

The genomic region encoding toxin gamma from the scorpion *Tityus serrulatus* contains an intron

B. Becerril^a, M. Corona^a, M.C. Mejía^a, B.M. Martin^b, S. Lucas^c, F. Bolivar^a, L.D. Possani^{a,*}

^a*Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Apdo. Postal 510-3, Cuernavaca, 62271 México*

^b*National Institute of Mental Health, Molecular Neurogenetics Unit, Clinical Neuroscience Branch, Building 49, B1EE16, Bethesda, MD 20892, USA*

^c*Instituto Butantan, Av. Vital Brazil 1500, São Paulo, Brazil*

Received 30 September 1993

The gene encoding toxin gamma from the scorpion, *Tityus serrulatus*, was amplified by PCR from genomic DNA employing synthetic oligonucleotides designed from the reported cDNA sequence. The nucleotide sequence of this gene reveals the presence of an intron of 475 base pairs (bp) which interrupts the region that encodes the signal peptide of the precursor toxin. A comparison of the intron boundary sequences of the gamma toxin gene with ones from other arachnid genes is also presented.

Scorpion toxin; Na⁺ channel; Genomic DNA clone; Nucleotide sequence; Arachnid intron; *Tityus serrulatus*

1. INTRODUCTION

Scorpion toxins are extremely useful tools for the study of ion channels [1]. Toxin gamma, the first chemically characterized toxin from *Tityus serrulatus* venom [2], has played an important role in the characterization of voltage-sensitive Na⁺ channels (see review in [3]). The complete amino acid sequence of this peptide was obtained by two independent groups [4,5]. An extended report on the modification of the Na⁺ channel gating mechanism by this peptide has previously been reported [6]. Toxin gamma [2], also called toxin VII [5], belongs to the class of β -scorpion toxins, binds to the 4th site of the Na⁺ channel and has been the subject of intense research because it is toxic to both mammals and insects [7]. Recently, emphasis has been placed on cloning the gene that encodes this peptide, with the aim of expressing recombinant variants for the study of the structure–function relationship of the toxin–channel interactions [8]. A complementary DNA sequence encoding toxin VII was reported [9], however, no information on the nucleotide sequence at the genomic DNA level is known. In this communication, we report the complete nucleotide sequence of the gene that encodes gamma toxin of *T. serrulatus*. Thus, this toxin is the first for which the complete peptide sequence, nucleotide messenger RNA (cDNA) sequence and genomic DNA sequence have been determined.

2. MATERIALS AND METHODS

2.1. Purification of DNA from *T. serrulatus*

Genomic DNA was purified from whole body as described [10].

2.2. Synthesis of oligonucleotides

The oligonucleotides used for PCR (oligo 1) nt 1–20; oligo 2, nt 766–790 (complementary, Fig. 1) and the one used to screen for the gamma toxin gene (oligo 3, corresponding to amino acid residues 24–30, Fig. 1), were synthesized as described [10].

2.3. Polymerase chain reaction (PCR) and detection of gamma toxin-related nucleotide sequences

The PCR was performed in a programmable heating chamber (Hybaid) using 30 rounds of temperature cycling (92°C for 1 min, 55°C for 1 min and 72°C for 2 min) followed by a 10-min step at 72°C. A sample of the PCR reaction was electrophoresed in a 1% agarose gel, blotted onto a nitrocellulose membrane and probed with a specific ³²P-labeled oligonucleotide (oligo 3, Fig. 1).

2.4. Cloning and characterization of PCR products

PCR products were purified from gel, blunt-ended with T4 DNA polymerase and ligated into the *EcoRV* site of pBluescript phagemid (Stratagene, La Jolla, CA, USA). The ligation reaction was used to transform competent *E. coli* DH5- α cells. Plasmid DNA from white colonies was purified and digested with *Bam*HI and *Hind*III. Digestions were electrophoresed in an agarose gel, blotted onto nitrocellulose and probed with radioactive oligo 3 (see section 2.2. above). Positive clones were sequenced using the Sequenase kit, version 2 (US Biochemical, Cleveland, OH, USA) on both strands. Oligos 1, 2 (see section 2.2. above), M13 –20, M13 reverse and 4 (complementary to nt 531–550, Fig. 1), were used for sequencing.

3. RESULTS AND DISCUSSION

3.1. Southern blot of PCR products

Electrophoresis and staining with ethidium bromide of PCR products revealed the presence of several faint

*Corresponding author. Fax: (52) (73) 172 388.

<u>TCGATCTGAACG</u> <u>ATG AAA</u> GGA ATG ATC TTG TTT ATT AGC TGC TTA TTG CTG ATC G	55
M K G M I L F I S C L L L I	-14
gtaagctgaattcagtttctaagtatatattttattatataatataatgtatattcgctttaacaat	123
acacccatcataatgagctgttgtaataatctatcatcatgctatcgatcacacaatcattgaaacg	191
cgtagtattacagtagtattgcccagtttttatgaaaatatttaggattggtaaagcagaattactttt	259
catcgtttcatcaggaagcagattattcccatctgtttgctactaaatgtggtaaagaaaatttatta	327
gtgtgtgttataatgcagattagcaattaaagaatataacaaacttatgtgaaatatataaagaaac	395
gttgacccaatttttaaatttgaaatgtgattctcgtaaaattcatgtgcgagtgccaacgttgatgtgc	463
tatagaaattttcatatttatttagattactcaaaatgctaattggacttttttattttgttaacatag	530
.....	
GC ATT GTC GTA GAA TGT AAA GAA GGT TAT CTC ATG GAT CAC GAA GGT TGC AAA	583
G I V V E C K E G Y L M D H E G C K	12
CTT AGT TGC TTT ATC AGA CCA TCG GGA TAC TGC GGC AGA GAA TGC GGA ATT AAA	637
L S C F I R P S G Y C G G R E C G G I K	30
AAG GGC TCA TCG GGC TAT TGC GCC TGG CCC GCG TGT TAC TGC TAC GGG CTT CCA	691
K G S S G Y C A W P A C Y C Y G L P	48
AAT TGG GTG AAA GTT TGG GAT AGA GCG ACG AAC AAA TGT GGC AAA AAA TAA ATC	745
N W V K V W D R A T N K C G K K end	64
<u>TGTTTCGCTGAAAACCCCTTTACAAATGAACTGTAATAAGTTTGGC</u>	790

Fig. 1. Nucleotide sequence of the genomic DNA encoding the gamma toxin precursor from *Tityus serrulatus*. The deduced amino acid sequence is given below the nucleotide sequence. Nucleotides and amino acids (taking the first residue of the mature toxin as +1) are numbered at the right side of corresponding sequences. The signal peptide is underlined. Oligonucleotides 1 and 2 used for PCR (nt 1–20 and 766–790 (complementary)) are double underlined. The broken line above nt 617–637 and dotted line above nt 531–550 represent oligos 3 and 4, respectively (see section 2.4.).

Exons 1 (nt 1–55) and 2 (nt 531–790) are written in capital letters. The intron (nt 56–530) is written in lower case letters.

bands of diverse size and a stronger one of approximately 800 bp. Southern blot of these bands, using radioactive oligo 3 (see section 2.2.) as a probe, showed that only the 800 bp band was specific (data not shown). According to the stretch flanked by oligos 1 and 2 (315 bp) on the cDNA [9], these results suggested the presence of an intron of approximately 485 bp. As shown in Fig. 1, it is actually 475 bp.

3.2. Nucleotide sequence determination

From the final screening (see section 2.4.), 3 clones were selected on the basis of the expected size and the strength of the autoradiographic signal. They were partially sequenced and finally, one of them which encoded gamma toxin, was completely sequenced on both strands using the oligonucleotides described in section 2.4. as sequencing primers. This nucleotide sequence is shown in Fig. 1. Analysis of this sequence identifies two exons (nt 1–55 and nt 531–790) and an intron (nt 56–530) with a size of 475 bp. To our knowledge, this is the first formal report of a scorpion gene in which the complete genomic nucleotide sequence is presented. Preliminary results obtained with a gene from *Androctonus australis* Hector [8], suggested the presence of an intron of similar size and position, but provided no nucleotide sequence [8]. The nucleotide sequence of the two exons from the gene encoding gamma toxin of *T. serrulatus* presented here, coincides exactly with the reported cDNA sequence for toxin VII [9], except for positions 745 and 760 in which a thymine (cDNA) is replaced by a cytosine (genomic DNA). These results are consistent

Table I

Comparison of intron boundaries from different genes of arachnids*

Intron	Sequence of intron boundaries		
	5' Splice donor	Intron size	3' Splice acceptor Ref.
1	gtaagctctca	6.1 kbp	tttttttttag [12]
2	gtaagtacat	8.6 kbp	tttttttatag [12]
3	gtaagtccaa	6.3 kbp	gtaactctag [12]
4	gtaagtatgt	2.1 kbp	tatttttttag [12]
5	gtaagtagtc	5.0 kbp	tcattccgcag [12]
6	gtattttctgc	6.9 kbp	catttttacag [12]
7	gtatgttaag	14.3 kbp	tttatttttag [12]
8	gtaagtgcga	2.7 kbp	cttcttttcag [12]
9	gtaagctgaa	475 bp	gttaacatag **
10	gttaggtccc	58 bp	taataatcag [13]
11	gtaagtttga	60 bp	aaattgctag [13]
12	gtatgttttg	60 bp	ttatttttttag [13]
13	gtaagtttga	65 bp	aaaatgttag [13]
14	gtacaaaatg	60 bp	catttttttag [13]
15	gtaagtttga	65 bp	aaatttatag [13]
16	gtaagtttgc	61 bp	atttttttttag [13]
Consensus	<u>gtaagtnnnn</u>		<u>nnnttttttag</u>

* Comparison of 5' and 3' intron boundaries from different genes of arachnids. Only the first 10 and last 10 nucleotides of the introns are shown. The proposed consensus for 5' and 3' intron boundaries are shown at the bottom of this table (n means any nucleotide). Introns 1–8 correspond to the gene encoding subunit e of hemocyanin from the tarantula *Eurypelma californicum*. Intron 9 corresponds to the gamma toxin encoding gene from the South American scorpion, *Tityus serrulatus*. Introns 10–16 correspond to the genes encoding cysteine proteases from the mites, *Dermatophagoides farinae* (introns 10–12), *Euroglyphus maynei* (introns 13–14) and *Dermatophagoides pteronyssinus* (introns 15–16).

**This work.

with the conclusion that the gene encoding toxin gamma [2], also called toxin VII [5], from *Tityus serrulatus*, has been cloned and sequenced. Also, it appears that native toxin gamma is in fact processed at the C-terminal region, after synthesis, by a similar mechanism to that described for toxin Cn4 of the scorpion, *Centruroides noxius* [11]. As shown in Fig. 1, prior to the termination codon, in the translated peptide there is a glycine followed by two basic residues (lysines). Thus, the cysteine at the end of the mature peptide may be amidated, after C-terminal processing as described [5] and suggested by Martin-Eauclaire et al. [9].

3.3. Intron sequence analysis

A search of the GenBank (release August, 1993) using the program Gene-Works, revealed the existence of four genes containing introns from other members of the group of arachnids. One from the tarantula (*Eurypelma californica*) encoding subunit e of hemocyanin [12] which contains eight introns and other three encoding cysteine proteases from the mites, *Euroglyphus maynei*, with two introns, *Dermatophagoides farinae* with three introns, and *Dermatophagoides pteronyssinus* with two introns (unpublished results of Kent et al. [13]). Comparison of the sequences from these fifteen introns with the intron of gamma toxin gene (Table I), reveals a diversity of sizes from 58 bp to 14.3 kbp. The only common sequences are those at the 5' and 3' boundaries. A consensus sequence is proposed for these boundaries in Table I. When these boundary consensus sequences are compared to other eukaryotic intron boundaries [14], it appears that they are a regular feature of all eukaryotic introns thus far reported.

Acknowledgements: This work was partially supported by grants from Howard Hughes Medical Institute (75191-527104), DGAPA-UNAM (IN 205893) to L.D.P., and CONACyT-Mexico (0018-N9105) to L.D.P. and F.P.

REFERENCES

- [1] Catterall, W. (1976) *J. Biol. Chem.*, 251, 5528-5536.
- [2] Possani, L.D., Alagon, A.C., Fletcher Jr., P.L. and Erickson, B.M. (1977) *Arch. Biochem. Biophys.* 180, 394-403. See also Possani, L.D., Martin, B.M., Mochca-Morales, J. and Svendsen, I. (1981) *Carlsberg Res. Commun.* 46, 195-205.
- [3] Lazdunski, M., Frelin, C., Barhanin, J., Lombet, A., Meiri, H., Pauron, D., Romey, G., Schmid, A., Schweitz, H., Vigne, P. and Vijverberg, H.P.M. (1986) *Ann. NY Acad. Sci.* 496, 204-220.
- [4] Possani, L.D., Martin, B.M., Svendsen, I., Rode, G.S. and Erickson, B.W. (1985) *Biochem. J.* 229, 739-750. See also Possani, L.D., Martin, B.M., Fletcher, M.D. and Fletcher Jr., P.L. (1991) *J. Biol. Chem.* 266, 3178-3185.
- [5] Bechis, G., Sampieri, F., Yuan, P.M., Brando, T., Martin, M.F., Diniz, C.R. and Rochat, H. (1984) *Biochem. Biophys. Res. Commun.* 122, 1146-1153.
- [6] Kirsch, G.E., Skattebøl, A., Possani, L.D. and Brown, A.M. (1989) *J. Gen. Physiol.* 93, 67-83.
- [7] Loret, E.P., Sampieri, F., Roussel, A., Granier, C. and Rochat, H. (1990) *Proteins* 8, 164-172.
- [8] Martin-Eauclaire, M.F., Delabre, M.L., Ceard, B., Ribeiro, A., Sogaard, M., Svensson, B., Diniz, C.R., Smith, L.A., Rochat, H. and Bougis, P.E. in: *Recent Advances in Toxinology Research*, vol. 1 (Gopalakrishnakone, P. and Tan, C.K. eds.) pp. 196-209, Singapore University Press, Singapore.
- [9] Martin-Eauclaire, M.F., Ceard, B., Ribeiro, A.M., Diniz, C.R., Rochat, H. and Bougis, P.E. (1992) *FEBS Lett.* 302, 220-222.
- [10] Becerril, B., Vázquez, A., García, C., Corona, M., Bolivar, F. and Possani, L.D. (1993) *Gene* 128, 165-171.
- [11] Vázquez, A., Becerril, B., Martin, B.M., Zamudio, F., Bolivar, F. and Possani, L.D. (1993) *FEBS Lett.* 320, 43-46.
- [12] Voll, W. and Voit, R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 5312-5316.
- [13] Kent, N., Hill, M.R., Holland, P.W. and Hart, B.J. Dept. Biochem. Univ. Oxford, South Parks Road, Oxford OX1 3QU, UK. (Unpublished).
- [14] Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiber, S. and Sharp, P.A. (1986) *Annu. Rev. Biochem.*, 55, 1119-1150.