Structure and organization of the cardiotoxin genes in *Naja naja sputatrix*

Ramkumar Lachumanan, Arunmozhiarasi Armugam, Chee-Hong Tan, Kandiah Jeyaseelan*

Department of Biochemistry, Faculty of Medicine, National University of Singapore, 10 Kent Ridge Crescent, 119260 Singapore

Received 19 May 1998; revised version received 6 July 1998

Abstract We report the genomic structure, organization and the presence of multiple isoforms of the gene encoding cardiotoxins (CTX) of *Naja naja sputatrix*. The cardiotoxin gene consists of six CTX isoforms, each (2.2 kb) having three exons and two introns. Two possible transcription initiation sites as well as consensus TATA boxes and transcription factor binding motifs, AP-2, NFIL-6/C/EBP, NF-κB and PuF have been identified in the 5′-region of the gene. The CTX gene isoforms show nucleotide variations at *specific segments* in exon 2 and exon 3, which correspond to the functional domains in the three-finger loop structure of the cardiotoxin molecule. The diverse functions of cardiotoxins together with our findings suggest that the cardiotoxin gene isoforms may have evolved under adaptive pressure through a positive Darwinian selection process.

Key words: Cardiotoxin; Gene structure; Multiple isoform; *Naja naja sputatrix*

1. Introduction

Snake venom cardiotoxins (CTX) are low molecular mass (6.5–7 kDa), highly basic ($pI > 10$) proteins cross-linked by four disulfide bridges [1,2]. These classes of toxins are found in the species of *Naja* (the cobras) and *Hemachatus* (the ringhals) [3]. Interestingly, cardiotoxins exhibit a wide variety of biological activities which include hemolysis, cytotoxicity, membrane depolarization in excitable cells, membrane fusion, selective killing of certain types of tumor cells, inhibition of protein kinase C activity and muscle contraction [4].

More than 52 cardiotoxins from various cobras of different origins have been reported so far and about eight of them possess isoforms of the cardiotoxin [5]. These cardiotoxins are highly conserved proteins and exhibit more than 90% amino acid sequence homology [6]. The very first report on cloning of a cardiotoxin cDNA was published by our laboratory [7] and since then about a dozen cardiotoxins and cardiotoxin-like cDNA sequences have been reported [8,9] and/or deposited in the GenBank database by other laboratories.

At the genomic level, only one nucleotide sequence of cardiotoxin from *Naja naja atra* had been reported recently [10]. Therefore, in pursuit of understanding of the structure and regulatory mechanisms of cardiotoxin gene(s) in cobras, we provide the first report of the presence of multiple copies of the cardiotoxin gene in the snake genome, their structure and the nature of the regulatory elements in the gene's promoter region. We present evidence that the spitting cobra genome has at least six CTX gene isoforms capable of encoding six types of functional CTX mRNAs.

2. Materials and methods

2.1. Materials

Restriction enzymes, *Taq* DNA polymerase, [γ-^{33}P]dATP and [α-^{32}P]dCTP were purchased from Amersham (UK). Advantage Genomic PCR kit, Universal GenomeWalker kit and AdvanTAge PCR cloning kit were obtained from Clontech (USA). Reverse transcriptase, T4 DNA ligase and T4 polynucleotide kinase were from Gibco BRL (USA), New England Biolabs (USA) and Amersham (UK) respectively. DNA sequencing reagents were from Perkin-Elmer (USA). Custom oligonucleotides were supplied by the National University Medical Institute (NUMI), National University of Singapore. All other chemicals were of molecular biology grade and were purchased either from Boehringer Mannheim (Germany), Sigma (USA), Merck (Germany) or BDH (UK).

2.2. Naja naja sputatrix

A Malayan spitting cobra (*Naja naja sputatrix*) was obtained from Singapore Zoological Gardens. Its identity was confirmed by its anatomical features.

2.3. Preparation of spitting cobra genomic DNA

High molecular weight Malayan spitting cobra genomic DNA was prepared according to the method of Blin and Stafford [11]. Freshly excised and frozen snake liver from a single snake was gently ground into fine powder using a mortar containing liquid nitrogen. Ten volumes of digestion buffer (Tris-HCl buffer, pH 8, 25 mM EDTA, 0.1 M NaCl and 0.5% SDS) were added together with proteinase K (100 μg/ml) and incubated at 50°C for 3 h, with frequent mixing. The sample was then gently extracted twice with TE buffer (10 mM Tris-HCl, pH 8 and 1 mM EDTA)-saturated phenol, once with phenol:chloroform (1:1) and finally with chloroform. Genomic DNA was precipitated from aqueous phase by addition of 0.1 vol of 3 M ammonium acetate and 2 vol of ethanol. The pellet was dissolved in 100 μl sterile TE buffer at 37°C overnight.

2.4. Genomic PCR

The AdvanTAge Genomic PCR kit (Clontech) was used to amplify cardiotoxin gene(s) from the highly intact genomic DNA of the snake liver. The kit employs two thermostable DNA polymerases, the first is *Tth* DNA polymerase which acts as primary polymerase, and the second DNA polymerase, Vent, provides 3′-5′ proofreading activity. Primers used were designed based on the highly conserved regions in the structural gene of *N. n. sputatrix* cardiotoxin cDNAs [12] as shown in Fig. 1A. The primer sets used were: forward (C1), 5′-CCTTGGTGGTGGTGACA-3′ and reverse (C2) 5′-AACATCAATACATCCCCTTTT-3′. The thermal profile involved a hot start at 94°C for 1 min followed by 30 cycles of 30 s at 94°C, 30 s at 59°C and 3 min at 68°C and a final extension at 68°C for 3 min after the final cycle. The PCR product was analyzed on agarose gel and the appropriate bands were excised and purified (Qiaquick gel extraction kit, Qiagen) followed by sub-cloning into pT-Adv vectors (Clontech AdvanTAge Cloning kit). About 55 clones carrying plasmids with the appropriate inserts were picked and sequenced on both strands using M13 and sequence specific sense and antisense primers.

2.5. Genome walking

To elucidate the 5′ and 3′ ends of the sequenced CTX gene, the Universal GenomeWalker kit (Clontech) was used according to the

*Corresponding author. Fax: (65) 7791453.
E-mail: bchjeya@nus.edu.sg

The genomic DNA sequences reported in this paper have been submitted to GenBank with accession numbers AF064096–AF064101.

PII: S0014-5793(98)00894-1

manufacturer's procedure. Briefly, the procedure involves the construction of adaptor-ligated libraries made by separate restriction digestion of highly intact genomic DNA with *Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Stu*I, followed by purification of the digested fragments and ligation to special adaptors provided in the kit. The adaptor-ligated genomic DNA fragments also known as GenomeWalker libraries were used as templates in 'DNA walking' experiments using genomic PCR kit (Clontech). Primers used for mapping the 5′ end of cardiotoxin gene(s) were the adaptor primer 1 (AP1), 5′-GTAATACGACTCACTATAGGGC-3′ provided in the kit and a 26-mer gene specific primer (C3), 5′-ATTGCCACCCGAGATCAAAGTTTGGT-3′. However, to map the 3′ end of the CTX gene(s), two sets of primers were used: AP1 and the gene-specific primer (C4), 5′-CATTATGCTTTCTGCTTCCTTCACAG-3′ for the primary PCR and the nested adaptor primer 2 (AP2), 5′-ACTATAGGGCACGCGTGGT-3′ provided in the kit and the nested gene specific primer (C5), 5′-TCCTGTCAAAAGGGGATGTATTGATG-3′ for the secondary PCR.

The amplification technique used was 'touchdown' PCR [13,14], which involved a single annealing/extension temperature that is several degrees higher than the T_m of the primers during the initial PCR cycles. This temperature is later reduced to the primers' T_m for the remaining PCR cycles. The cycling parameters used for both primary PCR were: 10 cycles of 24 s at 94°C and 3 min at 65°C ('touchdown') and 30 cycles of 25 s at 94°C and 3 min at 60°C followed by 10 min at 68°C after the final cycle. The cycling parameters for the secondary PCR were: 10 cycles of 25 s at 94°C and 3 min at 68°C and 30 cycles of 25 s at 94°C and 3 min at 65°C followed by a final extension of 68°C for 10 min after the last cycle. The PCR products were analyzed on agarose gel and the appropriate bands excised and purified (Qiaquick gel extraction kit, Qiagen) followed by sub-cloning into pT-Adv vectors (Clontech AdvanTAge Cloning kit). The inserts were later sequenced on both strands with M13 and sequence specific sense and antisense primers.

2.6. Sequence analysis

All double-stranded sequencing was performed with the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer) using an automated fluorescent DNA sequencer (Applied Biosystems, Model 373). Isolated plasmid clones were completely sequenced on both strands using vector specific (M13 forward and reverse primers) and gene specific primers where appropriate.

Nucleotide sequence homology searches of non-redundant databases of GenBank (National Center for Biotechnology Information) were performed using the BLAST program. DNA sequence alignments were carried out using the DNASIS software package from Hitachi Software Engineering. Nucleotide sequences reported in this paper have been assigned GenBank accession numbers AF064096–AF064101.

2.7. Primer extension reaction

Primer extension analysis was performed to identify the transcription initiation site(s) of the CTX gene(s). Total RNA was isolated from the *N. n. sputatrix* venom glands following the guanidinium isothiocyanate method [15]. 10 pmol of primer C3 (see above) was 5′ end 33P-labeled by T4 polynucleotide kinase using [γ-^{33}P]ATP with a specific activity >5000 Ci/mmol. The ^{33}P-labeled primers were purified on a NAP 5 column (Pharmacia) and annealed to RNA (10 μg) in hybridization buffer (0.15 M KCl, 0.01 M Tris-HCl, pH 8.3 and 1 mM EDTA) at 65°C and brought to room temperature slowly. The hybrid was precipitated in ethanol and dissolved in 9 μl of reverse transcription mix (10 mM Tris-HCl, pH 8.3, 10 mM dithiothreitol, 6 mM $MgCl_2$ and 1 mM of each of the four dNTPs). Primer extension reaction was carried out at 42°C for 1 h in the presence of 20 units of MMLV reverse transcriptase. The reverse transcribed products were extracted once with phenol-chloroform-isoamyl alcohol (25:24:1) saturated with TE buffer, then precipitated with ethanol and, re-suspended in 4 μl of sequencing gel loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.02% xylene cyanol FF). This was electrophoresed on a 6% denaturing polyacrylamide DNA sequencing gel. A set of dideoxynucleotide DNA sequencing reaction using ^{33}P-labeled primer C3 and 30 fmol of the pT-Adv template harboring the 5′-flanking region of the CTX gene were generated using the AmpliCycle sequencing kit (Perkin-Elmer) and were used as size markers.

2.8. Genomic Southern blot analysis

Genomic DNA, isolated as mentioned above, was digested with restriction enzymes and fragments resolved by electrophoresis in 0.8% agarose gel and blotted onto nitrocellulose membrane (Schleicher and Schuell). ^{32}P-labeled probes were prepared by random hexamer priming (Megaprime DNA labeling kit, Amersham) on plasmids harboring the CTX cDNA inserts. Blots were hybridized overnight at 65°C in 10 mM EDTA, pH 8, 0.42 M Na_2HPO_4, pH 7.2, 0.08 M NaH_2PO_4, pH 7.2, 7% SDS and 100 μg/ml sheared salmon sperm DNA with at least 10^6 cpm/ml of ^{32}P-labeled probes. Blots were washed for 1 h at 42°C and then for 20 min at 65°C, each with 2×SSC containing 0.1% SDS.

3. Results

3.1. Gene structure and organization

PCR amplification of *N. n. sputatrix* genomic DNA was carried out using C1 and C2 as sense and antisense primers respectively. These primers were designed based on the highly conserved region of the CTX cDNAs [12] (Fig. 1A).

Analysis of the genomic PCR product on 1% agarose gel electrophoresis revealed a single band of about 2 kb (Fig. 1C i). PCR amplification of CTX cDNA clones using the above primers gave a band of 167 bp only (data not shown). These results suggested the presence of intron(s) of approximately 1.8 kb in the CTX gene region flanked by the two primers.

The nucleotide sequence of the 5′ end of the CTX gene was elucidated by sub-cloning and sequencing the 0.8 kb genome walking PCR fragment of the *Dra*I library (lane 2, Fig. 1C ii) generated using AP1 and C3 as sense and antisense primers respectively (Fig. 1B). Likewise, the 3′ end of the CTX gene was elucidated by sub-cloning and sequencing the 0.9 kb genome walking PCR fragment of the *Dra*I library (lane 2, Fig. 1C iii) generated using the combination of C4 (sense)-AP1 (antisense) and the nested C5-AP2 primers in primary and secondary PCR reactions respectively (Fig. 1B). Sequencing and analysis of the PCR products from the above experiments on both strands using appropriate primers (see Section 2) gave the complete gene sequence of spitting cobra CTX.

The CTX gene consists of two introns of approximately 1.2 kb and 0.5 kb respectively with both showing classical gt/ag sequences [16] flanking the intron splice junctions (Fig. 2). The first intron splice site occurs within the CTX leader sequence region between nucleotides 634 and 635 (Fig. 2). The second intron splice site occurs between nucleotides 1983 and 1984 within the coding region of the sequence (Fig. 2). This splice junction splits the codon for valine (GTT) at amino acid residue 32 in the deduced protein sequence between the first and second bases, indicating a phase 1 intron. Thus, the CTX gene is composed of three exons. The first exon encodes the 5′-UTR and most of the leader sequence. The second exon encompasses 102 bases and encodes bits of the leader sequence and half of the mature toxin coding region and the third exon encodes the remainder of the coding region and the 3′-UTR region.

The nucleotide sequence comparison between CTX3 from *N. n. sputatrix* and CTX4 from *N. n. atra* shows high sequence homology in all segments of the gene with exception of exon 2 with only 87% sequence homology. The 3′-UTR of the gene contains a polyadenylation signal at position 2821–2826, and 194 nucleotides downstream of the translation stop codon.

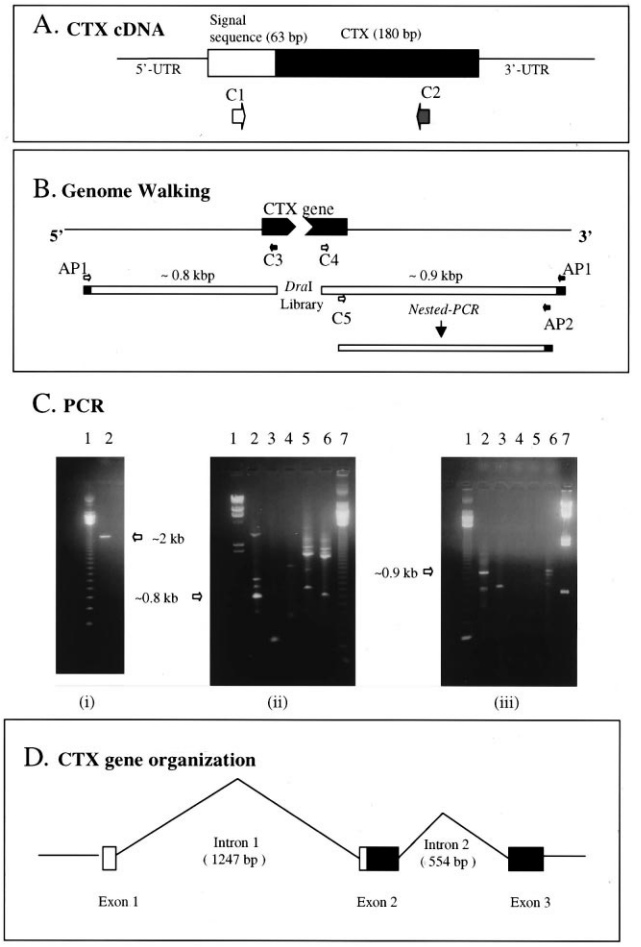

Fig. 1. Strategies for genomic and genome walking PCR and the organization of the CTX gene. A: Primers (C1 and C2) for genomic or long-distance PCR were designed based on the conserved regions of the CTX cDNAs reported earlier and their locations relative to the CTX cDNA are indicated. B: Schematic illustration on the genome walking strategies to map the 5′ and 3′ ends of the CTX gene. Gene specific primer C3 and adaptor-primer I (AP1) were used to amplify adaptor-ligated libraries to map the 5′ end of the CTX gene which includes the 5′ end of the signal peptide coding region, 5′-UTR and the 5′-flanking (promoter) region of the CTX gene. The 3′ end of the gene was mapped by two consecutive PCR reactions using the following primer pairs consecutively: C4/AP1 and C5/AP2. Only the fragments of *Dra*I library amplification, which were cloned and sequenced, are shown. C: Amplification products generated from the genomic PCR and genome walking PCR to elucidate the complete nucleotide sequence of the CTX gene. i: Amplification of partial CTX gene using genomic PCR. Lane 1: 123 bp ladder of DNA size markers; lane 2: PCR product using primers C1 and C2 (A). ii: Amplification of the 5′ end of the CTX gene, comprising the region encoding part of the signal peptide, the 5′-UTR and the 5′-flanking (promoter) region. Lanes 1 and 7: λ phage *Hin*dIII digest and 123 bp ladder of DNA size markers, respectively. Lanes 2–6: Genome walking PCR products from *Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Stu*I adaptor-ligated DNA libraries, respectively, using primers C3 and AP1 (B). The intense band from lane 2 (*Dra*I library) was purified, sub-cloned and sequenced. iii: Amplification of the 3′ end of the CTX gene, comprising the region encoding the 3′ end of the mature cardiotoxin, the 3′-UTR and the 3′-flanking region. Lanes 1 and 7: 123 bp ladder and phage λ *Hin*dIII digest DNA size markers, respectively. Lanes 2–6: Genome walking PCR products from *Dra*I, *Eco*RV, *Pvu*II, *Sca*I and *Stu*I adaptor-ligated DNA libraries, respectively, using primers AP2 and C5 (B). The intense band from lane 2 (*Dra*I library) was purified, sub-cloned and sequenced. D: Organization of the CTX gene. The exons are boxed and the introns are indicated as thin lines. Gene regions encoding the signal peptide are represented by open boxes while those encoding mature cardiotoxin are represented by shaded boxes.

→

3.2. Sequence analysis of CTX gene

The sequence analysis of the CTX gene revealed the presence of not one but six CTX gene isoforms in the genome of a single spitting cobra. The frequencies of appearance of the six isoforms in the 55 clones sequenced are almost similar, with 0.20, 0.16, 0.15, 0.18, 0.15 and 0.16 for CTX3, CTX2A, CTX4B, CTX5A, CTX6 and CTX7, respectively. Nucleotide and deduced amino acid sequence comparison among the six CTX gene isoforms (Fig. 2) shows that the introns are highly conserved while the second and, to a much lesser extent, the third exons, which encode the mature cardiotoxins, show nucleotide variations. These nucleotide variations are limited to specific segments on both exons, and the highly variable segment on exon 2 could be called the 'major variable segment' of the CTX gene. The nucleotide variations thus lead to variations in the amino acids being encoded, resulting in the identification of six isoforms of the CTX gene, encoding six putative cardiotoxins in a single spitting cobra.

3.3. Determination of the transcription initiation sites of the CTX gene

A sequence of about 500 bp of the 5′-flanking region was analyzed and used to define the transcriptional initiation site(s) of the CTX gene (which was characterized by the longest major extension product). The CTX transcription initiation site(s) were determined by primer extension using primer C3 which is complementary to the 5′ end of the toxin mRNA. Two major transcriptional products were observed from the primer extension reaction (Fig. 3). An alignment of these products with the gene sequence showed that the major primer extension product corresponds to the start sequence GGTA at position 516 (transcription initiation site 1, TIS1) (Fig. 2). The minor product (in terms of band intensity) is present 83 bp upstream of TIS1 at position 433 (TIS2) (Fig. 2). These sites did not correspond to the transcription start site previously reported in *N. n. atra* CTX4 gene [10]. Two putative TATA box motifs are located 31 and 69 bp immediately upstream of TIS1, while the nearest putative TATA box motif for TIS2 is located 68 bp upstream of it. No putative CAAT elements were identified in the sense strand upstream of both transcription initiation sites.

3.4. Genomic Southern blot analysis

Southern blot analysis of the *N. n. sputatrix* genome, using the cDNA encoding the CTX as the probe, detected 4–6 bands (data not shown). The multiple bands that hybridized to CTX cDNA under stringent conditions were detected in all the digest of genomic DNA, suggesting that at least six isoforms of the CTX gene exist in the spitting cobra genome.

3.5. Putative regulatory sequence elements of the gene promoter

Searches for regulatory sequences that have been identified as binding sites for the conserved DNA binding proteins in other genes, using TRANSFAC and TFD software packages, show that the CTX3 gene promoter contains sequences matching the following consensus sequences: AP-2 [17],

Fig. 2. Complete nucleotide sequence of the CTX3 gene including the introns and the 5′- and 3′-flanking regions. The promoter region is designated from 1 to 432, while the transcribed sequence is numbered from 433 onwards. Coding sequences of exons are represented in capitals, and the deduced amino acid sequence is given below them in single-letter code. The signal peptide of the toxin precursor is in uppercase while the mature toxin sequence is represented in boldfaced uppercase letters. Non-coding sequences of exons, intron sequences, and flanking sequences in the 5′ and 3′ regions of the CTX gene are given in lowercase letters. The two transcription initiation sites (TIS1 and TIS2), assigned by primer extension analysis, are indicated by arrows and the corresponding nucleotides are boldfaced and double underlined. The nucleotides constituting the TGA stop codon are indicated by asterisks. The consensus sequence for polyadenylation, AATAAA, at the 3′-flanking region of the gene is boldfaced. The 5′-flanking region of the gene contains recognition elements for nuclear factors, as revealed by database analysis. Consensus sites are underlined with the name of the transcription factor binding site indicated on top of the sequence. The orientation of the elements is shown by the signs, i.e. (+) for those elements identified on the plus (sense) strand, and (−) for those elements identified on the minus (antisense) strand. TATA box motifs are also boldfaced. Alignment of the mature toxin coding regions located in the second and third exons of the CTX2A, CTX4B, CTX5A, CTX6 and CTX7 with CTX3 gene sequence is indicated in the complete gene sequence. The amino acid sequences of the putative proteins encoded by the six CTX gene isoforms are represented by uppercase single-letter code. CTX3 was taken as the consensus sequence and any variations from this sequence (both nucleotide and amino acid) are indicated for the five other CTX gene isoforms. Nucleotide residues that match the CTX3 sequence are indicated by a hyphen '-', while amino acid residues that are different from CTX3 are indicated in boldfaced uppercase single-letter codes and boxed. The complete nucleotide sequence of CTX3 from *N. n. sputatrix* has been deposited in GenBank (NCBI) under the accession number AF064096. The partial CTX nucleotide sequences (spanning exon 2, intron 2 and exon 3) of CTX2A, CTX4B, CTX5A, CTX6 and CTX7, deposited in the GenBank database, have been assigned accession numbers AF064097–AF064101.

NFIL-6/C/EBP [18], and NF-κB [19], (Fig. 2). An additional interesting feature of the CTX gene promoter region is that it contains four putative AP-2 sites, two CAAT elements and one PuF [20] site in reverse orientation. It is currently unknown whether these elements are functionally active.

4. Discussion

4.1. Structure of the CTX gene

Although the presence of isoforms of the cardiotoxin molecule in a single venom source and multiple types of cardiotoxin cDNA in a venom gland had been reported [12,21], this is the first report that traces the source of all these variations to the genome of *N. n. sputatrix*.

Based on the DNA sequence information obtained through a combination of genomic PCR and genome walking experiments, the CTX coding sequence was assigned three exons with two intervening introns (Fig. 2). The entire sequenced CTX gene encompasses a 2887 bp region, including 579, 2044 and 264 bases in the 5′-non-coding, coding (with two

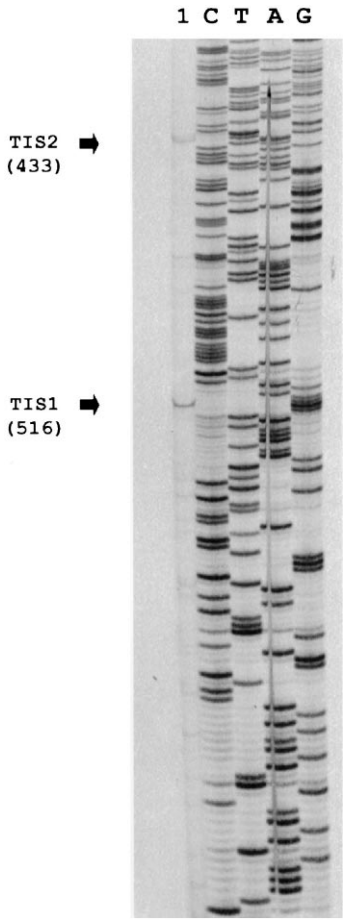

Fig. 3. Determination of the transcription initiation site(s) of the CTX gene by primer extension analysis. A synthetic primer (C3) was end-labeled and annealed to the total RNA from a spitting cobra venom gland. The annealed template/primer was extended with reverse transcriptase. The two major extension products from spitting cobra venom gland RNA are indicated by arrows (lane 1). The adjacent lanes are a sequence ladder of the corresponding genomic sequence of the non-coding strand obtained using the same primer (lanes C, T, A, G).

introns of 1247 and 554 bases each) and the 3′-non-coding regions, respectively (Fig. 2).

Exon 1 contains the TIS, 5′-UTR and most of the signal peptide coding gene sequence. The rest of the signal peptide and half of the mature cardiotoxin coding gene sequences are encoded by exon 2. The other half of the cardiotoxin coding region is found in exon 3 (as illustrated in Fig. 1D).

Determination of the transcription initiation site(s) by primer extension revealed the presence of two possible sites, TIS1 and TIS2. These sites differ from the single TIS reported recently in the *N. n. atra* CTX4 gene [10]. Two TATA box consensus motifs were indeed found 31 and 69 bp immediately upstream of TIS1 while the nearest putative TATA box motif was found 68 bp upstream of TIS2 in the CTX gene promoter region. Interestingly, these TATA box motifs lack the canonical upstream CCAAT elements. Three other potential TATA box consensus motifs were found in the promoter region, though these sites do not seem to initiate transcription of the CTX gene. The presence of multiple transcription initiation sites in animal toxin genes was first reported by our laboratory [22] in the potassium channel toxin HmK from sea anemones. Additionally we have demonstrated the presence of multiple transcription initiation sites in snake toxin genes in particular and in animal toxin genes in general.

On the question of which of the two transcription initiation sites, TIS1 or TIS2, is involved in gene transcription in vivo, it is possible that both are involved in initiating transcription. This is in accordance with an earlier report stating that a longer 5′-UTR increases the efficiency of translation over that of a shorter one and may be necessary for the efficient initiation of translation in some genes [23]. The primer extension signal for TIS1 is more intense as compared to the TIS2 band intensity, which leads to the speculation that the former is the transcription initiation site of choice (possibly AP-2 enhanced transcription initiation), though not the only one. The classical polyadenylation signal sequence [24] in the 3′-UTR was also detected in the CTX gene. However, no consensus mRNA degradation signal [25] immediately downstream of the polyadenylation signal sequence was present, which might explain the long half-life of CTX mRNAs observed during in vivo CTX gene expression studies (unpublished result).

4.2. Analysis of the CTX gene

The analysis of the genomic organization of the *N. n. sputatrix* CTX3 gene and its comparison with that of the CTX4 gene from *N. n. atra* revealed striking similarities. The exception, however, is the exon 2 homology comparison between both genes, which is quite low at 87% compared to the rest of the gene segments. The variations in exon 2 between the two genes are confined to the region of the CTX gene encoding the seven amino acid residues FMVATPK of the mature cardiotoxin molecule. Variations at a similar location on exon 2 were also observed among the five other *N. n. sputatrix* CTX gene sequences. This region does in fact correspond to the 'hyper-variable segment' reported earlier in CTX cDNAs [12].

Cloning and sequence analysis of the CTX genes from a single snake provide confirmatory evidence on the presence of *multiple isoforms* of the CTX gene. The similar frequency of appearance of all six isoforms in a sample of 55 clones sequenced strongly suggests that all CTX gene isoforms in the genome of *N. n. sputatrix* have been elucidated with the least possibility of missing out any rare isoforms. This is further corroborated from the Southern blot analysis of the genomic DNA from a single snake liver using the CTX cDNA probe, which gave multiple bands. Therefore, this is the first report on the presence of six isoforms of the CTX gene in the genome of the snake.

Comparison of nucleotide sequences amongst the CTX gene isoforms reveals the presence of a *major variable segment* on exon 2 (nucleotides 1963–1981 in Fig. 2) which encodes the amino acid residues at positions 25–31 of the mature cardiotoxin protein sequence. These amino acid residues in the 'hyper-variable region' are actually located at the tip of loop 2 in the cardiotoxin 3D structure. Therefore it is highly possible that the CTX gene isoforms evolved through gene duplications followed by mutations concentrated specifically at the *major variable segment*. In return, the CTX gene isoforms encode cardiotoxins that are highly homologous except for the residues located at the hyper-variable region of the cardiotoxin molecules. This scheme of evolution would have the added advantage of encoding a toxin molecule that maintains

its overall 3D structure, which is paramount for maintaining its physiological function, while providing room for variation on the 'active site' of the cardiotoxin molecule to effect its diverse pharmacological roles.

A similar mechanism of evolution was reported earlier in Crotalidae snakes [26]. Ohno and coworkers had suggested that the molecular evolution of snake toxins involves a mechanism of *accelerated evolution* to effect the functional diversity of snake toxins [27]. The accelerated evolution model fits the evolutionary mechanism of the CTX gene isoforms, where the introns and some stretches of exons are highly conserved with variations limited to *specific stretches* in the exons. In return, these isoforms encode cardiotoxins which have been shown to possess variable physiological activities [12] and could be important in manifestation of diverse pharmacological roles. The other region of variability seen on the CTX gene is located on a very small stretch on exon 3 (nucleotides 2577–2583). This variation divides the CTX gene isoforms into two groups; CTX3, CTX2A and CTX4B have an identical exon 3 but differ slightly from CTX5A, CTX6 and CTX7 which share an identical exon 3 amongst them (Fig. 2).

In conclusion, we report the complete structure, organization and the presence of multiple isoforms of the CTX gene in the genome of *N. n. sputatrix*. Analysis of the gene promoter region shows the presence of multiple transcription initiation sites. Further analysis of the gene sequence revealed the presence of multiple isoforms of the CTX gene with nucleotide variations almost exclusively in the *major variable segment* on exon 2 and a minor variable segment on exon 3. The variations in gene sequence would result in variations in transcribed mRNAs, which translate into putative multiple isoforms of cardiotoxins in the venom. It may be assumed that each cardiotoxin isoform encoded by the CTX genes would manifest a particular pharmacological or physiological activity. Indeed, distinct physiological functions have been found in *N. n. sputatrix* cardiotoxins in terms of cardiotoxicity and hemolytic activity [12,28] and protein kinase C inhibition [29]. It can be envisaged from the nucleotide substitutions having occurred only in the *selective segments* of the protein coding regions, that the CTX gene isoforms have probably evolved under adaptive pressure via positive Darwinian selection. The acquisition of these multiple cardiotoxin isoforms in the venom would be expected to confer a strong selective advantage on the snake for immobilizing various types of preys and for defense against its predators.

Acknowledgements: This work was supported by Research Grant RP 960324 from the National University of Singapore. R. Lachumanan is the recipient of a NUS Research Scholarship.

References

[1] Kumar, T.K.S., Lee, C.S. and Yu, C. (1996) in: Natural Toxins II (Tu, A.T. and Singh, B.R., Eds.), pp. 115–129, Plenum Press, New York.
[2] Yu, C., Bhaskaran, R. and Yang, C.C. (1994) J. Toxin Toxicol. Rev. 13, 291–315.
[3] Dufton, M.J. and Hider, R.C. (1991) in: Snake Toxins (Harvey, A.L., Ed.), pp. 259–302, Pergamon Press, New York.
[4] Kumar, T.K., Yang, P.W., Lin, S.H., Wu, C.Y., Lei, B., Lo, S.J., Tu, S.C. and Yu, C. (1996) Biochem. Biophys. Res. Commun. 219, 450–456.
[5] Kumar, T.K.S., Jayaraman, G., Lee, C.S., Arunkumar, A.I., Sivaraman, T., Samuel, D. and Yu, C. (1997) J. Biomol. Struct. Dynamics 15, 431–463.
[6] Brackenridge, R. and Dufton, M.J. (1987) J. Mol. Evol. 26, 274–283.
[7] Yeo, M.S.L., Jeyaseelan, K., Chung, M.C.M., Gopalakrishnakone, P., Tan, C.H. and Wong, H.A. (1993) Toxicon 31, 53–60.
[8] Chang, L.S., Wu, P.F. and Lin, J. (1996) Biochem. Biophys. Res. Commun. 219, 116–121.
[9] Chang, L.S., Lin, J. and Wu, P.F. (1996) Biochim. Biophys. Acta 1295, 1–4.
[10] Chang, L.S., Lin, J., Chou, Y.C. and Hong, E. (1997) Biochem. Biophys. Res. Commun. 239, 756–762.
[11] Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303–2306.
[12] Jeyaseelan, K., Armugam, A., Lachumanan, R., Tan, C.H. and Tan, N.H. (1997) Biochim. Biophys. Acta 1380, 209–222.
[13] Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) Nucleic Acids Res. 19, 4008.
[14] Roux, K.H. (1995) PCR Methods Appl. 4, 5185–5194.
[15] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.A. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[16] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[17] Imagawa, M., Chiu, R. and Karin, M. (1987) Cell 51, 251–260.
[18] Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimoto, T. (1990) EMBO J. 9, 1897–1906.
[19] Baeuerle, P.A. (1991) Biochim. Biophys. Acta 1072, 63–80.
[20] Postel, E.H., Berberich, S.J., Flint, S.J. and Ferrone, C.A. (1993) Science 261, 478.
[21] Tan, N.H. (1982) Arch. Biochem. Biophys. 218, 51–58.
[22] Gendeh, G.S., Chung, M.C. and Jeyaseelan, K. (1997) FEBS Lett. 418, 183–188.
[23] Kozak, M. (1991) Gene Exp. 1, 117–125.
[24] Proudfoot, N. (1991) Cell 64, 671–674.
[25] Shaw, G. and Kamen, R. (1986) Cell 46, 659–667.
[26] Nakashima, K., Nobuhisha, I., Deshimaru, M., Nakai, M., Ogawa, T., Shimohigashi, Y., Fukumaki, Y., Hattori, M., Sasaki, Y., Hattori, S. and Ohno, M. (1995) Proc. Natl. Acad. Sci. USA 92, 5605–5609.
[27] Ohno, M., Menez, R., Ogawa, T., Danse, J.M., Shimohigashi, Y., Fromen, C., Ducancel, F., Zinn-Juxtin, S., Le du, M.H., Boulain, J.C., Tamiya, T. and Menez, A. (1998) Prog. Nucleic Acid Res. Mol. Biol. 59, 307–364.
[28] Hider, R.C. and Khader, F. (1982) Toxicon 20, 175–179.
[29] Chiou, S.H., Raynor, R.L., Zheng, B., Chambers, T.C. and Kuo, J.F. (1993) Biochemistry 32, 2062–2067.