CMLS, Cell. Mol. Life Sci. 61 (2004) 1751–1763
1420-682X/04/141751-13
DOI 10.1007/s00018-004-4149-1
© Birkhäuser Verlag, Basel, 2004

CMLS Cellular and Molecular Life Sciences

Research Article

The scorpine family of defensins: gene structure, alternative polyadenylation and fold recognition

S. Zhu and J. Tytgat*

Laboratory of Toxicology, University of Leuven, E. Van Evenstraat 4, 3000 Leuven (Belgium),
e-mail: jan.tytgat@pharm.kuleuven.ac.be

Received 5 April 2004; accepted 17 May 2004

Abstract. Small cationic antimicrobial peptides (SCAMPs) as effectors of animal innate immunity provide the first defense against infectious pathogens. This class of molecules exists widely in invertebrate hemolymph and vertebrate skin secretion, but animal venoms are emerging as a new rich resource. Scorpine is a unique scorpion venom defensin peptide that has an extended amino-terminal sequence similar to cecropins. From the African scorpion *Opistophthalmus carinatus* venom gland, we isolated and identified several cDNAs encoding four new homologs of scorpine (named opiscorpines 1–4). Importantly, we show for the first time the existence of multiple opiscorpine mRNAs with variable 3′ untranslated regions (UTRs) in the venom gland, which may be generated by alternative usage of polyadenylation signals. The complete opiscorpine gene structure including its promoter region is determined by genomic DNA amplification. Two large introns were found to be located within the 5′ UTR and at the boundary of the mature peptide-coding region. Such a gene structure is distinct, when compared with other scorpion venom peptide genes. However, a comparative promoter analysis revealed that both opiscorpine and scorpion venom neurotoxins share a similar promoter organization. Sequence analysis and structural modeling allow us to group the scorpines and scorpion long-chain K-channel toxins together into one family that shares a similar fold with two distinct domains. The N-terminal cecropin-like domain displaying a clear antimicrobial activity implies that the scorpine family represents a group of real naturally occurring hybrids. Based on the phylogenetic analysis, a possible cooperative interaction between the N and C domains is elucidated, which provides an evolutionary basis for the design of a new class of anti-infectious drugs.

Key words. Scorpine gene; in silico comparative promoter analysis; domain hybrid; venom antimicrobial peptide; cecropin; defensin; innate immunity.

Innate immunity, as an important component of animal defense, is mediated by some small cationic antimicrobial peptides (SCAMPs) that are constitutively expressed or induced by microorganisms or their products [for reviews, see refs 1–9]. In the past 20 years, several hundred such peptides have been found in a wide range of organisms, which offer rich sources for screening and developing new anti-infectious drugs. On the basis of the structural features, these natural molecules can be divided into three distinct groups: (i) linear, α-helical peptides; (ii) cysteine-containing peptides with single or several disulfide bridges, and (iii) molecules composed of specific amino acids such as Pro, Gly or Trp [1–9]. Despite enormous diversity in sequence and structure, most SCAMPSs are molecules of < 100 residues with spatially separated hydrophobic and charged regions. This unique structural arrangement is thought to be crucial to their function.

Cecropins are the first SCAMPs fully characterized from insects (*Hyalophora cecropia*) and can be considered as

* Corresponding author.

representative of the linear, α-helical peptides [6]. This class of molecules is widely distributed in nature and found not only in arthropods, but also in nematodes and even in bacteria [6, 10, 11]. These SCAMPSs are 3–4 kDa in size, and consist of two α-helical segments linked by a short hinge. These molecules exhibit strong activity against a large spectrum of microbes, including Gram-negative bacteria, filamentous fungi, protozoan and metazoan parasites as well as viruses, and thus play a key role in host innate defense.

Cysteine-containing defensins represent another important family of antimicrobial peptides [5, 7, 9]. They are only one class of SCAMPSs conserved between the plant and metazoan innate immune response [12]. These molecules can be classified into three major groups based on their three-dimensional (3D) structure: (i) plant and invertebrate defensins characterized by three disulfide bridges and a common structural motif (cystine stabilized α-helical and β sheet; CS$\alpha\beta$); (ii) vertebrate α- and β-defensins containing a core of three antiparallel β strands stabilized by three intra-molecular disulfide bridges, and (iii) invertebrate and vertebrate minidefensins, a new group of SCAMPSs genetically related to α- and β-defensins but half their size. Defensins exhibit a broad-spectrum activity against Gram-positive bacteria, fungi and some enveloped viruses.

Interestingly, a unique invertebrate defensin of 75 amino acids (called scorpine) was recently found in the venom of the African scorpion *Pandinus imperator*, which has an extended amino-terminal sequence similar to cecropins, and thus was proposed as a hybrid of these two class of SCAMPs, cecropins and defensins [13]. The highly antiparasitic activity (100 times higher than shiva-3) makes it an attractive candidate for generating transgenic mosquitoes resistant to malaria [14]. In this work, we characterized four new scorpine-like peptides (designated as opiscorpines 1–4) by a cDNA cloning strategy and determined the first ever gene structure and promoter region of this family from the scorpion *Opistophthalmus carinatus*, a close relative of *Pandinus imperator* (both belonging to the family of Scorpionidae). Furthermore, alternative polyadenylation is also demonstrated in this gene family. The research using a synthetic peptide confirmed that the extended N-terminal segment is a real natural structural domain with obvious antimicrobial activity. Sequence analysis combined with structural prediction highlighted a common fold comprising two distinct domains shared by scorpine peptides and scorpion venom long-chain K-channel toxins.

Materials and methods

Preparation of total RNA and genomic DNA

Scorpions (*O. carinatus*) were collected in southern Africa. Preparation of total RNA and genomic DNA was described previously [15].

Isolation of cDNA and genomic clones

The rapid amplification of cDNA ends (RACE) technique was used to isolate cDNAs encoding scorpine-like peptides. For 3′ RACE, a degenerate primer (DP1, TGT CACGG(TA)AC(TC)AAATGCAA(AG)TG(TC)GG) was designed based on the amino acid sequence of scorpine (CHGTKCKCG). Total RNA was reverse transcribed into first-strand cDNAs using Superscript II reverse transcriptase and a universal oligo(dT)-containing adaptor primer (AP, Life Technologies). The cDNAs were amplified with Taq DNA polymerase (Promega) using the primer DP1 together with the AUAP primer (Life Technologies). For 5′ RACE, two reverse gene-specific primers (SP1, CTTTGGCTGCAATTAATAGGACAATG; SP2, CTGCAATTAATAGGACAATGGTAC) were designed based on the 3′ RACE sequencing results. We modified the conventional 5′ RACE method for directly isolating the 5′ sequence from the synthesized first-strand cDNAs (fig. 1A). Briefly, the first cDNA mixture was tailed with terminal transferase and dCTP, and then an amplification of the cDNAs was performed using oligo(dG) and the gene-specific primer SP1. The ratio between oligo(dG) and SP1 was set at 1:50 (asymmetric amplification). After 30 cycles, 1 µl of diluted PCR product was taken as template for a standard 30 cycles of amplification with oligo(dG) and the nested gene-specific primer (SP2). Finally, a forward primer (FLP, GAAG GTTTCGTTGATTCTTCTGCCTC) was synthesized and used together with AUAP to isolate complete cDNAs. To determine the exon-intron organization of opiscorpines, genomic DNA was amplified using the primers FLP and SP1 under standard PCR conditions [16].

Isolation of the promoter region

Thermal asymmetric interlaced PCR (TAIL-PCR) was used to amplify the upstream regulation region of the opiscorpine gene. This method consists of three consecutive PCR reactions with three reverse nested gene-specific primers (4U1, 5′-CGAAAAGGCTCTCGATAA CAATAC-3′; 4U2, 5′-GTCCAAGGAAGATAAGGGCA GTA-3′; 4U3, 5′-GGAGGCAGAAGAATCAACGA-3′) and a shorter arbitrary degenerate primer (AD1, 5′-(A/T/G/C)GTCGA(C/G)(A/T)GA(A/T/G/C)A(A/T)GAA-3′) (fig. 2A). The primer 4U1 anneals to the 5′ end of intron 2, and 4U2 and 4U3 anneal to exon 2. The PCR reaction was performed as described by Zhu et al. [15].

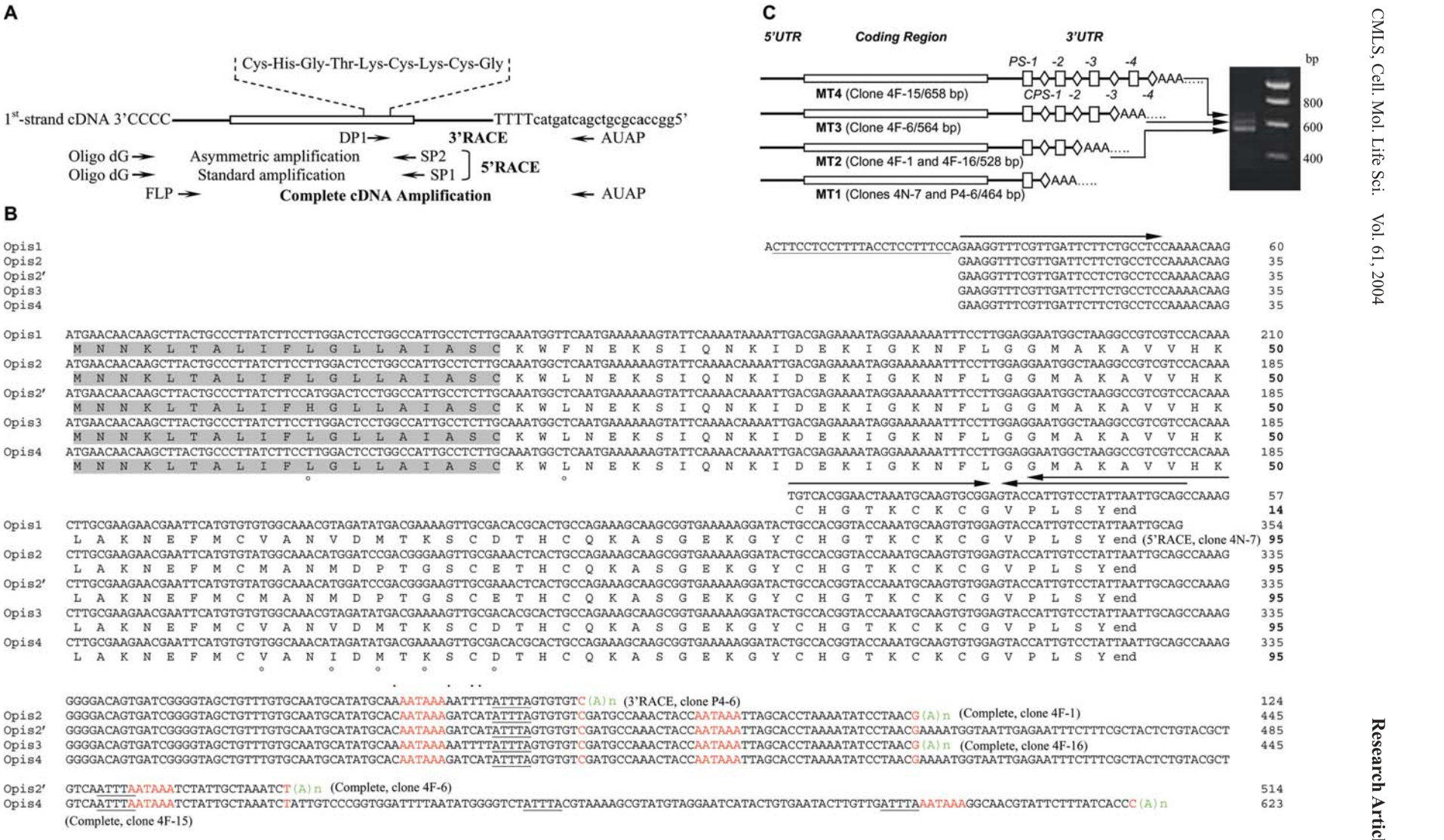

Figure 1. cDNA cloning of opiscorpines. (*A*) Isolation of the cDNAs by an improved RACE technique. Cys-His-Gly-Thr-Lys-Cys-Lys-Cys-Gly is the sequence of scorpine for the design of the degenerate primer (DP1). Primers for PCR amplification are represented by arrows. (*B*) Nucleotide and deduced amino acid sequences of opiscorpines. The signal peptide is shadowed. One pyrimidine tract of 23 nucleotides in the 5′ UTR and four ATTTA elements in 3′ UTR are undelined once. Four polyadenylation signals (AATAAA) are indicated in red, and cleavage/polyadenylation sites are in green. º indicates nonidentical amino acid sites; · indicates nonidentical nucleotide sites. (*C*) Schematic representation of alternative polyadenylation. UTR, untranslational region; PS, polyadenylation signal; CPS, cleavage and polyadenylation site; MT, mature transcript. Semi-quantitative RT-PCR detecting the relative abundance of four transcripts is shown to the right. *D*. Predicted RNA secondary structure of the 3′ UTR of the opiscorpine transcript.

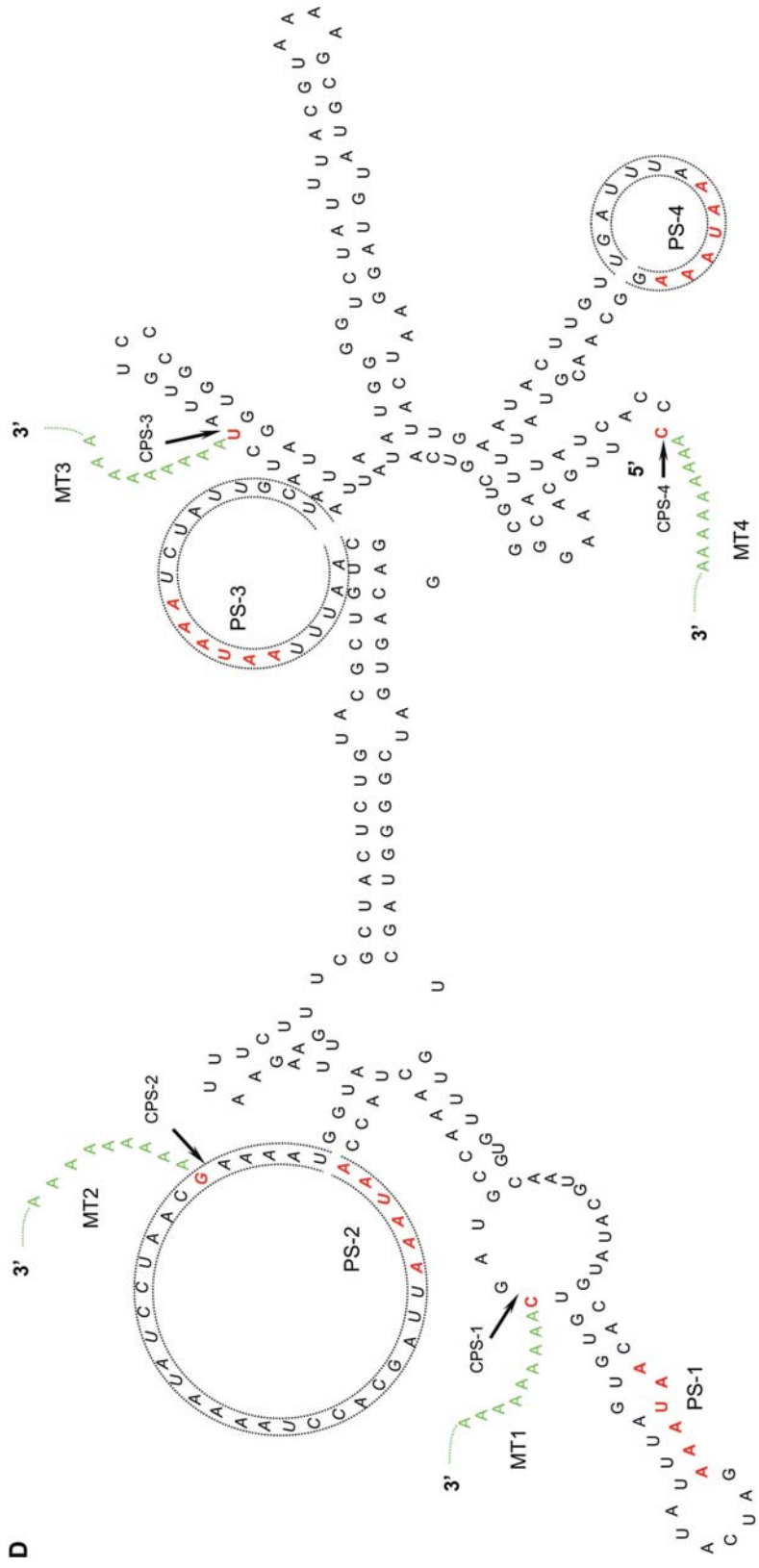

Figure 1 (continued)

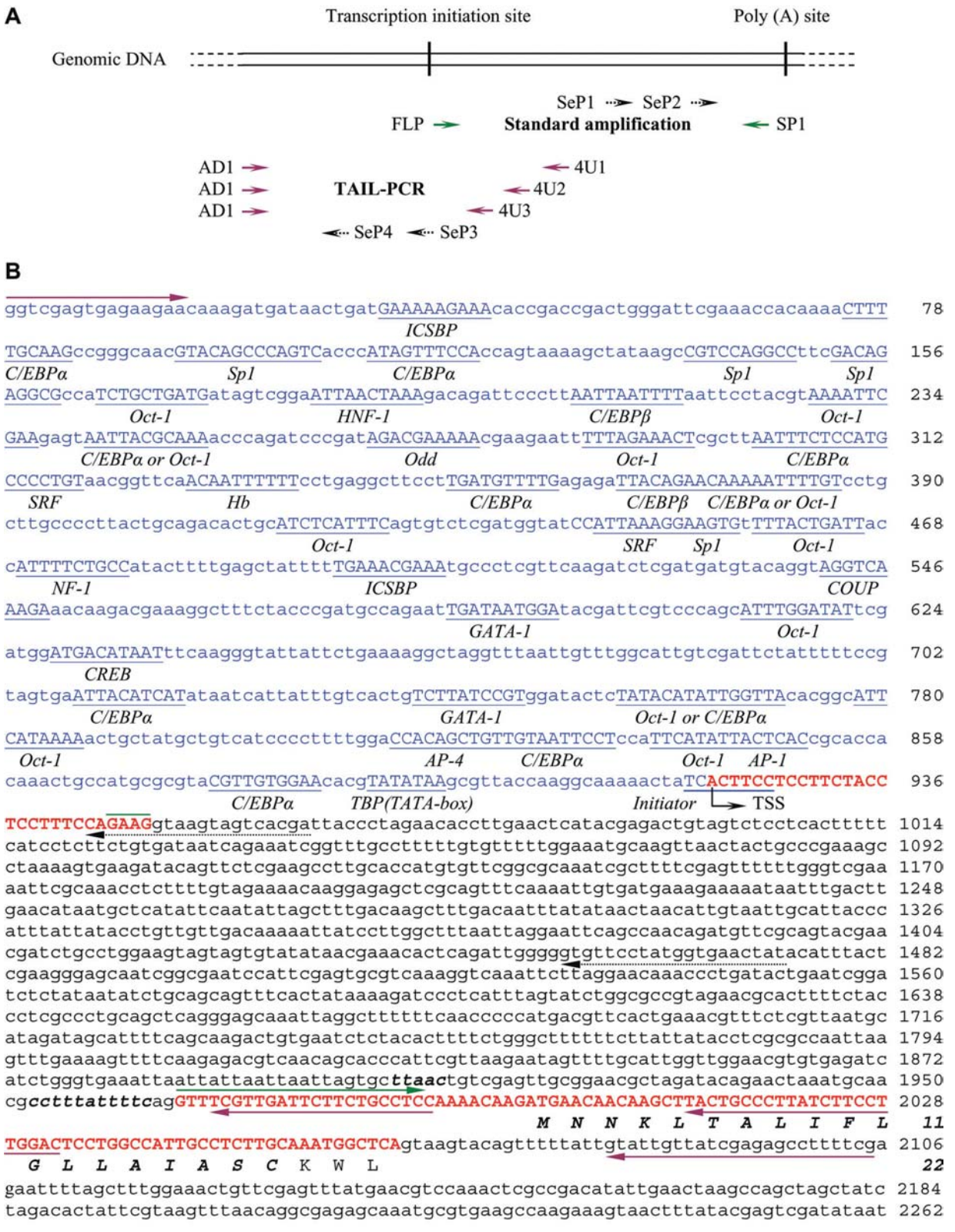

Figure 2. The opiscorpine gene and its promoter. (*A*) Strategies used for PCR amplification of genomic DNA and walking sequencing. (*B*) Nucleotide sequence. The upstream promoter is indicated in blue. The potential transcription factor-binding sites are underlined once and represented in upper case. The exon is represented by red upper case. The intron is represented by black lower case. The putative branch sites and the U2AF factor-binding region are italicized and in bold. The Alu-like sequence is shadowed. The repetitive element is in bold. Primers positions are indicated by arrows. (*C*) Schematic structure of the opiscorpine promoter and comparison with those of two additional scorpion venom peptide promoters. Two distinct modules are boxed and represented by different colors.

```
ttacgtagcaaatctgatggtatattctaagagagtacacttccgttgacgtttaagtcacgtgataccgcataatag 2340
tcataaggattctgtaaagccaatcatccgttattgaaaatgaacttattaattgctagaatcccttgcgactattat 2418
ctgtgtctagaaagtttcatgaacacgttcgtaattcgctgatagaaaatcgttcaactttcgtgtaggttattcaac 2496
ggcataaaatgcaccaaatatttcaaaggctgcgattacttccgtccaaagctgtgtaaatttgtcaagtttactgaa 2574
ctccgaagagaaaagttgggtccaagaacctccctcttgtgcagacagtagctcctgctttagtggggacagtccggg 2652
tctgttttatatctatcccagttcaactctatttaataaaaaaaaaaaaaaaaaaatgcatcgcaaacagcttttaatt 2730
gcatacaaatttataacttcgtgctaaaaaaagccgtgcatcagaaaaaataatgatgttatagaaatcgaagcaatg 2808
gccaaaagaacgattgtcattgttacgtaataactgacgtcacgataattgttacgtagtgataacattaatgcagaa 2886
ttgctagatatgatatttttttcttaaacttgtttgcatacagtcagtgaaaagaagttctgaattttaatataatttg 2964
caacttccttatcttaaacaaaactgtagttgttgattattacttataggtgataaactaactaacttccaattttaa 3042
ataatacgattaaattgacgcgaacatccacgttgtcccacttgccctggattctgtactgtacggcagtaagcgcag 3120
tattgtattgtattgtatttccttctatgcctctgttttgcttgctattatgggtcattttatagtcgaacatactca 3198
agttaattactatttctctttccagATGAAAAAAGTATTCAAAACAAAATTGACGAGAAAATAGGAAAAAATTTCCTT 3276
                          N  E  K  S  I  Q  N  K  I  D  E  K  I  G  K  N  F  L   40
GGAGGAATGGCTAAGGCCGTCGTCCACAAACTTGCGAAGAACGAATTCATGTGTGTGGCAAACGTAGATATGACGAAA 3354
 G  G  M  A  K  A  V  V  H  K  L  A  K  N  E  F  M  C  V  A  N  V  D  M  T  K   66
AGTTGCGACACGCACTGCCAGAAAGCAAGCGGTGAAAAAGGATACTGCCACGGTACCAAATGCAAGTGTGGAGTACCA 3432
 S  C  D  T  H  C  Q  K  A  S  G  E  K  G  Y  C  H  G  T  K  C  K  C  G  V  P   92
TTGTCCTATTAATTGCAGCCAAAG                                                        3456
 L  S  Y end                                                                      95
```

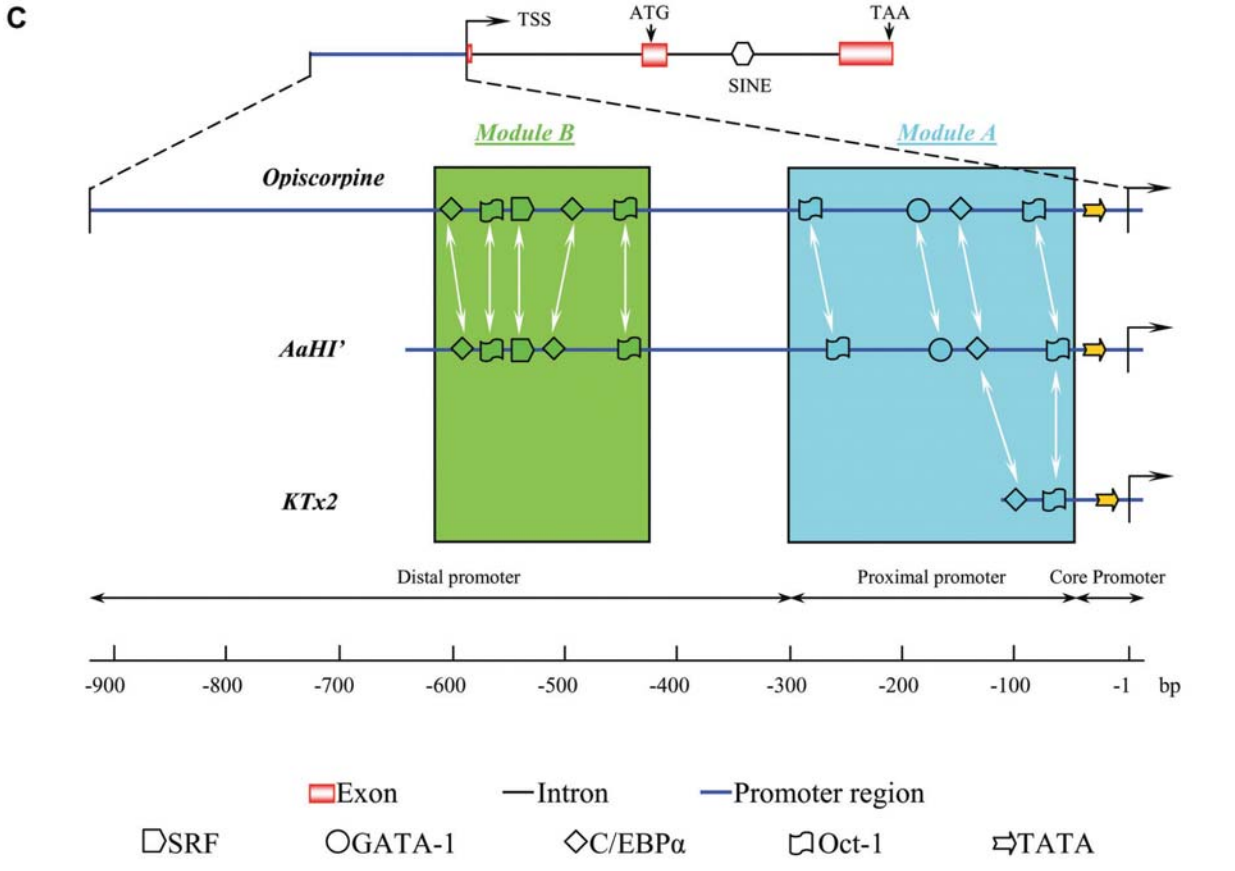

Figure 2 (continued)

DNA sequencing

PCR products derived from cDNA and genomic DNA were purified by high-speed centrifugation of agarose gel slices and ligated into the pGEM-T Easy Vector (Promega) for sequencing. *Escherichia coli* JM109 was used for plasmid propagation. Recombinant clones were analyzed by *Eco*RI digestion and gel electrophoresis. Positive clones were sequenced by the chain termination method from two directions using SP6 and T7 primers. A primer-walking strategy was used to sequence the genomic clones (SeP1, 5′-TTGCGACTATTATCTGTGTC-3′; SeP2, 5′-TTGATTATTACTTATAGGTG-3′; SeP3, 5′-ATAGTTCACCATAGGAACAC-3′; SeP4, 5′-TCGTGACTACTTACCTTCTG-3′) (fig. 2A).

The nucleotide sequences reported in this paper have been submitted to the GenBank database (http://www.ncbi.nlm.nih.gov) for release upon publication.

Semi-quantitative RT-PCR

Amplification was performed using the primer FLP and AUAP according to the method described in McPherson and Moller [16]. The template is derived from the first-strand cDNA mixture transcribed from mRNA primed by the universal oligo(dT)-containing adaptor primer (AP).

Chemical synthesis and CD spectra

Chemical synthesis, characterization and CD spectra of the amino-terminal segments (residues 1–9, named Jan-f1; residues 1–35, named Jan-f3) of opiscorpine 1 were performed according to Dai et al. [17].

Antibacterial assay

The inhibition zone assay described by Oh et al. [18] was used to determine the effect of Jan-f3 on *E. coli* JM109 (Promega).

Antifungal activity assay

Yeast strains used in this study were *Fusarium culmorum* IMI180420 and *Fusarium oxysporum* MUCL39789. Antifungal activity of the peptide sample against yeast strains was assayed by microscopic analysis of liquid cultures grown in microtiter plates as described previously [19]. The incubation temperature was 25 °C. Growth medium used was PDB (24 g/l potato dextrose broth; Sigma).

Bioinformatics

Protein secondary structure predictions were performed at the server (http://pbil.ibcp.fr) using the consensus prediction with a combination of the methods of DMP, DSC, GOR4, HNNC, PHD, Predator, SIMPA96 and SOPM. RNA secondary structure prediction was performed using RANdraw 1.12b software (http://rnadraw.base8.se) with default parameters. Fold assignment and template selection were carried out at the 3D-PSSM (three-dimensional position-specific scoring matrix) server (http://www.bmm.icnet.uk/servers/3dpssm). The nuclear magnetic resonance (NMR) structures of α2 peptide (a molecule structurally similar to the cecropin, PDB entry 1QP6) and sapacine (an insect defensin, PDB entry 1ICA) were selected for comparative modeling of amino- and carboxyl-terminal domains of scorpines and long-chain K-channel toxins, respectively. Initial backbone fitting and energy minimization steps were performed with the DeepView program (http://www.expasy.ch/spdbv) and further refined via submission to the Swiss-Model server (http://www.expasy.ch/swissmod/SWISS-MODEL.html). Evolutionary trees based on amino acid alignments of cecropins and invertebrate defensins as well as cecropin-like and invertebrate defensin-like domains in scorpines and long-chain K-channel toxins were constructed by using the neighbor-joining method implemented in MGEA 2.1 (http://www.megasoftware.net). Potential transcription factor (TF)-binding sites in upstream regulation regions of opiscorpine, AaHI′ and KT×2 gene promoters were searched using the AliBaba2 program, against the TRANSFAC database (http://www.cs.uni-magdeburg.de/grabe/alibaba2). The AliBaba2 program is believed to be the most specific tool for predicting TF sites in that it constructs specific matrices for each sequence analyzed instead of using predefined matrices for an unknown sequence.

Results and discussion

Isolation and identification of cDNAs encoding scorpine homologs from *O. carinatus*

The RACE technique was used to isolate full-length cDNAs encoding precursors of scorpine peptides from the scorpion *O. carinatus*. Sequencing of 3′ and 5′ RACE clones (P4-6 and 4N-7) allowed us to assemble a full-length cDNA encoding one new scorpine peptide, named opiscorpine 1. The cDNA is composed of three parts including a 5′ untranslated region (UTR), (ORF) and 3′ UTR. The 5′ UTR contains one pyrimidine tract of 23 nucleotides. At the 3′ UTR, one copy of a poly(A) signal (AATAAA) is located in the region 18 nucleotides upstream of the poly(A) addition. The ORF codes for a precursor of 95 residues composed of an amino-terminal signal peptide of 19 residues followed by a carboxyl-terminal part comprising the mature peptide of 76 residues. A complete amplification with primers FLP and AUAP identified three additional homologs with several amino acid changes mostly occurring in their C-terminal regions (fig. 1B). These four homologs share about 70% identity at the amino acid level with scorpine from *P. imperator* with one deletion/insertion mutation at position 55.

Alternative polyadenylation of opiscorpine transcripts

Sequence analysis showed that three clones (4F-16, 4F-6 and 4F-15) obtained from the complete amplification code for an identical 5′ UTR and highly similar coding sequences, but variable 3′ UTRs. Clones 4F-1 and 4F-16 have the same 3′ UTR length. Thus, together with clone P4-6 that contains the shortest 3′ UTR, we have isolated four cDNA species with variable 3′ UTRs (fig. 1C). Given the fact that the longest 3′ UTR (300 nt) contains four tandem poly(A) signals and that these signals are located in an ideal distance from cleavage sites (15–23 nt) [16], these four transcripts with variable 3′ UTRs are likely derived from alternative usage of the poly(A) signals. The result of semi-quantitative RT-PCR shows that

the smallest transcript is less abundant so that it cannot be detected by gel electrophoresis (fig. 1C).

Four ATTTA elements, which have been identified as destabilization sequences of some mRNAs [20], can be found within the longest 3′ UTR, but the abundance of this mRNA is obviously higher than the shortest one that contains only one ATTTA element. We can thus exclude its role in the mRNA destabilization. A similar case has also been observed in the mRNAs of the human *hap* gene [21]. Because the secondary structure of the 3′ UTR is crucial to the polyadenylation of pre-mRNA [22], we predicted the longest 3′ UTR structure (fig. 1D). The result revealed that all three polyadenylation signals (PS-2, PS-3 and PS-4) that are presumably responsible for the three larger transcripts are located within loops. By contrast, the signal (PS-1) for the smallest transcript is located in a stem. Given that a poly(A) signal (AATAAA) functions only when it is recognized by the cleavage and polyadenylation specificity factor (CPSF) [20], the location of AATAAA in a loop appears to facilitate its binding with the CPSF. A search of the GenBank database (http://www.ncbi.nlm.nih.gov) showed that the majority of scorpion venom peptide cDNAs possess a single polyadenylation signal and that most if not all AATAAAs were found in the loop (data not shown). This observation further emphasizes the importance of the location of AATAAA in the mRNA polyadenylation process.

The isolation of these transcripts in the absence of any immune challenge suggests that opiscorpines belong to constitutively expressed immune peptides. Obviously, this type of expression is energy consuming. To save energy, an alternative strategy is to control translational efficiency at the mRNA level by regulating mRNA stability. This strategy may have been adopted by the opiscorpine gene. By alternative usage of polyadenylation signals, several mRNAs with variable 3′ UTRs are produced to regulate the expression level of opiscopine by different stability, as observed in the case of the human PR264/SC35 splicing factor [23].

Intron-exon organization of opiscorpine gene

PCR amplifications were used to determine the intron-exon organization of the opiscorpine gene. The strategy used is schematically outlined in figure 2A. The amplification of genomic DNA, using primers FLP and SP1, produced one ~1.5-kb DNA fragment that was then cloned and sequenced. A comparison between genomic and cDNA sequences identified one phase one intron (1160 bp) located within a polar residue (Asn) codon of the mature peptide-coding region (fig. 2B). The exon sequence accurately matches opiscorpine 3 cDNA. When we tried to isolate the promoter region of the opiscorpine gene by TAIL-PCR, we found another large intron (1016 bp) which is located within the 5′ UTR. These two introns display a similar A+T content (~50%), and have a consensus GT-AG splice junction and a putative U2AF splicing factor-binding region that shares a similar sequence feature (pyrimidine tract) to that of other species. Their putative branch sites are located 54 and 18 nucleotides upstream of the 3′ splice sites, respectively. The presence of one intron in the untranslated region and mature peptide-coding region split by an intron are uncommon in scorpion venom peptide genes [24, 25].

Analysis of the two intron sequences revealed two distinct features in the second intron: (i) One copy of a repetitive element (GTATTGTATTGTATTGTATT) was found upstream of the branch site. This element is repeated four times by one pentamer (GTATT). Interestingly, a database search (http://www.ncbi.nlm.nih.gov) showed that this element is also present in genomes of several eukaryotic species, including *Lycopersicon esculentum*, *Nicotiana tabacum, Plasmodium falciparum, Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* (data not shown). In some of these genomes, this element repeats more than 20 times in one or two orientations. The widespread distribution in diverse species indicates its evolutionary conservation and putative important biological significance; (ii) A new member of the short interspersed element (SINE) family was also identified within the intron. Despite no obvious sequence similarity, this stretch of 210 nucleotides displays characteristics of the SINE: its sequence is flanked by direct repeats of 8 nucleotides and has an A-rich region in its 3′ end [26]. In the eight-nucleotide repeat, four are palindromic (TGCA). Two *Alu*I sites (AGCT) were also found within this sequence (fig. 2B), suggesting that this sequence might belong to an *Alu*-like mobile element. The biological roles of the two sequences await further investigation.

Identification of potential transcriptional factor TF-binding sites in the promoter region and in silico comparative promoter analysis

The promoter region of the opiscorpine gene was isolated by a TAIL PCR strategy (fig. 2A). After tertiary amplification had been performed, one ~2-kb fragment was recovered and sequenced. We defined the transcription start site (TSS) sequence (TC$\mathbf{A}^{+1}$CTTCC) by a direct mapping of cDNA into the genomic sequence, which exactly matches the eukaryotic initiator element consensus [(TC)C$\mathbf{A}^{+1}$N(TA)(TC)(TC)(TC)] (where $\mathbf{A}^{+1}$ is the base at which transcription starts, N is any of the four bases) [26]. The initiator is thought to be an alternative promoter element which is capable of replacing the function of the TATA box (one of the most prominent core elements in eukaryotic as well as prokaryotic promoters). However, one copy of the TATA box can also be found −30 to −24 bp upstream of the TSS. The combination of both initiator and TATA box in one promoter has been identified in several viral genes [27]. In addition to these two core

promoter elements, a computer-assisted analysis indicated that the opiscorpine promoter region contains numerous putative TF-binding sites (fig. 2B), among which C/EBPα (CCAAT/enhancer binding protein), Oct-1 (octamer-binding transcription factor), and Sp1 (specificity protein 1) were found in multiple copies. Although *in silico* prediction of a single binding site is of limited value on its own, the promoter context (promoter module) is crucial for TF-binding sites to fulfil their functions (e.g. activating or repressing) [28]. A closer analysis of the potential sites allows us to identify several interesting combinations which might be related to regulatory function [29] and should be subjected to further experimental research: (i) an AP1 site (activating protein 1) located close to the TSS; (ii) three copies of closely spaced Sp1 sites in the distal region which might be important in regulating the functions of both the TATA box and initiator, and (iii) several overlapping binding sites in the proximal and distal region, which includes three copies of [Oct-1][C/EBPα], one copy of [AP-4][C/EBPα] one copy of [SRF][Sp1], and one copy of [Oct-1][AP-1].

To provide a clue for the regulation mechanism of venom gland tissue-specific gene expression at the level of the promoter module, we performed a comparative promoter analysis [27]. In this analysis, we found that, surprisingly, despite the lack of any detectable sequence similarity, both opiscorpine and AaHI′ genes [29] share a similar promoter organization which comprises two distinct modules (A and B) separated by one spacer of about 200 bp (fig. 2C). Module A is located within the proximal promoter region composed of four TF sites ([Oct-1][GATA-1][C/EBPα][Oct-1]) whereas module B within the distal region is composed of five TF sites ([C/EBPα][Oct-1][SRF][C/EBPα][Oct-1]). Given that this conserved module structure nicely matches the assembly model of a promoter complex, we assume that it might represent one functional context which may be specific to scorpion venom peptide gene expression and regulation. Significantly, the lack of the nuclear factor κB response element (NFκB-RE) sequence (a critical site for immune induction, which is found in nearly all promoters of inducible SCAMP genes in animals [30]) in both opiscorpine and AaHI′ promoter regions confirms their constitutive-expression feature.

Scorpines and long chain K-channel toxins belong to one family of domain hybrids with sequence and structural similarities to cecropins and invertebrate defensins

A BLAST search on the amino acid sequence of opiscorpine precursors revealed a significant similarity to those of scorpion long-chain K-channel toxins (TsTXKβ, AaTXKβ and BmTXKβ2) [31, 32]. About 38% identity and 56% similarity (including the conservation replacement sites) were found between them when aligning these sequences using the CLUSTAL program (fig. 3A). Of these two groups of molecules, TsTXKβ and scorpine may be considered as their respective representative in that their natural products have been isolated and functions are also known [13, 33]. TsTXKβ is a selective blocker of voltage-gated noninactivating K-channels [33]. Based on comparison with the sequence of TsTXKβ, the precursors of AaTXKβ and BmTXKβ2 were identified to contain an amino-terminal signal peptide followed by a propeptide of 8 residues and a carboxyl-terminal part comprising the mature peptide. In contrast, the precursors of the scorpine family are processed only by removing an amino-terminal signal peptide [13]. However, we found that two cleavage sites, characterized in the scorpine and the long-chain K-channel toxin precursors, are completely conserved between them (fig. 3A).

If we assume that long-chain K-channel toxin precursors only cut off their own signal peptides, the similarity between these two classes of molecules is further strengthened by their secondary structure and 3D fold. Although the sequence similarity of scorpine, as a hybrid peptide, to cecropins and invertebrate defensins has been recognized by Conde et al. [13], the same case for the long-chain K-channel toxins (fig. 3B) was not mentioned by these authors. Secondary-structure prediction clearly indicates that the amino-terminal region of all the molecules (residues 1–35, numbered according to scorpines) is actually able to form two α helices disrupted by a Gly or Pro residue. This typical structural feature of cecropins is further supported by comparative protein modeling [34], which revealed an amphipathic structure with hydrophobic and hydrophilic clusters on two sides of the two helices (fig. 3B) and thus predicted a possible antimicrobial activity.

To experimentally confirm this prediction, we synthesized the amino-terminal segment of opiscorpine 1 (residues 1–35), named Jan-f3 (molecular weight = 3985.71). CD analysis revealed a typical α-helical structure in 60% trifluoroethanol (TFE) with negative ellipticities at 208 and 222 nm (fig. 4). In particular, this peptide displays an antifungal activity (for *F. culmorum* and *F. oxysporum*, 8.8 and 10.04 µM of 50% inhibitory concentration (IC_{50}) were respectively obtained when incubated in $^1/_2$ PDB medium), which was also observed for cecropins. The classical antibacterial method described by Hultmark was used to determine the antibacterial effect of Jan-f3 on *E. coli* JM109 (Promega). A clear inhibition zone was formed with a minimum inhibitory concentration (MIC) of ~12 µm. A shorter fragment (residues 1–9, named Jan-f1), corresponding to half of the first helix (fig. 3B), showed no antibacterial activity, suggesting that a complete helical structure is a prerequisite for the antimicrobial activity.

Taken together, we provide evidence in favor of the extended N-terminal segment in scorpines representing a

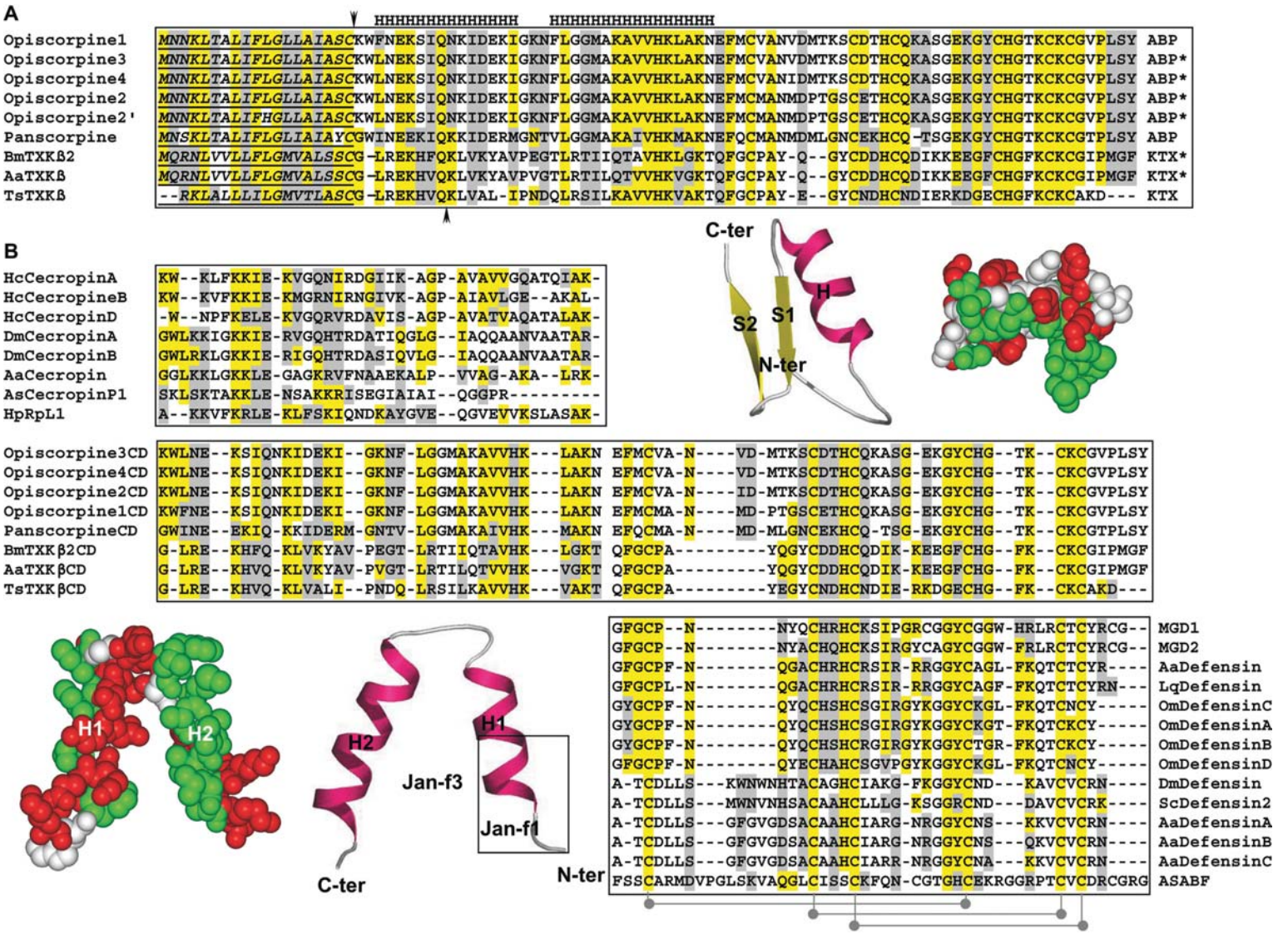

Figure 3. Identification of a family of hybrid peptides. (*A*) Precursor sequence alignment of scorpine peptides and several long-chain K-channel toxins from scorpion venoms. Identical residues are shaded in yellow, conservative replacement in gray. Two processing sites are indicated by arrows. H, α helix by secondary-structure prediction. Disulfide bridges are shown at the bottom. * indicates putative activities based on sequence similarity. (*B*) Comparison of amino- and carboxyl-terminal sequences of the scorpine family with cecropins and invertebrate defensins. The predicted 3D structures of two domains of opiscorpine 1 are shown as cartoon and CPK models (two in nonidentical orientation). As hightlighted by the CPK models, clustering of hydrophobic (green) and hydrophilic (red) amino acids into two distinct regions in the two domains provides a structural basis for their antimicrobial activity. H, α helix; S, β strand. Insects: Hc, *Hyalophora cecropia*; Dm, *Drosophila melanogaster*; Sc, *Stomoxys calcitrans*; Aa, *Aedes aegypti*. Tick: Om, *Ornithodoros moubata*. Scorpions: Opi, *Opistophthalmus carinatus* (this work); Bm, *Buthus martensii*; Ts, *Tityus serrulatus*; Aa, *Androctonus australis*; Pan, *Pandinus imperator*; Lq, *Leiurus quinquestriatus*. Mollusc: MGD1 and MGD2, *Mytilus galloprovincialis*. Nematode: As, *Ascaris suum*. Bacterium: Hp, *Helicobacter pylori* (for sequence sources, see the ref. 9–11, 13, 31, 32, 41).

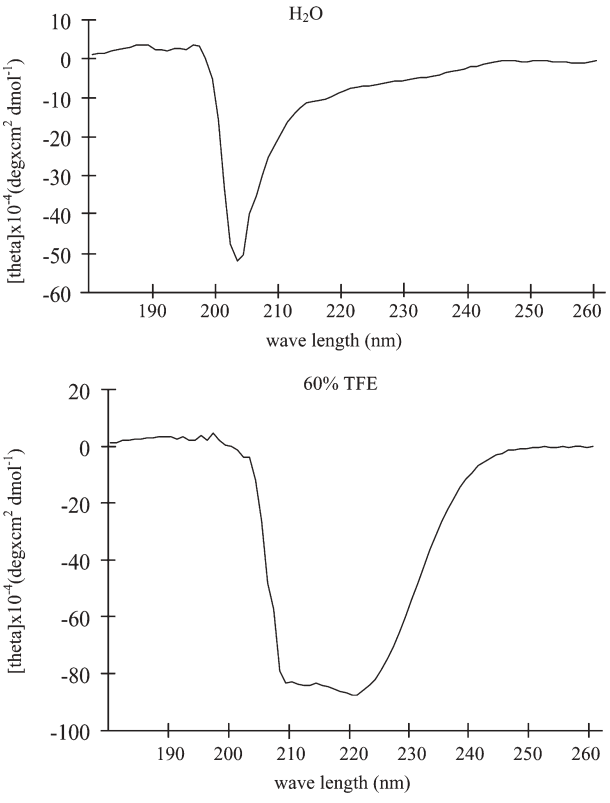

Figure 4. CD Spectra of Jan-f3. Spectra were taken at a peptide concentration of 0.1 g/l in water and 60% aqueous TFE.

natural structural and functional domain genetically related to cecropins. Given the fact that the carboxyl-terminal segments of these two classes of molecules possess a typical Cys pattern of invertebrate defensins ($CX_{4-15}\mathbf{CX_2HC}X_{6-9}GX_1CX_{4-9}\mathbf{CX_1C}$), it will induce a CS$\alpha\beta$ fold that is adopted by all the invertebrate defensins characterized so far [35, 36]. This is also supported by comparative protein modeling [34] that predicted a typical CS$\alpha\beta$ structure with clusters of hydrophobic and hydrophilic residues separated on two faces of the molecule, thereby forming an 'amphipathic' antimicrobial architecture (fig. 3B). Our analyses therefore suggest that the two domains within a single chain can be considered as spatially distinct structures that could conceivably fold and function in isolation. The immune peptides (scorpine and opiscorpines) and the scorpion venom neurotoxins (TsTXKβ, AaTXKβ and BmTXKβ2) may thus be grouped together into one family in which all members share a common fold with two distinct domains. Considering the remarkable sequence, fold and domain organization similarity to scorpines, further research is needed to evaluate a possible antimicrobial activity of the K-channel toxin members in this family. A detailed structural investigation may also prove helpful to provide a clue to explain their functional diversification.

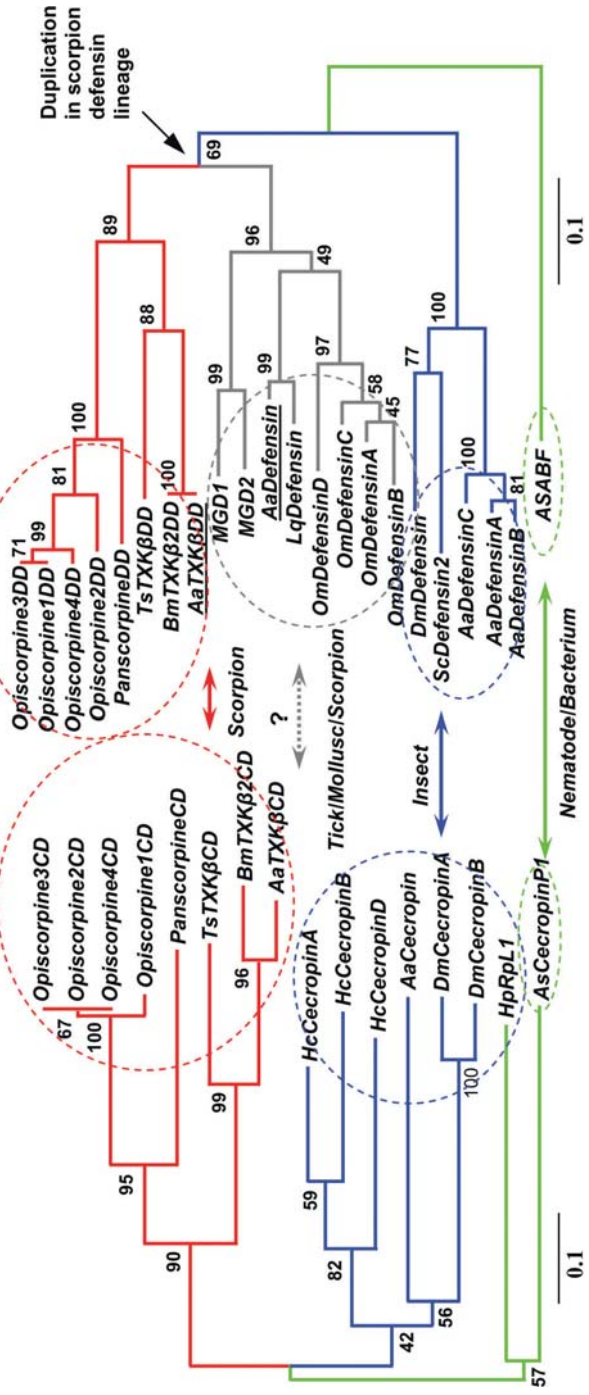

Figure 5. Similarity of phylogenetic trees. The cecropin family (left), and the invertebrate defensin family (right). In this analysis, the two domains in scorpines and long chain K-channel toxins were considered as independent proteins (CD, cecropin domain; DD, defensin domain). The dotted arrow represents lack of corresponding cecropin sequences in three species (tick, mollusc and scorpion). The amino acid sequences used are the same as in figure 3B. Trees were constructed using the neighbor-joining method based on the proportion of amino acid differences (p). The root was determined by using the nematode sequence as an outgroup. Numbers on the branches are bootstrap percentages. A possible gene duplication event in the scorpion defensin lineage, based on the existence of defensin (AaDefensin) and a defensin domain in the hybrid (*AaTXKβ*) in the scorpion *Androctonus australis* genome, is indicated by an arrow.

Co-evolution of two domains

Generally, it is accepted that if two proteins or two domains of one protein evolve in a consistent manner (co-evolution), one can logically conclude that the two proteins or the two domains will be functionally related (as a pair of interaction partners or involved in the same pathway) [37, 38]. The finding and identification of a family of single chain peptides (composed of scorpines and long-chain K-channel toxins) with two distinct domains provide a glimpse for a possible interaction between the two domains. This is supported by a phylogenetic analysis that reveals a clear co-evolutionary relationship between the trees of the two domains (fig. 5) with a correlation coefficient of 0.94 [38]. This proposal is further strengthened by the observation of Andersson et al. [11] who showed that *Ascaris* nematodes make these two types of SCAMPs (cecropin P1, and two *Ascaris suum* antibacterial factors (ASABFs) belonging to the invertebrate defensin family) conferring protection against infection. Other evidence also is available from the experimental data which indicates that the combination of linear peptides and defensins can produce a synergistic effect against microbes in vivo [39]. On the basis of these observations, it is now becoming clear that integrating two distinct SCAMPs with a co-evolutionary relationship into a single peptide (hybrid) may significantly improve the activity and antimicrobial spectrum by a cooperative manner and can offer obvious advantages in both drug design and clinical application.

Conclusions

In recent years, the body of evidence supporting the role of SCAMPs in local immune defense of animal venom glands has been growing. Such peptides have widely been reported in the venoms of several animal organisms, such as spider, solitary wasp, bee, ant and scorpion (for a review, see ref. 40). However, little genomic data regarding their gene structure and expression regulation elements are currently available. We isolated and determined the first complete gene structure as well as its promoter region in the animal venom defensin family, which provides us with a genetic basis for further elucidating the expression pattern differences between invertebrate venom- and hemolymph-derived immune peptides. The highly antiparasitic activity observed in the scorpine peptide [13] also allows construction of a rational vector by integrating the whole scorpine gene for generating transgenic mosquitoes to control malaria [14]. The recognition of one structural and functional domain in the N-terminal region of the scorpine family lays a foundation for further probing the action mode and evolutionary mechanism of these unique molecules. Given that the hybrid of two SCAMPs represents an apparent design advantage at both the structural and functional level, identification of such molecules in other species will be particularly interesting both for an understanding of functional innovation in SCAMPs from an evolutionary perspective and for a nature-guided development of a new class of anti-infectious drugs.

Acknowledgements. This work was supported in part by a K.U. Leuven post-doctoral fellowship to S. Z. We gratefully thank L. Dai for chemical synthesis and CD spectra of Jan-f3, and K. Thevissen for the antifungal tests. Furthermore, we thank Fons Verdonck for providing scorpions for our research and Lorenzo Prendini for scorpion identification.

1 Zasloff M. (2002) Antimicrobial peptides of multicellular organisms. Nature **415:** 389–395
2 Hancock R. E. W. and Scott M. G. (2000) The role of antimicrobial peptides in animal defenses. Proc. Natl. Acad. Sci. **97:** 8856–8861
3 Boman H. G. (2000) Innate immunity and the normal microflora. Immun. Rev. **173:** 5–16
4 Boman H.G. (2003) Antibacterial peptides: basic facts and emerging concepts. J. Intern. Med. **254:** 197–215
5 Lehrer R. I. and Ganz T. (2002) Defensins of vertebrate animals. Curr. Opin. Immunol. **14:** 96–102
6 Vizioli J. and Salzet M. (2002) Antimicrobial peptides from animals: focus on invertebrates. Trends Pharmacol. Sci. **23:** 494–496
7 Ganz T. (2003) Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. **3:** 710–720
8 Andreu D. and Rivas L. (1998) Animal antimicrobial peptides: an overview. Biopolymers **47:** 414–433
9 Dimarcq J., Bulet P., Hetru C. and Hoffmann J. (1998) Cysteine-rich antimicrobial peptides in invertebrates. Biopolymers **47:** 465–477
10 Pütsep K., Branden C. I., Boman H. G. and Normark S. (1999) Antibacterial peptide from *H. pylori.* Nature **398:** 671–672
11 Andersson M., Boman A. and Boman H. G. (2003) *Ascaris* nematodes from pig and human make three antibacterial peptides: isolation of cecropin P1 and two ASABF peptides. Cell. Mol. Life Sci. **60:** 599–606
12 Thomma B. P. H. J., Cammue B. P. A. and Thevissen K. (2002) Plant defensins. Planta **216:** 193-202
13 Conde R., Zamudio F. Z., Rodriguez M. H. and Possani L. D. (2000) Scorpine, an anti-malaria and anti-bacterial agent purified from scorpion venom. FEBS Lett. **471:** 165–168
14 Possani L. D., Corona M., Zurita M. and Rodriguez M. H. (2002) From noxiustoxin to scorpine and possible transgenic mosquitoes resistant to malaria. Arch. Med. Res. **33:** 398–404
15 Zhu S., Darbon H., Dyason K., Verdonck F. and Tytgat J. (2003) Evolutionary origin of inhibitor cystine knot peptides. FASEB J. **17:** 1765–1767
16 McPherson M. J. and Moller S. G. (2000) PCR, Cromwell Press, Trowbridge
17 Dai L., Corzo G., Naoki H., Andriantsiferana M. and Nakajima T. (2002) Purification, structure-function analysis, and molecular characterization of novel linear peptides from scorpion *Opisthacanthus madagascariensis*. Biochem. Biophys. Res. Commun. **293:** 1514–1522
18 Oh H., Hedberg M., Wade D. and Edlund C. (2000) Activities of synthetic hybrid peptides against anaerobic bacteria: aspects of methodology and stability. Antimicrob. Agents Chemother. **44:** 1694–1696
19 Thevissen K., Terras F. R. and Broekaert W. F. (1999) Permeabilization of fungal membranes by plant defensins inhibits fungal growth. Appl. Environ. Microbiol. **65:** 5451–5458

20 Zhao J., Hyman L. and Moore C. (1999) Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. **63:** 405–445

21 Qu X., Qi Y. and Qi B. (2002) Generation of multiple mRNA transcripts from the novel human apoptosis-inducing gene *hap* by alternative polyadenylation utilization and the translational activation function of 3′ untranslated region. Arch. Biochem. Biophys. **400:** 233–244

22 Klasens B. I. F., Thiesen M., Virtanen A. and Berkhout B. (1999) The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. Nucleic Acids Res. **27:** 446–454

23 Sureau A. and Perbal B. (1994) Several mRNAs with variable 3′ untranslated regions and different stability encode the human PR264/SC35 splicing factor. Proc. Natl. Acad. Sci. USA **91:** 932– 936

24 Dai L., Wu J., Gu Y., Lan Z., Ling M. and Chi C. (2000) Genomic organization of three novel toxins from the scorpion *Buthus martensi* Karsch that are active on potassium channels. Biochem. J. **346:** 805–809

25 Zhu S., Li W. and Zeng X. (2001) Precursor nucleotide sequence and genomic organization of BmTXKS1, a new scorpion toxin-like peptide from *Buthus martensii* Karsch. Toxicon **39:** 1291–1296

26 Lodish H., Berk A., Zipursky S. L., Matsudaira P., Baltimore D. and Darnell J. (2000) Molecular Cell Biology, New York

27 Werner T. (1999) Models for prediction and recognition of eukaryotic promoters. Mamm. Genome **10:** 168–175

28 Fessele S., Maier H., Zischek C., Nelson P. J. and Werner T. (2002) Regulatory context is a crucial part of gene function. Trends Genet. **18:** 60–63

29 Delabre M. L., Pasero P., Marilly M. and Bougis P. E. (1995) Promoter structure and intron-exon organization of a scorpion α-toxin gene. Biochemistry **34:** 6729–6736

30 Hoffmann J. A. and Reichhart J. M. (2002) *Drosophila* innate immunity: an evolutionary perspective. Nat. Immunol. **3:** 121–126

31 Zhu S., Li W., Zeng X., Jiang D., Mao X. and Liu H. (1999) Molecular cloning and sequencing of two 'short chain' and two 'long chain' K^+ channel-blocking peptides from the Chinese scorpion *Buthus martensii* Karsch. FEBS Lett. **457:** 509–514

32 Legros C., Ceard B., Bougis P. E. and Martin-Eauclaire M. F. (1998) Evidence for a new class of scorpion toxins active against K^+ channels. FEBS Lett. **431:** 375–380

33 Rogowski R. S., Krueger B. K., Collins J. H. and Blaustein M. P. (1994) Tityustoxin Kα blocks voltage-gated noninactivating K^+ channels and unblocks inactivating K^+ channels blocked by α-dendrotoxin in synaptosomes. Proc. Natl. Acad. Sci. USA **91:** 1475–1479

34 Duret L., Guex N., Peitsch M. C. and Bairoch A. (1998) New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. Genome Res. **8:** 348–353

35 Tamaoki H., Miura R., Kusunoki M., Kyogoku Y., Kobayashi Y. and Moroder L. (1998) Folding motifs induced and stabilized by distinct cystine frameworks. Protein Eng. **11:** 649–659

36 Bontems F., Roumestand C., Gilquin B., Ménez A. and Toma F. (1991) Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. Science **254:** 1521–1523

37 Valencia A. and Pazos F. (2003) Prediction of protein-protein interaction from evolutionary information. In: Structural Bioinformatics, pp. 411–426, Bourne P. E. and Weissig H. (eds), Wiley-Liss, Hoboken

38 Goh C.S., Bogan A. A., Joachimiak M., Walther D. and Cohen F. E. (2000) Co-evolution of proteins with their interaction partners. J. Mol. Biol. **299:** 283–293

39 Ong P. Y., Ohtake T., Brandt C., Strickland I., Boguniewicz M., Ganz T. et al. (2002) Endogenous antimicrobial peptides and skin infections in atopic dermatitis. N. Engl. J. Med. **347:** 1151–1160

40 Kuhn-Nentwig L. (2003) Antimicrobial and cytolytic peptides of venomous arthropods. Cell. Mol. Life Sci. **60:** 2651–2668

41 Zheng X. and Zheng L. (2002) Genomic organization and regulation of three cecropin genes in *Anopheles gambiae.* Insect Mol. Biol. **11:** 517–525

To access this journal online:
http://www.birkhauser.ch