The Gene Structure and Evolution of ku-wap-fusin (Kunitz Waprin Fusion Protein), a Novel Evolutionary Intermediate of the Kunitz Serine Protease Inhibitors and Waprins from *Sistrurus catenatus* (Massasauga Rattlesnake) Venom Glands

Robin Doley[1,2], Susanta Pahari[1,3], Md Abu Reza[1], Stephen P. Mackessy[4] and R. Manjunatha Kini*[,1,5]

[1]*Protein Science Laboratory, Department of Biological Sciences, National University of Singapore, Singapore 117543*

[2]*Department of Molecular Biology and Biotechnology, Tezpur University, Napaam-784 028, Tezpur, Assam, India*

[3]*CMR Institute of Technology, 132 AECS Layout, IT Park Road, Bangalore 560 037, India*

[4]*School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639-0017, USA*

[5]*Department of Biochemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0614, USA*

Abstract: Snake venom proteins belong to various families which differ in their primary and secondary structures. Recently, we found a novel transcript in the venom gland of *Sistrurus catenatus edwardsii* which was named "Ku-wap-fusin" because it encodes Kunitz-type serine protease inhibitors (Kunitz SPIs) and whey acidic protein (WAP) domains in tandem. We have now determined the gene structure of this unique transcript (from *S. c. edwardsii* and *S. c. tergeminus*) and analyzed its evolutionary relationships with genes encoding Kunitz-type serine protease inhibitors and waprins. Structurally, Ku-wap-fusin has four exons and three introns. Comparison of all gene structures (Kunitz SPI, WAP and fusin) shows that exon I in all of these genes is highly conserved and codes for the signal peptide. However, in the Ku-wap-fusin gene, there is an insertion of an exon, which codes for the WAP domain, in the intron II region of the Kunitz SPI gene. In the waprin gene, the exon II, encoding the Kunitz domain, has been lost. Ku-wap-fusin and waprin (whey acidic protein related proteins) genes therefore appear to have evolved *via* insertion/deletion of an exon. We propose a model for the evolution of these genes (using their gene structures) in which Kunitz, Ku-wap-fusin and waprin genes have evolved from a common ancestor. During the evolution of this protein family complex, the Ku-wap-fusin gene is likely an intermediate between the ancestral and the waprin genes.

Keywords: Kunitz-type serine protease inhibitor, Waprins, new exon insertion, toxin evolution Sistrurus catenatus.

1. INTRODUCTION

Natural products are an excellent source of pharmacologically active molecules, and snake venoms may contain 100 or more different proteins and peptides [1-3]. These proteins and polypeptides induce various pharmacological effects, including neurotoxicity, cardiotoxicity, pro-coagulant and/or anticoagulant effects, tissue necrosis and even death in the prey or victim. Over the past few decades, a plethora of protein toxins have been isolated and characterized from snake venoms [4]. Several studies have shown that the composition of venom varies among different snake species, and inter- and intraspecific variation resulting from an interplay of several different selective factors can produce venoms with different pharmacologies [5,6]. Many of these proteins are encoded by multigene families and contain functionally diverse isoforms. Venom proteins and polypeptides have evolved from their ancestral precursors through a process of gene duplication followed by accelerated point mutations in the protein coding region, and hence the crude venoms often contain multiple isoforms (for a review see [7]). This adaptive mode of protein evolution has been reported in toxin superfamilies such as phospholipases A_2 [8-12], three-finger toxins [13-15], serine proteases [16] and serine protease inhibitors (SPIs) [17]. However, in addition to gene duplication and accelerated evolution [7,9,10,16,18], insertion of new exons [19], extension of intron-exon boundary [20] and exchange of exon segments [21,22] also play an important role in the evolution of new genes with novel functions.

Most snake venom SPIs are structurally similar to members of the Kunitz/BPTI inhibitor family. They have approximately 60 amino acid residues, with a conserved fold stabilized by three disulphide bridges [23], and functionally they can be divided into either non-neurotoxic or neurotoxic homologs. Non-neurotoxic homologs inhibit either trypsin or chymotrypsin, while neurotoxic homologs act as calcium and potassium channel blockers which do not have protease inhibitory activity [17,24,25]. Non-neurotoxic Kunitz SPI tar-

*Address correspondence to this author at the Protein Science Laboratory, Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore; Tel: +65 6516 5235; Fax: +65 6779 2486;
E-mail: dbskinim@nus.edu.sg

1874-4044/10

get serine proteases through their reactive site (P1) [26]. In trypsin inhibitors, the P1 residue is positively charged and is either Lys or Arg, whereas in chymotrypsin inhibitors, large hydrophobic residues such as Leu, Phe or Tyr are present [26]. Dendrotoxin, calcicludine and the B chain of β-bungarotoxin are neurotoxic Kunitz SPI homologs and act as Ca^{2+} and K^{+} channel blockers [27-30]. Analysis of nucleotide sequences of Kunitz SPIs reveals that they are also encoded by a multigene family and have evolved through gene duplication and positive selection, resulting in multiple isoforms with different biological activities [17,23,31].

Recently we identified a new family of snake venom proteins which shows homology to whey acidic proteins and hence were named waprins (whey acidic protein related proteins) [32]. Since then, the presence of waprins has been reported from venoms of several Australian elapids as well as colubrid snake venoms [33-35]. Members of the waprin family are ~50 amino acid residue proteins with a core domain containing four disulphide bridges [32]. Functions of waprins have not been elucidated except for omwaprin, which shows selective antimicrobial properties [33].

Previously, we reported a novel fused gene from the cDNA library of *Sistrurus catenatus edwardsii* [36], and the putative mature protein from this novel transcript is 152 amino acids long. A comparison of this transcript with available sequences revealed that it is a hetero-multimeric multidomain protein with two structural domains. Its signal peptide and N-terminal domain are similar to Kunitz SPI, whereas it has a waprin domain at the C-terminal with an extended tail (Fig. **1**). The putative mature protein is cysteine-rich and contains 14 cysteine residues; the first six cysteine positions are located at positions identical with Kunitz SPI and β–bungarotoxin chain B, whereas the other six cysteine positions are identical to waprins. The 8th and 13th cysteines (corresponding to the 2nd and 7th of waprins, which form a disulphide bridge) are missing in this fused toxin.

The gene structures of both Kunitz SPIs and waprins have been reported recently [35,37,38]. Structurally, they have three exons interrupted by two introns, and comparison between these two sets of genes revealed that they may have evolved from a common ancestor. Since the fused gene contains the domains of these two proteins in a single transcript, in the present study we determined its gene structure to obtain a better understanding of their evolutionary relationship within a single species (*Sistrurus catenatus edwardsii*). We have also confirmed the presence of a homologous fused gene transcript, and determined the complete gene structure, in a related subspecies of *S. catenatus* (*S. c. tergeminus*). A detailed analysis of these gene structures reveal that waprins may have evolved from the same ancestor as Kunitz SPI, as recently reported [35], but that it occurred through this fused gene as an evolutionary intermediate.

2. MATERIAL AND METHODS

2.1. Materials

Snakes were collected from Lincoln County, CO, USA (*S. c. edwardsii*) and from Barton County, KS, USA (*S. c. tergeminus*) under permits from the Colorado Division of Wildlife and Kansas Wildlife and Parks. Extraction of venom glands and liver from *Sistrurus catenatus tergeminus* followed the methods reported previously for *S. c. edwardsii* [36]. RNeasy® RNA isolation kit, DNeasy® DNA extraction kit, Qiaquick gel extraction and PCR purification kit were purchased from Qiagen (Valencia, CA, USA). SMART RACE cDNA amplification kit was purchased from Clontech Laboratories Inc (Palo Alto, CA, USA). pGEM®-T Easy vector system was purchased from Promega Corporation (Madison, WI, USA). ExprepTM plasmid extraction kit was purchased from GeneAll Biotechnology (Seoul, Korea). Long PCR enzyme mix was obtained from Fermentas (Hanover, MD, USA). The ABI PRISM® BigDye® terminator cycle sequencing ready reaction kit was purchased from Perkin-Elmer (Foster City, CA, USA). Oligonucleotides were custom-synthesized from 1st BASE (Singapore). All other chemicals and reagents were of the purest grade available.

2.2. RNA Isolation

Total RNA was isolated from the venom gland tissues using the Qiagen RNeasy kit with on-column DNase digestion according to the manufacturer's instructions. The purity of RNA was verified by 1% agarose gel electrophoresis (stained with Biotium GelRed™, Hayward, CA, USA) and by checking the optical density by NanoDrop 1000 (Thermo Scientific, Wilmington, DE, USA).

2.3. Reverse Transcription and Polymerase Chain Reaction

Total RNA (100 ng) from *S. catenatus edwardsii* venom gland was reverse transcribed using the reverse transcriptase enzyme provided in the SMART RACE cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA, USA). Sense (FuG-F: 5′-ATGTCTTCTGGAGGTCTTCTGC-3′) and antisense (FuG-R: 5′-TCAAGTTGGAACACTAGATCC-3′) primers were designed from the cDNA sequence of the *S. catenatus edwardsii* fused gene transcript (DQ464286) reported in our previous study [36]. PCR was carried out in a final volume of 25 µl using 1 µl of reverse transcribed cDNA. PCR products were subjected to 1% agarose gel electrophoresis, visualized by Biotium GelRed™ staining and purified using a Qiagen PCR purification kit according to the manufacturer's instruction.

2.4. 5'- and 3'- Rapid Amplification of cDNA Ends (RACE)

5′ and 3′ RACE reactions were performed to identify the existence of mRNA containing either the Kunitz/BPTI domain or the WAP domain in *S. catenatus edwardsii* venom glands. Both 5'- and 3'-RACE libraries were created and amplifications were carried out with Clontech's SMART RACE cDNA amplification kit. Universal primer mix (UPM: 5′-CTAATACGACTCACTATAGGGCAAGCAGTGGTATC AACGCAGAGT-3′; and 5′-CTAATACGACTCACTATA GGGC-3′) from the SMART RACE kit and gene-specific oligonucleotides were used for amplification. The gene-specific oligonucleotides (5′GSP1: 5′-GAGGGCATCCT TGCACTCAGTCTTG-3′; and 3′GSP1: 5′-CCTGCTGAAA CCGGAGAGTGTAGAGG-3′) were designed based on the sequence data obtained from RT-PCR product from *S. catenatus edwardsii*. The 5'-RACE gene-specific reverse primer was designed from the WAP domain (exon III) and the

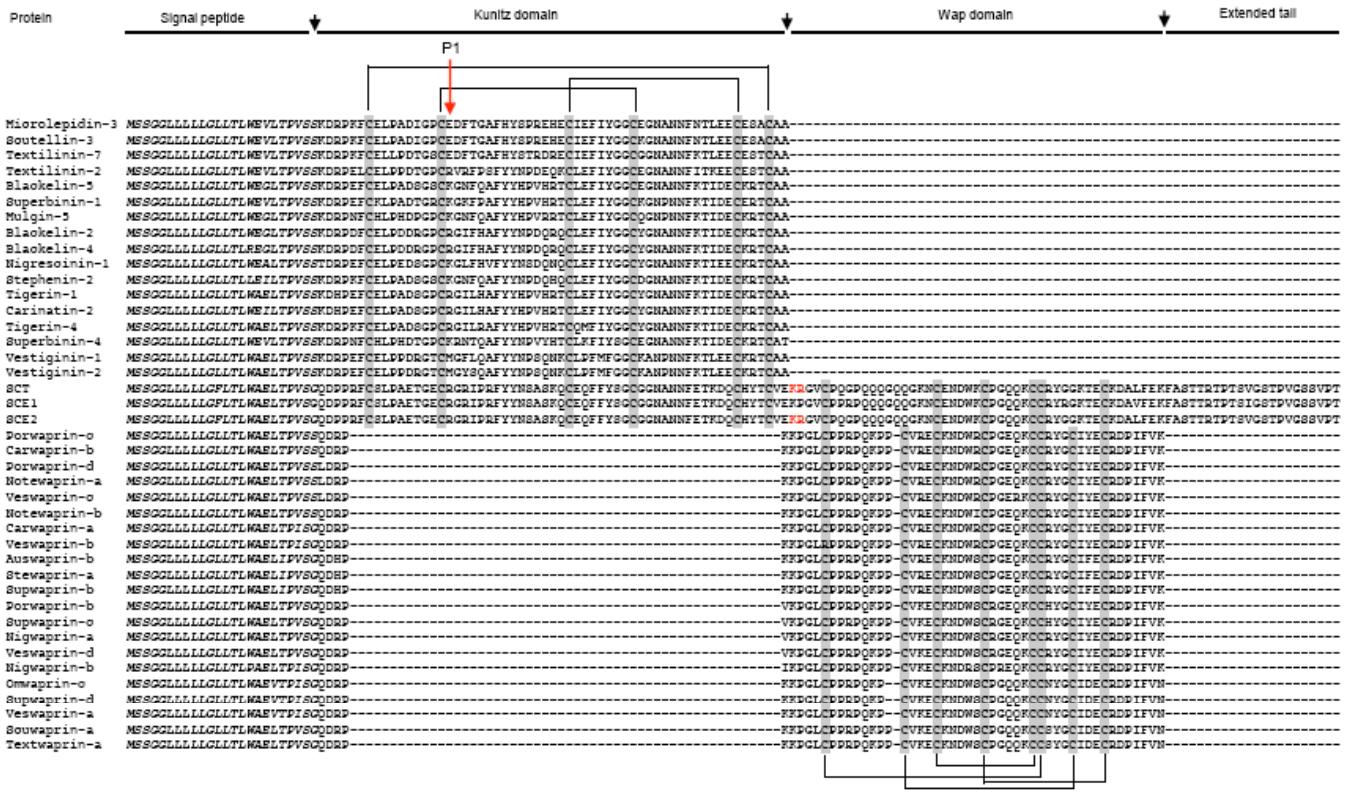

Fig. (1). Alignment of amino sequences of Kunitz-type serine protease inhibitors, Ku-wap-fusins and waprins. The amino acid sequences were either directly obtained or translated from gene sequences obtained from the NCBI database. Kunitz-type serine protease inhibitors used in the alignment are Microlepidin-3 (EU401843) from *Oxyuranus microlepidotus*, Scutellin-5 (EU401841) and Scutellin-3 (ACC77789) from *Oxyuranus scutellatus*, Textilinin-7 (EU401839) and Textilinin-2 (EU401838) from *Pseudonaja textilis textilis*, Blackelin-2 (EU401847), Blackelin-4 (EU401848) and Blackelin-5 (EU401849) from *Pseudechis porphyriacus*, Superbinin-1 (EU401856) and Superbinin-4 (EU401857) from *Austrelaps superbus*, Mulgin-5 (EU401846) from *Pseudechis australis*, Nigrescinin-1 (EU401850) from *Rhinoplocephalus nigrescens*, Stephenin-2 (EU401855) from *Hoplocephalus stephensii*, Carinatin-2 (EU401853) from *Tropidechis carinatus*, Tigerin-1 (EU401851) and Tigerin-4 (EU401852) from *Notechis scutatus*, Vestiginin-1 (EU401858) and Vestiginin-2 (EU401859) from *Demansia vestigiata*; Ku-wap-fusins are SCE1 (DQ464286) and SCE2 (FJ746496) from *Sistrurus catenatus edwardsii* and SCT (FJ686529) from *Sistrurus catenatus tergeminus*; Waprins are Porwaprin-b (EU401820), Porwaprin-C (EU401821) and Porwaprin-d (EU401822) from *Pseudechis porphyriacus*, Carwaprin-a (EU401827) and Carwaprin-b (EU401828) from *Tropidechis carinatus*, Notewaprin-a (EU401825) and Notewaprin-b (EU401826) from *Notechis scutatus*, Veswaprin-a (EU401836), Veswaprin-b (EU401833), Veswaprin-c (EU401834) and Veswaprin-d (EU4018) from *Demansia vestigiata*, Auswaprin-b (EU401819) from *Pseudechis australis*, Stewaprin-a (EU401829) from *Hoplocephalus stephensii*, Scuwaprin-a (EU401818), Supwaprin-b (EU401830) Supwaprin-c (EU401831) and Supwaprin-d (EU401835) from *Austrelaps superbus*, Nigwaprin-a (EU401823) and Nigwaprin-b (EU4018) from *Rhinoplocephalus nigrescens*, Omwaprin-c (EU401817) from *Oxyuranus microlepidotus*, and Textwaprin-a (EU401837) from *Pseudonaja textilis*. Sequence gaps are represented by "-". The predicted signal peptide is shown in italics, the cysteine residues are highlighted and the disulphide bond pattern is indicated for each domain. The signal peptide, the Kunitz domain, the waprin domain and the extended C-terminal end of Ku-wap-fusins are marked with solid lines. The position of the P1 residue in Kunitz-type serine protease inhibitors is marked with a red arrow, and the dibasic site in ku-wap-fusins SCE2 and SCT is highlighted with red letters.

3'-RACE gene-specific forward primer was designed from the Kunitz/BPTI domain (exon II). Primers were designed based on the requirements for touchdown PCR according to manufacturer's instruction.

Following the initial RACE reactions, 5 µl of the PCR product from each reaction was analyzed on a 1% agarose gel. A prominent band with a number of background bands was observed in both the 5′- and 3′-RACE reactions. The remaining PCR products from both RACE reactions were purified and used as templates for PCR re-amplification reactions. A second set of nested 5' and 3' gene-specific primers were designed several bases upstream and downstream, respectively, from the 5' GSP1 (5' GSP2: 5′-CTTCTGCT GCCCAGGACATTTCCAG-3′) and 3' GSP1 (3' GSP2: 5′-TTCCTCGTTTCTACTACAACTCGGCTTCAA-3′ oligonucleotides, respectively. These nested gene-specific RACE primers and nested universal primer (Nested UPM: 5′-AAGCAGTGGTATCAACGCAGAGT-3′) from the SMART RACE cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA, USA) were used for nested PCR amplification. PCR products were cloned into pGEM®-T Easy vectors. At least 8 clones for each of the 5′- and 3′-RACE products from *S. c. edwardsii* were sequenced from both sides.

2.5. Genomic DNA Extraction

Genomic DNA was extracted from *S. c. edwardsii* and *S. c. tergeminus* liver tissue using the DNeasy® Tissue kit

(Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. For each extraction, 30 mg liver tissue was used. Integrity of the isolated genomic DNA extracted was examined by 0.8% agarose gel electrophoresis and quantified by optical density measurements.

2.6. Isolation and Sequencing of Genomic Clones

gDNA PCR was carried out to amplify the fused gene from *S. c. edwardsii* and *S. c. tergeminus* from start to stop codon using the same FuG-F and FuG-R primers mentioned earlier. The PCR reaction mix contained 0.25-0.50 µg DNA as template and a final concentration of 1.5 U Fermentas long PCR enzyme, 1X PCR buffer, 15 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 for 2 min followed by 34 amplification cycles (94 °C for 1 min, annealing temperature of 60 °C for 40 s (depending on the primer Tm), extension at 68 °C for 3 min) and a final extension of 68 °C for 10 min. PCR products were fractionated on a 1.0 % agarose gel and the band of interest was excised, purified and ligated into pGEM®-T Easy vector. At least sixteen clones carrying the inserts were sequenced.

2.7. DNA Sequencing

All DNA sequencing reactions were performed using the ABI PRISM BigDye terminator cycle sequencing ready reaction kit according to manufacturer's instructions. DNA sequencing was carried out using an ABI PRISM® 3100 automated DNA sequencer.

2.8. Sequence Analysis and Phylogenetic Tree Construction

Sequence analysis was carried out using the BLASTn program at the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov). All venom-derived proteins showing homology with either the Kunitz SPI domain or the WAP domain of the fused toxins were used in these analyses. The intron-exon boundaries were determined by comparing the sequences of respective cDNAs and putative genes. Multiple protein sequence alignment was done using ClustalX v. 1.81 [39]. A Bayesian analysis of aligned sequences was performed with MrBayes v. 3.04b [40] using a variable rate of amino acid mutation model (equalin) and coupled Markov chain Monte Carlo (MCMC) sampling. The variable rate model was chosen because it resolved several polytomies obtained using the fixed rate model, but otherwise the tree topologies were essentially identical. A total of 10^6 replications were run, with a burn-in of 0.25. Tree branches were proportionally transformed for clarity using FigTree v. 1.2.1 (available at http://tree.bio.ed.ac.uk/software/figtree/).

2.9. Molecular Modeling

The three dimensional structures of ku-wap-fusin from Desert Massasauga (*Sistrurus catenatus edwardsii*) were predicted using the online I-TASSER server for protein 3D structure prediction [41]. The server predicts the folds and secondary structure by profile-profile alignment (PPA) threading techniques. Five models were obtained and the model with the lowest free energy was used for further analysis. Ribbon structure diagrams were created using the DS ViewerPro software for analysis.

3. RESULTS AND DISCUSSION

3.1. Gene Structure of Fused Toxin

Previously we reported the presence of a unique transcript containing both Kunitz and WAP domains from the venom glands of the Desert Massasauga Rattlesnake, *S. c. edwardsii* [36], and we subsequently identified a similar transcript from *S. c. tergeminus*, a closely related subspecies. Based on these observations, we hypothesized that these transcripts may have evolved through a fusion of modules from two individual genes encoding Kunitz and WAP domains separately [36]. Consequently, we generated 5′and 3′-RACE libraries to analyze for the presence of individual toxin mRNAs, to determine if these two genes exist individually in the *S. c. edwardsii* venom glands. In RACE libraries of venom glands, only the full length transcripts with both Kunitz and WAP domains in tandem were obtained (data not shown). Two isoforms, SCE1 (7 clones) [GenBank: DQ464286] and SCE2 (9 clones) of full length transcripts were obtained by both 3′ and 5′ RACE, and the newly identified isoform (SCE2) was submitted to GenBank database [FJ746496]. The mature proteins of these two isoforms SCE1 and SCE2 differ by seven amino acid residues (Fig. **1**), each due to single base substitution in the nucleotide sequence; the amino acid sequences of SCT and SCE2 are identical. Thus, this species contains only the fused transcripts, and the protein encoded by this transcript was named as "Ku-wap-fusin" (Kunitz and waprin fusion protein). The completed sequences of *S. c. edwardsii* and *S. c. tergeminus* Ku-wap-fusin genes were obtained using gDNA PCR and were submitted to GenBank database under accession numbers FJ686530 and FJ686529 respectively. The total nucleotide sequences (from start to stop codon) in these two genes were found to be 2899 bp and 2894 bp respectively (Fig. **2A**) and they show 98.2% identity. The exon and intron regions were identified by comparing the cDNA and gene sequences, following the GT-AG rule of splice-donor and -acceptor sites [42]. The fused genes consist of four exons interrupted by three introns (Fig. **2A**). Such multi-domain proteins containing Kunitz SPI and WAP domains have been reported from different organisms. For example, human WFIKKN protein contains WAP, follistatin/kazal, immunoglobulin, and two Kunitz and netrin domains, and it is thought to be a multidomain inhibitor of various types of proteases [43]. Similarly, proteins containing Kunitz and WAP domains have also been reported from *Caretta caretta*, *Caenorhabditis elegans, Homo sapiens* and *Maduca sexta* [43,44]. In all of these multidomain proteins, the WAP domain is followed by the Kunitz domain. In contrast, in Ku-wap-fusin, the Kunitz domain is followed by the waprin (WAP equivalent) domain. Therefore, it appears that Ku-wap-fusin is evolutionarily not closely related to other proteins containing Kunitz and WAP domains.

Comparison of gene structures of Kunitz SPI, Ku-wap-fusin and waprin genes reveals both addition and deletion of exons (Fig. **2B**). Nucleotide sequences of exon I of Ku-wap-fusin, waprin and Kunitz SPI genes show 84-92% identity. This exon encodes the signal peptide and four amino acids of the mature protein. The highly conserved signal peptides show ~90% identity among these toxins (shown in grey color), indicating that these toxins have evolved from a common ancestor. Unlike exon I, intron I is diverse in these

toxins, except for a highly similar region near the exon-intron boundary. The middle segments of Kunitz SPI intron I (shown in violet color, Fig. **2B**) appear to be replaced by another heterologous intronic segment (shown in green color, Fig. **2B**) in Ku-wap-fusin and waprin genes. Such deletion and addition of segments in the intronic region have been observed in other snake venom genes and are known to play an important role in the evolution and diversification of toxins [14]. Exon II in Kunitz SPIs and Ku-wap-fusin encodes the Kunitz domain and also shows a high degree of similarity. However, this domain and the intron II of Kunitz and Ku-wap-fusin seem to be deleted in the waprin gene (shown as dashed lines and box in Fig. **2B**). Again, exon III of Ku-wap-fusin and exon II of waprin are highly similar and encode the WAP domain (absent in the Kunitz gene). Exon III of Kunitz SPI genes in most cases encodes only one amino acid residue, the stop codon, and the 3′UTR, whereas exon III of waprin genes encodes only the 3′UTR (stop codons are found in exon II). However, in the Ku-wap-fusin gene, the stop codon and the 3′UTR are encoded by exon IV. The stop codon TGA, as seen in Kunitz SPI genes, has mutated to TCA in Ku-wap-fusin and results in an extension of the C-terminal end (Supplementary Fig. **S1**).

3.2. Evolution of New Classes of Protein Toxins

The origin and evolution of hetero-multimeric multi-domain proteins are of great interest. It is an established fact now that the assemblies of the heterologous modules or domains in chimeric proteins are carried out by intronic recombination [45,46]. Intronic recombination can successfully create the total or modular duplication, deletion and insertion in a gene and thus is considered as one of the most powerful factors affecting protein evolution. Comparison of the structures of normal and abnormal LDL-receptor genes reveals that misalignment and recombination involving the middle repetitive sequences (*Alu* repeats) of introns flanking the exons are responsible for the deletion and duplication of domains or entire sets of domains of multidomain proteins [45]. The same types of *Alu*-mediated intronic recombination are also responsible for affecting protein structure and function; examples include apolipoprotein B, β-hexosaminidase, antiplasmin and antithrombin. Intronic recombination can even mediate the fusion of genes located on different chromosomes [47]. Experimental evidence is now available to show that the group II introns can move to heterolologous sites by an alternative process of 'reverse splicing' (insertion of spliced intron into a different RNA) [48-50].

It was also suggested that insertion of exons might occur by the same mechanism of insertion of introns or by looping out–excision–reinsertion of modules [45]. During the process of intronic recombination, inclusion of an exon can produce a modular composition of intron-exon-intron. Such a modular composition of intron-exon-intron structure can

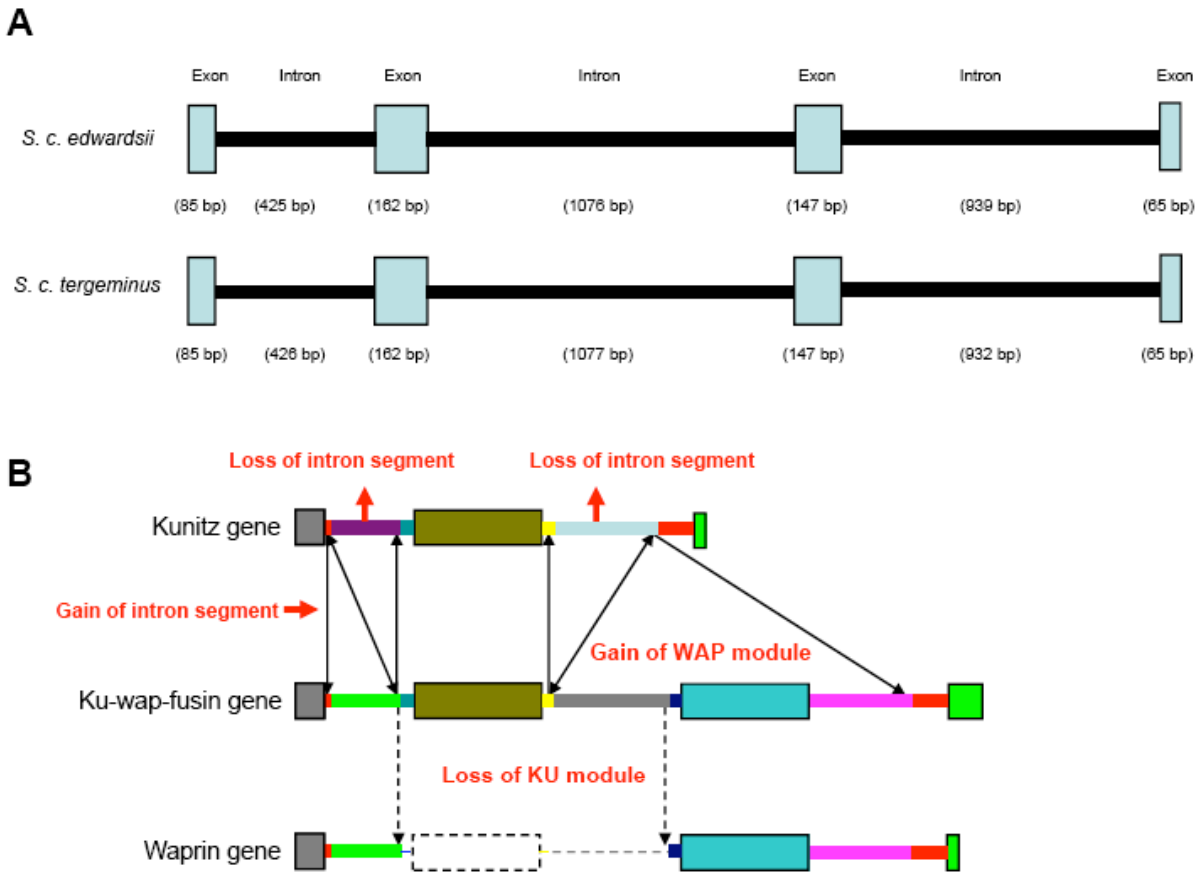

Fig. (2). Gene structure of ku-wap-fusins. **A**) Schematic representation of the gene structure of Ku-wap-fusins of *S. c. edwardsii* and *S. c tergeminus*. The exons and introns are represented by light turquoise boxes and black solid lines respectively. The number of base pairs is shown in parentheses. **B**) Schematic representation of insertion/deletion of new domains in Kunitz-type serine protease inhibitor, Ku-wap-fusin and waprin genes. The solid arrow indicates insertion of segment and a dashed arrow indicates deletion of segments.

then be re-inserted elsewhere in the genome and may give rise to new chimeric genes; this process can be called 'exon shuffling'. Exon shuffling can also result from alternate splicing followed by recombination.

During the evolution of new genes, selection appears to favor mutants created by intronic recombination which follow certain rules. According to the 'intron-phase compatibility rule' [51], the two intron partners involved in recombination must belong to the same intron phase class; otherwise, recombination will shift the reading frame and would lead to a loss of protein information downstream of the recombination point. A second rule, the 'symmetrical rule' [45,51], states that the two introns flanking the module are in the same phase. In other words, only the three symmetrical module groups are useful in exon shuffling (class 1-1, class 0-0 and class 2-2, depending on whether phase 1, phase 0 or phase 2 introns are found at both boundaries of the module). Non-symmetrical modules or exons (class 0-1, class 1-2 etc) are non-productive and useless to protein evolution by exon shuffling [45]. It must be noted that even symmetrical exons have a phase restriction: a class 0-0 exon can be inserted only in phase 0 introns, a 1-1 exon can be inserted in phase 1 intron and a 2-2 exon can be inserted only into phase 2 intron without disrupting the reading frame. Interestingly, both the Kunitz-type serine protease inhibitor module (KU) and the Whey protein module (WAP) are class 1-1 modules [45]. Moreover, all the introns in Ku-wap-fusin in both *S. c. edwardsii and in S. c. tergeminus* are in phase 1.

In general, an individual structural and folding domain of a multi-domain protein performs a specific, distinct function that may remain intact in the isolated domain. In the case of such functional autonomy these structural domains contribute to overall functions as independent functional domains [45]. Post-translational proteolytic processing of these multidomain proteins may gives rise to multiple functional domains which can act individually [52]. A well-studied example in snake venoms is the common precursor of metalloproteases and disintegrins. This common precursor undergoes post-translational processing in the Golgi apparatus and the individual domains are then separated (for a review see [53]). These proteolytically processed by-products are highly stable in the venom, and many metalloproteases and disintegrins have been purified from the crude venoms. Such processes contribute to the diversity of the toxins present in snake venom as well as to the biochemical and pharmacological properties of its venom.

Multidomain proteins like Ku-wap-fusin might also undergo such post-translational proteolysis, and the individual domains could act as separate toxins. The junction between the Kunitz and WAP domains in SCT and SCE2 has two basic residues, KR, which is a potential site for proteolytic processing, giving rise to two individual proteins, a Kunitz SPI and a waprin. However, in SCE1 this site has KP at the junction of the two domains and would not be processed. The Kunitz domain in Ku-wap-fusins has the conserved cysteine residues which will give the BPTI fold. The reactive site in Kunitz SPI has a highly conserved antiprotease site (P1), and the specificity towards serine proteases depends on the amino acid at this position. In Ku-wap-fusins, the P1 amino acid residue of the Kunitz domain is Arg (Fig. **1**), and it is thus likely to be a trypsin inhibitor. The other domain (after hypothetical proteolytic cleavage) is a waprin, which have been reported in many snake venoms; the biological roles of both Kunitz SPIs and waprins during envenomation are not yet known. Although the dibasic site is found in SCT and SCE2, proteolytic processing need not occur at this site. Hence, Ku-wap-fusins might also exist as full length proteins in the venom and could exhibit different biological activities than the single domain proteins. Molecular modeling of this fused toxin reveals that the Kunitz domain is structurally similar to other Kunitz SPI structure [54] with BPTI fold. Interestingly, although independent WAP domain folds similar to waprins, in the presence of Kunitz SPI domain, the WAP domain structure does not fold in the correct conformation. It is not clear whether it is the limitation of the modeling software or indeed ku-wap-fusin has unusual folding in the WAP domain.

3.3. Evolution of Ku-wap-fusin and Waprin Genes

In order to understand the evolutionary relationship of Ku-wap-fusin, waprins and Kunitz SPIs (see Fig. **1**), we constructed a Bayesian phylogenetic tree using the protein sequences (Fig. **3**). Posterior probabilities $\geq$0.75 are shown immediately preceding nodes. Boophilin (from *Boophilus microplus*) and BPTI (from *Bos taurus*) were identified as basal to all other sequences and were considered as an outgroup. The phylogenetic tree reveals that all of these genes have evolved from a common ancestor. They can be divided into four major groups: 1) elapid Kunitz SPIs; 2) the colubrid/viperid SPIs and Ku-wap-fusins; 3) the waprins; and 4) the neurotoxic Kunitz SPIs. The Ku-wap-fusins (fused protein) grouped with the colubrid and viperid Kunitz SPIs in a sister clade to the waprins, indicating both the intermediate nature of these fusion toxins and a somewhat closer relationship to the waprins than to the elapid Kunitz SPIs. We have also analyzed the phylogenetic relationship of mammalian proteins containing Kunitz and WAP domains. Interestingly, they cluster in a clade distinct from ku-wap-fusins (data not shown), indicating that ku-wap-fusins are not recruited from a reptilian counterpart of proteins containing Kunitz and WAP domains. It also indicates that ku-wap-fusins are not the ancestor of the mammalian counterparts. Thus both of these groups have evolved independently.

Recently, it has been reported that waprins and Kunitz SPIs have evolved from a common ancestor through gene duplication followed by diversification in the protein coding region [35]. Members of the Kunitz SPI gene family, such as β–bungarotoxin B chains and PILPs (protease inhibitor-like proteins), have also evolved from a common ancestor and diversified through an accelerated mode of evolution [38]; this adaptive mode of evolution has been reported in many snake venom superfamilies. However, Ku-wap-fusin and waprin genes seem to have evolved through intronic recombination resulting in deletion and insertion of intron segments as well as intron-exon-intron modules. Here we propose a model for the evolution of Ku-wap-fusin, Kunitz SPI and waprin genes (Fig. **4**). The genes encoding these proteins are obviously closely related; they have a highly conserved signal peptide (common exon I), and all known exons and non-coding regions are also highly conserved. Phylogenetic analysis indicates that the Ku-wap-fusins and the viperid and colubrid Kunitz SPIs form a clade sister to the waprins, and this cluster is a sister group to the elapid Kunitz SPIs. The

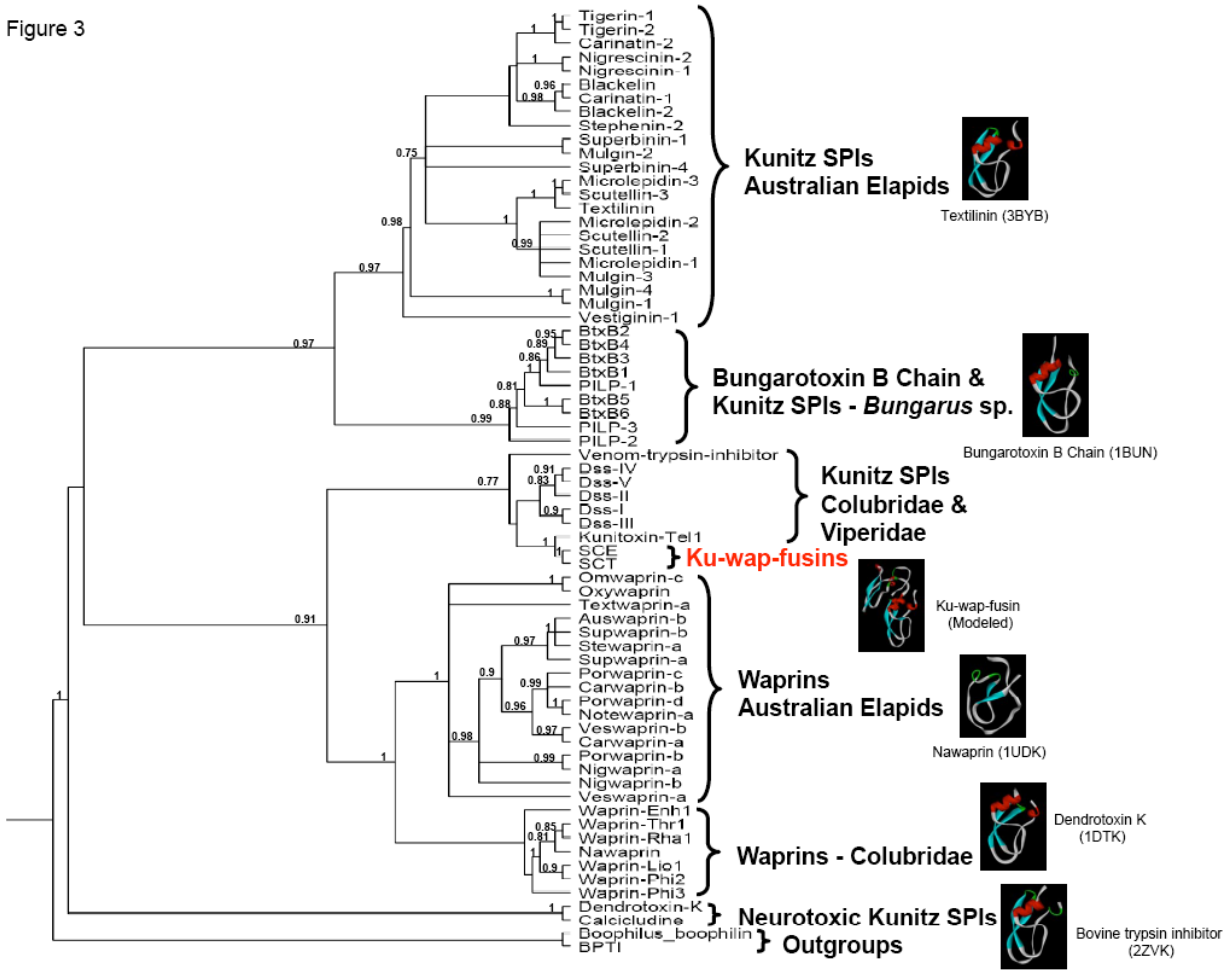

Fig. (3). Phylogenetic relationship of β-bungarotoxin B chain, Kunitz-type serine protease inhibitors, *Sistrurus* fused toxins (Ku-wap-fusins) and waprins. Protein sequences were aligned with ClustalX and a Bayesian analysis was performed with MrBayes, with a total of 10^6 generations, sampling every 100 generations, with a burn-in of 0.25. Posterior probabilities $\geq$0.75 are shown immediately preceding nodes, and proteins from different snake families are indicated. Microlepidin-1 (AY626930), Microlepidin-2 (AY626931) and Microlepidin-3 (EU401843) from *Oxyuranus microlepidotus*, Scutellin-1 (AY626928), Scutellin-2 (AY626929) and Scutellin-3 (EU401840) from *Oxyuranus scutellatus*, Textilinin (AF402324) from *Pseudonaja textilis textilis*, Blackelin (AY626934) and Blackelin-2 (EU401874) from *Pseudechis porphyriacus*, Superbinin-1 (EU401856) and Superbinin-4 (EU401857) from *Austrelaps superbus*, Mulgin-1 (AY626924), Mulgin-2 (AY626925), Mulgin-3 (AY626926) and Mulgin-4 (AY626927) from *Pseudechis australis*, Nigrescinin-1 (EU401850) and Nigrescinin (EF025517) from *Rhinoplocephalus nigrescens*, Stephenin-2 (EU401855) from *Hoplocephalus stephensii*, Carinatin-2 (EU401853) from *Tropediechis carinatus*, Tigerin-1 (EU401851) and Tigerin-2 (AY626933) from *Notechis scutatus*, Vestiginin-1 (EU401858) from *Demansia vestigiata* are Kunitz SPIs; SCE1 (DQ464286) and SCE2 (FJ746496) from *Sistrurus catenatus edwardsii* and SCT (FJ686529) from *Sistrurus catenatus tergeminus* are Ku-wap-fusins; Btx1 (CAA72809), Btx2 (CAJ18318), Btx3 (CAB44700), Btx4 (CAJ18319), Btx5 (CAJ18320) and Btx6 (CAJ18321) are bungarotoxin B chains from *Bungarus multicinctus*; PILP-1 (CAP74381), PILP-2 (CAP74382) and PILP3 (CAP74383) are protease inhibitor-like proteins from *Bungarus multicinctus*. Venom trypsin inhibitor (P24541) from *Eristocophis macmahonii*, DssI (ABD24040), DssII (ABD24041), DssIII (ABD24042), DssIV (ABD24043), and DssV (ABD24044) are Kunitz-type serine protease inhibitors from *Daboia russelli russelli*, Kunitoxin-Tel1 (ABU68487) from *Telescopus dhara;* Omwaprin-c (EU401817) and Oxywaprin (P83952) from *Oxyuranus microlepidotus*, Textwaprin-a (EU401837) from *Pseudonaja textiles*, Auswaprin-b (EU401819) from *Pseudechis australis*, Scuwaprin-a (EU401818) and Supwaprin-b (EU401830) from *Austrelaps superbus*, Stewaprin-a (EU401829) from *Hoplocephalus stephensii*, Porwaprin-b (EU401820), Porwaprin-C (EU401821) and Porwaprin-d (EU401822) from *Pseudechis porphyriacus*, Carwaprin-a (EU401827) and Carwaprin-b (EU401828) from *Tropediechis carinatus*, Notewaprin-a (EU401825) from *Notechis scutatus*, Veswaprin-a (EU401836) and Veswaprin-b (EU401833) from *Demansia vestigiata*, Nigwaprin-a (EU401823) and Nigwaprin-b (EU4018) from *Rhinoplocephalus nigrescens*, Waprin-Enh1 (ABU68545) from *Enhydris polylepis*, Waprin-Thr1 (ABU68540) from *Thrasops jacksoni*, Waprin-Rha1 (ABU68541) from *Rhabdophis tigrinus*, Nawaprin (P60589) from *Naja nigricolis* Waprin-Lio1 (ABU68544) from *Liophis poecilogyrus*, Waprin-Phi2 (ABU68543) and Waprin-Phi3 (ABU68546) from *Philodryas olfersii*, Dendrotoxin K (AAB26998) from *Dendroaspis polylepis*, calcicludine (AAB29942) from *Dendroaspis angusticeps* are waprins. BPTI (NP 991355) from *Bos taurus* and boophilin (CAC82582) from *Boophilus microplus* are Kunitz-type serine protease inhibitor used as outgroups in this analysis. Three-dimensional structures of different domain proteins are also shown. The structures were obtained from the PDB database except for ku-wap-fusin structure, which was obtained using I-TASSER modeling software. The ribbon model was created using the DS ViewerPro software.

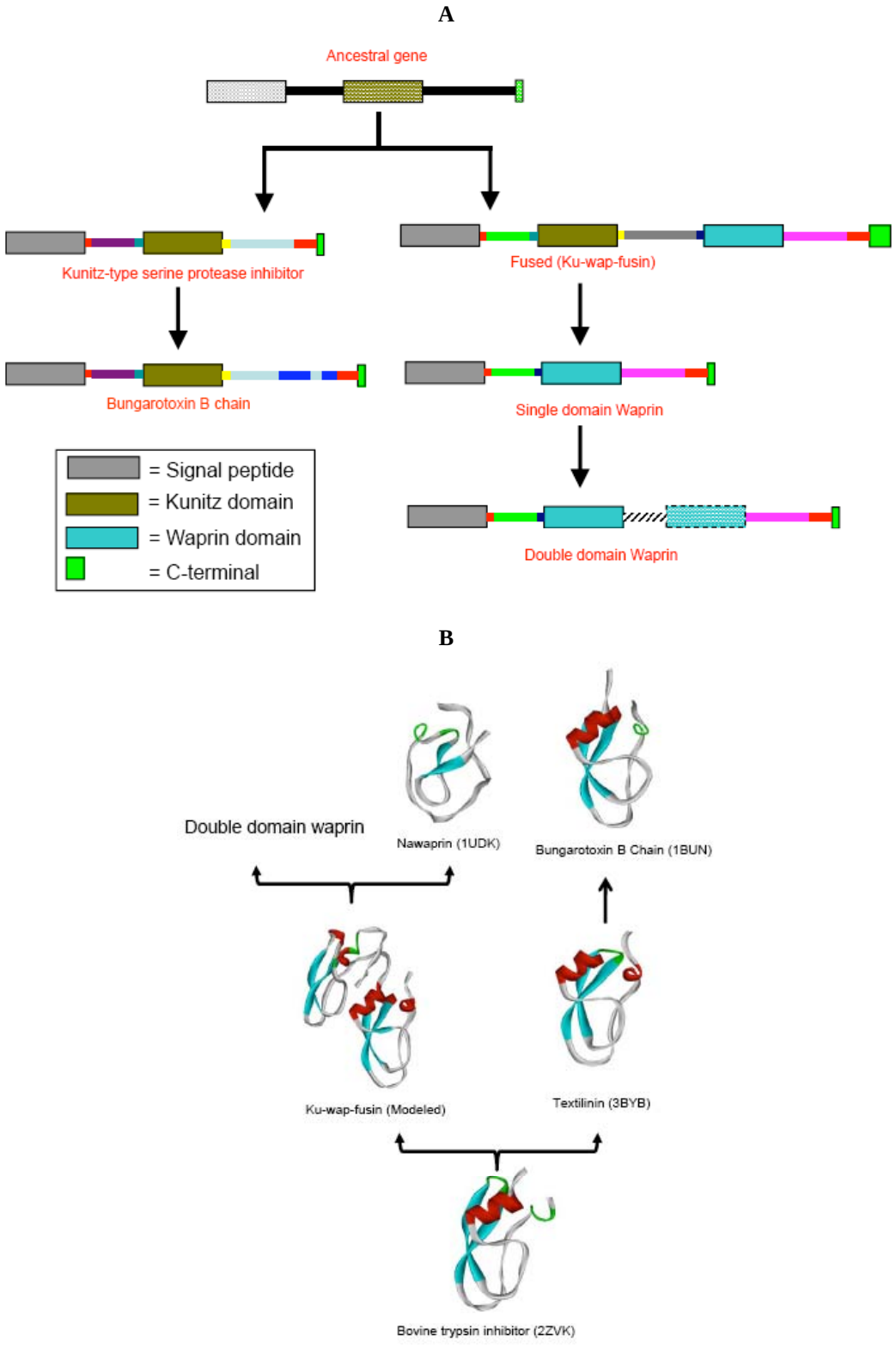

Fig. (4). A). Model for the evolution of Kunitz-type serine protease inhibitor, Ku-wap-fusin and waprin genes. Exons and introns are represented by colored boxes and solid lines, respectively, and exon domains are indicated. Similar/dissimilar nucleotide sequences in the introns are color coded. The ancestral gene in this hypothetical model is a Kunitz-type serine protease inhibitor gene. This ancestral gene has evolved to the present day Kunitz-type serine protease inhibitors and Ku-wap-fusins. β-bungarotoxin B chain has evolved through the present day Kunitz-type serine protease inhibitor, while waprins have evolved from the same ancestor through Ku-wap-fusins. Genes containing waprin domains in tandem might have evolved from a gene containing a single waprin domain through insertion of an additional exon (see text for details). **B**). Evolution of three-dimensional structures of different proteins. The structures were obtained from the PDB database except for ku-wap-fusin structure, which was obtained using I-TASSER modeling software. The ribbon model was created using the DS ViewerPro software.

neurotoxic Kunitz SPIs are basal to all of these clades, and it appears that the Kunitz SPI gene is the ancestral toxin family and that the other genes have evolved from this common ancestor. Within the elapid waprins, diversification appears to have occurred rapidly, as suggested by the short divergence lengths (Supplementary Fig. **S2**). Within this toxin superfamily, one lineage has evolved to the present-day elapid Kunitz SPI genes, β-bungarotoxin B chain and other neurotoxic homologs, whereas the other lineage includes Ku-wap-fusins, colubrid/viperid Kunitz SPIs and waprins. In the case of the β-bungarotoxin B chain, insertions of segments have occurred in the second intron, followed by accelerated evolution in exon II [38]. Ku-wap-fusin seems to have evolved from the same ancestor, with insertion of a WAP module in intron II and replacement of intronic segment in intron I. This Ku-wap-fusin gene appears to be an intermediate between the ancestor and the waprin gene. In the latter gene, the KU module [parts of intron I and exon II (of Kunitz SPI genes)] is deleted. Careful examination of the colubrid waprin-Phi1 (EU029742) from the venom of *Philodryas olfersii* [34] reveals that it has two WAP domains in tandem (Supplementary Fig. **S3**). Thus it seems that this gene is evolving through insertion of two WAP domains. In the TKDP (trophoblast Kunitz domain protein) gene family, insertion of three N-domains between the signal peptide and Kunitz domain have also been observed, and this insertion is through endogenous retroviral repetitive elements [55]. Similarly, in humans, WAP domain-containing proteins have also been reported to have evolved through repeated duplication of WAP domains [56]. Such retroviral repetitive elements were not observed in the Ku-wap-fusin gene. Comparison of the Ku-wap-fusin genes with the human WFDC2 gene [57] does not show any significant similarity. The human WFDC2 protein contains four disulphide core domains in tandem, whereas in Ku-wap-fusin, the Kunitz domain is followed by the WAP domain. In the human WFDC gene, the gene locus containing the WFDC2 gene has been subjected to a "birth-death" evolutionary event, where the generation of new genes occurs through a gene duplication event [57]. Alternatively, the evolution of a new gene with a new function has also been observed through intronic recombination and exon shuffling, and each exon encodes a distinct domain. Such events are responsible for the increase in genome size as well as the complexity of eukaryote genomes (for a review see [58]). In particular, these phenomenon are more evident in the evolution of extracellular multidomain proteins [59]. Hence Ku-wap-fusin (venom protein) is also likely to have evolved through such an event. In Ku-wap-fusin, the WAP domain acts also as a module, inserted into the ancestral gene of Kunitz SPI and giving rise to Ku-wap-fusin. Hence the gene superfamily including the waprins and Kunitz(-like) SPIs has obviously been subjected to extensive evolutionary reshaping, and this is likely yet another one of the factors responsible for the observed variation in composition of snake venoms.

CONCLUSION

Analysis of the gene structure of Ku-wap-fusin reveals that it has evolved through insertion of a WAP domain in the ancestral Kunitz gene. Our proposed model for evolution of this gene conforms well with established rules for intronic recombination. However, the exact mechanism of WAP domain insertion is not yet clear, but understanding such events will provide better insight into the evolution of these closely related proteins. The insertion of a WAP domain is apparently an independent event in this family of toxins, and further analysis of Ku-wap-fusin, the intermediate, will help explain the evolution of genes containing Kunitz and WAP domains in snake venom.

ACKNOWLEDGEMENTS

This work was supported by a grant to R.M. Kini from the Biomedical Research Council, Agency for Science, Technology and Research, Singapore. We thank M. McGlaughlin (UNC) for assistance with phylogenetic analyses.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Perkins, J.R.; Parker, C.E.; Tomer, K.B. The Characterization of Snake Venoms Using Capillary Electrophoresis in Conjunction With Electrospray Mass Spectrometry: Black Mambas. *Electrophoresis*, **1993**, *14*, 458-468.

[2] Perkins, J.R.; Tomer, K.B. Characterization of the Lower-Molecular-Mass Fraction of Venoms From Dendroaspis Jamesoni Kaimosae and Micrurus Fulvius Using Capillary- Electrophoresis Electrospray Mass Spectrometry. *Eur. J. Biochem*., **1995**, *233*, 815-827.

[3] Sanz, L.; Gibbs, H.L.; Mackessy, S.P.; Calvete, J.J. Venom Proteomes of Closely Related Sistrurus Rattlesnakes With Divergent Diets. *J. Proteome. Res*., **2006**, *5*, 2098-2112.

[4] King, G.F.; Gentz, M.C.; Escoubas, P.; Nicholson, G.M. A Rational Nomenclature for Naming Peptide Toxins From Spiders and Other Venomous Animals. *Toxicon*, **2008**, *52*, 264-276.

[5] Mackessy, S.P.; Williams, K.; Ashton, K. Characterization of the Venom of the Midget Faded Rattlesnake (Crotalus Oreganus Concolor): a Case of Venom Paedomorphosis? *Copeia*, **2003**, *4*, 769-782.

[6] Mackessy, S.P. Venom Composition in Rattlesnakes: Trends and Biological Significance. In *The Biology of Rattlesnakes*; Hayes, W. K., Beaman, K. R., Cardwell, M. D., Bush, S. P.,Eds.; Loma Linda University Press, Loma Linda: California, **2008**.

[7] Kordis, D.; Gubensek, F. Adaptive Evolution of Animal Toxin Multigene Families. *Gene*, **2000**, *261*, 43-52.

[8] Chijiwa, T.; Deshimaru, M.; Nobuhisa, I.; Nakai, M.; Ogawa, T.; Oda, N.; Nakashima, K.; Fukumaki, Y.; Shimohigashi, Y.; Hattori, S.; Ohno, M. Regional Evolution of Venom- Gland Phospholipase A2 Isoenzymes of Trimeresurus Flavoviridis Snakes in the Southwestern Islands of Japan. *Biochem. J*., **2000**, *347*, 491-499.

[9] Nakashima, K.; Ogawa, T.; Oda, N.; Hattori, M.; Sakaki, Y.; Kihara, H.; Ohno, M. Accelerated Evolution of Trimeresurus Flavoviridis Venom Gland Phospholipase A2 Isozymes. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 5964-5968.

[10] Nakashima, K.; Nobuhisa, I.; Deshimaru, M.; Nakai, M.; Ogawa, T.; Shimohigashi, Y.; Fukumaki, Y.; Hattori, M.; Sakaki, Y.; Hattori, S. Accelerated Evolution in the Protein- Coding Regions Is Universal in Crotalinae Snake Venom Gland Phospholipase A2 Isozyme Genes. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 5605-5609.

[11] Ogawa, T.; Nakashima, K.; Oda, N.; Shimohigashi, Y.; Ohno, M.; Hattori, S.; Hattori, M.;Sakaki, Y.; Kihara, H. Trimeresurus Flavoviridis Venom Gland Phospholipase A2 Isozymes. Genes Have Evolved Via Accelerated Substitutions. *J. Mol. Recognit*., **1995**, *8*, 40-46.

[12] Ogawa, T.; Nakashima, K.; Nobuhisa, I.; Deshimaru, M.; Shimohigashi, Y.; Fukumaki, Y.; Sakaki, Y.; Hattori, S.; Ohno, M. Accelerated Evolution of Snake Venom Phospholipase A2 Isozymes for Acquisition of Diverse Physiological Functions. *Toxicon*, **1996**, *34*, 1229-1236.

[13] Chang, L.S.; Huang, H.B.; Lin, S.R. The Multiplicity of Cardiotoxins From Naja Naja Atra (Taiwan Cobra) Venom. *Toxicon*, **2000**, *38*, 1065-1076.

[14] Fujimi, T.J.; 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*, 111-118.

[15] Gong, N.; Armugam, A.; Jeyaseelan, K. Molecular Cloning, Characterization and Evolution of the Gene Encoding a New Group of Short-Chain Alpha-Neurotoxins in an Australian Elapid, Pseudonaja Textilis. *FEBS Lett*., **2000**, *473*, 303-310.

[16] 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*, 83-88.

[17] Zupunski, V.; Kordis, D.; Gubensek, F. Adaptive Evolution in the Snake Venom Kunitz/BPTI Protein Family. *FEBS Lett.*, **2003**, *547*, 131-136.

[18] Ohno, M.; Menez, R.; Ogawa, T.; Danse, J.M.; Shimohigashi, Y.; Fromen, C.; Ducancel, F.; Zinn-Justin, S.; Le Du, M.H.; Boulain, J.C.; Tamiya, T.; Menez, A. Molecular Evolution of Snake Toxins: Is the Functional Diversity of Snake Toxins Associated With a 18 Mechanism of Accelerated Evolution? *Prog*. *Nucleic Acid Res*. *Mol*. *Biol*., **1998**, *59*, 307- 364.

[19] Pawlak, J.; Kini, R.M. Unique Gene Organization of Colubrid Three-Finger Toxins: Complete CDNA and Gene Sequences of Denmotoxin, a Bird-Specific Toxin From Colubrid Snake Boiga Dendrophila (Mangrove Catsnake). *Biochimie*, **2008**, *90*, 868-877.

[20] Tamiya, T.; Ohno, S.; Nishimura, E.; Fujimi, T.J.; Tsuchiya, T. Complete Nucleotide Sequences of CDNAs Encoding Long Chain Alpha-Neurotoxins From Sea Krait, Laticauda Semifasciata. *Toxicon*, **1999**, *37*, 181-185.

[21] Doley, R.; Pahari, S.; Mackessy, S.P.; Kini, R.M. Accelerated Exchange of Exon Segments in Viperid Three-Finger Toxin Genes (Sistrurus Catenatus Edwardsii; Desert Massasauga). *BMC Evol. Biol*., **2008**, *8*, 196.

[22] Doley, R.; Mackessy, S.P.; Kini, R.M. Role of Accelerated Segment Switch in Exons to Alter Targeting (ASSET) in the Molecular Evolution of Snake Venom Proteins. *BMC Evol. Biol.*, **2009**, *9*, 146.

[23] Chen, C.; Hsu, C.H.; Su, N.Y.; Lin, Y.C.; Chiou, S.H.; Wu, S.H. Solution Structure of a Kunitz-Type Chymotrypsin Inhibitor Isolated From the Elapid Snake *Bungarus Fasciatus*. *J*. *Biol*. *Chem*., **2001**, *276*, 45079-45087.

[24] Schweitz, H.; Heurteaux, C.; Bois, P.; Moinier, D.; Romey, G.; Lazdunski, M. Calcicludine, a Venom Peptide of the Kunitz-Type Protease Inhibitor Family, Is a Potent Blocker of High-Threshold Ca2+ Channels With a High Affinity for L-Type Channels in Cerebellar Granule Neurons. *Proc*. *Natl*. *Acad*. *Sci*. *USA*, **1994**, *91*, 878-882.

[25] Smith, L.A.; Lafaye, P.J.; LaPenotiere, H.F.; Spain, T.; Dolly, J.O. Cloning and Functional Expression of Dendrotoxin K From Black Mamba, a K+ Channel Blocker. *Biochemistry*, **1993**, *32*, 5692-5697.

[26] Laskowski, M., Jr.; Kato, I. Protein Inhibitors of Proteinases. *Annu*. *Rev*. *Biochem*., **1980**, *49*, 593-626.

[27] Gilquin, B.; Lecoq, A.; Desne, F.; Guenneugues, M.; Zinn-Justin, S.; Menez, A. Conformational and Functional Variability Supported by the BPTI Fold: Solution Structure of the Ca2+ Channel Blocker Calcicludine. *Proteins*, **1999**, *34*, 520-532.

[28] Harvey, A.L.; Karlsson, E. Dendrotoxin From the Venom of the Green Mamba, Dendroaspis Angusticeps. A Neurotoxin That Enhances Acetylcholine Release at Neuromuscular Junction. *Naunyn Schmiedebergs*. *Arch*. *Pharmacol*., **1980**, *312*, 1-6.

[29] Harvey, A.L.; Karlsson, E. Protease Inhibitor Homologues From Mamba Venoms: Facilitation of Acetylcholine Release and Interactions With Prejunctional Blocking Toxins. *Br*. *J*. *Pharmacol*., **1982**, *77*, 153-161.

[30] Joubert, F.J.; Taljaard, N. Snake Venoms. The Amino-Acid Sequence of Protein S2C4 From Dendroaspis Jamesoni Kaimosae (Jameson's Mamba) Venom. Hoppe Seylers. *Z*. *Physiol*. *Chem*., **1979**, *360*, 571-580.

[31] Filippovich, I.; Sorokina, N.; Masci, P.P.; de Jersey, J.; Whitaker, A.N.; Winzor, D.J.; Gaffney, P.J.; Lavin, M F. A Family of Textilinin Genes, Two of Which Encode Proteins With Antihaemorrhagic Properties. *Br*. *J*. *Haematol*., **2002**, *119*, 376-384.

[32] Torres, A.M.; Wong, H.Y.; Desai, M.; Moochhala, S.; Kuchel, P.W.; Kini, R.M. Identification of a Novel Family of Proteins in Snake Venoms. Purification and Structural Characterization of Nawaprin From Naja Nigricollis Snake Venom. *J*. *Biol*. *Chem*., **2003**, *278*, 40097-40104.

[33] Nair, D.G.; Fry, B.G.; Alewood, P.; Kumar, P.P.; Kini, R.M. Antimicrobial Activity of Omwaprin, a New Member of the Waprin Family of Snake Venom Proteins. *Biochem*. *J*. **2007**, *402*, 93-104.

[34] Fry, B.G.; Scheib, H.; van der, W.L.; Young, B.; McNaughtan, J.; Ramjan, S.F.; Vidal, N.; Poelmann, R.E.; Norman, J.A. Evolution of an Arsenal: Structural and Functional Diversification of the Venom System in the Advanced Snakes (Caenophidia). *Mol*. *Cell Proteomics*, **2008**, *7*, 215-246.

[35] St Pierre, L.; Earl, S. T.; Filippovich, I.; Sorokina, N.; Masci, P. P.; de Jersey, J.; Lavin, M.F. Common Evolution of Waprin and Kunitz-Like Toxin Families in Australian Venomous Snakes. *Cell Mol*. *Life Sci*., **2008**, *65*, 4039-4054.

[36] Pahari, S.; Mackessy, S.P.; Kini, M.R. The Venom Gland Transcriptome of the Desert Massasauga Rattlesnake (Sistrurus Catenatus Edwardsii): Towards an Understanding of Venom Composition Among Advanced Snakes (Superfamily Colubroidea). *BMC Mol*. *Biol*. **2007**, *8*,115.

[37] Chang, L.; Lin, S.; Huang, H.; Hsiao, M. Genetic Organization of Alpha-Bungarotoxins From Bungarus Multicinctus (Taiwan Banded Krait): Evidence Showing That the Production of Alpha-Bungarotoxin Isotoxins Is Not Derived From Edited MRNAs. *Nucleic Acids Res*., **1999**, *27*, 3970-3975.

[38] Chang, L.S.; Wang, J.J.; Cheng, Y.C.; Chou, W.M. Genetic Organization of Bungarus Multicinctus Protease Inhibitor-Like Proteins. *Toxicon*, **2008**, *51*, 1490-1495.

[39] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res*., **1994**, *22*, 4673-4680.

[40] Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian Phylogenetic Inference Under Mixed Models. *Bioinformatics*, **2003**, *19*, 1572-1574.

[41] Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics*, **2008**, *9*, 40.

[42] Breathnach, R.; Chambon, P. Organization and Expression of Eucaryotic Split Genes Coding for Proteins. *Annu*. *Rev*. *Biochem*., **1981**, *50*, 349-383.

[43] Trexler, M.; Banyai, L.; Patthy, L. A Human Protein Containing Multiple Types of Protease-Inhibitory Modules. *Proc*. *Natl*. *Acad*. *Sci*. *USA*, **2001**, *98*, 3705-3709.

[44] Nardi, J.B.; Martos, R.; Walden, K.K.; Lampe, D.J.; Robertson, H.M. Expression of Lacunin, a Large Multidomain Extracellular Matrix Protein, Accompanies Morphogenesis of Epithelial Monolayers in Manduca Sexta. *Insect Biochem*. *Mol*. *Biol*., **1999**, *29*, 883-897.

[45] Patthy L. Protein Evolution by Assembly of Modules. In *Protein Evolution*; Chapter 8. Blackwell Publishing: Malden, **2008**.

[46] Gilbert, W. Why Genes in Pieces? *Nature*, **1978**, *271*, 501-20.

[47] Babcock, M.; Pavlicek, A.; Spiteri, E.; Kashork, C.D.; Ioshikhes, I.; Shaffer, L.G.; Jurka, J.; Morrow, B.E. Shuffling of Genes Within Low-Copy Repeats on 22q11 (LCR22) by Alu- Mediated Recombination Events During Evolution. *Genome Res*., **2003**, *13*, 2519-2532.

[48] Mueller, M.W.; Allmaier, M.; Eskes, R.; Schweyen, R.J. Transposition of Group II Intron AI1 in Yeast and Invasion of Mitochondrial Genes at New Locations. *Nature*, **1993**, *366*, 174-176.

[49] Lambowitz, A.M.; Zimmerly, S. Mobile Group II Introns. *Annu*. *Rev*. *Genet*., **2004**, *38*, 1- 35.

[50] Belfort, M. An Expanding Universe of Introns. *Science*, **1993**, *262*, 1009-1010.

[51] Patthy, L. Intron-Dependent Evolution: Preferred Types of Exons and Introns. *FEBS Lett*., **1987**, *214*, 1-7.

[52] Kini, R.M.; Evans, H.J. 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*, 265-293.

[53] Fox, J.W.; Serrano, S.M. Insights into and Speculations About Snake Venom Metalloproteinase (SVMP) Synthesis, Folding and Disulfide Bond Formation and Their Contribution to Venom Complexity. *FEBS J*., **2008**, *275*, 3016-3030.

[54] Millers, E.K.; Trabi, M.; Masci, P.P.; Lavin, M.F.; de Jersey, J.; Guddat, L.W. Crystal Structure of Textilinin-1, a Kunitz-Type Ser-

ine Protease Inhibitor From the Venom of the Australian Common Brown Snake (Pseudonaja Textilis). *FEBS J*., **2009**, *276*, 3163-3175.

[55] Chakrabarty, A.; Green, J. A.; Roberts, R. M. Origin and Evolution of the TKDP Gene Family. *Gene*, **2006**, *373*, 35-43.

[56] Clauss, A.; Lilja, H.; Lundwall, A. A Locus on Human Chromosome 20 Contains Several Genes Expressing Protease Inhibitor Domains With Homology to Whey Acidic Protein. *Biochem. J*., **2002**, *368*, 233-242.

[57] Hurle, B.; Swanson, W.; Green, E.D. Comparative Sequence Analyses Reveal Rapid andcDivergent Evolutionary Changes of the WFDC Locus in the Primate Lineage. *Genome Res*., **2007**, *17*, 276-286.

[58] Patthy, L. Genome Evolution and the Evolution of Exon-Shuffling--a Review. *Gene*, **1999**, *238*, 103-114.

[59] Patthy, L. Modular Assembly of Genes and the Evolution of New Functions. *Genetica*, **2003**, *118*, 217-231.

Received: June 10, 2010 Revised: August 13, 2010 Accepted: August 30, 2010