UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

Vinícius Martins de Oliveira

Efeito das Interações Eletrostáticas e de pH no Enovelamento e Estabilidade de Proteínas

SÃO JOSÉ DO RIO PRETO - SÃO PAULO
2017

Vinícius Martins de Oliveira

Efeito das Interações Eletrostáticas e de pH no Enovelamento e Estabilidade de Proteínas

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de São José do Rio Preto.

Financiadora: CNPq - Proc. 141985/2013-5

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

São José do Rio Preto, 06 de Julho de 2017

Oliveira, Vinícius Martins de.
Efeito das interações eletrostáticas e de pH no enovelamento e estabilidade de proteínas / Vinícius Martins de Oliveira. -- São José do Rio Preto, 2017
135 f. : il., grafs., tabs.

Orientador: Vitor Barbanti Pereira Leite
Tese (doutorado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Biologia molecular. 2. Biofísica. 3. Proteínas - Estrutura. 4. Dinâmica molecular. 5. Proteínas – Enovelamento. 6. Mutações. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. II. Título.
CDU – 577-112

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Campus de São José do Rio Preto

Vinícius Martins de Oliveira

Efeito das Interações Eletrostáticas e de pH no Enovelamento e Estabilidade de Proteínas

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de São José do Rio Preto.

Financiadora: CNPq - Proc. 141985/2013-5

BANCA EXAMINADORA

Prof. Dr. Vitor Barbanti Pereira Leite
UNESP – São José do Rio Preto
Orientador

Prof. Dr. Jorge Chahine
UNESP – São José do Rio Preto

Prof. Dr. Luis Gustavo Dias
USP – Ribeirão Preto

Prof. Dr. Nelson Augusto Alves
USP – Ribeirão Preto

Prof. Dr. Pedro Geraldo Pascutti
UNESP – São José do Rio Preto

São José do Rio Preto, 06 de Julho de 2017

Dedico este trabalho à minha família.

Agradecimentos

Agradeço primeiramente a Deus, e a toda à minha família. Principalmente à minha mãe e minhas irmãs que sempre me ajudaram e apoiaram, e tenho certeza que não teria chegado até aqui sem elas. Agradeço também a todos os professores e funcionários do Departamento de Física do IBILCE, em especial ao Professor Vitor, que me orientou durante esses quatro anos de Doutorado com muita atenção e paciência, e quem tornou possível a realização deste trabalho.

À minha noiva Manu que esteve do meu lado nos momentos difíceis e que sempre me incentivou a continuar a estudar nesses quase sete anos em que estamos juntos. Com ela também aprendi a jogar basquete (mais ou menos) e a caçar morcegos. Agradeço muito aos meus amigos da turma de física biológica de 2007. Em especial ao Jesus, Daniel, Davi, Guilherme, Rodolfo e Tibúrcio que tornaram este curso nada fácil em algo bem mais tranquilo e, por vezes, divertido.

Queria agradecer também aos alunos e professores envolvidos no grupo do pão quentinho: professor Chahine, Ronaldo que sempre estiveram dispostos e abertos a nos ajudar desde a graduação. Aos colegas de pesquisa Vini, Tambonis, Fernando, Paulo e Bibi, os quais se tornaram bons amigos durante esses anos de doutoramento.

Ao apoio financeiro do CNPq e ao suporte computacional fornecido pelo GRIDUnesp.

"O cientista não é o homem que fornece as verdadeiras respostas; é quem faz as verdadeiras perguntas".

Claude Lévi-Strauss

Resumo

As interações eletrostáticas desempenham um papel crucial no enovelamento de proteínas. Além disso, essas interações entre resíduos carregados são fundamentais para a estabilidade proteica e para que as proteínas realizem suas funções biológicas. A distribuição de carga nessas macromoléculas é definida pelos tipos de resíduos que as compõem e pelas condições do ambiente a que elas estão sujeitas, como, por exemplo, concentração de sal e pH. Compreender como esses aspectos podem governar a estabilidade e o enovelamento de proteínas é de fundamental importância no desenvolvimento racional de aplicações biomédicas e biotecnológicas. No presente estudo, utilizou-se modelos computacionais simplificados visando investigar esses efeitos eletrostáticos nos processos de enovelamento, na estabilidade térmica e na presença e ausência de mutações do seguinte conjunto de proteínas: o domínio N-terminal da proteína ribossomal L9 (NTL9); a variante da proteína G (1PGB-QDD), a proteína *Cold Shock* A (CspA); e a proteína 5 do ácaro *Blomia tropicalis* (Blo t 5). A partir desses modelos computacionais, buscou-se avaliar os efeitos causados pelas mutações, pelo pH e pela concentração de sal na dinâmica e na estabilidade proteica. Os resultados obtidos via simulações foram comparados com dados experimentais e obtiveram uma excelente concordância, tendo em vista a simplicidade dos modelos utilizados. Além disso, os resultados computacionais permitiram obter uma enorme gama de informações sobre o enovelamento, as mudanças conformacionais e as mutações para essas proteínas em diferentes condições da solução.

Palavras-chave: Enovelamento de proteínas, mutações, dinâmica molecular, interações eletrostáticas, pH.

Abstract

Electrostatic interactions play a crucial role in protein folding. In addition, these interactions between charged residues are fundamental for protein stability and for proteins to perform their biological functions. The charge distribution in these macromolecules is defined by the types of residues composing them and by the conditions of the environment to which they are subject, such as salt concentration and pH. Understanding how these aspects can govern protein folding and stability is of fundamental importance in the rational development of biomedical and biotechnological applications. In the present study, simplified computational models were used to investigate these electrostatic effects in the folding process, in the thermal stability and in the presence or absence of mutations for a set of proteins. Using these models, it was possible evaluate the effects caused by these mutations, pH and salt concentration on protein dynamics and stability. The results obtained through simulations were compared with experimental data and it was obtained an excellent agreement, considering the simplicity of the models used. Finally, the computational results allowed to obtain a huge range of information about the folding, the conformational changes and the mutations in these proteins in different conditions of the solution.

Keywords: Protein folding, mutations, molecular dynamics, electrostatic interactions, pH.

Lista de Figuras

Lista de Tabelas

Sumário

Introdução Geral

O enovelamento é um processo biofísico que permite às proteínas alcançarem seu estado nativo e, desse modo, estarem aptas à realizar suas funções biológicas. O processo de enovelamento é complexo e, se realizado de forma errada, é potencialmente perigoso. Sabe-se que falhas no enovelamento são as causas primárias de doenças como Alzheimer, Parkinson, Huntington, entre outras *misfolding deseases* [1–3]. Apesar do enovelamento ser alvo de estudos há mais de meio século, ainda não existe um consenso geral da maneira em que as proteínas se enovelam [4]. Sendo assim, o presente estudo tem como principal objetivo analisar e quantificar o efeito das interações eletrostáticas e do pH no enovelamento de proteínas via simulação computacional.

Técnicas experimentais em biologia estrutural, tais como cristalografia de raio-X e ressonância magnética nuclear (NMR), foram responsáveis por fornecer configurações funcionais em alta resolução de uma enorme gama de proteínas. A resolução de estruturas proteicas nativas teve, e ainda tem, um papel fundamental para a compreensão do enovelamento e da função dessas biomoléculas [4, 5]. No entanto, o conceito de um mecanismo rígido que liga apenas a estrutura estática com a função da proteína, como "chave-fechadura", tem se mostrado inadequado ao não levar em conta a dinâmica proteica. A dinâmica que leva a proteína de seu estado desenovelado ao seu estado nativo e vice-versa, e que também permite que essa biomolécula acesse múltiplos estados conformacionais são complementares aos dados estruturais. Logo, a combinaçao dos dados estruturais e dinâmicos fornece as informações necessárias para o entendimento do enovelamento e da função da proteína [6]. Tais informações dinâmicas tem sido obtidas graças a recentes avanços em técnicas experimentais, tais como a ressonância de transferência de energia Föster (FRET) e a espectroscopia de *stopped flow* [7, 8]. Apesar da contribuição dessas técnicas no estudo do enovelamento, por vezes existe uma dificuldade de combinar a baixa resolução obtida nos resultados experimentais de dinâmica com a alta resolução estrutural. A simulação computacional em um nível de resolução atômico consegue ser a ponte entre os dados estruturais e dinâmicos.

As simulações de dinâmica molecular são ferramentas indispensáveis em bi-

ofísica molecular há mais de 30 anos. Tais simulações tratam a molécula como uma coleção de partículas clássicas interagindo através de uma função de energia potencial chamado campo de força [9]. As dinâmicas das moléculas são propagadas através do tempo pela integração numérica das equações de Hamilton, resultando em uma trajetória. A partir dessas trajetórias moleculares é possível obter informações termodinâmicas e cinéticas da proteína de interesse. Com o objetivo de integrar as simulações com dados experimentais, foram gerados modelos biomoleculares com uma função de energia potencial robusta e capaz de acessar longas escalas temporais. Esses modelos baseados em estruturas (SBM) tem como base teórica a teoria de superfície de energia e o princípio de mínima frustração [10].

Teoria da superfície de energia e princípio de mínima frustração

O enovelamento de proteínas é um processo auto-organizado por onde a proteína transiciona entre um *ensemble* altamente desorganizado (desenovelado) para um *ensemble* estruturado (enovelado ou estado nativo). Esta transição ocorre em uma escala temporal relativamente curta, implicando que a relação entre qualquer energia de estados não-nativos é pequena em relação a energia do estado enovelado (mínimo da energia livre) [11]. Além disso, pelo príncipio de mínima frustração dos estados, tem-se que a evolução levou a seleção de sequências de aminoácidos otimizadas, os quais interagem no estado nativo de forma minimamente frustrada, ou, maximamente consistente. Sendo assim, esta organização leva a superfície de energia livre do enovelamento de proteínas para uma forma afunilada, onde, em média, a energia diminui para estruturas que se aproximam da estrutura nativa.

Os modelos baseados em estruturas tem por base teórica o princípio de mínima frustração e a superfície de energia afunilada. A utilização desses modelos, também conhecidos como Modelos G$\bar{o}$ [12], permite simplificar não só o potencial de interação dos componentes da proteína, mas também os próprios componentes (ver em detalhes em *Methods*, Capítulo 1). Essas simplificações geram uma enorme economia de tempo computacional. No entanto, esse tipo de modelagem simplificada leva em conta somente as interações que levam a proteína ao seu estado nativo (interações nativas), deixando de fora quaisquer propriedades físico-químicas dos resíduos [13]. Essas propriedades produzem efeitos que não são tratados explicitamente em modelos G$\bar{o}$, como efeitos de solvatação, de hidrofobicidade, carga-carga, etc. Essas interações não-nativas são, por vezes, fundamentais para o entendimento do enovelamento, estabilidade, transições conformacionais e formação de estados intermediários em proteínas [13, 14]. Particularmente, esse trabalho tem como foco o estudo das interações

não-nativas eletrostáticas.

Interações eletrostáticas e pH

As interações eletrostáticas em proteínas tem sido estudadas por uma grande quantidade de trabalhos experimentais e teóricos [15–17]. Esses trabalhos evidenciam a importância de resíduos carregados no enovelamento, estabilidade e função das proteínas, principalmente quando esses resíduos se encontram no sítio ativo dessas biomoléculas [18–21]. Os resíduos ionizáveis podem sofrer alterações no seu estado de protonação não só devido a fatores ambientais como pH e força iônica, mas também devido à propria distância entre os grupos ionizáveis. Neste último caso, o processo de enovelamento e transição conformacional provocam variações drásticas nas distâncias entre os resíduos, tendo um grande efeito no estado de protonação deles. A maioria dos trabalhos computacionais nesta área trata a carga dos resíduos de forma fixa durante toda a simulação [16, 18, 22]. Devido a isso, efeitos como regulação de carga e a dependência do pH e da concentração de sal não são propriamente levados em conta.

Com o intuito de verificar esses efeitos, um modelo de simulação a pH constante (CpHMD) foi utilizado para um conjunto de proteínas que serão detalhadas nos próximos capítulos. Com esse método de simulação e com o cálculo das interações eletrostáticas via TKSA-MC foi possível obter concordância com uma série de resultados experimentais e obter uma enorme gama de informações sobre o enovelamento, as mudanças conformacionais e as mutações para essas proteínas em diferentes condições de pH e concentração de sal.

Objetivos

O objetivo principal deste trabalho é desenvolver e aplicar modelos computacionais simplificados que permitam analisar os efeitos das interações eletrostáticas e do pH no enovelamento, nas mudanças conformacionais e na estabilidade de proteínas. As questões específicas estudadas nesta tese foram divididas nas seguintes etapas:

1. **Desenvolvimento de modelos:** Desenvolver e testar dois modelos computacionais para os estudos realizados neste trabalho (ambos serão explicados em detalhes nos próximos capítulos).

    - Modelo TKSA-MC (*Tanford Kirkwood Solvent Acessibility - Monte Carlo*) calcula as interações eletrostáticas de cada resíduo da proteína individualmente, porém é um modelo estático e apenas fornece informações da proteína em seu estado enovelado.
    - CpHMD (*Constant pH Molecular Dynamics*) possibilita a obtenção de informações dinâmicas das proteínas a pH constante. Nesse modelo é possível analisar todos os estágios do enovelamento.

2. **Teste e aplicação dos modelos:**

    - Capítulo 1: Proteína NTL9 - Utilizada para testar o CpHMD a partir de comparações com experimentos.
    - Capítulo 2: Variante da proteína G - Utilização de TKSA-MC e do CpHMD para estudo da estabilidade sob diferentes condições ambientais.
    - Capítulo 3: Proteína cspA - Análise de estabilidade térmica via CpHMD comparada com experimentos.
    - Capítulo 4: Proteína Blo t 5 - Estudo da interação eletrostática proteína-ligante.

Referências Bibliográficas

[1] Monica Bucciantini, Elisa Giannoni, Fabrizio Chiti, Fabiana Baroni, Lucia Formigli, Jesús Zurdo, Niccolo Taddei, Giampietro Ramponi, Christopher M Dobson, and Massimo Stefani. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416(6880):507–511, 2002.

[2] Massimo Stefani and Christopher M Dobson. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of molecular medicine*, 81(11):678–699, 2003.

[3] Christopher M Dobson. Protein-misfolding diseases: Getting out of shape. *Nature*, 418(6899):729–730, 2002.

[4] S Walter Englander and Leland Mayne. The nature of protein folding pathways. *Proceedings of the National Academy of Sciences*, 111(45):15873–15880, 2014.

[5] Alan R Fersht and Valerie Daggett. Protein folding and unfolding at atomic resolution. *Cell*, 108(4):573–582, 2002.

[6] Jeffrey K Noel and José N Onuchic. The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules. In *Computational modeling of biological systems*, pages 31–54. Springer, 2012.

[7] Hoi Sung Chung, Kevin McHale, John M Louis, and William A Eaton. Single-molecule fluorescence experiments determine protein folding transition path times. *Science*, 335(6071):981–984, 2012.

[8] Robert L Baldwin and George D Rose. Molten globules, entropy-driven conformational change and protein folding. *Current opinion in structural biology*, 23(1):4–10, 2013.

[9] Stewart A Adcock and J Andrew McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5):1589–1615, 2006.

[10] J.D. Bryngelson and P.G. Wolynes. Spin-glasses and the statistical mechanics of protein folding. 84:7524–7528, 1987.

[11] Joseph D Bryngelson and Peter G Wolynes. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *The Journal of Physical Chemistry*, 93(19):6902–6915, 1989.

[12] Y. Ueda, H. Taketomi, and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. 7:445–459, 1975.

[13] Zhuqing Zhang and Hue Sun Chan. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proceedings of the National Academy of Sciences*, 107(7):2920–2925, 2010.

[14] Arash Zarrine-Afsar, Stefan Wallin, A Mirela Neculai, Philipp Neudecker, P Lynne Howell, Alan R Davidson, and Hue Sun Chan. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proceedings of the National Academy of Sciences*, 105(29):9999–10004, 2008.

[15] Jae-Hyun Cho, Satoshi Sato, Jia-Cherng Horng, Burcu Anil, and Daniel P. Raleigh. Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch. Biochem. Biophys.*, 469(1):20–28, January 2008.

[16] Ariel Azia and Yaakov Levy. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.*, 393(2):527–542, October 2009.

[17] Arash Zarrine-Afsar, Zhuqing Zhang, Katrina L. Schweiker, George I. Makhatadze, Alan R. Davidson, and Hue Sun Chan. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins: Struct., Funct., Bioinf.*, 80(3):858–870, 2012.

[18] Ohad Givaty and Yaakov Levy. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, 385(4):1087–1097, January 2009.

[19] Alexey V. Gribenko, Mayank M. Patel, Jiajing Liu, Scott A. McCallum, Chunyu Wang, and George I. Makhatadze. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc. Natl. Acad. Sci. USA*, 106(8):2601–2606, February 2009.

[20] Ji Guo Su, Wei Zu Chen, and Cun Xin Wang. Role of electrostatic interactions for the stability and folding behavior of cold shock protein. *Proteins: Struct., Funct., Bioinf.*, 78(9):2157–2169, July 2010.

[21] Debabani Ganguly, Weihong Zhang, and Jianhan Chen. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. *PLoS Comput. Biol.*, 9(11):e1003363, November 2013.

[22] Swarnendu Tripathi, Angel E. Garcìa, and George I. Makhatadze. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge–Charge Interactions. *J. Phys. Chem. B*, 119(41):13103–13112, October 2015.

Capítulo 1

Domínio N-terminal da proteína ribossomal L9 - NTL9

NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH-Dependence

Vinícius G. Contessoto†, Vinícius M. de Oliveira†, Sidney J. de Carvalho, Leandro C. Oliveira, Vitor B.P. Leite

†These authors contributed equally to this work

*This article has been published in Journal of Chemical Theory and Computation. Vol. 12, pages 3270 - 3277, 2016. DOI: 10.1021/acs.jctc.6b00399

Abstract

The folding process of the N-terminal domain of ribosomal protein L9 (NTL9) was investigated at constant pH computer simulations. Evaluation of the role of electrostatic interaction during folding was carried out by including a Debye-Huckel potential into a C_α structure-based model (SBM). In this study, the charges of the ionizable residues and the electrostatic potential are susceptible to the solution conditions, such as pH and ionic strength, as well as to the presence of charged groups. Simulations were performed under different pHs and the results were validated by comparing them with experimental values of pKa and with denaturation experiment data. Also, the free energy profiles, Φ-values and folding routes were calculated for each condition. It was shown how charges vary along the folding under different pH, which is subject to different scenarios. This study reveals how simplified models can capture essential physical features, reproducing experimental results, and presenting the role of electrostatic interactions before, during and after the transition state.

Keywords: Constant pH simulations, NTL9, Φ-value.

1.1 Introduction

Electrostatic interactions during the protein folding process have been extensively investigated by experiments and theory [1–3]. Charged residues may also play an important role in protein stability and function, especially when they are related to binding site recognition [4–8]. The charge distribution in a protein is defined by its residue type and environmental conditions. Some aspects that may alter the protonation state of a ionizable residue are enumerated here: (*i*) the solvent-accessible surface area of these residues; (*ii*) the distance between these charged groups; (*iii*) the ionic strength and the pH variations, which are responsible for modifications in environment characteristics [3, 9, 10]. Of all these aspects, the variation in solution pH is usually the one that most affects the degree of ionization of the residues.

pH variation effects can be observed in enzymes that have optimal enzymatic activity in a restricted range of pH [11, 12]. pH-regulation has been exhaustively studied experimentally [9, 10, 13, 14], but there are few computational studies on pH dependence in protein folding [15, 16]. In most computational works on electrostatic interactions in proteins, the charges are held fixed throughout the simulation [2, 5, 17], and effects, like charge regulation and pH dependence can not be properly taken into account during the folding process. These effects can be mapped in constant pH simulations [18], where specific residues may be protonated or deprotonated based on a given pH value. The constant-pH molecular dynamic method (CpHMD) are employed in this work in order to understand how pH modulates folding mechanisms. Thus, a Debye-Huckel [19, 20] potential was included into a C_α Structure-Based Model (SBM) [21–24] with titration of ionizable residues. The simulations were performed under different salt concentrations and pHs, in a similar range to the experimental ones [25].

The chosen protein in this study was the N-terminal domain from the ribosomal protein L9 - NTL9 [26]. NTL9 is a monomeric alpha beta [27] protein that contains 56 residues forming a three-stranded antiparallel beta sheet sandwiched between two alpha helices [28] (Figure 1.1). Its folding mechanism is a two-state pathway [29] and it does not interact with any cofactors to fold or contain disulfide bridges [30]. Experimental results showed an important non-native electrostatic interaction in the unfolded state of the NTL9 [31, 32]. The protein folds in a pH range from 1.0 to 12.0 and its folding is fully reversible [25]. These characteristics make the NTL9 a good system for studying pH-dependence and verifying the consistency of the model compared to experimental results. It is worthwhile noting that there are CpHMD studies in NTL9 protein, but they use different approaches with more complex potential models [33, 34]. In the present study, CpHMD in C_α SBM results are compared with experi-

mental ones. The pka values from other computational studies are also compared with the CpHMD+C_α SBM [33, 34] (Support Information - Table S1).

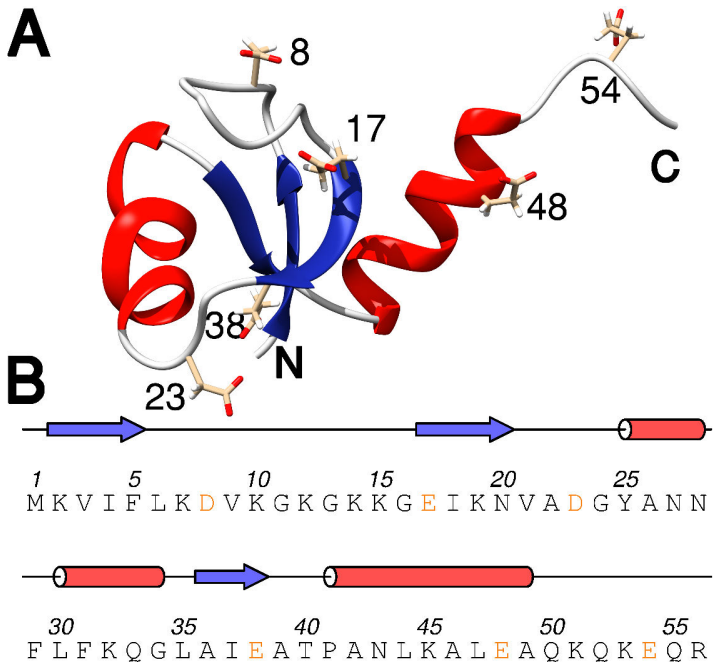


Figure 1.1: A - Structure of NTL9, pdb 1CQU, with highlighted acid side chain residues. B - Primary sequence of NTL9 with highlighted acid residues.

The present paper is organized into two sections: in the next section the methods are presented. The Structure-Based Model for protein folding and the monomer titration equation for each ionizable residue are described. The simulation details and analysis methods are also described. In the final section, the results are discussed. pK_a residues are calculated and compared to experimental data to verify whether the methodology is able to sample the charges of the protein properly. Denaturation curves and free energies calculations are performed for different pH values and compared to experiments. The folding routes are calculated to evaluate the pH-dependence and the importance of electrostatic interaction in each stage of the folding process. Finally, Φ-values and the frequency of contact formation indicate which are the important residues in the transition state in simulations with and without charges.

1.2 Methods

1.2.1 Structure-based C_α model

The NTL9 protein was modeled using a Structure-based model (SBM) in which the protein is coarse-grained in a C_α atom level of simplification [21–24, 35, 36].

The residues are represented by beads in α carbon positions and the Hamiltonian that gives the protein energy interaction is based on the geometry of its native state. This first approach does not take into account the charges in proteins and the potential energy surface reaches its minimum at this reference state [37]. The electrostatic interactions were introduced through the inclusion of point charges at the center of the beads representing basic/acid residues and treating the electrolyte solution according to the Debye-Hückel theory. The charge-charge interaction was given by a Coulomb screening potential (last term of Equation 1.1). This approach has been used successfully in other studies with fixed charge systems [2, 3, 17, 38]. The potential energy for a conformation Γ was calculated relative to the native conformation Γ_o, given by

$$\begin{aligned} V(\Gamma,\Gamma_o) = & \sum_{bonds} \epsilon_r (r-r_o)^2 + \sum_{angles} \epsilon_\theta (\theta-\theta_o)^2 \\ & + \sum_{dihedrals} \epsilon_\phi \left\{ [1-\cos(\phi-\phi_o)] + \frac{1}{2}[1-\cos(3(\phi-\phi_o))] \right\} \\ & + \sum_{contacts} \epsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \\ & + \sum_{electrostatics} K_{electrostatics} \frac{q_i q_j \exp^{-\kappa r}}{\epsilon_K r_{ij}} \end{aligned} \tag{1.1}$$

where r_0, θ_0 and ϕ_0 are defined by the values of the distance between two subsequent residues, angles formed by three subsequent residues and dihedral angles formed by four subsequent residues, respectively. All of them are obtained from NTL9 native structure (PDB-ID 1CQU). d_{ij} is the native distance between the pair contact i and j determined by CSU software (Contact of Structural Units) [39], $\epsilon_C = 1.0$ kcal/mol, $\epsilon_r = 100$ kcal/(mol Å^2), $\epsilon_\theta = 20$ kcal/mol, $\epsilon_{NC} = 1.0$ kcal/mol, $\sigma_{NC} = 4.00$ Å[21, 40] and $K_{electrostatics} = 332$ kcal Å/(mol e^2) [2, 5, 17, 38]. q_i and q_j are the charges from residues i and j respectively, κ is the inverse of Debye length [41] and the dieletric constant is $\epsilon_K = 80$.

1.2.2 Simulation Details

The constant-pH molecular dynamic method (CpHMD) adopted in this study combines a standard molecular dynamic simulation with the Metropolis Monte Carlo method for sampling protein protonation states [42–44]. Basically, at each given number of molecular dynamics steps, a titratable residue is randomly chosen and the changing of its protonation state is accepted or rejected according to the Metropolis criterion [45]. This algorithm was implemented using Espresso Package version 3.2 in

NVT ensemble with a Langevin thermostat [46, 47]. The protein was initialized in its crystal structure and with all titratable residues protonated. An equilibration process was carried out for 10^7 steps and the trajectory was generated from 10^9 steps with the integration time step having 0.5 fs. The results were stored every 1000 steps and Monte Carlo titration procedure was carried out every 2000 steps. The free energy profiles were obtained using the weighted histogram method (WHAM) [48–50], calculated employing 30 temperatures for each different pH value. The reaction coordinate used to describe the folding was defined as the sum of the native contacts (Q) in a structure Γ and a time t. A native contact was considered formed if the distance between the residues i and j ($j > i + 3$) was shorter than $1.2d_{ij}$ from the native structure.

1.2.3 Protonation/deprotonation Metropolis Criterion

The calculation of electrostatic energy variation ΔE associated with the protonation/deprotonation of a given residue is necessary for the implementation of the Metropolis criterion. This is generally done by the adoption of a small model compound with the same chemical titrable group of the residue and known pK_a, which takes into account local energetic contributions that can not be obtained from the adopted force field, such as quantum contributions to proton chemical bond [51, 52]. In this study, we used a compound model [53], which considers that local contributions are the same for the residue in the compound model and within the protein. ΔE is given by

$$\Delta E = \pm \xi k_B T \ln(10)(pH - pK_a) + \Delta E_{elec} \tag{1.2}$$

where the positive and negative signs are used for protonation or deprotonation, respectively, and the constant ξ is 1 or -1 for acid and basic residues, respectively. The first term is the contribution due the presence of a proton bath at a given concentration determined by pH; the second one is the contribution of the compound model; and the last one is the variation of the residue–residue electrostatic energy caused by changing the protonation state. The pK_a values adopted in this study are shown in Table I.

Φ-value analysis and contact formation frequency

Φ-value analysis is an approach to investigating the transition state ensemble of a two-state protein folding mechanism. These values indicate the importance of each residue during the folding process[21, 54–56]. Φ-values were calculated employing the free energy variation between the cases with and without Lennard Jones interactions involving the target residue[57]. The Φ-value for a residue k is obtained

Table I: pK_a of compound model

Residue	pK_a
Cter	3.6
ASP	4.0
GLU	4.5
HIS	6.3
Nter	7.5
LYS	10.6
ARG	12.0

by

$$\Phi^k = \frac{\Delta F^k_{TS} - \Delta F^k_U}{\Delta F^k_N - \Delta, F^k_U} \tag{1.3}$$

where ΔF^k_U,ΔF^k_{TS} and ΔF^k_N are its free energy values in the unfolded state (U), at transition state (TS) and at native state (N), respectively. The transition state is defined as values of native contacts fraction between 0.35 and 0.5 obtained from the free energy profile in the folding temperature simulation. The parameter measures the contribution from each residue mutated to the transition state (TS) relative to the native state (N).

Values close to 1 indicate that the specific residue forms native contacts in the transition state. It is considered important because it suggests that these native contacts, formed in the transition state, lead the protein to fold correctly. On the other hand, values close to 0 indicate a less important residue, because native contacts are formed when the protein is almost in the native state, not being essential to protein motif construction [58]. Values of $\Delta\Delta F$ lower than $0.5kT$ are discarded.

The contact formation frequency ($f(Q)$) is a statistical parameter calculated from the folding dynamics trajectory. Its value measures the formation of the native contact between the residue i and j that was made in the transition state. Both parameters are obtained employing a single simulation at the folding temperature.

Folding route calculation

The folding route is a measure that helps to describe the folding process along the energy landscape [59]. The route value indicates the specificity of each pathway that leads the protein to its native state among all possible pathways [60]. This calculation makes use of an ensemble of protein configurations sampled during the simulation and determines the ratio of accessible configurations between all possible configurations with the same amount of native contacts [61]. The function that

describes the route measure is given by

$$R(Q) = \sum_{i=1}^{M} \frac{\left\langle \left(\langle Q_i \rangle_Q - Q \right)^2 \right\rangle_Q}{MQ(1-Q)}, \tag{1.4}$$

where M is the total number of native contacts. The variable Q_i is equal to 1 when contact i is formed and equal to 0 if it is not formed for a particular conformation. The calculation of the average $\langle Q_i \rangle_Q$ is made using all configurations that have the same number of formed native contacts Q. The folding route measurement is normalized by the maximum number of possible routes, so $R(Q)$ are between 1 and 0. Values close to 1 mean a high specificity in the native contact formation leading to a few possible routes. On the other hand, values close to 0 have a high number of possible routes. In this case, the probability of a native contact being made is the same for all other native contacts. There is no preference in native contact formation, which means there are no specific pathways [62].

Titration curves

The ionization degree of each ionizable residue (α_i) was calculated by the average of the absolute valency ($|Z_i|$) of these residues throughout the configurations generated during the simulation $\alpha_i(pH) =< |Z_i(pH)| >$. The ionization degree was calculated for different pHs and then the titration curves were built. The generalized Henderson-Hasselbalch [63] equation $pH = pK_{ai} + \log\{\alpha_i(pH)/[1-\alpha_i(pH)]\}$ was adopted to fit the titration curves. It was possible to calculate the pK_a value of each acid residue [18].

1.3 Results and Discussion

Simulations were performed over a pH range from 0.5 to 12.0, at several ionic strengths and temperatures. The pK_a values of NTL9 were calculated fitting the individual titration curves with the Henderson-Hasselbalch equation (see Figure S1 in SI). Also, in order to test the generality of the approach used in this work, the pK_a was calculated for two additional proteins, the BBL and HEWL (see Table S1 and additional discussion in SI for more details). The uncertainty in the pK_a values was obtained by mean and standard deviation of three independent set simulations (all errors were of the order of 0.1 or lower pK_aunits and were omitted).

For NTL9, salt concentrations of C_s = 10, 100 and 750mM were used

Table II: Comparison between the experimental pK_a values[13] and the computational pK_a values for the acidic residues in NTL9

Residue	$C_s = 10$ mM		$C_s = 100$ mM		$C_s = 750$ mM	
	Exp pK_a	Comp pK_a [b]	Exp pK_a [a]	Comp pK_a [b]	Exp pK_a [a]	Comp pK_a [b]
$ASP-8$	**	0.6	3.0	1.9	3.2	3.9
$GLU-17$	**	1.4	3.6	3.0	3.8	4.4
$ASP-23$	**	2.0	3.1	3.2	3.2	4.0
$GLU-38$	**	2.4	4.0	3.6	4.1	4.4
$GLU-48$	**	2.4	4.2	3.5	4.4	4.4
$GLU-54$	**	2.7	4.2	3.4	4.2	4.3

** There are no experimental values of pK_a in this salt concentration available in literature;

[a] The experimental pK_a values for the acidic residues measured by Kuhlman et al.[13];

[b] The computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation.

($\kappa \approx 0.03$, 0.1 and 0.28 $Å^{-1}$, respectively). In the last two cases, the calculated pK_a values were compared with experiments, shown in Table I. For this comparison the definition of Goh *et al.* was used[33]. The error in calculated pK_a values is considered significant only in cases where the difference from the experimental pK_a is higher than 1.0 pK_a unit. For both salt concentrations, only ASP-8 presented a difference of -1.1 pK_a units at 100 mM of salt. This difference may be explained by the salt bridge interactions that ASP-8 forms with the amide backbone of near residues [33] and the computational model used, which did not have the necessary resolution to probe this kind of interaction. The proteins side chain explicitly simulated indicates an improvement in electrostatic calculation [64, 65], such as the formation of salt bridges interactions. In addition to the comparison with the experiment, the model used here was compared to different CpHMD approaches in more complex potential models with implicit solvent [66] and explicit solvent[33]. In both cases the NTL9 pK_a values calculated by CpHMD C_α SBM were very similar to those obtained via more complex simulations.

High electrostatic screening in $C_s = 750$ mM makes the pKa values of all residues close to the pKa values of ASP and GLU in isolated solution (Table I). This is because the interaction between the charged groups at this ionic strength is almost zero and their protonation behaviour is almost the same asa that isolated amino acids. Nevertheless, when ionic strength decreases, the electrostatic screening also decreases, allowing a higher interaction between charged residues. This higher interaction yields

to the variation in pKa values, which can be seen in Table I.

The folding/unfolding process is monitored by the fraction of native contact formation. Figure 1.2 shows the unfolded fraction as a function of temperature for pH 1.5 and pH 5.5 for three different salt concentrations of 100 mM (Fig. 1.2A), 10 mM (Fig. 1.2B) and 750 mM (Fig. 1.2C). In all three cases, NTL9 is more stable at pH 5.5 than at pH 1.5, denaturing at a higher temperature, which is in agreement with the experimental data [25]. An explanation for this stability difference could be the fact that, at pH 1.5, all residues tend to be protonated, generating only repulsive electrostatic interaction between positive charges of basic residues. This repulsion may hinder native contact formation and may suggest a decrease in protein stability. On the other hand, at pH 5.5, the acid residues tend to be deprotonated and the basic residues tend to be protonated. The electrostatic interactions between these charged residues are energetically favorable and suggest that these interactions may help the native contact formation compared to pH 1.5.

The stability difference of NTL9 for different pHs is expected to be related to the electrostatic interaction. Thus, at high ionic strength, this interaction tends to be screened and the stability difference between pH 1.5 and 5.5 is decreased. At the highest salt concentration (750 mM), the difference between the mean of total electrostatic energy ($\langle \Delta E_{elec} \rangle = \langle E_{elec}^{1.5} \rangle - \langle E_{elec}^{5.5} \rangle$) for pH 1.5 and 5.5 is about 0.5 K_BT, and the stability curves almost overlap. When the ionic strength is low, such as 10 mM, the electrostatic screening is also low and the pH effect becomes more evident. In this case, the value of $\langle \Delta E_{elec} \rangle$ is about 15.0 K_BT and the stability curves for pH 1.5 and 5.5 have the biggest separation in comparison with the other salt concentrations. NTL9 at an ionic strength of 100 mM presents intermediate behaviour in stability, with $\langle \Delta E_{elec} \rangle$ close to 4.0 K_BT.

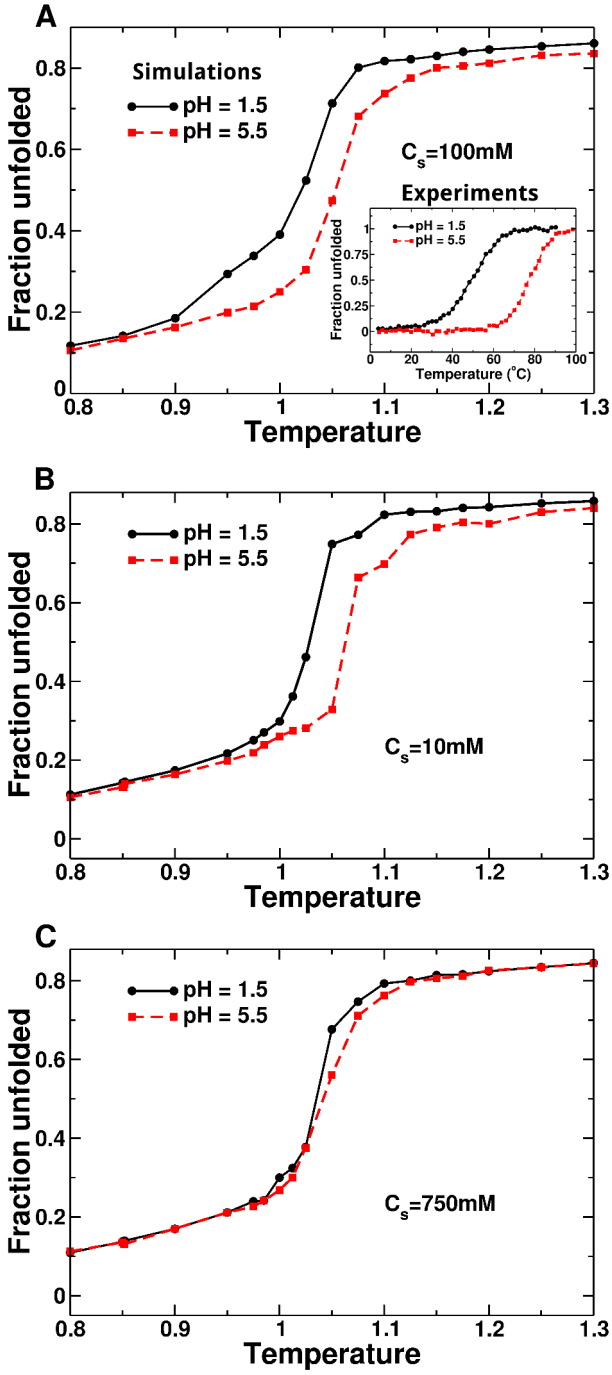


Figure 1.2: Fraction of unfolded configurations as a function of temperature in reduced units for two different pH conditions. The black circles represent the curve for pH 1.5 and the red squares for that pH 5.5. (A): Results obtained at $C_s = 100$ mM and inset graph shows experimental data of thermal unfolding curve of NTL9 adapted from Luisi, D. L. and Raleigh, D. P at same salt concentration[25]. (B) and (C): Results of simulations at $C_s = 10$ and 750 mM, respectively.

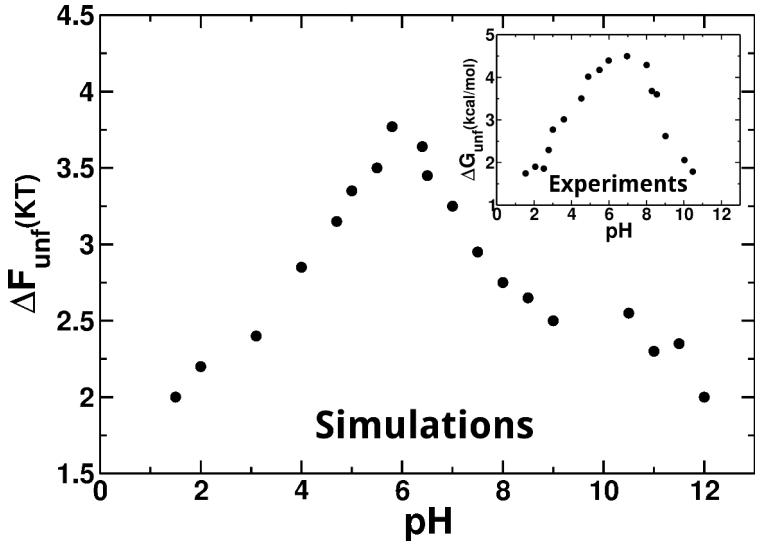


Figure 1.3: The Free energy unfolding barrier ΔF_{unf} as a function of the pH. The values were obtained from simulation analysis with $WHAM$ for 19 different pH values from 1.5 to 12.0 and $\kappa = 0.1$ Å^{-1} similar to experimental conditions. The inset plot is from experimental values for protein stability ΔG_{unf} as a function of pH adapted from Luisi, D. L. and Raleigh, D. P.[25].

In order to compare the computational and experimental results, the free energy barrier of unfolding was plotted as a function of pH (Figure 1.3). The value of the free energy barrier of unfolding is observed by the analysis of the free energy profile for each pH condition (Figure S2). Even with a very simplified model, the computational results were in good agreement with the experiments obtained by Luisi and coworkers [13, 25]. These results show a higher stability of NTL9 for intermediary pH values, close to 6.0. The reason for the higher stability near pH 6.0 seems to be due to the better balance of the interactions between the ionizable residues.

1.3.1 Effects of pH variation on the protein folding process

The folding route measurement was calculated to evaluate how the charge modulation may affect the protein folding mechanisms, which addresses how different pH scenarios may establish preferential routes or pathways [67]. $R(Q)$ as a function of Q is presented in Figure 1.4 for NTL9 without electrostatic interaction and with this interaction at pH 1.5 and pH 5.5. It was used a temperature near transition midpoint. At pH 5.5 the temperature in reduced units was 1.055, for pH 1.5 was 1.008 and in the case without charge T= 1.040.

The general behavior for all three cases was similar. At the beginning of folding with $Q \sim 0.3$ the route measurement was high, showing that some specific

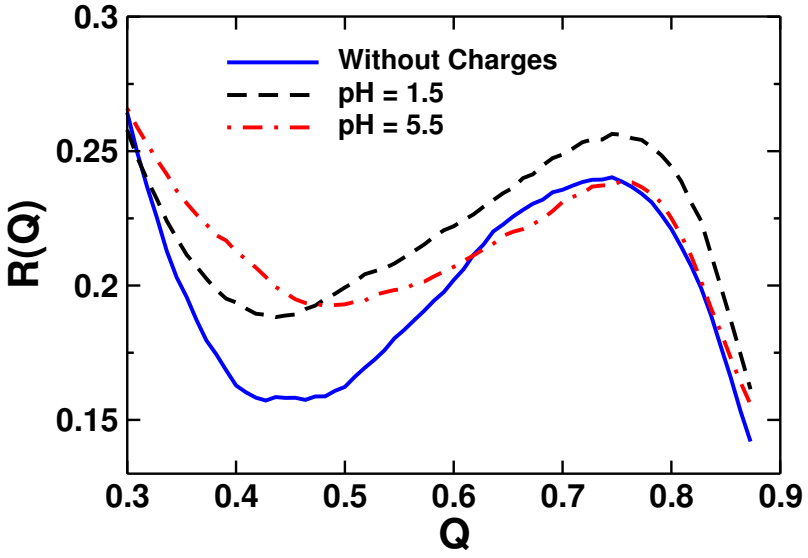


Figure 1.4: Folding Route measurement $R(Q)$ as a function of native contacts Q for different pH conditions. The continuous blue line represents the route for the standard C_α model without electrostatic interactions. The dashed black line represents the route for pH 1.5 and the dash-dot red line is the route measurement for pH 5.5.

contacts are formed with a high frequency even in the denatured state. There was no significant difference in the routes of the charged and uncharged protein. The reason $R(Q)$ is independent of the electrostatic interaction is the large distance between the charged groups in the denatured state and the electrostatic screening caused by $100mM$ of salt.

At intermediate values of Q (0.4 to 0.6) the protein enters the transition state and $R(Q)$ decreases. This reduction of $R(Q)$ at this folding stage in relation to the denatured state indicates a high degeneracy of the folding routes. However, $R(Q)$ for the charged NTL9 is higher than for the uncharged one in TS. Therefore, the presence of charges seems to induce a stronger order of contact formation than in the absence of electrostatic interaction, which is independent of pH. This result suggests that this interaction, be it predominantly repulsive (pH 1.5) or attractive (pH 5.5), makes the routes more specific. When NTL9 is close to the native state $Q \sim 0.75$ the route measurement increases again. At this stage of folding, almost all the native contacts are formed and the charged groups are close to each other. Thus, $R(Q)$ for pH 5.5 and without any electrostatic interaction is practically the same, but the higher value of $R(Q)$ occurs for pH 1.5. This behavior seems to be due to the proximity of positive charges and the repulsive interaction between them, which limits the number of possible routes to reach the native state.

1.3.2 Evaluation of the protein folding steps

It was shown that the biggest difference in the folding routes between the charged and the uncharged NTL9 occurs in the TS. Therefore, Φ-value measurements were used to analyse the native contacts formed in TS. Experimental Φ-values of three residues of NTL9 ($ASP8$, $LYS12$ and $GLU17$) were found in literature [68]. All of them have small Φ-values (near 0.2) and are not in good agreement with computational results. However, the experimental interpretation of these values is not simple and it requires caution if denatured state effects are significant. As a general rule, low Φ-values indicate that the native interaction of these residues are poorly developed in the transition state and are equally developed at each folding stage. According to Raleigh and coworkers [68] this is not the case and these residues are involved in non-native electrostatic interactions in both transition and denatured state.

In Figure 1.5A the $\Delta\Phi^k_{i,j}$ of each residue k is presented, which is defined as the difference between Φ^k at pH conditions i and j which can be pH 1.5, 5.5 or $\emptyset$ (model without charges). Figure 1.5A was also divided into three regions (left-blue from residues 1 to 20, middle-red from residues 20 to 39 and right-white from residues 40 to 56) because the $\Delta\Phi^k_{i,j}$ present distinct features. In the left region, from the N-terminal to residue 20, the $\Delta\Phi_{pH,\emptyset}$ tends to be positive and the peaks occur in $ASP8$ and $LYS12$ for both pH values. Experimental and computational results have showed that the electrostatic interaction between these residues promotes the contact formation even in the unfolded state [2, 32, 65]. Our analysis of the distance between $ASP8$ and $LYS12$ indicates that this contact formation also occurs in C_α model with charges in the denatured state. This contact formation suggests an increase in their Φ-values, as it has been observed. Moreover, this N-terminal part of the protein has the highest charge density and the contact formation of the others charged residues also becomes more important in the TS. In the middle region, from residue 20 to residue 39, $\Delta\Phi_{pH,\emptyset}$ tends to be close to zero or slightly negative. This behavior of $\Delta\Phi_{pH,\emptyset}$ is the opposite of what occurs in the first region and indicates that the contacts formed in this part of the protein become less important for the TS. In the right region, the $\Delta\Phi_{pH,\emptyset}$ of these residues is close to zero because most of them do not make any native contact, i.e. they do not produce variations higher than 0.5 kT when the perturbation analysis is done. This lack of contact is due to the fact that the residues form a kind of protein tail.

The difference between the contact frequencies between residues k and l, $\Delta f^{k,l}_{i,j}$ (using the same notation of $\Delta\Phi^k_{i,j}$) of NTL9 (Figure 1.5B and Figure 1.5C) shows that electrostatic interaction increases the frequency of contact formation within

the region near to the N-terminal, principally for pH 5.5. However, the frequency of contacts that this same region makes with residues 20 until the C-terminal tends to decrease. These results indicate that the interaction between the charged groups of the chain beginning drives the folding at the TS. In this folding stage the NTL9 tends to fold the most charged region first before folding the rest of the protein. For that reason the NTL9 is more routed when it is charged.

There is no significant difference between the $\Delta\Phi_{pH,\emptyset}$ values for pH 1.5 and pH 5.5 although there are practically no negative charges in pH 1.5. The most obvious differences in the region near to N-terminal occur in the case of $LYS14$ and $LYS15$ that have a higher $\Delta\Phi_{pH,\emptyset}$ for pH 5.5 than for pH 1.5. These differences occur due to the higher frequency of contact formation that these positively residues charged make with the negatively charged $ASP8$ in pH 5.5 (see figure 1.5D). The increase in this contact frequency also explains the peak of $\Delta\Phi_{5.5,\emptyset}$ that $ASP8$ has at this pH. However, the value of $\Delta\Phi_{1.5,\emptyset}$ for the uncharged $ASP8$ is also high, despite its contact formation frequency being low within the most charged region of NTL9. This can be explained by the increase in the frequency of contact formation that $ASP8$ make with $LEU35$. At pH 1.5, the residues in the region near the to N-terminal tend to make contact with the rest of the protein because of the electrostatic repulsion caused by the lack of negative charges. Therefore, most $\Delta\Phi_{pH,\emptyset}$ of the residues $20-39$ are slightly higher in pH 1.5 than in pH 5.5.

1.4 Conclusions

In this study, a computational method for studying NTL9 folding at constant pH was proposed. The method consists of a protein folding Molecular Dynamics of a Cα Structure-Based Model added to a screened Coulomb potential with titration of ionizable residues by Metropolis Monte Carlo criterion according to the pH of the simulation. All the statistical properties were sampled by molecular dynamics, and the protonation/deprotonation followed the Monte Carlo - Metropolis criterion. The pK_a values obtained from the simulations were close to the experimental results, suggesting that the charges were well sampled for all pH conditions. The NTL9 stability as a function of pH was analyzed and the computational results are in good agreement with the experiments. The free energy barrier had the same qualitative behavior, shown in Figure 1.3. The folding routes were calculated to evaluate the charge influence at each folding stage. The folding stage most altered by the electrostatic interaction was at the transition state. In this state, the interaction of the charged groups tends to limit the folding routes and the order of contact formation becomes more important. $\Delta\Phi_{pH,\emptyset}$

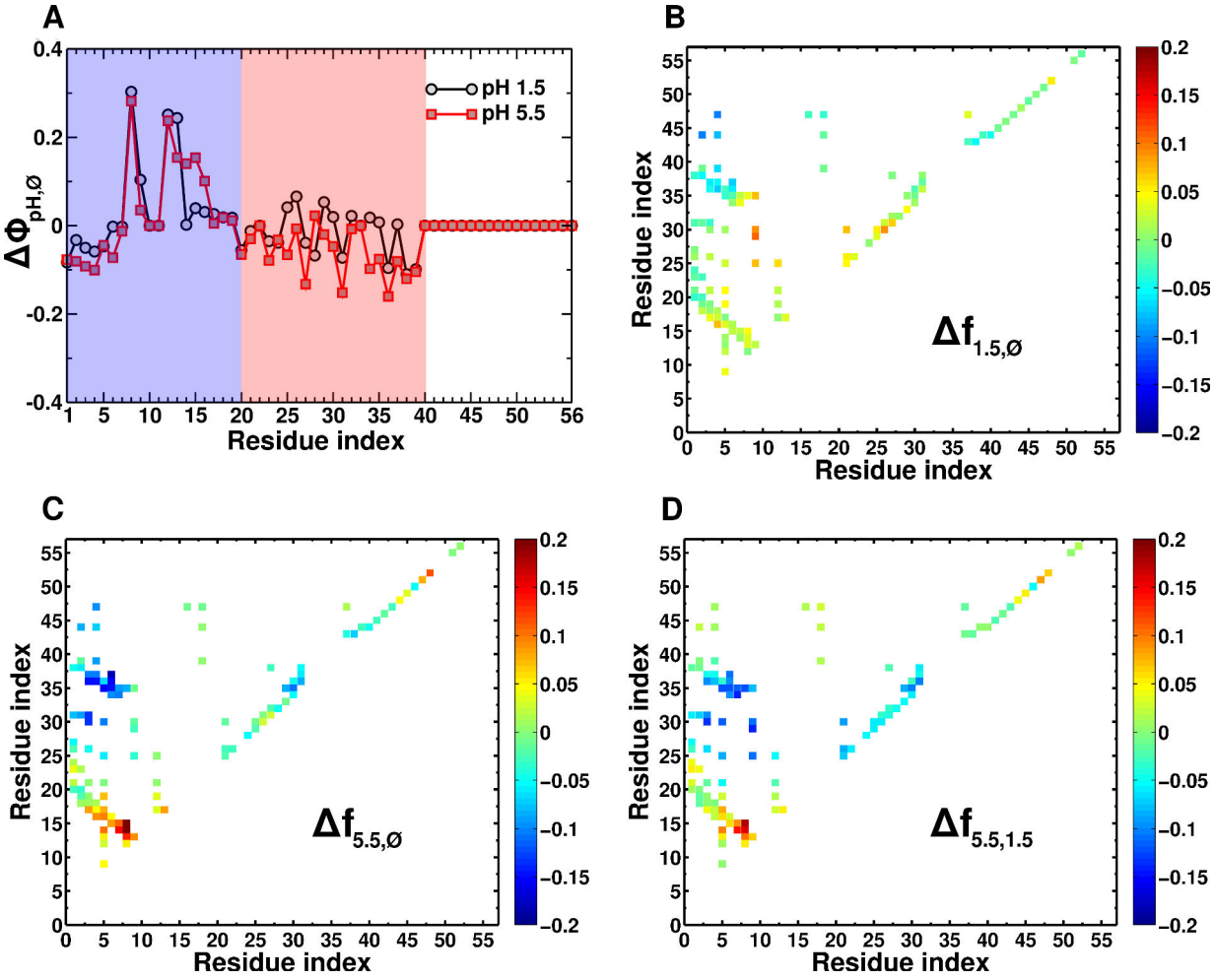


Figure 1.5: (A): $\Delta\Phi_{i,j}^{k}$ for each residue k, black line and circles for pH 1.5 and red squares for pH 5.5. (B),(C) and (D): $\Delta f_{i,j}^{k,l}$ for each pair of residues (k,l). The colorbar represents the $\Delta f_{i,j}^{k,l}$ value for each pair of native contact between residues k and l.

was used to better understand how the charges affect the TS. From residues 1 to 20 $\Delta\Phi_{pH,\emptyset}$ tends to increase and the most charged region of the protein has the highest increase in $\Delta\Phi_{pH,\emptyset}$ for both pH values. These results corroborate the route measurements and indicate the increased importance of specific contact formation by charged residues. Therefore, the method tested is suitable for providing qualitative and quantitative results that agree with experimental data. Furthermore, the computational results contributed to an understanding of the effect of charges and pH on the NTL9 folding. Accordingly, this method may be used for the study of other proteins which have their folding affected by electrostatic interaction and pH. This study shed light on the importance of considering variable charges, even in very simple coarse-grained models.

1.5 Acknowledgments

The authors would like to thank Margaret Cheung and José Nelson Onuchic for the helpful discussions. This work was financially supported by PROPe–UNESP and CNPQ (Grants 442352/2014-0, 2532/002/14-PROPe/CDC). VBPL was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2014/06862-7). VBPL, LCO and VMO were supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). VGC was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. We also thank the Center for Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) for computational resources.

1.6 Support Information

Five out of ten residues of HEWL presented at least 1 pK_a unit of difference from experiments (see Table S1). Distance differences between the C_α model and its all-atom native PDB structure were analyzed. $GLU35$ had the highest difference (2.9 pK_a units, underestimated). The side chain of this residue is almost not exposed to solvent (SASA $\sim$ 13 Å^2) and its self-solvation energy is low. In addition, the distance between the C_αs of $GLU35$ and the closest charged residue, $LYS33$, is about 5.7 Å, whereas the distance between the $GLU35$ carboxyl group and the $LYS33$ ammonium group is 13.2 Å. Considering that the electrostatic potential is calculated using a screened Coulomb with a 10 Å Debye length, the significant decrease in the distance between charges in the C_α model makes the interaction between $GLU35$ and $LYS33$ becomes more favorable. This interaction tends to keep $GLU35$ deprotonated and its pK_a value is underestimated by the model.

For the $ASP48$ and $ASP87$ residues, which have overestimated predicted pK_a values, the situation is reversed. The side chain of both residues is highly exposed to solvent (SASA $\sim$ 60 Å^2 and $\sim$ 73 Å^2, respectively). The C_α of $ASP48$ is far from $ARG61$ C_α, about 9.4 Å, while the $ASP48$ carboxyl group is almost three times closer to the $ARG61$ ammonium group ($\sim$ 3.2 Å). In the case of $ASP87$, $HIS15$ is the closest charged residue, and the carboxyl and ammonium group are distant about 11.0 Å. The distance of C_αs, however, is about 4.9 Å.

$ASP66$ is a special case, because, even having SASA 7 Å^2, the experimental pK_a is very low (1.2). This residue is near two positively charged residues, $ARG45$ and $ARG68$, which tends to keep $ASP66$ deprotonated. The ammonium group of $ARG45$ and $ARG68$ are distant from the $ASP66$ carboxyl group about 6.4 Åand 3.7 Å, respectively. These distances are smaller than the C_αs distances (11.2 Å and 5.5 Å, respectively). The increase in charged group distances in C_α model seems to be the cause of the overestimation in predicted pK_a value.

Lastly, $ASP101$ presents an underestimated predicted pK_a value. The closest charged residue of $ASP101$ is $LYS97$. The distance between the C_αs is $\sim$ 5.9 Å and between the carboxyl and ammonium group is $\sim$ 5.2 Å. There is no significant difference between the distances of the charges, therefore. We have no a plausible explanation for this discrepancy, but similar variations are also observed in other studies [66, 69].

Table III: Comparison between the experimental and computational pK_a values of 45-residue binding domain of 2-oxoglutarate dehydrogenase multi-enzyme complex BBL (PDB-ID 1W4H) and hen egg-white lysozyme HEWL (PDB-ID 2LZT).

Residue	Exp pK_a [a]	Comp pK_a [b]
BBL		
ASP-129	3.9	3.7
GLU-141	4.5	3.7
HIS-142	6.5	6.7
ASP-145	3.7	3.5
GLU-161	3.7	3.8
ASP-162	3.2	3.1
GLU-164	4.5	3.6
HIS-166	5.4	6.0
HEWL		
GLU-7	2.6 ± 0.1	3.2
HIS-15	5.5 ± 0.5	5.1
ASP-18	2.8 ± 0.0	2.3
GLU-35	6.1 ± 0.2	3.2
ASP-48	1.4 ± 0.2	3.5
ASP-52	3.6 ± 0.0	3.0
ASP-66	1.2 ± 0.4	3.1
ASP-87	2.2 ± 0.1	3.2
ASP-101	4.5 ± 0.3	3.0
ASP-119	3.5 ± 0.1	3.1

[a] The experimental pK_a values for the acidic residues measured by Arbely et al. [70] and Webb et al. [71];

[b] The computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation.

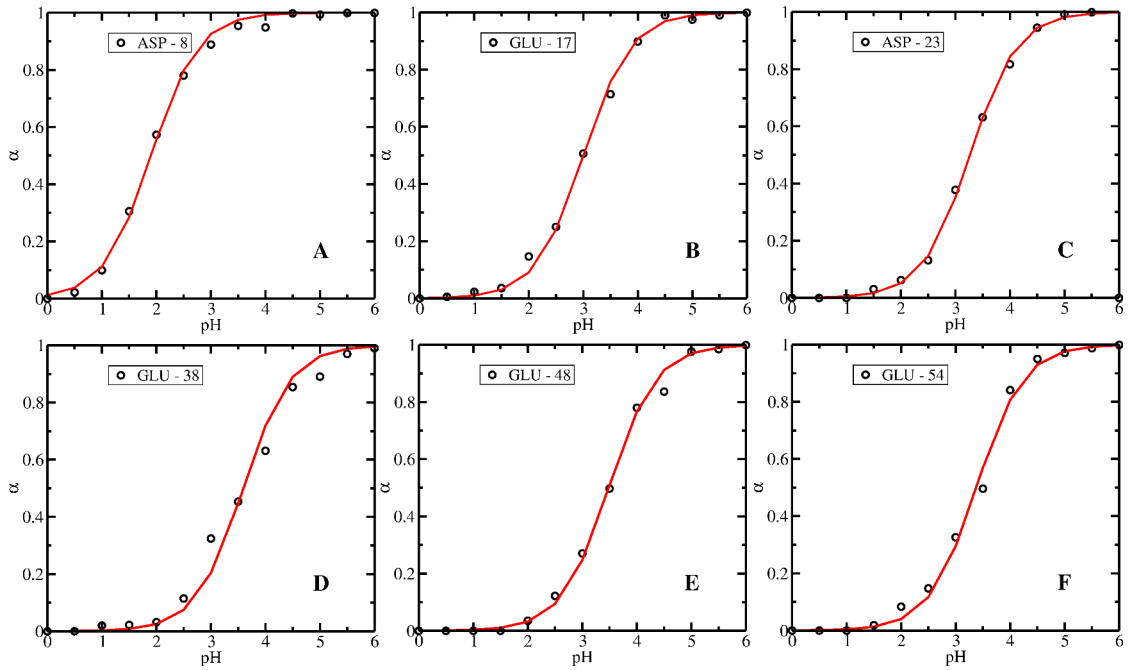


Figure 1.6: Titration curves: Ionization degree α as function of pH for each acid residue of NTL9 in ionic strength 100 mM. Red curves are the fit of Henderson-Hasselbalch equation. (A): Asp8; (B): Glu17; (C): Asp23; (D): Glu38; (E): Glu48 and (F): Glu54.

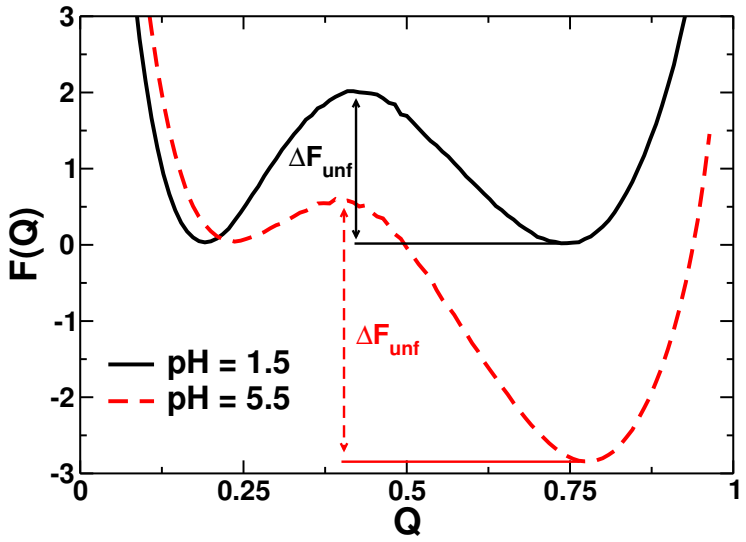


Figure 1.7: Free energy profiles F(Q) for NTL9 as function of native contacts Q. The black straight line represents the curve for pH 1.5. The red dashed line represents the curve for pH 5.5. The temperature used in both curves are the same evidencing the stabilization of the protein native state in pH 5.5 compared with pH 1.5. ΔF_{unf} is the free energy barrier of unfolding [25].

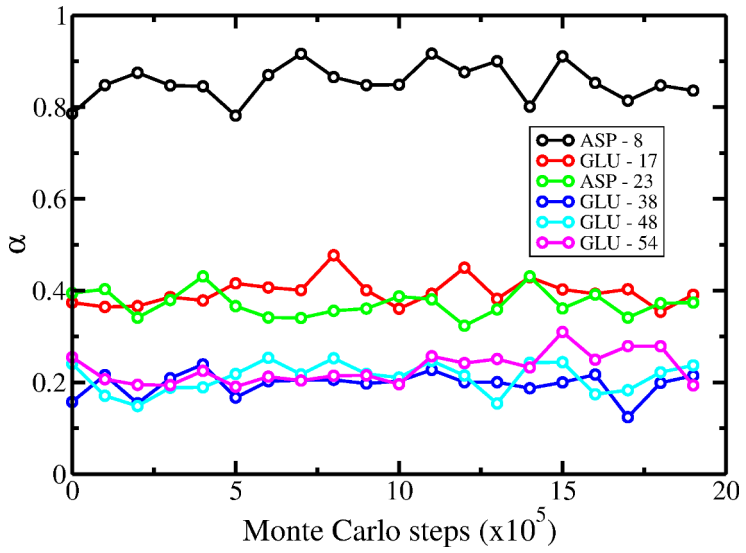


Figure 1.8: Values of ionization degree for each acid residue of NTL9 calculated within 10^5 steps. These results show a good convergence of the charges. Ionic strength was 100 mM and pH 3.0.

References

[1] Jae-Hyun Cho, Satoshi Sato, Jia-Cherng Horng, Burcu Anil, and Daniel P. Raleigh. Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch. Biochem. Biophys.*, 469(1):20–28, January 2008.

[2] Ariel Azia and Yaakov Levy. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.*, 393(2):527–542, October 2009.

[3] Arash Zarrine-Afsar, Zhuqing Zhang, Katrina L. Schweiker, George I. Makhatadze, Alan R. Davidson, and Hue Sun Chan. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins: Struct., Funct., Bioinf.*, 80(3):858–870, 2012.

[4] S. J. Davis, E. A. Davies, M. G. Tucknott, E. Y. Jones, and P. A. van der Merwe. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc. Natl. Acad. Sci. USA*, 95(10):5490–5494, May 1998.

[5] Ohad Givaty and Yaakov Levy. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, 385(4):1087–1097, January 2009.

[6] Alexey V. Gribenko, Mayank M. Patel, Jiajing Liu, Scott A. McCallum, Chunyu Wang, and George I. Makhatadze. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc. Natl. Acad. Sci. USA*, 106(8):2601–2606, February 2009.

[7] Ji Guo Su, Wei Zu Chen, and Cun Xin Wang. Role of electrostatic interactions for the stability and folding behavior of cold shock protein. *Proteins: Struct., Funct., Bioinf.*, 78(9):2157–2169, July 2010.

[8] Debabani Ganguly, Weihong Zhang, and Jianhan Chen. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. *PLoS Comput. Biol.*, 9(11):e1003363, November 2013.

[9] Stina Lindman, Wei-Feng Xue, Olga Szczepankiewicz, Mikael C. Bauer, Hanna Nilsson, and Sara Linse. Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.*, 90(8):2911–2921, April 2006.

[10] A. K. Stewart, C. E. Kurschat, and S. L. Alper. Role of nonconserved charged residues of the AE2 transmembrane domain in regulation of anion exchange by pH. *Pfluegers Arch.*, 454(3):373–384, June 2007.

[11] A. Warshel. Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry*, 20(11):3167–3177, May 1981.

[12] Ikram Ul Haq, Fatima Akram, Mahmood Ali Khan, Zahid Hussain, Ali Nawaz, Kaleem Iqbal, and Ali Javed Shah. CenC, a multidomain thermostable GH9 processive endoglucanase from Clostridium thermocellum: cloning, characterization and saccharification studies. *World J. Microbiol. Biotechnol.*, 31(11):1699–1710, November 2015.

[13] Brian Kuhlman, Donna L. Luisi, Paul Young, and Daniel P. Raleigh. pKa Values and the pH Dependent Stability of the N-Terminal Domain of L9 as Probes of Electrostatic Interactions in the Denatured State. Differentiation between Local and Nonlocal Interactions. *Biochemistry*, 38(15):4896–4903, April 1999.

[14] S. W. Englander. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.*, 29:213–238, 2000.

[15] Jana Khandogin, Jianhan Chen, and Charles L. Brooks. Exploring atomistic details of pH-dependent peptide folding. *Proc. Natl. Acad. Sci. USA*, 103(49):18546–18550, December 2006.

[16] Logan S. Ahlstrom, Sean M. Law, Alex Dickson, and Charles L. Brooks III. Multiscale Modeling of a Conditionally Disordered pH-Sensing Chaperone. *J. Mol. Biol.*, 427(8):1670–1680, April 2015.

[17] Swarnendu Tripathi, Angel E. Garcìa, and George I. Makhatadze. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge–Charge Interactions. *J. Phys. Chem. B*, 119(41):13103–13112, October 2015.

[18] Jana Khandogin and Charles L. Brooks. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.*, 89(1):141–157, July 2005.

[19] Donald Allan McQuarrie. *Statistical Mechanics*. University Science Books, 2000.

[20] V. M. de Oliveira and S. J. de Carvalho. Adsorption of pH-responsive polyelectrolyte chains onto spherical macroions. *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 37(8), August 2014.

[21] C Clementi, H Nymeyer, and J N Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.*, 298(5):937–953, May 2000.

[22] Paul C. Whitford, Jeffrey K. Noel, Shachi Gosavi, Alexander Schug, Kevin Y. Sanbonmatsu, and José N. Onuchic. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins: Struct., Funct., Bioinf.*, 75(2):430–441, May 2009.

[23] Vinícius G. Contessoto, Debora T. Lima, Ronaldo J. Oliveira, Aline T. Bruni, Jorge Chahine, and Vitor B. P. Leite. Analyzing the effect of homogeneous frustration in protein folding. *Proteins: Struct., Funct., Bioinf.*, 81(10):1727–1737, October 2013.

[24] Matheus R. de Mendonca, Leandro G. Rizzi, Vinicius Contessoto, Vitor B. P. Leite, and Nelson A. Alves. Inferring a weighted elastic network from partial unfolding with coarse-grained simulations. *Proteins: Struct., Funct., Bioinf.*, 82(1):119–129, January 2014.

[25] D. L. Luisi and D. P. Raleigh. pH-dependent interactions and the stability and folding kinetics of the N-terminal domain of L9. Electrostatic interactions are only weakly formed in the transition state for folding. *J. Mol. Biol.*, 299(4):1091–1100, June 2000.

[26] D. L. Luisi, B. Kuhlman, K. Sideras, P. A. Evans, and D. P. Raleigh. Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed alpha/beta protein: characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L9. *J. Mol. Biol.*, 289(1):167–174, May 1999.

[27] A G Murzin, S E Brenner, T Hubbard, and C Chothia. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, 247(4):536–540, April 1995.

[28] B. Kuhlman, D. L. Luisi, P. A. Evans, and D. P. Raleigh. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.*, 284(5):1661–1670, December 1998.

[29] B. Kuhlman, J. A. Boice, R. Fairman, and D. P. Raleigh. Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding. *Biochemistry*, 37(4):1025–1032, January 1998.

[30] Burcu Anil, Ying Li, Jae-Hyun Cho, and Daniel P. Raleigh. The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry*, 45(33):10110–10116, August 2006.

[31] Jae-Hyun Cho, Satoshi Sato, and Daniel P. Raleigh. Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol.*, 338(4):827–837, May 2004.

[32] Jae-Hyun Cho, Wenli Meng, Satoshi Sato, Eun Young Kim, Hermann Schindelin, and Daniel P. Raleigh. Energetically significant networks of coupled interactions within an unfolded protein. *Proc. Natl. Acad. Sci. USA*, 111(33):12079–12084, August 2014.

[33] Garrett B. Goh, Benjamin S. Hulbert, Huiqing Zhou, and Charles L. Brooks. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.*, 82(7):1319–1331, July 2014.

[34] Jana K. Shen. Uncovering Specific Electrostatic Interactions in the Denatured States of Proteins. *Biophys. J.*, 99(3):924–932, August 2010.

[35] Jeffrey K. Noel, Paul C. Whitford, Karissa Y. Sanbonmatsu, and José N. Onuchic. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.*, June 2010.

[36] Yunxiang Sun and Dengming Ming. Energetic frustrations in protein folding at residue resolution: A homologous simulation study of im9 proteins. *PLoS One*, 9(1):e87719, January 2014.

[37] H. Taketomi, Y. Ueda, and N. GO. Studies on protein folding, unfolding and fluctuations by computer simulation. i. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.*, 7(6):445–459, 1975.

[38] Franco O. Tzul, Katrina L. Schweiker, and George I. Makhatadze. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. USA*, 112(3):E259–E266, January 2015.

[39] V. Sobolev, A. Sorokine, J. Prilusky, E E Abola, and M. Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15(4):327–332, April 1999.

[40] N Koga and S Takada. Roles of native topology and chain-length scaling in protein folding: a simulation study with a go-like model. *J. Mol. Biol.*, 313(1):171–180, October 2001.

[41] Magnus Ullner, Cliff E. Woodward, and Bo Jönsson. A Debye–Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.*, 105(5):2056–2065, August 1996.

[42] António M. Baptista, Vitor H. Teixeira, and Cláudio M. Soares. Constant-ph molecular dynamics using stochastic titration. *J. Chem. Phys.*, 117(9):4184–4200, September 2002.

[43] John Mongan, David A. Case, and J. Andrew McCammon. Constant ph molecular dynamics in generalized born implicit solvent. *J. Comput. Chem.*, 25(16):2038–2048, 2004.

[44] John Mongan and David A. Case. Biomolecular simulations at constant ph. *Curr. Opin. Struct. Biol.*, 15:157–163, 2005.

[45] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, June 1953.

[46] H. J. Limbach, A. Arnold, B. A. Mann, and C. Holm. ESPResSo—an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174(9):704–727, May 2006.

[47] Axel Arnold, Olaf Lenz, Stefan Kesselheim, Rudolf Weeber, Florian Fahrenberger, Dominic Roehm, Peter Košovan, and Christian Holm. ESPResSo 3.1: Molecular Dynamics Software for Coarse-Grained Models. In Michael Griebel and Marc Alexander Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VI*, number 89 in Lecture Notes in Computational Science and Engineering, pages 1–23. Springer Berlin Heidelberg, 2013.

[48] Alan M. Ferrenberg and Robert H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61(23):2635, December 1988.

[49] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. THE weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13(8):1011–1021, October 1992.

[50] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. Multidimensional free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16(11):1339–1350, November 1995.

[51] Donald Bashford. Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci.*, 9:1082–1099, 2004.

[52] Arial Warshel, Pankaz A. Sharma, Mitsunori Kato, and William W. Parson. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta*, 1764:1647–1676, 2006.

[53] Alan Fersht. *Enzyme Structure and Mechanism.* W H Freeman & Co, New York, January 1985.

[54] Alan R. Fersht. Protein folding and stability: the pathway of folding of barnase. *FEBS Lett.*, 325(1–2):5–16, June 1993.

[55] A. R. Fersht. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.*, 5(1):79–84, February 1995.

[56] Alan R. Fersht and Satoshi Sato. phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA*, 101(21):7976–7981, May 2004.

[57] Paul C. Whitford, Osamu Miyashita, Yaakov Levy, and José N. Onuchic. Conformational transitions of Adenylate Kinase: switching by cracking. *J. Mol. Biol.*, 366(5):1661–1671, March 2007.

[58] Jeffrey K. Noel and José N. Onuchic. The many faces of structure-based potentials: from protein folding landscapes to structural characterization of complex biomolecules. In *Computational Modeling of Biological Systems*, pages 31–54. Springer US, 2012.

[59] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267(5204):1619–1620, March 1995.

[60] Steven S. Plotkin and José N. Onuchic. Investigation of routes and funnels in protein folding by free energy functional methods. *Proc. Natl. Acad. Sci. USA*, 97(12):6509–6514, June 2000.

[61] Leslie L. Chavez, José N. Onuchic, and Cecilia Clementi. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.*, 126(27):8426–8432, July 2004.

[62] Heiko Lammert, Jeffrey K. Noel, and José N. Onuchic. The Dominant Folding Route Minimizes Backbone Distortion in SH3. *PLoS Comput. Biol.*, 8(11):e1002776, November 2012.

[63] A. Onufriev, D. A. Case, and G. M. Ullmann. A novel view of pH titration in biomolecules. *Biochemistry*, 40(12):3413–3419, March 2001.

[64] Edward P. O'Brien, Bernard R. Brooks, and D. Thirumalai. Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. *J. Am. Chem. Soc.*, 134(2):979–987, January 2012.

[65] Tao Chen, Jianhui Song, and Hue Sun Chan. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.*, 30:32–42, February 2015.

[66] Jason A. Wallace and Jana K. Shen. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.*, 7(8):2617–2629, August 2011.

[67] Thomas R. Weikl and Ken A. Dill. Folding Rates and Low-entropy-loss Routes of Two-state Proteins. *J. Mol. Biol.*, 329(3):585–598, June 2003.

[68] Jae-Hyun Cho and Daniel P. Raleigh. Electrostatic Interactions in the Denatured State and in the Transition State for Protein Folding: Effects of Denatured State Interactions on the Analysis of Transition State Structure. *J. Mol. Biol.*, 359(5):1437–1446, June 2006.

[69] J. E. Nielsen and G. Vriend. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Proteins: Struct., Funct., Bioinf.*, 43(4):403–412, June 2001.

[70] Eyal Arbely, Trevor J. Rutherford, Hannes Neuweiler, Timothy D. Sharpe, Neil Ferguson, and Alan R. Fersht. Carboxyl pKa Values and Acid Denaturation of BBL. *J. Mol. Biol.*, 403(2):313–327, October 2010.

[71] Helen Webb, Barbara Mary Tynan-Connolly, Gregory M. Lee, Damien Farrell, Fergal O'Meara, Chresten R. Søndergaard, Kaare Teilum, Chandralal Hewage, Lawrence P. McIntosh, and Jens Erik Nielsen. Remeasuring HEWL pKa values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pKa calculations. *Proteins: Struct., Funct., Bioinf.*, 79(3):685–702, March 2011.

Capítulo 2

Variante da proteína G - 1PGB-QDD

Effects of pH and Salt Concentration on Stability of a Protein G Variant: Insights Using Coarse-Grained Models

Vinícius M. de Oliveira†, Vinícius G. Contessoto†, Fernando B. da Silva, Daniel L.Z. Caetano, Sidney J. de Carvalho, Vitor B.P. Leite

†These authors contributed equally to this work

*This article has been submmited to Journal of Chemical Theory and Computation.

Abstract

The importance of charge-charge interactions in the thermal stability of proteins is widely known. pH and ionic strength play a crucial role in these electrostatic interactions, as well as in the arrangement of ionizable residues in each protein-folding stage. In the present study, two coarse-grained models were used to evaluate the effect of pH and salt concentration on the thermal stability of a protein G variant (1PGB-QDD), which was chosen on account of the quantity of experimental data exploring these effects on its stability. One of these coarse-grained models, the TKSA, calculates the electrostatic free energy of the protein in the native state via the Tanford-Kirkwood approach for each residue. The other one, CpHMD-SBM, uses a Coulomb screening potential in addition to the structure-based model Cα. Both models simulate the system in constant pH. The comparison between the experimental stability analysis and the computational results obtained by these simple models showed a remarkable agreement. Through the TKSA method, the role of each charged residue in the protein's thermal stability was inferred. Using CpHMD-SBM, it was possible to evaluate salt and pH effects throughout the folding process. Finally, the computational pK_a values were calculated by both methods and presented a good level of agreement with the experiments. The present study provides new information and a comprehensive description of the electrostatic contribution to protein G stability.

Keywords: Constant pH simulations, ionic strength, protein stability, protein G variant.

2.1 Introduction

A knowledge of the factors that affect the thermal stability of proteins is of fundamental importance in understanding the basic principles that govern the behavior of these macromolecules as well as in the rational development of biotechnological applications.[1–3] It is widely known that a variation in pH and salt concentration can affect the thermal stability of proteins[4–7], indicating the importance of electrostatic interactions. Indeed, studies have shown the thermal stability of proteins can be enhanced by optimizing charge–charge interactions.[8–10] Establishing a relationship between the distribution of ionizable residues and the features of folding and thermal stability can therefore provide valuable information for protein engineering.

Several levels of approximation have been proposed for model proteins and for evaluating the electrostatic contributions of unfolding free energy[11–14]. In several of these studies, versions of structure-based models that include a fixed charge have been adopted to study the contribution of the electrostatic interaction in the unfolded state and the interaction between proteins and and charged macromolecules[8, 15, 16]. Recently, in order to study pH effects on the folding dynamics of the N-terminal domain of ribosomal protein L9 (NTL9)[17], the constant-pH molecular dynamic method (CpHMD)[17–19] was implemented and the results showed a good level of agreement with the experiments. Another alternative to evaluate the charge-charge contribution to protein stability is the Tanford-Kirkwood model (TK) [20], which enable the individual electrostatic interaction of each ionizable residue to be calculated [21]. Makhatadze and colleagues have managed to optimize proteins by the mutation of specific residues, which are predicted by the TK approach with a solvent accessibility modification, the Tanford-Kirkwood Solvent Accessibility (TKSA) method. [22–25] In the present study, both methods, CpHMD and TKSA, are used to study the importance of pH and salt concentration in 1PGB-QDD stability.

1PGB-QDD is a variant of the B1 domain of protein G with mutations T2Q, N8D, and N37D (Fig.2.1A). Protein G is a multi-domain protein present in the cell wall of G streptococcus containing the immunoglobulin G binding domains denoted as B1 and B2[26–30]. These proteins differ by six amino acids and do not have disulfide bonds[31, 32]. Also, experimental data showed that the protein folds/unfolds as a two-state model without kinetic intermediates[32]. PGB1-QDD has 56 amino acids (including the NH_2-terminal Met) and a regular α/β structure. The fold consists of a four β-sheet and one α-helix bundled against the sheet. The charged residues are almost entirely surface exposed [33] and display distinct pH-dependence protein stability [31]. Another important factor is that PGB1-QDD displays the highest stability at low salt

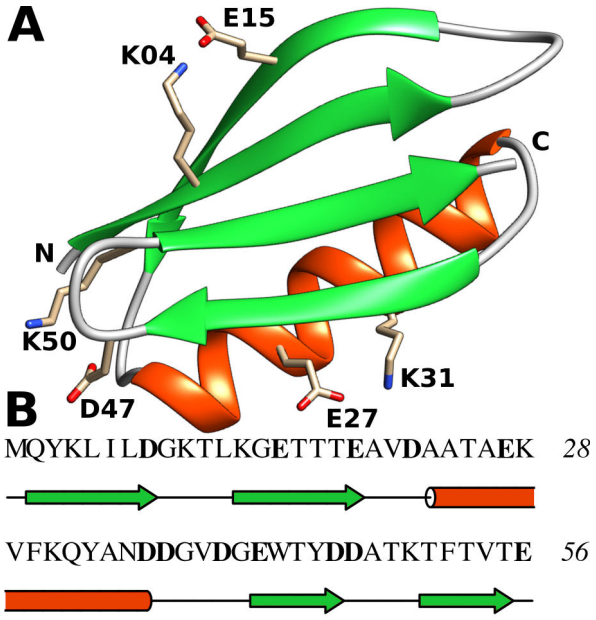


Figure 2.1: (A) Structure of PGB1-QDD built from wild-type protein with PDB ID:1PGB. The side-chains are shown for residues which have major contributions to protein stability. (B) Primary sequence of PGB1-QDD with highlighted acid residues, which have the highest charge variations in the studied pH range, and provide a useful insight into biological systems.

concentrations[4]. Due to its dependence on salt concentration and pH, it is an excellent model for studying electrostatic interactions.

Salt concentration and pH dependence in 1PGB-QDD stability were tested using the TKSA and CpHMD methods, and the results are in remarkable qualitative agreement with the experiments. The contribution of each ionizable residue to this protein stability was also calculated using the TKSA. From CpHMD simulations, it was possible to evaluate the dependence of the free energy unfolding barrier and thermal denaturation on pH. Electrostatic interaction influence at each folding stage was also analyzed. Finally, the pK_a values were calculated for the folded protein (using CpHMD and TKSA) and the unfolded protein (using only CpHMD). The computational pK_a values are in excellent agreement with the experimental values, considering the simplicity of the models. Thus, these coarse-grained models were found to be a computationally fast and cheap alternative to explore and provide useful insight into biological systems.

2.2 Methods

2.2.1 The Tanford-Kirkwood Model with Solvent Accessibility - TKSA

In the model proposed by Tanford and Kirkwood (TK), the protein is modeled by a spherical cavity with dielectric constant, ϵ_p and radius b, surrounded by an electrolyte solution treated according to the Debye-Hückel theory [20]. A solvent static accessibility parameter for each charged residue was incorporated by Shire *et al* into the TK model [34]. This modification was introduced to overcome the uncertainty of an adjustable charge-burial parameter beneath the dielectric interface and to allow for the irregular protein-solvent interface. With this change, the model proposed by Tanford and Kirkwood is now referred to as the Tanford-Kirkwood model with a solvent accessibility (TKSA) [34–36]. The interaction energy, U_{ij}, between two charged residues is given by

$$U_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}) \,, \tag{2.1}$$

where e is the elementary charge; b is the radius of the protein; a is the closest possible approach distance of an ion; A_{ij}, B_{ij} and C_{ij} are parameters obtained from the analytical solution of the Tanford and Kirkwood model, which are functions of the distance between charged residues, the dielectric constants, and the salt concentration [20, 37, 38]; and SA_{ij} is the average of the solvent accessible surface area of residues i and j [34, 35, 39].

Once the electrostatic energy of the ionizable residues of the protein is determined, it is necessary to sample all the possible protonation states [21, 40]. For a protein with n ionizable residues, there are $\chi = 2^n$ different protonation states. The free energy, $\Delta G_N(\chi)$, for a protein in its native state and for given state of protonation χ is given by

$$\Delta G_N(\chi) = -RT(\ln 10) \sum_{i=1}^{n} (q_i + x_i) pK_{int,i} + \frac{1}{2} \sum_{i,j=1}^{n} U_{ij}(q_i + x_i)(q_j + x_j), \tag{2.2}$$

where R is the ideal gas constant; T is the temperature; q_i is the charge of the ionizable residue i in its deprotonated state; x_i is 0 or 1, depending on the protonation state of the residue i; and $pK_{int,i}$ is the intrinsic pK_a of the ith ionizable residue.

The probability that the protein is in its native state with a particular state

of protonation, $\rho_N(\chi)$, is

$$\rho_N(\chi) = \frac{1}{Z_N} \exp\left[-\frac{\Delta G_N(\chi)}{RT} - \nu(\chi)(\ln 10)pH\right], \tag{2.3}$$

where $\nu(\chi)$ is the number of ionizable residues that are protonated in the state of protonation χ, and Z_N is the partition function for the protein in its native state.

The average total electrostatic energy, $\langle W_{qq}\rangle$ is given by the average over all protonation states, taking into account Eq. 2.3:

$$\langle W_{qq}\rangle = \sum_{\chi}\left[\frac{1}{2}\sum_{i,j=1}^{n} U_{ij}(q_i + x_i)(q_j + x_j)\right]\rho_N(\chi). \tag{2.4}$$

Ibarra-Molero and Makhatadze have shown that the contribution of electrostatic interaction energy to free energy, ΔG_{qq}, can be described by the negative of the average total electrostatic energy of the native state, that is, $\Delta G_{qq} \approx -\langle W_{qq}\rangle$ [21].

2.2.2 Constant-pH Molecular Dynamics Simulations

In the Structure-Based Model (SBM), the residues of a protein are represented by individual beads centered in C_α position[41–45]. The energy of the protein is given by a Hamiltonian equation in which only the native interactions based on its native topology are taken into account[41]. The first simplification of this model does not take into account the charged residues in the protein; the electrostatic interactions are therefore inserted by adding charged points at beads, which represent acid/basic residues [8, 15, 17, 46]. The interaction between charged residues is given by a Coulomb screening potential. The energy in a given configuration Γ with regard to configuration Γ_o, the native structure, is given by:

$$\begin{aligned} V(\Gamma,\Gamma_o) = &\sum_{bonds}\epsilon_r(r-r_o)^2 + \sum_{angles}\epsilon_\theta(\theta-\theta_o)^2 \\ &+ \sum_{dihedrals}\epsilon_\phi\left\{[1-\cos(\phi-\phi_o)] + \frac{1}{2}[1-\cos(3(\phi-\phi_o))]\right\} \\ &+ \sum_{contacts}\epsilon_C\left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10}\right] + \sum_{non-contacts}\epsilon_{NC}\left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \\ &+ \sum_{electrostatics} K_{electrostatics}\frac{q_i q_j \exp(-\kappa r_{ij})}{\epsilon_K r_{ij}}, \end{aligned} \tag{2.5}$$

where r_o represents the distance between two subsequent residues, θ_o and ϕ_o are the angles formed by three and four subsequent residues, respectively. All of them are taken from the native structure. Factor d_{ij} represents the distance between pairs of residues i and j ($i < j - 3$), $\epsilon_C = 1.0$ kcal/mol, $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$, $\epsilon_\phi = \epsilon_{NC} = \epsilon_C$, $\sigma_{NC} = 4$Å[41] and $K_{electrostatic} = 332$ kcalÅ/(mol e^2)[8, 16, 17]. The charges of residues i and j are q_i and q_j respectively, κ is the inverse of Debye length[47] and the dielectric constant is $\epsilon_K = 80$.

Protonation/Deprotonation Criterion

The constant-pH molecular dynamics method (CpHMD) was utilized in this study. It associates a standard molecular dynamics (MD) simulation with the Metropolis Monte Carlo method (MC) for sampling protein protonation states[17, 19, 48–50]. During the simulation, the molecular dynamic is regularly halted, a MC step is taken, and a random titratable residue is considered to change its protonation. The transition energy, ΔE, for this step is appraised according to Equation 4.2, in which the positive and negative signs mean protonation and deprotonation respectively. The acidic or basic residues are represented by ξ, where acidic residues present ξ = -1 and basic residues ξ = +1.

$$\Delta E = \pm \xi k_B T \ln(10)(pH - pK_a) + \Delta E_{elec} \tag{2.6}$$

In Equation 4.2, k_B is the Boltzmann constant, T is the temperature, pK_a is the value of the reference compound (reference pKa values are 3.6 for C-terminal, 4.0 for Aspartic acid, 4.5 for Glutamic acid, 6.3 for Histidine, 7.5 for N-terminal, 10.6 for Lysine and 12.0 for Arginine). The first term is the contribution due to the presence of a proton bath at a given concentration determined by pH; the second one is the contribution of the compound model; and the last one, ΔE_{elec}, is the electrostatic energy variation for the protonation state change of the titratable residue. The protonation/deprotonation is determined by the Metropolis criterion[51]. Thus, if the MC step is accepted, the protonation/deprotonation of the residue will change to the new state, and the MD will continue. However, if the new state is not accepted, there is no need for updating, and the residue continues with its unchanged protonation state.

2.2.3 Simulation Details

The CpHMD-SBM algorithm was implemented using Espresso Package version 3.2 in NVT ensemble and Langevin thermostat [52, 53]. In the folding simulations,

the protein was initialized in its native state with all titratable residues protonated. An equilibration process was executed with 10^7 steps and 10^9 steps to generate the trajectory with a time step of 0.5 fs. The data was stored every 1000 MD steps, and the Monte Carlo titration procedure was carried out every 2000 steps [17]. The free energy and specific heat profiles were obtained using the weighted histogram method (WHAM)[54–56], calculated from different simulations at 25 temperatures for pH ranging from 1.0 to 12.0 for high and low salt concentrations. The reaction coordinate used to describe the folding was defined as the sum of native contacts (Q) in a structure Γ and a time t. A native contact was considered formed when the distance between the residue i and j ($i < j-3$) was shorter than 1.2 d_{ij} from the native structure. The native contact map was determined by contact of structural units software (CSU) [57]. The mutant protein PGB1-QDD was built from PGB1 protein with PDB ID:1PGB using Modeller software version 9.17 [58]. The protein 3D structure graphic in Figure 2.1A was created with the UCSF Chimera package[59] and the protein sequence in Figure 2.1B was generated with Aline software[60]. The TKSA method was implemented in Python 2.7, and the solvent accessible surface area (SASA) was calculated with Surface Racer© software [61]. The TKSA method was performed for each of the 19 charged residues in PGB1-QDD and was simulated for pH from 1.0 to 12.0 at intervals of 0.5.

2.3 Results and Discussion

2.3.1 Electrostatic energy calculations – TKSA

In this section the contribution of electrostatic energy to protein free energy was calculated with the TKSA method. The sum of electrostatic energy of each charged residue is here referred to as ΔG_{elec}. The ΔG_{elec} of PGB1-QDD was calculated in a range of pH from 1.0 to 12.0, and the stability pH-dependent profile is presented in Figure 2.2.

The data presented in Figure 2.2 of ΔG_{elec} as a function of pH have similar behavior with slight differences for high and low salt concentrations in simulations via the TKSA method. The TKSA simulation results indicate that high levels of salt concentration do not eliminate the electrostatic interaction of surface charges and the pH dependence of protein native state stability is still strong [4]. The peak of TKSA curves indicates that the protein PGB1-QDD native state is more stable around its isoelectric point, which is close to pH 4.5 for both high and low salt concentrations. The TKSA result has a level of good agreement with experimental data [4] for low salt concentrations shown inset in Figure 2.2, and the correlation value between TKSA and

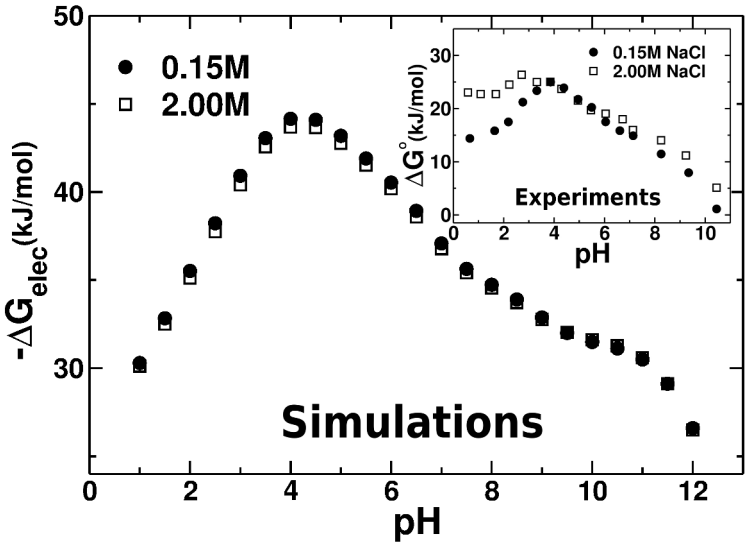


Figure 2.2: Electrostatic energy contribution to free energy native state stability ΔG_{elec} as a function of pH in kJ/mol. The values were calculated from TKSA simulation analysis for 23 different pH values from 1.0 to 12.0. The black circles represent values for 0.15M of a monovalent salt concentration in simulation and the white filled squares represent values of 2.0M of salt concentration. The inset plot shows the experimental results of the free energy difference between the unfolded and the native state ΔG° from 1PGB-QDD as a function of pH in kJ/mol. The black circles represent values for 0.15M of NaCl, and the white filled squares for 2.00M of NaCl (adapted from ref.[4])

experimental data is 0.83. On the other hand, for high salt concentrations, there is not a strong correlation between the TKSA and experimental results and the correlation value is 0.44. The weak correlation is due to the differences between the results for lower pH values. The behavior of experimental data for lower pH values does not present a significant stability difference from 2M of NaCl [4] shown inset in Figure 2.2. The correlation between TKSA and experimental data became stronger if only values of pH higher than 3.0 are taken into account and the new correlation value is then 0.95.

Another result obtained from TKSA simulation is the contribution of each charged residue to protein stability[21, 25, 62]. Figure 2.3 shows the TKSA analysis for pH values 2.5, 4.5, 7.5 and 10.0 for low salt concentrations. For pH 2.5 in Figure 2.3A there are eight residues with $\Delta G_{qq} < 0.0$ which contribute favorably to protein stability. The other eleven residues have energy values close to zero with $\Delta G_{qq} \approx 0.0$. The total electrostatic free energy ΔG_{elec} for pH 2.5 is -38.22 kJ.mol^{-1} and is calculated as the sum of ΔG_{qq} of all charged residues [21, 22, 62]. The major contribution to stability comes from six residues: K04, E15, E27, K31, D47, and K50. These six residues make and important energetic contribution to all pH values and they are responsible for 86% $\pm$ 6.2% of all favorable interaction for pH values 2.5, 4.5, 7.5, and 10.0 shown in Figure 2.3. The side chains of these residues are not fully exposed to solvent, with

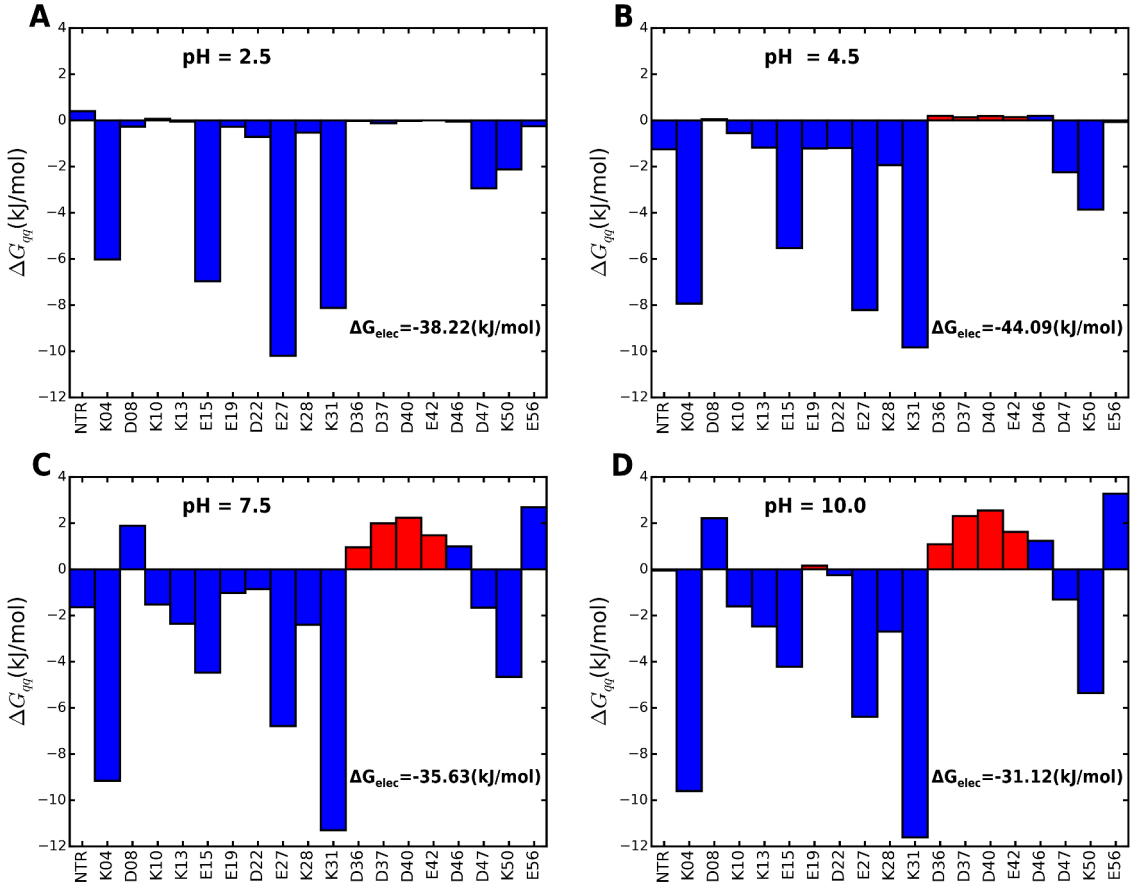


Figure 2.3: Charge–charge interaction energy ΔG_{qq} calculated by the TKSA model for each ionizable residue. A, B, C, and D represent the energy profiles of pH values 2.5, 4.5, 7.5 and 10.0, respectively. The red bars indicate the residues with the side chain exposed to solvent with SASA $\geq$ 50% and positive energy contribution to native state stability.

SASA$\leq$ 55%. However, the short distance between their ionizable groups of opposite charge (less than 2.95Å $\pm$ 0.56Å) promotes the high energetic contribution of these groups to protein stability (the side chains of K04, E15, E27, K31, D47, and K50 are highlighted in Figure 2.1A). At pH 4.5 in Figure 2.3B, similar to pH 2.5 analysis, the major electrostatic energy contribution comes from the six residues mentioned above. At pH 4.5 there are twelve residues with $\Delta G_{qq} < 0.0$ and the other seven residues have energy values close to zero, with $\Delta G_{qq} \approx 0.0$. The total electrostatic free energy ΔG_{elec} for pH 4.5 is -44.09 kJ.mol^{-1}, which is the pH which maximizes protein stability in experiments [4] and also in the TKSA simulation shown in Figure 2.2. For pH 7.5 shown in Figure 2.3C there are also 12 residues with $\Delta G_{qq} < 0.0$ similar to pH 4.5. There are seven residues with $\Delta G_{qq} > 0.0$ contributing with unfavorable electrostatic interactions to native state stability. These seven residues: D08, D36, D37, D40, E42, D46, and E56 are all acidic residues in their deprotonated state with charge -1. The protonation/deprotonation events are detailed in the pK$_a$ comparison section. The positive ΔG_{qq} values of these seven residues make the greatest contribution to

decreasing the native state stability of PGB1-QDD, and the total electrostatic free energy ΔG_{elec} for pH 7.5 is -35.63 kJ.mol^{-1}. In Figure 2.3C the red bars highlight the residues whose side chains are exposed to solvent more than 50% with SASA = 73% ± 14% for residues D36, D37, D40, and E42. The positive ΔG_{qq} energy contribution and the side chain exposed more than 50% to solvent indicate that mutations in these residues should lead the protein to a more stable native state based on optimization of the electrostatic interactions on the protein surface [21–25, 62–64]. The TKSA result for pH 10.0 is shown in Figure 2.3D with 9 residues $\Delta G_{qq} < 0.0$, 3 residues with ΔG_{qq} close to zero and 7 residues with $\Delta G_{qq} > 0.0$. The discussion of the favorable and unfavorable electrostatic energy contribution of each residue in pH 10.0 is similar to what has been discussed for pH 7.5. There is an increase of ΔG_{qq} for acid residues. The positive energy contribution passes from 12.2 kJ.mol^{-1} in pH 7.5 to 14.5 kJ.mol^{-1} in pH 10.0. The residue E19 had a significant variation from negative value $\Delta G_{qq} = -1.07$ kJ.mol^{-1} to a slightly positive contribution $\Delta G_{qq} = 0.15$ kJ.mol^{-1}. The N-terminal also had a significant variation due to the fact that it becomes deprotonated with charge 0 in pH 10.0 and its energy contribution is close to zero. The total electrostatic free energy ΔG_{elec} for pH 10.0 is -31.12 kJ.mol^{-1}

2.3.2 pH and ionic strength effects on 1PGB-QDD folding – CpHMD-SBM-C$_\alpha$

The previous section indicates that the drive interaction responsible for PGB1-QDD stability is electrostatic. However, the TKSA model takes into account only the static native structure of the protein in its calculations, while experimental data suggests that the unfolded state of PGB1-QDD is also important to its stability[4]. Thus, in this section, CpHMD-SBM simulations were utilized, which can explore the effects of the unfolded and transition states on protein stability. 1PGB-QDD was therefore simulated at four different pHs (2.5, 4.5, 7.5 and 10.0) and three ionic strengths (Low salt ≈ 0.01 M, intermediate salt ≈ 0.1 M and high salt ≈ 0.4 M).

Heat capacity (C_v) and free energy profiles of PGB1-QDD were calculated to explore the role of pH and ionic strength on protein folding thermodynamics (Figure 2.4). In these computational results, electrostatic screening at the high salt concentration is too strong, making the pH effect on protein stability almost insignificant. Even so, it is possible to see a slight increase in 1PGB-QDD stability at pH 4.5. As expected, the pH effect becomes more evident in the protein at low and intermediate salt concentrations, but the higher stability remains at pH 4.5. At the low salt concentration, the difference between the C_v peaks of each pH curve was maximized. The

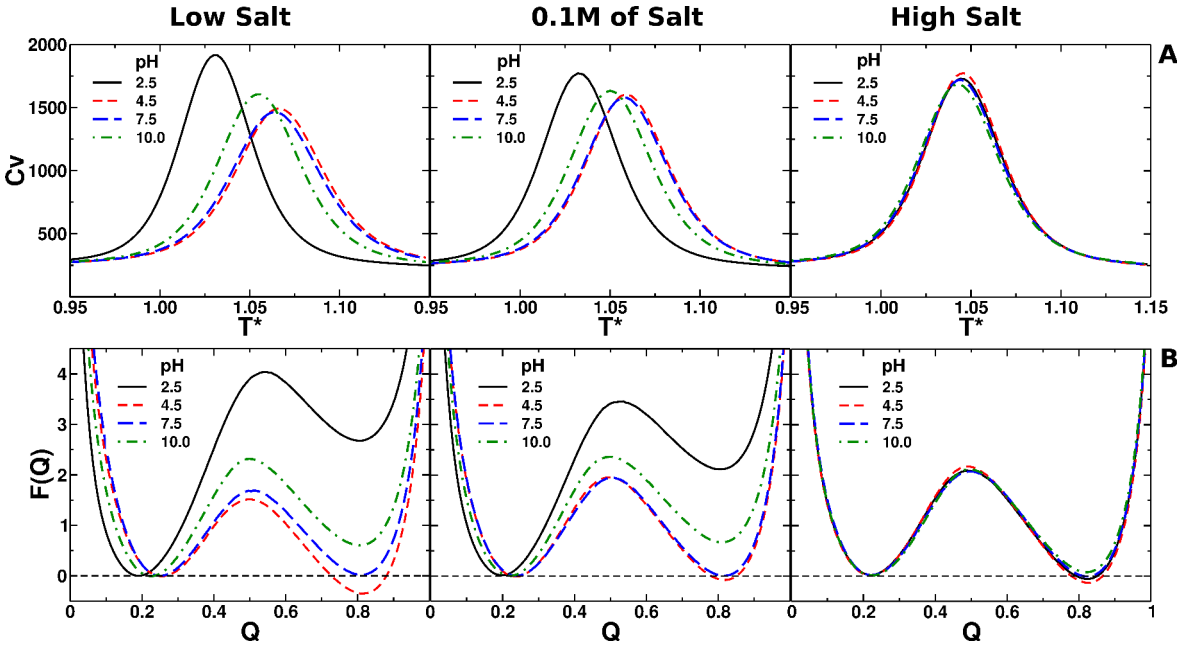


Figure 2.4: Thermodynamic properties of 1PGB-QDD folding. **A:** Heat capacities at constant volume (C_v) in low, intermediate, and high salt concentrations. Solid black lines for pH 2.5, dashed red lines for pH 4.5, dashed blue lines for pH 7.5, and dashed/doted green lines for pH 10.0. **B:** Free energy curves as a function of the reaction coordinate Q for the same three salt concentrations. All systems at T close to T_M of pH 7.5 in each ionic strength.

transition barrier between the folded and unfolded state in protein near T_M was not affected by salting, with the barrier close to 2.0 ϵ_C. In general, the thermodynamic properties calculated by CpHMD-SBM are in good agreement with the experimental results.

The comparison between experimental and computational T_M is presented in Figure 2.5. The general behavior T_M for computational results in each salt concentration shows a good level of agreement with the experimental data, with the T_M peak at pH 4.5 for all systems. At low pH, there is a decrease of T_M mostly in low and intermediate salt concentrations. In this case, both experiments and simulations agree that a high ionic strength contributes to 1PGB-QDD stability. However, when the protein has a pH higher than 4.5, the highest stability in the simulations occurs at low salt concentration, followed by intermediate and high salt concentrations. On the other hand, experimental results at pH 7.5 and 10.0 indicate that the electrostatic screening caused by a high salt concentration helps to stabilize 1PGB-QDD. This difference between experimental and computational results is caused by the electrostatic energy calculated in the simulation, and this interaction will be explored below.

The effect of electrostatic interaction and charge regulation on each folding stage of 1PGB-QDD can be evaluated by an analysis of Figure 2.6. This figure presents

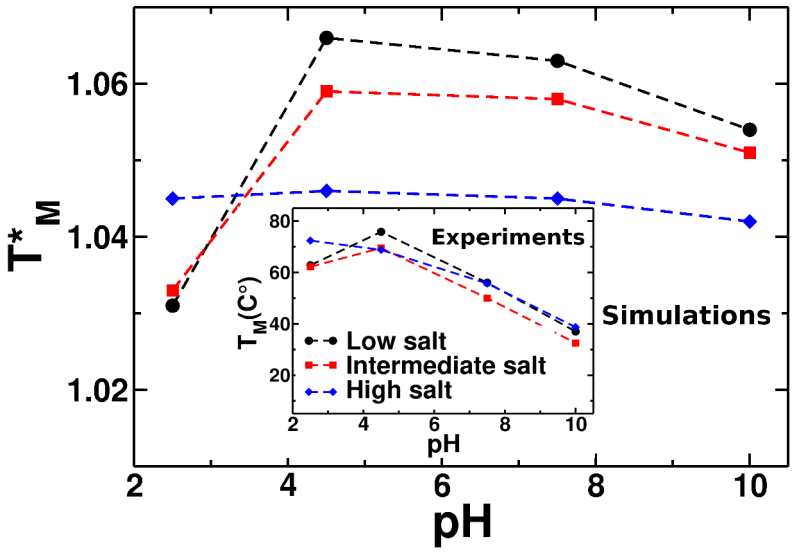


Figure 2.5: Values of melting temperature T_M^* in reduced temperature as a function of pH. Black circles for low salt, red squares for an intermediate ionic strength, and blue diamonds for high salt concentration. Inset graphic presents the experimental results of T_M as a function of pH in similar salt conditions to simulation (adapted from ref.[4]). Dashed lines connecting symbols help guide the eye.

a two dimensional map of the distribution of electrostatic energy as a function of Q. The net charge q in each highly populated region of distribution of electrostatic energy was also calculated and included in the Figure 2.6.

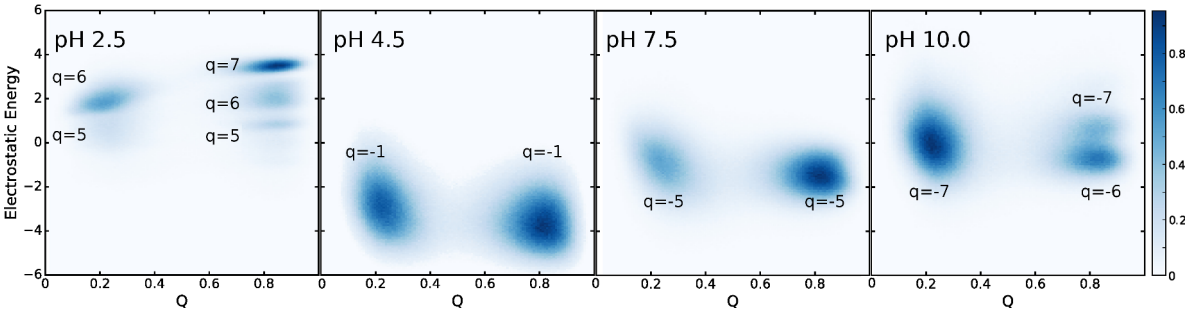


Figure 2.6: Two-dimensional map of the distribution of electrostatic energy as a function of the reaction coordinate Q for pH 2.5, 4.5, 7.5 and 10.0. The color map represents the probability distribution of the electrostatic energy and Q, normalized by its highest value. All the systems at $T \approx T_M$ and intermediate salt concentration. q represents the net charge of 1PGB-QDD in each minimum.

Previous results showed that the highest stability of 1PGB-QDD occurs at pH 4.5. This is because the protein in this condition has the lowest mean electrostatic energy $< E_{elec} >$ compared to the protein at pH 2.5, 7.5 and 10.0. For both folded and unfolded states, the $< E_{elec} >$ is near -4 ϵ_C and net charge $q = -1$, not favoring any state in comparison to each other. A similar discussion applies to the protein at pH 7.5: the folded and unfolded states have $< E_{elec} >$ close to -2 ϵ_C and $q = -5$

(experimental net charge[4] is -6). An increase in mean electrostatic energy promotes a loss of stability and is caused by this negative q, although, even with a high negative net charge, $< E_{elec} >$ is kept negative. This favorable energy explains the fact that low salt concentration promotes more stability in 1PGB-QDD at pH 7.5 than intermediate and high salt concentrations (Figure 2.5), which does not agree with the experimental results. Although the net charge is similar (experimental $q = -6$ and computational $q = -5$), the arrangement of charges in the C_α model can be different from the more realistic case, and creates more regions of favorable interaction. At pH 10.0 the same thing occurs: the protein has a high negative $q = -7$ or $q = -6$ and a negative value of $< E_{elec} >$ close to -1 ϵ_C.

Figure 2.6 also shows that the pH effect at pH 4.5 and 7.5 is strong enough to keep the protein net charge, irrespective of its conformational state. However, at the extreme pHs 2.5 and 10.0, some differences in electrostatic energy and net charge arise, depending on the folding stages. In these cases, the charge regulation is more evident, and the interaction between charged groups becomes fundamental for their protonation states. With the protein at pH 2.5, the 2D map presents two populated regions in the unfolded state, one with $< E_{elec} >$ close to 1 ϵ_C ($q = 5$) and other near 2.0 ϵ_C (q=6), the second being the most populated. In the folded state, in this same pH condition, three populated regions arise, the most populated with the highest $< E_{elec} >= 4\ \epsilon_C$ and $q = 7$, an intermediate region in $< E_{elec} >= 2\epsilon_C$ and $q = 6$, and the lowest populated region with $< E_{elec} >= 1\epsilon_C$ and $q = 5$. The loss of stability at this low pH condition may be associated with the fact that $< E_{elec} >$ of the unfolded protein is lower than $< E_{elec} >$ in folded state. Thus, the electrostatic interaction favors the unfolded state.

2.3.3 Comparison between computational and experimental pK_a

In this final section, the experimental pK_a and the pK_a calculated by TKSA (for folded state), and CpHMD (for folded and unfolded states) were compared. The computational pK_a was calculated fitting individual titration curves with the Henderson-Hasselbalch equation [17] and the results are presented in Table I. The errors in pK_a values were calculated using a set of three different simulations and an error lower than 0.1 pK_a unit was obtained for each ionizable residue. Thus, these error values were omitted from Table I.

Only differences higher than 1.0 pK_a unit between experimental and com-

Table I: Comparison between experimental [65] and computational pK_a values for the acidic residues in folded and unfolded 1PGB-QDD protein.

Residue	Experimental[a]		CpHMD[b]		TKSA[c]
	pK_a Folded	pK_a Unfolded	pK_a Folded	pK_a Unfolded	pK_a Folded
D08	4.9	3.7	3.5	3.5	4.2
E15	4.6	4.4	4.0	4.1	1.8
E19	3.9	4.5	4.0	4.4	4.0
D22	3.0	3.9	3.7	3.9	3.4
E27	4.8	4.4	3.5	3.7	0.6
D36	4.2	4.1	4.1	4.2	4.1
D37	6.5	4.2	4.1	4.3	4.0
D40	4.1	4.2	4.0	4.1	4.3
E42	4.8	4.4	4.6	4.7	5.0
D46	3.8	4.0	4.0	4.1	4.0
D47	3.1	3.9	3.6	3.8	2.9
E56	3.8	4.7	4.6	4.4	6.5
C-ter	3.1	3.6	**	**	3.5

[a]The experimental pK_a values for the acidic residues measured by Lindman et al.[65];
[b]The computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation, simulation using CpHMD.
[c]The computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation, simulation using TKSA.

putational values were considered significant [17]. Comparing the experimental pK_a of folded 1PGB-QDD protein and the values calculated using TKSA, only four residues presented a difference higher than 1.0 pK_a unit. This is a good level of agreement considering the simplicity of the model, which can output the results in less than one hour. E15 and E27 have a downshift in computational pK_a that may be caused by the overrated interaction calculated in this model that these residues make with K04 and K21, respectively. D37 has a large upshift in its experimental pK_a, which may be caused by desolvation effects associated with side-chain burial [65] that TKSA is not taking into account sufficiently. Finally, the residue E56 has two charged groups that cause a large electrostatic repulsion in the TKSA model; this repulsion may explain the high upshift in its computational pK_a value.

Using CpHMD, the comparison with the experimental results is even better, with only three residues of folded 1PGB-QDD presenting more than 1.0 pK_a unit of difference between the simulation and the experimental pK_a values. D27 has a similar

downshift in pK_a values to that obtained using TKSA; the reason for this difference is the high interaction that this residue makes with K04. In the case of D08, the downshift of 1.4 pK_a units obtained in computational results could be caused by a model limitation. Considering the native structure of the protein, the distance between the C_α of D08 and C_α K13 is near 5.1Å, whereas the distance between the carboxyl group of D08 and ammonium group of K13 is close to 8.3Å. This difference in the distances promotes a more favorable interaction between these residues in CpHMD, resulting in this disagreement between experimental and simulation results. In residue D37, as previously stated, desolvation effects have a strong impact on its experimental pK_a value and the CpHMD model does not treat this effect adequately. When the pK_a of the unfolded protein was considered, the agreement between experimental values and CpHMD values is remarkable: in no case does the difference between these values exceed 1.0 pK_a unit.

2.4 Conclusions

In the present study, the protein 1PGB-QDD was studied by two coarse-grained models with the objective of evaluating the pH and salting effect on protein stability. The results of this study were compared with experimental results. The first model tested here was TKSA, which analyses the effect of each ionizable residue in the electrostatic free energy of the protein in its native state. Using this simple model, it was possible to quantify the importance of each charged residue in 1PGB-QDD stability. The comparison between TKSA calculations and experimental results also indicates that the driving force of this stability is the electrostatic interaction. The second model used in this study was CpHMD-SBM, which was able to explore the protein folding process in constant pH [17]. With CpHMD-SBM simulations, the folded and unfolded states and the transition between these states could be studied. The pH effect in 1PGB-QDD stability explored by simulation was found to be in good agreement with the experimental data. Similarly, the salt effect was evaluated, and the simulations obtained the same general behavior in protein stability as the experiments. Finally, pK_a was calculated using TKSA in the folded protein and CpHMD-SBM for folded and unfolded 1PGB-QDD. The pK_a values calculated by the two models are near experimental values. Therefore, both methods were able to describe the effects of charge-charge interaction, pH and ionic strength on protein, despite the high level of simplification.

2.5 Acknowledgments

VMO was supported by the National Council for Scientific and Technological Development (CNPq - Grant Process No. 141985/2013-5). VGC was funded by Grant 2016/13998-8, São Paulo Research Foundation (FAPESP) and Higher Education Personnel Improvement Coordination (CAPES). FBS was supported by the Higher Education Personnel Improvement Coordination (CAPES). DLZC acknowledges the São Paulo Research Foundation (FAPESP, Grant Process No. 2013/13151-7) for the financial support. VBPL was supported by the National Council for Scientific and Technological Development (CNPq) and FAPESP Grant 2014/06862-7 and 2016/19766-1. We also thank the Center for Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) for computational resources.

References

[1] M Roca, H Liu, B Messer, and A Warshel. On the relationship between thermal stability and catalytic power of enzymes. *Biochemistry*, 46(51):15076–15088, 2007.

[2] J.W. Bye, L Platts, and R.J. Falconer. Biopharmaceutical liquid formulation: a review of the science of protein stability and solubility in aqueous environments. *Biotech. Lett.*, 36(5):869–875, 2014.

[3] K.V.M. Huber, K.M. Olek, A.C. Müller, C.S.H. Tan, K.L. Bennett, J Colinge, and G Superti-Furga. Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. *Nature Methods*, 12:1055–1057, 2015.

[4] Stina Lindman, Wei-Feng Xue, Olga Szczepankiewicz, Mikael C. Bauer, Hanna Nilsson, and Sara Linse. Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophys. J.*, 90(8):2911–2921, April 2006.

[5] M Yamasaki and H. Yano. Differential scanning calorimetric studies on bovine serum albumin: I. effects of ph and ionic strength. *Int. J. Biol. Macr.*, 12(4):263–268, 1990.

[6] K.S. Kim, S Kim, H.J. Yang, and D.Y. Kwon. Changes of glycinin conformation due to ph, heat and salt determined by differential scanning calorimetry and circular dichroism. *Int. J. Food Sci. Tech.*, 39(4):385–393, 2004.

[7] J.W. Bye and R.J. Falconer. Thermal stability of lysozyme as a function of ion concentration: A reappraisal of the relationship between the hofmeister series and protein stability. *Prot. Sci.*, 22(11):1563–1570, 2013.

[8] Swarnendu Tripathi, Angel E. Garcìa, and George I. Makhatadze. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J. Phys. Chem. B*, 119(41):13103–13112, October 2015.

[9] M Wunderlich, A Martin†, and F.X. Schmid. Stabilization of the cold shock protein cspb from bacillus subtilis by evolutionary optimization of coulombic interactions. *J. Mol. Biol.*, 2005.

[10] E Papaleoa, M Olufsena, L De Gioiab, and B.O. Brandsdala. Optimization of electrostatics as a strategy for cold-adaptation: A case study of cold- and warm-active elastases. *J. Mol. Graph. Mod.*, 26(1):93–103, 2007.

[11] Mihir S. Date and Brian N. Dominy. Modeling the influence of salt on the hydrophobic effect and protein fold stability. *Commun. Comput. Phys.*, 13(1):90–106, 2013.

[12] Manuel P. Luitz, Rainer Bomblies, Evelyn Ramcke, Aymelt Itzen, and Martin Zachariasa. Adenylylation of tyr77 stabilizes rab1b gtpase in an active state: A molecular dynamics simulation analysis. *Sci Rep.*, 6:19896, 2016.

[13] Ji-Guo Su, Xiao-Ming Han, Shu-Xin Zhao, Yan-Xue Hou, Xing-Yuan Li, Li-Sheng Qi, and Ji-Hua Wang. Impacts of the charged residues mutation s48e/n62h on the thermostability and unfolding behavior of cold shock protein: insights from molecular dynamics simulation with gö model. *J. Mol. Mod.*, 22:91, 2016.

[14] MA Coronado, IP Caruso, VN De Oliveira, VG Contessoto, VB Leite, RK Arni, and RJ Eberle. Cold shock protein a from corynebacterium pseudotuberculosis: Role of electrostatic forces in the stability of the secondary structure. *Protein Pept. Lett.*, 2017.

[15] Ariel Azia and Yaakov Levy. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.*, 393(2):527–542, October 2009.

[16] Ohad Givaty and Yaakov Levy. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, 385(4):1087–1097, January 2009.

[17] Vinícius G. Contessoto, Vinícius M. de Oliveira, Sidney J. de Carvalho, Leandro C. Oliveira, and Vitor B. P. Leite. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.*, 12(7):3270–3277, July 2016.

[18] Jana Khandogin and Charles L Brooks. Constant ph molecular dynamics with proton tautomerism. *Biophys. J.*, 89(1):141–157, 2005.

[19] John Mongan and David A. Case. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.*, 15(2):157–163, April 2005.

[20] Charles Tanford and John G Kirkwood. Theory of protein titration curves. i. general equations for impenetrable spheres. *J. Am. Chem. Soc.*, 79(20):5333–5339, 1957.

[21] B. Ibarra-Molero, V. V. Loladze, G. I. Makhatadze, and J. M. Sanchez-Ruiz. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry*, 38(25):8138–8149, June 1999.

[22] George I. Makhatadze, Vakhtang V. Loladze, Dmitri N. Ermolenko, XiaoFen Chen, and Susan T. Thomas. Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.*, 327(5):1135–1148, April 2003.

[23] Serge E. Permyakov, George I. Makhatadze, Rikard Owenius, Vladimir N. Uversky, Charles L. Brooks, Eugene A. Permyakov, and Lawrence J. Berliner. How to improve nature: study of the electrostatic properties of the surface of alpha-lactalbumin. *Protein Eng., Des. Sel.*, 18(9):425–433, September 2005.

[24] Samantha S. Strickler, Alexey V. Gribenko, Alexander V. Gribenko, Timothy R. Keiffer, Jessica Tomlinson, Tracey Reihle, Vakhtang V. Loladze, and George I. Makhatadze. Protein stability and surface electrostatics: a charged relationship. *Biochemistry*, 45(9):2761–2766, March 2006.

[25] Alexey V. Gribenko, Mayank M. Patel, Jiajing Liu, Scott A. McCallum, Chunyu Wang, and George I. Makhatadze. Rational stabilization of enzymes by computational redesign of surface charge–charge interactions. *Proc. Natl. Acad. Sci. USA*, 106(8):2601–2606, February 2009.

[26] Lars Björck, William Kastern, Gunnar Lindahl, and Kristina Widebäck. Streptococcal protein G, expressed by streptococci or by Escherichia coli, has separate binding sites for human albumin and IgG. *Mol. Immunol.*, 24(10):1113–1122, October 1987.

[27] U. Sjöbring, L. Björck, and W. Kastern. Protein g genes: structure and distribution of IgG-binding and albumin-binding domains. *Mol. Immunol.*, 3(3):319–327, mar 1989.

[28] L. Björck and G. Kronvall. Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. *J. Immunol.*, 133(2):969–974, August 1984.

[29] K. J. Reis, E. M. Ayoub, and M. D. Boyle. Streptococcal Fc receptors. I. Isolation and partial characterization of the receptor from a group C streptococcus. *J. Immunol.*, 132(6):3091–3097, June 1984.

[30] Sebastian Kmiecik and Andrzej Kolinski. Folding pathway of the b1 domain of protein g explored by multiscale modeling. *Biophys. J.*, 94(3):726–736, feb 2008.

[31] Patrick Alexander, Stephen Fahnestock, Timothy Lee, John Orban, and Philip Bryan. Thermodynamic analysis of the folding of the streptococcal protein g IgG-binding domains b1 and b2: why small proteins tend to have high denaturation temperatures. *Biochemistry*, 31(14):3597–3603, apr 1992.

[32] Patrick Alexander, John Orban, and Philip Bryan. Kinetic analysis of folding and unfolding the 56 amino acid igg-binding domain of streptococcal protein g. *Biochemistry*, 31(32):7243–7248, 1992.

[33] A. Gronenborn, D. Filpula, N. Essig, A Achari, M Whitlow, P. Wingfield, and G. Clore. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein g. *Science*, 253(5020):657–661, 1991.

[34] SJ Shire, GIH Hanania, and FR N_ Gurd. Electrostatic effects in myoglobin. hydrogen ion equilibriums in sperm whale ferrimyoglobin. *Biochemistry*, 13(14):2967–2974, 1974.

[35] Charles Tanford and Robert Roxby. Interpretation of protein titration curves. application to lysozyme. *Biochemistry*, 11(11):2192–2198, 1972.

[36] William H Orttung. Proton binding and dipole moment of hemoglobin. refined calculations. *Biochemistry*, 9(12):2394–2402, 1970.

[37] John G Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.*, 2(7):351–361, 1934.

[38] Fernando LuífS B Da Silva, Bo Jönsson, and Robert Penfold. A critical investigation of the tanford-kirkwood shceme by means of monte carlo simulations. *Protein Sci.*, 10(7):1415–1425, 2001.

[39] James J Havranek and Pehr B Harbury. Tanford–kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA*, 96(20):11145–11150, 1999.

[40] Donald Bashford and Martin Karplus. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.*, 95(23):9556–9561, 1991.

[41] C Clementi, H Nymeyer, and J N Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, 298(5):937–953, May 2000.

[42] Jeffrey K. Noel, Paul C. Whitford, Karissa Y. Sanbonmatsu, and José N. Onuchic. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.*, June 2010.

[43] Vinícius G. Contessoto, Debora T. Lima, Ronaldo J. Oliveira, Aline T. Bruni, Jorge Chahine, and Vitor B. P. Leite. Analyzing the effect of homogeneous frustration in protein folding. *Proteins: Struct., Funct., Bioinf.*, 81(10):1727–1737, October 2013.

[44] Matheus R. de Mendonca, Leandro G. Rizzi, Vinicius Contessoto, Vitor B. P. Leite, and Nelson A. Alves. Inferring a weighted elastic network from partial unfolding with coarse-grained simulations. *Proteins: Struct., Funct., Bioinf.*, 82(1):119–129, January 2014.

[45] Paulo Ricardo Mouro, Vinícius de Godoi Contessoto, Jorge Chahine, Ronaldo Junio de Oliveira, and Vitor Barbanti Pereira Leite. Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophys. J.*, 111(2):287–293, July 2016.

[46] Franco O. Tzul, Katrina L. Schweiker, and George I. Makhatadze. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. USA*, 112(3):E259–E266, January 2015.

[47] Magnus Ullner, Cliff E. Woodward, and Bo Jönsson. A Debye-Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.*, 105:2056–2065, August 1996.

[48] António M. Baptista, Vitor H. Teixeira, and Cláudio M. Soares. Constant-ph molecular dynamics using stochastic titration. *J. Chem. Phys.*, 117(9):4184–4200, September 2002.

[49] John Mongan, David A. Case, and J. Andrew McCammon. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.*, 25(16):2038–2048, December 2004.

[50] Sarah L. Williams, Patrick G. Blachly, and J. Andrew McCammon. Measuring the successes and deficiencies of constant pH molecular dynamics: A blind prediction study. *Proteins: Struct., Funct., Bioinf.*, 79(12):3381–3388, aug 2011.

[51] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, June 1953.

[52] Axel Arnold, Olaf Lenz, Stefan Kesselheim, Rudolf Weeber, Florian Fahrenberger, Dominic Roehm, Peter Košovan, and Christian Holm. ESPResSo 3.1: Molecular Dynamics Software for Coarse-Grained Models. In Michael Griebel and Marc Alexander Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VI*, number 89 in Lecture Notes in Computational Science and Engineering, pages 1–23. Springer Berlin Heidelberg, 2013.

[53] H. J. Limbach, A. Arnold, B. A. Mann, and C. Holm. ESPResSo—an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174(9):704–727, May 2006.

[54] Alan M. Ferrenberg and Robert H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61(23):2635, December 1988.

[55] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. THE weighted histogram analysis method for free[U+2010]energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13(8):1011–1021, October 1992.

[56] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. Multidimensional free[U+2010]energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16(11):1339–1350, November 1995.

[57] V Sobolev, A Sorokine, J Prilusky, E E Abola, and M Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15(4):327, 1999.

[58] Benjamin Webb and Andrej Sali. Comparative Protein Structure Modeling Using MODELLER. In *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., 2002.

[59] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13):1605–1612, October 2004.

[60] Charles Simon Bond and Alexander Wolfgang Schüttelkopf. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr.*, 65(Pt 5):510–512, May 2009.

[61] Tsodikov, M. Record, and Y. Sergeev. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.*, 23:600–609, 2002.

[62] Beatriz Ibarra-Molero and Jose M. Sanchez-Ruiz. Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins. *J. Phys. Chem. B*, 106(26):6609–6613, July 2002.

[63] Vakhtang V. Loladze and George I. Makhatadze. Energetics of charge-charge interactions between residues adjacent in sequence. *Proteins: Struct., Funct., Bioinf.*, 79(12):3494–3499, December 2011.

[64] George I Makhatadze, Vakhtang V Loladze, Alexey V Gribenko, and Maria M Lopez. Mechanism of Thermostabilization in a Designed Cold Shock Protein with Optimized Surface Electrostatic Interactions. *J. Mol. Biol.*, 336(4):929–942, February 2004.

[65] Stina Lindman, Mikael C Bauer, Mikael Lund, Carl Diehl, Frans AA Mulder, Mikael Akke, and Sara Linse. pk a values for the unfolded state under native conditions explain the ph-dependent stability of pgb1. *Biophys. J.*, 99(10):3365–3373, 2010.

Capítulo 3

Proteína *Cold shock* A - CspA

Cold shock protein A from *Corynebacterium pseudotuberculosis*: Role of electrostatic forces in the stability of the secondary structure

Monika A. Coronado, Icaro P. Caruso, Vinícius M. de Oliveira, Vinícius G. Contessoto, Vitor B.P. Leite, Liege A. Kawai, Raghuvir K. Arni, Raphael J. Eberle

*This article has been published in Protein and Peptide Letters. Vol. 24(4), 358-367, 2017. DOI 10.2174/0929866524666170207153808.

Abstract

The aim of this work was to understand the conformational stability of Cold shock protein A (CspA) from *C. pseudotuberculosis* (Cp), a nucleic acid binding protein, related to the contribution of pH and salt (NaCl). Differential scanning calorimetry and CD spectroscopy in combination with computational analysis were used to investigate the pH and NaCl influence to specify affected amino acids. Our approach identified a sodium binding site in CpCspA and at pH 8.0 a significant reduction in the β-sheet content was observed decreasing protein stability. The computational analyses identified His30 and His65 with the largest charge shifts. The identified ion binding is important for the conformational stability of the protein. His30 and His65 are part of hydrogen bonds and salt bridges that stabilize the protein fold. Both residues are strongly affected by pH variations and modify the protein conformation. The combination of experimental results together computational approaches generate new insights to interpret the behavior of proteins in different environment. Our results provide a new view of the structural elements responsible to control the conformational stability of CpCspA.

Keywords: Cold shock protein, *C. pseudotuberculosis*, ion binding, histidine, pH, secondary structure.

3.1 Introduction

Bacteria respond to changes in environmental parameters such as nutrient levels, oxygen availability, osmotic stress and temperature by triggering the activation of a specific set of genes that upregulate the production of cold shock proteins (Csps) and subsequently, by transiently restricting the production of non-cold-shock proteins. [1-3].

The initial phase of the cold shock response, the acclimation phase, is characterized by high levels of Csp expression which regulates the synthesis of other cold-stress inducible proteins [4,5]. At the end of the acclimation phase the synthesis of cold shock proteins decreases and remains constant and the levels of protein production are re-established [6].

The primary function of Csps is the enhancement of DNA transcription to support the expression of other cold shock induced genes [5,7,8]. This is achieved by binding of the Csps to short stretches of single stranded DNA with high affinity and specificity [9-12]. However, Csps have also been reported to bind unspecifically to ssRNA, leading to the suggestion that cold shock proteins may also function as RNA chaperones [13], and prevent the formation of cold-induced mRNA secondary structures [14]. Both mechanisms help to preserve cell viability during the cold shock process and to restore normal cell function.

Csps are encountered in psychrophilic, mesophilic, thermophilic and hyperthermophilic bacterial species and share highly conserved sequences and structures. The three-dimensional structures of a number of Csps have been determined, including *E. coli* CspA, *B. subtilis* CspB, and *B. caldolyticus* CspB and they possess a typical β-barrel fold, which consist of a three stranded N-terminal sheet (β1-β2-β3) and a two stranded C-terminal sheet (β4-β5). Csps contain two nucleic acid binding motifs (RNP1 and RNP2), which principally consist of aromatic and basic amino acid residues contributing to the binding of single stranded nucleic acids [15-18].

Notwithstanding, the high degree of conservation of sequence and structural features of Csps, studies have revealed considerable differences in the thermal stabilities of these proteins. *E. coli* CspA unfolds at 60 °C [16,19] whereas; *B. subtilis* CspB unfolds at 52 °C [10]. The thermophilic *B. caldolyticus* Csp exhibits a melting temperature (T_m) at 77 °C [17], twenty five degrees higher than the T_m of *B. subtilis* CspB. The proteins unfold reversibly in a monomolecular (N) $\leftrightarrow$ unfolded (U) two state reaction [20] and theoretical and experimental investigations have examined the role of point mutations [21], increasing salt concentration [17,22] and pH [23,24] on the

thermal stability of Csps.

Corynebacterium pseudotuberculosis (*C. pseudotuberculosis*) is a member of the heterogeneous CMNR-group of pathogens (Corynebacterium, Mycobacterium, Nocardia, and Rhodococcus), a cluster of gram-positive bacteria species [25]. *C. pseudotuberculosis* is the causative agent of caseous lymphadenitis (CLA), a disease encountered in sheep, goats, and equids (ulcerative lymphangitis) and cattle (cutaneous excoriated granulomas) and leads to drastically reduced yields of wool and milk, weight loss, carcass condemnation and eventually results the sacrifice of infected animals thus resulting in considerable economic loss all over the world [26,27] and isolated cases of infection in humans have been reported [28]. As a facultative intracellular parasite this bacterium is capable of survival and growth in macrophages and is thus able to evade detection by the host immune system [29]. Although *C. pseudotuberculosis* is a mesophilic bacterium, growing optimally at 37 °C, its genome includes *csp* genes, coding for cold shock proteins (Csp) and the expression, purification and characterization of *C. pseudotuberculosis* CpCspA has been reported [30].

Circular dichroism spectroscopy (CD) and differential scanning calorimetry (DSC) combined with computational analysis was utilized to determine the key factors responsible for secondary structure variations observed by varying the pH and salt concentration and the effect on the thermal stability of CpCspA.

3.2 Materials and Methods

3.2.1 In silico analysis

Sequence alignments of different Csp proteins were performed with the ClustalW2 online tools [31]. The atomic coordinates of CspA from *E. coli* (PDB: 2L15) were used as a template in comparative modeling by the satisfaction of spatial restraints as implemented in the program Modeller 9v13 [32] to model the structure of CpCspA (Uniprot: D9Q632). For the presentation of the sodium binding site CSP of *B. caldolyticus* (PDB: 1C9O) was used to model the structure of CpCspA.

For the prediction of salt bridges in the CpCspA homology model the ESBRI web server was used [33] and for the prediction of hydrogen bonds PDBsum was utilized [34].

3.2.2 Recombinant expression and purification of the CpCspA

CpCspA was cloned, expressed and purified as reported previously [30].

3.2.3 Circular dichroism spectroscopy (CD)

CD measurements were conducted using a Jasco J-701 spectropolarimeter (Jasco, USA). Far-UV spectra were measured in the 200 to 260 nm range. Protein concentration used for the far-UV CD measurements was 20 M. Cells of 0.1 cm path length were used for the measurements of the far-UV spectra. 15 repeat scans were obtained for each sample and five scans were conducted to establish the respective baseline. The averaged baseline spectrum was subtracted from the averaged sample spectrum. To investigate the influence of pH a protein sample was dialyzed against a 1:1000 higher volume of 20 mM K_2HPO_4/KH_2PO_4 for the pH range 8.0 to 6.0. Results are expressed as molar ellipticity $[\theta]$ (mdeg cm^2 $dmol^{-1}$) according to the equation:

$$[\theta]_\lambda = \theta/(c * l * 10 * n) \tag{3.1}$$

With θ being the measured ellipticity at wavelength λ (in deg), c is the protein concentration (in mol/L), l the path length of the cell (in cm) and n the number of amino acids. Data were analyzed with the program CDpro [35].

3.2.4 Differential scanning calorimetry (DSC)

The DSC experiments were performed using a N-DSC III (TA Instruments, USA) in the 20 – 90 °C range and a scan rate of 1 °C/min. For the DSC experiments the protein concentration was 109.6 M (1 mg/ml), the buffer for the measurements contained 50 mM K_2HPO_4/KH_2PO_4. Two conditions of pH and salt concentration were used: 1) pH range 8.0 – 6.0 without NaCl and 2) pH 7.5 with NaCl range 0 – 2.0 M. Both the calorimeter cells were loaded with the buffer solution, equilibrated at 20 °C for 10 min and scanned repeatedly as described above until the baseline was stable and reproducible. The sample cell was subsequently loaded with the protein solution and scanned in the same way. Baseline correction was conducted by subtracting the 'buffer vs. buffer' scan from corresponding 'protein vs. buffer' scan. The measurements were repeated twice.

For the estimation of the number of Na^+ ions released upon unfolding we used the following equation:

$$dTm/d(ln[a]) = -\Delta n(R.Tm^2/\Delta H(Tm)) \tag{3.2}$$

where Tm is the melting temperature, [a] is the ionic activity of sodium chloride, Δn is denaturational change in the number of bound ions, R is the gas constant, and ΔH(Tm) is the transition enthalpy change at the transition temperature Tm[36,37].

3.2.5 Computational Methods

A constant-pH molecular dynamic method [38] was adopted for protein folding studies which combines a standard molecular dynamic simulation using a structure-based model [39,40] with the Metropolis Monte Carlo method to sample the protein protonation states [41]. The simulation is conducted over a given number of molecular dynamics steps and the titratable residue is randomly chosen. The modification of its protonation state is accepted or rejected according to the Metropolis criterion.

Cα - structure-based model (protein folding molecular dynamic)

In this model the cold shock protein is modeled using a Cα - structure-based model (Cα-SBM) level of simplification [39,40]. The amino acid residues are represented by beads in the positions of the α-carbon atoms and the Hamiltonian which provides the protein energy interaction based on the geometry of its previously determined native state. In this first approach, the protein charge is not taken into consideration and the potential energy surface reaches its minimum in this reference state. The electrostatic interactions are introduced through the inclusion of point charges at the centers of the beads. The electrostatic interaction occurs via the Debye-Huckel potential.

Monomer's titration (Metropolis Monte Carlo)

The Monte Carlo method is implemented to protonate or deprotonate a selected titratable residue [42]. The Metropolis criterion is adopted to rule the acceptance of the monomer's titration. The values of the residues valence are changed (+1$\Leftrightarrow$0 for Arg/Lys/His and 0$\Leftrightarrow$-1 for Asp/Glu). The energy variation used in Metropolis criterion is given by the equation:

$$\Delta U = \Delta U_{el} + \xi k_B T \ln(10)(pH - pKa) \tag{3.3}$$

where ΔUel is the electrostatic energy variation between the charge configuration after and before the titration [43]. The constant ξ is 1 or -1 for acid and basic residues,

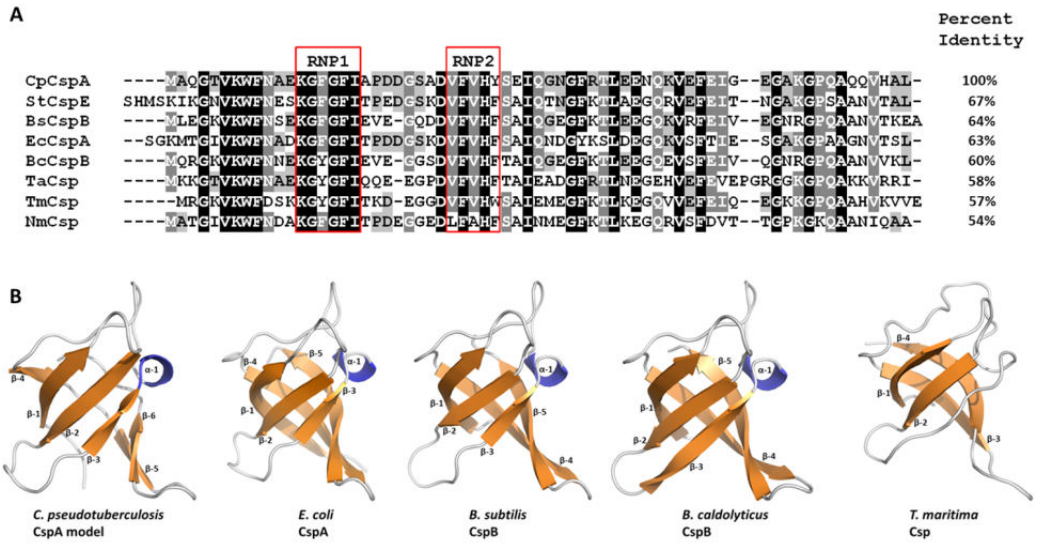


Figure 3.1: Sequence and structural comparison of different bacterial Csps. (A) Sequence alignment of CspA from *C. pseudotuberculosis* (GenBank:ADL20127.1), CspE from *S. typhimurium* (GenBank:AE006468.1; PDB:3I2Z), CspB from *B. subtilis* (GenBank:CAA42235.1; PDB:2I5M), CspA from *E. coli* (GenBank:AAA23617.1; PDB:2L15), CspB from *B. caldolyticus* (GenBank:CAA51790.1; PDB:1C9O), Csp from *T. aquaticus* (GenBank:EED09135.1; PDB:1MO0), Csp from *T. maritima* (GenBank:AAD36750.1; PDB:1G6P) and Csp from *N. meningitis* (GenBank:AAF41249.1; PDB:3CAM). (B) structural comparison of *C. pseudotuberculosis* CspA homology model, *E.coli* CspA and *B. subtilis* CspB (mesophilic), *B. caldolyticus* CspB (thermophilic) and *T. maritima* Csp (hyperthermophilic).

respectively. The intrinsic pKa values of the residues are: Asp (4.0), Glu (4.5), His (6.3), Lys (10.6) and Arg (12.0) [44] (Data S2 show simulation details and are deposited in supplementary material).

3.3 Results and Discussion

3.3.1 Sequence analysis and molecular modeling of CpCspA

A Blast search of the CpCspA sequence against the atomic coordinates currently deposited with the Protein Data Bank (http://www.rcsb.orgwww.rcsb.org) indicated 67% identity with CspE (PDB: 3I2Z) from *S. typhimurium* (Fig. 1A). The Csp sequence alignment across the bacterial genera highlights the conservation of the nucleotide binding sites RNP1 and RNP2 [45] which binds nucleic acids [10] and the exposed aromatic residues are crucial for the conformational stability of Csps [46].

The Csp structure is typically composed of five antiparallel β-strands that form a Greek-key β-barrel [15,16,18] as illustrated in figure 1B and the structural core is

highly conserved in the Csp protein family. The numbers and lengths of the β-strands in the Csps from different bacterial species as well as the number of hydrogen bonds and salt bridges varies and are probably the key factors for the observed differences in the thermal stability of these proteins [18,47-49].

3.3.2 Identification of the sodium binding site of CpCspA

The binding capacity for different monovalent salts such as NaCl, KCl or LiCl and their role in the stability of Csp proteins from *B. caldolyticus* has been demonstrated [17]. Salts may shield ionic interactions affecting macromolecular stability as well as intermolecular binding reactions and have the ability to modify Hofmeister effects and indirect salt effects on water activity. These effects may alter the hydrophobic force by increasing the surface tension of the solvent and stabilize the protein conformation [22], or they may stabilize peptide dipoles through specific ionic interactions [50,51]. Additionally, salts modify the ionic strength of the protein solution [50,52] and their stabilizing or destabilizing effects depends on the specific charge distribution within the protein [53].

To study the effect of NaCl on the thermal stability of CpCspA, differential scanning calorimetry (DSC) experiments were performed with variable salt concentrations. The structural database was searched to identify CSP structures with ligated ions, the sequences and the structure of these proteins were compared with the sequence and the modeled structure of CpCspA to identify potential ion binding sites that have an influence on protein stability.

DSC experiments revealed an increase in the T_m of CpCspA with increasing NaCl concentration (Fig. 2A). The difference in the melting points of the protein between the two extreme levels of NaCl concentration is approximately 13 °C.

The determination of the number of Na^+ ions released upon thermal unfolding of CspA, resulted in Δn=-0,7 and was approximated to 1.0 Na^+ ion per Csp molecule (Fig. 2B). Interestingly, at a NaCl concentration higher then 1.0 M the slope of the straight change and one Csp molecule release around 3.0 Na^+ ions. Proteins are stabilized by high concentrations of strongly hydrated anions [52], like NaCl, these ions can interact with the hydration water shell surrounding the protein molecule [54] and can be released during the thermal denaturation process.

The search of the PDB database encountered a CSP protein structure from *B. caldolyticus* (PDB: 1C9O), which possess a Na^+ binding site in each monomer coordinated by Val20 and Gly23 [17]. Sequence and model comparisons between CpCspA

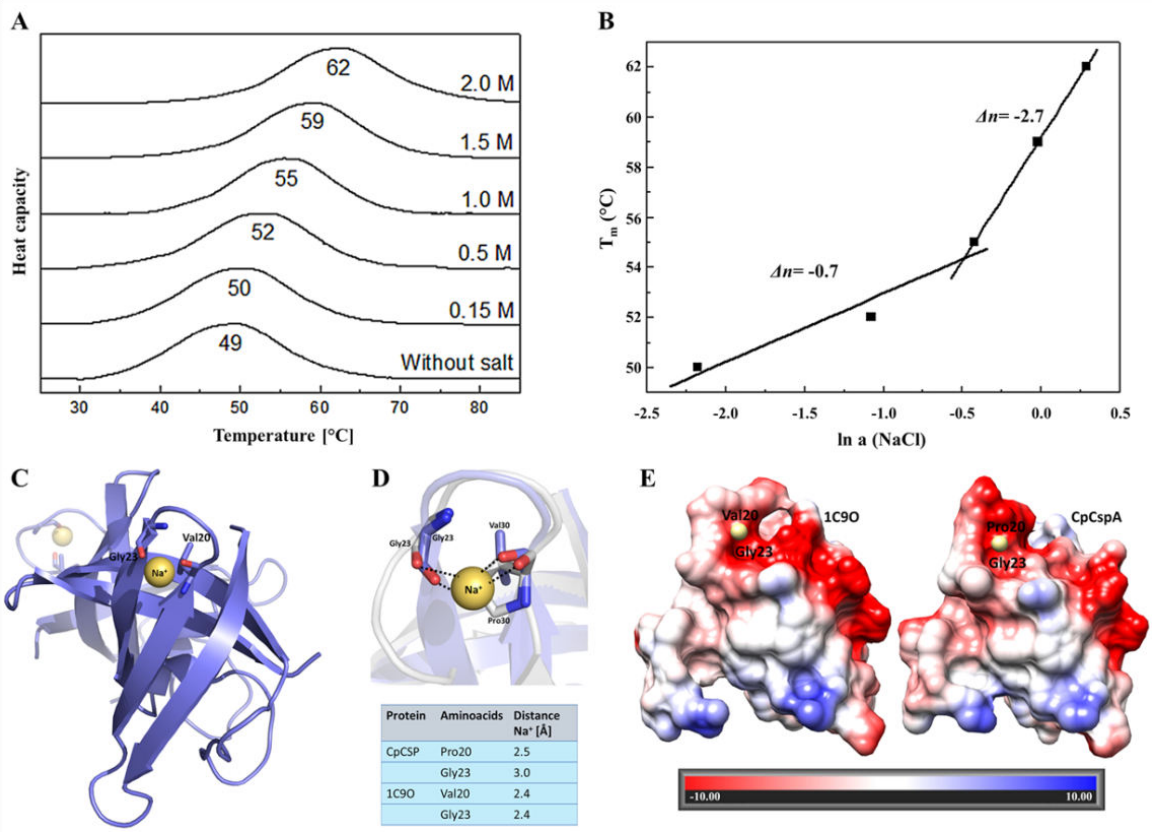


Figure 3.2: The influence of NaCl on the thermal stability of CpCspA and identification of a sodium binding site in CpCspA. (A) Determination of T_m under increasing NaCl concentrations. (B) Estimation of the number of Na^+ ions released upon thermal unfolding of CspA. (C) *B. caldolyticus* protein structure (PDB:1C9O), each monomer of the protein structure contain a Na^+ binding site, constructed by Val20 and Gly23. (D) Structural comparison of CpCspA homology model (grey) and 1C9O (blue), zoom of the Na^+ binding area showing the ion coordination. (E) Coulombic surfaces of *B. caldolyticus* (1C9O) and the CpCspA homology model, Na^+ ion in light yellow.

and 1C9O showed that Val20 is exchanged for Pro20 in CpCspA and that Gly23 is conserved (Fig. 2C and 2D). Structural overlays of the two proteins revealed that the O1 atom, which is important for Na^+ binding to Pro20 exhibits bond distances to Na^+ in the case of Val20 (Fig. 2D).

The interactions between Na^+, Val20 and Gly23 observed in 1C9O may serve as a stabilizing anchor of the β-hairpin. The identification of a similar binding area in CpCspA suggests Na^+ binding in this region which contributes to thermal stabilization of the protein as observed in DSC experiments upon increasing the NaCl concentration. The Na^+ binding site coordinated by Val20/Pro20, Gly23 is negatively charged in 1C9O and the CpCspA homolgy model (Fig. 2E), and form the basis of interaction with the positively charged sodium ion.

3.3.3 Impact of the pH on the secondary structure and thermal stability of CpCspA

Beside salt concentration, the pH of protein solution also has a significant influence on the ionic networks present on the protein surface and charged amino acids which influence the protein stability.

Conformational changes of CpCspA through pH titrations were monitored using far-UV CD spectroscopy (260-200 nm). The CD spectrum exhibits the CpCspA typical maxima at 220 nm and the minima at 205 nm (Fig. 3A) which are in agreement with the results of Lindae *et al.* [30] and indicates the presence of high β-strand content in the secondary structure of CpCspA.

The investigations of different pHs on electrostatic interactions of the protein surface and their stabilizing effect using CD and DSC experiments are shown in Fig. 3A and 3B.

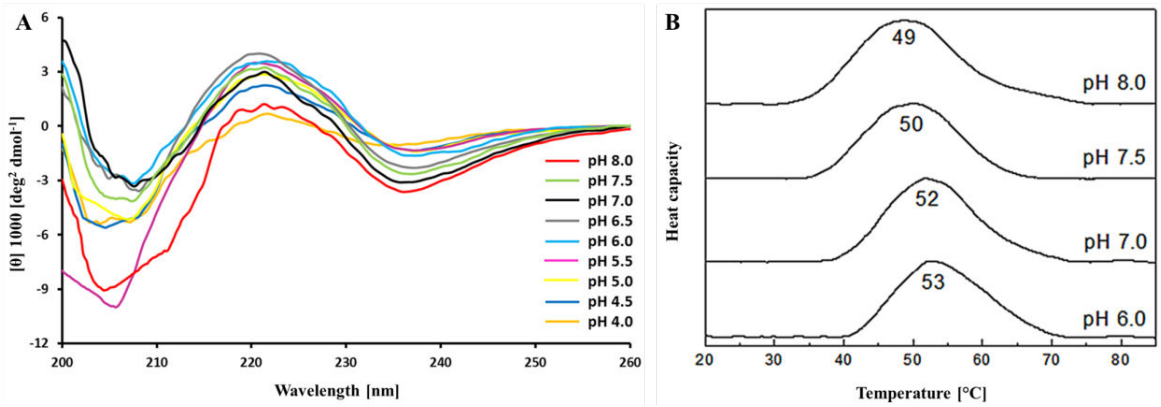


Figure 3.3: Biophysical investigation of CpCspA at different pHs. (A) far-UV CD spectrum of CspA from 200 to 260 nm wavelength; pH 8.0 to pH 6.0 dialyzed for 24h. (B) DSC thermogram of CpCspA, the thermogram indicates a reduction in T_m with increasing pH.

The results of the CD experiments indicate significant changes in the secondary structure of CpCspA at pH 8.0. Analysis of the CD results with the program CDpro [35] indicated a reduction of the β-strand content by about 30 % compared with the secondary structure composition of the protein at pH 6.0 (Fig. 3A). The secondary structure content of all tested pH for CpCspA summarized in tableS1 in the supplementary material.

The pH change affects the electrostatic interactions that stabilize the protein structure; especially the charges on ionizable groups and peptide dipoles [55]. Examples for ionizable groups in proteins are α-carboxyl, aspartic acid, glutamic acid, histidine, α-amino, cysteine, tyrosine, lysine, and arginine residues [56]. The ionization states of

these amino acids influence peripheral ion clusters, salt bridges, intramolecular ion paring and the lengths of the β-strands. These structural elements have an impact on the thermal stabilityof Csps which was described for thermophilic and hyperthermophilic bacteria [17,48].

To examine the influence of pH on the CpCspA stability, thermal unfolding experiments and T_m determinations at different pHs were carried out using DSC, these results indicate an increase of the CpCspA T_m from 49 °C (pH 8.0) to 53 °C (pH 6.0) (Fig. 3B), which is in agreement with the results of the CD experiments.

The reduction of the β-strand contents at pH 8.0 results in the reduced stability of CpCspA. Variations in β-sheet lengths may play a critical role in the thermal stability of Csps especially from thermophilic bacteria [57].

The influence of pH influence on the CpCspA molecule was simulated computationally. The homology model of CpCspA was used for several simulations that were carried out at pH 8.0, 7.0, 6.0, and without charges. The thermal stability for all these systems were evaluated by the comparison the heat capacity C_v (Fig. 4A), the value of the C_v peak showed a significant change, this value corresponds to the folding transition temperature (T_F). The CpCspA model without charges exhibited the lowest T_F, consequently, the lowest thermal stability. Thus, the presence of charged groups tends to stabilize CpCspA, at the pHs simulated here. The model with charges presented the lowest thermal stability at pH 8.0 and the highest thermal stability at pH 6.0. Quantitatively, the values of T_F in reduced units are 1.081, 1.078 and 1.073 for pH 6.0, 7.0, and 8.0, respectively. This represents a decrease of ˜1K for the pH 7.0 relative to pH 6.0 and ˜4K for the pH 8.0 relative to pH 6.0. The results of the theoretical model regarding the thermal stability under the influence of different pH are in good agreement with our results obtained from the DSC and CD experiments.

The corresponding free energy changes F(Q) as functions of the fraction of total native contacts Q and energy barriers between the folded (N) and unfolded (U) state of the CpCspA model are presented in Fig. 6B. Interestingly, the protein at pH 8.0 forms an intermediate state (I) which does not appear in pH 7.0, 6.0 or in the model without charges, but the shoulders in the curves are observable (Fig. 4B) and possibly correspond to very rapidly formed intermediate states.

As demonstrated by previous simulation experiments on the *B. subtilis* CspB the protein follows a rapid two-state folding mechanism after thermal or chemical unfolding with a highly unstable intermediate state [20,57]. Such an intermediate state is observable for CpCspA at pH 8.0 in the computational analysis (Fig. 4B and 5B). CpCspA at pH 8.0 possess decreased thermal stability and structural compactness as

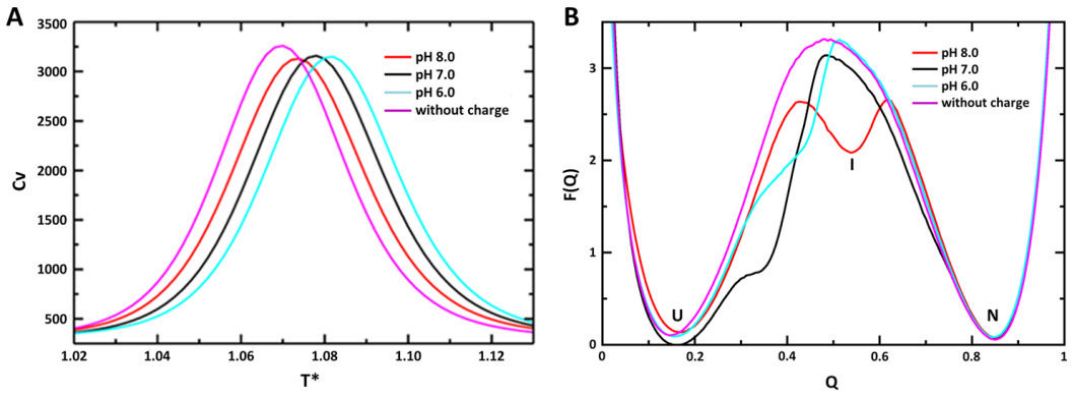


Figure 3.4: Thermodynamic properties of CpCspA at different pHs. (A) Heat capacity (Cv) at constant volume as function of temperature in reduced units (T*) for three different pHs and a system without charges. (B) Free energy curves as function of the reaction coordinate Q (fraction of total native contacts). U, I and N indicate the unfolded, the intermediate and the native state, respectively. All the systems with T close to the folding transition temperature T_F.

shown before by CD and DSC experiments.

The comparison of the distribution from the electrostatic energy of CpCspA at different pHs indicates a more negative average value for pH 6.0 than for pHs 7.0 and 8.0 (Fig. 5A). Thus, the charge-charge interaction contributes to the higher stability of the protein at pH 6.0. At pH 8.0 three peaks appear, while at pH 7.0 and 6.0 only two are visible, the comparison between the highest peaks for each pH (highlighted by dotted lines) indicated a change of ~$-2\varepsilon_c$ in the electrostatic energy between pH 6.0 and 7.0 and ~$-3\varepsilon_c$ between pH 6.0 and 8.0. The distribution of the electrostatic energy as a function of Q was plotted in a two dimensional map to evaluate the charge-charge interaction and their effects of each folding stage for pH 8.0, 7.0 and 6.0 (Fig. 5B-D, respectively).

The I state of the protein at pH 8.0 is not detectable either at pH 6.0 or at 7.0, resulting in only two peaks for these pHs in Fig. 5A. At pH 7.0 and 8.0 the N state of the protein is present in two different regions of the 2D map. These differences in the energetic ensembles in the same Q are caused by the higher ionization probability of specific residues at these pHs and depending on the protonation states of these amino acids resulting in a less stable N state of the protein.

Nevertheless, at all examined pHs the N state of the protein the electrostatic energy is more negative compared to the U state, and at pH 8.0 even the I state has a lower electrostatic energy. The higher energetic difference between the N state and the U state occurs at pH 6 (almost ~$2\varepsilon_c$), principally because the N state at this pH presents only one highly populated electrostatic energy level. All these electrostatic

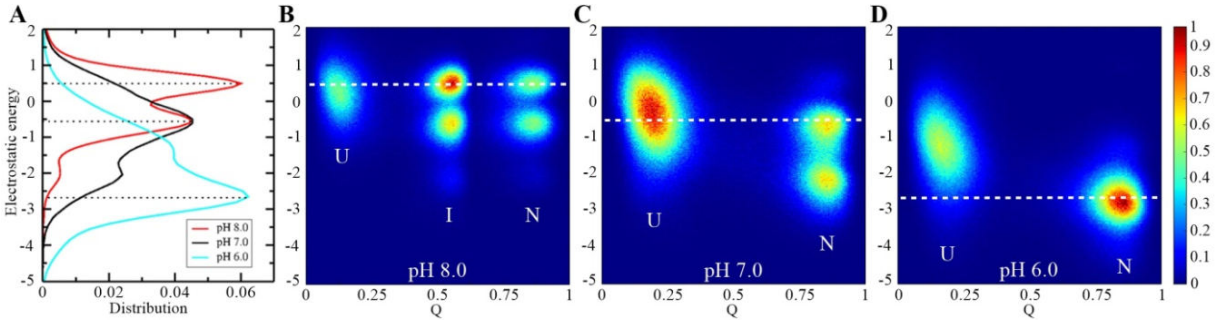


Figure 3.5: Electrostatic energy of CpCspA at different pHs. (A) Distribution of the electrostatic energy in the reduced unit $\varepsilon_c$for three different pHs. (B,C,D) Two dimensional map of the electrostatic energy as functions of the reaction coordinate $Q$for pH 8, 7 and 6, respectively. The color map represents the probability distribution of the electrostatic energy and Q, normalized by its highest value. All the systems with T ~ T_F.

features result in the higher stability of CpCspA at pH 6.0 than at pH 7.0 and 8.0.

The results of the CD, DSC experiments and computational analysis can be summarized to indicate that the pH of the protein solution has a very strong influence on the CpCspA secondary structure and thermal stability. However, at this stage we need to identify the role of key amino acids in the observed behavior.

3.3.4 His30 and His65 are pH-sensitive and advance the structural change?

Since pH changes in the solution affect the charges on ionizable groups [56], the observed destabilization at pH 8.0 and stabilization at pH 6.0 indicates the participation of an involved amino acid with a pKa value in this pH range, the pKa value of histidine residues is about 6.5 [56]. Various examples indicate that histidine residues play a role in the pH-induced conformational transitions of proteins [57-59].

CpCspA contain two histidines, His30 and His65. His30 is conserved and forms part of the RNP2 region and part of the β strand 3, which is in contrast to His65 which only present in CpCspA and the only other Csp that contains a histidine in this region is *T.maritima* (His61).

During the computational analysis, three amino acid residues show significant changes in their charges during the pH variations, other amino acids showed charge changes less than 1% (table S1 in the supplementary material present a list of all charges of the investigated amino acids).The Cα of the N terminal Met, His30 and His65 show a charge change, in which the strongest influence was observed for His30 and His65 at pH 8.0; the charge of both amino acids are near 0.0 (Fig. 6).

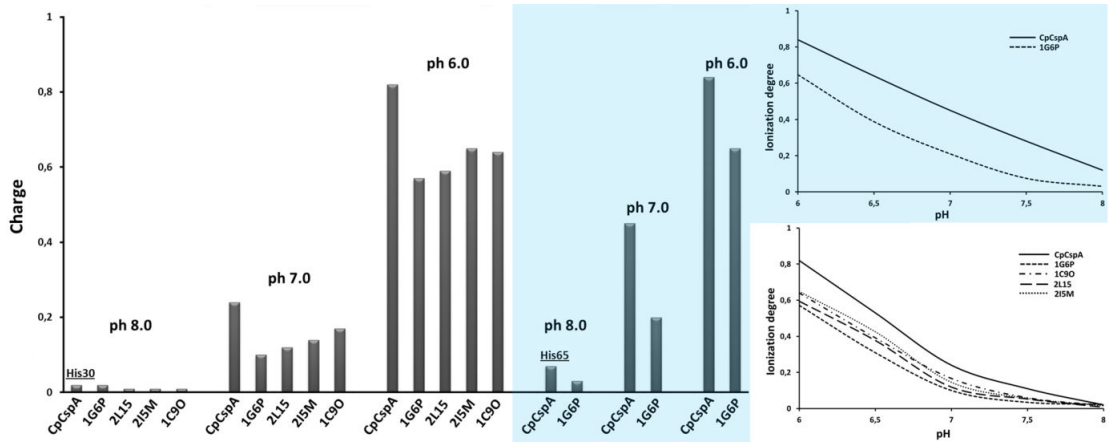
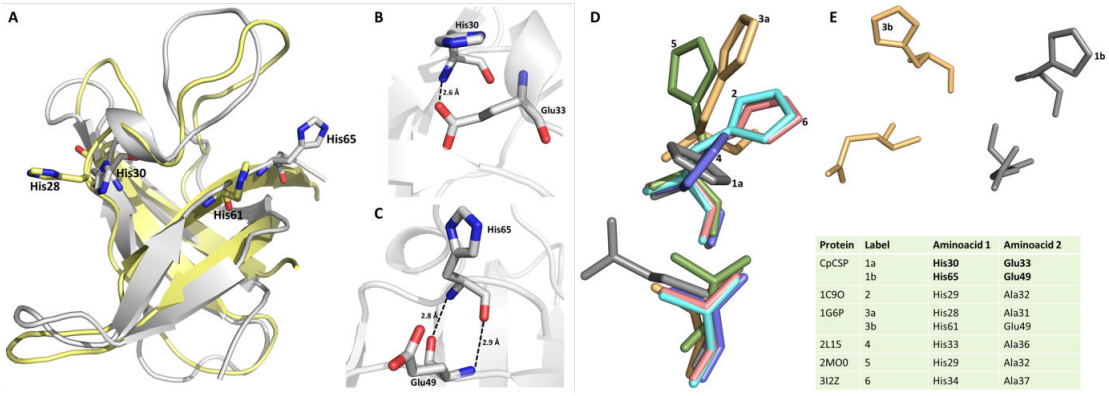


Figure 3.6: Charges of the histidine residues of Csp proteins from different species in variant pHs. The changes in the charge of His30 and His65 of CpCspA, His28 and His61 of Csp from *T. maritima*(PDB:1G6P), His33 of CspA from *E. coli*(PDB:2L15), His29 of CspB from *B. subtilis*(PDB:2I5M) and His29 of CspB from *B. caldolyticus*(PDB:1C9O) were observed. At pH 8 the charge of the histidine residues is near 0.0 and the maximum charge reached at pH 6.0.

Figure 6 shows that the conserved histidines in the Csp family undergo charge changes at different pHs. Especially at pH 8.0 the residue charges are near 0.0 and at pH 6.0 undergo the largest charge modification.

His30 is conserved in the Csp family and forms the C-terminal end of β-strand 3 (Fig. 7A), a web tool to predict salt bridges [33] indicated that His30 possess the possibility to form a salt bridge with Glu33 (Fig. 7B). This salt bridge is not present in the deposited Csp structures in the PDB, this amino acid is exchanged for alanine residues in all CSPs (Fig. 7D), the His30, Glu33 connection is unique for CpCspA. As consequence of the pH change the His30, Glu33 bond break and this region of the protein loses a stabilization factor.

The residue His65 is not conserved within the Csp family and is located in the C-terminal part of CpCspA (Fig. 7A). However, within the sequence of Csp from *T. maritima*(PDB: 1G6P), a histidine is present in the C-terminal region of the protein (His61), which is part of the long β-strand 4 (Fig. 7A). In 1G6P, His61 forms a salt bridge with Glu49; the distance between both residues is less than 3 Å (Fig. 7C). A hydrogen bond between His65 and Glu49 was detected in CpCspA using online database PDBsum [34] as a tool, most likely during a pH change the His65, Glu49 bond break. Additionally showed computational analysis that His65 has the potential to be involved in an electrostatic interaction, especially at pH 7.0 appears the highest charge change, which correspond to a favorable electrostatic interaction for this residue (Fig. A.1 is deposited in supplementary material).



Protein	Label	Aminoacid 1	Aminoacid 2
CpCSP	1a 1b	**His30** **His65**	**Glu33** **Glu49**
1C9O	2	His29	Ala32
1G6P	3a 3b	His28 His61	Ala31 Glu49
2L15	4	His33	Ala36
2MO0	5	His29	Ala32
3I2Z	6	His34	Ala37

Figure 3.7: Locations ofthe histidine residues in Csp family structures. (A) Structural comparison of the CpCspA homology model (grey) and the Csp of *T. maritima*(1G6P) (yellow), the positions of the histidine residues highlighted as sticks. (B) Detail view of the CpCspA region His30 and Glu33. (C) Detail view of the CpCspA region His65 and Glu49. (D) Overlay of the conserved His30 region of the Csp family and Glu33 of CpCspA and the alanine residues in homolog structures. (E) Overlay of the His61, Glu49 region from 1G6P and the His65, Glu49 region from CpCspA. The Color code for the histidine, alanine and glutamic acid residues is labeled for each species different.

Indeed His30 and His65 of CpCspA are pH sensitive and during a shift from pH 6.0 to pH 8.0, the charge of the amino acids will change.

3.4 Conclusions

Csp family proteins possess different amounts of beta strands and ionic networks, which influence the thermal stability of these proteins. The T_m of Csps from thermophilic and mesophilic bacteria were extensively studied and compared. However, less is known about the amino acids influenced by pH changes and effect the thermal stability from CpCspA.

We present the influence of different pHs and NaCl concentration on the secondary structure and thermal stability of CpCspA. Our approach identified one sodium binding site in CpCspA which are important for the protein stability.

At pH 8.0 CD spectroscopy experiments indicated a loss in the β-strand content, what was the first time described for this protein family. Additionally, thermal unfolding experiments and computational analysis showed that at pH 8.0 the protein exists in a low stability.

His30 and His65 which are involved in the formation of salt bridges and hydrogen bonds were identified by computational analysis and indicated the presence of significant changes in their charges.

3.5 Acknowledgements

This research was supported by grants from the CNPq (Science without Frontiers), FAPESP and CAPES. The authors want to thanks the PROPe UNESP. The authors would like to thank Sidney Jurado de Carvalho and Leandro Cristante de Oliveira for the helpful discussions. RJE is funded by FAPESP (2016/08104-8), VBPL was funded by FAPESP (2014/06862-7) and CNPq (Grants 442352/2014-0, 2532/002/14-PROPe/CDC). We also thank the Center of Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) for computational resources.

3.6 References

[1] Hébraud M, Potier P (1999) Cold shock response and low temperature adaptation in psychrotrophic bacteria. J Mol Microbiol Biotechnol 1: 211-219.

[2] Yamanaka K (1999) Cold shock response in *Escherichia coli*. J Mol Microbiol Biotechnol 1: 193-202.

[3] Cavicchioli R, Thomas T, Curmi PM (2000) Cold stress response in Archaea. Extremophiles 4: 321-331.

[4] Jones PG, van Bogelen RA, Neidhardt FC (1987) Induction of proteins in response to low temperature in *Escherichia coli*. J Bacteriol 169: 2092-2095.

[5] Jones PG, Krah R, Tafuri SR, Wolffe AP (1992) DNA gyrase, CS7.4, and the cold shock response in *Escherichia coli*. J Bacteriol 174: 5798-5802.

[6] Horn G, Hofweber R, Kremer W, Kalbitzer HR (2007) Structure and function of bacterial cold shock proteins. Cell Mol Life Sci 64: 1457-1470.

[7] La Teana A, Brandi A, Falconi M, Spurio R, Pon CL, Gualerzi CO (1991) Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS. Proc Natl Acad Sci USA 88: 10907-10911.

[8] Brandi A, Pon CL, Gualerzi CO (1994) Interaction of the main cold shock protein CS7. 4 (cspa) of *Escherichia coli* with the promoter region of hns. Biochimie 76: 1090-1098.

[9] Lopez MM, Yutani K, Makhatadze GI (1994) Interactions of the major cold shock protein of *Bacillus subtilis* CspB with single-stranded dna templates of different base composition. J Biol Chem 274: 33601-33608.

[10] Max KE, Zeeb M, Bienert R, Balbach J, Heinemann U (2006) T-rich dna single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. J Mol Biol 360: 702-714.

[11] Lopez MM, Makhatadze GI (2000) Major cold shock proteins, cspa from *Escherichia coli* and cspb from *Bacillus subtilis*, interact differently with single-stranded dna templates. BBA-Protein Struct M 1479: 196-202.

[12] Morgan HP, Estibeiro P, Wear MA, Max KE, Heinemann U, Cubeddu L, Gallagher MP, Sadler PJ, Walkinshaw MD (2007) Sequence specificity of single-stranded DNA-binding proteins: a novel DNA microarray approach. Nucleic Acids Res 35: e75.

[13] Jiang W, Hou Y, Inouye M (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. J Biol Chem 272: 196-202.

[14] Higgs PG (2000) RNA secondary structure: physical and computational aspects. Q Rev Biophys 33: 199-253.

[15] Schindelin H, Marahiel MA, Heinemann U (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. Nature 364: 164-168.

[16] Schindelin H, Jiang W, Inouye M, Heinemann U (1994) Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. Proc Natl Acad Sci USA 91: 5119-5123.

[17] Mueller U, Perl D, Schmid FX, Heinemann U (2000) Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. J Mol Biol 297: 975-988.

[18] Schnuchel A, Wiltscheck R, Czisch M, Herrler M, Willimsky G, Graumann P, Marahiel M, Holak TA (1993) Structure in solution of the major cold-shock protein from *Bacillus subtilis*. Nature 364: 169-171.

[19] Newkirk K, Feng W, Jiang W, Tejero R, Emerson SD, Inouye M, Montelione GT (1994) Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. Proc Natl Acad Sci USA 91: 5114-5118.

[20] Perl D, Welker C, Schindler T, Schröder K, Marahiel MA, Jaenicke R, Schmid FX (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. Nat Struct Mol Biol 5: 229-235.

[21] Perl D, Mueller U, Heinemann U, Schmid FX (2000) Two exposed amino acid residues confer thermostability on a cold shock protein. Nat Struct Mol Biol 7: 380-383.

[22] Dominy BN, Perl D, Schmid FX, Brooks CL (2002) The effects of ionic strength on protein stability: the cold shock protein family. J Mol Biol 319: 541-554.

[23] Wassenberg D, Welker C, Jaenicke R (1999) Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritime*. J Mol Biol 289: 187-193.

[24] Makhatadze GI, Marahiel MA (1994) Effect of pH and phosphate ions on self-association properties of the major cold-shock protein from *Bacillus subtilis*. Protein Sci 3: 2144-2147.

[25] Dorella FA, Pacheco LGC, Oliveira SC, Miyoshi A, Azevedo V (2006) *Corynebacterium pseudotuberculosis*: microbiology, biochemical properties, pathogenesis and molecular studies of virulence. Vet Res 37: 201-218.

[26] Paton MW, Rose IR, Hart RA, Sutherland SS, Mercy AR, Ellis TM, Dhaliwal JA (1994) New infection with *Corynebacterium pseudotuberculosis* reduces wool production. Aust Vet J 71: 47-49.

[27] Ayers JL (1977) Caseous lymphadenitis in goats and sheep: a review of diagnosis, pathogenesis, and immunity. J Am Vet Med Assoc 171: 1251-1254.

[28] Peel MM, Palmer GG, Stacpoole AM, Kerr TG (1997) Human lymphadenitis due to *Corynebacterium pseudotuberculosis*: report of ten cases from Australia and review. Clin Infect Dis 24: 185-191.

[29] McKean S, Davies J, Moore R (2005) Identification of macrophage induced genes of *Corynebacterium pseudotuberculosis* by differential fluorescence induction. Microbes Infect 7: 1352-1363.

[30] Lindae A, Eberle RJ, Caruso IP, Coronado MA, de Moraes FR, Azevedo V, Arni RK (2015) Expression, purification and characterization of cold shock protein A of *Corynebacterium pseudotuberculosis*. Protein Expres Purif 112: 15-20.

[31] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

[32] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006) Comparative Protein Structure Modeling With MODELLER. Curr Protoc Bioinformatics 15: 1-30.

[33] Costantini S, Colonna G, Facchiano AM (2008) ESBRI: a web server for evaluating salt bridges in proteins. Bioinformation 3: 137-138.

[34] Laskowski RA (2007), Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. Bioinformatics 23: 1824-1827.

[35] Sreerama N, Woody RW (2004) Computation and Analysis of protein circular dichroism spectra. Methods Enzymol 383: 318-351.

[36] Sturtevant JM (1987) Biochemical applications of differential scanning calorimetry. Annual review of physical chemistry 38: 463-488.

[37] Sanchez-Ruiz JM (2007) Ligand effects on protein thermodynamic stability. Biophysical chemistry 126: 43-49.

[38] Contessoto VG, Oliveira VM, Carvalho SJ, Oliveira LC, Leite VBP (2016) NTL9 folding in constant pH: The importance of Electrostatic interaction and pH-dependence. J Chem Theory Comput 12: 3270-3277

[39] Contessoto VG, Lima DT, Oliveira RJ, Bruni AT, Chahine J, Leite VBP (2013) Analyzing the effect of homogeneous frustration in protein folding. Proteins 81: 1727-1737.

[40] Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 298: 937-953.

[41] Mongan J, Case DA, McCammon JA (2004) Constant pH molecular dynamics in generalized Born implicit solvent. J Comput Chem 25: 2038-2048.

[42] Baptista AM, Teixeira VH, Soares CM (2002) Constant-pH molecular dynamics using stochastic titration. J Chem Phys 117: 4184-4200.

[43] Ullner M, Woodward CE, Jönsson B (1996) A Debye–Hückel theory for electrostatic interactions in proteins. J Chem Phys 105: 2056-2065.

[44] Fersht A (1984) Enzyme Structure and Mechanism, 2nd edn. W.H. Freeman, New York, NY.

[45] Schröder K, Graumann P, Schnuchel A, Holak TA, Marahiel MA (1995) Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif. Mol Microbiol 16: 699-708.

[46] Schindler T, Perl D, Graumann P, Sieber V, Marahiel MA, Schmid FX (1998) Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from *Bacillus subtilis*. Proteins 30: 401-406.

[47] Jin B, Jeong KW, Kim Y (2014) Structure and flexibility of the thermophilic cold-shock protein of *Thermus aquaticus*. Biochem Biophys Res Commun 451: 402-407.

[48] Kremer W, Schuler B, Harrieder S, Geyer M, Gronwald W, Welker C, Jaenicke R, Kalbitzer HR (2001) Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritime*. Eur J Biochem 268: 2527-2539.

[49] Chatterjee S, Jiang W, Emerson SD, Inouye M (1993) The backbone structure of the major cold-shock protein CS7.4 of *Escherichia coli* in solution includes extensive beta-sheet structure. J Biochem 114: 663-669.

[50] Record Jr MT, Zhang W, Anderson CF (1998) Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. Adv Protein Chem 51: 281-353.

[51] Baldwin RL (1996) How Hofmeister ion interactions affect protein stability. Biophys J 71: 2056-2063.

[52] von Hippel PH, Wong KY (1964) Neutral salts: the generality of their effects on the stability of macromolecular conformations. Science 145: 577-580.

[53] Kohn WD, Kay CM, Hodges RS (1997) Salt effects on protein stability: two-stranded alpha-helical coiled-coils containing inter- or intrahelical ion pairs. J Mol Biol 267: 1039-1052.

[54] Zhao L, Ma K, Yang Z (2015). Changes of water hydrogen bond network with different externalities. Int J Mol Sci 16: 8454-8489.

[55] Huang L, Shakhnovich EI (2012) Is there an en route folding intermediate for cold shock proteins? Protein Sci 21: 677-685.

[56] D'Auria G, Esposito C, Falcigno L, Calvanese L, Iaccarino E, Ruggiero A, Pedone C, Pedone E, Berisio R (2010) Dynamical properties of cold shock protein A from *Mycobacterium tuberculosis*. Biochem Biophys Res Commun 402: 693-698.

[57] Karshikov AD, Engh R, Bode W, Atanasov BP (1989) Electrostatic interactions in proteins: Calculations of the electrostatic term of free energy and the electrostatic potential field. Eur Biophys J 17: 287-297.

[58] Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN (2006) pK values of the ionizable groups of proteins. Protein Sci 15: 1214-1218.

[59] Armstrong KM, Baldwin RL (1993) Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges. Proc Natl Acad Sci USA 90: 11337-11340.

Capítulo 4

Proteína 5 do ácaro *Blomia tropicalis* - Blo t 5

Epitope charge distribution controls the antigenicity for binding human IgG4 isotype in a modified Blo t 5, the major allergen from dust mite *Blomia tropicalis*

Bárbara G Ávila; Laura F Tomaz; Carlos R Prudencio; Karine C Almeida; Deise A Silva; Odonírio Abrahão Junior; Vinícius M Oliveira; Vinícius G Contessoto; Vitor B Leite; Ronaldo J Oliveira; Pamela A Candido; Jair P Cunha-Júnior

*This article has been submitted to Immunobiology Journal

Abstract

Background: Blo t 5 is the major allergen from *Blomia tropicalis* (Bt), since 40-60% of Bt-sensitive patients show IgE reactivity to this allergen. A better understanding of the biomolecular interactions of the binding allergen-antibody epitope is required to produce new allergens for the diagnostics and treatment of allergic respiratory diseases.

Objective: To produce a modified allergen (mBlo t 5) with enhanced IgG1 and IgG4 reactivity profile with respect to the recombinant Blo t 5 (rBlo t 5) and evaluate a novel protocol to create less allergenic epitopes.

Methods: Specific reactivity of rBlo t 5 and mBlo t 5 allergens was probed with anti-Blo t 5 monoclonal antibody and levels of IgE, IgG1 and IgG4 were measured to both allergens in atopic and non-atopic patients. Molecular dynamics and charge-charge electrostatic calculations were performed to validate and to optimize the epitope of the designed mBlo t 5 allergen in relation to the natural rBlo t 5.

Results: Levels and positivity for IgE and IgG1 to both allergens were comparable in Bt-sensitive patients, contrasting with a noticeable increase of IgG4 levels to mBlo t 5. 69% of the samples had increase of IgG4 reactivity compared to 45% of increase for IgG1 and 44% of reduction for IgE to mBlo t 5. Thermodynamic analysis showed that mBlo t 5 is thermo-stable molecule as compared as rBlo t 5 ($\Delta F = 2.2kT$) and also its amino acids dynamics. The modified allergen mBlo t 5 epitope had the energy of the charged residues stabilized while rBlo t 5 epitope showed to be electrostatically unstable.

Conclusion: The modified Blo t 5 allergen produced a remarkable reaction with IgG4 antibodies in *Blomia tropicalis*-sensitized patients. The simplified computer-based analysis suggested that the electrostatic-stable epitope controls the improvement in antigenicity for binding to human IgG4 isotype. This study has the potential to be a powerful high-throughput approach to investigate further molecular modifications in order to design potential hypoallergenic candidates to be used in specific allergen immunotherapy.

Keywords: Blo t 5 - recombinant allergen - *Blomia tropicalis* - antibody isotypes - allergic rhinitis - *in silico* design - electrostatic stability

4.1 Introduction

Allergic respiratory diseases, such as asthma and rhinitis, consist of an immunoglobulin E (IgE)-mediated hypersensitivity reaction that affects more than 25% of the population, reaching endemic proportions. The socioeconomic impact of these diseases on the community is large, including cost of health care and lost work and school hours [1].

Epidemiological studies show that house dust mites (HDMs) of the genus *Dermatophagoides* (*D. pteronyssinus* and *D. farinae*) are considered a globally significant source of aeroallergens responsible for the sensitization of more than 50% of allergic patients [2,3]. However, in many tropical and subtropical regions, including Spain, India, Taiwan, Brazil, Colombia, the Philippines and Indonesia [4], *Blomia tropicalis* mite has been found in house dust and its allergens are also one important triggering factor of allergic respiratory diseases [5–7].

The introduction of molecular biological techniques in allergy research allowed the identification of more than 20 groups of HDM allergens. Nowadays, at least 21 IgE binding proteins obtained from *B. tropicalis* have been observed, and nine of these allergens were extensively characterized at molecular level, including the identification of cDNA sequences that encode these important allergenic proteins [8]. The frequency of human IgE reactivity to these allergens has been reported, ranging from 11% to 70% [9]. Among the *B. tropicalis* identified allergens, Blo t 5, a protein of 14 kDa (113 amino acids), is the major allergen, and epidemiological studies showed that up to 90% of the symptomatic (asthmatic and rhinitis) patients have strong IgE reactivity to this specific allergen [10,11].

Allergenic extracts obtained from natural sources have been used for over 100 years for the diagnosis and treatment of allergic diseases. During the manufacturing process of the extracts, other allergenic proteins in addition to the allergen of interest, including macromolecules and non-allergenic substances, often contaminate the product. Likewise, the different methods of production used by different manufacturers produce a heterogeneous extract, containing varying and even negligible amounts of the relevant allergen [12].

A variety of methods for detecting serum IgE antibodies to allergens use natural allergen extracts that cannot be fully standardized and may compromise the interpretation of the serological results [13]. For this reason, the use of recombinant allergens for diagnostic and therapeutic purposes is attractive for the standardization of new allergenic preparations and the evaluation of a defined allergen compound [14].

Currently, by using techniques of molecular biology, it is possible to produce allergens with the same immunological characteristics of the natural allergen or produce new molecules with modified properties for the purpose of reducing the allergenicity and increasing the immunogenicity of formulations for application on allergen-specific immunotherapy [12]. Thus, the aim of this study was to produce a modified form of the Blo t 5 allergen from *Blomia tropicalis,* to evaluate the reactivity profile of IgE, IgG1 and IgG4 antibodies from atopic and non-atopic patients and also develop and validate an *in silico* approach to predict potential modified epitopes with hypoallergenic properties.

Homologous allergens in amino acid sequence and 3D structure with significant differences in charged amino acid residues of the epitope with the same thermal molecular stability display different IgE affinity [15,16]. Thus, the protein electrostatic interactions of charged residues within IgE-binding epitope have to be taken into account to in order to design new molecules with hypoallergenic activity.

Previous works have shown that charged and surface exposed residues at the epitope region are extremely important for recognition and binding of antigens-antibodies [16–20] and the differences in distribution of charges in the epitope accounts for the difference in the IgE binding specificities among different HDM's allergens [21]. However, there is a lack of information regarding the biomolecular electrostatic interactions between these molecules. In order to evaluate the role of potential interactions and how they can modulate the antigen-antibody binding, a rapid computational model was proposed in this work. This method allows one to calculate the free energy of each ionizable residue of the antibodies using the Tanford-Kirkwood method [22,23]. Based on this method, it is possible to evaluate whether the electrostatic energy of each charged residue is favorable or unfavorable for the overall molecular stabilization. This kind of approach has been used with widely success in previous report demonstrating the increase in stability of a large number of proteins [24–27]. The new modified allergen mBlo t 5 demonstrated notable reactivity with IgG4 antibodies from *Blomia tropicalis*-sensitized patients in comparison with the natural recombinant Blo t 5 or the total *Blomia tropicalis* extract. The bridge between experimental and theoretical analysis performed was intended to shed some light on a better understanding of the antigen-antibody binding mechanism and to propose a method in which it could be used to evaluate molecular modifications in new allergens in further hypoallergenic specific immunotherapy studies.

4.2 Materials and Methods

Patients, skin prick test and serum samples

A total of 110 adult subjects were enrolled for the study. Fifty-five subjects (20 males and 35 females, age from 18 to 55 years) had allergic rhinitis with or without mild-to-moderate asthma and they were selected on basis of a clinical history of allergic symptoms in response to HDMs and positive skin prick test (SPT) to HDM allergens (*Blomia tropicalis* and/or *Dermatophagoides pteronyssinus* and/or *Dermatophagoides farinae*) at Allergy Unit of the Clinic Hospital of the Federal University of Uberlândia, Brazil. As non-atopic control group, 30 healthy subjects (5 males and 25 females, age from 19 to 59 years) were also selected based on the absence of allergy history and negative SPT to a panel of standardized aeroallergens. An additional control group was also included with 25 atopic subjects (7 males and 18 females, age from 19 to 60 years) on basis of a clinical history of allergy in response to HDM and positive SPT to *D. pteronyssinus* and/or *D. farinae*, however with negative SPT to *B. tropicalis* allergens. The Ethics Committee on Human Research at the Federal University of Uberlândia (Protocol CEP/UFU 174/10) approved all procedures and all subjects enrolled in this study signed informed consent form.

All subjects underwent SPT, as described elsewhere [28], with the following standardized commercial aeroallergen extracts: mites (*D. pteronyssinus*, *D. farinae* and *B. tropicalis*), cockroaches (*Blattella germanica* and *Periplaneta americana*), mold (*Alternaria alternata*) and pet danders (*Felis domesticus* and *Canis familiaris*). Histamine (10 mg/mL) and 50% glycerin saline solution were used as positive and negative controls, respectively. The commercial allergen extracts and control solutions were purchased from FDA Allergenic, RJ, Brazil. Reactions were read after 15 min and a mean wheal 3 mm-diameter larger than the negative control was considered to be positive. In parallel, blood samples (20 mL) were collected from all participants and sera were stored in aliquots at -20 °C until serological assays were performed.

Synthesis of artificial Blo t 5 genes

The gene sequence that encodes the *Blo t 5* allergen from *Blomia tropicalis* (537 bp) was obtained in the EMBL Nucleotide Sequence Database (EMBL: O96870) and the artificial genes were purchased from GenScript (www.genscript.com). This sequence was used to create two constructs encoding the full sequence of the Blo t 5 alllergen, one with the natural sequence of Blo t 5, the recombinant Blo t 5 (rBlo t

5) and another with a modified sequence producing a rearrangement on the Blo t 5 α-domains (mBlo t 5). The amino acid sequence of rBlo t 5 was divided in four regions, one disorder/unstructured N-domain followed by three α-domains (D, α1, α2 and α3). The α3 helix was shifted between D and α1 domains (D, α3, α1 and α2) to produce the modified allergen mBlo t 5 (Fig. 1). The synthetic DNA sequences were ligated into the expression vector pET14b (Novagen, USA) in order to promote the production of *His*-tag fusion proteins.

Molecular modeling and structure prediction of the modified allergen mBlo t 5

The three-dimensional structure prediction of the modified recombinant allergen (mBlo t 5) from *B. tropicalis* was performed by homology modeling with the MHONLINE server [29], which is a biological workflow that combines a specific set of programs for automated protein structure prediction and function annotation. Structure optimization and energy minimization were performed with OPLS Force Field [30] using Macromodel Software [31] and Maestro Graphical interface (www.schrodinger.com). The homology modeling of mBlo t 5 used the Blo t 5 nuclear magnetic resonance (NMR) structure coordinates (Protein Data Bank code 2JMH, model 1) as the model template (Fig. S1).

Production of recombinant Blo t 5 allergens from *Blomia tropicalis*

BL21 (DE3) *E. coli* strain (Novagen, USA) cells were transformed with 2.5 L (50 ng) of the plasmid pET14b by the heat shock method [20]. Transformed *E. coli* clones were selected on solid medium Luria Bertani (LB) containing appropriate antibiotic (50 g/mL ampicillin (Sigma-Aldrich Co., MO, USA) for 18 h at 37 °C. After, single colonies of BL21 (DE3) *E. coli* were inoculated in LB medium plus antibiotic and grown overnight at 37 °C. Then, the expression of recombinant proteins was inducted by 0.5 mM isopropyl-β-thiogalactopyranoside (IPTG) for 18 h at 30 °C. For purification of recombinant Blo t 5 protein (rBlo t 5), *E. coli* cells were harvested by centrifugation (4600 $\times$ g, 30 min at 4 °C) and then lysed by addition of lysozyme at 1 mg/mL in extraction buffer (50 mM NaH_2PO_4, 500 mM NaCl, 1mM PMSF protease inhibitor). Alternatively, for purification of mBlo t 5, the extraction buffer was supplemented with 6 M urea. His-tagged proteins were purified by nickel affinity chromatography column Ni-NTA Agarose (Qiagem GmbH), previously equilibrated with the binding

buffer (50 mM NaH_2PO_4, 500 mM NaCl, 40 mM imidazole pH 7.4) for rBlo t 5 or binding buffer plus 6 M urea for mBlo t 5. The column was washed with the binding buffer appropriated to each protein in order to remove unbound proteins. The His-tag proteins were eluted by addition of elution buffer (50 mM NaH_2PO_4, 500 mM NaCl, 500 mM imidazole, pH 7.4) for rBlo t 5 or elution buffer plus 6 M urea for mBlo t 5. All purified fractions were dialyzed in Amicon (Millipore, cutoff 3.5 kDa) and stored at -70 °C. The protein content was determined using the Lowry method [32].

Electrophoresis and immunoblotting of the recombinant Blo t 5 allergens

Samples of the rBlo t 5 and mBlo t 5 allergens from *B. tropicalis* were run in polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE) under nonreducing conditions. Samples were homogenized in SDS sample buffer, boiled at 95 °C for 5 min and loaded on 13.5% gel in parallel with molecular weight markers (BenchMark™ Protein Ladder 6-180 kDa, Invitrogen, Karlsuhe, Germany). The gel was stained with Coomassie blue or electrotransferred to nitrocellulose membranes (0.45 m, Bio-Rad, Hercules, CA, USA), as previously described [33], using a semi-dry transfer system (Trans-blot SD, Bio-Rad Laboratories Inc.)

Fractions of Ni-affinity chromatography were assayed with mouse monoclonal antibody (mAb) to Blo t 5 (clone 4G9 - Indoor Biotechnologies Inc., Charlottesville, VA) in order to confirm the production of the rBlo t 5 allergens. Membranes were blocked with 5% nonfat milk in phosphate-buffered saline pH 7.2 (PBS) plus 0.05% Tween-20 (PBS-T-M) for 2 h at room temperature and then incubated overnight at 4 °C with mouse mAb anti-Blo t 5 diluted 1:1000 in PBS-T-M or only PBS-T-M. A second step using biotinylated anti-mouse IgG (Kierkegaard and Perry Lab., Gaithersburg, USA) diluted 1:1000 was added and incubated for 2 h at room temperature. After washing with PBS-T, membranes were incubated for 1 h at room temperature with streptavidin-peroxidase (Sigma-Aldrich) diluted 1:1000 and revealed by 3,3 diaminobenzidine (DAB) tablets (Sigma-Aldrich) and 0.03% H_2O_2, until the appearance of bands when the reaction was stopped with the addition of distilled water.

Slot-blot of the recombinant Blo t 5 allergens

Nitrocellulose membranes were blotted with rBlo t 5 (2.5 g/slot) or mBlo t 5 (2.5 g/slot) on vacuum device (Bio-dot®SF apparatus, Bio-Rad, Hercules, CA, USA) to confirm the specific reactivity of the rBlo t 5 and mBlo t 5 allergens. As controls,

total Bt extract (10 g/slot; FDA Allergenic) and bovine serum albumin (5 g/slot; BSA, Sigma) were also blotted in the same conditions. Membranes were blocked with 5% nonfat milk in PBS plus 0.05% Tween-20 (PBS-T-M) for 2 h at room temperature and then incubated overnight at 4 °C with mouse mAb anti-Blo t 5 diluted 1:500 in PBS-T-M or only PBS-T-M. A second step using biotinylated anti-mouse IgG diluted 1:1000 was added and incubated for 2h at room temperature. After washing with PBS-T, membranes were incubated for 1 h at room temperature with streptavidin-peroxidase diluted 1:1000 and revealed by DAB and 0.03% H_2O_2.

In a second experiment, membranes were blocked with 5% BSA in PBS plus 0.05% Tween-20 (PBS-T-BSA) for 2 h at room temperature and incubated overnight at 4 °C with serum samples from atopic Bt+, atopic Bt [336?] and non-atopic subjects, diluted 1:2 in PBS-T-BSA or only PBS-T-BSA. A second step using biotinylated anti-human IgE diluted 1:3000 was added and incubated for 2 h at room temperature. After washing with PBS-T, membranes were incubated for 1 h at room temperature with streptavidin-peroxidase diluted 1:500 and revealed by DAB and 0.03% H_2O_2.

Measurement of specific IgE, IgG1 and IgG4 to *B. tropicalis* extract, rBlo t 5 and mBlo t 5 allergens

Levels of specific IgE to total Bt extract, rBlo t 5 and mBlo t 5 were measured by conventional ELISA as described elsewhere [34], with minor modifications. High-binding microtiter plates were coated overnight at 4 °C with crude Bt (10 μg/well) or rBlo t 5 (5 μg/well) or mBlo t 5 (5 μg/well). The plates were blocked with PBS containing 0.05% Tween 20 plus 1% bovine serum albumin (PBS-T-BSA) and incubated with serum samples diluted 1:2 (IgE) and 1:5 (IgG1 and IgG4) in PBS-T-BSA for 1 h at 37 °C. After washing, plates were incubated with biotinylated anti-human IgE (Kierkegaard and Perry Lab.) diluted 1:1000, or biotinylated anti-human IgG1 (Sigma) diluted 1:3000 or biotinylated anti-human IgG4 (Sigma) diluted 1:1000 for 1 h at 37 °C. After additional step of washing, the plates were incubated with streptavidin-peroxidase (Sigma) diluted 1:1000 for 30 min at room temperature. The assay was developed with 10 mM 2,2-azinobis-(3-ethyl-benzthiazoline) sulfonic acid (ABTS; Sigma) and 0.03% H_2O_2. The absorbance was determined in a plate reader at 405 nm. Antibody titers were reported as ELISA index (*EI*) and determined as follows: $EI = absorbance\ of\ test\ sample/cutoff$, where cutoff was calculated as the mean absorbance of negative control sera plus three standard deviations. $EI\ values > 1.2$ were considered to be positive in order to exclude borderline reactivity values close to $EI = 1.0$.

Structure-based models

Structure-based models (SBM), also known as Gō-models, are simplified molecular dynamics simulations that have been proved to be successful in studying protein folding and functional mechanisms with experimental quantities in agreement with theory such as experimental φ-values [35–37]. In the same way, the theoretical structure-based models show a strong correlation with experimental free energy barriers, protein diffusion coefficients and folding rates [38–45].

SBM interaction energy Hamiltonian and its parameters are based in the protein native structure. Structure-based models can simplify a protein by all of its heavy atoms or its alpha carbons. The simulation input files are generated for the SBM in the SMOG webserver [46] and the simulations are performed using the molecular dynamic package GROMACS, version 4.5.4 [47]. The order parameter used in the SBM simulations of a given structure Γ was the fraction of native contacts *(Q)*, which has been shown to be an excellent and reliable reaction coordinate to monitor protein folding/unfolding events ($Q = 0.0$ is unfolded and $Q = 1.0$ is folded) in theoretical studies with the native contact assumed to be formed if the residues i and j are in a distance smaller than $1.2\sigma^{ij}$ [38–45,48]. The potential functional form used in the all-atom and C_α simulations are described in the **Supporting Information** section.

All-atom SBM. Simulations were performed using time steps dt= 0.5 fs and snapshots were record at every 1 ps. Berendsen thermostat algorithm [49] was applied in the simulations to maintain coupling to an external bath temperature with constant coupling of 1 ps. Blo t 5 fluctuations in residue positions, given by the root mean square fluctuations (rmsf), were calculated after proper system equilibration and averaging over all-atom SBM trajectories of 100 ns at temperature $0.9T_f$ (in reduce units) (T_f is the folding temperature in which the protein changes its native to the denatured structure ensemble).

C_α SBM. Each simulation was executed starting from an unfolded and random configuration, simulated during 1 μs (10 times higher than all-atom runs) with time steps dt= 0.5 fs and snapshots were record at every 0.5 ps. Berendsen thermostat algorithm [49] was applied in the simulations to maintain proper coupling to an external temperature bath with constant coupling of 1 ps. Both Blo t 5 thermodynamic free energy and specific heat profiles were analyzed using the weighted histogram analysis method (WHAM) [50] over 24 constant temperature runs ranging from 0.8 to 1.5, which cover the transition folding temperature ($T_f \tilde{} 1.0$). T_f was defined as the temperature at the specific heat peak of each simulated protein.

Charge-charge interaction calculation - TKSA-MC

The charge-charge energy interaction between two charged residues was calculated using the Tanford-Kirkwood model with a Solvent Accessibility (TKSA) contribution [22,51] The charge-charge interaction calculation has been used to understand the thermo-stabilization of proteins and to redesign enzymes by optimization of the surface electrostatic interactions [25,52–56]. In this work, it was developed a similar yet less computer time and memory consuming methodology using Metropolis Monte Carlo simulation [57] in the canonic ensemble. This new procedure will be thereafter referred as the Tanford-Kirkwood Surface Accessibility - Monte Carlo or simply as the TKSA-MC procedure.

In this simple model, the protein is represented by a sphere with low dielectric constant $\epsilon = 4$ and radius b, which represents the radius of the protein. The interaction energy is between unit positive charges placed in the protonation sites of ionizable groups. The pairwise interaction energy between residues i and j is described by

$$U_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}) \,, \tag{4.1}$$

where e is the unit charge, Aij , Bij and Cij are parameters obtained from the Tanford-Kirkwood solution, which are associated with the dielectric constant of the protein and solvent, the distance between the charges and, Cij is the ionic term [22,23,52]. $SAij$ is the average of accessible surface area of the side chain from residue i and residue j and a is the ion sphere exclusion radius with a = b + 1.4 Å (mean cation and anion radii of a typical NaCl salt).

The protein charge distribution is modified at each Monte Carlo step by titration, where the values of the residues valency are changed ($0 \leftrightarrow +1$, for ARG/LYS/HIS, $-1 \leftrightarrow 0$ for ASP/GLU). The acceptances of the Monte Carlo steps follow the Metropolis criterion considering the energy variation [58,59].

$$\Delta E = \pm \xi k_B T \ln(10)(pH - pK_a) + \Delta E_{elec} \tag{4.2}$$

where the positive sign is used when the residue is protonated and the negative sign is used when the residue is deprotonated during the simulation. The residues intrinsic pKa values are: ASP = 4.0; GLU = 4.5; HIS = 6.3; LYS = 10.6; ARG = 12.0 [60,61]. It was performed 10^5 Monte Carlo steps for equilibration and 10^8 steps for production. The accessible surface area was calculated using the Surface

Racer software [62]. The surface diagrams were prepared with the Virtual Molecular Dynamic (VMD) package and the PDB2PQR server [63–65].

Statistical analysis

Statistical analyses were performed using the software GraphPad Prism version 5.0 (GraphPad Software Inc., San Diego, CA, USA). Differences between groups related to age were analyzed by the Kruskal-Wallis test and Dunn multiple comparison post-test. Differences observed in sex, SPT positivity and clinical diagnosis were analyzed by the Chi-square (χ^2) test or Fisher exact test, when appropriate. Differences in the mean wheal size were evaluated by ANOVA and Tukey multiple comparison post-test. Differences in the antibody levels between the allergens within each patients group were analyzed by the Friedman test and Dunn multiple comparisons post-test, and between patient groups for each allergen by the Kruskal-Wallis test and Dunn multiple comparison post-test. Differences in the seropositivity between allergens within each patient group were analyzed by the Fisher exact test. Correlation between the levels of antibody isotypes to rBlo t 5 and mBlo t 5 allergens were analyzed by the Spearman correlation coefficient. Values of $P < 0.05$ were considered statistically significant.

4.3 Results

Patient data

Demographic and clinical characteristics of the subjects included in the study are shown in Table 1. No significant differences were found for sex and age among atopic Bt+, atopic Bt– and non-atopic groups, although the females were the most frequent in each group ($P < 0.01$). In atopic Bt+ group, 96.4% and 94.5% were cosensitized to *D. pteronyssinus* and *D. farinae,* respectively, with only two patients monosensitized to *D. farinae* and three to *D. pteronyssinus*. In atopic Bt– group, 96% and 92% were sensitized to *D. pteronyssinus* and *D. farinae,* respectively, with only one patient monosensitized to *D. farinae* and two to *D. pteronyssinus*. No significant differences were found for SPT positivity and mean wheal size between the atopic groups as well as within each group. The most frequent diagnosis was rhinitis within the atopic Bt+ group (92.8%) that also was higher than the atopic Bt– group (44%) ($P < 0.0001$). On the other hand, rhinitis with asthma predominated in the atopic Bt– (40%) than atopic Bt+ (7.2%) patients ($P = 0.0008$) and only 16% of atopic Bt– patients were asymptomatic ($P < 0.05$).

Recombinant rBlo t 5 and the designed mBlo t 5 allergens

The sequences of the major IgE binding protein rBlo t 5 and the modified mBlo t 5 are shown in Fig. 1. The amino acid sequences of the rBlo t 5 protein, obtained in the Genbank by the accession number O96870, and the mBlo t 5, designed by a shift of the α3 helix of the rBlo t 5 (indicated by the red arrow) are shown in Fig. 1. In Fig. 1, Arabic numbers over the sequences refer to amino acid position in mature proteins, the total number of amino acids is indicated at the end of the sequences (134 amino acids) and the 17 initial amino acids (in italic), in each sequence, correspond to the signal peptide, which were not included in the modeling and simulation analysis. The rBlo t 5 protein structure (Protein Data Bank code 2JMH) consists of three α-helices (α1, α2, α3), forming a helix bundle, and a N-terminal disordered domain (D) (Fig. S1, left panel). A similar structure was modeled for mBlo t 5, with the three α-helices rearranged as in Fig. 1 (α3, α1, α2) and the N-terminal (D), also in a disordered state (Fig. S1, right panel). The modeled mBlo t 5 has 72% of sequence identity and 87% of structural similarity with respect to the wild-type protein, rBlo t 5. This high identity in sequence proposes that the modeled rBlot 5 structure (also with high structural identity) is a reasonable model to be used in the theoretical studies employed in the allergen proteins.

Expression, purification and immunodetection of rBlo t 5 and mBlo t 5 allergens

After the design of the constructs, *pET14b-rBlot5* and *pET14b-mBlot5* vectors were used to produce these recombinant allergens in BL21 (DE3) *E. coli* strain. Coomassie blue staining of gels revealed the presence of an additional band of 15 kDa in extracts of *E. coli/pET14b-rBlot5* induced with IPTG (Fig. 2A, lane 2). After purification of this sample in Ni-NTA agarose resin and separation in SDS-PAGE (Fig. 2A, lane 3), it was observed a strongly stained band (30 kDa) that could represent stable rBlo t 5 dimers in the presence of SDS, which was weakly visualized migrating around 15 kDa (monomer). One additional band was also detected migrating around 64 kDa that could represent stable rBlo t 5 tetramers. These findings were confirmed by using immunoblot probed with monoclonal anti-Blo t 5 antibody (Fig. 2B, lane 1), which revealed bands around 15, 30, and 60 kDa, especially the 30 kDa band (dimers). Two additional bands between 19 and 26 kDa were also detected in SDS-PAGE (Fig. 2A, lane 3), but no reactivity was found for those bands in immunoblot (Fig. 2B, lane 1). In extracts of *E. coli/pET14b-mBlot5* induced with IPTG (Fig. 2A, lane 5) an additional strongly stained band was visualized migrating around 15 kDa, and after

purification and electrophoresis procedures (Fig. 2A, lane 6), this band was maintained along with additional bands between 26 and 37 kDa that could represent mBlo t 5 dimers. This partially purified sample was probed in immunoblot with monoclonal anti-Blo t 5 antibody (Fig. 2B, lane 3) and only bands of 15 kDa (monomer) and 30 kDa (dimers) were stained, with more intense staining for the 30 kDa band.

The specific reactivity of rBlo t 5 and mBlo t 5 was also confirmed by a slot-blot analysis probed with the monoclonal antibody anti-Blo t 5 and positive IgE serum from atopic Bt+ patient compared to control groups (atopic Bt– and non-atopic patients). A strong reactivity was observed with the monoclonal antibody anti-Blo t 5 to both allergens, particularly to rBlo t 5, but no reactivity was found to total Bt extract as well as to irrelevant BSA protein (Fig. 2C, lane 1). Likewise, a strong IgE reactivity to both recombinant proteins as well as to the total Bt extract was detected in serum from atopic Bt+ patient (Fig. 2C, lane 3) whereas no IgE reactivity was observed in atopic Bt– (Fig. 2C, lane 4) and non-atopic (Fig. 2C, lane 5) sera as well as to irrelevant BSA protein. Controls with diluent only (Fig. 2C, lanes 2 and 6) showed no reactivity.

IgE, IgG1 and IgG4 reactivity to total Bt extract, rBlo t 5 and mBlo t 5

Levels of IgE, IgG1, and IgG4 to the total Bt extract and the recombinant allergens (rBlo t 5 and mBlo t 5) were measured by ELISA in sera of atopic patients (group Bt+ and Bt−) and non-atopic individuals (Fig. 3). In atopic Bt+ patients, the median levels of IgE to the total Bt extract were higher than both recombinant allergens (P <0.001) (Fig. 3A). In the atopic Bt− group, the median IgE levels to Bt total were higher than mBlo t 5 ($P < 0.001$), although borderline IgE reactivity was found for all allergens, with only two patients showing high IgE levels to rBlo t 5. In the non-atopic group, no IgE reactivity was found to all allergens analyzed (Fig. 3A). Also, the percentage of serum samples positive to the total Bt extract in atopic Bt+ patients was higher than the recombinant allergens ($P < 0.05$) (Fig. 3A). No significant difference in IgE positivity was found for the other groups.

IgG1 levels to Bt total, rBlo t 5 and mBlo t 5 were similar in both atopic Bt+ and non-atopic groups (Fig. 3B). In the atopic Bt− group, however, the median IgG1 levels to Bt total were higher as compared to rBlo t 5 and mBlo t 5, but with significant differences only for rBlo t 5 (P <0.05). The IgG1 positivity was similar between total Bt extract and recombinant allergens in all patient groups (Fig. 3B).

Regarding the IgG4 isotype (Fig. 3C), the median IgG4 levels to mBlo t 5 were higher than total Bt extract and rBlo t 5 in the atopic Bt+ group ($P < 0.05$). In atopic Bt– and non-atopic patients, no IgG4 reactivity was found for all allergens analyzed (Fig. 3C). The IgG4 positivity for mBlo t 5 was slightly higher than the other allergens, although significance level was not achieved. In the other groups, no significant difference was found for IgG4 positivity to all allergens.

Correlation and association of the levels of antibody isotypes between recombinant allergens in the atopic Bt+ group

A high positive correlation for IgE ($r_S = 0.7784$; $P < 0.0001$; Fig. 4A) and IgG1 ($r_S = 0.7229$; $P < 0.0001$; Fig. 4B), but low positive correlation for the IgG4 isotype ($r_S = 0.3783$; $P = 0.0044$; Fig. 4C) was found between the recombinant allergens. The association between positivity of IgE to rBlo t 5 and mBlo t 5 showed 29% double-positive samples and only 7% and 9% were single-positive IgE to mBlo t 5 and rBlo t 5, respectively (Fig. 4A). Likewise, 38% samples were double-positive IgG1 and only 4% and 14% were single-positive IgG1 to mBlo t 5 and rBlo t 5 (Fig. 4B). On the other hand, 35% samples were double-positive IgG4 and 27% were single-positive IgG4 to mBlo t 5 (Fig. 4C), showing that single positivity of IgG4 to mBlo t 5 was higher than those found for IgG1 and IgE ($P = 0.0010$).

Change of the antibody isotype reactivity to mBlo t 5 in relation to rBlo t 5

The individual serum samples of the atopic Bt+ patients positive to rBlo t 5 and/or mBlo t 5 were analyzed in parallel for each antibody isotype (Fig. 5A). A slight reduction of the median levels of IgE and increase of IgG1 reactivity from rBlo t 5 to mBlo t 5 was observed, although with no significant difference. On the other hand, the median levels of IgG4 to mBlo t 5 showed a significant increase in relation to rBlo t 5 ($P < 0.05$) (Fig. 5A). The change of antibody isotype reactivity to mBlo t 5 in relation to rBlo t 5 was analyzed and a great variation in the reactivity change was observed, with 44% of samples showing IgE reactivity reduction, from which 28% had reduction $\geq$20% (Fig. 5B). For IgG1, 45% of samples had increased reactivity, with 35% showing increase $\geq$20% whereas 69% of samples showed increased IgG4 reactivity, with 50% exhibiting increase $\geq$20% (Fig. 5B).

Structure-based model simulations of the recombinant allergens

It was possible to carry out coarse-grained molecular dynamics simulations in order to evaluate thermodynamic and kinetic changes in the designed epitope (red curves) against the original wild type protein (blue curves). Fig. 6 shows the structure based model (SBM) simulations using C_α and all-atom models for the thermodynamic and kinetic analysis, respectively. Fig. 6A presents the C_α-model results of the thermodynamic free energy profiles as a function of the reaction coordinate (*F(Q)*) at T_f and the specific heat ($C_v(T)$) in the inset. The reaction of folding/unfolding rBlo t 5 and mBlo t 5 was monitored by Q with the folded ($Q \tilde{} 0.8$) and unfolded ($Q \tilde{} 0.3$) populations, in Fig. 6A, separated by the transition state activation barrier at $Q \tilde{} 0.5$. In Fig. 6A, despite of the fact that the transition temperature from folding to unfolding (T_f), temperature at the peak of $C_v(T)$, of rBlo t 5 and mBlo t 5 were slightly different, both proteins showed to be two state proteins, with a very similar thermodynamic profile and thermal stability of $\Delta F = 2.2kT$ obtained from the simulations. This is an indication that the modeled sequence mBlo t 5 has a stable three-dimensional structure, possibly a functional protein due to the same thermodynamic characteristic of the recombinant rBlo t 5.

Kinetic analyses were performed by using the all-atom representation of the proteins in order to include the amino acid side chain information to the SBM simulations. Fig. 6B shows the root mean square fluctuations (rmsf) in position for each residue of rBlo t 5 and mBlo t 5 allergens obtained during the all-atom model simulations. The residue fluctuations were calculated averaging over trajectories of 200 ns at temperature $0.9T_f$ of each allergen. Dashed vertical lines in Fig. 6B mark the four major IgE binding amino acid residues of rBlo t 5 (Glu76, Asp81, Glu86 and Glu91), previously experimentally identified by the NMR approach [20]. The dynamics of both allergens are similar with high fluctuations for the N-terminals, stable helix bundle and high fluctuations for the binding epitope region (near the dashed lines). However, it is worth noting that the fluctuations of the mBlo t 5 binding epitope region (aligned to rBlo t 5), between the dashed lines, are one residue left-shifted with respect to the original rBlo t 5, although mBlo t 5 has almost the same binding epitope side chain dynamics of the natural rBlo t 5.

Electrostatic interaction optimization of the recombinant allergens

The contribution of the electrostatic energy for each ionizable residue of the rBlo t 5 and mBlo t 5 allergens was investigated by using the TKSA-MC simulations. In Fig. 7A, the average of the charge-charge electrostatic energy interaction of residues Glu76, Asp81, Glu86 and Glu91 have a positive energy value at the rBlo t 5 binding epitope region (from residue 76 to 91), which means that the electrostatic contribution due to these residues are not energetically favorable and it is a less stable region in comparison with the rest of the protein (red bars in Fig. 7A). This instability is associated with the proximity of the negative ionizable groups. The average distance between the negative charged groups is 12.88 Å, whereas in the positive groups the average is 30.60 Å. This difference is mainly caused by residues Glu76, Asp81 and Glu86, which are closer between each other than the other ionizable residues in the binding epitope. This negative charge concentration also causes a negative electrostatic potential in this region (surface diagram in Fig. 7C). The same electrostatic analysis was applied in the mBlo t 5 allergen. The electrostatic energy of the modified mBlo t 5 epitope region (aligned with rBlo t 5) has showed to be more energetic stable than the epitope region of the rBlo t 5 (Fig. 7B). This electrostatic energy stabilization in mBlo t 5 is associated with the inclusion of a positive residue in its epitope region, which causes a more homogeneous charge distribution and a more positive electrostatic potential (surface diagram in Fig. 7D) when compared to the rBlo t 5 binding epitope (Fig. 7C). It suggests that the increase of the stability and the decrease of the average electrostatic energy in the epitope region of mBlo t 5 could be one of the reasons of the change in antigenicity to the IgG4 and the decrease to IgE binding with respect to the recombinant allergen rBlo t 5.

Theoretical prediction and experimental measurements correlation

There is a strong correlation between the IgE binding reduction by the alanine point mutations in rBlo t 5 (data from the NMR experiment [20]) and the average electrostatic energy value ($<W_i>$) of each ionizable residue developed in this study (Fig. 8). The percentage of reduction in IgE binding is promoted by each of the six mutations in the epitopes residues among eight patients (only mutations in which induced more than 20% of reduction in IgE binding were considered for analysis). The correlation $R = 0.834$ suggests that the substitution of the residues that have a

non favorable electrostatic interaction by a neutral residue increases the electrostatic stability in the epitope region and induces the reduction of the interaction between rBlo t 5 and IgE.

4.4 Discussion

Prevalence of the allergic diseases has increased during the past 30 years [14] and exposure to HDM allergens is recognized as the most important risk factor for the development of respiratory allergy. A high incidence in the sensitization to *B. tropicalis* is strongly associated with allergic respiratory diseases in tropical and subtropical regions like Singapore, Malaysia and Brazil, where more than 90% patients with allergic rhinitis have positive SPT [4,20]

In the present study, we evaluated three groups of subjects consisting of patients sensitized to HDM allergens including *B. tropicalis* (atopic Bt+), to HDM allergens but not to *B. tropicalis* (atopic Bt–) and non-atopic subjects. We observed a high degree of cosensitization (>95%) by SPT among *B. tropicalis*, *D. pteronyssinus* and *D. farinae* in the atopic Bt+ group, indicating that the three HDM allergens are important in the sensitization and development of respiratory allergic diseases in genetically predisposed individuals of this region. In addition, monosensitization to *B. tropicalis* was not found within the atopic Bt+ group, reinforcing our previous findings that showed low sensitization (6%) to *B. tropicalis* only [66]. However, a previous study has reported allergenic differences between *B. tropicalis* and *D. pteronyssinus,* showing only low to moderate cross-reactivity between them [6], thus considering the sensitization to *B. tropicalis* as an independent and important cause of allergy [67]. In this context, rhinitis was predominantly found in our atopic Bt+ patients, emphasizing the relevance of the sensitization to *B. tropicalis* in the development of allergic rhinitis.

Likewise, the atopic Bt– group showed high cosensitization (>92%) between *D. pteronyssinus* and *D. farinae* allergens, supporting that these HDMs are source of major allergens worldwide and considering this patient group as an essential control in the evaluation of recombinant *B. tropicalis* allergens for in vitro allergy diagnosis. Also, our data showed a predominance of rhinitis associated with asthma in this group of atopic Bt– patients, suggesting that *D. pteronyssinus* and *D. farinae* rather than *B. tropicalis* allergens can be more engaged with this clinical diagnosis in our patients.

Current diagnostic and therapeutic strategy for allergic diseases is not totally appropriate due to the lack of highly purified allergens. Thus, recombinant allergen production presents an efficient procedure to obtain high yields for immunological

studies as well diagnosis and treatment of allergic disorders [68]. The group 5 allergen, Blo t 5, is the most prevalent among *B. tropicalis* allergens, and IgE reactivity to Blo t 5 was found in 40-60% of *B. tropicalis* [20,69]. In this context, we designed two constructs encoding the full sequence of the Blo t 5 allergen, one with the natural sequence of Blo t 5 (rBlo t 5) and another with a modified sequence in the Blo t 5 α-domains (mBlo t 5) and produced the respective recombinant allergens.

After partial purification, rBlo t 5 and mBlo t 5 could be visualized as monomers, dimers or tetramers in electrophoresis procedures and subsequently confirmed in immunoblots probed with specific monoclonal antibody to Blo t 5. These findings can be explained due to formation of stable oligomers during the purification step in affinity chromatography [70]. In addition, some rBlo t 5 components with intermediate molecular weights were stained in immunoblots, suggesting the formation of proteolytically degraded products, even with the use of protease inhibitors, indicating a probable activity of bacterial metalloproteases. It was interesting to note that the monoclonal antibody to Blo t 5 recognized both rBlo t 5 and mBlo t 5, indicating that the epitope of this clone (4G9) was preserved in mBlo t 5 despite of the rearrangement designed in the natural sequence. This finding was also noted in the modeled mBlo t 5 three-dimensional structure, which share a similar IgE binding epitope within rBlo t 5, despite of the slight differences displayed in their amino acid sequences.

Next, we demonstrated the specific reactivity of rBlo t 5 and mBlo t 5 probed with the monoclonal antibody to Blo t 5 and also Bt-specific IgE antibodies from atopic Bt+ patient, compared to control groups (atopic Bt– and non-atopic subjects) in slot-blot assays. The clone 4G9 detected strongly both allergens, mainly rBlo t 5, but no reactivity was found to total Bt extract, probably due to the poor representation of the native Blo t 5 allergen in this extract [71]. Likewise, it was observed IgE reactivity to both recombinant allergens and to total Bt extract, with more intense staining for rBlo t 5, in the serum from atopic Bt+ patient, whereas no IgE reactivity was detected in control groups.

The analysis of the reactivity of antibody isotypes to the total Bt extract and recombinant allergens from Blo t 5 by ELISA showed that IgE reactivity to the total Bt extract was predominant in atopic Bt+ patients as expected, since other *B. tropicalis* allergenic components of the extract can play a role in the sensitization of these patients. On the other hand, IgE levels and positivity to both rBlo t 5 (38%) and mBlo t 5 (35%) were similar, indicating that the modified form of Blo t 5 was not able to significantly change the IgE binding. In addition, the positivity of IgE to rBlo t 5 was similar to previous studies [69], reinforcing that it is a major sensitizing allergen from *B. tropicalis* in patients of our region. In the atopic Bt– patients, however, borderline IgE

reactivity to all allergens was detected, except for two patients with high IgE reactivity to rBlo t 5, indicating low or moderate cross-reactivity with *D. pteronyssinus* allergens, in accordance to previous reports that showed ˜40% sequence identity between the group 5 allergens from *B. tropicalis* and *D. pteronyssinus* [72]. Interestingly, it was also noted in the atopic Bt– patients that positivity and levels of IgE to mBlo t 5 decreased in comparison to rBlo t 5, and especially to the total Bt extract, indicating that the rearrangement in the mBlo t 5 molecule was able to reduce this cross-reactivity. Furthermore, the probable residual contaminants in recombinant allergen preparations seem to have not caused any interference in ELISA and immunoblotting results, since no IgE reactivity to these contaminants was found in non-atopic subjects.

Regarding the IgG1 isotype, a similar profile of IgG1 reactivity to total Bt extract and recombinant allergens in atopic Bt+ patients was observed, indicating that IgG1 isotypes might be more closely related to allergen exposure to repeated antigenic stimuli [66]. On the other hand, in the atopic Bt– patients, the IgG1 reactivity to both rBlo t 5 and mBlo t 5 decreased in comparison with the total Bt extract, indicating a reduction in the IgG1 cross-reactivity with other *D. pteronyssinus* allergens. In contrast, it was notable an increased IgG4 reactivity to mBlo t 5 in relation to rBlo t 5 and total Bt extract in atopic Bt+ patients, indicating that the mBlo t 5 allergen showed higher antigenicity to IgG4 than to other isotypes. These findings suggest that mBlo t 5 might be a valuable tool to be employed in allergen specific immunotherapy procedures to assure its potential effect to induce IgG4-producing B cell subsets.

The antibody isotype reactivity to rBlo t 5 and mBlo t 5 was also compared in atopic Bt+ patients, showing a high positive correlation for IgE and IgG1 antibodies to rBlo t 5 and mBlo t 5 in addition to high rates of concordant results between the recombinant allergens, since the percentage of single-positive IgE or IgG1 samples to mBlo t 5 was very low. These findings corroborate that the mBlo t 5 allergen still maintain the major structural features of rBlo t 5, including several preserved epitopes recognized by these isotypes. On the other hand, the lower positive correlation for IgG4 to rBlo t 5 and mBlo t 5 associated with a higher percentage of single-positive IgG4 samples to mBlo t 5 as compared to other isotypes suggest that the modified form of Blo t 5 has allowed a higher antigenicity to the IgG4 isotype. This feature is particularly attractive in approaches to produce new recombinant allergens with potential application in allergen-specific immunotherapy.

When serum samples of the atopic Bt+ patients positive to rBlo t 5 and/or mBlo t 5 were individually compared for each antibody isotype, the change in the median levels of IgE and IgG1was not significant, although there was a great individual variation in the antibody reactivity. In contrast, the median levels of IgG4 were sig-

nificantly increased from rBlo t 5 to mBlo t 5. This increase in the IgG4 antigenicity was notably reinforced when the degree of this change was analyzed, since 69% of the samples had increased IgG4 reactivity compared to 45% of increase for IgG1 and 44% of reduction for IgE. This IgG4 reactivity change occurred in 50% of samples in relation to 35% for IgG1 and 28% for IgE reactivity, considering as significant a reactivity change $\geq$ 20% as previously proposed [20]. These findings suggest that the modified Blo t 5 allergen, designed in the present study, showed a shift of antigenicity for the IgG4 isotype, thus making suitable its use as a potential candidate in immunotherapy studies. In this context, recombinant allergen oligomers, found here during the electrophoresis procedures, represent an interesting variation, because though maintain their IgE reactivity, the stable oligomers loose allergenic activity due to the altered presentation of the IgE epitopes [73].

Further theoretical studies, followed by experimental trials should be required to evaluate the hypoallergenic potential and functional activity of this mBlo t 5 allergen in basophil degranulation assays [69] as well as the cytokine profile secreted by peripheral blood mononuclear cells from atopic patients sensitized to *Blomia tropicalis* after *in vitro* stimulation with rBlo t 5 and mBlo t 5.

These experimental findings were corroborated by the theoretical approach, which also proposed the origins of the shift in antigenicity for mBlo t 5 in the molecular level. Computer-based approaches have been widely applied in conjunction with experimental and clinical assays and this combination is playing a key role on the frontier science researches in immunotherapy studies [73–75]. *In silico* methods can reduce significantly the blindness and the time consuming of experiments [76] and clinical trials [77]. Molecular dynamics simulation is able to dissect the relevant processes of biological mechanisms in their ultimate detail [78], although it could be very computer-time consuming requiring high-performance computer cluster facilities, even for small biomolecules [79]. In this context, here we characterize a novel approach to evaluate computer-designed epitopes by simplified models that could increase the allergen antigenicity for the IgG isotypes.

First, the thermodynamic and kinetics of the modified Blo t 5 allergen were characterized. We certified that the designed allergen has the same global features of rBlo t 5 protein: almost identical free energy profile and thermo-stability ($\Delta F = 2.2kT$) and backbone with side chains dynamics (similar rmsf for the projected epitope). These findings confirm that the modified allergen has its helical bundle dynamically stable and the IgE-binding epitope highly exposed to solvent due to the same fluctuation profile as observed in rBlo t 5, which agrees with the original allergen structure resolved by NMR as in protein dynamic analyses [20,72].

Second, the electrostatic interactions of the charge residues located at the IgE-binding epitope and at the rest of the protein are balanced. Previous works showed that specific charged residues located at the HDM allergens surface play a crucial role in the interaction between allergen-antibody [21,73,80] . In this study, the electrostatic interactions among each ionizable residue in the mBlo t 5 [72] were stabilized (less negative energy values) in comparison with rBlo t 5. The protein charge-charge interactions were determined by a novel TKSA-MC method, modified to be more computer-efficient. The average charge-charge interaction method has been used to understand the thermo-stabilization of proteins and to redesign enzymes by optimization of the surface electrostatic interactions [25,52–56].

Chan and co-workers made point mutations in the wild-type Blo t 5 by changing the ionizable residues in the epitope region by a neutral one (alanine) [20]. The allergen mutants were capable to decrease IgE binding and also to reduce cutaneous reaction in the SPT. It is interesting to observe in the electrostatic calculations that these effects were more significant when the mutation was made in the residues that show a positive electrostatic energy (unstable) determined by our method. The charged key residues in the natural allergen IgE-binding epitope (Glu76, Asp81, Glu86 and Glu91), determined by NMR [69], showed electrostatic energy stabilization in the new antibody-binding epitope, due to the insertion of a positive charge after rearrangement of the three α-helices.

This is a worthy validation of our model as the experimental and computer simulation converged in a great correlation ($R = 0.83$), considering that it is a simplified and low time-consuming model. The electrostatic energy stabilization in the epitope region, as well as in the rest of mBlo t 5, suggests a possible explanation for the shift in antigenicity for IgG4 isotype when compared to the rBlo t 5. It is recently known that refined differences in composition and distribution of charges in the epitope may account for the difference in the IgE binding specificities among several allergen groups in *B. tropicalis*, *D. pteronyssinus* and *D. farinae* [21]. Thus, the coarse-grained model along with the TKSA-MC simulation seems to be a powerful approach to investigate future *in silico* point mutations by high-throughput screening in order to support experimental design of new candidates in allergen-specific immunotherapy studies.

4.5 Acknowledgements

We thank Dr. Carlos Ueira (Laboratory of Nanobiotechnology, UFU) by supporting with the protocols for expression and purification of recombinant allergens

and GridUnesp network for the computational resources. We also thank the Brazilian funding agencies CNPq (441730/2014-0), CAPES, FAPEMIG (APQ-03046-10; PPM-00348-12, APQ-00941-14) and FAPESP (Proc. 2014/06862-7) for financial support.

4.5.1 Author Contributions

Conceived and designed the experiments: EAT CRP DAOS JPCJr. Performed the gene design of mBlo t 5 and protein purification: CRP BGMA LFT JPCJr. Performed all immunoassays: BGMA LFT KCA DAOS. Developed the molecular modeling and simulations of Blot 5 proteins: OAJ PAC VMO VGC VBPL RJO. Analyzed the data: EAT DAOS BGMA LFT CRP KCA JPCJr OAJ PAC VMO VGC RJO. Contributed reagents/materials/analysis tools: CRP OAJ VMO RJO JPCJr EAT. Wrote the manuscript: BGMA DAOS KCA OAJ VMO RJO JPCJr EAT.

4.6 References

[1] Masoli M, Fabian D, Holt S, Beasley R. The global burden of asthma: Executive summary of the GINA Dissemination Committee Report. Allergy Eur J Allergy Clin Immunol. 2004;59: 469–478.

[2] Kaw YC, Cheong N, Kuo I-C, Bee WL, Fong CY, Huang C-H, et al. The Blomia tropicalis allergens. Protein Pept Lett. 2007;14: 325–333.

[3] Vrtala S, Huber H, Thomas WR. Recombinant house dust mite allergens. Methods. 2014;66: 67–74.

[4] Fernandez-Caldas E, Puerta L, Mercado D, Lockey RF, Caraballo LR. Mite fauna, Der pI Der fI and Blomia tropicalis allergen levels in a tropical environment. Clin Exp Allergy. 1993;23: 292–297.

[5] Zhang L, Chew FT, Soh SY, Yi FC, Law SY, Goh DYT, et al. Prevalence and distribution of indoor allergens in Singapore. Clin Exp Allergy. 1997;27: 876–885.

[6] Chew FT, Yi FC, Chua KY, Fernandez-Caldas E, Arruda LK, Chapman MD, et al. Allergenic differences between the domestic mites Blomia tropicalis and Dermatophagoides pteronyssinus. Clin Exp Allergy. 1999;29: 982–988.

[7] Carvalho KDA, de Melo-Neto OP, Magalhães FB, Ponte JCM, Felipe FAB, dos Santos MCA, et al. Blomia tropicalis Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. BMC Immunol. 2013;14.

[8] Cheong N, Yang L, Lee BW, Chua KY. Cloning of a group 3 allergen from Blomia tropicalis mites. Allergy Eur J Allergy Clin Immunol. 2003;58: 352–356.

[9] Kuo IC, Cheong N, Trakultivakorn M, Lee BW, Chua KY. An extensive study of human IgE cross-reactivity of Blo t 5 and der p 5. J Allergy Clin Immunol. 2003;111: 603–609.

[10] Yeoh SM, Kuo IC, Wang DY, Liam CK, Sam CK, De Bruyne JA, et al. Sensitization profiles of Malaysian and Singaporean subjects to allergens from Dermatophagoides pteronyssinus and Blomia tropicalis. Int Arch Allergy Immunol. 2003;132: 215–220.

[11] Manolio TA, Barnes KC, Naidu RP, Levett PN, Beaty TH, Wilson AF. Correlates of sensitization to Blomia tropicalis and Dermatophagoides pteronyssinusin asthma in Barbados. Int Arch Allergy Immunol. 2003;131: 119–126.

[12] Geraldini M, Filho N a R, Fernandes F, Castro M, Seba J, Rubini NPM. Alérgenos recombinantes na prática da imunoterapia * Recombinant allergens in immunotherapy practice. 2008; 92–97.

[13] Smith HE, Hogger C, Lallemant C, Crook D, Frew AJ. Is structured allergy history sufficient when assessing patients with asthma and rhinitis in general practice? J Allergy Clin Immunol. 2009;123: 646–650.

[14] Yi FC, Chua K-Y, Cheong N, Shek LP, Lee BW. Immunoglobulin E reactivity of native Blo t 5, a major allergen of Blomia tropicalis. Clin Exp Allergy. 2004;34: 1762–1767.

[15] Tan KW, Jobichen C, Ong TC, Gao YF, Tiong YS, Wong KN, et al. Crystal Structure of Der f 7, a Dust Mite Allergen from Dermatophagoides farinae. PLoS One. 2012;7.

[16] Chruszcz M, Chapman MD, Vailes LD, Stura EA, Saint-Remy J-M, Minor W, et al. Crystal Structures of Mite Allergens Der f 1 and Der p 1 Reveal Differences in Surface-Exposed Residues that May Influence Antibody Binding. J Mol Biol. 2009;386: 520–530.

[17] Kosmoliaptsis V, Dafforn TR, Chaudhry AN, Halsall DJ, Bradley JA, Taylor CJ. High-resolution, three-dimensional modeling of human leukocyte antigen class I structure and surface electrostatic potential reveals the molecular basis for alloantibody binding epitopes. Hum Immunol. 2011;72: 1049–1059.

[18] Fukunaga A, Tsumoto K. Improving the affinity of an antibody for its antigen via long-range electrostatic interactions. Protein Eng Des Sel. 2013;26: 773–780.

[19] Chan SL, Ong ST, Ong SY, Chew FT, Mok YK. Nuclear magnetic resonance structure-based epitope mapping and modulation of dust mite group 13 allergen as a hypoallergen. J Immunol. 2006;176: 4852–4860.

[20] Chan SL, Ong TC, Gao YF, Tiong YS, Wang DY, Chew FT, et al. Nuclear magnetic resonance structure and IgE epitopes of Blo t 5, a major dust mite allergen. J Immunol (Baltimore, Md 1950). 2008;181: 2586–2596.

[21] Tan KW, Ong TC, Gao YF, Tiong YS, Wong KN, Chew FT, et al. NMR structure and IgE epitopes of Blo t 21, a major dust mite allergen from Blomia tropicalis. J Biol Chem. 2012;287: 34776–34785.

[22] Tanford C, Kirkwood JG. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. J Am Chem Soc. American Chemical Society; 1957;79: 5333–5339.

[23] Havranek JJ, Harbury PB. Tanford–Kirkwood electrostatics for protein modeling. Proc Natl Acad Sci U S A. 1999;96: 11145–11150.

[24] Spector S, Wang M, Carp SA, Robblee J, Hendsch ZS, Fairman R, et al. Rational modification of protein stability by the mutation of charged surface residues. Biochemistry. 2000;39: 872–879.

[25] Permyakov SE, Makhatadze GI, Owenius R, Uversky VN, Brooks CL, Permyakov EA, et al. How to improve nature: study of the electrostatic properties of the surface of alpha-lactalbumin. Protein Eng Des Sel PEDS. 2005;18: 425–433.

[26] Makhatadze GI, Loladze V V, Gribenko A V, Lopez MM. Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. J Mol Biol. 2004;336: 929–942.

[27] Naganathan AN. A rapid, ensemble and free energy based method for engineering protein stabilities. J Phys Chem B. 2013;117: 4956–4964.

[28] Oppenheimer J, Nelson HS. Skin testing: A survey of allergists. Ann Allergy, Asthma Immunol. 2006;96: 19–23.

[29] Capriles PVSZ, Guimarães ACR, Otto TD, Miranda AB, Dardenne LE, Degrave WM. Structural modelling and comparative analysis of homologous, analogous and specific proteins from Trypanosoma cruzi versus Homo sapiens: Putative drug targets for chagas' disease treatment. BMC Genomics. 2010;11.

[30] Jorgensen WL, Maxwell DS, Tirado-rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. 1996;7863: 11225–11236.

[31] Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, et al. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J Comput Chem. John Wiley & Sons, Inc.; 1990;11: 440–467.

[32] Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. J Biol Chem. 1951;193: 265–275.

[33] Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc Natl Acad Sci U S A. 1979;76: 4350–4354.

[34] Guimarães Junqueira De Queirós M, Oliveira Silva DA, Alves R, Fukuhara Chiba H, Soares De Amaral VB, De Almeida KC, et al. Mite-specific immunotherapy using allergen and/or bacterial extracts in atopic patients in Brazil. J Investig Allergol Clin Immunol. 2008;18: 84–92.

[35] Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: What determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. J Mol Biol. 2000;298: 937–953.

[36] Fersht AR. Characterizing transition states in protein folding: an essential step in the puzzle. Curr Opin Struct Biol. 1995;5: 79–84.

[37] Matouschek A, Kellis JT, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature. 1989;340: 122–126.

[38] Chavez LL, Onuchic JN, Clementi C. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. J Am Chem Soc. 2004;126: 8426–8432.

[39] Snow CD, Sorin EJ, Rhee YM, Pande VS. How well can simulation predict protein folding kinetics and thermodynamics? Annu Rev Biophys Biomol Struct. 2005;34: 43–69. doi:10.1146/annurev.biophys.34.04

[40] Socci ND, Onuchic JN, Wolynes PG. Diffusive dynamics of the reaction coordinate for protien folding funnels. J Chem Phys. 1996;104: 5860–5868.

[41] Yang S, Onuchic JN, Levine H. Effective stochastic dynamics on a protein folding energy landscape. J Chem Phys. 2006;125: 54910–54918.

[42] Best RB, Hummer G. Coordinate-dependent diffusion in protein folding. Proc Natl Acad Sci U S A. 2010;107: 1088–1093.

[43] Chahine J, Oliveira RJ, Leite VBP, Wang J. Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. Proc Natl Acad Sci U S A. 2007;104: 14646–14651.

[44] Oliveira RJ, Whitford PC, Chahine J, Wang J, Onuchic JN, Leite VBP. The origin of nonmonotonic complex behavior and the effects of nonnative interactions on the diffusive properties of protein folding. Biophys J. 2010;99: 600–608.

[45] Contessoto VG, Lima DT, Oliveira RJ, Bruni AT, Chahine J, Leite VBP. Analyzing the effect of homogeneous frustration in protein folding. Proteins Struct Funct Bioinforma. 2013;81: 1727–1737.

[46] Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. Nucleic Acids Res. 2010;38: W657–W661.

[47] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. J Comput Chem. 2005;26: 1701–1718.

[48] Oliveira RJ, Whitford PC, Chahine J, Leite VBP, Wang J. Coordinate and time-dependent diffusion dynamics in protein folding. Methods. 2010;52: 91–98.

[49] Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81: 3684–3690.

[50] Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. Phys Rev Lett. 1988;61: 2635–2638.

[51] Matthew JB, Gurd FRN. [18] Stabilization and destabilization of protein structure by charge interactions [Internet]. Methods in Enzymology. 1986. pp. 437–453.

[52] Ibarra-Molero B, Loladze V V, Makhatadze GI, Sanchez-Ruiz JM. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. Biochemistry. 1999;38: 8138–8149.

[53] Petersen MTN, Martel P, Petersen EI, Drabløs F, Petersen SB. [7] Surface and electrostatics of cutinases. In: Byron Rubin EAD, editor. Methods in Enzymology. Academic Press; 1997. pp. 130–154.

[54] Ibarra-Molero B, Sanchez-Ruiz JM. Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins. J Phys Chem B. 2002;106: 6609–6613.

[55] Strickler SS, Gribenko A V, Gribenko A V, Keiffer TR, Tomlinson J, Reihle T, et al. Protein stability and surface electrostatics: a charged relationship. Biochemistry. 2006;45: 2761–2766.

[56] Gribenko A V, Patel MM, Liu J, McCallum SA, Wang C, Makhatadze GI. Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. Proc Natl Acad Sci U S A. 2009;106: 2601–2606.

[57] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Phys. 1953;21: 1087–1092.

[58] Ullner M, Jönsson B. A Monte Carlo Study of Titrating Polyelectrolytes in the Presence of Salt. Macromolecules. 1996;29: 6645–6655.

[59] De Oliveira VM, de Carvalho SJ. Adsorption of pH-responsive polyelectrolyte chains onto spherical macroions. Eur Phys J E. 2014;37.

[60] Fersht A. Enzyme Structure and Mechanism. New York: W H Freeman & Co; 1985.

[61] Creighton TE. Proteins: Structures and Molecular Properties. Second Edition. New York: W. H. Freeman; 1992.

[62] Tsodikov, Record M, Sergeev Y. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J Comput Chem. 2002;23: 600–609.

[63] Humphrey W, Dalke A, Schulten K. : Visual molecular dynamics. J Mol Graph. 1996;14: 33–38. 64. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci. 2001;98: 10037–10041.

[64] Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. 2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 2007;35: W522–W525.

[65] Pereira EAL, Silva DAO, Cunha Jr. JP, Almeida KC, Alves R, Sung SJ, et al. IgE, IgG1, and IgG4 antibody responses to Blomia tropicalis in atopic patients. Allergy Eur J Allergy Clin Immunol. 2005;60: 401–406.

[66] Arruda LK, Vailes LD, Platts-Mills TA, Fernandez-Caldas E, Montealegre F, Lin KL, et al. Sensitization to Blomia tropicalis in patients with asthma and identification of allergen Blo t 5. Am J Respir Crit Care Med. American Thoracic Society - AJRCCM; 1997;155: 343–350.

[67] Lim LH, Li HY, Cheong N, Lee BW, Chua KY. High-level expression of a codon optimized recombinant dust mite allergen, Blo t 5, in Chinese hamster ovary cells. Biochem Biophys Res Commun. 2004;316: 991–996.

[68] Tsai J-J, Yi FC, Chua K-Y, Liu Y-H, Lee BW, Cheong N. Identification of the major allergenic components in Blomia tropicalis and the relevance of the specific IgE in asthmatic patients. Ann Allergy, Asthma Immunol. 2003;91: 485–489.

[69] Pramanik A, Hauf W, Hoffmann J, Cernescu M, Brutschy B, Braun V. Oligomeric Structure of ExbB and ExbB-ExbD Isolated from Escherichia coli As Revealed by LILBID Mass Spectrometry. Biochemistry. American Chemical Society; 2011;50: 8950–8956.

[70] Valenta R, Niespodziana K, Focke-Tejkl M, Marth K, Huber H, Neubauer A, et al. Recombinant allergens: What does the future hold? J Allergy Clin Immunol. 2011;127: 860–864.

[71] Naik MT, Chang C-F, Kuo I-C, Kung CC-H, Yi F-C, Chua K-Y, et al. Roles of structure and structural dynamics in the antibody recognition of the allergen proteins: an NMR study on Blomia tropicalis major allergen. Structure. 2008;16: 125–36.

[72] Campana R, Vrtala S, Maderegger B, Dall'Antonia Y, Zafred D, Blatt K, et al. Altered IgE epitope presentation: A model for hypoallergenic activity revealed for Bet v 1 trimer. Mol Immunol. 2011;48: 431–441.

[73] Tai H-Y, Zhou J-K, Chou H, Tam MF, Chen Y-S, Sheu S-Y, et al. Epitope Mapping and In Silico Characterization of Interactions between Der p 7 Allergen and MoAb WH9. PLoS One. 2013;8.

[74] Caravella JA, Wang D, Glaser SM, Lugovskoy A. Structure-Guided design of antibodies. Curr Comput Aided Drug Des. 2010;6: 128–138.

[75] Hakkaart GAJ, Aalberse RC, Van Ree R. Epitope mapping of the house-dust-mite allergen Der p 2 by means of site-directed mutagenesis. Allergy Eur J Allergy Clin Immunol. 1998;53: 165–172.

[76] Moreira PF de S, Gangl K, Vieira F de AM, Ynoue LH, Linhart B, Flicker S, et al. Allergen Microarray Indicates Pooideae Sensitization in Brazilian Grass Pollen Allergic Patients. PLoS One. Public Library of Science; 2015;10: e0128402.

[77] Cheng X, Ivanov I. Molecular Dynamics. In: Reisfeld B, Mayeno AN, editors. Computational Toxicology SE - 11. Humana Press; 2012. pp. 243–285.

[78] Zhang D, Chen C-F, Zhao B-B, Gong L-L, Jin W-J, Liu J-J, et al. A novel antibody humanization method based on epitopes scanning and molecular dynamics simulation. PLoS One. 2013;8.

[79] Ebo DG, Bridts CH, Mertens CH, Hagendorens MM, Stevens WJ, De Clerck LS. Analyzing histamine release by flow cytometry (HistaFlow): A novel instrument to study the degranulation patterns of basophils. J Immunol Methods. 2012;375: 30–38.

Table 1: Demographic and clinical characteristics of the subjects in study.

Characteristics	Groups			*p* value
	Atopic Bt+	Atopic Bt-	Non-atopic	
Number of individuals	55	25	30	
Sex				
Male	20	7	5	0.2476[a]
Female	35*	18*	25*	
Age (years)				
Median (range)	28 (18-55)	34 (19-60)	29 (19-59)	0.2532[b]
Positive SPT (n, %)				
B. tropicalis	55 (100%)	0	0	ND
D. pteronyssinus	53 (96.4%)	24 (96%)	0	1.0000[a]
D. farinae	52 (94.5%)	23 (92%)	0	0.6453[a]
Wheal size (mean ± SD, mm)				
B. tropicalis	6.3 ± 3.2	0	0	ND
D. pteronyssinus	7.1 ± 3.6	6.6 ± 3.1	0	0.5325[c]
D. farinae	7.2 ± 3.4	6.6 ± 3.2	0	0.4239[c]
Clinical diagnosis (n, %)				
Rhinitis	51 (92.8%)*	11 (44%)	0	<0.0001[a]
Rhinitis + Asthma	4 (7.2%)	10 (40%)	0	0.0008[a]
Asymptomatic	0	4 (16%)*	0	ND

Bt+: patients with positive skin prick test (SPT) to *Blomia tropicalis* extract;

Bt−: patients with negative SPT to *B. tropicalis* extract.

Values of *p* represent comparisons done horizontally between groups and asterisks (*) represent comparisons done vertically between the parameters analyzed (sex and clinical diagnosis) within each subject group. ND: not done.

[a] Chi-square test or Fisher exact probability test, when appropriate;

[b] Kruskal-Wallis and Dunn multiple comparison tests;

[c] ANOVA and Tukey multiple comparison tests.

4.7 Figures

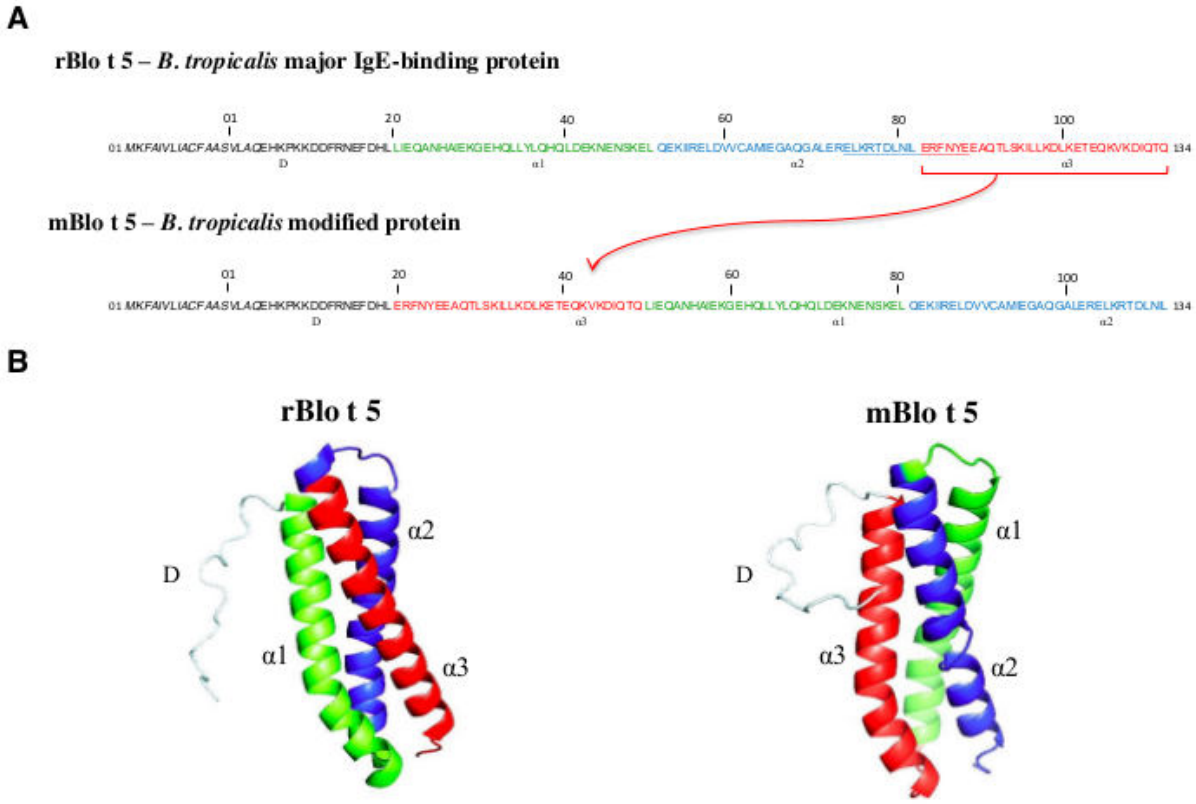


Figure 4.1: Amino acid sequences of *Blomia tropicalis* group 5 recombinant allergen, the natural sequence (rBlo t 5) and the modified allergen (mBlo t 5) constructed by a shift of the $\alpha 3$ helix of the rBlo t 5 (indicated by the red arrow). Colored letters represent the four domains, one disorder/unstructured N-domain followed by three α-helices (D, $\alpha 1$, $\alpha 2$ and $\alpha 3$), rearranged in the rBlo t 5 allergen to produce the modified protein (D, $\alpha 3$, $\alpha 1$ and $\alpha 2$).

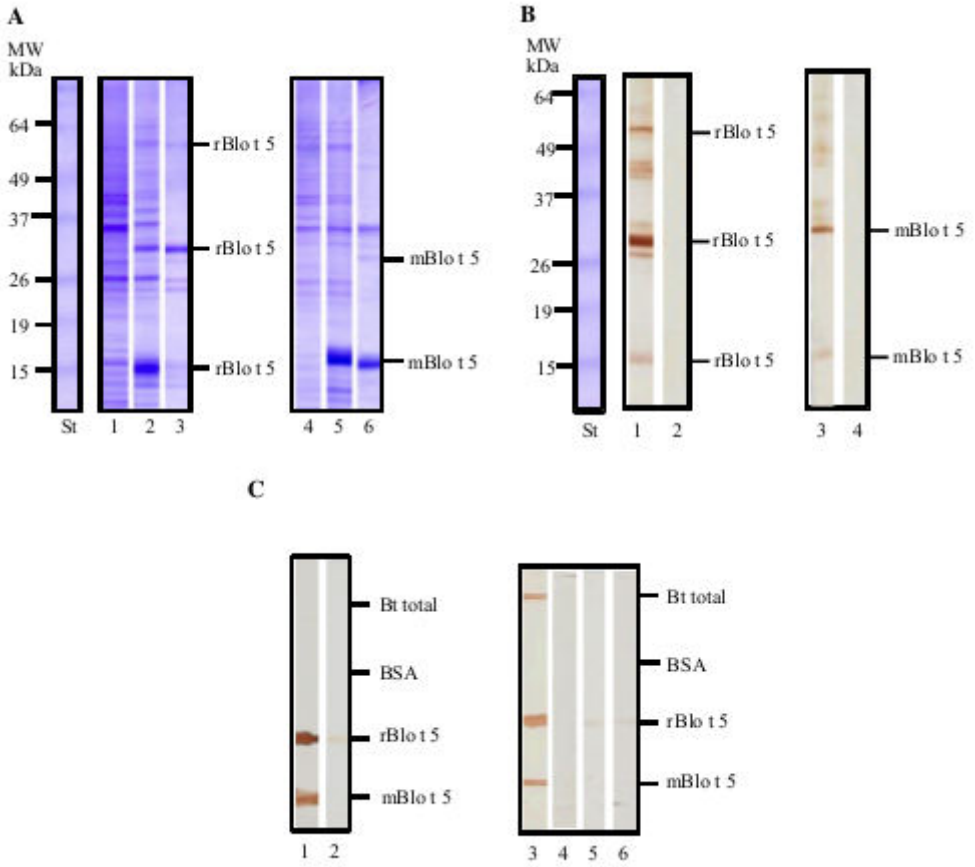


Figure 4.2: (**A**) Expression of the *Blomia tropicalis* recombinant Blo t 5 allergen (rBlo t 5) and the modified allergen (mBlo t 5) in *E. coli* BL21 transformed with pET14b containing the respective genes. The BL21 *E. coli* strain was treated with IPTG 0.5 mM as protein expression inductor. Samples were run in SDS-PAGE at 13.5% and gel was stained with Coomassie blue. Lane 1: Protein extract from BL21 *E. coli* /rBlo t 5 without IPTG; Lane 2: Protein extract from BL21 *E. coli*/rBlo t 5 induced with IPTG; Lane 3: Purified rBlo t 5 protein; Lane 4: Protein extract from BL21 *E. coli* /mBlo t 5 without IPTG; Lane 5: Protein extract from BL21/mBlo t 5 induced with IPTG; Lane 6: Purified mBlo t 5 protein. (**B**) Immunoblots of purified rBlo t 5 (lanes 1 and 2) and mBlo t 5 (lanes 3 and 4) probed with mouse monoclonal antibody anti-Blo t 5 (lanes 1 and 3) or diluent only (lanes 2 and 4) and revealed with biotinylated anti-mouse IgG, streptavidin-peroxidase and DAB. The molecular weight standard (St) expressed in kiloDaltons (kDa) and the position of the allergens (rBlo t 5 and mBlo t 5) are also indicated. (**C**) Slot-blots of rBlo t 5 (2.5 g) and mBlo t 5 (2.5 g) probed with mouse monoclonal antibody anti-Blo t 5 (lane 1) or diluent only (lane 2) and revealed with biotinylated anti-mouse IgG, streptavidin-peroxidase and DAB. Alternatively, membranes were probed with human serum from atopic Bt+ (lane 3), atopic Bt– (lane 4), non-atopic (lane 5) patients and only diluent (lane 6) and the reaction was revealed with biotinylated anti-human IgE, streptavidin-peroxidase and DAB. As controls, bovine serum albumin (BSA; 5 g) and *B. tropicalis* extract (Bt total; 10 g) were also blotted.

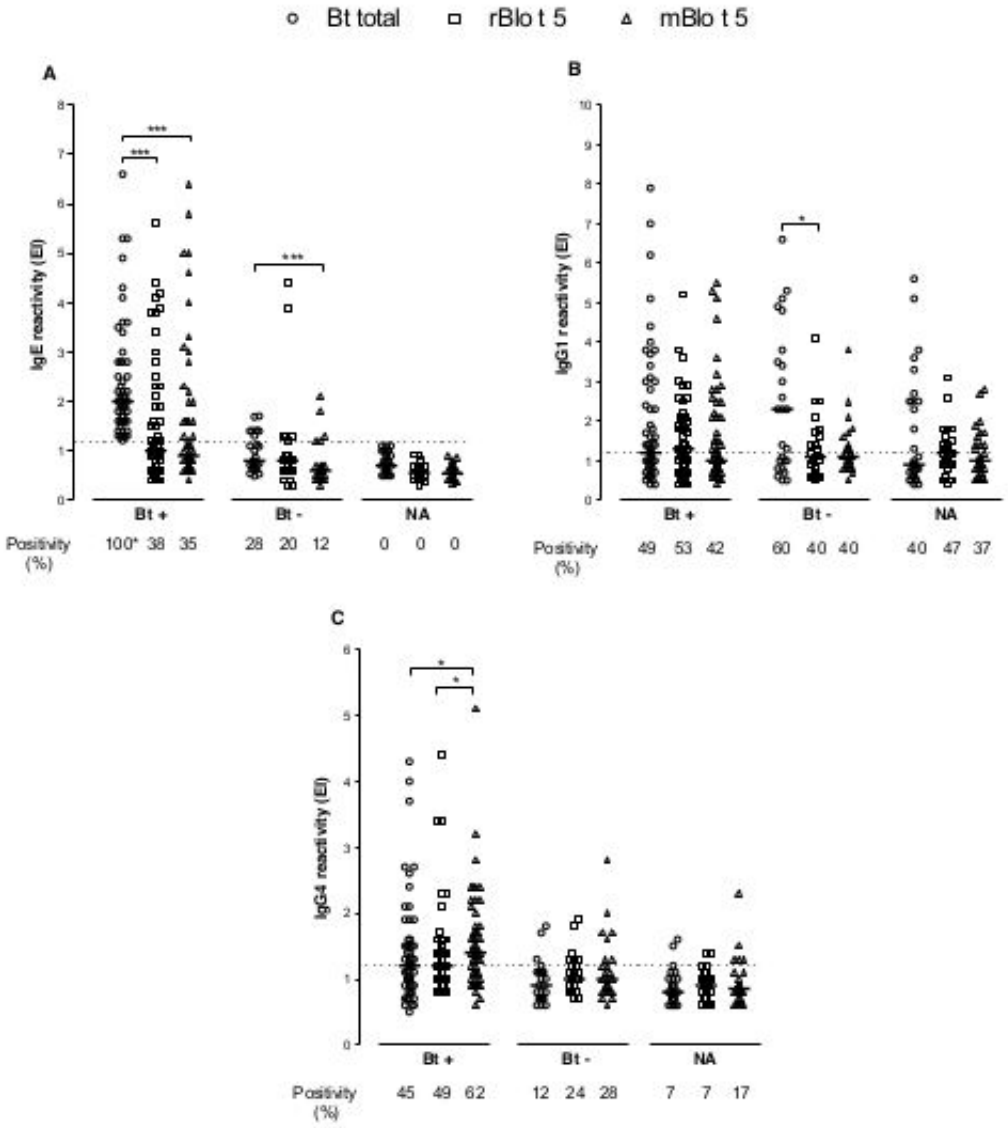


Figure 4.3: Levels of IgE (**A**), IgG1 (**B**) and IgG4 reactivity to (**C**) *Blomia tropicalis* extract (Bt total, circles), recombinant Blo t 5 (rBlo t 5, squares) and modified Blot 5 (mBlo t 5, triangles) allergens determined by ELISA in serum samples from atopic Bt+ ($n = 55$), atopic Bt– ($n = 25$) and non-atopic ($n = 30$) patients. Horizontal bars indicate the median and the dashed lines indicate the ELISA index cutoff value for each reaction ($EI > 1.2$). Comparisons of antibody levels between the allergens within each group were analyzed by the Friedman and Dunn multiple comparison tests ($^*P < 0.05$; *** $P < 0.001$). The positive percentages for each allergen in each patient group are also indicated and statistically significant differences in the positivity within of each group are indicated by asterisk (Fisher exact test; $P < 0.05$).

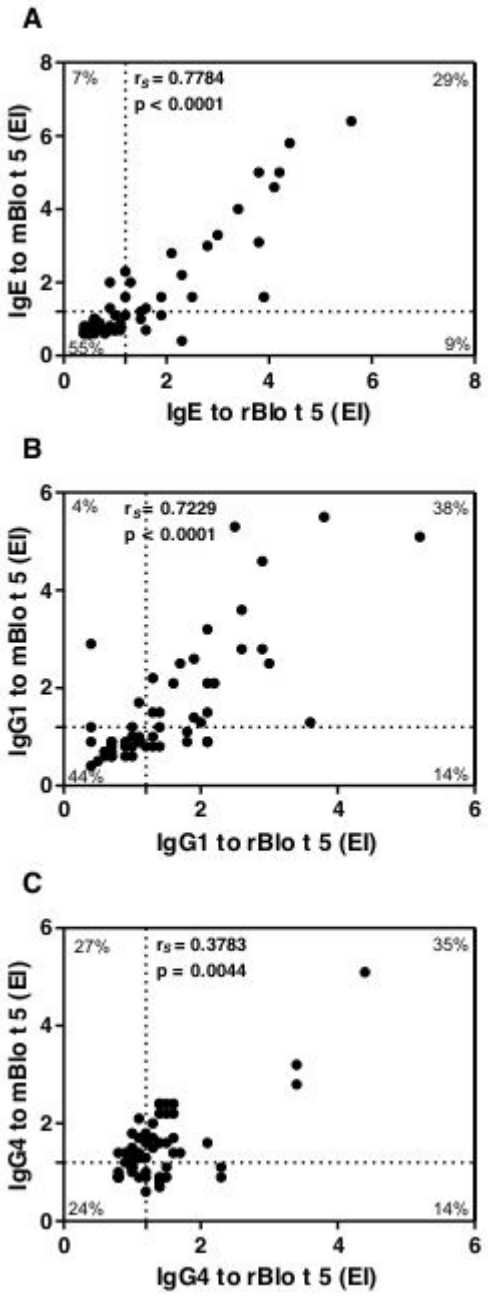


Figure 4.4: Correlation and association of the antibody isotype levels between the rBlo t 5 and mBlo t 5 allergens from *Blomia tropicalis*. Levels of IgE (**A**), IgG1 (**B**) and IgG4 (**C**) reactivity to rBlo t 5 and mBlo t 5 were determined by ELISA in 55 serum samples from atopic Bt+ patients. The dashed lines indicate the ELISA index (*EI*) cutoff value for each reaction ($EI > 1.2$). The Spearman correlation coefficient (r_S) and p value are indicated. Double-positive, double-negative or single-positive percentages are indicated in each correspondent corner.

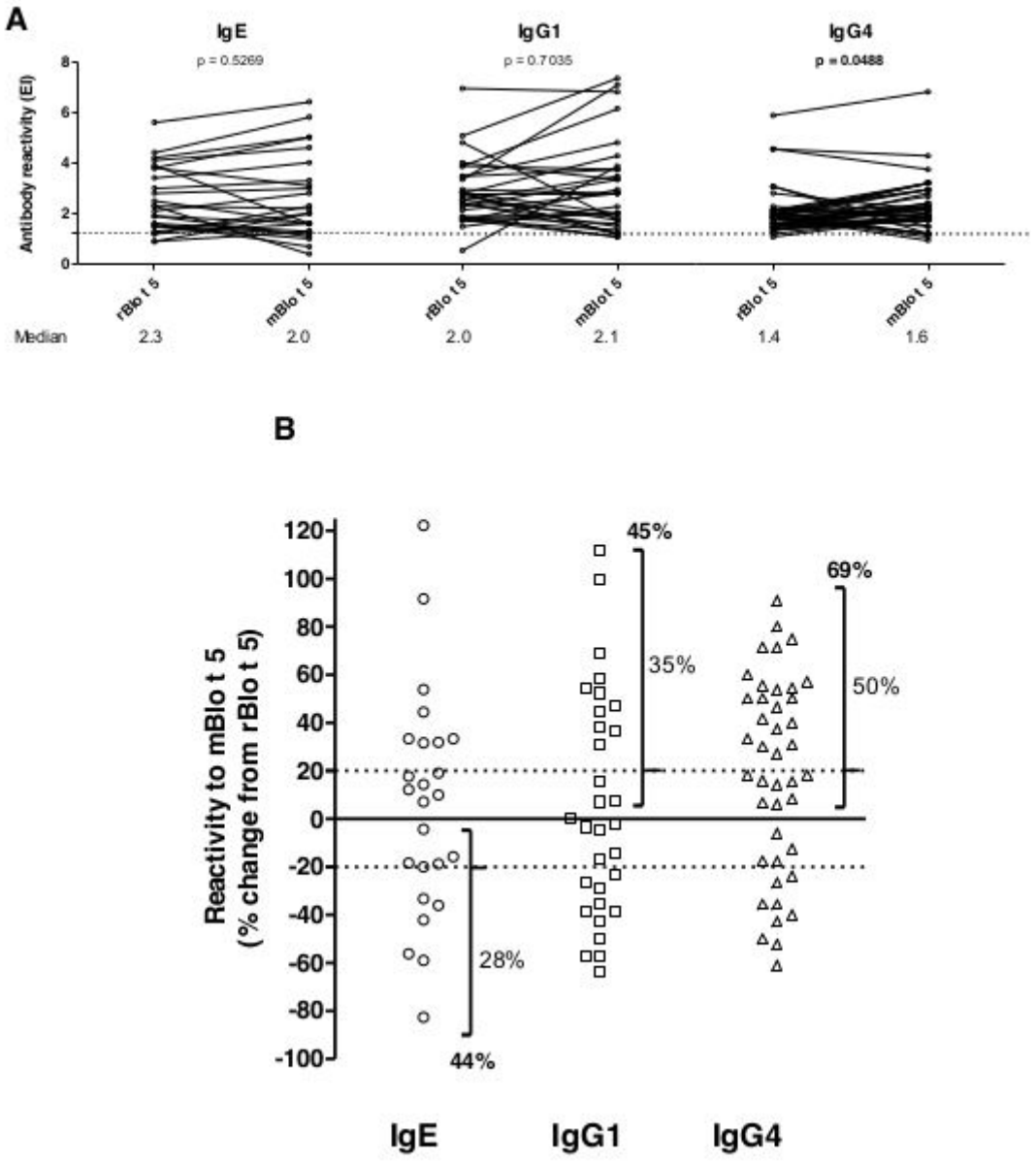


Figure 4.5: (**A**) Individual comparison between the levels of antibody isotypes (IgE, IgG1 and IgG4) to rBlo t 5 and mBlo t 5 determined by ELISA in serum samples from atopic Bt+ patients positive to rBlo t 5 and/or mBlo t 5 (IgE, $n = 25$; IgG1, $n = 31$; IgG4, $n = 42$). The dashed line indicates the ELISA index (EI) cutoff value for each reaction ($EI > 1.2$) and the median values are also indicated. Significant differences were determined by the Wilcoxon rank test ($P < 0.05$). (**B**) Percentages of change of IgE (circles), IgG1 (squares) and IgG4 (triangles) reactivity to mBlo t 5 in relation to rBlo t 5 in serum samples of atopic Bt+ patients positive to rBlo t 5 and/or mBlo t 5. The dashed lines indicate significant reduction or increase ($\geq$20%) and percentages of serum samples included in these ranges are indicated.

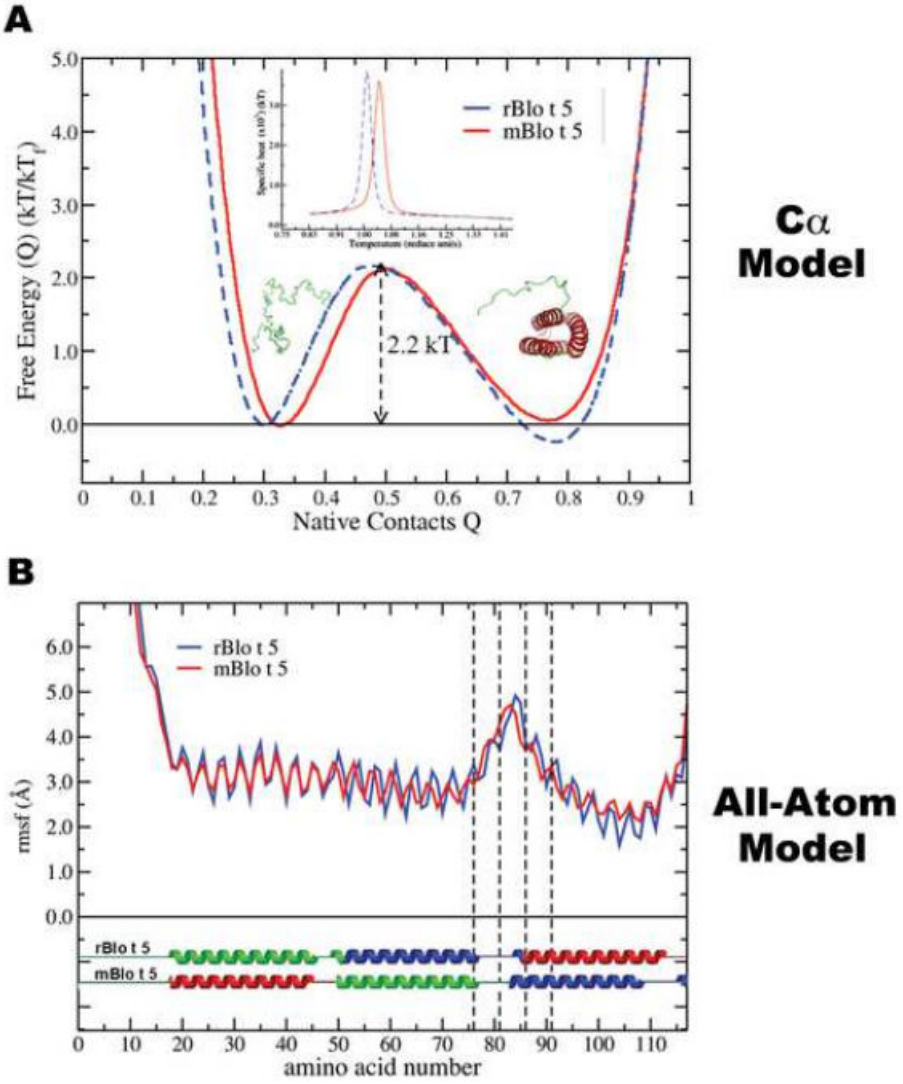


Figure 4.6: Structure-based model (SBM) simulations of the naturally allergen (rBlo t 5, blue curves) and the designed allergen (mBlo t 5, red curves). (**A**) The free energy profiles as a function of the reaction coordinate fraction of native contacts (Q) at the folding temperature (T_f) obtained with the C_α-model simulations. The insert shows the specific heat capacity as a function of the temperature. T_f was defined as the temperature at the specific heat peak of each simulated protein. Both proteins are at the unfolded state when $Q \tilde{} 0.3$ and at the folded state when at $Q \tilde{} 0.8$. The designed protein mBlo t 5 has the same thermal stability ($\Delta F = 2.2kT$) and a very similar thermodynamic free energy profile as the wild-type protein rBlo t 5. (**B**) The protein fluctuations in residue positions using the all-atom model simulations at constant temperature run ($0.9T_f$) given by the root mean square fluctuations (rmsf). Dashed vertical lines represent the four major charged IgE binding epitope residues of rBlo t 5 (Glu76, Asp81, Glu86 and Glu91) and the designed mBlo t 5 epitope residues (Asn76, Lys81, Lys86 and Leu91). The new allergen has similar residues and side chain dynamics as the wild-type allergen, except for one amino acid left shifted at the binding epitope region. Allergens secondary structures are also shown at the bottom in (**B**) with the domains colored as the sequences in Fig. 1. In (**A**) and (**B**), the temperature are show in reduce units and the energy in units of kT where k is the Boltzmann constant.

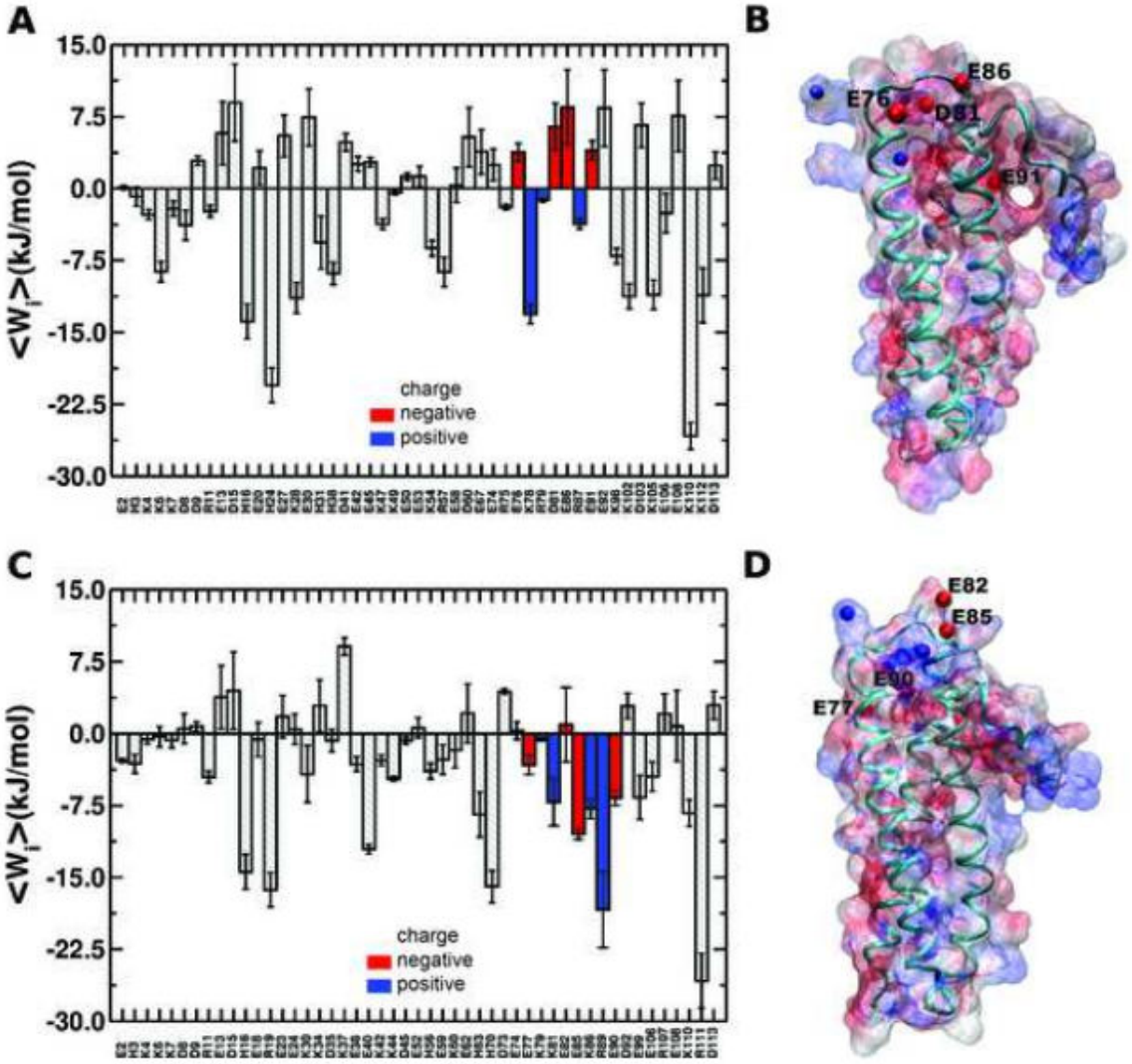


Figure 4.7: Electrostatic interaction optimization calculated in the protein charged residues by the Tanford-Kirkwood Surface Accessibility model implemented using Monte Carlo simulations (TKSA-MC). Average of the charge-charge electrostatic energy $< W_i >$ of each ionizable residue i for (**A**) the recombinant allergen rBlo t 5 and (**B**) the modified mBlo t 5. The charged residues of the binding epitopes of (**A**) rBlo t 5 (E76 to E91) and (**B**) mBlo t 5 (E77 to E90) are shown in blue (positive charges) and red (negative charges) filled bars. Ribbon and electrostatic surface diagrams are show for (**C**) rBlo t 5 and (**D**) mBlo t 5 with the key charged residues of the binding epitopes to IgE represented as spheres located at each amino acid side chain (negative charges are labeled). The designed allergen mBlo t 5 shows a more favorable and stable electrostatic energy, represented by the average charge-charge interaction energy reduction at the epitope region as well as in the overall protein, when compared with the wild-type allergen. In (**C**) and (**D**), all positive charged groups are shown in blue surface, negative charged residues are shown in red and neutral in white. Energy is in unit of kJ/mol and the error bars are the mean standard deviation of the Monte Carlo simulations.

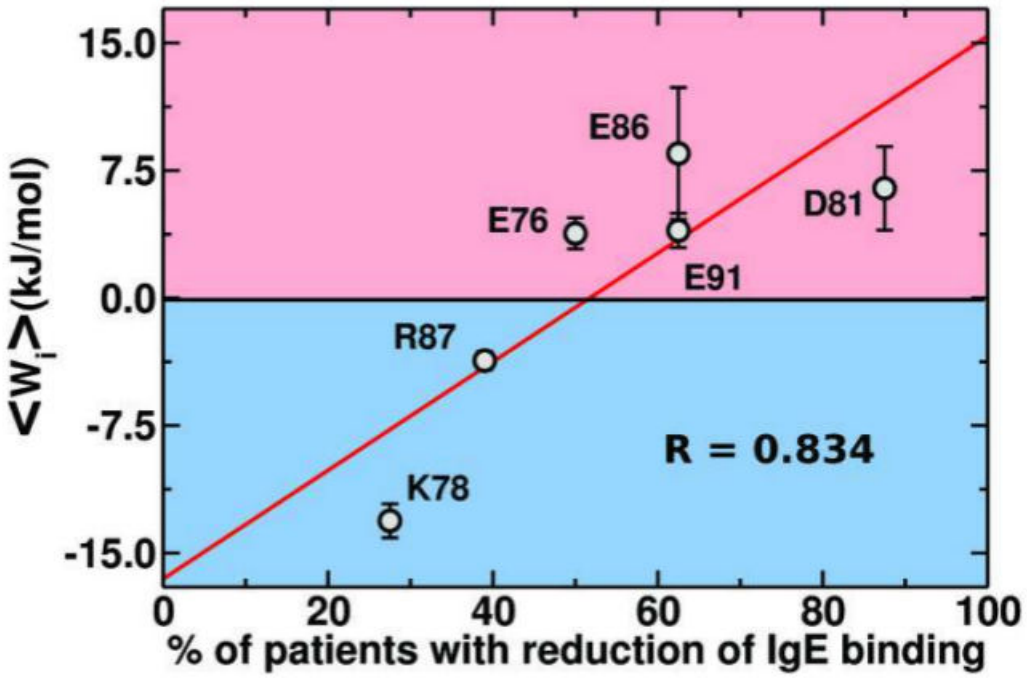


Figure 4.8: Correlation between theoretical and experimental measurements. Average of the charge-charge electrostatic energy interaction $< W_i >$ of the rBlo t 5 epitope residues as functions of the percentage of patients with IgE binding reduction induced by the mutation of these charged residues to alanine (experimental data extracted from Chan and co-workers [34]). The four major IgE binding residues (E76, D81, E86 and E91) are negatively charged and showed an unfavorable electrostatic energy interaction ($< W_i > > 0$) opposing to the minor binding residues (K78 and R87), which are positively charged and showed an favorable electrostatic energy interaction ($< W_i > < 0$). The linear fit correlation of the theoretical calculation to the experimental data shows a strong positive correlation with $R = 0.83$ (linear fit obtained by applying the Pearson correlation [39]). Energy is in unit of kJ/mol and the error bars are the mean standard deviation of the TKSA-MC simulations.

Conclusões Gerais

Nesta tese foram apresentados dois tipos de modelos para o estudo das interações eletrostáticas e de pH em proteínas, o TKSA-MC e o CpHMD. O primeiro trata as proteínas de forma estática e o segundo de forma dinâmica. Tanto individualmentel, quanto a combinação de ambos os modelos foram fundamentais para o entendimento de vários sistemas diferentes. A partir da utilização de TKSA-MC e CpHMD, foi possível caracterizar uma série de propriedades das proteínas, tais como: enovelamento, mudanças conformacionais, estabilidade, mutações e também permitiram inferir sobre a interação proteína-ligante (caso da Blo t 5).

Os resultados de CpHMD tiveram excelente concordância com os experimentos para as várias proteínas simuladas nesse trabalho. Os resultados dessas dinâmicas permitiram ampliar o entendimento das interações entre resíduos carregadas e o sistema de ionização que ocorre durante cada estágio do enovelamento das proteínas alvo. Esse modelo também mostrou-se eficiente para caracterizar os efeitos do pH no enovelamento e na estabilidade de proteínas. No caso da variante da proteína G 1PGB-QDD, não apenas o efeito do pH foi avaliado, mas também o efeito da força iônica. Esse efeito também apresentou boa concordância com os experimentos. Por fim, o CpHMD também permitiu obter valores teóricos de pKa próximos aos experimentais e, com uma acurácia, por vezes, tão boa quanto de modelos computacionais mais sofisticados.

As mudanças no perfil de energia eletrostática causadas por mutações e variações nas condições de pH e concentração de sal também foram calculadas pelo modelo de TKSA-MC para o estado nativo da proteína 1PGB-QDD. Esse cálculo permitiu obter a contribuição energética de cada resíduo ionizável para a estabilidade dessa proteína. Além disso, esse modelo foi importante no estudo da antigenicidade da Blo t 5, na qual mutações alteram o perfil de energia eletrostática de seu epítopo e a tornam mais estável. Com a análise das interações eletrostáticas do Blo t 5 selvagem e mutante, foi possível inferir sobre seu mecanismo de ligação com o anticorpo. Portanto, com os trabalhos apresentados nessa tese, fica evidente que tanto o TKSA-MC, quanto o CpHMD são ferramentas poderosas no estudo dos efeitos eletrostáticos e de pH em proteínas, podendo ser aplicadas em uma ampla gama de problemas biomédicos e biotecnológicos.

Apêndice A

Resolução da equação de Poisson-Boltzmann linearizada em coordenadas esféricas

A equação de Poisson-Boltzmann linearizada

Nesta tese, a solução da equação de Poisson-Boltzmann (PB) linearizada em coordenadas esféricas foi utilizada para adicionar um potencial eletrostático ao potencial do SBM. Neste apêndice, serão apresentadas a equação de PB, sua linearização e solução em coordenadas esféricas. A equação de PB descreve a distribuição do potencial elétrico em uma solução em relação a uma superfície carregada. Desse modo, é possível verificar como as interações eletrostáticas afetam as moléculas presentes nessa solução. Na equação de Poisson A.1, o potencial eletrostático ψ é relacionado com a densidade de carga ρ.

$$\nabla^2\psi(\vec{r}) = -\frac{\rho(\vec{r})}{\epsilon_s\epsilon_0} \tag{A.1}$$

onde ϵ_0 representa a constante de permissividade elétrica do vácuo e ϵ_s é a constante dielétrica do solvente em um meio contínuo e homogêneo. A densidade de carga ρ de uma solução contendo i-ésimas espécies carregadas pode ser definida em função da densidade numérica η_i dessas espécies:

$$\rho(\vec{r}) = ez_i\eta_i(\vec{r}) \tag{A.2}$$

onde e representa a carga elementar e z_i é a valência da i-ésima espécie carregada, respectivamente. Considerando que essa solução está em equilíbrio termodinâmico, assume-se que a distribuição dessas espécies é dada pela distribuição de Boltzmann, ou seja,

$$\eta_i(\vec{r}) = \eta_i^0 \exp\left[-\frac{z_i e\psi(\vec{r})}{k_B T}\right] \tag{A.3}$$

onde η_i^0 representa a densidade de referência da espécie carregada i, onde $\psi(\vec{r}) \to 0$; k_B é a constante de Boltzmann e T é a temperatura absoluta.

A equação de PB para o potencial ψ é obtida substituindo a equação A.2 e A.3 em A.1:

$$\nabla^2\psi(\vec{r}) = -\frac{e}{\epsilon_s\epsilon_0}\sum_i z_i\eta_i^0 \exp\left[-\frac{z_i e\psi(\vec{r})}{k_B T}\right] \tag{A.4}$$

Nos casos em que a energia potencial eletrostática é muito menor que a energia térmica ($z_i e\psi(\vec{r}) \ll k_B T$) tem-se

$$\exp\left[-\frac{z_i e\psi(\vec{r})}{k_B T}\right] \approx 1 - \frac{z_i e\psi(\vec{r})}{k_B T} \tag{A.5}$$

Substituindo-se A.5 em A.4, obtém-se

$$\nabla^2\psi(\vec{r}) = -\frac{e}{\epsilon_s\epsilon_0}\left[\sum_i z_i\eta_i^0 - \sum_i \frac{z_i^2\eta_i^0 e\psi(\vec{r})}{k_B T}\right] \tag{A.6}$$

A eletroneutralidade do sistema faz com que o termo $\sum_i z_i\eta_i^0$ da equação A.6 vá a zero. Assim, obtém-se a forma linearizada da equação de PB, ou equação de Debye-Hückel:

$$\nabla^2\psi(\vec{r}) = \left[\frac{e^2}{\epsilon_s\epsilon_0 k_B T}\sum_i z_i^2\eta_i^0\right]\psi(\vec{r}) = \kappa^2\psi(\vec{r}) \tag{A.7}$$

onde o parâmetro κ^{-1} tem dimensão de comprimento e é conhecido como *comprimento de Debye*. Este parâmetro está relacionado com a blindagem eletrostática entre os corpos carregados produzida pela solução eletrolítica. O comprimento de Debye é inversamente proporcional a concentração de sal da solução.

Solução da equação de Poisson-Boltzmann linearizada em coordenadas esféricas

A solução da equação de Poisson-Boltzmann linearizada em coordenadas esféricas foi utilizada no no método de CpHMD, pois descreve a interação eletrostática entre os resíduos carregados. Nesta seção, a resolução desta equação fornecerá o potencial eletrostático em duas regiões distintas: (i) na região onde não há presença de carga devido ao impedimento estérico causado pelo volume das espécies carregadas; (ii) na região onde existe uma densidade de

carga liquída promovida pelos íons em solução. Desse modo, tem-se as seguintes expressões:

$$\nabla^2\psi_1(\vec{r}) = \quad 0 \quad , \quad 0 < \vec{r} \leq a \tag{A.8}$$

$$\nabla^2\psi_2(\vec{r}) = \quad \kappa\psi_2(\vec{r}) \quad , \quad \vec{r} > a \tag{A.9}$$

onde a delimita cada uma das regiões do espaço. Aplicando agora as equações A.8 e A.9 em coordenas esféricas, onde o potencial eletrostático ψ dependerá apenas da coordenada radial r, obtém-se:

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2\frac{d\psi_1(r)}{dr}\right] = 0 \tag{A.10}$$

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2\frac{d\psi_2(r)}{dr}\right] = \kappa^2\psi_2(r) \tag{A.11}$$

A solução geral da equação A.10, conhecida por equação de Laplace em coordenas esféricas, é dada por:

$$\psi_1(r) = \frac{A}{r} + B \tag{A.12}$$

Já a solução geral para a equação A.11 é dada por

$$\psi_2(r) = \frac{C\exp(\kappa r)}{r} + \frac{D\exp(-\kappa r)}{r} \tag{A.13}$$

As constante de integração A, B, C e D podem ser determinadas por meio das seguintes condições de contorno:

$$\psi_2(r)|_{r\to\infty} = 0 \tag{A.14}$$

$$\psi_1(r)|_{r=a} = \psi_2(r)|_{r=a} \tag{A.15}$$

$$-\nabla\psi_1(r)|_{r=a} = -\nabla\psi_2(r)|_{r=a} \tag{A.16}$$

$$-\oint_{r=a}\nabla\psi_1(r)\cdot\vec{n}dA = \frac{ze}{\epsilon_0\epsilon_s} \tag{A.17}$$

Assim, de posse das condições de contorno A.14 a A.17, as equações A.12 e A.13 podem ser reescritas, respectivamente, como:

$$\psi_1(r) = \quad \frac{ze}{4\pi\epsilon_0\epsilon_s}\left[\frac{1}{r} - \frac{\kappa}{(\kappa a+1)}\right] \quad , \quad 0 < r \leq a \tag{A.18}$$

$$\psi_2(r) = \frac{ze}{4\pi\epsilon_0\epsilon_s r}\frac{\exp[-\kappa(r-a)]}{(\kappa a+1)} \quad , \quad r > a \tag{A.19}$$

Para o caso específico apresentado nesta tese, utilizou-se apenas a expressão A.19, pois assumiu-se que a região de exclusão muito é pequena. Assim, tomando-se o limite $a \to 0$, obtém-se a seguinte equação:

$$\psi_2(r) = \frac{ze}{4\pi\epsilon_s\epsilon_0}\frac{\exp(-\kappa r)}{r} \tag{A.20}$$

Apêndice B

Variação de energia associada ao processo titulação

O pH da solução nas simulações utilizando TKSA-MC e CpHMD foi mantido constante. Para tanto, utilizou-se a variação de energia associada ao processo de titulação em ambos os modelos. Assim, essa expressão de variação de energia é obtida partindo do potencial químico de uma espécie i em uma solução constituída por partículas interagentes:

$$\mu_i = \mu_i^0 + k_B T \ln(\eta_i) + k_B T \ln(\gamma_i) \tag{B.1}$$

onde μ_i representa o potencial químico padrão ou de referência da i-ésima espécie, η_i é a densidade numérica da espécie i e γ_i é o coeficiente de atividade dessa espécie. Assume-se que os íons não sofrem alterações estruturais em sua camada de hidratação. Desse modo, o trabalho necessário para adicionar ou modificar a posição de uma determinada espécie i será proporcional à variação da energia potencial eletrostática, ou seja,

$$k_B T \ln(\gamma_i) \propto z_i e \psi(r) = \Delta U_i \tag{B.2}$$

onde $\psi(r)$ é o potencial eletrostático, z_i é a valência da i-ésima espécie e e é a carga elementar. Assim, substituindo B.2 em B.1 tem-se:

$$\mu_i = \mu_i^0 + k_B T \ln(\eta_i) + \Delta U_i \tag{B.3}$$

Considerando que a concentração de prótons na solução é dada pela distribuição de Boltzmann:

$$[H] = [H^0] \exp\left[-\frac{\Delta U_i}{k_B T}\right] \tag{B.4}$$

onde $[H^0]$ representa a concentração dos prótons no *bulk* da solução. Utilizando agora a equação de Henderson-Hasselbalch associada à equação B.4, obtem-se:

$$\begin{aligned} pH - pK_i &= \log\left\{\exp\left[-\frac{\Delta U_i}{k_B T}\right]\right\} + \log\left(\frac{\alpha}{1-\alpha}\right) \\ 10^{pH-pK_i} &= \left(\frac{\alpha}{1-\alpha}\right)\exp\left[-\frac{\Delta U_i}{k_B T}\right] \end{aligned} \tag{B.5}$$

onde α representa o grau de ionização da espécie i. A equação B.5 pode ser reescrita da seguinte forma:

$$\Delta U_i = -k_B T \ln\left[\left(\frac{1-\alpha}{\alpha}\right) 10^{(pH-pK_i)}\right] \tag{B.6}$$

ou

$$\Delta U_i = \pm k_B T (pH - pK_i)\ln(10) \tag{B.7}$$

onde o sinal positivo refere-se ao processo de protonação e o sinal negativo ao processo de desprotonação. Quando na presença de outros corpos carregados, deve-se adicionar um termo referente à interação eletrostática entre a espécie que irá sofrer o processo de protonação/desprotonação e os demais corpos carregados ΔU_i^{el}. Logo, tem-se que:

$$\Delta U_i = \Delta U_i^{el} \pm k_B T (pH - pK_i)\ln(10) \tag{B.8}$$

Apêndice C

Trabalhos publicados e em andamento

Trabalhos publicados durante o período do doutorado

- A Coronado, M., P Caruso, I., M de Oliveira, V., G Contessoto, V., BP Leite, V., A Kawai, L., ... & J Eberle, R. (2017). Cold Shock Protein A from Corynebacterium pseudotuberculosis: Role of Electrostatic Forces in the Stability of the Secondary Structure. Protein and Peptide Letters, 24(4), 358-367.

- Caetano, D. L., Bossa, G. V., de Oliveira, V. M., Brown, M. A., de Carvalho, S. J., & May, S. (2016). Role of ion hydration for the differential capacitance of an electric double layer. Physical Chemistry Chemical Physics, 18(40), 27796-27807.

- Contessoto, V. G., de Oliveira, V. M., de Carvalho, S. J., Oliveira, L. C., & Leite, V. B. (2016). NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. Journal of Chemical Theory and Computation, 12(7), 3270-3277

- de Oliveira, V. M., & de Carvalho, S. J. (2014). Adsorption of pH-responsive polyelectrolyte chains onto spherical macroions. The European Physical Journal E, 37(8), 1-7

Trabalhos submetidos

- Epitope charge distribution controls the antigenicity for binding human IgG4 isotype in a modified Blo t 5, the major allergen from dust mite *Blomia tropicalis*. Artigo submetido ao *Immunobiology Journal*.

- Effects of pH and Salt Concentration on Stability of a Protein G Variant: Insights Using Coarse-Grained Models. Artigo submetido ao *Journal of Chemical Theory and Computation*.

- The Role of Ion-Specific Hydration Interactions on the Differential Capacitance of an Electrical Double Layer. Artigo submetido ao *Physical Chemistry Chemical Physics*.

Trabalhos em andamento

- Efeitos das Mutações Câncer-Ativadoras e de pH na Transição Funcional do Receptor de Estrogênio-α. Trabalho em colaboração com pesquisadores da Northeastern University (Boston-EUA). Processo Fapesp Sprint # 2014/50739-5

- Diferentes métodos de amostragem de estados de protonação aplicados no cálculo do modelo de Tanford-Kirkwood. Desenvolvimento de *Web Server* para executar cálculos de TKSA.

- Estudo das interações não-nativas no enovelamento dos domínios da α-espectrína - R15, R16 e R17. Trabalho em fase final de escrita.

- Otimização de enzimas relacionadas à segunda geração de bioetanol. Trabalho em colaboração com grupos experimentais do CTBE (Campinas-SP).

- Estudo da atividade da fosfolipase PLA2 do veneno de serpente. Trabalho em colaboração com o grupo experimental do professor Arni da UNESP (São José do Rio Preto-SP).