Promoter Structure and Intron−Exon Organization of a Scorpion α-Toxin Gene[†,‡]

M.-L. Delabre,[⊥] P. Pasero,[§] M. Marilley,[§] and P. E. Bougis*[,⊥]

Laboratoire de Biochimie, Unité de Recherche Associée 1455 du Centre National de la Recherche Scientifique, Université d'Aix-Marseille II, Faculté de Médecine Secteur Nord, 15 Bd. Pierre Dramard, 13916 Marseille Cedex 20, France, and Laboratoire de Génétique, Groupement de Recherche 976 du Centre National de la Recherche Scientifique, Université d'Aix-Marseille II, Faculté de Médecine, 27 Bd. Jean Moulin, 13385 Marseille Cedex 05, France

Received December 14, 1994[⊗]

ABSTRACT: The *Androctonus australis* scorpion venom contains α-toxins for which the complementary DNAs have been cloned [Bougis et al. (1989) *J. Biol. Chem. 264*, 19259−19265], targeting with high affinity the voltage-sensitive sodium channel. From a genomic library made of this species of scorpion, we have cloned and characterized the gene encoding the toxin AaH I′. The gene transcriptional unit is 793 base pairs long, and the gene has a single intron of 425 base pairs located near the end of the signal peptide of the toxin precursor. The transcription initiation site was determined by primer extension and corresponded to the nucleotide sequence AACAA. Upstream, a promoter region has been identified with positive acting sequence elements at consensus positions, such as a CCAAT box and a TATA box. In addition, putative elements for binding the transcriptional factors MAT-α2, Pit-1, and IEF1 are also present. Analysis of DNA curvature by computer modeling revealed a strong bending centered around the transcription initiation site of the gene. The bending angle (61°) estimated experimentally using polyacrylamide gel electrophoresis correlates well with the value predicted by computer modeling (66°). Other minor deflections of the helix axis cooperate for an overall curvature of nearly 90°, which is significantly stronger than similar structures already reported in eukaryotic cells. It is worth noting that the grooves relative to the CCAAT box and the TATA box lie along the inside of the DNA curve. This observation is in agreement with the previously reported correlation between DNA bending and promoter function.

The venoms of buthid scorpions contain long neurotoxins (60−70 amino acid residues, 4 disulfide bridges) that target voltage-sensitive sodium channels, thus impairing the initial rapid depolarization phase of the action potential in nerve and muscle (Miranda & Lissitsky, 1961; Rochat et al., 1979; Catterall, 1986). Yet, studies have focused on their purification and the characterization of their structural, immunological, and pharmacological properties [for reviews, see Granier et al. (1989) and Martin-Eauclaire and Couraud (1994)]. Scorpion toxins active on mammals are classed into α- or β-type, a distinction based on electrophysiological, ligand-binding, and ion-flux studies. Some other toxins are lethal to either insects or crustaceans (Zlotkin et al., 1993). We have recently described a genetic approach. From a library constructed from *Androctonus australis* venom gland mRNA, cDNAs have been cloned and sequenced, encoding the precursors of the α-type toxins AaH I,[1] AaH I′, AaH II, and AaH III and the toxins active on insects AaH IT1 and AaH IT2 (Bougis et al., 1989). In addition to the cleavage of the signal peptide, the maturation of these precursors depends on diverse additional C-terminal sequences that are removed to give the mature toxins: Gly-Arg in the case of the α-amidated toxin AaH II, and Arg in the case of the other toxins active on mammals. The toxins AaH IT1 and AaH IT2 do not possess additional C-terminal sequences. Similarly, other cDNAs encoding toxins from *Tityus serrulatus* (Martin-Eauclaire et al., 1992, 1994), *Centruroides noxius* (Becerril et al., 1993; Vasquez et al., 1993), *Buthotus judaicus* (Zilberberg et al., 1991), and *Leirus quinquestriatus hebraeus* (Zilberberg et al., 1992) have been cloned and also show such a C-terminal structural diversity. The origin of this C-terminal structural diversity is unknown. At the genomic level, Southern blot analysis of the genome of *A. australis*, using the cDNA encoding the toxin AaH II as the probe, detected a single *Eco*RI fragment of 2800 nucleotides, suggesting that the corresponding toxin gene is present as a single copy (Bougis et al., 1989).

Here, on screening a genomic library of *A. australis*, we cloned and sequenced the gene encoding the toxin AaH I′. Furthermore, we have analyzed the sequence-directed curvature of the AaH I′ gene by computer modeling and polyacrylamide gel electrophoresis. Bent DNA has often been associated with promoter function in prokaryotes [for reviews, see Hagerman (1990) and Perez-Martin et al.

[†] This work was supported in part by the Délégation à la Recherche et aux Etudes Techniques, the National Research Council/United States Army Medical Research Institute of Infectious Diseases Research Associateship from the National Academy of Sciences (concerning the construction of the genomic library) and the Association pour la Recherche contre le Cancer.

[‡] The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the Accession No. X76135.

* To whom correspondence should be addressed; Telephone (33) 91-96-20-70; Fax: (33) 91-69-88-39.

[⊥] Laboratoire de Biochimie.

§Laboratoire de Génétique.

[⊗] Abstract published in *Advance ACS Abstracts,* April 15, 1995.

[1] Abbreviations: AaH, *Androctonus australis* Hector; bp, base pair; EDTA, ethylenediamine tetraacetic acid disodium salt; Hepes, 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid; kb, kilo base; N = A + C + G + T; PBS, phosphate buffer saline; SDS, sodium dodecyl sulfate; SSC, saline sodium citrate buffer; TBE, tris:borate EDTA buffer; Tris, tris(hydroxymethyl)aminomethane; W = A + T; Y = C + T.

0006-2960/95/0434-6729$09.00/0

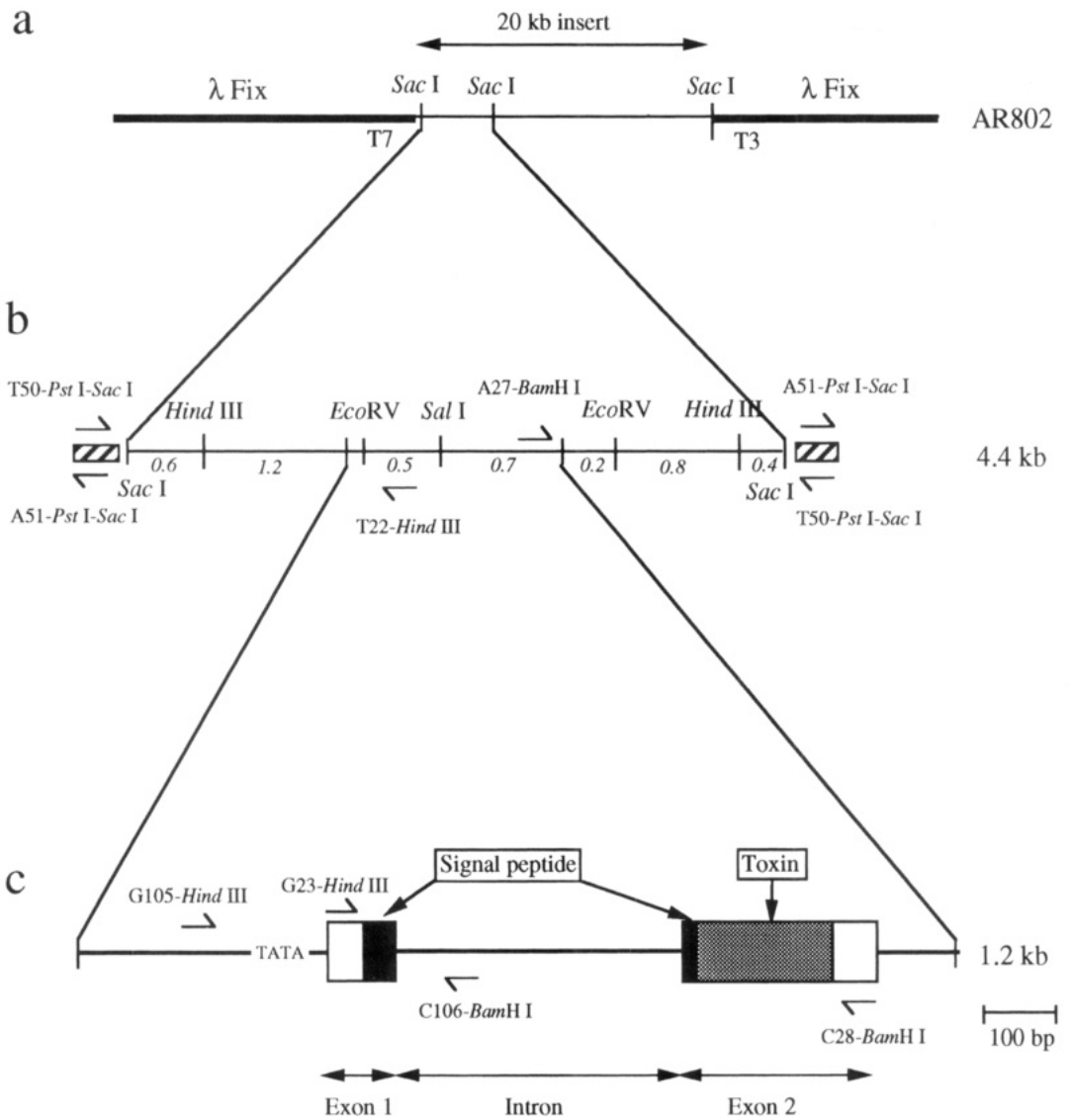

FIGURE 1: Strategy of analysis, partial restriction map, and organization of the AaH I′ toxin gene. The location of primers used to amplify by PCR parts of the gene are shown (see Materials and Methods). (a) *Sac*I restriction map of the 20 kb insert of the positive clone λFix-AR802; (b) restriction analysis of the 4.4 kb *Sac*I fragment; (c) organization of the gene. The exons are indicated by filled blocks and the 5′- and 3′-untranslated regions are indicated by open blocks. The line connecting the two filled blocks represents the single intron.

(1994)], but a stable deflection of the double-helix axis has rarely been reported in eukaryotic promoters. The promoter region of the AaH I′ gene displays a strong bending approximately 30 bp downstream to the transcription initiation site consistent with the idea that curved DNA structures are associated with efficient transcription, both in eukaryotic and prokaryotic cells.

MATERIALS AND METHODS

Standard recombinant DNA techniques were used as described in Sambrook et al. (1989). Scorpions of the species *A. australis* were collected in the area of Beni-Khedache, South Tunisia, and kept alive at the laboratory.

Preparation of the Genomic Scorpion DNA and Construction of the Genomic Library. High molecular mass scorpion genomic DNA was prepared as follows. Muscles of two scorpions were obtained from the claws, the dorsal part of the abdomen, and the tail. The tissue was gently homogenized by five cycles on a potter B at 800 rpm in PBS supplemented with 2 mM EDTA, pH 7.4, collected by centrifugation at 3000*g* for 5 min and washed twice with the same buffer. The centrifugation pellet (3 mL) was dissolved in 20 mL of lysis buffer (10 mM Tris-HCl, 400 mM NaCl, 2 mM EDTA, pH 8.0, 0.5% SDS) and incubated for 1 h at 37 °C in the presence of 0.2 mg/mL proteinase K (Sigma) and then incubated for an additional hour at 37 °C in the presence of 0.02 mg/mL heat-treated RNAse (Sigma). The sample was gently extracted (1:1 by volume) with phenol, phenol−chloroform−isoamyl alcohol (25:24:1), and finally chloroform. DNA was then precipitated by adding 0.5 vol of 7.5 M ammonium acetate and 2 vol of ethanol. The DNA was migrated on a 0.3% agarose gel in 1× TBE . The fragment size was greater than 50 kb. A λ-based vector (λFix) was used to construct a genomic library by the partial fill-in technique (Stratagene). The *Escherichia coli* strain used to amplify and to screen the library was P2392: [LE392 (P2 lysogen)], LE392: [e^{14-}*(mcrA), hsdR514, supE44, supF58, lacY1, or* δ(lac1ZY)6, galK2, galT22, metB1, trpR55]. The library was amplified up to 1.5×10^8 phage clones/mL.

DNA Hybridization. Probes were ^{32}P-labeled by nick translation (Boehringer Mannheim). Hybridization was done

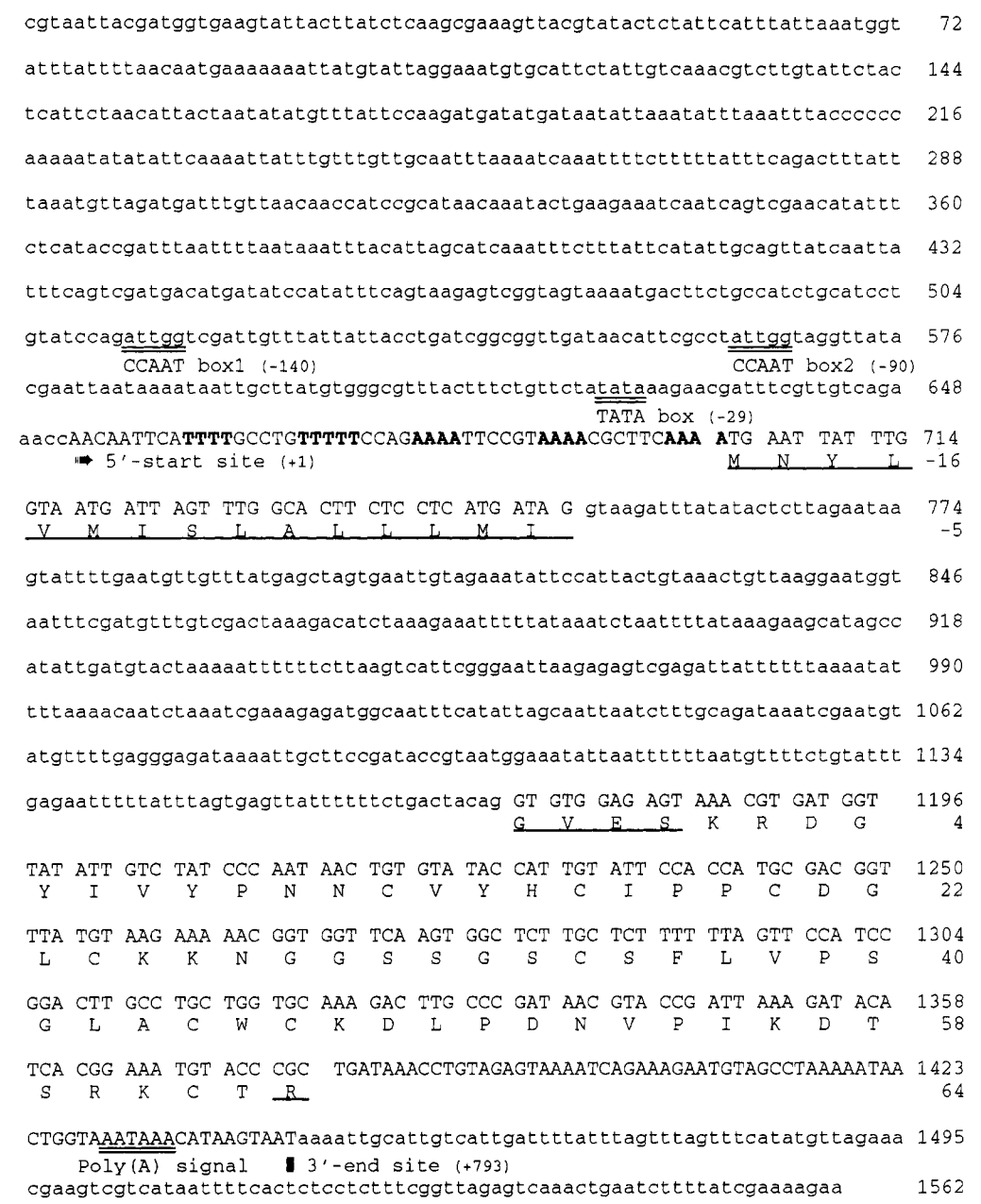

FIGURE 2: Complete nucleotide sequence of the AaH I' toxin gene including the intron and the 5'- and 3'-flanking regions. The sequence of the transcriptional unit is in capital letters. The transcriptional start site and the end site are noted below the sequence. The derived amino acid sequence is presented in the one-letter code below the codons and is numbered separately starting at the first amino acid of the native toxin. The signal peptide of the toxin precursor and the C-terminal Arg residue are underlined once. Both are cleaved to generate the native toxin. The polyadenylation signal AATAAA is underlined twice. Sequences that may be involved in gene regulation are underlined twice (TATA and CCAAT boxes). Stretches of $d(A)_n$ or $d(T)_n$ ($n > 3$) responsible for the stable bending of the AaH I' promoter are boldfaced.

at 42 °C in 50 mM Hepes, 1× Denhart's medium, 3× SSC, 20 mg/mL sonicated salmon testis DNA, 1 mg/mL yeast tRNA, and 50% formamide. Filters were washed at high stringency (0.2× SSC, 0.1% SDS) for 40 min at 53 °C. Restriction fragment analysis was performed under the same conditions of hybridization.

Polymerase Chain Reaction (PCR) Amplification and Sequencing Procedures. Gene-Amp (Perkin Elmer-Cetus) with native Taq DNA polymerase was used for PCR amplification of DNA. PCR reactions contained 20 ng of DNA and 1 μM each oligonucleotide primer. The thermal profile involved 25 cycles of 1 min at 94 °C, 1 min at 37 °C, and 1 min at 72 °C. According to the genomic region to be amplified, the following sets of primers were used: (i) the forward G23-*Hin*dIII (654−679) 5'-CCCAAGCTTACAATCTATTTTGCCTGCTTTTCCAG-3' and the reverse C28-*Bam*HI (1390−1415) 5'-CCCGGATCCAGGATACATTCTTTCTGATTTTACTC-3'; (ii) the forward T50-*Pst*I−*Sac*I 5'-TCCCCCGGGAACTGCAGGAGCT-3' and the reverse T22-*Hin*dIII (676−701) 5'-CCCAAGCTTTTGAAGCGTTTTACGGAATTTT-3'; iii) the forward A27-*Bam*HI (1377−1403) 5'-CCCGGATCCTGATAAACCTGTAGAGT-AAAATCAGA-3' and the reverse T50-*Pst*I−*Sac*I; (iiii) the forward G105-*Hin*dIII (449−430) 5'-CCCAAGCTTTTATTTCAGTCGATGACATG-3' and the reverse C106-*Bam*HI (907−927) 5'-CGCGGATCCCATCAATATGGCTATGCTTC-3'; the numbers in parentheses refer to the position of the primer in the gene nucleotide sequence (Figure 2). The two complementary primers A51-*Pst*I−*Sac*I 5'-CCTGCAGTTCCCGGGGGA-3' and T50-

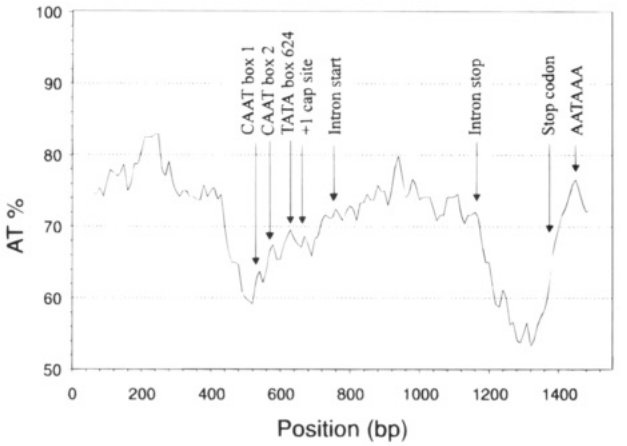

FIGURE 3: Profile of the A+T content of the nucleotide sequence of the AaH I′ gene. The A+T content was calculated for a 120 bp segment moving along the molecule in 10 bp steps. Specific features of the gene are indicated in regard to their position in the nucleotide sequence.

*Pst*I−*Sac*I were allowed to hybridize in order to be ligated to the 3′ overhangs of the 4.4 kb *Sac*I fragment. The primer T50-*Pst*I−*Sac*I was further used with the primer T22-*Hin*dIII or the primer T27-*Bam*HI to amplify the 5′ and 3′ regions flanking the structural part of the gene from the product of the ligation. The DNA fragments obtained were subsequently subcloned into either M13mp18, M13mp19, pUC 18, or pUC19 and sequenced by the dideoxy chain-termination method (Sanger et al., 1977) using Sequenase (United States Biochemicals) with universal and sequence-specific primers and [^{35}S]dATPαS (Amersham). To avoid PCR-base sequence artifacts, two or more individual clones were sequenced systematically.

Primer Extension Reactions. Poly (A)$^+$ mRNA was obtained from 10 fresh scorpion telsons (0.7 g). The animals were sacrificed 2 days after manual extraction of their venom to allow the toxin-producing cells of the venom glands to enter the secretory phase. Total RNA was first extracted from homogenized telsons using the guanidinium hot-phenol method (Sambrook et al., 1989), using vanadyl−ribonucleoside complex (Bethesda Research Laboratory) as an effective ribonuclease inhibitor. About 20 μg of poly (A)$^+$ mRNA was further selected using oligo(dT)-cellulose (Bethesda Research Laboratory) chromatography (Aviv & Leder, 1972). Ten picomoles of primer T22-*Hin*dIII (see above) was 5′-end ^{32}P-labeled by T4 polynucleotide kinase using [γ-^{32}P]-ATP having a specific radioactivity >5000 Ci/mmol (Amersham). The ^{32}P-labeled primer was purified on a NAP column (Pharmacia). The specific radioactivity achieved was 7.1×10^6 dpm/pmol. In a siliconized PCR tube with a final volume of 8 μL, 4 μg of poly (A)$^+$ mRNA (an estimated 6 fmol of target mRNA) was hybridized to 30 fmol of the ^{32}P-labeled primer (about 200 000 dpm) in the presence of 1× reverse transcriptase buffer. The control experiment, carried out under the same experimental conditions, contained 4 μg of total yeast RNA. The hybridization was performed according to the following thermal profile: 5 min at 70 °C, from 70 to 40 °C in 1 h, and from 40 to 30 °C in 15 min. Each reaction was made up to a final volume of 50 μL containing 5 μL of 10× reverse transcriptase buffer, 2 μL of 25 mM dNTP (Pharmacia), 35 U of RNAguard (Pharmacia), 2 μg of actinomicin D (4 mg/mL of 80% ethanol), and 100 U of reverse transcriptase (New England Biolabs).

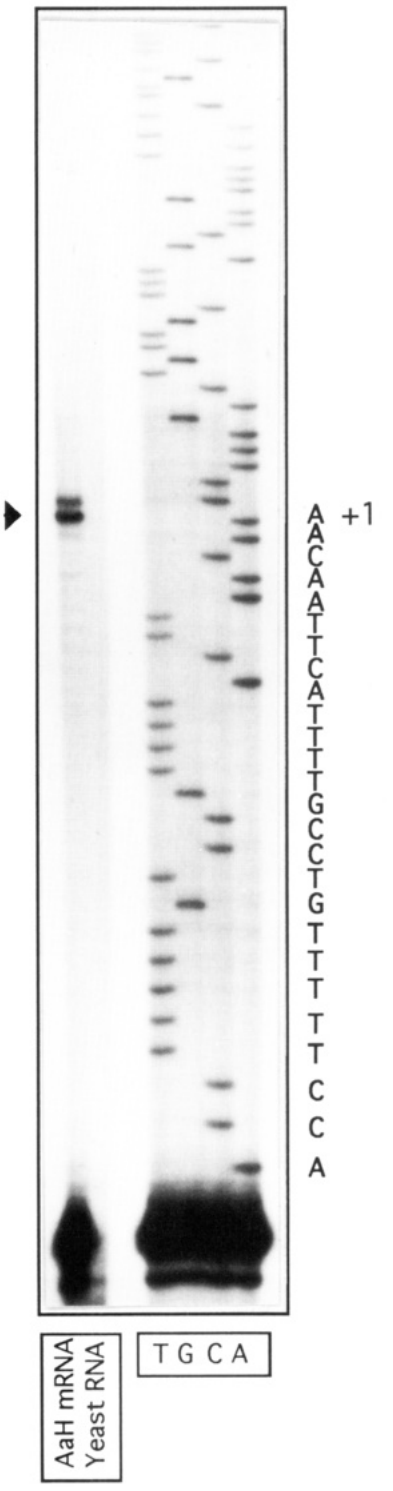

FIGURE 4: Determination of the transcription initiation site of the AaH I′ gene by primer extension analysis. The 5′-end ^{32}P-labeled primer T22-*Hin*dIII was hybridized with either 4 μg of poly (A+) mRNA from *A. australis* scorpion telsons or 4 μg of total yeast RNA (control experiment) and extended with reverse transcriptase. Opposite is the dideoxy sequencing reactions carried out with the double-stranded 1.8 kb DNA of the 5′-flanking region of the gene (see Figure 1b) using the same ^{32}P-labeled primer. On the right is noted the nucleotide sequence determined.

The reverse transcriptase reaction was carried out at 37 °C for 1 h and terminated by adding 5 μl of 4 M ammonium acetate and 50 mM EDTA, and 4 μL of 10% (w/v) SDS, 5 mg/mL yeast RNA. The samples were extracted once with phenol−chloroform−isoamyl alcohol (25:24:1) saturated with 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA, then

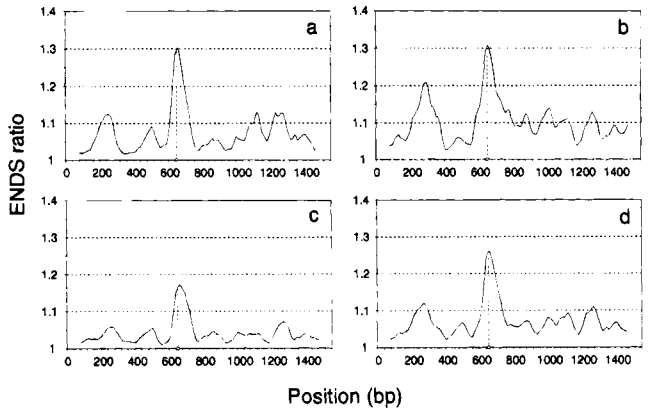

FIGURE 5: Sequence-directed curvature of the AaH I′ toxin gene. Plots of the variation of DNA curvature (ENDS ratio) of 120 bp fragments as functions of their position in the AaH I′ toxin gene are shown. Curvature maps were calculated with three independent models of DNA bending. (a) Dinucleotide wedge model of Trifonov (Bolshoy et al., 1991); (b) junction model of Crothers (Koo & Crothers, 1988); (c) wedge model of de Santis (de Santis et al., 1988); (d) average curvature map calculated from the three theoretical models of DNA bending. The position of the transcription initiation site is indicated by an open circle.

precipitated with ethanol, and resuspended in 5 μL of sequencing gel loading buffer for analysis on a 8 M urea−12% acrylamide DNA sequencing gel. A set of dideoxynucleotide DNA sequencing reactions using 7 × 10^6 dpm of the ^{32}P-labeled primer and 2 pmol of the pUC template harboring the 5′-flanking region of the gene of 1.8 kb was generated using Sequenase and was used for size markers.

Computer Modeling Analysis of the Sequence-Directed Curvature of DNA. The three-dimensional trajectory of the DNA molecule was calculated from its primary sequence as previously described (Pasero et al., 1993) using the tilt and roll angles given by the theoretical models of DNA curvature of Trifonov (Bolshoy et al., 1991), Crothers (Koo & Crothers, 1988), and de Santis (de Santis et al., 1988) and by the twist angles calculated by Kabsch et al. (1982). Curvature maps were calculated for a 150 bp segment moving along the molecule in 10 bp steps (Eckdahl & Anderson, 1987). The probability of occurrence of $(dA)_n$, $(dT)_n$ ($n > 3$), ApA, and TpT dinucleotides along the molecule was investigated by positional autocorrelation analysis as described by Marini et al. (1982).

Electrophoretic Analysis of DNA Curvature. A fragment of 516 bp centered around the bent motif of the AaH I′ promoter was amplified by PCR using primers G105-*Hin*dIII and C106-*Bam*HI and was submitted to electrophoresis on a 7% polyacrylamide gel (acrylamide−bis acrylamide, 37.5:1, by weight). Electrophoresis was carried out at constant voltage (5.3 V/cm) in a cold room at 7 °C for 24 h in 1× TBE buffer. The gel was stained with ethidium bromide (0.5 μg/mL) and photographed under UV light. λDNA digested with *Hin*dIII (Boehringer) and 123 bp ladder (Bethesda Research Laboratory) were used as molecular weight markers. An empirical relationship between the degree of DNA curvature and the altered electrophoretic mobility in polyacrylamide gels has been generated by Thompson and Landy (1988), using a set of DNA fragments with independently determined bending angles as standards. Here, the degree of curvature of the 516 bp fragment was estimated from mobility data (three independent gel electrophoresis) with the cosine function:

$$1/K = \cos \alpha/2$$

where K is the ratio of the observed length to the expected length of the DNA fragment and α is the bending angle of the fragment studied (Thompson & Landy, 1988).

Sequence Analysis of DNA. The sequence data analysis and the search for specific sequence motifs and for sequence homologies have been made using BISANCE: a French service for access to biomolecular sequence databases (Dessen et al., 1990).

RESULTS AND DISCUSSION

Isolation and Characterization of the Gene. About 9 × 10^5 λFix clones of the *A. australis* genomic library were screened with the ^{32}P-labeled *Bam*HI−*Pst*I fragment from the previously identified 368 bp cDNA (pcD633) encoding the toxin AaH I (Bougis et al., 1989). The clones λFix-AR802, λFix-AR812 and λFix-AR828 were obtained. The clone λFix-AR802, giving the most intense signal on hybridization with the probe, was investigated further. The overall genomic sequence of λFix-AR802 is about 20 kb (Figure 1a). From this sequence, a single *Sac*I fragment of 4.4 kb hybridized with the probe in a Southern blot experiment. We used the PCR DNA amplification technique to analyze this fragment. The primers G23-*Hin*dIII and C28-*Bam*HI, comprizing the most distal sequences of the AaH I cDNA (Figure 1c), amplified a single band of about 800 bp from either (i) the genomic DNA (as used to construct the library; (ii) the clone λFix-AR802; (iii) or the 4.4 kb *Sac*I fragment. Because the corresponding cDNA is 368 bp, this indicates that the amplified fragment contains about 400 bp of intron sequence(s).

We used the primers T22-*Hin*dIII, A27-*Bam*HI, T50-*Pst*I−*Sac*I and A51-*Pst*I−*Sac*I to amplify by PCR both the 5′- and 3′-flanking regions of the gene (1.8 and 1.6 kb in length, respectively, see Materials and Methods and Figure 1b). From these PCR amplified fragments, a sequence of 1562 nucleotides which comprizes the gene in full was determined and is shown in Figure 2. The gene contains a single intron (Figure 1c and Figure 2). The deduced amino acid sequence encoded the toxin AaH I′ and not toxin AaH I, the cDNA (pcD 633) of which was used to obtain the probe. Indeed, the toxin AaH I′ differs from the toxin AaH I only in a single mutation, V17-I, at position 1233 in the gene sequence (Bougis et al., 1989). In addition, the gene sequence contains two nucleotide differences at positions 1223 and 1370 compared to the pcD 639 cDNA sequence encoding the toxin AaH I′. Both differences are silent mutations, taking place at the third base of codons. It is most probable that these differences in the nucleotide sequence are due to the polymorphism usually observed between individuals. In this respect, it is to be noticed that the *A. australis* venom gland cDNA library used to clone the toxin precursors was made from 30 scorpions (Bougis et al., 1989) and that the genomic DNA library we have used herein was created using only two scorpions. At the protein level, such polymorphism has been already reported for the scorpion toxins (Martin-Eauclaire & Couraud, 1994). Indeed, due to the very small amount of venom that a scorpion produces, commercial venoms consist of pools of the venoms of many individuals. Finally, as expected, the gene contains the codon for the additional arginine residue found at the C-terminus of the precursor of the toxin AaH I′.

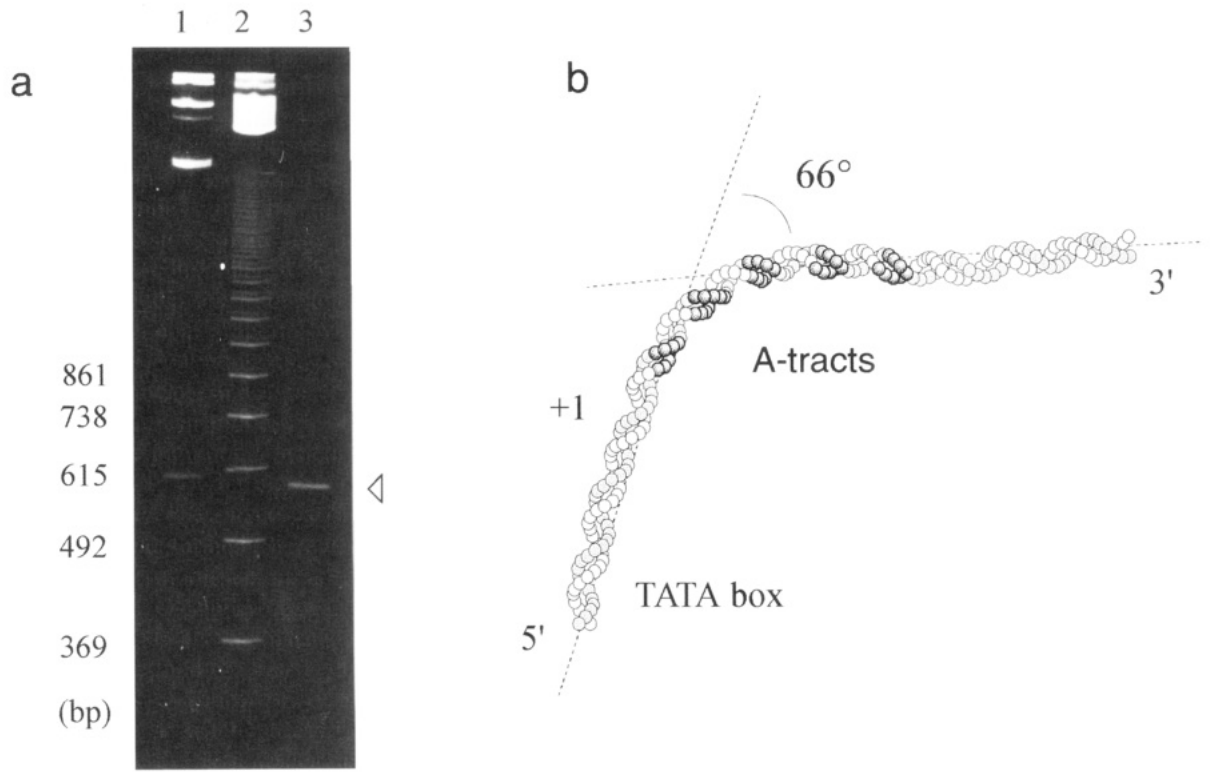

FIGURE 6: Polyacrylamide gel electrophoresis analysis of the intrinsic curvature of the AaH I′ promoter. (a) 7% polyacrylamide gel electrophoresis: Lane 1, λ DNA digested with *Hin*dIII; lane 2, 123 bp ladder; lane 3, AaH I′ promoter (516 bp fragment centered around the initiation site). With reference to the 123 bp ladder, the 516 bp AaH I′ fragment (arrow head) migrates as a DNA fragment of 598 bp ($K = 1.16$). (b) 3D model of the bent locus of the AaH I′ promoter calculated with the dinucleotide wedge model. Five $(dA/dT)_4$ tracts in phase with the helical pitch of DNA are shown (shaded spheres). Since it has been shown that the mobility of DNA fragments in polyacrylamide gels fits to a cosine function (Thompson & Landy, 1988), we assume that this gel retardation correspond to a bending angle of approximately 61° (see Materials and Methods). This estimate is fully consistent with the bending angle predicted by computer modeling (~66°).

The intron−exon junctions, which are typical of donor and acceptor splice sites, have been established in reference to the sequence of the pcD 639 cDNA encoding the toxin AaH I′. The coding part of the gene is divided into two exons and a single intron of 425 bp, which is located close to the end of the signal peptide just after the first base of a glycine codon at position −4 (Figure 1c; Figure 2). The rule that introns begin with GT and end with AG applies (Breathnach & Chambon, 1981). As most frequently observed in *Drosophila*, the exon−intron boundary sequences are AG/GTAAG (5′ splice site) and TACAG/RT (3′ splice site) (Mount et al., 1992). Figure 3 shows the A+T profile of the gene nucleotide sequence. The A+T content of the intron is 75% (33% A, 5% C, 19% G, and 42% T). This percentage is consistent with the structure of the introns of many other species, which are also A+T-rich (Csank et al., 1990). Of particular interest is the pyrimidine-rich region (up to 65%) associated with the 3′ splice site, which has been shown to play a critical role in the branch point during splicing (Ruskin et al., 1988). Upstream of the putative regulatory sequence elements of the gene promoter (see below), the nucleotide sequence is also A+T rich whereas the coding region of the gene is characterized by a low A+T content.

Determination of the Transcription Initiation Site of the Gene. The transcription initiation site was determined by primer extension using reverse transcriptase and the ^{32}P-labeled primer T22-*Hin*dIII, which is complementary to the 5′-end of the toxin mRNA. On the same sequencing gel was loaded a set of dideoxynucleotide DNA sequencing reactions, using the ^{32}P-labeled primer T22-*Hin*dIII and the 1.8 kb 5′-flanking region of the gene, for a direct correspondence between the primer extension product and the gene sequence (Figure 4). A major primer extension product is clearly observed, which corresponds to the start sequence AACAA at position 653. These data agree with the previous finding that the cDNAs encoding the toxins AaH I′, AaH I, and AaH II start with the sequence AACAA and consequently are to be considered as full length (Bougis et al., 1989). A minor extension product is also observed, which is one nucleotide longer.

Determination of the 3′-End of the Gene. The 3′-end of the gene has been defined by comparing its nucleotide sequence with that of pcD 639 (AaH I′) containing a poly(A) tail (Bougis et al., 1989). The nucleotide sequence at the 3′-end of the gene (AATAAA[n = 9]NAAA[n = 3]CATTGNCATTG) matches the class II consensus for recognition elements surrounding poly(A) addition sites involved in U4-RNA mediated processing according to Berget (1984). Accordingly, the transcriptional unit of the gene expends over 793 bp.

Putative Regulatory Sequence Elements of the Gene Promoter. Searches for regulatory sequences that have been identified as binding sites for conserved DNA-binding proteins in other genes (TFSDSITE, version 7.3) show that the gene promoter contains sequences matching the following consensus sequences: MAT-α2 (ATTTACAT) at position 385 (Ko et al., 1988), Pit-1 (WTATYCAT) at positions 408 and 450 (Elsholtz et al., 1990), and IEF1 (GCCATCTG) at position 491 (Ohlsson et al., 1988). In accordance with consensus promoter sequences, there are positive-acting sequence elements including a TATA box at position 624

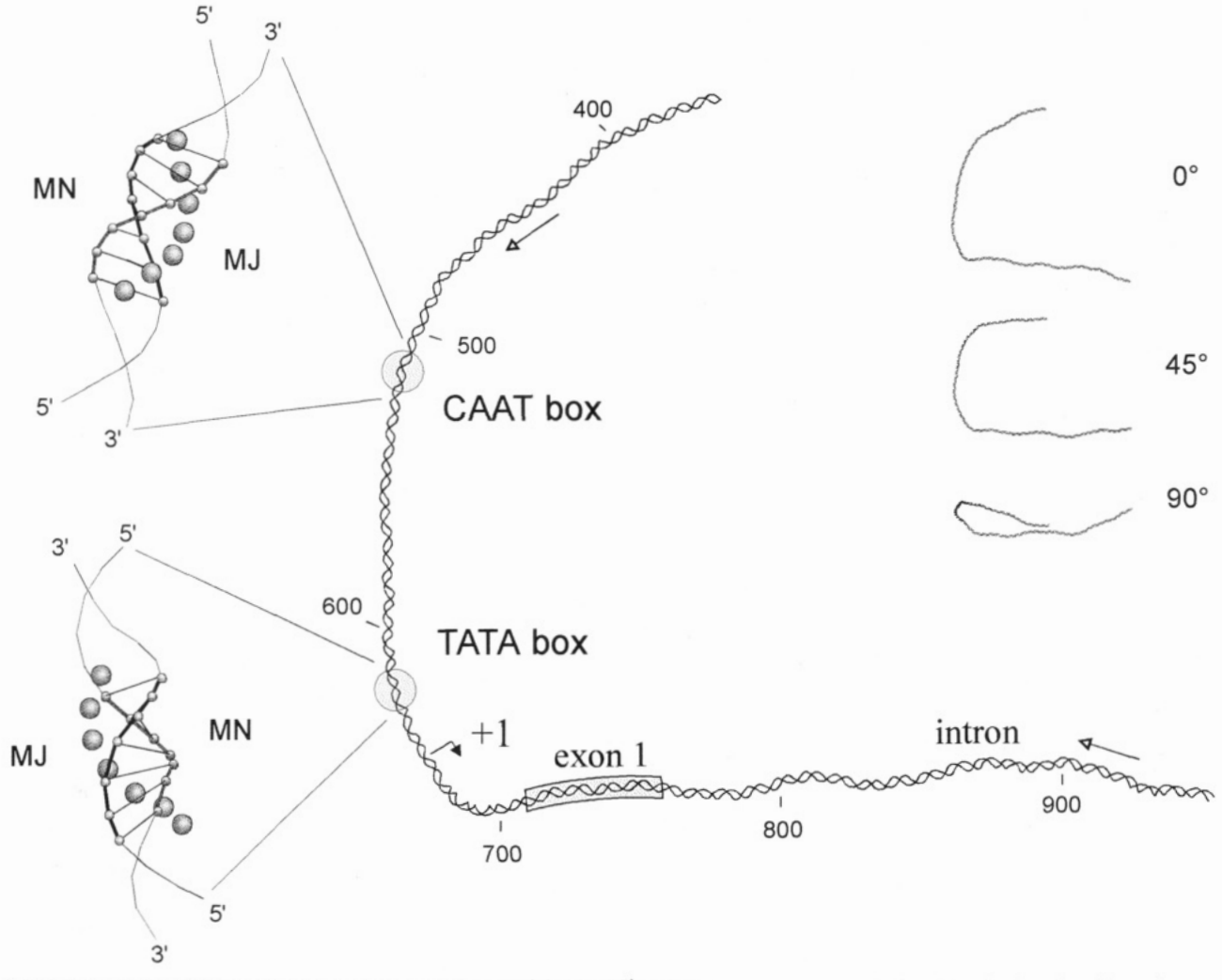

FIGURE 7: Two-dimensional projection of the spatial path of the AaH I′ toxin gene promoter as calculated by the dinucleotide wedge model of Trifonov (Bolshoy et al., 1991). Rotations of the fragment in a counter clockwise direction by 0°, 45°, and 90° indicate that all the bending elements lie approximately in one plane. The main bending element (66°) cooperates with other minor deflections of the helix axis and leads to an overall curvature of the promoter region of nearly 90°. Enlarged portions of the model show that the major groove (MJ) of the putative CCAAT box (at position 513) and the minor groove (MN) of the putative TATA box (at position 624) are turned toward the center of the curve. The primers G105-*Hin*dIII and C106-*Bam*HI, used to amplify by PCR the 516 bp fragment centered around the initiation site, are indicated by arrows.

(−29, relative to the start site) and two CCAAT boxes further upstream at position 513 and 563 (−140 and −90, relative to the start site), the first having the highest statistical score and the second being at the best consensus position (−70 to −90; Figure 2) (Breathnach & Chambon, 1981; Kadonaga et al., 1986).

Analysis of DNA Curvature. Several recent studies have shown that sequence-directed curvature and protein-mediated DNA bending play a key role in the regulation of gene expression [for reviews, see Hagerman (1990), Travers (1992), van der Vliet and Verrijzer (1993), and Perez-Martin et al (1994)]. In order to test whether bent DNA is also an important structural feature of the AaH I′ promoter, we have analyzed the structure of this DNA fragment by computer modeling using three independent theoretical models of DNA curvature (Bolshoy et al., 1991; Koo & Crothers, 1988; de Santis et al., 1988). These models generally appear to fit the experimental data very well (Eckdahl & Anderson, 1987; Muzard et al., 1990; Pasero et al., 1993), and computer modeling has been developed as a valid alternative to polyacrylamide gel electrophoresis to detect the curved regions of DNA fragments (Boffeli et al., 1992; Nair & Kulkarni, 1994). Here, curvature maps calculated with three theoretical models revealed a strong bending centered around position 670 (Figure 5). This region contains five $(dA/T)_4$ tracts perfectly phased with the helical pitch of DNA (Figure 2). Since stretches of $d(A)_n$ or $d(T)_n$ ($n > 3$) are the major cause for the intrinsic curvature of DNA [for reviews, see Hagerman (1990) and Calladine and Drew (1992)], we assume that these motifs are responsible for the stable bending of the AaH I′ promoter. Bent DNA displays an altered electrophoretic mobility in polyacrylamide gels (Diekmann, 1992). Here, we observed that a 516 bp fragment centered around the initiation site of the AaH I′ gene is significantly retardated on 7% polyacrylamide gels, confirming the theoretical results obtained by computer modeling (Figure 6). This anomalously low mobility is not observed on agarose gels (data not shown), and the bending angle estimated from the experimental data (61°) is in good agreement with the value predicted by computer modeling (66°).

What is the effect of DNA bending on promoter activity? Numerous analyses, both in prokaryotic and in eukaryotic cells, have shown that promoter strength is dependent in part

on the intensity of curvature (Plaskon & Wartell, 1987; Delic et al., 1991). Moreover, several studies have pointed out a correlation between the rotational orientation of the curved DNA relative to the promoter and its ability to enhance transcription (McAllister & Achberger, 1989; Bracco et al., 1989; Ohyama et al., 1992). Here, it is worth noting that the overall curvature of the AaH I′ promoter (nearly 90°, Figure 7) is significantly stronger than similar structures already reported in eukaryotic cells [for a review, see Hagerman (1990)]. Moreover, both the major groove of the putative CCAAT box (at position 513) and the minor groove of the TATA box—which interacts with the TATA box binding protein (Kim et al., 1993)—lie along the inside of the DNA curve (Figure 7). This provocative orientation of the DNA curvature relative to the cis-acting sequences of the promoter strongly suggest that this structural element is important for efficient transcription of the AaH I′ gene. It will be interesting to assess whether this spacial organization of the promoter is a conserved structural feature of scorpion toxin genes.

ACKNOWLEDGMENT

We wish to thank Prof. H. Rochat for his constant interest and support, Dr. M. El Ayeb from the Pasteur Institute of Tunis for providing the scorpions, Miss B. Céard for oligonucleotide synthesis, and Dr. B. Kerfelec for fruitful discussion.

REFERENCES

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408−1412.
Becerril, B., Vasquez, A., Garcia, C., Corona, M., Bolivar, F., & Possani, L. D. (1993) *Gene 128*, 165−171.
Berget, S. M. (1984) *Nature (London) 309*, 179−181.
Boffeli, D., de Santis, P., Palleschi, A., Risuleo, G., & Savino, M. (1992) *FEBS Lett. 300*, 175−178.
Bolshoy, A., McNamara, P., Harrington, R. E., & Trifonov, E. N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2312−2316.
Bougis, P. E., Rochat, H., & Smith, L. A. (1989) *J. Biol. Chem. 264*, 19259−19265.
Bracco, L., Kortlarz, D., Kolb, A., Diekmann, S., & Buc, H. (1989) *EMBO J. 8*, 4289−4296.
Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349−383.
Calladine, C. R., & Drew, H. R. (1992) *Understanding DNA*, Academic Press, London.
Catterall, W. A. (1986) *Annu. Rev. Biochem. 55*, 953−985.
Csank, C., Taylor, F. M., & Martindale, D. W. (1990) *Nucleic Acids Res. 18*, 5133−5141.
Delic, J., Onclercq, R., & Moisan-Coppey, M. (1991) *Biophys. Biochem. Res. Commun. 180*, 1273−1283.
de Santis, P., Palleschi, A., Savino, M. & Scipioni, A. (1988) *Biophys. Chem. 32*, 305−317.
Dessen, P., Fondrat, C.,Valencien, C., & Mugnier, C. (1990) *Comput Appl. Biosci. 6*, 355−356.
Diekmann, S. (1992) *Methods Enzymol. 212*, 30−46.
Eckdahl, T. T., & Anderson, J. N. (1987) *Nucleic Acids Res. 15*, 8530−8545.
Elsholtz, H. P., Albert, V. R., Treacy, M. N., & Rosenfeld, M. G. (1990) *Genes Dev. 4*, 43−51.
Granier, C., Novotny, J., Fontecilla-Camps, J. C., Fourquet, P., El Ayeb, M., & Bahraoui, E. M. (1989) *Mol. Immunol. 26*, 503−513.
Hagerman, P. J. (1990) *Annu. Rev. Biochem. 59*, 755−781.
Kabsch, W., Sander, C., & Trifonov, E. N. (1982) *Nucleic Acids Res. 10*, 1097−1104.
Kadonaga, J. T., Jones, K. A., & Tjian, R. (1986) *Trends Biochem. Sci. 11*, 20−23.
Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993) *Nature (London) 365*, 512−520.
Ko, H.-S., Fast, P., McBride, W., & Staudt, L. M. (1988) *Cell 55*, 135−144.
Koo, H., & Crothers, D. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1763−1767.
Marini, J. C., Levene, S. D., Crothers, D. M., & Englund, P. T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7664−7668.
Martin-Eauclaire, M.-F., & Couraud, F. (1994) in *Handbook of Neurotoxicology* (Chang, L. W., & Dyer, R. S., Eds.) pp 683−716, New York.
Martin-Eauclaire, M.-F., Céard, B., Ribeiro, A. M., Diniz, C. R., Rochat, H., & Bougis, P. E. (1992) *FEBS Lett. 302*, 220−222.
Martin-Eauclaire, M.-F., Céard, B., Ribeiro, A. M., Diniz, C. R., Rochat, H., & Bougis, P. E. (1994) *FEBS Lett. 342*, 181−184.
McAllister, C. F., & Achberger, E. C. (1989) *J. Biol. Chem. 264*, 10451−10456.
Miranda, F., & Lissitzky, S. (1961) *Nature (London) 190*, 443.
Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O., & Fields, C. (1992) *Nucleic Acids Res. 20*, 4255−4262.
Muzard, G., Théveny, B., & Révet, B. (1990) *EMBO J. 9*, 1289−1298.
Nair, T. M., & Kulkarni, B. D. (1994) *Biophys. Chem. 48*, 383−393.
Ohlsson, H., Karlsson, O., & Edlund T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4228−4231.
Ohyama, T., Nagumo, M., Hirota, Y., & Sakuma, S. (1992) *Nucleic Acids Res. 20*, 1617−1622.
Pasero, P., Sjakste, N., Blettry, C., Got, C., & Marilley, M. (1993) *Nucleic Acids Res. 21*, 4703−4710.
Perez-Martin, J., Rojo, F., & de Lorenzo, V. (1994) *Microbiol. Rev. 58*, 268−290.
Plaskon, P. R., & Wartell, R. M. (1987) *Nucleic Acids Res. 15*, 785−796.
Rochat, H., Bernard, P., & Couraud, F. (1979) in *Neurotoxins: Tools in Neurobiology* (Ceccarelli, B., & Clementi, F., Eds.) Vol. 3, pp 325−334, Raven Press, New York.
Ruskin, B., Zamore, P. D., & Green, M. R. (1988) *Cell 52*, 207−219.
Sambrook, J., Fritsh, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Thompson, J. F., & Landy, A. (1988) *Nucleic Acids Res. 16*, 9687−9705.
Travers, A. A. (1992) *Curr. Opin. Struct. Biol. 2*, 71−77.
van der Vliet, P. C., & Verrijzer, C. P. (1993) *BioEssays 15*, 25−32.
Vasquez, A., Becerril, B., Martin, B. M., Zamudio, F., Bolivar, F., & Possani, L. D. (1993) *FEBS Lett. 320*, 43−46.
Zilberberg, N., Zlotkin, E., & Gurevitz, M. (1991) *Toxicon 29*, 115−1158.
Zilberberg, N., Zlotkin, E., & Gurevitz, M.(1992) *Insect Biochem. Mol. Biol. 22*, 199−203.
Zoltkin, E., Gurevitz, M., Fowler, E., & Adams, E. (1993) *Arch. Insect Biochem. Physiol. 22*, 55−73.

BI942866L