FEBS 25458 FEBS Letters 508 (2001) 241–244

A naturally occurring non-coding fusion transcript derived from scorpion venom gland: implication for the regulation of scorpion toxin gene expression

Shunyi Zhu*, Wenxin Li, Zhijian Cao

Department of Biotechnology, College of Life Sciences, Wuhan University 430072, Wuhan, Hubei Province, PR China

Received 18 September 2001; accepted 10 October 2001

First published online 30 October 2001

Edited by Lev Kisselev

Abstract Scorpion venom glands synthesize and secrete a great number of low molecular mass toxic peptides for prey and defense. Many cDNAs and genomic genes encoding these toxins have been isolated and sequenced. However, their expression regulation mechanism is not yet known at present. During screening of a cDNA library prepared from venom glands of the scorpion *Buthus martensii* Karsch, we isolated a natural fusion cDNA composed of the 5′-untranslated region (UTR) and upstream coding sequence of a long-chain toxin transcript and the downstream coding sequence and 3′-UTR of a short-chain toxin transcript. The junction site is just the overlapping region of 11 nucleotides (GGCAAGGAAAT) between the two wild transcripts, and thus leads to the formation of an early stop codon, which will cause premature translation. Based on the above observations, combined with the genomic data, we proposed a characteristic regulation mechanism of scorpion toxin genes, in which *trans*-splicing and nonsense mediated mRNA decay are involved. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Scorpion toxin; Fusion transcript; Gene expression regulation; *Buthus martensii* Karsch

1. Introduction

Scorpions belong to one of the most ancient arthropods, appeared on earth about 400 million years ago. During the course of evolution, they have developed the venoms as effective weapons for prey and defense [1]. The venoms secreted by highly differentiated venom gland cells [2] contain rich bioactive components, of which the small toxic peptides (toxins) with molecular mass of 3–9 kDa are of the most important value, which have been used as experimental tools for the studies of the structure–function relationship of ion channels [3] and as natural scaffolds for protein engineering [4]. To date, primary structures of more than 200 toxins from different scorpion species have been determined. Based on sequence similarity, Possani et al. divide them into at least 26 subfamilies [5]. Pharmacologically, this class of molecules bind with ion (Na^+, K^+, Cl^- and Ca^{2+}) channels with high affinity and specificity and thus block and modify their gating mechanism [5]. Three-dimensional structure studies show all the scorpion toxins acting on Na^+, K^+ and Cl^- channels adopt a common structural motif, namely the cysteine-stabilized α-helix (CSH) motif, which involves a $CX_{(3)}C$ stretch of the α-helix bonded through two disulfides to a $CX_{(1)}C$ triplet of the β-strand, while the toxins active against ryanodine-sensitive Ca^{2+} channel, like Maurocalcine, fold following the inhibitor cystine knot fold (ICK), which consists of a cystine knot and a triple-stranded anti-parallel β-sheet [5,6].

Although some structural and functional data about scorpion toxins have been established, little is known about the expression regulation mechanism [7]. Our group has previously cloned and sequenced many toxin-coding cDNAs and genes from *Buthus martensii* Karsch, among which a cDNA for a long-chain K^+ channel toxin of 60 amino acids (BmTXKβ) [8] and a cDNA for a short-chain Cl^- channel toxin of 34 amino acids (BmKCT) [9] have been characterized and reported. Here we describe a fusion cDNA (BmTXKβ-BmKCT) sequence composed of the 5′-sequence of BmTXKβ cDNA and the 3′-sequence of BmKCT cDNA. We found that the junction site is just the overlapping region of 11 nucleotides (GGCAAGGAAAT) between two wild transcripts. In addition, we isolated the genomic regions encoding BmTXKβ and BmKCT transcripts. On the basis of these results, we analyzed the possible origin of BmTXKβ-BmKCT and its putative regulation function in gene expression of scorpion toxins.

2. Materials and methods

2.1. Construction and screening of the cDNA library of venom gland

The cDNA library prepared from the venom gland of the *B. martensii* Karsch have previously been reported [8]. The fusion clone was obtained and characterized by random sequencing of the clones containing small inserts from the cDNA library and comparing them with the sequences in GenBank database (http://www.ncbi.nlm.nih.gov/).

2.2. Isolation of genomic region encoding BmTXKβ, BmKCT and BmTXKβ-BmKCT

The genomic DNA for PCR template was prepared according to the method described previously [10]. PCR primers for amplification of the genomic regions encoding BmTXKβ, BmKCT and BmTXKβ-BmKCT transcripts were shown in Table 1.

PCR conditions: 95°C for 5 min, followed by 30 cycles of 95°C for 1 min, 55°C for 1–4 min and 72°C for 2 min, and at last 72°C for 10 min. The cloning of PCR product was done according to the method described previously [10z].

*Corresponding author. Present address: Laboratory of Toxicology, University of Leuven, E. Van Evenstraat 4, 3000 Leuven, Belgium.
E-mail address: shunyizhu@hotmail.com (S. Zhu).

Abbreviations: UTR, untranslated region; ORF, open reading frame; NMD, nonsense mediated mRNA decay

0014-5793/01/$20.00 © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
PII: S0014-5793(01)03067-8

Table 1
Oligonucleotide primers for the isolation of genomic regions[a]

	Forward primer	Reverse primer
BmTXKβ	(FP1)AAAAATATAAAGGAAAAGCTG	(RP1)AACTATTTGCGCAACTTCAG
BmKCT	(FP2)ATGAAGTTCCTCTACGGAATCG	(RP2)TCATATACGGTTACACAGACAT
BmTXKβ-BmKCT	(FP3)ATGATGAAACAACAGTTCTTC	(RP2)TCATATACGGTTACACAGACAT

[a]Reverse primers are the reverse complementary sequences.

2.3. Data analysis

Sequence comparison was completed using CLUSTAL X (1.8) software (Thompson et al., 1997), and further refined manually. RNA secondary structure prediction was performed using RNAdraw1.12b software (http://rnadraw.base8.se/) with default parameters. The nucleotide sequences of BmTXKβ-BmKCT (initially named BmTXKβ-SP in GenBank) transcript and the genomic region encoding BmTXKβ and BmKCT have been deposited in GenBank under accession numbers AF155365, AF419252 and AF380939, respectively.

3. Results and discussion

3.1. Analysis of nucleotide sequence and origin of BmTXKβ-BmKCT

By using random sequencing strategy, we isolated and characterized one fusion cDNA from the venom gland cDNA library. It is composed of 243 nucleotides with a poly(A) tail of 141 A. The first half sequence (1–118 nucleotides) corresponds to the 5′-untranslated region (UTR) and signal peptide- and pro-peptide-coding sequence of BmTXKβ cDNA, only a substitution (G→A) was found in position 101, whereas the second half exactly match the 3′-coding sequence and UTR of BmKCT cDNA (Fig. 1). The junction site is just the overlapping region of 11 identical nucleotides (GGCAAGGAAAT) of BmTXKβ and BmKCT transcripts, and thus leads to the formation of an early stop codon and subsequent four short open reading frames (ORFs) with no one greater than 28 amino acids.

To analyze the possible origin of this fusion transcript, we

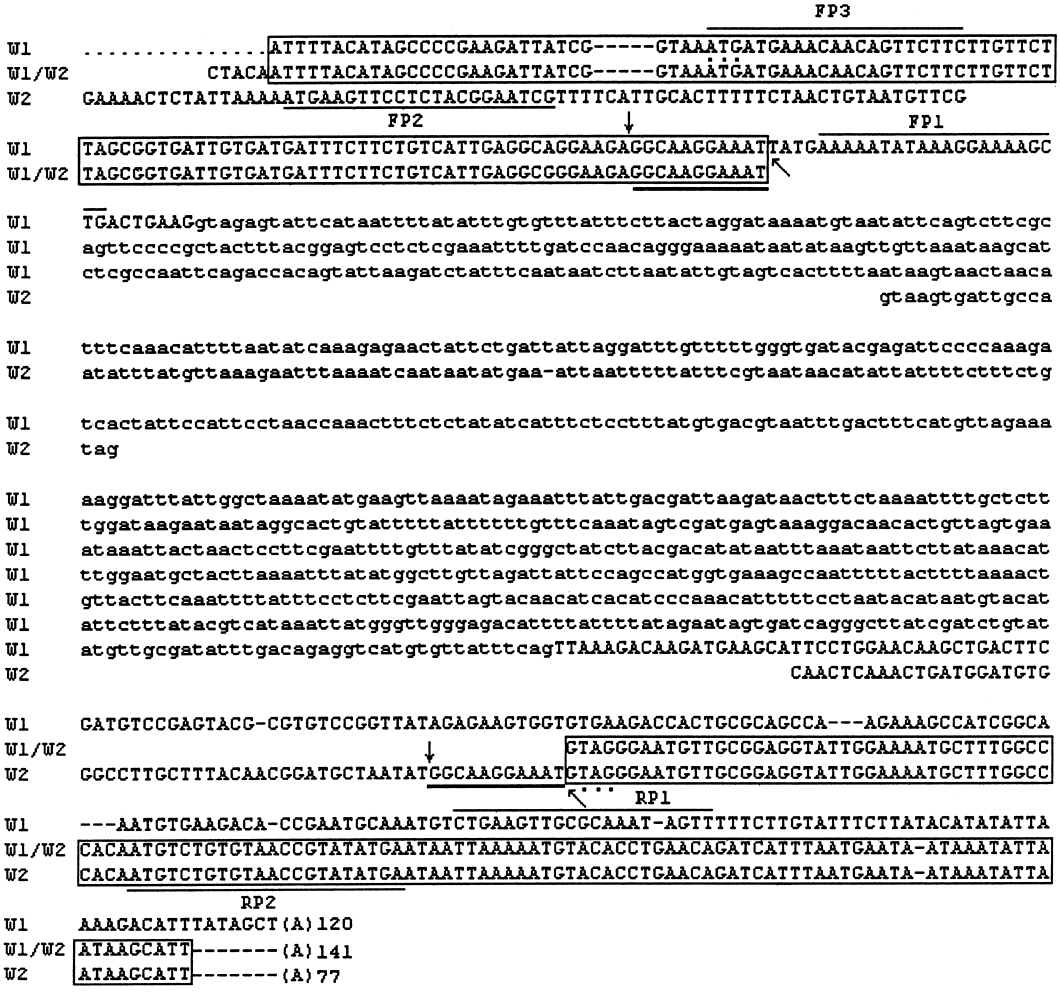

Fig. 1. Nucleotide sequences of BmTXKβ, BmKCT and BmTXKβ-BmKCT cDNAs and the genomic DNAs encoding the two wild transcripts. Exonic sequences are written in capitals, while intronic sequences are in lowercase. The identical nucleotides between BmTXKβ-BmKCT and its wild transcripts are boxed. The forward and reverse primers for the amplification of genomic DNA are underlined once. The overlapping sequence of 11 nucleotides is underlined thickly once. The arrows indicate two possible splicing sites. W1: BmTXKβ; W2: BmKCT; W1/W2: BmTXKβ-BmKCT.

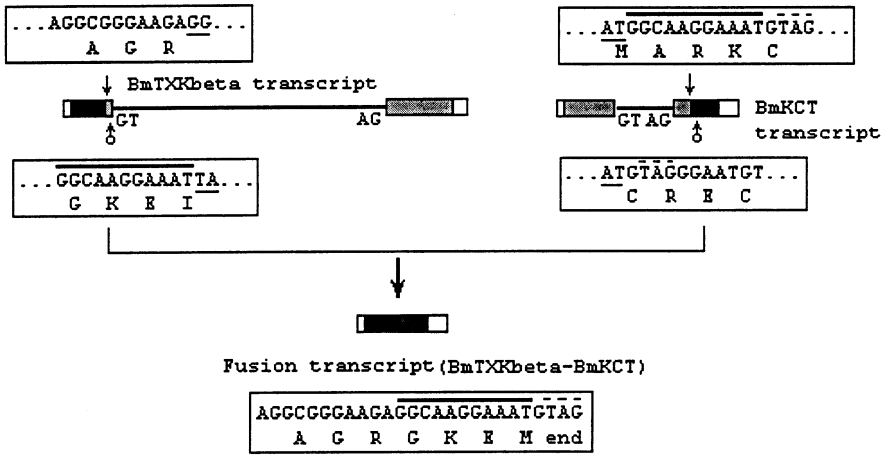

Fig. 2. Schematic representation of the putative *trans*-splicing of BmTXKβ and BmKCT genes. Two possible splicing sites are indicated by the symbols ↓ and ♂, respectively. The junction sites and early stop codon are shown. The boxes represent exons. Black or gray, coding region; white, non-coding region. The thick lines represent introns.

decided to amplify the genomic regions encoding BmTXKβ, BmKCT and BmTXKβ-BmKCT. The results showed that BmTXKβ and BmKCT shared similar genomic organization to that of other scorpion toxins, which contain one intron of 887 and 94 nucleotides respectively located in the coding regions with high A+T content, display identical boundaries, and agree with the GT/AG splice junctions (Fig. 1). But no PCR product was formed when the primers FP3 and RP2 (Table 1) were used to amplify the corresponding genomic DNA of the fusion transcript, which seemingly showed that the genes of BmTXKβ and BmKCT are not located in the same chromosome. The above results allow us to assume that BmTXKβ and BmKCT genes could independently be transcribed and then either spliced into two wild mature transcripts with coding capability by *cis*-splicing mechanism (a common mechanism for all the scorpion toxin pre-mRNA splicing characterized so far [11,12]) or fused into chimeric transcript with no coding capability by *trans*-splicing (Fig. 2). Due to the existence of the overlapping region in the junction site between BmTXKβ and BmKCT transcripts, there are two potential splicing donor and acceptor sites (GG/AT or TA/AT) for choice, which both do not match the consensus splicing site (GT/AG) (Fig. 2). The similar case was also found in the intergenic splicing of human *P2Y*$_{11}$ and *SSF1* genes, in which the splicing donor site is GC instead of GT [13].

3.2. Functional analysis of the overlapping nucleotides GGCAAGGAAAT

That the overlapping 11 nucleotides just occurred in the junction site (Fig. 1) suggests it is not an accidental phenomenon. Therefore, analyzing its potential role is of important value. Because almost all RNA processing events depend on RNA–RNA interactions [14,15] and it is also common for the snRNAs to pair with a short sequence in the pre-mRNA to participate splicing [14], we examined five snRNAs (U1, U2, U4, U5 and U6) to find sequences which could hybridize to the 11 nucleotides sequence. Only U2 RNA contained a complementary sequence to GGCAAGGAAAT (Fig. 3). Furthermore, the secondary structure of the region containing the complementary sequence is changeable in U2 [16]. Thus, we can hypothesize that U2 snRNA could involve the *trans*-splicing of BmTXKβ and BmKCT transcripts by base pairing, which leads to endonucleolytic cleavage between dsRNA and ssRNA as described in the 3′-end generation of histone H3 mRNA [14], and finally forming fusion molecule (Fig. 3). The single-strand state of GGCAAGGAAAT in the predicted secondary structures of the BmTXKβ and BmKCT tran-

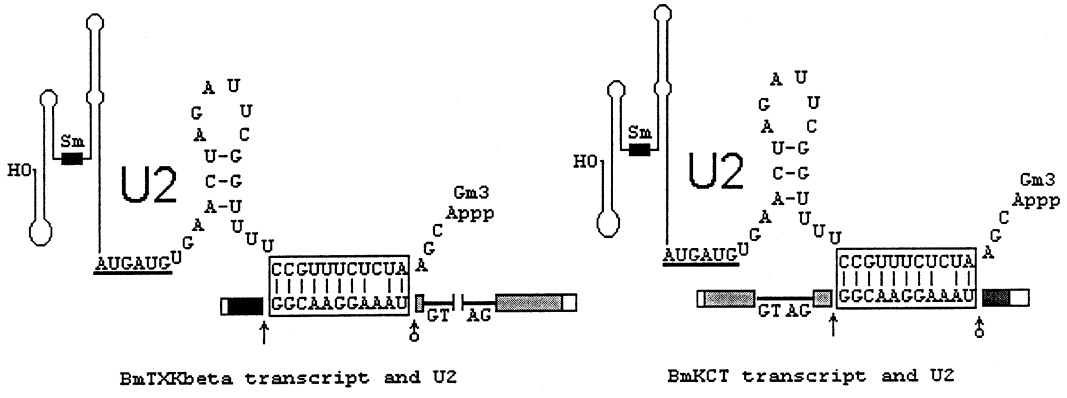

Fig. 3. Possible mode of hybridization between U2 snRNA and the overlapping site (GGCAAGGAAAT). The U2 data were taken from Phillip A. Sharp [16].

scripts facilitate its pairing with U2 (result not shown). The presence of similar sequences to GGCAAGGAAAT, which can also pair with U2, in other scorpion toxins affecting Na^+ and K^+ channels (AF276225, AF155364, S82286, AF156171, etc.), further supports our assumption. Recently, an exonic splicing enhancer (GAAGAAG) was found responsible for mammalian natural *trans*-splicing [17]. Whether GGCAAGGAAAT is of similar function remains unknown at present.

3.3. Implication for the regulation of gene expression of scorpion toxins

Previous studies have shown that some species such as yeast and mammal have evolved a nonsense mediated mRNA decay (NMD) mechanism, by which the aberrant mRNAs are rapidly degraded [18,19]. Moreover, several evidences have suggested that a class of seemingly normal mRNAs with alternate stop codons produced by alternative splicing patterns can be the substrates for NMD mechanism and thus downregulate the levels of the normal mRNAs [18,19]. Very likely, BmTXKβ-BmKCT is a substrate of NMD due to the existence of an early stop codon closer to the initiator codon of BmTXKβ transcript, and multiple short ORFs [18] (Fig. 1). Thus, a logical inference may be proposed to explain the biological purpose of the fusion transcript present in scorpion venom gland. In our opinion, the non-coding fusion transcript produced by the precise fusion of two different wild coding transcripts might be rapidly degraded after formation by NMD pathway and thus down-regulates the expression level of the two toxin genes. Certainly, this inference needs more experimental data for support.

Acknowledgements: This work was supported by grant number 39970897 from the National Natural Science Foundation of China.

References

[1] Froy, O., Sagiv, T., Poreh, M., Urbach, D., Zilberberg, N. and Gurevitz, M. (1999) J. Mol. Evol. 48, 187–196.
[2] Kanwar, U., Sharma, A. and Nagpal, N. (1981) J. Anim. Morphol. Physiol. 28, 206–209.
[3] Garcia, M.L., Gao, Y.D., McManus, O.B. and Kaczorowski, G.J. (2001) Toxicon 39, 739–748.
[4] Vita, C., Drakopoulou, E., Vizzavona, J., Rochette, S., Martin, L., Menez, A., Roumestand, C., Yang, Y., Ylisastigui, L., Benjouad, A. and Gluckman, J.C. (1999) Proc. Natl. Acad. Sci. USA 96, 13091–13096.
[5] Possani, L.D., Merino, E., Corona, M., Bolivar, F. and Becerril, B. (2000) Biochimie 82, 861–868.
[6] Mosbah, A., Kharrat, R., Fajloun, Z., Renisio, J.G., Blanc, E., Sabatier, J.M., Ayeb, M.E. and Darbon, H. (2000) Proteins 40, 436–442.
[7] Gurevitz, M., Froy, O., Zilberberg, N., Turkov, M., Strugatsky, D., Gershburg, E., Lee, D., Adams, M.E., Tugarinov, V., Anglister, J., Shaanan, B., Loret, E., Stankiewicz, M., Pelhate, M., Gordon, D. and Chejanovsky, N. (1998) Toxicon 36, 1671–1682.
[8] Zhu, S.Y., Li, W.X., Zeng, X.C., Jiang, D.H., Mao, X. and Liu, H. (1999) FEBS Lett. 457, 509–514.
[9] Zeng, X.C., Li, W.X., Zhu, S.Y., Peng, F., Zhu, Z.H., Wu, K.L. and Yang, F.H. (2000) Toxicon 38, 1009–1014.
[10] Zhu, S.Y., Li, W.X. and Zeng, X.C. (2001) Toxicon 39, 1291–1296.
[11] Delabre, M.L., Pasero, P., Marilly, M. and Bougis, P.E. (1995) Biochemistry 34, 6729–6736.
[12] Legros, C., Bougis, P.E. and Martin-Eauclaire, M.F. (1999) Perspect. Drug Discov. Des. 15, 1–14.
[13] Communi, D., Suarez-Huerta, N., Dussossoy, D., Savi, P. and Boeynaems, J.-M. (2001) J. Biol. Chem. 276, 16561–16566.
[14] Lewin, B. (1997) Genes VI, Oxford University Press, Oxford.
[15] Berget, S.M. (1984) Nature 309, 179–182.
[16] Sharp, P.A. (1994) Cell 77, 805–815.
[17] Caudevilla, C., Codony1, C., Serra, D., Plasencia, G., Román, R., Graessmann, A., Asins, G., Bach-Elias, M. and Hegardt, F.G. (2001) Nucleic Acids Res. 29, 3108–3115.
[18] Ruiz-Echevarria, M.J., Czaplinski, K. and Peltz, S.W. (1996) Trends Biochem. Sci. 21, 433–438.
[19] Hilleren, P. and Parker, R. (1999) Annu. Rev. Genet. 33, 229–260.