

Vinícius de Godoi Contessoto

Estudo de interações eletrostáticas no processo
de enovelamento e na estabilidade de proteínas
utilizando modelos simplificados

Departamento de Física

Instituto de Biociências, Letras e Ciências Exatas (IBILCE)

Universidade Estadual Paulista (UNESP)

São José do Rio Preto – SP

Agosto – 2016

Contessoto, Vinícius de Godoi.

Estudo de interações eletrostáticas no processo de enovelamento e na estabilidade de proteínas utilizando modelos simplificados / Vinícius de Godoi Contessoto. -- São José do Rio Preto, 2016

146 f. : il., 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. Enovelamento de proteínas. 4. Enzimas. 5. Bioetanol. 6. Dinâmica molecular. I. Leite, Vitor Barbanti Pereira. II. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU – 577.112

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

Vinícius de Godoi Contessoto

Estudo de interações eletrostáticas no processo de enovelamento e na estabilidade de proteínas utilizando modelos simplificados

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.

Orientador: Prof. Dr. Vitor Barbanti Pereira Leite

São José do Rio Preto - SP

2016

Vinícius de Godoi Contessoto

Estudo de interações eletrostáticas no processo de enovelamento e na estabilidade de proteínas utilizando modelos simplificados

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 do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto.

BANCA EXAMINADORA

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

Prof. Dr. Alexandre Suman de Araujo
UNESP – São José do Rio Preto - SP

Prof. Dr. Pedro Geraldo Pascutti
UFRJ – Rio de Janeiro - RJ

Prof. Dr. Antonio Caliri
USP – Ribeirão Preto - SP

Prof. Dr. Luis Gustavo Dias
USP – Ribeirão Preto - SP

São José do Rio Preto - SP

16 de agosto de 2016

Dedico esta tese à minha família

“The beauty of a living thing is not the atoms that go into it, but the way those atoms are put together.”

— Carl Sagan (1934–1996)

Agradecimentos

Agradeço primeiramente a Deus pelo dom da vida e por me permitir completar esta etapa.

Agradeço aos meus pais, Francisco Contessoto Neto e Claudete Ap^a Ferreira de Godoi Contessoto pela dedicação, incentivo e carinho.

À minha esposa Nayara Sousa de Alcântara pelo amor, amizade, carinho, compreensão e paciência.

Ao meu irmão Allan de Godoi Contessoto e minha cunhada Beatriz Curti pelo apoio.

Aos meus Avôs, Tios e Primos pelo incentivo.

Aos amigos de pós-graduação, Ícaro, Daniel, Alexandre, Josimar, Pedro, Guilherme, Miriam, Fernanda, Gabriela, Taísa, Dayane, Natália, pela ajuda constante durante esses anos.

Aos amigos do grupo: Ronaldo, Antonio, Vinícius Goiás, Gabriel, Fernando e Tiago, pela ajuda, companheirismo e discussões.

Aos professores do Departamento de Física pela qualidade de ensino proporcionada.

Aos servidores do Departamento de Física por toda ajuda prestada.

Ao programa de pós-graduação em Biofísica molecular que não me negou auxílio quando precisei e aos pesquisadores colaboradores.

À UNESP pelo acolhimento e infra-estrutura fornecida.

Ao Prof. Dr. Vítor Barbanti Pereira Leite pela orientação, amizade, paciência e ajuda no amadurecimento profissional.

À CAPES pelo auxílio financeiro.

A todos que de forma ou de outra contribuíram com esse trabalho e que esqueci de mencionar.

Sumário

Lista de Abreviaturas	iii
Lista de Símbolos	v
Lista de Figuras	ix
Lista de Tabelas	x
Resumo	xi
Abstract	xii
1 Organização da tese	1
2 Interações eletrostáticas no enovelamento	2
2.1 Introdução e motivação	2
2.2 Proteína estudada - NTL9	3
2.3 Resultados principais	4
2.4 Conclusões	8
2.5 Aplicação da metodologia - proteína CpCspA	10
3 Interações eletrostáticas na estabilidade de enzimas	12
3.1 Introdução	12
3.1.1 Enunciado do problema - Bioetanol	12
3.1.2 Objetivos e motivação	13
3.2 Materiais e métodos	14

3.2.1	Cálculos de energia eletrostática	14
3.2.2	Detalhes da aplicação do método	18
3.2.3	Enzimas estudadas	18
3.3	Resultados parciais	19
3.3.1	Xilanase A (XynA)	20
3.3.2	Xilanase A com mutações (G3_x4)	24
3.3.3	Xilanase XylB	28
3.4	Conclusões parciais	30
3.5	Aplicação da metodologia - Blo t 5	31
Referências		35
A Descrição do modelo de Tanford-Kirkwood		44
Anexo I - Artigo publicado na JCTC		47
Anexo II - Material suplementar do Anexo I		55
Anexo III - Artigo CpCspA submetido		61
Anexo IV - Artigo Blo t 5 submetido		79

Lista de Abreviaturas

B-factor Debye–Waller factor ou Temperature factor

Blot 5 *Blomia tropicalis*

C Cisteína - CIS

CM Centro de massa

CpHMD Constant-pH Molecular Dynamic Method

CsCspA Proteína Cold Shock A de *Corynebacterium pseudotuberculosis*

CTBE Laboratório Nacional de Ciência e Tecnologia do Bioetanol

CTR C-terminal

D Ácido aspártico - ASP

DSC Differential scanning calorimetry

E Ácido glutâmico - GLU

EC Enzyme Commission number

G Glicina - GLI

G3_x4 Xilanase A de *Bacillus subtilis* com as mutações Q7H, G13R, S22P e S179C

GH Glycoside Hydrolase family

H Histidina - HIS

IBILCE Instituto de Biociências, Letras e Ciências Exatas

IgE Imunoglobulina E

IgG Imunoglobulina G

JCTC Journal of Chemical Theory and Computation

NTL9 N-terminal da proteína ribossomal L9

NTR N-terminal

P Prolina - PRO

PDB Protein Data Bank

Q Glutamina - GLN

R Arginina - ARG

RMSF Root Mean Square Fluctuation

S Serina - SER

SASA Solvent-Accessible Surface Area

SBM- C_α Structure-Based Models C_α

TK Tanford-Kirkwood

TKSA Tanford-Kirkwood Solvent Accessibility

UFTM Universidade Federal do Triângulo Mineiro

UFU Universidade Federal de Uberlândia

UNESP Universidade Estadual Paulista “Júlio de Mesquita Filho”

WHAM Weighted Histogram Analysis Method

XylB Xilanase de *Bacillus pumilus*

XynA Xilanase A de *Bacillus subtilis*

Lista de Símbolos

A_{ij} Termo associado com região de baixa constante dielétrica

B_{ij} Termo associado com região de baixa e de alta constante dielétrica

C_s Concentração de sal

C_{ij} Termo associado com as interações com íons no solvente

E_{ij} Energia de interação eletrostática entre duas cargas positivas existentes no sítios de protonação dos grupos ionizáveis i e j

$F(Q)$ Energia livre em função da fração de contatos nativos

M Quantidade de grupos ionizáveis na proteína

N Número de *Avogadro*

Q Fração de contatos nativos

R Constante de Boltzmann dividida por mol

SA_{ij} Média da área acessível ao solvente entre o resíduo i e j

T Temperatura do ambiente

Z_D Função de partição de protonação para o estado desenovelado

Z_N Função de partição de protonação para o estado nativo

ΔF_{unf} Barreira de energia livre de desenovelamento

ΔG_{qq} Contribuição da interação carga-carga para a energia livre

$\Delta G_D(\chi)$ A energia livre da reação de protonação para um estado de protonação χ para o estado desenovelado

$\Delta G_N(\chi)$ A energia livre da reação de protonação para um estado de protonação χ para o estado nativo

χ Estado de protonação da proteína

κ Inverso do comprimento de Debye

$\langle W_{ij} \rangle$ Energia média da interação carga–carga entre dois grupos ionizáveis i e j

$\langle W_{q-q} \rangle$ Energia total da interação carga–carga do estado nativo

$\langle W_i \rangle$ Energia de interação carga–carga do grupo ionizável i com todos os outros grupos ionizáveis presentes na proteína

$\nu(\chi)$ Número de grupos ionizáveis protonados no estado de protonação χ

ϕ Ângulo azimute para o sistema de coordenadas esféricas

$\rho_D(\chi)$ Fração da proteína no estado desenovelado para o estado de protonação χ

$\rho_N(\chi)$ Fração da proteína no seu estado nativo para o estado de protonação χ

θ Ângulo polar para o sistema de coordenadas esféricas

e Carga elementar

k Constante de Boltzmann

$pK_{int,i}$ Valor do pK_a intrínseco para o grupo i

q_i Valor da carga elétrica para o grupo ionizável i

r Coordenada radial para o sistema de coordenadas esféricas

r_i Distância da carga i até o centro de massa

x_i Estado de protonação do resíduo i

z_j Grupo ionizável da proteína

$P_n^m(\cos \theta)$ Polinômios de *Legendre*

ϵ_p Região de baixa constante dielétrica - proteína

ϵ_s Região de alta constante dielétrica - solvente

φ_1 Potencial na região 1 do modelo de Kirkwood

φ_2 Potencial na região 2 do modelo de Kirkwood

φ_3 Potencial na região 3 do modelo de Kirkwood

$\vec{r}_j$ Posição do grupo ionizável da proteína

Lista de Figuras

1	Estrutura e sequência de aminoácidos da proteína NTL9	3
2	Curvas teóricas e experimentais da fração de desenovelamento em função da temperatura em pH 1.5 e 5.5	5
3	Perfil de energia livre $F(Q)$ para a proteína NTL9 em função da fração de contatos nativos Q	6
4	Barreira de energia livre de desenovelamento ΔF_{unf} em função do pH . . .	7
5	Curvas de calor específico para a proteína CsCspA	11
6	Representação do modelo de Tanford-Kirkwood	14
7	Estrutura da enzima Xilanase A de <i>Bacillus subtilis</i> (XynA)	19
8	Representação da energia de interação carga–carga de cada resíduo para a enzima selvagem XynA	20
9	Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima selvagem XynA e a enzima XynA com a mutação K99E	22
10	Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima selvagem XynA e a enzima XynA com a mutação D83R	23
11	Representação da energia livre de interação carga–carga de cada resíduo para a enzima G3_x4	24
12	Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima G3_x4 e a enzima G3_x4 com a mutação K99E	26
13	Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima selvagem G3_x4 e a enzima G3_x4 com a mutação D83R	27
14	Representação da energia livre de interação carga–carga de cada resíduo entre a enzima selvagem XylB e a enzima XylB com as mutações K47E e D126R	29

15	Representação da estrutura e da energia eletrostática para cada antígeno .	32
16	Correlação entre a energia eletrostática e a porcentagem de pacientes com redução na reatividade com o anticorpo IgE	34
17	Representação esquemática do modelo de Kirkwood	44

Lista de Tabelas

I	Comparação entre valores de pK_a experimentais e computacionais para proteína NTL9	8
II	Valores de pK_{int} dos grupos ionizáveis	16
III	Valores de ΔG_{qq} e de SASA para cada resíduo com contribuição desfavorável para a proteína XynA	21
IV	Valores de ΔG_{qq} e de SASA para cada resíduo com contribuição desfavorável para a proteína G3_x4	25

Resumo

Entender a contribuição das interações eletrostáticas no processo de enovelamento e na estabilidade do estado nativo de proteínas é de grande relevância em biofísica molecular. Este trabalho possui duas frentes de estudos. Na primeira etapa é apresentado o estudo das interações eletrostáticas no processo de enovelamento utilizando modelos baseados em estruturas com cargas em dinâmica molecular com pH constante. O estudo foi realizado na parte N-terminal da proteína ribossomal L9 (NTL9), uma proteína com o mecanismo de enovelamento de 2 estados, reversível e realizado desde pH 1.0 até 12.0. Foi possível comparar os resultados das simulações com os resultados experimentais presentes na literatura e os dados obtidos indicam que o modelo proposto é capaz de capturar informações fundamentais sobre o processo de enovelamento referentes à estabilidade e dependência com pH. Na segunda etapa é apresentado o estudo sobre as interações eletrostáticas na estabilidade de enzimas com interesse na produção de bioetanol de segunda geração. O objetivo final deste trabalho é adequar as enzimas às condições do reator para a produção de bioetanol. Neste trabalho foi utilizada uma metodologia capaz de indicar possíveis mutações, pela otimização da interação carga-carga na superfície da proteína, que levam ao aumento de termostabilidade. Este trabalho conta com a colaboração do grupo experimental do Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE) que realizam os testes experimentais e as validações das mutações propostas.

Palavras-chave: Mutações racionais, Otimização de enzimas, Bioetanol de segunda geração, Modelo de Tanford-Kirkwood, Método de dinâmica molecular com pH constante, Termoestabilidade.

Abstract

The understanding of electrostatic interactions in protein folding process and in native state stability is important to molecular biophysics area. This work has two areas of interest. At first step it is presented the study about electrostatic interactions in the protein folding process using structure-based models in a constant pH molecular dynamics. It was used the N-terminal domain of ribosomal protein L9 (NTL9), which folding mechanism is a two-state pathway, fully reversible and foldable in a pH range from 1.0 to 12.0. The simulation results were compared with experimental results from literature and the obtained data indicates that the proposed model is capable of capturing essential features of folding mechanism about stability and pH dependence. At second step, it is presented the study about electrostatic interactions in stability of enzymes related with second generation bioethanol production. The final goal of this work is to adjust the enzymes to reactor conditions of bioethanol production. It was employed a method that can suggest rational mutation, based on optimization of charge-charge interactions, that leads to thermostability increase. This work can count on the collaboration of an experimental group of Brazilian Bioethanol Science and Technology Laboratory (CTBE) that performs the wet lab tests and validates the suggested mutations.

Keywords: Rational mutations, Enzyme optimization, Second generation bioethanol, Tanford-Kirkwood model, Constant-pH molecular dynamic method, Thermostability.

1 Organização da tese

Esta tese procura entender e discutir a contribuição das interações eletrostáticas ao longo processo de enovelamento e na estabilidade do estado nativo de proteínas. Na seção 2 será apresentado o trabalho sobre as interações eletrostáticas no enovelamento de proteínas. O trabalho desta primeira etapa está recém publicado e encontra-se no Anexo I - Artigo publicado na JCTC. A aplicação da metodologia desenvolvida neste trabalho foi aplicada em outros estudos e os resultados obtidos são discutidos na seção 2.5. Na seção 3 é apresentado o estudo sobre as interações eletrostáticas na estabilidade de enzimas com interesse na produção de bioetanol de segunda geração. Este é um estudo em andamento, possui algumas conclusões parciais referente a parte teórica realizada e conta com a colaboração do grupo experimental do CTBE para os testes em laboratório. A metodologia para o cálculo das interações eletrostáticas também foi utilizada em estudos sobre a caracterização de epítomos alergênicos e os resultados desta etapa estão descritos na seção 3.5.

2 Interações eletrostáticas no enovelamento

Nesta seção será apresentado o estudo das interações eletrostática no processo de enovelamento utilizando modelos baseados em estruturas com cargas em dinâmica molecular com pH constante. Este é um trabalho recém publicado [1], desta forma, será realizada uma breve introdução ao assunto e serão apresentados os principais resultados deste trabalho. Os detalhes sobre os modelos utilizados e as discussões de outros resultados encontram-se presentes no Anexo I - Artigo publicado na JCTC.

2.1 Introdução e motivação

Entender a contribuição das interações eletrostáticas ao longo do processo de enovelamento é de grande importância para a compreensão desse mecanismo [2–4]. A distribuição de cargas em uma proteína é determinada pelos tipos de aminoácidos com grupos ionizáveis presentes na sua composição e pelas condições do ambiente em que ela se encontra. Os estados de protonação dos grupos ionizáveis são determinados pelas distâncias entre os grupos carregados, pela concentração de sal e pelo valor de pH do meio [5, 6].

Os estudos computacionais em enovelamento que abordam o uso de interações eletrostáticas, geralmente, utilizam valores de cargas fixas para os grupos ionizáveis ao longo de toda a simulação [7, 8]. Neste trabalho, os efeitos associados com a variação de pH em proteínas são estudados computacionalmente por simulações de dinâmica molecular com pH constante - *Constant-pH Molecular Dynamic Method (CpHMD)* [9]. Nas simulações CpHMD, um resíduo de aminoácido pode ser protonado ou desprotonado baseado em seu valor de pK_a e no valor do pH do ambiente. As simulações CpHMD foram implementadas no modelo baseado em estrutura - *Structure-Based Models C_α (SBM- C_α)* [10–15] com adição do potencial de *Debye-Huckel* [16, 17] e com titulações dos grupos ionizáveis.

2.2 Proteína estudada - NTL9

A proteína escolhida neste estudo foi a parte N-terminal da proteína ribossomal L9 (NTL9) [18]. A proteína NTL9 possui 56 resíduos de aminoácidos formando 3 fitas β entre 2 hélices- α [19] (Figura 1). Seu enovelamento acontece por um mecanismo de 2 estados [20], é completamente reversível e pode ser realizado desde pH 1.0 até 12.0 [21]. Estas características apresentadas sugerem a proteína NTL9 como uma boa candidata ao estudo de enovelamento com dependência do pH, sendo possível comparar os resultados computacionais da metodologia utilizada com resultados experimentais [21].

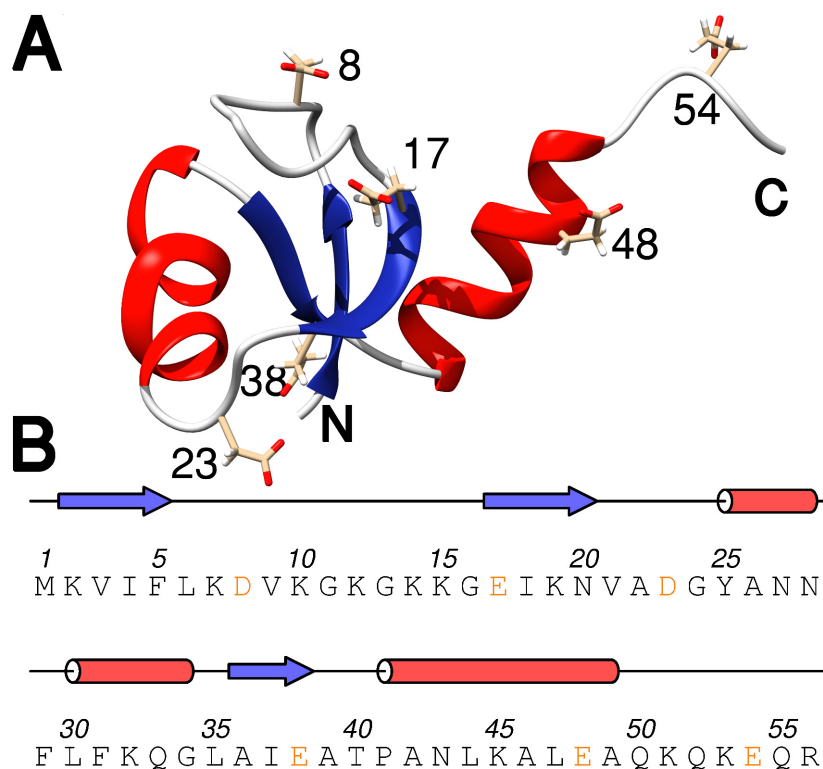


Figura 1: Figura retirada do artigo publicado na JCTC [1]. **A** - Estrutura da proteína NTL9, com PDB 1CQU e em destaque os resíduos ácidos. **B** - sequência de aminoácidos da proteína NTL9 com destaque para os resíduos ácidos. A escolha para destacar os resíduos ácidos é motivada pelos valores de pH utilizados nas simulações e no valores utilizados nos trabalhos experimentais que ocorrem em pH 1.5 e 5.5. Nestes valores de pH, os resíduos ácidos possuem a maior variação de carga.

2.3 Resultados principais

Os principais resultados apresentados são propriedades obtidas das simulações e que são comparadas com resultados experimentais [21]. As simulações foram realizadas em diferentes temperaturas para valores de pH entre 1.5 e 12.0. As análises termodinâmicas foram obtidas pela combinação das diferentes simulações utilizando o método dos múltiplos histogramas - *Weighted Histogram Analysis Method (WHAM)*.

Uma das primeiras análises obtidas das trajetórias das simulações é a fração de desenovelamento da proteína (Figura 2). Os dados são obtidos para diferentes simulações realizadas em diferentes temperaturas. Pela análise da figura 2 é possível identificar as temperaturas em que a proteína está enovelada (temperaturas baixas), as temperaturas em que a proteína se desenovela (temperaturas altas) e a temperatura de transição onde a fração de desenovelamento é de 0.5. O comportamento da curva de fração de desenovelamento é semelhante à função sigmóide para os resultados teóricos e experimentais obtidos para valores de pH 1.5 e 5.5. Tanto no resultado teórico, quanto no resultado experimental, as curvas referentes ao pH 5.5 estão deslocadas a direita em relação ao pH 1.5, indicando um aumento da termostabilidade da proteína com essa mudança de pH.

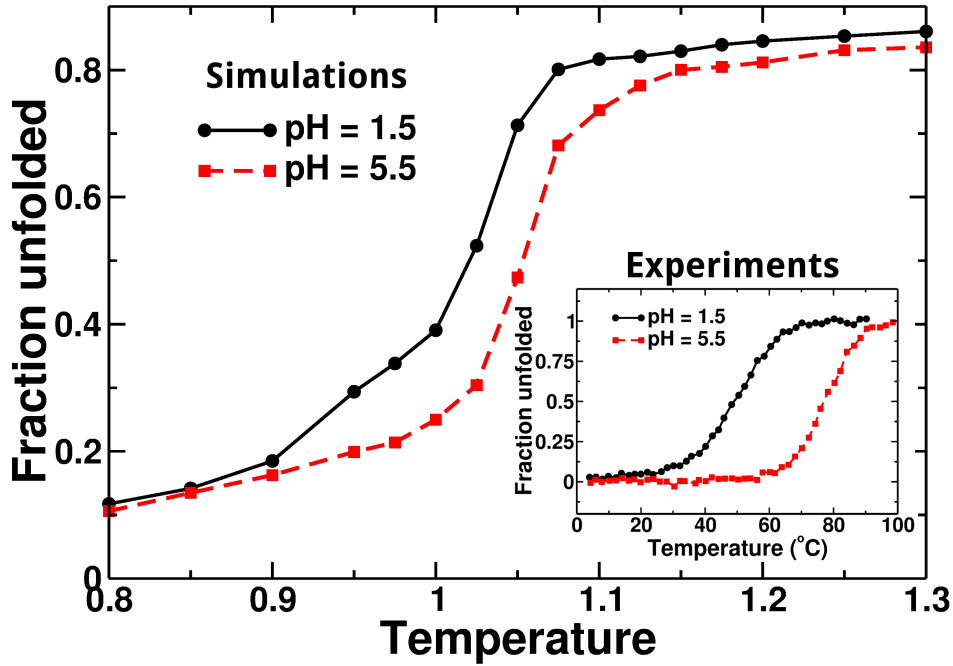


Figura 2: Fração de desenovelamento em função da temperatura em unidades reduzidas para dois valores de pH diferentes. A linha contínua com círculos pretos representa a curva para o pH 1.5 e a linha tracejada com quadrados vermelhos representa a curva para o pH 5.5. Inserido está o resultado experimental - adaptado de Luisi, D. L and Raleigh, D [21]. Figura retirada do artigo recém publicado [1].

Para uma melhor descrição sobre o aspecto da estabilidade da proteína NTL9 para diferentes pH é analisado o perfil de energia livre $F(Q)$ em função da fração de contatos nativos Q calculado pelo WHAM (Figura 3). A figura 3 destaca a barreira de energia livre de desenovelamento ΔF_{unf} para os valores de pH 1.5 e 5.5. O valor de ΔF_{unf} é obtido pela diferença do valor de energia livre entre o estado nativo ($Q \approx 0.75$) e o estado de transição da proteína ($Q \approx 0.4$) para os dois valores de pH simulados. É possível verificar a estabilização do estado nativo da proteína NTL9 para o valor de pH 5.5 em comparação ao pH 1.5. Esta estabilização está relacionada com a desprotonação dos resíduos ácidos em pH 5.5 contribuindo para a formação de interações menos desfavoráveis e diminuindo o valor da energia eletrostática total [1].

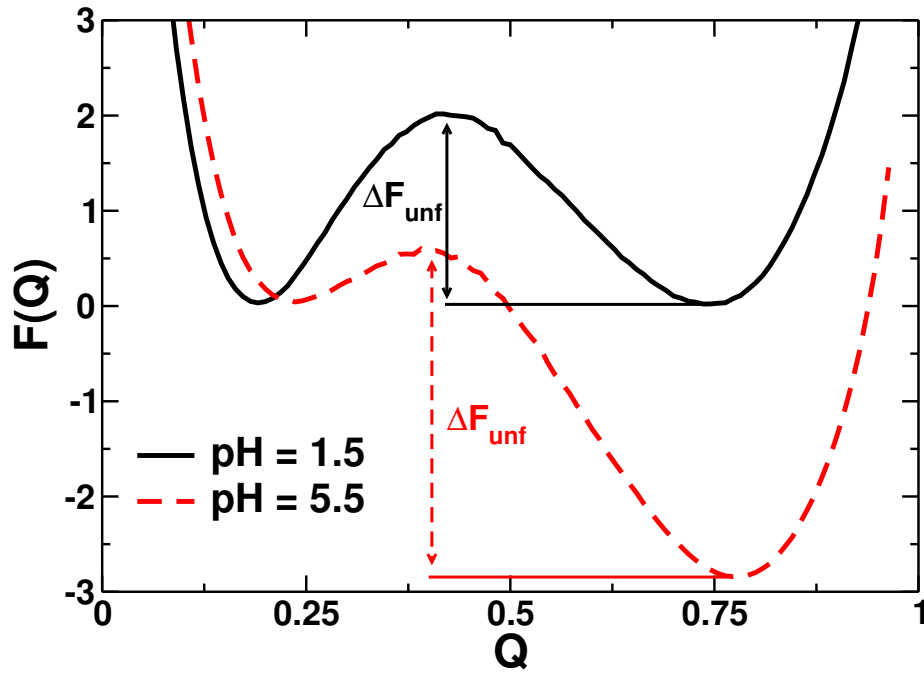


Figura 3: Perfil de energia livre $F(Q)$ para a proteína NTL9 em função da fração de contatos nativos Q . A linha contínua preta representa a curva para o valor de pH 1.5. A linha tracejada vermelha representa a curva para o valor de pH 5.5. ΔF_{unf} é a barreira de energia livre de desenovelamento [21]. Figura retirada do material suplementar artigo recém publicado [1].

A influência na variação dos valores de pH na estabilidade do estado nativo da proteína NTL9 pode ser verificada na figura 4. Na figura 4 os valores de ΔF_{unf} são calculados em diferentes condições de pH. É possível comparar os resultados computacionais com os resultados experimentais (gráfico inserido - figura 4). As curvas possuem o formato tipo sino e o pico de estabilidade nas duas curvas estão com valores próximos de pH entre 6.0 e 7.0. Vale ressaltar que esta condição de maior estabilidade do estado nativo representa o menor valor da energia eletrostática e que esses valores de pH não correspondem ao ponto isoelétrico da proteína que é de pH 10.5 [1, 21].

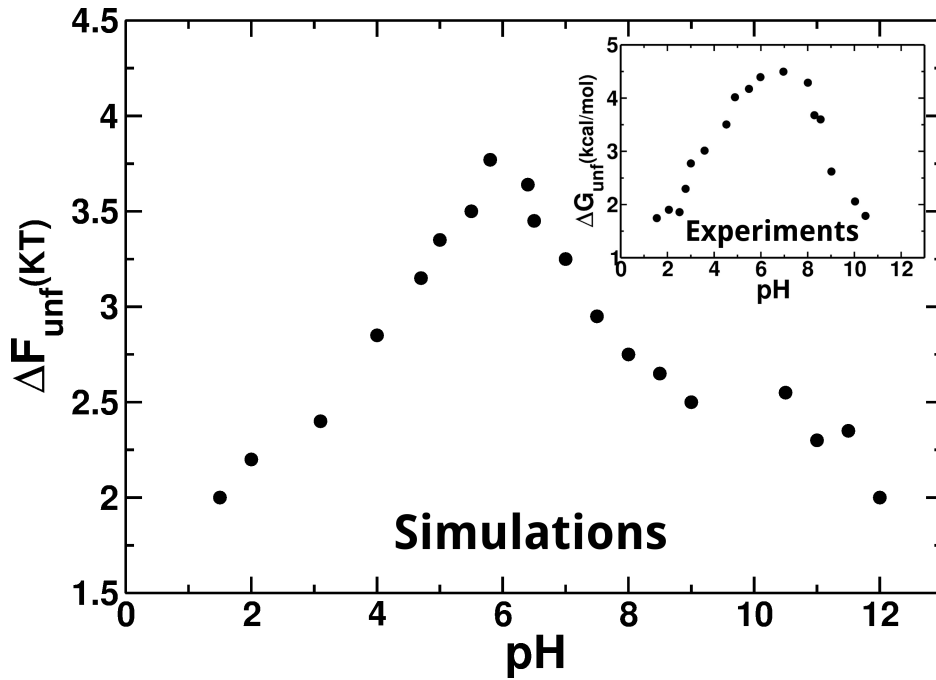


Figura 4: Barreira de energia livre de desenovelamento ΔF_{unf} em função do pH. O gráfico inserido é referente aos resultados experimentais de Luisi, D. L and Raleigh, D [21]. Figura retirada do artigo recém publicado [1].

Outra característica relevante obtida no estudo é o cálculo dos valores de pK_a dos grupos ionizáveis. Esta análise indica o quão bem o modelo é capaz de amostrar as variações de cargas para o sistema estudado e de determinar o grau de ionização de cada resíduo de aminoácido. Os valores são obtidos pelo ajuste da equação de Henderson–Hasselbalch nas curvas com o grau de ionização de cada resíduo [22]. As curvas de titulação do resíduos ácidos estão apresentados no Anexo II - Material suplementar do Anexo I. Os valores obtidos pelas simulações estão próximos aos valores obtidos pelos experimentos [23], divergindo em mais de 1 unidade de pK_a apenas para o resíduo *ASP* – 8 - Tabela I. Esta diferença entre os valores de pK_a computacional e experimental pode ser explicada pelas interações realizadas entre cadeia lateral deste resíduo (ausente neste modelo) e outros grupos ionizáveis da proteína [24].

Tabela I: Comparação entre valores de pK_a experimentais [23] e computacionais para proteína NTL9.

Resíduo	$C_s = 100 \text{ mM}$	
	Exp pK_a	Comp pK_a
<i>ASP</i> – 8	3.0	1.9
<i>GLU</i> – 17	3.6	3.0
<i>ASP</i> – 23	3.1	3.2
<i>GLU</i> – 38	4.0	3.6
<i>GLU</i> – 48	4.2	3.5
<i>GLU</i> – 54	4.2	3.4

2.4 Conclusões

Este estudo procurou entender a influência das interações eletrostáticas no processo de enovelamento de proteínas utilizando modelos simplificados. As metodologias e os modelos empregados neste estudo são conhecidos e bem estabelecidos pela comunidade científica quando utilizados de maneira independente. A inovação foi a utilizar esses métodos de maneira complementar para que as simulações tenham informações sobre os estados enovelados e desenovelados da proteína e que as interações eletrostáticas e mudanças de cargas sejam mapeadas em cada pH estudado.

A metodologia apresentada mostrou ser capaz de capturar de maneira satisfatória as características sobre o processo de enovelamento da proteína NTL9. As curvas de fração de desenovelamento (Figura 2) e de barreira de energia livre de desenovelamento (Figura 4) apresentaram uma boa concordância com resultados experimentais, indicando que foi possível amostrar os estados enovelados e desenovelados e obter informações das interações eletrostáticas nestes estados de maneira satisfatória.

Um aspecto relevante quanto à utilização do modelo é a ausência das cadeias laterais dos resíduos de aminoácidos e a aproximação de colocar as cargas na posição do carbono α da cadeia da proteína. Esta característica possui influência direta nos cálculos de valores de pK_a dos grupos ionizáveis das proteínas. Os cálculos de valores de pK_a para a proteína NTL9 estão próximos aos valores experimentais (Tabela I) divergindo apenas para o resíduo $ASP - 8$. Outras proteínas também foram utilizadas nos cálculos de valores de pK_a e os resultados e discussões sobre a importância das cadeias laterais estão no Anexo II - Material suplementar do Anexo I

2.5 Aplicação da metodologia - proteína CpCspA

A metodologia foi empregada em um estudo sobre a proteína cold shock A de *Corynebacterium pseudotuberculosis* (CsCspA). Este é um trabalho recém submetido em colaboração com o grupo experimental do Prof. Dr. Arni (UNESP-IBILCE). O estudo aborda as questões sobre a termoestabilidade da proteína CsCspA em função do pH. O principal resultado apresentado será a comparação entre os resultados teóricos e experimentais. Maiores detalhes sobre o trabalho está presente no Anexo III - Artigo CpCspA submetido.

Foram realizados experimentos de calorimetria de varredura diferencial - *Differential scanning calorimetry* (DSC) na proteína CsCspA, obtendo as curvas de calor específico para diferentes valores de pH entre 6.0 e 8.0 (Figura 5). As curvas de calor específico possuem um perfil tipo sino e o valor do seu pico representa a temperatura de *melting* ou temperatura de transição. Os resultados experimentais indicam o maior valor da temperatura de transição para o valor de pH 6.0 em relação aos demais (Figura 5B). As curvas de calor específico também são obtidas teoricamente com a análise do WHAM nas trajetórias simuladas em diferentes valores de pH (Figura 5A). Os resultados obtidos pelas simulações também indicam o maior valor na temperatura de transição para o valor de pH 6.0 em relação aos demais. Este resultado teórico está em acordo com os resultados experimentais, reforçando a característica de maior termoestabilidade da proteína CsCspA em pH 6.0.

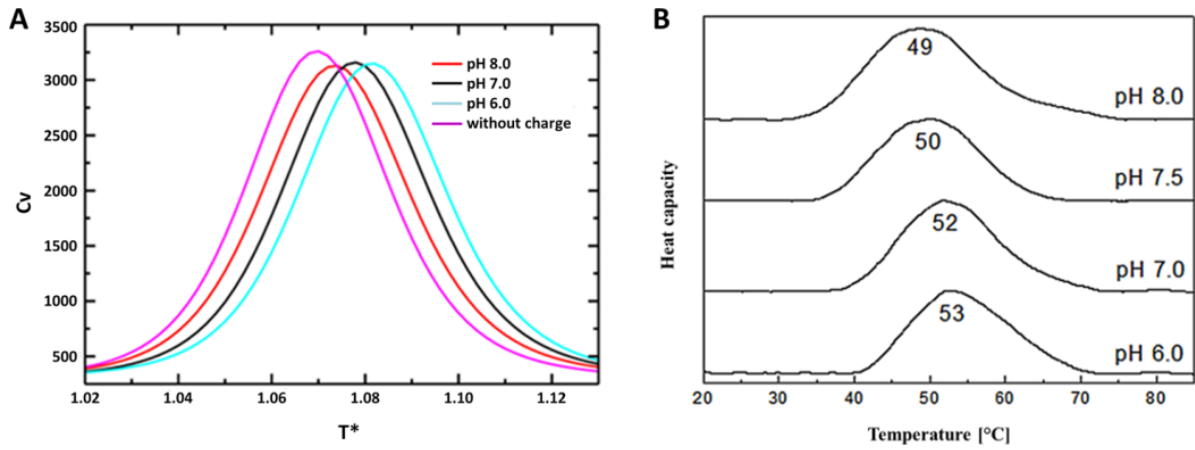


Figura 5: Curvas de calor específico para a proteína CsCspA. **A** – Curvas, obtidas pela análise do WHAM para diferentes valores de pH. Os valores da temperatura no eixo horizontal estão em unidades reduzidas. Também foram realizadas simulações para o modelo sem a utilização de cargas com o intuito de explorar a dependência com o pH na temperatura de transição - Adaptada da figura 4 do artigo submetido [25]. **B** – Curvas obtidas pelo experimento de DSC para diferentes valores de pH entre 6.0 e 8.0 - Adaptada da figura 3 do artigo submetido [25].

3 Interações eletrostáticas na estabilidade de enzimas

Nesta seção será apresentado o estudo das interações eletrostática na estabilidade de enzimas com interesse na produção de bioetanol de segunda geração. O estudo procura entender a contribuição para a energia eletrostática de cada resíduo com cadeia lateral carregada na estabilização do estado nativo da enzima e propor mutações nos resíduos que atuam de maneira desfavorável. Este é um trabalho em andamento e conta com a colaboração do grupo experimental do Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE) que realiza os ensaios experimentais de mutações e teste de atividade catalítica. Serão apresentados os resultados referente a parte teórica, os resultados experimentais conclusivos estão em andamento.

3.1 Introdução

3.1.1 Enunciado do problema - Bioetanol

A procura por fontes de energia limpas e renováveis está entre os grandes desafios para a ciência atual. A utilização de biocombustíveis como fonte de energia é vista como essencial na substituição dos combustíveis fósseis [26,27]. O Brasil produz o etanol proveniente da fermentação do caldo da cana-de-açúcar aparecendo como um dos maiores produtores de bioetanol de primeira geração do mundo [28,29]. Diante da perspectiva de aumento da demanda de etanol, há uma necessidade de encontrar meios que visam aumentar sua produção e que não gere um aumento na área plantada de cana-de-açúcar.

Um possível caminho, para este aumento de produtividade, está na utilização da porção da biomassa da cana-de-açúcar que não é utilizada diretamente no processo de fermentação. Esta biomassa é composta basicamente de celulose, hemicelulose e lignina [30,31] e corresponde, aproximadamente, a 2/3 da massa total da cana-de-açúcar [32]. A produção do bioetanol utilizando esta parte da biomassa constitui em transformar a celulose em unidades de carboidratos menores e solúveis. Este processo de transformação da celulose é realizado utilizando algumas enzimas específicas que catalisam o processo de hidrólise [33–36]. Esta nova abordagem caracteriza a segunda geração de produção de

bioetanol [37].

O desafio tecnológico atual está em aperfeiçoar este processo de hidrólise enzimática. A meta é tornar o processo economicamente viável e que seja realizado em escala industrial [38, 39]. Este avanço envolve a obtenção de enzimas específicas para realizar o processo de hidrólise com maior eficiência [40, 41]. A caracterização biofísica das enzimas e a compreensão dos detalhes moleculares associados à hidrólise são fundamentais para o avanço nesta área. A partir dessas abordagens, espera-se alcançar meios para a criação de métodos eficientes na geração de enzimas otimizadas.

3.1.2 Objetivos e motivação

A motivação deste estudo é trabalhar na otimização *in silico* da estabilidade de enzimas, propondo mutações sítio dirigidas que serão testadas experimentalmente. A aplicação de modelos teóricos-computacionais visam a avaliação e otimização de coquetéis enzimáticos utilizados na produção de bioetanol de segunda geração. A otimização tem como objetivo o aumento de atividade catalítica, o aumento de termoestabilidade e o controle de pH, adequando as enzimas para que seu funcionamento ótimo esteja de acordo com as condições dos reatores. A temperatura de funcionamento dos reatores é de aproximadamente 50°C e o valor de pH em torno de 5.0. Este esforço conta com a colaboração de grupos experimentais do CTBE, que realiza os ensaios enzimáticos, fornece alvos de interesse e *feedback* experimental.

A metodologia empregada neste trabalho, introduzido por G. Makhatadze e colaboradores, é baseada na otimização da interação carga-carga dos resíduos de aminoácidos e apresenta um elevado índice de acerto na predição de mutações que aumentam a termoestabilidade em proteínas globulares [42–44]. Este método realiza o cálculo da energia eletrostática entre os grupos ionizáveis da proteína, com base no pH escolhido [45, 46], e indica a contribuição energética de cada resíduo para a estabilização da proteína. O método sugere mutações sítio dirigidas em resíduos que contribuem de maneira desfavorável energeticamente [4, 47, 48]. Com a aplicação deste método é possível entender a influência das mutações envolvendo resíduos polares carregados e sugerir outras mutações visando o aumento de termoestabilidade.

Resultados experimentais obtidos pelo grupo experimental CTBE em estudos na enzima Xilanase A de *Bacillus subtilis* (XynA) apresentam uma série de mutações que aumentam a atividade catalítica e a termoestabilidade desta enzima em pH 6.0 [49, 50]. Este trabalho também procura entender, do ponto de vista físico, o efeito de cada mutação realizada na enzima para o aumento da termostabilidade no determinado pH, e sugerir novas mutações baseadas na otimização da interação carga–carga. Este trabalho também estuda a enzima Xilanase de *Bacillus pumilus* (XylB) utilizando a mesma abordagem.

3.2 Materiais e métodos

3.2.1 Cálculos de energia eletrostática

Os cálculos da interação carga–carga dos resíduos de aminoácidos serão realizados utilizando o modelo de Tanford-Kirkwood (TK) [51]. Neste formalismo a proteína é modelada como uma cavidade esférica de baixa constante dielétrica ϵ_p inserida em uma solução eletrolítica de constante dielétrica ϵ_s - Figura 6.

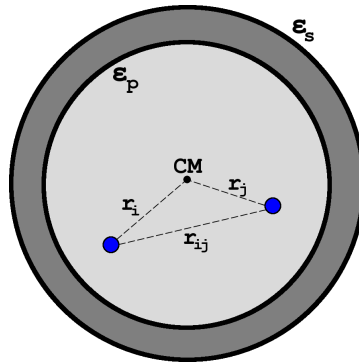


Figura 6: Representação do modelo de Tanford-Kirkwood. CM indica o centro de massa da proteína, r_i e r_j são as distâncias das cargas i e j até o centro de massa respectivamente, ϵ_p é constante dielétrica da proteína e ϵ_s é a constante dielétrica da solução.

O modelo *TK* é considerado um ótimo modelo para representar o cálculo da energia eletrostática das cargas em proteínas globulares. Para uma melhor aproximação, os valores da energia eletrostática encontrados para o modelo são corrigidos pela acessibilidade ao solvente de cada resíduo polar carregado da proteína [52, 53].

A energia de interação E_{ij} para duas cargas existentes nos sítios de protonação entre os grupos ionizáveis i e j é calculada como:

$$E_{ij} = e^2 \left(\frac{A_{ij} - B_{ij}}{2b} - \frac{C_{ij}}{2a} \right) (1 - SA_{ij}) \quad (1)$$

onde e é o valor da carga elementar, b é o valor do raio da proteína, a é o valor do raio da esfera de exclusão de íons ($a = b + 1.4\text{\AA}$). A grandeza SA_{ij} é a média da área acessível ao solvente - Solvent-Accessible Surface Area (SASA) dos resíduos i e j [54]. Os parâmetros A_{ij} , B_{ij} e C_{ij} são obtidos pela solução analítica do modelo de Tanford e Kirkwood [51, 55]. Os parâmetros são funções dependentes das posições das cargas, das constantes dielétricas da proteína e do solvente, e no caso de C_{ij} , da força iônica [56]. Os detalhes da solução do modelo e determinação dos parâmetros estão descritos no Apêndice A - Descrição do modelo de Tanford-Kirkwood.

Após a determinação da energia eletrostática entre os grupos ionizáveis da proteína é necessário realizar a amostragem entre os possíveis estados de protonação de cada resíduo com base no pH escolhido para o cálculo. No cálculo dos estados de protonação é utilizado a abordagem descrita por Bashford e Karplus [57]. Os detalhes do método para amostragem dos estados e as equações utilizadas foram descritas por B. Ibarra-Molero e G. Makhatadze [42].

Para uma proteína com n grupos ionizáveis, existem 2^n estados de protonação diferentes. Estes estados serão representados por um vetor χ de 2^n elementos x_i . Cada elemento x_i deste vetor χ pode assumir os valores 1 ou 0, dependendo se o resíduo está protonado ou não. A energia livre da reação de protonação $\Delta G_N(\chi)$ pode ser calculada para a proteína em seu estado nativo e para um estado de protonação χ de acordo com a equação abaixo:

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

onde q_i é o valor da carga do grupo i no seu estado desprotonado, R é a constante de Boltzmann dividida por mol e $pK_{int,i}$ é o valor do pK_a intrínseco para grupo i [58] - Tabela II.

Tabela II: Valores de pK_{int} dos grupos ionizáveis.

Resíduo	pK_{int}
CTR	3.6
ASP	4.0
GLU	4.5
HIS	6.3
NTR	7.5
LYS	10.6
ARG	12.0

A fração da proteína $\rho_N(\chi)$ no seu estado nativo para um determinado estado de protonação é dada por:

$$\rho_N(\chi) = \frac{1}{Z_N} \exp \left[-\frac{\Delta G_N(\chi)}{RT} - \nu(\chi) (\ln 10) pH \right] \quad (3)$$

onde $\nu(\chi)$ é o número de grupos ionizáveis protonados no estado de protonação χ e Z_N é a função de partição de protonação para a proteína no seu estado nativo descrita por:

$$Z_N = \sum_{\chi} \exp \left[-\frac{\Delta G_N(\chi)}{RT} - \nu(\chi) (\ln 10) pH \right] \quad (4)$$

A energia total média $\langle W_{q-q} \rangle$ da interação carga–carga do estado nativo é resultado da soma sobre todos os estados de protonação com o peso da fração $\rho_N(\chi)$:

$$\langle W_{q-q} \rangle = \sum_{\chi} \left[\frac{1}{2} \sum_{i,j=1}^n E_{ij} (q_i + x_i) (q_j + x_j) \right] \rho_N(\chi) \quad (5)$$

A equação 5 também pode ser entendida como a energia média $\langle W_{ij} \rangle$ da interação carga–carga entre dois grupos ionizáveis i e j :

$$\langle W_{q-q} \rangle = \frac{1}{2} \sum_{i,j=1}^n \langle W_{ij} \rangle \quad (6)$$

As equações 2, 3 e 4 descrevem as grandezas para o estado nativo da proteína. A contribuição da energia livre da reação de protonação para o estado desenovelado $\Delta G_D(\chi)$ não possui os termos de interação carga–carga e é descrita por:

$$\Delta G_D(\chi) = -RT (\ln 10) \sum_{i=1}^n (q_i + x_i) pK_{int,i} \quad (7)$$

A equação 7 é similar à equação 2 e as grandezas como a fração da proteína no estado desenovelado em um dado estado de protonação $\rho_D(\chi)$ e a função de partição de protonação no estado desenovelado Z_D podem ser calculadas substituindo o termo $\Delta G_N(\chi)$ por $\Delta G_D(\chi)$. Desta forma, a contribuição da interação carga–carga para a energia livre ΔG_{qq} [59] é dada por:

$$\Delta G_{qq} = -RT \ln \left(\frac{Z_D}{Z_N} \right) \quad (8)$$

O trabalho de B. Ibarra–Molero e G. Makhatadze [42] apresentou que a contribuição da interação carga–carga para a energia livre ΔG_{qq} pode ser descrito como, aproximadamente, o negativo da energia total média da interação carga–carga do estado nativo:

$$\Delta G_{qq} \approx -\langle W_{qq} \rangle = \frac{1}{2} \sum_{i=1}^n \left(\sum_{j=1}^n \langle W_{ij} \rangle \right) = -\frac{1}{2} \sum_{i=1}^n \langle W_i \rangle \quad (9)$$

Esta aproximação permite interpretar os valores de ΔG_{qq} em termos da contribuição individual de cada grupo ionizável apresentada no último termo da equação 9. A grandeza $\langle W_i \rangle$ é a energia de interação carga–carga do grupo ionizável i com todos os outros grupos ionizáveis presentes na proteína. A metodologia observa o valor de $\langle W_i \rangle$ para todos os grupos ionizáveis da proteína, identificando os resíduos que contribuem de maneira desfavorável para a energia e sugerindo mutações sítio dirigidas nestes grupos [60, 61]. As mutações são sugeridas no sentido de remover ou inverter a carga na posição do resíduos que contribui de maneira desfavorável e é esperado que as mutações aumentem a termoestabilidade da proteína [62, 63]. A aplicação desta metodologia se torna relevante no estudo de otimização de enzimas, uma vez que é possível prever as mutações com base nas condições escolhidas de pH e de temperatura.

3.2.2 Detalhes da aplicação do método

A aplicação do método para sugerir uma possível mutação para aumento de termoestabilidade não envolve apenas fatores energéticos para a escolha do resíduo. O resíduo escolhido deve contribuir de maneira desfavorável energeticamente e estar com a cadeia lateral mais de 50% exposta ao solvente. A mutação deve ser realizada distante em pelo menos 10Å em relação ao sítio catalítico para não atrapalhar atividade da proteína e que a mutação não modifique de maneira significativa a estrutura da proteína [64]. Uma futura ampliação do método está em propor mutações com inserções de resíduos polares carregados em posições otimizadas que serão previstas pelo modelo. Outra questão relevante envolve a limitação do número de grupos ionizáveis presentes na proteína alvo. O número de possíveis estados de protonação cresce de maneira exponencial com o número de grupos ionizáveis, tornando-se inviável seu cálculo [65].

3.2.3 Enzimas estudadas

Xilanase A (XynA) - GH11: A enzima Xilanase A de *Bacillus subtilis* (XynA) faz parte da família GH11 e realiza a função endo- β -1,4-xilanase (EC 3.2.1.8). Esta enzima possui 185 resíduos de aminoácidos totais, sendo 23 resíduos polares carregados (Figura 7). O peso molecular é de, aproximadamente, 20.5 kDa. O valor de pH ótimo é de 6.0 e sua temperatura ótima é 50°C [49, 50]. Resultados recentes indicam um aumento na atividade catalítica da enzima XynA em mutantes que tiveram a introdução de resíduos polares carregados na superfície [66]. Também foi estudada a enzima XynA com as quatro mutações realizadas pelo grupo experimental do CTBE - G3_x4 (Q7H, G13R, S22P e S179C) [49].

Endo- β -xilanase (XylB) - GH11: A enzima Xilanase de *Bacillus pumilus* (XylB) pertence a família GH11 e realiza a função endo- β -1,4-xilanase (EC 3.2.1.8) semelhante a enzima XynA. Esta enzima possui 213 resíduos totais, sendo 35 polares carregados e seu peso molecular é de, aproximadamente, 23.0 kDa. A enzima XylB não possui estrutura depositada no PDB, desta forma, foi realizada uma modelagem estrutural [67], com base na sequência primária cedida pelo CTBE, para se aplicar a metodologia TKSA.

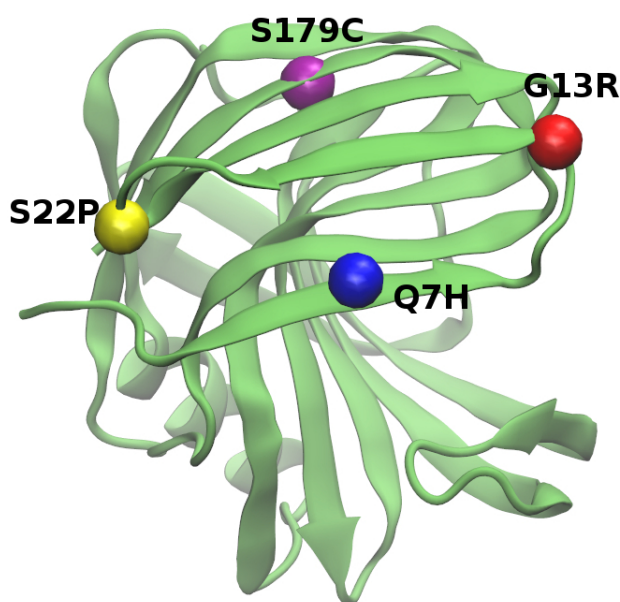


Figura 7: Estrutura da enzima Xilanase A de *Bacillus subtilis* (XynA) da família GH11 - PDB 1XXN. Em destaque os resíduos mutados no trabalho inicial do grupo experimental do CTBE [49]

3.3 Resultados parciais

Nesta seção serão apresentados os resultados teóricos obtidos pela aplicação do modelo TKSA para as enzimas estudadas. Os resultados não são conclusivos, pois os experimentos de validação experimental ainda estão em andamento no CTBE. Os valores da contribuição energética de cada resíduo serão apresentados em um gráfico de barras para cada proteína estudada. Em cada gráfico, as barras de erro são resultados de variações conformacionais realizadas na proteína. As variações das posições dos grupos ionizáveis são realizadas com base nos valores de *Root Mean Square Fluctuation (RMSF)* de cada grupo, convertidos do valor de *B-factor* da estrutura cristalográfica [68]. O valor de pH utilizado nas simulações foi de 6.0 e a temperatura simulada foi de 325K. A notação adotada será a mesma utilizada em comum na literatura [4, 42, 44, 45, 48, 60, 61, 65], ΔG_{qq} no eixo vertical em que cada barra será o valor de $-\langle W_i \rangle$ do resíduo i .

3.3.1 Xilanase A (XynA)

O resultado para os valores de energia eletrostática de cada resíduo da enzima XynA estão apresentados na figura 8. Para os 23 resíduos observados, 12 apresentaram contribuições favoráveis significativas com valores de $\Delta G_{qq} < 0$, 4 apresentaram contribuições, aproximadamente, nulas com valores de $\Delta G_{qq} \approx 0$ e 7 apresentaram contribuições desfavoráveis com valores de $\Delta G_{qq} > 0$.

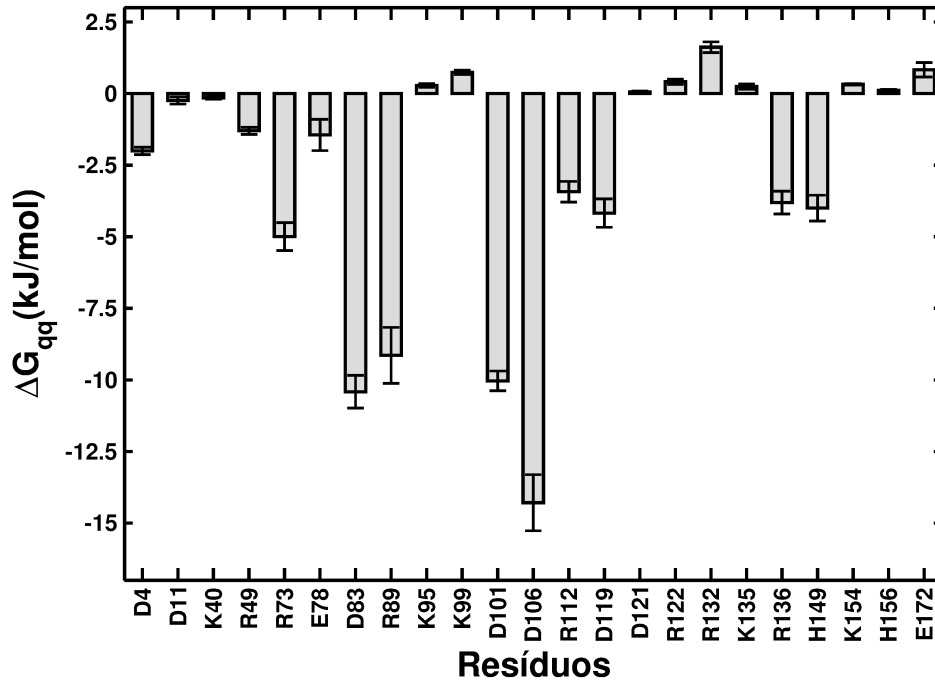


Figura 8: Representação da energia de interação carga–carga de cada resíduo para a enzima selvagem XynA. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos.

De acordo com a metodologia, o foco será dado nos 7 resíduos com contribuições desfavoráveis (K95, K99, R122, R132, K135, K154 e E172). Os valores de ΔG_{qq} e de SASA de cada resíduo estão descritos na tabela III. Os valores de SASA são normalizados pelo valor experimental da cadeia lateral do resíduo inserido no meio do tri-peptídeo Gli–X–Gli [69].

Tabela III: Valores de ΔG_{qq} e de SASA para cada resíduo com contribuição desfavorável para a proteína XynA.

Resíduo	$\Delta G_{qq}(\text{kJ/mol})$	SASA
K95	0.281	0.401
K99	0.744	0.752
R122	0.415	0.733
R132	1.624	0.092
K135	0.243	0.462
K154	0.321	0.721
E172	0.834	0.006

A tabela III destaca 3 resíduos que contribuem de maneira desfavorável para energia eletrostática e que estão com a cadeia lateral exposta ao solvente em mais de 50% (K99, R122 e K154). De acordo com metodologia, esses 3 resíduos são possíveis candidatos para serem substituídos na enzima e que devem levar ao aumento de termoestabilidade. Cada uma das três mutações foram realizadas de maneira independente e substituídas por um Ácido glutâmico - GLU (E) (K99E, R122E e K154E). Essas mutações são sugeridas com o intuito de inserir uma carga de valor contrário na posição desfavorável.

É apresentada na figura 9 os novos valores da energia eletrostática para a enzima XynA com a mutação K99E. A figura 9 mostra a sobreposição dos valores de energia em um gráfico de barras com intuito de comparar as variações de energia de cada grupos ionizável com o efeito da mutação. As mutações R122E e K154E não serão apresentadas, pois não mostraram variações significativas nos valores de energia.

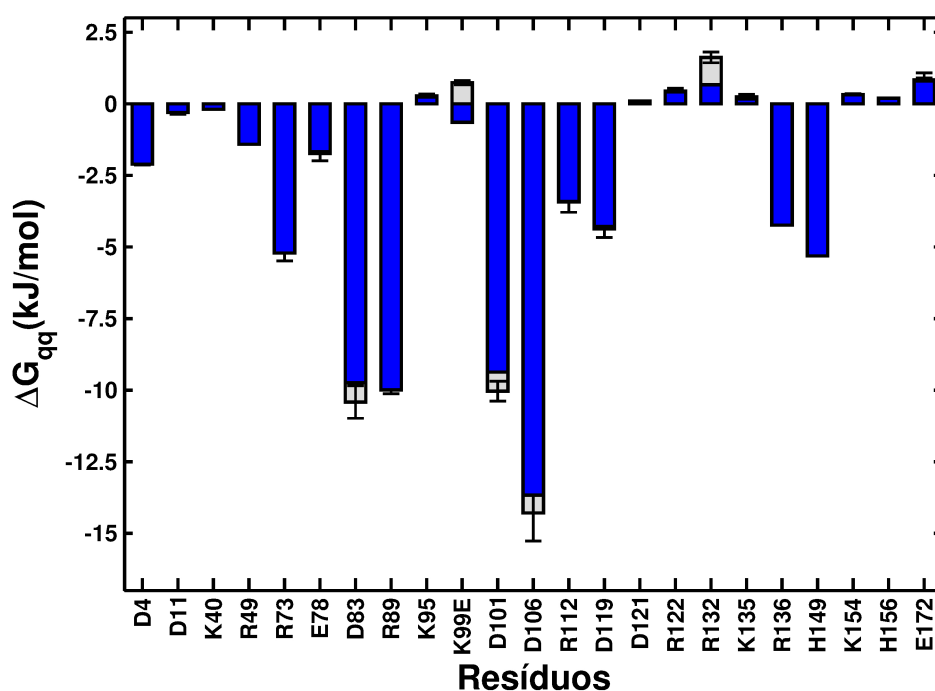


Figura 9: Comparação da energia livre de interação carga-carga de cada resíduo entre a enzima selvagem XynA e a enzima XynA com a mutação K99E. As barras azuis indicam os valores de energia de cada resíduo para a enzima com a mutação K99E. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos com destaque para a mutação K99E.

A análise da figura 9 mostra a diferença na estabilidade dos grupos ionizáveis devido a mutação K99E. A sobreposição das barras indica menores valores de energia para a enzima devido à mutação K99E e esta estabilização pode ser observada na maioria dos resíduos. É esperado que a mutação K99E promova uma termoestabilidade maior que a enzima selvagem XynA. Os testes experimentais de verificação do aumento da temperatura de transição e do aumento de atividade enzimática estão em andamento no CTBE.

Para complementar a validação da metodologia, foi sugerida uma mutação para desestabilizar a enzima XynA de maneira energética. O objetivo é realizar a mutação em um resíduo que contribui de maneira favorável para a energia eletrostática e que possua SASA menor que 50%. Desta forma, foi escolhido o resíduo D83 para ser substituído por uma Arginina - ARG (R) com o intuito de realizar a troca de carga na posição escolhida.

O resíduo D83 da enzima selvagem XynA tem valor de $\Delta G_{qq} = -10.413$ kJ/mol e SASA normalizada de 0.01. A figura 10 apresenta, em um gráfico de barras, a comparação entre os novos valores de energia devido à mutação D83R e a enzima selvagem XynA. Pela análise da figura 10 é possível observar a desestabilização da energia eletrostática na maioria dos resíduos observados. Os resíduos R132, K135 e R136 tiveram a maior variação de ΔG_{qq} , além do grupo mutado D83R. Os testes experimentais de validação estão em andamento no CTBE.

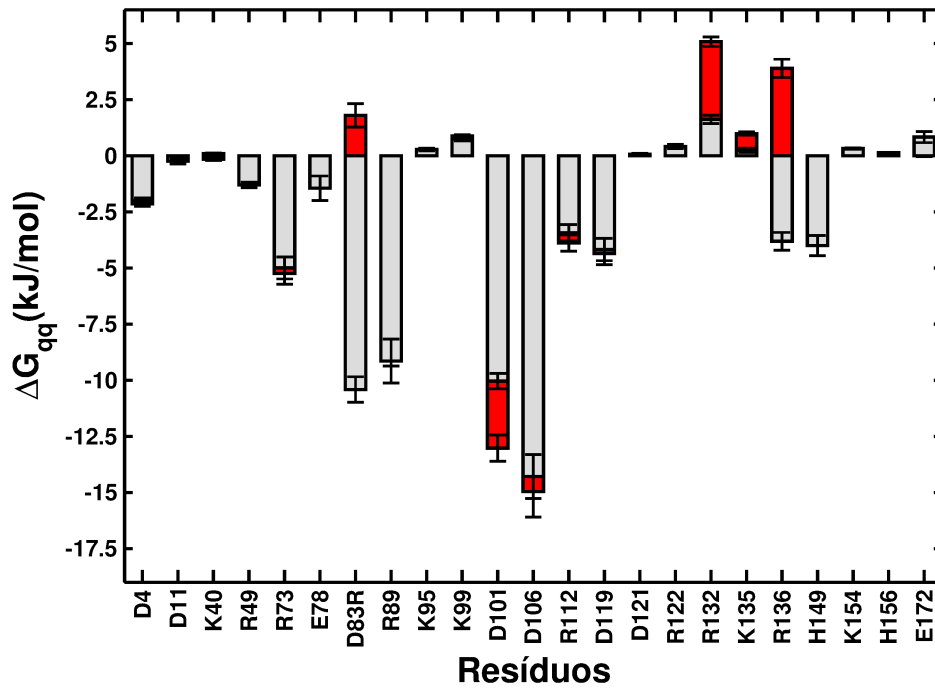


Figura 10: Comparação da energia livre de interação carga-carga de cada resíduo entre a enzima selvagem XynA e a enzima XynA com a mutação D83R. As barras em vermelho indicam os valores de energia de cada resíduo para a enzima com a mutação D83R. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos com destaque para a mutação D83R.

3.3.2 Xilanase A com mutações (G3_x4)

Nesta seção serão apresentados os resultados para a enzima Xilanase A de *Bacillus subtilis* com as mutações Q7H, G13R, S22P e S179C (G3_x4). O resultado para os valores de energia eletrostática de cada resíduo da enzima G3_x4 estão apresentados na figura 11. A enzima G3_x4 possui dois resíduos com cadeia lateral polar carregada a mais do que a enzima XynA (H7 e R13) com o número total de 25 resíduos carregados. Dentre os 25 resíduos observados, 14 apresentaram contribuições favoráveis significativas com valores de $\Delta G_{qq} < 0$, 5 apresentaram contribuições, aproximadamente, nulas com valores de $\Delta G_{qq} \approx 0$ e 6 apresentaram contribuições desfavoráveis com valores de $\Delta G_{qq} > 0$.

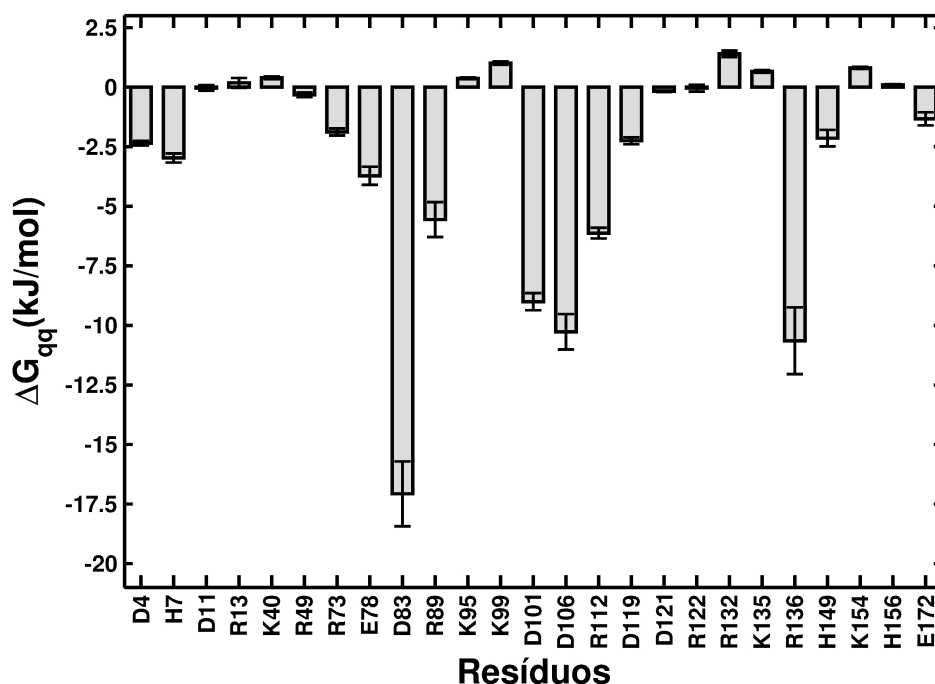


Figura 11: Representação da energia livre de interação carga–carga de cada resíduo para a enzima G3_x4. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos.

A adição dos grupos carregados H7 e R13 induziram a estabilização da energia eletrostática de outros resíduos carregados na enzima G3_x4 em relação a enzima XynA. A estabilização é mais acentuada nos resíduos E78, D83 E172, além do próprio grupo inserido H7. A estabilização devido a adição dos grupos carregados H7 e R13 pode ser uma

explicação para o aumento de termo-estabilidade observado pelo grupo experimental [49].

Similar à discussão realizada na seção 3.3.1, o foco será dado nos 6 resíduos com contribuições desfavoráveis (K40, K95, K99, R132, K135, K154). Os valores de ΔG_{qq} e de SASA de cada resíduo estão descritos na tabela IV.

Tabela IV: Valores de ΔG_{qq} e de SASA para cada resíduo com contribuição desfavorável para a proteína G3_x4.

Resíduo	$\Delta G_{qq}(\text{kJ/mol})$	SASA
K40	0.392	0.061
K95	0.364	0.415
K99	1.013	0.732
R132	1.402	0.116
K135	0.664	0.489
K154	0.321	0.523

A tabela IV destaca o resíduo K99 que contribui de maneira desfavorável para energia eletrostática e que está com a cadeia lateral mais exposta ao solvente. É apresentada na figura 12 os novos valores da energia eletrostática para a enzima G3_x4 com a mutação K99E. A figura 12 mostra a sobreposição dos valores de energia em um gráfico de barras com intuito de comparar as variações de energia de cada grupos ionizável com o efeito da mutação. A mutação K154E não será apresentada, pois não mostrou variações significativas nos valores de energia.

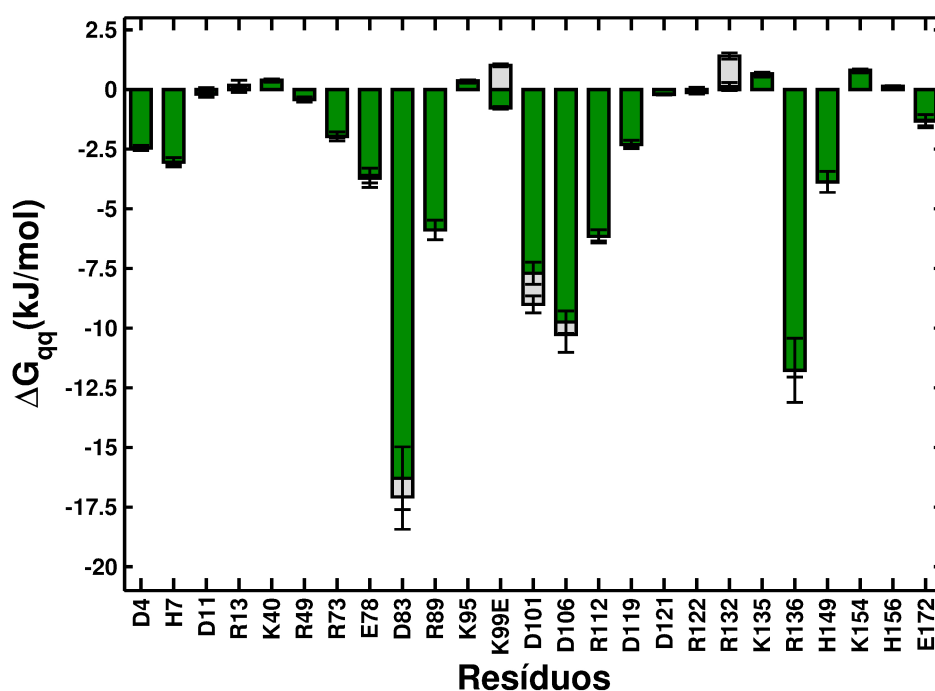


Figura 12: Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima G3_x4 e a enzima G3_x4 com a mutação K99E. As barras verdes indicam os valores de energia de cada resíduo para a enzima G3_x4 com a mutação K99E. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos com destaque para a mutação K99E.

A análise da figura 12, similar à figura 9, mostra a diferença na estabilidade dos grupos ionizáveis devido a mutação K99E. Houve um aumento na estabilidade devido à mutação e esta estabilização pode ser observada na maioria dos resíduos. É esperado que a enzima com a mutação K99E possua uma termo-estabilidade maior que a enzima selvagem XynA e maior que a enzima G3_x4. Os testes experimentais de verificação do aumento da temperatura de transição e do aumento de atividade enzimática estão em andamento no CTBE.

Também foi sugerida uma mutação para desestabilizar a enzima G3_x4 como teste de validação da metodologia. Da mesma forma realizada na enzima XynA, foi escolhido o resíduo D83 para ser substituído por uma Arginina - ARG (R) com o intuito de realizar a troca de carga na posição escolhida. O resíduo D83 da enzima G3_x4 tem valor de

$\Delta G_{qq} = -17.068$ kJ/mol e SASA normalizada de 0.01. A figura 13 apresenta a comparação entre os novos valores de energia devido à mutação D83R e a enzima G3_x4. É observado a desestabilização da energia eletrostática na maioria dos resíduos observados. Os resíduos K40, R132, K135 e R136 tiveram a maior variação de ΔG_{qq} , além do próprio grupo mutado D83R. Os testes experimentais de validação estão em andamento no CTBE.

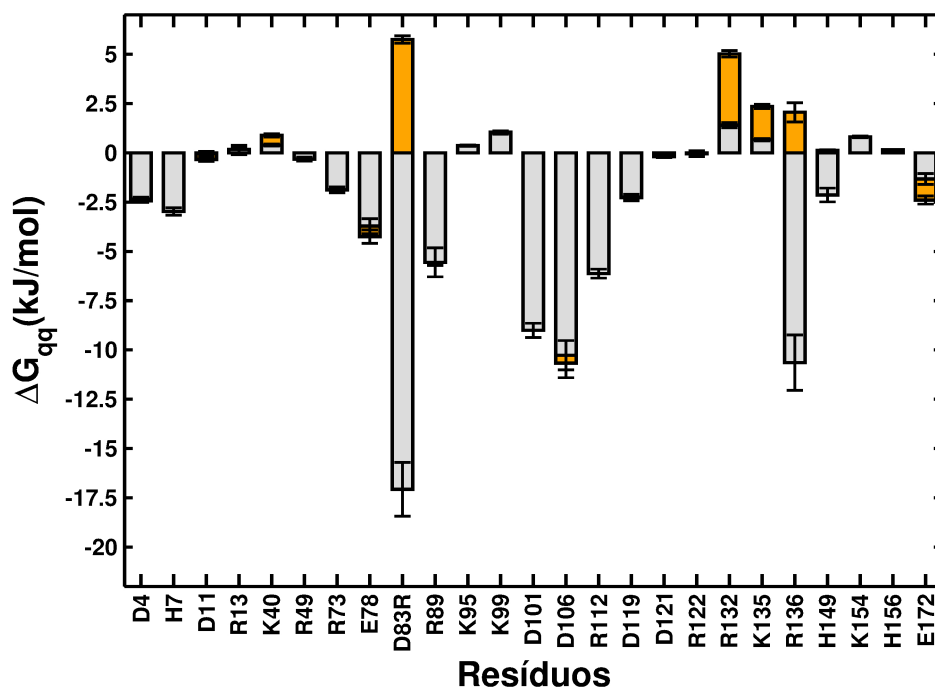


Figura 13: Comparação da energia livre de interação carga–carga de cada resíduo entre a enzima selvagem G3_x4 e a enzima G3_x4 com a mutação D83R. As barras em laranja indicam os valores de energia de cada resíduo para a enzima com a mutação D83R. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos com destaque para a mutação D83R.

3.3.3 Xilanase XylB

Nesta seção serão apresentados os resultados para a enzima Xilanase de *Bacillus pumilus* (XylB). As análises realizadas seguem o mesmo procedimento descrito nas seções 3.3.1 e 3.3.2. Para a enzima XylB foram realizadas 2 mutações simultaneamente, uma na posição K47 e outra na posição D126. As substituições foram a K47E e D126R, todas no sentido de inverter a carga na posição escolhida. A figura 14 apresenta o gráfico com a contribuição eletrostática para a enzima XylB e para a enzima com as duas mutações. A análise da figura 14 indica a estabilização energética de vários resíduos (K39, H49, H50, E83, K163, E172 e R 175), além dos resíduos mutados K47E e D126R. Os testes experimentais de caracterização da temperatura de transição e de atividade enzimática estão em andamento no CTBE para a enzima selvagem XylB e para a enzima XylB com as duas mutações K47E e D126R.

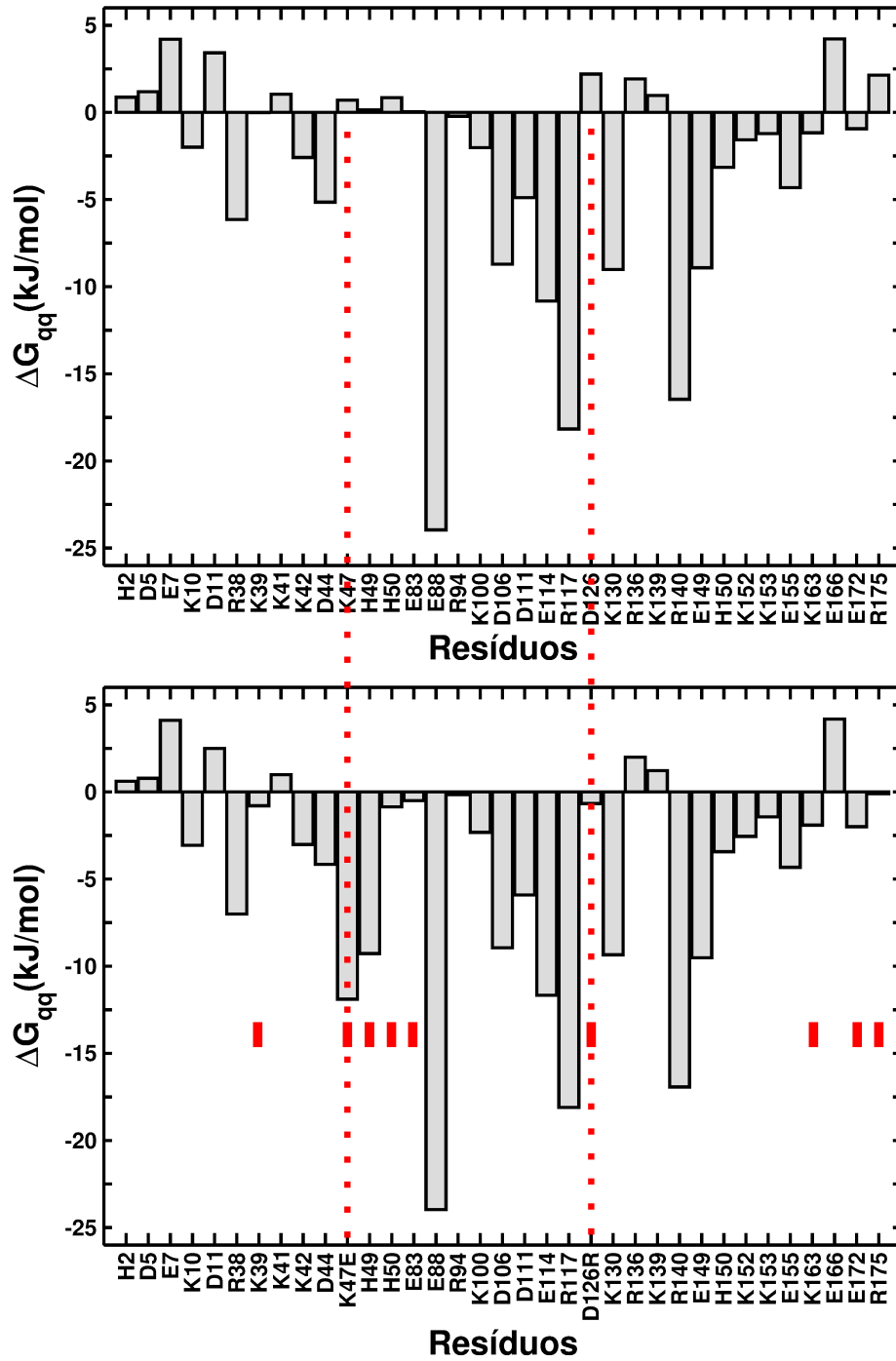


Figura 14: Representação da energia livre de interação carga-carga de cada resíduo entre a enzima selvagem XylB e a enzima XylB com as mutações K47E e D126R. O gráfico de cima se refere à enzima selvagem XylB e o gráfico de baixo se refere à enzima XylB com as mutações K47E e D126R. As linhas tracejadas vermelhas indicam as posições das mutações realizadas e as pequenas barras vermelhas no gráfico de baixo indicam os resíduos com a estabilização mais acentuada. O eixo vertical apresenta ΔG_{qq} em que cada barra representa o valor de $-\langle W_i \rangle$ do resíduo i . O eixo horizontal indica o tipo de resíduo e a sua respectiva posição na sequência de aminoácidos com destaque para as mutações K47E e D126R.

3.4 Conclusões parciais

Este trabalho procurou entender a contribuição das interações eletrostáticas na estabilidade do estado nativo de proteínas utilizando o modelo de Tanford-Kirkwood. Além de entender os efeitos das interações carga–carga, o objetivo deste trabalho é de sugerir mutações sítio-dirigidas que otimizam essa interação e que levam ao aumento da termoestabilidade. Os testes de validação dos códigos desenvolvidos foram efetuados com sucesso na comparação em diversos sistemas diferentes presentes na literatura [4, 42–48]. Uma das vantagens do método é a possibilidade da otimização ser baseada em um pH escolhido. Acredita-se que esta ferramenta possa auxiliar na otimização de enzimas com interesse na produção de bioetanol de segunda geração visando o aumento da termostabilidade e atividade. Os testes experimentais para a validação da metodologia e para concluir a proposta estão em andamento no CTBE.

3.5 Aplicação da metodologia - Blo t 5

A metodologia TKSA foi aplicada em um estudo sobre o antígeno do ácaro *Blomia tropicalis* (Blo t 5) que está relacionado com doenças alérgicas respiratórias como asma e rinite [70–72]. Este é um trabalho em colaboração com grupos teóricos e experimentais das universidades UFTM e UFU [73]. Serão apresentados os resultados teóricos obtidos pela aplicação da metodologia TKSA na região do epítipo do antígeno. Maiores detalhes sobre o trabalho estão presentes no Anexo IV - Artigo Blo t 5 submetido.

O mecanismo para a reação alérgica consiste na hiper-sensibilidade de controle do alérgeno mediado pela Imunoglobulina E (IgE) [74, 75]. Os resultados experimentais deste trabalho apresentam modificações realizadas no alérgeno que induzem uma redução de reatividade para anticorpos IgE e aumentam a antigenicidade para anticorpos Imunoglobulina G (IgG). Em termos práticos, as modificações no alérgeno diminuem a hiper-sensibilidade do sistema imunológico e, conseqüentemente, atenuam a reação alérgica [73]. Os efeitos das modificações são melhores descritos analisando a região do epítipo que está responsável pela interação do antígeno com os anticorpos. O resultado teórico apresenta os valores da energia eletrostática média $\langle W_i \rangle$ de cada grupo ionizável do antígeno selvagem (Blo t 5) e do antígeno modificado (mBlo t 5) (Figura 15). Para este trabalho foi utilizado o método de Monte-Carlo com o critério de Metropolis para a amostragem dos estado protonação - TKSA-MC [73].

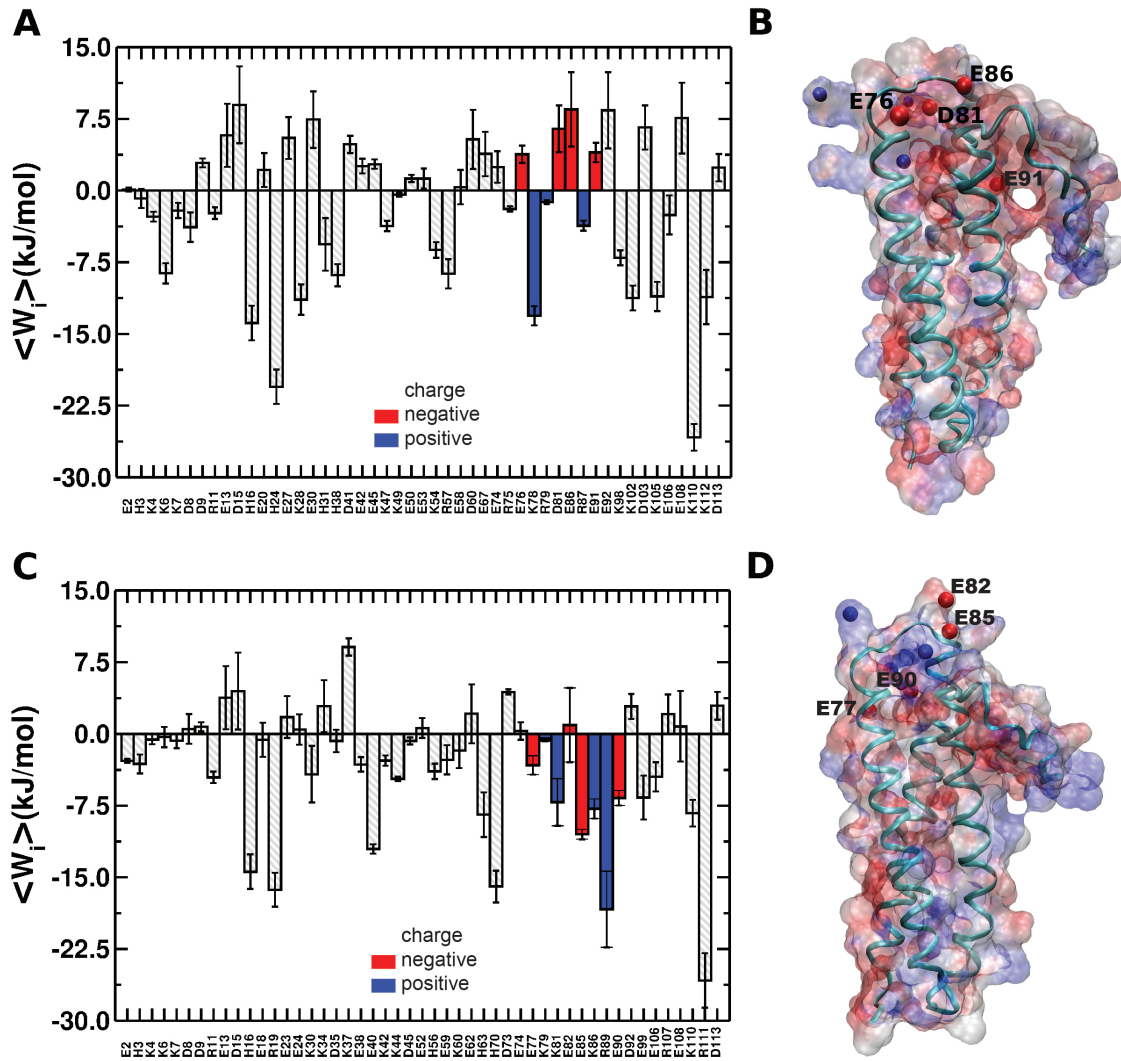


Figura 15: Representação da energia de interação eletrostática para cada resíduo dos antígenos Blo t 5 e mBlo t 5 e representação da estrutura de cada antígeno. Os gráficos descrevem a energia eletrostática $\langle W_i \rangle$ em kJ/mol no eixo vertical e o nome com a posição na sequência primária de cada grupo ionizável no eixo horizontal. As barras em cores indicam a região do epítipo para cada antígeno. A cor vermelha representa os resíduos ácidos e a cor azul representa os resíduos básicos. **A** e **B** são informações para o antígeno Blo t 5. **C** e **D** são informações para o antígeno mBlo t 5. Figura retirada do artigo submetido [73]

A região destacada em cores, na figura 15A e na figura 15C, descreve a região do epítopo para cada antígeno. As estruturas para os antígenos Blo t 5 e mBlo t 5 são apresentadas nas figuras 15B e 15D, respectivamente. Na figura 15A, quatro resíduos possuem energia eletrostática com valores positivos (E76, D81, E86 e E91), representando um conjunto de interações eletrostáticas desfavoráveis no epítopo do antígeno Blo t 5. As interações eletrostáticas na região do epítopo para a mBlo t 5 apresentam uma estabilização em relação ao antígeno Blo t 5 (Figura 15C).

Os detalhes sobre o mecanismo de ação de cada antígeno ainda não é conhecido, a sugestão é que a estabilização da energia eletrostática na região do epítopo contribui para reduzir resposta do sistema imunológico para anticorpos IgE. Foi observada uma correlação de 0.834 entre os valores de energia eletrostática de cada resíduo na região do epítopo para o antígeno Blo t 5 e a redução de reatividade ao anticorpo IgE (Figura 16). Os valores da redução de reatividade ao anticorpo IgE foram observados em pacientes que receberam os antígenos modificados para estudos. Neste trabalho, foram realizadas mutações pontuais nos resíduos ionizáveis da região do epítopo, substituindo o grupo carregado por uma Alanina [76]. Desta forma, cada uma das mutações foram testadas em pacientes, indicando a importância das interações eletrostáticas na interação entre o antígeno Blo t 5 e o anticorpo IgE.

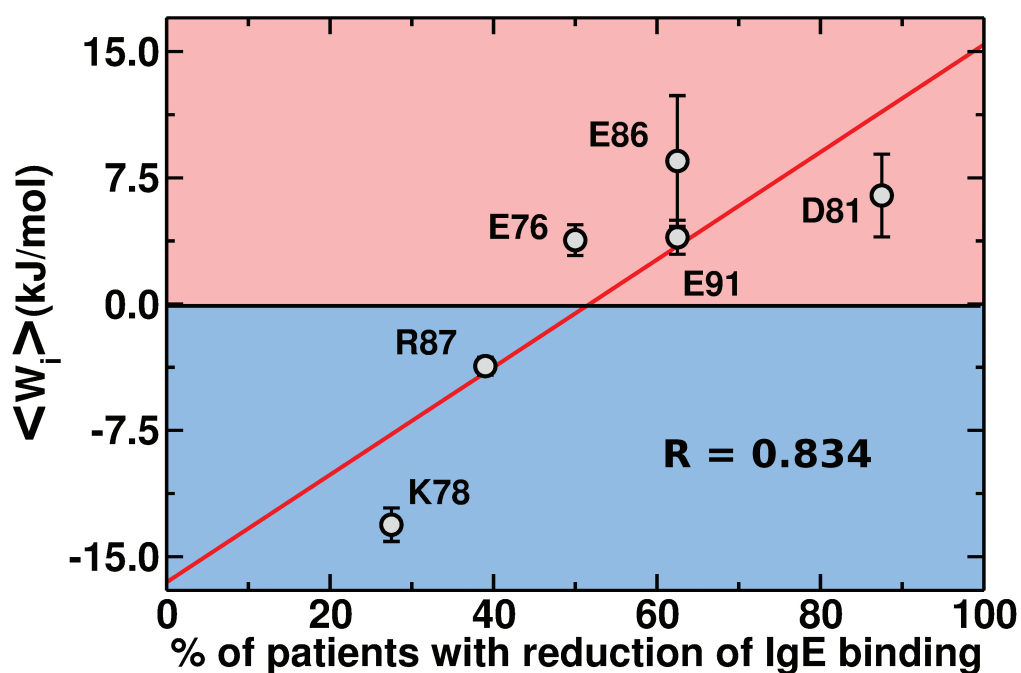


Figura 16: Correlação entre a energia eletrostática e a porcentagem de pacientes com redução na reatividade com o anticorpo IgE. No eixo vertical, $\langle W_i \rangle$ é o valor médio da energia eletrostática em kJ/mol. No eixo horizontal estão os valores em porcentagem dos pacientes que tiveram redução na reatividade com o anticorpo IgE. A correlação de 0.834 é obtida pelo ajuste linear dos pontos indicados pelo resíduos carregados existentes na região do epítipo e que são substituídos por Alaninas. Figura retirada do artigo submetido [73]

Referências

- [1] 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. *Journal of Chemical Theory and Computation*, 12(7):3270–3277, July 2016.
- [2] 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. *Archives of Biochemistry and Biophysics*, 469(1):20–28, January 2008.
- [3] Ariel Azia and Yaakov Levy. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *Journal of Molecular Biology*, 393(2):527–542, October 2009.
- [4] 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: Structure, Function, and Bioinformatics*, 80(3):858–870, 2012.
- [5] 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. *Biophysical Journal*, 90(8):2911–2921, April 2006.
- [6] 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. *Pflügers Archiv: European Journal of Physiology*, 454(3):373–384, June 2007.
- [7] Ohad Givaty and Yaakov Levy. Protein sliding along DNA: dynamics and structural characterization. *Journal of Molecular Biology*, 385(4):1087–1097, January 2009.
- [8] Swarnendu Tripathi, Angel E. García, and George I. Makhatadze. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge–Charge Interactions. *The Journal of Physical Chemistry B*, 119(41):13103–13112, October 2015.

- [9] Jana Khandogin and Charles L. Brooks. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophysical Journal*, 89(1):141–157, July 2005.
- [10] 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. *Journal of Molecular Biology*, 298(5):937–953, May 2000.
- [11] 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*, 75(2):430–441, May 2009.
- [12] 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: Structure, Function, and Bioinformatics*, 81(10):1727–1737, October 2013.
- [13] Matheus R. de Mendonça, 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: Structure, Function, and Bioinformatics*, 82(1):119–129, January 2014.
- [14] Vinícius de Godoi Contessoto. *Estudo do efeito da adição de frustração no enovelamento de proteínas utilizando modelos baseados em estruturas*. Dissertação de Mestrado, Universidade Estadual Paulista (UNESP), 2012.
- [15] Paulo Ricardo Mouro, Vinícius de Godoi Contessoto, Jorge Chahine, Ronaldo Juníode Oliveira, and Vitor Barbanti Pereira Leite. Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophysical Journal*, 111(2):287–293, July 2016.
- [16] Donald Allan McQuarrie. *Statistical Mechanics*. University Science Books, 2000.

- [17] V. M. de Oliveira and S. J. de Carvalho. Adsorption of pH-responsive polyelectrolyte chains onto spherical macroions. *The European Physical Journal E*, 37(8), August 2014.
- [18] 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. *Journal of Molecular Biology*, 289(1):167–174, May 1999.
- [19] 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. *Journal of Molecular Biology*, 284(5):1661–1670, December 1998.
- [20] 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.
- [21] 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. *Journal of Molecular Biology*, 299(4):1091–1100, June 2000.
- [22] Henry N. Po and N. M. Senozan. The Henderson-Hasselbalch Equation: Its History and Limitations. *Journal of Chemical Education*, 78(11):1499, November 2001.
- [23] 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.
- [24] 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.

- [25] Monika A. Coronado, Icaro P. Caruso, Vinícius M. de Oliveira, Vinícius G. Contessoto, Vitor B. P. Leite, Raghuvir K. Arni, and Raphael J. Eberle. Significance of ph and nacl concentration in the refolding process of cold shock protein a from *corynebacterium pseudotuberculosis*. *Spectrochimica Acta Part A*, 2016. Submitted - in revision.
- [26] Michelle Grayson. Biofuels. *Nature*, 474(7352):S1, June 2011.
- [27] Michael E. Himmel, Shi-You Ding, David K. Johnson, William S. Adney, Mark R. Nimlos, John W. Brady, and Thomas D. Foust. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science*, 315(5813):804–807, 2007.
- [28] Marcia Moraes. Perspective: Lessons from Brazil. *Nature*, 474(7352):S25, June 2011.
- [29] Martin Robbins. Policy: Fuelling politics. *Nature*, 474(7352):S22–S24, June 2011.
- [30] Edward M. Rubin. Genomics of cellulosic biofuels. *Nature*, 454(7206):841–845, 2008.
- [31] Katharine Sanderson. Lignocellulose: A chewy problem. *Nature*, 474(7352):S12–S14, June 2011.
- [32] Arthur J. Ragauskas, Charlotte K. Williams, Brian H. Davison, George Britovsek, John Cairney, Charles A. Eckert, William J. Frederick, Jason P. Hallett, David J. Leak, Charles L. Liotta, Jonathan R. Mielenz, Richard Murphy, Richard Templer, and Timothy Tschaplinski. The Path Forward for Biofuels and Biomaterials. *Science*, 311(5760):484–489, January 2006.
- [33] Michael E. Himmel, William S. Adney, John O. Baker, Richard Elander, James D. McMillan, Rafael A. Nieves, John J. Sheehan, Steven R. Thomas, Todd B. Vinzant, and Min Zhang. Advanced Bioethanol Production Technologies: A Perspective. In Badal C. Saha and Jonathan Woodward, editors, *Fuels and Chemicals from Biomass*, volume 666, pages 2–45. American Chemical Society, Washington, DC, May 1997.
- [34] Nathan Mosier, Charles Wyman, Bruce Dale, Richard Elander, Y.Y. Lee, Mark Holtzapple, and Michael Ladisch. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, 96(6):673–686, 2005.

- [35] Miranda Maki, Kam Tin Leung, and Wensheng Qin. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*, 5(5):500–516, 2009.
- [36] David B Wilson. Cellulases and biofuels. *Current Opinion in Biotechnology*, 20(3):295–299, June 2009.
- [37] Peter Fairley. Introduction: Next generation biofuels. *Nature*, 474(7352):S2–S5, June 2011.
- [38] Neil Savage. Fuel options: The ideal biofuel. *Nature*, 474(7352):S9–S11, June 2011.
- [39] Lee R. Lynd, Janet H. Cushman, Roberta J. Nichols, and Charles E. Wyman. Fuel Ethanol from Cellulosic Biomass. *Science*, 251(4999):1318 –1323, March 1991.
- [40] Zhanmin Fan, Joshua R. Werkman, and Ling Yuan. Engineering of a multifunctional hemicellulase. *Biotechnology Letters*, 31(5):751–757, January 2009.
- [41] Jonathan M. Galazka, Chaoguang Tian, William T. Beeson, Bruno Martinez, N. Louise Glass, and Jamie H. D. Cate. Cellodextrin Transport in Yeast for Improved Biofuel Production. *Science*, 330(6000):84 –86, 2010.
- [42] 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.
- [43] 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. *Journal of Molecular Biology*, 336(4):929–942, February 2004.
- [44] 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 engineering, design & selection: PEDS*, 18(9):425–433, September 2005.

- [45] Vakhtang V. Loladze and George I. Makhatadze. Energetics of charge-charge interactions between residues adjacent in sequence. *Proteins*, 79(12):3494–3499, December 2011.
- [46] Chi-Ho Chan, Cecily C. Wilbanks, George I. Makhatadze, and Kam-Bo Wong. Electrostatic contribution of surface charge residues to the stability of a thermophilic protein: benchmarking experimental and predicted pKa values. *PloS One*, 7(1):e30296, 2012.
- [47] 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.
- [48] 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. *Proceedings of the National Academy of Sciences*, 106(8):2601–2606, February 2009.
- [49] Roberto Ruller, Laila Deliberto, Tatiana Lopes Ferreira, and Richard J Ward. Thermostable variants of the recombinant xylanase A from *Bacillus subtilis* produced by directed evolution show reduced heat capacity changes. *Proteins*, 70(4):1280–1293, March 2008.
- [50] Roberto Ruller, Juliana Alponi, Laila Aparecida Deliberto, Letícia Maria Zanphorlin, Carla Botelho Machado, and Richard John Ward. Concomitant adaptation of a GH11 xylanase by directed evolution to create an alkali-tolerant/thermophilic enzyme. *Protein engineering, design & selection: PEDS*, 27(8):255–262, August 2014.
- [51] Charles Tanford and John G. Kirkwood. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *Journal of the American Chemical Society*, 79(20):5333–5339, October 1957.
- [52] James B. Matthews and Frank R. N. Gurd. [17] Calculation of electrostatic interactions in proteins. In Serge N. Timasheff C. H. W. Hirs, editor, *Methods in Enzymology*, volume 130 of *Enzyme Structure Part K*, pages 413–436. Academic Press, 1986.

- [53] James B. Matthew and Frank R. N. Gurd. [18] Stabilization and destabilization of protein structure by charge interactions. In Serge N. Timasheff C. H. W. Hirs, editor, *Methods in Enzymology*, volume 130 of *Enzyme Structure Part K*, pages 437–453. Academic Press, 1986.
- [54] James J. Havranek and Pehr B. Harbury. Tanford–Kirkwood electrostatics for protein modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20):11145–11150, September 1999.
- [55] John G. Kirkwood. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *The Journal of Chemical Physics*, 2(7):351–361, July 1934.
- [56] Fernando Luís B. Da Silva, Bo Jönsson, and Robert Penfold. A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations. *Protein Science : A Publication of the Protein Society*, 10(7):1415–1425, July 2001.
- [57] 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, November 1991.
- [58] Alan Fersht. *Enzyme Structure and Mechanism*. W H Freeman & Co, New York, January 1985.
- [59] A. S. Yang and B. Honig. On the pH dependence of protein stability. *Journal of Molecular Biology*, 231(2):459–474, May 1993.
- [60] 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. *Journal of Molecular Biology*, 327(5):1135–1148, April 2003.
- [61] Katrina L. Schweiker, Arash Zarrine-Afsar, Alan R. Davidson, and George I. Makhatadze. Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge–charge interactions. *Protein Science : A Publication of the Protein Society*, 16(12):2694–2702, December 2007.

- [62] V. V. Loladze, B. Ibarra-Molero, J. M. Sanchez-Ruiz, and G. I. Makhatadze. Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry*, 38(50):16419–16423, December 1999.
- [63] Vakhtang V. Loladze and George I. Makhatadze. Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. *Protein Science: A Publication of the Protein Society*, 11(1):174–177, January 2002.
- [64] Katrina L. Schweiker and George I. Makhatadze. A computational approach for the rational design of stable proteins and enzymes: optimization of surface charge-charge interactions. *Methods in Enzymology*, 454:175–211, 2009.
- [65] Beatriz Ibarra-Molero and Jose M. Sanchez-Ruiz. Genetic Algorithm to Design Stabilizing Surface-Charge Distributions in Proteins. *The Journal of Physical Chemistry B*, 106(26):6609–6613, July 2002.
- [66] Juliana Sanchez Alponi, Raquel Fonseca Maldonado, and Richard J. Ward. Thermostabilization of *Bacillus subtilis* GH11 xylanase by surface charge engineering. *International Journal of Biological Macromolecules*, 87:522–528, June 2016.
- [67] A. Fiser, R. K. Do, and A. Sali. Modeling of loops in protein structures. *Protein Science: A Publication of the Protein Society*, 9(9):1753–1773, September 2000.
- [68] Antonija Kuzmanic and Bojan Zagrovic. Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *Biophysical Journal*, 98(5):861–871, March 2010.
- [69] Laurence Lins, Annick Thomas, and Robert Brasseur. Analysis of accessible surface of residues in proteins. *Protein Science : A Publication of the Protein Society*, 12(7):1406–1417, July 2003.
- [70] F. Carswell. The relationship between mite allergen exposure and asthma severity. *Clinical & Experimental Allergy*, 25(2):99–101, February 1995.
- [71] R. Sporik, M. D. Chapman, and T a. E. Platts-Mills. House dust mite exposure as a cause of asthma. *Clinical & Experimental Allergy*, 22(10):897–906, October 1992.

- [72] Matthew Masoli, Denise Fabian, Shaun Holt, Richard Beasley, and Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*, 59(5):469–478, May 2004.
- [73] Bárbara G. M. Ávila, Laura F. Tomaz, Carlos R. Prudencio, Karine C. de Almeida, Deise Ap. O. Silva, Odonório Abrahão, Vinícius M. de Oliveira, Vinícius G., Vitor B. P. Leite, Pamela Ap. Candido, Ronaldo J. de Oliveira, Jair P. Cunha-Júnior, and Ernesto A. Taketomi. Epitope charge distribution controls the antigenicity for binding human ige4 isotype in a modified blo t 5, the major allergen from dust mite *blomia tropicalis*. *Immunobiology*, 2016. Submitted - in revision.
- [74] F. T. Chew, F. C. Yi, K. Y. Chua, E. Fernandez-Caldas, L. K. Arruda, M. D. Chapman, and B. W. Lee. Allergenic differences between the domestic mites *Blomia tropicalis* and *Dermatophagoides pteronyssinus*. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology*, 29(7):982–988, July 1999.
- [75] Kaw Yan Chua, Nge Cheong, I.-Chun Kuo, Bee Wah Lee, Fong Cheng Yi, Chiung-Hui Huang, and Lip Nyin Liew. The *Blomia tropicalis* allergens. *Protein and Peptide Letters*, 14(4):325–333, 2007.
- [76] Siew Leong Chan, Tan Ching Ong, Yun Feng Gao, Yuen Sung Tiong, De Yun Wang, Fook Tim Chew, and Yu Keung Mok. Nuclear magnetic resonance structure and IgE epitopes of Blo t 5, a major dust mite allergen. *Journal of Immunology (Baltimore, Md.: 1950)*, 181(4):2586–2596, August 2008.
- [77] David J. Griffiths. *Introduction to Electrodynamics*. Pearson, Boston, 4 edition edition, October 2012.

A Descrição do modelo de Tanford-Kirkwood

Neste apêndice A será apresentada a descrição dos parâmetros A_{ij} , B_{ij} e C_{ij} presentes na equação 1 utilizados no modelo de Tanford-Kirkwood (TK). Os parâmetros fazem parte do resultado do trabalho de John Kirkwood [55]. No trabalho de Kirkwood é realizado a solução de um sistema com cargas inseridas dentro de uma cavidade esférica de baixa constante dielétrica envolvida por uma solução eletrolítica com uma constante dielétrica maior. Em nosso modelo, a proteína será a região com baixa constante dielétrica ϵ_p e o solvente terá a constante dielétrica maior ϵ_s . O sistema envolve 3 regiões, a região 1 será a cavidade esférica de raio b e constante dielétrica ϵ_p que representa a proteína, a região 2 representa a área de exclusão de íons entre o raio a e o raio b e a região 3 representa a solução eletrolítica (Figura 17).

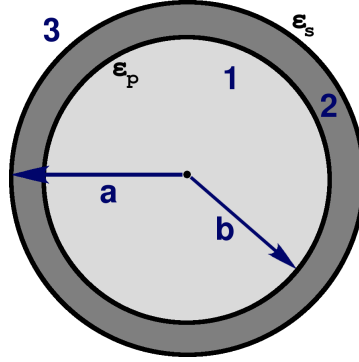


Figura 17: Representação esquemática para o trabalho de Kirkwood. Os números indicam as regiões que são abordadas no modelo e as letras representam os valores dos raios de cada região.

O potencial φ_1 referente à região 1 é descrito pela equação de Laplace [77]. O potencial eletrostático para qualquer ponto (r, θ, ϕ) dentro da esfera de raio b é dado por:

$$\varphi_1(r < b, \theta, \phi) = \sum_{j=1}^M \frac{z_j}{\epsilon_p |\vec{r} - \vec{r}_j|} + \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} B_{nm} r^n P_n^m(\cos \theta) \exp(im\varphi) \quad (10)$$

onde M é o número de grupos ionizáveis z_j da proteína na posição $\vec{r}_j$, $P_n^m(\cos \theta)$ são os polinômios de Legendre e B_{nm} é o parâmetro que deve ser determinado pelas condições de contorno. O primeiro termo da equação 10 está relacionado com a auto-energia da configuração de cargas do sistema e o segundo termo está associado com as interações das

cargas com a solução eletrolítica.

Para a região 2, o modelo assume que distância máxima de aproximação de um íon da solução com a proteína é referente a uma esfera de raio a , concêntrica com a esfera da proteína de raio b . O potencial φ_2 para a região 2 também deve satisfazer a equação de *Laplace* e é dado por:

$$\varphi_2(b < r < a, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \left[\frac{C_{nm}}{r^{n+1}} + G_{nm} r^n \right] P_n^m(\cos \theta) \exp(im\phi) \quad (11)$$

A solução para a região externa à esfera de raio a envolve uma distribuição de carga presente na solução eletrolítica. O potencial φ_3 para a região 3 deve satisfazer a equação de *Poisson* [77]. O modelo assume uma distribuição de carga média e deve satisfazer a condição:

$$\nabla^2 \varphi_3 - \kappa^2 \varphi_3 = 0 \quad (12)$$

$$\kappa^2 = (4\pi N e^2 / 1000 \epsilon_s k T) \sum_i c_i z_i^2 \quad (13)$$

onde κ é função da força iônica do eletrólito ($\sum_i c_i z_i^2$), do número de *Avogadro* N e da carga elementar e , k é a constante de *Boltzmann*, T é a temperatura do ambiente e ϵ_s é a constante dielétrica do meio. Diante dessas condições o potencial φ_3 é descrito por:

$$\varphi_3(r > a, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \frac{A_{nm}}{r^{n+1}} \exp(-\kappa r) K_n(\kappa r) P_n^m(\cos \theta) \exp(im\phi) \quad (14)$$

$$K_n(x) = \sum_{s=0}^n \frac{2^s n! (2n-s)!}{s! (2n)! (n-s)!} x^s \quad (15)$$

Os potenciais φ_1 , φ_2 e φ_3 e as constantes A_{nm} , B_{nm} , C_{nm} e G_{nm} devem ser determinados para cada valor de m e n . As interfaces entre as regiões devem satisfazer a condição de continuidade do potencial:

$$[\varphi_3 = \varphi_2]_{r=a} \quad (16)$$

$$[\varphi_2 = \varphi_1]_{r=b} \quad (17)$$

Para a região 2 e 3, como não há descontinuidade da constante dielétrica, é necessário satisfazer a condição:

$$\left[\vec{\nabla} \varphi_3 = \vec{\nabla} \varphi_2 \right]_{r=a} \quad (18)$$

Para a interface entre as regiões 1 e 2 é preciso considerar a descontinuidade da constante dielétrica, neste caso, é necessário ter a continuidade do vetor deslocamento elétrico e das componentes do campo elétrico:

$$\left[\frac{1}{r} \frac{\partial \varphi_2}{\partial \theta} = \frac{1}{r} \frac{\partial \varphi_1}{\partial \theta} \right]_{r=b} \quad (19)$$

$$\left[\frac{1}{r \sin \theta} \frac{\partial \varphi_2}{\partial \phi} = \frac{1}{r \sin \theta} \frac{\partial \varphi_1}{\partial \phi} \right]_{r=b} \quad (20)$$

$$\left[\epsilon_s \frac{\partial \varphi_2}{\partial r} = \epsilon_p \frac{\partial \varphi_1}{\partial r} \right]_{r=b} \quad (21)$$

A solução da aplicação das condições de contorno descreve os parâmetros A_{ij} , B_{ij} e C_{ij} utilizados no modelo Tanford-Kirkwood (TK). Os detalhes da solução dos parâmetros estão descritos abaixo:

$$A_{ij} = \frac{1}{\epsilon_p |\vec{r}_i - \vec{r}_j|} \quad (22)$$

$$B_{ij} = \frac{1}{b \epsilon_p} \sum_{n=0}^{\infty} \frac{(\epsilon_s - \epsilon_p)}{(\epsilon_s - n \epsilon_p / (n+1))} \left(\frac{r_i r_j}{b^2} \right)^n P_n(\cos \theta_{ij}) \quad (23)$$

$$C_{ij} = \frac{1}{a \epsilon_s} \left[\frac{a \kappa}{1 + a \kappa} + (a \kappa)^2 \sum_{n=1}^{\infty} \frac{\frac{2n+1}{2n-1} \left(\frac{\epsilon_s}{(n+1) \epsilon_s + n \epsilon_p} \right) \left(\frac{r_i r_j}{a^2} \right)^n P_n(\cos \theta_{ij})}{\frac{K_{n+1}(a \kappa)}{K_{n-1}(a \kappa)} + \frac{n(\epsilon_s - \epsilon_p)}{(n+1) \epsilon_s + n \epsilon_p} \left(\frac{(a \kappa)^2}{4n^2 - 1} \right) \left(\frac{b}{a} \right)^{2n+1}} \right] \quad (24)$$

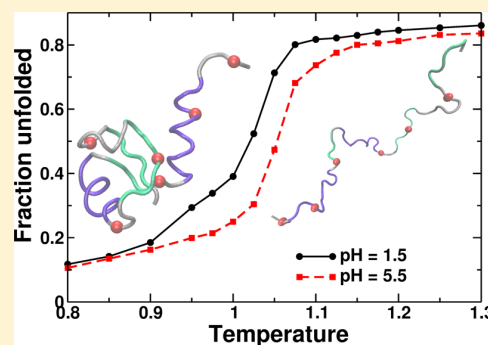
NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence

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

Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista (UNESP), São José do Rio Preto, São Paulo 15054-000, Brazil

Supporting Information

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–Hückel 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 pK_a 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.



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, such as charge regulation and pH dependence, cannot 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) is employed in this work in order to understand how pH

modulates folding mechanisms. Thus, a Debye–Hückel^{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 range similar 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 β sheet sandwiched between two α helices²⁸ (Figure 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 experimental ones. The pK_a values from other computational studies are also compared with the CpHMD + C_α SBM^{33,34} (Support Information (S1) Table S1).

Received: April 20, 2016

Published: June 21, 2016

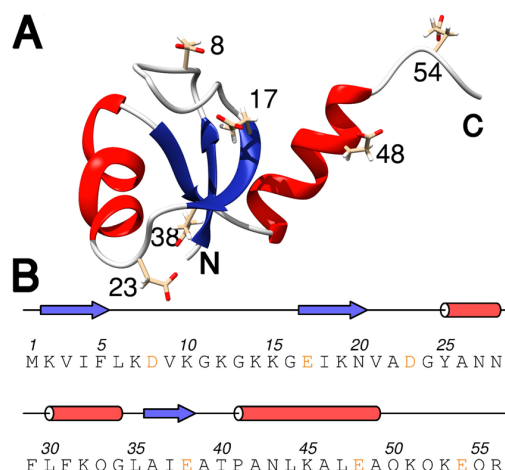


Figure 1. (A) Structure of NTL9, Protein Data Bank (PDB) 1CQU, with highlighted acid side chain residues. (B) Primary sequence of NTL9 with highlighted acid residues.

The present work 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.

2. METHODS

2.1. Structure-Based C_α Model. The NTL9 protein was modeled using 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 Debye–Hückel theory. The charge–charge interaction was given by a Coulomb screening potential (last term of eq 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

$$V(\Gamma, \Gamma_o) = \sum_{\text{bonds}} e_r(r - r_o)^2 + \sum_{\text{angles}} e_\theta(\theta - \theta_o)^2 + \sum_{\text{dihedrals}} e_\phi \left\{ [1 - \cos(\phi - \phi_o)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_o))] \right\} + \sum_{\text{contacts}} e_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{\text{noncontacts}} e_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} + \sum_{\text{electrostatics}} K_{\text{electrostatics}} \frac{q_i q_j \exp^{-\kappa r}}{\epsilon_K r_{ij}} \quad (1)$$

where r_o , θ_o , and ϕ_o 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 (Protein Data Bank ID (PDB ID) 1CQU). d_{ij} is the native distance between the pair contacts i and j determined by CSU software (Contact of Structural Units),³⁹ $e_C = 1.0$ kcal/mol, $e_r = 100$ kcal/(mol $\text{\AA}^2$), $e_\theta = 20$ kcal/mol, $e_{NC} = 1.0$ kcal/mol, $\sigma_{NC} = 4.00$ $\text{\AA}$,^{21,40} and $K_{\text{electrostatics}} = 332$ kcal $\text{\AA}/(\text{mol } e^2)$.^{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 dielectric constant is $\epsilon_K = 80$.

2.2. Simulation Details. The constant-pH molecular dynamic method 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 $1.2 d_{ij}$ from the native structure.

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 titratable group of the residue and known pK_a , which takes into account local energetic contributions that cannot 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

$$\Delta E = \pm \xi k_B T \ln(10)(\text{pH} - pK_a) + \Delta E_{\text{elec}} \quad (2)$$

where the positive and negative signs are used for protonation or deprotonation, respectively, and the constant ξ is 1 or -1 for acidic 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 1.

Table 1. 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

2.3.1. 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 by 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 by

$$\Phi^k = \frac{\Delta F_{TS}^k - \Delta F_U^k}{\Delta F_N^k - \Delta F_U^k} \quad (3)$$

where ΔF_U^k , ΔF_{TS}^k , and ΔF_N^k 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 the values of native contacts fractions 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 relative to the native state.

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 $0.5 k_B T$ 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 residues i and j that was made in the transition state. Both parameters are obtained by employing a single simulation at the folding temperature.

2.3.2. 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

$$R(Q) = \sum_{i=1}^M \frac{\langle (Q_i)_Q - Q \rangle_Q^2}{MQ(1-Q)} \quad (4)$$

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

2.3.3. 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 $\alpha_i(\text{pH}) = \langle |Z_i(\text{pH})| \rangle$. The ionization degree was calculated for different pHs, and then the titration curves were built. The generalized Henderson–Hasselbalch⁶³ equation $\text{pH} = pK_{ai} + \log \{ \alpha_i(\text{pH}) / [1 - \alpha_i(\text{pH})] \}$ was adopted to fit the titration curves. It was possible to calculate the pK_a value of each acid residue.¹⁸

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 the 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 the SI for more details). The uncertainty in the pK_a values was obtained by the mean and standard deviation of three independent set simulations (all errors were of the order of 0.1 or lower pK_a units and were omitted).

For NTL9, salt concentrations of $C_s = 10, 100$, and 750 mM were used ($\kappa \approx 0.03, 0.1$, and $0.28 \text{ \AA}^{-1}$, respectively). In the last two cases, the calculated pK_a values were compared with experiments, shown in Table 2. For this comparison the definition of Goh et al. was used.³³ The error in calculated pK_a

Table 2. Comparison between the Experimental pK_a Values¹³ and the Computational pK_a Values for the Acidic Residues in NTL9

residue	$C_s = 10 \text{ mM}$		$C_s = 100 \text{ mM}$		$C_s = 750 \text{ mM}$	
	exp pK_a	comp pK_a^k	exp pK_a^b	comp pK_a^k	exp pK_a^b	comp pK_a^k
ASP-8	^a	0.6	3.0	1.9	3.2	3.9
GLU-17	^a	1.4	3.6	3.0	3.8	4.4
ASP-23	^a	2.0	3.1	3.2	3.2	4.0
GLU-38	^a	2.4	4.0	3.6	4.1	4.4
GLU-48	^a	2.4	4.2	3.5	4.4	4.4
GLU-54	^a	2.7	4.2	3.4	4.2	4.3

^aThere are no experimental values of pK_a in this salt concentration available in literature; ^bThe experimental pK_a values for the acidic residues measured by Kuhlman et al.¹³ ^cThe computational pK_a values fitting the individual titration curves with the Henderson–Hasselbalch equation.

values is considered significant only in cases where the difference from the experimental pK_a is higher than a 1.0 pK_a unit. For both salt concentrations, only ASP-8 presented a difference of -1.1 pK_a units at 100 mM salt. This difference may be explained by the salt bridge interactions that ASP-8 forms with the amide backbone of near residues³³ 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 bridge 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 solvent⁶⁶ 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 pK_a values of all residues close to the pK_a values of ASP and GLU in isolated solution (Table 1). This is because the interaction between the charged groups at this ionic strength is almost zero and their protonation behavior is almost the same as that in 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 pK_a values, which can be seen in Table 2.

The folding/unfolding process is monitored by the fraction of native contact formation. Figure 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 (Figure 2A), 10 mM (Figure 2B), and 750 mM (Figure 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 the total electrostatic energy ($\langle \Delta E_{\text{elec}} \rangle = \langle E_{\text{elec}}^{1.5} \rangle - \langle E_{\text{elec}}^{5.5} \rangle$) for pH 1.5 and 5.5 is about $0.5 k_B T$, 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 $\langle \Delta E_{\text{elec}} \rangle$ is about $15.0 k_B T$ 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 behavior in stability, with $\langle \Delta E_{\text{elec}} \rangle$ close to $4.0 k_B T$.

In order to compare the computational and experimental results, the free energy barrier of unfolding was plotted as a function of pH (Figure 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

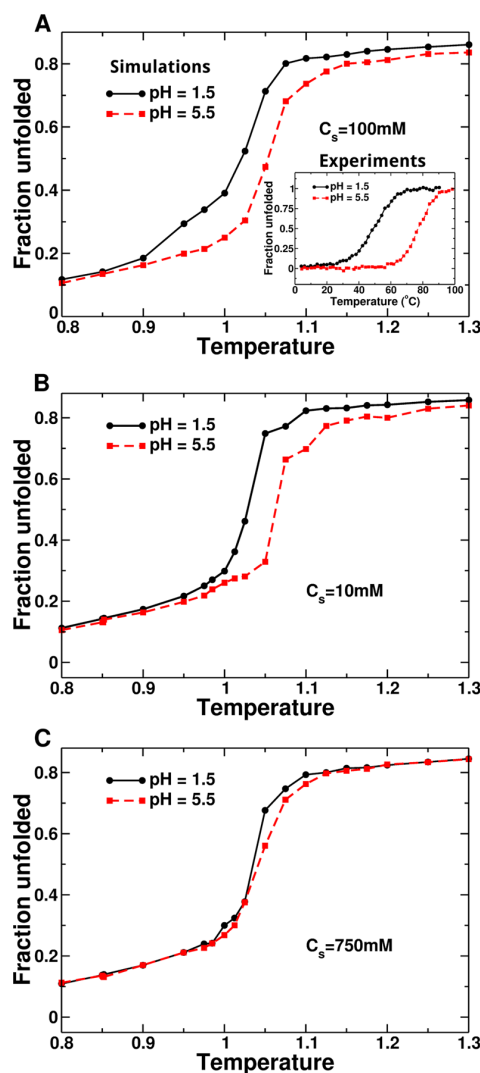


Figure 2. Fraction of unfolded configurations as a function of temperature in reduced units for two different pH conditions. The black circles represent the curve for pH 1.5 and the red squares that for pH 5.5. (A) Results obtained at $C_s = 100$ mM. Inset graph shows experimental data of the thermal unfolding curve of NTL9 from Luisi and Raleigh at same salt concentration (adapted with permission from ref 25; copyright 2000 Elsevier). (B and C) Results of simulations at $C_s = 10$ and 750 mM, respectively.

simplified model, the computational results were in good agreement with the experiments obtained by Luisi and co-workers.^{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.

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 4 for NTL9 without electrostatic interaction and with this interaction at pH 1.5 and pH 5.5. It was used at a temperature near the transition midpoint. At pH 5.5 the temperature in reduced units was

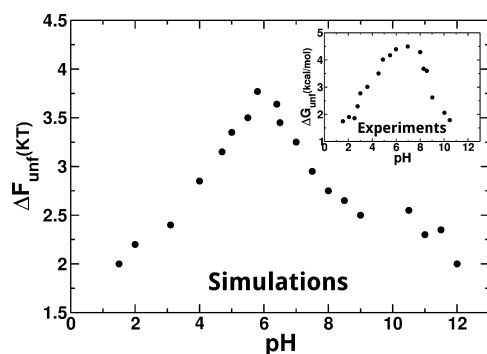


Figure 3. Free energy unfolding barrier ΔF_{unf} as a function of pH. The values were obtained from simulation analysis with WHAM for 19 different pH values from 1.5 to 12.0 and $\kappa = 0.1 \text{ \AA}^{-1}$ similar to experimental conditions. The inset plot is from experimental values for protein stability ΔG_{unf} as a function of pH (adapted with permission from ref 25; copyright 2000 Elsevier).

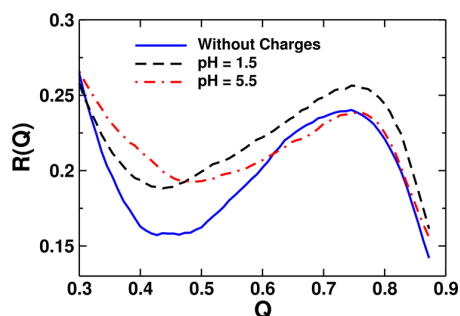


Figure 4. Folding route measurement $R(Q)$ as a function of native contacts Q for different pH conditions. The continuous blue line represents the route for the standard C_α model without electrostatic interactions. The dashed black line represents the route for pH 1.5, and the dashed–dotted red line is the route measurement for pH 5.5.

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 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 100 mM salt.

At intermediate values of Q (0.4–0.6) 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 of 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.

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 analyze the native contacts formed in TS. Experimental Φ -values of three residues of NTL9 (ASP-8, LYS-12, and GLU-17) were found in the 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 Cho and Raleigh,⁶⁸ this is not the case and these residues are involved in non-native electrostatic interactions in both transition and denatured states.

In Figure 5A the $\Delta\Phi_{ij}^k$ 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 5A was also divided into three regions (left, blue, from residues 1 to 20; middle, red, from residues 20 to 39; right, white, from residues 40 to 56) because the $\Delta\Phi_{ij}^k$ present distinct features. In the left region, from the N-terminal to residue 20, the $\Delta\Phi_{\text{pH},\emptyset}$ tends to be positive and the peaks occur in ASP-8 and LYS-12 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 ASP-8 and LYS-12 indicates that this contact formation also occurs in the 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 other charged residues also becomes more important in the TS. In the middle region, from residue 20 to residue 39, $\Delta\Phi_{\text{pH},\emptyset}$ tends to be close to zero or slightly negative. This behavior of $\Delta\Phi_{\text{pH},\emptyset}$ 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 $\Delta\Phi_{\text{pH},\emptyset}$ 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 , Δf_{ij}^{kl} (using the same notation of $\Delta\Phi_{ij}^k$) of NTL9 (Figure 5B,C) 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 $\Delta\Phi_{\text{pH},\emptyset}$ values for pH 1.5 and pH 5.5 although there are practically no

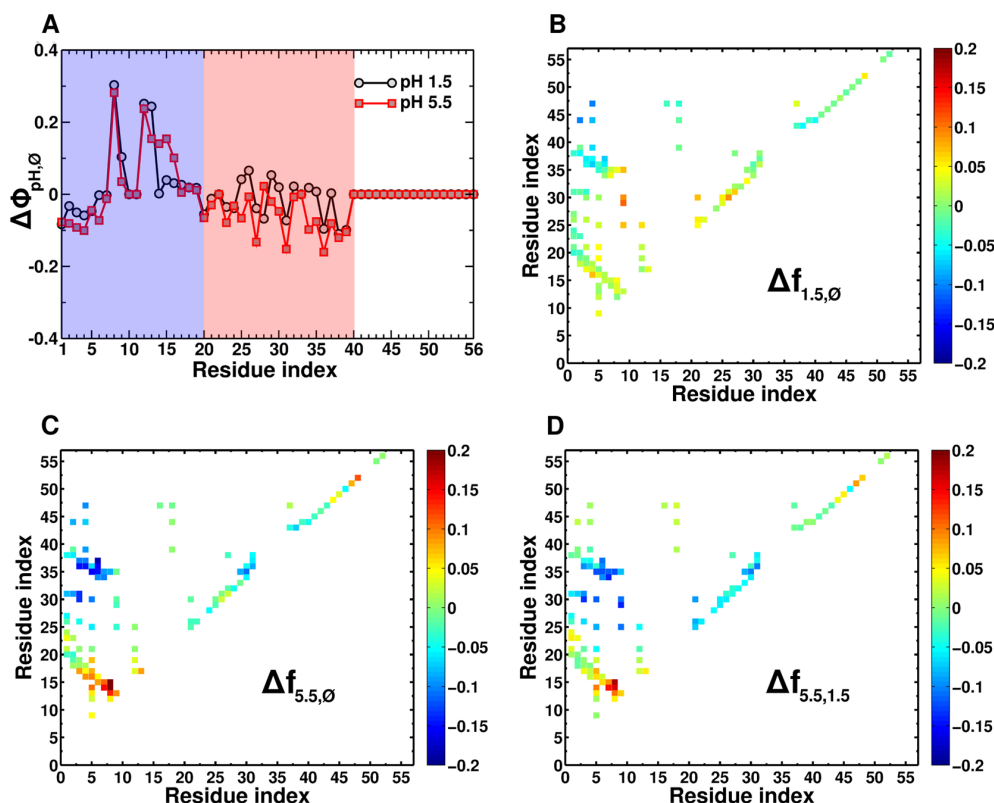


Figure 5. (A) $\Delta\Phi_{ij}^k$ for each residue k : black line and circles, for pH 1.5; red squares, for pH 5.5. (B–D) $\Delta f_{ij}^{k,l}$ for each pair of residues (k, l). The color bar represents the $\Delta f_{ij}^{k,l}$ value for each pair of native contacts between residues k and l .

negative charges in pH 1.5. The most obvious differences in the region near to the N-terminal occur in the cases of LYS-14 and LYS-15 that have a higher $\Delta\Phi_{pH,0}$ for pH 5.5 than for pH 1.5. These differences occur due to the higher frequency of contact formation that these positively charged residues make with the negatively charged ASP-8 in pH 5.5 (see Figure 5D). The increase in this contact frequency also explains the peak of $\Delta\Phi_{5.5,0}$ that ASP-8 has at this pH. However, the value of $\Delta\Phi_{1.5,0}$ for the uncharged ASP-8 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 ASP-8 makes with LEU-35. 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 $\Delta\Phi_{pH,0}$ of residues 20–39 are slightly higher in pH 1.5 than in pH 5.5.

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 C_α 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 of 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 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. $\Delta\Phi_{pH,0}$ was used to better understand how the charges affect the TS. From residues 1 to 20, $\Delta\Phi_{pH,0}$ tends to increase and the most charged region of the protein has the highest increase in $\Delta\Phi_{pH,0}$ 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.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00399.

Comparison between the experimental and computational pK_a values of the 45-residue binding domain of 2-

oxoglutarate dehydrogenase multienzyme complex BBL (PDB ID 1W4H) and hen egg-white lysozyme HEWL (PDB ID 2LZT) (Table S1), titration curves fitted by Henderson–Hasselbalch equation^{5,14} with the ionization degree α as a function of pH for each acid residue of NTL9 in ionic strength 100 mM (Figure S1), free energy profiles $F(Q)$ from NTL9 in function of native contacts Q with the free energy barrier of unfolding ΔF_{unf} shown (Figure S2), and values of ionization degree for each acid residue of NTL9 calculated within 10^5 steps to show the convergence of the charges (Figure S3) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: bleite@sjrp.unesp.br. Phone: +55 (17) 3221-2240.

Author Contributions

[†]V.G.C. and V.M.d.O. contributed equally to this work.

Funding

This work was financially supported by PROPe–UNESP and CNPQ (Grants 442352/2014-0 and 2532/002/14-PROPe/CDC). V.B.P.L. was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPEPS; Grant 2014/06862-7). V.B.P.L., L.C.O., and V.M.d.O. were supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ). V.G.C. was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Margaret Cheung and José Nelson Onuchic for helpful discussions. We also thank the Center for Scientific Computing (NCC/GridUNESP) of São Paulo State University (UNESP) for computational resources.

REFERENCES

- (1) Cho, J.-H.; Sato, S.; Horng, J.-C.; Anil, B.; Raleigh, D. P. *Arch. Biochem. Biophys.* **2008**, *469*, 20–28.
- (2) Azia, A.; Levy, Y. *J. Mol. Biol.* **2009**, *393*, 527–542.
- (3) Zarrine-Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 858–870.
- (4) Davis, S. J.; Davies, E. A.; Tucknott, M. G.; Jones, E. Y.; van der Merwe, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5490–5494.
- (5) Givaty, O.; Levy, Y. *J. Mol. Biol.* **2009**, *385*, 1087–1097.
- (6) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2601–2606.
- (7) Su, J. G.; Chen, W. Z.; Wang, C. X. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2157–2169.
- (8) Ganguly, D.; Zhang, W.; Chen, J. *PLoS Comput. Biol.* **2013**, *9*, e1003363.
- (9) Lindman, S.; Xue, W.-F.; Szczepankiewicz, O.; Bauer, M. C.; Nilsson, H.; Linse, S. *Biophys. J.* **2006**, *90*, 2911–2921.
- (10) Stewart, A. K.; Kurschat, C. E.; Alper, S. L. *Pflügers Arch.* **2007**, *454*, 373–384.
- (11) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
- (12) Haq, I. U.; Akram, F.; Khan, M. A.; Hussain, Z.; Nawaz, A.; Iqbal, K.; Shah, A. J. *World J. Microbiol. Biotechnol.* **2015**, *31*, 1699–1710.
- (13) Kuhlman, B.; Luisi, D. L.; Young, P.; Raleigh, D. P. *Biochemistry* **1999**, *38*, 4896–4903.
- (14) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213–238.
- (15) Khandogin, J.; Chen, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18546–18550.
- (16) Ahlstrom, L. S.; Law, S. M.; Dickson, A.; Brooks, C. L., III. *J. Mol. Biol.* **2015**, *427*, 1670–1680.
- (17) Tripathi, S.; Garcia, A. E.; Makhatadze, G. I. *J. Phys. Chem. B* **2015**, *119*, 13103–13112.
- (18) Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89*, 141–157.
- (19) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, USA, 2000.
- (20) de Oliveira, V. M.; de Carvalho, S. J. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2014**, *37*, 75.
- (21) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937–953.
- (22) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins: Struct., Funct., Genet.* **2009**, *75*, 430–441.
- (23) Contessoto, V. G.; Lima, D. T.; Oliveira, R. J.; Bruni, A. T.; Chahine, J.; Leite, V. B. P. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1727–1737.
- (24) de Mendonca, M. R.; Rizzi, L. G.; Contessoto, V.; Leite, V. B. P.; Alves, N. A. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 119–129.
- (25) Luisi, D. L.; Raleigh, D. P. *J. Mol. Biol.* **2000**, *299*, 1091–1100.
- (26) Luisi, D. L.; Kuhlman, B.; Sideras, K.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1999**, *289*, 167–174.
- (27) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536–540.
- (28) Kuhlman, B.; Luisi, D. L.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1998**, *284*, 1661–1670.
- (29) Kuhlman, B.; Boice, J. A.; Fairman, R.; Raleigh, D. P. *Biochemistry* **1998**, *37*, 1025–1032.
- (30) Anil, B.; Li, Y.; Cho, J.-H.; Raleigh, D. P. *Biochemistry* **2006**, *45*, 10110–10116.
- (31) Cho, J.-H.; Sato, S.; Raleigh, D. P. *J. Mol. Biol.* **2004**, *338*, 827–837.
- (32) Cho, J.-H.; Meng, W.; Sato, S.; Kim, E. Y.; Schindelin, H.; Raleigh, D. P. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12079–12084.
- (33) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 1319–1331.
- (34) Shen, J. K. *Biophys. J.* **2010**, *99*, 924–932.
- (35) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res.* **2010**, *38*, W657–W661.
- (36) Sun, Y.; Ming, D. *PLoS One* **2014**, *9*, e87719.
- (37) Taketomi, H.; Ueda, Y.; GO, N. *Int. J. Pept. Protein Res.* **1975**, *7*, 445–459.
- (38) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E259–E266.
- (39) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. *Bioinformatics* **1999**, *15*, 327–332.
- (40) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171–180.
- (41) Ullner, M.; Woodward, C. E.; Jönsson, B. *J. Chem. Phys.* **1996**, *105*, 2056–2065.
- (42) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- (43) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- (44) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163.
- (45) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (46) Limbach, H. J.; Arnold, A.; Mann, B. A.; Holm, C. *Comput. Phys. Commun.* **2006**, *174*, 704–727.
- (47) Arnold, A.; Lenz, O.; Kesselheim, S.; Weeber, R.; Fahrenberger, F.; Roehm, D.; Košovan, P.; Holm, C. In *Meshfree Methods for Partial Differential Equations VI*; Griebel, M.; Schweitzer, M. A., Eds.; Lecture Notes in Computational Science and Engineering 89; Springer: Berlin, Heidelberg, 2013; pp 1–23, DOI: [10.1007/978-3-642-32979-1_1](https://doi.org/10.1007/978-3-642-32979-1_1).
- (48) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635.
- (49) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

- (50) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
- (51) Bashford, D. *Front. Biosci., Landmark Ed.* **2004**, *9*, 1082–1099.
- (52) Warshel, A.; Sharma, P. A.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1647–1676.
- (53) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed.; W. H. Freeman: New York, 1985.
- (54) Fersht, A. R. *FEBS Lett.* **1993**, *325*, 5–16.
- (55) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1995**, *5*, 79–84.
- (56) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7976–7981.
- (57) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661–1671.
- (58) Noel, J. K.; Onuchic, J. N. In *Computational Modeling of Biological Systems*; Dokholyan, N., Ed.; Springer: New York, 2012; pp 31–54, DOI: [10.1007/978-1-4614-2146-7_2](https://doi.org/10.1007/978-1-4614-2146-7_2).
- (59) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619–1620.
- (60) Plotkin, S. S.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6509–6514.
- (61) Chavez, L. L.; Onuchic, J. N.; Clementi, C. J. *Am. Chem. Soc.* **2004**, *126*, 8426–8432.
- (62) Lammert, H.; Noel, J. K.; Onuchic, J. N. *PLoS Comput. Biol.* **2012**, *8*, e1002776.
- (63) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413–3419.
- (64) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. *J. Am. Chem. Soc.* **2012**, *134*, 979–987.
- (65) Chen, T.; Song, J.; Chan, H. S. *Curr. Opin. Struct. Biol.* **2015**, *30*, 32–42.
- (66) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.
- (67) Weikl, T. R.; Dill, K. A. *J. Mol. Biol.* **2003**, *329*, 585–598.
- (68) Cho, J.-H.; Raleigh, D. P. *J. Mol. Biol.* **2006**, *359*, 1437–1446.

Anexo II - Material suplementar do Anexo I

Supporting Information – 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,[†] and Vitor B. P. Leite^{*,†}

Instituto de Biociências, Letras e Ciências Exatas

UNESP - Univ. Estadual Paulista, Departamento de Física

São José do Rio Preto, SP, Brazil

E-mail: vleite@sjrp.unesp.br

^{*}To whom correspondence should be addressed

[†]Universidade Estadual Paulista

[‡]These authors contributed equally to this work

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 ($\text{SASA} \sim 13 \text{ \AA}^2$) and its self-solvation energy is low. In addition, the distance between the C_α s of *GLU35* and the closest charged residue, *LYS33*, is about 5.7 $\text{\AA}$, whereas the distance between the *GLU35* carboxyl group and the *LYS33* ammonium group is 13.2 $\text{\AA}$. Considering that the electrostatic potential is calculated using a screened Coulomb with a 10 $\text{\AA}$ 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 ($\text{SASA} \sim 60 \text{ \AA}^2$ and $\sim 73 \text{ \AA}^2$, respectively). The C_α of *ASP48* is far from *ARG61* C_α , about 9.4 $\text{\AA}$, while the *ASP48* carboxyl group is almost three times closer to the *ARG61* ammonium group ($\sim 3.2 \text{ \AA}$). In the case of *ASP87*, *HIS15* is the closest charged residue, and the carboxyl and ammonium group are distant about 11.0 $\text{\AA}$. The distance of C_α s, however, is about 4.9 $\text{\AA}$.

ASP66 is a special case, because, even having $\text{SASA} 7 \text{ \AA}^2$, 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 $\text{\AA}$ and 3.7 $\text{\AA}$, respectively. These distances are smaller than the C_α s distances (11.2 $\text{\AA}$ and 5.5 $\text{\AA}$, 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_α s is $\sim 5.9 \text{ \AA}$ and between the

Table S1: 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

^aThe experimental pK_a values for the acidic residues measured by Arbely et al.¹ and Webb et al.;²

^bThe computational pK_a values fitting the individual titration curves with the Henderson-Hasselbalch equation.

carboxyl and ammonium group is ~ 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.^{3,4}

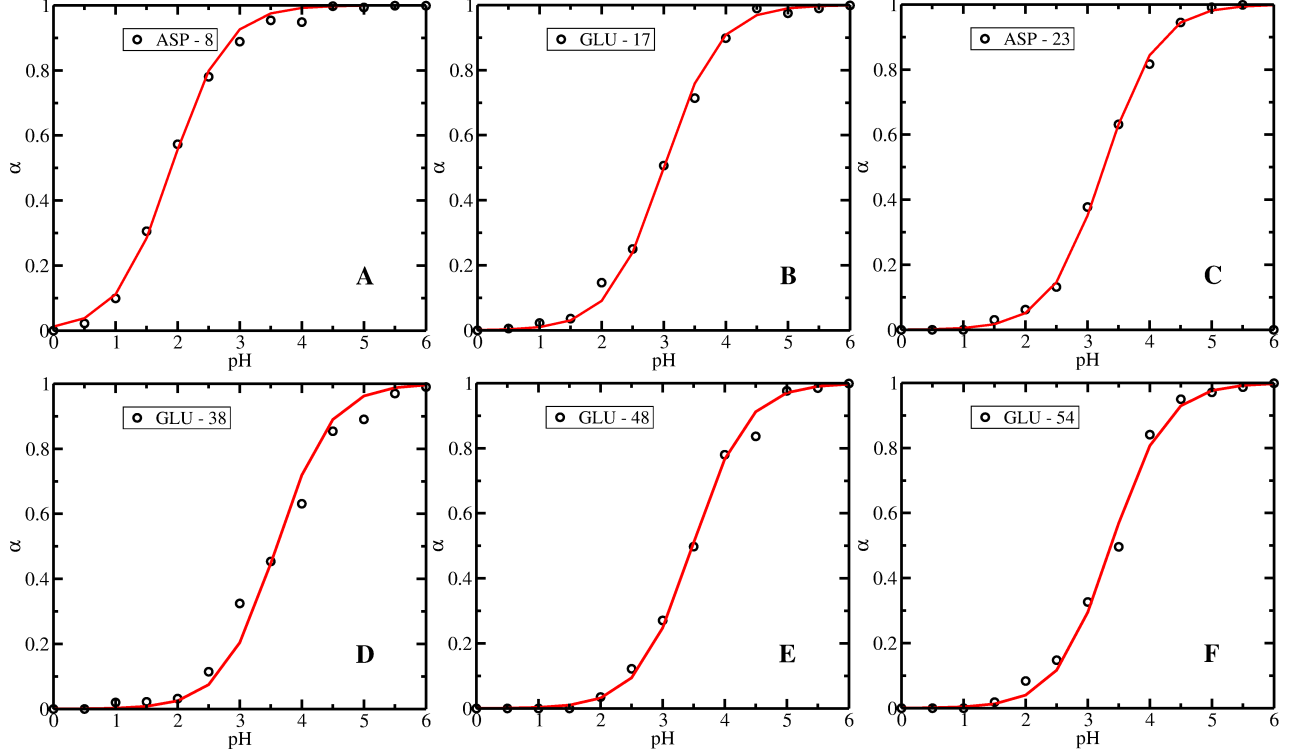


Figure S1: Titration curves: Ionization degree α as a function of pH for each acid residue of NTL9 in ionic strength 100 mM. Red curves are the fit of the Henderson-Hasselbalch equation. (A): Asp8; (B): Glu17; (C): Asp23; (D): Glu38; (E): Glu48 and (F): Glu54.

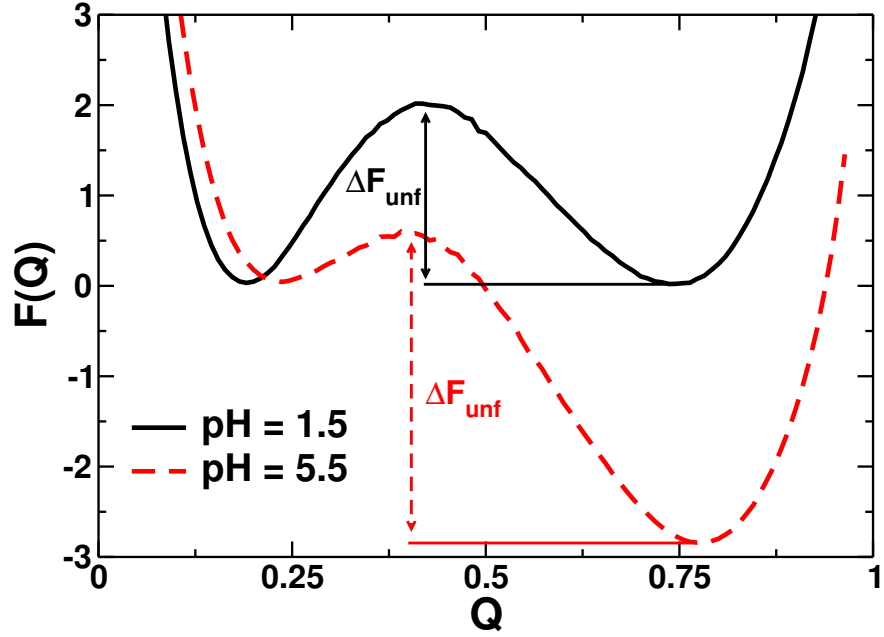


Figure S2: Free energy profiles $F(Q)$ for NTL9 as function of native contacts Q . The black straight line represents the curve for pH 1.5. The red dashed line represents the curve for pH 5.5. The temperature used in both curves are the same evidencing the stabilization of the protein native state in pH 5.5 compared with pH 1.5. ΔF_{unf} is the free energy barrier of unfolding.⁵

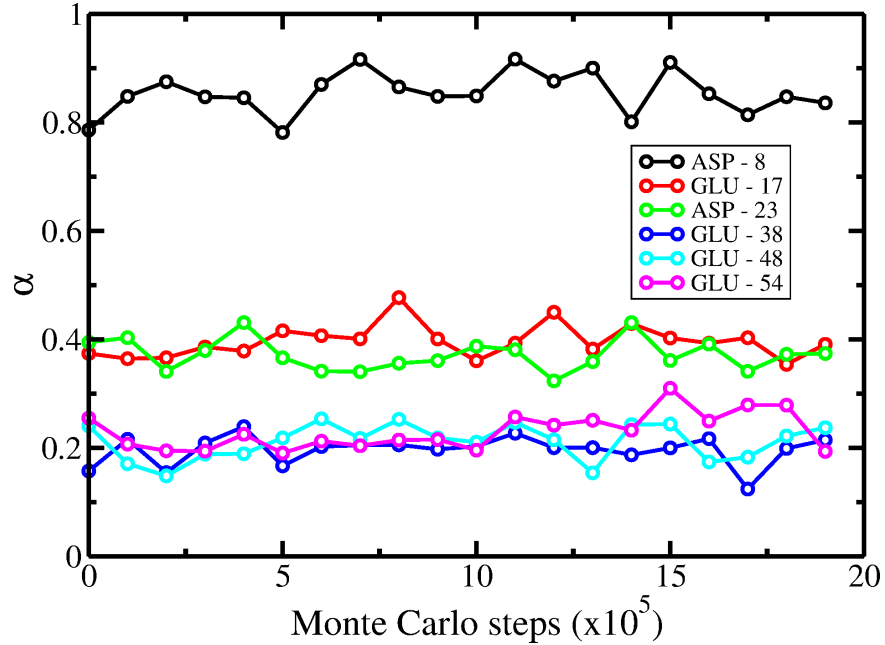


Figure S3: Values of ionization degree for each acid residue of NTL9 calculated within 10^5 steps. These results show a good convergence of the charges. Ionic strength was 100 mM and pH 3.0.

References

- (1) Arbely, E.; Rutherford, T. J.; Neuweiler, H.; Sharpe, T. D.; Ferguson, N.; Fersht, A. R. *J. Mol. Biol.* **2010**, *403*, 313–327.
- (2) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O’Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 685–702.
- (3) Nielsen, J. E.; Vriend, G. *Proteins: Struct., Funct., Bioinf.* **2001**, *43*, 403–412.
- (4) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.
- (5) Luisi, D. L.; Raleigh, D. P. *J. Mol. Biol.* **2000**, *299*, 1091–1100.

Anexo III - Artigo CpCspA submetido

Influence of electrostatic forces on the secondary structure and stability of cold shock protein A from *Corynebacterium pseudotuberculosis*

Monika A. Coronado^{1,2}, Icaro P. Caruso², Vinícius M. de Oliveira², Vinícius G. Contessoto², Vitor B.P. Leite², Raghuvir K. Arni^{1,2}, Raphael J. Eberle^{1,2#}

¹Multiuser Center for Biomolecular Innovation, Universidade Estadual Paulista (UNESP), São Jose do Rio Preto-SP, 15054-000, Brazil.

²Department of Physics, Universidade Estadual Paulista (UNESP), São Jose do Rio Preto-SP, 15054-000, Brazil.

corresponding author:

eberleraphael@gmail.com

Phone: +55-17-32212707 fax +55-17-32212247

Centro Multiusuário de Inovação Biomolecular e Departamento de Física,
Universidade Estadual Paulista (UNESP), Rua Cristóvão Colombo 2265
São Jose do Rio Preto-SP, 15054-000, Brazil.

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.

Abbreviations

CLA, caseous lymphadenitis; CD, circular dichroism spectroscopy; DSC, differential scanning calorimetry; T_m , melting temperature; heat capacity, C_v ; folding transition temperature, T_F ; N, folded state; U, unfolded state; I, intermediate state.

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 (β 1- β 2- β 3) and a two stranded C-terminal sheet (β 4- β 5). 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.

Materials and methods

***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].

Recombinant expression and purification of the CpCspA

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

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 µM. 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 K₂HPO₄/KH₂PO₄ for the pH range 8.0 to 6.0. Results are expressed as molar ellipticity $[\theta]$ (mdeg cm² dmol⁻¹) according to the equation:

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

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].

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 µM (1 mg/ml), the buffer for the measurements contained 50 mM K₂HPO₄/KH₂PO₄. 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:

$$dT_m/d(\ln[a]) = -\Delta n(R \cdot T_m^2 / \Delta H(T_m))$$

where T_m 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 $\Delta H(T_m)$ is the transition enthalpy change at the transition temperature T_m [36,37].

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 $C\alpha$ - structure-based model ($C\alpha$ -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:

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

where ΔU_{el} 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, respectively. The intrinsic pK_a 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).

Results and discussion

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 (www.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].

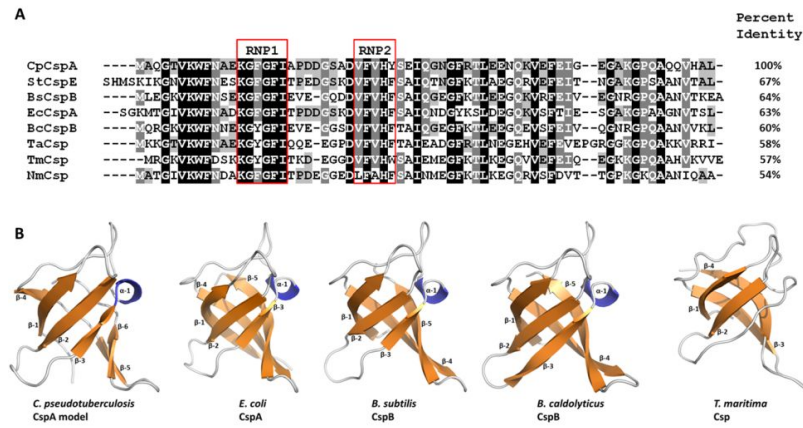


Fig. 1. Sequence and structural comparison of different bacterial Csp. (A) Sequence alignment of CspA from *C. pseudotuberculosis* (GenBank:ADL20127.1), CspE from *S. typhimurium* (GenBank:AE006468.1; PDB:3I2Z), CspB from *B. subtilis* (GenBank:CAA42235.1; PDB:2I5M), CspA from *E. coli* (GenBank:AAA23617.1; 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).

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 Csp 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].

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.

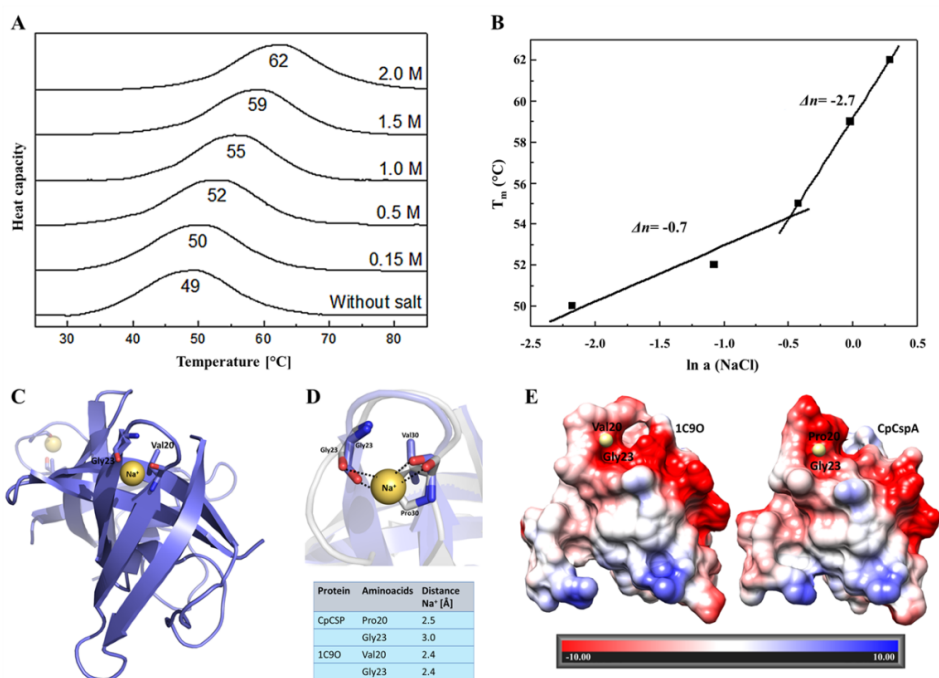


Fig. 2. The influence of NaCl on the thermal stability of CpCspA and identification of a sodium binding site in CpCspA. (A) Determination of T_m under increasing NaCl concentrations. (B) Estimation of the number of Na^+ ions released upon thermal unfolding of CspA. (C) *B. caldolyticus* protein structure (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.

The determination of the number of Na^+ ions released upon thermal unfolding of CspA, resulted in $\Delta n = -0.7$ and was approximated to 1.0 Na^+ ion per Csp molecule (Fig. 2B). Interestingly, at a NaCl concentration higher than 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 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 homology model (Fig. 2E), and form the basis of interaction with the positively charged sodium ion.

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.

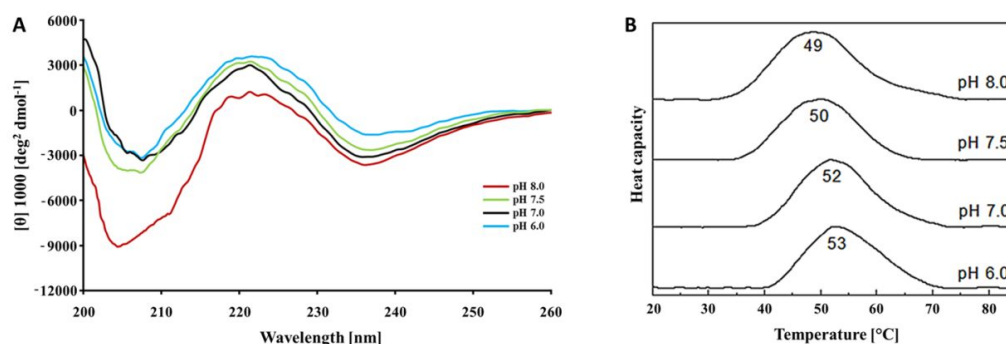


Fig. 3. Biophysical investigation of CpCspA at different pHs. (A) far-UV CD spectrum of CspA from 200 to 260 nm wavelength; pH 8.0 to pH 6.0 dialyzed for 24h. (B) DSC thermogram of CpCspA, the thermogram indicates a reduction in T_m with increasing pH.

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 pairing and the

lengths of the β -strands. These structural elements have an impact on the thermal stability of CspA 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 CspA especially from thermophilic bacteria [57].

The influence of pH 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 shown before by CD and DSC experiments.

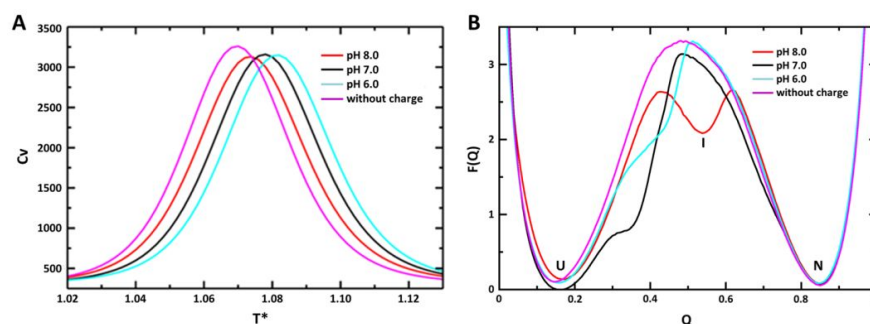


Fig. 4. Thermodynamic properties of CpCspA at different pHs. (A) Heat capacity (C_v) at constant volume as function of temperature in reduced units (T^*) for three different pHs and a system without charges. (B) Free energy curves as function of the reaction coordinate Q (fraction of total native contacts). 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 .

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 $\sim 2\epsilon_c$ in the electrostatic energy between pH 6.0 and 7.0 and $\sim 3\epsilon_c$ 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 $\sim 2\epsilon_c$), principally because the N state at this pH presents only one highly populated electrostatic energy level. All these electrostatic features result in the higher stability of CpCspA at pH 6.0 than at pH 7.0 and 8.0.

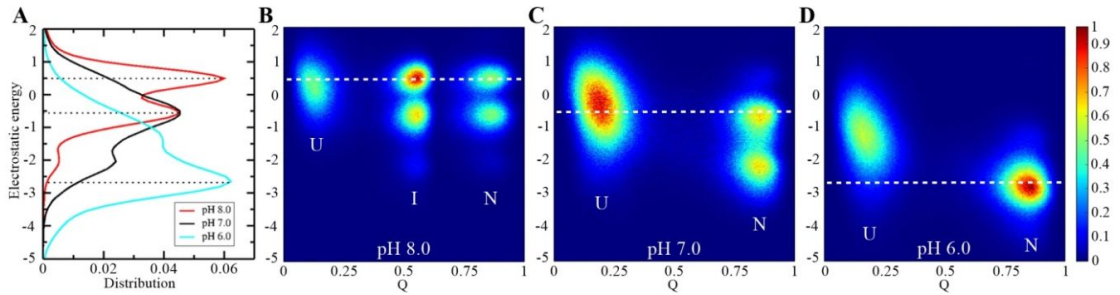


Fig. 5. Electrostatic energy of CpCspA at different pHs. (A) Distribution of the electrostatic energy in the reduced unit ϵ_c for three different pHs. (B,C,D) Two dimensional map of the electrostatic energy as functions of the reaction coordinate Q for pH 8, 7 and 6, respectively. The color map represents the probability distribution of the electrostatic energy and Q , normalized by its highest value. All the systems with $T \sim T_F$.

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.

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 C α 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).

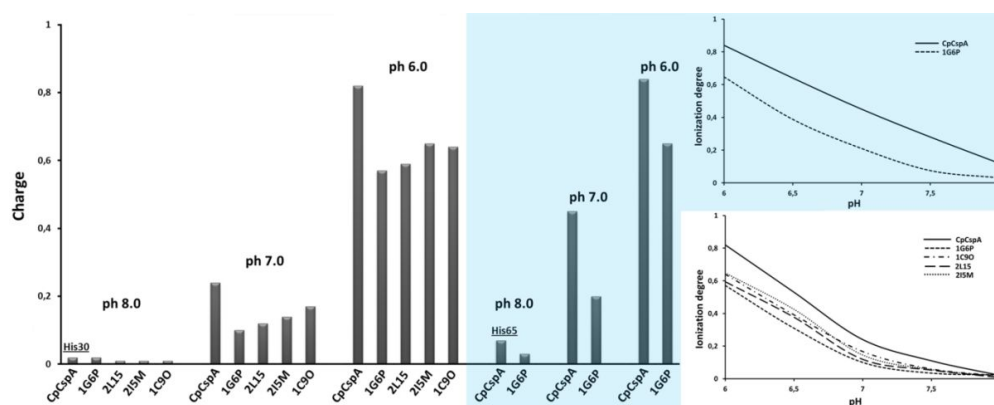


Fig. 6. Charges of the histidine residues of Csp proteins from different species in variant pHs. The changes in the charge of His30 and His65 of CpCspA, His28 and His61 of Csp from *T. maritima* (PDB:1G6P), His33 of CspA from *E. coli* (PDB:2L15), His29 of CspB from *B. subtilis* (PDB:2I5M) and His29 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.

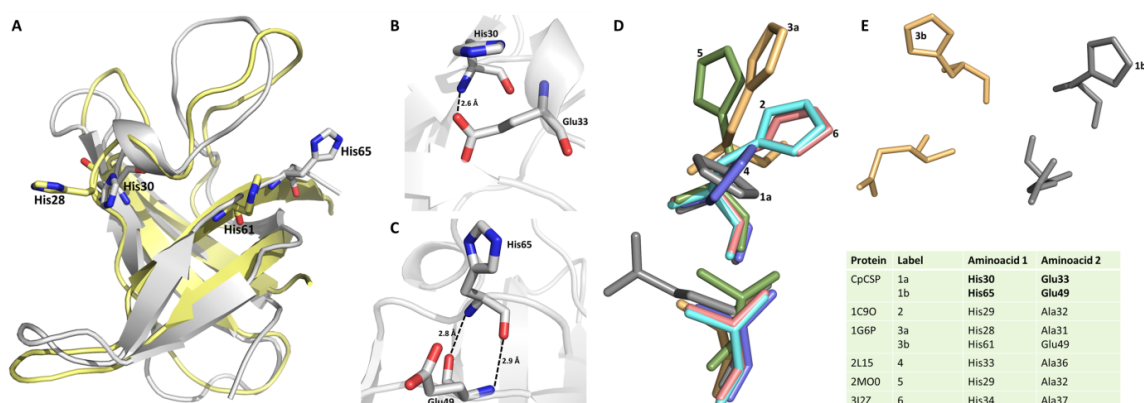


Fig. 7. Locations of the histidine residues in Csp family structures. (A) Structural comparison of the CpCspA homology model (grey) and the Csp of *T. maritima* (1G6P) (yellow), the positions of the histidine residues highlighted as sticks. (B) Detail view of the CpCspA region His30 and Glu33. (C) Detail 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.

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).

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.

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.

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.

Reference

- 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, Wilscheck R, Czisch M, Herrler M, Willmsky 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 maritima*. *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 maritima*. *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.

Supporting Information

Additional supporting information

Table A.1 Secondary structure analysis of the CD-data computed by CDpro.

Table A.2 Charge changes of charged amino acids in CpCspA.

Figure A.1 Charge changes of His35 and His65 of CpCspA as a function of Temperature.

Data A.1 Computational methods, model potential.

Anexo IV - Artigo Blo t 5 submetido

Manuscript

[Click here to download Manuscript Godoi et al. Manuscript.pdf](#) 

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 Godoi Melo Ávila¹, Laura Fontes Tomaz¹, Carlos Roberto Prudencio²,
Karine Cristine de Almeida¹, Deise Aparecida Oliveira Silva¹, Odonório Abrahão
Junior³, Vinícius Martins de Oliveira⁴, Vinícius de Godoi Contessoto⁴, Vitor
Barbanti Pereira Leite⁴, Pamela Aparecida Candido⁵, Ronaldo Junio de Oliveira⁵,
Jair Pereira Cunha-Júnior¹, Ernesto Akio Taketomi^{1*}**

¹Laboratório de Alergia e Imunologia Clínica, Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia, Av. Pará 1720, 38400-902, Uberlândia, MG, Brazil

²Centro de Imunologia, Instituto Adolfo Lutz. Av. Dr. Arnaldo 351, 10º Andar, 01246-902, São Paulo, SP, Brazil.

³Laboratório de Química Computacional e Medicinal, Instituto de Ciências Biológicas e Naturais, Universidade Federal do Triângulo Mineiro, Praça Manoel Terra, 330, 38025-015, Uberaba, MG, Brazil

⁴Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, R. Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, SP, Brazil

⁵Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Av. Dr. Randolfo Borges Junior, 1400, 38064-200, Uberaba, MG, Brazil

* Correspondence to: etaketomi@gmail.com (Ernesto Akio Taketomi)

Laboratory of Allergy and Clinical Immunology
Institute of Biomedical Sciences
Federal University of Uberlândia, Uberlândia, MG, Brasil
Tel:(55) 34 3225 8674

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 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 showed that the electrostatic-stable epitope controls the improvement in antigenicity for binding to human IgG4 isotype. This study suggests 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

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 acid), 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 an approach *in silico* technique 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 and also in thermal molecular stability display different IgE affinity [15,16]. Thus, the protein thermal stability and the

electrostatic interactions of charged residues within IgE-binding epitope have to be taken into account 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 overall of the molecular stabilization. This kind of approach has been used with widely success in previous report demonstrating the 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 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.

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% glycerined 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 allergen, 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, $\alpha 1$, $\alpha 2$ and $\alpha 3$). The $\alpha 3$ helix was shifted between D and $\alpha 1$ domains (D, $\alpha 3$, $\alpha 1$ and $\alpha 2$) 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. 1B).

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

BL21 (DE3) *E. coli* strain (Novagen, USA) cells were transformed with 2.5 μ 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 μ 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 induced 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 NaH₂PO₄, 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 (Qiagen GmbH), previously equilibrated with the binding buffer (50 mM NaH₂PO₄, 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 NaH₂PO₄, 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, Karlsruhe, Germany). The gel was stained with Coomassie blue or electrotransferred to nitrocellulose membranes (0.45 µm, Bio-Rad, Hercules, CA, USA), as previously described [33], using a semi-dry transfer system (Trans-blot SD, Bio-Rad Laboratories Inc.)

In order to confirm the production of the rBlo t 5 allergens, fractions of Ni-affinity chromatography were assayed with mouse monoclonal antibody (mAb) to Blo t 5 (clone 4G9 - Indoor Biotechnologies Inc., Charlottesville, VA). 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% H₂O₂, 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

To confirm the specific reactivity of the rBlo t 5 and mBlo t 5 allergens, nitrocellulose membranes were blotted with rBlo t 5 (2.5 µg/slot) or mBlo t 5 (2.5 µg/slot) on vacuum device (Bio-dot®SF apparatus, Bio-Rad, Hercules, CA, USA). As controls, total Bt extract (10 µg/slot; FDA Allergenic) and bovine serum albumin (5 µg/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% H₂O₂.

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[−] 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% H₂O₂.

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% H₂O₂. The absorbance was determined in a plate reader at 405 nm. Antibody titers were reported as ELISA index (*EI*) and determined as follows: $EI = \text{absorbance of test sample} / \text{cutoff}$, 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 Gromacs 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 I 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 μ s (10 times 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 and covering the folding temperature ($T_f \sim 1.0$). 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

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

where e is the unit charge, A_{ij} , B_{ij} and C_{ij} 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, the C_{ij} term, the ionic [22,23,52]. SA_{ij} 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 $a = b + 1.4 \text{ \AA}$ (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].

$$\Delta U = \Delta U^{sl} \pm K_b t \ln(10)(Ph - PK_a) \quad (2)$$

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 pK_a 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.

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 ($P < 0.05$).

Sequence and structure of the recombinant rBlo t 5 and the designed mBlo t 5 allergens

The sequence and structure of the major IgE binding protein rBlo t 5 and the modified mBlo t 5 are shown in Figure 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 $\alpha 3$ helix of the rBlo t 5 (indicated by the red arrow) are shown in Figure 1A. In Figure 1A, 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 ($\alpha 1$, $\alpha 2$, $\alpha 3$), forming a helix bundle, and a N-terminal disordered domain (D) (Fig. 1B, left panel). A similar structure was modeled for mBlo t 5, with the three α -helices rearranged as in Figure 1A ($\alpha 3$, $\alpha 1$, $\alpha 2$) and the N-terminal (D) slightly crooked, also in a disordered state (Fig. 1B, 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.

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-rBlo t 5* 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-mBlo t 5* 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. When this partially purified sample was probed in immunoblot with monoclonal anti-Blo t 5 antibody (Fig. 2B, lane 3), 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). When the change of antibody isotype reactivity to mBlo t 5 in relation to rBlo t 5 was analyzed, 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). Figure 6 shows the structure based model (SBM) simulations using the C_α and all-atom models for the thermodynamic and kinetic analysis, respectively. The C_α -model results in Figure 6A presents 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 ($Q \sim 0.8$) and unfolded ($Q \sim 0.3$) populations, in Fig. 6A, separated by the transition state activation barrier at $Q \sim 0.5$. 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 analysis was performed by using the all-atom representation of the proteins in order to include the amino acid side chain information to the SBM simulations. Figure 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 using the TKSA-MC simulations. In Fig. 7A, at the rBlo t 5 binding epitope region (from residue 76 to 91), the average of the charge-charge electrostatic energy interaction of the residues Glu76, Asp81, Glu86 and Glu91 have a positive energy value, 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 the 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. 7B). 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. 7C). This electrostatic energy stabilization in mBlo t 5 is associated with the inclusion of a positive residue in its epitope region (Fig. 7D), which causes a more homogeneous charge distribution and a more positive electrostatic potential when compared to the rBlo t 5 binding epitope. 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 point mutations in rBlo t 5 (data from the NMR experiment [20]) and the average electrostatic energy value ($\langle W_i \rangle$) 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 epitope's 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.

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 to this [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 the rearrangement designed in the natural

sequence. This finding was also noted in the mBlo t 5 three-dimensional structure, which share a similar IgE binding epitope within rBlo t 5, although slight differences are displayed in their amino acid sequence.

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.

When analyzing the reactivity of antibody isotypes to the total Bt extract and recombinant allergens from Blo t 5 by ELISA, 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 IgG1 was not significant, although there was a great individual variation in the antibody reactivity. In contrast, the median levels of IgG4 were significantly 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. Considering as significant a reactivity change $\geq 20\%$ as previously proposed [20], this IgG4 reactivity change occurred in 50% of samples in relation to 35% for IgG1 and 28% for IgE reactivity. 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].

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 very computer-time consuming, which requires 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 produce a decrease in IgE binding and also reduced cutaneous reaction in the skin prick test. 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.

This is a worthy validation to our model as the experimental and computer simulation converged in a good 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 design new candidates in allergen-specific immunotherapy studies.

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.

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.

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 Blo t 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.

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 pteronyssinus* 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, Lallemand 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, Degraive 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.34.040204.144447
40. Socci ND, Onuchic JN, Wolynes PG. Diffusive dynamics of the reaction coordinate for protein 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. Matthews 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.
65. 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.
66. 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.

67. 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.
68. 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.
69. 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.
70. 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.
71. 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.
72. 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.
73. 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.
74. 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.
75. Caravella JA, Wang D, Glaser SM, Lugovskoy A. Structure-Guided design of antibodies. *Curr Comput Aided Drug Des*. 2010;6: 128–138.
76. 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.
77. 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.
78. Cheng X, Ivanov I. Molecular Dynamics. In: Reisfeld B, Mayeno AN, editors. *Computational Toxicology SE* - 11. Humana Press; 2012. pp. 243–285.

79. 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.
80. 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.

Figure legends

Figure 1: (A) 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 $\alpha 2$). (B) On the left, rBlo t 5 three-dimensional structure (Protein Data Bank code 2JMH, model 1) and, on the right, the modeled mBlo t 5 allergen. The disorder domains are labeled in gray and the α -helices $\alpha 1$, $\alpha 2$ and $\alpha 3$ are colored in green, blue and red, respectively.

Figure 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 µg) and mBlo t 5 (2.5 µg) 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 µg) and *B. tropicalis* extract (Bt total; 10 µg) were also blotted.

Figure 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 Blo t 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 ($ELI > 1.2$). 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: 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 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 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 $Q \sim 0.3$ and at the folded state when at $Q \sim 0.8$. 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.

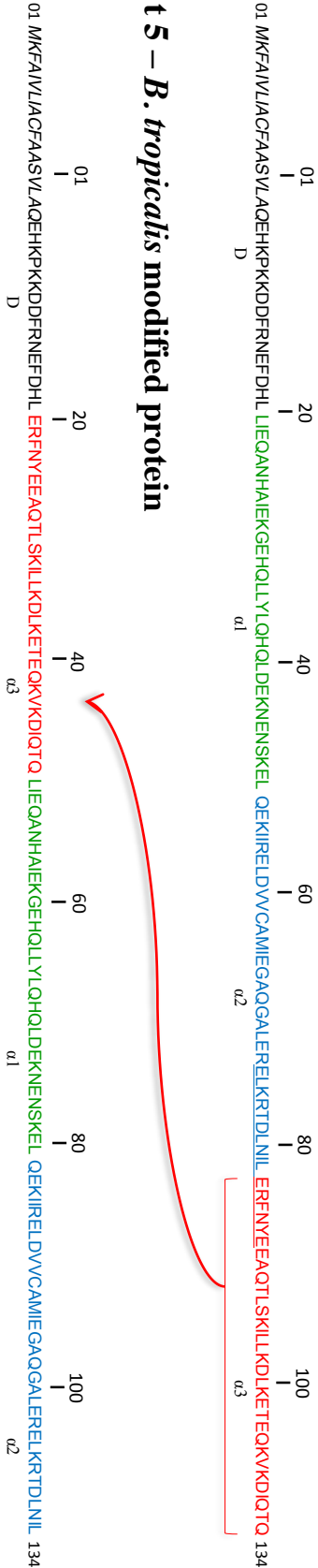
Figure 7: Electrostatic interaction optimization calculated in the protein charged residues by the Tanford-Kirkwood Surface Accessibility model implemented using Monte Carlo simulations (TKSA-MC). Average of the charge-charge electrostatic energy $\langle W_i \rangle$ of each ionizable residue i for **(A)** the recombinant allergen rBlo t 5 and **(C)** the modified mBlo t 5. The charged residues of the binding epitopes of **(A)** rBlo t 5 (E76 to E91) and **(C)** 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 **(B)** 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. 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 **(B)** 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 8: Correlation between theoretical and experimental measurements. Average of the charge-charge electrostatic energy interaction $\langle W_i \rangle$ 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 ($\langle W_i \rangle > 0$) opposing to the minor binding residues (K78 and R87), which are positively charged and showed an favorable electrostatic energy interaction ($\langle W_i \rangle < 0$). 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.

A

rBlo t 5 – *B. tropicalis* major IgE-binding protein



mBlo t 5 – *B. tropicalis* modified protein

B

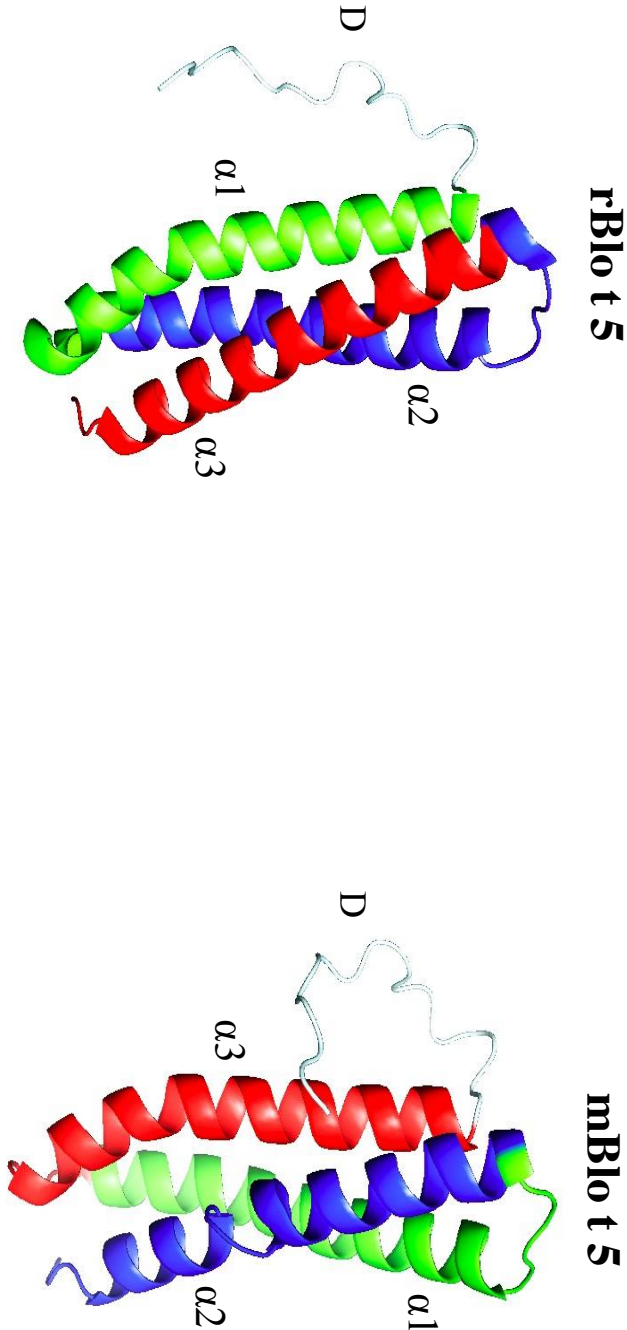


Figure 2

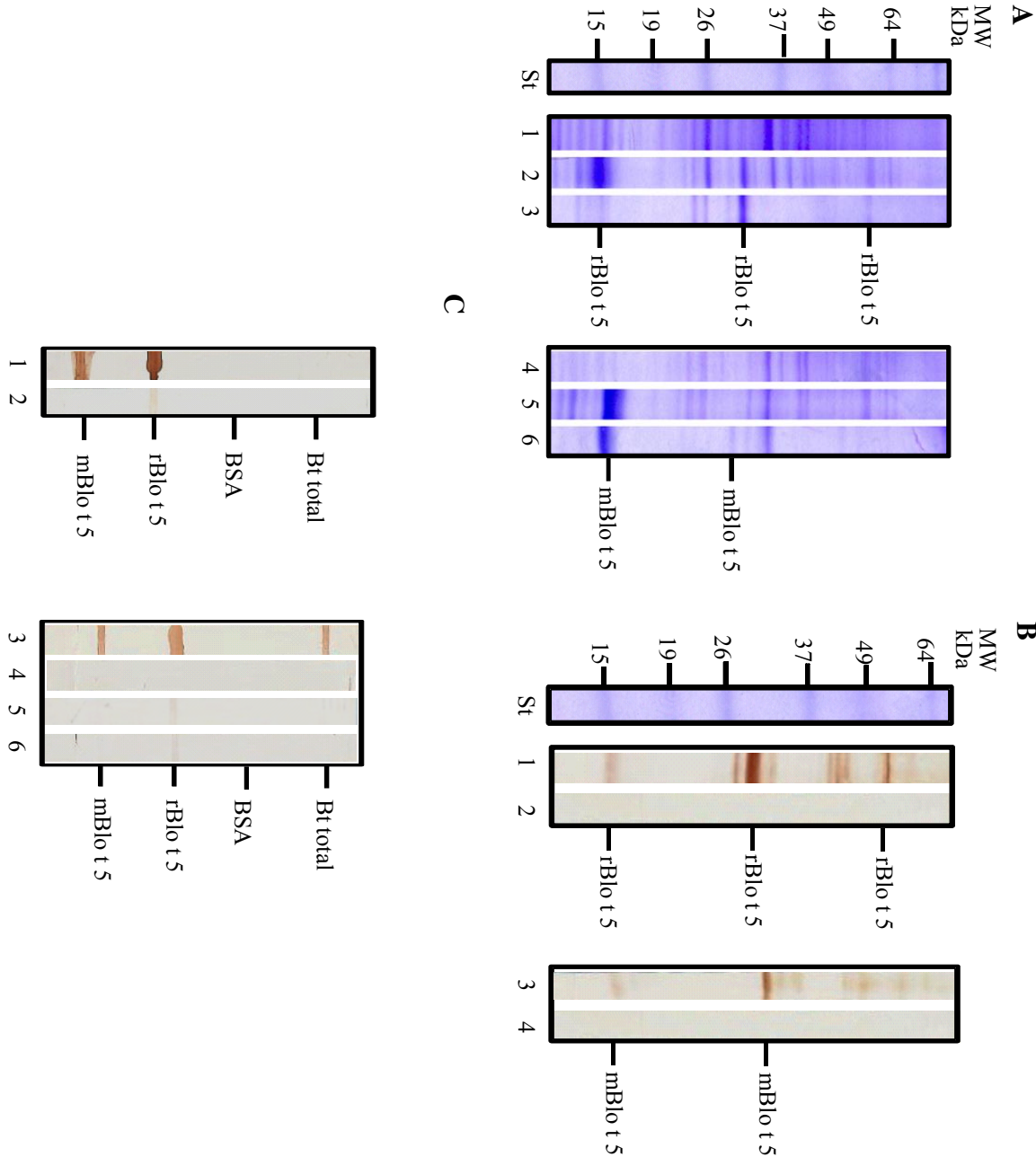


Figure 3

[Click here to download Figure Figure 3.pptx](#)

Figure 3

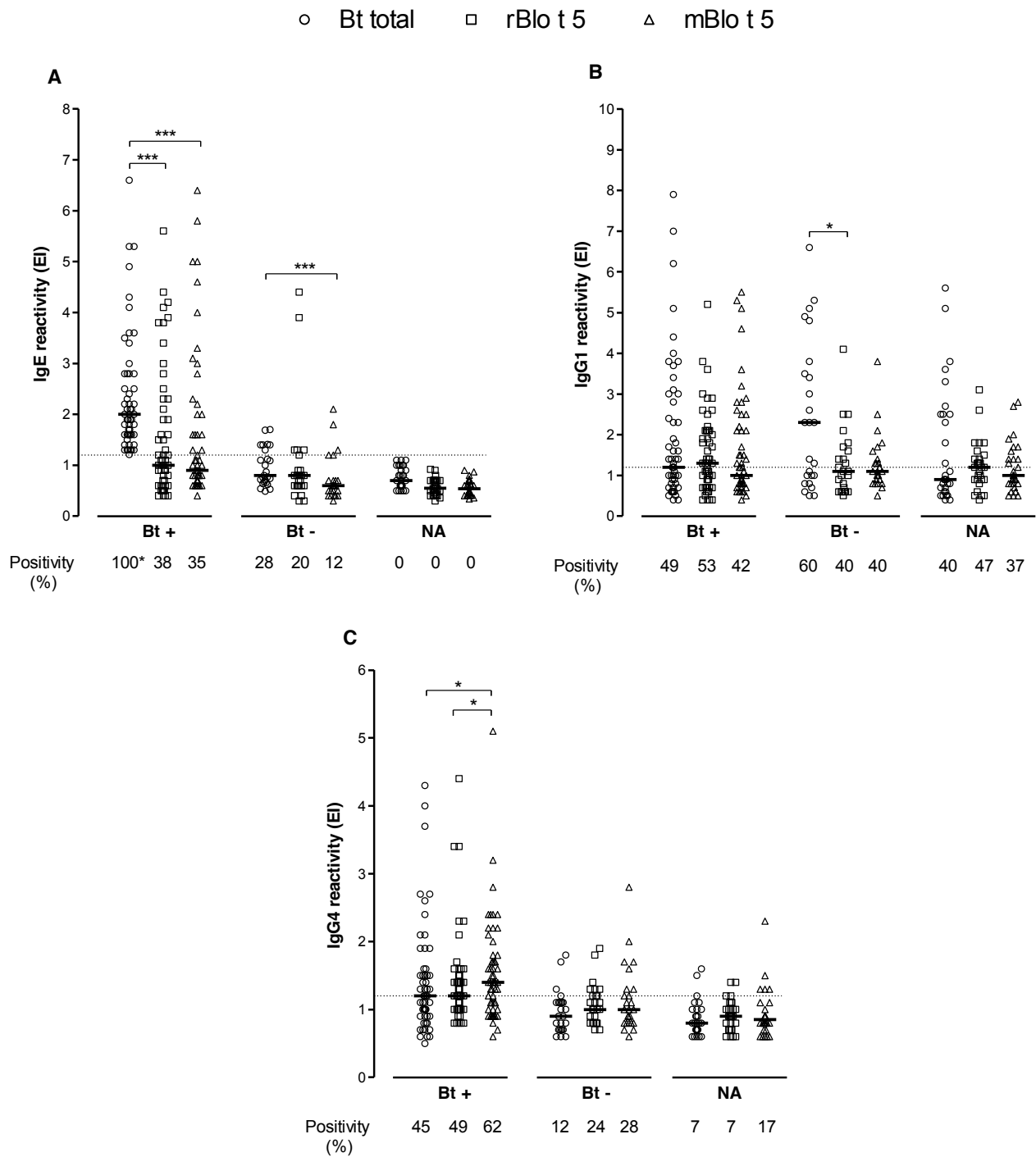


Figure 4

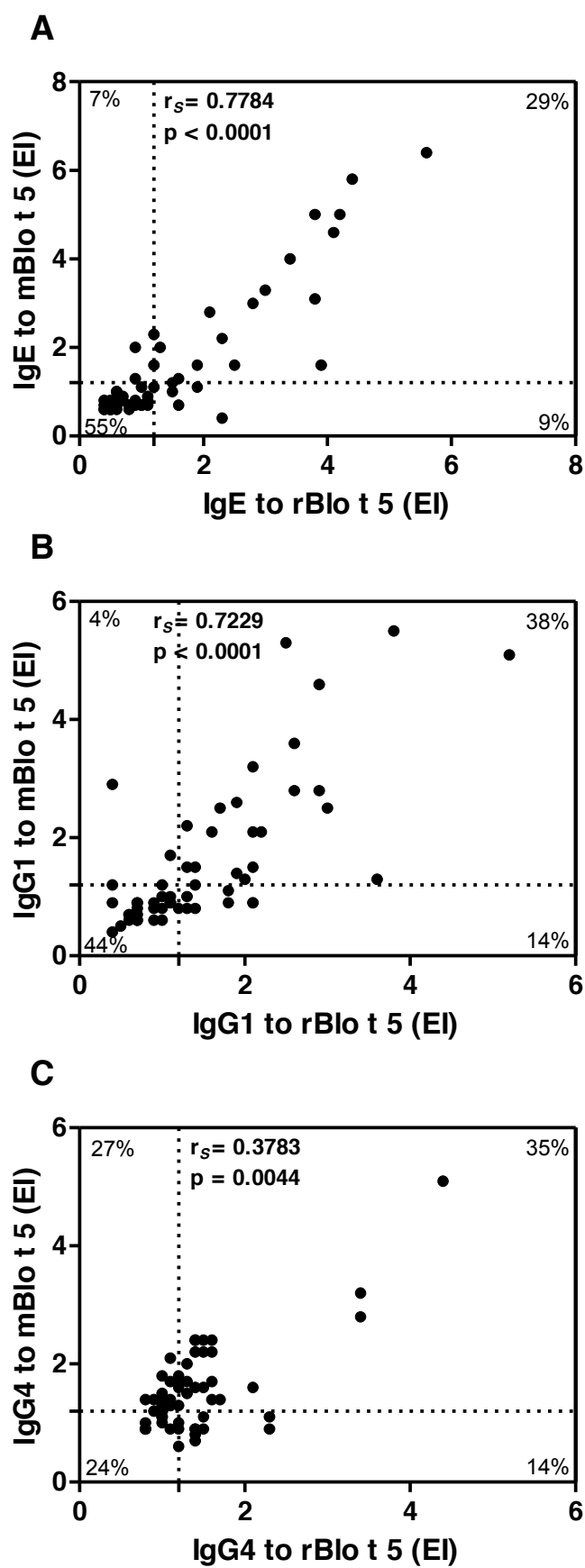


Figure 5

[Click here to download Figure Figure 5.pptx](#)

Figure 5

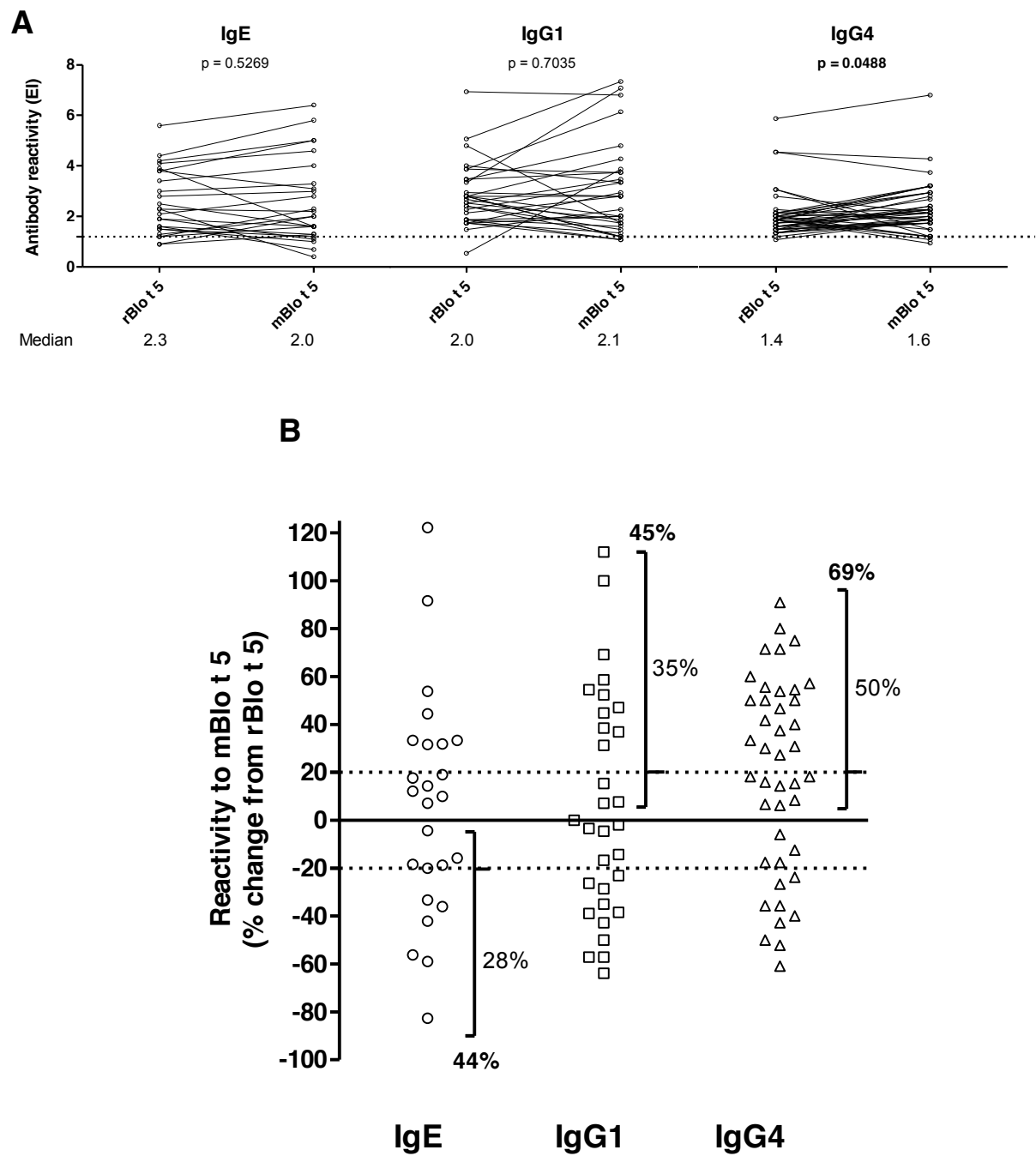


Figure 6

[Click here to download Figure Figure 6.png](#)

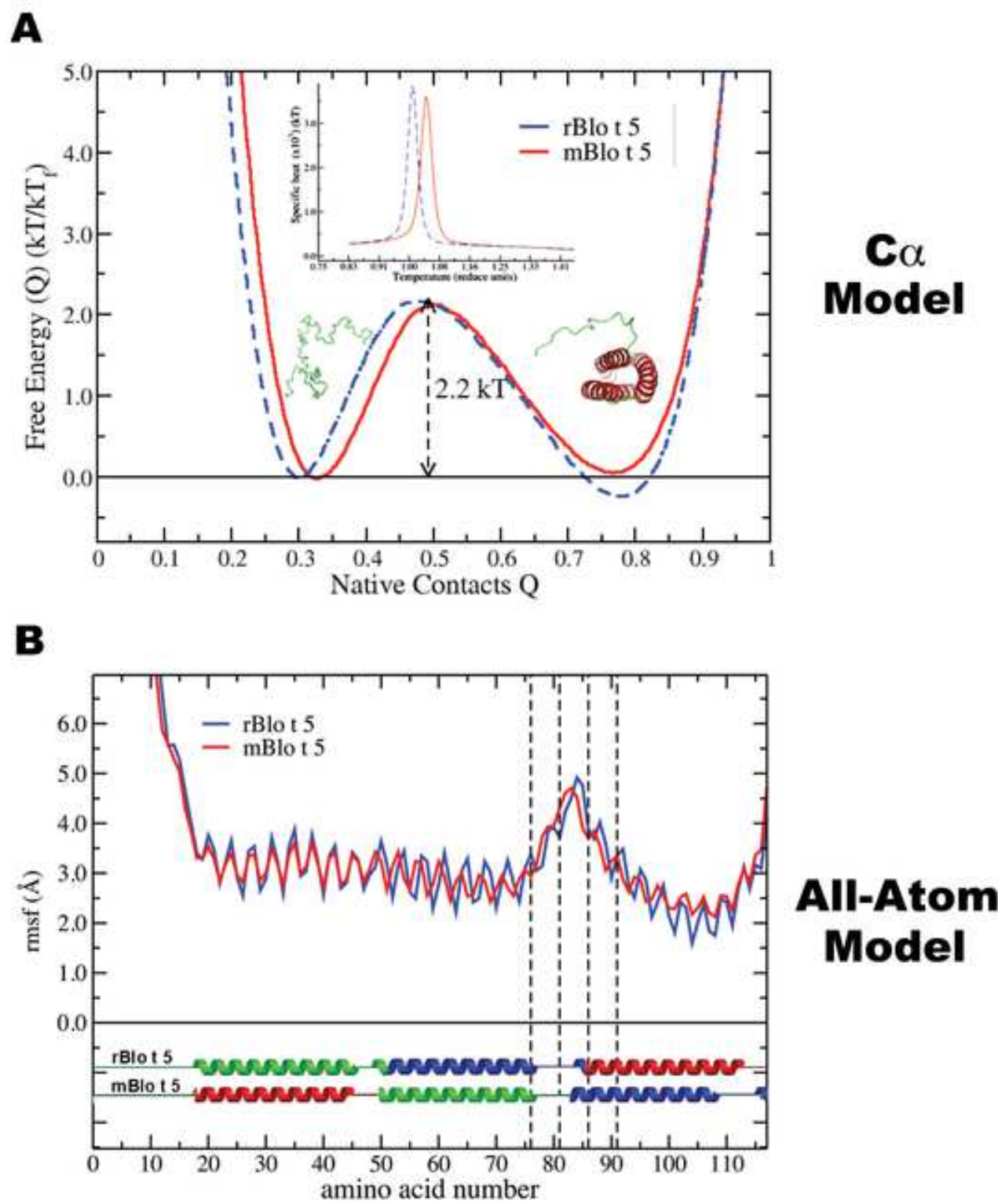


Figure 7

[Click here to download Figure 7.jpg](#)

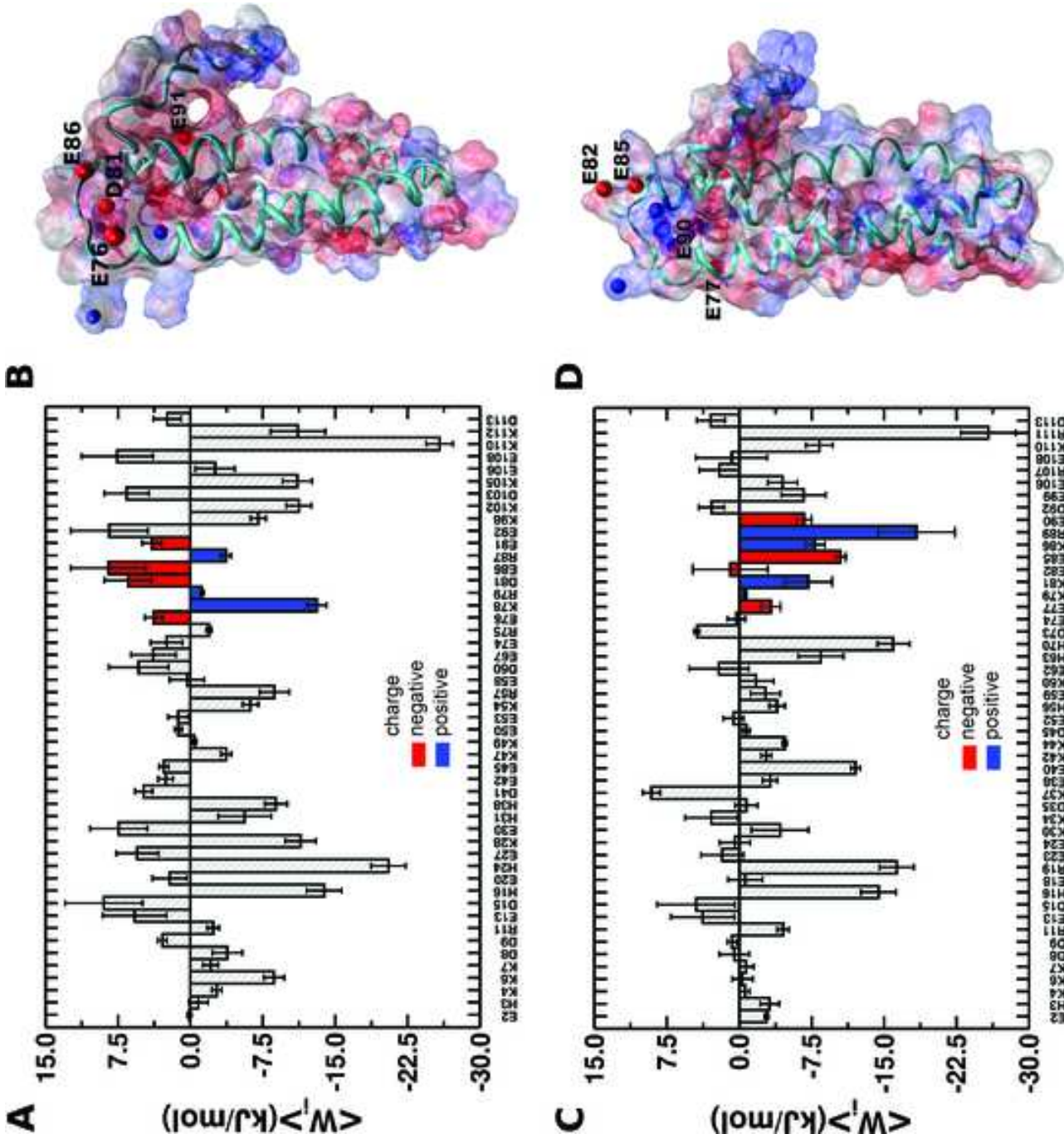


Figure 8

[Click here to download Figure 8.jpg](#)

