J. Biochem. 122, 1252-1259 (1997)

A Novel Neurotoxin, Cobrotoxin *b*, from *Naja naja atra* (Taiwan Cobra) Venom: Purification, Characterization, and Gene Organization[1]

Long-Sen Chang,*,[2] Yii-Cheng Chou,† Shinne-Ren Lin,* Bin-Nan Wu,* Jordge Lin,* Enjong Hong,* Yuh-Ju Sun,‡ and Chwan-Deng Hsiao‡

**Department of Biochemistry and Pharmacology, Kaohsiung Medical College, Kaohsiung, Taiwan; †Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan; and ‡Crystallography Laboratory, Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan*

Received for publication, August 18, 1997

A novel neurotoxin, cobrotoxin *b*, was isolated from *Naja naja atra* (Taiwan cobra) venom by successive chromatographies on gel filtration and SP-Sephadex C-25 columns. The yield of this novel toxin was 5% of that of cobrotoxin from the same venom. Its neurotoxicity determined as the inhibition of acetylcholine-induced muscle contractions was approximately 50% of that of cobrotoxin. Cobrotoxin *b* consists of 61 amino acid residues including 8 cysteine residues. Moreover, there are 12 amino acid substitutions between cobrotoxin *b* and cobrotoxin. The genomic DNA, with a size of 2,386 bp, encoding the precursor of cobrotoxin *b* was isolated from the liver of *N. naja atra*. The gene consists of three exons separated by two introns. This exon/intron structure is essentially the same as that reported for the cobrotoxin gene. Moreover, the nucleotide sequences of the two neurotoxin genes exhibit 92% identity. These results highly suggest that the cobrotoxin *b* and cobrotoxin genes are derived from a common ancestor. Comparative analyses of cobrotoxin *b* and cobrotoxin precursors showed that the protein-coding regions of the exons are more diverse than introns, except for in the signal peptide domain. This indicates that the protein-coding regions may have arised *via* accelerated evolution. BLAST searches for sequence similarity in the GeneBank databases showed that intron 1 of the cobrotoxin *b* and cobrotoxin genes encodes a small nucleolar RNA (snoRNA). However, the snoRNA gene is absent from the gene encoding the *Laticauda semifasciata* erabutoxin *c* precursor (*L. semifasciata* and *N. naja atra* are sea and land snakes, respectively). Since previous studies suggested the potential mobility of snoRNA genes during evolution, we propose that intron insertions or deletions of snoRNA genes occurred with the evolutionary divergence between the sea snake and land snake neurotoxins.

Key words: amino acid sequence, evolutionary divergence, genomic structure, snoRNA gene, Taiwan cobra neurotoxin.

Cobrotoxin is the main neurotoxic protein in *Naja naja atra* (Taiwan cobra) venom (*1*). It is comprises 62 amino acid residues in a single polypeptide chain, and adopts an all β-sheet three-finger loop structure (loops I, II, and III) held together by four disulfide bonds (*2*). In contrast to that several iso-neurotoxins (*i.e.* erabutoxins or κ-bungarotoxins) were isolated from the venom of *Laticauda semifasciata* or *Bungarus multicinctus* (*3, 4*), previous studies showed that cobrotoxin was the only neurotoxin isolated from *N. naja atra* venom. Alternatively, seven cardiotoxin analogues have been isolated from *N. naja atra* venom (*5, 6*). However, notable variation in the content of each neurotoxin or cardiotoxin was observed during the purification of these toxic proteins from the venom. To explain this phenomenon, one may propose that the venom was collected from a number of snakes which individually secreted different neurotoxins or cardiotoxins. Another possibility is that several neurotoxins and cardiotoxins genes co-exist in the snake genome, but that their transcription and/or translation are differently regulated. Nevertheless, there is no direct evidence supporting either proposition. Our recent studies showed that the nucleotide sequences of the cobrotoxin and cardiotoxin genes exhibit 84.2% identity. Moreover, the two genes exhibited virtually identical overall organizations, *i.e.* three exons separated by two introns, which are inserted in similar positions of the genes' coding regions (the sequence data for the genomic structures of cobrotoxin and cardiotoxin 4 have been deposited in the EMBL, GeneBank, and DDBJ nucleotide sequence databases under the accession numbers, Y12492 and Y12493, respectively). These results suggested that the cobrotoxin and cardiotoxin precursors may have arisen from a common ancestor, and were probably produced

[1] This work was supported by a grant, NSC 87-2314-B037-007, from the National Science Council, ROC (to L.S. Chang). The amino acid sequence data for cobrotoxin *b* have been deposited in the SWISS-PROT protein sequence databases under the accession number, P80958. The sequence data for the genomic structure of cobrotoxin *b* have been deposited in the EMBL, GeneBank, and DDBJ nucleotide sequence databases under the accession number, Y13399.

[2] To whom correspondence should be addressed. Fax: +886-7-3218309, E-mail: lschang@mail.nsysu.edu.tw

through gene duplications. The fact that seven cardiotoxins have been isolated from *N. naja atra* venom is likely to reflect that the cardiotoxin gene evolved at a more accelerated rate than the neurotoxin gene did. However, a novel neurotoxin gene was identified during our extensive studies on the genetic structures of cobrotoxin and cardiotoxins. This novel neurotoxin, designated as cobrotoxin *b*, was also successfully isolated from *N. naja atra* venom. This indicates that the venom glands of *N. naja atra* secrete at least two kinds of neurotoxins. However, the yield of cobrotoxin *b* is only approximately 5% of that of cobrotoxin. Since the *N. naja atra* genome contains both the cobrotoxin and cobrotoxin *b* precursors, the finding that the contents of the two isotoxins in venom are different supports that the neurotoxin genes of *N. naja atra* should be differently regulated. The results are presented in this paper.

MATERIALS AND METHODS

Preparation of Genomic DNA from N. naja atra Liver—*N. naja atra* livers were ground into a fine powder in liquid nitrogen. The genomic DNA was extracted from the powder in the presence of SDS and proteinase K (*7*).

PCR Amplification and Cloning—Two oligonucleotide primers in sense and antisense orientations based on the promoter and 3′-noncoding regions of the cobrotoxin and cardiotoxin genes with the forward sequence, 5′-GTCCAGGTGCCCAGGTTTTGTATG-3′, and the reverse one, 5′-GGATGGTCCTTGATGGATGAGAGC-3′, were synthesized.

PCR was carried out in 100 μl reaction buffer containing 1 μg genomic DNA, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM dNTP, 50 pmol primers, and 5 U Taq polymerase. A thermal cycler was used for 30 cycles of reactions, each cycle comprising denaturation at 94°C for 1 min, annealing at 65°C for 1 min and extension at 72°C for 1 min.

The PCR products were cloned into the pCRII vector according to the TA-cloning procedure (Invitrogen, San Diego, USA).

Primer Extension Analysis—Synthetic oligonucleotides (5′-CTAAGTCCAGGCACACGATTGTCACC-3′) corresponding to the antisense sequence in the first exon encoding the signal peptide of the cobrotoxin *b* gene were used in the primer extension analysis. Total RNA from *N. naja atra* venom glands was isolated by the guanidinium isothiocyanate/phenol/chloroform method (*8*). The primer (50 pmol) was hybridized to 6 ng of total RNA in the reverse transcription buffer (50 mM NaCl, 34 mM Tris-HCl, and 5 mM DTT, pH 8.3) at 85°C for 10 min, followed by at 45°C for 12 h. dATP, dGTP and dTTP, each at a final concentration of 0.25 mM, and 8 μCi of [α-^{32}P]dCTP were added to the annealing mixture, followed by incubation with 3 units of MMLV reverse transcriptase in a final volume of 10 μl at 42°C for 15 min. One microliter of the dNTP mixture (2 mM each) was then added to the reaction mixture, followed by chasing for another 15 min at 42°C. The reaction was stopped by adding 5 μl of 98% formamide containing 10 mM EDTA, 0.3% xylene cyanol, and 0.3% bromophenol blue. The cobrotoxin *b* gene served as a template for dideoxynucleotide sequencing with the same oligonucleotide primer. The sequencing reaction products were run parallel with the primer-extended fragments on a 6% polyacrylamide/urea sequencing gel.

DNA Sequencing—Exonuclease III-generated deletion subclones of the genomic DNA containing the cobrotoxin *b* precursor were prepared with an Erase-a-Base kit (Promega), following the manufacturer's instructions. Sequence analysis was carried out according to the dideoxy method with a sequencing kit (Sequenase sequencing system, USB), with labeling with [^{35}S]dATP (Amersham, >1,000 Ci/mmol). The reaction products were sequenced in a 6% polyacrylamide gel, which was dried and then exposed to a Kodak film for one day at room temperature.

Comparison of Nucleotide Sequences and a Homology Search—For the comparison and analysis of the determined nucleotide sequences, a software package (PC/GENE program; Stratagene, USA) was used for sequence alignment based on percent sequence identity. BLAST searches (*9*) in non-redundant databases were carried out *via* the internet using the GCG software package.

Purification of Cobrotoxin b—Cobrotoxin *b* was purified as follows: 5 g of lyophilized venom was dissolved in 10 ml of 1% acetic acid and then applied to a column of Sephadex G-50 (2×200 cm) at room temperature. The Sephadex column was eluted with 1% acetic acid at a flow rate of 40 ml/min. Fractions of 15 ml each were collected. The fractions containing proteins with molecular masses of lower than 20 kDa were pooled and lyophilized for further purification. The lyophilized sample was dissolved in 10 ml of 0.005 M sodium phosphate (pH 7.0) and then applied to a SP-Sephadex C-25 column (2×95 cm) equilibrated with the same buffer. The column was washed with the starting buffer (0.005 M phosphate buffer, pH 7.0) and then eluted with a salt gradient of 2,400 ml, from 0 to 0.5 M NaCl, in the same buffer. The flow rate was 60 ml/h, and fractions of 15 ml each were collected.

Amino Acid Sequence Determination of Cobrotoxin b—Amino acid sequencing was carried out with an Applied Biosystem 477 A protein sequencer. Reduction and *S*-carboxymethylation of cobrotoxin *b* were performed essentially according to the methods previously described (*10*). The reduced and *S*-carboxymethylated (RCM) protein was then subjected to automated Edman degradation to determine the N-terminal sequence of cobrotoxin *b* up to 40 residues. Moreover, RCM-cobrotoxin *b* was digested with *Staphylococcus aureus* V8 protease (protein : enzyme, 30 : 1, w/w) in 0.2 M ammonium bicarbonate (pH 7.8) at 37°C for 3 h. The hydrolysate was separated by HPLC on a Synchropak RP-P column (4.6 mm×25 cm) equilibrated with 0.1% trifluoroacetic acid and then eluted with a linear gradient of 7.5-25% acetonitrile over 70 min. Each peptide fraction derived from the hydrolysate of cobrotoxin *b* was subjected to determination of its amino acid composition and sequence.

Assaying of Neurotoxicity—Frogs of either sex, weighing between 80 and 100 g, were pithed, and then the rectus abdominis muscle was quickly excised. The rectus abdominis muscle was cut into two equal strips. One muscle strip was mounted in a 10 ml organ bath with one end fixed and the other end connected to a fore displacement transducer (Grass, Model FT03). The experiments were carried out at room temperature in Ringer's solution (comprising 112 mM NaCl, 5 mM KCl, 25 mM $NaHCO_3$, 1.25 mM $CaCl_2$, and 11.5 mM glucose), and bubbled with a mixture of 95% O_2 and 5% CO_2. The muscle strip was prestretched

to a basal tension of 2.0 g and then equilibrated for 60 min in aerated (95% O_2-5% CO_2) Ringer's solution before the experiments. Then 10 μM acetylcholine was added to induce contractions. After the maximal tension had been achieved, 1 μM cobrotoxin or cobrotoxin *b* was added to the bath to inhibit acetylcholine-induced contractions.

RESULTS

Construction and Sequencing of the Cobrotoxin b Gene—Although seven cardiotoxins have been isolated from *N. naja atra* venom (*5, 6*), the gene encoding the cardiotoxin 4 precursor is the only one for which the genetic structure has been determined. Our recent studies on the cardiotoxin 4 and cobrotoxin genes strongly suggested that they were derived from a common ancestral gene. However, the evolutionary process for the production of cardiotoxin variants still remained to be resolved. Elucidation of the genetic structures of cardiotoxin analogues may, in part, throw light on this problem. Comparative analyses of the genetic structures of *N. naja atra* cardiotoxin 4 and cobrotoxin showed that the nucleotide sequences of the promoter and 3'-untranslational regions are highly conserved. Thus, two primers were designed on the basis of the conserved regions for amplification of the genomic DNAs encoding the cardiotoxin precursors. PCR amplification of *N. naja atra* genomic DNA with the designed primers allowed the isolation of PCR fragments estimated to be

```
   1 GTCCAGGTGCCCAGGTTTTGTATGCAGAAGGGAAGCATTTGCCTTGAAGACAGGGAATGG
  61 TTTGTTCTTCAGCTGATCTGACCTGGCTTCCGTTTTGGTGTCGGGGGCAAAAAGGCCACA
 121 CCCCCTGGCCTTTCCTTTCCCATTGAATCTCCCTGTTCTTCTCTCTAGACATCCTAGATA
 181 TGTGGCCCTGCCCACCCACCCCATCTCAGACGGGTATAAATAAGTGTCTGCTTCCATCGT
 241 GGGCTACACTTTGCAGGCTCCAGAGAAGATTGCAAGATGAAAACTCTGCTGCTGACCTTG
                                          M  K  T  L  L  L  T  L
 301 CTGGTGGTGACAATCGTGTGCCTGGACTTAGGTAAGGCTACCAAACTTTGATTGTGGGTG
       L  V  V  T  I  V  C  L  D  L
 361 GCAATCAAACCTAAAACAAAGGGGCAGAGTAAGGGAGGGTATTCGAGGCAATTGTGGTCC
 421 CTTCTCTTTGGTGGGACTTTCTCACTGTTCTGGTCTGCATAAGGAGGGTCCATGAGAACC
 481 TTTGGGGGAGAAGAGGTCTCTGTTTGGAGAGCCGCCCATGAGACAGGGGCTTTTTGAGCC
 541 TGGCATTCTGGGGTAGACTGCCTCTGAAAGTCGAGGTTCCTTTGCGTAGTGAGTTCTGCC
 601 AAGGGATTACTCCAGGGAATGAGATGAAAAACGTTTCTGCTCACCTCTCTCTGCTTTGCC
 661 ATTCTTCAGTATTTTGGATGGTACAGCTAAAAGGGACCAGAGATGCCTCCTTCAAGTCAA
 721 GCCCTGGACATTCCTCAACATCAACCTGACATTGGTTGCTGCTGCTGCTGCAATTTCTGA
 781 GACTCCTCAAACTTTCATACAACCTTGGGGGTCTCCCATCCTAAGATTAACCAATGCCAA
 841 CCCTGCTTAGCTTCCACCTATATACCAGCTACTTAGGTTTCAATTTCGTTTTATTTCATG
 901 GCCTGTTTCAGAAGGGGTTAGGGTTTTTTTACTCCAGCTTTTATCCCAGCTATGGTGGTA
 961 CAGTGGTTCATTCTGCTTACTGCCAGCACCTGCAGTTCGGTAGTATGGTTGTCTCGGTTT
1021 TCCATCCTTCTGAGGTTGGTAAAATGAGAACCAAGATTTTTAGGGACCATAGGCTGACTC
1081 CATACATTGCCTAGAGAGAGCTGCAAAGGACTGTGAAGCAGTATAAAAGTCTAAGAGCTA
1141 TAGCTATGTAGTTTTCTAATGACCCAAGGGAGTGAAGATACCTAGCACTCCTTCCTCCCA
1201 TTTTCCCCACAGCAGCAATCCTGTGAGGTAGGATGGGCAGAGAGTGGGTGAATAGTCCAA
1261 AATCCCTTAGCTGTCTTTGATGCTTAAAGTGGGACTGGGACTCACTGTCTTCTGTTGATT
1321 GGCCCAGAGTCACCCACTGTTCTTCATGGCCAAAGTGGGACTAGAACTCTCTGTCTCCTG
1381 CTGATTAACCCAGAGTCACCCAGATGGCTTTGATGCCTAAGGTGAGGCCGGAACTCAGGT
1441 TTCCTATTTCTAGACCCTCTGATCTTTCTAGGGAAATTTTCACAGCTAACCTTGTATGAA
1501 AAGATCTATCATCTAATATGTAATAAAATTACAAGATCAGAAAAACAAATATCAAAGTAA
1561 CATTTCTATACATTGTATCTTTCATCCAGGATACACCCTGGAATGTCACAACCAACAATC
                                    G  Y  T  L  E  C  H  N  Q  Q  S
1621 ATCGCAAACTCCAACAACTAAAACTTGTTCAGGGGAGACCAATTGCTATAAAAAGTGGTG
       S  Q  T  P  T  T  K  T  C  S  G  E  T  N  C  Y  K  K  W  W
1681 GTCTGATCACCGTGGTAAGCACAGCTCTTATCTTCTTGCACAACATTTTTTTCTTCTTGA
       S  D  H  R
1741 ACCCAGACTGGAGCTCCCAAATACAACTGAGAACTGGAGGTCACATAGGACCTGTCTGTG
1801 TCATGCCGGGTTCCCCTGAGGTCACATCCTGTTGCCAGACAGACTAAGCAATGCAAGGAA
1861 AAGGGGTCTTGGACTGGGGTGCCCCATCTTTCTCAGCAATAGCCCAGACAGGAGGGGGAA
1921 GCGTTCTTTAGAGAAAGGGGGAGGCCAATGCAGGGAGAGCTGCAGGTTTGTTCTGGAATT
1981 GAGGGAAATTGGATGAAGCCATGAAACATCCAAGAAAGATGCCATTGCCCCTGAGTCTGT
2041 CAAAACATGAGACTCTCCCAGAAATCTCTTTTTCATTCATTTGCTCCTGAGTTTTGAGTA
2101 GAGGAGGAAGGAGGTCTGCCTTCCTAATGGGAGGGGATGCATGAATGTGTTCAGTGTGGG
2161 ATTACCATTTTGTTCAGCAGCAGCCACAGGTCTGTTCTTTGGTATTTTTAAAATAATCAT
2221 TCTACTTTGTGCTTCCTTCACAGGAACTATAATCGAAAGGGGATGTGGTTGCCCTAAAGT
                                G  T  I  I  E  R  G  C  G  C  P  K  V
2281 GAAGCCCGGTGTTAATCTTAACTGTTGCACAACAGACAGATGCAACAATTAGCTCTCCGA
       K  P  G  V  N  L  N  C  C  T  T  D  R  C  N  N end
2341 GTGGCTAAATTCCTTGAGTTTTGCTCTCATCCATCAAGGACCATCC
```

Fig. 1. **Nucleotide sequence of the cobrotoxin *b* gene.** The deduced amino acid sequence is presented below the coding parts of the exons, and the signal peptide is underlined. The TATA box is indicated by bold letters and underlining. The exon regions are underlined. The genetic structure of cobrotoxin *b* appears in the EMBL, GeneBank, and DDBJ nucleotide sequence databases under the accession number, Y13399.

about 2.4 kb in length (data not shown). The DNA fragments were then subcloned with a TA-cloning kit. More than 20 clones were selected for nucleotide sequencing. In addition to the clone containing the cardiotoxin VII precursor identified, another gene encoding a cobrotoxin analogue, designated as cobrotoxin *b*, was determined (Fig. 1). Comparison of the cDNA nucleotide sequence for cobrotoxin with the genomic structure of cobrotoxin *b* allowed us to assign intron/exon boundaries. Alignment of the deduced cDNA sequence of cobrotoxin *b* with that of cobrotoxin revealed that they exhibit 91.7% homology. The deduced amino acid sequence of cobrotoxin *b* revealed that this novel protein comprises 61 amino acid residues including 8 cysteine residues (Fig. 2). One of the consecutive Gly residues at positions 19 and 20 of cobrotoxin was absent in cobrotoxin *b*. Additionally, there were 12 amino acid substitutions between cobrotoxin and cobrotoxin *b*.

Separation and Characterization of Cobrotoxin b—Since the cobrotoxin *b* gene was found in the *N. naja atra* genome, efforts have been made to purify this novel protein from *N. naja atra* venom. The components of *N. naja atra* venom were separated by successive chromatographies on gel filtration and ion exchanger columns. After passage through a gel filtration column, the fractions containing the components with molecular weights of lower than 20 kDa were collected for further purification on an SP-Sephadex C-25 column, being separated into 15 fractions (Fig. 3). Each fraction was subjected to amino acid sequencing to determine the N-terminal sequence. The determined N-terminal sequence up to 40 residues revealed that fractions 6 and 8 contained cobrotoxin *b* and cobrotoxin, respectively (Fig. 3). Cobrotoxin *b* was further purified by HPLC on a SynChropak RP-P column (4.6 mm × 25 cm) eluted with a linear gradient of 5-50% acetonitrile over 80 min at a flow rate of 0.8 ml/min (data not shown). The purified cobrotoxin *b* behaved as a homogeneous protein on polyacrylamide gel electrophoresis (inset in Fig. 3), and had an apparent molecular mass of about 7 kDa, as revealed on SDS-

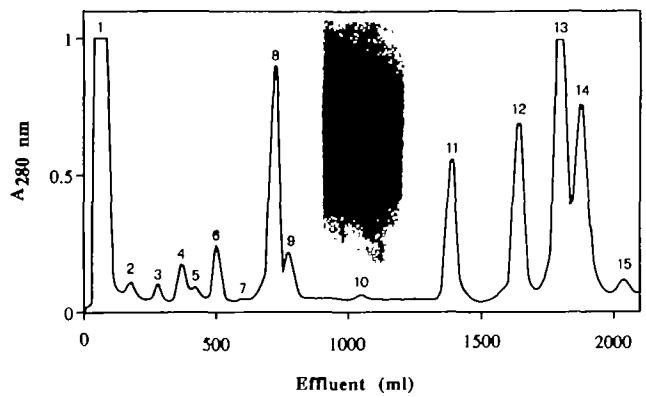

Fig. 3. **Chromatographic profile showing the separation of *N. naja atra* venom on a SP-Sephadex C-25 column.** The column was eluted with a salt gradient of 2,400 ml, from 0 to 0.5 M NaCl, in the same buffer. Peaks 6 and 8 are cobrotoxin *b* and cobrotoxin, respectively. Inset: Polyacrylamide gel electrophoresis of cobrotoxin *b* and cobrotoxin. Electrophoresis on a 7% polyacrylamide gel (pH 4.5) was carried out as described by Yang and Chang (*30*). Lane 1, cobrotoxin *b*; lane 2, cobrotoxin.

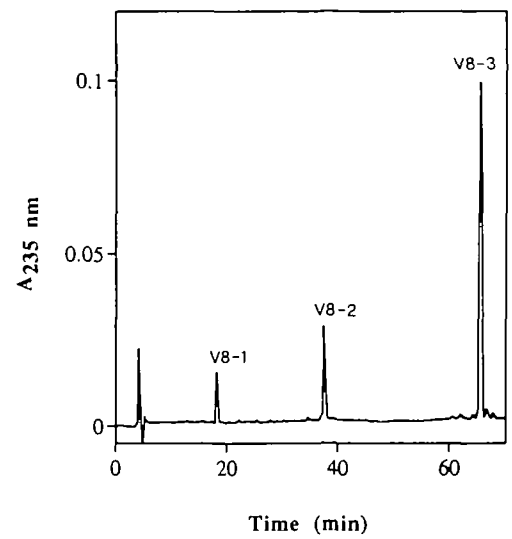

Fig. 4. **Separation of the peptide fragments of RCM-cobrotoxin *b* derived on hydrolysis with *S. aureus* V8 protease.** One milligram of proteolytic hydrolysate was applied to a SynChropak RP-P column (4.6 mm × 25 cm) equilibrated with 0.1% trifluoroacetic acid and then eluted with a linear gradient of 7.5-25% acetonitrile over 80 min. The flow rate was 0.8 ml/min and the effluent was monitored at 235 nm.

```
                                                                    20
LeuGluCysHisAsnGlnGlnSerSerGlnThrProThrThrThrGlyCysSerGlyGly
LeuGluCysHisAsnGlnGlnSerSerGlnThrProThrThrLysThrCysSerGly---
|<------------------------------------------ V8-1 ------------------------------------------

                                                                    40
GluThrAsnCysTyrLysLysArgTrpArgAspHisArgGlyTyrArgThrGluArgGly
GluThrAsnCysTyrLysLysTrpTrpSerAspHisArgGlyThrIleIleGluArgGly
-->|<------------------------------------------ V8-3 ------------------------------------------>|<--

                                                                    60
CysGlyCysProSerValLysAsnGlyIleGluIleAsnCysCysThrThrAspArgCys
CysGlyCysProLysValLysProGlyValAsnLeuAsnCysCysThrThrAspArgCys
-------------------------------------------V8-2 ------------------------------------------

AsnAsn      Cobrotoxin
AsnAsn      Cobrotoxin b
-->|
```

Fig. 2. **Alignment of the amino acid sequences of cobrotoxin and cobrotoxin *b*.** The amino acid substitutions between cobrotoxin *b* and cobrotoxin are underlined. V8-1, V8-2, and V8-3 represent the peptide fragments of RCM-cobrotoxin *b* derived on digestion with *S. aureus* V8 protease.

polyacrylamide gel electrophoresis.

Cobrotoxin *b* was subjected to reduction and *S*-carboxymethylation. The RCM-cobrotoxin *b* was digested with *S. aureus* V8 protease, and then the hydrolysate was separated into three peptide fractions (V8-1, V8-2, and V8-3) (Fig. 4). The results of amino acid composition and sequence determinations revealed that V8-1, V8-2, and V8-3 were the segments comprising positions 1-20, 38-61, and 21-37 of cobrotoxin *b* (Fig. 2), respectively. The amino acid sequence of cobrotoxin *b* is exactly the same as that deduced from the cobrotoxin *b* gene (Figs. 1 and 2). As shown in Fig. 5, cobrotoxin *b* as well as cobrotoxin effectively inhibited acetylcholine-induced muscle contractions. It is evident that cobrotoxin *b* is a neurotoxic protein like cobrotoxin. Alternatively, the degree of inhibition caused by the addition of cobrotoxin *b* was approximately 50% of that observed with the addition of cobrotoxin. This indicates that the neurotoxic activity of cobrotoxin *b* is weaker than that of cobrotoxin.

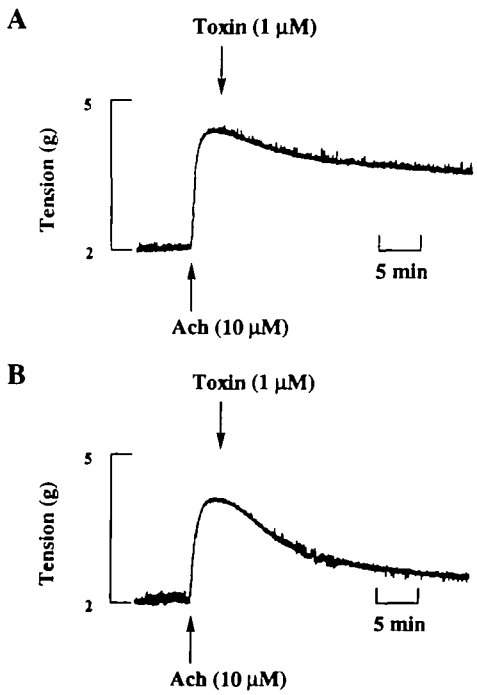

Fig. 5. **Typical recordings of the inhibitory effects of cobrotoxin and cobrotoxin *b* on acetylcholine-induced maximal contractions in isolated frog rectus abdominis muscle.** The added concentrations of acetylcholine and toxins (A, cobrotoxin *b*; B, cobrotoxin) were 10 and 1 μM, respectively.

Structural Organization of the Cobrotoxin b, Cobrotoxin, and Erabutoxin c Genes—The genomic DNA of cobrotoxin *b* is 2,386 bp long (Fig. 1). The transcription-initiation site of the cobrotoxin *b* gene was further determined by primer extension analysis (data not shown). It was found that the transcription-initiation site should be 29 bp upstream from the ATG start codon. The nucleotide sequence, TATAAATAA, corresponding to the consensus sequence of the TATA box was located 25-33 upstream from the transcription-initiation site. This result is exactly the same as those reported for the *N. naja atra* cobrotoxin gene and the *L. semifasciata* erabutoxin *c* gene (a neurotoxin genomic DNA from a sea snake) (*11*).

The gene containing the cobrotoxin *b* precursor have three exons, and the three exons are interrupted by two introns. The sequences of all the exon/intron junctions agree with the GT/AG rule (*12, 13*), and are in agreement with the consensus sequences reported for exon-intron boundaries (*14*). Exon 1 of the cobrotoxin *b* gene encodes a large part of the signal peptide. Exon 2 of the cobrotoxin *b* precursor encodes the remaining part of the sequence of the signal peptide (3 amino acid residues), and the N-terminal half of the mature cobrotoxin *b* (1-32 out of 61 amino acid residues). Exon 3 encodes the carboxyl-terminal half of the toxin and the following 3′-noncoding region. Structurally, the genomic DNAs of cobrotoxin and erabutoxin *c* also comprise three exons and two introns (Fig. 6). The cobrotoxin and erabutoxin *c* genes are similar to the cobrotoxin *b* gene in that the introns interrupt their coding regions in the same positions (according to their amino acid se-

TABLE I. **The levels of similarity of the cobrotoxin *b*, cobrotoxin, and erabutoxin *c* genes.**

Gene segment[a]	Similarity (%)		
	CBTXb/CBTX[b]	CBTX/ERTXc[b]	CBTXb/ERTXc[b]
5′-Flanking region	98.4	89.3	89.3
Exon 1	98.8	97.6	96.4
Intron 1	92.1	91.4	92.4
Exon 2	92.4	74.1	75.2
Intron 2	89.4	87.5	87.7
Exon 3	86.7	79.0	87.4

[a]Comparison of related parts of the *N. naja atra* cobrotoxin *b* gene sequence with the gene sequences of *N. Naja atra* cobrotoxin and *L. semifasciata* erabutoxin *c*. The intron and exon designations correspond to those of the cobrotoxin *b* gene. [b]CBTXb, CBTX, and ERTXc represent cobrotoxin *b*, cobrotoxin, and erabutoxin *c*, respectively.

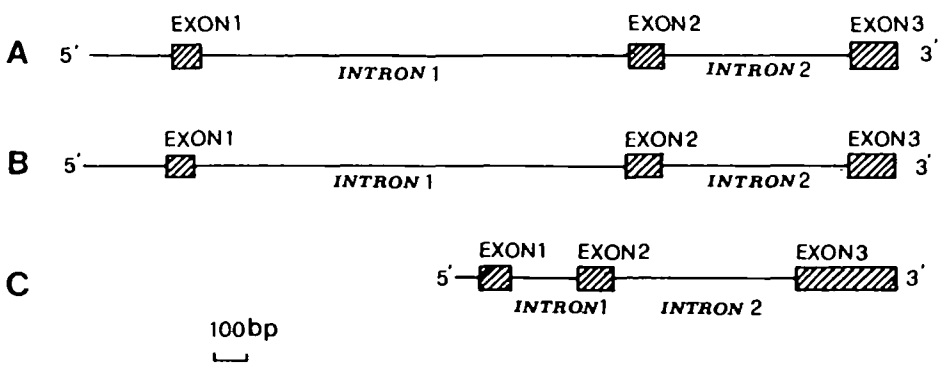

Fig. 6. **Gene organization of the cobrotoxin *b*, cobrotoxin, and erabutoxin *c* genes.** A, cobrotoxin *b* gene; B, cobrotoxin gene; C, erabutoxin *c* gene. The shorter length of the exon 3 regions of the cobrotoxin *b* and cobrotoxin genes is due to that only parts of their 3′-untranslation regions were determined in the present study.

quences). Sequence comparison of the cobrotoxin *b* gene with the cobrotoxin and erabutoxin *b* precursors revealed that the sequence identity of intron regions of these genes are higher than that observed for their protein-coding regions (Table I). Identical numbers of amino acid residues are encoded by exons 1 and 3 of these neurotoxin precursors. However, the number of amino acid residues encoded by the exon 2 region of the cobrotoxin *b* gene is one less than those of the cobrotoxin and erabutoxin *c* genes (Table II). The splice donor and acceptor sequences are highly conserved (Table II). Although similar sizes for intron 2 were noted (549, 551, and 538 bp for the cobrotoxin *b*, cobrotox-

TABLE II. **Comparison of the structural organizations of the cobrotoxin *b*, cobrotoxin, and erabutoxin *c* genes.**

	Exon number	Exon length	5′-splice donor sequence	Intron number	Intron length	3′-splice acceptor sequence
N. naja atra	1	84	CTTAGgtaag	1	1,258	tccagGATAC
cobrotoxin *b*	2	105	CCGTGgtaag	2	549	cacagGAACT
	3[a]	143				
N. naja atra	1	84	CTTAGgtaag	1	1,269	tccagGATAC
cobrotoxin	2	108	CCGTGgtaag	2	551	cacagGATAT
	3[a]	143				
L. semifasciata	1	84	CTTAGgtaag	1	197	tccagGATAC
erabutoxin *c*	2	108	CCGTGgtaag	2	538	cacagGAACT
	3[a]	301				

[a]The shorter length of the exon 3 regions of the cobrotoxin *b* and cobrotoxin precursors is due to that only a part of the 3′-untranslational region was determined in the present study.

```
CBTXb-I1  GTAAGGCTACCAAACTTTGATTGTGGGTGGCAATCAAACCTAAAACAAAG -50
          ||||||| | |||||||||||  |||||| ||||||||| |||||||||
ERTXc-I1  GTAAGGCCAGCAAACTTTGATCTTGGGTGACAATCAAACCCAAAACAAAG -50

          GGGCAGAGTAAGGGAGGGTATTCGAGGCAATTGTGGTCCCTTCTCTTTGG -100
          |||||||||||||||||||||| ||||||  |||||||||| ||  |||||
          GGGCAGAGTAAGGGAGGGTATTTGAGGCAGTTGTGGTCCCATCCTTTTGG -100

          TGGGACTTTCTCACTGTTCTGGTCTGCATAAGGAGGGTCCATGAGAACCT -150
          ||||||||||||| |||||||||||||||||||| ||||||| ||||||
          TGGGACTTTCTCACAGTTCTGGTCTGCATAAGGAGAGTCCATGGGAACCT -150

          TTGGGGGAGAAGAGGTCTCTGTTTGGAGAGCCGCCCATGAGACAGGGGCT -200
          ||||||||||||||||  ||
          TTGGGGGAGAAGAGGTC-ATG---------------- ----------- -170

          TTTTGAGCCTGGCATTCTGGGGTAGACTGCCTCTGAAAGTCGAGGTTCCT -250
          TTGCGTAGTGAGTTCTGCCAAGGGATTACTCCAGGGAATGAGATGAAAAA -300
          CGTTTCTGCTCACCTCTCTCTGCTTTGCCATTCTTCAGTATTTTGGATGG -350
          TACAGCTAAAAGGGACCAGAGATGCCTCCTTCAAGTCAAGCCCTGGACAT -400
          TCCTCAACATCAACCTGACATTGGTTGCTGCTGCTGCTGCAATTTCTGAG -450
          ACTCCTCAAACTTTCATACAACCTTGGGGGTCTCCCATCCTAAGATTAAC -500
          CAATGCCAACCCTGCTTAGCTTCCACCTATATACCAGCTACTTAGGTTTC -550
          AATTTCGTTTTATTTCATGGCCTGTTTCAGAAGGGGTTAGGGTTTTTTTA -600
          CTCCAGCTTTTATCCCAGCTATGGTGGTACAGTGGTTCATTCTGCTTACT -650
          GCCAGCACCTGCAGTTCGGTAGTATGGTTGTCTCGGTTTTCCATCCTTCT -700
          GAGGTTGGTAAAATGAGAACCAAGATTTTTAGGGACCATAGGCTGACTCC -750
          ATACATTGCCTAGAGAGAGCTGCAAAGGACTGTGAAGCAGTATAAAAGTC -800
          TAAGAGCTATAGCTATGTAGTTTTCTAATGACCCAAGGGAGTGAAGATAC -850
          CTAGCACTCCTTCCTCCCATTTTCCCCACAGCAGCAATCCTGTGAGGTAG -900
          GATGGGCAGAGAGTGGGTGAATAGTCCAAAATCCCTTAGCTGTCTTTGAT -950
          GCTTAAAGTGGGACTGGGACTCACTGTCTTCTGTTGATTGGCCCAGAGTC -1000
          ACCCACTGTTCTTCATGGCCAAAGTGGGACTAGAACTCTCTGTCTCCTGC -1050
          TGATTAACCCAGAGTCACCCAGATGGCTTTGATGCCTAAGGTGAGGCCGG -1100
          AACTCAGGTTTCCTATTTCTAGACCCTCTGATCTTTCTAGGGAAATTTTC -1150
          ACAGCTAACCTTGTATGAAAAGATCTATCATCTAATATGTAATAAAATTA -1200

          CAAGATCAGAAAAACAAATATCAAAGTAACATTTCTATACATTGTATCTT -1250
                                          |||||||||||| ||||||
          ------------------------------TTTCTATACATTATATCTT -189

          TCATCCAG -1258
          ||||||||
          TCATCCAG -197
```

Fig. 7. **Alignment of the first intronic sequences of cobrotoxin *b* and erabutoxin *c* genes.** CBTXb-I1 and ERTXc-I1 represent the first intronic sequences of the cobrotoxin and erabutoxin *c* genes, respectively. The first intronic sequence of the erabutoxin *c* gene is underlined.

in, and erabutoxin *c* genes, respectively), the size of intron 1 of the erabutoxin *c* gene is notably shorter than those of the cobrotoxin *b* and cobrotoxin genes (Fig. 6 and Table II). Intron 1 of the cobrotoxin *b*, cobrotoxin, and erabutoxin *c* genes is 1,258, 1,269, and 197 bp long, respectively. Obviously, during the evolution of snake neurotoxin genes the second intron remained in a more conserved state than the first intron. This seems to imply that intron 2 may play a more critical role in pre-mRNA stability and/or processing. Although notable variation in the size of intron 1 was found in these neurotoxin precursors, the first intronic sequences neighboring the splice sites are highly conserved. The sequence identity of intron 1 between the erabutoxin *c* gene and the cobrotoxin *b* gene (or cobrotoxin gene) is up to 91% (Table I). Alignment of the first intronic sequences of these genes showed that the middle region at intron 1 of the cobrotoxin *b* and cobrotoxin genes was absent from that of the erabutoxin *c* gene (Fig. 7).

DISCUSSION

In the present study, a novel neurotoxin was purified from *N. naja atra* venom, which exhibits approximately 50% neurotoxicity compared to that of cobrotoxin. Since the genes for cobrotoxin analogues are present independently in the *N. naja atra* genome, the genetic structures reported here clearly exclude *N. naja atra* neurotoxin messengers produced through alternative splicing. The observation that the yield of cobrotoxin *b* was approximately 5% of that of cobrotoxin reflects that the genes encoding the cobrotoxin and cobrotoxin *b* precursors should be differently regulated at the transcriptional and/or translational levels. Alternatively, the cobrotoxin *b* and cobrotoxin genes share the same exon-intron structures and exhibit a high degree of sequence identity, suggesting that the two genes arose from a common ancestor, and probably co-exist in one snake genome due to gene duplications. The exon regions of the cobrotoxin *b* and cobrotoxin genes are notably more diverse than their intron regions. This implies that the exon and intron regions evolved *via* different processes. The same has been noted for snake phospholipase A_2 isozymes genes (*15-20*). In view of the conserved size of intron 2 of neurotoxin precursors, it is evident that there was strong selection for intron 2 during the evolution of snake neurotoxins. On the contrary, notable variation in the size of intron 1 was observed for neurotoxin genes. This showed that deletion of a part of the intron 1 sequence from the presumed erabutoxin *c* gene (or the addition of this sequence to the cobrotoxin and cobrotoxin *b* genes) occurred after the evolutionary divergence of land and sea snakes. Thus, the middle region of intron 1 of the cobrotoxin *b* and cobrotoxin genes seems not to be crucial for pre-mRNA processing, as evidenced by the finding that this region is deleted in intron 1 of the erabutoxin *c* gene (Fig. 7). Recent studies showed that the first intron region of genes, including the platelet-derived growth factor A-chain gene and sea urchin arylsulfatase genes, contains elements that regulate their transcription (*21, 22*). The appearance of such essential elements in the first intron of the cobrotoxin *b* and cobrotoxin genes remained to be confirmed. Alternatively, BLAST searches for sequence similarity in Genebank databases showed that parts of the first intronic sequences of cobrotoxin (positions 770-864) and cobrotoxin *b* (positions 774-868) exhibit the structural hallmarks of a small nucleolar RNA (snoRNA). There are 12-nt and 15-nt tracts of complementarity to phylogenetically conserved sequences in 5S rRNA (Fig. 8). Meanwhile, the intronic sequences of cobrotoxin contain the characteristic box D (CUGA) motif, which is a structural feature typical of snoRNA genes. However, the sequence (UGAGA) of box C in this snoRNA gene has one base-deletion, and thus is different from those (UGAUGA) of other subsets of snoRNAs that exhibit great complementarity to invariant sequences in eukaryotic 18S and 28S rRNAs (*23-27*). The snoRNA gene in the cobrotoxin *b* gene retains the box C element, but lacks the characteristic sequence of box D. This is consistent with previous results showing that none of these boxes is universal and that the molecular structures of snoRNAs are diverse (*28, 29*). However, snoRNAs were found to be produced through intronic RNA processing (*23-27*). The results of Kenmochi *et al.* (*27*) suggested the potential mobility of the snoRNA genes during evolution. Moreover, the snoRNA gene is absent from the gene encoding the *L. semifasciata* erabutoxin *c* precursor. These results highly suggest that intron insertions and/or deletions with snoRNA genes occurred during the evolution of snake venom neurotoxins. At the same time, the same searches were carried out for intron 2 of the cobrotoxin and cobrotoxin *b* genes, but their second intronic sequences do not contain sequences complementary to the invariant sequences of 5S rRNA. Previous studies on the genes encoding snake venom phospholipase A_2 from different species revealed that the protein-coding regions of phospholipase A_2 isozymes genes are unusually varied, and that the introns of venom gland phospholipase A_2 isozymes genes had evolved at similar rates and are highly conserved (*15-20*). Our data indicating that the genes encoding neurotoxin precursors are in agreement with this observation. These results probably suggest that intron regions are superior to exon regions for assessing the evolutionary

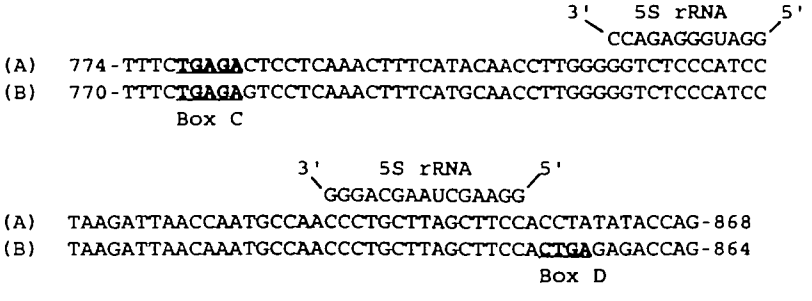

Fig. 8. **The sequence complementarity between 5S rRNA, and the intron 1 sequences of the cobrotoxin *b* and cobrotoxin genes.** The conserved sequences of 5S rRNA involved in the sequence complementarity are shown on the top. (A) The intron 1 sequence at positions 774-868 of the cobrotoxin *b* gene. (B) The intron 1 sequence at positions 770-864 of the cobrotoxin gene. The box C and D motifs are indicated by bold letters and underlining.

relationships of homologous proteins on the basis of their genomic structures.

The authors wish to express their sincere gratitude to Prof. C.C. Yang, Department of Life Sciences, National Tsing Hua University, for his encouragement during this study.

REFERENCES

1. Chiou, S.H., Hung, C.C., Huang, H.C., Chen, S.T., Wang, K.T., and Yang, C.C. (1995) Sequence comparison and computer modelling of cardiotoxins and cobrotoxin isolated from Taiwan cobra. *Biochem. Biophys. Res. Commun.* **206**, 22-32
2. Yu, C., Bhaskaran, R., Chuang, L.C., and Yang, C.C. (1993) Solution conformation of cobrotoxin: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry* **32**, 2123-2136
3. Endo, T. and Tamiya, N. (1991) Structure-function relationships of postsynaptic neurotoxins from snake venoms in *Snake Toxins* (Harvey, A.L., ed.) pp. 165-222, Pergamon Press, New York
4. Chang, L.S., Lin, J., Wu, P.F., Chang, C.C., and Hong, E. (1997) cDNA sequence analysis and expression of κ-bungarotoxin from Taiwan banded krait. *Biochem. Biophys. Res. Commun.* **230**, 192-195
5. Dufton, M.J. and Hider, R.C. (1991) The structure and pharmacology of elapid cytotoxins in *Snake Toxins* (Harvey, A.L., ed.) pp. 259-302, Pergamon Press, New York
6. Patel, H.V., Vyas, A.A., Vyas, K.A., Liu, Y.S., Chiang, C.M., Chi, L.M., and Wu, W.G. (1997) Heparin and heparin sulfate bind to snake cardiotoxin, sulfate oligosaccharides as a potential target for cardiotoxin action. *J. Biol. Chem.* **272**, 1482-1492
7. Sambrook, J., Fritsh, E.F., and Maniatis, J.T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
8. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
10. Chang, L.S., Lin, J., Kuo, K.W., Lin, S.R., and Chang, C.C. (1995) Characterization of epitopes in native and unfolded cobrotoxin: evidence showing an immunodominant C-terminal region related to the production of precipitating and non-precipitating antibodies against cobrotoxin. *J. Biochem.* **118**, 686-692
11. Fuse, N., Tsuchiya, T., Nonomura, Y., Menez, A., and Tamiya, T. (1990) Structure of snake short-chain neurotoxin, erabutoxin *c*, precursor gene. *Eur. J. Biochem.* **193**, 629-633
12. Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383
13. Csank, C., Taylor, F.M., and Martindale, D.W. (1990) Nuclear pre-mRNA introns: analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. *Nucleic Acids Res.* **18**, 5133-5141
14. Senapathy, P., Shapiro, M.B., and Harris, N.L. (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252-278
15. Nakashima, K.I., Ogawa, T., Oda, N., Mattori, M., Sakaki, Y., Kihara, H., and Ohno, M. (1993) Accelerated evolution of *Trimeresurus flavoviridis* venom gland phospholipase A_2 isozymes. *Proc. Natl. Acad. Sci. USA* **90**, 5964-5968
16. Nakashima, K.I., Nobuhisa, I., Deshimaru, M., Nakai, M., Ogawa, T., Shimohigashi, Y., Fukumaki, Y., Hattori, M., Sakaki, Y., Hattori, S., and Ohno, M. (1995) Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A_2 isozyme genes. *Proc. Natl. Acad. Sci. USA* **92**, 5605-5609
17. John, T.R., Smith, L.A., and Kaiser, I.I. (1994) Genomic sequences encoding the acidic and basic subunits of mojave toxin: unusually high sequence identity of non-coding regions. *Gene* **139**, 229-234
18. John, T.R., Smith, J.J., and Kaiser, I.I. (1996) A phospholipase A_2-like pseudogene retaining the highly conserved introns of mojave toxin and other snake venom group II PLA_2s, but having different exons. *DNA Cell Biol.* **15**, 661-668
19. Ogawa, T., Wakashima, K.I., Nobuhisa, I., Deshimaru, M., Shimohigashi, Y., Fukumaki, Y., Sakaki, Y., Hattori, S., and Ohno, M. (1996) Accelerated evolution of snake venom phospholipase A_2 isozymes for acquisition of diverse physiological functions. *Toxicon* **34**, 1229-1236
20. Kordis, D. and Gubensek, F. (1996) Ammodytoxin c gene helps to elucidate the irregular structure of crotalinae group II phospholipase A_2 genes. *Eur. J. Biochem.* **240**, 83-90
21. Nobuyoshi, M., Lin, X.W., Takimoto, Y., Deuel, T.F., and Wang, Z.Y. (1997) Transcription regulation of PDF A-chain gene by first intron elements. *Biochem. Biophys. Res. Commun.* **230**, 569-572
22. Sakamoto, N., Akasaka, K., Mitsunga-Nakatsubo, K., Takata, K., Nishitani, T., and Shimada, H. (1997) Two isoforms of orthodenticle-related proteins bind to the enhancer element of sea urchin arylsulfatase gene. *Dev. Biol.* **181**, 284-295
23. Nicoloso, M., Qu, L.H., Michot, B., and Bachellerie, J.P. (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2′-*O*-ribose methylation of rRNAs. *J. Mol. Biol.* **260**, 178-195
24. Qu, L.H., Nicoloso, M., Michot, B., Azum, M.C., Caizergues-Ferrer, M., Renalier, M.H., and Bachellerie, J.P. (1994) U21, a novel small nucleolar RNA with a 13 nt, complementarity to 28S rRNA, is encoded in an intron of ribosomal protein L5 gene in chicken and mammals. *Nucleic Acids Res.* **22**, 4073-4081
25. Qu, L.H., Henry, Y., Nicoloso, M., Michot, B., Azum, M.C., Renalier, N.H., Caizergues-Ferrer, M., and Bachellerie, J.P. (1995) U24, a novel intron-encoded small nucleolar RNA with two 12 nt long, phylogenetically conserved complementarities to 28 S rRNA. *Nucleic Acids Res.* **23**, 2669-2676
26. Renalieer, M.H., Nicoloso, M., Qu, L.H., and Bachellerie, J.P. (1996) SnoRNA U21 is also intron-encoded in *Drosophila melanogaster* but in a different host gene as compared to warm-blooded vertebrates. *FEBS Lett.* **379**, 212-216
27. Kenmochi, N., Higa, S.M., Yoshihama, M., and Tanaka, T. (1996) U14 snoRNAs are encoded in introns of human ribosomal protein S13 gene. *Biochem. Biophys. Res. Commun.* **228**, 371-374
28. Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNA. *Annu. Rev. Biochem.* **35**, 897-934
29. Riedel, N., Wise, J.A., Swerdlow, H., Mak, A., and Guthrie, C. (1986) Small nuclear RNAs from *Saccharomyces cerevisiae*: unexpected diversity in abundance, size and molecular complexity. *Proc. Natl. Acad. Sci. USA* **83**, 8097-8101
30. Yang, C.C. and Chang, L.S. (1990) The N-terminal amino group essential for the biological activity of notexin from *Notechis scutatus scutatus* venom. *Biochim. Biophys. Acta* **1040**, 35-42