CMLS, Cell. Mol. Life Sci. 62 (2005) 2257–2269
1420-682X/05/202257-13
DOI 10.1007/s00018-005-5200-6

CMLS Cellular and Molecular Life Sciences

Review

Phylogenetic distribution, functional epitopes and evolution of the CSαβ superfamily

S. Zhu[a,*], B. Gao[a] and J. Tytgat[b]

[a] Group of Animal Innate Immunity, State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology, Chinese Academy of Sciences, No. 25 Beisihuan-Xi Road, Beijing 100080 (P.R. China), e-mail: Zhusy@ioz.ac.cn
[b] Laboratory of Toxicology, Leuven University, E. Van Evenstraat 4, 3000 Leuven (Belgium)

Received 11 May 2005; received after revision 20 June 2005; accepted 28 June 2005
Online First 7 September 2005

Abstract. A superfamily of proteins often conserves a common structural scaffold but develops diverse biochemical and biological functions during evolution. The understanding of evolutionary mechanisms responsible for this diversity is of fundamental importance not only in structural genomics but also in nature-guided drug design. A superfamily of peptides with a conserved CSαβ structural motif provides a considerably intriguing example to approach such an issue. The peptides from this superfamily have wide origins, ranging from plants to animals, and exhibit diverse biological activities, varying from a sweet-tasting protein to antibacterial defensins and animal toxins targeting ion channels. This review describes the phylogenetic distribution and structural classification of this unique scaffold and provides new insights into its functional diversity from the perspective of sequence, structure and evolution.

Key words. Phylogeny; peptide; toxin; fold.

Introduction

According to hierarchical protein (or domain) classification systems, evolutionarily related proteins can roughly bedivided into four levels of units: family, superfamily, fold and class [1]. Proteins with detectable sequence similarity are grouped into a family that comprises the most basic classification unit of protein universe and defines the evolutionary relationship of all the members within a family. A clear evolutionary link among different families within a superfamily, however, often is difficult to detect due to low sequence similarity as well as functional diversity [2–4]. For the latter, it is even more challenging to fully understand the evolutionary mechanism responsible for this diversity.

A simple structural motif called the cysteine-stabilized α-helix and β-sheet (CSαβ), initially recognized by the group of Menez and later named by Cornet et al., defines a peptide superfamily [5, 6] and gives an excellent example in approaching such issues. Peptides belonging to this superfamily exhibit relatively diverse biochemical and biological functions. In most cases, these molecules serve a common function as defenders of their hosts, both for animals (e. g. scorpion toxins, plant defensins with enzyme inhibitor activities) and microorganisms (e. g. plant and animal antibacterial/antifungal defensins) [7–9]. This structural motif is stabilized by six conserved Cys residues which form three intramolecular disulfide bridges [5, 6]. In sequence, this motif corresponds to the

* Corresponding author.

consensus sequence C…CXXXC…C…CXC. Structurally, it is composed of a single α-helix and one β-sheet of two strands. The α-helix spanning the CXXXC is connected by two disulfide bridges (C2/C5 and C3/C6) to the C-terminal β-strand, whereas the third disulfide bridge (C1/C4) links the N-terminus to the first β-strand [5] (fig. 1). Previous structural analysis has confirmed that such a consensus cystine framework can induce and stabilize the formation of the CSαβ structural motif [10]. This finding offers a possibility for searching more new distant homologues containing this motif.

Bontems et al. first noticed the existence of this common motif between antibacterial insect defensins and scorpion toxins based on a limited structural comparison [5]. Subsequently, a great number of peptides with more diverse activities were found to possess this motif. The extensive distribution of this motif in some functionally diverse polypeptides suggests that it is a relatively stable and versatile scaffold for developing new biological activity. Indeed, the CSαβ motif, as a well-defined scaffold, has the potential to tolerate insertions, deletions and substitutions throughout the structure. Artificial grafting of an unrelated active site onto this scaffold did not significantly alter the global folding of this structural motif [11]. Thus it appears that the CSαβ motif may be an attractive candidate for engineering design, by which peptides with novel functions can be created.

In an early review, Menez et al. discussed the structural basis for functional diversity of animal toxins with the CSαβ motif and proposed that functional diversity of this motif was associated with the decoration of the fold with deletion, insertion and addition of disulfide bridges [12]. Here, we will attempt to use the information from recently obtained sequences and structures to present an overview of the CSαβ superfamily, focusing on the phylogenetic distribution, structural classification and evolution of functional diversity. The evolutionary origin of the unusually varied toxic CSαβ peptides is analyzed in detail from a structural point of view.

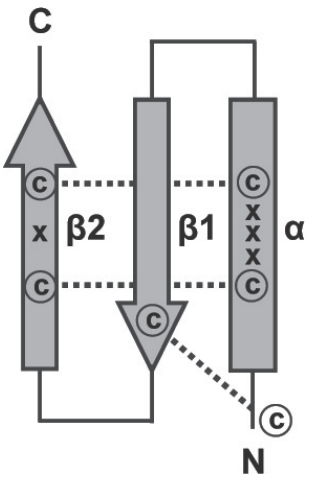

Figure 1. Topology of the CSαβ superfamily. The single α-helix and two β-strands are indicated with cylinder and arrows, respectively. Dotted lines represent disulfide bridges.

Phylogenetic distribution

More recently, phylogenetic analyses have been used successfully to study the molecular origin and evolution of 24 different types of snake toxins [13]. To perform similar analyses, we first determined the phylogenetic distribution of the CSαβ superfamily by using similarity search and regular expression pattern-based methods. The latter method fully utilized the conserved cysteine pattern of this superfamily to find remote homologues. Using the conserved signatures of the CSαβ superfamily (…C-X(2,18)-C-X(3)-C-X(2,10)-[GAPSIDERYW]-X-C-X(4-17)-CXC...), we performed regular expression pattern searches against several databases (Swiss-Prot, TrEMBL, TrEMBL-new and PDB) (http://us.expasy.org/) [14] and five genomes of model species *(Drosophila melanogaster, Anopheles gambiae, Apis mellifera, Caenorhabditis elegans, Caenorhabditis briggsae)* (http://www.ensembl.org/). Combined with BLAST searching of the GenBank database (http://www.ncbi.nlm.nih.gov/), we identified most, if not all, peptides with the CSαβ motif in these databases (fig. 2). WEBLOGO [15] was used to visualize the consensus of a given species according to the aligned amino acid sequences.

These results allow us to define the phylogenetic distribution of the CSαβ superfamily. Meanwhile, different bioactivities can now be mapped onto the tree of life, and the point of emergence of a given biological activity can be deduced from the resulting distribution. Our results demonstrate that the CSαβ fold is restricted to plants and animals belonging to Protostomia, including Ecdysozoa (arthropods and nematodes) and Lophotrochozoa (mollusks), but is lacking in Deuterostomia (e.g. vertebrates) [16]. Given that both Protostomia and Deuterostomia commonly originated from Bilateria [16], gene loss in Deuterostomia after divergence from Protostomia appears to be a plausible explanation. Furthermore, this loss can be expected in that vertebrates have gained two new types of defensins (α- and β-defensins) [17] to replace the functions of the CSαβ antimicrobial peptides. Alternatively, gene transfer [14] from plants to Protostomia may be another attractive scenario. Evidence for the lack of CSαβ in lower invertebrates will provide support for the latter hypothesis.

Insect

Insects are a major resource of CSαβ peptides with various antimicrobial activities. These molecules were originally isolated from the hemolymph of immunized larvae of the dipteran insect *Phormia terranovae* and later found in many other insect species, the majority

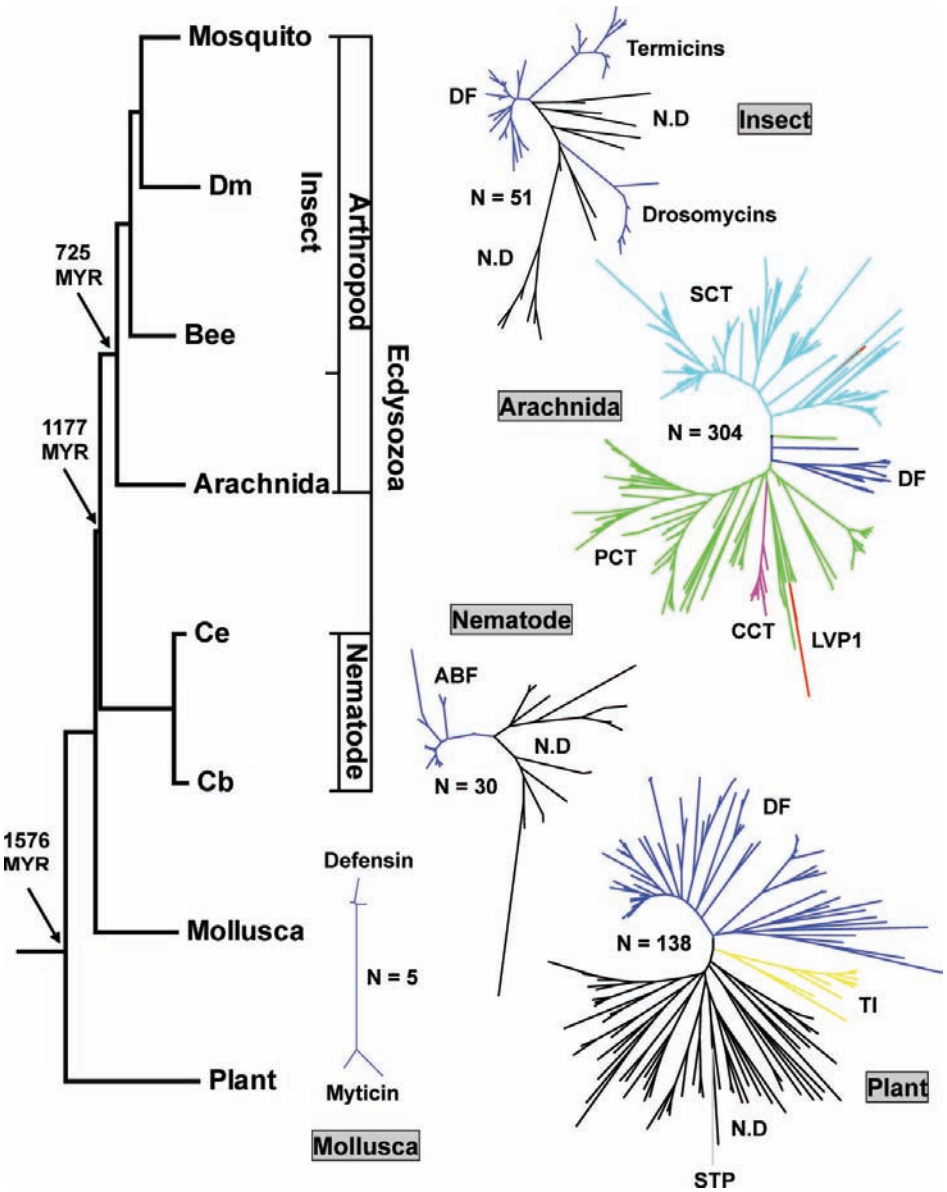

Figure 2. Phylogenetic distribution of the CSαβ superfamily. The known functions of some members are indicated by colors. The life tree is constructed based on Hsp90 sequences by using the Neighbor-Joining method. The divergence time is labeled in the left of the tree. MYR, million years; Dm, *D. melanogaster;* Ce, *C. elegans;* Cb, *C. briggsae.* The phylogenetic analysis was performed using the same method based on the amino acid sequences of the CSαβ peptides from different evolutionary lineages. N.D., not determined; SCT, Na-channel toxin; DF, defensin; CCT, Cl-channel toxin; PCT, K-channel toxin; STP, sweet-tasting protein; TI, trypsin inhibitor; LVP1, a lipolysis-activating protein from scorpion venom.

coming from different orders of the subclass Neoptera (Diptera, Coleoptera, Hemiptera and Hymenoptera) [18]. They are primarily active against Gram-positive bacteria, and some have the ability to form voltage-dependent channels in bacterial membrane that result in a loss of cytoplasmic potassium. Two types of antifungal peptides (termicin and drosomycin) [19–21] have also been characterized recently from insects. An exhaustive search characterized 71 peptides potentially adopting this fold, including those with known functions (antibacterial and antifungal) as well as some new members with unknown functions, most of which are derived from three model organisms *(D. melanogaster, Anopheles gambiae,* and *Apis mellifera)*. The consensus sequence for insect CSαβ motif is C-X(4,16)-C-X(3)-C-X(3,10)-[GY]-X-C-X(7-8)-CXC.

Arachnida

Arachnida, represented by scorpions and ticks, is the richest resources of CSαβ peptides [8, 22, 23] among which a total of 304 peptides were searched. Functionally, these peptides can be divided into toxins and defensins. The former are derived from scorpion venoms and are active against various ion channels, including K-, Cl-, Na-, and Ca-channels [9, 23], whereas the latter exhibit clear antibacterial activities [22]. With respect to size, toxins form two distinct groups: one being Na- and Ca-channel-specific toxins with approximately 60 residues and four disulfide bridges [24], and the other being K- and Cl-channel-specific toxins with approximately 30–40 residues and three to four disulfide bridges [25]. The defensins derived from Arachnida hemolymph as well as scorpion venoms are considered to be involved

in host innate immunity response [18, 23, 26]. They have sequences and activities similar to the so-called ancestral defensins found in phylogenetically distant invertebrates (mussels, arachnids and the dragonfly Aeschnea cyanea). In addition to toxins and defensins, scorpion venoms also contain other types of CSαβ peptides, such as LVP1, a dimer composed of two chains structurally related to Na-channel toxins and functionally acting as a nontoxic lipolysis-activating protein [27]. A reorganization of the fourth disulfide bridge is widely present in the CSαβ peptides from Arachnida (see below). The consensus sequence for the Arachnida CSαβ motif is C-X(2,10)-C-X(3)-C-X(4,11)-C-X(4-12)-CXC.

Nematode

The CSαβ-type antimicrobial peptides in nematodes were originally described in the intestinal parasitic nematode *A. suum* by Kato et al. They were named ASABF (*Ascaris suum* antibacterial factor) and were later also found in two model nematodes (*C. elegance* and *C. briggsae*) [28–31]. ASABFs are highly effective against bacteria and yeasts. Our approaches characterized a total of 30 CSαβ peptides, most of which have not been functionally identified. All the previously characterized members possess four disulfide bridges which are identical to those of mollusca. However, a search of the genome sequences of the two model nematodes reveals new types of ASABF peptides with typically three disulfide bridges identical to insect defensins as well as those with a reorganization of the fourth disulfide bridge (see below). Interestingly, in contrast to those from other sources, the nematode CSαβ peptide gene contains an intron located within the coding region of mature toxins rather than in the signal sequence. The consensus sequence for the nematode CSαβ motif is C-X(3,18)-C-X(3)-C-X(7,9)-[GS]-X-C-X(4-13)-CXC.

Mollusca

Only five antimicrobial peptides were found to adopt the CSαβ motif, and they were all isolated from the hemolymph of two mussels, a bivalve mollusk [32]. These molecules can be classified into two distinct groups based on shared features of their primary structure. The first group comprises three defensins which show some sequence similarities with ancestral defensins from some arthropods. The other group of peptides includes the myticins A and B characterized from hemocytes of Mytilus galloprovincialis [32]. Despite the lack of detectable sequence similarity between these two groups, all five molecules display an identical cysteine arrangement pattern. Functionally, defensins and myticins are essentially active against Gram-positive bacteria, including some pathogens for marine invertebrates, and are much less active against Gram-negative bacteria or fungi. The consensus sequence for the mollusca CSαβ motif is C-X(4,5)-C-X(3)-C-X(4,6)-C-X(1,2)-[GY]-X-C-X(7-8)-CXC-X(2)-C.

Plant

A total of 138 sequences were recovered from various plants. All these sequences possess a typical six-cysteine arrangement and thus can induce the CSαβ fold [6, 33, 34]. A phylogenetic analysis revealed three distinct clusters, including antifungal peptides, trypsin inhibitors and peptides with unknown functions. The sweet-tasting protein [35] is included in the last category. The majority of peptides belong to the recently cloned low molecular weight, cysteine-rich gene family (LCR) derived from the model organism *Arabidopsis* [34]. The functions of these genes are unknown at present. Some plant CSαβ peptides deleted a cysteine in their N- or C-termini, which might lead to the formation of a dimer by establishing an intermolecular disulfide bridge, as observed in scorpion venom LVP1 peptides [27]. For the plant CSαβ motif, a consensus sequence can be determined as follows: C-X(2,8)-C-X(3)-C-X(7,10)-[GASWF]-X-C-X(4-17)-CXC.

Structural classification and functional epitopes

Evidence supporting a common ancestor for all the CSαβ peptides comes from the observation of a commonality in gene organization, three-dimensional (3D) structure and biological activity [36]. However, rapid sequence divergence during evolution has prevented us from performing evolutionary analyses based on sequence similarity to correctly detect the evolutionary relationship among different families within the CSαβ superfamily. Instead, structural information may be useful for this task [37, 38]. We used a structure similarity score to analyze currently available structures stored in the PDB database (http://www.ncbi.nlm.nih.gov). An evolutionary tree was constructed for 66 CSαβ peptides using DALI similarity scores for protein structures [37] (http://www.cebl.auckland.ac.nz/pal-project). This tree divides these structures into two distinct clusters (fig. 3). Cluster I includes the short-chain toxins derived from scorpion venoms which act on K- and Cl-channels, whereas cluster II comprises several families with extremely diverse biological activities, including plant trypsin inhibitors and sweet-tasting proteins, scorpion Na-channel toxins and plant/animal antimicrobial defensins (fig. 3). The further clustering of the CSαβ peptides with similar functions in the structure-based evolutionary tree suggests the structural data of the CSαβ superfamily can well be used to elucidate the evolution of their functional diversity.

Cluster I

K-channel toxins

Despite different targets, all scorpion K-channel toxins (PCTs) analyzed are grouped together with the scorpion Cl-channel toxins (CCTs) in the structure-based evolutionary tree. The PCTs represent the most diverse CSαβ family which may mirror the diversity of its targets and indicate a possible co-evolutionary relationship between the PCTs and K-channels [39].

Undoubtedly, scorpion PCTs also show diverse variation in their functional surface locations with both pore and turret activities. Mutational analysis combined with computational methods have contributed new data with respect to residues directly involved in recognizing and blocking K-channels and identified four different interacting modes for the interaction of PCTs with various subtypes of K-channels [40, 41]. An exhaustive mutational study of AgTx2 highlighted a functional surface important for binding to the Kv channel [42]. This surface is mainly formed by residues associated with two β-strands and their connecting loop, which includes a dyad composed of a Lys residue plus an aromatic amino acid for blocking Kv1-type channels.

Unlike Na-channel toxins, which are modulators of Na-channel function mediated by two distinct domains in their molecular surface (see below), the blocking activities of PCTs often appear to be associated with their secondary elements. A remarkable division of functional surface associated with secondary structure elements results in their pharmacological diversification [43]. Generally, the helical domain is important for binding to calcium-activated K-channels. In contrast, the sheet domain provides major contact residues for interaction with Kv channels. To devolop multiple pharmacological targets, some PCTs (e. g. MTx and BmTx3) have evolved two binding sites in the distinct secondary structure elements of the same molecule [43, 44].

Cl-channel toxins

Cl-channel toxins are small basic peptides which can block small-conductance Cl-channels and are toxic to arthropods [45]. Recently, a new target, named matrix metalloproteinase-2 (MMP-2), for chlorotoxin was found, which is specifically upregulated in gliomas and related cancers, but is not normally expressed in brain [46]. In addition to three common disulfide bridges of the CSαβ motif, chlorotoxins evolved a fourth bridge to put the small N-terminal β strand close to the helical region of the molecule [45]. Interestingly, we noticed that this additional disulfide bridge is identical in position and arrangement pattern to that of a plant defensin with five disulfide bridges (Phd1) [47] and scorpion toxins targeting ERG (ether-a-go-go related gene) K-channels [48]. Despite a closer relationship between CCTs and PCTs, as revealed by the structure-based tree, obvious electrostatic potential surface differences between them might be partially responsible for their functional diversification [45]. The functional surface of chlorotoxins has still not been elucidated.

Cluster II

Plant trypsin inhibitors

Plant CSαβ peptides with trypsin activity possess disulfide bridges identical to those with other functions. The first nuclear magnetic resonance (NMR) structure of a trypsin inhibitor with the CSαβ motif, Atti-2, from *Arabidopsis*, was recently reported by Zhao et al. [49]. The reactive site located in loop 1 shows a conformation similar to other inhibitors with a completely different fold type, suggesting that the emergence of this inhibitory loop may represent a major evolutionary event which determines its unique function among from other plant defensins. In the tree, Atti-2 forms a single branch, separated from all other members, suggesting a role for structural remodeling in shaping the new function.

Scorpion Na-channel toxins

An unexpected result from our structure-based evolutionary analysis is that scorpion sodium channel toxins (SCTs) are closer to plant/animal defensins than to scorpion short-chain toxins (fig. 3). This observation may be useful for detecting the evolutionary origin and mechanism of ion-channel toxins.

SCTs are longer than PCTs and CCTs, which are composed of about 61–76 residues with typically four disulfide-bridges, of which three are involved in the formation of the CSαβ motif, whereas the fourth links the N- and C-termini [24, 50]. In terms of pharmacological features, SCTs can be divided into two distinct groups, called α- and β-toxins. These two groups modulate the sodium current differently by binding to two distinct receptor sites on the sodium channels. The α-toxins cause a slowing of the inactivation process of sodium currents and a prolongation of action potential by binding to receptor site 3 of the voltage-gated sodium channel. The β-toxins cause the voltage-gated sodium channel to shift its activation voltage dependence to more negative membrane potentials and cause a reduction of peak current amplitude by binding to receptor site 4. Although the α- and β-toxins display detectable sequence similarity, and a good structural superimposition can be performed in their core CSαβ motifs, a dramatic difference can be found in their C-tails [51] which may be associated with their pharmacological versatility.

Thanks to work of several groups [52–58], the functional sites of scorpion Na-channel toxins have been elucidated in detail (fig. 3). All the results show that despite some

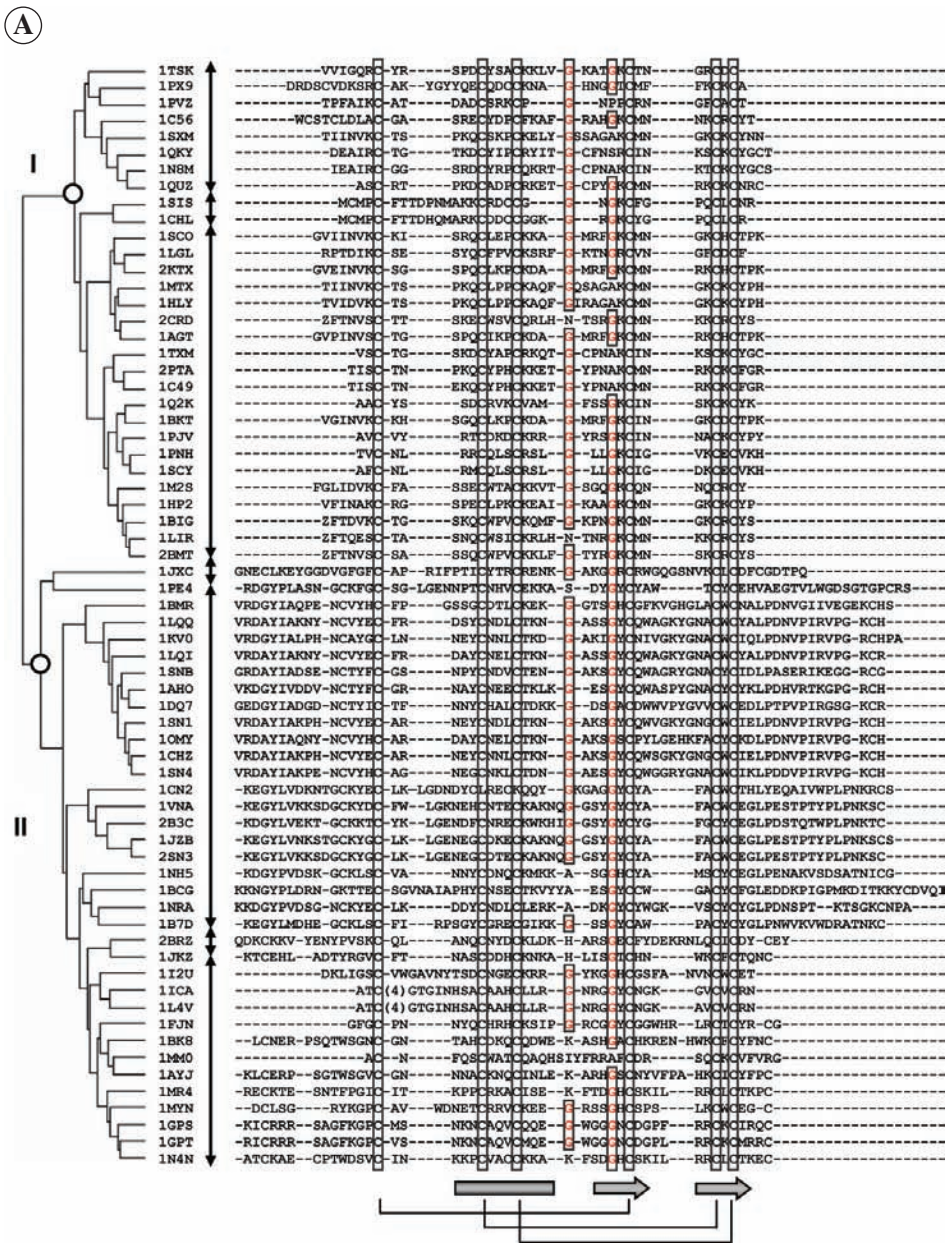

Figure 3. Structural classification and functional epitopes of the CSαβ peptides. (a) A complete linkage clustering tree was drawn using Pebble program with DALI structural similarity scores between the CSαβ structures. Antimicrobial peptides: 1fjn 1ica 1l4v 1mm0 1i2u 1myn 1gps 1gpt 1ayj 1bk8 1jkz 1mr4 1n4n. Na-channel toxins: 1aho 1sn1 1chz 1lqi 1lqq 1sn4 1kv0 1omy 1snb 1dq7 1bmr 1jzb 2sn3 1vna 2b3c 1cn2 1b7d 1nh5 1nra 1pe4 1bcg. K-channel toxins: 1q2k 1pjv 1pnh 1scy 1lir 2crd 2bmt 1big 1lgl 2ktx 1agt 1sco 1bkt 1hly 1mtx 1sxm 1hp2 1tsk 1c56 1m2s 1c49 2pta 1quz 1qky 1txm 1n8m 1acw 1wm7 1du9 1wm8 1pvz 1px9. Cl-channel toxins: 1chl 1sis. Sweet-tasting protein: 2brz. Trypsin inhibitor: 1jxc; (b) Functional epitopes of the representatives of the CSαβ superfamily based on mutational analysis [52–59]. Colors represent the functional surfaces. In the most examples the surfaces can be separated as two discrete sites.

changes in the location of functional surfaces, it appears that scorpions have adopted a common strategy, as observed in other venomous organisms, to design the functional surfaces of their toxins for highly efficient targeting of various types of Na-channels from phylogenetically diverse organisms, including both predator and prey. The functional sites are mainly composed of two distinct domains comprising one 'core domain' containing the so-called pharmacophore that is conserved across one pharmacological group [52–54]. In contrast, the 'variable domain' is most likely the determinant of the specificity of Na-channel subtypes. For α-toxins, the conserved core domain is situated in the loops connecting the secondary structure elements of the molecule core. The variable NC-domain is formed by a five-residue turn (residues 8–12) and a C-terminal segment (residues 56–64). For β-toxins, the functional surface includes: (i) a cluster of residues associated with the α-helix, which includes a putative 'hot spot'. This cluster is conserved among scorpion β-toxins and contains their pharmacophore; (ii) a hydrophobic cluster associated mainly with the β2 and β3 strands, which is likely to confer specificity for mammalian Na-channels; (iii) a single bioactive residue (Trp-58) in the C-tail; and (iv) a negatively charged residue (Glu-15) involved in voltage sensor trapping [54]. Anti-insect excitatory toxin, also classified as β-group due to similar pharmacological features, conserved its core domain to anti-mammal β-toxin, which is composed of the clustered residues around the main α-helical motif, whereas the variable domain responsible for specificity is

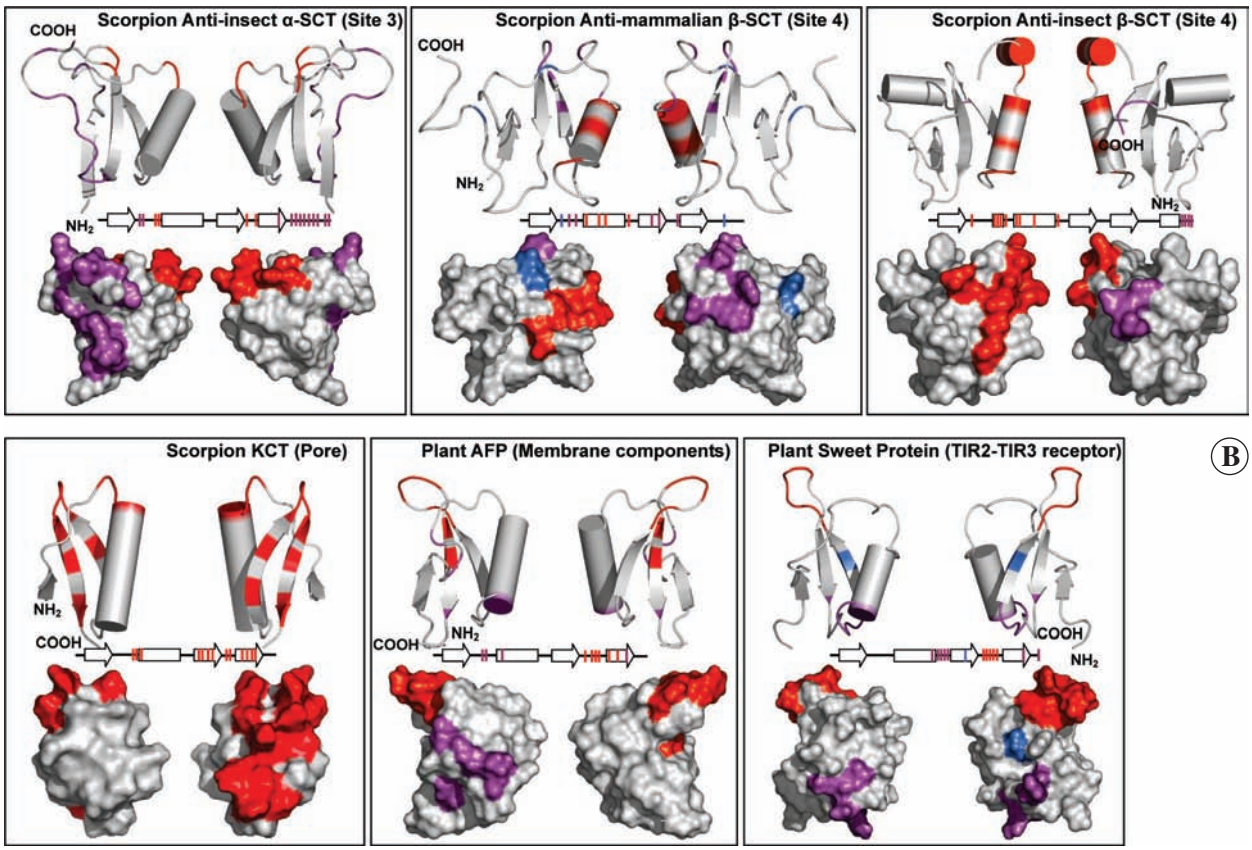

Figure 3B.

formed by a stretch of hydrophobic residues positioned on the C-tail [52].

Antimicrobial peptides

Defensins with the CSαβ motif are only one class of antimicrobial peptides conserved throughout the plant and animal kingdoms [6, 7, 9]. These molecules can be classified into two distinct groups on the basis of their diverse biological activities: antibacterial defensins derived from phylogenetically diverse organisms, including insects, scorpions and mussels, and antifungal peptides derived from plants and insects.

Little is known about the functional surface of defensins. Some structure-function relationship studies highlight the importance of the solvent-exposed site located in the loop and turn. This contrasts strongly with PCTs that use secondary elements as the major source of functional sites. Mutational analysis of a plant defensin, Rs-AFP2, identified its putative functional surfaces important for antifungal activity. This surface is primarily located in the loop connecting strands β2 and β3, the loop connecting the β1 strand and the α-helix, and the β3 strand. Spatially, these residues form two adjacent sites which probably facilitate the interaction with fungus in two separate regions [59]. The pattern of functional site distribution is similar to that of the SCTs and sweet-tasting protein [37, 52–54]. Interestingly, a recent study of a series of synthetic peptides revealed a key role of the same loop connecting the two β strands for targeting Gram-positive bacteria in a mussel defensin [60]. It thus appears that the highly exposed loop of the CSαβ structural motif is crucial for defensin activity against both bacteria and fungi.

Sweet-tasting protein

The fruit of *Pentadiplandra brazzeana* Baillon contains a small, sweet-tasting protein named brazzein. The NMR structure revealed that brazzein contains one short α-helix and three strands of an anti-parallel β-sheet held together by four disulfide bonds and thus displays the typical CSαβ fold [61]. Interestingly, in the structure-based tree, brazzein is clustered with defensins and shares the most significant structural similarity with the plant antifungal peptide Psd I (*Pisum sativum* defensin I) (fig. 3).

Mutational analysis on brazzein shows that the sweet determinants are located in at least two regions of the peptide surface, which are apparently associated with two loops connecting the secondary structure elements as well as the C-terminus, and thus suggest a multi-point interaction between brazzein and its receptor in which charge plays a significant role [37].

Remarkably, comparison of the functional surfaces reveals an interesting feature in terms of their location in the secondary structure elements of peptides from Cluster I and II (fig. 3). In most cases peptides from Cluster I developed their functional surfaces in the α-helix and β-strand regions. In contrast, those from Cluster II primarily used residues in the loops to form their functional surfaces separated as two discrete sites. The evolutionary significance behind it remains unsolved at present.

Evolution of functional diversity

Despite the difficulty in elucidating the evolutionary mechanism responsible for functional diversity of the CSαβ superfamily, it appears that sequence and structure information can provide some new clues to approach this question.

Role of disulfide bridges

The importance of disulfide bridges in shaping a protein structure has well been documented [62]. However, their role in the functional evolution of proteins is less well understood. This may partially be because traditionally disulfide bridge-formed cysteine residues are not subject to mutation considering their roles in protein structural stability. In most cases the disulfide bridges formed by specific pairing of cysteines are located in protein cores and obviously play a structural role. In contrast to proteins, cystines that form disulfide bridges in some small polypeptides (e. g. animal toxins) often expose their side chains to the molecular surface and thereby likely contribute to the functional site. The Janus-faced atracotoxins from spiders are examples in which double cystines form a vicinal disulfide bridge and are situated at the region of the functional sites [63]. Generally, the highly conserved disulfide bridges crucial for folding of a distinct structural element essentially confer protection from degradation by some proteinases rather than represent a functional role. However, additional disulfide bridges excluded in peptide folding may be involved in functional evolution. An excellent study in this respect was performed by Sun et al. who defined the respective roles of four disulfide bridges of BmKM1 (a scorpion Na-channel toxin) by mutational analysis [64]. Their results showed that three disulfide bridges (C2/C5, C3/C6 and C4/C7), which are highly conserved across the whole CSαβ superfamily, are involved in motif folding and play an obvious structural role. On the contrary, C1/C8 (the fourth disulfide bridge), which varies between different types of scorpion toxins, essentially contributes to the functional performance of scorpion toxins (fig. 4).

As observed in the CSαβ superfamily, the position alterations of the fourth disulfide bridge often correlate with functional diversification (fig. 4). It is especially clear in the case of scorpion CSαβ toxins. These molecules have a total of eight different pairings in their fourth disulfide bridges. The correlation between disulfide bridge pairings and unique functions includes: (i) pairing IV–1 is typical for SCTs (α- and β-groups), with the exception of excitatory toxins, in which a shift of Cys1 to Cys5 (pairing IV-7) relative to IV-1 might result in functional alteration; (ii) adding a disulfide bridge (pairing IV-5) relative to the PCTs with three disulfide bridges led to the emergence of new targets (ERG K-channel and Cl-channels); (iii) the scorpion venom lipolysis-activating peptide LVP1 provides another example with respect to this correlation. This peptide possesses a heterodimeric structure composed of α and β chains cross-linked by an interchain disulfide bridge. Both α and β chains of LVP1 share detectable sequence and structural similarity to scorpion SCTs. However, deletion of the first cysteine in the N-terminus resulted in the loss of the fourth intramolecular disulfide bridge. Instead, a dimer is formed by an intermolecular bridge linked by the two cysteines in the C-termini [29]. This change thus led to a loss of Na-channel toxin activity and a gain of a new function. Similarly, several SCT-like peptides deduced from their complementary DNA (cDNA) sequences display a de-

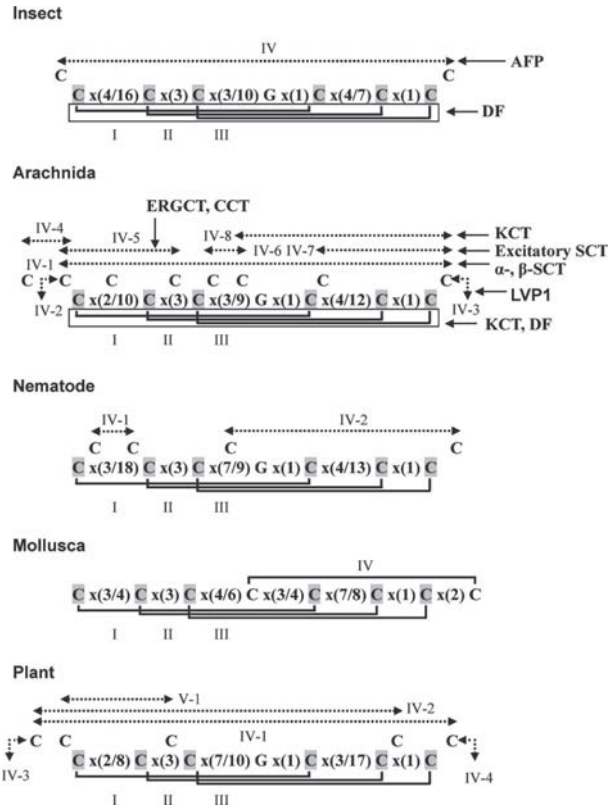

Figure 4. Disulfide bridge pattern in the CSαβ peptides of a given evolutionary lineage. Dotted lines represent the fourth disulfide bridge, which is variable between different members.

leted cysteine in the last C-terminus, which might have caused them to lose their toxic activity and to develop new functions [65].

In addition to scorpion venom peptides, similar cases are also found in CSαβ peptides from other species. For example, in *Drosophila* the development of the fourth disulfide bridge may be responsible for a functional change from targeting bacteria (defensins) to targeting fungi (drosomycins) [18]. In these examples, a clear paralogous relationship between the defensins and drosomycins highlights a possible role of disulfide bridge reorganization in functional evolution after gene duplication.

Given that the CSαβ peptides with identical disulfide bridges can exhibit diverse biological functions and that the fourth reorganized disulfide bridge is not often located in a functional site, this change alone cannot be considered to be an indicator of functional evolution. However, such a change may lead to functional alteration of a peptide by affecting the spatial distribution of key residues in functional sites.

Positive Darwinian selection

Adaptive molecular evolution driven by positive Darwinian selection was originally recognized in the antigen recognition site (ARS) of the major histocompatibility complex (MHC) loci and later frequently observed in molecules involved in both innate and acquired immunity [66–68]. For the effector molecules of the innate immunity, this mechanism is likely to be of considerable importance given that adaptive changes to those molecules have a direct role in targeting rapidly evolving pathogens. Most members of the CSαβ superfamily play a role in defense and immunity and thus represent excellent examples for detecting such evolutionary events. Using the maximum likelihood methods developed by Yang et al., Zhu et al. have provided clear statistical evidence for adaptive molecular evolution of scorpion Na-channel toxin genes [69]. Because sodium channels have many subtypes and are distributed in different tissues (cardiac and skeletal muscles, neurons), accelerated evolution driven by positive Darwinian selection is needed for expanding their target range to become a more efficient killer and have a better defense against predators. Another CSαβ example comes for the study of Bulmer and Crozier who found that termite antifungal peptides – termicins – evolved by duplication and diversifying selection [20, 70]. In termicins, selection appears to be largely directed at reducing the charge of three externally exposed sites which may interact with a fungal binding site that has shifted charge to achieve resistance.

In order to detect whether other types of CSαβ peptides have also been subjected to positive Darwinian selection, we analyzed several other gene families using maximum likelihood methods [20, 69]. These include Cl-channel toxin, K-channel toxin, antifungal peptide drosomycin and insect depressant Na-channel toxin genes (fig. 5). As expected, results again indicate that positive Darwinian selection has driven the adaptive molecular evolution of these families. These results will be useful in understanding the evolution of the CSαβ superfamily from a mechanistic perspective. Since previous studies have revealed that positively selected sites were frequently found in the functional surface of the proteins studied [69], it is reasonable to assume the same for these new examples, and thus provide a good starting point for further experimental research into the evolution of the functional sites of these CSαβ peptides.

Origin of toxic peptides

The evolutionary origin of scorpion toxins acting on various ion channels remains a mystery. Similarities in gene organization, 3D structure and precursor architecture provide strong evidence for a common ancestor for defensins and ion channel toxins derived from scorpions [71]. A closer relationship, as revealed by our structure-based evolutionary analysis, between scorpion Na-channel toxins and plant/animal antimicrobial defensins provides new insights into the origin of scorpion toxic peptides. When using antifungal defensins as probes to carry out BLAST searches, we not only retrieved homologues of antifungal defensins from plants and insects but also some scorpion Na-channel toxins with E-values ranging from 0.006 to 0.1. In particular, a scorpion depressant toxin named BmKITc (AF073899) shares about 50% sequence similarity with drosomycins in the region corresponding to the CSαβ motif (fig. 6). The wide existence of the CSαβ peptides with antimicrobial activity in both plants and animals suggests that their ancient feature can be traced to 1,576 million years ago when animals diverged from plants. Scorpions diverged from other Arachnida about 400 million years ago [72] when Na-channel toxins first emerged from antimicrobial defensins and likely represent the ancestor of all scorpion toxins. Subsequently, a CC deletion event, as suggested by Ceard et al. [65], produced the ancestor of short-chain toxins. Wide sequence divergence driven by positive Darwinian selection expanded the family into new pharmacological groups.

Compared with antifungal defensins, Na-channel toxins appear to have extended their C-termini with a 14-residue segment and reorganized their fourth disulfide bridges during evolution (fig. 6). Intriguingly, the starting point for such an extension begins at Leu, a position putatively for evolutionary modeling of scorpion short-chain toxin genes by a position-specific deletion mechanism [65]. Because mutational studies have confirmed that the extension C-terminal region in SCTs plays a crucial role for these molecules in interacting with their targets (fig. 3) [52, 53], we hypothesize that such an extension event

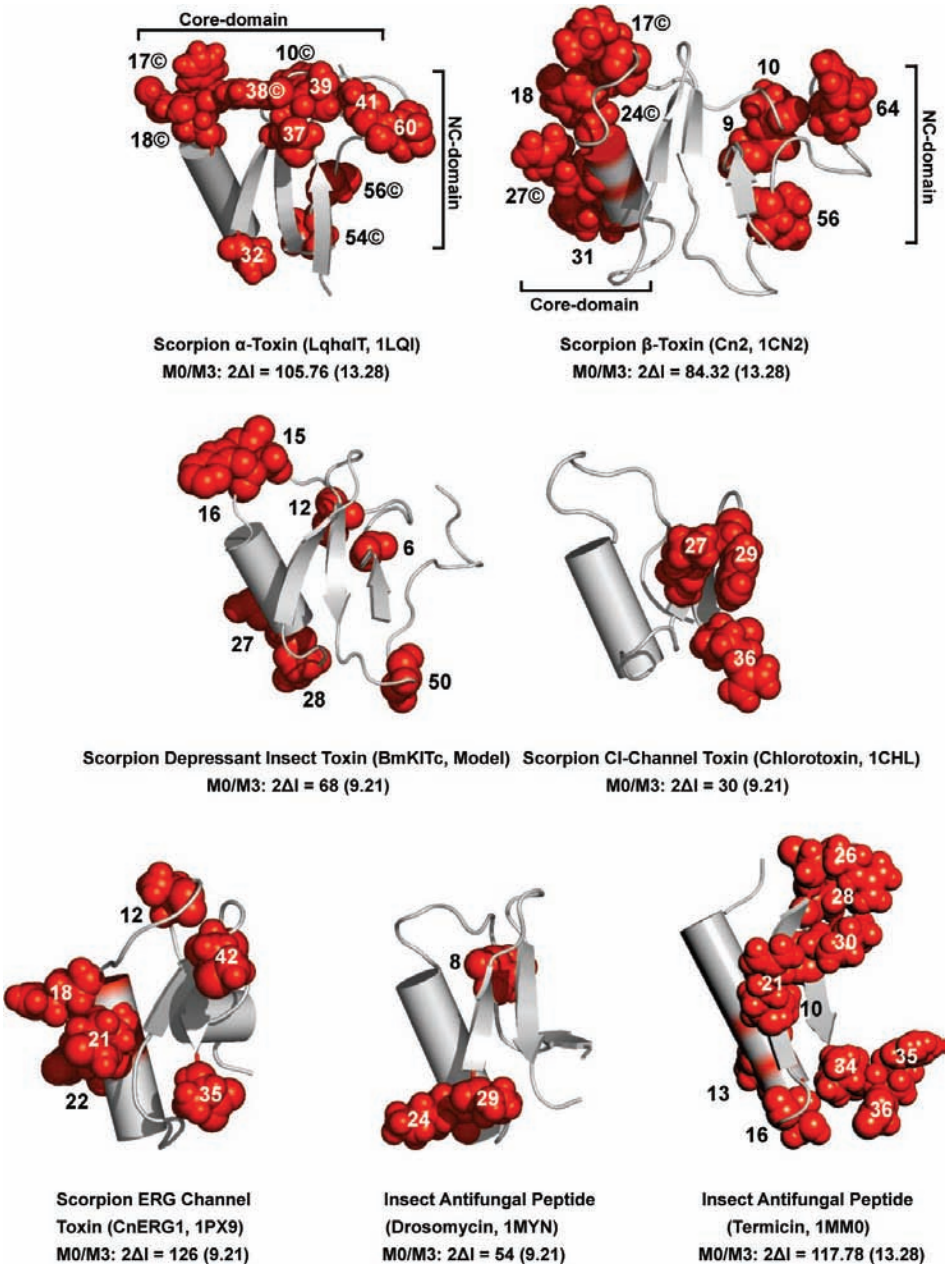

Figure 5. Adaptive evolution in the CSαβ peptides detected by maximum likelihood (ML) methods. Sites under positive Darwinian selection characterized by model M3 are highlighted in red sphere models. Data for scorpion α- and β-toxins are from [69] (the symbol © represents the functional sites identified by mutational analysis [53, 54]); for insect antifungal peptide, termicin is from [20]. Other data are from this study.

would result in two distinctive effects, namely disruption of the functional sites of defensin by partially masking the key loop and remodeling a new surface in this extension loop for targeting Na-channels. Certainly, this assumption needs to be further confirmed by experimental data.

Concluding remarks

As a versatile structural template, the CSαβ fold has developed many diverse biological functions through evolution to participate in both defense and predatory behavior. Several experimental studies can be logically followed from the observations presented here. They are (i) the discovery of new members and determination of the phylogenetic distribution of the CSαβ fold provides a research platform for further understanding of this unique fold from a phylogenetic and evolutionary perspective; (ii) the existence of wide positive Darwinian selection in the CSαβ fold suggests that adaptive amino changes in a conserved scaffold are a major force driving new functional emergence, and undoubtedly useful for nature-guided drug design; (iii) the recognition of a putative evolutionary route for the origin of toxic peptides provides a possibility for experimentally evolutionary mimicry to fulfill functional switch; (iv) by establishing one or two additional disulfide bridges or by displacing

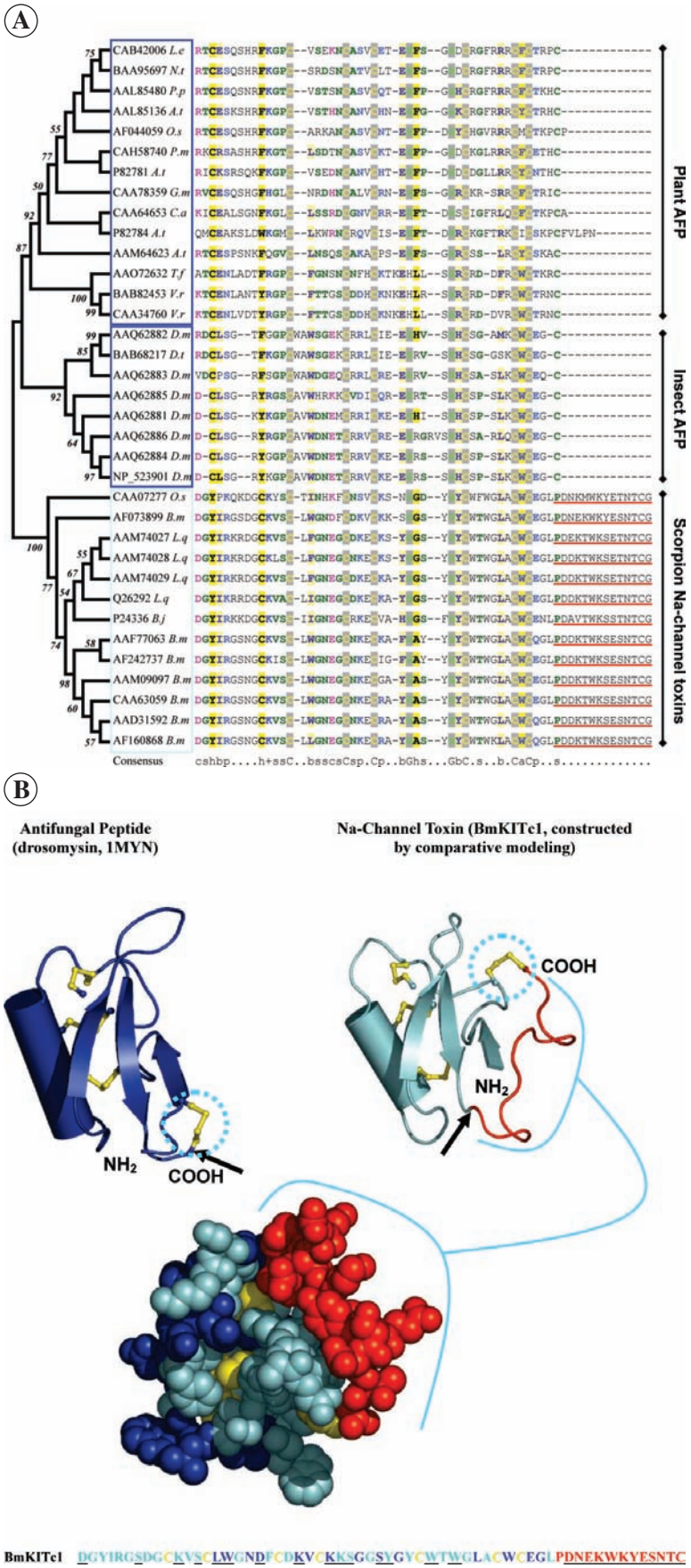

Figure 6. Evidence for the origin of scorpion toxins from antifungal defensins. (a) Phylogenetic analysis and sequence alignment. Organisms: L.e, L. esculentum;. N.t, N. tabacum; P.p, P. persica; A.t, A. thaliana; O.s, O. sativa; P.m, P. major; G.m, G. max; C.a, C. annuum; T.f, T. foenum; V.r, V. radiate; D.m, D. melanogaster; D.t, D. triauraria; B.m, B. martensii; L.q, L. quinquestriatus; B.j, B. judaicus. Consensus: a, aromatic; b, big; c, charged; h, hydrophobic; p, polar; s, small; +, positive. (b) Structural analysis emphasized a key extension event in the C-terminus of defensins (indicated by arrows) together with reorganization of the fourth disulfide bridges (indicated by dotted circles), which might lead to its functional change to Na-channel toxins. In BmKITc1 sequence and its model structure, identical residues (not including six conserved Cys residues) with drosomycin 5 are highlighted in blue. Residues on the molecular surface with >30% accessibility calculated by the Swiss-PDBViewer program (http://us.expasy.org/spdbv/) are underscored once. The extension 14-residue segment in BmKITc1is indicated in red.

two discontinuous functional sites which do not affect the global fold, we can try to develop new peptides with completely different biological activities.

Acknowledgments. Zhu would like to thank the support of the Bairen Plan from the Chinese Academy of Sciences, China.

1 Ponting C. P. and Russell R. R. (2002) The natural history of protein domains. Annu. Rev. Biophys. Biomol. Struct. **31:** 45–71
2 Anantharaman V., Aravind L. and Koonin E. V. (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. Curr. Opin. Chem. Biol. **7:** 12–20
3 Todd A. E., Orengo C. A. and Thornton J. M. (2002) Sequence and structural differences between enzyme and nonenzyme homologs. Structure **10:** 1435–1451
4 Nagano N., Orengo C. A. and Thornton J. M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. **321:** 741–765
5 Bontems F., Roumestand C., Gilquin B., Menez A., and Toma F. (1991) Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. Science **254:** 1521–1523
6 Cornet B., Bonmatin J. M., Hetru C., Hoffmann J. A., Ptak M. and Vovelle F. (1995) Refined three-dimensional solution structure of insect defensin A. Structure **3:** 435–448
7 Lay F. T. and Anderson M. A. (2005) Defensins-components of the innate immune system in plants. Curr. Protein Pep. Sci. **6:** 1–17
8 Landon C., Barbault F., Legrain M., Menin L., Guenneugues M., Schott V. et al. (2004) Anti-microbial peptides: from invertebrates to vertebrates. Immunol. Rev. **198:** 169–184
9 Thevissen K., Warnecke D. C., Francois I. E., Leipelt M., Heinz E., Ott C. et al. (2004) Defensins from insects and plants interact with fungal glucosylceramides. J. Biol. Chem. **279:** 3900– 3905
10 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
11 Vita C., Roumestand C., Toma F. and Menez A. (1995) Scorpion toxins as natural scaffolds for protein engineering. Proc. Natl. Acad. Sci. USA **92:** 6404–6408
12 Menez A., Bontems F., Roumestand C., Gilquin B. and Toma F. (1992) Structural basis for functional diversity of animal toxins. Proc.R. Soc. Edinburgh **99B:** 83–103
13 Fry B. G. (2005) From genome to venome: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res. **15:** 403–420
14 Zhu S., Darbon H., Dyason K., Verdonck F. and Tytgat J. (2003) Evolutionary origin of inhibitor cystine knot peptides. FASEB J. **17:** 1765–1767
15 Crooks G. E., Hon G., Chandonia J.-M. and Brenner S. (2004) WebLogo: a sequence logo generator. Genome Res. **14:** 1188–1190
16 Strickberger M. W. (2000) Evolution, 3rd edn., Jones and Bartlett Publishers, Boston, MA
17 Lehrer R. I. (2004) Primate defensins. Nat. Rev. Microbiol. **2:** 727–738
18 Dimarcq J. L., Bulet P., Hetru C. and Hoffmann J. (1998) Cysteine-rich antimicrobial peptides in invertebrates. Biopolymers **47:** 465–477
19 Da Silva P., Jouvensal L., Lamberty M., Bulet P., Caille A. and Vovelle F. (2003) Solution structure of termicin, an antimicrobial peptide from the termite Pseudacanthotermes spiniger. Protein Sci. **12:** 438–446
20 Bulmer M. S. and Crozier R. H. (2004) Duplication and diversifying selection among termite antifungal peptides. Mol. Biol. Evol. **21:** 2256–2264
21 Landon C., Sodano P., Hetru C., Hoffmann J. and Ptak M. (1997) Solution structure of drosomycin, the first inducible antifungal protein from insects. Protein Sci. **6:** 1878–1884
22 Nakajima Y., van der Goes van Naters-Yasui A., Taylor D. and Yamakawa M. (2001) Two isoforms of a member of the arthropod defensin family from the soft tick, Ornithodoros moubata (Acari: Argasidae). Insect Biochem. Mol. Biol. **31:** 747–751
23 Cociancich S., Goyffon M., Bontems F., Bulet P., Bouet F., Menez A. et al. (1993) Purification and characterization of a scorpion defensin, a 4kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins. Biochem. Biophys. Res. Commun. **194:** 17–22
24 Gurevitz M., Froy O., Zilberberg N., Turkov M., Strugatsky D., Gershburg E. et al. (1998) Sodium channel modifiers from scorpion venom: structure-activity relationship, mode of action and application. Toxicon **36:** 1671–1682
25 Possani L. D., Merino E., Corona M., Bolivar F. and Becerril B. (2000) Peptides and genes coding for scorpion toxins that affect ion-channels. Biochimie **82:** 861–868
26 Rodriguez de la Vega R. C., Garcia B. I., D Ambrosio C., Diego-Garcia E., Scaloni A. and Possani L. D. (2004) Antimicrobial peptide induction in the haemolymph of the Mexican scorpion Centruroides limpidus limpidus in response to septic injury. Cell. Mol. Life Sci. **61:** 1507–1519
27 Soudani N., Gharbi-Chihi J., Srairi-Abid N., Yazidi C. M., Planells R., Margotat A. et al. (2005) Isolation and molecular characterization of LVP1 lipolysis activating peptide from scorpion *Buthus occitanus tunetanus*. Biochim. Biophys. Acta **1747:** 47–56
28 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
29 Kato Y., Aizawa T., Hoshino H., Kawano K., Nitta K. and Zhang H. (2002) abf-1 and abf-2, ASABF-type antimicrobial peptide genes in *Caenorhabditis elegans*. Biochem. J. **361:** 221–230
30 Kato Y. and Komatsu S. (1996) ASABF, a novel cysteine-rich antibacterial peptide isolated from the nematode *Ascaris suum*. Purification, primary structure and molecular cloning of cDNA. J. Biol. Chem. **271:** 30493–30498
31 Zhang H. and Kato Y. (2003) Common structural properties specifically found in the CSαβ-type antimicrobial peptides in nematodes and mollusks: evidence for the same evolutionary origin? Dev. Comp. Immunol. **27:** 499–503
32 Mitta G., Vandenbulcke F. and Roch P. (2000) Original involvement of antimicrobial peptides in mussel innate immunity. FEBS Lett. **486:** 185–190
33 Thomma B. P. H. J., Cammue B. P. A. and Thevissen K. (2002) Plant defensins. Planta **216:** 193–202
34 Vanoosthuyse V., Miege C., Dumas C. and Cock J. M. (2001) Two large *Arabidopsis thaliana* gene families are homologus to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. Plant Mol. Biol. **16:** 17–34
35 Jin Z., Danilova V., Assadi-Porter F. M., Aceti D. J., Markley J. L. and Hellekant G. (2003) Critical regions for the sweetness of brazzein. FEBS Lett. **544:** 33–37
36 Froy O. and Gurevitz M. (1998) Membrane potential modulators: a thread of scarlet from plants to humans. FASEB J. **12:** 1793–1796
37 Dietmann S. and Holm L. (2001) Identification of homology in protein structure classification. Nat. Struct. Biol. **8:** 953–957
38 Dietmann S., Fernandez-Fuentes N. and Holm L. (2002) Automated detection of remote homology. Curr. Opin. Struct. Biol. **12:** 362–367

39 Rodriguez de la Vega R. C. and Possani L. D. (2004) Current views on scorpion toxins specific for K^+-channels. Toxicon **43:** 865–875

40 Rodriguez de la Vega R. C., Merino E., Becerril B. and Possani L. D. (2003) Novel interactions between K^+ channels and scorpion toxi ns. Trends Pharmacol. Sci. **24:** 222–227

41 Xu C. Q., Zhu S. Y., Chi C. W. and Tytgat J. (2003) Turret and pore block of K^+ channels: what is the difference? Trends Pharmacol. Sci. **24:** 446–448

42 MacKinnon R., Cohen S. L., Kuo A., Lee A. and Chait B. T. (1998) Structural conservation in prokaryotic and eukaryotic potassium channels. Science **280**: 106–109

43 Regaya I., Beeton C., Ferrat G., Andreotti N., Darbon H., De Waard M. et al. (2004) Evidence for domain-specific recognition of SK and Kv channels by MTX and HsTx1 scorpion toxins. J. Biol. Chem. **279:** 55690–55696

44 Huys I., Xu C. Q., Wang C. Z., Vacher H., Martin-Eauclaire M. F., Chi C. W. et al. (2004) BmTx3, a scorpion toxin with two putative functional faces separately active on A-type K^+ and HERG currents. Biochem. J. **378:** 745–752

45 Lippens G., Najib J., Wodak S. J. and Tartar A. (1995) NMR sequential assignments and solution structure of chlorotoxin, a small scorpion toxin that blocks chloride channels. Biochemistry **34:** 13–21

46 Deshane J., Garner C. G., and Sontheimer H. (2003) Chlorotoxin inhibits glioma cell invasion via matrix metalloproteinase-2. J. Biol. Chem. **278:** 4135–4144

47 Janssen B. J. C., Schirra H. J., Lay F. T., Anderson M. A. and Craik D. J. (2003) Structure of Petunia hybrida defensin 1, a novel plant defensin with five disulfide bonds. Biochemistry **42:** 8214–8222

48 Gurrola G. B., Rosati B., Rocchetti M., Pimienta G., Zaza A., Arcangeli A. et al. (1999) A toxin to nervous, cardiac, and endocrine ERG K^+ channels isolated from *Centruroides noxius* scorpion venom. FASEB J. **13:** 953–962

49 Zhao Q., Chae Y. K. and Markley J. L. (2002) NMR solution structure of ATTp, an *Arabidopsis thaliana* trypsin inhibitor. Biochemistry **41:** 12284–12296

50 Possani L. D., Becerril B., Delepierre M., and Tytgat J. (1999) Scorpion toxins specific for Na^+-channels. Eur. J. Biochem. **264:** 287–300

51 Gurevitz M., Gordon D., Ben-Natan S., Turkov M. and Froy O. (2001) Diversification of neurotoxins by C-tail 'wiggling': a scorpion recipe for survival. FASEB J. **15:** 1201–1205

52 Cohen L., Karbat I., Gilles N., Froy O., Corzo G., Angelovici R. et al. (2004) Dissection of the functional surface of an anti-insect excitatory toxin illuminates a putative hot spot common to all scorpion beta-toxins affecting Na^+ channels. J. Biol. Chem. **279:** 8206–8211

53 Karbat I., Frolow F., Froy O., Gilles N., Cohen L., Turkov M. et al. (2004) Molecular basis of the high insecticidal potency of scorpion alpha-toxins. J. Biol. Chem. **279:** 31679–31686

54 Cohen L., Karbat I., Gilles N., Ilan N., Benveniste M., Gordon D. et al. (2005) Common features in the functional surface of scorpion beta-toxins and elements that confer specificity for insect and mammalian voltage-gated sodium channels. J. Biol. Chem. **280:** 5045–5053

55 Wang C. G., Gilles N., Hamon A., Le Gall F., Stankiewicz M., Pelhate M. et al. (2003) Exploration of the functional site of a scorpion alpha-like toxin by site-directed mutagenesis. Biochemistry **42:** 4699–4708

56 Sun Y. M., Bosmans F., Zhu R. H., Goudet C., Xiong Y. M., Tytgat J. et al. (2003) Importance of the conserved aromatic residues in the scorpion alpha-like toxin BmK M1: the hydrophobic surface region revisited. J. Biol. Chem. **278:** 24125–24131

57 Liu L. H., Bosmans F., Maertens C., Zhu R. H., Wang D. C. and Tytgat J. (2005) Molecular basis of the mammalian potency of the scorpion alpha-like toxin, BmK M1. FASEB J. **19:** 594–596

58 Rogers J. C., Qu Y., Tanada T. N., Scheuer T., and Catterall W. A. (1996) Molecular determinants of high affinity binding of alpha-scorpion toxin and sea anemone toxin in the S3-S4 extracellular loop in domain IV of the Na^+ channel alpha subunit. J. Biol. Chem. **271:** 15950–15962

59 De Samblanx G. W., Goderis I. J., Thevissen K., Raemaekers R., Fant F., Borremans F. et al. (1997) Mutational analysis of a plant defensin from radish (*Raphanus sativus* L.) reveals two adjacent sites important for antifungal activity. J. Biol. Chem. **272:** 1171–1179

60 Romestand B., Molina F., Richard V., Roch P. and Granier C. (2003) Key role of the loop connecting the two beta strands of mussel defensin in its antimicrobial activity. Eur. J. Biochem. **270:** 2805–2813

61 Caldwell J. E., Abildgaard F., Dzakula Z., Ming D., Hellekant G. and Markley J. L. (1998) Solution structure of the thermostable sweet-tasting protein brazzein. Nat. Struct. Biol. **5:** 427–431

62 Chuang C.-C., Chen C.-Y., Yang J.-M., Lyu P.-C. and Hwang J.-K. (2003) Relationship between protein structures and disulfide-bonding patterns. Proteins **53:** 1–5

63 Maggio F. and King G. F. (2002) Scanning mutagenesis of a janus-faced atracotoxin reveals a bipartite surface patch that is essential for neurotoxic function. J. Biol. Chem. **277:** 22806–22813

64 Sun Y. M., Liu W., Zhu R. H., Goudet C., Tytgat J., and Wang D. C. (2002) Roles of disulfide bridges in scorpion toxin BmK M1 analyzed by mutagenesis. J. Pept. Res. **60:** 247–256

65 C ard B., Martin-Eauclaire M.-F. and Bougis P. E. (2001) Evidence for a position-specific deletion as an evolutionary link between long- and short-chain scorpion toxins. FEBS Lett. **494:** 246–248

66 Hughes A. L. and Nei M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170

67 Yang Z. (2005) The power of phylogenetic comparison in revealing protein function. Proc. Natl. Acad. Sci. USA **102:** 3179–3180

68 Mclnerney J. O., Littlewood D. T. J. and Creevey C. J. (2003) Detecting adaptive molecular evolution: additional tools for the parasitologist. Adv. Parasit. **54:** 359–379

69 Zhu S., Bosmans F. and Tytgat J. (2004) Adaptive evolution of scorpion sodium channel toxins. J. Mol. Evol. **58:** 145–153

70 Lamberty M., Zachary D., Lanot R., Bordereau C., Robert A., Hoffmann J. A. et al. (2001) Insect immunity. Constitutive expression of a cysteine-rich antifungal and a linear antibacterial peptide in a termite insect. J. Biol. Chem. **276:** 4085–4092

71 Froy O. and Gurevitz M. (2004) Arthropod defensins illuminate the divergence of scorpion neurotoxins. J. Pept. Sci. **10:** 714–718

72 Polis G. A. (ed.) (1990) The Biology of Scorpions, Stanford University Press, Stanford, Ca