Molecular cloning of the gene for the allatostatin family of neuropeptides from the cockroach *Diploptera punctata*  
(Juvenile hormone/endoproteolytic processing/neurobiology)

B. C. Donly*, Q. Ding†, S. S. Tobe†, and W. G. Bendena*

*Department of Biology, Queen’s University, Kingston, ON K7L 3N6, Canada; and †Department of Zoology, University of Toronto, Toronto, ON Canada M5S 1A1

Communicated by Mary Lou Pardue, June 16, 1993

**ABSTRACT**  
Allatostatins (ASTs) are insect neuropeptides that inhibit juvenile hormone biosynthesis by the corpora allata. We have isolated a cDNA from the cockroach *Diploptera punctata* that encodes a 41.5-kDa precursor polypeptide containing the AST family of peptides. Translation of the cDNA revealed a 370-amino acid pre-pro-peptide consisting of 13 AST-type peptides and appropriate processing sites for endoproteolytic cleavage and amidation. The 13 potential AST sequences are characterized by the C-terminal AST core-structure Phe-Gly-Leu-NH₂, with only one exception. Separating the clustered ASTs in the precursor, three acidic spacer regions are found. Contained within the largest of these two potentially related peptides that may also be processed. Southern blot analysis revealed the presence of a single copy of the AST gene per haploid genome, as well as the probability that the gene may be present in at least two allelic forms. In situ hybridization indicated the AST-encoding gene is expressed in neurosecretory cells of *D. punctata* brain.

Juvenile hormones (JHs) are sesquiterpenoid molecules produced by the corpora allata (CA) that play a central role in the metamorphosis and reproduction of most insect species. The presence of JH is required for maintenance of larval characteristics, whereas its absence results in the transition from larva to adult. In the adult female of most species, JH is also required for oocyte development (1).

JH biosynthesis is probably regulated, in part, by a family of brain-derived neuropeptide amides called allatostatins (ASTs), which inhibit JH production by the CA (2). Seven ASTs ranging in size from 8 to 18 amino acids have been purified from the cockroach *Diploptera punctata* (3–5). All share the common C-terminal sequence Tyr-Xaa-Phe-Gly-Leu-NH₂, which has been demonstrated in vitro to be the minimum sequence capable of inhibiting JH biosynthesis (6, 7). The cAMP-independent signal-transduction mechanism(s) through which ASTs exert their effect on the JH biosynthetic pathway is unclear (8). However, membrane receptors for ASTs have recently been identified (9). ASTs have also been identified and purified from muscle tissue, suggesting that these molecules may have other roles apart from regulation of JH production (10).

The recent characterization of several AST peptides from *D. punctata* has provided the opportunity for application of molecular approaches to isolate the gene responsible for production of these factors. Also, in the event that a family of small peptides might be coded for in the form of a single large precursor, as for the *Drosophila* Phe-Met-Arg-Phe-NH₂ (FMRFamide) group (18), gene characterization provides the opportunity to quickly identify additional members of the peptide family. We report here the characterization of a cDNA from the cockroach *D. punctata* that encodes such a precursor for the AST family of neuropeptides. The deduced precursor contains at least 13 potential ASTs, including all seven of the peptides previously isolated.

**MATERIALS AND METHODS**

**DNA Synthesis and Sequencing.** Oligonucleotides were synthesized on a Biosearch model 8750 DNA synthesizer at the Core Facility for Protein/DNA Chemistry (Queen’s University). Templates for DNA sequencing were prepared by PCR amplification of single-stranded cDNA followed by purification on 1% low-melting-temperature agarose gels. Resulting templates were sequenced by the dideoxynucleotide chain-termination method using either of two procedures: (i) Sequenase version 2.0 DNA sequencing kit (United States Biochemical) or (ii) Taq DyeDeoxy Terminator Cycle sequencing kit from Applied Biosystems (Core Facility for Protein/DNA Chemistry).

**RNA Isolation and cDNA Synthesis.** Brains were dissected from the heads of day 2–3 virgin female cockroaches (*D. punctata*) and frozen immediately in liquid nitrogen. Poly(A)⁺ RNA was directly extracted from 1000 brains by using the Pharmacia QuickPrep mRNA purification system and then stored under ethanol at −20°C.

Synthesis of first- and second-strand cDNA from the RNA was done by using the Zap-cDNA synthesis system (Stratagene). The resulting vector-ready cDNA was then ligated to a compatible plasmid, pBluescript II SK⁻ (Stratagene), using T4 DNA ligase (19). The circularized cDNA was diluted 5-fold with water and used as a pool for subsequent PCR.

**PCR Methods.** Standard PCR was done in 100-μl volumes containing 2.5 μl of cDNA (above), 1× PCR buffer (Perkin-Elmer/Cetus), 200 μM dNTPs, bovine serum albumin at 100 μg/ml, 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer/Cetus), and, when using degenerate primers, one AmpliWax PCR Gem. Unique oligonucleotide primers were added to 0.25 μM concentration, whereas degenerate ones were used at 2.5 μM. To produce larger quantities of clean DNA product for sequencing template, PCR reactions were routinely run as nested pairs, using a second set of oligonucleotide primers inside the first to improve specificity.

**Temperature-cycling profiles** varied according to the oligonucleotide primers being used but were generally based on the following model: 95°C for 2.5 min, 50–65°C for 1 min, and 72°C for 1 min per kb of expected product, followed by 34 cycles of 95°C for 45 sec, 50–65°C for 1 min, and 72°C as in

Abbreviations: AST, allatostatin; CA, corpora allata; JH, juvenile hormone.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. U00444).
the first cycle. A final 5-min extension at 72°C was also included.

**Southern Blotting.** High-molecular-weight DNA was purified from adult *D. punctata* essentially as outlined in ref. 20. Aliquots containing 10 μg of extracted DNA were digested with selected restriction enzymes and then separated on a 0.7% agarose gel. The DNA was denatured and transferred under vacuum to Hybond-N+ nylon membrane.

Hybridization was done at 68°C in 5× standard saline phosphate/EDTA (SSPE)/10× Denhardt’s solution/0.5% SDS/denatured salmon sperm DNA at 250 μg/ml for 16 hr (19). The hybridization probe was a random-primed 32P-labeled segment of the AST gene corresponding to bases 76–658 (see Fig. 2). The first wash contained 2× SSPE/0.1% SDS and was done at 60°C, followed by two more washes in 0.5× SSPE/0.1% SDS also at 60°C.

**In Situ Hybridization.** Brains from day 5 mated *D. punctata* females were fixed in 4% (wt/vol) paraformaldehyde and then desheathed by submersion in a solution of collagenase (Sigma, type II, 1 mg/ml) for 15 min before hybridization as described (21). The hybridization probe was the same AST gene segment described above labeled with digoxigenin (21).

**RESULTS**

**PCR Using Degenerate Oligonucleotides.** To isolate the coding sequence of an internal portion of the AST-encoding gene, oligonucleotides were designed to amplify between the ends of the largest AST peptide sequence, the octadecapeptide ASB2 (4). The degenerate sequences are diagrammed in Fig. 1 (primers 1 and 2), in relation to the amino acid sequence of the peptide. Each oligonucleotide carries a 5′ extension that provides restriction recognition sequences for cloning the fragments produced. PCR amplification was done by using as template a cDNA pool synthesized from RNA extracted from the brains of virgin female cockroaches. A product of 68 bp was isolated and cloned; a number of the resulting clones were then sequenced. A consensus sequence for the amplified region was derived from these and used to design a third oligonucleotide having only one degenerate position (primer 3 in Fig. 1), with sufficient specificity to allow 5′-end-specific PCR of the cDNA.

One-sided PCR was used to amplify the 5′ end of the sequence because the cDNA target had first been ligated to a plasmid vector. The plasmid provides a known reference point beyond the cDNA terminus where opposing primers can anneal (22). Therefore, one-sided PCR was performed by using two nested primer pairs: primers 2 and 3 (Fig. 1) in conjunction with two vector-based primers T3 and SK (Stratagene), respectively. In retrospect, our consensus primer (primer 3) was found to have three mismatches with the authentic gene sequence. However, annealing of the primer was sufficient to amplify a DNA fragment of ~380 bp, which upon sequencing was found to encode the appropriate N terminus for the peptide. This sequence was then used to design further pairs of specific oligonucleotide primers, enabling a similar application of one-sided PCR to derive fragments containing the 3′ portion of the gene.

**The AST Gene Sequence.** The sequence of the AST cDNA was derived by direct sequencing of PCR-amplified material on both strands. The entire sequence produced by this method is presented in Fig. 2. A variant of this sequence was also found in the cDNA (data not shown), which diverges from that shown in Fig. 2 at nt 1669. The alternative species is slightly shorter (1900 bp) and differs only in the extreme 3′ region, beyond nt 1669.

Analysis of the cDNA sequence reveals a single large open reading frame beginning at residue 96 and ending at residue 1206 (Fig. 2), which encodes an apparent precursor polypeptide of 370 amino acids. This pre-pro-AST is predicted to be 41.5 kDa in size, beginning with a methionine codon at nt position 96, which is positioned in a suboptimal translational context (23). Several potential sites for cleavage of a signal sequence are found clustered between amino acids 24 and 28 (24). The remainder of the open reading frame predicts a complex set of at least 13 AST-type peptides, each of which is bounded by potential dibasic endoproteolytic cleavage sites and ends with a glycine, which is the signal for C-terminal amidation by peptidyl-glycine α-amidating monooxygenase (PAM) (25) (Fig. 3). Seven of these potential peptides have previously been purified, providing evidence that these cleavage sites are used in vivo. The peptide sequences are displayed in Table 1 in the order in which they appear in the gene sequence and have been numbered accordingly. We believe this to be the most practical convention for future AST nomenclature.

All but one of the peptides includes the C-terminal amino acid sequence Phe-Gly-Leu, whereas 10 of the group are further alike in encoding tyrosine in the fifth position from the terminus (Table 1). The last of these peptides, AST-13, is unique in having isoleucine as the C-terminal residue. The N-terminal regions of the predicted peptides are rather more variable, suggesting the likelihood that address sequences for individual targeting, function, or degradation rate may be located here.

An unusual feature of the organization of the AST peptides in the precursor is their tendency to be clustered in groups, separated by spacer regions of extremely high acidity in the protein sequence (Fig. 3). Although such acidic regions have never been documented for an insect neuropeptide gene, a precedent has been found in the *Aplysia* FMRFamide gene (26). These regions have been postulated to serve a neutralization function, counterbalancing the high frequency of basic residues found in the neuropeptides and their accom-
Fig. 2. Nucleotide sequence of AST cDNA and deduced amino acid sequence of the peptide precursor. The numbering for each sequence is shown at right. The amino acid sequences for the 13 proposed AST peptides are shown in boldface type. Dibasic amino acid pairs are shaded, and glycine residues required for amiation are underlined. Potential polyadenylation signals are marked by thick bars. The most likely site for cleavage of a signal sequence is indicated by stars.

panying processing sites. In the AST precursor, the total numbers of acidic and basic amino acid residues are equal.

The third and largest acidic spacer region displays additional interesting features. Within this region three potential dibasic cleavage sites are found, processing of which could produce two additional peptides of 12 and 8 amino acids. However, neither of these sequences includes the necessary glycine as an amidation signal, but both do contain isoluecine as their C-terminal amino acid, in common with AST-13. Searches of the GenBank and SwissProt data bases for homologous sequences using these two peptides produced no significant matches.

The predicted AST precursor terminates at a TAA stop codon at position 1206 (Fig. 2). The untranslated nucleotide sequence that follows is extremely A + T rich and extends for >700 bp. Potential signals for polyadenylation are located near the 3' end of this region.

Genomic DNA Analysis. Copy number of the AST-encoding gene was examined by Southern analysis of D. punctata genomic DNA. Fig. 4 shows hybridization of a fragment of cDNA spanning residues 75–658 of the AST gene to restriction digests of total DNA. Digestion with the 4-base recognition sequence enzymes Mse I and Hpa II showed hybridization with smaller fragments consistent with the sizes predicted from the cDNA sequence (lanes M and H). An Xho I digest (lane X) showed hybrids at 14.2 kb and 11.4 kb. A combination digest with EcoRI and Xho I together reduced the sizes of these fragments by ~2.6 kb each (lane E/X). A recognition site for EcoRI is found in the cDNA sequence at residue 1163 (Fig. 2). The combination of EcoRI and Sal I
produced a single hybrid of ∼1.9 kb. These results suggest that the gene encoding AST is present in a single copy in the haploid D. punctata genome and, further, that between the upstream Sal I and Xho I sites is an allelic variation in the population such that one allele contains an extra 1.9 kb.

AST Gene Expression in Diploptera Brain. In situ hybridization analysis using brains from day 5 mated female D. punctata showed that the AST mRNA is strongly expressed by four large medial cells in the pars intercerebralis of the protocerebrum (Fig. 5).

DISCUSSION

We have determined the sequence of a cDNA encoding pre-pro-AST in the cockroach D. punctata. Because our initial attempts to amplify the AST gene from brain cDNA libraries using PCR were inconclusive, we chose, instead, to make use of double-stranded cDNA as the source material. As a result, sequence data were obtained directly from PCR-amplified DNA. To eliminate the risk of Taq-generated sequence errors, data were collected from both strands of the DNA, using independently amplified templates.

The occurrence of two products in PCR amplifications of the 3' end of the AST cDNA suggests that the AST mRNA might undergo alternative splicing. The affected region is contained entirely within the 3' untranslated sequence downstream of nt 1669 (Fig. 2). Both of the possible 3' combinations are extremely A + T rich. Highly A + U-rich sequences in the 3' regions of some mRNAs have been shown to affect message stability (27). It is, therefore, tempting to speculate that important information regulating relative mRNA degradation rates resides in these different 3' arrangements of the AST mRNA.

The deduced sequence of the AST precursor contains all the appropriate signals for an exported protein capable of being processed into 13 amidated AST-type peptides. The peptide sequences confirm the structures of seven previously isolated ASTs and predict at least six new candidates. These potential ASTs are largely consistent with the basic structural-core motif for AST activity of Tyr-Xaa-Phe-Gly-Leu-NH₂, with two notable exceptions. (i) In AST's 10-12, phenylalanine is substituted for tyrosine in the fifth position from the C terminus; the consequences of this alteration on activity remain to be studied. (ii) The final peptide (AST-13) contains isoleucine in place of leucine at the C terminus. This change may, in fact, have a relatively limited impact on the effectiveness of the peptide in inhibiting JH production. An analogous peptide variant of AST-2 (with isoleucine at the C terminus) was reported by Pratt et al. (4) to inhibit JH production in CA from 10-day-old female D. punctata to a similar degree as the authentic product ending with leucine.

Curiously, both potential peptides located within the third acidic spacer region also have C-terminal isoleucine residues. However, neither of these peptides would be expected to be amidated. Previous studies using synthetic ASTs modeled on isolated peptides have found amidation an absolute requirement for significant JH inhibition (ref. 7 and S.S.T., unpublished work). Making use of anti-AST antibodies, ASTs have been identified in nerves serving other insect tissues, particularly those projecting to the pulsatile organ muscle in D. punctata (10). This result suggests that other functions for this family of bioactive peptides will most probably be found. These two nonamidated AST candidates should prove useful in the search for such functions.

In D. punctata, retrograde filling of axons from the CA with nickel chloride (28) demonstrated that CA neurons are innervated by medial cells in the pars intercerebralis and lateral cells in the pars lateralis. Medial cells in the brain extend to the CA through the nervi corporis cardiace I and lateral cells extend to the CA through the paired nervi

Table 1. D. punctata AST peptides deduced from cDNA

<table>
<thead>
<tr>
<th>AST</th>
<th>Previous designation*</th>
<th>Primary structure</th>
<th>Isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASB2</td>
<td>L Y D</td>
<td>†</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>V Y S</td>
<td>†</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>K R L</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>P V Y</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>N P F</td>
<td>†</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>G L</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>S K M</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Y G F</td>
<td>†</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>G S L</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>G D G</td>
<td>†</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>L A F</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>F F P</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>F G L</td>
<td></td>
</tr>
</tbody>
</table>

*Previous designations are from refs. 3-5.

†AST peptides that have been isolated.

Fig. 4. Southern hybridization analysis of total cellular DNA (10 μg per lane) using a 32P-labeled fragment of the AST gene (nt 76–658 in Fig. 2). Genomic DNA samples were digested with various restriction enzymes (M, Mse I; H, Hpa II; E, EcoRI; S, Sal I; X, Xho I). DNA size markers (bases) are indicated at right.
corporis cardiaci II. Destruction of either medial or lateral cell regions or the axons that extend from them (29) relieves the inhibition of JH biosynthesis imposed on the CA of virgin females, thus suggesting that ASTs are produced from both cell types. Recently, the immunoreactivity of the D. punctata brain was described by using monoclonal antibody raised to AST-7 (30). Lateral neurosecretory cells and associated nervi corporis cardiaci II are the major source of AST immunoreactivity extending to the CA. However, in agreement with our in situ hybridization observations, four medial cells are strongly immunopositive. These immunopositive medial cells terminate within the protocerebrum in areas where lateral cells form arborizations, rather than extending to the CA. Although AST mRNA expression was not evident in lateral cells of the brain by whole-mount in situ hybridization, preliminary studies using brain sections have indicated that cells within the lateral and tritocerebral regions also express AST mRNA at low levels compared with the four positive medial cells (unpublished work).

We thank Tian Yu for dissections and Tina Inalsingh for sequencing support. This work has been funded in part by Insect Biotech Canada, one of 15 Networks of Centres of Excellence supported by the government of Canada and by National Sciences and Engineering Research Council Grants OGP0036481 (W.G.B.) and A9407 (S.S.T.).


**Fig. 5.** Cellular localization of expression of the D. punctata AST gene in whole (day 5 mated) female brains (frontal view). Four strongly hybridizing medial neurosecretory cells were found in all brains examined. (Bar = 40 μm.)