

UNIVERSIDADE DE MOGI DAS CRUZES

JOSÉ EDUARDO KROLL

**CARACTERIZAÇÃO DA INTERAÇÃO DA CATEPSINA B E
PAPAÍNA COM HEPARINA UTILIZANDO MÉTODOS
COMPUTACIONAIS**

Mogi das Cruzes, SP

2009

UNIVERSIDADE DE MOGI DAS CRUZES

JOSÉ EDUARDO KROLL

**CARACTERIZAÇÃO DA INTERAÇÃO DA CATEPSINA B E
PAPAÍNA COM HEPARINA UTILIZANDO MÉTODOS
COMPUTACIONAIS**

Dissertação apresentada ao programa de Mestrado em Biotecnologia da Universidade de Mogi das Cruzes como parte dos requisitos para a conclusão do mesmo.

Área de concentração: Ciências Exatas

Orientador: Prof. Dr. Cláudio Saburo Shida

Mogi das Cruzes, SP

2009

FICHA CATALOGRÁFICA

Universidade de Mogi das Cruzes - Biblioteca Central

Kroll, José Eduardo

Caracterização da interação da catepsina B e papaína com heparina utilizando métodos computacionais / José Eduardo Kroll. – 2009.

181 f.

Dissertação (Mestrado em Biotecnologia) -
Universidade de Mogi das Cruzes, 2009

Área de concentração: Ciências Exatas

Orientador: Prof. Dr. Cláudio Saburo Shida

1. Catepsina B 2. Papaína 3. Heparina 4.
Glicosaminoglicanos 5. Generalized Simulated
Annealing I. Shida, Cláudio Saburo

CDD 660.6

ATAS

ATA DA SESSÃO PÚBLICA DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO EM BIOTECNOLOGIA DA UNIVERSIDADE DE MOGI DAS CRUZES

Às catorze horas e trinta minutos do dia vinte e dois de dezembro de dois mil e oito, na Universidade de Mogi das Cruzes, realizou-se a defesa de dissertação "CARACTERIZAÇÃO DA INTERAÇÃO DA CATEPSINA B E PAPAÍNA COM HEPARINA UTILIZANDO MÉTODOS COMPUTACIONAIS" para obtenção do grau de Mestre pelo(a) candidato(a) **José Eduardo Kroll**. Tendo sido o número de créditos alcançados pelo(a) mesmo(a) no total de 50 (cinquenta), a saber: 26 unidades de crédito em disciplinas de pós-graduação e 24 unidades de crédito no preparo da dissertação, o(a) aluno(a) perfaz assim os requisitos para obtenção do grau de Mestre. A Comissão Examinadora estava constituída dos Senhores Professores Doutores Cláudio Saburo Shida, Ivarne Luis dos Santos Tersariol e Henrique Jesus Quintino de Oliveira da Universidade de Mogi das Cruzes, sob a presidência do(a) primeiro(a), como orientador(a) da dissertação. A Sessão Pública da defesa de dissertação foi aberta pelo Senhor Presidente da Comissão que apresentou o(a) candidato(a). Em seguida o(a) candidato(a) realizou uma apresentação oral da dissertação. Ao final da apresentação da dissertação, seguiram-se as arguições pelos Membros da Comissão Examinadora. A seguir a Comissão, em Sessão Secreta, conforme julgamento discriminado por cada membro, considerou o(a) candidato(a)

aprovado por unanimidade
(aprovado(a)/reprovado(a)) (unanimidade/maioria)

Mogi das Cruzes, 22 de dezembro de 2008.

Comissão Examinadora	Julgamento
<u>Cláudio Saburo Shida</u> Prof. Dr. Cláudio Saburo Shida	<u>aprovado</u> (aprovado(a)/reprovado(a))
<u>Ivarne Luis dos S Tersariol</u> Prof. Dr. Ivarne Luis dos S Tersariol	<u>APROVADO</u> (aprovado(a)/reprovado(a))
<u>Henrique Jesus Quintino Oliveira</u> Prof. Dr. Henrique Jesus Quintino Oliveira	<u>Aprovado</u> (aprovado(a)/reprovado(a))

“Se estamos mesmo sendo observados por seres extraterrestres, por que não ouvimos as risadinhas deles?”

(Orben's Comedy Fillers)

RESUMO

A catepsina B é uma cisteíno-protease encontrada no interior de lisossomos de células animais. Essa enzima apresenta uma atividade ótima em pH ácido, e é facilmente inativada em pH neutro e alcalino. A externalização dessa enzima por células carcinômicas induz a degradação da matriz extracelular, que apresenta um pH próximo ao neutro e é rica em glicosaminoglicanos (GAGs), os quais modulam a atividade e a estabilidade estrutural dessa enzima contra a desnaturação alcalina. Nesse trabalho, foi estudado a interação da catepsina B e papaína, um modelo de cisteíno-proteases, com o tetrasacarídeo, heparina, através de métodos computacionais utilizando o novo método *Generalized Simulated Annealing* e a Dinâmica Molecular. Os resultados obtidos sugerem que os GAGs modulam a atividade dessa classe de protease por neutralizarem os efeitos produzidos pela desprotonação das histidinas, e que essa neutralização não é somente o resultado de uma simples anulação de cargas, mas sim de um mecanismo complexo de interferência dinâmica resultante do contato do GAG com a protease, capaz de induzir a estrutura da enzima ao seu estado funcional em pH alcalino. Esse estudo irá contribuir no entendimento da participação das cisteíno-proteases em tumores e poderá ser o ponto de partida para o desenvolvimento de novos inibidores de diferentes naturezas químicas.

Palavras-chave: Catepsina B; Papaína; Heparina.

ABSTRACT

Cathepsin B is a cystein-protease found inside of animal cell lysosomes. This enzyme presents an optimum activity at acidic pH and is easily inactivated at neutral and alkaline pH. In some pathologies this enzyme is externalized by carcinomic cells inducing the degradation of the extracellular matrix, rich in glycosaminoglycans (GAGs). The interaction with GAG modules the activity and protected the enzyme against alkaline denaturation. In this work, it was studied the interaction of cathepsin B and papain, a cystein-protease model, with the heparin tetrasaccharide through computational methods using the Generalized Simulated Annealing Docking and Molecular Dynamics. The results suggest that the GAGs module this class of protease activity by neutralizing the effects produced by the histidines desprotonation, which is not only the result of a simple electrostatic charging neutralizing, but a complex dynamic mechanism of interference due to the contact between the GAG and the protease, capable to induce the enzyme structure to its functional state at an alkaline pH. This study will contribute to the knowledge of the role of cystein-proteases in tumors and it could be the start point to the developing of new inhibitors of different nature chemicals.

Keywords: Cathepsin B; Papain; Heparin.

LISTA DE TABELAS

Tabela 1: Diferenças entre o heparam sulfato e a heparina.....	25
Tabela 2: Alterações feitas no campo de força da heparina.....	68
Tabela 3: Energia potencial obtida nas diversas etapas do processo de minimização de energia da heparina.....	75
Tabela 4: Energia potencial obtida nas diversas etapas do processo de minimização de energia da papaína.....	76
Tabela 5: Energia potencial obtida nas diversas etapas do processo de minimização de energia da catepsina B.....	107
Tabela 6: Dados obtidos nos dockings (GSA) feitos entre a catepsina B e a heparina.....	108

LISTA DE ILUSTRAÇÕES

Figura 1: Unidades de dissacarídeos de maior e menor repetição no heparam sulfato e heparina, onde X= H ou SO ₃ , Y=Ac, SO ₃ ou H. Modificado de Capila & Linhardt (2002).....	23
Figura 2: Monossacarídeos encontrados no heparam sulfato e heparina.....	24
Figura 3: Representação de um tetrassacarídeo de heparina (sem os hidrogênios dos carbonos).....	25
Figura 4: A biossíntese do HS envolve uma série de modificações, incluindo a sulfatação e a epimerização do resíduo GlcA. Acredita-se que polimerização e a modificação ocorrem simultaneamente. (PAPS) 3´ -fosfoadenil-5´ -fosfosulfato é doador de grupos sulfato. Modificado de Varki <i>et al.</i> (1999).....	26
Figura 5: Representação do polipeptídeo da papaína; α -hélices e folha- β s são representadas por estruturas espirais e lisas, respectivamente. Os domínios L e R estão representados em verde e vermelho, respectivamente. O sítio da fenda se encontra no meio da parte superior da representação. Modificado de Musil <i>et al.</i> (1991).....	30
Figura 6: Sítios de interação do GAG com a catepsina B por GSA docking e dinâmica molecular (A e B); e (C) superfície eletrostática da superfície da protease – em azul as regiões de caráter básico.	35
Figura 7: (1) Número de pontes de hidrogênio dos complexos A e B; (2) Energia da interação entre o dissacarídeo e a catepsina B dos complexos A e B.....	36
Figura 8: Método Leap-Frog de integração. A velocidade obtida na metade do passo de tempo é utilizada para calcular a nova posição da partícula. Modificado de Spoel <i>et al.</i> (2005).....	40
Figura 9: Condições periódicas de contorno de um sistema em duas dimensões. Modificado de Spoel <i>et al.</i> (2005).....	42
Figura 10: Forma do potencial de interação de Lennard Jones (A) e de Coulomb (B) utilizando diferentes parâmetros. Modificado de Spoel <i>et al.</i> (2005).....	45
Figura 11: Princípio da distância de ligação (esquerda), e o potencial de distância de ligação (direita). Modificado de Spoel <i>et al.</i> (2005).....	47

Figura 12: Os átomos i, j e k estão na seqüência das ligações covalentes. O átomo j está no meio, e os átomos i e k estão nas extremidades. Modificado de Spoel <i>et al.</i> (2005).....	48
Figura 13: Princípio do ângulo de diedro impróprio . O ângulo de diedro impróprio é definido pelo ângulo entre os planos (i, j, k) e (j, k, l). Modificado de Spoel <i>et al.</i> (2005).....	48
Figura 14: Potencial de diedro impróprio. Modificado de Spoel <i>et al.</i> (2005).....	48
Figura 15: Princípio do ângulo de diedro próprio (esquerda) e o potencial de ângulo de diedro (direita). Modificado de Spoel <i>et al.</i> (2005).	49
Figura 16: Representação do caminho do algoritmo Steepest-Descent em uma dimensão. Modificado de Pascutti (2004).....	50
Figura 17: Representação do caminho do algoritmo dos gradientes conjugados em duas dimensões. Modificado de Pascutti (2004).....	51
Figura 18: (A) Representação gráfica do método de docking GSA; a linha tracejada representa a dimensão da energia de contato. (B) Representação de docking com temperatura constante. (C) Espaço de configurações e as respectivas energias (E).....	55
Figura 19: Esferas analisando a área de superfície acessível de um grupo de hemácias em duas dimensões através do método rolling-ball.....	56
Figura 20: Conceito da área de superfície acessível e da superfície de solvente excluído.....	57
Figura 21: Exemplo de ligações de hidrogênio.....	58
Figura 22: Exemplo de cadeia polipeptídica. Modificado de Lehninger, Nelson & Cox (2000).....	59
Figura 23: Gráfico Ramachandran mostrando os possíveis ângulos 'phi' e 'psi' e as respectivas estruturas secundárias. Modificado de Lehninger, Nelson & Cox (2000).....	59
Figura 24: Exemplo simplificado da aplicação do método MOD_GYR. A partícula em vermelho foi utilizada como centro.....	61
Figura 25: Representação gráfica do método de análise da superfície eletrostática.....	61
Figura 26: Seqüência primária da papaína (1BQI). resíduo básico; resíduo ácido; resíduo polar.....	63
Figura 27: Seqüência primária da Catepsina B (1HUC). resíduo básico; resíduo ácido; resíduo polar.....	64

Figura 28: Representação gráfica do Cluster do Laboratório de Modelagem Molecular que pertence ao Centro Interdisciplinar de Investigações Bioquímicas da UMC.....	65
Figura 29: Proteína centralizada com 1,2 nm de distância das paredes da caixa de simulação.....	69
Figura 30: Representação dos parâmetros utilizados no docking molecular.....	72
Figura 31: Complexo desfavorável entre a catepsina B e a heparina.....	73
Figura 32: Superfície eletrostática da papaína. Visão frontal (A) e visão traseira (B).....	77
Figura 33: Complexo papaína-heparina encontrado pelo THOR/GSA.....	78
Figura 34: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises RMSD anteriores (D).....	79
Figura 35: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 7 (B). Comparação entre as análises RMSD anteriores (C).....	80
Figura 36: Raio de giração em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de raio de giração anteriores (D).....	81
Figura 37: MOD_GYR por resíduo em função do tempo da papaína livre em pH 5 (A), livre em pH 7 (B), complexada em pH 7(C e D*), e para a estrutura da heparina livre em pH 7 (E) e complexada em pH 7(F e G*). *Obs: análise feita até 100 ns da dinâmica – gráfico fora da escala dos demais gráficos feitos até 60ns.....	82
Figura 38: Perfil da dinâmica de uma única oscilação da papaína (pH 7) complexada, baseado no MOD_GYR por resíduo.....	83
Figura 39: Representação em 2 dimensões da oscilação sofrida pela papaína com base nos resultados obtidos na análise feita pelo MOD_GYR por resíduo. “Frequência 2X” representa o dobro da frequência de “1X”	84
Figura 40: Análise MOD_GYR feita isoladamente para os resíduos 43 e 143 da papaína livre em pH 5 (A), papaína livre em pH 7 (B) e complexada em pH 7 (C).....	84
Figura 41: Comparação entre os perfis de oscilação da papaína em pH 5, pH 7 e complexada.....	85
Figura 42: Papaína e as regiões M1, M2 e M3 (conforme os resultados obtidos na figura 38). Os domínios R e L estão representados pelas respectivas letras.....	85

Figura 43: Transformada de Fourier aplicada para as oscilações MOD_GYR do resíduo 43 da papaína em pH 5 (A) e complexada em pH 7 (B).	86
Figura 44: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de energia potencial anteriores (D).....	87
Figura 45: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 7 (B). Comparação entre as análises anteriores (C).....	88
Figura 46: Conteúdo de estrutura α -hélice medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de conteúdo de estrutura α -hélice anteriores (D).....	89
Figura 47: Conteúdo de estrutura folha- β medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de conteúdo de estrutura folha- β anteriores (D).....	90
Figura 48: Diagrama helicoidal de duas estruturas α -hélices que sofreram transições estruturais durante a dinâmica da papaína complexada à heparina.....	91
Figura 49: Estruturas secundárias medidas em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Em azul, α -hélice, e em vermelho, folha- β	92
Figura 50: Acessibilidade ao solvente medida em função de cada resíduo na estrutura primária e do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Regiões mais escuras representam uma maior acessibilidade.....	94
Figura 51: Estrutura da papaína (vermelho, α -hélice, amarelo, folha- β) e, em destaque (azul), os resíduos que podem afetar as estruturas do tipo α -hélice através do aumento ou diminuição das respectivas acessibilidades.....	95
Figura 52: Acessibilidade ao solvente da HIS159 medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de acessibilidade da HIS159 anteriores (D).....	96
Figura 53: Comparação entre os períodos de tempo em que ocorreram a diminuição da acessibilidade da HIS159 e das ondas de oscilação MOD_GYR por resíduo correspondentes.....	96

Figura 54: Representação gráfica do complexo inicial (A), complexo no tempo de 40 ns (B), complexo no tempo de 84 ns (C) e superfície eletrostática da papaína no tempo de 0 ns (D).....	97
Figura 55: Número de ligações de hidrogênio entre a papaína e a heparina (A), e a energia de interação de curto alcance (Lennard Jones e Coulomb) entre essas mesmas moléculas (B).....	98
Figura 56: Distância dos resíduos da papaína à heparina em função do tempo (A). Apenas os átomos de oxigênio da heparina foram utilizados como referência para o cálculo de distância. Regiões de aparente maior importância na interação entre a papaína e a heparina baseado nos tempos em que foram verificados a diminuição da acessibilidade da HIS159 (B).....	99
Figura 57: Soma das distâncias (nm) entre os átomos de oxigênio da heparina e os resíduos da papaína dos domínios L e R, separadamente.....	100
Figura 58: Comparação do perfil de interação obtido no tempo exato de 40 ns (A) e 84 ns (B) com os perfis dos demais tempos.....	101
Figura 59: Relação entre o perfil de interação e o padrão de oscilação MOD_GYR da papaína complexada.....	102
Figura 60: Distância dos átomos de oxigênio da heparina às regiões 5-17 (A), 135-148 (B) e 173-188 (B) da papaína.....	102
Figura 61: Representação da papaína. As regiões 5-17, 135-148, 173-188 estão representadas em vermelho, azul e amarelo, respectivamente.....	103
Figura 62: Gráfico RMSD tridimensional entre as trajetórias da papaína em pH 5 e complexada (A) e entre a papaína em pH 7 e complexada (B). Somatória dos valores RMSD obtidos em função do tempo do complexo papaína-heparina: complexo versus papaína livre em pH 5 (C), e complexo versus papaína livre em pH 7(D).....	104
Figura 63: Gráfico suavizado (médias obtidas a cada 20 posições) e comparativo da somatória dos valores RMSD em função do tempo do complexo papaína-heparina (A), e gráfico da somatória dos valores RMSD entre a papaína em pH 7 e complexada em função do tempo do complexo e a sua relação com a acessibilidade do resíduo HIS159 (B).....	105
Figura 64: Distância entre os átomos de oxigênio da heparina e o resíduo HIS81 da papaína (A). Em destaque, no gráfico B, momentos de menor distância identificados entre as partes citadas anteriormente.....	106
Figura 65: Superfície eletrostática da catepsina B. Visão frontal (A) e visão traseira (B)....	108

Figura 66: Alguns complexos obtidos para a catepsina B e heparina através do THOR/GSA docking.....	109
Figura 67: Representação gráfica da trajetória de curta duração por MD criada a partir do primeiro complexo apresentado na figura 86.....	110
Figura 68: Gráficos de distância (A) e energia de contato (B) entre a heparina e a catepsina B baseados na dinâmica molecular do primeiro complexo apresentado na figura 66.....	110
Figura 69: Complexo catepsina B-heparina encontrado pelo THOR/GSA.....	111
Figura 70: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises RMSD anteriores (D).....	112
Figura 71: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 5 (B). Comparação entre as análises RMSD anteriores (C).....	113
Figura 72: Raio de giração em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C) . Comparação entre as análises de raio de giração anteriores (D).....	114
Figura 73: MOD_GYR por resíduo em função do tempo da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C).....	115
Figura 74: Detalhes do MOD_GYR por resíduo da catepsina B.....	116
Figura 75: Análise MOD_GYR por resíduo feita para os resíduos 80 e 210 da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7 (C).....	116
Figura 76: Transformada de Fourier aplicada para as oscilações MOD_GYR do resíduo 80 da catepsina B em pH 5 (A) e complexada em pH 7 (B).....	116
Figura 77: Energia potencial em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de energia potencial anteriores (D).....	117
Figura 78: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 5 (B). Comparação entre as análises de energia potencial anteriores (C).....	118
Figura 79: Conteúdo de estrutura α -hélice medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B), e complexada em pH 5 (C). Comparação entre as análises de conteúdo de estrutura α -hélice anteriores (D).....	120

Figura 80: Conteúdo de estrutura folha- β medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de conteúdo de estrutura folha- β anteriores (D).....	121
Figura 81: Estruturas secundárias em função do resíduo e do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Em azul, α -hélice, e em vermelho, folha- β	122
Figura 82: Acessibilidade ao solvente medida em função de cada resíduo e do tempo da trajetória da dinâmica molecular da estrutura da catepsina livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Regiões mais escuras representam uma maior acessibilidade.....	123
Figura 83: Acessibilidade ao solvente da HIS197 medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de acessibilidade da HIS197 anteriores (D).....	124
Figura 84: Representação gráfica do complexo inicial (A), complexo no tempo de 30 ns (B), complexo no tempo de 60 ns (C) e superfície eletrostática da catepsina B no tempo de 0 ns (D).....	125
Figura 85: Número de ligações de hidrogênio entre a catepsina B e a heparina (A), e a energia de interação de curto alcance (Lennard Jones e Coulomb) entre essas mesmas moléculas (B).....	126
Figura 86: Distância dos resíduos da catepsina B à heparina em função do tempo.....	127
Figura 87: Distância média da catepsina B à heparina em função do tempo.....	127
Figura 88: Representação da catepsina B. As regiões 50-70, 80-100, 120-150 estão representadas em vermelho, amarelo e azul, respectivamente.....	128
Figura 89: Degradação da matriz extracelular antagonizada pela catepsina B.....	129
Figura 90: Modelo para explicar a hipótese da afinidade dos aniões OH ⁻ pela superfície da catepsina B ou papaína em função da presença da heparina.....	132
Figura 91: Representação gráfica da interferência dinâmica sobre a energia estrutural da papaína.....	134
Figura 92: Modelo para ilustrar a hipótese da acessibilidade da histidina do sítio ativo da papaína sobre a inativação alcalina.....	137
Figura 93: Ilustração mostrando o provável efeito da heparina sobre as cisteíno-proteases. Obs: a enzima em preto está inativa, e em azul está ativa.....	138

LISTA DE ABREVIATURAS E SIGLAS

Å – Angstrom

ALA, A - Alanina

ARG, R - Arginina

ASA – Área de superfície acessível

ASN, N – Asparagina

ASP, D – Ácido aspártico

CD – Dicroísmo circular

CG – *Conjugated Gradient*

CP - Cisteíno-protease

CYS, C - Cisteína

FGF - Fator de crescimento de fibroblastos

GAG - Glicosaminoglicano

GlcA - D-glicosamina

GlcN – Ácido D-glicurônico

GlcNAc - Ácido D-glicurônico acetilado

GlcNS– D-Glicosamina sulfatada

GLN, Q – Glutamina

GLU, E – Ácido glutâmico

GLY, G – Glicina

GROMACS – *Groeningen Machine for Chemical Simulation*

GROMOS - *Groeningen Molecular Simulation*

GSA – *Generalized Simulated Annealing*

HIS, H – Histidina

HS – Heparan Sulfato

HSPG – Proteoglicano de heparan sulfato

IdoA – Ácido L-Idurônico

ILE, I - Isoleucina

Ks – Constante de dissociação do substrato

LEU, L - Leucina

LINCS - *Linear constraint solver*

LYS, K – Lisina

MC – Monte Carlo

MD – Dinâmica Molecular

MET, M - Metionina

MOD_GYR – Modificação do método de raio de giração

nm – Nanômetro

ns - Nanosegundo

PDB – *Protein data bank*

PG - Proteoglicano

PHE, F - Fenilalanina

pKa – Potencial da constante de dissociação ácida

PRO, P – Prolina

ps – Picosegundo

RF – *Reaction Field*

Rg – Raio de giração

SA – *Simulated Annealing*

SD - *Steepest Descent*

SER, S - Serina

SPC – *Simple point charge*

TGFP – Fator-P de transformação de crescimento

THR, T – Treonina

TRP, W - Triptofano

TYR, Y – Tirosina

VAL, V – Valina

SUMÁRIO

1. APRESENTAÇÃO.....	20
2. INTRODUÇÃO.....	22
2.1. Glicosaminoglicanos.....	22
2.1.1. heparam Sulfato & Heparina.....	22
2.1.1.1. Estrutura.....	23
2.1.1.2. Biossíntese.....	25
2.1.1.3. Funções Biológicas.....	27
2.2. Cisteíno-Proteases.....	28
2.2.1. Papaína e Catepsina B.....	28
2.2.1.1. Estrutura e atividade.....	29
2.2.1.2. Funções da Catepsina B.....	31
2.3. Cisteíno-proteases e Glicosaminoglicanos.....	32
2.3.1. Papaína e Catepsina B.....	33
2.3.1.1. Análise in silico: Catepsina B e Heparina.....	35
2.4. Teoria dos Métodos.....	36
2.4.1. O Método da Dinâmica Molecular.....	36
2.4.1.1. Integração das Equações de Movimento.....	39
2.4.1.2. Controle de Temperatura e Pressão.....	40
2.4.2. Condições Periódicas de Contorno.....	41
2.4.3. Campos de Força em Sistemas Atômicos.....	43
2.4.3.1. Interações de Ligações.....	47
2.4.4. Otimização da Geometria Molecular.....	49
2.4.4.1. “Steepest Descent”.....	50
2.4.4.2. “Conjugate Gradient”.....	51
2.4.5. “Docking” Molecular.....	52
2.4.5.1. Método de Monte Carlo.....	52
2.4.5.2. Método “Generalized Simulated Annealing”.....	53
2.4.6. Acessibilidade ao Solvente.....	56

2.4.7. Desvio da Raiz Média Quadrática.....	57
2.4.8. Distância Entre Átomos ou Grupos de Átomos.....	58
2.4.9. Ligações de Hidrogênio.....	58
2.4.10. Estrutura Secundária.....	59
2.4.11. Raio de Giração e MOD_GYR.....	60
2.4.12. Superfície Eletrostática.....	61
3. OBJETIVOS.....	62
4. MATERIAIS & MÉTODOS.....	63
4.1. Materiais.....	63
4.1.1. Equipamentos.....	65
4.2. Métodos.....	66
4.2.1. Preparação das Moléculas.....	66
4.2.2. Preparação do Ambiente de Simulação.....	68
4.2.3. Otimização da Geometria Molecular.....	69
4.2.4. Termalização e Dinâmica do Complexo.....	70
4.2.5. Parâmetros de Simulação.....	71
4.2.6. Montagem do Complexo.....	71
4.2.7. Análise de estabilidade e interação.....	72
5. RESULTADOS.....	75
5.1. Heparina.....	75
5.1.1. Preparação e Minimização de Energia.....	75
5.2. Papaína.....	76
5.2.1. Preparação e Minimização de Energia.....	76
5.2.2. Obtenção dos Complexos.....	76
5.2.3. Análise Estrutural.....	78
5.2.3.1. Desvio Médio Quadrático (RMSD).....	78
5.2.3.2. Raio de Giração.....	80
5.2.3.3. Raio de Giração Modificado por Resíduo (MOD_GYR).....	82
5.2.3.4. Energia Potencial.....	86
5.2.3.5. Estrutura Secundária.....	88
5.2.3.6. Acessibilidade.....	93
5.2.4. Análise da Interação.....	97

5.2.4.2. RMSD Entre Trajetórias.....	103
5.2.4.1. Distância da Heparina aos Resíduos Histidina	106
5.3. Catepsina B.....	107
5.3.1. Obtenção dos Complexos.....	107
5.3.2. Análise Estrutural.....	112
5.3.2.1. RMSD.....	112
5.3.2.2. Raio de Giração.....	113
5.3.2.3. Raio de Giração Modificado por Resíduo (MOD_GYR).....	115
5.3.2.4. Energia Potencial.....	117
5.3.2.5. Estrutura Secundária.....	119
5.3.2.6. Acessibilidade ao Solvente.....	123
5.3.3. Análise da Interação.....	125
6. DISCUSSÃO.....	129
6.1. Sobre os Resultados Preliminares.....	130
6.2. Análise da Estabilidade Dinâmica.....	132
6.3. análise da Estabilidade Estrutural.....	135
6.4. Análise da Interação de Contato.....	138
7. CONCLUSÕES & SUGESTÕES.....	140
REFERÊNCIAS.....	141
ANEXOS.....	153

1. APRESENTAÇÃO

A papaína e catepsina B são cisteíno-proteases encontradas no interior de lisossomos de células vegetais e animais, respectivamente. Por ser obtida facilmente, a papaína é utilizada como modelo da catepsina B. Essas enzimas apresentam uma atividade ótima em pH ácido, e são facilmente inativadas em pH neutro e alcalino (DICKINSON, 2002). Elas possuem uma atividade que deve ser estritamente controlada e isolada em compartimentos especiais. Falhas nesse controle podem levar à patologias (TWINING, 1994; CHAPMAN *et al.*, 1997).

Através de um descontrole celular, normalmente apresentado por células carcinômicas de mamíferos, as cisteíno-proteases, especialmente a catepsina B, podem ser superexpressadas, externalizadas e ancoradas na matriz extracelular, provocando a degradação dessa estrutura (BUCK *et al.*, 1992; SLOANE *et al.*, 1986; SZPADERSKA & FRANKFATER, 2001). A degradação da matriz extracelular pode ser um fator importante no processo de metástase, onde as células tumorais se deslocam para outras localizações, agravando a doença. Porém, a matriz extracelular é caracterizada por apresentar um pH próximo ao neutro. Nesse pH, a papaína e catepsina B são rapidamente inativadas, mas, devido a determinados fatores, essa inativação não ocorre tão rapidamente. Estudos demonstraram que alguns glicosaminoglicanos (GAGs) altamente sulfatados, encontrados na matriz extracelular, podem evitar que essas proteases sofram desestabilização estrutural e a desprotonação da histidina do sítio ativo, evitando, conseqüentemente, a perda da atividade da enzima (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001).

A interação entre as cisteíno-proteases e os glicosaminoglicanos é eletrostática (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001), o que sugere que o sítio de ligação encontrado nesse tipo de enzima é eletropositivo, uma vez que os glicosaminoglicanos são moléculas predominantemente eletronegativas. As interações eletrostáticas são capazes de neutralizar, atrair e repelir cargas encontradas próximas à região de contato intermolecular, remodelando, assim, as interações intramoleculares da enzima, e produzindo alterações estruturais relevantes de maior ordem. A interação das cisteíno-proteases com os glicosaminoglicanos e, a maior resistência da enzima ao pH alcalino, são caracterizadas, principalmente, pelo o aumento do conteúdo das estruturas α -hélices da enzima *in vitro*

(ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001). Observações detalhadas de diferentes variações estruturais podem explicar os mecanismos envolvidos na modulação da atividade das cisteíno-proteases pelos glicosaminoglicanos.

Neste trabalho, estudamos a interação da catepsina B e papaína com o tetrasacarídeo heparina por simulação computacional utilizando o novo método *Generalized Simulated Annealing* e a Dinâmica Molecular. Ambos esses métodos são aproximações teóricas dos métodos quânticos, e são baseados, principalmente, em modelos estatísticos e de interação. A Dinâmica Molecular é um método utilizado para simular a dinâmica de sistemas moleculares, e utiliza como princípio as equações de movimento de Newton, onde o sistema evolui linearmente em função do tempo. O *Generalized Simulated Annealing*, por outro lado, é um método estocástico, onde os movimentos moleculares ocorrem de forma aleatória com base na teoria de distribuição de Tsallis (MORET, 1998). Esse método pode ser utilizado na formação de complexos moleculares (*docking* molecular), onde as superfícies estruturais dos ligantes são varridas em busca de contatos de menor energia, os quais podem representar as interações reais que ocorrem *in vitro*. A aplicação da Dinâmica Molecular sobre esses complexos é importante para verificar a estabilidade do mesmo sobre a influência da entropia do sistema, assim como para observar interações e variações estruturais com possíveis implicações funcionais.

Dentro de determinadas limitações teóricas, é possível simular computacionalmente a possível interação entre a catepsina B e papaína com a heparina. A comparação dos resultados simulados com os resultados obtidos *in vitro*, como o aumento do conteúdo de α -hélices, pode auxiliar na validação dos métodos e no desenvolvimento de novas hipóteses com base em dados simulados. Estudos por simulação computacional, tendo como base estruturas de cisteíno-proteases cristalizadas com inibidores, têm sido realizados para compreender o mecanismo molecular de inibição do sítio ativo da enzima (YAMAMOTO *et al.*, 2000) com o objetivo de desenvolver novos inibidores (SELZER *et al.*, 1997; FENG *et al.*, 1996). O estudo da interação com GAGs irá contribuir no entendimento da participação das cisteíno-proteases em tumores e poderá ser o ponto de partida para o desenvolvimento ou aperfeiçoamento de diferentes tipos de inibidores.

2. INTRODUÇÃO

Neste capítulo é feita uma revisão sobre glicosaminoglicanos e as proteínas estudadas neste trabalho, além de introduzirmos a teoria dos métodos utilizados.

2.1. GLICOSAMINOGLICANOS

Os glicosaminoglicanos são heteropolissacarídeos que possuem uma grande diversidade de funções. Os GAGs são compostos por unidades de dissacarídeos que se repetem ao longo do polímero. A diversidade existente entre esses dissacarídeos caracteriza diferentes famílias de GAGs, sendo os de importância fisiológica o hialuronato, condroitina 4-sulfato, queratam sulfato, dermatam sulfato, heparam sulfato (HS) e a heparina. Esses GAGs, com exceção da heparina, são encontrados, em grande parte do tempo, ligados à proteínas, o que caracteriza os proteoglicanos, que podem ser classificados conforme a respectiva proteína de base. Os proteoglicanos, assim como os GAGs, possuem a propriedade de se ligarem a um grande número de proteínas (CONRAD, 1998; CARDIN & WEINTRAUB, 1989), caracterizando a diversidade e complexidade de suas funções.

Os GAGs possuem estruturas moleculares simples e funções mediadas por mecanismos físicos (mecânicos e eletrostáticos), inerentes à respectiva estrutura. Nessa seção, portanto, serão abordados os principais pontos relacionados ao HS e a heparina, incluindo descrições estruturais e funcionais.

2.1.1. HEPARAM SULFATO & HEPARINA

O heparam sulfato (HS) foi primeiramente identificado em preparações comerciais de heparina, onde as heparinas se encontravam pouco sulfatadas. Atualmente, sabe-se que o HS possui uma estrutura distinta da heparina e uma vasta gama de atividades biológicas. Eles também possuem origens distintas. A heparina é sintetizada no tecido conjuntivo pelos

mastócitos. Por outro lado, o HS, que é uma unidade integrante da grande maioria dos proteoglicanos, pode ser encontrado na matriz extracelular e, superfície celular de todos os órgãos e tecidos de mamíferos. (TOLEDO & DIETRICH, 1977; GOMES & DIETRICH, 1982)

Os proteoglicanos de HS (HSPG) podem ser classificados em famílias distintas. A caracterização é feita conforme a estrutura da proteína de base e dos respectivos GAGs. As duas principais famílias de HSPGs de membrana são os sindecans e glipicans (GALLAGHER & TURNBULL, 1992). A família perlecan é encontrada na matriz extracelular, em específico, na membrana basal. Outras proteínas, como o betaglicam, CD44 V-3 e o receptor TGFP 3, também podem ser glicosilados pelo HS (GALLAGHER & TURNBULL, 1992). As proteínas de base dos proteoglicanos são de grande importância no arranjo espacial das cadeias de GAGs e no ancoramento dos proteoglicanos em regiões específicas da superfície celular; importante quando há interações com moléculas intracelulares (SALMIVIRTA & JALKANEN, 1995).

2.1.1.1. Estrutura

O HS encontrado nos diferentes proteoglicanos podem ser diferentes quando comparados entre si. A estrutura do HS está diretamente relacionada à respectiva origem celular, e pode ser crítica em interações com proteínas (STRINGER & GALLAGHER, 1997). Mesmo possuindo uma estrutura molecular bastante variável, o HS, assim como a heparina, possuem um padrão estrutural único que os diferenciam dos demais GAGs.

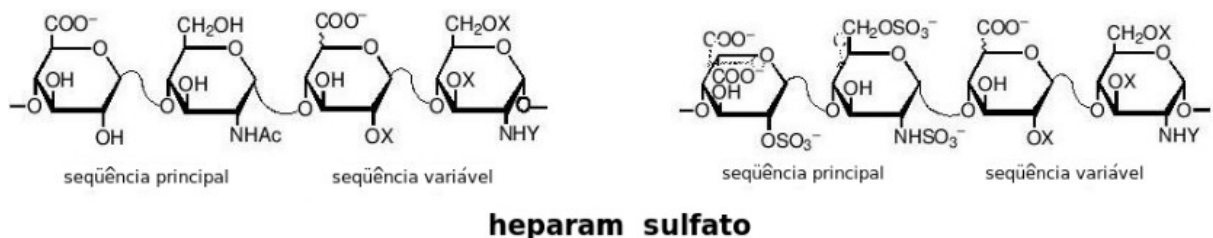


Figura 1: Unidades de dissacarídeos de maior e menor repetição no heparam sulfato e heparina, onde X= H ou SO₃, Y=Ac, SO₃ ou H. Modificado de Capila & Linhardt (2002).

Os polissacarídeos HS e heparina são compostos por monossacarídeos que alternam entre o ácido hexurônico [ácido D- glicurônico (GlcA) ou ácido L-idurônico (IdoA)] e

resíduos de D-glicosamina (GlcN), que podem ser N-sulfatados (NSO₃) ou N-acetilados (Nac; GALLAGHER & TURNBULL, 1992). O IdoA ocorre somente em conjunto com o GlcNSO₃, dois sacarídeos N-sulfatados que caracterizam o domínio S. Grupos O-sulfato podem ser encontrados no carbono 2 dos resíduos IdoA, e no carbono 6 e 3 do GlcNSO₃. Mais raramente, o GlcNAc pode ser 6-O-sulfatado quando próximo a um dissacarídeo N-sulfatado, e o GlcA pode ser também sulfatado no carbono 2 (STRINGER & GALLAGHER, 1997). Para maiores detalhes sobre as estruturas dos monossacarídeos presentes nos GAGs, ver figuras 1 e 2.

O tamanho dos domínios S e o padrão de O-sulfatação, os quais são características da origem celular dos HS, podem ser críticos em interações com proteínas (STRINGER & GALLAGHER, 1997). Esses domínios podem se ligar especificamente à proteínas extracelulares e moléculas sinalizadoras. A modulação da atividade dessas moléculas é o resultado de alterações conformacionais ou a habilidade dos domínios adjacentes do HS de intermediarem a interação entre duas ou mais moléculas (LEHNINGER, NELSON & COX, 2004).

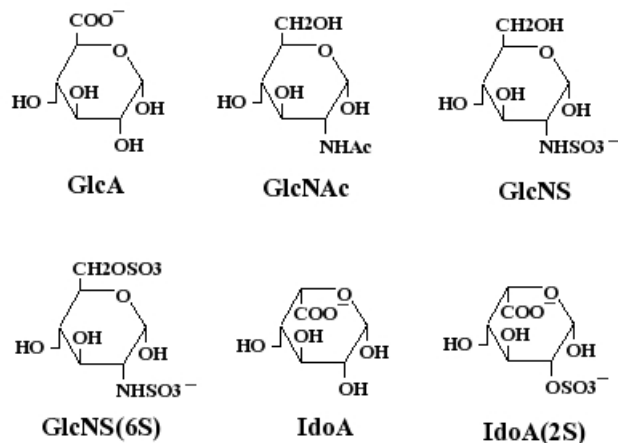


Figura 2: Monossacarídeos encontrados no heparam sulfato e heparina.

O dissacarídeo mais comum encontrado no HS é composto pelos resíduos GlcA e GlcNAc, que representam aproximadamente 50% dos dissacarídeos desse GAG. Por outro lado, a heparina encontrada nos pulmões é composta por 85% de resíduos IdoA(2S)-GlcNS(6S), enquanto a heparina da mucosa intestinal possui apenas 75%. Essas diferenças composicionais podem dificultar a qualificação de GAGs em heparina ou HS. Tem sido sugerido que um GAG pode ser qualificado como heparina somente quando a respectiva

quantidade de grupos N-sulfato exceder a de grupos N-acetil, e quando a quantidade de grupos O-sulfato exceder a de grupos N-sulfato (GALLAGHER & WALKER, 1985). Apesar dessas diferenças, a heparina (figura 3) é utilizada como um modelo para o heparam sulfato (MULLOY & FORSTER, 2000; CAPILA & LINHARDT, 2002), pois é facilmente disponível em grande quantidade. Maiores detalhes sobre as diferenças entre o HS e a heparina podem ser vistos na tabela 1.

Tabela 1: Diferenças entre o heparam sulfato e a heparina

Características	heparam Sulfato	Heparina
Tamanho	10 - 70 kD	10 - 12 kD
Sulfato/hexosamina	0.8 - 1.8	1.8 - 2.4
GlcN N-sulfatos	40 - 60%	≥ 85%
IdoA	30 - 50%	≥ 70%
Ligação à antitrombina (ver pág. 27)	0 - 0.3%	~ 30%
Local de síntese	todas células	mastócitos

Modificado de Varki *et al.* (1999)

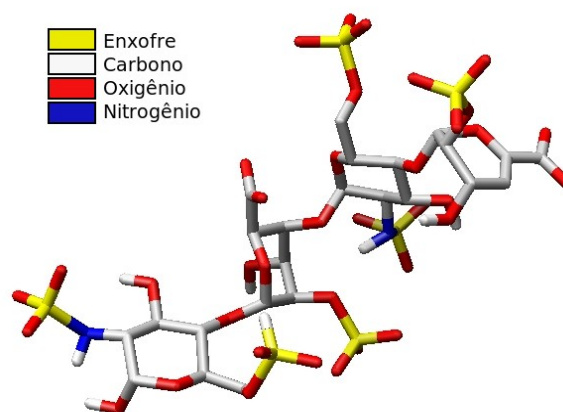


Figura 3: Representação de um tetrassacarídeo de heparina (sem os hidrogênios dos carbonos).

2.1.1.2. Biossíntese

As proteínas precursoras dos proteoglicanos de heparam sulfato são sintetizadas no retículo endoplasmático rugoso e translocadas para o complexo de Golgi, onde a biossíntese do HS ou heparina ocorre. O processo é iniciado pela xilosilação dos resíduos serina das seqüências -serina-glicina- (YANAGISHITA & HASCALL, 1992). A presença de grupos de aminoácidos ácidos, hidrofóbicos, e de unidades -serina-glicina- próximas, permitem que a

enzima GlcNAc transferase adicione o primeiro resíduo GlcNAc ao tetrasacarídeo intermediário: GAG - GlcA β 1,3-Gal β 1,3-Gal β 1,4-Xyl(\pm 2 - PO₄) - O -serina (ZHANG *et al.*, 1995). As unidades de monossacarídeos (GlcNAc e GlcA) são adicionadas ao terminal não redutivo da cadeia de HS crescente por uma única polimerase, formando a estrutura do polissacarídeo. (YANAGISHITA & HASCALL, 1992; SALMIVIRTA & JALKANEN, 1995).

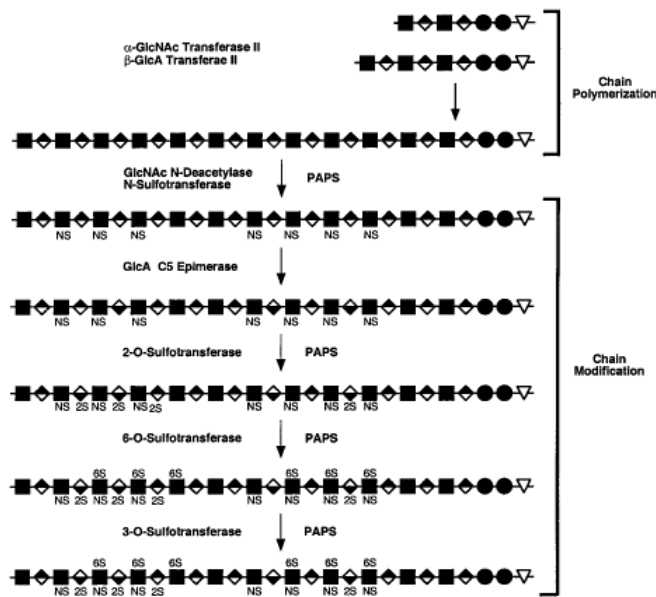


Figura 4: A biossíntese do HS envolve uma série de modificações, incluindo a sulfatação e a epimerização do resíduo GlcA. Acredita-se que polimerização e a modificação ocorrem simultaneamente. (PAPS) 3'-fosfoadenil-5'-fosfosulfato é doador de grupos sulfato. Modificado de Varki *et al.* (1999).

Assim que os monossacarídeos se polimerizam, eles sofrem uma série de reações modificadoras, catalisadas por pelo menos quatro famílias de sulfotransferases e uma epimerase. A GlcNAc N-deacetylase e a N-sulfotransferase, agem nos resíduos GlcNAc ao longo de determinadas regiões do polímero. Geralmente, o GlcNAc é deacetilado, e o seu grupo amino livre é rapidamente sulfatado, formando o resíduo GlcNSO₃; alguns dos resíduos GlcN deacetilados podem escapar da N-sulfatação. A epimerase, então, age nos resíduos GlcA adjacentes ao GlcNS, e os resíduos IdA recém gerados têm os seus átomos 2-O sulfatados. Em seguida, a sulfotransferase adiciona sulfato ao 6-OH dos resíduos GlcN adjacentes ao ácido urônico. Finalmente, determinadas combinações de resíduos sulfatados e epímeros do ácido urônico agem como alvo para a 3-O-sulfotransferase. Essas modificações tendem a ocorrer em apenas determinadas regiões do polímero, gerando regiões pobres e ricas em sulfato. A combinação específica de resíduos sulfatados e epímeros do ácido urônico na heparina e HS são importantes na especificidade de suas interações com outras moléculas (VARKI *et al.*, 1999). Para maiores detalhes, ver figura 4.

2.1.1.3. Funções Biológicas

O HS, assim como os proteoglicanos de heparam sulfato (HSPGs), são caracterizados por possuírem uma grande diversidade de funções ainda pouco exploradas. Entre as diversas funções dos HSPGs, podemos citar a aceleração da formação de complexos específicos de proteases e inibidores de proteases, e a mediação da interação entre citocinas e receptores celulares. Essas funções são basicamente essenciais para o controle da proliferação, diferenciação, adesão e migração celular (STRINGER & GALLAGHER, 1997). A atividade dos HSPGs é na maioria das vezes promovida por interações eletrostáticas entre os respectivos GAGs e proteínas ricas em aminoácidos básicos. Thompson *et al.* (1994) mostrou que dois terços dessas interações podem ser de ligações de hidrogênio, interações hidrofóbicas e de Van der Waals.

A heparina é um anticoagulante biológico natural. Visto que o HS e a heparina são estruturalmente bastante similares entre si, o HS também pode demonstrar propriedade anticoagulante. Eles promovem a anticoagulação por catalisarem, primariamente, a interação entre a protease inibitória ATIII e a protease pró-coagulante trombina, e entre a ATIII e o fator Xa (STRINGER & GALLAGHER, 1997).

No endotélio vascular, certas citocinas, como os interferons, podem interagir com o HS, facilitando a ativação de leucócitos. A atividade pró-coagulativa e anti-angiogênica do fator plaqueta 4 (PF4), são também reguladas pelo HS. O PF4 pode evitar ou romper a ligação do HS com várias proteínas efetoras, como a bFGF e ATIII. (STRINGER & GALLAGHER, 1997)

Os HSPGs de superfície celular são importantes no controle da atividade de determinados fatores de crescimento, como o fator de crescimento de fibroblastos (FGFs) (GALLAGHER & TURNBULL, 1992). A interação entre os HSs e os FGFs promove a concentração desse fator na superfície celular, induzindo uma maior interação entre os FGFs e seus respectivos receptores (SALMIVIRTA & JALKANEN, 1995). Nesse caso, o HSPG pode ser classificado como um co-receptor (LEHNINGER, NELSON & COX, 2004).

Os HSPGs de superfície celular são também responsáveis por ligar as células mesenquimais e epiteliais a uma variedade de componentes da matriz intersticial e membrana basal, incluindo a fibronectina, colágeno e trombospondina, aumentando, assim, a afinidade desses componentes pelos receptores de matriz, como as integrinas (SALMIVIRTA & JALKANEN, 1995). Nessa caso, os HSPGs são obviamente um pré-requisito para a adesão

celular (STRINGER & GALLAGHER, 1997).

2.2. CISTEÍNO-PROTEASES

A clivagem enzimática de ligações peptídicas é fundamental em diversos aspectos da vida e, por isso, as proteases representam aproximadamente 2% dos produtos genéticos da grande maioria dos organismos. As proteases podem ser encontradas comumente na digestão, coagulação do sangue, fibrinólise, pré-processamento de proteínas, função imune, desenvolvimento e apoptose celular (revisto em TWINING, 1994).

As proteases possuem uma grande diversidade em diversos aspectos. Elas são primariamente caracterizadas de acordo com o mecanismo utilizado na clivagem peptídica. As cisteíno-proteases (CPs), por exemplo, utilizam a cisteína como grupo nucleofílico no mecanismo de clivagem. As CPs são classificadas em ao menos seis grupos distintos, e totalizam 43 famílias, as quais metade são encontradas em vírus (RAWLINGS & BARRETT, 2000). Essas proteases, encontradas em diferentes organismos, podem apresentar estruturas similares, fruto da evolução convergente. Um exemplo clássico é a catepsina B e a papaína (CPs do grupo CA1), que são estruturalmente similares entre si e possuem como origem mamíferos e alguns vegetais, respectivamente.

2.2.1. PAPAÍNA E CATEPSINA B

As CPs são os membros mais conhecidos do grupo das CA já sintetizadas por técnicas bioquímicas, incluindo a enzima papaína (encontrada em vegetais), as catepsinas lisossomais B, H e L e a DPPI. Mais recentemente, a clonagem molecular tem sido utilizada na identificação de outras catepsinas CP, principalmente as de origem humana. O genoma humano mostrou possuir ao menos 11 catepsinas CP distintas: B, F, H, K, L, O, S, V, X, DPPI, e limfopaina. Essas enzimas são bastante similares à papaína. (KIRSCHKE, BARRETT & RAWLINGS, 1995)

2.2.1.1. Estrutura e atividade

A estrutura tridimensional de diversas enzimas similares à papaína (figura 5) já foram determinadas (KIRSCHKE *et al.*, 1995; TURK *et al.*, 1997; MCGRATH, 1999), com exceção da DPPI, a qual é oligomérica. Essas diferentes enzimas mostram possuir estruturas similares, e a grande maioria das mutações parecem ocorrer nos *loops*. Elas são normalmente bilobadas, ou seja, possuem dois domínios, os quais podem ser glicosilados. Esses domínios são designados de L e R, e são caracterizados por conterem as porções N e C-terminal da proteína, respectivamente. O sítio catalítico é encontrado entre os domínios, sendo a cisteína catalítica encontrada no domínio L e a histidina oposta no domínio R. Normalmente, a estrutura dos domínios é estabilizada por três ligações dissulfídicas. (revisto em DICKINSON, 2002)

Os substratos polipeptídicos se ligam ao longo da fenda encontrada entre os domínios da enzima. Os sítios de ligação do N-terminal, envolvidos na clivagem peptídica, são designados como S1, S2...etc; os sítios do C-terminal são designados como S1', S2'. Similarmente, P1, P2...; P1', P2'...etc, são utilizados para designar os resíduos do substrato correspondentes aos sítios da enzima (BARRETT, 1994). A maioria das enzimas possuem um resíduo hidrofóbico no sítio P2, onde a respectiva cadeia lateral se projeta no sítio S2, mas as CPs em geral variam bastante na acomodação desse resíduo aromático. As interações com os resíduos P3 e P2 dos substratos envolvem as cadeias laterais dos resíduos correspondentes, enquanto P2, P1, e P1' envolvem ambas as cadeias principais e laterais (TURK *et al.*, 1998; MCGRATH, 1999). A maioria das enzimas são endopeptidases, mas a catepsina B possui uma forte atividade carboxipeptidase (revisto em KIRSCHKE *et al.*, 1995). A catepsina B possui 18 resíduos de inserção próximos ao sítio da fenda, os quais formam a alça de oclusão. Ela restringe o acesso de potenciais substratos aos primeiros sítios e auxilia no ancoramento do grupo carboxil C-terminal. A flexibilidade da alça facilita a atividade endopeptidase (revisto em MORT & BUTTLE, 1997; MCGRATH, 1999).

A cisão das ligações peptídicas é feita por um ataque nucleofílico ao carbono carbonil, seguido por uma hidrólise ácido-base. As peptidases são geralmente agrupadas em cinco grupos principais (cisteína, serina, treonina, aspartato e metaloproteínase) de acordo com o mecanismo utilizado na geração do grupo nucleofílico no sítio ativo (BARRETT *et al.*, 1998). Nas cisteíno-proteases, um resíduo cisteína ativado é utilizado como nucleófilo e o resíduo histidina como doador de próton. Em algumas enzimas, um terceiro resíduo serve para

orientar o resíduo histidina (revisto em DICKINSON, 2002).

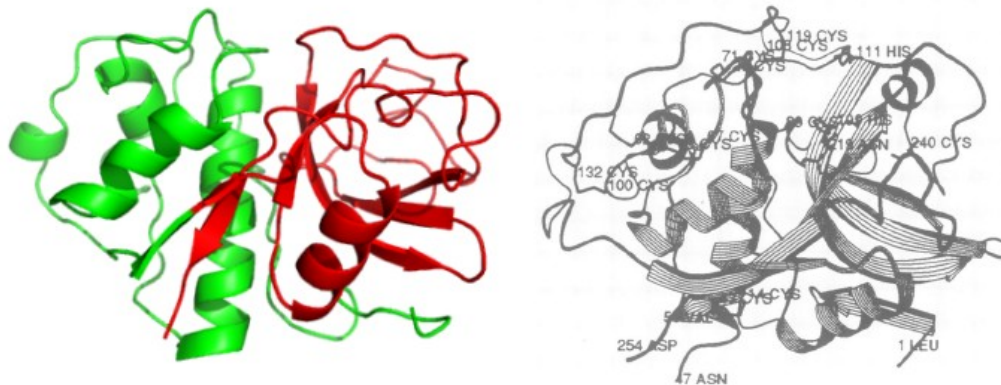


Figura 5: Representação do polipeptídeo da papaína; α -hélices e folha- β s são representadas por estruturas espirais e lisas, respectivamente. Os domínios L e R estão representados em verde e vermelho, respectivamente. O sítio da fenda se encontra no meio da parte superior da representação. Modificado de Musil *et al.* (1991).

As peptidases são enzimas cuidadosamente manipuladas pelas células, uma vez que falhas nesse processo podem levar à patologias. Além dos controles transcricionais e pós-transcricionais, diversos mecanismos são utilizados no controle da atividade das peptidases. (TWINING, 1994; CHAPMAN *et al.*, 1997). Grande parte das enzimas são sintetizadas na forma inativa e devem ser ativadas por clivagem proteolítica. Esse processo pode ocorrer automaticamente em condições específicas, como em um baixo pH. A liberação de peptidases das células é normalmente um processo controlado. As CPs são rapidamente inativadas pela oxidação da cisteína do sítio ativo e requer um meio redutor para a total atividade. Diversas CPs humanas são instáveis em pH neutro e requerem um pH ácido para a atividade total. Uma vez ativadas, a atividade da enzima pode ser perdida pela auto-degradação. Um controle que pode governar a atividade das peptidases é a presença de proteínas inibidoras que se ligam fortemente à enzima, bloqueando a ligação do substrato (revisto em DICKINSON, 2002).

As catepsinas B, H e L se encontram primariamente confinadas aos endossomos ou lisossomos, promovendo a degradação de proteínas que trafegam pelo sistema endossomal. Porém, atualmente é claro a idéia que as catepsinas B, H e L não são puramente lisossomais e podem ser liberadas das células em diversas circunstâncias. De acordo com essa situação, os HSPGs da superfície celular podem ancorar e concentrar um número de catepsinas B na membrana. Na presença de compostos tiol, a catepsina B é ativa entre o pH 5 e 6, a catepsina L é ativa entre o pH 3 e 6,5, e a catepsina H entre o pH 6,5-6,8 (KIRSCHKE *et al.*, 1995).

Nesses pHs, as catepsinas B, L e H podem degradar uma variedade de componentes da matriz extracelular, como os proteoglicanos, lâmina, e colágenos II, IX e XI (KIRSCHKE *et al.*, 1995). As catepsinas B, H e L são instáveis em pH neutro, e são irreversivelmente inativadas acima de pH 7 (BARRETT & KIRSCHKE, 1981). A catepsina L possui uma meia vida de apenas 1 minuto em pH 7,2 e 37 °C (WANG *et al.*, 1998), enquanto a catepsina B é 15 vezes mais estável (TURK *et al.*, 1995). O grau de degradação da catepsina B em pH neutro é minimizado na presença de certos substratos (BUCK *et al.*, 1992). A instabilidade dessas enzimas pode limitar severamente a degradação da matriz extracelular. Porém, já foi mostrado que algumas formas de catepsina B ligadas à membrana podem ser bastante resistentes à inativação em pH neutro

2.2.1.2. Funções da Catepsina B

As catepsinas B, H e L (BARRETT & KIRSCHKE, 1981; XING *et al.*, 1998) e seus respectivos mRNAs (SÖDERSTRÖM *et al.*, 1999) foram detectados em todos os tecidos e células examinados. Porém, há uma significativa variação nos níveis dessas enzimas nos diferentes tecidos e células (SÖDERSTRÖM *et al.*, 1999). A catepsina B é a catepsina mais abundante e expressada, e é encontrada em maiores concentrações nos macrófagos. Em relação ao mRNA, os maiores níveis são encontrados nos tecidos não esqueléticos, embora para a catepsina B as diferenças não são tão relevantes. Em ratos, elevados níveis de mRNA das catepsinas B e L são encontrados nos rins (DICKINSON, 2002). Há evidências de que os níveis de catepsina podem ser governados por processos pós-transcricionais, e os níveis de catepsina B, especificamente, podem estar associados com a diferenciação celular. (YAN *et al.*, 1998)

Até recentemente, a distribuição lisossomal das catepsinas B, H e L, indica que elas são enzimas essenciais para a degradação normal de proteínas celulares. Estudos mostraram que a inibição de diferentes CPs bloqueia 40% da degradação de proteínas. (SHAW & DEAN, 1980). Porém, a regulação por fatores de crescimento e variações nos níveis de expressão dessas proteases podem implicar na função de tecidos específicos. Ratos com deficiência em catepsina B apresentam, aparentemente, fenótipos normais (DEUSSING *et al.*, 1998).

As proteinases lisossomais, as quais incluem as CPs, estão envolvidas no

processamento de antígenos, e a importância individual de cada enzima continua incerta (VILLADANGOS & PLOEGH, 2000). O uso de inibidores específicos já evidenciaram o envolvimento das catepsinas B e D no processamento de antígenos (MAEKAWA *et al.*, 1998).

A catepsina L, e em menor extensão, a catepsina B, também estão envolvidas em eventos de remodelação de tecidos normais. Os níveis de mRNA da catepsina B aumentam em células apoptóticas do lúmen epitelial de próstata e glândulas mamárias; fato consistente com o processo de degradação da membrana basal, um evento que ocorre no início da morte celular (GUENETTE *et al.*, 1994). As CPs também já se mostraram essenciais para o desenvolvimento embrionário normal e na decidualização do útero. Porém, ratos deficientes em catepsina B e L, os quais aparentam se desenvolver normalmente durante a gestação, mostraram, novamente, a possibilidade da existência de diferentes outras enzimas e sistemas redundantes (SOL-CHURCH *et al.*, 1999). Esses sistemas e enzimas redundantes sugerem que a catepsina B pode ser eventualmente inibida com objetivos terapêuticos sem causar grandes efeitos colaterais no indivíduo.

2.3. CISTEÍNO-PROTEASES E GLICOSAMINOGLICANOS

Tem sido demonstrado que as CPs lisossomais, especialmente a catepsina B, podem participar da invasão tumoral pela degradação dos componentes da matriz extracelular (KATUNUMA, 1989 *apud* ALMEIDA *et al.*, 2001). Esse processo pode acontecer intracelularmente pela atividade heterofagossomal das células tumorais (SHEAHAN, SHUJA & MURNANE, 1989) ou extracelularmente pela catepsina B associada à superfície celular (DEL NERY *et al.*, 1997). A presença da catepsina B na membrana plasmática resulta na dissolução das proteínas da matriz extracelular, possibilitando a metástase das células tumorais (Buck *et al.*, 1992; BARRETT & KIRSCHKE, 1981). Um alto nível de catepsina B e mudanças na expressão da proteína, incluindo um padrão anormal de glicosilação, podem ser importantes na manutenção do fenótipo do carcinoma (SLOANE *et al.*, 1988). Alterações na expressão da catepsina, processamento, e localização celular, tem sido observado em diversos tecidos tumorais (HIRATA *et al.*, 1994).

Como revisto na seção 2.2.1.1, as catepsinas são encontradas principalmente em meios

ácidos (lisossomos e endossomos), e podem ser facilmente inativadas em um pH neutro como o encontrado na matriz extracelular; porém, não é isso o que ocorre naturalmente. Estudos recentes demonstraram que o tempo de vida da catepsina B pode ser prolongada na matriz extracelular por interações com GAGs da matriz extracelular, os quais podem modular as propriedades da enzima, explicando a grande relação da catepsina B no processo de metástase. (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001)

2.3.1. PAPAÍNA E CATEPSINA B

A heparina, assim como GAGs similares, aumentam o conteúdo α -hélice da papaína e catepsina B. Esse processo é acompanhado pela diminuição do conteúdo de folha- β e estruturas remanescentes da enzima. A heparina permite que a papaína e catepsina B sejam estáveis em pH alcalino. A presença da heparina reduz a perda do conteúdo de α -hélice induzido por um pH alcalino, aumentando a estabilidade da enzima, que é aparentemente reflexo de uma maior quantidade de estruturas hélices observadas. (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001)

A interação com a heparina e heparan sulfato é bastante específica. Outros GAGs sulfatados, como o dermatan sulfato e condroitim sulfato, não são capazes de aumentar a afinidade da papaína e catepsina B por substratos ou induzir o aumento do conteúdo de α -hélice. Por outro lado, o dextran sulfato, um polímero mais sulfatado, possui um maior efeito que a heparina na atividade da catepsina B. A característica eletrostática dessas ligações parece ser mais importante do que as características estruturais dos GAGs. (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001)

O domínio de ligação onde a heparina se liga na papaína ainda não foi demonstrado, porém, Cardin & Weintraub (1989) sugeriram que a seqüência 188-191 (RIKR) é um sítio de ligação putativo da heparina. Eles também propuseram que os sítios de ligação podem ocorrer em estruturas α -hélice e folha- β . O putativo sítio de ligação da heparina na região 188-199 da seqüência da papaína apresenta uma estrutura do tipo folha- β (KAMPHUIS *et al.*, 1984). Essa seqüência catiônica se encontra adjacente aos resíduos ASN¹⁷⁵ e TRP¹⁷⁷ e contribui para o campo eletrostático, influenciando na formação do par de íons cataliticamente ativos tiolato-imidazólio, onde o resíduo triptofano é também envolvido em interações entre a enzima e o substrato (BAKER & DRENTH, 1987)

Na catepsina B, a deleção do *loop* de oclusão resulta na diminuição da estabilidade da enzima (ILLY *et al.*, 1997). Illy *et al.* (1997) também mostrou que a ponte salina entre os resíduos HIS¹¹⁰ e ASN²² é um importante contato entre o *loop* de oclusão e a α -hélice central, e essa interação pode contribuir para a estabilidade desses elementos estruturais (MUSIL *et al.*, 1991; NÄGLER *et al.*, 1997; QURAIISHI *et al.*, 1999). Por outro lado, o resíduo HIS¹¹¹ não está envolvido em interações eletrostáticas com elementos da enzima (MUSIL *et al.*, 1991; NÄGLER *et al.*, 1997), mas a proteção contra a inativação alcalina é bastante dependente desse resíduo. A fraca inibição da catepsina B de membrana por compostos como o CA-030, os quais requerem interações com os resíduos HIS¹¹⁰ e HIS¹¹¹, está provavelmente relacionada às interações da enzima com o heparam sulfato da superfície celular (NÄGLER *et al.*, 1997).

A inativação induzida pelo pH alcalino, assim como o desestruturamento da catepsina B (TURK *et al.*, 1994) e L (DUFOUR, DIVE & TOMA, 1988; TURK *et al.*, 1993) mostraram ser um processo de primeira ordem, indicando que a inativação das CPs está relacionada à estabilidade da proteína. A desprotonação da histidina do sítio ativo (HIS¹⁵⁹ da papaína e HIS¹⁹⁷ da catepsina B), catalizada pelos aniões OH⁻, é considerada um evento crucial para a inativação das CPs em pH alcalino. A quebra do par de íons tiolato-imidazólio influencia a ionização e exposição ao solvente de alguns resíduos carregados localizados na interface interna, resultando em alterações conformacionais na enzima e na destabilização da α -hélice central (DUFOUR, DIVE & TOMA, 1988; TURK *et al.*, 1993).

A interação da heparina ou heparam sulfato com a papaína ou catepsina B pode alterar a orientação relativa da estrutura superficial da enzima. Essa mudança conformacional pode se dissipar para o resto da proteína através da estrutura terciária ou ligações dissulfídicas (LELLOUCH & LANSBURY, 1992). Os resíduos arginina e lisina demonstraram estabilizar os domínios de ligação da heparina de várias proteínas. A ligação da heparina pode estabilizar as estruturas α -hélice da papaína ou catepsina B por eliminar interações eletrostáticas, e os grupos negativamente carregados do sulfato e carboxilato da heparina podem neutralizar as cargas positivas que podem contribuir para a destabilização das α -hélices (FERRAN, SOBEL & HARRIS, 1992).

2.3.1.1. Análise *in silico*: Catepsina B e Heparina

Recentemente, o estudo com *docking* da heparina dissacarídea à catepsina B, utilizando o algoritmo *Generalized Simulated Annealing* (GSA), foi identificado dois sítios de interação do dissacarídeo. O potencial eletrostático superficial da catepsina B mostrou duas regiões onde é possível a interação do GAG com resíduos básicos. A primeira região se encontra próxima a uma série de resíduos básicos, ARG⁸³, ARG⁸⁴, LYS¹²⁵, LYS¹²⁸, LYS¹³⁹, LYS¹⁴² e ARG²⁵⁰, os quais fazem com que esta seja a região na superfície da protease com caráter mais básico (figura 6A). Este sítio é uma região que contém aminoácidos cujo prolongamento é justamente a alça de oclusão. Acredita-se que GAGs maiores poderiam interagir com esta região e a alça de oclusão, afetando assim a atividade da enzima. Para reforçar esta hipótese, experimentalmente verificou-se que diferentes tamanhos de GAGs modulam de modo diferente a atividade da enzima. (SHIDA *et al.*, artigo em preparação)

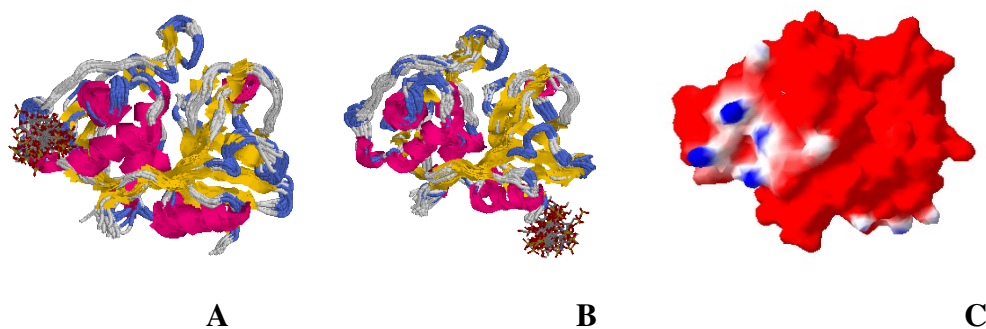


Figura 6: Sítios de interação do GAG com a catepsina B por GSA *docking* e dinâmica molecular (A e B); e (C) superfície eletrostática da superfície da protease – em azul as regiões de caráter básico.

O segundo sítio de interação (figura 6B), é a região do N-terminal da cadeia A da catepsina B e alguns outros resíduos básicos. A dinâmica molecular dos dois complexos (figura 6A e B) mostrou que o segundo complexo (B), mostrado na figura 6B, é energeticamente menos favorável e, apresenta, em média, um número menor de ligações de hidrogênio entre as duas moléculas que a do primeiro complexo (A) da figura 6A (figura 7). (SHIDA *et al.*, artigo em preparação)

Segundo Cardin & Weintraub (1989), as regiões de proteínas que interagem com GAG, tendem a apresentar dois padrões de distribuição, XBBXBX ou XBBBXXBX, onde B

representa resíduos com carga positiva (ARG, LYS) e X, resíduos neutros e apolares. Na catepsina B, este padrão é encontrado próximo à CYS²⁴⁰, LYS²³², ARG²³⁵ e HIS²³⁹. No entanto, nem o GSA *docking* nem a análise do potencial eletrostático superficial indicam que este seja um sítio de ligação (SHIDA *et al.*, artigo em preparação).

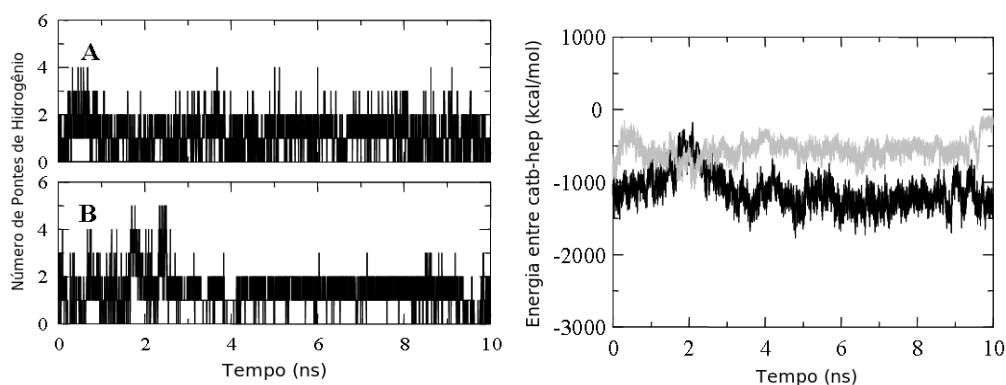


Figura 7: (1) Número de pontes de hidrogênio dos complexos A e B; (2) Energia da interação entre o dissacarídeo e a catepsina B dos complexos A e B.

2.4. TEORIA DOS MÉTODOS

2.4.1. O MÉTODO DA DINÂMICA MOLECULAR

A simulação por Dinâmica Molecular (MD) é um dos métodos utilizados para estudar sistemas macroscópicos através da evolução de um sistema ao nível molecular. As diferentes simulações moleculares existentes podem ser classificadas através do grau determinístico usado na geração das posições moleculares. Dos métodos estocásticos, como o de Monte Carlo, aos métodos puramente determinísticos, como o da Dinâmica Molecular, podemos encontrar diversos outros métodos, entre os quais, podemos citar o *Force-Biased Monte Carlo*, *Brownian Dynamics*, *General Langevin Dynamics* (GRØNBECH-JENSEN & DONIACH, 1994), *Dissipative Particle Dynamics* (HOOGERBRUGGE & KOELMAN, 1992; GROOT & WARREN, 1997), *Collisional Dynamics* (LEMAK & BALABAEV, 1996) e *Reduced Variable Molecular Dynamics* (TURNER *et al.*, 1995).

Resumidamente, a Dinâmica Molecular é um método que simula a dinâmica de sistemas moleculares através da integração das equações de movimento de Newton (clássica) de

cada partícula. Por utilizar uma aproximação clássica, as interações atômicas são calibradas através de estimativas experimentais e/ou teóricas. O movimento das partículas e energias envolvidas podem ser facilmente traduzidos em observáveis macroscópicas através da aplicação da mecânica estatística. Além do mais, a MD apresenta a seguinte característica: o sistema evolui em função do tempo. Isso classifica esse método como determinístico, ou seja, as conformações moleculares são sempre calculadas com base nas anteriores, descrevendo, assim, uma trajetória temporal previsível.

Entrando em detalhes no método da MD; o vetor \vec{r}^N denota uma série de subvetores que localizam o centro de massa de cada um dos átomos encontrados em um sistema, $\vec{r}^N = \{\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N\}$. As partículas do sistema possuem D graus de liberdade, e N partículas interagem entre si através de um potencial, $V(\vec{r}^N)$, que integra modelos de interações entre partículas (ver seção 2.4.3). A massa de cada partícula i é m_i , e F_i é a força total que atua sobre a partícula i no tempo t . A equação de Newton de movimento para cada partícula, $i=1, \dots, N$, pode ser escrita como:

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} \equiv m_i \ddot{\vec{r}}_i(t) = \vec{F}_i(t) \quad (1)$$

A força que a partícula j exerce sobre a partícula i é

$$f_{i,j} = -\frac{dV(r_{i,j})}{dr_{i,j}} \quad (2)$$

onde $r_{i,j} = \|r_i - r_j\|$ é a distância que separa as partículas i e j . A energia potencial total (E_p) do sistema é a soma da energia potencial (V) de todos os pares (equação 3), e a força total (F_i) que atua sobre cada partícula, i , é encontrada através da soma da força de todos os pares

$$E_p = \frac{1}{2} \sum_{i,j} V(r_{i,j}) \quad (3)$$

$$F_i = \sum_{j=1}^N f_{i,j'} (j \neq i) \quad (4)$$

Nota-se que, através da terceira lei de Newton, $\mathbf{f}_{ij} = -\mathbf{f}_{ji}$, cada interação entre os pares deve ser calculada apenas uma vez, porém, a equação (4) não descreve essa função e, portanto, quando o mesmo par é calculado, utiliza-se as forças \mathbf{F}_i e \mathbf{F}_j com sinais opostos.

Na MD, as novas posições dos átomos são obtidas através da solução da equação de movimento de Newton. Para solucionar a equação (1), utiliza-se a equação (4), sendo também necessário especificar as condições iniciais e limites espaciais do sistema de D dimensões. Isto resulta em uma série de equações $D \times N$ ordinárias diferenciais de segunda ordem e um total de $D \times N$ graus de liberdade. As novas posições e velocidades de cada um dos átomos do sistema são calculadas numericamente em função do tempo. O algoritmo pode ser representado da seguinte forma: (1) Especificação das condições iniciais (N , temperatura inicial, modelos de potenciais, conectividade molecular, passo do tempo, densidade, ...); (2) Construção da estrutura inicial do sistema e atribuição das velocidades iniciais para cada partícula; (3) Para cada passo de tempo de simulação, todas as forças, energias e propriedades opcionais (temperatura, pressão, ...) são computadas e integradas às equações de movimento. (4) As propriedades do sistema são amostradas em intervalos regulares; (5) Através das propriedades amostradas, o sistema é ajustado (temperatura, pressão etc) para satisfazer as condições especificadas.

Em uma MD de um sistema isolado e em equilíbrio, temos um volume V fixo, assim como um número fixo de átomos N . Por ser um sistema isolado, a energia total E é constante e, portanto, as variáveis N, V e E determinam os estados termodinâmicos. Os átomos do sistema interagem entre si através de modelos de potenciais. Suas posições são obtidas através da solução das equações de movimento para cada átomo. Normalmente, a dimensão é $D=3$, mas $D=2$ é também comum, e também há diversos exemplos de simulações em 4 dimensões (BEUTLER & VAN GUNSTEREN, 1994). Através da solução da equação 1, é possível obter as posições e velocidades de cada partícula em função do tempo. Através da aplicação da teoria cinética, mecânica estatística e da média do tempo sobre a trajetória dessas partículas, podemos computar propriedades macroscópicas. Na dinâmica molecular, essas propriedades podem representar propriedades estatísticas, como a temperatura, pressão, funções de distribuição de pares e propriedades dinâmicas, como os coeficientes de transporte.

Após completar a simulação, a trajetória é analisada, produzindo-se os resultados. Assume-se que através da média de um suficiente número de passos de tempo é possível obter

medidas correspondentes às médias derivadas do conjunto *NVE* (número de partículas, volume e energia). A inicialização de um sistema, análise e visualização de uma grande quantidade de dados pode ser desafiante, assim como os recursos computacionais necessários para produzir a própria simulação.

2.4.1.1. Integração das Equações de Movimento

O algoritmo *Leap-Frog* (HOCKNEY, GOEL & EASTWOOD, 1974) é utilizado na integração das equações de movimento, sendo implementado na grande maioria das dinâmicas moleculares. Ele utiliza as posições r no tempo t e velocidades v no tempo $t - \Delta t/2$, e atualiza as posições e velocidades através das forças $F(t)$ determinadas nas posições em função do tempo t :

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m} \Delta t \quad (5)$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t \quad (6)$$

O algoritmo *Leap-Frog* (figura 8) se baseia no algoritmo de Verlet (VERLET, 1967):

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m} \Delta t^2 + O(\Delta t^4) \quad (7)$$

O algoritmo é de terceira ordem em r e permite reversão no tempo. O método de Verlet, ao contrário do método *Leap-Frog*, não utiliza a média da velocidade ($v(t + \Delta t/2)$) e é portanto mais susceptível à erros. Por terem controle sobre as equações de velocidade, essas equações são normalmente modificadas para controlarem parâmetros de temperatura e pressão, e podem ser também estendidas para incluírem outras funções, como a restrição de movimento (*restraints*). (SPOEL *et al.*, 2005)

Nos algoritmos de Verlet ou *Leap-Frog*, assim como em outros métodos, as equações de movimento são resolvidas para cada átomo em função do espaço de tempo percorrido. O processo de avaliação das forças para a obtenção das acelerações é um dos processos que mais

consomem tempo computacional na dinâmica molecular, e depende diretamente da complexidade das funções de energia potencial. Após as forças serem obtidas, obtêm-se, portanto, a aceleração, a qual será utilizada posteriormente para obter as novas posições das partículas.

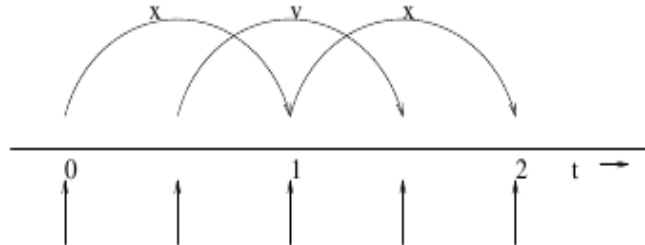


Figura 8: Método *Leap-Frog* de integração. A velocidade obtida na metade do passo de tempo é utilizada para calcular a nova posição da partícula. Modificado de Spael *et al.* (2005).

2.4.1.2. Controle de Temperatura e Pressão

A temperatura (T) do sistema é calculada com base na mecânica estatística, usando a energia cinética média de N partículas com N_f graus de liberdade

$$\frac{1}{2} k_b N_f T = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \quad (8)$$

onde k_b é a constante de Boltzmann, v_i é a velocidade da partícula i com massa m_i .

A expressão para a pressão (P) do sistema é baseada no teorema virial (CLAUSIUS, 1870). A pressão é expressa como a soma da parte ideal $PV = NkT$ e a soma dos produtos de todas as forças F_{ij} e distâncias entre as partículas r_{ij} :

$$PV = NkT + \frac{1}{3} \sum_{i < j}^N (r_{ij} F_{ij}) \quad (9)$$

No controle de temperatura de um sistema simulado, a velocidade das partículas podem ser reescaladas. Em cada passo de integração, a velocidade v é agora representada por λv , onde

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T^{ref}}{T(t-1/2 \Delta t)} - 1 \right) \right]^{\frac{1}{2}} \quad (10)$$

O objetivo da temperatura do sistema no instante $T(t-1/2 \Delta t)$ é alcançar a temperatura de referência T^{ref} . A constante τ_T descreve o quão rápido será o acoplamento da temperatura, e um valor elevado significa que haverá pequenas flutuações em torno da T^{ref} . O acoplamento da pressão P^{ref} é feita através do redimensionamento das coordenadas r para ηr da caixa de simulação, onde

$$\eta = \left[1 + \frac{\gamma \Delta t}{\tau_p} (P(t) - P^{ref}) \right]^{\frac{1}{2}} \quad (11)$$

e onde $P(t)$ é a pressão do sistema no tempo t . Novamente, τ_p descreve o grau de acoplamento. (BERENDSEN *et al.*, 1984)

2.4.2. CONDIÇÕES PERIÓDICAS DE CONTORNO

As condições periódicas de contorno são utilizadas nas simulações com o objetivo de minimizar o efeito produzido pelo tamanho finito do sistema. Uma das formas para solucionar esse problema é replicar o número de caixas e acomodá-las periodicamente ao longo da caixa principal, criando um sistema de dimensões pseudo-infinitas (figura 9). A escolha da caixa computacional deve permitir o acomodamento espacial de um número infinito de caixas, assim como o seu acomodamento em todas as direções do espaço. Existem diversos formatos de caixas (cubo, tetraedro, octaedro, etc) (FINCHAM & HEYES, 1985 *apud* HEDMAN, 2006), e a caixa do tipo cubo é a mais simples e mais utilizada.

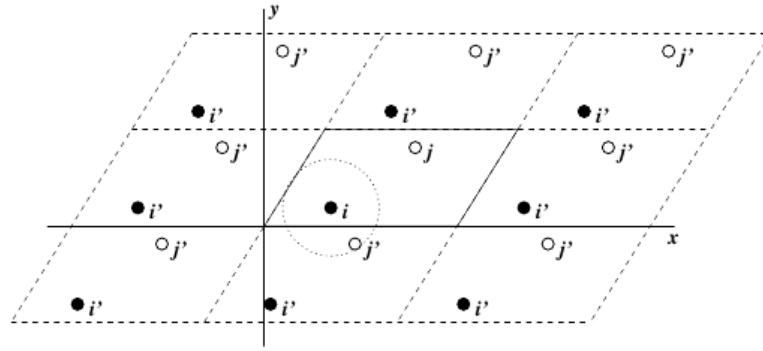


Figura 9: Condições periódicas de contorno de um sistema em duas dimensões. Modificado de Spoel *et al.* (2005)

As condições periódicas de contorno resultam em interações com um número infinito de vizinhos para cada uma das partículas do sistema. A energia potencial total de um sistema periódico em uma caixa cúbica com arestas de tamanho L é dado por

$$V_{tot} = \frac{1}{2} \sum_n \sum_{i,j} v_{eff}(r_{i,j} + nL) \quad (12)$$

onde o fator $\frac{1}{2}$ certifica que cada par é contato apenas uma vez, n é um vetor numérico, e o símbolo sobre a segunda somatória é um lembrete de que, quando $n=0$, o termo $i=j$ não deve ser contado e $r_{i,j} = \|r_i - r_j\|$. Para interações de curto alcance, é utilizado uma aproximação para restringir o número de interações através da aplicação de um raio de corte, r_c , em torno de cada partícula. Para interações de longo alcance, as imagens periódicas devem se levadas em consideração. Conforme a equação 12, para as interações de curto alcance é necessário apenas incluir o termo $n=0$. O raio de corte, r_c , define uma esfera ao redor de cada partícula e, por convenção, ela deve ter o tamanho máximo da caixa computacional. Em geral, se L é o tamanho da aresta da caixa de simulação (no caso de um cubo), $r_c < L/2$. Esses parâmetros evitam que a partícula interaja com ela mesma. (HEDMAN, 2006)

As partículas presentes na caixa computacional central é envolvida por imagens de partículas presentes em cada réplica periódica da caixa central. As partículas das imagens se movem exatamente igual à caixa central. Quando uma partícula foge do limite da caixa central durante o curso da simulação, a sua imagem periódica reaparece no lado oposto da caixa computacional central. (HEDMAN, 2006)

A distância inter-partícula utilizada na simulação é calculada através da convenção da “mínima imagem”. Ela assume que o tamanho da caixa de simulação é $L > 2r_c$, assegurando que todas as partículas j que interagem com a partícula i sejam determinadas apenas uma vez. A distância entre duas partículas i e j é a menor de todas as distâncias possíveis entre elas, incluindo todas as réplicas de imagens da partícula j . Como por exemplo, uma caixa cúbica com arestas de tamanho L restringe as coordenadas cartesianas em cada dimensão no intervalo $[-L/2, L/2]$ e, conseqüentemente, a diferença de cada coordenada, $\Delta_{i,j,\alpha}, \alpha = \{x, y, z\}$, que está no intervalo $[0, L]$. Podemos encontrar a mínima distância em cada coordenada através do menor valor absoluto de três valores mínimos possíveis $\{\Delta_{i,j,\alpha} - L, \Delta_{i,j,\alpha}, \Delta_{i,j,\alpha} + L\}, \alpha = \{x, y, z\}$. No caso de caixas computacionais diferentes, o critério utilizado no cálculo de distância da imagem mínima também será diferente. (HEDMAN, 2006)

Utilizando o critério de distância da imagem mínima, podemos assegurar que a distância entre duas partículas variará continuamente conforme o movimento de saída e reaparecimento das partículas na caixa computacional. Além do mais, as condições periódicas de contorno são importantes na restrição de flutuações não físicas da densidade. (HEDMAN, 2006)

2.4.3. CAMPOS DE FORÇA EM SISTEMAS ATÔMICOS

As interações inter-atômicas são fundamentais na simulação por MD. Do ponto de vista físico, todas as contribuições são importantes e derivadas de interações eletrônicas entre o núcleo e as nuvens eletrônicas (STONE, 1986). Algumas dessas contribuições são clássicas, como as interações de Coulomb, enquanto outras, como a dispersão, requerem uma explicação ao nível quântico. Esses modelos de interações possuem partes atrativas e repulsivas, e são normalmente aditivas. No geral, a energia potencial total do sistema é normalmente descrita como a soma de n termos, $n = 1, 2, 3, \dots$

$$V(r_1, r_2, \dots, r_N) = \sum_i v_1(r_i) + \sum_{i,j} v_2(r_i, r_j) + \sum_{i,j,k} v_3(r_i, r_j, r_k) + \dots \quad (13)$$

O primeiro termo do lado esquerdo representado na equação 13, descreve o efeito de um campo externo no sistema (campos elétricos ou magnéticos). O segundo termo é denominado potencial de par. Ele é somado entre todos os pares distintos de partículas. O terceiro termo é o potencial de 3 corpos e deve ser somado entre todas as triplas distintas. Termos de ordens maiores possuem valores mínimos comparados aos termos de 2 ou 3 corpos, e são conseqüentemente excluídos.

Na vasta maioria das simulações por MD, o cálculo da energia total é simplificado através de uma adição aditiva entre potenciais de pares. Em simulações que contém moléculas flexíveis, é comum adicionarmos termos que representam as ligações químicas, ângulos das ligações, diedros próprios e impróprios. Interações entre os átomos das moléculas são representados por potenciais aditivos. Essa aproximação empírica separa a energia potencial total do sistema em partes ligadas (*bonded*) e não ligadas (*non-bonded*).

Para um sistema isolado não há influências externas e, portanto, o primeiro termo da equação 13 é zero. O potencial de par é uma função da distância entre os pares $r_{i,j} = \|r_i - r_j\|$, e é construído de forma a incluir o verdadeiro efeito do potencial de pares e médias de termos de maior ordem; normalmente, também é incluído os efeitos eletrostáticos e dipolares. A energia potencial total do sistema é a soma do potencial de todos os pares distintos das partículas:

$$V_{eff} = \frac{1}{2} \sum_{i,j} v_{eff}(r_{i,j}) \quad (14)$$

As interações podem ser descritas utilizando diferentes atributos, mas através do ponto de vista computacional, é mais importante as forças de longo e curto alcance. Como exemplo de interações de curto alcance, temos o potencial de Lennard-Jones (eq. 15, figura 10-A), e como exemplo de interações de longo alcance, temos o potencial de Coulomb (eq. 16, figura 10-B).

$$V_{LJ}(r_{i,j}) = 4 \varepsilon_{i,j} \left(\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right) \quad \begin{array}{l} \sigma \text{ é raio aproximado do átomo e } \varepsilon \\ \text{é a profundidade do potencial.} \end{array} \quad (15)$$

$$V_c(r_{i,j}) = f \frac{q_i q_j}{\varepsilon_r r_{i,j}} \quad \begin{array}{l} q \text{ é a carga do átomo e } \varepsilon \text{ é a} \\ \text{constante dielétrica do meio.} \end{array} \quad (16)$$

Na MD, é comum se descrever uma interação como de curto alcance quando o seu potencial decai mais rápido que r^{-d} , onde d é a dimensionalidade do sistema. A contribuição para a energia potencial total de todas as partículas fora do raio de corte é bem definida. Como exemplo, em 3 dimensões, essa contribuição é proporcional à

$$\int_{r_c}^{inf} V(r)4\pi r^2 dr, \quad (17)$$

que é bem definida quando $V(r)$ decai mais rápido que r^{-3} .

Em interações de longo e curto alcance, somos induzidos a utilizar um simples raio de corte, mas essa prática pode criar efeitos não físicos no espaço do raio de corte da esfera. Para alguns sistemas, métodos de raios de cortes mais avançados, baseados no agrupamento de cargas, podem ser uma solução aceitável (STEINBACH & BROOKS, 1994). Porém, há um número crescente de casos em que as interações de longo alcance são bastante importantes, e devem ser consideradas com os devidos cuidados, especialmente em sistemas iônicos (ANÉZO *et al.*, 2003; PATRA *et al.*, 2003). Conseqüentemente, um número de aproximações alternativas tem sido desenvolvidas. Como exemplos, temos os métodos *lattice sum* (BATCHO & SCHLICK, 2001), *reaction field* (TIRONI *et al.*, 1995), *isotropic periodic sum* (WU & BROOKS, 2005), e métodos hierárquicos (PFALZNER & GIBBON, 1996 *apud* HEDMAN, 2006), como o *multigrid* (IZAGUIRRE, HAMPTON & MATTHEY, 2005) e o *fast multipole* (GREENGARD & ROKHLIN, 1987).

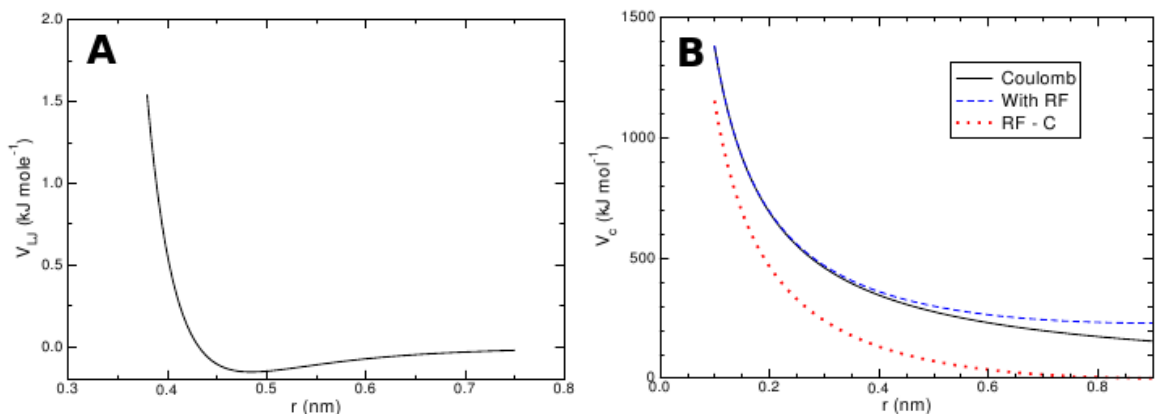


Figura 10: Forma do potencial de interação de Lennard Jones (A) e de Coulomb (B) utilizando diferentes parâmetros. Modificado de Spoel *et al.* (2005)

A diferença essencial entre essas aproximações é a forma no qual as interações após o raio de corte são tratadas. Nos métodos de corte de raio, as interações com as partículas presentes após o raio de corte possuem valor zero, possivelmente com a aplicação de uma constante de correção de longo alcance. No método *reaction field*, uma constante dielétrica é assumida após o raio de corte, e as interações de longo alcance são substituídas por interações *reaction field* (equação 18). No método *lattice sum*, as interações de longo alcance são calculadas utilizando a soma de todas as imagens da caixa de simulação. Além do mais, esses métodos consomem tempo e custos computacionais diferentes: o método de corte de raio é o mais barato, mas normalmente introduz problemas à simulação; o método *reaction field* possui um equilíbrio entre o tempo computacional e a acurácia, e o método *lattice sum*, o mais caro, é o que fornece a maior acurácia.

$$V_{erf} = f \frac{q_i q_j}{\epsilon_r r_{i,j}} \left[1 + \frac{\epsilon_{rf} - \epsilon_r}{2 \epsilon_{rf} + \epsilon_r} \frac{r_{i,j}^3}{r_c^3} \right] - f \frac{q_i q_j}{\epsilon_r r_c} \frac{3 \epsilon_{rf}}{2 \epsilon_{rf} + \epsilon_r} \quad (18)$$

ϵ_r e ϵ_{rf} são constantes dielétricas utilizadas antes e depois do raio de corte, respectivamente.

O método *lattice sum* mais prominente é o método de somatória Ewald. Esse método é referência para interações eletrostáticas em sistemas modelados com condições periódicas de contorno. Mesmo apesar do método de Ewald evitar o truncamento de interações de longo alcance, ainda há problemas dependentes do sistema, provenientes das condições periódicas de contorno (WU & BROOKS, 2005).

No método de Ewald, a energia eletrostática total de N partículas é representada por:

$$V = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_i^N \sum_j^N \frac{q_i q_j}{r_{i,j,n}} \quad (19)$$

onde $(n_x, n_y, n_z) = \mathbf{n}$ é o índice do vetor da caixa, e o asterisco indica que o termo $i=j$ deve ser omitido quando $(n_x, n_y, n_z) = (0, 0, 0)$. A distância $r_{i,j,n}$ é a distância real entre as cargas, e q é a carga da partícula. (SPOEL *et al.*, 2005).

2.4.3.1. Interações de Ligações

As interações de ligações são baseadas em uma lista fixa de átomos. Elas não são interações exclusivamente de pares, mas também incluem interações com 3 e 4 corpos. Existem as interações de distância de ligação (2 corpos), ângulo de ligação (3 corpos) e ângulo de diedro (4 corpos). Existem diversos modelos de potenciais para interações de ligações. O potencial mais comum é o potencial harmônico, o qual evolui exponencialmente em função da distância, ângulo ou diedro apresentado pelas partículas.

O potencial harmônico de distância de ligação é descrito pela distância da ligação entre os átomos i e j ligados covalentemente, e é representado por um potencial harmônico (equação 20), que cresce exponencialmente em função da distância entre as partículas (figura 11).

$$V_b(r_{i,j}) = \frac{1}{2} k_{i,j}^b (r_{i,j} - b_{i,j})^2 \quad (20)$$

b é a distância teórica ideal, r é a distância real, e k é a constante de força

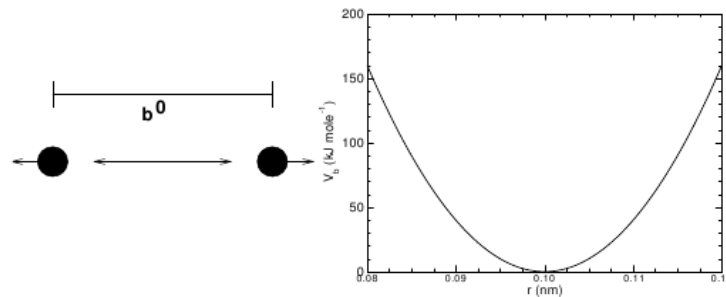


Figura 11: Princípio da distância de ligação (esquerda), e o potencial de distância de ligação (direita). Modificado de Spoel *et al.* (2005)

Temos também o potencial harmônico de ângulo de ligação, que é descrito pelo ângulo de ligação entre uma tripla de átomos $i-j-k$, e é também representado por um potencial harmônico sobre o ângulo entre ijk (equação 21; figura 12).

$$V_b(\theta_{i,j,k}) = \frac{1}{2} k_{i,j,k}^\theta (\theta_{i,j,k} - \theta_{i,j,k}^0)^2 \quad (21)$$

Θ^0 é o ângulo teórico ideal, Θ é ângulo real, e k é a constante de força

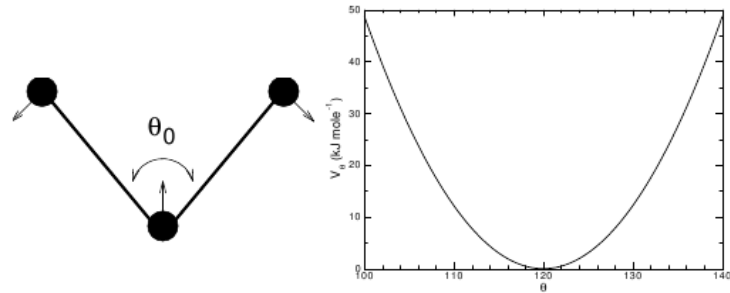


Figura 12: Os átomos i , j e k estão na seqüência das ligações covalentes. O átomo j está no meio, e os átomos i e k estão nas extremidades. Modificado de Spoel *et al.* (2005)

O diedro impróprio, um dos dois tipos de diedros, mantêm os grupos planares em planos (ex.: anéis aromáticos) ou previne as moléculas de alterarem a sua conformação quiral (equação 22, figuras 13 e 14).

$$V_{id}(\xi_{i,j,k,l}) = \frac{1}{2} k_{\xi} (\xi_{i,j,k,l} - \xi_0)^2 \quad (22)$$

ξ_0 é o ângulo diedral teórico ideal, ξ é ângulo diedral real, e k é a constante de força

Visto que esse potencial é harmônico, a periodicidade não é levada em consideração e, portanto, o diedro impróprio é melhor definido por um ângulo central de 0° que pode se estender até $\pm 180^\circ$ graus.

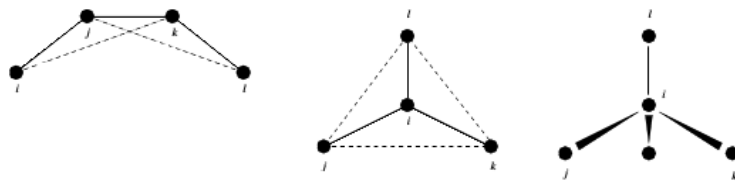


Figura 13: Princípio do ângulo de diedro impróprio. O ângulo de diedro impróprio é definido pelo ângulo entre os planos (i, j, k) e (j, k, l) . Modificado de Spoel *et al.* (2005)

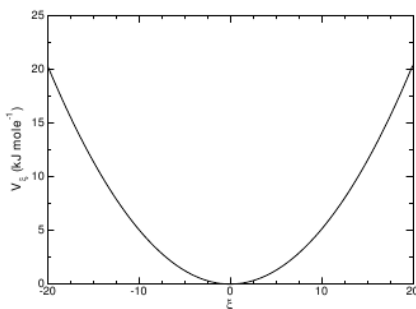


Figura 14: Potencial de diedro impróprio. Modificado de Spoel *et al.* (2005)

O outro tipo de diedro, o diedro próprio, é definido pelo ângulo existente entre os planos ijk e jkl , sendo que zero corresponde à configuração *cis*, o que significa que l e i estão do mesmo lado do plano (equação 23, figura 15).

$$V_d(\phi_{i,j,k,l}) = k_\phi (1 + \cos(n\phi - \phi_s)) \quad (23)$$

Φ_s é o ângulo diedral teórico ideal, Φ é ângulo diedral real, e k é a constante de força

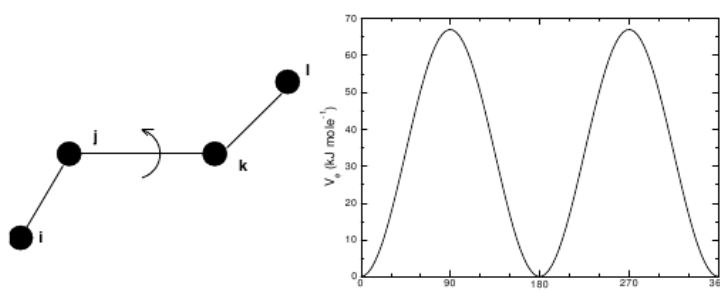


Figura 15: Princípio do ângulo de diedro próprio (esquerda) e o potencial de ângulo de diedro (direita). Modificado de Spoel *et al.* (2005).

2.4.4. OTIMIZAÇÃO DA GEOMETRIA MOLECULAR

A Otimização da Geometria Molecular ou Minimização de Energia é uma técnica que visa encontrar coordenadas para as partículas da configuração inicial da simulação com uma baixa energia potencial. O procedimento básico consiste em caminhar sobre a superfície de potencial na direção em que a energia decresce, levando o sistema à um mínimo de energia local. A minimização da energia cobre apenas uma pequena parte do espaço de configurações, porém, ela cria ajustes nas posições atômicas, relaxando distorções nas ligações químicas, nos ângulos entre ligações e nos contatos de van der Waals. A eliminação de deformações é importante na preparação de um sistema molecular sobre o qual pretende-se aplicar a Dinâmica Molecular. Durante a dinâmica são constantes as conversões entre as energias potencial e cinética, de maneira que um sistema com energia potencial inicialmente elevada apresentará temperaturas irreais nos primeiros passos da dinâmica, as quais poderão aprisioná-lo em configurações indesejáveis do espaço de fase e, com frequência, abortar a simulação. A prévia minimização da energia favorece a termalização e evita regiões “quentes” em um sistema no início de sua dinâmica. (PASCUTTI, 2004)

2.4.4.1. “Steepest Descent”

O *Steepest Descent* (SD) não é o algoritmo mais eficiente na busca da energia mínima local, porém é um método robusto e fácil de ser implementado. O método SD pode ser representado por:

$$r_{n+1} = r_n + \frac{F_n}{\max(|F_n|)} h_n \quad (24)$$

onde h_n é o deslocamento máximo e F_n é a força, ou o gradiente negativo do potencial V . A notação $\max(|F_n|)$ significa o maior valor absoluto dos componentes de força. As forças e energia são computadas novamente para as novas posições. Se $(V_{n+1} < V_n)$ as novas posições são aceitas e $h_{n+1} = 1.2 h_n$. Se $(V_{n+1} > V_n)$ as novas posições são rejeitadas e $h_{n+1} = 0.2 h_n$. O algoritmo é finalizado quando alcança um número limite de ciclos especificado pelo usuário, ou quando valor máximo absoluto dos componentes de força é menor que um valor especificado (SPOEL *et al.*, 2005). Resumindo esse processo, podemos dizer que quando o deslocamento da partícula encontra uma energia menor, o passo do deslocamento será redimensionado para mais, alcançando rapidamente coordenadas próximas ao mínimo, onde a partícula encontrará restrições de movimento em função da energia, e diminuirá gradativamente o passo de deslocamento até encontrar o ponto mínimo da energia local (figura 16).

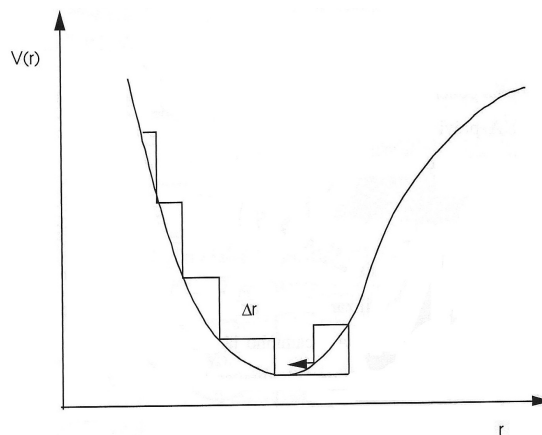


Figura 16: Representação do caminho do algoritmo *Steepest-Descent* em uma dimensão. Modificado de Pascutti (2004).

2.4.4.2. “Conjugate Gradient”

O método *Conjugated Gradient*, ou gradiente conjugado (CG), é mais lento que o SD nos primeiros estágios da minimização, mas se torna mais eficiente quando próximo à energia mínima, e possui parâmetros e critérios de parada iguais ao do SD. O CG é mais sofisticado que o SD, e leva em consideração o caminho já percorrido na busca do mínimo, ao contrário do SD, que utiliza a informação do gradiente apenas na coordenada atualizada (PASCUTTI, 2004). O método pode ser representado por:

$$r_{i,n+1} = r_{i,n} + \Delta r_{i,n} \quad (25)$$

onde,

$$\Delta r_{i,n} = \alpha_n \delta_{i,n} \quad e, \quad (26)$$

$$\delta_{i,n} = (F_{i,n} / |F_{i,n}|) + \delta_{i,n-1} (|F_{i,n}^2| / |F_{i,n-1}^2|) \quad (27)$$

onde α_n é um parâmetro ajustável a cada passo. O valor ideal de α_n é obtido por um processo de minimização do potencial na direção definida pelo vetor $\delta_{i,n}$. A vantagem da utilização do gradiente conjugado vem do fato de que a direção do gradiente no novo ponto será sempre ortogonal ao gradiente do ponto anterior e conseqüentemente a todos os anteriores. Essa propriedade determina um caminho mais direto ao fundo do poço de potencial (figura 17), evitando o retorno sobre caminhos já percorridos, ao contrário do algoritmo SD que utiliza somente a informação em um ponto. (PASCUTTI, 2004)

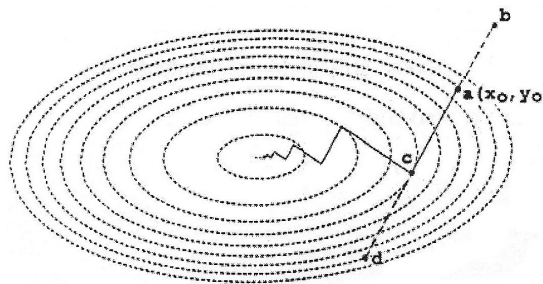


Figura 17: Representação do caminho do algoritmo dos gradientes conjugados em duas dimensões. Modificado de Pascutti (2004).

2.4.5. “DOCKING” MOLECULAR

Os métodos de *docking* molecular são uma extensão dos métodos de simulação molecular por utilizarem os mesmos modelos de potenciais. Os métodos de *docking* se baseiam principalmente nos métodos estocásticos de simulação, onde os novos estados são aceitos com base em probabilidades de distribuição, como a de Boltzmann, a mais comum e antiga, e a de Tsallis, que foi desenvolvida mais recentemente. O objetivo do *docking* molecular é induzir as partículas de um sistema computacional para um estado de menor energia local ou global. Do ponto de vista biológico, o estado de baixa energia pode representar a forma estrutural de como as moléculas biológicas interagem entre si, sendo de grande importância no estudo das interações moleculares, as quais, por exemplo, podem ser de grande valor no aprimoramento e desenvolvimento de novos fármacos.

O método *Generalized Simulated Annealing* (GSA), utilizado nesse trabalho, é um híbrido do método de Monte Carlo (MC) e do *Simulated Annealing* (SA). O GSA se baseia no efeito da temperatura, utilizando, inicialmente, temperaturas elevadas, as quais ampliam a escala de probabilidades para novos estados conformacionais, permitindo uma busca global e mais eficiente pelas regiões de menor energia (figura 18A). Essa temperatura é gradativamente minimizada, o que diminui a probabilidade de novas conformações, limitando a busca para um escala local. Métodos que utilizam a variação de temperatura na busca de estados de baixas energia são normalmente derivados do método *Simulated Annealing*, e métodos que calculam novos estados moleculares baseados em probabilidades são normalmente derivados dos métodos Monte Carlo e Metropolis, que é conhecido por utilizar a probabilidade de distribuição de Boltzmann (METROPOLIS *et al.*, 1953). O GSA, por outro lado, utiliza a probabilidade de distribuição de Tsallis (MORET, 1998).

2.4.5.1. Método de Monte Carlo

O método Monte Carlo (MC) é uma aproximação estocástica, e explora a energia do sistema através de testes aleatórios na geometria molecular. Matematicamente, ele especifica as coordenadas dos átomos X_O^N ; e gera novas coordenadas X_O^N aleatórias. Em seguida, é computado a probabilidade da transição $W(0, a)$, e é gerado um número aleatório R na escala $[0, 1]$. Se $W(0, a) < R$, as antigas coordenadas são mantidas. Por outro lado, se

$W(0, a) \geq R$, as novas coordenadas são aceitas. O ciclo, então, retorna para o início, sorteando novas coordenadas. Nesse método, as interações são independentes uma da outra, ou seja, o sistema não possui memória. A possibilidade do sistema reverter para o seu estado anterior é possível, assim como a escolha de um estado diferente. Como resultado da simulação estocástica, um grande número de configurações são acumuladas, e a energia potencial é calculada, fornecendo propriedades termodinâmicas do sistema. (LABANOWSKI, 1996)

2.4.5.2. Método “Generalized Simulated Annealing”

No método *Simulated Annealing*, o método MC é utilizado para gerar novos estados aleatórios, e aceita ou os rejeita de acordo com o critério de Metropolis. Esse método, inicialmente, especifica as coordenadas iniciais dos átomos, e seleciona determinados átomos i , movendo-os aleatoriamente $(\Delta X_1, Y_1, Z_1)$. Em seguida, a energia potencial do sistema ΔV é calculada entre o novo e antigo estado. Se $\Delta V < 0$, as novas coordenadas são aceitas. Por outro lado, é escolhido um número aleatório na escala $[0, 1]$ e, se $e^{-\Delta V/kT} < R$, as novas coordenadas são aceitas, ou, se $e^{-\Delta V/kT} \geq R$, as coordenadas originais são mantidas. A expressão $p = e^{-\Delta V/kT}$ é a probabilidade de distribuição de Boltzmann, onde T é a temperatura da simulação e k é a constante de Boltzmann. Em altas temperaturas, vários estados são aceitos, enquanto em baixas temperaturas, a probabilidade da maioria dos movimentos são rejeitadas. A idéia é sempre aceitar uma melhor configuração, e a tendência do sistema será sempre converger para uma energia mínima local, porém, quando configurações não ideais são aceitas, o sistema poderá superar a configuração mínima local e, talvez, atingir mínimos próximos ao global (figura 18-C). O SA, diferentemente do MC (figura 18-B), é caracterizado por utilizar uma temperatura inicial elevada e evoluir gradativamente para uma temperatura final baixa, encontrando mínimos locais, objetivo para o qual ele foi exclusivamente desenvolvido .

O GSA é idêntico ao SA, exceto por utilizar a probabilidade de distribuição baseada na entropia generalizada de Tsallis:

$$S_q = k_B \frac{\sum_i p_i^q}{q-1} \quad (28)$$

onde k_b é a constante de Boltzmann; e p_i é a probabilidade da configuração i . No limite $q \rightarrow 1$, a expressão de Boltzmann para a entropia é recuperada. A probabilidade de aceitação do novo estado é dado por:

$$P_{q_A}(\vec{x}_i \rightarrow \vec{x}_{i+1}) \begin{cases} 1 & \text{if } (E(\vec{x}_{i+1}) < E(\vec{x}_i)) \\ \frac{1}{\left[1 + (q_A - 1)(E(\vec{x}_{i+1}) - E(\vec{x}_i))/T_{q_T}^V\right]^{\frac{1}{q_A-1}}} & \text{if } (E(\vec{x}_{i+1}) \geq E(\vec{x}_i)) \end{cases} \quad (29)$$

onde a temperatura decai conforme

$$T_{q_T}^V = T_{q_T}(t) \frac{2^{q_T-1}}{(1-t)^{q_T-1} - 1} \quad (30)$$

onde t é o passo de “tempo” markhovian; x_t é a configuração no “tempo” t ; $E(x_t)$ é a energia da configuração x_t . A superfície de energia é governada pela distribuição:

$$g_{q_V}(\Delta\phi_t) = \left(\frac{q_V-1}{\Pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{1-\frac{q_V-1}{2}}{q_V-1}\right) \left(T_{q_T}(t)\right)^{\frac{1}{3-q_V}}}{\Gamma\left(\frac{1}{q_V-1} - \frac{1}{2}\right) \left[1 + (q_V-1) \frac{(\Delta\phi_t)}{\left(T_{q_T}(t)\right)^{\frac{1}{3-q_V}}}\right]^{\frac{1}{q_V-1} - \frac{1}{2}}} \quad (31)$$

onde Γ é a função gama; $\Delta\phi_t$ é a mudança de conformação; q_V , q_T , e q_A são parâmetros que devem ser otimizados antes do *docking*. O processo de geração de configurações é repetido até o sistema alcançar o mínimo global.

A estatística de distribuição deslocalizada de Tsallis provê movimentos de longo

alcance entre configurações espaciais distantes, porém, termodinamicamente significantes. Esses movimentos longos são aleatoriamente entrelaçados com movimentos de curto alcance, provendo boas amostras locais. A estatística de Tsallis permite atingir o mínimo global de energia de um sistema mais eficientemente que outros métodos estocásticos convencionais, e não serve para determinar propriedades de equilíbrio do sistema. (STRAUB & ANDRICIOAEI, 1999)

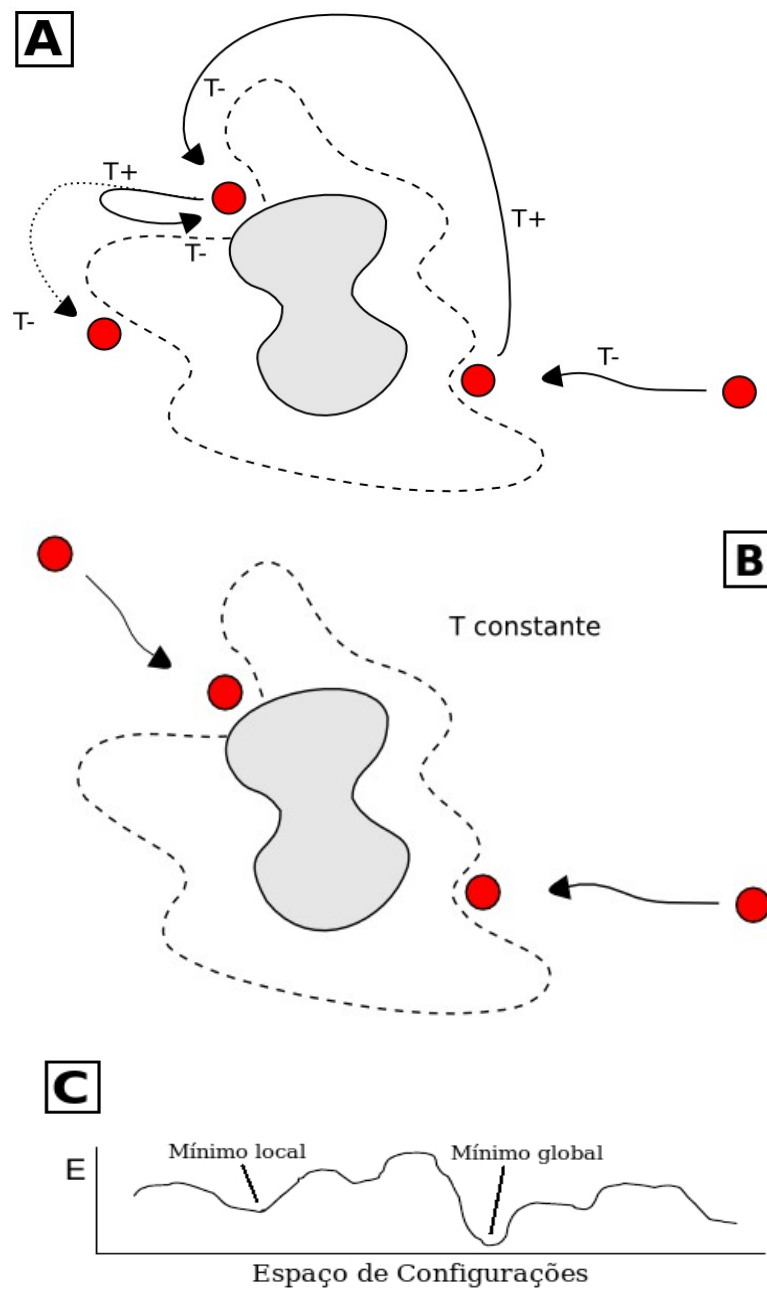


Figura 18: (A) Representação gráfica do método de *docking* GSA; a linha tracejada representa a dimensão da energia de contato. (B) Representação de *docking* com temperatura constante. (C) Espaço de configurações e as respectivas energias (E).

2.4.6. ACESSIBILIDADE AO SOLVENTE

A área de superfície acessível (ASA) é a superfície de uma biomolécula (proteína, DNA, etc.) acessível ao solvente. A método ASA foi primeiramente descrito por Lee & Richards (1971) e é algumas vezes denominada de superfície molecular de Lee-Richards. Ela é comumente calculada usando o algoritmo *rolling-ball* desenvolvido por Shrake & Rupley (1973). Esse algoritmo utiliza uma esfera de um determinado raio, que representa o solvente, e é utilizada para verificar a superfície da molécula, de forma semelhante ao apresentado na figura 19. A escolha do raio da esfera altera a área de superfície observada. Quando são utilizadas esferas de raios menores, a superfície será detectada com maiores detalhes e, apresentará, conseqüentemente, uma maior superfície.

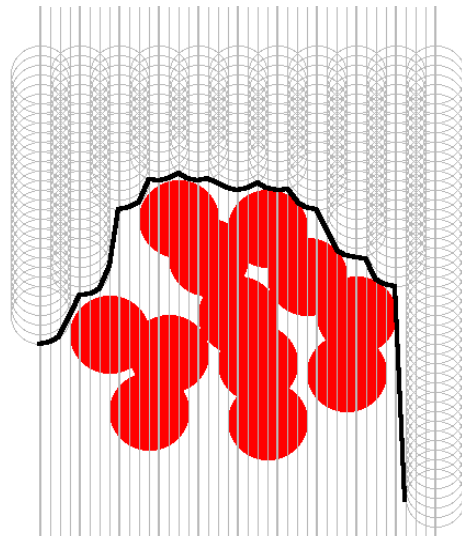


Figura 19: Esferas analisando a área de superfície acessível de um grupo de hemácias em duas dimensões através do método *rolling-ball*.

Um raio típico possui aproximadamente 1.4 \AA , que é o raio aproximado da molécula de água. Outro fator que afeta os resultados é o valor do raio de van der Waals dos átomos da molécula estudada. Como por exemplo, átomos de hidrogênio representados de forma implícita na molécula, necessitam de alterações no respectivo raio.

O ASA é semelhante ao conceito do método da superfície do solvente excluído (*solvent-excluded surface*, também conhecido como superfície molecular ou superfície de Connolly), que pode ser imaginado como uma bolha no solvente (figura 20). Esse conceito também pode ser calculado através do algoritmo *rolling-ball* desenvolvido por Richards (1977), que foi implementado em três dimensões por Connolly (1983) e Richmond (1984).

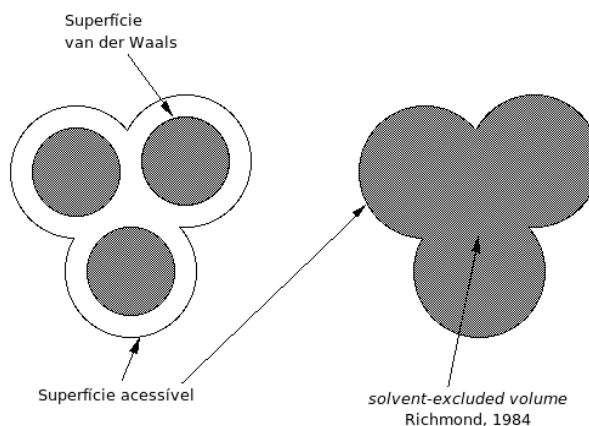


Figura 20: Conceito da área de superfície acessível e da superfície de solvente excluído.

No estudo de proteínas, a acessibilidade pode ser calculada para a cadeia lateral dos resíduos, e sua acessibilidade pode ser calculada de forma absoluta ou relativa, utilizando como referência os demais resíduos da proteína. A acessibilidade pode ser um bom parâmetro na análise de alterações em proteínas ao longo de uma dinâmica molecular, e pode indicar importantes alterações estruturais relacionadas à atividade da proteína.

2.4.7. DESVIO DA RAIZ MÉDIA QUADRÁTICA

O desvio da raiz média quadrática (RMSD) da estrutura de uma molécula pode ser calculado através da comparação com a estrutura de referência ($t=0$ ps). O RMSD é calculado através da fórmula:

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (32)$$

onde $M = \sum (N, i=1)$ e $r_i(t)$ é a posição do átomo i no tempo t . Ao invés de comparar todas as estruturas à uma única estrutura (ex.: tempo $t = 0$), pode-se utilizar como referência a estrutura do tempo $t_2 = t_1 - T$. Essa representação, em função de T , pode trazer vantagens na visualização da mobilidade da estrutura. Esse método é limitado, pois não é capaz de especificar tais variações. Portanto, ele é útil e aplicado, normalmente, na verificação da estabilidade da conformação geral de uma molécula em função do tempo.

2.4.8. DISTÂNCIA ENTRE ÁTOMOS OU GRUPOS DE ÁTOMOS

A distância entre dois centros de massas de átomos ou grupos de átomos em função do tempo é calculada através da equação da hipotenusa de Pitágoras para três dimensões:

$$d(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \quad (33)$$

onde d é a distância entre A e B; e x , y e z são as coordenadas tridimensionais de A e B.

2.4.9. LIGAÇÕES DE HIDROGÊNIO

A ligação de hidrogênio é o resultado de uma força dipolo-dipolo entre um átomo eletronegativo e um átomo de hidrogênio ligado a um átomo também eletronegativo (ex.: O, N, Cl e F; ex.: figura 21). A energia de uma ligação de hidrogênio (normalmente entre 5 e 30 kJ/mol) é comparável às ligações covalentes fracas (155 kJ/mol). Essas ligações podem ocorrer entre moléculas (intermolecular), ou entre diferentes partes de uma única molécula (intramolecular). A ligação de hidrogênio é uma ligação bastante forte do tipo van der Waals-Keesom, mas é considerada mais fraca que as ligações covalentes, iônicas e metálicas. Esse tipo de ligação ocorre em moléculas orgânicas (ex.: DNA) e inorgânicas (ex.: água).

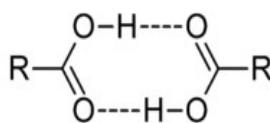


Figura 21: Exemplo de ligações de hidrogênio.

As ligações de hidrogênio são ligações de origem eletrostática, e são fortes e orientadas. Elas podem ocorrer em outros átomos eletronegativos (ex.: O, N, Cl e F) e a presença do hidrogênio é obrigatória. Defini-se como “doador” (D) o átomo eletronegativo que tem o átomo H ligado a ele, e “aceptor” (A) o átomo eletronegativo que atrai o átomo H.

Por ser orientada e possuir uma distância flexível, critérios devem ser utilizados para definir a existência de uma ponte de hidrogênio. Um critério bastante comum utiliza uma distância $\leq 0,35 \text{ nm}$ e um ângulo $\alpha \leq 60^\circ$.

2.4.10. ESTRUTURA SECUNDÁRIA

Em uma cadeia polipeptídica, as únicas rotações permitidas no esqueleto ou *backbone* ocorrem nas ligações $N-C_{\alpha}$ e $C_{\alpha}-C$. O esqueleto de uma cadeia polipeptídica pode ser vista como uma série de planos rígidos com planos consecutivos compartilhando um ponto comum de rotação em C_{α} . As ligações peptídicas rígidas limitam o número de conformações que podem ser assumidas por uma cadeia polipeptídica. (LEHNINGER, NELSON & COX, 2000)

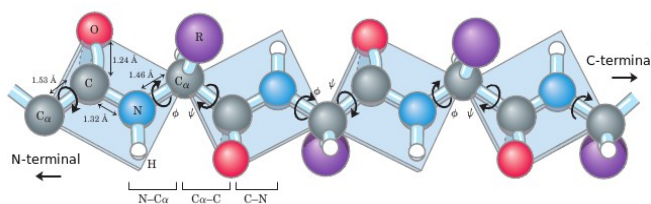


Figura 22: Exemplo de cadeia polipeptídica. Modificado de Lehninger, Nelson & Cox (2000).

Por convenção, as rotações das ligações $N-C_{\alpha}$ são denominadas de ϕ (phi), e as rotações das ligações $C_{\alpha}-C$ são denominadas de ψ (psi) (figura 22). Por convenção, novamente, ambas são definidas como 180° quando o polipeptídeo está em sua conformação estendida e todos os grupos peptídicos estão no mesmo plano. A combinação entre diferentes ângulos ϕ e ψ pode definir a estrutura secundária de um peptídeo (figura 23). (LEHNINGER, NELSON & COX, 2000)

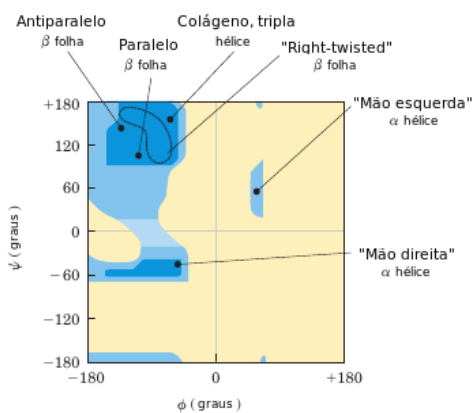


Figura 23: Gráfico Ramachandran mostrando os possíveis ângulos 'phi' e 'psi' e as respectivas estruturas secundárias. Modificado de Lehninger, Nelson & Cox (2000).

2.4.11. RAI0 DE GIRAÇÃO E MOD_GYR

O método de raio de giração é bastante simples e útil na análise da distribuição de massa de uma molécula ou de um objeto qualquer, assim como na estimacão de sua dimensão. O raio de giração é representado por:

$$Rg = \sum_{k=1}^N \sqrt{\frac{m(r_k)^2}{M}} \quad (34)$$

onde m é a massa do partícula k , r_k é a distância da partícula k do centro de gravidade do objeto, e M é a massa total do objeto. N Representa o número total de partículas do objeto, e a somatória é feita para todos as partículas de 1 à N (LANG, 1975). No caso, as partículas podem ser átomos, resíduos, grupos de átomos etc. Com base na equação 34, fica claro que, quanto maior a distância das partículas do centro de gravidade, maior será o raio de giração.

Baseando-se na teoria do raio de giração, foi desenvolvido, nesse trabalho, o método MOD_GYR, o qual o nome é a abreviacão de “MODificacão” e “GYRate” ou, em outras palavras, “modificacão do método de raio de giração”. O MOD_GYR é representado por:

$$MG_i = \sum_{j=1}^N \frac{R_{ij}^{2+n}}{2N} \quad (35)$$

onde R_{ij} é a distância entre as partículas i e j , N é o número total de partículas, e n é o grau de sensibilidade da análise. No caso, as partículas também podem ser átomos, resíduos ou grupos de átomos. Nesse caso, a grande diferença é que o método não utiliza parâmetros de massa, apenas posições geométricas. Outra grande diferença é que o MOD_GYR é calculado em função de todas as partículas, ou seja, a partícula i será o centro do objeto, diferentemente do raio de giração que utiliza um centro de gravidade fixo.

Teoricamente, através do MOD_GYR, é possível analisar a posição de um átomo ou resíduo em relação aos demais semelhantes, ou seja, por exemplo, é possível verificar se um resíduo se encontra no interior, exterior, distante ou próximo do centro geométrico de uma proteína (detalhes, figura 24). Da mesma forma, é possível analisar a ocupação espacial de

uma molécula, assim como no método de raio de giração original. Por todas essas qualidades, o método MOD_GYR pode ser utilizado em análises complexas, como por exemplo, na interação entre moléculas para a verificação de importantes alterações estruturais.

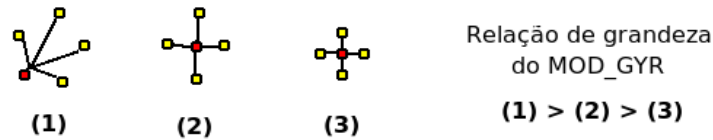


Figura 24: Exemplo simplificado da aplicação do método MOD_GYR. A partícula em vermelho foi utilizada como centro.

2.4.12. SUPERFÍCIE ELETROSTÁTICA

O método da superfície eletrostática é composto por dois métodos distintos, onde um deles é o método ASA (superfície acessível), que descreve espacialmente a superfície molecular em si, e o outro método descreve a influência eletrostática de todas as cargas da molécula sobre um determinado ponto presente na superfície molecular anteriormente calculada (figura 25). Com maiores detalhes, podemos dizer que a superfície eletrostática é a soma do campo elétrico de Coulomb de cada partícula eletricamente carregada sobre os diversos pontos da superfície molecular. O campo elétrico de Coulomb de cada partícula é dado por:

$$E = \frac{Q}{4 \pi \epsilon_0 r^2} \quad (36)$$

onde Q é a carga da partícula, r é a distância da partícula à um ponto específico da superfície molecular, e ϵ_0 é a constante dielétrica do meio.

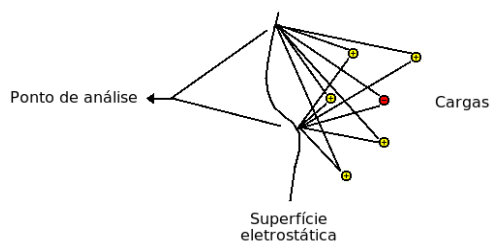


Figura 25: Representação gráfica do método de análise da superfície eletrostática.

3. OBJETIVOS

A catepsina B pode ser encontrada na matriz extracelular, onde ela pode ser facilmente inativada pelo pH alcalino do meio. Porém, na presença de GAGs, ela se torna resistente à inativação alcalina e apresenta como característica principal um aumento no conteúdo da estrutura α -hélice. Os GAGs são encontrados abundantemente na matriz extracelular, a qual, na presença da catepsina B, pode ser degradada, facilitando o processo de metástase tumoral. Os mecanismos envolvidos nesse processo são ainda desconhecidos e necessitam uma averiguação melhor detalhada. Nesse trabalho foi caracterizado a interação do modelo de GAG, heparina, com as cisteíno-proteases papaína e catepsina B através de métodos computacionais, como a Dinâmica Molecular e o novo *Generalized Simulated Annealing*. A vantagem de utilizar a Dinâmica Molecular é que ela atualmente apresenta um elevado nível de desenvolvimento. Atráves de sua aplicação é possível simular e visualizar a evolução gradativa de importantes variações estruturais das moléculas em função do tempo, permitindo também estudar a influência da entropia do sistema sobre interações intermoleculares, que é um fator de grande importância no estudo de complexos moleculares, como o encontrado entre as cisteíno-proteases e os glicosaminoglicanos. No caso do *docking* molecular, o *Generalized Simulated Annealing* foi utilizado por ser um dos métodos mais eficientes na busca de mínimos locais e, conseqüentemente, na busca do mínimo global, que muitas vezes pode representar o verdadeiro sítio de interação. A combinação desses métodos é capaz de reproduzir de forma aproximada os resultados visualizados *in vitro*, que, por sua vez, foram utilizados para validar os resultados simulados. Portanto, como objetivo, foi analisado de forma detalhada os mecanismos e os padrões de interação entre a enzima e a heparina responsáveis por variações estruturais importantes, como o aumento do conteúdo da estrutura α -hélice, que é um dos resultados mais relevantes observados experimentalmente *in vitro*. A corroboração entre os dados simulados e experimentais pode implicar em um maior nível de confiança sobre a validação dos métodos utilizados, possibilitando o desenvolvimento de novas hipóteses, baseadas também em dados simulados, os quais podem contribuir para um melhor entendimento dos mecanismos pelos quais a papaína e a catepsina B se tornam resistentes à inativação alcalina.

4. MATERIAIS & MÉTODOS

4.1. MATERIAIS

Nesse trabalho foram utilizadas estruturas de raio-X da proteína catepsina B humana (seqüência, figura 27), papaína (seqüência, figura 26) e heparina. Todas as estruturas foram obtidas do *Protein Data Bank* (PDB) com os respectivos códigos de acesso: 1HUC (MUSIL *et al.*, 1991), 1BQI (LALONDE *et al.*, 1998) e 1HPN (MULLOY *et al.*, 1993). Detalhadamente, a seqüência de resíduos de açúcar da estrutura da heparina é GlcNS(2S)-IdoA(S)-GlcNS(2S)-IdoA(S); todos esses resíduos se encontram em uma conformação fixa do tipo “cadeira”.

1	ILE	31	VAL	61	TYR	91	VAL	121	LEU	151	GLY	181	TRP	211	LYS
2	PRO	32	VAL	62	GLY	92	GLN	122	LEU	152	PRO	182	GLY	212	ASN
3	GLU	33	THR	63	CYS	93	ARG	123	TYR	153	CYS	183	GLU		
4	TYR	34	ILE	64	ASN	94	TYR	124	SER	154	GLY	184	ASN		
5	VAL	35	GLU	65	GLY	95	CYS	125	ILE	155	ASN	185	GLY		
6	ASP	36	GLY	66	GLY	96	ARG	126	ALA	156	LYS	186	TYR		
7	TRP	37	ILE	67	TYR	97	SER	127	ASN	157	VAL	187	ILE		
8	ARG	38	ILE	68	PRO	98	ARG	128	GLN	158	ASP	188	ARG		
9	GLN	39	LYS	69	TRP	99	GLU	129	PRO	159	HIS	189	ILE		
10	LYS	40	ILE	70	SER	100	LYS	130	VAL	160	ALA	190	LYS		
11	GLY	41	ARG	71	ALA	101	GLY	131	SER	161	VAL	191	ARG		
12	ALA	42	THR	72	LEU	102	PRO	132	VAL	162	ALA	192	GLY		
13	VAL	43	GLY	73	GLN	103	TYR	133	VAL	163	ALA	193	THR		
14	THR	44	ASN	74	LEU	104	ALA	134	LEU	164	VAL	194	GLY		
15	PRO	45	LEU	75	VAL	105	ALA	135	GLN	165	GLY	195	ASN		
16	VAL	46	ASN	76	ALA	106	LYS	136	ALA	166	TYR	196	SER		
17	LYS	47	GLN	77	GLN	107	THR	137	ALA	167	GLY	197	TYR		
18	ASN	48	TYR	78	TYR	108	ASP	138	GLY	168	PRO	198	GLY		
19	GLN	49	SER	79	GLY	109	GLY	139	LYS	169	ASN	199	VAL		
20	GLY	50	GLU	80	ILE	110	VAL	140	ASP	170	TYR	200	CYS		
21	SER	51	GLN	81	HIS	111	ARG	141	PHE	171	ILE	201	GLY		
22	CYS	52	GLU	82	TYR	112	GLN	142	GLN	172	LEU	202	LEU		
23	GLY	53	LEU	83	ARG	113	VAL	143	LEU	173	ILE	203	TYR		
24	SER	54	LEU	84	ASN	114	GLN	144	TYR	174	LYS	204	THR		
25	CYS	55	ASP	85	THR	115	PRO	145	ARG	175	ASN	205	SER		
26	TRP	56	CYS	86	TYR	116	TYR	146	GLY	176	SER	206	SER		
27	ALA	57	ASP	87	PRO	117	ASN	147	GLY	177	TRP	207	PHE		
28	PHE	58	ARG	88	TYR	118	GLN	148	ILE	178	GLY	208	TYR		
29	SER	59	ARG	89	GLU	119	GLY	149	PHE	179	THR	209	PRO		
30	ALA	60	SER	90	GLY	120	ALA	150	VAL	180	GLY	210	VAL		

Figura 26: Seqüência primária da papaína (1BQI). □ resíduo básico; □ resíduo ácido; □ resíduo polar.

1	LEU	41	ARG	81	TRP	121	GLY	161	GLU	201	ILE	241	GLU
2	PRO	42	ILE	82	THR	122	ASP	162	ILE	202	LEU	242	SER
3	ALA	43	CYS	83	ARG	123	THR	163	TYR	203	GLY	243	GLU
4	SER	44	ILE	84	LYS	124	PRO	164	LYS	204	TRP	244	VAL
5	PHE	45	HIS	85	GLY	125	LYS	165	ASN	205	GLY	245	VAL
6	ASP	46	THR	86	LEU	126	CYS	166	GLY	206	VAL	246	ALA
7	ALA	47	ASN	87	VAL	127	SER	167	PRO	207	GLU	247	GLY
8	ARG	48	VAL	88	SER	128	LYS	168	VAL	208	ASN	248	ILE
9	GLU	49	SER	89	GLY	129	ILE	169	GLU	209	GLY	249	PRO
10	GLN	50	VAL	90	GLY	130	CYS	170	GLY	210	THR	250	ARG
11	TRP	51	GLU	91	LEU	131	GLU	171	ALA	211	PRO	251	THR
12	PRO	52	VAL	92	TYR	132	PRO	172	PHE	212	TYR	252	ASP
13	GLN	53	SER	93	GLU	133	GLY	173	SER	213	TRP		
14	CYS	54	ALA	94	SER	134	TYR	174	VAL	214	LEU		
15	PRO	55	GLU	95	HIS	135	SER	175	TYR	215	VAL		
16	THR	56	ASP	96	VAL	136	PRO	176	SER	216	ALA		
17	ILE	57	LEU	97	GLY	137	THR	177	ASP	217	ASN		
18	LYS	58	LEU	98	CYS	138	TYR	178	PHE	218	SER		
19	GLU	59	THR	99	ARG	139	LYS	179	LEU	219	TRP		
20	ILE	60	CYS	100	PRO	140	GLN	180	LEU	220	ASN		
21	ARG	61	CYS	101	TYR	141	ASP	181	TYR	221	THR		
22	ASP	62	GLY	102	SER	142	LYS	182	LYS	222	ASP		
23	GLN	63	SER	103	ILE	143	HIS	183	SER	223	TRP		
24	GLY	64	MET	104	PRO	144	TYR	184	GLY	224	GLY		
25	SER	65	CYS	105	PRO	145	GLY	185	VAL	225	ASP		
26	CYS	66	GLY	106	CYS	146	TYR	186	TYR	226	ASN		
27	GLY	67	ASP	107	GLU	147	ASN	187	GLN	227	GLY		
28	SER	68	GLY	108	HIS	148	SER	188	HIS	228	PHE		
29	CYS	69	CYS	109	HIS	149	TYR	189	VAL	229	PHE		
30	TRP	70	ASN	110	VAL	150	SER	190	THR	230	LYS		
31	ALA	71	GLY	111	ASN	151	VAL	191	GLY	231	ILE		
32	PHE	72	GLY	112	GLY	152	SER	192	GLU	232	LEU		
33	GLY	73	TYR	113	SER	153	ASN	193	MET	233	ARG		
34	ALA	74	PRO	114	ARG	154	SER	194	MET	234	GLY		
35	VAL	75	ALA	115	PRO	155	GLU	195	GLY	235	GLN		
36	GLU	76	GLU	116	PRO	156	LYS	196	GLY	236	ASP		
37	ALA	77	ALA	117	CYS	157	ASP	197	HIS	237	HIS		
38	ILE	78	TRP	118	THR	158	ILE	198	ALA	238	CYS		
39	SER	79	ASN	119	GLY	159	MET	199	ILE	239	GLY		
40	ASP	80	PHE	120	GLU	160	ALA	200	ARG	240	ILE		

Figura 27: Sequência primária da Catepsina B (1HUC). ■ resíduo básico; ■ resíduo ácido; □ resíduo polar.

4.1.1. EQUIPAMENTOS

Nesse trabalho, foi utilizado um relevante poder computacional para as simulações moleculares. Como equipamentos utilizados, o mais importante deles é o *cluster*, construído com recursos da FAPESP concedidos ao prof. Dr. Cláudio S. Shida. O *cluster* nada mais é que uma rede de computadores interligados entre si através de uma rede de alta velocidade (figura 28), e se encontra no Laboratório de Modelagem Molecular do Centro de Investigações Bioquímicas (CIIB) da Universidade de Mogi das Cruzes. Esses computadores processam dados de forma paralela como em um processador *multicore*, ou seja, no caso, como um *multicore* de 13 processadores Pentium® HT de 3 GHz (32 bits) cada. Ao contrário dos *multicores* tradicionais, o *cluster* apresenta uma menor eficiência devido às limitações de velocidade da rede, visto que há uma grande troca de informações entre as diferentes máquinas. Relevantemente, foi também utilizado um Pentium® Quad-core®, que possui 4 núcleos de 2,4 GHz (64 bits). Em comparação, essa máquina apresenta aproximadamente a metade da capacidade computacional do *cluster* (estimativa feita durante as simulações), uma vez que a troca de informações entre os seus diferentes processadores é mais rápida. Supondo que fosse utilizado um único computador doméstico (ex.: 1 processador 3GHz - 32 bits) para as simulações moleculares, esse trabalho se estenderia por mais alguns anos, inviabilizando a pesquisa.

O sistema operacional utilizado em todas as máquinas foi o GNU Linux, especificamente a distribuição Ubuntu 7 e Fedora 4, os quais apresentam facilidade na obtenção e atualização de programas e bibliotecas, o que contribuiu para a otimização do tempo na utilização do sistema e no desenvolvimento da pesquisa em questão.

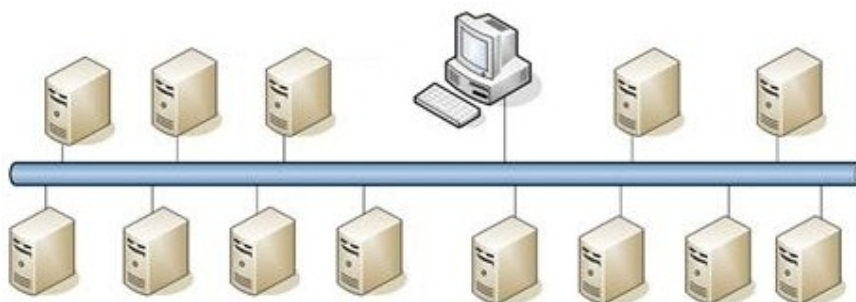


Figura 28: Representação gráfica do Cluster do Laboratório de Modelagem Molecular que pertence ao Centro Interdisciplinar de Investigações Bioquímicas da UMC.

4.2. MÉTODOS

O método utilizado nesse trabalho pode ser classificados em 4 partes principais: preparação das moléculas, preparação do ambiente de simulação, montagem do complexo (*docking*), otimização da estrutura molecular e dinâmica.

Nesse trabalho, foi estudado as enzimas cathepsina B e papaína em diferentes pHs. Para essas simulações, alterações na estrutura e campo de força foram feitas para cada molécula. Na seção “Preparação das Moléculas” (3.2.1) é descrito esse processo, assim como as alterações feitas no campo de força da heparina. Após a preparação das moléculas, partiu-se para a criação dos ambientes de simulação (3.2.2), criando limites geométricos e adicionando moléculas de água e íons. O ambiente de simulação é necessário para o início da dinâmica. Porém, para se iniciar uma dinâmica, é necessário a minimização prévia da energia do sistema, assim como descrito na seção “Otimização da Geometria Molecular” (3.2.3). O processo de dinâmica foi dividido em duas categorias, termalização e dinâmica do complexo. O processo de termalização é uma dinâmica comum, porém, ela é feita antes do processo do *docking* e serve para criar estruturas termalizadas das moléculas em questão. Essas estruturas termalizadas foram utilizadas conseqüentemente no *docking* molecular, assim como descrito na seção “Montagem do Complexo” (3.2.6). A dinâmica do complexo, por outro lado, é feita apenas para as moléculas complexadas obtidas através do *docking*. Os procedimentos utilizados para ambas essas simulações estão descritos na seção “Termalização e Dinâmica do Complexo” (3.2.4). Os parâmetros globais utilizados para todas as simulações, incluindo a minimização de energia, estão descritos na seção “Parâmetros de Simulação” (3.2.5).

Com o fim das simulações e a obtenção dos resultados, os dados foram analisados conforme descrito na seção “Análise de Estabilidade e Interação” (3.2.7). Devido ao grande número de informações, os dados foram analisados por categoria, conforme os diferentes métodos.

4.2.1. PREPARAÇÃO DAS MOLÉCULAS

Nesse trabalho, as proteínas papaína e catB foram estudadas sobre a influência de diferentes pHs: 5 e 7. O grupo do pH 5 representa essas enzimas em seu habitat natural, os

lisossomos, que apresentam um pH ácido em seu interior; e o grupo do pH 7, representa as proteínas quando presentes na matriz extracelular, um meio que possui um pH próximo ao neutro.

O método de Dinâmica Molecular não é capaz de simular explicitamente os efeitos causados pelas variações no pH do solvente. O impacto causado pelo pH ocorre ao nível quântico, e é capaz de alterar o campo de força das moléculas, assim como o processo de protonação e desprotonação (troca de elétrons). A protonação e desprotonação são os únicos parâmetros que nos permitem simular os efeitos do pH ao nível computacional por dinâmica molecular. Para implementar os efeitos resultantes do pH nas moléculas, são necessárias pequenas alterações na estrutura e parâmetros de campo de força, onde, dependendo do pH, determinados átomos serão protonados ou desprotonados manualmente, conforme o potencial da constante de dissociação ácida (pKa) teórico (*standard*) dos mesmos. As cargas e constantes de força também são redefinidas com base em dados experimentais, normalmente disponibilizadas pelo programa de simulação. Esses parâmetros são mantidos estáticos durante toda a simulação, uma vez que a dinâmica molecular não faz cálculos quânticos, ou seja, os parâmetros de campo de força não são atualizados, independentemente das condições em que as moléculas do sistema se encontram; quando alterações são necessárias ou desejadas, é preciso a intervenção humana. É necessário reforçar que a MD é um método limitado, uma aproximação grosseira da realidade.

Visto o problema discutido anteriormente, a catepsina B e papaína tiveram seus resíduos histidina des/protonados em função do pH. A histidina é o único resíduo que possui um pKa teórico (padrão) capaz de protonar ou desprotonar no intervalo entre o pH 5 e 7 (pKa = 6,0). Portanto, em pH 7, todas as histidinas foram desprotonadas (HISA) com exceção das histidinas 197 do sítio ativo da catepsina B e 159 do sítio ativo da papaína. Experimentalmente, a desprotonação da histidina do sítio ativo dessa classe de protease pode representar a inativação da atividade proteolítica da enzima e, portanto, em pH 7 a histidina do sítio ativo deve se encontrar protonada para a análise. Em pH 5, todas as histidinas foram protonadas (HISH). Complementando, a papaína utilizada nesse trabalho possui apenas duas histidinas: HIS⁸¹, HIS¹⁵⁹; e a catepsina B possui 8 histidinas: HIS⁴⁵, HIS⁹⁵, HIS¹⁰⁸, HIS¹⁰⁹, HIS¹⁴³, HIS¹⁸⁸, HIS¹⁹⁷, HIS²³⁷. O campo de força (parâmetros de interações) utilizado para essas enzimas foi o GROMOS 43a1, que é atualmente o padrão do GROMACS 3.3.x (LINDAHL, HESS & SPOEL, 2001), o programa de simulação utilizado nesse trabalho.

A heparina, assim como a catepsina B e papaína, possuem átomos des/protonáveis. Porém, a única região da heparina capaz de se des/protonar se encontra nos grupos sulfato, que possui um pKa bastante baixo. Portanto, a heparina teve os seus grupos sulfato totalmente desprotonados. O campo de força da heparina foi construído com dados do campo de força GROMOS 43a1 e dados da literatura, uma vez que o campo de força GROMOS é próprio para proteínas e não para moléculas de outra natureza. As constantes de força representadas pelas abreviações do campo de força GROMOS foram substituídas da seguinte forma:

Tabela 2: Alterações feitas no campo de força da heparina.

Constante GROMOS	Nova constante (dist./const. força)	Referência
gb_2	0.1020 / 1.8700e+07	RAGAZZI, 1997
gb_19	0.1444 / 6.1000e+06	RAGAZZI, 1997
gb_20	0.1438 / 8.7100e+06	RAGAZZI, 1997
gb_24	0.1600 / 8.3700e+06	RAGAZZI, 1997
ga_13	101.0 / 450.0	RAGAZZI, 1997
ga_11	117.0 / 421.0	RAGAZZI, 1997
ga_41	107.0 / 504.0	SPIESER, 1997
ga_41	107.0 / 518.0	SPIESER, 1997
ga_17	111.0 / 460.0	GROMOS
gd_14	180.0 / 3 / 3.77	GROMOS

4.2.2. PREPARAÇÃO DO AMBIENTE DE SIMULAÇÃO

Após a obtenção das enzimas devidamente parametrizadas em função do pH ou após a obtenção do complexo obtido através do *docking*, foi necessário preparar o ambiente para o início do processo de termalização ou para estudo da interação do complexo, respectivamente.

Foram utilizadas caixas cúbicas para as simulações. Essas caixas são as mais simples, porém, elas podem conter um número maior de moléculas de solvente, as quais podem auxiliar na minimização dos efeitos resultantes das condições periódicas de contorno. As dimensões da caixa foram definidas através do tamanho máximo da proteína ou complexo nas dimensões xyz , com um incremento adicional de 2 x 1,2 nm em cada dimensão (figura 29). A enzima ou complexo foi devidamente centralizado na caixa.

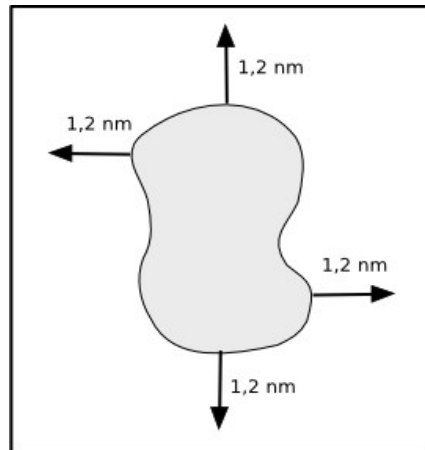


Figura 29: Proteína centralizada com 1,2 nm de distância das paredes da caixa de simulação.

Em seguida, o solvente, água (H_2O), modelo SPC (BERENDSEN *et al.*, 1981), foi adicionado de forma explícita à caixa de simulação. As moléculas de água foram adicionadas através de uma distribuição padrão independente de parâmetros de temperatura ou pressão, o que não é um problema, visto que, durante uma simulação, a caixa de simulação é capaz de se auto-ajustar, satisfazendo as condições físicas de referência.

Após a criação da caixa de simulação e adição do solvente, foi averiguado a carga total do sistema que, por convenção, deve ser zero. Determinados resíduos de uma proteína em um pH neutro, por exemplo, podem ser encontrados desprotonados, e o sistema apresentará, portanto, cargas negativas, uma vez que o hidrônio, responsável, nesse caso, por equilibrar a carga total do sistema, não existe do ponto de vista computacional, provavelmente por ainda não apresentar um comportamento fisicamente previsível. Portanto, para equilibrar essas cargas negativas é necessário a adição explícita de íons ao solvente. No caso de um sistema com carga total negativa, normalmente utiliza-se íons Na^+ , ou, no caso de um sistema com carga total positiva, utiliza-se normalmente o íon Cl^- .

4.2.3. OTIMIZAÇÃO DA GEOMETRIA MOLECULAR

Antes de iniciar o processo de termalização ou a dinâmica do complexo, foi necessário um rigoroso pré-processo de minimização de energia ou otimização da geometria molecular. Como já revisado anteriormente, a minimização corrige distorções nas ligações e interações eletrostáticas, além de tornar a energia do sistema mais homogênea, o que evita a formação de

regiões “quentes” ou energeticamente heterogêneas resultantes da adição aleatória do solvente no sistema pelo programa de simulação. A formação de regiões quentes pode favorecer o aumento abrupto da energia localizada, provocando distorções não físicas nas partículas e podendo ocasionar também o interrompimento da simulação; uma situação nada desejável.

A minimização do sistema consistiu em 3 passos principais: (1) minimização *SD* com restrição de átomos pesados (átomos pesados = tudo, exceto os hidrogênios) e moléculas do solvente; (2) minimização *SD* e (3) *CG*, ambas sem restrições. Todas as minimizações *SD* foram feitas por um período de até 1000 passos com um passo de integração de 0,001 ps. A minimização *CG* foi feita por um período de até 5000 passos com um passo de integração de 0,002 ps. A minimização de energia com restrição de átomos pesados e moléculas do solvente permite que a proteína ou o complexo converja para uma conformação de menor energia sem produzir grandes alterações na posição dos átomos pesados. Quando a minimização é feita sem restrições, as moléculas convergem mais rapidamente para um estado de menor energia, e o solvente é contraído e forçado sobre a proteína ou complexo. Após as minimizações iniciais feitas através do método *SD*, foi feita uma última minimização utilizando o método *CG* que, ao contrário do *SD*, é útil apenas no processo final de minimização da energia, onde o *SD* se demonstra ineficiente na localização da energia mínima local.

4.2.4. TERMALIZAÇÃO E DINÂMICA DO COMPLEXO

Para o processo inicial de termalização ou dinâmica do complexo, foi feita uma breve dinâmica de preparação por um tempo de 100 ps com restrição de átomos pesados e moléculas do solvente. Finalmente, foi feita uma dinâmica normal sem restrições por um tempo de 100 ns. Em ambos os casos foi utilizado um passo de integração de 0,002 ps. Uma simulação por 100 ns, dentro dos limites técnicos computacionais, é, neste caso, suficiente para a termalização das moléculas e para o estudo da interação do complexo.

4.2.5. PARÂMETROS DE SIMULAÇÃO

Para todos os processos de dinâmica molecular já citados anteriormente, foi utilizado o algoritmo LINCS (HESS *et al.*, 1997) para constranger todas as ligações com uma tolerância de 10^{-4} nm. As interações eletrostáticas de longo alcance foram tratadas através do método Poisson-Boltzmann *reaction field* (TIRONI *et al.*, 1995), com uma permissividade para a água de 54 (GARGALO *et al.*, 2000). As simulações foram feitas sobre condições constantes de temperatura, 295 K, e pressão, 1 atm, utilizando o algoritmo de acoplamento de Berendsen (BERENDSEN *et al.*, 1984) com atualização a cada 0,1 ps. Para as interações eletrostáticas de longo e curto alcance foi utilizado um raio de corte de 1.2 nm e 1.4 nm, respectivamente, com uma atualização de pares a cada 10 passos. O sistema foi simulado em uma caixa com condições periódicas de contorno.

4.2.6. MONTAGEM DO COMPLEXO

Para a montagem dos complexos, as enzimas catepsina B/papaína e o glicosaminoglicano, heparina, foram isolados de suas caixas de simulação e solvente. O programa utilizado para a montagem do complexo, ou *docking* molecular, foi o THOR/GSA (MORET *et al.*, 2002) que utiliza o método GSA. Os arquivos de estrutura e topologia dos programas THOR e GROMACS possuem grandes incompatibilidades entre si, e foi necessário o desenvolvimento de um complexo programa para a conversão entre esses diferentes formatos de arquivos (ANEXO1, “GRO2THOR Converter”).

O programa THOR/GSA é caracterizado por permitir um *docking* com moléculas parcialmente ou totalmente móveis. Essa característica permite, no caso da catpsina B ou papaína, que a proteína não perca a sua estrutura esqueleto (*backbone*) termalizada, sendo possível, também, especificar como móveis as cadeias laterais dos resíduos da proteína. Nesse trabalho, apenas as cadeias laterais dos resíduos arginina e lisina foram especificados como móveis, uma vez que eles são básicos (apresentam cargas positivas) e possuem uma longa cadeia lateral que pode facilmente interagir com os grupos negativamente carregados da heparina. Os parâmetros estruturais da heparina também foram alterados, e todos os átomos, com exceção dos átomos presentes nos anéis dos monossacarídeos, foram especificados como móveis (figura 30). Para a configuração espacial inicial do *docking*, as estruturas da heparina

e proteína foram colocadas a uma distância de 4 nm uma da outra. O *docking* foi feito com uma temperatura inicial de 200,0 K e com um número máximo de 1.000.000 ciclos. Complementando, os parâmetros utilizados no THOR/GSA foram: ParqVa=2.5, ParqAa=1.1, ParqTa=1.7, Nseriea=7, NRANa =-65140 e NslideGSAa=1.

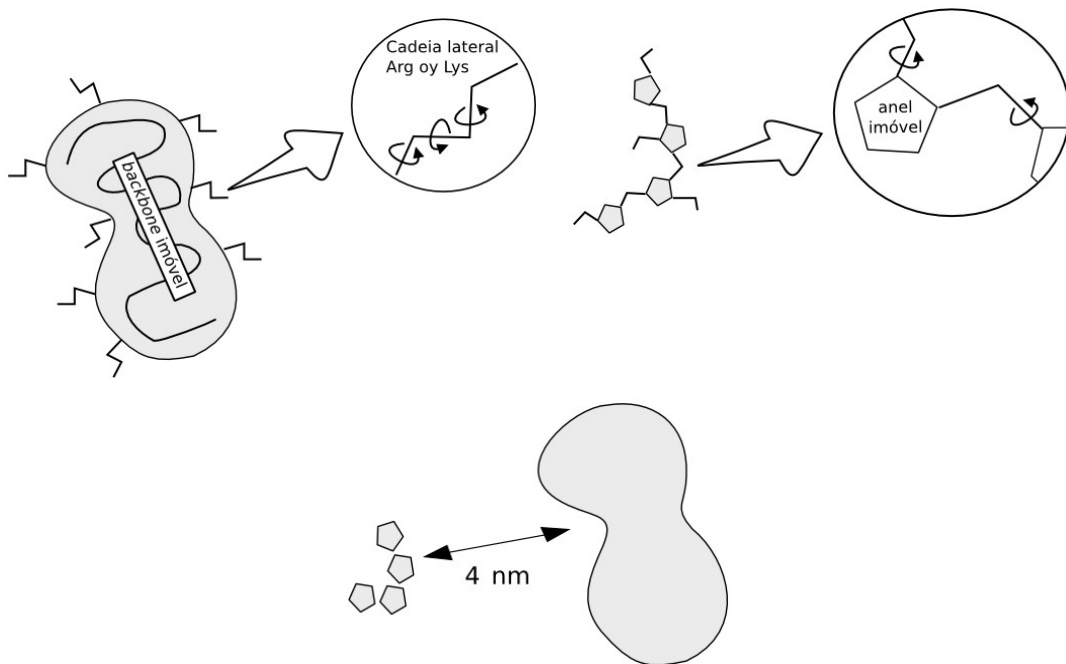


Figura 30: Representação dos parâmetros utilizados no *docking* molecular.

Após o *docking* molecular, os complexos mais relevantes foram selecionados utilizando como critério o valor da diferença entre a energia inicial e final, e a posição espacial do ligante. Complexos que apresentam uma energia pouco diferente da energia inicial, teoricamente, não são estáveis. A figura 31 mostra um complexo espacialmente desfavorável, embora a sua energia seja favorável. Ela é resultante de uma forte atração entre cargas opostas. Nesse trabalho, complexos dessa natureza foram descartados.

4.2.7. ANÁLISE DE ESTABILIDADE E INTERAÇÃO

Análise de variação estrutural: Para analisar a variação estrutural das moléculas foram utilizados os métodos RMSD, raio de giração e MOD_GYR (modificação do método de raio de giração, para maiores detalhes ver introdução, seção 2.4.11). Essas análises foram feitas para as enzimas antes e depois da formação do complexo, e são a indicação principal da ação

da heparina sobre a protease. Foram utilizados os subprogramas `g_rms` e `g_gyr` do pacote GROMACS para a análise RMSD e raio de giração, respectivamente. O MOD_GYR foi calculado através dos programas MOD_GYR_res e MOD_GYR_atom (ver ANEXO), desenvolvidos especificamente para a análise de resíduos e átomos, respectivamente. A heparina, como não possui resíduos, foi analisada, portanto, pelo programa MOD_GYR_atom. Para a análise comparativa entre as estruturas da enzima das diferentes trajetórias, foi utilizado o método tradicional RMSD. Nesse caso, por limitações do subprograma `g_rms` do GROMACS, foi criado um pequeno programa (ver ANEXO, “RMS3D_aux”). Esse programa extrai a estrutura da enzima de um único tempo de uma das trajetórias e a compara com a outra trajetória através do programa `g_rms`. Essa análise é feita em função do tempo em que a estrutura da enzima da trajetória é obtida. Para a somatória dos valores RMSD em função do tempo de uma das trajetórias, foi desenvolvido outro programa (ver ANEXO, “RMS3D_sum”), que simplesmente armazena os resultados em vetores, elevando cada um dos itens ao quadrado e somando-os posteriormente.

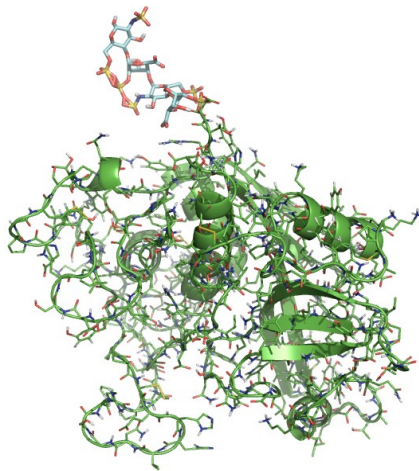


Figura 31: Complexo desfavorável entre a catepsina B e a heparina.

Estrutura secundária: Essa análise foi feita para as enzimas antes e depois da formação do complexo. O objetivo dessa análise foi verificar a influência da formação do complexo sobre a estrutura secundária da enzima, a qual mostrou, experimentalmente, relações entre o aumento/diminuição de determinados tipos estruturais e a atividade das proteases abordadas nesse estudo. A análise das estruturas secundárias foi feita através do programa DSSP (KABSCH & SANDER, 1983) implementado pelo GROMACS no subprograma DO_DSSP.

Acessibilidade dos resíduos: Essa análise também foi feita para as enzimas antes e depois da formação do complexo. A acessibilidade pode apontar importantes alterações na estrutura da proteína durante a interação, assim como a exposição de resíduos essenciais para a atividade proteolítica da catepsina B e papaína. Para essa análise, foi necessário o desenvolvimento de um programa que integrasse o GROMACS e o NACCESS ver. 2.1.1, programa responsável pelo cálculo da acessibilidade (ANEXO, “Solvent Access Module”). Para a análise feita pelo NACCESS, foi utilizado o algoritmo *rolling-ball* e uma esfera de análise de 1,4 Å de diâmetro, que representa a molécula de água.

Análise de interação: De forma mais específica, focando o perfil da interação entre a heparina e a protease, foi analisado a energia de interação de curto alcance (Lennard Jones e Coulomb), assim como, foi analisado o número de ligações de hidrogênio entre ambas as moléculas ao longo da simulação. Para essas análises foram utilizados os subprogramas *g_energy* e *g_hbond* do pacote GROMACS, respectivamente. Também foi medida a distância da heparina em relação aos resíduos da protease em função do tempo. Para essa análise foi necessário o desenvolvimento de um programa que funcionasse em conjunto com o subprograma *g_mindist* do pacote GROMACS (ANEXO, “MINDIST Module”). Para a análise do tipo “convolução” das distâncias entre a heparina e a enzima durante toda a dinâmica e um tempo de referência específico, foi desenvolvido uma fórmula (para o código fonte do programa, ver ANEXO, “New Convolute”), que é dada por:

$$C_{i,t} = |Dref_i - D_{i,t}| \times Dref_i \quad (37)$$

onde, *Dref* é a distância de referência do resíduo *i* à heparina, e *D* é a distância do resíduo *i* no tempo *t* à heparina. *Dref* foi multiplicado pela diferença das distâncias entre *Dref* e *D* para comparar os perfis de interação e salientar as distâncias mais próximas e distantes, as quais são de grande importância na caracterização da interação.

Representações gráficas: O programa NOC (CHEN & CANG, 2007) foi utilizado para representar e calcular a estrutura eletrostática das proteases. O programa pyMol (DELANO, 2007) foi utilizado para a renderização de todas as outras imagens de moléculas.

5. RESULTADOS

Nas seções seguintes será apresentado os resultados obtidos para as dinâmicas moleculares da heparina, papaína e catespina B e, especificamente, no caso das enzimas, em diferentes condições de pH, sem ou com ligante, assim como já descrito na seção de métodos (3.2).

5.1. HEPARINA

Essa seção foi criada para mostrar a preparação da heparina para a dinâmica molecular. Os demais resultados obtidos na dinâmica da heparina estão presentes nas subseções da papaína (seção 4.2).

5.1.1. PREPARAÇÃO E MINIMIZAÇÃO DE ENERGIA

Na preparação do sistema da heparina, 3.822 moléculas de água foram incluídas dentro dos limites da respectiva caixa de simulação. Para equilibrar as cargas do sistema, foram incluídos 8 íons Na^+ dispersos no solvente. As energias e forças obtidas para esse sistema nas etapas de minimização de energia estão presentes na tabela 3.

Tabela 3: Energia potencial obtida nas diversas etapas do processo de minimização de energia da heparina.

Heparina	Energia Potencial (kJ mol ⁻¹)
prep.* Steepest Descent	-2,5881284567e+05
Steepest Descent	-2,7893141108e+05
Conjugat Gradient	-2,7910643538e+05

*Obs: minimização por SD com restrição de átomos pesados.

5.2. PAPAÍNA

5.2.1. PREPARAÇÃO E MINIMIZAÇÃO DE ENERGIA

Na preparação dos sistemas da papaína, 10.045 moléculas de água foram incluídas na caixa de simulação da papaína em pH 5, 10.047 moléculas de água na caixa de simulação da papaína em pH 7 e 10.774 moléculas de água na caixa de simulação da papaína complexada à heparina. Para equilibrar a carga dos sistemas, foi necessário a inclusão de 11 íons Cl⁻ no sistema da papaína em pH 5, 10 íons Cl⁻ no sistema da papaína em pH 7 e 2 íons Cl⁻ no sistema da papaína complexada à heparina. As energias e forças obtidas para esses sistemas nas etapas de minimização de energia estão presentes na tabela 4.

Tabela 4: Energia potencial obtida nas diversas etapas do processo de minimização de energia da papaína.

Papaína – pH 5	Energia Potencial (kJ mol ⁻¹)
prep.* Steepest Descent	-5,2205962396e+05
Steepest Descent	-5,6861385498e+05
Conjugat Gradient	-5,6917668449e+05
Papaína – pH 7	
prep.* Steepest Descent	-5,2124504328e+05
Steepest Descent	-5,6799925086e+05
Conjugat Gradient	-5,6856978999e+05
Papaína complexo	
prep.* Steepest Descent	-5,7637621260e+05
Steepest Descent	-6,2249785278e+05
Conjugat Gradient	-6,2310416239e+05

*Obs: minimização por SD com restrição de átomos pesados.

5.2.2. OBTENÇÃO DOS COMPLEXOS

Apenas um complexo promissor foi obtido para a “papaína mais heparina”, uma vez que todas as tentativas de *docking* convergiram para as mesmas coordenadas espaciais. A energia inicial, antes da formação do complexo, foi de 39.726.912,93 unidades de energia

(UE). A energia final, obtida quando a heparina se estabilizou em uma região de menor energia ao longo da superfície da papaína, foi de 3.798.136,66 UE. A energia final apresentou um valor aproximadamente 10 vezes menor que a energia inicial.

A figura 32 mostra a superfície eletrostática da papaína. A região mais escura, em azul, é uma região eletropositiva, e é potencialmente um dos locais preferíveis para a ligação da heparina, uma molécula eletronegativa.

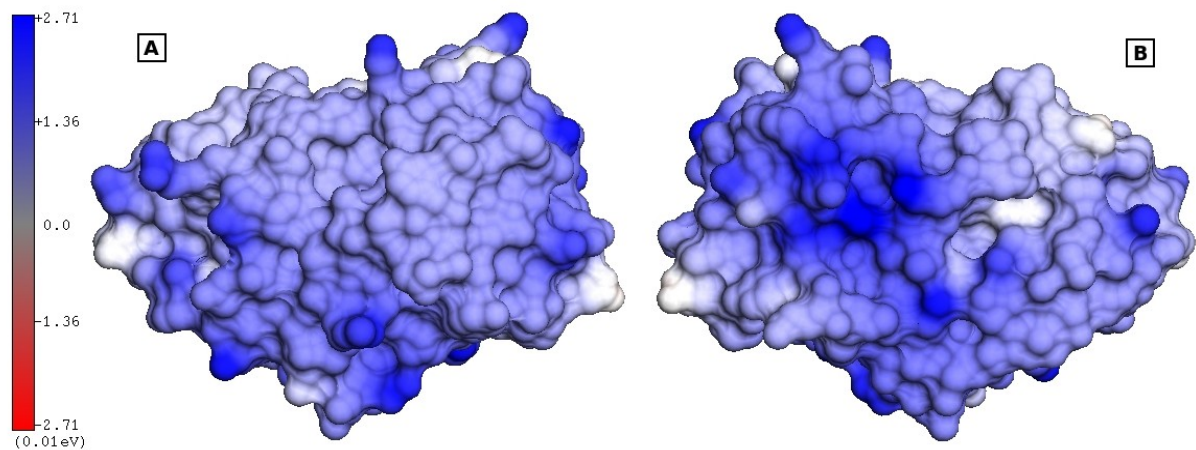


Figura 32: Superfície eletrostática da papaína. Visão frontal (A) e visão traseira (B).

Como era esperado, a heparina se ligou próximo à região mais eletropositiva da superfície da papaína (figura 33), ocupando e interagindo sobre uma vasta superfície. O sítio de ligação encontrado pela heparina se encontra na parte posterior da enzima e próximo à fenda presente entre os domínios L e R. Os resíduos da papaína encontrados em contato com os átomos de oxigênio da heparina dentro de um raio de 3 Å são a THR¹⁴, PRO¹⁵, VAL¹⁶, LYS¹⁷, ASN¹⁸, ASN⁴⁶, GLN⁴⁷, TYR⁴⁸, GLU⁵⁰, ARG⁸³, ASN⁸⁴, TYR⁸⁶, PRO⁸⁷, TYR⁸⁸, LYS¹⁷⁴, GLY¹⁷⁸, ASN¹⁸⁴ e ASN²¹². Desta lista, 3 resíduos são básicos (1 arginina e 2 lisinas), sendo que, no total, a papaína possui 22 resíduos dessa natureza. Detalhadamente, 50% (14/28) dos resíduos em contato com os átomos de oxigênio da heparina se encontram no domínio R, e 50% (14/28) se encontram no domínio L (ver figura 5). Embora a região de contato da heparina seja uma das mais eletropositivas da papaína, o açúcar entrou em contato com poucos resíduos básicos. Para confirmar se esse sítio de ligação é realmente um sítio putativo, os resultados obtidos para a dinâmica molecular deverão suportar a estabilidade do complexo e de suas respectivas variações espaciais.

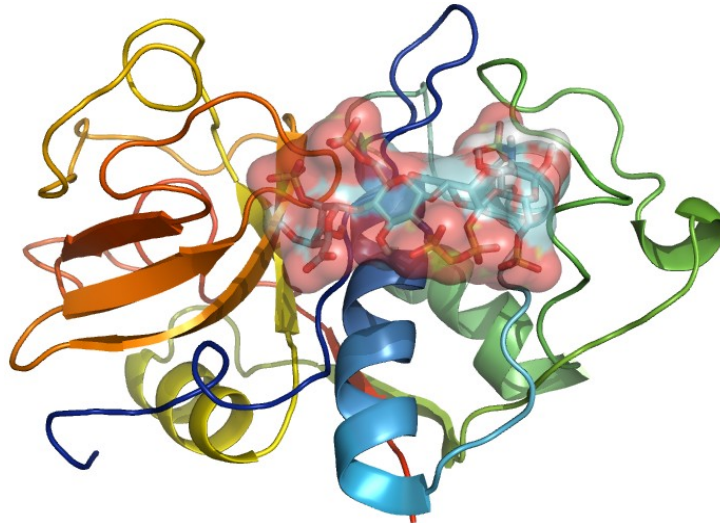


Figura 33: Complexo papaína-heparina encontrado pelo THOR/GSA.

5.2.3. ANÁLISE ESTRUTURAL

Nessa seção, em específico, foram feitas análises estruturais das moléculas papaína e heparina de forma isolada. Para as análises que envolvem especificamente as interações entre a papaína e a heparina, foi criada uma nova seção denominada “Análise da Interação” (seção 4.3).

5.2.3.1. Desvio Médio Quadrático (RMSD)

O desvio médio quadrático (RMSD) foi medido em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína (pH 7) complexada à heparina (figura 34). O RMSD também foi calculado para a heparina livre e complexada à papaína (figura 35). Análises do RMSD também foram feitas entre as trajetórias da papaína em diversas condições. Nesse caso, foi criada uma subseção especial na seção “Análise da Interação” (4.2.4), pois essa análise foi feita para testar hipóteses, utilizando dados de várias análises.

O gráfico RMSD da papaína livre em pH 5 (figura 34-A) mostrou que a estrutura da enzima se estabilizou após 20 ns de termalização. O gráfico RMSD da papaína livre em pH 7 (figura 34-B) também mostrou que a estrutura da enzima se estabilizou após 20 ns de termalização, apesar de pequenas flutuações ocorridas em torno de 40 ns. Em ambos os casos,

o RMSD atingiu rapidamente o valor médio da estabilização observado após os 20 ns iniciais da termalização. No caso da papaína complexada à heparina em pH 7 (figura 34-C), o valor do RMSD se elevou gradativamente, e encontrou a linha média de estabilização após 40 ns de dinâmica. Também foi verificado que a média do RMSD da papaína complexada foi menor que o RMSD medido para a papaína livre em pH5 e 7, ou seja, ao longo da dinâmica, a estrutura da papaína complexada variou relativamente pouco e lentamente em relação à estrutura inicial.

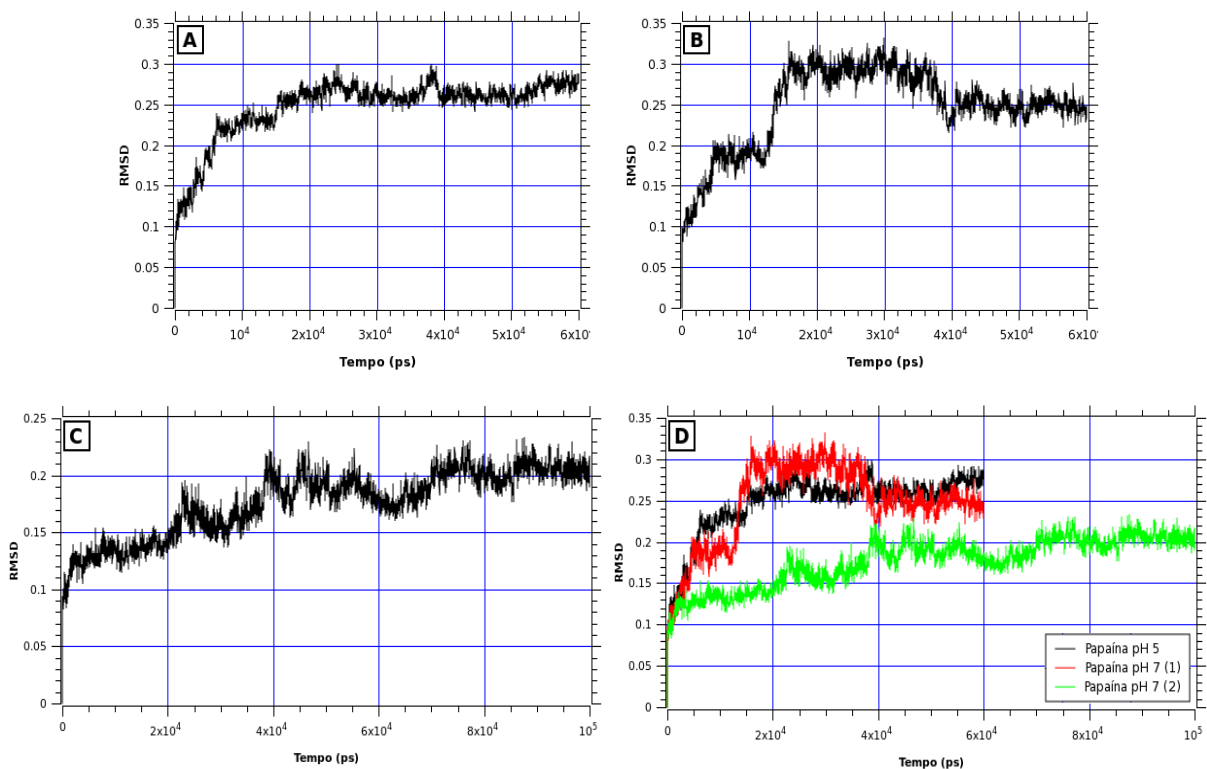


Figura 34: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises RMSD anteriores (D).

O RMSD da heparina livre e complexada (figura 35) mostrou que a estrutura da molécula evoluiu quase que imediatamente para o valor médio de estabilização em ambos os casos. Utilizando como referência os valores RMSD obtidos para a heparina livre (figura 35-A), verificou-se que o RMSD da heparina complexada à papaína apresentou valores menores de oscilação (figura 35-B), no entanto, essas oscilações se mantiveram por períodos maiores de tempo, sugerindo uma restrição estrutural e a existência de uma interação considerável entre a papaína e a heparina.

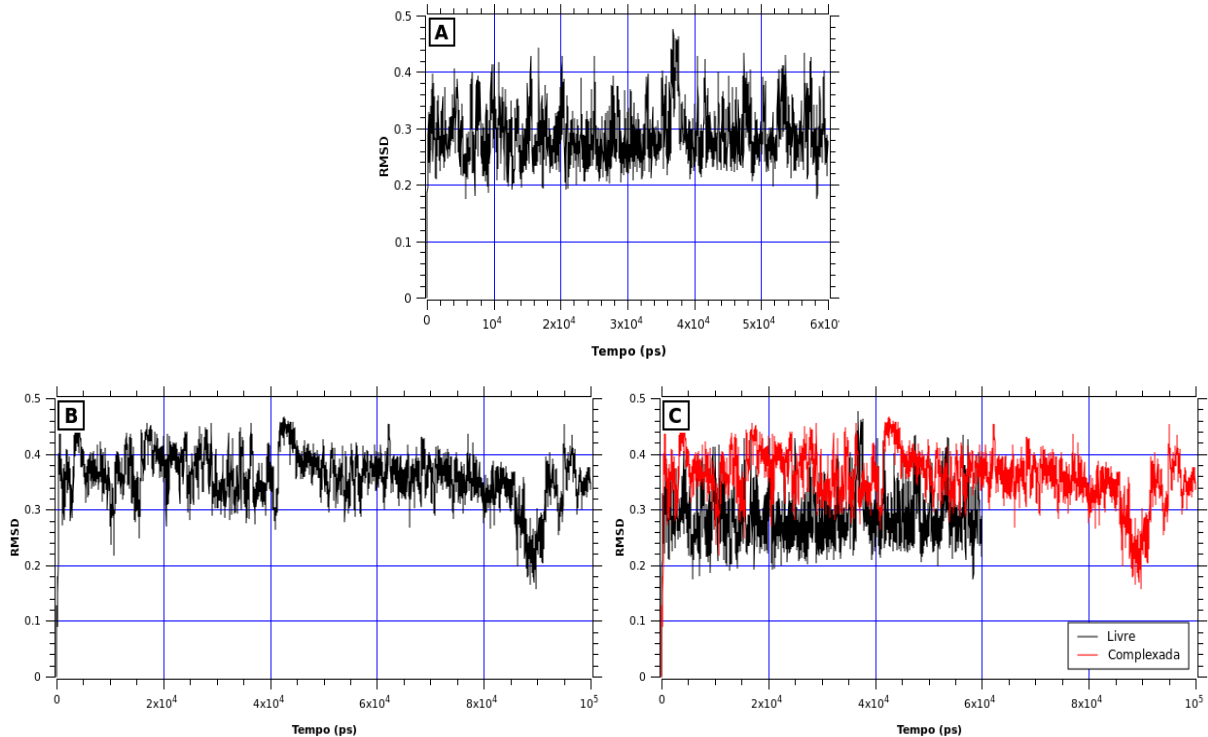


Figura 35: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 7 (B). Comparação entre as análises RMSD anteriores (C).

5.2.3.2. Raio de Giração

O raio de giração foi medido em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína complexada à heparina em pH 7 (figura 36). Os resultados da abordagem utilizando uma versão modificada do método de raio de giração, o MOD_GYR, são apresentados na seção seguinte (4.2.3.3).

Na análise da papaína livre em pH 5 (figura 36-A), a partir de 40 ns de dinâmica, os valores médios do raio de giração começam a se diferenciar do em pH 7. No pH5 a média do raio de giração foi de 1,611 nm, com um desvio padrão de 0,007 nm, e um valor máximo de 1,633 e mínimo de 1,604 nm. Para a papaína livre em pH 7 (figura 36-B), a partir do mesmo tempo, a média foi de 1,627 nm, com um desvio padrão de 0,006 nm, e um valor máximo de 1,650 e mínimo de 1,604. Para a papaína (pH 7) complexada à heparina (figura 36-C), também a partir do mesmo tempo, a média foi de 1,631 nm, com um desvio padrão de 0,007 nm, e um valor máximo de 1,654 e mínimo de 1,609 nm. Em ordem decrescente, a média do raio de giração pode ser ordenada da seguinte forma: papaína complexada, papaína livre em pH 7 e papaína livre em pH 5.

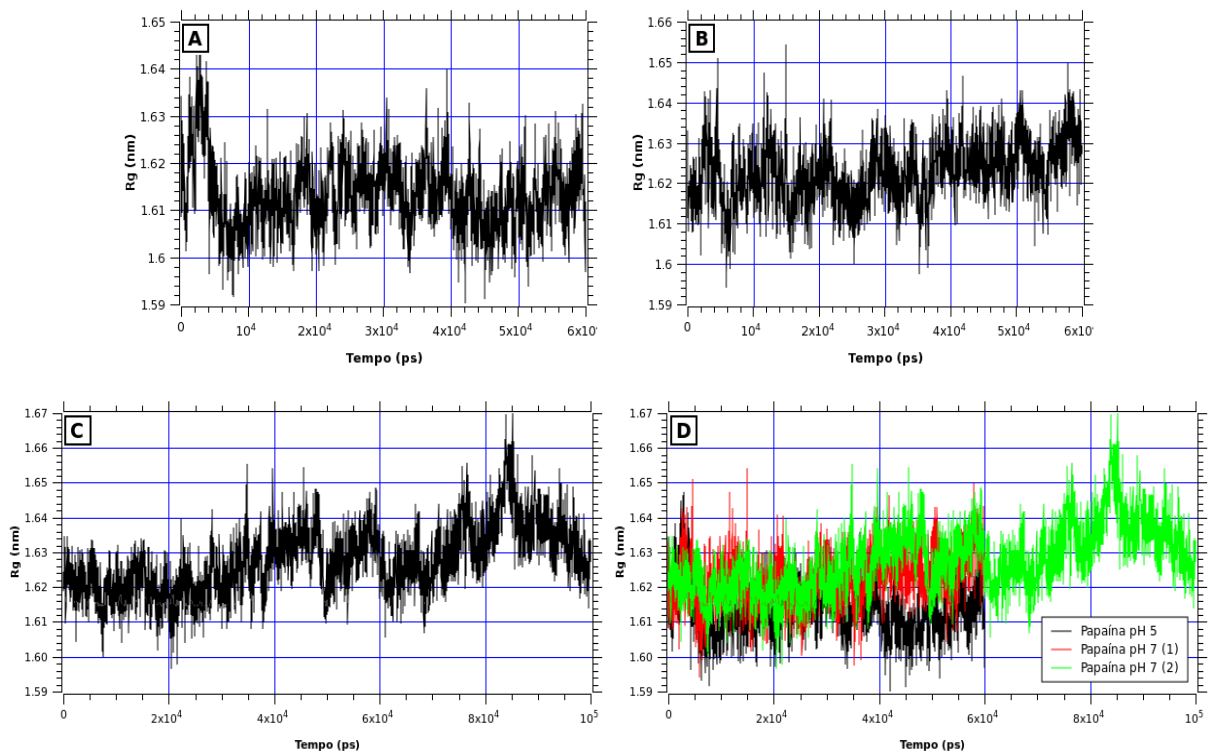


Figura 36: Raio de giração em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de raio de giração anteriores (D).

O raio de giração da papaína complexada apresentou valores máximos e uma média que superou os valores de raio de giração obtidos para a papaína livre em pH 5 e pH 7. Oscilações abruptas foram observadas no tempo aproximado de 38, 48, 60 e 85 ns. Outras oscilações também foram observadas, mas eram menos abruptas e de difícil identificação.

5.2.3.3. Raio de Giração Modificado por Resíduo (MOD_GYR)

O MOD_GYR foi medido em função do tempo por resíduo para a papaína em pH 5, pH 7 e complexada à heparina (figura 37-A-D), e por átomo para a heparina livre e complexada à papaína (figura 37-E-G). Esse tipo de análise, teoricamente, permite uma análise melhor detalhada das variações estruturais sofridas pela papaína e heparina.

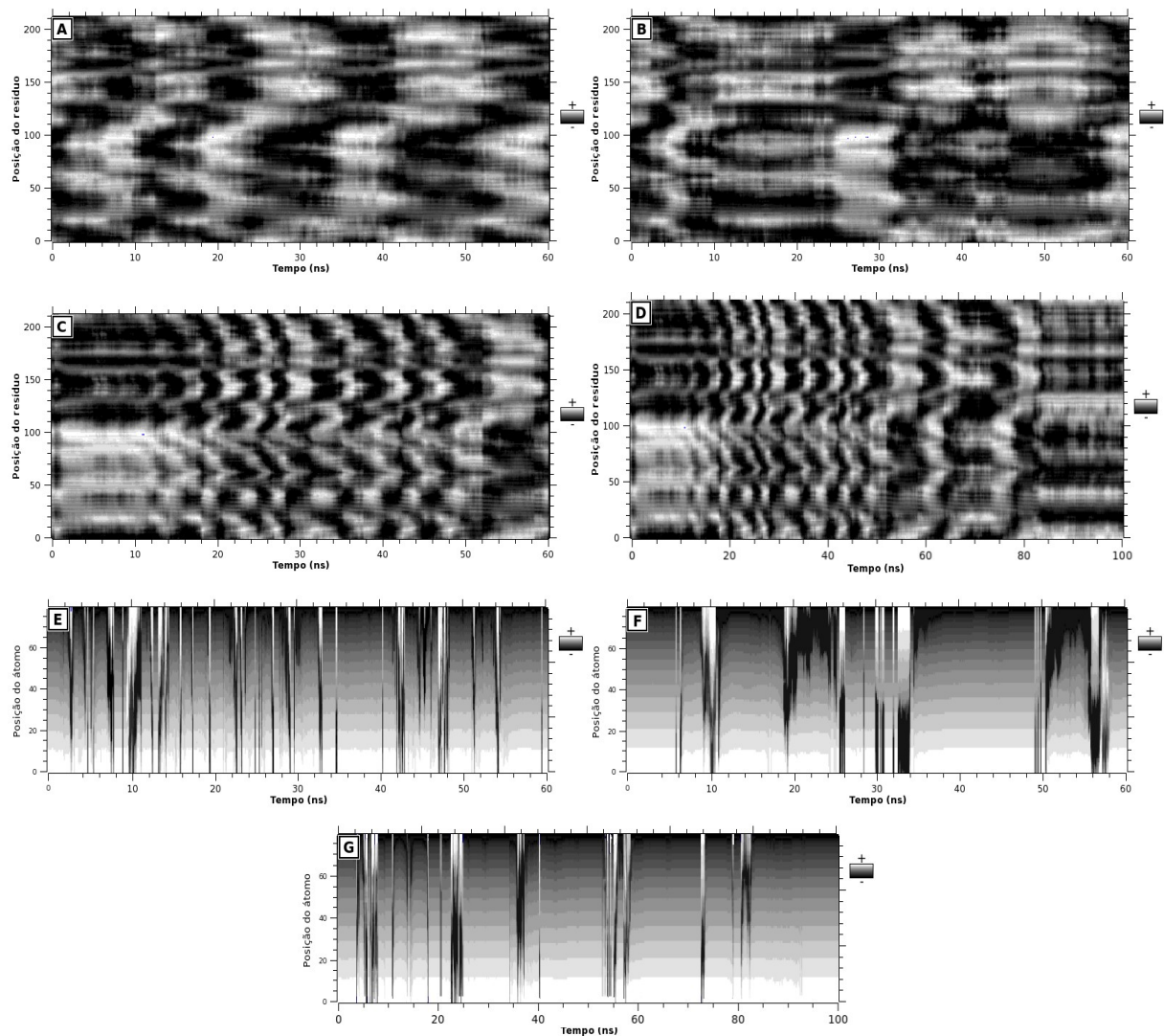


Figura 37: MOD_GYR por resíduo em função do tempo da papaína livre em pH 5 (A), livre em pH 7 (B), complexada em pH 7 (C e D*), e para a estrutura da heparina livre em pH 7 (E) e complexada em pH 7 (F e G*).
*Obs: análise feita até 100 ns da dinâmica – gráfico fora da escala dos demais gráficos feitos até 60ns.

As figuras 37 e 43 mostram que o MOD_GYR por resíduo da papaína apresentou oscilações periódicas com frequências dependentes do estado da enzima. Por ordem

decrecente de frequência de oscilação, temos a papaína complexada (figura 37-C), papaína em pH 5 (figura 37-A) e papaína em pH 7 (figura 37-B). No caso da heparina, livre ou complexada, não foi verificado um padrão de oscilação estrutural. No caso da heparina livre (figura 37-E), a estrutura analisada pelo MOD_GYR por resíduo apresentou diversas e rápidas oscilações, enquanto a heparina complexada à papaína (figura 37-F), apresentou poucas oscilações, as quais apresentaram um maior tempo relativo de duração. Um menor grau de oscilação, aparentemente, significa uma maior rigidez estrutural, e o tempo da oscilação está diretamente relacionado à velocidade de ocorrência das alterações estruturais.

Para um melhor entendimento sobre os resultados MOD_GYR por resíduo obtidos para a papaína, foi criado um pequeno modelo em 2 dimensões (figura 38 e 39) análogo à estrutura da papaína, o qual também inclui os domínios L e R. A criação desse modelo é de extrema importância no entendimento abstrato fornecido pelos resultados MOD_GYR, os quais podem ser difíceis de serem observados.

A figura 38 analisou o perfil da dinâmica de uma das oscilações sofridas pela papaína complexada à heparina em pH 7. Ficou claro que os domínios da enzima oscilam de forma alternada, em fases diferentes (figura 39 e 40), e que determinadas regiões do domínio L oscilam juntamente com o domínio R; essas regiões, especificamente, foram denominadas de M1 (18-24), M2 (61-66) e M3 (92-97) e estão representadas na figura 42. Em pH 5, a papaína não apresentou de forma clara e proporcional essas oscilações conjuntas e, em pH 7, essa relação ficou pouco definida devido ao aparente perfil instável de oscilação. Uma comparação entre esses perfis de oscilação está representada na figura 41. Aparentemente, a oscilação conjunta das regiões M1, M2 e M3 é o resultado da interação dessas regiões com a heparina.

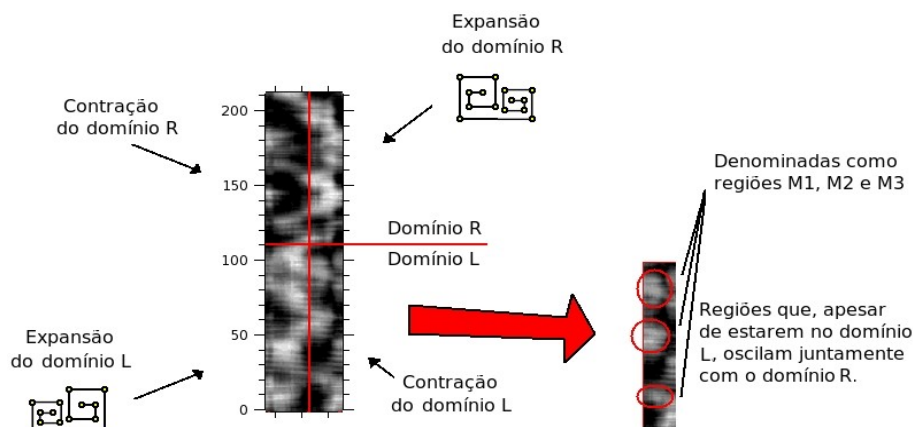


Figura 38: Perfil da dinâmica de uma única oscilação da papaína (pH 7) complexada, baseado no MOD_GYR por resíduo.

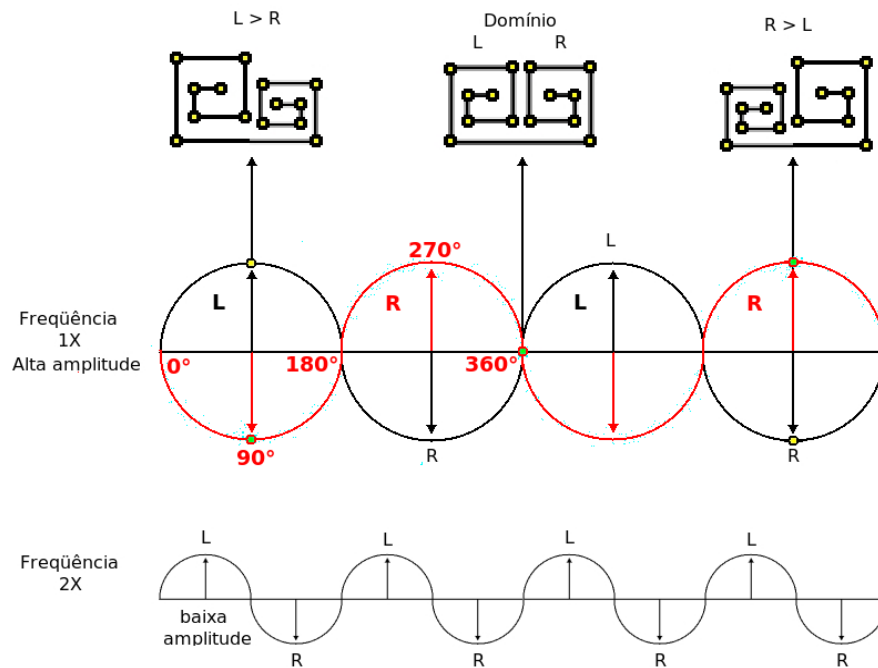


Figura 39: Representação em 2 dimensões da oscilação sofrida pela papaína com base nos resultados obtidos na análise feita pelo MOD_GYR por resíduo. “Frequência 2X” representa o dobro da frequência de “1X”.

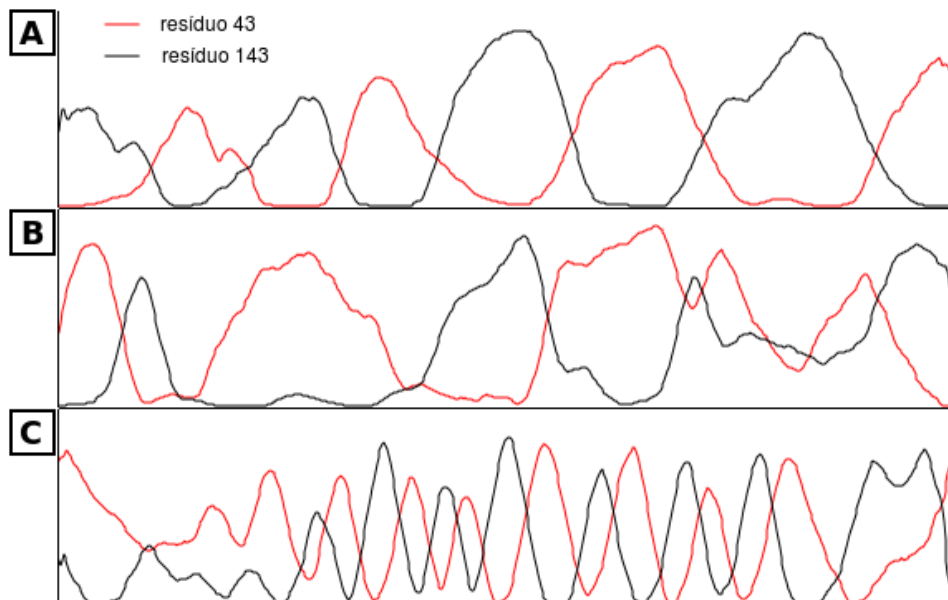


Figura 40: Análise MOD_GYR feita isoladamente para os resíduos 43 e 143 da papaína livre em pH 5 (A), papaína livre em pH 7 (B) e complexada em pH 7 (C).

O modelo em 2 dimensões (figura 39) mostrou que as ondas de oscilação e, especificamente, as amplitudes positivas e negativas, virtualmente, representam a expansão e a contração dos domínios L e R, ou vice-versa. Se considerarmos a amplitude do modelo de onda como o reflexo da maleabilidade da proteína, e que a frequência de oscilação é proporcional à amplitude, podemos dizer que, quanto maior a frequência de oscilação, menor será a maleabilidade da proteína.

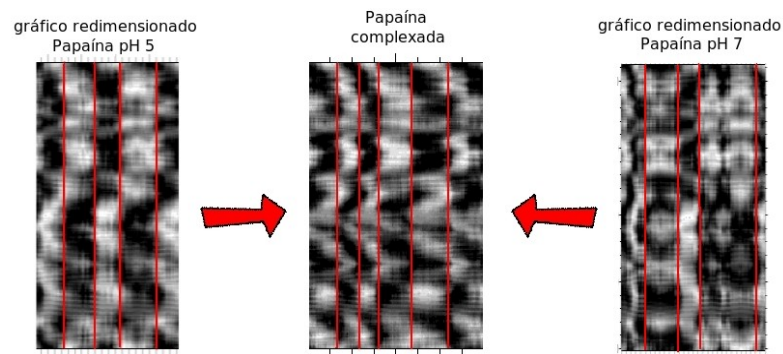


Figura 41: Comparação entre os perfis de oscilação da papaína em pH 5, pH 7 e complexada.

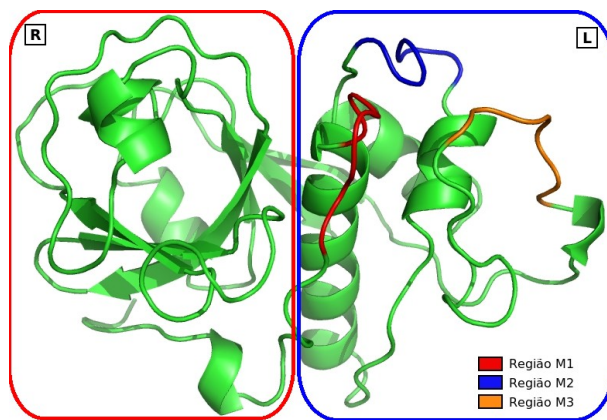


Figura 42: Papaína e as regiões M1, M2 e M3 (conforme os resultados obtidos na figura 38). Os domínios R e L estão representados pelas respectivas letras.

A frequência de oscilação MOD_GYR foi medida para o resíduo 43 da papaína complexada e da papaína em pH 5 utilizando a transformada de Fourier (figura 43-A e B). A frequência medida para a papaína em pH 5 foi de aproximadamente 0,065 oscilações/nanosegundo, e para a papaína complexada foi de aproximadamente 0,185 oscilações/nanosegundo, que representa um valor 2,85 vezes maior que o encontrado para a papaína em pH 5. A análise da frequência de oscilação da papaína em pH 7 não foi mostrada devido ao respectivo padrão oscilatório, o qual foi determinado como sendo pouco definido para a transformação de Fourier.

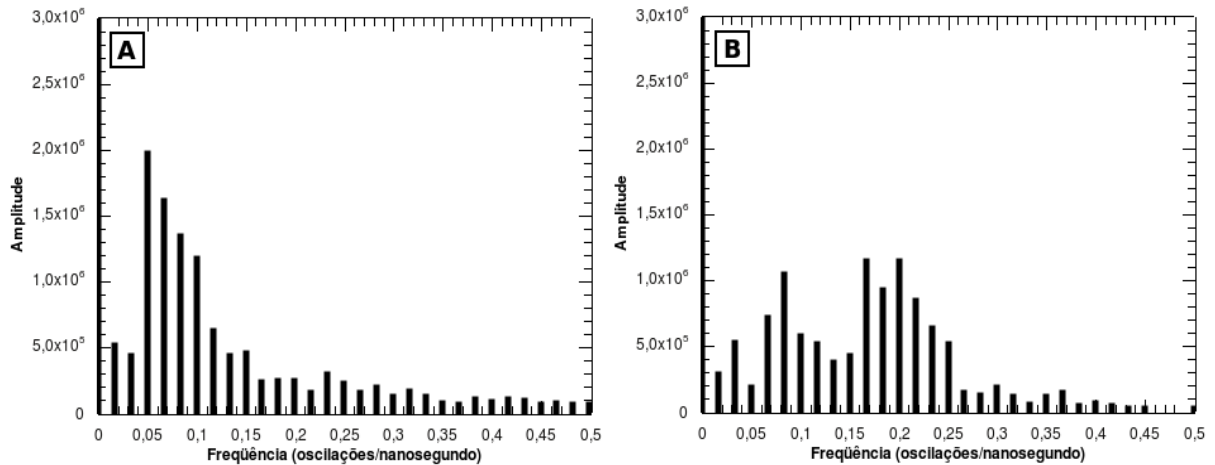


Figura 43: Transformada de Fourier aplicada para as oscilações MOD_GYR do resíduo 43 da papaína em pH 5 (A) e complexada em pH 7 (B).

5.2.3.4. Energia Potencial

A energia potencial foi medida em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína complexada com a heparina (figura 44). A energia potencial também foi medida para a heparina livre e complexada à papaína (figura 45). A energia potencial foi inferida para as moléculas isoladas, sem qualquer interferência das moléculas do solvente ou ligante. Esse potencial inclui os componentes gerais do campo de força, como a energia do potencial harmônico de distância, ângulo, diedro próprio e impróprio, e as energias eletrostáticas de curto e longo alcance.

A análise da energia potencial da papaína livre mostrou que em pH 5 e 7 (figura 44-A e B), a energia se manteve praticamente estável, apesar da presença de pequenas tendências, como o leve aumento da energia potencial observado no final do gráfico da papaína livre em pH 5. No caso da papaína complexada (figuras 44-C), a energia potencial aumentou gradativamente, mais rapidamente que o verificado para a papaína livre em pH 5, com um padrão bastante similar ao observado na análise do raio de giração correspondente, incluindo parte das oscilações abruptas, e sugerindo uma relação direta entre o raio de giração e a energia potencial da enzima. As oscilações abruptas verificadas na análise do raio de giração provavelmente possuem relações diretas com a energia potencial da molécula em questão.

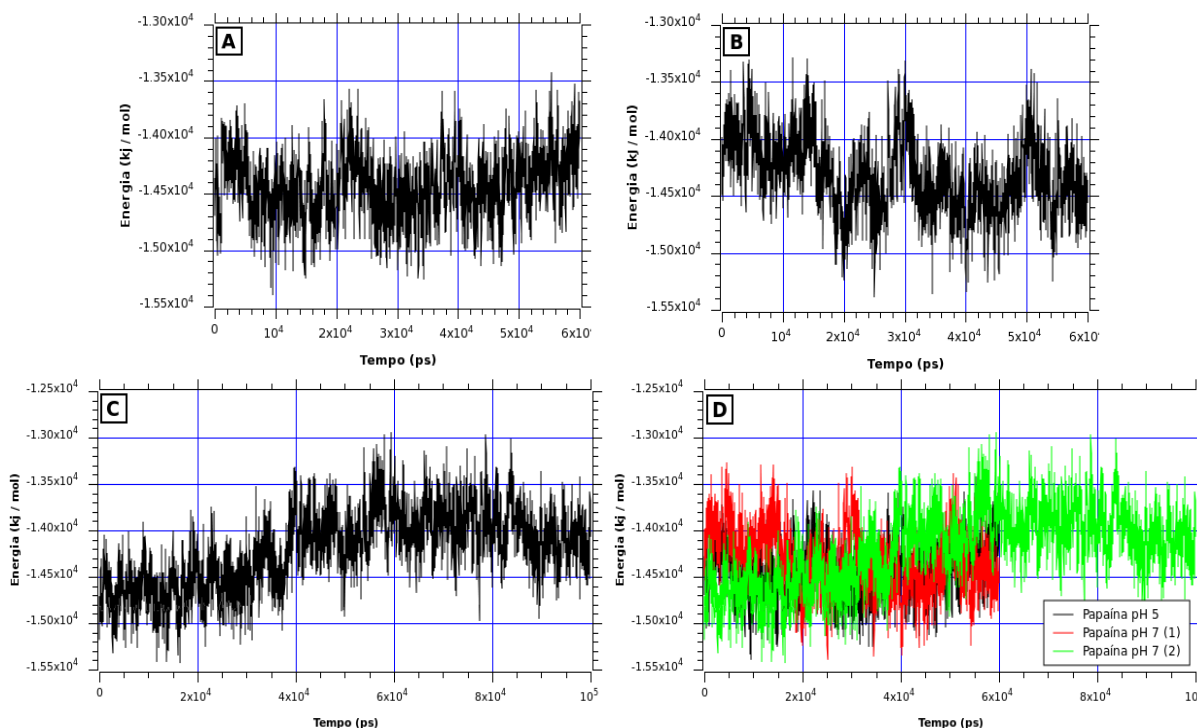


Figura 44: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de energia potencial anteriores (D).

Na análise da papaína livre em pH 5, a partir de 40 ns e até 60 ns de dinâmica, período em que a energia potencial aparentemente começou a se divergir entre as diferentes análises, a média da energia potencial foi de -14357,0 kJ/mol, com um desvio padrão de 285,4 kJ/mol, e um valor máximo de -13423,0 e mínimo de -15145,9 kJ/mol. Para a papaína livre em pH 7, a média foi de -14435,2 kJ/mol, com um desvio padrão de 303,6 kJ/mol, e um valor máximo de -13385,8 e mínimo de -15344,4 kJ/mol. Para a papaína complexada à heparina, a média foi de -13972,1 kJ/mol, com um desvio padrão de 305,3 kJ/mol, e um valor máximo de -12936,6 e mínimo de -14929,0 kJ/mol. Esses valores estatísticos aparentemente se mantiveram estáveis até o fim da dinâmica de 100 ns da papaína complexada. Em ordem decrescente de energia potencial, temos a papaína complexada, papaína livre em pH 5 e papaína livre em pH 7. A elevada energia potencial apresentada pela papaína complexada sugere a influência da heparina sobre estrutura da enzima. Essa influência pode ser considerada uma forma de mobilização, assim como sugerido na análise MOD_GYR por resíduo.

Quanto à energia potencial da heparina livre e complexada à papaína (figura 45), ela mostrou oscilar em torno de uma média bastante estável, sem alterações consideráveis.

Detalhadamente, a análise da heparina complexada, de 0 a 60 ns de dinâmica, a média da energia potencial foi de 375,5 kJ/mol, com um desvio padrão de 61,4 kJ/mol, e um valor máximo de 630,7 e mínimo de 211,7 kJ/mol. Para a heparina livre, no mesmo período de tempo, a média foi de 364,8 kJ/mol, com um desvio padrão de 62,2 kJ/mol, e um valor máximo de 862,3 e mínimo de 187,0 kJ/mol. Em ordem decrescente de energia potencial, temos a heparina complexada e a heparina livre. Assim como verificado para a papaína complexada, a heparina complexada também apresentou uma elevada energia potencial, a qual, logicamente, sugere a influência da papaína sobre a estrutura da heparina.

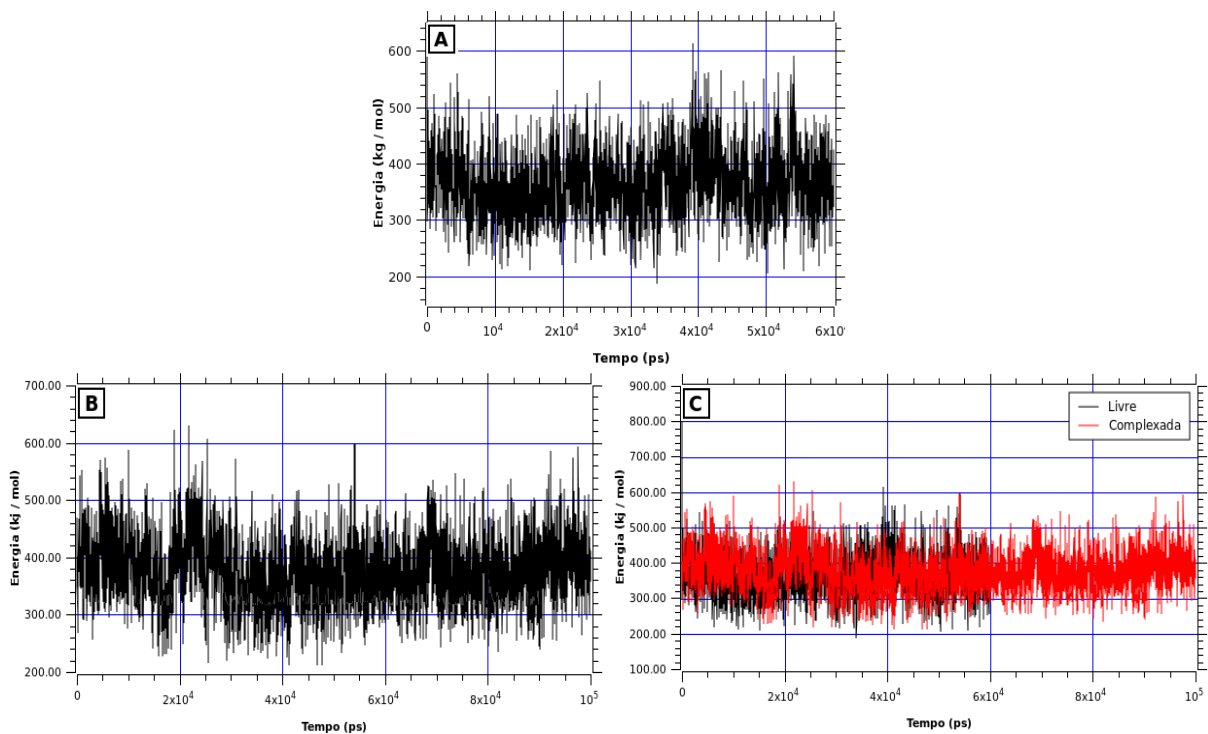


Figura 45: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 7 (B). Comparação entre as análises anteriores (C).

5.2.3.5. Estrutura Secundária

A estrutura secundária, α -hélice, foi medida em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína complexada à heparina (figura 46). Na análise da papaína livre em pH 5 (figura 46-A), a partir de 40 ns e até 60 ns de dinâmica, período em que o conteúdo de α -hélice aparentemente começou a se divergir entre as diferentes análises, a média do conteúdo de α -hélice foi de 47,0%, com um desvio padrão de 2,4%, e um valor

máximo de 56,0 e mínimo de 41,0%. Para a papaína livre em pH 7 (figura 46-B), no mesmo intervalo de tempo, a média foi de 45,6%, com um desvio padrão de 3,0%, e um valor máximo de 53,0 e mínimo de 37,0%. Para a papaína complexada à heparina (figura 46-C), também no mesmo intervalo, a média foi de 49,4%, com um desvio padrão de 2,1%, e um valor máximo de 54,0 e mínimo de 40,0%. Após 60 ns e até 100 ns, os valores obtidos para a papaína complexada se mostraram estáveis, apesar de uma grande oscilação verificada próxima ao tempo de 80 ns. Em ordem decrescente de conteúdo de α -hélice, temos a papaína complexada, papaína livre em pH 5 e papaína livre em pH 7. Além do mais, a análise da estrutura secundária da papaína mostrou que, em pH 5 e 7, o conteúdo de α -hélice diminuiu gradativamente até atingir uma linha média de estabilização, e esse processo ocorreu aparentemente durante os primeiros 30 ns da dinâmica. Inversamente, o conteúdo de α -hélice da papaína complexada, que apresentou um conteúdo inicial menor que o encontrado na papaína livre em pH 5 e 7, aumentou em função do tempo, superando o conteúdo dessa estrutura em relação aos demais casos a partir do tempo aproximado de 38 ns. O tempo de 38 ns corresponde à primeira oscilação abrupta verificada na análise do raio de giração e energia potencial.

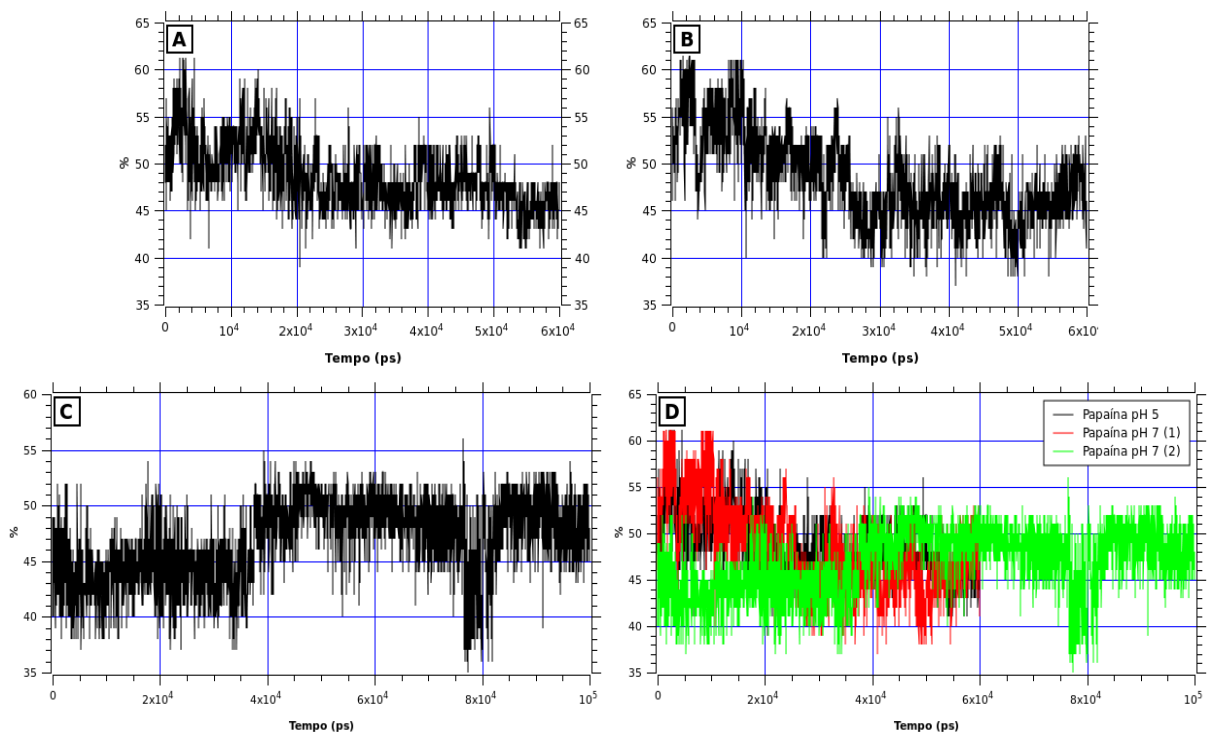


Figura 46: Conteúdo de estrutura α -hélice medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7 (C). Comparação entre as análises de conteúdo de estrutura α -hélice anteriores (D).

A estrutura secundária, folha- β , foi medida em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína complexada à heparina (figura 47). Na análise da papaína livre em pH 5 (figura 47-A), durante toda a dinâmica, a média do conteúdo de folha- β foi de 38,0%, com um desvio padrão de 4,6%, e um valor máximo de 48,0 e mínimo de 22,0%. Para a papaína livre em pH 7 (figura 47-B), a média foi de 34,9%, com um desvio padrão de 3,5%, e um valor máximo de 44,0 e mínimo de 22,0%. Para a papaína complexada à heparina em pH 7 (figura 47-C), a média foi de 35,5%, com um desvio padrão de 3,3%, e um valor máximo de 45,0 e mínimo de 18,0%. Após 60 ns e até 100 ns, os valores obtidos para a papaína complexada se mostraram bastante estáveis. Em ordem decrescente de conteúdo de folha- β temos a papaína livre em pH 5, papaína complexada, e papaína livre em pH 7.

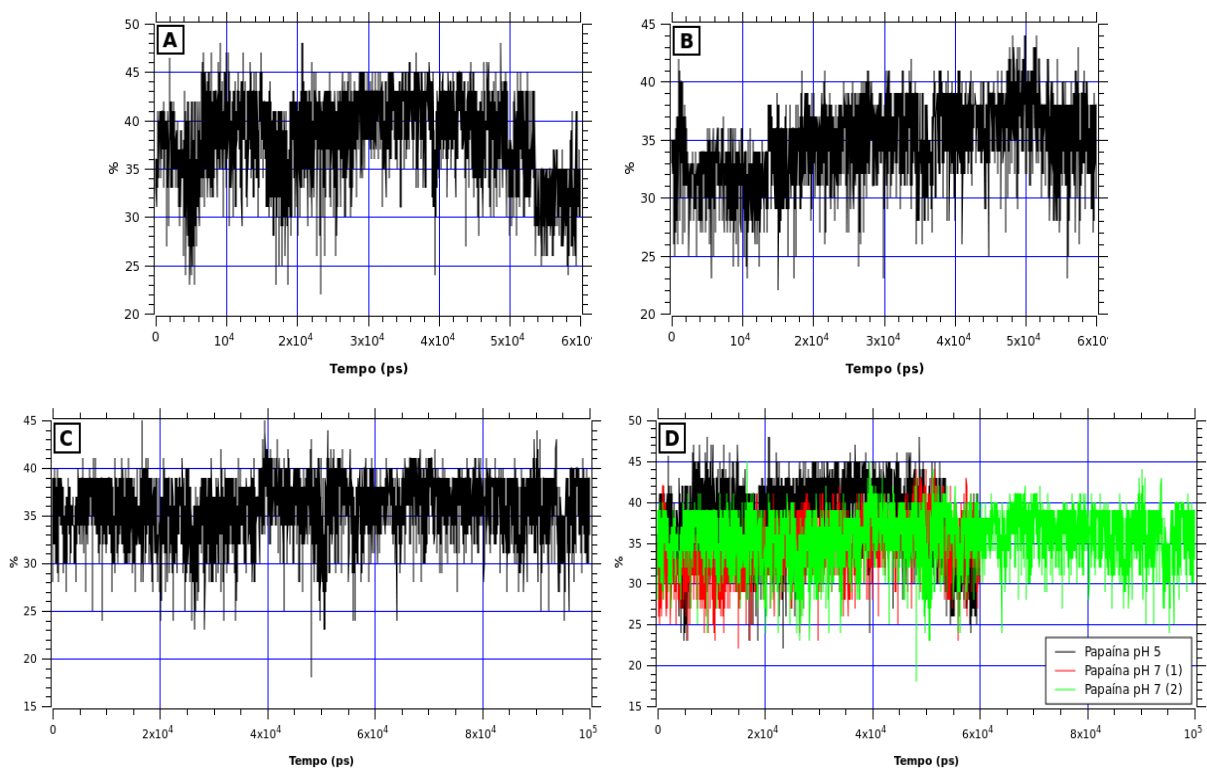


Figura 47: Conteúdo de estrutura folha- β medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de conteúdo de estrutura folha- β anteriores (D).

Todos os tipos de estruturas secundárias também foram analisadas em função dos resíduos e tempo (figura 49). No gráfico da papaína complexada (figura 49-C), a região formada pelos resíduos GLY¹³⁸, LYS¹³⁹, ASP¹⁴⁰, PHE¹⁴¹ e GLN¹⁴², apresentou uma clara

transição de uma estrutura mal definida para uma estrutura do tipo α -hélice a partir do tempo aproximado de 38 ns, onde, a partir desse momento, ela se manteve estável até o fim da dinâmica. No caso da papaína livre em pH 7 (figura 49-B), nessa mesma região, houve uma diminuição do conteúdo de α -hélice entre o intervalo aproximado de 25 e 50 ns. A papaína em pH 7 também apresentou uma diminuição no conteúdo α -hélice da região formada pelos resíduos ARG⁹⁶, SER⁹⁷, ARG⁹⁸, GLU⁹⁹ entre o intervalo aproximado de 48 e 56 ns. No caso da papaína livre em pH 5 (figura 49-A), essa última região citada não apresentou em momento algum estrutura do tipo α -hélice, porém, apresentou estrutura α -hélice na região formada pelos resíduos ASP⁶, TRP⁷, ARG⁸, GLN⁹ e LYS¹⁰ durante toda a dinâmica, ao contrário do verificado para a papaína nas demais situações.

Uma análise mais detalhada da estrutura α -hélice presente entre os resíduos ASP⁶ e LYS¹⁰, e ARG⁹⁶ e GLU⁹⁹ (figura 48), os quais apresentam resíduos básicos e ácidos, mostrou que os resíduos básicos, lisina (LYS ou K) e arginina (ARG ou R), não se encontram na mesma direção na hélice, mas sim, na direção dos resíduos ácidos, como o ácido glutâmico (GLU ou E) e ácido aspártico (ASP ou D). A direção da cadeia lateral de um resíduo em uma α -hélice é de grande importância na caracterização de uma estrutura estável e, nesses casos analisados, essa relação pareceu favorecer a formação e a estabilidade da estrutura α -hélice em questão. A distância entre os resíduos ASP⁶ e LYS¹⁰, e entre os resíduos ARG⁹⁶ e GLU⁹⁹ foi medida, porém, não houve variações relevantes que pudessem explicar o aumento ou diminuição do conteúdo de α -hélice dessas regiões (dados não apresentados).

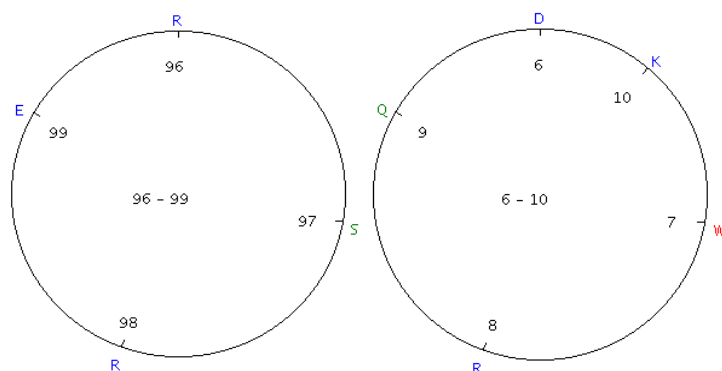


Figura 48: Diagrama helicoidal de duas estruturas α -hélices que sofreram transições estruturais durante a dinâmica da papaína complexada à heparina.

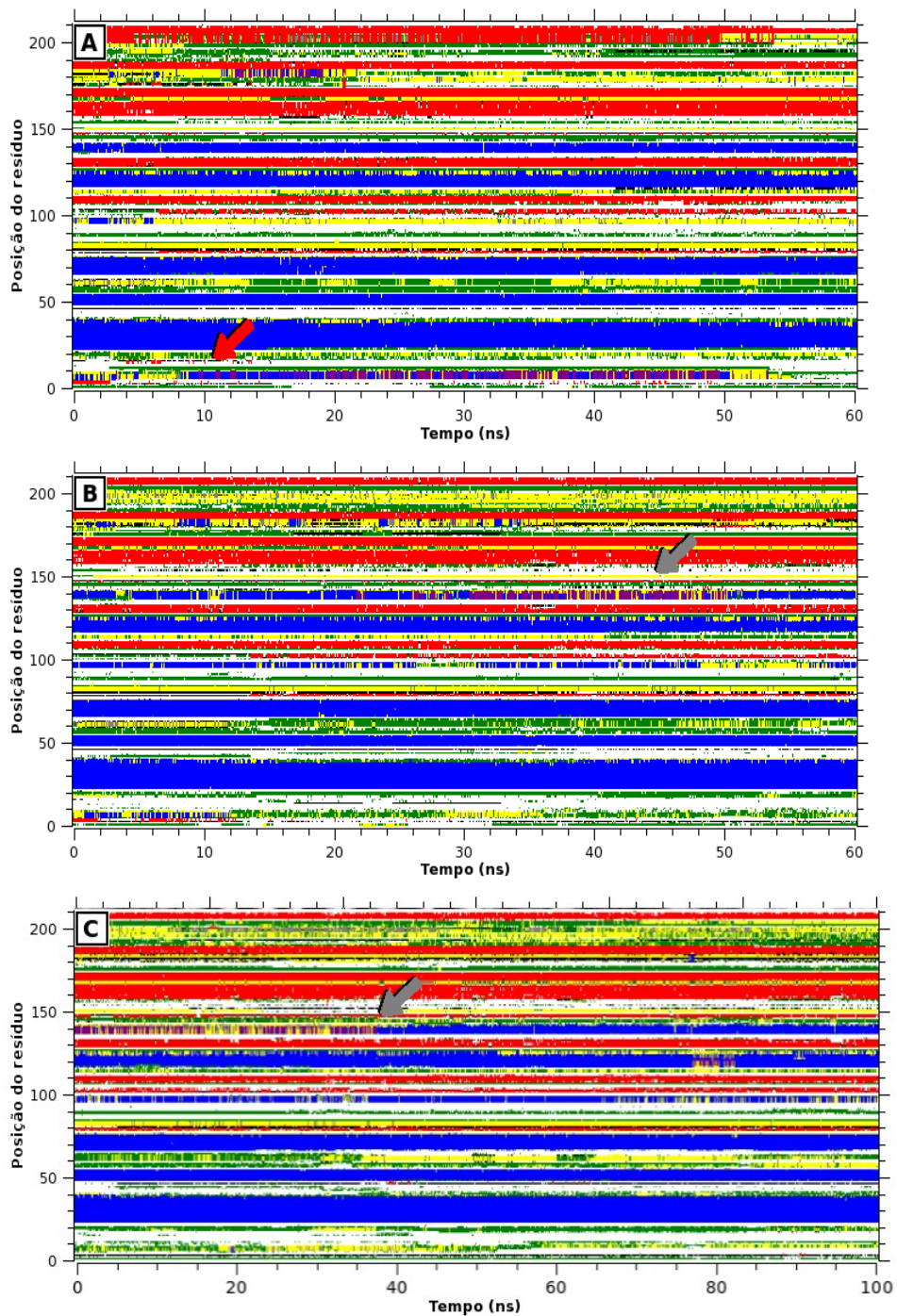


Figura 49: Estruturas secundárias medidas em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Em azul, α -hélice, e em vermelho, folha- β .

5.2.3.6. Acessibilidade

A acessibilidade ao solvente dos resíduos foi medida em função do tempo para a papaína livre em pH 5 e 7, e para a papaína complexada à heparina (figura 50). No gráfico de acessibilidade da papaína livre em pH 5 (figura 50-A), as diferenças de acessibilidade mais evidentes ocorreram para a TYR¹²³, ARG¹⁴⁵, LYS¹⁵⁶, GLN⁹ e LYS¹⁷. Nos três primeiros resíduos citados, a acessibilidade dos mesmos diminuiu ao longo do tempo. Nos últimos dois resíduos, houve um efeito contrário, ou seja, houve um aumento de acessibilidade ao longo da dinâmica. No caso da papaína livre em pH 7 (figura 50-B), as diferenças de acessibilidade mais evidentes ocorreram para a ARG⁵⁸, ARG⁵⁹, ARG⁹³, TYR⁹⁴, ARG⁹⁸ e LEU⁴⁵, onde apenas o último resíduo teve a sua acessibilidade aumentada ao longo da dinâmica. No caso da papaína complexada (figura 50-C), o processo procedeu de forma inversa, e ocorreu para os resíduos ARG⁵⁹, ARG⁹⁸, ARG¹⁴⁵ e LYS¹³⁹. A grande diferença nesse caso é que os três primeiros resíduos tiveram a sua acessibilidade aumentada, e o último resíduo teve a sua acessibilidade diminuída ao longo da dinâmica. Nesse caso, grande parte das transições de acessibilidade verificadas ocorreram no tempo aproximado de 38 ns. Esse tempo corresponde ao momento em que também foram verificadas transições em outros parâmetros de análise, como o raio de giração, conteúdo de α -hélice e energia potencial. No caso da papaína livre em pH 5 e 7, as transições de acessibilidade ocorreram em tempos diferentes.

Os resíduos da papaína complexada e livre que sofreram transições de acessibilidade ao solvente ao longo do tempo foram salientados na representação gráfica da papaína (figura 51). Pôde-se perceber que eles se encontram próximos ou em estruturas do tipo α -hélice. Além do mais, os momentos em que ocorreram as transições de acessibilidade são compatíveis com os momentos em que ocorreram as transições no conteúdo da estrutura secundária α -hélice, relevando a relação desses resíduos sobre a estabilidade e conteúdo dessa estrutura. Aparentemente, o aumento do conteúdo da estrutura α -hélice da papaína complexada parece ser o resultado de alterações estruturais drásticas, uma vez que a papaína livre em pH 7, que apresenta as mesmas condições da papaína complexada, não apresentou alterações no conteúdo dessa estrutura, o que sugere, novamente, a influência da heparina sobre a enzima.

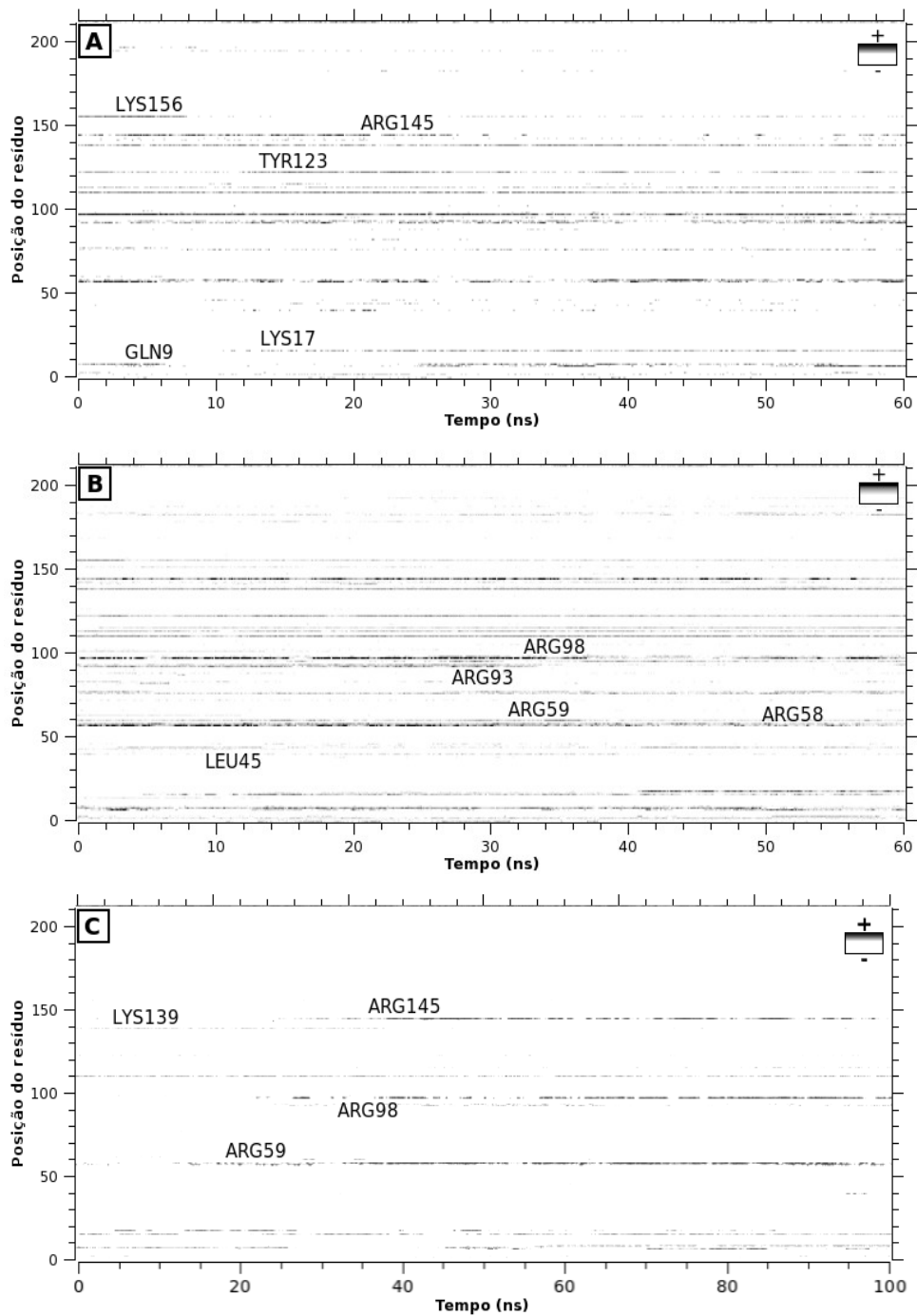


Figura 50: Acessibilidade ao solvente medida em função de cada resíduo na estrutura primária e do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Regiões mais escuras representam uma maior acessibilidade.

Analisando de forma mais detalhada a acessibilidade dos resíduos da papaína em função do tempo, foi verificado a acessibilidade do resíduo HIS¹⁵⁹, que não foi salientada na análise geral anterior, e é essencial para a atividade da enzima, como revisto diversas vezes nesse trabalho. A acessibilidade da HIS¹⁵⁹ foi medida em função do tempo para a papaína livre em pH 5, pH 7 e para a papaína complexada à heparina (figura 52).

A papaína livre em pH 5 é a que mostrou a menor acessibilidade do resíduo HIS¹⁵⁹ em função do tempo (figura 52-A). Após 20 ns de dinâmica, período em que as oscilações iniciais da dinâmica se estabilizaram, a acessibilidade desse resíduo apresentou uma média de 15,26 Å², um desvio padrão de 5,86 Å², e um valor máximo de 38,67 e mínimo de 1,16 Å². No caso da papaína livre em pH 7 (figura 52-B), após o mesmo tempo, a acessibilidade média da HIS¹⁵⁹ foi de 29,10 Å², com um desvio padrão de 8,7 Å², e um valor máximo de 57,15 e mínimo de 5,59 Å². Finalizando, a acessibilidade da HIS¹⁵⁹ da papaína complexada à heparina (figura 52-C), também no mesmo período, apresentou uma média de 30,10 Å², com um desvio padrão de 9,26 Å², e um valor máximo de 57,83 e mínimo de 2,4 Å². Por ordem decrescente de acessibilidade desse resíduo, temos a papaína em pH 5, papaína em pH 7 e a papaína complexada à heparina. Porém, no caso da papaína complexada, no intervalo aproximado entre 38 e 41 ns, 60 e 64 ns e 80 e 85 ns, houve uma queda drástica na acessibilidade da HIS¹⁵⁹, alcançando valores similares ou próximos ao encontrado na análise da papaína livre em pH 5, sendo que, no tempo entre 38 e 41 ns e, em menor grau, entre 80 e 85 ns, a queda da acessibilidade foi mais abrupta e elevada que a observada no tempo aproximado entre 60 e 64 ns. Foi verificado que os períodos de tempo em que ocorreram a diminuição da acessibilidade da HIS¹⁵⁹ são altamente compatíveis com os períodos de oscilação MOD_GYR por resíduo correspondentes (figura 53).

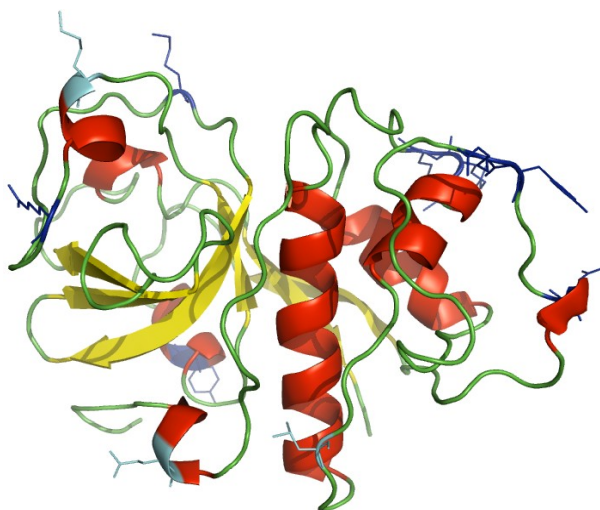


Figura 51: Estrutura da papaína (vermelho, α -hélice, amarelo, folha-B) e, em destaque (azul), os resíduos que podem afetar as estruturas do tipo α -hélice através do aumento ou diminuição das respectivas acessibilidades.

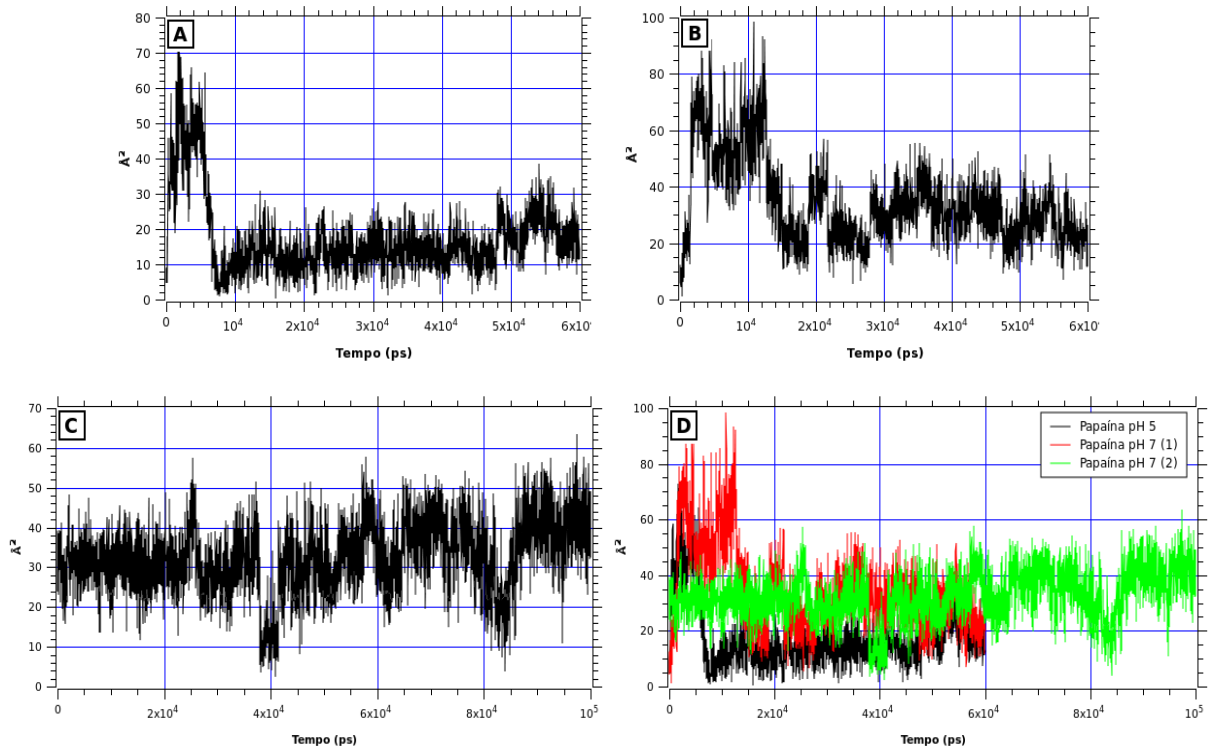


Figura 52: Acessibilidade ao solvente da HIS¹⁵⁹ medida em função do tempo da trajetória da dinâmica molecular da estrutura da papaína livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7(C). Comparação entre as análises de acessibilidade da HIS¹⁵⁹ anteriores (D).

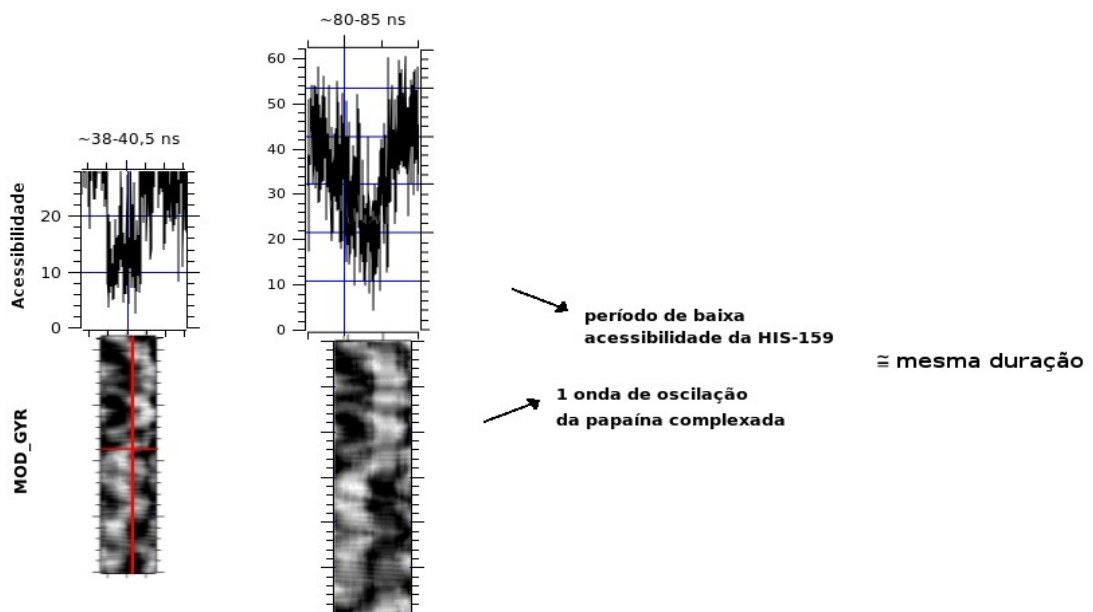


Figura 53: Comparação entre os períodos de tempo em que ocorreram a diminuição da acessibilidade da HIS¹⁵⁹ e das ondas de oscilação MOD_GYR por resíduo correspondentes.

5.2.4. ANÁLISE DA INTERAÇÃO

Para averiguar a estabilidade da interação entre a papaína e a heparina, foi medido o número de ligações de hidrogênio e a energia de interação, ambas em função do tempo (figura 55).

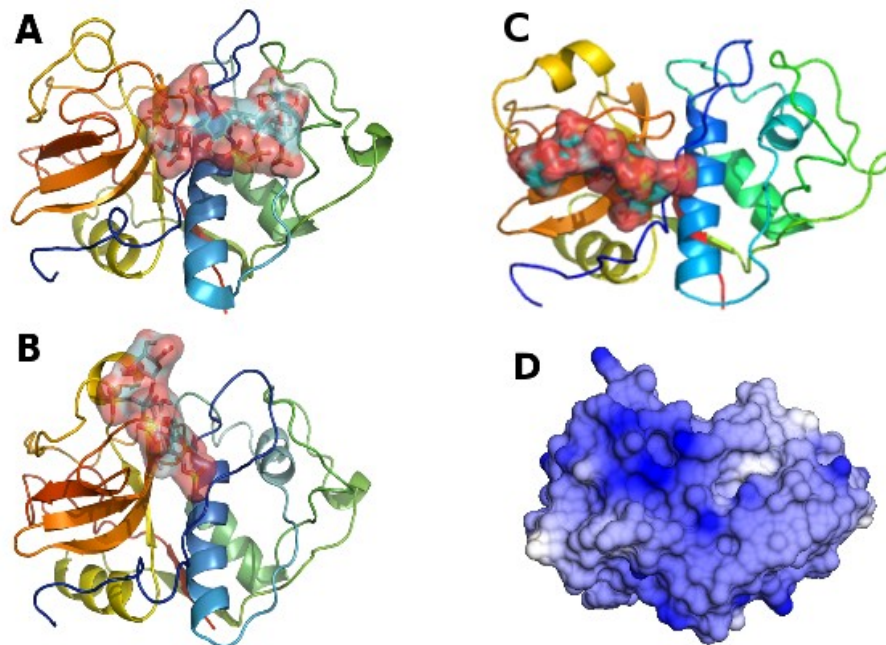


Figura 54: Representação gráfica do complexo inicial (A), complexo no tempo de 40 ns (B), complexo no tempo de 84 ns (C) e superfície eletrostática da papaína no tempo de 0 ns (D).

A interação do complexo papaína mais heparina se mostrou estável, uma vez que apresentou uma energia de interação negativa ao longo da dinâmica, o que significa que a heparina não foi expelida da papaína. A estabilidade da interação do complexo também foi suportada pelo número de ligações de hidrogênio, os quais mantiveram valores maiores do que zero durante grande parte do tempo, com algumas exceções verificadas nos tempos próximos de 12 e 50 ns. O complexo também foi analisado de forma visual (figura 54); em relação ao complexo inicial, os complexos obtidos nos tempos de 40 e 84 ns apresentaram uma leve diferença, a qual foi identificada como um deslocamento, o qual restringiu a heparina à região mais eletropositiva da papaína. Além do mais, nesses tempos em específico, foi verificado que a heparina entrou em contato com ambos os domínios da papaína.

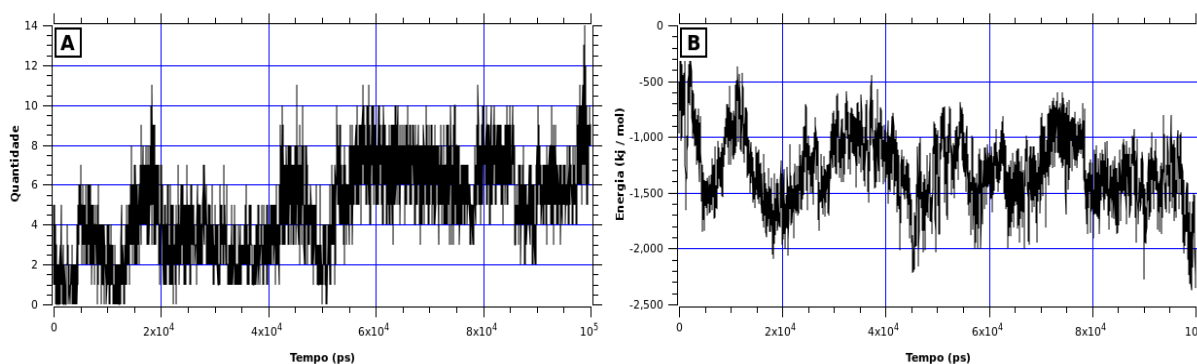


Figura 55: Número de ligações de hidrogênio entre a papaína e a heparina (A), e a energia de interação de curto alcance (Lennard Jones e Coulomb) entre essas mesmas moléculas (B).

Em função do tempo, também foi medido a distância entre os resíduos da papaína e os átomos de oxigênio da heparina (figura 56-A). O perfil de contato foi praticamente o mesmo durante toda a dinâmica, exceto por algumas variações aparentemente normais. As regiões de contato identificadas para papaína se encontram entre os resíduos 5-17 (domínio L), 135-148 e 173-188 (domínio R). Nos intervalos de tempo em que foram verificadas a diminuição da acessibilidade da HIS¹⁵⁹, foram também verificadas aparentes intensificações nas distâncias entre a papaína e a heparina (esses dados serão melhor detalhados posteriormente). Adicionalmente, conforme a figura 56-B, os resíduos histidina da enzima apresentaram uma proximidade estável com os átomos de oxigênio da heparina, que foi visualmente intensificada aproximadamente após 50 ns de dinâmica.

Os resíduos da papaína que entraram em contato com os átomos de oxigênio da heparina dentro de um raio de 3 Å no tempo de 40 ns são a GLN¹⁴², GLY¹⁷⁸, THR¹⁷⁹ e GLY¹⁸⁰, e, dentro de um raio de 5 Å, os resíduos são a LYS¹³⁹, GLN¹⁴², LEU¹⁴³, TYR¹⁴⁴, ARG¹⁴⁵, LYS¹⁷⁴, ASN¹⁷⁵, GLY¹⁷⁸, THR¹⁷⁹, GLY¹⁸⁰, TRP¹⁸¹ e ASN¹⁸⁴, todos encontrados no domínio R. Desses resíduos, 3 são básicos (1 arginina e 2 lisina). No tempo de 84 ns, os resíduos que entraram em contato com os átomos de oxigênio da heparina dentro de um raio de 3 Å são o TRP⁷, GLY¹¹, ALA¹², VAL¹³, LYS¹⁷, LYS¹⁷⁴, GLY¹⁸⁰, ASN¹⁸⁴, GLY¹⁸⁵ e TYR¹⁸⁶, e, dentro de um raio de 5 Å, os resíduos são o TRP⁷, GLN⁹, LYS¹⁰, GLY¹¹, ALA¹², VAL¹³, THR¹⁴, PRO¹⁵, LYS¹⁷, LYS¹⁷⁴, THR¹⁷⁹, GLY¹⁸⁰, TRP¹⁸¹, GLU¹⁸³, ASN¹⁸⁴, GLY¹⁸⁵, TYR¹⁸⁶ e ARG¹⁸⁸. Nesse último caso, a quantidade de resíduos encontrados em cada domínio é equivalente entre si (exatamente 50% para ambos os lados). Desses resíduos, 4 são básicos (1 arginina e 3 lisinas),

sendo 2 lisinas do domínio L e 1 arginina e 1 lisina do domínio R. As diferenças entre os perfis de contato obtidos nos tempos de 40 e 84 ns, de modo geral, são o resultado da maior aproximação da heparina à papaína em função do tempo, assim como representado na figura 57. Esse gráfico representa a soma das distâncias entre a heparina e os resíduos da papaína, devidamente separados conforme os domínios L e R da enzima. Nesse caso, os resultados obtidos apresentaram uma rápida relação com as frequências de oscilação MOD_GYR, mas não com a diminuição da acessibilidade da HIS¹⁵⁹.

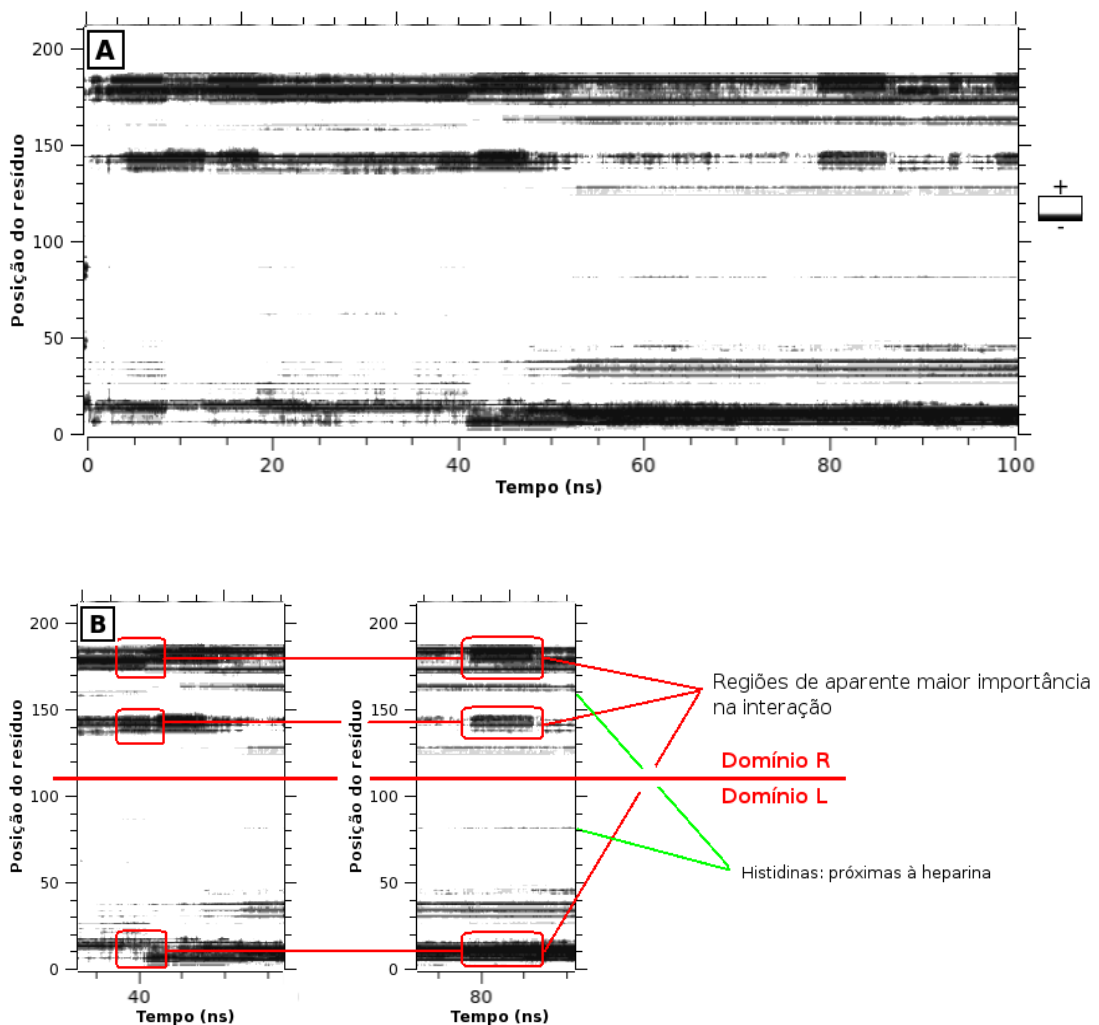


Figura 56: Distância dos resíduos da papaína à heparina em função do tempo (A). Apenas os átomos de oxigênio da heparina foram utilizados como referência para o cálculo de distância. Regiões de aparente maior importância na interação entre a papaína e a heparina baseado nos tempos em que foram verificados a diminuição da acessibilidade da HIS¹⁵⁹ (B).

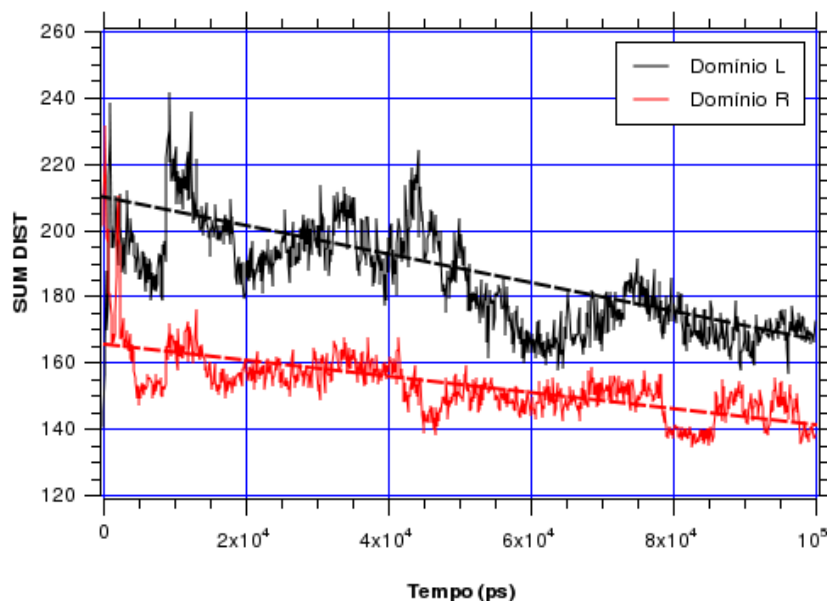


Figura 57: Soma das distâncias (nm) entre os átomos de oxigênio da heparina e os resíduos da papaína dos domínios L e R, separadamente.

Para verificar a existência de padrões de interação similares aos verificados nos tempos de 40 e 84 ns, foi feita uma relação do tipo “convolução” sobre toda a dinâmica do complexo (figura 58). Esse método foi desenvolvido nesse trabalho e foi devidamente descrito na seção de métodos (3.2.7). Nos gráficos a seguir, valores baixos, representados em preto, significam uma maior semelhança do perfil analisado com o perfil de referência.

Analisando os resultados obtidos na figura 58-A e B, ficou evidenciado que o perfil de interação verificado no tempo de 40 ns não é o mesmo perfil do tempo de 84 ns. Conforme os dados, o perfil do tempo de 40 ns se manteve o mesmo entre o tempo aproximado de 0 e 40 ns e, por outro lado, o perfil do tempo de 84 ns manteve o mesmo perfil entre o tempo aproximado de 50 e 100 ns. Conseqüentemente, o intervalo entre 40 e 50 ns apresentou um perfil diferente, ou melhor, um perfil de transição. Embora os perfis de referência, 40 e 84 ns, tenham se mantido os mesmos durante um longo intervalo de tempo, algumas regiões apresentaram intensidades de similaridade maiores que nos demais tempos. Essas regiões, claramente, são as regiões próximas aos tempos de referência. Interessantemente, essa intensificação se estendeu para os mesmos intervalos onde foram verificados a diminuição da acessibilidade da HIS¹⁵⁹, sugerindo a existência de perfis de interação específicos para a geração de determinadas alterações estruturais na enzima. Os resultados também sugerem que os diferentes perfis de interação podem criar efeitos estruturais similares. Adicionalmente, conforme o mapa de distância de contatos (figura 56-A), as regiões de interação mostraram

ser as mesmas durante toda a dinâmica, porém, o que os resultados aqui sugerem é que as distâncias dessas regiões à heparina não são as mesmas entre os perfis de referência. Isso pode ser consequência de pequenos deslocamentos e rotações da heparina ou de alterações estruturais sofridas pela papaína durante a dinâmica. Para a clarificação desses dados, foi medido a distância entre a heparina e as regiões encontradas entre os resíduos 5-17, 135-148 e 173-188 da papaína (figura 61), conforme verificado anteriormente no mapa de contatos. As distâncias verificadas (figura 60) comprovam as observações anteriores, ou seja, a grande diferença entre os perfis de interação dos tempos 40 e 84 ns é a distância entre a heparina e as regiões de contato da papaína. Também foi verificado que, no intervalo aproximado entre 80 e 85 ns, a tendência do perfil, com exceção da região 5-17, foi se aproximar parcialmente do perfil observado para o tempo de 40 ns, sugerindo um mecanismo similar para esses perfis na indução de alterações estruturais similares na papaína.

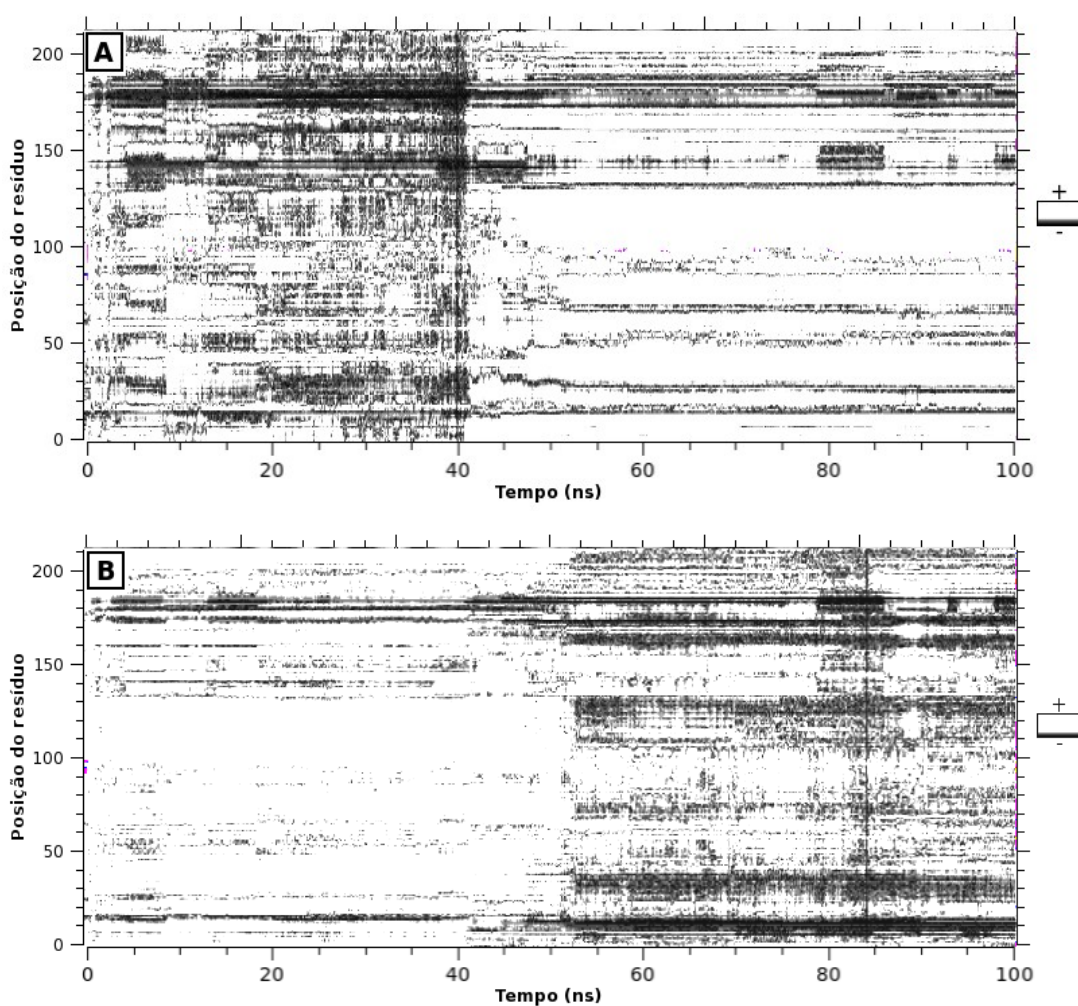


Figura 58: Comparação do perfil de interação obtido no tempo exato de 40 ns (A) e 84 ns (B) com os perfis dos demais tempos.

O gráfico de comparação de perfis, utilizando como referência o perfil obtido no tempo de 40 ns, foi comparado com o gráfico MOD_GYR por resíduo da papaína complexada. É interessante notar que os perfis mais similares ao perfil de 40 ns estão diretamente relacionados à alta frequência de oscilação MOD_GYR por resíduo, assim como detalhado na figura 59. Essa interação também apresentou fortes relações com as distâncias entre a heparina e as regiões 5-17 (menor contato), 135-148 e 173-188 (maior contato) da papaína. A distância da heparina para essas regiões é um fator que pode aumentar ou diminuir a frequência de oscilação MOD_GYR por resíduo, aparentemente por restringir a dinâmica da enzima.

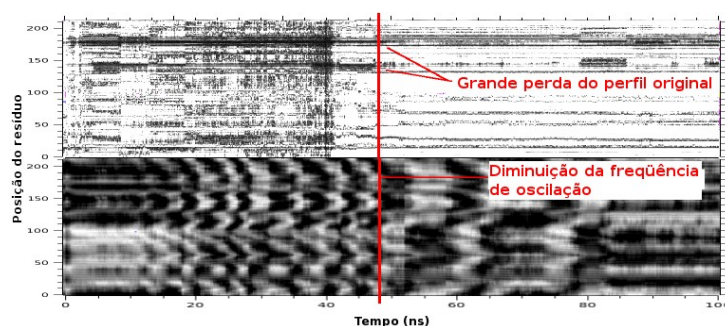


Figura 59: Relação entre o perfil de interação e o padrão de oscilação MOD_GYR da papaína complexada.

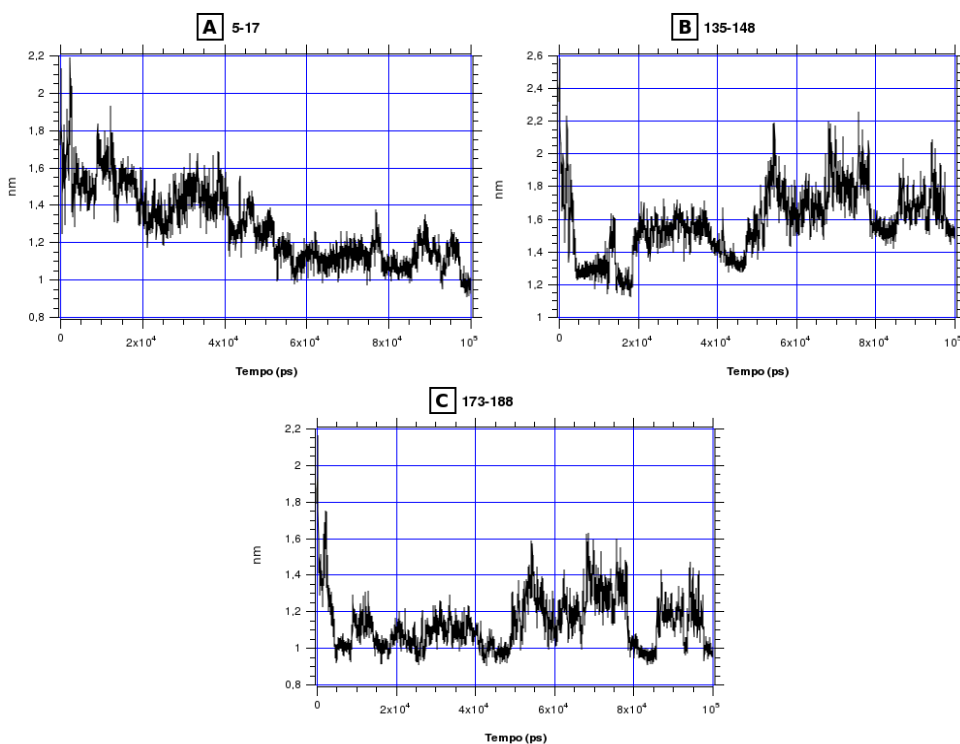


Figura 60: Distância dos átomos de oxigênio da heparina às regiões 5-17 (A), 135-148 (B) e 173-188 (B) da papaína.

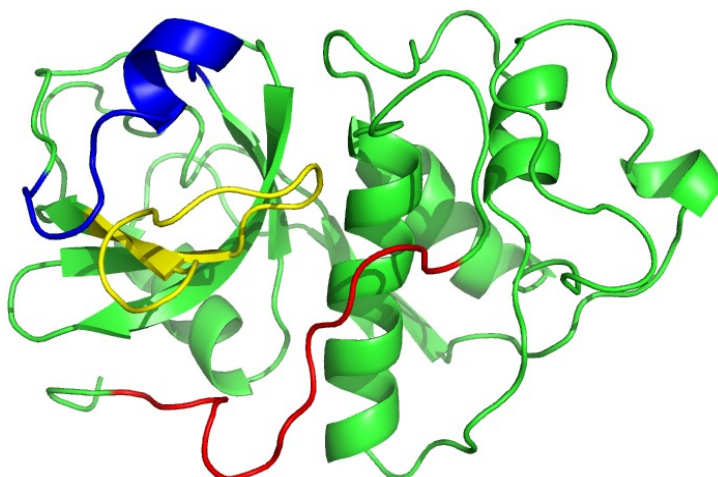


Figura 61: Representação da papaína. As regiões 5-17, 135-148, 173-188 estão representadas em vermelho, azul e amarelo, respectivamente.

5.2.4.2. RMSD Entre Trajetórias

Com base em todos os resultados apresentados anteriormente, relações importantes foram criadas. Nessa seção, em específico, a grande questão é conhecer melhor as alterações envolvidas no processo do aumento do conteúdo de α -hélice e da diminuição da acessibilidade da HIS¹⁵⁹, ambos processos ocorridos sobre a influência da heparina. As interações estruturais são complexas e difíceis de serem analisadas. No caso, foi feita uma análise RMSD entre as trajetórias da dinâmica molecular da papaína em diferentes situações: (1) “papaína livre em pH 5 *versus* papaína complexada” e (2) “papaína livre em pH 7 *versus* papaína complexada”.

Os gráficos RMSD tridimensionais (figura 62-A e B), são visualmente difíceis de serem analisados. Porém, de forma qualitativa, podê-se observar que o gráfico da “papaína livre em pH 7 *versus* papaína complexada” apresentou uma coloração mais escura, sugerindo uma maior similaridade entre as estruturas de ambas essas trajetórias, o que já era esperado, visto que a estrutura e o campo de força da papaína complexada é a mesma da papaína em pH 7. Por outro lado, o gráfico da “papaína livre em pH 5 *versus* papaína complexada” apresentou uma coloração mais clara, sugerindo diferenças maiores, porém, apresentou também, em menor grau, regiões escuras, as quais sugerem semelhanças estruturais. Para melhor analisar a influência das similaridades estruturais sobre os resultados observados anteriormente em outras seções, como a diminuição da acessibilidade da HIS¹⁵⁹, foi feita uma somatória do RMSD tridimensional em função do tempo da papaína complexada (figura 62-C

e D). A comparação da somatória (figura 63-A) de ambas as matrizes mostrou o que já era esperado, ou seja, a trajetória da papaína complexada é mais similar à trajetória da papaína livre em pH 7 do que à trajetória da papaína livre em pH 5. O gráfico de similaridade entre as trajetórias da papaína complexada e da papaína em pH 5 foram relacionados aos resultados da acessibilidade da HIS¹⁵⁹ (figura 63-B) e, interessante, foi verificado uma alta relação entre os momentos de maior similaridade estrutural e a baixa acessibilidade da HIS¹⁵⁹. Ou seja, quando a diminuição da acessibilidade da HIS¹⁵⁹ é observada para a papaína complexada, a sua estrutura apresenta grandes semelhanças com as estruturas da papaína livre em pH 5 livre.

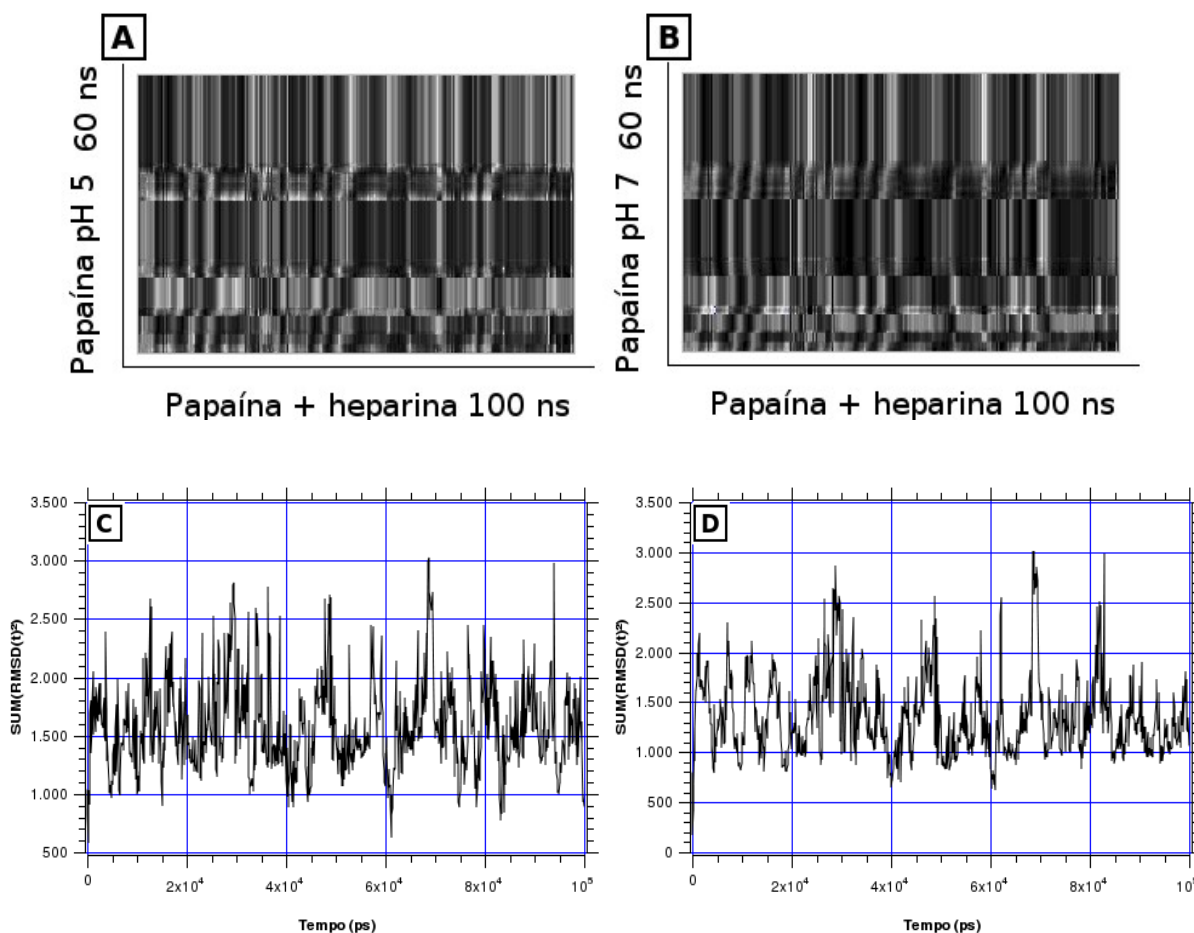


Figura 62: Gráfico RMSD tridimensional entre as trajetórias da papaína em pH 5 e complexada (A) e entre a papaína em pH 7 e complexada (B). Somatória dos valores RMSD obtidos em função do tempo do complexo papaína-heparina: complexo *versus* papaína livre em pH 5 (C), e complexo *versus* papaína livre em pH 7(D).

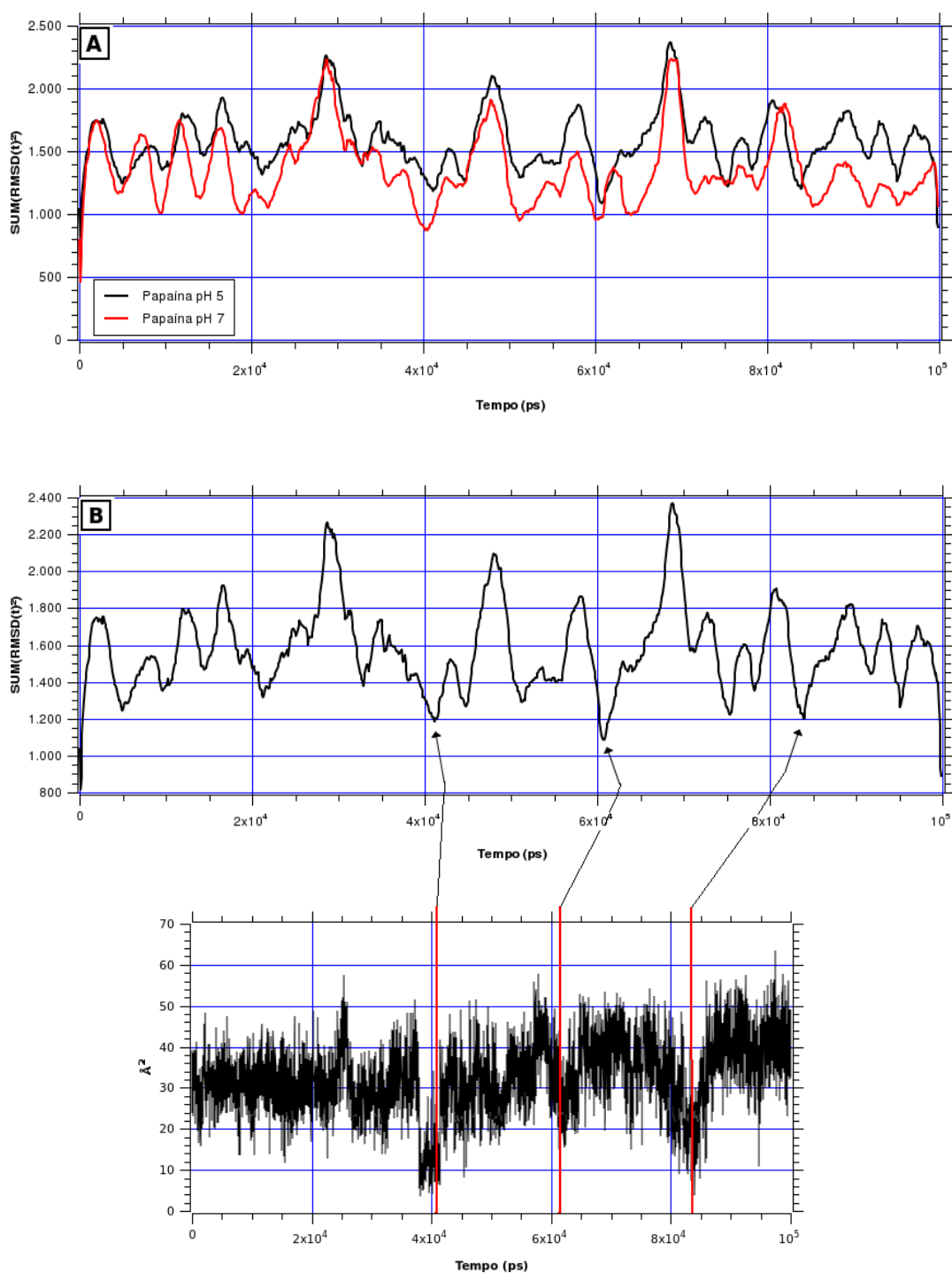


Figura 63: Gráfico suavizado (médias obtidas a cada 20 posições) e comparativo da somatória dos valores RMSD em função do tempo do complexo papaína-heparina (A), e gráfico da somatória dos valores RMSD entre a papaína em pH 7 e complexada em função do tempo do complexo e a sua relação com a acessibilidade do resíduo HIS¹⁵⁹ (B).

5.2.4.1. Distância da Heparina aos Resíduos Histidina

Para averiguar a possível neutralização dos efeitos eletrostáticos das histidinas desprotonadas da papaína pela heparina, foi calculado a distância entre os átomos de oxigênio da heparina e o resíduo HIS⁸¹ (figura 64), o único resíduo histidina encontrado na papaína além da histidina do sítio ativo.

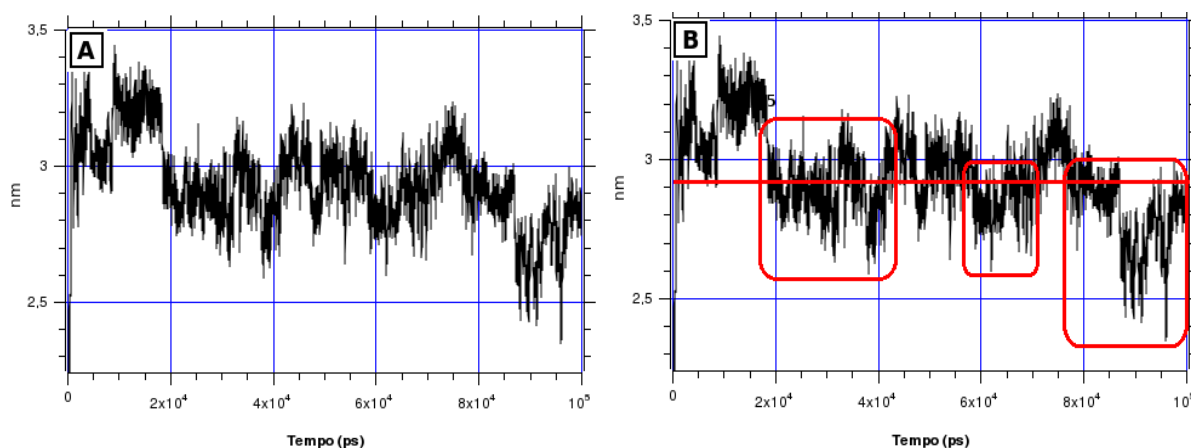


Figura 64: Distância entre os átomos de oxigênio da heparina e o resíduo HIS⁸¹ da papaína (A). Em destaque, no gráfico B, momentos de menor distância identificados entre as partes citadas anteriormente.

Analisando o gráfico de distância entre os átomos de oxigênio da heparina e o resíduo HIS⁸¹ (figura 64), foi possível verificar momentos de maior aproximação entre essas partes. Uma das aproximações de maior relevância foi verificada no intervalo entre 20 e 40 ns. Essa aproximação poderia explicar os efeitos estruturais observados nos tempos próximos de 40 ns. Outra importante aproximação ocorreu entre 58 e 70 ns, o que também explicaria os efeitos estruturais observados para os tempos próximos de 60 ns, embora não tenham sido consideradas anteriormente relevantes. Finalizando, no intervalo entre 75 ns e 100 ns, foi observado uma aproximação relevante da histidina pelos oxigênios da heparina, também explicando as alterações observadas para os tempos próximos de 84 ns. Além do mais, essa aproximação mostrou uma grande intensificação após o tempo de 85 ns sem apresentar relações com alterações estruturais relevantes, conforme os dados obtidos até o momento. No geral, todas as aproximações mostraram fortes relações com as alterações estruturais observadas para a papaína, sendo a principal delas o aumento da semelhança da estrutura da enzima complexada em pH 7 pela a respectiva estrutura encontrada em pH 5.

5.3. CATEPSINA B

Na preparação dos sistemas da catepsina B, 10.628 moléculas de água foram incluídas na caixa de simulação da catepsina B em pH 5 e 10.622 moléculas de água na caixa da catepsina B em pH 7. Para equilibrar as cargas dos sistemas, foi necessário a inclusão de 3 íons Na⁺ no sistema da catepsina B em pH 5 e 10 íons Na⁺ no sistema da catepsina B em pH 7. As energias e forças obtidas para esses sistemas nas etapas de minimização de energia estão presentes na tabela 5.

Tabela 5: Energia potencial obtida nas diversas etapas do processo de minimização de energia da catepsina B.

Catepsina B – pH 5	Energia Potencial (kJ mol ⁻¹)
prep. Steepest Descent	-5.5412303837e+05
Steepest Descent	-6.0402733260e+05
Conjugat Gradient	-6.0464378171e+05
Catepsina B – pH 7	
prep. Steepest Descent	-5.5504456906e+05
Steepest Descent	-6.0492972440e+05
Conjugat Gradient	-6.0553630405e+05

*Obs: minimização por SD com restrição de átomos pesados.

5.3.1. OBTENÇÃO DOS COMPLEXOS

Diversos complexos foram obtidos para a “catepsina B mais heparina”. Todos os *dockings* convergiram para diferentes coordenadas espaciais. As diferenças entre a energia inicial e final dos complexos se mostraram mínimas, sugerindo que não foi obtido um complexo promissor em pH 7 capaz se manter estável durante uma dinâmica molecular. Para maiores detalhes, a tabela 6 apresenta a relação entre essas energias, assim como a temperatura e o número do ciclo em que o complexo foi obtido.

A figura 65 mostra a superfície eletrostática da catepsina B. A região mais escura, em azul, é uma região bastante eletropositiva, e é potencialmente um dos locais preferíveis para a ligação da heparina, que é uma molécula bastante eletronegativa. Essa região é caracterizada

por se encontrar no domínio L e na face frontal da enzima.

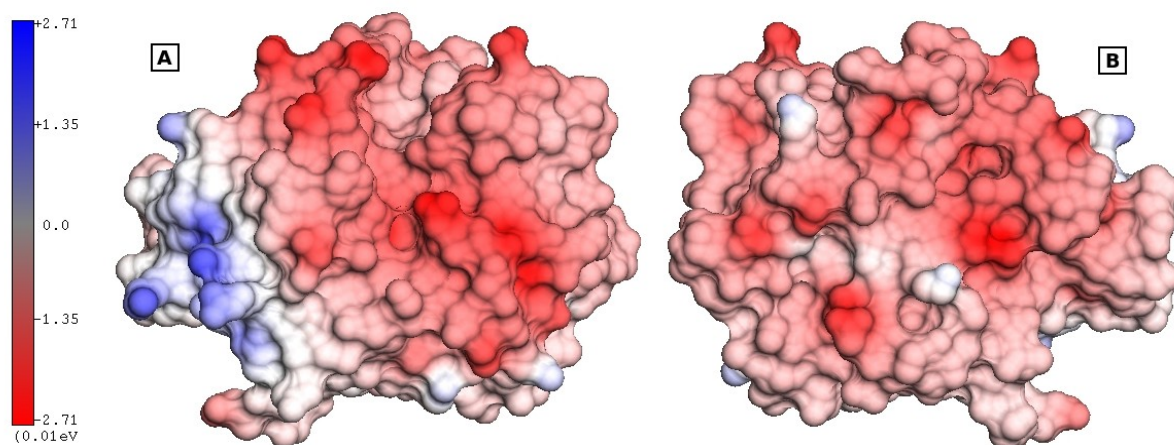


Figura 65: Superfície eletrostática da catepsina B. Visão frontal (A) e visão traseira (B).

Tabela 6: Dados obtidos nos *dockings* (GSA) feitos entre a catepsina B e a heparina.

Complexo	E. inicial (kcal/mol)	E. final (kcal/mol)	Temperatura (K)	Ciclo N°
1	682613.25	672563.65	0.181081	88352
2	682166.20	674296.03	0.106133	189490
3	682166.20	619063.54	4.71E-02	603937
4	682340.36	682017.77	3.92E-02	784990
5	682166.20	682059.19	0.146425	119666
6	682166.20	681981.00	0.111561	176462
7	680223.20	665867.38	8.02E-02	282589
8	680223.20	320270.045	6.14E-02	413663

Alguns dos complexos obtidos entre a catepsina B e a heparina estão representados na figura 66. Os complexos obtidos mostraram que a heparina pode se complexar à diversas regiões da catepsina B, como por exemplo, nos domínios L ou R, e nas faces frontal ou traseira. Grande parte do complexos mostraram uma maior afinidade por uma região eletrostaticamente negativa (domínio R, face traseira), o que sugere que a região eletropositiva da catepsina B não é necessariamente um sítio putativo de ligação, supostamente por não ser energeticamente favorável.

Para verificar a estabilidade dos complexos obtidos entre a catepsina B e a heparina, foram feitas análises preliminares através da dinâmica molecular para selecionar os melhores complexos, os quais deverão ser analisados posteriormente por uma MD de longa duração, assim como feito para a papaína.

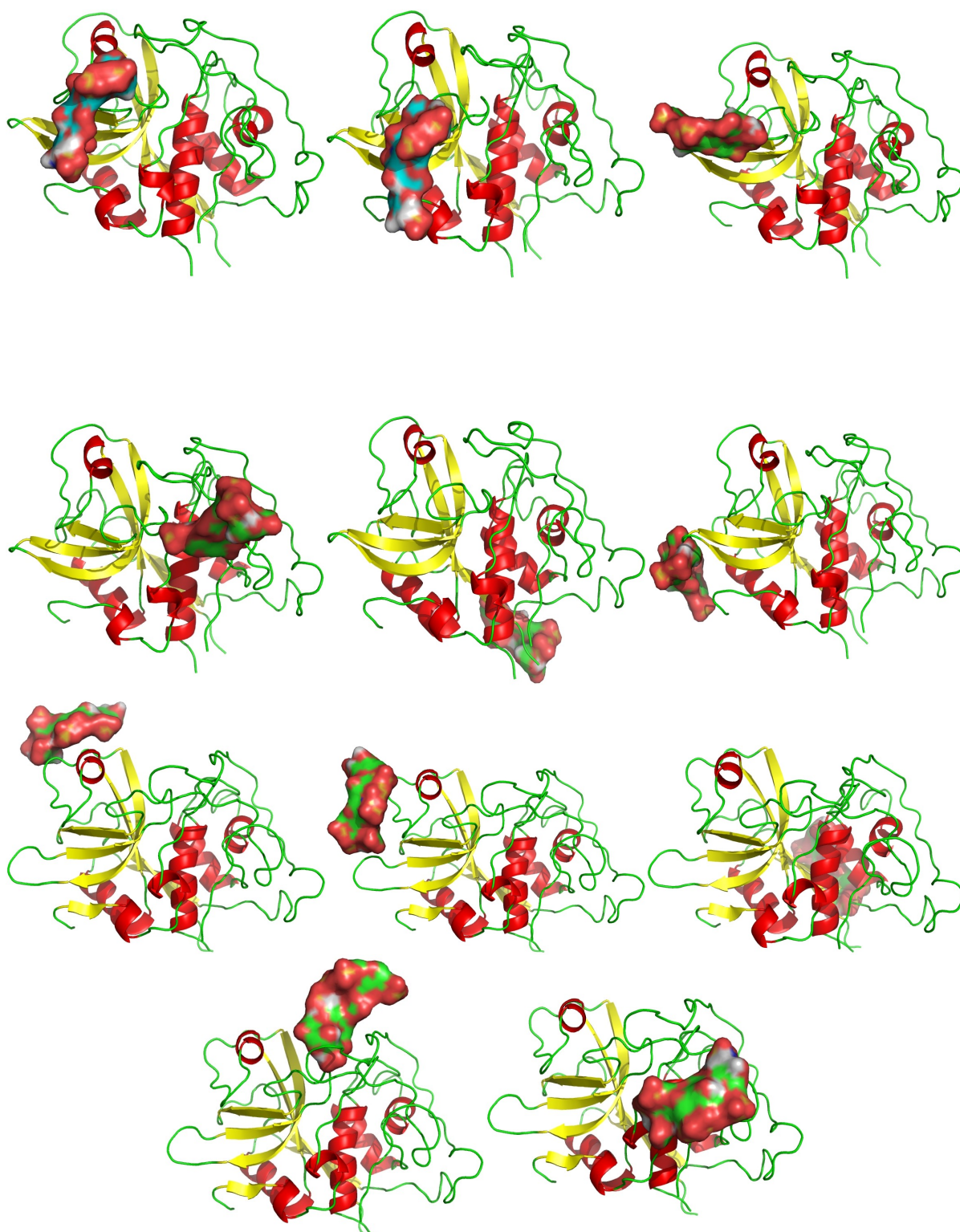


Figura 66: Alguns complexos obtidos para a catepsina B e heparina através do THOR/GSA *docking*.

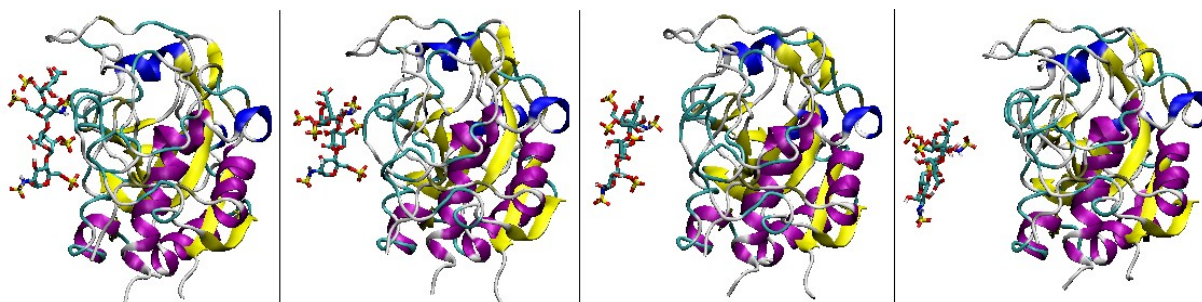


Figura 67: Representação gráfica da trajetória de curta duração por MD criada a partir do primeiro complexo apresentado na figura 86.

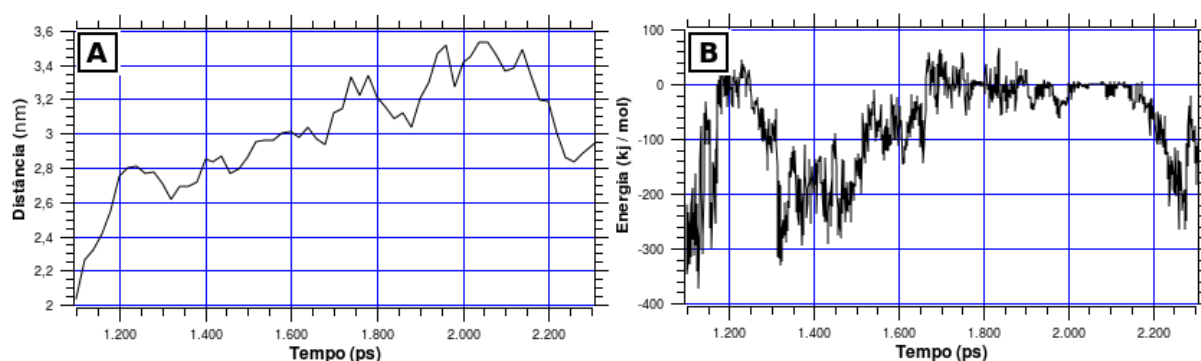


Figura 68: Gráficos de distância (A) e energia de contato (B) entre a heparina e a catepsina B baseados na dinâmica molecular do primeiro complexo apresentado na figura 66.

As análises preliminares por MD dos complexos obtidos pelo GSA mostraram que não foi encontrado um complexo estável entre a heparina e a catepsina B em pH 7. As figuras 67 e 68-A-B apresentam os resultados obtidos na MD do primeiro complexo apresentado na figura 66. Foi claramente evidenciado o distanciamento e o aumento da energia de contato entre a heparina e a catepsina B, sugerindo incompatibilidades e um sítio de ligação não putativo para esse complexo e os demais similares. Esses dados sugerem a necessidade de alterações metodológicas para esse caso. Uma tentativa de solução foi fazer o *docking* utilizando a catepsina B em pH 5, onde as histidinas se encontram protonadas e carregadas positivamente. O GSA, novamente, não identificou a região mais eletropositiva da enzima como um possível sítio de ligação, e repetiu os resultados já obtidos em pH 7. Insistentemente, o GSA foi parametrizado para forçar o *docking* da heparina na região mais eletropositiva da catepsina B, processo o qual ocorreu com sucesso, porém, sem uma minimização energética relevante. Interessantemente, através de dados preliminares, a MD desse complexo apresentou uma alta estabilidade, sugerindo um bom alvo para o estudo de longa duração por MD. Esse complexo, portanto, foi utilizado para as demais análises desse trabalho.

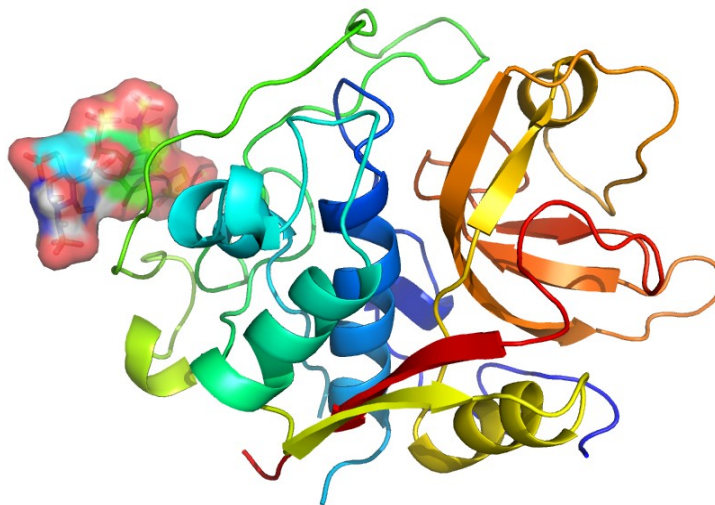


Figura 69: Complexo catepsina B-heparina encontrado pelo THOR/GSA

O complexo obtido em pH 5 mais promissor (figura 69) mostrou que a heparina se complexou próximo, mas não sobre à região mais eletropositiva da catepsina B, que se encontra predominantemente na parte frontal do domínio L. Os resíduos da catepsina B encontrados em contato com os átomos de oxigênio da heparina dentro de um raio de 5 Å são a GLU⁹³, HIS⁹⁵, PRO¹⁰⁰, LYS¹²⁵, SER¹²⁷, LYS¹²⁸, ILE¹²⁹, GLU¹³¹, PRO¹³², GLY¹³³ e GLN¹⁸⁷. Desses resíduos, 27% (3/11) são básicos (LYS e HIS). Dentro de um raio de 3 Å, o único resíduo da catepsina B encontrado em contato com os átomos de oxigênio da heparina foi a GLU¹³¹, que é um resíduo ácido que, no caso, deve ter interagido com o átomo enxofre de um dos sulfatos do açúcar, o que explicaria a proximidade dos átomos de oxigênio por esse resíduo ácido que apresenta uma carga eletrostaticamente negativa, assim como também apresentada pelos átomos de oxigênio. Aparentemente, através da metodologia utilizada no *docking* molecular desse trabalho, os resíduos ácidos também podem ser importantes para a estabilização da heparina na catepsina B, o que pode desclassificar regiões eletropositivas isoladas como sendo necessariamente um sítio putativo de ligação para a heparina. Essa informação pode explicar o fato dos complexos obtidos anteriormente terem apresentado contatos próximos entre a heparina e regiões eletronegativas da enzima. Porém, como já observado anteriormente nas análises preliminares, os complexos dessa natureza apresentaram energias de interação positivas ao longo da MD, o que caracteriza a expulsão do ligante da enzima. Ou seja, os resultados de *docking* por si próprios não são fisicamente relevantes.

5.3.2. ANÁLISE ESTRUTURAL

5.3.2.1. RMSD

O desvio médio quadrático (RMSD) foi medido em função do tempo para a catepsina B livre em pH 5, livre em 7 e complexada em pH 5 (sistema contendo 10044 moléculas de H₂O e 11 de Na⁺ 11) (figura 70). O gráfico RMSD da catepsina B em pH 5 (figura 70-A) mostrou que a estrutura da enzima se estabilizou a partir de aproximadamente 20 ns de termalização. O RMSD da catepsina B dos demais casos, livre em pH 7 (figura 70-B), e complexada em pH 5 (figura 70-C), apresentaram resultados bastante similares, e se estabilizaram após o mesmo tempo de termalização. As diferenças observadas entre os gráficos RMSD da catepsina B foram mínimas (figura 70-D). Não há evidências sugestivas de restrição estrutural sofrida pela estrutura da catepsina B complexada à heparina como a observada no estudo da papaína.

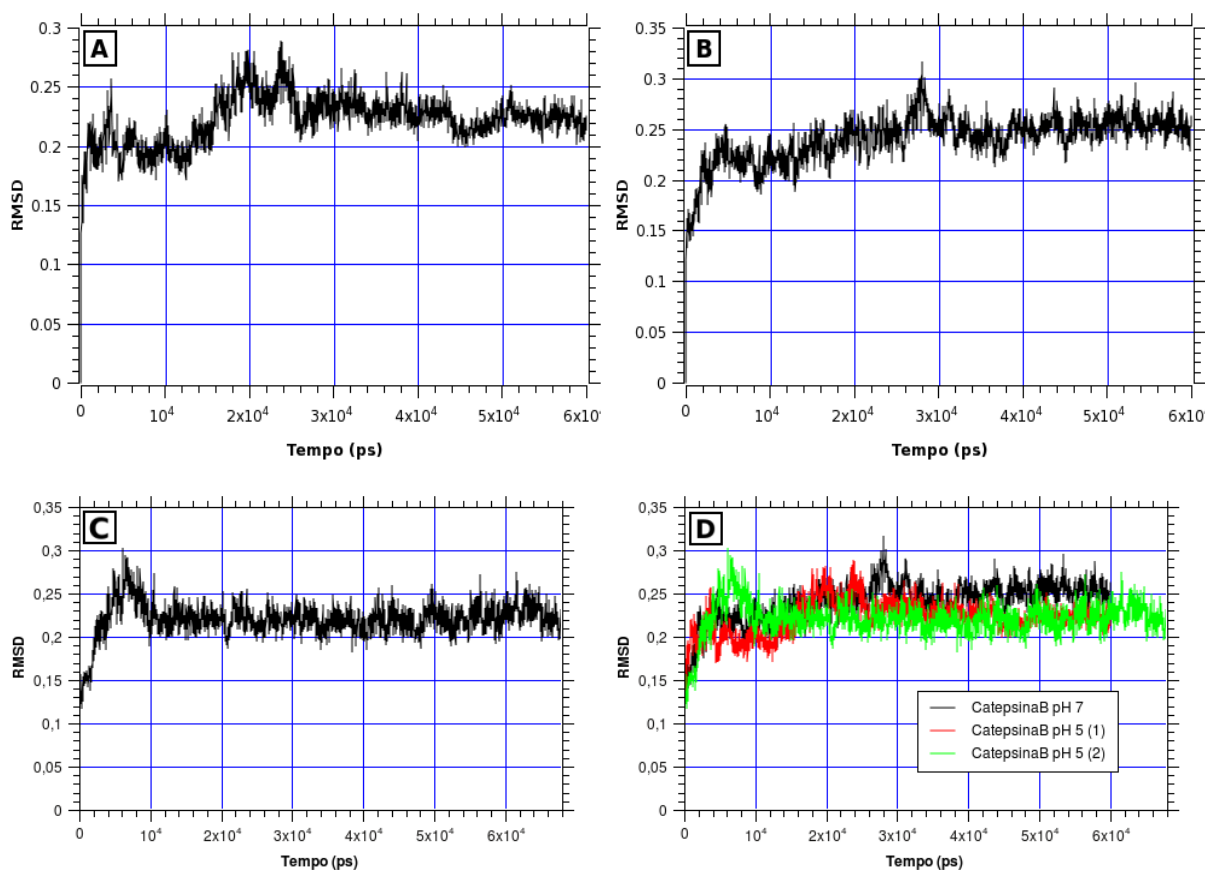


Figura 70: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises RMSD anteriores (D).

O gráfico RMSD da heparina complexada (figura 71-B) apresentou resultados semelhantes ao observado para a heparina no estudo da papaína. Utilizando como referência os valores RMSD obtidos para a heparina livre (figura 71-A), foi verificado que a heparina complexada, tanto à papaína quanto à catepsina B (figura 71-B), apresentou níveis menores de oscilação estrutural, que é provavelmente o resultado de um relevante contato intermolecular com a enzima, capaz de restringir parcialmente a dinâmica do açúcar.

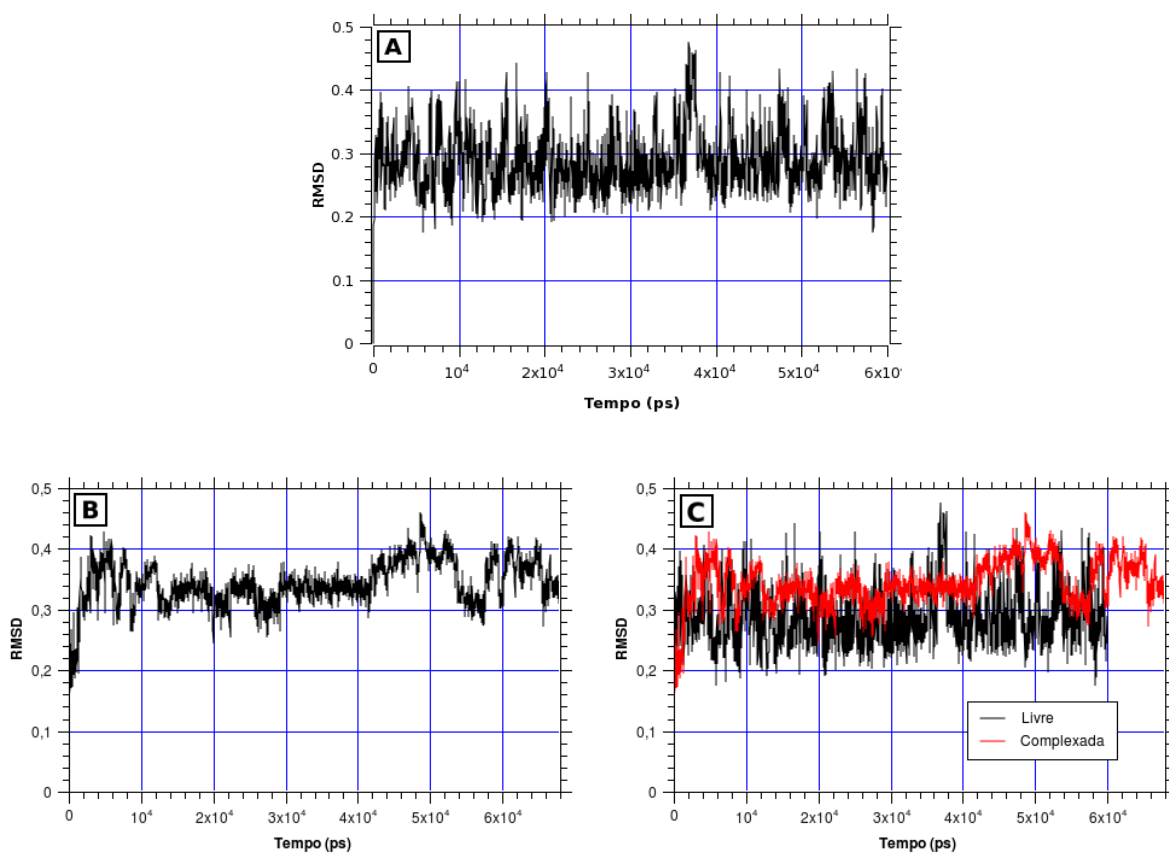


Figura 71: RMSD em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 5 (B). Comparação entre as análises RMSD anteriores (C).

5.3.2.2. Raio de Giração

O raio de giração foi medido em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 72). Na análise da catepsina B em pH 5 (figura 72-A), a partir de 40 ns de dinâmica, momento em que os valores das diferentes análises se

divergiram e aparentemente se estabilizaram, a média do raio de giração foi de 1,698 nm, com um desvio padrão de 0,005 nm, e um valor máximo de 1,718 e mínimo de 1,684 nm. Para a catepsina B em pH 7 (figura 72-B), a partir do mesmo tempo, a média foi de 1,740 nm, com um desvio padrão de 0,006 nm, e um valor máximo de 1,763 e mínimo de 1,721 nm. Para a catepsina B complexada em pH 5 (figura 72-C), entre 0 e 68 ns, a média foi de 1,732 nm, com um desvio padrão de 0,007 nm, e um valor máximo de 1,759 e mínimo de 1,700 nm. Em ordem decrescente das médias de raio de giração, temos a catepsina B livre em pH 7, complexada e livre em pH 5. O resultado da catepsina B livre em pH 7 foi mais similar ao observado para a catepsina B complexada em pH 5 (figura 72-D), o que sugere que essa enzima pode sofrer a ação da heparina mesmo em pH 5, que é, nesse caso, um pH ideal para uma ótima atividade enzimática.

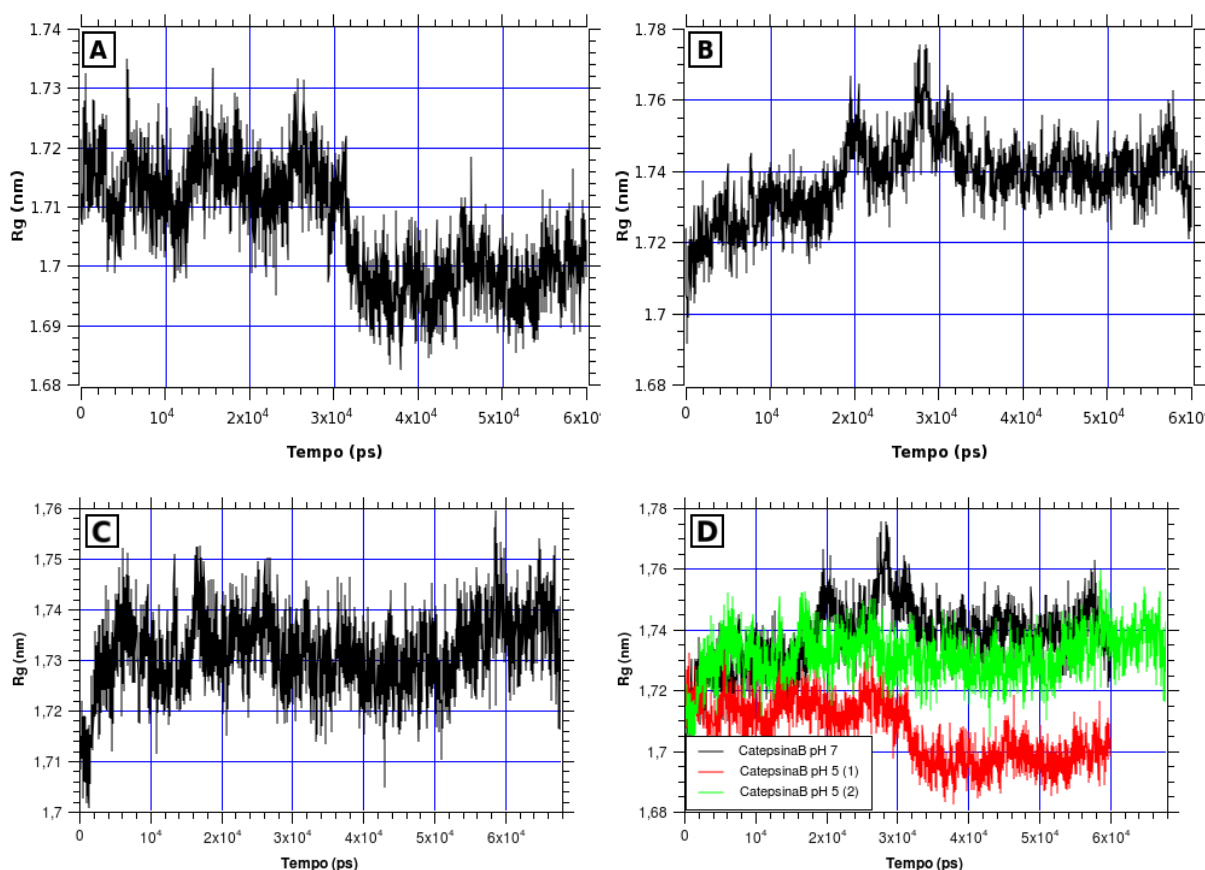


Figura 72: Raio de giração em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C) . Comparação entre as análises de raio de giração anteriores (D).

5.3.2.3. Raio de Giração Modificado por Resíduo (MOD_GYR)

O MOD_GYR por resíduo foi medido em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 73). Chama a atenção o fato de o MOD_GYR por resíduo da catepsina B, em todos os casos, apresentar diferentes frequências de oscilação na estrutura da enzima, assim como já verificado anteriormente para a papaína. Por ordem decrescente de frequência de oscilação temos a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5. Essa relação oscilatória entre os diferentes pHs, assim como o perfil, frequência, e estabilidade das oscilações (figuras 73, 74, 75 e 76), se mostraram altamente compatíveis com os resultados obtidos para a papaína. Porém, no caso da catepsina B complexada, foi verificado um aumento da instabilidade oscilatória. Aparentemente, esse é o resultado de uma baixa interferência da heparina na interação entre os domínios da enzima.

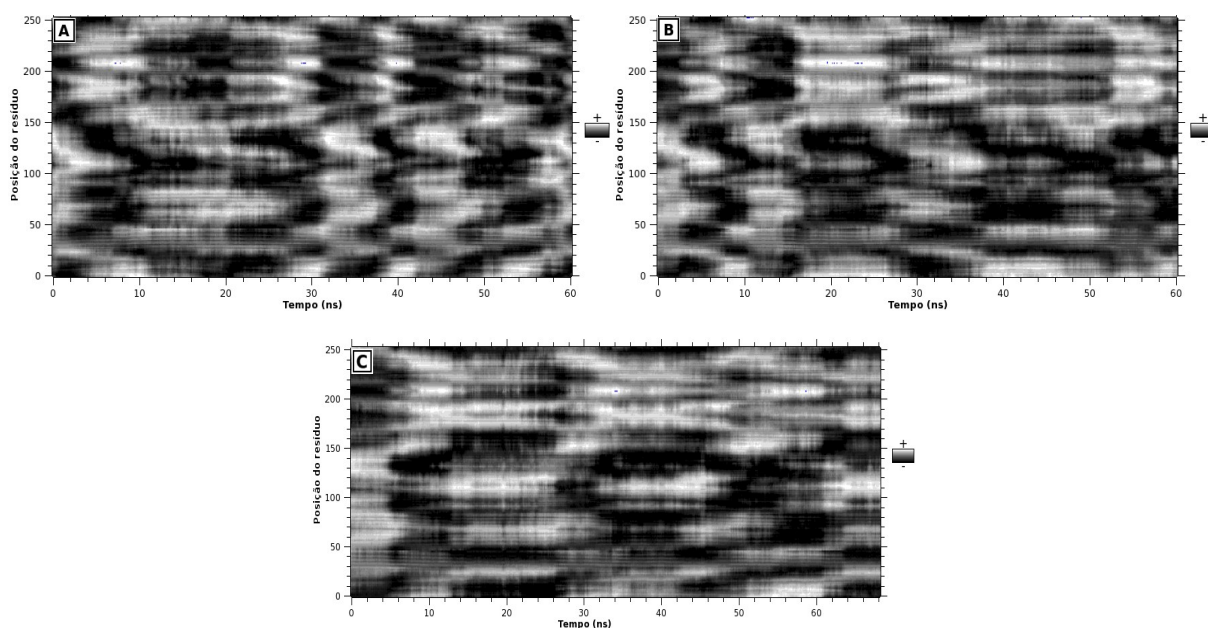
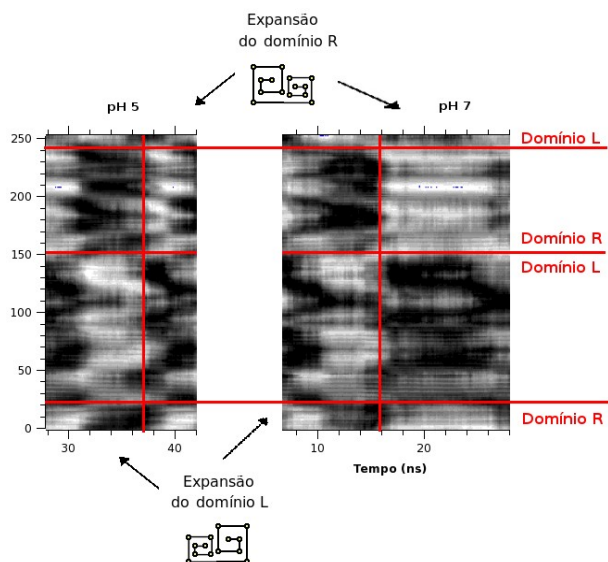


Figura 73: MOD_GYR por resíduo em função do tempo da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C).

Até esse ponto, os resultados sugerem que o pH é um fator que apresenta uma grande influência na estabilidade dinâmica dessa classe de enzima. Essa sugestão foi fortemente suportada pelas análises MOD_GYR por resíduo das simulações da papaína e catepsina B. Os resultados mostraram que um pH alcalino pode criar uma desestabilização dinâmica, e um pH ácido, contrariamente, mostrou-se capaz de acelerar e estabilizar o perfil dinâmico de ambas

as enzimas. A presença da heparina por sua vez, mostrou interferir positivamente ou negativamente sobre a dinâmica da enzima, criando possivelmente, em ambos os casos, uma diminuição da atividade enzimática.



Obs: A seqüência primária e os domínios da catepsina B não podem ser diretamente relacionados, uma vez que determinadas partes da seqüência podem ser encontradas em diferentes domínios, não seguindo, exatamente, uma separação linear; como observado na figura ao lado. O mesmo acontece para a papaína, porém, em um grau muito menor, ao ponto dessa característica ter sido desconsiderada nas respectivas análises.

Figura 74: Detalhes do MOD_GYR por resíduo da catepsina B.

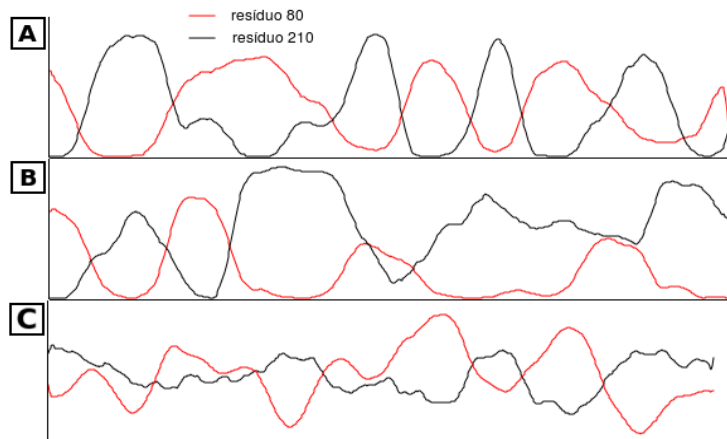


Figura 75: Análise MOD_GYR por resíduo feita isoladamente para os resíduos 80 e 210 da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 7 (C).

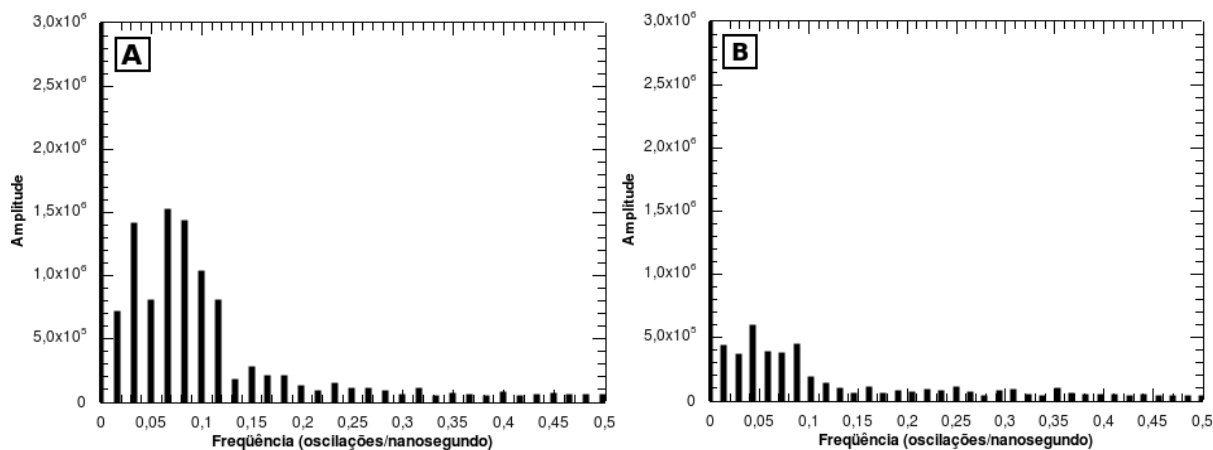


Figura 76: Transformada de Fourier aplicada para as oscilações MOD_GYR do resíduo 80 da catepsina B em pH 5 (A) e complexada em pH 7 (B).

5.3.2.4. Energia Potencial

A energia potencial foi medida em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 77). A análise da energia potencial da catepsina B mostrou que nesses pHs, a energia se manteve praticamente estável com pequenas tendências, como uma leve diminuição da energia potencial observada no final dos gráficos da catepsina B livre em pH 5 e 7. Após o tempo de 40 ns, os padrões observados no gráfico da energia potencial da catepsina B em todos os casos analisados são compatíveis com os padrões verificados na análise de raio de giração correspondentes. Porém, anteriormente ao tempo de 40 ns, nenhuma relação entre esses resultados foi identificada.

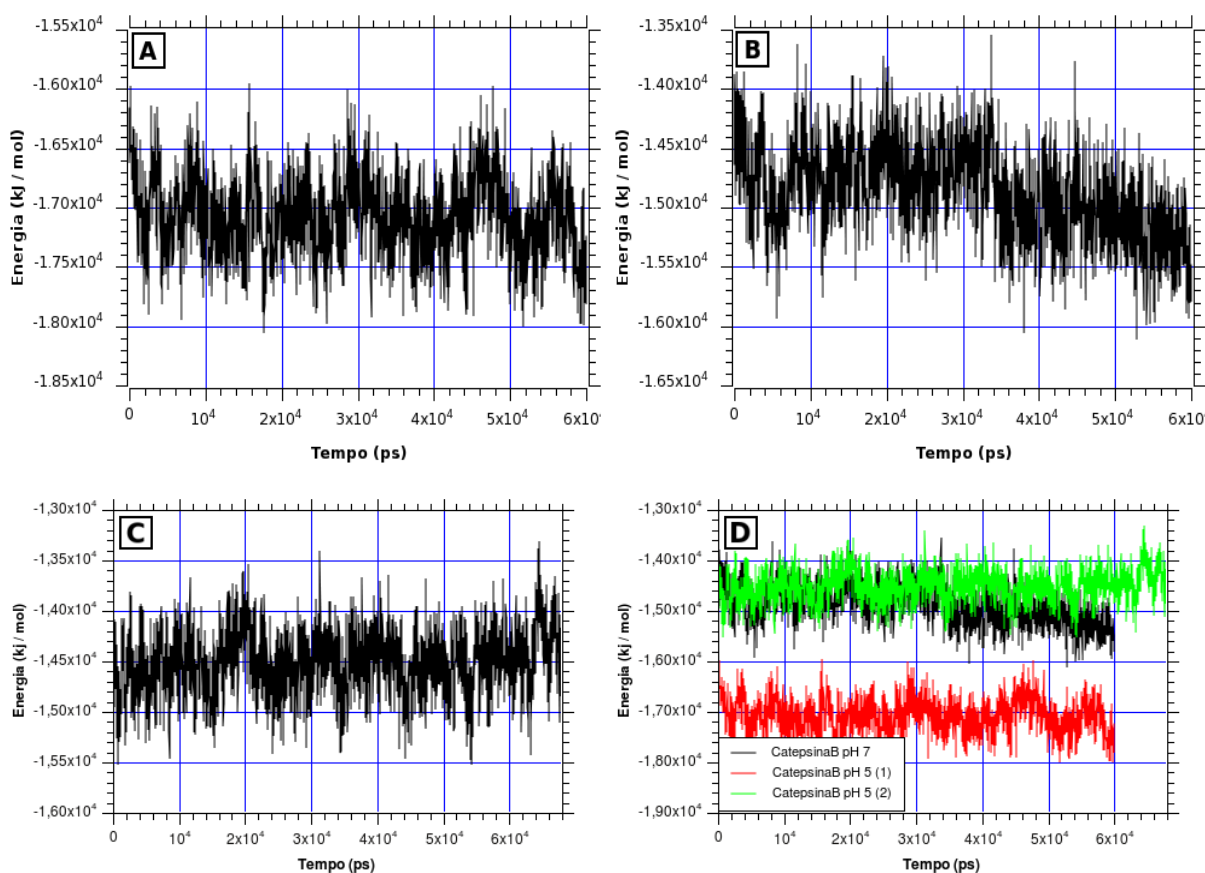


Figura 77: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de energia potencial anteriores (D).

Na análise da catepsina B livre em pH 5 (figura 77-A), durante toda a dinâmica, a média da energia potencial foi de -17072,8 kJ/mol, com um desvio padrão de 338,4 kJ/mol, e um valor máximo de -15954,1 e mínimo de -18053,9 kJ/mol. Para a catepsina B livre em pH 7 (figura 77-B), a média foi de -14873,9 kJ/mol, com um desvio padrão de 381,6 kJ/mol, e um valor máximo de -13542,4 e mínimo de -16112 kJ/mol. Para a catepsina B complexada em pH 5 (figura 77-C), a média foi de -14516,6 kJ/mol, com um desvio padrão de 319,0 kJ/mol, e um valor máximo de -13315,3 e mínimo de -15876,9 kJ/mol. Em ordem decrescente de energia potencial, temos a catepsina B complexada em pH 5, livre em pH 7 e livre em pH 5. A alta energia potencial apresentada pela catepsina B complexada sugere um desconforto estrutural e a influência da heparina sobre a enzima, assim como também observado no estudo da papaína. Porém, esses dados não explicam a baixa influência da heparina nos resultados RMSD da catepsina B. Aparentemente, os mecanismos que modulam as alterações estruturais da catepsina B são diferentes dos observados para papaína.

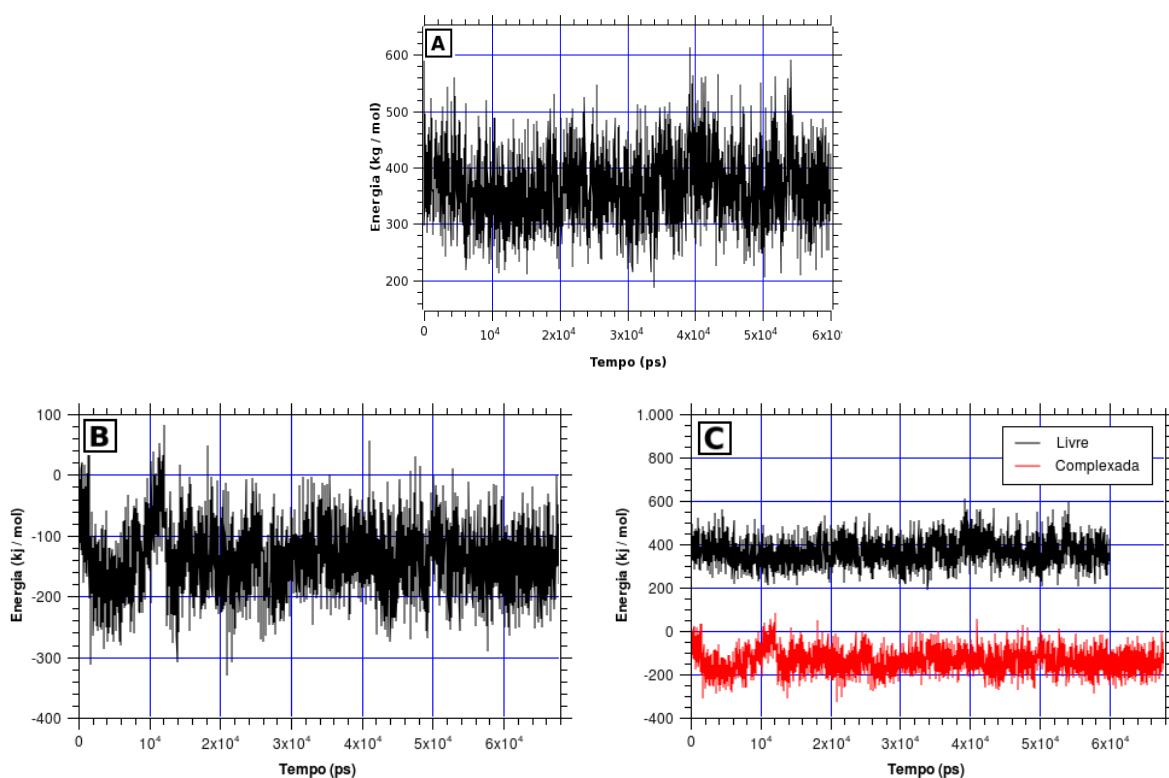


Figura 78: Energia potencial medida em função do tempo da trajetória da dinâmica molecular da estrutura da heparina livre em pH 7 (A) e complexada em pH 5 (B). Comparação entre as análises de energia potencial anteriores (C).

Quanto à energia potencial da heparina livre e complexada à catepsina B (figura 78), ela mostrou oscilar em torno de uma média bastante estável, sem alterações consideráveis. Detalhadamente, a análise da heparina complexada (figura 78-B), de 0 a 68 ns de dinâmica, a média da energia potencial foi de -139,2 kJ/mol, com um desvio padrão de 54,7 kJ/mol, e um valor máximo de 81,5 e mínimo de -329,6 kJ/mol. Para a heparina livre (figura 78-A), no mesmo período de tempo, a média foi de 364,8 kJ/mol, com um desvio padrão de 62,2 kJ/mol, e um valor máximo de 862,3 e mínimo de 187,0 kJ/mol. Em ordem decrescente de energia potencial, temos a heparina livre e a heparina complexada, sendo que a última apresentou valores negativos de energia, sugerindo que o contato da heparina com a catepsina B é capaz de aumentar a estabilidade estrutural do açúcar, diferentemente do observado para a heparina complexada à papaína.

5.3.2.5. Estrutura Secundária

O conteúdo de estrutura secundária, α -hélice, foi medido em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 79). Na análise da catepsina B em pH 5 (figura 79-A), após 30 ns de dinâmica, período em que houve a estabilização do conteúdo da estrutura α -hélice, a média obtida foi 54,3%, com um desvio padrão de 2,3%, e um valor máximo de 64,0 e mínimo de 44,0%. Para a catepsina B em pH 7 (figura 79-B), a média foi de 59,3%, com um desvio padrão de 2,9%, um valor máximo de 67,0 e mínimo de 45,0%. Para a catepsina B complexa em pH 5 (figura 79-C), a média foi de 44,6%, com um desvio padrão de 3,2%, um valor máximo de 56,0 e mínimo de 31,0%. Em ordem decrescente de conteúdo de α -hélice, temos a catepsina B livre em pH 7, livre em pH 5 e complexada em pH 5. Esse perfil se manteve estável durante toda a dinâmica. Os valores do conteúdo de α -hélice da catepsina B complexada em pH 5 se aproximaram mais dos valores obtidos para a catepsina B livre em pH 5, diferentemente da tendência observada anteriormente nas demais análises.

O conteúdo de folha- β foi medido em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 80). Na análise da catepsina B em pH 5 (figura 80-A), durante toda a dinâmica, a média do conteúdo de folha- β foi de 44,7%, com um desvio padrão de 5,2 %, e um valor máximo de 62,0 e mínimo de 27,0%. Para a catepsina B em pH 7 (figura 80-B), a média foi de 46,2%, com um desvio padrão de 4,4 %, um valor máximo de 61,0 e mínimo de 27,0 %. Para a catepsina B complexada em pH 5 (figura 80-C), a média foi de 41,1 %, com um desvio padrão de 5,6 %, uma valor máximo de 57,0 e mínimo de 21,0 %. Em ordem decrescente de conteúdo de folha- β , temos a catepsina B livre em pH 7, pH 5, e complexada em pH 5, diferentemente do observado para a papaína. Porém, os resultados em geral se mostraram bastante próximos entre si, sugerindo uma baixa influência do pH ou da heparina sobre esse tipo de parâmetro em específico.

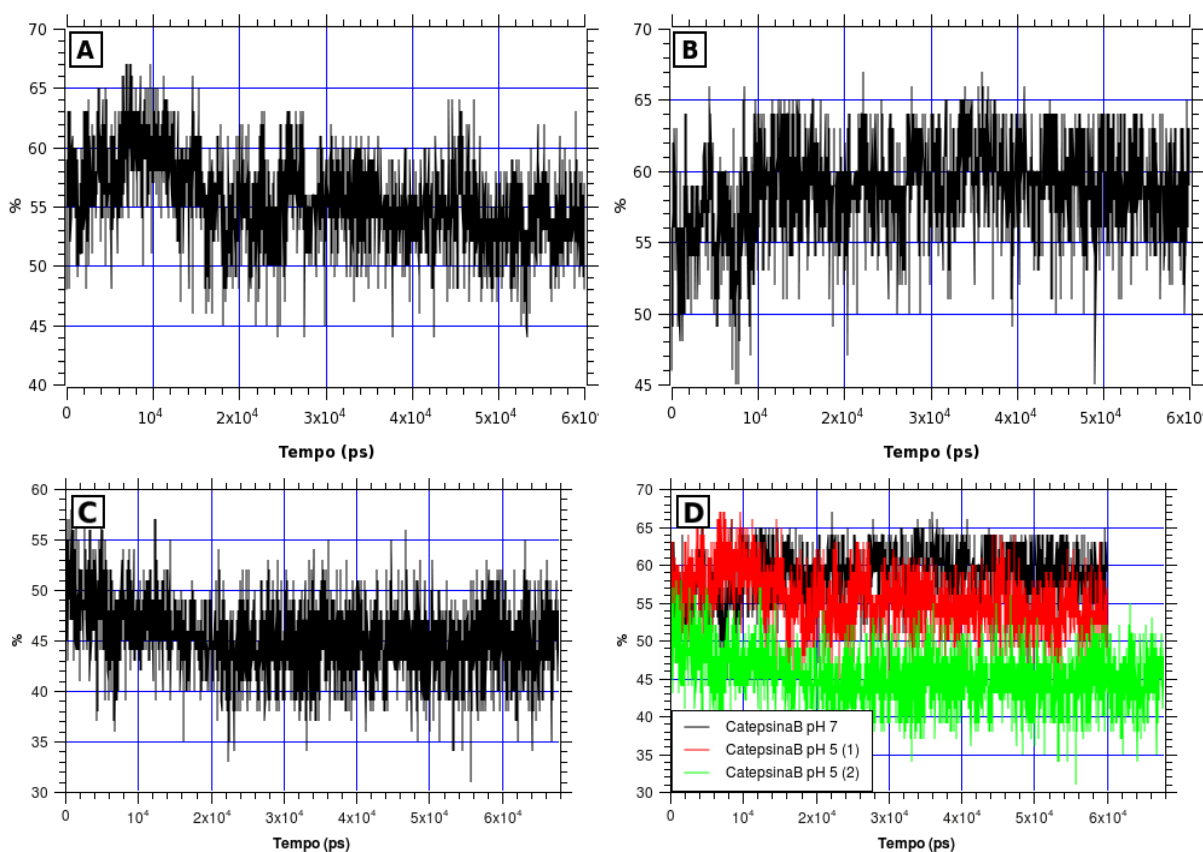


Figura 79: Conteúdo de estrutura α -hélice medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B), e complexada em pH 5 (C). Comparação entre as análises de conteúdo de estrutura α -hélice anteriores (D).

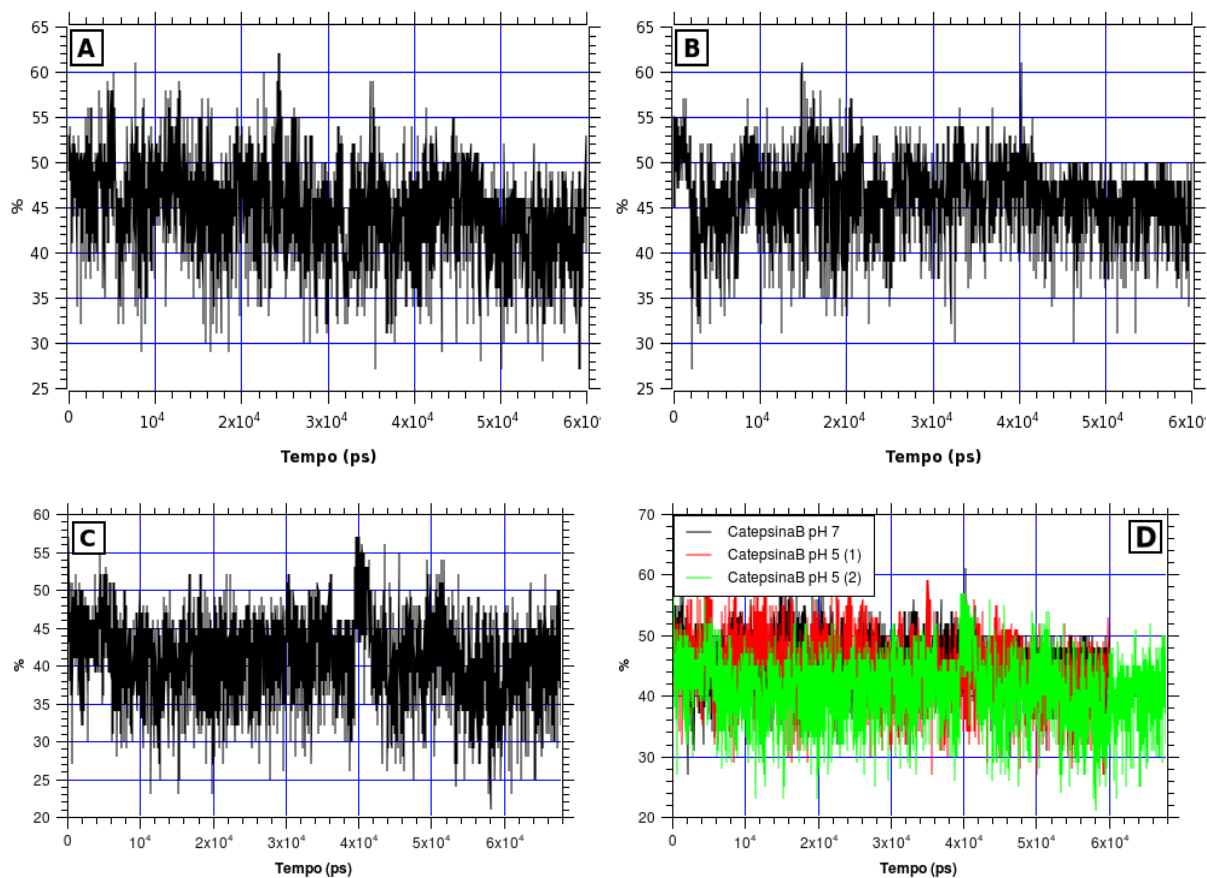


Figura 80: Conteúdo de estrutura folha- β medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de conteúdo de estrutura folha- β anteriores (D).

Diferentes tipos de estruturas secundárias também foram medidas em função dos resíduos e tempo (figura 81). Foram verificadas poucas variações estruturais de α -hélices e folhas- β em função do tempo. Em pH 5, a catepsina B livre e complexada apresentaram de forma relevante uma menor presença de α -hélice na região N-terminal inicial, especificamente entre os resíduos 1 e 20. As estruturas α -hélice da catepsina B complexada apresentaram uma diminuição generalizada de conteúdo, que é o provável resultado da influência “global” protagonizada pela heparina sobre a catepsina B. Esses dados suportam os resultados obtidos na análise do conteúdo das estruturas α -hélice e folha- β verificadas na seção anterior (4.3.2.5).

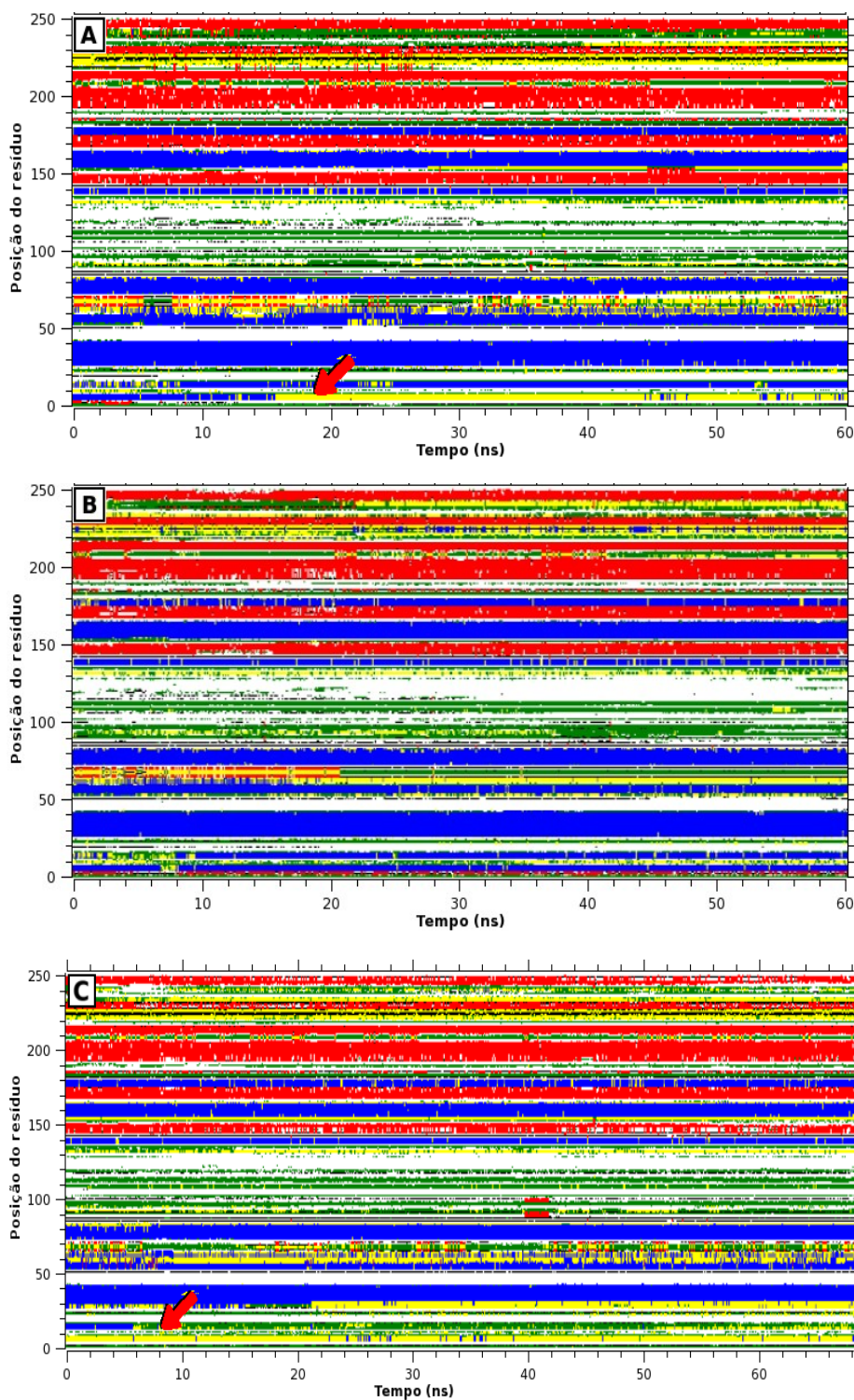


Figura 81: Estruturas secundárias medidas em função do resíduo e do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Em azul, α -hélice, e em vermelho, folha- β .

5.3.2.6. Acessibilidade ao Solvente

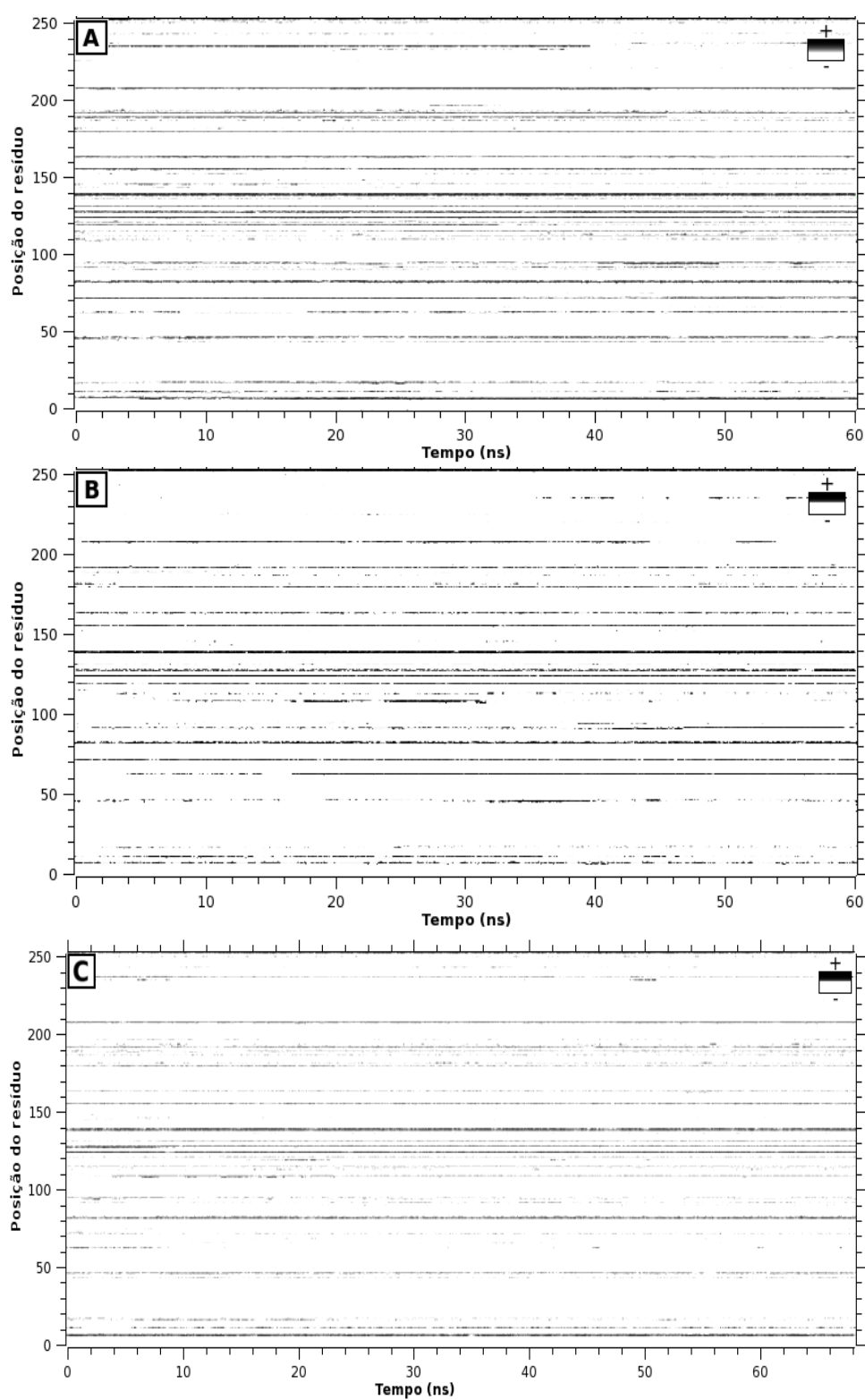


Figura 82: Acessibilidade ao solvente medida em função de cada resíduo e do tempo da trajetória da dinâmica molecular da estrutura da catepsina livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Regiões mais escuras representam uma maior acessibilidade.

A acessibilidade ao solvente dos resíduos da catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 foi medida em função do tempo (figura 82). No gráfico de acessibilidade da catepsina B em todos os casos, não foram verificadas variações conjuntas relevantes na acessibilidade dos resíduos, embora, conforme o gráfico, algumas variações sejam claramente visíveis, elas foram identificadas como casos isolados, o que significa que são ocorrências comuns não resultantes de transformações estruturais relevantes. Além do mais, esses mesmos resultados não demonstraram relações diretas e claras com os demais resultados apresentados para a catepsina B.

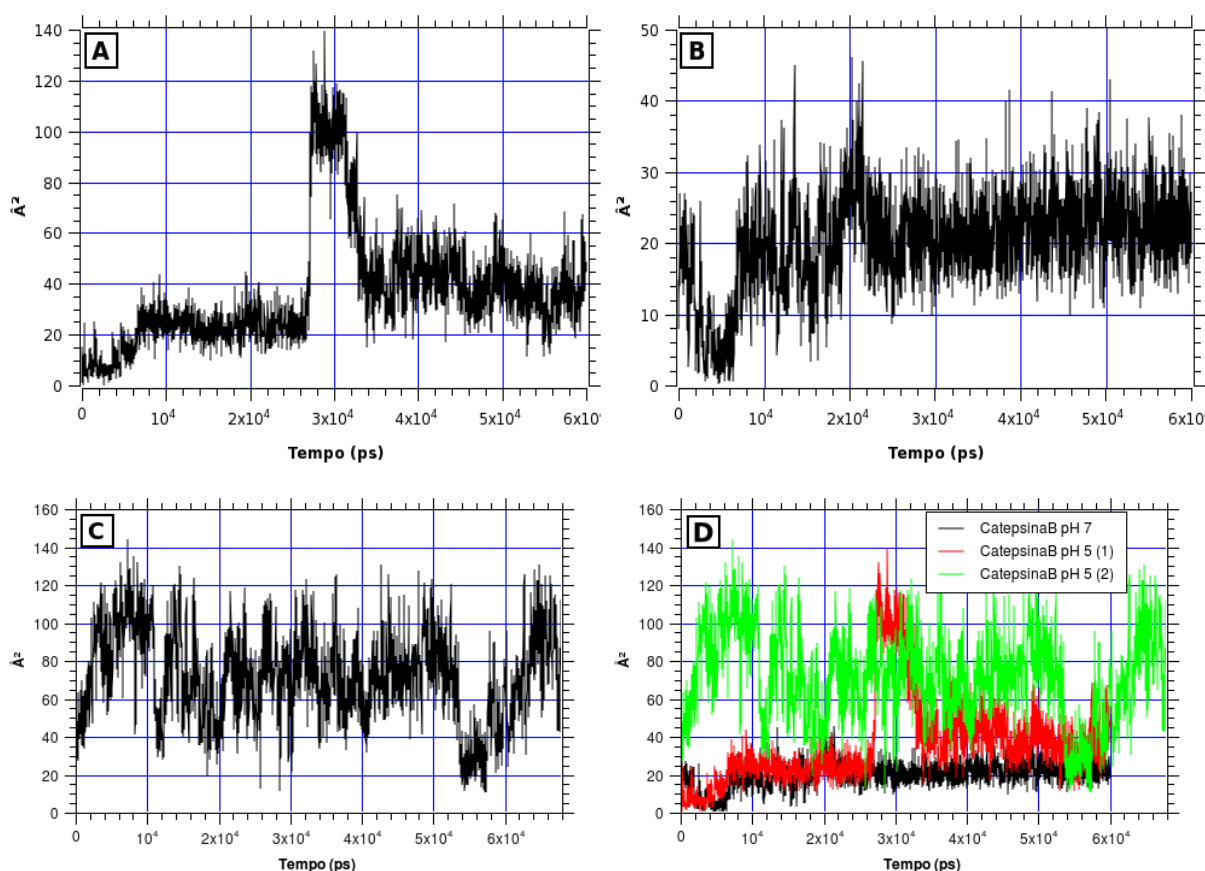


Figura 83: Acessibilidade ao solvente da HIS¹⁹⁷ medida em função do tempo da trajetória da dinâmica molecular da estrutura da catepsina B livre em pH 5 (A), livre em pH 7 (B) e complexada em pH 5 (C). Comparação entre as análises de acessibilidade da HIS¹⁹⁷ anteriores (D).

Analisando de forma mais detalhada a acessibilidade dos resíduos da catepsina B em função do tempo, foi verificada a acessibilidade da HIS¹⁹⁷, que é análoga à HIS¹⁵⁹ da papaína, e são importantes para a atividade dessas enzimas. A acessibilidade da HIS¹⁹⁷ foi medida em função do tempo para a catepsina B livre em pH 5, livre em pH 7 e complexada em pH 5 (figura 83). A catepsina B em pH 7 (figura 83-B) é a que mostrou a menor acessibilidade do

resíduo HIS¹⁹⁷ ao longo da dinâmica, ao contrário do verificado para a papaína. Após 36 ns de dinâmica, período em que as oscilações iniciais da dinâmica se estabilizaram, a acessibilidade desse resíduo apresentou uma média de 22,89 Å², um desvio padrão de 4,99 Å², e um valor máximo de 42,96 e mínimo de 9,74 Å². No caso da catepsina B em pH 5 (figura 83-A), após o mesmo tempo, a acessibilidade média da HIS¹⁹⁷ foi de 39,24 Å², com um desvio padrão de 10,38 Å², um valor máximo de 75,11 e mínimo de 11,39 Å². No caso da catepsina B complexada em pH 5 (figura 83-C), grandes oscilações foram observadas. A média foi de 71,70 Å², com um desvio padrão de 24,05 Å², um valor máximo de 143,96 e mínimo de 10,98 Å². Esse último resultado, especificamente, apresentou uma maior aproximação pelo resultado obtido para a catepsina B livre em pH 5.

5.3.3. ANÁLISE DA INTERAÇÃO

Para averiguar a estabilidade da interação entre a catepsina B e a heparina, foi medido o número de ligações de hidrogênio e a energia de interação, ambas em função do tempo (figura 85).

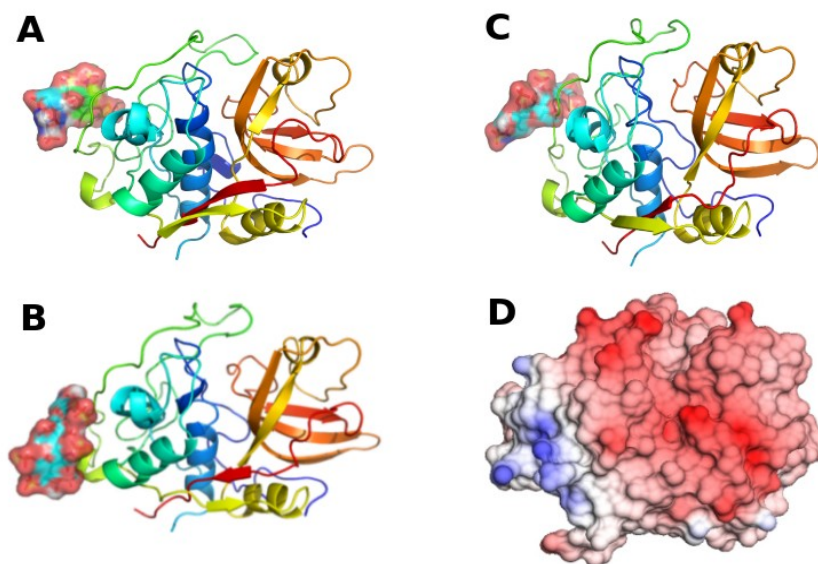


Figura 84: Representação gráfica do complexo inicial (A), complexo no tempo de 30 ns (B), complexo no tempo de 60 ns (C) e superfície eletrostática da catepsina B no tempo de 0 ns (D).

A interação do complexo catepsina B mais heparina se mostrou estável, uma vez que

apresentou uma energia de interação negativa ao longo da dinâmica, o que significa que a heparina não se desligou da catepsina B. A estabilidade da interação do complexo também foi suportada pelo número de ligações de hidrogênio, os quais mantiveram valores maiores do que zero durante todo o tempo. Esses dados sugerem um complexo mais estável que o observado para a papaína complexada à heparina. O complexo também foi analisado de forma visual (figura 84); em relação ao complexo inicial, os complexos obtidos nos tempos de 30 e 60 ns apresentaram leves diferenças, e não foram verificadas rotações e translações relevantes da heparina no sítio de ligação. Foi também verificado que a heparina entrou em contato apenas com o domínio L da catepsina B, próximo à região eletropositiva, distante do domínio oposto. Por não se ligar em ambos os domínios, esses dados suportam a idéia de que a heparina apresenta um mecanismo de modulação estrutural diferente do apresentado na análise da papaína.

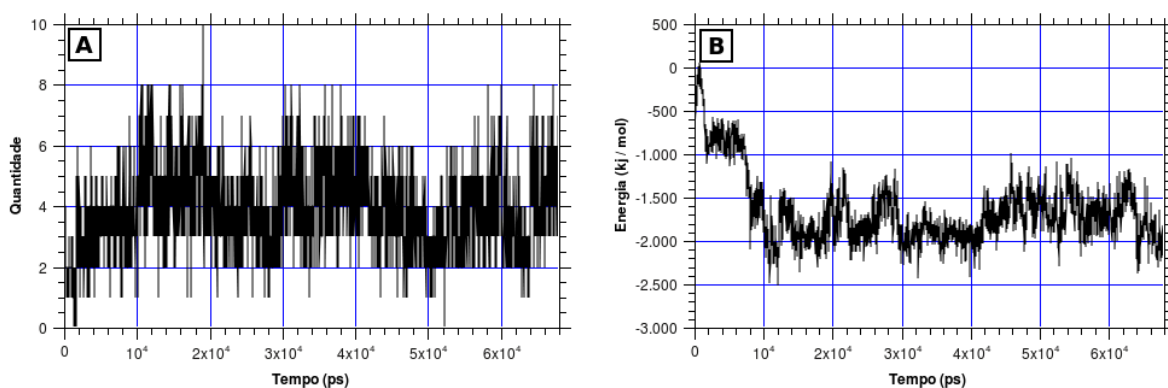


Figura 85: Número de ligações de hidrogênio entre a catepsina B e a heparina (A), e a energia de interação de curto alcance (Lennard Jones e Coulomb) entre essas mesmas moléculas (B).

Em função do tempo, também foi medido a distância entre os resíduos da catepsina B e os átomos de oxigênio da heparina (figura 86). O perfil de contato foi praticamente o mesmo durante toda a dinâmica, exceto por poucas variações, aparentemente irrelevantes. As regiões de contato identificadas para catepsina B se encontram aproximadamente entre os resíduos 50-70, 80-100 e 120-150, todos encontrados no domínio L da enzima (figura 88).

Os resíduos da papaína que entraram em contato com os átomos de oxigênio da heparina dentro de um raio de 3 Å no tempo de 30 ns, após o período de estabilização inicial, são a SER¹²⁷ e LYS¹²⁸, e dentro de um raio de 5 Å, os resíduos são a LYS⁸⁴, CYS¹²⁶, SER¹²⁷, LYS¹²⁸, ILE¹²⁹, TYR¹³⁸, LYS¹³⁹, LYS¹⁴². Desses resíduos, 4 resíduos são lisina, ou seja, 50%

dos resíduos em contato com a heparina no tempo de 30 ns são básicos. Atráves do mapa de contatos (figura 86), sugere-se que esse perfil de contato se manteve estável até o fim da dinâmica molecular do complexo.

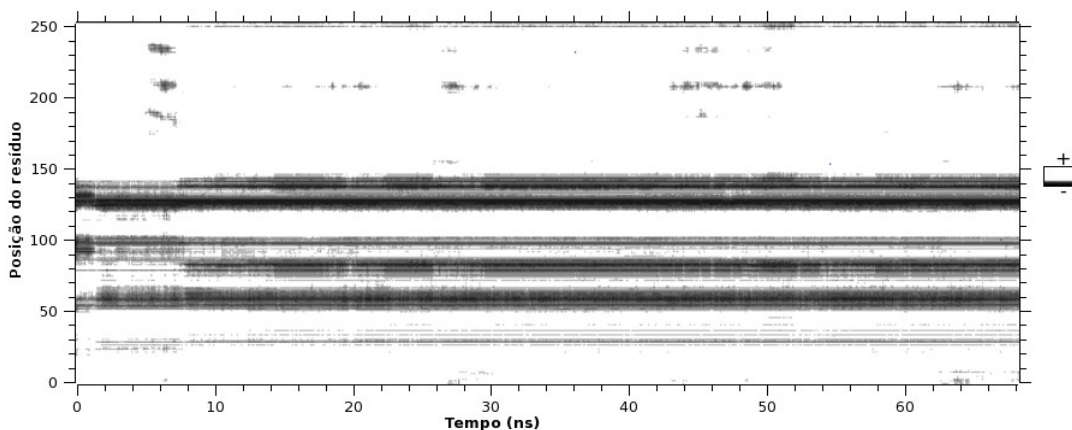


Figura 86: Distância dos resíduos da catepsina B à heparina em função do tempo.

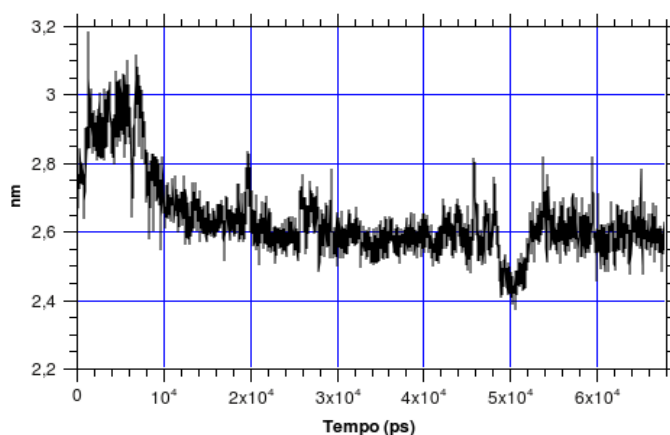


Figura 87: Distância média da catepsina B à heparina em função do tempo.

O gráfico de distância média entre os átomos de oxigênio da heparina e a catepsina B (figura 87) mostra que a distância da heparina se estabilizou a partir de 20 ns, apresentando a partir desse tempo uma elevada estabilidade ao longo da simulação. No tempo aproximado de 50 ns houve uma diminuição na distância entre as moléculas, porém, nenhuma relação com os demais resultados foi identificada. Esse gráfico, juntamente com os demais dessa seção, sugerem uma relação bastante estável entre a catepsina B e a heparina. Os dados também sugerem que, diferentemente do observado para o sítio de ligação da heparina na papaína, o sítio de ligação da heparina na catepsina B é dinamicamente estável, ou seja, é uma região não

reativa, incapaz de promover alterações estruturais através de “força mecânica” como observado na análise da papaína.

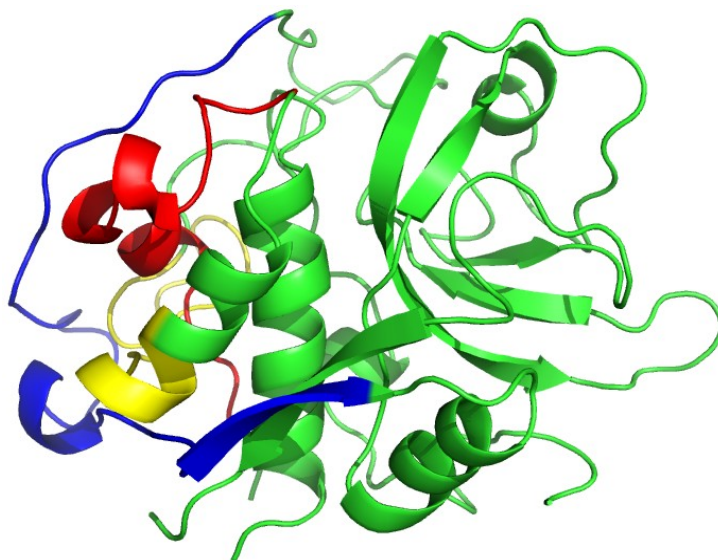


Figura 88: Representação da catepsina B. As regiões 50-70, 80-100, 120-150 estão representadas em vermelho, amarelo e azul, respectivamente.

6. DISCUSSÃO

A papaína e catepsina B são cisteíno-proteases que possuem uma atividade que deve ser estritamente controlada e isolada em compartimentos especiais. Falhas nesse controle podem levar à patologias (TWINING, 1994; CHAPMAN *et al.*, 1997). Em algumas linhagens de células carcinômicas, as cisteíno-proteases podem ser superexpressadas e externalizadas para a matriz extracelular, provocando a degradação dessa estrutura (BUCK *et al.*, 1992; SLOANE *et al.*, 1986; SZPADERSKA & FRANKFATER, 2001). A degradação da matriz extracelular é um fator importante no processo de metástase, pois permite às células tumorais se desprenderem e migrarem para outros tecidos, agravando a doença. A matriz extracelular é, porém, caracterizada por apresentar um pH próximo ao neutro. Nesse pH, *in vitro*, tanto a papaína quanto a catepsina B são rapidamente inativadas, mas, quando interagindo com GAGs, essa inativação não ocorre imediatamente. Estudos demonstraram que alguns GAGs altamente sulfatados, encontrados na matriz extracelular, podem evitar que essas proteases sofram desestabilização estrutural e a desprotonação da histidina do sítio ativo, evitando, conseqüentemente, a perda da atividade da enzima (figura 89) (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001). Neste projeto, estudamos a interação da catepsina B e papaína, um modelo de cisteíno-proteases, com o GAG heparina de alto peso molecular por simulação computacional utilizando o novo método *Generalized Simulated Annealing* e a Dinâmica Molecular. O estudo da interação com GAGs irá contribuir no entendimento da participação das cisteíno-proteases em tumores e poderá ser o ponto de partida para o desenvolvimento ou aperfeiçoamento de diferentes tipos de inibidores.

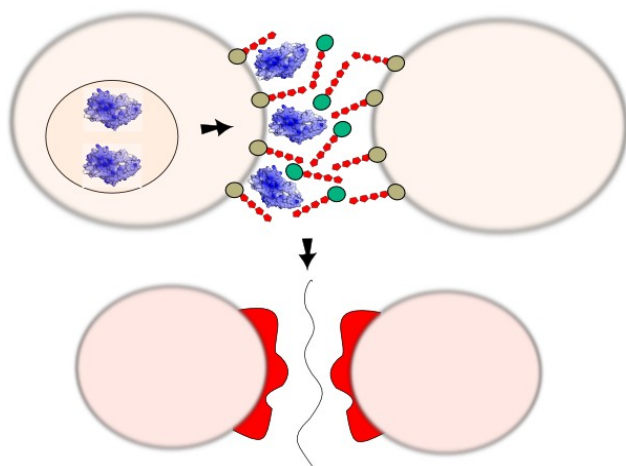


Figura 89: Degradação da matriz extracelular protagonizada pela catepsina B.

6.1. SOBRE OS RESULTADOS PRELIMINARES

A quantidade de íons adicionados na neutralização da carga total do sistema (proteína, solvente e heparina) mostra o caráter da interação eletrostática entre a papaína, catepsina B e heparina. A catepsina B e heparina apresentam um caráter geral eletronegativo, ao contrário da papaína, que tem um caráter eletropositivo. Em pH 7, onde as histidinas são encontradas desprotonadas, essas proteases devem apresentar, conseqüentemente, um caráter ainda mais eletronegativo. Esse resultado implica diretamente sobre a interação dessas proteases com a heparina, uma vez que a heparina é altamente eletronegativa e pode ser facilmente afetada por alterações na superfície eletrostática da enzima. Isso sugere que a interação entre a catepsina B e os GAGs deve ser favorecida por mecanismos capazes de induzirem um efeito de “pH ácido” na enzima, como por exemplo a protonação das histidinas, o que poderia facilitar a interação entre ambas essas moléculas. Quanto à papaína, que possui um caráter predominantemente eletropositivo, a interação com os GAGs deve teoricamente ser facilitada, uma vez que essas moléculas apresentam cargas opostas, as quais devem se atrair mutuamente.

A análise detalhada da superfície eletrostática da papaína em pH 7 mostrou que essa enzima apresenta um caráter totalmente eletropositivo (figura 32). Uma das regiões mais eletropositivas verificadas na superfície da papaína se encontra na parte traseira e no domínio R da enzima, onde se concentra a grande maioria das estruturas folha- β . De acordo com essas constatações, a heparina, que possui um caráter eletronegativo, através do *docking* pelo GSA, mostrou se complexar próxima à região mais eletropositiva da papaína citada anteriormente, que se encontra próxima aos resíduos THR¹⁴, PRO¹⁵, VAL¹⁶, LYS¹⁷, ASN¹⁸, ASN⁴⁶, GLN⁴⁷, TYR⁴⁸, GLU⁵⁰, ARG⁸³, ASN⁸⁴, TYR⁸⁶, PRO⁸⁷, TYR⁸⁸, LYS¹⁷⁴, GLY¹⁷⁸, ASN¹⁸⁴ e ASN²¹² (figura 33). Nesse caso, o caráter da interação é eletrostático, onde há atração e repulsão de cargas, que são responsáveis por definir a localização espacial do ligante ao longo da superfície da enzima. Esses resultados conferem com dados já publicados, que mostram que a interação da papaína pela heparina pode ser facilmente inibida pela adição de íons no solvente, sugerindo que a interação entre ambas essas moléculas é de caráter eletrostático (ALMEIDA *et al.*, 1999), e que o sítio de interação da heparina nessa enzima é aparentemente específico (ALMEIDA *et al.*, 1999; ALMEIDA *et al.*, 2001), uma vez que não foi obtido nenhum outro complexo entre a papaína e heparina. Contrariamente ao verificado na papaína,

a catepsina B apresentou uma superfície eletrostática bastante eletronegativa, com exceção de um pequena região identificada no domínio L na face frontal da enzima (figura 65), que pode ser um possível sítio de ligação. Esse sítio da catepsina B é também compatível com o descrito teoricamente e experimentalmente por outros estudos (SHIDA *et al.*, em preparação; ALMEIDA *et al.*, 2001). Esses dados sugerem que a carga total da enzima não influencia diretamente na interação com a heparina, mas sim que, cargas locais, geograficamente limitadas, podem ser suficientes para a interação com o GAG. Porém, o *docking* feito para a catepsina B e heparina em pH 7 não encontrou nenhum sítio putativo de ligação. Embora vários complexos tenham sido obtidos, poucas semelhanças espaciais foram encontradas entre eles (figura 66), assim como, a diferença entre a energia inicial e final desses complexos foi mínima, sugerindo, portanto, que nenhum dos complexos obtidos poderiam se manter estáveis em um sistema molecular durante uma dinâmica, assim como demonstrado por análises preliminares. Em pH 5, porém, essa relação deve se inverter, visto que a heparina pode se complexar à catepsina B com uma energia final mais favorável devido ao aumento da carga eletropositiva da enzima, uma vez que ela possui diversas histidinas, as quais são protonadas ou desprotonadas em função do pH, o que pode alterar de forma relevante a superfície eletrostática da enzima e a afinidade pelo ligante. O estudo feito em pH 5 comprovou essa hipótese, e apresentou um complexo altamente estável envolvendo os resíduos 50-70, 80-100 e 120-150 da catepsina B (figura 88), os quais são um potencial sítio de ligação para a heparina. Essa interação corrobora a hipótese de uma interação eletrostática, e releva a idéia de que a formação de um complexo estável entre a catepsina B e a heparina em pH 7 deve favorecer de alguma forma a desprotonação das histidinas, uma vez que o complexo em pH 5 mostrou uma maior estabilidade frente aos demais complexos em pH 7. Na prática, estudos sugerem que a interação do heparan sulfato com a catepsina B pode ocorrer ainda em pH ácido, visto que essas moléculas podem ser encontradas juntas nos lisossomos durante o tráfego intracelular (KRAMER, 1971; BIENKOWSKI & CONRAD, 1984; YANAGISHITA & HASCALL, 1984), o que poderia facilitar a formação inicial do complexo. Sugere-se que essa interação só poderia se manter estável em pH 7, na matriz extracelular, caso as histidinas da enzima não se desprotonassem, sugerindo, portanto, que os GAGs podem inibir a perda de atividade da catepsina B e papaína por manterem a estrutura da enzima em uma conformação funcionalmente ativa, possivelmente por evitarem a desprotonação das histidinas, assim como sugerido por Almeida *et al.* (1999). Além do mais, a possibilidade da

glicosilação das catepsinas por GAGs (DICKINSON, 2002) poderia criar interações constantes, explicando parte da estabilidade apresentada pela catepsina B em pH alcalino. Prematuramente, como possível mecanismo de proteção alcalina, sugere-se que os GAGs podem ser capazes de criarem um bloqueio espacial físico e eletrostático sobre a enzima, evitando o seu contato com o solvente e com os aniões OH^- (figura 90), explicando como a catepsina B poderia interagir de forma favorável com um GAG e como os resíduos histidinas poderiam evitar sua desprotonação. Estudos experimentais demonstraram que GAGs de maior massa molecular e altamente sulfatados, como a heparina e o dextran sulfato, apresentam os maiores efeitos sobre a catepsina B e papaína (ALMEIDA *et al.*, 2001). Em outras palavras, GAGs altamente sulfatados, bastante eletronegativos, podem expelir ânions da superfície da enzima com uma maior facilidade, e, GAGs de maior peso molecular, que ocupam um maior espaço físico, podem diminuir o contato da enzima pelo solvente e, conseqüentemente, pelos ânions OH^- .

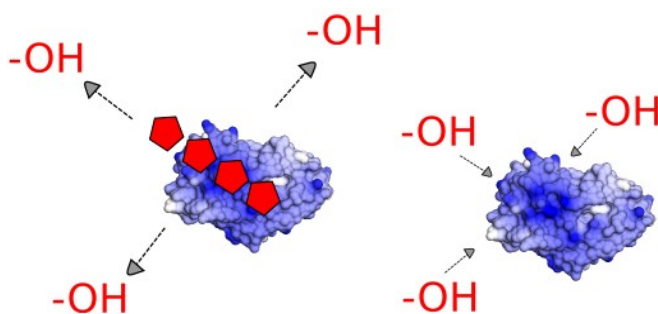


Figura 90: Modelo para explicar a hipótese da afinidade dos aniões OH^- pela superfície da catepsina B ou papaína em função da presença da heparina.

6.2. ANÁLISE DA ESTABILIDADE DINÂMICA

Foi observado pela análise de RMSD e MOD_GYR da dinâmica da papaína e heparina, que essas moléculas em seus estados livres apresentam uma maior maleabilidade estrutural. Os resultados mostraram que as estruturas da papaína e catepsina B oscilam dinamicamente de forma repetitiva, com frequências que dependem diretamente da condição na qual a enzima se encontra, ou seja, em pH 5 e 7 ou em presença ou ausência do ligante. Os resultados MOD_GYR por resíduo da papaína e catepsina B (figuras 37 e 73), apenas em

função do pH, mostraram que, quanto menor o pH, maior é a frequência de oscilação, e, quanto maior é o pH, menor é a frequência de oscilação MOD_GYR. Além da frequência de oscilação, o perfil dinâmico das enzimas apresentou um grau de desorganização acentuado em pH alcalino. Esses dados sugerem uma forte relação entre pH, perfil de dinâmica e, possivelmente, atividade. Além disso, o resultado da papaína na presença da heparina mostrou uma frequência de oscilação MOD_GYR por resíduo relativamente maior que a encontrada nos demais casos. No caso da catepsina B, por outro lado, foi observado uma baixa frequência e uma maior desestabilização MOD_GYR, sugerindo que os GAGs não induzem a papaína e a catepsina B aos respectivos perfis dinâmicos originais. Se considerarmos que as oscilação MOD_GYR por resíduo da papaína e catepsina B possuem relações diretas com a atividade catalítica, esses resultados sugerem que a alteração do perfil dinâmico induzido pela heparina pode diminuir a capacidade catalítica da papaína e catepsina B, assim como verificado experimentalmente (ALMEIDA *et al.* 1999; ALMEIDA *et al.* 2001). Porém, visto a modulação dinâmica promovida pela heparina sobre a papaína e, possivelmente, sobre a catepsina B, a presença do GAG, aparentemente, pode estabilizar de forma parcial os efeitos dinâmicos produzidos na protease por um pH alcalino, o que pode favorecer um leve aumento relativo na sua atividade, explicando, possivelmente, o deslocamento do perfil de atividade observado experimentalmente para a catepsina B (ALMEIDA *et al.*, 2001). Assim como, também foi observado na análise da papaína que a frequência de oscilação MOD_GYR é sensível ao perfil de interação com a heparina, e além disso, as diferenças observadas nos perfis dinâmicos não possuem relações com alterações estruturais relevantes na enzima. Esses dados sugerem uma independência entre os mecanismos de proteção contra a inativação alcalina e o aumento da estabilidade dinâmica observada experimentalmente para essas enzimas.

Seguindo o raciocínio sobre a interação entre a papaína e a heparina sobre os resultados MOD_GYR, se a frequência de oscilação e mobilidade estrutural estão relacionadas entre si, podemos dizer que a mobilidade dos resíduos da papaína é restringida pela heparina, o que corrobora com os resultados obtidos na análise RMSD. De forma mais detalhada, foi mostrado, em todos os casos, que os domínios da papaína e catepsina B oscilam com uma diferença de fase, ou seja, quando um domínio se “expande”, o outro se “contraí” (figuras 38, 39 e 40). Esse resultado salienta a individualidade dinâmica apresentada pelos domínios da catepsina B e papaína, além de relevar um importante fator na interação dessas

enzimas com os GAGs, uma vez que foi mostrado que a heparina pode apresentar contatos próximos com ambos os domínios da papaína (figura 56), o que sugere uma interação complexa envolvendo restrições na dinâmica estrutural, capaz de promover distorções estruturais e alterações na energia da proteína quando essa interação é intensificada (figura 91). Essas hipóteses foram corroboradas pela análise da energia potencial da papaína complexada e pelas oscilações alternadas verificadas para as regiões M1, M2 e M3, caracterizadas nesse trabalho, as quais demonstraram oscilar juntamente com o domínio oposto, sugerindo a influência da heparina e da oscilação alternada sobre a estrutura da papaína. Esses dados indicam que, na presença de GAGs, a oscilação alternada dos domínios da papaína e catepsina B pode ser o combustível ou a força dinâmica necessária para a promoção de alterações estruturais significativas, e que GAGs altamente sulfatados e de maior peso molecular podem apresentar interações de maior intensidade. Além do mais, resultados experimentais demonstraram que pontes salinas que contribuem para a interação entre ambos os domínios são naturalmente importantes para a estabilidade estrutural e atividade da catepsina B e papaína (MUSIL *et al.*, 1991; NÄGLER *et al.*, 1997; TURK *et al.*, 1994), sugerindo que os GAGs também podem ser um complemento para o aumento dessa interação, o que poderia favorecer a estabilidade estrutural dessas enzimas. Se a estabilidade das oscilações MOD_GYR também possuem relações diretas com a estabilidade estrutural, é possível sugerir que um pH ácido e a presença de um GAG também podem ser importantes para a estabilidade estrutural, o que auxiliaria na atividade da enzima em pH alcalino, assim como também verificado experimentalmente por Almeida *et al.* (1999). Resumindo, sugere-se que a estabilidade estrutural e catalítica deve possuir correlação com a estabilidade dinâmica da enzima, assim como representado pelos resultados MOD_GYR.

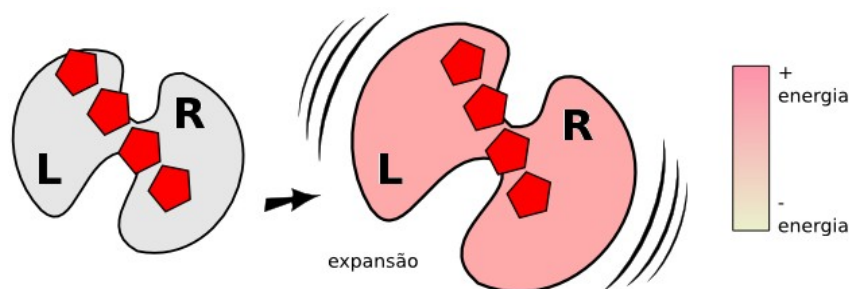


Figura 91: Representação gráfica da interferência dinâmica sobre a energia estrutural da papaína.

Na análise da catepsina B, os resultados do raio de giração e MOD_GYR mostraram

uma baixa influência da heparina sobre alterações estruturais na enzima. Atráves dos resultados MOD_GYR, foi possível observar que a heparina induz a papaína em pH 7 à uma dinâmica estrutural bastante similar à encontrada em pH 5 e, ainda, que a heparina induz a catepsina B em pH 5 à uma dinâmica bastante similar à encontrada em pH 7; ou seja, existe uma inversão de características com relação à papaína. Analisando a carga eletrostática, pode-se esperar que o complexo da catepsina B em pH 5 apresente as mesmas características observadas para a enzima livre em pH 7, uma vez que a presença da heparina pode equilibrar as cargas observadas entre ambas essas situações. Por outro lado, quanto ao provável complexo da papaína em pH 7, pode-se dizer que não há um resultado indicativo de neutralização de cargas eletrostáticas, uma vez que a carga eletrostática da papaína complexada em pH 7 e livre em pH 5 são bastante diferentes entre si. Além do mais, outra explicação possível para esse fenômeno de “inversão” é que, diferentemente do observado na papaína, o sítio de ligação da heparina na catepsina B se encontra no extremo do domínio L, o que não permite o contato do sacarídeo com ambos os domínios da enzima. A ausência desse padrão de contato sugere uma interação incapaz provocar alterações estruturais e dinâmicas significativas, uma vez que o sítio de ligação da heparina na catepsina B, aparentemente, não apresenta significativa atividade dinâmica. Esses dados validam o mecanismo de interação já sugerido para a papaína (interferência dinâmica), e sugere a existência de um diferente mecanismo para a interação entre a catepsina B e a heparina observada nesse trabalho. Esse mecanismo é provavelmente o resultado único de interações eletrostáticas, uma vez que a interferência dinâmica é aparentemente insignificante nesse caso.

6.3. ANÁLISE DA ESTABILIDADE ESTRUTURAL

Experimentalmente, o conteúdo de estrutura α -hélice da papaína e catepsina B aumenta na presença de GAGs sulfatados (ALMEIDA *et al.* 1999; ALMEIDA *et al.* 2001). Da mesma forma, teoricamente, foram verificadas variações no conteúdo de estrutura α -hélice da papaína complexada à heparina (figuras 46 e 49). Essas variações apresentaram correlação com a acessibilidade de determinados resíduos, os quais possuem uma natureza básica ou ácida, e localização predominante próxima ou em estruturas do tipo α -hélice (figuras 48 e 51). A principal região que apresentou variações na estrutura α -hélice se

encontra próxima ao sítio de ligação da heparina, especificamente próxima aos resíduos GLY¹³⁸, LYS¹³⁹, ASP¹⁴⁰, PHE¹⁴¹ e GLN¹⁴². Esses dados sugerem que os GAGs podem promover alterações estruturais localizadas por eliminarem interações eletrostáticas detrimentalis (FERRAN, SOBEL & HARRIS, 1992), o que não explica o grande aumento no conteúdo da estrutura α -hélice observado experimentalmente (ALMEIDA *et al.* 1999; ALMEIDA *et al.* 2001). Interessantemente, os resíduos que apresentaram variações na acessibilidade mostraram se encontrar em estruturas α -hélices, o que pode ser o possível resultado de torções resultantes da presença próxima do GAG nessas estruturas. Adicionalmente, o conteúdo da estrutura α -hélice se mostrou estável após o seu aumento, sugerindo que a simples presença do GAG no respectivo sítio de ligação, sem contatos específicos, é suficiente para a manutenção do conteúdo das α -hélices, provavelmente porque essas estruturas, especificamente na papaína, são naturalmente estáveis. Porém, para o aumento inicial do conteúdo dessa estrutura, foi verificado a necessidade de um perfil de contato específico, sugerindo também que os GAGs podem ser capazes de induzirem alterações estruturais significativas na papaína e, provavelmente, na catépsina B.

Quanto ao conteúdo da estrutura folha- β da papaína (figura 47), por outro lado, não foram verificadas variações relevantes e proporcionais ao aumento do conteúdo da estrutura α -hélice. Almeida *et al.* (1999) sugeriu que a diminuição do conteúdo de folha- β é o resultado do aumento do conteúdo de α -hélice, a qual poderia limitar as regiões de formação das estruturas folha- β . Conforme a estrutura da papaína, as folha- β se encontram no domínio R, o mesmo domínio em que foi encontrado o potencial sítio de ligação de GAGs. Porém, a pouca influência apresentada pela heparina sobre as estruturas folha- β da catépsina B e, principalmente, da papaína, sugere uma alta estabilidade dessa região.

Outra característica verificada foi a acessibilidade da histidina do sítio ativo da papaína encontrada nas diversas situações de análise. Experimentalmente, os GAGs diminuem o grau de desprotonação da histidina-159, aumentando a atividade da papaína por preservar o par de íons tiolato-imidazólio (ALMEIDA *et al.*, 1999). Nesse trabalho, foi verificado uma possível relação entre a baixa acessibilidade desse resíduo com a resistência da enzima à inativação alcalina. Através de interação forte nos contatos entre a papaína e a heparina, foram verificadas alterações na acessibilidade da histidina do sítio ativo (figura 52) que, em determinados momentos, se apresentou baixa, o que poderia diminuir o seu contato com o solvente e com os aniões OH⁻, mantendo a integridade do par de íons tiolato-imidazólio

e favorecendo a sua proteção contra a inativação alcalina (figura 92). Esses dados sugerem que a papaína em pH 7 sobre a ação da heparina é favorecida contra a inativação alcalina devido à posição estratégica apresentada pela histidina-159. Porém, essa mesma relação não foi verificada entre as estruturas da catepsina B nos diferentes pHs, o que sugere que a acessibilidade da histidina pode não ser um fator importante contra a inativação alcalina. Por outro lado, sugere-se que a localização espacial dessa histidina em relação aos demais resíduos é de maior importância, e que está aparentemente relacionada à atividade e estabilidade da enzima. Além disso, a análise de comparação entre as trajetórias da papaína complexada e papaína em pH 5 apresentou informações interessantes. Foi verificado que a diminuição da acessibilidade da histidina do sítio ativo é apenas um reflexo de uma alteração estrutural maior que mostrou induzir as estruturas da trajetória da papaína (pH 7) complexada à estruturas bastante similares às verificadas para a trajetória dessa enzima em pH 5 (figura 63). Esses dados sugerem que a papaína e, possivelmente, a catepsina B, são funcionalmente ativas em pH alcalino simplesmente porque os GAGs podem ser capazes de induzirem nessas enzimas alterações estruturais bastante similares às verificadas nas respectivas estruturas encontradas em pH ácido (figura 93), onde apresentam uma maior estabilidade e uma atividade máxima (KIRSCHKE *et al.*, 1995; ALMEIDA *et al.*, 2001). Esses dados também explicam porque o par de íons tiolato-imidazólio é preservado na interação entre a papaína e a heparina (ALMEIDA *et al.*, 1999), uma vez que a estrutura da enzima encontrada em pH ácido pode favorecer a sua estabilidade. Adicionalmente, foi mostrado que o período de tempo das diminuições observadas na acessibilidade da histidina do sítio ativo foram compatíveis com os respectivos períodos de oscilação MOD_GYR (figura 53), sugerindo uma forte influência da dinâmica e estabilidade da enzima sobre as alterações estruturais observadas. Possivelmente, existem conformações estruturais da enzima que favorecem a interação com o GAG e, de forma similar, existem conformações pouco favoráveis que interrompem a interação entre ambas essas moléculas, que é possivelmente o resultado do aumento da energia potencial da estrutura da enzima.

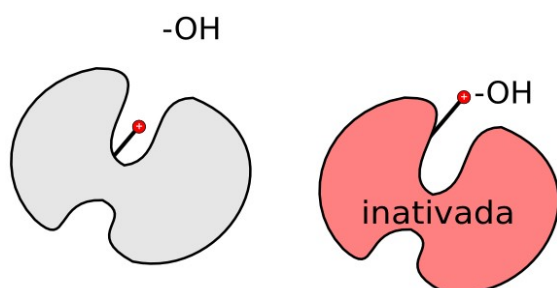
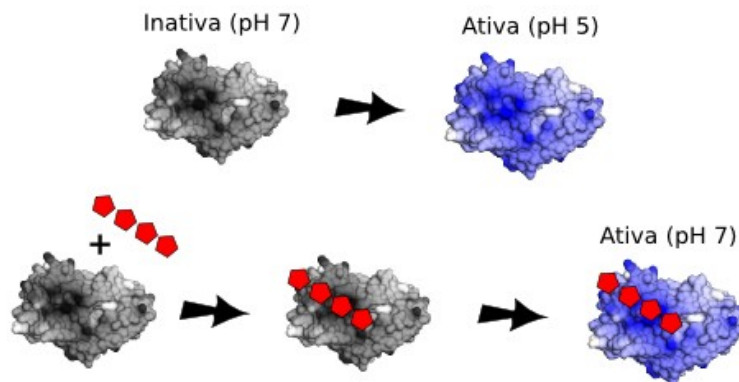


Figura 92: Modelo para ilustrar a hipótese da acessibilidade da histidina do sítio ativo da papaína sobre a inativação alcalina.

Figura 93: Ilustração mostrando o provável efeito da heparina sobre as cisteíno-proteases. Obs: a enzima em preto está inativa, e em azul está ativa.



6.4. ANÁLISE DA INTERAÇÃO DE CONTATO

Quanto ao perfil de interação entre a heparina e a papaína, sugerimos que ela é específica, uma vez que as regiões eletrostaticamente favoráveis se mostraram limitadas a uma única região, sendo ela encontrada no domínio R na face traseira da enzima, próxima aos resíduos 5-17, 135-148 e 173-188 (figura 54), e próxima aos resíduos 188-191 sugeridos por Cardin & Weintraub (1989). Todas as alterações estruturais identificadas significantes para a papaína mostraram ser o resultado da interação com a heparina, mostrando claramente a influência e a importância da interação eletrostática dos GAGs. Detalhadamente, durante toda a dinâmica molecular, a heparina mostrou proximidade por ambos os domínios da papaína, e esse complexo mostrou ser energeticamente estável, sugerindo que essa região é um sítio de ligação putativo. Esses dados sugerem que os GAGs podem interferir na oscilação dinâmica da estrutura da enzima por se ligarem em regiões específicas, podendo favorecer o aumento da frequência de oscilação, estabilidade dinâmica e drásticas alterações estruturais. Além disso, foi verificado também que não existe um único perfil de interação para a promoção dessas alterações estruturais observadas (figura 58), porém, a maior intensidade do contato do GAG pelas regiões 135-148 e 173-188 da papaína mostrou ser uma regra (figura 60). Esses dados sugerem que as alterações estruturais promovidas pelos GAGs são o resultado principal de forças eletrostáticas. Interações puramente eletrostáticas não levam em consideração de as características estruturais da enzima e do GAG, assim como também sugerido por Almeida *et al.* (2001). A hipótese da neutralização das cargas eletrostáticas foi corroborada pela análise de distância entre a HIS⁸¹ da papaína e os átomos de oxigênio da heparina (figura 64), que mostrou que a distância entre essas partes diminui quando interações estruturais significativas

foram observadas; embora não muito significativas, essas aproximações demonstraram ocorrer muito antes das alterações estruturais, sugerindo que as alterações estruturais observadas na enzima apareceram em consequência de interações estáveis desse resíduo com a heparina. Também foi verificado que a frequência de oscilação MOD_GYR não está diretamente relacionada com uma maior aproximação da heparina por ambos os domínios da papaína (figura 59 e 60), sugerindo que a simples aproximação do GAG por ambos os domínios não restringe a maleabilidade da enzima e aumenta a frequência de oscilação. Conseqüentemente, foi verificado que o aumento da oscilação MOD_GYR por resíduo da papaína é aparentemente o resultado do aumento da interação da heparina por regiões específicas da enzima, sendo elas a 135-148 e 173-188, e um menor contato pela região 5-17 (figura 60).

O perfil de interação entre a heparina e a catepsina B mostrou ser bastante estável, uma vez que as regiões eletrostaticamente favoráveis se mostraram limitadas a uma única região, que se encontra no domínio L entre a face frontal e traseira da enzima, próxima aos resíduos 50-70, 80-100 e 120-150 (figura 86), o que corrobora com resultados obtidos por Shida *et al.* (em preparação) em um estudo com dissacarídeos utilizando a mesma metodologia utilizada nesse trabalho. O perfil de contato não apresentou alterações significativas. Após o período inicial de estabilização da dinâmica, a energia de contato e o número de ligações de hidrogênio mostraram que essa é uma interação de alta estabilidade (figura 85), o que também explica a alta estabilidade e especificidade do perfil de contato (figura 86). Essa estabilidade é provavelmente o resultado de um sítio de ligação pouco dinâmico, o que explica a pouca atividade da heparina sobre alterações estruturais na catepsina B, assim como já sugerido anteriormente. Embora o complexo da catepsina B em pH 5 tenha sido analisado detalhadamente, pouco se sabe sobre os demais complexos em pH 7 que não apresentaram estabilidade. Porém, o conhecimento obtido através das análises de interação da papaína e catepsina B com a heparina foram consistentes e puderam ser correlacionados com dados experimentais.

7. CONCLUSÕES & SUGESTÕES

A combinação do GSA *docking* e simulações por MD mostrou que essa abordagem pode ser utilizada para estudar complexos entre receptores e ligantes, como a papaína e heparina, ou catepsina B e heparina, uma vez que suas estruturas são difíceis de serem obtidas através de cristalografia por raio-X ou por outras técnicas espectroscópicas. Através do *docking* pelo GSA, foi encontrado um putativo sítio de ligação de GAGs na região mais eletrostaticamente positiva da papaína e catepsina B, sugerindo uma interação eletrostática e específica. No caso da catepsina B, ainda não foi encontrado nenhum potencial sítio de ligação para a heparina em pH 7, apenas em pH 5. O estudo também sugere que a interação entre as proteínas e o GAG é dinâmica, o que significa que a heparina não fica ancorada em um único sítio e que ela pode interagir por uma certa área dos sítios putativos. Através das simulações por MD, resumidamente, verificou-se como a papaína e, possivelmente, a catepsina B, são funcionalmente ativas em pH alcalino pela interação com os GAGs, que podem ser capazes de induzirem nessas enzimas alterações estruturais e dinâmicas similares às verificadas nas respectivas estruturas encontradas em pH ácido, onde apresentam uma maior estabilidade e uma maior atividade. O mecanismo envolvido nesse processo é possivelmente o resultado da neutralização dos efeitos produzidos pela desprotonação das histidinas. Aparentemente, essa neutralização não é somente o resultado de uma simples anulação de cargas eletrostáticas, mas sim, também, de um mecanismo complexo de interferência dinâmica resultante do contato do GAG com a protease. Os dados obtidos nesse trabalho podem ser utilizados na interpretação de dados experimentais sobre os efeitos da interação dessa classe de proteases com GAGs, o que também possibilitará, futuramente, o desenvolvimento de novos tratamentos terapêuticos na inibição da metástase tumoral.

REFERÊNCIAS

- ALMEIDA, P. C.; NANTES I. L.; CHAGAS, J. R.; RIZZI, C. C. A.; ALARIO, A. F.; CARMONA, E.; JULIANO, L.; NADER, H. B.; & TERSARIOL I. L. S. Cathepsin B Activity Regulation, **J Biol Chem**, Estados Unidos, Vol. 276, N° 2, p. 944–951, 2001.
- ALMEIDA, P. C.; NANTES, I. L.; RIZZI, C. C. A.; JUDICE, W. A. S.; CHAGAS, J. R.; JULIANO, L.; NADER, H. B.; & TERSARIOL, I. L. S. Cysteine Proteinase Activity Regulation: A Possible Role of Heparin and Heparin-Like Glycosaminoglycans, **J Biol Chem**, Estados Unidos, Vol. 274, N° 43, p. 30433–30438, 1999.
- ANEZO, C.; VRIES, A. H.; HOLTJE, H. D.; TIELEMAN, D. P.; & MARRINK, S. J. Methodological Issues in Lipid Bilayer Simulations, **J Phys Chem**, Estados Unidos, Vol. 107, N° 35, p. 9424-9433, 2003.
- BAKER, E. N.; & DRENTH, J. **Biological Macromolecules and Assemblies**. 1. ed. Nova York: John Wiley & Sons, 1987.
- BARRETT, A. J. Classification of peptidases, **Methods Enzymol**, Estados Unidos, Vol. 244, N° 1, p. 1-15, 1994.
- BARRETT, A. J.; & KIRSCHKE, H. Cathepsin B, cathepsin H, and cathepsin L, **Methods Enzymol**, Estados Unidos, Vol. 80, N° 1, p. 535-561, 1981.
- BARRETT, A. J.; RAWLINGS, N. D.; & WOESSNER, J. F. **Handbook of proteolytic enzymes**. 1. ed. Nova York: Academic Press, 1998.
- BATCHO, P. F.; & SCHLICK, T. New splitting formulations for lattice summations, **J Chem Phys**, Estados Unidos, Vol. 115, N° 18, p. 8312-8326, 2001.
- BERENDSEN, H. .C.; POSTMA, J. P. M.; DINOLA, A.; & HAAK, J. R. Molecular dynamics with coupling to an external bath, **J Chem Phys**, Estados Unidos, Vol. 81, N° 8, p. 3684–3690, 1984.

- BERENDSEN, H. J. C.; POSTMA, J. P. M.; VAN GUNSTEREN, W. F.; & HERMANS, J. Interaction models for water in relation to protein hydration. In: **Intermolecular Forces**. 1. ed. Dordrecht : Pullman, B. ed., p. 331–342, 1981.
- BEUTLER, T. C.; & VAN GUNSTEREN, W. F. Molecular dynamics free energy calculation in four dimensions, **J Chem Phys**, Estados Unidos, Vol. 101, N° 2, p. 1417–1422, 1994.
- BIENKOWSKI, M. J.; & CONRAD, H. E. Kinetics of proteoglycan sulfate synthesis, secretion, endocytosis, and catabolism by a hepatocyte cell line, **J Biol Chem**, Estados Unidos, Vol. 250, N° 21, p. 12989-12996, 1984.
- BUCK, M. R.; KARUSTIS, D. G.; DAY, N. A.; HONN, K. V.; & SLOANE, B. F. Degradation of extracellular-matrix proteins by human cathepsin B from normal and tumour cells, **Biochem J**, Inglaterra, Vol. 282, N° 1, p. 273-278, 1992.
- CAPILA, I.; & LINHARDT, R. J. Heparin-Protein Interactions, **Angew Chem Int Ed Engl**, Alemanha, Vol. 41, N° 3, p. 390-412, 2002.
- CARDIN, A. D.; & WEINTRAUB, H. J. Molecular modeling of protein–glycosaminoglycan interactions, **Arteriosclerosis**, Estados Unidos, Vol. 9, N° 1, p. 21-32, 1989.
- CHAPMAN, H. A.; RIESE, R. J.; & SHI, G. P. Emerging roles for cysteine proteases in human biology, **Annu Rev Physiol**, Estados Unidos, Vol. 59, N° 1, p. 63-88, 1997.
- CHEN, M. E.; CANG, H. X.; & NYMEYER, H. NOC. Inst. Mol. Biol.; Florida State University, site: <<http://noch.sourceforge.net>>, acessado dia 20 de abril de 2007.
- CLAUSIUS, R. J. E. On a Mechanical Theorem Applicable to Heat, **Philosophical Magazine**, Inglaterra, Vol. 4, N° 40, p. 122-127, 1870.
- CONNOLLY, M. L. Analytical molecular surface calculation, **J Appl Crystallogr**, Dinamarca, Vol. 16, N° 5, p. 548–558, 1983.

- DEL NERY, E.; JULIANO, M. A.; LIMA, A. P. C. A.; SCHARFSTEIN, J.; & JULIANO, L. Kininogenase Activity by the Major Cysteiny Proteinase (Cruzipain) from *Trypanosoma cruzi*, **J Biol Chem**, Estados Unidos, Vol. 272, N° 41, p. 25713–25718, 1997.
- DELANO, W. L. The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, site: <<http://www.pymol.org>>, acessado dia 20 de abril de 2007.
- DEUSSING, J.; ROTH, W.; SAFTIG, P.; PETERS, C.; PLOEGH, H. L.; & VILLADANGOS, J. A. Cathepsins B and D are dispensable for major histocompatibility complex class II-mediated antigen presentation, **Proc Natl Acad Sci USA**, Estados Unidos, Vol. 95, N° 8, p. 4516-4521, 1998.
- DICKINSON, D. P. Cysteino Peptidases of Mammals: Their Biological Roles and Potential Effects in the Oral Cavity and Other Tissues in Health and Disease, **Crit Rev Oral Biol Med**, Estados Unidos, Vol 13, N° 3, p. 238-275, 2002.
- DUFOUR, E.; DIVE, V.; & TOMA, F. Delineation of chicken cathepsin L secondary structure; relationship between pH dependence activity and helix content, **Biochim Biophys Acta**, Holanda, Vol. 955, N° 1, p. 58-64, 1988.
- FENG, M. H.; CHAN, S. L.; XIANG, Y.; HUBER, C. P.; & LIM, C. The binding mode of an E-64 analog to the active site of cathepsin B, **Protein Eng**, Inglaterra, Vol. 9, p. 977-986, 1996.
- FERRAN, D. S.; SOBEL, M.; & HARRIS, R. B. Design and synthesis of a helix heparin-binding peptide, **Biochemistry**, Estados Unidos, Vol. 31, N° 21, p. 5010-5016, 1992.
- GALLAGHER, J. T. & TURNBULL, J. E. Mini review heparan sulphate in the binding and activation of basic fibroblast growth factor, **Glycobiology**, Inglaterra, Vol. 2, N° 6, p. 523-528, 1992.
- GALLAGHER, J. T.; & WALKER, A. Molecular distinctions between heparan Sulphate and Heparin: Analysis of sulphation patterns indicates heparan Sulphate and Heparin are separate families of N-sulphated polysaccharides, **Biochem J**, Inglaterra, Vol. 230, N° 3, p. 665-674, 1985.

- GARGALO, R.; OLIVA, B.; QUEROL, E.; & AVILÉS, F. X. Effect of the reaction field electrostatic term on the molecular dynamics simulation of the activation domain of procarboxypeptidase B, **Protein Eng**, Inglaterra, Vol. 13, N° 1, p. 21-26, 2000.
- GOMES, P. B.; & DIETRICH, C. P. Distribution of heparin and other sulfated glycosaminoglycans in vertebrates, **Comp Biochem Physiol B**, Inglaterra, Vol. 73, N° 4, p. 857-863, 1982.
- GREENGARD, L.; & ROKHLIN, V. A Fast Algorithm for Particle Simulations, **J Comput Phys**, Estados Unidos, Vol. 73, N° 2, p. 325-348, 1987.
- GRØNBECH-JENSEN, N.; & DONIACH, S. Long Time Overdamped Langevin Dynamics of Molecular Chains, **J Comput Chem**, Estados Unidos, Vol. 15, N° 9, p. 997-1012, 1994.
- GROOT, R.; & WARREN, P. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation, **J Chem Phys**, Estados Unidos, Vol. 107, N° 11, p. 4423-4435, 1997.
- GUENETTE, R. S.; MOOIBROEK, M.; WONG, K.; & WONG, P.; TENNISWOOD, M. Cathepsin B, a cysteine protease implicated in metastatic progression, is also expressed during regression of the rat prostate and mammary glands, **Eur J Biochem**, Inglaterra, Vol. 226, N° 1, p. 311-321, 1994.
- HEDMAN, F. Algorithms for Molecular Dynamics Simulations. Stockholms University: Dissertação de Mestrado, 2006.
- HESS, B.; BEKKER, H.; BERENDSEN, H. J. C.; & FRAAIJE, J. G. E. M. LINCS: A linear constraint solver for molecular simulations, **J Comput Chem**, Estados Unidos, Vol. 18, N° 12, p. 1463-1472, 1997.
- HIRATA I. Y.; CEZARI, M. H. S.; NAKAIE, C. R.; BOSCHCOV, P.; ITO, A. S.; JULIANO, M. A. & JULIANO, L. Internally quenched fluorogenic protease substrates: solid-phase synthesis and fluorescence spectroscopy of peptides containing ortho-aminobenzoyl/dinitrophenyl groups as donor-acceptor pairs, **Letters in Peptide Science**, Holanda, Vol. 1, N° 1, p. 299-308, 1994.

- HOCKNEY, R. W.; GOEL, S. P.; & EASTWOOD, J. Quiet high-resolution computer models of a plasma, **J Comput Phys**, Estados Unidos, Vol. 14, N° 1, p.148-158, 1974.
- HOOGERBRUGGE, P. J.; & KOELMAN, J. M. V. A. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, **Europhys Lett**, França, Vol. 19, N° 3, p. 155-160, 1992.
- ILLY, C.; QURAIISHI, O.; WANG, J.; PURISIMA, E.; VERNET, T.; & MORT, J. S. Role of the Occluding Loop in Cathepsin B Activity, **J Biol Chem**, Estados Unidos, Vol. 272, N° 2, p. 1197–1202, 1997.
- ISRAELACHVILI, J. M. **Intermolecular and Surface Forces**. 2. ed. London: Academic Press, 1991.
- IZAGUIRRE, J. A.; HAMPTON, S. S.; & MATTHEY, T. Parallel multigrid summation for the N-body problem, **J Parallel Distrib Comput**, Estados Unidos, Vol. 65, N° 8, p. 949-962, 2005.
- KABSCH, W.; & SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, **Biopolymers**, Estados Unidos, Vol. 22, N° 12, p. 2577-2637, 1983.
- KAMPHUIS, I. G.; KALK, K. H.; SWARTE, M. B.; & DRENTH, J. Structure of papain refined at 1.65 Å resolution, **J Mol Biol**, Inglaterra, Vol. 179, N° 2, p. 233-56, 1984.
- KIRSCHKE, H.; BARRETT, A. J.; & RAWLINGS, N. D. Proteinases 1: lysosomal cysteine proteinases, **Protein Profile**, Inglaterra, Vol. 2, N° 14, p. 1581-1643, 1995.
- KRAMER, P. M. Heparan sulfates of cultured cells. II. Acid-soluble and -precipitable species of different cell lines, **Biochemistry**, Estados Unidos, Vol. 10, N° 8, p. 1437-1445, 1971.
- LABANOWSKI, J. K. Molecular-Modeling (apostila). Ohio Supercomputer Center, site: <<http://www.ccl.net/cca/documents/molecular-modeling/node9.html>>, acessado dia 7 de fevereiro de 2008.

LALONDE, J. M.; ZHAO, B.; SMITH, W. W.; JANSON, C. A.; DESJARLAIS, R. L.; TOMASZEK, T. A.; CARR, T. J.; THOMPSON, S. K.; YAMASHITA, D. S.; VEBER, D. F.; & ABDEL-MEQUID, S. S. Use of papain as a model for the structure based design of cathepsin K inhibitors: crystal structures of two papain-inhibitor complexes demonstrate binding to S'-subsites, **J Med Chem**, Estados Unidos, V. 41, N° 23, p.4567-4643, 1998.

LANG S. **Equation**. 1. ed. Massachusetts: Addison-Wesley Publishing Company, 1975.

LEE, B.; & RICHARDS, F. M. The interpretation of protein structures: estimation of static accessibility, **J Mol Biol**, Inglaterra, Vol. 55, N° 3, p. 379-400, 1971.

LEHNINGER, A. L.; NELSON, D. L.; & COX, M. M. **Principles of Biochemistry**. 4. ed. Nova York: Worth Publishers, 2004.

LELLOUCH, A. C.; & LANSBURY, P. T. Jr. A peptide model for the heparin binding site of antithrombin III, **Biochemistry**, Estados Unidos, Vol. 31, N° 8, p. 2279-2285, 1992.

LEMAK, A.; & BALABAEV, N. Molecular dynamics simulation of a polymer chain in solution by collisional dynamics method, **J Comput Chem**, Estados Unidos, Vol. 17, N° 15, p. 1685-1695, 1996.

LINDAHL E.; HESS, B.; & VAN DER SPOEL, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis, **J Mol Model**, Alemanha, Vol. 7, N° 1, p. 306-317, 2001.

MAEKAWA, Y.; HIMENO, K.; ISHIKAWA, H.; HISAEDA, H.; SAKAI, T.; DAINICHI, T. Switch of CD4+ T cell differentiation from Th2 to Th1 by treatment with cathepsin B inhibitor in experimental leishmaniasis, **J Immunol**, Estados Unidos, Vol. 161, N° 1, p. 2120-2127, 1998.

MCGRATH, M.E. The lysosomal cysteine proteases, **Biophys Struct Mech**, Alemanha Vol. 28, N° 1, p. 181-204, 1999.

METROPOLIS, N.; ROSEMBLUTH, A.W.; ROSEMBLUTH, N.M.; TELER, A.N.; & TELLER, E. Equation of state calculations by fast computing machines, **J Chem Phys**, Estados Unidos, Vol. 21, N° 1, p. 1087-1092, 1953.

- MORET, M. A.; BISCH, P. M.; MUNDIM, K. C.; & PASCUTTI, P. G. New stochastic strategy to analyze helix folding, **Biophys J**, Estados Unidos, Vol. 82, N° 3, p. 1123-1132, 2002.
- MORET, M. A.; PASCUTTI, P. G.; BISCH, P. M.; & MUNDIM, K. C. Stochastic molecular optimization using generalized simulated annealing, **J Comput Chem**, Estados Unidos, Vol. 19, N° 1, p. 647-657, 1998.
- MORT, J.S.; & BUTTLE, D.J. Molecules in focus. Cathepsin B, **Int J Biochem Cell Biol**, Holanda, Vol. 29, N° 5, p.715-720, 1997.
- MULLOY, B.; & FORSTER, M. J. Conformation and dynamics of heparin and heparan sulfate, **Glycobiology**, Inglaterra, Vol. 10, N° 11, p. 1147-1156, 2000.
- MULLOY, B.; FORSTER, M. J.; JONES, C.; & DAVIES, D. B. N.m.r. and molecular-modelling studies of the solution conformation of heparin, **Biochem J**, Inglaterra, Vol. 293, N° 1, p. 849-858, 1993.
- MUSIL, D.; ZUCIC, D.; TURK, D.; ENGH, R. A.; MAYR, I.; HUBER, R.; POPOVIC, T.; TURK, V.; TOWATARI, T.; KATUNUMA, N.; & BODE, W. The refined 2.15 Å X-ray crystal structure of human cathepsin B : the structural basis for its specificity, **EMBO J**, Inglaterra, Vol. 10, N° 9, p. 2321-2330, 1991.
- NÄGLER, D. K.; STORER, A. C.; PORTARE, F. C. V.; CARMONA, E.; JULIANE, L.; & MÉNARD, R. Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts, **Biochemistry**, Estados Unidos, Vol. 36, N° 41, p. 12608-12615, 1997.
- NÄGLER, D. K.; TAM, W.; STORER, A. C.; KRUPA, J. C.; MORT, J. S.; & MÉNARD, R. Interdependency of sequence and positional specificities for cysteine proteases of the papain family, **Biochemistry**, Estados Unidos, Vol. 38, N° 15, p. 4868-4874, 1999.
- PASCUTTI, P. G. **Introdução à modelagem molecular** (Apostila). Laboratório de Física Biológica, UFRJ, 2004.

- PATRA, M.; KARTTUNEN, M.; HYVONEN, M. T.; FALCK, E.; LINDQVIST, P.; & VATTULAINEN, I. Molecular Dynamics Simulations of Lipid Bilayers: Major Artifacts Due to Truncating Electrostatic Interactions, **Biophys J**, Estados Unidos, Vol. 84, N° 6, p. 3636-3645, 2003.
- QURAIISHI, O.; NÄGLER, D. K.; FOX, T.; SIVARAMAN, J.; CYGLER, M.; MORT, J. S.; & STORER, A. C. The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide, **Biochemistry**, Estados Unidos, Vol.38, N° 16, p. 5017-23, 1999.
- RAWLINGS, N.D.; & BARRETT, A.J. MEROPS: the peptidase database, **Nucleic Acids Res**, Inglaterra, Vol 27, N° 1, p. 323-325, 2000.
- RICHARDS, F. M. Areas, volumes, packing and protein structure, **Annu Rev Biophys Bioeng**, Estados Unidos, Vol. 6, N° 1, p. 151-176, 1977.
- RICHMOND, T. J. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect, **J Mol Biol**, Inglaterra, Vol. 178, N° 1, p. 63-89, 1984.
- SALMIVIRTA, M.; & JALKANEN, M. Syndecan family of cell surface proteoglycans: developmentally regulated receptors for extracellular effector molecules, **Experientia**, Suíça, Vol. 51, N° 9, p. 863-872, 1995.
- SELZER, P. M.; CHEN, X.; CHAN, V. J.; CHENG, M.; KENYON, G. L.; KUNTZ, I. D.; SAKANARI, J. A.; COHEN, F. E.; & MCKERROW, J. H. Leishmania major: molecular modeling of cysteine proteases and prediction of new nonpeptide inhibitors, **Exp Parasitol**, Estados Unidos, Vol. 87, N° 1, p. 212-221, 1997.
- SHAW, E.; & DEAN, R. T. The inhibition of macrophage protein turnover by a selective inhibitor of thiol proteinases. **Biochem J**, Inglaterra, Vol. 186, N° 1, p. 385-390, 1980.
- SHEAHAN, K.; SHUJA, S.; & MURNANE, M. J. Cysteine Protease Activities and Tumor Development in Human Colorectal Carcinoma, **Cancer Res**, Estados Unidos, Vol. 49, N° 1, p. 3809-3814, 1989.
- SHRAKE, A.; & RUPLEY, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin, **J Mol Biol**, Inglaterra, Vol. 79, N° 2, p. 351-371, 1973.

- SLOANE, B. F.; ROZHIN, J.; JOHNSON, K.; TAYLOR, H.; CRISSMAN, J. D.; & HONN, K. V. Cathepsin B: association with plasma membrane in metastatic tumors, **Proc Natl Acad Sci USA**, Estados Unidos, Vol. 83, N° 1, p. 2483–2487, 1986.
- SLOANE, B. F.; ROZHIN, J.; LAH, T. T.; DAY, N. A.; BUCK, M.; RYAN, R. E.; CRISSMAN, J. D.; & HONN, K. V. Tumor cathepsin B and its endogenous inhibitors in metastasis, **Adv Exp Med Biol**, Estados Unidos, Vol. 233, N° 1, p. 259-268, 1988.
- SÖDERSTRÖM, M.; SALMINEN, H.; GLUMOFF, V.; KIRSCHKE, H.; ARO, H.; & VUORIO, E. Cathepsin expression during skeletal development, **Biochim Biophys Acta**, Holanda, Vol. 1446, N° 1, p. 35-46, 1999.
- SOL-CHURCH, K.; SHIPLEY, J.; BECKMAN, D. A.; MASON, R. W. Expression of cysteine proteases in extraembryonic tissue during mouse embryogenesis, **Arch Biochem Biophys**, Estados Unidos, Vol. 372, N° 1, p.375-381, 1999.
- SPOEL, D.; LINDAHL, E.; HESS, B.; BUUREN, A. R.; APOL, E.; MEULENHOF, P. J.; TIELEMAN, D. P.; SIJBERS, A. L. T. M.; FEENSTRA, K. A.; DRUNEN, R.; & BERENDSEN, H. J. C. **Gromacs User Manual version 3.3**, <www.gromacs.org>, 2005.
- STEINBACH, P. J.; & BROOKS, B. R. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation, **J Comput Chem**, Estados Unidos, Vol. 15, N° 7, p. 667–683, 1994.
- STONE, A. J. The Theory of Intermolecular Forces. **International Series of Monographs on Chemistry**, 1. ed., Vol. 32, Oxford: Clarendon Press, 1996.
- STRAUB, J. E.; & ANDRICIOAEI, I. Computational Methods Inspired by Tsallis Statistics: Monte Carlo and Molecular Dynamics Algorithms for the Simulation of Classical and Quantum Systems, **Brazilian Journal of Physics**, Brasil, Vol. 29, N° 1, p. 179-186, 1999.
- STRINGER, S. E.; & GALLAGHER, J. T. Molecules in focus: heparan Sulfate, **Int J Biochem**, Inglaterra, Vol. 29, N° 5, p. 709-714, 1997.

- SZPADERSKA, A.; & FRANKFATER, A. An intracellular form of cathepsin B contributes to invasiveness in cancer, **Cancer Res**, Estados Unidos, Vol. 61, N° 1, p. 3493-3500, 2001.
- THOMPSON, L. D.; PANTOLIANO, M. W. E.; & SPRINGER, B. A.; Energetic characterization of the basic fibroblast growth factor heparin interaction identification of the heparin binding domain, **Biochemistry**, Estados Unidos, Vol. 33, N° 14, p. 3831-3840, 1994.
- TIRONI, I. G.; SPERB, R.; SMITH, P. E.; & VAN GUNSTEREN, W. F. A. Generalized reaction field method for molecular dynamics simulations, **J Chem Phys**, Estados Unidos, Vol. 102, N° 13, p. 5451-5459, 1995.
- TOLEDO, O. M.; & DIETRICH, C. P. Tissue specific distribution of sulfated mucopolysaccharides in mammals, **Biochim Biophys Acta**, Holanda, Vol. 498, N° 1, p. 114-122, 1977.
- TURK, B.; BIETH, J. G.; BJÖRK, I.; DOLENC, I.; TURK, D.; CIMERMAN, N.; KOS, J.; COLIC, A.; STOKA, V.; & TURK, V. Regulation of the activity of lysosomal cysteine proteinases by pH-induced inactivation and/or endogenous protein inhibitors, cystatins, **Biol Chem Hoppe Seyler**, Alemanha, Vol. 376, N° 4, p. 225-230, 1995.
- TURK, B.; DOLENC, I.; TURK, V.; & BIETH, J. G. Kinetics of the pH-induced inactivation of human cathepsin L, **Biochemistry**, Estados Unidos, Vol. 32, N° 1, p. 375-80, 1993.
- TURK, B.; DOLENC, I.; ZEROVNIK, E.; TURK, D.; GUBENSEK, F.; & TURK, V. Human cathepsin B is a metastable enzyme stabilized by specific ionic interactions associated with the active site, **Biochemistry**, Estados Unidos, Vol. 33, N° 9, p. 14800-14806, 1994.
- TURK, B.; TURK, V.; & TURK, D. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors, **Biol Chem**, Alemanha, Vol. 378, N° 3, 4, p.141-150, 1997.
- TURK, D.; GUNCAR, G.; PODOBNIK, M.; & TURK, B. Revised definition of substrate binding sites of papain-like cysteine proteinases, **Biol Chem**, Alemanha, Vol 379, N° 2, p 137-147, 1998.

- TURNER, J. D.; WEINER, P.; ROBSON, B.; VENUGOPAL, R. H. S. III, & SINGH, R. Reduced variable molecular dynamics, **J Comput Chem**, Estados Unidos, Vol. 16, N° 10, p. 1271–1290, 1995.
- TWINING, S.S. Regulation of proteolytic activity in tissues. **Crit Rev Biochem Mol Biol**, Estados Unidos, Vol 29, N° 5, p. 315-383, 1994.
- VARKI, A.; CUMMINGS, R.; ESKO, J.; FREEZE, H.; HART, G.; & MARTH, J. **Essentials of Glicobiology**. 1. ed. California: The Cons. of Glyco., 1999.
- VERLET.; L. Computer experiments on classical fluids. I. Thermodynamical properties of lennard-jones molecules, **Phys Rev**, Estados Unidos, Vol. 159, N° 1, p. 98-103, 1967.
- VILLADANGOS, J. A.; & PLOEGH, H. L. Proteolysis in MHC class II antigen presentation: who's in charge?, **Immunity**, Estados Unidos, Vol. 12, N° 3, p. 233-239, 2000.
- WANG, B.; SHI, G. P.; YAO, P. M.; LI, Z.; & CHAPMAN, H. A.; BROMME, D. Human cathepsin F. Molecular cloning, functional expression, tissue localization, and enzymatic characterization, **J Biol Chem**, Estados Unidos, Vol. 273, N° 48, p. 32000-32008, 1998.
- WU, X.; & BROOKS, B. R. Isotropic periodic sum: A method for the calculation of long-range interactions, **J Chem Phys**, Estados Unidos, Vol.122, N° 1, p. 044-107, 2005.
- XING, R.; ADDINGTON, A. K.; & MASON, R.W. Quantification of cathepsins B and L in cells, **Biochem J**, Inglaterra, Vol. 332, N° 1, p. 499-505, 1998.
- YAMAMOTO, A.; TOMOO, K.; MIYAGAWA, H.; TAKAOKA, Y.; SUMIYA, S.; KITAMURA, K.; & ISHIDA, T. Molecular dynamics simulations of bovine cathepsin B and its complex with CA074, **Chem Pharm Bull**, Japão, Vol. 48, N° 1, p. 480-485, 2000.
- YAN, S.; SAMENI, M.; & SLOANE, B. F. Cathepsin B and human tumor progression, **Biol Chem**, Alemanha, Vol. 379, N° 2, p. 113-123, 1998.
- YANAGISHITA, M.; & HASCALL, V. C. Mini review. Cell surface heparan sulphate proteoglycans, **J Biol Chem**, Estados Unidos, Vol. 261, N° 7, p. 9451-9454, 1992.

YANAGISHITA, M.; & HASCALL, V. C. Proteoglycans synthesized by rat ovarian granulosa cells in culture. Isolation, fractionation, and characterization of proteoglycans associated with the cell layer, **J Biol Chem**, Estados Unidos, Vol. 259, N° 16, p. 10260-10269, 1984.

ZHANG, L.; DAVID G.; & ESKO J. D. Repetitive Ser-Gly sequences enhance heparan sulphate assembly in proteoglycans, **J Biol Chem**, Estados Unidos, Vol. 270, N° 45, p. 27127-27135, 1995.

ANEXOS

PARTE I

PROGRAMAS DESENVOLVIDOS

MOD_GYR_res.....	154
MOD_GYR_atom.....	157
Solvent Access Module.....	159
MINDIST Module.....	161
GRO2THOR Converter.....	162
New Convolution.....	178
Yet Another Map Plotter, YAMP.....	179
RMS3D_aux.....	180
RMS3D_sum.....	181

Nome do programa: MOD_GYR_res (modificação do método de raio de giração);

Linguagem de programação: Perl;

Descrição: calcula o MOD_GYR de cada resíduo da proteína em função do tempo;

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```
#!/usr/bin/perl

#Utilizar da seguinte forma, por exemplo:
#trjconv -f ../md60nsfinalpapahep_download.xtc -s ../md60nsfinalpapahep_download.tpr -o proteincarcs.gro
-dt 100 -sep | ngyr.pl

print "MOD_GYR (for residues) module created by José Eduardo Kroll, 2008.\n";

$passo = 100;          #ps espaço de tempo entre cada frame...
$total = 60000;       #ps tempo total da simulação
$pdnum = $total / $passo; #número de arquivos PDBs que serão criados
$protnum = 212;       #NUMERO DE RESIDUOS NA PROTEINA

$pasta = "/mnt/Bkp1/MESTRADO/SImulacao/papainaph7/resultdock/60/MODIF_GYRATE/";
$protnome = "proteincarcs";
$proteinanum = 0;
$proteinaname = $protnome . '_' . $proteinanum;
print $proteinaname;
$tempo = 0;

system ("rm -f *$protnome*.");
open (fileout, ">plot.out");
print fileout "Time Res Accessibility ResType\n";

while ($proteinanum <= $pdnum) { #$pdnum
    opendir (filesdir, $pasta);
    @files = readdir (filesdir);

    $freeway = 0;
    $numaux = $proteinanum + 1;
    $proteinanamaux = $protnome . '_' . $numaux;
    foreach $i (@files) {
        if ($proteinanum >= $pdnum - 1 || $i eq "$proteinanamaux.gro") {
            $freeway = 1;
            last;
        }
    }

    if ($freeway == 1) {
        foreach $i (@files) {
            if ($i eq "$proteinanam.gro") {
```

```

open (filersa, "+<$proteinanam.gro");
@ver = <filersa>;
print "Processing: $proteinanam.gro...";

@MATRIX = @ZERO;
@MATRIY = @ZERO;
@MATRIZ = @ZERO;

$initial = 1;
for $poscenter (2..$#ver-2) {
    $res = substr($ver[$poscenter], 0, 5);
    #print "$res\n";
    if ($initial) {
        $resaux = $res;
        $initial = 0;
        #print "QNTAS VEZES??";
    }

    #print "$res $resaux\n";
    if ($res != $resaux) {

        push (@MATRIX, $positionx);
        push (@MATRIY, $positiony);
        push (@MATRIZ, $positionz);
        #print "$positionx $positiony $positionz\n";

        $positionx = 0;
        $positiony = 0;
        $positionz = 0;
        $resaux = $res;

        goto processa;
    } else {
        processa:
        @aces = split (/s+/, $ver[$poscenter]);
        $positionx += $aces[4]; $positionx /= 2;
        $positiony += $aces[5]; $positiony /= 2;
        $positionz += $aces[6]; $positionz /= 2;
        #print "$resnum, XYZ: $aces[4] $aces[5] $aces[6]\n";
    }
}
push (@MATRIX, $positionx);
push (@MATRIY, $positiony);
push (@MATRIZ, $positionz);
#print "$positionx $positiony $positionz\n";

for $residnum (1..$#MATRIX+1) {
    $raio = 0;
    for $residnumaux (1..$#MATRIX+1) {

        if ($residnum != $residnumaux) {
            $raio += ((sqrt( ($MATRIX[$residnum-1]-
$MATRIX[$residnumaux-1])^2 + ($MATRIY[$residnum-1]-$MATRIY[$residnumaux-1])^2 +
($MATRIZ[$residnum-1]-$MATRIZ[$residnumaux-1])^2 ))^2)/2;
        }
    }
}

```


Nome do programa: MOD_GYR_atom (modificação do método de raio de giração);

Linguagem de programação: Perl;

Descrição: calcula o MOD_GYR de cada átomo de uma proteína ou molécula qualquer;

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```
#!/usr/bin/perl

#Utilizar da seguinte forma, por exemplo:
#trjconv -f ../md60nsfinalpapahep_download.xtc -s ../md60nsfinalpapahep_download.tpr -o proteincarcs.gro
-dt 100 -sep | ngyr.pl

print "MOD_GYR (for atoms) module created by José Eduardo Kroll, 2008.\n";

$passo = 100;          #ps espaço de tempo entre cada frame...
$ttotal = 60000;      #ps tempo total da simulação
$pdbnum = $ttotal / $passo;  #número de arquivos PDBs que serão criados
$protresnum = 79;     #NUMERO DE ÁTOMOS DA MOLÉCULA

$pasta = "/mnt/Bkp1/MESTRADO/SImulacao/papainaph7/resultdock/60/MODIF_GYRATE/";
$protname = "proteincarcs";
$proteinanum = 0;
$proteinanam = $protname . '_' . $proteinanum;
print $proteinanam;
$tempo = 0;

system ("rm -f *$protname*.");
open (fileout, ">plothep.out");
print fileout "Time Res Accessibility ResType\n";

while ($proteinanum<=$pdbnum) { #$pdbnum
    opendir (filesdir, $pasta);
    @files = readdir (filesdir);

    $freeway = 0;
    $numaux = $proteinanum + 1;
    $proteinanamaux = $protname . '_' . $numaux;
    foreach $i (@files) {
        if ($proteinanum>=$pdbnum-1 || $i eq "$proteinanamaux.gro") {
            $freeway = 1;
            last;
        }
    }
}

if ($freeway==1) {
    foreach $i (@files) {
        if ($i eq "$proteinanam.gro") {
```

```

open (filersa, "+<$proteinanam.gro");
@ver = <filersa>;
print "Processing: $proteinanam.gro...";

@MATRIX = @ZERO;
@MATRIY = @ZERO;
@MATRIZ = @ZERO;

$initial = 1;
for $poscenter (2..$#ver-1) {

    @aces = split (/s+/, $ver[$poscenter]);
    $positionx += $aces[4];
    $positiony += $aces[5];
    $positionz += $aces[6];
    push (@MATRIX, $positionx);
    push (@MATRIY, $positiony);
    push (@MATRIZ, $positionz);
}

for $residnum (1..$#MATRIX+1) {
    $raio = 0;
    for $residnumaux (1..$#MATRIX+1) {

        if ($residnum != $residnumaux) {
            $raio += ((sqrt( ($MATRIX[$residnum-1]-
$MATRIX[$residnumaux-1])^2
+ ($MATRIY[$residnum-1]-$MATRIY[$residnumaux-1])^2
+ ($MATRIZ[$residnum-1]-$MATRIZ[$residnumaux-1])^2 ))/2);
        }
    }

    $raio /= $protresnum;
    $raio_aux = sprintf ("%0.3f", $raio);
    $tempo_aux = sprintf ("%0.3f", $tempo);
    #print "CHEGOU?\n";
    print fileout "$tempo_aux, $residnum, $raio_aux \n";
}

close filersa;
close filesdir;

system ("rm -f $proteinanam.*");
print "OK \n";

$tempo += $tpasso;
$proteinanum++;
$proteinanam = $protname . '_' . $proteinanum;
last;
}
}
}
}
close;

```

Nome do programa: Solvent Access Module (utilizando como base o programa naccess);

Linguagem de programação: Perl;

Descrição: Calcula a acessibilidade de cada resíduo da proteína em função do tempo.

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```
#!/usr/bin/perl

print "Solvent Accessibility module created by José Eduardo Kroll, 2008.\n";

$passo = 20;          #ps  espaço de tempo entre cada frame...
$total = 60000;      #ps tempo total da simulação
$pdnum = $total / $passo;  #número de arquivos PDBs que serão criados

$pasta = "/home/kroll/ACS/";  #PATH onde se encontra esse programa
$protnome = "proteinarcs";
$proteinanum = 0;
$proteinanam = $protnome . '_' . $proteinanum;
print $proteinanam;
$tempo = 0;

system ("rm -f *$protnome*.");
open (fileout, ">plot_papahep.out");
print fileout "Time\t\tAcs Abs\tResCheck\n";

while ($proteinanum <= $pdnum) { # $pdnum
    opendir (filesdir, $pasta);
    @files = readdir (filesdir);

    $freeway = 0;
    $numaux = $proteinanum + 1;
    $proteinanamaux = $protnome . '_' . $numaux;
    foreach $i (@files) {
        if ($proteinanum >= $pdnum - 1 || $i eq "$proteinanamaux.pdb") {
            $freeway = 1;
            last;
        }
    }

    if ($freeway == 1) {
        foreach $i (@files) {
            if ($i eq "$proteinanam.pdb") {
                system (".naccess $proteinanam.pdb");

                open (filersa, "+<$proteinanam.rsa");
                @ver = <filersa>;
                print "Processing: $proteinanam.pdb...";
                $tempo_aux = sprintf ("%3f", $tempo);

                for $resnum (1..212) {
```

```
        @aces = split (/s+/, $ver[$resnum+2]);
        print fileout "$tempo_aux\t$resnum\t$aces[5]\t$aces[1]\n";
    }

    close filersa;
    close filesdir;

    system ("rm -f $proteinanam.*");
    print "OK \n";

    $tempo += $tpasso;
    $proteinanum++;
    $proteinanam = $protname . '_' . $proteinanum;
    last;
}
}
}
close;
```

Nome do programa: MINDIST Module;

Linguagem de programação: Perl;

Descrição: Calcula a distância entre átomos ou res. de moléculas em função do tempo.

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```
#!/usr/bin/perl

#ifinalpapahep.xtc -s md40ns_papahep.tpr -n indicehep.ndx -or mindteste.xvg -b 0 -e 0 < grupos.txt | ./mnd.pl

print "teste...";

$dt = 100;           #PASSO DE TEMPO
$tt = 60000;        #TEMPO TOTAL
$tf = $tt/$dt;      #TOTAL DE ARQUIVOS A SEREM CRIADOS

open (outfile, ">mndplot_HS.out");

for $tempo(0..$tt/$dt) {
    $realtempo = $dt * $tempo;
    system ("g_mindist -f md60nsfinalpapahep.xtc -s md40ns_papahep.tpr -n indicehep.ndx -or
mindteste.xvg -b $realtempo -e $realtempo < grupos.txt");

    open (mndfile, "+<mindteste.xvg");
    @lista1 = <mndfile>;

    for $residuos(1..212) {
        @linha = split /\s+/, $lista1[$residuos+18];
        print outfile "$realtempo, $residuos, $linha[2]\n";
    }
    close mndfile;

    system ("rm -f *.xvg*");
}

close;
```

Nome do programa: GRO2THOR Converter;

Linguagem de programação: Perl;

Descrição: Converte topologias e estruturas entre os programas THOR e GROMACS.

Observação: criado por José Eduardo Kroll, 2007; Versão final está em desenvolvimento.

Código fonte:

```
#!/usr/bin/perl
```

```
### ESTRUTURA PRINCIPAL DO PROGRAMA
#####

### INICIALIZA VARIÁVEIS GERAIS
INITVARS();
### TRANSFORMAR ARQUIVOS DE TOPOLOGIA
OPENTOPOLOGY();
### ABRE ARQUIVOS PARA LER E ESCREVER
OPENFILES();
### CARREGAR DADOS DO ARQUIVO TOP
LOADTOPDATA();
### CONTAR LIGAÇÕES DE H
COUNTHBONDS();
### IMPRIMIR TABELA BONDS H
BONDSPRINT(1);
### IMPRIMIR TABELA BONDS NÃO H
BONDSPRINT(2);
### IMPRIMIR ANGLE
ANGLEPRINT();
### IMPRIMIR IMPROPER DIHEDRE
IMPRODIPRINT();
### IMPRIMIR PROPER DIHEDRE
PRODIPRINT();
### IMPRIMIR VIZINHOS EXCLUIDOS
PRINTNEIGH1();
### IMPRIMIR TERCEIROS VIZINHOS
PRINTNEIGH3();
# IMPRIMIR LISTA DE MASSAS
PRINTMASSLIST();
# IMPRIMIR LISTA DE CONSTRAINT
PRINTCONSTR();
### CRIA ZMATRIZ COM BABEL E CARREGA DADOS
LOADZMT();
### IMPRIMIR ARQUIVO DE ESTRUTURA .IN
STRUCTPRINT();
### FECHA ARQUIVOS

#####
### FIM DA ESTRUTURA PRINCIPAL

##### SUBROTINAS #####
```

```
### INICIALIZA VARIÁVEIS GERAIS
```

```
sub INITVARS {
```

```
### ARQUIVOS QUE DEVEM SER ABERTOS -- GROMACS FILES
```

```
$grofile = "./papaph7.gro";
```

```
$stopfile = "./papaph7.top";
```

```
##saida THOR
```

```
$THOR_INout = "./output.IN";
```

```
$THOR_TOPout = "./output.TOP" ;
```

```
# NUMERO DO ATOMO INICIAL / ULTIMO NUMERO DA OUTRA ESTRUTURA
```

```
#ZERO É DEFAULT -- AGORA, APLICADO SOMENTE NA LISTA DE VIZINHOS
```

```
$INI_NATOM = 0;
```

```
## ARQUIVOS DE SISTEMA -- NAO PRECISA MEXER...
```

```
$xyzfile = "./system.xyz";
```

```
$zmtfile = "./system.zmt";
```

```
## ARQUIVOS DE CAMPO DE FORÇA...TODOS OS TIPOS DISPONIVEIS
```

```
$ffchoice = 1; #ESCOLHA DO CAMPO DE FORÇA A SER UTILIZADO
```

```
$forcefield[1] = "ffG43a1bon.itp"; # GROMOS 43a1
```

```
#$forcefield[1] = "/gro43a1.pl"; # GROMOS 43a1
```

```
#$forcefield[1] = "/gro43a1.pl";
```

```
#$forcefield[1] = "/gro43a1.pl";
```

```
#$forcefield[1] = "/gro43a1.pl";
```

```
#$forcefield[1] = "/gro43a1.pl";
```

```
#$forcefield[1] = "/gro43a1.pl";
```

```
@secoes = qw(moleculetype atoms bonds pairs angles dihedrals position_restraints system molecules);
```

```
#PROPRIEDADES DO ATOMOS, UTILIZADO NA LISTA DE MASSAS
```

```
#OS DEMAIS PARAMETROS ESTAO NA SECAO QUE IMPRIME A TABELA DE MASSAS
```

```
@atypelist = qw(O OM OA OW N NT NL NR5 NR5X NP C CH1 CH2 CH3 CR5 CR6 CB H HO HW  
HS S FE ZN NZ NE P OS CS1 NR6 NR6X CS2 SI NA CL CA MG H3 H2 HS);
```

```
$atypelist_propert[0] = "1 2 3 4 28"; #OXYGEN
```

```
$atypelist_propert[1] = "5 6 7 8 9 10 25 26 30 31"; #NITROGEN
```

```
$atypelist_propert[2] = "11 12 13 14 15 16 17 29 32"; #CARBON
```

```
$atypelist_propert[3] = "18 19 20 21 38 39"; #HYDROGEN
```

```
$atypelist_propert[4] = "22 40"; #SULFUR
```

```
$atypelist_propert[5] = "23 24 27 33 34 35 36 37"; #OUTROS COM CARACTERISTICAS
```

```
INDIVIDUAIS..
```

```
}
```

```
### TRANSFORMAR ARQUIVOS DE TOPOLOGIA
```

```
sub OPENTOPOLOGY {
```

```
open ($stopology, $forcefield[$ffchoice]) or die "Problemas com o arquivo de campo de força fonte...";
```

```
while (<$stopology>) {
```

```
    $linhaf = $_;
```

```
    chomp ($linhaf);
```

```
    @linedata = split(/\s+/, $linhaf);
```

```
    #printf $linedata[0] . "\n";
```

```
    if ($linedata[0] eq '#define') {
```

```
        $ffkind = substr($linedata[1], 0, 2);
```

```
        $ffnumb = substr($linedata[1], 3, length($linedata[1])-3);
```



```

}
### SECAO [ ATOMS ]
elseif ($POS_secao == 1) {
    @linedata = split(/\s+/, $linhaf);
    for $natom (1...8) { push (@atom, $linedata[$natom]) };
}

### SECAO [ BONDS ]
elseif ($POS_secao == 2) {

    ### TRADUZIR ALIAS DE ENERGIAS GB
    for ($nalisitem=0; $nalisitem<=#forcefield_gb; $nalisitem+=3) {
        $aliasff = ' '. $forcefield_gb[$nalisitem] . ' ';
        $linhaf .= ' ';
        if (' '.$linhaf.' ' =~ m{$aliasff}) {
            $energff = "$forcefield_gb[$nalisitem+1]
$forcefield_gb[$nalisitem+2]";
            $linhaf =~ s{$aliasff} {$energff};
        }
    }

    ### QUEBRAR LINHA PARA ARRAY
    @linedata = split(/\s+/, $linhaf);
    for $natom (1...5) { push (@bonds, $linedata[$natom]) };
    if ($linedata[3] != 2) { die "Bond function unknow in GROMACS topology file...
$linedata[3] \n" };

##### TABELA DE LIGAÇÕES PARA OS PARAMETROS DE ESTRUTURA E VIZINHOS...
    $atombond1 = $linedata[1];
    $atombond2 = $linedata[2];
    $atombondtable1[$atombond1] = $atombondtable1[$atombond1] . ' '. $atombond2;
    $atombondtable1[$atombond2] = $atombondtable1[$atombond2] . ' '. $atombond1;
    #####
}

### SECAO [ PAIRS ]
elseif ($POS_secao == 3) {
    @linedata = split(/\s+/, $linhaf);
    for $natom (1...2) { push (@pairs, $linedata[$natom]) };
    if ($linedata[3] != 1) { die "Pair function unknow in GROMACS topology file...
$linedata[3] \n" };
}

### SECAO [ ANGLES ]
elseif ($POS_secao == 4) {

    ### TRADUZIR ALIAS DE ENERGIAS GA
    for ($nalisitem=0; $nalisitem<=#forcefield_ga; $nalisitem+=3) {
        $aliasff = ' '. $forcefield_ga[$nalisitem] . ' ';
        $linhaf .= ' ';
        if (' '.$linhaf.' ' =~ m{$aliasff}) {
            $energff = "$forcefield_ga[$nalisitem+1]
$forcefield_ga[$nalisitem+2]";
            $linhaf =~ s{$aliasff} {$energff};
        }
    }

    ### QUEBRAR LINHA PARA ARRAY
    @linedata = split(/\s+/, $linhaf);

```

```

for $natom (1..6) { push (@angles, $linedata[$natom]) };
if ($linedata[4] != 2) { die "Angle function unknow in GROMACS topology file..."
$linedata[4] \n" };
}

### SECAO [ dihedrals ]
elseif ($POS_secao == 5) {
    @linedata = split(/\s+/, $linhaf);
    if ($linedata[5]==2) {

        ### TRADUZIR ALIAS DE ENERGIAS GI - DIEDRO IMPROPRIO
        for ($nalisitem=0; $nalisitem<=#forcefield_gi; $nalisitem+=3) {
            $aliasff = '$forcefield_gi[$nalisitem] .';
            $linhaf .= ' ';
            if ('.$linhaf.' =~ m{$aliasff}) {
                $energff = "$forcefield_gi[$nalisitem+1]
$forcefield_gi[$nalisitem+2]";
                $linhaf =~ s{$aliasff} {$energff};
            }
        }

        ### QUEBRAR LINHA PARA ARRAY
        @linedata = split(/\s+/, $linhaf);
        for $natom (1..7) { push (@diedralimp, $linedata[$natom]) };
    }
    elseif (($linedata[5]==1)) {

        ### TRADUZIR ALIAS DE ENERGIAS GD - DIEDRO PROPRIO
        for ($nalisitem=0; $nalisitem<=#forcefield_gd; $nalisitem+=4) {
            $aliasff = '$forcefield_gd[$nalisitem] .';
            $linhaf .= ' ';
            if ($linhaf =~ m{$aliasff}) {
                ### BUG FEIO NO PERL... $forcefield_gd[$nalisitem+3]
                NAO FUNCIONA DIREITO!!! NAO TEM LOGICA
                ### TESTE ISOLADOS (AQUI MESMO) MOSTRARAM
                QUE ELE FUNCIONA...
                $energff = "$forcefield_gd[$nalisitem+1]
$forcefield_gd[$nalisitem+2] $forcefield_gd[$nalisitem+3]";
                #print "$energff\n";
                $linhaf =~ s{$aliasff} {$energff};
            }
        }

        ### QUEBRAR LINHA PARA ARRAY
        @linedata = split(/\s+/, $linhaf);
        for $natom (1..8) { push (@diedralprop, $linedata[$natom]) };
    }
}
}
}

### CONTAR LIGACOES DE H E NAO H ----- OK!
sub COUNTHBONDS {
    $maxbondh = 0;
    $maxbondo = 0;

```

```

for ($ver=0; $ver<=$#bonds; $ver+=5) {
    ### CARREGA DADOS BONDS
    $atombond1 = $bonds[$ver];
    $atombond2 = $bonds[$ver+1];
    $atomtype1 = $atom[-7 + $atombond1 * 8];
    $atomtype2 = $atom[-7 + $atombond2 * 8];
    if (uc($atomtype1) eq 'H' || uc($atomtype2) eq 'H') {
        $maxbondh++;
    } else {
        $maxbondo++;
    }
}
}

### IMPRIMIR TABELA BONDS, OPCAO 1 = H, 2 = NAO H ----- FUNCIONANDO OK
sub BONDSPRINT {
    @opcao_subbnd = @_ ;

    if ($opcao_subbnd[0] == 1) {
        ### IMPRIMIR BONDS COM H
        print $hTOPOUTfile "\n      HOOK TABLE  I-J\n\n";
        print $hTOPOUTfile "LIGACOES COM HIDROGENIOS=" . ' ' x (5 - length($maxbondh)) .
"$maxbondh\n\n";
    } else {
        ### IMPRIMIR BONDS SEM H
        print $hTOPOUTfile "\nLIGACOES SEM HIDROGENIOS=" . ' ' x (5 - length($maxbondo)) .
"$maxbondo\n\n";
    }

    $posh=0; ##POSICAO LIG
    for ($ver=0; $ver<=$#bonds; $ver+=5) {

        ### CARREGA DADOS BONDS
        $atombond1 = $bonds[$ver];
        $atombond2 = $bonds[$ver+1];

        ### CARREGA RESPECTIVAS INF ATOMS
        $atomtype1 = $atom[-7 + $atombond1 * 8];
        $atomtype2 = $atom[-7 + $atombond2 * 8];

        #MODELO LIGACOES THOR
        #  5)--  25  26    N -  H    895.0000  1.0000
        #  6)--  29  30    OA - H    750.0000  1.0000
        #  7)--  33  34    N -  H    895.0000  1.0000
        #  8)--  45  38    HC - C    700.0000  1.0900

        ### DEPENDENDO DA OPCAO, ELE PULA AS LIGACOES COM H E VICE VERSA
        if (uc($atomtype1) ne 'H' && uc($atomtype2) ne 'H' && $opcao_subbnd[0] == 1) { next; }
        if (uc($atomtype1) eq 'H' || uc($atomtype2) eq 'H' && $opcao_subbnd[0] == 2) { next; }

        $posh++;
        ### IMPRIMIR LIGACOES

        $printbondener = sprintf("%.4f", $bonds[$ver+4]);
        $printbonddist = sprintf("%.4f", $bonds[$ver+3]);

        ### ALGUEM ENTENDE ESSA CONFUSAO?

```

```

    $outprint = ' ' x (6 - length($posh)) . "$posh -- " . ' ' x (6 - length($atombond1)) . $atombond1 .
      ' ' x (6 - length($atombond2)) . $atombond2 . ' ' x (4 - length($atomtype1)) .
"$atomtype1 - " .
      ' ' x (4 - length($atomtype2)) . $atomtype2 . ' ' x (16 - length($sprintbondener)) .
$sprintbondener .
      ' ' x (10 - length($sprintbonddist)) . $sprintbonddist . "\n";
    print $hTOPOUTfile $outprint;
  }
}

```

```

### IMPRIMIR TABELA ANGLES -----

```

```

sub ANGLEPRINT {

```

```

    print $hTOPOUTfile "\n      BOND TABLE  I-J-K\n\n";

```

```

    $posh=0;

```

```

    for ($sver=0; $sver<=$#angles; $sver+=6) {

```

```

        ### CARREGA DADOS BONDS

```

```

        $atombond1 = $angles[$sver];

```

```

        $atombond2 = $angles[$sver+1];

```

```

        $atombond3 = $angles[$sver+2];

```

```

        ### CARREGA RESPECTIVAS INF ATOMS

```

```

        $atomtype1 = $atom[-7 + $atombond1 * 8];

```

```

        $atomtype2 = $atom[-7 + $atombond2 * 8];

```

```

        $atomtype3 = $atom[-7 + $atombond3 * 8];

```

```

        #      BOND TABLE  I-J-K

```

```

        #

```

```

        #  1)--  1  2  3    NL - CH1 - CH2  110.0000 109.5000 --- QUAL ◆ QUAL??

```

```

        #  2)--  1  2  7    NL - CH1 - C   110.0000 109.5000

```

```

        #  3)--  2  3  4    CH1 - CH2 - CH1  110.0000 111.0000

```

```

        #  4)--  2  7  8    CH1 - C - O   120.0000 121.0000

```

```

        #  5)--  2  7 12    CH1 - C - N   120.0000 115.0000

```

```

        $posh++;

```

```

        ### IMPRIMIR LIGACOES

```

```

        $sprintbondener = sprintf("%.4f", $angles[$sver+5]);

```

```

        $sprintbonddist = sprintf("%.4f", $angles[$sver+4]);

```

```

        ### ALGUEM ENTENDE ESSA CONFUSAO?

```

```

        $outprint = ' ' x (6 - length($posh)) . "$posh -- " . ' ' x (6 - length($atombond1)) . $atombond1 .

```

```

          ' ' x (6 - length($atombond2)) . $atombond2 . ' ' x (6 - length($atombond3)) . $atombond3 .

```

```

          ' ' x (10 - length($atomtype1)) . "$atomtype1 - " . ' ' x (5 - length($atomtype2)) .

```

```

"$atomtype2 - " .

```

```

          ' ' x (5 - length($atomtype3)) . $atomtype3 . ' ' x (12 - length($sprintbondener)) .

```

```

          $sprintbondener . ' ' x (10 - length($sprintbonddist)) . $sprintbonddist . "\n";

```

```

        print $hTOPOUTfile $outprint;
    }
}

```

```

### IMPRIMIR TABELA DIEDROS ----- FUNCIONANDO OK

```

```

sub IMPRODIPRINT {

```

```

    print $hTOPOUTfile "\n      IMPROPER TORSION TABLE  I-X-Y-L\n\n";

```

```

$posh=0;
for ($ver=0; $ver<=$#diedralimp; $ver+=7) {

    ### CARREGA DADOS BONDS
    $atombond1 = $diedralimp[$ver];
    $atombond2 = $diedralimp[$ver+1];
    $atombond3 = $diedralimp[$ver+2];
    $atombond4 = $diedralimp[$ver+3];

    ### CARREGA RESPECTIVAS INF ATOMS
    $atomtype1 = $atom[-7 + $atombond1 * 8];
    $atomtype4 = $atom[-7 + $atombond4 * 8];

    #      IMPROPER TORSION TABLE  I-X-Y-L
    #
    #  1)--  2  1  7  3  CH1 - X - Y - CH2  80.0000  35.26439
    #  2)--  4  6  5  3  CH1 - X - Y - CH2  80.0000  35.26439
    #  3)--  7  2 12  8   C - X - Y -  O  40.0000  0.00000
    #  4)-- 12  7 13 16   N - X - Y - CH2  40.0000  0.00000
    #  5)-- 13 12 17 14  CH1 - X - Y - CH2  80.0000  35.26439
    #  6)-- 17 13 19 18   C - X - Y -  O  40.0000  0.00000

    $posh++;
    ### IMPRIMIR LIGACOES
    $printbondener = sprintf("%.4f", $diedralimp[$ver+6]);
    $printbonddist = sprintf("%.5f", $diedralimp[$ver+5]);

    ### ALGUEM ENTENDE ESSA CONFUSAO?
    $outprint = ' ' x (6 - length($posh)) . "$posh -- " . ' ' x (6 - length($atombond1)) . $atombond1 .
        ' ' x (6 - length($atombond2)) . $atombond2 . ' ' x (6 - length($atombond3)) . $atombond3 .
        ' ' x (6 - length($atombond4)) . $atombond4 . ' ' x (7 - length($atomtype1)) . "$atomtype1 -
X - Y - " .
        ' ' x (5 - length($atomtype4)) . $atomtype4 . ' ' x (10 - length($printbondener)) .
$printbondener .
        ' ' x (12 - length($printbonddist)) . $printbonddist . "\n";
    print $hTOPOUTfile $outprint;
}
}

### IMPRIMIR TABELA DIEDROS ----- OK!!!!
sub PRODIPRINT {

    print $hTOPOUTfile "\n      PROPER TORSION TABLE  X-I-J-Y\n\n";

    $posh=0;
    for ($ver=0; $ver<=$#diedralprop; $ver+=8) {

        ### CARREGA DADOS BONDS
        $atombond1 = $diedralprop[$ver];
        $atombond2 = $diedralprop[$ver+1];
        $atombond3 = $diedralprop[$ver+2];
        $atombond4 = $diedralprop[$ver+3];

        ### CARREGA RESPECTIVAS INF ATOMS
        $atomtype2 = $atom[-7 + $atombond2 * 8];

```

```

$atomtype3 = $atom[-7 + $atombond3 * 8];

#      PROPER TORSION TABLE  X-I-J-Y
#  1)--  1  2  3  4  X - CH1 - CH2 - Y  1.4000  0.0000  3
#  2)--  1  2  7 12  X - CH1 - C - Y   0.2400  0.0000  6
#  3)--  2  3  4  5  X - CH2 - CH1 - Y  1.4000  0.0000  3
#  4)--  9  1  2  7  X - NL - CH1 - Y   0.9000  0.0000  3
#  5)--  7 12 13 17  X - N - CH1 - Y   0.2400 180.0000  6

$posh++;
### IMPRIMIR LIGACOES
$printbondener = sprintf("%.4f", $diedralprop[$ver+6]);
$printbonddist = sprintf("%.4f", $diedralprop[$ver+5]);
$printbondunk = sprintf("%.0f", $diedralprop[$ver+7]);

### ALGUEM ENTENDE ESSA CONFUSAO?
$outprint = ' ' x (6 - length($posh)) . "$posh -- " . ' ' x (6 - length($atombond1)) . $atombond1 .
' ' x (6 - length($atombond2)) . $atombond2 . ' ' x (6 - length($atombond3)) . $atombond3 .
' ' x (6 - length($atombond4)) . $atombond4 . " X - " . ' ' x (5 - length($atomtype2)) .
"$atomtype2 -" .
' ' x (5 - length($atomtype3)) . "$atomtype3 - Y" . ' ' x (10 - length($printbondener)) .
$printbondener .
' ' x (10 - length($printbondist)) . $printbondist . ' ' x (5 - length($printbondunk)) .
$printbondunk . "\n";
print $hTOPOUTfile $outprint;
}
}

#IMPRIMIR E CALCULAR VIZINHOS EXCLUIDOS...duh! ----- FUNCIONANDO OK!!!
sub PRINTNEIGHE {

    print $hTOPOUTfile "\n      VIZINHOS EXCLUIDOS\n\n";

    ## PROCURANDO VIZINHOS EXCLUIDOS
    for ($ver=0; $ver<=$#atom; $ver+=8) {
        $neighbors_aux = "";
        $atompos = (($ver/8)+1);
        @atombondtable2 = split(/\s+/, $atombondtable1[$atompos]);
        $neighbors_aux = $neighbors_aux . " $atombondtable1[$atompos]";

        for ($ver2=0; $ver2<=$#atombondtable2; $ver2++) {
            @atombondtable3 = split(/\s+/, $atombondtable1[$atombondtable2[$ver2]]);
            $neighbors_aux = $neighbors_aux . " $atombondtable1[$atombondtable2[$ver2]]";

            for ($ver3=0; $ver3<=$#atombondtable3; $ver3++) {
                $neighbors_aux = $neighbors_aux . "
$atombondtable1[$atombondtable3[$ver3]]";
            }
        }
    }

    ## TIRAR ATOMOS REPETENTES - VIZINHOS EXCLUIDOS
    @neighbors = split(/\s+/, $neighbors_aux);
    for ($rrepet1=0; $rrepet1<=$#neighbors; $rrepet1++) {
        if ($neighbors[$rrepet1] <= $atompos && $neighbors[$rrepet1] != 0) {
            $neighbors[$rrepet1] = '0';
        }
    }
}

```

```

    }
    for ($rrepet2=0; $rrepet2<=$#neighbors; $rrepet2++) {
        if ($rrepet1 != $rrepet2) {
            if ($neighbors[$rrepet1] == $neighbors[$rrepet2] &&
$neighbors[$rrepet2] != 0) {
                $neighbors[$rrepet2] = '0';
                $rrepet2 = 0;
            }
        }
    }
}
## COLOCAR EM ORDEM CRESCENTE - VIZINHOS EXCLUIDOS
for ($rrepet1=0; $rrepet1<=$#neighbors; $rrepet1++) {
    for ($rrepet2=0; $rrepet2<=$rrepet1; $rrepet2++) {
        if ($rrepet1 != $rrepet2) {
            if ($neighbors[$rrepet2] >= $neighbors[$rrepet1]) {
                $neigh_aux = $neighbors[$rrepet1];
                $neighbors[$rrepet1] = $neighbors[$rrepet2];
                $neighbors[$rrepet2] = $neigh_aux;
            }
        }
    }
}
}

#
#     VIZINHOS EXCLUIDOS
#
# 1 9 2 3 4 7 8 9 10 11 12
# 2 12 3 4 5 6 7 8 9 10 11 12 13 16
# 3 9 4 5 6 7 8 9 10 11 12
# 4 3 5 6 7

$outprint = '' . ($atompos + $INI_NATOM);
@outprint_aux = @ZERO;
foreach $i (@neighbors) { if ($i!=0) { push (@outprint_aux, ($i + $INI_NATOM)); }}
$outprint = $outprint . ' ' . ($#outprint_aux + 1);
foreach $i (@outprint_aux) { $outprint = $outprint . " $i"; }

print $hTOPOUTfile $outprint ."n";
}
}

```

#IMPRIMIR E CALCULAR TERCEIROS VIZINHOS -- OK!!!!

sub PRINTNEIGH3 {

```

    print $hTOPOUTfile "\n      TERCEIROS VIZINHOS\n\n";

```

PROCURANDO TERCEIROS VIZINHOS

```

for ($ver=0; $ver<=$#atom; $ver+=8) {
    $neighbors_aux = "";
    $neighbors_3 = "";
    $atompos = (($ver/8)+1);
    @atombondtable2 = split(/\s+/, $atombondtable1[$atompos]);
    $neighbors_aux = $neighbors_aux . " $atombondtable1[$atompos]";

    for ($ver2=0; $ver2<=$#atombondtable2; $ver2++) {
        @atombondtable3 = split(/\s+/, $atombondtable1[atombondtable2[$ver2]]);
    }
}

```



```

$neighbors_aux = $neighbors_aux . " $atombondtable1[$atombondtable2[$ver2]]";

for ($ver3=0; $ver3<=#atombondtable3; $ver3++) {
    $neighbors_3 = $neighbors_3 . "
$atombondtable1[$atombondtable3[$ver3]]";
}
}

## ISOLAR OS AUTENTICOS TERCEIROS VIZINHOS - TERCEIROS VIZINHOS
@neighbors = split(/\s+/, $neighbors_aux);
@neighbors3 = split(/\s+/, $neighbors_3);
for ($rrepet1=0; $rrepet1<=#neighbors; $rrepet1++) {
    if ($neighbors[$rrepet1] <= $atompos && $neighbors[$rrepet1] != 0) {
        $neighbors[$rrepet1] = '0';
    }
    for ($rrepet2=0; $rrepet2<=#neighbors3; $rrepet2++) {
        if ($neighbors[$rrepet1] == $neighbors3[$rrepet2] &&
$neighbors3[$rrepet2] != 0) {
            $neighbors3[$rrepet2] = '0';
            $rrepet2 = 0;
        }
    }
}

## TIRAR ATOMOS REPETENTES - TERCEIROS VIZINHOS
for ($rrepet1=0; $rrepet1<=#neighbors3; $rrepet1++) {
    if ($neighbors3[$rrepet1] <= $atompos && $neighbors3[$rrepet1] != 0) {
        $neighbors3[$rrepet1] = '0';
    }
    for ($rrepet2=0; $rrepet2<=#neighbors3; $rrepet2++) {
        if ($rrepet1 != $rrepet2) {
            if ($neighbors3[$rrepet1] == $neighbors3[$rrepet2] &&
$neighbors3[$rrepet2] != 0) {
                $neighbors3[$rrepet2] = '0';
                $rrepet2 = 0;
            }
        }
    }
}

## COLOCAR EM ORDEM CRESCENTE - TERCEIROS VIZINHOS
for ($rrepet1=0; $rrepet1<=#neighbors3; $rrepet1++) {
    for ($rrepet2=0; $rrepet2<=$rrepet1; $rrepet2++) {
        if ($rrepet1 != $rrepet2) {
            if ($neighbors3[$rrepet2] >= $neighbors3[$rrepet1]) {
                $neigh_aux = $neighbors3[$rrepet1];
                $neighbors3[$rrepet1] = $neighbors3[$rrepet2];
                $neighbors3[$rrepet2] = $neigh_aux;
            }
        }
    }
}

#
#   TERCEIROS VIZINHOS
#
# 1 9 2 3 4 7 8 9 10 11 12

```

```

# 2 12 3 4 5 6 7 8 9 10 11 12 13 16
# 3 9 4 5 6 7 8 9 10 11 12
# 4 3 5 6 7

$outprint = ' ' . ($atompos + $INI_NATOM);
@outprint_aux = @ZERO;
foreach $i (@neighbors3) { if ($i!=0) { push (@outprint_aux, ($i + $INI_NATOM)); }}
$outprint = $outprint . ' ' . ($#outprint_aux + 1);
foreach $i (@outprint_aux) { $outprint = $outprint . " $i"; }

print $hTOPOUTfile $outprint . "\n";
}
}

# IMPRIMIR LISTA DE MASSAS ----- OK
sub PRINTMASSLIST {

# Massas dos Atomos pertencentes a Molecula
# 17 11 C 12.01100 6 4
# 18 1 O 15.99940 8 6
# 19 5 N 14.00670 7 5
# 20 18 H 1.00800 1 1
# natom nsei type mass full neleo

print $hTOPOUTfile "\n Massas dos Atomos pertencentes a Molecula\n\n";

for ($ver=0; $ver<=$#atom; $ver+=8) {
    $atompos = (($ver/8)+1);
    $atomtype1 = $atom[$ver+1];
    $atommass = $atom[$ver+7]; $atommass = sprintf("%.5f", $atommass);

#VARIAVEIS QUE DEPENDEM DO TIPO DO ATOMO
for ($finda=0; $finda <= $#atypelist; $finda++) {
    $atypelist_dt = 0;
    if (length($atypelist[$finda]) > 3) {
        $atypelist_aux = substr($atypelist[$finda], 1, 3);
        $atypelist_dt++;
    } else {
        $atypelist_aux = $atypelist[$finda];
    }
    if ($atypelist_aux eq $atomtype1) {
        $atomthorlist = 1 + $finda + $atypelist_dt;
        $firstletter = substr($atypelist[$finda], 1, 1);

# DEFINE A CARACTERISTICA DE CADA ATOMO (n DE ELETRONS, ELETRONS NO ULTIMO
# ORBITAL...)
for ($finda2=0; $finda2 <= $#atypelist_propert; $finda2++) {
    @atypelist_propert_aux = split(/\s+/, $atypelist_propert[$finda2]);
    for ($finda3=0; $finda3 <= $#atypelist_propert_aux; $finda3++) {
        if ($atypelist_propert_aux[$finda3] == $atomthorlist) {
            if ($finda2 == 0) { #OXYGEN
                $atomthorchg1 = 8; $atomthorchg2 = 6;
            } elsif ($finda2 == 1) { #NITROGEN

```

```

        $atomthorchg1 = 7; $atomthorchg2 = 5;
    } elsif ($finda2 == 2) { #CARBON
        $atomthorchg1 = 6; $atomthorchg2 = 4;
    } elsif ($finda2 == 3) { #HYDROGEN
        $atomthorchg1 = 1; $atomthorchg2 = 1;
    } elsif ($finda2 == 4) { #SULFUR
        $atomthorchg1 = 16; $atomthorchg2 = 6;
    } elsif ($finda2 == 5) { #OUTROS
        if ($atomthorlist == 23) {
            $atomthorchg1 = 26;
        } elsif ($atomthorlist == 24) {
            $atomthorchg1 = 30;
        } elsif ($atomthorlist == 27) {
            $atomthorchg1 = 15;
        } elsif ($atomthorlist == 33) {
            $atomthorchg1 = 14;
        } elsif ($atomthorlist == 34) {
            $atomthorchg1 = 11;
        } elsif ($atomthorlist == 35) {
            $atomthorchg1 = 17;
        } elsif ($atomthorlist == 36) {
            $atomthorchg1 = 20;
        } elsif ($atomthorlist == 37) {
            $atomthorchg1 = 12;
        }
    }
}
last;
}
}
}
}
}
}

$Soutprint = ' ' x (6 - length($atompos)) . $atompos .
' ' x (6 - length($atomthorlist)) . $atomthorlist .
' ' x (7 - length($atomtype1)) . $atomtype1 .
' ' x (15 - length($atommass)) . $atommass .
' ' x (5 - length($atomthorchg1)) . $atomthorchg1 .
' ' x (5 - length($atomthorchg2)) . $atomthorchg2 . "\n";
print $hTOPOUTfile $Soutprint;

}
}
}
}
}
}

### IMPRIMIR LISTA DE CONSTRAINTS ----- OK
sub PRINTCONSTR {

    $Soutprint = "\n" .
    '
    SHAKE - DATA SOLVENT

```

NUMBER OF CONSTRAINTS = 3

```

1    2    1.0000
1    3    1.0000
2    3    1.6330' . "\n";

```

```

    print $hTOPOUTfile $outprint;
}

```

CRIA ZMATRIZ COM BABEL E CARREGA DADOS ----- OK
sub LOADZMT {

TRANSFORMA ARQUIVO GRO PARA XYZ

```

$linha = 0;
while (<$hgrofile>) {
    $linhaf = $_;
    chomp ($linhaf);
    $linha++;

    if ($linha == 2) {
        $linhaf_aux = $linhaf;
        print $hxyzfile "$linhaf_aux\n\n";
    }

    if ($linha >= 3 && $linha <= $linhaf_aux+2) {
        @linedata = split(/\s+/, $linhaf);
        $X = $linedata[4] * 10;
        $Y = $linedata[5] * 10;
        $Z = $linedata[6] * 10;
        print $hxyzfile "H $X $Y $Z\n";
    }
}

```

TRANSFORMA O ARQUIVO XYZ PARA ZMATRIX UTILIZANDO O OPENBABEL
system (".babel -ixyz \$xyzfile -ofh \$zmtfile");

ABRE ARQUIVO DE ESTRUTURA ZMATRIZ

open (\$zmtfile, \$zmtfile) or die "Problemas com o arquivo de estrutura ZMATRIZ...";

LIMPA ARQUIVO ZMATRIZ !!! OPENBABEL TEM SAIDA DEFEITUOSA...

```

$linha = 0;
while (<$zmtfile>) {
    $linhaf = $_;
    chomp ($linhaf);
    $linha++;

    $letrafinal_aux = "";

    if ($linha == 2) { $linhaf_aux = $linhaf; }
    if ($linha >= 3 && $linha <= $linhaf_aux+2) {
        @zmt_data = split(/\s+/, $linhaf);
        for ($sver=0; $sver <= $#zmt_data; $sver++) {
            $pontos = 0;
            for ($sverpos=0; $sverpos <= length($zmt_data[$sver]); $sverpos++) {
                $letra = substr($zmt_data[$sver], $sverpos, 1);
            }
        }
    }
}

```



```

$chainletter = 'A';
$chainnumber = 1;

# DADOS OBTIDOS NO ARQUIVO ZMATRIZ
@zmatriz_aux = split(/\s+/, $zmatriz[(($ver/8)+1)];
for $i (0..7-$zmatriz_aux) {
    push (@zmatriz_aux, '0');
}
if (($ver/8)+1 == 1) { $zmatriz_aux[2] = 0; }

$zmatriz_aux[3] = sprintf("%.3f", $zmatriz_aux[3]);
$zmatriz_aux[5] = sprintf("%.2f", $zmatriz_aux[5]);
$zmatriz_aux[7] = sprintf("%.2f", $zmatriz_aux[7]);

# PEGA DADOS DA TABELA DE LIGACOES - QUEM LIGA COM QUEM...
@atombondtable2 = split(/\s+/, $atombondtable1[(($ver/8)+1)];
#POE ZEROS NOS LUGARES VAZIOS DA MATRIZ
for ($roda=6-(6-#$atombondtable2); $roda<=6; $roda++) { push (@atombondtable2, "0") };

# IMPRIMIR LINHA DO ATOMO
$coutprint = ' ' x (4 - length($atomtype1)) . $atomtype1 .
    ' ' x (8 - length($zmatriz_aux[3])) . $zmatriz_aux[3] . ' 0' .
    ' ' x (11 - length($zmatriz_aux[5])) . $zmatriz_aux[5] . ' 0' .
    ' ' x (10 - length($zmatriz_aux[7])) . $zmatriz_aux[7] . ' 0' .

    ' ' x (5 - length($zmatriz_aux[2])) . $zmatriz_aux[2] .
    ' ' x (5 - length($zmatriz_aux[4])) . $zmatriz_aux[4] .
    ' ' x (5 - length($zmatriz_aux[6])) . $zmatriz_aux[6] . ' 0' .

    ' ' x (7 - length($atomcharg)) . $atomcharg .

    ' ' x (5 - length($atombondtable2[1])) . $atombondtable2[1] .
    ' ' x (4 - length($atombondtable2[2])) . $atombondtable2[2] .
    ' ' x (4 - length($atombondtable2[3])) . $atombondtable2[3] .
    ' ' x (4 - length($atombondtable2[4])) . $atombondtable2[4] .
    ' ' x (4 - length($atombondtable2[5])) . $atombondtable2[5] .
    ' ' x (4 - length($atombondtable2[6])) . $atombondtable2[6] .

    ' ' x (4 - length($atompdbtype)) . $atompdbtype .
    ' ' x (5 - length($respdbtype)) . $respdbtype .

    ' ' x (2 - length($chainletter)) . $chainletter .
    ' ' x (4 - length($resposnum)) . $resposnum .
    ' ' x (5 - length($chainnumber)) . $chainnumber . "\n";
print $hINOUTfile $coutprint;
}

### IMPRIMIR RODAPE
print $hINOUTfile
'##### 0.000 0 0.00 0 0.00 0 0 0 0 0.000 0 0 0 0 0 0

Simulated Annealing Parameters
&GSAPAR Temper0a =000.0, ParqVa=0.0, ParqAa=0.0, ParqTa=0.0, Nseriea=0, NStopMaxa=0, NRANa
=-00000, NslideGSAa=0 &END'; }

##### FIM DAS SUBROTINAS #####

```

Nome do programa: New Convolution;

Linguagem de programação: BASIC;

Descrição: Faz um tipo de “convolução” sobre a interação entre duas moléculas.

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```

print "Carregando...";
open "../mndplot_HS.out" for input as #1           'Arquivo da Interação completa
open "../mdnat40ns_HS.out" for input as #2       'Arquivo da Interação de referência
open "distif2.out" for output as #3
print "OK"

%nresiduos = 212                                '212 é o número de resíduos da proteína
redim refer!(%nresiduos)
redim compa!(%nresiduos)

print "Carregando Ref. ...";
  do while not eof(2)
    input #2, vazio%, residuo%, access!
    refer!(residuo%) = access!
  loop
print "OK"

  do while not eof(1)
    input #1, tempo%, residuo%, access!
    compa!(residuo%) = access!

    if residuo% = %nresiduos then
      for res% = 1 to 212
        'FORMULA DO NEW CONVOLUTION DETALHADA NOS MÉTODOS
        distref! = (abs(refer!(res%) - compa!(res%))) * (refer!(res%))
        print #3, tempo% & ", " & res% & ", " & distref!
      next res%
    end if

  loop

print "Fim..."

```

Nome do programa: Yet Another Map Plotter, YAMP;

Linguagem de programação: BASIC;

Descrição: Faz a plotagem de mapas de distância (convolução, MOD_GYR etc)

Observação: criado por José Eduardo Kroll, 2007; Programa altamente customizável.

Código fonte:

```
screen 19
```

```
line (9, 501)-(10+601, 500-(213)), 14, B
```

```
open command$ for input as #1                                'Abre arquivo de mapa: tempo / resíduo / valor
  do while not eof(1)
    input #1, tempo%, residuo%, access!
    cor% = 16 + (access!*4)                                    'Reajustar as cores conforme a escala dos valores...
    if cor% > cormax% then cormax% = cor%
    pset (10+tempo%/100, 500-(residuo%)), cor%
  loop
close #1
```

```
line (9,51-16)-(41,50-(cormax%+1)), 14, b
```

```
for barra% = 16 to cormax%
```

```
  line (10,50-barra%)-(40,50-barra%), barra%
```

```
next barra
```

```
input pausa$
```


Nome do programa: RMS3D_aux;

Linguagem de programação: Perl;

Descrição: Auxilia na criação de dados RMSD entre trajetórias não compatíveis entre si.

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```
#!/usr/bin/perl

print "ok...";

$dt = 100;          #PASSO DE TEMPO
$tt = 60000;       #TEMPO TOTAL DO PH5
$tf = $tt/$dt;     #TOTAL DE ARQUIVOS A SEREM CRIADOS

open (outfile, ">rms3dplot_ph5xpaphep.out");

for $tempo(0..$tt/$dt) {
    $realtempo = $dt * $tempo;
    system ("trjconv -f ~/ACS/md60nsfinalpapa5.xtc -s ~/ACS/md60ns_papa5.tpr -o REFTEST.gro -e
    $realtempo -b $realtempo < grupos1.txt");
    system ("g_rms -f ~/ACS/md100nsfinalpapahep.xtc -s REFTEST.gro -n ~/ACS/md100ns_papahep.ndx -o
    rms3dteste.xvg -b 0 -e 100000 -dt $dt < grupos2.txt");

    open (mndfile, "+<rms3dteste.xvg");

    $residuo = 0;
    while (<mndfile>) {
        $linhaf = $_;
        chomp ($linhaf);
        $residuo++;

        if ($residuo >= 14) {
            $linhaf =~ s/^(s*)//s; ##### UTILIZAR ESSE COMANDO NO GRO2THOR
            @linha = split (/s+/, $linhaf);

            $var1 = sprintf("%.0f", $linha[0]);

            print outfile "$realtempo, $var1, $linha[1]\n";
        }
    }
    close mndfile;

    system ('rm -f "#"*"#");
    system ('rm -f nohup.out');
}

close;
```

Nome do programa: RMS3D_sum;

Linguagem de programação: BASIC;

Descrição: Faz a soma dos valores RMSD (3D) em função de uma das trajetórias e os plots.

Observação: criado por José Eduardo Kroll, 2008;

Código fonte:

```

screen 21

line (9, 501)-(10+601, 500-(213)), 14, B

pset (700, 1000-(400)), 12

dim score!(1000)

open command$ for input as #1

    do while not eof(1)
        input #1, tempo!, residuo!, access!
        cor% = 16 + ((access!^2)*2)
        if cor% > cormax% then cormax% = cor%
        'pset (1+tempo!/100, 500-(residuo!/100)), cor%
        pset (1+(residuo!/100), 700-(tempo!/100)), cor%

        score!(residuo!/100) = score(residuo!/100) + (access! ^ 2)

    loop
close #1

line (9,51-16)-(41,50-(cormax%+1)), 14, b

open "score.out" for output as #2
    for scoreprint% = 0 to 999
        print #2, scoreprint%, score!(scoreprint%)
    next scoreprint%

close #2

print "Finalizado..."

input pausa$

```