0270-6474/85/0507-1872$02.00/0

The Journal of Neuroscience
Vol. 5, No. 7, pp. 1872-1880
July 1985

Structure and Expression of the Egg-laying Hormone Gene Family in *Aplysia*[1]

ANNE C. MAHON, JOHN R. NAMBU, RONALD TAUSSIG, MALLADI SHYAMALA, ARTHUR ROACH, AND RICHARD H. SCHELLER[2]

Department of Biological Sciences, Stanford University, Stanford California 94305

Abstract

Transcription of the egg-laying hormone (ELH) gene family was examined by characterizing homologous cDNA clones from abdominal ganglion and atrial gland cDNA libraries. All cDNAs contain an exon that spans the coding region (exon III) and one or two additional exons. The tissue-specific expression of the ELH gene family was confirmed by the observation that exon III encodes the ELH precursor protein in the bag cell transcripts and either the A or B precursor proteins in the atrial gland transcripts. The cDNA clones also contain 5′ untranslated exons not present in the previously isolated genomic clones. One type of transcript has a 40-base pair segment, designated exon I, contiguous with exon III. A second type of transcript has an additional 149 base pairs of DNA, designated exon II, located between exons I and III. Several genomic clones containing exons I and II were isolated. DNA sequence analysis reveals that exons I and II are directly linked and that they are separated from exon III by an intervening sequence of at least 5 kilobases (kb). Consensus sequences for a putative promotor region and also for RNA splicing and polyadenylation were identified. From this work we can describe a prototype ELH gene complete with identified sequences necessary for the proper initiation of transcription and the subsequent processing of the transcript.

Egg laying in *Aplysia* is a stereotyped behavior goverened by a family of neuropeptides (Arch et al., 1978; Strumwasser et al., 1980; Scheller et al., 1983a). The A and B peptides are secreted by an exocrine tissue, the atrial gland (Arch et al., 1978; Heller et al., 1980; Schlesinger et al., 1981). *In vitro*, the A and B peptides cause depolarization of two groups of 400 electrically coupled neurons situated on the rostral margin of the abdominal ganglion (Blankenship, 1980; Strumwasser et al., 1980). These "bag cells" fire tonically for about 30 min, releasing a battery of peptides, including egg-laying hormone (ELH), into the vascularized connective tissue sheath which surrounds the ganglion (Frazier et al., 1967; Arch, 1972; Loh et al., 1975; Arch et al., 1978; Strumwasser et al., 1980). The bag cell peptides alter the electrical excitability of central neurons (Mayeri, 1979; Stuart and Strumwasser, 1980; Scheller et al., 1983b) and at the same time travel through the circulatory system to distant targets such as the ovotestis (Kupfermann, 1970; Coggeshall, 1970; Dudek and Tobe, 1978; Rothman et al., 1983). The combined hormonal, modulatory, and transmitter activities of these biologically active peptides are thought to generate the egg laying fixed action pattern.

The peptides involved in egg laying are encoded by a small gene family (Scheller et al., 1982, 1983a). Three of the approximately five members of the ELH gene family have been cloned, isolated, and characterized (Scheller et al., 1983a). These distinct but highly homologous genes include the linked ELH and peptide B genes (on clone ELH-1) and the peptide A gene (on clone ELH-18). In situ and *in vitro* hybridization studies reveal that the ELH gene family is expressed in the bag cells, the atrial gland, and a small network of interneurons distributed throughout the CNS (McAllister et al., 1983). Furthermore, within these tissues expression is limited to different members of the ELH gene family. For example, the predominant polyadenosine (poly A) mRNA species in the bag cells is homologous to the ELH gene, whereas in the atrial gland transcripts homologous to the peptide A and B genes are present (Scheller et al., 1983a).

Each member of the ELH gene family encodes a distinct protein containing identified biologically active peptides flanked by proteolytic cleavage sites. Although the genes are more than 90% homologous at the DNA level, divergence between sequences has resulted in significant changes in the amino acids which signal processing. Thus, different sets of peptides are predicted to be released from the precursors. Most significantly, although all precursors contain regions of both A/B peptide-like homology and ELH-like homology, each precursor releases a different one of these three peptides (Scheller et al., 1983a).

Several discrete RNA species homologous to the ELH genomic clones are present in the bag cells and the atrial gland (Scheller et al., 1982). These may be transcripts from distinct members of the ELH gene family or, alternatively, they may represent alternately processed RNAs following transcription of a single gene. To characterize these transcripts we have isolated cDNA clones complementary to the RNA present in the bag cells and the atrial gland. The structural analysis of the cDNA clones and of the genes encoding the transcripts is presented.

Materials and Methods

Isolation of ELH clones. ELH cDNA clones were isolated from two different cDNA libraries representing transcripts present in the abdominal ganglion and the atrial gland. The abdominal ganglion library, constructed in the phage vector λgt10, was generously provided by Michael Palazolo, James Schwartz, and Richard Axel. The atrial gland cDNA library was constructed in the plasmid pBR322 (Rowekamp and Firtel, 1980). Approximatley 100,000 phage plaques (abdominal library) or 10,000 bacterial colonies (atrial library) were absorbed to nitrocellulose filters. Two radiolabeled probes were employed to screen the filters following the procedure of Benton and Davis (1977). The probes were generated by restriction enzyme digestion of the

Received August 29, 1984; Revised December 3, 1984;
Accepted December 4, 1984

[1] We would like to thank Felix Strumwasser for the atrial gland cDNA library and Michael Palazolo, James Schwartz, and Richard Axel for the abdominal ganglion cDNA library. This work is supported by National Institutes of Health Grant 1 RO1 NS19542-01 to R. H. S. A. C. M. is a National Institutes of Health Postdoctoral Fellow.

[2] To whom correspondence should be addressed.

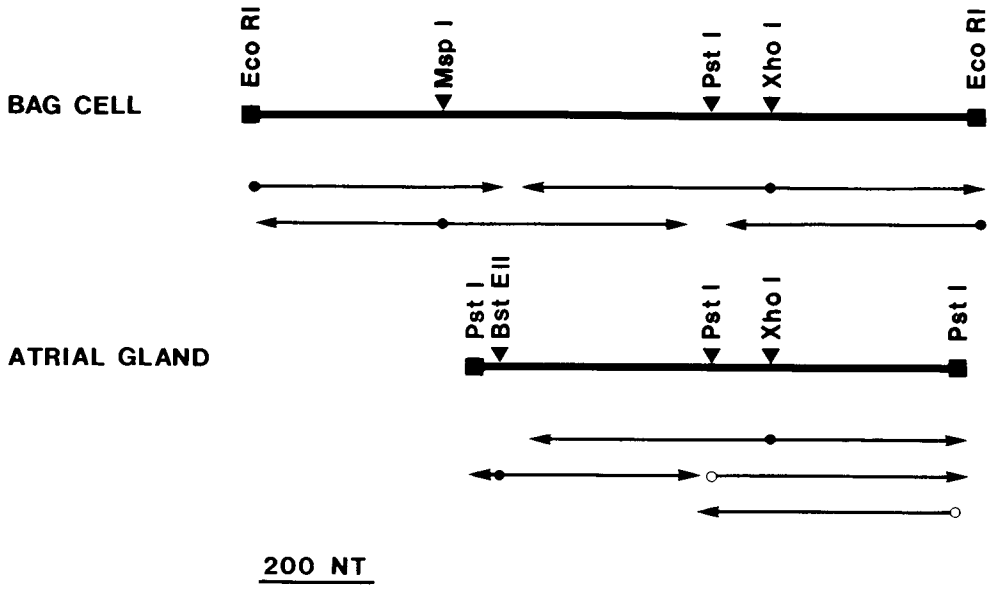

Figure 1. Strategy for sequencing the atrial gland and abdominal ganglion cDNA clones. The DNA sequence was determined from the indicated restriction enzyme sites using either the chain termination method of Sanger et al. (1977) (*open circles*) or the chemical modification and cleavage method of Maxam and Gilbert (1980) (*solid circles*). *Arrows* indicate the direction and length of the sequence determined.

genomic clone ELH-1 to yield a 1 kb Pst I fragment and a 470 base pair Pst I/Hinc II fragment containing the 3′ and 5′ ends of the ELH gene, respectively. These were labeled with ^{32}P by nick-translation (Rigby et al., 1977) to a specific activity of 10^7 to 10^8 cpm/μg. A heat- or base-denatured probe (2 × 10^6 cpm) was incubated with the filters for 18 hr at 65°C in a solution of 5× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 10 mM Tris (pH 7.5), 1mM EDTA (pH 8.0), 0.5% SDS, 5× Denhardt's (1× Denhardt's is 0.02% each of bovine serum albumin, Ficoll, and polyvinyl pyrolidone), and 10 μg/ml of sheared salmon sperm DNA. The filters were washed in several changes of 0.5× SSC and 0.05% SDS at 65°C for 3 to 4 hr, air dried, and autoradiographed for 24 to 72 hr at −70°C. Both probes yielded numerous positive plaques or colonies from both libraries. These were isolated and purified by rescreening at a lower phage or colony density.

ELH genomic clones were isolated from two different *Aplysia* genomic libraries. The construction of the first library, using the phage lambda cloning vector Charon 4, has been described previously (Scheller et al., 1982). The second library was similarly constructed using the phage cloning vector λJ1 (Mullins et al., 1984) and size-selected fractions of Mbo I partially digested DNA. Approximately 400,000 recombinants were screened using ^{32}P-labeled ELH cDNA clone J restriction fragments as probes. These nick-translated probes contained either exon I, exon I and II, or exon III sequences. Hybridizations were performed using the same solution as above and were carried out for 16 to 20 hr at 60 to 65°C with 2 × 10^6 to 1 × 10^7 cpm of probe. Several clones from the Charon 4 library hybridized only to exon I or exon I/II probes. The λJ1 library yielded clones that hybridized both to exon III and exon I/II probes. These clones were purified by rescreening at lower densities with the same probes.

Restriction endonuclease mapping. The positions of various restriction enzyme cleavage sites were determined by size analysis of single, double, and partial digestion products which were fractionated using agarose or polyacrylamide gel electrophoresis.

DNA sequencing. The atrial gland and abdominal ganglion cDNA insert segments and regions of ELH genomic clones containing 5′-exon sequences were subcloned into the plasmid pBR322 and transformed into *Escherichia coli* strain HB 101 using the $CaCl_2$ method (Maniatis et al., 1982). Recombinant plasmid DNA was prepared according to the method of Maniatis et al. (1982) and used as substrate for DNA sequence analysis. Subclones for all cDNA clones and genomic clones 16, 17, and 18 were digested with the appropriate restriction endonuclease, dephosphorylated with calf-intestinal alkaline phosphatase (Boehringer Mannheim), and 5′-end labeled with [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (Bethesda Research Laboratories). Labeled DNA was then cleaved with a second restriction enzyme to produce asymmetrically labeled fragments, and the products were isolated from 6 to 10% polyacrylamide gels. Direct nucleotide sequence analysis was then performed using the chemical modification and cleavage procedure of Maxam and Gilbert (1980). The products were separated on 0.35-mm-thick 20% and 6% polyacrylamide gels (Sanger and Coulson, 1978) and analyzed via autoradiography. In some cases DNA restriction fragments were subcloned into the phage M13 and were sequenced following the chain termination method of Sanger et al. (1977). The strategies employed for sequencing the cDNA clones and genomic clones are presented in Figures 1 and 5, respectively.

Genome blots. Sperm was extracted from the small hermaphroditic duct of a single *Aplysia californica* and incubated in 10 vol of 0.25 M EDTA (pH 8.0), 0.5% SDS, and 250 μg/ml of proteinase K (Sigma) for 3 hr at 65°C. Genomic DNA was isolated via sequential extraction of the digestion mixture with equal volumes of phenol, phenol/chloroform, and chloroform, followed by ethanol precipitation at room temperature. The purified DNA was resuspended in 10 mM Tris (pH 7.5), 1 mM EDTA (pH 8.0). Ten micrograms of DNA were then digested to completion with Eco RI, Hind III, or Pst I restriction enzymes. The products were separated by agarose gel electrophoresis and transferred to nitrocellulose filters (Schleicher and Schuell BA 85) via the technique of Southern (1975).

Restriction fragments generated from ELH cDNA clone J were utilized as probes in hybridization analysis; the 135-base pair Eco RI/Hinc II 5′-fragment and the 1.0-kilobase (kb) Hinc II/Eco RI coding fragment were ^{32}P labeled via nick-translation to a specific activity of 5 × 10^6 to 1 × 10^9 cpm/μg. Heat-denatured probe (2 × 10^6 cpm) was incubated with the filters for 18 hr at 65°C in the hybridization solution described above. The filters were washed in several changes of 0.5× SSC and 0.05% SDS at 65°C for 3 to 4 hr, air dried, and autoradiographed for 24 to 72 hr at −70°C.

Results

Analysis of abdominal ganglion and atrial gland cDNA clones. To isolate cDNA clones complementary to the ELH gene family transcripts, the abdominal ganglion and atrial gland cDNA libraries were screened using two radiolabeled DNA segments from the 5′ and 3′ regions of the ELH gene. Clones which hybridized to both probes were thought to represent close to full length transcripts. Approximately 12 clones from each library were isolated and the lengths of their cDNA inserts were determined. The clones containing the longest cDNA inserts were mapped using restriction endonucleases and their nucleotide sequences determined according to the scheme shown in Figure 1.

The nucleotide sequences of four representative abdominal ganglion cDNA clones are shown in Figure 2. The abdominal ganglion library was constructed with poly A mRNA from abdominal ganglia including the bag cell clusters. *In situ* hybridization studies have shown that greater than 90% of the ELH gene family expression in these tissues occurs in the bag cells (McAllister et al., 1983). Furthermore, under stringent conditions, RNA from this tissue preferentially hybridizes to the ELH gene on genomic clone ELH-1 (Scheller et al., 1983a). These data suggest that the bag cell clones represent mRNA transcribed from the ELH gene. For comparison, the nucleotide sequence of the ELH gene is indicated above the cDNA nucleotide sequences.

The cDNA inserts range from approximately 1.1 to 1.2 kb and

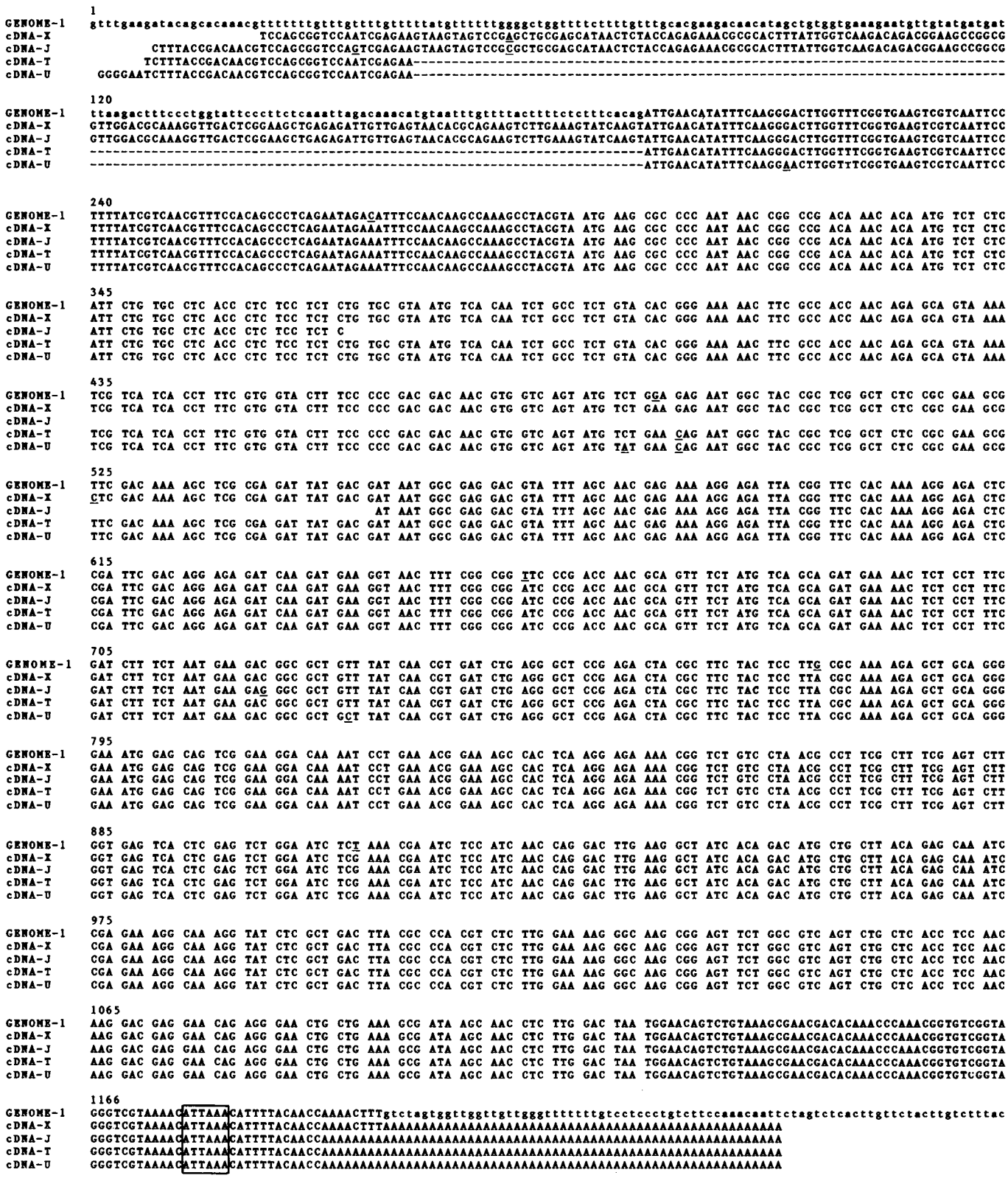

Figure 2. Nucleotide sequence of abdominal ganglion cDNA clones. The nucleotide sequences of four abdominal cDNA clones are presented beneath the sequence of genomic clone ELH-1. Regions homologous to the cDNA clones in the genomic clone are presented in *capital letters*. Regions of divergence are depicted in *lower case letters* in ELH-1. The *dashed lines* in cDNA clones *T* and *U* represent 149 bases absent in these clones but present in clones *J* and *X*. Single base changes are *underlined*. The first in-phase methionine codon (*ATG* at position 303) initiates the precursor protein and is followed by a 1119-nucleotide open reading frame. In the 3′ noncoding region the potential polyadenylation signal is surrounded by a *box*.

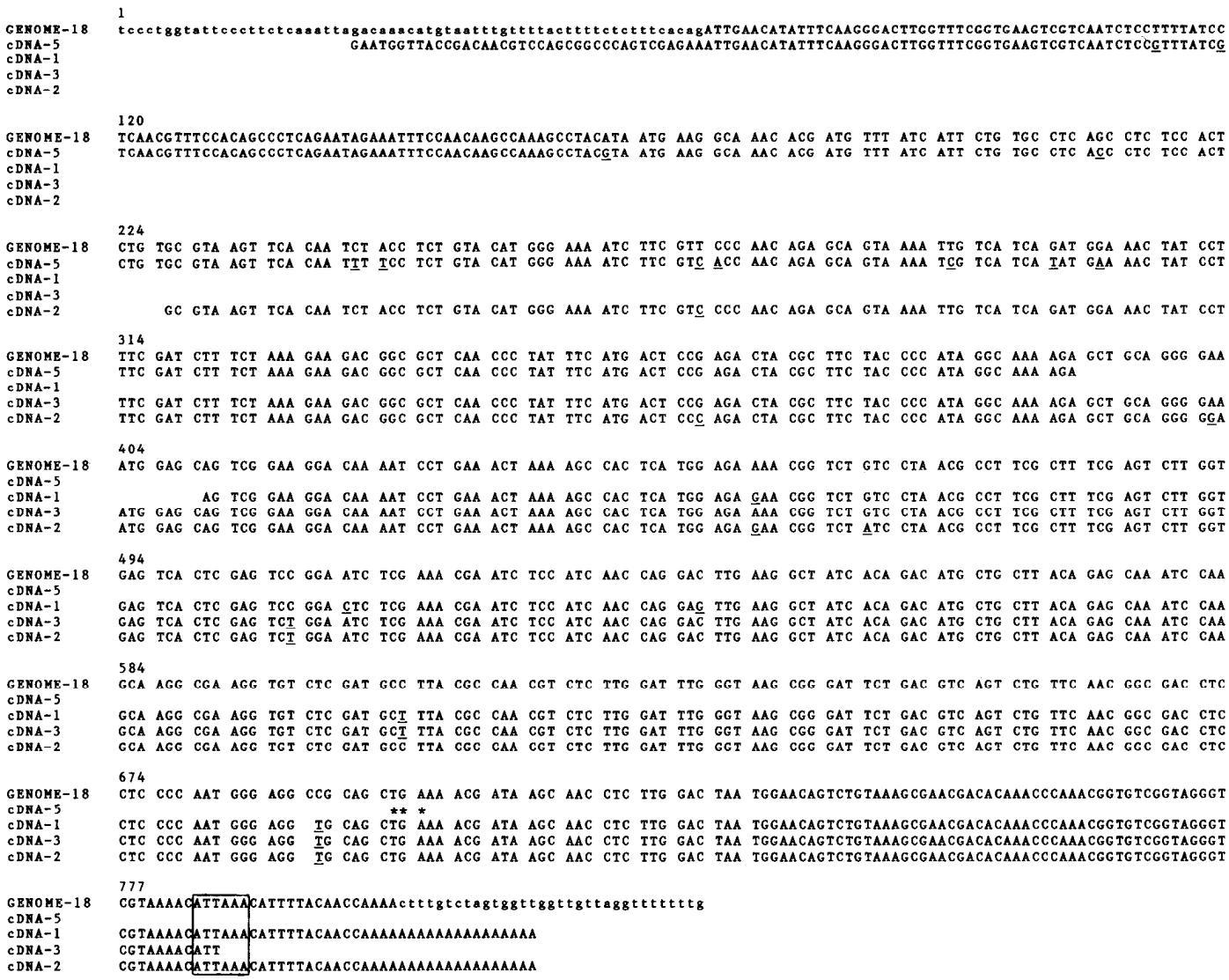

Figure 3. Nucleotide sequence of atrial gland cDNA clones. The nucleotide sequences of four atrial gland cDNA clones are presented beneath the sequence of genomic clone ELH-18. Regions of homology in the cDNA and genomic clones are in *capital letters*; regions where they diverge are in *lower case letters* in the genomic sequences. Single base changes are *underlined*. The initiator methionine codon is present at position 175 and is followed by a 548-basepair open reading frame in the genomic clone. A single base deletion in the cDNA clones results in the termination of the open reading frame after 519 nucleotides. The stop condon in this shifted frame is designated by the *asterisks*. The potential polyadenylation signal is surrounded by a *box*.

contain a core region of about 1 kb homologous to the previously characterized portion of the ELH gene (Scheller et al., 1983a). The genomic and cDNA clones diverge at the site of the poly A tail addition on the cDNAs. A nucleotide sequence common to all genes which give rise to polyadenylated transcripts, AUUAAA, is located at position 1178 (Proudfoot and Brownlee, 1976). The poly A tail is added either 13 (clones J, T, and U) or 21 (clone X) nucleotides after the AUUAAA sequence. The total length of the 3′ untranslated region (80 nucleotides) is short compared to the other neuropeptide genes we have studied (Nambu et al., 1983; Taussig et al., 1984).

The nucleotide sequence of the coding region is almost identical to the genomic sequence determined previously (Scheller et al., 1983a). The precise site of translation initiation is defined by the first methionine in the transcripts. Single base changes between the cDNA and genomic sequences (*underlined* in Fig. 2) often give rise to amino acid substitutions; however, none are noted in the regions encoding known peptides. Most of the differences are probably the result of polymorphisms. Similar studies on other peptide genes have shown that in some cases differential RNA processing can give rise to unique precursor proteins (Amara et al., 1982). We find no evidence of this and conclude that a single precursor protein is the prominent species in the bag cells of adult *Aplysia*.

The homology between the cDNA and genomic DNA sequences is conserved upstream of the initiator methionine for 110 nucleotides, after which, complete divergence occurs. The four clones shown in Figure 2 are present in either of two configurations in the 5′ untranslated region. These two configurations differ by the presence (clones X and J) or absence (clones T and U) of 149 nucleotides. Examination of the nucleotide sequences at the junctions between homologous and nonhomologous regions suggests that the ELH genes contain a previously unidentified 5′ sequence which is separated from the main body of the coding region by one or more introns. For simplicity of discussion we have defined exon I as the 5′-most region present in all cDNA clones but not in the genomic clone ELH-1. We defined exon II as the 149 nucleotides unique to clones X and J, and exon III as the remaining sequences which include the coding region.

The nucleotide sequences of four atrial gland cDNA clones are shown in Figure 3 along with the nucleotide sequence of the A gene from genomic clone ELH-18 (Scheller et al., 1983a). The atrial gland cDNA library does not contain full length clones; however, the nucleotide sequence of the four cDNAs presented reveals the salient features of these transcripts. The atrial and abdominal cDNAs share

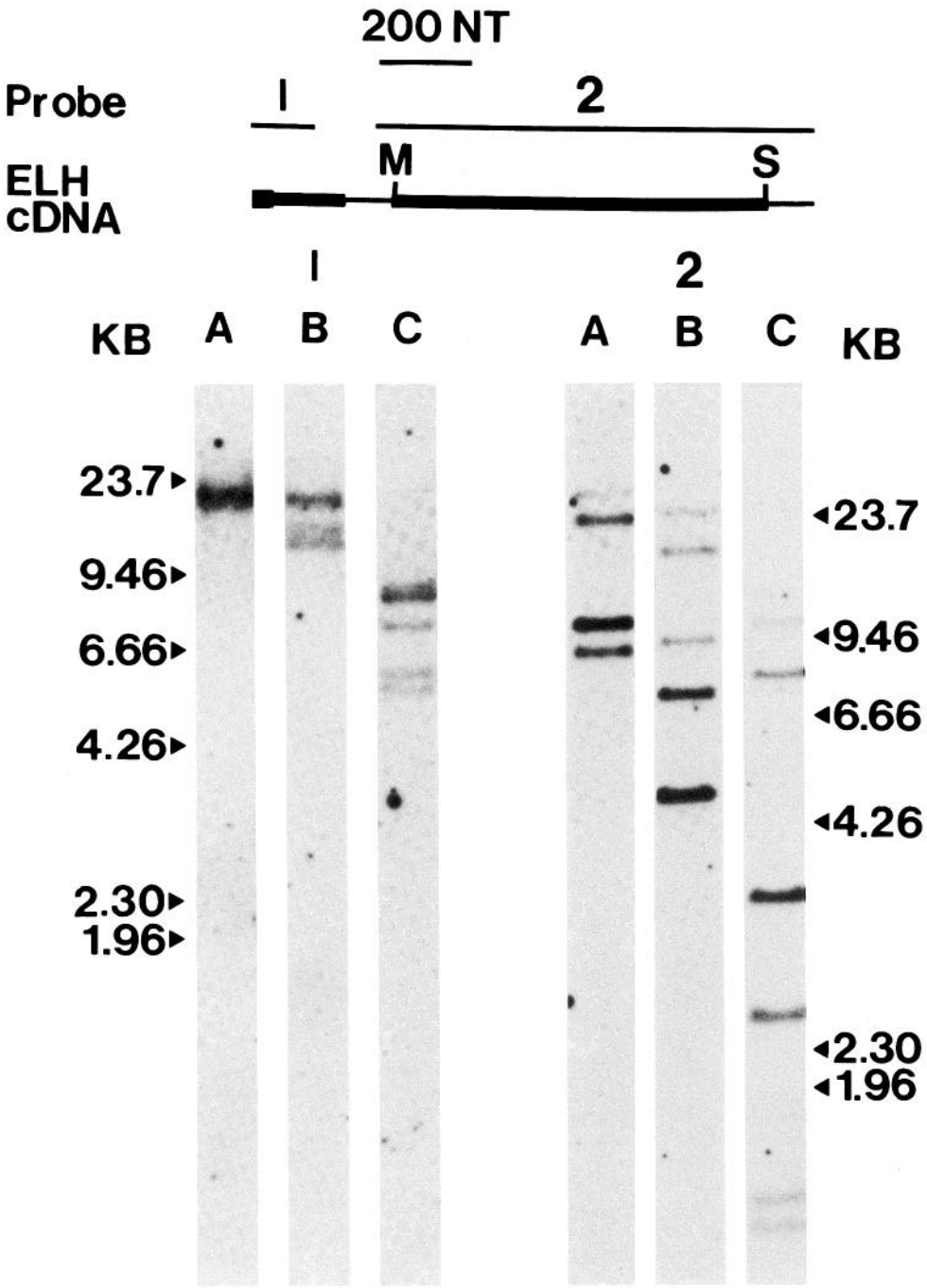

Figure 4. Genomic Southern blot analysis of ELH genes. Aliquots of genomic DNA from a single *Aplysia* specimen were digested to completion with the restriction enzymes Eco RI (*A*), Hind III (*B*), or Pst I (*C*). Ten micrograms of these digestion products were electrophoresed on 0.7% agarose gels and transferred to nitrocellulose filters. Two restriction fragments derived from ELH cDNA clone J were then radiolabeled with ^{32}P via nick-translation and used as probes in hybridization analysis. *Probe 1* consists of a 130-base pair EcoRI/Hinc II fragment and contains exons I and II. *Probe 2* is a 1.1-kb Hinc II/Eco RI fragment which contains only exon III. The positions of fragments 1 and 2 relative to ELH cDNA clone J and the resulting Southern blot pattern are shown. *Column 1*, genomic DNA blot using probe 1. *Column 2*, Genomic DNA blot using probe 2.

several common structural features. All have the same organization in the 3′ untranslated region and lose homology with their respective genes at precisely the same position in their 5′ ends. The leader sequence of the only atrial cDNA clone isolated that has this region is homologous to exon I in the bag cell transcripts. The rest of the atrial cDNA is homologous to the genomic DNA up to the site of the poly A tail. These results suggest that, like the abdominal ganglion transcripts, the atrial transcripts undergo RNA splicing 5′ to, but not within, the coding region.

Although the coding regions of the atrial clones are homologous to the previously characterized genomic clones, several significant differences between the gene and the transcripts have been observed. The three clones which extend far enough into the 3′ region all have a deletion of nucleotide 699 followed by a transition of nucleotide 700 from a C to a T. This causes a frame shift which gives rise to a cysteine codon immediately after the deletion and a stop codon further downstream. The cysteine residue may form a disulfide bond with the previously identified cysteine residue in the ELH region of the precursor (Figs. 3 and 8).

Peptides A and B are each 34 amino acids and differ from each other by four base changes which result in amino acid substitutions (Heller et al., 1980; Scheller et al., 1983a). Two of the clones encode peptide A (genomic clone ELH-18 and cDNA clone 2), one clone encodes peptide B (genomic clone ELH-1), and the other clone encodes a previously uncharacterized peptide (which we call peptide C) that is homologous to peptide B at three positions and peptide A at one position (cDNA clone 5). Other base changes do not occur in regions of the gene which encode known peptides.

Genomic organization of the ELH genes. Previous studies characterized three genomic clones containing exon III regions homologous to the bag cell and atrial transcripts (Scheller et al., 1983a). To complete the gene structure the genomic representation of exons I

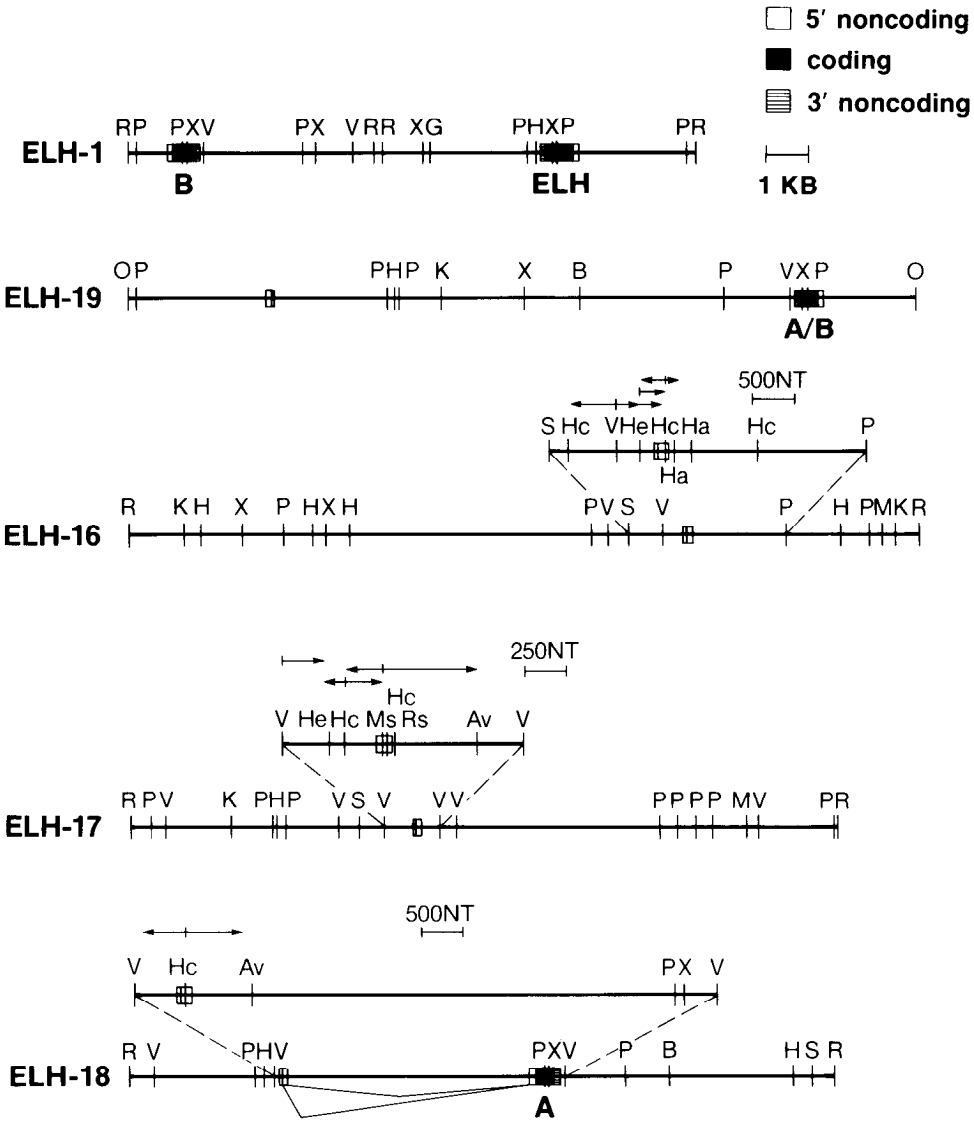

Figure 5. Organization of ELH genomic clones. The positions of restriction enzyme cleavage sites were deduced using single, double, and partial digests. The restriction maps for ELH-1 and ELH-18 have been presented previously (Scheller et al., 1983a). ELH-1 contains exon III regions for both an ELH and a B peptide gene linked in opposite orientations. ELH-18 contains exons I, II, and III from an A peptide gene. Both ELH-16 and ELH-17 possess only the 5′-exons I and II, while ELH-19 has exon III from a peptide A/B-like gene along with noncorresponding 5′-exon sequences. The direction of transcription for the genes was determined via DNA sequencing and restriction map analysis. Two alternate patterns for RNA splicing are presented for the A peptide gene of ELH-18. The strategy employed in sequencing 5′-exons I and II is indicated by *arrows*. *Av*, Ava II; *B*, Bam HI; *G*, Bg1 II; *H*, Hind III; *Ha*, Hha I; *Hc*, Hinc II, *He*, Hae III; *K*, Kpn I; *M*, Sma I, *Ms*, Msp I; *O*, Mbo I, *P*, Pst I; *R*, Eco RI; *Rs*, Rsa I; *S*, Sa1 I; *V*, Pvu II; *X*, Xho I.

and II was investigated. Southern blot analysis was performed on germ line DNA from an individual *Aplysia* using two probes representing exons I and II (probe 1) or exon III (probe 2) (Fig. 4). Genomic DNA was cut with three restriction enzymes and size fractionated on an agarose gel. Two enzymes, Eco RI and Hind III, do not cut within the cDNA transcripts or the coding region of the genomic clones. The other enzyme, Pst I, cuts one time within the transcribed regions. In all cases distinct hybridization patterns were observed with the two probes indicating that the untranslated and coding regions are sometimes on different restriction enzyme segments.

The results of this analysis are consistent with the proposal that there are five distinct ELH genes per haploid genome (Mahon and Scheller, 1983). In two of the six experiments, the resulting pattern produced five hybridizing fragments, and in no case, except where the enzyme is known to cleave within the gene, were more than five bands observed. Thus, the coding probe (2) generated five homologous fragments using Hind III cut DNA, and the 5′ probe revealed five hybridizing Pst I fragments. The observation of fewer than five bands in some cases is probably due to several similarly sized fragments containing exon sequences and/or to a single fragment containing more than one sequence.

In order to isolate the genomic DNA segments containing exons I and II, two *Aplysia* genomic libraries were screened. Several distinct clones were isolated from both a Charon 4 and a λJ1 library which contained ELH exons I, II, and III (Figure 5). In addition, ELH-18, the isolation of which has been described previously (Scheller et al., 1983a), was also found to contain ELH 5′ exon sequences. The identity, position, and direction of transcription of the ELH genes along the various clones were determined using restriction enzyme mapping, Southern blot hybridization, and DNA sequence analysis (Fig. 5). Two general categories of ELH genomic clones were identified based on the arrangement of the exons which they contained. The first contains a single region in the genomic DNA representing a portion of an ELH transcript. The second type of genomic clone has two of these regions linked together on the same piece of DNA. In these latter cases, the transcript regions are separated by at least 5 kb of DNA. The two Charon 4 clones, ELH-16 and ELH-17, are examples of genomic clones with a single unlinked transcribed region. Both possess only 5′-exon sequences and do not contain any coding region. In these clones and in all others characterized containing 5′ exon sequences, exons I and II are contiguous with no intervening sequences between them (Fig. 6). There are 5.5 and 10.0 kb of DNA downstream of the 5′-exons in ELH-16 and ELH-17, respectively, the minimum length of intervening sequence between these 5′-exons and any corresponding coding exons which might be present on a contiguous segment of genomic DNA.

Clones containing linked transcript regions could have two copies of the same exon or two distinct exons. ELH-1 has single copies of exon III from both an ELH and a B-peptide gene linked in opposite orientation. A distance of 8 kb separates the two coding regions, which are transcribed toward each other (from opposite strands of

```
            1
GENOME-16   CCCCTTTTTCCTTCCTTTTCCCTTAACAAATCCTGCCAAACCATTTCCCCTCTCTCTCTTTTTTTTCGCTTAACGACCTCCCCCCCCCCCTCCCCCCAC

            100
GENOME-16   TTTTTTTCATTCCCTTGAAACAAATCCTATCAAACCGATAATCCCTCCAGCTGCCGTATAAAAACCAATTCGCCCCCACCCCTTCCCCCAAAACTACCC
GENOME-17                                                                 GCCTTATAAAAACCAATTCGCCCCCACCCCTTCCCCCAAAACTACCC

            200
GENOME-16   CGAAACAAACTGTATCAAATGACCTTCCTCTACCAGTGTTACGCAAATGCACTCCCCCCCCC  TCACTCTAACCGCTTTTCCTCCCGCCCCCTCCACA
GENOME-17   CGAAACAAACTGTATCAAATGATCTCCCTCTACCAGTGTTACCCAACTGCAACCCCCCTCCCCTTCACT  AACCGTTTTTCCACCCACCCCCTCCACT

            300
GENOME-16   TCCCCCCCAAC     ACACACACACTTTTTCCATCCATCCCCTCCTCCACACCCTTCCCCCCCTCCCCCCCAAAGTGTCCACTCACCTTTAGAATGTA
GENOME-17   TGCCCCCCCCCCCCCCCACACACACTTTTTCCATCCATCCCCTCCTCCACACCCTTCTCCCCCCCCTCC   AAGTGTCCACTCACCTTTAGAATGTA
GENOME-18                                                                                         ACCTTTAGAATGTA

            400
GENOME-16   AGAGCCCGCGTTTGTGCCAGCAGGCCAGGCCAGGGTGACAGCGGACAGAGAG       ACAGAGAGAAAGAGAGAGAGAGGGACACACACACACCAC
GENOME-17   AGAGCCCGCGTTTGTGCCAGCAGGCCAGGCCAGGGTGACAGCGTGCAGAGAGATAGAGAGAAAGAGAG  AGAGAGGGACACCC    ACACACACC
GENOME-18   AGAGCCCGCGTTTGTGCTAG     CAAGCCAAGGTGACAGCGGG       CAGAGAGAAAGAGAGAGAGAGAGGGACACCC    ACACACACC

            500                      *********                                                     *******
GENOME-16   CACCACCACCACTACCAGCCCTCGTGAGGACAACGCATAAACCATCACACGTGCCCATAATTAACGGCTCCACGCCCTCTTGCTCGTGTATAAAACC A
GENOME-17    ACCACCACCACTACCAGCCCTCGTTAGGTCAACGCATAAACCATCACACCTGCCCATAATTAACGACTCCACGCCCTCTTGCTTGTGTATAAAACCTA
GENOME-18     CCACCACCACTACCAGCTCTCGTTAGGGCAACGCATAAACCATCGCACCTGCCCATAATTAACGACTCCACGCCCTCTTGCTTGTGTATAAAACCCA

                                                           EXON 1
            600
GENOME-16   GGGAAATCAGCCCCAGTCTCGGCATAAGTCTCGGGGGAATCTTTACCGACAACGTCCAGCGGTCCAA TCGAGAAGTAAGTAGTCCGAGCTGCGAGCAT
GENOME-17   GGGAAATCAGCCC AGTCTCGGCAGAACCCTCGGAGGAATCTTTACCGACAACGTCCAGCGG CCAAGTCGAGAAGTAAGTAGTCCGAGCTGCGAGCAT
GENOME-18   GGGAAATCAGCCC AGTCTCGGCAGAACTCTCGGGGGAATCTTTACCGACAACGTCCAGCG TCCCAGTCGAGAAGTAAGTAGTCCGAGCTGTGAGCAT

            700                                                EXON 2
GENOME-16   AACTCTACCAGAGAAACGCGCACTTTATTGGTCAAGACAGACGGAAGCCGGCGGTTGGACGCAAAGGTTGACTCGGAAGCTGAGAGATTGTTGAGTAAC
GENOME-17   AACTCTACCAGAGAAACGTGCGCTTTATTGGTCAAGATAGACGGAAGCCGGCGGTTGAAGGCAAAGGTTGACTCGGAAGCTGAGAGATTGTTGAGTAAC
GENOME-18   AACTCTACCAGAGAAACGCGCGCTTTATTGGTCAAGATAGACGGAAGCCGGCGGTTGAAGGCAAAGGTTGACTCGGAAGCTGAGAGATTGTTGAGTAAC

            800
GENOME-16   ACGCAGAAGTCTTGAAAGTATCAAGTGTGAGTACACTGTGACTTAAAGACAATATTTATCAAACCGAAACGTGTATTTAT  TTTTTAAATAGTTTTAG
GENOME-17   ACGCAGAAGTCTTGAAAGTATCAAGTGTGAGTACACTGTGACTTAAAGACAATATTTATCAAACCGAAACGTGTATTTAT  TTTTTAAA
GENOME-18   ACGCAGAAGTCTTGAAAGTATCAAGTGTGAGTACTCTGTGACTTAAAGACAACATTTATTAAACCGAAACGTGTATTTATATTTTTTAAATAGTTTTAG

            900
GENOME-16   TAGAGGTTTTACGGCA
GENOME-17
GENOME-18   TGGAGGTTTTACGGTGCTCACAACTGCATAAGGTCATACCCAGCCTGCACAAAGATGTGGACCCTACATGGGGCCAAACGGA
```

Figure 6. Nucleotide sequence of ELH exons I and II. The DNA sequence is presented for the 5′-exons as well as for flanking regions from ELH genomic clones 16, 17, and 18. Sequences were determined via the method of Maxam and Gilbert (1980) and are here aligned for maximum homology. *Underlined positions* indicate base pair differences, and *asterisks* above the sequence designate the putative TATA box (−40 basepairs) and CAAT box (−100 basepairs). Regions corresponding to exon I and II are delineated by *overhead bars* or *dots*, respectively. *Boxed GT* dinucleotides represent presumptive RNA splicing sites.

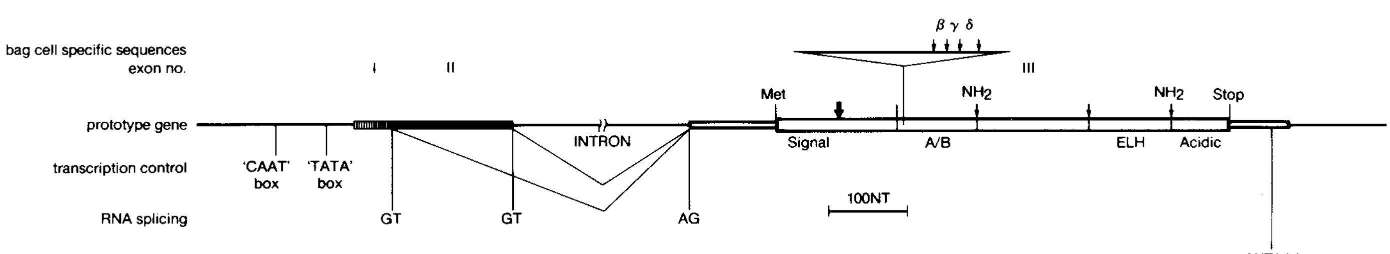

Figure 7. Schematic representation of ELH gene organization. Relative arrangement of ELH exons I, II, and III as well as possible RNA splicing pathways (including splice junction donor and acceptor sites) are depicted. Locations of the putative CAAT and TATA boxes and the polyadenylation consensus site are also shown. The neuropeptide encoding regions of both bag cell and atrial gland genes are presented and include the positions of the initiator methionine (*Met*), signal peptide cleavage (➡), proteolytic processing sites (↓), amidation signals (NH_2), and the stop codon. Bag cell-specific sequences include those specifying β-, γ-and δ-bag cell peptides.

DNA). The 5′-exons which complete the ELH and B gene transcripts are not present on ELH-1. Whether these regions are present on clones ELH-16, ELH-17, or on an unidentified genomic clone is currently under investigation. Two genomic clones have been characterized which have linked copies of different portions of the transcripts. The λJ1 clone, ELH-19, has both a single copy of 5′-exons and exon III for an A/B-like peptide gene. These regions are separated by more than 10 kb and represent different parts of two distinct transcripts since exons I and II are 3′ to the coding region. In contrast, we believe that genomic clone ELH-18, which has linked exon I, II, and III sequences, contains the complete transcription unit for an A gene. An intervening sequence of 5.3 kb separates the two regions, and the entire transcript spans at least 6.5 kb of genomic DNA. Based on the structures of the other clones, this distance probably represents a minimum size for members of the ELH gene family. Whether any size heterogeneity of the ELH genes has relevance to their expression is at present unknown.

Two genomic clones contain exons from distinct ELH genes, suggesting that many or even all of the members of the gene family are linked. This arrangement would be consistent with that of other

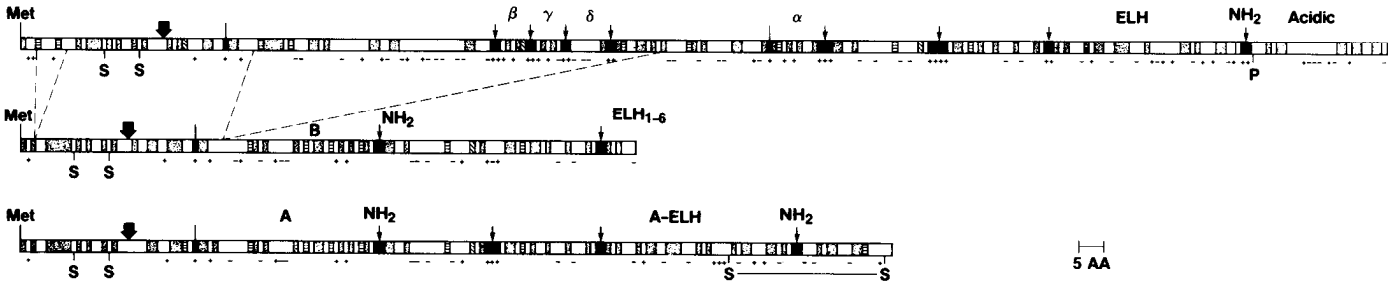

Figure 8. Neuropeptide precursor proteins encoding the ELH: A and B peptides. The positions of known peptide products including ELH, the A and B peptides, the acidic proteins, and α-, β-, γ-, and δ-bag cell peptides are indicated. Portions of the A and B peptide precursors corresponding to ELH encoding sequences are also given (ELH_{1-6}, A-ELH). The presumed signal sequence cleavage site is designated by a large arrow (➧); additional known or potential proteolytic processing sites occur at single arginine (|) and multiple basic (↓) residues. Other putative post-translational modifications include carboxy-terminal amidation (NH_2) and phosphorylation (P). Positions of charged amino acids (+/−), as well as hydrophobic residues (▨), histidines (▩), prolines (▤), and cysteines (S) are indicated.

multigene families (Kedes, 1979; Efstratiadis et al., 1980; Eberhardt et al., 1982) and further studies are currently being pursued to address this possibility.

In general, there is a good correlation between the sizes and numbers of the restriction fragments from the genome blots and from the genome clones that contain the 5′ and coding exons. It is not surprising that some discrepancies exist because the genomic libraries and the Southern blot studies used DNA from three different *Aplysia* subjects. A significant amount of DNA polymorphism is observed in these animals and related species (Britten et al., 1978; Scheller et al., 1982), making an exact correlation between animals difficult.

DNA sequence analysis. The positions of exons I and II on genomic clones ELH-16, ELH-17, and ELH-18 were determined by restriction mapping, hybridization experiments, and nucleotide sequencing (Fig. 6). It was discovered that all three of these clones contain both exons I and II (ELH-18 also contains exon III) and that these two exons are contiguous, with no intervening sequence between them. RNA-splicing consensus dinucleotides GT were found immediately following the 3′ ends of both exons I and II (Fig. 6, *boxes*). The corresponding AG acceptor dinucleotides are also present directly preceding exon III in the ELH, A, and B peptide genes. This is consistent with splicing pathways which could be used to generate both types of ELH cDNAs, those having exons I, II, and III (bag cell cDNAs J and X) and those having only exons I and III (bag cell cDNAs U and T and atrial gland cDNA 5) from a single gene type. We have not isolated genomic clones of other configurations such as that with exons I and III only or that in which exons I and II are not contiguous, suggesting that all RNAs arise through this mechanism.

The genomic sequences on clones ELH-16, ELH-17, and ELH-18 were quite homologous (94%) to each other throughout the transcribed region and were equally homologous to the corresponding regions of the cDNAs. In fact, the exon I and II sequence from ELH-16 exactly matches that of the bag cell cDNA clone X, suggesting that this is part of a transcribed ELH gene. The large degree of homology observed between both the bag cell and atrial transcripts in the 5′ noncoding region suggests that these sequences do not encode critical information for the tissue-dependent expression of the members of the ELH gene family. The significance of having two configurations of transcripts is unknown but is probably also unrelated to the tissue-dependent expression of the genes, since we recently identified cDNA clones containing exon II from an atrial gland cDNA library. It is possible that the transcripts have different stability conferred by exons I and II, or that one form of transcript predominates in a subset of cells.

The DNA sequence flanking the transcribed exons in the three genomic clones was also determined. All three contain a perfect TATAAAA box (Fig. 6, *asterisks*) centered at position −40 which is bounded on both sides by GC-rich regions. These sequences occur in roughly the same position in all eukaryotic genes which have been examined and are thought to be important in directing the specificity and efficiency of gene transcription (Goldberg, 1979; Breathnach and Chambon, 1981). At −100 basepairs upstream from exon I is a sequence bearing homology to another conserved region, the "CAAT" box (Fig. 6, *asterisks*). This second, less conserved consensus sequence is also proposed to be necessary for efficient transcription (Benoist et al., 1980; Breathnach and Chambon, 1981). Twenty-five nucleotides upstream from the proposed regulatory sequences, several insertion/deletion events lead to reduced homology between the three genes.

Discussion

From the data presented, we can define a prototype gene that represents all characterized members of the ELH gene family (Fig. 7). The gene consists of three exons which are separated by a single intervening sequence. Nucleotide sequences homologous to the consensus sequences for the proper initiation of transcription in eukaryotes (the "TATA" and "CAAT" boxes) are present upstream from exon I (Goldberg, 1979; Benoist et al., 1980; Breathnach and Chambon, 1981). We have also identified sequences which in all other eukaryotes are necessary for RNA processing. These include the GT donor and AT acceptor sequences which are present at RNA splice junctions (Lewin, 1980) and the sequence AUUAAA in the 3′ end of the gene which signals polyadenylation (Proudfoot and Brownlee, 1976). The major structural difference between the bag cell gene type and the atrial gland gene type is a 240-nucleotide insert in the A/B peptide coding region.

The ELH genomic clones exhibit a large degree of homology which extends for several kilobases past the transcribed regions around exon III (Scheller et al., 1982) and exons I and II. Such homology would be expected if the ELH gene family arose by duplication of a DNA segment which contained the complete transcription unit for a primordial ELH gene. It is striking that the homology between genes dramatically declines upstream from the TATA and CAAT sequences, a region which has been implicated in the tissue-specific control of gene expression in other systems. We suggested that the ELH genes diverged in the coding region following duplication of an ancestral gene which resulted in alternative proteolytic processing pathways to produce different sets of peptides from the precursors (Scheller et al., 1983a). Similarly, it is possible that changes in the DNA sequences upstream to transcription initiation may be important in the evolution of the tissue-specific expression of the gene family. A similar proposal has been made for the growth hormone gene family which displays similar structural characteristics (Eberhardt et al., 1982). The structural characterization of the cDNA and genomic clones presented here serves as a basis for eventual deletion analysis and transformation studies which may address these questions and provide insight into the evolution of the egg-laying behavior.

The peptide precursors encoded by the atrial and bag cell transcripts are shown in Figure 8. These proteins are very similar to those predicted from the DNA sequence analysis of the genomic clones which have been discussed previously (Scheller et al., 1983a). Analysis of the bag cell cDNAs identifies the initiation of translation, which was ambiguous in the genomic clones. The atrial cDNAs and the A gene differ in that the former encodes an additional cysteine residue in the carboxy-terminal end of the protein. It is likely that a disulfide linkage is formed between this residue which is on the acidic-like peptide and another cycteine residue present on the ELH-like peptide, a prediction supported by high pressure liquid chromatography data (B. Rothman and E. Mayeri, personal communication). The consequences of such a bridged molecule are unknown. The linkage could simply inactivate the the atrial ELH by blocking receptor binding, or alternatively, a new receptor with unique specificity for this hybrid molecule may have evolved. The present study confirms earlier hybridization experiments which suggested that the ELH gene is expressed preferentially in the bag cells, whereas the peptide A and B genes are expressed in the atrial gland (Scheller et al., 1983a).

In summary, in the present work we provide a structural analysis of the ELH gene family and demonstrate the tissue-dependent expression of these genes in the atrial gland and the bag cells. Furthermore, the precursor structure for the predominant neuropeptides present in these tissues is unambiguously defined. Finally, this work supports the proposal that the expression of unique genes in different tissues resulted from the duplication and divergence of a single ancestral gene and provides a starting point for further investigation into the genetic basis of the tissue-specific expression of the ELH gene family.

References

Amara, S. G., V. Jonas, M. G. Rosenfeld, E. J. Ong, and R. M. Evans (1982) Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature *298:* 240–244.

Arch, S. (1972) Biosynthesis of the egg-laying hormone (ELH) in the bag cell neurons of *Aplysia californica*. J. Gen Physiol. *60:* 1263–1278.

Arch, S., R. T. Smock, R. Gurvis, and C. McCarthy (1978) Atrial gland induction of the egg-laying response in *Aplysia californica*. J. Comp. Physiol. *128:* 67–70.

Benoist, C., K. O'Hare, R. Breathnach, and P. Chambon (1980) The ovalbumin gene: Sequence of putative control regions. Nucleic Acids Res. *8:* 127–142.

Benton, W. D., and R. W. Davis (1977) Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science *196:* 180.

Blankenship, J. E. (1980) Physiological properties of peptide-secreting neuroendocrine cells in the marine mollusc *Aplysia*. In *The Role of Peptides in Neuronal Function*, J. L. Barker and T. G. Smith, Jr., eds., pp. 160–187, Marcel Dekker, Inc., New York.

Breathnach, R., and P. Chambon (1981) Eucaryotic split genes. Annu. Rev. Biochem. *50:* 349–383.

Britten, R. J., A. Cetta, and E. H. Davidson (1978) The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell *15:* 1175–1186.

Coggeshall, R. E. (1970) A cytologic analysis of the bag cell control of egg laying in *Aplysia*. J. Morphol. *132:* 461–485.

Dudek, F. E., and S. S. Tobe (1978) Bag cell peptide acts directly on ovotestis of *Aplysia californica*: Basis for an *in vitro* bioassay. Gen. Comp. Endocrinol. *36:* 618–627.

Eberhardt, N. L., M. Selby, G. Cathala, M. Karin, A. Gutierrez-Hartmann, S. H. Mellon, N. C. Lan, D. Garner, and J. Baxter (1982) The growth hormone gene family: Structure, evolution, expression, and regulation. In *UCLA Symposia on Molecular and Cellular Biology*. Vol. XXVI: *Gene Regulation*, B. W. O'Malley, ed., pp. 235–251, Academic Press, New York.

Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot (1980) The structure and evolution of the human β-globin gene family. Cell *21:* 653–668.

Frazier, W. T., E. R. Kandel, I. Kupfermann, R. Waziri, and R. E. Coggeshall (1967) Morphological and functional properties of identified neurons in the abdominal ganglion of *Aplysia californica*. J. Neurophysiol *30:* 1288–1351.

Goldberg, M. L. (1979) Sequence analysis of Drosophila histone genes. Ph.D. thesis, Stanford University, Stanford, CA.

Heller, E., L. K. Kaczmarek, M. W. Hunkapiller, L. E. Hood, and F. Strumwasser (1980) Purification and primary sequence of two neuroactive peptides that cause bag cell afterdischarge and egg-laying in *Aplysia*. Proc. Natl. Acad. Sci. U. S. A. *77:* 2328–2332.

Kedes, L. H. (1979) Histone genes and histone messengers. Annu. Rev. Biochem. *48:* 837–870.

Kupfermann, I. (1970) Stimulation of egg-laying by extracts of neuroendocrine cells (bag cells) of abdominal ganglion of *Aplysia*. J. Neurophysiol. *33:* 877–881.

Lewin, B. (1980) Alternatives for splicing: Recognizing the ends of introns. Cell *22:* 324–326.

Loh, Y. P., Y. Sarne, and H. Gainer (1975) Heterogeneity of proteins synthesized, stored and released by the bag cells of *Aplysia californica*. J. Comp. Physiol. *100:* 283.

Mahon, A. C., and R. H. Scheller (1983) The molecular basis of a neuroendocrine fixed action pattern: Egg laying in *Aplysia*. Cold Spring Harbor Symp. Quant. Biol. *48:* 405–412.

Maniatis, T., E. F. Fritsch, and J. Sambrook (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., and W. Gilbert (1980) Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. *65:* 499–560.

Mayeri, E. (1979) Local hormonal modulation of neural activity in *Aplysia*. Fed. Proc. *38:* 2103–2108.

McAllister, L. B., R. H. Scheller, E. R. Kandel, and R. Axel (1983) *In situ* hybridization to study the origin and fate of identified neurons. Science *222:* 800–808.

Mullins, J. I., D. S. Brody, R. C. Binari and S. M. Cotter (1984) Viral transduction of c-myc gene in naturally occuring feline leukemias. Nature *308:* 856–858.

Nambu, J. R., R. Taussig, A. C. Mahon, and R. H. Scheller (1983) Gene isolation with cDNA probes from identified *Aplysia* neurons: Neuropeptide modulators of cardiovascular physiology. Cell *35:* 47–56.

Proudfoot, N. J., and G. G. Brownlee (1976) 3′ Non-coding region sequences in eukaryotic messeger RNA. Nature *263:* 211–214.

Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P Berg (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. *113:* 237.

Rothman, B. S., G. Weir, and F. F. Dudek (1983) Direct action of egg laying hormone on ovotestis of *Aplysia*. Gen. Comp. Endocrinol. *52:* 134–141.

Rowekamp, W., and R. A. Firtel (1980) Isolation of developmentally regulated genes from *Dictyostelium*. Dev. Biol. *79:* 409–412.

Sanger, F., and A. R. Coulson (1978) The use of thin acrylamide gels for DNA sequencing. FEBS Lett. *87:* 107–110.

Sanger, F., S. Nicklen, and A. R. Coulson (1977) Proc. Natl. Acad. Sci. U. S. A. *74:* 5463–5467.

Scheller, R. H., J. F. Jackson, L. B. McAllister, J. H. Schwartz, E. R. Kandel, and R. Axel (1982) A family of genes that codes for ELH, a neuropeptide eliciting a stereotyped pattern of behavior in *Aplysia*. Cell *28:* 707–719.

Scheller, R. H., J. F. Jackson, L. B. McAllister, B. S. Rothman, E. Mayeri, and R. Axel (1983a) A single gene encodes multiple neuropeptides mediating a stereotyped behavior. Cell *35:* 7–22.

Scheller, R. H., B. S. Rothman, and E. Mayeri (1983b) A single gene encodes multiple peptide transmitter candidates involved in a stereotyped behavior. Trends Neurosci. *6:* 340–345.

Schlesinger, D. H., S. P. Babirak, and J. E. Blankenship (1981) Primary structure of an egg-releasing peptide from the atrial gland of *Aplysia californica*. In *Symposium on Neurohypophyseal Peptide Hormones and Other Biologically Active Peptides*. D. H. Schlesinger, ed., pp. 137–150, Elsevier/North-Holland Biomedical Press, New York.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98:* 503–517.

Strumwasser, F., L. K. Kaczmarek, A. Y. Chiu, E. Heller, K. Jenning, and D. Viele (1980) Peptides controlling behavior in *Aplysia*. In *Peptides: Integrators of Cell and Tissue Function*, F. Bloom, ed., pp. 197–218, Raven Press, New York.

Stuart, D. K., and F. Strumwasser (1980) Neuronal sites of action of a neurosecretory peptide, egg-laying hormone, in *Aplysia californica*. J. Neurophysiol. *43:* 399–519.

Taussig, R., R. R. Kaldany, and R. H. Scheller (1984) A cDNA clone encoding neuropeptides isolated from *Aplysia* neuron L11. Proc. Natl. Acad. Sci. U. S. A., *81:* 4988–4992.