP Publishing, v. 143, n. 8, 2015.
- 49 FIRMAN, T.; GHOSH, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. The Journal of chemical physics, AIP Publishing, v. 148, n. 12, 2018.
- 50 ATKINS, P.; JONES, L.; LAVERMAN, L. Princípios de Química-: Questionando a Vida Moderna e o Meio Ambiente. [S.l.]: Bookman Editora, 2018.
- 51 HILL, T. L. Statistical mechanics: principles and selected applications. [S.l.]: Courier Corporation, 2013.
- 52 BOŽIČ, A. L.; PODGORNIK, R. Anomalous multipole expansion: Charge regulation of patchy inhomogeneously charged spherical particles. The Journal of chemical physics, AIP Publishing, v. 149, n. 16, 2018.
- 53 SASSI, A. P. et al. Monte carlo simulations of hydrophobic weak polyelectrolytes: Titration properties and ph-induced structural transitions for polymers containing weak electrolytes. The Journal of chemical physics, American Institute of Physics, v. 97, n. 11, p. 8767–8774, 1992.
- 54 ULLNER, M.; JÖNSSON, B.; WIDMARK, P.-O. Conformational properties and apparent dissociation constants of titrating polyelectrolytes: Monte carlo simulation and scaling arguments. The Journal of chemical physics, American Institute of Physics, v. 100, n. 4, p. 3365–3366, 1994.
- 55 MARKOVICH, T.; ANDELMAN, D.; PODGORNIK, R. Charged membranes: Poisson-boltzmann theory. DLVO paradigm and beyond. arXiv preprint arXiv, v. 1603, p. 09451, 2016.
- 56 MARKOVICH, T.; ANDELMAN, D.; PODGORNIK, R. Charged membranes: Poisson-boltzmann theory, dlvo paradigm and beyond. arXiv preprint arXiv:1603.09451, 2016.
- 57 SALINAS, S. R. Introdução a física estatística vol. 09. [S.l.]: Edusp, 1997.
- 58 LEMOS, N. A. Mecânica analítica. [S.l.]: Editora Livraria da Física, 2007.
- 59 THORNTON, S. T.; MARION, J. B. Dinâmica clássica de partículas e sistemas. [S.l.]: Cengage Learning, 2011.
- 60 POLAND, D. Statistical Mechanics (McQuarrie, Donald A.). [S.l.]: ACS Publications, 1977.
- 61 XIAO, T.; SONG, X. A molecular debye-hückel theory and its applications to electrolyte solutions. The Journal of chemical physics, AIP Publishing, v. 135, n. 10, 2011.
- 62 CHODANOWSKI, P.; STOLL, S. Polyelectrolyte adsorption on charged particles in the debye- hückel approximation. a monte carlo approach. Macromolecules, ACS Publications, v. 34, n. 7, p. 2320–2328, 2001.
- 63 DOI, M.; EDWARDS, S. F. The theory of polymer dynamics. [S.l.]: oxford university press, 1988. v. 73.
- 64 MUTHUKUMAR, M. Physics of Charged Macromolecules. [S.l.]: Cambridge University Press, 2023.
- 65 GHOSH, K.; CARRI, G. A.; MUTHUKUMAR, M. Configurational properties of a single semiflexible polyelectrolyte. The Journal of Chemical Physics, American Institute of Physics, v. 115, n. 9, p. 4367–4375, 2001.
- 66 ISRAELACHVILI, J. Intermolecular and Surface Forces; Israelachvili. [S.l.]: JN Academic Press: Cambridge, MA, USA, 2011.

-
- 67 EDWARDS, S. F.; JEFFERS, E. F. Size of a polymer molecule in solution. part 2.—semi-dilute solutions. Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics, Royal Society of Chemistry, v. 75, p. 1020–1029, 1979.
- 68 BUTKOV, E. Física matemática. [S.l.]: Livros Técnicos e Científicos, 1988. v. 1.
- 69 ARFKEN, G.; ROMAIN, J. Mathematical methods for physicists. Physics Today, AIP Publishing, v. 20, n. 5, p. 79–79, 1967.
- 70 MUTHUKUMAR, M. Adsorption of a polyelectrolyte chain to a charged surface. The Journal of chemical physics, American Institute of Physics, v. 86, n. 12, p. 7230–7235, 1987.
- 71 IRIMIA, M. et al. Self-consistent-field method for correlation calculation within density-matrix-functional theory. Physical Review A, APS, v. 108, n. 5, p. 052818, 2023.
- 72 MCCARTY, J. et al. Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins. The journal of physical chemistry letters, ACS Publications, v. 10, n. 8, p. 1644–1652, 2019.
- 73 CHEN, X. et al. Evaluation of the bioactivity of a mastoparan peptide from wasp venom and of its analogues designed through targeted engineering. International journal of biological sciences, Ivyspring International Publisher, v. 14, n. 6, p. 599, 2018.
- 74 CHOI, M. B.; LEE, Y.-H. The structure and antimicrobial potential of wasp and hornet (vespidae) mastoparans: A review. Entomological Research, Wiley Online Library, v. 50, n. 7, p. 369–376, 2020.
- 75 SOUZA, B. M. de et al. Structure–activity relationship of mastoparan analogs: effects of the number and positioning of lys residues on secondary structure, interaction with membrane-mimetic systems and biological activity. Peptides, Elsevier, v. 72, p. 164–174, 2015.
- 76 OSHIRO, K. G. et al. Deciphering the structure and mechanism of action of computer-designed mastoparan peptides. The FEBS Journal, Wiley Online Library, 2023.
- 77 HILPERT, K. et al. High-throughput generation of small antibacterial peptides with improved activity. Nature biotechnology, Nature Publishing Group US New York, v. 23, n. 8, p. 1008–1012, 2005.
- 78 KIM, Y. S.; CHA, H. J. Disperse distribution of cationic amino acids on hydrophilic surface of helical wheel enhances antimicrobial peptide activity. Biotechnology and bioengineering, Wiley Online Library, v. 107, n. 2, p. 216–223, 2010.
- 79 GOLKARAM, M.; SHIN, Y.; DUIN, A. V. Reactive molecular dynamics study of the ph-dependent dynamic structure of α -helix. The Journal of Physical Chemistry B, ACS Publications, v. 118, n. 47, p. 13498–13504, 2014.