

UNIVERSIDADE DE SÃO PAULO
PROGRAMA INTERUNIDADES DE PÓS GRADUAÇÃO EM BIOINFORMÁTICA

LARIANI APARECIDA DELBONI

**Estudo por simulação computacional em larga escala da complexação
de proteínas do leite**

Ribeirão Preto

2016

LARIANI APARECIDA DELBONI

**Estudo por simulação computacional em larga escala da complexação
de proteínas do leite**

Dissertação de Mestrado apresentada ao Programa Interunidades de Pós-graduação em Bioinformática da Universidade de São Paulo para obtenção do grau de Mestre em Ciências.

Área de concentração: Bioinformática
Orientador: Prof. Dr. Fernando Luís Barroso da Silva

Ribeirão Preto

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Delboni, Lariani Aparecida

Estudo por simulação computacional em larga escala da complexação de proteínas do leite. Ribeirão Preto, 2016.

141 p. : il. ; 30 cm

Dissertação de Mestrado, apresentada ao Programa Interunidades de Pós-graduação em Bioinformática. Área de concentração: Bioinformática.

Orientador: da Silva, Fernando Luís Barroso.

1. Proteínas do soro do leite. 2. Correlação de cargas em proteínas. 3. Modelo do solvente contínuo. 4. Energia livre. 5. Complexação.

FOLHA DE AVALIAÇÃO

Lariani Aparecida Delboni

Estudo por simulação computacional em larga escala da complexação de proteínas do leite

Dissertação de Mestrado apresentada ao Programa Interunidades de Pós-graduação em Bioinformática da Universidade de São Paulo para obtenção do grau de Mestre em Ciências.

Área de concentração: Bioinformática

Orientador: Prof. Dr. Fernando Luís Barroso da Silva

Aprovada em: 24/01/2017

Banca examinadora

Prof(a). Dr(a). Fernando Luís Barroso da Silva

Instituição: FCFRP - USP

Assinatura: Fernando L. B. da S.

Prof(a). Dr(a). Osvaldo de Freitas

Instituição: FCFRP - USP

Assinatura: [Assinatura]

Prof(a). Dr(a). Luciana Malavolta Quaglio

Instituição: Faculdade de Ciências Médicas da Santa Casa de São Paulo

Assinatura: [Assinatura]

AGRADECIMENTO

Grata a Deus, por estar aqui.

Agradeço a meus pais, Maria Aparecida e Áureo, por terem proporcionado e incentivado meus estudos desde a infância. Por todo o cuidado, atenção, disponibilidade e carinho, sem os quais a realização deste trabalho não seria possível.

À professora Ana Sílvia, professor Osvaldo e professor Caliri, componentes da banca examinadora da qualificação, meus sinceros agradecimentos pela leitura cuidadosa do trabalho e suas respectivas contribuições para a finalização deste texto.

A meu orientador, professor Fernando Barroso, um agradecimento especial por todo o tempo e paciência dispensados ao iniciar uma farmacêutica na área computacional. Obrigada por todas as oportunidades concedidas, todos os ensinamentos, apoio e broncas, e principalmente por confiar em mim. Foi um longo caminho, que sem seu incentivo e parceria, não teria sido possível.

Agradeço ao grupo de Química Teórica da Universidade de Lund na Suécia, em nome do professor Bo Jönsson, que sem dúvida alguma contribuiu para a concretização deste Mestrado.

Agradeço às secretárias do programa Interunidades em Bioinformática, Patrícia e Cristiane, responsáveis por todo o suporte burocrático neste Mestrado.

Um agradecimento especial também à professora Marilisa, pelo apoio e compreensão quanto ao cumprimento dos créditos e finalização deste trabalho.

Aos amigos e colegas, que de uma forma ou de outra me ouviram, incentivaram ou apoiaram nesta caminhada. Foram muitos, mas especialmente Alejandro, Junier, João Paulo e Maristela.

RESUMO

A estrutura e o comportamento de proteínas são de grande interesse a uma infinidade de áreas de conhecimento, com especial destaque para alimentação e saúde. As proteínas compõem uma fundamental classe de moléculas biológicas, com capacidade para interagir com diversas outras estruturas e formar complexos. Dois tópicos principais são abordados neste trabalho: (a) avaliação da formação de complexos entre três proteínas do soro do leite (α -lactoalbumina, β -lactoglobulina e lactoferrina) em diferentes condições experimentais de força iônica e pH da solução através de derivadas da energia livre e (b) avaliação de aspectos relacionados com a fundamentação teórica dos modelos moleculares com granularidade grossa e uso do dielétrico contínuo empregados para estudo do fenômeno da complexação molecular, através da análise, em larga escala, da correlação de cargas em proteínas disponíveis no banco de dados de proteínas. Os estudos teóricos de complexação foram realizados através do emprego de simulações Monte Carlo Metropolis, com o modelo do solvente contínuo, constante dielétrica homogênea e sal implicitamente descrito. Os estudos de complexação permitem compreender os mecanismos moleculares envolvidos, elucidando, inclusive, controvérsias experimentais e oferecendo orientação para aplicações. Já os resultados do estudo de correlação de cargas em estruturas experimentais de proteínas permitem justificar as aproximações do modelo.

Palavras chave: Proteínas do soro do leite, Correlação de cargas em proteínas, Modelo do solvente contínuo, Energia livre, Monte Carlo, Complexação.

ABSTRACT

The structure and behavior of proteins are of great interest for a number of science fields, with special emphasis to food industry and health. Proteins comprise a fundamental group of biological molecules, with the capacity of interacting with several other structures and creating complexes. Two major topics are covered in this study: (a) evaluation of complex formation among three whey proteins (α -lactalbumin, β -lactoglobulin and lactoferrin) under different experimental conditions of ionic strength and pH of solution through free energy derivatives and (b) evaluation of features related to the theoretical basis of the coarse-grained molecular models and the use of continuous dielectric which are employed to investigate the molecular complexation phenomenon, through analysis, on a large scale, of charge correlation in proteins available in the protein data bank. The theoretical studies of complexation were performed through the use of Monte Carlo Metropolis simulations, with continuous solvent model, homogeneous dielectric constant and salt implicitly described. The complexation studies allow to understand the molecular mechanisms involved, also elucidating experimental controversies and offering orientation to applications; whereas the results of the charge correlation study in experimental structure of proteins allow justify the approximations of the model.

Keywords: Whey proteins, Charge correlations in proteins, Continuum solvent model, Free energy, Monte Carlo, Complexation.

LISTA DE FIGURAS

Figura 1 – Sequência de etapas desde o leite até o produto final.....	3
Figura 2 – Esquema da estrutura de uma micela.....	6
Figura 3 – Molécula de β -LG (código PDB 1BEB).	7
Figura 4 – Molécula de α -LA (código PDB 1F6S), representada em “fita”.....	8
Figura 5 – Molécula de LF, (código PDB 1BLF).	10
Figura 6 – Algumas das aplicações mais clássicas do leite e suas proteínas.	12
Figura 7 – Esquema representando os quatro níveis de organização de uma proteína.	14
Figura 8 – Aminoácido isolado glicina.	15
Figura 9 – Aminoácidos de cadeia lateral com capacidade para alterar a carga líquida da molécula, conforme o grupo químico.	16
Figura 10 – Representação de proteína com momento de dipolo afetado por uma carga externa.	18
Figura 11 – Ilustração da alteração do estado de protonação de uma proteína próxima a uma superfície carregada, em função do potencial elétrico imposto por esta interface.....	20
Figura 12 – Esquema simplificado do efeito hidrofóbico e maximização da entropia.	22
Figura 13 – Esquema mostrando diferentes associações entre proteínas.....	24
Figura 14: Criação de um modelo para flores, desde um canteiro até uma flor genérica.....	28
Figura 15 – Representação de três dos principais níveis de modelagem para proteínas.....	30
Figura 16 – Comparação de superfície energética para estruturas de nível de modelagem atômico e de granularidade grossa.....	31
Figura 17 – Esquema comparando química computacional e bioinformática.	32
Figura 18 – Esquema do procedimento de correlação de cargas de uma proteína.....	36
Figura 19 – Esquema simplificado do método Monte Carlo Metropolis.....	40
Figura 20 – Representação esquemática das “raias” para o cálculo da função de distribuição radial.	41
Figura 21 – Esquema de cilindro onde ocorrem as simulações de complexação.....	43
Figura 22 – Curvas de titulação das proteínas α -LA, β -LG e LF, em diferentes concentrações de sal.	51

Figura 23 – Curvas de titulação sobrepostas das proteínas α -LA, β -LG e LF, com concentração de sal de 150 mM.	52
Figura 24 – Capacitâncias simuladas para as proteínas do soro do leite α -LA, β -LG e LF, em função do pH em diferentes concentrações de sal.....	53
Figura 25 – Sobreposição das curvas de capacitância das proteínas α -LA, β -LG e LF, em concentração de sal de 150 mM.	54
Figura 26 – Valores de momento de dipolo simulados para as proteínas do soro do leite α -LA, β -LG e LF, em função do pH em diferentes concentrações de sal.	55
Figura 27 – Curvas de potencial de energia livre de interação $\beta w(r)$ para os três sistemas estudados.	56
Figura 28 – Curvas de potencial de energia livre de interação $\beta w(r)$ para os três sistemas estudados comparando diferentes pHs.	57
Figura 29 – Curvas de potencial de energia livre de interação $\beta w(r)$ para os três sistemas estudados comparando diferentes condições salinas, assim como vdW.....	59
Figura 30 – Efeito do pH nas interações simuladas e estimadas entre os sistemas α -LA-LF e β -LG-LF.	60
Figura 31 – Representação da proteína carregada com as camadas concêntricas para criação do histograma.	65
Figura 32 – Esquema de uma célula esférica contendo solução eletrolítica em um meio descrito pela constante dielétrica ϵ	68
Figura 33 – Distribuição de probabilidade de cargas negativas e positivas confinadas em uma esfera de raio 20 Å.	70
Figura 34 – Distribuição de cargas positivas e negativas para todas as 172.622 cadeias de proteínas disponíveis no PDB.	73
Figura 35 – Distribuição de resíduos positivos e negativos na α -LA (<i>Bos taurus</i>) PDB id 1F6S, nas formas <i>apo</i> e <i>halo</i> , em pH aproximadamente neutro.	76
Figura 36 – Distribuição de resíduos positivos e negativos na κ -caseína, em pH aproximadamente neutro.....	77
Figura 37 – Distribuição de resíduos positivos e negativos na proteína priônica humana (PDB id 1HJM), em pH aproximadamente neutro.	78
Figura 38 – Distribuição de resíduos positivos e negativos das proteínas β -LG (PDB id 1BEB) (a) e LF (PDB id 1BLF) (b), em pH aproximadamente neutro.	79
Figura 39 – Distribuição de resíduos positivos e negativos nas proteínas, em pH aproximadamente neutro.....	80

Figura 40 – Distribuição de resíduos positivos e negativos. 81

LISTA DE TABELAS

Tabela 1 – Principais propriedades físico-químicas das proteínas estudadas.	49
Tabela 2 – Valores de carga líquida e variação de energia livre de ligação para os complexos α -LA – LF e β -LG – LF nos pH's 6,5 e 8.	61
Tabela 3 – Conjunto de dados obtidos do PDB para estudo de correlação de cargas, divididos em duas classificações, Biológica e Estrutural.	63
Tabela 4 – Características médias das diferentes categorias de proteínas extraídas do PDB.	71
Tabela 5 – Distribuição de carga líquida e massa molecular de algumas proteínas estudadas.	75
Tabela 6 – Distribuição de resíduos para a proteína precursora do príon (1HJM) e lisozima (2LZT) com correlação de cargas positiva e a proteína α -LA (1F6S) com correlação de cargas negativa.	82

LISTA DE ABREVIATURAS

BPTI	Inibidor da tripsina pancreática bovina
CN	Caseína
DH	Debye-Hückel
DM	Dinâmica Molecular
Ig	Imunoglobulina
LF	Lactoferrina
LJ	Lennard-Jones
LP	Lactoperoxidase
LSZ	Lisozima
MC	Monte Carlo
MCM	Monte Carlo Metropolis
PDB	Protein Data Bank (Banco de Dados de Proteínas)
RMN	Ressonância Magnética Nuclear
SA	Soroalbumina
TK	Tanford-Kirkwood
VAST	Vector Alignment Search Tool (algoritmo computacional – Ferramenta de busca de alinhamento de vetor)
vdW	van der Waals
α-LA	α -lactoalbumina
β-LG	β -lactoglobulina
ΔG	Varição de energia livre

SUMÁRIO

RESUMO.....	i
ABSTRACT	ii
LISTA DE FIGURAS	iii
LISTA DE TABELAS	iii
LISTA DE ABREVIATURAS.....	vii
1. Contextualização	1
2. Introdução	4
2.1 Leite.....	48
2.1.1 Proteínas do leite	5
2.1.2 Proteínas do soro do leite	6
2.1.3 Aplicações das proteínas do leite	10
2.2 Proteínas.....	13
2.2.1 Aspectos químicos e físicos das proteínas	13
2.2.2 Interação proteína-proteína.....	23
2.2.3 Banco de Dados de Proteínas (PDB).....	26
2.2.4 Modelagem molecular.....	27
2.2.5 Correlação de cargas das proteínas	33
3. Objetivos	37
4. Metodologia e modelo.....	38
4.1 Método Monte Carlo Metropolis.....	38
4.2 Potencial de força média	40
4.3 Complexação de proteínas do leite.....	42
5. Resultados e discussão	48
5.1 Propriedades físico-químicas de proteínas do soro do leite	48
5.2 Complexação molecular	55

6.	Estudo de correlação de cargas em estruturas de proteínas.....	62
6.1	Caracterização do modelo	62
6.1.1	Correlação de cargas em estruturas tridimensionais	62
6.1.2	Sistema referência para a correlação de cargas	67
6.2	Resultados e discussão	68
6.2.1	Sistema de referência	68
6.2.2	Análise das estruturas de proteínas do PDB.....	70
7.	Conclusões	83
8.	Referências.....	85
	APÊNDICE A – Disciplinas cursadas	106
	APÊNDICE B – Artigo já publicado	107
	APÊNDICE C – Artigo em preparação.....	119

1. Contextualização

Esta dissertação estudará a complexação de proteínas do leite, através de uma abordagem de Bioinformática.

O leite é um dos alimentos mais completos que existem, perfeito para suprir as necessidades nutricionais e fisiológicas do recém-nascido (THOMPSON; BOLAND; SINGH, 2009). Há milênios são produzidos seus derivados, mantendo características nutricionais e aumentando a durabilidade. Um dos componentes do leite de maior interesse por parte da indústria são as proteínas, em especial as proteínas do soro (THOMPSON; BOLAND; SINGH, 2009; MCSWEENEY; FOX, 2013), com novas aplicações sendo descobertas a cada dia (CHANASATTRU; DECKER; MCCLEMENTS, 2009; BOUYER et al., 2012; BRYANT; CAI; SINGH, 2013; SAĞLAM et al., 2014; GAJENDRAGADKAR; GOGATE, 2016; JAUREGI et al., 2016; WAGONER; FOEGEDING, 2017).

As proteínas do soro do leite são em sua maioria moléculas relativamente estáveis à degradação, com funções voltadas a eventos fisiológicos (THOMPSON; BOLAND; SINGH, 2009). Muitas vezes essas proteínas se ligam a outras estruturas, como outras proteínas, formando complexos moleculares. Esses complexos são estruturas que despertam grande interesse à indústria, como farmacêutica (COWAN-JACOB; MÖBITZ; FABBRO, 2009), cosmética (BRYANT; CAI; SINGH, 2013), mas especialmente à indústria alimentícia (SAĞLAM et al., 2014).

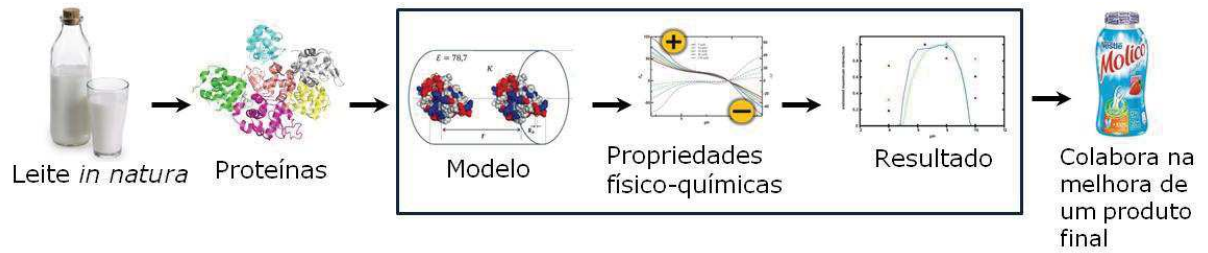
Nosso trabalho visa justamente observar e entender algumas das principais propriedades das proteínas do soro do leite, e como essas propriedades influenciam na formação de complexos de três proteínas do soro selecionadas. Tomando como exemplo a lactoferrina (LF)¹, sabe-se que é uma proteína catiônica, portanto, capaz de se ligar eletrostaticamente às outras proteínas, que são em sua maioria aniônicas (BÖSZE, 2008; THOMPSON; BOLAND; SINGH, 2009; MCSWEENEY; FOX, 2013). Porém, um estudo detalhado deste processo, com condições bem definidas faz-se necessário, principalmente para entender e quantificar os efeitos de sal e pH nesses processos.

¹ As abreviaturas dos nomes das proteínas do leite utilizadas neste trabalho seguem as regras propostas por Farrell (FARRELL et al., 2004).

Para obter nossos resultados, escolhemos como ferramenta de estudo as simulações computacionais (ARAI, 1995; HOLDER; KILWAY, 2005), quando possível comparando com dados experimentais. Visto a necessidade de um grande volume de simulações, em função de uma ampla variedade de diferentes condições experimentais, a abordagem empregada foi a bioinformática (GERSTEIN; JANSEN, 2000; LUSCOMBE; GREENBAUM; GERSTEIN, 2001), com o uso de conceitos da mecânica estatística (HILL, 1986) e uso do método Monte Carlo Metropolis (METROPOLIS; ULAM, 1949; METROPOLIS et al., 1953). Para a solução dos modelos, este é um método robusto e bastante adequado para a obtenção de valores de energia livre de interação entre moléculas, sendo esse nosso alvo. Assim, é possível obter um perfil de energia livre associada com a formação (ou não) de complexo entre duas proteínas, em função da distância de separação entre as mesmas. Da análise dessa grandeza física, pode-se entender os mecanismos moleculares e determinar as condições de complexação para uso prático.

A finalidade deste trabalho é obter valores conclusivos referentes à formação de complexos entre três das principais proteínas do soro do leite. O intuito dos resultados é que sejam úteis especialmente à indústria, promovendo melhorias nos produtos que contenham derivados do leite, notadamente bebidas esportivas em meio aquoso (DELBONI; DA SILVA, 2016). A formação de produtos estáveis por mais tempo com decorrente aumento no tempo de prateleira, a criação de novos alimentos ou fármacos que contenham proteínas do soro do leite (consequentemente levando consigo suas vantagens fisiológicas) são algumas das aplicações que os perfis de formação de complexos podem proporcionar.

Na Figura 1 é apresentada uma sequência baseada nesta abordagem de bioinformática, desde o leite *in natura* até um produto final industrializado. O leite possui suas diversas proteínas, que depois de selecionadas e caracterizadas, permitem definirmos os elementos mais importantes de um processo, propormos um modelo teórico (com o auxílio de simulações computacionais) para racionalizá-lo, e assim, obtermos propriedades químicas e físico-químicas a serem analisadas. A partir dessas propriedades, resultados acerca do comportamento das proteínas são obtidos e avaliados, até que possam ser empregados na melhoria de um produto já existente ou criação de um novo. Toda essa sequência compreende uma série de áreas de conhecimento distintas. O presente trabalho foca na terceira, quarta e quinta etapas, demarcadas pelo quadro, ou seja, a modelagem do sistema e sua solução computacional, com a obtenção de resultados práticos através de propriedades características das moléculas.

Figura 1 – Sequência de etapas desde o leite até o produto final.

Na sequência, partindo do leite (<https://ecomedicaonline.com/images/goatsmilk.jpg>), tem-se as proteínas (estrutura representada com PDB id 1F6S). A partir daí o modelo é construído e os resultados são obtidos, como ilustrado no quadro (DELBONI; DA SILVA, 2016), beneficiando um produto final (<http://www.paodeacucar.com/img/uploads/1/183/523183.jpg>).

2. Introdução

Neste tópico abordaremos os dois principais pontos de interesse desse trabalho, ou seja, o leite e suas proteínas, assim como uma extensa revisão bibliográfica da área.

2.1 Leite

O leite é um líquido secretado pelas glândulas mamárias das fêmeas de todas as espécies de mamíferos. Quimicamente, é uma solução coloidal composta basicamente por água, lipídios, açúcares, sais e proteínas. Possui os componentes ideais para o desenvolvimento dos recém-nascidos, como ácidos graxos essenciais, lactose, elementos inorgânicos, onde se encontram diferentes sais, proteínas, sendo a caseína (CN) a proteína em maior quantidade, além de uma coleção de diferentes vitaminas e hormônios (THOMPSON; BOLAND; SINGH, 2009).

É considerado um alimento completo (quase perfeito), no entanto, é sob medida para as devidas espécies, podendo ser deficiente nutricionalmente para as outras (THOMPSON; BOLAND; SINGH, 2009). A distribuição dos componentes do leite pode variar bastante, como por exemplo, animais que vivem em áreas geladas apresentam altas concentrações de gordura, enquanto o leite de outros animais apresenta somente traços. Ou seja, o leite de cada animal tem a distribuição de componentes diretamente relacionada à espécie, tipo de gestação, clima ou tempo de desenvolvimento do filhote. Ainda assim, entre oito e 10 mil anos atrás, após o ser humano evoluir a ponto de domesticar rebanhos, desenvolveu-se a prática de utilizar o leite de outras espécies como alimento, e suas características nutricionais deram suporte ao desenvolvimento dos seres humanos (EVERSHED et al., 2008).

Além das funções de nutrição que possibilitam o desenvolvimento do indivíduo, o leite promove a manutenção de muitas funções biológicas no organismo, proporcionadas principalmente por sua porção de proteínas e peptídeos. Os responsáveis por essas funções biológicas são imunoglobulinas, enzimas e inibidores enzimáticos, fatores de crescimento, hormônios e agentes antibacterianos (THOMPSON; BOLAND; SINGH, 2009). Mesmo com a especificidade entre o leite e as respectivas espécies, o leite de vaca tornou-se nos dias atuais o leite mais amplamente difundido para consumo humano em idade adulta. A variedade de

proteínas e outros nutrientes existentes no leite de vaca permite já há milênios, a produção de alimentos derivados, os chamados laticínios, como queijos, manteigas e iogurtes, dos mais diversos tipos. A maior justificativa para a produção dos derivados do leite é aumentar sua durabilidade, visto que o leite *in natura* mantém suas características por pouco tempo antes da deterioração. Mais recentemente outros usos vêm sendo constantemente descobertos, tanto na indústria de laticínios (YADA, 2004; SAĞLAM et al., 2014; KAPOOR; RAFIQ; SHARMA, 2015), quanto em outras das mais diversas, como farmacêutica (JACOBSON, 2015; MÜLLER, 2015; OMORPHOS; KAHN; KALASKAR, 2015), cosmética (STEIJNS; VAN HOOIJDONK, 2000; SECCHI, 2008), desenvolvimento de detergentes (VOJCIC et al., 2015), novos materiais (SU; KIM; LIU, 2014), entre outras.

2.1.1 Proteínas do leite

Dentre todos os constituintes do leite, as proteínas são os elementos que mais afetam as propriedades do leite em si, e dos laticínios em geral (KINSELLA; WHITEHEAD, 1989; KORHONEN et al., 1998; SMITHERS, 2008; WATSON; ZIBADI; PREEDY, 2010; MCSWEENEY; FOX, 2013). Por ser um alimento tão nobre, as primeiras publicações de estudos a respeito do leite datam do século 19, sendo a CN citada desde 1830. Em meio ácido o leite coagula, formando um precipitado constituído basicamente de CN. O líquido restante é chamado de soro, e composto pelos outros elementos do leite, assim como as proteínas do soro (THOMPSON; BOLAND; SINGH, 2009).

Hoje sabe-se que a CN, proteína mais abundante do leite, não é uma proteína única, e sim um conjunto de quatro diferentes moléculas pertencentes à mesma família, chamadas α_{s1} -CN, α_{s2} -CN, β -CN e κ -CN. Além da coagulação em meio ácido (as CNs são insolúveis em pH abaixo de 4,6, sendo que as proteínas do soro são solúveis), as CNs podem se diferenciar das proteínas do soro em vários outros aspectos, como maior resistência ao calor, alta quantidade de prolina nos resíduos (dificulta a formação de α -hélices e folhas- β , criando estruturas sem forma definida), sendo exclusivas do leite, enquanto muitas das proteínas do soro são originadas no sangue, com funções biológicas distintas. Além disso, as CNs estão

presentes no leite organizadas na forma de micelas², como ilustrado na Figura 2, e as proteínas do soro apresentam-se como monômeros ou dímeros, em sua maioria (THOMPSON; BOLAND; SINGH, 2009). O pH é um importante parâmetro nesse processo. Por exemplo, a β -lactoglobulina (β -LG) existe na forma monomérica (pH abaixo de 3,5 e acima de 7,5), dimerica (entre os pHs 5,5 e 7,5) e tetramérica (no intervalo de pH entre 3,5 e 5,5) (THOMPSON; BOLAND; SINGH, 2009).

Figura 2 – Esquema da estrutura de uma micela.

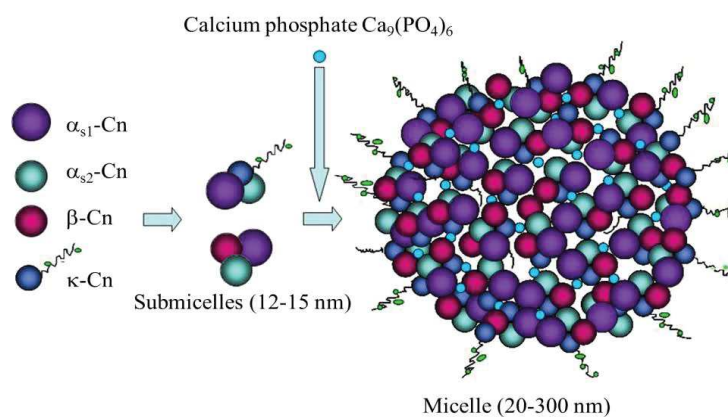


Figura adaptada do trabalho de revisão de Rebouillat, de 2015 (REBOUILLAT; ORTEGA-REQUENA, 2015). Os quatro subtipos de caseína se agregam e formam submicelas. Com o acréscimo do fosfato de cálcio, as submicelas também se agregam, formando assim a micela.

2.1.2 Proteínas do soro do leite

Visto a capacidade das proteínas do soro atuarem em regiões distintas do organismo, regulando outras funções fisiológicas além da nutrição (PIHLANTO-LEPPÄLÄ, 2000; GAUTHIER; POULIOT, 2003; RUTHERFURD-MARKWICK; GILL, 2005), seu estudo é também bastante difundido, principalmente nos dias atuais, com descobertas de novos usos (WOLZ; MERSCH; KULOZIK, 2016). Cerca de 20% de todas as proteínas do leite bovino são classificadas como proteínas do soro; dentre elas destacam-se: β -lactoglobulina (β -LG),

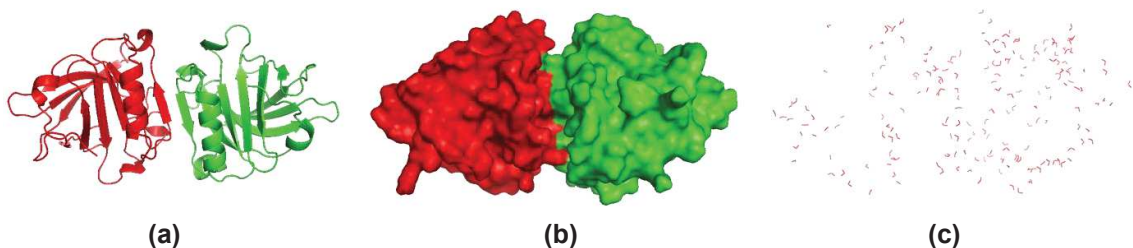
² Micelas são grandes agregados moleculares, com massa em torno de 10^8 Da, contendo cerca de 5000 moléculas cada. As estruturas formadas pelas caseínas contêm também íons Ca^{2+} , responsáveis pela manutenção do aglomerado (THOMPSON; BOLAND; SINGH, 2009; LAZZARO et al., 2017).

α -lactoalbumina (α -LA), imunoglobulina (Ig), soro albumina (SA), lactoferrina (LF), lactoperoxidase (LP) e lisozima (LSZ). Destas, a maior representante é a β -LG, responsável por 50% das proteínas do soro do leite (THOMPSON; BOLAND; SINGH, 2009). Neste texto, mesmo com toda a diversidade das proteínas do soro do leite, daremos destaque somente a três proteínas sendo elas a β -LG, a α -LA e a LF, pela sua abrangência e riqueza de propriedades físico-químicas.

β -lactoglobulina

A β -LG é uma proteína que ocorre em vários outros mamíferos além dos ruminantes, mas não é encontrada no leite humano. É a proteína do soro do leite mais abundante, representando 12% de todas as proteínas do leite bovino, com peso molecular de aproximadamente 18 kDa e 162 resíduos por monômero. Como já antecipado, pode estar presente no leite na forma de monômeros, dímeros ou tetrâmeros, sendo que o dímero ocorre no intervalo de pH entre 5,5 e 7,5. Desta forma, no pH natural do leite que é 6,6, a molécula pode ser considerada na forma de dímero, como observado na Figura 3. A β -LG possui dez variantes genéricas, de β -LG A até β -LG J, sendo que as formas mais comuns são as β -LG A e β -LG B (KONTOPIDIS; HOLT; SAWYER, 2004; THOMPSON; BOLAND; SINGH, 2009).

Figura 3 – Molécula de β -LG (código PDB 1BEB).



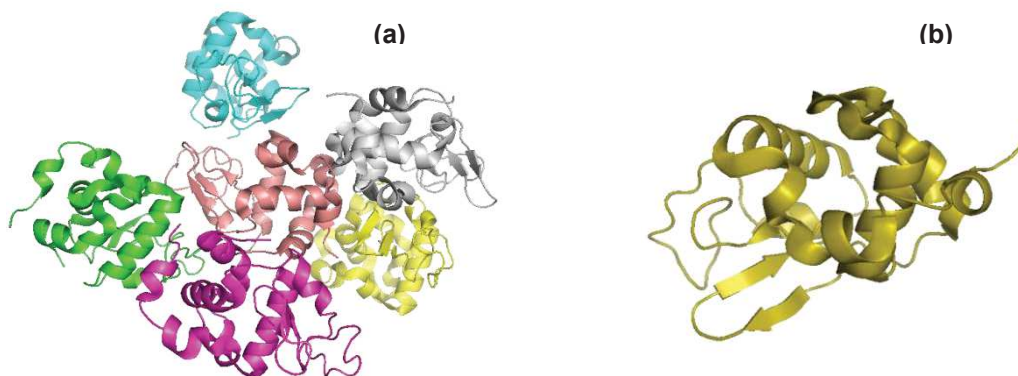
Estas representações moleculares, assim como todas as outras citadas neste texto, foram criadas pelo programa computacional de visualização molecular *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Em (a), representação em “fita” do complexo de duas cadeias de β -LG. As cores vermelha e verde representam as cadeias A e B. Em (b), as moléculas são apresentadas nas mesmas posições e mesmo esquema de cores, agora com a representação de superfície. Em (c), são apresentadas em vermelho todas as moléculas de água presentes no arquivo obtido do PDB.

A estrutura da β -LG é bastante estável e resistente à proteólise. Isso indica que a proteína não tem como principal função a nutrição, e sim outras funções biológicas no organismo. Uma das funções conhecidas da β -LG é o transporte de produtos, em função de suas propriedades de ligação, como por exemplo, vitaminas (KONTOPIDIS; HOLT; SAWYER, 2004). Especificamente no caso da vitamina A, pode ser transportada do estômago ao intestino, sendo que a presença da proteína impede a oxidação da vitamina. Da mesma forma, a β -LG pode se ligar a outras moléculas hidrofóbicas, atuando como transportadora. Outra importante função da proteína mais abundante do soro é se ligar a ácidos graxos, atuando como uma lipase (THOMPSON; BOLAND; SINGH, 2009; MCSWEENEY; FOX, 2013).

α -lactoalbumina

A α -LA compõe cerca de 20% das proteínas do soro do leite de vaca, e 3,5% do total de proteínas. Diferentemente da β -LG, a α -LA é a principal proteína no leite humano. Possui somente 123 resíduos, com uma massa de 14 kDa, e seu formato globular pode ser visto na Figura 4. Possui também mais de uma variante genética, mas somente uma delas é encontrada no leite bovino (THOMPSON; BOLAND; SINGH, 2009).

Figura 4 – Molécula de α -LA (código PDB 1F6S), representada em “fita”.



Representações criadas com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Em (a), temos a estrutura original obtida do PDB, sendo seis moléculas iguais juntas. Na figura, cada uma das seis cores representa uma das moléculas. Em nosso estudo de complexação, somente a primeira molécula (cadeia A) foi utilizada para os cálculos (b). O íon Ca^{2+} , o qual a α -LA se liga, não está representado nas figuras.

Esta é uma metaloproteína, pois contém um íon de Ca^{2+} em cada molécula, sustentado por quatro resíduos de ácido aspártico (ASP). O íon proporciona também grande estabilidade à proteína, principalmente em relação ao calor, caracterizando assim uma função biológica em detrimento da nutricional, da mesma forma com que ocorre com a β -LG. Porém, devido ao fato do íon estar ligado aos quatro resíduos de ASP, um pH abaixo de 5 (valor próximo ao pK_a do aminoácido) faz com que os resíduos se tornem protonados, perdendo assim a habilidade de ligação proteína-íon. Sem o íon a proteína torna-se mais sensível à temperatura, desnaturando-se e não voltando à forma inicial com o retorno da temperatura. Outra característica da α -LA é que ela é produzida nas glândulas mamárias, mas também pode ser encontrada no soro do sangue, diretamente relacionada a hormônios da gravidez; quanto maior a quantidade desses hormônios esteroidais no organismo, maior a quantidade de α -LA (THOMPSON; BOLAND; SINGH, 2009).

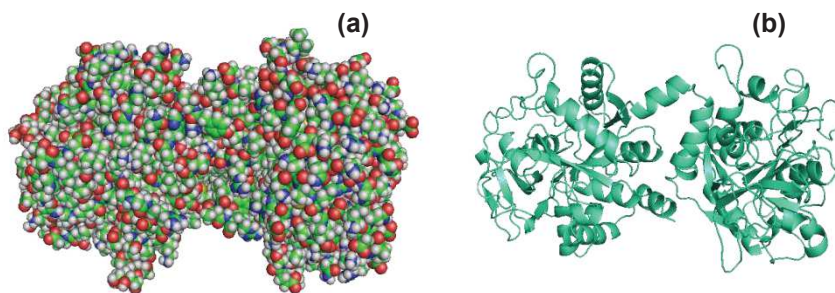
Lactoferrina

A LF é uma glicoproteína ligadora de ferro (possui dois íons Fe^{3+}), porém não apresenta a porção heme como no caso da hemoglobina. É uma proteína secretada pelas células epiteliais das glândulas mamárias, contudo, presente em diversos fluidos corporais além do leite, como saliva, lágrima, suor e sêmen. Assim como ocorre com outras proteínas do soro, como a Ig, a concentração de LF é bastante alta logo após o parto, para conferir maior proteção ao recém-nascido. Possui diversas funções biológicas em potencial, como melhorar a biodisponibilidade de ferro no organismo, sequestrar ferro de bactérias intestinais, tem ação antiviral, anti-inflamatória, antioxidante, imunomoduladora e anticarcinogênica (LÖNNERDAL; IYER, 1995; THOMPSON; BOLAND; SINGH, 2009; EL-LOLY; MAHFOUZ, 2011; TAKAYAMA, 2012; WAKABAYASHI et al., 2014). Com todos esses atributos, a LF torna-se uma proteína com grande potencial terapêutico, alvo de diversos estudos.

No leite de vaca, a lactoferrina juntamente com a lactoperoxidase e peptona compõem 0,23% das proteínas do soro. Já o leite humano contém alta concentração de LF (20% do total de nitrogênios), tornando-se interessante fortificar fórmulas infantis (baseadas em leite de

vaca) com a proteína. A LF possui um alto pI^3 , fazendo com que a proteína tenha carga positiva no pH natural do leite (6,6), enquanto a maioria das outras proteínas possui carga negativa neste mesmo pH. Isso faz com que a molécula tenha alta capacidade de se ligar eletrostaticamente a moléculas aniônicas (TAKAYAMA, 2012). As LFs bovina e humana têm 69% de homologia, com a humana possuindo três resíduos a mais (692, sendo 689 para a bovina). É uma proteína globular contendo dois domínios, como pode ser observado na Figura 5, e cada domínio é subdividido em dois subdomínios, onde se aloja um íon metálico (TAKAYAMA, 2012). Talvez uma das mais interessantes características da LF seja sua habilidade de formar complexos em função de suas características físico-químicas, como já visto desde 1971 (HEKMAN, 1971).

Figura 5 – Molécula de LF, (código PDB 1BLF).



Representações criadas com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Em (a), cada átomo da molécula está representado por uma esfera. As esferas verdes correspondem a átomos de carbono, as azuis nitrogênio, as vermelhas oxigênio, e as cinzas correspondem aos átomos de hidrogênio. Em (b), mantendo as mesmas posições e tamanhos, observa-se a representação em “fitas”, evidenciando as estruturas secundárias da proteína, ou seja, diversos segmentos de α -hélices, agrupados em dois domínios.

2.1.3 Aplicações das proteínas do leite

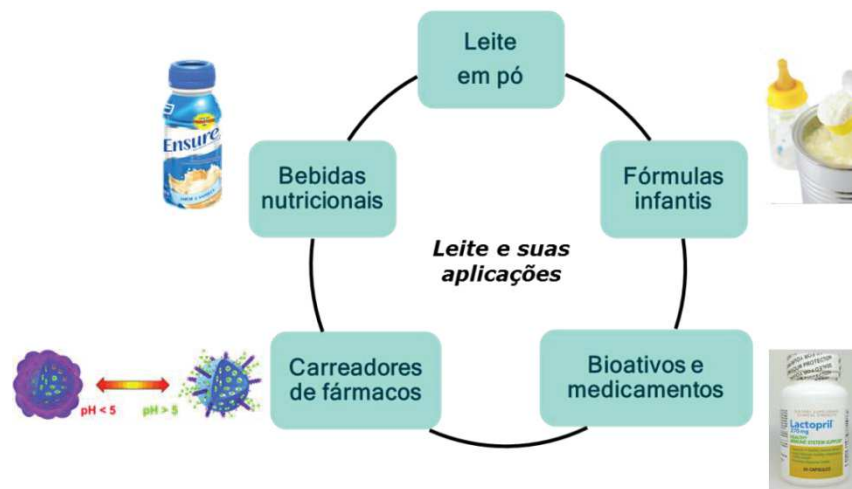
Proteínas em geral tem grande aplicabilidade na indústria, com as mais diversas características. Já há algum tempo a indústria de detergentes emprega enzimas em seus produtos, e estudos recentes ainda tentam melhorar esses produtos por modelagem molecular,

³ O pI ou ponto isolétrico é o pH em que a carga da proteína é nula, ou seja, a soma das cargas de todos os resíduos ionizáveis é zero (HITCHCOCK et al., 1931).

para aumentar a capacidade das enzimas de digerirem um número cada vez maior de partículas (VOJCIC et al., 2015). Como novos materiais, podem-se citar principalmente proteínas da soja e do leite, que por polimerização, podem fornecer novos materiais, com a vantagem de serem biodegradáveis (KAPLAN; MCGRATH, 2012); ou então as espidroínas, proteínas do fio da seda da aranha, que podem ser usadas como biomateriais, auxiliando o crescimento de células nervosas (ALLMELING et al., 2008) e cardíacas (TEPLENIN et al., 2015).

As proteínas do leite possuem várias aplicações em diversas áreas da indústria (BÖSZE, 2008; THOMPSON; BOLAND; SINGH, 2009; WATSON; ZIBADI; PREEDY, 2010; MCSWEENEY; FOX, 2013), incluindo odontologia (TENOVUO, 2002), medicina (VAN HOOIJDONK, 2000; PERMYAKOV et al., 2004; STEIJNS; TOMITA et al., 2009), biodefesa (CLARE; CATIGNANI; SWAISGOOD, 2003), alimentos (KINSELLA; WHITEHEAD, 1989; DALGLEISH, 1992; DE ROOS; WALSTRA; GEURTS, 1998; SYRBE; BAUER; KLOSTERMEYER, 1998; STEIJNS; VAN HOOIJDONK, 2000; DONALD, 2008; VARDHANABHUTI et al., 2009; DICKINSON, 2008, 2013), cosméticos (STEIJNS; VAN HOOIJDONK, 2000; SECCHI, 2008), farmacêutica (DICKINSON, 2008), entre outras indústrias (AUDIC; CHAUFER; DAUFIN, 2003). Na Figura 6 podemos ver outros exemplos de aplicações das proteínas do leite, como a proteína pura na forma de medicamento, fórmulas infantis acrescidas de proteínas, bebidas nutritivas, assim como carreadores de fármacos, que protegem o medicamento no pH estomacal. O crescente interesse em materiais biodegradáveis também reforça a necessidade de estudos relacionados às proteínas do leite, pois é um substituto viável aos polímeros derivados do petróleo (GHANBARZADEH; OROMIEHI, 2008; AZEVEDO et al., 2014; KHANZADI et al., 2015).

Figura 6 – Algumas das aplicações mais clássicas do leite e suas proteínas.



Lactoferrina purificada vendida como medicamento (Lactopril® – <http://lactopril.com/2011/05/09/lactoferrin-a-member-of-the-transferrin-family-of-iron-binding-glycoproteins/>); fórmulas infantis acrescidas de proteínas do leite de vaca, tornando-as mais completas (<http://cdn.womensunitedonline.com/articles/2011/06/baby-formula.jpg>); leite em pó em geral, com acréscimo ou diminuição de constituintes, de acordo com as necessidades do público; bebidas nutricionais, com funções biológicas em vários aspectos do organismo (http://www.farmacialeloir.com.ar/img/articulos/ensure_plus_sabor_vainilla_x_237ml_imagen1.jpg); carreadores de fármacos, onde micropartículas de proteínas protegem determinado medicamento até que ele chegue a seu destino final no organismo [Figura adaptada do trabalho de Guillet-Nicolas e colaboradores, de 2013 (GUILLET-NICOLAS et al., 2013)].

Focando na alimentação, um de seus usos é em bebidas com alta concentração proteica, assim como energéticos, que além de nutrir, tem a função de manter o produto estocado com fases mais estáveis (VARDHANABHUTI et al., 2009; RYAN; FOEGEDING, 2015). Em outros casos, a proteína pode ser introduzida em alimentos sólidos e semissólidos, como um ingrediente capaz de modificar a textura e proporcionar estabilidade (BATT; BRADY; SAWYER, 1994; DONALD, 2008; GIRARD; SCHAFFER-LEQUART, 2008; SAĞLAM et al., 2014). São também bons agentes estabilizantes de emulsões e gelificantes (KINSELLA; WHITEHEAD, 1989; DALGLEISH, 1992; HAVEA; SINGH; CREAMER, 2001; TCHOLAKOVA et al., 2006; DICKINSON, 2008, 2013; RYAN et al., 2012), com aplicação alimentícia ou cosmética. Já na saúde, exemplos clássicos são o desenvolvimento de drogas baseadas em sítios alvo de proteínas (JACOBSON, 2015), micropartículas formadas por proteínas, apresentando como uma das vantagens, uma maior afinidade com componentes celulares (OMORPHOS; KAHN; KALASKAR, 2015), ou proteínas que atuam diretamente

como drogas, como ocorre com a LF (PARODI, 2007) e as citocinas (MÜLLER, 2015), ambas com ação anticâncer.

Especificamente sobre as proteínas do soro, a exploração de suas propriedades moleculares pode contribuir para resolver deficiências tecnológicas que dificultam suas aplicações comerciais em grande escala (SMITHERS et al., 1996), sendo necessário para satisfazer a demanda de mercado (KINSELLA; WHITEHEAD, 1989). A pesquisa no desenvolvimento e manipulação de proteínas do leite, ou proteínas do leite interagindo com polissacarídeos, lipídeos ou outras proteínas pode ter um efeito lucrativo na economia e levar ao desenvolvimento de, por exemplo, sistemas inteligentes de encapsulação em nanoescala, para aplicações farmacêuticas e alimentícias (BATT; BRADY; SAWYER, 1994; DICKINSON, 2008).

Todos os exemplos citados estão diretamente relacionados ao estudo de proteínas. Assim, quanto mais e melhor pudermos entender seu comportamento e suas peculiaridades, teremos mais condições de usar essas moléculas e otimizar suas funcionalidades para aplicações específicas, assim como extrapolar o comportamento para sistemas semelhantes.

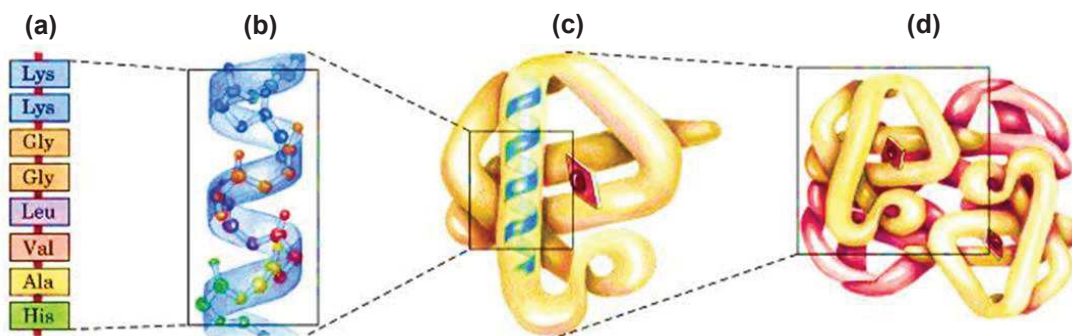
2.2 Proteínas

2.2.1 Aspectos químicos e físicos das proteínas

Ao lado dos ácidos nucleicos, carboidratos e lipídios, as proteínas compõem uma das principais classes de moléculas biológicas; quantitativamente, representa 50% do peso seco de uma célula (COZZONE, 2002). São compostas por quatro níveis estruturais, sendo (1) estrutura primária, correspondendo à sequência de aminoácidos unidos covalentemente, com um grupo N-terminal em uma extremidade e um grupo C-terminal na outra; (2) estrutura secundária, já organizada tridimensionalmente em α -hélices e folhas- β ; (3) estrutura terciária, o que corresponde ao arranjo tridimensional completo da cadeia, com os componentes da estrutura secundária; finalmente, (4) estrutura quaternária, que contempla proteínas compostas por mais de uma cadeia polipeptídica (NELSON; COX, 2002). Na Figura 7, pode-se visualizar cada uma das estruturas. Em (a), é apresentada a estrutura primária, podendo ser comparada a um segmento de um colar de pérolas; em (b), temos uma α -hélice, como

exemplo de estrutura secundária, formada pelo dobramento da estrutura primária; em (c), a união de vários segmentos de estrutura secundária forma a estrutura terciária, ou seja, o polipeptídeo. Finalmente em (d), a união de dois ou mais polipeptídeos forma uma estrutura quaternária.

Figura 7 – Esquema representando os quatro níveis de organização de uma proteína.



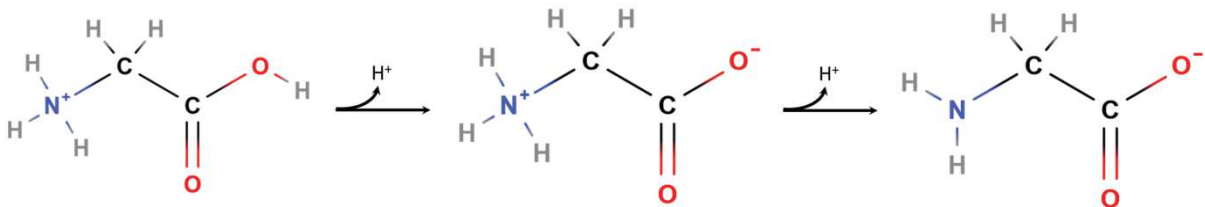
Esquema adaptado de Lehninger, edição de 2002 (NELSON; COX, 2002). Em (a), estrutura primária; (b), estrutura secundária; (c), estrutura terciária e (d) estrutura quaternária.

Outra categoria de proteínas são as proteínas estruturalmente desordenadas, caracterizadas por não apresentarem estrutura terciária estável (UVERSKY; DUNKER, 2010; UVERSKY, 2013). Muitas proteínas apresentam ao menos uma região desordenada (VAN DER LEE et al., 2014), sendo que as CNs (já citadas na seção anterior deste trabalho) compõem um exemplo clássico de proteínas intrinsecamente desordenadas, com uma típica formação de estruturas micelares (BÖSZE, 2008; THOMPSON; BOLAND; SINGH, 2009; JAIN; BHATTACHARYA; MUKHOPADHYAY, 2011).

A forma mais abrangente para se entender as proteínas é estudar seus aspectos físicos e químicos. A seqüência de aminoácidos é formada por blocos carregados positivamente, negativamente e neutros, de acordo com o perfil de protonação de cada resíduo de aminoácido e do meio. Na Figura 8 pode ser observado o perfil de protonação do aminoácido mais simples existente, a glicina (GLY), de forma isolada. Nele, em meio ácido, o aminoácido apresenta um grupo amino protonado (NH^{3+}) e carga $+1e$; com o aumento do pH do meio o hidrogênio da carboxila é perdido, adquirindo a molécula uma carga neutra, na forma de

zwitterion ⁴. Com o meio tornando-se ainda mais básico, um novo próton é perdido, agora do grupo amino carregado (NH_3^+), restando uma molécula carregada negativamente ($-1e$). Os pHs em que ocorrem as duas desprotonações são numericamente iguais aos pK_a 's dos aminoácidos. Porém, a proteína não apresenta resíduos isolados como ilustrado, e sim uma sequência de blocos interligados que, através de interações, são diretamente responsáveis pela estrutura tridimensional, carga líquida e demais propriedades da molécula, assim como sua interação com outras estruturas (COZZONE, 2002). Essa propriedade dá-se pelas cadeias laterais dos aminoácidos, que podem apresentar comportamento ácido ou básico, semelhante ao já citado na Figura 8.

Figura 8 – Aminoácido isolado glicina.



Estruturas construídas com o auxílio da plataforma MolView. *MolView.org*. Na sequência ocorrem duas desprotonações, conforme aumenta o pH do meio.

Na Figura 9 são apresentadas as cadeias laterais de todos os resíduos que podem apresenta-se na forma ácida ou básica. Os resíduos ácidos têm cargas que variam de -1 a $0e$, pois podem perder um próton; já os resíduos básicos apresentam carga de 0 a $+1e$, já que tem a capacidade de ganhar um próton. Ainda que essas reações pareçam bastante simples e intuitivas, o acoplamento entre a titulação de múltiplos sítios presentes em uma proteína torna esse processo intrinsecamente complexo. Essas reações são de suma importância, por serem as responsáveis pelo comportamento ácido-básico das proteínas, e conseqüentemente um dos pré-requisitos para o entendimento deste trabalho. O fato de muitos resíduos de aminoácidos serem carregados, faz com que mudanças no pH do meio altere a ionização desses grupos, permitindo variações drásticas na carga total da molécula, podendo levar inclusive à desnaturação da proteína (ENGLAND; HARAN, 2011), como já citado no caso da α -LA, que

⁴ Molécula eletricamente neutra, porém apresentando cargas opostas em diferentes pontos na mesma estrutura. O conceito aplica-se tanto para estruturas pequenas, como um aminoácido simples, como para estruturas poliméricas (LASCHEWSKY, 2014).

desnatura-se com facilidade após a perda do íon Ca^{2+} . Alterações nas concentrações de eletrólitos do meio, da mesma forma, podem provocar sérias variações na conformação molecular, afetando a função (KAUZMANN, 1959). A desnaturação também tende a expor resíduos hidrofóbicos, os quais acabam contribuindo para agregação proteica, não raramente indesejada.

Figura 9 – Aminoácidos de cadeia lateral com capacidade para alterar a carga líquida da molécula, conforme o grupo químico.

Resíduo	Forma ácida (doador de próton)	Forma básica (aceptor de próton)	pKa
Glu, Asp	$\text{R}-\text{COOH}$	$\text{R}-\text{COO}^-$	4,4 4,0
Lys, Arg	$\text{R}-\overset{\text{H}}{\overset{+}{\text{N}}}\text{H}_2$	$\text{R}-\ddot{\text{N}}\text{H}_2$	10,4 12,0
Cys	$\text{R}-\text{SH}$	$\text{R}-\text{S}^-$	9,5
His			6,3
Tyr			9,6

Esquema adaptado de Lehninger, edição de 2002 (NELSON; COX, 2002). Em rosa estão destacados os átomos de hidrogênio responsáveis pela variação de carga da cadeia lateral.

Os valores de pK_a referem-se à publicação de Nozaki e Tanford de 1967, sendo estes os valores de pK_a utilizados em nossas simulações (NOZAKI; TANFORD, 1967). Além das cadeias laterais citadas acima, os grupos N-terminal e C-terminal também tem a capacidade alterar a carga líquida da molécula, com pK_a 's de 7,5 e 3,8, respectivamente. Os valores de pK_a são referentes aos aminoácidos isolados, obtidos experimentalmente à temperatura de 25°C.

Já há muito tempo, as proteínas são estudadas por abordagens físicas, químicas e computacionais, além das ciências biológicas (SØRENSEN; LINDERSTRØM-LANG; LUND, 1927). Em 1952, Kirkwood e Shumaker estudavam as forças existentes entre as moléculas proteicas, considerando a eletrostática como a predominante (KIRKWOOD; SHUMAKER, 1952). Nos anos seguintes, vários trabalhos teóricos a respeito de titulação de proteínas foram publicados (TANFORD, 1957; TANFORD; KIRKWOOD, 1957; NOZAKI;

TANFORD, 1967). Inclusive a teoria quântica já foi empregada para a análise da distribuição de cargas dos resíduos de proteínas (DELRE; PULLMAN; YONEZAWA, 1963).

Considerando as características das estruturas proteicas já citadas, o entendimento de alguns fenômenos físicos são essenciais ao estudo dessas moléculas, como momento de dipolo, capacitância [ou parâmetro de regulação de cargas (LUND; JÖNSSON, 2005)], interações de van der Waals e hidrofobicidade.

Momento de dipolo

Muitas moléculas, mesmo que não tenham carga líquida, apresentam um dipolo elétrico, onde o átomo mais polar tende a atrair elétrons do menos polar para si. Sua unidade de medida é o Debye, sendo que 1 D é igual a $3,336 \cdot 10^{-30} \text{Cm}$ (ISRAELACHVILI, 2010). Para moléculas grandes, como as proteínas, define-se o dipolo elétrico em função das cargas dos aminoácidos ionizáveis (ver Figura 9). O momento de dipolo (μ) é caracterizado em proteínas como um desequilíbrio na distribuição espacial das cargas, podendo variar com as alterações de pH do meio (TAKASHIMA, 1999). Em nosso trabalho, utilizaremos μ em unidades reduzidas⁵, típicas dessa área (DA SILVA et al., 2016). Tradicionalmente, o μ de uma proteína é calculado como o somatório vetorial dos produtos de cada carga da proteína e suas respectivas posições vetoriais, como dado na Equação 1,

$$\mu = \left| \sum_{i=1}^{N_p} z_i e \mathbf{r}_i \right| \quad (1)$$

tal que N_p é o número de cargas (resíduos carregados) na proteína, z_i é a valência ou número de carga de cada aminoácido (adimensional), e é a carga elétrica elementar ($1,602 \times 10^{-19} \text{C}$) e \mathbf{r}_i é a posição do aminoácido, em unidades de comprimento. Neste trabalho empregamos unidades reduzidas (em função da carga elementar) e μ é tratado como o *número* do momento de dipolo e μ_p (*número* do momento de dipolo da proteína)

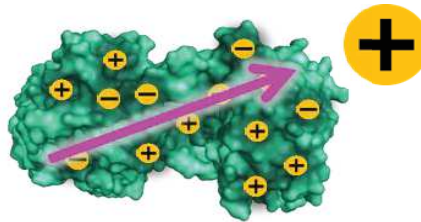
⁵ O emprego das unidades reduzidas é uma forma de simplificação. Neste caso, ao invés de representar o dipolo em unidades de D, apresentamos em unidades de comprimento (Å), mais usuais para caracterizar o estudo de proteínas em simulações computacionais.

$$\mu_P = \left| \sum_{i=1}^{N_P} z_i \mathbf{r}_i \right| \quad (2)$$

sendo a unidade resultante o comprimento, adotado em Å. A posição \mathbf{r}_i é um referencial arbitrário, mas escolhemos como referência o centro de massa da proteína.

Momentos de dipolo são responsáveis por parte das forças necessárias à estabilização de proteínas (BHARTI, 2014; SINGH et al., 2014). Podem ocorrer diversos tipos de interações, como dipolo-dipolo, em que polos opostos de uma mesma proteína ou proteínas diferentes se atraem; em dipolo-dipolo induzido, há um polo preexistente e outra região é induzida a se polarizar para combinar com a primeira, formando aí uma interação. Um exemplo dessa interação pode ser observado na Figura 10, em que uma carga externa afeta o dipolo de uma proteína. Por último, existe a interação dipolo induzido-dipolo induzido, pois quando dois átomos ficam próximos, cada um induz um dipolo no outro, por polarização da nuvem eletrônica (OZDAL; CAPANOGLU; ALTAY, 2013). Aqui, a única forma que uma carga externa afetará a proteína será através dos efeitos dessa carga no equilíbrio ácido-base dos aminoácidos ionizáveis.

Figura 10 – Representação de proteína com momento de dipolo afetado por uma carga externa.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000).

Forças de van der Waals (vdW)

Diferentemente de outras forças, que podem ou não estar presentes de acordo com as características da molécula ou do meio, as forças de vdW ou forças de dispersão estão sempre presentes, mesmo nas estruturas neutras (ISRAELACHVILI, 2010). Forças de vdW são forças de curto alcance, em relação à eletrostática, sendo que quanto mais próximas as

estruturas, maior a atração (ROTH; NEAL; LENHOFF, 1996). Estas interações estão intimamente relacionadas aos momentos de dipolo de moléculas orgânicas pequenas, já discutidos anteriormente, visto que os três tipos de interações entre dipolos são interações de van der Waals (OZDAL; CAPANOGLU; ALTAY, 2013).

Interações dipolo-dipolo são chamadas de forças de Keesom, dipolo-dipolo induzido são as forças de Debye, e dipolo induzido-dipolo induzido são as forças de dispersão de London. O meio o qual as moléculas estão inseridas pode influenciar nos dipolos, e conseqüentemente nas forças resultantes; dessa forma, pode-se dizer que o solvente tem grande importância na condução dessas forças, assim como para a eletrostática (ISRAELACHVILI, 2010). As interações de vdW entre aglomerados de átomos, como proteínas, ocorre como a soma das três contribuições (Keesom, Debye e London). Finalmente, a interação de vdW entre duas proteínas é a soma das interações entre cada um dos átomos (ROTH; NEAL; LENHOFF, 1996).

A maioria das análises das forças de vdW empregando corpos coloidais utiliza a aproximação simples de Hamaker (HAMAKER, 1937), que evidenciou que as contribuições de átomos individuais para a energia de dispersão global é aproximada pelo somatório de cada um dos pares. Como o μ considera a posição das cargas, o cálculo das energias de vdW entre proteínas depende do formato das moléculas, sendo que moléculas isotrópicas terão um padrão mais definido e fácil de calcular a energia. Já estruturas anisotrópicas, com reentrâncias e saliências, apresentam energias de vdW mais difíceis de serem calculadas (ROTH; NEAL; LENHOFF, 1996).

Parâmetro de regulação de cargas

O parâmetro de regulação de cargas (C), termo derivado da capacitância, mede a flutuação de cargas na proteína em função de sua interação com potenciais elétricos externos, evidenciado na Figura 11, onde cargas externas à molécula tem a capacidade de afetar a distribuição de cargas de uma proteína. A capacitância quantifica essa flutuação na proteína, sendo uma função intrínseca da proteína afetada pelo pH, concentração de sal, quantidade, tipo e disposição de resíduos carregados (LUND; JÖNSSON, 2005).

Figura 11 – Ilustração da alteração do estado de protonação de uma proteína próxima a uma superfície carregada, em função do potencial elétrico imposto por esta interface.

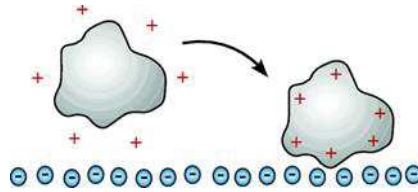


Figura adaptada de Lund, 2005 (LUND; ÅKESSON; JÖNSSON, 2005).

O mecanismo de regulação de cargas em biomoléculas evidencia que a carga total de uma proteína não é um valor único em um dado pH, mas pode flutuar em função de qualquer outro corpo carregado que esteja próximo (KIRKWOOD; SHUMAKER, 1952; LUND; ÅKESSON; JÖNSSON, 2005). Esse mecanismo é caracterizado pela migração de prótons da molécula, buscando uma adaptação de cargas a estruturas vizinhas, sendo inclusive um dos responsáveis pela interação entre proteínas (LUND; JÖNSSON, 2013). Para uma proteína, essa flutuação pode ser medida de forma relativamente simples pelo método Monte Carlo Metropolis

$$C_p \equiv \langle z_p^2 \rangle - \langle z_p \rangle^2 \quad (3)$$

onde C_p é a capacitância (ou parâmetro de regulação de cargas), intrínseca da proteína e z_p é o número de carga da proteína. Sendo C_p um valor específico da proteína, pode ser obtido por simulação computacional ou experimentalmente, a partir da derivada da curva de titulação da proteína (LUND; ÅKESSON; JÖNSSON, 2005; LUND; JÖNSSON, 2005)

$$C_p \equiv - \frac{1}{\ln 10} \frac{\partial z_p}{\partial \text{pH}} \quad (4)$$

Adicionalmente, pode também ser analiticamente estimada a partir do cálculo da capacitância de cada aminoácido,

$$C_i \equiv - \frac{10^{pH-pK_{a_i}}}{(1 + 10^{pH-pK_{a_i}})^2} \quad (5)$$

sendo que a capacitância ideal da proteína é a soma de todos os resíduos ionizáveis da mesma, juntamente com o C-terminal e o N-terminal.

Hidrofobicidade

Quando consideramos moléculas de soluto envoltas por um solvente, no caso a água, a disposição dessas moléculas está intimamente ligada a fatores entrópicos. As moléculas de água e do soluto se distribuem de forma a maximizar a entropia⁶. Ou seja, quanto menos moléculas de água forem necessárias para cercar o soluto, e mais moléculas de água estiverem livres no meio para formar ligações de hidrogênio, maior será a entropia do sistema, como pode ser observado na Figura 12 (ISRAELACHVILI, 2010). Desta forma, moléculas hidrofóbicas tendem a se agregar formando estruturas micelares, bolhas de ar, dímeros, entre outras estruturas, a fim de potencializar a entropia.

⁶ A entropia é uma medida do grau de desordem de um sistema. Quanto maior o valor da entropia, mais desorganizado e estável é o sistema (BERRY, 1995; CHANG, 2000; WEHRL, 1991).

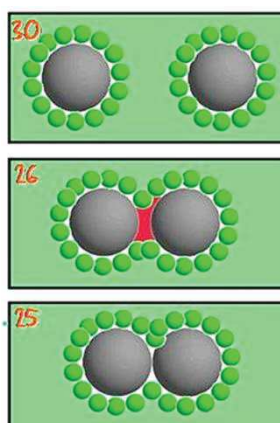
Figura 12 – Esquema simplificado do efeito hidrofóbico e maximização da entropia.

Figura adaptada de http://csb.stanford.edu/levitt/demo_lectures/lec1/Lecture1/Complex_interactions/pages/Explaining_Hydrophobicity.html. No primeiro quadro, observam-se duas moléculas hidrofóbicas (representadas em cinza), cercadas cada uma por 15 moléculas de água (representadas em verde), num total de 30 moléculas de água, necessárias para circundar (solvatar) as moléculas hidrofóbicas. Nos dois quadros seguintes, as moléculas hidrofóbicas se aproximam, fazendo com que moléculas de água sejam liberadas para o meio. No último quadro, as moléculas hidrofóbicas encontram-se na forma mais aproximada, e agora somente 25 moléculas de água são necessárias para solvatar a estrutura hidrofóbica formada. Em relação ao quadro inicial, cinco moléculas de água foram liberadas para o meio, estando livres para fazer ligações de hidrogênio.

Diferentemente do parâmetro de regulação de cargas, a hidrofobicidade de um aminoácido (que pode ser extrapolada a uma proteína), não é uma propriedade que pode ser medida tão facilmente. Uma das formas de ter a hidrofobicidade quantificada é medir a energia livre de transferência da cadeia lateral de um aminoácido da água para um meio apolar (EISENBERG et al., 1982; SALWINSKI; EISENBERG, 2003). Há várias escalas de hidrofobicidade, sendo que cada uma delas pode usar diferentes meios apolares. É possível calcular também momentos de dipolo hidrofóbicos, de uma sequência de aminoácidos, considerando sua distribuição de α -hélices e folhas- β (EISENBERG et al., 1982).

Acredita-se que este efeito seja um dos fatores responsáveis no processo de enovelamento de proteínas, sendo grande alvo de interesse (CHANPHAI; BEKALE; TAJMIR-RIahi, 2015).

2.2.2 Interação proteína-proteína

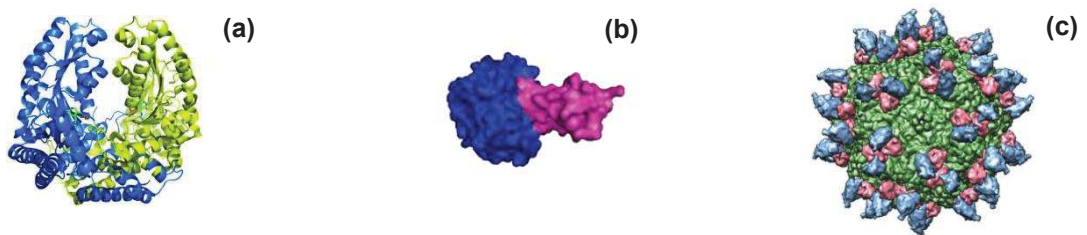
Forças intermoleculares são inevitavelmente explicadas a partir da eletrostática, tendo como fundamento a força eletromagnética (ISRAELACHVILI, 2010). Têm a capacidade de manter um complexo coeso, sendo diretamente responsáveis pelo correto funcionamento da maquinaria biológica em uma célula.

As proteínas consistem em uma das principais classes de moléculas biológicas, presentes em todos os processos vitais de um organismo. Elas podem interagir com diversos tipos de moléculas, entre elas, outras proteínas (ARKIN, 2005; RYAN; MATTHEWS, 2005; BIER et al., 2015). Essas interações podem apresentar um comportamento clássico [interação eletrostática coulombica entre cargas opostas (SHEINERMAN, 2000)], interações de carga-dipolo elétrico (DA SILVA; BOSTRÖM; PERSSON, 2014) ou interações mais elaboradas [como a descrita pelo fenômeno de correlação de cargas (LUND; JÖNSSON, 2005; JÖNSSON; LUND; DA SILVA, 2007a; ADŽIĆ; PODGORNIK, 2015)], assim como interações de van der Waals (vdW) (CHANPHAI; BEKALE; TAJMIR-RIABI, 2015). A análise desse comportamento para diferentes classes de proteínas é uma área muito promissora de estudos, visto a diversidade estrutural, funcional e de características físico-químicas dessas moléculas assim como a necessidade prática nas diversas áreas já citadas.

A interação proteína-proteína (contemplando regimes de atração e repulsão) é fundamental para a manutenção da vida nos processos biológicos de qualquer organismo. Uma das reações mais importantes que ocorrem entre proteínas é a formação de complexos (ZACHARIAS, 2010), ou seja, quando uma proteína se associa a outra, sem que ocorra uma ligação covalente. Uma proteína pode complexar íons, como o caso da calbindina ou calmodulina, que ligam íons de cálcio (FUJISE et al., 1995), ou a LF, que mantém em sua estrutura dois íons férricos com elevada afinidade (MOORE et al., 1997); pode interagir também com outra proteína idêntica (homo-associação), caso da β -LG, que em determinado intervalo de pH se mantém estável na forma de dímero (BROWNLOW et al., 1997). A interação pode ocorrer entre duas proteínas diferentes (hetero-associação), como entre a tripsina e seu inibidor BPTI (MARQUART et al., 1983), ou três proteínas diferentes, como no complexo entre a trombina e duas outras unidades distintas formando o inibidor hirudina (RYDEL et al., 1991); até mesmo um complexo formado por íons e proteínas idênticas e diferentes ao mesmo tempo, como o caso da hemoglobina, formada por duas subunidades

idênticas de uma molécula, mais duas subunidades idênticas de outra molécula, além da porção heme, onde se encontram os íons de ferro (PERUTZ, 1979). Pode-se também dizer que as estruturas complexadas mais fascinantes são os capsídeos virais, tamanha sua complexidade e organização (MACKENZIE, 1972; CHENG; BROOKS, 2013; FAUSTINO et al., 2014). Na Figura 13 estão ilustradas diferentes associações entre proteínas, sendo (a) uma homo-associação, (b) hetero-associação e (c) estrutura de um capsídeo.

Figura 13 – Esquema mostrando diferentes associações entre proteínas.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Em (a) encontra-se uma homo-associação, de duas proteínas idênticas (PDB id 3K40), em (b) uma hetero-associação (complexo tripsina-BPTI, PDB id 2PTC) e (c) parte da superfície de um capsídeo viral, com três tipos de proteínas, todas interligadas formando uma grande superfície (PDB id 3JCX).

O estudo dessas interações, sobretudo a definição das condições em que ocorrem as complexações, tem uma relevância cada vez maior. Informações acerca de complexos de proteínas são úteis em processos biotecnológicos, como métodos de diagnósticos em biologia molecular (TEMPLIN et al., 2002; CHERAK; TURNER, 2015; PERRIN et al., 2015); na área cosmética pode ser citado o caso da toxina botulínica, formada por um complexo de proteínas (BRYANT; CAI; SINGH, 2013), ou moléculas como a β -LG, atuando na estabilização de emulsões (CHANASATTRU; DECKER; MCCLEMENTS, 2009). O tópico de estabilização de formulações é bastante amplo, pois engloba tanto a área cosmética quanto as áreas farmacêutica e alimentar (SAĞLAM et al., 2014). Proteínas e seus complexos podem ser empregados com a função de estabilizantes e emulsificantes em formulações, substituindo surfactantes artificiais, potencialmente irritantes ao organismo (BOUYER et al., 2012). Várias proteínas tem essa propriedade, com destaque às proteínas do leite. As CN, por exemplo, são bastante eficazes como agentes emulsificantes, diminuindo a tensão interfacial como faria um surfactante, provocando um desarranjo na superfície do meio aquoso (DICKINSON;

DAVIES, 1999). Já as proteínas do soro do leite, destacadamente a α -LA e a β -LG, têm como principal característica suas cargas negativas, que por repulsão eletrostática mantêm as partículas das emulsões afastadas umas das outras (KIM; DECKER; MCCLEMENTS, 2002).

Complexos de proteínas do soro do leite

Proteínas e complexos de proteínas do soro do leite já foram bastante estudados por diversas técnicas experimentais, como filtração a gel (PUYOL et al., 1991), calorimetria (JU; HETTIARACHCHY; KILARA, 1999; RAEMY; LAMBELET; ROUSSET, 2005), espectrofotometria (MORAWETZ; HUGHES, 1952), microscopia de transmissão eletrônica (MUDGAL et al., 2011), fluorescência (NIGEN et al., 2009), eletroforese (SINGH; CREAMER, 2001; SEYREK; HATTORI; DUBIN, 2004; HATTORI et al., 2005; HAVEA; MUDGAL et al., 2011), espalhamento de luz (LI et al., 1996), observações microscópicas e de turbidez (NIGEN; CROGUENNEC; BOUHALLAB, 2009; RYAN et al., 2012; RYAN; FOEGEDING, 2015), cromatografia líquida de alta eficiência (TOPUZOĞULLARI et al., 2007), cromatografia por exclusão de tamanho (RYAN; FOEGEDING, 2015), potencial zeta (RYAN et al., 2012; RYAN; FOEGEDING, 2015), microscopia de força atômica (TOUHAMI; DUTCHER, 2009), titulação potenciométrica e turbidimétrica e espectroscopia de correlação de fótons (LAOS; BROWNSEY; RING, 2007). Sobre as interações a nível molecular, ainda se sabe pouco; há uma grande necessidade de se entender melhor esses eventos (DE WIT, 2009; WIJAYANTI; BANSAL; DEETH, 2014; RYAN; FOEGEDING, 2015; ZEILER; BOLHUIS, 2015) e poder aplicar essas informações em prol do mercado e da sociedade.

Um olhar atento na interação entre as proteínas revela comportamentos inesperados. Apesar das interações coulômbicas serem uma das mais importantes forças que governam a complexação de proteínas (TAKAHASHI, 1997; JÖNSSON; LUND; DA SILVA, 2007b), e tanto a α -LA quanto a β -LG poderem ter carga positiva e negativa no mesmo pH, a interação dessas moléculas com outras proteínas mostra algumas peculiaridades. Diferentes trabalhos demonstraram que a complexação com a β -LG ocorre, diferentemente das complexações com a α -LA (PUYOL et al., 1991). Usando filtração a gel, Puyol e colaboradores analisaram e compararam a interação da α -LA e da β -LG com ácido palmítico. Foi encontrado que a β -LG se liga ao ácido graxo, enquanto a α -LA não (PUYOL et al., 1991). Hekman também

levantou essas questões, especialmente sobre a habilidade da LF formar complexos facilmente, e quais forças atuam nessa complexação (HEKMAN, 1971). Outros estudos indicam que a LF se complexa com a β -LG, mas não com a α -LA (LAMPREAVE et al., 1990), mesmo sendo uma contradição às altas habilidades de ligação da LF.

2.2.3 Banco de Dados de Proteínas (PDB)

Para que uma simulação computacional clássica entre proteínas possa ser realizada, é necessária a descrição das moléculas em algum nível de resolução, que pode variar desde o nível atômico até mesmo como um corpo único carregado, passando por diversos níveis intermediários. O ponto de partida comum é o uso de coordenadas de cada um dos átomos que as compõem, para que seja construído um modelo baseado na estrutura tridimensional da macromolécula. Tipicamente, empregam-se dados de cristalografia de raio-X, e Ressonância Magnética Nuclear (RMN), as duas principais técnicas na obtenção experimental de estruturas. A diferença mais expressiva entre as técnicas é que a cristalografia de raio-X fornece coordenadas advindas de uma estrutura rígida, já a RMN provê coordenadas de uma estrutura em solução. Assim, algumas vezes pode-se escolher a estrutura que mais se adequa às necessidades de estudo, considerando-se também a resolução experimental das mesmas.

Desde o fim da década de 50, quando as primeiras coordenadas resolvidas por cristalografia de raio-X foram elucidadas, houve grande avanço e hoje um vasto número de estruturas está disponível. Para facilitar a utilização, todas as estruturas resolvidas encontram-se disponíveis em um banco de dados, com livre acesso. O chamado “Protein Data Bank” (PDB) (BERMAN et al., 2000) agrega todas as estruturas disponíveis, designando cada estrutura por um código de quatro caracteres. Cada código fornece as coordenadas de cada átomo da proteína, assim como de eventuais ligantes, solutos, águas cristalográficas e informações essenciais acerca da técnica utilizada (COUCH et al., 1995; WALLACE, 1996). No arquivo do PDB encontramos colunas designadas pelo tipo e número de cadeias, resíduos e átomos, assim como as coordenadas tridimensionais x, y e z de cada átomo, em Å, a ocupância, e o fator térmico. Ao final da sequência encontram-se linhas referentes a outras moléculas.

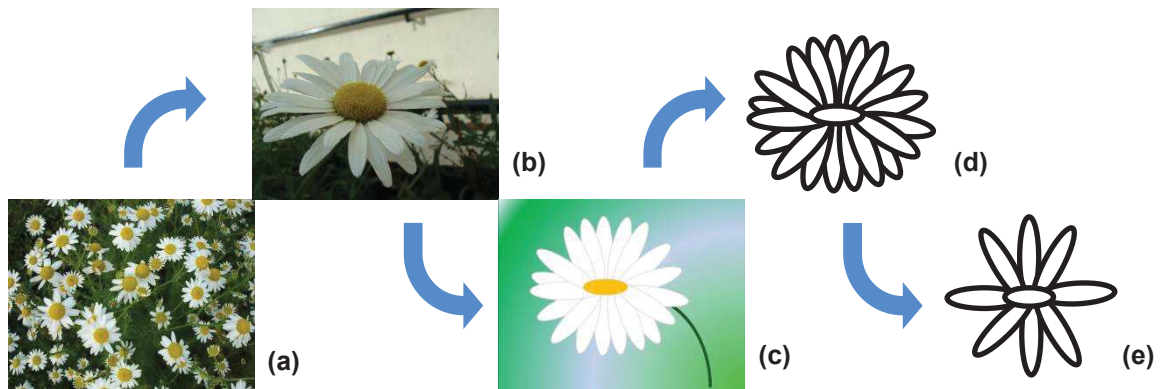
Em torno de duas décadas após o estabelecimento do PDB, o crescimento do número de estruturas resolvidas cresceu exponencialmente. Atualmente, há 122.583 códigos disponíveis no banco de dados, sendo 113.790 exclusivos de proteínas, segundo os dados do próprio PDB (BERMAN et al., 2000). Sabe-se, porém, que nem todos esses códigos pertencem a proteínas exclusivas, havendo redundâncias, com algumas proteínas resolvidas mais de uma vez, em função de melhores condições com o passar do tempo. Ainda assim, periodicamente novas estruturas estão disponíveis, provendo uma fonte gigantesca de matéria prima para estudos teóricos e experimentais.

2.2.4 Modelagem molecular

Um modelo é uma representação simplificada de um sistema físico real, mantendo apenas suas propriedades essenciais. As características a serem mantidas podem variar de acordo com o sistema, com o que se deseja analisar com esse modelo, com a precisão necessária e o custo computacional⁷ disponível. Na Figura 14, para ilustrar os diferentes níveis de analogia em um modelo, apresentamos um esquema de um modelo hipotético de margaridas, partindo da foto de um canteiro (a), até chegar ao esquema de uma flor genérica (e). No nível de complexidade mais alto (a), há muitas flores, interagindo umas com as outras e com o meio. Já no segundo nível (b) há somente uma flor, perdendo-se as interações existentes entre os indivíduos. A partir daí, as fotos nítidas perdem espaço para desenhos, até chegar a um esquema genérico. Os cinco níveis de complexidade adotados no esquema podem servir de analogia para os modelos moleculares. Da mesma forma que foi apresentado o modelo das margaridas, qualquer objeto de estudo pode ser tratado de forma semelhante, apresentando diferentes níveis de granularidade, escolhida em função do que se considera mais relevante para o que se quer estudar, com qual precisão e com quais recursos computacionais se tem disponíveis.

⁷ Custo computacional corresponde ao tempo de cálculo necessário para a obtenção dos resultados, de acordo com a complexidade do modelo utilizado, o número de passos necessário para cada simulação e as características do ambiente computacional (processador, memória, etc).

Figura 14: Criação de um modelo para flores, desde um canteiro até uma flor genérica.



Em (a), foto de um canteiro de margaridas; (b) foto de uma margarida; (c) desenho de uma margarida com alguns detalhes e cores; (d) desenho de uma margarida sem cores; (e) uma margarida simplificada, com poucas pétalas, podendo representar também outras flores. Entre (a) e (e) há vários níveis de complexidade, sendo que o modelo deve ser escolhido de acordo com as necessidades do estudo, assim como o nível de complexidade do resultado que se deseja obter.

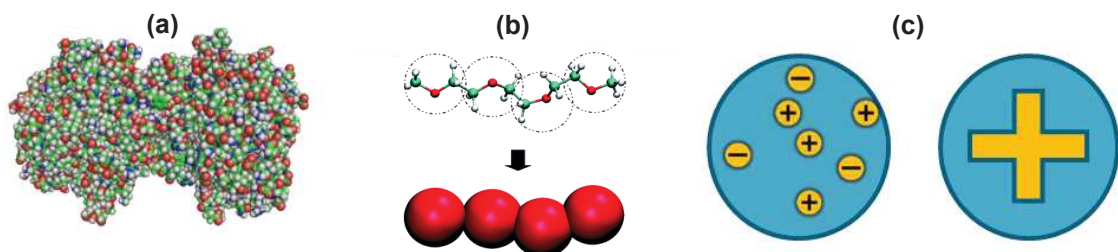
Existe uma diversidade de modelos teóricos com características e aplicações específicas (HALGREN; DAMM, 2001; MACKERELL, 2002; PONDER; CASE, 2003; TOZZINI, 2005). Modelos simplificados permitem discutir os aspectos considerados mais relevantes para um determinado sistema com um menor custo computacional. Oferecem também a possibilidade de direta comparação com teorias analíticas e a racionalização do processo que melhor se deseja compreender. A linha de modelos aqui empregados se iniciou com os trabalhos de Svensson e seus colaboradores (FUSHIKI et al., 1991; SVENSSON et al., 1993; JÖNSSON; SVENSSON, 1994; LUND; JÖNSSON, 2003), os quais aplicaram uma abordagem baseada em química coloidal para descrever as proteínas. Esses modelos foram gradativamente aperfeiçoados, com a incorporação de novos detalhes químicos para melhor reprodução de dados experimentais (LUND; JÖNSSON, 2003; DE CARVALHO; GHIOTTO; DA SILVA, 2006; TEIXEIRA; LUND; DA SILVA, 2010). A principal inovação foi a incorporação do processo de ionização dos aminoácidos, permitindo a realização de simulações a pH-constante (TEIXEIRA; LUND; DA SILVA, 2010), o que está na fronteira do conhecimento, visto que apenas recentemente a comunidade de teóricos começa a se preocupar em tornar as simulações mais realistas; algoritmos novos passaram a ser frequentemente publicados nos últimos anos e meses (JOHNSON; HOLT; ACKERS, 2009; SAKALLI; KNAPP, 2015; LEE; MILLER; BROOKS, 2016).

Há vários níveis de detalhamento possíveis na elaboração de um modelo molecular, os quais são determinados de acordo com as necessidades e características do sistema. O grau de granularidade está diretamente relacionado com o que se define como sítio (sendo átomos, conjunto de átomos, ou aminoácido). As interações físicas reais entre cada sítio do sistema são aproximadas e descritas por funções matemáticas conhecidas como “campos de força” (HALGREN; DAMM, 2001; NORRBY; BRANDT, 2001; MACKERELL, 2002; PONDER; CASE, 2003). São também chamadas de Hamiltonianas efetivas do sistema (DA SILVA, 2000). Os parâmetros dessas funções são obtidos através da reprodução de dados experimentais, sendo geralmente termodinâmicos. Dependendo da granularidade, em termos de solvente, o primeiro nível que pode ser destacado é o de Schrödinger, onde a definição sítio de interação (ou partículas) são os núcleos e os elétrons dos átomos (FRIEDMAN, 1977). Essa Hamiltoniana é extremamente detalhada, porém, tem a desvantagem de possuir um custo computacional proibitivo mesmo nos dias de hoje. Se compararmos com o esquema de modelo da Figura 14, pode ser associado ao primeiro nível de modelagem (a), que apresenta a foto do canteiro de margaridas, considerando todas as interações possíveis. O segundo nível de modelagem molecular é o nível de Born-Oppenheimer, também chamado de modelo microscópico, e tem como sítios de interação os centros dos átomos, tanto do soluto quanto do solvente (FRIEDMAN, 1977; DA SILVA, 2000). Este nível de modelagem é mais simplista em relação ao anterior, contudo, ainda é bastante custoso principalmente quando se necessita amostrar em várias diferentes condições experimentais. Na analogia com a Figura 14, Born-Oppenheimer pode ser classificado no terceiro nível, (c). Já o nível de modelagem de McMillan-Mayer, também chamado de modelo macroscópico, ou de contínuos de solvente, ou ainda de solvente implícito, é o de custo computacional mais acessível entre os citados. É derivado do nível de Born-Oppenheimer, porém com uma simplificação: as coordenadas e momento do solvente são substituídos por seu comportamento médio de blindagem das interações eletrostáticas (FRIEDMAN, 1977; DA SILVA, 2000). Solvente contínuo ou implícito é a situação onde o efeito médio do solvente é descrito por uma constante dielétrica. Assim, grande parte das partículas é eliminada, e o custo computacional torna-se mais acessível, mantendo a descrição do principal fenômeno. Na associação à Figura 14, McMillan-Mayer pode ser classificado no último nível, ou seja, (e), onde o detalhamento considerado é bastante baixo.

Para se modelar solutos (no caso, proteínas), há também várias alternativas, em diferentes níveis de granularidade, os quais vão desde modelos simplificados em rede até

descrição em detalhes atômicos, onde cada átomo é explicitamente descrito, tomando-se como base a estrutura cristalográfica (ou obtida por Ressonância Magnética Nuclear) da proteína (ZIELENKIEWICZ; RABCZENKO, 1988; TRAMONTANO, 2006; LEE; WU; ZHANG, 2009;). A proteína pode ser ainda modelada como um objeto único, com cargas pontuais em coordenadas definidas pela estrutura cristalográfica (Tanford & Kirkwood, 1957), ou, simplesmente, como um corpo único com uma carga líquida (DA SILVA, 2000). Modelos intermediários (ou mesoscópicos) também podem ser usados, descrevendo a proteína como uma coleção de esferas de Lennard-Jones com diferentes tamanhos e cargas que representam os seus aminoácidos (LUND; JÖNSSON, 2003). Na Figura 15 são ilustrados três níveis de modelagem para proteínas, desde o modelo atômico (microscópico) em (a), modelo mesoscópico em (b) e corpo único, em (c), chamado modelo macroscópico. Ao fazer uma analogia desses três níveis de modelagem aos níveis da Figura 14, o modelo microscópico pode ser associado a (a), nível de maior detalhamento. O modelo mesoscópico pode ser associado a (c), com um nível intermediário de detalhamento, e o nível macroscópico é comparado a (e), ou até mesmo a níveis de detalhamento ainda mais simplificados.

Figura 15 – Representação de três dos principais níveis de modelagem para proteínas.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Modelo atômico (microscópico) em (a), coarse grained (modelo mesoscópico), representado pela figura adaptada do trabalho de Prasitnok e Wilson (PRASITNOK; WILSON, 2013) em (b) e corpo único (modelo macroscópico) em (c), com várias cargas dispersas na superfície, ou contendo uma carga única. O modelo macroscópico é também chamado modelo de Tanford-Kirkwood (TK), ou seja, uma superfície esférica contendo cargas com localização definida.

Na Figura 16, obtida a partir do trabalho de revisão de Kmiecik e colaboradores (KMIECIK et al., 2016), pode-se visualizar de forma didática especificamente a diferença entre os níveis de modelagem atômico e “coarse-grained”. O nível de modelagem de

granularidade grossa facilita a amostragem, visto que a superfície energética é suavizada, fazendo com que os mínimos locais não dificultem a chegada ao mínimo global. Desta forma, uma estrutura proteica menos minuciosa é mais adequada a estudos que busquem resultados globais, ou seja, referentes a toda a proteína, e não de regiões específicas.

Figura 16 – Comparação de superfície energética para estruturas de nível de modelagem atômico e de granularidade grossa.

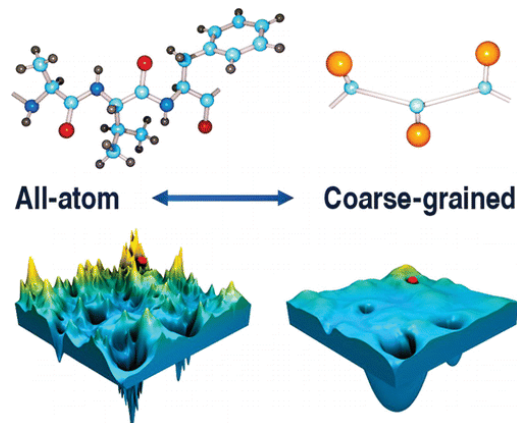


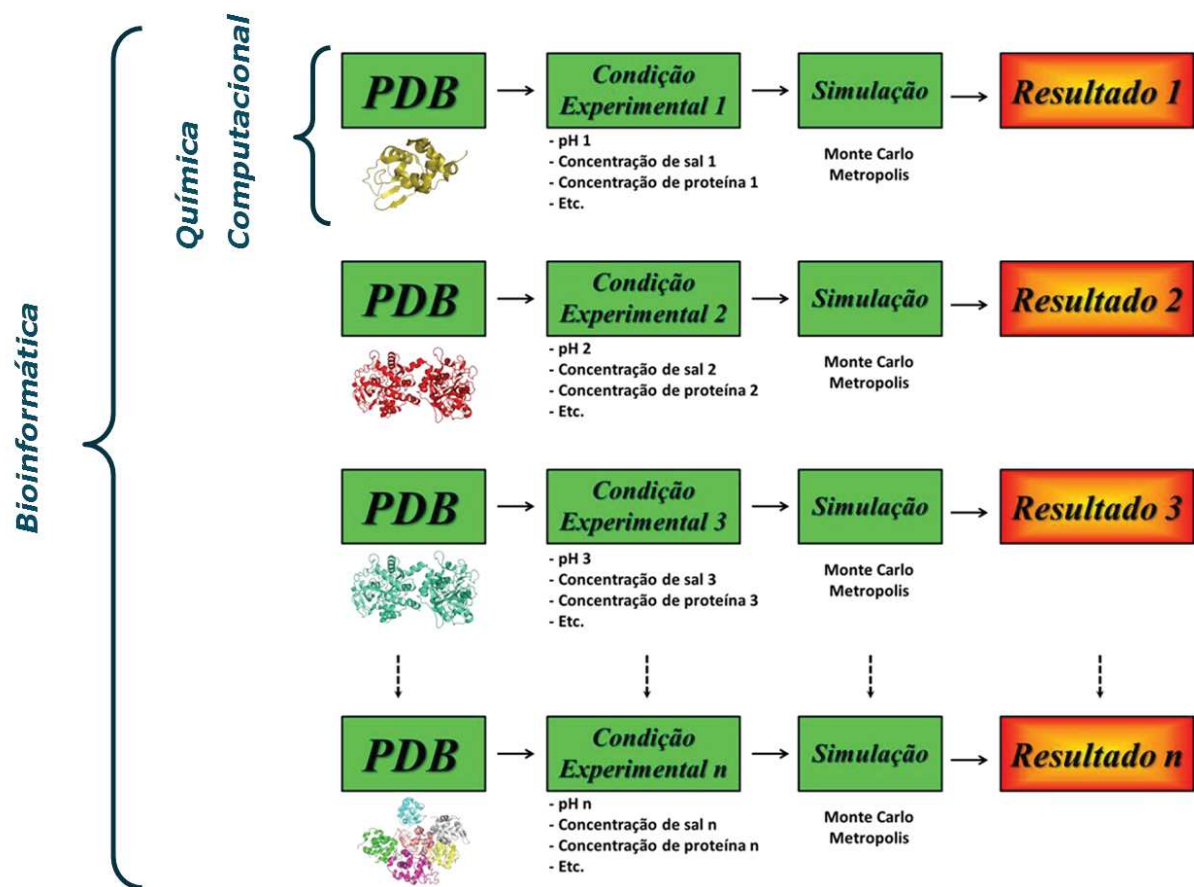
Ilustração obtida a partir do trabalho de revisão de Kmiecik e colaboradores (KMIECIK et al., 2016). Neste caso, quanto menor o detalhamento, mais claro é o resultado final.

Para que esses modelos possam ser resolvidos, algoritmos computacionais são essenciais. Inicialmente, esse processo era chamado de simulação computacional, e caracterizava uma área acadêmica conhecida como Química Computacional (ARAI, 1995; HOLDER; KILWAY, 2005). Hoje, pelo emprego de simulações em *larga escala*, a referência mais apropriada é Bioinformática Estrutural (BOGUSKI, 1994; MURRAY-RUST, 1994; HOFESTTIDT, 1995). A melhor e mais apropriada definição para isso foi cunhada por Mark Gerstein: “Bioinformática está relacionada com o entendimento da biologia em termos de suas moléculas (no sentido da físico-química) e aplicando técnicas de ‘informática’ para entender e organizar as informações associadas com estas moléculas, em larga-escala” (GERSTEIN; JANSEN, 2000; LUSCOMBE; GREENBAUM; GERSTEIN, 2001).

Com a abordagem em bioinformática, podemos não só resolver um problema em uma condição específica por vez, mas os cálculos podem ser sistematizados e otimizados para uma longa sequência de sistemas e condições experimentais, permitindo a obtenção de resultados em larga escala (ISHII et al., 2004; O’DRISCOLL; DAUGELAITE; SLEATOR, 2013; BOLOURI, 2014; SHIRAI et al., 2014), especialmente após o estabelecimento dos termos “omics” (BIEDA, 2012; GAHLAUT et al., 2013). Na Figura 17, temos um esquema, comparando as diferenças essenciais entre química computacional e bioinformática. Na

química computacional, uma molécula extraída do banco de dados de proteínas [PDB (BERMAN et al., 2000)] é submetida a condições experimentais específicas, e então efetuada a simulação computacional. Finalmente, um resultado específico para as condições escolhidas é obtido, sendo os cálculos feitos para casos individuais. Já em bioinformática, uma molécula extraída do PDB ou até mesmo todo o PDB pode ser simulado em muitas condições experimentais diversas. Podem ser variadas a temperatura, concentração de proteína, pH, concentração de eletrólitos, mutações, ou o que se fizer necessário. Em nossas simulações, as variações foram realizadas com diferentes valores de pH e concentração de eletrólitos na solução. Pode-se dizer que as duas abordagens se somam, visto que a química computacional dá maior ênfase a uma determinada condição específica, e a bioinformática permite uma visão mais abrangente, com diferentes níveis de análises.

Figura 17 – Esquema comparando química computacional e bioinformática.



A química computacional é englobada pela bioinformática, que pode simular condições experimentais diferentes simultaneamente.

O presente trabalho tem duas partes distintas e complementares. Na primeira parte, focamos no fenômeno da complexação; já na segunda parte, apresentamos novas justificativas para as aproximações intrínsecas dos modelos contínuos empregados. Em ambas as abordagens, fazemos uso intensivo de cálculos e análises computacionais em larga escala.

2.2.5 Correlação de cargas das proteínas

A questão de como modelar o interior de uma proteína em modelos contínuos, como os empregados aqui para a complexação de proteínas, é um tema polêmico na literatura (DEMCHUK; WADE, 1996; PENFOLD; WARWICKER; JÖNSSON, 1998; VICATOS; ROCA; WARSHEL, 2009; KUKIC et al., 2013). Tanford & Kirkwood (TANFORD; KIRKWOOD, 1957) introduziram o conceito de interface dielétrica ao proporem um modelo de duas constantes dielétricas para se estudar uma proteína em solução. No entanto, não há um critério objetivo da localização dessa interface e quais os valores de constante dielétrica deveriam ser utilizados. Dessa forma, acabam sendo empregados como parâmetros de ajustes (DE CARVALHO; FENLEY; DA SILVA, 2008).

Para melhor entender, discutiremos brevemente a estabilidade e o enovelamento proteico. As forças responsáveis pelo enovelamento e estabilidade das proteínas têm sido muito debatidas na literatura ao longo dos anos (HILL, 1955; TANFORD; KIRKWOOD, 1957; VICATOS; WARSHEL, 1981; DILL, 1990; AQVIST, 1991; MYERS; PACE, 1996; MIKI; KAKUYAMA; SODA, 1997; WARSHEL; WHITTEN; GARCÍA-MORENO E, 2000; CHI et al., 2003; ROCA; MESSER; WARSHEL, 2007; HILDEBRANDT et al., 2007; GVRITISHVILI; GRIBENKO; MAKHATADZE, 2008; PACE; GRIMSLEY; SCHOLTZ, 2009; ROCA; WARSHEL, 2009; JEFFERYS; KELLEY; STERNBERG, 2010; THIRUMALAI et al., 2010; STERPONE; MELCHIONNA, 2012). Hoje em dia parece haver um consenso de que as interações hidrofóbicas correspondem à maior interação responsável pelo menos no processo de enovelamento de proteínas (DILL et al., 2008). A importância da interação eletrostática direta é mais obscura tanto no enovelamento quanto na estabilidade das estruturas, mas ainda é frequente a afirmação de que estas interações são importantes nas proteínas (HILL, 1955; TANFORD; KIRKWOOD, 1957; PERUTZ, 1978; WARSHEL, 1981; DILL, 1990; WARSHEL; AQVIST, 1991; NISHIMURA; UVERSKY; FINK, 2001). Por exemplo, concentrando-se na estabilidade das proteínas, se as interações eletrostáticas são

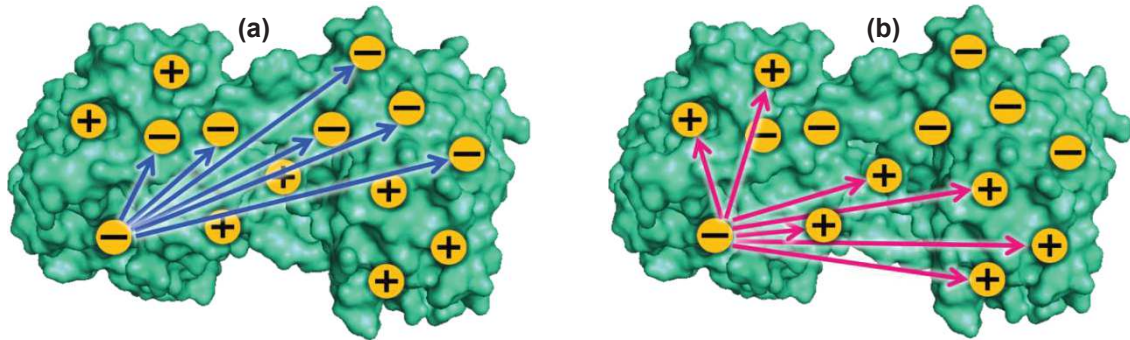
seu maior componente, seria de esperar uma distinta correlação entre grupos carregados nas estruturas terciárias da proteína. Isso foi investigado por Wada e Nakamura (WADA; NAKAMURA, 1981) que calcularam a energia de Coulomb entre cargas iguais e opostas, como uma função da distância de separação. As cargas nas proteínas foram definidas de acordo com certas regras e a energia de interação eletrostática foi calculada a partir do modelo de Tanford-Kirkwood (TK) (TANFORD; KIRKWOOD, 1957). Baseado em tal modelo, suas conclusões foram que cada carga é estabilizada por diversas unidades de $k_B T$ [k_B é a constante de Boltzmann ($1,3807 \times 10^{-23} \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$) e T é a temperatura em Kelvin]. Mais tarde, Barlow e Thorton (BARLOW; THORNTON, 1983) calcularam a distribuição das distâncias para os pares de íons, considerando os aminoácidos de cargas iguais e opostas, utilizando dados experimentais. Estes resultados indicam que resíduos carregados tem importância na estabilização da estrutura terciária de proteínas globulares. Infelizmente, os cálculos foram limitados a distâncias carga-carga bastante curtas, sendo menores de 10 Å, e somente 38 estruturas de proteínas foram estudadas. Este trabalho foi estendido por Gvritishvili e colaboradores (GVRITISHVILI; GRIBENKO; MAKHATADZE, 2008) para explorar as geometrias das pontes salinas em um conjunto representativo de 1500 coordenadas de proteínas, obtidas a partir do PDB. Suas conclusões reforçaram a importância das interações eletrostáticas na estabilidade da proteína. Reciprocamente, devido à anisotropia das cargas das proteínas, algumas proteínas apresentam uma preferência por ter cargas iguais próximas (DE VRIES; STUART, 2006; DE VOS et al., 2010). Outro conjunto de artigos de mesmo tema foi publicado por Warshel e colaboradores (ROCA; MESSER; WARSHEL, 2007; VICATOS; ROCA; WARSHEL, 2009). Empregando o método do dipolo semi-macroscópico de Langevin em sua versão de resposta linear aproximada, esses autores analisaram a contribuição eletrostática na estabilidade da proteína. Essas análises foram limitadas a um conjunto de dados contendo somente 45 estruturas de proteínas.

Hamiltonianas efetivas são aproximações, dando espaço para discussões e críticas. Esses modelos físicos de como modelar a resposta dielétrica do interior da proteína estão frequentemente sob intenso debate e opiniões controversas. Desde a introdução de uma interface dielétrica pelo modelo de TK citada anteriormente, muitos trabalhos discutiram o tema (KING; LEE; WARSHEL, 1991; WARSHEL; AQVIST, 1991; ANTOSIEWICZ; MCCAMMON; GILSON, 1994, 1996; SIMONSON; PERAHIA, 1995; DEMCHUK; WADE, 1996; SIMONSON; BROOKS, 1996; SHAM; CHU; WARSHEL, 1997; LÖFFLER; SCHREIBER; STEINHAUSER, 1997; PENFOLD; WARWICKER; JÖNSSON, 1998;

WARWICKER, 1999; DUDEV; LIM, 2000; SCHUTZ; WARSHEL, 2001; DA SILVA; JÖNSSON; PENFOLD, 2001; VARMA; JAKOBSSON, 2004; ARCHONTIS; SIMONSON, 2005; KO et al., 2005; HE; XU; PAN, 2007; DE CARVALHO; FENLEY; DA SILVA, 2008; VICATOS; ROCA; WARSHEL, 2009; KUKIC et al., 2013).

Nessa dissertação, para justificar o uso de um modelo uniforme para o tratamento do dielétrico das proteínas estudadas, realizamos uma análise no PDB para validar nossa escolha, estudando a correlação de cargas em estruturas de proteínas. Evitamos a construção de qualquer Hamiltoniana aproximada e baseamos nossos estudos nos dados estruturais experimentais diretos. Todas as estruturas terciárias de proteínas disponíveis no PDB foram usadas diretamente como fonte experimental, sem qualquer manipulação, sendo o maior conjunto de estruturas proteicas já analisadas em toda a literatura. Nós medimos as correlações de cargas de todas as estruturas, calculando a distribuição de probabilidade de cargas positivas e negativas em uma estrutura proteica em função de sua separação, r . As três distribuições são $P_{++}(r)$, $P_{--}(r)$, $P_{+-}(r)$, representadas na Figura 18. Em (a), é medida a distância entre uma carga fixa, e todas as outras cargas de mesmo sinal; em (b), é medida a distância da mesma carga, a todas as outras cargas de sinal oposto na proteína. E, o mesmo procedimento é repetido para todos os resíduos ionizáveis da proteína. Assim, nós usamos uma quantidade determinada experimentalmente, que é a distância de separação Euclidiana entre cargas de resíduos tituláveis como encontrado em coordenadas de proteínas disponíveis no PDB. Nós esperamos ver que $P_{+-}(r) > P_{++}(r), P_{--}(r)$ para pequenos r e o oposto para grandes r , ou seja, maior correlação de cargas de sinais opostos a curtas distâncias e de sinais iguais a longas distâncias. Nós também compararemos essas distribuições de cargas das proteínas para a função de correlação correspondente para uma solução eletrolítica simples confinada em uma esfera, que serviu de referencial teórico para embasar nossa interpretação. Nossos resultados serão apresentados no capítulo 6.

Figura 18 – Esquema do procedimento de correlação de cargas de uma proteína.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000). Em (a), medida a distância entre resíduo negativo e todos os outros de carga de mesmo sinal [$P_{--}(r)$]. Em (b), medidas as distância entre o mesmo resíduo e todas as outras cargas de sinal oposto [$P_{+-}(r)$].

3. Objetivos

Nossa meta principal é explorar e compreender os fundamentos físicos envolvidos na formação de complexos moleculares de proteínas do leite. Nossos objetivos específicos são:

- Obter um conjunto de informações sobre as interações intermoleculares envolvidas na formação de complexos moleculares, através de análises do *potencial de força média* [$w(r)$], para complexos entre as proteínas α -LA, β -LG e LF, em várias condições experimentais.
- Realização de um estudo de correlação de cargas em estruturas de proteínas, abrangendo todo o PDB, para discussão do conceito de constante dielétrica para modelar o interior de uma proteína.

4. Metodologia e modelo

O trabalho foi dividido em duas etapas: (a) complexação de macromoléculas (proteínas do soro do leite α -LA, β -LG e LF) empregando um modelo contínuo de solvente e (b) estudo da correlação de cargas das proteínas, tendo como material de trabalho todas as proteínas disponibilizadas pelo PDB. As duas etapas são complementares, sendo que o estudo da correlação de cargas visa justificar o tipo de constante dielétrica empregada para solvente e proteína no estudo de complexação das macromoléculas.

4.1 Método Monte Carlo Metropolis

Em comum às duas partes do trabalho, tanto na complexação molecular quanto na interpretação da correlação de cargas em estruturas de proteínas⁸, temos o método Monte Carlo (MC). MC é um método estocástico, ou seja, baseado na probabilidade de ocorrência de determinado evento. O título faz alusão ao cassino Monte Carlo em Mônaco, onde os jogos de azar são a grande atração. Analogamente, o método MC utiliza constantemente números pseudoaleatórios criados por computador, gerando soluções a integrais, resolvendo assim muitos problemas físicos e matemáticos (FISHMAN, 1999). Desde o surgimento do ENIAC (Electronic Numerical Integrator and Computer) em 1946 até os dias atuais, muita coisa na área da informática mudou, com o surgimento de máquinas cada vez mais poderosas, e o método MC ganhou poder de cálculo. Simulações moleculares de grandes estruturas proteicas, considerando suas coordenadas tridimensionais, já são realizadas com relativa facilidade (DA SILVA; JÖNSSON; PENFOLD, 2001; DE VRIES, 2004; DE CARVALHO; FENLEY; DA SILVA, 2008). Na simulação, os números pseudoaleatórios gerados correspondem a novas configurações, sendo que a cada “passo”, a estrutura pode estar em uma posição distinta. Quanto maior o número de “passos” ou ciclos, maior o espaço amostral percorrido no ambiente de simulação, conseqüentemente uma melhor estatística e resultados mais fundamentados. Uma limitação do método é a inexistência de trajetória, perdendo-se a obtenção de propriedades dinâmicas. Cada “passo” ocorre de forma independente dos demais

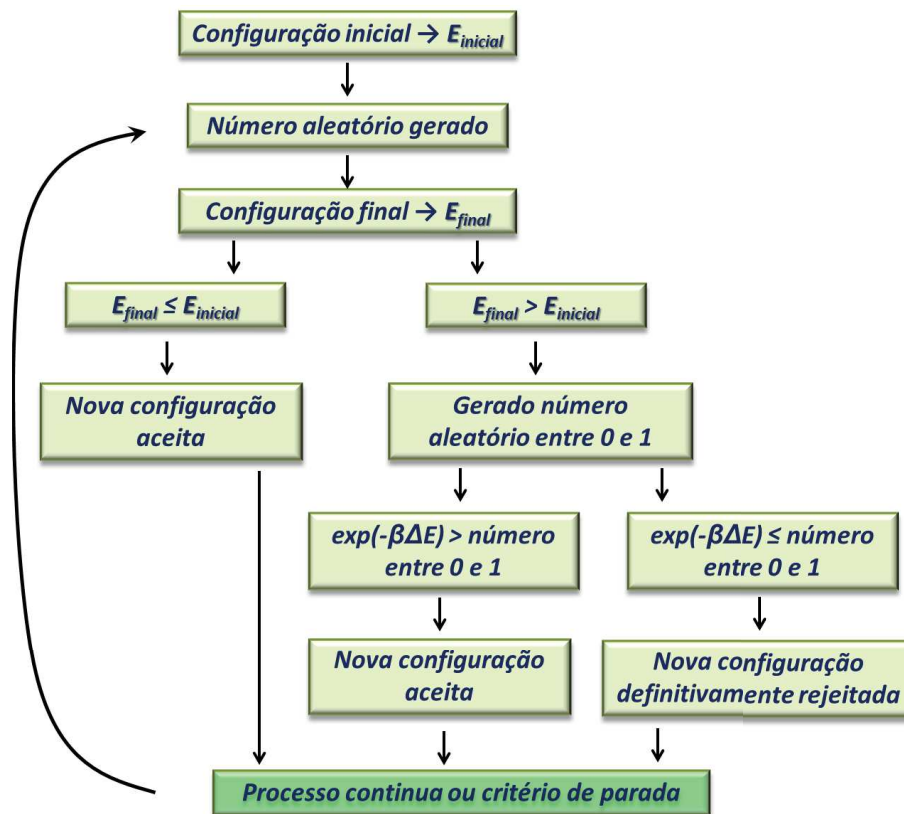
⁸ Note que apenas empregamos simulações no modelo teórico que serviu de referência para a interpretação do observado na análise, puramente experimental das correlações de carga.

e movimentos não físicos⁹ são permitidos, desde que satisfaçam aos critérios energéticos. Uma alternativa a essa técnica (quando em equilíbrio) é o método da Dinâmica Molecular (DM), um método determinístico, pautado na mecânica Newtoniana; ele pode reconhecer toda a trajetória de uma partícula, mas não será abordado nesse texto.

O método MC é ideal para simulações estatísticas, onde se avalia a energia das configurações, como é o caso de nosso trabalho. Na Figura 19 apresentamos um esquema simplificado do método MC, em que após estabelecer as energias final e inicial do sistema, a nova configuração, ou novo “passo”, só é aceito se houver diminuição de energia em relação à energia do passo anterior. Se a energia final for maior que a do passo inicial, então o critério de Metropolis é aplicado (METROPOLIS; ULAM, 1949). Neste critério temos o uso do fator de Boltzmann expresso por $\exp(-\Delta E/k_B T)$, onde ΔE é a energia resultante ($E_{final} - E_{inicial}$). Nele, toda vez que a energia final (E_{final}) for maior que a energia inicial ($E_{inicial}$), um novo número pseudoaleatório é gerado; se o fator de Boltzmann for maior que este número, a configuração é aceita. Ou, se o fator de Boltzmann for menor que o número gerado, a configuração é definitivamente rejeitada.

⁹ O termo “movimentos não físicos” refere-se a movimentos aleatórios que o sistema está sujeito, independente das leis da física. Os números pseudoaleatórios criados definem posições distintas no espaço amostral, mesmo que haja barreiras físicas.

Figura 19 – Esquema simplificado do método Monte Carlo Metropolis.



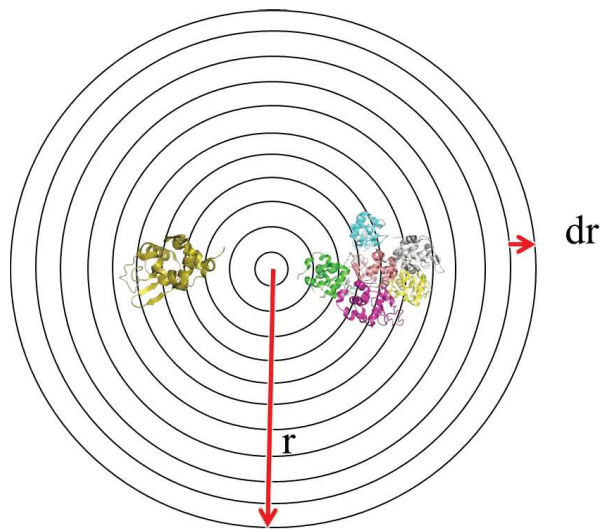
4.2 Potencial de força média

Nossos resultados finais para os estudos de complexação são baseados em valores de potencial de força média (PFM), podendo, resumida e simplificada, ser descrito como uma medida da maneira como partículas interagem em um meio $[w(r)]$. No contexto do trabalho, pode também ser chamado de energia livre de formação de complexos. Para chegar ao PFM ou energia livre de complexação, partimos da função de distribuição radial (FDR), também conhecida como função de correlação de pares $[g(r)]$.

A FDR descreve a variação da densidade de partículas em relação a um referencial. É calculada tridimensionalmente a partir de raias ou fatias com espessura dr , considerando a variação de volume (ΔV) e o número de partículas observadas em cada uma das raias, como pode ser observado na Figura 20. Em nosso trabalho, a FDR foi obtida em função da distância de separação r entre os centros geométricos das duas proteínas para todas as configurações

geradas segundo o critério de Monte Carlo Metropolis (MCM). Durante as simulações, essas diferentes posições das proteínas ao longo do eixo em que se encontram permitem a amostragem da frequência que cada distância de separação r foi observada. Isso é registrado em um histograma, o qual, devidamente normalizado, resulta em $g(r)$.

Figura 20 – Representação esquemática das “raias” para o cálculo da função de distribuição radial.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank – PDB, PDB id 1F6S) (BERMAN et al., 2000).

Através da Mecânica Estatística, podemos converter $g(r)$ para o PFM através da relação matemática:

$$w(r) = -k_B T \ln g(r) \quad (6)$$

evidenciando as unidades do PFM como $k_B T$. A relação entre $g(r)$ e $w(r)$ é inversa, ou seja, quando $g(r)$ é máximo, $w(r)$ é mínimo, e vice-versa. Quando negativo, quanto menor o valor obtido de $w(r)$, maior o potencial de interação de pares, ou seja, maior a interação entre as duas proteínas, levando a uma maior probabilidade de formação de complexo. Tipicamente, valores mínimos maiores que $1 k_B T$ negativo indicam a formação de complexos.

4.3 Complexação de proteínas do leite

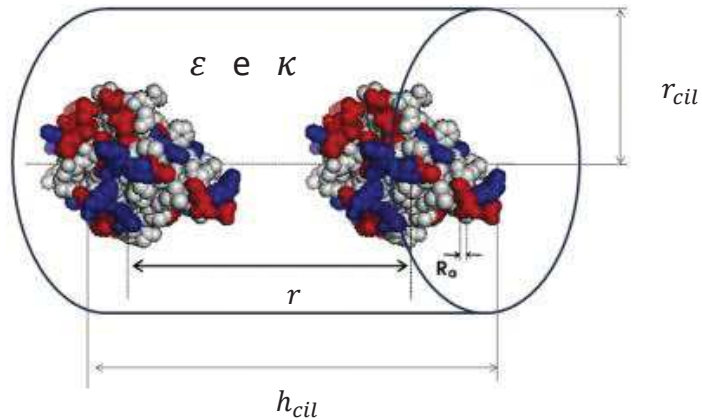
A mecânica estatística nos provê as bases teóricas para explorar as principais características da complexação de duas macromoléculas em uma solução eletrolítica (HILL, 1986). Os modelos “coarse grained” empregados juntamente com a mecânica estatística, permitem a exploração das principais características físicas de um sistema com uma quantidade reduzida de parâmetros. Desta forma, diversos modelos moleculares empregando as aproximações dos modelos contínuos de solvente foram idealizados para o estudo da interação proteína-proteína (LUND; JÖNSSON, 2003; KURUT et al., 2012; KURUT; DICKO; LUND, 2015) e resolvidos por simulação MCM (METROPOLIS et al., 1953; FRENKEL; SMIT, 1996;).

O modelo usado aqui é baseado em trabalhos prévios de outros autores (LUND; JÖNSSON, 2003; PERSSON et al., 2010; KURUT et al., 2012). As duas macromoléculas são simuladas considerando cada uma como uma coleção de esferas de Lennard-Jones (LJ) carregadas, com raios e valências definidos (R_{a_i} e z_{a_i} respectivamente) representando os aminoácidos, o chamado modelo mesoscópico. Estas estruturas são colocadas em uma célula aberta, cilíndrica e eletroneutra, de raio e comprimento definidos (r_{cil} e h_{cil}). Esta é uma versão simplificada do chamado modelo da célula (MARCUS; MARCUS, 1955; HILL, 1956; JÖNSSON, 1981), modificado convenientemente para melhorar a amostragem. As proteínas ficam livres para rotacionar como corpos rígidos (sem graus de liberdade internos) em qualquer direção, podendo também se deslocar para frente e para trás ao longo do eixo do cilindro. Para manter o sistema neutro, as cargas da proteína são neutralizadas na célula através da adição de contra íons, os quais assim como o sal adicionado ao meio, são modelados através do termo de blindagem $[exp(-\kappa r_{i,j})]$, onde κ é o inverso do comprimento de Debye e $r_{i,j}$ é a distância de separação entre as macropartículas.

Ao invés de moléculas de água explícitas, que resultariam em um custo computacional proibitivo para o estudo, foi adotado o modelo do dielétrico contínuo, ou seja, a água também é representada de forma implícita, sem uma estrutura molecular, através da sua constante dielétrica (permissividade dielétrica relativa), ϵ . O mesmo valor de ϵ é assumido para todas as espécies carregadas do sistema, como justificado no capítulo posterior (DA SILVA;

JÖNSSON; PENFOLD, 2001; DE CARVALHO; GHIOTTO; DA SILVA, 2006). O esquema do cilindro pode ser visualizado na Figura 21.

Figura 21 – Esquema de cilindro onde ocorrem as simulações de complexação.



No cilindro, raio e comprimento são definidos respectivamente por (r_{cil}) e (h_{cil}). As proteínas localizam-se internamente ao cilindro, a uma distância r uma da outra. ϵ e κ denotam a constante dielétrica do meio, e o sal implícito, respectivamente. Demais detalhes no texto (DELBONI; DA SILVA, 2016).

As coordenadas das proteínas são obtidas a partir do PDB e editadas antes dos cálculos. Tomando como exemplo a proteína α -LA bovina (código PDB 1F6S), ao ser extraída do PDB ela possui seis cadeias (Figura 4). Por ser encontrada na natureza na forma monomérica, as cinco últimas cadeias são deletadas, restando somente a primeira, ou seja, a cadeia A, que será empregada nos cálculos. Na natureza a proteína pode estar nas formas *apo* e *halo*, ou seja, sem e com o íon cálcio (Ca^{2+}), respectivamente. Para nossas investigações, utilizamos a α -LA em sua forma *halo*, mantendo assim o íon Ca^{2+} , presente na estrutura 1F6S, visto que sua ausência proporciona mudança configuracional da proteína (PEDERSEN et al., 2006). Além disso, a presença ou ausência do íon Ca^{2+} influencia na formação de complexos, como por exemplo, na complexação com a lisozima. Sem o íon, não há complexação (NIGEN et al., 2009).

A maioria dos cálculos de complexação foram realizados em pH 6,5, por ser bem próximo ao pH natural do leite, que é em torno de 6,6 (GIRARD; SCHAFFER-LEQUART, 2008). Neste pH, a β -LG é um dímero, como observado na Figura 3 (FOGOLARI et al., 2000; GOTTSCHALK et al., 2003). Assim, as duas cadeias da estrutura obtida por cristalografia de raio-X (código PDB 1BEB) foram usadas nos cálculos. Já as coordenadas da

LF (código PDB 1BLF) contém somente uma cadeia, que é a que usamos (Figura 5). Da mesma forma que para a α -LA, os dois íons ferro ($z_{Fe} = +3$) (FARRELL et al., 2004) presentes nas coordenadas do PDB foram incluídos nos cálculos de complexação. Em alguns casos, as coordenadas disponibilizadas pelo PDB apresentam alguns resíduos faltantes, que não foram adicionados em nossas simulações. Cargas fracionárias são atribuídas a cada esfera de LJ, de acordo com o pH da solução. As simulações ocorrem a pH constante, de modo que as cargas podem variar, devido às propriedades de equilíbrio ácido-base dos aminoácidos. Foi empregado um rápido esquema de titulação de prótons para permitir a titulação de resíduos ionizáveis, levando-se em conta as constantes de dissociação de cada aminoácido componente, ou seja, seu pK_a (NOZAKI; TANFORD, 1967). Mais detalhes podem ser encontrados na referência original (TEIXEIRA; LUND; DA SILVA, 2010).

Simulações adicionais foram realizadas com uma única proteína fixa no centro da célula, com a presença de eletrólitos em diferentes pHs, a fim de obter suas principais propriedades físico-químicas. Os resultados obtidos são comparáveis a cálculos empregando modelo atomístico e sal explícito.

A função de energia total do sistema é dada pela energia potencial Coulombica blindada [$u^{el}(r_{ij})$] e um termo de vdW [$u^{vdw}(r_{ij})$]. A $u^{el}(r_{ij})$ entre dois aminoácidos i e j é dada por

$$u^{el}(r_{ij}) = \frac{z_i z_j e^2}{4\pi\epsilon_0\epsilon r_{ij}} \exp(-\kappa r_{ij}) \quad (7)$$

onde ϵ_0 é a permissividade no vácuo ($\epsilon_0 = 8,854 \times 10^{-12} \text{C}^2/\text{Nm}^2$), z_i e z_j representam a valência de i e j respectivamente, e é a carga elementar, e r_{ij} é a separação entre elas. κ é o inverso do comprimento de Debye, sendo $[8\pi e^2 / (\epsilon_0 \epsilon k_B T) \sum_{ions} c_k z_k^2]^{1/2}$ onde c_k é o número de densidade da espécie k do eletrólito móvel (contra-íons e sal adicionado).

Uma energia potencial de LJ é usada para capturar os aspectos físicos fundamentais das duas proteínas excluindo seu volume (o termo repulsivo $1/r^{12}$ é devido ao princípio de exclusão de Pauli), uma interação média de vdW e uma contribuição efetiva do efeito hidrofóbico (estes dois últimos compõem o termo atrativo $1/r^6$). Interações hidrofóbicas específicas para descrever momentos hidrofóbicos, efeitos específicos de sal (Hofmeister), hidratação e ligações de hidrogênio são descritos apenas por aproximação. Somente as forças

principais são explicitamente descritas. Como é esperado para interações de vdW, os aminoácidos maiores atraem os outros mais fortemente que os menores, devido à dependência do tamanho de potencial, como mostrado na Equação 8. A parametrização do modelo é baseada em um argumento de Hamaker, e os parâmetros que utilizamos foram retirados de outras referências (PERSSON et al., 2010; KURUT et al., 2012;). Para quaisquer dois aminoácidos i e j , a energia potencial de interação de vdW é dada por

$$u^{vdw}(r_{ij}) = 4\varepsilon_{LJ} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (8)$$

onde ε_{LJ} assume o valor de $0,05005 k_B T$ ou $0,124 \text{ kJ/mol}$ (KURUT et al., 2012). Este valor corresponde a uma constante de Hamaker de cerca de $9 k_B T$ para pares de aminoácidos (LUND; JÖNSSON, 2003). O termo σ_{ij} ($= R_{a_i} + R_{a_j}$) é a distância de separação entre dois aminoácidos i e j em contato. Por exemplo, σ_{ij} para uma Alanina ($R_{a_{ALA}} = 3,1 \text{ \AA}$) e uma Arginina ($R_{a_{ARG}} = 4,0 \text{ \AA}$) é $7,1 \text{ \AA}$. Os tamanhos de aminoácidos também foram baseados em referências (PERSSON et al., 2010), onde os autores assumiram uma densidade de $0,9 \text{ g/mL}$ e usaram seu peso molecular para determinar um raio R_{a_i} adequado.

A energia de interação total do sistema para uma dada configuração é

$$U(\{\mathbf{r}_k\}) = u^{ele} + u^{vdw} = \frac{1}{2} \sum_{i=1}^{N_{aa}} \sum_{j=1}^{N_{aa}} [u^{ele}(r_{ij}) + u^{vdw}(r_{ij})] \quad (9)$$

onde $\{\mathbf{r}_k\}$ são as posições dos aminoácidos, $N_{aa} = N_{aa_{P1}} + N_{aa_{P2}}$ é o número total de aminoácidos das duas proteínas, $N_{aa_{P1}}$ é o número total de aminoácidos da proteína 1 e $N_{aa_{P2}}$ é o número total de aminoácidos da proteína 2.

A caixa de simulação cilíndrica possui um raio $r_{cil} = 100 \text{ \AA}$ e comprimento $h_{cil} = 500 \text{ \AA}$. A constante dielétrica e a temperatura da solução aquosa foram fixadas em $\varepsilon = 78,7$ e $T = 298K$, respectivamente. O pH da solução variou de 2 a 13. Os cálculos foram realizados

em diversas concentrações de sal (c_s): 5, 10, 20, 40, e 150 mM, correspondendo, respectivamente, a $1/\kappa$ igual a 43,0653; 30,4444; 21,5326; 15,2259 e 7,86261 Å. O número de passos MC para a equilibração e produção foram pelo menos 10^7 passos. Para quantificar especificamente as interações de vdW, algumas simulações foram realizadas apenas considerando todas as cargas como zero. Os parâmetros de deslocamento para translação e rotação foram ajustados para resultar em um aceite em torno de 30%. Devido às grandes barreiras de energia livre, maiores que $5 k_B T$ encontrado em alguns dos sistemas estudados, algumas amostragens foram obtidas a partir de múltiplas simulações, com restrições. Após uma simulação padrão inicial sem restrições, as grandes barreiras foram identificadas e usadas para definir as regiões críticas para uma amostragem adequada em simulações separadas. Nas rodadas subsequentes, a distância de separação entre as proteínas foi limitada em regiões específicas.

Durante as simulações, foram medidas propriedades das proteínas como a valência (z_p), a capacitância (C_p) [medida do mecanismo de flutuação de prótons (LUND; JÖNSSON, 2013)], e o momento de dipolo (μ_p). Para uma solução de pH diferente do pI, μ_p depende da origem das coordenadas. Para os cálculos, foi considerado o centro de massa da proteína. A FDR [$g(r)$] proteína-proteína foi amostrada com um histograma com fatia de 1 Å. A partir desta função foi obtido o PFM [$\beta w(r) = -\ln g(r)$] com baixo ruído, e usado para estimar a energia livre da complexação em cada condição físico-química. O ruído estatístico foi reduzido realizando-se múltiplas simulações independentes nas mesmas condições experimentais, e utilizando resultados médios.

A complexação pode ser estimada analiticamente baseada em uma teoria simples de Debye-Hückel (DH) para a energia livre (A_{DH}), onde a interação de cargas proteína-proteína é blindada por [$\exp(-\kappa r_{i,j})$]:

$$A_{DH}(r_{ij}) = \frac{z_i z_j e^2}{4\pi \epsilon_0 \epsilon r_{ij}} \exp(-\kappa r_{ij}) \quad (10)$$

Os valores para z_i e z_j são obtidos a partir de simulações de curvas de titulação, mas poderiam ser adquiridos também experimentalmente. Por uma questão de conveniência, assumiu-se uma distância de separação de 50 Å (estimativa para a distância de contato) para todos os cálculos A_{DH} . De acordo com Egan e colaboradores (EGAN et al., 2014), a energia

livre estimada é normalizada de 0 a 1 com a finalidade de preparar modelos de predições analíticas para as interações máximas ao longo dos pHs estudados. Após os cálculos de A_{DH} para todos os valores de pH pré-estabelecidos, o valor máximo observado ($A_{DH,max}$) foi usado para normalizar todos os outros ($A_{DH}/A_{DH,max}$). Foram utilizados somente os valores de pH em que é observado uma interação atrativa entre as proteínas.

5. Resultados e discussão

5.1 Propriedades físico-químicas de proteínas do soro do leite

Diferentes forças podem dirigir a complexação de proteínas, como a atração de Coulomb para proteínas opostamente carregadas, atração dipolar para moléculas com alto momento de dipolo, interações hidrofóbicas para proteínas com grande superfície hidrofóbica, ou as chamadas “interações eletrostáticas sensíveis”, quando a proteína tem uma grande capacitância devido ao mecanismo de regulação de cargas (DA SILVA; JÖNSSON, 2009; LUND; JÖNSSON, 2013). O conhecimento das propriedades eletrostáticas da proteína como sua carga [$q_p = \sum_{i=1}^{N_p} z_i e$, onde N_p é o número de cargas (resíduos carregados) na proteína] e o momento de dipolo (μ_p , tratado no item 3.1.1, Equação 2) em diferentes valores de pH e concentração de sal fornece uma visão inicial sobre os mecanismos de complexação, ao menos no regime Coulombico puro. Assim, é esperado atração entre proteínas com cargas de sinais opostos. Isso acontece quando o pH da solução está acima do pI de uma das proteínas e abaixo do pI da outra. Quanto maiores os valores de μ_p e C_p , mais forte será a atração. Seguindo um argumento simples da teoria de DH, o sal blindas as interações carga-carga por $\exp(-\kappa r_{ij})$. Para outras interações, o fator de blindagem será mais severo, no caso de interação dipolo induzido-dipolo induzido, $\exp(-2\kappa r_{ij})$, ou seja, o parâmetro de DH muda de acordo com a interação considerada.

As propriedades físico-químicas da α -LA, β -LG e LF no pI e em diferentes concentrações de sal estão listadas na Tabela 1.

Tabela 1 – Principais propriedades físico-químicas das proteínas estudadas.

Propriedade	α -LA (PDB id 1F6S)					β -LG* (PDB id 1BEB)					LF (PDB id 1BLF)				
	Concentração de sal					Concentração de sal					Concentração de sal				
	(mM)					(mM)					(mM)				
	5	10	20	40	150	5	10	20	40	150	5	10	20	40	150
C_p	1,2	1,2	1,3	1,4	1,7	3,3	3,6	4,1	4,7	6,3	4,0	4,5	5,2	6,0	8,0
μ_p	66	66	67	67	70	131	133	135	137	143	188	197	212	231	282
pI			5,0					4,8					9,6		
N_p			123					324					689		
R_p			28					42					53		

Capacitância e momento de dipolo das proteínas (C_p e μ_p , respectivamente) em diferentes forças iônicas para as proteínas investigadas nos pontos isoelétricos (pI). N_p e R_p são número de aminoácidos da proteína e o raio estimado em Å, respectivamente.

* A β -LG encontra-se na forma dimérica, em que são utilizadas as cadeias A e B.

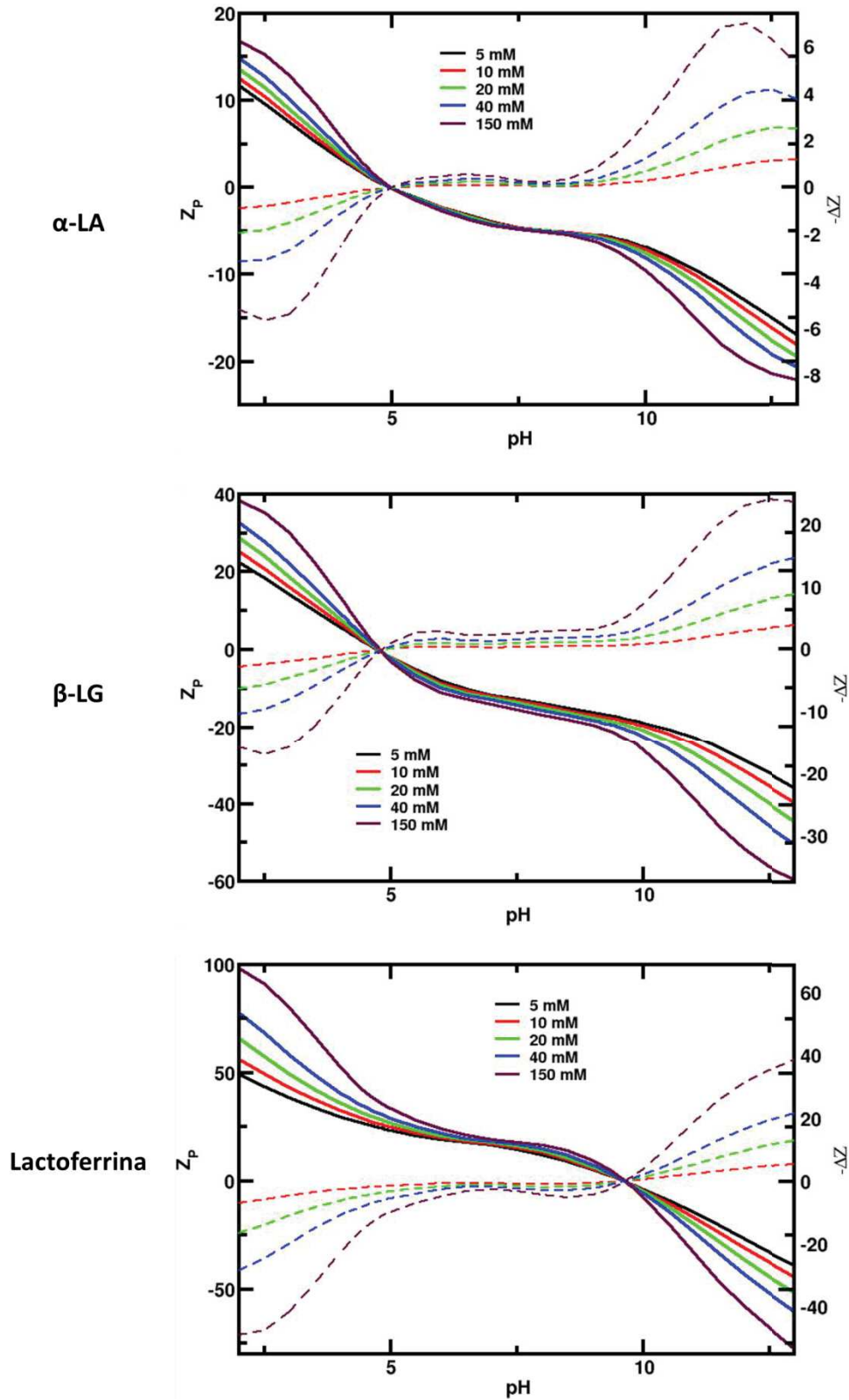
Estes dados são originados de simulações MC com uma única proteína em solução eletrolítica. Os valores experimentais de pI disponíveis na literatura podem variar muito como já mostrado por Medda e colaboradores (MEDDA et al., 2012), em parte devido a incertezas experimentais e também devido a condições variáveis, como força iônica, concentração de proteína, variações genéticas, dentre outras. Diferentes valores de pI tem sido citados para a α -LA bovina, por exemplo, 4,2 – 4,5 (JENG et al., 1997; FARRELL et al., 2004; THOMPSON; BOLAND; SINGH, 2009), 4,6 – 4,9 (MCSWEENEY; FOX, 2013), 5,0 (VALETTE; THOMAS; DÉJARDIN, 1999); já para a β -LG, 4,6 (GOTTSCALK et al., 2003), 4,8 (FOGOLARI et al., 2000), 5,1 para ambas variações genéticas (FARRELL et al., 2004), 5,2 (THOMPSON; BOLAND; SINGH, 2009; MCSWEENEY; FOX, 2013), 5,1 – 5,5 (MEDDA et al., 2012); finalmente, para a LF, 8,7 (TAKAYAMA, 2012), 8,8 (FARRELL et al., 2004) e aproximadamente 9,0 (THOMPSON; BOLAND; SINGH, 2009). Os resultados das simulações realizadas estão próximos dos valores medidos experimentalmente 5,0; 4,8 e 9,6 para α -LA, β -LG e LF, respectivamente.

A Tabela 1 mostra que os valores de pI para a α -LA (5,0) e β -LG (4,8) são bastante similares. Por outro lado, a β -LG tem um maior momento de dipolo (no pI e em baixa força iônica $\mu_{\alpha-LA} = 66 < \mu_{\beta-LG} = 131$) e uma maior capacitância (no pI e em baixa força iônica $C_{\alpha-LA} = 1,2 < C_{\beta-LG} = 3,3$) que facilita a interação atrativa com outras moléculas

carregadas. A LF é uma proteína altamente básica, ($pI = 9,6$), o que confirma sua tendência à atração Coulombica com moléculas aniônicas, como, por exemplo, a α -LA e a β -LG quando o pH da solução está entre 4,8 e 9,6. Também, devido a outras possíveis interações físicas, como íon-dipolo, mecanismo de regulação de cargas e vdW, a janela de pH pode ser inclusive maior. O momento de dipolo da LF é duas vezes o momento de dipolo observado para a β -LG em concentração fisiológica de sal. Esta proteína também tem uma alta C_P ($C_{LF} = 4$ em $c_s = 5 \text{ mM}$; $C_{LF} = 8$ em $c_s = 150 \text{ mM}$), resultados que sugerem que a LF é uma molécula que pode formar complexos facilmente com outros objetos carregados, justamente concordando com observações experimentais. A dependência do sal para o pI não é observada em nossos dados devido a aproximações no sistema de titulação escolhido. Simulações MCM foram realizadas com uma proteína com descrição atômica e um esquema de titulação com íons móveis explícitos (KESVATERA et al., 1999) mostrando variações de cerca de 0,3 unidades de pH entre 5 e 150 mM para a α -LA e a β -LG. O pI da LF se mantém insensível com o aumento da força iônica, enquanto o μ_P é substancialmente afetado, variando de 188 Å em $c_s = 5 \text{ mM}$ a 282 Å em $c_s = 150 \text{ mM}$ em nossos cálculos. Isso indica que um aumento na força iônica tende também a aumentar a interação atrativa íon-dipolo. A combinação desses altos valores de μ_{LF} e C_{LF} é responsável por facilitar a união da LF com outras moléculas.

Gráficos de carga (em unidades de carga elementar e) da proteína em função do pH em diferentes concentrações de sal são mostrados na Figura 22. Se por um lado os valores de pI não são virtualmente afetados pela adição de sal, como pode ser visto na figura, as diferenças nos números de carga das proteínas ($\Delta z_P = \langle z_{sal = 150 \text{ mM}} \rangle - \langle z_{ref = 5 \text{ mM}} \rangle$) podem ser muito grandes em alguns valores de pH. Comparando o sistema em 5 e 150 mM percebe-se diferenças maiores que 15, 38 e 72 em números absolutos, para a α -LA, a β -LG e LF, respectivamente. Pode ser observada a mesma tendência de aumento da carga da proteína com o aumento da força iônica da solução, como visto experimentalmente para a lisozima (KUEHNER et al., 1999).

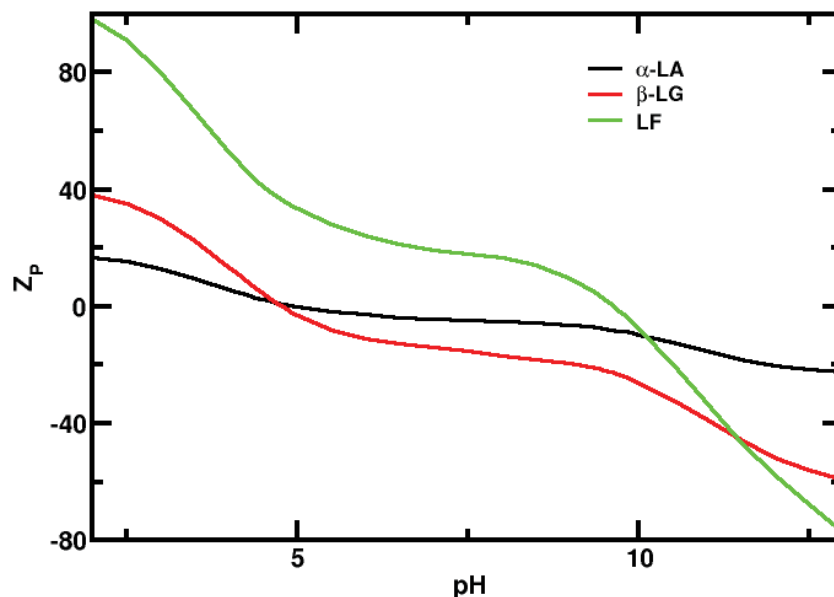
Figura 22 – Curvas de titulação das proteínas α -LA, β -LG e LF, em diferentes concentrações de sal.



Valores de número de carga ($\langle Z_p \rangle$) simulados para as proteínas, em função do pH em diferentes concentrações de sal. As diferenças entre as cargas nas diferentes concentrações de sal ($\Delta Z = \langle Z_{sal} \rangle - \langle Z_{Ref} \rangle$) também podem ser visualizadas no gráfico, no eixo da direita. O sistema referência foi o de 5 mM de sal.

Na Figura 23 pode-se observar as curvas de titulação referentes às três proteínas [pH 6,5 e concentração de sal fisiológica (150 mM)], agora sobrepostas, mostrando de forma evidente a disparidade de cargas de cada molécula. No caso da LF, a carga da proteína se mantém positiva por um intervalo de pH bastante grande em relação às outras proteínas, mostrando também uma maior abrangência de cargas ao longo do eixo dos valores de pH. Já a α -LA apresenta a menor variação de cargas em relação às outras proteínas do soro estudadas.

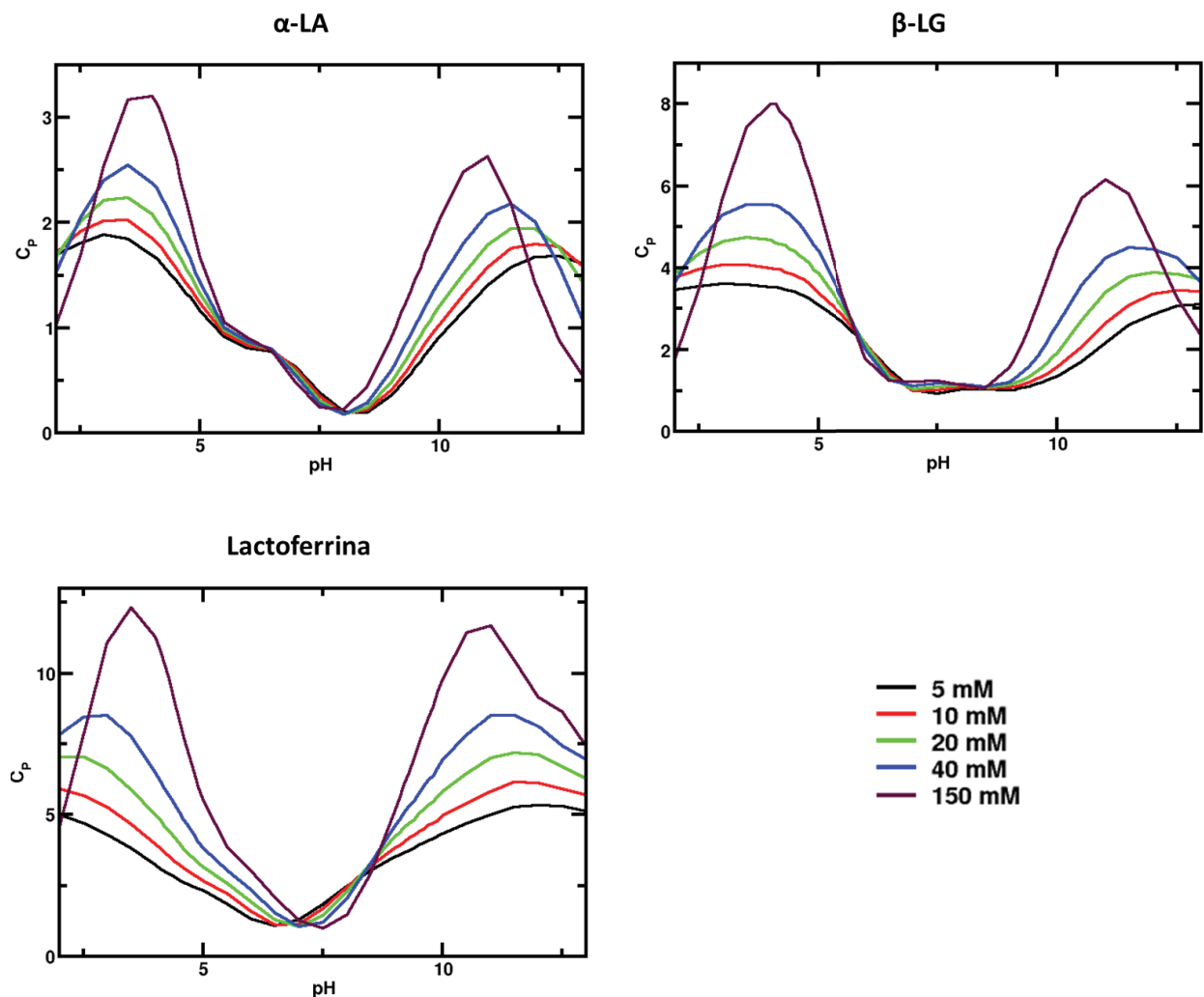
Figura 23 – Curvas de titulação sobrepostas das proteínas α -LA, β -LG e LF, com concentração de sal de 150 mM.



Os parâmetros de flutuação de cargas das proteínas variam com o pH em diferentes concentrações de sal, como dado na Figura 24. A capacitância varia com o pH como uma consequência do número de resíduos de aminoácidos que titulam ao redor de cada pH. As três proteínas do soro estudadas mostram qualitativamente um comportamento similar, ou seja, a C_p tem valores maiores nos extremos de pH, isto é, nos regimes mais ácidos e mais básicos, com um mínimo ao redor do pH 8. Em geral, a LF tem a mais alta capacitância, seguida da β -LG. O efeito do sal é mais pronunciado quando a capacitância tem valores mais altos, como no caso da LF. Nela, as diferenças na C_p são mais visíveis, mesmo para concentrações baixas e moderadas de sal. O comportamento geral observado é que o aumento da força iônica aumenta a capacitância da proteína, particularmente nos valores de pH onde há uma tendência

a um pico. Esse pico é a consequência de um grande número de certos resíduos com pK_a intrínseco similar ao pH da solução. Flutuações de cargas são maiores quando $pH \cong pK_a$. Uma proteína rica em ácido aspártico terá um pico no pH 4, por exemplo, visto que $pK_a^{asp} \approx 4,0$ (NOZAKI; TANFORD, 1967).

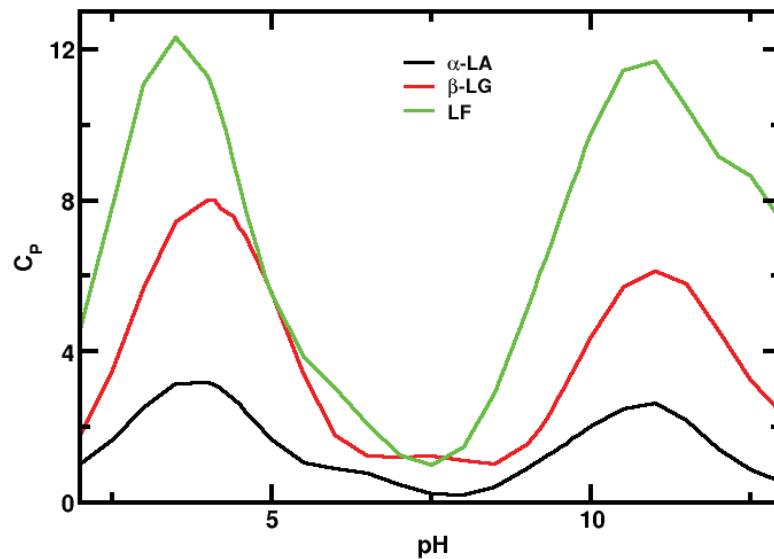
Figura 24 – Capacitâncias simuladas para as proteínas do soro do leite α -LA, β -LG e LF, em função do pH em diferentes concentrações de sal.



Na Figura 25 observa-se a sobreposição das curvas de capacitância em função do pH referentes às três proteínas [concentração de sal fisiológica (150 mM)]. Assim como ocorre com as cargas, a sobreposição de curvas mostra a LF com as maiores alterações na

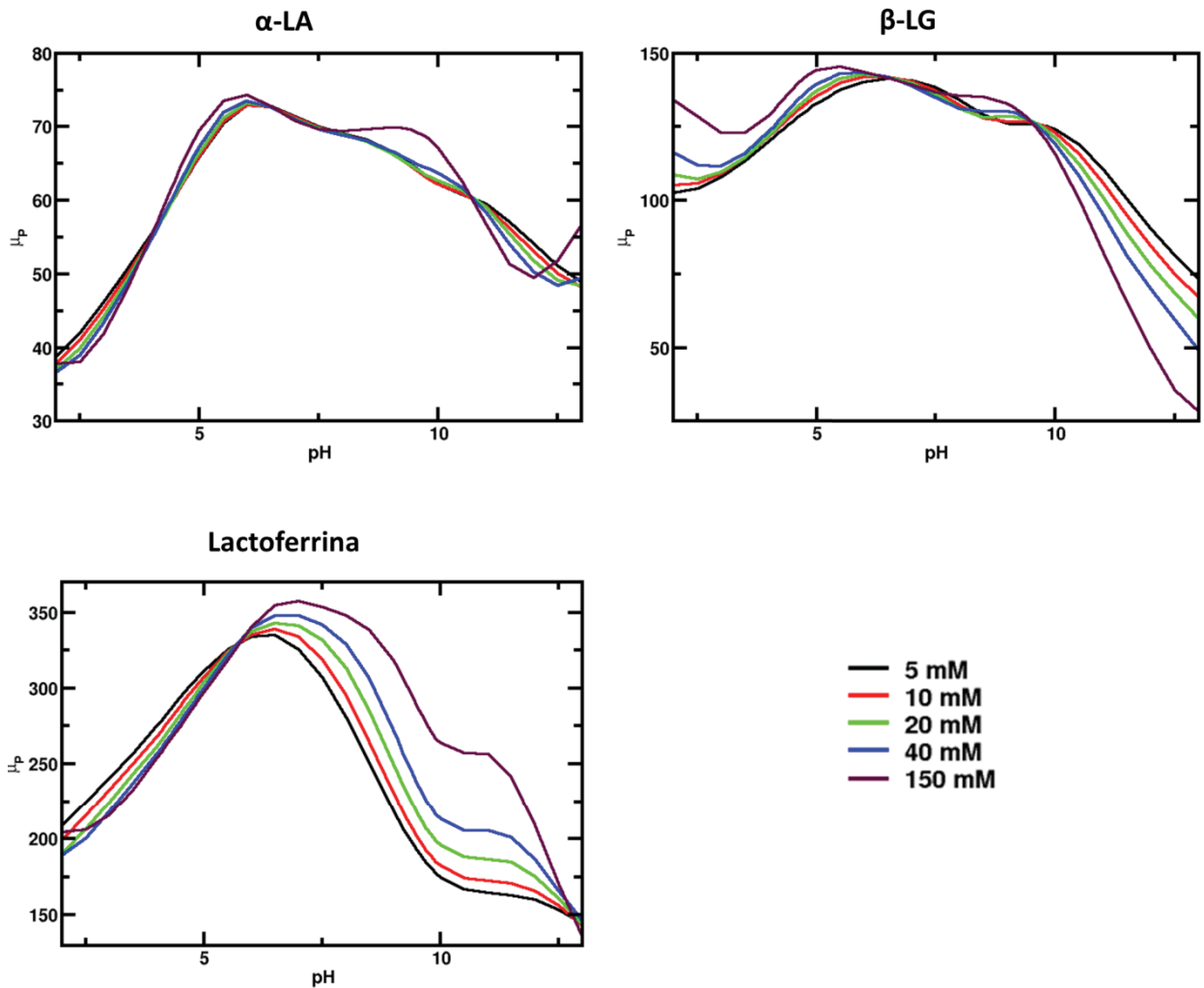
capacitância, e a α -LA apresenta-se como a proteína mais estável em relação ao pH, comparada às outras proteínas estudadas.

Figura 25 – Sobreposição das curvas de capacitância das proteínas α -LA, β -LG e LF, em concentração de sal de 150 mM.



Um gráfico similar para o momento de dipolo variando com o pH em diversas concentrações de sal é dado na Figura 26. Momentos de dipolo podem variar substancialmente com o pH para todas as proteínas estudadas. A α -LA tem o menor μ_p enquanto a β -LG tem valores intermediários e a LF os maiores. Considerando os resultados já discutidos sobre momento de dipolo e capacitância, pode-se concluir antes mesmo das simulações de complexação que a LF forma complexos fortes com a β -LG. O efeito do sal é mais pronunciado para valores de pH muito altos ou muito baixos. Cada proteína responde de uma maneira particular ao aumento da força iônica.

Figura 26 – Valores de momento de dipolo simulados para as proteínas do soro do leite α -LA, β -LG e LF, em função do pH em diferentes concentrações de sal.

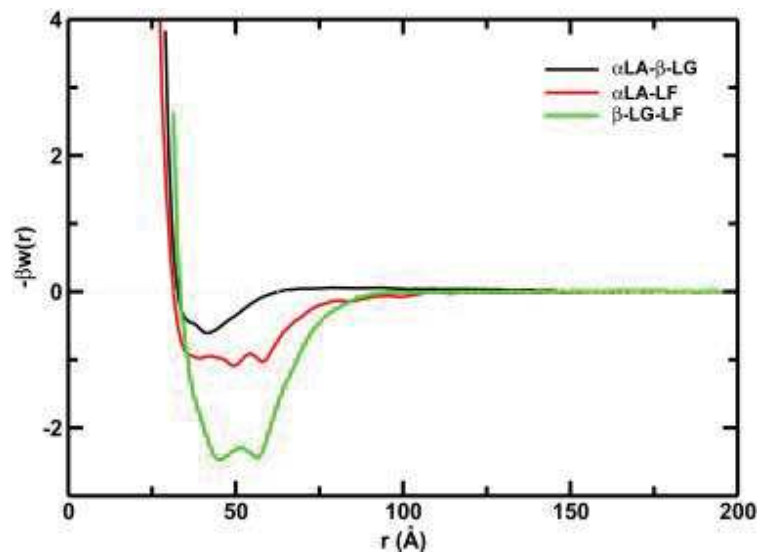


5.2 Complexação molecular

O processo de dimerização foi investigado através de derivadas da energia livre obtidas em simulações MCM. As energias livre de interação calculadas são dadas por potenciais de força média [$\beta w(r)$] para os três sistemas estudados (α -LA- β -LG, α -LA-LF e β -LG-LF) em pH 6,5 e força iônica fisiológica e são apresentadas na Figura 27. Neste valor de pH natural do leite ou em suas proximidades, todos os sistemas mostram um claro mínimo de energia ao redor de 50 Å, que pode indicar uma complexação baseada em critérios termodinâmicos [$\beta w(r) < 0$]. Entretanto, a associação entre α -LA e β -LG mostra um poço de potencial de $-0,6 k_B T$ na distância de separação de 41,6 Å. Sendo esse valor menor que a energia térmica, $1 k_B T$, essa é uma associação instável. Assim, assumiremos que esse

complexo é dissolvido nessas condições de solução. Da mesma forma, a magnitude do mínimo de energia livre para o complexo α -LA-LF é da ordem de $1 k_B T$, que pode explicar porque esse complexo foi observado em alguns experimentos e não em outros (HEKMAN, 1971; LAMPREAVE et al., 1990; PUYOL et al., 1991). E finalmente, a interação entre β -LG e LF, como já mencionado anteriormente, é bastante pronunciada, inclusive nesta concentração de sal.

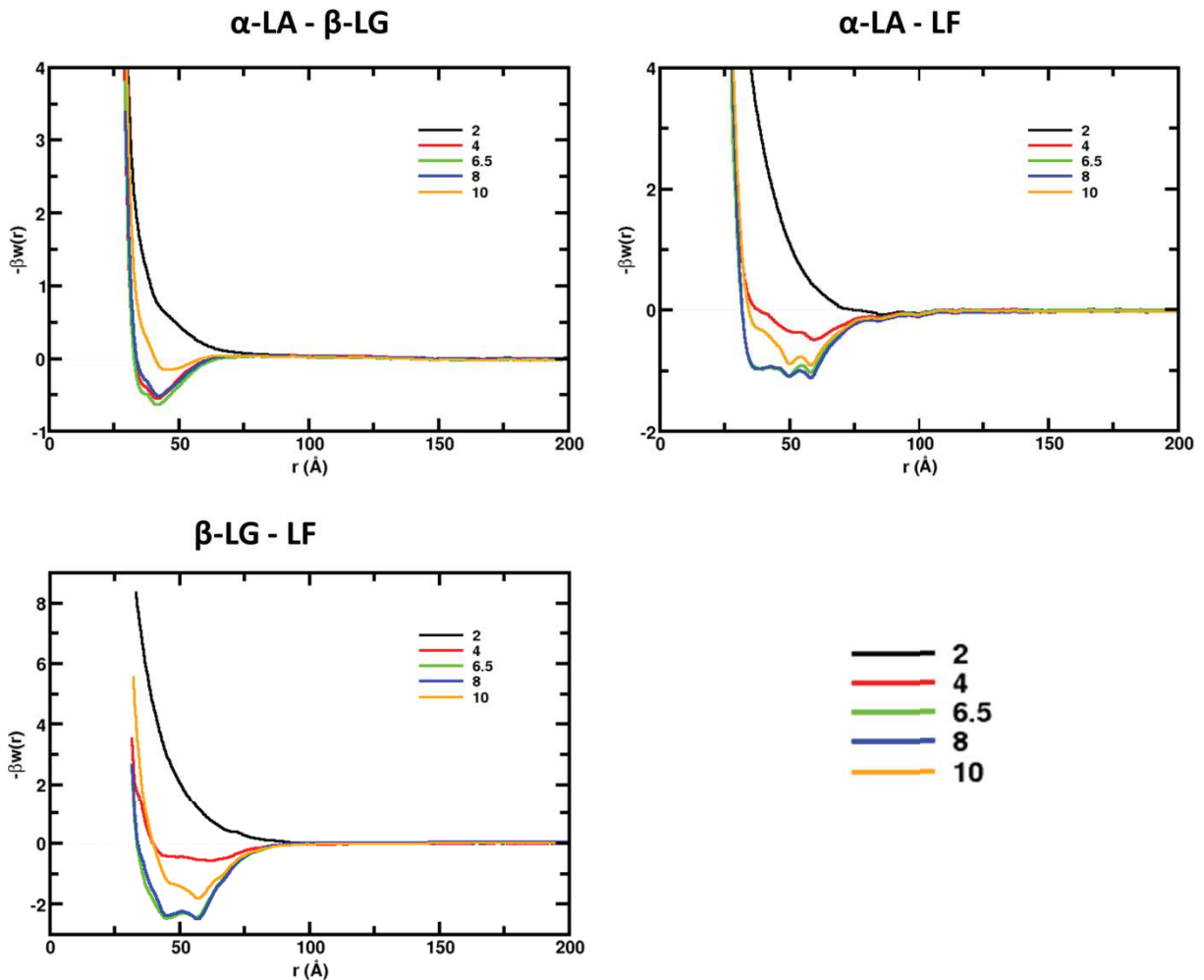
Figura 27 – Curvas de potencial de energia livre de interação $[\beta w(r)]$ para os três sistemas estudados.



Simulações de energia livre de interação $[\beta w(r)]$ entre os centro de massa das proteínas. O pH da solução foi fixado em 6,5 e a concentração de sal 150 mM.

A investigação seguinte foi a respeito dos efeitos do sal e pH na complexação. Na Figura 28 pode-se ver que para todos os sistemas, o valor de pH ótimo está próximo do pH natural do leite, ou seja, pH 6,5, justamente o pH em que os mínimos de energia livre foram observados. Atrações fracas podem ser vistas para o complexo α -LA- β -LG entre os pH's 4 e 10. Seguindo o critério de que uma energia livre negativa pode resultar em um complexo, este é o intervalo de pH onde os complexos α -LA- β -LG, α -LA-LF e β -LG-LF seriam formados em concentrações de sal fisiológicas. Uma visão mais realista, como mencionado acima, seria assumir um mínimo de $1 k_B T$ que diminui com as janelas de pH de 6,5 a 8,0 e 6,5 a 10,0 para os complexos α -LA-LF e β -LG-LF, respectivamente. Nenhuma associação seria observada para a α -LA- β -LG em qualquer situação de pH.

Figura 28 – Curvas de potencial de energia livre de interação $[\beta w(r)]$ para os três sistemas estudados comparando diferentes pHs.

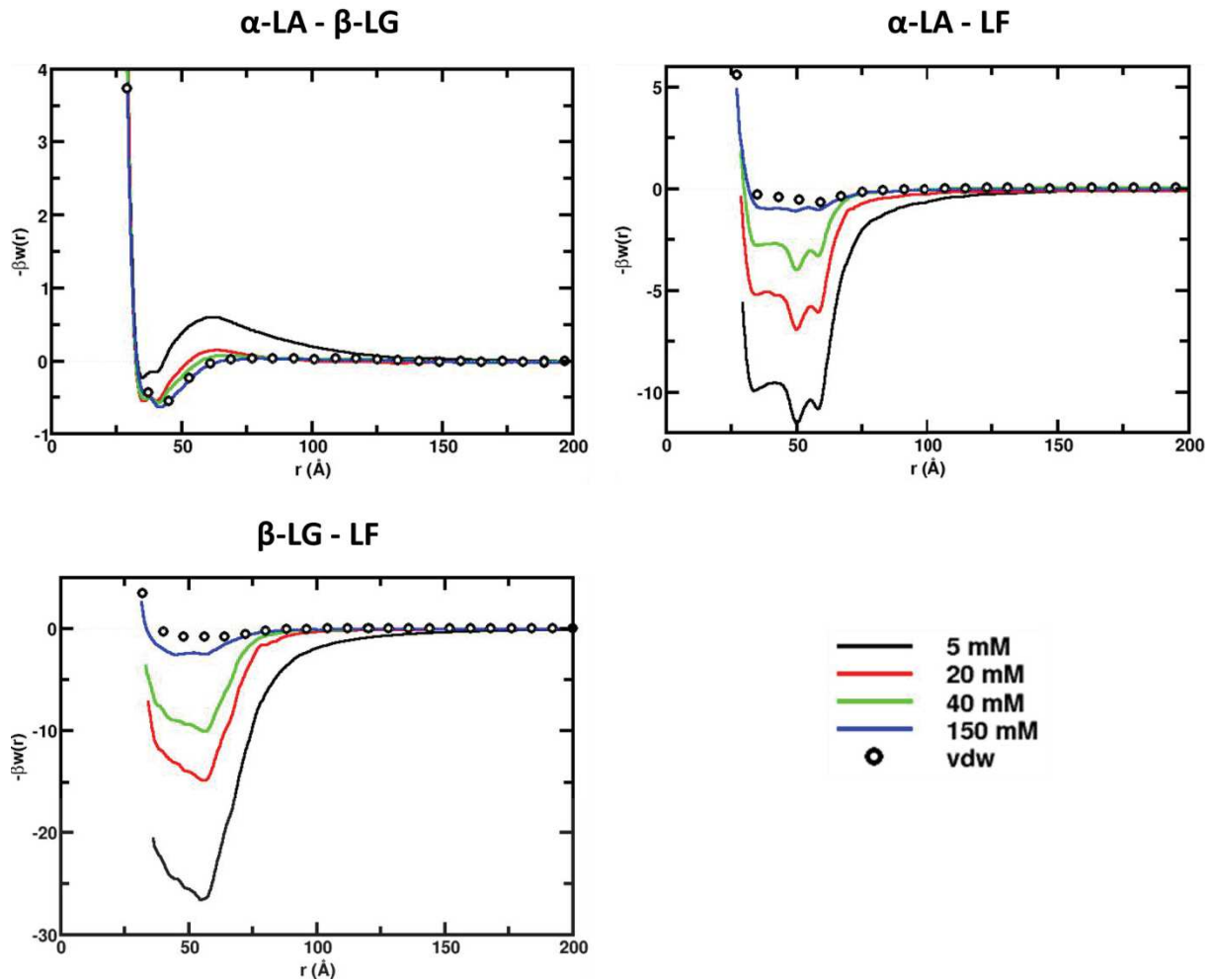


Simulações de energia livre de interação $[\beta w(r)]$ entre os centro de massa das proteínas em diferentes valores de pH.

O efeito do sal é examinado na Figura 29 em pH 6,5. Neste pH e em uma solução a 150 mM de sal, as cargas simuladas das proteínas são $z_{\alpha-LA} = -3,7$, $z_{\beta-LG} = -12,5$ e $z_{LF} = +21,2$. O mínimo atrativo em um complexo de proteínas é reduzido com a adição de sal quando as proteínas são opostamente carregadas. Da mesma forma, para proteínas de cargas iguais a repulsão é reduzida. Temos também uma curva apresentando a contribuição de vdW simulada para a energia livre. Isso nos permite distinguir a principal força que governa a complexação, entre interações eletrostática e vdW. Para o sistema α -LA- β -LG, a blindagem do sal é mais severa com a interação eletrostática (multipolo-multipolo e o mecanismo de regulação de cargas). Em 150 mM, o mínimo de energia livre para esse sistema é unicamente devido à vdW. Todas as interações eletrostáticas, incluindo a interação repulsiva de Coulomb

devido às cargas iguais das duas proteínas, foram rastreadas. Da mesma forma, para os outros sistemas, devido às suas cargas de sinais opostos, o grande momento de dipolo e capacitância da LF, assim como seu grande mínimo de energia é sempre aumentado com respeito à curva de vdW. Em baixas concentrações de sal, os sistemas α -LA-LF e β -LG-LF tem uma tendência a se associarem fortemente. O mínimo de energia livre medida é menor que -10 e $-20 k_B T$ para α -LA-LF e β -LG-LF respectivamente. Em todas as condições estudadas, β -LG interage mais fortemente com LF do que α -LA. Considerando a alta concentração de espécies salinas (sódio, potássio, cloreto, sulfato, fosfato, citrato, cálcio e magnésio) no leite natural (THOMPSON; BOLAND; SINGH, 2009), a maioria desses complexos não seria observada nesse meio. Da mesma forma, em altas concentrações de proteína, as interações eletrostáticas são mais blindadas reduzindo a associação dessas proteínas. Uma pequena quantidade de sal em uma alta concentração de proteína pode ter uma força iônica equivalente a um sistema com alta concentração de sal. Nessas condições, a concentração máxima de sal em que a complexação pode ocorrer é diminuída. Isso pode explicar por que foi observado o complexo β -LG-LF no intervalo de pH entre 5,0 e 7,5, com uma baixa concentração crítica de sal (100 mM, e uma concentração de proteína de cerca de 1g/L) (ANEMA; DE KRUIF, 2014). Uma alta concentração de proteínas (20 g/L) na presença de agregados diminui ainda mais essa concentração máxima de sal, como observado por outros autores (YAN et al., 2013).

Figura 29 – Curvas de potencial de energia livre de interação $[\beta w(r)]$ para os três sistemas estudados comparando diferentes condições salinas, assim como vdW.



Simulações de energia livre de interação $[\beta w(r)]$ entre os centro de massa das proteínas em diferentes forças iônicas. O pH foi fixado em 6,5. As curvas formadas pelos círculos abertos são dados obtidos por simulações MC, assumindo que todos os aminoácidos das proteínas sejam neutros.

Finalmente, a Figura 30 conclui todo o estudo de complexação deste trabalho, reunindo dados obtidos através de simulação computacional (MCM), e estimados analiticamente, utilizando a teoria simples de DH, segundo a Equação 10. Para o cálculo analítico, foram utilizados os dados de carga encontrados pelas curvas de titulação das proteínas em condições fisiológicas de sal (150 mM), juntamente com uma constante dielétrica (ϵ) de 78,7 e a distância de separação entre as proteínas fixa em 50 Å. As linhas pontilhadas representam o cálculo analítico, sendo que os círculos são referentes às simulações MCM, realizadas nos pH's 4, 6,5 8 e 10. Os valores de interação obtidos nos dois modelos foram normalizados para 1, a fim de se fazer a comparação. No pH 4 vemos que

cada um dos sistemas tem uma interação diferente, assim como nos pH's 8 e 10. Já no pH 6,5 há sobreposição dos pontos, mostrando que a maior interação dos sistemas ocorre nesse pH. Sabe-se que algumas interações são bastante fracas, como o sistema α -LA- β -LG, e outras mais fortes, como o sistema β -LG-LF, porém sempre no mesmo pH. Analisando as curvas referentes ao resultado analítico, somente para os dois sistemas com maior interação, pode-se ver uma “janela” de pH para essas complexações. Considerando exclusivamente a eletrostática, observa-se que o pH onde ocorreria a maior interação seria o pH 8, porém com interação muito próxima ao pH 6,5. Com isso entende-se que para estes sistemas de proteínas do soro do leite, a interação eletrostática carga-carga é a chave para a complexação. As outras interações, consideradas na simulação, são insignificantes para o sistema.

Figura 30 – Efeito do pH nas interações simuladas e estimadas entre os sistemas α -LA-LF e β -LG-LF.

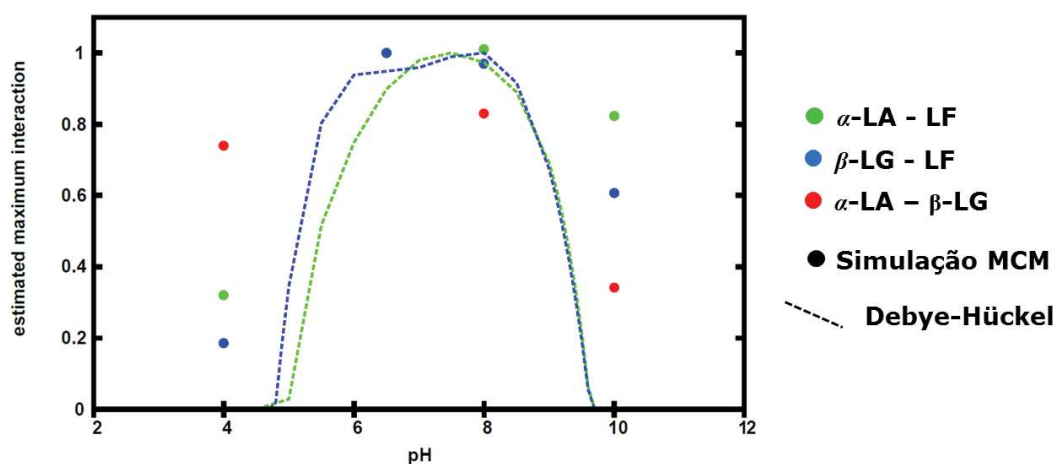


Figura comparando dados a partir da simulação MCM e da teoria simples de DH. Nos pH's de maior interação entre as moléculas as interações Coulombicas são o grande destaque.

Os resultados obtidos a partir da Figura 30 podem ser úteis à indústria, a fim de manipular o pH dos produtos para que as proteínas do soro fiquem complexadas ou não, de acordo com as necessidades. Na Tabela 2 são apresentados valores de carga e da energia livre de interação entre agregados (ΔG_{Agre}) caracterizados na Figura 30. Essa estimativa foi feita através da equação analítica (Equação 10), nos pH's 6,5 e 8, e em condições fisiológicas de sal. As cargas referem-se à carga líquida dos complexos (α -LA - LF e β -LG - LF), e o ΔG_{Agre} é a variação da energia livre de complexação estimada considerando somente as interações coulômbicas do homocomplexo. A distância média de separação utilizada para os complexos

foi de 75 Å, visto que pode variar de 50 a 100 Å, de acordo com o posicionamento das proteínas. Como os complexos de proteínas possuem a mesma carga (complexos iguais) o ΔG_{Agre} será sempre positivo, ainda que bastante próximo a zero. Com esses valores de ΔG de ligação, conclui-se que não há complexação entre os homopares de proteína-proteína, ou seja, não há agregação ou formação de grandes aglomerados proteicos nessas condições. Isso é bastante útil à indústria, especialmente para o caso de bebidas produzidas com as proteínas do soro do leite. Assim, podem conservar um aspecto homogêneo sem formação de agregados proteicos, mantendo uma boa aparência por mais tempo, para o produto pronto no local de venda.

Tabela 2 – Valores de carga líquida e variação de energia livre de ligação para os complexos α -LA – LF e β -LG – LF nos pH's 6,5 e 8.

Complexo	pH 6,5		pH 8	
	Carga	ΔG	Carga	ΔG
α -LA - LF	16	0,001745	9,8	0,000655
β -LG - LF	9	0,000552	-2	0,000027

6. Estudo de correlação de cargas em estruturas de proteínas

6.1 Caracterização do modelo

6.1.1 Correlação de cargas em estruturas tridimensionais

Diversos conjuntos de dados de coordenadas atômicas de proteínas obtidos a partir do PDB, contendo proteínas com diferentes propriedades físico-químicas, classificação estrutural e funções biológicas, foram usadas neste estudo. Alguns desses conjuntos contêm um grande número de coordenadas de proteínas, e no maior deles, todas as estruturas do PDB foram usadas. O maior conjunto é composto por 172.622 proteínas, correspondendo a 72.966 códigos PDB. A fim de avaliar possíveis artefatos devido à redundância (HOLM et al., 2008), fizemos nossos cálculos também com conjuntos de cadeias do PDB não redundantes, resultante do servidor VAST (Vector Alignment Search Tool), (GIBRAT; MADEJ; BRYANT, 1996). A seguir, apresentamos a Tabela 3 com os conjuntos de dados divididos entre classificação biológica e estrutural.

Tabela 3 – Conjunto de dados obtidos do PDB para estudo de correlação de cargas, divididos em duas classificações, Biológica e Estrutural.

<i>Classificação Biológica</i>	<i>Classificação Estrutural</i>
Alérgenos	Grupo PISCES 1 ^(b)
Apoptose	Grupo PISCES 2 ^(c)
Ligação de Cálcio (<i>apo</i>)	RMN ^(d)
Ligação de Cálcio (<i>halo</i>)	Somente Alfa ^(e)
Coagulação	Somente Beta ^(f)
Hormônios	Vicatos (VICATOS; ROCA; WARSHEL, 2009)
Imunoglobulinas	Todo o PDB ^(g)
Membranas	Somente cadeia A de "Todo o PDB"
Oxirredutase	Grupo não redundante ^(h)
Fotossíntese	Grupo não redundante selecionado ⁽ⁱ⁾
Ligação de RNA	
Sinalização	
Toxinas	
Transferase	
Transporte ^(a)	
Virus	

(a) Incluindo transporte de oxigênio, prótons, lipídios, metais, retinol e açúcares.

(b) Conjunto de dados gerados pelo servidor PISCES (<http://dunbrack.fccc.edu/PISCES.php>) a) porcentagem de identidade de sequência: $\geq 25\%$, b) resolução: 0,0~1Å, c) fator R: 0,3, d) comprimento da sequência: 40~10000, e) estruturas não obtidas por raio-X: excluídas, f) estruturas somente com cálcio: excluídas e g) somente 185 estruturas do PDB foram usadas.

(c) Mesmos parâmetros de (c), com uma resolução menos refinada 0,0~3,0Å. Foram usadas 7638 estruturas do PDB.

(d) Estruturas selecionadas a partir do servidor PDB (<http://www.rcsb.org/pdb/home/home.do>) com os seguintes critérios: a) método experimental em solução RMN, b) as estruturas não apresentam DNA, RNA ou híbridos e c) somente estruturas de eucariotos foram utilizadas.

(e) Os mesmos parâmetros de (e), agora com o método experimental de raio-X, e adição de: a) resolução de 1,499Å ou menos e b) a classe das proteínas é somente α [como classificado pelo SCOP (HUBBARD et al., 1997)].

(f) O mesmo que em (f), porém a classe de proteínas é somente β .

(g) Todos os 72.966 códigos existentes no PDB na data de elaboração do trabalho foram utilizadas.

(h) O conjunto de cadeias não redundantes do PDB foi fornecido pelo servidor VAST (Vector Alignment Search Tool), disponível no seguinte endereço: <http://structure.ncbi.nlm.nih.gov/Structure/VAST/nrpdb.html> com níveis de não redundância iguais a 10^{-7} . Neste contexto, a redundância refere-se às muitas proteínas que foram cristalizadas mais de uma vez, existindo assim mais de um código PDB referente à mesma estrutura.

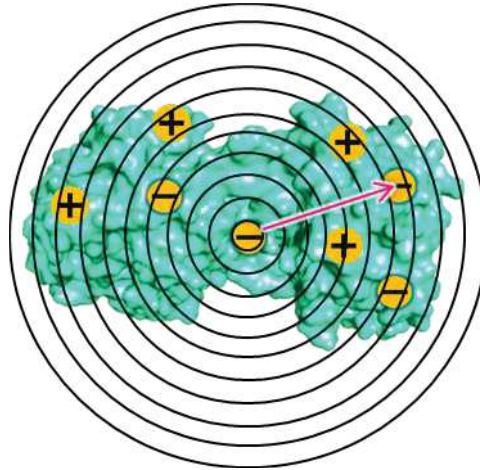
- (i) A partir do conjunto não redundante acima, foram selecionadas coordenadas de proteínas que não apresentassem heteroátomos (exceto água).

Esta classificação em grupos já sugerida na literatura (GUTTERIDGE; THORNTON, 2005) é interessante, a fim de explorar particularmente padrões relacionados com a função biológica. As coordenadas de todos os átomos das estruturas tridimensionais das proteínas foram usadas sem qualquer manipulação. As proteínas foram assumidas como sendo monômeros, isto é, quando mais de uma cadeia distinta estava disponível no arquivo de coordenadas PDB, cada uma das cadeias foi tratada como uma proteína individual, e as análises geométricas foram realizadas para todas essas estruturas. O efeito das múltiplas cópias das proteínas (quando há várias estruturas idênticas no mesmo arquivo PDB) foi investigado realizando as análises com somente a primeira cadeia simples selecionada a partir do total. Os heteroátomos, como íons, ligantes, metais e moléculas de solvente não foram considerados (exceto quando indicado), devido ao fato de nossa ênfase aqui ser a estrutura da proteína. Outra razão para se remover íons metálicos é a necessidade de se promover um critério único para as estruturas, visto que no PDB podem-se encontrar proteínas em sua forma *halo* (presença de íons metálicos) e *apo* (ausência de íons metálicos). Por outro lado, proteínas com ligação de cálcio foram investigadas em ambas as formas a fim de descrever os efeitos da presença do metal. Também a fim de comparações, foram realizadas análises para um conjunto de coordenadas de proteínas não redundantes que não contém heteroátomos (exceto a água) no arquivo de coordenadas. Em todos os casos, os átomos faltantes nas estruturas foram desconsiderados. O arquivo PDB pode conter incertezas (DAVIS; TEAGUE; KLEYWEGT, 2003), e, antes de submeter as estruturas às análises, os arquivos foram corrigidos. Coordenadas com átomos estatisticamente desordenados foram desconsideradas, e somente foram escolhidas as formas “A” para as simulações, ou seja, somente as coordenadas com maior credibilidade. Proteínas que existem em várias isoformas foram consideradas somente pela estrutura dada pelo arquivo PDB.

A distribuição das cargas em cada proteína foi avaliada simplesmente medindo a distância entre todas as cargas a partir das coordenadas da proteína. Posteriormente a distribuição foi normalizada para ter sua integral igual a um. Foi criando um histograma $P(R)$ com 0,5 Å de distância entre as camadas, representando esferas concêntricas, conforme pode ser visualizado na Figura 31. Esta é uma forma aproximada de representação, visto que quanto mais longe do centro a esfera estiver, maior o volume da camada, criando diferentes

volumes e conseqüentemente diferentes densidades para as cargas. Contudo, o modelo é bastante útil para comparar as mesmas camadas entre cada uma das estruturas, já que a aproximação ocorre de forma idêntica a todas.

Figura 31 – Representação da proteína carregada com as camadas concêntricas para criação do histograma.



Representação da proteína criada com o auxílio do programa computacional *PyMOL* (SCHRÖDINGER, 2002). A estrutura da proteína foi obtida a partir do Banco de Dados de Proteínas (Protein Data Bank - PDB) (BERMAN et al., 2000).

Nós esperamos observar uma probabilidade de distribuição para resíduos opostamente carregados, P_{+-} , maior que para resíduos igualmente carregados, P_{++} ou P_{--} a curtas distâncias de separação. Em grandes distâncias, o oposto seria verdade, $P_{++}, P_{--} > P_{+-}$, com a probabilidade de distribuição normalizada. Neste texto, $P_{+-} > P_{++}, P_{--}$ será referido como uma correlação de cargas positiva, enquanto o inverso será uma correlação de cargas negativa, ou anti-correlação. Aos resíduos tituláveis foram atribuídas cargas referentes a uma solução em pH 7 (PACE; GRIMSLEY; SCHOLTZ, 2009; TEIXEIRA; LUND; DA SILVA, 2010), fazendo com que algumas proteínas tenham cargas consideráveis. Dessa forma, utilizamos o campo de força reduzido, desconsiderando as vizinhanças. Seguimos aqui o mesmo procedimento usado por Barlow e Thornton (BARLOW; THORNTON, 1983). Considerar todos os resíduos tituláveis em um mesmo pH fixo é uma forma de aproximação, porém útil para manter todas as estruturas proteicas estudadas em um mesmo padrão. Outros critérios também foram testados, como diferentes campos de força, mas os resultados se mantiveram similares. Destacamos que os histogramas são baseados no sinal da carga e não na sua

magnitude, a qual é pH dependente. Assim, a função de distribuição P_{+-} é bastante insensível à carga líquida.

Para quantificar a correlação em um grande conjunto de dados, com milhares de coordenadas de proteínas, é conveniente introduzir um critério numérico. O coeficiente (ξ) foi definido como uma função de parâmetros de correlação ξ_{+-} e ξ_{++} , ξ_{--} , como

$$\xi = \frac{\xi_{+-}}{\xi_{++,-}} = \frac{\int_0^{r_c} P_{+-}(r_{+-}) dr_{+-}}{0,5(\int_0^{r_c} P_{++}(r_{++}) dr_{++} + \int_0^{r_c} P_{--}(r_{--}) dr_{--})} \quad (11)$$

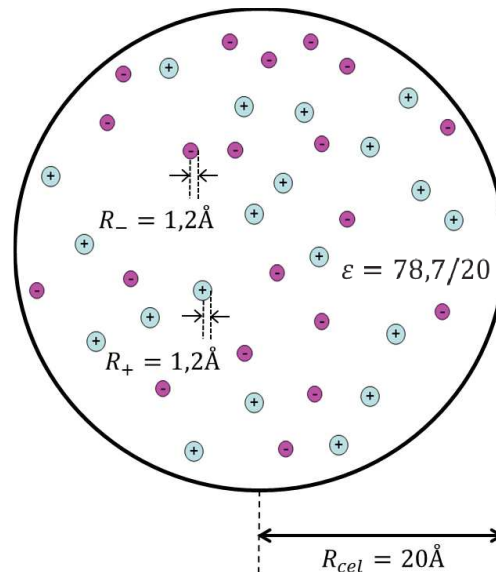
onde r_c (distância máxima computada) é assumido como 10 Å, baseado em trabalhos prévios (BARLOW; THORNTON, 1983; KUMAR; NUSSINOV, 2002). Alterando este parâmetro de ± 1 Å, o resultado final não varia mais que 10%. Se $\xi < 1$, o sistema apresenta uma correlação de cargas negativa, ou uma anti-correlação. Por outro lado, com $\xi \geq 1$ o sistema tem uma correlação de cargas positiva, ou seja, há uma preferência de se ter carga positiva e negativa próximo a cada uma. Este critério permite que sejam analisadas quaisquer coordenadas de proteínas, incluindo todo o PDB. O coeficiente de correlação médio $\langle \xi \rangle$ será relatado para diferentes classes e conjuntos de proteínas.

A fim de revelar as nuances que fazem cada proteína única, no sentido biológico, enfatizamos algumas estruturas específicas escolhidas aleatoriamente, com e sem correlação de cargas. Também incluímos estruturas de modelagem molecular da α_{s1} -CN (KUMOSINSKI; BROWN; FARRELL, 1994), α_{s2} -CN (FARRELL et al., 2009), β -CN (KUMOSINSKI; BROWN; FARRELL, 1993a) e κ -CN (KUMOSINSKI; BROWN; FARRELL, 1993b). Como já citado neste texto, as caseínas são particularmente interessantes em função de sua flexibilidade, não apresentando uma estrutura terciária definida; especificamente para estes sistemas, foram utilizados modelos computacionais (KUMOSINSKI; BROWN; FARRELL, 1993a, 1993b, 1994; FARRELL et al., 2009). As proteínas apresentadas nas tabelas foram classificadas de acordo com o ξ calculado a partir de seus dados estruturais. Quando mais de uma cadeia de proteínas estava disponível no arquivo PDB, somente a primeira foi utilizada.

6.1.2 Sistema referência para a correlação de cargas

Visando ter um sistema de referência para interpretação das observações experimentais, também foi estudada uma solução eletrolítica confinada a uma célula esférica eletroneutra de raio $R_{cel} = 20 \text{ \AA}$, como um modelo muito simples de cargas em uma proteína, como pode ser visualizado na Figura 32. Simulações adicionais também foram realizadas com uma célula maior, $R_{cel} = 30 \text{ \AA}$, e em ambas as células os íons livres foram descritos pelo modelo primitivo, ou seja, uma esfera rígida carregada imersa em um solvente representado por um dielétrico contínuo (HILL, 1986). Cada íon móvel k com valência z_k é tratado explicitamente como uma esfera rígida com um raio de $1,2 \text{ \AA}$, e o solvente é um meio dielétrico desestruturado caracterizado por uma permissividade dielétrica relativa, ϵ_r . O sistema foi amostrado por simulações Monte Carlo Metropolis (METROPOLIS et al., 1953) e os cálculos foram executados com 20 pares de íons e duas constantes dielétricas diferentes ($\epsilon_r = 20$ e $78,7$), em temperatura ambiente. Após o equilíbrio, ao menos 10^6 ciclos de simulações foram realizados (período de produção), onde as probabilidades P_{+-} , P_{++} e P_{--} foram obtidas. As observações encontradas nesse sistema simples fornecem uma boa base para a interpretação dos resultados observados para as coordenadas de proteínas disponíveis no PDB.

Figura 32 – Esquema de uma célula esférica contendo solução eletrolítica em um meio descrito pela constante dielétrica ϵ



Íons positivos e negativos são representados respectivamente por esferas de cores azuis e rosa. Demais detalhes estão apresentados no texto.

6.2 Resultados e discussão

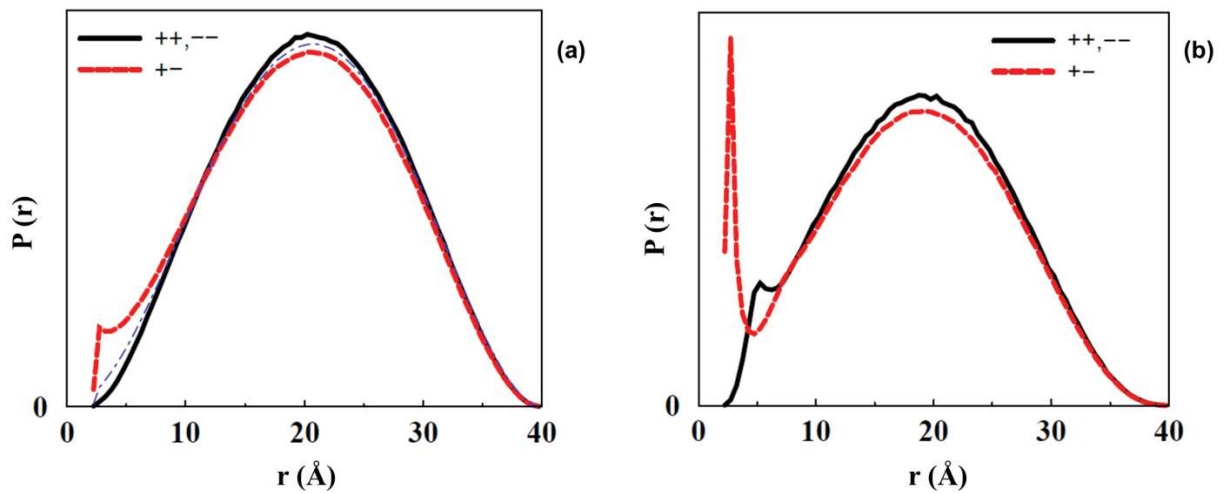
6.2.1 Sistema de referência

Iniciamos nossos resultados do estudo de correlação de cargas pelo sistema de referência, o qual servirá para termos um referencial teórico para interpretação das análises seguintes. Esse sistema de referência, adotado para todas as nossas análises de correlação de cargas de proteínas, é a única parte deste estudo de correlação de cargas cuja amostragem dos dados foi obtida por simulações MCM. Nessa análise, admitimos que podemos ter uma noção do comportamento de cargas elétricas no interior de uma proteína valendo-se da analogia com pares de íons confinados em uma caixa de simulação esférica.

Para um eletrólito consistindo de 20 pares de íons confinados em uma célula esférica eletroneutra, observamos, nas simulações MCM, que há uma correlação positiva, ou seja, $P_{+-} > P_{++,--}$, a curta separação, como pode ser demonstrado na Figura 33a. A correlação positiva é bastante fraca e limitada para separações menores que 10 \AA , para uma esfera de confinamento de 20 \AA de raio. Essa fraca correlação é observada em um meio de alta

permissividade dielétrica relativa, $\epsilon = 78,7$. Uma distribuição de probabilidade para pontos aleatórios na mesma esfera é mostrada para comparação, caracterizada como a curva azul. Com uma permissividade dielétrica relativa reduzida para a esfera (que aqui imita uma proteína), é possível observar um aumento da correlação positiva de cargas com um forte pico localizado na distância correspondente ao contato de caroço duro entre dois íons – veja Figura 33b. A correlação é tão forte que leva à formação de pequenas camadas de íons, formando máximos na distribuição de cargas iguais na distância de separação entre 5 a 6 Å, como pode ser também observado na Figura 33b. Dessa forma, sabemos que a redução da constante dielétrica leva a um aumento expressivo nessa correlação. Logo, se observarmos correlações elevadas nas análises a seguir, das cargas em estruturas de proteínas disponíveis no PDB, temos evidência forte de que o interior das proteínas deva ter uma resposta dielétrica equivalente a um meio dielétrico de baixa constante dielétrica. No entanto, caso não se observe correlações significativas, tem-se evidência adicional (e nesse caso, baseado na análise direta de dados experimentais) de que o interior das proteínas pode ser modelado por uma constante dielétrica próxima a da água, como já argumentado por vários autores (PENFOLD; WARWICKER; JÖNSSON, 1998; DA SILVA; JÖNSSON; PENFOLD, 2001; DE CARVALHO; GHIOTTO; DA SILVA, 2006; DE CARVALHO; FENLEY; DA SILVA, 2008).

Figura 33 – Distribuição de probabilidade de cargas negativas e positivas confinadas em uma esfera de raio 20 Å.



Os íons são modelados como esferas rígidas carregadas, com um raio de 1,2 Å. A constante dielétrica da esfera em (a) é 78,7, e em (b) é 20. A linha azul pontilhada em (a) mostra uma distribuição randômica de cargas positivas e negativas em uma esfera.

6.2.2 Análise das estruturas de proteínas do PDB

Nas análises seguintes, os resultados são exclusivamente baseados na observação da distribuição de cargas elétricas diretamente em estruturas experimentais de proteínas disponíveis no PDB. Não empregamos qualquer Hamiltoniana e/ou efetuamos simulações.

A Tabela 4 compila os resultados de diferentes classes de proteínas, onde podem ser observadas moléculas tanto com correlação de cargas positiva ou negativa. Independentemente da estrutura ou função biológica, a maioria das proteínas (87%) de todo o PDB apresentam correlação positiva. Algumas classes de proteínas, como as imunoglobulinas, oxirredutases, transferases, proteínas de transporte e vírus, apresentam uma correlação positiva mais forte que a média observada para o PDB como um todo. Das classes de proteínas avaliadas, as que se ligam a cálcio compõe a única classe que na média apresenta uma correlação de cargas negativa (onde cargas de mesmo sinal preferem estar próximas umas das outras). Isso possivelmente ocorre devido à necessidade dessas proteínas terem cargas negativas concentradas no sítio de ligação para atrair, com bastante afinidade, íons de cálcio, como é observado, por exemplo, para a calbindina D_{9K} (KESVATERA et al., 1994).

Notamos também que todas as classes as quais as proteínas foram classificadas apresentam carga média negativa. Essas cargas são referentes às proteínas em pH 7, e correspondem a uma média de todos os possíveis valores existentes.

Tabela 4 – Características médias das diferentes categorias de proteínas extraídas do PDB.

<i>Classes de proteínas</i>	N_{estrut} ^(a)	$\langle N_{aaP} \rangle$	$\langle z_P \rangle$	$\langle \xi \rangle$	$\%_{+-}$
Alérgenos	136	186	-7	1,35	78
Apoptose	744	164	-4	1,31	81
Ligação de Cálcio (<i>apo</i>)	260	156	-13	0,92	33
Ligação de Cálcio (<i>halo</i>)	260	156	-2	1,11	53
Coagulação	130	186	-7	1,07	65
Hormônios	1478	134	-3	1,42	73
Imunoglobulinas	512	188	-6	1,46	95
Membranas	2563	193	-9	1,44	84
Oxirredutase	9662	322	-8	1,38	94
Fotossíntese	707	207	-5	1,40	80
Ligação de RNA	1442	149	-1	1,50	75
Sinalização	3903	185	-5	1,32	84
Toxinas	2063	173	-6	1,35	82
Transferase	20824	306	-8	1,34	94
Transporte	9944	200	-6	1,37	87
Virus	1259	327	-10	1,44	95
Grupo PISCES 1	223	179	-5	1,42	83
Grupo PISCES 2	16893	216	-7	1,38	87
RMN	1631	94	-1	1,33	69
Somente Alfa	98	176	-1	1,40	91
Somente Beta	134	163	-2	1,44	92
Vicatos (VICATOS; ROCA; WARSHEL, 2009)	41 ^(b)	128	-4	1,38	81
Todo o PDB	172622	244	-7	1,35	87
Somente cadeia A de "Todo o PDB"	68912	258	-6	1,35	89
Grupo não redundante	11356	187	-5	1,38	83
Grupo não redundante selecionado	3816	135	-3	1,40	77

Na tabela são apresentados o número total de estruturas (N_{estrut}), o número médio de resíduos ($\langle N_{aaP} \rangle$) e carga média da proteína ($\langle z_P \rangle$) em unidades de carga elementar e . Coeficiente de correlação médio $\langle \xi \rangle$ e porcentagem ($\%_{+-}$) de correlação de cargas positiva ($\xi \geq 1$) em curtas distâncias.

- (a) Todas as cadeias disponíveis no arquivo PDB foram consideradas como estruturas de proteínas individuais e utilizadas no trabalho.
- (b) Os dados originais da referência de Vicatos são compostos por 45 estruturas. As estruturas com formas mutantes não foram consideradas, somente as formas nativas.

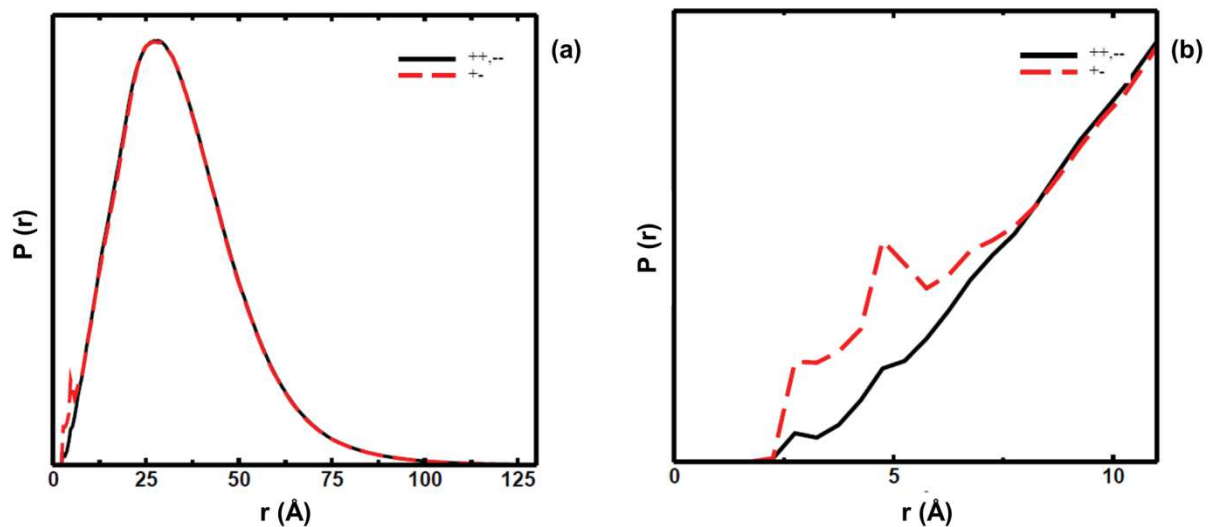
O mesmo padrão observado para estruturas do PDB ($\%_{+-} = 87\%$) também é visto para as estruturas de proteínas geradas pelo servidor PISCES (WANG; DUNBRACK, 2003) ($\%_{+-} = 83$ e 87%), usando estruturas obtidas por raio-X. Os resultados encontrados para as estruturas de proteínas com resolução menor que $1,0 \text{ \AA}$ (correspondendo ao primeiro conjunto de PISCES na Tabela 4) são praticamente os mesmos obtidos com resolução menos refinada ($< 3,0 \text{ \AA}$). Na classificação estrutural SCOP, (HUBBARD et al., 1997) essencialmente também não há diferença entre os dados usando as estruturas α ou β ($\%_{+-} = 91$ e 93% , respectivamente). As estruturas α e β (α/β) seguem a mesma tendência, com $\%_{+-} = 92,9\%$, sendo que $\alpha + \beta$ tem a probabilidade levemente menor em ter aminoácidos positivos e negativos próximos um do outro, com $\%_{+-} = 84,2\%$. Aparentemente, estruturas obtidas a partir da técnica de RMN tem uma tendência em apresentar mais estruturas com correlação negativa do que as obtidas por raio-X ($\%_{+-} = 69,3\%$); isso provavelmente pode ser explicado pelo fato das estruturas obtidas por RMN serem normalmente menores que as obtidas por raio-X. Esses resultados podem ser comparados com os obtidos a partir de estruturas do PISCES, usando os dados RMN. Todavia, porcentagem similar foi observada no grupo de proteínas de coagulação ($\%_{+-} = 65\%$). Agrupando as estruturas de todo o PDB de acordo com sua carga líquida, temos que $\%_{+-}$ fica em um intervalo entre 67% e 95% . Proteínas positivamente carregadas tem uma leve tendência a apresentar mais áreas carregadas, que significa menores valores observados para $\%_{+-}$. Estruturas obtidas a partir de simulações de trajetórias de dinâmica molecular e modelagem por homologia também apresentam as mesmas características (BRASIL; DELBEM; DA SILVA, 2013).

Os valores médios obtidos do conjunto total de estruturas do PDB nos dão $\%_{+-} = 87\%$, mas essa porcentagem é composta por uma larga amplitude, de 33 a 95 , dependendo do critério usado para formar o conjunto de dados. Assim, selecionando uma proteína simples ou uma pequena classe de proteínas, poderíamos encontrar um $\%_{+-}$ menor que 50% . Porém, as proteínas usadas por Vicatos e seus colaboradores (VICATOS; ROCA; WARSHEL, 2009) parece ser heterogênea o suficiente para refletir os valores médios maiores que 50% . Analisando para todas as estruturas do PDB, encontra-se que 13% das proteínas, nada menos que 22.432 cadeias de proteínas, mostram uma correlação carga-carga negativa! Visto o grande número de dados, não é possível cobrir aqui todos os aspectos funcionais de cada estrutura proteica do PDB. É interessante notar que analisando 191 enzimas, Gutteridge e

Thornton puderam observar a relação entre dois resíduos carregados e sua função biológica. Eles descreveram que em um sítio ativo de enzima, resíduos de cargas idênticas são frequentemente observados e importantes para a função catalítica (GUTTERIDGE; THORNTON, 2005).

Um gráfico de $P_{+-}(r)$ e $P_{++,-}(r)$ para todas as estruturas de proteínas disponíveis no PDB é mostrado na Figura 34. Pode ser observado que a correlação positiva é moderada e observada somente para curtas distâncias ($r < 8 \text{ \AA}$). Deve-se lembrar de que este é um resultado médio, e que o gráfico oculta sistemas com correlação carga-carga mais distintas. Porém, nenhuma estrutura individual apresentou um forte pico como observado no sistema referência, com uma baixa constante dielétrica (Figura 33b). Esta observação pode explicar porque em vários trabalhos da literatura é importante adotar altas constantes dielétricas (entre 35 e 40) para descrever as interações eletrostáticas nas proteínas (VICATOS; ROCA; WARSHEL, 2009). Outras investigações sugerem um panorama similar (WARWICKER, 1999; DOMINY; MINOUX; BROOKS, 2004).

Figura 34 – Distribuição de cargas positivas e negativas para todas as 172.622 cadeias de proteínas disponíveis no PDB.



Em (a), temos todas as distâncias de separação, e em (b), apresentamos uma aproximação do primeiro gráfico, entre as distâncias de zero a aproximadamente **11 Å**.

A tabela 4 compila dois conjuntos individuais de proteínas, sendo o conjunto superior com correlação de cargas positiva e o conjunto inferior com correlação de cargas negativa. Considerando os dois conjuntos de dados, temos que $\%_{+-} = 72\%$. Das 25 estruturas estudadas em detalhes, 18 delas pertencem ao primeiro conjunto, usando o critério ξ . Novamente, não encontramos qualquer relação entre a presença de correlação de cargas positiva, função biológica e propriedades físico-químicas, como peso molecular ou tamanho. Ambos os grupos contém, por exemplo, proteínas do leite e príons. Apesar do fato da LSZ e da α -LA terem origem genética comum e tamanho similar (MCKENZIE; WHITE, 1991), elas estão classificadas em diferentes grupos.

Informações complementares adicionais, como o número médio de resíduos de aminoácidos ($\langle N_{aa_p} \rangle$) para cada categoria de proteínas e carga média da proteína ($\langle z_p \rangle$) foram também dadas na Tabela 4, como uma indicação da diversidade de propriedades físico-químicas das proteínas nos diferentes conjuntos de dados. Na Tabela 5 os pesos moleculares foram associados às cargas e às estruturas.

Tabela 5 – Distribuição de carga líquida e massa molecular de algumas proteínas estudadas.

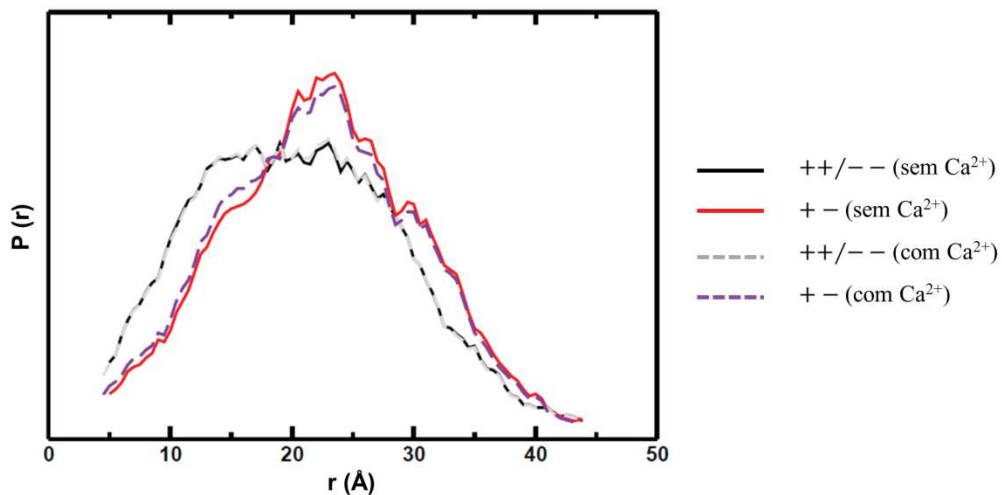
Proteína	Código PDB	z_p	M_m (kDa)
Acetilcolinesterase	1DX6	-13	61,3
Proteína de ligação à l-arabinose	1ABE	-7	33,3
β -caseína	(KUMOSINSKI; BROWN; FARRELL, 1993a)	-6	24,0
Hemoglobina A (deoxi, cadeia α) ^(b)	1HBB	+9	15,2
Lactoferrina	1BLF	+2	76,2
β -lactoglobulina	1BEB	-11	18,3
Álcool desidrogenase hepática	8ADH	+7	39,8
Lisozima	2LZT	+6	14,3
Proteína precursora do príon	1HJM	-10	12,6
Proteína periplasmática ligadora de maltose	1EZO	-20	40,7
Papaína	9PAP	-9	23,5
Pectina metil esterase	1GQ8	-1	34,2
Ribonuclease T1	IRGC	-15	11,1
Soroalbumina	1AO6	-16	66,6
Proteína S (príon na forma amilóide)	2RNM	+7	8,7
Termolisina	8TLN	-32	34,4
Titina (domínio A77-A78)	3LPW	-6	21,7
Inibidor da tripsina	4PTI	+2	6,5
α_{s1} -caseína	(KUMOSINSKI; BROWN; FARRELL, 1994)	-15	23,0
α_{s2} -caseína	(FARRELL et al., 2009)	-18	25,0
κ -caseína	(KUMOSINSKI; BROWN; FARRELL, 1993b)	-9	19,0
α -lactoalbumina (<i>Bos Taurus</i>)	1F6S	-8	14,2
α -lactoalbumina (<i>Capra hircus</i>)	1HFY	-7	14,2
Proteína S (príon na forma não amilóide)	2KJ3	+7	8,7
Tripsina	1C1P	-1	23,3

Códigos PDB, carga líquida z_p em unidades de carga elementar (e) e peso molecular M_m (kDa) de algumas proteínas foram investigados. Tabela agrupada em dois conjuntos específicos, sendo o primeiro conjunto composto por proteínas com correlação positiva em curtas distâncias. Já o segundo conjunto mostra ausência de correlação, ou correlação negativa entre os aminoácidos carregados a curtas distâncias.

A Figura 35 mostra a distribuição de probabilidade de resíduos carregados na α -LA (*Bos taurus*). Surpreendentemente, há uma forte correlação negativa estendendo-se até 20 Å de separação. Isto é provavelmente uma consequência da existência de aglomerados de cargas, citados anteriormente (DE VRIES; WEINBRECK; DE KRUIF, 2003; DE VRIES,

2004). A inclusão do íon de cálcio na análise praticamente não afetou os resultados. Os dados classificados como “com Ca^{2+} ” foram obtidos incluindo a carga do cálcio no cálculo das distribuições de cargas. Para um dado arquivo de coordenadas de proteínas, a localização dos grupos carregados da proteína não mudará se os heteroátomos forem incluídos ou não, quando se usa os dados estáticos do PDB, como usamos aqui. Todas as cargas permanecerão nas mesmas posições iniciais e terão as mesmas contribuições na análise geométrica. Considerando o número de cargas nas coordenadas de uma proteína, um raciocínio matemático simples pode mostrar que a ausência de uma carga terá um efeito bem menor nos resultados finais como observado nesta figura. A α -LA (código PDB 1F6S) contém 48 possíveis grupos carregados (7 glu, 13 asp, 12 lys, 1 arg, 3 his, 8 cys e 4 tyr), dependendo do pH da solução (HARRIS; TURNER, 2002). Em pH neutro, cisteínas e tirosinas são excluídas, resultando em 630 distâncias entre pares avaliados (não considerando o íon Ca^{2+}); incluindo o cálcio, esse valor aumenta 6%.

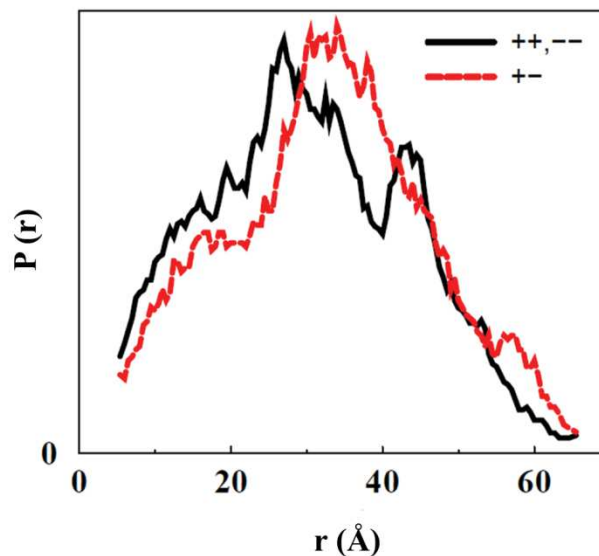
Figura 35 – Distribuição de resíduos positivos e negativos na α -LA (*Bos taurus*) PDB id 1F6S, nas formas *apo* e *halo*, em pH aproximadamente neutro.



Outra proteína do leite, a κ -CN, mostra o mesmo comportamento, com uma extensão da correlação negativa até quase a distância de 30 Å, como observado na Figura 36. No caso da α -LA, utilizamos coordenadas determinadas por raio-X (CHRYSINA; BREW; ACHARYA, 2000), enquanto para a κ -CN, as coordenadas foram determinadas por estudos computacionais (KUMOSINSKI; BROWN; FARRELL, 1993b). Outra α -LA, (código PDB

1HFY) originária de outro organismo (*Capra hircus*) comporta-se da mesma forma que a α -LA resultante do *Bos taurus* (1F6S). Um padrão similar é encontrado para outras proteínas como a α_{s1} -CN, α_{s2} -CN, tripsina, bem como príons.

Figura 36 – Distribuição de resíduos positivos e negativos na κ -caseína, em pH aproximadamente neutro.



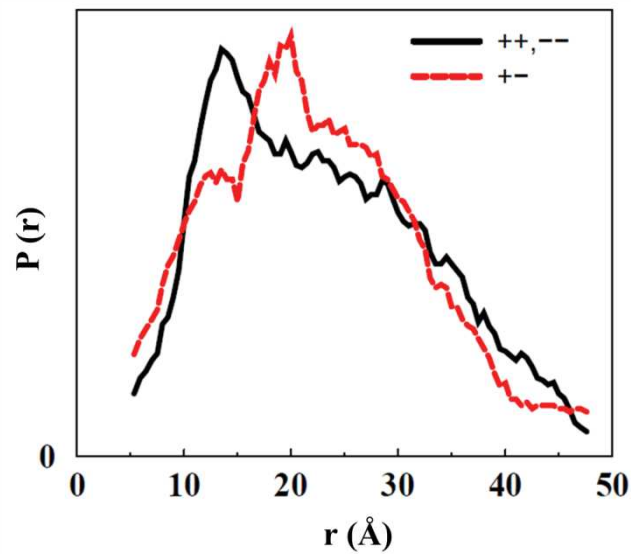
As coordenadas são do estudo computacional de Kumosinski e colaboradores (KUMOSINSKI; BROWN; FARRELL, 1993b).

Nessas comparações temos em média as duas funções de distribuição, $P_{++}(r)$ e $P_{--}(r)$. Uma análise mais minuciosa revela que $P_{--}(r) > P_{++}(r)$ em curtas distâncias. Uma possível interpretação é que os resíduos de ácido glutâmico e aspártico frequentemente criam sítios de ligação para cátions bivalentes, enquanto os resíduos básicos são encontrados mais frequentemente na superfície da proteína. Outra possibilidade imaginável é atribuir essa análise ao fato de que em pH 7 (pH em que foram feitos todos os nossos estudos de correlação de cargas) todos os grupos os quais as proteínas foram classificadas apresentaram carga média negativa. Assim, pode-se concluir que há maior quantidade de resíduos negativos do que positivos na média dessas proteínas, levando a $P_{--}(r) > P_{++}(r)$.

Algumas proteínas mostram uma preferência por dois resíduos opostamente carregados para estarem próximos, formando pares de íons como observado em estudos prévios (WADA; NAKAMURA, 1981; BARLOW; THORNTON, 1983). Por exemplo, a

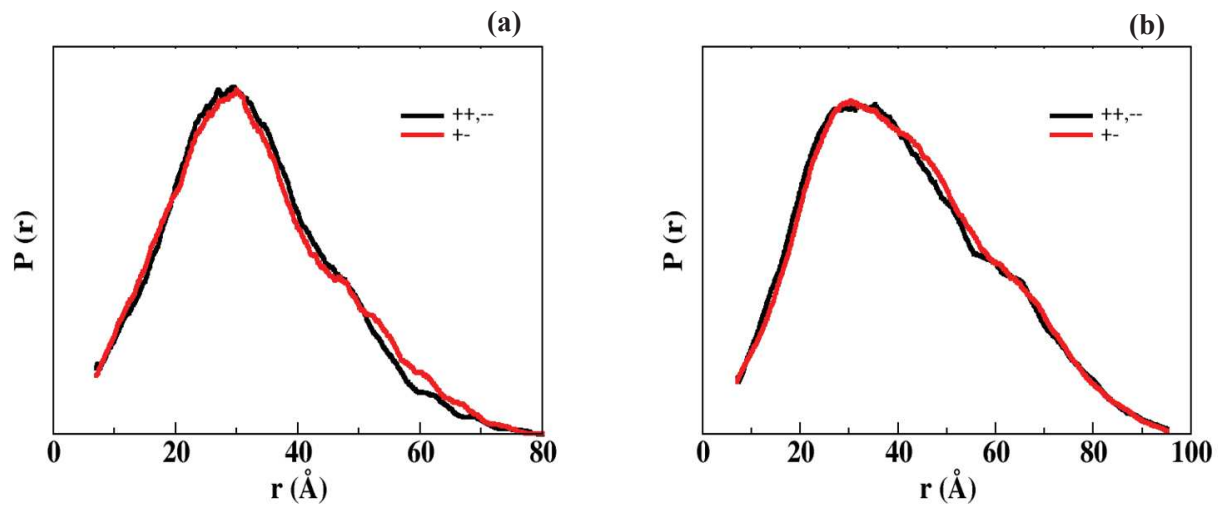
proteína priônica humana tem claramente este comportamento para curtas separações, como mostrado na Figura 37. Todavia, em separações intermediárias, há uma forte correlação negativa.

Figura 37 – Distribuição de resíduos positivos e negativos na proteína priônica humana (PDB id 1HJM), em pH aproximadamente neutro.



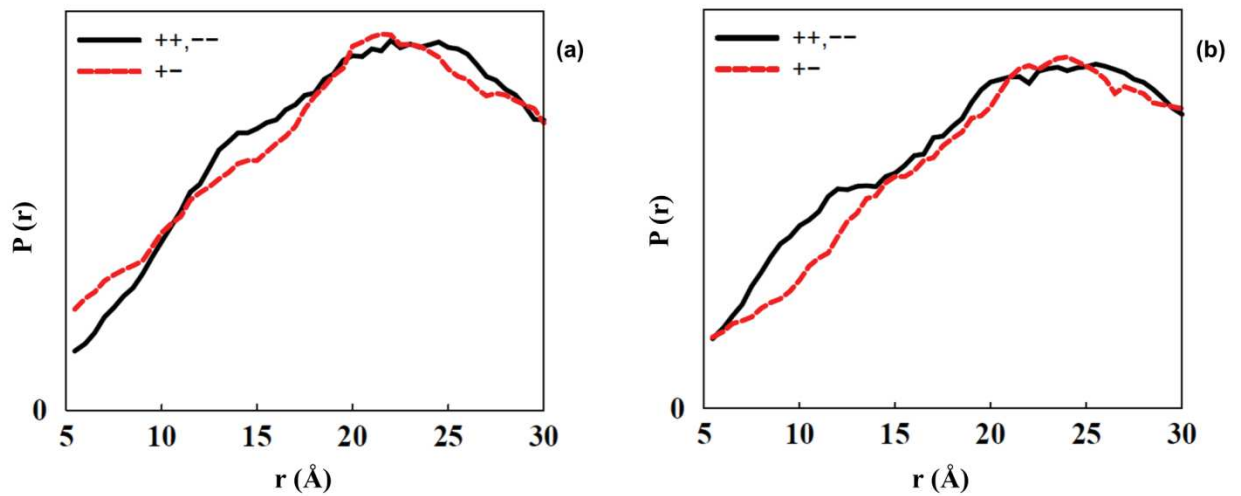
A proteína do soro do leite, α -LA, apresenta uma correlação de cargas negativa, enquanto outras proteínas do soro, como a β -LG e LF mostram uma correlação de cargas positiva bastante fraca ou ausente, como pode ser visualizado na Figura 38.

Figura 38 – Distribuição de resíduos positivos e negativos das proteínas β -LG (PDB id 1BEB) (a) e LF (PDB id 1BLF) (b), em pH aproximadamente neutro.



As distribuições de probabilidade para todas as proteínas dos conjuntos superior e inferior da Tabela 5 foram agrupadas e o resultado final pode ser visto na Figura 39. Em geral, a correlação positiva é normalmente pequena e presente somente em curtas separações. Há uma inversão do comportamento para separações médias ($10 < r < 20$ Å). O intervalo de exibição na Figura 39 foi limitado a 30 Å para que as diferenças sejam mais claramente visualizadas.

Figura 39 – Distribuição de resíduos positivos e negativos nas proteínas, em pH aproximadamente neutro.



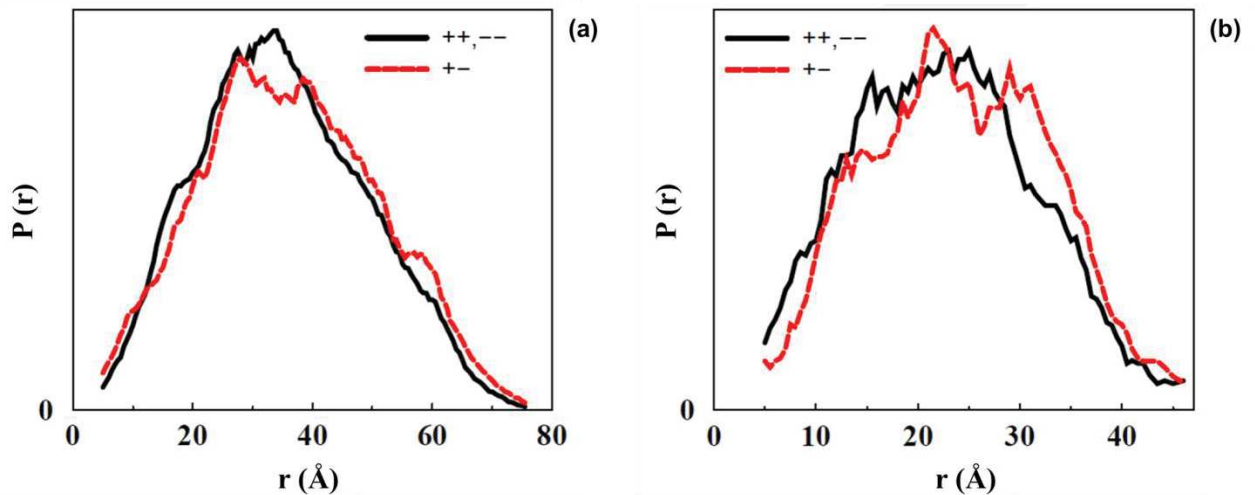
Em (a), proteínas do primeiro grupo da Tabela 5. Em (b), proteínas do segundo grupo da Tabela 5.

Percebe-se que algumas proteínas não tem uma correlação de cargas positiva. Isso pode ser explicado pela forma de cristalização das proteínas, que ocorre em altas concentrações de sal, onde as interações eletrostáticas são blindadas. Por exemplo, a cristalização da β -LG ocorre em uma solução 2 M de $(NH_4)_2SO_4$. Nessa solução, κ^{-1} é cerca de 1 Å e as interações eletrostáticas são efetivamente blindadas (HILL, 1986). Essa explicação pode ser parcialmente testada simulando nosso sistema de referência com 20 pares de íons, já discutido anteriormente, acrescido agora de um potencial de Coulomb blindado, com $\kappa^{-1} = 1$ Å, de acordo com a Equação 7 (já apresentada na seção Metodologia e modelo), ao invés da interação de Coulomb habitual. O resultado desta simulação não mostra correlação entre os íons, nem em alta ou baixa constante dielétrica do meio. Para estes casos com forte blindagem eletrostática devido a altas concentrações de sal, as probabilidades de distribuição simuladas coincidem com a distribuição randômica na Figura 33a. Assim, a ausência de correlação de cargas positiva ou até mesmo o aparecimento de correlações negativas não deve ser visto como uma surpresa para estruturas determinadas por raio-X, já que em condições normais de cristalização, a influência das interações eletrostáticas é menos significativa.

Aceitando esta explicação, deve-se esperar um comportamento levemente diferente para estruturas de proteínas determinadas pela técnica de RMN, já que a determinação destas requer condições fisiológicas de sal, com conseqüente menor blindagem eletrostática. A

Tabela 4 mostra, todavia, que estruturas determinadas por RMN normalmente apresentam uma fraca correlação de cargas, 69%, do que a média de 87%. A Figura 40 mostra dois exemplos de figuras determinadas por RMN; no primeiro deles, Figura 40a, a proteína ligadora de açúcar é uma proteína relativamente grande, com uma fraca correlação de cargas positiva. No segundo exemplo visualizado na Figura 40b, a tripsina, determinada também por RMN, mostra uma forte correlação de cargas negativa. A estrutura da proteína priônica humana utilizada na Figura 37 também foi obtida por RMN e mostra uma correlação positiva.

Figura 40 – Distribuição de resíduos positivos e negativos.



(a) Proteína ligadora de açúcar (PDB id 1EZO) e (b) tripsina (PDB id 1C1P).

Determinação de estruturas por RMN em baixas concentrações de sal, e a explicação da blindagem pode não ser válida para todos os casos. Todavia, as concentrações de proteína são normalmente bastante elevadas nos experimento de RMN, o que também poderia originar uma blindagem significativa (LINSE; JÖNSSON; CHAZIN, 1995).

A fim de procurar alguma relação entre a distribuição de resíduos e a existência de correlação de cargas positiva e negativa nas proteínas, analisamos também as porcentagens de cada um dos resíduos para três estruturas, apresentadas na Tabela 6. Foram utilizadas a proteína precursora do príon (PDB id 1HJM) e a LSZ (PDB id 2LZT), ambas com correlação de cargas positiva, e a α -LA (PDB id 1F6S), com anti-correlação de cargas. As porcentagens de cada um dos resíduos foram calculadas individualmente, e também agrupadas de acordo com a literatura (NELSON; COX, 2002) correspondendo a resíduos apolares, polares,

aromáticos, básicos e ácidos. Analisando as porcentagens encontradas, não foi possível relacionar as estruturas 1HJM e 2LZT, mesmo as duas mostrando correlação positiva. Outra comparação que fizemos foi entre as moléculas 2LZT e a 1F6S, apresentando correlações positiva e negativa, respectivamente. Mesmo essas duas proteínas apresentando origem evolutiva comum, não encontramos grandes discrepâncias na distribuição dos resíduos. A comparação entre as três moléculas inicialmente investigadas não permite detectar diferenças ou semelhanças significativas que possam justificar os padrões de correlação das três proteínas.

Tabela 6 – Distribuição de resíduos para a proteína precursora do príon (1HJM) e lisozima (2LZT) com correlação de cargas positiva e a proteína α -LA (1F6S) com correlação de cargas negativa.

		1HJM			1F6S			2LZT		
		AA	%		AA	%		AA	%	
Apolares	GLY	5	4,81		6	4,88		12	9,30	
	ALA	2	1,92		3	2,44		12	9,30	
	VAL	7	6,73	26,92	6	4,88	30,08	6	4,65	35,66
	LEU	2	1,92		13	10,57		8	6,20	
	MET	7	6,73		1	0,81		2	1,55	
	ILE	5	4,81		8	6,50		6	4,65	
Polares	SER	5	4,81		7	5,69		10	7,75	
	THR	9	8,65		7	5,69		7	5,43	
	CYS	2	1,92	31,73	8	6,50	30,89	8	6,20	34,11
	PRO	3	2,88		2	1,63		2	1,55	
	ASN	7	6,73		8	6,50		14	10,85	
	GLN	7	6,73		6	4,88		3	2,33	
Aromáticos	PHE	3	2,88		4	3,25		3	2,33	
	TYR	11	10,58	13,46	4	3,25	9,76	3	2,33	9,30
	TRP	0	0,00		4	3,25		6	4,65	
Básicos (+)	LYS	3	2,88		12	9,76		6	4,65	
	ARG	8	7,69	14,42	1	0,81	13,01	11	8,53	13,95
	HIS	4	3,85		3	2,44		1	0,78	
Ácidos (-)	ASP	5	4,81	13,46	13	10,57	16,26	7	5,43	6,98
	GLU	9	8,65		7	5,69		2	1,55	
Total		104	100	100	123	100	100	129	100	100

7. Conclusões

Empregando simulações em um vasto número de diferentes condições experimentais, confirmamos que as proteínas do soro do leite tem suas características atrativas ótimas no pH natural do leite, em condições fisiológicas. Sal e pH podem ser usados para manipular a força relativa da energia livre mínima de associação.

Através de simulações MC foi definida a janela de pH nas condições fisiológicas para os complexos α -LA-LF e β -LG-LF. Encontramos que não se espera associação para os complexos α -LA- β -LG nessas condições. A energia livre de interação computada entre α -LA-LF revelou uma atração fraca entre elas em contraste com a forte atração do outro complexo, β -LG-LF, explicando inclusive incertezas experimentais. As forças que governam as complexações com a LF são predominantemente devido a uma associação de interações eletrostáticas (carga-carga, carga-dipolo e interação carga induzida devido às flutuações de cargas). Os resultados presentes podem guiar aplicações práticas na indústria, para as proteínas do soro, por exemplo, para produzir bebidas estáveis.

Dessa forma, conseguimos mostrar também como a bioinformática pode contribuir no entendimento de problemas práticos, guiando sua funcionalização.

Na segunda etapa do trabalho, um grande número de estruturas de proteínas foi utilizado, tanto elucidadas por raio-X quanto RMN, para calcular as correlações entre resíduos carregados em todas as estruturas de proteínas disponíveis no PDB. O resultado geral é que a maioria das proteínas (não todas) exibe uma correlação positiva entre resíduos opostamente carregados, mas somente a curtas distâncias. Com base nessa análise, concluímos que as interações eletrostáticas são importantes na estabilização da maioria das estruturas terciárias de proteínas.

As condições de cristalização para proteínas são normalmente muito severas, com altas concentrações de sal, levando a forte blindagem eletrostática, o que poderia ser explicado pela falta de correlações de cargas encontrada em torno de 13% das proteínas existentes no PDB. Estes resultados podem ser úteis para uma análise geral das forças e importância das interações eletrostáticas nas proteínas. Pode ser inclusive que aspectos funcionais sejam mais importantes que a minimização da energia livre eletrostática.

Desta parte inicial, o resultado mais importante é a não observação de picos bem definidos na correlação de cargas, como seria esperado em um meio de baixa constante

dielétrica. Pode-se então deduzir que a utilização de um dielétrico contínuo de alta constante dielétrica e único para uma proteína, é mais condizente com a realidade do que considerar a proteína como um meio com baixa constante dielétrica, em modelos de solvente contínuo, sendo uma justificativa experimental para as aproximações assumidas nos modelos empregados na complexação. Tal resultado soma-se a várias outras evidências anteriormente reportadas na literatura.

8. Referências

- ADŽIĆ, N.; PODGORNIK, R. Charge regulation in ionic solutions: Thermal fluctuations and Kirkwood-Schumaker interactions. **Physical Review E**, v. 91, n. 2, p. 22715, 2015.
- ALLMELING, C. et al. Spider silk fibres in artificial nerve constructs promote peripheral nerve regeneration. **Cell Proliferation**, v. 41, n. 3, p. 408–420, 2008.
- ANEMA, S. G.; DE KRUIF, C. G. K. Complex coacervates of lactotransferrin and β -lactoglobulin. **Journal of colloid and interface science**, v. 430, p. 214–20, 15 set. 2014.
- ANTOSIEWICZ, J.; MCCAMMON, J. A.; GILSON, M. K. Prediction of pH-dependent properties of proteins. **Journal of molecular biology**, v. 238, n. 3, p. 415–36, 6 maio 1994.
- ANTOSIEWICZ, J.; MCCAMMON, J. A.; GILSON, M. K. The determinants of pKas in proteins. **Biochemistry**, v. 35, n. 24, p. 7819–33, 18 jun. 1996.
- ARAI, Y. Computational chemistry on catalyst technology and international collaboration. **Catalysis Today**, v. 23, n. 4, p. 439–448, abr. 1995.
- ARCHONTIS, G.; SIMONSON, T. Proton binding to proteins: a free-energy component analysis using a dielectric continuum model. **Biophysical journal**, v. 88, n. 6, p. 3888–904, 2005.
- ARKIN, M. Protein-protein interactions and cancer: Small molecules going in for the kill. **Current Opinion in Chemical Biology**, v. 9, n. 3, p. 317–324, 2005.
- AUDIC, J.-L.; CHAUFER, B.; DAUFIN, G. Non-food applications of milk components and dairy co-products: A review. **Le Lait**, v. 83, n. 6, p. 417–438, nov. 2003.
- AZEVEDO, V. M. et al. Whey protein isolate biodegradable films: Influence of the citric acid and montmorillonite clay nanoparticles on the physical properties. **Food Hydrocolloids**, v. 43, p. 252–258, 2014.
- BARLOW, D. J.; THORNTON, J. M. Ion-pairs in proteins. **Journal of molecular biology**, v. 168, n. 4, p. 867–885, ago. 1983.
- BATT, C. A.; BRADY, J.; SAWYER, L. Design improvements of β -lactoglobulin. **Trends in Food Science & Technology**, v. 5, n. 8, p. 261–265, ago. 1994.
- BERMAN, H. M. et al. The Protein Data Bank. **Nucleic acids research**, v. 28, n. 1, p. 235–42, 1 jan. 2000.

- BERRY, S. Entropy, irreversibility and evolution. **Journal of theoretical biology**, v. 175, n. 2, p. 197–202, 12 jul. 1995.
- BHARTI, B. **Adsorption, Aggregation and Structure Formation in Systems of Charged Particles**. Cham: Springer International Publishing, 2014.
- BIEDA, M. Kepler for ‘Omics Bioinformatics. **Procedia Computer Science**, v. 9, n. 0, p. 1635–1638, 2012.
- BIER, D. et al. Stabilization of Protein–Protein Interactions in chemical biology and drug discovery. **Progress in Biophysics and Molecular Biology**, v. 119, n. 1, p. 10–19, 2015.
- BOGUSKI, M. S. Bioinformatics. **Current Opinion in Genetics & Development**, v. 4, n. 3, p. 383–388, jun. 1994.
- BOLOURI, H. Modeling genomic regulatory networks with big data. **Trends in Genetics**, v. 30, n. 5, p. 182–191, 2014.
- BÖSZE, Z. **Bioactive Components of Milk**. New York, NY: Springer New York, 2008. v. 606.
- BOUYER, E. et al. Proteins, polysaccharides, and their complexes used as stabilizers for emulsions: Alternatives to synthetic surfactants in the pharmaceutical field? **International Journal of Pharmaceutics**, v. 436, n. 1–2, p. 359–378, 2012.
- BRASIL, C. R. S.; DELBEM, A. C. B.; DA SILVA, F. L. B. Multiobjective evolutionary algorithm with many tables for purely ab initio protein structure prediction. **Journal of computational chemistry**, v. 34, n. 20, p. 1719–34, 2013.
- BROWNLOW, S. et al. Bovine β -lactoglobulin at 1.8 Å resolution — still an enigmatic lipocalin. **Structure**, v. 5, n. 4, p. 481–495, abr. 1997.
- BRYANT, A.-M.; CAI, S.; SINGH, B. R. Comparative immunochemical characteristics of botulinum neurotoxin type A and its associated proteins. **Toxicon: official journal of the International Society on Toxinology**, v. 72, p. 126–32, 2013.
- CHANASATTRU, W.; DECKER, E. A.; MCCLEMENTS, D. J. Influence of glycerol and sorbitol on thermally induced droplet aggregation in oil-in-water emulsions stabilized by β -lactoglobulin. **Food Hydrocolloids**, v. 23, n. 2, p. 253–261, mar. 2009.
- CHANG, R. **Physical chemistry for the chemical and biological sciences**. Sausalito, Calif: Univ. Science Books, 2000.

CHANPHAI, P.; BEKALE, L.; TAJMIR-RIAAHI, H. A. Effect of hydrophobicity on protein–protein interactions. **European Polymer Journal**, v. 67, p. 224–231, jun. 2015.

CHENG, S.; BROOKS, C. L. Viral capsid proteins are segregated in structural fold space. **PLoS computational biology**, v. 9, n. 2, p. 1–12, 2013.

CHERAK, S. J.; TURNER, R. J. Influence of GTP on system specific chaperone – Twin arginine signal peptide interaction. **Biochemical and Biophysical Research Communications**, v. 465, n. 4, p. 753–757, 2015.

CHI, E. Y. et al. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. **Pharmaceutical Research**, v. 20, n. 9, p. 1325–1336, 2003.

CHRYSINA, E. D.; BREW, K.; ACHARYA, K. R. Crystal Structures of Apo- and Holo-bovine α -Lactalbumin at 2.2-Å Resolution Reveal an Effect of Calcium on Inter-lobe Interactions. **Journal of Biological Chemistry**, v. 275, n. 47, p. 37021–37029, 24 nov. 2000.

CLARE, D. A.; CATIGNANI, G. L.; SWAISGOOD, H. E. Biodefense properties of milk: the role of antimicrobial proteins and peptides. **Current pharmaceutical design**, v. 9, n. 16, p. 1239–1255, 2003.

COUCH, G. S. et al. Annotating PDB files with scene information. **Journal of Molecular Graphics**, v. 13, n. 3, p. 153–158+193, 1995.

COWAN-JACOB, S. W.; MÖBITZ, H.; FABBRO, D. Structural biology contributions to tyrosine kinase drug discovery. **Current opinion in cell biology**, v. 21, n. 2, p. 280–7, abr. 2009.

COZZONE, A. J. Proteins : Fundamental Chemical Properties. **Life Sciences**, p. 1–10, 2002.

DA SILVA, F. L. B. **Statistical Mechanical Studies of Aqueous solutions and Biomolecular Systems**. [s.l.] Lund University, 2000.

DA SILVA, F. L. B. et al. Electrostatics analysis of the mutational and pH effects of the N-terminal domain self-association of the major ampullate spidroin. **Soft Matter**, v. 12, n. 25, p. 5600–5612, 2016.

DA SILVA, F. L. B.; BOSTRÖM, M.; PERSSON, C. Effect of charge regulation and ion-dipole interactions on the selectivity of protein-nanoparticle binding. **Langmuir: the ACS journal of surfaces and colloids**, v. 30, n. 14, p. 4078–83, 15 abr. 2014.

DA SILVA, F. L. B.; JÖNSSON, B. Polyelectrolyte–protein complexation driven by charge regulation. **Soft Matter**, v. 5, n. 15, p. 2862, 2009.

- DA SILVA, F. L. B.; JÖNSSON, B.; PENFOLD, R. A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations. **Protein science: a publication of the Protein Society**, v. 10, n. 7, p. 1415–1425, 2001.
- DALGLEISH, D. G. Bovine milk protein properties and the manufacturing quality of milk. **Livestock Production Science**, v. 35, n. 1–2, p. 75–93, 1992.
- DAVIS, A. M.; TEAGUE, S. J.; KLEYWEGT, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. **Angewandte Chemie (International ed. in English)**, v. 42, n. 24, p. 2718–36, 2003.
- DE CARVALHO, S. J.; FENLEY, M. O.; DA SILVA, F. L. B. Protein–Ion Binding Process on Finite Macromolecular Concentration. A Poisson–Boltzmann and Monte Carlo Study. **The Journal of Physical Chemistry B**, v. 112, n. 51, p. 16766–16776, 25 dez. 2008.
- DE CARVALHO, S. J.; GHIOTTO, R. C. T.; DA SILVA, F. L. B. Monte Carlo and modified Tanford-Kirkwood results for macromolecular electrostatics calculations. **Journal of Physical Chemistry B**, v. 110, n. 17, p. 8832–8839, 2006.
- DE ROOS, A. L.; WALSTRA, P.; GEURTS, T. J. The association of lysozyme with casein. **International Dairy Journal**, v. 8, n. 4, p. 319–324, 1998.
- DE VOS, W. M. et al. Field theoretical analysis of driving forces for the uptake of proteins by like-charged polyelectrolyte brushes: effects of charge regulation and patchiness. **Langmuir: the ACS journal of surfaces and colloids**, v. 26, n. 1, p. 249–59, 2010.
- DE VRIES, R. Monte Carlo simulations of flexible polyanions complexing with whey proteins at their isoelectric point. **The Journal of chemical physics**, v. 120, n. 7, p. 3475–3481, 2004.
- DE VRIES, R.; STUART, M. C. Theory and simulations of macroion complexation. **Current opinion in colloid & interface science**, v. 11, n. 5, p. 295–301, 2006.
- DE VRIES, R.; WEINBRECK, F.; DE KRUIF, C. G. Theory of polyelectrolyte adsorption on heterogeneously charged surfaces applied to soluble protein–polyelectrolyte complexes. **The Journal of chemical physics**, v. 118, n. 10, p. 4649–4659, 2003.
- DE WIT, J. N. Thermal behaviour of bovine beta-lactoglobulin at temperatures up to 150 °C. A review. **Trends in Food Science and Technology**, v. 20, n. 1, p. 27–34, 2009.
- DELBONI, L. A.; DA SILVA, F. L. B. On the complexation of whey proteins. **Food Hydrocolloids**, v. 55, p. 89–99, 2016.

- DELRE, G.; PULLMAN, B.; YONEZAWA, T. Electronic Structure of the Alpha-Amino Acids of Proteins. I. Charge Distributions and Proton Chemical Shifts. **Biochimica et biophysica acta**, v. 75, p. 153–82, 1963.
- DEMCHUK, E.; WADE, R. C. Improving the Continuum Dielectric Approach to Calculating pK_as of Ionizable Groups in Proteins. **The Journal of Physical Chemistry**, v. 100, n. 43, p. 17373–17387, jan. 1996.
- DICKINSON, E. Interfacial structure and stability of food emulsions as affected by protein–polysaccharide interactions. **Soft Matter**, v. 4, n. 5, p. 932, 2008.
- DICKINSON, E. Stabilising emulsion-based colloidal structures with mixed food ingredients. **Journal of the Science of Food and Agriculture**, v. 93, n. 4, p. 710–721, 2013.
- DICKINSON, E.; DAVIES, E. Influence of ionic calcium on stability of sodium caseinate emulsions. **Colloids and Surfaces B: Biointerfaces**, v. 12, n. 3–6, p. 203–212, 1999.
- DILL, K. A. Dominant forces in protein folding. **Biochemistry**, v. 29, n. 31, p. 7133–7155, ago. 1990.
- DILL, K. A. et al. The Protein Folding Problem. **Annual Review of Biophysics**, v. 37, n. 1, p. 289–316, 2008.
- DOMINY, B. N.; MINOUX, H.; BROOKS, C. L. An electrostatic basis for the stability of thermophilic proteins. **Proteins: Structure, Function, and Bioinformatics**, v. 57, n. 1, p. 128–141, 22 jun. 2004.
- DONALD, A. M. Aggregation in β -lactoglobulin. **Soft Matter**, v. 4, n. 6, p. 1147, 2008.
- DUDEV, T.; LIM, C. Metal Binding in Proteins: The Effect of the Dielectric Medium. **The Journal of Physical Chemistry B**, v. 104, n. 15, p. 3692–3694, abr. 2000.
- EGAN, T. et al. Cold-set whey protein microgels as pH modulated immobilisation matrices for charged bioactives. **Food chemistry**, v. 156, p. 197–203, 1 ago. 2014.
- EISENBERG, D. et al. Hydrophobic Moments and Protein Structure. **Faraday Symp. Chem. Soc.**, v. 17, p. 109–120, 1982.
- EL-LOLY, M. M.; MAHFOUZ, M. B. Lactoferrin in Relation to Biological Functions and Applications: A Review. **International Journal of Dairy Science**, v. 6, n. 2, p. 79–111, 1 fev. 2011.
- ENGLAND, J. L.; HARAN, G. Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. **Annual Review of Physical Chemistry**, v.

62, p. 257–277, 2011.

EVERSHED, R. P. et al. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. **Nature**, v. 455, n. 7212, p. 528–531, 2008.

FARRELL, H. M. et al. Nomenclature of the proteins of cows' milk—sixth revision. **Journal of dairy science**, v. 87, n. 6, p. 1641–74, 2004.

FARRELL, H. M. et al. Review of the chemistry of α S2-casein and the generation of a homologous molecular model to explain its properties. **Journal of Dairy Science**, v. 92, n. 4, p. 1338–1353, abr. 2009.

FAUSTINO, A. F. et al. Dengue virus capsid protein interacts specifically with very low-density lipoproteins. **Nanomedicine: nanotechnology, biology, and medicine**, v. 10, n. 1, p. 247–55, jan. 2014.

FISHMAN, G. S. **Monte Carlo: concepts, algorithms, and applications**. New York, NY [u.a.]: Springer, 1999.

FOGOLARI, F. et al. Electrostatic properties of bovine beta-lactoglobulin. **Proteins**, v. 39, n. 4, p. 317–330, 2000.

FRENKEL, D.; SMIT, B. **Understanding Molecular Simulation: From Algorithms to Applications**. 1st. ed. Orlando, FL, USA: Academic Press, Inc., 1996.

FRIEDMAN, H. L. Introduction. **Faraday Discussions of the Chemical Society**, v. 64, p. 7, 1977.

FUJISE, N. et al. Distribution of the calcium binding proteins, calbindin D-28K and parvalbumin, in the subicular complex of the adult mouse. **Neuroscience Research**, v. 22, n. 1, p. 89–107, mar. 1995.

FUSHIKI, M. et al. Electrostatic interactions in protein solution--a comparison between Poisson-Boltzmann and Monte Carlo calculations. **Biopolymers**, v. 31, n. 10, p. 1149–1158, 1991.

GAHLAUT, A. et al. Proteomics & metabolomics: Mapping biochemical regulations. **Drug Invention Today**, v. 5, n. 4, p. 321–326, 2013.

GAJENDRAGADKAR, C. N.; GOGATE, P. R. Intensified recovery of valuable products from whey by use of ultrasound in processing steps – A review. **Ultrasonics Sonochemistry**, v. 32, p. 102–118, set. 2016.

GAUTHIER, S. F.; POULIOT, Y. Functional and Biological Properties of Peptides Obtained

by Enzymatic Hydrolysis of Whey Proteins. **Journal of Dairy Science**, v. 86, n. July 2002, p. E78–E87, 2003.

GERSTEIN, M.; JANSEN, R. The current excitement in bioinformatics - Analysis of whole-genome expression data: How does it relate to protein structure and function? **Current Opinion in Structural Biology**, v. 10, n. 5, p. 574–584, 2000.

GHANBARZADEH, B.; OROMIEHI, A. R. Biodegradable biocomposite films based on whey protein and zein: Barrier, mechanical properties and AFM analysis. **International Journal of Biological Macromolecules**, v. 43, n. 2, p. 209–215, 2008.

GIBRAT, J. F.; MADEJ, T.; BRYANT, S. H. Surprising similarities in structure comparison. **Current opinion in structural biology**, v. 6, n. 3, p. 377–85, 1996.

GIRARD, M.; SCHAFFER-LEQUART, C. Attractive interactions between selected anionic exopolysaccharides and milk proteins. **Food Hydrocolloids**, v. 22, n. 8, p. 1425–1434, 2008.

GOTTSCHALK, M. et al. Protein self-association in solution: the bovine beta -lactoglobulin dimer and octamer. **Protein science: a publication of the Protein Society**, v. 12, n. 11, p. 2404–2411, 2003.

GUILLET-NICOLAS, R. et al. pH-responsive nutraceutical-mesoporous silica nanoconjugates with enhanced colloidal stability. **Angewandte Chemie (International ed. in English)**, v. 52, n. 8, p. 2318–22, 18 fev. 2013.

GUTTERIDGE, A.; THORNTON, J. M. Understanding nature's catalytic toolkit. **Trends in biochemical sciences**, v. 30, n. 11, p. 622–9, 2005.

GVRITISHVILI, A. G.; GRIBENKO, A. V; MAKHATADZE, G. I. Cooperativity of complex salt bridges. **Protein science: a publication of the Protein Society**, v. 17, n. 7, p. 1285–90, 2008.

HALGREN, T. A.; DAMM, W. Polarizable force fields. **Current opinion in structural biology**, v. 11, n. 2, p. 236–42, 2001.

HAMAKER, H. C. The London—van der Waals attraction between spherical particles. **Physica**, v. 4, n. 10, p. 1058–1072, 1937.

HARRIS, T. K.; TURNER, G. J. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. **IUBMB life**, v. 53, n. 2, p. 85–98, fev. 2002.

HATTORI, T. et al. Characterization of polyanion-protein complexes by frontal analysis continuous capillary electrophoresis and small angle neutron scattering: Effect of polyanion flexibility. **Analytical Biochemistry**, v. 342, n. 2, p. 229–236, 2005.

- HAVEA, P.; SINGH, H.; CREAMER, L. K. Characterization of heat-induced aggregates of beta-lactoglobulin, alpha-lactalbumin and bovine serum albumin in a whey protein concentrate environment. **The Journal of dairy research**, v. 68, n. 3, p. 483–97, 15 ago. 2001.
- HE, Y.; XU, J.; PAN, X.-M. A statistical approach to the prediction of pKa values in proteins. **Proteins: Structure, Function, and Bioinformatics**, v. 69, n. 1, p. 75–82, 22 jun. 2007.
- HEKMAN, A. Association of lactoferrin with other proteins, as demonstrated by changes in electrophoretic mobility. **Biochimica et biophysica acta**, v. 251, n. 3, p. 380–7, 28 dez. 1971.
- HILDEBRANDT, A. et al. Electrostatic potentials of proteins in water: a structured continuum approach. **Bioinformatics**, v. 23, n. 2, p. e99–e103, 15 jan. 2007.
- HILL, T. L. Approximate calculation of the electrostatic free energy of nucleic acids and other cylindrical macromolecules. **Archives of biochemistry and biophysics**, v. 57, n. 1, p. 229–39, 1955.
- HILL, T. L. **Statistical mechanics : principles and selected applications**. New York: Dover Publications, 1956.
- HILL, T. L. **An introduction to statistical thermodynamics**. New York: Dover Publications, 1986.
- HITCHCOCK, B. Y. D. I. et al. the Isoelectric Point of a Standard Gelatin. **The Journal of general physiology**, v. 2, p. 685–699, 1931.
- HOFESTTIDT, R. Computer Science, Supercomputing and Biology-Bioinformatics. **Simulation Practice and Theory**, v. 2, n. 4–5, p. P7–P42, 1995.
- HOLDER, A. J.; KILWAY, K. V. Rational design of dental materials using computational chemistry. **Dental materials: official publication of the Academy of Dental Materials**, v. 21, n. 1, p. 47–55, jan. 2005.
- HOLM, L. et al. Searching protein structure databases with DaliLite v.3. **Bioinformatics (Oxford, England)**, v. 24, n. 23, p. 2780–1, 2008.
- HUBBARD, T. J. et al. SCOP: a structural classification of proteins database. **Nucleic acids research**, v. 25, n. 1, p. 236–9, 1997.
- ISHII, N. et al. Toward large-scale modeling of the microbial cell for computer simulation. **Journal of Biotechnology**, v. 113, n. 1–3, p. 281–294, 2004.

- ISRAELACHVILI, J. N. **Intermolecular and Surface Forces**. [s.l.] Elsevier Science, 2010.
- JACOBSON, K. A. New paradigms in GPCR drug discovery. **Biochemical Pharmacology**, v. 98, n. 4, p. 541–555, 2015.
- JAIN, N.; BHATTACHARYA, M.; MUKHOPADHYAY, S. Chain collapse of an amyloidogenic intrinsically disordered protein. **Biophysical Journal**, v. 101, n. 7, p. 1720–1729, 2011.
- JAUREGI, P. et al. Astringency reduction in red wine by whey proteins. **Food chemistry**, v. 199, p. 547–55, 15 maio 2016.
- JEFFERYS, B. R.; KELLEY, L. A.; STERNBERG, M. J. E. Protein Folding Requires Crowd Control in a Simulated Cell. **Journal of Molecular Biology**, v. 397, n. 5, p. 1329–1338, 2010.
- JENG, S. Y. et al. Characterization and partial purification of bovine alpha-lactalbumin and beta-casein produced in milk of transgenic mice. **Journal of dairy science**, v. 80, n. 12, p. 3167–3175, 1997.
- JOHNSON, M. L.; HOLT, J. M.; ACKERS, G. K. Biothermodynamics. In: **Methods in Enzymology**. Methods in Enzymology. Vol. 466 ed. [s.l.] Elsevier Science, 2009. p. 455–475.
- JÖNSSON, B. **The Thermodynamics of Ionic Amphiphilic–Water Systems – A Theoretical Analysis**. [s.l.] Lund University, 1981.
- JÖNSSON, B.; LUND, M.; DA SILVA, F. L. B. Chapter 9. Electrostatics in Macromolecular Solutions. In: **Food Colloids**. Cambridge: Royal Society of Chemistry, 2007a. p. 129–154.
- JÖNSSON, B.; LUND, M.; DA SILVA, F. L. B. Electrostatics in macromolecular solutions. **Food Colloids: Self-Assembly and Material Science**, v. 302, p. 129–154, 2007b.
- JÖNSSON, B.; SVENSSON, B. Monte Carlo Simulation of Electrostatic Interactions in Biomolecules. In: DONIACH, S. (Ed.). **Statistical Mechanics, Protein Structure, and Protein Substrate Interactions**. Boston, MA: Springer US, 1994. p. 75–85.
- JU, Z. Y.; HETTIARACHCHY, N.; KILARA, A. Thermal properties of whey protein aggregates. **Journal of Dairy Science**, v. 82, n. 9, p. 1882–1889, 1999.
- KAPLAN, D.; MCGRATH, K. **Protein-Based Materials**. [s.l.] Birkhäuser Boston, 2012.
- KAPOOR, S.; RAFIQ, A.; SHARMA, S. Protein Engineering and Its Applications in Food Industry. **Critical reviews in food science and nutrition**, v. 8398, n. January, p. 00–00, 11 jun. 2015.

- KAUZMANN, W. Some factors in the interpretation of protein denaturation. **Advances in protein chemistry**, v. 14, p. 1–63, 1959.
- KESVATERA, T. et al. Binding of Ca²⁺ to calbindin D9k: structural stability and function at high salt concentration. **Biochemistry**, v. 33, n. 47, p. 14170–6, 1994.
- KESVATERA, T. et al. Ionization behavior of acidic residues in calbindin D(9k). **Proteins: Structure, Function and Genetics**, v. 37, n. 1, p. 106–115, 1999.
- KHANZADI, M. et al. Physical and mechanical properties in biodegradable films of whey protein concentrate–pullulan by application of beeswax. **Carbohydrate Polymers**, v. 118, p. 24–29, 2015.
- KIM, H. J.; DECKER, E. A.; MCCLEMENTS, D. J. Role of postadsorption conformation changes of b-lactoglobulin on its ability to stabilize oil droplets against flocculation during heating at neutral pH. **Langmuir**, v. 18, n. 20, p. 7577–7583, 2002.
- KING, G.; LEE, F. S.; WARSHEL, A. Microscopic simulations of macroscopic dielectric constants of solvated proteins. **The Journal of Chemical Physics**, v. 95, n. 6, p. 4366–4377, 1991.
- KINSELLA, J. E.; WHITEHEAD, D. M. Proteins in Whey: Chemical, Physical, and Functional Properties. **Advances in Food and Nutrition Research**, v. 33, n. C, p. 343–438, 1989.
- 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**, v. 38, n. 10, p. 863–871, 1 out. 1952.
- KMIECIK, S. et al. Coarse-Grained Protein Models and Their Applications. **Chemical Reviews**, v. 116, n. 14, p. 7898–7936, 27 jul. 2016.
- KO, J. et al. Prediction of active sites for protein structures from computed chemical properties. **Bioinformatics**, v. 21, n. Suppl 1, p. i258–i265, 2005.
- KONTOPIDIS, G.; HOLT, C.; SAWYER, L. Invited review: beta-lactoglobulin: binding properties, structure, and function. **Journal of dairy science**, v. 87, n. 4, p. 785–96, 2004.
- KORHONEN, H. et al. Impact of processing on bioactive proteins and peptides. **Trends in Food Science and Technology**, v. 9, n. 8–9, p. 307–319, 1998.
- KUEHNER, D. E. et al. Lysozyme net charge and ion binding in concentrated aqueous electrolyte solutions. **Journal of Physical Chemistry B**, v. 103, n. 8, p. 1368–1374, 1999.

- KUKIC, P. et al. Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. **Journal of the American Chemical Society**, v. 135, n. 45, p. 16968–16976, 13 nov. 2013.
- KUMAR, S.; NUSSINOV, R. Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins. **Biophysical Journal**, v. 83, n. 3, p. 1595–1612, set. 2002.
- KUMOSINSKI, T. F.; BROWN, E. M.; FARRELL, H. M. Three-dimensional molecular modeling of bovine caseins: an energy-minimized beta-casein structure. **Journal of dairy science**, v. 76, n. 4, p. 931–45, abr. 1993a.
- KUMOSINSKI, T. F.; BROWN, E. M.; FARRELL, H. M. Three-dimensional molecular modeling of bovine caseins: a refined, energy-minimized kappa-casein structure. **Journal of dairy science**, v. 76, n. 9, p. 2507–20, set. 1993b.
- KUMOSINSKI, T. F.; BROWN, E. M.; FARRELL, H. M. Predicted Energy-Minimized α s1-Casein Working Model. In: **Molecular Modeling**. [s.l: s.n.]. p. 368–390.
- KURUT, A. et al. Anisotropic Interactions in Protein Mixtures: Self Assembly and Phase Behavior in Aqueous Solution. **The Journal of Physical Chemistry Letters**, v. 3, n. 6, p. 731–734, 2012.
- KURUT, A.; DICKO, C.; LUND, M. Dimerization of Terminal Domains in Spiders Silk Proteins Is Controlled by Electrostatic Anisotropy and Modulated by Hydrophobic Patches. **ACS Biomaterials Science & Engineering**, v. 1, n. 6, p. 363–371, 2015.
- LAMPREAVE, F. et al. Interaction of bovine lactoferrin with other proteins of milk whey. **International journal of biological macromolecules**, v. 12, n. 1, p. 2–5, 1990.
- LAOS, K.; BROWNSEY, G.; RING, S. Interactions between furcellaran and the globular proteins bovine serum albumin and β -lactoglobulin. **Carbohydrate Polymers**, v. 67, n. 1, p. 116–123, 2007.
- LASCHEWSKY, A. Structures and Synthesis of Zwitterionic Polymers. **Polymers**, v. 6, n. 5, p. 1544–1601, 23 maio 2014.
- LAZZARO, F. et al. Gradual disaggregation of the casein micelle improves its emulsifying capacity and decreases the stability of dairy emulsions. **Food Hydrocolloids**, v. 63, p. 189–200, fev. 2017.
- LEE, J.; MILLER, B. T.; BROOKS, B. R. Computational scheme for pH-dependent binding free energy calculation with explicit solvent. **Protein science: a publication of the Protein Society**, v. 25, n. 1, p. 231–43, jan. 2016.

- LEE, J.; WU, S.; ZHANG, Y. Ab Initio Protein Structure Prediction. In: RIGDEN, D. J. (Ed.). **From Protein Structure to Function with Bioinformatics**. Dordrecht: Springer Netherlands, 2009. p. 3–25.
- LI, Y. et al. Light Scattering Studies of the Binding of Bovine Serum Albumin to a Cationic Polyelectrolyte. **Biopolymers**, v. 38, p. 527–533, 1996.
- LINSE, S.; JÖNSSON, B.; CHAZIN, W. The effect of protein concentration on ion binding. **Proceedings of the National Academy of Sciences**, v. 92, n. May, p. 4748–4752, 1995.
- LÖFFLER, G.; SCHREIBER, H.; STEINHAUSER, O. Calculation of the dielectric properties of a protein and its solvent: theory and a case study. **Journal of molecular biology**, v. 270, n. 3, p. 520–34, 1997.
- LÖNNERDAL, B.; IYER, S. Lactoferrin: Molecular structure and biological function. p. 93–110, 1995.
- LUND, M.; ÅKESSON, T.; JÖNSSON, B. Enhanced protein adsorption due to charge regulation. **Langmuir**, v. 21, n. 18, p. 8385–8388, 2005.
- LUND, M.; JÖNSSON, B. A mesoscopic model for protein-protein interactions in solution. **Biophysical journal**, v. 85, n. 5, p. 2940–2947, 2003.
- LUND, M.; JÖNSSON, B. On the Charge Regulation of Proteins. **Biochemistry**, v. 44, n. 15, p. 5722–5727, 2005.
- LUND, M.; JÖNSSON, B. Charge regulation in biomolecular solution. **Quarterly reviews of biophysics**, v. 46, n. 3, p. 265–81, 2013.
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? A proposed definition and overview of the field. **Methods of information in medicine**, v. 40, n. 4, p. 346–58, 2001.
- MACKENZIE, J. Physical Properties of Blue-Green Virus and. **Vacuum**, v. 504, p. 497–504, 1972.
- MACKERELL, A. D. Protein Force Fields. In: **Encyclopedia of Computational Chemistry**. Chichester, UK: John Wiley & Sons, Ltd, 2002. p. 2191–2200.
- MARCUS, R. A.; MARCUS, R. A. Calculation of thermodynamic properties of polyelectrolytes. **J. Chem. Phys.**, v. 23, p. 1057, 1955.
- MARQUART, M. et al. The geometry of the reactive site and of the peptide groups in trypsin,

trypsinogen and its complexes with inhibitors. **Acta Crystallographica Section B Structural Science**, v. 39, n. 4, p. 480–490, 1 ago. 1983.

MCKENZIE, H. A.; WHITE, F. H. Lysozyme and alpha-lactalbumin: structure, function, and interrelationships. **Advances in protein chemistry**, v. 41, p. 173–315, 1991.

MCSWEENEY, P. L. H.; FOX, P. F. **Advanced Dairy Chemistry**. Boston, MA: Springer US, 2013.

MEDDA, L. et al. Hofmeister challenges: Ion binding and charge of the BSA protein as explicit examples. **Langmuir**, v. 28, n. 47, p. 16355–16363, 2012.

METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087–1092, 1953.

METROPOLIS, N.; ULAM, S. The Monte Carlo method. **Journal of the American Statistical Association**, v. 44, n. 247, p. 335–341, 1949.

MIKI, Y.; KAKUYAMA, K.; SODA, K. Protein stability; optimization of electrostatic contributions by partially neutralizing surface ionic charges. **Bio Systems**, v. 44, n. 1, p. 69–77, 1997.

MOORE, S. A et al. Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. **Journal of molecular biology**, v. 274, n. 2, p. 222–236, 1997.

MORAWETZ, H.; HUGHES, W. L. The Interaction of Proteins with Synthetic Polyelectrolytes. I. Complexing of Bovine Serum Albumin. **The Journal of Physical Chemistry**, v. 56, n. 1, p. 64–69, 1952.

MUDGAL, P. et al. Effect of disulfide interactions and hydrolysis on the thermal aggregation of β -lactoglobulin. **Journal of agricultural and food chemistry**, v. 59, n. 5, p. 1491–7, 9 mar. 2011.

MÜLLER, D. Antibody fusions with immunomodulatory proteins for cancer therapy. **Pharmacology & Therapeutics**, v. 154, p. 57–66, 2015.

MURRAY-RUST, P. Bioinformatics and drug discovery. **Current Opinion in Biotechnology**, v. 5, n. 6, p. 648–653, dez. 1994.

MYERS, J. K.; PACE, C. N. Hydrogen bonding stabilizes globular proteins. **Biophysical Journal**, v. 71, n. 4, p. 2033–2039, out. 1996.

NELSON, D. L.; COX, M. M. **Lehninger princípios de bioquímica**. 3. Ed. ed. São Paulo: Sarvier, 2002.

- NIGEN, M. et al. Molecular interaction between apo or holo α -lactalbumin and lysozyme: Formation of heterodimers as assessed by fluorescence measurements. **Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics**, v. 1794, n. 4, p. 709–715, abr. 2009.
- NIGEN, M.; CROGUENNEC, T.; BOUHALLAB, S. Formation and stability of α -lactalbumin-lysozyme spherical particles: Involvement of electrostatic forces. **Food Hydrocolloids**, v. 23, n. 2, p. 510–518, 2009.
- NISHIMURA, C.; UVERSKY, V. N.; FINK, A. L. Effect of salts on the stability and folding of staphylococcal nuclease. **Biochemistry**, v. 40, n. 7, p. 2113–28, 2001.
- NORRBY, P.-O.; BRANDT, P. Deriving force field parameters for coordination complexes. **Coordination Chemistry Reviews**, v. 212, n. 1, p. 79–109, 2001.
- NOZAKI, Y.; TANFORD, C. Examination of titration behavior. **Methods in Enzymology**, v. 11, n. C, p. 715–734, 1967.
- O'DRISCOLL, A.; DAUGELAITE, J.; SLEATOR, R. D. “Big data”, Hadoop and cloud computing in genomics. **Journal of Biomedical Informatics**, v. 46, n. 5, p. 774–781, 2013.
- OMORPHOS, N. P.; KAHN, L.; KALASKAR, D. M. Design of extracellular protein based particles for intra and extra-cellular targeting. **Colloids and Surfaces B: Biointerfaces**, v. 136, p. 440–448, 2015.
- OZDAL, T.; CAPANOGLU, E.; ALTAY, F. A review on protein–phenolic interactions and associated changes. **Food Research International**, v. 51, n. 2, p. 954–970, maio 2013.
- PACE, C. N.; GRIMSLEY, G. R.; SCHOLTZ, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. **The Journal of biological chemistry**, v. 284, n. 20, p. 13285–9, 2009.
- PARODI, P. W. A role for milk proteins and their peptides in cancer prevention. **Current pharmaceutical design**, v. 13, n. 8, p. 813–28, 2007.
- PEDERSEN, J. B. et al. Towards control of aggregational behaviour of α -Lactalbumin at acidic pH. **Journal of Fluorescence**, v. 16, n. 4, p. 611–621, 2006.
- PENFOLD, R.; WARWICKER, J.; JÖNSSON, B. Electrostatic Models for Calcium Binding Proteins. **The Journal of Physical Chemistry B**, v. 102, n. 43, p. 8599–8610, out. 1998.
- PERMYAKOV, S. E. et al. No Need To Be HAMLET or BAMLET To Interact with Histones: Binding of Monomeric α -Lactalbumin to Histones and Basic Poly-Amino Acids. **Biochemistry**, v. 43, n. 19, p. 5575–5582, maio 2004.

- PERRIN, E. et al. Genomes analysis and bacteria identification: The use of overlapping genes as molecular markers. **Journal of Microbiological Methods**, v. 117, p. 108–112, 2015.
- PERSSON, B. A. et al. Molecular evidence of stereo-specific lactoferrin dimers in solution. **Biophysical Chemistry**, v. 151, n. 3, p. 187–189, 2010.
- PERUTZ, M. Electrostatic effects in proteins. **Science**, v. 201, n. 4362, p. 1187–1191, 29 set. 1978.
- PERUTZ, M. F. Regulation of Oxygen Affinity of Hemoglobin: Influence of Structure of the Globin on the Heme Iron. **Annual Review of Biochemistry**, v. 48, n. 1, p. 327–386, jun. 1979.
- PIHLANTO-LEPPÄLÄ, A. Bioactive peptides derived from bovine whey proteins. **Trends in Food Science & Technology**, v. 11, n. 9–10, p. 347–356, set. 2000.
- PONDER, J. W.; CASE, D. A. Force fields for protein simulations. **Advances in protein chemistry**, v. 66, p. 27–85, 2003.
- PRASITNOK, K.; WILSON, M. R. A coarse-grained model for polyethylene glycol in bulk water and at a water/air interface. **Physical Chemistry Chemical Physics**, v. 15, n. 40, p. 17093, 2013.
- PUYOL, P. et al. Interaction of Bovine .BETA.-Lactoglobulin and Other Bovine and Human Whey Proteins with Retinol and Fatty Acids. **Agricultural and Biological Chemistry**, v. 55, n. 10, p. 2515–2520, 1991.
- RAEMY, A.; LAMBELET, P.; ROUSSET, P. Calorimetric information about food and food constituents. In: **The Nature of Biological Systems as Revealed by Thermal Methods**. Dordrecht: Kluwer Academic Publishers, 2005. p. 69–98.
- REBOUILLAT, S.; ORTEGA-REQUENA, S. Potential Applications of Milk Fractions and Valorization of Dairy By-Products: A Review of the State-of-the-Art Available Data, Outlining the Innovation Potential from a Bigger Data Standpoint. **Journal of Biomaterials and Nanobiotechnology**, v. 6, n. 3, p. 176–203, 2015.
- ROCA, M.; MESSER, B.; WARSHEL, A. Electrostatic contributions to protein stability and folding energy. **FEBS Letters**, v. 581, n. 10, p. 2065–2071, maio 2007.
- ROTH, C. M.; NEAL, B. L.; LENHOFF, A. M. Van der Waals interactions involving proteins. **Biophysical journal**, v. 70, n. 2, p. 977–87, 1996.
- RUTHERFURD-MARKWICK, K. J.; GILL, H. S. Immunomodulating activity of protein

- concentrates derived from bovine milk whey in mice. **Nutrition Research**, v. 25, n. 2, p. 157–166, 2005.
- RYAN, D.; MATTHEWS, J. Protein–protein interactions in human disease. **Current Opinion in Structural Biology**, v. 15, n. 4, p. 441–446, 2005.
- RYAN, K. N. et al. Stability and mechanism of whey protein soluble aggregates thermally treated with salts. **Food Hydrocolloids**, v. 27, n. 2, p. 411–420, 2012.
- RYAN, K. N.; FOEGEDING, E. A. Formation of soluble whey protein aggregates and their stability in beverages. **Food Hydrocolloids**, v. 43, p. 265–274, 2015.
- RYDEL, T. J. et al. Refined structure of the Hirudin-thrombin complex. **Journal of Molecular Biology**, v. 221, n. 2, p. 583–601, set. 1991.
- SAĞLAM, D. et al. Design, properties, and applications of protein micro- and nanoparticles. **Current Opinion in Colloid & Interface Science**, v. 19, n. 5, p. 428–437, 2014.
- SAKALLI, I.; KNAPP, E.-W. pK(A) in proteins solving the Poisson-Boltzmann equation with finite elements. **Journal of computational chemistry**, v. 36, n. 29, p. 2147–57, 5 nov. 2015.
- SALWINSKI, L.; EISENBERG, D. Computational methods of analysis of protein-protein interactions. **Curr Opin Struct Biol**, v. 13, p. 377–382, 2003.
- SCHRÖDINGER, L. **The PyMOL Molecular Graphics System, Version 1.1r1**. [s.l: s.n.]. Disponível em: <<https://www.pymol.org/>>.
- SCHUTZ, C. N.; WARSHEL, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? **Proteins**, v. 44, n. 4, p. 400–17, 1 set. 2001.
- SECCHI, G. Role of protein in cosmetics. **Clinics in Dermatology**, v. 26, n. 4, p. 321–325, jul. 2008.
- SEYREK, E.; HATTORI, T.; DUBIN, P. L. Frontal Analysis Continuous Capillary Electrophoresis for Protein–Polyelectrolyte Binding Studies. In: **Capillary Electrophoresis of Proteins and Peptides**. New Jersey: Humana Press, 2004. p. 217–228.
- SHAM, Y. Y.; CHU, Z. T.; WARSHEL, A. Consistent Calculations of p K a 's of Ionizable Residues in Proteins: Semi-microscopic and Microscopic Approaches. **The Journal of Physical Chemistry B**, v. 101, n. 22, p. 4458–4472, maio 1997.
- SHEINERMAN, F. Electrostatic aspects of protein–protein interactions. **Current Opinion in Structural Biology**, v. 10, n. 2, p. 153–159, 2000.

- SHIRAI, H. et al. Antibody informatics for drug discovery. **Biochimica et biophysica acta**, v. 1844, n. 11, p. 2002–2015, 2014.
- SIMONSON, T.; BROOKS, C. L. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics. **Journal of the American Chemical Society**, v. 118, n. 35, p. 8452–8458, jan. 1996.
- SIMONSON, T.; PERAHIA, D. Microscopic Dielectric Properties of Cytochrome c from Molecular Dynamics Simulations in Aqueous Solution. **Journal of the American Chemical Society**, v. 117, n. 30, p. 7987–8000, ago. 1995.
- SINGH, S. N. et al. Dipole-dipole interaction in antibody solutions: correlation with viscosity behavior at high concentration. **Pharmaceutical research**, v. 31, n. 9, p. 2549–58, set. 2014.
- SMITHERS, G. W. et al. New opportunities from the isolation and utilization of whey proteins. **Journal of dairy science**, v. 79, n. 8, p. 1454–1459, 1996.
- SMITHERS, G. W. Whey and whey proteins-From “gutter-to-gold”. **International Dairy Journal**, v. 18, n. 7, p. 695–704, 2008.
- SØRENSEN, S. P.; LINDERSTRØM-LANG, K.; LUND, E. THE INFLUENCE OF SALTS UPON THE IONISATION OF EGG ALBUMIN. **The Journal of general physiology**, v. 8, n. 6, p. 543–99, 7 mar. 1927.
- STEIJNS, J. M.; VAN HOOIJDONK, A. C. Occurrence, structure, biochemical properties and technological characteristics of lactoferrin. **The British journal of nutrition**, v. 84 Suppl 1, n. August, p. S11–S17, 2000.
- STERPONE, F.; MELCHIONNA, S. Thermophilic proteins: insight and perspective from in silico experiments. **Chem. Soc. Rev.**, v. 41, n. 5, p. 1665–1676, 2012.
- SU, R. S. C.; KIM, Y.; LIU, J. C. Resilin: Protein-based elastomeric biomaterials. **Acta Biomaterialia**, v. 10, n. 4, p. 1601–1611, 2014.
- SVENSSON, B. et al. Binding of Ca²⁺ to calmodulin and its tryptic fragments: theory and experiment. **Biochemistry**, v. 32, n. 11, p. 2828–34, 1993.
- SYRBE, A.; BAUER, W. J.; KLOSTERMEYER, H. Polymer Science Concepts in Dairy Systems—an Overview of Milk Protein and Food Hydrocolloid Interaction. **International Dairy Journal**, v. 8, n. 3, p. 179–193, mar. 1998.
- TAKAHASHI, T. Significant role of electrostatic interactions for stabilization of protein assemblies. **Advances in Biophysics**, v. 34, p. 41–54, 1997.

- TAKASHIMA, S. Computation of the dipole moment of protein molecules using protein databases. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v. 148, n. 1–2, p. 95–106, mar. 1999.
- TAKAYAMA, Y. **Lactoferrin and its Role in Wound Healing**. Dordrecht: Springer Netherlands, 2012.
- TANFORD, C. Theory of Protein Titration Curves. II. Calculations for Simple Models at Low Ionic Strength. **Journal of the American Chemical Society**, v. 79, n. 20, p. 5340–5347, out. 1957.
- TANFORD, C.; KIRKWOOD, J. G. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. **Journal of the American Chemical Society**, v. 79, n. 20, p. 5333–5339, out. 1957.
- TCHOLAKOVA, S. et al. Coalescence stability of emulsions containing globular milk proteins. **Advances in Colloid and Interface Science**, v. 123–126, n. SPEC. ISS., p. 259–293, 2006.
- TEIXEIRA, A. A. R.; LUND, M.; DA SILVA, F. L. B. Fast proton titration scheme for multiscale modeling of protein solutions. **Journal of Chemical Theory and Computation**, v. 6, n. 10, p. 3259–3266, 2010.
- TEMPLIN, M. F. et al. Protein microarray technology. **Drug Discovery Today**, v. 7, n. 15, p. 815–822, 2002.
- TENOVUO, J. Clinical applications of antimicrobial host proteins lactoperoxidase, lysozyme and lactoferrin in xerostomia: efficacy and safety. **Oral diseases**, v. 8, n. 1, p. 23–29, 2002.
- TEPLENIN, A. et al. Functional analysis of the engineered cardiac tissue grown on recombinant spidroin fiber meshes. **PLoS ONE**, v. 10, n. 3, p. 1–14, 2015.
- THIRUMALAI, D. et al. Theoretical Perspectives on Protein Folding. **Annual Review of Biophysics**, v. 39, n. 1, p. 159–183, abr. 2010.
- THOMPSON, A.; BOLAND, M.; SINGH, H. **Milk Proteins From Expression to Food**. Burlington: Elsevier, 2009.
- TOMITA, M. et al. Twenty-five years of research on bovine lactoferrin applications. **Biochimie**, v. 91, n. 1, p. 52–57, 2009.
- TOPUZOĞULLARI, M. et al. Molecular-weight distribution and structural transformation in water-soluble complexes of poly(acrylic acid) and bovine serum albumin. **European**

Polymer Journal, v. 43, n. 7, p. 2935–2946, 2007.

TOUHAMI, A.; DUTCHER, J. R. pH-induced changes in adsorbed β -lactoglobulin molecules measured using atomic force microscopy. **Soft Matter**, v. 5, n. 1, p. 220–227, 2009.

TOZZINI, V. Coarse-grained models for proteins. **Current Opinion in Structural Biology**, v. 15, n. 2, p. 144–150, 2005.

TRAMONTANO, A. The role of molecular modelling in biomedical research. **FEBS Letters**, v. 580, n. 12, p. 2928–2934, 2006.

UVERSKY, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. **Protein science : a publication of the Protein Society**, v. 22, n. 6, p. 693–724, jun. 2013.

UVERSKY, V. N.; DUNKER, A. K. Understanding protein non-folding. **Biochimica et biophysica acta**, v. 1804, n. 6, p. 1231–64, jun. 2010.

VALETTE, P.; THOMAS, M.; DÉJARDIN, P. Adsorption of low molecular weight proteins to hemodialysis membranes: Experimental results and simulations. **Biomaterials**, v. 20, n. 17, p. 1621–1634, 1999.

VAN DER LEE, R. et al. Classification of intrinsically disordered regions and proteins. **Chemical Reviews**, v. 114, n. 13, p. 6589–6631, 2014.

VARDHANABHUTI, B. et al. Interactions between β -lactoglobulin and dextran sulfate at near neutral pH and their effect on thermal stability. **Food Hydrocolloids**, v. 23, n. 6, p. 1511–1520, ago. 2009.

VARMA, S.; JAKOBSSON, E. Ionization states of residues in OmpF and mutants: effects of dielectric constant and interactions between residues. **Biophysical journal**, v. 86, n. 2, p. 690–704, 2004.

VICATOS, S.; ROCA, M.; WARSHEL, A. Effective approach for calculations of absolute stability of proteins using focused dielectric constants. **Proteins: Structure, Function, and Bioinformatics**, v. 77, n. 3, p. 670–684, 15 nov. 2009.

VOJCIC, L. et al. Advances in protease engineering for laundry detergents. **New biotechnology**, v. 32, n. 6, p. 629–34, 2015.

WADA, A.; NAKAMURA, H. Nature of the charge distribution in proteins. **Nature**, v. 293, n. 5835, p. 757–8, 1981.

WAGONER, T. B.; FOEGEDING, E. A. Whey protein–pectin soluble complexes for

- beverage applications. **Food Hydrocolloids**, v. 63, p. 130–138, fev. 2017.
- WAKABAYASHI, H. et al. Lactoferrin for prevention of common viral infections. **Journal of Infection and Chemotherapy**, v. 20, n. 11, p. 666–671, 2014.
- WALLACE, R. W. PDB funding. **Drug Discovery Today**, v. 1, n. 8, p. 315–315, 1996.
- WANG, G.; DUNBRACK, R. L. PISCES: a protein sequence culling server. **Bioinformatics (Oxford, England)**, v. 19, n. 12, p. 1589–91, 2003.
- WARSHEL, A. Electrostatic basis of structure-function correlation in proteins. **Accounts of Chemical Research**, v. 14, n. 9, p. 284–290, set. 1981.
- WARSHEL, A.; AQVIST, J. Electrostatic energy and macromolecular function. **Annual review of biophysics and biophysical chemistry**, v. 20, p. 267–98, 1991.
- WARWICKER, J. Simplified methods for pKa and acid pH-dependent stability estimation in proteins: removing dielectric and counterion boundaries. **Protein science: a publication of the Protein Society**, v. 8, n. 2, p. 418–25, 1999.
- WATSON, R. R.; ZIBADI, S.; PREEDY, V. R. **Dietary components and immune function**. New York: Humana Press, 2010.
- WEHRL, A. The many facets of entropy. **Reports on Mathematical Physics**, v. 30, n. 1, p. 119–129, ago. 1991.
- WHITTEN, S. T.; GARCÍA-MORENO E, B. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. **Biochemistry**, v. 39, n. 46, p. 14292–304, 21 nov. 2000.
- WIJAYANTI, H. B.; BANSAL, N.; DEETH, H. C. Stability of Whey Proteins during Thermal Processing: A Review. **Comprehensive Reviews in Food Science and Food Safety**, v. 13, n. 6, p. 1235–1251, 2014.
- WOLZ, M.; MERSCH, E.; KULOZIK, U. Thermal aggregation of whey proteins under shear stress. **Food Hydrocolloids**, v. 56, p. 396–404, maio 2016.
- YADA, R. Y. **Proteins in Food Processing**. 1 Edition ed. Cambridge: Elsevier, 2004.
- YAN, Y. et al. Heteroprotein complex coacervation: bovine β -lactoglobulin and lactoferrin. **Langmuir: the ACS journal of surfaces and colloids**, v. 29, n. 50, p. 15614–23, 17 dez. 2013.
- ZACHARIAS, M. **Protein-protein Complexes: Analysis, Modeling and Drug Design**. [s.l.]

Imperial College Press, 2010.

ZEILER, R. N. W.; BOLHUIS, P. G. Exposure of thiol groups in the heat-induced denaturation of β -lactoglobulin. **Molecular Simulation**, v. 41, n. 10–12, p. 1006–1014, 13 ago. 2015.

ZIELENKIEWICZ, P.; RABCZENKO, A. Methods of molecular modelling of protein-protein interactions. **Biophysical chemistry**, v. 29, n. 3, p. 219–224, 1988.

APÊNDICE A – Disciplinas cursadas

Janus - Sistema Administrativo da Pós-Graduação



Universidade de São Paulo
Interunidades em Bioinformática
Documento sem validade oficial

FICHA DO ALUNO

95131 - 4887035/1 - Lariani Aparecida Delboni

Sigla	Nome da Disciplina	Início	Término	Carga Horária	Cred.	Freq.	Conc.	Exc.	Situação
IBI5026-1/3	Biofísico-Química Computacional	04/08/2014	07/11/2014	120	8	100	B	N	Concluída
6015015-1/1	Forças Moleculares (Faculdade de Ciências Farmacêuticas de Ribeirão Preto - Universidade de São Paulo)	13/08/2014	26/11/2014	90	6	93	A	N	Concluída
6045850-2/1	Introdução à Proteômica (Faculdade de Ciências Farmacêuticas de Ribeirão Preto - Universidade de São Paulo)	02/03/2015	08/03/2015	30	2	100	A	N	Concluída
RBQ5752-4/2	Estrutura e Função de Biomoléculas (Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo)	02/03/2015	28/04/2015	90	0	-	-	N	Pré-matrícula indeferida
RBP5767-4/2	Seminários em Biologia Celular e Molecular II (Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo)	05/03/2015	18/06/2015	30	2	75	A	N	Concluída
5935938-1/3	Tópicos em Química Computacional (Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - Universidade de São Paulo)	10/03/2015	23/06/2015	90	6	100	A	N	Concluída
5935915-2/1	Cinética: Técnicas e Abordagens Aplicadas à Enzimologia (Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - Universidade de São Paulo)	01/04/2015	03/06/2015	45	3	88	B	N	Concluída
Atividade do Programa	Publicação do trabalho "on the complexation of whey proteins", na revista Food Hydrocolloids, páginas 88-99; volume 55, Land - Suécia. (2)	27/07/2015	04/11/2015	-	4	100	-	-	-
5935889-7/2	Físico-Química Avançada II (Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto - Universidade de São Paulo)	11/08/2015	24/11/2015	120	8	100	A	N	Concluída
6045845-3/1	Tópicos Especiais em Ciências Farmacêuticas (Faculdade de Ciências Farmacêuticas de Ribeirão Preto - Universidade de São Paulo)	05/09/2015	11/09/2015	15	1	100	A	N	Concluída
RBQ5778-1/3	Estrutura e Função de Proteínas (Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo)	21/10/2015	07/12/2015	60	4	100	A	N	Concluída

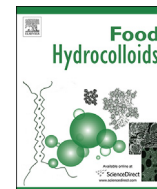
	Créditos mínimos exigidos		Créditos obtidos
	Para exame de qualificação	Para depósito da dissertação	
Disciplinas:	40	40	44
Estágios:			
Total:	40	40	44

Créditos Atribuídos à Dissertação: 56

APÊNDICE B – Artigo já publicado

Referência

DELBONI, L. A.; DA SILVA, F. L. B. On the complexation of whey proteins. **Food Hydrocolloids**, v. 55, p. 89–99, 2016.



On the complexation of whey proteins



Lariani Aparecida Delboni^{a, b}, Fernando Luís Barroso da Silva^{a, b, *}

^a Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do café, s/no. – Universidade de São Paulo, 14040-903, Ribeirão Preto, SP, Brazil

^b Department of Theoretical Chemistry – Lund University, POB 124 – S–221 00, Lund, Sweden

ARTICLE INFO

Article history:

Received 27 July 2015

Received in revised form

31 October 2015

Accepted 4 November 2015

Available online 14 November 2015

Keywords:

Milk proteins

Electrostatic interactions

pH

Ionic strength

Molecular modeling

Heteroprotein association

ABSTRACT

Milk proteins have a rich diversity of physical chemistry and biodegradable properties which makes them appealing for different food and pharmaceutical applications. Theoretical coarse grained models and numerical simulations were employed here in order to gain novel insight into the understanding of the fundamental mechanisms of the process of milk proteins complexation in a diversity of environmental conditions. The interactions between α -lactalbumin, β -lactoglobulin and lactoferrin were investigated by means of Monte Carlo simulations. The comparison between the free energies associated with the complexation of α -lactalbumin–lactoferrin and β -lactoglobulin–lactoferrin at different pH and ionic strengths let us to explain why is experimentally observed the later complex and not the α -lactalbumin–lactoferrin complex.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Since probably the domestication of cattle and other animals after the eighth millennium BC, milk has gained a vital role on men survival and development (Evershed et al., 2008). Milk uses have a great impact in our present daily lives and important applications in various areas (Bösze, 2008; McSweeney and Fox, 2013; Thompson, Boland, & Singh, 2009; Watson, Zibadi, & Preedy, 2010), including odontology (Tenovuo, 2002), medicine (Permyakov, Pershikova, Khokhlova, Uversky, & Permyakov, 2004; Steijns & van Hooijdonk, 2000; Tomita et al., 2009), biodefense (Clare, Catignani, & Swaisgood, 2003), food (Aly, Ros, & Frontela, 2013; Dalgleish, 1992; Dickinson, 2008; Dickinson, 2013; Donald, 2000; Kinsella & Whitehead, 1989; de Roos, Walstra, & Geurts, 1998; Steijns & van Hooijdonk, 2000; Syrbe, Bauer, & Klostermeyer, 1998; Vardhanabhuti, Yucel, Coupland, & Foegeding, 2009), cosmetics (Secchi, 2008; Steijns & van Hooijdonk, 2000), pharmaceuticals (Dickinson, 2008), and others industries (Audica, Chaufer, & Daufin, 2003). There also have been considerable increase on its importance specially due to the

interest in developing biodegradable materials. Some of these applications are also essential to many academic areas.

In addition to water, lipids and sugar, milk contains several proteins with a large diversity of physical chemistry properties which let them be used in a quite wide number of applications ranging from nutritional to functional and biological functions (Thompson et al., 2009). From all milk constituents, these proteins are particularly interesting for industrial applications. For instance, they are highly effective emulsion stabilizing agents and gelling ingredients (Dalgleish, 1992; Dickinson, 2008, 2013; Havea, Singh, & Creamer, 2001; Kinsella & Whitehead, 1989; Ryan et al., 2012; Tcholakova, Denkov, Ivanov, & Campbell, 2006) responsible to texturing a variety of foods (Batt, Brady, & Sawyer, 1994; Donald, 2000; Girard & Schaffer-Lequart, 2007), used in sports drinks to retain clarity and improve nutritional functions in beverages (Ryan & Foegeding, 2015; Vardhanabhuti et al., 2009), effective microencapsulating agents in food (Egan, Jacquier, Rosenberg, & Rosenberg, 2013; Egan, O'Riordan, O'Sullivan, & Jacquier, 2014; Eratte, Wang, Dowling, Barrow, & Adhikari, 2014), used to the design of infant formula with similar functional effects than human milk (Aly et al., 2013), as models in medicine for *in-vitro* studies on human tear film deposits for the development of optical lens (Luensmann & Jones, 2008; Soltys-Robitaille, Ammon, Valint, & Grobe, 2001), and large-scale protein purification on biotechnology (Wang, Gao, & Dubin, 1996).

* Corresponding author. Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do café, s/no. – Universidade de São Paulo, 14040-903, Ribeirão Preto, SP, Brazil.

E-mail address: fernando@fcfp.usp.br (F.L. Barroso da Silva).

Together with caseins, bovine whey (or milk serum) proteins [β -lactoglobulin (β -LG), α -Lactalbumin (α -LA), immunoglobulin (IgG), serum albumin (SA), lactoferrin (LF), lactoperoxidase (LP) and lysozyme (LSZ)] (Kinsella & Whitehead, 1989; Korhonen, 1998) are of considerable importance on dairy products (McSweeney and Fox, 2013; Smithers, 2008; Watson et al., 2010). The exploitation of the properties of the whey proteins may contribute to solve some technological shortcomings that makes difficult their widespread commercial applications (Smithers et al., 1996), despite the fact that there is a need to satisfy the market demand (Kinsella & Whitehead, 1989). In fact, the food marketplace requires whey proteins to be continuously manipulated in order to improve their functionality (Clare & Daubert, 2011). As already pointed out by other authors, the research on the improvement and manipulation of milk proteins, milk protein–polysaccharide and milk protein–protein interactions may have an effect on economy and lead to the development of intelligent nanoscale encapsulation systems for food and pharmaceutical applications (Batt et al., 1994; Dickinson, 2008).

The major whey proteins are β -LG, α -LA, SA and LF. From these proteins, LF seems to be the most versatile one both in terms of its physical chemistry properties and applications. Just to name a few, LF might be involved in several physiological functions as such as inflammation, bacteriostasis and bactericidal effects, autoimmune diseases, growth, and absorption of iron from breast milk (El-Loly & Mahfouz, 2011; Lönnerdal & Iyer, 1995; Takayama, 2012; Wakabayashi, Oda, Yamauchi, & Abe, 2014). Particularly in the food technology, heteroprotein coacervates of β -LG and LF are of interest in new food products (Anema & Kruif, 2014; Yan et al., 2013). Perhaps one of the most exciting features of LF is its ability to easily form complexes, as noticed by Hekman already in 1971 (Hekman, 1971).

A closer look at these proteins interactions often reveals some rather unexpected behavior. Despite the fact that electrostatic interaction is one of the most important driven force for protein complexation (Jönsson, Lund, & da Silva, 2007; Takahashi, 1997) and that both α -LA and β -LG acquire positive and negative charges at virtually the same pH, the interaction of these molecules with other proteins shows some peculiarities. Different reports have demonstrated that the complexation with β -LG occurs but not with α -LA (Puyol, Perez, Ena, & Calvo, 1991). Using gel filtration, Puyol and co-authors have analysed and compared the interaction of α -LA and β -LG with palmitic acid. They found that while β -LG binds to the fatty acid, α -LA does not (Puyol et al., 1991). Hekman has also raised some questions not conclusively solved about the ability of LF to easily forms complexes and what forces play a role in it (Hekman, 1971). In spite of the fact that LF does form hetero-complexes with β -LG, it is not clear if LF may associate with α -LA (Lampreave et al., 1990) which is an apparent contradiction with LF binding abilities. Recent studies show the complexation of β -LG and LF not only for the 1:1 stoichiometry but also observed the formation of clusters suggested as β -LG₂–LF– β -LG₂ (Kizilay et al., 2014; Yan et al., 2013). These reports also discussed the physical chemistry differences between the heteroprotein association and the self-aggregation processes.

Although whey proteins and whey protein complexes have been extensively reported by means of several experimental techniques (e.g. gel filtration (Puyol et al., 1991), calorimetry (Ju, Hettiarachchy, & Kilara, 1999; Raemy, Lambelet, & Rousset, 2005), spectrophotometry (Morawetz & Hughes, 1952), transmission electron microscopy (Mudgal, Daubert, Clare, & Foegeding, 2011), scanning electron microscopy (Egan et al., 2013), fluorescence (Nigen, Tilly, Croguennec, Drouinkucma, & Bouhallab, 2009b), laser scanning microscopy (Kizilay et al., 2014), electrophoresis (Hattori, Bataldar, Kato, Bohidar, & Dubin, 2005; Havea et al., 2001; Mudgal et al.,

2011; Seyrek, Hattori, & Dubin, 2004), rheology (Kizilay et al., 2014), light scattering (Li, Mattison, Dubin, Havel, & Edwards, 1996; Yan et al., 2013), small angle neutron scattering (Kizilay et al., 2014), turbidity and microscopic observations (Anema & Kruif, 2014; Nigen, Croguennec, & Bouhallab, 2009a; Ryan et al., 2012; Ryan & Foegeding, 2015; Yan et al., 2013), high performance liquid chromatography (Egan et al., 2014; Topuzogullari, Cimen, Mustafaeva, & Mustafaev, 2007), size exclusion chromatography (Ryan & Foegeding, 2015; Yan et al., 2013), ζ -potential (Anema & Kruif, 2014; Ryan et al., 2012; Ryan & Foegeding, 2015), atomic force microscopy (Touhami & Dutcher, 2009), potentiometric and turbidimetric titration and photon correlation spectroscopy (Laos, Brownsey, & Ring, 2007)), much less is known about milk protein–protein interactions at the molecular level, where is the origin of any functional characteristics of a protein. In fact, there is a well supported need for a better understanding of their molecular interactions (Ryan & Foegeding, 2015; Wijayanti, Bansal, & Deeth, 2014; de Wit, 2009; Zeiler & Bolhuis, 2015).

Theoretical coarse grained models and numerical simulations offer an opportunity to gain novel insight into the understanding of the fundamental mechanisms of the process of milk proteins complexation in a diversity of environmental conditions. As such, these techniques are a valuable tool for the nascent soft matter approach of the food science offering a rational approach to describe, explain and control complexation mechanisms that affect the functions. Moreover, due to the fact that these proteins have also important physiological functions, these tools are as well of relevance to the Biological sciences.

In this article, we illustrate the applicability of such molecular modeling approaches on food science and technology addressing the complexation issues of α -LA and β -LG interactions with itself and with LF. Doing that we demonstrate the physical reasons for the more favorable interactions of LF with β -LG but not so strong with α -LA (Lampreave et al., 1990). Ionic strength and pH effects are also discussed.

2. Model and methodology

Statistical mechanics provides an appealing theoretical framework in order to explore the main features of the complexation of two macromolecules in an electrolyte solution (Hill, 1986). Particularly interesting here is the fact that this approach provides free energy derivatives as a function of the macromolecules separation at different environmental conditions. Coarse grained models within this framework also offer the possibility to explore the main physical features of a system with a reduced number of parameters. With this in mind, several molecular models within the continuum solvent description have been devised to study protein–protein interactions (Kurut, Persson, Åkesson, Forsman, & Lund, 2012, 2015; Lund & Jönsson, 2003) and solved by Monte Carlo (MC) simulations (Frenkel & Smit, 1996; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953).

Monte Carlo methods are numerical integration tools that have become a common procedure to compute thermodynamic averages in theoretical biochemistry during the last 30 years. Applied to molecular systems, they provide a set of relevant configurations for a given physical model, that can be used to calculate detailed structural and thermodynamical properties. Compared with other popular approaches such as the Poisson-Boltzmann equation, MC simulations have the advantage to *exactly* (within statistical errors) solve the given model, providing a deeper understanding of the system at hand (Allen & Tildesley, 1989; Frenkel & Smit, 1996).

The model used here is based in previous works (Kurut et al., 2012; Lund & Jönsson, 2003; Persson, Lund, Forsman, Chatterton, & Åkesson, 2010). Two macromolecules build up by a collection

of charged Lennard-Jones (LJ) spheres of radii (R_{ai}) and valence Z_{ai} mimicking aminoacids (a mesoscopic approach) are placed in an electroneutral open cylindrical cell of radius r_{cyl} and height l_{cyl} . This is a simplified version of the so-called *cell model* (Jönsson, 1981; Marcus, 1955), conveniently modified for an efficient sampling of the system within a reasonable cpu time. Such binary protein-protein model used here permits to focus on the heteroprotein association without interference of self-aggregation, a common issue on experiments (Anema & Kruif, 2014; Kizilay et al., 2014; Yan et al., 2013).

Proteins are freely to rotate as rigid bodies (with no internal degree of freedom) in any direction, and let to translate back and forward along the axis in which their centers are laying. Counterions and added salt particles are implicitly taken into account via a screening term [$\exp(-\kappa r_{ij})$], where κ is the inverse Debye length, and r_{ij} is the interparticle separation distance]. Solvent enters the calculation through its averaged screening behavior. Instead of explicit water molecules which would result in prohibitive computational costs for this kind of studies, a dielectric continuum model is assumed. This means that water is mimicked by a structureless dielectric medium characterized by a relative dielectric permittivity ϵ . The same uniform static dielectric permittivity ϵ_s is assigned to all charged species interior (de Carvalho, Ghiotto, & da Silva, 2006; da Silva, Jönsson, & Penfold, 2001). A scheme of the model is given in Fig. 1.

Protein coordinates are obtained from the RCSB Protein Data Bank (PDB) (Berman et al., 2000), and edited before the calculations. Six chains are available for the x-ray structure of bovine α -LA (PDB id 1F6S). We have deleted the last five chains of the 1F6S file and selected chain A for our calculations. Since our main measurements are averaged thermodynamical properties as such as free energies, the present calculations were performed with holo α -LA (Ca^{2+} loaded) in order to prevent us from the fact that this protein undergoes large structural changes after the loss of the Ca^{2+} (Pedersen, Fojan, Sorensen, & Petersen, 2006). Moreover, the presence of Ca^{2+} seems to be important for α -LA. For instance, its apo form (Ca^{2+} unloaded) does not complex with lysozyme (Nigen, Tilly et al., 2009b).

Most of the complexation calculations are carried out at pH 6.5 in order to be closer to the natural pH of milk (≈ 6.6) (Girard & Schaffer-Lequart, 2007). At this pH, β -LG is a dimer (Fogolari

et al., 2000; Gottschalk, Nilsson, Roos, & Halle, 2003). Therefore, chains A and B available from the x-ray structure of β -LG (PDB id 1BEB) were used. β -LG can be found in ten genetic variant forms (Thompson et al., 2009), being the most common the β -LG-A dimer (Asp64) and the β -LG-B dimer (Gly64). In the PDB file 1BEB, a mixture of genetic variants A and B is given. The amino acid difference is rather small and the effects in electrostatic interactions is consequently of minor general importance. For example, the predicted titration curves differ by only ca. 0.1 units in agreement with experiments. The LF coordinates (PDB id 1BLF) contain only one chain that was used by us. The two iron ions ($Z_{Fe} = +3$) (Farrell Jr et al., 2004) presented in this coordinate PDB file were included in the complexation calculations. Missing residues were not added on the structures. Fractional charges are assigned to each protein LJ sphere accordingly to the solution pH. During the constant pH simulations, such charges could change due the acid-base equilibrium properties of the aminoacids. A fast proton titration scheme is followed to allow ionizable residues to titrate. We have taken the dissociation constant of the aminoacid model compounds (pK_0) from ref. (Nozaki & Tanford, 1967). Details are given elsewhere (Teixeira, Lund, & da Silva, 2010).

Additional simulation runs with a single protein fixed at the center of the cell in the presence of electrolytes (described within the Debye-Hückel framework) were also carried out at different pH and salt concentrations in order to acquire their main physical chemistry properties. The outcomes are comparable with data from calculations with an atomistic model and explicit mobile ions.

The total system energy function is given by a screened Coulombic potential energy [$u^{el}(r_{ij})$] and a van der Waals term [$u^{vdw}(r_{ij})$]. The $u^{el}(r_{ij})$ between any two aminoacids i and j is given by,

$$u^{el}(r_{ij}) = \frac{z_i z_j e^2}{4 \pi \epsilon_0 \epsilon r_{ij}} \exp(-\kappa r_{ij}) \quad (1)$$

where ϵ_0 is the vacuum permittivity ($\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/\text{Nm}^2$), z_i and z_j denote the valency of i and j , respectively, $e = 1.602 \times 10^{-19} \text{ C}$ is the elementary charge, and r_{ij} their separation. κ , the inverse Debye length, is given by $[8\pi e^2 / (\epsilon_0 \epsilon k_B T) \sum_{ions} c_k z_k^2]^{1/2}$, where $k_B (= 1.3807 \times 10^{-23} \text{ J.mol}^{-1} \text{ K}^{-1})$ is Boltzmann constant, T is the temperature in Kelvin units and c_k is the number density of the mobile electrolyte specie k (counter-ions and added salt).

A LJ potential energy is used to capture the general fundamental physical aspects of both the protein excluded volume (the repulsive $1/r^{12}$ term due the Pauli-exclusion principle), an averaged van der Waals (vdw) interaction and an effective contribution of the hydrophobic effect (these last two composed the attractive $1/r^6$ term). No specific hydrophobic “interaction” to handle hydrophobic patches, salt specific effects (Hofmeister), hydration and hydrogen bonds are captured by such approximation. Only the main driven forces are described. As it is expected for vdw interactions, large aminoacids will attract each other more than smaller ones due the potential size dependency – see Eq. (2). The model parametrization is based on a Hamaker argument, and the parameters used here, assumed to be universal for any protein system, were taken from refs. (Kurut et al., 2012; Persson et al., 2010). For any two aminoacids i and j , this vdw interaction potential energy is given by,

$$u^{vdw}(r_{ij}) = 4\epsilon_{LJ} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2)$$

where ϵ_{LJ} is, following ref. (Kurut et al., 2012), assumed here as equal to $0.05005 k_B T (= 0.124 \text{ kJ/mol})$. This should correspond to a

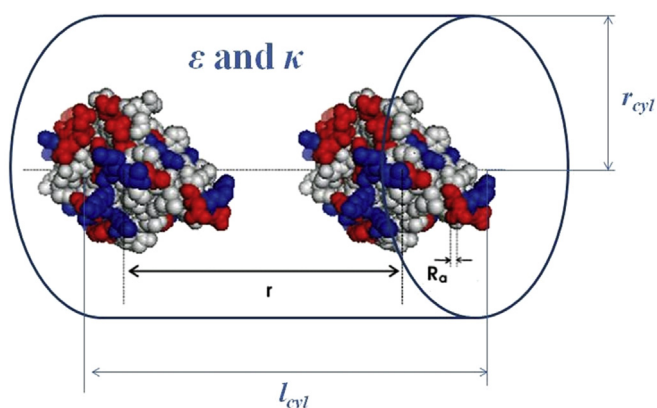


Fig. 1. Schematic representation of the model system. Two proteins build up by a collection of charged LJ spheres of radii (R_{ai}) and valence Z_{ai} mimicking aminoacids are seen surrounded by counterions and added salt particles implicitly described by the inverse Debye length κ . Positive and negatively charged protein aminoacids (radii R_{ai}) are represented in blue and red, respectively. These two proteins center of mass are apart from each other by a r separation at this configuration. The solvent is represented by its static dielectric constant ϵ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Hamaker constant of ca. $9 k_B T$ for aminoacids pairs (Lund & Jönsson, 2003). The term σ_{ij} ($= R_{ai} + R_{aj}$) is the separation distance of two aminoacids i and j at contact. For instance, σ_{ij} for ALA ($R_{aALA} = 3.1 \text{ \AA}$) and ARG ($R_{aARG} = 4.0 \text{ \AA}$) is 7.1 \AA . Aminoacids sizes were taken from ref. (Persson et al., 2010) where the authors assumed that they have a density of 0.9 g/ml and use their molecular weights to determine the proper R_{ai} .

The total system interaction energy for a given configuration $\{U(\{\mathbf{r}_k\})\}$ is

$$U(\{\mathbf{r}_k\}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(u^{ele}(r_{ij}) + u^{vdw}(r_{ij}) \right) \quad (3)$$

where $\{\mathbf{r}_k\}$ are aminoacids positions, $N = N_{p1} + N_{p2}$ is the total number of aminoacids, N_{p1} is the number of aminoacids of protein 1 and N_{p2} is the number of aminoacids of protein 2.

Calculations were performed with the Faunus biomolecular simulation package (Stenqvist, Thuresson, Kurut, Vácha, & Lund, 2013), where the fast titration scheme was already implemented (Teixeira et al., 2010). The cylindrical simulation box was set up with radius $r_{cyl} = 100 \text{ \AA}$ and height $l_{cyl} = 500 \text{ \AA}$ in all calculations. The aqueous solution dielectric constant and temperature were fixed at $\epsilon = 78.7$ and $T = 298 \text{ K}$, respectively. Solution pH was varied from 2 to 13. Calculations were performed at several salt concentrations (c_s): 5, 10, 20, 40 and 150 mM which correspond, respectively, to $1/\kappa$ equals to 43.0653, 21.5326, 15.2259 and 7.86261 \AA . The number of MC steps for production was at least 10^7 steps after equilibration. In order to quantify the vdW contributions to the full interactions, additional simulation runs assigning zero to all protein charges were carried out as well.

The displacement parameters for translation and rotation were adjusted to give ca. 30–50% of acceptance. Due to large free energy barriers ($>5k_B T$) found on some systems with a complex free energy landscape, some samplings were done in multiple simulation runs with constraints. After an initial standard simulation (with no constraints), these large barriers were identified and used to define the critical regions for a proper sampling in separated simulations. Subsequent runs where the protein–protein separation distance was restricted within specific regions were carried out. This protocol provides an alternative to the penalty function method (da Silva, Lund, Jönsson, & Åkesson, 2006).

During the simulations, protein properties as the net valence (Z_p), the capacitance ($C_p \equiv \langle Z^2 \rangle_0 - \langle Z \rangle_0^2$), which is a measurement of the proton fluctuation mechanism (Lund & Jönsson, 2013), and the dipole moment number (μ_p) were measured by the direct computation of the charges at every generated configuration. For solution pH \neq pI (isoelectric point), μ will depend on the coordinate origin. Here, we have used the protein's center of mass. The protein–protein radial distribution function $[g(r)]$ was also sampled with a bin size of 1 \AA . From this function, the angularly averaged potential of mean force $[\beta w(r) = -\ln g(r)]$, where $\beta = 1/k_B T$ was obtained with low noise, and used to estimate the free energy of the complexation at each physical chemistry condition. Statistical noise was reduced by running and averaging multiple independent simulation at the same experimental conditions.

The complexation can be analytically estimated based on a simple Debye–Hückel theory argument for the free energy (A_{DH}), where the protein–protein charges interaction is screened out by $\exp(-\kappa r_{ij})$:

$$A_{DH}(r_{ij}) = \frac{z_i z_j e^2}{4 \pi \epsilon_0 \epsilon r_{ij}} \exp(-\kappa r_{ij}) \quad (4)$$

The values for z_i and z_j are taken from the simulated titration

curves. They could also be obtained from experiments. For the sake of convenience, we assumed a separation distance of 50 \AA (an estimation for the contact distance) for all A_{DH} calculations. Following Egan and co-workers (Egan et al., 2014), the estimated free energy is normalized from 0 to 1 for the purpose of preparing such analytical prediction model for the maximum interaction over the studied pH range. After calculating A_{DH} for all pH values, the maximum observed $A_{DH,max}$ is used to normalize all the others ($A_{DH}/A_{DH,max}$). Only pH values at which an attractive interaction between proteins is observed are used.

3. Results and discussion

3.1. Physical chemistry properties of the proteins

Different driven forces can result in protein complexation [e.g. Coulombic attraction for oppositely charged proteins, dipolar attraction for molecules with high dipole moments, hydrophobic “interactions” for proteins with a large hydrophobic surface, the so-called “sensitive electrostatic interactions” when the protein has a large charge capacitance due to a charge regulation mechanism (Da Silva & Jönsson, 2009; Lund & Jönsson, 2013), etc.]. The knowledge of the protein electrostatic properties such as the protein net charge ($q_p = \sum_{i=1}^{N_p} z_i e$, where N_p is the number of protein aminoacids) and the dipole moment number ($\langle \mu \rangle_0 = \left| \sum_{i=1}^{N_p} z_i \mathbf{r}_i \right|$, where \mathbf{r}_i is aminoacid i position) at different pH values and salt concentrations provides an initial insight about the complexation molecular mechanisms, at least on the pure Coulombic regime. In this case, it is expected an attraction when both proteins have charges with opposite signals. This happens when the solution pH is above one of the protein's pI and below the other's pI. As larger are μ_p and C_p stronger attractions might be seen due dipolar and charge regulation mechanisms, respectively. Following a simple Debye–Hückel theory argument, salt will screen the protein charge–charge interactions by $\exp(-\kappa r_{ij})$. For the other interactions, the screening factor will be more severe [e.g. $\exp(-2\kappa r_{ij})$ for the induced–induced interaction] (da Silva, Boström, & Persson, 2014; Striolo et al., 2002).

These physical chemistry properties of α -LA, β -LG and LF at pI and different ionic strengths are listed in Table 1. This data is from the MC simulations with a single protein in the electrolyte solution. The experimental pI data available in the literature can be rather scattered as shown before (Medda et al., 2012) partly due to direct experimental uncertainties but also due to varying conditions (ionic strengths, protein concentrations, genetic variants, etc.). Different pI values have been reported for bovines α -LA [e.g. 4.2–4.5 (Farrell Jr et al., 2004; Jeng, Bleck, Wheeler, & Jime, 1997; Thompson et al., 2009), 4.6–4.9 (McSweeney and Fox, 2013) 5.0 (Valette, Thomas, & Déjardin, 1999)], β -LG [e.g. 4.6 (Gottschalk et al., 2003), 4.8 (Fogolari et al., 2000), 5.1 – for both genetic variants A and B (Farrell Jr et al., 2004), 5.2 (McSweeney and Fox, 2013; Thompson et al., 2009), 5.1–5.5 (Medda et al., 2012))] and LF [e.g. 8.7 (Takayama, 2012), 8.8 (Farrell Jr et al., 2004), ≈ 9 (Thompson et al., 2009)]. The simulation outcomes (5.0, 4.8 and 9.6 for α -LA, β -LG and LF, respectively) are closer to these measurements.

Table 1 reveals that the pI values for α -LA (5.0) and β -LG (4.8) are similar. On the other hand, β -LG has larger dipole moments (at pI and low ionic strength, $\mu_{\alpha-LA} = 66 < \mu_{\beta-LG} = 131$) and even larger protein charge capacitance (at pI and low ionic strength, $C_{\alpha-LA} = 1.2 < C_{\beta-LG} = 3.3$) which facilitates its attractive interactions with other charged molecules. LF is a highly basic protein (pI = 9.6) which confirms its tendency to Coulombically attract anionic molecules, e.g. α -LA and β -LG when the solution pH is within the 4.8–9.6 pH window. Of course, due to the other possible physical interactions (ion–dipole, the charge regulation mechanism and

Table 1

Main physical chemistry properties of the studied proteins. Charge capacitance (C_p) and dipole moment number (μ_p) at different ionic strengths for the investigated proteins at their isoelectric points (pI). N_p is the number of aminoacids. R_p (given in Å units) is an estimate of the protein radius. μ is in Å units. * Dimer (chain A and B). PDB ids are 1F6S, 1BEB, and 1BLF, respectively, for α -LA, β -LG and LF.

property	α -LA					β -LG*					LF				
	Salt concentrations (mM)														
	5	10	20	40	150	5	10	20	40	150	5	10	20	40	150
C_p	1.2	1.2	1.3	1.4	1.7	3.3	3.6	4.1	4.7	6.3	4.0	4.5	5.2	6.0	8.0
μ_p	66	66	67	67	70	131	133	135	137	143	188	197	212	231	282
pI			5.0					4.8					9.6		
N_p			123					324					689		
R_p			28					42					53		

vdw), this pH window will be even wider. The dipole moment of LF is about 2 times the one observed for β -LG at physiological ionic strength. This protein also has a high C ($C_{LF} = 4$ at $c_s = 5$ mM; $C_{LF} = 8$ at $c_s = 150$ mM). These results suggest that LF is in fact a molecule designed to easily form complexes with other charged objects which is in agreement with experimental observations — see, e.g., Hekman (1971); Kizilay et al. (2014). The pI salt dependence is not observed in the present data due the approximations in the chosen titration numerical scheme where anisotropic protein-salt interactions are neglected (Teixeira et al., 2010). More computationally expensive Monte Carlo simulations carried out with an atomistic protein description and a titration scheme with explicit mobile ions (Kesvatera, Jönsson, Thulin, & Linse, 1999) show shifts of ca. 0.3 pH units from 5 to 150 mM for α -LA and β -LG. In these simulations with a explicit salt model, LF's pI keeps its insensitive characteristic with the increase of ionic strength. For both models, LF's μ is substantially affected by the salt concentration. Here, in the implicit salt model, we can see that at pI LF's μ changes from 188 at $c_s = 5$ mM to 282 (in Å units) at $c_s = 150$ mM. This indicates that an increase in the ionic strength tends also to increase the (attractive) ion-dipole interactions although its screening factor is more severe than the one for the charge-charge term. The combination of these high values of μ_{LF} and C_{LF} is responsible to facilitate the assembly of LF with other molecules.

Plots of the protein net charge as a function of pH at different salt concentrations are shown in Fig. 2. If by one side the pI values were virtually not affected by the addition of salt, as can be seen in this figure, the differences on the protein net charges ($\Delta Z = \langle Z_{salt=150mM} \rangle - \langle Z_{ref=5mM} \rangle$) can be quite large at some pH values. Comparing the system at 5 and 150 mM shows differences as large as 15, 38 and 72, in absolute numbers, for α -LA, β -LG and LF, respectively. It can be observed the same trend of increasing the protein net charge with increasing solution ionic strength as experimentally seen for Hen-egg-white lysozyme (Kuehner et al., 1999). Experiments on bovine serum albumin via potentiometric titration and electrophoretic light scattering also show that the surface charge of this protein increases as the concentration of NaCl electrolyte is raised (Medda et al., 2012).

The protein capacitances varying with pH at different salt concentrations are given in Fig. 3. This variation with pH is a consequence of the number of aminoacid residues that titrate around each pH. The three studied whey proteins show qualitatively a similar behavior for their capacitances, i.e. C has the larger values at both acidic and basic regimes with a minimum around pH 8. In general, LF has the highest capacitances followed by β -LG. The salt effect is more determined when C has higher values, being LF the molecule where the differences on C for the low and moderate salt systems are more considerable. The general observed behavior is that increasing ionic strength raises the protein capacitance particularly at pH values where it has a tendency to peak. This peak is a consequence of a large number of a certain residue with

intrinsic pK_a similar to the solution pH. Charge fluctuations are largest when $pH \cong pK$. A protein rich in aspartic acid (asp) will peak at pH 4 while a protein rich in lysine (lys) will peak at pH 10.4 [$pK_a^{asp} \approx 4.0$ and $pK_a^{lys} \approx 10.4$ (Nozaki & Tanford, 1967)].

A similar plot for the dipole moment number varying with pH at several salt concentrations is given in Fig. 4. Dipole moments can substantially vary with pH for all the studied proteins. α -LA has the smallest μ_p while β -LG has intermedium values and LF the largest numbers. From dipole moments and also the already discussed capacitance, one can conclude even before performing any molecular simulation that LF will form stronger complexes with β -LG due a combination of electrostatic interactions (Coulombic attraction, ion-dipole interaction and charge regulation). The salt effect is more marked at low or high pH conditions. Each protein μ_p responds in a particular manner to the increase in ionic strength.

3.2. Complexation

The dimerization process was investigated by means of free energy derivatives. The calculated free energy of interactions as given by the potentials of mean force [$\beta w(r)$] for the three studied systems (α -LA- β -LG, α -LA-LF and β -LG-LF) at pH 6.5 and physiological ionic strength are given at Fig. 5. At this typical natural pH of milk and critical salt regime (150 mM), all these systems show a clear minimum around 50 Å which can indicate a complexation based on a simple thermodynamical criterion [$\beta w(r) < 0$]. However, the association between α -LA and β -LG showed a potential depth of $-0.6 k_B T$ at the separation distance of 41.6 Å. Being smaller than the thermal energy ($1 k_B T$), this indicates an unstable association. Therefore, we assume that this complex is dissolved at this solution condition. In a similar manner, the strength of the free energy minimum for the α -LA-LF system is of the order of $1 k_B T$ which can explain why this complex was observed in some experiments and not in others (Hekman, 1971; Lampreave et al., 1990; Puyol et al., 1991). Conversely, the interaction between β -LG and LF, as anticipated in the previous section, is more pronounced even at this salt concentration. Such outcome supports the LF(β -LG)₂ clusters proposed by Schmitt and co-authors (Kizilay et al., 2014; Yan et al., 2013).

We next studied the pH and salt effects in these association processes. In Fig. 6a–c, the free energy of interactions are reported for the case where the salt concentration is fixed at 150 mM and pH is varied. It can be seen that for all the three systems, the optimal pH value is closer to the natural pH of milk. It is at pH 6.5 that we observed the minimum free energy values. Weaker attractions can be seen for α -LA- β -LG between pH 4 and 10. Following the criterion that a negative free energy would result in a complex, this would be the pH window where α -LA- β -LG, α -LA-LF and β -LG-LF complexes would be found at physiological salt concentrations. A more realist view as mentioned above would be to assume a minimum of $1 k_B T$ which decreases such pH window to 6.5–8.0 and

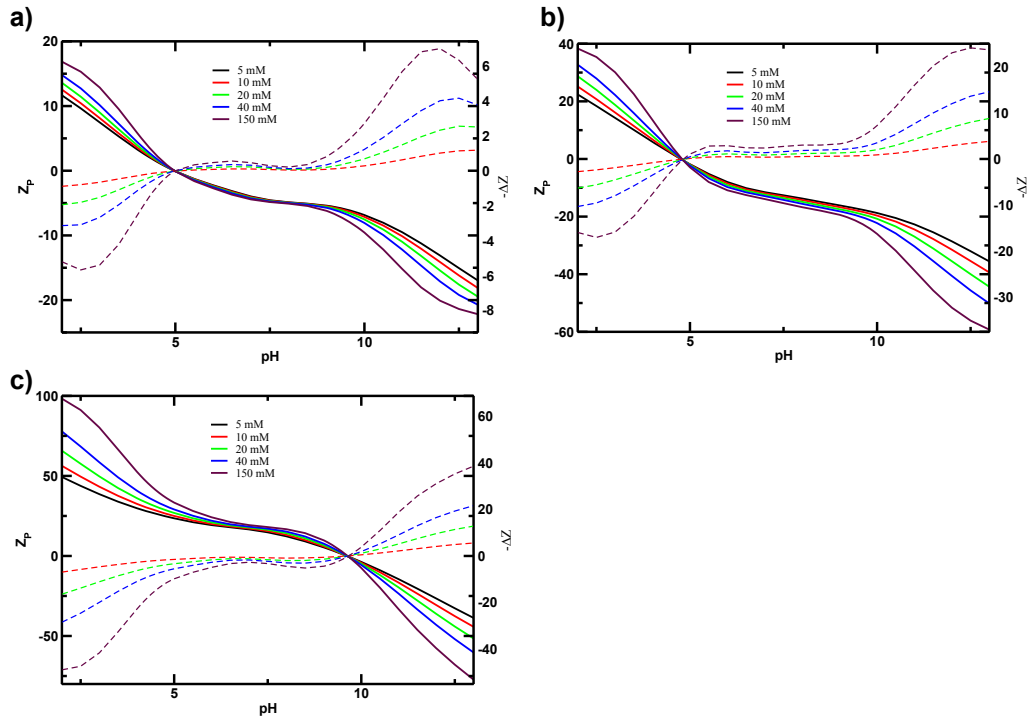


Fig. 2. The simulated charge number ($\langle Z_p \rangle$) of the whey proteins α -LA, β -LG and LF as a function of pH at different salt concentrations. The differences between charge numbers at different salt concentrations ($\Delta Z = \langle Z_{salt} \rangle - \langle Z_{ref} \rangle$) are also displayed in the plots (right axis). The system at 5 mM of salt was assumed as the reference state. a) Top panel: Data for α -LA. b) Left panel: Data for β -LG. c) Right panel: Data for LF. .

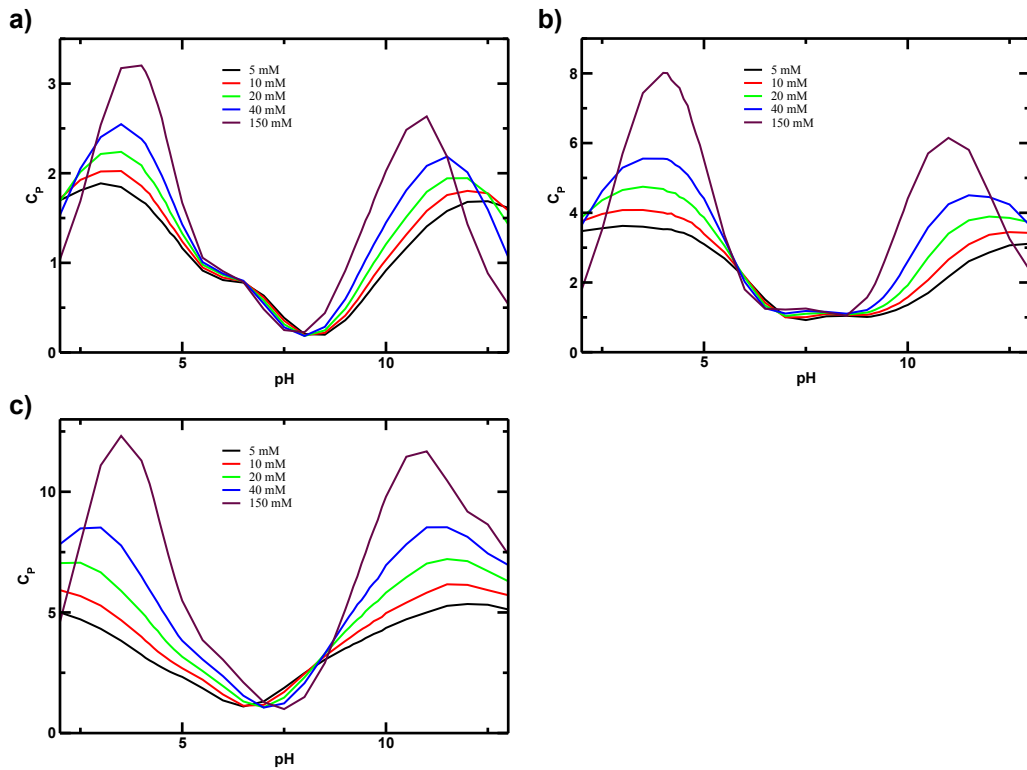


Fig. 3. The simulated capacitances of the whey proteins α -LA, β -LG and LF as a function of pH at different salt concentrations. a) Top panel: Data for α -LA. b) Left panel: Data for β -LG. c) Right panel: Data for LF. .

6.5–10.0 for α -LA–LF and β -LG–LF complexes, respectively. No association would be observed for α -LA– β -LG at any solution pH.

Salt effect is scrutinized in Fig. 7a–c while the solution pH was kept constant at 6.5. Salt has a clear screening effect in all studied

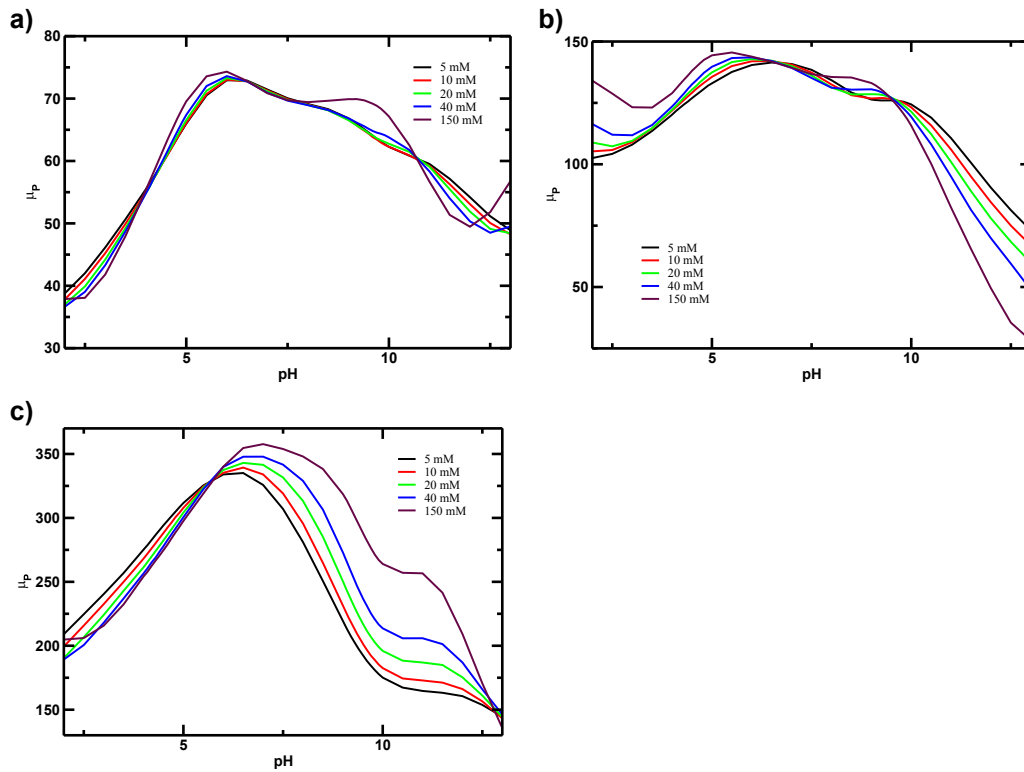


Fig. 4. The simulated dipole moment number of the whey proteins α -LA, β -LG and LF as a function of pH at different salt concentrations. a) Top panel: Data for α -LA. b) Left panel: Data for β -LG. c) Right panel: Data for LF.

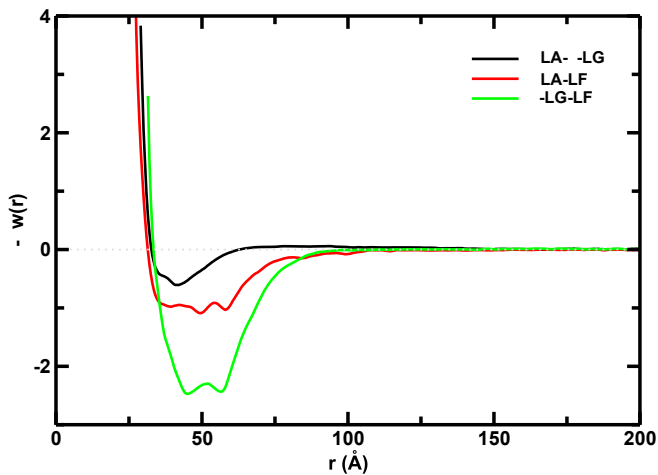


Fig. 5. The simulated free energy of interactions [$\beta w(r)$] between the centers of mass of the proteins. The solution pH was fixed at 6.5. Data for the systems α -LA– β -LG (top panel), α -LA–LF (left panel) and β -LG–LF (right panel) at 150 mM salt concentration.

systems. At this pH and 150 mM salt solution, the simulated proteins net charge numbers are $Z_{\alpha-LA} = -3.7$, $Z_{\beta-LG} = -12.5$ and $Z_{LF} = +21.2$. The attractive minima in the protein complexes are reduced upon addition of salt when proteins are oppositely charged. Similarly, for like charged proteins (α -LA and β -LG), the repulsion is reduced. We added a curve showing the simulated van der Waals contribution to the free energy of interaction. This permits us to distinguish the main driven force among electrostatic interactions and vdw. At low salt, the complexation mechanism is governed by electrostatic interactions with a predominant role of the protein charges. Conversely, as expected, for the α -LA– β -LG

system, the salt screening is too severe with the electrostatic interactions (multipolar-multipolar interactions and the charge regulation mechanism). At 150 mM, the free energy minimum for this system is solely due to vdw. All electrostatic interactions (including the repulsive coulombic interaction due the like charge of both proteins) were screened out. For the other systems, due their charges of opposite signals, the LF's large dipole moment and its high capacitance, the depth of the minimum is always bigger and more highlighted. At low salt concentrations, the α -LA–LF and β -LG–LF systems have a market tendency to be strongly associated. The measured free energy minimum is smaller than -10 and $-20 k_B T$ for α -LA–LF and β -LG–LF, respectively. In all studied conditions, β -LG interacts more strongly with LF than α -LA does. Considering the high concentration of salt species (Sodium, Potassium, Chloride, Sulfate, Phosphate, Citrate, Calcium and Magnesium) in natural milk (Thompson et al., 2009), most of these complexes will however not be observed in this medium. In a similar manner, at high protein concentrations, electrostatic interactions are more screened out reducing the association of these proteins. A small amount of salt in a high protein concentration can have an equivalent ionic strength as a high salt concentrated system. In this condition, the maximum salt concentration where the complexation can occur is decreased. This might explain why Anema & de Kruif observed that β -LG–LF complexes were formed in the 5–7.3 pH window with a smaller critical salt concentration (of about 100 mM at a protein concentration of ca. 1 g/L) in their work (Anema & Kruif, 2014). A higher protein concentration (20 g/L) and the presence of aggregates decrease even more this maximum salt concentration as observed by other authors (Yan et al., 2013).

The simulated protein net charges at 150 mM as a function of pH were used as input for Eq. (4). This analytical estimated free energies of interactions are displayed at Fig. 8. The pH windows at

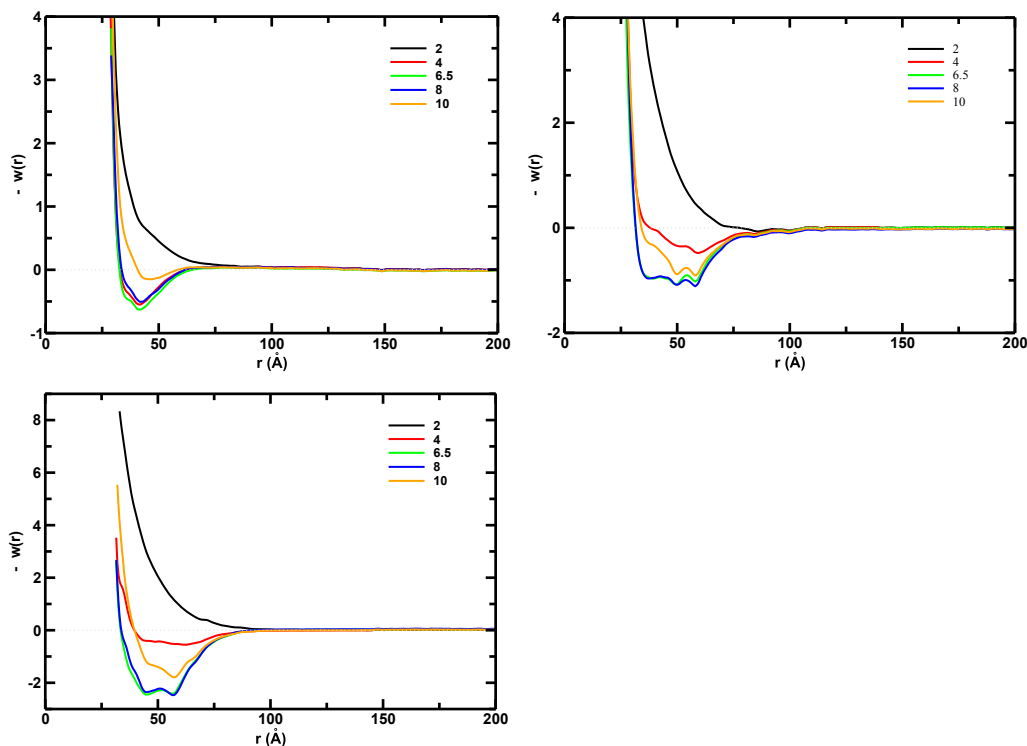


Fig. 6. The simulated free energy of interactions [$\beta w(r)$] between the centers of mass of the proteins at different solution pHs. Data for the systems α -LA- β -LG (top panel), α -LA-LF (left panel) and β -LG-LF (right panel) at 150 mM salt concentration.

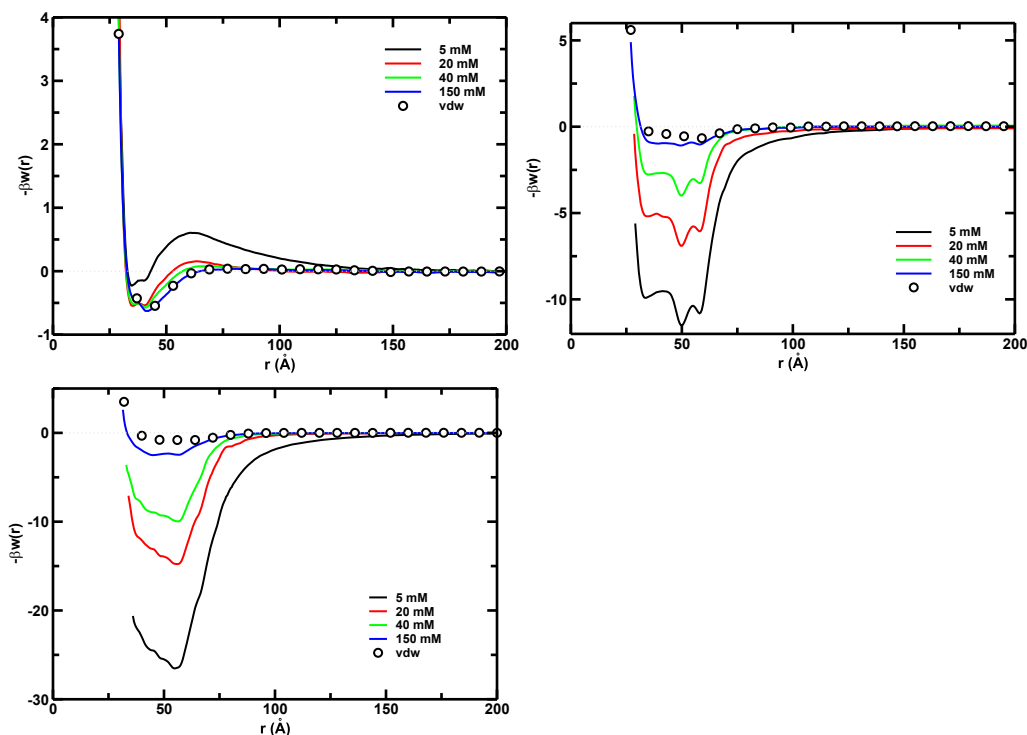


Fig. 7. The simulated free energy of interactions [$\beta w(r)$] between the centers of mass of the proteins at different ionic strengths. Data for the systems α -LA- β -LG (top panel), α -LA-LF (left panel) and β -LG-LF (right panel) at pH 6.5. The open circle curves are data obtained from MC simulations with all protein aminoacids assumed to be neutral.

which an attractive interaction between α -LA-LF and β -LG-LF predicted by Eq. (4) are 5.0 – 9.6 and 4.8 – 9.6, respectively. Even in this crude estimation, β -LG better interacts with LF than α -LA. Such

analytical tool only predicts a complexation for α -LA- β -LG at pH 4.8 – 4.9 (data not shown in the plot). These are results from a pure Coulombic charge-charge regime. In the simulated free energies of

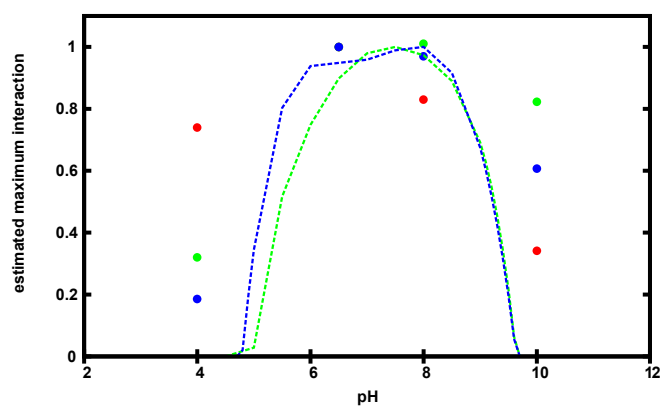


Fig. 8. Effect of pH on the simulated (filled symbols) and estimated (dashed line) interaction between α -LA-LF (green) and β -LG-LF (blue). The salt concentration was fixed at 150 mM. The red symbols are data obtained from MC simulations for the α -LA- β -LG. The corresponding estimated values by Eq. 4 are pH 4.8 – 4.9 (data not shown in the plot). All free energy of interactions are normalized – see text for details.

interactions (Figs. 5–7), multipolar, charge regulation and vdW interactions were also present in the model. Comparing the behavior of the analytical A_{DH} with the simulated data normalized in the same way reveals that these interactions are indeed responsible to enhance the attraction. The complexation is then observed at a wider pH window. This feature is similar to the experimental and predicted data reported by Egan et al. (2014). Nevertheless, it is interesting to point out that for these protein systems (α -LA- β -LG, α -LA-LF and β -LG-LF) protein net charges dominate their interactions making them to behave as colloidal particles at several pHs. This is not an universal pattern. In the literature, it can be seen cases where the charge regulation mechanism dominates the interaction [e.g. ref. (Da Silva & Jönsson, 2009; da Silva et al., 2006)] or the dipole-dipole interactions [e.g. ref. (Kurut, Dicko, & Lund, 2015)]. There are also situations where each of these interactions is the main drive force depending on the experimental conditions (da Silva et al., 2014).

4. Concluding remarks

Whey proteins were designed to have their optimal attractive features at milk's natural pH under physiological conditions. Salt and pH can be used to manipulate the relative strength of the association free energy minima. By means of Monte Carlo simulations we have defined the pH window at physiological conditions for α -LA-LF and β -LG-LF complexes. We found that virtually no association is expected to α -LA and β -LG at this relatively high salt concentration regime for applications in food systems. The computed free energy of interaction between α -LA and LF revealed the weaker attraction between them in contrast with the strongest attraction measured for β -LG and LF explaining experimental uncertainties. Reducing the ionic strength wider the pH windows reported here. The driven force for the studied complexation with LF is due an association of electrostatic interactions (charge-charge, charge-dipole and induced charge interactions due charge fluctuations) where protein charge-charge plays a key role. The present results can guide practical applications of whey proteins, for instance, to produce stable beverages.

Acknowledgments

This work has been supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes 2405/13-0),

Fundação de Amparo a Pesquisa do Estado de São Paulo (Fapesp 2010/50425-0) and the Swedish Research Council through a Linnaeus grant. FLBDS thanks also the support of the University of São Paulo through the NAP-CatSinQ (Research Core in Catalysis and Chemical Synthesis) and the hospitality of the University of Lund. Both authors express their deepest gratitude to Prof. Bo Jönsson for fruitful discussions and his critical reading of the manuscript. It is also a pleasure to acknowledge Dr. Björn Persson for his initial assistance with Faunus package.

References

- Allen, M. P., & Tildesley, D. J. (1989). *Computer simulation of liquids*. Oxford: Oxford University Press.
- Aly, E., Ros, G., & Frontela, C. (2013). Structure and functions of lactoferrin as ingredient in infant formulas. *Journal of Food Research*, 2(4), 25–36.
- Anema, S. G., & Kruij, C. G. D. (2014). Complex coacervates of lactotransferrin and β -lactoglobulin. *Journal of Colloid and Interface Science*, 430, 214–220.
- Audica, J.-L., Chauferat, B., & Daufin, G. (2003). Non-food applications of milk components and dairy co-products: a review. *Le Lait*, 83, 417–438.
- Batt, C. A., Brady, J., & Sawyer, L. (1994). Design improvements of β -lactoglobulin. *Trends in Food Science & Technology*, 5(8), 261–265.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Research*, 28, 235–242.
- Bösze, Z. (Ed.). (2008). *Bioactive components of milk* (1st ed.). New York: Springer Science Business Media.
- de Carvalho, S. J., Ghiotto, R. T., & da Silva, F. L. B. (2006). Monte Carlo and modified Tanford-Kirkwood results for macromolecular electrostatics calculations. *Journal of Physical Chemistry B*, 110, 8832–8839.
- Clare, D. A., Catignani, G. L., & Swaisgood, H. E. (2003). Biodefense properties of milk: the role of antimicrobial proteins and peptides. *Current Pharmaceutical Design*, 9, 1239–1255.
- Clare, D. A., & Daubert, C. R. (2011). Expanded functionality of modified whey protein dispersions after transglutaminase catalysis peptides. *Journal of Food Science*, 76(4), C576–C584.
- Da Silva, F. L. B., & Jönsson, B. (2009). Polyelectrolyte-protein complexation driven by charge regulation. *Soft Matter*, 5(15), 2862–2868.
- Dalgleish, D. G. (1992). Bovine milk protein properties and the manufacturing quality of milk. *Livestock Production Science*, 35, 75–93.
- Dickinson, E. (2008). Interfacial structure and stability of food emulsions as affected by protein–polysaccharide interactions. *Soft Matter*, 4, 932–942.
- Dickinson, E. (2013). Stabilising emulsion-based colloidal structures with mixed food ingredients. *Journal of the Science of Food and Agriculture*, 93, 710–721.
- Donald, A. M. (2000). Aggregation in β -lactoglobulin. *Soft Matter*, 4, 1147–1150.
- Egan, T., Jacquier, J.-C., Rosenberg, Y., & Rosenberg, M. (2013). Cold-set whey protein microgels for the stable immobilization of lipids. *Food Hydrocolloids*, 31, 317–324.
- Egan, T., O'Riordan, D., O'Sullivan, M., & Jacquier, J.-C. (2014). Cold-set whey protein microgels as pH modulated immobilisation matrices for charged bioactives. *Food Chemistry*, 156, 197–203.
- El-Loly, M. M., & Mahfouz, M. B. (2011). Lactoferrin in relation to biological functions and applications: a review. *International Journal of Dairy Science*, 6(2), 79–111.
- Eratte, D., Wang, B., Dowling, K., Barrow, C. J., & Adhikari, B. P. (2014). Complex coacervation with whey protein isolate and gum arabic for the microencapsulation of omega-3 rich tuna oil. *Food & Function*, 5, 2743–2750.
- Evershed, R. P., Payne, S., Sherratt, A. G., Copley, M. S., Coolidge, J., Urem-Kotsu, D., et al. (2008). Earliest date for milk use in the near east and southeastern europe linked to cattle herding. *Nature*, 455, 528–531.
- Farrell, H. M., Jr., Jimenez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., et al. (2004). Nomenclature of the proteins of cows—milk—sixth revision. *Journal of Dairy Science*, 87, 1641–1674.
- Fogolari, F., Ragona, L., Licciardi, S., Romagnoli, S., Michelutti, R., Ugnolini, R., et al. (2000). Electrostatic properties of bovine β -lactoglobulin. *Proteins-Structure Function And Genetics*, 39, 317–330.
- Frenkel, D., & Smit, B. (1996). *Understanding molecular Simulation: From algorithms to applications*. San Diego: Academic Press.
- Girard, M., & Schaffer-Lequart, C. (2007). Attractive interactions between selected anionic exopolysaccharides and milk proteins. *Food Hydrocolloids*, 22, 1425–1434.
- Gottschalk, M., Nilsson, H., Roos, H., & Halle, B. (2003). Protein self-association in solution: the bovine β -lactoglobulin dimer and octamer. *Protein Science*, 12, 2404–2411.
- Hattori, T., Bataldar, S., Kato, R., Bohidar, H., & Dubin, P. (2005). Characterization of polyanionprotein complexes by frontal analysis continuous capillary electrophoresis and small angle neutron scattering: effect of polyanion flexibility. *Analytical Biochemistry*, 342(2), 229–236.
- Havea, P., Singh, H., & Creamer, L. K. (2001). Characterization of heat-induced aggregates of beta-lactoglobulin, alpha-lactalbumin and bovine serum albumin in a whey protein concentrate environment. *Journal of Dairy Research*, 68(3), 483–497.
- Hekman, A. (1971). Association of lactoferrin with other proteins, as demonstrated

- by changes in electrophoretic mobility. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 251(3), 380–387.
- Hill, T. L. (1986). *An introduction to statistical thermodynamics*. New York: Dover Publications Inc.
- Jeng, S.-Y., Bleck, G. T., Wheeler, M. B., & Jime, R. (1997). Characterization and partial purification of bovine α -lactalbumin and β -casein produced in milk of transgenic mice. *Journal of Dairy Science*, 80, 3167–3175.
- Jönsson, B. (1981). *The thermodynamics of ionic Amphiphilic–Water systems — a theoretical analysis*. Ph.D. thesis. Lund, Sweden: Lund University.
- Jönsson, B., Lund, M., & da Silva, F. L. B. (2007). Electrostatics in macromolecular solution. In E. Dickinson, & M. E. Leser (Eds.), *Food colloids: Self-assembly and material science* (pp. 129–154). Londres: Royal Society of Chemistry.
- Ju, Z. Y., Hettiarachchy, N., & Kilara, A. (1999). Thermal properties of whey protein aggregates. *Journal of Dairy Science*, 82(9), 1882–1889.
- Kesvatera, T., Jönsson, B., Thulin, E., & Linse, S. (1999). Ionization behavior of acidic residues in calbindin D_{9k}. *Proteins-Structure Function And Genetics*, 37, 106–115.
- Kinsella, J. E., & Whitehead, D. M. (1989). Proteins in whey: chemical, physical, and functional properties. *Advances in Food and Nutrition Research*, 33, 343–438.
- Kizilay, E., Seeman, D., Yan, Y., Du, X., Dubin, P. L., Donato-Capel, L., et al. (2014). Structure of bovine β -lactoglobulin and lactoferrin coacervates. *Soft Matter*, 10, 7262–7268.
- Korhonen, H. (1998). Impact of processing on bioactive proteins and peptides. *Trends in Food Science & Technology*, 9(8–9), 307–319.
- Kuehner, D. E., Engmann, J., Fergg, F., Wernick, M., Blanch, H. W., & Prausnitz, J. M. (1999). Lysozyme net charge and ion binding in concentrated aqueous electrolyte solutions. *Journal Of Physical Chemistry B*, 103, 1368–1374.
- Kurut, A., Dicko, C., & Lund, M. (2015). Dimerization of terminal domains in spiders silk proteins is controlled by electrostatic anisotropy and modulated by hydrophobic patches. *ACS Biomaterials Science & Engineering*, 1(6), 363–371.
- Kurut, A., Persson, B. A., Åkesson, T., Forsman, J., & Lund, M. (2012). Anisotropic interactions in protein mixtures: self assembly and phase behavior in aqueous solution. *Journal of Physical Chemistry Letters*, 3(6), 731–734.
- Lampreave, F., Piñeiro, A., Brock, J. H., Castillo, H., Sánchez, L., & Calvo, M. (1990). Interaction of bovine lactoferrin with other proteins of milk whey. *International Journal of Biological Macromolecules*, 12(1), 2–5.
- Laos, K., Brownsey, G., & Ring, S. (2007). Interactions between furcellaran and the globular proteins bovine serum albumin and β -lactoglobulin. *Carbohydrate Polymers*, 67(1), 116–123.
- Li, Y., Mattison, K. W., Dubin, P. L., Havel, H. A., & Edwards, S. L. (1996). Light scattering studies of the binding of bovine serum albumin to a cationic polyelectrolyte. *Biopolymers*, 38(4), 527–533.
- Lönnerdal, B., & Iyer, S. (1995). Lactoferrin: molecular structure and biological function. *Annual Review of Nutrition*, 15, 93–110.
- Luensmann, D., & Jones, L. (2008). Albumin adsorption to contact lens materials: a review. *Cont Lens Anterior Eye*, 31(4), 179–187.
- Lund, M., & Jönsson, B. (2003). A mesoscopic model for protein-protein interactions in solution. *Biophysical Journal*, 85, 2940–2947.
- Lund, M., & Jönsson, B. (2013). Charge regulation in biomolecular solution. *Quarterly Reviews of Biophysics*, 46, 265–281.
- Marcus, R. A. (1955). Calculation of thermodynamic properties of polyelectrolytes. *Journal of Chemical Physics*, 23, 1057.
- McSweeney, P. L., & Fox, P. F. (Eds.). (2013). *Advanced dairy chemistry - Proteins: Basic aspects* (4th ed.). New York: Springer Science Business Media.
- Medda, L., Barse, B., Cugia, F., Boström, M., Parsons, D. F., Ninham, B. W., et al. (2012). Hofmeister challenges: Ion binding and charge of the bsa protein as explicit examples. *Langmuir*, 28, 16355–16363.
- Metropolis, N. A., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1097.
- Morawetz, H., & Hughes, W. L. (1952). The interaction of proteins with synthetic polyelectrolytes. I. Complexing of bovine serum albumin. *The Journal of Physical Chemistry*, 56(1), 64–69.
- Mudgal, P., Daubert, C. R., Clare, D. A., & Foegeding, E. A. (2011). Effect of disulfide interactions and hydrolysis on the thermal aggregation of β -lactoglobulin. *Journal of Agricultural and Food Chemistry*, 59, 1491–1497.
- Nigen, M., Croguennec, T., & Bouhallab, S. (2009). Formation and stability of α -lactalbumin–lysozyme spherical particles: Involvement of electrostatic forces. *Food Hydrocolloids*, 23(2), 510–518.
- Nigen, M., Tilly, V., Croguennec, T., Drouinukma, D., & Bouhallab, S. (2009). Molecular interaction between apo or holo α -lactalbumin and lysozyme: formation of heterodimers as assessed by fluorescence measurements. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1794(4), 709–715.
- Nozaki, Y., & Tanford, C. (1967). Examination of titration behavior. *Methods Enzymol*, 11, 715–734.
- Pedersen, J. B., Fojan, P., Sorensen, J., & Petersen, S. B. (2006). Towards control of aggregational behaviour of α -lactalbumin at acidic pH. *Journal of Fluorescence*, 16(2), 611–621.
- Permyakov, S. E., Pershikova, I. V., Khokhlova, T. I., Uversky, V. N., & Permyakov, E. A. (2004). No need to be HAMLET or BAMLET to interact with histones: binding of monomeric α -lactalbumin to histones and basic poly-amino acids. *Biochemistry*, 43(19), 5575–5582.
- Persson, B., Lund, M., Forsman, J., Chatterton, D. E. W., & Åkesson, T. (2010). Molecular evidence of stereo-specific lactoferrin dimers in solution. *Biophysical Chemistry*, 3(3), 187–189.
- Puyol, P., Perez, D. M., Ena, J. M., & Calvo, M. (1991). Interaction of bovine beta-lactoglobulin and other bovine and human whey proteins with retinol and fatty acids. *Agricultural and Biological Chemistry*, 55(10), 2515–2520.
- Raemy, A., Lambelet, P., & Rousset, P. (2005). Calorimetric information about food and food constituents. In D. Lőrinczy (Ed.), *The nature of biological systems as revealed by thermal methods* (pp. 69–98). Netherlands: Springer.
- de Roos, A. L., Walstra, P., & Geurts, T. J. (1998). The association of lysozyme with casein. *International Dairy Journal*, 8, 319–324.
- Ryan, K. N., & Foegeding, E. A. (2015). Formation of soluble whey protein aggregates and their stability in beverages. *Food Hydrocolloids*, 43, 265–274.
- Ryan, K. N., Vardhanabhuti, B., Jaramillo, D. P., van Zanten, J. H., Coupland, J. N., et al. (2012). Stability and mechanism of whey protein soluble aggregates thermally treated with salts. *Food Hydrocolloids*, 27, 411–420.
- Secchi, G. (2008). Role of protein in cosmetics. *Clinics in Dermatology*, 26(4), 321–325.
- Seyrek, E., Hattori, T., & Dubin, P. L. (2004). Frontal analysis continuous capillary electrophoresis for protein-polyelectrolyte binding studies. In M. A. Strega, & A. L. Lagu (Eds.), *Methods in molecular biology: Capillary electrophoresis of proteins and peptides* (pp. 217–228). Humana Press Inc., Ch. 11.
- da Silva, F. L. B., Boström, M., & Persson, C. (2014). Effect of charge regulation and ion-dipole interactions on the selectivity of protein-nanoparticle binding. *Langmuir*, 30(14), 4078–4083.
- da Silva, F. L. B., Jönsson, B., & Penfold, R. (2001). A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations. *Protein Science*, 10, 1415–1425.
- da Silva, F. L. B., Lund, M., Jönsson, B., & Åkesson, T. (2006). On the interaction between protein and polyelectrolyte. *Journal of Physical Chemistry B*, 110(9), 4459–4464.
- Smithers, G. (2008). Whey and whey proteins: from gutter-to-gold. *International Dairy Journal*, 18(7), 695–704.
- Smithers, G. W., Ballard, F. J., Copeland, A. D., De Silva, K. J., Dionysius, D. A., Francis, G. L., et al. (1996). New opportunities from the isolation and utilization of whey proteins. *Journal of Dairy Science*, 79(8), 1454–1459.
- Soltys-Robitaille, C., Ammon, D. M., Jr., Valint, P. L., Jr., & Grobe, G. L., III (2001). The relationship between contact lens surface charge and in-vitro protein deposition levels. *Biomaterials*, 22(24), 3257–3260.
- Steijns, J. M., & van Hooijdonk, A. C. M. (2000). Occurrence, structure, biochemical properties and technological characteristics of lactoferrin. *British Journal of Nutrition*, 84(Supplement, S1), 11–17.
- Stenqvist, B., Thuresson, A., Kurut, A., Vácha, R., & Lund, M. (2013). Faunus – a flexible processor for Monte Carlo simulation. *Mol Sim*, 39(14–15), 1233–1239.
- Striolo, A., Bratko, D., Wu, J. Z., Elvassore, N., Blanch, H. W., & Prausnitz, J. M. (2002). Forces between aqueous nonuniformly charged colloids from molecular simulation. *Journal of Chemical Physics*, 116, 7733–7743.
- Syrbe, A., Bauer, W., & Klostermeyer, H. (1998). Polymer science concepts in dairy systems—an overview of milk protein and food hydrocolloid interaction. *International Dairy Journal*, 8(3), 179–193.
- Takahashi, T. (1997). Significant role of electrostatic interactions for stabilization of protein assemblies. *Advances in Biophysics*, 34, 41–54.
- Takayama, Y. (2012). *Lactoferrin and its role in wound healing* (1st ed.). New York: Springer Science Business Media.
- Tcholakova, S., Denkov, N. D., Ivanov, I. B., & Campbell, B. (2006). Coalescence stability of emulsions containing globular milk proteins. *Advances in Colloid and Interface Science*, 123–126, 259–293.
- Teixeira, A. A., Lund, M., & da Silva, F. L. (2010). Fast proton titration scheme for multiscale modeling of protein solutions. *Journal of Chemical Theory and Computation*, 6(10), 3259–3266.
- Tenovuo, J. (2002). Clinical applications of antimicrobial host proteins lactoperoxidase, lysozyme and lactoferrin in xerostomia: efficacy and safety. *Oral Diseases*, 8(1), 23–29.
- Thompson, A., Boland, M., & Singh, H. (2009). *Milk proteins: From expression to food*. USA: Academic Press/Elsevier.
- Tomita, M., Wakabayashi, H., Shin, K., Yamauchi, K., Yaeshima, T., & Iwatsuki, K. (2009). Twenty-five years of research on bovine lactoferrin applications. *Biochimie*, 91(1), 52–57.
- Topuzogullari, M., Cimen, N., Mustafaeva, Z., & Mustafaev, M. (2007). Molecular-weight distribution and structural transformation in water-soluble complexes of poly(acrylic acid) and bovine serum albumin. *European Polymer Journal*, 43(7), 2935–2946.
- Touhami, A., & Dutcher, J. R. (2009). pH-induced changes in adsorbed β -lactoglobulin molecules measured using atomic force microscopy. *Soft Matter*, 5(1), 220–227.
- Valette, P., Thomas, M., & Déjardin, P. (1999). Adsorption of low molecular weight proteins to hemodialysis membranes: experimental results and simulations. *Biomaterials*, 20, 1621–1634.
- Vardhanabhuti, B., Yucler, U., Coupland, J. N., & Foegeding, E. A. (2009). Interactions between β -lactoglobulin and dextran sulfate at near neutral pH and their effect on thermal stability. *Food Hydrocolloids*, 23(6), 1511–1520.
- Wakabayashi, H., Oda, H., Yamauchi, K., & Abe, F. (2014). Lactoferrin for prevention of common viral infections. *Journal of Infection and Chemotherapy*, 20, 666–671.
- Wang, Y.-F., Gao, J. Y., & Dubin, P. L. (1996). Protein separation via polyelectrolyte coacervation: selectivity and efficiency. *Biotechnology Progress*, 12(3), 356–362.
- Watson, R. R., Zibadi, S., & Preezy, V. R. (Eds.). (2010). *Dietary components and immune function* (1st ed.). New York: Springer Science Business Media.
- Wijayanti, H. B., Bansal, N., & Deeth, H. C. (2014). Stability of whey proteins during thermal processing: a review. *Comprehensive Reviews in Food Science and Food*

- Safety*, 13, 1235–1251.
- de Wit, J. (2009). Thermal behaviour of bovine β -lactoglobulin at temperatures up to 150 C. A review. *Trends in Food Science & Technology*, 20, 27–34.
- Yan, Y., Kizilay, E., Seeman, D., Flanagan, S., Dubin, P. L., Bovetto, L., et al. (2013). Heteroprotein complex coacervation: bovine β -Lactoglobulin and lactoferrin. *Langmuir*, 29, 15614–15623.
- Zeiler, R. N. W., & Bolhuis, P. G. (2015). Exposure of thiol groups in the heat-induced denaturation of β -lactoglobulin. *Mol Sim*, 41(10–12), 1006–1014.

APÊNDICE C – Artigo em preparação

Electrostatic Correlations in Proteins: a large-scale analysis on the Protein Data Bank
LARIANI APARECIDA DELBONI, BO JÖNSSON AND FERNANDO LUÍS BARROSO
DA SILVA

Electrostatic Correlations in Proteins: a large-scale analysis on the Protein Data Bank

LARIANI APARECIDA DELBONI,^a BO JÖNSSON^b AND FERNANDO LUÍS BARROSO DA
SILVA^{a,b*}

^a*Department of Physics and Chemistry, Faculty of Pharmaceutical Sciences at Ribeirão Preto – University of São Paulo, BR-14040–903, Av. do café, s/no., Ribeirão Preto, SP, Brazil, and* ^b*Department of Theoretical Chemistry, Chemical Center, Lund University, POB 124, S-221 00 Lund, Sweden. E-mail: flbarros@ncsu.edu*

Abstract

Electrostatic interactions are often seen as being important for protein structure and stability. In order to scrutinize this idea, the present study avoided the construction of any approximate physical model and it was based on the study of direct structural experimental data. This has been done by calculating the probability distributions for equally and oppositely charged groups directly from all the three dimensional coordinates available at the Protein Data Bank – 72,966 proteins were analysed.

Most of the protein coordinates show some positive correlation, but then only at *very short* separations – the formation of so-called salt bridges. However, approximately 13% of the investigated structures show negative charge-charge correlations. That is, it is more probable to find a positive or negative charge close to another equally charged one than an oppositely

charged atom. No remarkable difference in behaviour is found between x-ray and nmr determined structures. Structures obtained from molecular dynamics simulation trajectories and homology modeling also show the same characteristics.

Based on such analysis, it can be tempted to conclude that electrostatic interactions play a role in stabilizing the tertiary structure of proteins, but they can also under some conditions be effectively screened. In the latter case it could be that functional aspects are more important than a minimization of the electrostatic free energy.

1. Introduction

The driving forces behind protein folding and stability have been much debated in the literature over the years (Hill, 1955; Tanford & Kirkwood, 1957; Warshel, 1981; Dill, 1990; Warshel & Åqvist, 1991; Myers & Pace, 1996; Miki *et al.*, 1997; Whitten & E., 2000; Chi *et al.*, 2003; Roca *et al.*, 2007; Hildebrandt *et al.*, 2007; Gvritishvili *et al.*, 2008; Vicatos *et al.*, 2009; Pace *et al.*, 2009; Jefferys *et al.*, 2010; Thirumalai *et al.*, 2010; Sterpone & Melchionna, 2012). Today there seems to be a consensus that hydrophobic interactions is a major component at least in the folding process. (Dill *et al.*, 2008) The importance of direct electrostatic interactions is more obscure both in the protein folding and protein stability issues. Still it is often stated that "electrostatic interactions are important in proteins". (Hill, 1955; Tanford & Kirkwood, 1957; Warshel, 1981; Dill, 1990; Warshel & Åqvist, 1991; Perutz, 1978; Nishimura *et al.*, 2001) For instance, focusing on protein stability, if electrostatic interactions are a major component in protein stability, one would expect a distinct correlation between charged groups in tertiary protein structures. This has been investigated by Wada and Nakamura, (Wada & Nakamura, 1981) who calculated the Coulomb energy between equal and opposite charges as a function of their separation. The charges in the protein were

defined according to certain rules. The electrostatic interaction energy was calculated from the Tanford-Kirkwood (TK) model. (Tanford & Kirkwood, 1957) Based on such a model, their conclusion was that each charge is stabilized by several units of kT (k is the Boltzmann constant and T is the temperature). Later, Barlow and Thornton (Barlow & Thornton, 1983) avoided the construction of an *approximate* energy function, but instead they calculated ion-pair distance distributions for like and oppositely charged residues directly from experimental data. Their results indicate that charged residues play a role in stabilizing the tertiary structure of globular proteins. Unfortunately, their calculations were limited to rather short charge-charge distances, $< 10 \text{ \AA}$, and only a small number (38) of protein structures were studied. This work has been recently extended by Gvritishvili and colleagues (Gvritishvili *et al.*, 2008) to explore the geometries of salt bridges in a representative nonredundant set of 1,500 protein coordinates from the Protein Data Bank (PDB). (Berman *et al.*, 2000) Their conclusions reinforce the importance of the electrostatic interactions for protein stability. Conversely, due to the charge anisotropy of proteins, some proteins do show a preference to have equally charged residues close to each other in what has been called "charged patches". (de Vries & Stuart, 2006; de Vos *et al.*, 2010) Another set of papers has been published by Warshel and collaborators. (Roca *et al.*, 2007; Vicatos *et al.*, 2009) Employing the semi-macroscopic Langevin dipole method in its linear response approximation version, these authors analyzed the electrostatic contributions to protein stability. Their analyses were limited to a dataset containing a relatively small number (45) of protein structures.

Effective Hamiltonians contain some degree of approximation and offer room for discussion and criticisms. These physical models are often under intense debate and strong controversial opinions of how to model the dielectric response of the protein interior can be found in the literature. Since the introduction of a dielectric interface by the TK model, many works recurrently discuss this theme. (Demchuk & Wade, 1996; Penfold *et al.*, 1998; Warshel & Åqvist, 1991; King *et al.*, 1991; Antonsiewicz *et al.*, 1994; Simonson & Perahia, 1995; Simon-

son & Brooks III, 1996; Antonsiewicz *et al.*, 1996; Löffler *et al.*, 1997; Sham *et al.*, 1997; Warwicker, 1999; da Silva *et al.*, 2001; Autreto *et al.*, 2003; Schutz & Warshel, 2001; Dudev & Lim, 2000; Varma & Jakobsson, 2004; Archontis & Simonson, 2005; Ko *et al.*, 2005; He *et al.*, 2007; de Carvalho *et al.*, 2008; Vicatos *et al.*, 2009; Kukic *et al.*, 2013).

In this communication we avoid the construction of any approximate Hamiltonian and base our study on direct structural experimental data. All protein tertiary coordinates available on the PDB are directly used as our experimental source with no further manipulation. This is the largest protein structures dataset analysed so far in the literature. We measured the correlations on *all* ranges by calculating the probability distributions of positive and negative charges in a protein structure as a function of their separation, R – the three distributions are $P_{++}(R)$, $P_{--}(R)$ and $P_{+-}(R)$. Thus, we use an experimentally determined quantity, that is the Euclidean separation distance between charges of titrable residues as found in protein coordinates available on the PDB. We expect to see that $P_{+-}(R) > P_{++}(R)$, $P_{--}(R)$ for small R , while the opposite should be true at large R . We will also compare these protein charge distributions to the corresponding correlation functions for a simple electrolyte solution confined in a sphere.

2. Approach

2.1. Charge correlations in proteins

Several datasets of protein coordinates obtained from the Protein Data Bank (PDB) (Berman *et al.*, 2000) containing proteins with different physical chemical properties, structural classification and biological functions were used in the present study – see Table 1. Some of these contain a vast number of protein coordinates. In one of them, the whole PDB was used. In order to evaluate possible artifacts due the redundancy, (Holm *et al.*, 2008) we have also carried out calculations with a non-redundant PDB chain set provided by the VAST (Vector Alignment Search Tool) server. (Gibrat *et al.*, 1996) Other datasets were specific to some bio-

logical function as given in Table 1. This is an interesting classification in order to explore particularly patterns related with the biological function as suggested before. (Gutteridge & Thornton, 2005) The coordinates of all protein atoms present in the three dimensional structure were used without any manipulation. Proteins were assumed to be monomers, i.e. when more than one chain was available in the PDB coordinate file, each chain was treated as one individual protein, and the geometrical analysis were performed for all them. The effect of such multiple protein copies was investigated by carrying out analysis with only one selected single chain from this collection. Heteroatoms (ions, ligands, metals, and solvent molecules) were not considered (unless otherwise stated) due the fact that the emphasis here is on the protein structure. Another reason to remove metal ions is the need for a uniform criterion since metalloproteins can be found in PDB in the halo (presence of metal ions in the coordinate file) and apo (absence of metal ions in the coordinate file) forms. Conversely, calcium binding proteins were investigated in both forms in order to describe the effects of such assumption. For the same reason, we have also performed analysis for a protein non-redundant coordinates set that do not contain heteroatoms (except water) in the coordinates files. In all cases, missing atoms were not considered or added. The PDB file might contain uncertainties, (Davis *et al.*, 2003) and, before analysis, the files were corrected. When atoms that are statically disordered were presented in the structures, the form labeled "A" was chosen. Proteins that exist in more isoforms were considered only by the structure given in the PDB file.

The probability distribution of charges in each protein was evaluated by simply measuring the distance between all charges in the protein coordinate creating a histogram for $P(R)$ with a bin size equal to 0.5 \AA . $P(R)$ was normalized in order to have its integral equal to one. We expect to see that the probability distribution for oppositely charged residues, P_{+-} , is larger than the one for equally charged residues, P_{++} or P_{--} at short separation distances. At large distances the opposite will be true, $P_{++}, P_{--} > P_{+-}$, since the probability distributions are normalized. We will refer to $P_{+-} > P_{++}, P_{--}$ as a *positive charge correlation*, while if the

opposite is true, we will use the term *negative charge correlation*. Titratable residues were assigned charges corresponding to a solution pH of 7, (Pace *et al.*, 2009; Teixeira *et al.*, 2010) which means that some of the proteins will have a considerable net charge. We have followed here the same procedure used by Barlow and Thornton. (Barlow & Thornton, 1983) Other criteria were also tested resulting in similar outcomes. However, the P_{+-} distribution function is surprisingly insensitive to the net charge as judged from simulations of an electrolyte solution in a sphere (see "reference system" below).

In order to quantify the correlation in a large dataset with several thousands of protein coordinates, it is convenient to introduce a numerical criterion. We have defined the ratio (ξ) as a function of the correlation parameters ξ_{+-} and $\xi_{++,--}$,

$$\xi = \frac{\xi_{+-}}{\xi_{++,--}} = \frac{\int_0^{r_c} P_{+-}(r_{+-}) dr_{+-}}{0.5 \left(\int_0^{r_c} P_{++}(r_{++}) dr_{++} + \int_0^{r_c} P_{--}(r_{--}) dr_{--} \right)} \quad (1)$$

where r_c is assumed to be equal to 10 Å based on previous works. (Barlow & Thornton, 1983; Kumar & Nussinov, 2002) Changing r_c within ± 1 Å will alter the final results by less than 10 %. If $\xi < 1$, the system presents a negative correlation or an anti-correlation. Otherwise ($\xi \geq 1$), the system has a positive charge correlation, i.e. there is a preference to have a positive and a negative charges close to each other. This criterion allows us to analyse any number of protein coordinates including the whole PDB. The mean correlation ratio $\langle \xi \rangle$ is reported in Table 1 for the different protein classes and sets.

In order to reveal the nuances that make each protein unique (in the Biological sense), we have emphasized some particular protein structures randomly picked up with and without charge-charge correlation – see Table 2, where their PDB identities are listed. We have also included molecular modeling structures of α -s1-casein, (Kumosinski *et al.*, 1994) α -s2-casein, (Farrell *et al.*, 2009) β -casein, (Kumosinski *et al.*, 1993b) and κ -casein. (Kumosinski *et al.*, 1993a) Caseins are particularly interesting because of their flexibility and the lack of a well defined tertiary structure. For these systems we have instead used computer mod-

els. (Kumosinski *et al.*, 1994; Farrell *et al.*, 2009; Kumosinski *et al.*, 1993*b*; Kumosinski *et al.*, 1993*a*) The proteins presented in Table 2 were classified accordingly to the calculated ξ from their structural data. When more than one chain was available in the PDB coordinate file, only the first one was used.

Additional complementary information such as the mean number of residues (N_{aa}), the mean protein net charge ($\langle Z_{net} \rangle$), were also given in Table 1 as an indication of the diversity of physical chemistry properties of the proteins in the different datasets. In Table 2, absolute numbers were provided for these quantities together with the molecular weight.

2.2. Reference system

In order to have a reference system for comparison we have also studied an electrolyte solution confined to an electroneutral spherical cell of radius $R_{cell} = 20\text{\AA}$ as a very simple model of the charges in a protein. Additional simulations were run with $R_{cell} = 30\text{\AA}$. The free ions are described by the primitive model. (Hill, 1960) Each mobile ion k with valence z_k is treated explicitly as a hard sphere of radius 1.2 (\AA), while the solvent is a structureless dielectric medium characterized by a relative dielectric permittivity ϵ_r . The system was sampled by Metropolis Monte Carlo (MC) simulations. (Metropolis *et al.*, 1953) Calculations were carried out with 20 ion pairs and different dielectric permittivities ($\epsilon_r = 20, 40, 60$ and 78.7) at room temperature. After the equilibration, at least 10^6 simulation cycles were performed during the production run where the probabilities P_{+-} , P_{++} and P_{--} were sampled. The observations found in this simple system provide a background for the interpretation of the outcomes observed for the protein coordinates available at the PDB.

3. Results and discussion

Let us start with the results from the simple reference model. For an electrolyte consisting of 20 ion pairs in a spherical electroneutral cell, there is a positive charge correlation, *i.e.*

$P_{+-} > P_{++,--}$, at short separation - see Figure 1. The positive correlation is rather weak and limited to separations less than 10 Å for a sphere of 20 Å radius. The weak correlation can be understood from the high dielectric permittivity, $\epsilon_r = 78$ used in the simulation. The probability distribution for random points in the same sphere is shown for comparison. With a reduced dielectric permittivity for the sphere (protein) we see an increased positive charge correlation with a very strong peak at a distance corresponding to hard core contact between two ions. The correlation is so strong that it actually leads to small layering of ions manifested in small maximum in the like charge distribution at 5-6 Å separation – see Figure 1b.

Table 1 compiles the results for the different protein classes. Proteins with both positive and negative charge correlations can be found in all classes. Independent of structure and biological function most of the proteins (87%) of the whole PDB present positive correlation. Some protein classes show stronger positive correlation (e.g. immunoglobins, oxidoreductases, transferases, transport proteins and virus) than the average behavior observed for the whole PDB. Calcium binding proteins are the only class that on *average* present a negative charge correlation. This is probably due to the need of these proteins to have several negative charges defining the binding sites as it is observed for the two calcium binding sites of the calbindin D_{9K} . (Kesvatera *et al.*, 1996)

The same pattern observed for the PDB ($\Gamma_{+-} = 87\%$) structures is also seen for the protein structures ids generated by the PISCES (Wang & Dunbrack, 2003) server ($\Gamma_{+-} = 83$ and 87 %) with structures determined by x-ray. The results obtained by protein coordinates with resolution smaller than 1.0Å (see PISCES set 1 in Table 1) are virtually the same obtained by a less refined resolution ($< 3.0\text{Å}$). In terms of the SCOP structural classification, (Hubbard *et al.*, 1997) there is essentially no difference between the datasets with only α ($\Gamma_{+-} = 91\%$) or with only β structures ($\Gamma_{+-} = 93\%$). Structures alpha and beta (a/b) follow the same trend ($\Gamma_{+-} = 92.9\%$) while the structures alpha and beta (a+b) have a slightly smaller probability to show positive and negative charged aminoacids close to each other ($\Gamma_{+-} = 84.2\%$).

Apparently, NMR structures ($\Gamma_{+-} = 69.3\%$) have a tendency to show more structures with negative correlation than the ones obtained by x-ray, probably because the NMR structures are usually smaller than the x-ray ones. This can be seen comparing the PISCES with the NMR datasets. However, similar percentage was observed by the coagulation protein group ($\Gamma_{+-} = 65\%$). Clustering the protein structures from all the PDB accordingly to their net charge reveals Γ_{+-} in the interval of 67 to 95 %. Positively charged proteins have a slight tendency to contain more charged patches, which means smaller values observed for Γ_{+-} . Structures obtained from molecular dynamics simulation trajectories and homology modeling also show the same characteristics. (Brasil *et al.*, 2013)

The mean values obtained from the full set of PDB structures gives $\Gamma_{+-} = 87\%$, but this percentage have wide spread numbers in the range from 33 to 95 depending on the criterion used to build-up the dataset. Therefore, selecting a single protein or a small class of proteins could mean that Γ_{+-} turns out to be less than 50%. However, the proteins used by Vicatos and co-authors (Vicatos *et al.*, 2009) seem to contain enough heterogeneity to reflect the mean value larger than 50%. Looking upon the all PDB structures, we find that 13% of the proteins, or surprisingly, 22, 432 protein chains, show a negative charge-charge correlation. As it is neither our aim nor is possible to cover functional aspects of such high number of protein coordinates in the present manuscript, the list of these PDB entries in a table format is provided as a supplementary material. Nevertheless, it is interesting to note that analysing a non-redundant and much smaller set of 191 enzymes, Gutteridge & Thornton could observe the relation between two like charges residues and the biological function. They have described that in enzyme active sites identical charge residues are often observed and important for the catalytic function. (Gutteridge & Thornton, 2005)

A plot of the $P_{+-}(R)$ and $P_{++,--}(R)$ for all the protein structures available at the PDB is shown in Figure 2. It can be observed that the positive correlation is moderate and only observed for at short separation ($R < 8 \text{ \AA}$). Being an averaged result, of course, this plot

hides systems with more distinct charge-charge correlations. Yet, no single protein structure presented the very strong peak observed in the reference system with a low dielectric response - see Figure 1. This observation might explain why it is important to adopt high dielectric constants (35 – 40) in order to describe electrostatic interactions in proteins. (Vicatos *et al.*, 2009) Other investigations have suggested a similar picture. (Warwicker, 1999; Dominy *et al.*, 2004)

Table 2 compiles two sets of individual proteins – set A with positive charge correlation and set B with negative charge correlation. In the dataset reported in Table 2, $\Gamma_{+-} = 72.0\%$. Out of the 25 structures studied in detail, 18 of them belong to set A using the ξ criterion. Again, we did not find any relation between the presence of positive charge correlations and the biological function and physical chemistry properties such as molecular weight or size. Both groups contain, for example, milk and prion proteins. Despite the fact that lysozyme and α -lactalbumin have a common genetic origin and similar size, (McKenzie & Whitejr, 1991) they appear in different groups.

Figure 3 shows the probability distribution of charged residues in α -lactalbumin (*bos Taurus*). Surprisingly enough, there is a strong negative correlation extending up to about 20 Å separation. This is probably a consequence of the existence of "charged patches" reported before. (de Vries *et al.*, 2003; de Vries, 2004) The inclusion of the Calcium ions in the analysis has virtually no effect. The data labeled "with Ca²⁺" was obtained including the pair distances between charged groups on the protein *and* between the charged groups on the protein and the charged heteroatom (Calcium, in this case). For a given protein coordinates file, the location of the protein charged groups will not change if one includes or does not include the heteroatoms, when using the static PDB data as we do here. All the charges will remain at the same initial positions and have the same contribution in geometrical analysis. Considering the number of charges in a protein coordinate, a simple mathematical reasoning can show that the absence of one charge will have a rather minor effect on the final outcomes as observed in

this Figure. α -lactalbumin (PDB id 1f6s) contains 48 possible charged groups (7 glu, 13 asp, 12 lys, 1 arg, 3 his, 8 cys and 4 tyr) depending on the solution pH. (Harris & Turner, 202) At neutral pH, cysteines and tyrosines are excluded which gives 630 pair distances evaluations (without Calcium) in our analysis. Including Calcium, the number of pairs distances increases to only 6% more.

Another milk protein, κ -casein, shows the same behaviour with an extension of the negative correlation up to almost 30 Å – see Fig. 4. In the case of α -lactalbumin we have used coordinates from an x-ray determination, (Chrysina *et al.*, 2000) while for κ -casein protein coordinates were taken from the computational study by Kumosinski et al. (Kumosinski *et al.*, 1993a). Another α -lactalbumin (pdb ID 1HFY) from a different organism (*capra Hircus*) behaves as the α -lactalbumin from *Bos Taurus* (1F6S). A similar pattern is found for other proteins like α -s1-casein, α -s2-casein, trypsin as well as a prion.

In these comparisons we have averaged the two distribution functions, $P_{++}(R)$ and $P_{--}(R)$. A more thorough analysis reveals that $P_{--}(R) > P_{++}(R)$ at short separations. A possible interpretation is that glutamic and aspartic acid residues often create binding sites for divalent cations, while the basic residues more often are found on the surface of the protein.

Some proteins do show a preference for two oppositely charged residues to be close in space forming ion-pairs as observed in previous studies. (Wada & Nakamura, 1981; Barlow & Thornton, 1983) For instance, the human prion protein has clearly this behaviour for short separations as shown in Fig. 5. However, at intermediate separations there is a strong negative correlation. The whey protein, α -lactalbumin, shows a negative charge correlation, while other whey proteins, β -lactoglobulin and lactoferrin (LF) show a weak positive or no charge correlation.

The probability distributions for all proteins from set A and B were pooled and the final result can be seen in Figure 6. In general, the positive correlation is usually small and present only in short separations. There is an inversion of the behaviour for medium separations ($10 <$

$R < 20\text{\AA}$). The display interval in Figure 6 has been limited to 30\AA in order to more clearly visualize the differences.

How can we understand these results where some proteins do not have a positive charge correlation? One obvious explanation is that the crystallization of proteins take place under high salt condition, where the electrostatic interactions are fully screened out. For example, the crystallization of β -lactoglobulin took place in 2 M $(\text{NH}_4)_2\text{SO}_4$ solution. In such a solution the Debye-Hückel screening length, κ^{-1} , is about 1\AA and electrostatic interactions are effectively screened out. (Hill, 1960) This explanation can be partly tested by simulating our reference system with 20 ion pairs discussed initially but now interacting via a screened Coulomb potential with $\kappa^{-1} = 1\text{\AA}$,

$$u(r_{ij}^{SC}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \exp(-\kappa r_{ij}) \quad (2)$$

instead of the ordinary Coulomb interaction. The result of such a simulations show no correlation whatsoever between the ions, neither at high nor at low dielectric permittivity of the medium. For these cases with strong electrostatic screening due to the high salt concentration the simulated probability distributions coincide with the random distribution in Figure 1a. Thus, the absence of positive charge correlation or even the appearance of negative correlations should not come as a surprise for x-ray determined structures, since under typical crystallization conditions, the influence of electrostatic interactions is marginal.

Accepting this explanation we should expect a slightly different behaviour for protein structures determined via nmr experiments, since these are normally carried out under physiological salt conditions with less electrostatic screening. Table 1 shows, however, that nmr determined structures normally show weaker charge correlations, 69%, than the average of 87%. Figure 7 shows two examples of nmr determined structures: i) the sugar binding protein in Figure 7a is a fairly big protein with a weak positive charge correlation. Figure 7b with an nmr determined structure of trypsin (PDB id 1C1P), shows a strongly negative charge corre-

lations. The human prion protein structure used in Fig. 5 was also obtained by nmr and shows a positive correlation. Structure determination with nmr is usually performed at rather modest salt conditions and the "screening explanation" may not be valid in this case. However, the protein concentration is usually rather high in nmr studies, which could possibly give rise to a significant screening.

4. Conclusion

We have used a large dataset of protein coordinates from x-ray and nmr structural determinations to calculate the correlation between the charged residues in all proteins available in the PDB. The general result is that *most* (but not all) of the proteins exhibit a positive correlation between oppositely charged residues but then only at short separations between the charged residues. Based on such analysis, we are tempted to conclude that electrostatic interactions play a role in stabilizing the tertiary structure of proteins, but they can also under some conditions be effectively screened. In the latter case it could be that functional aspects are more important than a minimization of the electrostatic free energy.

The crystallization conditions for proteins are usually very harsh with high salt concentrations leading to strong electrostatic screening, which could explain the lack of charge correlation observed in about 13% of the proteins found in the PDB. The present findings might serve useful for a general analysis of the strengths and importance of electrostatic interactions in proteins.

Appendix A SUPPLEMENTARY MATERIAL

Due to the large amount of data, the full list of all PDB ids used in each dataset is available as

supplementary materials.

Acknowledgements This work has been supported in part by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Capes/MCT* (2405/13-0), *Fundação de Amparo à Pesquisa do Estado de São Paulo* (Fapesp 2010/50425-0) and the *Swedish Research Council* through a Linnaeus grant.

References

- Antonsiewicz, J., McCammon, J. A. & Gilson, M. K. (1994). *J. Mol. Biol.* **238**, 415–436.
- Antonsiewicz, J., McCammon, J. A. & Gilson, M. K. (1996). *Biochemistry*, **35**, 7819–7833.
- Archontis, G. & Simonson, T. (2005). *Biophys. J.* **88**, 3888–3904.
- Autreto, P. A. S., Figueiredo, F. V., Nonato, M. C. & da Silva, F. L. B. (2003). *Braz. J. Pharm. Sci. suppl.***2:39**, 203.
- Barlow, D. & Thornton, J. (1983). *Journal of Molecular Biology*, **168**(4), 867–885.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Research*, **28**, 235–242.
- Brasil, C. R., Delbem, A. C. & da Silva, F. L. B. (2013). *J. Comput. Chem.* **34**(20), 1719–1734.
- Calzolari, L. & Zahn, R. (2003). *Journal of Biological Chemistry*, **278**(37), 35592–35596.
- de Carvalho, S. J., Fenley, M. O. & Da Silva, F. L. B. (2008). *The Journal of Physical Chemistry B*, **112**(51), 16766–16776.
- Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. (2003). *Pharmaceutical Research*, **20**(9), 1325–1336.
- Chrysin, E. D., Brew, K. & Acharya, K. R. (2000). *Journal of Biological Chemistry*, **275**(47), 37021–37029.
- Davis, A. M., Teague, S. J. & Kleywegt, G. J. (2003). *Angewandte Chemie International Edition*, **42**(24), 2718–2736.
- Demchuk, E. & Wade, R. C. (1996). *J. Phys. Chem. B*, **100**, 17373–17387.
- Dill, K. A. (1990). *Biochemistry*, **29**(31), 7133–7155.
- Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. (2008). *Annual Review of Biophysics*, **37**(1), 289–316.
- Dominy, B. N., Minoux, H. & Brooks, C. L. (2004). *Proteins*, **57**(1), 128–141.
- Dudev, T. & Lim, C. (2000). *J. Phys. Chem. B*, **104**, 3692–3694.
- Farrell, H. M., Malin, E. L., Brown, E. M. & Mora-Gutierrez, A. (2009). *Journal of Dairy Science*, **92**(4), 1338–1353.
- Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). *Current opinion in structural biology*, **6**(3), 377–385.
- Gutteridge, A. & Thornton, J. M. (2005). *TRENDS in Biochemical Sciences*, **30**, 622–629.
- Gvritshvili, A. G., Gribenko, A. V. & Makhatadze, G. I. (2008). *Protein science*, **17**(7), 1285–1290.
- Harris, T. K. & Turner, G. J. (202). *Life*, **53**, 85–98.
- He, Y., Xu, J. & Pan, X.-M. (2007). *Proteins: Struc., Func., and Bioinformatics*, **69**, 75–82.
- Hildebrandt, A., Blossey, R., Rjasanow, S., Kohlbacher, O. & Lenhof, H.-P. (2007). *Bioinformatics*, **23**(2), e99–e103.
- Hill, T. L. (1955). *Arch. Biochem. Biophys.* **57**, 229–239.
- Hill, T. L. (1960). *An Introduction to Statistical Thermodynamics*. Reading, USA: Addison-Wesley Publishing Company, Inc.
- Holm, L., Käriäinen, S., Rosenström, P. & Schenkel, A. (2008). *Bioinformatics*, **24**(23), 2780–2781.

- Hubbard, T. J., Murzin, A. G., Brenner, S. E. & Chothia, C. (1997). *Nucleic acids research*, **25**(1), 236–239.
- Jefferys, B. R., Kelley, L. A. & Sternberg, M. J. E. (2010). *Journal of Molecular Biology*, **397**(5), 1329–1338.
- Katz, B. A., Clark, J. M., Finer-Moore, J. S., Jenkins, T. E., Johnson, C. R., Ross, M. J., Luong, C., Moore, W. R. & Stroud, R. M. (1998). *Nature*, **391**(6667), 608–612.
- Kesvatera, T., Jönsson, B., Thulin, E. & Linse, S. (1996). *Biochemistry*, **33**, 14170–14176.
- King, G., Lee, F. S. & Warshel, A. (1991). *J. Chem. Phys.* **95**, 4366–4377.
- Ko, J., Murga, L. F., Wei, Y. & Ondrechen, M. J. (2005). *Bioinformatics*, **21**(suppl 1), i258–i265.
- Kukic, P., Farrell, D., McIntosh, L. P., García-Moreno, Jensen, K. S., Toleikis, Z., Teilum, K. & Nielsen, J. E. (2013). *J. Am. Chem. Soc.* **135**(45), 16968–16976.
- Kumar, S. & Nussinov, R. (2002). *Biophysical Journal*, **83**(3), 1595–1612.
- Kumosinski, T. F., Brown, E. M. & Farrell, H. M. (1994). *ACS Symposium Series 576*, pp. 368–390.
- Kumosinski, T. F., Brown, E. M. & Jr, H. M. F. (1993a). *J. Dairy Sci.* **76**, 2507–2520.
- Kumosinski, T. F., Brown, E. M. & Jr, H. M. F. (1993b). *J. Dairy Sci.* **76**, 931–945.
- Löffler, G., Screiber, H. & Steinhauser, O. (1997). *J. Mol. Biol.* **270**, 520–534.
- McKenzie, H. & Whitejr, F. (1991). *Lysozyme and alpha-Lactalbumin: Structure, Function, and Interrelationships*, vol. 41 of *Advances in Protein Chemistry*, pp. 173–315. Elsevier.
- Metropolis, N. A., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1097.
- Miki, Y., Kakuyama, K. & Soda, K. (1997). *Bio Systems*, **44**(1), 69–77.
- Mueller, G. A., Choy, W. Y., Yang, D., Forman-Kay, J. D., Venters, R. A. & Kay, L. E. (2000). *Journal of Molecular Biology*, **300**(1), 197–212.
- Myers, J. K. & Pace, C. N. (1996). *Biophysical journal*, **71**(4), 2033–2039.
- Nishimura, C., Uversky, V. N. & Fink, A. L. (2001). *Biochemistry*, **40**(7), 2113–2128.
- Pace, C. N., Grimsley, G. R. & Scholtz, J. M. (2009). *Journal of Biological Chemistry*, **284**(20), 13285–13289.
- Penfold, R., Warwicker, J. & Jönsson, B. (1998). *J. Phys. Chem. B*, **102**, 8599–8610.
- Perutz, M. F. (1978). *Science*, **201**, 1187–1191.
- Roca, M., Messer, B. & Warshel, A. (2007). *FEBS letters*, **581**(10), 2065–2071.
- Schutz, C. N. & Warshel, A. (2001). *Proteins: Struc., Func., and Genetics*, **44**, 400–417.
- Sham, Y. Y., Chu, Z. T. & Warshel, A. (1997). *J. Phys. Chem. B*, **101**, 4458–4472.
- da Silva, F. L. B., Jönsson, B. & Penfold, R. (2001). *Prot. Sci.* **10**, 1415–1425.
- Simonson, T. & Brooks III, C. L. (1996). *J. Am. Chem. Soc.* **118**, 8452–8458.
- Simonson, T. & Perahia, D. (1995). *J. Am. Chem. Soc.* **117**, 7987–8000.
- Sterpone, F. & Melchionna, S. (2012). *Chem. Soc. Rev.* **41**(5), 1665–1676.
- Tanford, C. & Kirkwood, J. G. (1957). *J. Am. Chem. Soc.* **79**, 5333–5339.
- Teixeira, A. A., Lund, M. & da Silva, F. L. (2010). *Journal of Chemical Theory and Computation*, **6**(10), 3259–3266.
- Thirumalai, D., O'Brien, E. P., Morrison, G. & Hyeon, C. (2010). *Annual Review of Biophysics*, **39**(1), 159–183.
- Varma, S. & Jakobsson, E. (2004). *Biophys. J.* **86**(2), 690–704.
- Vicatos, S., Roca, M. & Warshel, A. (2009). *Proteins*, **77**(3), 670–684.
- de Vos, W. M., Leermakers, F. A., de Keizer, A., Cohen Stuart, M. A. & Kleijn, M. M. (2010). *Langmuir*, **26**(1), 249–259.
- de Vries, R. (2004). *J. Chem. Phys.* **120**(7), 3475–3481.
- de Vries, R. & Stuart, M. C. (2006). *Curr. Opin. Colloid. Interf. Sci.* **11**, 295–301.
- de Vries, R., Weinbreck, F. & de Kruijff, C. G. (2003). *J. Chem. Phys.* **118**(10), 4649–4659.
- Wada, A. & Nakamura, H. (1981). *Nature*, **293**, 757.
- Wang, G. & Dunbrack, R. L. (2003). *Bioinformatics (Oxford, England)*, **19**(12), 1589–1591.
- Warshel, A. (1981). *Acc. Chem. Res.* **14**, 284–290.
- Warshel, A. & Åqvist, J. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 267–298.
- Warwicker, J. (1999). *Prot. Sci.* **8**, 418–425.
- Whitten, S. T. & E., B. G.-M. (2000). *Biochemistry*, **39**, 14292–14304.

Table 1. Number of protein coordinates extracted from PDB (N_{coord}), mean number of residues (N_{aa}), mean protein net charge in units of e ($\langle Z_{\text{net}} \rangle$) and structural datasets. The mean correlation ratio ξ and the percentage Γ_{+-} of proteins in each class that show a positive charge-charge correlation ($\xi \geq 1$) at short separation.

Protein class	$N_{\text{coord}}^{(a)}$	$\langle N_{\text{aa}} \rangle$	$\langle Z_{\text{net}} \rangle$	$\langle \xi \rangle$	Γ_{+-}
Allergen	136	186	-7	1.35	78
Apoptosis	774	164	-4	1.31	81
Calcium binding (apo)	260	156	-13	0.92	33
Calcium binding (halo)	260	156	-2	1.11	53
Coagulation	130	186	-7	1.07	65
Hormone	1,478	134	-3	1.42	73
Immunoglobulin	512	188	-6	1.46	95
Membrane	2,563	193	-9	1.44	84
Oxidoreductase	9,662	322	-8	1.38	94
Photosynthesis	707	207	-5	1.40	80
RNA binding	1,442	149	-1	1.50	75
Signaling	3,903	185	-5	1.32	84
Toxin	2,063	173	-6	1.35	82
Transferase	20,824	306	-8	1.34	94
Transport ^(b)	9,944	200	-6	1.37	87
Virus	1,259	327	-10	1.44	95
<hr/>					
PISCES set 1 ^(c)	223	179	-5	1.42	83
PISCES set 2 ^(d)	16,893	216	-7	1.38	87
NMR ^(e)	1,631	94	-1	1.33	69
Only alpha ^(f)	98	176	-1	1.40	91
Only beta ^(g)	134	163	-2	1.44	92
Vicatos (Ref. (Vicatos <i>et al.</i> , 2009))	41 ^(h)	128	-4	1.38	81
<hr/>					
All PDB ⁽ⁱ⁾	172,622	244	-7	1.35	87
Only chain A from "All PDB" set	68,912	258	-7	1.35	89
A non-redundant set ^(j)	11,356	187	-5	1.38	83
A selected non-redundant set ^(l)	3,816	135	-3	1.40	77

^(a) All available chains in the PDB file where considered as individual protein structures and used here.

^(b) Including oxygen, proton, electron, lipid, metal, retinol and sugar transportation.

^(c) Dataset generated by the PISCES server (<http://dunbrack.fccc.edu/PISCES.php>) with the following parameters: a) sequence percentage identity: $\leq 25\%$, b) resolution: $0.0 \sim 1.0 \text{ \AA}$, c) R-factor: 0.3, d) sequence length: $40 \sim 10000$, e) non X-ray entries: excluded, g) CA-only entries: excluded, and g) cull PDB by chain. Only 185 PDB's IDs were used.

^(d) The same as ^(c) with a less refined resolution ($0.0 \sim 3.0 \text{ \AA}$). 7, 638 PDB's IDs were used.

^(e) Structures selected via the PDB server (<http://www.pdb.org>) with the criterion: a) experimental method is solution NMR, b) there is a protein chain but not any DNA or RNA or hybrid, and c) taxonomy is just Homo sapiens and Eukaryota.

^(f) The same as ^(e) with the experimental method criterion changed to X-Ray, and the addition of a) the resolution is 1.499 \AA or less, and b) the protein class is only alpha (as classified by SCOP (Hubbard *et al.*, 1997)).

^(g) The same as ^(f) but here the protein class is only beta.

^(h) The original benchmark dataset considered in ref. (Vicatos *et al.*, 2009) contains 45 protein structures. Both *Staphylococcal Nuclease* (PDB id 1TR5) and *Ubiquitin* (PDB id 1AAR) appeared in three different mutant forms. Here, we have considered only the original wild type structure as given by the PDB file.

⁽ⁱ⁾ 72, 966 PDB's IDs were used.

^(j) A non-redundant PDB chain set provided by the VAST (Vector Alignment Search Tool) server (<http://structure.ncbi.nlm.nih.gov/Structure/VAST/nrpdb.html>) with the non-redundancy level equals to 10^{-7} .

^(l) From the non-redundant set below, we have selected protein coordinates that do not contain heteroatoms (except water).

Table 2. PDB entries, protein net charge in units of e , Z_{net} and molecular weight, M_w , for some specific proteins investigated here. ^(a) These are grouped into two sets, A and B, where the former shows a positive charge correlation at short separation, while the proteins in set B show no correlation or negative correlation between the charged amino acids.

Protein	PDB entry	Z_{net}	M_w (kDa)
acetylcholinesterase	1DX6	-13	61.3
l-arabinose-binding protein	1ABE	-7	33.3
β -casein	Ref. (Kumosinski <i>et al.</i> , 1993b)	-6	24.0
hemoglobin A (deoxy, α chain) ^(b)	1HBB	+9	15.2
lactoferrin	1BLF	+2	76.2
β -lactoglobulin	1BEB	-11	18.3
apo-liver alcohol dehydrogenase	8ADH	+7	39.8
lysozyme	2LZT	+6	14.3
major prion protein precursor	1HJM	-10	12.6
maltose-binding periplasmic protein	1EZO	-20	40.7
papain	9PAP	-9	23.5
pectin methylesterase	1GQ8	-1	34.2
ribonuclease T1	1RGC	-15	11.1
serum albumin	1AO6	-16	66.6
small S protein (prion in amyloid form)	2RNM	+7	8.7
thermolysin	8TLN	-32	34.4
titin (A77-A78 domain)	3LPW	-6	21.7
trypsin inhibitor	4PTI	+2	6.5
α -s1-casein	Ref. (Kumosinski <i>et al.</i> , 1994)	-15	23.0
α -s2-casein	Ref. (Farrell <i>et al.</i> , 2009)	-18	25.0
κ -casein	Ref. (Kumosinski <i>et al.</i> , 1993a)	-9	19.0
α -lactalbumin (<i>bos Taurus</i>)	1F6S	-8	14.2
α -lactalbumin (<i>capra Hircus</i>)	1HFY	-7	14.2
small S protein (prion)	2KJ3	+7	8.7
trypsin	1C1P	-1	23.3

^(a) The molecular weight of the four caseins was estimated from their chemical components.

^(b) The other chains of hemoglobin (B, C and D) also show positive correlation.

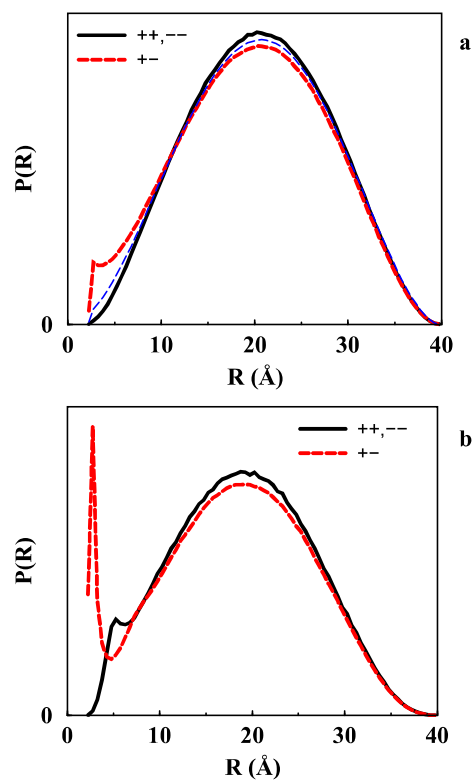


Fig. 1. The probability distribution of negative and positive charges confined in a sphere of radius 20 Å. The ions are modelled as charged hard charged spheres with a radius of 1.2 Å. The curves are obtained from a Metropolis Monte Carlo simulation (Metropolis *et al.*, 1953) within the primitive model. (Hill, 1960) The sphere dielectric permittivity is a) 78 and b) 20. The blue dot-dashed line in a) shows the random distribution of positive and negative charges in a sphere.

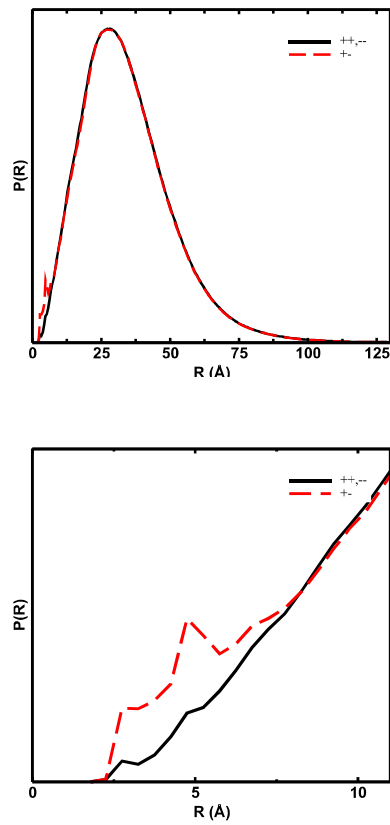


Fig. 2. The distribution of positive and negative charges for all the 172,622 protein chains available at the PDB. a) all separations (top panel) and b) short separations (bottom panel).

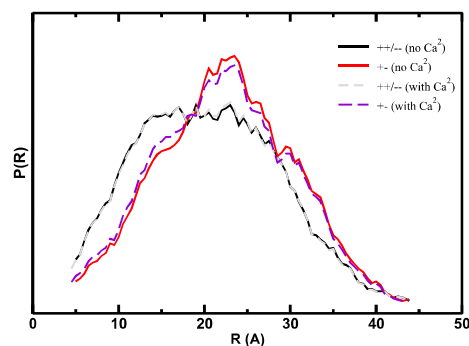


Fig. 3. The distribution of positive and negative residues in α -lactalbumin (bos taurus) in both halo and apo forms at approximate neutral pH. Coordinates taken from PDB entry 1F6S. (Chrysina *et al.*, 2000)

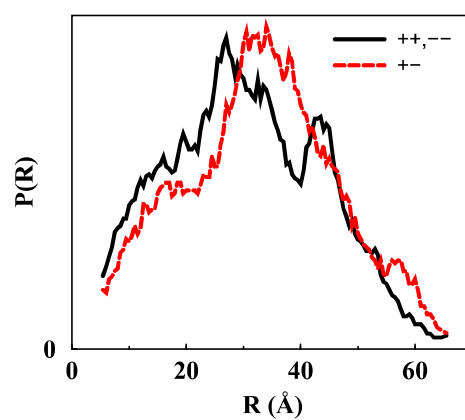


Fig. 4. The distribution of positive and negative residues in κ -casein at approximate neutral pH. Coordinates taken from the computational study by Kumosinski *et al.* (Kumosinski *et al.*, 1993a).

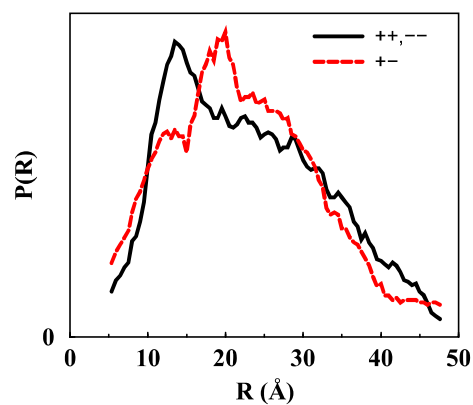


Fig. 5. The distribution of positive and negative residues in human prion protein at approximate neutral pH. Coordinates taken from PDB entry 1HJM. (Calzolari & Zahn, 2003)

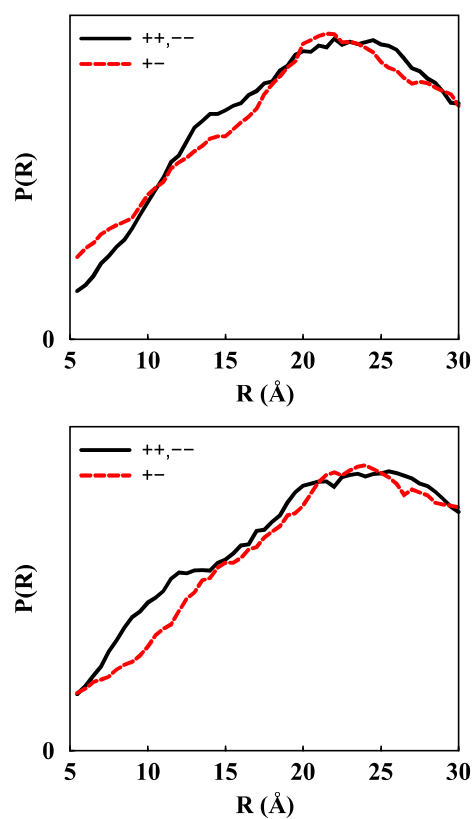


Fig. 6. a) The distribution of positive and negative residues in proteins from set A at approximate neutral pH. Coordinates taken from PDB entry as listed on Table 2 (top panel) b) Same as a) but for charged residues from set B (bottom panel).

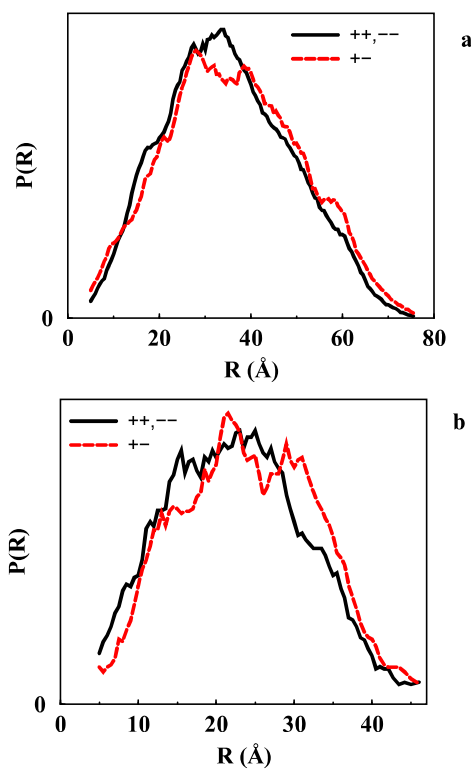


Fig. 7. a) The distribution of positive and negative residues in the Sugar Binding Protein (top panel). Coordinates taken from PDB entries 1EZO (Mueller *et al.*, 2000). b) The distribution of positive and negative residues in trypsin (bottom panel). Coordinates taken from PDB entries 1C1P (Katz *et al.*, 1998).

Synopsis

A large-scale geometrical analysis on protein structure coordinates were performed. Surprisingly, about 13% of the proteins found in the Protein Data Bank show a *negative* charge-charge correlation.
