Volume 320, number 1, 43–46 FEBS 12270 March 1993

Primary structure determination and cloning of the cDNA encoding toxin 4 of the scorpion *Centruroides noxius* Hoffmann

A. Vázquez[a], B. Becerril[a], B.M. Martin[b], F. Zamudio[a], F. Bolívar[a] and L.D. Possani[a]

[a]*Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Cuernavaca 62271, México* and [b]*National Institute of Mental Health, Molecular Neurogenetics Unit, Clinical Neuroscience Branch, Building 10 3N256 Bethesda, MD 20892, USA*

Received 15 February 1993

A peptide (toxin II-10), shown to be a Na^+ channel blocker, was purified from the venom of the scorpion *Centruroides noxius* Hoffmann and sequenced by Edman degradation. It has 66 amino acid residues with the C-terminal residue (asparagine) amidated, as demonstrated by mass spectrometry. In addition, we report the cloning and the nucleotide sequence of the cDNA (CngtV) that codes for this toxin. We discuss the mechanism for processing the precursor peptide to its final form and compare the primary structure to that of other Na^+ channel toxins. Two distinct groups of toxins seem to emerge from this comparison, suggesting a structure–function relationship of these peptides towards the recognition of either mammalian or insect tissues.

Scorpion toxin; Na^+channel; cDNA clone; Nucleotide sequence; Peptide processing

1. INTRODUCTION

Scorpion toxins are exquisite tools for the study of ion channels [1]. At least two classes of different families of peptides have been purified and characterized from the venom of the scorpions of the family, Buthidae [2]. Among the toxins specific for the Na^+ channel is toxin II-10 from the Mexican scorpion *Centruroides noxius* Hoffmann. Initially, this peptide was purified by Sephadex G-50 gel filtration followed by ion-exchange chromatography on a carboxymethyl-cellulose column [2,3]. Its effect on Na^+ channels of excitable membranes was documented in squid axon membranes [4,5], and brain synaptosomes [6]. In this communication we describe the determination of its complete amino acid sequence, by automatic Edman degradation [7] of reduced and alkylated toxin, and of its peptides purified by high performance liquid cromatography (HPLC). We also describe the cloning of the cDNA encoding this toxin (CngtV), and discuss a possible mechanism for its post-translational processing.

2. MATERIALS AND METHODS

2.1. *Purification of the toxin Cn4.*

Venom from the scorpion *C. noxius* was obtained in the laboratory by electrical stimulation [3] and toxin Cn4 was purified by the same methods as previously described [3], except for the addition of an HPLC step at the end of the purification procedure, using published protocols [8].

Correspondence address: L.D. Possani, Instituto de Biotecnología-UNAM, Av. Universidad 2001, Apto. Postal 510–3, Cuernavaca 62271, Mexico. Fax: (52) (73) 172 388.

2.2. *Amino acid sequence determination*

Highly purified toxin was reduced and carboxymethylated (RC-toxin) as reviewed [2], prior to sequencing on a Millipore 6600 ProSequencer. RC-toxin Cn4 was digested (66 nmol) using protease V8 from *Staphylococcus aureus* (Boehringer-Mannheim). Another sample (30 nmol) was cleaved with trypsin (TPCK treated) from Sigma Co. (St. Louis, MO). Both digests were separated by HPLC. Several peptides obtained by this procedure were sequenced in order to obtain the primary structure of this toxin. The C-terminal peptide was also analysed by mass spectrometry (data not shown) in order to confirm the amidation of the last residue. Only analytical grade reagents were used, as earlier described [8].

2.3. *Construction and screening of a* C. noxius *cDNA library*

Isolation of total RNA, purification of the poly(A)$^+$ RNA, synthesis and construction of the cDNA library in λgt11 were performed according to the instructions supplied in the commercial kits utilized (Amersham, RPN.1264, RPN.1511, RPN 1256Y/Z, respectively). The screening of the library, and the conditions for pre-hybridization and hybridization were performed as described [9].

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequence determination of toxin Cn4*

Fig. 1 shows the complete amino acid sequence of toxin II-10, which we propose to name Cn4 (fourth toxin completely sequenced from *C. noxius*), following the nomenclature proposed in our recent publication [10]. Direct automatic Edman degradation confirmed the sequence of the N-terminal region of toxin Cn4, as previously published by our group [3,5], permitting unequivocally the identification of the first 45 residues. Tryptic digestion of RC-toxin produced at least 16 pep-

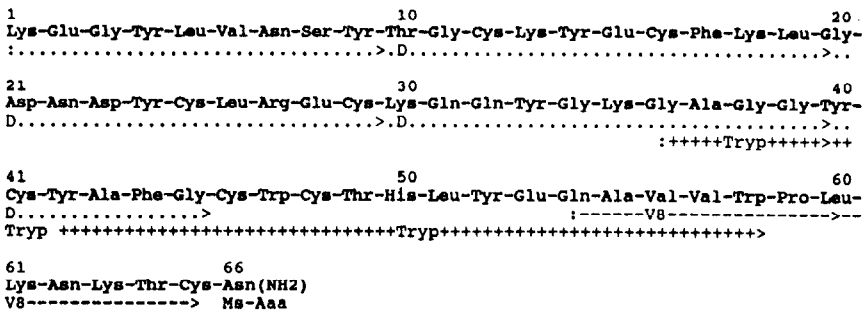

Fig. 1. Complete amino acid sequence of toxin Cn4. Results of direct Edman degradation (shown with D separated by points), allowed us to identify the first 45 residues. Tryptic digestion (labeled Tryp separated by crosses) indicates the amino acid sequence corresponding to the peptide eluted at 48.35 in the HPLC. This sequence starts with Gly at position 36 and goes to Trp at position 58. Protease V8 digestion (labeled with V8 separated by dashes) corresponds to the C-terminal region, with amino acid sequence corresponding to Gln at position 54 to the end of the sequence. The last residue (asparagine amide) was positioned based on Ms (mass spectrometry) and Aaa (amino acid analysis).

tides (data not shown), from which the peptide eluting at 48.35 min corresponded to the intermediate overlapping segment of the primary structure. HPLC separation of protease V8-cleaved RC-toxin (data not shown) produced at least 13 different peptides, from which the peptide eluting at 35.09 min was shown to correspond to the C-terminal region of the toxin (Fig. 1). The C-terminal V8 peptide was covalently attached to Sequelon AA (Millipore Corp.) following the manufacturer's instructions, and subjected to N-terminal sequence analysis on a Milligen 6600 ProSequencer using TFA 100 cycles. Amino acid residues 54–65 were unambiguously identified. In order to confirm that the entire peptide was sequenced, a second aliquot was subjected to amino acid analysis. The amino acid composition revealed the presence of a second aspartic acid residue (either Asp or Asn) not detected in the sequence. Our failure to identify the C-terminal residue would suggest that it was not attached to the aryl amide membrane. One would expect that aspartic acid and asparagine to be covalently attached to the membrane and therefore be identified, albeit in low yield. Thus, we surmised that the missing residue was asparagine amide. The peptide was covalently attached to the membrane through the carboxyl group of carboxymethylated-cysteine (CM-cys) of position 65, an attachment site previously identified for other CM-cys containing peptides (B.M. Martin, unpublished). Mass spectral analysis of the peptide with electrospray ionization indicated a molecular mass of 1,557 ± 1, consistent with an asparagine amide at the C-terminal residue. These conclusions are further supported by the results of cloning and nucleotide sequencing of the cDNA encoding for this toxin, as discussed below.

3.2. *Isolation and sequencing of the cDNA*

Using duplicate filters to screen the cDNA library with a degenerate oligonucleotide probe designed from a conserved region of the *Centruroides* toxins, we detected several positive signals with different intensities. One of these clones was isolated, subcloned in pBluescript (Stratagene, La Jolla, CA, USA), and both strands sequenced (Fig. 2), using the Sequenase version 2 kit (USB, USA). The resulting amino acid sequence deduced from this clone corresponded to toxin Cn4, the complete amino acid sequence of which is reported in this work (Fig. 1). The nucleotide sequence shown in Fig. 2 displayed an open reading frame coding for a precursor of 86 amino acid residues with a signal peptide of 19 amino acid residues, the mature peptide of 66

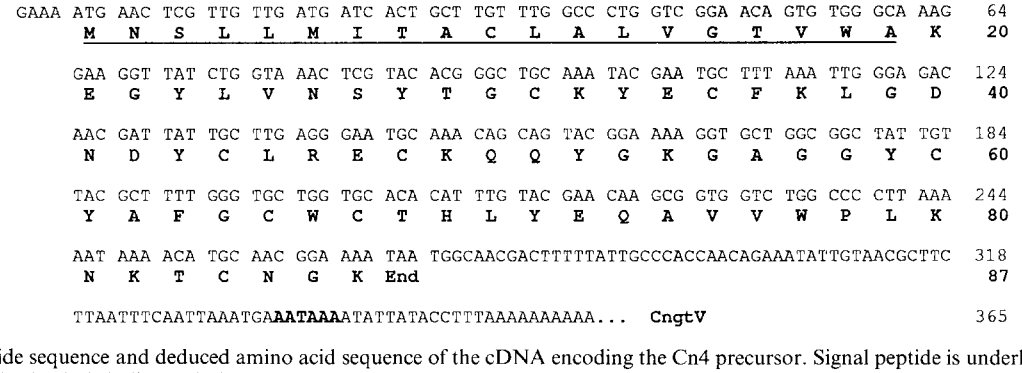

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the cDNA encoding the Cn4 precursor. Signal peptide is underlined. Polyadenylation signal (AATAAA) is indicated in bold characters. This nucleotide sequence will appear in the GenBank nucleotide sequence Database under accession number L05063. Nucleotides and amino acids are numbered at the right side of the corresponding sequences.

GROUP	TOXIN	AMINO ACID SEQUENCE	a	b
I	Cn4	KEGYLVNSYTGCKYECFKLGDNDYCLRECKQQYGKGAGGYCYAFGCWCTHLYEQAVVWPLKNKTCN	100%	(+)
I	Clt1	KEGYLVNHSTGCKYECFKLGDNDYCLRECRQQYGKGAGGYCYAFGCWCTHLYEQAVVWPLPNKTCS	92%	(10)
I	Cn3	KEGYLVELGTGCKYECFKLGDNDYCLRECKARYGKGAGGYCYAFGCWCTQLYEQAVVWPLKNKTCR	89%	(10)
I	CssII	KEGYLVSKSTGCKYECLKLGDNDYCLRECKQQYGKSSGGYCYAFACWCTHLYEQAVVWPLPNKTCN	88%	(14)
I	Cn2	KEGYLVDKNTGCKYECLKLGDNDYCLRECKQQGYKGAGGYCYAFACWCTHLYEQAIVWPLPNKRCS	83%	(10)
I	CssIII	KEGYLVSKSTGCKYECLKLGDNDYCLRECKQQYGKSSGGYCYAFACWCEALPDHTQVW-VPNK-CT	71%	(14)
II	CsEv2	KEGYLVNKSTGCKYGCLKLGENEGCDKECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNK-CSS	56%	(14)
II	CsEv1	KEGYLVKKSDGCKYDCFWLGKNEHCNTECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNK-CS	56%	(14)
II	CsEv3	KEGYLVKKSDGCKYGCLKLGENEGCDTECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNKSC	56%	(14)
II	Cn1	KDGYLVDA-KGCKKNCYKLGKNDYCNRECRMKHRGGSYGYCYGFGCYCEGLSDSTPTWPLTNKTC	56%	(10)
II	CsEI	KDGYLVEK-TGCKKTCYKLGENDFCNRECKWKHIGGSYGYCYGFGCYCEGLPDSTQTWPLPNK-CT	56%	(14)
		****** *** * ** * * ** **** * * * * ** *		

Fig. 3. Amino acid sequence comparison of several toxins from the genus *Centruroides* that affect Na^+ channels Gaps (-) were introduced for maximizing similarities. (a) Percent of similarity taking Cn4 as 100%. The gaps have been taken as differences for the calculus. (b) Reference: +this paper. Clt1, is toxin 1 from *C. limpidus tecomanus*. Cn1, Cn2, Cn3, Cn4 are toxins from *C. noxius* Hoffmann. CsEv1, CsEv2, CsEv3 and CsEI are toxins from *C. sculpturatus* Ewing. CssII and CssIII are toxins from *C. suffusus suffusus*. *Amino acid residues that are invariant in the same position in all the sequences.

residues and two additional residues at the C-terminus. These results support previous reports [11,12] which proposed that scorpion toxins are synthesized as precursors of the mature toxin. Two processing steps seem to occur for toxin Cn4: (i) elimination of the signal peptide, and (ii) processing of the C-terminal region. At the N-terminal region, the signal peptide is certainly excised by a signal peptidase, while at the C-terminal part of the peptide the additional two amino acid residues (Gly and Lys) are removed and the Gly residue donates its amino group for the α-amidation of asparagine, as has been described for the processing of the inactive precursors of secretory peptides that contain sites for proteolysis and α-amidation, including a toxic peptide from a cDNA library of *Tityus serrulatus* [13].

3.3. *Primary structure comparison*

The comparison of the amino acid sequence of toxins purified and sequenced from *Centruroides* scorpion venoms (Fig. 3) shows that they constitute a cluster of closely related peptides, sharing a high degree of similarity in their primary structures (at least 56%), as well as in their functional specificities through their recognition of Na^+ channels [4–6,14]. By comparing the different binding properties and electrophysiological data, the toxins affecting Na^+ channels were classified as: (i) α-toxins, that mainly act at the inactivation mechanism, and (ii) β-toxins that modify the activation mechanism of the Na^+ channel [14]. Toxin Cn4 has been shown by Carbone et al. [5] to affect the activation mechanism of Na^+ channels of squid axon, hence it belongs to the β-type toxins. It is very toxic to mice and rats, and binds to brain synaptosomes with an affinity of 2–10 nM [15]. Toxin Cn2 was poorly effective towards Na^+ channels of squid axon [5], but was shown to be very toxic to mice [10], binding to their brain synaptosomes with K_d's in the range of the low nanomoles [6,15]. Because it affects the activation mechanism, it is also a β-type toxin. The toxins from the venom of *C. sculpturatus* Ewing (CsEv1, 2 and 3) have the characteristic effects of α-type toxins and are very toxic to insects or crustaceans but have a small toxic effect in mammals [16].

Analysis of the sequences shown in Fig. 3 permits us to sort out several general features of the primary structure that can be correlated with the distinct pharmacological effects observed. Since the total number of amino acids and the positions of all cysteines are conserved it is very likely that the general folding of all toxic molecules are the same (the positions of 27 of the 66 amino acid residues are identical). The N-terminal region of all toxins is highly conserved. Larger differences are observed at the C-terminal part of the toxins, which provides support for a tentative classification into two distinct groups: group I, for which Cn4 is the prototype, is composed of peptides mainly toxic to mammals, displaying at least 70% similarity in the primary structure, with only two prolines in the last 15 amino acid residues of the sequence. Group II, consisting of peptides mainly toxic to insects and crustaceans, with about 56% primary structure similarity to Cn4, containing more than 2 proline residues at the C-terminal region, with CsEV2 as a representative toxin. Possibly, the structural constraints caused by the proline residues (α-imino acid) near the C-terminus could hamper an appropriate folding during interaction with the channel proteins, as suggested by others [16]. Final evaluation of this model, distinguishing the toxins based on their specificity for either mammalian or insect targets, depends upon development of more comprehensive structural data for the comparison of these motifs.

Acknowledgements: The Ph.D. scholarship obtained by A.V. from DGAPA-UNAM is greatly acknowledged. We express our gratitude to Dr. Mark Duncan of the Biomedical Mass Spectrometry Unit, University of New South Wales, Australia, for the mass spectral data, and Mr. Bill Eliason of NIMH for helpful discussions. This work was partially supported by grants from Howard Hughes Medical Institute (75191-527104), DGAPA-UNAM (IN202689 and IN300991) to L.D.P., and CONACyT-Mexico (0018-N9105) to L.D.P. and F.B.

REFERENCES

[1] Rochat, H. Bernard, P. and Couraud, F. (1979) In: Advances in Cytopharmacology, vol. 3 (Ceccarelli, B. and Clementi, F. eds.) pp. 325–334, Raven Press, NY.

[2] Possani, L.D. (1984) In: Handbook of Natural Toxins, vol. 3 (Tu, A.T. ed.) pp. 513–550, Marcel Dekker, NY.

[3] Possani, D.L., Dent, M.A.R., Martin, B.M., Maelicke, A. and Svendsen, I. (1981) Carlsberg Res. Commun. 46, 207–214.

[4] Carbone, E., Wanke, E., Prestipino, G. Possani, L.D. and Maelicke, A. (1981) Nature 296, 90–91.

[5] Carbone, E., Prestipino, G., Franciolini, F. Dent, M.A.R. and Possani, L.D. (1984) J. Physiol. (Paris) 79, 179–184.

[6] Sitges, M., Possani, L.D. and Bayon, A. (1987) J. Neurochem. 48, 1745–1752.

[7] Edman, P. and Begg, G. (1967) Eur J. Biochem. 1, 80–91.

[8] Possani, L.D., Martin, B.M., Svendsen, I., Rode, G.S. and Erickson, B.W. (1985) Biochem. J. 229, 739–750.

[9] Becceril, B., Vázquez, A., García, C. Corona, M., Bolivar, F. and Possani, L.D. (1993) Gene (in press).

[10] Zamudio, F.Z., Saavedra, R., Martin, B.M., Gurrola-Briones, G., Herion, P. and Possani, L.D. (1992) Eur. J. Biochem. 204, 281–192.

[11] Bougis, P.E., Rochat, H. and Smith, L.A. (1989) J. Biol. Chem. 264, 19259–19265.

[12] Gurevitz, M., Urbach, D., Zlotkin, E. and Zilberberg, N. (1990) Toxicon 29, 1270–1272.

[13] Bradbury, A., Finnie, M.D. A. and Smith, D.G. (1982) Nature 298, 686–688; see also Martin-Eauclaire, M.F., Ceard. B., Ribeiro, A.M., Diniz, C.R., Rochat, H. and Bougis, P.E. (1992) FEBS Lett. 302, 220–222.

[14] Meves, H., Simard, M.J. and Watt, D.D. (1984) J. Physiol. Paris 79, 185–191; same authors (1986) Ann. NY Acad. Sci. 479, 113–132.

[15] Ramírez, A.N., Gurrola, G.B., Valdivia, H.H. and Possani, L.D. (1991) Proceedings Xth World Congress on Animal, Plant and Microbial Toxins, Singapore. Abstract 301, p. 355.

[16] Loret, E.P., Sampieri, F., Roussel, A., Granier, C. and Rochat, H. (1990) Proteins 8, 164–172.