Precursors of *Androctonus australis* Scorpion Neurotoxins

STRUCTURES OF PRECURSORS, PROCESSING OUTCOMES, AND EXPRESSION OF A FUNCTIONAL RECOMBINANT TOXIN II

Pierre E. Bougis, Hervé Rochaç, and Leonard A. Smith

From the Toxinsology Department, Pathology Division, United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21701-5011 and the Université d'Aix-Marseille II, Faculté de Médecine Secteur-Nord, Biochimie, Centre National de la Recherche Scientifique URA179-Institut National de la Santé et de la Recherche Médicale U172, Bd. P. Dramard, 13326 Marseille Cédex 15, France

From a cDNA library made from telsons of scorpions of the species *Androctonus australis*, full-length cDNAs of about 370 nucleotides encoding precursors of toxins active on mammals or on insects have been isolated using oligonucleotide probes. Sequence analysis of the cDNAs revealed that the precursors contain signal peptides of about 20 amino acid residues. In addition, precursors of toxins active on mammals have extensions at their COOH-terminal ends: Arg or Gly-Arg. The processing steps required to generate toxins from their respective precursors are thus not identical for all of them. Southern blot analysis performed at the genomic level with a cDNA encoding toxin II suggested a single copy gene having a minimum size of 2800 base pairs. Finally, in an attempt to successfully express an animal toxin, monkey kidney COS-7 cells transfected with a plasmid harboring a cDNA encoding toxin II transiently expressed a recombinant toxin having the immunological and biological properties of toxin II.

The Buthid scorpion venoms are known to contain single-chain basic proteins of 60–70 amino acid residues (1, 2) tightly reticulated by four disulfide bridges (3). The acute toxicity of these proteins is due to their high binding affinity for voltage-sensitive sodium channels, thus impairing the initial rapid depolarization phase of the action potential in nerve and muscle. The pharmacology of these neurotoxins has been reviewed elsewhere (4). Interestingly, they may be species-specific, and a venom may contain toxins selectively active on mammals and others lethal to either insects (5) or crustaceans (6).

During the last 30 years, many studies have focused on the purification and characterization of the structural and immunological properties of numerous scorpion neurotoxins. Five structural groups of toxins active on mammals have been defined according to both sequence and antigenic homologies (7, 8). From their binding properties and pharmacology, they have also been classified into α- or β-type toxins (8, 9). Chemical modification has allowed the determination of critical amino acid residues involved in the toxic activity (10). Five main regions appear responsible for antigenicity and have been mapped by using homologous toxins (11), chemically modified toxins (12), and/or synthetic peptides (reviewed in Ref. 13). Thus, while scorpion neurotoxins are polymorphic proteins, conserved domains exist in which putative toxic sites have been proposed to be located (14). The first three-dimensional structure of a “variant” protein with mild toxicity was solved by x-ray crystallography analysis a few years ago (15), and very recently that of the most potent among the α-toxins has been determined at high resolution (16).

For improved serotherapy of Buthid scorpion stings as well as the development of a possible vaccine, we wished to pursue a study of structure-activity relationships in scorpion neurotoxins by carrying out specific amino acid substitutions in toxin molecules using recombinant DNA techniques. Considering the lack of knowledge of the molecular biology of scorpion neurotoxins, as a first step we cloned and sequenced cDNAs encoding precursors of *Androctonus australis* neurotoxins. The venom of this scorpion species living in North Africa has been extensively studied. Toxins comprising less than 2% of venom dry weight were purified from animals collected in Algeria (area of Chellala) and sequenced: AaH I (17), AaH I' (17), AaH I" (17), AaH II (18), AaH III (19), toxins active on mammals, and AaH IT (20), a toxin active on insects. The venom of animals collected in Tunisia (area of Tozeur) differed by the absence of AaH I" and the presence of a second toxin active on insects: AaH TT2; AaH IT1 corresponded to AaH IT on the basis of amino acid composition (21). The processing steps required to generate toxins from precursors were analyzed by comparing their amino acid sequences. Southern blot analysis was performed at the genomic level. Furthermore, the present paper describes experiments in which COS-7 cells were transfected with a construct containing an *A. australis* toxin II cDNA to investigate the ability of COS-7 cells to express a biologically active recombinant toxin.

EXPERIMENTAL PROCEDURES

Materials—The following reagents were obtained from the indicated sources: [35S] and [32P]-labeled nucleotides, DuPont-New England
Purification and Characterization of Telson mRNA—Scorpions of the species *A. australis* were collected in the area of Beni-Khedache (Tunisia) for the Pasteur Institute of Tunis. Animals were transported alive to the United States Army Medical Research Institute of Infectious Diseases where they were killed 2 days after manual extraction of their venom to allow the toxin-producing cells of the venom glands to enter the secretory phase. The origin of the mRNA was the telson, the last segment of the tail containing the two venom glands. Total RNA was extracted from homogenized telsons using an guanidinium hot phenol method (22). Vanadyl-ribonuclease complex was used as an effective ribonuclease inhibitor. Poly(A)+ mRNA was further selected using oligo(dt)-cellulose chromatography (23). The intactness of mRNA and its ability to serve as a template for full-length cDNA transcription was measured using 1 pg of mRNA, 0.4 pg of oligo(dt)16 as primer, and 200 units of Moloney murine leukemia virus reverse transcriptase. 32P-Labeled primary transcripts were submitted to electrophoresis on a 1.5% alkaline agarose gel, blotted onto Whatman DE-81 paper, and exposed to X-AR film with intensifying screens for 48 h at −70 °C. 

Construction of a cDNA Library—The Okayama-Berg cloning and expression system (24) was used to generate the library. From 9 pg of telson mRNA and 2 pg of oligo(dt)-tailed pcDV-1 plasmid primer, about 8 × 10⁶ E. coli K12 MC1061 transformants were generated, which constituted the cDNA library used in this work.

Screening of the cDNA Library—For each screening experiment, 400,000 clones from the cDNA library were analyzed by oligonucleotide probes having sequence homology to regions coding for toxins. Probes were synthesized on a Biosearch 8700 DNA synthesizer by β-cyanoethyl phosphoramidite chemistry and were 32P end-labeled by T4 polynucleotide kinase. High and low density screenings of bacterial colonies for recombinant plasmids were performed on nitrocellulose filters (25). Filters were prehybridized for 2 h at 37 °C in 6 × SSC, pH 7.0, containing 1 × Denhardt’s, 0.5% SDS, 100 μg/ml sheared and denatured salmon sperm DNA, and 0.05% sodium pyrophosphate. Filters were then hybridized in 6 × SSC, pH 7.0, containing 1 × Denhardt’s, 20 μg/ml yeast tRNA, 0.05% sodium pyrophosphate, and the 32P end-labeled oligonucleotide probe for 16 h at 37 °C. Subsequent washes were performed in 6 × SSC, pH 7.0, containing 0.05% sodium pyrophosphate at 37 °C for 1 h and once at 47 °C for 5 min before autoradiography using X-AR film with intensifying screens for 36 h at −70 °C.

DNA Sequence Analysis—Inserts within BamHI and PstI restriction sites adjacent to the poly(A/T) and poly(G/C) tracts of the pcD vector were subcloned into both M13mp18 and M13mp19. Nucleotide sequence was determined by the dideoxy sequencing method (26). Sequenase, [5S]deoxyadenosine 5'- (α-thio)triphosphate, and the 7'-base universal M13 primer were used. Because of difficulties in sequencing through the poly(A/T) tract (except in the case of pcD-401) and because of limited enzyme restriction sites within cDNA sequences, the following primers were synthesized and used to prime M13 clones: clones pcD-634, -635, -639 subcloned into M13mpl8, S'CAAAGGATATTGCTGCT3'; clones pcD-644, -645, -648 subcloned into M13mpl8, 5'-CAAGGATATTGCTGCT3';

Table I

Synthetic oligonucleotides used for hybridization probes

Shown are portions of the amino acid sequences of AaH I (17), AaH II (17), AaH III (19), AaH II (18), and AaH IT (20) including their reverse translation in all possible nucleotide sequences. I is inosine.

<table>
<thead>
<tr>
<th>8</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>AaH I</td>
<td>Tyr Pro Asn Asn Cys Val Tyr His Cys Val Pro Pro</td>
</tr>
<tr>
<td>5' TAC CCI AAC AAC TGT GTI TAC CAC TGT GTI CCI CC 3'</td>
<td>T T T T</td>
</tr>
<tr>
<td>AaH II</td>
<td>Tyr Cys Asn Glu Glu Cys Thr Lys Leu Lys Gly Glu Ser Gly Tyr Cys Gln Trp Ala</td>
</tr>
<tr>
<td>5' TAC TGC AAC GAA TGC AC 3'</td>
<td>T T T G G T</td>
</tr>
<tr>
<td>AaH I</td>
<td>Tyr Pro Asn Asn Cys Val Tyr His Cys Val Pro Pro</td>
</tr>
<tr>
<td>5' GGA TAC TGC CAA TGG GC 3'</td>
<td>C T T G G</td>
</tr>
<tr>
<td>AaH III</td>
<td>Asn Ser Lys</td>
</tr>
<tr>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>AaH II</td>
<td>Tyr Cys Asn Glu Glu Cys Thr Lys Leu Lys Gly Glu Ser Gly Tyr Cys Gln Trp Ala</td>
</tr>
<tr>
<td>Asn Gln Cys Thr Lys Val His Tyr Ala Asp Lys Gly</td>
<td></td>
</tr>
<tr>
<td>5' AAC CAA TGG ACI AAA GTI CAC TAT GCI GAC AAA GG 3'</td>
<td>G G G G</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
</tr>
</tbody>
</table>
clones pcD-402, -403 subcloned into M13mp18, 5'AGTTGAAAG-GTGAGGAGT3'; clone pcD-402 subcloned into M13mp19, 5'TACG-TAGCTATGCGGCG3'; clones pcD-635, -634, -639 subcloned into M13mp19, 5'TCGGTACCTAATGCGGCG3'; clones pcD-644, -645, -648 subcloned into M13mp19, 5'TGACATTAGACCGAAGC3'.

Sequence Analysis—Programs from Intelligenetics including SEQ, PEF, QUEST, and IPIND, were used on a VAX from Digital Equipment.

DNA Transfection—Transfection of cesium chloride-purified plasmid was performed by the DEAE-dextran method using chloroquine as described by Luthman and Magnusson (27), COS-7 (SV40-transformed African green monkey kidney) cells obtained from the American Type Culture Collection (CRL 1651, passage number: 9) were maintained at 37 °C under 5% CO₂ in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Cells were plated onto 75-cm² plates and were subconfluent at the time of transfection.

Post-transfection, cells were fed each day with 5 ml of Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Cells were plated on top of the nucleotide sequence.

Sequences are aligned for maximum homology with that of clone pcD-633, and differing nucleotides in each sequence are indicated by symbols placed on top of nucleotide sequence.

The predicted protein sequences are given below the nucleotide sequences and are numbered with the NH₂-terminal amino acid residue of the toxin. Signal peptide sequences are underlined. A potential polyadenylation signal of AATAAA is italicized.

**Clones pcD-402, -403 subcloned into M13mp18, 5'AGTTGAAAG-GTGAGGAGT3'; clone pcD-402 subcloned into M13mp19, 5'TACG-TAGCTATGCGGCG3'; clones pcD-635, -634, -639 subcloned into M13mp19, 5'TCGGTACCTAATGCGGCG3'; clones pcD-644, -645, -648 subcloned into M13mp19, 5'TGACATTAGACCGAAGC3'.

Sequence Analysis—Programs from Intelligenetics including SEQ, PEF, QUEST, and IPIND, were used on a VAX from Digital Equipment.

DNA Transfection—Transfection of cesium chloride-purified plasmid was performed by the DEAE-dextran method using chloroquine as described by Luthman and Magnusson (27). COS-7 (SV40-transformed African green monkey kidney) cells obtained from the American Type Culture Collection (CRL 1651, passage number: 9) were maintained at 37 °C under 5% CO₂ in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Cells were plated onto 75-cm² plates and were subconfluent at the time of transfection.

Post-transfection, cells were fed each day with 5 ml of Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum.

Immunopurification and Characterization of the Recombinant Toxin—CNBr-activated Sepharose 4B from Pharmacia LKB Biotechnology Inc. was used to prepare an anti-AaH II immunosorbent column of 1-ml capacity with 10 mg of antibodies purified from rabbit AaH II-immune serum (28) using an ImmunePure column from Pierce. After elution with 0.1 N formic acid containing 0.15 M NaCl, the immunosorbed recombinant toxin was desalted on Sephadex G-10 in the presence of 0.1 N acetic acid containing 0.1% bovine serum albumin to minimize loss and then lyophilized. AaH II was purified as previously described (2) and was iodinated with ¹²⁵I by the lactoperoxidase method and recovered by immunoprecipitation (29). AaH II-specific immunosassay was essentially as described elsewhere (30), except that 40 x 10⁻⁶ M of AaH II-specific antibodies and 6 x 10⁻¹² M of ¹²⁵I-AaH II were used to increase the sensitivity of the assay. ¹²⁵I-AaH II-receptor assay was performed on rat brain synaptosomes, as previously described (31).

In Vivo Assay—Toxicity on male C57/BL6 mice from Evic-Ceba, 33 Blanquefort, France, weighing 20 ± 3 g was assayed by intracerebroventricular injection as described by Haley and McGirr (35). Southern Blot Analysis—High molecular weight genomic DNA was prepared from A. australis scorpion muscle tissue. After digestion with appropriate restriction endonucleases of 10 µg of DNA, samples were submitted to electrophoresis in 0.8% agarose gel, blotted onto nitrocellulose membranes, and probed with the BamHI/PstI insert of clone pcD-402 encoding AaH II. The probe was ¹²⁵I-labeled using the nick translation kit from Boehringer Mannheim. The filter was
poly(A)+ mRNA were obtained. To both characterize and probe mixtures representing all possible complementary nucleotides, we did not select telson mRNAs by size to construct a toxins of about 65 amino acid residues should be close to 360
by the first and second structural groups, respectively. By
400,000 clones were plated and hybridized to oligonucleotide
sponding in size to about 360 and 1100 bases, the former
oligo(pdT)-primed reverse transcriptase. After electrophore-
assay the mRNAs, primary transcripts were synthesized by

AaH 1', AaH 111, and two iso-AaH IT. For each experiment,
other screening experiments were performed to clone AaH I,
were subsequently determined in full (see "Experimental Pro-
cedures")

RESULTS

Isolation and Sequencing of cDNAs Encoding Toxins—
Starting with 30 fresh telsons giving 2 g of tissue, 42 pg of
cDNAs encoding toxins active on insects and predicted amino acid sequence of toxin precursors. The nucleotide sequences beginning with the 5' end of the cDNA inserts are represented in the 5' to 3' direction and numbered on top. Sequences are aligned for maximum homology with that of clone pcD-644, and differing nucleotides in each sequence are indicated by symbols placed on top of the nucleotide sequence. The predicted protein sequences are given below the nucleotide sequences and are numbered starting at the NH2-terminal amino acid residue of the toxin. Signal peptide sequences are underlined. A potential polyadenylation signal of AA-
TAAA is italicized.

washed extensively in 0.1 × SSC containing 0.1% sodium dodecyl
sulfate.

Scorpion Neurotoxin Precursors

Fig. 2. Nucleotide sequence of cDNAs encoding toxins active on insects and predicted amino acid sequence of toxin precursors. The cDNA sequences begin with the same sequence: AACAA. In the case of pcD-633, clones encoding AaH I' and AaH III were also obtained. Clones chosen at random were colony-purified and analyzed for insert content. The size of inserts within the BamHI and PstI restriction sites suggested that the percentage of full-length cDNAs was high. The

Fig. 1 shows the nucleotide sequences of representative cDNAs encoding toxins active on mammals. cDNA sequences were compared and aligned for maximum homology with that of clone pcD-633. Clones pcD-633 (AaH I), pcD-639 (AaH I'), and pcD-402 (AaH II) contained complete cDNAs beginning with the same sequence: AACAA. In the case of pcD-634 (AaH III) the first four bases of the insert were most probably

structural group (AaH I, AaH I', and AaH III). Using duplicate filters to screen the cDNA library, we obtained about 100 positive clones each time. During the cloning for AaH I, clones showing low intensity hybridization were pooled separately from those of high intensity. Clones encoding AaH I' and AaH III were also obtained. Clones chosen at random were colony-purified and analyzed for insert content. The size of the nucleotide sequences of inserts of a total of 11 pcD clones were subsequently determined in full (see "Experimental Procedures").

The first toxin to be cloned successfully was AaH II. Two other screening experiments were performed to clone AaH I, AaH I', AaH III, and two iso-AaH IT. For each experiment, 400,000 clones were plated and hybridized to oligonucleotide probe mixtures representing all possible complementary nucleotide sequences corresponding to selected amino acid sequence regions of AaH II, AaH I, and AaH IT (Table I). Probes for AaH II were designed to maximize the amino acid sequence differences between AaH I and AaH II, belonging to the first and second structural groups, respectively. By contrast, the probe for AaH I was designed to maximize sequence homologies between all toxins belonging to the first structural group (AaH I, AaH I', and AaH III). Using duplicate filters to screen the cDNA library, we obtained about 100 positive clones each time. During the cloning for AaH I, clones showing low intensity hybridization were pooled separately from those of high intensity. Clones encoding AaH I' and AaH III were also obtained. Clones chosen at random were colony-purified and analyzed for insert content. The size of inserts within the BamHI and PstI restriction sites suggested that the percentage of full-length cDNAs was high. The nucleotide sequences of inserts of a total of 11 pcD clones were subsequently determined in full (see "Experimental Procedures").

Fig. 2 shows the nucleotide sequences of cDNAs encoding
Toxins active on mammals

pcD-401, -402, -403 (AaH II)

pcD-633, -635, -636 (AaH I)

pcD-639 (AaH I')

pcD-634 (AaH III)

Toxins active on insects

pcD-644, -648 (AaH IT1)

pcD-645 (AaH IT2)

Fig. 3. Schematic representation of the structure of toxin precursors deduced from the cDNAs and processing outcomes. From top to bottom are shown, respectively, the sequences of cDNA, toxin precursor, and toxin. The poly(A)* is not represented. The open reading frame is boxed, and the peptide signal domain is dashed. The COOH-terminal extra domain is in black. ▲, represents a site of proteolytic cleavage.

toxins active on insects. Inserts of pcD-644 (AaH IT1) and pcD-648 (AaH IT1) differed once at position 98 by a substitution: C/T corresponding in the open reading frame to the first base of Asn and Tyr codons. The third clone, pcD-645 (AaH IT2), had an insert differing from that of pcD-644 by four substitutions: two in the 5'-noncoding region; T/A at position 170 corresponding to the first base of Asn and Tyr codons; A/C at position 302 corresponding to the first base of Thr and Pro codons.

All cDNA sequences displayed one major open reading frame of about 310 nucleotides. At the 5' end, multiple potential initiation codons were observed. In each case, the 5' proximal ATG codon should have been responsible for the initiation of translation (33). At the 3' end, a putative polyadenylation signal (AATAAA) was found 10-16 nucleotides upstream of the poly(A)* tail (34). It appeared that the 5'- and 3'-nontranslated regions were approximately of the same length.

Primary Structures of Toxin Precursors—For all cDNAs, the open reading frame demonstrated a coding capacity for toxin precursors because they were larger in size than toxins that have been previously purified and characterized from the venom (Fig. 3).

Precursors of toxins active on mammals possessed sequences extended at both NH2 and COOH termini (Fig. 1). The N-terminal initiating Met was the first amino acid residue of a series of highly hydrophobic amino acid residues, suggesting the presence of a signal peptide of 19 residues. Homology among signal peptide sequences was very high. Signal peptide cleavage occurred for each precursor at Ser. Thus, the predicted NH2-terminal residue of each toxin was identical to the one chemically determined on toxins isolated from the venom. By contrast, COOH-terminal amino acid residues predicted from the cDNA nucleotide sequences did not correlate with the COOH-terminal residues chemically determined on toxins isolated from the venom. The toxin precursors ended with additional amino acid residues as compared with toxins: Gly-Arg for AaH IT I and Arg for AaH I, AaH I', and AaH III (Fig. 3). One discrepancy was observed between the sequence encoded by clone pcD-634 and the AaH III sequence previously determined by peptide sequencing (19): Asp at position 8 instead of Asn.

Three different precursors of toxins active on insects were found (Fig. 2). For each of them, the signal peptide was 18 amino acid residues long ending with Gly. These precursors had no additional amino acid residue at their COOH-terminal ends (Fig. 3) when compared to the amino acid sequence of AaH IT previously determined by peptide sequencing (20). The two precursors encoded by clones pcD-644 and pcD-648 had peptide signals substituted at position -2, but each gave the same toxin containing one difference with AaH IT: Glu at position 25 instead of Gln. Clone pcD-645 contained a cDNA coding for an isotoxin differing from AaH IT by two additional substitutions: Tyr at position 23 instead of Asn; Pro at position 67 instead of Thr.

Expression of a Recombinant Toxin in COS-7 Cells—The cloning vector was a transient eukaryotic expression system containing an SV40 origin of replication (24). We used the system to further prove that clone pcD-403 encoding AaH II directed the expression in COS-7 cells of a biologically active recombinant toxin. The culture medium was collected between 24 and 168 h post-transfection and was assayed for the appearance of AaH II-related antigen by a specific radioimmunoassay (Fig. 4a). The optimum quantity of plasmid used to transfect COS-7 cells was 5 µg, and expression was maximum (0.2 µg/10⁶ cells) at 120 h. Culture medium from COS-7 cells transfected in the absence of DNA or transfected with the 5' truncated clone pcD-461 had background values. These results demonstrated that clone pcD-403 was capable of directing the expression of AaH II immunoreactive material in COS-7 cells.

Immunological and Biological Properties of the Recombinant AaH II—To purify the recombinant AaH II secreted by COS-7 cells, the 120-h cell culture medium was chromatographed on an immunoaffinity column. To test whether or not the recombinant AaH II was as active in vivo as AaH II, the affinity-purified product was injected intracerebroventricularly in mice. Mice died with symptoms identical to those observed with mice dying from AaH II. From the results (Table II), the concentration of the recombinant AaH II was estimated to be 6.9 × 10⁻⁸ M, with the assumption that its activity was the same as that determined for AaH II (LD₅₀ = 0.5 ng/mouse). The immunopurified recombinant AaH II was further characterized on the basis of its binding properties to...
performed on cell culture medium. After being purified by immunoaffinity chromatography, the recombinant toxin was further characterized by AaH II-specific immunoassays performed on cell culture medium. After being purified by immunoaffinity chromatography, the recombinant toxin was further characterized by AaH II-specific (b) immunoassay and (c) receptor assay. For both assays, O depicts the standard curve. Ten-fold serial dilutions of the sample of recombinant AaH II, at an initial concentration of $6.9 \times 10^{-8}$ M as estimated from the in vivo experiment of Table II, were tested in both assays, and x represents the results of such experiments.

**Table II**

<table>
<thead>
<tr>
<th>Dilution factor of recombinant AaH II sample</th>
<th>Dead/injected</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>2/2</td>
</tr>
<tr>
<td>1/5</td>
<td>3/5</td>
</tr>
<tr>
<td>1/10</td>
<td>0/3</td>
</tr>
<tr>
<td>AaH II</td>
<td>Dead/injected</td>
</tr>
</tbody>
</table>

AaH II-specific antibodies (Fig. 4b) and rat brain voltage-sensitive sodium channels (Fig. 4c). The same serial dilutions of the sample were tested in parallel. Results established that the recombinant AaH II had the same ability as AaH II to compete in both assays with $^{125}$I-AaH II. Thus, it was most probable that the antigenic and toxic sites of the recombinant AaH II were closely related or identical to those of AaH II.

**Genomic Distribution of AaH II Sequence**—To determine the size and copy number of the AaH II gene, a Southern blot analysis was performed using high molecular weight DNA from scorpion muscle and a $^{32}$P-labeled insert of clone pcD-402. The results of this experiment are shown in Fig. 5. A single band was observed using the following restriction enzymes: EcoRI, BglII, SphI, SmaI, XhoI, Stul, and NruI; the cDNA probe contained no restriction sites for these enzymes. Thus, the results indicated a unique copy gene for AaH II having a minimum size of 2800 base pairs.

**Discussion**

Size analysis of the pool of mRNAs obtained from the scorpion telson indicates that two main mRNA populations exist and that our initial assumption that 360 bases should be the expected size for monocistronic mRNAs encoding scorpion neurotoxin precursors has been experimentally verified. Indeed, by using synthetic oligonucleotides as screening probes, we isolated numerous full-length cDNAs of about 370 base pairs encoding precursors of *A. australis* scorpion neurotoxins.

The comparison of the amino acid sequences of both toxin precursors (as predicted by cDNAs) and toxins (as determined from peptide sequencing) gives an insight into the structure of precursors and the processing they require in order to generate toxins (Fig. 3). There exist an almost complete homology among signal peptide sequences within toxins active on mammals as well as within toxins active on insects. By contrast, the signal peptide sequences are very different when comparing precursors of toxins from both sets. These findings reinforce the classification of scorpion neurotoxins based on amino acid compositions and sequences and CD spectra in which the toxins active on insects are placed apart from all the toxins active on mammals (35). For precursors of toxins active on insects, cleavage of the signal peptide by a signal protease is enough to generate toxins. For precursors of toxins active on mammals, two possibilities exist. For AaH II, in
addition to the cleavage of the signal peptide, the removal of the COOH-terminal dipeptide Gly-Arg is required together with the \(\alpha\)-amidation of the His (18). Precursors of secretory peptides synthesized as part of large and inactive precursor proteins contain sites for proteolysis and \(\alpha\)-amidation. These are often marked by the sequence -X-Gly-\(B-B\); \(X\) is the COOH-terminal amino acid residue in the mature peptide that is \(\alpha\)-amidated, and \(B\) is either Lys or Arg (36). Thus, the additional dipeptide Gly-Arg is in agreement with these observations although it has one basic residue instead of the usual two or none at all as in the case of melittin from bee venom (37). The toxins active on mammals that are not \(\alpha\)-amidated (AaH I, AaH I', and AaH III) require the removal of the signal peptide but also the cleavage of the COOH-terminal Arg residue most probably by a specific exopeptidase. We have no explanation for the fact that different processing steps seem to exist for scorpion neurotoxins. It is worth noting that the removal of the AaH I signal peptide leads to the amino acid sequence of AaH I', an isoform of AaH I differing by an additional Arg residue at the COOH-terminal end (17). However, as stated previously, AaH I' has only been found in the venom of animals of Chelulaa origin. We may suppose that the cDNA encoding AaH I' compared with the encoding for AaH I has been mutated at position 306, as is the case for AaH III; the stop codon TGA being replaced by the Arg codon CGA preceding the second stop codon TAA. For both AaH III and AaH IT, discrepancies exist between the sequences derived from the cDNA and those previously determined by peptide sequencing. In the case of AaH III, we have sequenced two samples of toxin purified from the venom of animals of both Chelulaa and Tozaur origins and found in both cases Asp at position 8 as predicted by the cDNA sequence. The amino acid sequence of AaH III published previously (19) has to be corrected accordingly. In the case of AaH IT, the discrepancies observed have to be confirmed by peptide sequencing of AaH IT1 and AaH IT2. From the amino acid compositions of AaH IT1 and AaH IT2 (21) it seems likely that the sequences encoded by clones pCD-644 and pCD-645 correspond to those of AaH IT1 and AaH IT2, respectively. All the preceding observations may be of concern if we consider, as has been previously suggested, that subspecies of \textit{A. australis} may exist as \textit{A. australis} Hector in Algeria and \textit{A. australis} garzoni in Tunisia (38).

The expression of a recombinant AaH II was achieved by transfection of COS-7 monkey kidney cells with an expression plasmid harboring a cDNA encoding AaH II. As far as we know, this is the first successful expression of a recombinant animal toxin. The rationale for choosing a mammalian expression system was that the biological activity of AaH II is, \textit{stricto sensu}, dependent on the correct formation of four disulfide bridges and the proper folding of the protein (39). The recombinant AaH II expressed by the COS-7 cells was characterized in three different ways: by \textit{in vivo} assay, immunosassay, and receptor assay. The results obtained support the conclusion that COS-7 cells expressed a recombinant protein that behaves as AaH II in each of the three assays. Therefore, we assume that the recombinant toxin that monkey kidney cells transiently expressed upon their transfection is the mature form of AaH II. The fact that the recombinant toxin was secreted from the COS-7 cells in a biologically active form implies that the signal peptide had been removed. Many bioactive peptides have carboxyl-terminal \(\alpha\)-amide residues and, in general, the presence of an \(\alpha\)-amide is critical for biological activity. We do not know what influence the \(\alpha\)-amidation of the COOH-terminal His of AaH II has upon the biological activity of this toxin. Only a biochemical analysis of the recombinant toxin can certify its identity to AaH II and address the accuracy of the double processing of the toxin precursor. The amount of recombinant toxin obtained did not permit such an analysis. We are, at present, attempting to increase the levels of expression of the recombinant toxin in other host systems to answer these questions.

Acknowledgments—Dr. Mohamed El Ayeb is recognized for providing the scorpions. We thank Drs. Linda Lee and Catherine Loridan for their kind help, Dr. Marie-France Martin for assistance with lethality assays, and Pascal Manneville for peptide sequencing of AaH III. We are grateful to Drs. Ariane Monneron and Catherine Ronin for their comments on the manuscript.