FEBS Letters 494 (2001) 246–248 FEBS 24769

Correspondence

Evidence for a position-specific deletion as an evolutionary link between long- and short-chain scorpion toxins

Brigitte Céard, Marie-France Martin-Eauclaire, Pierre E. Bougis*

First published online 26 March 2001

Scorpions (Arthropoda, Chelicerata, Scorpionida, approximately 1500 species) evolved over 400 million years and many of their morphological aspects lead them to be considered as living fossils. They thrive in their natural biotope mainly due to specialised venom glands that secrete numerous neurotoxins and other non-toxic peptides. Most scorpion toxins can be classified as either long-chain toxins (60–70 amino acid residues), which target voltage-gated Na^+ channels, or short-chain toxins (less than 40 amino acid residues), which target K^+ channels [1]. All these toxins share a common structural motif [2], named CSH for cysteine-stabilised α-helix. Thus, scorpion venoms can be evolutionary eyed as successful combinatorial libraries of peptides for which the availability of specific targets outlines some of them as neurotoxins.

Initial studies on the genetics of scorpion toxins led to the suggestion of a one gene–one toxin organisational model, with the idea of a putative common ancestor leading to a hypervariable gene family [3]. A unique intron, located close to the end of the signal peptide, seems to be a conserved feature [4]. Furthermore, the amino acid sequences are more highly conserved in the signal peptides than in the toxins themselves. We took advantage of this conservation and found new clones that, we believe, define an evolutionary link between long- and short-chain toxins.

We used a two-step subtractive procedure to screen for new clones in the *Androctonus australis* (North Africa) cDNA library [3]. The initial screening, which was performed with a specific probe for part of the signal peptide (LVMISLALL; 5′-TTG GTA ATG ATT AGT TTG GCA CTT CTC-3′) of toxins AaH I (M27701), AaH I′ (M27702), AaH II (M27704), and AaH III (M27703), gave approximately 170 positive clones. The second screening, which used a mixture of probes specific to toxin AaH II, and toxins AaH I and AaH I′ (5′-TTG CCC GAT CAT GTA CGT ACT-3′ and 3′-TTG CCC GAT AAC GTA CCG ATT-5′, respectively), selected 10 of these clones by subtractive selection. Six of them had unique nucleotide sequences. According to the cloning strategy used, they are not related to any known toxin cDNA but possess an almost identical signal peptide.

The amino acid sequence encoded by pcD-1003 (AJ308439) matches the one determined by Edman sequencing for AaH IV (AAB30626). Its nucleotide sequence begins with AACAA, which is the transcription initiation site of the AaH I′ gene [4]. The C-terminal ends with basic amino acid residues, KR. This ending is not unusual and has been described in detail for *Androctonus* toxin precursors [3]. Comparison with the mature sequence showed that only the arginine residue was cleaved. Clones pcD-1005 (AJ308440) and pcD-985 (AJ308441) encoded two long-toxin homologues. The first is 80% similar to AaH II and contained five additional lysine residues, which increase the net positive charge from one to seven. This probably explains why this molecule was not found by use of cationic exchangers. Its amino acid sequence ended with NGR, suggesting that the asparagine residue could undergo α-amidation [3]. The second displayed 55% and 60% sequence identity with the BeM9 (P09981) and BeM14 (P09982) toxins from *Buthus eupeus* (Asia). These two new sequences also ended with an additional arginine residue.

The most interesting finding of our work concerned the three last clones, pcD-1008 (AJ308442), pcD-996 (AJ308443) and pcD-993 (AJ308444), which encode shorter peptides (53 amino acid residues). Closer comparison of their nucleotide sequences with pcD-1005 and pcD-985 revealed a unique 'CC' deletion at the proline codon of the LPD sequence that resulted in a frame shift. This frame shift resulted in a premature stop codon, such that protein translation ended by LR. As mentioned above, the presence of the arginine residue is also canonical. In the initial reading frame the putative amino acid sequences were consistent with long-chain toxins (LPDNAPIYDESKQCTRR, pcD-1008; LPDNAPIHDESKQCTRR, pcD-996 and LPDNSPIYAESKQCTRR, pcD-993). These truncated toxin homologues have never previously been described in *A. australis* venom.

It is known that position-specific codon conservation exists in hypervariable gene families, as recently documented in conotoxins [5]. We have aligned the amino acid sequences of 48 long-chain scorpion toxins from the SwissProt database and visualised at the best the consensus sequence (77 positions) using LOGO (Fig. 1a). As expected, cysteine residues were well-conserved, as they are responsible for the tight packing of the molecule. The C_1–C_8 bridge put together the N- and C-terminal regions. The toxin core sequence comprising the cysteine residues, C_2–C_7, contained a CSH motif: an α-helix linked to two anti-parallel β-sheets by three disulfide bridges. Accordingly, both long- and short-chain toxins can be defined by the PROSITE (http://www.expasy.ch/prosite/) signature: **C**-{C}(2,10)-**C**-{C}(3)-**C**-{C}(5,9)-[GPAS]-{C}-**C**-X-{C}(3,9)-**C**-[WY]-**C**. This signature gave 103 entries searching SwissProt and TrEMBL sequence data banks. These included a few γ-thionines and insect defensins, which are known to have structural identity with the scorpion toxin fold. There was also high variability in the inter-cysteine loops, which indicates that adaptive evolution has taken place as for the conotoxins, in which diversification operates at an extraordinary high rate [6]. However, there were position-specific conserved codons within the loops and, in peculiar, the leucine codon of the LPD sequence. If we look at the three-dimensional (3D) structures of both long-chain (AaH II) and short-chain (KTX) toxins, it can be seen that this peculiar leucine residue is at a 'hinge' position between the two structures (Fig. 1b). In an evolutionary context, we think that it is an interesting observation and believe that this is an evolutionary link between long- and short-chain scorpion toxins. Indeed, posi-

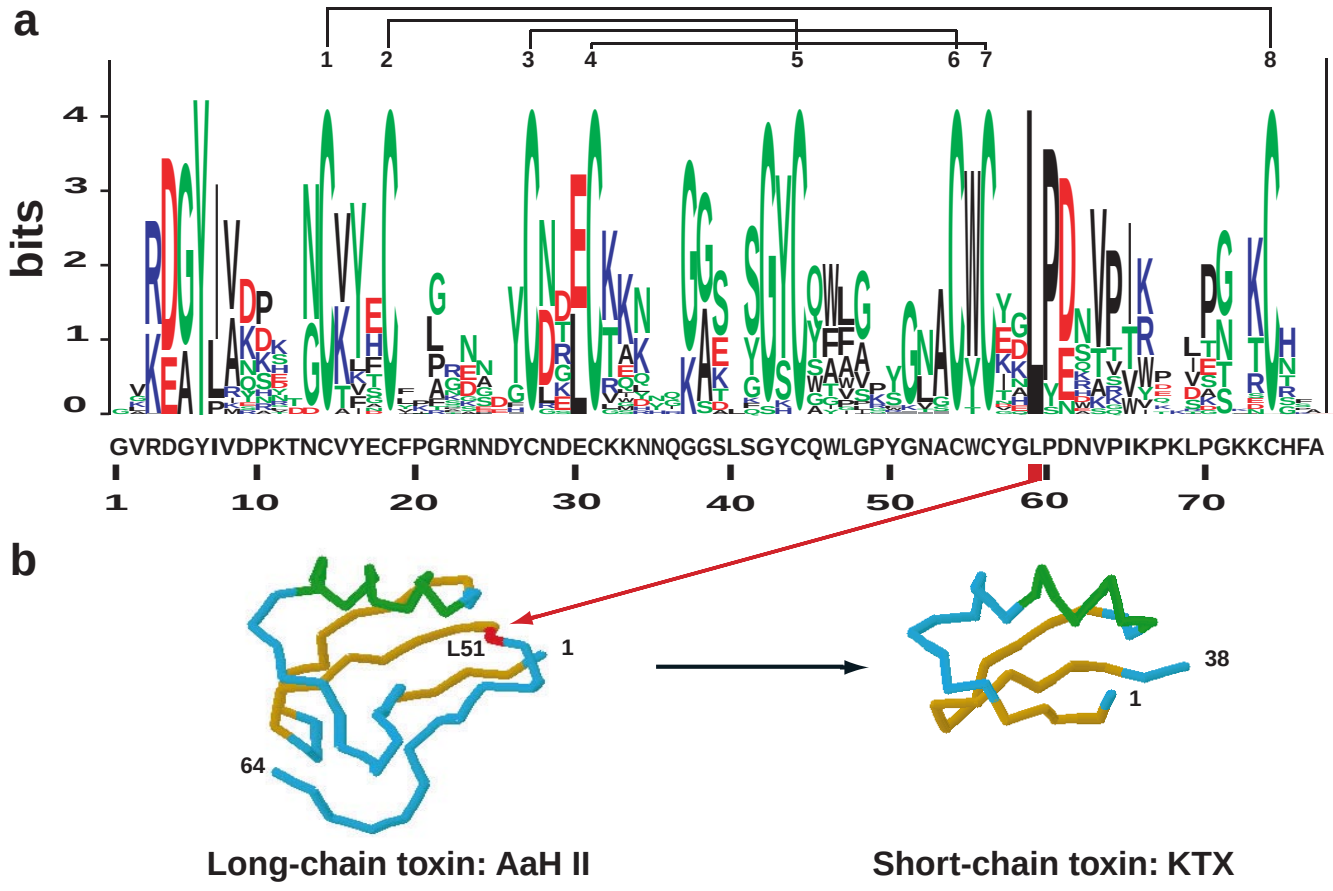

Fig. 1. a: Sequence logo (http://www.cbs.dtu.dk/gorodkin/appl/plogo.html) from alignment of 48 long-chain scorpion toxin sequences taken from SwissProt. Note the high variability in the inter-cysteine loops. Numbering is reported under the consensus sequence. The disulfide bridge organisation is indicated on the top of the sequence logo. b: Comparison of the respective structures of the long-chain toxin AaH II (pdb: 1AHO) and the short-chain toxin KTX (pdb: 2KTX) taken as examples. The drawings have been obtained using Cn3D viewer.

tional deletion/insertion mechanisms are known to be responsible for mutational or evolutionary modelling of genes. The 'CC' deletion observed here may result from such a mechanism.

To support this hypothesis of an evolutionary link between long- and short-chain toxins we constructed using PHYLIP (http://evolution.genetics.washington.edu/phylip.html) a phylogenetic tree, which takes the amino acid core sequence (CSH motif, positions 11–49 of AaH I) of all the *A. australis* toxins and homologues yet described into account. The rational was to align only the common structural part of long- and short-chain toxins. From the distinct clades observed, it is noteworthy that toxins active on insects, AaH IT1 (M27706) and AaH IT2 (M27707), have evolved distantly. The other toxins form three clades. The first two only contain long-chain toxins, clustering with the previously described AaH I and AaH II groups assuming both sequence and immunological criteria [1]. Truncated products of clones pcD-1008, pcD-996 and pcD-993 cluster with the short-chain toxin, KTX2 (S74733). Accordingly, this phylogenetic tree sustains our hypothesis reasonably.

Scorpion venoms, as those from sea cones, bees, spiders and snakes, are very successful combinatorial peptide libraries. On a structural basis, these peptides seem to have evolved under specific 3D constraints, mostly originating from the strict conservation of the cysteine residues. Indeed, like sea urchins, core/loop structures promote multiple point-to-point contacts with large target proteins (ionic channels, enzymes) in a much better way than achieved by scrambled peptides. The mechanism by which position-specific codon conservation occurs among DNA hypervariable regions is unknown. A 'lithographic mask' model has been proposed [5]. Another issue could be a structure quality control, occurring at the endoplasmic reticulum level and pushing incorrectly folded toxins resulting from mutational events (particularly at cysteine codons) out of the productive pathway. In this study, the position-specific deletion reported enlightens for a conserved leucine residue a structural role, in the context of scorpion toxin gene modelling during evolution.

This work was supported in part by a grant from the DRET.

References

[1] Martin-Eauclaire, M.F. and Couraud, F. (1995) in: Neurotoxi-

cology Handbook II: Effects and Mechanisms, Vol. 22, pp. 683–716, Chang Dyer, NY.
[2] Bontems, F., Roumestand, C., Boyot, P., Gilquin, B., Doljansky, Y., Menez, A. and Toma, F. (1991) Eur. J. Biochem. 196, 19–28.
[3] Bougis, P.E., Rochat, H. and Smith, L.A. (1989) J. Biol. Chem. 264, 19259–19265.
[4] Delabre, M.L., Pasero, P., Marilley, M. and Bougis, P.E. (1995) Biochemistry 340, 6729–6736.
[5] Conticello, S.G., Pilpel, Y., Glusman, G. and Fainzilber, M. (2000) Trends Genet. 16, 57–59.
[6] Duda Jr., T.F. and Palumbi, S.R. (1999) Proc. Natl. Acad. Sci. USA 96, 6820–6823.

*Corresponding author. Fax: (33)-4-91 69 88 39.
E-mail: bougis.p@jean-roche.univ-mrs.fr

UMR 6560 CNRS, Université de la Méditerranée, Institut Jean Roche, Faculté de Médecine Secteur Nord, 15 Bd. Pierre Dramard, F-13916 Marseille Cedex 20, France

PII: S0014-5793(01)02336-5