

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Mudanças estruturais na proteína príon celular induzidas por  
alteração de pH**

Dissertação de Mestrado

Helen Nathalia Thompson

Porto Alegre

2012

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**Mudanças estruturais na proteína príon celular induzidas por  
alteração de pH**

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre.

Helen Nathalia Thompson

Orientador: Dr. Hubert Karl Stassen

Coorientador: Dr. Paulo Augusto Netz

Porto Alegre. Maio de 2012.

## Declaração de Autoria e Orientação

A presente dissertação foi realizada inteiramente pela autora Helen Nathalia Thompson no período entre março de 2010 e abril de 2012, no Instituto de Química da Universidade Federal do Rio Grande do Sul sob orientação do Professor Doutor Hubert K. Stassen e coorientação do Professor Doutor Paulo A. Netz. A dissertação foi julgada aprovada para a obtenção do título de Mestre em Química pela seguinte banca examinadora:

### Comissão Examinadora:

-----  
Prof. Dr. Maximiliano Segala

-----  
Prof. Dr. Paolo Roberto Livotto

-----  
Prof. Dr. Laurent Emmanuel Dardenne

-----  
Prof. Dr. Hubert K. Stassen

-----  
Prof. Dr. Paulo A. Netz

-----  
Mestranda Helen Nathalia Thompson

*Dedico esta dissertação aos meus pais.*

"(...)

*O alargamento nos singulares últimos cem anos das múltiplas ramificações do conhecimento, tanto em extensão quanto em profundidade, confrontou-nos com um dilema difícil. Sentimos claramente que só agora começamos a adquirir material confiável para reunir tudo o que se sabe em uma só totalidade. Mas, por outro lado, tornou-se quase impossível para uma só mente dominar por completo mais que uma pequena porção especializada desse conhecimento.*

*Não vejo outra saída para esse dilema além de alguns de nós nos aventurarmos a embarcar numa síntese de fatos e teorias, ainda que munidos de conhecimento incompleto e de segunda mão sobre alguns deles, e sob o risco de parecermos tolos.*

(...)"

*Erwin Schrödinger- trecho extraído  
do prefácio do livro "O que é vida?".*

## ***Agradecimento***

*Ao meu orientador, professor Hubert K. Stassen, pelo apoio, confiança, paciência e por ter participado ativamente da confecção do presente trabalho.*

*Ao meu coorientador, professor Paulo A. Netz, por ter me apresentado ao universo dos prions e pelas valiosas discussões realizadas ao longo de todo o desenvolvimento desta dissertação.*

*Ao professor Paolo R. Livotto, por me orientar na disciplina de estágio docente, me dando tranquilidade neste primeiro contato com a docência.*

*Ao professor Paulo F. B. Gonçalves, pela disposição em esclarecer dúvidas recorrentes no estudo de metodologias de química computacional.*

*Ao professor Hermes Luís Neubauer de Amorim, que participou da banca do exame de qualificação da presente dissertação, pelas sugestões que agregaram conhecimento na construção deste trabalho.*

*Aos professores Hermes Luís Neubauer de Amorim, Leandra Franciscato Campo, Paolo R. Livotto, Maximiliano Segala e Laurent Emmanuel Dardenne, por terem aceitado prontamente o convite para participar da banca de avaliação desta dissertação.*

*Aos professores Paolo R. Livotto, Maximiliano Segala e Laurent Emmanuel Dardenne, que foram selecionados para compor a banca de avaliação da presente dissertação, pelas críticas construtivas, discussões e sugestões que agregaram melhorias à versão final deste trabalho.*

*Aos colegas de laboratório, pelo conhecimento compartilhado.*

*Ao meu pai, por me servir de inspiração profissional e por sempre ter incentivado os meus estudos.*

*A minha mãe, eterno porto seguro.*

*Aos meus irmãos, familiares e amigos pela lealdade e suporte nos momentos difíceis.*

*Ao meu namorado, Thiago, que me acompanha desde os primeiros anos da graduação, pelos constantes carinho, incentivo e compreensão.*

*À CAPES, pela concessão da bolsa de mestrado.*

**MUITO OBRIGADA.**

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1 PRÍONS.....	2
1.1.1 A História dos Príons .....	2
1.1.2 Encefalopatias Espongiformes Transmissíveis.....	5
1.1.3 PrP <sup>C</sup> x PrP <sup>SC</sup> .....	7
1.2 HIPÓTESE DE INDUÇÃO CONFORMACIONAL .....	9
1.3 MODELAGEM COMPUTACIONAL.....	10
<b>2. FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>13</b>
2.1 PROTEÍNAS.....	14
2.1.1 Aminoácidos: blocos fundamentais de construção das proteínas .....	14
2.1.2 Classificação Estrutural de Proteínas .....	17
2.1.2.1 Estrutura Primária .....	17
2.1.2.2 Estrutura Tridimensional .....	18
2.1.2.2.1 Estrutura Secundária: .....	19
2.1.2.2.2 Estruturas Terciária e Quaternária .....	21
2.1.3 Dobramento de proteínas e Estabilidade da forma nativa .....	23
2.1.4 Forças que contribuem para a estabilidade de proteínas.....	25
2.2 BANCOS DE DADOS .....	28
2.3 DINÂMICA MOLECULAR .....	28
2.4 CAMPOS DE FORÇA.....	31
2.4.1 ENCAD.....	32
2.4.2 GROMOS .....	34
2.4.3 AMBER .....	36
2.4.4 OPLS.....	38

2.4.5	Comparação dos Campos de Força GROMOS, AMBER, OPLS e ENCAD.....	39
2.5	INTERAÇÕES ELETROSTÁTICAS DE LONGO ALCANCE .....	40
2.5.1	GRF .....	42
2.5.2	SPME.....	42
2.6	MÚLTIPLAS SIMULAÇÕES DE DINÂMICA MOLECULAR .....	43
<b>3.</b>	<b>OBJETIVOS.....</b>	<b>46</b>
3.1	OBJETIVO GERAL .....	47
3.2	OBJETIVOS ESPECÍFICOS.....	47
<b>4.</b>	<b>METODOLOGIA .....</b>	<b>48</b>
4.1	PROTOCOLO DE SIMULAÇÃO.....	49
4.2	ANÁLISES.....	50
<b>5.</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>56</b>
5.1	COMPARAÇÃO DE CAMPOS DE FORÇA.....	57
5.1.1	GROMOS, AMBER e OPLS .....	57
5.1.2	GROMOS e AMBER: estendendo as trajetórias .....	61
5.1.3	Conclusões Parciais .....	66
5.2	INFLUÊNCIA DA FORMA DE PROTONAÇÃO DOS AMINOÁCIDOS.....	67
5.2.1	Investigação dos estados de protonação da Histidina (HisA, HisB e HisH) .....	67
5.2.2	Conclusões Parciais .....	77
5.3	INFLUÊNCIA DE DIFERENTES CONDIÇÕES INICIAIS.....	79
5.3.1	Utilização de diferentes conformações iniciais de RMN.....	79
5.3.2	Randomização das velocidades iniciais .....	83
5.3.3	Conclusões Parciais .....	86
5.4	INVESTIGAÇÃO DAS INTERAÇÕES ELETROSTÁTICAS DE LONGO ALCANCE.....	88
5.4.1	SPME x GRF .....	88



5.4.2	Conclusões Parciais .....	93
<b>6.</b>	<b>CONCLUSÕES FINAIS .....</b>	<b>94</b>
<b>7.</b>	<b>PERSPECTIVAS .....</b>	<b>97</b>
<b>8.</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>99</b>
<b>9.</b>	<b>ANEXOS.....</b>	<b>104</b>
<b>9.1</b>	<b>ANEXO A1: Propriedades e convenções relacionadas com os 20 aminoácidos comuns.....</b>	<b>105</b>
<b>9.2</b>	<b>Anexo A2: Comparação entre os modelos de carga de diferentes campos de força para o aminoácido Serina.....</b>	<b>106</b>

## Lista de Figuras

<b>Figura 1:</b> A imagem mostra quatro mulheres afetadas com uma forma avançada de kuru, necessitando de estacas de madeira para se manterem de pé. As três meninas sentadas também estão afetadas pela doença. <sup>(3)</sup> .....	3
<b>Figura 2:</b> Estrutura primária da proteína PrP <sup>C</sup> de camundongo (adaptado de Brown, 2001). <sup>(27)</sup> .....	8
<b>Figura 3:</b> Diferenças esquemáticas entre as estruturas terciárias de PrP <sup>C</sup> (predita por RMN) e PrP <sup>Sc</sup> (isoforma especulada do príon humano) (adaptado de <a href="http://www.stanford.edu/group/virus/prion/prion2.html">http://www.stanford.edu/group/virus/prion/prion2.html</a> ). <sup>(37)</sup> .....	9
<b>Figura 4:</b> <b>a)</b> Estrutura geral de um $\alpha$ -aminoácido; <sup>(50)</sup> <b>b)</b> Isômeros L e D dos $\alpha$ -aminoácidos; <b>c)</b> Configuração S, que prevalece nos $\alpha$ -aminoácidos L (com exceção da glicina que é aquiral) encontrados em proteínas, na qual ocorre um giro de sentido anti-horário (do substituinte de maior prioridade para o de menor prioridade). <sup>(49)</sup> .....	14
<b>Figura 5:</b> Curva de titulação, indicando os valores de pK <sub>a</sub> dos grupos $\alpha$ -amino e $\alpha$ -carboxílico e o pK <sub>a</sub> do grupo ionizável da cadeia lateral para o aminoácido histidina. <sup>(50)</sup> .....	16
<b>Figura 6:</b> <b>a)</b> A formação da ligação peptídica (destacada em amarelo) ocorre através de uma reação de condensação entre o grupo $\alpha$ -carboxila de um aminoácido e o grupo $\alpha$ -amino de outro aminoácido, com a conseqüente liberação de uma molécula de água; <b>b)</b> Representação da estrutura primária de um fragmento protéico empregando o código de 3 letras para cada aminoácido. <sup>(50)</sup> .....	17
<b>Figura 7:</b> Esquema representando a configuração <i>trans</i> dos aminoácidos num fragmento polipeptídico, com os ângulos $\phi$ e $\Psi$ indicados, assim como os comprimentos de ligação característicos das ligações peptídicas. <sup>(50)</sup> .....	18
<b>Figura 8:</b> <b>a)</b> Modelo de $\alpha$ -hélice, indicando as ligações de hidrogênio entre os grupos CO do aminoácido (n) e NH do aminoácido (n+4) e o passo da hélice (5,4 Å); <sup>(50)</sup> <b>b)</b> representação esquemática de uma $\alpha$ -hélice. <sup>(53)</sup> .....	19
<b>Figura 9:</b> Arranjos de folhas- $\beta$ : <b>a)</b> paralela com um período (ou a distância entre as cadeias laterais de dois resíduos de aminoácidos consecutivos de uma fita- $\beta$ ) característico de 6,5 Å; <b>b)</b> antiparalela, com um período de 7,0 Å. <sup>(50)</sup> .....	20
<b>Figura 10:</b> Exemplo de uma estrutura supersecundária comum, o motivo $\beta\alpha\beta$ . <sup>(50)</sup> .....	22
<b>Figura 11:</b> Funil de energia livre representando o processo de dobramento de uma proteína (figura adaptada do livro Princípios de Bioquímica de Lehninger). <sup>(50)</sup> .....	24

<b>Figura 12:</b> Esquema do método de integração <i>Leap-Frog</i> . Este método é assim denominado, pois as velocidades e as posições das partículas são atualizadas em valores sistematicamente intercalados, como se as velocidades saltassem como “rãs” sob as posições e vice-versa (imagem adaptada do Manual do GROMACS 4.0). <sup>(61)</sup> .....	30
<b>Figura 13:</b> Esquema bidimensional da replicação da caixa de simulação empregando as condições de contorno periódicas (PBC) e a convenção da “imagem mínima”. O círculo em azul indica o raio de corte aplicado em uma dada partícula. <sup>(69)</sup> .....	40
<b>Figura 14:</b> Representação esquemática da formação da ligação de hidrogênio (LH) estabelecida entre os grupamentos CO (i) e NH (i + n) na cadeia polipeptídica que define os diferentes tipos de n-voltas: <b>a)</b> 3-volta; <b>b)</b> 4-volta; <b>c)</b> 5-volta (imagens adaptadas do artigo de Kabsch e Sander) <sup>(81)</sup> .....	53
<b>Figura 15:</b> Representação esquemática das ligações de hidrogênio (LH) na estruturas: <b>a)</b> Segmento $\beta$ paralelo; <b>b)</b> Segmento $\beta$ anti-paralelo (imagens adaptadas do artigo de Kabsch e Sander) <sup>(81)</sup> .....	54
<b>Figura 16:</b> Esquema da curvatura para as <i>dobras (bends)</i> , indicando um ângulo ( $>70^\circ$ ) entre a direção dos 3 primeiros resíduos e a direção dos 3 últimos resíduos do fragmento da cadeia polipeptídica em questão (imagem adaptada do artigo de Kabsch e Sander) <sup>(81)</sup> .....	55
<b>Figura 17:</b> Gráficos de: <b>a)</b> RMSD ( $C\alpha$ ) em função do tempo de simulação; <b>b)</b> RMSF ( $C\alpha$ ) em função do resíduo, obtido sobre os últimos 5 ns de simulação para cada um dos campos de força GROMOS, AMBER e OPLS.....	57
<b>Figura 18:</b> Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para os campos de força: <b>a)</b> AMBER; <b>b)</b> GROMOS; <b>c)</b> OPLS.....	58
<b>Figura 19:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: <b>a)</b> AMBER; <b>b)</b> GROMOS; <b>c)</b> OPLS.....	60
<b>Figura 20:</b> Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV): <b>a)</b> GROMOS em $t = 0$ ns; <b>b)</b> AMBER em $t = 0$ ns; <b>c)</b> OPLS em $t = 0$ ns; <b>d)</b> GROMOS em $t = 30$ ns; <b>e)</b> AMBER em $t = 45$ ns; <b>f)</b> OPLS em $t = 30$ ns. ....	61
<b>Figura 21:</b> Gráficos de RMSD de $C\alpha$ em função do tempo de simulação para as trajetórias estendidas dos campos de força AMBER e GROMOS.....	62
<b>Figura 22:</b> Gráficos de RMSF ( $C\alpha$ ) em função do resíduo para as simulações estendidas dos campos de força AMBER e GROMOS representando diferentes intervalos de tempo de simulação: <b>a)</b> 50 a 60 ns; <b>b)</b> 90 a 100 ns. ....	62

<b>Figura 23:</b> Gráfico de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para : <b>a)</b> AMBER; <b>b)</b> GROMOS .....	63
<b>Figura 24:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação para: <b>a)</b> AMBER; <b>b)</b> GROMOS.....	64
<b>Figura 25:</b> Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV) para: <b>a)</b> GROMOS em t = 50 ns; <b>b)</b> GROMOS em t = 80 ns; <b>c)</b> GROMOS em t = 100 ns; <b>d)</b> AMBER em t = 50 ns; <b>e)</b> AMBER em t = 80 ns; <b>f)</b> AMBER em t = 100 ns. Sobreposição de imagens (SPDBV) correspondentes aos instantes de tempos de 50, 60, 70, 80, 90 e 100 ns para: <b>g)</b> GROMOS; <b>h)</b> AMBER. ....	65
<b>Figura 26:</b> Representação bioquímica do aminoácido histidina com a cadeia lateral: <b>a)</b> carregada positivamente; <b>b)</b> não-carregada (imagens adaptadas do livro Princípios de Bioquímica de Lehninger). <sup>(50)</sup> .....	67
<b>Figura 27:</b> Representação dos estados iônicos e tautoméricos da cadeia lateral do aminoácido histidina. O equilíbrio entre as diferentes formas da histidina se deve à ressonância que ocorre no grupamento imidazol da cadeia lateral. Esta ressonância permite que qualquer um dos átomos de nitrogênio do anel seja protonado, dando origem a duas formas neutras possíveis para o aminoácido histidina: os tautômeros $\delta$ e $\epsilon$ nos quais, respectivamente, os nitrogênios $\delta$ e $\epsilon$ são protonados. ....	68
<b>Figura 28:</b> Representações do aminoácido histidina conforme o campo de força GROMOS: <b>a)</b> HISH, com nitrogênios ND1 e NE2 protonados e com uma carga resultante de +1; <b>b)</b> HISB, com o nitrogênio NE2 protonado e com carga resultante nula; <b>c)</b> HISA, com o nitrogênio ND1 protonado e com carga resultante nula. HISH é a representação da histidina carregada, enquanto HISA e HISB são ambas formas neutras da histidina. (Adaptado de Biomolecular Simulation: The GROMOS 96 Manual and User Guide) <sup>(85)</sup> .....	69
<b>Figura 29:</b> Gráfico de RMSD ( $C\alpha$ ) em função do tempo para as simulações: Protonado, Neutro, Neutro_HisA, Neutro_HisB, Neutro_HisA_HisB, Neutro_HisB_HisA e Neutro_HisH. ....	71
<b>Figura 30:</b> Gráfico de RMSF ( $C\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 15 ns de cada simulação) para as simulações: Protonado, Neutro, Neutro_HisA, Neutro_HisB, Neutro_HisA_HisB, Neutro_HisB_HisA e Neutro_HisH.....	72
<b>Figura 31:</b> Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP): <b>a)</b> Neutro; <b>b)</b> Protonado; <b>c)</b> Neutro_HisA; <b>d)</b> Neutro_HisB; <b>e)</b> Neutro_HisA_HisB; <b>f)</b> Neutro_HisB_HisA; <b>g)</b> Neutro_HisH.....	73

<b>Figura 32:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: <b>a)</b> folhas $\beta$ ; <b>b)</b> $\alpha$ -hélices. As curvas nos gráficos referem-se às simulações de: Protonado, Neutro, Neutro_HisA, Neutro_HisB, Neutro_HisA_HisB, Neutro_HisB_HisA e Neutro_HisH.....	74
<b>Figura 33:</b> Gráficos de: <b>a)</b> RMSD ( $C\alpha$ ) em função do tempo de simulação; <b>b)</b> RMSF ( $C\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 30 ns de simulação). As curvas nos gráficos se referem às simulações de Protonado_13 e Protonado_7 .....	80
<b>Figura 34:</b> Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para: <b>a)</b> Protonado_13; <b>b)</b> Protonado_7.....	81
<b>Figura 35:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação para: <b>a)</b> Protonado_7; <b>b)</b> Protonado_13.....	82
<b>Figura 36:</b> Imagens correspondentes a determinados instantes de tempo para as simulações (criadas com o programa SPDBV) de Protonado_13 em: <b>a)</b> $t = 0$ ns; <b>b)</b> $t = 9$ ns; <b>c)</b> $t = 20$ ns; e de Protonado_7 em: <b>d)</b> $t = 0$ ns; <b>e)</b> $t = 9$ ns; <b>f)</b> $t = 20$ ns.....	82
<b>Figura 37:</b> Gráficos de: <b>a)</b> RMSD ( $C\alpha$ ) em função do tempo de simulação; <b>b)</b> RMSF ( $C\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 10 ns de cada simulação). As curvas nos gráficos referem-se às simulações de Neutro_HisB, Neutro_HisB_1, Neutro_HisB_2, Protonado e Protonado_1.....	84
<b>Figura 38:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: <b>a)</b> folha- $\beta$ ; <b>b)</b> $\alpha$ -hélice. As curvas nos gráficos referem-se às simulações de Neutro_HisB, Neutro_HisB_1, Neutro_HisB_2, Protonado e Protonado_1. ...	85
<b>Figura 39:</b> Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para: <b>a)</b> Neutro_HisB; <b>b)</b> Neutro_HisB_1; <b>c)</b> Neutro_HisB_2; <b>d)</b> Protonado; <b>e)</b> Protonado_1.. ..	86
<b>Figura 40:</b> Gráfico de RMSD ( $C\alpha$ ) em função do tempo para as simulações de Protonado_SPME, Neutro_SPME, Protonado_GRF e Neutro_GRF.....	89
<b>Figura 41:</b> Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP): <b>a)</b> Protonado_SPME; <b>b)</b> Neutro_SPME; <b>c)</b> Protonado_GRF; <b>d)</b> Neutro_GRF.....	90
<b>Figura 42:</b> Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: <b>a)</b> Protonado_SPME; <b>b)</b> Neutro_SPME; <b>c)</b> Protonado_GRF; <b>d)</b> Neutro_GRF.. ..	91

**Figura 43:** Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV) para SPME: **a)** Protonado em  $t = 0$  ns; **b)** Protonado em  $t = 50$  ns; **c)** Neutro em  $t = 0$  ns; **d)** Neutro em  $t = 50$  ns; e GRF: **e)** Protonado em  $t = 0$  ns; .....92

## Lista de Tabelas

<b>Tabela I:</b> Número de moléculas e íons de cada um dos sistemas simulados.....	57
<b>Tabela II:</b> Número de moléculas e íons de cada um dos sistemas simulados, e a forma de protonação dos resíduos de aminoácidos histidinas, ácidos aspárticos e glutâmicos em cada uma das simulações. ....	70
<b>Tabela III:</b> Simulações e seus respectivos valores de semente de gerador de velocidades iniciais.....	83
<b>Tabela IV:</b> Número de íons e moléculas de cada sistema simulado. ....	88

## Acrônimos e Siglas Empregadas

<b>1B10</b>	Código PDB ID para a proteína príon celular do Hamster Sírio.
<b>2PrP</b>	Código PDB ID para a proteína príon celular do Hamster Sírio.
<b>AMBER</b>	Campo de força ou pacote de programas de simulação.
<b>BSE</b>	"Bovine Spongiform Encephalopathy".
<b>CD</b>	Dicroísmo circular.
<b>CHARMM</b>	Campo de força ou pacote de programas de simulação; do inglês, "Chemistry at HARvard Macromolecular Mechanics".
<b>CJD</b>	Doença de "Creutzfeldt-Jakob".
<b>DNA</b>	Ácido desoxirribonucléico.
<b>DM</b>	"Dinâmica Molecular".
<b>DSSP</b>	"Dictionary of Protein Secondary Structure".
<b>EEB</b>	Encefalopatia espongiforme bovina.
<b>ENCAD</b>	Campo de força ou pacote de programas de simulação; do inglês, "ENergy Calculations And Dynamics".
<b>FFI</b>	"Familial Fatal Insomnia".
<b>FTIR</b>	Espectroscopia no infravermelho com transformada de Fourier.
<b>F3C</b>	Modelo clássico para a água desenvolvido para o campo de força ENCAD.
<b>GPI</b>	Glicosilfosfatidilinositol; glicolipídios que ancoram proteínas na face externa da membrana plasmática.
<b>GQT</b>	Grupo de Química Teórica.
<b>GRF</b>	Campo de reação generalizado; do inglês, "Generalized Reaction Field".
<b>GROMOS</b>	Campo de força; do inglês, GRONingen MOlecular Simulation.
<b>GSS</b>	Doença de "Gerstmann-Straussler-Scheinker".
<b>HF</b>	Método aproximado para a resolução das equações de Schrödinger; do inglês, "Hartree-Fock".



<b>LH</b>	Ligação de Hidrogênio.
<b>OPLS</b>	Campo de força; do inglês, "Optimized Parameters for Liquid State Simulations".
<b>PBC</b>	Condição de contorno periódica; do inglês, "Periodic Boundary Conditions".
<b>PDB</b>	Banco de dados de estruturas biológicas; do inglês, "Protein Data Bank".
<b>pH</b>	Potencial hidrogeniônico.
<b>pI</b>	Ponto isoelétrico.
<b>pKa</b>	Potencial de dissociação de um ácido; tendência de um grupo de fornecer prótons.
<b>PRION</b>	"Proteinaceous Infectious Only particle".
<b>PrP</b>	"Proteína Prion celular".
<b>PrP-res</b>	Proteína Prion resistente a proteases.
<b>PrP-sen</b>	Proteína Prion Celular sensível a proteases.
<b>RF</b>	Campo de reação; do inglês "Reaction Field".
<b>RMN</b>	Ressonância magnética nuclear.
<b>RMSD</b>	"Root Mean Square Deviation".
<b>RMSF</b>	"Root Mean Square Fluctuation".
<b>SEP</b>	Superfície de energia potencial.
<b>SPC</b>	Modelo clássico para a água desenvolvido para o campo de força GROMOS.
<b>SPDBV</b>	Programa para visualizar e analisar proteínas; do inglês, Swiss-PdbViewer.
<b>SPME</b>	Método para o cálculo das interações eletrostáticas de longo alcance; do inglês, "Smooth Particle-Mesh Ewald".
<b>TIP3P</b>	Um dos modelos clássicos para a água utilizado no campo de força OPLS.
<b>TSE</b>	"Transmissible Spongiform Encephalopathy".
<b>vCJD</b>	Nova variante da doença de Creutzfeldt-Jakob.

## Resumo

Os príons são proteínas que causam um grupo de doenças neurodegenerativas invariavelmente fatais, sendo uma das mais conhecidas a encefalopatia espongiforme bovina (ou doença da vaca louca). A proteína príon celular (PrP<sup>c</sup>), rica em estrutura  $\alpha$ -helicoidal, sofre uma mudança na sua estrutura secundária produzindo a proteína patológica (PrP<sup>Sc</sup>; o príon) na qual prevalecem folhas- $\beta$ . Devido a falta de dados de estruturais de alta resolução dos príons, simulações de DM podem ser particularmente úteis para estudar o redobramento de PrP. Estudos experimentais e computacionais, descritos na literatura, indicam que a utilização de pH ácido é capaz de criar alguma instabilidade estrutural, produzindo um ganho de estrutura- $\beta$  na região N-terminal antes desestruturada.

Este trabalho se propõe a investigar computacionalmente as mudanças estruturais na proteína príon celular do hamster Sírio induzidas por alteração de pH. Para isso, foi avaliada a influência do uso de diferentes campos de força (GROMOS, AMBER e OPLS), diferentes estados de protonação dos resíduos de histidina, diferentes condições iniciais e diferentes métodos de cálculo de interações eletrostáticas de longo alcance (GRF e SPME). A partir da evolução temporal das estruturas secundárias, foi observada uma forte dependência dos resultados com o uso de diferentes parâmetros de simulação. De fato, a tendência de pH descrita na literatura não foi claramente observada neste trabalho. Isso pode estar associado com a necessidade de se investir mais em múltiplas simulações de dinâmica molecular para quantificar com maior precisão o comportamento estrutural dos fragmentos protéicos em cada pH de estudo.

**Palavras-chave: Príon; Simulações de Dinâmica Molecular; Campos de força; pH; Interações eletrostáticas de longo alcance.**

## Abstract

Prions are proteins that cause a group of invariably fatal neurodegenerative diseases, one of the most known being bovine spongiform encephalopathy (or mad cow disease). The cellular prion protein ( $\text{PrP}^c$ ), rich in  $\alpha$ -helical structure, undergoes a change in its secondary structure producing the pathological protein ( $\text{PrP}^{\text{Sc}}$ , the prion) in which  $\beta$ -sheet structure prevails. Because of the lack of high-resolution prion structural data, MD simulations can be particularly useful to study PrP misfolding. Experimental and computational studies, described in literature, indicate that the use of low pH is capable to create some structural instability, producing a gain of  $\beta$ -structure content in the otherwise unstructured N-terminal region.

This work aims to investigate computationally structural changes in the cellular prion protein of Syrian hamster induced by pH change. For this, we evaluated the influence of different force fields (GROMOS, AMBER and OPLS), different protonation states of histidine residues, different initial conditions and different methods for calculating long-range electrostatic interactions (GRF and SPME). From the time evolution of the secondary structures, we observed a strong dependence on the simulation parameters. In fact, the pH tendency described in literature was not clearly observed in this work. It may be associated with the need to invest more in multiple molecular dynamics simulations to quantify more accurately the structural behavior of the protein fragments in each pH study.

**Key-words: Prion; Molecular Dynamics Simulations; Force fields; pH; Long-range electrostatics.**

## **1. INTRODUÇÃO**

## 1.1 PRÍONS

### 1.1.1 A História dos Príons

Apesar de existirem relatos das doenças priônicas datados do Império Romano, a primeira descrição dessas doenças remonta ao século XVIII<sup>(1)</sup>, na Europa. Mais especificamente, foi descrita pela primeira vez em 1755, na Grã-Bretanha, em um relatório publicado no Diário das Câmaras. Essa doença, então denominada de doença das tremuras, afetava pequenos ruminantes, principalmente os ovinos e, mais raramente, os caprinos. A baixa ocorrência dessa doença tornava-a alvo de importância econômica e sanitária secundária, tendo sido, por isso, praticamente esquecida por quase 200 anos.<sup>(2)</sup>

Discussões sobre o assunto ressurgiram em 1936, quando dois investigadores da escola veterinária de Toulouse, Cuillé, J. e Chelle, P-L. demonstraram que a doença das tremuras era transmissível entre carneiros e que era veiculada por um agente infeccioso não convencional. Esses mesmos pesquisadores mostraram que era impossível detectar a partícula patogênica antes do surgimento dos primeiros sinais clínicos e que a evolução da doença era acompanhada por longos períodos de incubação, que podiam chegar a anos. Com isso, surgiu pela primeira vez a noção de “vírus lento”. Com a eclosão da Segunda Guerra Mundial e a morte prematura de Chelle os avanços nessas pesquisas foram barrados e caíram no esquecimento.<sup>(2)</sup>

Em 1950, o veterinário Bjorn Sigurdsson, de Reiquejavique, retomou o trabalho dos investigadores franceses. Analisando amostras de cérebro de carneiros infectados, Sigurdsson constatou longos tempos de incubação, resistência a tratamentos físico-químicos e, inclusive, tamanhos de partículas infecciosas incomuns aos vírus clássicos. Com base nessas análises classificou esta patologia de doença viral de “vírus lento”. Anos mais tarde, o veterinário americano, William Hadlow, espantou-se com as semelhanças clínicas e anatomopatológicas entre essa doença ovina e uma patologia humana conhecida como Kuru.<sup>(2)</sup>

Kuru, que significa arrepio, é a denominação dada a uma desordem neurológica que atingiu aldeias indígenas da Tribo dos montes Fore, na Papua-Nova Guiné (ver Figura 1). Esta doença atingia cerca de 10% da população das aldeias e caracterizava-se por movimentos oculares anormais, tremores e demência. No estágio terminal da doença, o indivíduo ficava deficiente, sendo incapaz de gesticular ou engolir. Em 1955, o Dr. Vincent Zigas, impressionado pela alta incidência destas perturbações neurológicas afetando aldeias indígenas, alertou as autoridades australianas. Em 1957, acompanhado pelo microbiólogo e

pediatra americano Carleton Gajdusek, Zigas estudou os casos clínicos de centenas de mulheres e crianças afetadas pela estranha doença, nas tribos da Nova Guiné.<sup>(2)</sup>



**Figura 1:** A imagem mostra quatro mulheres afetadas com uma forma avançada de kuru, necessitando de estacas de madeira para se manterem de pé. As três meninas sentadas também estão afetadas pela doença.<sup>(3)</sup>

Através de seus estudos, Zigas e Gajdusek constataram que a ocorrência da doença se devia à práticas de canibalismo adotadas pelas tribos. Em tais práticas, aos homens eram reservados os músculos, símbolo de força e virilidade, e às mulheres e crianças cabiam o cérebro e as vísceras. Os resultados desses estudos foram publicados por Zigas e Gajdusek, jogando luz às questões neuropatológicas e epidemiológicas desta doença.<sup>(4)</sup> Assim que as práticas de canibalismo foram proibidas, a ocorrência de Kuru caiu exponencialmente. No entanto, a doença matou cerca de três mil indivíduos em uma população de, aproximadamente, trinta e cinco mil indígenas da Papua Nova-Guiné.<sup>(2)</sup>

Os estudos de Gajdusek chamaram a atenção de Hadlow, que reconheceu características comuns entre a doença das tremuras e o Kuru.<sup>(5)</sup> No entanto, foi o grupo de Gajdusek que deu provas quanto a semelhança das duas doenças. Gajdusek convencido da semelhança entre essas doenças, dedicou seu tempo a experimentos de inoculação de tecidos cerebrais de pacientes mortos por Kuru em cérebros de macacos. Após longos períodos de incubação, os macacos eram afetados pela encefalopatia espongiiforme proveniente de humanos. Estes resultados foram publicados em 1966 na revista Nature<sup>(6)</sup>, conduzindo a comunidade científica da época a apoiar a descoberta da primeira doença humana causada por um vírus lento.<sup>(2)</sup>

O grupo de Gajdusek deu, no ano de 1968, outra grande contribuição ao estudo das doenças neurodegenerativas ao, instigados pelas semelhanças entre o Kuru e a doença de Creutzfeldt-Jakob (conhecida desde 1921), inocular cérebros de macacos com a doença de Creutzfeldt-Jakob, mostrando assim a ruptura de barreira entre espécies.<sup>(7)</sup> Tais descobertas de transmissibilidade de doenças neurológicas, rendeu a Gajdusek o Prêmio Nobel de Medicina em 1976.<sup>(2)</sup>

Neste momento, ainda não se tinha conhecimento da natureza do agente patogênico causador destas doenças. Partindo do pressuposto de se tratar de um agente viral, um grupo liderado por Thykave Alper no Hospital Hammersmith, em Londres, empregou metodologias físicas, mais especificamente a irradiação, para estimar o tamanho da partícula infecciosa. Os resultados apontaram para um agente patogênico de tamanho muito menor que o esperado para um vírus. Com o auxílio de um radiobiólogo francês, Raymond Latarjet, do Instituto Curie, Alper verificou que a amostra infecciosa era sensível à radiações de 280 nm, correspondente ao pico de absorção de proteínas.<sup>(2, 8)</sup>

A essa altura, o físico e matemático inglês J. S. Griffith já havia proposto, em 1967,<sup>(9)</sup> um modelo teórico, apresentando a hipótese de uma única proteína ser a causadora destas doenças neurodegenerativas.<sup>(10)</sup> Mas foram estudos experimentais posteriores, realizados por um jovem entusiasta pesquisador, Stanley Prusiner, que, de fato, decifraram a natureza destas partículas infecciosas.<sup>(2)</sup>

Stanley Prusiner, formado pela Faculdade de Medicina da Universidade da Pensilvânia, em 1968, começou a se interessar pelas doenças neurodegenerativas quando fazia residência em neurologia. Foi em 1972 que Prusiner teve seu primeiro contato com a CJD (Doença de Creutzfeldt-Jakob), acompanhando o caso de uma paciente sua que faleceu em menos de 2 meses após o início dos sintomas. Este contato impactante com os efeitos da doença inspirou Prusiner a iniciar uma empreitada para determinar a natureza das partículas responsáveis por essas devastadoras doenças.<sup>(11)</sup>

Inúmeros foram os pesquisadores que falharam na difícil tarefa de extrair o agente infeccioso a partir de cérebros de carneiros infectados. No entanto, Prusiner, impulsionado pelo otimismo da juventude, realizou tal feito. Em 1981, Prusiner e seus colegas provaram que a infectividade de amostras de cérebros de hamsters, inoculados com a doença das tremuras, dependia de uma proteína hidrófoba. Provaram, também, que as propriedades de resistência do agente infeccioso correspondiam a de uma proteína e que o comportamento químico indicava que a amostra era desprovida de ácidos nucleicos.<sup>(2, 12, 13)</sup> Neste contexto, foi

concebido o conceito de PRION (**PR**oteinaceous **I**nfectious **ON**ly particle), sugerido pelo grupo de Prusiner, ganhador do Prêmio Nobel de Medicina em 1997.<sup>(10)</sup>

Infelizmente, após estes consideráveis avanços, o interesse dado às doenças priônicas voltou a cair, ressurgindo fortemente somente em 1985, quando se teve notícia de uma nova doença que afetava as manadas de bovinos britânicos. Um ano depois, veterinários britânicos identificaram a doença como encefalopatia espongiforme bovina, também conhecida como doença da vaca louca. Estimativas oficiais apontam que, até o ano 2000, cerca de 185 mil bovinos sucumbiram à doença, sendo que nas primeiras semanas morreram em torno de mil animais por semana.<sup>(2)</sup>

Investigações epidemiológicas mostraram que a fonte de contaminação provinha de farinhas suplementares adicionadas na alimentação dos bovinos. Estas farinhas eram produzidas a partir de carcaças e restos de carnes, oriundas de matadouros, notadamente, de carneiros. Em 1988 o governo de Margareth Thatcher interditou o uso de farinhas animais. No entanto, os fabricantes das farinhas obtiveram um prazo de cinco semanas para escoar seus estoques. Nesse período, cerca de centenas de animais continuaram a ser infectados por dia.<sup>(2)</sup>

Em 1995 artigos científicos foram publicados relatando a existência de uma forma atípica da doença de Creutzfeldt-Jakob, que afetava pessoas com menos de 30 anos.<sup>(14)</sup> Um ano depois, foram publicados artigos sugerindo que príons bovinos poderiam ter sido transmitidos à espécie humana.<sup>(11, 15)</sup> Estes e outros estudos repercutiram no quadro econômico europeu, fazendo com que no ano 2000, mais de 4 milhões de bovinos tenham sido sacrificados, o que representou um custo de 6,8 bilhões de euros ao Reino Unido, causando uma crise no setor pecuário deste país.<sup>(2)</sup>

### **1.1.2 Encefalopatias Espongiformes Transmissíveis**

O termo “príons” é empregado para designar proteínas infecciosas que causam um grupo de doenças neurodegenerativas invariavelmente fatais.<sup>(11)</sup> Este grupo de doenças, caracterizado por desencadear no cérebro um aspecto esponjoso, é classificado de Encefalopatias Espongiformes Transmissíveis (TSEs, da sigla em inglês).<sup>(1)</sup>

Na década de 90, estudos revelaram que a mesma proteína encontrada em animais doentes também estava presente em animais sadios. A esta proteína (encontrada principalmente nos neurônios), natural no organismo de mamíferos, se deu o nome de “proteína príon celular”.<sup>(11)</sup> Mais tarde, foi mostrado que ambas, a proteína natural e a



patológica, possuíam a mesma sequência de aminoácidos, diferindo somente em suas estruturas secundárias e terciárias.<sup>(11, 16, 17)</sup>

As doenças priônicas (designação usada para agrupar as doenças que possuem o príon como agente causal em comum)<sup>(11)</sup> atingem animais e humanos. Nos animais as formas mais conhecidas são: o scrapie (ou doença das tremuras), que atinge ovelhas e cabras, e a encefalopatia espongiforme bovina (sigla, EEB) ou doença da vaca louca. Menos conhecidas, outras formas dessas encefalopatias já foram verificadas em animais em cativeiro (grandes felinos, ruminantes e primatas), em cervos, em antílopes selvagens, em minks (da família dos mustelídeos) e até mesmo em gatos domésticos.<sup>(1)</sup>

Nos seres humanos estas encefalopatias incluem o Kuru, a doença de Gerstmann-Straussler-Scheinker (sigla, GSS), a insônia familiar fatal (sigla, FFI), a doença de Creutzfeldt-Jakob (sigla em inglês, CJD) que atinge pessoas entre 50 e 60 anos e a nova variante da doença de Creutzfeldt-Jakob (sigla, vCJD) que atinge jovens de 30 anos.<sup>(18)</sup> Existem evidências de que a vCJD é causada pelo agente da encefalopatia espongiforme bovina e acredita-se que a transmissão aos humanos ocorreu por via alimentar através da ingestão de carne de animais infectados pela EEB.<sup>(10)</sup>

São conhecidas três formas de doenças priônicas em humanos: hereditária, esporádica (de origem desconhecida) e transmissível<sup>(1, 11, 19)</sup>. A forma transmissível corresponde a 5%, enquanto a hereditária é responsável por 15% dos casos registrados de portadores dessas doenças.<sup>(10)</sup>

Além das práticas de canibalismo que propagaram a disseminação do Kuru em aldeias indígenas, outras fontes foram importantes na transmissibilidade das doenças priônicas na espécie humana. Entre elas merecem destaque as causadas por procedimentos médicos, tais como transplante de órgãos (em especial, dura-máter), o uso de derivados hormonais preparados a partir de material humano (destacando-se o tratamento de crianças com hormônio do crescimento) e material cirúrgico (em especial para cirurgias no cérebro).<sup>(10)</sup>

Atualmente, o risco de transmissão via tratamento com hormônio do crescimento foi eliminado, uma vez que agora esse hormônio é produzido em bactérias. Da mesma forma, as metodologias de esterilização e reutilização de materiais cirúrgicos, empregados em biópsias de tecido cerebral contaminado por príons, foram aperfeiçoadas.<sup>(10)</sup> A poucos anos, visando evitar a propagação destas doenças, a Organização Mundial da Saúde divulgou normas rígidas nesse sentido. Estas normas também encontram-se disponíveis na página do Ministério da Saúde brasileiro.<sup>(20)</sup>

As doenças por príons hereditárias são causadas por mutações no gene Prnp. Este gene, localizado no cromossomo 20, é o responsável pela produção da proteína príon celular presente no organismo humano. Mutações neste gene desencadeiam a produção de uma proteína defeituosa (o príon). Além disso, como a alteração está impressa no DNA, terá 50% de chances de ser transmitida aos descendentes do portador da mutação. As formas hereditárias humanas mais conhecidas das doenças priônicas são: GSS e CJD.<sup>(11)</sup>

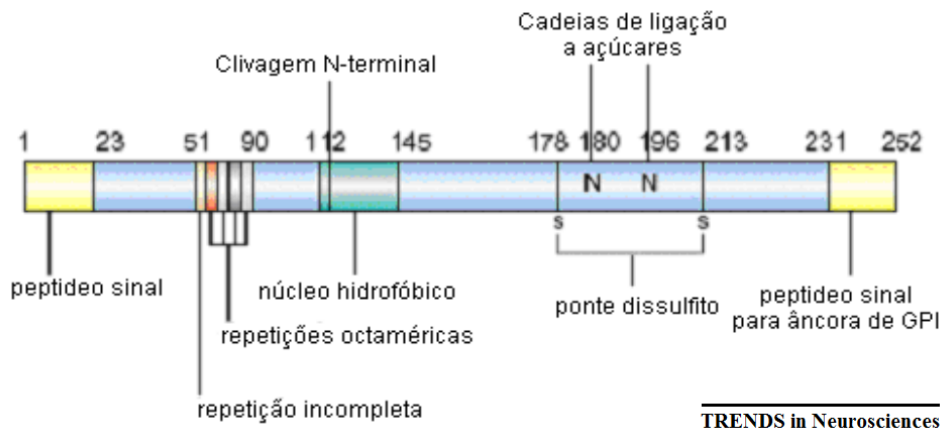
### 1.1.3 PrP<sup>C</sup> x PrP<sup>Sc</sup>

A distinção entre as isoformas de PrP foi feita através de dois critérios: a solubilidade em detergentes e a resistência às proteases. Assim, a proteína príon celular, sensível às proteases e solúvel em detergentes, foi designada PrP-sen (-sen de sensível). Já a proteína patológica (príon), insolúvel em detergentes e com alta resistência às proteases, foi denominada PrP-res (-res de resistente)<sup>(16)</sup>. Mais comumente, denomina-se a isoforma normal da proteína príon celular de PrP<sup>C</sup> (c = celular) e a isoforma anormal de PrP<sup>Sc</sup> (Sc = scrapie; com referência à primeira das doenças priônicas identificadas em animais)<sup>(21, 22)</sup>.

PrP<sup>C</sup> é definida como uma sialoglicoproteína de membrana ancorada via glicosilfosfatidilinositol (GPI), de distribuição ubíqua<sup>1</sup>, de forma que embora PrP<sup>C</sup> seja altamente expressa no sistema nervoso central, também pode ser encontrada em componentes do sistema imune (medula óssea, células sanguíneas e órgãos linfóides) e, em menor quantidade, nos rins, músculos, trato gastrointestinal, fígado, pulmão, coração e nas glândulas mamárias<sup>(23, 24, 25)</sup>.

A sequência primária da proteína príon celular possui cerca de 250 aminoácidos (dependendo da espécie) e sua massa molecular varia entre 18-33 KDa de acordo com o seu nível de glicosilação (mono, di ou não-glicosilada).<sup>(26, 27)</sup> É possível identificar vários domínios distintos em PrP<sup>C</sup>, sendo que na região N-terminal tem-se um peptídeo sinal (de 22 aminoácidos), 4 cópias completas de uma sequência de octapeptídeos repetidos (OR), ricos em prolina e glicina, e uma cópia incompleta, além de um segmento hidrofóbico entre os resíduos 112 e 145 que possui um sítio para a clivagem da proteína durante seu metabolismo. A região ocupada pelos octapeptídeos repetidos de sequência completa (resíduos 51-90) é capaz de se coordenar preferencialmente a íons cobre e essa ligação está relacionada à atividade antioxidante de PrP<sup>C</sup>.<sup>(27)</sup> Na região C-terminal, os resíduos de cisteína, 178 e 213, realizam um ligação de dissulfeto (conferindo estabilidade à estrutura) e os resíduos de asparagina, 180 e 196, são possíveis sítios de glicosilação para PrP<sup>C</sup>. Adicionalmente, um

peptídeo (de 23 aminoácidos) sinalizador para âncora de GPI encontra-se na região C-terminal, conforme se observa na Figura 2, a seguir.<sup>(27)</sup>



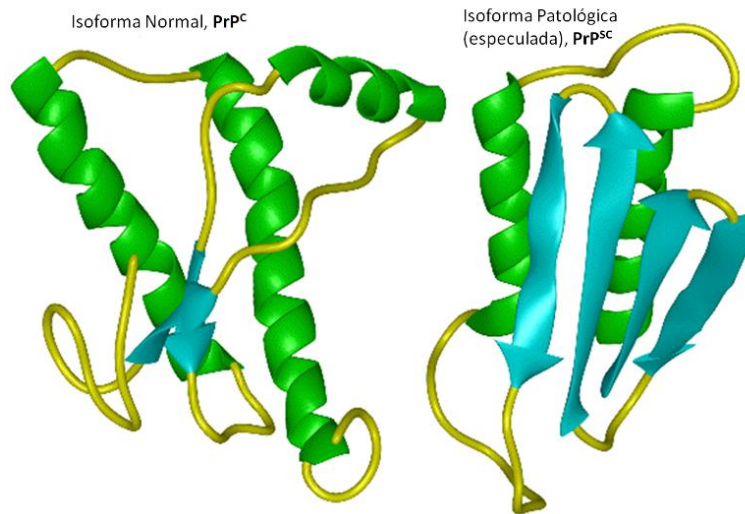
**Figura 2:** Estrutura primária da proteína PrP<sup>C</sup> de camundongo (adaptado de Brown, 2001).<sup>(27)</sup>

O precursor de PrP<sup>C</sup> é enviado ao retículo endoplasmático, onde ocorre a clivagem dos peptídeos N- e C-terminal, permitindo a adição da âncora de GPI (resíduos 121-231)<sup>(28)</sup> nesta última região. A proteína é, então, direcionada para o complexo de Golgi no qual sofre glicosilação dos resíduos 180 e 196, adição de ácido siálico nas cadeias oligossacarídicas e a formação da única ponte dissulfeto da estrutura (entre os aminoácidos 178 e 213). Dessa forma, é gerada a proteína madura com conformação adequada que será direcionada à membrana plasmática.<sup>(27)</sup>

Em relação à estrutura terciária, análises de RMN para fragmentos da proteína príon celular (resíduos 121-231) de hamsters<sup>(17)</sup>, ratos<sup>(29)</sup>, bois<sup>(30)</sup> e humanos<sup>(31)</sup> permitiram a obtenção de estruturas de alta resolução para PrP<sup>C</sup>.<sup>(32)</sup> Os resultados apontaram para uma conservação de estrutura secundária nas diferentes espécies de mamíferos analisadas.<sup>(16)</sup> O que era esperado considerando que essas sequências exibem identidade de sequência superior à 90%.<sup>(33)</sup> Dessa forma, foi possível identificar em PrP<sup>C</sup> uma região N-terminal (resíduos 23-124) desordenada, altamente flexível<sup>(11, 34, 35)</sup>, e uma região C-terminal (resíduos 125-231) composta de 3  $\alpha$ -hélices e uma folha beta anti-paralela adjacente à primeira  $\alpha$ -hélice.<sup>(36)</sup>

Estudos de espectroscopia ótica foram realizados comparando as duas isoformas de PrP. Através de análises de infravermelho com transformada de Fourier (sigla em inglês – FTIR) e dicroísmo circular (CD) foi encontrada uma diferença notável no conteúdo de estrutura secundária destas isoformas. PrP<sup>C</sup> apresentou baixo conteúdo de folhas- $\beta$  (em torno de 3 %) e alto conteúdo de  $\alpha$ -hélices (cerca de 40%). Enquanto PrP<sup>Sc</sup> mostrou ser composta

de 17-30 % de  $\alpha$ -hélices e 43-54% de folhas- $\beta$ .<sup>(11, 17)</sup> A Figura 3 apresenta uma comparação entre as estruturas tridimensionais de PrP<sup>C</sup> e PrP<sup>Sc</sup>.<sup>(37)</sup>



**Figura 3:** Diferenças esquemáticas entre as estruturas terciárias de PrP<sup>C</sup> (predita por RMN) e PrP<sup>Sc</sup> (isoforma especulada do príon humano) (adaptado de <http://www.stanford.edu/group/virus/prion/prion2.html>).<sup>(37)</sup>

No que diz respeito a estrutura de PrP<sup>Sc</sup>, esta é desconhecida, uma vez que ainda não foi possível purificar suficientemente amostras de material infeccioso para a obtenção de dados estruturais de alta resolução. Tentativas de purificar PrP<sup>Sc</sup>, obtida de amostras de cérebro de animais infectados por príon, resultaram em misturas heterogêneas e estados de agregação não-propícios para a caracterização via técnicas de RMN e raios-X. Como alternativa, foram empregadas técnicas de baixa resolução, como microscopia eletrônica, com o intuito de estabelecer características dos agregados de príons.<sup>(16, 32, 38)</sup>

## 1.2 HIPÓTESE DE INDUÇÃO CONFORMACIONAL

Existem duas teorias para explicar o mecanismo de neurodegeneração das doenças priônicas: a teoria de ganho de função de PrP<sup>Sc</sup> e a teoria de perda de função de PrP<sup>C</sup>. Atualmente, existem evidências experimentais que suportam ambas as teorias, mas nenhuma foi definitivamente comprovada. Ao que se sabe estas teorias não são mutuamente exclusivas.<sup>(24, 39)</sup>

No que diz respeito à forma de propagação da doença no organismo, a ideia mais aceita atualmente baseia-se na hipótese de indução conformacional. Segundo esta hipótese, a proteína alterada, PrP<sup>Sc</sup>, induz a mudança de conformação da proteína normal, PrP<sup>C</sup>. Este

processo implica na conversão de porções  $\alpha$ -helicoidais a folhas- $\beta$ . Assim, neste processo auto-catalítico, proteínas alteradas seguem convertendo proteínas normais. Estas proteínas alteradas, por sua vez, formam pequenos agregados ou protofibrilas que precipitam nas células cerebrais, dando ao cérebro uma aparência esponjosa.<sup>(35, 40)</sup>

Embora os sistemas de conversão *in vitro* reproduzam em parte o processo de propagação dos príons (explicando questões como a especificidade entre espécies), mostra-se pouco eficaz na produção de PrP<sup>Sc</sup>, requerendo um excesso (cerca de 50 vezes) de PrP<sup>Sc</sup> em relação à PrP<sup>C</sup>. Esta condição é justamente oposta à observada em sistemas *in vivo*.<sup>(40, 41)</sup> Além do mais, PrP<sup>Sc</sup> gerado *in vitro* não apresenta infectividade.<sup>(41, 42)</sup> Portanto, a possibilidade de co-fatores celulares participarem do processo de conversão não pode ser descartada.<sup>(40)</sup> Estudos feitos por James e colaboradores sugerem que neste processo de transição conformacional haveria o envolvimento de uma chaperona molecular (proteína acompanhante) que formaria um complexo com o intermediário metaestável PrP\*, diminuindo a barreira de energia de ativação entre PrP<sup>C</sup> e PrP<sup>Sc</sup>, facilitando assim a formação da proteína alterada (PrP<sup>Sc</sup>).<sup>(18, 34)</sup>

A hipótese de indução conformacional tem um caráter revolucionário, pois rompe com ideia aceita, até então, em biologia molecular de que somente vírus e bactérias (por conterem material genético) poderiam desencadear doenças infecciosas. Assim, os príons abrem um precedente na história das ciências biológicas, por se tratar do único grupo de doenças capaz de assumir caráter tanto infeccioso como hereditário.<sup>(10, 11, 43)</sup>

### 1.3 MODELAGEM COMPUTACIONAL

Modificando-se as condições do ambiente *in vitro*, PrP<sup>C</sup> sofre conversão a PrP<sup>Sc</sup>. Esta conversão *in vitro* pode ser observada em ambiente redutor e em pH ácido. Estes monômeros de PrP<sup>Sc</sup> acabam formando estruturas fibrilares estáveis, resistentes a proteases, que se assemelham a agregados e placas amilóides. Este mecanismo descrito *in vitro* possui relevância *in vivo* devido aos ciclos de endocitose a que a proteína PrP<sup>C</sup> está sujeita na presença de cobre (sendo que essa via transita pelo retículo endoplasmático).<sup>(44, 45, 46)</sup>

Ao ciclar constitutivamente entre a superfície celular e compartimentos endocíticos, a proteína PrP<sup>C</sup> é submetida a ambientes com diferentes valores de pH. Na superfície celular a PrP<sup>C</sup> é submetida a um pH neutro, enquanto em ambiente endocítico vivencia um pH mais baixo (em torno de 5,0, mas podendo chegar à 4,3). Desta forma, tentar compreender a

influência do pH na transição conformacional de proteína normal à infecciosa pode levar à compreensão de onde exatamente ocorre essa conversão estrutural.<sup>(21, 45, 46)</sup>

Estudos experimentais de influência do pH na mudança conformacional foram realizados com formas de PrP<sup>C</sup> de humanos e murines. Os resultados apontaram para a perda de hélices e ganho de estrutura  $\beta$  na região de pH 4,4-6.<sup>(33, 45, 46)</sup> Baseados nesses resultados e considerando a falta de dados estruturais disponíveis da proteína alterada, DeMarco e Daggett propuseram um modelo protofibrila *in silico*: o modelo espiral.<sup>(32)</sup>

A unidade estrutural básica, a proteína alterada, utilizada para a formação deste modelo protofibrila foi produzida por Alonso e colaboradores<sup>(33)</sup> via simulações de dinâmica molecular (DM) utilizando o campo de forças ENCAD.<sup>(47)</sup> Para tanto, foi utilizado o fragmento (resíduos 109-219) de PrP<sup>C</sup> do hamster Sírio (resolvido por RMN)<sup>(18)</sup> depositado no banco de dados do “Protein Data Bank” (PDB). Os resíduos His, Glu e Asp desta proteína foram protonados (promovendo um pH ácido) e a estrutura resultante foi utilizada como ponto de partida para uma simulação de 10 ns.<sup>(33)</sup> A estrutura correspondente ao tempo de 8 ns foi usada como unidade básica para a formação do modelo protofibrila.<sup>(32)</sup>

Alonso e colaboradores<sup>(33)</sup> realizaram também uma simulação de DM, por 10 ns, em meio neutro (com os resíduos Asp e Glu carregados e com o resíduo de His neutro) para o mesmo fragmento de PrP<sup>C</sup> do hamster Sírio. Com isto, foram analisadas as diferenças de estrutura secundária da proteína devido à influência do pH do sistema. De fato, verificou-se que a simulação em pH ácido promoveu o aumento de folhas- $\beta$  na região N-terminal, enquanto a simulação em pH neutro não promoveu nenhuma alteração estrutural significativa.<sup>(33)</sup>

De maneira geral, a caracterização de propriedades conformacionais de proteínas em nível atômico é um grande desafio e torna-se particularmente difícil quando trata-se do processo de conversão PrP<sup>C</sup>  $\rightarrow$  PrP<sup>Sc</sup>. Neste contexto, as simulações computacionais via DM apresentam-se como uma alternativa viável para este tipo de investigação.<sup>(33)</sup>

A dinâmica molecular é uma técnica computacional na qual são determinados os movimentos das partículas de um sistema, conhecendo-se o potencial de interação entre as partículas e as equações que regem seu movimento. Com isto, é possível estudar a evolução temporal das configurações dos constituintes do sistema. Partindo das sequências de posições geradas, são determinadas as propriedades macroscópicas do sistema, respeitando os princípios fundamentais de Mecânica Estatística.<sup>(48)</sup>

A metodologia DM é uma ferramenta para o estudo de sistemas nos quais os efeitos de temperatura devem ser considerados.<sup>(48)</sup> Como o processo de conversão PrP<sup>C</sup>  $\rightarrow$  PrP<sup>Sc</sup> é

associado com substancial alteração conformacional e não envolve reações químicas, é ideal para ser investigado por simulações DM.

O presente trabalho tem o objetivo de realizar estudos de influência de pH no processo de conversão estrutural da proteína príon celular empregando simulações de dinâmica molecular. Para isso será investigado o uso de diferentes parâmetros de simulação e sua influência nos resultados das simulações.

## **2. FUNDAMENTAÇÃO TEÓRICA**



## 2.1 PROTEÍNAS

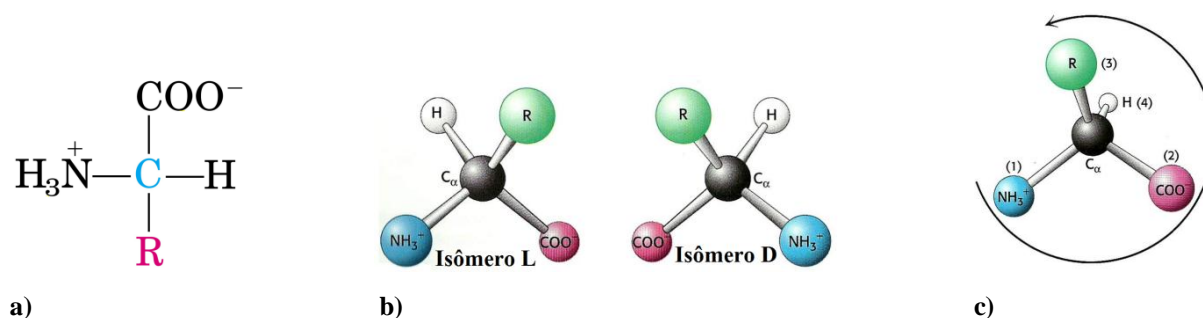
Proteínas são polímeros lineares feitos de unidades monoméricas denominadas aminoácidos, que se unem ponta a ponta,<sup>(49)</sup> e são as macromoléculas mais abundantes das células. De fato, milhares de diferentes tipos de proteínas podem ser encontradas em uma única célula.<sup>(50)</sup> Praticamente todas as transformações moleculares que definem o metabolismo celular são mediadas pela catálise protéica, o que coloca as proteínas no centro de ação dos processos biológicos.<sup>(51)</sup>

A impressionante versatilidade das proteínas nos sistemas vivos lhes conferem uma gama de funções (dinâmicas e estruturais) cruciais em essencialmente todos os processos biológicos. Essas macromoléculas estão presentes no organismo de seres vivos como catalisadores, hormônios, anticorpos, transportadores, fibras musculares e antibióticos. As proteínas também podem fornecer apoio mecânico e proteção imunitária, transmitir impulsos nervosos e controlar o crescimento e a diferenciação celular.<sup>(49, 50)</sup>

### 2.1.1 Aminoácidos: blocos fundamentais de construção das proteínas

Os aminoácidos são as unidades estruturais fundamentais das proteínas. Vários aminoácidos estão entre os compostos orgânicos que, acredita-se, surgiram nos primórdios da história da Terra. Assim, como moléculas antigas e ubíquas, os aminoácidos foram captados pela evolução para uma variedade de propósitos nos sistemas vivos.<sup>(49, 51)</sup>

Existem 20 aminoácidos comuns, que são aqueles para os quais existe pelo menos um códon no código genético, denominados de  $\alpha$ -aminoácidos por serem ligados a um grupo amino primário  $-\text{NH}_2$ . A única exceção é a prolina, que possui um grupo amino secundário ( $-\text{NH}-$ ), mas por questão de uniformidade também é referida como um  $\alpha$ -aminoácido.<sup>(50, 51, 52)</sup>



**Figura 4:** a) Estrutura geral de um  $\alpha$ -aminoácido;<sup>(50)</sup> b) Isômeros L e D dos  $\alpha$ -aminoácidos; c) Configuração S, que prevalece nos  $\alpha$ -aminoácidos L (com exceção da glicina que é aquiral) encontrados em proteínas, na qual ocorre um giro de sentido anti-horário (do substituinte de maior prioridade para o de menor prioridade).<sup>(49)</sup>

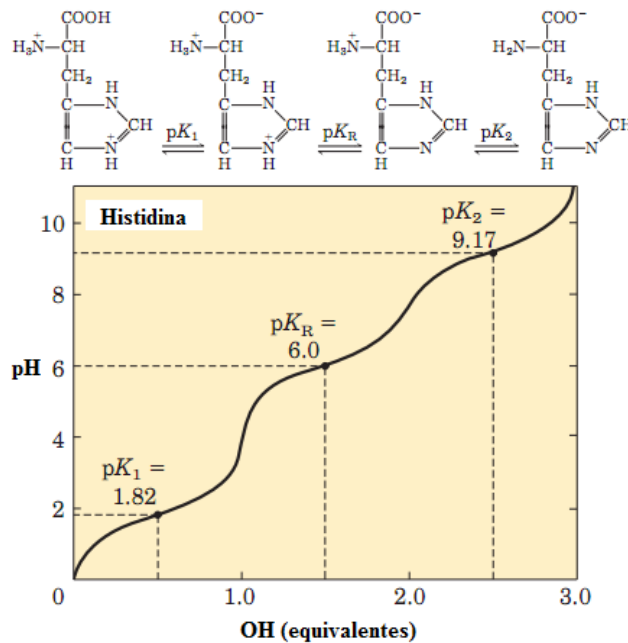
Os  $\alpha$ -aminoácidos são constituídos de um átomo central de carbono (o carbono- $\alpha$ ) ligado a um grupo amina (ou  $\alpha$ -amino), um grupo carboxil (ou  $\alpha$ -carboxil), um átomo de hidrogênio e um grupamento R diferenciado que corresponde à cadeia lateral. Em solução em pH neutro, os aminoácidos comuns ocorrem predominantemente na forma de íons dipolares (denominados zwitteríons), com a amina protonada ( $-\text{NH}_3^+$ ) e a carboxila desprotonada ( $-\text{COO}^-$ ), como mostrado na Figura 4a. Em função da presença de quatro grupamentos diferentes conectados ao carbono- $\alpha$  tetraédrico, os  $\alpha$ -aminoácidos (com exceção da glicina, cujo grupo R é um átomo de hidrogênio) são quirais. Isto é, podem existir na forma de isômeros L e D, que são imagens especulares um do outro (Figura 4b).<sup>(49, 50)</sup>

Na construção das proteínas são empregados somente os aminoácidos L, que na grande maioria dos casos possui uma configuração absoluta S (Figura 4c). Não há uma explicação satisfatória para que prevaleça o isômero L nos aminoácidos. Porém, parece plausível que a seleção de L tenha sido arbitrária e, uma vez feita, foi fixada nos primórdios da história evolutiva.<sup>(49, 51)</sup>

Conhecer as propriedades químicas dos aminoácidos é fundamental para entender sua bioquímica. Os aminoácidos diferem uns dos outros em suas cadeias laterais (ou grupos R), as quais variam em estrutura, tamanho e carga elétrica, e influenciam a solubilidade dos aminoácidos em água. Desta forma, é possível agrupar os aminoácidos em cinco classes principais com base nas propriedades de seus grupos R, em particular, suas polaridades ou tendências a interagir com água em pH fisiológico (próximo ao pH 7). Essas cinco classes são: grupos R apolares (alifáticos), aromáticos, polares não-carregados, carregados positivamente (ou básicos) e carregados negativamente (ou ácidos).<sup>(50)</sup>

A partir da curva de titulação ácido-base (que representa a adição ou a remoção gradual de prótons) para cada um dos 20 aminoácidos são obtidos os valores de  $\text{pK}_a$  dos grupos  $\alpha$ -amino e  $\alpha$ -carboxílico, além do  $\text{pK}_a$  de grupos ionizáveis das cadeias laterais (presentes em alguns aminoácidos).<sup>(50)</sup>

A Figura 5 apresenta a curva de titulação para o aminoácido histidina, cujos três grupos ionizáveis são titulados com uma base forte (NaOH). O gráfico possui três estágios distintos que correspondem à desprotonação dos três grupos diferentes na histidina. O grupo  $\alpha$ -carboxílico possui um  $\text{pK}_1$  de 1,82, que corresponde ao valor do seu primeiro ponto de inflexão. Conforme a titulação prossegue outros pontos de inflexão são alcançados em pH 6,0 e 9,17 correspondendo, respectivamente, aos  $\text{pK}_a$  do grupo ionizável da cadeia lateral e do grupo  $\alpha$ -amino.<sup>(50)</sup>



**Figura 5:** Curva de titulação, indicando os valores de  $pK_a$  dos grupos  $\alpha$ -amino e  $\alpha$ -carboxílico e o  $pK_a$  do grupo ionizável da cadeia lateral para o aminoácido histidina.<sup>(50)</sup>

O ponto isoelétrico (pI) corresponde ao pH no qual a carga elétrica líquida de um aminoácido qualquer é zero. Para a histidina, o pI (igual à 7,59) reflete a natureza do grupo R ionizável e o seu valor equivale à média dos valores de  $pK_a$  dos grupos  $\alpha$ -amino e imidazol (que contribuem com uma carga líquida de +1 contrabalanceando a carga -1 devida ao grupo  $\alpha$ -carboxílico desprotonado). Assim, pela Figura 5 é fácil ver que em qualquer pH acima de seu pI, a histidina terá uma carga líquida negativa (o mesmo raciocínio é válido para os demais aminoácidos) e em qualquer pH abaixo do seu pI, possui uma carga líquida positiva.<sup>(50)</sup>

Para os 20 aminoácidos comuns, os valores dos  $pK_a$  dos grupos  $\alpha$ -carboxílicos ( $pK_1$ ) situam-se em uma pequena faixa em torno de 2,2, ao passo que os valores de  $pK_a$  dos grupos  $\alpha$ -amino ( $pK_2$ ) estão próximos de 9,4. A tabela no ANEXO A1, apresenta os valores de  $pK_a$  e pI, entre outras propriedades para cada um dos 20 aminoácidos comuns.<sup>(51)</sup>

No que diz respeito à nomenclatura, os aminoácidos comuns são designados com abreviações de 3 letras e símbolos de 1 letra. Este último código é universalmente aceito e foi proposto por Margareth O. Dayhoff (1925-1983; considerada por muitos como a fundadora da bioinformática) como tentativa de reduzir o tamanho dos arquivos de dados usados para descrever sequências de aminoácidos.<sup>(50)</sup>

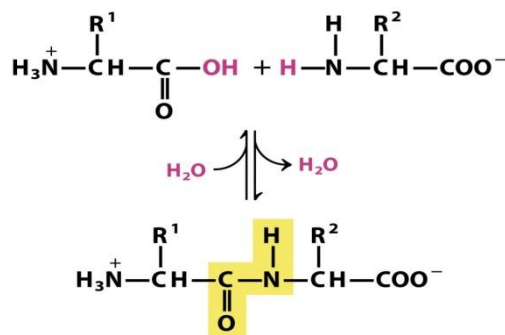
## 2.1.2 Classificação Estrutural de Proteínas

Polímeros compostos de 2, 3, alguns (3-10) e muitos aminoácidos são conhecidos, respectivamente, como dipeptídeos, tripeptídeos, oligopeptídeos e polipeptídeos. Após incorporados a um peptídeo, os aminoácidos individuais são chamados de resíduos de aminoácidos. A proteína com a maior cadeia polipeptídica de que se tem conhecimento é a tinina (presente nas fibras musculares), com 29.926 resíduos.<sup>(51)</sup>

Atualmente, conhecem-se as sequências completas de aminoácidos de mais de 20.000 proteínas. Do mesmo modo que todas as moléculas poliméricas, as proteínas podem ser descritas em termos de níveis de organização estrutural, sendo eles: estrutura primária, estrutura secundária, estrutura terciária e estrutura quaternária. Os próximos subcapítulos são destinados a explicar a essência dessa classificação estrutural.<sup>(49, 51)</sup>

### 2.1.2.1 Estrutura Primária

A estrutura primária de uma proteína consiste da sequência de aminoácidos da sua cadeia polipeptídica (ou cadeias, no caso de ser constituída por mais de uma). Os resíduos de aminoácido dessa cadeia são unidos covalentemente por ligações peptídicas (Figura 6a) significativamente estáveis, com uma meia vida ( $t_{1/2}$ ) média de cerca de 7 anos sob a maioria das condições intracelulares. A estrutura primária (Figura 6b) também inclui a localização de ligações dissulfeto na cadeia polipeptídica.<sup>(50, 51, 52)</sup>



a) Formação da ligação peptídica;



b) Estrutura primária.

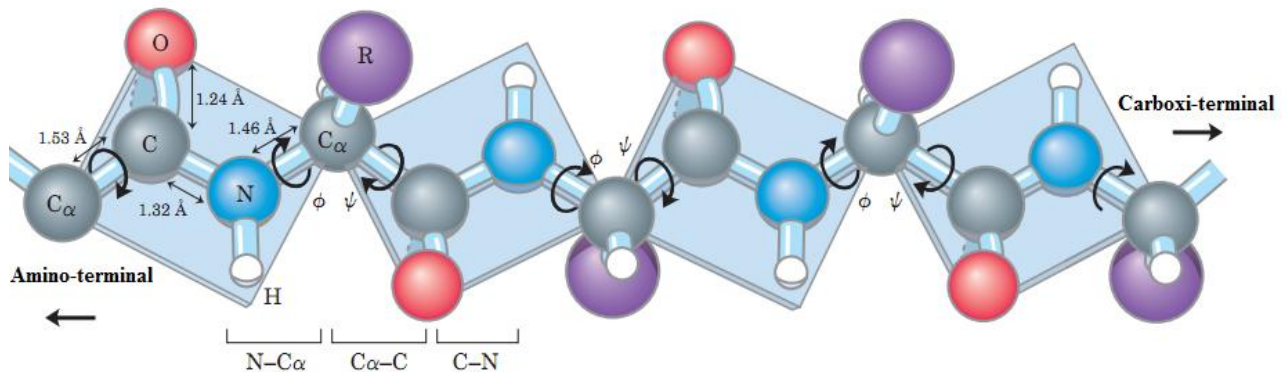
**Figura 6:** a) A formação da ligação peptídica (destacada em amarelo) ocorre através de uma reação de condensação entre o grupo  $\alpha$ -carboxila de um aminoácido e o grupo  $\alpha$ -amino de outro aminoácido, com a consequente liberação de uma molécula de água; b) Representação da estrutura primária de um fragmento protéico empregando o código de 3 letras para cada aminoácido.<sup>(50)</sup>

A posição que um aminoácido ocupa na cadeia polipeptídica de uma proteína influencia as propriedades dessa, uma vez que cada um dos 20 aminoácidos comuns possui propriedades físicas e químicas singulares.<sup>(51)</sup>

### 2.1.2.2 Estrutura Tridimensional

A descrição da estrutura tridimensional de uma proteína é dada por níveis mais altos de organização estrutural (secundário, terciário e quaternário) que serão apresentados mais adiante neste texto. Mas antes de estudar estas estruturas, é importante revisar as propriedades geométricas dos grupos peptídicos que são base para todos os níveis estruturais superiores.<sup>(51)</sup>

O grupo peptídico de um polipeptídeo possui uma estrutura planar rígida, ou seja, as ligações peptídicas são essencialmente planas. Isto ocorre, pois a ligação peptídica tem um considerável caráter de dupla ligação, o que impede que ocorra uma rotação em torno dela. A distância C-N em uma ligação peptídica é de 1,32Å, que está entre os valores esperados para uma ligação C-N simples (1,49Å) e para uma ligação dupla C=N (1,27Å), conforme apresentado na Figura 7.<sup>(49, 51)</sup>



**Figura 7:** Esquema representando a configuração *trans* dos aminoácidos num fragmento polipeptídico, com os ângulos  $\phi$  e  $\Psi$  indicados, assim como os comprimentos de ligação característicos das ligações peptídicas.<sup>(50)</sup>

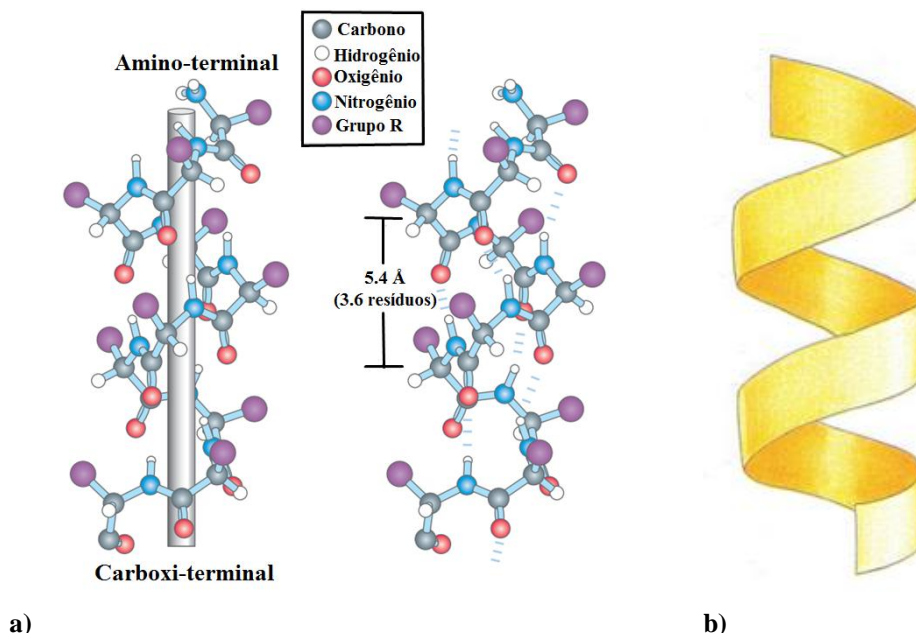
Uma ligação peptídica pode ocorrer em uma conformação *trans*, na qual os dois carbonos- $\alpha$  estão em lados opostos da ligação peptídica, ou em uma conformação *cis*, com os dois carbonos- $\alpha$  para o mesmo lado. Em proteínas, a conformação *trans* de ligações peptídicas é mais favorável e mais comum em relação à conformação *cis*. Isso se deve ao impedimento estereoquímico entre as cadeias laterais dos resíduos adjacentes que decorre da conformação *cis*.<sup>(49, 51)</sup>

### 2.1.2.2.1 Estrutura Secundária:

A conformação de uma cadeia polipeptídica é descrita pelos ângulos de torção em torno das ligações  $C_{\alpha}$ -N (denominado de ângulo phi,  $\phi$ ) e  $C_{\alpha}$ -C (denominado de ângulo psi,  $\Psi$ ), sendo C o carbono carbonílico. Certas combinações desses ângulos produzem colisões estéricas entre o oxigênio carbonílico, o hidrogênio amídico e as cadeias laterais (grupos R) dos resíduos de aminoácidos. Os valores permitidos das combinações de  $\phi$  e  $\Psi$  são definidos por um gráfico bidimensional, denominado diagrama de Ramachandran.

#### A) Estruturas Secundárias Regulares: $\alpha$ -hélices e folhas- $\beta$

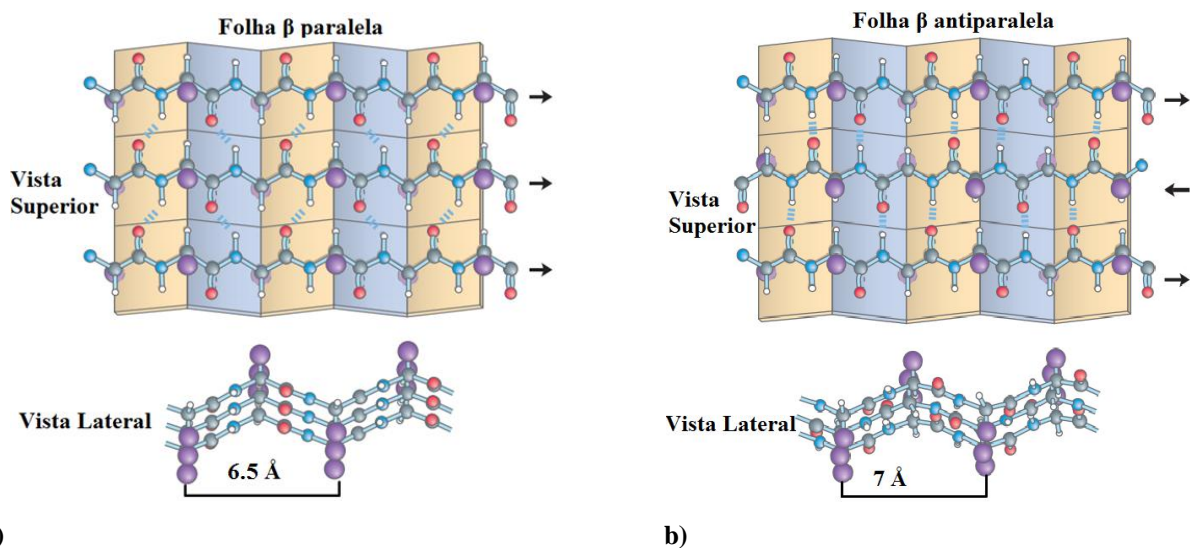
O termo estrutura secundária se refere a qualquer segmento de uma cadeia polipeptídica que possui um determinado arranjo espacial dos átomos da cadeia principal, sem considerar a relação deste com os demais segmentos da proteína.<sup>(50)</sup> Duas estruturas secundárias regulares de proteínas muito comuns e de fácil reconhecimento são as  $\alpha$ -hélices e as folhas- $\beta$  (ou folhas- $\beta$  pregueadas). Essa periodicidade, ou regularidade, se deve ao fato de que essas estruturas possuem sequências de resíduos com valores repetidos dos ângulos torcionais,  $\phi$  e  $\Psi$ , e um padrão regular de ligações de hidrogênio entre os grupos N-H e C=O da cadeia polipeptídica.<sup>(49, 51)</sup>



**Figura 8:** a) Modelo de  $\alpha$ -hélice, indicando as ligações de hidrogênio entre os grupos CO do aminoácido (n) e NH do aminoácido (n+4) e o passo da hélice (5,4 Å);<sup>(50)</sup> b) representação esquemática de uma  $\alpha$ -hélice.<sup>(53)</sup>

A  $\alpha$ -hélice (Figura 8) é o arranjo mais simples que a cadeia polipeptídica pode assumir, dada a rigidez de suas ligações peptídicas. Esta estrutura é firmemente enrolada em torno de um eixo imaginário (longitudinal em relação ao centro da hélice), de forma que as cadeias laterais dos resíduos de aminoácidos do segmento são projetadas para fora. Uma  $\alpha$ -hélice assume tipicamente para os ângulos phi e psi os valores de  $\Phi = -57^\circ$  e  $\Psi = -47^\circ$ , respectivamente.<sup>(50)</sup> As ligações de hidrogênio da cadeia principal ocorrem entre o átomo de oxigênio do grupo C=O de um resíduo (n) e o átomo de hidrogênio do grupo N-H de um resíduo (n+4), resultando em uma forte ligação de hidrogênio.<sup>(51)</sup>

Em uma  $\alpha$ -hélice, cada resíduo de aminoácido relaciona-se com o seguinte por um deslocamento (denominado *translação*) de 1,5 Å ao longo do eixo da hélice e por uma rotação de 100°. Como consequência disso, tem-se 3,6 resíduos de aminoácidos por volta de hélice. Desta forma, aminoácidos a 3 e 4 unidades de distância estão próximos um do outro em uma  $\alpha$ -hélice, enquanto aminoácidos afastados de duas unidades na sequência linear são posicionados em lados opostos da hélice (tornando improvável a possibilidade de contato entre eles). O passo de uma  $\alpha$ -hélice é equivalente ao produto da translação (1,5 Å) pelo número de aminoácidos por volta (3,6), ou seja, é igual a 5,4 Å.<sup>(49, 51)</sup>



**Figura 9:** Arranjos de folhas- $\beta$ : **a)** paralela com um período (ou a distância entre as cadeias laterais de dois resíduos de aminoácidos consecutivos de uma fita- $\beta$ ) característico de 6,5 Å; **b)** antiparalela, com um período de 7,0 Å.<sup>(50)</sup>

A *folha- $\beta$* , que lembra um conjunto de pregas, é outra estrutura secundária muito presente em proteínas e é composta por dois ou mais segmentos de cadeia polipeptídica (denominados de fitas- $\beta$ ). Neste arranjo, as ligações de hidrogênio são formadas entre os

segmentos adjacentes da cadeia polipeptídica, de forma que as cadeias laterais de aminoácidos adjacentes na cadeia apontam em sentidos opostos. Estes segmentos não necessariamente estão próximos na sequência linear de aminoácidos da proteína.<sup>(49, 50, 51)</sup> Diferente da  $\alpha$ -hélice, uma fita- $\beta$  é quase que completamente estendida. A forma como as fitas- $\beta$  se arranjam define dois tipos de folhas- $\beta$ : paralela e antiparalela (Figura 9). No primeiro caso, as cadeias polipeptídicas adjacentes se estendem no mesmo sentido e os valores ideais dos ângulos torcionais são  $\Phi = -119^\circ$  e  $\Psi = +113^\circ$ . No segundo, as cadeias se estendem em sentidos opostos e os ângulos torcionais ideais correspondem a  $\Phi = -139^\circ$  e  $\Psi = +135^\circ$ .<sup>(49, 50, 51)</sup>

### **B) Estruturas Secundárias Irregulares: voltas e alças**

Segmentos de estruturas secundárias regulares,  $\alpha$ -hélices e folhas- $\beta$ , comumente são conectados por pedaços que mudam abruptamente de direção.<sup>(51)</sup> Quase um terço dos resíduos de aminoácidos encontra-se em voltas ou alças onde ocorrem a inversão da cadeia polipeptídica. As voltas  $\beta$  ( $\beta$  turns ou volta em forma de U) conectam as extremidades de dois segmentos adjacentes de uma folha- $\beta$ . Estas estruturas são bem comuns e envolvem 4 resíduos de aminoácidos em uma volta de  $180^\circ$  (onde uma ligação de hidrogênio ocorre entre o oxigênio do grupo carbonílico do primeiro aminoácido e o hidrogênio do grupo amino do quarto aminoácido). As voltas  $\beta$  são encontradas, em geral, nas proximidades da superfície de proteínas, pois nessa região os grupos peptídicos dos aminoácidos centrais (da volta  $\beta$ ) podem estabelecer ligações de hidrogênio com a água.<sup>(50)</sup>

Em outros casos, estruturas mais elaboradas são responsáveis pela reversão da cadeia polipeptídica. As alças (também conhecidas como alças  $\Omega$ , ômega, em referência à forma de ferradura desta estrutura) são frequentemente rígidas e bem definidas. A maioria das proteínas com mais de 60 resíduos possui uma ou mais alças  $\Omega$  (com cerca de 6 a 16 resíduos de aminoácidos cada). As alças  $\Omega$ , assim como as voltas  $\beta$ , são encontradas, em geral, na superfície de proteínas.<sup>(49, 51)</sup>

#### **2.1.2.2 Estruturas Terciária e Quaternária**

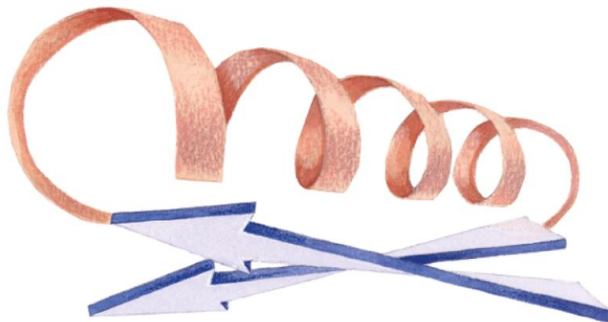
A estrutura terciária de uma proteína descreve o dobramento dos elementos estruturais secundários e especifica as posições de cada átomo na proteína, incluindo aqueles das cadeias laterais. Ou seja, é o arranjo tridimensional de todos os átomos da proteína. A estrutura quaternária refere-se ao arranjo de cadeias polipeptídicas de proteínas multicadeias, cujas subunidades são associadas de forma não-covalente. Experimentalmente a estrutura



tridimensional de uma proteína é obtida por cristalografia de raios X ou por ressonância magnética nuclear (RMN).<sup>(51, 52)</sup>

Para entender completamente a estrutura tridimensional de uma proteína é preciso analisar seu padrão de dobramento. *Motivo* (estrutura supersecundária ou dobramento) é um padrão de dobramento identificável que envolve dois ou mais elementos de estrutura secundária e a conexão (ou conexões) entre eles. Um motivo não é um elemento estrutural hierárquico entre as estruturas secundária e terciária, mas sim um padrão de dobramento que pode representar somente uma parte de uma proteína ou abranger toda a cadeia polipeptídica.<sup>(50)</sup>

Os motivos são padrões de dobramentos encontrados em muitas proteínas, sendo que um dos mais comuns é o  $\beta\alpha\beta$  (Figura 10), no qual uma  $\alpha$ -hélice conecta duas fitas- $\beta$  paralelas de uma folha- $\beta$ . Outra estrutura supersecundária comum é o motivo *grampo  $\beta$*  que consiste em fitas antiparalelas conectadas por voltas  $\beta$  firmes. Esses padrões de combinação de estruturas secundárias estão presentes em muitas proteínas e com frequência exibem funções semelhantes.<sup>(49, 51)</sup>



**Figura 10:** Exemplo de uma estrutura supersecundária comum, o motivo  $\beta\alpha\beta$ .<sup>(50)</sup>

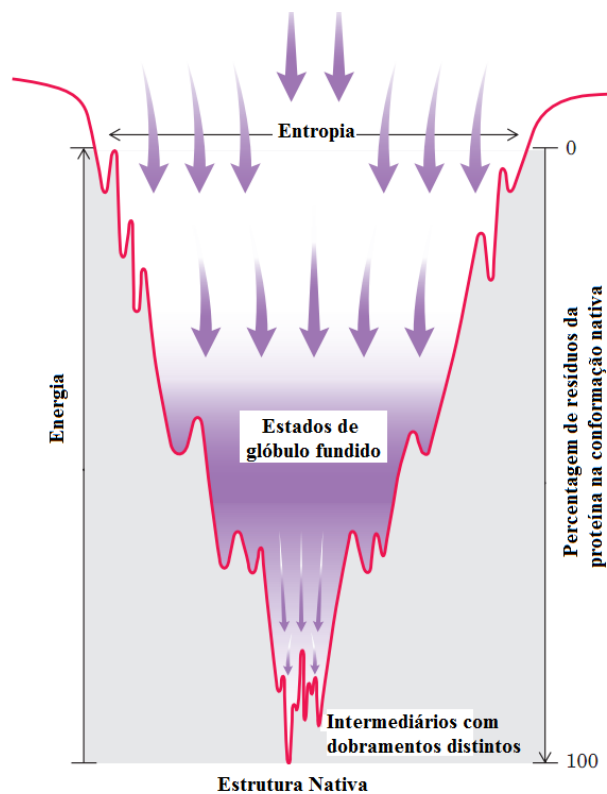
Outro termo que descreve padrões estruturais é o *domínio*. Este representa uma parte da proteína que é independentemente estável ou que se movimenta como uma entidade isolada em relação ao restante da macromolécula. Um domínio possui, em média, de 100 a 150 resíduos de aminoácidos. De modo geral, proteínas com cerca de pelo menos 200 resíduos de aminoácidos se dobras em um ou mais domínios diferentes que podem exercer diferentes funções, enquanto proteínas pequenas comumente possuem um único domínio. Considera-se que domínios com padrões de dobramento semelhantes possuem o mesmo motivo, ainda que as estruturas secundárias que constituem esses motivos ( $\alpha$ -hélices e fitas- $\beta$ ) possuam diferentes números de resíduos.<sup>(50)</sup>

### 2.1.3 Dobramento de proteínas e Estabilidade da forma nativa

Na células vivas, as proteínas são construídas a partir dos aminoácidos em uma velocidade muito alta. Mas como uma proteína chega em sua conformação nativa? Uma possibilidade à priori seria a de uma busca ao acaso (ou seja, testar todas as conformações possíveis) para encontrar a conformação energeticamente mais estável. Considerando uma proteína pequena contendo 100 resíduos e assumindo que cada um dos resíduos de aminoácidos (previamente estabelecidos) possa assumir, em média, 10 conformações diferentes, seriam possíveis  $10^{100}$  conformações diferentes para essa cadeia polipeptídica. Admitindo-se ainda que a proteína se dobra espontaneamente (através de um processo aleatório no qual são testadas todas as conformações possíveis em torno de cada uma das ligações de seu esqueleto até encontrar sua forma nativa), se cada conformação fosse testada no menor tempo possível ( $\approx 10^{-13}$  segundos; tempo de uma vibração molecular), seriam necessários cerca de  $10^{77}$  anos para testar todas as possíveis conformações desta proteína. Esse tempo seria incompatível com a vida humana, já que diariamente um organismo produz milhares de proteínas em suas formas biologicamente ativas, responsáveis pela manutenção de inúmeras funções biológicas. Desta forma, fica claro que o dobramento de proteínas não é um processo completamente aleatório, de tentativa e erro. <sup>(49, 50)</sup>

A enorme diferença entre este tempo de enovelamento calculado e o real é chamada de *paradoxo de Levinthal*, que revela que as proteínas se enovelam por uma via parcialmente definida, constituída de intermediários parcialmente corretos entre a proteína desdobrada e sua estrutura nativa. Embora o mecanismo de dobramento de uma cadeia polipeptídica seja muito complicado (e nem todos os princípios envolvidos neste processo sejam conhecidos), existem alguns modelos plausíveis, destacando-se entre eles o modelo de “nucleação e condensação”, baseado num dobramento hierárquico. Neste modelo, regiões locais da cadeia polipeptídica assumem primeiramente estruturas secundárias favorecidas (como as  $\alpha$ -hélices e as folhas- $\beta$ ). Conforme vão se formando, estas estruturas secundárias (ou, sítios de iniciação) parcialmente dobradas podem interagir, resultando em estruturas supersecundárias. Como resultado destas interações é produzido um *estado de glóbulo fundido*, que é um intermediário condensado contendo muitos dos elementos de estrutura secundária da conformação nativa, mas muitas interações de estrutura terciária incorretas. Rearranjos dentro do glóbulo fundido conduzem às interações corretas (de médio e longo alcance) entre os diferentes sítios de iniciação. Este processo segue até que sejam formados domínios completos e todo o polipeptídeo tenha se dobrado. <sup>(49, 50, 52)</sup>

Termodinamicamente, a superfície de energia livre para o processo global do enovelamento de proteínas pode ser vista como um funil (ver Figura 11). Inicialmente, um polipeptídeo não-dobrado possui muitas possibilidades conformacionais (alta entropia). À medida que o dobramento ocorre, o estreitamento do funil representa a diminuição do número de espécies conformacionais presentes, com a consequente diminuição de energia livre e entropia. As pequenas depressões ao longo das paredes do funil de energia livre representam intermediários semiestáveis, que podem tornar o processo de dobramento um pouco mais lento. Estes intermediários permanecem temporariamente aprisionados até que, por energia de ativação térmica aleatória, consigam sobrepor a barreira de energia livre e possam progredir para uma conformação mais estável, de mais baixa energia. No fundo do funil, o conjunto de intermediários dobrados é reduzido a uma única conformação nativa (ou um pequeno conjunto de conformações nativas). De fato, muitas vias de dobramento podem levar a este mesmo mínimo de energia.<sup>(49, 50, 51)</sup>



**Figura 11:** Funil de energia livre representando o processo de dobramento de uma proteína (figura adaptada do livro *Princípios de Bioquímica* de Lehninger).<sup>(50)</sup>

A estabilidade termodinâmica é assimetricamente distribuída na estrutura da proteína, ou seja, as moléculas possuem algumas regiões de alta estabilidade e outras de baixa estabilidade. Essas últimas permitem que a proteína altere sua conformação entre dois ou mais

estados. As flutuações de estabilidade das regiões internas da proteína são, comumente, fundamentais para sua função.<sup>(50)</sup>

Nem sempre, porém, uma proteína irá se dobrar espontaneamente. Para contornar este problema, as células possuem proteínas que facilitam o processo de dobramento, entre elas estão incluídas as cis-trans-prolil isomerases (que convertem ligações peptídicas cis e trans de resíduos prolina), as proteínas dissulfeto isomerase (que catalisam a formação de ligações dissulfeto de cistina) e as chaperonas moleculares. Estas últimas são proteínas que interagem com polipeptídeos parcialmente dobrados, facilitando o processo de dobramento. Na verdade, as chaperonas não alteram o resultado final do dobramento, mas sim impedem a formação de agregados protéicos antes que as cadeias polipeptídicas a serem agregadas tenham concluído seu processo de dobramento.<sup>(50, 52)</sup>

Duas classes de chaperonas vêm sendo muito estudadas, ambas presentes em organismos de bactérias a humanos. A primeira classe é da família de proteínas Hsp70 (proteínas de choque térmico) que possuem massas moleculares em torno de 70 kDa e são abundantemente encontradas em células à altas temperaturas. Essas proteínas ligam-se a polipeptídeos à medida que esses são sintetizados nos ribossomos. Desta forma, as superfícies hidrofóbicas que estariam normalmente expostas ao solvente são bloqueadas, protegendo a proteína de agregação até que a cadeia inteira tenha sido sintetizada e o dobramento possa ocorrer. Além disso, estas chaperonas protegem as proteínas de desnaturação pela temperatura. A segunda classe são as chaperoninas (ou família Hsp60, com massa molecular em torno de 60 kDa), que são estruturas longas, cilíndricas, com múltiplas subunidades que se ligam a polipeptídeos não-dobrados em seu estado de glóbulo fundido. Estima-se que de 10 a 15% das proteínas celulares, em *Escherichia Coli*, precisam de chaperoninas assistindo seu processo de dobramento em condições celulares normais.<sup>(50, 52)</sup>

#### **2.1.4 Forças que contribuem para a estabilidade de proteínas**

A estabilidade de uma proteína se refere à tendência que essa possui em manter a sua conformação nativa. Em condições fisiológicas a variação de energia livre que separa as formas dobrada e não-dobrada de uma proteína é cerca de apenas 20 a 65 kJ/mol, o que faz com que as proteínas nativas sejam apenas marginalmente estáveis. A elevada entropia conformacional da proteína não-dobrada (relacionada ao grande número de conformações possíveis dessa estrutura) aliada às ligações de hidrogênio que os grupos da cadeia polipeptídica estabelecem com o solvente (água), tendem a manter o estado não-dobrado.

Contrabalanceando estes efeitos, interações covalentes (ligações de dissulfeto) e interações fracas não-covalentes (ligações de hidrogênio, interações iônicas e de van der Waals) estabilizam a conformação nativa.<sup>(50, 52)</sup>

O ambiente intracelular altamente redutor da maioria das células, evita a formação das ligações dissulfeto. Isto faz com que muitas proteínas não possuam essas ligações. Assim, para as proteínas intracelulares da maioria dos organismos, as interações fracas são especialmente importantes para o dobramento das cadeias polipeptídicas. Em comparação às ligações covalentes, que necessitam de 200 a 460 kJ/mol para serem rompidas, as interações não-covalentes são relativamente fracas (sendo rompidas com cerca de apenas 4 a 30 kJ/mol). No entanto, mesmo sendo individualmente fracas, o grande número de interações não-covalentes de uma proteína faz com que essas predominem como forças estabilizadoras que conduzem o dobramento protéico. Portanto, de maneira geral quanto maior o número de interações fracas, mais estável é a conformação da proteína.<sup>(50, 52)</sup>

As ligações de hidrogênio se formam quando um átomo de hidrogênio (ligado covalentemente a um átomo eletronegativo, denominado doador de hidrogênio) interage com um átomo eletronegativo (denominado aceptor de hidrogênio). A força de uma ligação de hidrogênio (tipicamente de 4-29 kJ/mol) depende da distância entre os átomos doador e aceptor, assim como da orientação estabelecida entre eles e o átomo de hidrogênio. As relações geométricas correspondentes a energias mais altas são aquelas em que os átomos doador e aceptor estão colinearmente dispostos. As ligações de hidrogênio parecem contribuir pouco para a estabilidade da conformação nativa. No entanto, a presença de um grupo aceptor (ou doador) de ligação de hidrogênio que não possua um par para se ligar no núcleo hidrofóbico de uma proteína possui um efeito tão desestabilizador que conformações contendo este grupo são termodinamicamente insustentáveis. Isto, aliado ao fato de que as ligações de hidrogênio apresentam-se otimizadas em estruturas secundárias ( $\alpha$ -hélices e fitas- $\beta$ ), faz com que elas sejam importantes no processo de dobramento de proteínas.<sup>(50, 52)</sup>

As interações eletrostáticas (ou ligações iônicas) entre grupos carregados de cadeias laterais podem exercer tanto um efeito estabilizante quanto desestabilizante nas conformações protéicas e são importantes na ligação de substratos à proteínas. Essas interações podem ser repulsivas (entre cargas de mesmo sinal) ou atrativas (entre cargas de sinais opostos), e sua ocorrência depende do pH do meio (que altera o estado de protonação dos resíduos de aminoácidos da cadeia polipeptídica). A força de uma ligação iônica aumenta à medida que essa se desloca para um ambiente com constante dielétrica mais baixa. Uma vez que a água possui um elevado valor de constante dielétrica ( $\epsilon = 78,8$ ; à 25 °C)<sup>(54)</sup>, as interações

eletrostáticas nela são relativamente fracas em comparação às do interior de uma proteína, onde a constante dielétrica é baixa. De maneira geral, as ligações iônicas ocorrem geralmente próximas do núcleo hidrofóbico e podem estabilizar significativamente a estrutura de uma proteína, além de limitar sua flexibilidade estrutural e conferir-lhe uma singularidade estrutural. A maioria dos grupos carregados, no entanto, permanecem na superfície (onde não interagem com outros grupos carregados da proteína devido à alta constante dielétrica da água), sendo estabilizados por ligações de hidrogênio e interações polares com a água. Essas interações geram forças dominantes que colocam a maioria dos grupos carregados de uma proteína no lado de fora de estruturas dobradas.<sup>(50, 52)</sup>

As interações de van der Waals, por sua vez, são as mais fracas dentre as interações não-covalentes (atuando entre dipolos temporários) e envolvem qualquer resíduo de aminoácido da cadeia polipeptídica. Particularmente entre grupamentos apolares essas são as interações predominantes. Para chegar à conformação nativa, podem ser necessárias milhares de interações fracas de van der Waals. Desta forma, é substancial a contribuição total dessas interações para a estabilidade de uma estrutura dobrada.<sup>(50, 52)</sup>

Por fim, o efeito hidrofóbico é de fundamental importância na estabilização da estrutura nativa. Esse efeito se deve às interações entre a água circundante e os grupos não-polares da cadeia polipeptídica. Quando uma proteína é colocada em um ambiente aquoso, uma camada de solvatação se forma ao redor da macromolécula. A ausência de ligações de hidrogênio entre os grupos não-polares e as moléculas de água (mais do que interação favorável entre os grupamentos não-polares) é o fator principal que leva à aproximação entre as cadeias laterais não-polares de uma proteína. Essa aproximação, por sua vez, promove a expulsão de moléculas de água que circundam as cadeias laterais não-polares, diminuindo, assim, a área superficial da cadeia polipeptídica exposta ao solvente (ou seja, diminui a camada de solvatação). Com isso, a entropia do sistema aumenta (em função dos graus de liberdade translacionais das moléculas de água). Este aumento de entropia da solução aquosa é termodinamicamente favorável e age como a força motriz para a aproximação de grupos não-polares em solvente aquoso. Desta forma, as cadeias laterais de aminoácidos hidrofóbicos tendem a se agrupar no interior das proteínas (“escondendo-se” das moléculas de água), levando à formação de um núcleo protéico altamente empacotado.<sup>(50, 52)</sup>

Em resumo, as interações fracas (não-covalentes) são fundamentais para estabilizar a estrutura de uma proteína, sendo o efeito hidrofóbico um dos principais contribuintes nesse processo de estabilização. As ligações de hidrogênio, as ligações iônicas e as interações de van der Waals são otimizadas nas estruturas termodinamicamente mais estáveis.<sup>(50)</sup>

## 2.2 BANCOS DE DADOS

Atualmente, o número de estruturas tridimensionais conhecidas de proteínas está na casa das dezenas de milhares e mais que dobra a cada dois anos.<sup>(50)</sup> Essas estruturas vêm sendo depositadas em bancos de dados disponíveis na Internet.<sup>(51)</sup> O principal banco de dados de estruturas de macromoléculas biológicas é o *Protein Data Bank* (PDB).<sup>(55)</sup> Este contém estruturas de proteínas e ácidos nucleicos. As estruturas depositadas no PDB são determinadas experimentalmente por técnicas como cristalografia por difração de raio X e ressonância magnética nuclear (RMN).<sup>(56)</sup>

Para identificar as estruturas deste banco de dados é empregado um código, chamado PDB ID, de quatro caracteres. Os arquivos de dados no PDB contêm as coordenadas espaciais de cada átomo cuja posição foi determinada. Estas coordenadas atômicas podem ser convertidas em uma imagem da macromolécula com a ajuda de programas de visualização (como o SPDBV<sup>(57)</sup>, o RASMOL<sup>(58)</sup>, entre outros). Arquivos de dados adicionais fornecem informações de como as estruturas foram determinadas e sua precisão.<sup>(50)</sup>

## 2.3 DINÂMICA MOLECULAR

As várias forças atuantes na estabilização de proteínas e a sua usual representação como uma estrutura única podem dar a falsa impressão de que proteínas têm estruturas fixas e rígidas. Na verdade, cada átomo em uma proteína está em movimento constante (apesar das limitações impostas pelas estruturas secundária e terciária e pelas ligações covalentes), como moléculas num fluido. As flutuações verificadas em cadeias laterais individuais, domínios ou até mesmo subunidades inteiras podem estar relacionadas a movimentos intramoleculares aleatórios ou podem ocorrer como resposta a um sinal, como a ligação de uma pequena molécula.<sup>(51, 52)</sup>

Cálculos teóricos de dinâmica molecular descrevem as mudanças nas coordenadas de átomos de uma proteína, indicando que sua estrutura nativa provavelmente consiste em uma gama de conformações rapidamente interconversíveis com estabilidades similares. O comportamento dinâmico de proteínas é a base para mudanças conformacionais que determinam fenômenos como, por exemplo, a inibição de uma enzima por um ligante ou droga e a geração de efeitos alostéricos em hemoglobinas. Assim, as proteínas são moléculas flexíveis que apresentam oscilações significativas do ponto de vista estrutural.<sup>(51, 52)</sup>

Uma das principais ferramentas para o estudo de proteínas do ponto de vista atômico-molecular é a dinâmica molecular (sigla DM). Com esta técnica podem ser estudados desde

sistemas homogêneos como gases, fluidos supercríticos, líquidos, soluções, a sistemas pouco (ou nada) homogêneos como biomembranas, lipídios, ácidos nucleicos e proteínas.<sup>(48)</sup>

Em dinâmica molecular, sucessivas configurações de um sistema são geradas pela integração numérica das equações de movimento de Newton. O resultado é uma trajetória que especifica como as posições e as velocidades das partículas do sistema variam com o tempo,  $t$ .<sup>(59)</sup> Entre os vários algoritmos existentes para integrar as equações clássicas está o algoritmo *Leap-Frog* (pulo da rã). Este algoritmo, que foi empregado no presente trabalho, consiste em atualizar as velocidades em tempos intermediários àqueles em que são atualizadas as posições, conforme mostrado nas equações a seguir:<sup>(59, 60)</sup>

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v}\left(t + \frac{\Delta t}{2}\right) \quad (1)$$

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\Delta t}{m} \mathbf{f}(t) \quad (2)$$

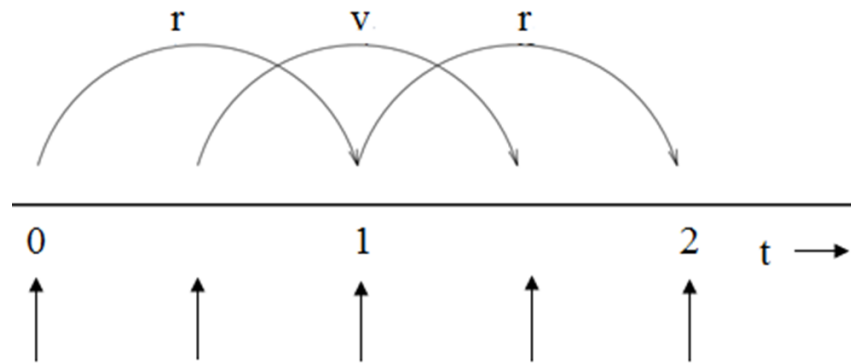
Sendo  $\mathbf{r}$  o vetor de posição da partícula;  $\mathbf{v}$  a velocidade da partícula;  $\mathbf{f}$  a força atuando na partícula;  $m$  a massa da partícula; e  $\Delta t$  o passo temporal. Para iniciar o algoritmo são necessários os valores de posições e velocidades em duas etapas consecutivas. Primeiramente, as velocidades,  $\mathbf{v}\left(t + \frac{\Delta t}{2}\right)$ , num meio passo temporal futuro,  $\left(t + \frac{\Delta t}{2}\right)$ , são calculadas a partir das velocidades,  $\mathbf{v}\left(t - \frac{\Delta t}{2}\right)$ , das partículas num meio passo temporal passado,  $\left(t - \frac{\Delta t}{2}\right)$ , e das forças que atuam nas posições das partículas num instante atual,  $t$ . Então, as posições,  $\mathbf{r}(t + \Delta t)$ , em um passo temporal futuro,  $(t + \Delta t)$ , são deduzidas a partir das velocidades que acabaram de ser calculadas,  $\mathbf{v}\left(t + \frac{\Delta t}{2}\right)$ , e das posições,  $\mathbf{r}(t)$ , no instante de tempo atual,  $t$ , usando a equação.<sup>(59)</sup>

Em resumo, são utilizadas a força atuando sobre uma partícula em um instante  $t$  e a velocidade num instante  $t - 0,5\Delta t$  para atualizar a velocidade para um instante  $t + 0,5\Delta t$ . A seguir, essa nova velocidade  $t + 0,5\Delta t$  é utilizada juntamente com a posição do instante  $t$  para atualizar a posição da partícula para um instante  $t + \Delta t$ . Para esta nova posição no instante  $t + \Delta t$  é calculada a força atuando sob esta partícula e, então, esta nova força  $t + \Delta t$  é empregada juntamente com a velocidade  $t + 0,5\Delta t$  para atualizar novamente a velocidade (agora para um instante  $t + 1,5\Delta t$ ), e assim por diante. Para inicializar o algoritmo, as posições iniciais das partículas,  $r(t)$ , são obtidas das estruturas experimentalmente determinadas



depositadas em bancos de dados, como o PDB, e as velocidades iniciais são atribuídas de acordo com uma função de distribuição de Maxwell-Boltzmann.<sup>(48)</sup>

O nome “pulo de rã” vem do fato de que as velocidades são calculadas em tempos intermediários àqueles nos quais são calculadas as posições. Assim, as velocidades e as posições ficam saltando intercaladamente umas sobre as outras, como se fossem rãs, conforme mostra o esquema da Figura 12.<sup>(61)</sup>



**Figura 12:** Esquema do método de integração *Leap-Frog*. Este método é assim denominado, pois as velocidades e as posições das partículas são atualizadas em valores sistematicamente intercalados, como se as velocidades saltassem como “rãs” sob as posições e vice-versa (imagem adaptada do Manual do GROMACS 4.0).<sup>(61)</sup>

A força (empregada no cálculo das velocidades na Equação 2) atuando em uma partícula ( $i$ ) se relaciona com o gradiente da função energia potencial, que atua nas coordenadas desta partícula, pela seguinte equação:<sup>(59)</sup>

$$\mathbf{f}_i = -\nabla V(\mathbf{r}_i) \quad (3)$$

Sendo  $\mathbf{f}_i$  a força que atua em uma partícula  $i$  qualquer,  $\nabla$  o operador gradiente e  $V(\mathbf{r}_i)$  a energia potencial que atua em uma dada posição da partícula  $i$  em um dado instante de tempo. Essa energia potencial depende somente das coordenadas das partículas.

Uma vez que cálculos exatos de energia potencial e força para as partículas de um dado sistema de estudo não são acessíveis, metodologias alternativas surgiram como forma de sanar esses problemas. Dessa forma, foram desenvolvidos modelos de “campos de força” com o objetivo de fornecer uma aproximação satisfatória para a função energia potencial.

## 2.4 CAMPOS DE FORÇA

Embora o Hamiltoniano quântico para um sistema biomolecular possa ser precisamente descrito, a resolução das equações quânticas para esse ainda é proibitiva em questão de custo computacional. De maneira alternativa, as equações de movimento podem ser adequadamente resolvidas empregando um Hamiltoniano ( $H$ ) descrito por mecânica clássica. Desta forma, o Hamiltoniano pode ser representado como uma soma das energias cinética ( $K$ ) e potencial ( $V$ ) totais do sistema.<sup>(62)</sup>

$$H(\mathbf{p}, \mathbf{r}; m) = K(\mathbf{p}; m) + V(\mathbf{r}) \quad (4)$$

Sendo:  $\mathbf{p}$  = momento ( $\mathbf{p} = m\mathbf{v}$ );  $\mathbf{r}$  = posição das do sistema;  $m$  = massas das partículas do sistema.

A energia cinética total é descrita como a soma das energias cinéticas de todas as partículas do sistema, sendo função do momento,  $\mathbf{p}$ , e das massas,  $m$ , e independente das coordenadas atômicas do sistema.<sup>(62)</sup>

$$K(\mathbf{p}; m) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} = \sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 \quad (5)$$

Sendo  $N$  o número total de partículas do sistema de estudo;  $m_i$  a massa da partícula  $i$ ; e  $\mathbf{v}_i$  a velocidade que a partícula  $i$  em um determinado instante de tempo da simulação. Essas velocidades podem ser obtidas a partir do emprego dos algoritmos de integração das equações de movimento clássicas, conforme descrito no capítulo anterior.

A energia potencial total, por sua vez, é uma função mais complicada que envolve diferentes interações entre as partículas de um sistema, se relacionando às forças que atuam sob essas partículas, conforme já definido na Equação (3).<sup>(62)</sup> Uma vez que a função,  $V(\mathbf{r})$ , descrita em (3) não possui uma forma analítica conhecida, ela pode ser aproximada pela aditividade do potencial entre pares:<sup>(63)</sup>

$$V(\mathbf{r}) = \sum_{i=1}^{N-1} \sum_{j>i}^N V_{ij}(\mathbf{r}) \quad (6)$$

Sendo  $V(\mathbf{r})$  a função potencial total descrita pelo campo de força;  $V_{ij}(\mathbf{r})$  o potencial atuando entre as partículas  $i$  e  $j$ ; e  $N$  o número total de partículas do sistema de estudo.

O potencial total também pode ser expresso como a soma de um termo relativo aos potenciais ligados ( $V_{\text{lig}}$ ) e um termo para os potenciais não-ligados ( $V_{\text{não-lig}}$ ), conforme apresentado na equação a seguir:<sup>(48)</sup>

$$V(\mathbf{r}) = V_{\text{Total}} = V_{\text{lig}} + V_{\text{não-lig}} \quad (7)$$

Com:

$$V_{\text{lig}} = V_r + V_\theta + V_\omega + V_\phi \quad (8)$$

$$V_{\text{não-lig}} = V_{\text{vdW}} + V_{\text{Coul}} \quad (9)$$

Sendo:  $V_r$  o potencial de interação para o estiramento de ligação entre dois átomos (interações tipo 1-2);  $V_\theta$  o potencial de interação para deformações angulares entre três átomos;  $V_\omega$  o potencial de interação que representa a energia para alterar o ângulo  $\omega$  entre dois planos de átomos formados por um átomo central ligado a outros três (diedro impróprio);  $V_\phi$  o potencial torcional de rotação em torno de uma ligação (diedro próprio; interações 1-4);  $V_{\text{vdW}}$  o potencial de van de Waals para descrever interações inter ou intramoleculares de longo alcance; e  $V_{\text{Coul}}$  o termo que trata as interações eletrostáticas.

Assim, um campo de força nada mais é do que a forma funcional da energia potencial total aliada ao conjunto de parâmetros que a definem. Obviamente, essa função e seus parâmetros variam conforme o campo de força empregado (cuja qualidade depende evidentemente do sistema e do tipo de propriedade estudados).<sup>(62)</sup> Entre os inúmeros campos de força usados em DM, os mais utilizados para os sistemas biológicos são: CHARMM, AMBER, OPLS, ENCAD e GROMOS.<sup>(48)</sup>

Detalhamentos dos campos de força GROMOS 96 53a6<sup>(62)</sup>, AMBER99 $\phi$ <sup>(64)</sup>, OPLS-AA<sup>(65)</sup> (empregados no presente trabalho) e ENCAD<sup>(47)</sup> (empregado por Alonso e colaboradores<sup>(33)</sup> nos estudos de influência de pH no processo de conversão estrutural da proteína príon celular) estão descritos nos próximos subcapítulos.

#### 2.4.1 ENCAD

O campo de força ENCAD (do inglês, *ENergy Calculations And Dynamics*) foi desenvolvido a partir da década de 1970, e seu esquema, desde então, permaneceu praticamente o mesmo. O que tem sido modificado, ao longo dos anos, são os detalhes dos cálculos de energia (como a truncagem), a escolha dos graus de liberdade e o modelo de água.<sup>(47)</sup> A energia potencial total deste campo de força assume a seguinte forma funcional:

$$V_{lig} = \sum_{n=1}^{N_r} K_{rn} (r_n - r_{0n})^2 + \sum_{n=1}^{N_\theta} K_\theta (\theta_n - \theta_{0n})^2 + \sum_{\text{diedral}} K_\phi^i \{1 - \cos[n^i (\phi_i - \phi_0^i)]\} \quad (10)$$

$$V_{\text{n\~{a}o-lig}} = \sum_{i=1}^{N-1} \sum_{j=1}^N \left[ A_{sc} \epsilon^{ij} \left( \frac{r_0^{ij}}{r_{ij}} \right)^{12} - 2 \epsilon^{ij} \left( \frac{r_0^{ij}}{r_{ij}} \right)^6 - S_{vdw}^A(r_{ij}) \right] + 332 \sum_{i=1}^{N-1} \sum_{j=1}^N \left[ \frac{q_i q_j}{r_{ij}} - S_{el}^A(r_{ij}) \right] \quad (11)$$

Os tr\~{e}s termos apresentados na Equa\~{c}\~{o} (10) descrevem as intera\~{c}\~{o}es ligadas atuando entre \~{a}tomos separados por uma, duas e tr\~{e}s liga\~{c}\~{o}es covalentes, respectivamente. O primeiro termo \~{e} um potencial harm\~{o}nico e descreve as liga\~{c}\~{o}es, sendo que  $r_n$  representa a dist\~{a}ncia entre dois \~{a}tomos ligados, com um valor de equil\~{b}rio,  $r_{0n}$ , e uma constante de for\~{c}a de liga\~{c}\~{o}  $K_{rn}$ . O segundo termo \~{e} tamb\~{e}m um potencial harm\~{o}nico e descreve o \~{a}ngulo entre tr\~{e}s \~{a}tomos, sendo  $\theta_n$  este \~{a}ngulo,  $\theta_{0n}$  o seu valor de equil\~{b}rio e  $K_\theta$  a constante de for\~{c}a do \~{a}ngulo de liga\~{c}\~{o}.  $N_r$  e  $N_\theta$  representam, respectivamente, o n\~{u}mero total de liga\~{c}\~{o}es e de \~{a}ngulos de liga\~{c}\~{o}. O terceiro termo descreve os \~{a}ngulos diedrais (pr\~{o}prios e impr\~{o}prios) e \~{e} um potencial trigonom\~{e}trico, sendo representado por uma ou mais fun\~{c}\~{o}es cossen\~{o}ides, sendo  $\phi_i$  o \~{a}ngulo diedral,  $\phi_0^i$  o seu valor de equil\~{b}rio,  $K_\phi^i$  a altura da barreira torcional do  $i$ -\~{e}simo \~{a}ngulo diedral e  $n^i$  a periodicidade.

Os dois termos em (11) representam a parte da energia potencial que lida com as intera\~{c}\~{o}es n\~{a}o-ligadas entre pares de \~{a}tomos separados por tr\~{e}s ou mais liga\~{c}\~{o}es ao longo da estrutura covalente. As intera\~{c}\~{o}es de van der Waals s\~{a}o representadas por potenciais de Lennard-Jones (12/6), sendo  $r_{ij}$  a separa\~{c}\~{a}o entre os \~{a}tomos,  $r_0^{ij}$  o par\~{a}metro de di\~{a}metro de Lennard-Jones e  $\epsilon^{ij}$  o par\~{a}metro de energia de Lennard-Jones.  $A_{sc}$  \~{e} um fator de escala usado para reduzir a repuls\~{a}o de van der Waals de forma a compensar a redu\~{c}\~{a}o da atra\~{c}\~{a}o provocada pelo truncamento do potencial. O termo eletrost\~{a}tico \~{e} representado por um potencial de Coulomb, que pode ser atrativo ou repulsivo dependendo dos sinais dos par\~{a}metros das cargas parciais,  $q_i$  e  $q_j$ , dos \~{a}tomos. Os potenciais de van der Waals e eletrost\~{a}tico s\~{a}o modificados, respectivamente, pelas fun\~{c}\~{o}es Shifting  $S_{vdw}^A$  e  $S_{el}^A$ , que compensam a truncagem da intera\~{c}\~{a}o \~{a} dist\~{a}ncia de *cutoff*,  $r_c$ .<sup>(47)</sup>

O campo de for\~{c}a ENCAD inclui todos os \~{a}tomos de hidrog\~{e}nio na descri\~{c}\~{a}o do seu modelo (sendo por isso, um modelo *all-atom*) e sua parametriza\~{c}\~{a}o empregou simula\~{c}\~{o}es com solvente expl\~{i}cito. Al\~{e}m disso, utiliza um modelo de \~{a}gua, conhecido como F3C (com um \~{a}ngulo e duas liga\~{c}\~{o}es flex\~{i}veis). Estudos mais extensivos mostraram, ainda, que este modelo de \~{a}gua \~{e} capaz de reproduzir dados termodin\~{a}micos, estruturais e din\~{a}micos da \~{a}gua.<sup>(47)</sup>

Entre as vantagens deste modelo, destacam-se: a conservação da energia como forma de detectar e prevenir erros; a inclusão de todos os graus de liberdade (hidrogênios explicitamente considerados; modelo de água flexível; etc.); e o uso de truncamento para ser computacionalmente mais eficiente e permitir longas simulações sem recursos especiais. Assim, visando um protocolo para simular trajetórias longas, este campo de força se concentra mais em aperfeiçoar os detalhes de cálculos (conservação de energia, flexibilidade total, truncamento, etc.) do que em tentar encontrar os melhores valores para os parâmetros que descrevem a função de energia potencial total.<sup>(47)</sup>

Por fim, o campo de força ENCAD tem sido notavelmente utilizado em estudos de dobramento de proteínas, mas também é empregado com sucesso em simulações de água pura, íons em água e DNA em solução.<sup>(47)</sup>

## 2.4.2 GROMOS

A expressão para o potencial total,  $V_{\text{Total}}$ , para o campo de força GROMOS (*GRO*ningen *MO*lecular *SI*mulation, do inglês) 96 53a6<sup>(62)</sup> está separada nos termos ligado e não-ligado, conforme apresentado a seguir:

$$V_{\text{lig}} = \frac{1}{4} \sum_{n=1}^{N_r} K_{r_n} (r_n^2 - r_{0n}^2)^2 + \frac{1}{2} \sum_{n=1}^{N_\theta} K_{\theta_n} (\cos \theta_n - \cos \theta_{0n})^2 + V_{\text{diedral}} \quad (12)$$

$$V_{\text{diedral}} = \frac{1}{2} \sum_{\text{impróprio}} K_{\xi_n} (\xi_n - \xi_{0n})^2 + \sum_{\text{próprio}} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)] \quad (13)$$

$$V_{\text{não-lig}} = \sum_{i=1}^{N-1} \sum_{j=1}^N \left[ \left( \frac{C12_{ij}}{r_{ij}^{12}} \right) - \left( \frac{C6_{ij}}{r_{ij}^6} \right) \right] + \sum_{i=1}^{N-1} \sum_{j>i}^N \frac{1}{4\pi\epsilon\epsilon_0} \left[ \left( \frac{q_i q_j}{r_{ij}} \right) \right] \quad (14)$$

$$C12_{ij} = (C12_{ii} \cdot C12_{jj})^{\frac{1}{2}} \quad (15)$$

$$C6_{ij} = (C6_{ii} \cdot C6_{jj})^{\frac{1}{2}} \quad (16)$$

Sendo:  $K_{\xi_n}$  = constante de força de diedro impróprio;  $\xi_n$  = ângulo diedro impróprio;  $\xi_{0n}$  = ângulo de diedro impróprio de equilíbrio;  $K_{\varphi_n}$  = altura da barreira torcional do diedro próprio;  $\delta_n$  = é o deslocamento de fase, restrito à 0 ou  $\pi$ ;  $m_n$  = multiplicidade do ângulo diedro próprio;  $\varphi_n$  = ângulo diedro próprio;  $\epsilon$  é a constante dielétrica do meio;  $\epsilon_0$  é a permissividade no vácuo.

Em (12) o primeiro termo é um potencial biquadrático (não-harmônico) e o segundo termo é um potencial harmônico no cosseno. O terceiro termo em (12), que descreve os diedros, é decomposto em dois termos em (13). O primeiro termo em (13) modela os diedros impróprios, descritos por um potencial harmônico. Enquanto o segundo termo em (13) modela os diedros próprios, descritos por potenciais trigonométricos.

O primeiro termo em (14) modela as interações de VDW (via potenciais de Lennard-Jones). O campo de força GROMOS é do tipo *united-atom* (ou seja, os hidrogênios diretamente ligados a um carbono alifático e esse carbono são descritos como um único sítio) e os parâmetros  $C6_{ij}$  e  $C12_{ij}$  para o par de átomos  $ij$  são obtidos de  $C12_{ii}$ ,  $C12_{jj}$ ,  $C6_{ii}$  e  $C6_{jj}$  por regras de combinação geométricas (definidas para cada tipo de átomo), apresentadas nas Equações (15) e (16). O segundo termo em (14) modela as interações eletrostáticas via potenciais de Coulomb.

A filosofia do campo de força GROMOS assume um compromisso de descrever precisamente a energia de interação como função da conformação molecular, ao mesmo tempo que emprega uma forma funcional relativamente simples (que minimiza o número de parâmetros necessário, limitando o custo computacional de avaliar a energia potencial). Os parâmetros das interações ligadas (comprimentos de ligação, ângulos de ligação, etc.) são obtidos a partir de dados espectroscópicos e de cristalografia para moléculas pequenas. Já as interações não-ligadas (eletrostáticas e de van der Waals) são muito mais difíceis de parametrizar.

Os valores dos parâmetros de interações não-ligadas, para o campo de força GROMOS, foram inicialmente obtidos a partir de dados de cristalografia e de polarizabilidades atômicas. Estes valores, por sua vez, foram ajustados de modo que as distâncias experimentais e as energias de interação (de pares individuais de átomos de grupos funcionais) minimizassem a energia de conformação.

Com o aumento do poder computacional, tornou-se viável ainda a utilização de abordagens de mecânica estatística para parametrizar as interações não-ligadas, de forma a reproduzir propriedades termodinâmicas, tais como a densidade, o calor de vaporização e a entalpia livre de solvatação de pequenas moléculas na fase condensada em temperatura e pressão fisiológicas. Além disso, dados obtidos a partir de métodos quânticos *ab initio* e dados experimentais foram utilizados para ajustar, por exemplo, a distribuição de ângulos torcionais em torno de ligações específicas.

O conjunto de parâmetros 53A6 foi desenvolvido visando a obtenção de um campo de força aplicável tanto para simulações de proteínas como de membranas. Para isso, muitos

parâmetros foram redefinidos com relação às versões anteriores: todos os tipos de átomos, ligações, ângulos de ligação, diedros próprios e impróprios foram reenumerados e vários tipos novos foram adicionados. Além disso, a parametrização 53A6 foi especificamente adaptada para reproduzir a entalpia livre de hidratação (que é uma propriedade termodinâmica-chave em muitos dos processos biomoleculares) em água SPC<sup>(66)</sup> para um conjunto de análogos de aminoácidos (com uma média de erro absoluto baixa). O comportamento de partição destes análogos entre água e um ambiente hidrofóbico também é corretamente reproduzido, assim como as entalpias livre de solvatação para os solventes ciclohexano e água. Assim, dadas as melhorias nas propriedades de hidratação e solvatação, o emprego deste conjunto de parâmetros 53A6, em princípio, deve produzir melhores resultados (em relação aos conjuntos de parâmetros anteriores) para as constantes de interação entre biomoléculas em fase aquosa e para simulações de dobramento de polipeptídeos.

### 2.4.3 AMBER

O potencial total,  $V(r)$ , para o campo de força AMBER99 $\phi^{(64)}$  é descrito pela soma dos potenciais ligado e não-ligado, apresentados, respectivamente, nas Equações (17) e (18) a seguir:

$$V_{\text{lig}} = \frac{1}{2} \sum_{n=1}^{N_r} K_{r_n} (r_n - r_{0n})^2 + \frac{1}{2} \sum_{n=1}^{N_\theta} K_{\theta_n} (\theta_n - \theta_{0n})^2 + \sum_{i=1}^{N_\Phi} \sum_{n=1}^5 K_{\Phi_i} [\cos(\Phi_i - \pi)]^n \quad (17)$$

$$V_{\text{não-lig}} = \sum_{i=1}^{N-1} \sum_{j=1}^N \left\{ 4\epsilon^{ij} \alpha_{LJ,ij} \left[ \left( \frac{r_0^{ij}}{r_{ij}} \right)^{12} - \left( \frac{r_0^{ij}}{r_{ij}} \right)^6 \right] \right\} + \sum_{i=1}^{N-1} \sum_{j=1}^N \alpha_{\text{Coul},ij} \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \quad (18)$$

Sendo:  $N_\phi$  é o número total de diedros próprios,  $K_{n,i}$  é a altura da barreira torcional do  $i$ -ésimo ângulo diedral;  $\Phi_i$  é o  $i$ -ésimo diedro próprio;  $\alpha_{\text{Coul},ij}$  e  $\alpha_{LJ,ij}$  são nulos para interações 1-2 e 1-3, para interações 1-4 assumem, respectivamente, valores padrão de 0,5 e 0,8333 e, por fim, assumem valor unitário para as demais interações.

O primeiro e o segundo termo em (17) são potenciais harmônicos, enquanto o terceiro termo é um potencial de Rieckaert-Bellemans<sup>(61)</sup>. O primeiro termo em (18) é um potencial de Lennard-Jones e o segundo termo é um potencial de Coulomb.

O desenvolvimento do campo de força AMBER começou na primeira metade da década de 1980. Em sua versão original, este campo de força possuía uma representação (*united-atom*) das cadeias alifáticas nas quais um carbono, e os hidrogênios diretamente ligados a ele, era representado como um único sítio. Na parametrização original, as cargas foram derivadas de métodos de química quântica ao nível HF STO-3G (ajustando as cargas parciais atômicas à potenciais eletrostáticos quânticos). As constantes de força e comprimentos e ângulos de ligação foram tomados de estruturas cristalinas e adaptados para se ajustar aos modos de frequências normais de um dado número de peptídeos. As constantes de força de torção foram ajustadas para combinar com as barreiras torcionais extraídas de experimentos ou de cálculos quânticos.<sup>(67)</sup>

Em 1990, em função do aumento do poder computacional, foram realizadas melhorias nos parâmetros e o campo de força foi estendido para outras moléculas que não proteínas. Assim, surgiu o campo de força AMBER94, desenvolvido para obter potenciais adequados para as simulações em fase condensada (ou seja, levou em consideração as interações soluto-solvente e solvente-solvente). Empregando metodologias quânticas mais refinadas e tamanhos de fragmentos maiores, este campo de força forneceu um procedimento geral para um cálculo rápido de cargas (empregando HF/6-31G\*) para todos os 20 aminoácidos.<sup>(67)</sup>

No que diz respeito aos parâmetros de Lennard-Jones, esses foram estabelecidos, primeiramente, em referência às densidades e calores de vaporização de simulações em estado líquido. Desta forma, somente um conjunto pequeno de parâmetros (12/6) foram necessários para conseguir uma concordância razoável com os experimentos.<sup>(67)</sup> Os últimos parâmetros ajustados (como acontece na parametrização da maioria dos campos de força) foram os dos potenciais torcionais suaves sobre ligações simples. Isto se deve ao fato de que o perfil energético de rotação de ângulos torcionais depende fundamentalmente de interações não-ligadas. Um conjunto chave de parâmetros torcionais são aqueles para os ângulos  $\Phi$  e  $\Psi$  da cadeia polipeptídica, uma vez que esses afetam muitos aminoácidos e influenciam pesadamente as energias relativas de hélices, folhas- $\beta$  e voltas de proteínas. Os parâmetros do AMBER94 foram ajustados para representar pontos de mapas de dipeptídeos para a glicina e a alanina (computados através de metodologias quânticas). Algumas contabilizações de efeitos de longo alcance foram feitas no AMBER96, no qual os potenciais  $\Phi$  e  $\Psi$  foram ajustados para tetrapeptídeos assim como dipeptídeos com energias conformacionais obtidas via mecânica quântica.<sup>(67)</sup>

Por fim, o campo de força AMBER99 $\phi$  (utilizado no presente trabalho) nasceu da necessidade de corrigir alguns pontos cruciais da versão AMBER99. Esta apresentava baixo



teor helicoidal em comparação com o campo de força AMBER-94, devido a barreiras adicionadas sobre o grau de liberdade rotacional  $\phi$ . Os desenvolvedores do AMBER resolveram, então, remover esses obstáculos no AMBER-99 empregando o potencial de torção  $\phi$  do AMBER-94 com o objetivo de melhor reproduzir dados experimentais termodinâmicos e cinéticos de hélices.<sup>(64)</sup>

#### 2.4.4 OPLS

A expressão para a função potencial total para o campo de força OPLS-AA (do inglês, *Optimized Parameters for Liquid State Simulations All-Atom*) é a soma dos potenciais não-ligado ( $V_{\text{não-ligado}}$ ), de estiramento de ligação ( $V_r$ ), de ângulo de ligação ( $V_\theta$ ) e de torção ( $V_{\text{torcional}}$ ). O potencial não-ligado representa a soma das contribuições dos potenciais de Coulomb e de Lennard-Jones tanto para as interações inter como intramoleculares, conforme a equação a seguir:<sup>(65)</sup>

$$V_{\text{lig}} = \sum_{n=1}^{N_r} K_{rn} (r_n - r_{0n})^2 + \sum_{n=1}^{N_\theta} K_{\theta n} (\theta_n - \theta_{0n})^2 + V_{\text{diedral}} \quad (19)$$

$$V_{\text{diedral}} = \sum_i \left\{ \frac{V_1^i}{2} [1 + \cos(\phi_i)] + \frac{V_2^i}{2} [1 - \cos(2\phi_i)] + \frac{V_3^i}{2} [1 + \cos(3\phi_i)] \right\} \quad (20)$$

$$V_{\text{não-ligado}} = V_{\text{VDW}} + V_{\text{Coul}} = \sum_{i=1}^{N-1} \sum_{j=1}^N \left\{ 4\epsilon_{ij} \left[ \left( \frac{r_0^{ij}}{r_{ij}} \right)^{12} - \left( \frac{r_0^{ij}}{r_{ij}} \right)^6 \right] f_{ij} \right\} + \sum_{i=1}^{N-1} \sum_{j=1}^N \frac{1}{4\pi\epsilon_0} \left( \frac{q_i q_j}{r_{ij}} \right) \quad (21)$$

Sendo: o coeficiente  $f_{ij}$  igual a zero, para pares i-j conectados por uma ligação (interações 1-2) ou por um ângulo de ligação (interações 1-3).  $f_{ij} = 0,5$  para interações 1-4 (átomos separados por 3 ligações) e  $f_{ij} = 1,0$  para todos os outros casos.<sup>(65)</sup>

O primeiro e o segundo termo em (19) são potenciais harmônicos. O potencial de diedros em (20) é um potencial trigonométrico cossenoidal com a altura da barreira torcional do i-ésimo ângulo diedral  $V$  ( $V_1^i$ ,  $V_2^i$  e  $V_3^i$ ). Na Equação (21), o primeiro termo é representado por um potencial de Lennard-Jones e o segundo, por um potencial de Coulomb.

A primeira versão do campo de força OPLS surgiu na década de 1980 como um modelo *united-atom* (OPLS-UA) para simular a água e cerca de 40 líquidos orgânicos. Este modelo dava grande ênfase à etapa de derivação das interações não ligadas, que foram sistematicamente ajustadas para reproduzir as propriedades termodinâmicas dos líquidos

simulados. As aplicações iniciais para proteínas tomaram os tipos de átomos e os parâmetros de valência (ligação, ângulo e diedros) do campo de força AMBER 1984, resultando no campo de força que ficou conhecido como AMBER/OPLS. Mais tarde, uma versão all-atom, conhecida como OPLS-AA, foi desenvolvida para proteínas, mantendo praticamente a mesma filosofia de derivação de cargas e parâmetros de van der Waals de simulações em líquidos puros.<sup>(65, 67)</sup>

As cargas do OPLS foram, em princípio, obtidas empiricamente e com o passar dos anos foram sendo amplamente ajustadas para reproduzir propriedades de líquidos orgânicos. Os parâmetros dessas cargas foram escolhidos com a intenção de serem facilmente transferidos a outras moléculas com grupamentos quimicamente similares. Os parâmetros dos termos de estiramento e ângulos de ligação foram amplamente tomados do campo de força *all-atom* AMBER 1986, enquanto os parâmetros torcionais foram desenvolvidos de modo consistente pelo ajuste a perfis energéticos de HF 6-31G\*. Este nível de precisão está em boa concordância com dados experimentais, mas pequenas melhorias foram obtidas com a inclusão de correções de correlação MP2. De fato, reajustes desta natureza foram feitos para as cadeias laterais carregadas de aminoácidos.<sup>(65, 67)</sup>

#### **2.4.5 Comparação dos Campos de Força GROMOS, AMBER, OPLS e ENCAD**

De maneira geral, os campos de força da fase condensada biomolecular AMBER, ENCAD, GROMOS e OPLS têm uma forma semelhante da função potencial de interação, mas diferem em suas filosofias de parametrização e nos valores dos seus parâmetros. Isto se deve à grande variedade de formas pelas quais podem ser obtidos estes parâmetros. De fato, uma gama de propriedades (geométricas, energéticas, dielétricas, dinâmicas, etc.) de moléculas pequenas podem ser ajustadas contra diferentes conjuntos de dados experimentais e de mecânica quântica. Desta forma, conjuntos de parâmetros diferentes podem produzir resultados muito diferentes quando aplicados em sistemas biomoleculares grandes e complexos.<sup>(62)</sup>

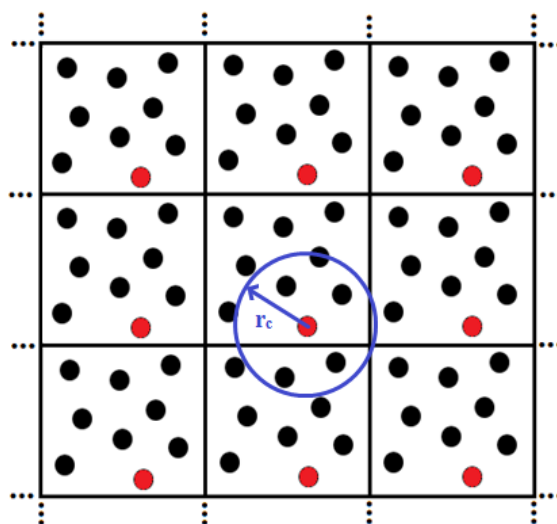
Ao longo dos anos os campos de força OPLS e AMBER vêm convergindo para um modelo eletrostático comum. Isto reflete o uso de protocolos e conjuntos de teste de parametrizações cada vez mais similares (uma vez que a parametrização do OPLS se inspirou na do AMBER), mas também indica uma busca incansável, e ainda em curso, por um modelo de carga ótimo. Por outro lado, o campo de força GROMOS utiliza um conjunto bem diferente de cargas para os aminoácidos (ver tabela no ANEXO A2). Por fim, comparar a

qualidade de campos de força é uma tarefa difícil, pois não há garantia de que um campo de força ajustado para ter bons resultados com um grupo de moléculas funcione bem com outro grupo.<sup>(67)</sup>

## 2.5 INTERAÇÕES ELETROSTÁTICAS DE LONGO ALCANCE

Uma vez que um sistema simulado é muito menor do que um sistema real, efeitos de superfície poderiam distorcer os resultados da simulação de dinâmica molecular. Para contornar problemas desta natureza, o sistema simulado imita o limite termodinâmico empregando a técnica de condições de contorno periódicas” (PBC, do inglês *Periodic Boundary Conditions*). Essa técnica produz réplicas da caixa (preenchida com as partículas a serem simuladas) em todas as direções, eliminando, com isso, as superfícies e tornando o sistema virtualmente infinito.<sup>(68)</sup>

Para o sistema assim definido, uma partícula  $i$ , posicionada na caixa central, pode interagir com todas as partículas  $j$  e as imagens periódicas de  $j$  (entre elas as imagens periódicas da própria partícula  $i$ ). Para evitar essa contagem múltipla de interações (que poderiam afetar os cálculos de potenciais e forças), utiliza-se a “convenção da imagem mínima” que restringe as interações da partícula  $i$ . Desta forma, a partícula  $i$  só irá interagir com a cópia mais próxima de cada partícula  $j$ .<sup>(68)</sup>



**Figura 13:** Esquema bidimensional da replicação da caixa de simulação empregando as condições de contorno periódicas (PBC) e a convenção da “imagem mínima”. O círculo em azul indica o raio de corte aplicado em uma dada partícula.<sup>(69)</sup>

O uso combinado destas técnicas (PBC e “convenção da imagem mínima”, ver Figura 13) afeta o alcance dos potenciais de interação. De fato, as contribuições dos potenciais de curto alcance (como os potenciais de Lennard-Jones) podem ser negligenciadas a partir de uma certa distância, denominada “raio de corte” ( $r_c$ ). Esse procedimento de truncamento dos potenciais anula as interações além dessa distância. Computacionalmente, o valor do raio de corte empregado numa simulação de dinâmica molecular é definido pelo usuário. No caso de simulações de sistemas biológicos (como proteínas em solução), o raio de corte deve ser menor ou igual à metade da diferença entre o comprimento da caixa cúbica e o tamanho da macromolécula na direção considerada. O truncamento do potencial, no entanto, produz duas importantes fontes de erro: <sup>(68, 69)</sup>

- 1) Introduz uma descontinuidade nas forças;
- 2) Negligencia as interações além do raio de corte;

Para o potencial de Lennard-Jones a primeira fonte de erro é contornada empregando funções de “*Shift* ou *Switch*”, que tornam contínuo o potencial e suas derivadas no raio de corte. Enquanto a segunda fonte de erro pode ser calculada exatamente por: <sup>(68)</sup>

$$V = \sum_{r_{ij} < r_c} V(r_{ij}) + \frac{N\rho}{2} \int_{r_c}^{\infty} v(\mathbf{r})g(\mathbf{r})4\pi r^2 d\mathbf{r} \quad (22)$$

Sendo  $v(\mathbf{r})$  o potencial de interação entre partículas e  $g(\mathbf{r})$  a função de distribuição dos pares a uma dada distância  $\mathbf{r}$ . Assim, no primeiro termo da equação, para uma distância entre pares de partículas dentro do raio de corte, as interações são calculadas por um potencial de Lennard-Jones  $V(r_{ij})$  definido pelo campo de força. Enquanto no segundo termo da equação, a integral contabiliza as interações entre pares de partículas separadas por uma distância maior do que o raio de corte.

No caso das interações de longo alcance (como é o caso das interações eletrostáticas), no entanto, estas duas fontes de erro não podem ser corrigidas pelo uso das técnicas acima descritas. <sup>(68, 69)</sup> Desta forma, é necessário o uso de outros métodos para o tratamento das interações de longo alcance. Dentre eles encontram-se os métodos GRF (do inglês, *Generalized Reaction Field*) <sup>(70)</sup> e SPME (do inglês, *Smooth Particle Mesh Ewald*) <sup>(71)</sup>, que são brevemente apresentados a seguir.

### 2.5.1 GRF

No método conhecido como “campo de reação” (do inglês, *Reaction Field*; sigla RF) uma esfera, com um raio equivalente à distância de corte, é construída ao redor de cada partícula  $i$ . A forma de calcular as interações eletrostáticas depende da distância ( $r_{ij}$ ) que separa essa partícula  $i$  de uma partícula  $j$  qualquer. Assim, interações que ocorrem dentro do raio de corte (ou seja,  $r_{ij} \leq r_c$ ) são explicitamente calculadas a partir da função potencial eletrostática definida pelo campo de força. Para interações entre partículas separadas por uma distância maior que o raio de corte ( $r_{ij} > r_c$ ), considera-se que o meio se comporta como um contínuo dielétrico que produz um campo de reação na cavidade da esfera. Esse campo, por sua vez, interage com a partícula  $i$ , modificando o potencial e as forças que nela atuam. Para levar em conta a força iônica do meio emprega-se o método de “campo de força generalizado” (do inglês, *Generalized Reaction Field*; sigla GRF).<sup>(59, 68, 69, 70)</sup>

O método do campo de reação representa simplesmente uma correção no potencial de interação eletrostático entre pares de partículas, promovendo apenas um ligeiro aumento no custo computacional. No entanto, alguns erros podem ser introduzidos no caso de simulações de sistemas altamente carregados.<sup>(68)</sup> Para esses casos, recomenda-se o uso do método SPME<sup>(71)</sup> (ou qualquer um dos derivados do método da soma de Ewald), cujos princípios são tratados na sequência.

### 2.5.2 SPME

A Soma de Ewald foi um dos primeiros métodos usados para calcular as interações eletrostáticas de longo alcance em simulações de dinâmica molecular. Originalmente, no entanto, este método foi desenvolvido para calcular a energia de redes cristalinas. A posterior aplicação deste em dinâmica molecular decorreu do uso da técnica PBC nas simulações. Com isso, cada caixa de simulação podia ser entendida (do ponto de vista eletrostático) como uma célula de um sistema cristalino.

O método da Soma de Ewald emprega uma combinação de distribuições gaussianas e transformadas de Fourier para reescrever o potencial eletrostático como a soma de três termos: um termo constante para cada configuração do sistema, um termo de convergência rápida no espaço real e outro termo de convergência rápida no espaço recíproco (ou espaço de transformada). Desta forma, o potencial eletrostático, que originalmente era uma função de decaimento lento (descrito por um potencial de Coulomb), assume uma forma de rápida convergência. Esse método, no entanto, possui um alto custo computacional, de modo que seu

uso não é recomendável para o tratamento de sistemas muito grandes (como os sistemas de natureza biológica).<sup>(48, 68)</sup>

Uma alternativa computacionalmente mais rápida é dada pelo método SPME que também transforma o potencial eletrostático em uma soma de dois termos de rápida convergência. No entanto, no método SPME as cargas originais são removidas de suas posições e são projetadas sob uma malha fixa e definida na forma de cargas parciais.<sup>(71, 72)</sup>

Desta forma, para uma densidade de carga original,  $\rho(\mathbf{r})$ , no ponto  $\mathbf{r}$  do sistema, define-se a densidade de carga,  $\rho_M$ , que é projetada no ponto  $x_p$  da malha, como:<sup>(72)</sup>

$$\rho_M(\mathbf{r}_p) = \frac{1}{h} \int_0^L W(\mathbf{r}_p - \mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} \quad (23)$$

Onde o prefator  $(1/h)$  meramente assegura que  $\rho_M$  é, de fato, uma densidade;  $h$  é o espaçamento dos pontos na malha;  $L$  refere-se às dimensões da caixa de simulação;  $\mathbf{r}_p$  da a posição dos pontos na malha;  $\mathbf{r}$  da a posição da densidade de carga na caixa de simulação;  $W(\mathbf{r}_p - \mathbf{r})$  é uma função de interpolação, ou seja, ela interpola a densidade de carga,  $\rho(\mathbf{r})$ , em uma densidade de carga na malha,  $\rho_M$ . No método SPME, essa função de interpolação é representada por uma função de *spline* B-cardinal.<sup>(72)</sup>

## 2.6 MÚLTIPLAS SIMULAÇÕES DE DINÂMICA MOLECULAR

A superfície de energia potencial (SEP) descrevendo uma proteína contém múltiplos mínimos similares, com conformações levemente diferentes, de forma que o sucesso de estudos de dinâmica molecular (DM) requer uma exploração efetiva da SEP.<sup>(73)</sup>

Qualquer sistema com caráter multimínimo (possuindo barreiras de SEP) pode ter suas barreiras energéticas não sendo atravessadas com frequência suficiente para se obter uma distribuição de equilíbrio em tempos computacionalmente acessíveis. Se a média temporal de uma quantidade depende das condições iniciais, o sistema é não-ergódico na escala de tempo e temperatura da simulação. Além do mais, propriedades diferentes podem convergir em escalas de tempo distintas. Portanto, é necessário estudar mais de uma propriedade para determinar a ergodicidade de uma simulação. Logo, o problema de amostragem é de grande importância.<sup>(73)</sup>

Inúmeras técnicas foram desenvolvidas visando reforçar o método de amostragem via DM. Auffinger e colaboradores (1995)<sup>(74)</sup> utilizaram a abordagem de múltiplas pequenas

trajetórias e concluíram que essas poderiam substituir uma única trajetória longa na exploração do espaço conformacional da molécula.

Caves e colaboradores<sup>(73)</sup> desenvolveram um estudo de múltiplas simulações de dinâmica molecular para a cambrina, realizando 10 simulações (T1-T10) de 120 ps cada, partindo de uma mesma conformação de referência, X, alterando apenas as velocidades atômicas iniciais geradas. Sendo que 3 das 10 simulações (T2, T5 e T9 → T2x, T5x e T9x) foram estendidas à 1 ns e uma (T1 → T1x), à 5 ns.

Os resultados obtidos mostraram que diferentes trajetórias podem amostrar a mesma região local do espaço conformacional, mas seguem caminhos diferentes. Desta forma, estender a trajetória por um tempo maior não garante, necessariamente, amostragem mais extensiva, pois a trajetória pode permanecer presa em uma região do espaço conformacional por longo tempo. Isto sugere que conformações geradas por uma única trajetória (mesmo muito longa) podem não ser representativas de toda a gama de conformações acessíveis termicamente para o sistema.

Caves e colaboradores ressaltam ainda que se deve tomar cuidado ao decidir se uma simulação foi estendida por tempo suficiente, pois algumas propriedades podem ser mais sensíveis a mudanças estruturais do que outras. Assim, não se deve concluir sobre a ergodicidade de uma simulação baseada na convergência de uma única propriedade. Este problema de uma adequada exploração do espaço de fases acentua-se com a complexidade dos sistemas biomoleculares.

Uma vez que muitas simulações de DM utilizam como estrutura de partida estruturas na vizinhança daquelas determinadas por difração de raios X ou RMN, o ambiente (contatos cristalinos, solvente, sais, etc) do experimento pode não ser bem reproduzido ou mesmo não ser explicitamente considerado no modelo usado nas simulações. Logo, uma estrutura experimentalmente determinada pode não estar em um mínimo, mas sim situada próxima a um pico da SEP empírica. Nesse último caso, quando realizada a dinâmica molecular, a estrutura pode se mover em direção a uma ou mais regiões de baixa energia. Em situações como essa, Caves e colaboradores acreditam que a manipulação das condições iniciais (empregando múltiplas trajetórias curtas) tem o efeito de aumentar a amostragem do espaço conformacional. Por outro lado, se a estrutura de partida (obtida do experimento) estiver próxima a um mínimo espera-se que a sensibilidade à randomização das velocidades iniciais diminua devido aos efeitos de confinamento da SEP. Este ponto pode gerar implicações no protocolo de simulações de DM de proteínas à temperatura ambiente.<sup>(73)</sup>

É importante, ainda, ressaltar que um estudo de influência do pH realizado com a proteína príon celular humana foi publicado por Campos e colaboradores no ano de 2010.<sup>(75)</sup> Neste estudo foram realizadas réplicas, empregando a randomização das velocidades iniciais, para cada um dos pH investigados. Este trabalho já estava apontando para a necessidade de produzir múltiplas simulações de dinâmica molecular nas investigações de influência de pH no processo de conversão estrutural da proteína príon celular.



### **3. OBJETIVOS**

### 3.1 OBJETIVO GERAL

O objetivo geral deste trabalho é produzir modelos de estruturas alterada, PrP<sup>Sc</sup>, e neutra, PrP<sup>C</sup>, baseando-se no modelo proposto por Alonso e colaboradores<sup>(33)</sup>, com o intuito de utilizar estas estruturas em investigações futuras de agregação protéica.

### 3.2 OBJETIVOS ESPECÍFICOS

1 - Avaliar o efeito do uso de diferentes campos de força (GROMOS96 53a6<sup>(62)</sup>, AMBER99φ<sup>(64)</sup>, OPLS-AA<sup>(65)</sup>) para estudar a influência do pH baixo no processo de conversão da proteína príon celular, com o intuito de verificar a reprodutibilidade da unidade estrutural básica do modelo espiral proposta por Alonso e colaboradores<sup>(33)</sup>.

2 - Investigar a estrutura em pH neutro, avaliando o efeito do uso das diferentes formas de histidina (HisA, HisB e HisH) disponíveis para o campo de força GROMOS. Esta etapa do procedimento foi realizada com o intuito de minimizar o conteúdo de fitas-β na região N-terminal da estrutura simulada em pH neutro.

3 - Avaliar a influência de diferentes condições iniciais nas simulações de DM, partindo de:

- a) diferentes conformações (diferentes estruturas do RMN);
- b) diferentes conjuntos de velocidades iniciais (através da randomização dessas velocidades).

4 - Comparar os resultados dos estudos de influência de pH empregando diferentes algoritmos GRF<sup>(70)</sup> (*Generalized Reaction Field*) e SPME<sup>(71)</sup> (*Smooth Particle Mesh Ewald*); para o tratamento das interações eletrostáticas de longo alcance.

## **4. METODOLOGIA**

## 4.1 PROTOCOLO DE SIMULAÇÃO

No trabalho de Alonso e colaboradores<sup>(33)</sup>, foi estudado o efeito do pH na conversão de PrP<sup>C</sup> → PrP<sup>Sc</sup>, tendo sido empregado, para tanto, o campo de força ENCAD. No presente trabalho, buscou-se identificar as condições ideais para o estudo de influência do pH (neutro e ácido) no processo de conversão estrutural da proteína príon celular. Para tanto, foram avaliados os efeitos de utilização de diferentes campos de força, de diferentes estados de protonação de resíduos histidina, da alteração das condições iniciais de simulação (pelo uso de diferentes modelos contidos no RMN e pela randomização das velocidades iniciais) e do uso de diferentes métodos para o cálculo das interações eletrostáticas de longo alcance.

Num primeiro momento foi obtida a estrutura do fragmento protéico de resíduos 109-219, resolvida por RMN em solução, do banco de dados *Protein Data Bank* (código PDB: 2PrP)<sup>(18)</sup>. A seguir é apresentada a sequência primária de aminoácidos da proteína príon celular do hamster Sírio correspondente ao fragmento de resíduos 109-219:

MKHMAGAAAAGAVVGGLGGYMLGSAMSRPMMHFGNDWEDRYRENMNRYPNQVYYRP  
VDQYNNQNNFVHDCVNITIKQHTVTTTTKGENFTETDIKIMERVVEQMCTTQYQ

Empregou-se o programa *pdb2gmx*, do pacote *software* GROMACS 4.05<sup>(76)</sup>, para gerar o arquivo de topologia, o arquivo de coordenadas cartesianas da macromolécula biológica e para selecionar o campo de força. Ainda nesta etapa, foi necessário manipular os estados de protonação dos resíduos His, Glu e Asp para que correspondessem ao pH desejado (detalhamentos dos estados de protonação desses resíduos serão apresentados nos capítulos de Resultados e Discussão do presente trabalho). Para cada um dos sistemas simulados foi, então, realizado o procedimento descrito a seguir.

O sistema foi centrado em uma caixa cúbica, com as faces distando de no mínimo 1,5 nm da proteína. A caixa foi preenchida com moléculas de água (do tipo SPC<sup>(66)</sup> para o campo de força GROMOS96 53a6<sup>(62)</sup> e TIP3P<sup>(77)</sup> para os campos de força AMBER99ϕ<sup>(64)</sup> e OPLS-AA<sup>(65)</sup>). Com isto, o arquivo de topologia foi alterado, gerando-se um novo arquivo de coordenadas, desta vez contendo proteína e água. Foram adicionados, ainda, íons Na<sup>+</sup> e Cl<sup>-</sup> para neutralizar as cargas do sistema e garantir um ambiente fisiológico.

A etapa seguinte consistiu na minimização de energia do sistema proteína-solvente. Para isto, foi utilizado o algoritmo de minimização *Steepest Descent*. Na sequência foi realizada uma dinâmica molecular com restrição de posição (*position restraint*) de 200 ps, na qual a macromolécula foi restrita, mas as moléculas do solvente tiveram liberdade de movimento.

A última etapa da simulação consistiu na dinâmica molecular propriamente dita, na qual todos os graus de liberdade foram amostrados. Para isto foram definidos alguns parâmetros, comuns a todas as simulações:

- passo de integração de 0,002 ps;
- para corrigir as interações de longo alcance foi utilizado o algoritmo o SPME<sup>(71)</sup> (com um raio de corte equivalente à 1,3 nm para o cálculo das interações eletrostáticas e de van der Waals) nas simulações apresentadas nos capítulos 5.1 a 5.3 e o algoritmo GRF<sup>(70)</sup> (com um raio de corte equivalente à 1,4 nm para o cálculo das interações eletrostáticas e de van der Waals, e uma constante dielétrica igual a 54) em algumas simulações apresentadas no capítulo 5.4;
- a temperatura de acoplamento foi mantida pelo uso do algoritmo de Nose-Hoover<sup>(78)</sup>;
- a temperatura de acoplamento empregada foi de 298 K;
- a pressão da simulação foi mantida constante pelo uso do algoritmo de Parrinello-Rahman<sup>(79)</sup>;
- a pressão do sistema simulado foi mantida em 1 bar;
- os comprimentos de todas as ligações de equilíbrio foram mantidos pelo uso do algoritmo LINCS<sup>(80)</sup>;

Uma vez terminadas as simulações foram feitas as seguintes análises com base nas trajetórias dessas: evolução da estrutura média (RMSD; do inglês, *Root Mean Square Deviation*), mobilidade relativa de segmentos (RMSF; do inglês, *Root Mean Square Fluctuation*) e análises de estruturas secundárias.

Todas as simulações e posteriores análises foram realizadas no laboratório do Grupo de Química Teórica (GQT), localizado no Instituto de Química da Universidade Federal do Rio Grande do Sul. Atualmente o GQT conta com um *cluster* de 13 nodos, permitindo cálculos paralelos, três estações de trabalho e um laboratório computacional com dez estações de trabalho. Para os sistemas empregados neste trabalho, simulações com duração de 5 ns levaram cerca de 2 dias de cálculos em máquina para serem realizadas.

## 4.2 ANÁLISES

De posse das trajetórias das simulações de DM, foram realizadas análises empregando programas do *software* GROMACS 4.05<sup>(76)</sup>. As análises realizadas foram: RMSD, RMSF e estrutura secundária.

O fator de temperatura é uma medida de mobilidade em sistemas cristalinos, relacionado às vibrações térmicas dos átomos num cristal em torno das suas posições de equilíbrio. Modificando a densidade eletrônica, estas oscilações alteram a capacidade de espalhamento do átomo, interferindo na resolução dos picos de difração (que podem ser associados a átomos cujas posições não são muito bem definidas na célula unitária). Desta forma, estes picos representam uma medida indireta de mobilidade.<sup>(48)</sup>

Cristais de sistemas complexos possuem caracteristicamente valores altos de mobilidade (vibração térmica) em sua rede cristalina. O fator B (obtido experimentalmente por cristalografia) pode alcançar valores de 50 Å<sup>2</sup> para soluções saturadas de proteínas e outras macromoléculas biológicas. De fato, este é um valor relativamente alto se comparado aos valores tipicamente obtidos para cristais inorgânicos e orgânicos, da ordem de 1 e 20 Å<sup>2</sup>, respectivamente. O fator B de temperatura se relaciona ao deslocamento quadrático médio RMSD por uma fórmula simples:<sup>(48)</sup>

$$B = \frac{8}{3} \pi (\text{RMSD})^2 \quad (24)$$

Desta forma, pode-se relacionar o RMSD (ou o desvio quadrático médio das posições atômicas de uma proteína) obtido de uma trajetória de DM com um observável experimental (B). No presente trabalho, foram empregadas particularmente duas definições de RMSD (dentre as inúmeras variantes existentes) que são importantes análises no contexto de simulações de proteínas: evolução temporal do RMSD, ou simplesmente RMSD, e a média de RMSD por átomo, também conhecido como RMSF.

O RMSD calcula o desvio padrão médio quadrático para todos os átomos da proteína, conforme a fórmula a seguir:<sup>(48,61)</sup>

$$\text{RMSD}(t_1) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_{i0}\|^2 \right]^{\frac{1}{2}} \quad (25)$$

$$M = \sum_{i=1}^N m_i \quad (26)$$

Sendo  $r_i(t_1)$  e  $r_{i0}$  as posições do átomo  $i$ , respectivamente, nos tempos  $t_1$  e  $t_0$ ;  $m_i$  a massa do átomo  $i$ ;  $N$  o número de átomos da macromolécula; e  $M$  o somatório da massa de todos os átomos da macromolécula, conforme apresentado na Equação (26).

Ou seja, o RMSD é utilizado para estimar os desvios médios da estrutura da proteína, em relação à estrutura de referência (onde  $t_0 = 0$  ns), no decorrer da simulação. Assim, é vislumbrada a evolução da mudança estrutural total da molécula ao longo da simulação. O RMSD pode ser calculado para os átomos da cadeia principal, para os átomos de carbono alfa ( $C\alpha$ ), ou outras opções disponíveis no programa `g_rms` do pacote GROMACS.

O RMSF calcula o desvio padrão das posições atômicas ao longo da simulação com relação às posições de uma estrutura tomada como referência, conforme apresenta a fórmula a seguir:<sup>(48, 61)</sup>

$$\text{RMSF} = \text{RMSD}_i = \left[ \frac{1}{S \cdot \Delta t} \sum_{j=1}^S \|r_i(t_j) - r_{i0}\|^2 \right]^{\frac{1}{2}} \quad (27)$$

Sendo  $r_i(t_j)$  e  $r_{i0}$  as posições do átomo  $i$ , respectivamente, nos tempos  $t_j$  e  $t_0$ ;  $S$  o número de configurações consideradas para o átomo  $i$ ; e  $\Delta t$  o intervalo de tempo entre cada passo gravado na trajetória de DM. Assim, multiplicar  $S$  por  $\Delta t$  resulta no tempo total de simulação. Ou seja, o RMSF calcula a média dos valores de RMSD de todos os passos de uma trajetória de DM para cada átomo (ou resíduo) especificamente, dando uma ideia das flutuações de átomos (resíduos ou regiões) da proteína ao longo da simulação<sup>(48)</sup>.

Para a realização de análises bem sucedidas relacionando sequências de aminoácidos e estrutura de proteínas, é necessário uma definição precisa de estrutura secundária. As análises estruturais, do presente trabalho, foram realizadas utilizando o programa DSSP<sup>(81)</sup> (do inglês *Dictionary of Protein Secondary Structure*) do pacote GROMACS. Partindo de uma definição de ligações de hidrogênio, o DSSP define as *n-voltas* e os *segmentos  $\beta$*  (do inglês, *n-turns* e *bridges*, respectivamente); e, baseado nestas, define as  *$\alpha$ -hélices* (do inglês,  *$\alpha$ -helix*) e as *fitas- $\beta$  pareadas* (do inglês,  *$\beta$ -ladders*). Entre as características geometricamente definidas, encontram-se as *dobras* (do inglês, *bends*). A seguir são apresentadas algumas definições estruturais empregadas no programa DSSP<sup>(81)</sup>:

**Ligações de Hidrogênio (LH)**: as ligações de hidrogênio em proteínas apresentam pequena sobreposição de funções de onda, sendo bem descritas por um modelo eletrostático:

$$E = q_1 q_2 \left[ \left( \frac{1}{r(ON)} \right) + \left( \frac{1}{r(CH)} \right) + \left( \frac{1}{r(OH)} \right) + \left( \frac{1}{r(CN)} \right) \right] \quad (28)$$

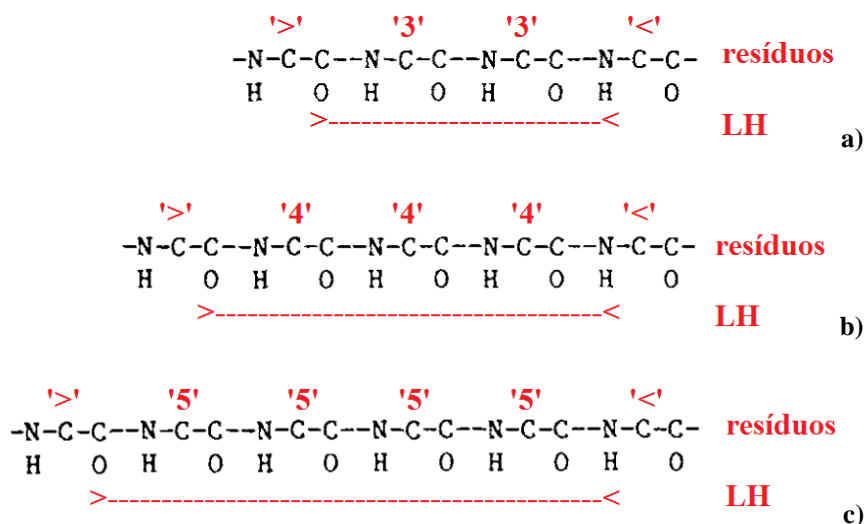
Na Equação (28)  $q_1$  e  $q_2$  são as cargas parciais, sendo que para os átomos de carbono (C), oxigênio (O), nitrogênio (N) e hidrogênio (H) têm-se, respectivamente, as seguintes cargas parciais:  $+q_1$ ,  $-q_1$ ,  $-q_2$  e  $+q_2$ . Enquanto  $r(*)$  é a distância entre o par de átomos indicado entre parênteses.

**Estruturas secundárias definidas com base em padrões de ligações de hidrogênio:**

**n-Voltas (ou do inglês, n-Turns):** o padrão básico das ligações de hidrogênio (LH), envolvidas nas n-voltas, é do tipo (i, i + n). A ligação de hidrogênio se estabelece do resíduo CO (i) ao resíduo NH (i + n):

$$n\text{-voltas (i) = LH (i, i + n) \quad , \text{ sendo } n = 3,4,5.$$

**Hélices:** uma hélice mínima é definida por 2 n-voltas consecutivas. Com base nos tipos de n-voltas são definidos três tipos de hélices mínimas: 3-hélice, 4-hélice e 5-hélice (conforme mostrado na Figura 14). Hélices maiores são definidas pela sobreposição de hélices mínimas. Assim, a sobreposição de 3-hélices gera uma  $3_{10}$ -hélice; a sobreposição de 4-hélices gera uma  $\alpha$ -hélice; e a sobreposição de 5-hélices gera uma  $\pi$ -hélice.



**Figura 14:** Representação esquemática da formação da ligação de hidrogênio (LH) estabelecida entre os grupamentos CO (i) e NH (i + n) na cadeia polipeptídica que define os diferentes tipos de n-voltas: **a)** 3-volta; **b)** 4-volta; **c)** 5-volta (imagens adaptadas do artigo de Kabsch e Sander)<sup>(81)</sup>.



**Segmento  $\beta$  (ou do inglês,  $\beta$ -Bridge):** um *segmento  $\beta$*  ocorre entre os resíduos  $i$  e  $j$ , se existirem duas ligações de hidrogênio características de estrutura  $\beta$ , em particular:

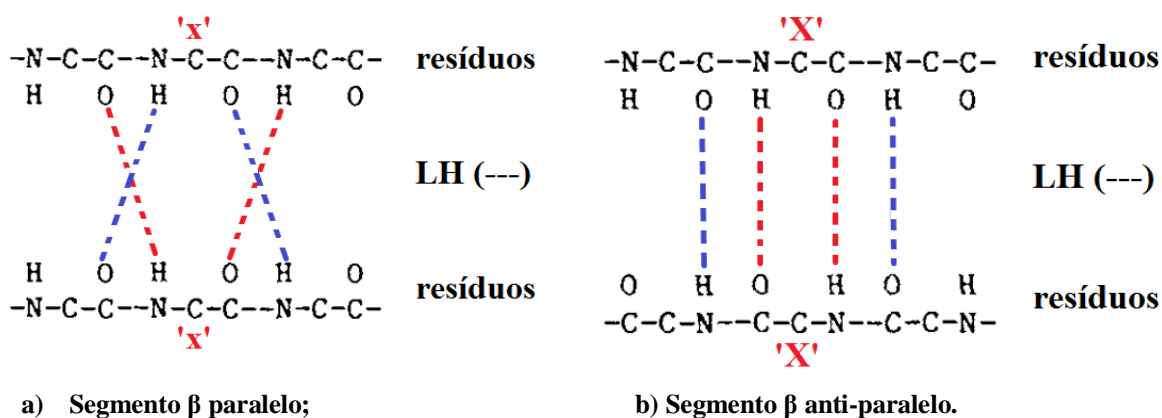
Segmento  $\beta$  paralelo (i,j) = [LH (i - 1, j) and LH (j, i + 1)]

Segmento  $\beta$  anti-paralelo (i,j) = [LH (i, j) and LH (j, i)]

**Fitas- $\beta$  pareadas e Folhas- $\beta$  (ou do inglês,  $\beta$ -ladders e  $\beta$ -sheet, respectivamente):**

- **Fitas- $\beta$  pareadas:** conjunto de um ou mais *segmentos  $\beta$* , de igual tamanho, consecutivos;
- **Folhas- $\beta$ :** conjunto de uma ou mais *fitas- $\beta$  pareadas* conectadas por resíduos compartilhados.

De maneira equivalente, pode-se dizer que duas fitas conectadas constituem uma estrutura que aqui denominamos de *fitas- $\beta$  pareadas* (ou  $\beta$ -ladder). Assim, uma estrutura formada de três *fitas pareadas* possui duas *fitas- $\beta$  pareadas* ( $\beta$ -ladders) e uma *folha- $\beta$*  ( $\beta$ -sheet). A Figura 15, a seguir, apresenta um esquema das ligações de hidrogênio nos *segmentos  $\beta$  paralelo e anti-paralelo*.

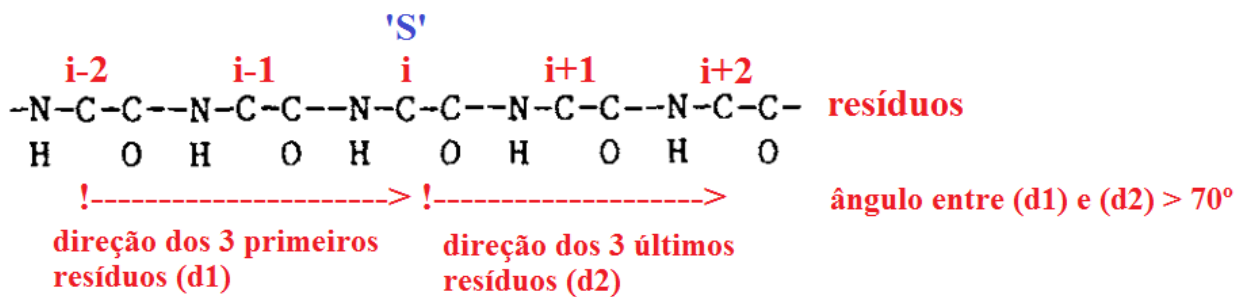


**Figura 15:** Representação esquemática das ligações de hidrogênio (LH) na estruturas: **a)** Segmento  $\beta$  paralelo; **b)** Segmento  $\beta$  anti-paralelo (imagens adaptadas do artigo de Kabsch e Sander)<sup>(81)</sup>.

**Estrutura secundária geometricamente definida:**

**Dobras (ou do inglês, bends):** são espécies com alta curvatura. Define-se a curvatura da cadeia, em um resíduo central  $i$  de 5 resíduos, como o ângulo entre a direção da cadeia principal dos três primeiros resíduos e, a dos três últimos. Para existir uma *dobra* no resíduo

central  $i$ , é necessário uma curvatura de pelo menos  $70^\circ$ , conforme apresentado na Figura 16, a seguir.



**Figura 16:** Esquema da curvatura para as *dobras (bends)*, indicando um ângulo ( $>70^\circ$ ) entre a direção dos 3 primeiros resíduos e a direção dos 3 últimos resíduos do fragmento da cadeia polipeptídica em questão (imagem adaptada do artigo de Kabsch e Sander)<sup>(81)</sup>.

Com base nessas definições de estrutura secundária é possível realizar análises estruturais das proteínas simuladas. Na análise de DSSP é produzido um gráfico que relaciona o tipo de estrutura secundária que cada resíduo da cadeia polipeptídica (eixo y) possui ao longo do tempo de simulação (eixo x). Na análise de *SCOUNT*, é produzido um gráfico que relaciona o número de resíduos (eixo y) envolvidos em cada tipo de estrutura secundária ao longo do tempo (eixo x) de simulação. Por fim, as imagens de estruturas de proteínas apresentadas neste trabalho foram produzidas utilizando o programa SPDBV<sup>(57)</sup>.

## **5. RESULTADOS E DISCUSSÃO**

## 5.1 COMPARAÇÃO DE CAMPOS DE FORÇA

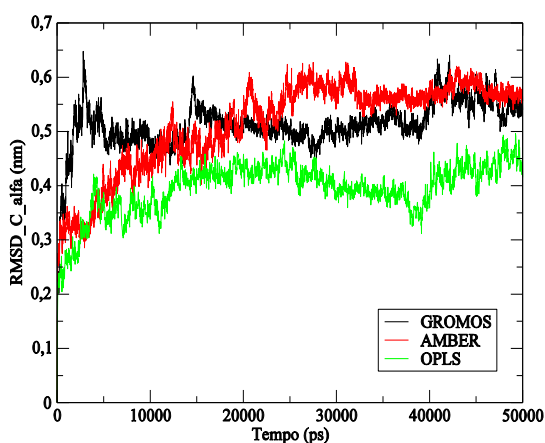
### 5.1.1 GROMOS, AMBER e OPLS

Com o objetivo de estabelecer um protocolo de simulação adequado para o processo de alteração conformacional da proteína príon celular simulada sob pH ácido ( $\text{pH} < 4$ ), foi avaliado o desempenho (durante 50 ns) de três dos mais utilizados campos de força para o estudo de sistemas biológicos: GROMOS 96 53a6<sup>(62)</sup>, AMBER99 $\phi$ <sup>(64)</sup> e OPLS-AA<sup>(65)</sup>. Para garantir o pH baixo dos sistemas, os resíduos Asp, Glu e His foram protonados, assumindo os tipos ASPH, GLUH e HISH no campo de força GROMOS. A Tabela I apresenta a composição dos diferentes sistemas simulados.

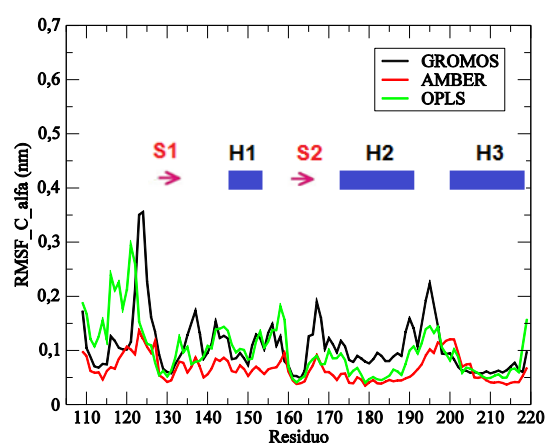
**Tabela I:** Número de moléculas e íons de cada um dos sistemas simulados.

Simulação	Solvente			
	(H <sub>2</sub> O)	Proteína	Na <sup>+</sup>	Cl <sup>-</sup>
AMBER	10951	1	31	45
GROMOS	10892	1	30	44
OPLS	13153	1	39	53

Observando o gráfico de RMSD ( $C\alpha$ ), Figura 17a, verifica-se que os campos de força GROMOS e OPLS convergem nos últimos 10 ns aos valores de, respectivamente, 0,55 e 0,45 nm, enquanto AMBER apresenta convergência em 0,57 nm a partir de 30 ns de simulação.

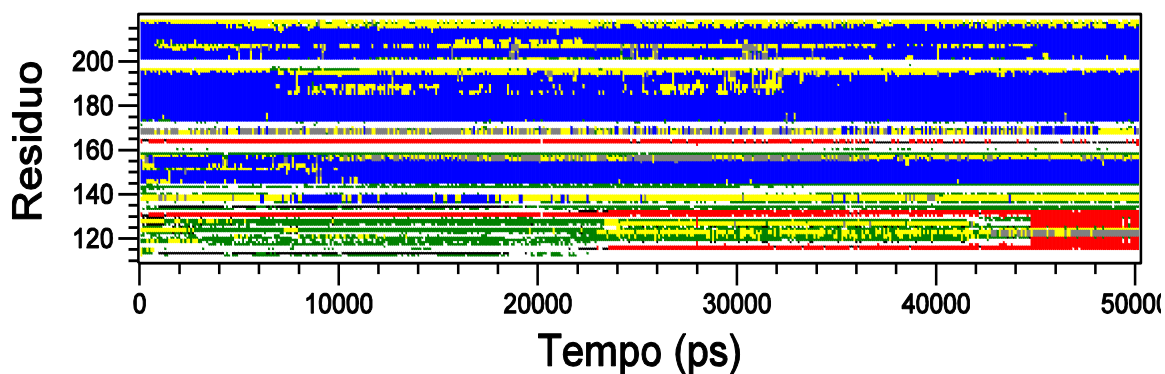


a) RMSD ( $C\alpha$ );

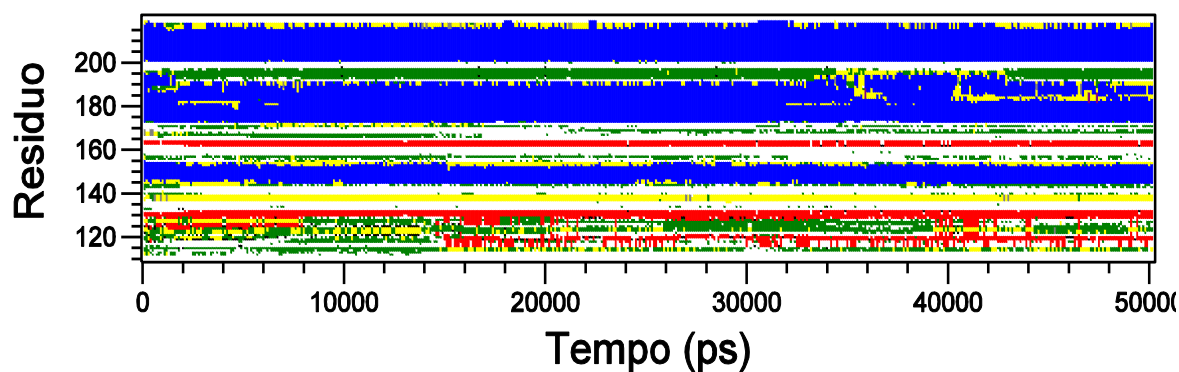


b) RMSF ( $C\alpha$ ).

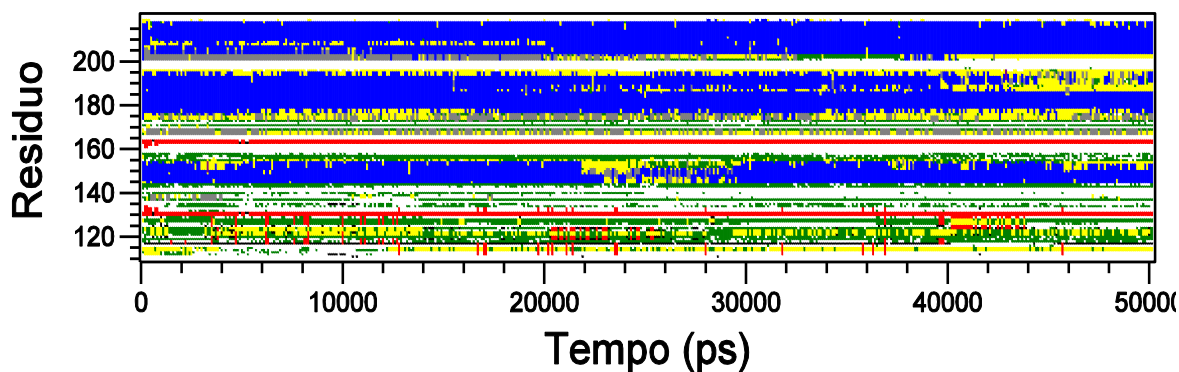
**Figura 17:** Gráficos de: a) RMSD ( $C\alpha$ ) em função do tempo de simulação; b) RMSF ( $C\alpha$ ) em função do resíduo, obtido sobre os últimos 5 ns de simulação para cada um dos campos de força GROMOS, AMBER e OPLS. No gráfico de RMSF ( $C\alpha$ ), as flechas rosas indicam a localização das fitas- $\beta$  originais (S1 e S2; presentes na estrutura inicial de 2PRP) e as caixas azuis indicam a localização das hélices (H1, H2 e H3) nas estruturas protéicas simuladas.



a) AMBER;



b) GROMOS;



c) OPLS.

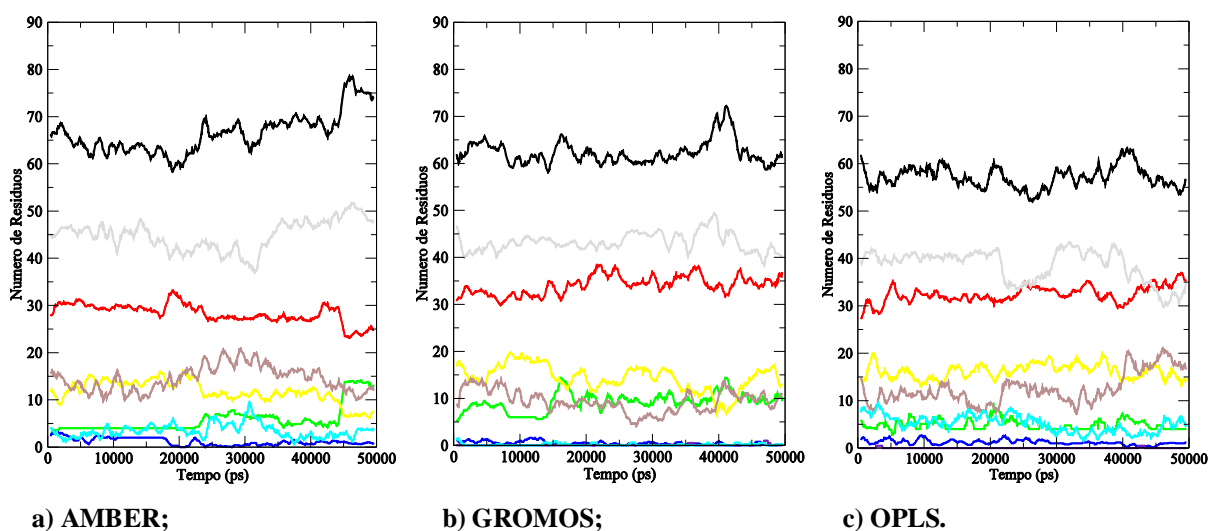
**Figura 18:** Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para os campos de força: a) AMBER; b) GROMOS; c) OPLS. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber:  alça (coil), ■ folha- $\beta$  ( $\beta$ -sheet), ■ segmento  $\beta$  ( $\beta$ -bridge), ■ dobra (bend), ■ volta (turn), ■  $\alpha$ -hélice ( $\alpha$ -helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).

No gráfico de RMSF ( $C\alpha$ ), Figura 17b, as flechas rosas indicam a localização das fitas- $\beta$  originais (S1 e S2; presentes na estrutura obtida do banco de dados PDB: 2PRP) e as caixas azuis indicam a localização das hélices (H1, H2 e H3) nas estruturas protéicas simuladas. As mobilidades observadas nas curvas de RMSF ( $C\alpha$ ) podem ser melhor compreendidas analisando-as em conjunto com os gráficos de estrutura secundária (DSSP), Figura 18, no intervalo de 45 a 50 ns. No intervalo de 109-130 aa (aminoácidos) a simulação com o campo de força OPLS possui um patamar de alta flexibilidade correspondente a sua região N-terminal desestruturada. Nesse mesmo intervalo, o campo de força GROMOS apresenta um pico (Figura 17b) em torno de 125 aa indicando uma dobra (*bend*) flexível conectando as duas fitas- $\beta$  (S1 e S3) na região N-terminal. Já o campo de força AMBER apresenta baixa flexibilidade no intervalo de 109-130 aa (Figura 17b) devido à formação de uma folha- $\beta$  consistente.

De maneira geral, os três campos de força apresentam mobilidades intermediárias nos intervalos 140-160 e 170-190 aa (Figura 17b) que compreendem, respectivamente, as  $\alpha$ -hélices 1 e 2 (H1 e H2). Observa-se ainda que as curvas de RMSF ( $C\alpha$ ) na região de 140-190 aa para os campos de força GROMOS e OPLS apresentam maior flexibilidade em comparação à curva do campo de força AMBER. Isso se deve ao fato de que os campos de força GROMOS e OPLS apresentam uma maior desestruturação das  $\alpha$ -hélices 1 e 2 (H1 e H2), enquanto o campo de força AMBER possui essas mesmas  $\alpha$ -hélices bem conservadas (como pode ser visto nos gráficos da Figura 18). Por fim, nas curvas de RMSF ( $C\alpha$ ) dos três campos de força, a  $\alpha$ -hélice 3 (H3) está associada a uma região de baixa mobilidade (intervalo 205-215 aa), enquanto as fitas- $\beta$ , S1 e S2, estão relacionadas a vales de baixas mobilidades centrados, respectivamente, em 130 aa e 160 aa.

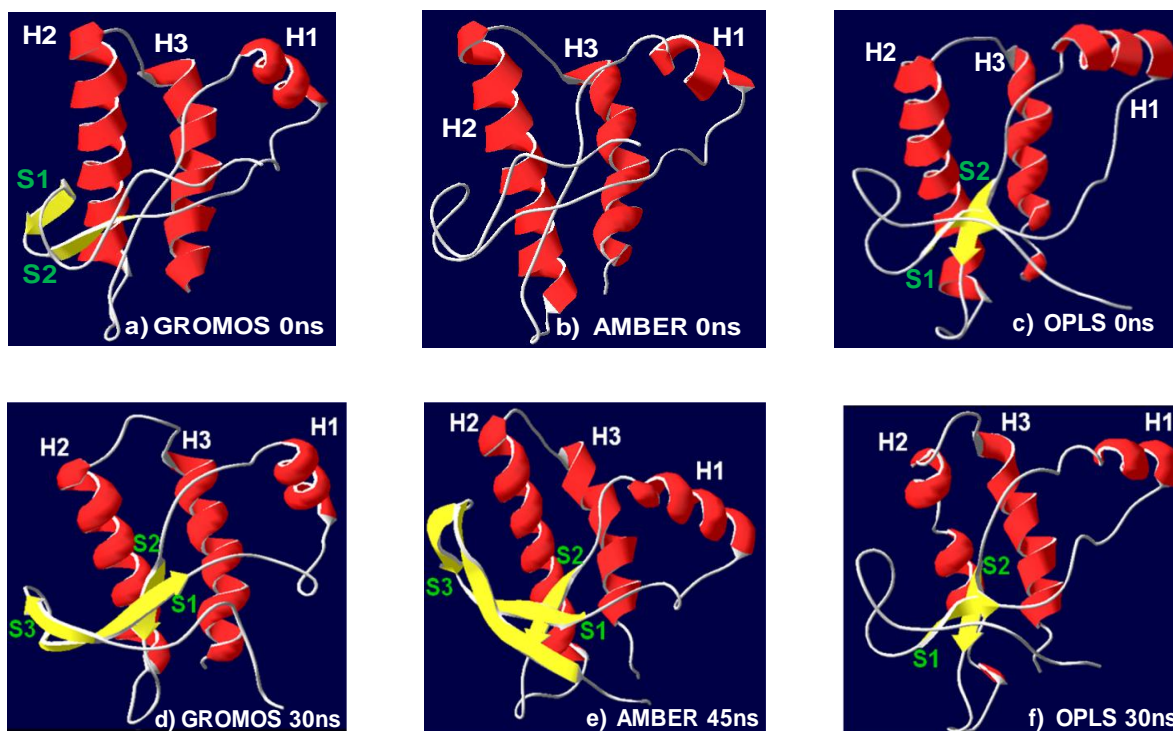
Analisando em conjunto os gráficos de estrutura secundária das Figuras 18 (DSSP) e 19 (SCOUNT), constata-se que, para o campo de força AMBER, o decaimento do conteúdo  $\alpha$ -helicoidal observado nos primeiros 30 ns de simulação é substituído por um acréscimo nos 20 ns posteriores. Assim, ao término da simulação, tem-se um ganho de conteúdo  $\alpha$ -helicoidal em relação à estrutura de partida. Observa-se ainda, para o campo de força AMBER, que a partir de 23 ns ocorre a formação de uma fita- $\beta$  compondo um núcleo de folhas- $\beta$  consistente na região N-terminal. Este núcleo sofre um incremento significativo de resíduos de aminoácidos nos últimos 5 ns de simulação. Para o campo de força GROMOS, verifica-se o surgimento de uma fita- $\beta$  estável (na região N-terminal) que promove um acréscimo de conteúdo de folhas- $\beta$  a partir de 15 ns de simulação. O campo de força OPLS não promove a formação de uma fita- $\beta$  na região N-terminal, mantendo seu conteúdo de folhas- $\beta$

aproximadamente constante e inalterado ao longo da simulação. Por fim, os campos de força GROMOS e OPLS apresentam diminuição (mais acentuadamente o OPLS) do conteúdo  $\alpha$ -helicoidal em função, principalmente, da desestruturação dos resíduos das hélices 2. Sendo que somente o campo de força GROMOS apresentou uma dobra (*bend*) no lugar dos resíduos finais da hélice 2 durante quase todo o intervalo de tempo de simulação.



**Figura 19:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: **a)** AMBER; **b)** GROMOS; **c)** OPLS. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: ■ *Structure*, ■ alça (coil), ■ folha- $\beta$  ( $\beta$ -sheet), ■ segmento  $\beta$  ( $\beta$ -bridge), ■ dobra (*bend*), ■ volta (turn), ■  $\alpha$ -hélice (A-helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).

A Figura 20 apresenta imagens dos fragmentos protéicos para os campos de força GROMOS, AMBER e OPLS em diferentes instantes de tempo de simulação. Essas imagens permitem visualizar a evolução das estruturas ao longo das simulações. De fato, essas imagens corroboram os resultados das análises de estrutura secundária (Figuras 18 e 19), indicando um alto conteúdo de folhas- $\beta$  para o campo força AMBER, seguido pelo campo de força GROMOS, e um conteúdo de folhas- $\beta$  praticamente inalterado (em relação ao início da simulação) para o campo de força OPLS.



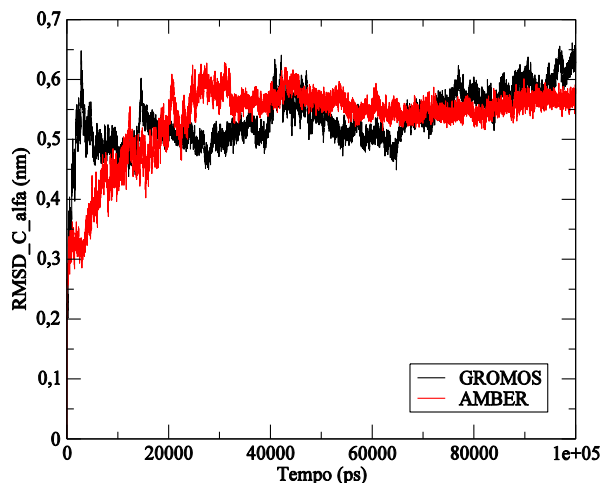
**Figura 20:** Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV): **a)** GROMOS em  $t = 0$  ns; **b)** AMBER em  $t = 0$  ns; **c)** OPLS em  $t = 0$  ns; **d)** GROMOS em  $t = 30$  ns; **e)** AMBER em  $t = 45$  ns; **f)** OPLS em  $t = 30$  ns.

### 5.1.2 GROMOS e AMBER: estendendo as trajetórias

Como o AMBER foi o campo de força que apresentou o maior acréscimo de folhas- $\beta$  na região N-terminal e uma vez que esse acréscimo acentuado surgiu somente nos últimos 5 ns, é importante estender a trajetória desta simulação. Com isto pretende-se verificar se o núcleo de folhas- $\beta$  formado é, de fato, estável e se o conteúdo  $\alpha$ -helicoidal permanecerá aumentando ou se retornará ao patamar observado nos primeiros 30 ns de simulação. O campo de força GROMOS, por sua vez, além de apresentar um aumento significativo de folhas- $\beta$  (embora em menor extensão que o campo de força AMBER) na região N-terminal, também apresentou uma diminuição de conteúdo  $\alpha$ -helicoidal nos últimos 10 ns de simulação.

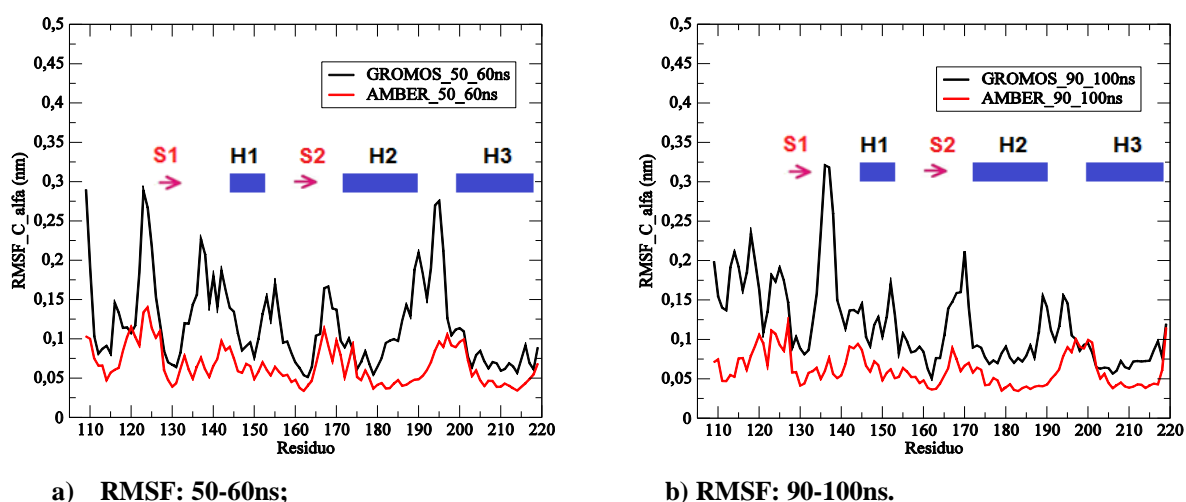
Assim, as trajetórias das simulações dos campos de força AMBER e GROMOS foram estendidas, até 100 ns, com o intuito de verificar se a tendência de formação de folhas- $\beta$  na região N-terminal seria mantida. O campo de força OPLS não foi incluído nesta etapa de extensão das trajetórias por ter apresentado baixa tendência à formação de uma nova fita- $\beta$  na região N-terminal.





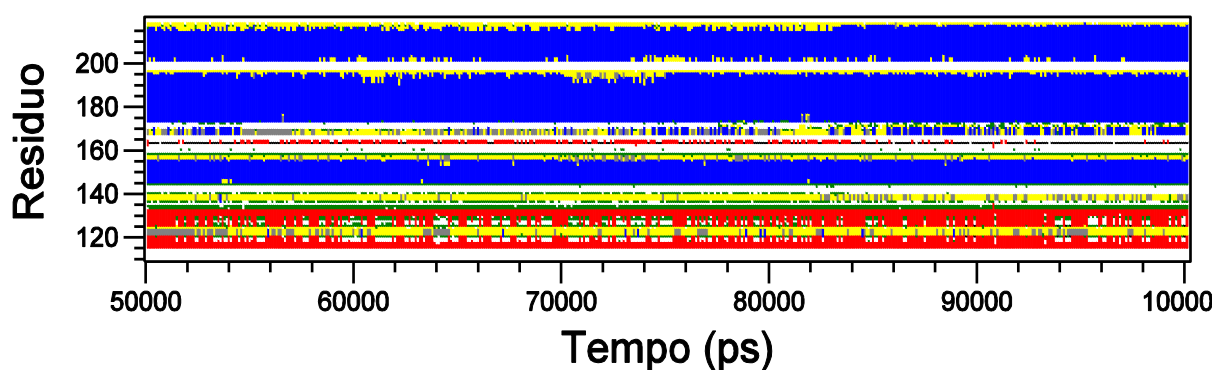
**Figura 21:** Gráficos de RMSD de  $C\alpha$  em função do tempo de simulação para as trajetórias estendidas dos campos de força AMBER e GROMOS.

Analisando o gráfico de RMSD ( $C\alpha$ ), Figura 21, observa-se que para o campo de força AMBER o patamar que havia sido alcançado em 50 ns (em torno de 0,57 nm) é basicamente mantido até o final da simulação. No entanto, o campo de força GROMOS que até 50 ns havia apresentado um patamar de convergência (em torno de 0,55 nm) inferior ao do AMBER, termina a simulação com um valor de RMSD ( $C\alpha$ ) superior ao de AMBER. De fato, a partir de 60 ns de simulação, a curva de RMSD ( $C\alpha$ ) do campo de força GROMOS apresenta um perfil crescente, indicando que a estrutura continua se alterando em relação à estrutura de partida (de 0 ns) e termina a simulação com um valor de cerca de 0,65 nm.

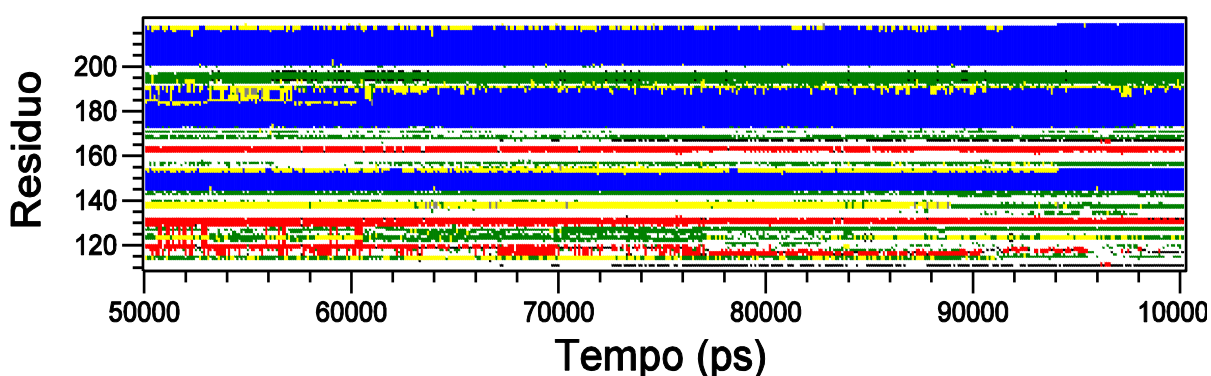


**Figura 22:** Gráficos de RMSF ( $C\alpha$ ) em função do resíduo para as simulações estendidas dos campos de força AMBER e GROMOS representando diferentes intervalos de tempo de simulação: **a)** 50 a 60 ns; **b)** 90 a 100 ns. No gráfico de RMSF ( $C\alpha$ ), as flechas rosas indicam a localização das fitas- $\beta$  originais (presentes na estrutura inicial de 2PRP) e as caixas azuis indicam a localização das hélices (H1, H2 e H3) nas estruturas proteicas simuladas.

Analisando os gráficos de RMSF ( $C\alpha$ ), Figura 22, observa-se que o perfil da curva do campo de força AMBER entre 50-60 ns (ver Figura 22a) e 90-100 ns (ver Figura 22b) é basicamente o mesmo. Isto indica que o fragmento protéico simulado por este campo de força não está mais sofrendo significativas mudanças estruturais no intervalo de 50 à 100 ns de simulação.



a) AMBER;

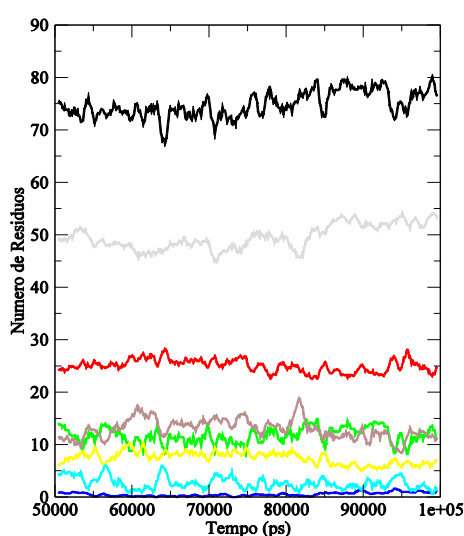


a) GROMOS.

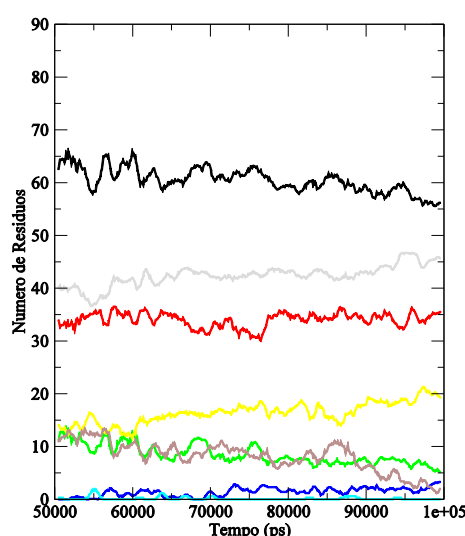
**Figura 23:** Gráfico de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para : **a)** AMBER; **b)** GROMOS. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: □ alça (coil), ■ folha- $\beta$  ( $\beta$ -sheet), ■ segmento  $\beta$  ( $\beta$ -bridge), ■ dobra (bend), ■ volta (turn), ■  $\alpha$ -hélice (A-helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).

A curva de RMSF ( $C\alpha$ ) do campo de força GROMOS, por outro lado, mostra mobilidades bastante diferentes nos intervalos de 50-60 ns (ver Figura 22a) e 90-100 ns (ver Figura 22b). A Figura 22a mostra que de 50-60 ns o campo de força GROMOS apresenta um pico de alta mobilidade, em torno de 195 aa, em sua curva de RMSF ( $C\alpha$ ), o que confere com a maior desestruturação dos resíduos finais da  $\alpha$ -hélice 2 (H2) neste intervalo (conforme

mostra a Figura 23b). Este pico, no entanto, é visivelmente suavizado na curva de RMSF ( $C\alpha$ ) no intervalo de 90-100 ns (Figura 22b), uma vez que nesse intervalo há uma regeneração de resíduos, que voltam a assumir estrutura  $\alpha$ -helicoidal. Analisando ainda a curva de GROMOS, Figura 22b, observa-se que a região N-terminal é marcada por um patamar de alta mobilidade, o que confere com o fato de que nos 10 ns finais de simulação ocorre uma perda significativa de conteúdo de folhas- $\beta$  nessa região (conforme se pode observar pela Figura 23b em torno de 120 aa). Verifica-se ainda, pela Figura 23b, que os resíduos finais da  $\alpha$ -hélice 1 (H1) e parte dos resíduos finais da  $\alpha$ -hélice 2 (H2) do campo de força GROMOS são reestruturados até o final da simulação.



a) AMBER;

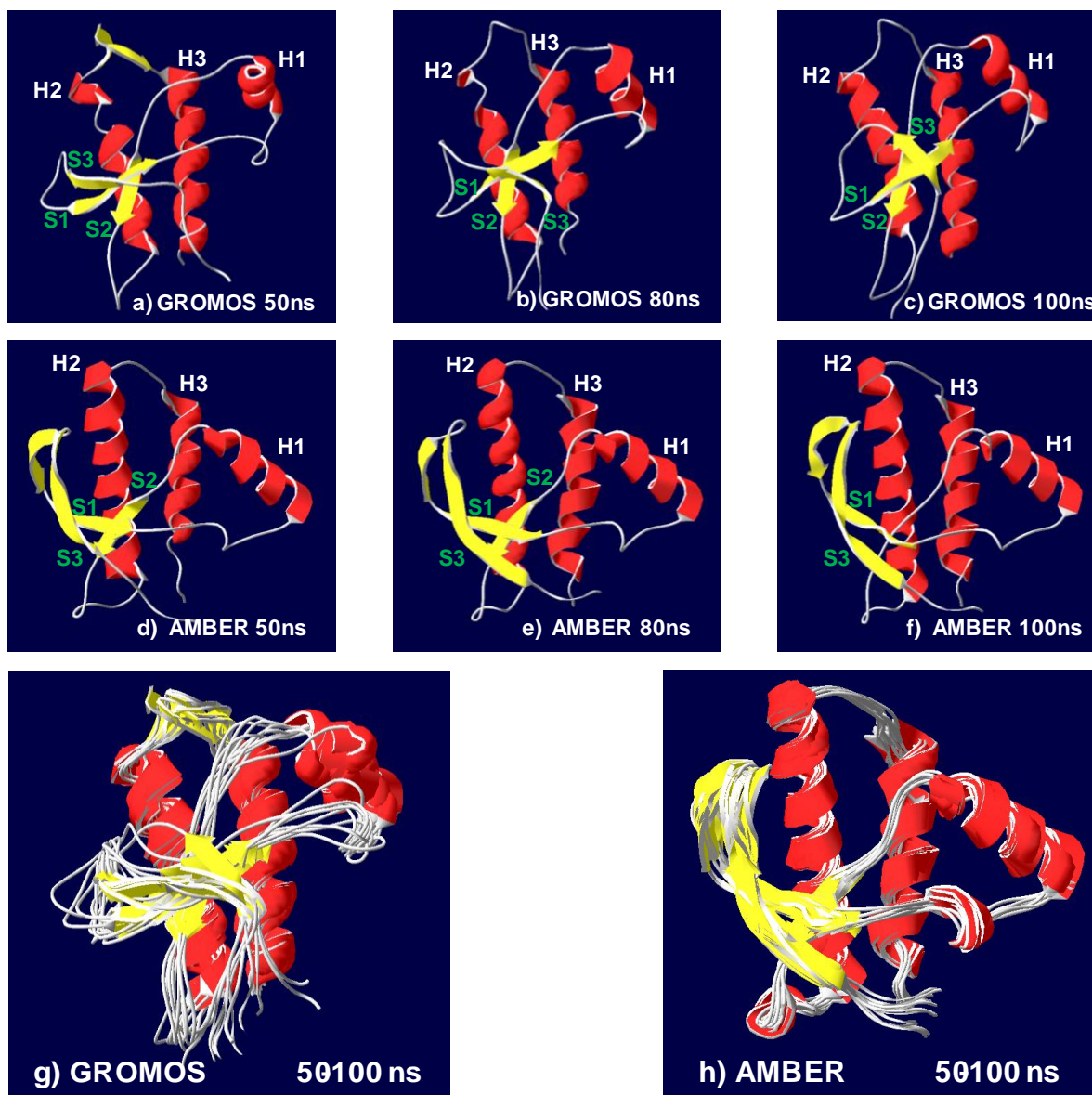


b) GROMOS.

**Figura 24:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação para: **a)** AMBER; **b)** GROMOS. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: **■** *Structure*, **■** alça (coil), **■** folha- $\beta$  ( $\beta$ -sheet), **■** segmento  $\beta$  ( $\beta$ -bridge), **■** dobra (bend), **■** volta (turn), **■**  $\alpha$ -hélice (A-helix), **■** 5-hélice (5-helix) e **■** 3-hélice (3-helix).

Avaliando a Figura 24a, observa-se que o conteúdo de folhas- $\beta$  do campo de força AMBER é constante ao longo do intervalo de 50-100 ns de simulação, enquanto o seu conteúdo  $\alpha$ -helicoidal aumenta ao final desse intervalo. No que diz respeito ao campo de força GROMOS, verifica-se pela Figura 24b que o seu conteúdo de folhas- $\beta$  decai continuamente no intervalo de 50-100 ns de simulação. Por outro lado, o conteúdo  $\alpha$ -helicoidal do campo de

força GROMOS aumenta continuamente ao longo desse intervalo de simulação, assim como ocorre com o AMBER.



**Figura 25:** Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV) para: **a)** GROMOS em  $t = 50$  ns; **b)** GROMOS em  $t = 80$  ns; **c)** GROMOS em  $t = 100$  ns; **d)** AMBER em  $t = 50$  ns; **e)** AMBER em  $t = 80$  ns; **f)** AMBER em  $t = 100$  ns. Sobreposição de imagens (SPDBV) correspondentes aos instantes de tempos de 50, 60, 70, 80, 90 e 100 ns para: **g)** GROMOS; **h)** AMBER.

As imagens da Figura 25 mostram a evolução dos fragmentos protéicos simulados com os diferentes campos de força. Comparativamente, verifica-se pelas imagens (da Figura 25a até a Figura 25f) que o campo de força AMBER apresenta maior conteúdo de folhas- $\beta$  e

de  $\alpha$ -hélices, ao longo da simulação, do que o campo de força GROMOS. As Figuras 25g e 25h, por sua vez, mostram imagens sobrepostas coletadas de 50 a 100 ns (a cada 10 ns) de simulação para cada um dos campos de força. Por estas imagens, verifica-se que, de fato, a estrutura simulada pelo campo de força AMBER oscila muito menos que a simulada pelo GROMOS como havia sido anteriormente discutido nas análises deste capítulo.

### 5.1.3 Conclusões Parciais

Pela análises de RMSD ( $C\alpha$ ), RMSF ( $C\alpha$ ) e de estrutura secundária apresentadas na Seção 5.1.1, observou-se uma perda mais acentuada de resíduos  $\alpha$ -helicoidais nas simulações realizadas com os campos de força GROMOS e OPLS. Por outro lado, somente os campos de força GROMOS e AMBER apresentaram aumento de folhas- $\beta$  em função do surgimento de uma fita- $\beta$  na região N-terminal. Embora essas alterações conformacionais não forneçam um núcleo de folhas- $\beta$  ideal, os resultados para este intervalo de tempo de simulação indicam que o campo de força GROMOS é o mais adequado para avaliar o processo de conversão estrutural da proteína príon celular simulada sob baixo pH. Isso porque a simulação com o campo de força GROMOS foi a única que resultou em diminuição de conteúdo  $\alpha$ -helicoidal e aumento de folhas- $\beta$  no intervalo de tempo avaliado.

Na Seção 5.1.2 foram apresentados os resultados da extensão das trajetórias das simulações com os campos de força GROMOS e AMBER (escolhidos para esta etapa por terem apresentado maior rendimento na formação de folhas- $\beta$ ). As análises indicaram que o campo de força AMBER apresentou maior conteúdo de folhas- $\beta$  e de  $\alpha$ -hélices do que o GROMOS. De fato, o campo de força AMBER apresentou pouca oscilação estrutural no intervalo de 50-100 ns de simulação, enquanto o campo de força GROMOS não manteve as folhas- $\beta$  na região N-terminal até o final da simulação.

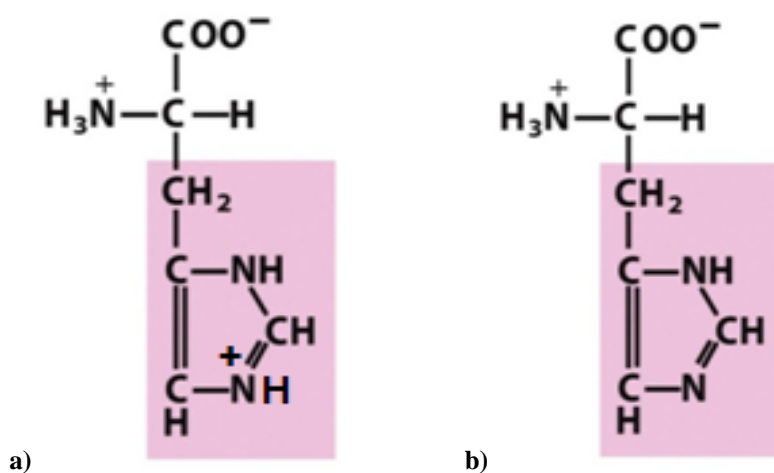
Em suma, o campo de força AMBER foi mais eficiente no quesito formação de folhas- $\beta$ , enquanto o GROMOS foi mais eficiente na desestruturação das  $\alpha$ -hélices. O ideal seria um campo de força que apresentasse os dois eventos simultaneamente, conforme a ideia original de transição conformacional da proteína príon celular. Para elucidar melhor as diferenças de comportamento estrutural entre estes campos de força é importante que também sejam comparados os seus desempenhos na simulação do fragmento protéico em pH neutro.

## 5.2 INFLUÊNCIA DA FORMA DE PROTONAÇÃO DOS AMINOÁCIDOS

### 5.2.1 Investigação dos estados de protonação da Histidina (HisA, HisB e HisH)

Continuando com o objetivo de estabelecer um protocolo de simulação, buscou-se estudar os estados de protonação dos resíduos histidina e seu efeito sobre a estrutura simulada em pH neutro. Esta etapa do procedimento foi inicialmente desenvolvida com o intuito de minimizar o conteúdo de fitas- $\beta$  na região N-terminal da estrutura simulada em pH neutro.

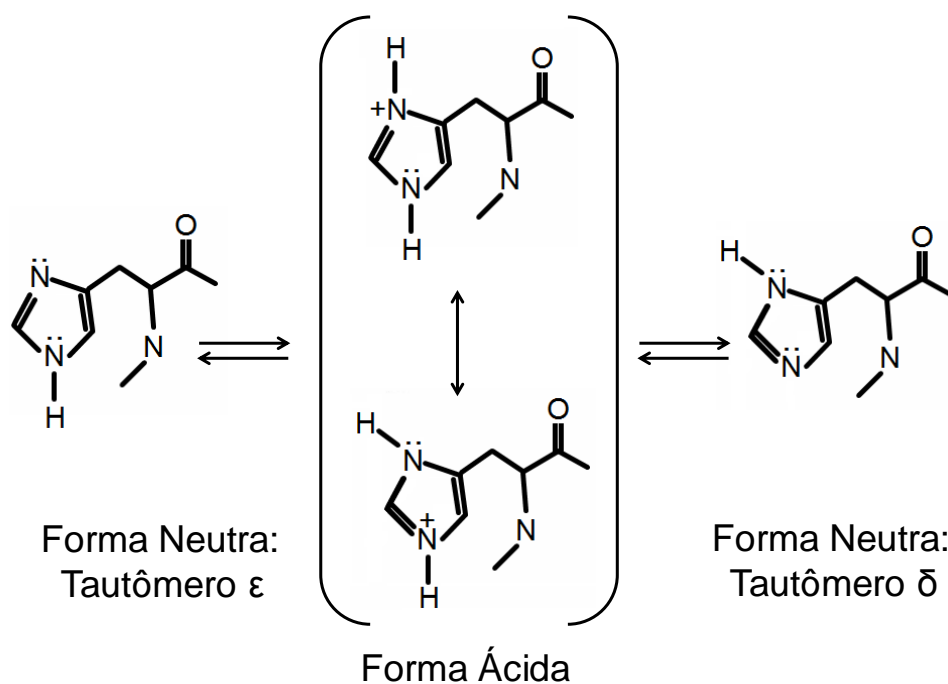
O aminoácido histidina, juntamente com a lisina e a arginina, pertence ao grupo de cadeia lateral R positivamente carregada (categoria também conhecida como a de aminoácidos básicos). A cadeia lateral da histidina é composta pelo grupo aromático imidazol e pode estar tanto carregada positivamente (forma protonada) como não carregada em pH 7 (ver Figura 26), uma vez que é o único aminoácido que possui uma cadeia lateral com  $pK_a$  próximo da neutralidade<sup>(50)</sup>. Assim, em pH neutro a histidina pode ou não estar carregada, dependendo do ambiente químico ao seu redor.<sup>(82)</sup>



**Figura 26:** Representação bioquímica do aminoácido histidina com a cadeia lateral: **a)** carregada positivamente; **b)** não-carregada (imagens adaptadas do livro Princípios de Bioquímica de Lehninger).<sup>(50)</sup>

Além da possibilidade de existir nas formas neutra e carregada positivamente em pH neutro, a histidina possui ainda dois tipos de representação estrutural para a forma neutra. Isso ocorre em função de a cadeia lateral da histidina possuir ressonância no grupo imidazol. Desta forma, qualquer um dos átomos de nitrogênio do anel pode ser protonado, dando origem às duas formas neutras para a histidina: os tautômeros  $\delta$  e  $\epsilon$  (ver Figura 27) nos quais,

respectivamente, os nitrogênios  $\delta$  e  $\epsilon$  são protonados. Obviamente, esses tautômeros possuem diferentes valores de  $pK_a$ , sendo o tautômero  $\delta$  mais ácido que o tautômero  $\epsilon$  (o  $pK_a$  do tautômero  $\delta$  é de 6,6 enquanto o do tautômero  $\epsilon$  é de 7,0).<sup>(83)</sup> Desta forma, dadas as peculiaridades do aminoácido histidina, deve-se tomar cuidado ao selecionar computacionalmente o seu estado de protonação para representar um dado pH (ou faixa de pH).

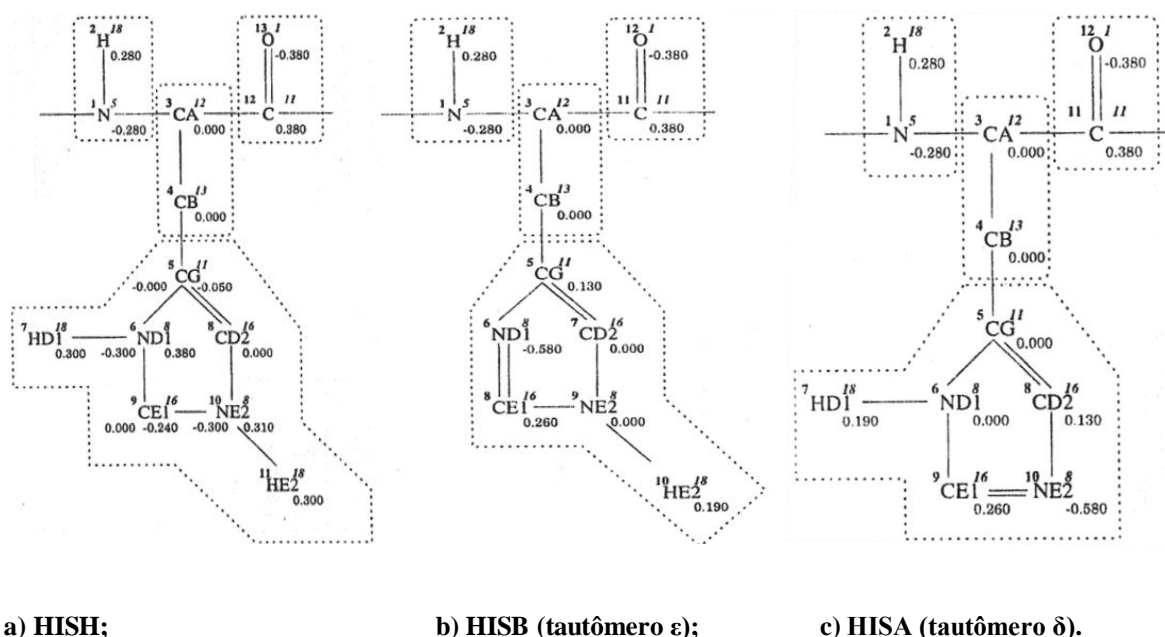


**Figura 27:** Representação dos estados iônicos e tautoméricos da cadeia lateral do aminoácido histidina. O equilíbrio entre as diferentes formas da histidina se deve à ressonância que ocorre no grupamento imidazol da cadeia lateral. Esta ressonância permite que qualquer um dos átomos de nitrogênio do anel seja protonado, dando origem a duas formas neutras possíveis para o aminoácido histidina: os tautômeros  $\delta$  e  $\epsilon$  nos quais, respectivamente, os nitrogênios  $\delta$  e  $\epsilon$  são protonados.

De fato, a questão da protonação dos resíduos histidina em pH neutro é bastante delicada e uma estrutura protéica pode, inclusive, possuir parte de seus resíduos histidina neutros e parte protonados em pH neutro. Na literatura há trabalhos nos quais a forma da histidina protonada (HisH; em conjunto com os ácidos aspárticos e glutâmicos desprotonados) foi empregada para representar todos os resíduos histidina de um fragmento (resíduos 125-228) da proteína príon celular humana simulada em um pH levemente ácido (em torno de

5).<sup>(84)</sup> Por outro lado, um estudo sistemático de manipulação de  $pK_a$  de resíduos histidina, realizado por Langella e colaboradores<sup>(83)</sup>, indica que a proteína príon celular humana (fragmento 125-228) apresenta dois de seus resíduos histidina (155 e 187) neutros e dois outros resíduos histidina (140 e 177) preferencialmente protonados sob condições fisiológicas (pH 7,0). Desta forma, a seleção do estado de protonação dos resíduos histidina pode influenciar os resultados das simulações computacionais da proteína príon celular em pH neutro.

No presente trabalho, foi empregado o campo de força GROMOS 96 53a<sup>(62)</sup> para investigar os diferentes estados de protonação da cadeia lateral dos resíduos de aminoácidos histidina e sua influência no conteúdo de estrutura secundária do fragmento de proteína príon celular do hamster Sírio. Para isso, foram realizadas várias simulações com diferentes estados de protonação dos 4 resíduos histidina (localizados nas posições 111, 140, 177 e 187) presentes no fragmento protéico.



**Figura 28:** Representações do aminoácido histidina conforme o campo de força GROMOS: a) HISH, com nitrogênios ND1 e NE2 protonados e com uma carga resultante de +1; b) HISB, com o nitrogênio NE2 protonado e com carga resultante nula; c) HISA, com o nitrogênio ND1 protonado e com carga resultante nula. HISH é a representação da histidina carregada, enquanto HISA e HISB são ambas formas neutras da histidina. (Adaptado de Biomolecular Simulation: The GROMOS 96 Manual and User Guide)<sup>(85)</sup>.



No campo de força GROMOS, a histidina possui duas representações neutras (HISA e HISB) e uma ácida (ou protonada; HISH). Assim, o próton pode estar (conforme apresentado na Figura 28) no nitrogênio ND1 (HISA), no nitrogênio ND2 (HISB) ou em ambos (HISH). Por padrão estas seleções são feitas automaticamente de forma a produzir a conformação ideal para o estabelecimento de ligações de hidrogênio. Com base em critérios geométricos, especificados pelo ângulo hidrogênio-doador-aceptor e pela distância doador-aceptor máximos, as ligações hidrogênio são definidas<sup>(38)</sup>.

**Tabela II:** Número de moléculas e íons de cada um dos sistemas simulados, e a forma de protonação dos resíduos de aminoácidos histidinas, ácidos aspárticos e glutâmicos em cada uma das simulações.

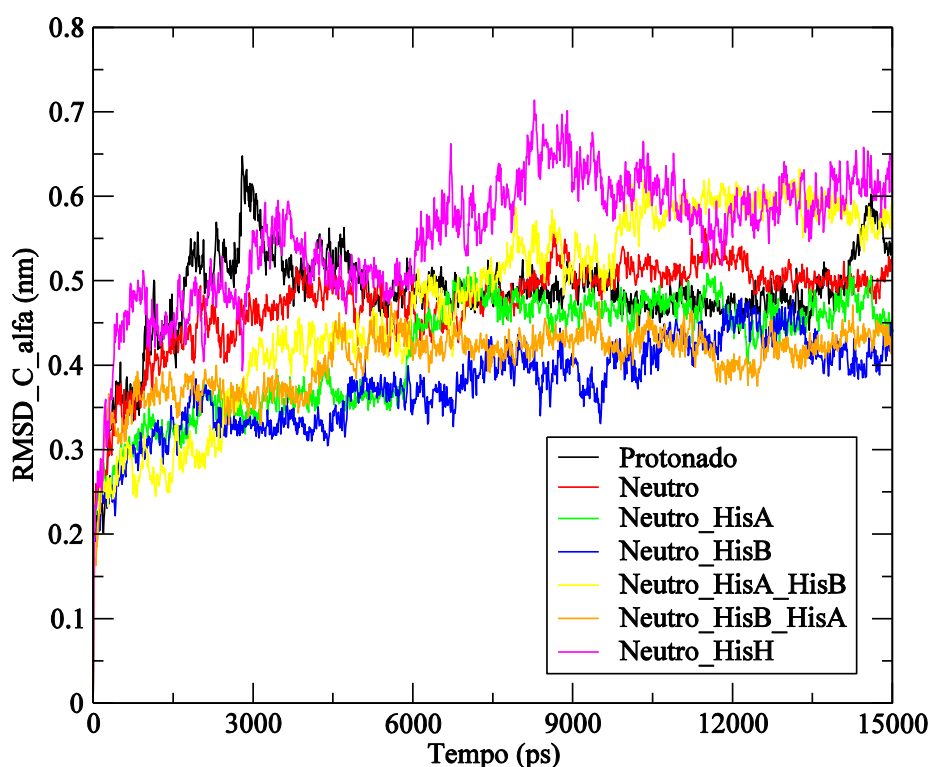
Simulação	Solvente (H <sub>2</sub> O)	Proteína	Na <sup>+</sup>	Cl <sup>-</sup>	HIS 111	HIS 140	HIS 177	HIS 187	ASP's	GLU's
<b>Neutro</b>	13161	1	40	39	HisB	HisB	HisA	HisB	Asp	Glu
<b>Neutro_HisB</b>	13158	1	40	39	HisB	HisB	HisB	HisB	Asp	Glu
<b>Neutro_HisA</b>	13161	1	40	39	HisA	HisA	HisA	HisA	Asp	Glu
<b>Neutro_HisA_HisB</b>	13158	1	40	39	HisA	HisA	HisB	HisB	Asp	Glu
<b>Neutro_HisB_HisA</b>	13158	1	40	39	HisB	HisB	HisA	HisA	Asp	Glu
<b>Neutro_HisH</b>	13161	1	39	42	HisH	HisH	HisH	HisH	Asp	Glu
<b>Protonado</b>	10892	1	30	44	HisH	HisH	HisH	HisH	AspH	GluH

\*Sendo HisB (tautômero  $\epsilon$ ), HisA (tautômero  $\delta$ ) e HisH (histidina carregada) as três formas de protonação possíveis para os aminoácidos histidina; Asp e Glu, os ácidos desprotonados; e AspH e GluH, os ácidos protonados. Esses estados de protonação são definidos conforme o padrão dos resíduos do GROMOS 96 53a6<sup>(62)</sup>.

Inicialmente foi simulada uma estrutura em pH neutro na qual os quatro resíduos histidina foram protonados conforme o padrão do GROMACS (resíduos 111, 140 e 187: HISB; resíduo 177: HISA). A esse sistema foi dado o nome de **Neutro**. Na sequência, visando explorar as possíveis combinações de estados de protonação dos resíduos histidina em pH neutro, foram produzidos outros sistemas para simulação: **Neutro\_HisB** (resíduos 111, 140, 177 e 187 = HISB), **Neutro\_HisA** (res. 111, 140, 177 e 187 = HISA), **Neutro\_HisA\_HisB** (resíduos 111 e 140 = HISA; resíduos 177 e 187 = HISB), **Neutro\_HisB\_HisA** (resíduos 111 e 140 = HISB; resíduos 177 e 187 = HISA) e **Neutro\_HisH**. Todos os sistemas neutros acima mencionados foram produzidos empregando o modelo 13 do RMN e possuíam os resíduos Asp e Glu desprotonados, ou seja negativamente carregados. A Tabela II, acima, apresenta os componentes de cada um dos sistemas estudados neste capítulo.

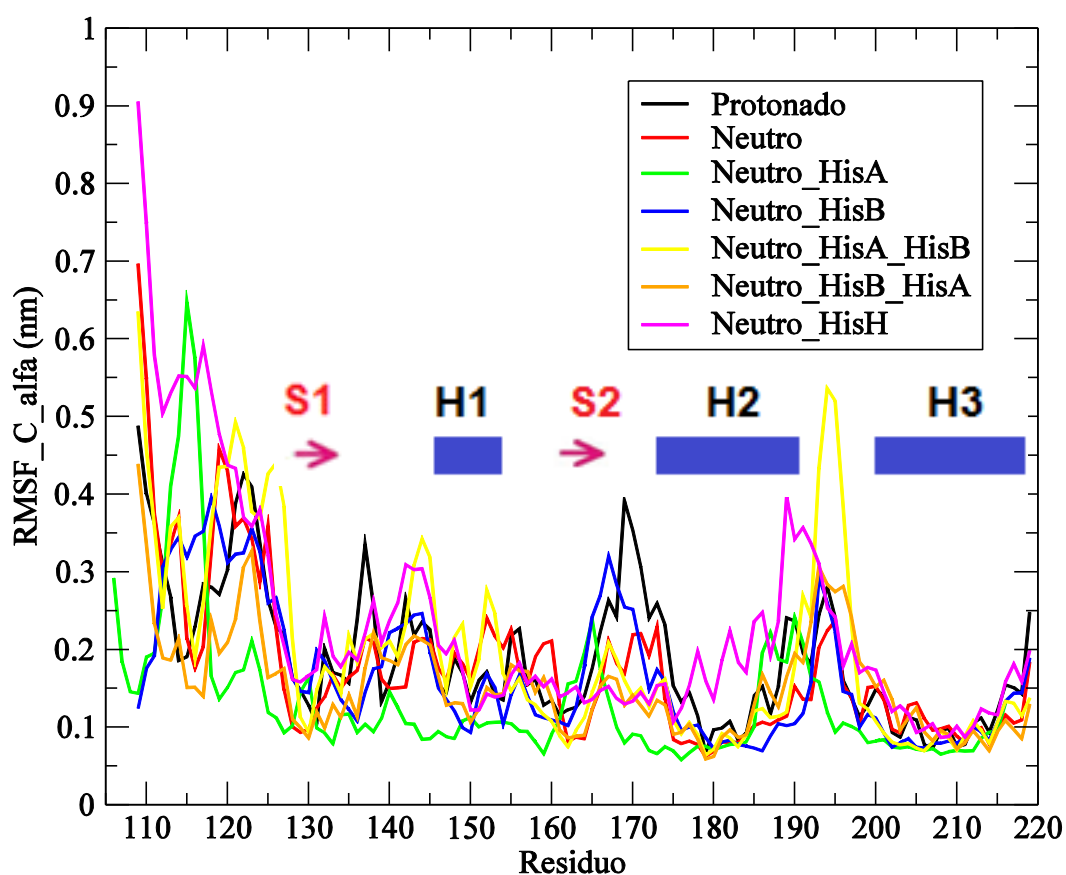
Pensando rigorosamente nos valores de  $pK_a$ , embora no presente texto tenha-se denominado o sistema com os quatro resíduos histidina protonados de **Neutro\_HisH**, este sistema corresponde a um pH levemente ácido (entre 4,5-5,9; pois nesse intervalo de pH as histidinas são carregadas positivamente e os ácidos glutâmico e aspártico estão desprotonados, ou negativamente carregados, de acordo com a representação da estrutura **Neutro\_HisH**). Desta forma, a denominação **Neutro\_HisH** é uma simples nomenclatura e o fragmento protéico correspondente não representa um sistema em pH neutro.

Para manter um controle das diferenças estruturais, os resultados das simulações dos sistemas “**Neutros**” foram comparados com o do sistema **Protonado** (resíduos 111, 140, 177 e 187 = HISH; resíduos Asp e Glu protonados, correspondendo a um valor de pH igual ou menor que 3) já apresentado no Seção 5.1 (correspondendo à simulação do campo de força GROMOS) deste trabalho. Todas as simulações aqui apresentadas empregaram a estrutura 13 do RMN<sup>(18)</sup> e o campo de força GROMOS 96 53a6<sup>(62)</sup>. As análises apresentadas a seguir foram realizadas para simulações de DM com durações de 15 ns.

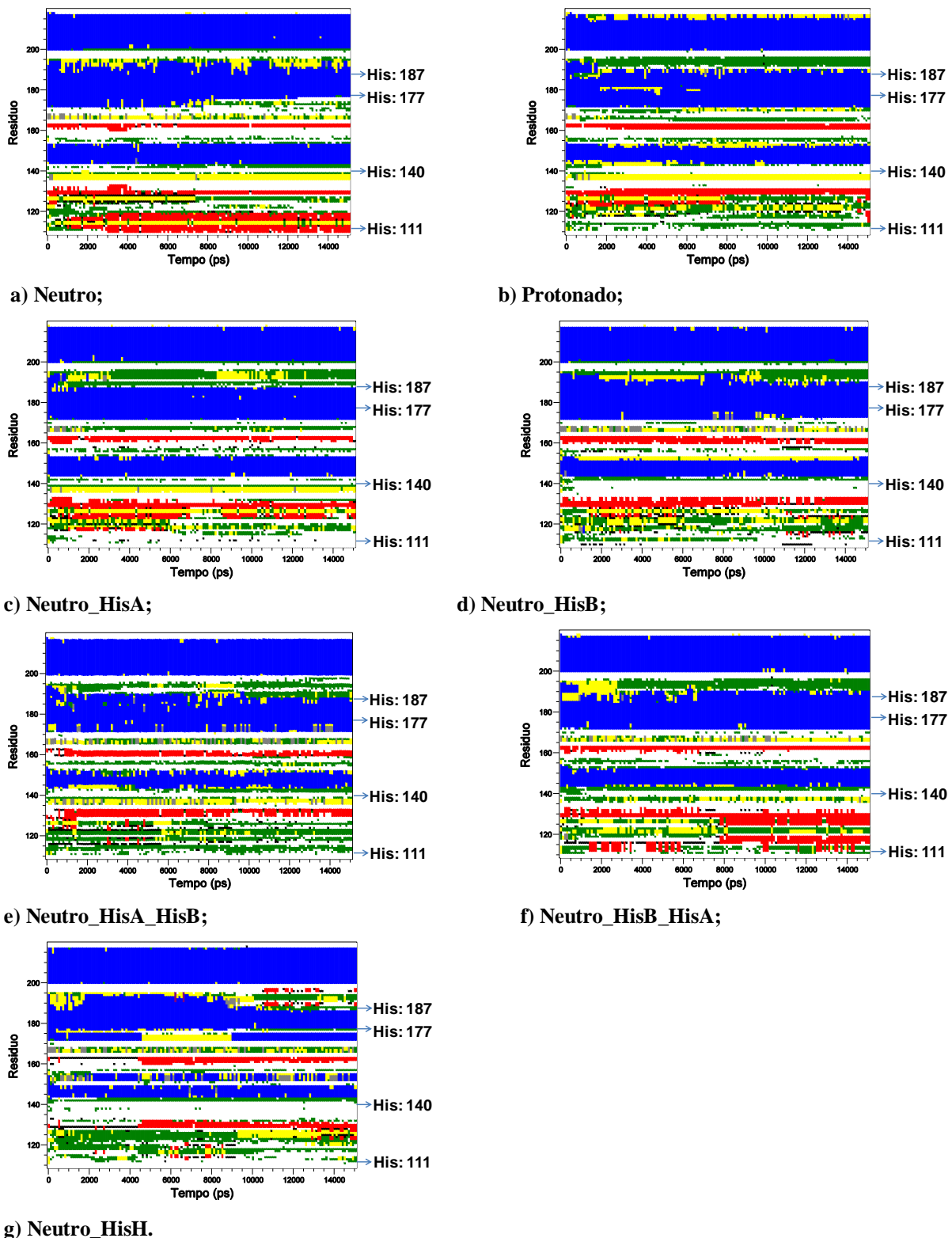


**Figura 29:** Gráfico de RMSD ( $C\alpha$ ) em função do tempo para as simulações: Protonado, Neutro, Neutro\_HisA, Neutro\_HisB, Neutro\_HisA\_HisB, Neutro\_HisB\_HisA e Neutro\_HisH.

A manipulação das histidinas resultou em curvas de RMSD (Figura 29) e RMSF (Figura 30) bastante distintas entre as diferentes simulações. As curvas de RMSD ( $C\alpha$ ) de **Neutro\_HisA\_HisB** e **Neutro\_HisH** apresentam valores superiores, as de **Neutro** e **Neutro\_HisA** mostram valores similares e as de **Neutro\_HisB** e **Neutro\_HisB\_HisA** apresentam valores inferiores à curva de **Protonado**. Pelos modelos de influência de pH da literatura<sup>(33, 75, 84)</sup>, quanto menor o pH do sistema, maior o valor de RMSD associado. Ou seja, a curva de RMSD ( $C\alpha$ ) da estrutura simulada em pH neutro deve ser inferior à da estrutura em pH ácido (ou especificamente, a simulação denominada de **Protonado**). Conforme este modelo, podemos excluir a simulação de **Neutro\_HisA\_HisB** como possível candidato à estrutura neutra.

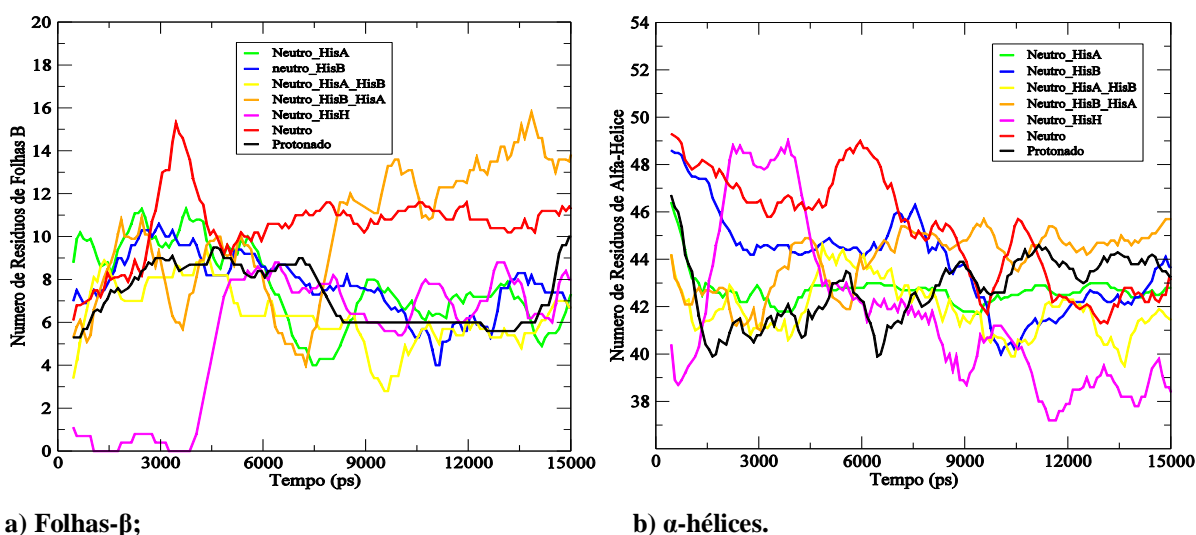


**Figura 30:** Gráfico de RMSF ( $C\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 15 ns de cada simulação) para as simulações: Protonado, Neutro, Neutro\_HisA, Neutro\_HisB, Neutro\_HisA\_HisB, Neutro\_HisB\_HisA e Neutro\_HisH. As flechas rosas indicam a localização das fitas- $\beta$  originais (presentes na estrutura inicial de 2PRP) e as caixas azuis indicam a localização das  $\alpha$ -hélices (H1, H2 e H3) nas estruturas protéicas simuladas.



**Figura 31:** Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP): **a)** Neutro; **b)** Protonado; **c)** Neutro\_HisA; **d)** Neutro\_HisB; **e)** Neutro\_HisA\_HisB; **f)** Neutro\_HisB\_HisA; **g)** Neutro\_HisH. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber:  alça (coil),  folha-β (β-sheet),  segmento β (β-bridge),  dobra (bend),  volta (turn),  α-hélice (A-helix),  5-hélice (5-helix) e  3-hélice (3-helix).

Analisando os gráficos de estrutura secundária (DSSP) da Figura 31, observa-se que, de fato, as simulações **Neutro**, **Neutro\_HisA** e **Neutro\_HisB\_HisA** produzem um conteúdo de folhas- $\beta$  mais acentuado que **Protonado** (que é a estrutura simulada em pH ácido; mais ácido que **Neutro\_HisH**) na região N-terminal. Como este resultado está em oposição ao que sugere os modelos de estudo do e influência de pH, estas simulações são excluídas como potenciais modelos para o pH neutro. Analisando ainda os gráficos de estrutura secundária, Figura 31, observa-se que há uma similaridade entre as regiões 145-219 de **Neutro\_HisA** e **Neutro\_HisB\_HisA** e entre as de **Neutro\_HisB** e **Neutro\_HisA\_HisB**. Mas não parece ter uma semelhança entre as regiões 109-145 para as simulações **Neutro\_HisA** e **Neutro\_HisA\_HisB** e para **Neutro\_HisB** e **Neutro\_HisB\_HisA**. Isso mostra que o estado global de protonação da proteína influencia no conteúdo estrutural da região N-terminal, mas as simulações precisam ser estendidas para verificar melhor este fato.



**Figura 32:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: **a)** folhas  $\beta$ ; **b)**  $\alpha$ -hélices. As curvas nos gráficos referem-se às simulações de: Protonado, Neutro, Neutro\_HisA, Neutro\_HisB, Neutro\_HisA\_HisB, Neutro\_HisB\_HisA e Neutro\_HisH.

Desta forma, segundo as análises de RMSD ( $C\alpha$ ) e estrutura secundária (Figuras 31 e 32), verifica-se que a única estrutura (entre as estruturas em pH neutro) que apresentou baixo conteúdo folhas- $\beta$  na região N-terminal e RMSD ( $C\alpha$ ) inferior à estrutura **Protonada** foi **Neutro\_HisB**. De fato, analisando os gráficos de número de resíduos envolvidos em tipos de

estrutura secundária ao longo do tempo de simulação, Figura 32, observa-se que **Neutro\_HisB** é a única simulação que apresenta simultaneamente menor conteúdo de folhas- $\beta$  e maior conteúdo  $\alpha$ -helicoidal em comparação com **Protonado**.

Assim, pode-se, em primeira análise, selecionar **Neutro\_HisB** como o melhor candidato a representar a estrutura em pH neutro. Esta estrutura ainda está distante do que sugere o modelo de pH neutro, pois o valor de RMSD ( $C\alpha$ ) e RMSF ( $C\alpha$ ) são somente levemente inferiores à da estrutura **Protonada**, porém o gráfico de estrutura secundária, Figura 32 (SCOUNT), já consegue distinguir mais adequadamente os dois pH de estudo (Protonado: pH < 4; Neutros: pH  $\approx$  7). De fato, escolhendo a simulação **Neutro\_HisB** a distinção entre os pH (neutro e baixo) ainda não está em uma situação ideal, mas já mostra resultados mais de acordo com os modelos de influência do pH que se tem descrito na literatura (em comparação com o que se tinha empregando a simulação **Neutro** para pH neutro).

Conforme os estudos de Alonso e colaboradores<sup>(33)</sup> (que trabalharam com estruturas equivalentes ao que aqui denominamos de **Protonado** e **Neutro**, ou pH 7 e pH menor que 4, empregando simulações de 10 ns), além de **Protonado** apresentar valores de RMSD ( $C\alpha$ ) e RMSF ( $C\alpha$ ) superiores aos de **Neutro**, também apresenta maior conteúdo de folhas- $\beta$  e menor conteúdo de  $\alpha$ -hélices em relação a **Neutro**. No presente trabalho, no entanto, **Neutro** apresentou um conteúdo de folhas- $\beta$  superior e um conteúdo de  $\alpha$ -hélices inferior ao de **Protonado**. Além disso, as curvas de RMSD ( $C\alpha$ ) e RMSF ( $C\alpha$ ) obtidas para **Neutro** mostraram-se similares às de **Protonado**.

Um estudo de influência de pH foi ainda realizado por Campos e colaboradores<sup>(75)</sup>. Estes trabalharam com os valores de pH 2, 4, 5, 6 e 7 empregando o príon humano (fragmento 90-231) e constataram que quanto maior o pH, maiores são os valores das curvas de RMSD ( $C\alpha$ ) e RMSF ( $C\alpha$ ). Aplicando esse resultado às simulações realizadas no presente trabalho, equivaleria a dizer que a curva de RMSD ( $C\alpha$ ) de **Protonado** (pH menor que 4) apresenta um valor de convergência maior que o de **Neutro\_HisH** (pH em torno de 5) que é superior à de **Neutro**. No presente trabalho, no entanto, constatou-se que a curva de RMSD ( $C\alpha$ ) de **Neutro\_HisH** (pH em torno de 5) é superior à de **Protonado** (pH menor que 4) que é similar à de **Neutro** (pH 7). No que diz respeito às curvas de RMSF ( $C\alpha$ ), essas apresentaram valores similares. Ou seja, o presente trabalho não conseguiu reproduzir a tendência de pH verificada nas simulações de Campos e colaboradores<sup>(75)</sup>.

O grupo de Langella e colaboradores<sup>(84)</sup>, por sua vez, realizou um estudo de influência de pH (empregando simulações de 10 ns com o fragmento 125-228 do príon humano) empregando estruturas com estados de protonação equivalentes ao que aqui denominamos de

**Neutro** (correspondente a um pH 7) e **Neutro\_HisH** (correspondente a um pH levemente ácido, em torno de 5). O grupo de Langella posteriormente desenvolveu também um estudo detalhado de manipulação dos estados de protonação dos resíduos histidina<sup>(83)</sup> (localizados nas posições 140, 155, 177 e 187 no fragmento protéico 125-228 do príon humano). Neste estudo foram realizadas 5 simulações de dinâmica molecular: PrPOH (todas histidinas neutras), PrP2H (histidinas 140 e 177 protonadas, demais neutras), PrP3H (histidina 187 neutra, demais protonadas), PrP3H\* (histidina 155 neutra, demais protonadas) e PrP4H (todas as histidinas protonadas). Todas essas simulações apresentaram diferentes comportamentos, inclusive possuindo diferentes curvas de RMSD (C $\alpha$ ). As simulações **Neutro** e **Neutro\_HisH**, do presente trabalho, correspondem, respectivamente, às simulações PrPOH e PrP4H.

Os resultados de ambos os estudos do grupo de Langella<sup>(83, 84)</sup> mostraram que **Neutro\_HisH** apresentava uma curva de RMSD (C $\alpha$ ) superior à de **Neutro**, indicando que a protonação dos resíduos histidina implica em maiores alterações da estrutura proteica. Além disso, foi constatado que **Neutro** apresentava pouca oscilação em relação à estrutura de partida, conservando bem as 3 hélices. Enquanto **Neutro\_HisH** desnaturou o final da hélice 2 (H2) através da formação de uma dobra (*bend*) nessa região e apresentou um aumento de resíduos da fita- $\beta$  na região N-terminal no intervalo de 4-6 ns de simulação. No presente trabalho, **Neutro\_HisH** também apresentou uma curva de RMSD ( $\alpha$ ) superior à de **Neutro**, assim como o surgimento de uma dobra (*bend*) no lugar dos resíduos finais da hélice 2 (a partir de 9 ns de simulação). No entanto, **Neutro** apresentou ao longo de toda a simulação um conteúdo de fitas- $\beta$  na região N-terminal superior ao de **Neutro\_HisH**. Por outro lado, **Neutro\_HisH** apresentou uma diminuição mais acentuada de número de resíduos  $\alpha$ -helicoidais.

Uma das alterações estruturais importantes de se destacar na simulação de **Neutro\_HisH** é a formação de uma dobra (*bend*) no final da  $\alpha$ -hélice 2 (H2). A justificativa que Langella<sup>(83)</sup> e colaboradores dão para essa alteração seria a formação de uma ligação salina entre a histidina 187 e o ácido glutâmico 196. De fato, ambos os resíduos estão presentes na mesma posição da sequência do fragmento protéico do hamster Sírio empregado no presente trabalho. Desta forma, a formação desta ligação salina parece ser o motivo da desestruturação da  $\alpha$ -hélice observada em **Neutro\_HisH**. Para confirmar isso, é necessário realizar uma análise de distância deste par de resíduos em função do tempo de simulação.

Os resultados do presente trabalho, por outro lado, mostram que **Neutro\_HisH** não é a única simulação que apresentou desestruturação da  $\alpha$ -hélice 2 (H2) devido à formação de uma

dobra (*bend*) em tal região. Na verdade, a mesma alteração é observada em **Neutro\_HisA**, **Neutro\_HisB**, **Neutro\_HisA\_HisB** e **Neutro\_HisB\_HisA** (inclusive em **Protonado**). Isso indica que diferentes combinações de formas neutras da histidina também proporcionam o mesmo efeito que a protonação do resíduo histidina na desestruturação desta  $\alpha$ -hélice (H2). De fato, **Neutro** foi a única simulação que não apresentou a formação de uma dobra (*bend*) no lugar dos resíduos finais da  $\alpha$ -hélice 2 (H2).

Feitas essas considerações, retoma-se o caso da escolha da melhor simulação para representar o pH neutro. Da literatura, conhece-se os diferentes tautômeros ( $\delta$  e  $\epsilon$ ) que são formas de representação da histidina neutra. A protonação padrão do GROMACS que resultou na simulação **Neutro** (resíduos 111, 140 e 187 = HISB; resíduo 177 = HISA) corresponde a representação preferencial que os resíduos histidinas assumem na forma neutra. O tautômero  $\epsilon$  (HISB) é preferencialmente selecionado para representar todos os resíduos histidina da proteína príon celular do hamster sírio (com exceção do resíduo 177, que é preferencialmente representado pelo tautômero  $\delta$  ou HISA). Desta forma, embora **Neutro\_HisB** tenha apresentado resultados mais condizentes com os modelos de influência de pH (ou seja, menor conteúdo de folhas- $\beta$  e maior conteúdo de  $\alpha$ -hélices, além de uma curva de RMSD inferior à de **Protonado**), **Neutro** oferece a representação dos resíduos histidina na forma neutra mais condizente com o que se observa na literatura.

### 5.2.2 Conclusões Parciais

Os resultados deste capítulo indicam que a alteração dos resíduos histidina mostrou ser capaz de produzir substancial diferença no comportamento estrutural dos fragmentos protéicos simulados. Desta forma, a manipulação dos resíduos histidina é um ponto crucial que deve ser considerado nos estudos de influência de pH. De fato, o estado de protonação destes resíduos parece estar intrinsecamente associado com a transição conformacional da proteína príon celular, como já havia sido relatado na literatura.

Por fim, embora **Neutro\_HisB** tenha sido escolhido, inicialmente, como o melhor candidato para representar a simulação em pH neutro, essa estrutura ainda está distante dos moldes do estudo de influência de pH. Além disso, é conhecido da literatura que a forma neutra mais provável de representação dos quatro resíduos histidina, presentes no fragmento protéico de estudo, é aquela que condiz com a da simulação que aqui denomina-se **Neutro**. Esta última, no entanto, não apresentou resultados coerentes com os modelos de influência de pH. Por esse motivo, estender esta trajetória é um compromisso que deve ser cumprido antes



de se tirar conclusões mais rigorosas a cerca da relação entre as estruturas em pH neutro e ácido. Outra possibilidade para contornar esta problemática é manipular as condições iniciais das simulações, empregando diferentes estruturas de partida (obtidas do RMN) e randomizando as velocidades iniciais. Desta forma, será possível verificar a reprodutibilidade dos resultados obtidos.

### 5.3 INFLUÊNCIA DE DIFERENTES CONDIÇÕES INICIAIS

Para verificar a reprodutibilidade dos resultados, optou-se por realizar estudos de alteração das condições iniciais das simulações de influência do pH. Com isso, tem-se a intenção de excluir a possibilidade de as discrepâncias observadas entre as diferentes simulações estarem associadas à existência de flutuações estruturais maiores que o erro sistemático implícito das simulações.

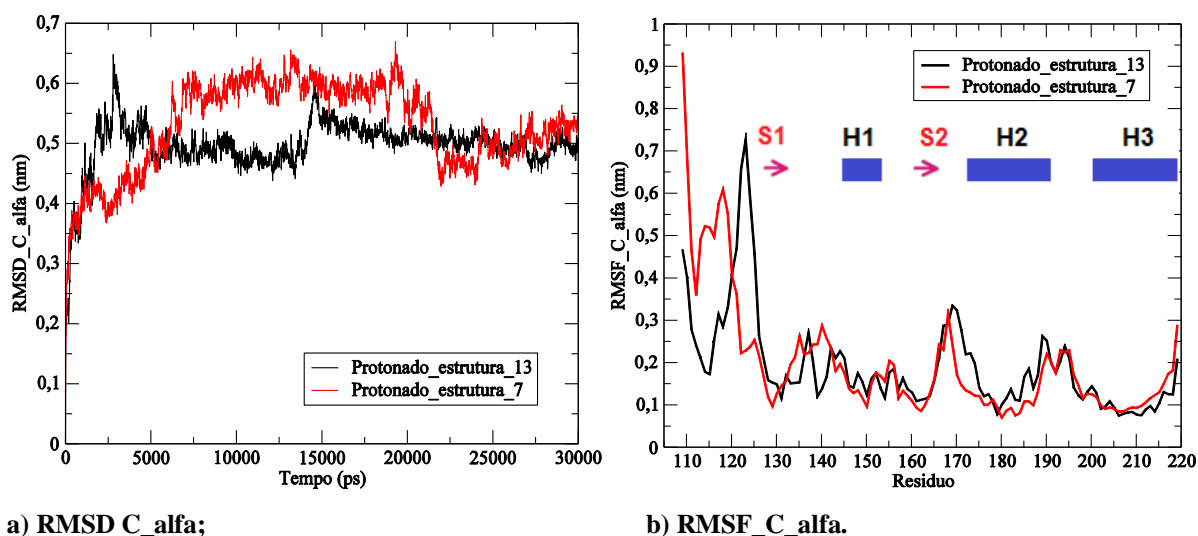
A trajetória de uma simulação de dinâmica molecular possui um erro intrinsecamente associado, assim como cada uma das análises. O erro associado a uma trajetória pode ser estimado analisando-se o intervalo de convergência de uma curva de RMSD em função do tempo de simulação. Assim, estabelece-se um valor médio que representa a conformação média ao redor do qual se tem flutuações da estrutura global da macromolécula sob estudo. O erro na trajetória também pode ser estimado analisando-se diferentes blocos ao longo da trajetória (dentro de um intervalo de convergência). Para estimar o erro em uma análise (como é o caso das análises Scout, etc.) é preciso analisá-la também em um intervalo de convergência. Desta forma, pode-se estabelecer um valor médio ao redor do qual a análise flutua. Além disso, também pode ser estimado o erro dentro de um conjunto de simulações, realizando-se réplicas (ou tréplicas, etc.) do sistema simulado, de forma que toma-se um valor médio para a propriedade analisada no conjunto replicado e estabelece-se um faixa ao redor da qual esse valor flutua.

No presente trabalho as alterações estruturais associadas ao uso de diferentes condições iniciais (empregando diferentes estruturas de RMN, ou seja diferentes conformações iniciais, e randomizando as velocidades iniciais) não serão avaliadas de um ponto de vista quantitativo, mas sim apenas qualitativo. Isso significa que aqui tem-se somente a intenção de avaliar possíveis distintos comportamentos estruturais nas diferentes simulações. Portanto, no presente trabalho não serão estimados os erros associados às análises e trajetórias.

#### 5.3.1 Utilização de diferentes conformações iniciais de RMN

Buscando verificar a influência de um conjunto de coordenadas iniciais diferentes, foram utilizados como estruturas de partida para as simulações de DM os modelos 13 (já apresentado pela denominação **Protonado** nas Seções 5.1 e 5.2) e 7 do RMN (PDB: 2PRP)<sup>(18)</sup>, que vamos aqui identificar, respectivamente, por **Protonado\_13** e **Protonado\_7**. O RMSD entre os modelos 7 e 13 (de 2PRP) é de: 0,311 nm, considerando somente os átomos

de carbono- $\alpha$  (111 átomos  $C_\alpha$ ); e de 0,403 nm, considerando todos os 891 átomos dos fragmentos protéicos. Ambas as estruturas tiveram seus resíduos Asp, Glu e His previamente protonados (gerando para os dois primeiros cadeias laterais neutras e para o último, cadeia lateral carregada positivamente) de forma a produzir sistemas em pH ácido. Estas estruturas foram, então, submetidas a dinâmicas moleculares de 30 ns cada empregando o campo de força Gromos 96 53 a6<sup>(62)</sup>.

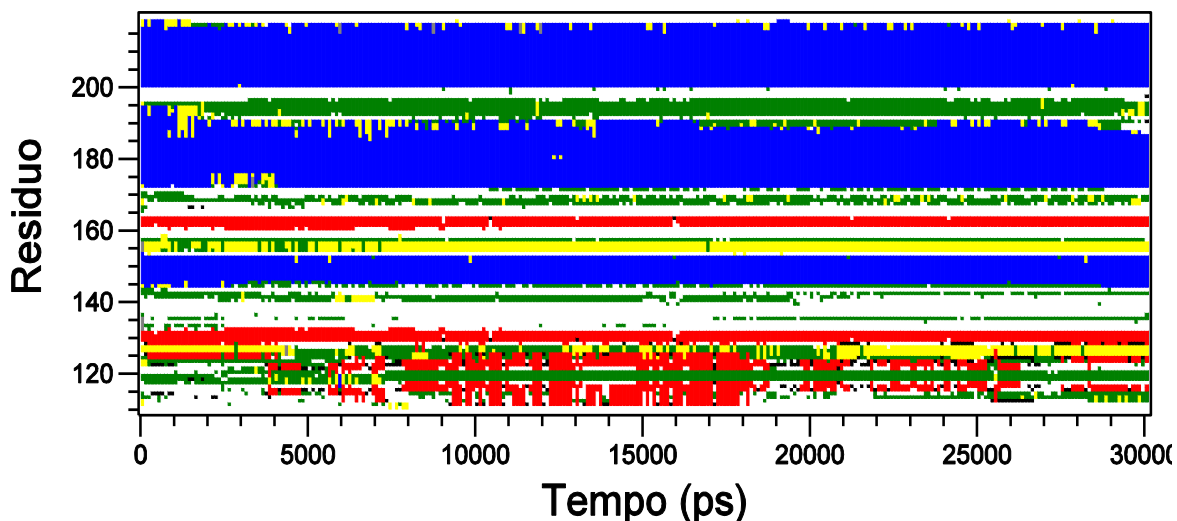


**Figura 33:** Gráficos de: a) RMSD ( $C_\alpha$ ) em função do tempo de simulação; b) RMSF ( $C_\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 30 ns de simulação). As curvas nos gráficos se referem às simulações de Protonado\_13 e Protonado\_7. As flechas rosas indicam a localização das fitas- $\beta$  originais (S1 e S2; presentes na estrutura inicial de 2PRP) e as caixas azuis indicam a localização das hélices (H1, H2 e H3) nas estruturas protéicas simuladas.

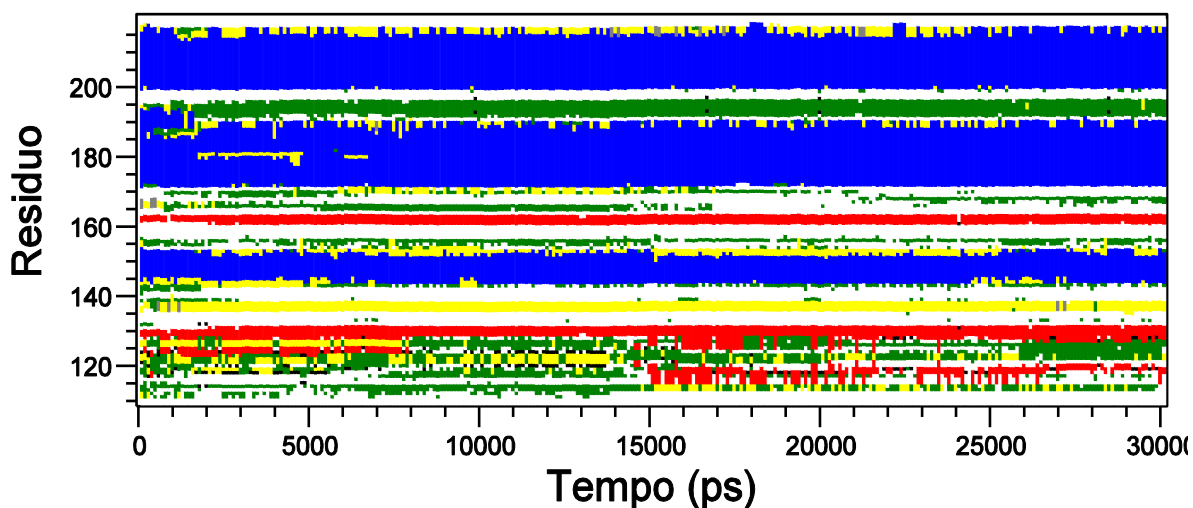
Analisando o gráfico de RMSD ( $C_\alpha$ ), Figura 33a, verifica-se que as curvas convergem para um valor próximo de 0,52 nm a partir de 22 ns. O gráfico de RMSF ( $C_\alpha$ ), Figura 33b, apresenta boa concordância entre as curvas no intervalo de 145-219 aa. No entanto, ocorre uma discrepância significativa entre as curvas para a região 109-145, indicando mobilidades (estimadas sobre todo o tempo de simulação) distintas dos resíduos N-terminais. Esse resultado concorda com os diferentes conteúdos de folhas- $\beta$  presente na região N-terminal ao longo das duas simulações (conforme pode ser visto pela Figura 34).

Avaliando em conjunto os gráficos de estrutura secundária das Figuras 34 e 35, verifica-se que ambas as simulações diminuem o conteúdo  $\alpha$ -helicoidal, principalmente pela desestruturação dos resíduos finais das  $\alpha$ -hélices 2 (H2). Isso ocorre devido a formação de

uma dobra (ou *bend*) em ambas as simulações. Observa-se também, pela Figura 35, que a curva de estrutura secundária folhas- $\beta$  de **Protonado\_7** atinge um pico em torno de 17 ns, mas acaba regredindo nos 13 ns posteriores, terminando a simulação com um valor similar ao que apresenta a curva de **Protonado\_13**.

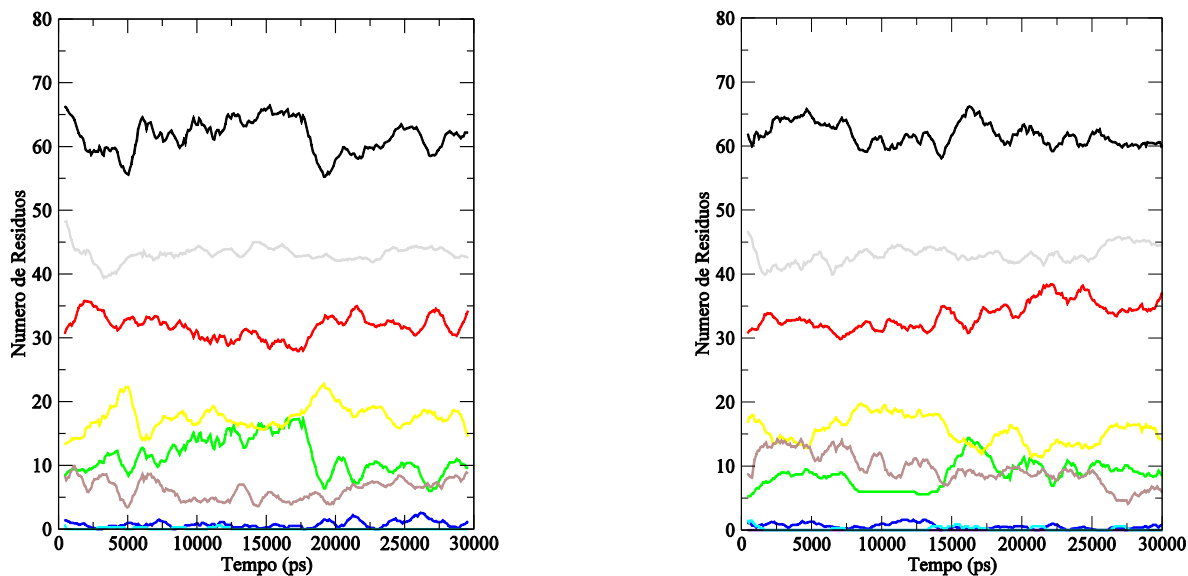


a) Protonado\_7;



b) Protonado\_13.

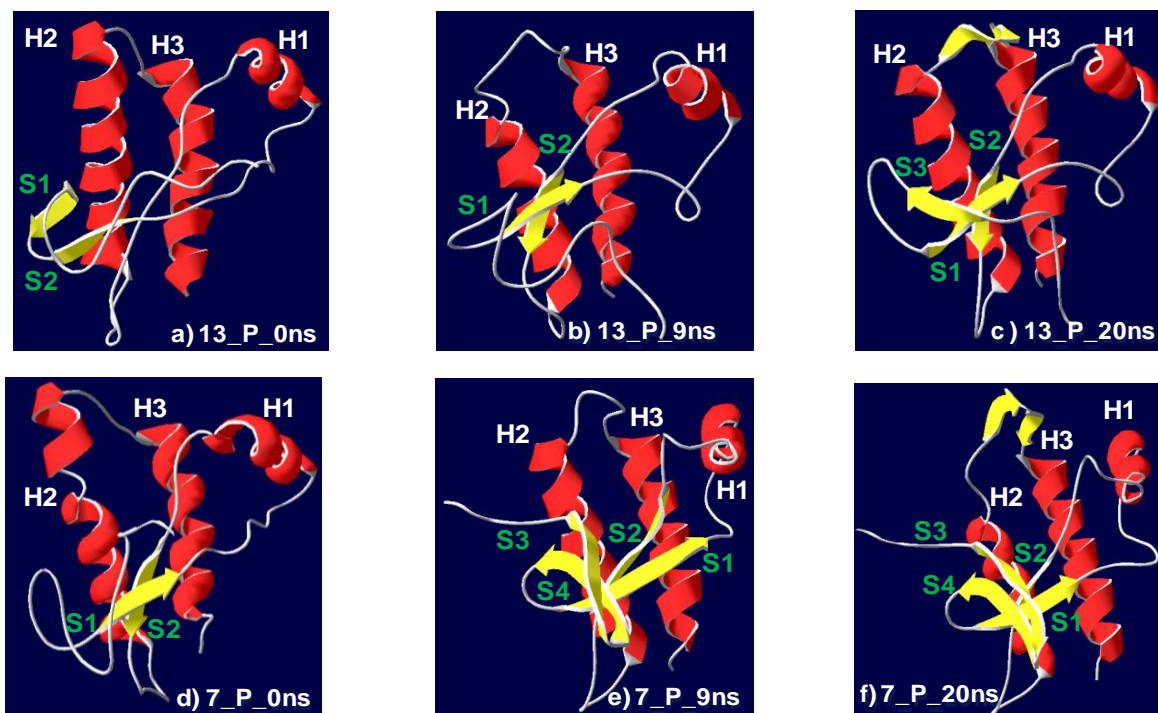
**Figura 34:** Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para: **a)** Protonado\_13; **b)** Protonado\_7. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber:  $\square$  alça (coil),  $\blacksquare$  folha- $\beta$  ( $\beta$ -sheet),  $\blacksquare$  segmento  $\beta$  ( $\beta$ -bridge),  $\blacksquare$  dobra (bend),  $\blacksquare$  volta (turn),  $\blacksquare$   $\alpha$ -hélice (A-helix),  $\blacksquare$  5-hélice (5-helix) e  $\blacksquare$  3-hélice (3-helix).



a) Protonado\_7;

b) Protonado\_13.

**Figura 35:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação para: a) Protonado\_7; b) Protonado\_13. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: ■ *Structure*, ■ alça (coil), ■ folha-β (β-sheet), ■ segmento β (β-bridge), ■ dobra (bend), ■ volta (turn), ■ α-hélice (A-helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).



**Figura 36:** Imagens correspondentes a determinados instantes de tempo para as simulações (criadas com o programa SPDBV) de Protonado\_13 em: a)  $t = 0$  ns; b)  $t = 9$  ns; c)  $t = 20$  ns; e de Protonado\_7 em: d)  $t = 0$  ns; e)  $t = 9$  ns; f)  $t = 20$  ns.

A Figura 36 apresenta imagens correspondentes aos instantes de tempo de 0, 9 e 20 ns para as simulações de **Protonado\_7** e **Protonado\_13**. Estes instantes de tempo foram escolhidos por serem representativos da evolução das estruturas simuladas. As imagens correspondentes a 9 ns enfatizam a significativa diferença de conteúdo de folhas- $\beta$  na região N-terminal nas simulações consideradas. Esta diferença é perceptivelmente diminuída nas imagens correspondentes a 20 ns das respectivas simulações. Desta forma, as imagens da Figura 36 corroboram o que havia sido discutido anteriormente nas análises do presente capítulo.

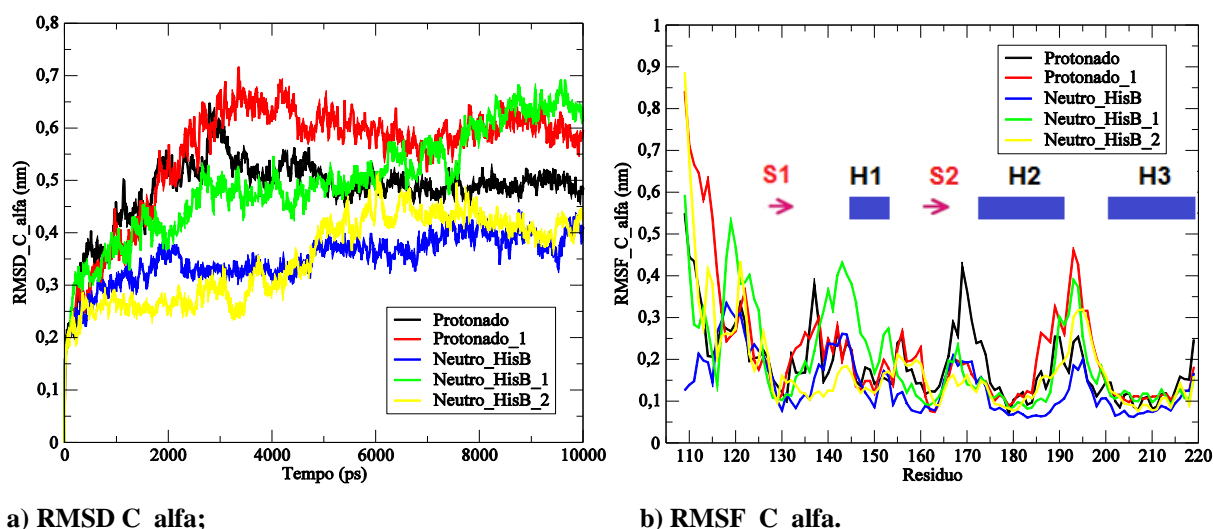
### 5.3.2 Randomização das velocidades iniciais

Nesta etapa, buscou-se estudar a influência da randomização das velocidades iniciais. Para isto, as simulações **Neutro\_HisB** (escolhida por ter se mostrado numa primeira análise como o melhor candidato à representar a simulação em pH neutro) e **Protonado** (modelo 13 do RMN; simulação em pH baixo já apresentada nas seções anteriores) foram randomizadas através da alteração do valor da semente do gerador de velocidades (*gen\_seed*, definida em um arquivo \*.mdp que contém os parâmetros da simulação). A Tabela III, a seguir, apresenta o valor dessa semente de gerador de velocidades empregada em cada uma das simulações apresentadas neste capítulo.

**Tabela III:** Simulações e seus respectivos valores de semente de gerador de velocidades iniciais.

Simulação	Gen_seed
<b>Neutro_HisB</b>	173529
<b>Neutro_HisB_1</b>	116247
<b>Neutro_HisB_2</b>	231100
<b>Protonado</b>	173529
<b>Protonad_1</b>	116247

As simulações randomizadas em pH neutro (**Neutro\_HisB\_1** e **Neutro\_HisB\_2**) e ácido (**Protonado\_1**) mantiveram os estados de protonação dos resíduos de aminoácidos equivalentes, respectivamente, aos de **Neutro\_HisB** e **Protonado** (já apresentados nas seções anteriores do presente trabalho). As dinâmicas moleculares tiveram duração de 10 ns cada e empregaram o campo de força GROMOS96 53a6<sup>(62)</sup>.

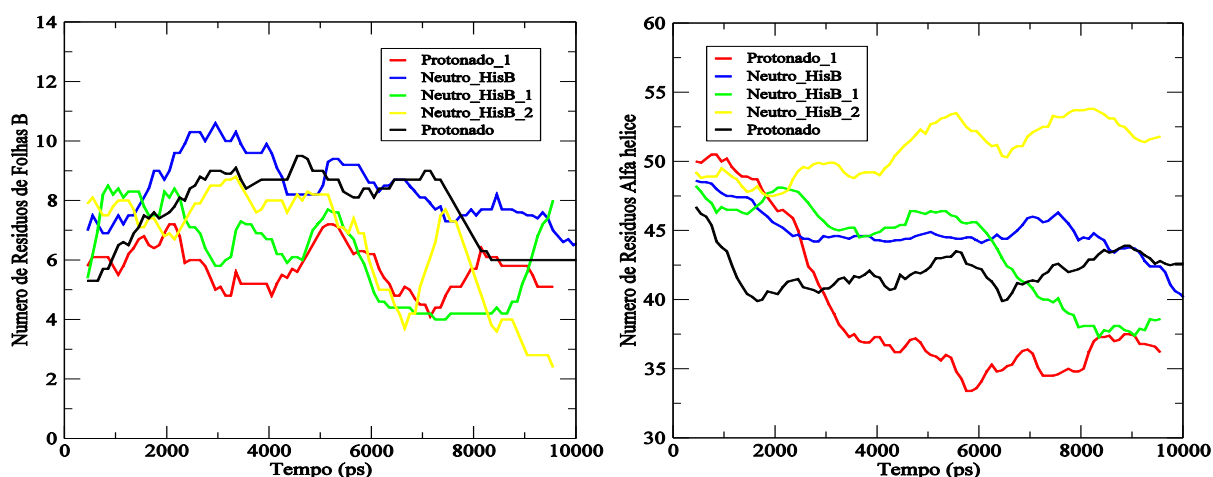


a) RMSD C\_alfa;

b) RMSF\_C\_alfa.

**Figura 37:** Gráficos de: **a)** RMSD ( $C\alpha$ ) em função do tempo de simulação; **b)** RMSF ( $C\alpha$ ) em função do resíduo de aminoácido (obtido sobre os 10 ns de cada simulação). As curvas nos gráficos referem-se às simulações de Neutro\_HisB, Neutro\_HisB\_1, Neutro\_HisB\_2, Protonado e Protonado\_1. No gráfico de RMSF ( $C\alpha$ ), as flechas rosas indicam a localização das fitas- $\beta$  originais (S1 e S2; presentes na estrutura original de 2PRP) e as caixas azuis indicam a localização das hélices (H1, H2 e H3) nas estruturas protéicas simuladas.

Analisando os gráficos de RMSD ( $C\alpha$ ) e RMSF ( $C\alpha$ ), Figura 37, são observadas significativas diferenças entre as curvas em pH neutro e entre as curvas em pH baixo. No que diz respeito, especificamente, às curvas de RMSD ( $C\alpha$ ), observa-se que **Neutro\_HisB** e **Neutro\_HisB\_2** convergiram nos últimos 2 ns de simulação a um valor próximo de 0,4 nm. A curva de **Neutro\_HisB\_1**, no entanto, apresentou um comportamento mais similar à curva de **Protonado\_1**, tendo ambas as curvas convergido a um valor próximo de 0,6 nm nos últimos 2 ns de simulação. A curva de **Protonado**, por sua vez, convergiu a um valor de 0,45 nm (a partir de 5 ns de simulação), apresentando um valor bastante distinto de **Protonado\_1**.



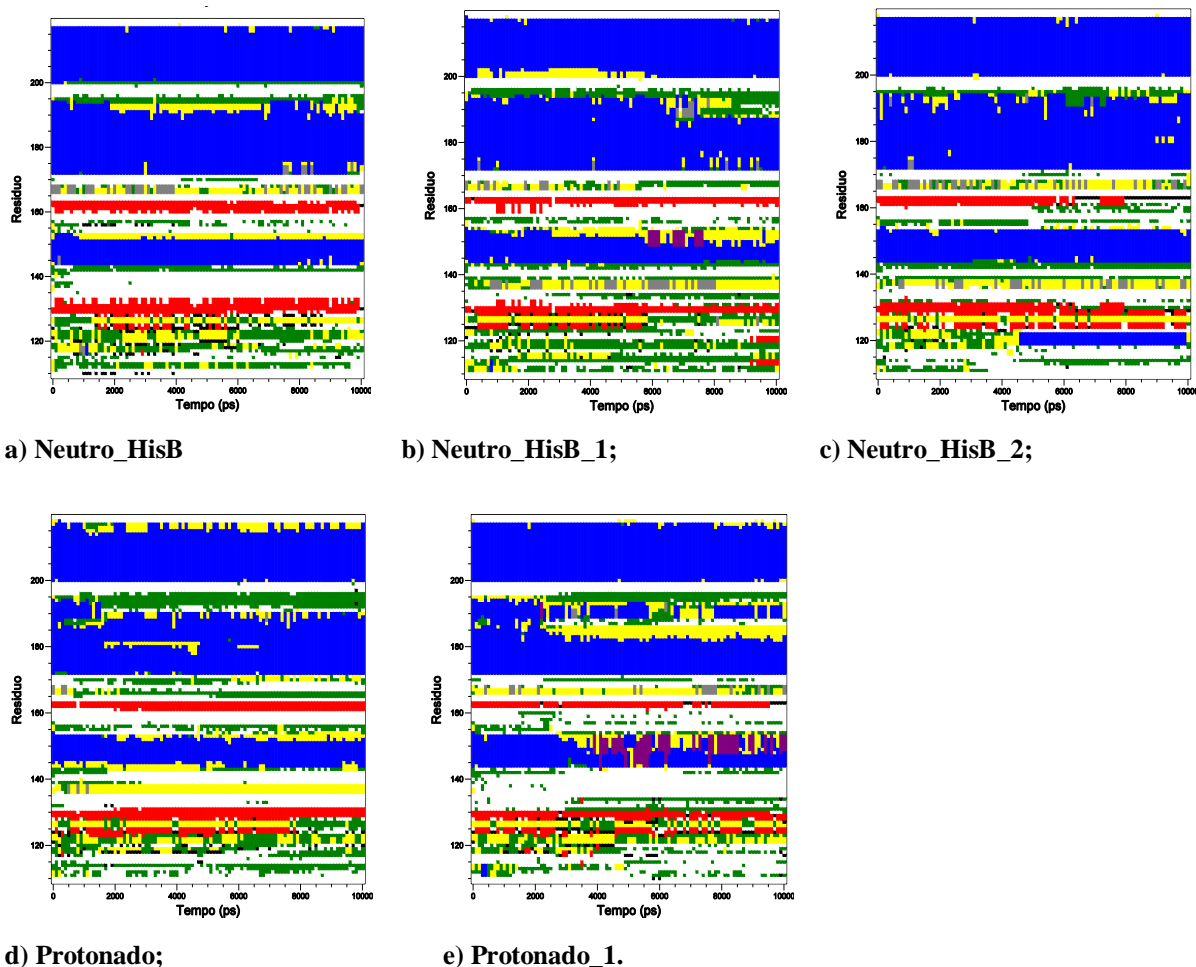
a) Folha- $\beta$ ;

b)  $\alpha$ -hélice.

**Figura 38:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: a) folha- $\beta$ ; b)  $\alpha$ -hélice. As curvas nos gráficos referem-se às simulações de Neutro\_HisB, Neutro\_HisB\_1, Neutro\_HisB\_2, Protonado e Protonado\_1.

Observando os gráficos de estrutura secundária, Figuras 38 e 39, verifica-se que as simulações em pH neutro apresentaram diferentes conteúdos de folhas- $\beta$  na região N-terminal. No que diz respeito ao conteúdo  $\alpha$ -helicoidal as simulações também mostraram diferenças, sendo que **Neutro\_HisB\_2** surpreendentemente exibiu a formação de uma nova  $\alpha$ -hélice na região N-terminal a partir de 4,5 ns de simulação. As diferentes simulações em pH neutro apresentaram diferentes níveis de desestruturação das  $\alpha$ -hélices 1 e 2 (H1 e H2). Para as estruturas em pH ácido, percebe-se que **Protonado\_1**, ao término da simulação, apresentou menor número de resíduos de folhas- $\beta$  e  $\alpha$ -hélice (esse último, em função de uma maior desestruturação das  $\alpha$ -hélices 1 e 2) em comparação à **Protonado**.





**Figura 39:** Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP) para: **a)** Neutro\_HisB; **b)** Neutro\_HisB\_1; **c)** Neutro\_HisB\_2; **d)** Protonado; **e)** Protonado\_1. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: □ alça (coil), ■ folha- $\beta$  ( $\beta$ -sheet), ■ segmento  $\beta$  ( $\beta$ -bridge), ■ dobra (bend), ■ volta (turn), ■  $\alpha$ -hélice (A-helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).

### 5.3.3 Conclusões Parciais

O uso de diferentes conjuntos de coordenadas iniciais (diferentes modelos de RMN) apresentou discrepâncias estruturais. No entanto, estas diferenças foram amenizadas nos últimos 10 ns das simulações. Isto mostra que o prolongamento destas simulações parece ter compensado o emprego de diferentes conformações iniciais. Por outro lado, a randomização das velocidades iniciais produziu diferenças estruturais mais significativas, como o surpreendente surgimento de uma nova  $\alpha$ -hélice na região N-terminal em **Neutro\_HisB\_2**,

diferentes níveis de desestruturação da  $\alpha$ -hélices 1 e 2 (H1 e H2) entre as estruturas neutras e a desestruturação mais acentuada das  $\alpha$ -hélices 1 e 2 em **Protonado\_1**. Isto indica que a randomização pode ter promovido a exploração de bacias conformacionais levemente diferentes, sobretudo separadas por barreiras energéticas que dificilmente seriam transpostas pelo simples prolongamento das simulações. Para corroborar estes resultados é, no entanto, necessário estender as simulações.

Estes resultados mostram que a realização de múltiplas simulações de dinâmica molecular, através da alteração das condições iniciais, é de grande importância nos estudos de influência de pH no processo de conversão estrutural da proteína príon celular. Sobretudo, os resultados indicam que não se deve tirar conclusões muito rigorosas pela análise de uma única simulação em um determinado pH. Desta forma, investir em condições iniciais distintas é um compromisso que deve ser seriamente considerado nestes estudos de influência de pH na alteração estrutural da proteína príon celular.

## 5.4 INVESTIGAÇÃO DAS INTERAÇÕES ELETROSTÁTICAS DE LONGO ALCANCE

### 5.4.1 SPME x GRF

Nos capítulos anteriores foi constatado que a simulação em pH neutro (denominada de **Neutro** na Seção 5.2) apresentou maior conteúdo de folhas- $\beta$  do que a simulação em pH ácido (denominada de **Protonado** na Seção 5.2). Esse resultado é, de fato, contrário ao que se tem descrito na literatura<sup>(33, 75, 84)</sup>. Em função disso, optou-se por estender a trajetória da simulação em pH neutro para verificar se a relação entre os comportamentos estruturais das simulações nos diferentes pH seria mantida.

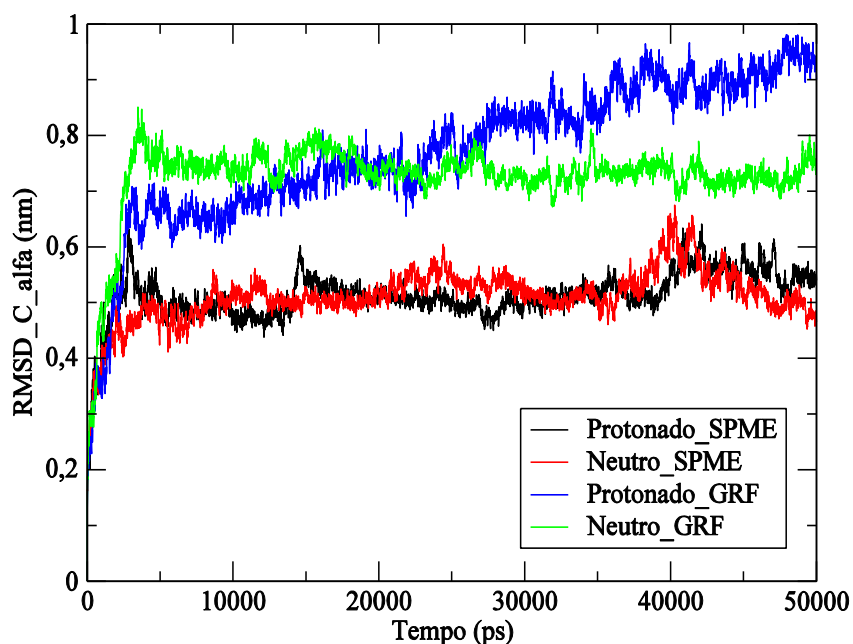
Além disso, optou-se também por realizar as simulações em pH neutro e ácido empregando um outro método para o cálculo das interações eletrostáticas de longo alcance: o método GRF (*Generalized Reaction Field*)<sup>(70)</sup>. De fato, é conhecido na literatura o fato de que o método SPME<sup>(71)</sup> (empregado em todas as simulações dos capítulos anteriores) estabiliza mais as estruturas simuladas do que o método GRF. Isso significa que o método GRF permitiria (segundo a literatura) visualizar maiores alterações estruturais na proteína prion celular, o que é justamente o que se procura neste trabalho.

Desta forma, a presente seção apresenta uma comparação das simulações realizadas em pH neutro e ácido empregando os métodos SPME e GRF para o cálculo das interações eletrostáticas de longo alcance. Os sistemas denominados "Protonados" tiveram seus resíduos Asp, Glu e His protonados. Enquanto nos sistemas denominados de "Neutros" os resíduos His foram mantidos neutros (resíduos 111, 140, 187 = HisB; resíduo 177 = HisA) e os resíduos Asp e Glu foram desprotonados. As simulações aqui discutidas empregaram o campo de força GROMOS 96 53a6<sup>(62)</sup> e foram estendidas por 50 ns. A Tabela IV, a seguir, apresenta os números de íons e moléculas de um cada um dos sistemas estudados.

**Tabela IV:** Número de íons e moléculas de cada sistema simulado.

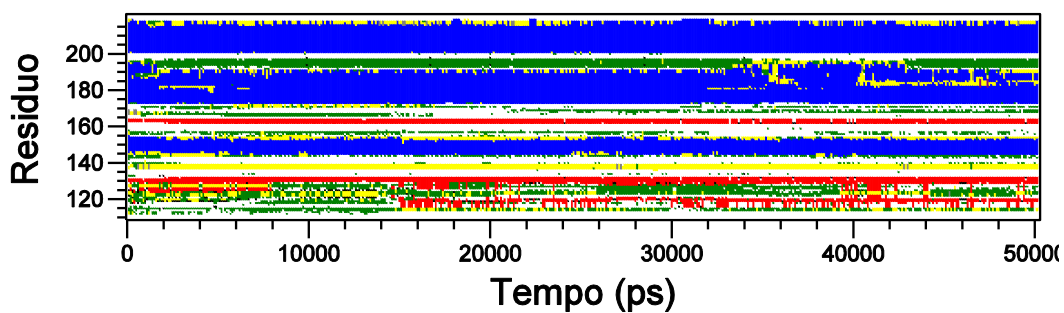
Simulação	Solvente (H <sub>2</sub> O)	Proteína	Na <sup>+</sup>	Cl <sup>-</sup>
<b>Protonado_SPME</b>	10892	1	30	44
<b>Neutro_SPME</b>	13161	1	40	39
<b>Protonado_GRF</b>	13163	1	40	54
<b>Neutro_GRF</b>	13161	1	40	39

Analisando o gráfico de RMSD ( $C\alpha$ ), Figura 40, observa-se claramente que o método GRF permite maiores alterações estruturais do que o método SPME. Além disso, o método GRF também conseguiu nitidamente distinguir os comportamentos dos diferentes pH de estudo. A curva de **Neutro\_GRF** convergiu a um valor de 0,75 nm já a partir de 5 ns de simulação, enquanto a curva de **Protonado\_GRF** apresentou um comportamento ascendente em todo o intervalo de tempo de simulação. A curva de **Protonado\_GRF** ultrapassou a curva de **Neutro\_GRF** em torno de 20 ns e terminou a simulação com um valor de RMSD ( $C\alpha$ ) em torno de 0,95 nm. O método SPME, por outro lado, apresenta um conteúdo de alteração estrutural similar para os diferentes pH, mesmo após o prolongamento das simulações. Assim, ao final das simulações **Neutro\_SPME** e **Protonado\_SPME** apresentam valores de RMSD ( $C\alpha$ ) próximos de 0,5 nm (sendo o valor de RMSD de **Protonado\_SPME** levemente superior ao de **Neutro\_SPME**).

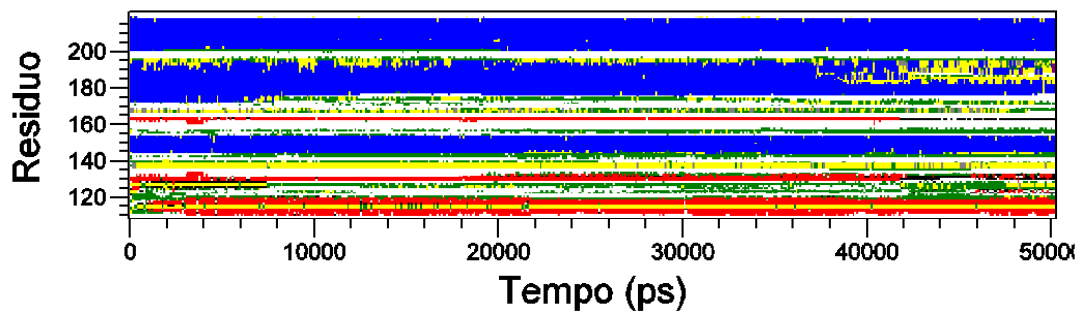


**Figura 40:** Gráfico de RMSD ( $C\alpha$ ) em função do tempo para as simulações de Protonado\_SPME, Neutro\_SPME, Protonado\_GRF e Neutro\_GRF.

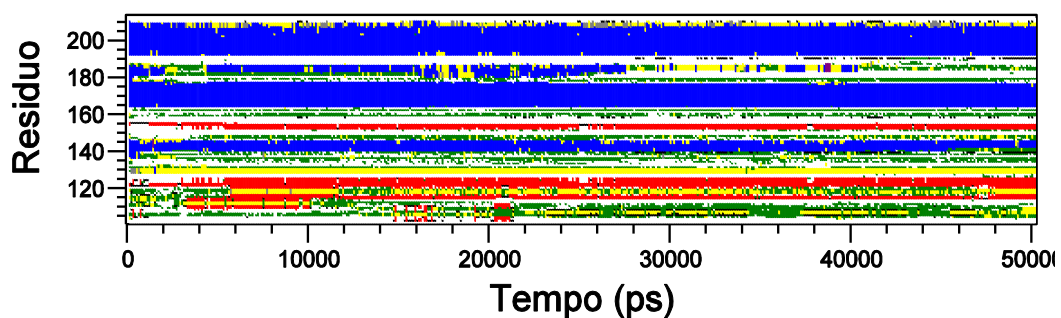
Avaliando em conjunto as análises de estrutura secundária das Figuras 41 e 42, observa-se que **Neutro\_SPME** promove uma maior desestruturação das  $\alpha$ -hélices do que **Protonado\_SPME**. O primeiro termina a simulação com cerca de 35 resíduos  $\alpha$ -helicoidais, enquanto o segundo, com cerca de 40 resíduos. De fato, a maior desestruturação dos resíduos  $\alpha$ -helicoidais de **Neutro\_SPME** ocorre nos últimos 8 ns de simulação, devido a uma maior desestruturação da  $\alpha$ -hélice 2 (H2). Além disso, **Neutro\_SPME** e **Protonado\_SPME** seguem



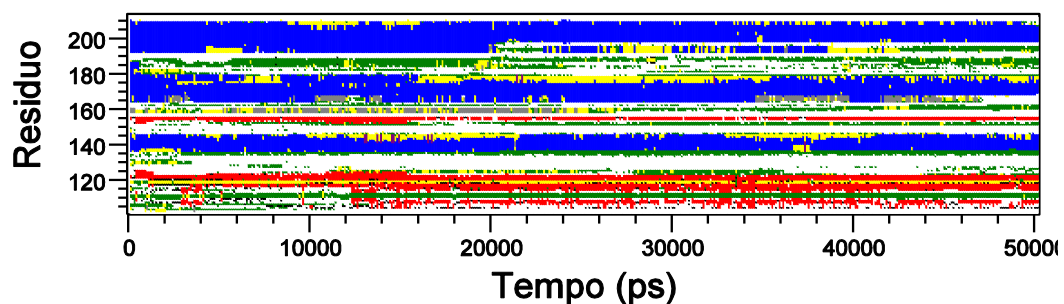
a) Protonado\_SPME;



b) Neutro\_SPME.

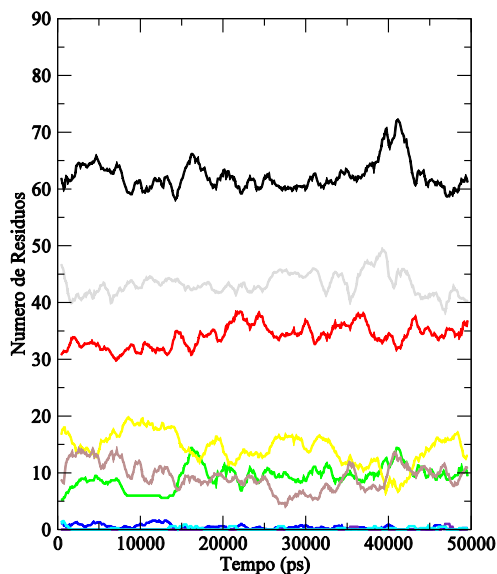


c) Protonado\_GRF;

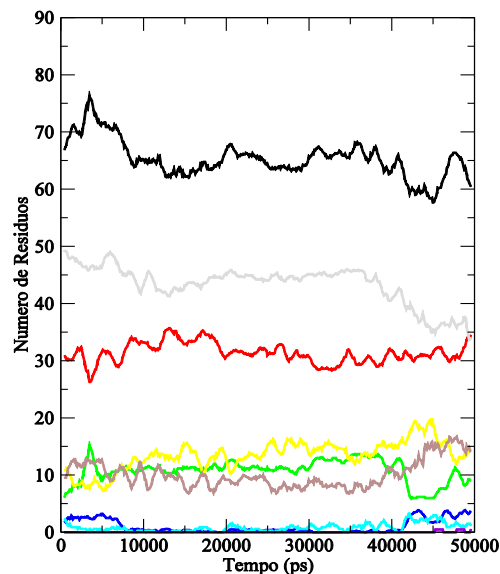


d) Neutro\_GRF.

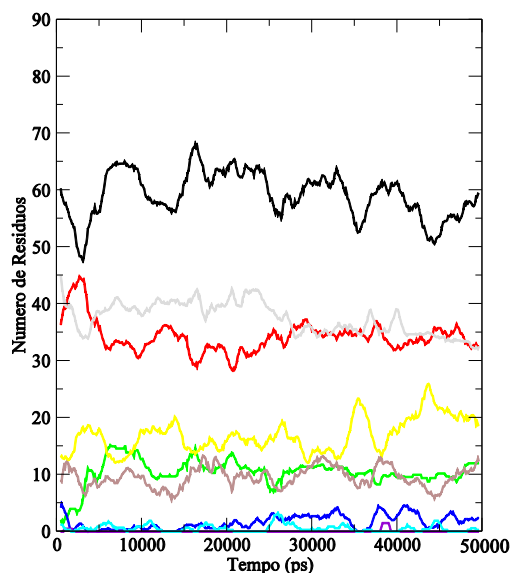
**Figura 41:** Gráficos de estrutura secundária dos resíduos de aminoácidos em função do tempo de simulação (criados empregando o programa DSSP): **a)** Protonado\_SPME; **b)** Neutro\_SPME; **c)** Protonado\_GRF; **d)** Neutro\_GRF. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: □ alça (coil), ■ folha- $\beta$  ( $\beta$ -sheet), ■ segmento  $\beta$  ( $\beta$ -bridge), ■ dobra (bend), ■ volta (turn), ■  $\alpha$ -hélice ( $\alpha$ -helix), ■ 5-hélice (5-helix) e ■ 3-hélice (3-helix).



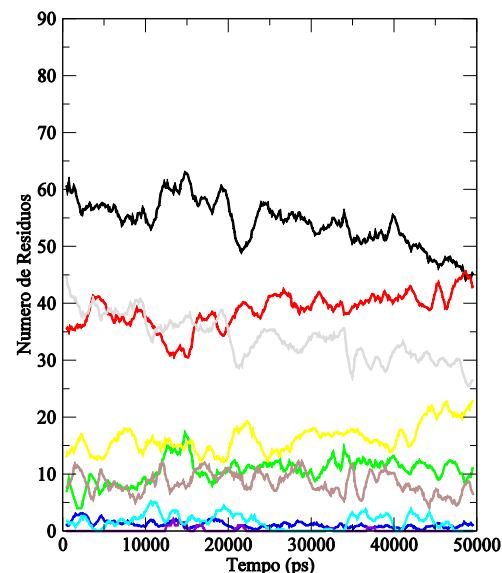
a) Protonado\_SPME;



b) Neutro\_SPME;



c) Protonado\_GRF;



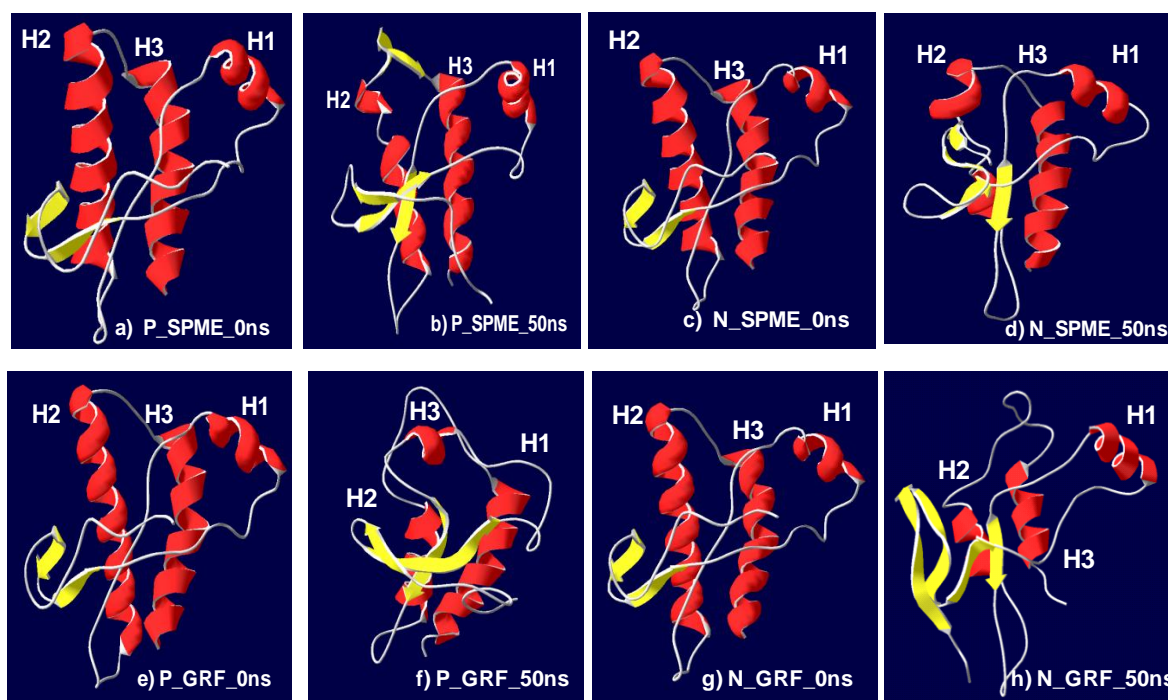
d) Neutro\_GRF.

**Figura 42:** Gráficos de número de resíduos envolvidos em tipos de estrutura secundária em função do tempo de simulação: **a)** Protonado\_SPME; **b)** Neutro\_SPME; **c)** Protonado\_GRF; **d)** Neutro\_GRF. Para cada cor nos gráficos corresponde uma estrutura secundária, a saber: **■** Structure, **■** alça (coil), **■** folha- $\beta$  ( $\beta$ -sheet), **■** segmento  $\beta$  ( $\beta$ -bridge), **■** dobra (bend), **■** volta (turn), **■**  $\alpha$ -hélice (A-helix), **■** 5-hélice (5-helix) e **■** 3-hélice (3-helix).

com um conteúdo similar de folhas- $\beta$  (em torno de 10 resíduos) até que nos 8 ns finais da simulação **Neutro\_SPME** tem seu número de resíduos diminuído. Isso se deve a desestruturação das fitas- $\beta$  (S1 e S2) que ocorre nos últimos 8 ns de simulação para **Neutro\_SPME**.

No que diz respeito às simulações com GRF, observa-se que **Neutro\_GRF** apresentou uma maior desestruturação de  $\alpha$ -hélices do que **Protonado\_GRF** (ver Figuras 41 e 42). O primeiro acabou a simulação com cerca de 25 resíduos  $\alpha$ -helicoidais, enquanto o segundo, com cerca de 33 resíduos. Isso se deve a uma maior desestruturação das  $\alpha$ -hélices 2 e 3 (H2 e H3) que ocorre a partir de 20 ns na simulação de **Neutro\_GRF**. Além disso, **Neutro\_GRF** apresentou cerca de 1 ou 2 resíduos de folhas- $\beta$  a mais do que **Protonado\_GRF**, destacadamente nos últimos 10 ns de simulação.

Comparativamente, os dois métodos de cálculo de interações eletrostáticas de longo alcance (GRF e SPME) apresentaram simulações em pH neutro com maior desestruturação de resíduos  $\alpha$ -helicoidais do que as simulações em pH ácido. Sendo que a simulação **Neutro\_GRF** foi, dentre todas as quatro simulações, a que mais desestruturou as  $\alpha$ -hélices.



**Figura 43:** Imagens correspondentes a determinados instantes de tempo das simulações (criadas com o programa SPDBV) para SPME: **a)** Protonado em  $t = 0$  ns; **b)** Protonado em  $t = 50$  ns; **c)** Neutro em  $t = 0$  ns; **d)** Neutro em  $t = 50$  ns; e GRF: **e)** Protonado em  $t = 0$  ns; **f)** Protonado  $t = 50$  ns; **g)** Neutro em  $t = 0$  ns; **h)** Neutro em  $t = 50$  ns.

As imagens da Figura 43 mostram visualmente a maior desestruturação das  $\alpha$ -hélices promovida pelo uso do método GRF, corroborando o que foi anteriormente discutido nas análises do presente capítulo.

## 5.4.2 Conclusões Parciais

Os resultados deste capítulo apontam que a simulação em pH neutro resultou em perda mais acentuada de conteúdo  $\alpha$ -helicoidal do que a simulação em pH ácido, independente do método para o cálculo das interações eletrostáticas de longo alcance utilizado. Verificou-se, ainda, que o método GRF propiciou alteração estrutural substancialmente maior dos fragmentos protéicos simulados em ambos os pH, em relação ao método SPME. Essa alteração mais acentuada deve-se, em grande parte, a uma maior desestruturação das  $\alpha$ -hélices. Além disso, o método GRF também promoveu um aumento de folhas- $\beta$ . Desta forma, o método GRF mostrou ser o mais adequado para realizar estudos de conversão estrutural da proteína príon celular. Por fim, é importante destacar que **Neutro\_GRF** promoveu uma diminuição mais acentuada de conteúdo  $\alpha$ -helicoidal e um maior acréscimo (ainda que de apenas 1 ou 2 resíduos) de folhas- $\beta$  do que **Protonado\_GRF**, contraindo os resultados de estudos de influência de pH descritos na literatura<sup>(33, 75, 84)</sup>.



## **6. CONCLUSÕES FINAIS**

O presente trabalho se propôs a investigar computacionalmente a influência do pH no processo de conversão estrutural da proteína príon celular do hamster Sírio. Para isso foram avaliados o uso de diferentes campos de força, a influência dos estados de protonação dos resíduos histidina, a influência de diferentes condições iniciais e o emprego de diferentes métodos para o cálculo das interações eletrostáticas de longo alcance.

Os resultados da Seção 5.1 indicam uma marcada diferença no conteúdo de estrutura secundária nos fragmentos protéicos simulados com os diferentes campos de força. De fato, os campos de força GROMOS e OPLS promoveram uma diminuição mais acentuada de resíduos  $\alpha$ -helicoidais. Enquanto GROMOS e AMBER aumentaram o conteúdo de folhas- $\beta$  do fragmento protéico sob estudo e, em função disso, tiveram suas trajetórias estendidas até 100 ns de simulação. Como resultado dessa extensão de trajetórias, o campo de força AMBER apresentou um maior conteúdo de folhas- $\beta$  (com a formação de um núcleo de folhas- $\beta$  estável na região N-terminal), assim como um maior conteúdo de  $\alpha$ -hélices em comparação ao campo de força GROMOS. Este último não foi capaz de manter a nova fita- $\beta$  (S3) até o término da simulação. Os resultados, portanto, assinalam uma significativa discrepância entre as diferentes simulações, ressaltando que o uso de diferentes campos de força influencia significativamente os estudos de alteração conformacional da proteína príon celular.

A Seção 5.2 apresentou indícios de que os resíduos de aminoácidos histidina presentes no fragmento protéico podem ter um papel importante no processo de conversão estrutural da proteína príon celular. Para compreender melhor os diferentes resultados obtidos, em função do uso de diferentes estados de protonação dos resíduos histidina, é necessário monitorar com quais resíduos de aminoácido essas histidinas estão interagindo ao longo das simulações.

Na Seção 5.3 foi realizado um estudo qualitativo utilizando diferentes condições iniciais. Os resultados indicam que o uso das diferentes estruturas de RMN (os modelos 7 e 13 de 2PRP), como ponto de partida para a simulação de dinâmica molecular, provocou diferenças de conteúdo de estrutura secundária (principalmente na região N-terminal) que foram compensadas com o prolongamento das simulações. Por outro lado, a randomização das velocidades iniciais provocou diferenças de estrutura secundária mais significativas. Isso ressalta a importância de se investir seriamente em múltiplas simulações de dinâmica molecular (através da manipulação das condições iniciais) na investigação de mudanças estruturais da proteína príon celular induzidas por alteração de pH. Por fim, a Seção 5.4 mostrou que o método GRF propiciou alteração estrutural substancialmente maior dos fragmentos protéicos simulados em ambos os pH, em relação ao método SPME.

De maneira geral, a definição do protocolo de simulação parece influenciar dramaticamente os resultados dos estudos de influência de pH no processo de conversão estrutural da proteína príon celular. Por isso, um cuidado especial deve ser tomado ao se escolher as condições dessas simulações. Além disso, a tendência de pH verificada nos estudos experimentais e computacionais<sup>(33, 75, 84)</sup> descritos na literatura não foram adequadamente reproduzidos no presente trabalho. Isso, no entanto, não invalida os resultados apresentados na literatura. De fato, essas diferenças podem indicar a necessidade de se investir mais em estudos de múltiplas simulações de dinâmica molecular. Assim, de um ponto de vista quantitativo, seria possível quantificar mais adequadamente o comportamento estrutural dos fragmentos protéicos em cada pH de estudo.

## **7. PERSPECTIVAS**

A partir dos resultados obtidos no desenvolvimento da presente dissertação, são destacadas, a seguir, algumas possibilidades de estudos futuros:

- 1) Estender as trajetórias das simulações das Seções 5.2 e 5.4.
- 2) Produzir a simulação em pH neutro para o campo de força AMBER.
- 3) Investir mais em estudos de múltiplas simulações de dinâmica molecular para quantificar com maior precisão o comportamento estrutural dos fragmentos protéicos em cada pH de estudo.

## **8. REFERÊNCIAS BIBLIOGRÁFICAS**

1. Doenças priônicas; *Ciência Hoje* **2005**, 37, 19.
2. Lledo, P.-M.; *História das vacas loucas*; Instituto Piaget; 2002; 1ª ed.; pg. 1-160.
3. <http://cienciahoje.uol.com.br/colunas/deriva-genetica/microbios-sem-dna> (acessado em 01/04/2012 às 14h).
4. Gajdusek, D. C.; Zigas, V.; *N. Engl. J. Med.* **1957**, 257, 974.
5. Hadlow, W. J.; *Lancet.* **1959**, 2, 289.
6. Gajdusek, D. C.; Gibbs, C. J.; Alpers, M. P.; *Nature* **1966**, 209, 794.
7. Gibbs, C. J.; Gajdusek, D. C.; Asher, D. M.; Alpers, M. P.; Beck, E.; Daniel, P. M.; Matthews, W. B.; *Science* **1968**, 161, 388.
8. Alper, T.; Cramp, W. A.; Haig, D. A.; Clarke, M. C.; *Nature* **1967**, 214, 764.
9. Griffith, J. S.; *Nature* **1967**, 215, 1043.
10. Linden, R.; *Ciência Hoje* **2003**, 33, 18.
11. Prusiner, S.; *Proc. Natl. Acad. Sci. USA* **1998**, 95, 13363.
12. Bolton, D. C.; McKinley, M. P.; Prusiner, S. B.; *Science* **1982**, 218, 1309.
13. Prusiner, S. B.; Bolton, D. C.; Groth, D. F.; Bowman, K. A.; Cochran, S. P.; McKinley, M. P.; *Biochemistry* **1982**, 21, 6942.
14. Bateman, D.; Hilton, D.; Love, S.; Zeidler, M.; Beck, J.; Collinge, J.; *Lancet.* **1995**, 346, 1155.
15. Bruce, M. E.; Will, R. G.; Ironside, J. W.; McConnell, I.; Drummond, D.; Suttie, A.; McCardle, L.; Chree, A.; Hope, J.; Birkett, C.; *Nature* **1997**, 389, 498.
16. Moore, R.; Taubner, L.; Priola, S.; *Curr. Opin. Struc. Biol.* **2009**, 19, 14.
17. a) Pan, K.M, Baldwin, M.;Nguyen, J.;Gasset, M.;Serban, A.; Groth, D.; Mehlhorn, I.; Huang, Z.; Flettericks, R. J.; Cohen, F. E.; Prusiner, S. B.; *Proc. Natl. Acad. Sci. USA* **1993**, 90, 10962.  
b) Caughey, B. W., Dong, A., Bhat, K. S., Ernst, D., Hayes, S. F.; Caughey, W. S.; *Biochemistry* **1991**, 30, 7672.
18. James, T.; Liu, H.; Ulyanov, N.; *Proc. Natl. Acad. Sci. USA* **1997**, 94, 10086.
19. Wadsworth, J.D.F.; Jackson, G.S.; Hill, A.F.; Collinge, J.; *Curr. Opin. Genet. Dev.* **1999**, 9, 338.
20. [http://portal.saude.gov.br/portal/saude/profissional/visualizar\\_texto.cfm?idtxt=31725&janela=1](http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=31725&janela=1) (acessado em 07/04/2012, 17:50).
21. DeMarco, M. L.; Daggett, V.; *C. R. Biol.* **2005**, 328, 847.
22. Foster, J. D.; Wilson, M.; Hunter, N.; *Vet. Rec.* **1996**, 139, 512.
23. Aguzzi, A.; Polymenidou, M.; *Cell* **2004**, 116, 313.

24. Linden, R.; Martins, V. R.; Prado, M. A. M.; Cammarota, M.; Izquierdo, I.; Brentani, R. R.; *Physiol. Rev.* **2008**, 88, 673.
25. Wechselberger, C.; Wurm, S.; Pfarr, W.; Höglinger, O.; *Exp. Cell Res.* **2002**, 281, 1.
26. Harris, D. A.; *Curr. Issues Mol. Biol.* **1999**, 1, 65.
27. Brown, D. R.; *Trends Neurosci.* **2001**, 24, 85.
28. Prusiner, S. B.; Scott, M.R.; *Annu. Rev. Genet.* **1997**, 31, 139.
29. Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R.; Wuthrich, K.; *Nature* **1996**, 382, 180.
30. Garcia, F. L.; Zahn, R.; Riek, R.; Wuthrich, K.; *Proc. Natl. Acad. Sci. USA* **2000**, 97, 8334.
31. Zahn, R.; Liu, A.; Luhrs, T.; Riek, R.; von Schroetter, C.; Lopez Garcia, F.; Billeter, M.; Calzolari, L.; Wider, G.; Wuthrich, K.; *Proc. Natl. Acad. Sci. USA* **2000**, 97, 145.
32. DeMarco, M.; Daggett, V.; *Proc. Natl. Acad. Sci. USA* **2004**, 101, 2293.
33. Alonso, D.; DeArmond, S.; Cohen, F.; Daggett, V.; *Proc. Natl. Acad. Sci. USA* **2001**, 98, 2985.
34. Daggett, V.; *Curr. Opin. Biotech.* **1998**, 9, 359.
35. DeMarco, M.; Silveira, J.; Caughey, B.; *Biochemistry* **2006**, 45, 15573.
36. Harris, D.A.; True, H.L.; *Neuron* **2006**, 50, 353.
37. <http://www.stanford.edu/group/virus/prion/prion2.html> (acessado em 14/04/2012 às 19:41).
38. Wille, H.; Michelitch, M. D.; Guénebaud, V.; Supattapone, S.; Serban, A.; Cohen, F. E.; Agard, D. A.; Prusiner, S. B.; *Proc. Natl. Acad. USA* **2002**, 99, 3563.
39. Aguzzi, A.; Baumann, F.; Bremer, J.; *Annu. Rev. Neurosci.* **2008**, 31, 439.
40. Harris, D.A.; *Clin. Microbiol. Rev.* **1999**, 12, 429.
41. Caughey, B.W.; Dong, A.; Bhat, K.S.; Ernst, D.; Hayes, S.F.; Caughey, W.S.; *Biochemistry* **1991**, 30, 7672.
42. Hill, A.F.; Antoniou, M.; Collinge, J.; *J. Gen. Virol.* **1999**, 80, 11.
43. Kovacs, G.G.; Budka, H.; *Am. J. Pathol.* **2008**, 172, 555.
44. Pauli, P.C; Harris, D.A.; *J. Biol. Chem.* **1998**, 273, 33107.
45. Swietnicki, W.; Petersen, R.; Gambetti, P.; Surewicz, W. K.; *J. Biol. Chem.* **1997**, 272, 27517.
46. Hornemann, S.; Glockshuber, R.; *Proc. Natl. Acad. Sci. USA* **1998**, 95, 6010.
47. Levitt, M.; Hirschberg, M. Sharon, R.; Daggett, V.; *Comput. Phys. Commun.* **1995**, 91, 215.



48. Martínez, L.; Borin, I. A.; Skaf, M. S.; Fundamentos de Simulação por Dinâmica Molecular. In: Morgon, N.H.; Coutinho, K. Métodos de Química Teórica e Modelagem Molecular. São Paulo: Livraria da Física, 2007 , p. 413.
49. Berg, J. M.; Tymoczko, J. L.; Stryer, L.; Bioquímica; Guanabara Koogan: Rio de Janeiro, 2008.
50. Nelson, D.L.; Cox, M.M.; Princípios de Bioquímica de Lehninger; Artmed: Porto Alegre, 2011.
51. Voet, D.; Voet, J. G.; Pratt, C.W.; Fundamentos de Bioquímica, A vida em nível molecular; Artmed: Porto Alegre, 2008.
52. Devlin, T. M.; Manual de bioquímica com correlações clínicas; Blücher: São Paulo, 2007.
53. <http://biologiabc1.blogspot.com.br/2010/10/estructuras-de-las-proteina.html> (acessado em 17/04/2012 às 16:40)
54. Floriano, W. B.; Nascimento, M. A. C.; Braz. J. Phys. **2004**, 34, 38.
55. <http://www.rcsb.org/pdb/home/home.do> (acessado em 20/04/2012, 17:51).
56. Lesk, A.; Introdução à Bioinformática; Artmed: Porto Alegre, 2008.
57. Guex, N.; Peitsch, M. C.; Electroforesis **1997**, 8, 2714.
58. a) Sayle, R.; Milner-White, E. J.; Trends Biochem. Sci. **1995**, 20, 374.  
b) Bernstein, H. J.; Trends Biochem. Sci. **2000**, 25, 453.
59. Leach, A.; Molecular Modelling: Principles and Aplications; Pearson: Harlow, 2001.
60. Scherer, C.; Métodos computacionais da física; Livraria da Física: São Paulo, 2005.
61. Apol, E.; Apostolov, R.H.; Berendsen, J.C.; Buuren, A.; Bjelkmar, P.; van Drunen, R.; Feenstra, A.; Groenhof, G.; Kasson, P.; Larsson, P.; Meulenhoff, P.; Murtola, T.; P'all, S.; Pronk, S.; Schulz, R.; Shirts, M.; Sijbers, A.; Tieleman, P.; Hess, B.; Spoel, D.; Lindahl, E.; Gromacs User Manual 4.5.4; 2010.
62. Oostenbrink, C.; Villa, A.; Mark, A.E.; Van Gunsteren W.F.; J. Comp. Chem. **2004**, 25, 1656.
63. McQuarrie, D. A.; Statistical Mechanics; HarperCollinsPublishers: Nova York ,1976.
64. a) Sorin, Eric J.; Pande, Vijay S.; Biophys. J. **2005**, 88, 2472.  
b) Wang, J.; Wolf, R.M.; Caldwell, J.M.; Kollman, P.A.; Case, D.A.; J. Comp. Chem. **2004**, 25, 1157.
65. a) Jorgensen, W. L. ; Maxwell, D. S.; Tirado-Rives, J.; J. Am. Chem. Soc. **1996**, 118, 11225.  
b) Kaminski, G. A.; Friesner, R. A. ;Tirado-Rives, J.; Jorgensen, W. L.; J. Phys. Chem. B **2001**, 105, 6474.

- 66.** Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Hermans, J.; D.; Reidel Publishing Company; Dordrecht **1981**, 33, 1342.
- 67.** Ponder, J.W.; Case, D.A.; *Adv. Prot. Chem.* **2003**, 66, 27.
- 68.** Amorim, H.; Netz, P. A.; Molecular dynamics simulations of proteins: theory and applications; artigo submetido à publicação.
- 69.** Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oosterbrink, C.; Schenk, M.; Trzesniak, D.; Van de Vegt, N. F. A.; Yu, H. B.; *Angew. Chem. Int.* **2006**, 45, 4065.
- 70.** Tironi, I.G.; Sperb, R.; Smith, P.E.; W.F. van Gunsteren; *J. Chem. Phys.* **1995**, 102, 5451.
- 71.** Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.; *J. Chem. Phys.* **1995**, 103, 8577.
- 72.** Deserno, M.; Holm, C.; *J. Chem. Phys.* **1998**, 109, 7678.
- 73.** Caves, L. S. D.; Evanseck, J. D.; Karplus, M.; *Protein Sci.* **1998**, 7, 649.
- 74.** Auffinger, P.; Louise-May, S.; Westhof, E; *J. Am. Chem. Soc.* **1995**, 117, 6720.
- 75.** Campos, S.R.R.; Machuqueiro, M.; Baptista, A.M.; *J. Phys. Chem. B* **2010**, 114, 12692.
- 76.** Ziff, B.; *Comput. Phys.* **1998**, 12, 385.
- 77.** Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.; *J. Chem. Phys.* **1983**, 79, 926.
- 78.** Nose, S.; *Mol. Phys.* **1984**, 52, 525.
- 79.** Parrinelo, M.; Rahman, A.; *J. Appl. Phys.* **1981**, 52, 7182.
- 80.** Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M.; *J. Comp. Chem.* **1997**, 18, 1463.
- 81.** Kabsch, W.; Sander, C.; *Biopolymers* **1983**, 22, 2577.
- 82.** Ferreira, R. S.; Olia, G.; Andricopulo, A. D.; *Quim. Nova* **2011**, 34, 1770.
- 83.** Langella, E.; Improta, R.; Crescenzi, O.; Barone, V.; *Proteins* **2006**, 64, 167.
- 84.** Langella, E.; Improta, R.; Barone, V.; *Bioph. J.* **2004**, 87, 3623.
- 85.** Van Gunsteren, W. F.; *Molecular Simulation: The GROMOS 96 Manual and User Guide*; Hochschulverlag AG an der ETH Zurich, 1996.

## **9. ANEXOS**

**9.1 ANEXO A1: PROPRIEDADES E CONVENÇÕES RELACIONADAS COM OS 20 AMINOÁCIDOS COMUNS.**

Aminoácido	Abreviação/		Valores de pK <sub>a</sub>			pI	Ocorrência em proteínas (%)
	Símbolo	M <sub>r</sub> *	pK <sub>1</sub> (-COOH)	pK <sub>2</sub> (-NH <sub>3</sub> <sup>+</sup> )	pK <sub>R</sub> (grupo R)		
<b>Grupos R alifáticos, apolares</b>							
Glicina	Gly G	75	2,34	9,6		5,97	7,2
Alanina	Ala A	89	2,34	9,69		6,48	5,2
Prolina	Pro P	115	1,99	10,96		5,97	6,6
Valina	Val V	117	2,32	9,62		5,98	9,1
Leucina	Leu L	131	2,36	9,6		6,02	5,3
Isoleucina	Ile I	131	2,36	9,68		5,74	2,3
Metionina	Met M	149	2,28	9,21			
<b>Grupos R alifáticos, aromáticos</b>							
Fenilalanina	Phe F	165	1,83	9,13		5,48	3,9
Tirosina	Tyr Y	181	2,2	9,11	10,07 (fenol)	5,66	3,2
Triptofano	Trp W	204	2,38	9,39		5,89	1,4
<b>Grupos R polares, não carregados</b>							
Serina	Ser S	105	2,21	9,15		5,68	6,8
Treonina	Thr T	119	2,11	9,62		5,87	5,9
Cisteína	Cys C	121	1,96	10,28	8,18 (sulfidril)	5,07	1,9
Asparagina	Asn N	132	2,02	8,8		5,41	4,3
Glutamina	Gln Q	146	2,17	9,13		6,65	4,2
<b>Grupos R carregados positivamente</b>							
Lisina	Lys K	146	2,18	8,95	10,53(ε-NH <sub>3</sub> <sup>+</sup> )	9,74	5,9
Histidina	His H	155	1,82	9,17	6 (imidazol)	7,59	2,3
Arginina	Arg R	174	2,17	9,04	12,48 (guanidina)	10,76	5,1
<b>Grupos R carregados negativamente</b>							
Aspartato	Asp D	133	1,88	9,6	3,65 (β-COOH)	2,77	5,3
Glutamato	Glu E	147	2,19	9,67	4,25 (γ-COOH)	3,22	6,3

\* Tabela adaptada do livro Princípios de Bioquímica de Lehninger.<sup>(50)</sup>

**9.2 ANEXO A2: COMPARAÇÃO ENTRE OS MODELOS DE CARGA DE DIFERENTES CAMPOS DE FOÇA PARA O AMINOÁCIDO SERINA**

	<b>AMBER 84</b>	<b>AMBER 94/99</b>	<b>OPL-UA</b>	<b>OPLS-AA</b>	<b>GROMOS 96</b>
<b>N</b>	-0,463	-0,4157	-0,57	-0,50	-0,28
<b>HN</b>	0,252	0,2719	0,37	0,30	0,28
<b>CA</b>	0,035	-0,0249	0,20	0,14	0,00
<b>HA</b>	0,048	0,0843		0,06	
<b>C</b>	0,616	0,5973	0,50	0,50	0,38
<b>O</b>	-0,504	-0,5679	-0,50	-0,50	-0,38
<b>CB</b>	0,018	0,2117	0,265	0,145	0,15
<b>HB</b>	0,119	0,0352		0,06	
<b>OG</b>	-0,55	-0,6546	-0,70	-0,683	-0,548
<b>HO</b>	0,31	0,4275	0,435	0,418	0,398

\* Tabela adaptada do artigo de Ponder e Case.<sup>(67)</sup>

*"Esta é a lição: nunca desista - nunca, nunca, nunca, nunca."*

*Winston Churchill em uma palestra em 1941.*