



**Universidade de Brasília**

**Programa de Pós-Graduação em Biologia Animal**

**Construção da biblioteca de cDNA da glândula de  
peçonha do escorpião *Opisthacanthus cayaporum* e  
clonagem de genes que codificam para componentes da  
peçonha**

**Édelyn Cristina Nunes Silva**

**Orientadora: Prof<sup>a</sup> Dr<sup>a</sup> Elisabeth N. Ferroni Schwartz**

**Dissertação apresentada ao  
Programa de Pós-Graduação em  
Biologia Animal da Universidade  
de Brasília como parte dos  
requisitos para a obtenção do  
título de Mestre.**

**Brasília, 2008**



UNIVERSIDADE DE BRASÍLIA  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

Dissertação de Mestrado  
**Édelyn Cristina Nunes Silva**

Título:

**“Construção da biblioteca de cDNA da glândula de peçonha do escorpião *Opisthacanthus cayaporum* e clonagem de genes que codificam para componentes da peçonha”**

Comissão examinadora:

***Prof<sup>a</sup>.Dr<sup>a</sup>. Elisabeth N. Ferroni Schwartz***  
***Presidente/Orientadora***  
***CFS/UnB***

***Prof. Dr. Luciano Paulino da Silva***  
***Embrapa/Cenargen***

***Prof<sup>a</sup>. Dra. Ildinete Silva Pereira***  
***CEL/UnB***

***Prof<sup>a</sup>. Dr<sup>a</sup>. Márcia Renata Mortari***  
***Suplente, CFS/UnB***

**Brasília, julho de 2008**

## **DEDICATÓRIA**

A Deus

Ao meu noivo Flávio

A minha amada Mãe, Matilde

A minha irmã Évelyn

A Dra. Elisabeth Schwartz

e ao Dr. Lourival Possani

## AGRADECIMENTOS

Esta dissertação de Mestrado só teve êxito porque ela é fruto do trabalho de muitas pessoas, de dois países, vários laboratórios, duas universidades... Assim tentarei agradecer a todos por essa vitória!

Meus sinceros agradecimentos ao Prof. Dr. Lourival Possani do Departamento de *Medicina Molecular y Bioprocessos, Instituto de Biotecnología da Universidad Autónoma do México*, em Cuernavaca / Morelos, pela oportunidade de viajar ao México, conhecer esse maravilhoso país e trabalhar em seu renomado laboratório onde pude realizar grande parte dessa dissertação.

Agradeço muitíssimo à minha orientadora Prof<sup>a</sup> Dra Elisabeth Schwartz, por acreditar no meu trabalho, me dar a oportunidade de conhecimento, me auxiliar sempre em todas as etapas desse trabalho, por ser fonte de enorme conhecimento e trabalho. Sem ela esse trabalho nem sairia do papel.

Ao laboratório de Biologia Molecular da Universidade de Brasília, às professoras Dras Andréa Maranhão, Ildinete Pereira, ao aluno de doutorado Marciano, por sempre me auxiliarem na obtenção dos dados para esse trabalho, por serem gentis e acessíveis para comigo.

Agradeço ao Dr. Carlos Bloch Júnior e Dr. Luciano Paulino da Silva do Laboratório de Espectrometria de Massa, Embrapa/Cenargen, pela colaboração e inestimáveis ajudas.

Ao professor Dr. Osmindo Júnior por sempre estar do meu lado, mesmo nos momentos de terrorismo. A sua esposa Janaína pelos “toques”, momentos de descontração e inestimável ajuda.

Ao professor Dr. Carlos Alberto Schwartz pelas sempre oportunas dicas, pelo terrorismo que no fundo ajudava muito.

À professora Dra. Márcia Mortari pelas ótimas conversas profundas sobre profissão, pelos conhecimentos compartilhados, pelos momentos de descontração, por ser sempre estar disponível para uma palavra amiga. Obrigada também a sua ótima família: Gustavo e Luzitano.

Como agradecer à Thalita e Naty? Seriam rolos de papel descrevendo a alegria que vocês me proporcionam, a base firme e otimista quando tudo tava dando errado,

companheiras nas horas divertidas e trágicas. Grandes amigas, esse trabalho tem um pedacinho de vocês.

Agradeço ao admirável Jimmy Guerrero Vargas, que tem guerreiro não só no nome, que é exemplo de vida para mim, de força de vontade, responsabilidade e trabalho.

Aos amigos do laboratório, Carol Barbosa, Carla Vizzoto, Solange, Poliana, Rafael Félix, Rafael Melani, Luís, Washigton, Cezinha, Leandro, Thompson, Professor Antônio Sebben, Flávio Dourado, Flávio Caixeta, pelos papos não-científicos, momentos de descontração e pela amizade nestes dois anos.

Agradeço aos amigos mexicanos: Maria Juanita, Cynthia, Verna, Verônica, River, Blanca, Fred, Ligia Gonzáles, pelos dias maravilhosos que passei no México, pelo apoio no desenvolvimento desse trabalho, pela imensa ajuda no laboratório e pelos formidáveis passeios. *Muchas gracias, ustedes están en mi corazón!*

A amiga Turca Figueim, minha companheira de quarto na casa da Yoli, agradeço por ser ótima companhia e por me mostrar o México que eu não conhecia.

Aos amigos brasileiros Clara, Giu, Vanessa, Gabi por compreenderem minha ausência por uns tempos e serem ótimas companheiras.

Ao meu amado noivo Flávio, por estar comigo em todos os momentos, por me auxiliar na escrita desse trabalho, por agüentar 40 dias sem mim, por me receber do México com todo o amor acumulado, por pesquisar na Internet sobre toxinas antimicrobianas, mesmo sem saber o que era isso.... E por transformar minha vida em um conto de fadas!

Agradeço imensamente a minha querida Mamis, Matilde, por ser a base da minha vida, por ter investido em mim, por ser meu exemplo de luta e perseverança, pelo apoio no desenvolvimento desse Mestrado e por me amar sempre. Amo você Mamis!

A minha querida irmã Évelyn, por me acordar pra fazer a dissertação, ser minha companheira por toda a vida, minha amiga incondicional...

Agradeço ao Cnpq pelo financiamento e bolsa de estudos de Mestrado.

Ao programa de Pós-graduação em Biologia Animal e ao Instituto de Biologia da Universidade de Brasília pelo apoio e oportunidade de trabalho.

E mais importante, agradeço à Deus que me deu a dádiva da vida, rodeou ela de amigos e anjos-da-guarda e que permitiu a conquista de mais essa vitória.

# SUMÁRIO

<b>INTRODUÇÃO</b> .....	1
1. Escorpiões.....	1
2. Composição da peçonha de escorpiões.....	2
2.1 Canal para Na <sup>+</sup> voltagem-dependente e toxinas escorpiônicas moduladoras de canal para Na <sup>+</sup> (NavTx).....	3
2.2 Canal para K <sup>+</sup> e toxinas escorpiônicas bloqueadoras de canal para K <sup>+</sup> .....	5
2.3 Toxinas escorpiônicas moduladoras de canal para Ca <sup>2+</sup> e moduladoras de canal para Cl <sup>-</sup> .....	6
2.4 Toxinas sem pontes dissulfeto.....	7
✓ Antimicrobianas.....	7
✓ Imunomoduladores.....	8
✓ Potencializadores de Bradicinina.....	9
✓ Hemolíticos.....	9
3. Estudos de proteômica e transcriptoma de peçonha de escorpiões.....	9
4. <i>Opisthacanthus cayaporum</i> .....	11
<b>RELEVÂNCIA CIENTÍFICA</b> .....	14
<b>OBJETIVOS</b> .....	15
✓ Gerais.....	15
✓ Específicos.....	15
<b>MATERIAIS E MÉTODOS</b> .....	16
1. Soluções e Meios de cultivo.....	16
2. Coleta e manutenção dos escorpiões.....	16
3. Construção da Biblioteca de cDNA.....	16
3.1 Extração do RNA total.....	16
3.2 Construção do Banco de cDNA.....	17
3.3 Ligação ao vetor pDNR-LIB (Clontech/UK).....	20
4. Transformação bacteriana.....	20
5. Seqüenciamento automático de DNA.....	21

6. Caracterização da Cayaporina.....	22
5.1 Obtenção da Cayaporina.....	23
5.2 Determinação da seqüência de aminoácidos .....	23
5.3 Atividade Biológica.....	23
<b>RESULTADOS</b> .....	25
1. Biblioteca de cDNA.....	25
2. Seqüência de nucleotídeos obtidos.....	25
Seqüências relacionadas a prováveis toxinas.....	29
Seqüências com cisteínas (cys) que formariam pontes dissulfeto..	29
Prováveis fosfolipases.....	41
Seqüências sem cisteínas.....	44
Seqüências de nucleotídeos sem função atribuída.....	50
3. Descrição do peptídeo Cayaporina.....	54
<b>DISCUSSÃO</b> .....	57
1. Biblioteca de cDNA.....	57
2. Seqüências de nucleotídeos obtidos.....	59
Prováveis Toxinas.....	59
Seqüências com cisteínas que formariam com pontes dissulfeto..	61
Toxinas sem cisteínas.....	65
Componentes do processamento celular.....	66
✓ Enzimas.....	66
✓ Proteínas estruturais.....	67
✓ Metabolismo.....	67
✓ Gene mitocondrial.....	67
Transcritos sem função atribuída.....	68
Cayaporina.....	69
<b>CONCLUSÃO</b> .....	70
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	71
<b>ANEXOS</b> .....	78

## ABREVIATURAS

<b>ACN</b> .....	Acetonitrila
<b>ED50</b> .....	Dose efetiva para o efeito farmacológico em 50% dos espécimes testados
<b>HPLC</b> .....	Cromatografia Líquida de alta eficiência
<b>IBAMA</b> .....	Instituto Brasileiro do Meio Ambiente e Recursos Naturais
<b>LD-PCR</b> .....	<i>Long distance</i> PCR
<b>MMLV</b> .....	<i>Moloney murine leukemia vírus</i>
<b>PCR</b> .....	Reação em cadeia da Polimerase
<b>RP-HPLC</b> .....	Cromatografia Líquida de alta eficiência em coluna de fase reversa
<b>rpm</b> .....	Rotações por minuto
<b>TFA</b> .....	Ácido Trifluoroacético
<b>SMART</b> .....	<i>Switching Mechanism At 5' end of RNA transcript</i>
<b>dNTP Mix</b> .....	Desoxirribonucleotídeos Fosfatados: Adenina (dATP), Citosina (dCTP), Guanina (dGTP) e Tirosina (dTTP)
<b>DTT</b> .....	Ditiotreitol ou Reagente de <i>Cleland</i>
<b>BLAST</b> .....	<i>Basic Local Alignment Search Tool</i>
<b>ATCC</b> .....	<i>American Type Culture Collection</i>
<b>MIC</b> .....	Concentração Inibitória Mínima para inibição de 100% do crescimento de bactérias
<b>EST</b> .....	<i>Expressed sequence tag</i> (seqüências expressa)
<b>ORF</b> .....	<i>Open Reading Frame</i>
<b>NDBP</b> .....	Peptídeos sem pontes dissulfeto



## ÍNDICE DE FIGURAS E TABELAS

<b>Figura 1:</b> Representação esquemática da morfologia geral dos escorpiões.....	2
<b>Figura 2:</b> Representação esquemática do canal para Na <sup>+</sup> voltagem-dependente.....	4
<b>Figura 3:</b> Representação esquemática do canal para K <sup>+</sup> voltagem-dependente.....	6
<b>Figura 4:</b> Escorpião brasileiro <i>Opisthacanthus cayaporum</i> .....	12
<b>Figura 5:</b> Mapa do vetor pDNR-LIB (Clontech/UK).....	19
<b>Figura 6:</b> Freqüência de clones em relação ao tamanho do inserto em pares de bases.....	26
<b>Figura 7:</b> Intervalo de distribuição do tamanho das 130 ESTs obtidas.....	26
<b>Figura 8:</b> Esquema mostrando o agrupamento das 130 ESTs da biblioteca de cDNA.....	27
<b>Figura 9:</b> Distribuição das 130 ESTs obtidas.....	29
<b>Figura 10:</b> Seqüências de nucleotídeos e predição da seqüência de aminoácidos de Ocy10.....	30
<b>Figura 11:</b> Alinhamento da seqüência de aminoácidos de Ocy10 com a insetotoxina LIT1 isolada do escorpião <i>Liocheles australasiae</i> .....	31
<b>Figura 12:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de OcyKtx1.....	32
<b>Figura 13:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy9 que apresentou 85% de identidade com a Omtx1 de <i>Opisthacanthus madagascariensis</i> .....	32
<b>Figura 14:</b> Alinhamento da seqüência de aminoácidos de OcyKtx1 e Ocy9 com as Toxinas para canal de potássio Omtxs e KappaTxs.....	33
<b>Figura 15:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy11.....	33
<b>Figura 16:</b> Comparação da seqüência de aminoácidos de Ocy11 com a seqüência de TXLP1, que foi isolada do escorpião <i>Buthus martensii</i> com o peptídeo LaI isolado da peçonha do escorpião <i>Liocheles australiasiae</i> .....	34
<b>Figura 17:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de OcyKtx2 que apresentou identidade com a toxina para canal de potássio IsTX.....	35
<b>Figura 18:</b> Comparação da seqüência de aminoácidos predita de OcyKtx2 com as toxinas pertencentes à subfamília 6 das $\alpha$ KTxs.....	36
<b>Figura 19:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy40.....	37
<b>Figura 20:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy41.....	38
<b>Figura 21:</b> Comparação da seqüência de aminoácidos de Ocy40 e Ocy41 com precursor nomeado Venom peptide 2 identificado no escorpião <i>Buthus martensii</i> .....	38
<b>Figura 22:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy6.....	39
<b>Figura 23:</b> Comparação da seqüência de aminoácidos de Ocy6 com o peptídeo nomeado Is6 isolado do carrapado <i>Ixodes scapularis</i> .....	39
<b>Figura 24:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy7.....	40

<b>Figura 25:</b> Comparação da seqüência de aminoácidos de Ocy7 com o precursor nomeado Opiscorpina2 isolado do escorpião <i>O. carinatus</i> , com a Scorpina isolada da peçonha do escorpião <i>Pandinus imperator</i> e com Ocyescorpina.....	40
<b>Figura 26:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos da Ocyfosfolipase1.....	42
<b>Figura 27:</b> Alinhamento da seqüência de aminoácidos Ocyfosfolipase1 com a Fosfolipase C.....	43
<b>Figura 28:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocyfosfolipase2.....	44
<b>Figura 29:</b> Alinhamento da seqüência de aminoácidos de Ocyfosfolipase2 com a seqüência da Phaiodactylipina.....	44
<b>Figura 30:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos Ocy1.....	45
<b>Figura 31:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy2.....	46
<b>Figura 32:</b> Alinhamento da seqüência de aminoácidos de Ocy1 e Ocy2 com os peptídeos antimicrobianos pertencentes à subfamília 5 dos NDBPs.....	46
<b>Figura 33:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy4, que apresentou identidade com o peptídeo antimicrobiano Pandinin2.....	47
<b>Figura 34:</b> Alinhamento da seqüência de aminoácidos de Ocy4 com os peptídeos antimicrobianos pertencentes à subfamília 4 dos NDBPs.....	48
<b>Figura 35:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos da Cayaporina.....	49
<b>Figura 36:</b> Alinhamento da seqüência de aminoácidos da Cayaporina com os componentes da subfamília 3 dos NDBPs.....	49
<b>Figura 37:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy12.....	50
<b>Figura 38:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy13.....	51
<b>Figura 39:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy14.....	52
<b>Figura 40:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy15.....	52
<b>Figura 41:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy16.....	53
<b>Figura 42:</b> Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy17.....	54
<b>Figura 43:</b> Purificação da cayaporina em HPLC.....	55
<b>Figura 44:</b> Espectro, Fragmentação do peptídeo Cayaporina e Seqüência parcial de aminoácidos obtida por Edman.....	56
<b>Tabela 1:</b> Peptídeos escorpiônicos sem pontes dissulfeto.....	8
<b>Tabela 2:</b> Identificação das 39 seqüências nucleotídicas e prováveis funções.....	27
<b>Tabela 3:</b> Comparação de alguns dados obtidos na biblioteca de cDNA de <i>O. cayaporum</i> com os outros trabalhos de transcriptoma.....	58

## RESUMO

A peçonha de escorpiões é uma complexa mistura de peptídeos que modulam canais iônicos, antimicrobianos, potencializadores de bradicinina, hemolíticos, e com outras atividades biológicas, sendo importante alvo de estudos para o desenvolvimento de novos fármacos. Os estudos de caracterização proteômica, principalmente com escorpiões da família Buthidae, têm descrito várias toxinas, contudo há pouco conhecimento sobre os processos celulares da glândula de peçonha. A metodologia de busca por ESTs tem sido utilizada em vários trabalhos porque é um método rápido e confiável para a detecção dos genes expressos na glândula, que sintetizam produtos conhecidos ou não. O presente trabalho é o primeiro estudo de caracterização dos transcritos da glândula de peçonha do escorpião brasileiro *Opisthacanthus cayaporum* (Ischnuridae), por meio da construção de uma biblioteca de cDNA. Foram analisadas 130 ESTs, resultando em 67 seqüências de nucleotídeos distintas com alta qualidade, sendo formadas por em média 535 pares de bases (variando entre 139 e 902 nucleotídeos). Das 130 ESTs, 30% são prováveis toxinas de *O. cayaporum*, (7 contigs e 8 singletos), 36% correspondem a proteínas relacionadas a diferentes processos celulares (16 singletos e 4 contigs), 17% são seqüências que não foi possível atribuir função, pois apresentaram expectância  $>10^{-5}$ , 15% seqüências que não foi encontrada identidade com peptídeos depositados no GenBank e 6% correspondem a seqüências caracterizadas de genomas. Foram descritas 15 seqüências distintas com identidade a toxinas, sendo 9 semelhantes à toxinas bloqueadoras de canais para potássio, 4 toxinas sem pontes dissulfeto e com atividade antimicrobiana e 2 seqüências com identidade a fosfolipases. As seqüências de aminoácidos obtidas a partir da caracterização da biblioteca de cDNA, neste trabalho, apresentaram distribuição de massas moleculares semelhante àquelas descrita pela análise proteômica de *O. cayaporum*. A seqüência nomeada Cayaporina (NDBP 3.7), de massa teórica calculada de 4675 Da, foi purificada da peçonha bruta e teve sua seqüência de aminoácidos obtida por seqüenciamento *de novo* e por degradação de Edman. Esse peptídeo possui atividade antimicrobiana contra *Escherichia coli*, em concentração de 28 $\mu$ M e também inibiu o crescimento em 56% de *Staphylococcus aureus*, não tendo atividade hemolítica em hemácias humanas. Este trabalho demonstrou a riqueza e diversidade de componentes da peçonha dessa espécie, e foi caracterizado o primeiro peptídeo antimicrobiano não hemolítico isolado da peçonha de *O. cayaporum*.

**Palavras-chave:** transcriptoma, biblioteca de cDNA, peptídeos antimicrobianos, *Opisthacanthus cayaporum*.

## ABSTRACT

Scorpions venom are a complex mixture of peptides that modulate ion channels, antimicrobials, bradykinin-potentiating, hemolytic, and other biological activities, and they are a potential source of pharmacological agents and physiological tools. The proteomics characterization, specially with the family Buthidae, have described several toxins, but there is little knowledge about the cellular processes of the gland. The ESTs approach has been used in several works because it is a rapid and reliable method for gene discovery, which synthesizes products known or not. This work is the first study to characterize the transcripts of the venom gland of the scorpion Brazilian *Opisthacanthus cayaporum* (Ischnuridae), by constructing the library of cDNA. We analyzed 130 ESTs, resulting in 67 unique sequences with high quality and is formed by an average of 535 bp (ranging between 139 and 902bp). Of the 130 ESTs, 30% are toxins-like sequences (7 contigs and 8 singlets), 36% are sequences involved in gene and protein expression (16 singlets and 4contigs), 17% are sequences that was not possible to assign function and that may be new toxins have not described, 15% sequences that was not found identity with peptides deposited in GenBank and 6% are characterized by sequences of genomes. Fifteen unique sequences had identity with the toxins, nine similar to potassium-channel-specific toxin, four NDBPs with antimicrobial activity and two sequences with identity with fosfolipases. Data obtained by the library of cDNA were linked with the molecular masses of *O. cayaporum* proteomics characterization, and observed that there is similarity in the distribution of molecular masses. The sequence named Cayaporina, theoretical mass calculated is 4675 Da, was purified from crude venom and had its sequence of amino acids obtained by *de novo* sequencing and Edman degradation. This peptide has antimicrobial activity against *Escherichia coli*, a concentration of 28  $\mu$ M also inhibited the growth in 56% of *Staphylococcus aureus*, not having hemolytic activity in human erythrocyte. This study demonstrated the richness and diversity of components of the *O. cayaporum* venom, and was characterised the first antimicrobial peptide not hemolytic isolated from the *O. cayaporum* venom, appointed Cayaporina (systematic name: NDBP 3.7).

**Key words:** transcriptome, cDNA library, antimicrobial peptide, *Opisthacanthus cayaporum*.

# INTRODUÇÃO

## 1. Escorpiões

Escorpiões estão entre os animais mais antigos da Terra (seu registro fóssil mais antigo data do período Siluriano) e têm sobrevivido há mais de 400 milhões de anos. Durante esse período evolucionário, os escorpiões têm mantido seus padrões morfológicos, o que pode ser devido ao desenvolvimento de aparatos de proteção e inoculação de peçonha eficientes que permitiram sua sobrevivência (Froy *et al.*, 1999).

Os escorpiões pertencem ao filo Arthropoda, sub-filo Chelicerata, classe Arachnida e à ordem Scorpiones. Existem 18 famílias às quais pertencem cerca de 1500 espécies e sub-espécies localizadas em todo o mundo, predominantemente em regiões tropicais e subtropicais. Em geral possuem hábitos noturnos e furtivos, podendo ser encontrados sob fendas e buracos tanto em ambiente urbano quanto em matas, desertos e zonas úmidas (Barnes, 2005; Lourenço, 2003).

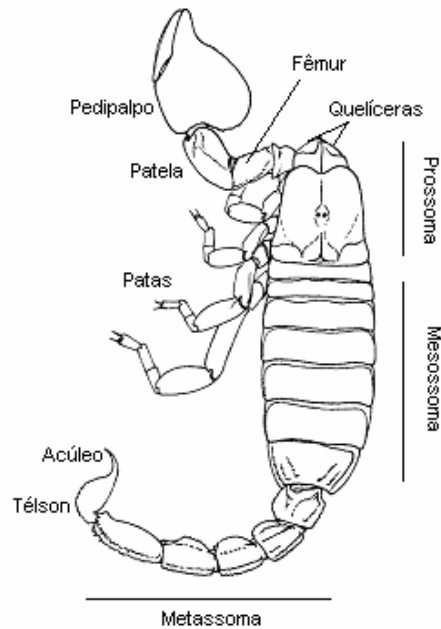
Os escorpiões possuem o corpo dividido em prossoma e opistossoma sendo que este último é dividido em mesossoma e metassoma (Figura 1). Na região de prossoma estão localizados dois olhos frontais e até cinco olhos em cada lateral. A esta região estão ligados os 4 pares de patas, 1 par de quelíceras e 1 par de pedipalpos. O mesossoma é formado por sete segmentos e o metassoma é formado por 5 sendo que no último forma-se o télson que contém um par de glândulas produtoras de peçonha e um acúleo (Barnes, 2005; Lourenço, 2003).

O Brasil conta com quatro famílias de escorpiões:

- ✓ Bothriuridae: Gêneros *Bothiurus*, *Brachistoternus*, *Urophonius* e *Thestylus*;
- ✓ Buthidae: Gêneros *Tityus*, *Ananteris*, *Rhopaluros*, *Microtityus* e *Isometrus*;
- ✓ Chactidae: Gêneros *Brotheas*, *Broteochactas*, *Teuthraustes*, *Chactopis* e *Chactas*;
- ✓ Ischunuridae (sinônimo Hemiscorpiidae): Gênero: *Opisthacanthus*.

No Brasil, acidentes por escorpiões são considerados de importância médico-sanitária, não só pela incidência, mas também pela potencialidade da peçonha de algumas espécies que pode levar a quadros clínicos graves, às vezes fatais, principalmente em crianças. As

principais espécies causadoras de acidentes são pertencentes à família Buthidae, principalmente *Tityus serrulatus*, *T. stigmurus* e *T. bahiensis* (Lourenço, 2003).



**Figura 1: Representação esquemática da morfologia geral dos escorpiões.** Visão dorsal (Lourenço, 2003, com adaptações).

## 2. Composição da peçonha de escorpiões

A peçonha de escorpiões é uma mistura complexa de centenas de componentes de origem apócrina, produzida por um par de glândulas exócrinas altamente especializadas que estão localizadas no último segmento pós-abdominal, nomeado de télson. A peçonha é utilizada na captura e imobilização das presas, como mecanismo de defesa e como demonstrado para a espécie *Opisthacanthus madagascariensis* no acasalamento (Yamaji *et al.*, 2004).

A peçonha de escorpiões contém enzimas, lipídeos, nucleotídeos, amins biogênicas, peptídeos e vários outros componentes ainda não identificados (Batista *et al.*, 2004). Os componentes de maior abundância encontrados na peçonha são os peptídeos responsáveis pelos efeitos neurotóxicos relacionados à picada e, em estudos com várias espécies de escorpiões, já foram descritos mais de 350 desses peptídeos (para revisão ver: Tan *et al.*,

2006). Todos esses peptídeos, que divergem funcionalmente, podem ter evoluído a partir de um gene ancestral comum através de derivações genéticas como polimorfismo, duplicação gênica, *trans-splicing* ou *splicing* alternativo (Zhijian *et al.*, 2006).

As neurotoxinas escorpiônicas podem ser classificadas com base em diferentes critérios, como tamanho dos peptídeos, sua ordem cronológica de descrição, a toxicidade ao homem, ou à especificidade íon-canal (Catterall *et al.*, 2007). De acordo com este último critério, esses peptídeos podem ser divididos em 4 famílias: família 1 contém os peptídeos que modulam a atividade dos canais para  $\text{Na}^+$  (para revisão ver: Rodriguez de la Vega & Possani, 2005); família 2 que compreende peptídeos que bloqueiam canais para  $\text{K}^+$  (para revisão ver: Rodriguez de la Vega & Possani, 2004); família 3 contém peptídeos que inibem canais para  $\text{Cl}^-$  (DeBin *et al.*, 1993; Zhijian *et al.*, 2006; Fu *et al.*, 2007); e família 4 que inclui peptídeos que modulam os canais para  $\text{Ca}^{2+}$  sensíveis a ryanodina (para revisão ver: Valdivia *et al.*, 1989).

As toxinas escorpiônicas possuem alto grau de letalidade, sendo comparáveis às toxinas de serpentes, são estáveis, sendo resistentes a mudanças do meio como, por exemplo, ao aumento de pH, uso de enzimas proteolíticas, agentes desnaturantes ou elevação da temperatura.

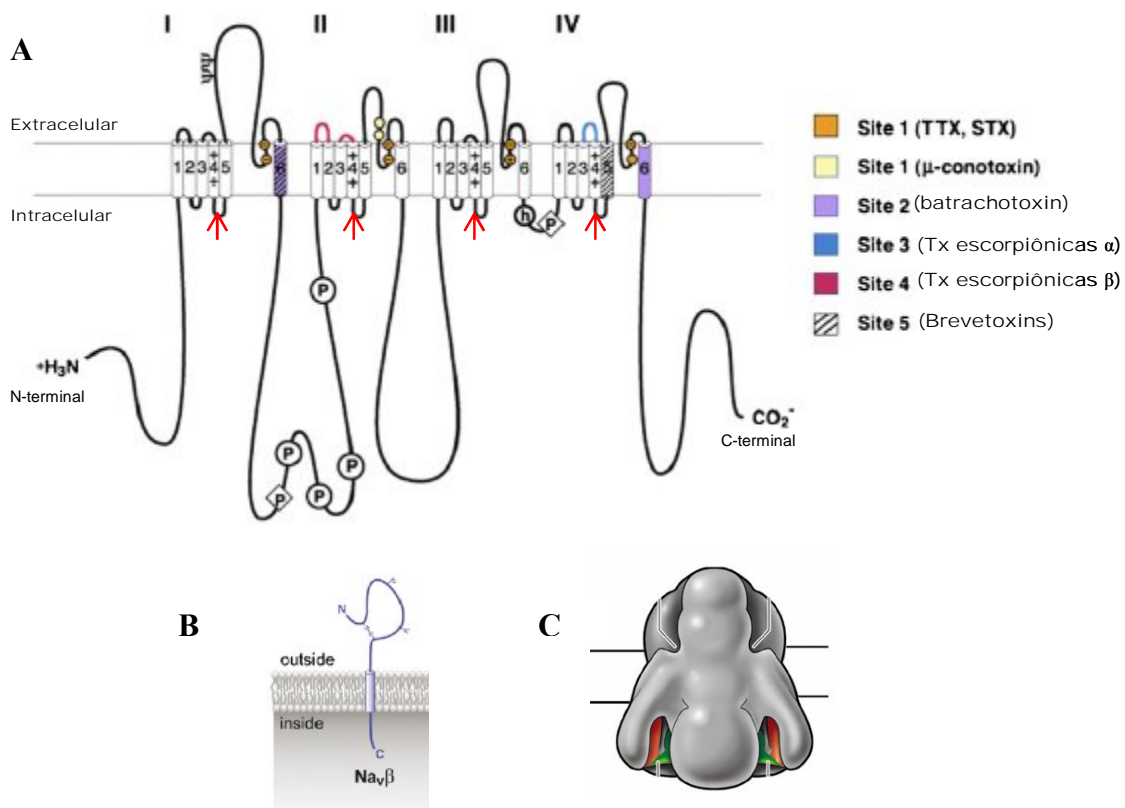
Atualmente, a alta especificidade e seletividade das toxinas animais que agem em canais iônicos têm despertado o interesse e estudos com essas toxinas, para o desenvolvimento de novos fármacos, além de serem importantes ferramentas no estudo de caracterização estrutural e funcional dos canais iônicos (Becerril, *et al.*, 1997; Catterall *et al.*, 2007). Essas toxinas também são importantes alvos para o desenvolvimento de novos fármacos, produção em larga escala de anti-venenos eficientes e o aprimoramento dos conhecimentos sobre evolução e composição das peçonhas.

### **2.1. Canal para $\text{Na}^+$ voltagem-dependente e toxinas escorpiônicas moduladoras de canal para $\text{Na}^+$ voltagem-dependente ( $\text{Na}_v\text{Tx}$ )**

Os canais para  $\text{Na}^+$  voltagem-dependentes ( $\text{Na}_v$ ) são responsáveis pelo rápido influxo de íons  $\text{Na}^+$  na fase de despolarização (fase ascendente) do potencial de ação em neurônios, células musculares e endócrinas. Através de estudos de reconstituição funcional dos canais para  $\text{Na}^+$  com neurotoxinas, sabe-se que os canais  $\text{Na}_v$  são compostos por duas

subunidades: subunidade- $\alpha$  com 220-260 kDa que contém o sensor de voltagem e os elementos de formação do poro, e uma ou duas subunidades- $\beta$  auxiliares com 33-36 kDa (Cestéle & Catterall, 2000).

As subunidades- $\alpha$  dos  $\text{Na}_v$  estão compostas por quatro domínios homólogos, sendo que cada um contém seis segmentos hidrofóbicos transmembrânicos (Figura 2). O sensor de voltagem do canal está localizado no segmento S4 que contém repetidas seqüências de resíduos de aminoácidos carregados positivamente seguido por dois resíduos hidrofóbicos. Esses sítios são responsáveis pela mudança conformacional que leva à abertura do canal e que é causada pela mudança da polarização da membrana (Catterall *et al.*, 2007). Foram funcionalmente caracterizadas nove  $\alpha$ -subunidades, designadas  $\text{Na}_v1.1$  a  $\text{Na}_v1.9$ , e estas possuem seqüências de aminoácidos, dos segmentos transmembrânicos, com mais de 70% de identidade entre si.



**Figura 2: Representação esquemática do canal para  $\text{Na}^+$  voltagem-dependente.** **A:** Diagrama representando a subunidade- $\alpha$  de um  $\text{Na}_v$  de mamífero. Os cilindros representam os 6 segmentos em alfa-hélice transmembrânicos de cada um dos quatro domínios (I a IV). O poro é formado pelos segmentos S5 e S6 de cada domínio. Os sítios de ligação de algumas toxinas estão indicados em cores. **B:** subunidade- $\beta$  auxiliar de  $\text{Na}_v$  (Catterall *et al.*, 2007, com modificações). **C:** Estrutura tridimensional esquemática da subunidade- $\alpha$  de um canal para  $\text{Na}^+$  de mamífero, obtida por eletromicrografia. (Catterall & Yu, 2003, com adaptações).

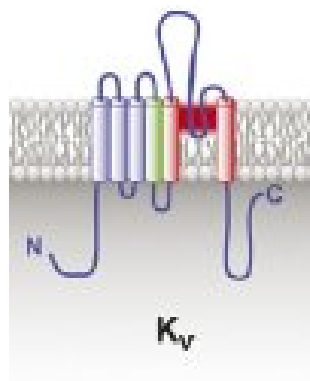


As  $\beta$ -subunidades dos  $\text{Na}_v$  são importantes elementos na família de proteínas associadas a canais iônicos e foram identificadas apenas em vertebrados ( $\text{Na}_v\beta 1$  a  $\text{Na}_v\beta 4$ ). Essas subunidades possuem apenas um segmento transmembrânico e um longo domínio N-terminal extracelular e interagem com diferentes sítios das subunidades- $\alpha$ , modulando suas funções (Catterall *et al.*, 2007).

As toxinas escorpiônicas que modulam canais para  $\text{Na}^+$ ,  $\text{Na}_v\text{Tx}$ , possuem de 60-70 resíduos de aminoácidos unidos por 4 pontes dissulfeto e podem ser classificadas em  $\alpha$  ou  $\beta$ -toxinas de acordo com seus efeitos eletrofisiológicos (Rodriguez de la Vega & Possani, 2005). As  $\alpha$ -toxinas se ligam ao sítio 3 dos canais para  $\text{Na}^+$ , tornando mais lento ou bloqueando o mecanismo de inativação desses canais, e as toxinas do tipo  $\beta$  se ligam ao sítio 4 independentemente do potencial de membrana e afetam a ativação do canal para sódio (Rodriguez de la Vega & Possani, 2005). Os peptídeos moduladores de canais para  $\text{Na}^+$  estão sendo estudados visando a compreensão do funcionamento dos canais para  $\text{Na}^+$  e as interações entre eles e seus ligantes. Já foram descritas 430 toxinas de escorpião para canal para sódio (Banco de dados Scorpion2: <http://sdmc.i2r.a-star.edu.sg/scorpion>).

## **2.2. Canal para $\text{K}^+$ e toxinas escorpiônicas moduladoras de canal para $\text{K}^+$ voltagem-dependente ( $\text{K}_v\text{Tx}$ )**

Os canais para potássio voltagem-dependentes ( $\text{K}_v$ ) são ativados na fase de despolarização e o fluxo de íons  $\text{K}^+$  através desses canais leva à repolarização da membrana, retornando assim ao potencial de repouso. Os canais para potássio voltagem-dependentes são compostos por quatro subunidades alfa, com massa molecular de 70 kDa, que são compostas por seis segmentos transmembrânicos que são homólogos aos encontrados em canais para  $\text{Na}^+$  (Figura 3). Os  $\text{K}_v$ s são conhecidos por sua diversidade, conhece-se 40 diferentes canais  $\text{K}_v$  que estão distribuídos em 12 subfamílias ( $\text{K}_v 1$  a  $\text{K}_v 12$ ) classificadas pela da similariedade apresentada entre suas seqüências de aminoácidos. Essa diversidade de canais para potássio permite que neurônios e outras células excitáveis controlem suas propriedades de sinalização elétrica por meio da expressão de diferentes combinações de subunidades de  $\text{K}_v$  (Catterall *et al.*, 2007).



**Figura 3: Representação esquemática do canal para  $K^+$  voltagem-dependente.** Diagrama representando um Kv. Em vermelho tem-se os segmentos S5 e S6 que formam o poro do canal, o sensor de voltagem está em verde (segmento S4), em azul os segmentos transmembrânicos em alfa-hélice.

Os peptídeos que atuam em canais para potássio (KTxs) são estabilizados por 3 ou 4 pontes dissulfeto, possuem de 23 a 64 resíduos de aminoácidos e podem ser classificados em 4 subfamílias:  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$ -Ktxs (Tytgat *et al.*, 1999). O número de toxinas novas descritas, específicas para canais de  $K^+$  tem aumentado significativamente nos últimos anos, porém poucos desses peptídeos tiveram sua atividade estudada em canais para  $K^+$ . Entre os mais de 40 tipos diferentes de canais para potássio descritos, apenas um número pequeno de canais tem um peptídeo identificado que seja capaz de afetar sua função (Rodriguez de la Vega & Possani, 2004). Atualmente, 290 peptídeos de escorpião foram descritos como toxinas para canal para potássio (Banco de dados Scorpion2: <http://sdmc.i2r.a-star.edu.sg/scorpion>).

### **2.3. Toxinas escorpiônicas moduladoras de outros canais iônicos (canais para $Cl^-$ e canais para $Ca^{2+}$ voltagem-dependentes)**

Os canais para  $Ca^{2+}$  voltagem-dependente são importantes na condução do sinal elétrico, na despolarização da membrana celular, na contração, secreção, neurotransmissão e outros em outros mecanismos (Catterall, 2005). Esses canais são formados por uma subunidade  $\alpha_1$  (com 212-250 kDa) semelhante à encontrada em canais para  $Na^+$ , com 4 domínios e 6 segmentos transmembrânicos, e subunidades menores nomeadas  $\alpha_2$ ,  $\beta$ ,  $\gamma$  e  $\delta$ . Foram caracterizados 5 toxinas para canais para  $Ca^{2+}$  em vertebrados (Banco de dados Scorpion2: <http://sdmc.i2r.a-star.edu.sg/scorpion>).

As toxinas escorpiônicas que atuam em canais para Cl<sup>-</sup> são compostos de 35-38 resíduos de aminoácidos estabilizados por 4 pontes dissulfeto e se diferenciam estruturalmente das toxinas para canais de Na<sup>+</sup> e K<sup>+</sup>, contudo há poucos estudos sobre as toxinas dessa família. Nesta família, destaca-se o peptídeo nomeado *Clorotoxina*, isolado da peçonha do escorpião *Leiurus quinquestriatus* (Buthidae). A *clorotoxina* bloqueia canais para Cl<sup>-</sup> em epitélio de camundongo e apresentou ligação específica em canais para Cl<sup>-</sup> em células da glia com modificações patológicas (Zhijian *et al.*, 2006). Uma toxina com 68% de identidade a essa clorotoxina foi descrita no escorpião *Buthus martensii* e foi nomeada rBmK Cta (Fu *et al.*, 2007). A toxina rBmK Cta apresentou capacidade de inibir o crescimento de gliomas humanos de forma dose-dependente, em testes *in vivo* e *in vitro*, sendo este um resultado importante para o desenvolvimento de novas drogas antitumorais (Fu *et al.*, 2007). Atualmente têm-se descrito 17 toxinas de escorpião para canal para Cl<sup>-</sup> (Banco de dados Scorpion2: <http://sdmc.i2r.a-star.edu.sg/scorpion>).

#### **2.4. Toxinas escorpiônicas sem pontes dissulfeto**

Centenas de peptídeos com pontes dissulfeto vêm sendo descritos da peçonha de várias espécies de escorpiões, contudo há pouco conhecimento acerca dos peptídeos sem pontes dissulfeto. Esses peptídeos são interessantes alvos de estudos devido às suas atividades como: antimicrobianos, imunomoduladores, hemolíticos, antitumorais etc. Segundo a atividade biológica, o tamanho do peptídeo e a similaridade entre estruturas primárias, esses peptídeos podem ser classificados em 6 subfamílias (Zeng *et al.*, 2005) (Tabela 1).

##### **✓ Antimicrobianas**

Os peptídeos antimicrobianos podem desestabilizar estruturalmente as membranas celulares, modificar sua permeabilidade iônica e influenciar na função e sinalização celular interagindo com proteínas da transdução de sinal. Estes peptídeos, atuando de forma não específica, inibem o crescimento de vários microrganismos como bactérias gram-positivas e negativas, leveduras e fungos filamentosos, e são importantes alvos para o desenvolvimento de novos antibióticos (Zeng *et al.*, 2005).

A presença de antimicrobianos na peçonha de escorpiões sugere que a glândula pode ser protegida por esses peptídeos, atuando no sistema imune local. Por meio de

experimentos com inoculação bacteriana na glândula do escorpião *Buthus martensii* e monitoramento da expressão gênica de dois peptídeos antimicrobianos, BmKb1 e BmKn2, foi constatado o aumento na expressão dos genes para esses peptídeos quando havia infecção bacteriana na glândula (Gao *et al.*, 2007).

**Tabela 1: Peptídeos escorpiônicos sem pontes dissulfeto.** A classificação apresentada refere-se ao modelo sugerido por Zeng *et al.*, 2005, em que o primeiro número indica a subfamília e o segundo a posição na subfamília, de acordo com a data de descrição do peptídeo. “PB” Potencializador de Bradicinina, “Antm” Antimicrobiano, “Hem” Hemolítico, “Antf” Antifúngico, “Imun” Imunomodulador.

Peptídeo	Nº aa	Espécie	Atividade Biológica	Classificação
Peptide T	13	<i>Tityus serrulatus</i>	PB	1.1
K 12	21	<i>Buthus occitanus</i>	PB	1.2
TsHpt-1	25	<i>Tityus serrulatus</i>	PB	1.3
Bs10	34	<i>Buthus indicus</i>	Desconhecida	2.1
Hadrurina	41	<i>Pandinus imperator</i>	Antm e Hem	3.1
Parabutoporina	45	<i>Parabuthus schlechteri</i>	Antm, Imun, Hem e Antf	3.2
BmKbpp	47	<i>Buthus martensii</i>	Antm	3.3
Pandinina1	44	<i>Pandinus imperator</i>	Antm, Hem e Antf	3.4
Opistorporina1	44	<i>Opisthophthalmus carinatus</i>	Antm, Imun, Hem e Antf	3.5
Opistorporina2	44	<i>Opisthophthalmus carinatus</i>	Antimicrobiano	3.6
Pandinina2	24	<i>Pandinus imperator</i>	Antm, Hem e Antf	4.1
BmKb1	18	<i>Buthus martensii</i>	Antm	4.2
BmKn1	13	<i>Buthus martensii</i>	Desconhecida	5.1
IsCT1	13	<i>O. madagascariensis</i>	Antm e Hem	5.2
IsCT2	13	<i>O. madagascariensis</i>	Antm e Hem	5.3
BmKn2	13	<i>Buthus martensii</i>	Antm	5.4
BmKa1	37	<i>Buthus martensii</i>	Desconhecida	6.1
BmKa2	50	<i>Buthus martensii</i>	Desconhecida	6.2

### ✓ Imunomoduladores

Alguns peptídeos antimicrobianos isolados de escorpião como as Opistorporinas1 e 2 e a parabutoporina, em baixas concentrações podem estimular e em concentrações maiores podem inibir a atividade de neutrófilos humanos, além de aumentar a concentração

intracelular de proteínas que levam à degranulação de granulócitos humanos (Willems *et al.*, 2002).

#### ✓ **Potencializadores de Bradicinina**

A bradicinina, juntamente com a angiotensina II, é um importante fator na regulação da pressão sanguínea, agindo como agente hipotensor. O peptídeo T isolado do escorpião *Tityus serrulatus* (Ferreira *et al.*, 1993) e o peptídeo K12, isolado da espécie *Buthus occitanus* (Meki *et al.*, 1995), foram caracterizados como potencializadores de bradicinina. Em recente trabalho, utilizando abordagem de caracterização proteômica, foi descrita uma nova família de hipotensinas do escorpião *Tityus serrulatus*. O peptídeo TsHpt-1, pertencente a essa família, apresentou atividade hipotensora em ratos normotensos, na forma nativa ou sintética mesmo sem a presença de bradicinina (Verano-Braga *et al.*, 2008). Estes peptídeos podem vir a ser utilizados para o desenvolvimento de novas drogas anti-hipertensivas.

#### ✓ **Hemolíticos**

Os peptídeos de escorpião sem pontes dissulfeto que apresentam atividade hemolítica, como os peptídeos IsCT1 e 2, podem ser utilizados no tratamento de coágulos sanguíneos (Zeng *et al.*, 2005).

### **3. Estudos de caracterização proteômica e transcriptoma em animais peçonhentos**

Os primeiros trabalhos de caracterização de componentes da peçonha de escorpiões brasileiros foram realizados com a espécie *Tityus serrulatus* (Diniz *et al.*, 1966; Gomes & Diniz, 1966). Depois desses trabalhos foram realizados estudos de caracterização proteômica com outras espécies do gênero *Tityus*, como por exemplo, *T. serrulatus* (Pimenta *et al.*, 2001), *T. cambridgei* (Batista *et al.*, 2004), *T. costatus* (Diego-Garcia *et al.*, 2005), *T. pachyurus* (Barona *et al.*, 2006) e *T. discrepans* (Batista *et al.*, 2006).

Outros trabalhos interessantes de caracterização proteômica foram realizados, como por exemplo, a caracterização da primeira neurotoxina do escorpião *Tityus fasciolatus* (Wagner *et al.*, 2003) e a caracterização da menor toxina bloqueadora de canal para potássio, com 23 resíduos de aminoácidos, isolada de *Tityus cambridgei* (Batista *et al.*, 2000).

A identificação de numerosas toxinas de escorpiões, utilizando-se ferramentas de análise de caracterização proteômica e de busca direcionada ao genoma através de oligonucleotídeos iniciadores, têm fornecido uma vasta coleção de seqüências relacionadas aos componentes da peçonha. Contudo estas seqüências se referem aos produtos de síntese que serão secretados nas glândulas. Desta forma, há pouco conhecimento sobre os processamentos que normalmente estão presentes na expressão, síntese e secreção das proteínas que formam a mistura da peçonha nas glândulas. A busca não dirigida através do transcriptoma de glândula de peçonha revela, além de novos componentes da peçonha e seus precursores, proteínas que estão envolvidas nos processos celulares da glândula na produção da peçonha, permitindo assim um maior conhecimento sobre a produção de toxinas animais (Schwartz *et al.*, 2007).

A metodologia de identificação das ESTs (*expressed sequence tag*) tem sido utilizada em vários trabalhos porque é um método rápido e confiável para a detecção dos genes expressos na glândula, que sintetizam produtos conhecidos ou não (Magalhães *et al.*, 2006; Schwartz *et al.*, 2007). As ESTs são seqüências geradas da biblioteca de cDNA que correspondem aos RNA mensageiros, que potencialmente serão traduzidos em componentes da peçonha ou componentes celulares da glândula. Essa análise também pode ser nomeada de transcriptoma da glândula (Jiang *et al.*, 2008).

As bibliotecas de cDNA são obtidas da transcrição *in vitro* dos RNAs mensageiros sintetizados, que podem corresponder às proteínas produzidas pelo tecido e, no caso de escorpiões, pela glândula de peçonha. Os fragmentos de cDNA são ligados a vetores que são inseridos em bactérias, por transformação, para a amplificação do cDNA. As bactérias que tiverem o vetor, logo o inserto de cDNA, serão selecionadas e seus plasmídios purificados e seqüenciados, gerando as seqüências de nucleotídeos (Vasconcelos *et al.*, 2003). Os prováveis produtos destas seqüências nucleotídicas podem ser identificados como componentes da peçonha através da comparação com os dados de caracterização proteômica da peçonha bruta.

A determinação das estruturas primárias das toxinas é limitada pela dificuldade de se obter material suficiente, tendo em vista a pequena quantidade de peçonha obtida em escorpiões. Para superar essa dificuldade pode-se utilizar a metodologia da identificação de ESTs (Jing *et al.*, 2008). Além desta vantagem, a utilização dessa metodologia permite a

determinação dos componentes da peçonha que são mais freqüentemente sintetizados, pois em uma biblioteca de cDNA não-amplificada, a freqüência de ESTs reflete a abundância dos seus respectivos RNAm na glândula (Magalhães *et al.*, 2006). Outra aplicação interessante para a utilização de ESTs é o desenho de novos oligonucleotídeos iniciadores para a identificação de seqüências genômicas com a identificação das regiões de retirada de íntrons (por *splicing*) dos cDNAs (Vasconcelos *et al.*, 2003).

O primeiro transcriptoma realizado em um escorpião não-Buthidae foi realizado por Schwartz e colaboradores (2007) com a glândula de peçonha do escorpião mexicano *Hadrurus gertschi* (Caraboctonidae). Neste trabalho, foram caracterizados 160 ESTs, incluindo 8 novas seqüências de toxinas, peptídeos citolíticos e componentes da glândula que atuam nos processos celulares, o que pode ser importante em estudos comparativos entre a produção de peçonha em animais Buthidae e não-Buthidae (Schwartz *et al.*, 2007).

A utilização de abordagens toxinômicas (genomas, proteomas e transcriptomas de escorpiões) pode levar ao total esclarecimento acerca da diversidade de toxinas, sendo importante ferramenta para o avanço nessa área de pesquisa (Zhijian *et al.*, 2006).

O presente trabalho é o primeiro estudo de busca não dirigida de genes transcritos da glândula de uma espécie de escorpião brasileira não Buthidae.

#### **4. *Opisthacanthus cayaporum***

O gênero *Opisthacanthus* pertence à família Ischnuridae que possui 12 gêneros e 81 espécies. Esta família está distribuída pela África, sudeste da Ásia, Austrália, América do Sul e ilhas adjacentes. O gênero *Opisthacanthus* ocorre no Caribe, Américas Central e do Sul, África e Madagascar.

Foram descritas duas espécies do gênero *Opisthacanthus* no Brasil: *Opisthacanthus cayaporum*, por Vellard (1932) (Figura 3) e *Opisthacanthus borboremai* por Lourenço & Fé, 2003. A espécie *O. cayaporum* é endêmica do Centro Sul do Brasil (sul do Pará, nordeste de Mato Grosso e Tocantins), é encontrada em cupinzeiros do gênero *Armitermes* e seus indivíduos se organizam de forma gregária (Lourenço, 1980; 1981; 2003). O tamanho desses escorpiões varia de 55 a 70 mm, possuem pedipalpos desenvolvidos e

fortes, metassoma pequeno e fraco em comparação com o restante do corpo, apresentando telson pequeno e de coloração amarela (Lourenço, 2003) (Figura 4).



**Figura 4:** Escorpião brasileiro *Opisthacanthus cayaporum*. Esta espécie se organiza de forma gregária e é endêmica do Centro Sul do Brasil.

O gênero a que pertence esta espécie é muito antigo (Lourenço, 1981), sendo também encontrado na África e em Madagascar. É possível que suas toxinas sejam mais ativas em artrópodes (crustáceos e insetos) e pouco tóxicas para os espécimes dos demais filos. Os estudos com escorpiões do gênero *Opisthacanthus* descrevem peptídeos com atividades antimicrobianas, antifúngicas e hemolíticas e foram realizados com a espécie africana *Opisthacanthus madagascariensis* (Dai *et al.*, 2001; 2002). Yamaji e colaboradores, em 2004, descreveram uma nova toxina sexo-específica bloqueadora de canal para potássio, com 41 resíduos de aminoácidos, isolada da peçonha de macho de *O. madagascariensis*, que pode estar relacionada ao comportamento dos macho dessa espécie de inocular peçonha na fêmea no acasalamento.

A caracterização proteômica da peçonha bruta de *O. cayaporum* foi realizada por Schwartz *et al.*, (2008). Nesse trabalho a peçonha foi fracionada por HPLC e foram obtidas 80 frações cromatográficas, das quais foram detectados 94 componentes por espectrometria de massa, compreendidas de 229,2 a 61.144,0 Da. Dezesete peptídeos foram purificados e tiveram suas seqüências N-terminais determinadas por degradação automática de Edman. O efeito farmacológico da peçonha bruta de *O. cayaporum* foi testado em nervos de baratas e ratos. Foi obtido uma ED50 de 1.1mg/ml para nervo de inseto, sendo que em baixas concentrações (0,22mg/mL) a peçonha bruta promove a redução do potencial de ação no nervo de barata. Foi encontrada baixa atividade neurotóxica a essa concentração em nervo



de rato, sendo que 3mg/ml de peçonha bruta causa uma redução de 9% no potencial de ação em nervo de rato. Esses dados refletem a grande diversidade de componentes na peçonha de *O. cayaporum* e que esta é mais ativa para insetos do que para mamíferos, como havia sido proposto por Lourenço, 1981 (Schwartz *et al.*, 2008).

## RELEVÂNCIA CIENTÍFICA

Há a estimativa de que haja em torno de 100 diferentes peptídeos (de 2000 Da a 8000 Da) na peçonha de cada escorpião, o que daria aproximadamente 100.000 peptídeos da peçonha de 1.500 espécies distintas de escorpiões (Tan *et al.*, 2006). Esses dados contrastam com o número aproximado de 400 peptídeos descritos de escorpiões, o que representa apenas 0,4%. A maioria dessas toxinas foi isolada de escorpiões pertencentes à família Buthidae, devido a várias espécies desta família serem responsáveis pelos casos de escorpionismo em humanos.

Há poucos estudos toxicológicos com espécies não-Buthidae, contudo há dados que demonstram que vários peptídeos isolados dessas espécies possuem estrutura primária única e atividades biológicas como, por exemplo, antimicrobianos. Além disso, o estudo da composição química da peçonha de espécies não-Buthidae é muito importante para que se possa esclarecer a origem e evolução das toxinas escorpiônicas (Miyashita *et al.*, 2007).

Este trabalho é a primeira caracterização molecular dos componentes transcritos da glândula de peçonha de um escorpião não-Buthidae brasileiro. Além disso, por meio dos dados obtidos com a biblioteca de cDNA e caracterização proteômica de *O.cayaporum*, foi caracterizado o primeiro peptídeo antimicrobiano não-hemolítico desta espécie, nomeado Cayaporina.

## **OBJETIVOS**

### **Objetivo Geral:**

- ✓ Isolar e caracterizar clones de cDNA que codificam para peptídeos da peçonha do escorpião *O. cayaporum*.

### **Objetivos Específicos:**

- ✓ Construir uma biblioteca de cDNA da glândula de peçonha de *O. cayaporum*;
- ✓ Identificar e caracterizar as ESTs;
- ✓ Clonagem de genes que codificam para componentes da peçonha de *O. cayaporum*;
- ✓ Complementar os resultados obtidos por caracterização proteômica da peçonha bruta de *O. cayaporum*, visando a futura purificação e caracterização biológica de peptídeos da peçonha.

## MATERIAL E MÉTODOS

### 1. Soluções e Meios de Cultivo

Meio Luria Bertani (LB):

- ✓ 10g/L Triptona/Bacto peptona
- ✓ 5 g/L Extrato de levedura
- ✓ 5 g/L NaCl

pH ajustado para 7.0.

Cloranfenicol: 0,03mg/ml;

Solução A: 0.12% ácido trifluoroacético – (TFA) em água MilliQ;

Solução B: 0.10% TFA em acetonitrila- (ACN);

Brometo de Etídio: 0,1µg/ml.

### 2. Coleta e manutenção dos exemplares de *Opisthacanthus cayaporum*

Em torno de 3 exemplares de *Opisthacanthus cayaporum* foram coletados em Palmas, Tocantins, sob licença emitida pelo IBAMA (nº. 048/2007-CGFAU). Esses animais foram mantidos em terrários adequados no Laboratório de Toxinologia, da Universidade de Brasília, sob ciclo dia/noite, com água *ad libitum* e sendo alimentados periodicamente com baratas.

### 3. Construção da Biblioteca de cDNA

#### 3.1 Extração do RNA total

Os escorpiões tiveram a peçonha extraída 5 dias antes da retirada dos télsons. A biblioteca de cDNA foi construída a partir do RNA total extraído por meio do kit *SV RNA total Isolation System* (Promega-USA) da glândula de peçonha de 2 animais adultos (1 macho e 1 fêmea). Os télsons foram cortados e macerados na presença de 1ml do tampão de lise do kit. Após a homogeneização acrescentou-se 350µL do tampão de diluição do kit, inverteu-se por 3 vezes, e o material foi incubado a 70°C por 3 minutos. O produto foi centrifugado por 10 minutos a 10000 rpm.

Acrescentou-se 200µL de etanol a 95% ao lisado, homogeneizou-se com pipetas livres de RNase, e centrifugou-se por 1 minuto a 10000 rpm na coluna de purificação

própria do kit (*Spin Column Assembly*). Após esse procedimento, foram acrescentados 600µL da solução de lavagem (*RNA wash solution*) à coluna de purificação e centrifugou-se a 10000 rpm por 1 minuto.

O produto foi incubado por 15 minutos a 20-25°C com 50µL da solução de DNase presente no kit (40µL de *Yellow Core Buffer*, 5µL 0,09M MnCl<sub>2</sub> e 5µL da enzima DNase). Após esse período foram acrescentados 200µL da solução *Stop DNase* e centrifugado por 1 minuto. Adicionamos 600µL da solução de lavagem de RNA com etanol e centrifugou-se por mais 1 minuto. Acrescentamos 250µL desta solução de lavagem e centrifugamos por 2 minutos nas mesmas condições. Foram acrescentados 100µL de água *nuclease-free*. O RNA total purificado foi guardado a -70°C. Não foi realizada quantificação deste RNA.

### 3.2 Banco de cDNA

Para a construção da biblioteca de cDNA foi utilizado o Kit Creator™ SMART™ cDNA Library Construction (Clontech/UK). O kit SMART foi desenvolvido para a construção de bibliotecas a partir de pequenas quantidades de RNA e possui dois protocolos: LD-PCR e *Primer extension*, utilizamos o protocolo LD-PCD pois utilizamos RNA total.

Para a síntese da primeira fita de cDNA, ao RNA total obtido foram acrescentados 1µL do oligonucleotídeo SMART IV, 1µL *CDS III/3' PCR Primer* e 2µL de água deionizada, formando um volume total de 5µL. Foi dado um breve *spin* nessa mistura e esta foi incubada por 2 min a 72°C, e colocou-se o tubo no gelo por 2 min. O tubo foi homogenizado brevemente e acrescentou-se: 2µL de 5X First-Strand Buffer, 1µL DTT (20 mM), 1µL de dNTP Mix (1mM) e 1µL da transcriptase reversa MMLV, formando um volume final de 10µL. A mistura foi gentilmente homogenizada, e incubada por 1 hora a 42°C, e após esse período o tubo foi colocado no gelo. Após esse procedimento tem-se a fita simples de cDNA com sua extremidade 5' completa e com uma seqüência complementar para o oligonucleotídeo SMART IV, que foi o adaptador para a reação de LD-PCR. Essa característica permitiu que somente os cDNAs com esse adaptador fossem amplificados.

A reação de LD-PCR para a síntese da segunda fita foi realizada com os parâmetros: 95°C por 1 min (desnaturação), 26 ciclos por 10 min, 95°C por 15 segundos e 68°C por 6 min (anelamento). Os reagentes utilizados foram (totalizando um volume final de 100µL):

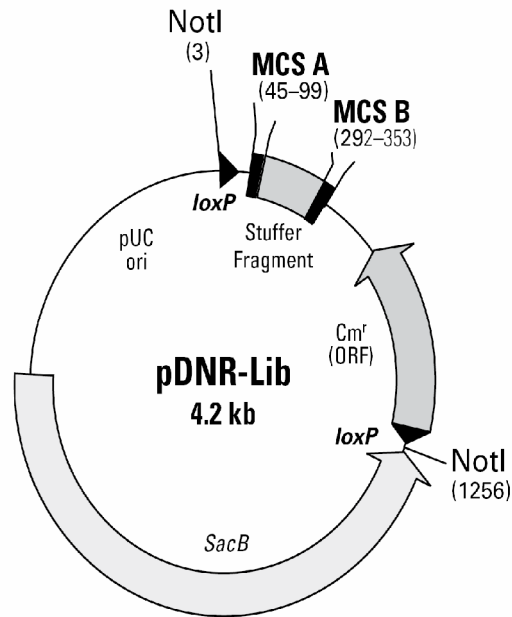
- ✓ 2µL do cDNA fita simples;
- ✓ 80µL de água deionizada;
- ✓ 10µL 10X *Advantage 2 PCR Buffer*;
- ✓ 2µL 50X dNTP Mix
- ✓ 2 µL 5' PCR Primer
- ✓ 2 µL CDS III/3' PCR Primer
- ✓ 2 µL 50X *Advantage 2 Polymerase Mix*

O produto da PCR (5µL) foi analisado em gel de agarose a 1% corado com brometo de etídio (0,1µg/ml).

Para a inativação da atividade da DNA polimerase, o cDNA dupla fita foi tratado com 2µL da proteinase K (20µg/µL) por 20 min a 45°C. Após esse período acrescentou-se 50µL de água deionizada e 100µL de fenol:clorofórmio:álcool isoamílico (25:24:1). O tubo foi invertido gentilmente (de 1 a 2 min). A mistura foi centrifugada por 5 min a 14.000 rpm. A fase aquosa (sobrenadante) foi transferida para um tubo estéril. Acrescentou-se 100µL de clorofórmio: álcool isoamílico (24:1) ao tubo que foi invertido gentilmente por 1 min. Centrifugou-se novamente e o sobrenadante foi transferido para outro tubo. Foi adicionado 10µL de acetato de sódio, 3M e 1.3µL de glicogênio (20µg/µL) e 260µL de etanol a 95% a temperatura ambiente. A mistura foi centrifugada por 20 min a 14000 rpm. O sobrenadante foi descartado e ao pellet foi acrescentado 100µL de etanol 80%, centrifugou-se novamente. O pellet foi deixado secando (aproximadamente por 10 min) a temperatura ambiente. O pellet foi ressuspendido com 79µL de água deionizada.

Para a digestão com a enzima de restrição SfiI, acrescentou-se a um tubo: 79µL do cDNA, 10µL Sfi Buffer 10X, 10µL SfiI Enzyme e 1µL BSA 100X, formando um volume total de 100µL. Misturou-se bem e incubou-se por 2 horas a 50°C. Adicionou-se 2µL de xileno cianol a 1%, mistura-se bem. Esta mistura foi fracionada em coluna CHROMA SPIN-400 (Clontech/UK). Os produtos da digestão foram verificados em gel de agarose a 1% corados com brometo de etídio, com marcador de bases 1Kb plus DNA ladder (Invitrogen). Às frações selecionadas foram acrescidos 1/10 vol. de Acetato de Sódio (3M,

pH 4.8), 1.3µL de glicogênio (20 mg/ml) e 2.5 vol. de etanol 95% gelado, incubou-se a guardado a 20°C durante a noite. O tubo foi centrifugado por 20 min a 14000 rpm e o sobrenadante foi descartado. O pellet foi deixado secando por aproximadamente 10 min. O pellet foi ressuspendido por 7µL de água deionizada e misturado gentilmente. O cDNA estava pronto para ser ligado ao vetor.



#### MCS A

loxP 45  
 TTA TCA GTC GAC GGT ACC GGA CAT ATG CCC GGG AAT TCG GCC ATT ACG GCC TGC AGG ATC C Stuffer Fragment...  
 Sall NdeI SmaI EcoRI SfiI PstI BamHI

#### MCS B

292  
 Stuffer Fragment... GG ATC CGG CCG CCT CGG CCC TCG AGA AGC TTT CTA GAC CAT TCG TTT GGC GCG CGG GCC CAG TAG GTA AGT GAA  
 BamHI SfiI XhoI HindIII XbaI Bsp120I  
 STOPs

**Figura5. Mapa do Vetor pDNR-LIB (Clontech/UK).** Figura retirada do Creator™ SMART™ cDNA Library Construction Kit User Manual ([www.clontech.com](http://www.clontech.com)). Este vetor possui gene de resistência a cloranfenicol (Cm<sup>r</sup>). “SacB” indica o gene para sucrase provindo do *Bacillus subtilis*. Esses genes são utilizados para a seleção dos clones. “Stuffer fragment” indica a posição em que o inserto de cDNA será ligado.

### 3.3 Ligação ao vetor pDNR-LIB (Clontech/UK)

O vetor utilizado para a construção da biblioteca de cDNA foi o pDNR-LIB (4.2 Kb) contido no kit Creator™ SMART™ cDNA Library Construction (Clontech/UK) (Figura 5).

Acrescentou-se em um tubo 1.5µL do cDNA, 1µL do vetor pDNR-LIB (0.1 µg/µL), 0,5µL de ATP (10mM), 0,5µL de T4 DNA Ligase e 1µL de água deionizada, formando um volume final de 5µL. A amostra foi homogenizada suavemente e incubada durante a noite a 16°C. Adicionou-se 95µL de água deionizada e 1,5µL de glicogênio misturou-se bem. Foi acrescentado 280µL de etanol 95% gelado. Incubou-se por 1 hora a -70°C. Após este período, o tubo foi centrifugado por 20 min a 15000 rpm. O etanol foi cuidadosamente retirado e o pellet foi ressuspensionado em 5µL de água deionizada. O produto de ligação, cDNA+vetor, foi guardado a -20°C para utilização para a transformação bacteriana.

### 4. Transformação bacteriana por choque térmico e purificação dos plasmídios

Foram realizadas 4 transformações bacterianas. O título calculado refere-se à primeira transformação e a porcentagem de recombinação obtida foi dada através de todos os resultados analisados em géis de agarose dessas 4 transformações. Na primeira transformação foi utilizado 2µL do sistema de ligação (cDNA+vetor pDNR-Lib) e nas outras 3 utilizou-se 1µL, para a transformação por choque térmico de células quimiocompetentes *Escherichia coli* linhagem DH5α. As células *E. coli* quimiocompetentes (OD 0,55 nm) foram transformadas por choque térmico em gelo/42°C. As bactérias transformadas foram plaqueadas em 3 placas de petri LB ágar+ cloranfenicol, em diluições de 1:5, 1:10 e 1:20 e incubadas em estufa a 37°C por 12 horas (protocolos de preparação de células quimiocompetentes e transformação bacteriana conforme Azevedo *et al.*, 2003).

Após a incubação em placas de petri com meio LB/ágar com cloranfenicol (0,03mg/ml), o título do banco foi calculado utilizando-se a seguinte fórmula: Título do banco de cDNA = nº de colônias x fator de diluição x 10<sup>3</sup> / µL de células usadas

As colônias crescidas em meio LB/ágar+cloranfenicol foram usadas como templates em PCR utilizando os primers T7 e M13. As condições da PCR foram, para uma reação de volume final de 10µL:



94°C por 4 min  
30 ciclos  $\left[ \begin{array}{l} 94^\circ\text{C por 40 segundos} \\ 55^\circ\text{C por 40 segundos} \\ 72^\circ\text{C por 40 segundos} \end{array} \right]$   
72°C por 7 minutos  
4°C por 30 minutos

Os resultados da PCR foram analisados em gel de agarose a 0,8% com brometo de etídio a 0,1µg/ml, as bandas obtidas foram comparadas às do marcador de tamanho de pares de bases, 1Kb plus DNA ladder (Invitrogen), para estimar o tamanho dos insertos (Azevedo *et al.*, 2003). As colônias que apresentaram insertos  $\geq 400$ pb foram recuperadas em meio líquido LB com cloranfenicol (0,03mg/ml).

Essas colônias recuperadas foram submetidas à lise alcalina conforme Azevedo *et al.*, (2003) e os plasmídios obtidos foram ressuspensos em água tetradestilada com RNase (10µg/ml) e mantidos a -20°C. Foram feitos géis de agarose a 0,8% com brometo de etídio para a quantificação e verificação da qualidade desses plasmídios para que fossem enviados ao sequenciamento automático de DNA.

## 5. Sequenciamento automático de DNA e análises de bioinformática

Para o sequenciamento automático de DNA utilizou-se o “primer” T7 (5 pM). Algumas seqüências incompletas foram reseqüenciadas utilizando-se o “primer” M13 reverso. O seqüenciador automático de DNA utilizado foi *Apply Biosystem* Modelo 3100 (Foster city/CA) e foram seguidas as instruções do fabricante.

As seqüências nucleotídicas obtidas foram submetidas ao programa CrossMatch que identifica e marca as regiões correspondentes ao DNA de *E. coli* e de vetor, por comparação a um arquivo de seqüência de vetores, e ao programa Phred. O *software* Phred classifica as seqüências de acordo com a qualidade apresentada. Utilizamos pontuação Phred  $\geq 20$  para atribuir à base nucleotídica a classificação “boa qualidade” (essa pontuação significa que há 1 chance em 100 para que a base esteja incorreta) (<http://www.unb.br/ib/cel/biomol>). As ESTs foram submetidas ao programa CAP3 que baseia-se na similaridade, freqüência e na qualidade das bases nucleotídicas para a

formação de seqüências *contig* (nomeadas *Contigs*), e as que não constituem *contigs*, permanecem singletos, ou seqüências únicas (*singlets*) (Huang & Madan, 1999). A seqüência *contig* é utilizada para descrever um conjunto de seqüências relacionadas, mas não idênticas, que foram agrupadas analisando o nucleotídeo mais freqüente da mesma posição em todas as seqüências que formam um *contig* (Vasconcelos *et al.*, 2003). Todas essas análises em bioinformática foram realizadas simultaneamente, com os parâmetros dados pelos programas (<http://www.unb.br/ib/cel/biomol/>).

As seqüências nucleotídicas com alta qualidade foram submetidas à predição da seqüência de aminoácidos, utilizando-se a ferramenta Bio Edit, e à identificação das regiões ORFs, que marcam o início da tradução, com o códon ATG (Metionina), e o término e desacoplamento do complexo ribossômico com os códons: TAA, TAG ou TGA (*stop codons*). Foram identificadas também as seqüências que possuem a poliadenilação na região C-terminal, bem como a seqüência sinal, AATAAA para a poliadenilação.

Todas as seqüências (*contigs* e singletos) foram submetidas aos programas de busca de similaridade em banco de dados não redundantes de proteínas, por meio do algoritmo BLAST, a fim de atribuir prováveis funções. Foram utilizados os subtipos de Blast: blastn, que analisa a seqüência de nucleotídeo comparando-a com seqüências de nucleotídeos depositadas e blastx onde a seqüência é traduzida nas seis possíveis fases de leitura e a análise é feita contra banco de proteínas (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) com valores de expectância (*e-value*) ajustados para  $<10^{-5}$  para a identificação das possíveis funções desses ESTs (Altschul *et al.*, 1990). As regiões de prováveis peptídeos sinais foram preditas com o programa SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) e os alinhamentos múltiplos foram obtidos com o programa CLUSTALW2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).

As massas moleculares teóricas monoisotópicas foram obtidas pelo programa Compute pI/Mw tool ([http://www.expasy.ch/tools/pi\\_tool.html](http://www.expasy.ch/tools/pi_tool.html)) e foram comparadas com as massas moleculares obtidas da caracterização proteômica da peçonha de *O. cayaporum* (Schwartz *et al.*, 2008). Todas as seqüências nucleotídicas obtidas neste trabalho serão depositadas no GeneBank.

## 6. Caracterização da Cayaporina

### **6.1 Obtenção da Cayaporina**

A peçonha bruta de 15 escorpiões adultos machos e fêmeas foi coletada por estimulação elétrica, solubilizada em água Milli-Q e submetida à centrifugação a 10000 rpm por 10 min. Foi retirada uma alíquota do sobrenadante para quantificação por espectrofotometria a 280 nm, e o restante liofilizado e armazenado a -20 °C. A peçonha bruta foi fracionada em HPLC em coluna de fase-reversa C<sub>18</sub> (Vydac, Hisperia, CA). Os componentes foram eluídos através de um gradiente linear de 0% a 60% da solução B, por 60 min, em um fluxo de 1ml/min. A detecção foi feita por absorvância a 230 nm. As frações eluídas foram coletadas manualmente e secadas a vácuo. A fração de interesse foi posteriormente recromatografada em coluna C<sub>18</sub> analítica com um gradiente linear de 0-60% de Solução B em 40 minutos, com um fluxo de 1 ml/min.

### **6.2 Determinação da seqüência de aminoácidos por espectrometria de massa**

As frações provenientes da cromatografia de fase reversa tiveram suas massas moleculares determinadas por um espectrômetro de massa MALDI-TOF/TOF *UltraFlex II* (Bruker Daltonics, Germany). Uma alíquota de cada fração foi misturada na proporção de 1:3 com uma solução de matriz (ácido  $\alpha$ -ciano-4-hidroxicinâmico) e depositada em uma placa do tipo Anchorchip (600 mm) que foi deixada secar a temperatura ambiente, e depois submetida ao espectrômetro de massa, utilizando-se a calibração externa sob modo de operação linear e refletido (positivo). A seqüência de aminoácidos foi obtida por sequenciamento *de novo* e confirmada por sequenciamento automático de Edman (*Beckman LF 3000 Protein Sequencer-Palo Alto, CA, USA*).

### **6.3 Atividade Biológica**

O peptídeo purificado por HPLC foi submetido ao teste antimicrobiano contra bactérias Gram-negativa (*Escherichia coli* -ATCC 25922) e Gram-positiva (*Staphylococcus aureus*-ATCC 29213) para a determinação do MIC. As bactérias foram incubadas em meio LB por 6 horas em 37°C sob agitação contínua, e foram feitas leituras de absorvância a 600nm (OD<sub>600</sub>). O cultivo foi monitorado até atingir a leitura de 0.5 a 600nm.

As culturas de bactérias foram incubadas em meio LB a 37°C em microplacas com 96 poços na presença de diferentes concentrações do peptídeo, em uma diluição seriada,

com o volume final de 50 $\mu$ l. Após 12hs a OD<sub>600</sub> foi mensurada a fim de determinar a porcentagem de inibição causada pelo peptídeo (Moerman *et al.*, 2002). Os controles negativos e positivos foram poços somente com meio de cultura e poços com cultivo de bactérias, respectivamente.

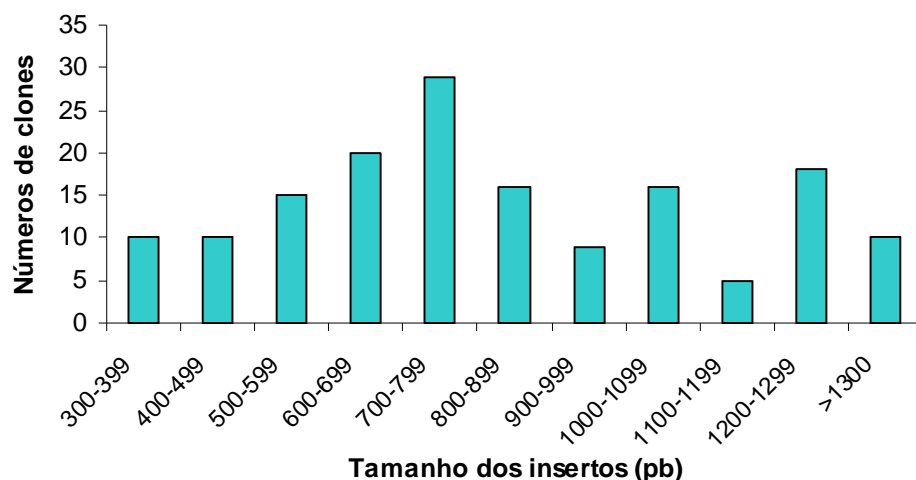
A atividade hemolítica do peptídeo foi determinada utilizando uma suspensão a 3% de hemácias humanas em solução de NaCl a 0.9%. Essa suspensão foi incubada por 1 hora a temperatura ambiente em placa com concentrações decrescentes de peptídeo. Após o período de incubação as placas foram analisadas quanto à presença ou ausência de hemólise, utilizando como controle positivo água destilada e como controle negativo a suspensão de hemácias em salina.

## RESULTADOS

### 1. Biblioteca de cDNA

Foram realizadas quatro transformações por choque térmico em células quimiocompetentes de *E. coli* DH5a com os cDNAs obtidos na construção da biblioteca. O título médio do banco foi de 2233,33 cfu/mL (unidades formadoras de colônias por mL) com uma porcentagem de clones recombinantes de 66,3%. Destas quatro transformações, 243 colônias foram recuperadas, e obtivemos 159 colônias que apresentaram insertos de tamanho  $\geq 300$ pb.

A distribuição por tamanho dos 159 clones obtidos pode ser observada na Figura 6. O tamanho mais freqüente de inserto encontrado foi de 700 pb (19% dos clones).



**Figura 6: Frequência de clones em relação ao tamanho do inserto em pares de bases (pb).** No eixo “x” tem-se o tamanho do inserto, em um intervalo de 100pb. No eixo “y” temos o número de clones, em um intervalo de 5. Obtivemos um total de 159 clones com inserto, sendo que o tamanho mais freqüente encontrado foi de 700 pb.

Dos 159 clones obtidos, os plasmídios de 150 deles foram purificados e seqüenciados. Os nove clones restantes não foram seqüenciados por apresentarem insertos menores que 400pb.

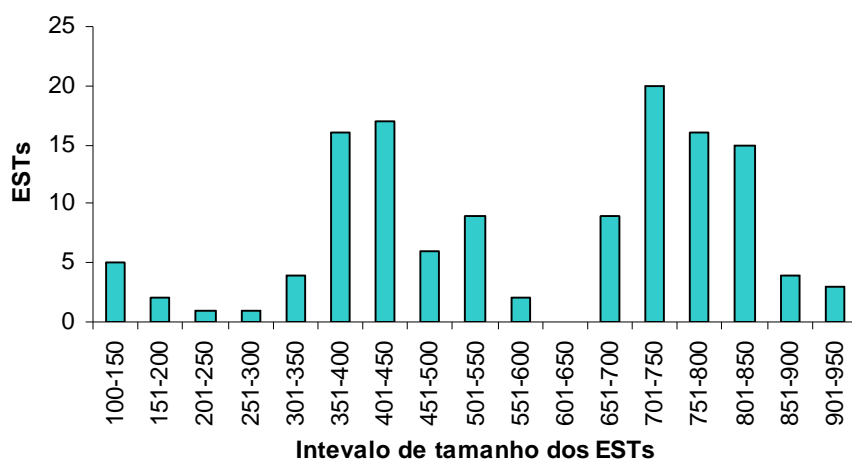
### 2. Seqüências dos cDNAs

Dos 150 plasmídios sequenciados, foram obtidas 149 seqüências que foram submetidas ao programa Phred que resultou em 130 ESTs que atenderam aos parâmetros utilizados.

As 130 ESTs obtidas possuem em média 535 nucleotídeos (variando entre 139 e 902 nucleotídeos), com pode ser visto na Figura 7. Observa-se no gráfico que os clones apresentaram uma distribuição bimodal, sendo os intervalos mais freqüentes de 351 a 450 pb e de 701 a 850pb.

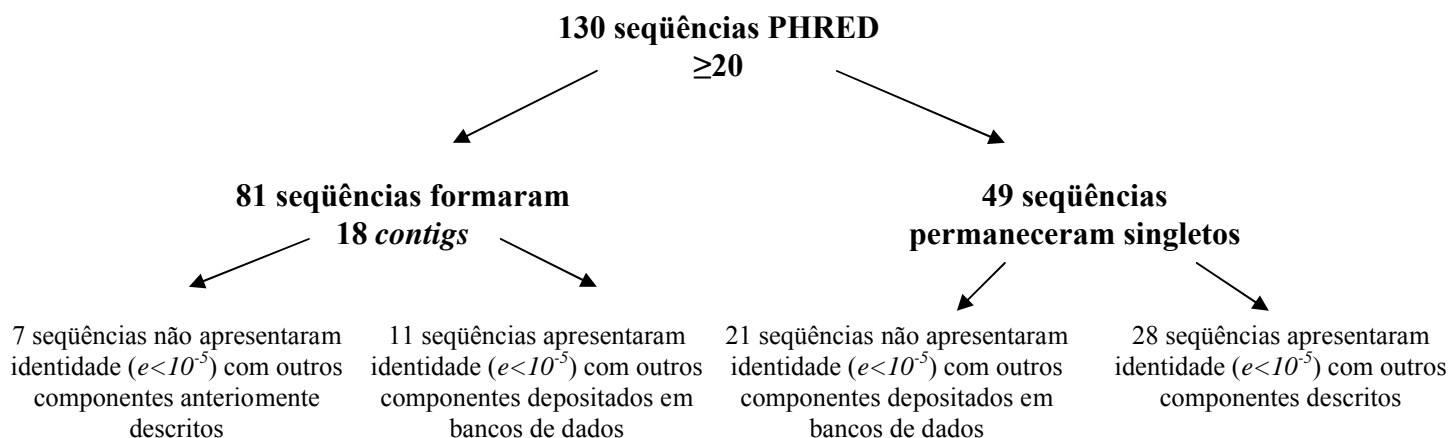
Dos 130 clones, 49 são singletos e 81 compõem 18 *contigs*. Em resumo, após as análises de bioinformática, temos 67 seqüências nucleotídicas distintas com alta qualidade, conforme esquematizado na figura 8.

Das 67 seqüências nucleotídicas distintas, 22% não apresentam regiões ORFs (9 singletos e 6 *contigs*) e 43% (3 *contigs* e 26 singletos) não possuem poli-A na região 3'UTR.



**Figura 7: Intervalo de distribuição do tamanho dos 130 ESTs obtidos.** No eixo “x” tem-se o tamanho dos ESTs, em número de nucleotídeos, em um intervalo de 50pb. No eixo “y” temos o número de ESTs, em um intervalo de 5. Obtivemos um total de 130 ESTs e observa-se que estes estão mais distribuídos entre os intervalos 351-450 e 701-850 nucleotídeos.

Das 67 seqüências nucleotídicas distintas, em 28 seqüências não foi possível identificar uma provável função (estão apresentadas em Anexo). Os resultados de similaridade e prováveis funções das 39 seqüências restantes estão mostradas na Tabela 2.



**Figura 8: Esquema mostrando o agrupamento das 130 ESTs da biblioteca de cDNA.** Resultados da análise de similariedade com o banco de dados, por busca “Blastn” e “Blastx”

**Tabela 2: Identificação das 39 seqüências nucleotídicas e prováveis funções**

Nomenclatura	Clones	Identidade e Expectância	Descrição e Análise (Blastn ou Blastx)
<b>Seqüências com identidade a componentes de peçonhas</b>			
Ocy10	<i>contig 1</i>	65%, $5^{-5}$	Insetotoxina LIT1- <i>Liocheles australiasiae</i> (Blastx)
Ocy11	<i>contig 2</i>	42%, $5^{-6}$	TXLP1 – <i>Buthus martensii</i> (Blastx)
Ocy1 (NDBP 5.5)	<i>contig 3</i>	77%, $3^{-9}$	IsCt antimicrobiano – <i>Opisthacanthus madagascariensis</i> (Blastx)
OcyKtx1	<i>contig 4</i>	64%, $6^{-3}$	Omtx3 - <i>Opisthacanthus madagascariensis</i> (Blastx)
Ocy2 (NDBP 5.6)	<i>contig 5</i>	71%, $1^{-8}$	IsCt antimicrobiano – <i>Opisthacanthus madagascariensis</i> (Blastx)
Cayaporina (NDBP 3.7)	<i>contig 13</i>	72%, $2^{-18}$	Opistoporin3 – <i>Opisthophthalmus carinatus</i> (Blastx)
OcyKtx2 ( $\alpha$ KTx 6.15)	<i>contig 17</i>	47%, $4^{-5}$	alfa-Ktx 6.11- <i>Opisthacanthus madagascariensis</i> (Blastx)
Ocy4 (NDBP 4.3)	12	65%, $2^{-5}$	Pandinin2 – <i>Pandinus imperator</i> (Blastx)
Ocy40	26	30%, $8^{-3}$	AAL15172 Venom peptide 2 - <i>Buthus martensii</i> (Blastx)
Ocy9	77	85%, $4^{-6}$	Omtx3 - <i>Opisthacanthus madagascariensis</i> (Blastx)
<b>Ocy7*</b>	<b>90</b>	52%, $3^{-11}$	Precursor de Opiscorpina 2- <i>Opisthophthalmus carinatus</i> (Blastx)
Ocy6	210	36%, $5^{-9}$	Is6- <i>Ixodes scapularis</i> (Blastx)
Ocy41	223	33%, $4^{-4}$	AAL15172 Venom peptide 2 - <i>Buthus martensii</i> (Blastx)

<b>Ocy fosfolipase1</b>	<b>112</b>	34%, 9 <sup>-12</sup>	Fosfolipase C – <i>Strongylocentrotus purpuratus</i> (Blastx)
<b>Ocy fosfolipase2</b>	<b>122</b>	63%, 5 <sup>-21</sup>	Fosfolipase – <i>Anuroctonus phaiodactylus</i> (Blastx)
<b>ESTs com identidade a componentes do processamento celular na glândula de peçonha</b>			
<b>Ocy3</b>	<b>contig 6</b>	59%, 1 <sup>-22</sup>	Proteína Ribossomal – <i>Scorpius maurus palmatus</i> (Blastx)
<b>Ocy5</b>	<b>contig 14</b>	44%, 4 <sup>-6</sup>	Domínio de proteína THUMP – <i>Mus musculus</i> (Blastn)
<b>Ocy18</b>	<b>contig 15</b>	73%, 5 <sup>-11</sup>	Gene mitocondrial – <i>Biphyllus lunatus voucher</i> (Blastn)
<b>Ocy20</b>	<b>01</b>	71%, 1 <sup>-61</sup>	Ferritina- <i>Ornithodoros parkeri</i> (Blastx)
<b>Ocy21</b>	<b>07</b>	63%, 4 <sup>-80</sup>	Hemocianina, subunidade F- <i>Nephila inaurata madagascariensis</i> (Blastx)
<b>Ocy22</b>	<b>11</b>	56%, 3 <sup>-31</sup>	Proteína ativada por Cálcio – <i>Apis melífera</i> (Blastx)
<b>Ocy23</b>	<b>22</b>	55%, 1 <sup>-19</sup>	Proteína similar à prefoldin – <i>Canis familiaris</i> (Blastn)
<b>Ocy25</b>	<b>68</b>	70%, 4 <sup>-49</sup>	Proteína do tipo tsunagi- <i>Apis melífera</i> (Blastx)
<b>Ocy26</b>	<b>73</b>	44%, 1 <sup>-25</sup>	NADH, desidrogenase- <i>Centruroides limpidus</i> (Blastn)
<b>Ocy27</b>	<b>91</b>	31%, 2 <sup>-12</sup>	Elastase - <i>Macaca mullata</i> (Blastx)
<b>Ocy28</b>	<b>93</b>	96%, 3 <sup>-114</sup>	Actina- <i>Rhipicephalus microplus</i> (Blastx)
<b>Ocy30</b>	<b>130</b>	38%, 3 <sup>-32</sup>	Sulfotransferase- <i>Ixodes scapularis</i> (Blastx)
<b>Ocy31</b>	<b>142</b>	82%, 1 <sup>-34</sup>	Gene ribossomal- <i>Pandinus imperator</i> (Blastn)
<b>Ocy32</b>	<b>172</b>	28%, 3 <sup>-14</sup>	Serina protease- <i>Mus musculus</i> (Blastx)
<b>Ocy33</b>	<b>195</b>	35%, 6 <sup>-24</sup>	Inibidor de serpin peptidase - <i>Danio rerio</i> (Blastx)
<b>Ocy34</b>	<b>228</b>	61%, 2 <sup>-6</sup>	Domínio de nuclease, Staphylococcal- <i>Danio rerio</i> (Blastx)
<b>Ocy35</b>	<b>232</b>	61%, 8 <sup>-11</sup>	Calmodulina - <i>Culex pipiens quinquefasciatus</i> (Blastx)
<b>ESTs que apresentam identidade com genes pertencentes a genomas seqüenciados</b>			
<b>Ocy19</b>	<b>contig 16</b>	37%, 3 <sup>-10</sup>	Alérgeno – <i>Lycosa singoriensis</i> (Blastn)
<b>Ocy24</b>	<b>31</b>	37%, 1 <sup>-21</sup>	Possível alérgeno – <i>Aedes aegypti</i> (Blastn)
<b>Ocy29</b>	<b>129</b>	40%, 9 <sup>-4</sup>	Provável proteína salivar- <i>Ixodes scapularis</i> (Blastn)
<b>Ocy36</b>	<b>19</b>	58%, 9 <sup>-18</sup>	Proteína hipotética – <i>Apis melífera</i> (Blastn)
<b>Ocy37</b>	<b>102</b>	54%, 6 <sup>-25</sup>	Proteína hipotética – <i>Apis melífera</i> (Blastn)
<b>Ocy38</b>	<b>108</b>	62%, 3 <sup>-10</sup>	Proteína hipotética – <i>Apis melífera</i> (Blastn)
<b>Ocy39</b>	<b>192</b>	62%, 3 <sup>-10</sup>	Proteína hipotética – <i>Danio rerio</i> (Blastn)

\* Em vermelho tem-se os ESTs que não apresentaram regiões ORFs.

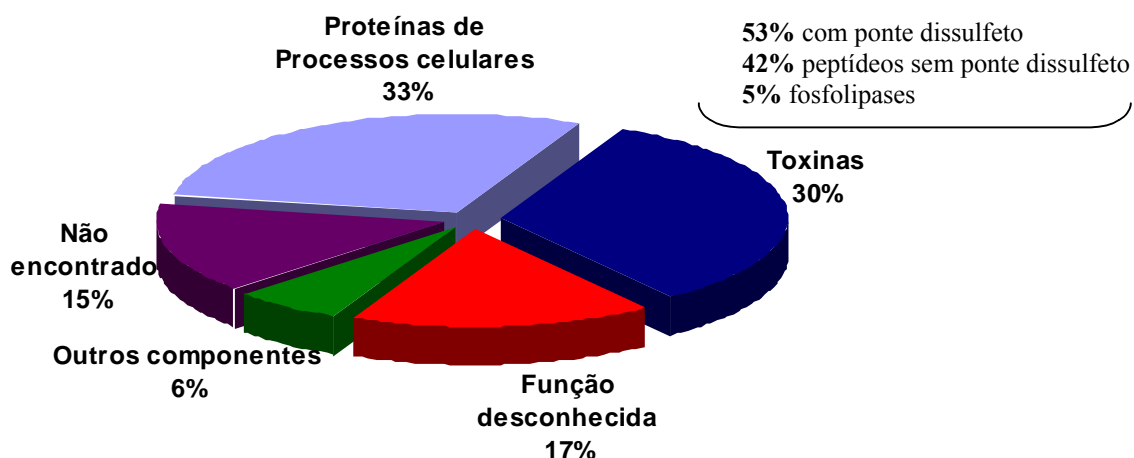
As seqüências de nucleotídeos que apresentaram identidade a componentes do processamento celular e as que apresentaram identidade com genes identificados de genomas estão apresentadas em Anexo.



Os dados encontrados na busca por similaridade demonstram que, de um total de 130 clones, 30% são prováveis toxinas de *O. cayaporum*, sendo 7 *contigs* e 8 singletos, ou seja 15 seqüências distintas (Figura 8). Dentre as prováveis toxinas, 53% correspondem a toxinas com pontes dissulfeto (4 *contigs* e 5 singletos), 42% são prováveis toxinas sem pontes dissulfeto (3 *contigs* e 1 singletos) e 2 singletos que são prováveis fosfolipases, correspondendo a 5%. Em resumo, tem-se 15 prováveis seqüências de toxinas que serão descritas detalhadamente mais adiante.

Do total de ESTs obtidos, 36% correspondem a proteínas relacionadas a processos celulares (16 singletos e 4 *contigs*) (Figura 9). Essas 20 seqüências distintas foram classificadas em 4 categorias, de acordo com a provável função exercida:

- ✓ **Gene mitocondrial:** Ocy3 (que é composto por 24 clones) e Ocy18 (que compreende 3 clones) e Ocy31
- ✓ **Proteínas estruturais:** Ocy5 (com 2 clones), Singletos Ocy28, Ocy25, Ocy23 e Ocy21;
- ✓ **Enzimas:** Singletos Ocy34, Ocy32, Ocy30, Ocy27 e Ocy26;
- ✓ **Metabolismo:** Singletos Ocy20, Ocy35, Ocy33, e Ocy22.



**Figura 9: Distribuição dos 130 ESTs obtidos.** A categoria “Toxinas” é referente às 15 seqüências distintas de prováveis toxinas, “outros componentes” compreendem as seqüências que apresentaram identidade com componentes de genomas. “não encontrado” refere-se às 11 seqüências que não apresentaram nenhuma identidade com proteínas no GenBank. “Função desconhecida” compreende as 17 seqüências que não obtiveram expectância mínima para a atribuição de funções.

Além das 15 prováveis toxinas, serão apresentadas detalhadamente 6 seqüências (cor que estão agrupadas na categoria “Função desconhecida”, e que apresentam provável peptídeo sinal, resíduos de cisteína, sinal de poliadenilação (AATAAA) no respectivo cDNA e poliadenilação na porção 3’, e que poderiam representar novas toxinas.

## Seqüências relacionadas a prováveis Toxinas

### 2.1.1. Seqüências com cisteínas (cys) que formariam pontes dissulfeto

Foram identificados do banco de cDNA de *O. cayaporum*, 9 seqüências que apresentaram identidade com toxinas, depositadas em bancos de dados, que possuem cisteínas potencialmente formadoras de pontes dissulfeto. As seqüências encontradas similares a toxinas com pontes dissulfeto foram: Ocy10, Ocy11, OcyKtx1, OcyKtx2, Ocy40, Ocy41, Ocy9, Ocy6 e Ocy7.

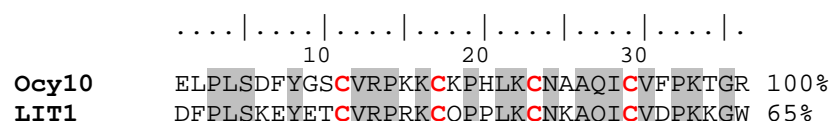
A seqüência nomeada Ocy10 (*contig 1*) representada por dois clones, possui 384 nucleotídeos (Figura 10). A seqüência de aminoácidos predita para Ocy10 possui um peptídeo sinal de 20 resíduos de aminoácidos, seguida de uma seqüência de 55 resíduos de aminoácidos, incluindo 4 resíduos de cisteínas que formariam 2 pontes dissulfeto.

1	ATG AAC TTT GCA ACC AAG ATA GTT ATC CTT CTC CTG GTG GCA GCC	45
1	<i>M N F A T K I V I L L L V A A</i>	15
46	CTG ATC TTG GCC GTT ACA AGT GAA AAG GGG GAC AGT TCT TCT GAC	90
16	<i>L I L A V T S E K G D S S S D</i>	30
91	GAT AAT GAA GCT AAG GAG ACG GAA GGA GAA CTT CCA CTG AGT GAC	135
31	<i>D N E A K E T E G E L P L S D</i>	45
136	TTT TAC GGA TCT TGC GTT CGC CCG AAG AAA TGC AAA CCT CAT CTG	180
46	<b>F Y G S C V R P K K C K P H L</b>	60
181	AAG TGC AAC GCC GCA CAG ATT TGC GTG TTT CCC AAA ACG GGG CGG	225
61	<b>K C N A A Q I C V F P K T G R</b>	75
226	TAG CCGCCTTCCAAATTTTTTTTTTCGCGAATTACGAACCGGCATGTGATTAACCGTAA	
76	stop	
	<u>ATAAAGCTCATATTAATAATTC</u> -polyA-3'	

**Figura 10: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy10.** O peptídeo sinal está marcado em azul, e a provável seqüência do sinal de poliadenilação está em duplo sublinhado. O provável propeptídeo está sublinhado e o peptídeo maduro em negrito.

A seqüência deduzida de aminoácidos de Ocy10 apresentou 65% de identidade (com expectância de 5<sup>-5</sup>) com a insetotoxina nomeada LAIT1 isolada do escorpião *Liocheles australasiae* pertencente à família Ischnuridae (Matsushita *et al.*, 2007).

A partir do resultado do alinhamento da seqüência de aminoácidos predita para Ocy10 com a do peptídeo LAIT1 (Figura 11), pode-se inferir que o peptídeo em *O. cayaporum* seja sintetizado na forma de um pró-peptídeo e que o provável peptídeo maduro seria constituído de 36 resíduos de aminoácidos (Figura 10). Considerando-se a seqüência de 36 resíduos de aminoácidos e as 2 pontes de dissulfeto formadas, a massa teórica monoisotópica calculada para esse peptídeo maduro seria de 4031.07 Da. Essa massa teórica não foi encontrada na caracterização proteômica de *O. cayaporum*.



**Figura 11: Alinhamento da seqüência de aminoácidos de Ocy10 com a insetotoxina LIT1 isolada do escorpião *Liocheles australasiae*** (Matsushita *et al.*, 2007). À direita tem-se a porcentagem de identidade da toxina LIT1 com Ocy10. Marcado em cinza têm-se os resíduos de aminoácidos que são comuns entre as seqüências, e em vermelho os 4 resíduos de cisteína, que formariam 2 pontes dissulfeto.

Obtivemos a seqüência nomeada OcyKtx1 (*contig* 4) representado por 2 clones, e que possui 359 nucleotídeos. A seqüência de aminoácidos predita para OcyKtx1 apresenta um peptídeo sinal com 26 resíduos seguido por uma seqüência de 44 aminoácidos contendo 4 cisteínas que formariam 2 pontes dissulfeto (Figura 12).

Foi encontrado 65% de identidade (expectância de 6<sup>-6</sup>) de OcyKtx1 com a toxina para canal de potássio Omtx2 isolada da peçonha do escorpião africano *Opisthacanthus madagascariensis*, pertencente à família Ischnuridae (Chagot *et al.*, 2005).

Ocy9 (Singleto) é composto por 347 nucleotídeos e apresenta peptídeo sinal predito com 26 resíduos de aminoácidos seguidos por uma seqüência de 39 resíduos, que contém 4 cisteínas podendo formar 2 pontes dissulfeto (Figura 13).

Ocy9 apresentou 85% de identidade com a Omtx1 que é uma toxina para canal de potássio isolada do escorpião africano *Opisthacanthus madagascariensis*. As Om-toxinas (Omtxs 1, 2, 3 e 4) foram testadas em canais para potássio Kv1.1, Kv1.2 de ratos e o canal Kv1.3 de humanos, e diminuíram a amplitude da corrente de K<sup>+</sup> nesses canais (Chagot *et al.*, 2005).

1	ATG GAG TCT TCC AGA AAG TCG TAT GTG CTG ATG TTG TTC CTG GCA	45
1	<b>M E S S R K S Y V L M L F L A</b>	15
46	TTT GTA ATT ATG AAT GTC TGT AGC GTA AGT GGA GAA CCG AAG GAT	90
16	<b>F V I M N V C S V S G</b> <u>E P K D</u>	30
91	GGA GAA ATA GCA GGC TTT GAA ATG GAA GAA GCA CGT TAC GAT GCT	135
31	<u>G E I A G F E M E E A R</u> <b>Y D A</b>	45
136	TGT GTT AAC GCA TGT CTT GAA CAT CAT CCC AAC GTA AGA GAA TGC	180
46	<b>C V N A C L E H H P N V R E C</b>	60
181	GAA GAG GCA TGT AAA AAT CCT GTT CCC CCC TAA ATGGCGAAAGAAGGA	
61	<b>E E A C K N P V P P</b> stop	

GTCAAGGTTTCATTTTCCACGATATGTAAAGACATACCAGAATAAAACACAGTAACATAG

TACTTCA-polyA-3'

**Figura 12: Sequência de nucleotídeos e predição da sequência de aminoácidos de OcyKtx1.** O provável propeptídeo está sublinhado, o peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A provável sequência sinal de poliadenilação está em duplo sublinhado.

1	ATG AAG ACT TCT AAG ATG ATT TGT GCT TTT CTA TTG GTT CTG GTT	45
1	<b>M K T S K M I C A F L L V L V</b>	15
46	GTC GGA ACA TTC AAT GAT ATA AGT GGA GCA TAT GGA GAA TAC GTG	90
16	<b>V G T F N D I S G A</b> <u>Y G E Y V</u>	30
91	GAG GAC CAG CAC AGC TTC AAA ATT GAA AGG CGT TTT CCA CCA TGC	135
31	<u>E D Q H S F K I E R R</u> <b>F P P C</b>	45
136	GTA GAA GTA TGC GTT CAG CAT ACT GGC AAT GTA AAA GAA TGT GAA	180
46	<b>V E V C V Q H T G N V K E C E</b>	60
181	GCA GCA TGC GGT GAA TAA TGAGGTGCGGAAGCTAATTTCTCATGATGCTTT	
61	<b>A A C G E</b> stop	

ATGTAAACGCATAGGTGACCGAGACTTAATAAAGATACTACAAAAATTTCT-polyA-3'

**Figura 13: Sequência de nucleotídeos e predição da sequência de aminoácidos de Ocy9 que apresentou 85% de identidade com a Omtx1 de *Opisthacanthus madagascariensis* (Chagot et al., 2005).** Em azul tem-se o provável peptídeo sinal, em sublinhado o propeptídeo e em negrito o provável peptídeo maduro. A provável sequência sinal do Poli-A está em duplo sublinhado.

A partir do alinhamento feito com a sequência de aminoácidos predita para OcyKtx1 e Ocy9 com as das Om-toxinas e Kappa-toxinas (Figura 14), pode-se inferir que OcyKtx1 seja sintetizado na forma de um pró-peptídeo e que o peptídeo maduro seja formado por 28 resíduos de aminoácidos (Figura 12). Assim considerando os 28 aminoácidos e as 2 pontes dissulfeto formadas, a massa monoisotópica teórica calculada para essa sequência seria de 3134.36 Da. Da mesma forma, infere-se que Ocy9 seja sintetizado na forma de pró-peptídeo e que o peptídeo maduro seja constituído por 24 resíduos (Figura 13).

Considerando-se esses 24 resíduos e as 2 pontes dissulfeto formadas, a massa monoisotópica teórica calculada para Ocy9 seria de 2544.96 Da. Essa massa molecular teórica não foi encontrada na caracterização proteômica de *O. cayaporum*.

Ocy11 é representado por 9 clones (*contig* 2) e composto por 551 nucleotídeos (Figura 15). Foi possível prever a seqüência de aminoácidos desse *contig* que possui 20 resíduos de aminoácidos de peptídeo sinal seguidos por uma seqüência de 78 resíduos de aminoácidos incluindo 8 resíduos de cisteína que poderiam formar 4 pontes dissulfeto.

	..... ..... ..... ..... ..... .....		
		10	20
<b>OcyKtx1</b>	YDA <b>CVNA</b> CL <b>EHHPNVRE</b> CEEA <b>CKNPVPP</b>	100%	45%
<b>Ocy9</b>	FPP <b>CVEV</b> CVQHTGNVKE <b>CEAA</b> CGE----	45%	100%
<b>Omtx1</b>	-DP <b>CYEV</b> CLQQHGNVKE <b>CEEA</b> CKHPVE-	61%	62%
<b>Omtx2</b>	-DP <b>CYEV</b> CLQQHGNVKE <b>CEEA</b> CKHPVEY	59%	62%
<b>Omtx3</b>	NDP <b>CYEV</b> CLQHTGNVKA <b>CEEA</b> CQ-----	62%	69%
<b>Omtx4</b>	-DP <b>CYEV</b> CLQQHGNVKE <b>CEEA</b> CKHP---	52%	62%
<b>KappaTx1</b>	GH <b>AC</b> YRNCWREGNDEET <b>CKERC</b> -----	27%	18%
<b>KappaTx2</b>	GH <b>AC</b> YRNCWREGNDEET <b>CKERC</b> G-----	26%	21%
<b>KappaKTx1.3</b>	GFG <b>CYRS</b> CWKAGHDEET <b>CKKE</b> CS-----	17%	17%

**Figura 14: Alinhamento da seqüência de aminoácidos de OcyKtx1 e Ocy9 com as Toxinas para canal de potássio Omtxs (Chagot *et al.*, 2005) e KappaTxs (Srinivasan *et al.*, 2002).** À primeira coluna direita tem-se a porcentagem de identidade das Omtx e KappaTx com OcyKtx1, e na segunda coluna as identidades com Ocy9. Em vermelho tem-se os 4 resíduos de cisteína, que podem formar 2 pontes dissulfeto.

1	ATG	AAA	ATC	GCC	TGT	ACC	TTG	GTA	CTT	TTC	GTA	ATG	CTT	AGG	TGT	45
1	<b>M</b>	<b>K</b>	<b>I</b>	<b>A</b>	<b>C</b>	<b>T</b>	<b>L</b>	<b>V</b>	<b>L</b>	<b>F</b>	<b>V</b>	<b>M</b>	<b>L</b>	<b>R</b>	<b>C</b>	15
46	TAC	GTA	AAT	GCT	CGT	AAT	ATT	CCA	GGG	ACC	TGC	AGG	ACG	CAC	ACA	90
16	<b>Y</b>	<b>V</b>	<b>N</b>	<b>A</b>	<b>R</b>	<b>N</b>	<b>I</b>	<b>P</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>R</b>	<b>T</b>	<b>H</b>	<b>T</b>	30
91	GGA	ATA	ATT	CTG	CTT	TCC	GGT	GAA	GAG	TGG	AAG	GAT	CCT	AAT	CAC	135
31	<b>G</b>	<b>I</b>	<b>I</b>	<b>L</b>	<b>L</b>	<b>S</b>	<b>G</b>	<b>E</b>	<b>E</b>	<b>W</b>	<b>K</b>	<b>D</b>	<b>P</b>	<b>N</b>	<b>H</b>	45
136	TGT	TCA	ACA	TAC	AGA	TGC	TCA	ATT	TTC	GAT	GGA	GAA	GCA	GAA	CTT	180
46	<b>C</b>	<b>S</b>	<b>T</b>	<b>Y</b>	<b>R</b>	<b>C</b>	<b>S</b>	<b>I</b>	<b>F</b>	<b>D</b>	<b>G</b>	<b>E</b>	<b>A</b>	<b>E</b>	<b>L</b>	60
181	GAG	GGT	GTG	ACA	TGT	GCA	GCC	TAC	CAC	GTA	CCA	CCC	CAT	TGT	AGA	225
61	<b>E</b>	<b>G</b>	<b>V</b>	<b>T</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>Y</b>	<b>H</b>	<b>V</b>	<b>P</b>	<b>P</b>	<b>H</b>	<b>C</b>	<b>R</b>	75
226	CTC	GTT	TCA	GCA	GCC	AAT	GAA	TTA	TAT	CCT	CAG	TGT	TGT	CCA	ACT	270
76	<b>L</b>	<b>V</b>	<b>S</b>	<b>A</b>	<b>A</b>	<b>N</b>	<b>E</b>	<b>L</b>	<b>Y</b>	<b>P</b>	<b>Q</b>	<b>C</b>	<b>C</b>	<b>P</b>	<b>T</b>	90
271	GTC	ATT	TGT	AGT	GAT	AAG	TCG	CGC	TGA	TCT	GTTA	ACC	ATTC	CAG	CTAAGT	
91	<b>V</b>	<b>I</b>	<b>C</b>	<b>S</b>	<b>D</b>	<b>K</b>	<b>S</b>	<b>R</b>	stop							

CATTGGAAGTTGCAATAAAGCCAATACATTAATCTGAT-polyA-3'

**Figura 15: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy11.** O provável peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A provável seqüência sinal de poliadenilação está em duplo sublinhado.



OcyKtx2 apresentou 39% de identidade (com uma expectância de 4<sup>-4</sup>) com a toxina IsTX isolada do escorpião africano *O.madagascariensis* (Yamaji *et al.*, 2004). A toxina IsTX pertence à subfamília 6 de  $\alpha$ KTxs (a toxina IsTX é nomeada  $\alpha$ KTx 6.11) que é caracterizada por possuir estrutura tridimensional com uma  $\alpha$ -hélice e duas folhas- $\beta$  estabilizadas por quatro pontes dissulfeto, e foi a primeira toxina descrita como macho-específica em escorpiões (Yamaji *et al.*, 2004). OcyKtx2 também apresentou identidade de 33% (expectância de 4<sup>-4</sup>) com a toxina OcKtx3 (pertencente à subfamília 6 das  $\alpha$ KTxs e nomeada  $\alpha$ KTx 6.8) isolada do escorpião *Opisthophthalmus carinatus* (Scorpionidae) (Zhu *et al.*, 2004).

A Figura 18 apresenta do alinhamento feito com OcyKtx2 e as 14 toxinas pertencentes à subfamília 6 das  $\alpha$ KTxs. A partir desses dados podemos inferir que OcyKtx2 possui um peptídeo maduro composto por 43 resíduos de aminoácidos (Figura 17) e considerando também as 4 pontes dissulfeto formadas, a massa molecular monoisotópica calculada para OcyKtx2 seria de 5084.25 Da. Portanto essa seqüência foi classificada como OcyKtx2- $\alpha$ KTx 6.15. Essa massa molecular não foi encontrada na caracterização proteômica de *O. cayaporum*.

1	ATG AAT TTA AAA TTG GCC TTG GTG TTG CTG CTT ACA GTC ATT AAT	45
1	<b>M N L K L A L V L L L T V I N</b>	15
46	GTA GGA ATG CTG CCT GGT GCG ACG TCA AAT GGG AAT ATC AAA ACA	90
16	<b>V G M L P G A T S N G N I K T</b>	30
91	GAT ATT AAA TGC TAC CGT AAC AGC CAT TGC AAC TTT CAT TGC GAG	135
31	<b>D I K C Y R N S H C N F H C E</b>	45
136	AAG AGT TAC TAT TGC CAA GGA AGT AAA TGT GTG CGT AAA CGA TGC	180
46	<b>K S Y Y C Q G S K C V R K R C</b>	60
181	AAC TGC TAC AAC TGT CCA TTG TGA TAAAGTCTTCCACAGTTTAGAGTGAAA	
61	<b>N C Y N C P L stop</b>	
	AGACTTCTGGAAAATGGATCTTATGGAAAATCCAACCTTCTGTCACT <u>AAATAA</u> AGCTCGCC	
	TGAGCACAAACATGCATTAC-polyA-3'	

**Figura 17: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de OcyKtx2 que apresentou identidade com a toxina para canal de potássio IsTX.** O provável peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A seqüência sinal de poliadenilação está em duplo sublinhado.





1	ATG TGG AAG GCT GTA ATA TTC TGT CTC GTC ATT GTA GCA TCT GTC	45
1	<b>M W K A V I F C L V I V A S V</b>	15
46	TGT GGA GAA GAA GAC AAT AAG GAA GAC AGG AGA AAA ATA AAC CTT	90
16	<b>C G E E D N K E D R R K I N L</b>	30
91	TTA TTT TCG GAC GAT GGC AGG AGG ATT TTG AAG TGT TGG TTC ACC	135
31	<b>L F S D D G R R I L K C W F T</b>	45
136	TAT GTT TTA TAT TAC TGG CCC AGT AGT GAT AAG GCT ACC AAA CAG	180
46	<b>Y V L Y Y W P S S D K A T K Q</b>	60
181	GCA GCT CTG GAA AAG ATG TTC GAG TGT ATG AAA AAG GGA TTG GAG	225
61	<b>A A L E K M F E C M K K G L E</b>	75
226	GCT TCG TCT GCC TAA ACTATTACACAGCCGTACAAGAAAATATAGAAAGGAACA	
76	<b>A S S A</b> stop	
	TTTAAAGACGACAAAGTTCAAGTTAATGTAAGTCAAGAACTGATTTGGAATAATAAAGT	
	<u>TTATTATTGCNGA-polyA-3'</u>	

**Figura 19: Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy40.** O provável peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A seqüência sinal de poliadenilação está em duplo sublinhado.

Ocy40 e Ocy41 apresentaram respectivamente 30% e 33% identidade e 4<sup>-4</sup> de expectância com o precursor nomeado *Venom peptide2* (número no Gene Bank AAL15172) caracterizada por clonagem da glândula de peçonha do escorpião asiático *Buthus martensii*, e que não teve sua atividade biológica testada (Figura 21).

A partir dos resultados obtidos com o alinhamento de Ocy40 e Ocy41 com o peptídeo *Venom peptide2*, pode-se inferir que o Ocy40 possui um peptídeo maduro composto por 62 resíduos de aminoácidos, e o Ocy41 um peptídeo maduro de 61 aminoácidos (Figura 21). Considerando-se os 62 aminoácidos do Ocy40 e a ponte dissulfeto formada, a provável massa monoisotópica calculada desse singlet seria de 7349.64 Da. Da mesma forma podemos inferir que a massa monoisotópica do Ocy41 seria de 7185.52 Da. Essas massas moleculares não foram encontradas na caracterização proteômica de *O. cayaporum*.

```

1   ATG TGG AAG GCT GTA ATA TTC TGT CTC GTC ATT GTA GCA TCT GTC 45
1   M W K A V I F C L V I V A S V 15

46  TGT GGA GAA GAA GAC AGT AAG GAA GAC AGG AGA ATA ATA AAC CTT 90
16  C G E E D S K E D R R I I N L 30

91  TTA TTT TCG GAG GAT GGT AGA AAG ATT TTG GAG TGT TGG TTC ACC 135
31  L F S E D G R K I L E C W F T 45

136 TAT GTT TTA TAT TAC TGG CCC AGT AGT GAC AAG GCT ACC AAA TAT 180
46  Y V L Y Y W P S S D K A T K Y 60

181 GCA GCT CTG GAA AAG ATG TTC GAG TGT ATG AAA AAG GGA TTG GAG 225
61  A A L E K M F E C M K K G L E 75

226 GCT TCG TCT GGC TAA ACTATTACACGTCCGTACAAGAAAATATAGAAAGG ACA
76  A S S G stop

TTTAAAGACGACAAAGTTCAAGTTAATGTAAGTCAAGAACTGACTTGAATAAAATAAAGT

TTATTATTGC-polyA-3'

```

**Figura 20: Sequência de nucleotídeos e predição da sequência de aminoácidos do Ocy41.** O provável peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A sequência sinal de poliadenilação está em duplo sublinhado.

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          10          20          30          40          50
Ocy40 -----EDNKEDRRKINLLFSDDGRRILKCWFTYVLYYWPSSDKATKQ
Ocy41 -----EDSKEDRRIINLLFSEdGRKILECWFTYVLYYWPSSDKATKY
AAL15172 DAWADAWADAWEKDEEEESRFFFNFIFSAEGRKILQCLATFGFSYSLTTDMAAKL

.....|.....|.....|.....
          60          70
Ocy40   AALEKMFECMKKGLEASSA 100% 88%
Ocy41   AALEKMFECMKKGLEASSG 88% 100%
AAL15172 TAQEKLCNCTANAIAKST- 30% 33%

```

**Figura 21: Comparação da sequência de aminoácidos de Ocy10 e Ocy41 com precursor nomeado *Venom peptide 2* (número no GeneBank: AAL15172) identificado no escorpião *Buthus martensii*.** A porcentagem de identidade para o Ocy40 está na primeira coluna à direita, e para o Ocy41 na segunda coluna à direita. Em vermelho tem-se os 2 resíduos de cisteína que formariam 1 ponte dissulfeto e em cinza os resíduos de aminoácidos consensos entre as seqüências.

Caracterizamos também a seqüência nomeada Ocy6 que é composta por 556 nucleotídeos. Através da predição da seqüência de aminoácidos, inferimos que esta seqüência possui um peptídeo sinal de 24 resíduos de aminoácidos seguido por uma seqüência de 94 resíduos, incluindo 10 cisteínas que formariam 5 pontes dissulfeto (Figura 22). Ocy6 apresentou 32% de identidade ( $e$ -value:  $6^{-7}$ ), com a seqüência nomeada Is6 caracterizada a partir da biblioteca de cDNA do carrapato *Ixodes scapularis* que não teve sua atividade biológica caracterizada (Packila & Guilfoile, 2002).

A partir dos resultados obtidos com o alinhamento do provável peptídeo maduro de Ocy6 com o peptídeo maduro de Is6, podemos inferir que em *O. cayaporum*, este peptídeo é composto por 94 resíduos de aminoácidos, com 10 cisteínas que formariam 5 pontes dissulfeto (Figura 23). Consideram-se estes 94 resíduos e as pontes dissulfeto formadas, a massa monoisotópica calculada para Ocy6 seria de 10287 Da. Essa massa molecular não foi encontrada na caracterização proteômica de *O. cayaporum*.

```

1   ATG AAT GCG ATA GCT GTC TAC GGT TGG TGT ATT GTA CTG TTG GTG 45
1   M N A I A V Y G W C I V L L V 15

46  ATA AGT GTA CCA TTT GTA CAC GCC ACT CCA CCT CCT TAC ACG GAA 90
16  I S V P F V H A T P P P Y T E 30

91  CAA AGG CAC GTT GAT CAC CTC GGA CAG ATA ATA CCT TAT CAA AGC 135
31  Q R H V D H L G Q I I P Y Q S 45

136 GAT TTG CCA GAT GGA GCA GTG TGC CTC AAC AGC AGA GAA TGC TCC 180
46  D L P D G A V C L N S R E C S 60

181 AGT GGA TGC TGT AGG GTT GCA TTG GGT TAT CAG CGC TGC TCT CCG 225
61  S G C C R V A L G Y Q R C S P 75

226 AGA GCA CAG CAA GGA GAA AAA TGC AGA GAT AGC CAG GTT AAA GGC 270
76  R A Q Q G E K C R D S Q V K G 90

271 GGA GTG TAT GTG TTT TAC TGC CCC TGC GAG AGT GGA GCG GGT GAT 315
91  G V Y V F Y C P C E S G A G D 105

316 TGC CAG GAA GGA GTT TGC TCG TAT GAA GGC GTG TTC TAC TAA AGA
106 C Q E G V C S Y E G V F Y stop

AGGCAAGACAAGGCACTGTTGACCGCTTCTATCCAATGAAGTCTCTTACTGTGCCTACA

AGAATGCATTTACTTACTATCAGTATCCTAAAGTTAACGATGTAATCCTTTTAAATAAAGT

ACTTCCGAAATGAAT-polyA-3'

```

**Figura 22: Sequência de nucleotídeos e predição da sequência de aminoácidos de Ocy6.** O provável peptídeo maduro e o peptídeo sinal estão assinalados respectivamente em negrito e em azul. A sequência sinal de poliadenilação está em duplo sublinhado.

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
          10          20          30          40          50
Ocy6  PPPYTEQRHVDHLGQIIPYQSDLPDGAVCLNSRECSSGCCRVAL-----GYQRCSSPRAQ
Is6    -----QQVATG----PRGQDLADGTICSTGDECASKCCLKHFTVTGSDGPAQCHVKSD

          60          70          80
Ocy6  QGEKCRDSQVKGGVYVFYCPESGAGDCQEGVCSYEGVFY---- 100%
Is6    LGESCSDDQVKGGSVNHCPCSX--GSCENNICTVENTDEDKDD 32%

```

**Figura 23: Comparação da sequência de aminoácidos de Ocy6 com o peptídeo nomeado Is6 (número no GeneBank: AA085923) isolado do carrapado *Ixodes scapularis* (Packila & Guilfoile, 2002).** A porcentagem de identidade para o Singlet 210 está na primeira coluna à direita, Em vermelho têm-se os 10 resíduos de cisteína que formariam 5 pontes dissulfeto e em cinza os resíduos de aminoácidos consensos entre as seqüências.

Ocy7 é constituído por 409 nucleotídeos, contudo não foi possível determinar pela predição da seqüência de aminoácidos, a região de peptídeo sinal. Esta seqüência apresentou identidade de 52%, expectância de 3<sup>-11</sup> ao precursor da seqüência nomeada Opiscorpina 2 (número no GenBank: AAQ94353) identificada por cDNA da glândula de peçonha do escorpião *Opisthophthalmus carinatus* (Zhu & Tytgat 2004).

```

1   GGG CTG ATT CAC AAG GTG ACC AAG GTA CAG CAG TTA TGT GCA TTC   45
1   G L I H K V T K V Q Q L C A F   15

46  AAC CAA GAT ATG GCT GGA TGG TGC GAG AAG AGT TGC CAG GCT GCG   90
16  N Q D M A G W C E K S C Q A A   30

91  GAA GGT AAG AAT GGA TAC TGC CAC GGT ACC AAA TGC AAA TGT GGA   135
31  E G K N G Y C H G T K C K C G   45

136 AAG CCT TTA AGC TAC AGA AGA AAA TAG TGACCATGACAATATTTCCCTGAC
46  K P L S Y R R K stop

TTGGTAGGGTACATCCGTCCTTCGTTATTTTTTTTTTTGTTTTTCAGCGTTTTCTTGGTAA

TTTTCTTGGATCTATGATCGCGAACAAGGATGCGCTTTAATGAAGCTTATTAGTAATATT

AAAATATCCAAAATTGTTGTA CTACACAATTTTAATAAGCACCAATAAAGGACTTTTGA

AAAGTGTGCG-polyA-3'

```

**Figura 24: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy7.** Não foi possível determinar a provável peptídeo sinal. Em negrito tem-se o provável peptídeo maduro. O sinal de poliadenilação está em duplo sublinhado.

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....
                      10      20      30      40      50
Ocy7                -----GLIHKVTKVQQLCAFNQDMAGWCEKSCQAAEGKN
Scorpina            GWINEEKIQKKIDERMGN TVLGGMAKAI VHKMAKNEFQCMANMDMLGNCEKHCOTSE--K
Ocyscorpina         GWINEEKIQKKIDEPLXXGFVRGVAXGLI-----
AAQ94353            KWLNEKSIQNKIDEKIGKNFLGGMAKAVVHKLAKNEFMCMANMDPTGSCETHCQKASGEK

.....|.....|.....|...
                      60
Ocy7                GYCHGTKCKCGKPLSYRRK   100%
Scorpina            GYCHGTKCKCGTPLSY---   54%
Ocyscorpina         -----   6%
AAQ94353            GYCHGTKCKGVPLSY---   52%

```

**Figura 25: Comparação da seqüência de aminoácidos de Ocy7 com o precursor nomeado Opiscorpina2** (número no GeneBank: AAQ94353) isolado do escorpião *O. carinatus*, com a Scorpina isolada da peçonha do escorpião *Pandinus imperator* (Conde *et al.*, 2000) e com Ocyscorpina. A porcentagem de identidade para Ocy7 está na primeira coluna à direita. Em vermelho tem-se os 6 resíduos de cisteína que formariam 3 pontes dissulfeto. Em cinza os aminoácidos consensos.

Fazendo-se a predição de aminoácidos a partir do primeiro códon, tem-se uma seqüência de 53 resíduos de aminoácidos, incluindo 6 cisteínas que formariam 3 pontes dissulfeto (Figura 24 e 25). Assim identificamos que a seqüência de Ocy7 não está completa na região N-terminal. Foi feito o alinhamento também com a seqüência de aminoácidos não completa do peptídeo nomeado Ocyescorpina isolado da peçonha de *O. cayaporum* (Schwartz *et al.*, 2008). A massa monoisotópica de Ocy7 não foi calculada.

### 2.1.2 Prováveis Fosfolipases

Foram obtidas duas seqüências (Singletos 112 e 122) que apresentaram identidade (expectâncias menores que  $10^{-5}$ ), com proteínas caracterizadas como fosfolipases e depositadas no GenBank. Estas seqüências foram nomeadas Ocyfosfolipase1 e 2.

A Ocyfosfolipase1 não está completa na porção 5' (Figura 26) e através da seqüência predita de aminoácidos, não foi possível determinar a seqüência de peptídeo sinal, estando essa seqüência incompleta na região N-terminal. (Figura 27). Considerando-se os 260 aminoácidos e as pontes dissulfeto formadas pelos resíduos de cisteína, a massa monoisotópica calculada para Ocyfosfolipase1 seria de 30183.05 Da.

Ocyfosfolipase1 apresentou identidade de 40% e expectância de  $3^{-50}$  à Fosfolipase C (número no GenBanK: ACA05826) caracterizada do genoma do peixe ósseo *Paralichthys olivaceus*.

496	TTA GAA GAT GAG TGT AAG AGT CCT GAT CTG TAG TTT GTT AGG TTT	540
	<b>F V R F</b>	4
541	TTT CAC AGT TTA CAG TCC AGA CCT GAT GCA GAA GCA TTA TTC AAA	585
5	<b>F H S L Q S R P D A E A L F K</b>	19
586	GAG TAT AAT GCC AGT AAT TCC ACT ACT ATG GGT GCT GAA GAG CTT	630
20	<b>E Y N A S N S T T M G A E E L</b>	34
631	ATG AAT TTC TTC ATA AGA GCA AAG ATG ACT AAA GTA ACA TTG GAG	675
35	<b>M N F F I R A K M T K V T L E</b>	49
676	AAT GCA AAA CTC TTA TTG AAA GGT ATG AAC CAG ATA AGA CTA AAC	720
50	<b>N A K L L L K G M N Q I R L N</b>	64
721	TGG AAG GAC ATC TTT CAC CGA AAG GGT TTA TAC AAA TGT TAC TTT	765
65	<b>W K D I F H R K G L Y K C Y F</b>	79
766	CGA AAG ATC ATG CCA TTT TTT CTG TTA ATC CAG TCC TTT GGG TGT	810
80	<b>R K I M P F F L L I Q S F G C</b>	94
811	CAA GAC ATG AAC CAG CCA TTA TGT AAT TAC TTT GTT GCT TCT CCT	855
95	<b>Q D M N Q P L C N Y F V A S P</b>	109
856	CAT AAC ACA TAT TTA ATA CAA GAT CAA CTG TTA GGA GAT AGC AGT	900
110	<b>H N T Y L I Q D Q L L G D S S</b>	124
901	GTG GAA GGA TAT AAT AGA GCA TTA GAA CAA GGA TGT AGA TGT GTG	945
125	<b>V E G Y N R A L E Q G C R C V</b>	139
946	GAA TTG GAT TGT TGG GAT GGT CCA AAT GAT GAA CCA GTG GTA TAT	990
140	<b>E L D C W D G P N D E P V V Y</b>	154
991	CAT GGA TAT ACA TTA ACA TCT AAA GTC CTA TTT AGG GAT GTA CTT	1035
155	<b>H G Y T L T S K V L F R D V L</b>	169
1036	GAT GCA ATC AAA TTG ACA GCA TTT AAA ACC TCA CAA TTT CCT GTG	1080
170	<b>D A I K L T A F K T S Q F P V</b>	184
1081	ATT CTC TCA TTG GAA AAC CAT TGC AGT GTT AAA CAG CAA ATA AGG	1125
185	<b>I L S L E N H C S V K Q Q I R</b>	199
1126	ATG GCA GAT CAT CTT GTG AAT ATT TTA GGT GAA TAC TTA CAT GCA	1170
200	<b>M A D H L V N I L G E Y L H A</b>	214
1171	GAT CCT GTG AAA CCA GAT GAA AGA GAG TTG CCA TCA CCA GCA TCC	1215
215	<b>D P V K P D E R E L P S P A S</b>	229
1216	TTA GCA CGT AAG ATT TTG ATA AAG AAC AAA AAC TTG CTG ATA ATT	1260
230	<b>L A R K I L I K N K N L L I I</b>	244
1261	CTC AAA CCA AAC TTG ATG ATA GCG ATA GCG AAG AAG ATT TAC CAG	1305
245	<b>L K P N L M I A I A K K I Y Q</b>	259
1306	AAG CAG TGA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAACATG	
260	<b>K Q stop</b>	

**Figura 26: Sequência de nucleotídeos e predição da sequência de aminoácidos da Ocyfosfolipase1.** Não foi possível determinar o provável peptídeo sinal. Em negrito o provável peptídeo maduro com 261 resíduos de aminoácidos.

Para Ocyfosfolipase2 não foi possível determinar a seqüência de peptídeo sinal através da seqüência predita de aminoácidos (Figura 28). Esta seqüência apresentou identidade de 46% e expectância de  $7^{-24}$  à fosfolipase nomeada *Pha1- Phaiodactylipin* (número no GenBank: ABQ43128) caracterizada através do cDNA do escorpião mexicano *Anuroctonus phaiodactylus*, Chactidae (Valdez-Cruz *et al.*, 2007). Ocyfosfolipase2 não está completa na região N-terminal, e considerando apenas os 118 aminoácidos, e as pontes dissulfeto formadas pelos resíduos de cisteína a massa monoisotópica calculada para Ocyfosfolipase2 seria de 13121.9 Da (Figura 29).

```

Ocyfosf.1      -----FVRFHSLQSRPD---AEAL
ACA05826      ...KIIANLNNLSTQQKTEHWIISCLKADKDKDDKLSQSEMKNFLRSINIEVEDDYAEML

Ocyfosf.1      FKEYNASNS-----TT--MGAEELMNF FIRAKMTK
ACA05826      FQQCDKSKSGYLAGVEIQHFYDLLTHREEINVIYGYAKTTGFMSPGNLVEFLMKEQREK

Ocyfosf.1      VTLENAKLLKGMN-QIRLNWKDIFHRKGLYKCYFRKIMPFFLLIQSFGCQDMNQPLCNY
ACA05826      ATVALAHEIIEKYEPDEHVKAKRLLSKDGFMYMHPEAMILNLDHKEVCQDMNQPLNHY

Ocyfosf.1      FVASPHNTYLIQDQLLGDSSVEGYNRALEQGCRCVELDCWDGPNDEPVYHGYTLTSKVL
ACA05826      FISSSH-TYLMEDQLKGPSSTEAYVRALLKGCRCVELDCWDGSDDEPVIYHGYTLTSKIL

Ocyfosf.1      FRDVLDAIKLTAFKTSQFPVILSLENHCSVKQQIRMADHLVNLG EYLHADPVKPD-RE
ACA05826      FKDTIKAIKEYAFKTS DYPVILSLEN-CSVEQQKVM AHMSSILGSALVTTPLGDDMPTN

Ocyfosf.1      LPSPASLARKILIK-----NKN
ACA05826      FPSPEELKGFLLK GKRLNKLEATFASEAAAADDDVTEEDENDEDEQREEEQRK KKKK

Ocyfosf.1      LLIIILKPNLMIAIAKKIYQKQ----- 100%
ACA05826      LKLAKELSDMVIYCKSVHFN GFEDAVKNLSFYEMSSFKEGKAVKLAEDSANAY... 40%

```

**Figura 27: Alinhamento da seqüência de aminoácidos Ocyfosfolipase1 com a Fosfolipase C (número no GeneBank: ACA05826).** A porcentagem de identidade para Ocyfosfolipase1 está na primeira coluna à direita. Em vermelho tem-se os 4 resíduos de cisteína que formariam 2 pontes dissulfeto. O peptídeo maduro da fosfolipase C (ACA05826) é composto por 596 resíduos, sendo que o alinhamento foi feito apenas com a região que apresenta similaridade com Ocyfosfolipase1.

```

1   CCA GGA AAT ATT GCG ACA GAC GAA TCC GAC CTT GGT ACT CTC GAA 45
1   P  G  N  I  A  T  D  E  S  D  L  G  T  L  E  15

46  GCC GAT AGG TGT TGC CGC ACT CAT GAC CAC TGC GAT AAT ATA CCA 90
16  A  D  R  C  C  R  T  H  D  H  C  D  N  I  P  30

91  GCT GGT GAA ACG AAG TAT GGA ATA ACT AAC GAT TCC TTT TTC ACC 135
31  A  G  E  T  K  Y  G  I  T  N  D  S  F  F  T  45

136 AAG TTG AAC TGC AAA TGT GAA GAA TCT TTC GAT CGC TGT TTG AAC 180
46  K  L  N  C  K  C  E  E  S  F  D  R  C  L  N  60

181 CAA GCT TAT GAG AAT GCA GAG GAT TCA GCC AAG GCA TCG ACC AAA 225
61  Q  A  Y  E  N  A  E  D  S  A  K  A  S  T  K  75

226 AAC TTG CGA AAT CTT TAC TTT TCC CTC TAC TCC CCA GAC TGC TAC 270
76  N  L  R  N  L  Y  F  S  L  Y  S  P  D  C  Y  90

271 GAT TTG CAA TGT AAC GCG AGG AAG TCT GGA AGG AAT GCA GGA TGT 315
91  D  L  Q  C  N  A  R  K  S  G  R  N  A  G  C  105

316 GAA AAT GGA ATT GCC ACT TGG AAG GAA GGT TAC AAT GTC TAG ATA 360
106 E  N  G  I  A  T  W  K  E  G  Y  N  V  stop

ATCTTGAAGCTTCCACATAGAAAGGAATTAATAATTAATCTTTAATTTTATTGAGTTGG
TGACATAACTTGTTTCAGTGTAAAGGCTTCAGCAAGCTTCCTTCAATGGGAGTATTCT
CTTTATTCTTGTGTGGCTCTCGATAATATTAATCTTGTTTACTTATGCCGGATGTGAGAA
AAGATCAATATTGTCTCCTTTGCTATAGATGAAATAAAATTCAAATA-polyA-3'

```

**Figura 28: Sequência de nucleotídeos e predição da sequência de aminoácidos de Ocyfosfolipase2.** Não foi possível determinar o provável peptídeo sinal. O sinal de poliadenilação está em duplo sublinhado.

```

Ocyfosf.2 -----PGNIATDESDLGTLEADRCCRTHDHCDNIPAGETKYGITNDSFFTKLNC
ABQ43128  FLIVPDTKWCGNNNIAANYSDLGPLEADKCCRDHDHCDHIASGETKYGLENKGLFTIILNC

Ocyfosf.2  KCEESFDRCLNQAYEN--AEGSAKASTKNLRNLYFSLYSPDCYDLQCNARRSGRNAGCED
ABQ43128  DCDEAFDHCLNEISNNFTMDIRQKGAENVWSYFQWYNANCYRLYCKDEKSARDEACTN

Ocyfosf.2  GIATWKEGYNV- 100%
ABQ43128  QYAVVKKNFTVQ 46%

```

**Figura 29: Alinhamento da sequência de aminoácidos de Ocyfosfolipase2 com a sequência da *Phaiodactylipin* (número no GenBank: ABQ43128).** Esta sequência foi isolada do escorpião *Anuroctonus phaiodactylus* (Valdez-Cruz *et al.*, 2007). A porcentagem de identidade para o Singlet 122 está na primeira coluna à direita. Em vermelho tem-se os 9 resíduos de cisteína que formariam 4 pontes dissulfeto.

### 2.1.3 Sequências sem cisteínas

Obtivemos 4 diferentes seqüências nucleotídicas que apresentaram identidade (e expectância menor que  $10^{-5}$ ) com toxinas sem pontes dissulfeto, caracterizadas de outras espécies de escorpiões: Ocy1, Ocy2, Ocy4 e a Cayaporina.



Ocy1 (*contig*, representado por 9 clones) possui 420 nucleotídeos. A seqüência de aminoácidos predita para Ocy1 possui um peptídeo sinal de 23 resíduos de aminoácidos, seguida de uma seqüência de 45 resíduos de aminoácidos (Figura 30).

Ocy2 (*contig*, representado por 3 clones) possui 345 nucleotídeos conforme Figura 25. A partir da seqüência de aminoácidos predita, infere-se que este possui um peptídeo sinal de 25 aminoácidos e segue uma seqüência de 45 resíduos (Figura 31).

Ocy1 apresentou 75% de identidade (com expectância de  $3^{-9}$ ) com o peptídeo antimicrobiano nomeado IsCT1 isolado do escorpião africano *Opisthacanthus madagascariensis* (Dai *et al.*, 2002). Ocy2 também apresentou identidade com o peptídeo antimicrobiano IsCT1 (identidade de 91% e expectância de  $1^{-8}$ ). IsCT1 e IsCT2 são peptídeos curtos com 13 resíduos de aminoácidos que juntamente com os peptídeos antimicrobianos BmKn1 e BmKn2 (identificados a partir do cDNA do escorpião *Buthus martensii* por Zeng e colaboradores em 2004), fazem parte da subfamília 5 dos NDBPs (Zeng *et al.*, 2005).

1	ATG AAG GCC CAG CTC TGC ATC CTC CTT ATT GCA CTA GTT CTG TTT	45
1	<u>M K A Q L C I L L I A L V L F</u>	15
46	CAA ACG TTT TCC CAA TCG GAT GCG ATT CTC TCG GCT ATT TGG AGT	90
16	<u>Q T F S Q S D A I L S A I W S</u>	30
91	GGA ATT AAG AGC CTC TTC GGA AGA AGA GGA TTG AAT GAC CTA GAT	135
31	<u>G I K S L F G R R G L N D L D</u>	45
136	GAC CTC GAT GAG CTG TTC GAT GGA GAA ATC TCA CAG GCC GAC GTC	180
46	<u>D L D E L F D G E I S Q A D V</u>	60
181	GAC TTT CTG AAC GAA CTA ATG CGA TAG TTTCAATGTAATCATAATGGACAT	
61	<u>D F L N E L M R stop</u>	
TCTACGTTATTCTCTGTTTTTCATTTTCGCGATATCCTGACACAATGTAATATTTCTACAT		
TAAGAAATCAATAAAATATTTTTCTGCAG-polyA-3'		

**Figura 30: Seqüência de nucleotídeos e predição da seqüência de aminoácidos Ocy1.** Esta seqüência apresentou 66% de identidade com o peptídeo antimicrobiano IsCT de *Opisthacanthus madagascariensis*. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de Poli-A está em duplo sublinhado. O provável propeptídeo está sublinhado, e o peptídeo maduro em negrito.

A partir dos resultados do alinhamento das seqüências de aminoácidos preditas para Ocy1 e 2 com os peptídeos IsCT1 e IsCT2 e com os outros componentes da subfamília 5 dos NDBPs (Figura 32), podemos inferir que os peptídeos em *O. cayaporum* sejam

provavelmente sintetizados na forma de um pró-peptídeo e que os prováveis peptídeos maduros sejam constituídos de 13 resíduos de aminoácidos (Figura 30 e 31).

1	ATG	AAG	ACT	CAG	TTT	GCC	ATC	CTT	ATG	ATT	GCT	GTA	GTC	CTC	ATG	45
1	<u>M</u>	<u>K</u>	<u>T</u>	<u>Q</u>	<u>F</u>	<u>A</u>	<u>I</u>	<u>L</u>	<u>M</u>	<u>I</u>	<u>A</u>	<u>V</u>	<u>V</u>	<u>L</u>	<u>M</u>	15
46	CAA	ATG	CTT	GTC	CAA	ACT	GAA	GGT	GGA	ATC	TTG	GGA	AAA	ATT	TGG	90
16	<u>Q</u>	<u>M</u>	<u>L</u>	<u>V</u>	<u>Q</u>	<u>T</u>	<u>E</u>	<u>G</u>	<u>G</u>	<u>I</u>	<u>L</u>	<u>G</u>	<u>K</u>	<u>I</u>	<u>W</u>	30
91	GAA	GGA	GTC	AAG	AGT	CTC	ATT	GGA	AAG	AGA	GGA	CTG	AAA	AAG	CTG	135
31	<u>E</u>	<u>G</u>	<u>V</u>	<u>K</u>	<u>S</u>	<u>L</u>	<u>I</u>	<u>G</u>	<u>K</u>	<u>R</u>	<u>G</u>	<u>L</u>	<u>K</u>	<u>K</u>	<u>L</u>	45
136	GAT	CAA	CTA	GAT	GAC	ACA	TTC	GAT	TCC	GAT	TTG	TCC	GAT	GCA	GAT	180
46	<u>D</u>	<u>Q</u>	<u>L</u>	<u>D</u>	<u>D</u>	<u>T</u>	<u>F</u>	<u>D</u>	<u>S</u>	<u>D</u>	<u>L</u>	<u>S</u>	<u>D</u>	<u>A</u>	<u>D</u>	60
181	GTT	AAG	CTC	CTC	AGG	GAA	ATG	TTC	AAA	TAA	GATGTACAGAATTAATTCA					
61	<u>V</u>	<u>K</u>	<u>L</u>	<u>L</u>	<u>R</u>	<u>E</u>	<u>M</u>	<u>F</u>	<u>K</u>	stop						

AAGTACATATAGCACTATTTCGAGG-polyA-3'

**Figura 31: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy2.** Esta seqüência apresentou 91% de identidade com o peptídeo antimicrobiano IsCT de *Opisthacanthus madagascariensis*. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de Poli-A está em duplo sublinhado. O provável propeptídeo está sublinhado, e o peptídeo maduro em negrito.

Considerando-se a seqüência de 13 resíduos de aminoácidos, a massa teórica monoisotópica calculada para Ocy1 seria de 1433.83 Da e para Ocy2 de 1454.89 Da. Segundo a classificação dos NDBPs proposta por Zeng *et al.*, 2005 as seqüências Ocy1 e Ocy2 pertencem à subfamília 5, sendo nomeados respectivamente de Ocy1 (NDBP 5.5) e Ocy2 (NDBP 5.6).

		..... ..... .....		
			10	
5.5	<b>Ocy1</b>	ILSAIWSGIKSLF	100%	66%
5.6	<b>Ocy2</b>	ILGKIWEGVKSLI	66%	100%
5.2	<b>IsCT1</b>	ILGKIWEGIKSLF	75%	91%
5.3	<b>IsCT2</b>	IFGAIWNGIKSLF	75%	66%
5.1	<b>BmKn1</b>	FIGAVAGLLSKIF	25%	25%
5.4	<b>BmKn2</b>	FIGAIARLLSKIF	25%	25%

**Figura 32: Alinhamento da seqüência de aminoácidos de Ocy1 e Ocy2 com os peptídeos antimicrobianos pertencentes à subfamília 5 dos NDBPs.** Têm-se na primeira coluna à direita a porcentagem de identidade do *contig* 5 com as toxinas antimicrobianas, e na segunda coluna a porcentagem de identidade do *contig* 3 com os outros peptídeos. Na primeira coluna à esquerda tem-se a classificação sistemática das toxinas dentro da subfamília 5 dos NDBPs (Zeng *et al.*, 2005). Os antimicrobianos IsCT1 e 2 foram isolados do escorpião *O. madagascariensis* (Dai *et al.*, 2001 e 2002) e os peptídeos BmKn1 e BmKn2 do escorpião *Buthus martensii* (Zeng *et al.*, 2001).

Ocy4 (Singleto) é composto por 490 nucleotídeos e através da predição da seqüência de aminoácidos, infere-se que este é composto por um peptídeo sinal de 22 resíduos de aminoácidos seguido por uma seqüência de 53 resíduos de aminoácidos (Figura 33).

Ocy4 apresentou identidade de 69% com o peptídeo antimicrobiano nomeado Pandinina2, isolado da peçonha do escorpião *Pandinus imperator*, família Scorpionidae (Corzo *et al.*, 2001). A Pandinina2 e o peptídeo antimicrobiano BmKb1, isolado do escorpião *Buthus martensii* (Zeng *et al.*, 2004), pertencem à família 4 dos NDPBs (Figura 34).

A partir do alinhamento feito com Ocy4 e os componentes da subfamília 4 dos NDBPs, podemos inferir que este peptídeo em *O. cayaporum* é sintetizado na forma de um pró-peptídeo de 26 resíduos e que o peptídeo maduro possui 27 aminoácidos (Figura 33). Desse modo, considerando-se esses 27 resíduos, a provável massa molecular de Ocy4 seria de 2946.72 Da. Ocy4 foi classificado como NDBP 4.3 (Zeng *et al.*, 2005). Essa massa molecular não foi encontrada na caracterização proteômica de *O. cayaporum*.

1	ATG CAG TAC AAA ACA TTT TTG GTG ATA TCT CTG GCC TAC CTC TTA	45
1	<u>M Q Y K T F L V I S L A Y L L</u>	15
46	GTG GCT GAC GAA GCA GCG GCA TTT TGG GCA ACC TTA GCC AAA GGT	90
16	<u>V A D E A A A F W A T L A K G</u>	30
91	GCA TTG AAG CTG ATA CCA ACA ATA GCT AAC GCG TTT TCT TCC AAA	135
31	<b>A L K L I P T I A N A F S S K</b>	45
136	TCG AAG AAG AGG AGG GAA ATC AAT AAT GTC TTT GAG CCA TAC CAT	180
46	<b>S K K R R E I N N V F E P Y H</b>	60
181	GAG AAC CTA GAT TTA GAA TTA GAA CGC TTC CTA TCA CAA CTT CAA	225
61	<u>E N L D L E L E R F L S Q L Q</u>	75
226	TAA CCTTGGTGTATTTCAGAAAGAATTGGATCGTCTTATATGTAAAAAAAAAAAAA	
76	stop	
	CAGTAACCTA <u>AATAAA</u> AGTATTTAAATTTGC-polyA-3'	

**Figura 33: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy4, que apresentou identidade com o peptídeo antimicrobiano Pandinin2.** O provável propeptídeo está sublinhado, o peptídeo maduro em negrito, provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal para Poli-A está em duplo sublinhado.

	..... ..... ..... ..... ..... ..	
	10 20	
4.3 <b>Ocy4</b>	FWATLAKGALKLIPTIANAFSSKSKKR	100%
4.1 <b>Pandinin2</b>	FWGALAKGALKLIPSL---FSSFSSKSD	69%
4.2 <b>BmKb1</b>	FLFSLIPSAISGLISA---FK-----	22%

**Figura 34: Alinhamento da seqüência de aminoácidos de Ocy4 com os peptídeos antimicrobianos pertencentes à subfamília 4 dos NDBPs.** Na primeira coluna à esquerda tem-se a classificação do peptídeo segundo Zeng e colaboradores, 2005. Marcado em cinza têm-se os resíduos de aminoácidos que são comuns entre as seqüências. Na primeira coluna à direita estão as porcentagens de identidade do Singlet 12 com os outros peptídeos. O antimicrobiano Pandinina2 foi isolado de *Pandinus imperator* (Corzo *et al.*, 2001) e o peptídeo BmKb1 de *Buthus martensii* (Zeng *et al.*, 2004).

Adicionalmente, caracterizamos a seqüência nomeada Cayaporina (*contig* representado por 3 clones) que possui 411 nucleotídeos. A partir da seqüência de aminoácidos que foi predita para essa seqüência, infere-se que este seja composto por um peptídeo sinal de 22 resíduos de aminoácidos seguido por uma seqüência de 59 resíduos (Figura 35).

A Cayaporina apresentou identidade de 69% (expectância de  $2^{-18}$ ) com o peptídeo antimicrobiano nomeado Opistoporin1 que foi isolado da peçonha do escorpião *Opistophtalmus carinatus* por Moerman e colaboradores em 2002.

Foram também isoladas da peçonha de *O. carinatus*: a Opistoporina2 que se diferencia da Opistoporina1 apenas em um resíduo de aminoácido e não teve sua atividade biológica testada (Moerman *et al.*, 2002). As Opistoporinas3 e 4 que estão disponíveis no Gene Bank (Swiss-Prot) sob os códigos Q5VJS8 e Q5VJS9 respectivamente, foram caracterizadas a partir de clonagem gênica de *O. carinatus*. As Opistoporinas1 e 2 juntamente com os peptídeos antimicrobianos Hadrurina, isolada do escorpião *Hadrurus aztecus* (Torres-Larios *et al.*, 2000), Parabutoporina, da peçonha do escorpião *Parabuthus schlechteri* (Verdonck *et al.*, 2000) e Pandinina1, caracterizada do escorpião *Pandinus imperator* (Corzo *et al.*, 2001) além do peptídeo BmKb1p caracterizado a partir do cDNA do escorpião *Buthus martensii* formam a subfamília 3 dos NDBPs (Zeng *et al.*, 2005). Os peptídeos Hadrurina, Parabutoporina, Pandinina1 e Opistoporina1 apresentaram atividade antimicrobiana e hemolítica, e o peptídeo BmKb1p não teve sua atividade hemolítica testada. Parabutoporina e Opistoporina1 apresentaram também atividade antifúngica e imunomoduladora (Zeng *et al.*, 2005).



A massa molecular da Cayaporina foi encontrada no *fingerprint* realizado da peçonha bruta de *O. cayporum* (Schwartz *et al.*, 2008). Este peptídeo foi purificado e sua atividade sua atividade antimicrobiana testada, como será descrito adiante.

## 2.2 Seqüências nucleotídicas sem função atribuída

Foram encontradas 6 seqüências nucleotídicas que não apresentaram identidade (com expectância menor que  $10^{-5}$  para que se possa ser atribuída função) com seqüências depositadas nos bancos de dados pesquisados. Essas seqüências são: Ocy 12 (*contig*, representado por 4 clones) e os singletos Ocy 13, Ocy14, Ocy15, Ocy16 e Ocy17.

Ocy12 é composto por 413 nucleotídeos. A partir da predição da seqüência de aminoácidos podemos inferir que o peptídeo sinal é composto por 23 resíduos e é seguido por uma seqüência de 45 aminoácidos (Figura 37). Considerando esses 45 aminoácidos, a massa monoisotópica calculada dessa seqüência seria de 5384.83 Da, essa massa molecular não foi encontrada no *fingerprint* de *O. cayporum*.

1	ATG	AAG	CTA	CGC	GCG	CTG	CTG	TTA	GTG	ACG	GTT	ATT	CGC	TTC	GCT	45
1	M	K	L	R	A	L	L	L	V	T	V	I	R	F	A	15
46	ATC	ATC	TCT	TCG	TCA	TAT	TGC	CAA	GAT	ATC	TAC	CAA	CTC	GAA	GAC	90
16	I	I	S	S	S	Y	C	Q	D	I	Y	Q	L	E	D	30
91	GGG	TAT	GCG	CCG	ATT	GAA	AAG	CGG	AAT	CCA	GTA	CTG	ACC	TGG	GCA	135
31	G	Y	A	P	I	E	K	R	N	P	V	L	T	W	A	45
136	GCC	GGG	AAA	GCA	GCG	GGA	TGG	GGC	TGG	AAA	AAA	TGG	AGG	AAC	CAC	180
46	A	G	K	A	A	G	W	G	W	K	K	W	R	N	H	60
181	CGT	AAA	AAT	CAC	CGC	GGT	CGC	CGC	TAA	TAAAAAGGATTACAAACCTGGATC						
61	R	K	N	H	R	G	R	R	stop							
GTTAACAAAAGAAACATGAAAGTACGAACATGAACTGGATTTATGTAATTGTTTAATAAT																
TAATCAATTTTCAATAAACACACTGAGTA-polyA-3'																

**Figura 37: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy 12.** Não foi encontrada similaridade com outro peptídeo conhecido. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de Poli-A está em duplo sublinhado.

Ocy13 é composto por 825 nucleotídeos e pela predição da seqüência de aminoácidos, infere-se que o peptídeo sinal é composto por 20 aminoácidos seguido por uma seqüência de 155 resíduos de aminoácidos incluindo 4 resíduos de cisteína que

formariam 2 pontes dissulfeto (Figura 38). Considerando-se os 155 aminoácidos e as pontes dissulfeto formadas, a massa monoisotópica calculada seria de 17415.71 Da.

Ocy14 é composto por 393 nucleotídeos. Através da predição da seqüência de aminoácidos, infere-se que o peptídeo sinal é composto por 26 aminoácidos seguido por uma seqüência de 26 resíduos de aminoácidos (Figura 39). A massa monoisotópica calculada dessa seqüência seria de 3114.54 Da, considerando os 26 aminoácidos.

1	ATG	ATG	TCC	CGT	ATT	GTT	GGA	TCT	GCC	ATG	TTA	ATT	TCT	TTT	ACT	45
1	<i>M</i>	<i>M</i>	<i>S</i>	<i>R</i>	<i>I</i>	<i>V</i>	<i>G</i>	<i>S</i>	<i>A</i>	<i>M</i>	<i>L</i>	<i>I</i>	<i>S</i>	<i>F</i>	<i>T</i>	15
46	GCT	TTG	GCT	CAG	TGC	TGC	CAC	CAT	TCG	ATA	AAT	ACA	TCT	GCT	GCC	90
16	<i>A</i>	<i>L</i>	<i>A</i>	<i>Q</i>	<i>C</i>	<i>C</i>	H	H	S	I	N	T	S	A	A	30
91	AAT	GCA	AGT	ATC	GCC	AAT	TCC	AGC	ACA	GGC	AAT	GAT	CTT	CAA	GAT	135
31	N	A	S	I	A	N	S	S	T	G	N	D	L	Q	D	45
136	TCC	TGT	ATG	ATG	CTT	GAT	CCC	AAC	GAT	TAT	GTA	GAT	CAA	CTA	ATA	180
46	S	<i>C</i>	M	M	L	D	P	N	D	Y	V	D	Q	L	I	60
181	GTG	GAA	CTG	AAT	AAT	AAG	CCC	ATC	GAC	ACC	GTC	CCT	TTG	CCC	AAT	225
61	V	E	L	N	N	K	P	I	D	T	V	P	L	P	N	75
226	AAC	TCC	GGA	TCC	ACC	TTT	TTT	AAA	GGT	CAA	ATG	TGG	AGT	CTC	AGT	270
76	N	S	G	S	T	F	F	K	G	Q	M	W	S	L	S	90
271	AGC	TTG	AGA	AGG	ACC	GGT	AAT	GCT	ACA	GTG	CTG	TGC	AAT	CAG	TCA	315
91	S	L	R	R	T	G	N	A	T	V	L	<i>C</i>	N	Q	S	105
316	TCG	ATC	ACA	ATA	GAG	GTG	TTG	TTG	TCT	TTC	GAA	CAA	CTG	AAA	GGG	360
106	S	I	T	I	E	V	L	L	S	F	E	Q	L	K	G	120
361	CAT	TAC	AAA	TGG	GAA	GGA	AAG	GTT	CTT	TTC	GAA	ACC	TTT	TCC	GGG	405
121	H	Y	K	W	E	G	K	V	L	F	E	T	F	S	G	135
406	TTT	ATA	AAC	ATT	AAA	GTG	GAT	GCA	GGG	GAT	ATA	TTG	TTC	CGA	GTG	450
136	F	I	N	I	K	V	D	A	G	D	I	L	F	R	V	150
451	AAA	TCA	CAC	CGC	TGG	AAC	GAT	ACT	GGG	CTA	ACA	TTA	GAT	AGG	CTA	495
151	K	S	H	R	W	N	D	T	G	L	T	L	D	R	L	165
496	AGT	ATG	AAA	ATA	TGT	CTA	GAA	TTC	GTA	CTG	AAG	TGA	CAG	GGT	TGG	GAG
166	S	M	K	I	<i>C</i>	L	E	F	V	L	K	stop				

TTTTCACCTGNTCTANTAGCATATACTGAGCCTGCAGTTAATTTCNNTTAGAAGTCGTTT  
 TGCTAGATGGATAGAGGCCACTGAAAGATGCTATGGGATGGACTTCGTGATATGAAATTT  
 CGATGTCTGACCTGCAGTATCATAGCAAGTTCTGTGGTATCATGATGAGACAACATTGTATAG  
 GATTCGGACCAACATTGTAATAGGTTTTCCAGTATCATTTACATGCTCATTGAATAAATAGTT  
 GGGTGTATGAAGATATATTCTCCGAATGTATGGGTTACACAAATAAAGCCAAGCCTTC-

poly a -3'

**Figura 38: Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy13.** Não foi encontrada similaridade com outro peptídeo depositado nos bancos de dados. O provável peptídeo sinal está marcado em azul e o sinal de poli-A está em duplo sublinhado.

1	ATG AAT GCC CTT GTC TGC ATC GTC CCT ACT GTC CAT TTT ATG TGT	45
1	<i>M N A L V C I V P T V H F M C</i>	15
46	CGA CTA CTG TCC CTC CTT CAC GCT ACT GTC GCA ACT ATT TGG AAT	90
16	<i>R L L S L L H A T V A T I W N</i>	30
91	GGA AAT CAC AGC CAC TTC AAA CCA CGA GAA ACG CAT CAC CTA AAC	135
31	G N H S H F K P R E T H H L N	45
136	GAC CTC TAT CCA TGT ATT CTA TGA AAAAAATCTCTGAAGTCGAATCCGACTT	180
46	D L Y P C I L stop	

ACTGAAACATCTACTGCCTACGATACTTCGTCAACATACTGGACATATCTACAATATGCC

GGTTATCATTACACGACATCCTGACACAATGTAATATTTCTACTTTTTAGAAATCAATA

AAATATTTTTCTGC-polyA-3'

**Figura 39: Seqüência de nucleotídeos e predição da seqüência de aminoácidos do Ocy14.** Não foi encontrada similaridade com outro peptídeo depositado nos bancos de dados. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de poli-A está em duplo sublinhado.

1	ATG AAA ACC CAA CTG GTT GTA ATC ATT CTC TTG GTT TGT GCT CTG	45
1	<i>M K T Q L V V I I L L V C A L</i>	15
46	CTT CCA ACT GAA GGA ATG CCG ATA GAA GAG GAG CGA GTG CCT GAT	90
16	<i>L P T E G M P I E E E R V P D</i>	30
91	TTA GCC GAC TAT GCT CTG TGT GTT GCG AAG TGT AAA TTA ATG GAG	135
31	L A D Y A L <b>C</b> V A K <b>C</b> K L M E	45
136	AGC CAA AAA TTA TTG CAA TGC CTT ATG ACC TGT TTA AAC CAC CTC	180
46	S Q K L L Q <b>C</b> L M T <b>C</b> L N H L	60
181	TTC GGC TGA ATGCTACTTTTGAAAGTTCT <u>AAATAAAAT</u> TATTCTCAGAAACC-polyA-3'	
61	F G stop	

**Figura 40: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy15.** Não foi encontrada identidade com outro peptídeo depositado nos bancos de dados. Em vermelho tem-se 4 cisteínas, que formariam 2 pontes dissulfeto. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de poli-A está em duplo sublinhado.

Ocy15 é composto por 371 nucleotídeos e a partir dos dados da predição da seqüência de aminoácidos, inferimos que o peptídeo sinal é composto por 21 aminoácidos seguido por uma seqüência de 41 resíduos de aminoácidos incluindo 4 cisteínas que formariam 2 pontes dissulfeto (Figura 40). Considerando essas pontes dissulfeto e os 41 resíduos de aminoácidos, a massa monoisotópica calculada seria de 4660.30 Da.

Ocy16 é composto por 358 nucleotídeos. A partir dos dados da predição da seqüência de aminoácidos, inferimos que o peptídeo sinal é composto por 21 aminoácidos seguido por uma seqüência de 50 resíduos de aminoácidos, com 2 cisteínas que formariam uma ponte dissulfeto (Figura 41). A massa monoisotópica calculada seria de 5817.98 Da, considerando



a ponte formada e os 50 aminoácidos, essa massa molecular não foi encontrada na caracterização proteômica de *O. cayaporum*.

1	ATG	CCT	CAG	CTT	CTT	TTA	CTG	GCG	TTG	ATA	CTC	TTG	GCA	ATT	CTT	45
1	<i>M</i>	<i>P</i>	<i>Q</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>A</i>	<i>L</i>	<i>I</i>	<i>L</i>	<i>L</i>	<i>A</i>	<i>I</i>	<i>L</i>	15
46	GCA	AAG	ACC	ATG	GCA	ATG	ATG	GAA	AAC	AAG	AGA	GTA	GAG	ACT	TTT	90
16	<i>A</i>	<i>K</i>	<i>T</i>	<i>M</i>	<i>A</i>	<i>M</i>	M	E	N	K	R	V	E	T	F	30
91	GGC	GAG	CTC	GAA	AAG	AGG	TCG	GAA	ACT	TCT	GGA	AAG	GAA	CGA	GAA	135
31	G	E	L	E	K	R	S	E	T	S	G	K	E	R	E	45
136	ACA	GAC	CGC	ATC	ATC	GGG	GCG	GTC	AAA	CCC	TTC	ATA	AAG	TGC	TTC	180
46	T	D	R	I	I	G	A	V	K	P	F	I	K	<b>C</b>	F	60
181	AGG	GCC	CCC	TGT	CCT	GGG	AAA	AGA	AGC	TAT	TTC	TGA	GCTGACTTAGTA			
61	R	A	P	<b>C</b>	P	G	K	R	S	Y	F	stop				

CTATTAATATATCTGTAAGTTGATGCCTAGAGAAATAAACGTTATGACTGCTTAAATAA

ATTACTAAAAATAAACCATAAACAT

**Figura 41: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy16.** Não foi encontrada similaridade com outro peptídeo depositado nos bancos de dados. Em vermelho tem-se as 2 cisteínas. O provável peptídeo sinal está marcado em azul, e as prováveis seqüências do sinal para Poli-A estão em duplo sublinhado.

Ocy17 é composto por 368 nucleotídeos. A partir da predição da seqüência de aminoácidos, infere-se que o peptídeo sinal é composto por 20 aminoácidos seguido por uma seqüência de 35 resíduos de aminoácidos com 4 cisteínas (Figura 42). Considerando as cisteínas que formariam 2 pontes dissulfeto e os 35 aminoácidos, a massa molecular monoisotópica calculada seria de 4017.93 Da.

1	ATG	AGG	ACG	CTG	CTG	TTG	GTA	CTA	CTG	ATA	ATC	TCA	GTC	GTG	TGC	45
1	<i>M</i>	<i>R</i>	<i>T</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>V</i>	<i>L</i>	<i>L</i>	<i>I</i>	<i>I</i>	<i>S</i>	<i>V</i>	<i>V</i>	<i>C</i>	15
46	ATT	GCC	CTG	AGT	GAT	GCA	AAG	CTT	GGA	TGT	TAT	TGT	CCA	GGA	AGA	90
16	<i>I</i>	<i>A</i>	<i>L</i>	<i>S</i>	<i>D</i>	A	K	L	G	<b>C</b>	Y	<b>C</b>	P	G	R	30
91	GGA	AGA	CCT	GGT	CCG	TGG	TTA	CCG	ACT	TGC	TTT	TAC	TGC	CCT	CCA	135
31	G	R	P	G	P	W	L	P	T	<b>C</b>	F	Y	<b>C</b>	P	P	45
136	AAA	TAT	CCG	TGG	AAT	TGG	CCT	GTT	CGT	GGA	TAA	TAGATAACTTGAGTC				
46	K	Y	P	W	N	W	P	V	R	G	stop					

TCCAGTGTTCAATAATATACAGCACTTAACAGTTGATTCTCTCAATGTATTTGTTTCTC

TTAAAAGAAATAACGCTTGTAACCTATAATTAGATGTCAAAATAAAATTTTGTGGGT

TTAGAGCATG-polyA-3'

**Figura 42: Seqüência de nucleotídeos e predição da seqüência de aminoácidos de Ocy17.** Não foi encontrada similaridade com outro peptídeo depositado nos bancos de dados. O provável peptídeo sinal está marcado em azul, e a provável seqüência do sinal de Poli-A está em duplo sublinhado. Em vermelho tem-se 4 cisteínas, que formariam 2 pontes dissulfeto.

### 3. Descrição do peptídeo Cayaporina

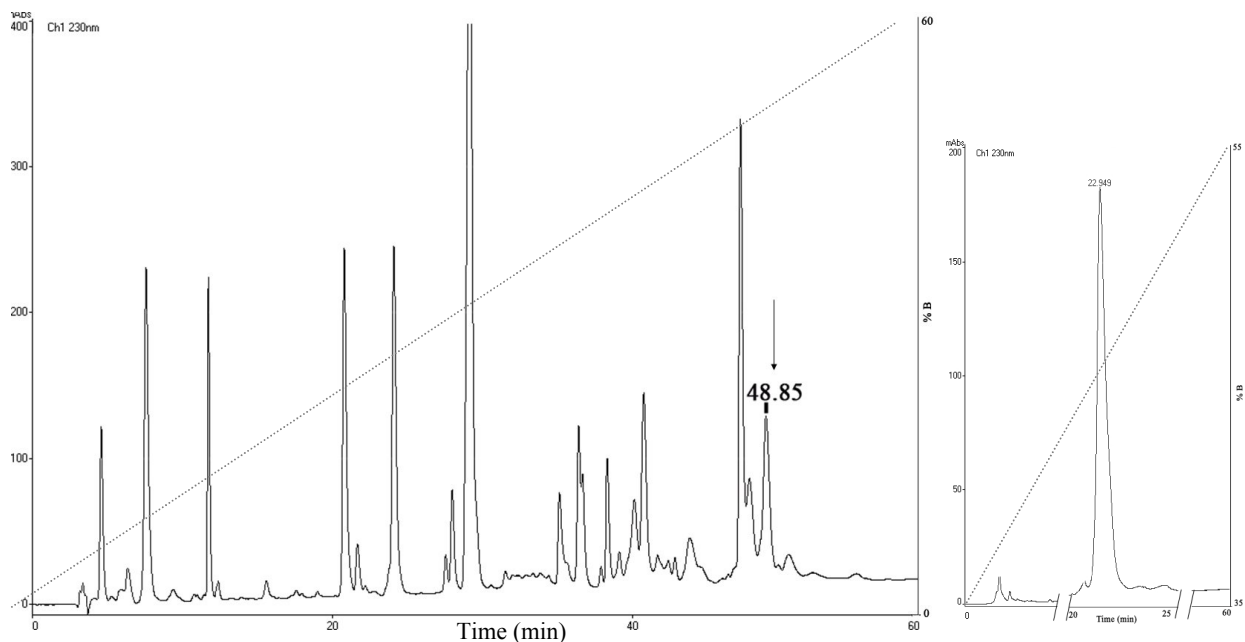
Como descrito, a seqüência de 411 nucleotídeos que apresentou identidade aos antimicrobianos denominados Opistoporinas (Moerman *et al.*, 2002) foi nomeada cayaporina. A massa molecular monoisotópica teórica calculada para essa seqüência foi de 4675.53 Da e foi encontrada na caracterização proteômica de peçonha bruta de *O. cayaporum* (Schwartz *et al.*, 2008).

Na cromatografia de peçonha bruta a fração referente à Cayaporina, elui a 48,85% de solução B. Essa fração foi recromatografada também em coluna C<sub>18</sub>, obtendo-se a cayaporina. A figura 42 mostra a cromatografia obtida em HPLC da peçonha bruta e a recromatografia da fração correspondente à Cayaporina.

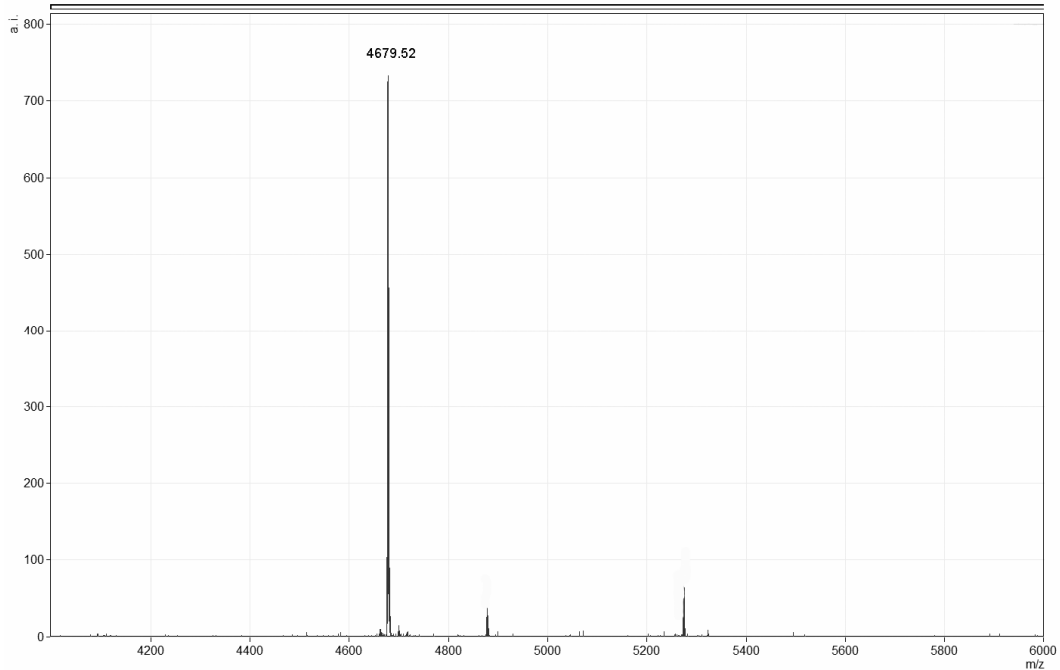
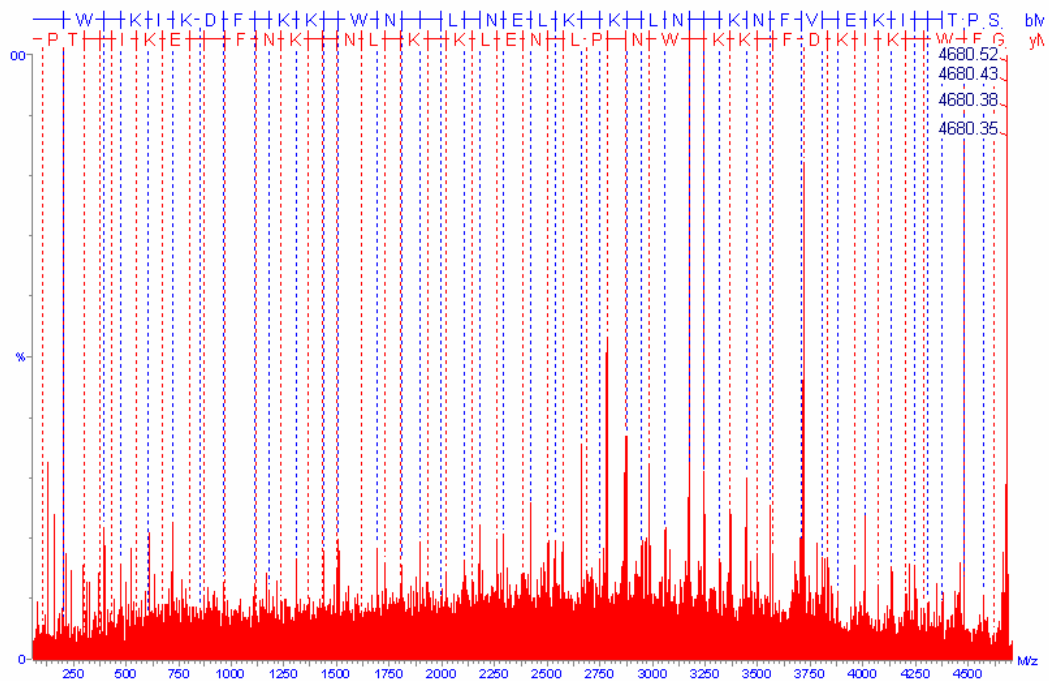
O peptídeo obtido na recromatografia foi analisado por espectrometria de massa em MALDI-TOF/TOF, sendo fragmentada e interpretada por sequenciamento *de novo* (figura 42). A massa molecular obtida por espectrometria foi de 4675 Da, e a seqüência de aminoácidos foi confirmada utilizando-se sequenciamento automático de Edman (Figura 43).

O peptídeo purificado, Cayaporina, foi testado quanto à capacidade de inibir o crescimento de bactérias Gram-negativa, (*E. coli* -ATCC 25922), e Gram-positiva (*S. aureus*-ATCC 29213). O MIC obtido da Cayaporina em *E. coli* foi de 28 µM. Nesta concentração, a cayaporina inibiu em 56% o crescimento de *S. aureus*.

Na concentração de 28 $\mu$ M, a Cayaporina não apresentou atividade hemolítica em suspensão de hemácias humanas.



**Figura 43: Purificação da cayaporina em HPLC.** Foi aplicado 1 mg de peçonha bruta de *O. cayaporum* em HPLC e separado utilizando-se coluna de fase-reversa analítica C<sub>18</sub>, eluido com um gradiente linear de acetonitrila (0 a 60% em 60 min). A fração indicada com uma seta (com tempo de retenção de 48.89) foi novamente separada utilizando-se uma coluna de fase-resersa C<sub>18</sub> analítica, em gradiente linear de 35 a 55% de acetonitrila em 40 min. O composto mostrado na janela menor, com o tempo de retenção de 22.95 corresponde à cayaporina purificada.

**A****B****C** GFWSKIKDFAKKAWNSPLANE . . .

**Figura 44: Espectro, fragmentação do peptídeo Cayaporina e Seqüência parcial de aminoácidos obtida por Edman. A:** Espectro de massa mostrando o pico referente à cayaporina purificada. **B:** Fragmentação e seqüência de aminoácidos, predita com massa molecular de 4680.52 Da seqüenciamento orientado pela seqüência da cayaporina obtida por cDNA. **B:** Seqüência de aminoácidos obtida por seqüenciamento automático de Edman.

## DISCUSSÃO

Os transcriptomas das glândulas de veneno de vários animais, como serpente, (Zhang *et al.*, 2006), peixes (Magalhães *et al.*, 2006), aranhas (Jiang *et al.*, 2008) e escorpiões (Schwartz *et al.*, 2007) têm contribuído para conhecer a composição da peçonha desses animais. Esses trabalhos também identificaram alguns transcritos com diferentes funções, o que possibilita vislumbrar o metabolismo de produção da peçonha nas glândulas altamente especializadas.

Os dados obtidos com essa metodologia podem ser utilizados para a busca direcionada de toxinas em estudos de caracterização proteômica para a identificação da estrutura gênica para o conhecimento do processamento da toxina na glândula, em análises filogenéticas e de biogeografia, em estudos de evolução das toxinas, e outras contribuições para a toxinologia.

### 1. Biblioteca de cDNA

Os animais foram ordenhados 5 dias antes da extração do RNA total porque segundo Paine e colaboradores, (1992) há um aumento considerável na produção de RNAm após esse intervalo. Como descrito por Yamaji *et al.*, (2004), há componentes sexo-específicos na peçonha do escorpião *Opisthacanthus madagascariensis* e por essa razão, para a construção da biblioteca de cDNA em *O. cayaporum*, utilizamos os télsons de dois espécimes, um macho e uma fêmea, afim de caracterizarmos os componentes transcritos na glândula dessa espécie.

O título médio do banco de cDNA não-direcionado de *O. cayaporum* (2233,33 cfu/mL) foi bem menor do que o obtido em outros trabalhos, como o transcriptoma não-direcionado de *H. gertschi* que obteve um título de  $2,8 \times 10^5$  cfu/mL (Schwartz *et al.*, 2007) e o trabalho de busca direcionada de clones de Jiang *et al.*, (2008), que obteve título de  $1.0 \times 10^6$  cfu/mL. A porcentagem média clones recombinantes foi de 63,8%, também foi menor do que em outros trabalhos como de Schwartz *et al.*, (2007) que obteve 99% de clones recombinantes. Houve a diminuição nas porcentagens de recombinação calculadas das

quatro transformações, sendo que a primeira obteve-se porcentagem de 90,1% de recombinação, e na quarta transformação de 30,3%.

A construção de biblioteca de cDNA com altas porcentagens de recombinantes e alto título, depende da eficiência de várias etapas e reagentes. A escolha do kit de purificação do RNA, a qualidade das células *E. coli* DH5 $\alpha$ , a transformação por choque térmico que possui menor eficiência do que a transformação por eletroporação (Azevedo *et al.*, 2003) são fatores que podem explicar os baixos índices de título e de recombinantes que foram obtidos nesse trabalho.

A tabela 3 compara alguns dados obtidos com a biblioteca de cDNA de *O. cayaporum* com outros trabalhos de análise de transcritos da glândula de peçonha.

**Tabela 3. Comparação de alguns dados obtidos na biblioteca de cDNA de *O. cayaporum* com outros trabalhos de transcriptoma.** O “tamanho médio das EST” inclui as seqüências de pro-peptídeos, peptídeos sinal e poliadenilação, excluindo-se as regiões de vetor. “EST analisadas/seq. distintas” refere-se ao número de EST de alta qualidade que foram analisados e o número de ESTs que são distintos. “Sem função atribuída” refere-se as seqüências que não obtiveram os parâmetros de predição de função. “seq sem identidade encontrada” refere-se as seqüências que não apresentaram similariedade com outro componente depositado dos bancos de busca.

Espécie	Referência	EST analisadas/seq. distintas	Tamanho médio das EST	Porcentagem dos clones segundo a provável função exercida			
				Toxinas	Componentes de processos celulares	Sem função atribuída	Seq sem identidade encontrada
<i>Opisthacanthus cayaporum</i>		130 / 67	535 pb	30%	33%	17%	15%
<i>Hadrurus gertschi</i>	Schwartz <i>et al.</i> , 2007	147 / 68	919 pb	31,3%	17,7%	1,4%	19,7%
<i>Thalassophryne nattereri</i>	Magalhães <i>et al.</i> , 2006	775 / 516	390 pb	18,8%	14,3%	6,8%	32%
<i>Deinagkistrodon acutus</i>	Zhang <i>et al.</i> , 2006	8696 / 2332	398 pb	40,16%	39,85%	20%	Não informado
<i>Ornithoctonus huwena</i>	Jiang <i>et al.</i> , 2008	468 / não informado	508 pb	68%	13%	19%	Não informado

O número de EST analisados da biblioteca de *O. cayaporum* está mais próximo ao número analisado do transcriptoma no escorpião mexicano *H. gertschi*, em que foi utilizado o mesmo kit de extração de RNA total, e o mesmo kit para a construção da biblioteca de cDNA (Schwartz *et al.*, 2007). A porcentagem de seqüências distintas em relação ao total

de ESTs analisados para *O. cayaporum* foi de 51%, o que está de acordo com os outros trabalhos analisados: 42,5 para *H. gertschi*, escorpião mexicano; 66% para *T. nattereri*, peixe brasileiro; e 26% para *D. acutus*, serpente chinesa (Schwartz *et al.*, 2007; Magalhães *et al.*, 2006 e Zhang *et al.*, 2006).

Mesmo com o baixo título e com porcentagem de clones recombinantes de 66%, as porcentagens de seqüências de prováveis toxinas (30%), de componentes do processamento glandular e de seqüências sem funções atribuídas, estão de acordo com o relatado em outros trabalhos (Magalhães *et al.*, 2006; Zhang *et al.*, 2006; Schwartz *et al.*, 2007; Zhang *et al.*, 2006 e Jiang *et al.*, 2008).

Apesar de neste trabalho termos clonado e caracterizado poucas seqüências, em relação ao total previsto de peptídeos da peçonha, conseqüentemente de transcritos de uma glândula (Becerril *et al.*, 2000), este estudo demonstra a diversidade e riqueza de componentes e processos moleculares presentes na glândula de *O. cayaporum*.

## **2. Seqüências de nucleotídeos obtidas**

Por termos utilizado uma biblioteca de cDNA não amplificada, a freqüência das seqüências obtidas reflete da presença do respectivo RNAm na glândula (Magalhães *et al.*, 2006 e Jiang *et al.*, 2008). Assim podemos inferir que as seqüências de nucleotídeos que foram caracterizadas representam os transcritos presentes em diferentes níveis de abundância. Das prováveis toxinas, a mais freqüente encontrada foi a provável toxina para canal de potássio Ocy11. Das seqüências que apresentaram identidade com proteínas de processamento celular, a mais freqüente encontrada foi Ocy3 que é similar à proteína ribossomal de *Scorpius maurus palmatus*.

### **2.1 Prováveis Toxinas**

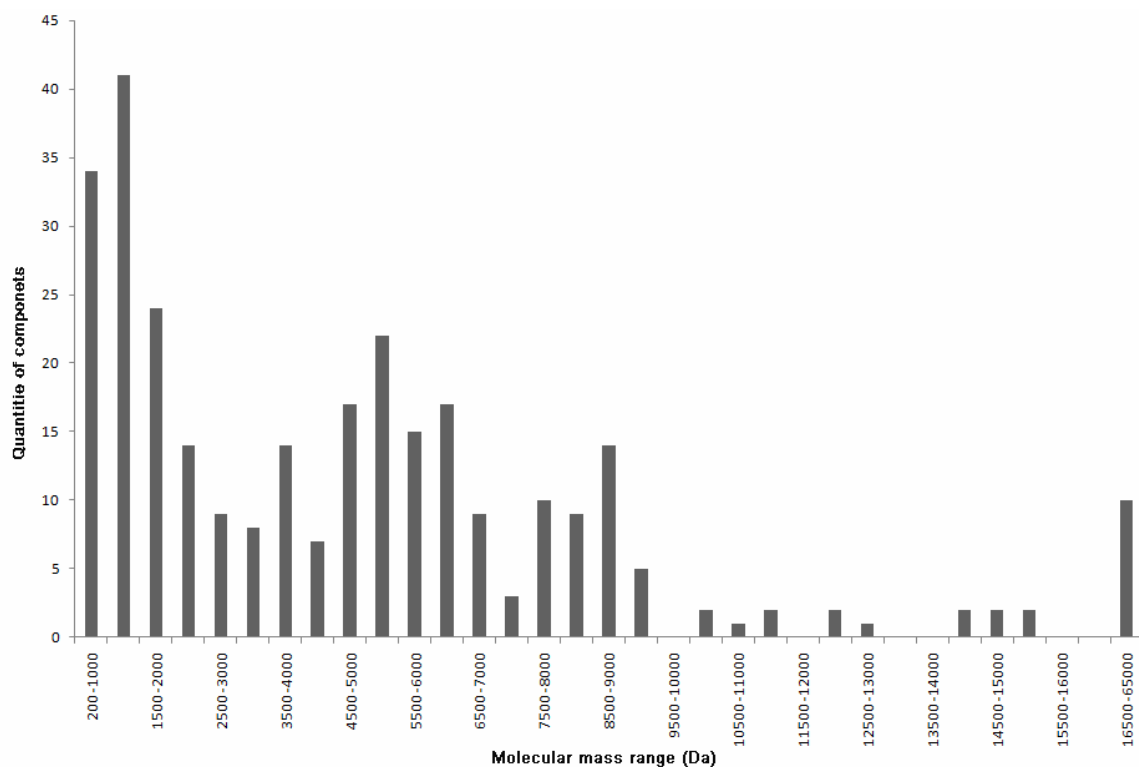
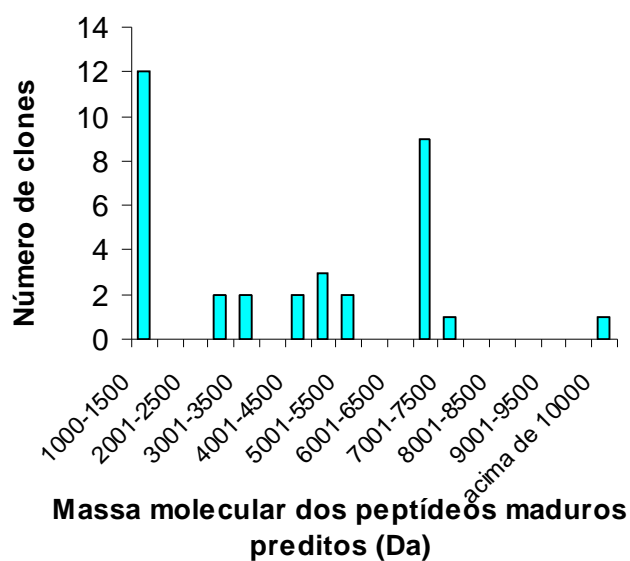
Os dados de massa molecular obtidos em trabalhos de caracterização proteômica combinados com os dados das bibliotecas de cDNA fornecem meios rápidos e altamente consistentes para o seqüenciamento *de novo*. Essa metodologia foi realizada por Jiang e colaboradores, 2008, com a aranha *Ornithoctonus huwena*, onde obtiveram importantes dados acerca da produção de toxinas, bem como o seqüenciamento de novos componentes.

Os dados de massa molecular teórica das prováveis toxinas (regiões de peptídeos maduros preditos) da biblioteca de cDNA de *O. cayaporum* foram relacionados com os dados de massa molecular obtidos para peçonha da mesma espécie (Schwartz *et al.* 2008). A figura 44 “A” mostra a distribuição de massas moleculares dos componentes da peçonha e “B” a distribuição das massa teóricas das prováveis toxinas do transcriptoma.

Mesmo com o pequeno número de clones, em relação aos dados de massa molecular do trabalho de proteômica, foi encontrado padrão semelhante de distribuição de massas. Há uma distribuição elevada de toxinas preditas com massas moleculares em torno de 1000 a 1500 Da, que no presente trabalho são prováveis peptídeos antimicrobianos. As toxinas preditas de 7001 a 7500 Da, correspondem a massas moleculares de prováveis toxinas para canais iônicos e fosfolipases com massas acima de 10000 Da. Assim as massas moleculares calculadas das prováveis toxinas obtidas na biblioteca de cDNA estão condizentes com os resultados obtidos na caracterização dos peptídeos da peçonha de *O. cayaporum*:

- ✓ Prováveis antimicrobianos: 1000 a 1500 Da;
- ✓ Prováveis toxinas para canais iônicos: 7001 a 7500 Da;
- ✓ Possíveis fosfolipases: acima de 10000 Da;



**A****B**

**Figura 44. Comparação da distribuição de massas moleculares . A:** Histograma de distribuição das massas moleculares obtidas por espectrometria de massa da peçonha de *O. cayaporum* (retirado do trabalho de Schwartz *et al.* 2008) **B:** distribuição das massas moleculares calculadas a partir das seqüências de nucleotídeos de prováveis toxinas da biblioteca de *O. cayaporum*. Observa-se semelhanças quanto ao padrão de distribuição das massas.

### 2.1.1 Seqüências com cisteínas que formariam pontes dissulfeto

### ✓ Prováveis toxinas para canais iônicos

Das 8 seqüências nucleotídicas que foram caracterizadas, 4 (Ocy10, OcyKtx1, Ocy9 e OcyKtx2) apresentaram identidade com peptídeos que foram isolados de aracnídeos, sendo que OcyKtx1 apresentou identidade com o peptídeo com atividade bloqueadora de canal para potássio. As outras 4 seqüências, Ocy6, Ocy7, Ocy 40 e Ocy41, apresentaram identidade a peptídeos ou seqüências que não tiveram a sua atividade biológica testada.

Ocy10 apresentou 65% de identidade ao peptídeo LaIT1 que foi isolado da peçonha do escorpião *Liocheles australasiae* (Ischnuridae). LaIT1 possui massa molecular de 4200 Da, quatro cisteínas formando 2 pontes dissulfeto e que se mostrou tóxica para grilos (dose de 1 µg/inseto) e sem efeito tóxico em ratos (Matsushita *et al.*, 2007). A toxicidade encontrada para grilos foi comparada à encontrada para outras insetoxinas de cadeia curta, como em toxinas de aranha, sugerindo que o alvo de ação desse peptídeo seja provavelmente os canais para K<sup>+</sup>, Cl<sup>-</sup> ou Ca<sup>2+</sup>. Esse peptídeo não apresentou homologia com nenhum outro peptídeo conhecido, sugerindo que este pertença a uma nova subfamília de insetoxinas com estrutura de pontes dissulfeto única (Matsushita *et al.*, 2007). Assim Ocy10 será alvo de busca na peçonha de *O. cayaporum*, para purificação e testes biológicos, a fim de determinar sua atividade biológica.

A seqüência Ocy11 apresentou 42% de identidade (*e-value* de 4<sup>-4</sup>), ao precursor nomeado TXLP1 (Número no GeneBank: AAK61816). Esse precursor foi identificado do escorpião *Buthus martensii*, Buthidae. Ocy11 também apresentou identidade de 28% ao peptídeo La1, contudo a expetância obtida foi de 0,052. La1 foi isolada da peçonha do escorpião da família Ischnuridae *Liocheles australasiae* (Miyashita *et al.*, 2007). La1 não teve sua atividade biológica testada, mas apresentou seqüência primária e disposição das 8 cisteínas semelhantes à toxinas escorpiônicas. Sendo que *L. australasiae* pertence à mesma família de *O. cayaporum*, é mais provável que a massa teórica calculada para Ocy11 seja de 8554.98 Da e que o provável peptídeo maduro seja composto por 78 resíduos de aminoácidos, baseando-se no alinhamento feito com Ocy11 (Figura 10). Essa massa molecular foi encontrada na caracterização proteômica de *O. cayaporum*, em dois tempos de retenção distintos, e esse peptídeo será purificado e terá sua atividade biológica testada.

OcyKtx1 e Ocy9 apresentaram identidade (65 e 85% respectivamente) à Omtx2, que é uma toxina bloqueadora de canal para K<sup>+</sup> isolada de *O. madagascariensis* (Ischnuridae) (Chagot *et al.*, 2005). As Om-toxinas (Omtx1 a Omtx4) são peptídeos com estrutura tridimensional incomum dentre os peptídeos moduladores de canais iônicos, apresentando duas  $\alpha$ -hélices estabilizadas por duas pontes dissulfeto, estrutura que só foi encontrada em *k*-hefutoxinas isoladas do escorpião *Heterometrus fulvipes*, família Scorpionidae (Srinivasan *et al.*, 2002) e as toxinas BgK e ShK isoladas de anêmona marinha (códigos no *Protein Data Bank*-PDB: 1BGK e 1ROO, respectivamente). As Om-toxinas apresentaram capacidade de ligação com os canais para K<sup>+</sup>: Kv1.1, Kv1.2 de ratos e Kv1.3. de humanos (Chagot *et al.*, 2005).

A massa molecular teórica calculada de OcyKtx1, 3134.36 Da, foi encontrada no *fingerpring* de *O. cayaporum* (Schwartz *et al.*, 2008), e a fração cromatográfica correspondente foi purificada e a atividade em canais para potássio está sendo estudada no Laboratório de Toxinologia da UnB.

As seqüências OcyKtx1 e Ocy9 serão melhor investigadas a fim de descrever suas estruturas tridimensionais e suas atividades biológicas, e compará-las com as Om-toxinas. Outro estudo a ser feito sobre essas toxinas, é a determinação de uma origem genética em comum, como a que foi sugerida entre as Om-toxinas e as *k*-hefutoxinas (pertencentes à subfamília  $\delta$ -Ktx) (Chagot *et al.*, 2005).

OcyKtx2 foi classificada como  $\alpha$ Ktx 6.15, pertencente à subfamília 6, de acordo a identidade apresentada e os alinhamentos feitos com as outras toxinas dessa subfamília. OcyKtx2 apresentou identidade de 39% à toxina sexo-específica de *O. madagascariensis* nomeada IsTX que foi classificada como  $\alpha$ Ktx 6, e possui estrutura tridimensional que apresenta: uma simples  $\alpha$ -hélice e duas folhas- $\beta$  conectadas por 4 pontes dissulfeto (Yamaji *et al.*, 2004). OcyKtx2 também será alvo de busca na peçonha bruta para purificação e identificação da sua atividade biológica.

Ocy6 apresentou apenas 32% de identidade com o peptídeo Is6 de *Ixodes scapularis*. O precursor nomeado Is6 foi caracterizado a partir da biblioteca de cDNA de *I. scapularis*, que foi construída utilizando-se machos na fase de acasalamento e alimentação (que freqüentemente ocorre ao mesmo tempo), a fim de identificar os componentes responsáveis pela maturação dos gametas e pela resposta imune. Foi encontrado o peptídeo Is6 apenas

uma vez no transcriptoma do carrapato, sendo que esse peptídeo está diretamente relacionado ao acasalamento dessa espécie (Packila & Guilfoile, 2002). Mesmo com a baixa identidade com o peptídeo Is6, a seqüência Ocy6 é importante alvo de investigação, pois além de possuir prováveis 5 pontes dissulfeto e seqüência completa com peptídeo sinal, possui estrutura primária não encontrada em outros peptídeos depositados no GenBank.

Ocy7 apresentou identidade e expectância considerável com o peptídeo nomeado Opistoporin2 caracterizado pelo cDNA do escorpião *Opisthophthalmus carinatus* (Scorpionidae). As opistoporinas (de 1 a 4) representam uma classe de pequenos peptídeos catiônicos antimicrobianos (SCAMPs – *Small cationic antimicrobial peptides*) que podem inserir na membrana plasmática formando poros que desestabilizam a membrana e fornecem a primeira defesa contra patógenos em vários organismos vertebrados, invertebrados ( artrópodos, nematóides) e em bactérias (Zhu & Tytgat, 2004).

Dentre as classes de SCAMPs têm-se as defensinas com conformação estrutural de  $\alpha$ -hélice e folha- $\beta$  (CS $\alpha\beta$ ) estabilizadas por 6 cisteínas que formam 3 pontes dissulfeto. Essas defensinas inibem o crescimento de um largo espectro de microrganismos, bactérias Gram-positivas, fungos filamentosos e alguns vírus envelopados. Em escorpiões, foi descrito também o peptídeo nomeado Scorpina com 75 resíduos de aminoácidos, que foi isolado da peçonha do escorpião *Pandinus imperator* (Scorpionidae) e apresentou atividade anti-malárica (Conde *et al.*, 2000). A scorpina apresentou 54% de identidade a seqüência Ocy7.

Estes trabalhos fornecem dados que reforçam a importância farmacológica ao estudo de componentes da peçonha de escorpiões, com várias aplicações clínicas apontadas. O peptídeo transcrito por Ocy7 caracterizado de *O. cayaporum*, possivelmente pertence a essa família de SCAMPs, com 6 cisteínas formando 3 pontes disulfeto, 53 aminoácidos, sendo importante alvo para purificação na peçonha e caracterização biológica. Para precursor do peptídeo Opiscorpina2 foram encontrados 3 sinais de polyadenilação (AATAAA), o que não ocorreu em Ocy7, onde só foi encontrado um sítio.

### ✓ **Prováveis fosfolipases**

Foram encontrados dois componentes com atividades fosfolipásicas na peçonha bruta de *O. cayaporum*, cujas seqüências N-terminais foram descritas (Schwartz *et al.*, 2008). Na

biblioteca de cDNA de *O. cayaporum* foram encontradas seqüências de nucleotídeos que podem codificar para fosfolipases: Ocyfosfolipase1 e Ocyfosfolipase2. Essas duas seqüências foram reseqüenciadas com primer reverso M13, contudo não foi possível determinar a seqüência completa das mesmas, com provável peptídeo sinal. Assim sugere-se que estes fragmentos de cDNA não estão completos, sendo necessário novas clonagens e seqüenciamentos para a determinação das seqüências completas.

Ocyfosfolipase1 apresentou identidade de 34% (*e-value*:  $9^{-12}$ ) a seqüência de Fosfolipase C caracterizada do genoma do peixe ósseo *S. purpuratus* (Número no GenBank: NP001008790). As fosfolipases C são importantes enzimas na transdução de sinal dependentes de  $Ca^{2+}$  e atuam em funções celulares como por exemplo, proliferação, apoptose, neurotransmissão, tráfico vesicular e regulação endócrina. Não há relatos na literatura de fosfolipases C caracterizadas de escorpiões.

Ocyfosfolipase2 apresentou 46% de identidade a fosfolipase A2 heterodimérica nomeada phaiodactylipina (*e-value*:  $5^{-21}$ ) caracterizada do cDNA de *Anuroctonus phaiodactylus* (Valdez-Cruz *et al.*, 2007). Fosfolipases A2 de peçonha de escorpiões (ScpPLA2 – *Scorpion venom phospholipases A2*) são em geral expressas na forma de um pro-peptídeo na região N-terminal, com um segmento interno de 5 ou 6 resíduos e uma região curta de C-terminal que é retirada para a formação da enzima madura, heterodimérica (Valdez-Cruz *et al.*, 2007). As ScpPLA2 são enzimas que catalizam a hidrólise em fosfolipídios sendo mediadores envolvidos em vários processos, como no processo inflamatório e hemostático. Alguns patógenos tem fosfolipases como fator de virulência, aumentando o dano tecidual. Muitas peçonhas animais contêm PLA2s, que atuam nos efeitos citolíticos, neurotóxicos, miotóxicos, nos distúrbios na coagulação sanguínea, edema, que estão presentes nos acidentes com esses animais (Schwartz *et al.*, 2007).

## **2.1.2 Toxinas sem pontes dissulfeto**

### **✓ Prováveis antimicrobianos**

Os peptídeos antimicrobianos isolados da peçonha de escorpiões são importantes fontes para o desenho de novas moléculas bioativas, e para o desenvolvimento de novos fármacos que minimizem a resistência desenvolvida pelos microrganismos às drogas

atualmente utilizadas (Zeng *et al.*, 2005). Neste trabalho foram caracterizadas 4 seqüências distintas de nucleotídeos que apresentaram identidade a peptídeos antimicrobianos de escorpião. Por essa identidade, essas seqüências de *O. cayaporum* serão buscadas na peçonha e serão testadas quanto à capacidade antimicrobiana, como o que foi realizado com o peptídeo Cayaporina.

Ocy1 e Ocy2 apresentaram identidade aos peptídeos IsCT1 e IsCT2 de *Opisthacanthus madagascariensis* que apresentaram atividade antimicrobiana em bactérias Gram-positivas e Gram-negativas (MIC, entre 0,7-17  $\mu\text{M}$  e 3.3->150  $\mu\text{M}$  respectivamente), e fraca atividade hemolítica em eritrócitos de carneiro (Dai *et al.*, 2002). O alinhamento feito com os precursores de IsCT1, IsCT2, Ocy1 e Ocy2 que permitiu inferir os peptídeos maduros para Ocy1 e Ocy2, pois os peptídeos maduros de IsCT1 e IsCT2 são sintetizados com a exclusão de um pro-peptídeo de 32 resíduos na porção C-terminal (Dai *et al.*, 2002). O antimicrobiano IsCT1 possui massa molecular de 1502.89 Da enquanto que o peptídeo antimicrobiano IsCT2 de 1464.81 Da, sendo que estes se diferem apenas por 3 resíduos de aminoácidos (Zeng *et al.*, 2005). Essas massas moleculares estão próximas às encontradas para os prováveis peptídeos maduros de Ocy1 e Ocy2, 1433.83 Da e 1454.89 Da respectivamente, sendo que essas duas seqüências se diferem por 5 resíduos de aminoácidos. Contudo essas massas moleculares teóricas não foram encontradas no *fingerprint*.

Ocy4 apresentou identidade ao peptídeo Pandinina2 isolado do escorpião *Pandinus imperator* (Scorpionidae) que é composto por 24 aminoácidos, tem massa molecular de 2610.46 Da e apresentou atividade antimicrobiana em bactérias Gram-positivas e Gram-negativas (MICs de 2.4-4.8  $\mu\text{M}$  e 19.1-38.2  $\mu\text{M}$  respectivamente). Além disso, possui atividade antifúngica com um MIC de 19.1  $\mu\text{M}$  e em uma concentração de 22.2  $\mu\text{M}$  causa 50% de hemólise em eritrócitos (Corzo *et al.*, 2001).

## **2.2 Componentes do processamento glandular**

As 47 seqüências de nucleotídeos que apresentaram identidade com componentes do processamento glandular foram classificadas em 6 grupos de acordo com a provável função exercida. As seqüências obtidas estão no Anexo e em parênteses está a porcentagem de cada grupo em relação às seqüências de componentes celulares obtidas:

✓ **Enzimas (11%)**

Este grupo engloba as seqüências de nucleotídeos que apresentaram identidade a enzimas. Por exemplo, foi identificada a seqüência Ocy27 que apresentou identidade (31%, *e-value*:  $2^{-12}$ ) com a enzima *elastase* de *Macaca mulata* (número no GenBank: XP001083219) que quebra a elastina, que juntamente com o colágeno, fornece a conexão dos tecidos.

✓ **Proteínas estruturais (14%)**

Ocy23 apresentou identidade de 55% (*e-value*:  $1^{-19}$ ) com a proteína do tipo *prefoldina* que em eucariotos é utilizada para a formação de moléculas de actina nos citoesqueletos (número no GenBank: XP848520). Ocy28 apresentou 96% de identidade, (*e-value*:  $3^{-114}$ ), com a actina caracterizada do ácaro *Rhipicephalus microplus* (número no GenBank: AAP81256).

✓ **Metabolismo (9%)**

Neste grupo estão proteínas que participam de alguma via metabólica celular, como a seqüência Ocy20 que apresentou identidade de 71%, (*e-value*:  $1^{-61}$ ), a *ferritina* de *Ornithodoros parker* (número no GenBank: AY156569), envolvida nas vias de ativação da imunidade inata. No trabalho de Magalhães *et al.*, (2006), com transcriptoma do peixe *Thalassophryne nattereri* foram encontrados ESTs que também apresentaram identidade à ferritina. Outra seqüência, Ocy35 apresentou identidade com a proteína do tipo calmodulina. Esta proteína se liga a diversos alvos e assim afeta diferentes funções celulares como: inflamação, apoptose, contração muscular, resposta imune e outras.

Em trabalhos de transcriptoma, é comum encontrar proteínas que estão relacionadas à produção energética da glândula para constante síntese protéica. Em *O. cayaporum* temos, por exemplo, Ocy26 que apresentou 44% de identidade com NADH desidrogenase de *Buthus martensii* (número no GenBank: YP214898).

✓ **Gene mitocondrial (65%)**

Este grupo contém: Ocy3 (seqüência *contig* composta por 24 clones), Ocy18 (*contig* que compreende 3 clones) e Ocy31. Ocy3 apresentou 70% de identidade ao gene para RNA ribossomal 16S da mitocôndria, caracterizado do escorpião *Scorpio maurus palmatus* (número no GenBank: AY156569). Ocy18 também apresentou identidade (73%, *e-value*: 5<sup>-11</sup>) ao gene que codifica para RNA ribossomal da mitocôndria (número no GenBank: DQ202525, caracterizado do Coleóptera *Biphyllus lunatus*). Como em outros trabalhos de transcriptoma (Zhang *et al.*, 2006) é comum obter seqüências nucleotídicas que são transcritas em proteínas relacionadas à produção energética das células, devido à intensa síntese protéica glandular. Ocy31 é composta por 677 nucleotídeos e apresenta identidade de 82% e expectância de 1<sup>-54</sup> com o gene ribossomal 28S de *Pandinus imperator* (número no GenBank: AY210830) (Mallat *et al.*, 2004). A presença desse transcrito na glândula indica a presença de síntese protéica intensa, o que também foi encontrado em outros trabalhos (Zhang *et al.*, 2006; Magalhães *et al.*, 2006 e Schwartz *et al.*, 2007).

### 2.3 Transcritos sem função atribuída

Foram descritas 6 seqüências nucleotídicas que não apresentaram identidade com outros peptídeos ou seqüências depositados no GenBank, porém apresentam características que sugerem que poderiam ser toxinas.

A presença de prováveis peptídeos sinais nesses ESTs sugere que estes codifiquem para toxinas (Magalhães *et al.*, 2006). Os prováveis peptídeos sinal dessas seqüências possuem identidade em torno de 20% entre si, tamanho médio de 21 resíduos de aminoácidos e o ponto de clivagem geralmente possui aminoácidos com cadeia lateral neutra (A, P, V, L, I, M, F e W). Essas características também são encontradas nas análises feitas com peptídeos sinal de toxinas escorpiônicas, sugerindo que possam ser toxinas ainda não descritas (Jing *et al.*, 2008 e Magalhães *et al.*, 2006).

Todas as 6 seqüências deste grupo apresentaram sinal de poliadenilação (AATAAA) e região de poli-A, indicando que estão completas em sua porção C-terminal. A presença de 3 sinais de poliadenilação na seqüência Ocy16 também foi encontrada nas Opiscorpinas de *Opisththalmus carinatus*, sendo que neste trabalho foi sugerido que a síntese desses peptídeos seria através de variáveis poliadenilações (Zhu & Tytgat, 2004). A presença de



cisteínas que formariam pontes dissulfeto, como nas seqüências Ocy16, Ocy17, Ocy15 e Ocy13, também é indício que estas seqüências possam ser novas toxinas. Outra característica indicativa dessa possível função, são as massas moleculares teóricas, 5384.83 Da, 3114.54 Da, 4660.30 Da, 5817.98 Da e 4017.93 (Ocy12, Ocy14, Ocy15, Ocy16 e Ocy17, respectivamente) que estão dentro da faixa de massas moleculares de toxinas escorpíonicas: 2000-7000 Da (Tan *et al.* 2006).

## 2.4 Cayaporina

Como descrito, a fração correspondente à massa molecular de 4675 Da foi purificada da peçonha de *O. cayaporum* e teve suas atividades antimicrobiana e hemolíticas testadas.

A cayaporina apresentou 69% de identidade ao peptídeo antimicrobiano Opistoporina1, isolado da peçonha de *O. carinatus* (Moerman *et al.*, 2002). A Opistoporina1 é composta por 44 resíduos de aminoácidos, possui massa molecular de 4833.65 Da e apresentou atividade antimicrobiana contra bactérias Gram-positivas (MIC de 12.5->50  $\mu$ M, para *S. aureus* o MIC foi >50  $\mu$ M) e Gram-negativas (MIC de 1.6-50  $\mu$ M, sendo que para *E. coli* o MIC foi de 12,5 $\mu$ M), além de atividade antifúngica (IC50 de 0.8-3.1  $\mu$ M) (Moerman *et al.*, 2002).

Em comparação aos peptídeos antimicrobianos pertencentes à subfamília 3, o MIC obtido para a cayaporina foi maior, 28 $\mu$ M, contudo dentre os componentes dessa subfamília apenas a Cayaporina não possui atividade hemolítica em hemácias humanas. Esses dados são promissores e importantes para o desenvolvimento de novas moléculas antimicrobianas.

## CONCLUSÃO

Foi realizada a primeira análise dos transcritos da glândula de peçonha de um escorpião brasileiro não-Buthidae, fornecendo a primeira visão global dos genes transcritos da glândula de peçonha de *O. cayaporum*.

A análise de ESTs permitiu a caracterização de várias seqüências demonstrando a riqueza e diversidade de componentes da peçonha dessa espécie, contribuindo assim para a descrição de novos fármacos, estudos de evolução de toxinas e filogenética das espécies de escorpião, bioengenharia de toxinas, e outras aplicações.

Foram analisados 130 ESTs da biblioteca de cDNA, sendo 68 seqüências distintas. Quinze seqüências de nucleotídeos distintas foram caracterizadas como prováveis toxinas da peçonha, por suas identidades a outras toxinas depositadas no GenBank. Outras 6 seqüências podem ser novas toxinas, pertencentes à famílias ainda não descritas. Quarenta e sete seqüências de prováveis componentes do processamento glandular foram apresentadas, além de seqüências que não foi determinada a provável função exercida. Foi observado também que as massas moleculares das prováveis toxinas estão distribuídas semelhantemente à distribuição encontrada para os peptídeos da peçonha de *O. cayaporum* (Schwartz *et al.*, 2008).

Com os dados da biblioteca de cDNA cruzados com a caracterização proteômica da peçonha bruta, foi caracterizado o primeiro peptídeo antimicrobiano não hemolítico isolado da peçonha de *O. cayaporum*, nomeado Cayaporina.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUL, S.F., GISH W., MILLER W., MYERS, E.W., LIPMAN D.J. **Basic local alignment search tool.** *J.Mol Biol*, 215: 403-410. 1990.
- AZEVEDO M.O., FELIPE M.S.S., BRIGIDO M.M., MARANHÃO A.Q., DE-SOUZA M.T. **Técnicas básicas em Biologia Molecular.** 1ª edição: editora Universidade de Brasília. 212p. 2003.
- BARNES. **Zoologia dos Invertebrados.** 7ª edição: editora Roca.1168pgs. 2005.
- BARONA J., BATISTA C.V., ZAMUDIO F.Z., GOMEZ-LAGUNAS F., WANKE E., OTERO R., POSSANI L.D. **Proteomic analysis of the venom and characterization of toxins specific for Na<sup>+</sup> and K<sup>+</sup> channels from the Colombian scorpion *Tityus pachyurus*.** *Biochim Biophys Acta.* 1764: 76-84.
- BATISTA C.V., D'SUZE G., GOMEZ-LAGUNAS F., ZAMUDIO F.Z., ENCARNACION S., SEVCIK C., POSSANI L.D. **Proteomic analysis of *Tityus discrepans* scorpion venom and amino acid sequence of novel toxins.** *Proteomics.*6:3718-3727. 2006.
- BATISTA C.V.F., GOMEZ-LAGUNAS F., LUCAS S., POSSANI L.D. **Tc1 from *Tityus cambridgei* is the first member of a new subfamily of scorpions toxins that blocks K<sup>+</sup> channels.** *Febs Letters* 486: 117-120. 2000.
- BATISTA C.V.F., POZO L., ZAMUDIO F.Z., CONTRERAS S., BECERRIL B., WANKE E., POSSANI L.D. **Proteomics of the venom from the Amazonian scorpion *Tityus cambridgei* and the role of prolines on mass spectrometry analysis of toxins.** *Journal of Chromatography.* 803: 55-66. 2004.
- BECERRIL, B., MARNGONI, S., POSSANI, L.D. **Toxins and genes isolated from scorpions of the genus *Tityus*.** *Toxicon*, 35, Issue: 6, 821:835. 1997.
- CATTERALL W A. & YU F.H. **Overview of the voltage-gated sodium channel family.** *Genome Biology.* 4:207. 2003
- CATTERALL W.A., CESTÉLE S., YAROV-YAROVY V., YU F. H., KONKI K., SCHEUER T. **Voltage-gated ion channels and gating modifier toxins.** *Toxicon.* 49: 124-141. 2007
- CESTÉLE S. & CATTERALL W.A., **Molecular mechanisms of neurotoxin action on voltage-gated sodium channels.** *Biochimie.* 82: 883-892.
- CHAGOT B., PIMENTEL C., DAÍ L., PIL J., TYTGAT J., NAKAJIMA T., CORZO G., DARBON H., FERRAT G. **A unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*.** *Biochem. J.* 388: 263-271. 2005

- CONDE R., ZAMUDIO F.Z., RODRIGUEZ M.H., POSSANI L.D. **Scorpina, na anti-malaria and anti-bacterial agent purified from scorpion venom.** *FEBS Lett.* 471:165-168. 2000.
- CORZO G., ESCOUBAS P., VILLEGAS E., BARNHAM K.J., HE W., NORTON R.S., NAKAJIMA T. **Characterization of unique amphipathic antimicrobial peptides from venom of the scorpion *Pandinus imperator*.** *Biochem. J.* 359: 35-45. 2001.
- DAI L., CORZO G., NAOKI H., ANDRIANTSIFERANA M., NAKAJIMA T. **Purification, structure-function analysis, and molecular characterization of novel linear peptides from scorpion *Opisthacanthus madagascariensis*.** *Biochim. Biophys. Res. Commun.* 293: 1514-1522. 2002.
- DAI L., YASUDA A., NAOKI H., CORZO G., ANDRIANTSIFERANA M., NAKAJIMA T. **IsCT, a novel cytotoxic linear peptide from scorpion *Opisthacanthus madagascariensis*.** *Biochem. Biophys. Res. Commun.* 293: 1514-1522. 2001
- DEBIN J. A., MAGGIO J.E., STRICHARTZ G.R. **Purification and characterization of *clorotoxina*, a chloride channel ligand from the venom of the scorpion.** *Am J Physiol Cell Physiol.* 264. 1993.
- DIEGO-GARCIA E., BATISTA C.V., GARCIA-GOMEZ B.I., LUCAS S., CANDIDO D.M., GOMEZ-LAGUNAS F., POSSANI L.D. **The Brazilian scorpion *Tityus costatus* Karsh: genes, peptides and function.** *Toxicon.*45: 273-283.
- DINIZ C. R., GOMEZ M. V., ANTONIO A., CORRADO A.P. **Chemical properties and biological activity of *Tityus serrulatus* venom.** *Mem. Inst. Butantan.*33: 435-456. 1966.
- FERREIRA L.A.F., ALVES E.W., HENRIQUES O.B. **Peptide T a novel bradykinin-potentiating peptide isolated from *Tityus serrulatus* scorpion venom.** *Toxicon.* 31: 941-947. 1993.
- FROY O., SAGIV T., POREH M., URBACH D., ZILBERBEG N., GUREVITZ M. **Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels.** *J.Mol. Evol.* 48: 187-196. 1999.
- FU Y.J., YIN L.R., LIANG A.J., ZHANG C.F., WANG W., CHAI B.F., YANG J.Y., FAN I.J. **Therapeutic potential of *clorotoxina*-like neurotoxin from the Chinese scorpion for human gliomas.** *Neuroscience Letters.* 412: 62-67. 2007.
- GAO B., TIAN C., ZHU A. **Inducible antibacterial response of scorpion venom gland.** *Peptides.* 28: 2299-2305. 2007.
- GOMEZ M.V. & DINIZ C.R. **Separation of toxin components from the brazilian scorpion *Tityus serrulatus* venom.** *Mem. Inst. Butantan.* 33: 899-902. 1966.

- HUANG X., MADAN, A. **CAP3: a DNA sequence assembly program.** *Genome Res*, 9: 868-877. 1999.
- HUYS I. & TYTGAT J. **Evidence for a function-specific mutation in the neurotoxin, parabutoxin 3.** *Eur. J. Neurosci.* 17:1786-1792 .2003.
- JIANG L., PENG L., CHEN J., ZHANG Y., XIANG X., XIONG X., LIANG S. **Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*.** *Toxicon*. 2008.
- KOSCHAK A., BUGIANESI R.M., MITTERDORFER J., KACZOROWSKI G.J., GARCIA M.L. KNAUS H.G. **Subunit composition of brain voltage-gated potassium channels determined by hongotoxin-1, a novel peptide derived from *Centruroides limbatus* venom.** *J. Biol. Chem.* 273: 2639-2644.1998.
- LOURENÇO W. R. **Scorpions of Brazil.** *Museum national d' histoire naturelle, Paris:* Les éditions de l'If, Paris-França. 306 p. 2003.
- MAGALHÃES G.S., JUNQUEIRA-DE-AZEVEDO I.L.M., LOPES-FERREIRA M., LORENZINI D.M., HO P.L., MOURA-DA-SILVA A.M. **Transcriptome analysis fo expressed sequence tags from the venom glands of the fish *Thalassophryne nattereri*.** *Biochimie.* 88: 693-699. 2006.
- MALLAT J.M., GAREY J.R., SCHULTZ J.W. **Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin.** *Mol. Phylogenet. Evol.* 31: 178-191. 2004.
- MATSUSHITA N., MIYASHITA M., SAKAI A., NAKAGAWA Y., MIYAGAWA H. **Purification and characterization of a novel short-chain insecticidal toxin with two disulfide bridges from the venom of the scorpion *Liocheles australasiae*.** *Toxicon*, 50: 861-867. 2007.
- MEKI A.M.A., NASSAR A.Y., ROCHAT H. **A bradykinin-potentiating peptide (peptide K12) isolated from the venom of Egyptian scorpion *Buthus occitanus*.** *Peptides.* 16: 1359-1365. 1995.
- MIYASHITA M., OTSUKI J., HANAI Y., NAKAGAWA Y., MIYAGAWA H. **Characterization of peptide components in the venom of the scorpion *Liocheles australasiae* (Hemiscorpiidae).** *Toxicon*, 50(3): 428-437. 2007.
- MOERMAN L., BOSTEELS S., NOPPE W., WILLEMS J., CLYNEN E., SCHOOF S. **Antibacterial and antifungal properties of alpha-helical, cationic peptides is the venom of scorpions from Southern Africa.** *Eur. J. Biochem.* 268: 1-12.2002.
- NCBI BLAST. Disponível em: <http://www.ncbi.nlm.nih.gov/blast>

- OLAMENDI-PORTUGAL T., GOMEZ-LAGUNAS F., GURROLA G.G., POSSANI L.D. **A novel structural class of K<sup>+</sup>-channel blocking toxin from the scorpion *Pandinus imperator*.** *Biochem. J.* 315:977-981.1996.
- OLAMENDI-PORTUGAL T., GOMEZ-LAGUNAS F., GURROLA G.G., POSSANI L.D. **Two similar peptides from the venom of the scorpion *Pandinus imperator*, one highly effective blocker and the other inactive on K<sup>+</sup> channels.** *Toxicon.* 36: 759-770.1998.
- PACKILA M. & GUILFOILE P.G. **Mating, male *Ixodes scapularis* express several genes including those with sequence similarity to immunoglobulin-binding proteins and metalloproteases.** *Experimental and Applied Acarology.* 27: 151-160. 2000.
- PAINE M.J., DESMOND H.P., THEAKSTON R.D., CRAMPTON J.M. **Gene expression in *Echis carinatus* (carpet viper) venom glands following milking.** *Toxicon.* 30: 379-386. 1992.
- PIMENTA A. M., STOCHKLIN R., FAVREAU P., BOUGIS P.E., MARTIN-EAUCLAIRE M.F. **Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae).** *Rapid commun Mass Spectrom.* 15: 1562-1572. 2001.
- POSSANI L.D., MERINO E., CORONA M., BOLIVAR F., BECERRIL B. **Peptides and genes coding for scorpion toxins that affect ion-channels.** *Biochimie.* 82: 861-868. 2000.
- ROCHAT H., KHARRAT R., SABATIER J.M., MANSUELLE P., CREST M., MARTIN-EAUCLAIRE M.F., SAMPIERI F., OUGHIDENI R., MABROUD K., JACQUET G., VAN RIETSCHOTEN J., EL AYEB M. **Maurotoxin, a four disulfide bridges scorpion toxin acting on K<sup>+</sup> channels.** *Toxicon.* 36:1609-1611. 1998.
- RODRÍGUEZ DE LA VEGA, R. C., & POSSANI, L. D. **Current views on scorpion toxins specific for K<sup>+</sup> channels.** *Toxicon,* 43: 865-875. 2004.
- RODRÍGUEZ DE LA VEGA, R. C., & POSSANI, L. D. **Overview of scorpion toxins specific for Na<sup>+</sup> channels and related peptides: biodiversity, structure-function relationships and evolution.** *Toxicon.* 1-14. 2005.
- SAVARIN P., ROMI-LEBRUN R., ZINN-JUSTIN S., LEBRUN B., NAKAJIMA T., GILQUIN B., MENEZ A. **Structural and functional consequences of the presence of a fourth disulfide bridge in the scorpion short toxins: solution structure of the potassium channel inhibitor HsTX1.** *Protein Sci.* 8:2672-2685.1999.
- SCHWARTZ E.F., CAMARGOS T. S., ZAMUDIO F.Z., SILVA L.P., BLOCH JR C., CAIXETA F., SCHWARTZ C.A., POSSANI L.D. **Mass spectrometry analysis, amino acid sequence and biological activity of venom components from the**

**Brazilian scorpion *Opisthacanthus cayaporum*. *Toxicon*, in press, Accepted Manuscript, Available online 4 April 2008.**

SCHWARTZ E.F., DIEGO-GARCIA E., RODRIGUEZ DE LA VEJA R.C., POSSANI L.D. **Transcriptome analysis of the venom gland of the Mexican scorpion *Hadrurus gertschi* (Arachnida: Scorpiones).** *BMC Genomics*. 8:119. 2007.

SRINIVASAN K.N., SIVARAJA V., HUYS I., SASAKI T., CHENG B., KUMAR T.K., SATO K., TYTGAT J., YU C., CHENG SAN B., RANGANATHAN S., BOWIE H.J., KINI R.M., GOPALAKRISHNAKONE P. **k-Hefutoxin1, a novel toxin from the scorpion *Heterometrus fulvipes* with a unique structure and function: importance of the functional diad in potassium channel selectivity.** *J. Biol.Chem.* 277, 30040-30047. 2002.

TAN P.T., VEERAMANI A, SRINIVASAN, K.N., RANGANATHAN S., BRUSIC V. **SCORPION2: a database for structure-function analysis of scorpion toxins.** *Toxicon*. 47: 356-363. 2006

TORRES-LARIOS A., GURROLA G.B., ZAMUDIO F.Z., POSSANI L.D. **Hadrurin, a new antimicrobial peptide from the venom of scorpion *Hadrurus aztecus*.** *Eur. J. Biochem.* 267: 5023-5031. 2000.

VALDEZ-CRUZ N.A., SEGOVIA L., CORONA M., POSSANI L.D. **Sequence analysis and phylogenetic relationship of genes encoding heterodimeric phospholipases A2 from the venom of the scorpion *Anuroctonus phaiodactylus*.** *Gene, Section Evolutionary Genomics*. 396: 149-158. 2007.

VALDIVIA, HH; KIRBY, MS; LEDERER, WJ; CORONADO, R. **Inhibition of dihidropyridine-sensitive calcium channels by the plant alkaloid ryanodine** *FEBS Letters*, 244, ( 2,) 27 :333-337. 1989.

VASCONCELOS, S.S. **Uma investigação: EST (*Expressed Sequences Tags*) podem ser usados no desenvolvimento de marcadores baseados em introns?** Dissertação de Mestrado, Universidade de Brasília. 2003.

VERANO-BRAGA T., ROCHA-RESENDE C., SILVA D.M., IANZER D., MARTIN-EAUCLAIRE M.F., BOUGIS P.E., LIMA M.E., SANTOS R.A.S., PIMENTA A.M.C. ***Tityus serrulatus* Hypotensins: A new family of peptides from scorpion venom.** *Bioch and Biophys Research Commun.* 371: 515-520. 2008.

VERDONCK F., BOSTEELS S., DESMET J., MOERMAN L., NOPPE W., WILLEMS J., TYTGAT J., VAN DER WALT., J. **A novel class of pore-forming peptides in the venom of *Parabuthus schlechteri* Purcell.** *Cimbebasia*. 16: 247-260. 2000)

- WAGNER S., CASTRO M.S., BARBOSA J.A., FONTES W., SCHWARTZ E.F., SEBEN A., RODRIGUES-PIRES O.J., SOUSA M.V., SCHWARTZ C.A. **Purification and primary structure determination of the first biotive peptide isolated from the venom of the Brazilian scorpion *T. fasciolatus*.** *Toxicon*. 41: 737-45. 2003.
- WILLEMS J., NOPPE W., MOERMAN L., VAN DER WALT J., VERDONCK F. **Cationic peptides from scorpion venom can stimulate and inhibit polymorphonuclear granulocytes.** *Toxicon*. 40: 1679-1683. 2002.
- WONG G.W., YASUDA S., MOROKAWA N., LI L., STEVENS R.L. **Mouse chromosome 17A3 contains 13 genes that encode functional tryptic-like serine proteases with distinct tissue and cell expression patterns.** *J.Biol.Chem.* 279: 2438-2452. 2004.
- YAMAJI N., DAI L., SUGASE K., ANDRIANTSIFERANA M., NAKAJIMA T., IWASHITA T. **Solution structure of IsTX-A male scorpion toxin from *Opisthacanthus madagascariensis* (Ischnuridae).** *Eur. J. Biochem.* 271:3855-3864. 2004.
- ZENG X.C., CORZO G., HAHIN R. **Scorpion Venom Peptides without Disulfide Bridges.** *IUBMB Life* 57:13-21. 2005.
- ZENG X.C., LI W.X., PENG F., ZHU Z.H. **Cloning and characterization of a novel cDNA sequence encoding the precursor of a novel venom peptide (BmKbpb) related to a Bradykinin-potentiating peptide from Chinese scorpion *Buthus martensii* Karsh.** *IUBMB Life*. 49: 207-210. 2000.
- ZENG X.C., LI W.X., WANG S.X., ZHU S. Y., LUO F. **Precursor of a novel scorpion venom peptide (BmKn1) with no disulfide bridge from *Buthus martensii* Karsh.** *IUBMB Life*. 51: 117-120. 2001.
- ZENG X.C., WANG S.X., ZHU Y., ZHU S.Y., LI W.Z. **Identification and functional characterization of novel scorpion venom peptides with no disulfide bridge from *Buthus martensii* Karsch.** *Peptides* 25: 143-150. 2004.
- ZHANG B., LIU Q., YIN W., ZHANG X., HUANG Y., LUO Y., QIU P., SU X., YU J., HU S., YAN G. **Transcriptome analysis of *Deinagkistrodon acutus* venomous gland focusing on cellular structure and functional aspects using expressed sequence tags.** *BMC Genomics*. 7: 152. 2006
- ZHIJIAN C., FENG L., YINGLIANG W., XIN M., WENXIN L. **Genetic mechanisms of scorpion venom peptide diversification.** *Toxicon*. 47: 348-355. 2006.
- ZHU S., HUYS I., DYASON K., VERDONCK F., TYTGAT J. **Evolutionary trace analysis of scorpion toxins specific for K-channels.** *Proteins* 54(2): 361-370. 2004.



ZHU, S. & TYTGAT J. **The Scorpina family of defensins: gene structure, alternative polyadenylation and fold recognition.** *CMLS celular and Molecular Life Sciences.* 61:1751-1763. 2004.

## ANEXOS

### 1. ESTs com identidade a componentes do processamento celular glandular

#### ✓ Ocy3

Seqüência de nucleotídeos com identidade de 59%, expectância de  $1^{-22}$  a proteína ribossomal de *Scorpius maurus palmatusi* (número no GenBank: AY156570).

```
GGGGGTATTAATTAAGTATAAATTTAACAAAAAATTTTTTATTTAGTAAGAATTTGAAAAATTA AAAAGTTTGAACCGCGA
GGGAAAAATAATATAATAATAGAAGAAGCAAATTTTCGTACCTTTTGTATCAGGGAGATACAATAAAATTTAAAATTAATA
TTTCCCGAATTGAATTGAGCTAATATAATTAATTAAGATGTTGCAAAATTC AAAGAAAAATTATATTAGAAATGAAATGT
TATTCGAAATGCACGATATCTGGTTC TTGAAAATAGATTTTATATCATCTATTA AAAAATAGAATAATTTTTTAGGGATAA
TCTTAATAAATATCTAAATATAAAAAGTTTTTTATAGAAGGAATAAGAGTTTCCATAGGTTTTATACTAAGTTTTATTTA
AAAATTTT TAGAAGGTTTAAAGA ACTTTACAATAAAAAATATAGAATATAATGTTTTATAAGTAGGATTATAATATAAC
TATTTTAGAAATTTAATTACTTTAATAATATTAATTTCCTTTAAGGAACTCGAACTTATTTAGCCTGTTTAACAAAACAT
GGTTTT CAGTAATAATTGAAAATCCACCTGCTCTATGCTAATGTAATGGCCGCAGTATTATGACTGTGCAAAGGTAGCAT
AATCATTAGTCTTTTAAATTTGGGACTGGAATGAACGGTTAGACTAAACAAAATTTGTCTCAAAGGAATATTCTGAATTTTAT
ATTTTAGTGA AAAAGCTTAAATTCCTTTATAGACAAGAAGACCCTTTCAAGCTTTACTGAATCAATCATAATCTTAAAA
TGAAAGTATTCTGTTTATCTGGGGCAGTTAATAATATAATTTTATTTTAAAATCGTGATGATTTATCTAGTAATATCTAAA
ATTTTAATTA
```

#### ✓ Ocy5

Seqüência de nucleotídeos com identidade de 44%, expectância de  $4^{-6}$  ao domínio da proteína THUMP de *Mus musculus* (número no GenBank: XM001478881).

```
GGGCACAAACCGAAGTGGAAGGCGATTTTAGAGATTACCACTTAGCTGTCGTTGTGGAAATTCGCAGACAGTTTGTTC
ATAGGTATCTTGAAAGACTATATTTCCATTTAAAAAATACAACCTTATCAGAACTGACGTTACCGCTACACGACTCTGCCGAA
AGCAGTAAAGTTGATAGAAAACACGGTCAACTAATCCTTTAGTTAGGGAAATTTGATGTTACCAATGAACTAGCAACAGTGACT
TCCCATGATCAAAGTAAATTAACAGTGTAGGGAAGCCACTATGCTTTGTAAATTAAGTGTA AAAAATAGTAGGAATAAATC
TATAATTACAATGAAAAA AAAAAAAAAAAAAAAAAAAAAAACCTTGTGGCCGCTCGGTCCTCAAAA
```

#### ✓ Ocy18

Seqüência de nucleotídeos com identidade de 73%, expectância de  $5^{-11}$  a gene mitocondrial de *Biphyllus lunatus voucher* (número no GenBank: DQ202525).

```
GGGGGAAGTCTGTTTCGACTTTTAAAATTTTCATGATCTGAGTTCAAATCGGCGTAAGCCAGATTGGTTTCTATCTTAAGA
ATAAAAAATGTTTTAGTACGAAAGGACCAAAACAAAAGAGAACTACTTTGAAAAA AAAAAAAAAAAAAAAAAAAAAAAA
CATGTCG
```

✓ **Ocy20**

Seqüência de nucleotídeos com identidade de 71%, expectância de 1<sup>-61</sup> à ferritina do ácaro *Ornithodoros parkeri* (número no GenBank: ABR23389).

```

1      GGG GAT ACG CAG AAG GGG TAG CTA GGT GTT TTG CAT CAA CAG TGA 45
46     TTG ATC GAG CAT CTA CTG CCA TTT ATA GAC TAT ACA TAT TTT TTT 90
91     TTT AAC TTT TTG AAA TTT GTT TGA TTT TGT TGC CAA TAT TTT GAA 135
136    AGC AAG ATG TCC CGT CCA GGA TGT AGG CCA TCT AGC CAA GTA CGA 180
        M   S   R   P   G   C   R   P   S   S   Q   V   R   13

181    CAG AAC TAT GAC CAA GAA TGT GAA ACG GCA ATC AAC AAA CAA ATC 225
14     Q   N   Y   D   Q   E   C   E   T   A   I   N   K   Q   I   28

226    AAC ATG GAG CTA TAT GCC TCT TAT GTT TAT CTA TCT ATG GCC TAC 270
29     N   M   E   L   Y   A   S   Y   V   Y   L   S   M   A   Y   43

271    CAC TTT GAC CGA AAT GAC GTT GCA TTA CGC GGA TTT CAT AAA TTT 315
44     H   F   D   R   N   D   V   A   L   R   G   F   H   K   F   58

316    TTC AAA AAG ATG AGT AAC GAG GAA AGA GAG CAC GCA GAC AAG TTG 360
59     F   K   K   M   S   N   E   E   R   E   H   A   D   K   L   73

361    CTG AAA TTC CAA AAC CAA AGA GGT GGC CGA GTT CTT TTT GAA GAT 405
74     L   K   F   Q   N   Q   R   G   G   R   V   L   F   E   D   88

406    ATC GAG CGG CCT GAG AAA GAT GAC TGG GGA TGC GGA TTG GAC GCA 450
89     I   E   R   P   E   K   D   D   W   G   C   G   L   D   A   103

451    ATG CAA GCA GCC CTG ACG CTG GAG AAA AAT GTG AAT GAA TCC TTA 495
104    M   Q   A   A   L   T   L   E   K   N   V   N   E   S   L   118

496    CTG AAA CTG CAT AAT CTG GCA ACT GAT AAA GGA GAT CCC CAG TTG 540
119    L   K   L   H   N   L   A   T   D   K   G   D   P   Q   L   133

541    TGC GAC TTT CTC GAG AGC CAT TAC CTC ACC GAA CAA GTG GAA TCA 585
134    C   D   F   L   E   S   H   Y   L   T   E   Q   V   E   S   148

586    ATA AAA GAA TTT GGG AGA CTA CAT CAC TTA CCC TTA ACG GGT TGG 630
149    I   K   E   F   G   R   L   H   H   L   P   L   T   G   W   163

631    ACA AGG CCT TGG AGA ATA TAT GTT TGA CCACGAACTCTTGACAGTTAAAC
164    T   R   P   W   R   I   Y   V   *
      GTTCCATTAATATGCTTCACTGTTTCCACATATATAAGAAGGGTTCCTGTTAATTAAC
      GAATCTAACCTG      753
  
```

**Alinhamento de Ocy20 com a seqüência da ferritina (ABR23389)**

```

Ocy20      MSRPGCRPSSQVRQNYDQECETAINKQINMELYASYVYLSMAYHFDRNDVALRGFHKFFK 100%
ABR23389   MTQP-----RQNYHTECEARINKQINMELYASYVYLSMAYYFDRDDVALRGFHKFFK 71%

Ocy20      KMSNEEREHADKLLKLFQNRGRVLFEDIERPEKDDWGCGLDAMQAALTLEKNVNESLLK
ABR23389   KSSHEEKEHAEKLMKYQNRGRVVLQPIQKPSQDEWGTGLEAMQASLELEKTVNQALLD

Ocy20      LHNLATDKGDPQLCDFLESHYLTEQVESIKEFG-RLHHLPLTGWTRPWRIYV-----
ABR23389   LHRIATDHNDAQLCDFLESEYLKEQVEAIKELSDYITNLKRVGPGLGEYMFDKETLDGEDS
  
```

✓ **Ocy21**

Apresentou 63% de identidade a hemocianina, subunidade F de *Nephila inaurata madagascariensis* (número no GeneBank:CAD68056)

Alinhamento de Ocy21 com Hemocianina (CAD68056)

```
Ocy21          MGVDKQKRL LPLFKHLTSLTTEQLPVDERDPRLKDVGV LQRGKLFSCFHEDHLL EAEKL 60
CAD68056       MPAQDKQRR I LPLFEHLASL TRAVLPPEERDPRLKRLGR LPRGTLFSCFHTEHL IEAEEL 60

Ocy21          FTVLFQAKDFDDLIQLCQQARDIVNEGLFVFAVSVAVLHREDCKGVTVPPIQEIFPDRFV 120
CAD68056       FETLYAAKDFEDFMRLAEQARDIVNEGMFVYSTSVALLHRDDCRGVTVPPIQEIFPDRFI 120

Ocy21          PAETINQAQKFD RQRANDDPVVVKIQETGNILDPEYHLAYFREDIETTP-----TIGTG 174
CAD68056       PSETINQAIKADLRRADDSPIVDVIQTGNILDPEYKLAYFREDIGANAHHWHWHIVYPA 180

Ocy21          TWSTR----- 179
CAD68056       TWRPEIMGKVKDRKGELFYMHQQMCARYDCDRLSTGLQRMIPYHNFNEKLEGYSPLTS 240
```

	GGG GGA CAG AGG CCA CAC TAG CTG GGC GCT GAG TTC ACA TCT GTT	45
46	ACG GGA TTT GTA GAC CTA TAA GCC ATG GGA GTT CAG GAC AAG CAG M G V Q D K Q	90 7
91	AAG CGT CTC TTA CCC CTT TTC AAA CAC TTG ACA TCT CTT ACT ACC K R L L P L F K H L T S L T T	135 22
136	GAA CAG TTA CCA GTC GAT GAG CGT GAC CCT AGA TTG AAG GAT GTT E Q L P V D E R D P R L K D V	180 37
181	GGT GTC CTT CAG AGA GGC AAG CTC TTC TCC TGT TTT CAC GAA GAC G V L Q R G K L F S C F H E D	225 52
226	CAT CTT CTG GAA GCC GAG AAA CTT TTC ACT GTC CTT TTC CAG GCT H L L E A E K L F T V L F Q A	270 67
271	AAG GAT TTC GAT GAT TTG ATC CAA CTG TGC CAA CAA GCC CGT GAC K D F D D L I Q L C Q Q A R D	315 82
316	ATT GTG AAC GAA GGT CTG TTT GTA TTC GCT GTC TCT GTT GCT GTC I V N E G L F V F A V S V A V	360 97
361	TTG CAT CGT GAA GAC TGC AAA GGA GTC ACT GTT CCA CCT ATC CAA L H R E D C K G V T V P P I Q	405 112
406	GAA ATA TTC CCC GAT AGG TTC GTC CCC GCC GAA ACC ATC AAC CAA E I F P D R F V P A E T I N Q	450 127
451	GCC CAG AAA TTT GAT CGC CAA AGA GCC AAT GAC GAC CCA GTA GTT A Q K F D R Q R A N D D P V V	495 142
496	GTC AAG ATT CAA GAG ACC GGT AAC ATT TTG GAC CCA GAA TAC CAT V K I Q E T G N I L D P E Y H	540 157
541	CTA GCT TAC TTC CGT GAG GAC ATC GAA ACC ACG CCC ACC ATT GGC L A Y F R E D I E T T P T I G	585 172
586	ACT GGC ACT TGG TCT ACC CGC TAA ATG GAG AAC AGA TGT CAT GGG T G T W S T R *	630
631	AAA GCC AAA GAC AGA NNA GGA GAA CTA TTT TAC TAC ATG CAT CAG	675
676	CAA ATG TGT GGC AGA TAC GAT TGC GAT AGG CTG TCA TCA GGG TTG	720
721	AAA CGT ATG ACT CCA TTC CNA CAC TTC GAC GAA	753

✓ **Ocy22**

Apresentou 56% de identidade (*e-value*:  $3^{-31}$ ) a proteína ativada por cálcio de *Apis mellifera* (número no GenBank: XP392192).

3	GGA GAA GTA GGG TTG GCA TAT ATA AAT TTT GCC CGA AAT AGT ATG	47
	M 0	
48	GAA CAA CTA CGT CAG AAA GTC ACA GAT GAA TTA TGG GTT TTG AGC	92
1	E Q L R Q K V T D E L W V L S	15
93	CTG TTT GAA TGG TAT GCA TCT CAG ATG CAG CTA CTT TGT ACC TGG	137
16	L F E W Y A S Q M Q L L C T W	30
138	CTC TCA GAG CGG ATT GAT CGA GGT CTG CAT TCA TTT CAA GTA TCA	182
31	L S E R I D R G L H S F Q V S	45
183	GCT TTA TCT CAT ATG GGA AAG AAG CTT TAT TCT GAT TTT GAG CTA	227
46	A L S H M G K K L Y S D F E L	60
228	CAA GGT ATC GAA GAA GAG AAG TTA AAT TCA AAA ATT TAT CAA ACA	272
61	Q G I E E E K L N S K I Y Q T	75
273	ATT AGT AGT CGC CTG CAG TTG GAA AAT GCT ACA GTG TCT GTA TCA	317
76	I S S R L Q L E N A T V S V S	90
318	GAG GCT GGG CAT CGA GAA AAA GAG GAT TCT GAA GAG AGG ATA CCC	362
91	E A G H R E K E D S E E R I P	105
363	TAG TTG ACA AGG GAG TCA CGC TTG TCA GGA AGT TTT CCA GAG GAT	407
106	*	
408	GGG TAT AGT AGC CCA GTG TCT TCA GGC AGT GGT GTT GTG TTA GGT	452
453	TGC CAG TTG TCA TTG TGA TAT ACC AGC AGA TGG TAA GCC AAA TCA	497
498	TCT GGT AAT ACT TTC TGT GTG ATC TAA ACA AAC TTT AAG CAG AAA	542
543	GAA ACA GAA GGT TGT ATA TAC ACC AAG ACA GCA GTG AGA ACT TTT	587
588	GTT TAA GTG ATC TAA TCA TTT GTT GAG TTT TAT TTA TGT GTA TTT	632
633	GGG NTG CTT TAT ATC AGC TTC CTG TGT GAG TGG ATG TAT GTA TCG	677
678	TTA GTT AGA GAA AGG AAT AGN ATA CAT GCT ACT TAT TCC TTT TTG	722
723	AAT ATG CAG CAT CTC TTT GGG AAA TTG GGT GCT GGT TAT TAA AAT	767
768	TAA ATT CTC GGA TAA TTG TAA 788	

### ✓ Ocy23

Apresentou identidade de 55% ( $e\text{-value}: 1^{-19}$ ) à proteína similar à prefoldina de *Canis familiaris* (número no Gene Bank: XP848520)

3	GGG ATG ACG TTT AAA TGT ATT GCA CTA GAA ATT TTG TGG TGA ATT	47
48	AAA TGC GAT ATC AGT AAC GGT AGT AAT TTA GTA ATT TTC GCG TAA	92
93	CGT GGT TCA ACG TCT ATT CTA ATG GCA GCG AAT ATG CCT GTC GAT	137
	M A A N M P V D	7
138	ATG GAA TTG AAA AAG GCA TTT AAA GAA TTA CAG CAA CAA ATG ATA	182
8	M E L K K A F K E L Q Q Q M I	22
183	GAA ACA AGT CAG AAG CTG AAA ATG TCA GAT ATA CAA ATA GAA AAC	227
23	E T S Q K L K M S D I Q I E N	37
228	TTG AAG CGG ACT ATA CAC CAT TCG AAA TTA ACC GAG AGT GAA GTT	272
38	L K R T I H H S K L T E S E V	52
273	TCA ACA ATG CCT GAA GAT ACA AGA ATG TAC ATG GGT GTC GGA AGG	317
53	S T M P E D T R M Y M G V G R	67
318	ATG TTT TTG TTG ACC GAT ATA TAT TAC ATA TAT TCT GGT CTT TAA	362
68	M F L L T D I Y Y I Y S G L *	
363	GAA TTA GTT AAT GCT TCA TAA TAT ATT ATT TAA ATT TAT AGT GTA	407
408	ATA CAG TTT TCC TCT TTA GTT ATT TGT AAA ATA TTT GTA AAC ATT	452
453	ATT ATA GAT TTG TAT TTT CCA AAT ATG TAA ATC CTT ATT CAT TGC	497
498	TAT CTT CAT TAT ATT GTC ATT ATT ATG CTT TTT TTA TGT TAT CTA	542
543	ATT ATA ATA TTT TCT ATT CTT CAT AAT CCT TTT TCT TTA TTT AAT	587
588	TAT CAT TTT TTT TAC ATC TTT ATA TAT ACT TAT CTT CTT ATT TAT	632
633	AAT ATT CTA TAT TTA TAA ACT ATC CTA ATC TCC TTT TAT GTT TGT	677
678	TTA TAA TGT TTA GTT TTA ACC TCT CTT ATC TTG AAN CAT TAT ATT	722
723	CCT ATT TAT TTA TTA CTT TTT CTA TGT ATA TCT TTC TCT ACT TAA	767
768	TCA TTT CTA CCT CAT CAT GAC TTA TCA CTT ATT TTA TAC ACC TAC	812
813	CGT TAT CTC TAT ACT ATA TAA TCT	836

### Alinhamento de Ocy23 e proteína similar à prefoldina (XP848520)

XP848520	MAA--PVDLELKKAFTELQAKVIDTQQQVKLADIQIEQLNRTKKHAHLTDTEIMTLVDET	58
Ocy23	MAANMPVDMELKKAFKELQQQMIETSQKLMKSDIQIENLKRTHHSLKLTSESEVSTMPEDT	60
XP848520	NMYEGVGRMFILQSKEVIHSQLEKQKIAEEKIKELEQKSYLERSVKEAEDNIREMLMA	118
Ocy23	RMVMGVGRMFLLDIYYIYSGL-----	82
XP848520	RRAQ 122	
Ocy23	----	

### ✓ Ocy25

Apresentou identidade de 70% a proteína tipo tsunagi de *Apis mellifera* (GeneBank: XP395245).

2	GGA	TGC	GCA	TTA	AGT	ATA	GGC	GAG	GAA	CGT	ATT	TTG	TTT	GCG	CTG	46
47	TGG	ATC	GCG	TTG	ATT	TGT	GCA	GTG	TTG	GAA	CGT	TTC	ATT	ATT	TTT	91
92	AGT	AAC	TTG	TAC	TTT	ACG	AGA	AGT	TAA	CAG	ACT	ATG	GCA	GAT	ATT	136
												M	A	D	I	4
137	CTG	GAT	ATA	GAC	GGT	TTA	GTT	GAA	GAA	TTT	GAA	GTT	GAC	GAA	GAA	181
5	L	D	I	D	G	L	V	E	E	F	E	V	D	E	E	19
182	GGT	GAT	CAT	GGT	GTA	CAG	AAA	TTG	AAG	GAA	AAA	GCC	AAG	AAA	CGT	226
20	G	D	H	G	V	Q	K	L	K	E	K	A	K	K	R	34
227	AAA	GGC	AGG	GGA	TTT	GGA	ACA	GGA	ACA	GCA	AGG	GAA	CAT	ATT	GGA	271
35	K	G	R	G	F	G	T	G	T	A	R	E	H	I	G	49
272	GAA	TAT	GGG	GCA	GTG	GAG	TCT	GAA	AGT	AGT	GAT	CAA	CCT	GGG	CCA	316
50	E	Y	G	A	V	E	S	E	S	S	D	Q	P	G	P	64
317	CAG	AGA	TCT	GTG	GAA	GGA	TGG	ATT	TTA	TTC	ATA	ACA	GGA	GTG	CAT	361
65	Q	R	S	V	E	G	W	I	L	F	I	T	G	V	H	79
362	GAA	GAA	TCA	CAA	GAG	GAT	GAC	ATA	TAC	GAC	AAG	TTT	GCT	GAG	TTT	406
80	E	E	S	Q	E	D	D	I	Y	D	K	F	A	E	F	94
407	GGC	GAA	ATA	AAA	AAT	ATA	CAT	TTA	AAC	CTT	GAC	CGA	AGA	ACT	GGA	451
95	G	E	I	K	N	I	H	L	N	L	D	R	R	T	G	109
452	TTT	CTG	AAG	GGA	TAT	GCT	TTA	GTA	GAA	TAT	GAA	ACC	TTT	TAT	GAG	496
110	F	L	K	G	Y	A	L	V	E	Y	E	T	F	Y	E	124
497	GCA	CAA	GCT	GCA	ATG	GAA	GCT	TTA	AAC	GGC	AGC	GAA	ATT	CTT	GGC	541
125	A	Q	A	A	M	E	A	L	N	G	S	E	I	L	G	139
542	CAG	AAT	ATA	AGT	GTT	GAT	TGG	GCT	TTC	GTA	AAA	GGT	CCC	CTG	AAA	586
140	Q	N	I	S	V	D	W	A	F	V	K	G	P	L	K	154
587	AAA	AGT	CAC	AGT	ACA	GGA	GCC	GAT	CAT	TTG	AGA	GAG	GAG	ACG	GTG	631
155	K	S	H	S	T	G	A	D	H	L	R	E	E	T	V	169
632	AAG	AGG	TTG	TGT	CGT	CTG	CAA	ACA	AAA	TTT	GTG	ATG	TTT	TTG	CTT	676
170	K	R	L	C	R	L	Q	T	K	F	V	M	F	L	L	184
677	CAA	TCG	TGT	AGT	TTT	AAG	CTG	TTA	GAA	TTT	CGA	AAC	AAA	AAA	AAA	721
185	Q	S	C	S	F	K	L	L	E	F	R	N	K	K	K	199
722	AAA	AAA	ATA	AAC	CTG	CCG	CCG	CTC	GTC	CCG	AAA	GAT	TCT	AAC	ATC	766
200	K	K	I	N	L	P	P	L	V	P	K	D	S	N	I	214
767	TTT	GGC	CCG	GGT	CAA	AAG	TAA	TGA	CGG	TCT	AAT	GTT	CCA	AGA	ACC	811
215	F	G	P	G	Q	K	*									
812	TGG	CGG	CAT	GGT	TGA	TTC	CCA		832							



Alinhamento de Ocy25 e proteína do tipo tsunagi de *Apis mellifera* (XP395245).

```

XP395245      MADVLDIDN-AEEFEVDDDEDDQGIARLKEKAKRRKGRGHGAEVSTRDDVGHYESMNVVE 59
Ocy25         MADILDIDGLVEEFEVDEEGDHGVQKLKEKAKRRKGRGFTG--TAREHIGEYG--AVES 56

XP395245      EDDDEPGPQRSVEGWILFINSVHEEAQEDDIQDKFSEFGQIKNLHLNLDLDRRTGFLKGYAL 119
Ocy25         ESSDQPGPQRSVEGWILFITGVHEESQEDDIYDKFAEFGELKNIHLNLDLDRRTGFLKGYAL 116

XP395245      VEYETFKEAQAAKDALNGTEILGQPIISVDWCFVKGPKKIKKRS---HRRR----- 166
Ocy25         VEYETFYEAQAAMEALNGSEILGQNISVDWAFVKGPLKKSHTGADHLREETVKRLCRLQ 176

XP395245      -----
Ocy25         TKFVMFLLQSCSFKLLLEFRNKKKKKINLPPLVPKDSNIFGPGQK 220
    
```

✓ **Ocy26**

Identidade 44%, expectância de 1<sup>-25</sup> com NADH desidrogenase de *Centruroides limpidus* (GenBank: YP214898).

2	GGC	AGG	GGA	AAT	TGA	ATT	TAT	TTT	AAT	CTC	CAA	AAT	TAA	TGT	TAT	46
47	GTA	TTA	ACT	ATC	TCC	TGG	ATT	TTT	GGA	GCT	TTA	ATA	ATA	AGA	ACT	91
92	TTT	GGA	TTT	TTA	GTT	TCT	TTT	CAT	CCA	TTG	GTT	ATG	GGA	GTT	TTT	136
												M	G	V	F	4
137	TTG	ATA	GTT	GGG	ACA	GGA	TTG	ACA	GCT	TTT	AGA	ACG	TAT	TTT	ATT	181
5	L	I	V	G	T	G	L	T	A	F	R	T	Y	F	I	19
182	TTT	GGA	TAT	TCT	TGG	TAT	TCT	TAT	ATT	ATA	GTT	TTA	GTG	TTT	TTA	226
20	F	G	Y	S	W	Y	S	Y	I	I	V	L	V	F	L	34
227	GGT	GGA	ATG	TTA	ATT	TTA	TTT	ATT	TAT	ATT	GCA	AGT	TTA	GCT	TCA	271
35	G	G	M	L	I	L	F	I	Y	I	A	S	L	A	S	49
272	AAT	GAG	TTT	GTA	AAA	TTG	AAA	AGC	TTT	TTT	TGG	TTT	TGG	GGA	TTT	316
50	N	E	F	V	K	L	K	S	F	F	W	F	W	G	F	64
317	TTA	TTA	TTT	ATT	GTT	CCT	ATT	AGT	GAT	TGA	GTT	TTG	TTG	AAG	GAT	361
65	L	L	F	I	V	P	I	S	D	*						
362	GAA	AAT	TCT	GTA	GTT	AAG	GTA	TAT	GAG	AAT	TTT	CCT	ACT	TCT	ATT	406
407	GTT	ACT	TTG	TTT	CTT	GCT	TCT	TAT	TTG	TTG	TTT	ACT	TTT	GTT	GGG	451
452	ATT	ATT	AAG	CTT	GTG	GAC	ATT	GAA	ATG	GGG	CCT	TTA	AGG	GAT	TTT	496
497	CGC	AAA	ANN	AAA	NNN	NAN	NAN	NNA	AAN	NAA	CAT				529	

Alinhamento de Ocy26 e NADH desidrogenase (YP214898). Em azul o peptídeo sinal predito.

```

YP214898      MFGVLVLLFVFFFFSTHPLVMGLILVVLSTVVFVSEHLLFFVLGVGWFSYVFFLVFLGGVLVL 60
Ocy26         -----MGVFLIVGTGLTAFRTYFIFGYSWYSYIIIVLVFLGGMLIL 40

YP214898      FIYVASLAANEPAASLEPKYVLSFLVFFVVFAGSWLFCYQEFFIWRFYLFPPFSLVIIIFLGA 120
Ocy26         FIYIASLASNEFVKLK-----SF-----FWFWGFLLEFI----- 68

YP214898      YLLMALVAVASICKVKEGPLREMV 144
Ocy26         -----VPISD----- 73
    
```

✓ **Ocy27**

Identidade de 31%, *e-value* de  $2^{-12}$  com elastase de *Macaca mullata* (XP001083099).

1	ATG	ACT	CTT	CTA	CTC	GTA	CTT	GGG	TTA	GCA	GTG	TTG	GCA	TGT	GTC	45
1	M	T	L	L	L	V	L	G	L	A	V	L	A	C	V	15
46	ATA	TTA	CCG	CAT	GGA	CTT	CAT	GCA	CGA	TAC	CTT	GAC	GAT	CAA	GAT	90
16	I	L	P	H	G	L	H	A	R	Y	L	D	D	Q	D	30
91	GCC	GAA	AGC	GGC	AGA	GTT	TTC	CGC	GGA	AGA	TTT	GCT	AAT	CAG	CAA	135
31	A	E	S	G	R	V	F	R	G	R	F	A	N	Q	Q	45
136	GAA	TTT	CCG	TGG	ATG	GTT	CAT	TTG	CAA	ATA	TCT	AAA	GGA	AAT	AAC	180
46	E	F	P	W	M	V	H	L	Q	I	S	K	G	N	N	60
181	ATG	GCC	TCT	ATT	TGT	GGA	GGT	TCT	ATC	ATT	TCG	AAA	AAT	TAC	GTA	225
61	M	A	S	I	C	G	G	S	I	I	S	K	N	Y	V	75
226	CTA	ACT	GCC	GCA	CAT	TGT	GTC	TGC	TCC	AAT	GCA	ACT	AAG	AAA	ACA	270
76	L	T	A	A	H	C	V	C	S	N	A	T	K	K	T	90
271	TAC	GCA	GAC	GTG	AGT	GCC	ATT	ACG	GGA	AGA	ATA	GGG	CAT	ATC	AAC	315
91	Y	A	D	V	S	A	I	T	G	R	I	G	H	I	N	105
316	AGG	TCT	GAT	GCC	ATC	GTA	GTG	AAA	TTC	AGT	CAA	AGA	ATA	GTA	CAC	360
106	R	S	D	A	I	V	V	K	F	S	Q	R	I	V	H	120
361	GAG	AGT	TAT	GAC	GAA	GAT	TAC	AAT	GCA	GAC	ATC	GCT	TTA	CTG	AAG	405
121	E	S	Y	D	E	D	Y	N	A	D	I	A	L	L	K	135
406	TTC	GAA	ACG	AGT	TTG	ACA	AAC	TAT	AAC	AAG	AGC	ATT	GAT	AGA	ATA	450
136	F	E	T	S	L	T	N	Y	N	K	S	I	D	R	I	150
451	TGT	CTT	GCT	GAC	AAA	GGA	AAG	AAA	TTT	CCA	AAT	AAG	CAG	GCC	GTA	495
151	C	L	A	D	K	G	K	K	F	P	N	K	Q	A	V	165
496	ATT	CAA	ATG	GGA	TGG	GGA	CGA	TTT	GAT	ATG	GTA	CTA	AGG	GAA	ACA	540
166	I	Q	M	G	W	G	R	F	D	M	V	L	R	E	T	180
541	GCC	CTA	CCC	TTA	ACA	CTG	CCG	TGG	TTA	CTT	ATT	GGA	TAA	TCC	AGA	585
181	A	L	P	L	T	L	P	W	L	L	I	G	*			
586	CTG	CGA	AAA	ACG	AAT	GGA	ATC	ATA	CGC	TGA	ACC	TGG	ACA	TTG	TGT	630
631	GTC	AGT	AAT	TTG	AGT	GGA	GAG	AAA	ATC	TGT	GGG	GTG	ACT	CTG	GAG	675
676	CCC	TTT	AAT	GTC	ATG	AAT	GGT	GGC	AGA	TAA	705					

## Alinhamento de Ocy27 com Elastase de *Macaca mullata* (XP001083099)

```

XP001083099      --MITTLLLSALVAGALSCGVPT-----YPPSVTRVVGEEATPNSWPWQVSLQYTSNGK  53
Ocy27            MTL LLVLGLAVLACVILPHGLHARYLDDQDAESGRVFRGRFANQQEFPMVHLQISKGNN  60

XP001083099      WYHTCGGSLITKNWVLTAAHCISSSRTYRVVLGRHNLN-----IAESGSLAVSVSKTVVH  108
Ocy27            MASICGGSIISKNYVLTAAHCVCSNATKKTYADVSAITGRIGHINRSDAIVVKFSQRIVH  120

XP001083099      QNWN SNQVSKGYDIALVKLANPVS-LTDKIQLACLPPAGTILPNNYPCYVTGWGRL-RT  165
Ocy27            ESYDEDYN---ADIALLKFFETSLTNYNKSIDRICLADKGGKFFPNKQAVIQMGWGRFYGT  177

XP001083099      NGPLPDV LQQGRL LVVDYATCSSSGWGSTVKTSMICAGGDGKISSCNGDSGGPLNCQAA  225
Ocy27            KGNSPT---HCRGYLLDNPDC EKR--MESYAEPGMLCVSNLSGE-NLWGDS-S-FNVMNG  235

XP001083099      DGRWEVHGIGSLT SVLGCNYYMPSIFTRVSNYNDWINSVIANN  269
Ocy27            GR----- 237
    
```

✓ **Ocy28**

Apresentou identidade de 96%, expectância de  $3^{-114}$  a actina de *R. microplus* (AAP81256).

1	GGG	GGA	GAT	GGT	GTT	TCC	CAT	ACT	GTA	CCA	ATC	TAC	GAA	GGC	TAC	45
46	GCC	CTC	CCC	CAC	GCC	ATT	CTC	CGT	CTA	GAT	TTG	GCT	GGA	CGA	GAT	90
91	TTA	ACT	GAC	TAC	TTA	ATG	AAG	ATC	CTT	ACT	GAG	AGA	GGC	TAT	TCT	135
						M	K	I	L	T	E	R	G	Y	S	10
136	TTC	ACA	ACT	ACA	GCC	GAG	AGA	GAG	ATT	GTG	CGT	GAT	ATC	AAG	GAG	180
11	F	T	T	T	A	E	R	E	I	V	R	D	I	K	E	25
181	AAA	CTA	TGC	TAC	GTT	GCT	CTG	GAT	TTC	GAA	CAG	GAG	ATG	GCA	GCA	225
26	K	L	C	Y	V	A	L	D	F	E	Q	E	M	A	A	40
226	GCC	GCT	TCC	TCT	TCA	TCT	TTG	GAA	AAG	TCT	TAC	GAA	TTG	CCT	GAC	270
41	A	A	S	S	S	S	L	E	K	S	Y	E	L	P	D	55
271	GGA	CAA	GTT	ATT	ACC	ATT	GGA	AAC	GAA	AGG	TTC	AGG	TGC	CCA	GAA	315
56	G	Q	V	I	T	I	G	N	E	R	F	R	C	P	E	70
316	GCA	TTG	TTC	CAG	CCA	TCT	TTC	TTG	GGT	ATG	GAG	TCC	TGT	GGA	ATC	360
71	A	L	F	Q	P	S	F	L	G	M	E	S	C	G	I	85
361	CAC	GAG	ACC	ACC	TAC	AAC	TCA	ATT	ATG	AAG	TGC	GAT	GTC	GAC	ATC	405
86	H	E	T	T	Y	N	S	I	M	K	C	D	V	D	I	100
406	CGT	AAA	GAC	TTG	TAT	GCC	AAT	ACC	GTT	CTG	TCA	GGA	GGT	ACC	ACC	450
101	R	K	D	L	Y	A	N	T	V	L	S	G	G	T	T	115
451	ATG	TAC	CCC	GGA	ATT	GCC	GAC	CGT	ATG	CAG	AAA	GAA	ATC	ACC	GCT	495
116	M	Y	P	G	I	A	D	R	M	Q	K	E	I	T	A	130
496	TTA	GCC	CCA	AGT	ACA	ATG	AAG	ATC	AAG	ATC	ATT	GCA	CCA	CCA	GAA	540
131	L	A	P	S	T	M	K	I	K	I	I	A	P	P	E	145
541	AGG	AAG	TAC	TCC	GTC	TGG	ATC	GGA	GGT	TCC	ATC	CTG	GCA	TCA	CTT	585
146	R	K	Y	S	V	W	I	G	G	S	I	L	A	S	L	160
586	TCT	ACT	TTT	CCA	CAG	ATG	TGG	ATT	TCC	CAA	GCA	GAA	TAC	GAA	AAA	630
161	S	T	F	P	Q	M	W	I	S	Q	A	E	Y	E	K	175
631	TCT	GTT	CCT	AGT	ATT	GTG	CAC	CGC	CAG	TGT	TTC	TAA	GGN	ATA	TAC	675
176	S	V	P	S	I	V	H	R	Q	C	F	*				
676	TTT	TTC	ATC	TAG	ACT	TTA	AAA	ACT	GGC	AGT	ATG	ATT	TTG	GGC	GGG	720
721	CTG	TTT	CAC	GAG	AAC	CAG	AGA	GTG	TAT	TTT	CAC	CCT	AAA	AGC	CCC	765
766	GAG	GGA	CGT	TTA	AGT	TTC	CAA	TTT	GGA	GCT	TGA	AGA	ATG	AAA	TTG	810
811	CGG	GCC	CTG	CTC	TAA	TGA		828								

## Alinhamento de Ocy28 com actina de *R.microplus*

```

AAP81256      MCDDEVAALVVDNGSGMCKAGFAGDDAPRAVFPSPVIGRPRHQGVMVGMGQKDSYVGDEAQ 60
Ocy28      -----

AAP81256      SKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKM 120
Ocy28      -----

AAP81256      TQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVLDSGDGVSHTVPIYEGYALPHAILRLD 180
Ocy28      -----

AAP81256      LAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKS 240
Ocy28      -----MKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMAAAASSSSLEKS 50
*****:*****

AAP81256      YELPDGQVITIGNERFRCPEALFQPSFPGMESCGIHETTYNSIMKCDVDIRKDLIANTVL 300
Ocy28      YELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTYNSIMKCDVDIRKDLIANTVL 110

AAP81256      SGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFPQMWISK 360
Ocy28      SGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFPQMWISQ 170

AAP81256      QEYDESGPSIVHRKCF 376
Ocy28      AEYEKSVPSIVHRQCF 186

```

### ✓ Ocy30

Identidade de 38%, expectância de  $3^{-32}$  com sulfotransferase (GeneBank: DQ066225).

```

AAAACAGTGACGTTAACGTACCTTCACATCGCCATCAGTTGAACGTACTIONACAATGAGCGAAGAAAAGTTATATGATGACA
AGTGCAGGGAATGCCTATTCCAAATGGGTTTCCAGCAGAAGAATCAAAGAAGCCATGAAATATAAGCCAAGACCAGGAG
AGATATTCATTGTTGCTTATCCTAGAGTGGTACAACATGGATGCAGTACATAGTCTTGCTGATTCTTCGACGTGGGAAGC
TATCTGAAGATCCAATGGACTTCTATGTGGGATGTCGGTACTTAGAAATGACAGGATCTGAAGCAACCGATTATCTGGCCG
CTTCTGGCGTTGCAAGAAGCCACTTACCCTTACCTGATGCCGATGCTAAGGATGCTAAATACATCTACGTTTTCAGAA
ACCCGGAGATTGTCCGTATCTTACTATAAACAACCTGAACGAATCCAGTTTGGATGACCCCGGTGATAGCAAATTTTATG
AATCTTCGAGAGCTTCATGGCAGGTGAAGTGCCCTTATGGTGATTTATTTGATCACGTTGCTGGCTGGTATGAGCGTAAGG
ACGACCCGAATGGTTATATTGTACATACGAAAAGTCTCAGGAAGATACAAAGGGGAGATCCTCANNNATAGCAAATTTN
ATGGGGATGACCATGAGAATG

```

### ✓ Ocy31

Identidade de 82%, *e-value* de  $1^{-34}$  a gene ribossomal de *Pandinus imperator* (GeneBank: AY210830).

```

CTGAATTTAAGCATATAAGTAAGCGGAGGAAGAGAAAAGCAACAGCGATTCCCCTAGTAACGGCGAGCGAA
CGGGGAAAAGCCCAGCGCTGATTCCCTTGTCCTCTGGGTAACGGGAAATGTGGCGTTTGGCAGGTCGTATC
CCGGCGGACGGACGGTGGTCAAGTCCCCCTGACCGGGGCTCTGATCCTGAGCGTTTGGCTAGGCCCTTAGA
TGCCGCGTCAGGCTCTGCCGATTCGACCTCTCTGGAGTCTGTTTGTCTGTTTGTGCTTCCCTCATTCTTCGC
CCTTTTTTTGCTTTTACTGTCAATAGGTTTTAGACTTGAATCTATCACATTCATTTACTATCTATTTTCATAA
TTTTGCTTTTGTAGTTGTGATTTCATATCTTAATTTTCAATTCAATTTAATAATATTCTGTTTTAGTTATCTT
TGTTTATTATTCACATTTAAATATCTTATGAATATATTATCAAATATATTATGTAATAAGTTAATTATTTTG
TTATTTTCTCAACCTATTACACCCTTGCTGTCAATATTACCCTTATATTTTACTTCTCTCTATAACATCTCTGT
GAATTTTTGTTATCTTAATATTTTCTTATCACTAATTGTATTATCTTTTATATTTTTTAATA

```

✓ **Ocy32**

Apresentou identidade de 28% a Serine Protease de *Mus musculus* (GenBank: YP437990).

2	GGG GAG CTT TTA TGT AAG CAA ATG TTC CCG AAT TCA CAC AGT TTC	46
	<i>M F P N S H S F</i>	8
47	ACG ATG AGG AGA TGT AGC ATA TTC TTC GTT TTT CCC GCT CTG TTG	91
9	<i>T M R R C S I F F V F P A L L</i>	23
92	ATC CCT GCT CAT CTT TCT CCA GTT GGA GAA GAG AGC AGG ATA TTT	136
24	<i>I P A H L S P V G E E S R I F</i>	38
137	AGA GGA AGA GAG GCA CAA GCT GGT GAA TTT CCT TGG ATG GTG TTC	181
39	<i>R G R E A Q A G E F P W M V F</i>	53
182	GTA AGG CTG ACT GAT GAA CAG AAC TGC GGC GGT TCC ATT ATT TCA	226
54	<i>V R L T D E Q N C G G S I I S</i>	68
227	CCC TCG TAT ATG TTG ACG GCA GCT CAC TGC GTG AAA GGA AGG TCT	271
69	<i>P S Y M L T A A H C V K G R S</i>	83
272	TTG ACT GAT TTG AGA GCA GTC GTT GGC TCA GTC GAT CGG GAA CAA	316
84	<i>L T D L R A V V G S V D R E Q</i>	98
317	GCA CCT ATG TTA CAG CTT GAA GAG TAT CTG ATA CAC CCA TAA TAC	361
99	<i>A P M L Q L E E Y L I H P *</i>	
362	GGT CGG GTA AGA AGA CAC CTT GGA GAT TTG GCA CTG CTC AAA TTT	406
407	AAT AGA CCA CTC TAA TTT ACA GAT TTT ATT AAG CCC TTT TGC ATA	451
452	GGG AAA ATC GAT TCT TTC GTT CTC CCT GGA AAT GAA GTG TTG CAA	496
497	ATG GGT AGG GGT AGA GAT CGT CAT GAC TGC GCA TTT GTA TCA AGA	541
542	AAC TAA AAG TGA CAG AGT AGG AAG GTT ATT TCT GAA ATT TAT GGC	586
587	AAA ACC TTC TTG AAC ACT ACA TTT ACG AAT GCT TAT TAT AGG TAG	631
632	ATT TGG GTT AGA ATA CGA ATT AAG GGA GTG TGG GAG TTA ATC TGG	676
677	ATT CCT CCC TTT CCC ACG TCA TAA TGT TGG TGA TAT TGG ACG GCT	721

✓ **Ocy33**

Identidade de 35%, expectânica de 6<sup>-24</sup> com Inibidor de serpin peptidase de *Danio rerio* (GenBank: NP956235).

ATGATGTTTCAGACGGGAAAATTTTTATACGGTATATTCCCAGGTCATGGATTGCTGCATTAGAGCTGCCT  
TACTCGGAACCCAATGTCAGTATGATCATTATACTACCTAAGAGTATATCTCGTCTGAAAGAGATGTTACGT  
GAACAGGACGGAAAAATACTGATTAACATCAGAAGCAAATGAATAAATTGGACGTGGATATTTGTCTGCC  
AAAATTTACGATTGAGAACGGTAAGTCAATAGAAGAGAAGCTATCTTCCTTAGGAATCCGACGGCTGTTCC  
ATCCTTCATCTGCAAATCTGGATGTGATGTCCCCTAACTCCGAGTTGTACGTCCGAAAGATAATCCACAGCG  
TANAGGTCCAAGTGGCAGAAAGCGGAACTAAAGGTCAAGCTTCTTCATCGTCCATCATTGAGGGACGTTCC  
ATCGCCTCATCTTTCTGTGCGGATCATCCTTTCCTTCTTCGTCTCGAAGACACAGAAACCAATTCTATTCTAT  
TTTGGGGTGAATACAATGGACTCGAGTAA

✓ **Ocy34**

Identidade de 61%, expectância de  $2^{-6}$  com Inibidor de domínio de Nuclease de *Danio rerio* (GenBank: NM182865).

2	GGG TTA CAA AAA ATC ATC ACT GAG TTC AAA GCA GAA GAA CAA GTT	46
47	GCA CGA AGC AAC AGG AGG AAT ATC TGG CGC TAT GGT GAT TTT GAT	91
92	GAC GAC GAT GCA GTA GAG TTT GGA TAT AAG GCT TGA ATG AAG TTT	136
		3
137	ACA TTG CAG TGT CAG ATA GAA GGA GGG CAG CCA TGT ACA TCA TAT	181
4	T L Q C Q I E G G Q P C T S Y	18
182	ATA ATA TCT TAT CTA CAT ATT TTG TGG ATG TTA ATA TTT TTC TCT	226
19	I I S Y L H I L W M L I F F S	33
227	AGA AAA TTT TTT GAC AAT TTC TTT TAC CTC AAA GAA AAT CTA TTT	271
34	R K F F D N F F Y L K E N L F	48
272	ACT ATT ATG AAT TTT ACA AAT TTA TTT GTA CAA AAA CCA CGT TTG	316
49	T I M N F T N L F V Q K P R L	63
317	AAG CAT GCA AAT GTT TTT ATG TCA AAG ACA TAA CAT CTG TAC TAT	361
64	K H A N V F M S K T *	
362	CAG TAA TAA CTG AAA AAG CTG TCC TTT GAC GCT GAC ACA TTC TGA	406
407	TGG GTG ATG GAC ATG TAA AAC CAG TGT ATT AGA ATG GAC CAG TTT	451
452	TAA TGC CAA TGA TGG TGC AGC ATT ACA AAC TCG TGA TCT CAG TCA	496
497	CTA CCT TAC CCA ATG AAT TAA AAT GCT TTC ATT ATA ATT AGC AAT	541
542	GGC ATT TTG TAA TGC TTC CAG AAA ACA TCC AGG TTT ACA AGT ACA	586
587	ATT TGG AAA AGT ATG AAG TTT ATT TGG GAA AGG TAT AAA ATT TTT	631
632	AGA TTG TTC CAT GTG TGC TGN TGC TAC CCT TGG TGC TGA TAT TAA	676
677	AAC TGG TGT ATA TGN CAA TTT GTT	700

✓ **Ocy35**

Identidade de 61%, expectância de  $8^{-11}$  com Calmodulina de *Culex pipiens quinquefasciatus* (GenBank: XM001868588).

GGGGCTCACAGATGAAGAAGTAGATGAAATGATCCGTGAAGCTGACATTGATGGAGATGGTCAAGTGAAT  
TATGAAGAATTTGTCACAATGATGACATCAAAGTAAAGAGACCCCTCTTCATTTGTAATTAATAATCTTTT  
TACTTCATATTTTGCTTTCATAACGTTTGGTGTAAAATAGCTTCAACATTTAATTGGTCTCGATGTTGCTCT  
GTAAAACTTTTGTACCATATTAAGAAAACGCCAGATGCATGAATGTGACCTGTCCCTCCATTTTCTTGTC  
TGTGTTCCATCATACTAGACTGTCTGTACTTTTGTAAATTTTCAACATTCATTAAGTTAGTCTTTATGTCA  
CANANNANAAAAAAAAAAAAAAAAAAAAAAAAACCTTGT

## 2. ESTs que apresentaram identidade à genes de genomas seqüenciados

✓ **Ocy19**

Seqüência de nucleotídeos com identidade de 37%, expectância de  $3^{-10}$  a alérgeno de *Lycosa singoriensis* (número no GenBank: EU247109).

TGGGGACACTTGTGGAAGTTATACTGTGATTTCTCTCGGAATTCACCAGCCACCATGGATTTGTTTGGGTT  
GATATCTTGCATGCTGCTACTGGTTCCTCCATATACGTCAGGCGACATTCATGGCAACTTCCCTTAGATAAT  
ATGAGGCGGGATTCGTTACGAACGTCCTTTCCAGAGGAGTTGCTGCTAGCGAGAAACGTAGCTTCGAAAC  
TGAAAACGCCCCTGAAATCTACAAACGCTACAGTACTCATCACATCTTTTGCTCGAATTCGATTTGCAATGC  
CAAGCAACGACCCTTACTCGGAAGTCTAAAGGACAACTTACTGACGATGCACAACGAACTACGCAGCAGTT  
TAGTGAGAGGTACAACACTCACTTCTCACAGTTACCTCATGCAGCAAACACGATGGAATTGGAATGGGAC  
GATGAACTAGCTCGGATGGCAGCAGCTGAAGCTCAACTTTGCCAAAAAGCCTAGATGAGATACCAAAGAC  
AGCCATAGCAGTAGGGAACTTCTCTTCGATTGGGCAAACCTTCTCCGCTATGAGCATGGGATGAATGACTT  
CATGGAACCTGTGGAATTGGAGCGACNACATTCGTTCTGGGTACATTAACCTCATTTCAATTTGAAGGATAT  
TTATGAAC



✓ **Ocy24**

Identidade de 37% (*e-value*: 1<sup>-21</sup>) a possível alérgeno identificado do genoma de *Aedes aegypti* (número no GeneBank: XP00165586).

1	GGG	ATG	CAC	AAT	GAA	TTT	AGA	AAT	AAA	ATT	GCT	CTT	GGA	AAA	GAA	45
0		M	H	N	E	F	R	N	K	I	A	L	G	K	E	14
46	ACC	ACA	CCC	CAG	CAG	CAA	CCT	CCA	GCA	GCA	AAC	ATG	ATG	CAG	ATG	90
15	T	T	P	Q	Q	Q	P	P	A	A	N	M	M	Q	M	29
91	GAA	TGG	GAT	GAT	GAA	TTG	GCC	AGA	ATT	GCG	CAA	GCT	CAT	GCT	AAT	135
30	E	W	D	D	E	L	A	R	I	A	Q	A	H	A	N	44
136	CTC	TGT	AAG	TTT	GAA	CAC	GAC	ACA	GGG	CCC	CAA	CGA	CAA	GTA	GAT	180
45	L	C	K	F	E	H	D	T	G	P	Q	R	Q	V	D	59
181	AAC	TTC	ATT	GTT	GGG	CAA	AAT	CTG	TTC	ATC	ACC	ATG	ATG	TCG	AAA	225
60	N	F	I	V	G	Q	N	L	F	I	T	M	M	S	K	74
226	AAG	ATA	GAT	TGG	AGA	AAA	GCT	GAC	ATG	TGG	TAT	ACC	TCT	GAA	GTT	270
75	K	I	D	W	R	K	A	D	M	W	Y	T	S	E	V	89
271	CAA	TAC	TTC	AAT	CCA	CAA	TAT	AGG	AAA	CCG	TTT	GTA	TTT	GGT	GGC	315
90	Q	Y	F	N	P	Q	Y	R	K	P	F	V	F	G	G	104
316	GCA	TAT	GGG	CAT	TTC	AGC	CAG	ATA	ATA	TGG	GCT	GAG	ACA	TGG	AAA	360
105	A	Y	G	H	F	S	Q	I	I	W	A	E	T	W	K	119
361	GTA	GGT	TGT	GGA	GTA	GCT	TGG	TAC	TAT	GAT	CAT	AGC	CAG	AAA	ATG	405
120	V	G	C	G	V	A	W	Y	Y	D	H	S	Q	K	M	134
406	GAT	ACA	ATG	CTA	TAT	ACC	TGC	AAT	TAT	GGA	CCT	GCC	GGA	AAC	ATG	450
135	D	T	M	L	Y	T	C	N	Y	G	P	A	G	N	M	149
451	CAG	GGA	GCA	ACT	GTT	TAC	ATT	AAA	GGA	GAA	CCA	TGC	TCT	CAA	TGT	495
150	Q	G	A	T	V	Y	I	K	G	E	P	C	S	Q	C	164
496	CCA	AAG	AAC	ACT	CAA	TGC	TCA	AAT	GAA	TAC	AAG	GGC	CTT	TTG	CAA	540
165	P	K	N	T	Q	C	S	N	E	Y	K	G	L	L	Q	179
541	TCA	ATG	ACT	CCA	GAT	GGC	CCC	CAG	AGA	GAT	TTT	TCA	CAT	CTT	CAG	585
180	S	M	T	P	D	G	P	Q	R	D	F	S	H	L	Q	194
586	ATG	AAT	TCC	TTT	TAT	ACT	GCA	ATT	TTA	GTG	TCA	ATG	AAC	AAC	GTG	630
195	M	N	S	F	Y	T	A	I	L	V	S	M	N	N	V	209
631	AGT	GCA	AAA	TGG	TAA	AAT	TAC	TGG	ATC	GAA	GTC	ATT	TNG	AAC	AAG	675
210	S	A	K	W	*											
676	AAA	AAT	ATC	AGT	GNT	GAA	TAT	AAA	CAG	TAG	TCC	TTA	ATG	GTG	GGA	720
721	ATC	AAT	TAC	ATC	AGT	TAG	GGA	AAG	CCA	GAT	GAT	ACG	GAA	ATG	CCG	765
766	TTC	AAT	ATG	GAT	TTT											780

## Alinhamento de Ocy24 com possível alérgeno de *Aedes aegypti* (XP001655867)

```

XP001655867      MFTVNESGLGSTDKYDILEEHNKLRQAVAQGLIPG---QPGAENMLEMQWDELARKAQD 57
Ocy24            -----MHNEFRNKIALGKETTPQQQPPAANMMQMEWDELARIAQA 41

XP001655867      WADQCIFKHD--PKKKLGRFTMGQNLALIWSSASLDDDRSDFPGRIRKWFDEVKKYSFG 115
Ocy24            HANLCKFEHDTGPQRQVDNFIVGQNLFITMMSKKIDWRKADMWYTSEVQYFNPQYRKPF- 100

XP001655867      AKFSLATGHYMQLIWADTYLVGCGYSYHYSGNK--YNKLYVCNYGPAGNVQGVQVPYITGH 173
Ocy24            -VFGGAYGHFSQIIWAETWKVCGVAWYYDHSQKMDTMLYTCNYGPAGNMQGATVYIKGE 159

XP001655867      PSCEQHG-LQKSSSTYPGLCAAGAGAGEVDPYHHDNVIPQNYYSYQASSNNHENQHYTSAP 232
Ocy24            PCSQCPKNTQCSNEYKGLLQSMTPDGPQRDFSHLQMN--SFYTAILVSMNNVSAKW---- 213
    
```

### ✓ Ocy29

Identidade de 40%, *e-value* de provável proteína salivar identificada do genoma de *Ixodes scapularis*.

```

3      GGG GTG TAA ACG TCG TCT GAC ATT GCA AAG ATG CCC ATA ACA AGA 47
                                           M   P   I   T   R   4

48     GCT GTG TTT TAC TGG GTG ATG TTA AAC GTT TGT ATA ATC TTC ACC 92
5      A   V   F   Y   W   V   M   L   N   V   C   I   I   F   T   19

93     TTC TCT AGC GGA TAT ACG TAC CTT CGC CGC CAG AAA TAT AGG GCA 137
20     F   S   S   G   Y   T   Y   L   R   R   Q   K   Y   R   A   34

138    TGT GCA ACG TCG CCC TGT GTG GAC CAT TTA GGA AAC GAC CAC GAA 182
35     C   A   T   S   P   C   V   D   H   L   G   N   D   H   E   49

183    TAC AAC GAC TCG TGG TTC CAG GAT GAG AGA TGC GAG GAG CAC ACT 227
50     Y   N   D   S   W   F   Q   D   E   R   C   E   E   H   T   64

228    TGT GTC GAG TAT CAT GGA ATA GCA TAT ATA CAA GCT TAC GGC TGC 272
65     C   V   E   Y   H   G   I   A   Y   I   Q   A   Y   G   C   79

273    GGA GTA ACG GAT GCC GCA CCG GGA TGC AAG CTA GAA AGA GGC AAG 317
80     G   V   T   D   A   A   P   G   C   K   L   E   R   G   K   94

318    GGC TCG TAT CCG GAC TGC TGT GAG GAA GAA GTT TGC TAA GTG CCT 362
95     G   S   Y   P   D   C   C   E   E   E   V   C   *

363    TTC TCC TTA GAA CAT TTT GGA AGG TAG AAC CAT TGG ATG TTG CGA 407
408    GAG AAG GAG GCA TTG CAC AAG AAA AAG CAT ATT ACA GTG TAT TAT 452
453    TAT GCT CAT TTC ATA ATT TGT ATA TGA TAT ATT ATC TCT TAA TAT 497
498    CAA TTA CAA CCT CCT TCT TTA TAG AAA TTA TAA TTA AAT TTT CAT 542
543    TAT TTT AAA CTC TTC TCA CCT CCT CTT GGA CCC CAT AAA ATA TTC 587
588    TTA AAA CCT TTT CTC ACG TGA TTA CNT ATT TCT TAT TAC TTA TTA 632
633    TAT AAC ATG TTA CAT AAT ATA TTT ATT TTA TAC TTA TTC TAT NTA 677
    
```

678 TTT TAT TCT CAA TTC TTG TGT ATA TAT TAA TTA ACT TTA TTT ATT 722

Alinhamento de Ocy29 com provável proteína salivar de *Ixodes scapularis*. Em azul os peptídeos sinal preditos.

Ocy29 MPITRAVFYWMLNVCIIFTFSSSGYTYLRRQKYRACATSPCVDHLGNDHEYNDSWFQDER 60  
 AAY6651 MNFLEALFCG-FVALALIEHISMG--ALPAYGEVEVVDGKCR-YMNRVQDGEDMHLEDP 56

Ocy29 CEEHTCVEYHGIAYIQAYGCGVTDAAPGCKLERGKGSYPDCCEEEVC----- 107  
 AAY6651 CQTLKCEAKKGS--FAILGCGIMSSKPGCKLEKGTGVFPQCCYKSVCSKTIKKRSFV 111

✓ **Ocy36**

Identidade de 58%, *e-value* de  $9^{-18}$  com proteína hipotética de *Apis mellifera*

(GenBank: XM623658)

3	GGG CGT GAG TCT GGG CAG AAA GGT GTT GAG GTT GCT AGT ACT TGA	47
48	TTG TCT TCA TAA GGT GCT CCG TGT ATC AGG AAA ATG TTC ACT TGC	92
	M F T C	3
93	GCA AAA TAT CTC GCA CCA TTA TCA AGA ACC TTG GTA GCA ACT GGT	137
4	A K Y L A P L S R T L V A T G	18
138	ACT CGT AGC CTT GCT AGA CCA ATT AGT TCT GTA CTA ATT AAG CCA	182
19	T R S L A R P I S S V L I K P	33
183	AGG GAA GAG ATT ACA GCC ATG GAT CCA TCA ATT TCA GGA GTA AAT	227
34	R E E I T A M D P S I S G V N	48
228	CTT AAT ATT ATA TCT ACT CCG TTA ATG TCA CAG ATT ACA CGT AAT	272
49	L N I I S T P L M S Q I T R N	63
273	CTC CAA ACT ACT TCC ATA TGT AGG GAT ATT GAT TCT GCT GCA AAG	317
64	L Q T T S I C R D I D S A A K	78
318	TTT ATT GGA GCT GGG GCA GCT ACT GTT GGG GTA GCT GGC TCA GGT	362
79	F I G A G A A T V G V A G S G	93
363	GCC GGA ATT GGA AGT GTA TTT GGG AGT CCT ATC ATT GGA TAT GCC	407
94	A G I G S V F G S P I I G Y A	108
408	AGA TAC CCA TCA CTT ATA CAG CAA CTT TTT TCA TAT GCC ATT CTA	452
109	R Y P S L I Q Q L F S Y A I L	123
453	GGG TTT GCA TTA TCT GAA GCC ATG GGT CTG TTT TGT TTG ATG ATG	497
124	G F A L S E A M G L F C L M M	138
498	GTA TTC TTG CTA CTG TTT GCC TTC TAA TGG TAG TTT GCC GGT CAA	542
139	V F L L L F A F *	
543	TGC TGA TCA TTA GCC AAT TTT ATT CAT CTA TTC TTT TAC TCC AGG	587
588	TAC CAG TAA TTG GTT TGC TTT CAT GTG TAC TTA ACA TGA ATA GAC	632
633	AAA TAA CTA ATG TTC TGG TAT TAC CCA CTG TGA ACT TAA CTT CAC	677
678	TTT TTT TTT TTA AAT GCT CGA ACC CTC TCC TTG TTT TCC TAA TTC	722
723	CTG TCA TAT TAA TTC CTC TAA TAT GAT TAG NCT TTC CAT CAN	764

✓ **Ocy37**

Identidade de 54%, *e-value* de  $6^{-25}$  com proteína hipotética de *Apis mellifera*  
(GenBank: XM396838)

2	GGG GAG TTG GCT	ATG GCC TGC CGC	TTG CAG CTT CTG	TTA GTG ACT	46
		M A C R	L Q L L	L V T	11
47	ATG AAC CTG GTG	TAT TAT AAT ACC	ACA ACC AAT CAT	AAT TGT TGT	91
12	M N L V	Y Y N T	T T N H	N C C	26
92	TTA AGA AAC TCA	ACT ATA ATT CAC	CAG AAA GGT GCA	CCA TAT ATA	136
27	L R N S	T I I H	Q K G A	P Y I	41
137	CTG GGC CTT GTC	AGT TTC ACT ATA	TCA TTG ACA ACA	GTG TAT GTT	181
42	L G L V	S F T I	S L T T	V Y V	56
182	ATT TGC TAC TGT	GTG AGA AGT CAT	TTC CAA AGC GTG	TCG CTT TTT	226
57	I C Y C	V R S H	F Q S V	S L F	71
227	CAT ATT TAG AAG	ACT TGT CGA ACG	AGT TTT ATA GTC	TGT ACG GTC	271
72	H I *				
272	AAT GAA TTC TAG	TAA TCA NTC GAC	CTT ACC ACT TTA	TCT GAA ATT	316
317	GAC TAT TAC ATC	CAT AAA GCA TAA	AAG TCT TAT GTT	GAT GCA AGA	361
362	TCT CTC TAA AAT	TCT GCC ACA CTT	AAT TAC TTA TTT	GCC AAA ATG	406
407	TGT ATA GGG ATA	ATG ATA GGT CAT	TAT CCA TGA TGC	TTT GCT AAA	451
452	AAG TCC TTG CTC	TGT CAG TAT TTG	GAT AAA AAA GTA	AGG TAA TTT	496
497	TGT CGA TCC TTA	TCG CTT TAT TTA	TAT TAT CTG ACC	CTC CCT TCT	541
542	TTT CAT TTC GAA	ATC CCA TAT GCC	CAA TTT GGG TTT	TGG TGG GTG	586
587	GTA AAA GGT TTG	TAT TTT TCC CTT	ATT TTT CTA TCC	CTC TAA AAA	631
632	TTT TAT TGG GGT	ATA TTA ATT TGG	TGA GAA ATA AAA	TTT AAC CCG	676
677	GGT GTT AAT AAT	TAC TGG TAA TTT	TTA AAT AAA TTT	TTC TTT ATT	721
722	TAA AAT TCT ATA	TTT ATT TGA TAA	TAA CGG GAC CTG	TAT TAT ATG	766
767	TAT CTT TTA ATT	CAC CCC CTC CCC	790		

✓ **Ocy38**

Identidade de 62%, *e-value* de  $3^{-10}$  com proteína hipotética de *Apis melífera*  
(GenBank: XM001120845)

```

1      GGG GAT AAT CCA TGG TGA AAG AGA ATG TTG TGG AAA AGT GTT AGC 45
46     TCC TTT CCA GCA GCA ACG TGT GAT AGC TTA TTT TGC AAC CTG TGA 90
91     AAG TAT CGC TAG AAT AAA ACA TCG ATG TCA GGT GTG AGA CAT GGA 135
          M  S  G  V  R  H  G  7

136    AAC ACT GAA TAT AAC GAC GAT CCA GTG GAG ACA TTA CTT AAT AAA 180
8      N  T  E  Y  N  D  D  P  V  E  T  L  L  N  K  22

181    ACT GGA TGC GCC GAT TTG CAT TAC GCC GTT CAG GAA TGC ATA GCT 225
23     T  G  C  A  D  L  H  Y  A  V  Q  E  C  I  A  37

226    GAA CAC AGA GAT TGG AGA AAG TGC CAA GAA CAA GTG AAA AAT TCC 270
38     E  H  R  D  W  R  K  C  Q  E  Q  V  K  N  S  52

271    GAA TGT GTA TAG AAN ATA ATT CAA AGA AAA AAC AAT AAT TTT ATA 315
53     E  C  V  *

316    GAC CTT TAT GTT AAA TAT GTT CCA CAA ATA AAC TAG AGA TAG AAT 360
361    GAT TCA TTG TTC ACA CAA AAN ANA NNN NNN NAA NAA AAA AAA ACC 405
406    TTG TCG GCC GCC TCG GCC TTC GAA AAG CTT TTC TAA ACA TTC GTT 450
451    TGG CCC GCC GGC CCA GTT AGG AAG TGA ACA TGG GCC TAC CTG TTT 495
496    CTT AGA AAT CCT GGC CAA GGA CAG TGG TTG GAT TCT CCC CAA AAA 540
541    AAA CCC GCG GAA CGA GGT TGA AAA ACA AGG GGT CGA GCT TTG GGT 585
586    TTT AAG GAC CAG GGG TCA TTA AAT ACC CCC CGG CCT CCC AAA GTG 630
631    TAT TTA ATA TTT CGC ATG GCC CCA AGG GTG TAA TTA TCC GGG TCC 675
676    CCC TTC CCC GTA TAT TTC CGG AGA GGA AAA TTC TTT TCC TTT AAA 720
721    AGG AAA CCC GGG GGG 735

```

Alinhamento Ocy38 e proteína hipotética de *Apis melífera* (XP001120845)

```

XP001120845      MMIPKLNSSQQNIEDPVEEMLKKTGCMFHYQVQECIAETQDWRKQCQEQVRFKICMDKY 60
Ocy38            -MSGVRHGNTEYNDDPVETLLNKTGCDLHYAVQECIAEHRDWRKQCQEQVKNSE-CV--- 55

XP001120845      QRKREESYSN 70
Ocy38            -----

```

✓ **Ocy39**

Identidade de 62%, *e-value* de  $3^{-10}$  com proteína hipotética de *Danio rerio* (GenBank: XM001334104)

3	GGG TAA CAA TAG AGA AGA TAA ATC GTC GGG TTT GTC ATT TTA CCG	47
48	CTT TCC AAC TGA CAA GGA AAG AAG GCA AGC ATG GAT TAA TGC TAT	92
93	AAA ACG TGA TAA TTG GAC GCC GTC GCC ATA TTC ACG TTT ACG CAG	137
138	TGT CCA CTT TAT GTC AGA TAT GTC ATT TCT TGG ACC ACT TGA CTG	182
183	CTA AAC GTA AGA AGC TGA CCC GAG TCC AAT TGA CTC AGA TCA TCT	227
228	CCC AAA ACA GGT GAA AAT ATT GCC TTT GGT AGG CTA TAA GGA TCA	272
273	TCA ATG TTC AGA ATT TTC AGT TTT TCT ACG TAC CGT TGC TTC ACT	317
	M F R I F S F S T Y R C F T	13
318	ATA CTA CCA AAA TGT TCT ACC TGC TTT GAT AAC TTC GTT TTC ATC	362
14	I L P K C S T C F D N F V F I	28
363	CTT CTC ACC CAT ACA CCG CAA CAA GAA ACT AGA CGT CCA AAA TGG	407
29	L L T H T P Q Q E T R R P K W	43
408	CGA CAA CAA CAT TGG CAC GTG ACA TCT ATG ACA TCA CTG CAA AAC	452
44	R Q Q H W H V T S M T S L Q N	58
453	CTC TAT AAG TTT GTT CCG AAT GTT GCA ACA ATA CTC GTG CAC TGC	497
59	L Y K F V P N V A T I L V H C	73
498	CAA ACC TGC GCT ATA ACG ATG ACG TAA CAA TGC TAG ACG ATT ACT	542
74	Q T C A I T M T *	
543	CAT CGA TGT GTG TAA CAA TGT AGT TCG TAA TAA TTT AAG GAC CAA	587
588	AAA TAT TTT TCC TGG ATC GAT AAG CGA CTA GAA GAC TGG TGT GAT	632
633	TCT TGA ATA GGG GTT TCT TCT TTN CTG TAT TTA GAA GGG CTG GCT	677
678	GGN TGG GGG AAA GAC ACA GCA	698