ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene xx (2006) xxx – xxx
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Ohanin, a novel protein from king cobra venom: Its cDNA and genomic organization

Yuh Fen Pung [a], Sanjeed Vijaya Kumar [a], Nandhakishore Rajagopalan [a], Bryan G. Fry [a,b], Prakash P. Kumar [a,c,*], R. Manjunatha Kini [a,d,*]

[a] Department of Biological Sciences, Faculty of Science, National University of Singapore, 117543 Singapore
[b] Australian Venom Research Unit, Level 8, School of Medicine, University of Melbourne, Parkville, Victoria 3010, Australia
[c] Temasek Life Sciences Laboratory, National University of Singapore, 117604 Singapore
[d] Department of Biochemistry, Virginia Commonwealth University Medical Center, Richmond VA 23298-0614, USA

Received 7 September 2005; received in revised form 2 December 2005; accepted 6 December 2005

Received by J.A. Engler

Abstract

Ohanin, from king cobra venom, is a novel protein which induces hypolocomotion and hyperalgesia in mice [Pung, Y.F., Wong, P.T.H., Kumar, P.P., Hodgson W.C., Kini, R.M., 2005. Ohanin, a novel protein from king cobra venom induces hypolocomotion and hyperalgesia in mice. J. Biol. Chem. 280, 13137–13147]. It is weakly similar to PRY-SPRY domains (B30.2-like domain). Here we report the complete cDNA and genomic organization of ohanin. Interestingly, cDNA sequence does not show significant sequence similarity to any known sequences, including those of B30.2-like domain-containing proteins. Its full-length cDNA sequence of 1558 bp encodes for prepro-ohanin with a propeptide segment at the C-terminal. Ohanin is the first member of a new subfamily of proteins containing B30.2-like domain with short N-terminal segment. We named this subfamily as vespryns. There are two mRNA subtypes differing in their 5′-untranslated regions. Southern hybridization study shows that ohanin is encoded by a single gene. Its genomic sequence is 7086 bp with five exons and four introns, and the two types of mRNAs are generated by alternative splicing of exon 2. Our results indicate that ohanin and vespryns may have evolved from the same ancestral gene as B30.2 domain.
© 2006 Elsevier B.V. All rights reserved.

Keywords: B30.2-like domain; PRY-SPRY domains; Vespryns; Alternative splicing

1. Introduction

Snake venom is a cocktail of pharmacologically active peptides and proteins. It is rich in enzymes, such as phospholipases A_2, metalloproteinases, serine proteinases, L-amino acid oxidases, phosphodiesterases, acetylcholinesterases and nucleases (Karlsson, 1979; Torres et al., 2003). It also contains non-enzymatic proteins, such as three-finger toxins, serine proteinase inhibitors, helveprins/CRISPs, C-type lectin related proteins and waprins (Mochca-Morales et al., 1990; Kini, 2002; Torres et al., 2003; Yamazaki et al., 2003). However, snake venom proteins are still far from being completely cataloged and there is room to isolate and characterize novel snake venom proteins both structurally and functionally.

One such example is the recently isolated venom protein, ohanin, from king cobra (*Ophiophagus hannah*) (Pung et al., 2005). It is the first member of a new family of snake venom proteins. Unlike other snake venom proteins, which are rich in Cys residues and disulfide bonds, ohanin has a single Cys residue. Thus ohanin is unique. It shows homology (49% similarity and 38% identity) to consensus sequence of PRY-SPRY domains. Ohanin lowers the locomotor activity and induces

Abbreviations: SPRY domain, domain with unknown function in Ryanodine receptors and *Dictyostelium discoideum*; B30.2 domain, Domain which was named after the B30.2 exon maps within the Human Class I Histocompatibility complex.

* Corresponding authors. Kini is to be contacted at Protein Science Laboratory, Department of Biological Sciences, Faculty of Science, National University of Singapore, 14 Science Drive 4, Singapore 117543. Tel.: +65 6874 5235; fax: +65 6779 2486. Kumar, Plant Morphogenesis Laboratory, Department of Biological Sciences, Faculty of Science, National University of Singapore, 14 Science Drive 4, Singapore 117543. Tel.: +65 6874 2879; fax: +65 6779 2486.

E-mail addresses: dbskumar@nus.edu.sg (P.P. Kumar), dbskinim@nus.edu.sg (R.M. Kini).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.12.002

ARTICLE IN PRESS

hyperalgesia in mice (Pung et al., 2005). Both these biological activities are mediated through its effect on the central nervous system. We proposed that ohanin could serve both offensive and defensive roles by slowing down the locomotor activity and inducing pain in both preys and predators (Pung et al., 2005).

Protein sequencing of ohanin showed that ohanin has a PRY domain followed by a partial SPRY domain. To further characterize this novel protein, we have cloned and sequenced its cDNA. In this paper, we show that ohanin is synthesized as a precursor protein and the prepro-protein has the complete SPRY domain with the presence of the C-terminal propeptide segment. The maturation of ohanin appears to occur by a cleavage at the dibasic Arg–Arg site. The results also show the presence of two ohanin mRNA subtypes differing in their 5′-untranslated regions. Analysis of genomic DNA sequence indicates that alternative splicing in ohanin gene leads to these two subtypes. Data presented here show that ohanin belongs to a new subfamily of proteins which contain the B30.2-like domain. We have named the protein subfamily as vesspryns. This is the first report on the cDNA and genomic organization of this novel protein family.

2. Materials and methods

2.1. Materials

King cobra venom glands and liver tissue were frozen in liquid nitrogen immediately after dissection and kept in −80 °C until used. All the reagents and kits used for molecular biology study were as follows: oligonucleotides (1st Base Pte Ltd, Singapore), Platinum *Taq* polymerase, dNTP mix, λ*Hind*III and 1 Kb plus ladder (GIBCO BRL®, Carlsbad, CA), restriction endonucleases (New England Biolabs®, Beverly, MA), pGEMT-easy vector (Promega, Madison, WI), RNeasy® Mini kit, QIAGEN® OneStep RT-PCR kit, DNeasy® Tissue kit, QIAprep® Miniprep kit and QIAEX II Gel Extraction kit (Qiagen GmbH, Hilden, Germany), SMART™ RACE cDNA amplification kit, Universal GenomeWalker™ kit and BD Advantage™ 2 Polymerase Mix (Clontech Laboratories Inc., Palo Alto, CA), PCR DIG probe synthesis kit, positively charged nylon membrane, DIG™ Easy Hyb buffer, anti-DIG alkaline phosphatase antibody and chemiluminescent substrate CDP-*Star*™ (Roche Diagnostics GmbH, Deutschland, Germany), ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Version 3.0) (PE-Applied Biosystems, Foster City, CA) and Luria Bertani broth and agar (Q.BIOgene, Irvine, CA).

2.2. Isolation of total RNA

Total RNA was isolated from king cobra venom gland using the RNeasy® Mini kit. For each extraction, 30 mg venom gland tissue was first pulverized in liquid nitrogen and further homogenized for 20 to 30 s using a Heidolph DIAX600 homogenizer (Schwabach, Germany) in the presence of 600 μl Buffer RLT. The integrity of the RNA extracted was examined by denaturing agarose gel electrophoresis.

2.3. Reverse transcription-polymerase chain reaction (RT-PCR)

To generate gene-specific sequence, RT-PCR was performed using the QIAGEN® OneStep RT-PCR kit. In brief, RT-PCR mixture contained 2 μl of QIAGEN OneStep RT-PCR Enzyme Mix and the final concentration of 250 ng total RNA as template, 0.4 mM dNTP mix and 0.6 μM degenerate primers in a total volume of 50 μl. The degenerate primers used were: RT1 (sense primer) 5′-GGNAAYTGGCARAARGCNGA-3′ and RT2 (antisense primer) 5′-CCACCANARNCCYTTYTG-CCA-3′. The reverse-transcription and amplification conditions were: reverse-transcription at 50 °C/30 min; initial PCR activation step at 95 °C/15 min; immediately followed by 30 cycles of 3-step thermal cycling profile of denaturation at 94 °C/1 min, annealing at 50 °C/1 min, extension at 72 °C/2 min and a final extension at 72 °C/10 min. The PCR products were fractionated by 1.5% agarose gel electrophoresis. The most intense bands were excised and purified using QIAEX II Gel Extraction kit before ligation to pGEMT-easy vector. Eight clones were sequenced with T7 (sense) 5′-GTAATACGACT CACTATAGGGC-3′ and SP6 (antisense) 5′-TATTTAGGTGA CACTATAG-3′ primers using the dideoxy chain termination method (Sanger et al., 1977) on an automated ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Version 3.0) was used to carry out the cycle sequencing reaction. Data were analyzed using the Sequencing Analysis 3.7 (Sample Manager) software (Applied Biosystems, Foster City, CA).

2.4. Construction of 5′- and 3′-RACE cDNA libraries

The 5′- and 3′-RACE cDNA libraries were constructed using the SMART™ RACE cDNA amplification kit according to the manufacturer's protocol.

2.5. Isolation and sequencing of cDNA clones

The 5′- and 3′-RACE-Ready cDNA libraries were constructed using SMART™ RACE kit For cDNA amplification, the 5′-RACE reaction mix consisted of 2.5 μl 5′-RACE-Ready-cDNA, 5 μl Universal Primer Mix (UPM) containing Long primer 5′-CTAATACGACTCACTATAGGGCAAGCAGTGG TATCAACGCAGAGT-3′ and Short primer 5′-CTAATAC GACTCACTATAGGGC-3′; and the final concentration of 1.5 U Platinum *Taq* polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix and 0.2 μM GSP2 (antisense primer) 5′-CTTCCCAGCTAACCCAACAGCCCATTCCC-3′ in a total volume of 50 μl. The 3-step thermal cycling profile was as follows: 1 cycle of hot start at 94 °C/1 min; 30 cycles of denaturation at 94 °C/30 s, annealing at 67 °C/30 s, extension at 72 °C/2 min and followed by a final extension of 72 °C/10 min. The 3′-RACE reaction mix, which yielded the full-length cDNA, consisted of 2.5 μl 3′-RACE-Ready-cDNA, 5 μl UPM and the final concentration of 1.5 U Platinum *Taq* polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix and 0.2 μM

GSP1 (sense primer) 5′-GATCATTTGATCCAGAGAAGACA CAGTCTC-3′ in a total volume of 50 μl. The 3-step thermal cycling profile was as follows: 1 cycle of hot start at 94 °C/1 min; 30 cycles of denaturation at 94 °C/30 s, annealing at 68 °C/30 s, extension at 72 °C/3 min, followed by a final extension of 72 °C/10 min. The PCR products were fractionated by 1.5% agarose gel electrophoresis. The most intense bands for both the 5′- and 3′-RACE amplifications were excised, purified and ligated into pGEMT-easy vector. At least 16 clones from both the 5′- and 3′-RACE libraries were fully sequenced using T7, SP6 and internal primers.

2.6. Isolation of genomic DNA

Genomic DNA was isolated from king cobra liver tissue using the DNeasy® Tissue kit. For each extraction, 25 mg liver tissue was pulverized in liquid nitrogen using a mortar and pestle pre-cooled at −80 °C. The integrity of the genomic DNA extracted was examined by 0.8% agarose gel electrophoresis.

2.7. Southern blot hybridization

The digoxigenin (DIG)-labeled double-stranded DNA probes were prepared using the PCR DIG Probe Synthesis kit. The PCR reaction mix contained 0.5 μl ohanin cDNA in pGEMT-easy vector as template and a final concentration of 1.3 U Enzymes mix, 0.2 mM PCR DIG Labeling mix, 0.2 mM dNTP mix, 1× PCR buffer and 0.2 μM primers in a total volume of 25 μl. The gene-specific primers used were: P11 (sense) 5′-GCTGATGTGACGTTTGACTCAAACACA-3′ and P12 (antisense) 5′-AAGCCACCAGAGGCCCTTTTGCCA-3′. The 3-step thermal cycling involved a hot start at 95 °C/2 min followed by 30 cycles of 95 °C/30 s, 63 °C/30 s, 72 °C/40 s and a final extension of 72 °C/7 min. The PCR product was purified and the concentration was determined by A_{260}. Southern hybridization was performed according to the DIG System User's Guide (Roche Diagnostics GmbH, Deutschland, Germany). For each lane, 10 μg of genomic DNA was used. Detection of the signal was performed using the substrate CDP-*Star*™.

2.8. Construction of GenomeWalker libraries

The GenomeWalker libraries were constructed using the Universal GenomeWalker kit according to the manufacturer's protocol. The libraries were made with 2.5 μg of genomic DNA restricted by *Dra*I, *Eco*RV, *Pvu*II and *Stu*I, respectively.

2.9. Isolation and sequencing of genomic clones

Genomic organization of ohanin was studied using both the gDNA PCR and 'genome walking' approaches. As a first step, gDNA PCR was performed using gene-specific primers designed from the coding region of ohanin cDNA. The primers used were as follows: gDNAsigpep (sense) 5′-ATGCTCC TGTTCACACTATGCTTT-3′ and gDNAstop (antisense) 5′-C CCTGTTTTAATGAAGAATTGTAACCTCTTA-3. The PCR reaction mix contained 0.5 μl gDNA as template and a final concentration of 1.5 U Platinum *Taq* polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix and 0.2 μM primers in a total volume of 25 μl. The 3-step thermal cycling involved a hot start at 94 °C/2 min followed by 30 cycles of 94 °C/1 min, 60 °C/45 s, 72 °C/3 min and a final extension of 72 °C/10 min. The PCR products were fractionated on a 1.5% agarose gel and the band of interest was excised, purified and ligated into pGEMT-easy vector. Sixteen clones carrying the inserts were sequenced.

Secondly, gDNA amplification was further performed to obtain the genomic sequence which corresponds to the 3′-UTR of the cDNA. The gene-specific primers used were: gDNA3UTR1 (sense) 5′-CTATATAGGGGCACGTGTTT CACTC-3′ and gDNA3UTR2 (antisense) 5′-TACTAACAGT GAGACTTTATTAGTAG-3′. The PCR reaction mix contained 0.5 μl gDNA as template and a final concentration of 1.5 U Platinum *Taq* polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix and 0.2 μM primers in a total volume of 25 μl. The 3-step thermal cycling involved a hot start at 94 °C/1 min followed by 30 cycles of 94 °C/1 min, 60 °C/30 s, 72 °C/3 min and a final extension of 72 °C/10 min. The amplified fragment was purified, cloned and sequenced.

Finally, the genomic sequence that corresponds to the 5′-UTR of the cDNA was obtained from 'genome walking' and two additional gDNA PCR amplifications. The first 'genome walk' involved two sets of primers: adaptor primer 1 (AP1-sense) 5′-GTAATACGACTCACTATAGGGC-3′ provided in the kit and a 25-mer gene-specific primer gDNA5UTR1 (antisense) 5′-CTTT CTGCCAATTCCCAGGAGGTGA-3′; and the nested PCR adaptor primer 2 (AP2-sense) 5′-ACTATAGGGC ACGCGT GGT-3′ provided in the kit and a 27-mer gene-specific primer gDNA5UTRnes2 (antisense) 5′-AGCCA GAGCCTTTCCAC CATTTTCCTG-3′. Primary and nested PCRs were performed as recommended by the BD GenomeWalker™ Kit user's manual with the following modifications. The 25.0 μl reaction mixture consisted of 0.5 μl of DNA template (either from each library or from primary PCR products), 1× PCR buffer, 0.2 mM dNTPs, 0.2 μM appropriate adaptor primers, 0.2 μM of appropriate gene-specific primers, 1× BD Advantage™ 2 polymerase mix. The thermal cycling profile used was as follows: 7 cycles of 94 °C/2 s, 72 °C/3 min; 32 cycles of 94 °C/2 s, 67 °C/4 min followed by a final extension of 67 °C/7 min. The PCR products were purified, cloned and sequenced.

A second gDNA PCR amplification was performed using the gene-specific primers 1-gDNA5UTR (sense) 5′-CGGA TCCTGCACAATAGTTTTATCTCC-3′ and 2-gDNA 5UTR (antisense) 5′-GACATGACCATTACAAACTCTAGC TTC-3′. The PCR reaction mix contained 0.5 μl gDNA as template and a final concentration of 1.5 U Platinum *Taq* polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP mix and 0.2 μM primers in a total volume of 25 μl. The 3-step thermal cycling involved a hot start at 94 °C/2 min followed by 30 cycles of 94 °C/20 s, 62 °C/30 s, 72 °C/3 min and a final extension of 72 °C/10 min. The amplified fragment was purified, cloned and sequenced.

The final gDNA PCR amplification was performed using the gene-specific primers 9-gDNA5UTR (sense) 5′-GATCATTT GATCCAGAGAAGACACAGT-3′ and 6-gDNA5UTR

(antisense) 5′-CAAACAAATGGTCATAAGCTGAGGTC TAC-3′. The PCR reaction mix contained 0.5 µl gDNA as template and a final concentration of 1× BD Advantage™ 2 polymerase mix, 1× PCR buffer, 0.2 mM dNTP mix and 0.2 µM primers in a total volume of 25 µl. The 2-step thermal cycling involved a hot start at 95 °C/1 min followed by 30 cycles of 95 °C/15 s, 67 °C/4 min, and a final extension of 67 °C/4 min. The amplified fragment was purified, cloned and sequenced.

3. Results

3.1. Cloning and sequencing of ohanin cDNA

The amount of total RNA obtained from the king cobra venom gland was low (~4 µg per 30 mg fresh weight), but the quality of RNA was relatively good (data not shown). We used a combination of RT-PCR and RACE techniques to obtain the full-length cDNA of ohanin. To isolate gene-specific sequences, RT-PCR was first performed using the total RNA as template. Degenerate primers, RT1 and RT2, were designed based on ohanin's amino acid sequence (Pung et al., 2005). Analysis of the amplified fragment (~300 bp) revealed that all eight randomly selected clones encoded for ohanin (data not shown).

Two bands of approximately 550 and 600 bp were obtained from the 5′-RACE amplification in an attempt to clone the 5′-coding region and 5′-UTR (Fig. 1A). Sequence analysis of 16 independent clones derived from the bands demonstrated the existence of two cDNA subtypes. We further designed a sense primer, GSP1, from the beginning of the 5′-UTR sequence for the 3′-RACE amplification. GSP1 and UPM yielded the full-length cDNA of ohanin (Fig. 1B). Sixteen individual clones were fully sequenced. Of these randomly selected cDNA clones,

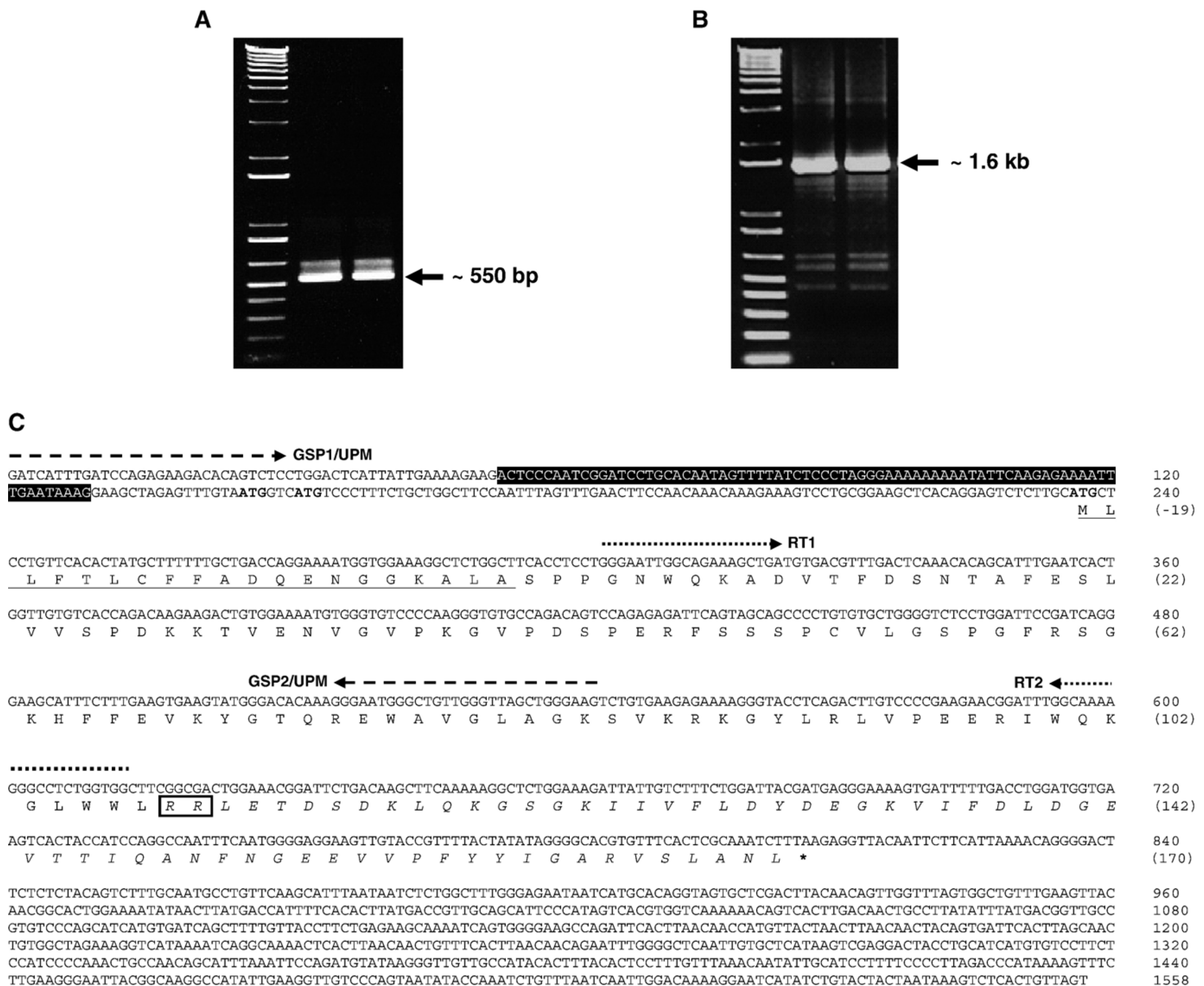

Fig. 1. Cloning and sequencing of ohanin cDNA. (A) 5′-RACE amplification. Partial coding region of ohanin together with its 5′-UTR was obtained from the 5′-RACE amplification using GSP2 and UPM. (B) 3′-RACE amplification. The 3′-RACE amplification which yielded the full-length cDNA of 1558 bp exclusive of poly-A tail was obtained using GSP1 and UPM. (C) Nucleotide sequence and deduced amino acid sequence (gb: AY351433). Nucleotide sequence is presented in the 5′- to 3′-orientation. Deduced amino acid sequence by reverse-translation from the putative open reading frame is shown: the three ATGs are indicated in bold; the putative signal peptide is underlined; dibasic processing site is boxed and propeptide segment is marked in italics. The stop codon is marked by an asterisk and the polyadenylation signal, AATAAA, is underlined twice. The missing stretch of nucleotides in type II cDNA is shaded black (See text for details).

A

```
Pro-ohanin   : ..ADVTFDSNTAFESLVVSPDKKTVENVGVPK-GVPDSPERFSSSPCVLGSPGF------RSGKHIFEVKY---
Thaicobrin   : ..ADVTFDSNTAFESLVVSPDKKTVENVGVSQ-VAPDNPERFDGSPCVLGSPGF------RSGKHIFEVKY---
RFP          : ..ADVTLDPETAHPNLVLSEDRKSVKFVETRLRDLPDTPRRFTFYPCVLATEGF------TSGRHYWEVEV---
BTN          : ..VDVTLDPDTAHPHLFLYEDSKSVRLEDSRQ-ILPDRPERFDSWPCVLGRETF------TSGRHYWEVEV---
PRY-SPRY     : ..VDVTLDPDTAHPYLSLSEDRKSVRYGDLKQ-SLPDNPERFDHYPCVLGSQGF-------SGRHYFEVEVFTG
Beta-SNTX    : ..CELTLDPETAHQVLTLSEGNKKAVSGNTKS--PTDHLEKFSHFQQVMCTKGL-------SGRHYWELEWS--
Alpha-SNTX   : ..CDLTFDRNTINNWISLSDNDTFAASEHGKRQNYPKHPERFVSFNQVLCNEGL-------MGKHYWEVEWN--
Enterophilin : ..ANVTLDPYTAHPALILSEE-RRVSMGEKPQ-DLPRSQMRFESLPCVLGKQSF------SSERHIWEVKV---
SOCS box     : ..LDQILDMPAAGLAVQLRHAWNPEDR-SLNVFVKDDDRLTFHRHPVAQSTDGIRGKVGHARGLHAWQIHNPAR

                   LDP motif                                                      WEVE motif

Pro-ohanin   : GTQREWAVGLAGKSVKRK---GYLRLVPEERIWQKGL-----WWLRRLETDSDKLQKGSG-KII-VFLDYDEGK
Thaicobrin   : GTQREWAVGLAGKSVKRK---GYLRLVPEERIWQKGL-----WWLG
RFP          : GDKTHWAVGVCRDSVSRK---GELTPLPETGYWRVRLWNGDKYAATTTPFTPLHIKVKPKRVG--IFLDYEAGT
BTN          : GDRTDWAIGVCRENVVKK---GFDPMTPDNGFWAVELYGNGYW-ALTPLRTSLRLAGPPRRVG--VFLDYDAGD
PRY-SPRY     : GDKGHWRVGWATKSVPRG---GFRLLGEDKGSWGYDGDGGKKYHNSEFPEYGLPFQEPGDVIGC--FLDLEAGT
Beta-SNTX    : G---YVGAGVTYKGIGRKTSTSDSSLGKNEKSWLFEYSTKSGYQQIHNSKKTRVTVSSTGFKLLGVYLDWPAGT
Alpha-SNTX   : G---YIDVGIAYISIPRKEIDFASAFGYNTYSWVLSYNPKIGYIERHKKREYNVRAPNPGFKRLGLFLDWRYGS
Enterophilin : NSCTGWDLGICRSNVMRK---ERTYIKPEDGFWAIRFYNHEYWALTSPKTQLTLKKPPGKVC---IFLDFEDQR
SOCS box     : QRGTHAVVGVATARAPLHSVGYTALVGSDSESWGWDLGRSRLYHDGKNRPGVAYPAFLGPDEAFAL-PDSLLVV

                                                                                 LDYE motif

                                                                            IDENTITIES (%)
Pro-ohanin   : VIFDLDGEVTTIQANFNGEEVVPFYYIGARVSLANL                             100  (100)
Thaicobrin   :                                                                  93   (93)
RFP          : LSFYNVTDRSHIYTFTDTFTEKLWPLFYPGIRAGRK..                            36   (51)
BTN          : ISFYNMSNGSLIFPYTSISFSGPLRPFFCLWSCGKK..                            31   (46)
PRY-SPRY     : ISFYKNGKYLGLAFFDVTFSGPLYPAVSLGNGGSV                               35   (43)
Beta-SNTX    : LSFYMVNKAWVTHLHTFHTKFNEAVYPAFLIGDAQQ..                            21   (38)
Alpha-SNTX   : ISFYAVSSDEVHHLHTFKTKFTEPVYPAFSIGPAGN..                            24   (36)
Enterophilin : ISFYNMTDNSHIYTFSQGAFYGSLRPFFRLWSKDSG..                            23   (34)
SOCS box     : LDMDEGTLSFIVDGQYLGVAFRGLKGKKLYPVVSAV..                            15   (22)
```

B

B30.2-like Domain-containing Families	Structural Organization	Examples
Ring-Bbox-Coiled coil-B30.2 subfamily	R – B – Coiled-coil – B30.2	Ring finger protein, TRIM5α, TRIM39, Sjögren's syndrome nuclear antigen A/Ro (SS-A/Ro)
Butyrophilin-B30.2 subfamily	IgV – IgC – Tm – B30.2	Btn1a1
Enterophilin-B30.2 subfamily	Ezrin/radixin/moesin – B30.2	Enterophilin-1
Suppressors of cytokine signaling (SOCS)-B30.2 Subfamily	(dotted) – B30.2 – SOCS	SPRY domain-containing SOCS box protein 2 (SSB-4)
Toxin-B30.2 Subfamily	(dotted) – B30.2	α- and β-subunits of stonustoxin
Vesspryns (Venom PRY-SPRY domain-containing Proteins)	(dotted) – B30.2	Ohanin/pro-hanin, Thai cobrin, Ohanin-like protein from *Naja naja atra*

Fig. 2. B30.2-like domain-containing proteins. (A) Sequence alignment of B30.2-like domains. The domains are from Pro-ohanin (gb: AY351433), Thaicobrin (sp: P82885), RFP (Ring finger protein, gb: NM172016), BTN (Butyrophilin, gb: NP038511), PRY-SPRY domains (conserved sequences of PRY-SPRY domains were provided by CDD data base), Beta subunit of SNTX (β-Stonustoxin, gb: Q91453), Alpha subunit of SNTX (α-Stonustoxin, gb: Q98989), Enterophilin (gb: AF126833) and SPRY domain-containing SOCS box protein 4 (Suppressors of cytokine signaling: NP660116). Identical and conserved residues are shaded black and grey. The Arg–Arg dibasic cleavage site of pro-ohanin is indicated in bold and propeptide segment is marked in italics. Three conserved LDP, WEVE and LDYE motifs as described by Henry et al. (13, 14) are boxed. Gaps (–) were introduced for optimal alignment and maximum homology for the sequences. The arrows indicate the boundary of PRY and SPRY domains. The numbers in parentheses represent the percentage of similarity between the B30.2-like domain of pro-ohanin with other B30.2-like domain-containing proteins. The substitutions among the following groups of amino acid residues are considered as conserved changes: Y, F, and W; S and T; V, L, I and M; H, R and K; D, E, N and Q; A and G. (B) Schematic representation of proteins possessing B30.2-like domain. B30.2-like domains are shaded black and the unidentified domains are dotted.

ARTICLE IN PRESS

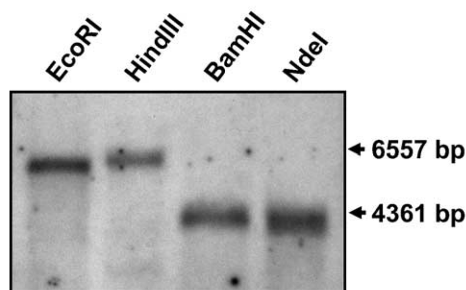

Fig. 3. Genomic Southern blot of ohanin. Genomic DNA of king cobra (10 μg each lane) was digested with *Eco*RI, *Hind*III, *BamH*I or *Nde*I enzymes. Southern hybridization shows the presence of one single band in all four digests. Thus ohanin is encoded by a single gene in the king cobra genome. The migration position of λ*Hind*III marker is indicated.

two clones were of type I and 14 clones were of type II. Our results demonstrate that type I cDNA has a longer 5′-UTR region (236 bp), whereas type II differed by a 76-bp deletion from position 54 to 129 (Fig. 1C; segment shaded black). Except for the missing segment in the 5′-UTR of type II cDNAs, no other sequence differences were observed between the two types of ohanin cDNAs.

The full-length cDNA (type I with 1558 bp) of ohanin (gb: AY351433) and its deduced amino acid sequence are shown in Fig. 1C. The cDNA (types I and II) encodes for a putative open reading frame (ORF) of 190 amino acid residues beginning with a start codon (ATG) at position 236 (type I numbering); ending with a termination codon (TAA) at position 808. The poly-adenylation signal, AATAAA, is located 14 nucleotides upstream of the poly-A tail (Proudfoot and Brownlee, 1976). The ORF is larger than the ohanin protein sequence we reported previously (Pung et al., 2005). It has a signal peptide of 20 amino acid residues, followed by 107 amino acids of mature ohanin and a 63 amino acid-long propeptide segment at the C-terminal (Fig. 1C). Thus ohanin is synthesized as a prepro-protein in the venom gland. We have named the precursor of ohanin as pro-ohanin. The cleavage at the dibasic Arg–Arg site probably results in the removal of C-terminal propeptide segment leading to mature ohanin (Fig. 1C).

3.2. Amino acid alignment of B30.2-like domains of pro-ohanin and other proteins

Comparison of the full-length cDNA sequence using the nucleotide BLAST algorithm (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997) did not display any significant sequence similarity with other nucleotide sequences deposited in the GenBank data base. This was surprising as the protein sequence showed weak similarity to B30.2-like domains (Pung et al., 2005). A Conserved Protein Domain Data base (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al., 2003) search, however, revealed that the deduced pro-ohanin sequence (from residue 9 to 170) shares an overall identity of 15% to 36% and similarity of 22% to 51% with PRY-SPRY domains (B30.2-like domains). Fig. 2A shows the alignment of B30.2-like domains of pro-ohanin with other B30.2-like domain-containing proteins.

The B30.2 domain, which maps within the B30.2 exon of the Human Class I Histocompatibility complex, is a conserved domain of 160 to 170 amino acid residues (Vernet et al., 1993; Henry et al., 1997a). One main characteristic of B30.2-like domains is the existence of three highly conserved motifs, LDP, WEVE, and LDYE (Henry et al., 1997b, 1998). It should be noted that mature ohanin has the first two motifs, whereas LDYE motif is found within the propeptide segment (Fig. 2A). This conserved domain, usually located at the C-terminal, is associated with N-terminal domains of different protein families (Fig. 2B). These protein families include RBCC proteins (RING-finger, B-box and Coiled-coil domain proteins) (Vernet et al., 1993; Orimo et al., 2000; Meyer et al., 2003), butyrophilin-related proteins (Ogg et al., 1996; Rhodes et al., 2001), as well as α- and β-subunits of stonustoxin (Ghadessy et al., 1996) as shown in Fig. 2B. In addition, SOCS box proteins (Suppressors Of Cytokine Signaling proteins) (Hilton et al., 1998; Yao et al., 2005) and enterophilin-related proteins (Gassama-Diagne et al., 2001) also contain the B30.2-like domain (Fig. 2B). The deduced pro-ohanin sequence, from residue 9 to 170, shows weak similarity to the complete B30.2-like domain proteins, suggesting that pro-ohanin may have been evolved from the same ancestral gene as the B30.2 domain. Unlike all other subfamilies, ohanin/pro-ohanin has an extremely short N-terminal segment of 8 residues. In addition to ohanin, two other proteins, namely Thaicobrin (sp: P82885) and an ohanin-like protein (*Naja naja atra*) (Li et al., 2004) from cobra venoms, have been identified so far in this subfamily. We named this new family of snake venom proteins as vesppryns (Venom PRY-SPRY domain-containing proteins).

3.3. Southern blot hybridization

As mentioned above, there are two mRNA subtypes for ohanin. It was of interest to determine whether these two

Fig. 4. Gene structure of ohanin. (A) Strategy used for cloning and sequencing of ohanin gene. The fragments obtained from gDNA PCR amplification are indicated by solid arrows; whereas the fragment from 'genome walking' is indicated by dashed arrow. Step 1: The region corresponding to the coding region of ohanin was amplified using the primers gDNAsigpep and gDNAstop. Step 2: The 3′-UTR was further amplified using the primers gDNA3UTR1 and gDNA3UTR2. Step 3: As for 5′-UTR, primer pairs used for 'genome walking' were gDNA5UTR1 and AP1 for primary amplification; gDNA5UTRnes2 and AP2 for secondary amplification. Steps 4 and 5: Primer pairs, 1-gDNA5UTR and 2-gDNA5UTR, as well as 9-gDNA5UTR and 6-gDNA5UTR, were used to obtain the remaining 5′-UTR region from the genomic DNA. (B) Ohanin gene sequence. Using both the genomic DNA PCR and 'genome walking' strategies, the full-length genomic sequence of 7086 bp was obtained. Exon–intron boundaries were determined based on cDNA and genomic sequences. Exons are shaded grey and indicated by upper case letters while introns are indicated by lower case letters. The missing exon in type II cDNA is shaded black. The three ATGs are indicated in bold; the putative signal peptide is underlined; dibasic processing site is boxed; propeptide segment is marked in italics, the stop codon is indicated by an asterisk and the polyadenylation signal, AATAAA, is underlined twice. (C) Genomic organization of ohanin. Ohanin gene comprises of five exons and four introns. Exons 1 to 5 have the sizes of 53, 76, 95, 96 and 1238 bp, respectively. The introns are 1160, 1743, 1292 to 1333 bp, respectively. In the case of alternative splicing, the whole exon 2 is excluded producing a shorter transcript of 1482 bp. The complete cDNA was named type I, while the shorter cDNA corresponding to the alternative splicing (missing exon 2) was named type II cDNA.

mRNAs are products of two independent genes or derived through alternative splicing. As a first step, we performed the genomic Southern hybridization experiment. King cobra genomic DNA was digested with *Eco*RI, *Hind*III, *Bam*HI and *Nde*I, separately. These genomic DNA digests were hybridized with a 297-bp DIG-labeled probe designed from nucleotide 319

A

(~1.8 kb) 4
(~1.9 kb) 1
(~1.4 kb) 5
(~1.65 kb) 3
(~750 bp) 2

B

```
GATCATTTGATCCAGAGAAGACACAGTCTCCTGGACTCATTATTGAAAAGAAGgtaagatcttacaacgcagctttcttccagaaatgctgattacttacccttgcttgtgaatcctgta   120
atagcttgggctgaagatagaatgtggaaaggatcttttattttccaagggctcaaaaagctttttgcggcagttgacacacaaagttctgcattttatcagaactcaaattttgaagt   240
aaaatgggaatatttctgcctttttacctcgttaactcctagaaaaacatattgaagaaaaaggctagactagagcattattcactatccatcagctcaattatttgttgtgaaatgaac   360
cttgtagagatcatcagtgattaaatgcttcatcagctgttggtgtatacatgccagatcattttcaggcataatcctgcacaataggtttttttttaaaaaactttattaaatttttatt   480
acattgaaaaacaacaaaacaaaaaacataaaaaacaaaaaaaacataagacataaaaacattgaaaaaaaacattaaaagtatatcactcaatgacaactacttgcggtgataattacc   600
taacaatcatatgtgcatgactaacaaatgttaacttgtgcataaatatatttttattctatctatatccaatatttactgattacctatctttattctgtctccgctctatccaactgt   720
accattgttcctatactttgtaaaatattgaatcttctttgtgtttaaactcaacgatagtttatccacttcggcacagctcataatttttttgtaagaccaattcatctgctggcatat   840
gtggctgtttccagcatagagcaaaagcaatcctagcagcagtcaatatatgtattaataaatatcttgtgctttttacatttttttaaagaattcccaatagaaagagttctggttcta   960
gatctaactgttcattagtaatttcctgaagccatttgtggatttgagccaatatttgtgagcttctgggcaagaccaccacatatggtagtatgttccctctttttttttttacacttc   1080
caacacatgagactcattttagaatacatttcgctaatctcgaagggggtaagtgccatctgtaaaacatttgtatgtgtttcttgtaagttactgattttgttaatttgtggttt   1200
ctattccatatagACTCCCAATCGGATCCTGCACAATAGTTTTATCTCCCTAGGGAAAAAAAAATATTCAAGAGAAAATTTGAATAAAGgtagacctcagcttatgaccatttgtttgaa   1320
gttatgctggcccccaaaaagagtttcttatgactcatctttgaagttatgtcacactaccctgtgttcatgtgactggaatgtgggcatttggcaaccagcttgcatttaagaggttg   1440
ccgtgtcctgcaatcatgtgactgtgatttgcaatcttcgcagccagcttctgatcaagcaaagtgaattggggaagccagatttgcttaatgacagtgtgattcatttaatgaccatgt   1560
aatttgcttaacaactgaggaagttacttaagagccgtggtggcacagtggttagaatgcagtactgcaggctactcctgccagcagttcggttctcaccagactcaaggttgactcag   1680
ccttccatccttccaaggtcggtaaaatgaggatccagactgttgggggcaatatgctagctctgtaaactgcttagagagggctgtaaagcagtataaagcagtatataagtctaagtg   1800
ctattgctattaatgaccacagtaaaattgatcataaaattgggtccagtcacatgatcatttacttagcaactgcatttgcttagtgactgaaattccggtcccaattaaattgagaac   1920
tatctgtagactcatcaacaagttgagatattggaattgagaactttgggactggtagtcacaagttgcggcttagattttgagtgattaaaataaactcttgagaaatgccattccccc   2040
acccccccaccccatttttcttctattggcatattcattattttcctctaaaaaagtggaataatgaggaaatgctaattagaatttataaaaccacagagataaacaataatgctgat   2160
tttttaaagtttaaactgattacttttgtttacacttaaaactgagttctgtaggaaacagacttccagcgaagaaaacccaatccaaagagaacctactggaattcaatttctcaccat   2280
cctgaactcaaaactataaagtgacaaaagccttccaagacctatcaaactgctgaaactgcactatgcacataccactgaaaacaggctgtatatggtactaaaaaaacacgcacacta   2400
ctttatgatgctcatgaaagcattgcatggcttcagcggtttgtgggctcaaggaaggagaatgtcatctgccattttgtgtgagatcatgttaccaggcagccatttgctgtttaagt   2520
actgtaacaagtcggccatctgccatttttgaaacaggcaagcccattaatatctaggggttgactttaaaagtcagatgttcttaaatgggggacttcctgcactaaacagataaaattc   2640
tagcagaagacttcgtaaactgttacacaaggtcttttggaacctgttgtgtggaattgtgcaacactttgcatgtccttctatgaaaacctgtgctgcaatccttcctgaaatgttggc   2760
tgagtgttcaaccaactcactgcaaaatagaaggacagttttggatatcaatccttttgtagttcacaaaagtcttcaaaatattatttattggtcactcacctgagaatgaactgtgtc   2880
agatgactcacagaaaaacagaataaaaaaatgatgacatggtataaaacagaatgaaaaaacgatgacattatatgtcatgctctggattcacttgaactaagtttttaaatccatttaa   3000
tcatttactcattttctctattccttctgtagGAAGCTAGAGTTTGTAATGGTCATGTCCCTTTCTGCTGGCTTCCAATTTAGTTTGAACTTCAACAACAAAGAAAGTCCTGCGGAAG   3120
CTCACAGgtaaagaaatatatatatatatatcatgattgtggtttgcttttgctctttcaacactttcagctctcttttgtgcaaaaatatcttggcaactccagtgtccccttttgatca   3240
tatggtgcacactttccccagtctgatcaaacactccccttccattaggtatgtttgctgccttctagttatatattttttaaaagatgggataaaaatataaagatatgcctttcata   3360
tgttgcaaaacatttttctttctgctttggtcatcactgagctatagcttataactatatgcttttttgctcatattttgttcatgtactggtaaagtcttgatcactatagattgctgggaa   3480
agacagatttatcgtagaagggaaaaggaaatatatcccattgtgcatttatcagccagtcagccagtcttctccaaatatctcctccaaacaaatgaaacagaatactatatttggcta   3600
accagggaacaaaagttagaacattcccaagacctgcaagacattataggcataaacaatagcactcagacctatataccacttcacagtgctttacagccctctccaagagtttacagt   3720
gtcagcatattgcccccaacaatcagagtcctcatttttaccgacctcagaaggatggaaggctgagtcaaccatgagctggtgagaatttgaacttgaactgctggcagtcggcagaattaacctc   3840
aatgctgcattctaaccactgtgccaccacaactcttgacttatgcccataattgaacccacaatttcagttgtaagtcatgacagtcattacacatcatatgattacacccgatcttat   3960
tacctttttgtagcaatcataaggcttacgtagtggttgtaaaaagaacactatggttgttaagcaaatgccacagttgttaaacaaacccattgtttgttatggggagtttttgccg   4080
gaagctggaagtagatgctggtttttgttaaaaatgccataaatggtggtcatgtgaccacgggacatggcaaaaaggtcatcaaaatgcactcatggccatgggcagagttgtggatct   4200
gggtggatgtatgtgtgtggatccagatattggtacttcaaatccaggtagtaagtagctttttggaatccattgtaagttgaggactacctgtacaggacatgtgacagcaagtgtta   4320
aattatgcccagtttttctgtaaaccaaactttaccatctataagacagtataaacaatgccttacctattcttgcctttgtgttctgtccatgtttagGAGTCTCTTGCATGCTCCTGT   4440
                                                                                                               M  L  L   (-18)

TCACACTATGCTTTTTTGCTGACCAGGAAAATGGTGGAAAGGCTCTGGCTTCACCTCCTGGGAATTGGCAGAAAGgtaagagatggcttaaataagagtgtccttctccaggatcagtcg   4560
F  T  L  C  F  F  A  D  Q  E  N  G  G  K  A  L  A  S  P  P  G  N  W  Q  K                                                  (8)

ggtcaggtgaacaataatgccagacaattacattgcaattaagtttggatttgctagatatttcctggcctcctttgtatgatccaacatgagtactcctaacatcagagtctcttgttg   4680
agatagaaatcaaataaataaatctattgtggatggaagggagaacagagtcatgtgaaatttaaccaacaaacattggctttcccagcttgcacctcactatagcttcagctgtgtggg   4800
cattttcctttattctgacactgtataagctgtcaggagtcatcttggtgagatgggtggcgtagaagtctaataataaataaatagagtcaaatgtggccttgagatgtataaggtctg   4920
ctaggctttgtcatctcagagcagtgtttatcaacctctgctagttggagtgcagaatagtttttaaatgggctagcaggatacaagaaacagtgcaacagtgaccccataactgtagg   5040
cagcccaagtctcctttttgaatggtcagaagcctttttgacaatgattcttgaaaactaaccgcctgtaatcaattacttgcaaaaaaaaagggtagtggtaaaaatgaaccatattg   5160
gagaactgaggatcacagcatttgttgcgtcatagcaatagcaatggcatttagacttatacaccactttacaccactctctaagtggttacagagtcagcatgctgcccccaacaatct   5280
gggtcctcattttaccaacctcagaaggatggaaggctcagtcaaccttgagcctggtgagaatcgaactccaggctgtgagcagagtttgcctacaataccccattctacccactgtgt   5400
caccacagctcatatcacatatacatgagggtgtgcctaaacatacattattttacattacatactgtacacacattattcttttgttcataaatgtgactatttgggagagaagtgtga   5520
gggatttttgtttctagaaagccacagtgcacctcttacctatcactccactttgacttgttttaggagagcagagttttacaacatcattgaagaagcagtcagcctttattgcacta   5640
tatatctttacataaagccctattaaaaagccttgtgtttatgcatcatataaatatgtagaagaatgcagacatgggaaatgaagccgggatatttttgtggcctttcttgcccctttaa   5760
tttgaggtgggtgggtaataacattttcagtgatgttgctctcctagagtttggtttggcttctgatgttctatctttctctctgcagCTGATGTGACGTTTGACTCAAACACAGCATTT   5880
                                                                                         A  D  V  T  F  D  S  N  T  A  F   (19)

GAATCACTGGTTGTGTCACCAGACAAGAAGACTGTGGAAAATGTGGGTGTCCCCAAGGGTGTGCCAGACAGTCCAGAGAGATTCAGTAGCAGCCCCTGTGTGCTGGGGTCTCCTGGATTC   6000
 E  S  L  V  V  S  P  D  K  K  T  V  E  N  V  G  V  P  K  G  V  P  D  S  P  E  R  F  S  S  S  P  C  V  L  G  S  P  G  F   (59)

CGATCAGGGAAGCATTTCTTTGAAGTGAAGTATGGGACACAAAGGGAATGGGCTGTTGGGTTAGCTGGGAAGTCTGTGAAGAGAAAAGGGTACCTCAGACTTGTCCCCGAAGAACGGATT   6120
 R  S  G  K  H  F  F  E  V  K  Y  G  T  Q  R  E  W  A  V  G  L  A  G  K  S  V  K  R  K  G  Y  L  R  L  V  P  E  E  R  I   (99)

TGGCAAAAGGGCCTCTGGTGGCTTCGGCGACTGGAAACGGATTCTGACAAGCTTCAAAAAGGCTCTGGAAAGATTATTGTCTTTCTGGATTACGATGAGGGAAAAGTGATTTTTGACCTG   6240
 W  Q  K  G  L  W  W  L [R  R] L  E  T  D  S  D  K  L  Q  K  G  S  G  K  I  I  V  F  L  D  Y  D  E  G  K  V  I  F  D  L   (139)

GATGGTGAAGTCACTACCATCCAGGCCAATTTCAATGGGGAGGAAGTTGTACCGTTTTACTATATAGGGGCACGTGTTTCACTCGCAAATCTTTAAGAGGTTACAATTCTTCATTAAAAC   6360
 D  G  E  V  T  T  I  Q  A  N  F  N  G  E  E  V  V  P  F  Y  Y  I  G  A  R  V  S  L  A  N  L  *                           (180)

AGGGGACTTCTCTCTACAGTCTTTGCAATGCCTGTTCAAGCATTTAATAATCTCTGGCTTTGGGAGAATAATCATGCACAGGTAGTGCTCGACTTACAACAGTTGGTTTAGTGGCTGTTT   6480
GAAGTTACAACGGCACTGGAAAATATAACTTATGACCATTTTCACACTTATGACCGTTGCAGCATTCCCATAGTCACGTGGTCAAAAAACAGTCACTTGACAACTGCCTTATATTTATGA   6600
CGGTTGCCGTGTCCCAGCATCATGTGATCAGCTTTTGTTACCTTCTGAGAAGCAAAATCAGTGGGGAAGCCAGATTCACTTAACAACCATGTTACTAACTTAACAACTACAGTGATTCAC   6720
TTAGCAACTGTGGCTAGAAAGGTCATAAAATCAGGCAAAACTCACTTAACAACTGTTTCACTTAACAACAGAATTTGGGGCTCAATTGTGCTCATAAGTCGAGGACTACCTGCATCATGT   6840
GTCCTTCTCCATCCCCAAACTGCCAACAGCATTTAAATTCCAGATGTATAAGGGTTGTTGCCATACACTTTACACTCCTTTGTTTAAACAATATTGCATCCTTTTCCCCTTAGACCCATA   6960
AAAGTTTCTTGAAGGGAATTACGGCAAGGCCATATTGAAGGTTGTCCCAGTAATATACCAAATCTGTTTAATCAATTGGACAAAAGGAATCATATCTGTACTACTAATAAAGTCTCACTG   7080
TTAGTA                                                                                                                   7086
```

(*continued on next page*)

ARTICLE IN PRESS

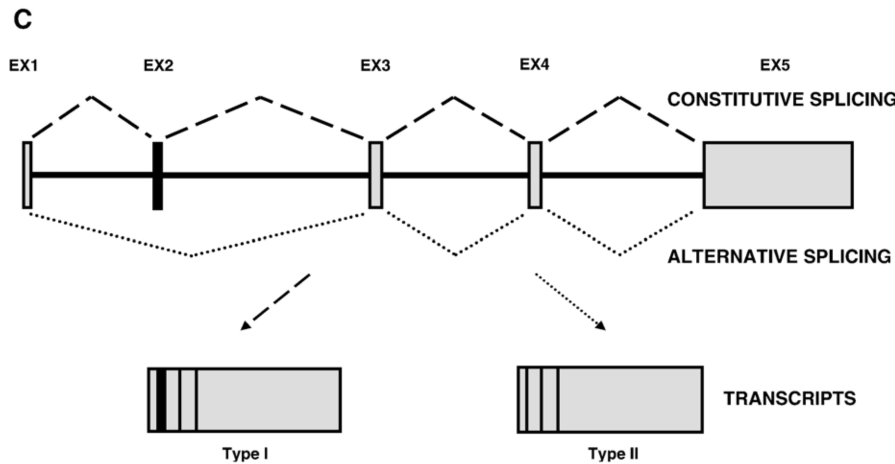

Fig. 4 (*continued*)

to 616 of its cDNA (Fig. 1C). We observed a single band in all four digests (Fig. 3), suggesting that ohanin is encoded by a single gene in the king cobra genome.

3.4. Cloning and sequencing of ohanin gene

To determine the genomic organization of ohanin gene, genomic DNA PCR and 'genome walking' approaches were used. Ohanin cDNA sequence (Fig. 1C) was used to map the exon–intron boundaries (Fig. 4). In the first amplification, gDNA-sigpep and gDNAstop were used to amplify its coding region. The resultant fragment was ~1.9 kb (Fig. 4A). Our attempts to PCR amplify the 3′-UTR region of the genomic DNA yielded another ~750 bp band as shown in Fig. 4A.

We tried to amplify the 5′-UTR region from the genomic DNA using primers designed from the transcription start site to the signal peptide region. However, no band was obtained after several attempts. This led us to suspect that our primers may have been interrupted by the presence of intron(s) or the thermal cycling profile used was still not optimal. Hence, genome walker libraries were constructed. As shown in Fig. 4A, the 'genome walk' was performed using antisense primers, gDNA5UTR1 and gDNA5UTRnes2 with adaptor primers (AP1 and AP2) from the kit. The resultant ~1.65 kb fragment was fully sequenced.

We obtained another ~1.8 kb further upstream by a gDNA PCR performed using primers 1-gDNA5UTR and 2-gDNA 5UTR designed from the 5′-region of the cDNA and the previously obtained ~1.65 kb genomic DNA fragment (Fig. 4A).

Table 1
The exon–intron boundaries of ohanin gene

Exon	Size (bp)	Intron	Size (bp)	5′-donor splice site	3′-acceptor splice site
1	53	1	1160	...AGAAG*gt*aag...	...tat*ag*ACTCC...
2	76	2	1743	...TAAAG*gt*aga...	...tgt*ag*GAAGC...
3	95	3	1292	...CACAG*gt*aaa...	...ttt*ag*GAGTC...
4	96	4	1333	...GAAAG*gt*aag...	...tgc*ag*CTGAT...
5	1238				

With the optimized thermal cycling profile, we further generated another fragment of ~1.4 kb corresponding to the transcription start site region of the cDNA using primers 9-gDNA5UTR and 6-gDNA5UTR (Fig. 4A).

Thus, we have obtained a total of 7086 bp of the gene sequence, spanning from 5′-UTR to 3′-UTR regions of ohanin cDNA (Fig. 4A and B). Sequences flanking the splice junctions were determined for all the exons and introns of ohanin (Table 1). The donor and acceptor splice sites of the exon–intron boundaries conform to the rule that intron begins with GT and ends with AG (Breathnach and Chambon, 1981). Ohanin gene contains five exons and four introns. Out of five exons identified, the coding region of ohanin is made up from two exons. Exons 1, 2 and 3 encode mainly the 5′-UTR region. Interestingly, exon 2 is spliced out in one of the mRNA subtypes (Figs. 1C and 4B and C). Exon 4 comprises of the remaining 5′-UTR region (11 bp), signal peptide and the first eight amino acid residues of ohanin. Exon 5 encodes for ohanin spanning from residues 9 to 107, the propeptide segment as well as the sequence corresponding to the 3′-UTR (Fig. 4B and C).

4. Discussion

Ohanin is a novel protein isolated from king cobra venom (Pung et al., 2005). It induces hypolocomotion and hyperalgesia in mice via both intraperitoneal and intracerebroventricular injection. Both the pharmacological actions are probably mediated through its effect on the central nervous system (Pung et al., 2005). This is a unique toxin which shows similarity to PRY-SPRY domains. cDNA cloning and sequencing show that ohanin is synthesized as a prepro-protein in the venom gland (Fig. 1). Further, there are two types of mRNA encoding ohanin; they differ in their 5-UTR. Interestingly, ohanin cDNA sequence shows no homology to other nucleotide sequences deposited in the GenBank data base including those of the B30.2-like domain-containing proteins. However, the deduced protein sequence shares weak similarity to the PRY-SPRY domains (Fig. 2). Unlike other subfamily members of B30.2 domain-containing

proteins, ohanin has a relatively short N-terminal extension (Fig. 2B). This new subfamily of venom proteins, vespryns, consists of ohanin (sp: P83234), Thaicobrin (sp: P82885) and an ohanin-like protein from *N. naja atra* (Li et al., 2004).

4.1. Analysis of ohanin cDNA

It should be noted that in addition to the start codon (AUG) at position 236, two other in-frame putative start codons are found further upstream in the same ORF at positions 146 and 152 (Fig. 1C). In general, the AUG closest to the 5′-end of an mRNA is the start signal for protein synthesis (Kozak, 1981, 1984). The context surrounding AUG is also crucial for the determination of the translation start site (Kozak, 1989). Based on the survey of 699 vertebrate mRNAs, GCC(A/G)CCaugG was identified as the most consensus sequence for eukaryotic translation start site (Kozak, 1987a,b). Kozak (1989) also highlighted the importance of purine at position −3 and G at position +4. The absence of purine at position −3 (usually an A) will impair the translation initiation more profoundly than any other nucleotides in the consensus sequence as it will lead to leaky scanning. In the absence of purine at position −3, however, G at the position +4 is essential for efficient translation (Kozak, 1989). A careful analysis at the three possible start codons revealed that none of them have the consensus sequence (Kozak, 1987a,b, 1989). Only the first AUG at position 146 fits well with the (−3, +4) rule with the presence of G at both positions. However, no suitable signal peptide required for its secretion into the venom can be identified with the first two AUG sites. Hence we propose that the third AUG at position 236 is the start codon for ohanin, although there is a possibility that the other two AUGs can be utilized as the start codon (see below).

Signal peptide is a hallmark of all secreted proteins. Since ohanin is secreted in the snake venom, it is expected to have a suitable signal peptide. Signal peptides generally have three common structurally and functionally distinct building blocks; a short, positively charged N-terminal (n-region) for its penetrability; a central hydrophobic region (h-region) that generally extends across the lipid head-groups of membranes and also helps in positioning the more polar C-terminal (c-region) in an exposed and extended conformation to be accessible for signal peptidase cleavage (Von Heijne, 1986, 1998). Based on the statistical analysis of 161 nonhomologous signal peptide sequences from the collection of 450 eukaryotic sequences, Von Heijne (1986, 1998) has suggested the (−3, −1) rule. According to this rule, the residue in the −3 position must not be aromatic (Phe, His, Tyr, Trp), charged (Asp, Glu, Lys, Arg) or large and polar (Asn, Gln). In addition, the residue in −1 position must be small (Ala, Ser, Gly, Cys, Thr or Gln). Recently determined crystal structure of *E. coli* signal peptidase (Paetzel et al., 1998) further confirms the importance of three building blocks and (−3, −1) rule of signal peptide as proposed by Von Heijne (1986, 1998).

With the three potential translation start sites, signal peptides with 50, 48 and 20 amino acid residues are possible. The signal peptide prediction software (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen et al., 2004) indicates that the cleavage site for prepro-ohanin is between Ala and Ser with all three AUGs as starting codons. Amino terminal sequencing by Edman degradation of ohanin (Pung et al., 2005) supports this predicted cleavage site. However, only the third AUG results in a 20-residue signal peptide, whereas the other two will generate unusually long signal peptides of 50- and 48-residue, respectively. Further the 20-residue signal peptide has the highest probability of 0.532 compared to 0.311 and 0.257 for 50- and 48-residue signal peptides. Since all toxin genes sequenced have 16- to 27-residue signal peptides (Smith, 1990), we propose that the third AUG acts as the start codon for ohanin. The putative 20-residue signal peptide has all the distinctive structural features; a basic n-region, a central hydrophobic h-region and a more polar c-region. In addition, Ala residues at position −3 and −1 also conform to Von Heijne's (−3, −1) rule.

4.2. Analysis of ohanin gene

Despite the presence of two types of mRNAs, there is only a single gene which encodes ohanin as indicated by Southern hybridization (Fig. 3). This gene has five exons and four introns. Genomic DNA sequence analysis indicates that the heterogeneity at the 5′-UTR of the two types of mRNAs are due to the occurrence of alternative splicing from a single gene (Fig. 4B). About half of all mammalian genes are estimated to have more than one splice forms (Mironov et al., 1999; Brett et al., 2000; Croft et al., 2000; Kan et al., 2001; Okazaki et al., 2002).

So far five major forms of alternative splicing have been identified. They consist of exon-skipping, alternative 3′ splice-site, alternative 5′ splice-site, mutually exclusive exons and intron retention (Ast, 2004). Analysis of ohanin genomic DNA indicates that the splicing pattern belongs to exon-skipping as the missing segment in type II cDNA is encoded solely by exon 2 (Fig. 4). The implication of alternative splicing of ohanin gene at the 5′-UTR is not clear at this time. However, alternative splicing at the 5′-UTR is probably a mechanism that allows for the use of several differently regulated promoters for the same gene as demonstrated by most of the splicing events in mammalian genes (Mironov et al., 1999).

In this study, we have also shown that a single intron is inserted right before residue 9 of ohanin, leaving the last exon coding for PRY-SPRY domains (B30.2-like domain) intact (Fig. 4B and C). Interestingly, at the nucleotide level, this 'B30.2 exon' does not show sequence similarity to either coding or 3′-UTR regions. Similar genomic organization among ohanin and other B30.2-like domain-containing proteins indicates these proteins may have evolved from the same ancestral gene.

4.3. Molecular evolution of ohanin/pro-ohanin

It is interesting to note that all the snake venom proteins reported so far are highly similar to non-venomous body proteins. For example, sarafotoxins are similar to endothelins, Cobra venom factors are similar to complement C3, and three-finger toxins are similar to proteins such as CD59. Thus, it has been postulated that snake toxins arise from recruitment events of genes from various body protein families during evolution (Fry, 2005). Under positive selection pressure, B30.2 domain

may have been selected to duplicate with MHC class I gene. Subsequently, this domain has been recruited in the venom gland. The basis for the multiple and independent recruitment events of certain protein families for the use as toxins are unclear. However, it is hypothesized that the chosen protein families are likely to be favored in the snake's adaptive evolution, particularly for its feeding habits. These protein families must also be beneficial or 'economical' for use as stable molecular scaffolds to incorporate various functional motifs on their surface to exert different pharmacological actions.

4.4. Functional implication(s) of B30.2-like domain

The B30.2-like domains are the only common domains found in transmembrane proteins (BTNs), intracellular proteins (RFPs, SSA/Ro) and secreted proteins (ohanin/pro-ohanin, Thaicobrin, α- and β-subunits SNTX) (Fig. 2B) (Henry et al., 1997b, 1998). However, the biological implication(s) of this C-terminal domain in the diverse functionally and structurally unrelated protein families is not known. Recently, TRIM5α, originally identified from rhesus monkey, was found to confer resistance to retroviral infection (HIV type 1) (Sawyer et al., 2005). Functional studies using chimeric TRIM5α have further shown that a 13 amino acid residue 'patch' in the SPRY domain of TRIM5α is responsible for the inhibitory effect (Sawyer et al., 2005). Mature ohanin with an 8-amino acid residue N-terminal segment, followed by PRY and partial SPRY domains, induces hypolocomotion and hyperalgesia in mice (Pung et al., 2005). Thus PRY-SPRY domains are sufficient for its biological activity. Currently, we are studying the mechanism of action and structure–function relationships of ohanin.

Acknowledgement

We thank Francis Tan Chee Kuan and Carol Han Ping from the Department of Biological Sciences, National University of Singapore for their invaluable discussions and technical help. YFP and NR are the recipients of research scholarships from the National University of Singapore.

References

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

Ast, G., 2004. How did alternative splicing evolve? Nat. Rev. Genet. 5, 773–782.

Bendtsen, J.D., Nielsen, H., Von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: signalP 3.0. J. Mol. Biol. 340, 783–795.

Breathnach, R., Chambon, P., 1981. Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349–383.

Brett, D., et al., 2000. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett. 474, 83–86.

Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., Mattick, J.S., 2000. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat. Genet. 24, 340–341.

Fry, B.G., 2005. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. 15, 403–420.

Gassama-Diagne, A., et al., 2001. Enterophilins, a new family of leucine zipper proteins bearing a B30.2 domain and associated with enterocyte differentiation. J. Biol. Chem. 276, 18352–18360.

Ghadessy, F.J., et al., 1996. Stonustoxin is a novel lethal factor from stonefish (*Synanceja horrida*) venom: cDNA cloning and characterization. J. Biol. Chem. 271, 25575–25581.

Henry, J., et al., 1997a. Cloning, structural analysis, and mapping of the B30 and B7 multigenic families to the major histocompatibility complex (MHC) and other chromosomal regions. Immunogenetics 46, 383–395.

Henry, J., Ribouchon, M.T., Offer, C., Pontarotti, P., 1997b. B30.2-like domain proteins: a growing family. Biochem. Biophys. Res. Commun. 235, 162–165.

Henry, J., Mather, I.H., McDermott, M.F., Pontarotti, P., 1998. B30.2-like domain proteins: update and new insights into a rapidly expanding family of proteins. Mol. Biol. Evol. 15, 1696–1705.

Hilton, D.J., et al., 1998. Twenty proteins containing a C-terminal SOCS box form five structural classes. Proc. Natl. Acad. Sci. U. S. A. 95, 114–119.

Kan, Z., Rouchka, E.C., Gish, W.R., States, D.J., 2001. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res. 11, 889–900.

Karlsson, E., 1979. Chemistry of protein toxins in snake venoms. In: Lee, C.Y. (Ed.), Snake Venoms. Springer-Verlag, NY, pp. 159–212.

Kini, R.M., 2002. Molecular moulds with multiple missions: functional sites in three-finger toxins. Clin. Exp. Pharmacol. Physiol. 29, 815–822.

Kozak, M., 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 9, 5233–5252.

Kozak, M., 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12, 857–872.

Kozak, M., 1987a. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.

Kozak, M., 1987b. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. 196, 947–950.

Kozak, M., 1989. The scanning model for translation: an update. J. Cell Biol. 108, 229–241.

Li, S., et al., 2004. Proteomic characterization of two snake venoms: *Naja naja atra* and *Agkistrodon halys*. Biochem. J. 15, 119–127.

Marchler-Bauer, A., et al., 2003. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res. 31, 383–387.

Meyer, M., Gaudieri, S., Rhodes, D.A., Trowsdale, J., 2003. Cluster of *TRIM* genes in the human MHC class I region sharing the B30.2 domain. Tissue Antigens 61, 63–71.

Mironov, A.A., Fickett, J.W., Gelfand, M.S., 1999. Frequent alternative splicing of human genes. Genome Res. 9, 1288–1293.

Mochca-Morales, J., Martin, B.M., Possani, L.D., 1990. Isolation and characterization of helothermine, a novel toxin from *Heloderma horridum horridum* (Mexican beaded lizard) venom. Toxicon 28, 299–309.

Ogg, S.L., Komaragiri, M.V., Mather, I.H., 1996. Structural organization and mammary-specific expression of the butyrophilin gene. Mamm. Genome 7, 900–905.

Okazaki, Y., et al., 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573.

Orimo, A., et al., 2000. Molecular cloning of testis-abundant finger protein/ring finger protein 23 (RNF23), a novel RING-B box-coiled coil-B30.2 protein on the class I region of the human MHC. Biochem. Biophys. Res. Commun. 276, 45–51.

Paetzel, M., Dalbey, R.E., Strynadka, N.C.J., 1998. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. Nature 396, 186–190.

Proudfoot, N.J., Brownlee, G.G., 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211–214.

Pung, Y.F., Wong, P.T.H., Kumar, P.P., Hodgson, W.C., Kini, R.M., 2005. Ohanin, a novel protein from king cobra venom induces hypolocomotion and hyperalgesia in mice. J. Biol. Chem. 280, 13137–13147.

Rhodes, D.A., Stammers, M., Malcherek, G., Beck, S., Trowsdale, J., 2001. The cluster of *BTN* genes in the extended major histocompatibility complex. Genomics 71, 351–362.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

Sawyer, S.L., Wu, L., Emerman, M., Malik, H.S., 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. U. S. A. 102, 2832–2837.

ARTICLE IN PRESS

Smith, L.A., 1990. Cloning, characterization and expression of animal toxin genes for vaccine development. J. Toxicol., Toxin Rev. 9, 243–283.

Torres, A.M., Wong, H.Y., Desai, M., Moochhala, S., Kuchel, P.W., Kini, R.M., 2003. Identification of a novel family of proteins in snake venoms: purification and structural characterization of nawaprin from *Naja nigricollis* snake venom. J. Biol. Chem. 278, 40097–40104.

Vernet, C., et al., 1993. Evolutionary study of multigenic families mapping close to the human MHC class I region. J. Mol. Evol. 37, 600–612.

Von Heijne, G., 1986. A new method for predicting signal sequence cleavage site. Nucleic Acids Res. 14, 4683–4690.

Von Heijne, G., 1998. Life and death of a signal peptide. Nature 396, 111–112.

Yamazaki, Y., Hyodo, F., Morita, T., 2003. Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch. Biochem. Biophys. 412, 133–141.

Yao, S., et al., 2005. Backbone 1H, 13C and 15N assignments of the 25 kDa SPRY domain-containing SOCS box protein 2 (SSB-2). J. Biomol. NMR 31, 69–70.