FEBS 18098 FEBS Letters 402 (1997) 45–49

Genomic organization of the KTX_2 gene, encoding a 'short' scorpion toxin active on K^+ channels

Christian Legros, Pierre E. Bougis, Marie-France Martin-Eauclaire*

Laboratoire de Biochimie, UHR 6560 du CNRS, Ingénierie des Protéines, Institut Fédératif INSERM Jean Roche, Faculté de Médecine Nord, Université de la Méditerranée, Bd Pierre Dramard, F-13916 Marseille, Cedex 20, France

Received 24 November 1996

Abstract **A single intron of 87 bp, close to the region encoding the C-terminal part of the signal peptide, was found in the gene of the 'short' scorpion toxin kaliotoxin 2 of *Androctonus australis* acting on various types of K^+ channels. Its A+T content was particularly high (up to 86%). By walking and ligation-mediated PCR, the promoter sequences of the kaliotoxin 2 gene of *Androctonus australis* were studied. The transcription unit of the gene is 390 bp long. Consensus sequences were identified. The genes of 'short' scorpion toxins active on K^+ channels are organized similarly to those of the 'long' scorpion toxins active on Na^+ channels and not like those of structurally related insect defensins, which are intronless.**

Key words: Scorpion toxin; Potassium channel; Gene; Promoter

1. Introduction

Scorpion venoms contain potent neurotoxins which specifically modify ionic channel functions [1]. They are therefore useful tools for investigating the mechanisms of ion conduction and channel selectivity. Toxins active on K^+ channels are minor components of venoms. They are single chain polypeptides of 30–40 amino acid residues, with three disulfide bridges [2]. Available three-dimensional structures show that 'long' toxins (60–70 amino acid residues, with four disulfide bridges) active on Na^+ channels and 'short' toxins share a common structural motif also found in the insect defensins: an α-helix linked to two anti-parallel β-sheets by two disulfide bridges. These bridges appear to conserve the relative position of the β-sheets and the α-helix [3]. A 370 bp cDNA encoding the kaliotoxin 2 (KTX_2) precursor (a 37 amino acid residues peptide) purified from the North African scorpion *Androctonus australis* and acting as a high affinity blocker of K^+ channels, was previously obtained by PCR amplification [4]. The present work is the second step towards analyzing the genomic organization of the scorpion toxin genes. The gene corresponding to this cDNA was amplified from the genomic DNA of *Androctonus australis* and its sequence determined. By walking and ligation-mediated PCR, the promoter sequences of the KTX_2 gene were studied.

2. Materials and methods

2.1. Preparation of the scorpion genomic DNA

High molecular mass scorpion genomic DNA was purified from muscle tissue of scorpions collected in various areas of Tunisia as previously described [5]. Three animals from Beni-Khedache and one from Tozeur were used separately.

2.2. Amplification of the genomic parts encoding precursors

The oligonucleotides used for PCR were the following: forward primer K1, 5′-ACTGGAGCTCTCATTCGAAAATGAAGGTG-3′; reverse primer K2, 5′-AAACTGCAGTCACTTTGGTGTACAATCGCA-3′. The forward and reverse primers were preceded by a *Sac*I and a *Pst*I site, respectively, to allow ligation into pBluescript. PCR was used to amplify the genomic sequence encoding KTX_2 (primers K1 and K2) precursors. PCRs were performed in a final volume of 100 μl containing 200 ng scorpion genomic DNA (Beni-Khedache) as template, 100 nM of each primer, 2.5 units of thermostable DNA polymerase Biotaq (Eurobio) and 1× supplied buffer according to the manufacturer's instructions. 10 μl of each sample was then analyzed by 2% agarose gel electrophoresis in TBE buffer and the PCR products were visualized under UV light by ethidium bromide staining. The band with the appropriate size was recovered, purified using the Wizard PCR Prep DNA Purification System (Promega) according to the manufacturer's instructions and reamplified by a second PCR in the same conditions.

2.3. Sequence of PCR products

The purified PCR product corresponding to the coding region of the KTX_2 gene was ligated into the vector pBlueScript SK^- (Stratagene). *E. coli* XL-1 Blue was used for plasmid propagation. The recombinant clones were analyzed by standard techniques [6] and sequenced [7].

2.4. Walking along the 5′ region of the KTX_2 gene

Ligation-mediated PCR (LMPCR) was used to obtain sequences adjacent to the genomic DNA sequenced encoding KTX_2. *Androctonus australis* (Tozeur) genomic DNA (500 ng) was digested with *Mbo*I and extracted using Geneclean kit (Bio 101). Primer extension was carried out with 0.5 pmol of 5′-biotinylated reverse primer (5′-biotin-TCACTTTGGTGTACAATCGCATTTGC-3′, corresponding to the C-terminus of KTX_2), 1 unit of Extra-Pol II DNA polymerase, 250 μM each dNTP, and 1× supplied buffer in a final volume of 40 μl. First strand synthesis involved a thermal cycle of 5 min at 95°C, 30 min at 62°C, and 15 min at 72°C. The primer extended DNA was immobilized on streptavidin-coated magnetic beads (Dynal) according to the manufacturer's instructions. Immobilized DNA was ligated to 80 pmol of common linker [8] in 60 μl with 6 Weiss units of T4 DNA ligase in 1× supplied buffer (New England Biolabs) at 16°C for 16 h. The supernatant was removed by decanting on a magnetic separation stand (Dynal) to retain the magnetic beads. Washing and elution of the non-biotinylated strand were performed as described previously [9]. After ethanol precipitation, PCR was carried out using 20% of total eluted DNA. The primers used for PCR amplifications were the following: for the first amplification, reverse gene-specific primer, 5′-GCATTTGCCATTCATGCATTTTCCAAATCT-3′ corresponding to the amino acid sequence Arg-23 to Cys-32 of KTX_2; forward primer, the large oligomer of the common linker. High fidelity Pwo DNA polymerase (Boehringer Mannheim) was used according to the manufacturer's instructions. The specific product revealed by Southern blot analysis was recovered, purified using the Wizard PCR Prep DNA system (Promega) and reamplified by a second PCR reaction with a nested gene-specific primer, 5′-CCAAATCTCATTCCAGCATCCTTGCATGG-3′ corresponding to the amino acid sequence Pro-16 to Gly-25 and the same forward primer as the first PCR.

2.5. Walking along the 3′ region of the KTX_2 gene

A second primer extension using genomic DNA digested with *Msp*I was performed with a 5′-biotinylated forward primer (5′-biotin-CATTCGAAAATGAAGGTGTTTTCCG-3′ corresponding to the

*Corresponding author. Fax: (33) 04 91 69 88 39.

0014-5793/97/$17.00 © 1997 Federation of European Biochemical Societies. All rights reserved.
PII S0014-5793(96)01492-5

5′ flanking region of the cDNA of KTX_2) as described below. Extended product was amplified twice with forward primer 5′-AGAATTCCAGTGTCATGT-3′ corresponding to the amino acid sequence Val-1 to Cys-7 of KTX_2.

2.6. Analysis of the LMPCR product

An aliquot of the LMPCR product was electrophoresed, transferred to an Duralon-UV membrane (Stratagene) and hybridized to a gene-specific biotinylated probe. The biotinylated probe was prepared by random priming of DNA of a plasmid containing the KTX_2 cDNA. Hybridized probe was detected using chemiluminescence detection kit (New England Biolabs). Purified PCR products of second amplification were partially sequenced using the Sequenase kit (Amersham) and inserted into pBluescript SK^+ (Stratagene). The recombinant clones were analyzed with standard techniques [6] and sequenced on both strands.

3. Results

3.1. Determination of the genomic region corresponding to the KTX_2 cDNA

A 310 bp DNA fragment encoding the KTX_2 precursor was amplified from scorpion genomic DNA. The PCR product was inserted between the *Sac*I and *Pst*I sites in pBluescript. Four clones containing insert were sequenced on both strands and gave the same sequence (data not shown). The gene con-

```
                                 gatcagctgttgttttcacttt   22

tgtttgctgttgattagcttcatattatgttttctgagagtatttatagtccagt   77
                                                   cagt    4
                                                   cagt    4

¤ o ¤ oo ¤  oooo o             o    ¤ ¤¤¤¤¤¤¤ o
-aattcgttcatattgtattcactgataatattttggattcatttgcaaaaatat  131
gaatacaat-ataaacgaatcactgataatactttg-a---att-gtaaaaatat   53
gatt-cgtttatattgtattcactgataatactttgaatt-----gtaaaaatat   53

 oo o       o    o                o   o
tgtgtaaggtaacttttattttcttaactaactgttcgatatttttaatcagttc  186
tacgaaaggtaattttttgttttcttaactaacttttctatatttttaatcagttc  108
tacgaaaggtaattttttgttttcttaactaacttttctatatttttaatcagttc  108

   ooo          o        o  o                     oo  o
taatggtttaataattcacttttatagacttaaatatatctttattcatcggaca  241
taaccATTTAATAATTGACTTTTATGGATATAAATATATCTTTATTCATTCGAAA  163
taaccATTTAATAATTGACTTTTATGGATATAAATATATCTTTATTCATTCGAAA  163
     Δ 5'-start site (+1)

             o    o    o    o                        o
ATGAAGGTGTTTTTCGCAATTTTAATAACTCTCTTCGTCTGTTCAATGAgtaagt  296
 M  K  V  F  F  A  I  L  I  T  L  F  V  C  S  M          -7
ATGAAGGTGTTTTCCGCAGTTTTGATAATTCTCTTCGTCTGTTCAATGAgtaatt  218
ATGAAGGTGTTTTCCGCAGTTTTGATAATTCTCTTCGTCTGTTCAATGAgtaatt  218
 M  K  V  F  S  A  V  L  I  I  L  F  V  C  S  M          -7

ooo o   o                o ¤  o o    ¤      o      o
tgcatttttttattaatttatatattct-tacgcaaaa-tttaatgattcatgaat  349
acgaattttcattaatttatatattttatatgtaaaaacttaataattcattaat  273
acgaatttttattaatttatatattttatatgtaaaaacttaataattcattaat  273

         o  o¤¤      o   o          o        ¤¤¤    ooo
taagcataggttcttttaacattctagTTATAGGCATTTATGGAGGGGTGGAGA  404
                            I  I  G  I  Y  G  G  V  E     +3
taagcatatgttg--tttaatattttagTTATTGGAATTAATGCA---GTGAGAA  323
taagcatatgttg--tttaatattttagTTATTGGAATTAATGCA---GTGAGAA  323
                            I  I  G  I  N  A     V  R     +2

 ooooo  oo     oooo    oo              o
TAAATGTGAAATGTACTGGTTCTCCTCAATGTTTAAAGCCATGCAAGGATGCTGG  459
I  N  V  K  C  T  G  S  P  Q  C  L  K  P  C  K  D  A  G  +21
TTCCAGTGTCATGTAAACATTCTGGTCAATGTTTAAAACCATGCAAGGATGCTGG  378
TTCCAGTGTCATGTAAACATTCTGGTCAATGTTTAAAACCATGCAAGGATGCTGG  378
I  P  V  S  C  K  H  S  G  Q  C  L  K  P  C  K  D  A  G  +20

AATGAGATTT                                                469
 M  R  F                                                  +24
AATGAGATTTGGAAAATGCATGAATGGCAAATGCGATTGTACACCAAAGTGA TT   432
AATGAGATTTGGAAAATGCATGAATGGCAAATGCGATTGTACACCAAAGTGA TT   432
 M  R  F  G  K  C  M  N  G  K  C  D  C  T  P  K End      +37

TTTTCTTCCATAAAATATTTTCAATGTGTAATAGTTAAATATTGAAAATAAAACT  487
TTTTCTTCCATAAAATATTTTCAATGTGTAATAGTTAAATATTGAAAATAAAACT  487

TATTTCGGAAACAATTaattttaatttcttatatatgtgcttttcatttctgtta  542
TATTTCGGAAACAATTaattttaatttcttatatatgtgcttttcatttctgtta  542
              · 3'-end site (+390)
tgaataatcatgtagaat                                        560
tgaataatcatgtagaat                                        560
```

Fig. 1. The KTX_2 gene. Nucleotide sequence of the KTX_2 gene including the intron and the 5′ and 3′ flanking regions (second and third line). The nucleotide sequence of part of the KTX_1 gene is shown on the first line. The sequence of the transcriptional unit is in capitals. The deduced amino acid sequence is presented in the one-letter code below the codons and is numbered separately starting at the first amino acid of the native toxin. Sequences are aligned for maximum identity and differing nucleotides are indicated by open circles (above the nucleotide sequence) and deletions by dashed circles. Putative TATA boxes are underlined twice. The transcriptional start and stop sites are indicated below the sequence.

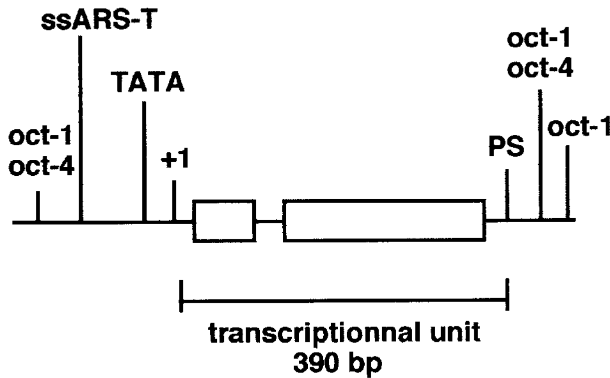

Fig. 2. Schematic representation of the KTX_2 gene. The promoter sequences are indicated (see Section 3). PS, polyadenylation signal.

sists of two exons disrupted by a single intron of 87 bp. The intron is close to the end of the sequence encoding the signal peptide, after the first base of a serine codon at position −6. The intron begins with GT and ends with AG, consistent with previously reported introns [10].

3.2. Determination of the nucleic sequences flanking the KTX_2 gene

The regions upstream and downstream of the 310 bp fragment encoding the KTX_2 gene were amplified by LMPCR [9]. Genomic DNA of the Tozeur scorpion was digested with *Mbo*I and used as the template for primer extension with forward 5′-biotinylated oligonucleotide encoding the C-terminus of KTX_2. Two PCR fragments of 480 bp and 520 bp were obtained, but only the smallest was specific as assessed by Southern blot analysis (data not shown). This product was isolated and sequenced. Three different nucleic acid sequences were obtained (Fig. 1). Two encoded the KTX_2 precursor and differed only in their 5′ part (nt 7–27): they did not contain the *Mbo*I restriction site (GATC) on their 5′ sides as was expected. Surprisingly, the third kind fragment was larger (500 bp) than the others (470 bp) and encoded part of the N-terminus of the KTX_1 precursor [4]. The 5′ side of the putative KTX_1 gene contains the expected *Mbo*I site. A similar experiment with the reverse 5′-biotinylated oligonucleotide encoding for the 5′ non-coding part of KTX_2 cDNA allowed the amplification of a single 470 bp fragment. This genomic fragment was assumed to be the 3′ part of the KTX_2 gene including the putative polyadenylation signal.

The cap site, as determined by primer extension of the KTX_2 gene (data not shown), was 50 nucleotides upstream from the Met initiation codon and was preceded by the bases aacc, like the cap site of the α-toxin AaH I′ [5]. A search for consensus eukaryotic promoter sequences, using the TESS basic Query program, found a putative TATA box upstream of the cap site at position −23, relative to the start site. Surprisingly, no CAAT box was found further upstream. The

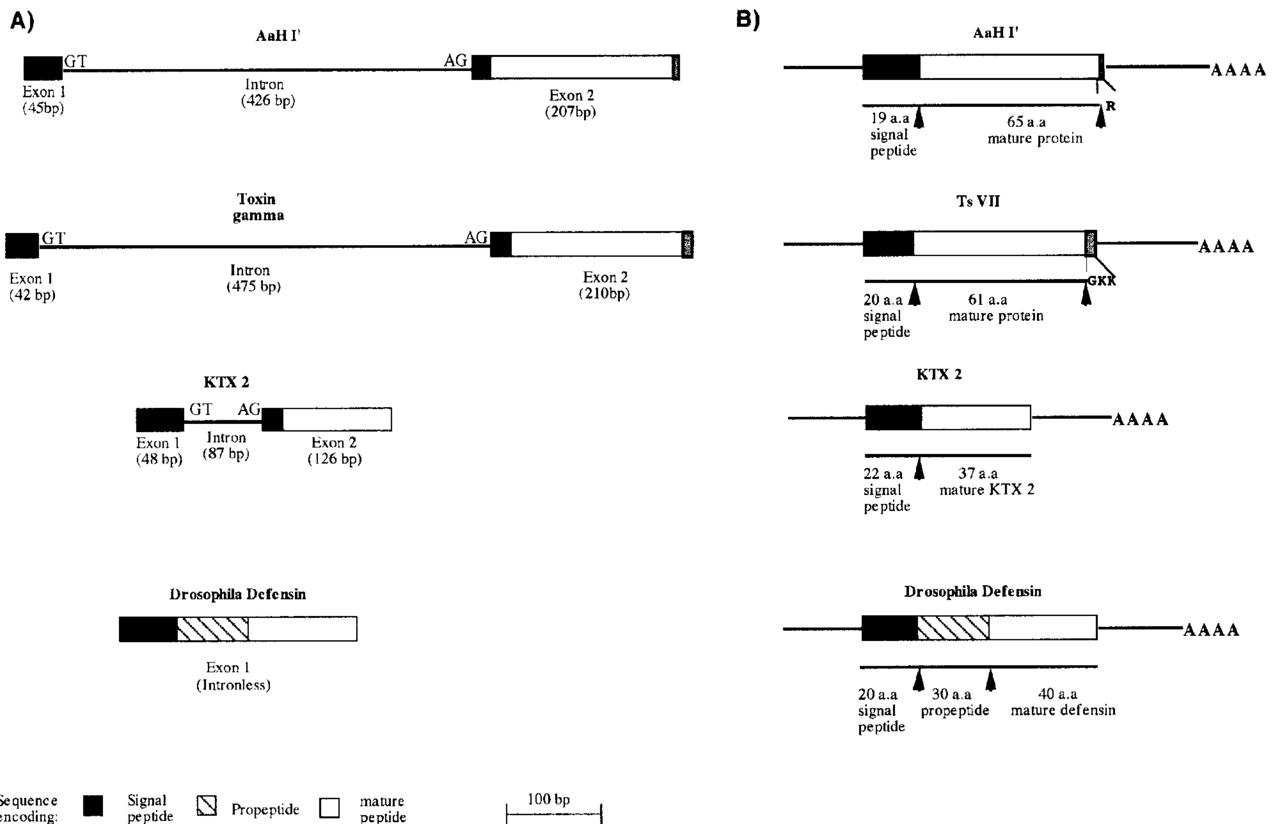

Fig. 3. Domain organization of scorpion toxin genes and cDNAs. A: Genes of toxins active on voltage-dependent Na^+ channels, i.e. from top to bottom, the α-toxin AaH I′ from the Tunisian *Androctonus australis* [4]; the β-toxin γ from the Brazilian scorpion *Tityus serrulatus* [14]; gene of toxins active on the voltage-dependent K^+ channel, i.e. KTX_2 (this work); gene of the defensin from the fly *Drosophila megalogaster* [15]. B: cDNAs of toxins active on voltage-dependent Na^+ channels, i.e. from top to bottom, the α-toxin AaH I′ from the Tunisian *Androctonus australis* [19]; the β-toxin γ from the Brazilian scorpion *Tityus serrulatus* [20]; cDNA of toxins active on the voltage-dependent K^+ channel, i.e. KTX_2 [4]; gene of the defensin from the fly *Drosophila megalogaster* [15].

Table 1
Comparison of the scorpion toxin gene introns

Scorpion gene	A+T content of intron (%)	5′ Splice donor	Intron size (bp)	3′ Splice acceptor
Ts γ	73	gtaagctgaa	475	gttaacatag
AaH I′	75	gtaagatta	425	ctgactacag
Ts IV-5	53	gtaagatttt	347	taaattttag
KTX_2	86	gtaattacga	87	aatattttag
KTX_1	78	gtaatttgca	87	aacattctag
Consensus	high A+T content	gtaagtnnnn		nnntttttag

Comparison of sequenced scorpion toxin genes: genes of toxins active on voltage-dependent Na^+ channels, i.e. from top to bottom, the β-toxin γ from the Brazilian scorpion *Tityus serrulatus* [14]; the α-toxin AaH I′ from the Tunisian *Androctonus australis* [4] and the α-toxin Ts IV-5 from the Brazilian scorpion *Tityus serrulatus* [13]; gene of toxins active on voltage-dependent K^+ channels (this work); A+T content of intron sequences; 5′ and 3′ intron boundaries of the toxins compared to the proposed arachnid consensus sequence for 5′ and 3′ intron boundaries [14]. Sizes of the various introns.

proximal upstream region was searched for putative *cis*-regulatory sequences: Octamer-1 and -4 (Oct-1 and Oct-4) binding sites [11] and a ssARS-T binding site [12] were found (Fig. 2). The introns were in positions similar to those described for the genes of the α-toxins AaH I′ [5] and Ts IV-5 [13], and the β-toxin Ts γ [14], i.e. in the sequence encoding the C-terminal part of the signal peptide. Fig. 3 shows the organization of the short toxin precursor and genes compared to those of the long α- and β-scorpion toxin gene and insect defensin. There is no significant similarity between the introns other than the introns of KTX_1 and KTX_2 (75% identity) and their lengths differ. The insect defensin gene is intronless. Table 1 shows the A+T composition of the introns. All the introns are rich in A+T (up to 86%). The 5′ and 3′ intron boundaries were assigned according to the cDNA sequences and are consistent with the consensus sequence for intron borders of genes from arachnids, i.e. $GTAAGT(n)_4$ and $(n)_3TTTTTAG$ [14].

4. Discussion

We report the first nucleotide sequence and the organization of the genes encoding 'short' scorpion toxins active on K^+ channels. This allows their comparison with those of 'long' scorpion toxins active on Na^+ channels and with insect defensins.

The KTX_1 [4] found in the venom of animals collected in Chellala (Algeria) and identical to the KTX of *Androctonus mauretanicus*, was not found in the venom of the animals collected in Tozeur (Tunisia). However, a gene encoding KTX_1 was found in the genome of these animals. The absence of a TATA box in the promoter region may indicate that the gene is not functional. A TATA box was found in the upstream sequences of the genes encoding KTX_2 (this work) and AaH I′ [5] which are really expressed.

The upstream sequences of the KTX_2 gene were compared to those of the *Drosophila* defensin gene: the upstream sequences of the *Drosophila* defensin gene have no CAAT box, and possess sequence motifs homologous to response elements for transcription factors of the promoters of genes encoding acute-phase proteins in mammals [15]. They both exhibit an Oct-1 site.

The scorpion toxins genes are different in length, and their 5′ untranslated regions including insertion sequences are not homologous. In contrast, in the nucleotide sequences of genes and cDNAs encoding the phospholipases A2 found in snake venoms [16], the protein-coding regions are much more variable than the untranslated regions, including introns, which are highly conserved ($>79\%$).

The introns of scorpion toxins genes have very high A+T content. A high A+T content has been suggested to be a prerequisite for intron splicing [17]. It is surprising and rare to find an intron at the end of the signal peptide [13]. The significance of this is unclear.

Foldons, which are quasi-independently folding units of large proteins, have been recently compared with the exons and structural modules of 30 proteins of 56–374 residues [18]. The exon length distribution was very similar to that of foldons. The average size of foldons (38 amino acid residues) is close to the average size of exons (35 amino acid residues). For some proteins, the foldon junctions were close to the intron positions. However, it was difficult for the authors to conclude for all the proteins studied that there is a strong correlation between foldons and exons. In our case, the second exon is the same size as the foldable unit of KTX_2 (37 amino acid residues), which constitutes a structural motif in itself.

Acknowledgements: We thank Prof. H. Rochat for his constant interest and support.

References

[1] Martin-Eauclaire, M.-F. and Couraud, F. (1995) in: Handbook of Toxicology (Chang, L.W. and Dyer, E.S., Eds.) pp. 683–716, Marcel Dekker, New York.
[2] Miller, C. (1995) Neuron 15, 5–10.
[3] Bontems, F., Roumestand, C., Boyot, P., Quilquin, B., Doljansky, Y., Menez, A. and Toma, F. (1991) Eur. J. Biochem. 196, 19–28.
[4] Laraba-Djebari, F., Legros, C., Crest, M., Céard, B., Romi, R., Mansuelle, P., Jacquet, G., Van Rietschoten, J., Gola, M., Rochat, H., Bougis, P.E. and Martin-Eauclaire, M.-F. (1994) J. Biol. Chem. 269, 32835–32843.
[5] Delabre, M.-L., Pasero, P., Marilley, M. and Bougis, P.E. (1995) Biochemstry 34, 6729–6736.
[6] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[7] Sanger, F., Nicklen, D. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[8] Mueller, P.R. and Wold, B. (1989) Science 246, 780–786.
[9] Warshawsky, D. and Miller, L. (1994) BioTechniques 16, 792–798.
[10] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[11] Pruijn, J.M., Van Driel, W., Van Miltenburg, R.T. and Van der Vliet, P.C. (1987) EMBO J. 6, 3771–3778.
[12] Schmidt, A.M.A., Herterich, S.U. and Krauss, G. (1991) EMBO J. 10, 981–985.
[13] Corona, M., Zurita, M., Possani, L.D. and Becerril, B. (1996) Toxicon 34, 251–256.

[14] Becerril, B., Corona, M., Mejia, M.C., Martin, B.M., Lucas, S., Bolivar, F. and Possani, L.D. (1993) FEBS Lett. 335, 6–8.
[15] Dimarcq, J.L., Hoffmann, D., Meister, M., Bulet, P., Lanot, R., Reichhart, J.-M. and Hoffmann, J.A. (1994) Eur. J. Biochem. 221, 201–209.
[16] Ogawa, T., Kitajima, M., Nakashima, K.-i., Sakaki, Y. and Ohno, M. (1995) J. Mol. Evol. 41, 867–877.
[17] Googall, G.J. and Filipowicz, W. (1989) Cell 58, 473–483.
[18] Panchenko, A.R., Luthey-Schulten, Z. and Wolynes, P. (1996) Proc. Natl. Acad. Sci. USA 95, 2008–2013.
[19] Bougis, P.E., Rochat, H. and Smith, L.A. (1989) J. Biol. Chem. 264, 19259–19265.
[20] Martin-Eauclaire, M.-F., Céard, B., Ribeiro, A.M., Diniz, C.R., Rochat, H. and Bougis, P.E. (1992) FEBS Lett. 302, 220–222.