UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE BIOQUÍMICA

"FISIOLOGIA MOLECULAR DIGESTIVA DE Musca domestica (DIPTERA)"

MARCELO HENRIQUE PETERES PADILHA

TESE DE DOUTORADO

DR. WALTER RIBEIRO TERRA
ORIENTADOR

SÃO PAULO – SP
2009

Data de Depósito na SPG: ___/___/2009

MARCELO HENRIQUE PETERES PADILHA

FISIOLOGIA MOLECULAR DIGESTIVA DE *Musca domestica* (DIPTERA)

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutor em Ciências (Bioquímica)

Orientador: Prof. Dr. Walter Ribeiro Terra

São Paulo –SP

2009

Marcelo Henrique Peteres Padilha

Fisiologia molecular digestiva de *Musca domestica* (Diptera)

Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
Título de Doutor em Ciências (Bioquímica)

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____
Instituição: _____
Assinatura:_____

Prof. Dr._____
Instituição: _____
Assinatura: _____

Prof. Dr. _____
Instituição: _____
Assinatura:_____

Prof. Dr._____
Instituição: _____
Assinatura:_____

Prof. Dr._____
Instituição: _____
Assinatura:_____

Dedico...

Ao Seu Padilha e Dona Érica, meus pais...

AGRADECIMENTOS

Agradeço, todas essas pessoas pelos momentos de distração e alegria....

Ao Prof. Dr. Walter Ribeiro Terra, por toda a orientação e amizade no desenvolver desse trabalho.

A professora Dra. Clélia Ferreira pela discussão e ajuda na realização desse trabalho.

Ao Dr. Alberto de Freitas Ribeiro e aos técnicos Ms. Waldir Caldeira e Márcio Valentin Cruz pela contribuição no estudo de imunocitoquímica.

Aos amigos do Laboratório de Bioquímica de Insetos: Dr. Fernando Ariel Genta, Dra. Maria Cícera Pereira da Silva, Dra. Fabiane Chaves Cançado, Walciane Silva, Fábio Kendi Tamaki, Daniela Beton, Ivan Bragatto, Natália Moreira, André Coppe Pimentel, Bruno Silvestre Lira e as técnicas Luiza Nakabayashi, Maria Ivanilde Marcelino e Cristiane Cardoso.

Aos amigos da Secretaria de Pós Graduação: Cibele, Milton e Emiliano.

A minha namorada Betânia Carvalho Bernardo.

A minha irmã Márcia Padilha e seu marido José Rafael Dias.

A minha tia Waltraude Peters.

Ao meu primo Milton Padilha e sua esposa Eliane Francisca Padilha.

A minha prima Silvia Padilha.

Ao Departamento de Bioquímica do Instituto de Química da USP pela excelente infraestrutura para realização deste trabalho.

A FAPESP e CNPq pelo apoio financeiro.

RESUMO

Padilha, M.H.P. Fisiologia molecular digestiva de *Musca domestica* (Diptera), 2009, 118 páginas. Tese – Programa de Pós-graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo-SP, Brasil.

A mosca domestica (*Musca domestica*) é um dos insetos largamente distribuído e conhecido pelo homem. A larva de *M. domestica* possui no conteúdo luminal do ventrículo anterior e médio uma atividade proteolítica com pH ótimo entre 3,0 – 3,5 e propriedades cinéticas similares a catepsina-D. Três cDNAs codificantes para preprocatepsina-D (ppCAD1, ppCAD2 e ppCAD3) foram clonados a partir de uma biblioteca de cDNA ventricular de larvas de *M. domestica*. As sequências possuem o peptídeo sinal, o propeptídeo e a enzima madura contendo os resíduos catalíticos e todos os resíduos de ligação ao substrato conservados, achados em uma catepsina-D lisossomal bovina. Um cladograma de sequências de aminoácidos de catepsinas-D de insetos e vertebrados depositados no GENBANK formou um grande grupo dividido em duas ramificações monofiléticas: Uma com sequências de vertebrados e a outra com sequências de catepsinas-D lisossomais de insetos incluindo a ppCAD1. A sequência do pepsinogênio humana, ppCAD2, ppCAD3 e uma sequência de *D. melanogaster* são excluídas desse grande grupo indicando uma função não lisossomal para essas sequências. ppCAD3 deve corresponder a uma catepsina-D digestiva encontrada no conteúdo luminal em larvas do inseto devido: (1) Análise por RT-PCR indicam que os transcritos codificantes para a CAD3 são expressos no ventrículo anterior e porção proximal do ventrículo médio. (2) pCAD3 recombinante após autoativação sob condições ácidas possui um pH ótimo entre 2,5 – 3,0 que é próximo ao pH luminal do ventrículo médio onde essa enzima atua. (3) Imunoblots das proteínas de diferentes tecidos e marcadas com o anticorpo preparado contra a pCAD3 foi positiva apenas nos tecidos e conteúdo do ventrículo anterior e médio. (4) CAD3 é localizada pela técnica de imuno-ouro no interior de vesículas de secreção e próxima a microvilosidades nas células do ventrículo anterior e médio. Esses dados suportam a idéia que ao se adaptar a um hábito detritivo, a CAD digestiva da mosca (e de outros Diptera Cyclorrhapha)

resultou do mesmo gene ancestral da catepsina-D lisossomal intracelular, da mesma forma que se acredita que ocorreu com a pepsina em vertebrados.

Uma limitada quantidade de informações de sequências de DNA e aspectos moleculares de *M. domestica* é disponível até o presente momento. Nós então propusemos gerar sequências de ESTs a partir de uma bibilioteca de cDNA ventricular da larva desse inseto. Um total de 826 ESTs randomicamente selecionados presentes no ventrículos de larvas de *M. domestica* foram seqüenciados e analisados com programas de bioinformática e separado em 323 *clusters*. As sequências foram manualmente anotadas e separadas em 3 categorias: (S) produtos provavelmente secretados, (H) produtos de metabolismo em geral (*housekeeping*) e (U) produtos sem função conhecida. Cento e sessenta clusters (423 ESTs) codificavam para proteínas secretadas tais como: lisozimas, lipases, tripsinas, quimotripsinas, dipeptidades, carboxipeptidase e α-amilases. Cento e trinta e dois clusters (190 ESTs) codificavam para sequências de metabolismo em energético, síntese de proteínas, transdução de sinal e outras funções celulares. Noventa e cinco clusters (231 ESTs) codificaram para proteínas sem similaridades com proteínas conhecidas no banco de dados. Estudos de expressão, localização e imunocitoquímica de sequências alvos deverão no futuro nos fornecer um melhor entendimento da fisiologia digestiva da larva de *Musca domestica* em detalhes moleculares.

Uma enzima lipolítica (LipMD) foi identificada no item anterior. A sequência de aminoácidos obtidas é homóloga a lipases e contém os três bem conservados resíduos de aminoácidos que compõe a tríade catalítica (Ser^{183}, His^{273} e Asp^{208}) para esse grupo de enzimas. O fragmento de cDNA codificante para a LipMD foi clonado em vetor pAE (Ramos et al., 2004) e expresso em *E. coli* produzindo uma enzima com massa molecular de 37,3 kDa. A LipMD recombinante foi purificada e é capaz de hidrolizar tributirina e um grande número de substratos (pNP-acetato a pNP-esterato) e possui um pH ótimo próximo de 7,5. Analise por RT-PCR mostrou que os transcritos codificantes para a LipMD são expressos apenas no ventrículo anterior. Western-blots após SDS-PAGE das proteínas de vários tecidos e marcados com o anticorpo produzido contra a LipMD revelou a ocorrência da enzima principalmente no conteúdo luminal do ventrículo anterior. A LipMD é localizada pela técnica de imuno-ouro no interior de vesículas de secreção próximas

a microvilosidades no interior das células do ventrículo anterior. Esta enzima pode atuar como uma enzima lipolítica digestiva secretada nessa região do ventrículo de larvas de *M. domestica*.

Palavras-chave: Inseto, *Musca domestica*, catepsina-D digestiva, ESTs, lipase/esterase.

ABSTRACT

Padilha, M.H.P. Molecular physiology of *Musca domestica* (Diptera), 2009, 118 p. PhD Thesis – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo – SP, Brazil.

The house fly, *Musca domestica* is one the best known and most widely distributed insects known to humans. *M. domestica* larvae display in anterior and middle midgut contents, a proteolytic activity with pH optimum of 3.0-3.5 and kinetical properties like cathepsin-D. Three cDNAs coding for preprocatepsin D-like proteinases (ppCAD1, ppCAD2, ppCAD3) were cloned from a *M. domestica* midgut cDNA library. The sequences encoding the signal peptide, propeptide and mature enzyme having all conserved catalytic and substrate binding residues found in bovine lysosomal cathepsin-D. A cladogram of sequences of insect and vertebrate cathepsin-D-like proteinases deposited on GENBANK form a large grouping divided into two monophyletic branches: one with vertebrate and the other with insect lysosomal sequences including ppCAD1. Human pepsinogen, ppCAD2, ppCAD3, and a sequence from *Drosophila melanogaster* are excluded indicating a non-lysosomal function for them. CAD3 should correspond to the digestive CAD found in enzyme assays because: (1) The mRNA for CAD3 is expressed (RT-PCR) only in the anterior and proximal middle midgut. (2) Recombinant pCAD3, after auto activation has a pH optimum of 2.5-3.0 that is close to the luminal pH of *M. domestica* midgut. (3) Immunoblots of proteins from different tissues and stained with antiserum prepared against recombinant pCAD3 were positive only for the anterior and middle midgut tissue and contents. (4) CAD3 is localized with immunogold labeling inside secretory vesicles and around microvilli in anterior and middle midgut cells. The data support the view that on adapting to a detritivorous habit *M. domestica* digestive CAD (and of other Diptera Cyclorrhapha) resulted from the same archetypical gene as the intracellular cathepsin-D, paralleling what happened with vertebrates.

A limited amount of data regarding DNA sequences and molecular aspects of *Musca domestica* species is avaliable. We proposed to generate ESTs sequences from a cDNA library constructed from larval midguts. A total of 826 randomly

selected midgut – derived cDNAs were sequenced and assembled based on their similarities into 323 clusters. The sequences were classified into three categories: (S) probably secretory products, (H) housekeeping products and (U) products with unknown cell localization and function. One hundred and sixty clusters (423 ESTs) encode putative secreted proteins such as lysozymes, lipases, trypsins, chymotrypsins, dipeptidases, carboxypeptidases A and α-amylases. One hundred and thirty two clusters (190 ESTs) encode housekeeping sequences associated with energy metabolism, protein syntesis, signal transduction and other cellular functions. Ninety five clusters (213 ESTs) encode proteins with no similarity with known proteins. Expression and high-throughput bioassay screening of target sequences must provide us with a better understanding of the digestive physiology of *Musca domestica* midgut larvae, in molecular detail.

A lipolytic enzyme (LipMD) was identified from expressed sequence tags (EST) constructed from midgut larvae cDNA library. The deduced amino acid sequence is homologous to lipases and contained the three well-conserved amino acid residues, Ser^{183}, His^{273}, and Asp^{208}, which form the catalytic triad of the enzyme. The cDNA fragment encoding for LipMD was cloned into a pAE vector (Ramos et al., 2004) and expressed in *E. coli* producing an enzyme with a molecular mass of 37.3 kDa. The recombinant LipMD was purified and was able to hydrolyse tributyrin and a broad range of substrates, from C_2 to C_{18} p-nitrophenyl-esters and displayed an optimal pH of approximately 7.5. RT-PCR analysis in tissue homogenates (anterior, middle and posterior midguts, hemolymph, fat body and Malpighian tubules) showed that LipMD mRNA transcripts were expressed only in anterior midgut. Western-blots after SDS–PAGE of proteins from different tissues and stained with anti-LipMD serum revealed that the enzyme occurs mainly in the anterior midgut lumen. LipMD is localized with immunogold labeling inside secretory vesicles and around microvilli in anterior midgut cells. LipMD is a candidate to be the digestive lipolytic enzyme found in that midgut region.

Keywords: Insect, *Musca domestica,* digestive cathepsin-D, ESTs, lipase/esterase.

ABREVIATURAS

BLAST: *Basic Local Aligment Search*

BSA: albumina de soro bovino

ppCADs: preprocatepsinas-D *de Musca domestica*

DEPC: dietilpirocarbonato

dNTP: desoxirribonucleotídeo 5' trifosfato

Abs: absorbância ou turbidância (este último para quantificação de número de bactérias)

DTT: ditiotreitol

EDTA: etileno diamin tetra acetato

E-64: 4-guanidino butano

EGTA: ácido etilenoglicol-bis-(beta-aminoetiléter)N,N'-tetracético

FPLC: *fast performance liquid chromatography*

Hepes: ácido N-2-hidroxietilpiperazina-N'-2-etanosulfônico

IPTG: isopropyl-β-D-galactopiranosídeo

kb: quilobases

kDa: quilodalton

LB: meio de cultura de Luria-Bertani

MM: massa molecular

ORF: fase aberta de leitura (*Open Reading Frame*)

PCR: reação em cadeia de polimerização

pb: pares de nucleotídeos

pI: ponto isoelétrico

PVDF: polivinilideno diflúor

pNP: para-nitrofenolato (4-nitrofenolato)

RNAse: ribonuclease A

rpm: rotações por minuto

RT-PCR: PCR precedido de reação com transcriptase reversa

SDS: dodecil sulfato de sódio

SDS-PAGE: eletroforese em gel de poliacrilamida contendo SDS

TCA: ácido tricloroacético

TAE: tampão tris-ácido acético-EDTA

Tris: (hidroximetil) aminometano

Tributirina: tributiril-glicerol

Trioleína: trioleil-glicerol

X-gal: 5-bromo-4-cloro-3-indoyl-β-D-galactopiranosídeo

SUMÁRIO

1. INTRODUÇÃO 1
1.1. Considerações iniciais 1
1.2. Fisiologia molecular da digestão nos Diptera superiores 2
1.2.1. Diversidade bacteriana no trato intestinal de larvas de *Musca domestica* .. 4
1.3. Tamponamento intestinal na larva de *M. domestica* 5
1.4. Aspártico proteinase e seus clãs 5
1.4.1. Clã AA das aspártico proteinases 6
1.4.2. Família A1 7
1.4.3. Catepsina-D (E.C. 3.4.23.5) 8
1.4.4. Especificidade das catepsinas-D 8
1.4.5. Aspártico proteinase digestiva de *M. domestica* 8
1.5. Sequênciamento randômico de ESTs 9
1.5.1. Geração dos ESTs 10
1.5.2. Erros associados com a geração de ESTs 11
1.5.3. Processamento dos ESTs 11
1.5.4. Pesquisa por similaridade em banco de dados 14
1.6. Enzimas lipolíticas - lipases e carboxiesterases 14
1.7. OBJETIVOS 17
2. PROCEDIMENTOS EXPERIMENTAIS 18
2.1. Animais 18
2.2. Extração do RNA 18
2.3. Biblioteca de cDNA ventricular de *Musca domestica* clonada em fago λ zap II 18
2.4. Obtenção de cDNAs que codificam aspártico proteinases em *Musca domestica* através da técnica de PCR por homologia 19
2.4.1. Desenho do primer degenerado para aspártico proteinase 19
2.4.2. PCR por homologia 19
2.5. Eletroforese em gel de agarose 21
2.6. Purificação dos produtos de PCR a partir de géis de agarose 21
2.7. Ligação dos fragmentos de cDNA em plasmídeo vetor 21
2.8. Transformação de bactérias competentes 22

2.9. Purificação de plasmídeos a partir de cultura de bactérias..... 22
2.10. Reação de sequênciamento de DNA 22
2.11. Clonagem em vetor pAE das sequências codificantes para as proteínas recombinantes pCAD1, pCAD3 e LipMD..... 23
2.12. Expressão da pCAD1, pCAD3 e LipMD recombinantes em vetor pAE..... 25
2.13. Lise das células induzidas..... 25
2.14. Purificacão das proteínas recombinantes (pCAD1, pCAD3 e LipMD) através de uma cromatografia de afinidade 25
2.15. Auto-ativação da pCAD3 recombinante 26
2.16. Determinação da atividade enzimática da CAD3 recombinante 26
2.17. Análise da degradação da hemoglobina pela CAD3 recombinante 27
2.18. Eletroforese em placas de gel de poliacrilamida com SDS (SDS-PAGE) 27
2.19. RT-PCR semi-quantitativo..... 28
2.20. Construção da biblioteca de cDNA de ventriculo total de larvas de *Musca domestica* clonada em plasmídeo 28
2.21. Amplificação e sequênciamento randômico dos clones de cDNA..... 29
2.22. Análises de bioinformática 30
2.23. Análise filogenética 30
2.24. Medida da atividade lipolítica usando pNP-ésteres..... 31
2.25. Efeito da temperatura e pH sobre a estabilidade e atividade da LipMD recombinante..... 31
2.26. Estabilidade da LipMD frente a solventes orgânicos..... 32
2.27. Produção dos anticorpos policlonais 32
2.28. Western blot 33
2.29. Transferência da CAD3 (madura) de *Musca domestica* para membranas de PVDF e obtenção da sequência N-terminal 34
2.30. Experimentos de Dicroismo Circular (CD)..... 34
2.31. Imunocitolocalização por microscopia eletrônica de transmissão 35
2.32. Meios de cultura..... 36
2.32.1. LB Líquido 36
2.32.2. Placas LB – Ágar 36

2.32.3. Determinação da atividade lipolítica da LipMD recombinante sobre placas de ágar-tributirina e ágar-trioleína 36
3. RESULTADOS 37
4.1. Estudo das aspártico proteinases intestinais de *Musca domestica* 37
4.1.1. Amplificação dos fragmentos de cDNAs que codificam aspártico proteinases utilizando a técnica de PCR por homologia 37
4.1.2. Identificação, clonagem e sequênciamento dos cDNAs que codificam aspártico proteinases ventriculares em *Musca domestica* 39
4.1.2.1. ppCAD1 41
4.1.2.2. ppCAD2 41
4.1.2.3. ppCAD3 44
4.1.3. Clonagem, sequênciamento e expressão da pCAD1 recombinante 47
4.1.4. Clonagem, sequênciamento e expressão da pCAD3 recombinante 50
4.1.4.1. Auto-ativação da pCAD3 recombinante..... 54
4.1.4.2. Análise da atividade proteolítica da CAD3 recombinante 56
4.1.4.3. Determinação do pH ótimo da CAD3..... 56
4.1.4.4. Análise da degradação da hemoglobina bovina pela CAD3 recombinante 58
4.1.5. Análise filogenética, expressão e localização das pCAD1, 2 e 3 58
4.1.6. Imunocitolocalização da pCAD3 em *Musca domestica*..... 62
4.2. Sequênciamento randômico de ESTs ventriculares de larvas de *Musca domestica* 62
4.2.1. Sequênciamento randômico dos ESTs ventriculares de larvas de *Musca domestica* 64
4.2.2. Organização das informações do transcriptoma de *M. domestica*..... 66
4.2.3. Descrição dos clusters pertencentes a categoria S 69
4.2.3.1. Descrição dos clusters pertencentes a categoria S sem função conhecida 69
4.2.4. Descrição dos clusters pertencentes a categoria housekeeping (H)..... 69
4.2.5. Descrição dos clusters pertencentes a categoria U 70
4.3. Identificação da ORF lipolítica ventricular de *Musca domestica* 78
4.3.1. Clonagem e expressão da LipMD recombinante 81
4.3.2. Efeito do pH e da temperatura sobre a atividade da LipMD..... 86

4.3.3. Especificidade pelo substrato da LipMD recombinante 89
4.3.4. Análise da expressão e localização da LipMD 90
4.3.5. Imunocitolocalização da LipMD em *Musca domestica* 92
4.3.6. Distribuição de enzimas lipolíticas em larvas de *Musca domestica* 93
5. DISCUSSÃO 94
5.1. Estudo das aspártico proteinases intestinais de *Musca domestica* 94
5.2. Sequênciamento randômico de ESTs versus modelo digestivo de larvas de *Musca domestica* 97
5.2.1. Lisozimas 98
5.2.2. Digestão de carboidratos 98
5.2.3. Digestão de proteínas 99
5.2.4. Peritrofinas 100
5.2.5. Digestão de lipídeos 100
5.3. Estudo de uma enzima lipolítica no tubo digestivo de larvas de *Musca domestica* 100
6. REFERÊNCIAS 104
7. SÚMULA CURRICULAR 119

LISTA DE FIGURAS

Figura 1... 3
Figura 2.. 6
Figura 3. 12
Figura 4. 20
Figura 5.. 35
Figura 6.. 38
Figura 7. 38
Figura 8. 42
Figura 9. 43
Figura 10.. 45
Figura 11. 46
Figura 12. 47
Figura 13.. 48
Figura 14.. 49
Figura 15. 50
Figura 16. 51
Figura 17.. 52
Figura 18. 53
Figura 19.. 54
Figura 20.. 55
Figura 21. 56
Figura 22.. 57
Figura 23. 57
Figura 24.. 58
Figura 25. 60
Figura 26.. 61
Figura 27. 61
Figura 28.. 63
Figura 29.. 65
Figura 30.. 66
Figura 31. 67

Figura 32. 68
Figura 33.. 79
Figura 34.. 81
Figura 35.. 82
Figura 36.. 83
Figura 37.. 84
Figura 38.. 85
Figura 39. 85
Figura 40.. 86
Figura 41.. 87
Figura 42.. 87
Figura 43.. 88
Figura 44.. 88
Figura 45.. 89
Figura 46.. 90
Figura 47.. 91
Figura 48.. 91
Figura 49.. 92

LISTA DE TABELAS

Tabela 1.....4
Tabela 2.....24
Tabela 3.....40
Tabela 4.....68
Tabela 5.....71
Tabela 6.....74
Tabela 7.....77
Tabela 8.....93

FISIOLOGIA MOLECULAR DIGESTIVA DE Musca domestica

1. *INTRODUÇÃO*

1.1. *Considerações iniciais*

Os insetos são, sem dúvida, o grupo predominante da Terra. Eles representam 75% das espécies animais e 57% das espécies vivas, se considerarmos também as plantas e os protozoários (Grimaldi e Engel, 2005). Estima-se que os insetos consomem 10% da biomassa vegetal (Schoonhoven *et al.,* 2005) anualmente, além de serem transmissores de doenças em plantas e animais, inclusive o homem. O controle dos insetos é, pois, necessário e se faz principalmente com o uso de inseticidas químicos. Como são crescentes os problemas ambientais causados por esses inseticidas, novas formas de controle de insetos tem sido pesquisadas.

Após a conscientização de que o tubo digestivo é uma enorme interface relativamente pouco protegida entre o inseto e o meio ambiente, uma compreensão da função intestinal foi considerada essencial para o desenvolvimento de novos métodos de controle agindo através do intestino, tais como o uso de plantas transgênicas para controlar insetos fitófagos (Law *et al.,* 1992). Aproximadamente nessa época, o nosso laboratório escolheu a estratégia de estudar detalhadamente insetos-modelo dispostos em pontos relevantes da árvore filogenética, a fim de produzir generalizações que pudessem valer para insetos de interesse aplicado que carecessem de estudos básicos. Esses estudos, que foram revistos várias vezes à medida que avançavam (Terra, 1988; 1990; Terra e Ferreira, 1994; 2005), avaliavam a organização do processo digestivo através do estudo da distribuição de enzimas nas diferentes partes do intestino e dos fluxos de fluidos intestinais, associados a análises morfológicas macro e ultraestruturais dos tecidos intestinais e do isolamento e caracterização de enzimas representativas. Esses estudos mostraram que o processo digestivo nos insetos organiza-se em padrões filogeneticamente determinados e permitiram o levantamento de hipóteses sobre o papel da membrana peritrófica (estrutura anatômica, quitino-proteica, que separa o alimento da superfície intestinal no interior do intestino dos insetos) (Terra, 2001;

Bolognesi *et al.,* 2008) e das membranas perimicrovilares (membrana lipoprotéica que recobre as microvilosidades intestinais como dedo de luva e que terminam em fundo cego) dos Hemiptera (percevejos, cigarras e pulgões) (Terra e Ferreira, 2005).

O desenvolvimento do setor passa pela ampliação do conhecimento dos eventos moleculares subjacentes aos fenômenos fisiológicos intestinais, o que requer conhecimento das enzimas digestivas em detalhe estrutural, assim como de proteínas associadas a outras funções intestinais tais como: transportadores, receptores, bombas, proteínas relacionadas aos mecanismos de secreção, etc. O objetivo geral do laboratório é procurar as causas estruturais e cinéticas das peculiaridades das enzimas digestivas que possibilitam os insetos contornar as defesas das plantas e adaptar-se a dietas peculiares.

1.2. *Fisiologia molecular da digestão nos Diptera superiores*

A digestão nos insetos ocorre no intestino médio (ventrículo) de forma compartimentada (Figura 1). A digestão inicial dos polímeros (amido, proteína, etc...) ocorre no interior da membrana peritrófica (estrutura anatômica quitino-proteíca que reveste o alimento no interior do ventrículo). Os oligômeros resultantes difundem-se para o espaço luminal exterior à membrana peritrófica onde são atacados por outras enzimas (carboxipeptidase, aminopeptidases) e, finalmente, na digestão final os dímeros resultantes são hidrolisados por enzimas imobilizadas na superfície do epitélio do intestino médio. Essa imobilização ocorre devido ao fato das enzimas serem integrantes de membrana em alguns insetos ou por estarem firmemente associados ao glicocálix celular em outros. Ocorre uma circulação endo-ectoperitrófica (isto é entre os espaços interiores e exteriores à membrana peritrófica) de enzimas de digestão inicial, o que leva a uma grande economia dessas enzimas, pois impede a sua excreção junto com o alimento não digerido. Essa circulação ocorre graças à secreção de água na região posterior do intestino médio (ventrículo) e sua absorção na região anterior. Dependendo do grupo filogenético, ocorrem variações nesse esquema (Terra, 1990; Terra e Ferreira, 1994, 2005).

Os Diptera superiores que incluem a mosca doméstica apresentam peculiaridades digestivas que aparentemente resultam de adaptações para digerir uma dieta que consiste principalmente de bactérias (ver item 1.2.1.). Os eventos digestivos a seguir estão resumidos seguindo os trabalhos mais relevantes (Espinoza-Fuentes e Terra, 1987; Espinoza-Fuentes *et al.,* 1987; Terra *et al.,* 1988). As larvas ingerem alimento rico em bactéria. No ventrículo anterior ocorre uma diminuição no conteúdo de amido do bolo alimentar graças a ação de uma α-amilase. O bolo passa agora para o ventrículo médio onde as bactérias são mortas pela ação combinada de baixo pH, uma lisozima especial e uma aspártico proteinase similar a uma catepsina-D. Finalmente, o material liberado das bactérias é digerido no ventrículo posterior, como ocorre no ventrículo inteiro da maioria dos insetos de outros grupos taxonômicos. Fluxos em contracorrente ocorrem no ventrículo posterior ocasionados pela secreção de fluido na parte distal do ventrículo posterior e sua absorção no ventrículo médio. O ventrículo médio tem células especializadas para tamponar os conteúdos luminais na zona ácida, além de funcionar na absorção de fluido.

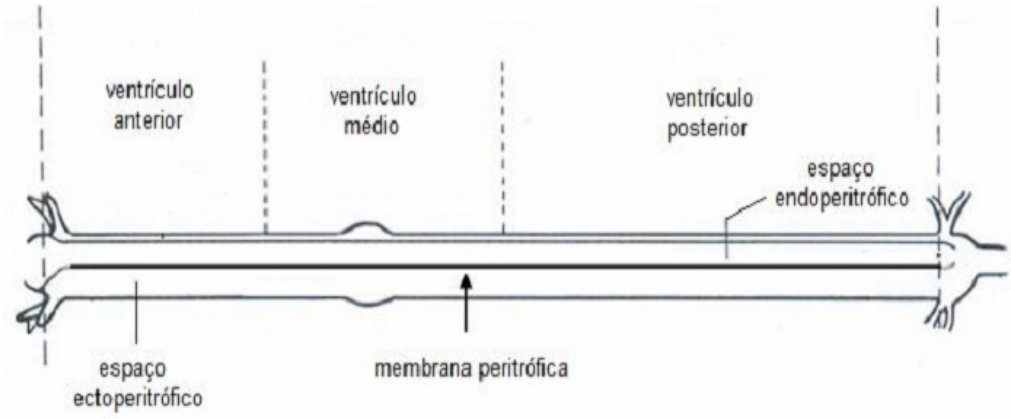

Figura 1. Representação esquemática do tubo digestivo de larvas de Musca domestica. No ventrículo anterior o amido é parcialmente digerido no interior da membrana peritrófica (espaço endoperitrófico) e bactérias são mortas no ventrículo médio pela ação conjunta de um pH ácido, uma lisozima especial e uma aspártico proteinase similar a cathepsina-D. No ventrículo posterior os oligômeros formado juntamente com as enzimas digestivas migram para o espaço ectoperitrófico e são

translocados para a região proximal do ventrículo posterior onde a digestão terminal e a absorção dos alimentos ocorre na superfície da célula. As enzimas digestivas retornam para o espaço endoperitrófico renovando-se o ciclo digestivo.

1.2.1. ***Diversidade bacteriana no trato intestinal de larvas de*** *Musca domestica*

Zurek *et al.,* 2000 trabalhando com larvas de *Musca domestica* (terceiro estágio larval) demonstrou a diversidade bacteriana encontrada no trato intestinal nas larvas desse inseto. Onze espécies de bactérias de oito gêneros incluindo *Staphylococcus, Streptococcus, Lactococcus, Morganella, Providencia, Yersinia, Sphingomonas* e *Bacillus* foram identificados quando as larvas foram tratadas em uma dieta composta de serragem de madeira molhada misturada com fezes de peru (dieta 1). Já em uma dieta composta por silagem de milho (dieta 2) foi verificada a presença de catorze espécies de bactérias de nove gêneros incluindo *Serratia, Providencia*, *Microbacterium, Xanthobacter, Sphingobacterium, Corynebacterium, Clavibacter, Ochrobacter* e *Gordona* (Tabela 1).

Tabela 1. Identificação das bactérias isoladas do trato intestinal de larvas de M. domestica tratadas em dois tipos de dietas. Dieta 1: serragem de madeira molhada misturada com fezes de peru. Dieta 2: silagem de milho (Zurek et al., 2000).

Dieta 1	***Identificação***	***Dieta 2***	***Identificação***
MdT1	*Morganella morganii*	MdS1	*Serratia marcescens*
MdT2	*Staphylococcus epidermidis*	MdS2	*Sphingobacterium spiritivorum*
MdT3	*NI*	MdS3	*Xanthobacter flavus*
MdT4	*Sphingomonas capsulata*	MdS4	*Microbacterium liquafaciens*
MdT5	*Bacillus sp.*	MdS5	*Providencia stuartii*
MdT6	*Bacillus coagulans*	MdS6	*Gordona amarae*
MdT7	*Providencia rettgeri*	MdS7	*Corynebacterium aquaticum*
MdT8	*Yersinia pseudotuberculosis*	MdS8	*Providencia rettgeri*
MdT9	*Lactococcus garviae*	MdS9	*Microbacterium esteraromaticum*
MdT10	*Staphylococcus lentus*	MdS10	*Microbacterium lactium*
MdT11	*Streptococcus sanguis*	MdS11	*Ochrobacterium anthropi*
		MdS12	*Clavibacter michiganese*
		MdS13	*Corynebacterium seminale*
		MdS14	*Microbacterium barkeri*

** NI: bactéria não identificada*

1.3. ***Tamponamento intestinal na larva de*** *M. domestica*

O pH do intestino médio (ventrículo) da larva de *M. domestica* é 6,1 na região anterior, 3,1 na média e 6,8 na posterior (Espinoza-Fuentes e Terra, 1987). O processo de acidificação da região média e a neutralização do conteúdo da região posterior foi estudado por Regel e Terra (1995) (Figura 2). Para isso eles determinaram os valores de pH luminais e as concentrações de amônia, cloreto e fosfato na presença e ausência de ouabaína e vanadato no ventrículo das larvas de *M. domestica* e, juntamente com dados já publicados e referidos acima, eles propuseram os seguintes mecanismos de tamponamento. A anidrase carbônica nas células oxínticas do ventrículo médio produz ácido carbônico que se dissocia em próton e bicarbonato. Bicarbonato é transportado para a hemolinfa, enquanto que o próton é ativamente translocado para o lúmen ventricular por uma ATPase tipo P similar a H^+/K^+-ATPase gástrica de mamíferos (Pedersen e Carafoli, 1987). Íons cloretos seguem os prótons translocados por transportador provavelmente específico. No ventrículo posterior, amônia difunde-se para o lúmen tornando-se protonada e neutralizando o conteúdo. O íon amônio é trocado por Na^+ por uma Na^+/K^+-ATPase microvilar (revisão: Emery *et al.,* 1998). Dentro das células, o íon amônio forma amônia que difunde-se para o lúmen enquanto o próton vai para a hemolinfa, neutralizando o bicarbonato originado a partir do ventrículo médio.

1.4. Aspártico proteinase e seus clãs

As aspártico proteinases são ativas em pH ácido, hidrolisam ligações peptídicas internas em proteínas e atacam alguns substratos sintéticos cromofóricos (Dunn *et al.,* 1986) ou substratos fluorescentes com apagamento intramolecular (Pimenta *et al.,* 2001).

Aspártico proteinases são assim chamadas pelo fato do resíduo catalítico Asp ser ligado e ativado por moléculas de água durante a catálise. Na grande maioria das enzimas desse grupo, dois resíduos de Asp atuam como resíduos catalíticos sendo que em alguns casos outros aminoácidos podem substituir o segundo Asp.

Uma notável característica das aspártico proteinases é que todas as enzimas desse grupo são endopeptidases, embora não há razões obvias pela qual esse tipo de mecanismo catalítico não deva atuar como exopeptidases (Barrett *et al.,* 2004).

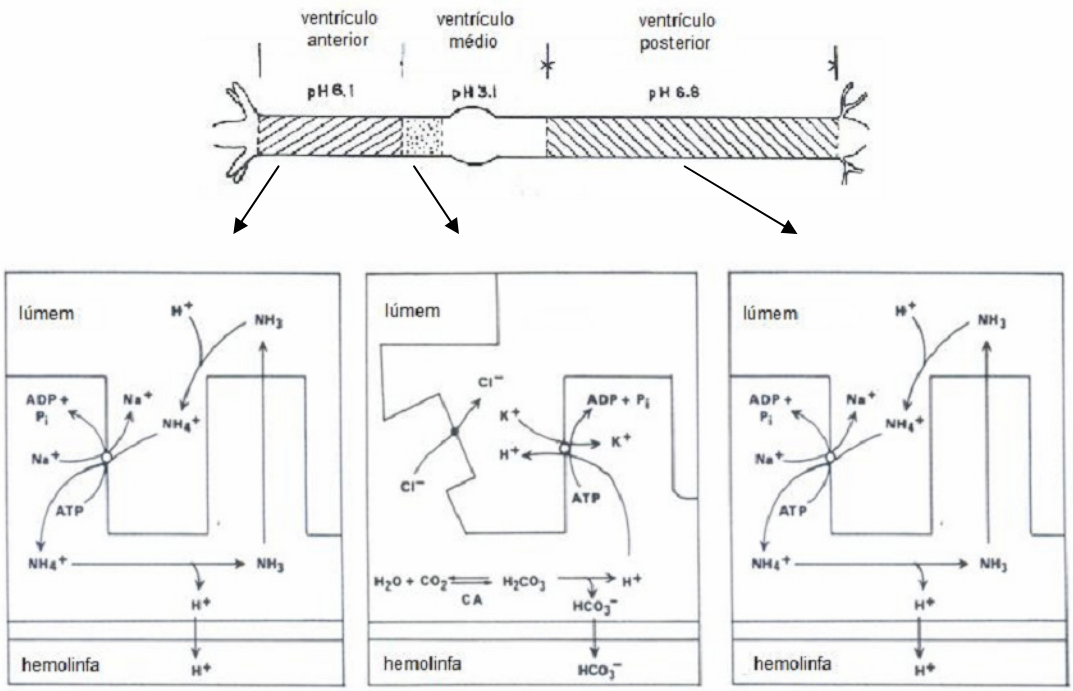

Figura 2. Representação esquemática da movimentação dos íons supostos para ser responsável para a manutenção do pH ventricular em larvas de Musca domestica. Anidrase carbônica associadas a células oxínticas no ventrículo médio produz ácido carbônico que se dissocia em próton e íon bicarbonato. O íon bicarbonato é transportado para a hemolinfa e os prótons são transportados para o ventrículo médio acidificando essa região. Íons cloretos seguem o mesmo movimento formando ácido clorídrico. NH_3 difundem-se pelas células no ventrículo anterior e posterior e dentro do conteúdo luminal tornam-se protonadas aumentando o pH naquelas regiões. NH_4^+ é então trocado por Na^+ por uma Na^+/K^+ ATPase. Dentro da célula o íon NH_4^+ dissocia-se em NH_3 onde se difunde novamente para o conteúdo luminal e os prótons são transferidos para a hemolinfa renovando-se o ciclo (Regel e Terra, 1995).

1.4.1. Clã AA das aspártico proteinases

O clã AA das aspártico proteinases tem por característica reunir famílias de enzimas que possuem dois Asp como resíduos catalíticos (em alguns casos o último resíduo pode ser substituído por His) e que apresentem estrutura tridimensional do tipo duplo β barril (ou dímeros de β barril). Esse clã contém a grande família A1, onde as enzimas mais estudadas são as pepsinas e catepsinas-D. As estruturas 3D para um grande número de enzimas dessa família (A1) mostram que elas são bilobares com o sítio ativo localizado entres os lobos (ou melhor, entre os dois domínios). Essa característica é comum para muitas peptidases sendo que na família A1 um lobo evoluiu a partir do outro durante o evento de duplicação gênica. Cada lobo possui 1 Asp catalítico sendo que os dois lobos são homólogos entre si. A similaridade de aminoácidos dos dois lobos é muito baixa, embora a estrutura tridimensional dos lobos seja muito similar.

Essa descoberta aumenta a questão de como o produto do gene original pode ser funcional como peptidase se a enzima original possui apenas 1 lobo com apenas 1 Asp catalítico. Uma possível resposta foi fornecida quando a estrutura 3D da retrospepsina da família A2 foi resolvida, mostrando que ela era constituída por apenas um lobo, com 1 Asp catalítico, muito similar a estrutura de um dos lobo da pepsina. A similaridade estrutural das endopeptidases da família A1 e A2 suportam a hipótese de que elas são originadas de um ancestral comum e por isso elas são agrupadas em um mesmo clã (Barrett *et al.,* 2004).

1.4.2. Família A1

As proteinases da família A1 são encontradas apenas em eucariotos, e como todas aspártico proteinases elas são endopeptidases. A grande maioria é ativa sobre pH ácido, sendo poucas que apresentarem atividade em pH neutro. Exemplos de enzimas de mamíferos incluem enzimas digestivas como a pepsina e quimosina e as intracelulares catepsinas-D e E.

A homologia dos dois lobos da molécula das enzimas da família A1 é refletida pelo fato dos dois Asp catalíticos ocorrerem em pequenas regiões muito conservadas. Essa região é comumente resumida em Asp – Thr – Gly (motivo “DTG”), sendo que a descrição completa é Xhh – Xhh – Asp – Xpp – Gly – Xpp,

onde Xhh são resíduos hidrofóbicos e Xpp é o aminoácido Ser ou Thr (Barrett *et al.,* 2004).

1.4.3. Catepsina-D (E.C. 3.4.23.5)

Proteinases tendo pH ótimo em meio ácido foram encontrados em extratos de tecidos há quase 90 anos. Estas enzimas foram nomeadas como catepsinas (Gr. *kathepsein*, "para digerir") (Barrett 1977). Devido ao grande número de proteinases ácidas encontradas, o termo catepsina-D foi aplicado para distinguir esta enzima de outras proteinases ácidas celulares, sendo que a partir de então todas as catepsinas foram diferenciadas a partir de uma letra (Press *et al.,* 1960). Esta enzima pertence à família A1 do clã AA das aspártico proteinases sendo identificada pela IUBMB por E.C. 3.4.23.5 (Barrett 1977). Artigos demonstrando purificações de catepsinas-D mostram que existem mais de 12 diferentes formas dessa enzima onde se diferem em tamanhos, números de subunidades e ponto isoelétrico (Sapolsky e Woessner, 1972). Keilvá (1970) e Cunningham e Tang (1976) usando modificação covalente sítio dirigida demonstraram que a catepsina-D possui um sítio ativo similar a pepsina (família A1 do clã AA) caracterizando-a como aspártico proteinase.

Estudos sobre o isolamento do cDNA que codifica a catepsina-D (Faust *et al.,* 1985), do estudo de sua biosíntese (Horst e Hasilik, 1991), ativação (Richo e Conner, 1994), especificidade pelo substrato (Scarborough *et al.,* 1993) e na determinação de sua estrutura (Baldwin *et al.,* 1993; Metcalf e Fusek, 1993) fazem da catepsina-D uma das proteínas lisossomais mais estudadas.

1.4.4. Especificidade das catepsinas-D

Como outras aspártico proteinases, a catepsina-D pode acomodar até 7 resíduos de aminoácidos em seu sítio ativo. A especificidade por seus substratos é muito similar a pepsina na qual há preferência em hidrolisar ligações peptídicas entre dois resíduos hidrofóbicos (Barrett, 1977).

1.4.5. Aspártico proteinase digestiva de M. domestica

O primeiro registro de aspártico proteinases em insetos foi feito por Greenberg e Paretsky (1955), que encontraram uma forte atividade hidrolítica a pH 2,5-3,0 em homogeneizados de corpos de *M. domestica*. Lemos e Terra (1991) mostraram que a enzima ocorre no intestino é similar à catepsina-D. Enzimas similares foram descritas em percevejos e besouros (Terra e Ferreira, 2005), mas todas foram só parcialmente purificadas e caracterizadas. Esses estudos devem ser estendidos de forma que a origem, especificidade e estrutura das enzimas digestivas similares a catepsina-D de insetos possam ser esclarecidas.

1.5. Sequênciamento randômico de ESTs

Os insetos incluem a maior variedade de espécies em todo reino animal e possuem uma vasta diversidade de genes desconhecidos que podem ser melhor explorados utilizando-se técnicas de biologia molecular. Atualmente, as técnicas de sequênciamento randômico de ESTs (transcriptoma) contribuíram significativamente para compreensão da base genética de vários insetos, permitindo identificar genes cujas proteínas são fortes candidatos a desenvolvimento de drogas visando o controle desses insetos (Calvo, *et al.,* 2004; Alarcon-Chaidez, *et al.,* 2007; Srinivasan, *et al.,* 2004).

Geração de ESTs (transcriptoma) em grande ou pequena escala é importante por várias razões. Primeiro eles representam a oportunidade de identificar e comparar diferentes genes expressos coletivamente entre vários estágios de desenvolvimento, incluindo ovos e tecidos (como tubos digestivos e glândulas salivares). As sequências obtidas por essas fontes podem ser comparados com seus ortólogos em *Drosophila melanogaster*, *Aedes aegypti* e *Anopheles gambiae* podendo também indicar indícios de adaptações evolutivas que foram necessárias para se atingir o atual processo fisiológico (Kirschener, 2005).

Etiquetas de sequências expressas (*Expressed Sequence Tags* - EST) fornecem uma evidência direta dos transcritos. Estes geralmente possuem entre 200-600 pb, sendo sequênciados randomicamente a partir das extremidades 5' ou 3' dos clones obtidos em uma biblioteca de cDNA. Em 1991, ESTs foram usados como fonte primária para a descoberta de genes humanos (Adams *et al.,* 1991). Atualmente é verificado um crescimento exponencial na geração de ESTs sendo estes depositados em bancos de dados públicos para uma grande diversidade de

organismos. Hoje os ESTs permitem a descoberta de genes, anotação complementares de genomas, caracterização de polimorfismo em um único nucleotídeo (SNP) e facilita também a análise proteômica (Jongeneel, 2000; Dong *et al.,* 2005 e Rudd, 2003).

1.5.1. Geração dos ESTs

Em uma célula viva, a todo momento podem ser encontradas moléculas de RNAs. Um subconjunto dessas moléculas é resultado do processo de transcrição de genes expressos (em atividade) naquele momento. As sequências obtidas a partir do mRNAs das células são chamadas de etiquetas de sequências expressas (*Expressed Sequence Tag*) ou simplesmente EST. Como o mRNA não pode ser clonado diretamente eles são transformados em cDNA pela ação de uma transcriptase reversa sendo que os cDNA resultantes são clonados em vetores e então inseridos em um organismo vetor, geralmente bactérias (*E. coli*) formando então as bibliotecas de cDNAs que representam os genes expressos em determinado tipos de células, tecidos ou organismos inteiros.

A bactéria é então induzida a se reproduzir gerando milhões de bactérias idênticas e consequentemente milhões de cópias da sequência alvo. As bactérias são separadas uma a uma em placas contendo meios de cultura, contendo antibiótico de seleção que depois são usadas na leitura propriamente dita do material genético. Esses cDNAs são sequênciados randomicamente a partir das extremidades 5' ou 3'. Os ESTs resultantes podem ser redundantes (repetidos) e suas sequências podem estar completas ou não. Bonaldo *et al.,* 1996 fornece uma descrição detalhada sobre construção e normalização de bibliotecas de cDNAs visando reduzir a redundâncias desses ESTs.

Os ESTs podem ser sequênciados por diversos métodos como o de Sanger (Sanger *et al.,* 1977) e o *Dye terminator sequencing*. Nesses casos, a idéia é criar, através de processos químicos, sequências de nucleotídeos que correspondam a todos os prefixos da sequência original. Esses prefixos recebem um marcador no último nucleotídeo (terminador). Como prefixos diferentes possuem massas diferentes, quando colocados em um gel e atraídos por um campo elétrico (eletroforese em gel), cada um deles se moverá a uma velocidade inversamente

proporcional a sua massa molecular. Assim é possível descobrir a ordem original dos nucleotídeos.

Porém os ESTs carregam, além de informação genética, muito ruído. Um outro agravante são os erros de leituras causados pelo próprio método de sequenciamento. Essas técnicas produzem como saída, um conjunto de sinais contínuos resultantes da leitura de luz em diferentes frequências, refletida pelos marcadores moleculares. Esses sinais são posteriormente interpretados a fim de gerar uma sequência de letras correspondentes aos nucleotídeos. Tal interpretação de sinais é conhecida como *base calling*. Entretanto nem sempre é possível interpretar estes sinais sem ambiguidade, dando origem a bases marcadas como desconhecidas (geralmente usa-se a letra N). Mesmo quando o algoritmo de *base calling* chega a alguma conclusão a respeito das bases presentes no sinal, essas bases estão sempre condicionadas a uma informação de qualidade, que pode ser alta ou baixa que em muitas vezes, leva ao descarte do EST.

1.5.2. Erros associados com a geração de ESTs

Uma típica sequência de ESTs (Figura 3) é apenas uma pequena cópia do mRNA e apresentam erros significativos especialmente no início e no final das sequências. A qualidade é geralmente significativa no meio da sequência (Figura 3A). Sequências do vetor no ínicio ou no final das sequência são extirpados durante o processamento dos dados dos ESTs. Os ESTs que foram sequênciados apenas uma vez, a leitura pode ser suscetível a erros. Geralmente a qualidade das sequência iniciais de um ESTs são ruins (em torno de 20% da sequência ou 50-100 pb) sendo que a qualidade aumenta gradualmente, indo para o meio da sequência e então novamente a qualidade diminui para o final (Aaronson *et al.,* 1996) ou seja a qualidade da sequência é significantemente melhor no meio da sequência (região altamente informativa (Figura 3B).

1.5.3. Processamento dos ESTs

A análise de ESTs é um processo feito em várias etapas, envolvendo diferentes técnicas em cada uma. Após realizados todos os procedimentos *in vitro*

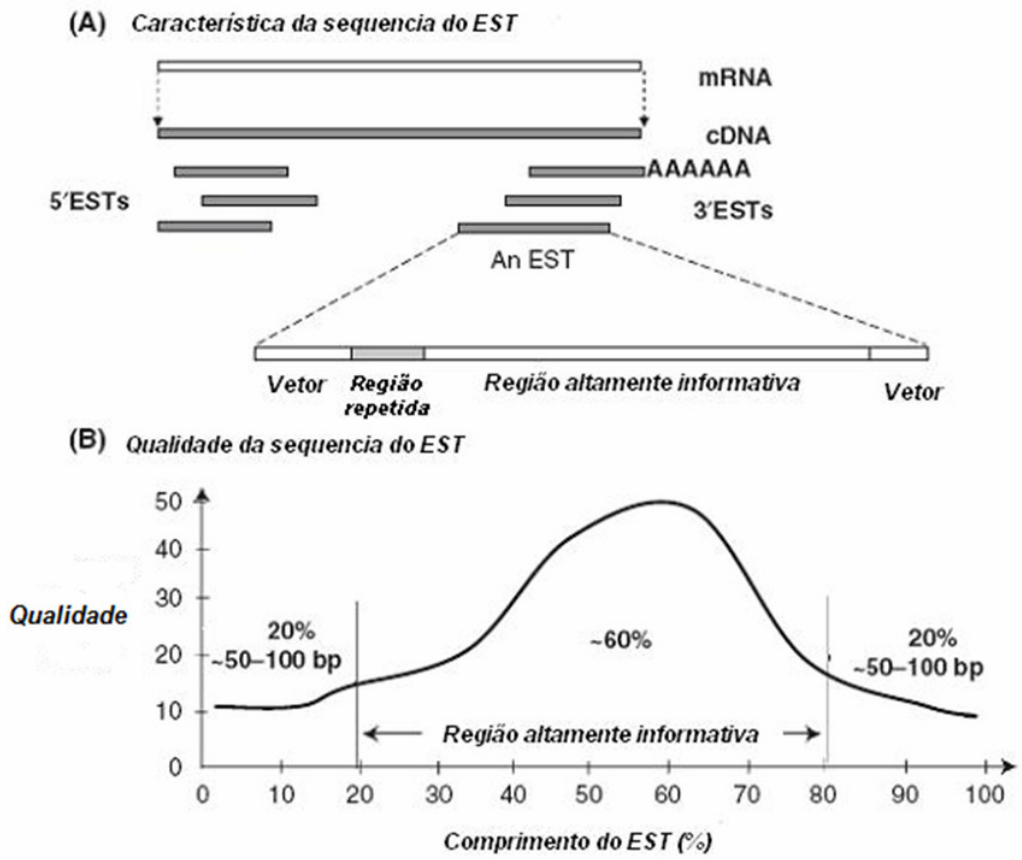

Figura 3. Característica da sequência dos ESTs. A. Geralmente nas extremidades das sequência do EST estão contaminadas com a sequências do vetor seguido de regiões de baixa complexidade. B. Nota inferida pelo Phred (Ewing et al., 1998) são plotadas em função do comprimento da sequência do EST hipotético como mostrado em A.

(amplificação, clonagem, sequênciamento), os dados de saída devem ser analisados, dando início aos procedimentos *in silico*.

A primeira tarefa na análise de ESTs é a extração das bases a partir dos sinais de saída do sequenciador. São quatro sinais: um para cada frequência lida pelo sequenciador. Como cada frequência de luz corresponde a uma base diferente e as bases passam pelo leitor uma a uma, em tese basta sobrepor os sinais e identificar os picos para determinar a ordem das bases. Na prática, porém, podem ocorrer muitos erros como sobreposição de picos e ruídos de fundo, que interferem e

podem causar ambiguidades na leitura. Por esse motivo, algoritmos mais sofisticados são empregados nesta tarefa.

A ferramenta mais popular de *base calling* é o Phred (Ewing *et al.,* 1998), (Ewing e Green, 1998) (www.phrap.com/phred/). Uma nota inferida pelo Phred fornece um valor referente à qualidade do nucleotídeo onde o maior valor corresponde uma sequência confiável. Phred examina o pico emitido por cada base e infere uma nota para a sua qualidade que varia de 4 – 40 e pode ser determinada pela equação: $q=-10.\log_{10}(P)$ onde q é a nota inferida pelo progarama e P é a probabilidade de erro. Uma nota de 20 representa uma chance de 1/100 se ser incorreta ou 99% de chance de estar correta. Se a nota for 30 denota 1/1000 a chance de ser incorreta ou 99,9% de probabilidade da base estar correta.

O propósito da *clusterização* dos ESTs é fazer a sobreposição dos ESTs a partir de um mesmo transcrito de um único gene dentro de um único *cluster*. O caminho simples para a *clusterização* dos ESTs é pela medida da similaridade do pareamento entre as sequências. Então essas distâncias são convertidas em valores binários verificando se existe uma similaridade significativa ou não, sendo que o pareamento das sequências pode ser aceito ou rejeitado a partir dos *clusters* que estão sendo montados (Burke *et al.,* 1999).

Diferentes etapas para a *clusterização* dos ESTs são descritas em detalhes por Ptitsyn e Ride (2005). No geral existem duas técnicas para a *clusterização* dos ESTs: estringente e *loose*. O método de *clusterização* estringente é conservativo usando um único passo de agrupamento dos ESTs, resultando em *clusters* com relativa precisão, mas gera pequenas sequências consenso cobrindo uma pequena região dos genes expressos. Em contraste o método da clusterização *loose* é "liberal" e repete baixas qualidades dos ESTs alinhando várias vezes para gerar menor precisão, porém fornecem sequências consensos maiores.

Consequentemente, ocorre uma maior cobertura das informações das sequências dos genes expressos, mas corre o risco de agrupar sequências diferentes no processo de *clusterização*. Phrap (Ewing e Green, 1998) e CAP3 (Huang e Madan, 1999) estão entre os mais usados programas para *clusterização* (agrupamento) das sequências.

1.5.4. Pesquisa por similaridade em banco de dados

Uma vez que as sequências consenso são obtidas a partir do conjunto de ESTs, possíveis funções podem ser inferidas por similaridade de sequências depositadas em banco de dados através de ferramentas disponíveis de forma livre na internet. Diferentes ferramentas de BLAST (Alschul *et al.,* 1997) disponíveis no NCBI (http://www.ncbi.nlm.nih.gov/) servem como ferramentas universais para a pesquisa por similaridades em banco de dados. BLASTN pode ser usado para pesquisar ESTs contra sequências de nucleotídeos depositadas no banco de dados e BLASTX para compará-la em um banco de dados de proteínas. Este último traduz a sequência consenso de ESTs (*query*) em proteínas nos 6 *frames* de leitura permitindo comparar essas 6 possibilidades contra um banco de dados de proteínas.

1.6. Enzimas lipolíticas - lipases e carboxiesterases

De forma resumida lipases são definidas como enzimas que preferem catalisar a hidrólise de triacilgliceróis com cadeia carbônica longa (>10 átomos de carbonos, usando trioleína (óleo de oliva) como substrato padrão) e atuam na interface água-lipídeo e possuem papel essencial na transferência de lipídeos entre organismos tais como durante a digestão e mobilização das reservas energéticas armazenadas.

Já as carboxiesterases são definidas como enzimas que preferem catalisar a hidrólise de triacilgliceróis solúveis em água, que contenham cadeias carbônicas curtas (<10 átomos de carbonos), sendo tributirina seu substrato padrão.

Lipídeos são essenciais para os organismos vivos. Eles são as principais moléculas de armazenamento de energia, possuem papel estrutural em membranas e também estão envolvidos em eventos de sinalização celular. A participação de lipídeos em todas essas funções requerem enzimas lipolíticas durante esse processo. Enzimas lipolíticas incluem duas grandes classes de enzimas: carboxiesterases (EC 3.1.1.1, carboxiéster hidrolases) e lipases (EC 3.1.1.3, triacilglicerol hidrolases). A identificação dessas enzimas envolvidas nesses processos e seus mecanismos regulatórios sobre sua atividade ou acesso a substratos continuam pouco conhecidos em insetos devido à falta de técnicas

sensíveis, complexidade da digestão relatada para solubilização dos lipídeos e a hidrólise de lipídeos de membranas que pode causar a lise celular sendo que os ácidos graxos produzidos nesta digestão pode causar efeito antibiótico no organismo, dificultando o trabalho nessa área.

Essas enzimas lipolíticas presentes no trato intestinal possuem um papel crítico na habilidade do animal utilizar lipídeos em sua dieta. Enzimas lipolíticas são também requeridas para mobilizar lipídeos a partir das reservas energéticas das células (Gibbons *et al.,* 2000). Os triacilgliceróis não atravessam membranas por transporte passivo e não são solúveis em água ou camadas de fosfolipídeos. Essas moléculas não podem atravessar as membranas microvilares intestinais o que requer a hidrólise dos triacilgliceróis por lipases/carboxiesterases antes da absorção ou reesterificação dos ácidos graxos pelo intestino (Desnuelle *et al.,* 1986).

Em *Drosophila melanogaster* a atividade lipolítica é maior nas células do intestino que em outros tecidos (Smith *et al.,* 1994). A maior região de absorção de lipídeos em insetos são o ceco gástrico e a porção anterior do ventrículo (Dow, 1987). Três sequências codificantes para enzimas lipolíticas (FlyBase 3.1 – Berkeleg Drosophila Genome Project – www.fruitfly.org/annot/) CG6271, CG6277, CG6283) são expressas no ventrículo anterior em *D. melanogaster*, expressão essa reprimida em uma dieta rica em açúcar (Zinke *et al.,* 2002). Essas proteínas possivelmente estão envolvidas na lipólise primária de lipídeos na dieta do animal.

Comparações entre caboxiesterases e lipases revelam notáveis similaridades de sequências e radicais diferenças entre especificidades por diferentes substratos ou funções fisiológicas. Pequenas diferenças estruturais do sítio ativo parecem ter impacto sobre as especificidades por seus substratos (Dolinsky *et al.,* 2004).

Comparação da estrutura tridimensionais das carboxiesterases e lipases revelam características conhecidas como dobramento das α/β hidrolases (Ollis *et al.,* 1992; Holmquist, 2000). Basicamente, esse dobramento consiste em uma folha β central cercadas por números variados de α-hélices que acomodam a tríade catalítica composta por serina, histidina e ácido aspártico. O resíduo de serina catalítico encontra-se em um motivo conservado “Gly – X – Ser - X – Gly”, onde X é qualquer aminoácido exceto prolina (Holmquist, 2000).

Lipases pancreáticas de mamíferos operam juntamente com colipases (pequenas proteínas de aproximadamente 12 kDa) e sais biliares (que atuam como

emulsificantes) permitindo uma melhor conformação do sítio ativo dessas enzimas durante a catálise sendo que os produtos da hidrólise são absorvidos pelas células do intestino (van Tilbeurge *et al.,* 1999).

Uma óbvia diferença entre lipases de insetos e de mamíferos é que lipases de insetos faltam o domínio PLAT (sítio de ligação de colipases) (Winkler *et al.,* 1990) além do fato de insetos não terem colipases. Se existe outra proteína com função similar a colipases em insetos, esta ainda está para ser descoberta (van Tilbeurgh *et al.,* 1999). Outra diferença é que em insetos também não existem sais biliares. Porém é sabido que existem estratégias alternativas, tais como, a formação de micelas com a utilização de fosfolipídeos em pHs alcalino facilitando a hidrólise dos lipídeos presentes na dieta (Canavoso *et al.,* 2001).

1.7. OBJETIVOS

Os objetivos desta tese, distribuíram-se em 3 vertentes principais:

1. Estudo molecular das aspártico proteinases (similares a catepsina-D) que atuam no processo digestivo durante a proteólise ácida no ventrículo médio de larvas de *Musca domestica*. Esse objetivo consistiu na clonagem de cDNAs codificantes para catepsinas-D, expressão em sistema heterólogo, purificação, caracterização, distribuição ao longo do tubo digestivo e estudos de imunocitoquímica.

2. Sequênciamento randômico de ESTs ventriculares (transcriptoma) em larvas de *Musca domestica*. As sequências de cDNA que despertaram interesses serão no futuro estudadas e caracterizadas separadamente.

3. Estudo molecular de uma enzima lipolítica cujo cDNA foi obtido no objetivo anterior. Esse objetivo consistiu na clonagem desse cDNA, expressão em sistema heterólogo, purificação, caracterização, distribuição e estudos de imunocitoquímica.

2. PROCEDIMENTOS EXPERIMENTAIS

2.1. Animais

Larvas de *Musca domestica* foram criadas em uma mistura fermentada de ração de porco e palha de arroz (1:2, v/v) a uma temperatura média de 25ºC sob luz constante. As larvas usadas estavam no terceiro estágio larval e se alimentavam ativamente.

2.2. Extração do RNA

Larvas de *Musca domestica* foram mantidas previamente em gel de amido 10% por 2 horas visando eliminar às bactérias presentes no trato intestinal. Essas larvas foram lavadas em água destilada, secadas e imobilizadas em gelo e dissecadas utilizando uma solução de NaCl 150mM, pinças e placas de vidro previamente fornadas ou tratadas com DEPC (0,1% (v/v)). O tecido dissecado era imediatamente mantido sobre gelo seco e posteriormente armazenado em –80ºC.

O RNA total foi extraído do tecido alvo usando Trizol (Molecular Research Center) (Chomczynski and Sacchi, 1987) e ressuspenso em 100 μL de água tratada com DEPC (0,1% (v/v)). O RNA mensageiro foi extraído a partir da amostra anterior por cromatografia em coluna de oligo (dT) utilizando o kit *Oligotex Direct mRNA Micro Kit* (Qiagen), seguindo rigorosamente o protocolo do fabricante. Uma alíquota de cada amostra foi quantificada por medida de absorbância em espectrofotômetro (260 nm e 280 nm) e checado sua integridade por eletroforese em gel de agarose 1% em tampão TAE.

2.3. Biblioteca de cDNA ventricular de *Musca domestica* clonada em fago λ zap II

Para construir a biblioteca de cDNA o RNA total foi extraído do epitélio do tubo digestivo da larva de *M. domestica* e enviado a Stratagene. O RNA foi dividido em duas amostras iguais e usado para a síntese de cDNA, a qual foi realizada utilizando-se o iniciador poli T e um iniciador randômico. Esta estratégia foi usada

para aumentar a probabilidade de ambas extremidades 5' e 3' do cDNA estarem representadas. Finalmente, os dois conjuntos de cDNA foram misturados (1:1) e inseridos não direcionalmente no vetor λ zap II. O titulo da biblioteca é 1,1 $.10^6$ pfu/mL.

2.4. Obtenção de cDNAs que codificam aspártico proteinases em Musca domestica através da técnica de PCR por homologia

2.4.1. Desenho do primer degenerado para aspártico proteinase

A partir do alinhamento de 16 aspártico proteinases (Tang e Wong, 1987) é verificado duas regiões muito conservadas no interior do sítio ativo, em torno dos 2 ácidos aspárticos catalíticos (Figura 4). Nesse alinhamento verificamos a sequência de aminoácido consenso D_{33}TGSSNLW onde o D_{33} é o primeiro resíduo catalítico usando o sistema de numeração de uma catepsina-D humana (Baldwin *et al.,* 1993). Baseado nessa sequência (D_{33}TGSSNLW) nós desenhamos o *primer* degenerado CADdeg – 5' GGH TCS TCS AAY YTS TGG - 3' onde H: A, C ou T, S: G ou C e Y: C ou T.

2.4.2. PCR por homologia

O *primer* degenerado CADdeg e o *primer* T3 (Tabela 2) específico do braço do fago foram usados em um PCR tendo como molde uma biblioteca de cDNA de *Musca dometica* clonada em fago λ zap II. O PCR foi realizado da seguinte forma: 94ºC por 5 minutos (para lise dos fagos), seguidos por 30 ciclos de 94ºC por 30 segundos (desnaturação), 57ºC por 30 segundos (pareamento) e 68ºC por 2 minutos (elongamento).

A partir do sequênciamento do produto do PCR encontrado, foram desenhados novos *primers* a partir de cada extremidade 3' das sequências obtidas, que combinados com o *primer* T7 (específico do outro braço do fago) em um outro PCR permitiu a amplificação da sequência completa das aspártico proteinase de *Musca domestica.*

Resíduos do sítio ativo	***D_{33}***	***D_{231}***
Enzimas		
Pepsinogênio de porco	FTVIFDTGSSNLWVPSV	QAIVDTGTSLLT
Pepsinogênio humana	FTVVFDTGSSNLWVPSV	QAIVDTGTSLLT
Pepsinogênio de macaco	FTVIFDTGSSNLWVPSV	QAIVDTGTSLLT
Pepsinogênio de galinha	FSVIFDTGSSNLWVPSI	QAIVDTGTSLLV
Progastricina humana	FLVLFDTGSSNLWVPSV	QAIVDTGTSLLT
Progastricina de macaco	FLVLFDTGSSNLWVPSV	QAIVDTGTSLLT
Procimosina de boi	FTVLFDTGSSDFWVPSI	QAILDTGTSKLV
Renina humana	FKVVFDTGSSNVWVPSS	LALVDTGASYIS
Renina fígado	FKVIFDTGSANLWVPST	AVVVDTGSSFIS
Renina Glândula submaxilar	FKVIFDTGSANLWVPST	EVVVDTGSSFIS
Catepsina D de porco	FTVVFDTGSSNLWVPSI	EAIVDTGTSLIV
Catepsina D humana	FTVVFDTGSSNLWVPSI	EAIVDTGTSLMV
Rizopuspepsinogênio	FNLDFDTGSSDLWIAST	DGILDTGTTLLI
Penicilopepsinogênio	LNLNFDTGSADLWVFST	SGIADTGTTLLL
Endotiopepsina	LNLDFDTGSSDLWVFSS	DGIADTGTTLLY
Proprotease A	FKVIFDTGSSNLWVPSN	GAAIDTGTSLIT
Consenso	***DTGSSNLW***	***DTGTSLL***
CADdeg	***5' G G H T C S T C S A A Y Y T S T G G 3'***	
Tradução	***G S S N L W***	

Figura 4. Desenho do primer degenerado CADdeg. A partir do alinhamento de 16 aspártico proteinases (Tang e Wong, 1987) são mostradas as regiões conservadas em torno dos resíduos catalíticos do sítio ativo (D_{33} e D_{231} baseado no sistema de

numeração para a catepsina D humana) (Baldwin et al., 1993). A sequência de aminoácidos consenso em torno do primeiro resíduo catalítico foi utilizada para desenhar o primer degenerado CADdeg.

2.5. Eletroforese em gel de agarose

Os géis foram preparados em uma concentração final 1% (m/v) de agarose em tampão TAE 1X. As amostras de DNA foram misturadas com tampão de amostra (azul de bromofenol 0,25%, xileno cianol 0,25%, glicerol 30% em água) na proporção de 1:10, respectivamente. O azul de bromofenol, devido a sua coloração, auxilia a estabelecer visualmente o fim da eletroforese.

Após a eletroforese, o gel foi tratado com brometo de etídeo (0,1 μg/mL) para visualização das amostras de DNA, pois este complexado com o brometo de etídeo emite luz laranja na presença da luz ultravioleta (312 nm).

2.6. Purificação dos produtos de PCR a partir de géis de agarose

Os produtos amplificados em reação de PCR foram recuperados diretamente dos géis de agarose usando o protocolo do produto QIAEX II (Quiagen), o qual baseia-se na dissolução da agarose em um tampão de alta força iônica e ligação do DNA em uma resina de troca iônica e eluição do DNA por água destilada seguida de centrifugação.

2.7. Ligação dos fragmentos de cDNA em plasmídeo vetor

Para clonagem dos fragmentos de cDNAs obtidos a partir da amplificação por PCR, foi utilizado o kit de ligação *pGEM-T Easy vector System* (Promega). Foi utilizada uma mistura de reação que consistia de 1 unidade de T4 DNA ligase, tampão apropriado, plasmídeo (pGEM-T) e o produto de PCR (purificado do gel de agarose). O plasmídeo e o produto de PCR foram utilizados em uma concentração de 1:3. Essa mistura foi incubada por 16 horas a 4ºC. O produto de ligação foi utilizado para transformar bactérias competentes. Para inserir fragmentos de cDNAs ao vetor pAE, tanto o inserto quanto o vetor foram digeridos com as enzimas de restrição apropriadas, purificados do gel de agarose e a seguir foram misturados

numa proporção de 3 volumes de inserto para 1 volume de plasmídeo adicionando 1 unidade de T4 DNA ligase (Gibco) e tampão apropriado. A amostra foi incubada a 16ºC por 16 horas e a seguir usada para transformar bactérias competentes.

2.8. Transformação de bactérias competentes

Alíquotas de 1 μL dos respectivos plasmídeo usados nessa tese foram adicionadas a 30μL de células competentes e mantidas em gelo por 30 minutos. Em seguida as células foram submetidas a um choque térmico a 42ºC por 30 segundos e novamente transferidas para o gelo por mais 2 minutos. Adicionamos então, 1 mL de meio LB, e essa solução foi incubada por 1 hora a 37ºC, para recuperação das bactérias e expressão das proteínas envolvidas na resistência ao antibiótico de seleção. Em seguida, a mistura era plaqueada em LB ágar com antibiótico apropriado e incubadas *overnight* a 37ºC, obtendo-se dessa forma colônias isoladas.

2.9. Purificação de plasmídeos a partir de cultura de bactérias

Os plasmídeos foram purificados utilizando-se o protocolo do kit *Wizard Miniprep* (Promega) a partir de 3,0 mL de cultura de bactérias cultivadas durante 16 horas a 37°C. O rendimento e pureza dos plasmídeos purificados foram avaliados através de determinação da absorbância das amostras a 260 e 280nm.

2.10. Reação de sequênciamento de DNA

O sequênciamento de DNA foi realizado na central de sequênciamento do Departamento de Bioquímica do Instituto de Química da USP – São Paulo, SP.

O para isso foi utilizado o aparelho ABI 377 DNA Sequencer (Applied Biosystems) e a técnica foi executada conforme recomendação do fabricante. A reação de sequênciamento foi realizada utilizando 2 μL da mistura de reação *Big Dye terminator mix* (Perkin Elmer), 200 – 500 ng de DNA, 10 pmols de oligonucleotídeos, $MgCl_2$ 1mM e tampão Tris 40mM pH 9,0 em um volume final de 15μL. A reação foi incubada por 1 minutos a 96ºC e em seguida submetida a 35

ciclos (96ºC por 45 segundos, 50ºC por 30 segundos e 60ºC por 4 minutos). Ao final, o DNA foi precipitado com etanol/acetato de sódio/glicogênio de acordo com a orientação do fabricante, e ressuspenso em tampão para aplicação no sequênciador automático.

2.11. Clonagem em vetor pAE das sequências codificantes para as proteínas recombinantes pCAD1, pCAD3 e LipMD

Para expressão da pCAD1, pCAD3 e LipMD na forma de proteínas recombinantes em bactérias foi escolhido um vetor chamado pAE (Ramos *et al.,* 2004). O cDNA clonado neste vetor tem sua expressão sob o controle da T7 RNA polimerase e a proteína é sintetizada fusionada a um peptídeo contendo 6 resíduos de histidina adjacentes que liga-se especificamente a uma resina cromatográfica de níquel. A estratégia adotada para a clonagem em vetor pAE passa pela amplificação por PCR do cDNA que codifica separadamente as proteínas pCAD1, pCAD3 e LipMD, utilizando como molde o plasmídeo pGEM-T contendo os respectivos insertos. Para tal foram desenhados *primers* específicos correspondentes às extremidades 5' e 3' do cDNA das sequências que codificam cada enzima. Adicionalmente, cada *primer* contém em suas extremidades sequências de nucleotídeos adicionais específicas para endonucleases de restrição que uma vez digeridas adquirem extremidade coesivas ao vetor pAE previamente digerido com as mesmas endonucleases de restrição.

Depois, os produtos de PCR (inserto) e o plasmídeo pAE digeridos foram submetidos a uma reação de ligação conforme descrito no item 2.7., e uma alíquota dessa reação foi usado para transformar células XL1-BLUE competentes. A partir de colônias isoladas das células transformadas foi preparada uma cultura em meio líquido que serviu como fonte de extração e purificação do plasmídeo pAE contendo as respectivas sequências trabalhadas nessa tese. Posteriormente estes plasmídeos foram separadamente utilizados para transformar células competentes de expressão (*E. coli*) e assim originar a linhagem utilizada para expressão das proteínas recombinantes.

Tabela 2. *Primers* usados nessa tese

Primer	***sequencia de nucleotídeos (5'→3')***
CADdeg	*GGH TCS TCS AAY YTS TGG
T3	ATT AAC CCT CAC TAA AGG GA
T7	TAA TAC GAC TCA CTA TAG GG
SP6	AGC TAT TTA GGT GAC ACT ATA
M13Fw	GTA AAA CGA CGG CCA GTA GA
M13Rv	GGG AAA CAG CTA TGA CCA TG
CAD1Ct	ATT TGG CAA TGG CAA AAC CAA CAC G
CAD2Ct	TTA CTT AGC CAA AGC AAA ACC
CAD3Ct	TTA GAC AGC TTC AGC AAA GC
pCAD1Nt_BamHI	CGC GGA TCC GAG GAG AAA TTA ATT CGA GTT CCC
pCAD1Ct_EcoRI	CGC GAA TTC CTT ATT TGG CAA TGG CAA AAC CAA CAC G
pCAD3Nt_BamHI	CGC GGA TCC GAT GTG GTT CGT GTT CCA AT
pCAD3Ct_HindIII	CGC AAG CTT TAG ACA GCT TCA GCA AAG C
RT CAD1Fw	GAG GAG AAA TTA ATT CGA GTT CCC
RT CAD1Rev	CAA GCC GGG TTC ACT CAA AGC
RT CAD2Fw	ACG CTC GTT CAA GTT CCC ATA
RT CAD2Rev	GTT GGC TGA TCA TGT TGT AC
RT CAD3Fw	GAT GTG GTT CGT GTT CCA AT
RT CAD3Rev	CTG TTA CCG GGT TCG TTC ATG
RT LipMDFw	GGA CAT GAA ATC TCG GCT ACA
RT LipMDRev	GTT GGG ATA GAA AGC ACC TTT
LipMDNt_PstI	CGC CTG CAG CTT CGT TAT TCC GGA GGA GGA A
LipMDCt_HindIII	CGC AAG CTT TAA CCC AAT ACA CCA TAA GGA TAC
rRNA MDFw	TCA GTG GGC AGG TTA GAC TT
rRNA MDRev	CCT TGT GTA TCA GGG TTT ATC

* H: A, C ou T, S: G ou C e Y: C ou T.
** Os nucleotídeos sublinhados indicam sítios de enzimas de restrição

2.12. Expressão da pCAD1, pCAD3 e LipMD recombinantes em vetor pAE

Bactérias Origami(DE3) foram transformadas separadamente com os plasmídeos pAE tendo como inserto a sequência codificante para as proteínas pCAD1 e pCAD3. Bactérias BL21(DE3)pLysS foram transformadas com o plasmídeo pAE tendo como inserto a sequência codificante para a proteína LipMD madura. As colônias de bactérias transformadas em cada caso foram crescidas em meio LB líquido contendo ampicilina (50µg/mL) e cultivadas a 37°C ou 20°C sob agitação de 150 rpm, até atingirem Abs_{600} = 0,6 – 1,0 e então induzidas com IPTG (0,5 mM) por tempos e temperaturas distintas, sob agitação de 150 rpm. Após esse procedimento as células eram sedimentadas por centrifugação a 7.000xg por 30min a 4ºC e então armazenadas a –20ºC. Foram coletadas alíquotas de 20µL de bactérias não-induzidas e induzidas a fim de verificar através de SDS-PAGE se houve produção da proteína recombinante.

2.13. Lise das células induzidas

As bactérias sedimentadas no item anterior correspondente a 50 mL de células induzidas (contendo a proteína recombinante) eram ressuspensas em 1 mL de tampão A (tampão fosfato de sódio 20mM, NaCl 150mM e imidazol 40mM pH 7,4). Após esse procedimento as células eram sonicadas (3 pulsos de 15 segundos com intervalos de 45 segundos) em *output* 4 em um sonicador *Branson* (com microtip) e centrifugadas a 7000xg por 30 minutos a 4ºC. O sobrenadante e o precipitado eram armazenados a –20ºC para posterior purificação da proteína recombinante ou análise em SDS-PAGE.

2.14. Purificacão das proteínas recombinantes (pCAD1, pCAD3 e LipMD) através de uma cromatografia de afinidade

Bactérias induzidas foram lisadas em tampão A (tampão fosfato de sódio 20 mM, NaCl 150mM, imidazol 40mM pH 7,4) e centrifugadas conforme descrito no item 2.13. O sobrenadante foi injetado em uma coluna cromatográfica de afinidade (HisTrap HP, 1 mL, GE) sendo que a proteína recombinante (contendo um peptídeo

de fusão contendo 6 resíduos de histina adjacentes) permanecia ligada a matriz e as proteínas que não fossem de interesse eram eluídas pelo mesmo tampão. As respectivas proteínas recombinantes ligadas a matriz foram eluídas em tampão B (tampão fosfato de sódio 20mM, NaCl 150 mM, imidazol 500mM pH 7,4), sendo que as frações recolhidas foram dialisadas e analisadas em SDS-PAGE para ser verificado o grau de pureza das respectivas proteínas recombinantes.

2.15. Auto-ativação da pCAD3 recombinante

A otimização nas condições de auto-ativação (conversão da proenzima pCAD3, na forma madura, CAD3) ocorreu conforme o método descrito por Westling *et al.,* 1999. Basicamente quantias iguais da proenzima pCAD3 foram incubadas em tampões, de diferentes pHs, por tempos distintos a 30°C. A reação foi interrompida pela adição de pepstatina A (10µM, Sigma). Alíquotas dessa preparação foram analisadas em SDS-PAGE (12%) para a verificação da ativação pelo deslocamento da massa molecular.

2.16. Determinação da atividade enzimática da CAD3 recombinante

Hemoglobina bovina (Sigma) foi desnaturada em HCl 0,06 N. A essa mistura foi adicionada tampão citrato-fosfato de sódio 100mM pH 2,5 de forma que a concentração final de hemoglobina fosse 1%. Os ensaios foram montados utilizando 20µL (0,35 µg) da CAD3 recombinante, juntamente com 20µL de hemoglobina desnaturada na ausência ou presença de inibidores de várias classes de proteinases (pepstatina A: aspártico proteinases (Sigma), leupeptina (Sigma): cisteína proteinase, E-64 (Sigma): cisteína proteinase, todos na concentração final de 20 µM). Essas reações foram incubadas em banho termostatizado (30ºC) por diferentes períodos de tempo. Cada reação foi interrompida pela adição de 200µL de tampão fosfato de sódio 200 mM pH 7,0. A atividade proteolítica foi também determinada conforme descrito em Yasuda et al., (1999) usando como substrato MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2, [MOCAc, (7-methoxicoumarina-4-il) acetil; Dnp, N-2,4-dinitrofenil; comprado da Calbiochem] em tampão citrato-fosfato de sódio 100 mM pH 3,0.

Este substrato de fluorescência apagada se torna fluorescente quando hidrolisado (excitação: 328 nm e emissão: 393 nm). Em cada determinação de atividade enzimática, os tubos contendo 10µL de substrato (5µM) foram misturados com 10 µL da CAD3 recombinante sendo esses tubos incubados a 30°C por ao menos 4 periodos de tempos onde a velocidade inicial foi calculada. As reações foram interrompidas pela adição de 1,0 mL de TCA (5%). Todas as análise foram executadas sobre condições tais que a atividade enzimática foi proporcional a concentração de proteína e ao tempo. Controle sem a enzima e outros sem os substratos foram incluídos.

O pH ótimo da CAD3 (madura) foi determinado pela análise de 0,35 µg da enzima em tampão citrato-fosfato de sódio 100mM nos pHs 2,0; 2,5; 3,0; 3,5; 4,0; 4,5; e 5,0 usando hemoglobina desnaturado ou MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2 (5 µM) como substrato. Incubações foram realizadas a 30°C como descrito anteriormente.

2.17. Análise da degradação da hemoglobina pela CAD3 recombinante

A atividade recombinante foi medida usando hemoglobina bovina (Sigma). Cem microgramas de hemoglobina foram incubados com 10 µL (0,35 µg) da CAD3 (madura) recombinante sozinha ou na presença de inibidores de proteinases em tampão citrato-fosfato de sódio 100mM pH 2,5 por 18 horas a 30ºC. A mistura reacional foi analisada em SDS-PAGE (12%) sob condições desnaturantes e corado com Coomassie Blue R-250. Pepstatina A, inibidor de aspártico proteinase; leupeptina, inibidor de serina proteinase e E-64, inibidor de cisteína proteinase, foram comprados da Sigma (USA) e todos eles utilizados na concentração final de 20 µM.

2.18. Eletroforese em placas de gel de poliacrilamida com SDS (SDS-PAGE)

As amostras de proteínas foram diluídas com tampão de amostra (Tris-HCl 120 mM pH 6,8, SDS 5% (m/v), β-mercaptoetanol 1,3mM, EDTA 1,0 mM, glicerol 20% (v/v) e azul de bromofenol 0,01% (m/v)). Antes da aplicação, as amostras foram fervidas junto com o tampão acima descrito por 5 minutos. O gel de poliacrilamida

12% com SDS 1% foi preparado segundo Laemmli (1970). A coloração dos géis foi feita com Coomassie Blue R-250 0,1% (p/v) preparado em uma solução de metanol 40% (v/v) e ácido acético 10% (v/v).

2.19. RT-PCR semi-quantitativo

Larvas de *Musca domestica* foram previamente mantidas em gel de amido 10% por 2 horas visando eliminar às bactérias presentes no trato intestinal. Essas larvas foram então imobilizadas em gelo e dissecadas em uma solução contendo 150mM NaCl usando luvas novas, pinças, tubos e vidrarias previamente tratados com DEPC (0,1% (v/v)). Os tecidos dissecados (ventrículo anterior, médio e posterior; hemolinfa; corpo gorduroso e túbulos de Malpighi) foram mantidos em gelo seco e armazenados a -80°C. O RNA total foi extraído com Trizol (ver item 2.2.) e usado para a síntese dos cDNAs correspondentes, utilizando-se da transcriptase reversa fornecida pelo kit *SuperScript First-Strand Synthesis for RT-PCR* (Invitrogen). O cDNA resultante foi usado como molde para a amplificação das sequências por PCR com *primers* específicos para as respectivas sequências (CAD1, CAD2, CAD3 e LipMD), tendo-se como controle constitutivo o RNA ribossomal do mesmo inseto (AY123346.1) (Tabela 2).

A reação de PCR foi realizada usando a Taq DNA polimerase (Invitrogen) em meio tamponado contendo 1,5 mM $MgCl_2$. As condições de PCR foram: 25 ciclos de 30 segundos 94°C (desnaturação), 30 segundos a 55°C (anelamento) e 1 minuto a 72°C (síntese). Os produtos de PCR obtidos foram analisados em gel de agarose 1%. O número de ciclos foi escolhido após várias tentativas, assim que a amplificação estava em fase log, resultando em bandas claras e visíveis.

O tamanho dos produtos do RT-PCR foi compatível com os tamanhos teóricos esperados, assim eliminando a possibilidade de contaminação por DNA genômico na preparação.

2.20. Construção da biblioteca de cDNA de ventriculo total de larvas de Musca domestica clonada em plasmídeo

Uma biblioteca unidirecional de cDNA foi construída a partir do RNA mensageiro isolado do ventrículo total de larvas de *Musca domestica*, utilizando o

kit *SMART cDNA library Construction Kit* (Clontech), utilizando o método de síntese por LD-PCR fornecido pelo fabricante. Resumidamente a fita simples do cDNA foi sintetizada tomando-se como molde o mRNA conforme descrito no item 2.2. A esse material foram adicionados, 10µM do *primer* SMART IV (5' - AAG CAG TGG TAT CAA CGC AGA GTG GCC ATT ACG GCC GGG – 3'), 10µM do CDS III/3' PCR *primer* (5' – ATT CTA GAG GCC GAG GCG GCC GAC ATG – (dT)30 (A/G/C) N – 3') e aquecidos por 2 minutos a 72ºC seguido pela adição de 1 µL da transcriptase reversa (*PowerScript Reverse Transcriptase*) em Tris 50mM pH 8,3, $MgCl_2$ 6mM, KCl 57mM, DTT 1mM e 1 mM dos nucleotídeos dATP, dCTP, dGTP e dTTP incubando essa mistura a 42ºC por 1 hora.

A amplificação do cDNA fita dupla foi executada tendo como molde uma alíquota da reação anterior combinados com o *primer* 5' PCR *Primer* (5' – AAG CAG TGG TAT CAA CGC AGA GT – 3") e CDS III/3' PCR *Primer* utilizando o kit *Advantage PCR Kit* (Clontech) em termociclador (GeneAmp 9700 (Perkin Elmer, Massachusetts, USA)) por 95ºC: 20 segundos seguidos de 22 ciclos de 95ºC: 5 segundos e 68ºC: 6 minutos. O produto desse PCR foi digerido com a enzima de restrição *SfiI* visando adicionar extremidades coesivas ao cDNA formado *in vitro*. O produto dessa digestão foi fracionado em uma coluna de cromatografia (CHROMA SPIN-400) e analisado em gel da agarose 1% sendo recolhidas às frações cujos produtos possuíssem o tamanho entre 300 – 2000 pb.

As frações escolhidas foram reunidas e ligadas no vetor pDNR-LIB (previamente linearizado contendo extremidades coesivas para ligação do inserto). Bactérias *Electromax™ DH10B™* competentes foram transformadas com o produto dessa ligação por eletroporação (2,0 kV, 200 Ω, 25 µF) e plaqueadas em placas de LB-ágar contendo cloranfenicol (30µg/mL) e incubadas a 37ºC durante 18 horas.

2.21. Amplificação e sequênciamento randômico dos clones de cDNA

Colônias isoladas obtidas no item anterior foram selecionadas randomicamente e incubadas a 37ºC por 16 horas em 100µL de meio LB líquido contendo cloranfenicol (30µg/mL) em placas de crescimento de bactérias contendo 96 poços. Um microlitro dessa cultura foi utilizados para amplificar os clones por PCR usando os *primers* M13Fw e M13Rv (Tabela 2) em termociclador GeneAmp 9700 (Perkin

Elmer, Massachusetts, USA): 95ºC: 5 minutos (para romper a parede celular bacteriana) seguido de 30 ciclos de 95ºC: 30 segundos, 55ºC: 30 segundos, 68ºC: 3 minutos. Posteriormente cada reação foi analisada em gel de agarose 1% para verificar a presença do inserto. Um microlitro do produto do PCR foi utilizado na reação de sequênciamento sendo executada conforme descrito no item 2.10.

2.22. Análises de bioinformática

Todas as análise de bioinformática foram realizadas em um microcomputador Pentium 4, 2,26 GHz, utilizando os sistemas operacionais Windows XP (Seattle, WA, USA) e Linux (Debian Testing com kernel 2,6).

Os cromatogramas foram analisados e separados em singletos e contíguos utilizando os programas phred, phrap e consed (Ewing *et al.,* 1998; Gordon *et al.,* 1998). BLAST client 3 foi utilizado para analisar as sequências obtidas, por BLAST (de forma automatizada) contra sequências depositadas no GENBANK (banco de dados não redundante) (Wheeler *et al.,* 2000). As sequências foram manualmente anotadas e caracterizadas em termos de GO (The Gene Ontology Consortium, 2000). Sítio de hidrólise do peptídeo sinal e sítios de N e O - glicosilação foram determinados com os programas *SignalP* 3.0 (http://www.cbs.dtu.dk/services/SignalP/), *NetNGlyc* 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) e *NetOGlyc* 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc/) respectivamente. Massa molecular teórica e ponto isoelétrico teórico foram determinados por *ProtParam tool* (http://www.expasy.org/tools/protparam.html). As traduções das sequências de aminoácidos presentes nessa tese foram determinadas usando *Translate tool* (http://www.expasy.org/tools/dna.html).

2.23. Análise filogenética

Análises filogenéticas usadas nessa tese foram executadas com os programas Clustal X (Thompson *et al.,* 1997) e MEGA 4 (Tamura, Dudley, Nei e Kumar 2007). As análises das distâncias foram feitas com o algorítimo *Neighbor – Joining* (Saitou e Nei, 1987). *Bootstraps* (Felsenstein, 1985; Hills e Bull, 1993) foram determinados baseados em 1000 réplicas.

2.24. Medida da atividade lipolítica usando pNP-ésteres

A atividade lipolítica foi realizada conforme descrito em Winkles e Stuckmann, (1979) com pequenas modificações. O substrato na forma de emulsão consistia em 90 mL de tampão fosfato de sódio 100 mM pH 7,5 contendo 207 mg de desoxicolato de sódio, 100 mg de goma arábica misturado com 10 mL de pNP-ésteres (3 mM) dissolvidos em acetonitrila (pNP-acetato (2C), butirato (4C), valerato (5C), caprilato (8C), caprato (10C), laurato (12C), miristato (14C), palmitato (16C) e estearato (18C)) e solubilizados em Triton X-100 (0,36% (v/v)). Cinquenta microlitros dos respectivos substratos (concentração final de ensaio 150 μM) foram misturados com 50μL da enzima recombinante purificada ou amostras de proteínas do conteúdo luminal de larvas de *M. domestica*. A quantificação do p-nitrofenolato liberado durante a reação foi medido em espectrofotômetro a 410nm. Uma unidade é definida como a quantia de enzima necessária para a produção de 1μmol de p-nitrofenolato liberado por minuto.

Todas as análise foram executadas sobre condições tais que a atividade enzimática foi proporcional a concentração de proteína e ao tempo. Controle sem a enzima e outros sem os substratos foram incluídos.

2.25. Efeito da temperatura e pH sobre a estabilidade e atividade da LipMD recombinante

O efeito da temperatura na atividade da LipMD foi determinado pela incubação da enzima em tampão fosfato de sódio 100 mM pH 7,5 (ver item 2.24.) a 25°C, 30°C, 35°C, 37°C, 42°C, 45°C, 47°C, 50°C e 60°C (1 hora), tendo como substrato pNP-miristato em uma concentração final de 150 μM. A estabilidade da LipMD frente a temperatura também foi investigada. Alíquotas da enzima foram incubadas por 1 hora nas respectivas temperaturas em tampão fosfato de sódio 100 mM pH 7,5 sendo que a atividade remanescente foi medida a 30°C. O pH ótimo da LipMD recombinante foi determinado usando pNP-miristato como substrato em uma concentração final de 150 μM a 30°C (ver item 2.24.) em diferentes valores de pHs. Os tampões utilizados foram: fosfato de sódio 100 mM pHs 6 – 8 e Bis-Tris-Propano 100 mM pHs 6 – 9. Para verificar a estabilidade da enzima frente ao pH, 5 μL da

enzima pura (1μg) foram incubados a 30°C por 1 hora com 45 μL dos respectivos tampões listados acima em vários valores de pHs. As amostras foram então diluídas (diluição final = 1/10) em tampão fosfato de sódio 200 mM pH 7,5 e a atividade remanescente foi medida conforme descrito no item 2.24.

2.26. Estabilidade da LipMD frente a solventes orgânicos

A estabilidade da LipMD recombinante frente a solventes orgânicos (metanol, etanol, acetona, 2-propanol e acetonitrila) na concentração de 50% foi examinada pela incubação da enzima por 1 hora a 30°C, sendo que a atividade remanescente foi determinada usando pNP-miristato como substrato em uma concentração final de 150 μM, conforme descrito no item 2.24.

2.27. Produção dos anticorpos policlonais

Antes das imunizações, foram retirados 5 mL de sangue de cada coelho, com o intuito de se obter o soro pré-imune. Para a produção do anticorpo, foi feita uma primeira aplicação de 1 mg da proteína recombinante pCAD3 e 0,5 mg da LipMD com adjuvante completo de Freud (Sigma) na proporção de 1:1 (v/v). Uma segunda aplicação foi realizada 28 dias após a primeira, aplicando os mesmos valores em massa de proteína, mas agora com adjuvante incompleto de Freud (Sigma). Após 15 dias da segunda imunização, o sangue do coelho foi retirado para se obter o soro imune. Para se obter o anticorpo semi-purificado, o sangue imune das respectivas preparações foram deixados por 1 hora a 37°C e em seguida a 4°C por 19 horas. O mesmo procedimento foi realizado com o sangue pré-imune.

Os respectivos materiais foram então centrifugados (3000 rpm, 15 minutos a 4°C) e ao sobrenadante (soro) foi adicionada uma solução de sulfato de amônio saturada preparada em tampão fosfato de sódio 200 mM pH 6,5 na proporção de 1:1 (v/v). Após permanecer por 18 horas a temperatura ambiente a suspensão obtida foi centrifugada (12000 rpm, 15 minutos a 4°C). O precipitado foi então ressuspenso em uma solução 50% saturada de sulfato de amônio pH 6,5 e novamente centrifugado nas mesmas condições anteriores. Esse passo foi então repetido por mais uma vez e o precipitado final obtido foi ressuspendido em NaCl

125 mM. As respectivas preparações foram então dialisadas por 20 horas contra 1000 volumes de NaCl 125 mM com uma troca da salina. Os dialisados foram aliquotados e mantidos a -20°C até o uso.

2.28. Western blot

Amostras das proteínas recombinantes (pCAD3 e LipMD) ou proteínas presentes do conteúdo luminal ou epitélio de larvas de *Musca domestica* foram submetidas a uma eletroforese em SDS-PAGE (12%). O gel resultante foi submetido ao método de transferência de proteínas para membranas de nitrocelulose conforme descrito por Towbin *et al.,* 1979. O gel de poliacrilamida e a membrana de nitrocelulose foram equilibrados por 15 minutos em tampão Towbin (Tris-HCl 25 mM pH 8,6, glicina 192 mM e metanol 20%). Em seguida foi realizada uma transferência eletroforética das proteínas contidas nos géis para as membranas de nitrocelulose (45µm). Foi utilizado um sistema de transferência de proteínas semi-seco (Bio Rad) na presença do tampão Towbin nas condições indicadas pelo fabricante do aparelho (10 V, 5,5 mA/cm^2 de gel, 20 minutos). A eficácia da transferência das proteínas foi monitorada por marcadores de massa moleculares pré-corados (Low-Range, Bio Rad) aplicados na eletroforese. Antes de serem imunoensaiadas, as membranas de nitrocelulose foram deixadas imersas em uma solução de 5% de leite em pó desnatado dissolvido em tampão TBST (Tris-HCl 50 mM pH 7,5, NaCl 0,15M, Tween 20 0,05% (v/v)) por 1 hora a temperatura ambiente. As membranas foram então incubadas por 2 horas com o anticorpo para a pCAD3 e LipMD separadamente diluído 2000 vezes em TBST (contendo 5% de leite em pó). As membranas foram então lavadas com TBST por 10 minutos por 6 vezes e incubadas por mais 2 horas com uma solução de anti-IgG de coelho (Sigma) acoplada a peroxidase diluída 20.000 vezes em TBST. A seguir, as membranas de nitrocelulose foram lavadas por 6 vezes de 10 minutos com TBST e então reveladas com o método fornecido pelo o k*it Imobbilon Western Chemiluminescent HRP substrate* (Millipore). Nesse método a atividade enzimática da peroxidase leva a emissão de luminescência que é detectada utilizando-se filmes específicos (*Hyperfil–ECL*, Amersham) através de tempos de exposição de 5 a 25 segundos executada conforme instruções do fabricante.

2.29. Transferência da CAD3 (madura) de Musca domestica *para membranas de PVDF e obtenção da sequência N-terminal*

Preparações de CAD3 previamente ativadas (ver item 2.15.) foram inibidas por pepstatina A (10 μM) e dialisadas em água milli-Q e secadas por centrifugação a vácuo. Um total de 80 μg dessa preparação foi ressuspendida em tampão de amostra e submetida a eletroforese em SDS-PAGE (12%). As proteínas presentes no gel foram transferidas para membranas de PVDF segundo o método descrito em Matsudaira (1987). A membrana de PVDF foi imersa por alguns segundos em metanol 100% e então lavada em água milli-Q. Depois o gel foi incubado juntamente com a membrana de PVDF em tampão de transferência (CAPS 10 mM pH 11, metanol 10% (v/v)) por 20 minutos. A transferência das proteínas foi realizada usando a cuba de transferência *Mini Trans-Blot* (Bio Rad). As condições da eletroforese foram de 15V por 30 minutos. Após a transferência, a membrana foi corada com uma solução 0,1% (m/v) de Coomassie Blue R-250 em metanol 50% (v/v) por 5 minutos sob agitação.

As proteínas transferidas eram então visualizadas pela descoloração da membrana em metanol 50% (v/v). As bandas correspondentes a CAD3 (madura) foram recortadas e submetidas a sequeciamento dos resíduos de aminoácidos N-terminal em um sequenciador do tipo PPSQ/23 (Shimadzu Corp., Kyoto, Japan) com a colaboração da Dra Izaura Yoshico Hirata do Departamento de Biofísca da Universidade Federal de São Paulo (UNIFESP).

2.30. Experimentos de Dicroismo Circular (CD)

Ensaios de dicroismo circular (CD) para a LipMD foram realizados em um espectropolarímetro J-720 (Jasco, Inc). A LipMD recombinante pura foi utilizada na concentração de 31 μM usando cubetas de quartzo de 0,5 mm. Foram realizadas varreduras de 190 a 245 nm em meio tamponado (tampão tris-HCl 5 mM pH 8,0).

A elipsidade molar residual (ϴ) (deg x cm^2 x $dmol^{-1}$) foi calculada utilizando a seguinte fórmula:

$$\frac{[\text{elipsidade } \Theta \text{ (mdeg) x massa molecular da LipMD (Da)}]}{[10 \text{ x caminho da cubeta (cm) x concentração da proteína (mg/mL) x (n° aminoácidos -1)}]}$$

2.31. Imunocitolocalização por microscopia eletrônica de transmissão

A imunocitolocalização por microscopia eletrônica foi feita usando os anticorpos produzido separadamente contra as enzimas pCAD3 e LipMD. O ventrículo de larvas de *Musca domestica* foi dissecado e dividido em 8 regiões (Figura 5) que foram transferidas separadamente para solução de fixação (glutaraldeído 5%, paraformaldeído 4% em tampão fosfato de sódio 100 mM pH 7,4). Os tecidos fixados foram desidratados em soluções com concentrações crescentes de etanol a temperatura ambiente e finalmente embebidos em resina acrílica L. R. White (Agar Aids, USA) por 24 horas a 50°C. Secções finas de tecidos foram obtidas em ultramicrótomo e transferidas para telas de níquel. A cada um dos cortes foi aplicada uma gota de TBS (Tris 0,1 M, NaCl 0,3 M, NaN_3 0,1% (m/v)) contendo albumina sérica bovina 1% (TBS-BSA). Após 30 minutos foi adicionado separadamente os respectivos anticorpos (anti-pCAD3 e anti-LipMD) em uma diluição de 1:200 para ambos os casos em TBS-BSA e incubado por 90 minutos a temperatura ambiente. Em seguida, o corte foi lavado com TBS-BSA e incubado com o anticorpo anti-IgG de coelho, produzido em cabra, conjugado a partículas de 10 nm de ouro coloidal. Esse último foi diluído na proporção de 1:15 em tampão TBS-BSA. Os cortes foram finalmente lavados com água milli-Q e corados com acetato de urânio e citrato de chumbo (Reynolds, 1963). Controles tiveram o anticorpo primário omitidos ou utilizaram o soro pré-imune. Todos os cortes foram visualizados em um microscópio eletrônico de transmissão Phillips 301.

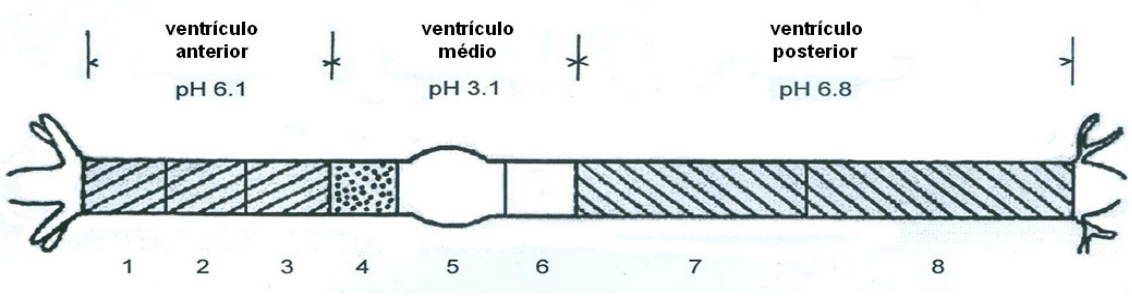

Figura 5. Representação esquemática do ventrículo de larvas de M. domestica, mostrando as principais regiões com seus respectivos pHs luminais baseado em Terra et al., (1988). As principais regiões do ventrículo foram subdivididas como

monstrado. Todas as regiões contém células colunares porém a região 4 contém também células intersticiais e oxínticas.

2.32. Meios de cultura

2.32.1. LB Líquido

10g NaCl, 10g bactotriptona e 5g de extrato de levedura foram dissolvidos em 1,0 L de água milliQ e em seguida autoclavados.

2.32.2. Placas LB – Ágar

Foram usados 20g bactoágar para 1,0 L de meio LB. Após a autoclavagem a mistura foi espalhada em placas de Petri. Se necessário foi adicionado o antibiótico apropriado antes do meio solidificar.

2.32.3. Determinação da atividade lipolítica da LipMD recombinante sobre placas de ágar-tributirina e ágar-trioleína

O método foi realizado conforme descrito em Kouger e Jaeger (1987) com pequenas modificações. Para a determinação de atividade de carboxiesterase o método consiste em produção de placas de LB-ágar com tampão fosfato 100 mM pH 7,5 contendo Triton X-100 (0,36% (v/v)) e tributirina (1% (v/v)) sendo autoclavado por 20 minutos. Alíquotas de 2 μL (0,35 μg) da LipMD recombinante purificada foram aplicadas nessas placas contendo a emulsão polimerizada. A presença de halos translúcidos indica a hidrólise do substrato (tributirina) pela enzima recombinante. Para verificar a atividade de lipase, foi utilizado placas de LB-ágar contendo a emulsão acima descrita, porém tendo como substrato trioleína (2,5 % (v/v)) combinada com Rodamina B (0,001 % (v/v)). A hidrólise da trioleína pode ser monitorada pela presença de halos fluorescentes na presença de radiação ultra violeta (312 nm). As placas utilizadas nesse experimento foram incubadas a 30°C por 20 minutos (placas contendo tributirina) ou 15 horas (contendo trioleína). Como

controle negativo foi utilizada a mesma preparação da enzima recombinante, porém fervida a 100 °C por 10 minutos.

3. RESULTADOS

4.1. Estudo das aspártico proteinases intestinais de Musca domestica

4.1.1. Amplificação dos fragmentos de cDNAs que codificam aspártico proteinases utilizando a técnica de PCR por homologia

Após a comparação das sequências de aminoácidos de aspártico proteinases de vários organismos, incluindo pepsinogênio de mamíferos, reninas e catepsinas-D (Tang e Wong, 1987), desenhamos um *primer* degenerado (CADdeg) que reconhece a sequência consenso presente no sítio ativo das aspártico proteinases de eucariotos. Quando o *primer* degenerado CADdeg foi combinado com o *primer* T3 (específico do braço do fago) em um PCR, utilizando como molde uma biblioteca de cDNA de *Musca domestica* clonada em fago λ zap II, uma banda de aproximadamente 1,0 kb foi amplificada (Figura 6).

O produto de PCR foi purificado a partir do gel de agarose 1%, clonado em vetor pGEM-T e transformado em células competentes XL1-BLUE. Cinquenta e sete colônias foram crescidas em meio LB líquido e tiveram extraídos seus respectivos plasmídeos que foram posteriormente sequênciados. Os cromatogramas obtidos nesse sequênciamento foram analisados com os programas Phred, Phrap e Consed (Ewing *et al.,* 1998; Gordon et. al., 1998). A sequência consenso formada por 2 ou mais sequências é nomeada por contíguo ou por singleto se formada por um única sequência. Foram encontrados 3 contíguos distintos cujos *frames* de leitura codificavam sequências de aminoácidos que apresentavam identidade com aspártico proteinases similares a catepsinas-D de insetos. Tomando como molde a extremidade 3' desses 3 contíguos, desenhamos separadamente para cada contíguo *primers* específicos reversos (CAD1Ct, CAD2Ct e CAD3Ct) (Tabela 2), a partir do *stop códon* de cada sequência que combinados com o *primer* T7 (específico do outro braço do fago) teve como produto de PCR uma banda de aproximadamente 1,2 kb em cada situação (Figura 7). Os respectivos produtos de

PCR foram separadamente purificados do gel de agarose 1% e clonados no plasmídeo pGEM-T. As respectivas construções foram separadamente sequênciadas e verificadas a presença das *ORFs* correspondentes para as 3 aspártico proteinases similares a catepsinas-D (ppCAD1, ppCAD2 e ppCAD3).

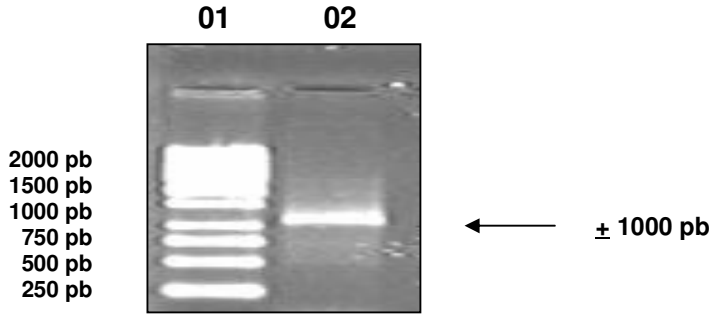

Figura 6. Eletroforese em gel de agarose 1% do produto de PCR, onde foram utilizados os primers CADdeg e T3, tomando como molde uma biblioteca de cDNA de Musca domestica. O gel foi tratado com brometo de etídeo e as bandas de DNA visualizadas com luz ultravioleta. 01. Padrão de DNA, 02. produto do PCR.

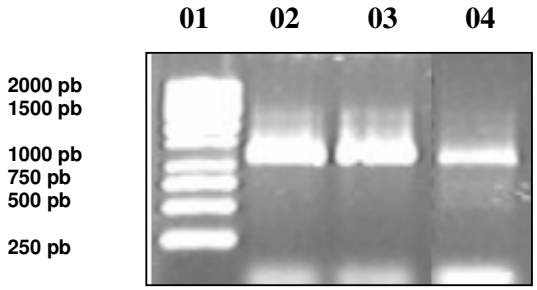

Figura 7. Eletroforese em gel de agarose 1% dos produtos de PCR onde foram obtidos os sequência completas (cDNAs) que codificam as proteínas ppCAD1, ppCAD2 e ppCAD3, tomando como molde uma biblioteca de cDNA de Musca domestica. O gel foi tratado com brometo de etídeo e as bandas de DNA visualizadas em luz ultravioleta. 01. Padrão de DNA; 02. produto do PCR com os primers CAD1Ct + T7; 03. produto do PCR com os primers CAD2Ct + T7; 04. produto do PCR com os primer CAD3Ct + T7.

4.1.2. Identificação, clonagem e sequênciamento dos cDNAs que codificam aspártico proteinases ventriculares em Musca domestica

Três cDNAs que codificam aspártico proteinase de *Musca dometica* similares a catepsinas-D (ppCAD1, ppCAD2 e ppCAD3) foram obtidos a partir de uma biblioteca de cDNA do epitélio do tubo digestivo, clonada em fago λ zap II, utilizando a sequência consenso do banco de dados, através de PCRs por homologia. As sequências de aminoácidos deduzidas a partir daqueles cDNAs incluem a sequência para o peptídeo sinal, propeptideo e enzima madura.

Aparentemente a mais expressa ppCAD é a ppCAD3, pois em um total de 57 sequências identificadas por BLAST achadas nos procedimentos de varredura com *primer* degenerado, 30 sequências correspondem à ppCAD3, 10 à ppCAD2 e 17 à ppCAD1. Todas as ppCADs possuem os ácidos aspárticos catalíticos Asp_{33} (juntos com T_{34} e G_{35}) e Asp_{229} (juntos com T_{230} e G_{231}) (numeração para uma catepsina-D bovina, Metcalf e Fusek, 1993) (Figura 8). A catepsina-D bovina, pepsinogênio humano e todas as ppCADs possuem também os resíduos conservados que formam a fenda de ligação do substrato na catepsina-D bovina (Metcalf e Fusek, 1993): V_{30}, D_{33}, G_{35}, S_{36}, Y_{78}, G_{79}, S_{80}, F_{124}, F_{129}, I_{132}, Y_{203}, D_{229}, G_{231}, T_{232}, S_{233} (numerção para a catepsina-D bovina, Metcalf e Fusek, 1993).

O sítio de hidrólise do propeptídeo *in vitro* previsto por homologia através do alinhamento das sequências de aminoácidos são: ppCAD1, Leu_{49} – Arg_{50}; ppCAD2, Leu_{40} - Lys_{41}; ppCAD3 Leu_{44} – Arg_{45}. O sítio de hidrólise *in vitro* da liberação do propeptídeo da pCAD3 durante a auto-ativação sobre condições ácidas foi confirmado pelo sequenciamento N-terminal da CAD3 madura (ver item 4.1.4.1.).

4.1.2.1. ppCAD1

A sequência de cDNA que contém a *ORF* codificante para a ppCAD1 (preproCAD1) mostra que esta possui 1439 pares de bases. O códon da metionina que inicia a transcrição do gene encontra-se na posição (82 – 84) e o códon de término de transcrição na posição (1252 –1254). Este cDNA possui uma *ORF* que localiza-se entre a base 82 e 1254, codificando 390 aminoácidos com massa molecular teórica de 42,42 kDa (Figura 9). O peptídeo sinal foi identificado entre os

resíduos Ala^{20} e Glu^{21} da ppCAD1, que resulta em uma proenzima 370 aminoácidos tendo um ponto isoelétrico (pI) e massa molecular (MM) teórica de 5,63 e 40,21 kDa, respectivamente. A análise da sequência de aminoácidos revela que a enzima possui os dois ácidos aspárticos catalíticos (Asp^{89} e Asp^{275}) característicos do único sítio ativo das aspártico proteinases (Tabela 3).

Tabela 3. Massas moleculares, pontos isoelétrico teóricos, número de aminoácidos, resíduos catalíticos e sítio de hidrólise do peptídeo sinal das sequências derivadas dos ppCADs clonadas apartir de uma biblioteca de cDNA intestinal de larvas de Musca domestica.

Enzima	N° de acesso GenBank	MM teórica (kDa)	pI teórico	N° de amino ácidos	Resíduos catalíticos	Peptídeo sinal
ppCAD1	EF193385	42,42	5,64	390	Asp^{89} e Asp^{275}	Ala^{20} - Glu^{21}
pCAD1		40,21	5,63	370		
ppCAD2	EF193384	42,04	4,67	379	Asp^{87} e Asp^{272}	Ala^{17} - Tre^{18}
pCAD2		40,11	4,65	362		
ppCAD3	EF 151165	41,30	4,29	379	Asp^{87} e Asp^{272}	Ala^{17} - Asp^{18}
pCAD3		39,67	4,25	362		
pCAD3*		40,86	4,50	371		
CAD3**		36,51	3,99	335		

*ProtParam é uma ferramenta da internet (http://av.expasy.org.tools/protparam.html), onde vários parâmetros foram computados a partir da sequência de aminoácidos de cada proteína. *pCAD3 fusionada a um peptídeo de fusão contendo $(His)_6$. **MM da CAD3 madura* in vitro*. As numerações usadas nessa tabela para identificação dos resíduos catalíticos ou sítio de hidrólise do peptídeo sinal são as das próprias sequências. No GENBANK a sequência ppCAD1 (EF193385) está identificada como AspMD03; a ppCAD2 (EF193384) como AspMD02 e ppCAD3 (EF151165) como AspMD01.*

A sequência de aminoácidos da ppCAD1 possui 1 sítio potencial de N-glicosilação no resíduo de aminoácido N_{126}. Análise do gene transcrito através de

BLASTX no GENBANK produziu uma *ORF*, revelando ser uma catepsina-D com alta identidade com aspártico proteinases de artrópodes e eucariotos. Essa análise mostra que a ppCAD1 possui 48% de identidade com a sequência de aminoácidos de uma catepsina-D de *Bos taurus* (AAB26186.1), 73% com uma catepsina-D de *Aedes aegypti* (XP_001657556.1), 40% de identidade com uma sequência de pepsinogênio de *Homo sapiens* (AAA60061.1) e por fim 46 % e 45 % de identidade com a ppCAD2 (ABM69085) ppCAD3 (ABL84270) de *Musca domestica,* respectivamente.

4.1.2.2. ppCAD2

A sequência de cDNA que contém a *ORF* codificante para a ppCAD2 (preproCAD2) mostra que esta possui 1162 pares de bases. O códon da metionina que inicia a transcrição do gene encontra-se na posição (7 – 9) e o códon de término de transcrição na posição (1144 –1146). Este cDNA possui uma *ORF* que localiza-se entre a base 7 e 1146, codificando 379 aminoácidos com uma massa molecular teórica de 42,04 kDa (Figura 10). O peptídeo sinal foi identificado entre os resíduos Ala^{17} e Tre^{18} da ppCAD2, que resulta em uma proenzima com 362 aminoácidos tendo um ponto isoelétrico (pI) e massa molecular (MM) teórica de 4,65 e 40,11 kDa respectivamente. A análise da sequência de aminoácidos revela que a enzima possui os dois ácidos aspárticos catalíticos (Asp^{87} e Asp^{272}) característicos do único sítio ativo das aspártico proteinases (Tabela 3).

Análise do gene transcrito através de BLASTX no GENBANK produziu uma *ORF* revelando ser uma catepsina-D com alta identidade com aspártico proteinases de artrópodes e eucariotos. Essa análise mostra que a ppCAD2 possui 38% de identidade com a sequência de aminoácidos de uma catepsina-D de *Bos taurus* (AAB26186.1), 47% com uma catepsina-D de *Aedes aegypti* (XP_001657556.1), 37% de identidade com uma sequência de pepsinogênio de *Homo sapiens* (AAA60061.1) e por fim 46 % e 53 % de identidade com a ppCAD1 (ABM69086) ppCAD3 (ABL84270) de *Musca domestica,* respectivamente.

```
BtCAD      1   ------------------------------------------------------------GP 2
ppCAD1     1   MIKFVILLSLVSLLLPESFAEEKLIRVPIQKIKSARKHFYEVGTELQQLRLTY-----------GAG-GV 58
ppCAD2     1   -----MWKLCVVLWSALLLAEATLVQVPITKVKETKSKANEIRKLK--AKYGGTP--------KAEIRDL 55
ppCAD3     1   -----MLKSLGMLAIVLAVAIADVVRVPIHRHENFVKTSKDIRSEKAVLRSKYN---------LPQPRDV 56
AappCAD    1   ----MLIKSIIALVCLAVLAQADFVRVQLHKTESARQHFRNVDTEIKQLRLKY-----------NAVSGP 55
HsppPEPS   1   -------MKWLLLLGLVALSECIMYKVPLIRKKSFRRTLSERGLLKDFLKKHNLNPARKYFPQWKAPTLV 63

                                          *
                                       a  b cd
BtCAD      3   IPELLKNYMDAQYYGEIGIGTPPQCFTVVFDTGSANLWVPSIHCKLLDIACWTHRKYNSDKSSTYVKNGT 72
ppCAD1     59  TPEPLSNYLDAQYYGPISIGTPPQDFKVVFDTGSSNLWVPSKKCHLTNIACLMHNKYDATKSKTFKQNGT 128
ppCAD2     56  VVEKLFNYVDDSYYGKITIGTPGQEFLVLFDTGSSNLWVPVAPCSADNAACENHNTYDPSASSTHVKKGE 125
ppCAD3     57  SDEPLENSLNMKYYGDITIGTPPQKFVVLFDTGSSNLWVPSSHCWIWDIACKKHNQYNHDDSSTYVKNGE 126
AappCAD    56  VPEPLSNYLDAQYYGAITIGTPPQSFKVVFDTGSSNLWVPSKECSFTNIACLMHNKYNAKSSTFEKNGT 125
HsppPEPS   64  DEQPLENYLDMEYFGTIGIGTPAQDFTVLFDTGSSNLWVPSVYCS--SLACTNHNRFNPEDSSTYQSTSE 131

               efg                                          h    i  j
BtCAD      73  TFDIHYGSGSLSGYLSQDTVSVPCNPSSSSPGGVTVQRQTFGEAIKQPGVVFIAAKFDGILGMAYPRISV 142
ppCAD1     129 EFAIHYGSGSLSGYLSTDTVNI---------GGLDIKDQTFAEALSEPGLVFVAAKFDGILGLGYSSISV 189
ppCAD2     126 SFSIQYGSGSLSGYLVEDTVDV---------EGLKIKKQVFAAATNEPGETFVYAPFDGIMGMGFKSIAV 186
ppCAD3     127 LISISYGSGSMSGFLSQDDVTV---------EGLTIKNQVFAEAMNEPGNSFTDANFDGIFGMAYQSLAE 187
AappCAD    126 AFHIQYGSGSLSGYLSTDTVGL---------GGVSVTKQTFAEAINEPGLVFVAAKFDGILGLGYSSISV 186
HsppPEPS   132 TVSITYGTGSMTGILGYDTVQV---------GGISDTNQIFGLSETEPGSFLYYAPFDGILGLAYPSISS 192

                                                                                k
BtCAD      143 NNVLPVFDNLMQQKLVDKNVFSFFLNRDPKAQPGGELMLGGTDSKYYRGSLMFHNVTRQAYWQIHMDQLD 212
ppCAD1     190 DGVKPPFYAMYEQGLISQPIFSFYLNRDPKAPEGGEIIFGGSDPDHYTGDFTYLPVTRKAYWQIKMDSAS 259
ppCAD2     187 DDVTPPWYNMISQHLISEKVFSFYLARRGTSDEGGVMVVGGNDDRYYEGDFHYVPVSEQGYWQFEMAEAH 256
ppCAD3     188 DNVVPPFYNMFAQGLVDANMFSFLLNRDGTSTDGGQMILGGVDSSLYTGDITYVPVSSQGYWQFEVTSGA 257
AappCAD    187 DGVVPVFYNMFNQGLIDAPVFSFYLNRDPSAAEGGEIIFGGSDSNKYTGDFTYLSVDRKAYWQFKMDSVK 256
HsppPEPS   193 SGATPVFDNIWNQGLVSQDLFSVYLSADDQS--GSVVIFGGIDSSYYTGSLNWVPVTVEGYWQITVDSIT 260

                          *
                          l mno
BtCAD      213 VGSSLTVCKGGCEAIVDTGTSLIVGPVEEVRELQKAIGAVPLIQ-GEYMIPCEKVSSLPEVTVKLGGKDY 281
ppCAD1     260 MG-DLKCAKGGCQVIADTGTSLIALPPSEATSINQAIGGTPIMG-GQYMVACEDIPKLPVIKFVLGGKTF 327
ppCAD2     257 VN-GVRICDR-CQAIADTGTSLIAVPTDKYEEIQKEIGATFSYDTYEYMLDCSKIDDLPVVTFRLGDGTF 324
ppCAD3     258 IK-GQSICDN-CQAIADTGTSLIVAPSDAYNTLNAEIGATYNEDDGNYYVDCSAVDSLPDVTFVIGGTTF 325
AappCAD    257 VG-DTEFCNNGCEAIADTGTSLIAGPVSEVTAINKAIGGTPIMN-GEYMVDCSLIPKLPKISFVLGGKSF 324
HsppPEPS   261 MNGEAIACAEGCQAIVDTGTSLLTGPTSPIANIQSDIGASENSD-GDMVVSCSAISSLPDIVFTINGVQY 329

BtCAD      282 ALSPEDYALKVSQAETTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRDQNRVGLAEAARL 346
ppCAD1     328 ELEGKDYVLRIAQMGKTICLSGFMGIDIPPPNGPLWILGDVFIGKYYTEFDMGNDRVGFAIAK-- 390
ppCAD2     325 TLEGRDYVIKS---DDNQCSSAFE-----DGGTDFWILGDVFIGKYYTTFDAEHNRVGFALAK-- 379
ppCAD3     326 TLPASAYIVTV---DGN-CMSSFT-----YMGTDFWILGDVFIGKYYTVFDFANNRVGFAEAV-- 379
AappCAD    325 DLEGADYVLRVAQMGKTICLSGFMGIDIPPPNGPLWILGDVFIGKYYTEFDMGNDRVGFATAV-- 387
HsppPEPS   330 PVPPSAYILQS----EGSCISGFQGMNLPTESGELWILGDVFIRQYFTVFERANNQVGLAPVA-- 388
```

Figura 8. Alinhamento das sequências de aminoácidos deduzidos a partir das sequências de cDNAs das preprocatepsinas-D e pepsina de vários organismos. BtCAD (AAB26186.1) de Bos taurus; ppCAD1 (ABM69086), ppCAD2 (ABM69085), ppCAD3 (ABL84270) de Musca domestica; AappCAD (XP_001657556.1) de Aedes aegypti; HsppPEPS (AAA60061.1) de Homo sapiens. As letras em caracteres minúsculos (a-o) indicam os resíduos conservados pertencentes aos subsítios que compõe a fenda de ligação do substrato. Os dois ácidos aspárticos catalíticos são indicados pelos asteriscos. Caixas pretas com letras em branco mostram resíduos

idêntico e caixas cinzas com letras em preto mostram resíduos com cadeias laterais semelhantes.

```
1    tttccatcgggaaaacaacaacttatttgttcgaagccaagatcaagtagaaaaataaat   60
61   ccacaaataaaaagtggaacaatgattaagtttgtcattttattaagcttagtcagctta  120
1                           M  I  K  F  V  I  L  L  S  L  V  S  L    13
121  ttgctcccggagagctttgctgaggagaaattaattcgagttcccatccagaaaataaaa  180
14    L  L  P  E  S  F  A  E  E  K  L  I  R  V  P  I  Q  K  I  K    33
181  tcggcacgaaaacatttctatgaggtaggaacggaattgcaacaactacgtttgacatac  240
34    S  A  R  K  H  F  Y  E  V  G  T  E  L  Q  Q  L  R  L  T  Y    53
241  ggagccggaggtgttacacccgaacctctatcaaattatttggatgcgcaatactatggc  300
54    G  A  G  G  V  T  P  E  P  L  S  N  Y  L  D  A  Q  Y  Y  G    73
301  ccaatttcaattggcacaccaccacaagatttcaaggtagttttcgataccggctcttcg  360
74    P  I  S  I  G  T  P  P  Q  D  F  K  V  V  F  D* T  G  S  S    93
361  aatctgtgggttcctagcaaaaaatgtcatttgaccaatattgcttgcctcatgcataat  420
94    N  L  W  V  P  S  K  K  C  H  L  T  N  I  A  C  L  M  H  N   113
421  aaatacgatgccaccaagtccaagacttttaagcaaaacggtacggaatttgccatacat  480
114   K  Y  D  A  T  K  S  K  T  F  K  Q •N  G  T  E  F  A  I  H   133
481  tatggatctggtagtctgtctggttacttgtctacggatacggttaacattggcggcttg  540
134   Y  G  S  G  S  L  S  G  Y  L  S  T  D  T  V  N  I  G  G  L   153
541  gatattaaggaccagacatttgccgaagctttgagtgaacccggcttggtctttgttgct  600
154   D  I  K  D  Q  T  F  A  E  A  L  S  E  P  G  L  V  F  V  A   173
601  gccaaatttgatggtattttgggcttgggttacagttcaatatcggtggatggtgttaag  660
174   A  K  F  D  G  I  L  G  L  G  Y  S  S  I  S  V  D  G  V  K   193
661  ccaccattctatgctatgtatgaacagggtttgatttcgcaaccaattttctcattctat  720
194   P  P  F  Y  A  M  Y  E  Q  G  L  I  S  Q  P  I  F  S  F  Y   213
721  ttaaatcgcgatcccaaagccccagaagggggtgaaatcattttcggtggttctgatcct  780
214   L  N  R  D  P  K  A  P  E  G  G  E  I  I  F  G  G  S  D  P   233
781  gatcactataccggtgattttacatatttgccagtaacacgcaaggcttattggcaaatc  840
234   D  H  Y  T  G  D  F  T  Y  L  P  V  T  R  K  A  Y  W  Q  I   253
841  aaaatggactcggcatcaatgggtgatttgaaatgtgccaaaggaggctgtcaagtaatt  900
254   K  M  D  S  A  S  M  G  D  L  K  C  A  K  G  G  C  Q  V  I   273
901  gctgataccggcacctctctcattgctttacctcccagtgaagctacatctattaaccaa  960
274   A  D* T  G  T  S  L  I  A  L  P  P  S  E  A  T  S  I  N  Q   293
961  gctattggtggaactccgattatgggtggtcaatatatggtggcctgtgaggatatcccc 1020
294   A  I  G  G  T  P  I  M  G  G  Q  Y  M  V  A  C  E  D  I  P   313
1021 aagttgcctgttattaaatttgttcttggtggcaaaacatttgagttggaaggtaaagac 1080
314   K  L  P  V  I  K  F  V  L  G  G  K  T  F  E  L  E  G  K  D   333
1081 tatgtattgcgtattgctcaaatgggtaaaacaatttgcttgtccggtttcatgggcatc 1140
334   Y  V  L  R  I  A  Q  M  G  K  T  I  C  L  S  G  F  M  G  I   353
1141 gacattccaccacccaatggacccctatggattttgggagatgttttcattggcaaatac 1200
354   D  I  P  P  P  N  G  P  L  W  I  L  G  D  V  F  I  G  K  Y   373
1201 tatacagaatttgatatgggcaatgaccgtgttggttttgccattgccaaataagcagca 1260
374   Y  T  E  F  D  M  G  N  D  R  V  G  F  A  I  A  K  *         390
1261 gctgtctgtgaatccaccagaagtgaagtttctttgttaattttatcaaatgttttgaat 1320
1321 tttatttgtattggtaatgtgtatgatggctttgttaaaagtaaatgtaaaatgatggac 1380
1381 tgtatcgtattaaatgcaaaacaacaacaataaagattaaatttgatcaatttaaaaaa  1439
```

*Figura 9. Sequência de aminoácidos deduzida a partir da sequência de nucleotídeos da ppCAD1. O peptídeo sinal está em itálico. O sinal de poliadenilação (AATAAA) está em vermelho. Os aminoácidos (D*TG) do sítio catalítico conservado estão em verde. A sequência consenso reconhecida pelo primer degenerado (CADdeg) está sublinhada. O sítio de N-glicosilação está marcado pelo símbolo (•).*

4.1.2.3. ppCAD3

A sequência de cDNA que contém a *ORF* codificante para a ppCAD3 (preproCAD3) mostra que esta possui 1397 pares de bases. O códon da metionina que inicia a transcrição do gene encontra-se na posição (31 – 33) e o códon de término de transcrição na posição (1168 –1170). Este cDNA possui uma *ORF* que localiza-se entre a base 31 e 1170, codificando 379 aminoácidos massa molecular teórica de 41,3 kDa (Figura 11). O peptídeo sinal foi identificado entre os resíduos Ala^{17} e Asp^{18} da ppCAD3, que resulta em uma proenzima de 362 aminoácidos tendo um ponto isoelétrico (pI) e massa molecular (MM) teórica de 4,25 e 39,67 kDa, respectivamente. A análise da sequência de aminoácidos revela que a enzima possui os dois ácidos aspárticos catalíticos (Asp^{87} e Asp^{272}) característicos do único sítio ativo das aspártico proteinases (Tabela 3). Análise do gene transcrito através de BLASTX no GENBANK produziu uma *ORF* revelando ser uma catepsina-D com alta identidade com aspártico proteinases de artrópodes e eucariotos. Essa análise mostra que a ppCAD3 possui 40% de identidade com a sequência de aminoácidos de uma catepsina-D de *Bos taurus* (AAB26186.1), 49% com uma catepsina-D de *Aedes aegypti* (XP_001657556.1), 42% de identidade com uma sequência de pepsinogênio de *Homo sapiens* (AAA60061.1) e por fim 45 % e 53 % de identidade com a ppCAD1 (ABM69086) e ppCAD2 (ABM69085) de *Musca domestica,* respectivamente.

```
1    taaaaaatgtggaaattgtgtgttgtgctatggagtgcgctccttttagccgaggccacg 60
1          M  W  K  L  C  V  V  L  W  S  A  L  L  L  A  E  A  T  18
61   ctcgttcaagttcccataactaaagttaaggaaaccaaaagcaaggccaatgaaatacgt 120
19    L  V  Q  V  P  I  T  K  V  K  E  T  K  S  K  A  N  E  I  R  38
121  aaattgaaggccaaatacggtggcactcccaaggcggaaatcagagatttggtggtggaa 180
39    K  L  K  A  K  Y  G  G  T  P  K  A  E  I  R  D  L  V  V  E  58
181  aaactcttcaactatgtcgatgactcgtactatggaaaaattaccattggcacacctggt 240
59    K  L  F  N  Y  V  D  D  S  Y  Y  G  K  I  T  I  G  T  P  G  78
241  caagaattcttggttttgttcgataccggctcctccaatttgtgggtccccgtagctccc 300
79    Q  E  F  L  V  L  F  D* T  G  S  S  N  L  W  V  P  V  A  P  98
301  tgttctgccgacaatgctgcctgtgaaaatcacaatacctacgatcccagtgcctcgtcc 360
99    C  S  A  D  N  A  A  C  E  N  H  N  T  Y  D  P  S  A  S  S  118
361  actcacgtcaagaaaggtgaatccttttccattcaatatggatccggtagcttgagtggt 420
119   T  H  V  K  K  G  E  S  F  S  I  Q  Y  G  S  G  S  L  S  G  138
421  tatctggttgaagataccgttgatgttgaaggtttgaagattaagaaacaagtctttgct 480
139   Y  L  V  E  D  T  V  D  V  E  G  L  K  I  K  K  Q  V  F  A  158
481  gctgccaccaatgaacctggcgagaccttcgtctatgctcccttcgatggtatcatgggt 540
159   A  A  T  N  E  P  G  E  T  F  V  Y  A  P  F  D  G  I  M  G  178
541  atgggcttcaagtcgattgctgttgatgatgtcactccaccatggtacaacatgatcagc 600
179   M  G  F  K  S  I  A  V  D  D  V  T  P  P  W  Y  N  M  I  S  198
601  caacatttgatttccgaaaaggtattctctttctacttggcccgtcgtggtacctctgat 660
199   Q  H  L  I  S  E  K  V  F  S  F  Y  L  A  R  R  G  T  S  D  218
661  gagggtggtgttatggttgtgggtggcaatgatgatcgttattatgaaggcgatttccac 720
219   E  G  G  V  M  V  V  G  G  N  D  D  R  Y  Y  E  G  D  F  H  238
721  tatgttcctgtctccgaacagggttactggcaatttgaaatggctgaagcccatgtcaat 780
239   Y  V  P  V  S  E  Q  G  Y  W  Q  F  E  M  A  E  A  H  V  N  258
781  ggtgttagaatttgtgaccgttgccaagctattgctgatactggtacctccctcatcgcc 840
259   G  V  R  I  C  D  R  C  Q  A  I  A  D* T  G  T  S  L  I  A  278
841  gtgcccacagacaaatacgaagaaatccaaaaggaaatcggtgccacctttagctatgac 900
279   V  P  T  D  K  Y  E  E  I  Q  K  E  I  G  A  T  F  S  Y  D  298
901  acctacgaatatatgttggattgttcgaagatcgatgatttgcccgttgttaccttccgt 960
299   T  Y  E  Y  M  L  D  C  S  K  I  D  D  L  P  V  V  T  F  R  318
961  ttgggcgatggtaccttcaccttggaaggtcgtgattatgttattaaatccgatgacaac 1020
319   L  G  D  G  T  F  T  L  E  G  R  D  Y  V  I  K  S  D  D  N  338
1021 caatgctcttctgcctttgaggatggtggtactgatttctggattttgggtgatgtcttc 1080
339   Q  C  S  S  A  F  E  D  G  G  T  D  F  W  I  L  G  D  V  F  358
1081 attggcaaatactacacaacattcgatgctgaacataatcgtgttggttttgctttggct 1140
359   I  G  K  Y  Y  T  T  F  D  A  E  H  N  R  V  G  F  A  L  A  378
1141 aagtaaagcattgcattgcatg 1162
379   K  *
```

*Figura 10. Sequência de aminoácidos deduzida a partir da sequência de nucleotídeos da ppCAD2. O peptídeo sinal está em itálico. Os aminoácidos (D*TG) do sítio catalítico conservado estão em verde. A sequência consenso reconhecida pelo primer degenerado (CADdeg) está sublinhada.*

```
1    gcacgaggattaaatcaaaagtgcttaataatgttaaaatcgttgggtatgttagccata 60
1                                     M  L  K  S  L  G  M  L  A  I  10
61   gtcttggctgtggccaccgctgatgtggttcgtgttccaattcacaggcatgaaaacttt 120
11    V  L  A  V  A  T  A  D  V  V  R  V  P  I  H  R  H  E  N  F  30
121  gtcaagaccagcaaggatatacgttcggaaaaggctgtcttgcgcagcaagtataatttg 180
21    V  K  T  S  K  D  I  R  S  E  K  A  V  L  R  S  K  Y  N  L  50
181  ccccagccccgtgatgtctccgatgagcctttggaaaattccttgaatatgaaatattat 240
41    P  Q  P  R  D  V  S  D  E  P  L  E  N  S  L  N  M  K  Y  Y  70
241  ggtgacataaccattggtacaccacctcagaaatttgttgttttgttcgacaccggctcc 300
61    G  D  I  T  I  G  T  P  P  Q  K  F  V  V  L  F  D* T  G  S  90
301  tccaatttgtgggttccctctagccactgctggatctgggatattgcctgcaagaaacac 360
81    S  N  L  W  V  P  S  S  H  C  W  I  W  D  I  A  C  K  K  H  110
361  aaccaatataaccatgacgattccagcacctatgtaaagaacggcgaacttatctccatc 420
101   N  Q  Y  N  H  D  D  S  S  T  Y  V  K  N  G  E  L  I  S  I  130
421  tcttatggctccggcagtatgagcggtttcttgtcccaggatgatgttaccgttgaagga 480
121   S  Y  G  S  G  S  M  S  G  F  L  S  Q  D  D  V  T  V  E  G  150
481  ttgactattaagaaccaagtattcgctgaagccatgaacgaacccggtaacagtttcact 540
141   L  T  I  K  N  Q  V  F  A  E  A  M  N  E  P  G  N  S  F  T  170
541  gatgccaatttcgatggtatttttggcatggcttaccaatctttggcggaggacaatgtt 600
161   D  A  N  F  D  G  I  F  G  M  A  Y  Q  S  L  A  E  D  N  V  190
601  gtgcctcccttctacaatatgttcgctcagggcttggtcgatgccaatatgttctcattc 660
181   V  P  P  F  Y  N  M  F  A  Q  G  L  V  D  A  N  M  F  S  F  210
661  cttttgaaccgtgatggcacctctaccgatggtggtcaaatgattttgggtggtgttgat 720
201   L  L  N  R  D  G  T  S  T  D  G  G  Q  M  I  L  G  G  V  D  230
721  tccagcttgtataccggcgatattacctatgttcccgtttcttctcaaggctattggcaa 780
221   S  S  L  Y  T  G  D  I  T  Y  V  P  V  S  S  Q  G  Y  W  Q  250
781  ttcgaagttacctctggtgccattaagggtcaaagcatttgtgataactgccaagctatt 840
241   F  E  V  T  S  G  A  I  K  G  Q  S  I  C  D  N  C  Q  A  I  270
841  gctgataccggcaccagcttgattgtggctccttccgatgcttacaacaccctgaacgcc 900
261   A  D* T  G  T  S  L  I  V  A  P  S  D  A  Y  N  T  L  N  A  290
901  gaaatcggtgccacttacaatgaggatgatggcaactactatgttgattgttccgccgtt 960
281   E  I  G  A  T  Y  N  E  D  D  G  N  Y  Y  V  D  C  S  A  V  310
961  gatagcttgcccgatgtcactttcgtcattggtggtactaccttcactttgcctgccagt 1020
301   D  S  L  P  D  V  T  F  V  I  G  G  T  T  F  T  L  P  A  S  330
1021 gcctacattgttaccgtcgatggcaattgcatgtcttcctttacctacatgggcactgat 1080
321   A  Y  I  V  T  V  D  G  N  C  M  S  S  F  T  Y  M  G  T  D  350
1081 ttctggattttgggtgatgtcttcattggcaaatactacaccgtcttcgatttcgccaat 1140
341   F  W  I  L  G  D  V  F  I  G  K  Y  Y  T  V  F  D  F  A  N  370
1141 aaccgtgttggctttgctgaagctgtctaaaatgagtgaggggaagaaaaatcaaaaaat 1200
361   N  R  V  G  F  A  E  A  V  *                                379
1201 ccgttttagcctgacctgttaagtttgccgtcatgtctgaatgtcgtgasatgtgcatgg 1260
1261 ctgaatgcatgcatgtcctgctgctgtgtaatggtcatgtgtcatgtatatagctatgct 1320
1321 gacaataaagctaatgctaaagtccagttagttagttaggaaaaaaaaaaaaaaaaaaaa 1380
1381 aaaaaaaaaaaaaaaaa                                            1397
```

*Figura 11. Sequência de aminoácidos deduzida a partir da sequência de nucleotídeos da ppCAD3. O peptídeo sinal está em itálico. O sinal de poliadenilação (AATAAA) está em vermelho. Os aminoácidos (D*TG) do sítio catalítico conservado estão em verde. A sequência consenso reconhecida pelo primer degenerado (CADdeg) está sublinhada.*

4.1.3. Clonagem, sequênciamento e expressão da pCAD1 recombinante

A sequência de cDNA codificante para a pCAD1 foi amplificada através de uma reação de polimerização em cadeia (PCR), usando como molde o plasmídeo pGEM-T (contendo o cDNA codificante para a pCAD1), juntamente com os oligonucleotídeos pCAD1Nt_BamHI e pCAD1Ct_EcoRI (ver item 2.11), contendo os respectivos sítios de restrição para a endonuclease de restrição. O produto do PCR foi digerido com as enzimas *BamHI* e *EcoRI* e posteriormente clonado no vetor pAE (Ramos *et al.,* 2004) previamente digeridos com as mesmas enzimas, resultando no plasmídeo pAE/pCAD1 (Figura 12).

A pCAD1 recombinante terá em seu N-terminal um peptídeo de fusão que contém 6 resíduos de histidina, (massa molecular teórica total 40,21 kDa) para facilitar sua purificação por cromatografia de afinidade em coluna de níquel.

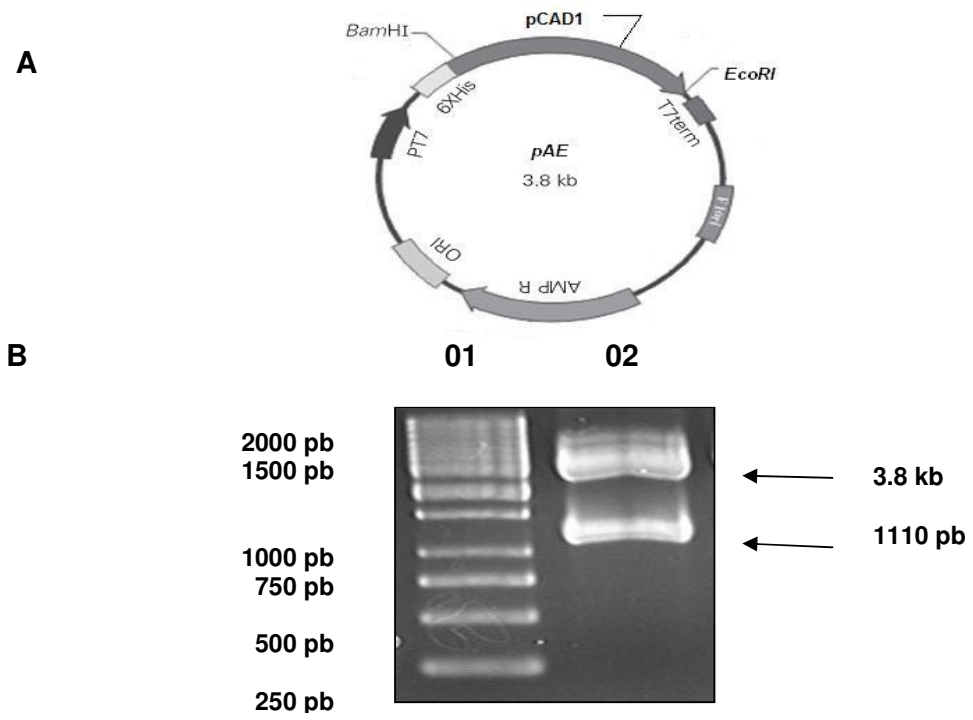

Figura 12. A. Esquema da construção pAE/pCAD1 que consiste do plasmídeo pAE (Ramos et al., 2004) com o cDNA que codifica a pCAD1 inserido. B. Eletroforese em gel de agarose 1% da digestão da construção pAE/pCAD1 com enzimas de restrição BamHI e EcoRI. 01. Padrão de DNA, 02. produto da digestão.

A construção pAE/pCAD1 foi usado para transformar bactérias Origami(DE3). As bactérias transformadas foram crescidas a 37ºC em 150mL de LB líquido contendo ampicilina 50µg/mL até uma Abs_{600} entre 0,6 -1,0. Essas bactérias foram divididas em 3 frascos (50mL de bactérias cada) e induzidas com IPTG 0,5mM. Desses 3 frascos, 2 foram induzidos por 4 horas, um a 37ºC e o outro a 26ºC sendo que o terceiro frasco induzido por 19 horas a 20ºC. As proteínas das células controle, induzidas e não induzidas foram analisadas em SDS-PAGE (Figura 13). Nota-se que a indução com IPTG levou a intensa produção de um proteína de massa molecular de 42 kDa. Para verificarmos se a proteína recombinante apresentava-se solúvel, células induzidas nas 3 diferentes temperaturas (37º, 26ºC e 20ºC) foram lisadas por sonicação sendo que o lisado foi submetido a centrifugação a 7000g por 30 minutos. O sobrenandante recuperado corresponde ao material solúvel enquanto que o precipitado ao material insolúvel bacteriano.

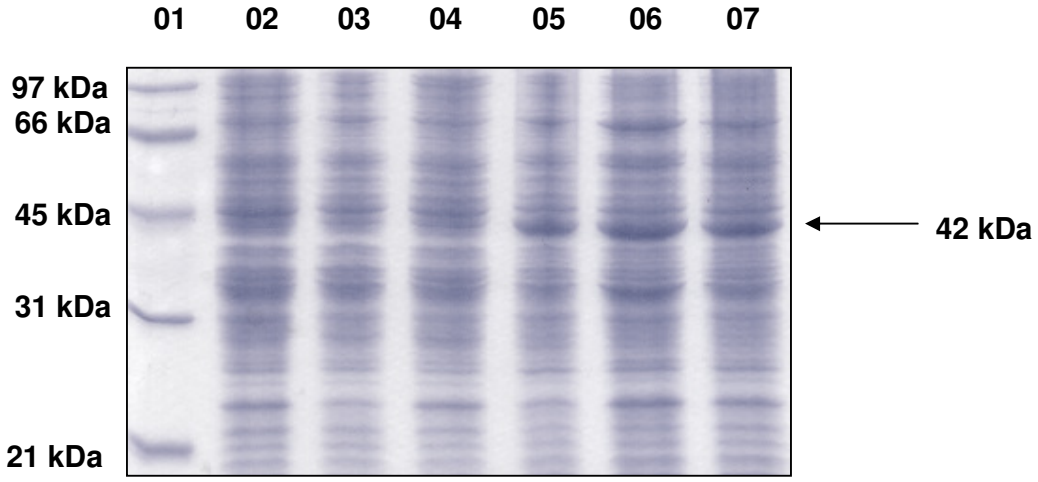

Figura 13. Análise em SDS-PAGE da expressão da pCAD1 com o plasmídeo pAE em células Origami(DE3). A indução da expressão foi feita com IPTG 0,5 mM em três diferentes temperaturas (37ºC e 26ºC por 4 horas e 20ºC por 19 horas). Temos nas seguintes raias: 01. padrão de massa molecular, 02. pAE não induzido a 37ºC, 03. pAE induzido a 37ºC, 04. pAE/pCAD1 não induzido a 37ºC 05. pAE/pCAD1 induzido a 37ºC, 06. pAE/pCAD1 induzido a 26ºC, 07. pAE/pCAD1 induzido a 20ºC. O gel continha poliacrilamida (12%) e foi corado com Coomassie Blue R-250.

Conforme pode ser visto na figura 14, o SDS-PAGE revela que a proteína recombinante produzida (pCAD1) é recuperada parcialmente na forma solúvel na temperatura de indução a 20ºC por um periodo de tempo de indução de 19 horas. A fração solúvel do lisado bacteriano induzido contendo a pCAD1 recombinante foi injetado em uma coluna cromatográfica de afinidade (HisTrap HP, 1 mL, GE), que possui afinidade pela cauda de histidina fusionada a extremidade N-terminal da pCAD1. A proteína recombinante foi então eluída na presença de um tampão B contendo alta concentração de imidazol (ver item 2.14.). Uma alíquota das frações coletadas foram analisada em SDS-PAGE e mostrou que a enzima recombinante pCAD1 apresentava-se na forma pura (Figura 15). A pCAD1 foi submetida a uma série de experimentos de auto-ativação, porém foi incapaz de hidrolizar hemoglobina, apresentando-se sem atividade nesse sistema de expressão.

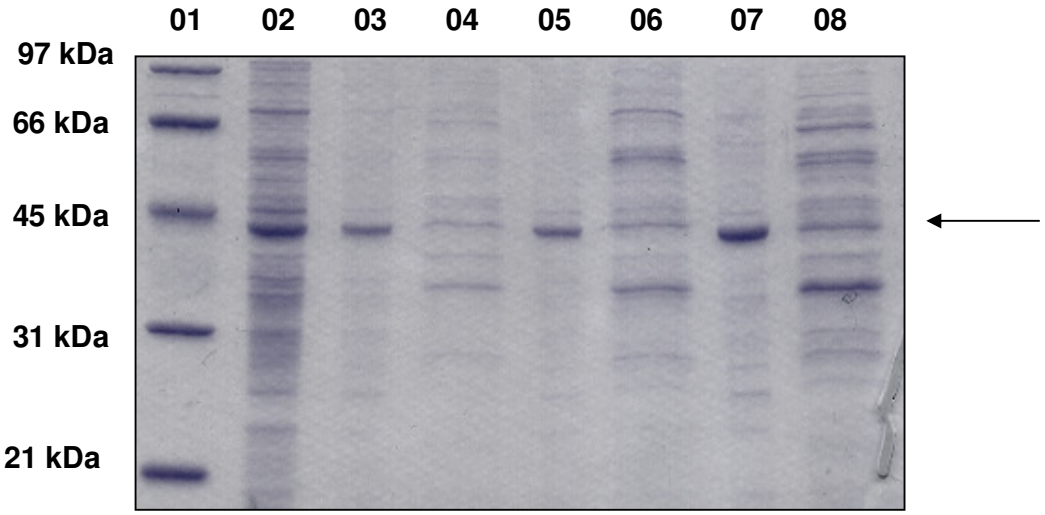

Figura 14. Análise em SDS-PAGE (12%) das bactérias Origami(DE3) contendo a contrução pAE/pCAD1 induzidas por IPTG 0,5 mM. Temos nas seguintes raias: 01. padrão de massa molecular, 02. Bactérias induzidas a 37°C por 4 horas, 03. precipitado das bactérias induzidas a 37°C por 4 horas, lisadas e centrifugadas, 04. sobrenadante das bactérias induzidas a 37° por 4 horas lisadas e centrifugadas. 05. precipitado das bactérias induzidas a 26° por 4 horas, lisadas e centrifugadas, 06. sobrenadante das bactérias induzidas a 26° por 4 horas, lisadas e centrifugadas, 07. precipitado das bactérias induzidas a 20° por 19 horas, lisadas e centrifugadas,

08. sobrenadante das bactérias induzidas a 20° por 19 horas, lisadas e centrifugadas. A seta indica a presença da pCAD1 recombinante na fração solúvel do lisado bacteriano.

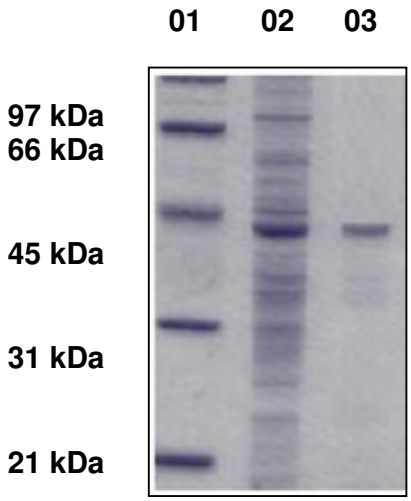

Figura 15. Análise em SDS-PAGE das frações da cromatografia de afinidade para purificação da pCAD1 recombinante. Esta enzima foi expressa com o peptídeo de fusão ligado a extremidade N-terminal da proteína utilizando o vetor pAE e foi purificada através de uma cromatografia de afinidade em coluna de Ni^{2+} (HisTrap HP, 1 mL, GE) Raia 01. padrão de massa molecular, 02. bactérias induzidas (pAE/pCAD1/Origami(DE3)); 03. pCAD1 purificada. O gel continha poliacrilamida (12%) e foi corado com Coomassie Blue R-250.

4.1.4. Clonagem, sequênciamento e expressão da pCAD3 recombinante

A sequência de cDNA codificante para a pCAD3 foi amplificada através de uma reação de polimerização em cadeia (PCR) usando como molde o plasmídeo pGEM-T (contendo o cDNA codificante para a pCAD3) juntamente com os oligonucleotídeos pCAD3Nt_BamHI e pCAD3Ct_HindIII (ver item 2.11.). O produto do PCR foi digerido com as enzimas *BamHI* e *Hind III* e posteriormente clonado no vetor pAE (Ramos *et al.,* 2004) previamente digeridos com as mesmas enzimas, resultando no plasmídeo pAE/pCAD3 (Figura 16).

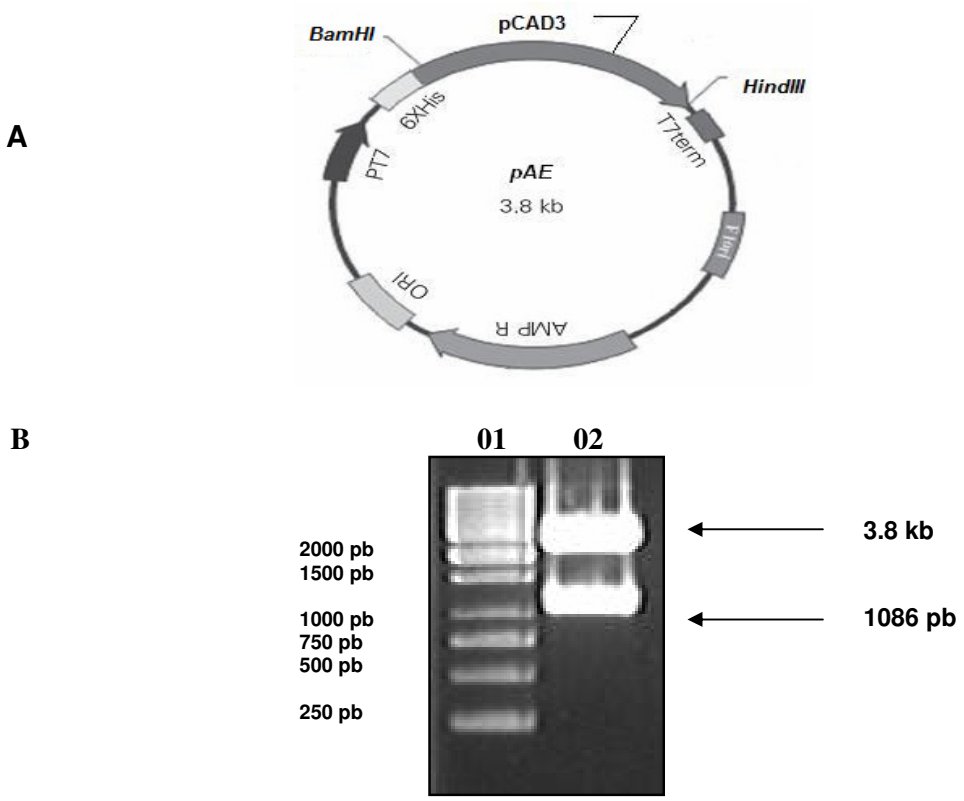

Figura 16. A. Esquema da construção pAE/pCAD3 que consiste do plasmídeo pAE (Ramos et al., 2004) com o cDNA que codifica a pCAD3 inserido. B. Eletroforese em gel de agarose 1% da digestão da construção pAE/pCAD3 com enzimas de restrição BamHI e HindIII. 01. Padrão de DNA, 02. produto da digestão

A clonagem nesse vetor está em fase com um peptídeo (codificado pelo plasmídeo) composto por 9 resíduos de aminoácidos contendo entre eles 6 resíduos de histidina adjacentes. Portanto a pCAD3 recombinante terá em seu N-terminal um peptídeos de fusão que contém 6 resíduos de histidina, tendo um massa molecular teórica (proenzima + peptídeo de fusão) de 40,86 kDa, na qual facilitará sua purificação por cromatografia de afinidade em coluna de níquel. A construção pAE/pCAD3 foi usada para transformar bactérias Origami(DE3), (que produzem T7 RNA polimerase de forma indutível por IPTG) que tornam-se resistentes a ampicilina podendo produzir de forma induzida a proteína recombinante. As bactérias transformadas foram crescidas a 37ºC em 150 mL de LB líquido contendo ampicilina 50µg/mL até uma Abs_{600} 0,6 -1,0. Essas bactérias

foram divididas em 3 frascos (50mL de bactérias cada) e induzidas com IPTG 0,5 mM. Desses 3 frascos, 2 foram induzidos por 4 horas, um a 37ºC e o outro a 26ºC sendo que o terceiro frasco induzido por 16 horas a 20ºC. As proteínas das células, induzidas e não induzidas e os controles foram analisadas em SDS-PAGE (Figura 17).

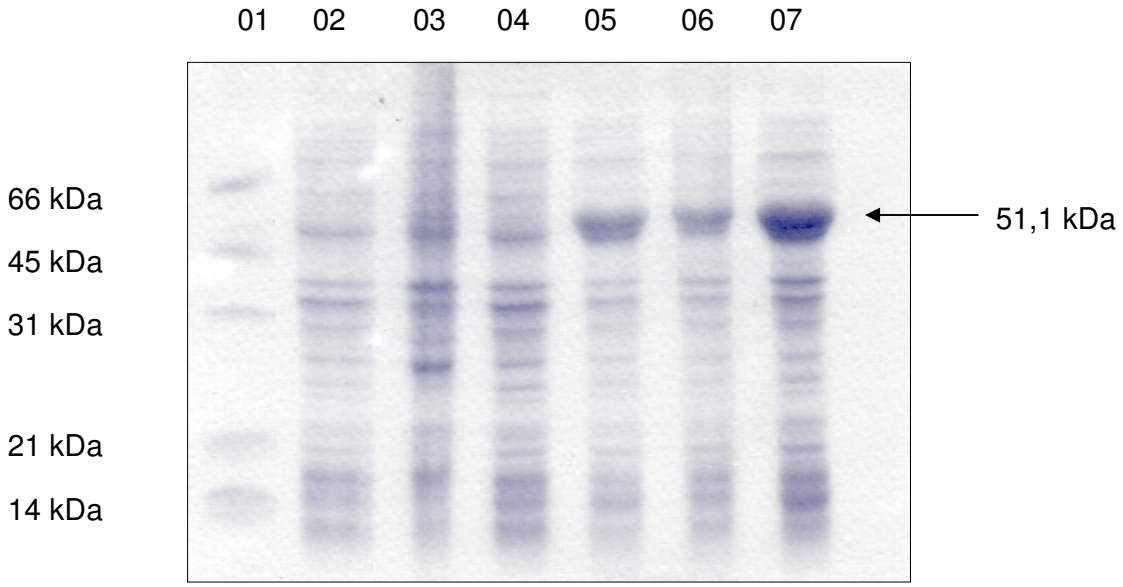

Figura 17. Análise em SDS-PAGE da expressão da pCAD3 com o plasmídeo pAE em células Origami(DE3). A indução da expressão foi feita com IPTG 0,5 mM em três diferentes temperaturas (37ºC e 26ºC por 4 horas e 20ºC por 16 horas). Temos nas seguintes raias: 01. padrão de massa molecular, 02. pAE não induzido a 37ºC, 03. pAE induzido a 37ºC, 04. pAE/pCAD3 não induzido a 37ºC 05. pAE/pCAD3 induzido a 37ºC, 06. pAE/pCAD3 induzido a 26ºC, 07. pAE/pCAD3 induzido a 20ºC. O gel continha poliacrilamida (12%) e foi corado com Coomassie blue R-250.

Nota-se que a indução com IPTG levou a intensa produção de um proteína de massa molecular de 51,1 kDa que deve corresponder a proteína pCAD3. Portanto a linhagem pCAD3/Origami(DE3), parece ser um sistema de expressão adequado para a pCAD3. Porém é sabido que muitas proteínas produzidas em *E. coli* acabam por se tornarem-se insolúveis formando corpos de inclusão. Como para os objetivos deste projeto é preciso que a proteína recombinante esteja solúvel e ativa, tornou-se necessário verificar a solubilidade da proteína recombinante produzida. Para isso células induzidas na temperatura de 20ºC por 16 horas foram lisadas por sonicação

sendo que o lisado foi submetido a centrifugação a 7000g por 30 minutos. O sobrenandante recuperado corresponde ao material solúvel, enquanto que o precipitado ao material insolúvel bacteriano. Conforme pode ser visto na figura 18, o SDS-PAGE revela que a proteína recombinante produzida (pCAD3) é recuperada na forma parcialmente solúvel quando induzida a temperatura de 20ºC por um periodo de tempo de 16 horas.

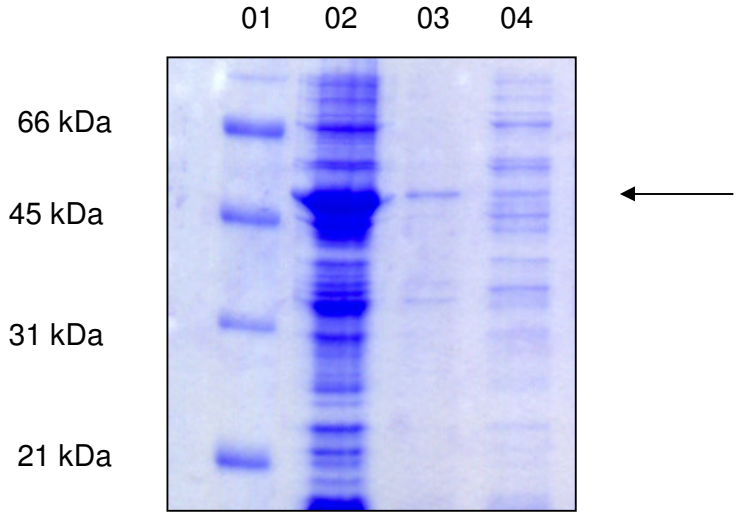

Figura 18. Análise em SDS-PAGE das bactérias Origami(DE3) contendo a contrução pAE/pCAD3 induzidas por IPTG 0,5 mM, a uma temperatura de 20ºC durante 16 horas. As bactérias induzidas foram lisadas e centrifugadas sendo que o precipitado e o sobrenadante foram analisados em SDS-PAGE (12%). Temos nas seguintes raias: 01. padrão de massa molecular, 02. Bactérias induzidas (pAE/pCAD3/Origami(DE3)), 03. precipitado das bactérias induzidas, lisadas e centrifugadas, 04. sobrenadante das bactérias induzidas, lisadas e centrifugadas. A seta indica a presença da pCAD3 recombinante na fração solúvel do lisado bacteriano.

A fração solúvel do lisado bacteriano induzido contendo a pCAD3 recombinante foi injetado em uma coluna cromatográfica de afinidade (HisTrap HP, 1 mL, GE) que possui afinidade pela cauda de histidina fusionada a extremidade N-terminal da pCAD3. A proteína recombinante foi então eluída na presença do tampão B contendo alta concentração de imidazol (ver item 2.14.) sendo que uma alíquota das frações coletadas foi utilizada em um ensaio enzimático tendo como

substrato hemoglobina desnaturada dissolvida em tampão citrato-fosfato de sódio 100 mM pH, 3,5 (ver item 2.16.) (Figura 19).

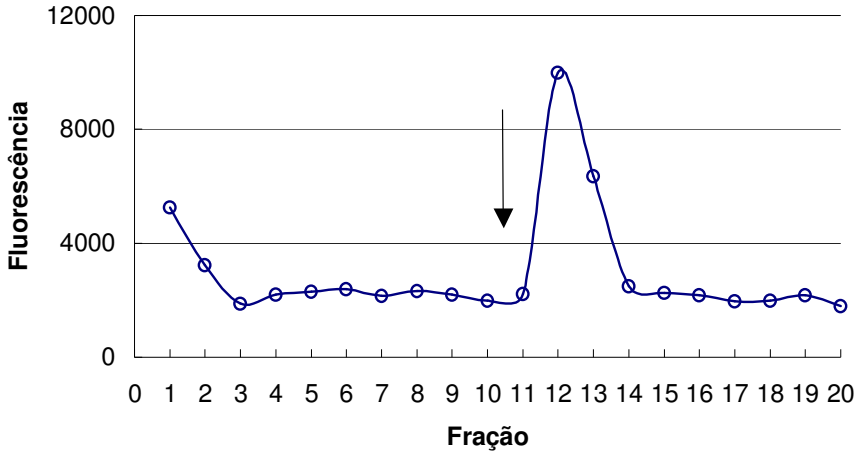

Figura 19. Cromatografia de afinidade em coluna HisTrap HP, 1 mL, GE, utilizando a fração solúvel do lisado de bactérias Origami(DE3) induzidas transformadas com o vetor pAE (Ramos et al., 2004) codificando a pCAD3. A atividade recombinante foi detectada utilizando hemoglobina desnaturada em tampão citrato-fosfato de sódio 100 mM, pH 3,5 onde o produto da hidrólise foi detectado por fluorescamina e medido em unidades de fluorescência. A seta indica o início da eluição das proteínas ligadas especificamente a coluna, utilizando o tampão B contendo imidazol 500 mM (ver item 2.14.).

Frações da cromatografia que apresentaram atividade recombinante foram reunidas e dialisadas em tampão fosfato de sódio 20 mM pH 7,0 por um período de tempo de 18 horas em câmara fria, visando a eliminação do tampão B contendo imidazol (pH 7,4; 0,5M). Posteriormente, uma alíquota da amostra foi analisada em SDS-PAGE, mostrando que a enzima recombinante pCAD3 apresentava-se na forma pura (Figura 20).

4.1.4.1. Auto-ativação da pCAD3 recombinante

A proenzima (pCAD3) previamente purificada apresenta uma auto-ativação máxima em valores de pH entre 3-4 e completa após 30 minutos a 30°C. Esta

reação de auto-ativação foi interrompida pela adição de pepstatina A (10 μM). Uma alíquota dessa preparação contendo a CAD3 madura foi analisada em SDS-PAGE (Figura 21).

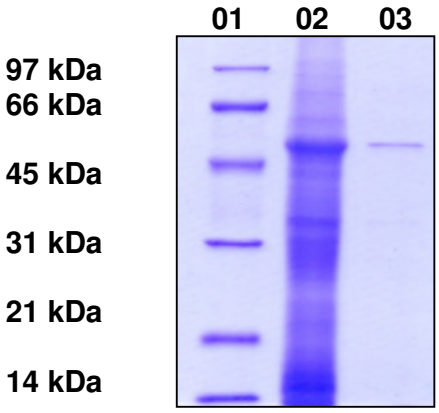

Figura 20. Análise em SDS-PAGE das frações da cromatografia de afinidade para purificação da pCAD3 recombinante. Esta enzima foi expressa com o peptídeo de fusão ligado a extremidade N-terminal da proteína utilizando o vetor pAE e foi purificada através de uma cromatografia de afinidade em coluna de Ni^{2+} (HisTrap HP, 1 mL, GE). Raia 01. padrão de massa molecular, 02. bactérias induzidas (pAE/pCAD3/Origami(DE3)), 03. pCAD3 recombinante purificada. O gel continha poliacrilamida (12%) e foi corado com Coomassie blue R-250.

A ativação é seguida pelo deslocamento da massa molecular de 51,1 para 46,9 kDa, correspondendo a perda do propeptídeo durante a auto-ativação *in vitro* (Figura 21). Embora a massa da pCAD3 e CAD3 quando determinada por SDS-PAGE diferem dos valores teóricos (Tabela 3), a diferenças entre eles (4,2 kDa) concorda com o deslocamento teórico (4,35 kDa). A massa molecular da pCAD3 (com a cauda de histidina) foi determinado por espectrometria de massa (*Ettan MALDI-TOF Pro*, Amerscham Biociences, na Unidade Analítica do Instituto de Química da USP, São Paulo, SP) como sendo de 41,09 kDa (dados não mostrados), que é bastante próximo do valor da massa molecular teórica.

A sequência N-terminal obtida por degradação de Edman da proteína CAD3 madura (sequência: R_{45}SKYNLPQP...) demonstrou que o sítio de hidrólise do propeptídeo *in vitro* encontra-se entre os resíduos Leu^{44} – Arg^{45}.

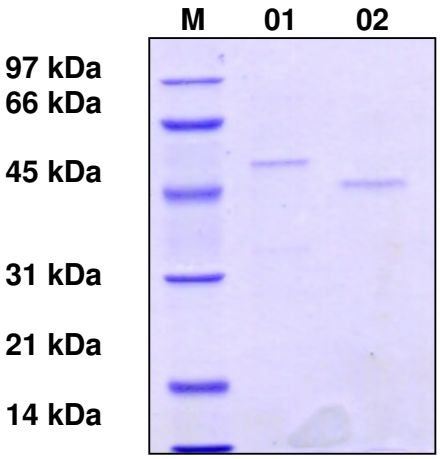

Figura 21. Análise em SDS-PAGE (12%) da pCAD3 antes (raia 01) e após (raia 02) a auto-ativação sob condições ácidas. Nota-se o deslocamento da massa molecular da enzima recombinante em função da remoção do propeptídeo in vitro. O gel foi corado com Coomassie Blue R-250.

4.1.4.2. Análise da atividade proteolítica da CAD3 recombinante

Pepstatina A (20µM), inibidor específico das aspártico proteinases foi usado para categorizar a atividade proteolítica recombinante da CAD3. Essa atividade foi determinada conforme descrito no item 2.16. com hemoglobina como substrato. Verificamos uma emissão de fluorescência constante com o passar do tempo nas reações contendo pepstatina A (20µM), indicando uma completa inibição da atividade proteolítica da CAD3 recombinante. Já a atividade proteolítica da CAD3 recombinante na ausência desse inibidor demonstra um aumento gradual durante o tempo de incubação a 30ºC (Figura 22).

4.1.4.3. Determinação do pH ótimo da CAD3

A atividade enzimática da CAD3 recombinante (madura) foi investigada em diferentes valores de pH, tendo como substrato hemoglobina bovina desnaturada e o substrato de fluorescência apagada MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2. A CAD3 apresentou uma atividade máxima em pH 2,5 com hemoglobina desnaturada e pH 3,0 com o substrato MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2 (Figura 23).

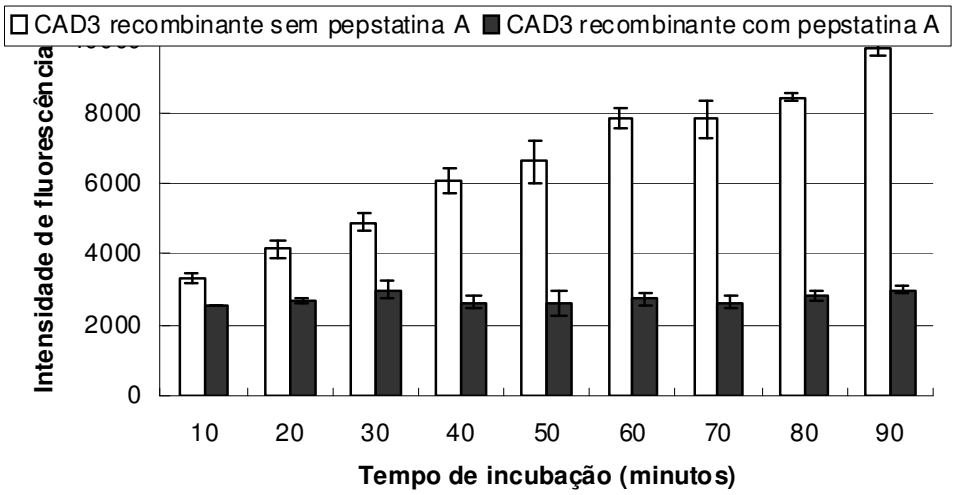

Figura 22: Perfil da hidrólise da hemoglobina desnaturada pela ação da CAD3 recombinante na presença e na ausência de pepstatina A (20 μM). Os dados foram plotados em unidades de fluorescência. Cada ponto representa a média aritmética de três reações.

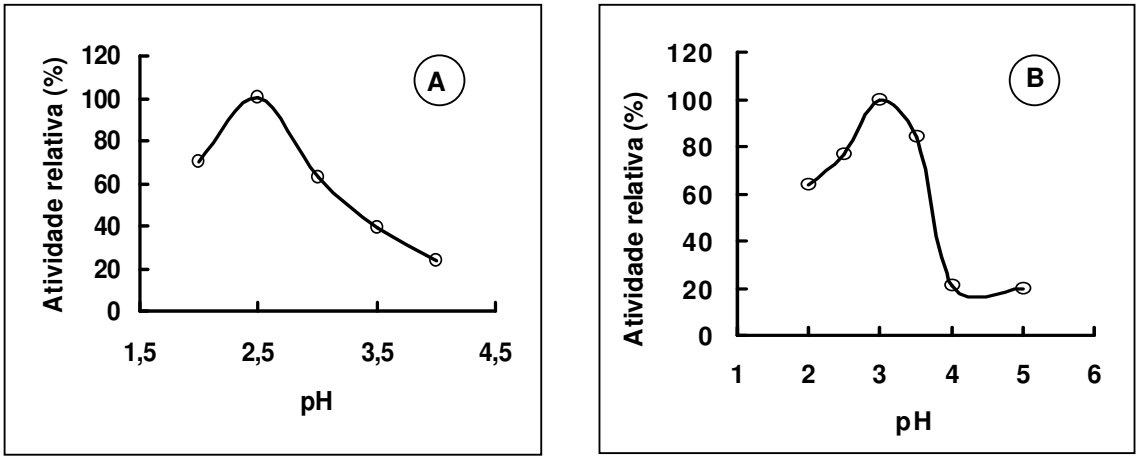

Figura 23. Efeito do pH sobre a atividade da CAD3 recombinante (madura) tendo como substrato hemoglobina desnaturada (A) e MOCAc-GKPILFFRLK-(Dnp)-D-R-NH_2 (B). Para cada valor de pH a atividade foi analisada a 30°C. A atividade relativa é expressa como porcentagem em relação a maior atividade obtida sobre a faixa de pH testada.

4.1.4.4. Análise da degradação da hemoglobina bovina pela CAD3 recombinante

A habilidade da CAD3 em hidrolisar hemoglobina bovina *in vitro* foi analisada em SDS-PAGE (Figura 24). O ensaio foi realizado em tampão citrato-fosfato de sódio 100 mM pH 2,5. A hemoglobina bovina migra com duas bandas mais intensas (setas grossas) representando o tetrâmero (~64 kDa) e o dímero (~32 kDa). O tetrâmero e o dímero foram completamente degradados pela CAD3 recombinante. A hidrólise da hemoglobina bovina foi inibida por pepstatina (20µM) (linha 03) onde leupeptina (20µM) e E-64 (20µM) (linha 04 e 05) não inibiu a enzima recombinante.

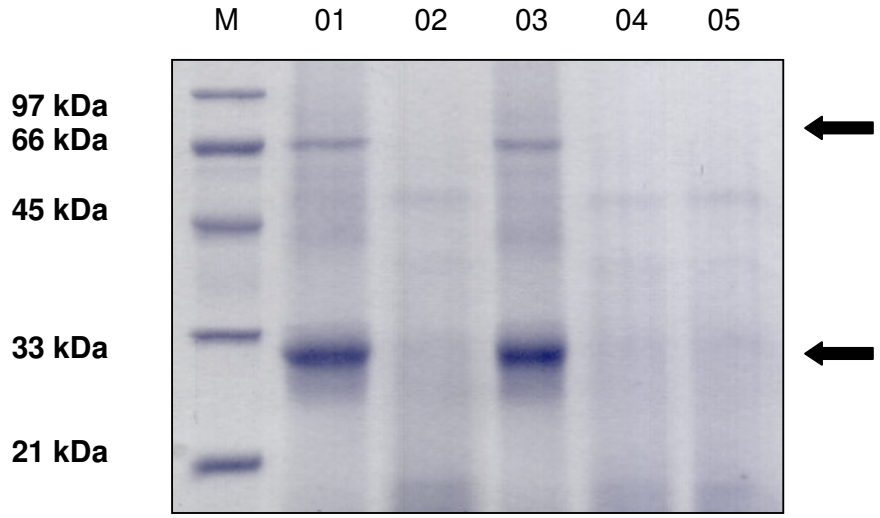

Figura 24. SDS-PAGE dos produtos gerados a partir de hemoglobina sob ação da CAD3 recombinante. Linha 01. Hb sozinha; Linha 02. CAD3 recombinante e Hb; Linha 03. CAD3 recombinante, Hb e pepstatina A (20 µM); Linha 04. Hb, CAD3 recombinante e leupeptina (20 µM); Linha 05. CAD3, Hb e E-64 (20 µM). As setas indicam a posição do tetrâmero (~64 kDa) e dímero (~32 kDa) da hemoglobina.

4.1.5. Análise filogenética, expressão e localização das pCAD1, 2 e 3

Um cladograma demonstrando a proximidade entre membros das aspártico proteinases foi gerado pelo pareamento da sequência de aminoácidos de várias aspártico proteinases, incluindo as ppCAD1, ppCAD2 e ppCAD3 com o programa MEGA 4 (Tamura, Dudley, Nei e Kumar 2007). As sequências utilizadas nessa análise foram: *Drosophila melanogaster* catepsina-D - 1 (número de acesso no GENBANK – NP_652013); *Drosophila melanogaster* catepsina-D - 2 (NP_609235);

Bombyx mori catepsina-D (AAP50847); *Aedes aegypti* aspártico proteinase (XP_001657556); *Anopheles gambiae* aspartico proteinase (XP_307784); *Tribolium castaneum* catepsina-D (XP_966517); *Apis melifera* catepsina–D (XP_392857); ppCAD3 catepsina-D (ABL84270), ppCAD2 catepsina-D (ABM69085) e ppCAD1 catepsina-D (ABM69086) de *Musca domestica*; *Canis familiaris* catepsina-D (NP_001020792); *Sus scrofa* catepsina-D (AAV90625); *Homo sapiens* catepsina-D (AAP36305); *Homo sapiens* pepsinogênio A (AAA60061); *Mus musculus* catepsina-D (NP_034113); *Ratus norvegegicus* catepsina-D (P24268); *Gallus gallus* catepsina-D (NP_990508) e *Podarcis sicula* catepsina-D (CAA08878) (Figura 25).

As sequências de aminoácidos de catepsinas-D de insetos e vertebrados formaram um grande grupo (Figura 25, *bootstrap* 89) onde apenas pepsinogênio A humana, ppCAD2 e ppCAD3 e a sequência de *D. melanogaster* – 2 foram excluídos. Esse grande grupo divide-se em duas ramificações monofiléticas: uma com sequências de catepsinas-D de vertebrados (*bootstrap* 100) e outra de insetos (*bootstrap* 99).

Análise por RT-PCR semi-quantitativo mostrou que os mRNAs codificantes para CAD1 foram expressos em todos os tecidos analisados (ventrículo anterior, médio e posterior, hemolinfa, corpo gorduroso e túbulos de Mapighi) (Figura 26).

Diferentemente, os mRNAs codificantes para as CAD3 e CAD2 foram expressos apenas no ventrículo anterior e porção proximal do médio onde atividade de aspártico proteinase similar a catepsina-D foram achadas no conteúdo luminal nas larvas de *Musca domestica* (Espinoza-Fuentes e Terra, 1987; Lemos e Terra, 1991). Essas duas últimas proteínas são óbvias candidatas a serem enzimas digestivas extracelulares desse inseto.

Adicionalmente, um Western blot após SDS-PAGE e revelado pelo método do kit *Imobbilon Western Chemiluminescent HRP substrate* (Millipore) mostrou que o anti-pCAD3 reconhece a proteína recombinante pCAD3 (zimógeno) e CAD3 (madura), juntamente com uma proteína de massa molecular semelhante a CAD3 (madura) presente nas proteínas do conteúdo luminal e epitélio do ventrículos anterior e médio em larvas de *Musca domestica* (Figura 27).

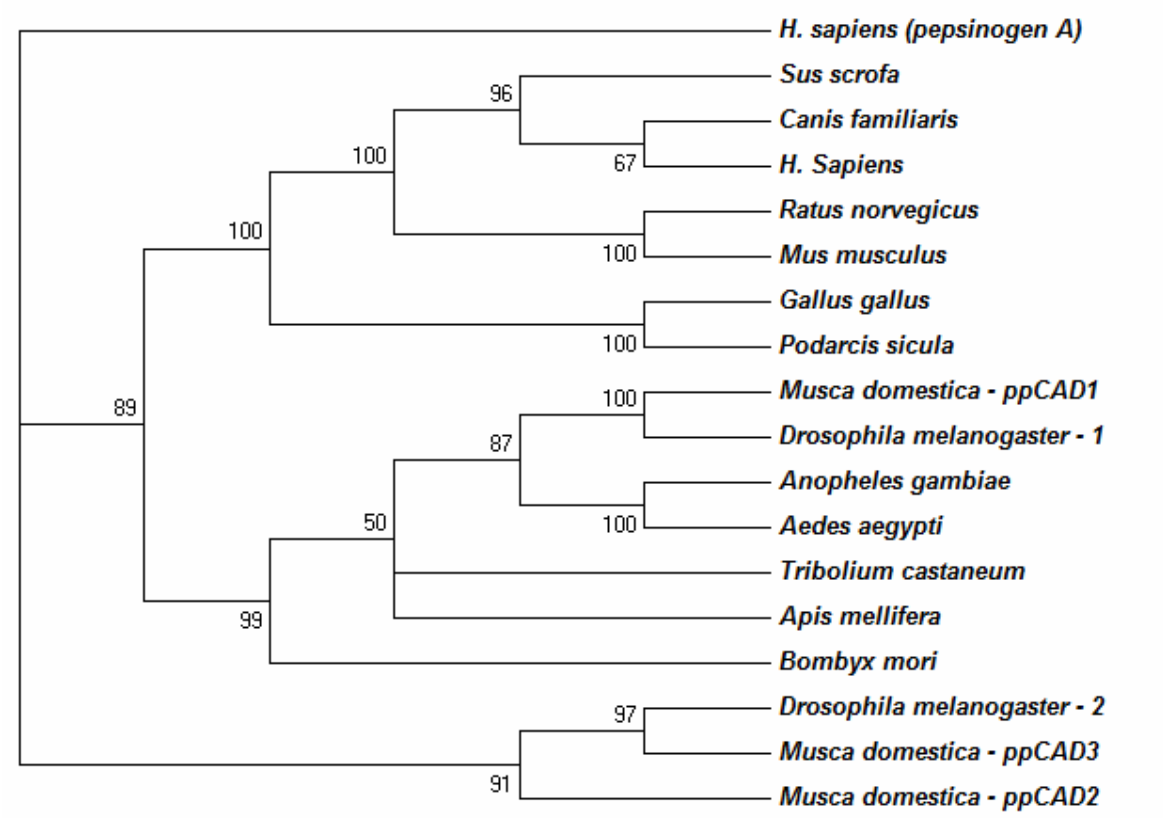

Figura 25. Análise filogenética das sequências de aspártico proteinases similares a catepsinas-D. O cladograma dessas enzimas foi gerado pela comparação do pareamento das sequências de aminoácidos deduzidas. Temos uma ramificação monofilética contendo a maioria das sequências, incluindo ppCAD1 e outras catepsinas-D lisossomais bem conhecidas. As sequências das enzimas ppCAD2, ppCAD3, pepsinogênio A e D. melanogaster-2 estão excluídas do ramo principal. Valores de bootstrap são baseados em 1000 replicatas.

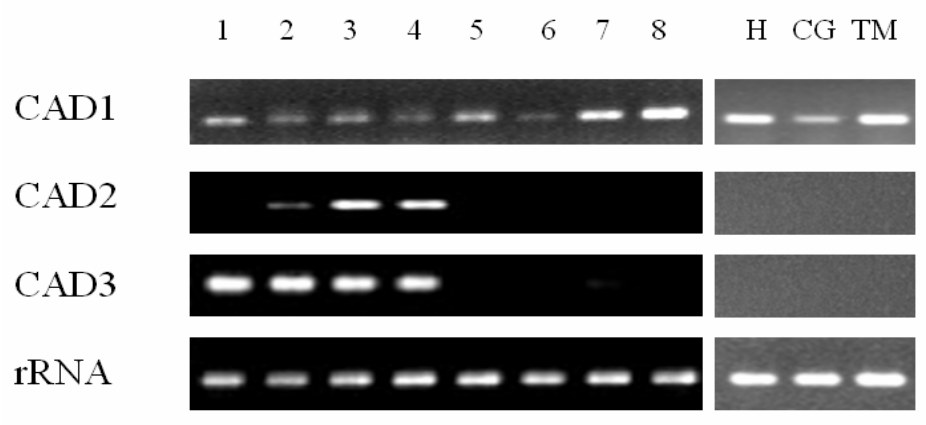

Figura 26. Padrão de expressão dos mRNAs para as CADs em diferentes tecidos no tubo digestivo em larvas de Musca domestica são mostrados por RT-PCR semi-quantitativo. Quantias iguais de RNA total de diferentes tecidos foram extraídos e usados como amostra em uma reação de RT. O produto dessas reações foram usados em uma reação de PCR com primers específicos para cada CAD e para o rRNA (AY123346.1) de Musca domestica (ver Tabela 2). Os produtos da amplificação foram analisados em gel de agarose 1%. As bandas amplificadas apareceram para os respectivos tecidos quando os correspondentes mRNAs eram transcritos. 1, 2 e 3 (ventrículo anterior), 4, 5 e 6 (ventrículo médio), 7 e 8 (ventrículo posterior) (ver Figura 5), H. hemolinfa, CG. corpo gorduroso, TM. túbulos de Malpighi. O número de ciclos durante o PCR foi escolhido após várias tentativas, assim que a amplificação estava em fase log e resultando em uma clara e visível banda.

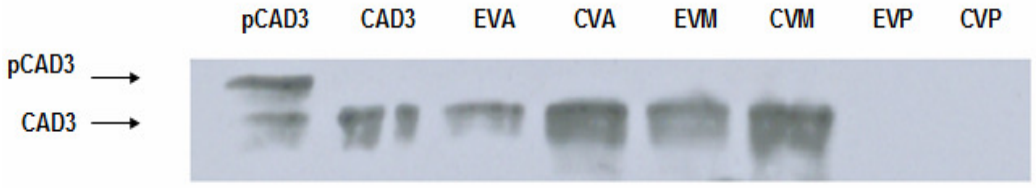

Figura 27. Análise por Western Blot após SDS-PAGE (12%) das proteínas recombinante (pCAD3 e CAD3), do conteúdo luminal e epitélio marcadas com o anticorpo anti-pCAD3. A detecção padrão com anti-IgG (coelho) acoplado com

peroxidase e revelado com kit "Imobbilon Western Chemiluminescent HRP substrate" (Millipore) sendo visualizado após exposição com o filme "High-performance luminescence detection film" (GE). EVA. epitélio do ventrículo anterior, CVA. conteúdo luminal do ventrículo anterior, EVM. epitélio do ventrículo médio, CVM. conteúdo luminal do ventrículo médio, EVP. epitélio do ventrículo posterior, CVP. conteúdo luminal do ventrículo posterior. Em cada linha foram aplicadas 40 µg de proteínas totais.

4.1.6. Imunocitolocalização da pCAD3 em Musca domestica

O anticorpo produzido (anti-pCAD3) foi usado em uma imunocitolocalização em microscopia eletrônica de transmissão dessa enzima no ventrículo de larvas de *Musca domestica*. As marcações são evidenciadas por partículas de ouro coloidal ligadas ao anticorpo secundário que reconhece o anticorpo anti-pCAD3. As marcações ocorreram de forma predominante no interior de pequenas vesículas de secreção localizadas no ápice das células e elementos do Golgi nas regiões 1 e 3 (ventrículo anterior) e na região proximal da região 4 (ventrículo médio) (Figura 28).

As células responsáveis pela secreção mencionadas anteriormente possuem todas as características morfológicas associadas a atividade secretória tais como a presença de elementos de Golgi, retículo endoplasmático rugoso e vesículas de secreção localizados no ápice da célula. Pelo resultado obtido é possível verificar que o ventrículo anterior de larvas de *Musca domestica* é o grande responsável pela secreção da CAD3. Também fica claro que essa enzima é secretada pelo animal, descartando a possibilidade dessa enzima ser de origem bacteriana.

4.2. Sequênciamento randômico de ESTs ventriculares de larvas de Musca domestica

Visando esclarecer aspectos moleculares da fisiologia digestiva de larvas de *M. domestica*, realizamos um sequênciamento randômico de ESTs a partir de uma biblioteca de cDNA ventricular de larvas de *Musca domestica,* pois dessa forma sequências obtidas nesse procedimento poderiam codificar para proteínas de interesse fisiológico digestivo. Sequências que despertaram interesse serão no

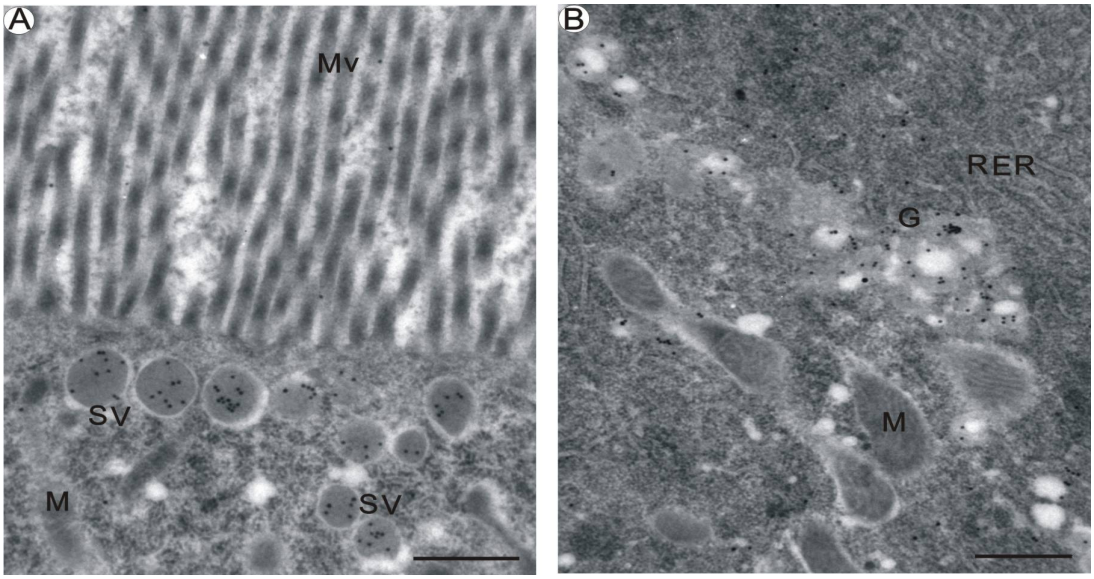

Figura 28. Imunocitolocalização por microscopia eletrônica de transmissão usando o anticorpo anti-pCAD3 nas células do ventrículo de larvas de Musca domestica. SV. vesículas de secreção, Mv. Microvilosidades, M. Mitocôndria, G. Complexo de Golgi, RER. Retículo endoplasmático rugoso. As escalas apresentadas equivalem a 0,5 μm. Marcações são evidenciadas com o anticorpo em vesículas de secreção localizadas no ápice das células. Nenhuma marcação foi observada nos controles com o soro pré-imune ou quando o anticorpo primário (anti-pCAD3) foi omitido.

futuro expressas na forma de proteína recombinante, caracterizadas e imunocitolocalizadas permitindo a formulação de um modelo mais refinado do processo fisiológico digestivo do inseto modelo. A princípio tínhamos em mãos uma biblioteca de cDNA normalizada, clonada em fago λ zap II, construída de forma comercial pela empresa norte americana Stratagene. Ao realizarmos os procedimentos experimentais (dados não mostrados nessa tese) verificamos de forma trágica um erro técnico cometido durante a construção desta biblioteca. Ao realizarmos o sequênciamento randômico dos ESTs presentes nessa biblioteca, constatamos que 70% dos clones sequenciados codificavam para RNA ribossomais, tornando-a inviável para obtenção de sequências de interesse ao nosso trabalho.

Frente a essas dificuldades, construímos uma nova biblioteca de cDNA. Existem vários métodos de construção de bibliotecas de alta qualidade clonada em

fagos. Mas como o nosso objetivo é gerar uma biblioteca para sequênciarmos randomicamente poucos ESTs (aproximadamente 1000), nós optamos por um caminho menos complicado. Construímos então uma biblioteca de cDNA clonada em plasmídeos conforme descrito em Calvo et. al., 2004 (ver item 2.20.).

4.2.1. Sequênciamento randômico dos ESTs ventriculares de larvas de Musca domestica

Nós sequenciamos os ESTs de forma randômica presentes em uma biblioteca de cDNA ventricular de larvas de *Musca domestica.* Após sequências de baixa qualidade serem manualmente removidas, obtivemos 826 ESTs sequênciados a partir da extremidade 5' originado a partir de diferentes clones.

Os ESTs foram analisados com os programas Phred, Phrap e Consed Linux (Ewing *et al.,* 1998; Gordon et. al., 1998), produzindo 99 contíguos e 244 singletos, totalizando 343 sequências únicas ou *clusters*. As sequências são nomeadas por contíguo (se formado por 2 ou mais ESTs) ou por singleto (formada por um único EST), mas para a simplicidade desta tese nós iremos usar a denominação de *cluster* ou sequência única, para identificar as sequências em ambos os casos.

A distribuição dos ESTs formadores de contíguos estão sumarizados na figura 29 sendo que 46 contíguos são formados por apenas 2 ESTs, enquanto que o contíguo com o maior número de sequências apresenta 83 ESTs.

Porém não se pode esperar que cada sequência única ou *cluster* seja um transcrito diferente, pois muitos dos ESTs que transmitem a mesma mensagem podem não se sobrepor ou podem estar presentes no banco de ESTs representando *splicings* alternativos, pré-mRNAs, polimorfismos ou regiões de baixa qualidade que dentro das condições de estringência adotadas podem levar a formação de contíguos diferentes de um mesmo transcrito. Postulando-se que cada sequência única ou *cluster* corresponda a um transcrito diferente e considerando que *Musca domestica* possui um genoma com o tamanho e a característica similar ao de *Drosophila melanogaster* (13647 genes e 18471 transcritos diferentes – FlyBase 3.1 – Berkeleg Drosophila Genome Project – www.fruitfly.org/annot/) é provável que as presentes informações cubram aproximadamente 2,5% dos genes e 1,9% dos transcritos de *Musca domestica*.

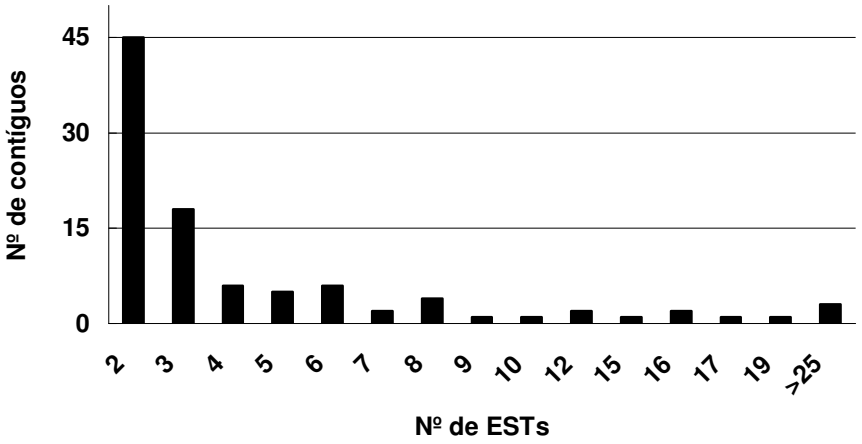

Figura 29. Distribuição e número de ESTs que formaram cada contíguo.

Os 343 *clusters* foram comparados usando os programas BLASTX e BLASTN (Altschul *et al.,* 1997) contra o banco de dados não redundante do GENBANK (Benson *et al.,* 2000) de forma automatizada, utilizando o programa BLAST client 3 (Wheeler *et al.,* 2000).

Análise por BLASTN (banco de dados não redundante) mostrou que 66,10% dos ESTs e 44,31% das sequências únicas de *Musca domestica* apresentou similaridade com algum tipo de sequência depositada no GENBANK (valor de corte do *e-value* < 10^{-10}). Como esperado, o organismo mais similar foi *D. melanogaster* seguido de *D. pseudoobscura* (Figura 30).

Análise por BLASTX (banco de dados não redundante) mostrou que grande maioria dos ESTs (82,92%) e das sequência únicas (78,13%) obtidas no transcriptoma de *M. domestica* possuem similaridade com proteínas depositadas no GENBANK. Em comparação, as sequências de *Drosophila melanogaster* e *D. pseudoobscura* se mostraram mais similares as sequências de *M. domestica* (Figura 31) (nota de corte do e-*value* < 10^{-6}). Estes resultados confirmam (como o esperado) uma grande aproximidade filogenética entre esses Diptera, pois os resultados dos BLASTN e BLASTX resultaram em maciça presença no GENBANK das sequências desses organismos. Nós não descartamos a hipótese de que esse quadro possa mudar no futuro quando sequências adicionais ou até genomas completos de outros Diptera Cyclorrhapha forem sendo depositados no banco de dados.

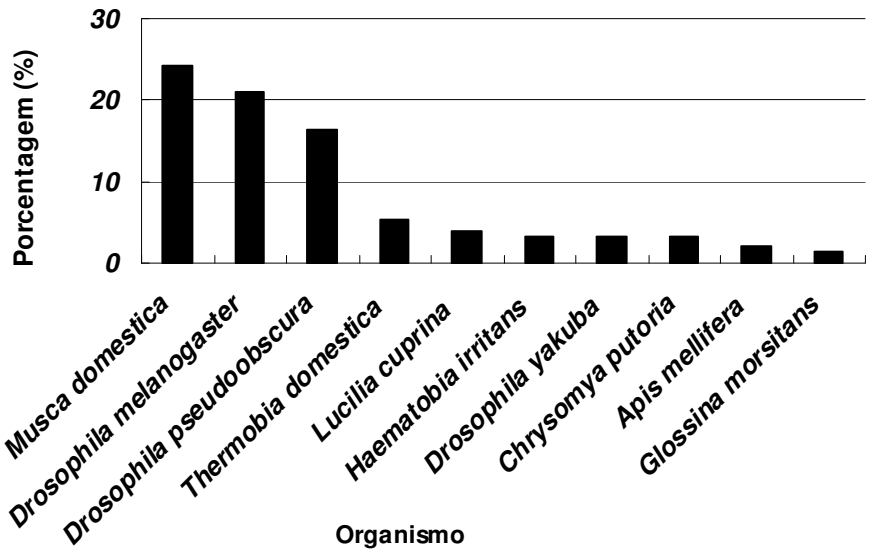

Figura 30. Análise por BLASTN *das sequências únicas ou clusters obtidas no transcriptoma ventricular de larvas de Musca domestica. No gráfico estão demonstrados os organismos que apresentaram maiores similaridades com as sequências de Musca domestica.*

4.2.2. Organização das informações do transcriptoma de M. domestica

Todas as 343 sequências únicas ou *clusters* foram manualmente anotados e separados em 3 categorias: categoria S: São aquelas sequências codificante para possíveis produtos secretados e eventualmente envolvidos no processo digestivo. Para classificar as sequências dentro dessa categoria analisávamos a sequência de aminoácidos do organismo mais próximo obtido por BLASTX (valor de corte do *e-value* < 10^{-6}). Caso a sequência de aminoácidos possuísse peptídeo sinal, nós assumíamos que o *cluster* correspondente ao nosso organismo (*M. domestica*) fosse também um produto secretado. Categoria H: aqueles possivelmente associados a processo metabólicos (*housekeeping),* como por exemplo enzimas do metabolismo energético, proteínas envolvidas nos processos estruturais, transporte, regulação, entre outros.

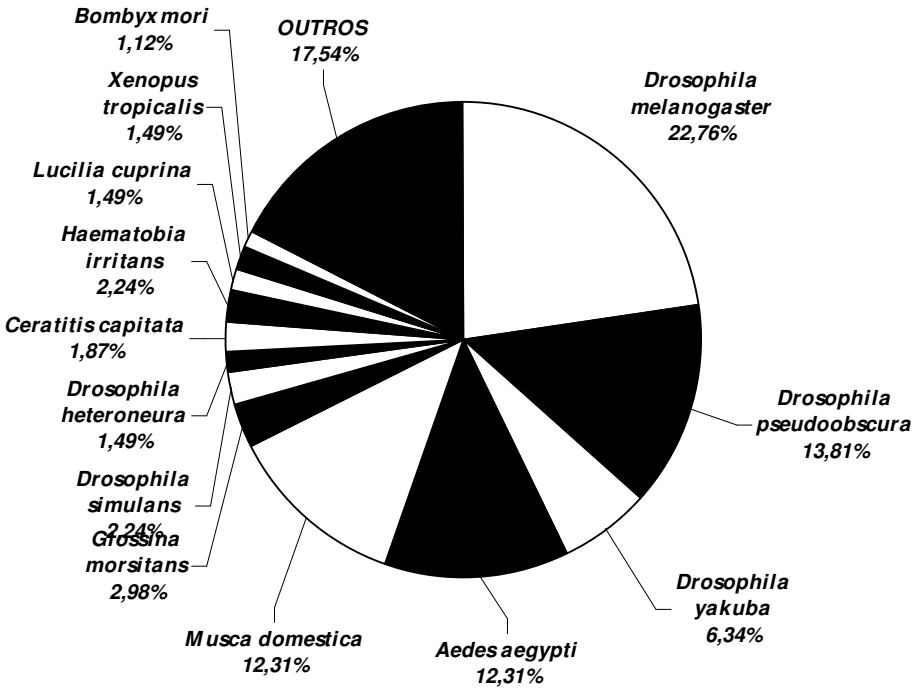

Figura 31. Análise por BLASTX *das sequências únicas ou clusters obtidas no transcriptoma ventricular de larvas de* Musca domestica*. No gráfico estão demonstrados os organismos que apresentaram maiores similaridades com as sequências de* Musca domestica*.*

Categoria U: aqueles que codificam proteínas sem função conhecida (*unknown*) ou que apresentavam baixos caracteres de confiabilidade (*e-values* altos) obtidos por BLAST.

Cento e trinta e dois *clusters* contendo 190 ESTs (23,0% do transcriptoma) foram classificados na categoria H (Figura 32). Esses *clusters* têm a média de 1,44 ESTs por *cluster*. Em contra partida 116 *clusters* contendo 423 ESTs (51,21% do transcriptoma; média de 3,65 ESTs por *cluster*) foram classificados na categoria S. Finalmente 95 *clusters* com 213 ESTs (25,79% do transcriptoma, média de 2,24 ESTs por *cluster*) foram classificados como categoria U (Tabela 4).

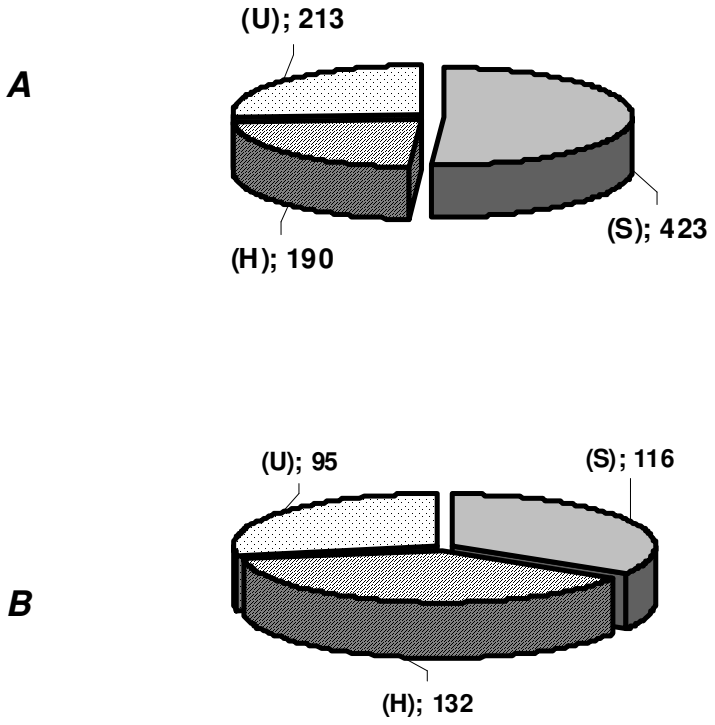

Figura 32. Número de ESTs (A) e sequências únicas (clusters) (B) obtidos a partir de 826 clones sequênciados randomicamente a partir de uma biblioteca de cDNA ventricular de larvas de Musca domestica. As sequências foram classificadas como provavelmente secretadas (S), metabolismo (H) ou sem função conhecida (U). O número de ESTs e sequências únicas (clusters) estão indicadas no diagrama.

Tabela 4. Perfil das sequências encontradas no intestino médio de larvas de Musca domestica.

Tipo de transcrito	Nº clusters	Nº ESTs	EST / cluster	cluster (%)	EST (%)
Secretado (S)	116	423	3,65	33,82	51,21
Metabolismo (H)	132	190	1,44	38,48	23,00
Não conhecido (U)	95	213	2,24	27,70	25,79
Total	343	826			

4.2.3. Descrição dos clusters pertencentes a categoria S

Foram identificados 116 *clusters* formados por 423 ESTs pertencentes a categoria S (Tabela 5). Os *clusters* podem variar de 1 a 83 ESTs sendo que a maioria desses *clusters* pertence a família de proteínas bem conhecidas tais como: lisozimas (15 *clusters*, 161 ESTs), catepsinas-D (6 *clusters*, 9 ESTs), lipases (5 *clusters*, 14 ESTs), peritrofinas (4 *clusters*, 8 ESTs), carboxipeptidases (3 *clusters*, 5 ESTs), dipeptidase (1 *cluster*, 1 EST), tripsinas (7 *clusters*, 25 ESTs), quimotripsinas (6 *clusters*, 100 ESTs) e enzimas que atuam na digestão de carboidratos (6 *clusters*, 7 ESTs). Existem também sequências que apresentaram similaridade apenas com sequências de proteínas hipotéticas (sem função conhecida) depositada no banco de dados sendo que estas estão descritas a seguir.

4.2.3.1. Descrição dos clusters pertencentes a categoria S sem função conhecida

Foram obtidos 63 conjuntos de sequências, totalizando 93 ESTs que apresentam similaridade apenas com proteínas hipotéticas depositadas em bancos de genomas (*D. melanogaster* e *D. pseudoobscura*) e não apresentaram similaridade com proteínas caracterizadas de outros organismos sendo que no futuro essas sequências podem ser identificadas melhor e caracterizadas.

4.2.4. Descrição dos clusters pertencentes a categoria housekeeping (H)

Cento e trinta e dois *clusters* formados por 190 ESTs pertencentes a categoria H foram identificados no transcriptoma ventricular de *Musca domestica* (Tabela 6). Dezessete *clusters* formados por 21 ESTs correspondem a genes envolvidos na síntese e secreção de proteínas incluindo aqueles que codificam para proteínas ribossomais. Quarenta e quatro *clusters* (63 ESTs) estão associados ao metabolismo energético incluindo várias enzimas mitocondriais (citocromos e ATPsintases), enzimas da via glicolítica (triose fosfato isomerase, fosfofrutoquinase e piruvato quinase), gliconeogênese (lactato desidrogenase B) e várias desidrogenases. Nove *clusters* (10 ESTs) estão associados com possíveis fatores de transcrição e tradução incluindo fatores de elongação, fatores de iniciação de transcrição, fatores de regulação nuclear, deoxiribonucleases, ligase e sintetases.

Trinta e nove *clusters* (68 ESTs) estão associados aos caminhos de sinalização e transporte incluindo glutationa S transferase, metalothioneina B, receptores e ATPase envolvidas no transporte de íons tal como a Vacuolar H^+-ATPase. Quatro *clusters* (5 ESTs) estão associados a produtos possivelmente envolvidos no *folding* de proteína tais como chaperonas, cofator A e imunofilinas. Doze *clusters* (13 ESTs) estão associados a funções estruturais tais como actina, fibrinogênio, filamina e tetraspaninas.

4.2.5. Descrição dos clusters pertencentes a categoria U

Nós obtivemos 213 ESTs formando 95 *clusters* (Tabela 7) e anotados como sequências não conhecidas. Destes, 20 *clusters* (72 ESTs) apresentaram similaridade com proteínas de *Drosophila melanogaster*, *Drosophila pseudoobscura* e outros organismos, porém essas proteínas não possuem função e localização conhecida nesses insetos. O restante das sequências não apresentam similaridade significativa com proteínas depositadas no banco de dados, sendo que estas podem representar novas proteínas (única em processos digestivo em Diptera) ou artefatos de PCR (erros durante o processo de sequênciamento), sendo que estes itens nós não entraremos em maiores detalhes nesta tese.

Tabela 5. Proteínas provavelmente secretadas

Nº do cluster	*Nº ESTs*	*Identificação*	*Proteína*	*Organismo*	*E value*	*Identidade*
Lisozimas						
1	83	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	0E+00	100%
2	17	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	6E-73	99%
3	16	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	4E-73	99%
4	15	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	2E-72	90%
5	12	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	2E-72	89%
6	8	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	3E-72	89%
7	1	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	2E-06	92%
8	1	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	4E-16	92%
9	1	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	5E-54	92%
10	1	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	2E-11	92%
11	1	sp|Q7YT16|	Lysozyme 1	*Musca domestica*	2E-18	92%
12	2	gb|AAQ20047.1	Lys-rich lysozyme 2	*Musca domestica*	2E-62	89%
13	1	gb|AAQ20047.1	Lys-rich lysozyme 2	*Musca domestica*	4E-34	87%
14	1	gb|AAQ20047.1	Lys-rich lysozyme 2	*Musca domestica*	1E-54	87%
15	1	gb|AAQ20047.1	Lys-rich lysozyme 2	*Musca domestica*	8E-13	87%
Lipase						
16	9	gb|EAT41287.1	Lipase	*Aedes aegypti*	7E-46	42%
17	2	gb|EAT41287.1	Lipase	*Aedes aegypti*	2E-47	51%
18	1	ref|NP_477331.1	Lip3 CG8823-PA	*Drosophila melanogaster*	3E-97	84%
19	1	gb|EAT35484.1	Lipase 1 precursor	*Aedes aegypti*	4E-39	41%
20	1	gb|EAT34384.1	Lysosomal acid lipase	*Aedes aegypti*	1E-11	39%
Catepsina D						
21	2	gb|ABL84270.1	Aspartic proteinase AspMD01	*Musca domestica*	3E-88	78%
22	1	gb|ABL84270.1	Aspartic proteinase AspMD01	*Musca domestica*	8E-47	64%
23	1	gb|ABL84270.1	Aspartic proteinase AspMD01	*Musca domestica*	6E-33	64%
24	1	gb|ABL84270.1	Aspartic proteinase AspMD01	*Musca domestica*	2E-05	78%
25	1	gb|ABM69085.1	Aspartic proteinase AspMD02	*Musca domestica*	2E-17	86%
26	3	gb|ABM69086.1	Aspartic proteinase AspMD03	*Musca domestica*	6E-72	79%
Hidrólise de carboidratos						
27	2	gb|AAO13691.1	alpha-Amylase	*Ceratitis capitata*	2E-73	70%
28	1	gb|AAO13691.1	alpha-Amylase	*Ceratitis capitata*	1E-75	79%
29	1	dbj|BAE86928.1|	alpha-Glucosidase	*Apis mellifera*	2E-22	38%
30	1	gb|AAO13692.1	Amylase-related protein	*Ceratitis capitata*	2E-30	60%
31	1	gb|AAO13691.1	alpha-Amylase	*Ceratitis capitata*	1E-08	79%
32	1	emb|CAA60858.1	Maltase-like	*Anopheles gambiae*	2E-17	51%
Carboxipeptidase A						
33	1	gb|AAC05137.1	Carboxypeptidase A	*Drosophila heteroneura*	8E-13	63%
34	2	gb|AAC05137.1	Carboxypeptidase A	*Drosophila heteroneura*	1E-43	50%
35	2	gb|AAC05137.1	Carboxypeptidase A	*Drosophila heteroneura*	4E-43	51%
Quimotripsina						
36	71	gb|AAF91345.1	Chymotrypsin-like	*Glossina morsitans*	1E-63	69%
37	16	gb|AAF91345.1	Chymotrypsin-like	*Glossina morsitans*	2E-55	54%
38	1	gb|AAF91345.1	Chymotrypsin-like	*Glossina morsitans*	3E-08	78%
39	1	gb|AAY59001.1|	Proteolytic lectin	*Glossina austeni*	3E-17	35%

Tabela 5. (continuação)

Nº do cluster	***Nº ESTs***	***Identificação***	***Proteína***	***Organismo***	***E value***	***Identidade***
40	1	gb|AAL93243.1	Larval chymotrypsin-like	*Aedes aegypti*	5E-22	55%
41	10	gb|AAF91345.1	Chymotrypsin-like	*Glossina morsitans*	3E-44	51%
Dipeptidase						
42	1	ref|NP_650318.1	Dipeptidase B CG9285-PA	*Drosophila melanogaster*	2E-23	71%
Peritrofinas						
43	5	gb|AAK01057.1	Peritrophin-15	*Chrysomya bezziana*	5E-28	60%
44	1	gb|AAK01057.1	Peritrophin-15 precursor	*Chrysomya bezziana*	2E-22	59%
45	1	gb|AAK01057.1	Peritrophin-15 precursor	*Chrysomya bezziana*	7E-13	46%
46	1	sp|Q25255|	Peritrophin-44 precursor	*Lucilia cuprina*	4E-14	43%
Tripsinas						
47	5	gb|AAN60085.1	Phosphotrypsin	*Glossina fuscipes*	2E-61	76%
48	3	sp|P35044|	Trypsin alpha-4	*Lucilia cuprina*	5E-78	86%
49	2	gb|AAB66370.1	Dvtry-1 trypsin precusor	*Drosophila virilis*	4E-85	84%
50	4	gb|AAB66370.1	Dvtry-1 trypsin	*Drosophila virilis*	2E-48	42%
51	6	sp|P35043|	Trypsin alpha-3	*Lucilia cuprina*	6E-35	70%
52	2	sp|P35043|	Trypsin alpha-3	*Lucilia cuprina*	2E-34	61%
53	3	gb|AAC05141.1	Trypsin	*Drosophila heteroneura*	5E-37	56%
Proteínas sem função conhecida						
54	1	gb|ABN49347.1	IP18055p	*Drosophila melanogaster*	6E-15	45%
55	1	ref|XP_001028745.1	Hypothetical protein	*Tetrahymena thermophila*	7E-32	58%
56	1	gb|EAT47864.1	Hypothetical protein AaeL	*Aedes aegypti*	4E-06	30%
57	1	emb|CAD89179.1	Peptidoglycan recognition	*Drosophila melanogaster*	3E-12	93%
58	3	gb|AAY66614.1	Putative secreted protein	*Ixodes scapularis*	2E-06	68%
59	2	gb|AAY66614.1	Putative secreted protein	*Ixodes scapularis*	3E-07	68%
60	1	gb|AAN71654.1	SD12036p	*Drosophila melanogaster*	8E-42	79%
61	1	ref|XP_647616.1	GA21464-PA	*Dictyostelium discoideum*	8E-04	36%
62	1	ref|XP_001352635.1	GA10936-PA	*Drosophila pseudoobscura*	1E-05	57%
63	2	ref|XP_001357932.1	GA14124-PA	*Drosophila pseudoobscura*	1E-35	44%
64	2	emb|CAL26565.1	CG7953	*Drosophila simulans*	2E-15	61%
65	2	emb|CAL26565.1	CG7953	*Drosophila simulans*	5E-69	61%
66	2	ref|XP_001355598.1	GA21954-PA	*Drosophila pseudoobscura*	4E-45	45%
67	2	ref|XP_001353851.1	GA19618-PA	*Drosophila pseudoobscura*	2E-36	82%
68	3	ref|XP_001353851.1	GA19618-PA	*Drosophila pseudoobscura*	4E-67	83%
69	1	ref|XP_001353851.1	GA19618-PA	*Drosophila pseudoobscura*	7E-10	66%
70	1	ref|XP_001353851.1	GA19618-PA	*Drosophila pseudoobscura*	2E-64	82%
71	1	gb|AAK07478.2	GA16452-PA	*Glossina morsitans*	1E-47	68%
72	1	dbj|BAE20328.1	CG5770	*Triticum aestivum*	1E-05	28%
73	1	ref|XP_001353851.1	GA19618-PA	*Drosophila pseudoobscura*	8E-15	82%
74	2	gb|AAQ09855.1	CG9568	*Drosophila yakuba*	7E-13	33%
75	2	ref|XP_001359181.1	GA21004-PA	*Drosophila pseudoobscura*	2E-33	53%
76	2	ref|XP_001356450.1	GA11277-PA	*Drosophila pseudoobscura*	1E-111	72%
77	2	ref|XP_001360670.1	GA12205-PA	*Drosophila pseudoobscura*	2E-31	58%
78	6	ref|XP_001360670.1	GA12205-PA	*Drosophila pseudoobscura*	6E-27	48%
79	3	ref|XP_001356235.1	GA21464-PA	*Drosophila pseudoobscura*	2E-60	71%
80	1	ref|XP_001356235.1	GA21464-PA	*Drosophila pseudoobscura*	2E-06	40%
81	3	ref|NP_609682.1	CG7916-PA	*Drosophila melanogaster*	7E-37	57%
82	2	gb|AAQ09817.1	CG5770	*Drosophila yakuba*	1E-08	32%
83	2	ref|XP_001358274.1	GA19482-PA	*Drosophila pseudoobscura*	3E-58	68%
84	1	ref|XP_001358274.1	GA19482-PA	*Drosophila pseudoobscura*	7E-47	47%
85	3	gb|AAM52606.1	GH05321p	*Drosophila melanogaster*	3E-35	66%
86	2	ref|XP_001356963.1	GA16452-PA	*Drosophila pseudoobscura*	6E-08	43%
87	1	ref|XP_001356963.1	GA16452-PA	*Drosophila pseudoobscura*	1E-07	43%

Tabela 5. (continuação)

Nº do cluster	***Nº ESTs***	***Identificação***	***Proteína***	***Organismo***	***E value***	***Identidade***
88	1	ref|NP_650481.1	CG5399-PA	*Drosophila melanogaster*	3E-09	29%
89	1	ref|NP_611097.2	CG8446-PA	*Drosophila melanogaster*	2E-26	63%
90	1	ref|NP_648526.2	CG5883-PA	*Drosophila melanogaster*	7E-07	26%
91	1	ref|XP_001350885.1	PFG377 protein	*Plasmodium falciparum*	1E-04	45%
92	1	ref|NP_609267.1	CG9568-PA	*Drosophila melanogaster*	2E-06	35%
93	1	ref|XP_001355740.1	GA17576-PA	*Drosophila pseudoobscura*	8E-61	68%
94	1	ref|XP_001353329.1	GA20426-PA	*Drosophila pseudoobscura*	1E-25	67%
95	1	ref|XP_001356449.1	GA17302-PA	*Drosophila pseudoobscura*	5E-17	45%
96	1	ref|XP_001356234.1	GA20684-PA	*Drosophila pseudoobscura*	3E-22	53%
97	1	ref|NP_647939.1	CG13704-PA	*Drosophila melanogaster*	3E-15	51%
98	1	ref|XP_001355340.1	GA15367-PA	*Drosophila pseudoobscura*	4E-05	31%
99	1	ref|XP_001355125.1	GA11365-PA	*Drosophila pseudoobscura*	2E-35	46%
100	1	ref|NP_608776.1	CG2772-PA	*Drosophila melanogaster*	1E-07	43%
101	1	ref|XP_001355049.1	GA11993-PA	*Drosophila pseudoobscura*	2E-22	53%
102	1	ref|XP_001358273.1	GA19495-PA	*Drosophila pseudoobscura*	1E-40	55%
103	1	ref|XP_001360723.1	GA19113-PA	*Drosophila pseudoobscura*	9E-07	24%
104	1	ref|XP_001357978.1	GA16136-PA	*Drosophila pseudoobscura*	9E-36	44%
105	1	ref|NP_572282.1	CG6048-PA	*Drosophila melanogaster*	7E-45	58%
106	1	gb|AAI10935.1	MGC132092 protein	*Xenopus laevis*	0.007	28%
107	1	ref|XP_001358275.1	GA19487-PA	*Drosophila pseudoobscura*	1E-15	54%
108	1	gb|AAQ09884.1	CG8661	*Drosophila yakuba*	8E-09	55%
109	3	gb|AAQ09884.1	CG8661	*Drosophila yakuba*	6E-13	31%
110	1	ref|NP_650599.1	CG31265-PA	*Drosophila melanogaster*	2E-17	67%
111	1	ref|NP_572479.1	CG15347-PA	*Drosophila melanogaster*	1E-41	45%
112	1	ref|XP_001359327.1	GA17826-PA	*Drosophila pseudoobscura*	7E-06	41%
113	1	ref|XP_001361680.1	GA18134-PA	*Drosophila pseudoobscura*	4E-05	33%
114	1	ref|NP_651526.1	CG6271-PA	*Drosophila melanogaster*	5E-16	47%
115	1	gb|AAR10016.1	CG6467	*Drosophila yakuba*	9E-55	81%
116	1	ref|NP_650603.1	CG5246-PA	*Drosophila melanogaster*	1E-38	56%

Tabela 6. Proteína de metabolismo (Housekeeping)

Nº do cluster	*Nº ESTs*	*Identificação*	*Proteína*	*Organismo*	*E value*	*Identidade*
Folding de proteínas						
117	1	gb|EAT44178.1	Tubulin-specific chaperone	*Aedes aegypti*	2E-13	65%
118	1	gb|EAT42311.1	Cofactor A, putative	*Aedes aegypti*	1E-20	71%
119	2	ref|NP_001037356.1	45 kDa immunophilin FKBP45	*Bombyx mori*	4E-12	51%
120	1	ref|NP_001037356.1	45 kDa immunophilin FKBP45	*Bombyx mori*	4E-13	49%
Armazenagem						
121	1	ref|NP_732904.2	Lipid storage droplet-1 CG10374-PA	*Drosophila melanogaster*	1E-07	61%
122	1	gb|AAL65406.2	Ferritin light-chain	*Glossina morsitans*	2E-65	73%
123	1	gb|AAL65406.2	Ferritin light-chain	*Glossina morsitans*	1E-36	73%
124	4	gb|ABC48949.1	Ferritin heavy chain-like	*Glossina morsitans*	4E-72	84%
Inibidores						
125	1	ref|NP_524953.1	Serine protease inhibitor	*Drosophila melanogaster*	2E-37	51%
126	1	gb|EAT33888.1	Diazepam binding inhibitor, putative	*Aedes aegypti*	3E-14	54%
127	1	gb|ABM55609.1	Diazepam binding inhibitor-like protein	*Maconellicoccus hirsutus*	6E-10	46%
Proteína de síntese e secreção						
128	1	gb|AAF31449.1	60S Acidic ribosomal protein P0	*Sarcophaga crassipalpis*	2E-04	96%
129	3	emb|CAA70259.1	Ribosomal protein P2	*Ceratitis capitata*	6E-23	55%
130	1	emb|CAA64319.1	Ribosomal protein L9	*Drosophila melanogaster*	2E-46	86%
131	1	gb|ABM55458.1	Ribosomal protein S11-2	*Xenopsylla cheopis*	4E-19	92%
132	1	ref|NP_523595.1	Ribosomal protein S26	*Drosophila melanogaster*	5E-32	88%
133	1	ref|NP_651740.1	Ribosomal protein S8	*Drosophila melanogaster*	5E-57	90%
134	1	gb|ABF18041.1	Ribosomal protein L10	*Aedes aegypti*	5E-21	60%
135	1	ref|NP_611685.1	Ribosomal protein S16	*Drosophila melanogaster*	8E-28	75%
136	1	ref|NP_511073.1	Ribosomal protein S6	*Drosophila melanogaster*	3E-90	79%
137	1	ref|NP_476963.1	Ribosomal protein L27A	*Drosophila melanogaster*	2E-45	63%
138	1	ref|NP_524687.1	Ribosomal protein L30	*Drosophila melanogaster*	1E-12	87%
139	1	ref|NP_651876.1	Ribosomal protein L6	*Drosophila melanogaster*	4E-05	84%
140	1	ref|NP_476631.1	Ribosomal protein L19	*Drosophila melanogaster*	1E-77	90%
141	1	ref|NP_651460.2	Ribosomal protein L34a	*Drosophila melanogaster*	2E-50	89%
142	2	ref|NP_609179.2	Ribosomal protein L36A	*Drosophila melanogaster*	2E-46	88%
143	1	ref|NP_524315.2	Ribosomal protein S25	*Drosophila melanogaster*	6E-15	95%
144	2	gb|ABM55591.1	Ubiquitin/ribosomal protein	*Maconellicoccus hirsutus*	6E-32	66%
Proteínas estruturais						
145	2	gb|ABG46341.1	Actin	*Aedes albopictus*	1E-61	99%
146	1	gb|ABF18522.1	Actin depolymerizing factor	*Aedes aegypti*	2E-28	84%
147	1	gb|ABF06543.1	Actin	*Ips paraconfusus*	4E-53	100%
148	1	gb|EAT33964.1	f-Actin capping protein alpha	*Aedes aegypti*	5E-41	78%
149	1	gb|ABD39699.1	Putative actin	*Lygus lineolaris*	5E-71	97%
150	1	gb|AAA51580.1	gamma-Actin	*Homo sapiens*	1E-17	64%
151	1	ref|NP_001072519.1	Fibrinogen C domain	*Xenopus tropicalis*	2E-20	51%
152	1	ref|NP_476691.1	Proteasome 29kD subunit	*Drosophila melanogaster*	2E-32	69%
153	1	gb|EAT41490.1	Muscle lim protein	*Aedes aegypti*	5E-33	81%
154	1	gb|AAF25614.1	Filamin1	*Drosophila melanogaster*	1E-63	61%
155	1	ref|NP_523630.1	Tetraspanin 42Ed	*Drosophila melanogaster*	5E-33	46%
156	1	ref|NP_523614.1	Decondensation factor 31	*Drosophila melanogaster*	1E-13	71%
Metabolismo energético						
157	1	gb|AAC39072.1	Triose phosphate isomerase	*Drosophila simulans*	2E-56	83%
158	1	ref|NP_724891.2	Phosphofructokinase	*Drosophila melanogaster*	5E-71	81%
159	1	gb|EAT44544.1	Aldo-keto reductase	*Aedes aegypti*	1E-37	46%
160	1	ref|ZP_01557919.1	Glycosyl transferase, family 2	*Burkholderia ambifaria*	5E-05	34%
161	1	gb|EAT44639.1	Glucose dehydrogenase	*Aedes aegypti*	8E-25	60%
162	1	ref|NP_032465.2	Ketohexokinase	*Mus musculus*	3E-23	56%

Tabela 6. (continuação)

Nº do cluster	***Nº ESTs***	***Identificação***	***Proteína***	***Organismo***	***E value***	***Identidade***
163	1	gb|ABF18243.1	Glyoxylase	*Aedes aegypti*	8E-17	72%
164	1	gb|ABF18352.1	Farnesoic acid O-methyl transferase	*Aedes aegypti*	8E-30	49%
165	1	gb|AAH90087.1	Wdr4-prov protein	*Xenopus tropicalis*	1E-26	39%
166	1	gb|AAC15808.1	Pyruvate kinase	*Drosophila melanogaster*	2E-36	70%
167	1	gb|ABF18352.1	Farnesoic acid O-methyl transferase	*Aedes aegypti*	5E-28	47%
168	1	gb|EAT40260.1	Mitochondrial NADH	*Aedes aegypti*	9E-34	63%
169	1	ref|YP_245509.1	NADH dehydrogenase subunit 1	*Haematobia irritans*	1E-43	64%
170	1	ref|YP_245497.1	NADH dehydrogenase subunit 2	*Haematobia irritans*	2E-14	52%
171	2	ref|NP_075454.1	NADH dehydrogenase subunit 3	*Cochliomyia hominivorax*	6E-09	41%
172	1	ref|NP_085090.1	NADH dehydrogenase subunit 4	*Chrysomya putoria*	3E-48	47%
173	1	ref|YP_245504.1	NADH dehydrogenase subunit 5	*Haematobia irritans*	5E-44	55%
174	1	emb|CAI52438.1	Cytochrome oxidase, subunit I	*Musca domestica*	6E-04	83%
175	1	gb|AAD39883.1	Cytochrome oxidase subunit I	*Musca domestica*	8E-88	75%
176	3	gb|AAS99749.1	Cytochrome c oxidase subunit I	*Musca domestica*	1E-128	80%
177	1	gb|AAS99749.1	Cytochrome c oxidase subunit I	*Musca domestica*	5E-33	75%
178	1	gb|AAO63466.1	Cytochrome oxidase subunit II	*Musca domestica*	6E-41	83%
179	1	gb|AAO63466.1	Cytochrome oxidase subunit II	*Musca domestica*	2E-92	80%
180	8	ref|YP_245502.1	Cytochrome c oxidase subunit III	*Haematobia irritans*	1E-108	80%
181	1	ref|YP_245502.1	Cytochrome c oxidase subunit III	*Haematobia irritans*	1E-63	77%
182	1	sp|O16805	Cytochrome P450 4d1	*Drosophila simulans*	7E-10	59%
183	4	ref|YP_245508.1	Cytochrome b	*Haematobia irritans*	5E-61	60%
184	1	gb|ABG34549.1	Cytochrome P450 CYP4D35	*Musca domestica*	4E-36	51%
185	1	gb|AAC33299.1	CYP9 cytochrome P450	*Drosophila mettleri*	2E-35	54%
186	1	ref|NP_991129.1	ADP-ribosylation factor-like 10C	*Danio rerio*	2E-87	84%
187	1	gb|AAN31393.1	Fatty acyl-CoA desaturase	*Musca domestica*	2E-18	80%
188	1	gb|EAT40616.1	Oxidoreductase	*Aedes aegypti*	8E-25	42%
189	1	gb|EAT42652.1	Uridine phosphorylase	*Aedes aegypti*	4E-30	52%
190	1	gb|EAT41184.1	Ubiquilin 1,2	*Aedes aegypti*	4E-36	52%
191	1	gb|AAP13345.1	Hexamerin L1	*Musca domestica*	6E-89	76%
192	1	ref|NP_988882.1	Lactate dehydrogenase B	*Xenopus tropicalis*	1E-22	51%
193	1	ref|NP_733395.1	prolyl-4-hydroxylase-alpha	*Drosophila melanogaster*	5E-45	51%
194	1	sp|Q06450	ATP synthase epsilon chain	*Ipomoea batatas*	6E-10	54%
195	4	ref|NP_075452.1	ATP synthase F0 subunit 6	*Cochliomyia hominivorax*	4E-58	68%
196	3	ref|NP_075452.1	ATP synthase F0 subunit 6	*Cochliomyia hominivorax*	4E-45	68%
197	1	ref|NP_724885.1	14-3-3zeta CG17870-PA, isoform A	*Drosophila melanogaster*	1E-22	46%
198	2	gb|AAR23787.1	Superoxide dismutase (SOD)	*Musca domestica*	1E-128	89%
199	1	ref|NP_524597.1	Saposin	*Drosophila melanogaster*	5E-58	52%
200	1	ref|NP_001008416.1	Kallikrein 1	*Bos taurus*	4E-05	59%
Sinalização e transporte						
201	5	gb|EAT44078.1	Neurotransmitter gated ion channel	*Aedes aegypti*	7E-14	35%
202	2	gb|AAC28086.1	Nitrate transporter NTL1	*Arabidopsis thaliana*	6E-26	68%
203	1	ref|NP_608577.2	Transport and Golgi organization	*Drosophila melanogaster*	2E-29	54%
204	5	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	8E-18	83%
205	4	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	3E-18	62%
206	1	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	6E-16	62%
207	1	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	1E-11	62%
208	3	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	3E-18	83%
209	2	ref|NP_524413.1	Metallothionein B CG4312-PA	*Drosophila melanogaster*	2E-46	83%
210	1	gb|AAF68570.1	Myosin II regulatory light chain	*Drosophila simulans*	5E-50	75%
211	2	gb|EAT44568.1	Mitochondrial benzodiazepine receptor	*Aedes aegypti*	5E-45	64%
212	1	sp|P46437|	Glutathione S-transferase	*Musca domestica*	6E-81	75%
213	1	gb|AAD54937.1	Glutathione S-transferase 6A	*Musca domestica*	2E-45	88%
214	1	ref|NP_524696.2	Microsomal glutathione S-transferase	*Drosophila melanogaster*	5E-37	63%
215	1	gb|AAD54937.1	Glutathione S-transferase 6A	*Mus musculus*	1E-88	88%
216	1	sp|Q70PU1	Peptidoglycan-recognition	*Drosophila simulans*	2E-52	92%
217	8	emb|CAD89184.1	Peptidoglycan recognition protein SC2	*Drosophila melanogaster*	1E-71	78%

Tabela 6. (continuação)

Nº do cluster	***Nº ESTs***	***Identificação***	***Proteína***	***Organismo***	***E value***	***Identidade***
218	3	emb|CAD89185.1	Peptidoglycan recognition protein SC2	*Drosophila melanogaster*	1E-34	43%
219	1	emb|CAD89179.1	Peptidoglycan recognition protein SC2	*Drosophila melanogaster*	1E-54	85%
220	1	emb|CAD89179.1	Peptidoglycan recognition protein SC2	*Drosophila melanogaster*	2E-08	42%
221	1	gb|AAU09451.1	Vacuolar H+-ATPase B2	*Dasyatis sabina*	1E-88	96%
222	1	gb|EAT35734.1	Vacuolar ATP synthase	*Aedes aegypti*	2E-32	72%
223	1	ref|NP_477002.1	Mitochondrial acyl carrier protein	*Drosophila melanogaster*	5E-39	61%
224	1	gb|EAT38079.1	Ganglioside induced differentiation	*Aedes aegypti*	4E-16	73%
225	1	ref|XP_646321.1	Acyl-CoA binding protein	*Dictyostelium discoideum*	8E-16	51%
226	1	gb|EAT43142.1	Importin	*Aedes aegypti*	9E-47	52%
227	2	emb|CAB43739.2	Titin	*Drosophila melanogaster*	4E-16	66%
228	1	gb|ABF51487.1	Thymosin isoform 2	*Bombyx mori*	8E-25	72%
229	2	ref|NP_651707.1	Odorant-binding protein	*Drosophila melanogaster*	2E-36	43%
230	1	ref|NP_651707.1	Odorant-binding protein 99a	*Drosophila melanogaster*	4E-31	43%
231	1	gb|AAY66886.1	p8 Nuclear protein	*Ixodes scapularis*	7E-21	65%
232	2	gb|EAT36371.1	Late endosomal/lysosomal MP1	*Aedes aegypti*	2E-44	69%
233	1	emb|CAJ83813.1	CHK1 checkpoint	*Xenopus tropicalis*	6E-04	67%
234	1	gb|EAT43966.1	Sorting nexin	*Aedes aegypti*	2E-53	76%
235	1	ref|NP_898880.1	Vacuolar protein sorting 28 isoform 2	*Homo sapiens*	5E-39	66%
236	1	gb|EAT43097.1	Mitochondrial processing peptidase	*Aedes aegypti*	3E-57	85%
237	1	gb|ABM55613.1	Death-associated protein-like protein	*Maconellicoccus hirsutus*	2E-11	46%
238	2	ref|NP_524938.1	Lethal (2) 35Bg CG4180-PA	*Drosophila melanogaster*	5E-49	59%
239	1	ref|NP_690845.1	Mitochondrial protein	*Saccharomyces cerevisiae*	2E-18	63%
Fatores de transcrição e tradução						
240	1	emb|CAF25506.1	Transcription factor AP-1	*Oncorhynchus mykiss*	1E-29	67%
241	1	gb|ABG88916.1	Elongation factor-1 gamma	*Aliatypus plutonis*	3E-38	52%
242	1	gb|EAT47702.1	Ubiquitin protein ligase	*Aedes aegypti*	8E-73	93%
243	1	gb|AAF20167.1	Tryptophanyl-tRNA synthetase	*Drosophila melanogaster*	6E-09	64%
244	1	gb|AAF20167.1	Tryptophanyl-tRNA synthetase	*Drosophila melanogaster*	9E-88	64%
245	2	gb|EAT42072.1	Deoxyribonuclease I	*Aedes aegypti*	7E-28	47%
246	1	gb|EAT39810.1	pre-mRNA processing factor	*Aedes aegypti*	3E-12	58%
247	1	gb|ABB54730.1	Argonaute-2	*Drosophila yakuba*	8E-19	68%
248	1	gb|AAA28650.1	Djun gene product	*Drosophila melanogaster*	1E-13	39%

Tabela 7. Proteínas sem função conhecida (Unknown)

Nº do cluster	*Nº ESTs*	*Identificação*	*Proteína*	*Organismo*	*E value*	*Identidade*
249	26	emb|CAM36311.1	Hypothetical protein	*Thermobia domestica*	3E-04	77%
250	12	emb|CAM36311.1	Hypothetical protein	*Thermobia domestica*	2E-05	77%
251	8	ref|NP_648081.1	CG8629-PA	*Drosophila melanogaster*	1E-15	48%
252	6	emb|CAM36311.1	Hypothetical protein	*Thermobia domestica*	2E-05	77%
253	3	gb|AAR10043.1	D. melanogaster CG11151	*Drosophila yakuba*	1E-44	87%
254	2	ref|NP_730777.1	CG32230-PA, isoform A	*Drosophila melanogaster*	5E-34	78%
255	2	ref|XP_001353336.1	GA21218-PA	*Drosophila pseudoobscura*	4E-45	79%
256	1	ref|XP_001359327.1	GA17826-PA	*Drosophila pseudoobscura*	7E-11	48%
257	1	ref|NP_608706.1	CG9894-PB, isoform B	*Drosophila melanogaster*	2E-17	48%
258	1	gb|AAL28701.1	LD12396p	*Drosophila melanogaster*	5E-10	61%
259	1	ref|XP_001361503.1	GA20920-PA	*Drosophila pseudoobscura*	5E-12	50%
260	1	ref|NP_648082.1	CG15829-PA	*Drosophila melanogaster*	9E-08	43%
261	1	ref|NP_611700.2	CG13510-PA	*Drosophila melanogaster*	6E-12	39%
262	1	ref|NP_647706.1	CG13807-PA	*Drosophila melanogaster*	9E-19	68%
263	1	ref|NP_572469.1	CG15343-PA	*Drosophila melanogaster*	1E-23	43%
264	1	ref|XP_001361458.1	GA10191-PA	*Drosophila pseudoobscura*	3E-23	42%
265	1	ref|NP_652443.1	CG13315-PA	*Drosophila melanogaster*	6E-13	60%
266	1	gb|AAA98835.1	Ref(2)P protein	*Drosophila melanogaster*	8E-12	43%
267	1	ref|XP_001352912.1	GA19896-PA	*Drosophila pseudoobscura*	2E-32	87%
268	1	ref|XP_001356996.1	GA18177-PA	*Drosophila pseudoobscura*	3E-62	68%
269	19	***nenhum alinhamento***				
270	7	***nenhum alinhamento***				
271	7	***nenhum alinhamento***				
272	6	***nenhum alinhamento***				
273	6	***nenhum alinhamento***				
274	6	***nenhum alinhamento***				
275	5	***nenhum alinhamento***				
276	4	***nenhum alinhamento***				
277	3	***nenhum alinhamento***				
278	3	***nenhum alinhamento***				
279	3	***nenhum alinhamento***				
280	2	***nenhum alinhamento***				
281	2	***nenhum alinhamento***				
282	2	***nenhum alinhamento***				
283	2	***nenhum alinhamento***				
284	2	***nenhum alinhamento***				
285	2	***nenhum alinhamento***				
286	2	***nenhum alinhamento***				
287	2	***nenhum alinhamento***				
288	1	***nenhum alinhamento***				
289	1	***nenhum alinhamento***				
290	1	***nenhum alinhamento***				
291	1	***nenhum alinhamento***				
292	1	***nenhum alinhamento***				
293	1	***nenhum alinhamento***				
294	1	***nenhum alinhamento***				
295	1	***nenhum alinhamento***				
296	1	***nenhum alinhamento***				
297	1	***nenhum alinhamento***				
298	1	***nenhum alinhamento***				
299	1	***nenhum alinhamento***				
300	1	***nenhum alinhamento***				
301	1	***nenhum alinhamento***				
302	1	***nenhum alinhamento***				
303	1	***nenhum alinhamento***				
304	1	***nenhum alinhamento***				
305	1	***nenhum alinhamento***				

Tabela 7. (continuação)						
Nº do cluster	Nº ESTs	Identificação	Proteína	Organismo	E value	Identidade
306	1	***nenhum alinhamento***				
307	1	***nenhum alinhamento***				
308	1	***nenhum alinhamento***				
309	1	***nenhum alinhamento***				
310	1	***nenhum alinhamento***				
311	1	***nenhum alinhamento***				
312	1	***nenhum alinhamento***				
313	1	***nenhum alinhamento***				
314	1	***nenhum alinhamento***				
315	1	***nenhum alinhamento***				
316	1	***nenhum alinhamento***				
317	1	***nenhum alinhamento***				
318	1	***nenhum alinhamento***				
319	1	***nenhum alinhamento***				
320	1	***nenhum alinhamento***				
321	1	***nenhum alinhamento***				
322	1	***nenhum alinhamento***				
323	1	***nenhum alinhamento***				
324	1	***nenhum alinhamento***				
325	1	***nenhum alinhamento***				
326	1	***nenhum alinhamento***				
327	1	***nenhum alinhamento***				
328	1	***nenhum alinhamento***				
329	1	***nenhum alinhamento***				
330	1	***nenhum alinhamento***				
331	1	***nenhum alinhamento***				
332	1	***nenhum alinhamento***				
333	1	***nenhum alinhamento***				
334	1	***nenhum alinhamento***				
335	1	***nenhum alinhamento***				
336	1	***nenhum alinhamento***				
337	1	***nenhum alinhamento***				
338	1	***nenhum alinhamento***				
339	1	***nenhum alinhamento***				
340	1	***nenhum alinhamento***				
341	1	***nenhum alinhamento***				
342	1	***nenhum alinhamento***				
343	1	***nenhum alinhamento***				

4.3. Identificação da ORF lipolítica ventricular de *Musca domestica*

Através de um sequênciamento randômico de ESTs, a partir de uma biblioteca de cDNA ventricular de *Musca domestica,* identificamos uma sequência contendo uma *ORF* codificante para uma enzima lipolítica (lipase), (ver Tabela 5, cluster número 16).

A escolha dessa sequência chamada LipMD foi baseada no fato dessa sequência ser formada por um maior número de ESTs (9 ESTs), sugerindo uma participação funcional mais expressiva no tubo digestivo de *Musca domestica*. A

```
   1  AGGGGGATATTATTTCGTCTAACCATGAGATATTTCGTTGTATTCGCATTTTGTGTAGCT  60
   1                          M  R  Y  F  V  V  F  A  F  C  V  A   12
  61  GCTGCTTGTGCCTTCGTTATTCCGGAGGAGGAACGCGTCAATGGTGAAAACGGTTGGTAC  120
  13   A  A  C  A  F  V  I  P  E  E  E  R  V  N  G  E  N  G  W  Y   32
 121  GTTCCCCAAGAAGACGGATCCCTGGAATGGGTTAATCGCGATGATGCTGAAGCCTATTTG  180
  33   V  P  Q  E  D  G  S  L  E  W  V  N  R  D  D  A  E  A  Y  L   52
 181  GATCAAGCCCAACAATTCGAGGGTCGTTTAACCACCAATCCTGTTAATTTCTATTTGTAC  240
  53   D  Q  A  Q  Q  F  E  G  R  L  T  T  N  P  V  N  F  Y  L  Y   72
 241  ACCAAGTCCAATCCCAATAACGGACATGAAATCTCGGCTACAAGCTCATCGATAAATGGA  300
  73   T  K  S  N  P  N  N  G  H  E  I  S  A ▲T  S  S  S  I •N  G   92
 301  TCTCCTTTCAATGCCCGCAACCCCACCAGAATCACCATCCACGGTTGGACTTCCAGCAAA  360
  93   S  P  F  N  A  R •N  P  T  R  I  T  I  H  G  W  T  S  S  K   112
 361  GATGATTACGTGAACTATGGTATTCGTGATGCCTACTTGGCAAAGGGTGACTACAACATG  420
 113   D  D  Y  V  N  Y  G  I  R  D  A  Y  L  A  K  G  D  Y  N  M   132
 421  ATTACCGTTGACTGGGGTCGTGCCCGTTCTGTTGATTATGCCTCATCGGTCTTGGCTGTT  480
 133   I  T  V  D  W  G  R  A  R  S  V  D  Y  A  S  S  V  L  A  V   152
 481  CCTGGTGTTGGTCAAAAAGTTGCCGCCTTGGTCGATTTCTTGGTTAAGAACCACGGCTTG  540
 153   P  G  V  G  Q  K  V  A  A  L  V  D  F  L  V  K  N  H  G  L   172
 541  AAATTGGCTGATTTGGAAATTATTGGTCACAGCTTGGGTGCCCATGTTGCTGGTTTTACC  600
 173   K  L  A  D  L  E  I  I  G  H  S  L  G  A  H  V  A  G  F  T   192
 601  GGCAAGAACATTGCCAGTGGTCAAGCTCATGCCATTATCGGTTTGGACCCTGCCCTTCCT  660
 193   G  K  N  I  A  S  G  Q  A  H  A  I  I  G  L  D  P  A  L  P   212
 661  TTGTACAGCTACGACAAACCAACTCAACGTTTATCTAGCACCGATGCCTACTATGTTGAA  720
 213   L  Y  S  Y  D  K  P  T  Q  R  L  S  S  T  D  A  Y  Y  V  E   232
 721  TCTATCCAGACCAACGGTGGTACTTTGGGTTTCTTGAAGCCCATCGGTAAAGGTGCTTTC  780
 233   S  I  Q  T  N  G  G  T  L  G  F  L  K  P  I  G  K  G  A  F   252
 781  TATCCCAACGGTGGCAAGACTCAACCTGGTTGCGGCGTTGATCTTACCGGTAGCTGTTCC  840
 253   Y  P  N  G  G  K  T  Q  P  G  C  G  V  D  L  T  G  S  C  S   272
 841  CATAGCCGCGCTGTTACCTACTACGTCGAAGCCGTTTCCATTAATGATTTCGGCACCATT  900
 273   H  S  R  A  V  T  Y  Y  V  E  A  V  S  I  N  D  F  G  T  I   292
 901  AAGTGTCCAAACTATGAAGATGCTGTTGCCAAGAGCTGTGGTTCCACTTACAGTTCCGTT  960
 303   K  C  P  N  Y  E  D  A  V  A  K  S  C  G  S  T  Y  S  S  V   312
 961  CGTATGGGTGCTACCTCTAATGCCTATATGGTCGCTGGTGATTTCTATGTTCCAGTCAAT  1020
 323   R  M  G  A  T  S  N  A  Y  M  V  A  G  D  F  Y  V  P  V  N   332
1021  AGCAAGTATCCTTATGGTGTATTGGGTTAATTTGGTTATGTTGACATGCTAAGTTTTGAA  1080
 333   S  K  Y  P  Y  G  V  L  G  *                                 341
1081  AATAAAGTATGATTGATTAGAGCCAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1140

      AAAAAA                                                        1146
```

Figura 33. Sequência de aminoácidos deduzidas a partir da sequência de nucleotídeos para a LipMD. O peptídeo sinal está em vermelho. Os resíduos de aminoácidos da tríade catalítica (Ser^{183}, Asp^{208} e His^{273}) estão em azul. O pentapeptídeo conservado (Gly – Xaa – Ser – Xaa – Gly) na família das α/β hidrolases está sublinhado. Os prováveis sítios de N-glicosilação estão marcados pelos pelo símbolo (•) e o sítio de O-glicosilação pelo símbolo (▲). O sinal de poliadenilação está marcado em negrito.

sequência de nucleotídeos da LipMD e sua sequência de aminoácidos deduzidas são mostrada na figura 33. Análise desta sequência mostra que este cDNA contém uma *ORF* codificante para a LipMD possuindo 1146 pb de comprimento. O códon inicial está predito entre os nucleotídeos 25 – 27 e seu *stop* códon localizado entre os nucleotídeos 1048 – 1050. O cDNA da LipMD possui um *ORF* localizada entre os nucleotídeos 25-1050 codificando uma proteína contendo 341 aminoácidos com massa molecular teórica de 36,82 kDa. A extremidade 3' não traduzida contém 96 pb e termina com uma cauda de poli A composta por 40 nucleotídeos começando 20 pb depois do sinal de poliadenilação (AATAAA) em eucarioto.

Um peptídeo sinal foi identificado sendo seu sítio de hidrólise entre os resíduos Ala^{16} e Phe^{17} resultando em uma proteína madura com 325 aminoácidos com uma massa molecular e ponto isoelétrico teóricos de 35,06 kDa e 5,5 respectivamente. Análise estrutural revela que a proteína deduzida possui os resíduos Ser^{183}, Asp^{208} e His^{273} como tríade catalítica típica das proteínas pertencentes à família das α/β hidrolases na qual incluem-se as lipases e esterases. O resíduo catalítico Ser^{183} que atua como nucleófilo encontra-se no interior do pentapeptídeo conservado Gly – X – Ser^{183} – X – Gly sendo um motivo característico das enzimas pertencentes a superfamília das α/β hidrolases. A sequência de aminoácidos deduzida possui 2 potenciais sítios de N-glicosilação nos resíduos de aminoácidos nas posições N_{91} e N_{99}. Possui também um potencial sítio de O-glicosilação no resíduo T_{86}.

Alinhamentos por BLASTX (http://blast.ncbi.nlm.nih.gov/Blast.cgi) e por domínios conservados (http://www.ebi.ac.uk/Tools/InterProScan/) revelou que o cDNA codificante para a LipMD compartilha alta identidade com lipases de insetos depositadas no banco de dados. A LipMD possui 63%, 64% e 62% de identidade com três sequências de aminoácidos deduzidas para lipases de *Drosophila melanogaster* (CG6271, CG6277 e CG6283 respectivamente (Zinke et. al., 2002)), 31% de identidade com uma lipase de *Aedes aegypti* (XP001652527.1) e 30% de identidade com uma lipase de *B. mori* (BAC00960.1, Ponnuvel *et al.,* 2003) (Figura 34).

```
LipMD   1   ----------------------------------------------MRYFVVFAFCVAAACAFVIPEEERV 25
LipDM1  1   ------------------------------------------------------------------LEESERV 7
LipDM2  1   ------------------------------------------------------------------IKESERV 7
LipDM3  1   -----------------------------------------------------------------PIEERV 6
LipAA   1   MRLLALFLSVAVAVAVALPIEELGPESVPFEEYGQEHIPNEEIGEEFIPNEDEGIEHIPYEDEGIEHMPV 70
LipBM   1   ---------------------------------------------------------------------- 1

LipMD   26  NGENGWYVPQEDGSLEWVNRDDAEAYLDQAQQFEGR-LTTNPVNFYLYTKSNPNNGHEISATSSSINGSP 94
LipDM1  8   NGENGWYVPQQDGSFEWVDMDVAEEWMEAQEKLEGRGLTTVPVNFYLFTPKNPSSSKHIYATTKSISKSN 77
LipDM2  8   HGENGWYVPQEDGTSEWVDMDVAEQWMEAQELLESRGLTTVPVKFYLYTSSNPTKGKKITASTKSIDASS 77
LipDM3  7   NGENGWFIPKLDGSFEWMDMQDAEDLLANGAQMEGR-ISTNAVNFYVYTKSNPTDGKEIKAKSGSVEDSH 75
LipAA   71  EQSPWSLVPDGNGHLHLVNTDPYNLAGQEEPEAYFNADTQTIFRLFTRANRNTNGHVIQLGNAGSLG-SF 139
LipBM   1   -------MPDGEGVPHLVDLEEP----AEEDILMSRNGANNQYWLFTRRNQN-NHQVITNGNVNSIRNSN 58

LipMD   95  FNARNPTRITIHGWTSSKDDYVNYGIRDAYLAKGDYNMITVDWGRARS-VDYASSVLAVPGVGQKVAALV 163
LipDM1  78  FNPAHPTRFVIHGWTQSYLNSMNSDIRKAFLSKGDYNVIVVDWARARS-VDYATSVMAVAATGKKVAKMI 146
LipDM2  78  FNSAHPTRFVIHGWTQSYTASMNKDIRSAWLSRGDYNVIVVDWARARS-VDYATSVLAVAATGKKVAKMI 146
LipDM3  76  FNKDHGTRFVIHGWTQRYSDDMNTRITKAWLSKGDYNVIVVDWARARS-VDYASSVLAVPGAGGKVGEMI 144
LipAA   140 WNGGRQTRLIIHGWNNNGGSPVNTQIRNAYLDRGDFNVITVDWGAGAQNPNYLTARNHINAVGATVARFV 209
LipBM   59  YNGNLPLFVIVHGWNSNGNSAVNTMIRPALLAVSDCNVIVVDW-RGLANGLYNTAVNGVPSVGQFLGNFL 127
                          *                   *
LipMD   164 DFIVKNHGLKLADLEIIGHSLGAHVAGFTGKNIASGQAHAIIGLDPALPLYSYDKPTQRLSSTDAYYVES 233
LipDM1  147 NFLKDNHGLNLNDVYVIGHSLGAHVAGYAGKNT-DGQVHTIIGLDPALPLFSYNKPNKRLNSDDAWYVES 215
LipDM2  147 NFLKDNHGLNLNDLYVIGHSLGAHVAGYAGKNT-DGQVHTIIGLDPALPLFSYNKPNKRLNSDDAWYVES 215
LipDM3  145 KYLHDHHGLDYDSLEVIGHSLGAHVAGYAGKTVGDKRVHTIVGLDPALPLFSYDKPAKRLSTDDAHYVES 214
LipAA   210 DFINQSGGMSFNNVYITGHSLGGHIAGIAGKRVTRGRLHSVIALDPALPLFSINAPNERVAPTDANYVEV 279
LipBM   128 VWLINNGGGNWGRVHLIGFSLGAHVVGNAGRQAG-GRPNRVTGLDPAGPRWGGN--NQALNRNAGAYVEA 194
                                                *
LipMD   234 IQTNGGTLGFLKPIGKGAFYPNGGKT-QPGCGVDLTGSCSHSRAVTYYVEAVSINDFGTIKCPNYEDAVA 302
LipDM1  216 IQTNGGTLGFLKPIGKGAFYPNGGKT-QPGCPLDVTGACSHGRSTTYYAEAVSEDNFGTMKCGDYEEAVA 284
LipDM2  216 IQTNGGTLGFLKPIGKGAFYPNGGKT-QPGCGLDLTGACSHGRSTTYYAEAVSEDNFGTMKCGDYEEAVS 284
LipDM3  215 IQTNGGKLGFLKPIGKGAFYPNGGKS-QPGCGLDATGSCSHARSVLYYAEAVTEDNFGSIKCHDYEDAVA 283
LipAA   280 IHTNAGLLGFDLPIGQADFYPNGGRS-QPGCGVDVAGTCAHSRAWEFFAESVRTGRFNSVRCANYDQILN 348
LipBM   195 IHTDGGLLGIFDRIAHGDFYPNGGRNPQPGCRVS---TCSHSRAYELYASTVRHNRFVGRLCNNLNQAQN 261

LipMD   303 KSCGSTYSSVRMG-ATSNAYMVAGDFYVPVNSKYPYGVLG 341
LipDM1  285 KECGSTYSSVRMG-ADTNAYMVEGDFYVPVNSKAPFGMI- 322
LipDM2  285 KECGSTYSSVRMG-ADTNAYMVEGDYYVPVNSKAPFGMI- 322
LipDM3  284 KNCGSTYSSVRMG-AITNAYMVEGDFYVPVNSEAPFGKI- 321
LipAA   349 NNCVSSGPNRNMGGEPSNIGIASGVYHLTTNAGSPFARG- 387
LipBM   262 NQCS--GGTFNMGNAVFGK-RGNGIYGLRTGSSWPF---- 294
```

Figura 34. Alinhamento das sequências de aminoácidos de várias proteínas pertencentes a família das α/β hidrolases similares a lipases. As proteínas são originárias de Musca domestica (LipMD, sequência não depositada no banco de dados), LipDM1 (CG6271), LipDM2 (CG6277) e LipDM3 (CG6283) de Drosophila melanogaster (Zinke et al., 2002), LipAA (XP001652527.1) de Aedes aegypti, LipBM (BAC00960.1) de B. mori. A tríade catalítica composta pelos aminoácidos Ser, Asp e His característicos das proteínas das α/β hidrolases estão marcados pelos asteriscos.

4.3.1. Clonagem e expressão da LipMD recombinante

A sequência de cDNA contendo a *ORF* codificante para a LipMD madura foi amplificada através de uma reação de polimerização em cadeia (PCR) usando como molde uma biblioteca de cDNA ventricular de *Musca domestica* juntamente com os oligonucleotídeos LipMDNt_PstI e LipMDCt_HindIII (ver item 2.11. e Tabela 2), (Figura 35).

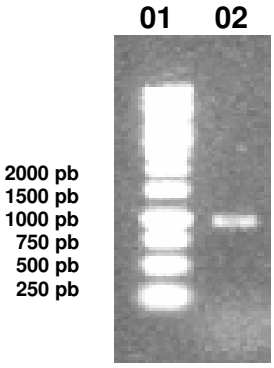

Figura 35. Eletroforese em gel de agarose 1% do produto de PCR onde foram utilizados os primers específicos LipMDNt_PstI e LipMDCt_HindIII, para a amplificação da LipMD madura, tomando como molde uma biblioteca de cDNA de Musca domestica. O gel foi tratado com brometo de etídeo e as bandas de cDNA visualizadas com luz ultravioleta. 01. Padrão de DNA, 02. cDNA codificante para a LipMD madura contendo os sítios de restrição Pst I e Hind III.

O produto do PCR foi digerido com as enzimas *Pst I* e *Hind III* e posteriormente clonado no vetor pAE (Ramos *et al.,* 2004), previamente digerido com as mesmas enzimas, resultando no plasmídeo pAE/LipMD (Figura 36). A construção pAE/LipMD foi usado para transformar *E. coli* (BL21(DE3)pLysS), (que produzem T7 RNA polimerase de forma indutível por IPTG), podendo produzir de forma induzida a proteína recombinante. As bactérias transformadas foram crescidas a 20ºC em 500 mL de LB líquido, contendo ampicilina 100µg/mL até uma Abs_{600} entre 0,6-1,0. Essas bactérias foram induzidas com IPTG na concentração final de 0,5 mM em temperatura de 20ºC por 17 horas. As proteínas das células controle (células transformadas apenas com o plasmídeo pAE sem inserto), não induzidas e induzidas foram analisadas também em SDS-PAGE (Figura 37). Nota-se que a indução com IPTG levou a intensa produção de uma proteína de massa molecular de 37,3 kDa que deve corresponder a proteína LipMD. Portanto a linhagem pAE/

LipMD/BL21(DE3)pLysS, parece ser um sistema de expressão adequado para a LipMD.

A

B

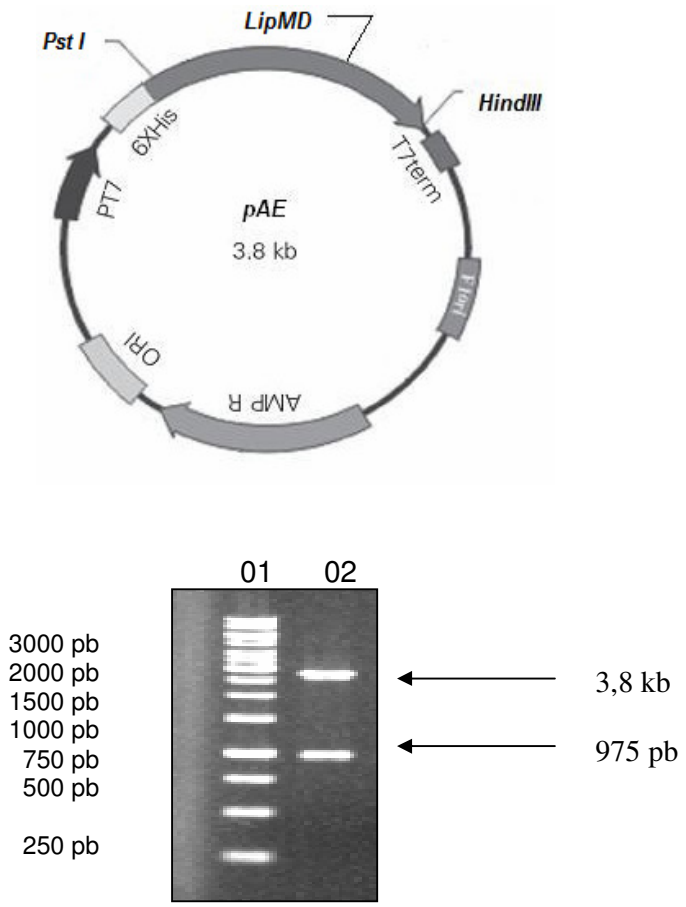

Figura 36. A. Esquema da construção pAE/LipMD que consiste no plasmídeo pAE (Ramos et al., 2004) com o cDNA que codifica a LipMD madura inserido. B. Eletroforese em gel de agarose 1% da digestão da construção pAE/LipMD com as enzimas de restrição Pst I e Hind III. 01. Padrão de DNA, 02. produto da digestão.

Cem mL de células induzidas à temperatura de 20ºC por 17 horas foram centrifugadas e seu sedimento ressuspenso em 1mL de tampão A (ver item 2.14.). As células foram lisadas por sonicação (ver item 2.13.) e centrifugadas a 7000g, 4°C por 30 minutos. O sobrenandante recuperado corresponde ao material solúvel, enquanto que o precipitado, ao material bacteriano insolúvel. Conforme pode ser visto na figura 38, o SDS-PAGE revela que a proteína recombinante produzida

(LipMD) é recuperada na forma solúvel quando a indução ocorre nas condições mensionadas. A partir da fração solúvel obtida, a proteína recombinante foi injetada em uma coluna cromatográfica de afinidade (HisTrap HP, 1 mL, GE) que se liga a cauda de histidina fusionada à extremidade N-terminal da LipMD. A proteína recombinante foi então eluída na presença de um tampão B, contendo uma alta concentração de imidazol (ver item 2.14.). Uma alíquota de cada fração coletada foi ensaiada tendo como substrato pNP-palmitato dissolvida em tampão fosfato de sódio 100mM pH 7,0 (ver item 2.24.). Frações da cromatografia que apresentaram atividade recombinante foram recolhidas e dialisadas em tampão Tris-HCl 5 mM pH 8,0 por um período de tempo de 18 horas em câmara fria, visando a eliminação do tampão na qual foram eluídas.

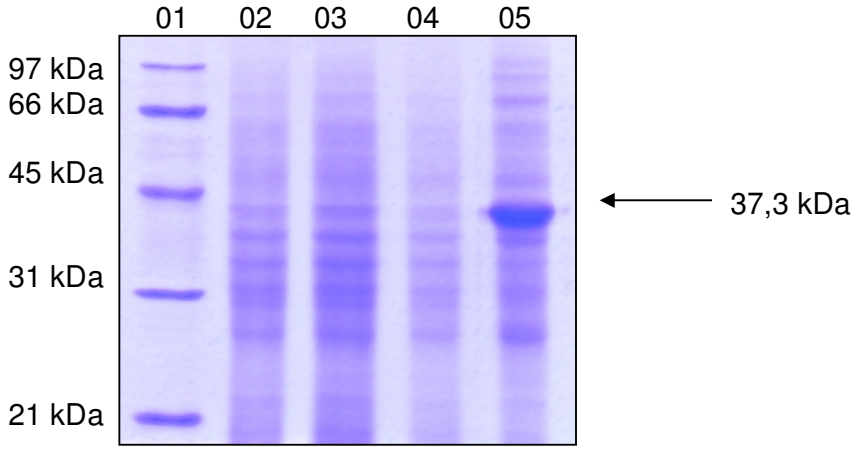

Figura 37. Análise em SDS-PAGE da expressão da LipMD em células BL21(DE3) pLysS. A indução da expressão da enzima e do controle foram feitas com IPTG 0,5 mM a temperatura de 20ºC por 17 horas. Temos nas seguintes raias: 01. padrão de massa molecular, 02. pAE não induzido, 03. pAE induzido a 20ºC (17 horas), 04. pAE/LipMD não induzido 05. pAE/LipMD induzido a 20ºC (17 horas). A seta indica a LipMD produzida. O gel continha poliacrilamida (12%) e foi corado com Coomassie Blue R-250.

Uma alíquota dessa amostra dialisada foi analisada em SDS-PAGE, onde verificamos o sucesso do procedimento de purificação da enzima recombinate (Figura 39). Análise da estrutura secundária por dicroismo circular para a LipMD na

forma pura, estimou que a enzima recombinante possui 37% de alfa hélices, 26% de folhas betas e 38% de *random coil* (Figura 40).

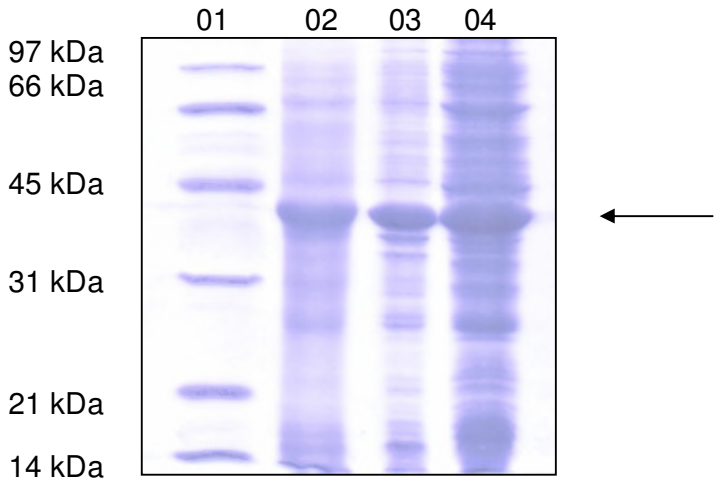

Figura 38. Análise em SDS-PAGE das bactérias BL21(DE3)LysS contendo a contrução pAE/LipMD induzidas por IPTG 0,5 mM, a uma temperatura de 20ºC durante 17 horas. As bactérias induzidas foram lisadas e centrifugadas sendo que o precipitado e o sobrenadante foram aplicados em SDS-PAGE (12%). Temos nas seguintes raias: 01. padrão de massa molecular, 02. Bactérias induzidas, 03. precipitado das bactérias induzidas, lisadas e centrifugadas, 04. sobrenadante das bactérias induzidas, lisadas e centrifugadas. A seta indica a presença da LipMD recombinante na fração solúvel do lisado bacteriano.

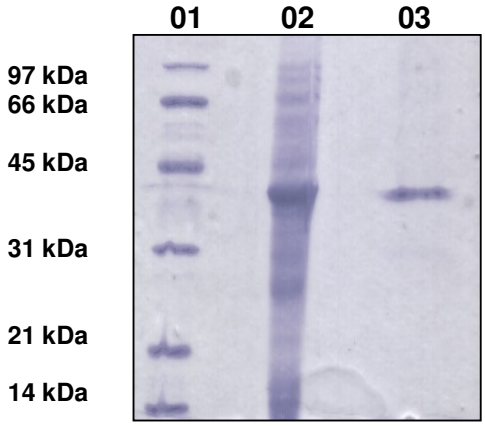

Figura 39. Análise em SDS-PAGE das frações da cromatografia de afinidade para purificação da LipMD recombinante. Esta enzima foi expressa com o peptídeo de fusão ligado a extremidade N-terminal da proteína utilizando o vetor pAE (Ramos et

al., 2004) e foi purificada através de uma cromatografia de afinidade em coluna de Ni^{2+} (HisTrap HP, 1mL, GE). Raia 01. padrão de massa molecular, 02. bactérias induzidas (pAE/LipMD/BL21(DE3)pLysS), 03. LipMD purificada. O gel continha poliacrilamida (12%) e foi corado com Coomassie Blue R-250.

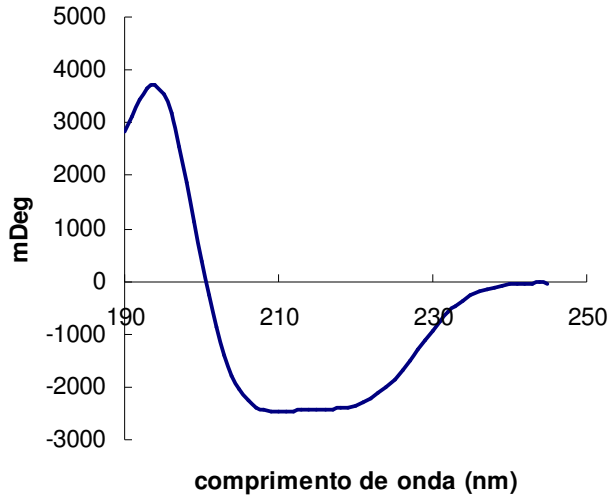

Figura 40. Espectro de dicroismo circular da enzima recombinante LipMD.

4.3.2. Efeito do pH e da temperatura sobre a atividade da LipMD

Usando pNP-miristato como substrato, a LipMD recombinante mostrou-se mais ativa em pH 7,5, apresentando-se estável na região do pH ótimo e com baixa atividade abaixo de pH 6 e acima de 9 (Figura 41). Quando examinado o efeito da temperatura sobre a atividade em pH 7,5, uma máxima atividade é observada a 42°C quando a duração do ensaio era de 60 minutos (Figura 42). Após a pré-incubação a 53°C por 20 minutos, a enzima apresentou uma atividade relativa em torno de 50%, indicando que o tempo de meia vida da atividade dessa enzima é de 20 minutos a 53°C (Figura 43). Após o tratamento com solventes orgânicos (50% v/v), a atividade lipolítica da enzima ficou estável em acetonitrila, com um aumento da atividade em relação ao solvente controle água (provavelmente o acetonitrila aumenta a solubilidade do substrato que se encontrava na forma de emulsão), mas sua atividade diminuiu de forma significativa durante o tratamento com os solventes metanol, etanol, 2-propanol e acetona (Figura 44).

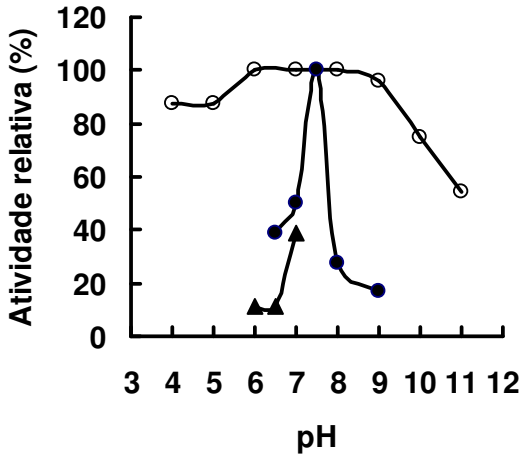

Figura 41. Efeito do pH sobre a atividade da LipMD recombinante (•) tendo como substrato pNP-miristato (150µM). Para cada valor de pH a atividade foi analisada a 30°C. A atividade relativa é expressa como porcentagem em relação a maior atividade obtida sobre a faixa de pH testada. A curva de estabilidade (○) frente ao pH representa a atividade remanescente após a pré incubação no período de 1 hora nos respectivos pHs indicados.

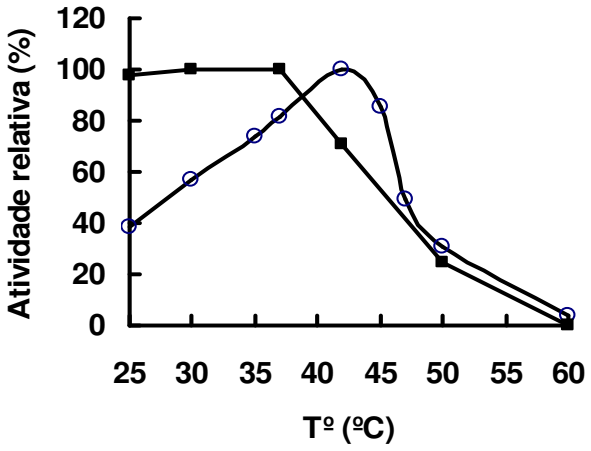

Figura 42. A. Efeito da temperatura sobre a atividade (○) e estabilidade (■) da LipMD recombinante. A atividade enzimática foi analisada para cada temperatura mostrada no gráfico. Para os experimentos de estabilidade térmica a enzima foi incubada nas respectivas temperaturas pelo período de 1 hora e então a atividade remanescente foi determinada.

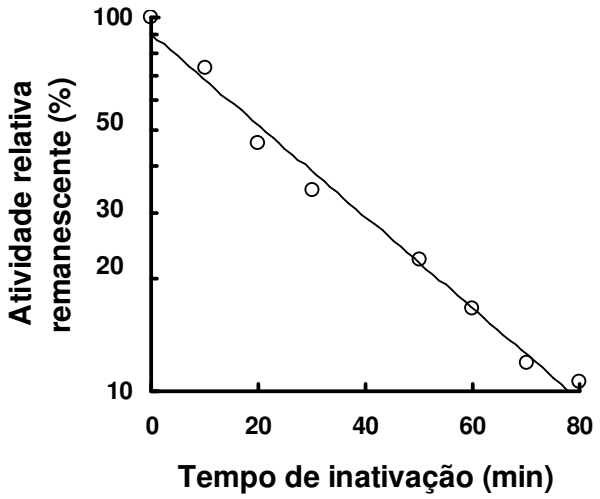

Figura 43. Inativação térmica da LipMD recombinante purificada. A inativação térmica foi realizada na temperatura de 53ºC sendo que a atividade remanescente foi detectada com pNP-miristato 150 µM em tampão fosfato de sódio 100 mM pH 7,5. A enzima recombinante apresentou um tempo de meia vida de aproximadamente 20 minutos nessas condições.

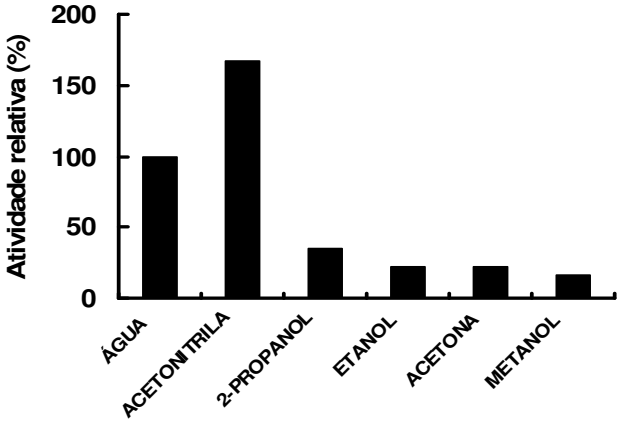

Figura 44. Estabilidade da LipMD frente a vários solventes orgânicos (50% v/v). A atividade relativa é expressa como porcentagem em relação ao solvente água.

4.3.3. Especificidade pelo substrato da LipMD recombinante

A atividade da LipMD recombinante purificada foi claramente detectada por análise em placas ágar-tributirina (Figura 45). Contudo nenhuma atividade lipolítica foi detectada através do método específico e sensível com placas de Rhodamina B / trioleína. Em correlação com esses resultados a LipMD foi capaz de hidrolisar um grande número de pNP-acil-ésteres (pNP-acetato (2C) à pNP-estearato (18C)). A LipMD mostrou alta atividade sobre pNP-acil-ésteres de curta cadeia carbônica (4C e 5C) e baixa atividade para pNP ésteres de média (8C - 12C) e longa (14C - 18C) cadeia carbônica (Figura 46). Nossos resultados parecem indicar que a enzima LipMD não seja uma verdadeira lipase, mas sim uma carboxiesterase, embora que a enzima possua uma ampla especificidade sobre pNP- acil-ésteres (C2 a C18).

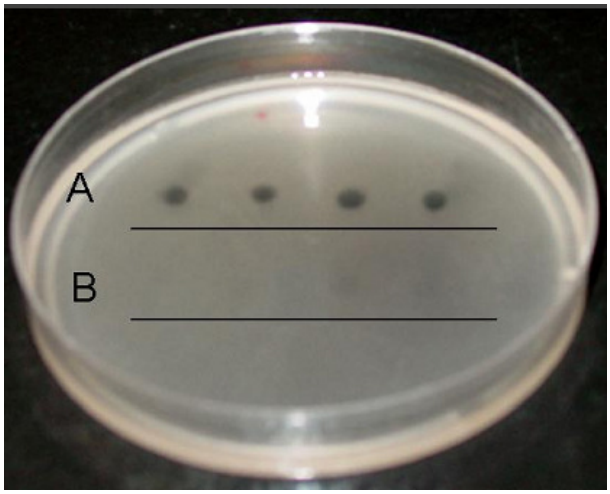

Figura 45. Identificação da atividade lipolítica da LipMD sobre placa de ágar-tributirina. A. Incubação (em cada ponto) de 2 µL (0,35µg) da LipMD ativa. B. Incubação (em cada ponto) de 2 µL (0,35µg) da LipMD desnaturada (fervida por 10 minutos a 100°C). A placa foi incubada por 20 minutos a 30°C. A formação dos halos translúcidos indicam a hidrólise da tributirina pela LipMD recombinante.

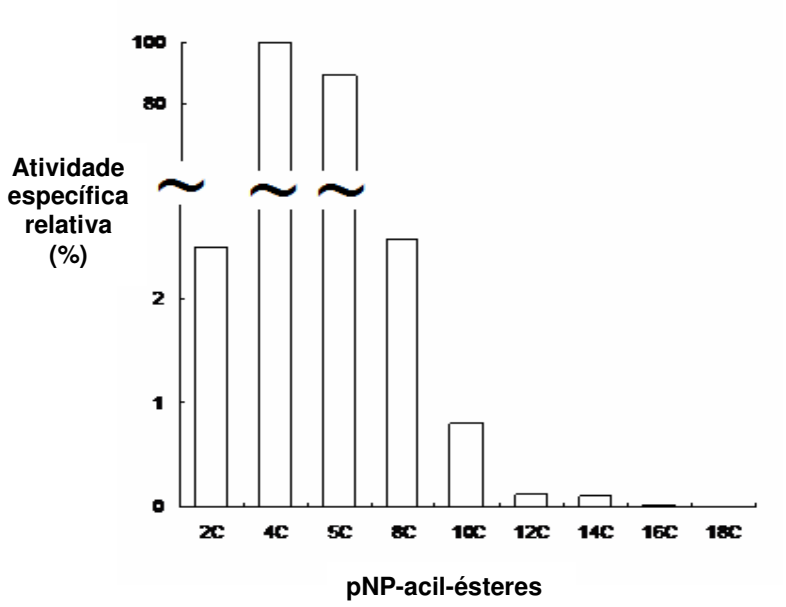

Figura 46. Atividade específica relativa para a LipMD recombinante tendo como substratos pNP-acil-ésteres: pNP-acetato, pNP-butirato, pNP-valerato, pNP-caprilato, pNP-caprato, pNP-laurato, pNP-miristato, pNP-palmitato e pNP-estearato.

4.3.4. Análise da expressão e localização da LipMD

Para determinar a expressão e a localização dos transcritos correspondentes a LipMD no tubo digestivo de larvas de *Musca domestica,* análise por RT-PCR foi executada. Os resultados revelam que os transcritos correspondentes para a LipMD foram expressos de forma predominante no ventrículo anterior e pouco expresso no ventrículo médio. Já nos tecidos, ventrículo posterior, hemolinfa, corpo gorduroso e túbulos de Malpighi não foram verificada a presença desses transcritos (Figura 47).

Adicionalmente análise por *Western blot* demonstrou o reconhecimento do soro anti-LipMD pela LipMD recombinante. Reconheceu também uma outra proteína contidas no conteúdo luminal do ventrículo anterior de larvas de *Musca domestica*, de massa molecular semelhante à LipMD recombinante (Figura 48).

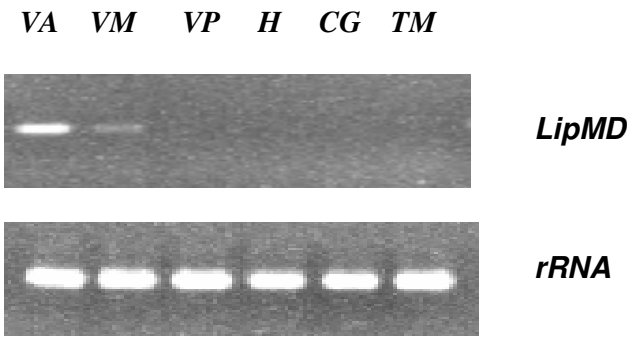

Figura 47. Padrão de expressão dos mRNAs codificantes para a LipMD em diferentes tecidos intestinais de larvas de Musca domestica são mostrados por RT-PCR semi-quantitativo. Iguais quantias de RNA total de diferentes tecidos foram extraídos e usados como amostra em uma reação de RT. O produto dessas reações foram usados em uma reação de PCR usando primers específicos para a LipMD e para o rRNA (AY123346.1) de Musca domestica (ver Tabela 2). Os produtos da amplificação foram analisados em gel de agarose 1%. As bandas amplificadas apareciam nos respectivos tecidos quando os correspondentes mRNAs eram transcritos. VA, ventrículo anterior; VM, ventrículo médio; VP. ventrículo posterior; H, hemolinfa; CG. corpo gorduroso; TM, túbulos de Malpighi. O número de ciclos durante o PCR foi escolhido após várias tentativas, assim que a amplificação estava em fase log e resultando em uma clara e visível banda.

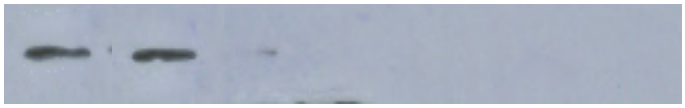

Figura 48. Análise por Western Blot após SDS-PAGE (12%) das proteínas do conteúdo luminal marcadas com o anticorpo anti-LipMD. A detecção padrão com anti-IgG (coelho) acoplado com peroxidase, revelados com "Imobbilon Western Chemiluminescent HRP substrate" (Millipore) e visualizado após exposição com o filme "High-performance luminescence detection film" (GE). CVA. conteúdo luminal do ventrículo anterior, CVM. conteúdo luminal do ventrículo médio, CVP. conteúdo

luminal do ventrículo posterior, HM. hemolinfa, CG. corpo gorduroso, TM. túbulos de Malpigh. Em cada linha foram aplicadas 20 μg de proteínas totais.

4.3.5. Imunocitolocalização da LipMD em Musca domestica

O anticorpo produzido (anti-LipMD) foi usado em uma imunocitolocalização em microscopia eletrônica de transmissão dessa enzima no ventrículo de larvas de *Musca domestica*. As marcações são evidenciadas por partículas de ouro coloidal ligadas ao anticorpo secundário que reconhece o anticorpo anti-LipMD. As marcações ocorreram de forma predominante no interior de pequenas vesículas de secreção localizadas no ápice das células do ventrículo anterior (região 1, ver Figura 5) próximo às microvilosidades (Figura 49). Nenhuma marcação foi verificada nas outras regiões analisadas. Pelo resultado obtido é possível verificar que o ventrículo anterior de larvas de *Musca domestica* é o grande responsável pela secreção da LipMD. Também fica claro que essa enzima é secretada pelo animal, descartando a possibilidade dessa enzima originar-se de bactérias comensais.

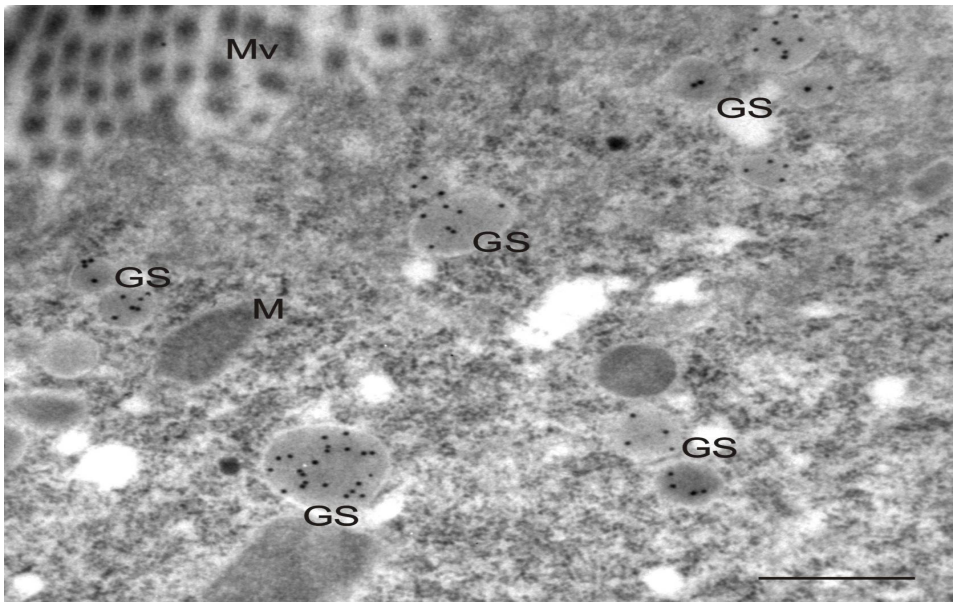

Figura 49. Imunocitolocalização por microscopia eletrônica de transmissão usando o anticorpo anti-LipMD nas células do ventrículo anterior de larvas de Musca domestica. GS: grânulos de secreção, Mv: Microvilosidades, M: mitocôndria. A escala apresentada equivale a 0,5 μm. Marcações são evidenciadas com os

anticorpos em grânulos de secreção localizadas no ápice das células próximo a microvilosidades. Nenhuma marcação foi observada nos controles com o soro pré-imune ou quando o anticorpo primário (anti-LipMD) foi omitido.

4.3.6. Distribuição de enzimas lipolíticas em larvas de Musca domestica

Larvas de *M. domestica* foram dissecadas em gelo e analisadas a atividade lipolítica presentes em seu conteúdo luminal nos ventrículos anterior, médio e posterior. Para isso foram analisando em dois substratos: pNP-butirato (substrato com cadeia carbônica curta e solúvel em água) e pNP-palmitato (substrato de cadeia carbônia longa formando uma solução emulsificada por detergente (Triton X-100)) e os valores das atividades encontrada (mU/animal) estão demonstrada na tabela 8. Verificamos uma maior atividade lipolítica (esterase/lipase) para ambos substratos analisados no conteúdo luminal do ventrículo anterior.

Tabela 8. Atividade lipolítica (mU/animal) encontrada no conteúdo luminal de larvas de Musca domestica.

Substrato	***ventrículo anterior***	***ventrículo médio***	***ventrículo posterior***
pNP-butirato	8,9 ± 0,9	0,6 ± 0,1	1,5 ± 0,3
pNP-palmitato	1,5 ± 0,2	0,32 ± 0,05	1,0 ± 0,1

Proteínas dos respectivos conteúdos luminais de 10 larvas foram solubilizadas em 500 µL água bidestilada. Condições de análise: pNP-butirato ou pNP-palmitato (150µM) dissolvidos em tampão fosfato de sódio 100 mM pH 7,5 contendo deoxicolato de sódio e goma arábica solubilizados em Triton X-100 (0,36%). A atividade demonstrada refere-se a mU/animal. Os valores são a média ± desvio padrão da média, baseados em três diferentes preparações.

5. DISCUSSÃO

5.1. Estudo das aspártico proteinases intestinais de *Musca domestica*

Neste trabalho nós clonamos e sequênciamos três aspártico proteinases similares a catepsinas-D (ppCAD1, ppCAD2 e ppCAD3), do ventrículo total de *Musca dometica*.

A atividade de proteinase ácida em larvas de *Musca domestica* ocorre principalmente no epitélio e conteúdo do ventrículo anterior e médio correspondendo à metade do ventrículo total da larva (Espinoza-Fuentes e Terra, 1987). Essa atividade proteolítica apresenta um pH ótimo entre 3,0 – 3,5. Essa atividade é abolida por pepstatina não sendo afetada por inibidor de tripsina de soja e inibidores de cisteína proteinases. Essa proteinase ocorre em múltiplas formas com massa molecular que pode variar de 40 – 80 kDa e com valores de pI próximo de 5. Essa enzima reage de forma lenta com diazoacetilnorleucina com $CuSO_4$ e hidrolisa hemoglobina melhor que albumina (Lemos e Terra, 1991). Essas propriedades são similares a catepsinas-D (Barrett *et al.,* 2004). Essa atividade pode ser originada a partir de lisossomos, como achado para outras enzimas com propriedades similares (Barrett, 1977). Porém, a digestão extracelular ácida de proteínas ocorre em vertebrados pela ação de uma catepsina-D ou uma enzima evoluída a partir de um mesmo gene ancestral tal como pepsina.

Tang e Wong (1987) sugerem que todas as aspártico proteinases evoluíram a partir de uma enzima primordial comum. Assim, um cladograma das sequências de aminoácidos das catepsinas-D de insetos e vertebrados foi construído para predizer a relação filogenética entre as ppCADs e outras aspártico proteinases.

No cladograma resultante, ocorre um grande grupo (*bootstrap* 89) que se divide em duas ramificações: uma com sequências de catepsinas-D de vertebrados (*bootstrap* 100) e outra de insetos (*bootstrap* 99). Nessa última ramificação estão presentes a sequência ppCAD1 e a sequência de *A. aegypti*, a catepsina-D lisossomal mais bem estudada em insetos (Cho e Raikhel, 1992). Apenas pepsinogênio humano, ppCAD2 e ppCAD3 e a sequência de *D. melanogaster* – 2 foram excluídos daquele grande grupo, indicando uma função não lisossomal para essas sequências. Esses dados estão de acordo com os achados de Lemos e Terra (1991), que sugerem que as catepsinas-D digestivas de *M. domestica* e de outros

Diptera Cyclorrapha resultaram do mesmo gene ancestral da catepsina-D lisossomal.

Das três CADs, CAD3 deve corresponder a catepsina-D digestiva presente no conteúdo luminal do ventrículo anterior e médio devido: (1) CAD3 possui um pI teórico de 3,99 e massa molecular de 36,5 kDa (que pode formar dímero como a catepsina-D lisossomal de mosquito (Cho e Raikhel, 1992) e parece ser a CAD mais expressa como visto nas frequência de ESTs achados durante a varredura das sequências com *primers* degenerados. (2) Baseado na análise por RT-PCR, os mRNAs codificantes para a CAD3 é expresso apenas no ventrículo anterior e porção proximal do ventrículo médio, onde a atividade de catepsina-D é encontrada em larvas do inseto. (3) A pCAD3 após auto-ativação em condições ácidas (condições otimizadas: pH 3,5 por 30 minutos, 30°C) hidrolisa hemoglobina desnaturada com um pH ótimo de 2,5 e MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2 com um pH ótimo de 3,0, ambos próximos do pH luminal do ventrículo médio de larvas de *M. domestica* onde essa enzima é ativa. Esse deslocamento do pH ótimo com hemoglobina desnaturada por ácido usada como substrato, foi também observada em outras catepsinas-D (Barrett *et al.,* 2004). (4) Imunoblots das proteínas de diferentes tecidos marcados com o anticorpo policlonal preparado contra a pCAD3 recombinante, revela que a tradução ocorre apenas no ventrículo anterior e médio, onde a enzima é achada em ensaios enzimáticos em homogenizadas de larvas de *M. domestica*. (5) CAD3 é localizada através de ensaios de imunocitolocalização (utilizando-se a técnica de imuno-ouro) no interior de vesículas de secreção no ápice das células próximo as microvilosidades (no interior da célula) no ventrículo anterior e porção proximal do ventrículo médio. Baseados nessas informações, CAD3 pode ser classificada como uma enzima digestiva. Assim como a CAD3, a CAD2 pode ser considerada também como uma enzima digestiva, porém expressa de forma minoritária.

O pH de 3,1 no ventrículo médio de larvas de *Musca domestica* (Regel e Terra, 1995) favorece a atividade da CAD3 (pH ótimo 2,5 com hemoglobina des naturada e pH 3,0 com o substrato MOCAc-GKPILFFRLK-(Dnp)-*D*-R-NH_2) que pode corresponder a atividade proteolítica de aspártico proteinase encontradas no conteúdo luminal de larvas de *M. domestica* (Espinoza-Fuentes e Terra, 1987; Lemos e Terra, 1991).

A molécula da catepsina-D é composta por dois lobos que são muito similares entre si. Essa homologia entre os dois lobos é refletida pelo fato dos dois ácidos aspárticos catalíticos ocorrerem em regiões muito conservadas no sítio ativo. Essa região comumente possui a sequência Asp – Thr – Gly (D*TG), sendo que o sítio de ligação do substrato é uma fenda localizada entre os dois lobos. As aspártico proteinases (Rawlings e Barrett, 1995) utilizam como resíduos catalíticos dois aspartatos durante a catálise. Aspártico proteinases em eucariotos, incluindo proteinases ácidas intracelulares, enzimas digestivas e proteinases extracelulares regulatórias possuem uma grande especificidade por seus substratos. Essa especificidade depende de dois resíduos de ácido aspárticos, conservados em todas aspártico proteinases inclusive em uma catepsina-D lisossomal de mosquito e carrapato (Cho e Raikhel, 1992; Boldbaatar *et al.,* 2006 respectivamente).

O sítio de hidrólise do peptídeo sinal da ppCAD1, ppCAD2 e ppCAD3 foi determinado usando o programa *Signal IP prediction* (Nielsen *et al.,* 1997) e todas as ppCADs possuem os ácidos aspárticos catalíticos Asp_{33} (juntos com T_{34} e G_{35}) e Asp_{229} (juntos com T_{230} e G_{231}) (numeração para uma catepsina-D bovina, Metcalf e Fusek, 1993) onde são indispensáveis para a atividade catalítica da enzima (Barrett *et al.,* 1998).

A catepsina-D bovina, pepsinogênio humana e todas as ppCADs possuem também os resíduos conservados que formam a fenda de ligação do substrato na catepsina-D bovina (Metcalf e Fusek, 1993): V_{30}, D_{33}, G_{35}, S_{36}, Y_{78}, G_{79}, S_{80}, F_{124}, F_{129}, I_{132}, Y_{203}, D_{229}, G_{231}, T_{232}, S_{233} (numeração para a catepsina-D bovina, Metcalf e Fusek, 1993). Os sítios de hidrólise dos propeptídeos previstos por homologia através de alinhamento das sequências de aminoácidos são: pCAD1, Leu_{49} – Arg_{50}; pCAD2, Leu_{40} - Lys_{41}; pCAD3 Leu_{44} – Arg_{45}. O sítio de hidrólise *in vitro* da liberação do propeptídeo da pCAD3 durante a auto-ativação sobre condições ácidas foi confirmado pelo sequenciamento N-terminal da CAD3 madura.

O processamento *in vivo* das aspártico proteinases deve ocorrer por vários mecanismos. Análise das sequências de diferentes formas de catepsinas-D indicaram que a remoção do propeptídeo deve ocorrer por quebra autocatalítica formando uma pseudocatepsina-D e devido a ação de uma outra proteinase, remove os resíduos N-terminal restantes do propeptídeo (Richo e Conner, 1994). Essa idéia poderia ser verificada comparando-se a sequência N-terminal da CAD3 recombinante com a enzima correspondente purificada no homogenizado de larvas

de *Musca domestica* por técnicas cromatográficas. Porém, foram infrutíferas as tentativas de purificação dessa enzima direto do inseto (dados não mostrados nessa tese), impossibilitando esse tipo de análise, na qual a CAD3 ativada *in vitro* poderia na realidade corresponder a uma pseudoCAD3.

A expressão do cDNA que codifica as enzimas pCAD1 e pCAD3 foram realizadas em *E. coli* sob controle do promotor *lac*. Ambas as enzimas recombinantes foram expressas na forma solúvel sendo que a pCAD1 apresentou-se inativa cataliticamente nessas condições de expressão. Porém a pCAD3 uma vez ativada sob condições ácidas encontra-se ativa, hidrolisando hemoglobina e sendo inibida por pepstatina, possuindo uma massa molecular de 41,09 kDa (proenzima contendo a cauda de histidina), determinada por espectrometria de massa (*Ettan MALDI-TOF Pro*, Amercham Biociences, na Unidade Analítica do Instituto de Química da USP, São Paulo, SP).

Esses achados suportam as informações de Espinoza-Fuentes e Terra 1987; Lemos e Terra 1991 onde as catepsinas-D digestiva de *M. Domestica* e de outras Diptera Cyclorrhapha são resultados de duplicação gênica e evolução independente de um gene ancestral intracelular, assim como é verificado nas pepsinas de mamíferos, onde acredita-se que as pepsinas evoluiram a partir de uma catepsina-D lisossomal intracelular.

5.2. ***Sequênciamento randômico de ESTs*** *versus* ***modelo digestivo de larvas de*** *Musca domestica*

Foi construída uma biblioteca unidirecional de cDNA de ventrículo total de larvas de *Musca domestica*. Os ESTs foram obtidos de forma randômica e após as sequências de baixa qualidade serem manualmente removidas obtivemos um total de 826 ESTs. As sequências foram analisadas em programas de bioinformática - Phred e Phrap (Ewing *et al.,* 1998; Gordon et. al., 1998) obtendo-se 323 sequências únicas ou *clusters* que foram analisadas por BLASTN e BLASTX. Essas sequências foram separadas em três categorias: S (que possuem peptídeo sinal); H (metabolismo) e por fim a categoria U (proteínas sem função/localização conhecidas). A seguir foram discutidas apenas as sequências da categoria S. As demais sequências serão avaliadas no futuro, desde que potencialmente relevantes para o estudo da fisiologia digestiva de *Musca domestica*.

5.2.1. Lisozimas

Lisozimas (EC 3.2.1.17) catalisam a hidrólise de ligações glicosídicas do tipo β-1,4 entre ácido N-acetil murâmico e N-acetilglicosamina de peptideoglicanos presentes na parede celular bacteriana, causando a morte das bactérias por lise. As lisozimas são parte do mecanismo de defesa contra bactérias e já foi descrita na maioria dos animais (Jolles e Jolles, 1984) incluindo insetos (Dunn, 1986).

Quinze *clusters* (161 ESTs) codificando para lisozimas foram identificados no transcriptoma do intestino médio de larvas de *Musca domestica,* indicando que vários genes expressos para lisozimas existem no genoma desse inseto. Esse fenômeno é também observado em *Drosophila melanogaster* que possui vários genes expressos para várias lisozimas em seu intestino médio (Kylsten *et al.,* 1992).

Como visto na tabela 5, uma grande expressão de mRNA codificante para essas lisozimas é verificada (19,5% do total dos ESTs), indicando sua importante participação na fisiologia desse inseto.

Larvas de *Musca domestica* matam as bactérias (maior constituinte de seu alimento) na porção média de seu intestino médio pela ação combinada em pH ácido (3,1), de uma lisozima e uma enzima similar a catepsina-D que deve corresponde a CAD3 (ver item 5.1.), (Espinoza-Fuentes e Terra, 1987; Lemos e Terra, 1991).

5.2.2. Digestão de carboidratos

O intestino médio de larvas de *Musca domestica* secretam enzimas tais como maltases ou α-glicosidases (Jordão e Terra, 1989-b), juntamente com α-amilases (Terra *et al.,* 1988) que atuam na digestão de carboidratos.

No transcriptoma do intestino médio de larvas de *Musca domestica* foram encontrados 6 *clusters* representados por 7 ESTs com similaridade para enzimas que atuam na digestão de carboidratos em outros Diptera. Quatro diferentes *clusters* (5 ESTs) mostraram similaridade com genes que codificam para α-amilases de *Ceratites capitata*. Dois *clusters* possuem similaridade com α-glicosidases (maltases), que possam atuar na digestão de oligodextrinas originadas na degradação do amido pela α-amilase.

A digestão de amido no ventrículo de larvas de *M. domestica* inicia-se no interior da membrana peritrófica sob ação de α-amilases. Os oligossacarídeos produzidos pela ação dessas enzimas tornam-se suficientemente pequenos que difundem-se para o interior do espaço ectoperitrófico no ventrículo posterior. Então os oligossacarídeos são hidrolisados pela ação de α-glicosidases (maltases) formando resíduos de glicose que são absorvidos pelo intestino e também pode suportar o crescimento de bactérias na qual é o principal alimento de larvas de *Musca domestica*. (Espinoza-Fuentes e Terra, 1987; Terra *et al.,* 1988; Jordão e Terra, 1989-b).

5.2.3. Digestão de proteínas

Seis *clusters* (9 ESTs) codificando para aspártico proteinases similar a catepsinas-D, sete *clusters* (25 ESTs) codificando para tripsinas, seis *clusters* (100 ESTs) codificando para quimotripsinas, três *clusters* (5 ESTs) codificando para carboxipeptidase A e um *cluster* (1 EST) codificando para dipeptidase, foram identificados no transcriptoma do intestino médio de larvas de *Musca domestica*.

Em larvas de *M. domestica* as proteínas são primeiramente digeridas por um catepsina-D no ventrículo médio que deve corresponder a CAD3 (ver item 5.1.), embora a maior parte da digestão de proteínas ocorre na porção posterior do ventrículo (Espinoza-Fuentes e Terra, 1987; Lemos e Terra, 1991). No ventrículo posterior a digestão de proteínas e oligopeptídeos ocorrem pela ação conjunta de proteinases tais como tripsinas (Espinoza-Fuentes e Terra, 1987; Lemos e Terra, 1992; Jordão *et al.,* 1996) e carboxipeptidase A (Jordão e Terra, 1989-a) e quimotripsina (Kendi *et al.,* em preparação). Os oligopeptideos produzidos por essas enzimas se tornam suficientemente pequenos e difundem-se para o interior do espaço ectoperitrófico e pelo fluxo de fluídos intestinais (Espinoza-Fuentes e Terra, 1987) são translocados e voltam a porção proximal do ventrículo posterior sendo hidrolizados por carboxipeptidase A, aminopeptidase e dipeptidase na superfície das microvilosidases das células (Jordão e Terra, 1991) sendo que os aminoácidos resultantes são finalmente absorvidos pelo intestino.

5.2.4. Peritrofinas

Quatro *clusters* (8 ESTs) codificando para peritrofinas foram identificados no transcriptoma do intestino médio de *Musca domestica*. Três *clusters* (7 ESTs) codificam para peritrofinas do tipo peritrofina-15 e 1 *cluster* (1EST) para peitrofina-44.

Membrana peritrófica é uma estrutura anatômica que envolve o bolo alimentar na maioria dos insetos. Essa membrana é formada por proteínas (peritrofinas) interligadas por fibras de quitina que compartimentaliza o tubo digestivo em espaço endoperitrófico (interior do conteúdo luminal) e espaço ectoperitrófico (espaço entre a membrana peritrófica e o epitélio do intestino)

Peritrofinas são proteínas integrantes pertencentes a membrana peritrófica constituídas por vários domínios. O maior domínio (peritrofina A – *domain*) é rico em resíduos de cisteína com propriedades de ligação de quitina tendo uma sequência consenso: $CX_{13-20}CX_{5-6}CX_{9-19}CX_{4-14}$ (onde X é qualquer aminoácido exceto cisteína) que possui vários resíduos de aminoácidos aromáticos conservados. Variação neste domínio de ligação de quitina são peritrofinas B e peritrofinas C possuindo uma sequência consenso: $CX_{12-13}C_{20-21}CX_{12}CX_{8}CX_{7-12}$ e $CX_{8-9}CX_{19-21}CX_{10-11}CX_{12-13}CX_{11}C$ respectivamente. Outro tipo de domínio que ocorrem em peritrofinas são domínios ricos em prolina/treonina que são severamente glicosilada e similar a mucinas. Peritrofinas pode ter um (peritrofinas-15 de *Lucillia cuprina*) ou vários (peritrofinas-44 de *Lucillia cuprina*) domínios de ligação de quitina (Terra e Ferreira, 2005).

5.2.5. Digestão de lipídeos

Cinco *clusters* (14 ESTs) codificando para lipases foram identificados no transcriptoma do intestino médio de *Musca domestica*. Uma discussão mais detalhada sobre enzimas lipolíticas é demonstrada a seguir.

5.3. Estudo de uma enzima lipolítica no tubo digestivo de larvas de Musca domestica

Através de um sequênciamento randômico de ESTs identificamos uma sequência contendo uma *ORF* (ver Tabela 5, cluster número 16) codificante para uma enzima lipolítica (lipase). A escolha desta sequência chamada LipMD foi

baseada no fato desse transcrito ser formado por um maior número de ESTs (9 ESTs) sugerindo uma participação funcional mais expressiva no tubo digestivo de *Musca domestica*.

A proteína LipMD possui um peptídeo sinal composto por 16 resíduos de aminoácidos localizados na extremidade N-terminal onde a posição do sítio de clivagem foi predito entre Ala^{16} – Phe^{17} (*Signal IP 3.0 Server*, (http:// www.cbs.dtu.dk/services/Signal IP/).

A sequência de aminoácido deduzida para a LipMD possui a tríade catalítica composta pelos resíduos Ser^{183}, Asp^{208} e His^{273} no interior do sítio ativo sendo que o resíduo catalítico Ser^{183} encontra-se no interior do pentapeptídeo conservado (Gly – X_{aa} – Ser – X_{aa} – Gly), característico das enzimas pertencentes a super família das α/β hidrolases (Lowe *et al.,* 1989).

Essa nova enzima lipolítica intestinal de larvas de *Musca domestica* foi expressa com sucesso em *E. coli* e purificada através de uma cromatografia de afinidade em coluna de Ni^{2+}. A proteína recombinante apresentou-se solúvel e ativa hidrolisando tributirina mas não trioleína (óleo de oliva) e demonstrou preferência para pNP-acil-ésteres de curta cadeia carbônica (4C e 5C), indicando a atividade de uma carboxiesterase e não de uma lipase verdadeira.

Alinhamentos por BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) e por domínios conservados (http://www.ebi.ac.uk/Tools/InterProScan/) com outras proteínas de insetos e artrópodes depositadas nos bancos de dados, indicaram que a LipMD seria uma lipase. Isso poderia indicar um artefato durante a anotação dessas sequências tidas como molde depositadas no banco de dados ou sugere uma característica especial ainda não descrita para lipases de insetos. Esses dados obtidos requerem atenção, pois até o presente momento não existem dados moleculares na literatura sobre especificidades por substratos, estrutura 3D, estudos de estéreo especificidades ou outros tipos de dados cinéticos sobre lipases de insetos. Os dados existentes na literatura sobre esse assunto abordam apenas dados filogenéticos (com sequências obtidas em genomas de Diptera), informações em análise de expressão de ESTs, que combinada com limitadas informações fisiológicas disponíveis, relatam baixos níveis de identidades com lipases de mamíferos (Horne *et al.,* 2009).

Enzimas lipolíticas (lipases e carboxiesterases) presentes no trato intestinal possuem um papel crítico na habilidade do animal utilizar lipídeos em sua dieta. Enzimas lipolíticas são também requeridas para mobilizar lipídeos a partir das reservas energéticas das células (Gibbons *et al.,* 2000). Os triacilgliceróis não atravessam membranas por transporte passivo e não são solúveis em água ou camadas de fosfolipídeos. Essas moléculas não podem atravessar as membranas microvilares intestinais e a hidrólise dos triacilgliceróis por lipases/carboxiesterases seguida pela absorção ou reesterificação dos ácidos graxos pelo intestino é requirida (Desnuelle *et al.,* 1986).

Esterases denominadas carboxiesterases hidrolisam ligações ésteres de triacilgliceróis de cadeia carbônica curta. Essas enzimas, ao contrário das lipases, atacam moléculas que são completamente solúveis em água. Ela hidrolisa também triacilgliceróis de cadeia longa insolúveis em água na presença de surfactantes mas a velocidades bem menores que lipases (Terra e Ferreira, 2005).

A LipMD deve corresponder a uma enzima lipolítica secretada no conteúdo luminal do ventrículo anterior em larvas de *Musca domestica* devido:

(1) Baseado na análise por RT-PCR, o mRNA codificante para a LipMD é expresso apenas no ventrículo anterior. (2) Imunoblots das proteínas de diferentes tecidos marcados com o anticorpo policlonal preparado contra a LipMD recombinante revela que a tradução ocorre apenas no ventrículo anterior. (3) LipMD é localizada através de ensaios de imunocitolocalização (utilizando-se a técnica de imuno-ouro) no interior de vesículas de secreção no ápice das células próximo as microvilosidades (no interior da célula) no ventrículo anterior. Em *Drosophila melanogaster* a atividade lipolítica é maior nas células do intestino que em outros tecidos (Smith *et al.,* 1994). A maior região de absorção de lipídeos em insetos são o ceco gástrico e a porção anterior do ventrículo (Dow, 1987) sendo que três sequências codificantes para enzimas lipolíticas identificadas como lipases (FlyBase 3.1 – Berkeleg Drosophila Genome Project – www.fruitfly.org/annot/) CG6271, CG6277, CG6283) são conhecidas para serem expressas no ventrículo anterior em *D. melanogaster*, onde sua expressão pode ser reprimida em uma dieta rica em açúcar (Zinke *et al.,* 2002). A LipMD demonstrou possuir alta identidade com essas três sequências lipolíticas de *D. Melanogaster* (Figura 34). Essas proteínas possivelmente estão envolvidas na lipólise primária de lipídeos na dieta desses animais. (4) A LipMD foi capaz de hidrolisar um grande número de pNP-acil-ésteres

(pNP-acetato (2C) à pNP-estearato (18C)). (5) O pH de 6,9 no ventrículo anterior de larvas de *Musca domestica* (Regel e Terra, 1995) favorece a atividade da LipMD (pH ótimo 7,5) que pode corresponder a atividade lipolítica encontradas nessa região do tubo digestivo. (6) Ensaios enzimáticos realizados com as proteínas encontradas no conteúdo luminal demonstraram uma maior atividade lipolítica no ventrículo anterior em larvas de *Musca domestica* concordando com os resultados obtidos nos itens 1, 2 e 3. Essas informações sugerem que a LipMD possa atuar como uma enzima lipolítica secretada no conteúdo luminal do tubo digestivo de larvas de *Musca domestica*.

Diptera Cyclorrhapha alimenta-se preferencialmente de bactérias e em seus intestinos, enzimas digestivas podem ter duplo papel, na degradação do bolo alimentar assim como no auxílio da degradação das bactérias em sua dieta. Bactérias contém lipopolisacarídeos sobre a sua superfície e enzimas lipolíticas podem estar envolvidos na quebra desses componentes auxiliando no rompimento da cápsula microbial ou mesmo na digestão do conteúdo lipídico presente no bolo alimentar (Boulanger *et al.,* 2006).

Até o presente momento não existem dados moleculares na literatura sobre especificidades por substratos, estrutura 3D, estudos de estéreo-especificidades ou outros tipos de dados cinéticos sobre lipases de insetos (Horne *et al.,* 2009). Nossos dados apresentados nessa tese podem contribuir para o início do surgimento de uma nova e extensa área da bioquímica de insetos que abordarão aspectos moleculares dessa classe enzimática. Os nossos resultados combinados com os futuros estudos empíricos poderão nos levar a um melhor entendimento molecular do metabolismo de lipídeos em insetos.

6. REFERÊNCIAS

Aaronson, J.S., Eckman, B., Blevins, R.A., *et al.,* 1996. Toward the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence data. Genome Res. 6, 829–845.

Adams, M.D., Kelley, J.M., Gocayne, J.D., et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 252, 1651–1656.

Alarcon-Chaidez, F. J., Sun, J., Wikel, S. K., 2007. Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). Insect Biochem. Mol. Biol. 37, 48-71.

Aldridge, W.N., 1993. The esterases: perspectives and problems. Chem. Biol. Interact. 87, 5-13.

Altschul, S.F., Lipman, D.J., 1990. Protein database searches for multiple alignments. Proc. Natl. Acad. Sci. (USA) 87, 5509 - 5513.

Altschul, S.F, Madden, T.L, Schaffer, A.A, *et al.,* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Alstad, D.N., Andow, D.A., 1995. Managing the evolution of insect resistance to transgenic plants. Science, 268, 1894-1896.

Arpigny, J.L., Jaeger, K.E., 1999. Bacterial lipolytic enzymes: classification and properties. Biochem. J. 343, 177-183.

Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M., Erickson, J.W., 1993. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. Proc. Natl. Acad. Sci. (USA) 90, 6796–6800.

Boulanger, N., Bulet, P., Lowenberger, C., 2006. Antimicrobial peptides in the interactions between insects and flagellate parasites. Trends. Parasit. 22, 262–268.

Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol. 10, 411-421.

Barrett, A.J., 1977. Cathepsin D and other carboxyl proteinases. In: Proteinases in Mammalian Cells and Tissues (Barrett, A.J., ed.). Amsterdam: North-Holland Biomedical Press, 209–248.

Barrett, A.J., Rawlings, N.D., Wasner, J.F., 1998. Handbook of proteolytic enzymes. Academic Press, London.

Barrett, A.J., Rawlings, N.D., Wasner, J.F., 2004. Handbook proteolytic enzymes. Second Edition. Academic Press, London.

Baudet, C., Dias, Z., 2005. New EST Trimming Strategy. In: Lecture Notes on Bioninformatics, Springer. 206–209.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. Wheeler, D. L., 2000. GENBANK. Nucl. Acids Res. 28, 15-18.

Beyer, B.M., Dunn, B.M., 1996. Self-activation of recombinant human lysosomal procathepsin D at a newly engineered cleavage junction, “short” pseudocathepsin D. J. Biol. Chem. 271, 15590–15596.

Blum, H., Beir H., Gross, H., 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrilamide gels. Electrophoresis 8, 93 – 99.

Bonaldo, M.F., Lennon G, Soares M.B., 1996. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res. 6, 791–806.

Bornia, M., Nielsen, S., Engel, A., Agree, P., 1999. Cellular and molecular biology of the aquaporin water channels. Annu. Rev. Biochem., 68, 425-458.

Boldbaatar, D., Sikasunge, C. S., Battsetseg, B., Xuan, X., Fujisaki, K., 2006. Molecular cloning and functional characterization of an aspartic protease from the hard tick *Haemaphysalis longicornis*. Insect Biochemistry and Molecular Biology 36, 25-36.

Bolognesi, R., Terra, W.R., Ferreira, C., 2008. Peritrophic membrane role in enhancing digestive efficiency Theoretical and experimental models. Journal of Insect Physiology 54, 1413-1422.

Bornscheuer, U.T., 2002. Microbial carboxyl esterase: classification properties and application in biocatalysis. FEMS Microbiol. Rev. 26, 73-81.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analyt. Biochem. 72, 248-254.

Burke, J., Davison, D., Hide, W., 1999. d2_*cluster*: a validated method for *cluster*ing EST and full-length cDNAsequences. Genome Res. 9,1135–1142.

Calvo, E., Andersen, J., Francischetti, I. M., Capurro, M.D., deBianchi, A.G., James, A.A., Ribeiro, J.M.C., Marinotti, O., 2004. The transcriptome of adult female *Anopheles darlingi* salivary glands. Insect Molecular Biology 13, 73-88.

Canavoso, L.E., Jouni, Z.E., Karnas, K.J., Pennington, J.E., Wells, M.A., 2001. Fat metabolism in insects. Annu. Rev. Nutrition 21, 23–46.

Cereghino, J.L., Cregg, J.M., 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol. Rev. 24, 45-66.

Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, M.K., Doctor, B. P., 1993. Relationship betweewn sequence conservation and three-dimentional structure in large family of esterases, lipases and related proteins. Protein Sci. 2, 366-382.

Cho, W.L., Raikhel, A.S., 1992. Cloning of cDNA for mosquito lysosomal aspartic protease. Sequence analysis of an insect lysosomal enzyme similar to cathepsins D and E. Journal of Biological Chemistry 267, 21823–21829.

Chomczynski, P., Sacchi, N., 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Connor, G.E., 1998. Proteolytic enzymes. In: Barrett, A.J., Rawlings, D., Woessner, J.F. (Eds.), Handbook of Proteolytic Enzymes. Academic Press, Oxford, 828–836.

Conner, G.E., Richo, G., 1992. Isolation and characterization of a stable activation intermediate of the lysosomal aspartyl protease cathepsin D. *Biochemistry* 31, 1142–1147.

Conner, G.E., Udey, J.A., 1990. Expression and refolding of recombinant human fibroblast procathepsin D. *DNA Cell Biol.* 9, 1–9.

Cunningham, M., Tang, J., 1976. Purification and properties of cathepsin D from porcine spleen. J. Biol. Chem. 251, 4528–4536.

Desnuelle, P., Sjostrom, H., Norén, O., 1986. Molecular and cellular basis of digestion, Elsevier, Amsterdam.

De Veau, E.J.I., Schultz, J.C., 1992. Reassessment of interaction between gut detergents and tannis in Lepdoptera and significance for gypsy moth larvae. J. Chem Ecol. 18, 1437 – 1453.

Djamgoz, M.B., Ready P.D., Billingsley, P.F., Emery, A.M., 1998. Insect Na^+/K^+-ATPase. J. Insect Physiol., 44, 197-209.

Dolinsky, V.W, Gilham, D., Alam, M., Vance, D.E., Lehner, R., 2004. Triacylglycerol hydrolase: role in intracellular lipid metabolism. Cell Mol. Life Sci. 61, 1633-1651.

Dong, Q., Kroiss, L., Oakley, F.D., *et al.,* 2005. Comparative EST analyses in plant systems. Methods Enzymol. 395, 400–408.

Dow, J.A.T., 1987. Insect midgut function. Adv. Insect Physiol. 19, 188–328.

Dunn, B.M., Jimenez, M., Parten, B.F., Valler, M.J., Rolph, C.E., *et al.,* 1986. A sysmatic series of synthetic chromophoric substrates for aspartic proteinases. Biochem. J. 237, 899-906.

Emery, A.M., Billingsley, P.F., Ready, P.D.Djamgoz, M.B.A., 1998. Insect Na^+/K^+-ATPase. J. Insect Physiol., 44, 197-209.

Escher, S.A., Rasmunson-Lestander, A., 1999. The Drosophila glucose transporter gene: cDNA sequence, phylogenetic comparisons analysis of functional sites and secondary structures. Hereditas 130, 95-103.

Espinoza-Fuentes, F.P., Terra, W.R., 1987. Physiological adaptations for digesting bacteria. Water fluxes and distribution of digestive enzymes in *Musca dometica* larval midgut. Insect Biochem. 17, 809-817.

Espinoza-Fuentes, F.P., Ribeiro, A.F., Terra, W.R., 1987. Microvillar and secreted digestive enzymes from *Musca dometica* larvae. Subcellular fractionation of midgut cells with electron microscopy monitoring. Insect Biochem. 17, 819-827.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194.

Ewin, B., Hillier, L., Wendl, M., Green, P., 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175 – 185.

Faust, P.L., Kornfeld, S., Chirgwin, J.M., 1985. Cloning and sequence analysis of cDNA for human cathepsin D. Proc. Natl. Acad. Sci. (USA) 82, 4910–4914.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783 – 791.

Garesse, R., Castell, J.V., Vallejo, C.G., Marco, R., 1979. A fluorescamine-based sensitive method for assay of proteinases, capable of detecting the initial cleavage steps of a protein. Eur. J. Biochem. 99, 253-259.

Gibbons, G.F, Islam, K., Pease, R.J., 2000. Mobilisation of triacylglycerol stores. Biochim. Biophys. Acta, 1483, 37-57

Gillott, C., 1995. Entomology 2nd ed. Plenum Press, New York, 798.

Gordon, D., Abajian, C., Green, P., 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8, 195-202.

Greenberg, B., Paretsky, D., 1955. Proteolytic enzymes in the housefly, *Musca dometica*. Ann. Entomol. Soc. Am. 48, 46-50.

Grimaldi, D., Engel M.S., 2005. Evolution of the insects. Cambrigde University. Press, New York**.**

Harrop, S.A., Prociv, P., Brindley, P.J., 1996. Acasp, a Gene Encoding a Cathepsin D-like Aspartic Protease from the Hookworm *Ancylostoma caninum*. Biochemical and biophysical research communications 227, 294-302.

Hillis, D.B., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182 – 192.

Holmquist, M., 2000. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr. Protein Pept. Sci. 1, 209-235.

Horst, M., Hasilik, A., 1991. Expression and maturation of human cathepsin D in baby-hamster kidney cells. Biochem. J. 273, 355–361.

Horne, I., Haritos, V.S., Oakeshott, J.G., 2009. Comparative and functional genomics of lipases in holometabolous insects Insect Biochem. Mol. Biol. 39, 547-567.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.

Jaeger, K.E., Eggert, T., 2002. Lipases for biotechnology. Curr. Opin. Biotechnol. 13, 390-397.

Jongeneel, CV., 2000. Searching the expressed sequence tag (EST) databases: panning for genes. Brief Bioinform. 1, 76–92.

Jolles, P., Jolles, J., 1984. What´s new in lysozime research? Molec. cell biochem. 63, 164-189.

Jordão, B.P., Terra, W. R., 1989-a. Distribution, properties, and functions of midgut carboxypeptidases and dipeptidases from *Musca domestica* larvae. Arch. Insect Biochem. Physiol. 11, 231-236.

Jordão, B.P., Terra, W. R., 1989-b. Final digestion of starch in *Musca domestica* larvae. Distribution and properties of midgut α-D-glucosidases and glucoamylase. Insect Biochem. 19, 285-292.

Jordão, B.P., Terra, W.R., 1991. Regional distribution and substrate specificity of digestive enzimes involved in terminal digestion in *Musca domestica* hind-midguts. Archives of Insect Biochemistry and Physiology. 17, 157-168.

Jordão, B.P., Terra, W.R., Ferreira, C., 1995. Chemical determinations in microvillar membranes purified form brush-borders isolated from the larval midgut of one Coleoptera and two Diptera species. Insect Biochem. Molec. Biol. 25, 417-426.

Jordão, B.P., Terra, W.R., Ribeiro, A.F., Lehane, M.J., Ferreira, C., 1996. Trypsin secretion in *Musca dometica* larval midguts: a biochemical and immunocytochemical study. Insect Biochem. Molec. Biol. 26, 337-346.

Keilvá, H., 1970. Inhibition of cathepsin D by diazoacetylnorleucine methyl ester. 25, 312-314.

Kirschner, M.W., 2005. The meaning of systems biology. Cell 121, 503 – 504.

Koike, M., Nakanishi, H., Saftig, P., Ezaki, *et al*., 2000. Cathepsin D deficiency induces lysosomal storage with ceroid lipofuscin in mouse CNS neurons. J. Neurosci. 20, 6898–6906.

Kouger, G., Jaeger, K.E., 1987. Specific and sensitive plate assay for bacterial lipases. Applied an Enviromental Microbiology 53, 211-213.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA 4: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5, 150-163.

Kylsten, P., Kimbrell, D. A., Samakovlis, C., Hultmark, D., 1992. The lysozyme locus in *Drosophila melanogaster*: different genes are expressed in midgut and salivary glands. Molec. Gen. Genet. 232, 335 – 343.

Laemmli, U.K., 1970. Cleveage of strutural proteins during assembly of the head of bacteriophage T7. Nature 227, 680-685.

Law, J.H., Ribeiro, J.M., Wells, M. A., 1992. Biochemical insights derived from insect diversity. Annu. Rev. Biochem. 61, 87-111.

Lemos, F.J.A., Terra, W.R., 1991. Properties and intracellular distribution of a cathepsin D-like proteinase active at the acid region of *Musca dometica* midgut. Insect Biochem. 21, 457-465.

Lemos, F.J.A., Terra, W.R., 1992. A high yield preparation of *Musca dometica* larval midgut microvilli and the subcellular distribution of amylase and trypsin. Insect Biochem. Molec. Biol. 22, 433-438.

Lemos, F.J.A., Ribeiro, A.F., Terra, W.R., 1993. A bacteria-digesting midgut-lysozyme from *Musca dometica* (Diptera) larvae. Purification, properties and secretory mechanism. Insect Biochem. Molec. Biol. 23, 533-541.

Lowe M.E., Rosenblum J.L., Struass, A.W., 1989. Cloning and characterization of human pancreatic lipase cDNA. *J. Biol. Chem.* 264, 20042–20048

Lowe, M.E., Jennens, M.L., 1995. A glutamic acid in colipase is essential for reactivation of pancreatic lipase. *Pediatr. Res.* 37, 127A (Abstr.).

Ludewing, U., 2004. Electroneutral ammonium transport by basolateral rhesus B glycoprotein. J. Physiol. 559, 751-759.

Malde, K., Coward, E., Jonassem, I., 2005. A graph based algorithm for generating EST consensus sequences. Bioinformatics. 21, 1371–1375.

Marini, A. M., Matassi, G., Raynal, V., André B., Cartron J. P., Cherif-Zahar B., 2000. The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast. Nat. Genet. 26, 341-344.

Matsuidara, P., 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262, 10035 – 10038.

Mayordomo, I. Randez-Gil, F., Prieto, J.A., 2000. Isolation, purification, and characterization of a cold-active lipase from *Aspergillus nidulans*. J. Agric. Food Chem. 48, 105-109.

Marana, S.R., Jacobs-Lorena, M., Terra, W.R., Ferreira, C., 2001. Amino acid residues involved in substrate binding and catalysis in an insect digestive β-glycosidase. Biochim. Biophys. Acta 1545, 41-52.

Marana, S.R., Terra, W.R., Ferreira, C., 2002. The role of amino-acid residues Q39 and E451 in the determination of substrate specificity of the *Spodoptera frugiperda* β-glycosidase. Eur. J. Biochem. 269, 3705-3714.

Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., *et al.,* 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30, 281–283.

Metcalf, P., Fusek, M., 1993. Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. EMBO J. 12, 1293–1302.

Nielsen, H., Engelbrecht, J., Brunak, S., Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and predictions of their cleavage sites. Protein Eng. Des. Sel. 10, 1–6.

Nishimura, Y.,Takeshima, H., Sakaguchi, M., Mihara, K., Omura, T., Kato, K., 1995. Expression of rat cathepsin D cDNA in *Saccharomyces cerevisia*e: implications for intracellular targeting of cathepsin D to vacuoles. *J. Biochem. (Tokyo)* 118, 168–177.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., 1992. The alpha/beta hydrolase fold. Protein Eng. 5, 197-211.

Pedersen, P.L., Carafolli, E., 1987. Ion motive ATPase. II. Energy coupling and work out put. Trends Biochem. Sci. 12, 186-189.

Pimenta, D.C., Oliveira, A., Juliano, M.A., Juliano, L., 2001. Substrate specificity of human cathepsin D using internally quenched fluorescent peptides derived from reactive loop of kallistatin. Biochim. Biophys. Acta. 1544, 113-122.

Ponnuvel, K.M., Nakazawa, H., Furukawa, S., Asaoka, A., Ishibashi, J., Tanaka, H., Yamakawa, M., 2003. A lipase isolated from the silkworm *Bombyx mori* shows antiviral activity against nucleopolyhedrovirus. J. Virol. 77, 10725–10729.

Press, E.M., Porter, R. R., Cebra, J.J., 1960. The isolation and properties of a proteolytic enzyme, cathepsin D, from bovine spleen. Biochem. J. 74, 501-514.

Ptitsyn, A., Hide, W., 2005. CLUSTAL: a new algorithm for EST *cluster*ing. BMC Bioinformatics. 6 [Suppl 2], S3.

Pullikuth, A.K., Filippov, V., Gill, S.S., 2003. Phylogeny and cloning of ion transporters in mosquitoes. J. Exp. Biol. 206, 3857-3868.

Ramos, C. C. R., Abreu, P.A.E., Nascimento, A.L.T.O., Ho, P.L., 2004. A high-copy T7 *Escherichia coli* expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. Brazilian Journal Medical and Biologycal Research 37, 1103-1109.

Rawlings, N.D., Barrett, A.J., 1995. Families of aspartic peptidases, and those of unknown catalytic mechanism. Methods Enzymol. 248, 105–120.

Regel, R., Terra, W.R., 1995. pH buffering in *M. domestica* midguts. Comp. Biochem. Physiol. 112, 559-564.

Reynolds, T.M., 1963. Chemistry of nonenzymic browning I. The reaction between aldoses and amines. Adv. Food Res. 12, 1-52.

Richo, G.R., Conner, G.E., 1994. Structural requirements of procathepsin D activation and maturation. J. Biol. Chem. 269, 14806–14812.

Rudd, S., 2003. Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant. Sci. 8, 321–329.

Saitou, N., Nei, M., 1987. Neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406 – 425.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., Coulson, A., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences (USA). 74, 5463–5467.

Santos, C.D., Ribeiro, A.F., Ferreira, C., Terra, W.R., 1984. The larval midgut of the cassava hornworm (*Erinnyis ello*). Ultrastructure, fluid fluxes and the secretory activity in relation to the organization of digestion. Cell Tissue Res. 237, 565-574.

Sapolsky, A.I., Woessner, J.F.Jr., 1972. Multiple forms of cathepsin D from bovine uterus. J. Biol. Chem. 247, 2069–2076.

Satoh, T., Hosokawa, M., 1998. The mammalian carboxylesterases: from molecules to functions. Annu. Rev. Pharmacol. Toxicol. 38, 257-288.

Scarborough, P.E., Guruprasad, K., Topham, C., Richo, G.R., *et al.*, 1993. Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling. Protein Sci. 2, 264–276.

Shah, D.M., Rommens, C.M.T., Beachy, R.N., 1995. Resistance to disease and insect in transgenic plants: progress and application to agriculture. Tibtech., 13, 362-368.

Schoonhooven, L.M., Van Loon J.J. A., Dicke, M., 2005. Insect-plant biology. Oxford University Press, Oxford.

Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., Bucher, P., 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 3, 265-274.

Smith, G.M., Rothwell, K., Wood, S.L., Yeaman, S.J., Bownes, M., 1994. Specificity and localization of lipolytic activity in adult *Drosophila melanogaster*. Biochem. J. 304, 775–779.

Srinivasan, P., Abraham, E.G., Glosh, A.K., *et al.,* 2004. Analysis of the Plasmodium and Anopheles Transcriptome during oocyst difeerentiation. The Journal of Biological Chemistry. 279, 5581 – 5587.

Stefanini, M., De Martino, C., Zamboni, L., 1967. Fixation of ejaculated spermatozoa for electron microscopy. Nature 216, 173 – 174.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596-1599 (Publication PDF at http://www.kumarlab.net/publications).

Tang, J., Wong, R.N., 1987. Evolution in the structure and function of aspartic proteases. J. Cell. Biochem. 33, 53–63.

Tatusov, R.L, Fedorova, N.D., Jackson, J.D., *et al.,* 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 4, 41.

Terra, W.R., 1988. Physiology and biochemistry of insect digestion: an evolutionary perspective. Braz. J. Med. Biol. Res. 21, 675-734.

Terra, W.R., Espinoza-Fuentes, F.P., Ribeiro, A.F., Ferreira, C., 1988. The larval midgut of the housefly (*Musca dometica*): ultrastructure, fluid fluxes and ion secretion in relation to the organization of digestion. J. Insect Physiol., 34, 463-472.

Terra, W.R., 1990. Evolution of digestive systems of insects. Annu. Rev. Entomol. 35, 181-200.

Terra, W.R., 2001. The origins and functions of the insect peritrophic membrane and peritrophic gel. Arch. Insect Biochem. Physiol. 47, 47-61.

Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. Comp. Biochem. Physiol., 109, 1-62.

Terra, W.R., Ferreira, C., 2005. Biochemistry of digestion. In: Gilbert, L.I., Iatrou, K., Gill, S.S. (Eds.), Comprehensive Molecular Insect Science, Vol. 4, Oxford: Elsevier, 171-224.

Terra, W.R., Ferreira, C., Jordão, B.P., Dillon, R.J., 1996. Digestive enzmes. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, 153–194.

Terra, W.R., Costa, R.H., Ferreira, C., 2006. Plasma membranes from insect midgut cells. An. Acad. Brasil. Cienc. 78, 255-269.

The Gene Ontology Consortium, 2000. Gene Ontology: tool for the unification of biology. Nature Genet. 25, 25-29.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

van Tilbeurgh, H., Bezzine, S., Cambillau, C., Verger, R., Carriere, F., 1999. Colipase: structure and interaction with pancreatic lipase. Biochim. Biophys. Acta 1441, 173–184.

Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. (USA) 76, 4350-4354.

Tsuchida, K., Wells, M. A., 1998. Digestion absorption transport and storage of fat during the last stadium of *Manduca sexta*: changes in the role lipophorin in the delivery of dietary lipids to the body. Insect Biochem. 18, 263 – 268.

Zinke, I., Schutz, C.S., Katzenberger, J.D., Bauer, M., Pankratz, M.J., 2002. Nutrientcontrol of gene expression in Drosophila: microarray analysis of starvation and sugar-dependent response. EMBO J. 21, 6162–6173.

Zurek, L., Schal, C., Watson, D.W., 2000. Diversity and contribution of the intestinal bacterial community to the development of *Musca domestica* (Diptera: Muscidae) larvae. J. Med. Entomol. 37, 924-928.

Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T., Ikemura, T., 1991. Codon usage tabuled from GENBANK genetic sequence data. Nucleic Acids Res. 19, Suppl, 1981–1986.

Weihrauch, D., 2006. Active ammonia absorption in the midgut of the Tobacco hornworm *Manduca sexta* L.: transport studies and mRNA expression analysis of a Rhesus-like ammonia transporter. Insect Biochemistry and Molecular Biology 36, 809-821.

Westling, J. Cipullo, P., Hung, S. H., Saft, H. Dame, J. B., Dunn, B. M., 1999. Active site specificity of plasmepsin II. Protein Sci. 8, 2001-2009.

Wheeler, D. L., Chappey, C., Lash, A. E., Leipe, D. D., Schuler, T. L., Tatusova, T. A., Rapp, B.A., 2000. Database resources of the National Center for Biotechnology Information. Nucl. Acids Res. 28, 10-14.

Winkler, U.K., Stuckmann, M., 1979. Glycogen, hyaluronate and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. J. Bacteriol. 138, 663-670.

Winkler, F.K., D'Arcy, A., Hunziker, W., 1990. Structure of human pancreatic lipase. Nature 343, 771–774.

Winlkins, M.R., Lindskog, I., Gasteiger E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F., Appel R.D., 1997. Detailed peptide characterization using PEPTIDE MASS – a Worl Wide Web accessible tool. Electrophoresis 18, 403 – 408.

Yasuda, Y., Kageyama, T., Akamine A., Shibata, M., 1999. Characterization of new fluorogenic substrates for the rapid and sensitive assay of cathepsin E and cathepsin D. J. Biochem. 125, 1137-1143.

7. SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Marcelo Henrique Peteres Padilha
Local e data de nascimento: Brasília, DF - 17/02/1982
E-mail: marcelaopadilha@yahoo.com.br

Formação

Graduação em Química

Universidade Estadual de Londrina (UEL), 1999 - 2003

Mestrado em ciências – Bioquímica

Universidade de São Paulo – USP, São Paulo, SP
Instituto de Química, Departamento de Bioquímica
Orientador: Dr. Sandro Roberto Marana
Período: 2003 - 2005

Doutorado em ciências – Bioquímica

Universidade de São Paulo – USP, São Paulo, SP
Instituto de Química, Departamento de Bioquímica
Orientador: Dr. Walter Ribeiro Terra
Período: 2005 - 2009

Ocupação

Bolsista de Mestrado, CNPq, 2003 – 2005.
Bolsista de Doutorado, CNPq, 2005 – 2009.

Publicações

PADILHA, M.H.P., PIMENTEL, A.C., RIBEIRO, A.F., TERRA, W.R., 2009. Sequence and function of lysosomal and digestive cathepsin D-like proteinases of *Musca domestica* midgut enzymes. Insect Biochemistry and Molecular Biology (*no prelo*).

Resumos publicados em anais de congressos

PADILHA, M.H.P., PIMENTEL, A.C., TERRA, W.R., 2009. The cathepsin D-like proteinase from the midgut of *Musca domestica* larvae: sequence, properties, localization and function. XXXIV FEBS Congress. República Tcheca, Praga.

TAMAKI, F.K., PADILHA, M.H.P., TERRA, W.R., 2009. Distribution, Purification, Partial Characterization, Sequencing, Cloning and Expression of *Musca domestica* Chymotrypsin. XXXVIII Reunião Anual da SBBq, Águas de Lindóia, SP.

PADILHA, M.H.P., TERRA, W.R., 2009. Cloning, Expression, Characterization and Distribution of a Lipolytic Enzyme from *Musca Domestica* Midgut Larvae. XXXVIII Reunião Anual da SBBq, Águas de Lindóia, SP.

PADILHA, M.H.P., PIMENTEL, A.C., TERRA, W.R., 2008. Recombinant Expression and Distribution of Three Cathepsin-D-like Aspartic Proteinases from *Musca domestica* Midgut Larvae. XXXVII Reunião Anual da SBBq, Águas de Líndóia, SP.

PADILHA, M.H.P., TERRA, W.R., 2008. Analyis of Expressed Sequence Tags from *Musca domestica* midgut larvae. XXXVII Reunião Anual da SBBq, Águas de Lindóia. XXXVII Reunião Anual da SBBq, 2008.

PADILHA, M.H.P., PIMENTEL, A.C., TERRA, W.R., 2007. Molecular cloning and partial characterization of three aspartic proteinase from *Musca domestica* larvae. XXXVI Annual Meeting of the SBBq and 10th IUBMB Conference, Salvador, BA.

PADILHA, M.H.P., MARANA, S.R., 2005. Molecular basis of aglycone specificity in a beta-glycosidase. XXXIV Reunião Anual da SBBq, Águas de Lindóia, SP.

MENDONÇA, L.M.F., MARANA, S.R., PADILHA, M.H.P., 2004. Molecular basis of pH optimum determination in beta-glycosidases. XXXIII Reunião Anual da SBBq, Caxambu, MG.

Monitorias

Participação do Estágio Supervisionado em docência do Programa de Aperfeiçoamento de Ensino (PAE) junto à disciplina de Bioquímica Experimental (QBQ-316), ministrada aos alunos da Farmácia (noturno) no 2º semestre de 2004.

Participação do Estágio Supervisionado em docência do Programa de Aperfeiçoamento de Ensino (PAE) junto à disciplina de Bioquímica Experimental (QBQ-316), ministrada aos alunos da Farmácia (noturno) no 2º semestre de 2005.