

Alessandra Finardi de Souza

**Caracterização de Metaloproteinases PIII a partir do DNA genômico de
*Bothrops jararaca***

Dissertação apresentada ao
Programa de Pós-Graduação
Interunidades em
Biotecnologia USP/ Instituto
Butantan/ IPT, do Instituto de
Ciências Biomédicas da
Universidade de São Paulo,
para a obtenção do Título de
Mestre em Biotecnologia.

São Paulo
2011

Alessandra Finardi de Souza

**Caracterização de Metaloproteinases
PIII a partir do DNA genômico de
*Bothrops jararaca***

Dissertação apresentada ao Programa de Pós-Graduação Interunidades em Biotecnologia USP/ Instituto Butantan/ IPT, do Instituto de Ciências Biomédicas da Universidade de São Paulo, para a obtenção do Título de Mestre em Biotecnologia.

Área de Concentração: Biotecnologia

Orientadora: Prof. Dra. Itamar Romano Garcia Ruiz

São Paulo
2011

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

© reprodução total

Souza, Alessandra Finardi.

Caracterização de metaloproteinase PIII a partir do DNA genômico de *Bothrops jararaca* / Alessandra Finardi Souza. -- São Paulo, 2011.

Orientador: Itamar Romano Garcia Ruiz.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan. Área de concentração: Biotecnologia. Linha de pesquisa: Biologia molecular.

Versão do título para o inglês: Characterization of Metalloproteinases PIII from genomic DNA of *Bothrops jararaca*.

Descritores: 1. Veneno 2. Genes de metaloproteinases 3. Exon 4. Intron 5. *Bothrops jararaca* I. Ruiz, Itamar Romano Garcia II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação em Biotecnologia USP/IPT/Instituto Butantan III. Título.

ICB/SBIB085/2011

UNIVERSIDADE DE SÃO PAULO
Programa de Pós-Graduação Interunidades em Biotecnologia
Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

Candidato(a): Alessandra Finardi Souza.

Título da Dissertação: Caracterização de metaloproteinase PIII a partir do DNA genômico de *Bothrops jararaca*.

Orientador(a): Itamar Romano Garcia Ruiz.

A Comissão Julgadora dos trabalhos de Defesa da **Dissertação de Mestrado**, em sessão pública realizada a/...../.....,

Aprovado(a)

Reprovado(a)

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:

*Dedico este trabalho as minhas três mães: minha
mãe Cecília, minha tia Vera, minha avó Adélia,
que sempre me deram incentivo para continuar
meus estudos. E ao meu namorado Gregório pelas
horas de paciência durante a elaboração deste
trabalho.*

Agradecimento

A Prof. Dra. Itamar Romano Garcia Ruiz do laboratório de Genética do Instituto Butantan, pela orientação, os ricos ensinamentos passados durante os anos que estive no laboratório.

Aos funcionários da Genética: Dona Maria, Nair, Diva e Zenaide.

Aos companheiros “das antigas”: Losanges, Juliana, Luciana, Anne e Gabriela.

Aos companheiros “das novas”: Daniel, Marcel, e principalmente a Flávia e a Cristiane que deram uma grande força no final deste trabalho.

Ao pessoal do Laboratório de Genética da Dra. Irina Kerkis: Dr. Carlos Maranduba, Babila, Lislei e Rui por disponibilizar o uso de equipamentos.

Ao pessoal do Laboratório de Biotecnologia: Prof. Dr. Paulo Lee Ho e Leonardo Kobashi, pelos sequenciamentos.

Ao Dr. Durvanei Augusto Maria, do Laboratório de Bioquímica pelo contato com o pessoal da Applied pelos meus últimos seqüenciamentos.

Ao meu atual chefe Dr. Eduardo Finger e minha amiga de trabalho Thaissa pela paciência e compreensão para finalizar minha Dissertação.

E a todos os meus familiares.

A Capes pelo apoio financeiro.

E todas as pessoas que de maneira direta ou indireta contribuíram para o andamento desse trabalho.

Muito Obrigada!!!

O “sábio não é o homem que fornece as respostas verdadeiras; é o que formula as perguntas verdadeiras”.

Lévi-Strauss

Resumo

Souza AF. Caracterização de Metaloproteinases PIII a partir do DNA genômico de *Bothrops jararaca* [Dissertação (Mestrado em Biotecnologia)]. São Paulo: Instituto de Ciências Biomédicas da Universidade de São Paulo; 2011.

O veneno de *Bothrops jararaca* contém uma série de componentes, entre eles as metaloproteinases hemorrágicas jararagina e bothropasina. Os cDNAs dessas toxinas mostram 97% de identidade. As diferenças, distribuídas ao longo de seus cDNAs, sugerem que estes mRNAs não resultam de *splicing* alternativo. O objetivo deste trabalho foi caracterizar os genes codificadores da jararagina e bothropasina pela identificação de exons e introns no DNA genômico. DNA foi extraído do sangue de um exemplar de *B. jararaca*; os *primers* para PCR foram baseados nos cDNAs publicados. Os produtos de amplificação foram clonados e seqüenciados revelando a sequência dos genes TOX1 com 12535 pb e TOX2 com 12268 pb. Quatorze exons e treze introns foram identificados em ambos os genes. Comparação entre as sequências mostrou pontos de mutação, inserções e deleções nos exons, e principalmente nos introns dos dois genes. Este constitui o primeiro relato na literatura sobre a identificação de exons e introns nos genes codificadores de jararagina e bothropasina.

Palavras chaves: Veneno. Genes de metaloproteinases. Introns. Exons. *Bothrops jararaca*.

Abstract

Souza AF. Characterization of Metalloproteinases PIII from genomic DNA of *Bothrops jararaca* [Master Thesis (Biochemistry)]. São Paulo: Instituto de Ciências Biomédicas da Universidade de São Paulo; 2011.

The *Bothrops jararaca* venom contains a number of components, including hemorrhagic metalloproteinases as jararhagin and bothropasin. The cDNA of these toxins show 97% identity. The differences distributed along the cDNAs length suggest that these mRNAs do not result from alternative splicing. This study aimed to characterize the genes that encode for jararhagin and bothropasin through the identification of exons and introns in genomic DNA. DNA was extracted from the blood of a *B. jararaca* specimen; PCR primers were based on published cDNA sequences. Amplification products were cloned and sequenced revealing the TOX1 gene is about 12,535 bp long, and TOX2 is 12,268 bp. Fourteen exons and thirteen introns were identified in both genes. Comparison of the sequences showed point mutations, insertions and deletions in exons, and particularly in introns. This is the first report in the literature on the identification of exons and introns in genes encoding for jararhagin and bothropasin.

Keywords: Venom. Metalloproteinases genes. Introns. Exons. *Bothrops jararaca*.

Lista de Ilustrações

Figura 1. Família das zinco-metaloproteínas	16
Figura 2. Estrutura dos multi-domínios das metaloproteínas	19
Figura 3. Representação esquemática do precursor da bothropasina	20
Figura 4. Estrutura da integrina	26
Figura 5. Alinhamento dos aminoácidos da jararagina e bothropasina	29
Figura 6. Estrutura tridimensional da bothropasina	30
Figura 7. Organização genômica de íntrons de serpentes	32
Figura 8. Mecanismo de <i>splicing</i> nuclear	33
Figura 9. Construção de um vetor utilizando íntrons	35
Figura 10. Sequência de clonagem <i>pGEM-T easy Vector</i>	41
Figura 11. Esquema do <i>pGEM-T easy Vector</i>	41
Figura 12. PCR Pbd-Pbr	45
Figura 13. Minipreps Pbd-Pbr	46
Figura 14. Digestão <i>EcoRI</i> Pbd-Pbr	46
Figura 15 Sequenciamento 5' UTR e Peptídeo Senal	49
Figura 16. PCR PbirD-Pair	50
Figura 17. Minipreps PbirD-Pair	50
Figura 18. PCR das minipreps PbirD-Pair	51
Figura 19. Sequenciamento PbirD-Pair	53
Figura 20. PCR Paid-Par	53
Figura 21. Minipreps Paid-Par	53
Figura 22. PCR das minipreps Paid-Par	54
Figura 23. Sequenciamento Pair-Par	54
Figura 24. Minipreps P1d-P1r	55
Figura 25. PCR das minipreps P1d-P1r	56
Figura 26. Digestão <i>EcoRI</i> P1d-P1rr	56
Figura 27. Sequenciamento P1d-P1r	58
Figura 28. PCR P1d-P1rr	58
Figura 29. Minipreps P1d-P1rr	58
Figura 30. PCR das minipreps P1d-P1rr	59
Figura 31. Digestão <i>EcoRI</i> P1d-P1rr	59
Figura 32. Sequenciamento P1d-P1rr	60
Figura 33. Sequenciamento P0d-P0rr	61
Figura 34. PCR P0dd-P0rr	63
Figura 35. Minipreps P0dd-P0rr	63
Figura 36. PCR das minipreps P0dd-P0rr	64
Figura 37. Digestão <i>EcoRI</i> P0dd-P0rr	64
Figura 38. Sequenciamento P0dd-P0rr	66
Figura 39. Sequenciamento P0dd-P0r	68
Figura 40. PCR P0dr-P2r	68
Figura 41. Minipreps P0dr-P2r	69
Figura 42. PCR das minipreps P0dr-P2r	69
Figura 43. Digestão <i>EcoRI</i> P0dr-P2r	69
Figura 44. Sequenciamento P0dr-P2r	71
Figura 45. PCR P4d-P4ir	71
Figura 46. Minipreps P4d-P4ir	71
Figura 47. PCR das minipreps P4d-P4ir	72
Figura 48. Sequenciamento P4d-P4ir	72

Figura 49. PCR P11d-P11r.....	73
Figura 50. Minipreps P11d-P11r.....	73
Figura 51. Digestão <i>EcoRI</i> P11d-P11r.....	74
Figura 52. Sequenciamento P11d-P11r.....	77
Figura 53. PCR P9d-P9r.....	77
Figura 54. Minipreps P9d-P9r.....	77
Figura 55. PCR das minipreps P9d-P9r.....	78
Figura 56. Sequenciamento P9d-P9r.....	78
Figura 57. PCR P6d-P6r.....	80
Figura 58. Minipreps P6d-P6r.....	80
Figura 59. PCR das minipreps P6d-P6r.....	80
Figura 60. Digestão <i>EcoRI</i> P6d-P6r.....	81
Figura 61. Sequenciamento P6d-P6r.....	82
Figura 62. PCR P7d-P7r.....	83
Figura 63. Minipreps P7d-P7r.....	83
Figura 64. PCR das minipreps P7d-P7r.....	83
Figura 65. Sequenciamento P7d-P7r.....	84
Figura 66. PCR P12d-P12r.....	85
Figura 67. Minipreps P12d-P12r.....	85
Figura 68. PCR das minipreps P12d-P12r.....	86
Figura 69. Sequenciamento P12d-P12r.....	86
Figura 70. PCR P12d-P10r.....	87
Figura 71. Minipreps P12d-P10r.....	88
Figura 72. Sequenciamento P12d-P10r.....	90
Figura 73. PCR P8id-P8r.....	90
Figura 74. Minipreps P8id-P8r.....	90
Figura 75. PCR das minipreps P8id-P8r.....	91
Figura 76. Sequenciamento P8id-P8r.....	91
Figura 77. PCR P5d-P5r.....	92
Figura 78. Minipreps P5d-P5r.....	92
Figura 79. PCR das minipreps P5d-P5r.....	93
Figura 80. Sequenciamento P5d-P5r.....	94
Figura 81. PCR P3d-P3r.....	95
Figura 82. Minipreps P3d-P3r.....	96
Figura 83. PCR das minipreps P3d-P3r.....	96
Figura 84. Sequenciamento P3d-P3r.....	97
Figura 85. Esquema exons e introns TOX1 (S) e TOX2 (I).....	98

Lista de Quadro

Quadro 1 - <i>Primers</i> para amplificação	39
Quadro 2 - Clones P1d-P1r	55
Quadro 3 - Alterações exon 3	100
Quadro 4 - Alterações exon 4	101
Quadro 5 - Alterações exon 5	102
Quadro 6 - Alterações exon 7	102
Quadro 7 - Alterações exon X	103
Quadro 8 - Alterações exon Y	103
Quadro 9 - Alterações exon Z	104
Quadro 10 - Alterações exon K	104
Quadro 11 - Alterações na região 3' UTR	105
Quadro 12 - Alterações nos introns	107

Sumário

1 INTRODUÇÃO	14
1.1 Composição do veneno	14
1.2 Metaloproteinase de veneno (SVMPs)	15
1.3 A Disintegrin And Metalloproteases (ADAMS)	20
1.4 Variabilidade do veneno ofídico	21
1.4.1 Evolução das Fosfolipases A2	22
1.4.2 Evolução das Lectinas Tipo-C	23
1.4.3 Evolução das Desintegrinas	23
1.4.4 Evolução das Metaloproteinases	24
1.5 Integrinas e Toxinas	25
1.6 Atividade das metaloproteinases sobre tumores	26
1.7 Jararagina e Bothropasina	28
1.8 Estrutura genômica de genes referentes a outros tipos de toxinas	31
1.9 Introns	33
1.9.1. Origem e função dos introns	34
2 OBJETIVO	36
3 JUSTIFICATIVA	37
4 MATERIAIS E MÉTODOS	38
4.1 Extração de sangue de <i>Bothrops jararaca</i>	38
4.2 Técnica de extração do DNA genômico	38
4.3 Estratégia de amplificação	39
4.4 Eletroforese e eluição	40
4.5 Clonagem molécula	40
4.5.1 Preparação de bactérias competentes	41
4.5.2 Transformação de bactérias competentes	42
4.5.3 Plaqueamento	42
4.5.4 Obtenção de DNA de plasmídeo recombinante (miniprep)	42
4.5.5 Análise do DNA recombinante	43
4.6 PCR dos clones para sequenciamento	43
4.7 Sequenciamento do DNA	43
4.8 Análise das sequências	44
5 RESULTADOS	45
5.1 Região 5' não traduzida (UTR) e Peptídeo Sinal	45
5.2 Pró-domínio	49
5.3 Domínio catalítico	63
5.4 Domínio desintegrina	78
5.5 Domínio rico em cisteína	86
5.6 Região 3' não traduzida	95
6 DISCUSSÃO	99
6.1 Análise dos exons	99
6.1.1 Peptídeo sinal	99
6.1.2 Pró domínio	100
6.1.3 Domínio catalítico	101
6.1.4 Domínio desintegrina	103
6.1.5 Domínio rico em cisteína	103
6.1.6 Região 3' UTR	104
6.2 Análise dos introns	106
7 CONCLUSÃO	108

REFERÊNCIAS	109
APÊNDICE - Reagentes	118
ANEXO A – Alinhamento dos cDNAs da Jararagina e Bothropasina	121
ANEXO B – Alinhamento da TOX1 e TOX2	125
ANEXO C – Alinhamento de aminoácidos de metaloproteinases	143
ANEXO D – Aminoácidos	145
ANEXO E – Sequências no <i>GenBank</i>	146

1 INTRODUÇÃO

1.1 Composição do veneno

O veneno é produzido por um par de glândulas especializadas localizadas no maxilar das serpentes. Seus componentes parecem ser bastante comuns e similares uns aos outros dentro da mesma família. Toxinas neurotóxicas são geralmente encontradas nos venenos de Hydrophidae e Elapidae, causando morte por bloqueio da transmissão neuromuscular. Já as serpentes das outras famílias, Viperidae e Crotalidae, causam choque, coagulação intravascular, hemorragia local e sistêmica, edema e necrose tecidual, devido à ação de toxinas hemorrágicas e mionecróticas (Bjarnason e Fox, 1995; Fox e Serrano, 2005).

Os constituintes inorgânicos dos venenos incluem: cálcio, cobre, ferro, potássio, magnésio, manganês, sódio, fósforo, cobalto e zinco. A presença e quantidades destes metais nos venenos variam de acordo com a espécie. Acredita-se que a função de alguns destes componentes, tais como cálcio, magnésio e manganês, seja contribuir para a estabilização de proteínas do veneno, enquanto outros, como zinco, cobre, ferro e cobalto, participam de mecanismos catalíticos de determinados componentes enzimáticos presentes no veneno, tais como as metaloproteinases (Bjarnason e Fox, 1995).

Os principais componentes encontrados no veneno de *Bothrops jararaca* são: fosfolipases, lectinas, serinoproteases, desintegrinas, peptídeos potenciadores da bradiginina e metaloproteinases (Russel et al., 1981; Tu, 1988).

As fosfolipases A2 (PLA2) são enzimas lipolíticas, que produzem lisofosfatídeos, ácidos graxos livres (principalmente o ácido aracdônico) e mediadores pró-inflamatórios como PAF (Fator Ativador de Plaquetas) (Ohno et al; 2003). O veneno de *B. jararaca* apresenta uma baixa atividade fosfolipásica quando comparada ao veneno de outras espécies do gênero *Bothrops* (Moura-da-Silva et al., 1991).

As Lectinas tipo-C são proteínas não-enzimáticas que se ligam a mono e oligossacarídeos, encontradas em diversos animais. São geralmente proteínas de domínios múltiplos que contêm uma ou mais cópias de uma sequência de

aproximadamente 120 aminoácidos, altamente conservada, denominada domínio reconhecedor de carboidrato (CRD). Na presença de Ca^{2+} , as lectinas tipo-C iniciam uma gama de processos biológicos como adesão, endocitose e neutralização de patógenos. As lectinas tipo-C do veneno de serpentes, contêm os CRDs, portanto ligam-se somente a açúcares (Weis et al., 1998).

As serinoproteases presentes no veneno botrópico são caracterizadas principalmente por atuarem sobre o fibrinogênio, e são chamadas de tipo-trombina. Hidrolisam somente a região N-terminal da cadeia $\text{A}\alpha$ do fibrinogênio, liberando apenas o fibrinopeptídeo A, diferentemente do que ocorre com a trombina endógena que atua também na cadeia $\text{B}\beta$ do fibrinogênio, liberando o fibrinopeptídeo B (Kamiguti e Sano-Martins, 1995).

Os Peptídeos Potenciadores da Bradicinina (BPPs) são uma classe de inibidores naturais da enzima conversora da angiotensina (ACE), que apresentam efeito anti-hipertensivo. A ACE é uma peptidase presente na membrana citoplasmática de células endoteliais responsável pela conversão de angiotensina I em angiotensina II e pela degradação da bradicinina (Orning et al., 1991).

1.2 Metaloproteinases de venenos (SVMPs)

As SVMPs (*Snake Venom Metalloproteases*) são membros da superfamília das proteases dependentes de zinco, as chamadas zinco-metaloproteinases, dividida conforme a estrutura primária de seus sítios catalíticos em (Hopper, 1994):

- Zincinas: possuem a sequência HEXXH como motivo de ligação ao zinco.
- Inverzincinas: possuem um motivo invertido da ligação ao zinco, HXXEH.
- Carboxipeptidase: com motivo de ligação ao zinco HXXE.
- DD-carboxipeptidases: com motivo de ligação ao zinco HXH.

As SVMPs pertencem ao subgrupo das reprotinas, caracterizadas por possuírem um resíduo de ácido aspártico após a terceira histidina ligante de zinco (Fig. 1).

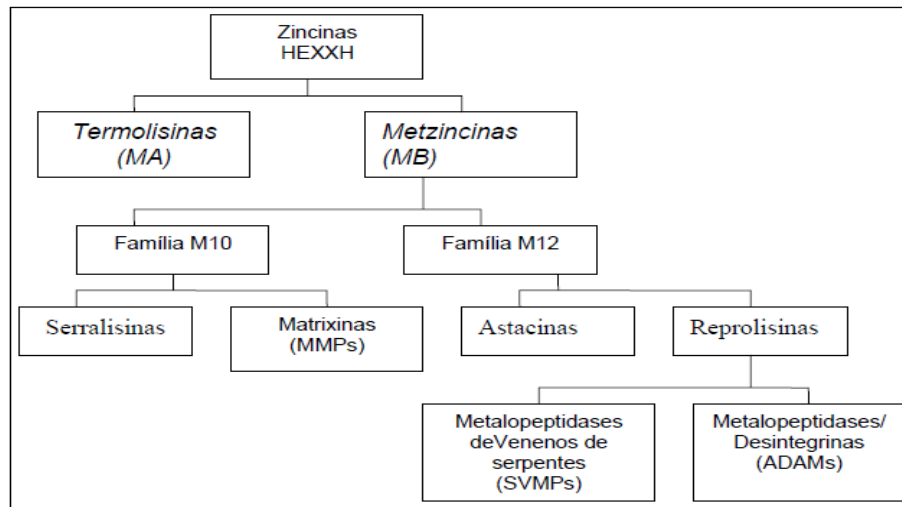


Figura 1. Família das zinco-metaloproteínas
Fonte: Hooper, 1994.

As SVMPs têm sido associadas às atividades hemorrágicas, fibrinolíticas e inibidoras da agregação plaquetária, decorrentes do envenenamento (Bjarnason e Fox, 1995). A hemorragia é causada pela ação das metaloproteínas que degradam o colágeno e outros componentes da lâmina basal dos vasos capilares. Como consequência os capilares são rompidos, promovendo equimose e sangramento. O veneno possui uma ação sobre os fatores de coagulação, alterando o tempo de coagulação. Os venenos do gênero *Bothrops* diminuem o tempo de coagulação, por atuarem sobre os fatores da cascata de coagulação e sobre a trombina (Queiroz e Petto, 1984; Baramova et al., 1990).

O edema local é outra típica manifestação de envenenamento bothrópico, independente da via de inoculação. O edema é acompanhado de dor, que pode variar de discreta a intensa. Isto é causado, provavelmente, pela combinação de elementos tais como: efeito direto do veneno sobre os vasos e liberação de mediadores endógenos como histaminas, cininas e prostaglandinas devido à ação dos componentes dos venenos sobre os mastócitos, cininogênios e fosfolípidios, respectivamente. Em alguns casos, o edema é responsável pela elevação da pressão intersticial hidrostática nos compartimentos musculares. Além da sua ação sobre células musculares e microvasculatura, o veneno bothrópico também afeta artérias, causando trombose e lesão das paredes arteriais. Lesão arterial causa isquemia e futura necrose, também afetando os nervos intramusculares (Barravieira, 1994).

As metaloproteinases fibrinolíticas não hemorrágicas também foram descritas em muitos venenos, como a Proteinase J, que está presente no veneno botrópico, e é um exemplo de metaloproteinase dependente de zinco que não apresenta atividade hemorrágica (Hite et al., 1994; Tanizaki et al., 1989).

A análise comparativa da estrutura primária das metaloproteinases hemorrágicas e não hemorrágicas de venenos de serpentes evidencia um alto grau de similaridade e não fornece nenhum indício para elucidação de características que justifiquem a especificidade por um determinado substrato ou a presença/ausência de atividade hemorrágica (Terada et al., 1999). Em vista disso, os efeitos hemorrágicos podem estar associados não somente com a estrutura protéica da SVMPs, mas podem estar relacionados com algumas modificações pós-traducionais (Bjarnason e Fox, 1995).

As hemorraginas podem ser divididas em dois grupos baseados em seu tamanho: pequenas (~25 KDa) e grandes (~90 KDa). As duas hemorraginas principais do veneno de *B. jararaca* são a jararagina e a bothropasina, descritas por Paine et al. (1992) e Mandelbaum et al. (1982), respectivamente.

Recentemente uma nova classificação das SVMP foi proposta baseando-se nas características dos precursores das metaloproteinases de venenos, bem como nos produtos gerados a partir desses precursores após modificações pós-traducionais durante o processo de síntese (Fox e Serrano, 2008) (Fig. 2).

A biossíntese dos venenos de serpentes ocorre em glândulas especializadas, onde a maioria das proteínas é sintetizada no citoplasma das células secretoras. Estas proteínas vão para o retículo endoplasmático rugoso, onde perdem o peptídeo sinal e assumem sua forma protéica através da formação de pontes de dissulfeto, glicosilação e multimerização, como a dimerização de desintegrina (P-IIId/P-IIe) e adição de domínios tipo-lectina (P-IIId). No complexo de Golgi estas proteínas são transportadas em vesículas até o lúmen da glândula secretora, onde a acidificação do meio e de um tripeptídeo de pyrol-glutamato, provavelmente inibem a atividade proteolítica destas enzimas no lúmen da glândula mantendo a integridade celular das proteínas do veneno (Fox e Serrano, 2008; Marques-Porto et al., 2008).

Podemos encontrar no veneno proteinases maduras resultantes da biossíntese de precursores da classe P-Ia, formada pelo domínio catalítico contendo a sequência estendida para a ligação do átomo de zinco com massa molecular entre 20 e 25 KDa. Estas proteinases apresentam um número variável de cisteínas na sua

sequência (4 a 7 resíduos), formando de 2 a 3 pontes dissulfeto. Algumas metaloproteinases pertencentes a este grupo são: atrolisina-C e BaP1, e etc (Fox e Serrano, 2008).

A classe P-II possui o peptídeo sinal, pró-domínio, domínio desintegrina na porção carboxil do domínio catalítico. Estes precursores podem gerar proteínas maduras formadas somente pelo domínio catalítico ou somente o domínio tipo-desintegrina (P-IIa), como a Atrolisina-E, e também proteínas que conservam o domínio desintegrina na proteína madura como as metaloproteinases jerdonitina (monomérica classe P-IIIb) e bilitoxina-1 (dimérica classe P-IIc). Estes precursores podem gerar desintegrinas verdadeiras, contendo a sequência RGD (Arg-Gli-Asp), como na Controtrostatina e a Acostatina, desintegrinas homodímeras e heterodímeras, das classes P-IIId e P-IIe, respectivamente (Fox e Serrano, 2008) (Fig. 2).

As proteinases classe P-III têm massa molecular entre 50 e 110 KDa que apresentam na sua forma madura os domínios: catalítico, tipo-desintegrina e rico em cisteína na região C-terminal (Bjarnason e Fox, 1995). Podem apresentar-se na forma monomérica (P-IIIa) como a jararagina, bothropasina e HF3 e também na forma dimérica (P-IIIc), como VAP1 (Fox e Serrano, 2008).

Ainda podemos encontrar no veneno uma proteína madura composta somente pelos domínios tipo-desintegrina e rico em cisteína, formada pelo processamento proteolítico de precursores da Jararagina da classe P-IIIa, como a Jararagina-C (Moura-da-Silva et al., 2003; Fox e Serrano, 2008).

A antiga classe P-IV, hoje denominada de classe P-IIId, é composta por enzimas que apresentam uma cadeia contendo os domínios encontrados nas proteinases da classe P-III, e mais duas cadeias de lectinas ligadas ao domínio rico em cisteína por pontes de dissulfeto. São representantes destas classes as proteinases: RVV-X (*Russell's viper venom factor X*) e VLFXA (*Vipera lebetina venom factor X activator*), ambas com atividade de ativação do fator X da coagulação (Fox e Serrano, 2008) (Fig. 2).

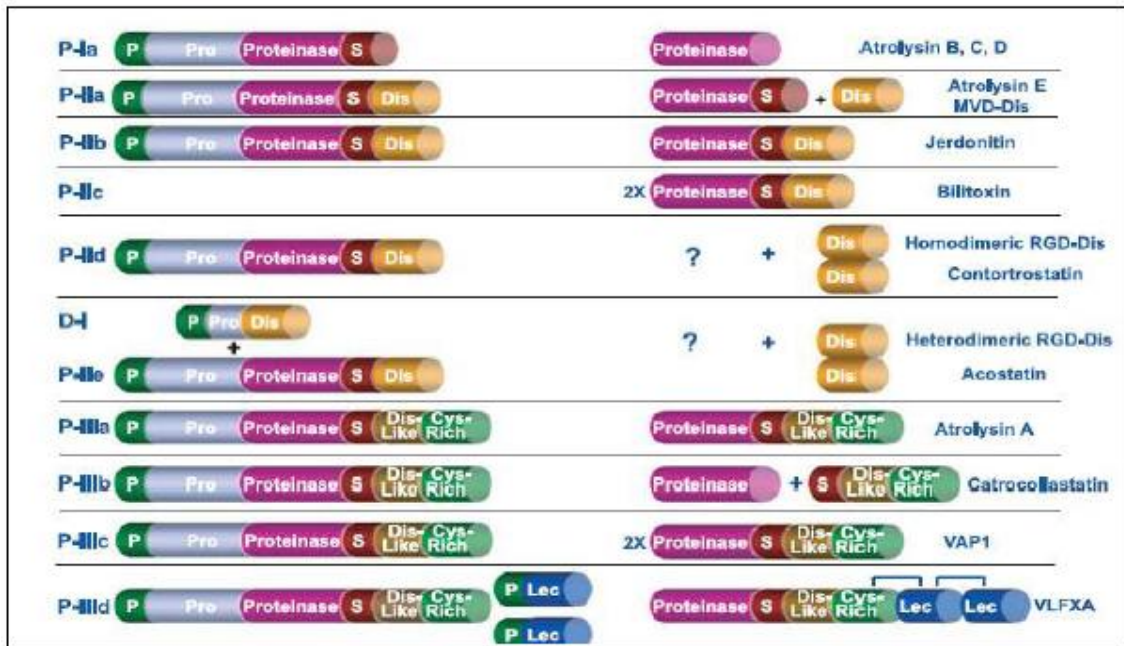


Figura 2. Estrutura dos multi-domínios das metaloproteínas. À esquerda, SVMP na forma inativa, à direita na forma madura. Verde escuro: peptídeo sinal, cinza: pró-domínio, lilás: domínio metaloproteínase, vermelho: espaçador, marrom: tipo-desintegrina, verde claro: rico em cisteína, azul: lectina.
Fonte: Fox e Serrano, 2008.

A estrutura molecular das metaloproteínas classe P-III pode ser composta pelos domínios: peptídeo sinal, pró-domínio (ambos presentes somente no precursor), catalítico (enzimático), desintegrina e rico em cisteína (Kini e Evans, 1992) (Fig. 3).

- Peptídeo sinal: tem como função garantir a correta localização subcelular ou secreção da proteína.
- Pró-domínio: mantém a molécula em estado inativo até ser secretada no lúmen da glândula de veneno. Esta região contém resíduos de aminoácidos altamente conservados, localizado cerca de 20 posições antes do resíduo N-terminal da proteína madura. Estes regulam a atividade enzimática através da interação entre a sulfidrina do resíduo Cys do motivo PKMCGVTQ, com o zinco presente no sítio ativo. Esta interação resulta na inativação do domínio catalítico, um mecanismo chamado *cysteine-switch*, dando origem à proteína madura (Botos et al., 1995)
- Domínio catalítico (Metaloprotease): é o domínio de ligação de íons metálicos. No caso da jararagina, o íon Zn^{2+} apresenta a sequência de aminoácidos HExxHxxGxxHD, altamente conservada nessa família de enzimas responsável pela coordenação do metal próximo à região CIMxP da volta da metionina ou *met-turn*,

característica das metazincinas. Neste domínio são encontrados 7 resíduos de cisteínas (Fox e Serrano, 2008).

- Domínio desintegrina: as desintegrinas verdadeiras contêm um *loop* com o motivo RGD (Arg-Gly-Asp), em conformação apropriada mantida por pontes dissulfetos. Na jararagina e bothropasina, e motivo RGD é substituído por ECD (Glu-Cys-Asp); *ApIVMP-II* de *Agkistrodon* o motivo é KGD (Lys-Gly-Asp). Devido a esta mudança são chamadas de “tipo desintegrina”. Neste domínio são encontradas 15 cisteínas (Jia e Perez, 2010; Fox e Serrano, 2008).

- Domínio rico em cisteína: possui uma grande quantidade de resíduos de cisteínas, que estão interligados por pontes dissulfeto. Este domínio é observado em toxinas hemorrágicas e não hemorrágicas e sua função parece estar relacionada ao direcionamento dessas proteínas aos seus alvos celulares e plasmáticos (Serrano et al., 2007; Jia et al., 2000).

A atividade enzimática da jararagina e bothropasina é dependente da manutenção de suas pontes dissulfetos e da presença do átomo de zinco no seu centro ativo (Assakura, 2000).

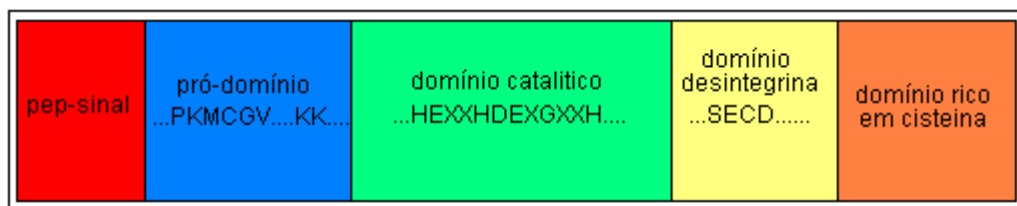


Figura 3. Representação esquemática do precursor da bothropasina.

Fonte: Assakura, 2000.

1.3 A Disintegrin And Metalloproteases (ADAMs)

As desintegrinas apresentam homologia com proteínas expressas em vários tecidos de mamíferos, denominadas ADAMs, para designar a presença, em sua estrutura, dos domínios desintegrinas e metaloproteinases. As ADAMs são uma família de proteínas de superfície celular as quais combinam características de moléculas de adesão e de proteases (Wolfsberg et al., 1995). As ADAMs são, portanto, proteínas homólogas às SVMPS da classe P-III encontradas em mamíferos e outros animais e estão relacionadas a diversos processos fisiológicos como fertilização e diferenciação celular. A diferença mais importante entre as ADAMs e as SVMPS reside no fato de que as primeiras são proteínas transmembrana, enquanto que as últimas são secretadas, apesar de formas solúveis das ADAMs já terem sido

descritas. Outra diferença importante entre as ADAMs e SVMPs é que estas últimas são proteases ativas, enquanto que cerca da metade das ADAMs possui alterações nas suas sequências relacionadas ao zinco, o que resulta na ausência de atividade proteolítica. As ADAMs 1, 8, 9 e 10 contêm resíduos do sítio de Zn para encaixe entre os subdomínios (Kamiguti et al., 1998; Stone et al., 1999).

A estrutura básica das ADAMs está filogeneticamente bem conservada e as mudanças estruturais são produtos da adaptação necessária à sua respectiva função (Yamamoto et al., 1999). Assim, por exemplo, as ADAMs designadas de 1 a 7 se expressam de maneira principal nos órgãos reprodutores, e seu papel durante e espermatogênese é a fusão esperma-óvulo (Blobel, 2005). Cada ADAM é expressa em distintas etapas da espermatogênese. A ADAM 9, ou MDC9 (*Metalloprotease-like, Disintegrin-like and Cysteine rich*) tem sido encontrada em vários órgãos, entre os quais, pulmão a as glândulas mamárias. A ADAM 11 é encontrada em células de órgãos reprodutores e não reprodutores, enquanto as ADAMs 12 e 19, também conhecidas como Meltrinas α e β respectivamente, estão localizadas na musculatura embrionária (estágio neonatal) e tecido ósseo de embriões e adultos. A ADAM 17, chamada de TACE (*Tumor necrosis factor Alfa Converting Enzyme*), e ADAM 10, conhecida por MADAM (*bovine Mammalian A Desintegrin And Metalloprotease*), que cumprem funções importantes no processamento da forma acoplada à superfície celular do precursor do fator de necrose tumoral α (TNF- α), na liberação da forma madura da citocina (Yamamoto et al., 1999).

1.4 Variabilidade dos venenos ofídicos

A maior parte dos venenos, porém, é composta por proteínas e peptídeos com uma variada atividade biológica, compreendendo cerca de 90 a 95 % de peso seco do veneno, é responsável pela totalidade dos efeitos biológicos gerados (Markland, 1998). O fato de existirem transcritos variados é esperado, já que numerosos estudos propõem a ocorrência de processos acelerados de evolução em venenos de muitas espécies de Viperidae (Fry et al., 2003).

A diversidade de proteínas relacionadas é, em geral, resultado de genes parálogos que se duplicam e passam por um processo acelerado de evolução. Genes parálogos são presentes em um mesmo organismo, geralmente famílias multigênicas, que evoluíram para novas funções dentro de um mesmo genoma,

antes ou depois da especiação. Os genomas contêm porções significativas (20 – 80 %) de genes que pertencem à família de parálogos. Genes ortólogos estão presentes em diferentes organismos que se originam de um ancestral comum antes da especiação; eles retêm a mesma função ancestral, tornando a transferência funcional possível (Koonin, 2005; Deshimaru et al., 1996). Isto é comparável com o modelo de nascimento e morte de multigenes dentro das famílias protéicas, que inclui três tipos de eventos: duplicação de genes ou domínios (nascimento), eliminação de genes ou domínios (morte) e inovação, ou seja, novos genes que surgem de sequências não codificantes. Surpreendentemente muitos elementos de transposição, do tipo *retrotransposon*, foram encontrados nas sequências que codificam as toxinas, que estariam contribuindo para um grande número de eventos de nascimentos por duplicação gerando novas atividades tóxicas e diversificação estrutural dentro das famílias de proteínas (Fry e Wuster, 2004; Kordis e Gubensek, 1997).

O aparecimento no mesmo veneno de uma diversidade de proteínas pertencentes à mesma família, porém diferindo umas das outras em seus efeitos farmacológicos, provavelmente resultou da duplicação de um gene ancestral seguida de mutações que levaram ao acúmulo de substituições nucleotídicas não-sinônimas nos exons conferindo novas funções que foram selecionadas durante a evolução. Esta hipótese sugere um papel importante da seleção no que diz respeito à manutenção de alta variação funcional das proteínas de venenos dentro das populações (Richman, 2000). Entretanto, apesar de compartilhar efeitos gerais dentro de uma mesma família de serpentes, os venenos apresentam diferenças de composição dependendo de cada espécie. Datry et al. (1996) demonstraram que os componentes de veneno apresentam uma variação geográfica considerável mesmo dentro de cada espécie de serpente, associada à sua dieta.

1.4.1 Evolução das fosfolipases A2

As isoformas das fosfolipases A2 (PLA2) constituem a maioria dos componentes tóxicos encontrados nos venenos *Trimemeresurus flavoviridis* (Ohno et al., 2003).

O veneno de *Trimemeresurus flavoviridis* contém isoenzimas (PLA2) com uma variedade de atividades fisiológicas. Comparando o cDNA e os genes, verificou-se

que os exons são muito mais diversificados do que os introns, e que o número de substituições não-sinônimas é maior que as sinônimas. Estas características indicam que as isoenzimas PLA2 da glândula de veneno da *T. flavoviridis* evoluíram através de uma evolução acelerada, resultando na diversidade de funções (Fukagawa et al., 1993; Deshimaru et al., 1996).

1.4.2 Evolução de Lectinas tipo-C

A grande variedade de lectinas tipo-C encontrada nos venenos de serpentes pode ser agrupada em três grupos principais, lectinas tipo C, cadeia A, e cadeia B, de proteínas similares a lectina tipo-C, com base na sequência de nucleotídeos dos exons (Ogawa, 2005)

Dados de substituições de nucleotídeos em íntrons comparados aos exons de lectinas revelaram uma maior substituição nos exons que nos introns. Assim, pode-se considerar que esses genes também evoluíram de maneira acelerada, e as serpentes adquiriram em seus venenos uma diversidade de proteínas com atividades biológicas diversas para garantir sua sobrevivência em meio a presas de diversas espécies e características biológicas (Tani et al., 2002).

1.4.3 Evolução das desintegrinas

As desintegrinas fazem parte de uma família de pequenos polipeptídeos (40-100 aminoácidos) ricos em cisteína, descritas inicialmente como potentes inibidoras do receptor plaquetário de fibrinogênio, integrina $\alpha_{IIb}\beta_3$ (Barczyk et al., 2010). Segundo o comprimento e o número de pontes dissulfeto do polipeptídico dividem-se em quatro grupos: curtas, médias, longas e diméricas (Calvete et al., 2007).

As desintegrinas de classe PII são inibidores de integrinas que evoluíram pela neofuncionalização dos domínios desintegrina de genes PIII-SVMP duplicados. As PIII-SVMPs são amplamente distribuídas nas cinco famílias das Colubroidae, Viperidae, Elapidae, Atractaspididae e Colubridae, enquanto que as desintegrinas da classe PII são encontradas apenas nos venenos das serpentes da família Viperidae. Este fato sugere que as desintegrinas emergiram após a separação das famílias Viperidae e Elapidae, porém, antes da separação das subfamílias das Viperidae

(Viperinae e Crotalinae), a cerca de 10 a 20 milhões de anos atrás (Junqueira-de-Azevedo et al., 2006; Cidade et al., 2006).

A minimização de genes pela perda de introns e regiões codificantes, e da estrutura de proteínas pela perda de pontes de dissulfeto sustentam a hipótese da divergência das desintegrinas. Esses autores utilizaram abordagens como Inferência Filogenética e Máxima Probabilidade baseadas na substituição de códons para analisar a evolução da família das desintegrinas. Vários resíduos estão localizados dentro do *loop* de ligação na integrina, e na cauda C-terminal há duas regiões que formam o epítipo de conformação funcional. Um motivo ancestral de reconhecimento de integrinas RDECD poderia ter dado origem, por exemplo, aos módulos RGD e ECD (Juárez et al., 2008).

Ainda no mesmo trabalho, Juárez et al., (2008) compararam a taxa de substituições sinônimas e não sinônimas em genes duplicados. O excesso de substituições não sinônimas sobre as sinônimas indica uma adaptação evolutiva que cria uma grande diversidade de proteínas.

1.4.4 Evolução das metaloproteinases

A evolução molecular das SVMPs foi inicialmente considerada quando as primeiras sequências de cDNAs estavam disponíveis no início da década de 90. As sequências de cDNAs mostraram claramente a relação entre as SVMPs e as ADAMs e que deu origem ao grupo das Reprolisinas para descrever este grupo de proteínas. O aumento do número de sequências, originadas de bibliotecas de cDNAs depositadas em bancos de dados, foi revelando que os venenos de serpente são misturas complexas, contendo vários tipos de SVMP codificadas pelas classes P-I, P-II e P-III de precursores relacionados com as ADAMs; estes geram proteomas complexos no veneno, resultado do processamento e oligomerização de cadeias polipeptídicas. Precursores de sequências SVMPs também foram comparados com as matrizes de metaloproteinases de mamíferos (MMP), gerando uma árvore filogenética que sugere a existência de um ancestral comum entre as SVMPs e MMPs de mamíferos (Fox e Serrano, 2009). A árvore evolutiva das metaloproteinases/desintegrinas ricas em cisteína (MDCs), MMPs de mamíferos e venenos de serpentes, pelo alinhamento das sequências completas de proteínas precursoras, origina diversos grupos. No caso das serpentes, foi gerado um grupo

que representa as desintegrinas com o *motif* RGD, as toxinas hemorrágicas de cadeia longa, como a jararagina e bothropasina, e as hemorraginas de baixo peso molecular, como, Atrolisina E, B e C. A análise sugere que essas proteínas são formadas por “módulos” relativamente conservados. Este estudo também sugere que as metaloproteinases de veneno derivam de um gene ancestral comum após a divergência entre Répteis e Mamíferos (Moura-da-Silva et al., 1996).

De acordo com Calvete et al. (2005) a distribuição monofilética proteínas de mamíferos e proteínas de veneno de serpentes indica que as SVMPs evoluíram relativamente tarde durante a evolução de um gene ancestral comum por especiação (duplicação de genes, seguida por divergência de cópias através da seleção Darwiniana Positiva).

1.5 Integrinas e toxinas

As integrinas estão relacionadas com a integridade das células e tecidos. Constituem uma família de glicoproteínas de superfície. Como consequência da união do ligante às integrinas formam-se complexos de adesão focal no citoplasma, sob a membrana. Funcionam como uma conexão entre a MEC e o citoesqueleto, constituindo-se em regiões transdutoras de sinais que regulam a migração e a proliferação. As funções das integrinas como mediadoras de adesão e fusão celular (célula-célula e célula-matriz), pela ativação e inibição de sinais, se resumem em dois tipos de sistemas: (1) sinalização *inside-out*, originado no interior da própria célula, (2) sinalização *outside in*, proveniente do meio externo para a célula, modulando assim sua função (Coppolino e Dedhar, 2000).

Os receptores integrina são heterodímeros formados pelas subunidades α e β localizadas na membrana (Fig. 4).

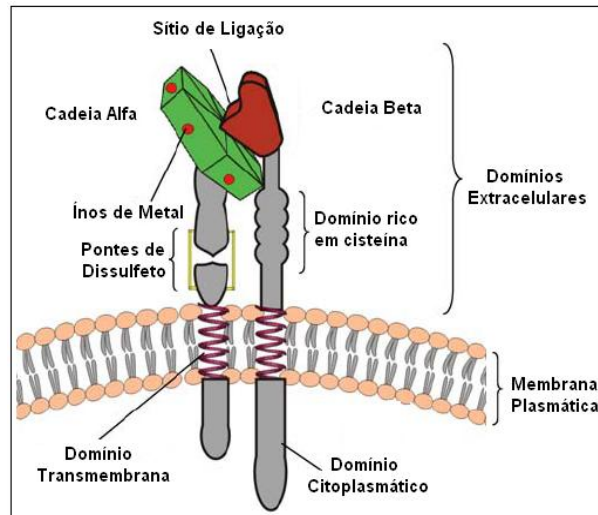


Figura 4. Estrutura da integrina.
Fonte: Kuphal et al., 2005.

As desintegrinas podem interromper uma ligação reversível célula-matriz extracelular, mediada por integrinas. Inibem a agregação plaquetária por se ligarem especificamente ao receptor de fibrinogênio, e são capazes de inibir a adesão de células tumorais à matriz extracelular por competição específica pelos receptores de integrinas (Tripathi et al., 1994; Chiang et al., 1994; Sheu et al., 1997).

O domínio desintegrina do veneno de serpente liga-se nas integrinas, podendo promover transdução de sinal, induzir reorganização do citoesqueleto e alterações do complexo de adesão focal. Estes eventos podem causar diminuição da habilidade de migração, invasão e formação de metástase e angiogênese (Marcinkiewicz et al., 2003).

1.6 Atividade das metaloproteinases sobre tumores

Existem vários estudos que demonstram a ação das desintegrinas sobre os tumores, através da ligação do domínio desintegrina com as integrinas. Com isso, as células tumorais não podem migrar através das proteínas da Matriz Extracelular nem constituir novos focos distantes (metástases). A adesão mediada por integrinas é crucial para a disseminação metastática. Este processo segue as seguintes etapas: 1) disseminação das células tumorais a partir do tumor primário; 2) interação destas células às integrinas das plaquetas; 3) adesão das células tumorais ao endotélio vascular e/ou membrana basal sub-endotelial; 4) proliferação e supressão da apoptose; 5) angiogênese (Akiyama et al., 1995).

A salmosina (*Agkistrodon halys brevicaudus*), uma desintegrina-RGD, reduziu *in vitro* a proliferação de células B16F10 com o colágeno de forma dose dependente; *in vivo* inibiu a colonização de tumores no pulmão (Kang et al., 2000). A desintegrina ristostatina que contém a sequência RGD é capaz de inibir em 87 % a formação de metástases pulmonares em camundongo (Danen et al., 1998). A desintegrina-RGD rhodostomina (*Calloselasma rhodostoma*), que se liga à integrina $\alpha_2\beta_1$ em células endoteliais, inibiu de maneira dose-dependente a resposta angiogênica induzida por bFGF (*basic fibroblast growth factor*), um dos principais fatores de indução de angiogênese, incluindo proliferação, migração, invasão e diferenciação em HUVECs (*human umbilical vein endothelial cells*) *in vivo* e *in vitro* (Yeh et al., 2002).

A catrocollastatina-C, halyssetina-C e jararagina-C inibem a agregação plaquetária induzida por colágeno, sendo que a jararagina-C também inibe a agregação induzida por ADP e, ainda, a adesão de células transfectadas com a integrina $\alpha_2\beta_1$, e das células ECV-304 e HUVEC ao colágeno tipo I (Moura-da-Silva et al., 2001). Alternagina-C (*Bothrops alternatus*) inibe a adesão mediada pelo colágeno tipo I em células transfectadas (K562- $\alpha_2\beta_1$), de forma dose dependente (Souza et al., 2000), e também de fibroblastos humanos, e de células tumorais das linhagens ECV-304, HeLa e MDA-MB-231 a este mesmo ligante (Cominetti et al., 2004).

Corrêa Jr. et al. (2002), em nosso grupo no Laboratório de Genética, no Instituto Butantan, verificaram o efeito da jararagina na forma nativa e na forma com o domínio catalítico inativado por orto-fenantrolina (remoção do átomo de zinco), em linhagens de células de melanoma humano. Foram observadas alterações *in vitro* na morfologia, adesão, migração e invasão das células pré-tratadas, bem como a diminuição significativa do número de metástases no pulmão dos camundongos, injetados com células malignas pré-tratadas com a toxina. Maria et al. (2003) observaram que a injeção semanal em camundongos com doses sub-letais de jararagina (12 ng) provoca significativa leucopenia, aumento do número de mastócitos e eritrócitos na fase inicial, mas não causa hemorragia nem inflamação.

Camundongos inoculados com as linhagens de carcinoma de cólon SW613-B3 (não tumorigênica) e SW613-12A1 (tumorigênica) tratadas com bothropasina apresentaram lesões cutâneas com alopecia e deposição de pigmentos escurecidos. A bothropasina nativa também se mostrou eficaz na inibição da adesão celular nessas linhagens SW613-B3 e SW613-12A1 (Ribeiro, 2003).

1.7 Jararagina e Bothropasina

A análise do transcriptoma da *Bothrops jararaca* identificou que das toxinas encontradas no veneno 52,6 % correspondem às metaloproteínas, 28,5 % às serino proteases, 8,3 % às lectinas tipo-C e aos peptídeos potenciadores de bradicininas 6,2 % (Cidade et al., 2006).

Conforme já mencionado, a jararagina e a bothropasina são SVMPs da classe P-IIIa isoladas do veneno de *Bothrops jararaca*, subfamília Reprolisina, com atividade catalítica hidrolase do tipo endopeptidase, dependentes de zinco (Bjarnason e Fox, 1995). Possuem ação hemorrágica e capacidade de inibir a agregação plaquetária, ao bloquear a ligação do colágeno à integrina $\alpha 2\beta 1$ presente na membrana de plaquetas. A jararagina se liga ao domínio I da cadeia $\alpha 2$ e/ou cliva a cadeia $\beta 1$ desta integrina. A jararagina interfere na função plaquetária de duas maneiras distintas: (1) por degradação dos diferentes receptores plaquetários e proteínas adesivas na hemostasia; e (2) interferência não enzimática (mediada pelo domínio desintegrina) com função de receptores da adesão plaquetária (Muniz et al., 2008).

A jararagina foi isolada por Paine et al. (1992) e caracterizada como uma metaloproteína/desintegrina de alto peso molecular, 52 KDa, com 421 aminoácidos na proteína madura. Esta análise foi feita a partir de *pool* de cDNAs de glândulas de veneno, de aproximadamente 100 serpentes.

A bothropasina é uma metaloprotease hemorrágica de 48 kDa, isolada pela primeira vez em 1982 (Mandelbaum et al., 1982). Assim como a jararagina, a bothropasina também é constituída por 421 resíduos de aminoácidos e apresenta atividade hemorrágica. Assakura et al. (2003), analisaram a sequência de aminoácidos a partir de *pool* do tecido glandular de 5 espécimes de *Bothrops jararaca*.

A jararagina e a bothropasina apresentam alto grau de identidade de sequência em relação aos seus cDNAs, sendo 95,2% na região pró-domínio, 91,2% no domínio catalítico, 98,9% no domínio desintegrina e 100% na região rica em cisteína (ANEXO A). As diferenças de sequência entre a bothropasina e a jararagina estão distribuídas ao longo de seus cDNAs. Alguns autores levantaram a hipótese de que as diferenças entre os seus cDNAs, assim como os que codificam outras

metalo-proteínas semelhantes, são decorrentes de um processo evolutivo acelerado (Assakura, 2000).

Assakura et al. (2003) tentaram a expressão da pró-bothropasina recombinante no vetor *pPIC9* em sistema eucariótico utilizando levedura *P. pastoris*. O precursor da bothropasina recombinante compreende os domínios: pró, catalítico, tipo desintegrina e rico em cisteína. Não houve a expressão da proteína recombinante, o que foi atribuído à ação da enzima proteolítica da própria *P. pastoris*. Em seguida tentou-se a expressão da pró-bothropasina recombinante em *E. coli*, e o vetor escolhido foi *pGEX-4T*. O *Western Blot* indicou ausência da expressão. Foi testada a expressão de domínios isolados da bothropasina. Somente com o Pró-domínio foram obtidos 5 mg/L em cultura do. Apenas o domínio catalítico teve uma acentuada queda no nível de expressão de 25-20 µg/L. Não foi experimentada a expressão dos domínios tipo-desintegrina e rico em cisteína.

Moura-da-Silva et al. (1999) expressaram em *E. coli* o domínio tipo desintegrina e rico em cisteína da jararagina mostrando que a proteína recombinante inibe a agregação plaquetária induzida por colágeno de forma dose dependente.

A análise da sequência de aminoácidos mostrou 19 resíduos não idênticos entre as toxinas jararagina e bothropasina (Fig. 5), o que para alguns autores seria evidência de serem isoformas, codificadas por um mesmo gene. Além disso, apenas os cDNAs foram descritos, não havendo até a realização deste trabalho, nenhum conhecimento sobre o DNA genômico, com a caracterização dos exons e introns.

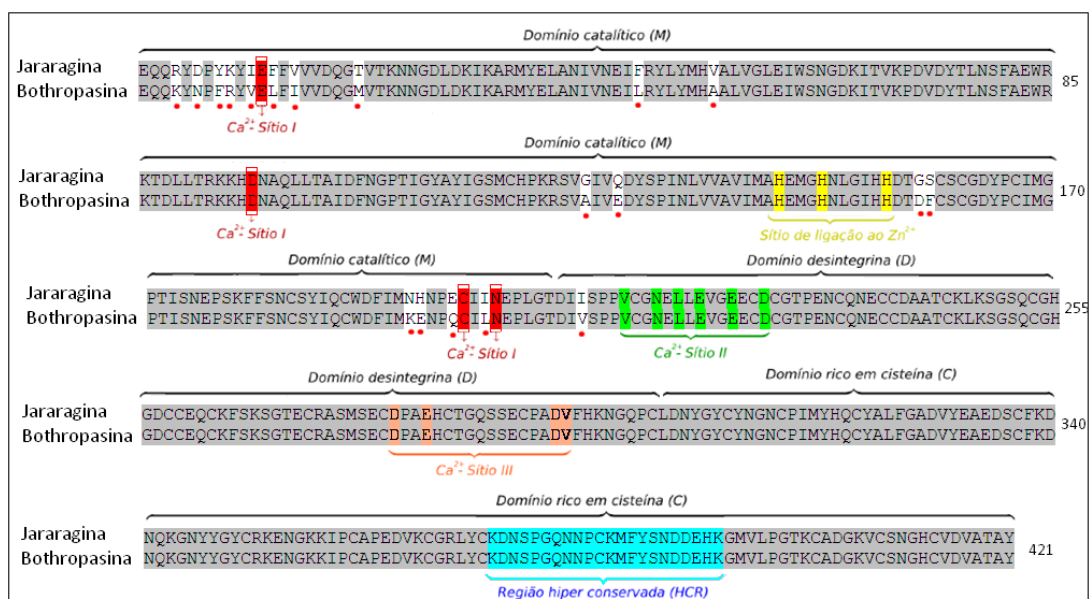


Figura 5. Alinhamento dos aminoácidos da jararagina e bothropasina. São encontrados 19 resíduos não idênticos.

Fonte: Muniz, 2007.

A análise da estrutura tridimensional do domínio catalítico da bothropasina mostra um esqueleto de dois subdomínios organizados de forma semelhante a outras SVMP, incluindo os sítios de ligação de cálcio e zinco (Fig. 6). O domínio metaloproteinase é formado 6 folhas β , 5 fitas paralelas (S1, 2, 3, 4 e 6) e uma anti-paralela (S5), e 5 hélices α (H1, 2, 3, 4 e 5). O domínio desintegrina é composto principalmente pelo *loop* estabilizado por 7 pontes dissulfeto e dois íons de cálcio. A região ECD está no *loop*, estruturalmente relacionada com a região RGD das desintegrinas, derivada das SVMP PII. O motivo ECD é estabilizado por pontes de dissulfeto Cys²⁷⁷-Cys³¹⁰ (entre os domínios desintegrina e rico em cisteína) e por um íon de cálcio. As diferenças entre os domínios desintegrina, e tipo-desintegrina, podem estar envolvidas na seleção de ligação ao alvo, o que pode gerar diversidade de substratos ou especificidade para o domínio catalítico (Muniz et al., 2008).

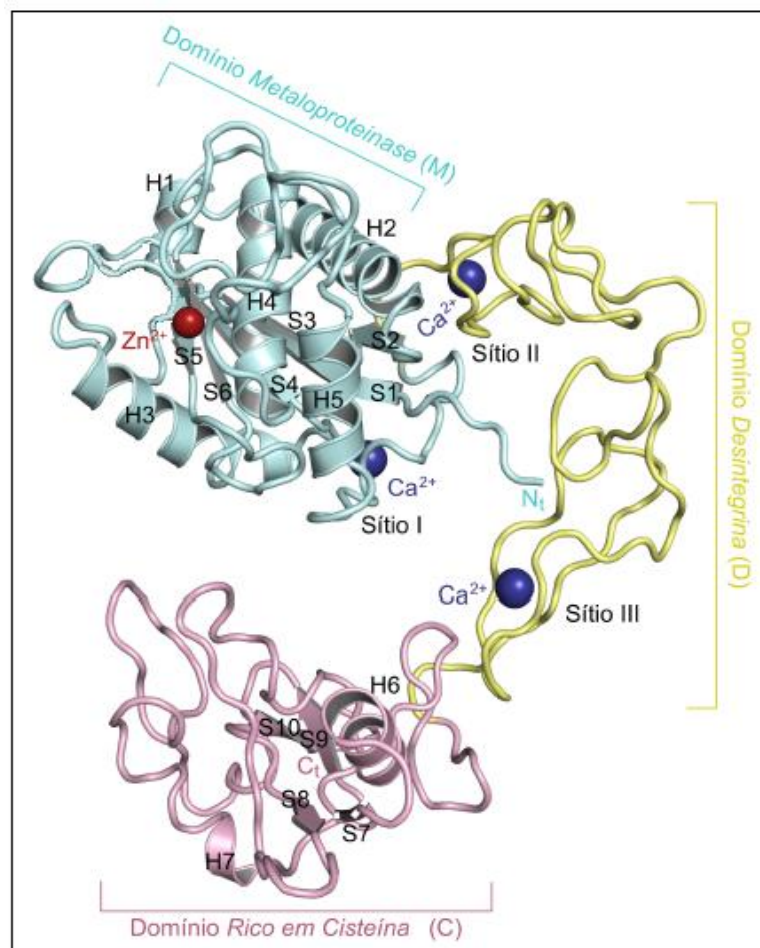


Figura 6 Estrutura tridimensional da bothropasina. Domínio Metaloproteinase apresenta 6 folhas β e 5 hélices α e o domínio rico em cisteína 4 folhas β e 2 hélices α .
Fonte: Muniz et al., 2008.

1.8 Estrutura genômica de genes referentes a outros tipos de toxinas

Muitas toxinas hemorrágicas e neurotóxicas de várias espécies de serpentes já foram isoladas e parcialmente caracterizadas através do sequenciamento de nucleotídeos dos cDNAs que as codificam, ou totalmente caracterizadas a partir do DNA genômico.

A crotamina é um polipeptídeo pequeno, encontrado no veneno da *Crotalus durissus terrificus*. O gene que codifica essa toxina tem 1.8 Kb, constituído por 3 exons (50 pb, 124 pb e 13 pb) e 2 introns (898 pb e 145 pb). A identificação dos exons e introns foi feita com base no cDNA que possui ~340 nucleotídeos incluindo a sequência do peptídeo sinal (Rádis-Baptista et al., 2003).

As neurotoxinas BM10-1 e BM10-1-like, isoladas do veneno da *Bungarus multicinctus*, foram caracterizadas a partir do produto de PCR de DNA genômico. A análise completa mostrou que ambas têm a mesma estrutura: 3 exons separados por 2 introns. O 1º intron tem 1233/1178 pb e o segundo 547/539 pb, em BM10-1 e BM10-1-like, com 98,8% e 95% de identidade de sequência, respectivamente. Quanto aos exons, o 1º (84 pb) e o 2º (114 pb) de ambas as neurotoxinas são essencialmente idênticos, mas a sequência do 3º exon apresenta vários pontos de mutação e tamanhos diferentes: o exon da BM10-1 apresenta 93 pb enquanto que a BM10-1L tem 84 pb (Chang et al., 2003).

Fujimi et al. (2003) estudaram onze genes codificadores de toxinas em várias espécies de serpentes (gêneros *Laticauda*, *Bungarus* e *Naja*). Mostraram que o 1º intron era dividido em sub-regiões ou módulos, comuns aos diferentes genes em diferentes combinações. Sugeriu a existência de um íntron hipoteticamente presente em um ancestral comum. A estrutura *three-finger* dessas neurotoxinas diverge de um ancestral comum através de dois tipos de precursor (longas e curtas). Em uma etapa posterior da evolução cada gene teria sofrido um acúmulo de mutações nos exons, especialmente no exon 2 que pode ter causado aumento na diversidade e funções (Fig. 7).

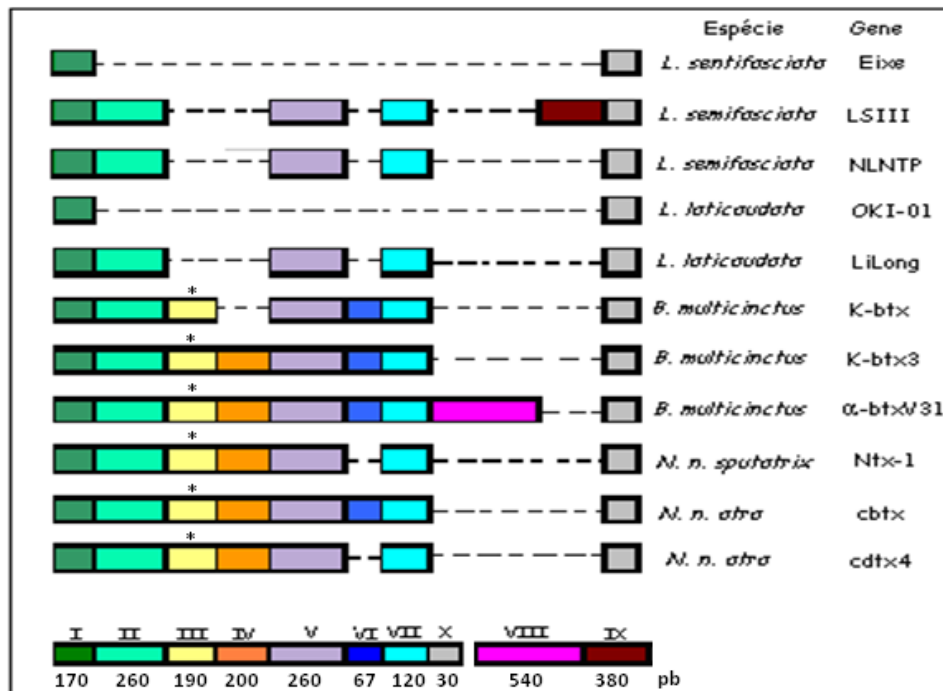


Figura 7. Organização genômica do 1º intron de diferentes genes codificadores de toxina de cinco espécies de serpentes de gênero *Laticauda*, três do gênero *Bungarus* e três do gênero *Naja*. A análise desses introns revelou a possibilidade da existência de 10 sub-regiões (I a X); 8 sub-regiões estariam num intron ancestral comum, as demais (VIII e IX) teriam sido adicionadas em duas espécies. As espécies de serpentes citadas possuem números variados dessas sub-regiões. As regiões I e X são as únicas conservadas em todas as espécies citadas. Os asteriscos indicam regiões que contêm seqüências que codificam snoRNA.

Fonte: Fujimi et al., 2003.

Proteínas *three-finger* são estruturalmente relacionadas e exibem várias funções. A análise filogenética da toxina-1 da truta arco-íris (*Oncorhynchus mykiss*) é classificada como neurotoxina de cadeia de curta (60-65 resíduos), contém 3 exons e 2 introns, como outros membros da família *three-finger*. Este é o primeiro relato de um gene tipo-toxina descrito em teleósteos (Georgaka et al., 2007).

Serpentes elapides australianas, como a *Tropidechis carinatus*, têm dois genes parecidos, mas que desempenham papéis diferentes. A trocari-D existe no veneno e atua como toxina, enquanto o fator de coagulação FX existe no plasma e desempenha um papel na homeostasia. Os genes que codificam essas proteínas, expressos de forma altamente específica, têm oito exons idênticos. Os íntrons têm 92-99 % de identidade de seqüência. O primeiro intron do gene trocari-D é quase idêntico ao do gene FX, com exceção de duas deleções (nas posições 255 e 1406) e três inserções (nas posições 214, 1975 e 2174). Quanto ao promotor gênico, observa-se que no caso do FX humano, FX murino, e FX de *Tropidechis*, existem vários elementos comuns, principalmente considerando FX humano e de

Tropidechis. Esses elementos também aparecem no promotor do gene trocarin-D que, além disso, apresenta uma inserção de 264 bp. Esta inserção poderia ser responsável pela especificidade de expressão do gene trocarin-D na glândula de veneno (Reza et al., 2007).

1.9 Introns

Genes eucarióticos apresentam-se divididos em exons e introns. Aproximadamente 97 % do genoma humano compreendem DNA que não codifica proteínas conhecidas até agora, incluindo os introns. Após a transcrição do gene em pré-mRNA ocorre o processo de remoção dos introns e emenda dos exons em uma molécula única de RNA maduro. O processo, denominado *splicing* dá-se graças ao *spliceossomo*, formado de pequenos RNAs nucleares (snRNAs) e proteínas, encontrado no núcleo das células eucarióticas. Os introns são removidos dos pré-mRNAs através de duas quebras e duas religações fosfodiéster entre nucleotídeos dos exons e do intron, ou seja, duas transesterificações (Fig. 8). O intron em forma de laço é removido (Lewin, 2000). A molécula resultante é composta somente por exons incluindo regiões não traduzidas em 5' e 3'.

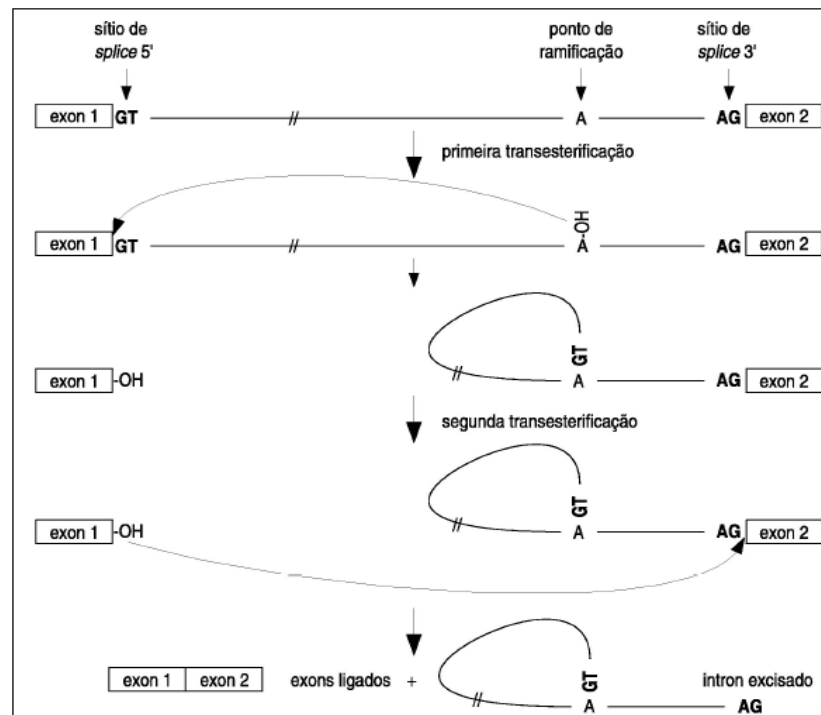


Figura 8. Mecanismo de *Splicing* nuclear
Fonte: Lewin, 2000.

São conhecidos três tipos de introns: os introns de *splicing* nuclear, e os introns auto-catalíticos, grupos I (encontrados em genes mitocondriais de fungos, genes mitocondriais e de cloroplastos de vegetais superiores e genes de eubactérias) e o grupo II (encontrados em genes que codificam rRNA no núcleo de eucariotos inferiores, em genes mitocondriais de fungos, em genes do fago T4 e em eubactérias) (Daí e Zimmerly, 2003; Haugen et al., 2005; Rogers et al., 1989).

A maioria dos genes é editada da mesma forma, indicando que a célula reconhece precisamente a borda entre exons e introns. Vários organismos têm a borda exon/intron marcada pela presença de nucleotídeos conservados: *gt* na extremidade 5' do intron e *ag* na extremidade 3' (~ 99,6% das bordas exon/intron de mamíferos tem esta marcação). Em mamíferos o sítio de ramificação (*branch site*) que abriga a adenina pode ser bastante degenerado e sua posição pode ser bem variável em relação à distância dos sítios de *splice* em 5' e 3' (Sheth et al., 2006).

1.9.1 Origem e funções dos introns

Desde a descoberta dos introns em 1977, questiona-se qual sua função, quando e como eles surgiram e porque introns de *splicing* nuclear existem em eucariotos e não em procariotos. Basicamente, existem duas hipóteses para explicar a origem dos introns: introns-*early* e introns-*late*. A primeira hipótese assume que os introns já estavam presentes nos genes do progenitor (ancestral comum de procariotos e eucariotos). Os introns teriam sido perdidos nos procariotos por um processo de redução de tamanho do genoma, resultado da pressão seletiva em favor da replicação rápida do DNA. Já os introns-*late* teriam sido adicionados posteriormente durante a evolução, após a divergência entre eucariotos e procariotos (Mattick, 1994).

Em relação às suas funções, recentemente estão sendo encontradas inúmeras sequências intrônicas que codificam microRNAs, que são responsáveis por mediar o silenciamento gênico através de interferência (RNAi) (Lin e Ying, 2006). Os miRNAs foram primeiramente descobertos em *Caenorhabditis elegans* (Hannon, 2002) como fragmentos nativos de RNA que modulam uma grande variedade de vias de regulação gênica de vários organismos, demonstrando a manutenção evolutiva do silenciamento gênico (Ying e Lin, 2006). Muitos snRNAs (*small nuclear*

RNAs) inclusive os que participam do *splicing* são codificados por introns (Lewin, 2000)

Nott et al. (2003) verificaram em eucariotos superiores que a remoção dos introns pelos spliceossomos pode afetar a expressão gênica em diferentes níveis, incluindo poliadenilação, exportação do mRNA, eficiência de tradução e a taxa de decaimento do mRNA. Foi construído um sistema repórter baseado na luciferase para monitorar os efeitos de introns individuais e suas posições dentro do gene, analisando a expressão dessa proteína em células HeLa. Foi feita a análise quantitativa da expressão de um vetor contendo parte (exon 6 + intron 6 + exon 7) do gene TPI (*triose phosphate isomerase*) humana posicionado em duas regiões diferentes (5' e 3'). O resultado revelou que a presença do intron promove o aumento da atividade do gene repórter dependendo de sua posição na sequência. Após 42 h de transfecção o vetor com o intron na posição 5' foi expresso 29 vezes mais do que na posição 3', um aumento de 8 vezes, comparado com o controle (sem intron) (Fig. 9). Este experimento demonstra a importância dos introns na expressão gênica.

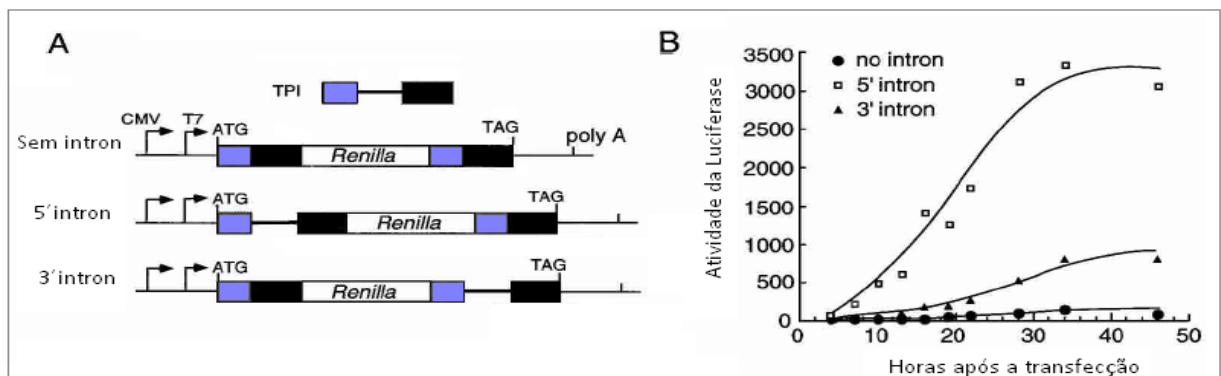


Figura 9. Construção de um vetor utilizando introns. Foi feita uma construção colocando os exons 6 e 7 da *triose fostafo isomerase* (TPI) em ambas extremidades do sistema repórter *Renilla*/luciferase. a) sem o intron 6; b) com intron 6 em 5'; c) com intron 6 em 3'. B. Monitoramento da atividade da luciferase de 4 a 46 h após a transfecção, indicando que o vetor com o intron em posição 5' tem uma maior atividade da luciferase.

Fonte: Nott et al., 2003.

2 OBJETIVO

Clonagem, sequenciamento e caracterização dos genes de metaloproteinase da classe P-IIIa, a partir do DNA genômico de *Bothrops jararaca*, utilizando como referência os cDNAs da bothropasina e jararagina.

3 JUSTIFICATIVA

A obtenção da sequência completa dos genes codificadores da jararagina e bothropasina a partir do DNA genômico, portanto incluindo os introns, visa esclarecer se ambas são codificadas pelo mesmo gene, ou por genes oriundos de duplicação, contribuindo para a compreensão da evolução desta família gênica.

4 MATERIAIS e MÉTODOS

4.1 Extração de sangue de *Bothrops jararaca*

Sangue (2,5 ml) proveniente de um exemplar de *Bothrops jararaca* foi extraído previamente por técnicos especializados conforme normas do setor de Herpetologia do Instituto Butantan. Este sangue foi estocado no gelo em tubo cônico com EDTA até o momento de uso.

4.2 Técnica para extração do DNA genômico

O sangue de *Bothrops jararaca* foi centrifugado a 1000 rpm (centrífuga Sorvall RC5B, rotor AS 600) durante 10 minutos, 4 °C. O soro foi desprezado. Ao *pellet* foram adicionados 10 ml de solução digestiva (NaCl 0,4 M; EDTA 0,1 M pH 8,0; proteinase K 100 µg/ml e SDS 1%), e a suspensão foi incubada a 50 °C por 3 h; em seguida a concentração de NaCl foi corrigida para 1 M. Esta solução foi transferida para um tubo Corex e extraída com igual volume da solução fenol/clorofórmio/álcool isoamílico (50:48:2). Em seguida, a solução ficou sob suave agitação por 30 min e foi centrifugada a 10.000 rpm, 4 °C. O sobrenadante foi transferido para um novo tubo Corex e precipitado durante 12 horas a -20 °C com dois volumes de etanol 100%. Após este período, a solução foi novamente centrifugada a 10.000 rpm a 4 °C por 15 min. O sobrenadante foi desprezado e o *pellet* lavado com 10 ml de etanol 70%, centrifugado a 10.000 rpm a 4 °C por 15 min e seco em um sistema a vácuo. O *pellet* foi ressuspendido em 2 ml de TE (Tris 10 mM, EDTA 1 mM), 1 µl de RNase 10 mg/ml e incubado a 37 °C por 1 h. Em seguida foi adicionado NaCl para concentração final 1 M. A solução foi novamente extraída com igual volume de fenol/clorofórmio/álcool isoamílico, e submetida a suave agitação manual por 30 min. Após isto, foi centrifugada a 10.000 rpm durante 15 min a 4 °C. O sobrenadante foi transferido para um novo tubo Corex, precipitado com 2 volumes de etanol 100% e mantido a -20 °C *overnight*, centrifugado a 10.000 rpm por 15 min a 4 °C, e desprezado o sobrenadante. O *pellet* foi finalmente lavado em etanol 70%, centrifugado, seco nas mesmas condições acima, ressuspendido em TE, quantificado em espectrofotômetro e gel de agarose 1%, aliquoteado e mantido a 4 °C.

4.3 Estratégia de amplificação

Partes contíguas dos genes em questão foram amplificadas pela técnica de PCR a partir de DNA genômico, utilizando *primers* desenhados conforme as regiões que codificam os respectivos domínios, tendo por base as sequências publicadas dos cDNAs de jararagina e bothropasina (Quadro 1). Os *primers* foram determinados a partir de sequências comuns a ambos os cDNAs para garantir o anelamento dos dois supostos genes, jararagina e bothropasina. As condições de PCR foram estabelecidas inicialmente conforme o número de (G+C) e (A+T) presentes nos *primers* e depois foram otimizadas. Foram utilizados os termocicladores *Gene* e *PCR Sprint Thermal Cycler (Thermo Eletronic Corporation)*. Condições do PCR: 94 °C 1 min; 54 °C 1 min; 72 °C 2 min (35 ciclos); 72 °C 5 min (extensão, 1 ciclo). Reação: *buffer* 10x, 50 mM MgCl₂, 10 mM dNTPmix, 50-100 ng DNA, 5 U *Taq DNA Polymerase* (qsp 30 µl). A concentração estoque dos *primers* foi 100 µM, e a utilizada no PCR foi 10 µM.

Quadro 1- Sequências dos primers.

<i>Primers</i>	Seqüências 5'-3'	T °C anelamento	Região
Pbd	GCACGAGAGAACTCAGATTGG	64 °C	Exon/ pep-sinal
Pbr	ATTCACGTTCCCAGATTCCA	58 °C	Exon/ pep-sinal
Pbdd	GCATGTTAAAGTGGGAGAG	56 °C	Intron/ pep-sinal
Pbrr	GGACGTACAATCACAACCTG	56 °C	Intron/ pep-sinal
PbirD	GGAATCTGGGAACGTGAATG	60 °C	Exon/pró-domínio
Pad	GAGCAGTTCAGCCAAAGTA	60 °C	Exon/pró-domínio
Pair	GGTGAAGGACCACTGGCT	62 °C	Exon/pró-domínio
Paid	GCCAGTGGTCCTTCACCT	57 °C	Exon/pró-domínio
Par	GTCAGCATCATTCTCGATG	60 °C	Exon/ pró-domínio
P1d	CATGGACGCATCGAGAATGA	60 °C	Exon/ pró-domínio
P1r	ACCCATCAAAAAGGCCTCTC	60 °C	Exon/ pró-domínio
P1rr	CTCAGAGCTGAGTTTCATTC	58 °C	Intron/pró-domínio
P0d	GTAGAAAAGGAGGATGAGGC	60 °C	Exon/ pró-domínio
P0dd	GTAGTTTGGAGATTGAA	48 °C	Intron/pró-domínio
P0rr	CATTTGTTCTTCCTGCTTAC	56 °C	Intron/pró-domínio
P0r	GAATGTATGAACTTGCCAAC	56 °C	Exon/ catalítico
P0dr	GAATGTATGAACTTGCCAACA	56 °C	Exon/ catalítico
P2r	GAGATAAGATTACCGTGAAG	56 °C	Exon/ catalítico
P2dx	CACTGATATCTTCTTTCTCTC	58 °C	Intron/catalítico
P4d	GAGATAAGATTACCGTGAAG	56 °C	Exon/ catalítico
P4ir	GTCAATTGCTGTGAGTAACTG	58 °C	Exon/ catalítico
P4irD	GTTACTCACAGCAATTGA	52 °C	Exon/ catalítico
P4r	CAATAAATCTTGTGGTTGC	52 °C	Exon/ catalítico
P11d	GCATTCATCATGACACAG	52 °C	Exon/ catalítico
P11Dr	CTGTGTCATGATGAATGC	56 °C	Exon/ catalítico
P11r	GGATATAACTACAATTGCTG	54 °C	Exon/ catalítico

P11dd	CTTGGGGTTTTTCCTGCCT	56 °C	Intron/ catalítico
P11rr	CTGGTGGCCATTAATAGAC	58 °C	Intron/ catalítico
P9d	CAGCAATTGTAGTTATATCC	54 °C	Exon/ catalítico
P9r	AGTGCCACAGTCACATTC	54 °C	Exon/ desintegrina
P6d	GAATGTGACTGTGGCACT	54 °C	Exon/ desintegrina
P6r	GCTCACAACAGTCTCCATG	56 °C	Exon/ desintegrina
P7d	CATGGAGACTGTTGTGAGC	58 °C	Exon/ desintegrina
P7r	CCAGTGCAGTGTTTCAGCC	58 °C	Exon/ desintegrina
P12d	GGCTGAACACTGCACTGG	58 °C	Exon/ desintegrina
P12r	GTAACCGTAGTTATGTAGG	54 °C	Exon/ desintegrina
P10d	CTCTCTTTGGTGCAGATGTT	58 °C	Exon/ cys-rich
Pydd	CTTACATTGACACTCGTC	52 °C	Intron/ cys-rich
Pyrr	GATGGTTATCTTCAAGCTG	54 °C	Intron/ cys-rich
P10r	CTTCTGGTGCACATGGAATC	60 °C	Exon/ cys-rich
P8d	GATTCCATGTGCACCAGAAG	60 °C	Exon/ cys-rich
P8id	GCAAAGATAATTCACCTGG	58 °C	Exon/ cys-rich
P8idR	CCAGGTGAATTATCTTTGC	54 °C	Exon/ cys-rich
P8r	GCACATTTTGTTCAGGAAG	58 °C	Exon/ cys-rich
P5d	TCCTGGAACAAAATGTGCAG	58 °C	Exon/ cys-rich
P5r	TCAAAGAGATCCATCTGCC	60 °C	Exon/ não traduzida
P3d	CCATCTGCCTGCATCTTACT	60 °C	Exon/ não traduzida
P3r	GCTGCTTCCATCAATAAACT	56 °C	Exon/ não traduzida

4.4 Eletroforese e eluição

A técnica de eletroforese foi usada tanto para quantificação de DNA quanto para resolver os fragmentos amplificados por PCR. O gel de agarose 1,5% em tampão TEB (Tris 0,9 M; EDTA 0,02 M; ácido bórico 0,9 M) foi utilizado para a migração dos produtos sob tensão 10 V/cm, corado com brometo de etídeo (EtBr) e fotografado utilizando câmara fotográfica Kodak *El Logic 100 Imaging System* sob luz UV. Os fragmentos amplificados foram eluídos do gel conforme instruções do *Kit Biotools Bioclean* (*Biotools* do Brasil) e utilizados para clonagem e sequenciamento.

4.5 Clonagem molecular

A fim de obter clones para sequenciamento, estoque, e para futuros estudos, os fragmentos amplificados foram clonados em vetor *pGEM T-easy* (*Promega*) em *E. coli* XL1-Blue e DH-5 α conforme instruções do fabricante. Os insertos foram ligados aos plasmídeos abertos em presença de ligase e usados para transformação das bactérias competentes. Os clones recombinantes selecionados em placas contendo ampicilina foram submetidos à mini-preparação (miniprep) a fim de extrair os plasmídeos recombinantes. Estes foram analisados pela digestão com a enzima

de restrição *Eco* RI, para avaliar a presença do inserto, e também submetidos à PCR, usando *primers* universais (T7 e SP6) que anelam nos plasmídeos a fim de confirmar a presença e tamanho do inserto para posterior sequenciamento (Fig. 10 e 11).

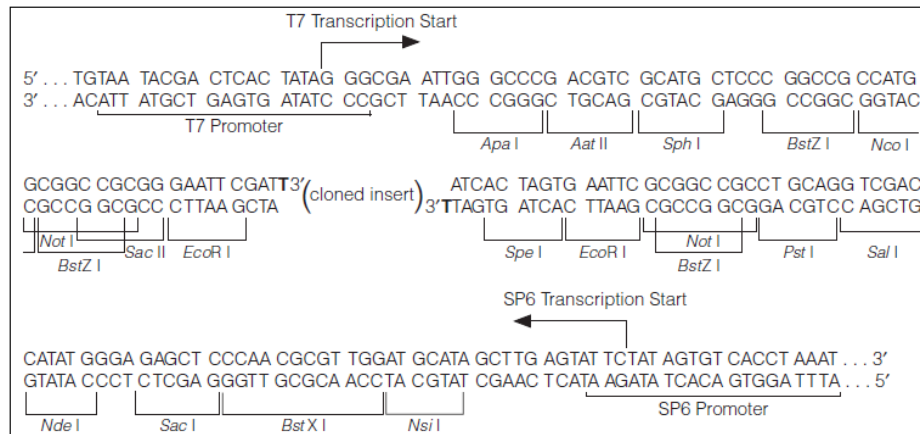


Figura 10. Sequência dos sítios de clonagem do plasmídeo *pGEM T-easy Vector* (Promega).

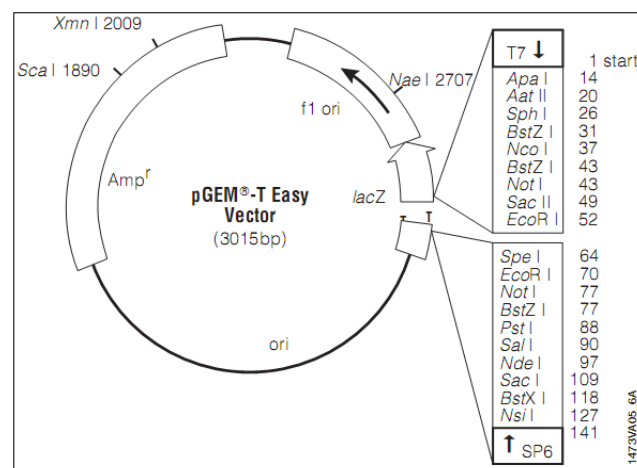


Figura 11. Esquema do *pGEM T-Easy Vector*.

4.5.1 Preparação de bactérias competentes

Um pré-inóculo de *E. coli* (uma colônia) foi feito em 5 ml de meio LB (triptona 1%, extrato de levedura 0,5%, NaCl 170 mM). A cultura foi mantida em *shaker* a 37 °C *overnight*. A seguir 1,0 ml desta cultura foi inoculado em 50 ml de meio LB e a cultura foi mantida sob agitação até atingir D.O.=0,3 a 600 nm. As bactérias foram centrifugadas a 5.000 rpm por 5 min (rotor Sorvall AS 600). O *pellet* contendo bactérias foi ressuscitado em 50 ml CaCl_2 50 mM. As bactérias foram incubadas no gelo por 30 min, centrifugadas a 5.000 rpm por 5 min, ressuscitadas em 1,25 ml de CaCl_2 50 mM, e mantidas na geladeira até o uso, no máximo uma semana.

4.5.2 Transformação de bactérias competentes

Os plasmídeos recombinantes (5 μ l) mais 95 μ l de TE foram adicionados a 100 μ l de bactérias competentes. A mistura foi mantida por 20 min a 0 °C, e depois submetida a choque térmico por 5 min a 42 °C. Em seguida, foram adicionados 800 μ l de meio LB (pré-aquecido a 37 °C), e a cultura foi mantida a 37 °C por 45 min.

4.5.3 Plaqueamento

Foram utilizadas para o plaqueamento placas de petri com meio LB-agar 1,5% contendo ampicilina (100 μ g/ml). Foram aplicados sobre as placas LB/amp 100 μ l de bactérias transformadas pelos plasmídeos recombinantes. Após o plaqueamento, as placas foram mantidas em estufa a 37 °C *overnight*. Como controles foram usadas bactérias transformadas com plasmídeos íntegros (controle da transformação), plasmídeos não ligados (controle da ligação), e apenas bactérias (controle da ampicilina).

4.5.4 Obtenção de DNA de plasmídeo recombinante (miniprep)

Uma colônia de bactérias selecionada em placa com ampicilina foi inoculada em 3 ml de meio LB contendo 50 μ g/ml de ampicilina e mantida *overnight* sob agitação. Desta cultura, foram coletados 1,5 ml, centrifugados por 1 min, e desprezado o sobrenadante. O *pellet* foi ressuscitado em 300 μ l de solução de lise (Triton X-100 2,5%, sucrose 4%, Tris-HCl 25 mM pH 8, EDTA 25 mM, 4 μ g lisozima 10 ml/ml). Após fazer um orifício na tampa do tubo *ependorf* para evitar perda de material, a solução foi fervida a 100 °C por 1 min, e centrifugada 10 min a 10.000 rpm. O DNA bacteriano foi retirado com um palito. Ao sobrenadante foram acrescentados 10 μ l de DEP (dietil pirocarbonato) para eliminar as nucleases; a solução foi incubada por 15 min a 65 °C e centrifugada a 10.000 rpm por 10 min. O sobrenadante foi transferido para um novo tubo. Os plasmídeos foram precipitados com NaCl, concentração final 100 mM, e um volume de isopropanol. A seguir centrifugados por 15 min a 10.000 rpm. O sobrenadante foi descartado e adicionado álcool 70% ao *pellet* para retirar o excesso de NaCl, centrifugado por 10 min a 10.000 rpm e desprezado o sobrenadante. Após secagem o DNA plasmidial foi

ressuspendido com TE (Tris 10 mM, EDTA 1 mM) e 1 µl RNAase (concentração final 10 µg/ml). Os plasmídeos foram analisados em gel de agarose 1,2%.

4.5.5 Análise do DNA recombinante

Os DNAs recombinantes foram submetidos a digestões simples com enzima de restrição *Eco RI* (*New England Biolabs*) para verificar a presença do inserto no plasmídeo. Foram utilizados 100 ng de DNA plasmidial (produto da miniprep), 0,25 µl (5 U) de *Eco RI* (200 U/µl), 3 µl de tampão de digestão *EcoRI* (10x) para uma reação de volume final 30 µl, completando com água Milli-Q. Após 3 h no banho-maria a 37 °C, a reação foi interrompida deixando o tubo a 4 °C. O DNA digerido foi colocado para migrar em gel de agarose 1,2%.

4.6 PCR dos clones para sequenciamento

Os clones foram amplificados por PCR, utilizando os *primers* que fazem parte das sequências do plasmídeo que flanqueiam o inserto (SP6 e T7). Além disso, foram usados outros *primers* necessários para *walking* ao longo do cDNA.

SP6 direto: 5' - ATT TAG GTG ACA CTA TAG AAT - 3'

T7 reverso: 5' - TAA TAC GAC TCA CTA TAG GGC - 3'

Condições: 94 °C 1 min; 54 °C 1 min; 72 °C 1 min (35 ciclos); 72 °C 5 min (extensão).

4.7 Sequenciamento de DNA

O sequenciamento dos fragmentos eluídos de PCR e/ou das minipreps de clones selecionados foi feito utilizando individualmente os *primers* SP6 ou T7 presentes no plasmídeo e flanqueando o inserto, e *kit* para amplificação *Big Dye Terminator*, versão 3.1 (*Applied Biosystems*). Condições: 96 °C, 10 s; 52 °C, 20 s; 60 °C, 4 min (40 ciclos). O sequenciamento foi feito em sequenciador *ABI Prism 3100*.

4.8 Análise das sequências

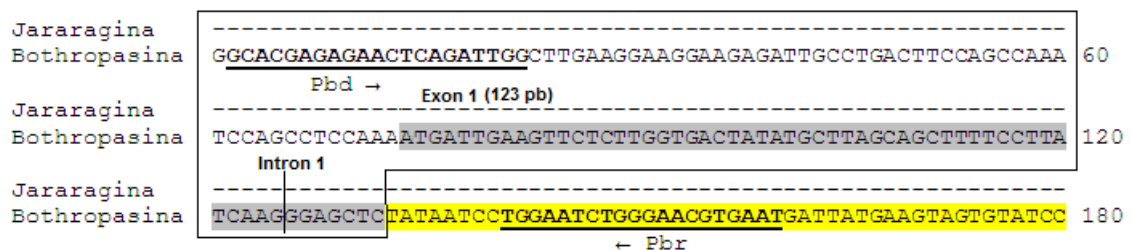
As seqüências obtidas foram comparadas entre si e com os cDNAs de referência utilizando programas como *BLAST* (*Basic Local Aligment Search Tool*), *Clustal W* e outros para pesquisa em banco de dados.

5 RESULTADOS

As sequências dos cDNAs da jararagina e bothropasina foram alinhadas utilizando o programa *Multiple Sequence Alignment by Clustal W* (<http://align.genome.jp/>), para analisar a identidade entre as sequências (ANEXO A).

5.1 Região 5' não traduzida (UTR) e Peptídeo Sinal

A região 5' UTR (bases sem cor) compreende do nucleotídeo 1 ao 72. O domínio Peptídeo Sinal (bases marcadas com a cor cinza) compreende do nucleotídeo 73 ao 131 (utilizando como base o cDNA da bothropasina). Apenas a sequência do peptídeo sinal da bothropasina havia sido publicada. Os *primers* utilizados para sua amplificação encontram-se na região 5' UTR e no domínio catalítico. Foi identificado um intron no final do Peptídeo Sinal.



Para a amplificação dessa região foram utilizados os *primers* Pbd presente na região 5' não traduzida e Pbr no domínio catalítico. O PCR teve como amplificação uma banda de ~1.8 Kb o que indica a presença de um intron, pois o produto esperado era de 140 pb (Fig. 12).

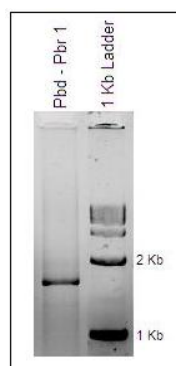


Figura 12. PCR Pbd-Pbr.

A banda foi recortada, purificada e ligada no plasmídeo *pGEM T-easy Vector*. Os plasmídeos recombinantes foram usados para transformar bactérias *E. coli*

DH5 α . Foram analisadas as minipreps de 3 colônias em gel de agarose 1,0%, comparando-se com o padrão de migração de um plasmídio íntegro (pUC19) sem inserto, que tem uma padrão de corrida mais rápido comparado com os plasmídeos recombinantes (Fig. 13).

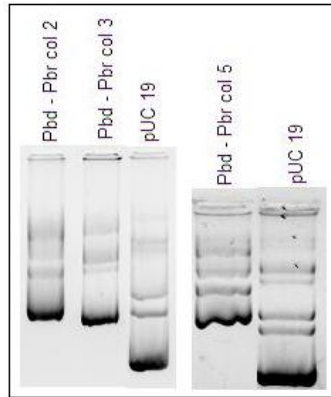


Figura 13. Minipreps Pbd-Pbr.

A confirmação da presença de inserto foi feita por digestão com *Eco* RI que faz parte do sítio múltiplo de clonagem. Não foi feito PCR para confirmar o tamanho região devido a *Taq* DNA polimerase disponível não amplificar um fragmento de ~2 Kb. Para confirmar o tamanho do inserto, foi utilizado como controle o pUC 19 digerido (sem inserto), representado pela banda de 3 kb. A digestão das minipreps mostra o tamanho do inserto (~1.8 kb). Os clones 2, 3 e 5 mostram insertos do tamanho esperado (Fig. 14).

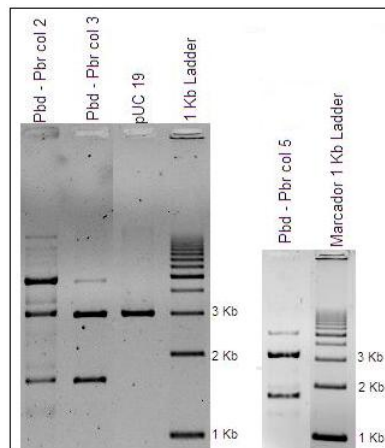


Figura 14. Digestão *Eco*RI Pbd-Pbr.

A partir dos clones Pbd-Pbr foram feitos *Nested PCRs* com os *primers* Pbdd-Pbr, obtendo-se sub- clones que foram analisados em conjuntos.

Os sequenciamentos foram feitos a partir dos fragmentos obtidos das minipreps dos clones selecionados Pbd-Pbr col 2, 3 e 5 e seus sub- clones, utilizando os *primers* SP6 e T7. Os clones 2 e 5 mostraram completa identidade de

sequência, representados pela sequência superior (S). O clone 3 corresponde à sequência inferior (I). O Exon 1 tem 123 pb em ambas as bandas; o intron 1 na sequência S tem 1441 pb e na sequência inferior 1440 pb. Foram identificados 22 pb divergentes entre as sequências. Notar a sequência receptora universal *gt* em 5' e *ag* em 3' do intron 1 (sublinhado) (Fig. 15).

	Pbd →		Exon 1			
B	<u>G</u> CACGAGAGA	<u>ACTCAGATTG</u>	GCTTGAAGGA	AGGAAGAGAT	TGCCTGTCTT	CCAGCCAAAT
C	<u>G</u> CACGAGAGA	<u>ACTCAGATTG</u>	GCTTGAAGGA	AGGAAGAGAT	TGCCTGTCTT	CCAGCCAAAT
	*****	*****	*****	*****	*****	*****
B	AGCCTCCAAA	AATGATTGAA	GTTCTCTTGG	TGACTATATG	CTTAGCAGCT	TTTCCTTATC
C	AGCCTCCAAA	AATGATTCAA	GTTCTCTTGG	TGACTATATG	CTTAGCAGCT	TTTCCTTATC
	*****	*****	*****	*****	*****	*****
B	AAG	Intron 1				
S	AAGgtaagat	gttctctttg	gttctcttgt	tcagattcct	actgctaaaa	gactattgca
I	AAgGtaagat	gttctctttg	gttttcttgt	tcagattcct	actgcaaaaa	gactattgca
	*****	*****	*** *****	*****	***** **	*****
S	cccaacagat	tgctatggtg	atggggtgtg	ttttgggatt	tatttttgac	aattaaccaa
I	cccaacagat	tgctacggtg	atggggtttg	ttttgggatt	tatttttgac	aattaaccaa
	*****	***** **	***** **	***** **	*****	*****
S	aatttgctcc	atttcagttt	ctacagattt	agcaatagaa	tggctctcaag	ggtcacattt
I	aatttgctcc	atttcagttt	ctacagattt	agcaatagaa	tggctctcaag	ggtcacattt
	*****	*****	*****	*****	*****	*****
S	gttctaaatt	gacataaatg	gtttcttggg	attttcttaa	attttgtaa	gcatgagca
I	gttctaaatt	gacataaatg	gttccttggg	attttcttaa	attttgtaa	gcatgagca
	*****	*****	*** *****	*****	*****	*****
				Pbd →		
S	gatcctaaga	aaaaaagtcc	taagatcttc	agaagaaaag	<u>catgttaaag</u>	<u>tgggagagaa</u>
I	gatcctaaga	aaaaaagtcc	taagatcttc	agaagaaaag	<u>catgttaaag</u>	<u>tgggagagaa</u>
	*****	*****	*****	*****	*****	*****
S	catttttcag	tgatgtgagt	tttaaaaaag	gtttagagga	ctgaagagga	cagagagaca
I	catttttcag	tgatgtgagt	tttaaaaaag	gtttagagga	ctgaagagga	cagagagaca
	*****	*****	*****	*****	*****	*****
S	gagtttgtgg	caagaaaatt	tcagaaagca	aacaaaatgt	taaagaaaaa	taaagtgttt
I	gagtttgtgg	caagaaaatt	tcagaaagca	aacaaaatgt	taaagaaaaa	taaagtgttt
	*****	*****	*****	*****	*****	*****
S	ttggtttttaa	ttttctttca	ataggtcaat	tagccatcct	gtaattaaac	ttacataata
I	ttggtttttaa	ttttctttca	ataggtcaat	tagccatcct	gtaattaaac	ttacataata
	*****	*****	*****	*****	*****	*****
S	ttcagccaaa	caaaagctat	tgctattgct	acagagaatc	ttcaaactac	agtcaagata
I	ttcagccaaa	caaaagctat	tgctattgct	acagagaatc	ttcaaactac	agtcaagata
	*****	*****	*****	*****	*****	*****


```

S atatgacatt aacttttctc aatgcaagac tttgaattat cttaaattat ttcttccctt
I atatgacatt aacttttctc aatgcaagat tttgaattat cttaaattat ttcttccctt
*****
S gtttgattca attgattcag aagttgccaa acacaactaa gggttagaga cttatagtaa
I gtttgattca attgattcag aagttgccaa acacaactaa gggttagaga cttatagtaa
*****
S acttgcttgc tagtaggtgg agttgaacc ccccatgat tctgggagca tcaggtagaa
I acttgcttgc tagtaggtgg agttgaacc ccccatgat tctgggagca tcaggtagaa
*****
S atgggtaggt acataaatag atgtgaggaa gatgtattct tctgtctcgg gttgctcatc
I atgggtaggt acataaatag atgtgaggaa gatgtattct tctgtctcgg gttgctcatc
*****
S agtctgagaa aacacatttc cagggcccaa cgctagtttc tctctttctc agtctctctc
I agtctgagaa aacacatttc cagggcccaa cgctagtttc tctctttctc agtctctctc
*****
S tctgtctcaa ttcctatagg tgttgggcta gcagtttctt tcccgtaacc tgtgggttatt
I tctgtctcaa ttcctatagg tgttgggcta gcagtttctt tcccgtaacc tgtgggttatt
*****
S gatacttgag tttattcttt tgtttccact tgcaactata agcacaacct taccctccc
I gatacttgag tttattcttt tgtttccact tgcaactata agcacaacct taccctccc
*****
S ttcattaat ctgctaacia atcttgggtct cagttgtgat tgtaactcca agagaacgtg
I ttcattaat ctgctaacia atcttgggtct cagttgtgat tgtaactcca agagaacttg
*****
S cagtgtcagg ctgccatttt cacaaaataa gttaagggaac aatatcagg tgaccatcca
I cagtgtcaga ctgccatttt cacaaaataa gctaagggaac aatatcagg tgaccatcca
*****
S atgtccatat ggggaccagc tccaaattgt tcagctatta gtcaaattct gaaatcacta
I atgtccatat ggggaccaac tccaaattgt tcagctatta gtcaaattct gaaatcacta
*****
S cattatttac atgttcttag gtgttccggg gataattatg atgttgtggg ttgtccatag
I tattatttac atgttcttag atgtctcagt gataattatg atgccgtggg ttgtccatag
*****
S attcttttga aatggaaga aaggaaagtc agcaaaacta ttaggatggc agtaaaatag
I attcttttaa aatggaaga aaggaaagtc agcaaaacta ttaggatggc agtaaaatag
*****
S cagttcagct gaaaggaatt actaatcagc tgcacttgca tcccagaga aaaggatttc
I cagttcagct gaaaggaatt actaatcagc tgcacttgca tcccagaga aaaggatttc
*****
S tgtgctgtgt cccactagaa aaagattcca ttcttctactg atgaaccgaa taacaaattg
I tgtgctgtgt cccactagaa aacgtttcca ttcttctactg at-aaccgga taacaaattg
*****

```

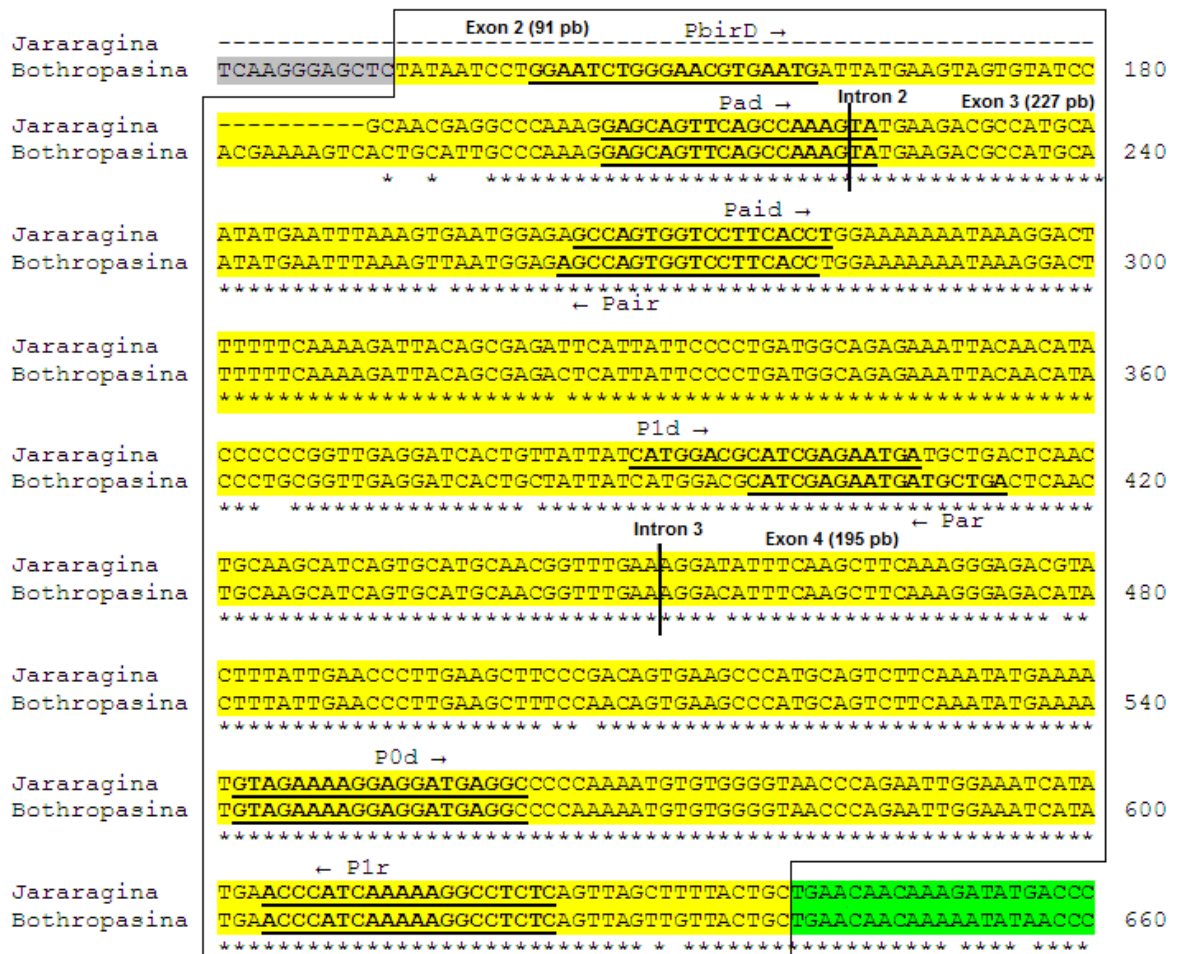
```
S tgtttgtggat gtgaaaatgg tagaataaca tcccaaagta aaccaaacct tcttttcttt
I tgtttgtggat gtgaaaatgg tagaatcaca tcccaaagta aaccaaacct tcttttcttt
*****      *****      *****  **      *****      *****      *****
```

```
Exon 2
S ttagGGAGCT C
I ttagGGAGCT C
***** *
```

Figura 15. Sequenciamento 5' UTR e Peptídeo Sinal. Os asteriscos mostram a identidade entre as sequências. Em negrito estão os *primers* utilizados para amplificar a região. (B) Bothropasina; (S) banda superior; (I) banda inferior. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

5.2 Pró-domínio

O pró-domínio é mostrado no box amarelo, compreendendo do nucleotídeo 132 ao 638, usando como base a sequência dos cDNAs na bothropasina. Os asteriscos mostram identidade de sequências; a falta deles indica divergência.



O primeiro par de *primers* utilizados para amplificar este domínio foi o PbirD-Pair. O fragmento esperado era de 142 pb, mas foram encontradas três bandas: banda 1 (~550 pb), banda 2 (~850 pb), e banda 3 (~950 pb) (Fig. 16).



Figura 16. PCR PbirD-Pair.

As bandas estavam muito próximas para serem recortadas. A clonagem foi feita a partir do PCR contendo todas as três bandas. Para análise dos clones foram feitas minipreps (Fig. 17).

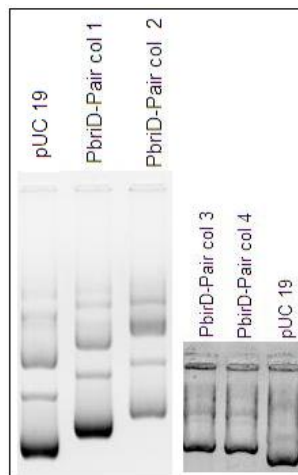


Figura 17. Minipreps PbirD-Pair.

Para confirmar a presença dos insertos do tamanho desejado foram feitos PCRs com os *primers* SP6 e T7 (Fig. 18).

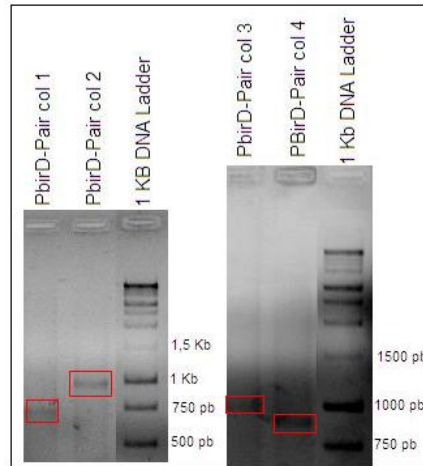


Figura 18. PCR das minipreps PbirD-Pair. Os PCRs confirmaram a presença do inserto da banda 1 (colônia ou clone 1), banda 2 (clone 4), e banda 3 (clones 2 e 3).

Foram sequenciados os clones das colônias 1, 4 e 3 que correspondem às bandas 1, 2 e 3, respectivamente. O consenso foi definido pelo PCR reverso e direto dos clones. O exon 2 tem 91 pb em todos os clones, o intron 2 das bandas 1, 2 e 3 (C1, C2, C3) contém 570 pb, 850 pb, e 974 pb, respectivamente. Notar a sequência doadora universal *gt* em 5', e receptora *ag* em 3' do intron 2 (sublinhado) (Fig. 19).

	PBirD →		Exon 2			
B	<u>GGAATCTGGG</u>	<u>AACGTGAATG</u>	ATTATGAAGT	AGTGTATCCA	CGAAAAGTCA	CTGCATTGCC
J						G CAACGAGGCC
C1	<u>GGAATCTGGG</u>	<u>AACGTGAATG</u>	ATTACGAAGT	AGTGTATCCA	CGAAAAATCA	CTGCATTGCC
C2	<u>GGAATCTGGG</u>	<u>AACGTGAATG</u>	ATTATGAAGT	AGTCTATCCA	CGAAAAATCA	CTGAATTGCC
C3	<u>GGAATCTGGG</u>	<u>AACGTGAATG</u>	ATTATGAAGT	AGTCTATCCA	CGAAAAGTCA	CTGCATTGCC
	*****	*****	*****	*****	*****	* * **
B	CAAAGGAGCA	G TTCAGCCAA	AG			
J	CAAAGGAGCA	G TTCAGCCAA	AG	Intron 2		
C1	CAAAGGAGCA	G TTCAGCCAA	AG <u>g</u> taagaaa	aagatctttc	tttcatcaac	aaattatt--
C2	CAAAGGAGCA	G TTCAGCCAA	AG <u>g</u> taagaaa	aagatctttc	tttcatcaac	aaatgatttc
C3	CAAAGGAGCA	G TTCAGCCAA	AG <u>g</u> taaggaa	aagatctttc	tttcatcaac	aaattatttc
	*****	*****	*****	*****	*****	*****
C1	-----	-----	-----	-----	-----	-----
C2	ttgtcagccc	acagaaatgt	tgttattcca	ttgctgccat	ctaattggta	tctggatttt
C3	ttgtcagccc	acagaaatgt	tgttattcct	ttgctgccat	ctaattggta	tctggatttt
	*****	*****	*****	*****	*****	*****
C1	-----	-----	-----	-----	-----	-----
C2	gcattacaat	ctatgttcct	gcttgaatta	tttatgtatt	aattaaagat	taattaatac
C3	gcattgcaat	ctatgttcct	gcttgaatta	tttatgtatt	aattaaaaaa	aattatatgg
	*****	*****	*****	*****	*****	* * **
C1	-----	-----	-----	-----	-----	-----
C2	ccaaattgtg	cccaattcac	acaaaattga	ggagattcct	ggcaatttac	aaatagaatt
C3	cc-----g	cccaattcac	acaaaactga	ggagattcct	ggcaatttac	aaatagaact
	**	*	*****	*****	*****	*****

```

C1 -----
C2 tcctaataata tataaggttag tataactcgt aaagtgttgg gagacaccca gccgtaccgt
C3 tcctaataatg tataaggttag tataactcat aaagtgttga gagacaccca ggtgtaccgt
*****
***** * *****

C1 -----gga attagccatg agtcatttca tagctgcaac agatccagga
C2 gaggagggat cgagtgggga attagccatg agtcatttca tagctgcaac agatccagga
C3 gaggagggat caagtaggga actagccatg agtcatttca tagctgcaac agatccagaa
***** * ***** * *****

C1 ctgacatcct cccccactcc aggaaattcc attaaagttt atctatctgt tctgaagaga
C2 ctgacatcct cccccacttc aggaaattcc atgaaagttt atctatatgt tctgaagaga
C3 ctgacatcct cccccactcc aggaaattcc attaaagttt atctatctgt tctgaagaga
***** * ***** ** *****

C1 aggggagcag tctataggga gattcctgca ttataggcag aatacagtaa atgatttaat
C2 agggaagcag tctataggga gatttctgca ttataggcag aatacagtaa agtgtttaat
C3 agggaagcag tctataggga gattcctgca ttataggcag aatacagtaa agtacttaat
**** ***** ***** ***** * *****

C1 ttaaactctg catgtctaca tgcacatcaatt agaagtataa aagatgt--- -----
C2 ttaaactctg catctctaca tgcacacattt agaaatgtaa aagatgtggt tctaggggaa
C3 ttagaatctg catgtctacc tgcacac--- ----atataa aaggtgtggt tctaggggaa
*** ***** ***** * ***** * ***** *****

C1 -----atag tggttacaac tcaagattag ttctgagaat gagattcaag tct-----
C2 gatggtataa tggttacaac tcaagattag ttctgagaat gagattcaag tct-----
C3 gatggtatag aggttacaac gcaagattat ttctgagaat gagattcaag tctatcgctg
***** ***** ***** ***** *****

C1 -----ctttt gagtcagtct ctctctgtcc cagctcaacc tcattcacca agatggtcaa
C2 -----ctttt gagtcagttt ctct--gtcc cagctcaacc tcattcaccc agatggtcaa
C3 gatgactttt gagtcagtct ctct--gtcc cagctcacc tcattcacca agatggtcaa
***** ***** * ***** ***** *****

C1 ttagagaatg aaatcttgaa cttatgaagg aa--t-gtta ttttaagatc ttacagttca
C2 ttagagaatg aaatcttgaa cttatgcaga a---t-gtta ttttaagatc gtacagttca
C3 ttagagaatg agatcttgaa cttatgcaga aagatcgtta taagatctta cagttgacat
***** * ***** ***** * * ***** * * *

C1 -----
C2 -----
C3 cactagatga aatactctaa atctaatagaa -taatttggc tatggcaaat tattcattag

C1 -----
C2 -----
C3 atttaaattg atacatgtgt agagttgtta agaggataca caacttagtt gcttaattct

Exon 3
B TATGAA GACGCCATGC AATATGAATT
J TATGAA GACGCCATGC AATATGAATT
C1 aagcact- -agcctgaat tctttgcttt tcagTATGAA GACGCCATGC AATATGAATT
C2 aagcacc- -agcctgaac tctttgcttt tcagTATAAA GACGCCATGC AATATGAATT
C3 ttaagaagta aagtctgaac actttgcttt tcagTATGAA GACACCATGC AATATGAATT
*** * ** ***** ***** ***** *****

```

← Pair

B TAAAGTTAAT GGAG**AGCCAG TGGTCCTTCA** CC
 J TAAAGTGAAT GGAG**AGCCAG TGGTCCTTCA** CC
 C1 TAAGGTGAAT GGAA**AGCCAG TGGTCCTTCA** CC
 C2 TAAGGTGAAT GGAG**AGCCAG TGGTCCTTCA** CC
 C3 TAAAGTGAAT GGAG**AGCCAG TGGTCCTTCA** CC
 ***** ***** ***** **

Figura 19. Sequenciamento PbirD-Pair. As denominações C1, 2 e 3 são provisórias, até a obtenção das sequências definitivas.

O produto de amplificação com os *primers* Paid-Par foi 150 pb indicando não haver intron nesta região (Fig. 20).

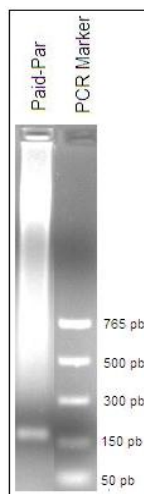


Figura 20. PCR Paid-Par.

A clonagem foi feita a partir do produto de PCR, tendo sido analisados seis clones pela técnica de miniprep. Os clones foram comparados com o pUC 19 íntegro. Devido o produto ligado ao plasmídeo ser muito pequeno, não aparece diferença entre os plasmídeos recombinantes e o pUC (Fig. 21).

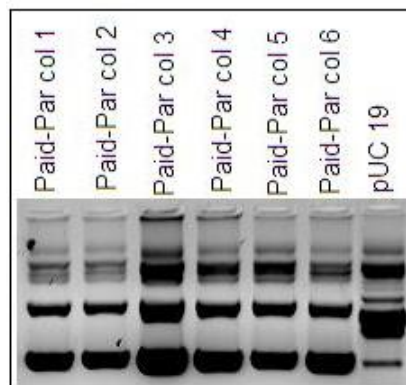


Figura 21. Minipreps Paid-Par.

Para confirmar o tamanho dos insertos foi feito um PCR com os *primers* SP6 e T7. Os clones Paid-Par col 2 e 6 não apresentam o inserto, os demais têm o inserto do tamanho esperado (Fig 22).

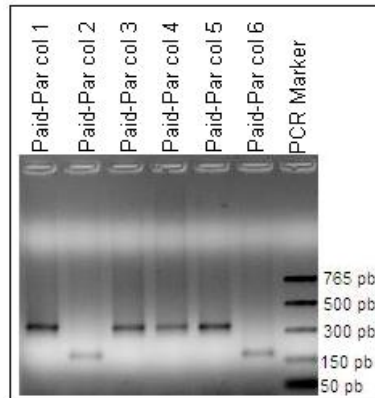


Figura 22. PCRs das minipreps Paid-Par.

Os clones Paid-Par col 1, 3, 4 e 5 foram sequenciados (Fig. 23). Os 4 clones apresentaram a mesma sequência. O tamanho total do exon 3 é 227 pb. Alguns pontos de mutação nas sequências obtidas dos clones diferem das sequências da jararagina e da bothropasina.

Exon 3 (cont.) Paid →

B	<u>GCCAGTGGTC</u> <u>CTTCACCT</u> GG	AAAAAAATAA	AGGACTTTTT	TCAAAAGATT	ACAGCGAGAC	
J	<u>GCCAGTGGTC</u> <u>CTTCACCT</u> GG	AAAAAAATAA	AGGACTTTTT	TCAAAAGATT	ACAGCGAGAT	
C	<u>GCCAGTGGTC</u> <u>CTTCACCT</u> GG	AAAAAAATAA	AGGACTTTTT	TCAGAAGATT	ACAGCGAGAT	
	*****	*****	*****	*** *****	*****	
B	TCATTATTCC	CCTGATGGCA	GAGAAATTAC	AACATACCCT	GCGGTTGAGG	ATCACTGCTA
J	TCATTATTCC	CCTGATGGCA	GAGAAATTAC	AACATACCCC	CCGGTTGAGG	ATCACTGTTA
C	TCATTATTCC	CCTGATGACA	GAGAAATTAC	AACAAACCCT	CTGGTTGAGG	ATCACTGCTA
	*****	***** *	*****	**** *	*****	***** *

← Par

B	TTATCATGGA	CG <u>CATCGAGA</u>	<u>ATGATGCTGA</u>
J	TTATCATGGA	CG <u>CATCGAGA</u>	<u>ATGATGCTGA</u>
C	TTATCATGGA	CA <u>CATCGAGA</u>	<u>ATGATGCTGA</u>
	*****	* *****	*****

Figura 23. Sequenciamento Paid-Par. B (bothropasina), J (jararagina), C (consenso dos clones).

Esta região foi inicialmente amplificada com os *primers* P1d-P1r e clonada (Quadro 2) pelo aluno de Mestrado Marcelo Moreira Tavares de Souza, e posteriormente confirmada neste trabalho.

Quadro 2 - Clones P1d-P1r. Nomeados pelo aluno Marcelo M. T. Souza. Alguns clones tiveram seu nome alterado e corrigido após a análise do PCR (alteração entre parênteses).

P1d - P1r
p1 inf 4
p1 inf 6 (sup)
p1 inf 7
p1 inf 8 (sup)
p1 inf 10
p1 mix 2
p1 mix 6
p1 inf C
p1 inf E (sup)
p1 inf F (sup)

Foram feitas minipreps a partir dos clones que estavam em estoque (*stabs*). Os plasmídeos recombinantes de sete colônias foram analisados após a miniprep e comparados com o plasmídeo integro ou marcador (Fig. 24).

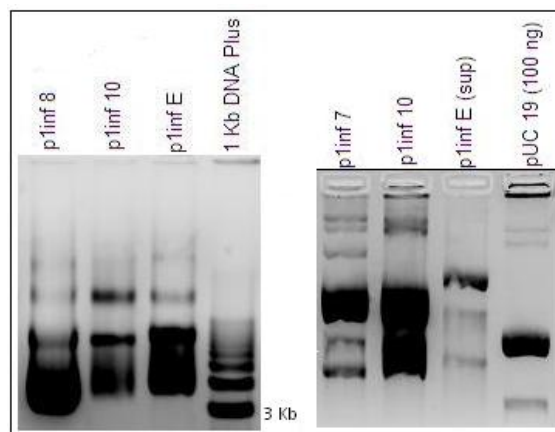


Figura 24. Minipreps P1d-P1r.

Os tamanhos dos insertos foram analisados e confirmados por PCR utilizando os *primers* SP6 e T7 (Fig. 25), e digestão *Eco* RI (Fig. 26).

O PCR indicou que os clones p1inf 10 e p1inf 7 têm um inserto de ~750 pb, os clones p1inf 8, p1inf E e p1inf 6 têm um inserto de ~1 Kb. Notar que a denominação original dos clones está incorreta, pois inf 8, inf E e inf 6 são na realidade sup8, sup E e sup 6 (inf = banda menor ou inferior; sup = banda maior ou superior no gel). Conferir com a Tabela 2. A digestão dos clones p1inf 7 e p1inf 10 confirmou a presença de inserto de ~600 pb. Os clones p1inf8 e p1inf E confirmaram a presença de um inserto de ~ 850 pb. O DNA do clone p1inf6 não foi digerido.

S ttgtcacaga taaaatgagg gacatcagtt tttcaagact tctgcagaat taagaaagac
 I ttgtcacaga taaaatgagg gacgtcagtt tttcaagact tctgcagaat taagaaagac

 S tttgatagaa taccaagaag agctcctgcc agtcactaga gacaatagga agctaaatca
 I tttgattgaa taccaagaag agctgctgcc agtcactaga gacaatagga agctaaataa

 S gatctgtcta aaggagcttc caagcccctg tctcctttac atataggaat tgatgctaca
 I gatctgtcta a-ggagcttc caagcccctg tctcctttac atataggagg tgatgctaca

 S ggtagagaag ataaataggt ttcaaattgc ataccttgct agaaaggat tccatgcaaa
 I ggtagagaag ataaataggt ttcaaattgc ataccttgct agaaaggat tccatgcaaa

 S agtatctaaa atgaatactg tacaatatt tcatggttgc actgagcgtg tgcaatctct
 I agtatctaaa a----- tacaatatt tcatggctgc attgagtgctg tgcaatctct

 S tatatctttc tgatcatagc tgtatcattg agaagtatga gaaagcgata tgagtgctga
 I tatatctttc tgttcatagc tgtatctttg agaggatga gaaagtata tgagtgctga

 S ggaaagaaaa taaaaatgtc agtgagtggt aagaaggaat ttaaaaacac agggaaatggt
 I ggaaagaaaa taaaaatgtc agtgagtggt aagaagggat tgaaaaacac agggaaatggt

 S tacttggtta gttaaaaatg ggtctcagag ctgagtttca ttcccaacta gctaacatca
 I gacttggtta gttaaaaata ggtctcagag ctgagtttca ttcccaacta gctaacatca

 S tcagtgacag tgatctctga cctaacaata ttctcttctg ttgcttctcc atctctgatc
 I ccagtgacag ttatctctga cctaacaata ttctcttctg ttgcttctcc atctctgatc

Exon 4

B AGGA CATTTCAGC TTCAAAGGGA GACATACTTT
 J AGGA TATTTCAGC TTCAAAGGGA GACGTACTTT
 S tacccttttc cactattggt ttacgaAGGA CATTTCAGC TTCAAAGGGA GACGTACTTT
 I a-cccttttc cac-attggt ttacagAGGA CATTTCAGC TTCAAAGGGA GACGTACTTT

 B ATTGAACCT TGAAGCTTC CAACAGTGAA GCCCATGCAG TCTTCAAATA TGAAAATGTA
 J ATTGAACCT TGAAGCTTCC CGACAGTGAA GCCCATGCAG TCTTCAAATA TGAAAATGTA
 S ATTGAACCT TGAAGCTTCC CGACAGTGAA GCCCATGCAG TCTTCAAATA TGAAAATGTA
 I ATTGAACCT TGAAGCTTCC CGACAGTGAA GCCCATGCAG TCTTCAAATA TGAAAATGTA

 B GAAAAGGAGG ATGAGGCCCC CAAAATGTGT GGGGTAACCC AGAATTGGAA ATCATATGAA
 J GAAAAGGAGG ATGAGGCCCC CAAAATGTGT GGGGTAACCC AGAATTGGAA ATCATATGAA
 S GAAAAGGAGG ATGAGGCCCC CAAA-TGTGT GGGGTAACCC AGAATTGGAA ATCATATGAA
 I GAAAAGGAGG ATGAGGCCCC CAAAATGTGT GGGGTAACCC AGAATTGGAA ATCATATGAA

← P1r

B CCCATCAAAA AGGCCTCTC
 J CCCATCAAAA AGGCCTCTC
 S CCCATCAAAA AGGCCTCTC
 I CCCATCAAAA AGGCCTCTC

Figura 27. Sequenciamento P1d-P1r. As denominações S e I são provisórias, até a obtenção das seqüências definitivas.

O produto da amplificação do par de *primers* P1d – P1rr é ~600 pb (Fig. 28). A banda foi recortada do gel de agarose 1,5%, ligada no plasmídeo *pGEM T-easy Vector* e usado para transformar as bactérias competentes DH-5 α .

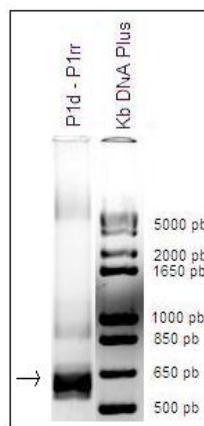


Figura 28. PCR P1d-P1rr. A banda mostrada pela flecha foi recortada, purificada e ligada no plasmídeo.

Os clones de 7 colônias foram analisados por miniprep. O produto das minipreps foi resolvido em gel de agarose 1,0 % e comparado com pUC 19 íntegro (sem inserto). O clone P1d – P1rr col 10 apresentou uma banda menor que os demais, indicando não ter o inserto (Fig. 29).

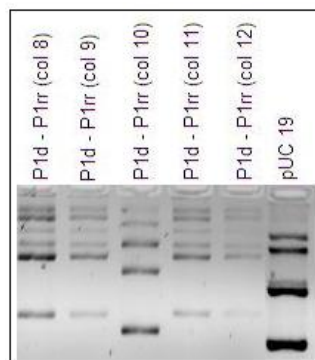


Figura 29. Minipreps P1d-P1rr.

Para confirmar o tamanho dos insertos foi feito um PCR com os *primers* SP6 e T7. A análise do padrão de migração do PCR permitiu identificar os plasmídeos com os insertos desejados (~ 800 pb), exceto o clone P1d - P1rr col 10 (Fig. 30). A

digestão com *Eco* RI confirmou os tamanhos dos insertos observados nos PCRs (Fig. 31).

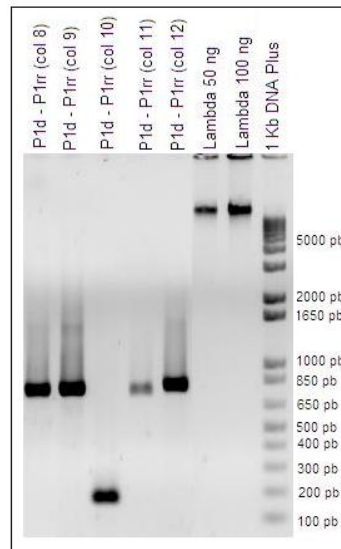


Figura 30. PCR das Minipreps P1d-P1rr.

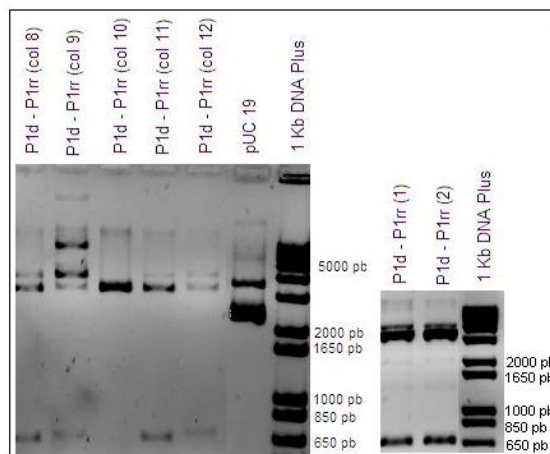


Figura 31. A digestão *Eco* RI P1d-P1rr.

O produto das minipreps dos clones 2, 8 e 9 da região P1d-P1rr foi sequenciado utilizando o *primer* T7 (direto). O alinhamento gerou o consenso (C) entre os três clones. Notar a sequência doadora universal *gt* em 5' do introns 3 (sublinhado) (Fig. 32).

Exon 3 P1d →

B	<u>CATGGACGCA</u> <u>TCGAGAATGA</u>	TGCTGACTCA	ACTGCAAGCA	TCAGTGCATG	CAACGGTTTG
J	<u>CATGGACGCA</u> <u>TCGAGAATGA</u>	TGCTGACTCA	ACTGCAAGCA	TCAGTGCATG	CAACGGTTTG
C	<u>CATGGACGCA</u> <u>TCGAGAATGA</u>	TGCTGACTCA	ACTGCAAGCA	TCAGTGCATG	CAACGGTTTG
	*****	*****	*****	*****	*****

B AA

J AA **Intron 3**

C AAgtaagata gtctctaatt ttttatttat tttattaaga ataatacact ggtcttggag
**

C ttgtcacaga taaaatgagg gacatcagtt tttcaagact tctgcagaat taagaaagac

C tttgatagaa taccaagaag agctcctgcc agtcactaga gacaatagga agctaaatca
 C gatctgtcta aaggagcttc caagcccctg tctcctttac atataggaat tgatgctaca
 C ggtagagaag ataaataggt ttcaaattgc ataccttgct agaaaggtat tccatgcaaa
 C agtatctaaa atgaatactg taaaaatatt tcatggttgc actgagcgtg tgcaatctct
 C tatactcttc tgatcatagc tgtatcattg agaagtatga gaaagcaata tgagtgctga
 C ggaaagaaaa taaaaatgtc agtgagtgtt aagaaggaat ttaaaaacac agggaatggt
 C tacttgggta gttaaaaatg ggtctcagag ctgagtttca ttc
 ← P1rr

Figura 32. Sequenciamento do P1d-P1rr. Confirmação do intron 3.

Os *primers* P0d localizados no exon 4 e P0rr (no intron) foram úteis para identificar o tamanho do exon presente no pré-domínio e início do catalítico. O exon 4 tem 193 pb. O intron 4 tem 523 pb e 521 pb nas bandas S e I, respectivamente. O exon 5 tem 61 pb. Notar a sequência doadora universal *gt* em 5' dos introns 4 e 5, e receptora *ag* em 3' do intron 4 (sublinhado).

Esta região foi amplificada, clonada e sequenciada pelo aluno de Mestrado Marcelo Moreira Tavares de Souza (Fig. 33).

Exon 4 P0d →

B	<u>GTAGAAAAGG</u>	<u>AGGATGAGGC</u>	CCCCAAAATG	TGTGGGGTAA	CCCAGAATTG	GAAATCATAT
J	<u>GTAGAAAAGG</u>	<u>AGGATGAGGC</u>	CCCCAAAATG	TGTGGGGTAA	CCCAGAATTG	GAAATCATAT
S	<u>GTAGAAAAGG</u>	<u>AGGATGAGGC</u>	CCCCAAA-TG	TGTGGGGTAA	CCCAGAATTG	GAAATCATAT
I	<u>GTAGAAAAGG</u>	<u>AGGATGAGGC</u>	CCCCAAAATG	TGTGGGGTAA	CCCAGAATTG	GAAATCATAT
	*****	*****	*****	*****	*****	*****

B	GAACCCATCA	AAAAGGCCTC	TCAGTTAGTT	GTTACTGCTG	AA	
J	GAACCCATCA	AAAAGGCCTC	TCAGTTAGCT	TTTACTGCTG	AA	Intron 4
S	GAACCCATCA	AAAAGGCCTC	TCAGTTAGTT	GTTACTGCTG	<u>AAgtaagctc</u>	tcatattaaa
I	GAACCCATCA	AAAAGGCCTC	TCAGTTAGTT	CTTACTCCTG	<u>AAgtaagctc</u>	tcatattaaa
	*****	*****	*****	*	*****	*****

S	cctggtttga	gaacagggat	tcatatactt	gtttctaaaa	tataacgtaa	gagagaaatt
I	cctcgtttga	gaacagggat	tcatttactt	gtttctaaaa	tataacgtaa	gagagaaatt
	***	*****	****	*****	*****	*****

S	ctttgtgggg	gggtaattag	ataatca-ca	aagggggaca	ccccccattt	atattttttt
I	cttttttggg	gggtaattag	atagaattca	aaagggagga	ccccccattt	ctattttttt
	***	* **	*****	***	**	* **

S	attctagcca	tggataaaaa	gaaataatgg	aaatatgatg	aatagaaaat	acatgttcca
I	attctagcta	tgggataaaaa	gaaataatgg	gaatatgatg	aattgaaaat	ac-tg-tcca
	*****	* **	*****	*****	***	*****

S	ggattatagc	gttttctttg	atcatgcaaa	gtagtttggg	agatttgaat	caaatgggag
I	ggattatcgc	gttt-ctttg	atcatgcaaa	gtagtttggg	agatttgaat	caaatgggag
	*****	**	*****	*	*****	*****

```

S ttaaataaat ctccaaatca tttcaacttt ctaaataaat ttttaaaaaat aatcaaatta
I ttaaataaat ctccaaatca tgtcaacttc ctaaataaat ttttaaaaaat aatcaaatta
***** * ***** ***** ***** ***** *****
S tcaatttgga tcaatttgca tccaaagtga gagttccttc cttccttcct tccttttggt
I tcaatttgga tcaatttgca tccaaagtga gagttccttc cttccttcct tccttttggt
***** ***** ***** ***** * *****
S ggcaaaaagat gaaactgcct aatacatatc tagaagttca aaatctcact ttttt-atta
I ggcaaaaagat gaaactgcct aatacatatc tagaagttca aaatctc-ct ttttttatta
***** ***** ***** ***** ***** ** *****
S aaatgttaaa attgacatga aacaacagtt tcaacattct tgttggtgc tgggagcaaa
I aaatgttaaa attgacatga aacaacaatt tcaacattct tgttggtgc tgggagcaaa
***** ***** ***** ** ***** *****
                                Exon 5
B                                CAA CAAAAATATA ACCCCTCAG ATACGTTGAG
J                                CAA CAAAGATATG ACCCCTACAA ATACATTGAG
S cagccttttc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA ATACGTTGAG
I cagcctttgc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA ATACGTTGAG
***** ** * ***** ***** ***** ** *****
                                Intron 5
B CTTTTCATAG TTGTGGACCA AGGAATGG
J CTTTTCGTAG TTGTGGACCA AGGAACCG
S CTTTTCATAG TTGTGGACCA AGGAATGGgt aagtaacctg gatatacttc cattttcttt
I CTTTTCATAG TTGTGGACCA AGGAATGGgt aagtaacctg gatatacttc cattttcttt
***** ** ***** ***** ** *****
                                ← P0rr
S ctgcattgga gccaaccaat atttacaag ttaagagtca tttgtaata tactctaatt
I ctgcattgga cccaaccaat atttacaag ttaagagtca tttgtaata tactctaatt
***** ***** ***** ***** ***** *****

```

Figura 33. Sequenciamento P0dd-P0rr. Sequenciamento do final do Pró-Domínio e início do domínio catalítico. As denominações S e I são provisórias, até a obtenção das seqüências definitivas.

5.3 Domínio Catalítico

O Domínio Catalítico é mostrado no box em verde; a seqüência sem cor está em processo de sequenciamento. Tendo como base as seqüências dos cDNAs, esse domínio é codificado pelos nucleotídeos 639 até 1307 (bothropasina). Os asteriscos mostram identidade de seqüência; a falta deles indica divergência entre os cDNAs. Os traços verticais denotam a presença de introns identificados neste trabalho, conforme detalhado abaixo.

Jararagina	← P1r	Intron 4	Exon 5 (61 pb)	470
Bothropasina				660
Jararagina		intron 5	Exon 6 (71 pb)	530
Bothropasina				720
Jararagina	← P0r	P0dr →	intron 6	590
Bothropasina				780
Jararagina			Exon 7 (parcial - 181 pb)	650
Bothropasina				840
Jararagina		P4d →		710
Bothropasina				900
Jararagina		← P4ir	P4irD →	770
Bothropasina				960
Jararagina			P13d →	830
Bothropasina				1020
Jararagina		← P4r		890
Bothropasina				1080
Jararagina		← P11dR	P11d →	950
Bothropasina				1140
Jararagina			intron V	1010
Bothropasina			Exon X (178 pb)	1200
Jararagina				1070
Bothropasina				1260
Jararagina			P9d →	1130
Bothropasina			← P11r	1320
Jararagina			P6d →	1130
Bothropasina				1320

Foi desenhado o par de *primers* P0dd – P0rr a partir do intron sequenciado na região catalítica. A análise do produto do PCR mostrou 3 bandas, nomeadas 1, 2 e 3, com tamanhos de ~680 pb, ~640 pb, e ~480 pb, respectivamente (Fig. 34).

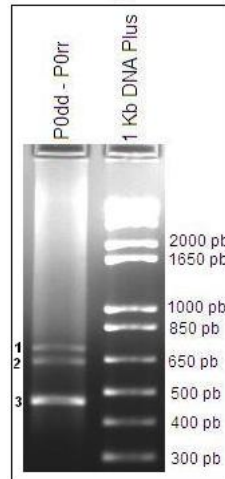


Figura 34. PCR P0dd-P0rr.

As 3 bandas foram recortadas, ligadas em plasmídeos e usadas para transformar as bactérias competentes. Foram analisados 7 clones P0dd – P0rr 1 (banda 1), 6 clones P0dd – P0rr 2 (banda 2) e 2 clones P0dd – P0rr 3 (banda 3) (Fig 35).

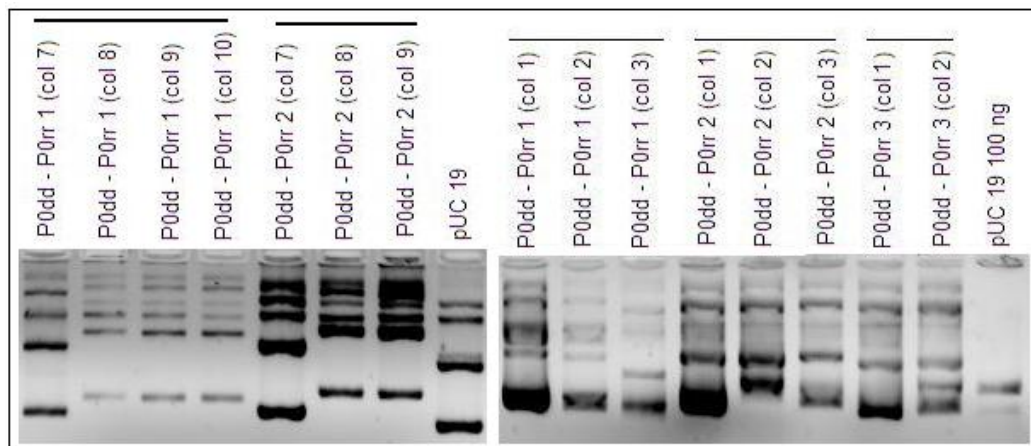


Figura 35. Minipreps P0dd-P0rr. As diferentes colônias foram comparadas com o pUC 19 íntegro.

A análise dos PCRs mostrou que no momento de recortar as bandas houve uma mistura de material entre as 3 bandas, sendo possível visualizar este resultado no PCR. Os clones P0dd – P0rr 1 col 7 e P0dd – P0rr 2 col 7 não tinham inserto (o plasmídeo fechado gera uma banda de 180 pb referente ao próprio plasmídeo) (Fig. 36).

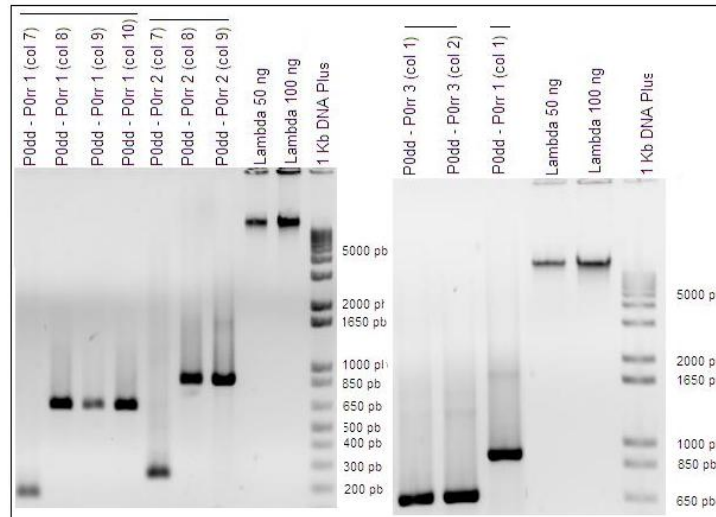


Figura 36. PCR das minipreps P0dd-P0rr.

A digestão com *EcoRI* permitiu separar as bandas pelo tamanho, conforme esperado. Os clones P0dd – P0rr 3 col 1 e 2, P0dd – P0rr 1 col 8, 9 e 10 têm o inserto de ~480 pb (correspondente à banda 3). O clone P0dd – P0rr 1 col 2 tem o fragmento de ~640 pb (banda 2); os clones P0dd – P0rr 1 col 1, P0dd – P0rr 2 col 1, 2, 3, 8 e 9 têm a banda de ~680 pb (banda 1). Os clones P0dd – P0rr 1 col 3 e 7, P0dd – P0rr 2 col 7 não têm inserto (Fig. 37).

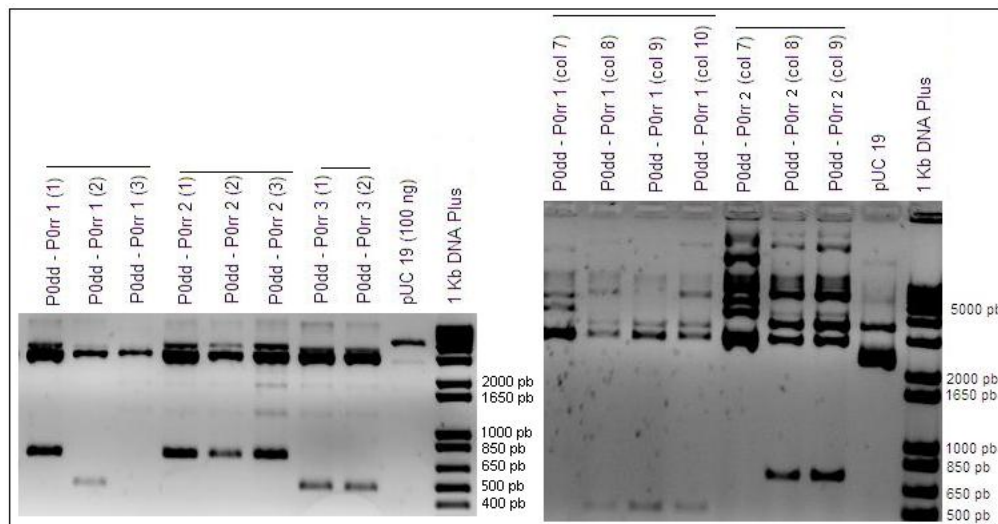


Figura 37. Digestão *EcoRI* P0dd-P0rr.

Foi feito o sequenciamento das 3 bandas obtendo-se os consensos (Cons). O exon 5 tem 61 pb. Para análise do consenso as sequências foram divididas em duas partes: a banda 3, e as bandas 1 e 2 que mostraram ter maior identidade entre si. Os asteriscos indicam a identidade entre as sequências. Notar a sequência receptora universal *ag* em 3' do intron 4, a sequência doadora universal *gt* em 5' do intron 5 (sublinhado) (Fig. 38).

				P0dd →	Intron 4
P0dd-P0rr3-	COL1			<u>gtagtttgg</u>	<u>agatttgaat</u>
P0dd-P0rr3-	COL2			<u>gtagtttgg</u>	<u>agatttgaat</u>
P0dd-P0rr3-	COL5			<u>gtagtttgg</u>	<u>agatttgaat</u>
Cons				*****	*****
P0dd-P0rr2-	COL8			<u>gtagtttgg</u>	<u>agatttgaat</u>
P0dd-P0rr1-	COL8			<u>gtagtttgg</u>	<u>agatttgaat</u>
P0dd-P0rr1-	COL10			<u>gtagtttgg</u>	<u>agatttgaat</u>
Cons				*****	*****
3-C1	caaaatggag	ttaaataaat	ctccaaatca	tttcaacttt	ctaaatcaat
3-C2	caaaatggag	ttaaataaat	ctccaaatca	tttcaacttt	ctaaatcaat
3-C5	caaaatggag	ttaaataaat	ctccaaatca	tttcaacttt	ctaaatcaat
Cons	*****	*****	*****	*****	*****
2-C8	caaaatggag	attaaatgaa	tctccaaatc	atgtcaactt	cctaaatcaa
1-C8	caaaatggag	attaaatgaa	tctccaaatc	atgtcaactt	caaaatcaaa
1-C10	caaaatggag	attaaatgaa	tctccaaatc	atttcaactt	tctaaatcaa
Cons	*****	*****	*****	** *****	*****
3-C1	ttttaaaaat	aatcaaatta	tcaatttgga	tcaatttgca	tccaaagtga
3-C2	ttttaaaaat	aatcaaatta	tcaatttgga	tcaatttgca	tccaaagtga
3-C5	ttttaaaaat	aatcaaatta	tcaatttgga	tcaatttgca	tccaaagtga
Cons	*****	*****	*****	*****	*****
2-C8	tttttaaaaa	taatcaaatt	tcaatttgga	tcaatttgca	tccaaagtga
1-C8	tttttaaaaa	taatcaaatt	tcaatttgga	tcaatttgca	tccaaagtga
1-C10	tttttaaaaa	taatcaaatt	tcagtttgga	tcaatttgca	tccaaagtga
Cons	*****	*****	*** *****	*****	*****
3-C1	gagttccttc	cttccttctc	tccttttggt	ggcaaaagat	aaaactgcct
3-C2	gagttccttc	cttccttctc	ttcttttggt	ggcaaaagat	gaaactgcct
3-C5	gagttccttc	cttccttctc	tcctctaggt	ggcaaaagat	gaaactgcct
Cons	*****	*****	* * * * *	*****	*****
2-C8	gagttccttc	cttccttctc	ttcttttggt	ggcaaaagat	gaaactgcct
1-C8	gagttccttc	cttccttctc	ttcttttggt	ggcaaaagat	gaaactgcct
1-C10	gagttccttc	cttccttctc	ttcttttggt	ggcaaaagat	gaaactgcct
Cons	*****	*****	*****	*****	*****
3-C1	aatacatatc	tagaagttca	aaatctcact	ttttttatta	aaatgttaaa
3-C2	aatacatatc	tagaagttca	aaatctcact	ttttttatta	aaatgttaaa
3-C5	aatacatatc	tagaagttca	aaatctcact	ttttttatta	aaatgttaaa
Cons	*****	*****	*****	*****	*****
2-C8	aatacatatc	tagaagttca	aaatctc-ct	ttttttatta	aaatgttaaa
1-C8	aatacatatc	tagaagttca	aaatctc-ct	ttttttatta	aaatgttaaa
1-C10	aatacatatc	tagaagttca	aaatctc-ct	ttttttatta	aaatgttaaa
Cons	*****	*****	*****	*****	*****
3-C1	attgacatga	aacaacagtt	tcaacattct	tggtggctgc	tgggagcaaa
3-C2	attgacatga	aacaacaatt	tcaacattct	tggtggctgc	tgggagcaaa
3-C5	attgacatga	aacaacagtt	tcaacattct	tggtggctgc	tgggagcaaa
Cons	*****	*****	** *****	*****	*****
2-C8	attgacatga	aacaacaatt	tcaacattct	tggtggctgc	tgggagcaaa
1-C8	attgacatga	aacaacaatt	tcaacattct	tggtggctgc	tgggagcaaa
1-C10	attgcacaga	aacaacagtt	tcaacattct	tggtggctgc	tgggagcaaa
Cons	*****	*****	** *****	*****	*****

Exon 5 (61 pb)

```

B          CAA CAAAAATATA ACCCCTTCAG
J          CAA CAAAGATATG ACCCCTACAA
3-C1      caggcctttc tttattgttc tgttttagCAA CAAAGATATA ACCCCGACAA
3-C2      cagccttttc tttattgttc tgttttagCAA CANAGATATA ACCCCTACAA
3-C5      cagccttttc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA
Cons      *****
2-C8      cagcccttgc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA
1-C8      cagcccttgc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA
1-C10     cagccttttc tttattgttc tgttttagCAA CAAAGATATA ACCCCTACAA
Cons      *****

B          ATACGTTGAG CTTTTCATAG TTGTGGACCA AGGAATGG
J          ATACATTGAG TTTTTCGTAG TTGTGGACCA AGGAACCG      Intron 5
3-C1      ATACGTTGAG TTTTTCATAG TTGTGGACCA AGGAATGGta agtaacctgg
3-C2      ATACGTTGAG TTTTTCATAG TTGCGGACCA AGGAATGGta agtaacctgg
3-C5      ATACGTTGAG TTTTTCATAG TTGTGGACCA AGGAATGGta agtaacctgg
Cons      **** *
2-C8      ATACGTTGAG TTTTTCATAG TTGTGGACCA AGGAATGGta agtaacctgg
1-C8      ATACGTTGAG TTTTTCATAG TTGTGGACCA AGGAATGGta agtaacctgg
1-C10     ATACGTTGAG TTTTTCATAG TTGTGGACCA AGGAATGGta ggtaacctgg
cons      **** *

3-C1      atatctttcc attttctttc tgcattggag ccaaccaata tttacaaagt
3-C2      atatctttcc attttctttc tgcattggac ccaaccaata tttacaaagt
3-C5      atatctttcc attttctttc tgcattggag ccaaccaata tttacaaagt
Cons      *****
2-C8      atatctttcc attttctttc tgcattggac ccaaccaata tttacaaagt
1-C8      atatctttcc attttctttc tgcattggac ccaaccaata tttacaaagt
1-C10     atatctttcc attttctttc tgcattggag ccaaccaata tttacaaagt
Cons      *****

                                     ← P0rr
3-C1      taagagtcac ttgtaaatat accctaattt gcacatttgt tcttctgct
3-C2      taagagtcac ttgtaaatat actctaattt gcacatttgt tcttctgct
3-C5      taagagtcac ttgtaaatat actctaattt gcacatttgt tcttctgct
Cons      *****
2-C8      taagagtcac ttgtaaatat actctaattt gcacatttgt tcttctgct
1-C8      taagagtcac ttgtaaatat actctaattt gcacatttgt tcttctgct
1-C10     taagagtcac ttgtaaatat actctaattt gcacatttgt tcttctgct
cons      *****

3-C1      tac
3-C2      tac
3-C5      tac
Cons      ***
2-C8      tac
1-C8      tac
1-C10     tac
Cons      ***

```

Figura 38. Sequenciamento P0dd-P0rr.

O par de *primers* P0dd, localizado no intron, e P0r, no exon, foi utilizado para definir o tamanho do intron. Esta região foi amplificada, clonada e sequenciada pelo

aluno de Mestrado Marcelo Moreira Tavares de Souza e confirmada neste trabalho. Exon 5 tem 61 pb, o intron 5 tem 546 pb e o exon 6 tem 71 pb. Notar a seqüência doadora universal *gt* em 5' do intron 5, e receptora *ag* em 3' dos introns 4 e 5 (sublinhado) (Fig. 39).

Intron 4 P0dd →

S	<u>g</u> ttagttt <u>g</u> g <u>ag</u> atttt <u>g</u> aat	caaaatggag	ttaaataaat	ctccaaatca	tttcaacttt
I	<u>g</u> gtagttt <u>g</u> g <u>ag</u> atttt <u>g</u> aat	caaaatggag	ttaaataaat	ctccaaatca	tgtcaacttc
	* *****	* *****	* *****	* *****	* *****

S	ctaaatcaat	ttttaaaaaat	aatcaaatta	tcaattttgga	tcaattttgca	tccaaagtga
I	ctaaatcaat	ttttaaaaaat	aatcaaatta	tcaattttgga	tcaattttgca	tccaaagtga
	*****	*****	*****	*****	*****	*****

S	gagttccttc	cttccttcct	tcctttttggt	ggcaaaagat	gaaactgcct	aatacatatc
I	gagttccttc	cttccttcct	tcctttttggt	ggcaaaagat	gaaactgcct	aatacatatc
	*****	*****	* *****	*****	*****	*****

S	tagaagttca	aaatctcact	ttttt-atta	aaatggttaa	attgacatga	aacaacagtt
I	tagaagttca	aaatctc-ct	ttttttatta	aaatggttaa	attgacatga	aacaacaatt
	*****	*****	*****	*****	*****	*****

B					CAA	
J					CAA	
S	tcaacattct	tgttggctgc	tgggagcaaa	cagccttttc	tttattggtc	tgttttagCAA
I	tcaacattct	tgttggctgc	tgggagcaaa	cagcctttgc	tttattggtc	tgttttagCAA
	*****	*****	*****	*****	*****	*****

Exon 5 (61 pb)

B	CAAAAATATA	ACCCCTTCAG	ATACGTTGAG	CTT TTCATA	GTTGTGGACC	AGGAATGG
J	CAAAGATATG	ACCCCTACAA	ATACATTGAG	TTT TTCGTA	GTTGTGGACC	AGGAACCG
S	CAAAGATATA	ACCCCTACAA	ATACGTTGAG	TTTTTCATAG	TTGTGGACCA	AGGAATGG <u>gt</u>
I	CAAAGATATA	ACCCCTACAA	ATACGTTGAG	TTTTTCATAG	TTGTGGACCA	AGGAATGG <u>gt</u>
	****	*****	*****	*****	*****	*****

Intron 5

S	aagtaacctg	gatatacttc	catttttcttt	ctgcattgga	gccaaccaat	atttacaag
I	aagtaacctg	gatatacttc	catttttcttt	ctgcattgga	ccaaccaat	atttacaag
	*****	*****	*****	*****	*****	*****

S	ttaagagtca	tttgtaaata	tactctaatt	gcacatttgt	tcttctgct	taccatctaa
I	ttaagagtca	tttgtaaata	tactctaatt	gcacatttgt	tcttctgct	taccatctaa
	*****	*****	*****	*****	*****	*****

S	atattttttt	atccccacacc	ggttaaacag	ttctgtaggt	ttctttgaat	cattcatgca
I	atattttttt	atccccacacc	ggttaaacag	ttctgtaggt	ttctttgaat	cattcatgca
	*****	*****	*****	*****	*****	*****

S	ctcttttgttt	gtccaaattg	tacaacacac	gaagaaaaag	atgtatagtg	aagaatgctt
I	ctcttttgttt	gtccaaattg	tacaacacac	gaagaaaaag	atgtatagtg	aagaatgctt
	*****	*****	*****	*****	*****	*****

S	atctcacaca	tctctgaaca	tggaaatata	gcatccaatt	taggaaagca	tttcagtagg
I	atctcacaca	tctctgaaca	tggaaatata	gcatccaatt	taggaaagca	tttcagtagg
	*****	*****	*****	*****	*****	*****

```

S ttgaccaaatt ttttctccca ccaggaaaaa agcccatgaa tgctgctgaa aagttaaaaa
I ttgaccaaatt ttttctccca ccaggaaaaa agcccatgaa tgctgctgaa aagttaaaaa
*****

S ccttaatata ttaatggtac aaatatagat taaaaaatca gcacaaaatc agcttcgata
I ccttaatata ttaatggtac aaatatagat taaaaaatca gcacaaaatc agcttcgata
*****

S aaaattatta ttatatcaac aaattcaatg atcctttcac tgatatcttc tttctctccc
I aaaattatta ttatatcaac aaattcaatg atcctttcac tgatatcttc tttctctccc
*****

S tcctttctttc ccttcgtcct ataattagtg atctcatctt agacattttt atcttttctg
I tcctttctttc ccttcgtcct ataattagtg atcttaatact taacattttt atcttttctg
*****
Exon 6 (71 pb) ← P0r
B TCACAA AAAACAATGG CGATTTAGAT AAGATAAAAG CAAGAAATGTA TGAACTTGCC
J TCACAA AAAACAATGG CGATTTAGAT AAGATAAAAG CAAGAAATGTA TGAACTTGCC
S taggTCACAA AAAACAATGG CGATTTAGAT AAGATAAAAG CAAGAAATGTA TGAACTTGCC
I taggTCACAA AAAACAATGG CGATTTAGAT AAGATAAAAG CAAGAAATGTA TGAACTTGCC
*****

B AAC
J AAC
S AAC
I AAC
***

```

Figura 39. Sequenciamento P0dd-P0rr. Sequenciamento parcial do domínio catalítico (continuação). As denominações S e I são provisórias, até a obtenção das seqüências definitivas.

Os *primers* P0dr–P2r encontrados no exon tiveram como produto da amplificação do PCR uma banda de ~750 pb (Fig. 40), indicando a presença de um intron, pois o tamanho esperado era 110 pb.

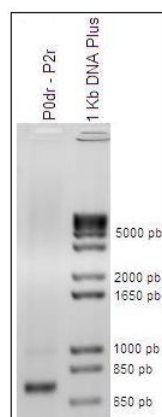


Figura 40. PCR P0dr-P2r.

A banda de interesse foi recortada, purificada e clonada. Os clones foram selecionados, 9 dos quais foram analisados por miniprep e comparados com pUC 19 (Fig. 41).

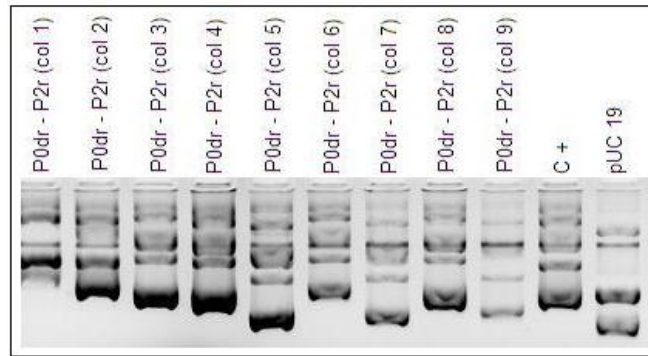


Figura 41. Minipreps P0dr-P2r.

Houve uma grande variação no tamanho dos produtos das minipreps. Para identificar a presença e o tamanho do inserto foi feita PCR. Os clones 4, 7 e 9 não têm inserto; os demais têm inserto do tamanho esperado (Fig. 42).

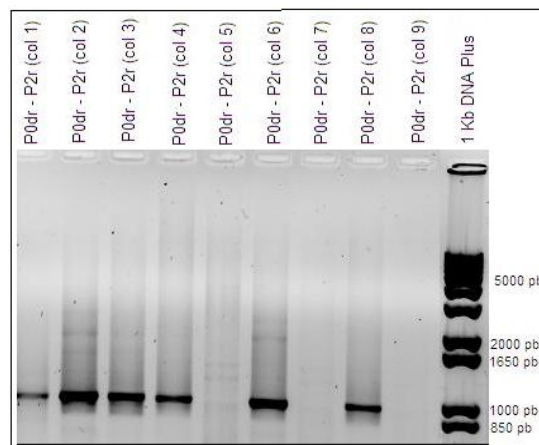


Figura 42. PCR das Minipreps P0dr-P2r.

A digestão com *EcoRI* mostrou duas bandas, de ~450 pb e ~350 pb, indicando existir um sítio *EcoRI* no meio da sequência, pois o tamanho esperado era ~800 pb (Fig. 43).

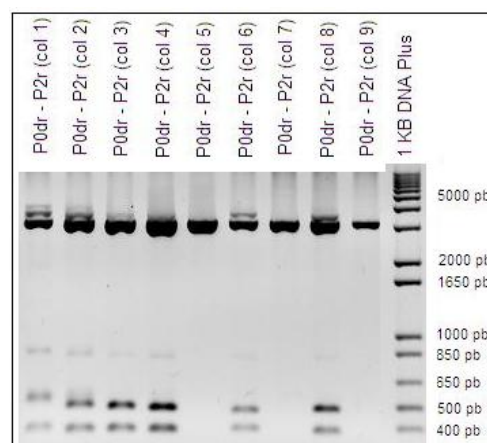


Figura 43. Digestão *EcoRI* P0dr-P2r.

Exon 7

```

B      AGA TTTTCAGATA CCTGTATATG CATGCAGCAC TGGTTGGCCT AGAAATTTGG
J      AGA TTTTCAGATA CCTGTATATG CATGTAGCAC TGGTTGGCCT AGAAATTTGG
S acctagcAGA TTTTAAGATA CCTGTATATG CATGCAGCAC TGGTTGGCCT AGAAATTTGG
I acctagcAGA TTTTAAGATA CCTGTATATG CATGCAGCAC TGGTTGGCCT AGAAATTTGG
  *****  *****  *****  *****  *****  *****  *****
          ← P2r
B TCCAATGGAG ATAAGATTAC CGTGAAG
J TCCAATGGAG ATAAGATTAC CGTGAAG
S TCCAATGGAG ATAAGATTAC CGTGAAG
I TCCAATGGAG ATAAGATTAC CGTGAAG
  *****  *****  *****  *****

```

Figura 44. Sequenciamento P0dr-P2r. O sítio *EcoRI* está indicado no box em cinza. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

Os *primers* P4d–P4ir localizados no exon 7 tiveram como produto da amplificação do PCR uma banda de ~ 130 pb (Fig. 45).

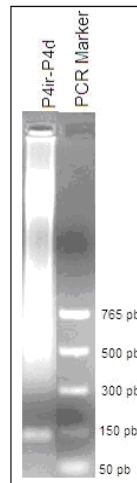


Figura 45. PCR P4d-P4ir.

A análise do PCR indicou não haver presença de intron. A banda foi recortada e clonada. Foram analisados 6 clones por miniprep (Fig. 46).

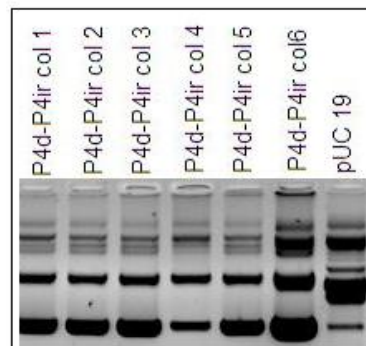


Figura 46. Minipreps P4d-P4ir.

Para identificar os clones com os tamanhos desejados foram feitos os PCRs a partir das minipreps com os *primers* SP6 e T7. Todos os clones apresentaram a banda do tamanho desejado exceto a colônia 10 (Fig. 47).

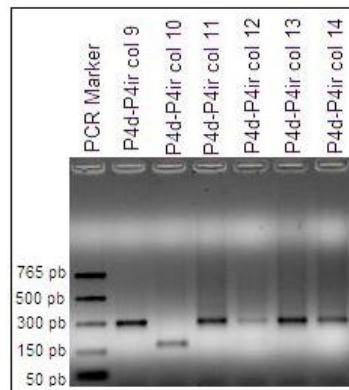


Figura 47. PCR das Minipreps P4d-P4ir.

O sequenciamento de quatro clones a partir de produto das minipreps (Fig. 48) indicou 100% de identidade.

	Exon 7		P4d →			
B	<u>GAGATAAGAT</u>	<u>TACCGTGAAG</u>	CCGGACGTGG	ATTATACTTT	GAATTCATTT	GCAGAATGGA
J	<u>GAGATAAGAT</u>	<u>TACCGTGAAG</u>	CCGGACGTGG	ATTATACTTT	GAATTCATTT	GCAGAATGGA
C	<u>GAGATAAGAT</u>	<u>TACCGTGAAG</u>	CCGGACGTGG	ATTATACTTT	GAATTCATTT	GCAGAATGGA
	*****	*****	*****	*****	*****	*****
						← P4ir
B	GAAAAACAGA	TTTGCTGACT	CGCAAAAAAC	ATGATAATGC	<u>TCAGTTACTC</u>	<u>ACAGCAATTG</u>
J	GAAAAACAGA	TTTGCTGACT	CGCAAAAAAC	ATGATAATGC	<u>TCAGTTACTC</u>	<u>ACAGCAATTG</u>
C	GAAAAACAGA	TTTGCTGACT	CGCAAAAAAC	ATGATAATGC	<u>TCAGTTACTC</u>	<u>ACAGCAATTG</u>
	*****	*****	*****	*****	*****	*****
B	<u>AC</u>					
J	<u>AC</u>					
C	<u>AC</u>					
	**					

Figura 48. Sequenciamento P4d-P4.

Para a amplificação dessa região foram utilizados os *primers* P11d presente na e P11r. O PCR gerou uma banda ~1.8 Kb indicando a presença de um intron, pois o produto esperado era 108 pb, como mostrado na Fig. 49.

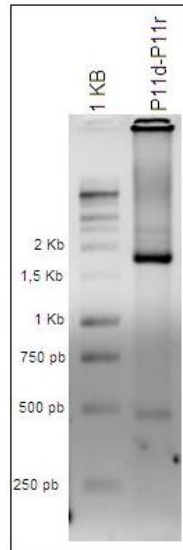


Figura 49. PCR P11d-P11r

A banda de ~1,8 Kb foi recortada do gel e clonada. Foram analisados 6 clones por miniprep. O clone P11d-P11r col 6 não tem inserto; os demais (exceto a col 7) têm inserto. Todos foram comparados ao pUC 19 (Fig. 50).

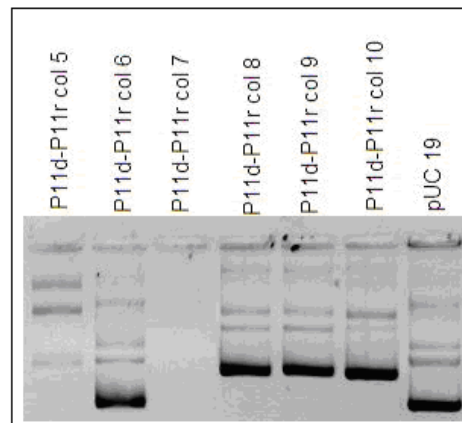


Figura 50. Minipreps P11d-P11rr.

Para confirmar o tamanho do inserto foram feitas digestões com *EcoRI*. Os clones P11d-P11r col. 5, 8 e 9 indicaram a presença de uma banda de ~ 1.8 Kb; a col 10 não apresentou inserto (Fig. 51).

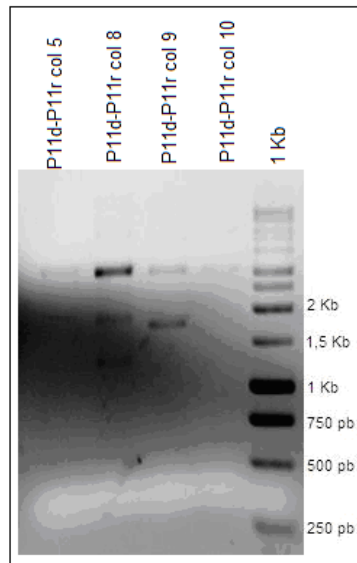


Figura 51. Digestão *EcoRI* P11d-P11r.

A partir dos clones P11d-P11rr foram feitos *Nested PCRs* com os *primers* P11dd-P11rr, obtendo sub-clones que foram analisados em conjuntos.

Os sequenciamentos foram feitos a partir dos fragmentos obtidos das minipreps dos clones Pbd–Pbr col 5, 8 e 9 e seus sub-clones, utilizando os *primers* SP6 e T7. Os clones 8 e 9 mostraram completa identidade de sequência, representados pela sequência superior (S). O clone 5 corresponde à sequência inferior (I). O Exon V tem 67 pb (sequenciamento parcial) em ambas as bandas. O intron v na sequência S tem 1758 pb e na sequência inferior 1694 pb. Foram identificados 103 pb divergentes entre as sequências (Fig. 52).

	P11d →		Exon V (parcial - 67 pb)			
B	<u>GCATTCATCA</u>	<u>TGACACAG</u>	AT TTCTGTTCTT	GTGGTGGTTA	CCCATGCATT	ATGGGTCCCG
J	<u>GCATTCATCA</u>	<u>TGACACAG</u>	GT TCCTGTTCTT	GTGGTGGTTA	CCCATGCATT	ATGGGTCCCG
S	<u>GCATTCATCC</u>	<u>TGACACAG</u>	AT TTCTGTTCTT	GTGGTGGTTA	CCCATGCATT	ATGGGTCCCG
I	<u>GCATTCATCA</u>	<u>TGACACAG</u>	GT TCCTGTTCTT	GTGGTGGTTA	CCCATGCATT	ATGGGTCCCG
	*****	*****	* *****	*****	*** *****	*****
B	GGATAAG					
J	GGATAAG	Intron v				
S	GGATAAGgta	aggattgagg	aaatcttctt	aatgcctttc	caatcaagtt	atTTTTaaat
I	GGATAAGgta	aggattgagg	aaatctt-tt	aatgcctttc	caatcaagtt	atTTTTaaat
	*****	*****	*****	*****	*****	*****
S	gcttgcaatt	aattggtcaa	ttaaataatac	tgagcaactt	acaaaaatta	aatggttaggg
I	gcttccagtt	aattggttaa	ttaaataatac	tgag-aactt	acaaaaatta	aatggttaggg
	*** ** *	*****	*****	*** *****	*****	*****
S	ggttacaaaa	aaagagggct	ttgaccattt	gtgtgtcagt	ttaccatgtg	tagttaaanaat
I	agttacaaaa	aaagagggct	ttggtcattt	gtgtgtcagt	ttaccatgtg	tagttaaanaat
	*****	*****	*** *****	*****	*****	*****

```

S  gtgttgatt  taaaaaggta  gaactgtcct  cagcatttga  atagccatcc  ttcaggagct
I  gtgtcggatt  taaaaag-ta  gaaatattct  caacaattga  atagccatcc  ttcaggagct
   **** * **  ***** **  *** * * **  ** *****  *****
S  tcctccaaaa  cgccagtagt  ttaaagaaac  agcataggaa  gcatagccaa  agaaacccaa
I  tcctccaaaa  caccag-agt  ttaaagaaac  agcgtaggaa  gcatagccaa  agaaacccaa
   ***** * **** **  *****  *** *****  *****
S  tattaatatc  tgtgcatgga  taaaagagga  aaccaagtta  gacattaaaa  ctttcctaaa
I  tattaatatc  tgtg-atgga  taaaagagga  aaccaagtta  gacgtt----  ---tcctaaa
   ***** **** *****  *****  *** **  *****
S  gtgtgatgta  gaaagggcaa  aaaattcttc  tgaatcaggc  aacagcagga  agaaaaggat
I  gtgtgaagca  aaaaggggaa  aaaattattc  tgaaccaggc  aacagcagga  agaaaaggat
   ***** * *  ***** **  ***** ***  *** *****
S  aaattagcat  cttggggttt tctgcctcg  ccctttcatc  aaggtagaat  acgttacatc
I  aaattagcat  cttggggttt tctgcctcg  ccctttcatc  aaggtagaat  acgttacatc
   ***** ***** *****  ***** *****  *****
S  tctgtcatat  attgcagtct  tctgaaatca  gcaggtagca  acgtacagga  agttattgac
I  tctgtcatat  attgcagtct  tctgaaatca  gcaggtagca  acgtacagga  agtcattgac
   ***** ***** *****  ***** *****  *** *****
S  ttacaaccat  ttgttttagca  accgttcaaa  gttacagtgg  ccctgagaac  agtgacttac
I  ttacaaccat  ttgtt-----  ---gttcaaa  gttacaatgg  ccctgagaac  agtgacttac
   ***** *****  ***** *****  ***** *****
S  gactggttct  cgcaattaca  gccatgcgag  cctcccttca  gtcacagaat  cgtgtattta
I  gactagtcct  tgcaattacg  gccatgcgag  cctcccttca  gtcacagaac  cgtgtattta
   **** ** **  ***** *****  ***** *****  *****
S  tgacgcttgc  agcatcctgc  attcatggga  ttctcattgg  ggactttccc  agctgtcttc
I  tgacgcttgc  agcatcctgc  attcatggga  tcctcattgg  ggactttccc  agctgtcttc
   ***** ***** *****  * ***** *****
S  ccatatgcaa  agccaatgga  gcaagctggg  cccttaatg  attgtgtggt  tcacttaaca
I  ccatatgcaa  agccaatgga  gcaagctggg  cccttaatg  actgtgtgat  tcacttaaca
   ***** ***** *****  ***** * ***** *  *****
S  aagtggcaaa  gccggttgta  aaacggagca  cgactcactt  aacaactggc  ttgctgggca
I  aagcggcaac  actggttgta  aaactgagca  cgactcactt  aacaactggc  tagctgggca
   ** ***** * *****  *** ***** ***** ***** *****
S  gtggaaattc  tgccccagtt  gtgacag--a  agtcaaggac  taccaatggt  ttcttccggg
I  atggaaattc  tgccccatt  gtgacagcca  agtcaaggac  taccatatt  ttcttctggg
   ***** ***** **  ***** *  ***** ***** ***
S  tggaagttta  gcccttactt  catcaagaag  gaaaaagaaa  aaggaaaaag  gaagcaaaat
I  tggaagttta  gcccttactt  catcaagaag  gaaaaagaaa  aaggaaaaag  gaagcaaaat
   ***** ***** ***** ***** ***** *****
S  cttctcagca  attgtctggt  gcaaaaactcc  ccacaacttt  tgtgtaatga  gttttgtgta
I  cttctcagca  attgtctggt  gcaagactcc  ccataattta  ttggtaatgt  gttttgtgta
   ***** ***** ***** ***** * ***** *****

```

S ggctcagtga gcatcccggc cattgatcaa tttctggttt cttcggctaa agcccagagga
 I ggctcagtga gcatcccggc tattgatcaa tttctgtttt ctttggctaa agcctgaaga

 S cttcattgcc tcccatttca tgcaagggca aggaggccag gaagtatgga gaatgccacg
 I cttcattgcc tcccatttca tgcaagggca aggaggccag gaagtatgga gaatgccatg

 S tagcctttcc tgtccttcct gaa-ggattt -ccgggcac atgattcctc ttttaatactt
 I tagcctttcc tgtccttcct aaatggattt gccgggcac atgattcctc ttttaatactt

 S ctagacactt tgtggtggtt tgtacggaag actttgttta gaaacaatt ttgctca-at
 I ctagacactt tatggtggtt tgtacagaag actttgctta gaaacaatt ttgctcatal

 S taacccccgcc aaaggggtct ggtgtaactt tgaccttggga agaagaagac catcttgtct
 I taacccccaca aaaggggtct ggcggatatt tgaccttggga agaagataac catcttgtct

 S atttaatggc caccagcaga ttctctttgg cgttcgtggt tccaccccat aacttgagag
 I atttaatggc caccagcaga ttctctttgg tgttcttgggt tccatccctt aatttgaggg

 S ccaatgatta tattcccaat gatttgaatt aacattagga aggtgctttc ctgggcttat
 I ccaatgatta tattcccaat gatttgaatt aacattagga aggtgctttg ctgggcttat

 S gatgtcacia gatgcagctc acaactaata tcacatggag ttctcattgc aaatgtattt
 I gatgtcacia gatgcagctc acaactaata tcacatggag ttctcattgc aaatgtattt

 S tgttaaaaaa gaaaggaaag ttagtggaat taaaactgag ggagtataat ctgggactaa
 I tgttagaaaa gaaaggaaag ttagtggaat taaaactgag ggagcataat ctgggaccaa

 S gatcaagggt tacgaagggt tctcttatct attctggttag ctttagagaa aaaaatcact
 I gatcaagggt tacgaagggt tctcttatct attttggttag ttttagagaa aaaaatcact

 S gcatttcttc attagccaac cttttttttc ttatatgttt ttgcaatgaa ttctgttcct
 I gcatttcttc attagccatc cttttttttc ttatatgttt ttgcaatgaa ttctgttcct

 B Exon X (178 pb)
 J C AATGAACCTT
 S agtccaaata tgggtgatgt catggtctta ttttggttca tctccacagC AATGAACCTT
 I agtccaaata tggatgatgt catggtctta ttctggttca tgtctacagC GATGAACCTT

← P11r

B CGAAATTTT CAGCAATTGT AGTTATATCA

J CGAAATTTT CAGCAATTGT AGTTATATCA

S CGAAATTTT CAGCAATTGT AGTTATATCA

I CGAAATTTT CAGCAATTGT AGTTATATCA

Figura 52. Sequenciamento completo P11d-P11r. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

Os *primers* P9d-P9r localizados no exon X tiveram como produto da amplificação do PCR uma banda de ~150 pb (Fig. 53).

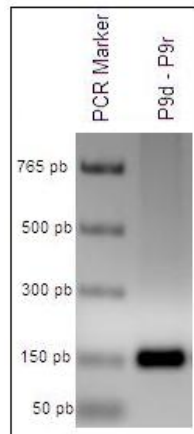


Figura 53. PCR P9d-P9r foi ~150 pb, próximo ao esperado, indicando não possuir intron.

A partir do produto de PCR foi feita a clonagem, tendo sido analisadas 7 colônias pela técnica de miniprep (Fig. 54).

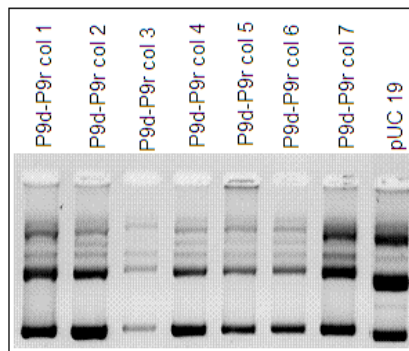


Figura 54. Minipreps P9d-P9r.

Devido á pouca diferença de migração entre o pUC controle e os clones foram feitos PCRs para confirmar a presença de introns. Todas as colônias apresentaram a banda com o tamanho esperado (Fig. 55).

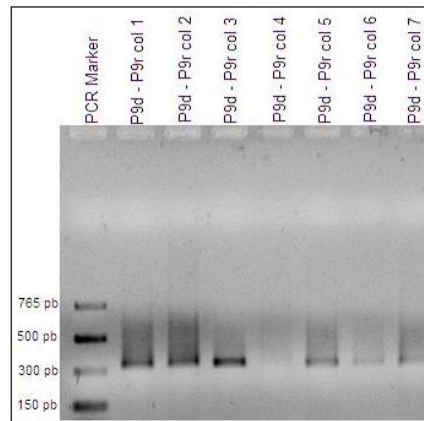


Figura 55. PCR das Minipreps P9d-P9r.

Foram sequenciadas as 7 colônias. A banda S é consenso das colônias 1 e 5; a banda I é o consenso das colônias 2, 3, 4, 6 e 7 (Fig. 56).

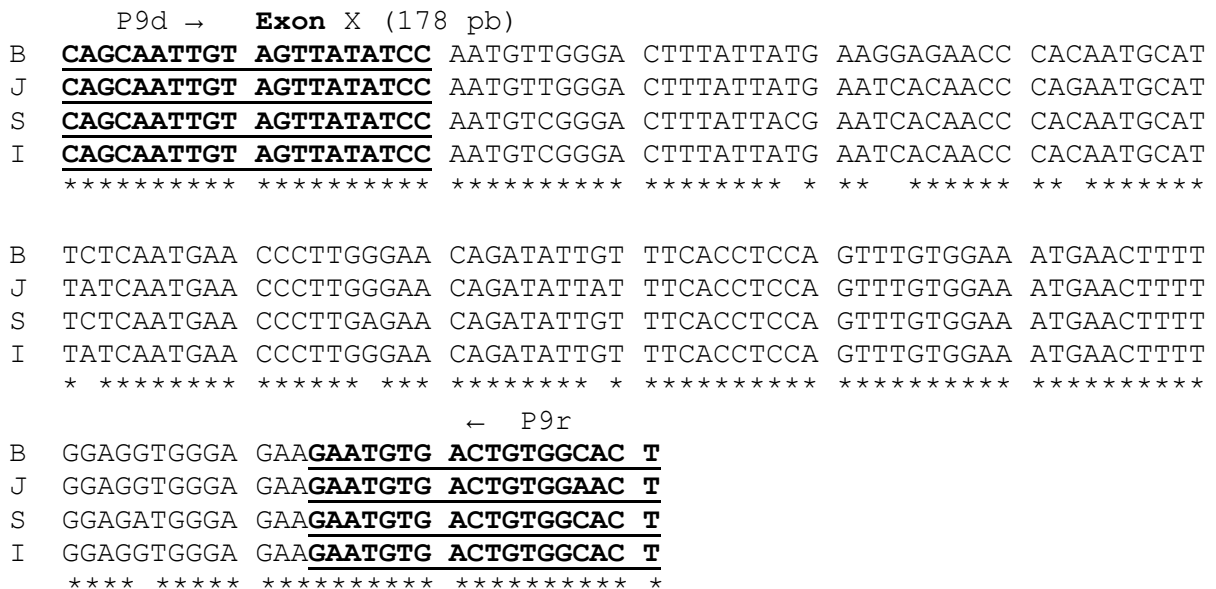


Figura 56. Sequenciamento P9d-P9r. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

5.4 Domínio Desintegrina

O domínio desintegrina é mostrado no box azul, compreendendo do nucleotídeo 1308 ao 1586, usando como base o cDNA da bothropasina. Os asteriscos mostram identidade de sequências; a falta deles indica divergência entre os cDNAs da jararagina e bothropasina. As barras verticais indicam presença de introns identificados neste trabalho.



Para amplificar o início desta região foram utilizados os *primers* P6d – P6r (Fig. 57), obtendo-se 2 bandas com mais de 1 Kb que foram nomeadas 1 e 2, indicando a presença de um intron, pois o produto esperado era de 106 pb.

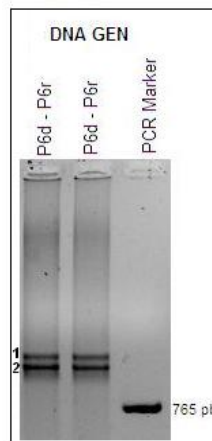


Figura 57. PCR P6d-P6r.

As bandas foram recortadas juntas, purificadas e clonadas. Os clones foram analisados após miniprep, e comparados com pUC 19 íntegro (Fig. 58).

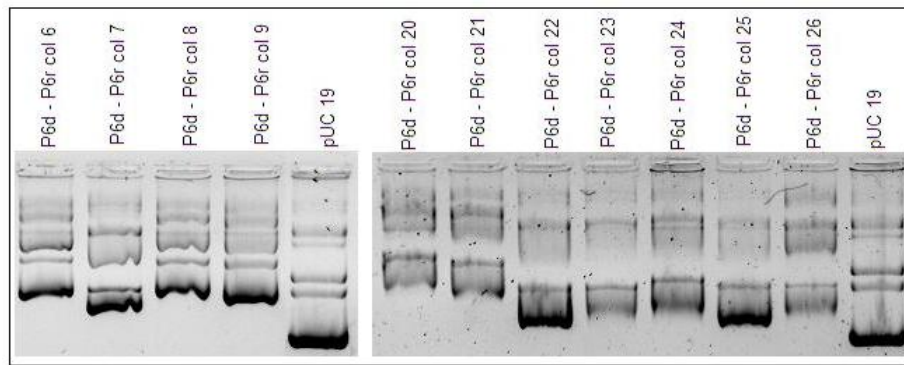


Figura 58. Minipreps P6d-P6r.

Foram feitos PCRs utilizando os *primers* SP6 e T7. Os clones P6d–P6r col 6, 8, 9, e 20 têm o inserto correspondente à banda 2. O clone P6d–P6r col 21 apresentou uma banda na altura de 1 Kb, correspondente à banda 1 (Fig. 59).

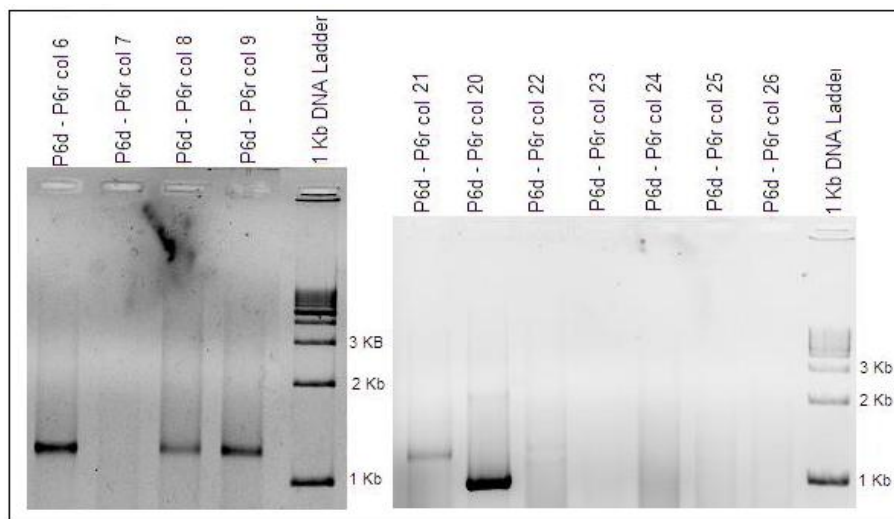


Figura 59. PCRs das Minipreps P6d-P6r.

A digestão com *EcoRI* (Fig. 60) confirmou o resultado do PCR. Os clones P6d–P6r col 6, 8, 9 e 20 mostraram ter o inserto correspondente à banda 2. Os clones P6d–P6r col 21, 24, 25 e 26 contêm a banda1. Os demais clones não tinham inserto.


```

S agcaatgggg catggcaaac atctcaagaa gaacacctcc cacccattct cttggtctat
I -----

S aaagatgagg tggtagaaat atactttcag tattgaaaga tttttctaca gtcaccttac
I -----

S aatcagggta atgttaatac tcatggttgg tgcttcttct ctggactctc aaaggctggc
I ----- -tggattct- aaaggctggc
                **** *
                *****

S atgattacga gaaaatttct catgaataag aaaggattgg ggctgtaagt tacgtgggct
I atgattatga gaaaattttt catgaataag aaaggattgg ggctgtaagt tacgtgggct
  ***** * *****
  *****

S tcacaattag gaaatgagga tatttttggtt ttattcttgt cagagggaaa tatcaggtaa
I tcacaattag gaaatgagga tatttttggtt ttattcttgt cagagggaaa tatcaggtaa
  ***** * *****
  *****

S ggctctttct cagagagatg tcattatctg tggcaataac aaacaatttg tgcatttgct
I ggctctttct cagagagatg tcattatctg cggcaataac aaacaatttg tgcatttgct
  ***** * *****
  *****

S agcatgaacc cataagaggg aacacattgc agaaatgtct ttttccttga caatttcctt
I agcatgaacc cataagaggg aacacattgc agaaatgtct ttttccttga cagtttcctt
  ***** * *****
  *****

S cttcaaaata caccaataaa aaaagaccag atagaaaatt ctctgtcatc atttgatatg
I cttcaaaata caccaataaa aaaagaccag gtagaaaatt ctctgacatc atttgatagc
  ***** * *****
  *****

                Exon W (90 pb)
B                AATTGTC AAAATGAGTG CTGCGATGCT GCAACGTGTA AACTGAAATC
J                AATTGTC AAAATGAGTG CTGCGATGCT GCAACGTGTA AACTGAAATC
S ttttggtttt cagAATTGTC AAAATGAGTG CTGCGATGCT GCAACGTGTA AACTGAAATC
I ttttggtttt cagAATTGTC AAAATGAGTG CTGCGATGCT GCAACGTGTA AACTGAAATC
  ***** * *****
  *****

                ← P6r P7d →
B AGGGTCACAG TGTGGACATG GAGACTGTTG TGAGC
J AGGGTCACAG TGTGGACATG GAGACTGTTG TGAGC
S AGGGTCACAG TGTGGACATG GAGACTGTTG TGAGC
I AGGGTCACAG TGTGGACATG GAGACTGTTG TGAGC
  ***** * *****
  *****

```

Figura 61. Sequenciamento P6d-P6r. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

Para amplificação foram utilizados os *primers* P7d – P7r, obtendo-se somente uma banda, com mais de 1 Kb, indicando a presença de um intron, pois o produto esperado era de 76 pb (Fig. 62).

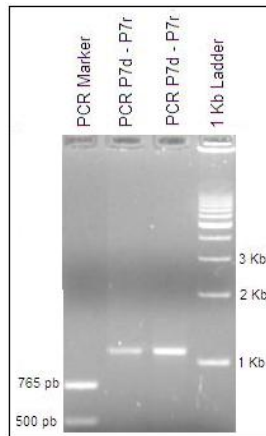


Figura 62. PCR P7d-P7r.

O produto de PCR foi ligado ao plasmídeo e usado para transformar as bactérias competentes. Os clones foram analisados por miniprep (Fig. 63). Os clones P7d – P7r col 4, 5 e 6 foram comparados com pUC 19 íntegro confirmando a presença do inserto.

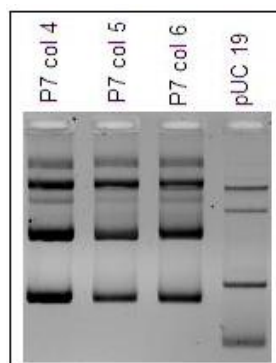


Figura 63. Minipreps P7d-P7r

O inserto foi identificado por PCR utilizando os *primers* SP6 e T7. Os clones P7d – P7r col 4, 5 e 6 apresentam a banda de tamanho esperado (Fig. 64).

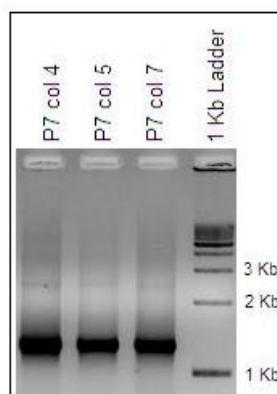


Figura 64. PCR das minipreps P7d-P7r.

Os sequenciamentos foram feitos a partir das minipreps. O Consenso (C) dos 3 clones P7d- P7 mostrou 100 % de identidade. Foi identificado um intron de 1022 pb, e o tamanho do Exon W foi definido, 90 pb. Notar a seqüência doadora universal *gt* em 5', e receptora *ag* em 3' do intron y (sublinhado) (Fig. 65).

P7d → **Exon W (90 pb)**

B CATGGAGACT GTTGTGAGCA ATGCAAA
 J CATGGAGACT GTTGTGAGCA ATGCAAA **Intron w**
 C CATGGAGACT GTTGTGAGCA ATGCAAAgta agacttgttt atgtttaaaca ccaggagaga
 ***** ***** *****

C ttttattctg ctccatacta gccatacaga aatggctggt tactaacct cttcctctcc
 C ctttcttcca gtttatttta cccttatgaa catatccata gggaagataa ttaacaaaa
 C tttcagcctt gtctcaatct caaatgcact ctttcagcat gttaaacgt atctgtgaaa
 C ataatatatt tcttctttga ctgagattgc atggaaacc agtttaaaca aggataggca
 C atatttgaga ttggtgccct aactcagctt cctgactttc tggaagcttc taagaggctc
 C ctggtaatgc tgtgacattt tcttctctga gccttttagg atggaaatag gtgcaggaga
 C cttctggaag taaagttgct tttttccca ccttaagttc tctacctgct ctctaaaagc
 C tctaaattca ggtgttttga tggctgcagg accatgaaaa gagaggtgca agttactcat
 C tgcttctttc tatgtgggat cccagttgac tctgtaatga actttttgag cagagtggcc
 C caaacattt tgttatttcc atatttccat cacaagccta gattcacaac aagagaaggg
 C aaccacatgt ttttcagcat gagacagaaa atgctatgaa tgcttcttcc catgtaaaga
 C aataaaatac atcatgagat gttcagcaat tcaacttttg ctgcttttac atgacagccc
 C acttgatttt ccctttatgg tcagccaaca ggtagaactt gtctttcagg aattgagcct
 C ttcattgcat tcatttcccc acagcaaata agacagagtg ggacttctag gccccacaca
 C gagttgtagc agggaaggga tgccttgctt ggtgatcctc aagacagatg aagaggaggt
 C tttgaaatgt gttgtgaacc atggtttgac tctttgatct ctgctgctga agagtgatag

Exon Y

B T TTAGCAAATC AGGAACAGAA TGCCGGGCAT
 J T TTAGCAAATC AGGAACAGAA TGCCGGGCAT
 C ctgggagtat ttttgattct caccacagT TTAGCAAATC AGGAACAGAA TGCCGGGCAT
 * ***** ***** *****

← P7r

B CAATGAGTGA ATGTGACCCG GCTGAACACT GCACTGG
 J CAATGAGTGA ATGTGACCCG GCTGAACACT GCACTGG
 C CAATGAGTGA ATGTGACCCG GCTGAACACT GCACTGG
 ***** ***** ***** *****

Figura 65. Sequenciamento P7d-P7r.

Para amplificação desta região foram utilizados os *primers* P12d – P12r, obtendo-se uma banda de tamanho esperado 170 pb, indicando não possuir intron (Fig. 66).



Figura 66. O produto do PCR P12d-P12r.

O produto da amplificação com os *primers* P12d – P12r foi recortado, ligado no plasmídeo e usado para transformar as bactérias competentes *E. coli* DH-5 α . Os 2 clones foram analisados por miniprep (Fig. 67).

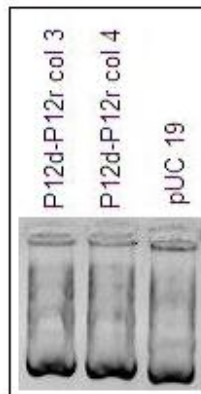


Figura 67. Minipreps P12d-P12r.

Para confirmar a presença do inserto foi feito PCR com os *primers* SP6 e T7 (Fig. 68).

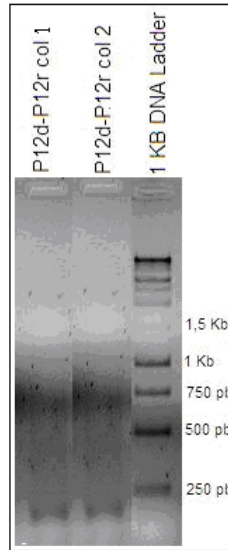


Figura 68. PCR das minipreps P12d-P12r. Confirmação da presença e tamanho do inserto.

As duas colônias P12d-P12r foram sequenciadas (Fig. 69). O consenso mostrou identidade com as seqüências da bothropasina (B), e jararagina (J), exceto por um C (cisteína) que está presente no cDNA da bothropasina.



Figura 69. Sequenciamento P12d-P12r.

5.5 Domínio Rico em Cisteína

O domínio rico em cisteína é mostrado no box rosa, compreendendo do nucleotídeo 1587 ao 1899. A região sem cor está em processo de análise (base o cDNA da bothropasina). Os asteriscos mostram identidade de seqüências; a falta deles indica divergência entre os cDNAs.

		Exon Y (206 pb)	P10d →	
Jararagina		CGGTTACTGCTACAATGGGAATTGCCCCATCATGTATCACCAATGTTATGCTCTCTTTGG		1430
Bothropasina		CGGTTACTGCTACAATGGGAATTGCCCCATCATGTATCACCAATGTTATGCTCTCTTTGG		1620

	Intron y	Exon Z (238 pb)		
Jararagina		TGCAGATGTTTATGAGGCTGAAGATTCATGCTTCAAAGATAACCAGAAAGGCAATTATTA		1490
Bothropasina		TGCAGATGTTTATGAGGCTGAAGATTCATGCTTCAAAGATAACCAGAAAGGCAATTATTA		1680

			P8d → ← P10r	
Jararagina		TGGCTACTGCAGAAAGGAAAATGGTAAAAAGATTCCATGTGCACCAGAAGATGTAAAATG		1550
Bothropasina		TGGCTACTGCAGAAAGGAAAATGGTAAAAAGATTCCATGTGCACCAGAAGATGTAAAATG		1740

			← P8idR P8id →	
Jararagina		TGGCAGGTTATACTGCAAAGATAATTCACCTGGACAAAATAATCCTTGCAAGATGTTCTA		1610
Bothropasina		TGGCAGGTTATACTGCAAAGATAATTCACCTGGACAAAATAATCCTTGCAAGATGTTCTA		1800

			← P8r P5d →	
Jararagina		TTCCAACGATGATGAACATAAGGGAATGGTTCTTCTGGAACAAAATGTGCAGATGGAAA		1670
Bothropasina		TTCCAACGATGATGAACATAAGGGAATGGTTCTTCTGGAACAAAATGTGCAGATGGAAA		1860

	Intron z	Exon K (36 pb)		
Jararagina		GGTGTGCAGCAACGGGCATTGTGTTGATGTGGCTACAGCCTACTAGTCAACCTCTGGCTT		1730
Bothropasina		GGTGTGCAGCAACGGGCATTGTGTTGATGTGGCTACAGCCTACTAGTCAACCTCTGGCTT		1920

Os *primers* utilizados para amplificar o início do domínio rico em cisteína foram P12d, localizado no final do domínio desintegrina, e P10r. O fragmento esperado era de 251 pb, mas o produto de amplificação foi ~2 Kb, indicando a presença de um intron (Fig.70).

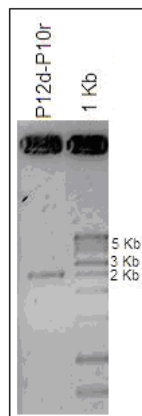


Figura 70. PCR P12d-P10r.

A banda foi recortada e clonada. Os clones foram analisados por miniprep e comparados com pUC 19 (Fig. 71).

C agtctattac aaattttcag catggtttta tattgaaaa gaaactgaaa tactgcacac
 C aatatttcca aagcattata aagtggact aacacttcat gtaatcttga ttttatccct
 C ctgtagtgt agcctaatta tagtgtaaaa cgattctgcc actattgcaa aagtgtagcc
 C tagattatgt gggatTTTT gcagttgcaa cacactgctg gctcttattt aagtgattgt
 C ccactagggc tccaagattc ctctcacagt tactaatatt gagccaggta tcacctcttc
 C tgtgtatttg ttttttcttt cctgaatgta gattggtttg ttttttact ttttcgctag
 C aaatttaatt ttcttagatt gggctcaatg ttcaaaagat cttgtggat ctttaagcct
 C atcttctgga gagattgcta ttcctgccag tatgggtgtc tctgcaaatt tattgagttc
 C cccttccacc ctcttgctta aattgtttat gaagggtcag gtatcatgtg atgaagattg
 C ggaatctgaa ttacaaatca gaacgttgct caacttgcag gattaattga tcaattgtta
 C ttttgatcag gtaaatagca taaatgaggt aacattccaa agaatagtaa aagctcagta
 C ctttgcaagt aatccaattg tattcaaga ggtttgtgcc caggaattga caggtctcaa
 C actaagaatg acacatgatg ggtatcttca agctgaattc tttcttgttt cattcctata
 C gtcaaaatct tgccagataa ggactctact aattggcttt tacagttacc ctctgtaaac
 C tattagggag ggagaggtcg tcaccctctt tttctcacac tttatatcca ggcagagttg
 C cctctgactc ctctggattt cacccttccc ttctgggaat gcatttcttc acggcattcc
 C agcacaggaa tgtacccatg atagagctaa agttgaaaaa tgtggacaca gaaagaaatc
 C tggattcaga ttcccctaca gtcattgatt taccacatgg tcttctgcaa gttgcattgt
 C agtgacttca gctttgtgtc cataaaatag aataagcaag tttagtgtaa cactgtatgc
 C gatagtgaag gcttcataat gaggaatagt taaataagta agtcaatatt aactaaatta
 C tttagaacac ctctgaaatc acttttagcca ttctttttcc tgctggcatg ttacacattc

Exon H

B ATGTT TATGAGGCTG AAGATTCATG
 J ATGTT TATGAGGCTG AAGATTCATG
 C atgtatttgt cttgggtcct cactccctct tacagATGTT TATGAGGCTG AAGATTCATG
 ***** *****

 B CTTCAAAGAT AACCAGAAAG GCAATTATTA TGGCTACTGC AGAAAGGAAA ATGGTAAAAA
 J CTTCAAAGAT AACCAGAAAG GCAATTATTA TGGCTACTGC AGAAAGGAAA ATGGTAAAAA
 C CTTCAAAGAT AACCAGAAAG GCAATTATTA TGGCTACTGC AGAAAGGAAA ATGGTAAAAA
 ***** ***** ***** ***** ***** *****

← P10r

B **GATTCCATGT GGACCAGAAG**

J **GATTCCATGT GCACCAGAAG**

C **GATTCCATGT GCACCAGAAG**

Figura 72. Sequenciamento parcial do Exon H. Em negrito e grifado estão os *primers* utilizados para amplificação.

Foram utilizados para amplificação os *primers* P8id-P8r, resultando em um produto de PCR esperado de ~140 pb, indicando não possuir intron (Fig. 73).



Figura 73. PCR P8id-P8r.

O produto de PCR foi ligado em um plasmídeo e inserido em bactérias competentes *E. coli* DH 5 α . Os clones foram analisados por miniprep (Fig. 74).

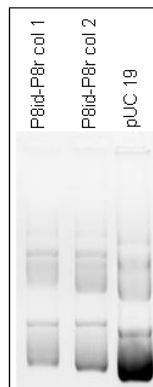


Figura 74. Minipreps P8id-P8r.

A presença do inserto nos clones P8id-P8r foi confirmada por PCR com os *primers* SP6 e T7 (Fig 75).

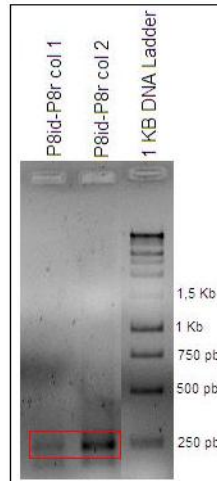


Figura 75. PCR das minipreps P8id-P8r.

O sequenciamento P8id-P8r referente aos 2 clones analisados mostrou 100 % de identidade entre eles. Alguns pontos de mutação foram observados em relação às sequências da bothropasina (B) e da jararagina (J) (Fig. 76).



Figura 76. Sequenciamento P8id-P8r.

Para amplificar o final desta região foram utilizados os *primers* P5d, presente no domínio rico em cisteína, e P5r, localizado no início da região 3'-não-traduzida. O produto do PCR esperado era de 161 pb, mas o PCR apresentou 3 bandas, nomeadas de 1 (~750 pb), 2 (~600) e 3 (570 pb) (Fig. 77).

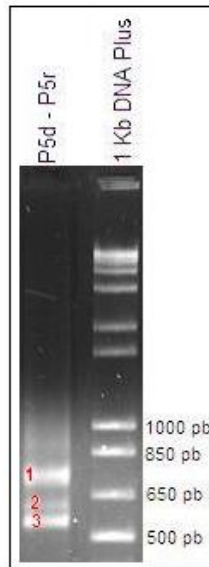


Figura 77. O PCR P5d-P5r

As bandas foram cortadas juntas (*pool* das três bandas), purificadas e clonadas. As colônias selecionadas foram analisadas por miniprep, e foram comparadas com pUC 19 íntegro (Fig. 78).

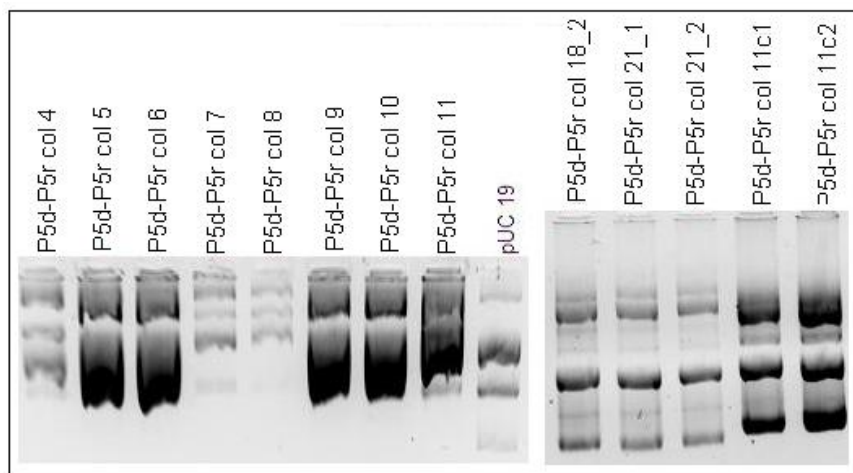


Figura 78. Minipreps P5d-P5r. Foram feitas 13 minipreps P5d-P5r para localizar as 3 bandas.

As minipreps mostraram uma variação de tamanho, sugerindo ter as três bandas. Para verificar o tamanho dos insertos foi feito PCR com os *primers* SP6 e T7. Os clones P5d-P5r col 10 e 11 não continham inserto. Os clones P5d-P5r col 11, 11c1 apresentaram o inserto de 750 pb (banda 1). Os clones P5d-P5r col 7, 8, 18_1 e 18_2 apresentaram o inserto de 600 pb (banda 2). Os clones P5d-P5r col 4, 5, 6, 21_1 e 21_2 apresentaram o inserto de 570 pb (banda 1) (Fig. 79).

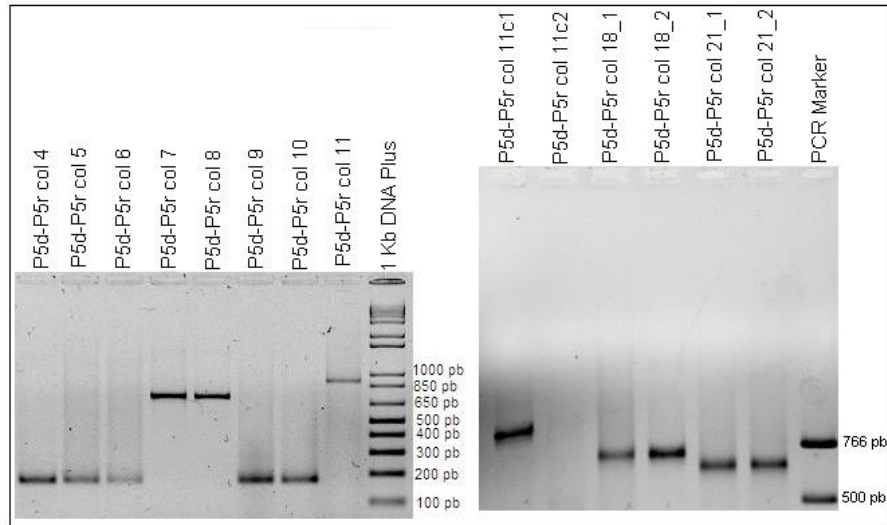


Figura 79. PCRs das Minipreps P5d-P5r

Foram feitos os sequenciamentos referentes às três bandas clonadas. O tamanho exato das três bandas pode ser definido pela sequência: 1 (753 pb) clones P5d-P5r col 11c2 e 11c1; 2 (479 pb) clones P5d-P5r col 21-1 e 21-2; 3 (550 pb) clones P5d-P5r col 8d e 18-2 (Fig. 80).

P5d → **Exon Z**

```

B      TCCTGGAA CAAAATGTGC AGATGGAAAG
J      TCCTGGAA CAAAATGTGC AGATGGAAAG
1      TCCTGGAA CAAAATGTGC AGAGGGAAAG
3      TCCTGGAA CAAAATGTGC AGAGGGAAAG
2      TCCTGGAA CAAAATGTGC AGAGGGAAAG
***** ***** *** *****

B GT
J GT intron z
1 GTtcgtaaaa gattacctct catctttgtg ttctaaagtc tgattc---- -----t-aa
3 GTtagtaaaa gattacctct aatctgtgtg ctctaaagtg tgattccatg gggtgatcaa
2 GTtagtaaaa gattacctct aatctgtgtg ctctaaagtc tgattccatg gggtgatcaa
  *** ***** ***** ***** ***** ***** * **

1 ggggtgatca ctaattcttt ttgcaatcg atcaatgctg aaagctggcc tagattttca
1 aagtacagag aatctgcata gaagaaaag tatccatcca tctatccttc ttttggttgt
1 tgttattttg cttttccttc aaagaacaac cccaacgaat gaggtggatg tccagggctg
1 tttcttttctt ccaagaccaa aatgcttggc cttctcagag ctttgtgctt ctctctcac-
1 ----- ----- ----- --agcatctt ctggacagac -----a-gag
2 taaaataatt ttgcaatcga tcaatgctgt aaagcatctt ctggacagac tgtcaa-gag
3 taaaataatt ttgcaatcga tcaatgctgt aaagcatctt ctggacggac tgtcaa-gag
          ***** ***** *** * **

```

```

1 acagtccttt ctctgaactt gtaggcatct ccatagggaa ggtggaagag agaaat-aga
3 acagtccttt ctctgaactt gtaggcatct ccatagggaa ggtggaagag agaaatgaga
3 acagtccttt ctctgaactt gtaggcatct ccatacggat ggtggaagag agaaatgaga
*****
1 aaatcagggc agagctgggt gtgacctaac aatgaagcca cttagaaatg ctgaagcaaa
2 aaatcagggc agagctgggt gtgacctaac aatgaagcca cttagaaatg ctgaagcaaa
3 aaatcagggc agagctgggt gggacctaac aaagaagcca tttagaaatg ctgaagcaaa
*****
1 cccaaaactt acctgaaagg atcaggaa-t cacttcccct tgatttttta- -----g
2 cccaaaactt acctgaaagg atcatgaa-t cacttcccct tgatttttta- -----g
3 cccaaagcct cccttaaagg atca-gaagt cacttcccct tgatttttta- tactatatag
*****
1 aacctgaaag aagtttgggt tagttataag gaatgacaga gtttgggtg- tgtagaaaag
2 aacctgaaag aagtttgggt tagttataag gaatgacaga gtttgggtg- tgtagaaaag
3 aacctgaaag aagtttcggt tagtt----- -----ca-- ----- ----gaaa-g
*****
                                           Exon K
J                                           GTGCAG CAACGGGCAT
B                                           GTGCAG CAACGGGCAT
1 tgctgcctta ctccattgat aatctcttgc tttgactttc aggtCTGCAG CAACAGGCAG
2 tgctgcctta ctccattgat aatctcttgc tttgactttc aggtCTGCAG CAACAGGCAG
3 tgctgtctta ttccattgat catctcttgc tttgactttc aggtCTGCAG CAACGGGCAT
*****
J TGTGTTGATG TGGCTACAGC CTA CTACTAGTCA ACCTCTGGC- -----T TTGATTTTGG
B TGTGTTGATG TGGCTACAGC CTA CTACTAGTCA ACCTCTGGC- -----T TTGATTTTGG
1 TGTGTTGATG TGA CTACTACAGC CTA CTACTAATCA ACCTCTGGCT TCTCTCAGAT TTGATTTTGG
2 TGTGTTGATG TGA CTACTACAGC CTA CTACTAATCA ACCTCTGGCT TCTCTCAGAT TTGATTTTGG
3 TGTGTTGATG TGGCTACAGC CTA CTACTAGTCA ACCTCTGGC- -----T TTGATTTTGG
*****
                                           ← P5r
B AGATCCTCCT TCCAGAAGGT TTGGCTTCTC TCAAGTCCAA AGAGATCCAT CTGCC
J AGATCCTCCT TCCAGAAGGT TTGGCTTCTC TCAAGTCCAA AGAGATCCAT CTGCC
1 AGATCCTCCT TCCAGAATGT TTGGCTTCCC TGTAGTCCAA AGAGATCCAT CTGCC
2 AGATCCTCCT TCCAGAATGT TTGGCTTCCC TGTAGTCCAA AGAGATCCAT CTGCC
3 AGATCCTCCT AATCGAAT-- -----TCCC
*****

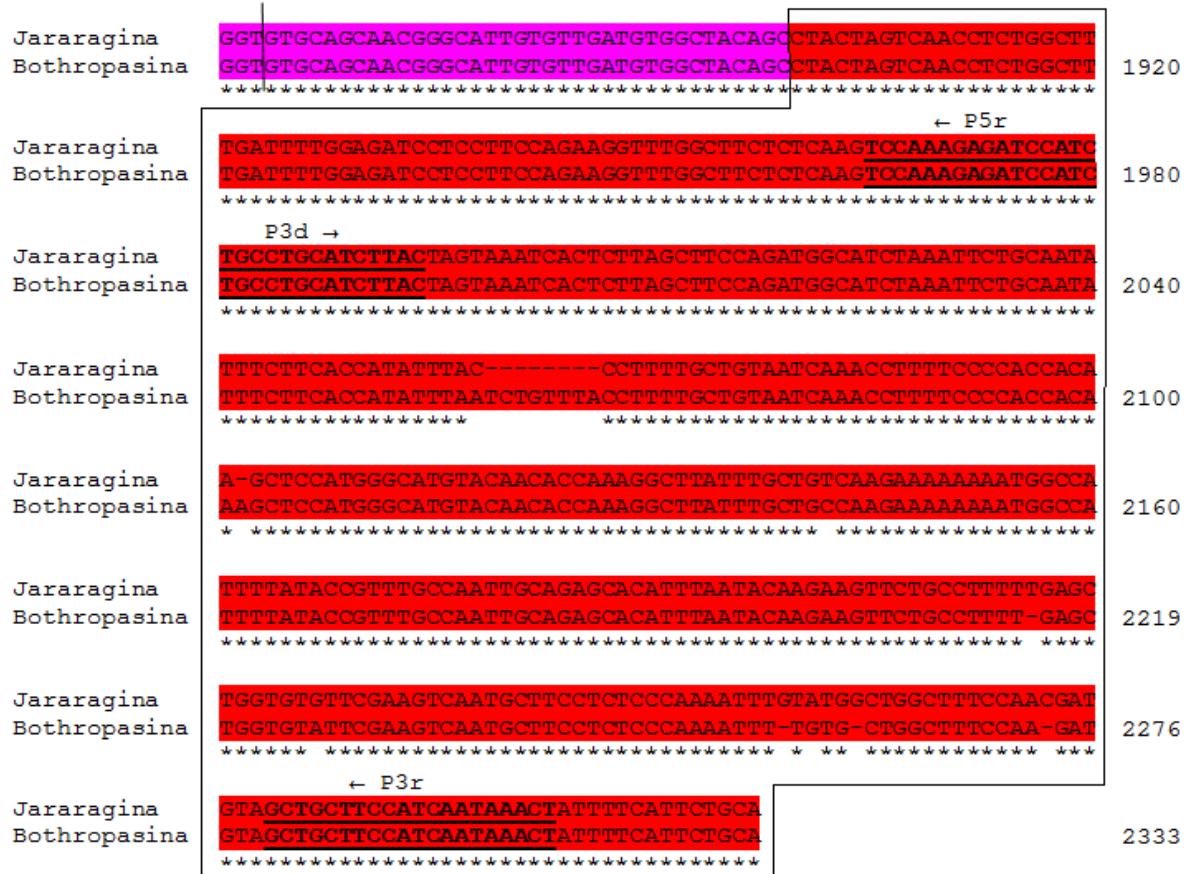
```

Figura 80. Sequenciamento P5d-P5r.

O resultado do sequenciamento indicou, como já observado anteriormente, a existência de pelo menos 2 genes diferentes. Nesse sequenciamento a sequência superior (S) foi denominada C1 e a inferior C3. Os clones col 21-1 e 21-2 (C2) mostraram em particular divergência na região 3' com relação aos demais consensos

5.6 Região 3'-não-traduzida

A região 3'-não-traduzida é mostrada no box vermelho compreendendo do nucleotídeo 1900 ao 2333, com base no cDNA da bothropasina. Os asteriscos mostram identidade de sequências; a falta deles indica divergência entre os cDNAs.



Para amplificar esta região foram utilizados os primers P3d-P3r, obtendo-se duas bandas como produto do PCR: a banda de ~300 pb foi nomeada inferior (I), e a banda de ~350 pb, superior (S) (Fig. 81).

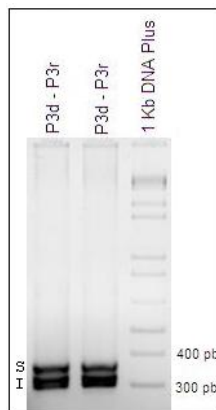


Figura 81. PCR P3d-P3r.

As bandas foram cortadas juntas, purificadas e clonadas, e os clones submetidos à miniprep. Foram analisados 11 clones, comparados com pUC 19 íntegro ou marcador (Fig. 82).

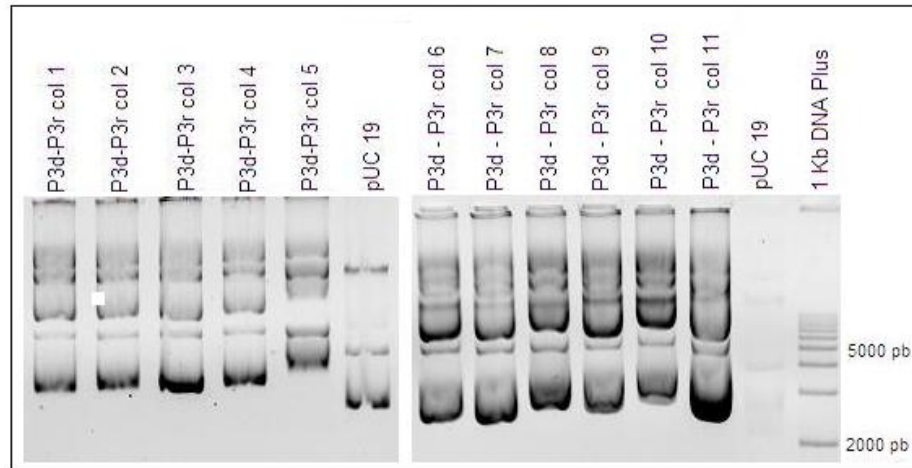


Figura 82. Minipreps P3d-P3r.

Para identificar os tamanhos dos insertos foram feitos PCRs com os *primers* SP6 e T7. Os clones P3d-P3r col 2, 5, 4, 8 e 11 apresentaram a banda inferior. Os clones P3d-P3r col 3, 6, 7 e 10 apresentaram a banda superior. O clone P3d-P3r col 9 apresentou uma banda muito fraca, não sendo possível definir o seu tamanho (Fig. 83).

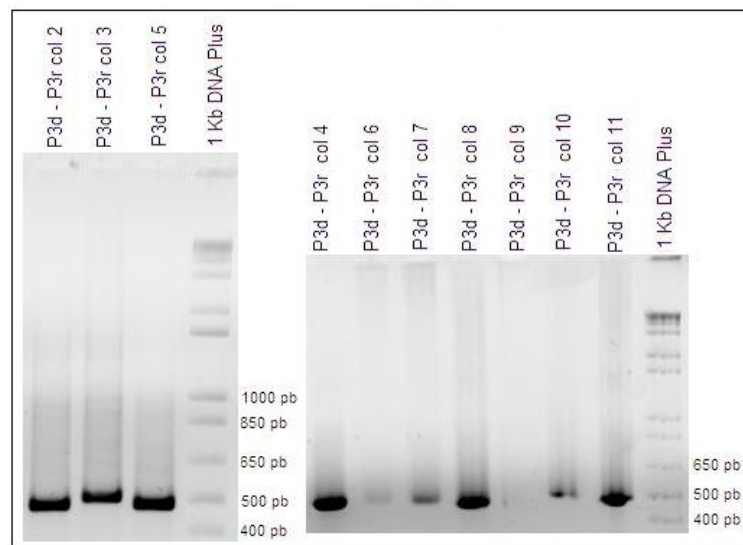


Figura 83. PCRs das Minipreps P3d-P3r.

Os 5 clones referentes à banda I e os 4 clones da banda S, foram sequenciados obtendo-se os consensos da banda inferior e da superior. A banda inferior (I) apresentou um tamanho de 297 pb e a superior (S) 320 pb (Fig. 84).

P3d →

```

B CCATCTGCCT GCATCTTACTT AGTAAATCAC TCTTAGCTTC CAGATGGCAT CTAAATTCTG
J CCATCTGCCT GCATCTTACTT AGTAAATCAC TCTTAGCTTC CAGATGGCAT CTAAATTCTG
S CCATCTGCCT GCATCTTACTT AGTAGATCAC TCTTAGCTTC CAGATGGCAT CCAAATTCCG
I CCATCTGCCT GCATCTTACTT AGTAAATCAC TCTTAGCTTC CATATGGAAT CTAAATTACG
*****
*****
*****
*****
*****
*****

B CAATATTTCT TCACCATATT TAATCTGTTT ACCTTTTGCT GTAATCAAAC CTTTTCCCA
J CAATATTTCT TCACCATATT TAC----- -CCTTTTGCT GTAATCAAAC CTTTTCCCA
S CAATGTTTCT TCTCCATATT TAATCTGTTT ACCTTTTGCT GTAATCAAAC CTTTTCCCG
I CAATATTTAT TCTCCATATT TAATCTGTTT ACCTTTTGCT GTAATCAAAC CTTTTCCCA
****
*****
**
*****
***
*****
*****
*****

B CCACAAAGCT CCATGGGCAT GTACAACACC AAAGGCTTAT TTGCTGCCAA GAAAAAAT
J CCACAA-GCT CCATGGGCAT GTACAACACC AAAGGCTTAT TTGCTGTCAA GAAAAAAT
S CCACAAAGCT CCATGGACAT GTACAACACC AAGAGCTTAT TTGCTGTCAA -AAAAAATAT
I CCATAAAGCT CCATAGGCAT GTACAACACC AAAGGCTT-- -----
***
*****
*****
*
***
*****
*****
*****

B GGCCATTTTA TACCGTTTGC CAATTGCAGA GCACATTTAA TACAAGAAGT TCTGCCTTTT
J GGCCATTTTA TACCGTTTGC CAATTGCAGA GCACATTTAA TACAAGAAGT TCTGCCTTTT
S GGCCATTTTA --CCGTTTGC CAATTGT-AA GCACATTTAA TGCAACAAAT TCTGGCTTTT
I ---TATTTTA --CCGTTTGC CAATTGCCAA GCACATTTAA TGCAACAAGT TCTGCCTTTT
*****
*****
*****
**
*****
*
***
**
*
*****
*****

B -GAGCTGGTG TATTCGAAGT CAATGCTTCC TCTCCCAAAA TTT-TGTG-C TGGCTTTCCA
J TGAGCTGGTG TGTTTGAAGT CAATGCTTCC TCTCCCAAAA TTTGTATGGC TGGCTTTCCA
S -GGTTGGTG TATTCGAAGT CGATGCTTCC TCTCCCAAAA TTC-CATG-C TGCCTTTCCA
I -GAGCTGGTG TATTCGAAGT CAATGCTTCC TCTCCCAAAA TTT-CACG-C TGGCTTTCCA
*
*
*****
*
***
***
*
*****
***
*****
**
*
*
**
*****

← P3r
B A-GATGTAGC TGCTTCCATC AATAAACT
J ACGATGTAGC TGCTTCCATC AATAAACT
S A-CATGTAGC TGCTTCCATC AATAAACT
I A-CATGTAGC TGCTTCCATC AATAAACT
*
*****
*****
*****

```

Figura 84. Sequenciamento P3d-P3r. As denominações S e I são provisórias, até a obtenção das sequências definitivas.

Para a visualização de todas as sequências obtidas, correspondentes aos domínios pré-definidos, foi elaborado um esquema com os introns e exons definidos até o momento. O esquema montado é provisório, pois só será possível definir os introns de cada toxina quando a sequência completa do genoma de *Bothrops jararaca* for publicada (Fig. 85).

A sequência superior (S) foi denominada de TOX1 e a sequência inferior (I) de TOX2. Partes das sequências já foram depositadas no Genbank e podem ser acessadas conforme os números indicados no ANEXO E.

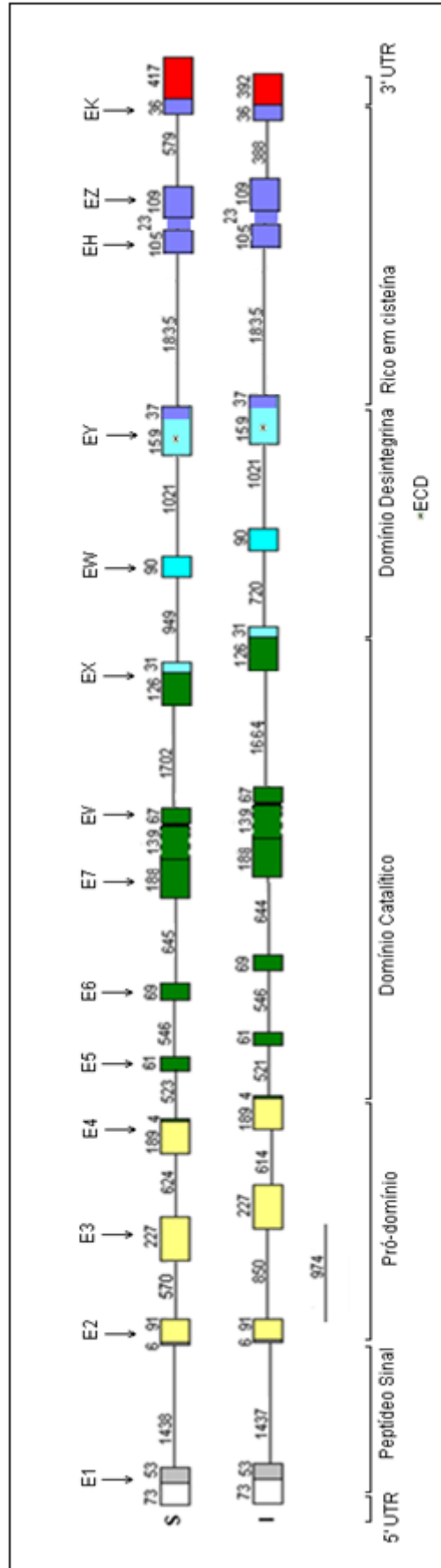


Figura 85. Esquema de exons e introns definida até o momento (suposição da ordem dos introns).

6 DISCUSSÃO

O estudo do DNA genômico da *Bothrops jararaca* é de suma importância para a Pesquisa Fundamental, fornecendo subsídios inéditos na literatura para o entendimento da evolução gênica das metaloproteinases ofídicas. Do ponto de vista médico, o conhecimento da atividade biológica de cada toxina presente no veneno, e seus domínios protéicos, permite pressupor o desenvolvimento de estratégias mais eficientes para o tratamento de casos de envenenamento, ou ainda viabilizá-las como agente terapêutico, como por exemplo, inibidor de metástases.

Algumas diferenças entre a jararagina e bothropasina estão distribuídas ao longo de seus cDNAs, sugerindo que essas proteínas não são isoformas, i.e., não resultam de *splicing* alternativo de um mesmo mRNA precursor, pois os cDNAs de ambas pareiam, não havendo perda de exons.

Até o presente momento foram identificados 14 exons (4 exons parciais) e 13 introns, com um total de 12535 pb no caso da TOX1 e 12268 pb para TOX2. Analisando a parte consenso entre as toxinas, no total encontram-se 322 posições diferentes entre elas, sendo 272 nos íntrons, e 41 na região 3' UTR. Na parte codificante da TOX 1 e TOX2 encontram-se 55 posições diferentes, levando a 9 mutações sinônimas e 22 não sinônimas. Foram encontrados ainda 10 aa iguais à sequência da bothropasina e 10 aa iguais à jararagina (ANEXO B). Esses resultados não indicam ter havido um processo de evolução acelerada. A diversificação das metaloproteinases com domínios desintegrina e rico em cisteínas (MDC) de mamíferos e de venenos de serpentes é decorrente de duplicações gênicas, seguidas pela divergência das cópias (Moura-da Silva et al., 1996). O mesmo processo muito provavelmente contribuiu para a evolução da jararagina e bothropasina, pertencentes à família das Reprolisinas (Fig. 1).

Este trabalho é o primeiro a identificar exons e introns de metaloproteinases, portanto não há como comparar a sequência das toxinas estudadas com a de outras toxinas.

6.1 Análise dos domínios proteicos / exons

6.1.1 Peptídeo Sinal

O peptídeo sinal é codificado pelo exon 1 e pela região 5' do exon 2, sendo conservado nas quatro sequências de toxinas comparadas, bothropasina, jararagina, TOX1 e TOX2.

6.1.2 Pró-domínio

O pró-domínio é codificado pelos exons 2, 3 e 4. O exon 2 da TOX1 e TOX2 tem a sequência igual à da bothropasina. O alinhamento de TOX1, TOX2, bothropasina, bothrostatina e HF3, também presentes no veneno de *Bothrops jararaca*, mostra total identidade de sequências, indicando que essas metaloproteinases têm módulos conservados, como previamente sugerido na Fig. 5 (ANEXO C).

A posição 315, no início do exon 3, tem G em TOX1, TOX2 e jararagina, e T na bothropasina. Há diferenças entre TOX1 e TOX2 em apenas duas posições (482, 501). Outras cinco posições diferem tanto da jararagina como da bothropasina, e outras ainda são iguais à jararagina ou bothropasina. Em consequência, os aminoácidos (aa) mudam, tendo sido identificadas 4 mutações sinônimas (aa 61, 97, 103 e 501) e 5 não-sinônimas (79, 90, 96, 98 e 108) (Quadro 3).

Quadro 3 - Alterações no exon 3. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição de nucleotídeos referente à sequência da bothropasina.

Exon 3 (227 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
315	G	T	G	G	61	V	V	V	V	Sinônima
367	A	A	G	G	79	K	K	E	E	Não-sinônima
383	T	C	T	T	84	T	I	I	I	aa = J
401	G	G	A	A	90	G	G	D	D	Não-sinônima
418	T	T	A	A	96	Y	Y	N	N	Não-sinônima
423	C	T	T	T	97	P	P	P	P	Sinônima
424-5	CC	GC	CT	CT	98	A	P	L	L	Não-sinônima
441	T	C	C	C	103	C	C	C	C	Sinônima
455	G	G	A	A	108	R	R	H	H	Não-sinônima
482	C	C	C	T	117	A	A	A	V	TOX2: não-sinônima
501	C	C	C	T	123	N	N	N	N	Sinônima

Nesta região foi identificado o motivo *cysteine switch*, conservado na TOX1, TOX2, jararagina e bothropasina (ANEXOS B e C).

Quanto ao exon 4, tanto na TOX1 como TOX2, foram identificadas uma mutação sinônima (aa 135), duas mutações iguais à jararagina (aa 144 e 145), e duas mutações iguais à bothropasina (aa 128 e 187). Nas posições 187 e 189 a TOX2 têm uma mutação não-sinônima, e a TOX1 é igual à jararagina (Quadro 4).

6.1.3 Domínio Catalítico

Quadro 4 - Alterações no exon 4. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à seqüência da bothropasina.

Exon 4 (203 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	J	TOX1	TOX2	Mutação
514	T	C	C	C	128	<i>H</i>	<i>Y</i>	<i>H</i>	<i>H</i>	aa = b
547	G	A	G	G	135	<i>T</i>	<i>T</i>	<i>T</i>	<i>T</i>	Sinônima
562	C	T	C	C	144	<i>S</i>	<i>P</i>	<i>P</i>	<i>P</i>	aa = j
565	G	A	G	G	145	<i>N</i>	<i>D</i>	<i>D</i>	<i>D</i>	aa = j
689	C	T	T	T	186	<i>V</i>	<i>A</i>	<i>V</i>	<i>V</i>	aa = b
691	T	G	G	C	187	<i>V</i>	<i>F</i>	<i>F</i>	<i>L</i>	TOX2: não-sinônima
697	G	G	G	C	189	<i>A</i>	<i>A</i>	<i>A</i>	<i>P</i>	TOX2: não-sinônima

O domínio catalítico é codificado pelos exons 4 (4 últimas bases), 5, 6, 7 e X. Faltam 227 pb para sequenciar, tendo como referência o cDNA. Devido a este *gap* ainda não foi possível numerar os exons e introns do segmento. Por isso os exons estão nomeados com letras maiúsculas.

O exon 5 apresenta 9 posições iguais à jararagina ou bothropasina, com alterações nos aminoácidos correspondentes. As posições 193, 197, 198 e 202 conferem com a jararagina; e 195, 200, 204 e 210, com a bothropasina (Quadro 5).

Quadro 5 - Alterações no exon 5. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à sequência da bothropasina.

Exon 5 (61 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
650	G	A	G	G	193	<i>K</i>	<i>R</i>	<i>R</i>	<i>R</i>	aa = j
655	G	A	A	A	195	<i>N</i>	<i>D</i>	<i>N</i>	<i>N</i>	aa = b
722	A	T	A	A	197	<i>F</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	aa = j
725	A	G	A	A	198	<i>R</i>	<i>K</i>	<i>K</i>	<i>K</i>	aa = j
730	A	G	G	G	200	<i>V</i>	<i>I</i>	<i>V</i>	<i>V</i>	aa = b
736	T	C	T	T	202	<i>L</i>	<i>F</i>	<i>F</i>	<i>F</i>	aa = j
742	A	G	A	A	204	<i>I</i>	<i>V</i>	<i>I</i>	<i>I</i>	aa = b
761-2	CC	TG	TG	TG	210	<i>M</i>	<i>T</i>	<i>M</i>	<i>M</i>	aa = b

O exon 6, constituído por 69 pb, é totalmente conservado em relação à jararagina e bothropasina.

O exon 7 foi sequenciado parcialmente. O aa da posição 236 na TOX1 e TOX2 são iguais à bothropasina (Quadro 6).

Quadro 6 - Alterações no exon 7. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à sequência da bothropasina.

Exon 7 (182 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
840	A	C	A	A	236	<i>L</i>	<i>F</i>	<i>L</i>	<i>L</i>	aa = b

A sequência parcial do exon V (67 pb) não apresentou nenhuma alteração em relação à jararagina e bothropasina.

O exon X codifica o final do domínio catalítico, mostrando 10 posições de nucleotídeos diferentes entre as sequências. Os aa 383, 396 e 412 da TOX2 são iguais à jararagina e bothropasina, mas na TOX1 as mutações são não-sinônimas. Os aa 384 e 385 são iguais à jararagina, e os aa 389 e 401 iguais à bothropasina, tanto na TOX1 como na TOX2. O aa 391 da TOX1 é igual à bothropasina e a TOX2 da jararagina, e o aa 419 é uma mutação sinônima, (este último aa já faz parte do domínio desintegrina) (Quadro 7).

Quadro 7 - Alterações no exon 7. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à sequência bothropasina. O sinal () representa o nucleotídeo conservado na posição 1286.

Exon X (156 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
1280	T	T	C	T	383	<i>M</i>	<i>M</i>	<i>T</i>	<i>M</i>	TOX1: não-sinônima
1284	T	G	T	T	384	<i>K</i>	<i>N</i>	<i>N</i>	<i>N</i>	aa = j
1285_7	C_G	G_C	C_C	C_C	385	<i>E</i>	<i>H</i>	<i>H</i>	<i>H</i>	aa = j
1294	C	G	C	C	388	<i>Q</i>	<i>E</i>	<i>Q</i>	<i>Q</i>	aa = b
1303	A	C	C	A	391	<i>L</i>	<i>I</i>	<i>L</i>	<i>I</i>	TOX1= b; TOX2 = j
1318	G	G	A	G	396	<i>G</i>	<i>G</i>	<i>R</i>	<i>G</i>	TOX1: não-sinônima
1330	A	G	G	G	400	<i>V</i>	<i>I</i>	<i>V</i>	<i>V</i>	aa = b
1306	G	G	A	G	412	<i>V</i>	<i>V</i>	<i>M</i>	<i>V</i>	TOX1: não-sinônima
1389	A	C	C	C	419	<i>G</i>	<i>G</i>	<i>G</i>	<i>G</i>	Sinônima

6.1.4 Domínio Desintegrina

O domínio desintegrina é codificado pela parte 3' do exon X (últimas 31 bases) e o exon W, de 90 pb, encontra-se conservado. O exon Y, onde se encontra o *motivo* ECD encontra-se conservado. No exon Y foi encontrada somente uma alteração de base, que não leva a alteração do aa (posição 477) (Quadro 8).

Quadro 8 - Alterações no exon Y. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posições referentes às sequências da bothropasina.

Exon Y (138 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
1563	T	C	C	C	477	<i>S</i>	<i>S</i>	<i>S</i>	<i>S</i>	Sinônima

6.1.5 Domínio Rico em Cisteína

A análise parcial do exon H (105 pb) mostra queo mesmo se encontra conservado em relação à jararagina e bothropasina.

O domínio rico em cisteína é codificado pelos exons Z e K. No exon Z foram identificadas 13 posições diferentes entre a TOX1 e TOX2 em relação à jararagina e bothropasina. Estas alterações levam a 8 mutações não-sinônimas (573, 474, 575, 578, 579, 580, 581 e 494), e uma sinônima (577) (Quadro 9).

Quadro 9 - Alterações no exon Y. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à sequência da bothropasina.

Exon Z (109 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
1791	G	G	T	T	573	<i>K</i>	<i>K</i>	<i>N</i>	<i>N</i>	Não-sinônima
1792-4	ATG	ATG	TAC	TAC	574	<i>M</i>	<i>M</i>	<i>Y</i>	<i>Y</i>	Não-sinônima
1796	T	T	A	A	575	<i>F</i>	<i>F</i>	<i>Y</i>	<i>Y</i>	Não-sinônima
1864	C	C	A	A	577	<i>S</i>	<i>S</i>	<i>S</i>	<i>S</i>	Sinônima
1865-6	AA	AA	TC	TC	578	<i>N</i>	<i>N</i>	<i>S</i>	<i>S</i>	Não-sinônima
1869-70	AT	AT	GG	GG	579	<i>D</i>	<i>D</i>	<i>R</i>	<i>R</i>	Não-sinônima
1871	G	G	C	C	580	<i>D</i>	<i>D</i>	<i>H</i>	<i>H</i>	Não-sinônima
1877	C	C	A	A	581	<i>H</i>	<i>H</i>	<i>N</i>	<i>N</i>	Não-sinônima
1914	T	T	G	G	594	<i>D</i>	<i>D</i>	<i>E</i>	<i>E</i>	Não-sinônima

Há uma região hiper-conservada (HCR) de 22 aa (Muniz et al. 2008) no exon Z, que no caso da TOX1 e TOX 2 apresenta 6 aa diferentes.

No exon K foram identificadas 4 posições alteradas levando a mutação-não sinônima (aa 601, 602 e 607) e uma mutação sinônima (aa 597) (Quadro 10).

Quadro 10 - Alterações no exon K. j (jararagina), b (bothropasina), TOX1 (sequência superior), TOX 2 (sequência inferior). Em itálico estão os aa (ANEXO D). *Posição referente à sequência da bothropasina.

Exon K (138 pb)										
Posição nt*	j	b	TOX1	TOX2	Posição aa*	b	j	TOX1	TOX2	Mutação
1923	G	G	C	C	597	<i>V</i>	<i>V</i>	<i>V</i>	<i>V</i>	Sinônima
1933	G	G	A	A	601	<i>G</i>	<i>G</i>	<i>R</i>	<i>R</i>	Não-sinônima
1938	T	T	G	G	602	<i>H</i>	<i>H</i>	<i>Q</i>	<i>Q</i>	Não-sinônima
1951	G	G	A	A	607	<i>A</i>	<i>A</i>	<i>T</i>	<i>T</i>	Não-sinônima

6.1.6 Região 3' UTR

A região 3' UTR da TOX1 tem 417 pb e na TOX2, 392 pb. Em comparação com a jararagina que têm 350 pb, e a bothropasina com 354 pb, esta região encontra-se bastante alterada, tanto em relação ao número como à sequência de bases (Quadro 11). Esta região não codifica aminoácidos.

Diversos genes de serino proteases de *Trimeresurus flavoviridis* apresentam os introns e as regiões UTRs altamente conservadas. Já as regiões que codificam a proteína madura têm alta taxa de alterações, sugerindo que estas toxinas sofreram evolução acelerada (Deshimaru et al., 1996). Diferentemente, neste trabalho as regiões que têm maior taxa de alteração são os introns e as regiões UTR.

Quadro 11 - Alterações 3'UTR. Foram identificadas 41 alterações entre as sequências e uma inserção de 10 pb, comparada com a jararagina e a bothropasina. *Posição referente à sequência da bothropasina.

3' URT					
Posição nt*	j	b	TOX1	TOX2	Mudança
1965	G	G	A	A	Alterações nas TOX 1 e TOX2
Entre 77-78	-	-	TTCTCTCAGA	TTCTCTCAGA	Inserção de 10 pb
2006	G	G	T	T	Alterações nas TOX1 e TOX2
2017	T	T	C	C	Alterações nas TOX1 e TOX2
2020-1	CA	CA	GT	GT	Alterações nas TOX1 e TOX2
2059	A	A	G	A	Alteração na TOX1
2077	G	G	G	T	Alteração na TOX2
2082	C	C	C	A	Alteração na TOX2
2086	T	T	C	T	Alteração na TOX1
2092	C	C	C	A	Alteração na TOX2
2093	T	T	G	G	Alterações nas TOX1 E TOX2
2099	A	A	G	A	Alteração na TOX1
2103	C	C	C	A	Alteração na TOX2
2107	A	A	T	T	Alterações nas TOX1 e TOX2
2118 – 25	-	TCTGTTTA	TCTGTTTA	TCTGTTTA	TOX1 e TOX2 = b
2154	A	A	G	A	Alteração na TOX1
2159	C	C	C	T	Alteração na TOX2
2162	-	A	A	A	Alteração na TOX2
2169	G	G	G	A	Alteração na TOX2
2171	G	G	A	G	Alteração na TOX1
2187	A	A	G	A	Alteração na TOX1

2200	T	C	T	-	TOX1 =j
2217	C	C	C	T	Alteração na TOX2
2224-5	TA	TA	-	-	Alterações nas TOX1 e TOX2
2240	C	C	T	C	Alteração na TOX1
2242	G	G	-	-	Alterações nas TOX1 e TOX2
2255	A	A	G	G	Alterações nas TOX1 e TOX2
2259	G	G	C	C	Alterações nas TOX1 e TOX2
2262	G	G	A	G	Alteração na TOX1
2268	C	C	G	C	Alteração na TOX1
2271	T	-	-	-	TOX1 = b
2273	A	A	G	A	Alteração na TOX1
2275	C	C	T	C	Alteração na TOX1
2282	G	A	A	A	TOX1 =j
2286	G	G	G	A	Alteração na TOX2
2292	A	A	G	A	Alteração na TOX1
2304	C	C	T	C	Alteração na TOX1
2310	T	T	C	T	Alteração na TOX1
2311	G	-	-	-	TOX1 = b
2115	T	T	C	C	Alterações nas TOX1 e TOX2
2316	G	-	-	-	TOX1 = b
2120	G	G	C	G	Alteração na TOX1

A análise dos aa pelo programa *ExpPASy Proteomics Server* (<http://expasy.org>) mostrou que a jararagina, bothropasina, TOX1 e TOX2, são hidrofóbicas, mas há uma pequena variação: jararagina -0,570, bothropasina -0,451, TOX1 -0,493 e a TOX2 -0,477. Esta pequena variação no grau de hidrofobicidade pode alterar as características físico-químicas das proteínas. No caso dos dois *gaps* que ainda não foram sequenciados, foi utilizada a sequência de aa da bothropasina. A 1º região não seqüenciada, tem uma diferença de 4 aa entre a jararagina e a bothropasina. A 2º região não seqüenciada não tem diferenças.

Todas as cisteínas foram mantidas, o que favorece a manutenção da estrutura secundária.

6.2 Análise dos introns

O intron 1 da TOX1 tem 1438 pb e TOX2 1437 pb, encontrando-se 21 posições diferentes ao se estabelecer a sequência consenso.

Quanto ao intron 2, foram identificadas 3 sequências distintas após a determinação dos consensos. As bandas 1, 2 e 3 apresentaram 570 pb, 850 pb e

974 pb, respectivamente. Há três seqüências de introns disponíveis, não sendo ainda possível afirmar se pertencem à TOX1, TOX2 ou ainda a outra metaloproteinase. Entretanto há alta identidade de seqüência na região consenso (ANEXO B).

O intron 3 de TOX1 e TOX2 tem 624 pb e 614 pb, respectivamente. Além da diferença de tamanho entre as seqüências, há 20 posições diferentes entre elas.

O intron 4 de TOX1 e TOX2 tem 523 pb e 521 pb, respectivamente. Além disso, são encontradas 24 posições diferentes entre elas.

O intron 5 de TOX1 e TOX2 tem 546 pb, ambos com a mesma seqüência de bases.

O intron 6 de TOX1 e TOX2 tem 645 pb e 644 pb, respectivamente. Foram encontradas 7 posições alteradas, inclusive um microsátélite com a repetição (TTA)₁₇ e um sítio *EcoRI* (GAATTC).

O intron v da TOX1 e TOX2 tem 1718 e 1702 pb, respectivamente. Foram encontradas 105 posições alteradas entre elas.

O Intron x de TOX1 e TOX2 tem 949 pb e 720 pb, respectivamente. Neste intron encontra-se a maior diferença em relação ao tamanho das seqüências, e 6 posições diferentes na seqüência consenso.

O intron w tem 1021 pb e intron y tem 1835 pb em ambas as toxinas.

O intron z de TOX1 e TOX2 tem 579 pb, com 5 posições diferentes na seqüência consenso (Quadro 12).

Quadro 12 - Alterações encontradas nos introns. Foram identificadas um total de 251 posições alteradas na comparação entre TOX1 e TOX2.

Intron	1	2	3	4	5	6	v	x	w	y	z
TOX1	1438	974	624	523	546	645	1718	949	1021	1835	579
TOX2	1438	850	614	521	546	644	1702	720	1021	1835	388
TOX1/TOX2	21	63	20	24	-	7	105	6	-	-	5

7 CONCLUSÃO

Os dados deste trabalho são indicativos da existência de no mínimo dois genes diferentes oriundos de duplicação gênica que codificam as toxinas jararagina e bothropasina. Ainda não foi possível relacionar diretamente qual deles codifica qual toxina devido à presença de mutações não-sinônimas. Devido a isto, foram nomeadas provisoriamente como TOX1 e TOX2, até a obtenção das sequências definitivas.

Devido ao fato dos cDNAs de jararagina e bothropasina terem sido obtidos a partir de *pool* de espécimens, não está descartada a hipótese de polimorfismo populacional, não identificado nas sequências dos clones de cDNA previamente publicados. O sequenciamento do genoma de *Bothrops jararaca* deverá esclarecer devidamente o número real de genes codificadores de metaloproteinasas nessa espécie de Ofídio. A semelhança de sequência entre diferentes metaloproteinasas já sugere que elas são formadas por módulos relativamente conservados.

Referências

Akiyama SK, Olden K, Yamada KM. Fibronectina and integrins in invasion and metastasis. *Cancer Metast Rev.* 1995;14(3):173-89.

Assakura MT. Clonagem molecular, análise da seqüência e expressão da bothropasina, uma metaloendopeptidase isolada do veneno da *Bothrops jararaca*. [tese (Doutorado)]. São Paulo: Escola Paulista de Medicina Universidade Federal de São Paulo; 2000.

Assakura MT, Silva CA, Mentele R, Camargo AC, Serrano SM. Molecular cloning and expression of structural domains of bothropasin, a P III metalloproteinase from the venom of *Bothrops jararaca*. *Toxicon.* 2003;41(2):217-27.

Baramova EN, Shannon JD, Bjarnason JB, Fox JW. Identification of the cleavage sites by hemorrhagic metalloproteinases in tipe IV collagen. *Matrix.* 1990;10(2):91-7.

Barczyk M, Carracedo S, Gulberg D. Integrins. *Cel Tissue Res.* 2010;339(1):269-80.

Barraviera B. Venenos animais uma visão integrada. Rio de Janeiro: EPUB; 1994. Capítulo 63, p. 97-105.

Bjarnason JB, Fox JW. Snake venom metalloendopeptidases: reprotolysins. *Methods Enzymol.* 1995;248:345-68.

Blobel CP. ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol.* 2005;6(1):32-43.

Botos I, Scapozza JD, Fox JW, Meyer EF. Strcture based analysis of inhibitor binding to HT-d. *Acta Cryst.* 1995;D51:597-604.

Calvete JJ, Marcinkiewicz C, Monleón D, Esteves V, Celda B, Juárez P. Snake venom desintegrin: evolution of structure and function. *Toxicon.* 2005;45(8):1063-74.

Calvete JJ, Juárez P, Sanz L. Snake venomomics. Strategy and applications. *J Mass Spectrom.* 2007;42(11):1405-14.

Chang L-S, Chung C, Liou J-C, Chang C-W, Yang C-C. Novel neurotoxins from Taiwan banded krait (*Bungarus multicinctus*) venom: purification, characterization and gene organization. *Toxicon*. 2003;42(3):323-30.

Chiang HS, Swaim MW, Huang TF. Characterization of platelet aggregation induced by human colon adenocarcinoma cells and its inhibition by snake venom peptides, trigramin and rhodostomin. *Br. J. Haematol*. 1994;87(2):325-31.

Cidade DAP, Simão TA, Dávila AMR, Wagner G, Junqueira-de-Azevedo ILM, Ho PL, Bom C, Zingali RB, Albano RM. *Bothrops jararaca* venom gland transcriptome: Analyse of the gene expression pattern. *Toxicon*. 2006;48:437-61.

Cominetti MR, Terruggi CH, Ramos OH, Fox JW, Mariano-Oliveira A, De Freitas MS, Figueiredo CC, Morandi V, Seliestre-de-Araujo HS. Alternagin-C, a disintegrin-like protein, induces vascular endothelial cell growth factor (VEGF) expression and endothelial cell proliferation *in vitro*. *J Biol Chem*. 2004;279(18):18247-55.

Coppolino MG, Dedhar S. Bi-directional signal transduction by integrin receptors. *Int. J Biochem. Cell Biol*. 2000;32(2):171-88.

Corrêa MC Jr, Maria DA, Moura-da-Silva AM, Pizzocaro KF e Ruiz IR. Inhibition of melanoma cells tumorigenicity by the snake venom toxin jararhagin. *Toxicon*. 2002;40(6):739-48.

Dai L, Zimmerty S. ORF-less reverse-transcriptase-encoding group II introns in archaeobacteria, with a pattern of homing into related group II intron ORF. *RNA*. 2003;9(1):14-9.

Daltry JC, Wüster W, Thorpe RS. Diet and snake venom evolution. *Nature*. 1996;379(6565):537-40.

Danen EH, Marcinkiewicz C, Cornelissen IM, van Kraats AA, Pachter JA, Ruiten DJ, Niewiarowski S, van Muijen GN. The disintegrin eristostatin interferes with integrin $\alpha\beta 1$ function and experimental metastasis of human melanoma cells. *Exp Cell Res*. 1998;238(1):188-96.

Deshimaru M, Ogawa T, Nakashima K, Nobuhisa I, Chijiwa T, Shimohigashi Y, Fukumaki Y, Niwa M, Yamashina I, Hattori S, Ohno M. Accelerated evolution of crotalinae snake venom gland serine proteases. *FEBS Lett*. 1996;397(1):83-8.

Jia Y, Perez JC. Molecular cloning and characterization of cDNAs encoding metalloproteinase from snake venom gland. *Toxicon*. 2010;5(2-3):462-9.

Fox JW, Serrano SM. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprotolysin family of metalloproteinases. *Toxicon*. 2005;45(8):969-85.

Fox JW, Serrano SM. Insights into and speculation about snake venom metalloproteinases (SVMP) synthesis, folding and disulfide bond formation and their contribution to venom complexity. *FEBS J*. 2008;275:3016-30.

Fox JJ, Serrano SMT. Timeline of key events in snake venom metalloproteinase research. *Journal of Proteomics*. 2009(72):200-9.

Fry BG, Lumsden NG, Wuster W, Wickramaratna JC, Hodgson WC, Kini RM. Isolation of neurotoxin (alpha-colubritoxin) from a nonvenomous colubrid: evidence for early origin of venomous snakes. *J Mol Evol*. 2003 Oct;57(4):446-52.

Fry BG, Wuster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol*. 2004 May;21(5):870-83.

Fujimi TJ, Nakajyo T, Nishimura E, Ogura E, Tsuchiya T, Tamiya T. Molecular evolution and diversification of snake toxin genes, revealed by analysis of intron sequences. *Gene*. 2003;313:111-8.

Fukagawa T, Nose T, Shimohigashi Y, Ogawa T, Oda N, Nakashima K, Chang CC, Ohno M. Purification, sequencing and characterization of single amino acid-substituted phospholipase A2 isozymes from *Trimeresurus graminus* (green habu snake) venom. *Toxicon*. 1993;31(8):957-67.

Georgakakou E, Nastopoulos V, Eleftheriou S, Zarkadis IK. A toxin-like gene in rainbow trout: cloning, expression, and gene organization. *Toxicon*. 2007;49(7):1002-9.

Hannon GJ. RNA interference. *Nature*. 2002;418(6894):244-21.

Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. *Trends in Genet*. 2005;21(2):111-9.

Hite LA, Jia LG, Bjarnason JB, Fox JW. cDNA sequences for four snake venom metalloproteinases: structure, classification, and their relationship to mammalian reproductive proteins. *Arch Biochem Biophys*. 1994;308(1):182-91.

Hooper NM. Families of zinc metalloproteases. *FEBS Lett*. 1994;354(1):1-6.

Jia LG, Wang XM, Shannon JD, Bjarnason JB, Fox JW. Inhibition of platelet aggregation by the recombinant cysteine-rich domain of the hemorrhagic snake venom metalloproteinase, atrolysin A. *Arch Biochem Biophys*. 2000;373(1):281-326.

Jia Y, Perez JC. Molecular cloning and characterization of cDNAs encoding metalloproteinase from snake venom gland. *Toxicon*. 2010;5(2-3):462-9.

Juárez P, Comas I, González-Candelas F, Calvete JJ. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol Biol Evol*. 2008;25(11): 2391-407.

Junqueira-de-Azevedo ICM, Ching ATC, Carvalho E, Faria F, Nishiyama JR, Yataku M, Ho, PL, Diniz MRV. *Lachesis muta* (Viperidae) cDNA reveal diverging pitviper molecules and scaffolds typical of snake (Elapidae) venom: implication in snake toxin repertoire evolution. *Genetic (Austin), EUA*. 2006;173(2):877-89.

Kamigute AS, Sano Martins IS. South American snake venoms affecting homeostasis. *J Toxicol Toxin Rev*. 1995;14:359-74.

Kamiguti AS, Zuzel M, Theakston RD. Snake venom metalloproteinases and disintegrins: interactions with cells. *Braz J Med Biol Res*. 1998;31(7): 853-62.

Kang IC, Kim DS, Jang Y, Chung KH. Suppressive mechanism of salmosin, a novel desintegrina in B16 melanoma cell metastasis. *Biochem Biophys Res Commun*. 2000;275(1):169-73.

Kini RM, Evans HJ. Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. *Toxicon*. 1992;30(3):265-93.

Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*. 2005;39:309-38.

Kordis D, Gubensek F. Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 gene and in genomes of Viperidae snake. *Eur J Biochem.* 1997 Jun 15;246(3):772-9.

Kuphal S, Bauer R, Bosserhoff AK. Integrin signaling malignant melanoma. *Cancer Metastasis Rev.* 2005;24(2):195-222.

Lewin B. Genes VII. New York: Oxford University Press; 2000. *Nuclear splicing*; p. 685-718.

Lin SL, Ying SY. Gene silencing *in vitro* and *in vivo* using intronic microRNAs. *Methods Mol Biol.* 2006;342:295-312.

Mandelbaum FR, Reichel AP, Assakura MT. Isolation and characterization of a proteolytic enzyme from the venom of the snake *Bothrops jararaca* (Jararaca). *Toxicon.* 1982;20(6):955-72.

Marcinkiewicz C, Weinreb PH, Calvete JJ, Kisiel DG, Mousa SA, Tuszynski GP, Lobb RR. Obtustatin: a potent selective inhibitor of alpha1beta1 integrin *in vitro* and angiogenesis *in vivo*. *Cancer Res.* 2003;63(9):2020-3.

Markland FS. Snake venoms and the hemostatic system. *Toxicon.* 1998;36(12):1749-1800.

Marques-Porto R, Lebrum I, Pimenta DC,. Self-proteolysis regulation in *Bothrops jararaca* venom: the metallopeptidases and their intrinsic peptidic inhibitor. *Comp Biochem Physiol C Toxicol Pharmacol.* 2008;147:424-33.

Maria DA, Vassão RC, Ruiz IR. Haematopoietic effects induced in mice by the snake venom toxin jararhagin. *Toxicon.* 2003;42(6): 579-85.

Mattick JS. Introns: evolution and function. *Curr Opin Genet Dev.* 1994;4(6):823-31.

Moura-da-Silva AM, Desmond H, Laing GD, Theakston RD. Isolation and comparison of myotoxins isolated from venoms of different species of *Bothrops* snake. *Toxicon.* 1991;29:713-23.

Moura-da-Silva AM, Theakston RD, Crampton JM. Evolution of disintegrin cysteine-rich and mammalian matrix-degrading metalloproteinases: gene duplication and

divergence of a common ancestor rather than convergent evolution. *J Mol Evol.* 1996;43(3):263-9.

Moura-da-Silva AM, Línica A, Della-Casa MS, Kamiguti AS, Ho PL, Crampton JM, Theakston RDG. Jararhagin ECD-containing disintegrin domain: expression in *Escherichia coli* and inhibition of the platelet-collagen interaction. *Archives of Biochemistry and Biophysics.* 1999;369:295-301.

Moura-da-Silva AM, Marcinkiewicz C, Marcinkiewicz M, Niewiarowski S. Selective recognition of a integrin $\alpha_2\beta_1$ by jararhagin, a metalloproteinase/disintegrin from *Bothrops jararaca* venom. *Thromb Res.* 2001;102(2):153-9.

Moura-da-Silva AM, Della-Casa MS, David AS, Assakura MT, Butera D, Lebrum I, Shannon JD, Serrano SMT, Fox JW. Evidence for heterogeneous form of the snake venom metalloproteinase jararhagin: a factor contributing to snake venom variability. *Archives of Biochemistry and Biophysics.* 2003;409:395-401.

Muniz JR. Estrutura Tridimensional da Bothropasina uma metaloproteinase/Desintegrina do veneno de *Bothrops jararaca*. [tese (Doutorado)]. São Paulo: Instituto de Física de São Carlos, Universidade de São Paulo, 2007.

Muniz JR, Ambrosio AL, Selistre-de-Araújo HS, Cominetti MR, Moura-da-Silva AM, Oliva G, Garrett RC, Souza DH. The three-dimensional structure of bothropasin, the main hemorrhagic factor from *Bothrops jararaca* venom: insights for a new classification of snake venom metalloprotease subgroups. *Toxicon.* 2008;52(7):807-16.

Nott A, Meislin SH, Moore MJ. A quantitative analyse of intron effects on mammalian gene expression. *RNA.* 2003;9(5):607-17.

Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M. Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. *Toxicon Review.* 2005;45(1):1-14.

Ohno M, Chijiwa T, Oda-Ueda N, Ogawa T, Hattori S. Molecular evolution of myotoxic phospholipases A2 from snake venom. *Toxicon.* 2003;42(8):841-54.

Orning L, Krivi G, Bild G, Gierse J, Aykent S, Fitzpatrick FA. Inhibition of leukotriene A4 hydrolase/aminopeptidase by captopril. *J Biol Chem.* 1991;266(25):16507-11.

Paine MJ, Desmond HP, Theakston RD, Crampton JM. Purification, cloning, and molecular characterization of a high molecular weight hemorrhagic metalloproteinase, jararhagin, from *Bothrops jararaca* venom. Insights into the disintegrin gene family. *J Biol Chem*. 1992;267(32):22869-76.

Queiroz LS, Petta CA. Histopathological changes caused by venom of urutu snake (*Bothrops alternatus*) in mouse skeletal muscle. *Rev Inst Med Trop São Paulo*. 1984;26:247-53.

Rádis-Baptista G, Kubo T, Oguiura N, Svartman M, Almeida TM, Batistic RF, Oliveira EB, Vianna-Morgante AM, Yamane T. Structure and chromosomal localization of the gene for crotamine, a toxin from the South American rattlesnake *Crotalus durissus terrificus*. *Toxicon*. 2003;42(7):747-52.

Reza MA, Swaruo S, Kine RM. Structure of two genes encoding parallel prothrombin activators in *Tropidechis carinatus* snake: gene duplication and recruitment of factor X gene to the venom gland. *J Thromb Haemost*. 2007;5(1):117-26.

Richman A. Evolution of balanced genetic polymorphism. *Mol Ecol*. 2000;9(12):1953-63.

Ribeiro GRH, Instabilidade do genoma, mutação/expressão gênica e tumorigênica de linhagens de carcinoma de cólon humano antes e após tratamento com toxina ofídica. [dissertação (Mestrado)]. São Paulo: Universidade Estadual Paulista "Julio de Mesquita Filho"; 2003.

Rogers JH. How were introns inserted into nuclear genes? *Trends Genet*. 1989;5:213-16.

Russel FE, Snyder CCMD. Snake Venom Poisoning. *Plastic and Reconstructive Surgery*. 1981;68(6):974.

Serrano SM, Wang D, Shannon JD, Pinto AF, Polanowska-Grabowska RK, Fox JW. Interaction of the cysteine-rich domain of snake venom metalloproteinase with the A1 domain of Von Willebrand factor promoter site-specific proteolysis of Von Willebrand factor and inhibition of Von Willebrand factor mediated platelet aggregation. *FEBS J*. 2007;274:3611-21.

Sheu JR, Yen MH, Kan YC, Hung WC, Chang PT, Luk HN. Inhibition of angiogenesis *in vitro* and *in vivo*: comparison of the relative activities of triflavin, an Arg-Gly-Asp

containing peptide and anti-alpha (v) beta 3 integrin monoclonal antibody. *Biochim Biophys Acta*. 1997;1336(3):445-54.

Sheth N Roca X, Hastings ML, Roeder T, Krainer AR. Comprehensive splice-site analysis using comparative genomics. *Nuclei Acid Res*. 2006;34(4):3955-67.

Souza DH, Iemma MR, Ferreira LL, Faria JP, Oliva ML, Zingali RB, Niewiarowski S, Selistre-de-Araujo HS. The disintegrin-like domain of the snake venom metalloprotease alternagin inhibits alpha2beta1 integrin-mediated cell adhesion. *Arch Biochem Biophys*. 2000;384(2):341-50.

Stone AL, Kroeger M, Sang QX. Structure-function analysis of the ADAM family of desintegrin-like and metalloproteinase-containing proteins. *J Protein Chem*. 1999;18(4):447-64.

Tanizaki MM, Zingali RB, Kawazaki H, Imajoh S, Yamazaki S, Suzuki K, Purification and some characteristics of a zinc metalloproteinases from the venom of *Bothrops jararaca* (jararaca). *Toxicon*. 1989;27(7):747-55.

Tani A, Ogawa T, Nose T, Nikandrov NN, Deshimaru M, Chijiwa T, Chang CC, Fukumaki Y, Ohno M. Characterization, primary structure and molecular evolution of anticoagulant protein from *Agkistrodon actus* venom. *Toxicon*. 2002;40(6):803–13.

Terada S, Hori J, Fujimura S, Kimoto E. Purification and amino acid sequence of brivilysin L6, a non-hemorrhagic metalloprotease from *Agkistrodon halys brevicaudus* venom. *J Biochem*. 1999;125(1):64-9.

Trikha M, De Clerck YA, Markland FS. Contortrostatin, a snake venom disintegrin, inhibits beta 1 integrin-mediated human metastatic melanoma cell adhesion and block experiment metastatic. *Cancer Res*. 1994;54(18):4993-8.

Tu AT. *Venoms: Chemistry and Molecular Biology*. New York: John Willyand Sons. *Snake venoms: general background and composition*. 1988; p.1-9.

Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. *Immunol Rev*. 1998;163:19–34.

Wolfsberg TG, Primakoff P, Myles DG, White JM. ADAM, a novel family of membrane proteins containing A Desintegrin And Metalloprotease domain:

multipotential functions in cell-cell and cell matrix interaction. *J Cell Biol.* 1995;131(2):275-8.

Yamamoto S, Higuchi Y, Yoshiyama K, Shimizu E, Kataoka M, Hijiya N, Matsuura K. ADAM family proteins in the immune system. *Immunol Today.* 1999;20(6):278-84.

Yeh RH, Yan X, Cammer M, Bresnick AR, Lawrence DS. Real time visualization of protein kinase activity in living cells. *J Biol Chem.* 2002;277(13):11527-32.

APÊNDICE - Reagentes

Ampicilina

estoque: 50µg/µl

uso: 50µg/ml

Azul de Bromofenol

Xilenocianol 0,25 % - 0,0125 g (azul claro)

Azul de Bromofenol 0,25 % - 0,0250 g (azul escuro)

Glicerol 50 % - 2,5 ml

EDTA 10 mM - 100 µl do EDTA 0,5 M

Completar com 5,0 ml de água bidestilada. Armazenar na geladeira.

Brometo de Etídeo

1 g de brometo de etídeo para 100 ml de água.

Dissolver com peixinho. Usar luvas, máscara e forrar a balança. Armazenar em frasco âmbar, na geladeira 4° C. (MANIATIS, pag B11).

CaCl₂ 50 mM

PM: 7,351

Pesar 0,7351 g e dissolver em 100 ml de água Milli-Q. Armazenar na geladeira.

EDTA 2 M

pH 8,0

EDTA – 14,88 g

Dissolver em 150 ml de água bidestilada, acertar o pH com NaOH e completar o volume para 200 ml com água bidestilada. Armazenar a 4° C.

EDTA 50 mM

pH 8,0

EDTA – 3,72 g Na₂H₂(C₁₀H₁₂N₂O₈)H₂O (PM 372,24 g/L)

Dissolver em 150 ml de água bidestilada, acertar o pH com NaOH e completar o volume para 200 ml de água bidestilada. Armazenar a 4° C.

LB

Bacto triptona – 10 g

Extrato de levedura (bacto yeast extrate) – 5 g

NaCl – 5 g

NaOH – 1 pastilha

Volume final: 1 litro de água bidestilada. Autoclavar 20 minutos, pH 7,5.

Lisozima

Solução estoque: 50 mg/ml

Solução utilizada na miniprep: 10 mg/ml

20 µl (solução 50 mg/ml) + 80 µl de água milli-Q = 100 µl (10 mg/ml).

Placas de Agar

15 g de Agar para 1 litro de LB

Autoclavar 20 minutos. Esperar esfriar para colocar a ampicilina (somente nas placas onde serão semeados os plasmídeos recombinantes). Colocar em cada placa 25 ml.

Ampicilina 50 µg / ml.

Save Money

Tris 200 mM pH 9

MgCl₂ 5 mM

STAB

Difco nutrient broth* – 10 g

NaCl – 5 g

Difco agar – 6 g

O Difco nutrient broth pode ser substituído por:

Bacto beef (yeast) extract – 3 g

Bacto peptona – 5 g

Volume final: 1 litro de água bidestilada. Autoclavar 20 minutos. Colocar 5 ml em cada tubo (estes também são autoclavados).

TE

Tris HCl – 10 mM (250 μ L de Tris HCl 2M)

EDTA – 1 mM (100 μ L de EDTA 0,5 M)

Volume final – 50 ml (utilizar água Milli-Q)

TEB 1x

Tris 0,9 M – 109 g

EDTA 0,02M – 40 ml a 0,5 M

Ac. Bórico – 55 g

Pesar separadamente, colocar em um Becker grande 800 ml de água bidestilada, deixar agitando com um peixinho. Completar para 1 litro.

TEST (miniprep)

Triton x 2,5 % (25 mL)

Sucrose 4 % (4 g)

Tris HCl 25 mM, pH 8,0 (1,25 mL de Tris HCl 2M)

EDTA 25 mM (5,0 mL de EDTA 0,5 M)

Completar com 100 ml de água Milli-Q.

ANEXO A – Alinhamento dos cDNAs da Jararagina e Bothropasina

Alinhamento dos cDNAs da jararagina e bothropasina usando o programa Clustal W

Seqüência 1: Jararagina (J) 2118 pb (X68251)

Seqüência 2: Bothropasina (B) 2333 pb (AFO56025)

As linhas verticais indicam a localização dos íntrons já identificados. Os domínios estão representados por diferentes cores: cinza (peptídeo sinal), amarelo (pró-domínio), verde (domínio catalítico), azul (domínio desintegrina), rosa (domínio rico em cisteína) e vermelho (região 3' UTR). Os *primers* utilizados (negrito) estão indicados por setas (direto e reverso). As regiões sem cor na região do domínio catalítico e rico em cisteína estão em fase de sequenciamento

```

                Pbd →
J -----
B GCACGAGAGAACTCAGATTGGCTTGAAGGAAGGAAGAGATTGCCTGACTTCCAGCCAAA 60
                Exon 1
J -----
B TCCAGCCTCCAAAATGATTGAAGTTCTCTTGGTGACTATATGCTTAGCAGCTTTTCCTTA 120

                Exon2          ← Pbr      PbirD →
J -----
B TCAAGGCAGCTCTATAATCCTTGGAATCTGGGAACGTGAATGATTATGAAGTAGTGTATCC 180

                Pad →          Exon 3
J -----GCAACGAGGCCCAAAGGAGCAGTTCAGCCAAAGTATGAAGACGCCATGCA 50
B ACGAAAAGTCACTGCATTGCCCAAAGGAGCAGTTCAGCCAAAGTATGAAGACGCCATGCA 240
          * *  *****

                ← Pair
J ATATGAATTTAAAGTGAATGGAGAGCCAGTGGTTCCTTCACCTGGAAAAAATAAAGGACT 110
B ATATGAATTTAAAGTGAATGGAGAGCCAGTGGTTCCTTCACCTGGAAAAAATAAAGGACT 300
          *****

                Paid →
J TTTTTCAAAAAGATTACAGCGAGATTCATTATTCGCCCTGATGGCAGAGAAATTACAACATA 170
B TTTTTCAAAAAGATTACAGCGAGACTCATTATTCGCCCTGATGGCAGAGAAATTACAACATA 360
          *****

                Pld →          ← Par
J CCCCCGGTTGAGGATCACTGTTATTATCATGGACGCATCGAGAATGATGCTGACTCAAC 230
B CCCTGCGGTTGAGGATCACTGCTATTATCATGGACGCATCGAGAATGATGCTGACTCAAC 420
          *** *****

                Exon 4
J TGCAAGCATCAGTGCATGCAACGGTTTGAAGGATATTTCAAGCTTCAAAGGGAGACGTA 290
B TGCAAGCATCAGTGCATGCAACGGTTTGAAGGACATTTCAAGCTTCAAAGGGAGACATA 480
          *****

J CTTTATTGAACCCTTGAAGCTTCCCACAGTGAAGCCCATGCAGTCTTCAAATATGAAAA 350
B CTTTATTGAACCCTTGAAGCTTCCAACAGTGAAGCCCATGCAGTCTTCAAATATGAAAA 540
          *****

```

P0d →

J **TGTAGAAAAGGAGGATGAGGC**CCCCAAAATGTGTGGGGTAACCCAGAATTGGAAATCATA 410
 B **TGTAGAAAAGGAGGATGAGGC**CCCCAAAATGTGTGGGGTAACCCAGAATTGGAAATCATA 600

← P1r | EXON 5

J **TGAACCCATCAAAAAGGCCTCTC**AGTTAGCTTTTACTGCT**TGAACAACAAAGATATGACCC** 470
 B **TGAACCCATCAAAAAGGCCTCTC**AGTTAGTTGTTACTGCT**TGAACAACAAAATATAACCC** 660

| EXON 6

J **CTACAAATACATTGAGTTTTTCGTAGTTGTGGACCAAGGAACCGT**CACAAAAACAATGG 530
 B **CTTCAGATACGTTGAGCTTTTCATAGTTGTGGACCAAGGAATGGT**CACAAAAACAATGG 720
 ** ** *****

← P0r P0dr → | EXON 7

J **CGATTTAGATAAGATAAAAAGCAAGAATGTATGAACTTGCCAACATTGTGAATGAGATTTT** 590
 B **CGATTTAGATAAGATAAAAAGCAAGAATGTATGAACTTGCCAACATTGTGAATGAGATTTT** 780

← P2r

J **CAGATACCTGTATATGCATGTAGCACTGGTTGGCCTAGAAATTTGGTCCAATGGAGATAA** 650
 B **AAGATACCTGTATATGCATGCAGCACTGGTTGGCCTAGAAATTTGGTCCAATGGAGATAA** 840

P4d →

J **GATTACCGTGAAG**CCGGACGTGGATTATACTTTGAATTCATTTGCAGAATGGAGAAAAAC 710
 B **GATTACCGTGAAG**CCGGACGTGGATTATACTTTGAATTCATTTGCAGAATGGAGAAAAAC 900

← P4irD P4ir →

J **AGATTTGCTGACTCGCAAAAAACATGATAATGCTCAGTTACTCACAGCAATTGA**CTTCAA 770
 B **AGATTTGCTGACTCGCAAAAAACATGATAATGCTCAGTTACTCACAGCAATTGA**CTTCAA 960

J TGGACCAACTATAGGATATGCTTACATAGGCAGCATGTGCCACCCGAAGCGTTCTGTAGG 830
 B TGGACCAACTATAGGATATGCTTACATAGGCAGCATGTGCCACCCGAAGCGTTCTGTAGC 1020

← P4r

J AATTGTTGAGGATTATAGCC**CAATAAATCTTGTGGTTGC**TGTTATAATGGCCCATGAGAT 890
 B AATTGTTGAGGATTATAGCC**CAATAAATCTTGTGGTTGC**TGTTATAATGGCCCATGAGAT 1080

Exon V P11d →

J GGGTCATAAATTTGG**GCATTCATCATGACACAG**TTTCTGTTCTTGTGGTGATTACCCATG 950
 B GGGTCATAAATCTGG**GCATTCATCATGACACAG**ATTTCTGTTCTTGTGGTGATTACCCATG 1140

| Exon X ← P11r P9d →

J **CATTATGGGTCCCACGATAAGCAATGAACCTTCGAAATTTTT**CAGCAATTGTAGTTATAT 1010
 B **CATTATGGGTCCCACGATAAGCAATGAACCTTCGAAATTTTT**CAGCAATTGTAGTTATAT 1200

J **CCAATGTTGGGACTTTATTATGAATCACAACCCAGAATGCATTATCAATGAACCCCTTGGG** 1070
 B **CCAATGTTGGGACTTTATTATGAAGGAGAACCCACAATGCATTCTCAATGAACCCCTTGGG** 1260

P6d →

J **AACAGATATTATTTACCTCCAGTTTGTGGAATGAACTTTTGGAGGTGGGAGAAGAATG** 1130
 B **AACAGATATTGTTTACCTCCAGTTTGTGGAATGAACTTTTGGAGGTGGGAGAAGAATG** 1320

← P9r | Exon W

J **TGACTGTGGA**ACTCCTGAAAA**TTGTCAA**AATGAGTGCTGCGATGCTGCAACGTGTA**AACT** 1190
 B **TGACTGTGG**CACTCCTGAAAA**TTGTCAA**AATGAGTGCTGCGATGCTGCAACGTGTA**AACT** 1380

← P6r P7d → | Exon Y

J **GAAATCAGG**TCACAGTGTGG**CATGGAGACTGTTGTGAGC**AA**TGCAA**ATTTAGCAA**ATC** 1250
 B **GAAATCAGG**TCACAGTGTGG**CATGGAGACTGTTGTGAGC**AA**TGCAA**ATTTAGCAA**ATC** 1440

← P7r P12d →

J **AGGAACAGA**ATGCCGGGCATCAATGAGTGAATGTGACCC**GGCTGAACACTGC**ACTGG**CC**A 1310
 B **AGGAACAGA**ATGCCGGGCATCAATGAGTGAATGTGACCC**GGCTGAACACTGC**ACTGG**CC**A 1500

← P12r

J **ATCTTCTG**AGTGTCTCCTGCAGATGTCTTCCATAAGAATGGACAACCAT**GCCTAGATA**ACT**A** 1370
 B **ATCTTCTG**AGTGTCTCCTGCAGATGTCTTCCATAAGAATGGACAACCAT**GCCTAGATA**ACT**A** 1560
 *** *****

P10d →

J **CGGTTACT**CGCTACAATGGGAATTGCC**CCATCATGTATC**ACCAATGTTATGCT**CTCTTTGG** 1430
 B **CGGTTACT**CGCTACAATGGGAATTGCC**CCATCATGTATC**ACCAATGTTATGCT**CTCTTTGG** 1620

| Exon H

J **TGCAGATG**TTTATGAGGCTGAAGATTCATGCTTCAAAGATAACCAGAAAGGCAATTAT**T**A 1490
 B **TGCAGATG**TTTATGAGGCTGAAGATTCATGCTTCAAAGATAACCAGAAAGGCAATTAT**T**A 1680

P8d → ← P10r

J **TGGCTACT**GCAGAAAGGAAAATGGTAAAA**GATTCCATGTGCACCAGA**AGATGTAAAATG 1550
 B **TGGCTACT**GCAGAAAGGAAAATGGTAAAA**GATTCCATGTGCACCAGA**AGATGTAAAATG 1740

Exon Z P8id →

J **TGGCAGGT**TATACT**GCAAAGATAATTCACCTGG**ACAAAATAATCCTTGCAAGATGTT**C**T**A** 1610
 B **TGGCAGGT**TATACT**GCAAAGATAATTCACCTGG**ACAAAATAATCCTTGCAAGATGTT**C**T**A** 1800

P5d → ← P8r

J **TTCCAACG**ATGATGAACATAAGGGAATGGTTCT**TTCTGGAACAAAATGTGCAG**ATGG**AA** 1670
 B **TTCCAACG**ATGATGAACATAAGGGAATGGTT**TTCTGGAACAAAATGTGCAG**ATGG**AA** 1860

| Exon K

J **GGTGTGC**AGCAACGGGCATTGTGTTGATGTGGCTACAG**CTACTAGTCAACCTCTGG**C**TT** 1730
 B **GGTGTGC**AGCAACGGGCATTGTGTTGATGTGGCTACAG**CTACTAGTCAACCTCTGG**C**TT** 1920

← P5r

J **TGATTTT**GGAGATCCTCCTTCCAGAAGGTTTGGCTTCTCTCAAG**TCCAAAGAGATCC**AT**C** 1790
 B **TGATTTT**GGAGATCCTCCTTCCAGAAGGTTTGGCTTCTCTCAAG**TCCAAAGAGATCC**AT**C** 1980

P3d →

J **TGCCTGC**ATCTT**ACT**TAGTAAATCACTCTTAGCTTCCAGATGGCATCTAAATTCTGCA**ATA** 1850
 B **TGCCTGC**ATCTT**ACT**TAGTAAATCACTCTTAGCTTCCAGATGGCATCTAAATTCTGCA**ATA** 2040

J **TTTCTT**CACCATATTTAC-----CCTTTTGCTGTAATCAAACCTTTTCCC**ACCACA** 1902
 B **TTTCTT**CACCATATTTAATCTGTTTACCTTTTGCTGTAATCAAACCTTTTCCC**ACCACA** 2100

J **A-GCTCC**ATGGGCATGTACAACACCAAAGGCTTATTTGCTGTCAAGAAAA**AAATGG**CC**A** 1961
 B **AAGCTCC**ATGGGCATGTACAACACCAAAGGCTTATTTGCTGTCAAGAAAA**AAATGG**CC**A** 2160
 * *****

J TTTTATACCGTTTGCCAATTGCAGAGCACATTTAATACAAGAAGTTCTGCCTTTTGAGC 2021
 B TTTTATACCGTTTGCCAATTGCAGAGCACATTTAATACAAGAAGTTCTGCCTTTT-GAGC 2219

J TGGTGTGTTTGAAGTCAATGCTTCCTCTCCAAAATTTGTATGGCTGGCTTTCCAACGAT 2081
 B TGGTGTATTCGAAGTCAATGCTTCCTCTCCAAAATTT-TGTG-CTGGCTTTCCAA-GAT 2276
 ***** * ** *****

← P3r

J GTAGCTGCTTCCATCAATAAACTATTTTCATTCTGCA 2218
 B GTAGCTGCTTCCATCAATAAACTATTTTCATTCTGCA 2333

S agagtttgtg gcaagaaaat ttcagaaagc aaacaaaatg ttaaagaaaa ataaagtgtt
 I agagtttgtg gcaagaaaat ttcagaaagc aaacaaaatg ttaaagaaaa ataaagtgtt

 S tttggtttta attttctttc aataggtcaa ttagccatcc tgtaattaaa cttacataat
 I tttggtttta attttctttc aataggtcaa ttagccatcc tgtaattaaa cttacataat

 S attcagccaa acaaaaagcta ttgctattgc tacagagaat cttcaaacta cagtcaagat
 I attcagccaa acaaaaagcta ttgctattgc tacagagaat cttcaaacta cagtcaagat

 S aatatgacat taacttttct caatgcaaga ttttgaatta tcttaaatta tttcttcct
 I aatatgacat taacttttct caatgcaaga ctttgaatta tcttaaatta tttcttcct

 S tgtttgattc aattgattca gaagttgcc aacacaacta agggttagag acttatagta
 I tgtttgattc aattgattca gaagttgcc aacacaacta agggttagag acttatagta

 S aacttgcttg ctagtaggtg gagttgaacc cctcccatga ttctgggagc atcaggtaga
 I aacttgcttg ctagtaggtg gagttgaacc cctcccatga ttctgggagc atcaggtaga

 S aatgggtagg tacataaata gatgtgagga agatgtattc ttctgtctcg ggttgctcat
 I aatgggtagg tacataaata gatgtgagga agatgtattc ttctgtctcg ggttgctcat

 S cagtctgaga aaacacattt ccagggccca acgctagttt ctctctttct cagtctctct
 I cagtctgaga aaacacattt ccagggccca acgctagttt ctctctttct cagtctctct

 S ctctgtctca attcctatag gtggtgggct agcagtttct ttcccgtacc ctgtgggtat
 I ctctgtctca attcctatag gtggtgggct agcagtttct ttcccgtacc ctgtgggtat

 S tgatacttga gtttattctt ttgtttccac ttgcaactat aagcacaacc ttaccctcc
 I tgatacttga gtttattctt ttgtttccac ttgcaactat aagcacaacc ttaccctcc

 S cttccattaa tctgctaaca aatcttggtc tcagttgtga ttgtacgtcc aagagaacgt
 I cttccattaa tctgctaaca aatcttggtc tcagttgtga ttgtacgtcc aagagaactt

 S gcagtgctcag gctgccattt tcacaaaata agttaaggaa caaatatcag gtgaccatcc
 I gcagtgctcag actgccattt tcacaaaata agctaaaggaa caaatatcag gtgaccatcc

 S aatgtccata tggggaccag ctccaaattg ttcagctatt agtcaaattc tgaatcact
 I aatgtccata tggggaccac ctccaaattg ttcagctatt agtcaaattc tgaatcact

 S acattattta catgttctta ggtgttccgg tgataattat gatgttggtg gttgtccata
 I atattattta catgttctta gatgtctccag tgataattat gatgccgtgg gttgtccata

← Pbr

```

S gattcttttg aaaatggaag aaaggaaagt cagcaaaact attaggatgg cagtaaaata
I gattctttta aaaatggaag aaaggaaagt cagcaaaact attaggatgg cagtaaaata
*****
S gcagttcagc tgaaggaat tactaatcag ctgcacttgc attcccagag aaaaggattt
I gcagttcagc tgaaggaat tactaatcag ctgcacttgc attcccagag aaaaggattt
*****
S ctgtgctgtg tcccactaga aaaagattcc attcttcact gatgaaccgg ataacaaatt
I ctgtgctgtg tcccactaga aaacgtttcc attcttcact gat-aaccgg ataacaaatt
*****
S gtgttggtgga tgtgaaaatg gtagaataac atcccaaagt aaaccaaacc ttcttttctt
I gtgttggtgga tgtgaaaatg gtagaatcac atcccaaagt aaaccaaacc ttcttttctt
*****
      | Pró-domínio
      | Exon 2
      | PbirD →
B      GAGC TCTATAATCC TGGAATCTGG GAACGTGAAT GATTATGAAG TAGTGTATCC
b      S   S   I   I   L   E   S   G   N   V   N   D   Y   E   V   V   Y   P
1 tttaggGAGC TCTATAATCC TGGAATCTGG GAACGTGAAT GATTACGAAG TAGTGTATCC
2 tttaggGAGC TCTATAATCC TGGAATCTGG GAACGTGAAT GATTACGAAG TAGTGTATCC
3 tttaggGAGC TCTATAATCC TGGAATCTGG GAACGTGAAT GATTACGAAG TAGTGTATCC

B ACGAAAAGTC ACTGCATTGC CCAAAGGAGC AGTTCAGCCA AAG
b  R  K  V  T  A  L  P  K  G  A  V  Q  P  K
J      GCAACGAGGC CCAAAGGAGC AGTTCAGCCA AAG
j      A  T  R
1 ACGAAAAGTC ACTGCATTGC CCAAAGGAGC AGTTCAGCCA AAGgtaagaa aaagatcttt
2 ACGAAAAGTC ACTGCATTGC CCAAAGGAGC AGTTCAGCCA AAGgtaagaa aaagatcttt
3 ACGAAAAGTC ACTGCATTGC CCAAAGGAGC AGTTCAGCCA AAGgtaagaa aaagatcttt
*****

1 ctttcatcaa caaattatt- -----
2 ctttcatcaa caaatgattt cttgtcagcc cacagaaatg ttgttattcc attgctgcc
3 ctttcatcaa caaattattt cttgtcagcc cacagaaatg ttgttattcc tttgctgcc
*****

1 -----
2 tctaattggtt atctggattt tgcattacaa tctatgttcc tgcttgaatt atttatgtat
3 tctaattggtt atctggattt tgcattgcaa tctatgttcc tgcttgaatt atttatgtat
*****

1 -----
2 taattaaaga ttaattaata cccaaattgt gcccaattca cacaaaattg aggagattcc
3 taattaaaaa aaattatatg gcc----- gcccaattca cacaaaactg aggagattcc
*****

1 -----
2 tggcaattta caaatagaat ttcctaatat atataggtta gtataactcg taaagtgttg
3 tggcaattta caaatagaac ttcctaatat gtataggtta gtataactca taaagtgttg
*****

1 -----
2 ggagacaccc agccgtaccg tgaggagggg tcgagtgggg aattagccat gagtcatttc
3 agagacaccc aggtgtaccg tgaggagggg tcaagtaggg aactagccat gagtcatttc
*****

```



```

1 atagctgcaa cagatccagg actgacatcc tccccactc caggaaattc cattaaagtt
2 atagctgcaa cagatccagg actgacatcc tccccactt caggaaattc catgaaagtt
3 atagctgcaa cagatccaga actgacatcc tccccactc caggaaattc cattaaagtt
  *****

```

```

1 tatctatctg ttctgaagag aaggggagca gtctataggg agattcctgc attataggca
2 tatctatctg ttctgaagag aaggggagca gtctataggg agatttctgc attataggca
3 tatctatctg ttctgaagag aaggggagca gtctataggg agattcctgc attataggca
  *****

```

```

1 gaatacagta aatgatttaa tttaaactct gcatgtctac atgcatcaat tagaagtata
2 gaatacagta aagtgtttaa tttaaactct gcatctctac atgcatcact tagaaatgta
3 gaatacagta aagtacttaa tttagaatct gcatgtctac ctgcatca-- ----atata
  *****

```

```

1 aaagatgt-- -----ata gtggttacia ctcaagatta gttctgagaa
2 aaagatgtgt ttctagggga agatggtata atggttacia ctcaagatta gttctgagaa
3 aaaggtgtgt ttctagggga agatggtata gaggttacia cgcaagatta tttctgagaa
  *****

```

```

1 tgagattcaa gtct----- -----cttt tgagtcagtc tctctctgtc ccagctcaac
2 tgagattcaa gtct----- -----cttt tgagtcagtt tctct--gtc ccagctcaac
3 tgagattcaa gtctatcgct ggatgacttt tgagtcagtc tctct--gtc ccagctcacc
  *****

```

```

1 ctcatcacc aagatggtca attagagaat gaaatcttga acttatgaag gaa--t-gtt
2 ctcatcacc gagatggtca attagagaat gaaatcttga acttatgcag aa---t-gtt
3 ctcatcacc aagatggtca attagagaat gagatcttga acttatgcag aaagatcggt
  *****

```

```

1 attttaagat cttacagttc a----- -----
2 attttaagat cgtacagttc a----- -----
3 ataagatctt acagttgaca tcactagatg aaatactcta aatctaata a-taatttgg

```

```

1 -----
2 -----
3 ctatggcaaa ttattcatta gatttaaatt gatacatgtg tagagttggt aagaggatac

```

```

B
b
J
1 ----- ---aagcact --agcctgaa ttctttgctt ttcagT
2 ----- ---aagcacc --agcctgaa ctctttgctt ttcagT
3 acaacttagt tgcttaattc ttttaagaagt aaagtctgaa cactttgctt ttcagT

```

Exon 3

← Pair Paid →

```

B ATGAAGACGC CATGCAATAT GAATTTAAAG TTAATGGAGA GCCAGTGGTC CTTCACCTGG
b  E D A M Q Y E F K V N G E P V V L H L E
J ATGAAGACGC CATGCAATAT GAATTTAAAG TGAATGGAGA GCCAGTGGTC CTTCACCTGG
j
C ATGAAGACGC CATGCAATAT GAATTTAAGG TGAATGGAAA GCCAGTGGTC CTTCACCTGG
c
  *****

```

B AAAAAAATAA AGGACTTTTT TCAAAAAGATT ACAGCGAGAC TCATTATTCC CCTGATGGCA
 b K K N G L F S K D Y S E T H Y S P D G R
 J AAAAAAATAA AGGACTTTTT TCAAAAAGATT ACAGCGAGAT TCATTATTCC CCTGATGGCA
 j K I G
 C AAAAAAATAA AGGACTTTTT TCAGAAGATT ACAGCGAGAT TCATTATTCC CCTGATGACA
 c E I D
 ***** ***** *** ***** ***** ***** ***** **
 Pld → ← Par
 B GAGAAATTAC AACATACCCT GCGGTTGAGG ATCACTGCTA TTATCATGGA CGCATCGAGA
 b E I T T Y P A V E D H C Y Y H G R I E N
 J GAGAAATTAC AACATACCCC CCGGTTGAGG ATCACTGTTA TTATCATGGA CGCATCGAGA
 j Y P P C R
 S GAGAAATTAC AACAAACCCT CTGGTTGAGG ATCACTGCTA TTATCATGGA CACATCGAGA
 s L P L C H
 A
 I GAGAAATTAC AACAAACCCT CTGGTTGAGG ATCACTGCTA TTATCATGGA CACATCGAGA
 i L P L C H
 ***** ***** ***** ***** ***** ***** * *****

B ATGATGCTGA CTCAACTGCA AGCATCAGTG CATGCAACGG TTTGAA
 b D A D S T A S I S A C N G L K
 J ATGATGCTGA CTCAACTGCA AGCATCAGTG CATGCAACGG TTTGAA
 j A **Intron 3**
 S ATGATGCTGA CTCAACTGCA AGCATCAGTG CATGCAACGG TTTGAAgtaa gatagtttct
 I ATGATGCTGA CTCAACTGTA AGCATCAGTG CATGCAATGG TTTGAAgtaa gatagtttct
 i V N
 ***** ***** * ***** ***** ***** *****

S aatcttttat ttattttatt aagaataata cactggctct ggagttgtca cagataaaat
 I aatcttttat ttattttatt aagaataata cactggctct ggagttgtca cagataaaat
 ***** ***** ***** ***** ***** *****

S agaataccaa gaagagctcc tgccagtcac tagagacaat aggaagctaa atcagatctg
 I tgaataccaa gaagagctgc tgccagtcac tagagacaat aggaagctaa ataagatctg
 ***** ***** * ***** ***** ***** *****

S gagggacatc agtttttcaa gacttctgca gaattaagaa agactttgat tctaaaggag
 I gagggacgtc agtttttcaa gacttctgca gaattaagaa agactttgat tctaa-ggag
 ***** ** ***** ***** ***** ***** *****

S cttccaagcc cctgtctcct ttacatatag gaattgatgc tacaggtaga gaagataaat
 I cttccaagcc cctgtctcct ttacatatag gaggtgatgc tacaggtaga gaagataaat
 ***** ***** ***** ** ***** ***** *****

S aggtttcaaa ttgcatacct tgctagaaaag gtattccatg caaaagtatc taaaatgaat
 I aggtttcaaa ttgcatacct tgctagaaaag gtattccatg caaaagtatc taaa-----
 ***** ***** ***** ***** ***** *****

S actgtacaaa tatttcatgg ttgcactgag cgtgtgcaat ctcttatatc tttctgatca
 I ----tacaaa tatttcatgg ctgcattgag tgtgtgcaat ctcttatatc tttctgttca
 ***** ***** **** ***** ***** ***** *****

S tagctgtatc attgagaagt atgagaaaagc gatatgagtg ctgaggaaaag aaaataaaaa
 I tagctgtatc tttgagaggt atgagaaaagc gatatgagtg ctgaggaaaag aaaataaaaa
 ***** ***** ** ***** ***** ***** *****

S tgtcagtgag tgtaagaag gaatttaaaa acacagggaa tgtttacttg gttagttaaa
 I tgtcagtgag tgtaagaag ggattgaaaa acacagggaa tgttgacttg gttagttaaa
 ***** * ***** ***** ***** **** ***** *****

← Plrr

S aatgggtctc agagctgagt ttcattccca actagctaac atcatcagtg caggtgatct
 I aataggtctc agagctgagt ttcattccca actagctaac atcaccagtg caggttatct
 ***** ***** ***** ***** ***** ***** *****

S ctgacctaac aatattctct tctgttgctt ctccatctct gatctaccct tttccactat
 I ctgacctaac aatattctct tctgttgctt ctccatctct gatc-accct tttccac-at
 ***** ***** ***** ***** ***** ***** ***** *

Exon 4

B AGGACATTTT AAGCTTCAAA GGGAGACATA CTTTATTGAA CCCTTGAAGC
 b G H F K L Q R E T Y F I E P L K L
 J AGGATATTTT AAGCTTCAAA GGGAGACGTA CTTTATTGAA CCCTTGAAGC
 j Y T
 S tgttttacga AGGACATTTT AAGCTTCAAA GGGAGACGTA CTTTATTGAA CCCTTGAAGC
 s H T
 I tgttttacag AGGACATTTT AAGCTTCAAA GGGAGACGTA CTTTATTGAA CCCTTGAAGC
 i H T
 ***** **** ***** ***** ***** ***** *****

P0d →

B TTTCCAACAG TGAAGCCCAT GCAGTCTTCA AATATGAAAA TGTAGAAAAG GAGGATGAGG
 b S N S E A H A V F K Y E N V E K E D E A
 J TTCCCACAG TGAAGCCCAT GCAGTCTTCA AATATGAAAA TGTAGAAAAG GAGGATGAGG
 j P D
 S TTCCCACAG TGAAGCCCAT GCAGTCTTCA AATATGAAAA TGTAGAAAAG GAGGATGAGG
 s P D
 I TTCCCACAG TGAAGCCCAT GCAGTCTTCA AATATGAAAA TGTAGAAAAG GAGGATGAGG
 i P D
 ** * * * * ***** ***** ***** ***** ***** ***** *****

Cysteine switch

← P1

B CCCCCAAAT GTGTGGGGTA ACCCAGAATT GGAAATCATA TGAACCCATC AAAAAGGCCT
 b P K M C G V T Q N W K S Y E P I K K A S
 J CCCCCAAAT GTGTGGGGTA ACCCAGAATT GGAAATCATA TGAACCCATC AAAAAGGCCT
 S CCCCCAAAT GTGTGGGGTA ACCCAGAATT GGAAATCATA TGAACCCATC AAAAAGGCCT
 I CCCCCAAAT GTGTGGGGTA ACCCAGAATT GGAAATCATA TGAACCCATC AAAAAGGCCT
 ***** ***** ***** ***** ***** ***** *****

Catalítico

B CTCAGTTA GTTGTTACTG CTGAA
 b Q L V V T A E
 J CTCAGTTA GCTTTTACTG CTGAA
 j A F A
 S CTCAGTTA GTTGTTACTG CTGAAgtaag ctctcatatt aaacctggtt tgagaacagg
 s V F A
 I CTCAGTTA GTTCTTACTC CTGAAgtaag ctctcatatt aaacctggtt tgagaacagg
 i V L P
 ***** * * ***** ***** ***** ***** ***** *****

Intron 4

S gattcatata cttgtttcta aaatataacg taagagagaa attctttgtg ggggggtaat
 I gattcattta cttgtttcta aaatataacg taagagagaa attctttttt ggggggtaat
 ***** * ***** ***** ***** ***** ***** *****

S tagataatca -caaaggggg acacccccca tttatatttt tttattctag ccatgggtata
 I tagatagaat tcaaaagggg ggacccccca tttctatttt tttattctag ctatgggata
 ***** **** ***** ***** ***** ***** ***** *****

S aaagaaataa tggaaatatg atgaatagaa aatacatggt ccaggattat agcgttttct
 I aaagaaataa tgggaatatg atgaattgaa aatac-tg-t ccaggattat cgcgttt-ct
 ***** ** * ***** ** * ***** ** * ***** ** *

← P0dd

S ttgatcatgc aaagttagtt tggagatttg aatcaaaatg gagttaaatg aatctccaaa
 I ttgatcatgc aaagttagtt tggagatttg aatcaaaatg gagttaaatg aatctccaaa
 ***** ** * ***** ***** ***** ***** ***** ***** *****

S tcatttcaac tttctaaatc aattttttaa aataatcaaa ttatcaattt ggatcaattt
 I tcatgtcaac ttcctaaatc aattttttaa aataatcaaa ttatcaattt ggatcaattt
 **** * ** * ***** ***** ***** ***** ***** ***** *****

S gcatccaaag tgagagttcc ttccttcctt ccttcctttt ggtggcaaaa gatgaaactg
 I gcatccaaag tgagagttct ttccttcctt ccttcctttt ggtggcaaaa gatgaaactg
 ***** ***** ***** ***** ***** ***** ***** *****

S cctaatacat atctagaagt tcaaaatctc actttttt-a ttaaaatggt aaaattgaca
 I cctaatacat atctagaagt tcaaaatctc -cttttttta ttaaaatggt aaaattgaca
 ***** ***** ***** ***** * ***** ***** ***** *****

S tgaacaaca gtttcaacat tcttggtggc tgctgggagc aaacagcctt ttctttattg
 I tgaacaaca atttcaacat tcttggtggc tgctgggagc aaacagcctt tgctttattg
 ***** ***** ***** ***** ***** ***** ***** *****

Exon 5

B CAACAAAAT ATAACCCTT CAGATACGTT GAGCTTTTCA TAGTTGTGGA
 b Q Q K Y N P F R Y V E L F I V V D
 J CAACAAAGAT ATGACCCTA CAAATACATT GAGTTTTTCG TAGTTGTGGA
 j R D Y K I F V
 S ttctgtttag CAACAAAGAT ATAACCCTA CAAATACGTT GAGTTTTTCA TAGTTGTGGA
 s R N Y K V F I
 I ttctgtttag CAACAAAGAT ATAACCCTA CAAATACGTT GAGTTTTTCA TAGTTGTGGA
 ***** ** * ***** ** * ***** ** * ***** ** *

B CCAAGGAATG G
 b Q G M V
 J CCAAGGAACC G
 j T

Intron 5

S CCAAGGAATG Ggtaagtaac ctggatatct ttccattttc tttctgcatt ggagccaacc
 s M
 I CCAAGGAATG Ggtaagtaac ctggatatct ttccattttc tttctgcatt ggacccaacc
 ***** * ***** ***** ***** ***** ***** *****

← P0rr

S aatatttaca aagttaagag tcatttgtaa atatactcta atttgcacat ttgtttctcc
 I aatatttaca aagttaagag tcatttgtaa atatactcta atttgcacat ttgtttctcc
 ***** ***** ***** ***** ***** ***** ***** *****

S tgcttaccat ctaaatttat ttttatccca caccggtaa acagttctgt aggtttcttt
 I tgcttaccat ctaaatttat ttttatccca caccggtaa acagttctgt aggtttcttt
 ***** ***** ***** ***** ***** ***** ***** *****

S gaatcattca tgcactcttt gtttggtccaa attgtacaac acacgaagaa aaagatgtat
 I gaatcattca tgcactcttt gtttggtccaa attgtacaac acacgaagaa aaagatgtat
 ***** ***** ***** ***** ***** ***** ***** *****

S agtgaagaat gcttatctca cacatctctg aacatggaaa tatagcatcc aatttaggaa
 I agtgaagaat gcttatctca cacatctctg aacatggaaa tatagcatcc aatttaggaa

S agcatttcag taggttgacc aaatttttct cccaccagga aaaaagccca tgaatgctgc
 I agcatttcag taggttgacc aaatttttct cccaccagga aaaaagccca tgaatgctgc

S tgaaaagtta aaaaccttaa tatattaatg gtacaaatat agattaataaa atcagcacaa
 I tgaaaagtta aaaaccttaa tatattaatg gtacaaatat agattaataaa atcagcacaa

S aatcagcttc gataaaaatt attattatat caacaaattc aatgatcctt tcaactgatat
 I aatcagcttc gataaaaatt attattatat caacaaattc aatgatcctt tcaactgatat

S cttcttttctc tccctccttc tttcccttcg tcctataatt agtgatctca tcttagacat
 I cttcttttctc tccctccttc tttcccttcg tcctataatt agtgatctat acttagacat

Exon 6

← P0r

B TC ACAAAAAACA ATGGCGATTT AGATAAGATA AAAGCAAGAA
 b T K N N G D L D K I K A R M
 J TC ACAAAAAACA ATGGCGATTT AGATAAGATA AAAGCAAGAA
 S ttttatcttt tctgtaggTC ACAAAAAACA ATGGCGATTT AGATAAGATA AAAGCAAGAA
 I ttttatcttt tctgtaggTC ACAAAAAACA ATGGCGATTT AGATAAGATA AAAGCAAGAA

P0dr →

B **TGTATGAACT TGCCAACATT** GTGAATG
 b Y E L A N I V N E
 J **TGTATGAACT TGCCAACATT** GTGAATG **Intron 6**
 S **TGTATGAACT TGCCAACATT** GTGAATGagg taagaggaca tggtaaatat atttcagttg
 I **TGTATGAACT TGCCAACATT** GTGAATGagg taagaggaca tggtaaatat atttcagttg

S tggattttaa acatctctca gccccagccc agattctcac gtgttcacac aacacacttg
 I tggattttaa acatctctca gccccagccc agattctcac gtgttcacac aacacacttg
 * *****

S ttttcctgcc tgagcaaggg gttggactag atgactccaa ggtcccctct aacttattat
 I ttttcctgcc tgagcaaggg gttggactag atgactccaa ggtcccctct aacttattat

Microsatélite

S tattattatt attattatta ttattattat tattattatt attaagatgg atttctacaa
 I tattattatt attattatta ttattattat tattattatt attaagatgg atttctacaa

S gaatgcaatc tttgctggct ttgagtgagg caaactaga tctttgggtca gctgggtgatc
 I gaatgcaatc tttgctggct ttgagtgagg caaactaga tctttgggtca gctgggtgatc

S cagcatgtga ggccatccca gaaaatgggt tattcagtaa ctaggttaat aagcctctta
 I cagcatgtga ggccatccca gaaaatgggt tattcagtaa ctaggttaat aagcctctta

Eco RI

S tgggaatgca gccactgtgt tgggagaacg cagatccaaa agtgatgcaa gaattcaatc
 I tgggaatgca gccactgtgt tgggagaacg cagatccaaa agtgcctgcaa gaattcaatc

S gaactatatac tgtagctgga ttctttcagg tcatggttga tggttccgat cttggagatt
 I gaactatatac tgtagctgga ttctttcagg tcatggttga tggttctgat cttggagatt

S agaaagaggt ggacagagtc agcctattcc aagctcagta gattatgaaa ctgatctctg
 I agaaagagat ggacagagtc agcctattcc aagctcagta gattatgaaa ctgatctctg

S agaagaagcc tgcttataga agccaatgtc agttgctttt taatcttgaa atgtttctca
 I agaagaagcc tgcttataga agccaatgtc agttgctttt taatcttgaa atgtttctca

Exon 7

B AGATTTTA
 b I L
 J AGATTTTC
 j F

S aagtttactt gagtttactt tgagtttctg attgatgcca tttccacctg acAGATTTTA
 I a-gtttactt gagtttactt tgagtttctg attgatgcca tttccacctg acAGATTTTA
 * *****

P4d → ← P2r

B AGATACCTGT ATATGCATGC AGCACTGGTT GGCCTAGAAA TTTGGTCCAA TGGAGATAAG
 b R Y L Y M H A A L V G L E I W S N G D K
 J AGATACCTGT ATATGCATGT AGCACTGGTT GGCCTAGAAA TTTGGTCCAA TGGAGATAAG
 S AGATACCTGT ATATGCATGC AGCACTGGTT GGCCTAGAAA TTTGGTCCAA TGGAGATAAG
 I AGATACCTGT ATATGCATGC AGCACTGGTT GGCCTAGAAA TTTGGTCCAA TGGAGATAAG

B ATTACCGTGA AGCCGGACGT GGATTATACT TTGAATTCAT TTGCAGAATG GAGAAAAACA
 b I T V K P D V D Y T L N S F A E W R K T
 J ATTACCGTGA AGCCGGACGT GGATTATACT TTGAATTCAT TTGCAGAATG GAGAAAAACA
 S ATTACCGTGA AGCCGGACGT GGATTATACT TTGAATTCAT TTGCAGAATG GAGAAAAACA

← P4ir P4irD →

B GATTTGCTGA CTCGCAAAAA ACATGATAAT GCTCAGTTAC TCACAGCAAT TGAC
 b D L L T R K K H D N A Q L L T A I D
 J GATTTGCTGA CTCGCAAAAA ACATGATAAT GCTCAGTTAC TCACAGCAAT TGAC
 C GATTTGCTGA CTCGCAAAAA ACATGATAAT GCTCAGTTAC TCACAGCAAT TGAC

P13d →

B TTCAATGGACCAACTATAGGATATGCTTACATAGGCAGCATGTGCCACCCGAAGCGTTCTGTAG
 J TTCAATGGACCAACTATAGGATATGCTTACATAGGCAGCATGTGCCACCCGAAGCGTTCTGTAG

← P4r

B CAATTGTTGAGGATTATAGCCCAATAAATCTTGTGGTTGCTGTTATAATGGCCCATGAGATGGG
 J GAATTGTTGAGGATTATAGCCCAATAAATCTTGTGGTTGCTGTTATAATGGCCCATGAGATGGG

B TCATAATCTGG
 J TCATAAATTGG

← P11dr P11d → **Exon V** (parcial - 67 pb)

B GCATTCATCA TGACACAGAT TTCTGTTCTT GTGGTGGTTA CCCATGCATT ATGGGTCCCG
b I H H D T D F C S C G D Y P C I M G P T
J GCATTCATCA TGACACAGGT TCCTGTTCTT GTGGTGGTTA CCCATGCATT ATGGGTCCCG
j G
S GCATTCATCC TGACACAGAT TTCTGTTCTT GTGGTGGTTA CCCATGCATT ATGGGTCCCG
s D
I GCATTCATCA TGACACAGGT TCCTGTTCTT GTGGTGGTTA CCCATGCATT ATGGGTCCCG
i G
***** *
B GGATAAG
j I S
J GGATAAG **Intron v**
S GGATAAGgta aggattgagg aaatcttctt aatgcctttc caatcaagtt atttttaaat
I GGATAAGgta aggattgagg aaatctt-tt aatgcctttc caatcaagtt atttttaaat
***** **
S gcttgcaatt aattgttcaa ttaaataatac tgagcaactt acaaaaatta aatggtaggg
I gcttccagtt aattgttaaa ttaaataatac tgag-aactt acaaaaatta aatggtaggg
**** ** ** ***** ** *****
S ggttacaaaa aaagagggct ttgaccattt gtgtgtcagt ttaccatgtg tagttaaaat
I agttacaaaa aaagagggct ttggtcattt gtgtgtcagt ttaccatgtg tagttaaaat
***** ***** ** *****
S gtgttgatgatt taaaaaggta gaactgtcct cagcatttga atagccatcc ttcaggagct
I gtgtcggatt taaaaag-ta gaaatattct caacaattga atagccatcc ttcaggagct
**** * ** ***** ** ** * * ** ** *****
S tcctccaaaa cgccagtagt ttaaagaac agcataggaa gcatagccaa agaaacccaaa
I tcctccaaaa caccag-agt ttaaagaac agcgtaggaa gcatagccaa agaaacccaaa
***** * **** ** ***** *****
S tattaatatc tgtgcatgga taaaagagga aaccaagtta gacattaataa ctttcctaaa
I tattaatatc tgtg-atgga taaaagagga aaccaagtta gacgtt---- ---tcctaaa
***** ***** ***** ***** ** ** *****
S gtgtgatgta gaaagggcaa aaaattcttc tgaatcaggc aacagcagga agaaaaggat
I gtgtgaagca aaaaggggaa aaaattattc tgaaccaggc aacagcagga agaaaaggat
***** * * ***** ** ***** *****
P11dd →
S aaattagcat cttggggttt tctgctcg ccctttcatc aaggtagaat acgttacatc
I aaattagcat cttggggttt tctgctcg ccctttcatc aaggtagaat acgttacatc
***** ***** ***** ***** *****
S tctgtcatat attgcagtct tctgaaatca gcaggtagca acgtacagga agttattgac
I tctgtcatat attgcagtct tctgaaatca gcaggtagca acgtacagga agtcattgac
***** ***** ***** ***** *****
S ttacaaccat ttgttttagca accgttcaaa gttacagtgg ccctgagaac agtgacttac
I ttacaaccat ttgtt---- ---gttcaaa gttacaatgg ccctgagaac agtgacttac
***** ** ***** ***** ***** *****
S gactggttct cgcaattaca gccatcgcag cctcccttca gtcacagaat cgtgtattta
I gactagtcct tgcaattacg gccatcgcag cctcccttca gtcacagaac cgtgtattta
**** ** * ***** ***** ***** ***** *****

S tgacgcttgc agcatcctgc attcatggga ttctcattgg ggactttccc agctgtcttc
 I tgacgcttgc agcatcctgc attcatggga tcctcattgg ggactttccc agctgtcttc

 S ccatatgcaa agccaatgga gcaagctggg ccccttaatg attgtgtggt tcaacttaaca
 I ccatatgcaa agccaatgga gcaagctggg ccccttaatg actgtgtgat tcaacttaaca

 S aagtggcaaa gccggttgta aaacggagca cgactcactt aacaactggc ttgctgggca
 I aagcggcaac actgggttgta aaactgagca cgactcactt aacaactggc tagctgggca
 *** *****
 S gtggaaattc tgccccagtt gtgacag--a agtcaaggac taccaatggt ttcttccggg
 I atggaaattc tgccccatt gtgacagcca agtcaaggac tacccatatt ttcttctggg
 ***** **
 S tggaagttta gcccttactt catcaagaag gaaaaagaaa aaggaaaaag gaagcaaaat
 I tggaagttta gcccttactt catcaagaag gaaaaagaaa aaggaaaaag gaagcaaaat

 S cttctcagca attgtctggt gcaaaactcc ccacaacttt tgtgtaatga gttttgtgta
 I cttctcagca attgtctggt gcaagactcc ccataattta ttggtaatgt gttttgtgta
 ***** ** * * *
 S ggctcagtga gcatcccggc cattgatcaa tttctggttt cttcggctaa agcccgagga
 I ggctcagtga gcatcccggc tattgatcaa tttctgttt ctttggctaa agcctgaaga
 ***** ** *
 S cttcattgcc tcccatttca tgcaagggca aggaggccag gaagtatgga gaatgccacg
 I cttcattgcc tcccatttca tgcaagggca aggaggccag gaagtatgga gaatgccatg
 ***** ** *
 S tagcctttcc tgtccttcct gaa-ggattt -ccgggcac atgattcctc ttaataactt
 I tagcctttcc tgtccttcct aaatggattt gccgggcac atgattcctc ttaataactt
 ***** ** *****
 S ctagacactt tgtggtgttt tgtacggaag actttgttta gaaacaatt ttgctca-at
 I ctagacactt tatggtgttt tgtacagaag actttgctta gaaacaatt ttgctcatat
 ***** * *****
 S taaccccgcc aaaggggtct ggtgtaactt tgaccttga agaagaagac catcttgtct
 I taacccaca aaaggggtct ggcggatatt tgaccttga agaagataac catcttgtct
 ***** * *****
 S atttaatggc caccagcaga ttctctttgg cgttcgtggt tccaccccat aacttgagag
 I atttaatggc caccagcaga ttctctttgg tgttcttgg tccatccctt aatttgaggg
 ***** *****
 S ccaatgatta tattccaat gatttgaatt aacattagga aggtgctttc ctgggcttat
 I ccaatgatta tattcctaat gatttgaatt aacattagga aggtgctttg ctgggcttat
 ***** ** *****
 S gatgtcacia gatgcagctc acaactaata tcacatggag ttctcattgc aatgtatatt
 I gatgtcacia gatgcagctc acaactaata tcacatggag ttctcattgc aatgtatatt
 ***** *****

← P11rr

S tgttaaaaaa gaaaggaaag ttagtggaat taaaactgag ggagtataat ctgggactaa
 I tgttagaaaa gaaaggaaag ttagtggaat taaaactgag ggagcataat ctgggaccaa
 ***** **

S gatcaagggg tacgaagggg tctcctatct attctgtag ctttagagaa aaaaatcact
 I gatcaagggg tacgaagggg tctcctatct atttctgtag ttttagagaa aaaaatcact
 ***** **

S gcatttccttc attagccaac ctttttttttc ttatatgttt ttgcaatgaa ttctgttcct
 I gcatttccttc attagccatc ctttttttttc ttatatgttt ttgcaatgaa ttctgttcct
 ***** *

Exon X (178 pb)

B C AATGAACCTT
 b N E P S
 J C AATGAACCTT
 S agtccaaata tgggtgatgt catggtctta ttttggttca tctccacagC AATGAACCTT
 I agtccaaata tgggatgatgt catggtctta ttctggttca tgtctacagC GATGAACCTT
 ***** **

P9d → ← P11r

B CGAAATTTTT CAGCAATTGT AGTTATATCC AATGTTGGGA CTTTATTATG AAGGAGAACC
 b K F F S N C S Y I Q C W D F I M K E N P
 J CGAAATTTTT CAGCAATTGT AGTTATATCC AATGTTGGGA CTTTATTATG AATCACAACC
 j M N H
 S CGAAATTTTT CAGCAATTGT AGTTATATCC AATGTCGGGA CTTTATTACG AATCACAACC
 s T N H
 I CGAAATTTTT CAGCAATTGT AGTTATATCC AATGTCGGGA CTTTATTATG AATCACAACC
 I M N H
 ***** **

B CACAATGCAT TCTCAATGAA CCCTTGGGAA CAGATATTGT TTCACCTCCA GTTTGTGGAA
 b Q C I L N E P L G T D I V S P P V C G N
 J CAGAATGCAT TATCAATGAA CCCTTGGGAA CAGATATTAT TTCACCTCCA GTTTGTGGAA
 j E I G I
 S CACAATGCAT TCTCAATGAA CCCTTGAGAA CAGATATTGT TTCACCTCCA GTTTGTGGAA
 s Q L R V
 I CACAATGCAT TATCAATGAA CCCTTGGGAA CAGATATTGT TTCACCTCCA GTTTGTGGAA
 i Q I G I
 ** ***** *

Desintegrina

P6d → ← P9r

B ATGAACTTTT GGAGGTGGGA GAAGGAATGTG ACTGTGGCAC TCCTGAA
 b E L L E V G E E C D C G T P E
 J ATGAACTTTT GGAGGTGGGA GAAGGAATGTG ACTGTGGAAC TCCTGAA **Intron x**
 j V G
 S ATGAACTTTT GGAGATGGGA GAAGGAATGTG ACTGTGGCAC TCCTGAAgta agtctttttt
 s V G
 I ATGAACTTTT GGAGGTGGGA GAAGGAATGTG ACTGTGGCAC TCCTGAAgta agtctttttt
 i V G
 ***** **

S taaaaaatc aacaagaata tgcagaattt tctcaaaaat gggaatcctt gaaaaatcac
 I taaaaaatc aacaagaata tgcagaattt tctcaaaaat gggaatcctt gaaaaatcac
 ***** **

```

S ctggcctttgt aagtggtttg agccatccaa aaggccaatc tgtgaatggc atagatttgt
I ctggcctttgt aagtggtttg agccatccaa aaggccaatc tgtgaatggc atagatttgt
*****
S gtcctttatg tacatgcatg tatgaagggg tttcttggat ttggctaadc agttgtagag
I gtcctttatg tacatgcatg tatgaagggg tttcttggat ttggctaadc agttgtagag
*****
S gaatggagaa ctggtatctc tctaccattt ggagaagacc acatattttt taaaatgggt
I gaatggagaa ctggtatctc tctaccattt ggagaagacc acatattttt taaaatgggt
*****
S gattgaccac tccaagaaaa tcttccttc ctgaaaactc ccattttggt gatatggcca
I gattgaccac tccaagaaaa tcttccttc ctgaaaactc ccattttggt gatatggcca
*****
S cattatctgt cccactattt tcttgagcca ctttctcca taactgatta tctttaatct
I cattatctgt cccactattt tcttgagcca ctttctcca taactgatga tctttaatct
*****
S atgctctgat cctaataata ttaataaga acagtaattt acatttactc aatggctcatg
I atgctctgat cctaataata ttaataaga acagtaattt acatttactc aatggctcatg
*****
S taagatgtag aatagcccg cagaagcaatg gggcatggca aacatctcaa gaagaacacc
I -----
S tcccacccat tctcttggtc tataaagatg aggtggtaga aatatacttt cagtattgaa
I -----
S agatttttct acagtcacct tacaatcagg gtaatgttaa tactcatggg tgggtgcttct
I -----
S tctctggact ctcaaaggct ggcattgatta cgagaaaatt tctcatgaat aagaaaggat
I tggatt---- ct-aaaggct ggcattgatta tgagaaaatt tctcatgaat aagaaaggat
*****
S tggggctgta agttacgtgg gcttcacaat taggaaatga ggatattttt gttttattct
I tggggctgta agttacgtgg gcttcacaat taggaaatga ggatattttt gttttattct
*****
S tgtcagaggg aaatatcagg taaggctctt tctcagagag atgtcattat ctgtggcaat
I tgtcagaggg aaatatcagg taaggctctt tctcagagag atgtcattat ctgtggcaat
*****
S aacaaacaat ttgtgcattt gctagcatga acccataaga gggaacacat tgcagaaatg
I aacaaacaat ttgtgcattt gctagcatga acccataaga gggaacacat tgcagaaatg
*****
S tctttttcct tgacaatttc cctcttcaaa atacaccaat aaaaaagac cagatagaaa
I tctttttcct tgacagtttc cctcttcaaa atacaccaat aaaaaagac caggtagaaa
*****

```

Exon W (90 pb)

B AATT GTCAAAATGA GTGCTGCGAT
 b N C Q N E C C D
 J AATT GTCAAAATGA GTGCTGCGAT
 S attctctgtc atcatttgat atgttttggt tttcagAATT GTCAAAATGA GTGCTGCGAT
 I attctctgac atcatttgat acgttttggt tttcagAATT GTCAAAATGA GTGCTGCGAT
 ***** * ***** * ***** ***** ***** *****
 ← P6r P7d →
 B GCTGCAACGT GTAAACTGAA ATCAGGGTCA CAGTGTGGAC **ATGGAGACTG TTGTGAGCAA**
 b A A T C K L K S G S Q C G H G D C C E Q
 J GCTGCAACGT GTAAACTGAA ATCAGGGTCA CAGTGTGGAC **ATGGAGACTG TTGTGAGCAA**
 S GCTGCAACGT GTAAACTGAA ATCAGGGTCA CAGTGTGGAC **ATGGAGACTG TTGTGAGCAA**
 I GCTGCAACGT GTAAACTGAA ATCAGGGTCA CAGTGTGGAC **ATGGAGACTG TTGTGAGCAA**
 ***** ***** ***** ***** ***** *****

B TGCAAA
 b C K
 J TGCAAA **Intron w**
 C TGCAAAgtaa gacttgttta tgtttaacac caggagagat tttattctgc tccatactag
 C ccatacagaa atggctgttt actaacctc ttctctccc cttcttcag tttattttac
 C cttatgaac atatccatag ggaagataat ttaacaaat ttcagccttg tctcaatctc
 C aaatgcactc tttcagcatg ttaaaacgta tctgtgaaaa taatatattt cttctttgac
 C tgggtgcccta actcagcttc ctgactttct ggaagcttct aagaggtccc tggtaatgct
 C tgagattgca tggaaacca gtttaaacaa ggataggcaa tatttgagat gtgacatttt
 C cttctctgag ctttttagga tggaaatagg tgcaggagac ttctggaagt aaagttgctt
 C ttttccccac cttaagtttc tacctgctct ctaaagctc taaattcagg tgttttgatg
 C gctgcaggac catgaaaaga gaggtgcaag ttactcattg cttctttcta tgtgggatcc
 C cagttgactc tgtaatgaac tttttgagca gagtggcca aacattttg ttatttccat
 C atttccatca caagcctaga ttcacaacaa gagaagggaa ccacatgttt ttcagcatga
 C gacagaaaat gctatgaatg cttcttccca tgtaaagaaa taaaatacat catgagatgt
 C tcagcaattc actttttgct gcttttacat gacagccac ttgattttcc ctttatggtc
 C agccaacagg tagaacttgt ctttcaggaa ttgagccttt cattgcattc atttccccac
 C agcaaataag acagagtggg acttctaggc cccacacaga gttgtagcag ggaagggatg
 C ctttgcttg tgatcctcaa gacagatgaa gaggaggttt tgaaatgtgt tgtgaacat
 C ggtttgactc tttgatctct gctgctgaag agtgatagct gggagtattt ttgattctca

Exon Y (196 pb)

B TTT AGCAAATCAG GAACAGAATG CCGGGCATCA ATGAGTGAAT GTGACCCGGC
 b F S K S G T E C R A S N S E C D P A
 J TTT AGCAAATCAG GAACAGAATG CCGGGCATCA ATGAGTGAAT GTGACCCGGC
 C cccacagTTT AGCAAATCAG GAACAGAATG CCGGGCATCA ATGAGTGAAT GTGACCCGGC
 ***** ***** ***** ***** * ***** *****

← P7r P12d →

B **TGAACACTGC ACTGG**CCAAT CCTCTGAGTG TCCTGCAGAT GTCTTCCATA AGAATGGACA
 b E H C T G Q S S E C P A D V F H K N G Q
 J **TGAACACTGC ACTGG**CCAAT CCTCTGAGTG TCCTGCAGAT GTCTTCCATA AGAATGGACA
 C **TGAACACTGC ACTGG**CCAAT CCTCTGAGTG TCCTGCAGAT GTCTTCCATA AGAATGGACA
 ***** ***** ***** ***** ***** *****

← P12r

B ACCATG**CCTA GATAACTACG GTTACT**GTCTA CAATGGGAAT TGCCCATCA TGTATCACCA
 b P C L D N Y G Y C Y N G N C P I M Y H Q
 J ACCATG**CCTA GATAACTACG GTTACT**GTCTA CAATGGGAAT TGCCCATCA TGTATCACCA
 C ACCATG**CCTA GATAACTACG GTTACT**GTCTA CAATGGGAAT TGCCCATCA TGTATCACCA

Rico em Cisteina

B ATGTTATGCT CTCTTTGGTG CAG
 b C Y A L F G A D
 J ATGTTATGCT CTCTTTGGTG CAG **Intron y**
 C ATGTTATGCT CTCTTTGGTG CAGgtaagtc aaccttagga tttcagaata gtggtggtat
 ***** ***** ***

C aaaagattaa gacagtcatt gccattaatg aaagatattt tctcctttgt atccataata
 C aatcatcatt agattttcttt tctataactg agtttgaagg aaccttggag gtcttctagt
 C ccacctccct gctcaagcag gaagccttat accatttcca atctcttctt taaaaaacct
 C ccagtgttg aacaccacaga acttctgaag gtcttttcca gaaagaatat ttttttttaa
 C gattaaagcc tccattcttt gtaagcttta gagcccagag atggactttg ggggaaggat
 C gcttttgtga gttattcata cgttcccaa ttggagaaag atctgagcca ctcttaatga
 C aagataatgt aaattgaaca tgggactatg cgctaagagt attcaaacct gcaactgaag
 C ttttccttac atttggccat aataantgaa gaacagatgt tggggggttg tcatt**cttac**
 C **attgacaactc gtc**attttaat gcagtctata caaaatttca gcatg cattt taatgccag
 C tctattacaa attttcagca tggttttata ttgaaaaaga aactgaaata ctgcacacaa
 C tatttccaaa gcattataaa gtggtactaa cacttcatgt aatcttgatt ttatccctct
 C gttagtgtag cctaattata gtgtaaaacg attctgccac tattgcaaaa gtgtagccta
 C gattatgtgg gattttttgc agttgcaaca cactgctggc tcttatttaa gtgattgtcc
 C actagggctc caagattcct ctcacagtta ctaatattga gccaggtatc acctcttctg
 C tgtatttgtt ttttctttcc tgaatgtaga ttggtttgtt tttttacttt ttcgctagaa
 C atttaatttt cttagattgg gctcaatggt caaaagatcc ttgtggatcc ttaagcctat

Pyd →

C cttctggaga gattgctatt cctgccagta tgggtgttctc tgcaaattta ttgagttccc
 C cttccaccct cttgtctaaa ttgtttatga agggtcaggt atcatgtgat gaagattggg
 C aatctgaatt acaaatcaga acgttgctca acttgcatga ttaattgatc aattgttatt
 C ttgatcaggt aaatagcata aatgaggtaa cttccaaaga atagtaaaag ctcagtactt
 C tgcaagtaat ccaattgtat tcaaagaggt ttgtgcccag gaattgacag gtctcaaact
 C agaatgacac atgatggta tttcaagctg aattctttct tgtttcattc ctatagtcaa
 C aatcttgcca gataaggact ctactaattg catttacaat taccctctgt aaactattag
 C ggaggagatg tcgtcaccct ctttttctca cactttatat ccaggcagag ttgcctetta
 C ctcctctgga tttcaccctt cccttctggg aatccatttc ttcacggcat tccagcacag
 C gaatgtatcc atgatagagc taaagttgaa aaatgtggac acagaaagaa atctggattc
 C agattcccct acagtcattg atttaccaca tgggtcttctg caaggacat tgtagtgact
 C tctcctttgt gtcaagaaaa tagaataagc aagtttagtg taacactgta tgcgatagtg
 ← P10rr

C aaagcttcat aatgaggaat agttaataa gtaagtcaat attaactaaa ttatttagaa
 C cacctctgaa atcactttag ccattctttt tctgctggc atgttacaca ttcattgtatt

Exon H

B AT GTTTATGAGG CTGAAGATTC ATGCTTCAAA
 b V Y E A E D S C F K
 J AT GTTTATGAGG CTGAAGATTC ATGCTTCAAA
 C tgtcttgggt cctcactccc tcttacagAT GTTTATGAGG CTGAAGATTC ATGCTTCAAA
 ** ***** ***** *****

← P10r

B GATAACCAGA AAGGCAATTA TTATGGCTAC TGCAGAAAGG AAAATGGTAA AAAGATTCCA
 b D N Q K G N Y Y G Y C R K E N G K K I P
 J GATAACCAGA AAGGCAATTA TTATGGCTAC TGCAGAAAGG AAAATGGTAA AAAGATTCCA
 C GATAACCAGA AAGGCAATTA TTATGGCTAC TGCAGAAAGG AAAATGGTAA AAAGATTCCA
 ***** ***** ***** ***** ***** ***** *****

P8d →

B TGTGGACCAG AAG ATGTAAAATGTGGCAGGTTATACT
 b C A P E D
 J TGTGCACCAG AAG ATGTAAAATGTGGCAGGTTATACT
 C TGTGCACCAG AAG
 ***** **

P8id →

Exon Z

B GCAAAGATAA TTCACCTGGA CAAAATAATC CTTGCAAGAT GTTCTATTCC AACGATGATG
 b K D N S P G Q N N P C K M F Y S N D D E
 J GCAAAGATAA TTCACCTGGA CAAAATAATC CTTGCAAGAT GTTCTATTCC AACGATGATG
 j K M F S N D D
 C GCAAAGATAA TTCACCTGGA CAAAATGATC CTTGCAATTA CTACTATTCA TCCAGGCATG
 c N Y Y S S R H
 ***** ***** ***** ***** * ***** * **

← P8r P5d →

B AACATAAGGG AATGGTTCTT CCTGGAACAA AATGTGCAGA TGGAAAGGT
b H K G M V L P G T K C A D G K V
J AACATAAGGG AATGGTTCTT CCTGGAACAA AATGTGCAGA TGGAAAGGT
j H **Intron z**
S AAAATAAGGG AATGGTTCTT CCTGGAACAA AATGTGCAGA GGGAAAGGTt agtaaaagat
s N E
I AAAATAAGGG AATGGTTCTT CCTGGAACAA AATGTGCAGA GGGAAAGGTt agtaaaagat

S tacctctcat ctttgtgttc taaagtctga ttc----- ---t-aaggg gtgatcacta
I tacctctaat ctgtgtgctc taaagtgtga ttccatgggg tgatcaa
***** * **
S attcctttttt gcaatcgatc aatgctgaaa gctggcctag attttcaaag tacagagaat
S ctgcatagaa gaaaaagtat ccatccatct atccttcttt tggttgttgt tattttgctt
S ttccttcaaa gaacaacccc aacgaatgag gtggatgtcc agggctgttt ctttcttcca
S agacccaaat gcctggcctt ctgagagcct tgtgcttctc tctcac---- -----
S -----
I taaaataatt ttgcaatcga tcaatgctgt aaagcatctt ctggacagac tgatcaagaga

S cagtcctttc tctgaacttg taggcatctc catagggag gtggaagaga gaaatgagaa
I cagtcctttc tctgaacttg taggcatctc catagggag gtggaagaga gaaatgagaa

S aatcagggca gagctggttg tgacctaac atgaagccac ttagaaatgc tgaagcaaac
I aatcagggca gagctggttg tgacctaac atgaagccac ttagaaatgc tgaagcaaac

S ccaaaactta cctgaaagga tcaggaatca cttccccttg atttttagaa cctgaaagaa
I ccaaaactta cctgaaagga tcatgaatca cttccccttg atttttagaa cctgaaagaa

S gtttgggtta gttataagga atgacagagt ttggttgtgt agaaaagtgc tgccttactc
I gtttgggtta gttataagga atgacagagt ttggttgtgt agaaaagtgc tgccttactc

Exon K

B GTGCAGCAA CGGGCATTGT GTTGATGTGG
b C S N G H C V D V A
J GTGCAGCAA CGGGCATTGT GTTGATGTGG
S cattgataat ctcttgcttt gactttcagg tCTGCAGCAA CAGGCAGTGT GTTGATGTGA
s T
I cattgataat ctcttgcttt gactttcagg tCTGCAGCAA CAGGCAGTGT GTTGATGTGA

3' UTR

B CTACAGCCTA CTAGTCAACC TCTGGC---- -----TTTG ATTTTGGAGA TCCTCCTTCC
b T A Y
J CTACAGCCTA CTAGTCAACC TCTGGC---- -----TTTG ATTTTGGAGA TCCTCCTTCC
S CTACAGCCTA CTAAATCAACC TCTGGCTTCT CTCAGATTTG ATTTTGGAGA TCCTCCTTCC
I CTACAGCCTA CTAAATCAACC TCTGGCTTCT CTCAGATTTG ATTTTGGAGA TCCTCCTTCC

← P5r P3r →

B AGAAGGTTTG GCTTCTCTCA AGTCCAAAGA GATCCATCTG CCTGCATCTT ACTAGTAAAT
 J AGAAGGTTTG GCTTCTCTCA AGTCCAAAGA GATCCATCTG CCTGCATCTT ACTAGTAAAT
 S AGAATGTTTG GCTTCCCTGT AGTCCAAAGA GATCCATCTG CCTGCATCTT ACTAGTAAAT
 I AGAATGTTTG GCTTCCCTGT AGTCCAAAGA GATCCATCTG CCTGCATCTT ACTAGTAGAT

B CACTCTTAGC TTCCAGATGG CATCTAAATT CTGCAATATT TCTTCACCAT ATTTAATCTG
 J CACTCTTAGC TTCCAGATGG CATCTAAATT CTGCAATATT TCTTCACCAT ATTTAC----
 S CACTCTTAGC TTCCAGATGG CATCCAAATT CCGCAATGTT TCTTCTCCAT ATTTAATCTG
 I CACTCTTAGC TTCCATATGG AATCTAAATT ACGCAATATT TATTCTCCAT ATTTAATCTG

B TTTACCTTTT GCTGTAATCA AACCTTTTCC CCACCACAAA GCTCCATGGG CATGTACAAC
 J ---CCTTTT GCTGTAATCA AACCTTTTCC CCACCACAA- GCTCCATGGG CATGTACAAC
 S TTTACCTTTT GCTGTAATCA AACCTTTTCC CCGCCACAAA GCTCCATGGA CATGTACAAC
 I TTTACCTTTT GCTGTAATCA AACCTTTTTC CCACCATAAA GCTCCATAGG CATGTACAAC

B ACCAAAGGCT TATTTGCTGC CAAGAAAAAA AATGGCCATT TTATACCGTT TGCCAATTGC
 J ACCAAAGGCT TATTTGCTGT CAAGAAAAAA AATGGCCATT TTATACCGTT TGCCAATTGC
 S ACCAAGAGCT TATTTGCTGT CAA-AAAAAA TATGGCCATT TTA--CCGTT TGCCAATTGT
 I ACCAAAGGCT T----- -----TATT TTA--CCGTT TGCCAATTGC

B AGAGCACATT TAATACAAGA AGTTCTGCCT TTT-GAGCTG GTGTATTCTGA AGTCAATGCT
 J AGAGCACATT TAATACAAGA AGTTCTGCCT TTTT-GAGCTG GTGTGTTCTGA AGTCAATGCT
 S A-AGCACATT TAATGCAACA AATTCTGGCT TTT-GGGTTG GTGTATTCTGA AGTCGATGCT
 I A-AGCACATT TAATGCAACA AGTTCTGCCT TTT-GAGCTG GTGTATTCAA AGTCAATGCT
 * *****

← P5r

B TCCTCTCCCA AAATTT-TGT G-CTGGCTTT CCAA-GATGT AGCTGCTTCC ATCAATAAAC
 J TCCTCTCCCA AAATTTGTAT GGCTGGCTTT CCAACGATGT AGCTGCTTCC ATCAATAAAC
 S TCCTCTTCCA AAATTC-CAT G-CTGCCTTT CCAA-CATGT AGCTGCTTCC ATCAATAAAC
 I TCCTCTCCCA AAATTT-CAC G-CTGGCTTT CCAA-CATGT AGCTGCTTCC ATCAATAAAC

B TATTTTCATT CTGCA
 J TATTTTCATT CTGCA
 S TATTTTCATT CTGCA
 I TATTTTCATT CTGCA

ANEXO C – Alinhamento de aminoácidos de metaloproteínas

Alinhamento dos aminoácidos de metaloproteínas do veneno de *Bothrops jararaca* comparadas à TOX1 e TOX2 deste estudo. As posições de nucleotídeos divergentes estão em azul. Os diferentes domínios estão separados por uma linha vertical. As regiões *Cysteine switch*, ECD e Região hiper conservada estão indicadas por boxes. Em cinza as cisteínas conservadas. Jararagina (Jara), Bothropasina (Both), Bothrostatina (Bothr).

	Peptídeo sinal	Pró-domínio	
TOX1	MIEVLLVTICLAAFPYQGSSII	LESIGNVNDYEVVYPPKVTALPKGAVQPKYEDAMQYEFK	
TOX2	MIEVLLVTICLAAFPYQGSSII	LESIGNVNDYEVVYPPKVTALPKGAVQPKYEDAMQYEFK	
Jara	-----	-----ATR	PKGAVQPKYEDAMQYEFK
Both	IEVLLVTICLAAFPYQGSSII	LESIGNVNDYEVVYPRKVTALPKGAVQPKYEDAMQYEFK	
HF3	MIQVLLVTICLAAFPYQGSSII	LESIGNVNDYEVVYARKVTALPKGAVQPKYEDTMQYELK	
Bothr	MIEVLLVTICLAAFPYQGSSII	LESIGNVNDYEVIIYPRKVTALPKGAVQPKYEDAMQYELK	
	-----:----- -----:----- -----:----- -----:----- -----:-----		
	10	20	30 40 50 60
TOX1	VNGEPVVLHLEKNKGLFS	EDYSEIHYS	SPDREITTNPLVEDHCCYYHGH
TOX2	VNGEPVVLHLEKNKGLFS	EDYSEIHYS	SPDREITTNPLVEDHCCYYHGH
Jara	VNGEPVVLHLEKNKGLFS	KDYSEIHYS	PDGREITTYPPVEDHCCYYHGRIENDADSTASIS
Both	VNGEPVVLHLEKNKGLFS	KDYSEIHYS	PDGREITTYPAVEDHCCYYHGRIENDADSTASIS
HF3	VNGEPVVLHLEKNK	QLFSKDYSEIHYS	PDGREITTYPPVEDHCCYYHGRIENDADSTASIS
Bothr	VNGEPVVLHLEKNKGLFS	KDYSEIHYS	PDGRKITTNPVEDHCCYYHGRIENDADSTASIS
	-----:----- -----:----- -----:----- -----:----- -----:-----		
	70	80	90 100 110 120
			<i>Cysteine Switch</i>
TOX1	ACNGLKGHF	FKLQRETYFIEPLKLSN	SEAHAVFKYENVEKEDEAPKMC
TOX2	ACNGLKGHF	FKLQRETYFIEPLKLSN	SEAHAVFKYENVEKEDEAPKMC
Jara	ACNGLKGY	FKLQRETYFIEPLKLP	DSEAHAVFKYENVEKEDEAPKMC
Both	ACNGLKGHF	FKLQRETYFIEPLKLSN	SEAHAVFKYENVEKEDEAPKMC
HF3	ACNGLKGHF	FKLQGETYFIEPLKLP	NSEAHAVFKYENVEKEDEV
Bothr	ACNGLKGHF	FKLQGETYLI	EPLKLSN
	-----:----- -----:----- -----:----- -----:----- -----:-----		
	130	140	150 160 170 180
			<i>Catalítico</i>
TOX1	KASQLV	FTAEQQRYNPYKY	VEFFIVVDQGMVTKNNGDL
TOX2	KASQLV	LTPEQQRYNPYKY	VEFFIVVDQGMVTKNNGDL
Jara	KASQLA	FTAEQQRYDPYKY	IEFFVVVDQGT
Both	KASQLV	VVTAEQQRYNPF	FRYVELFIVVDQGMVTKNNGDL
HF3	KASQLV	VVTAEQQRYNH	YKYIELVILADYRMVTKNNGDL
Bothr	KASQ	SNLTP	EHQR---
	-----:----- -----:----- -----:----- -----:----- -----:-----		
	190	200	210 220 230 240
TOX1	MHAALV	GLEIWSNGDKITVKPDVDY	TLNSFAEWRKTDLLTRKKHDNAQLLTAID
TOX2	MHAALV	GLEIWSNGDKITVKPDVDY	TLNSFAEWRKTDLLTRKKHDNAQLLTAID
Jara	MHVALV	GLEIWSNGDKITVKPDVDY	TLNSFAEWRKTDLLTRKKHDNAQLLTAID
Both	MHAALV	GLEIWSNGDKITVKPDVDY	TLNSFAEWRKTDLLTRKKHDNAQLLTAID
HF3	IRIALV	GIEIWSNAD	LSNVTL
Bothr	IDIAL	TG	VEIWSNK
	-----:----- -----:----- -----:----- -----:----- -----:-----		
	250	260	270 280 290 300

TOX1 IHHDTDFCSCGDYPCIMGP
 TOX2 IHHDTGFCSCGDYPCIMGP
 Jara GYAYIGSMCHPKRSVGIVQDYSPINLVVAVIMAHMGHNLGIHHDTGSCSCGDYPCIMGP
 Both GYAYIGSMCHPKRSVAIVEDYSPINLVVAVIMAHMGHNLGIHHDTDFCSCGDYPCIMGP
 HF3 GIANIASMNCNQNKSVGVMVDYSPINLVVAVIMAHMGHNLGINHDTGSCSCGGYSCIMAP
 Bothr GLAYWGSMDPKRSTAVIEDHSETDLLVAVTMAHELGHNLGIRHDTGSCSCGGYSCIMAP
 -----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
 310 320 330 340 350 360

Desintegrina

TOX1 TISNEPSKFFSNCSYIQCWDFITNHNPOCILNEPLRTDIVSPPVCGNELLEMGEEDCGT
 TOX2 TISNEPSKFFSNCSYIQCWDFIMNHNPOCILINEPLGTDIVSPPVCGNELLEVGEEDCGT
 Jara TISNEPSKFFSNCSYIQCWDFIMNHNPECIINEPLGTDIISPVPVCGNELLEVGEEDCGT
 Both TISNEPSKFFSNCSYIQCWDFIMKENPOCILNEPLGTDIVSPPVCGNELLEVGEEDCGT
 HF3 EISDQPSKLFSNCSKQAYQRYINYYKPOCILNEPLRTDIVSPPVCGNELLEMGEEDCGS
 Bothr VISHDIAKYFSDCSYIQCWDFIMKDNPOCILNKQLRTDTVSTPVS-G-KNFGAGEEDCGT
 -----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
 370 380 390 400 410 420

TOX1 PENCQNECCDAATCKLKSGSQCGHGDCCEQCKFSKSGTECRASNSECDFPAEHCTGQSSEC
 TOX2 PENCQNECCDAATCKLKSGSQCGHGDCCEQCKFSKSGTECRASNSECDFPAEHCTGQSSEC
 Jara PENCQNECCDAATCKLKSGSQCGHGDCCEQCKFSKSGTECRASMSECDFPAEHCTGQSSEC
 Both PENCQNECCDAATCKLKSGSQCGHGDCCEQCKFSKSGTECRASMSECDFPAEHCTGQSSEC
 HF3 PRNCRDPCCDAATCKLHSWVECESGECDDQCRFKGAGTECRAARSECDFIAESCTGQSADC
 Bothr PG---NPPCDAVTCKLRPGAQCAEGLCCDQCRFMKEGTVCRRAR-GDDMDDYCNGISAGC
 -----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
 430 440 450 460 470 480

Rico em Cisteína

TOX1 PADVFHKNQGPCLDNYGYCYNGNCPIMYHQCYALFGADVYEAEDSCFKDNQKGNYYGYCR
 TOX2 PADVFHKNQGPCLDNYGYCYNGNCPIMYHQCYALFGADVYEAEDSCFKDNQKGNYYGYCR
 Jara PADVFHKNQGPCLDNYGYCYNGNCPIMYHQCYALFGADVYEAEDSCFKDNQKGNYYGYCR
 Both PADVFHKNQGPCLDNYGYCYNGNCPIMYHQCYALFGADVYEAEDSCFKDNQKGNYYGYCR
 HF3 PTDDFKRNGQPCLHNYGYCYNGNCPIMYHQCYALFGSNATVAEDGCFEFNENGDKYFYCR
 Bothr PRNPFHA-----
 -----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
 490 500 510 520 530 540

Região hiper conservada

TOX1 KENGKKIPCAPEDKDNSPGQNNPCNYYSRHEHKGMVLPGTKCAEGKVCNSN
 TOX2 KENGKKIPCAPEDKDNSPGQNNPCNYYSRHEHKGMVLPGTKCAEGKVCNSN
 Jara KENGKKIPCAPEDVKCGRLYCKDNSPGQNNPCKMFYSNDDEHKGMVLPGTKCADGKVCNSN
 Both KENGKKIPCAPEDVKCGRLYCKDNSPGQNNPCKMFYSNDDEHKGMVLPGTKCADGKVCNSN
 HF3 KQSGVNIPCAQEDVKCGRLFCHT----KKHPCDYKYSEDPDY-GMVDNGTKCADGKVCNSN
 Bothr -----
 -----:-----|-----:-----|-----:-----|-----:-----|-----:-----|
 550 560 570 580 590 600

TOX1 RQCVDVTTAY
 TOX2 RQCVDVTTAY
 Jara GHCVDVTTAY
 Both GHCVDVATAY
 HF3 GHCVDVATAY
 Bothr -----
 -----:-----|-----
 610

ANEXO D - Aminoácidos

Tabela de aminoácidos.

		2 ^a base			
		U	C	A	G
1 ^a base	U	UUU (Phe/F) Fenilalanina	UCU (Ser/S) Serina	UAU (Tyr/Y) Tirosina	UGU (Cys/C) Cisteína
		UUC (Phe/F) Fenilalanina	UCC (Ser/S) Serina	UAC (Tyr/Y) Tirosina	UGC (Cys/C) Cisteína
		UUA (Leu/L) Leucina	UCA (Ser/S) Serina	UAA "Ocre" (Stop)	UGA "Opala" (Stop)
		UUG (Leu/L) Leucina	UCG (Ser/S) Serina	UAG "Âmbar" (Stop)	UGG (Trp/W) Triptofano
	C	CUU (Leu/L) Leucina	CCU (Pro/P) Prolina	CAU (His/H) Histidina	CGU (Arg/R) Arginina
		CUC (Leu/L) Leucina	CCC (Pro/P) Prolina	CAC (His/H) Histidina	CGC (Arg/R) Arginina
		CUA (Leu/L) Leucina	CCA (Pro/P) Prolina	CAA (Gln/Q) Glutamina	CGA (Arg/R) Arginina
		CUG (Leu/L) Leucina	CCG (Pro/P) Prolina	CAG (Gln/Q) Glutamina	CGG (Arg/R) Arginina
	A	AUU (Ile/I) Isoleucina	ACU (Thr/T) Treonina	AAU (Asn/N) Asparagina	AGU (Ser/S) Serina
		AUC (Ile/I) Isoleucina	ACC (Thr/T) Treonina	AAC (Asn/N) Asparagina	AGC (Ser/S) Serina
		AUA (Ile/I) Isoleucina	ACA (Thr/T) Treonina	AAA (Lys/K) Lisina	AGA (Arg/R) Arginina
		AUG (Met/M) Metionina, Start	ACG (Thr/T) Treonina	AAG (Lys/K) Lisina	AGG (Arg/R) Arginina
	G	GUU (Val/V) Valina	GCU (Ala/A) Alanina	GAU (Asp/D) Ácido aspártico	GGU (Gly/G) Glicina
		GUC (Val/V) Valina	GCC (Ala/A) Alanina	GAC (Asp/D) Ácido aspártico	GGC (Gly/G) Glicina
		GUA (Val/V) Valina	GCA (Ala/A) Alanina	GAA (Glu/E) Ácido glutâmico	GGA (Gly/G) Glicina
		GUG (Val/V) Valina	GCG (Ala/A) Alanina	GAG (Glu/E) Ácido glutâmico	GGG (Gly/G) Glicina

ANEXO E – Sequências no GenBank

Sequências depositadas no GenBank, referentes aos diferentes domínios e números de acesso.

Bothrops jararaca 5' UTR TOX1 – FJ90435

Bothrops jararaca 5' UTR TOX2 – FJ90436

Bothrops jararaca Pró-domínio TOX1 – FJ90437

Bothrops jararaca Pró-domínio TOX2 – FJ90438

Bothrops jararaca Domínio catalítico TOX1 – FJ90439

Bothrops jararaca Domínio catalítico TOX2 – FJ90440

Bothrops jararaca Domínio desintegrina TOX1 – FJ90441

Bothrops jararaca Domínio desintegrina TOX2 – FJ90442

Bothrops jararaca Domínio rico em cisteína TOX1 – FJ90443

Bothrops jararaca Domínio rico em cisteína TOX2 – FJ90444

Bothrops jararaca 3' UTR TOX1 – FJ90445

Bothrops jararaca 3' UTR TOX2 – FJ90446