Molecular Characterization of Neurally Expressing Genes in the \textit{para} Sodium Channel Gene Cluster of Drosophila

Chang-Sook Hong\textsuperscript{1} and Barry Ganetzky

Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, Wisconsin 53706

Manuscript received August 30, 1995
Accepted for publication December 11, 1995

ABSTRACT
To elucidate the mechanisms regulating expression of \textit{para}, which encodes the major class of sodium channels in the Drosophila nervous system, we have tried to locate upstream \textit{cis}-acting regulatory elements by mapping the transcriptional start site and analyzing the region immediately upstream of \textit{para} in region 14D of the polytene chromosomes. From these studies, we have discovered that the region contains a cluster of neurally expressing genes. Here we report the molecular characterization of the genomic organization of the 14D region and the genes within this region, which are: calnexin (\textit{Cnx}), actin related protein 14D (\textit{Arp14D}), calcineurin A 14D (\textit{CnaA14D}), and chromosome associated protein (\textit{Cap}). The tight clustering of these genes, their neuronal expression patterns, and their potential functions related to expression, modulation, or regulation of sodium channels raise the possibility that these genes represent a functionally related group sharing some coordinate regulatory mechanism.

The distinctive signaling properties of individual neurons are primarily established by the specific types of ion channels they express. Voltage-sensitive sodium channels mediate the rapid phase of action potentials and therefore have a critical role in the generation and propagation of electrical signals in neurons (reviewed by Catterall 1993). Like other ion channels, sodium channel polypeptides are encoded by members of a multigene family in Drosophila and mammals that have different spatial and temporal patterns of expression (Salkoff et al. 1987; Loughney et al. 1989; Ramaswami and Tanouye 1989; Mandel 1992; Hong and Ganetzky 1994). In Drosophila, two sodium channel structural genes have been identified, \textit{para} and \textit{Dsc} (Salkoff et al. 1987; Loughney et al. 1989; Ramaswami and Tanouye 1989). The \textit{para} locus appears to encode the predominant class of sodium channels expressed in most or all neurons in the central and peripheral nervous systems (CNS and PNS) at all developmental stages, whereas expression of \textit{Dsc} is very limited until the pupal and adult stages when its expression becomes much more widespread and overlaps with \textit{para} expression at least in the CNS (Tseng-Crank et al. 1991; Amichot et al. 1993; Hong and Ganetzky 1994). The differential expression of \textit{para} and \textit{Dsc} suggests that they have different effects on neuronal excitability and that normal function of the nervous system depends on expression of the correct sodium channel gene in selected subsets of neurons. Normal neuronal activity also requires that these channel genes be expressed at the correct levels. For example, gene dosage studies demonstrated that changes in the level of \textit{para}\textsuperscript{+} expression, in either direction, had profound effects on neuronal excitability and further suggested that the level of \textit{para} expression was regulated primarily at the transcriptional level (Stern et al. 1990). Despite the importance of understanding how individual neurons acquire their characteristic signaling properties, the regulatory mechanisms ensuring selection of the right channel gene for transcription at the right time and in the right amount are not well understood.

To study this problem, we have been interested in elucidating the transcriptional control of the \textit{para} locus. Achieving this goal has been complicated by the size and complexity of this gene. The open reading frame of \textasciitilde 6 kb consists of a minimum of 35 exons distributed over \textasciitilde 60 kb of genomic DNA. The mature transcript is \textgtr 15 kb, indicating the presence of long untranslated segments at the 5' and/or 3' ends. In addition, extensive alternative splicing of the \textit{para} transcript is capable of generating a minimum of several hundred distinct coding sequences (Loughney et al. 1989; Thackeray and Ganetzky 1994, 1995). As a first step in elucidating the mechanisms that regulate \textit{para} expression, we set out to map the transcriptional start site and to try to locate upstream \textit{cis}-acting regulatory elements.

In the course of this analysis, we have carried out a detailed molecular analysis of the \textasciitilde 60 kb genomic segment immediately upstream of the \textit{para} locus in region 14D of the polytene chromosomes. Here we report the identification, molecular characterization, genomic organization, and expression pattern of the next four genes upstream of \textit{para}, which are calnexin (\textit{Cnx}), actin-
related protein 14D (Arp14D), calcineurin A 14D (CnnA14D) and chromosome associated protein (Cap). The distance separating adjacent pairs of genes in this region ranges from 1 to 4 kb. Two of these genes, Arp14D and CnnA14D, have been previously identified on the basis of other studies (BROWN et al. 1994; FVERBERG et al. 1994) and the other two are reported here for the first time. At least three of these genes and possibly Cap as well, have functions that are plausibly related to the expression, modulation, or regulation of sodium channels. Moreover, for some of those genes that were examined, we were unable to identify upstream regulatory elements capable of conferring the endogenous expression pattern on reporter constructs. On the basis of these results, we discuss the possibility that the genes in the para cluster represent a group of functionally related genes that share some coordinate regulatory mechanism.

MATERIALS AND METHODS

Fly stocks: Drosophila melanogaster were grown on standard medium at 25°. The wild-type strain used for all in situ hybridization and Northern blot analyses was Canton Special (CS). w, \(A2-\Delta(99B)/TM6\) or \(TM3\) embryos were used as recipients for germ-line transformation.

Isolation and sequencing of cDNA clones: Genomic DNA fragments upstream were used to screen the adult head cDNA libraries, a ZAP library (kindly provided by T. SCHWARZ, Stanford University) and a XEXLX library (PALAZZOLLO et al. 1990). Genomic clones, \(\lambda 12.3\) and cos\(5.1\) were isolated from the previous work (LOUGHNEY et al. 1989) and LC, MA and PC were provided kindly by R. MAssod (CNRS, Marseille, France). Library screening, radioactive labeling of DNA probes, hybridizations, and preparation of DNA were performed using standard methodology (SAMBROOK et al. 1989). After purification of phage, cDNA clones (ZAP library) in \(\lambda\) plasmids were autoexcised using the Exassist helper phage system (Stratagene) or cDNA inserts (XEXLX library) were released by Sa/Al/Apa double restriction digestion and then subcloned into \(\lambda\) plasmids vectors. Each cDNA was mapped using restriction enzymes (New England Biolabs; Promega).

For sequencing cDNAs, single-stranded template DNA was prepared by polyethylene glycol precipitation (SAMBROOK et al. 1989) and sequenced using the dideoxy nucleotide chain-termination methods with \(\alpha\)-labeled dATP (Sequenase version 2.0 DNA sequencing kit; United States Biochemical). Sequence information was read using MacVector 3.5 (IBI) software and if necessary, a composite DNA sequence including a single open reading frame was assembled. Computer searches of databases and alignment and analysis of sequences were carried out using the BLAST (ALTSCHUL et al. 1990) server at National Center for Biotechnology Information (NCBI) and the PILEUP and DISTANCES programs in the Genetics Computer Group software package (DEVEREUX et al. 1984). Pairwise protein alignments were done using the MEGALIGN program in the DNASTAR software (HEIN 1990).

Southern blot analysis: Southern analysis of cDNAs or genomic clones was performed using \(\alpha\)-labeled probes or non-radioactive probes. Southern blotting of radiolabeled probes was at high stringency as described in SAMBROOK et al. (1989). Nonradioactive Southern analysis was performed using enhanced chemiluminescence detection (ECL; Amersham) and probe labeling, hybridization, washes and detection of hybridizing bands were carried out according to manufacturer's instructions.

Primer extension analysis and Northern blot analysis: Total RNA was isolated from CS adults, third-instar larvae and staged embryos and poly(A)+ RNA selection was performed as described in MCLEAN et al. (1990) except that tissues were homogenized using a Brinkman homogenizer. Primer-extension analysis was done as described in DOMDEV et al. (1984) with some modifications. Oligonucleotide primers were \(\alpha\)-labeled at the 5’ end with T4 polynucleotide kinase (New England Biolabs). cDNA was synthesized in a 50-\(\mu\)l reaction mixture containing ~10\(^{6}\) cpm of \(\alpha\)-labeled primer; 10 \(\mu\)g poly(A)+ RNA or 100 \(\mu\)g total RNA; 500 \(\mu\)M each dNTP; 5 U of RNasin; 10 mM DTT; and 100 U of murine leukemia virus (MuLV) reverse transcriptase in the buffer supplied by manufacturer (GIBCO-Bethesda Research Labs) for 90 min at 37°. The mixture was preincubated at 65° for 5 min and then chilled on ice before adding MuLV reverse transcriptase. The RNA was then hydrolyzed by adding 12.5 \(\mu\)l of 0.5 M NaOH and incubating at 100° for 3 min, then chilled on ice. The NaOH was neutralized by adding 12.5 \(\mu\)l of 0.5 M HCl and 12.5 \(\mu\)l of 1 M Tri-HCl (pH7.4). Primer-extension products were ethanol-precipitated and analyzed on sequencing gels. Northern blot hybridization was performed at high stringency according to the manufacturer's instructions (Hybond-N; Amersham); the blots were kindly provided by C. MESSIAH, R. ORDIWAY and L. PALLANCE (University of Wisconsin, Madison).

In situ hybridization: Nonradioactive \(\alpha\)-in situ localization of RNA in whole embryos and larval tissues was done essentially as described in HONG and GANETZKY (1994). Cloned DNA probes were digested into small pieces using multiple restriction enzymes before random priming.

Germ-line transformation and immunostaining of \(\beta\)-galactosidase: Genomic DNA fragments were cloned into the pCaSpeR-ATG-lacZ transformation vector (THUMMEI et al. 1990). P-element mediated transformation was carried out as described in SANTAMARIA (1986) and SPRADLING (1986) with some modifications. Dechorionated embryos of \(w, \Delta 2-\Delta(99B)/TM6\) or \(TM3\) were injected with the plasmids at 1000 \(\mu\)g/ml in 5 mM KC1, 0.1 mM PO\(4\) (pH7.8), with 3% Durkee green food coloring. Transgenic stocks were kept as homozygotes or established over the appropriate balancers. Immunostaining of \(\beta\)-galactosidase was performed as described in HONG and GANETZKY (1994). Embryos of transformants were incubated for 2 hr with anti-\(\beta\)-galactosidase antibody (Boehringer-Mannheim; diluted 1:2000), washed, and incubated for 1 hr with horseradish peroxidase-conjugated goat anti-mouse IgG (Boehringer-Mannheim; diluted 1:500). Detection was done according to manufacturer's instructions (Boehringer-Mannheim). Mounting and photographing were as described in HONG and GANETZKY (1994).

Accession numbers: The GenBank accession numbers for the sequences reported here are: U30492 for the Cap sequence, U30493 for the CnnA14D sequence, U30604 for the promoter region of CnnA14D, U30603 for the genomic segment containing the 3’ UTR of CnnA14D and the 5’ UTR of Arp14D, and U30466 for the Cnx sequence.

RESULTS AND DISCUSSION

Delimitation of neuronal expression pattern by embryo in situ hybridization: We initiated a molecular analysis of the 14D region with the aim of characterizing the transcriptional start site of the para sodium channel gene. In previous work, a set of overlapping cDNAs representing the complete 6-kb open reading frame of
para was isolated (Loughney et al. 1989; Ramaswami and Tanouye 1989). However, on Northern blots the size of the para transcript is ~15 kb (Thackeray and Ganetzky 1994), indicating that there are long untranslated segments at the 5' and/or 3' ends. Genomic DNA encompassing the entire upstream region of para should be contained within the extensive chromosome walk carried out by Surdej et al. (1990), which extends ~500 kb upstream beyond the para translational start site. The rudimentary (r) locus has been located on this walk at ~130 kb upstream of para (Surdej et al. 1990), providing an outer limit for the extent of the para locus.

We previously characterized the embryonic expression pattern of the para transcript by tissue in situ hybridization and found that beginning at stage 13 it is present throughout the entire CNS and PNS (Hong and Ganetzky 1994). To obtain an approximate limit of the 5' end of the para transcript, we used upstream genomic DNA probes for tissue in situ hybridization to determine which of these probes detected the neural expression pattern characteristic of para. One concern was whether a genomic probe that contained a small exon and a large intron would still be able to detect the transcript by in situ hybridization. To test this, we used a known para intron probe for in situ hybridization and found exactly the same embryonic expression pattern as with a cDNA probe (data not shown), except that, as expected, the intron probe hybridized exclusively to nuclei.

We proceeded to use genomic clones for in situ hybridization and found that three overlapping clones, A12.3, cos35.1 and PC, extending ~75 kb upstream from the para translational start site, detected transcripts with a neuronal expression pattern similar to that of para, whereas, genomic probes MA and LC, further upstream, did not detect this pattern (Figure 1). These results delimited the para transcriptional initiation site to within 75 kb upstream of the translational start site. The genomic probes that detected the neuronal expression pattern were then used to screen an adult head cDNA library. Over 40 independent cDNAs were isolated and characterized by restriction mapping, Southern blot hybridization, and sequence analysis. From these studies we were able to map the 5' end of the para transcript (C.-S. Hong and B. Ganetzky, unpublished data). Unexpectedly, we discovered that all the isolated cDNAs representing neurally expressed transcripts were not derived from the para locus but represented the next four genes upstream of para and that all four are transcribed in the same direction as para. Two of these genes, actin-related protein (Arp14D) and calcineurin A (Cna14D) were also identified recently in other laboratories using different approaches (Brown et al. 1994; Fyrberg et al. 1994). Two other genes, Drosophila homologues of calnexin (Cnx) and chromosome-associated polypeptide (Cap), are described here for the first time. Below we provide the results of our analyses of each of these genes.

Characterization of Cnx: The initiation site of the para transcript was pinpointed to position 15 kb on the molecular map (Figure 2) on the basis of cDNA analysis and primer extension experiments (C.-S. Hong and B. Ganetzky, unpublished observations). Less than 4 kb upstream from this site, we identified another transcription unit defined by two overlapping cDNAs with a combined length of 2 kb (Figure 2). Because the cDNAs detect a transcript of ~3 kb on Northern blots (Figure 3A) and a poly(A) tail is present on one of them, the cDNA sequence is apparently incomplete by ~1 kb at the 5' end. Although the gene defined by these cDNAs is clearly distinct from para, the embryonic expression pattern as revealed by tissue in situ hybridization (Figure 4B) is remarkably similar to that of para (Figure 4A; Hong and Ganetzky 1994). In particular, expression occurs throughout the CNS and PNS beginning well after early neurogenesis and continuing until the completion of embryogenesis. It will be of interest to determine whether the similar expression patterns of this gene and para is maintained at later developmental stages as well.

Analysis of partial cDNA sequence revealed an incomplete open reading frame of ~1 kb. The ORF begins upstream from the amino terminus because an initiating methionine is lacking. In addition, our sequence is incomplete at the carboxy terminus. Database searches show that the deduced sequence shares ~30% amino acid identity with mammalian and Drosophila calreticulins (Figure 5). Calreticulin is a major Ca^{2+}-binding protein of the ER lumen and is thought to function as a buffer in the storage and release of Ca^{2+} (Smith and Koch 1989). The cDNA sequence encodes a conserved calreticulin-like segment that includes four repeats of one consensus motif (IPDPXAXKPEXWDE) and three repeats of a second consensus motif (GKWXAPLIXPNP) present in a segment that has been shown in mammalian calreticulins to bind Ca^{2+} with high affinity (Figure 5).

However, the greatest similarity of the deduced ORF, 58% amino acid identity, was found with another mammalian relative of calreticulin known as calnexin (Figure 5). Consequently, the gene product appears to be a Drosophila homologue of calnexin and we have named the gene Cnx. Sequence alignments suggest that the Drosophila calnexin sequence shown in Figure 5 lacks ~150 amino acids at the amino terminus and ~110 amino acids, including a membrane-spanning segment and an ER retention signal, at the carboxy terminus.

Mammalian calnexin was originally isolated as a Ca^{2+}-binding phosphoprotein from endoplasmic reticulum (ER) membranes (Wada et al. 1991) and is thought to function as a molecular chaperone that binds transiently to newly synthesized glycoproteins retaining in-
FIGURE 1.—Molecular map and embryonic in situ hybridization pattern of genomic clones from region 14D on the X chromosome. An EcoRI map of a genomic segment that extends from the para translational start site to the 5' region of r is shown. The map is based on four overlapping cosmids clones (LC, MA, PC and cos35.1) and a phage clone (λ12.3) shown at the top. The para translational start site is very close to coordinate position 0 kb and the orientation of para transcription is marked by an arrow. The 5' end of r, whose orientation is opposite that of para, is located at about coordinate -135 kb (LURDEJ et al. 1990). Aligned with the genomic map at the bottom are the staining patterns of embryos hybridized in situ with DNA probes from the cosmids MA, PC, and cos35.1 digested into small pieces using multiple restriction enzymes before random priming. Note that genomic DNA probes from coordinates 0 kb to about -75 kb hybridize in situ in the CNS and PNS (arrow and arrowhead, respectively), whereas no such pattern is observed in embryo hybridized with probes from beyond the -80 kb region.

FIGURE 2.—Molecular map of region 14D on the X chromosome showing the alignment of cDNAs from the four genes upstream of the para locus. Coordinates are as in Figure 1. All five genes shown are transcribed in the same direction as indicated. Boxes represent exons deduced by comparison of restriction maps of cDNA and genomic clones and by hybridization of cDNAs to genomic restriction fragments. The 5' ends of Cap and Cnx were not mapped precisely. The approximate distances between genes are shown. The distance between CmnA14D and Arp14D was determined precisely by sequencing.

correctly or incompletely folded proteins in the ER (reviewed by BERGERON et al. 1994). Because many glycosylated proteins, e.g., ion channels, are expressed specifically or primarily in neurons, it will be of interest to determine if Drosophila calnexin functions as a molecular chaperone for these proteins. In particular, the very similar expression patterns of Cnx and para raise the question of whether calnexin might act as a chaperone for the glycosylated sodium channel polypeptide. In mammals, multiple isoforms of calnexin have been found including some that are expressed in a tissue-specific manner (OHSKO et al. 1994; WATANABE et al.
regions of the ventral ganglion (Figure 4F). In situ hybridization at later stages has not been examined.

Sequence analysis of the cDNAs revealed a complete ORF of 395 aa, which is identical with that of Arp14D, one of five previously cloned actin-related protein genes in Drosophila (Fyrberg et al. 1994). Actin related proteins (ARPs) are divergent relatives of conventional actin that have been identified in a broad range of phyla (Schroer et al. 1994). Whereas conventional actins share 70–95% amino acid identities, most ARPs are 35–55% identical to actin. The role of ARPs is currently not known, although their divergence from conventional actins and among themselves in intermolecular contact sites, including the myosin-binding site, suggest that they may interact with a variety of different proteins other than myosin (Frankel et al. 1994). Arp14D shares ~46% identity with conventional Drosophila actins but has greater similarity (65% identity) with the Act2 protein of Saccharomyces cerevisiae (Fyrberg et al., 1994; data not shown). Both proteins have conserved amino acids in the ATP and Ca²⁺-binding core domain but lack myosin-binding sites (Schwob and Martin 1992; data not shown).

We analyzed the genomic region between Arp14D and its immediate 5′ neighbor (see below) to initiate studies of the transcriptional regulation of Arp14D. The transcriptional start site was mapped by primer extension experiments. From embryonic and adult poly(A)⁺ RNA, a single extension product was synthesized using a primer that hybridized to the 5′ end of our Arp14D cDNA (Figure 6A). The size of this product indicated that transcription of Arp14D began 38 bp upstream from the 5′ end of our cDNA. The sequence of the identified initiation site (ATGCATCT) has a good match with the consensus [ATCA(G/T)T(C/T)] sequence for Drosophila transcriptional start sites (Hultmark et al. 1986). Although no apparent TATA box is present in the region 30 bp upstream from the start site, a TA-rich region is present at 43 bp upstream (Figure 7A).

The distance between the transcriptional start site of Arp14D and the polyadenylation signal of the next gene upstream (see below) is 1759 bp. To delimit segments within this region important for Arp14D expression, we inserted varying extents of this genomic DNA including ~300 bp of the untranslated leader sequence upstream of the lacZ reporter in the pCaSpeR-ATG-lacZ vector (Thummel et al. 1988). Several independent transformants were produced for each construct and the embryonic pattern of β-galactosidase expression was examined by immunostaining. Four constructs, RKII, NK, RK, and SK, containing 1694, 1449, 1187, and 126 bp, respectively, of 5′-flanking DNA (Figures 7 and 8), were examined. The RKII construct produced faint staining in the CNS, hindgut, and proventriculus (Figure 8) and essentially the same expression pattern was observed for the NK and RK constructs. In contrast, transformants

Figure 3.—Northern blot analysis of poly(A)⁺ RNA isolated from Cs adults and larvae. Northern blots of adult RNA were probed with cDNA clones from each of the genes indicated. The blot probed with the Cap cDNA (D) includes both adult (a) and larval (l) RNA. The size of the bands for blots A–C are marked on the left, and for blot D on the right.
carrying the smallest construct, SK, had uniform strong staining in the entire embryo. These results suggest the possibility that normal expression pattern of Arpl4D may depend on some silencer element(s) located between 1187 and 126 bp upstream of the transcription start site that restrict expression to certain tissues. Expression of Arpl4D in the nervous system may also require the activity of upstream enhancing elements. Database searches of the entire upstream sequence of Arpl4D identified only one stretch with similarity to other known transcriptional regulatory elements. This sequence between 887 and 902 bp upstream of the transcription start site matches well with an element (element I) in the 5′ flanking region of the Drosophila dopa decarboxylase (Ddc) gene (Figure 7B). This element is highly conserved between the Ddc genes of D. melanogaster and D. virilis and is necessary but not sufficient for expression of this gene in the CNS (Scholnick et al. 1986; Johnson and Hirsh 1990). Expression of Ddc in the hypoderm is not influenced by element I (Scholnick et al. 1986). Sequences closely related to element I and the Arpl4D upstream element have been found in several other genes expressed in the nervous system including the rat type II sodium channel gene, SCG10, and human dopamine β-hydroxylase gene (Malk et al. 1990; Mori et al. 1992; Ishiguro et al. 1993). However, the effect of the element on expression of those genes has not yet been investigated. The role of this element as well as that of the putative silencer element(s) in expression of Arpl4D will require more detailed functional dissection of the region between 126 and 1187 bp upstream of the transcriptional start site.

Characterization of CnnA14D: Ten independent overlapping cDNAs hybridized to the genomic DNA at coordinates −63 through −34 kb, immediately upstream of Arpl4D (Figure 2). On Northern blots, these

FIGURE 4.—Embryonic and larval expression patterns of the four genes upstream of para determined by whole-mount in situ hybridization. The embryos are oriented with anterior to the left and ventral down. The larval ganglia are oriented with anterior up. (A) Stage 16 embryo hybridized with pura probe is shown for comparison with expression patterns of the other genes. Note that strong staining appears in the antennomaxillary complex (arrow) and ventral ganglion of the CNS (arrowhead). (B) Cnx expression is seen in the antennomaxillary complex and throughout the entire CNS and PNS in a pattern very similar to pura. (C) In stage 16 embryos, Arpl4D is expressed in the CNS and hindgut (small arrowhead) and proventriculus (large arrowhead). (D) CnnA14D expression is detected throughout the CNS and PNS in a pattern similar to pura. (E) Cup is expressed in the CNS, gonad (arrow) and antennomaxillary complex. (F) In the third instar larva, Arpl4D expression is detected in the brain hemispheres and the ventral ganglion (arrows). (G) Cup expression also appears throughout the larval brain and ventral ganglion, as well as in the wing and leg discs (H) and in the salivary gland (I) of third instar larvae.

...
Analysis of the *para* gene cluster

FIGURE 5.—Alignment of the deduced amino acid sequence of a partial *Cnx* cDNA sequence (DCALNEXIN) with those of a mouse calnexin (MCALNEXIN; SCHREIBER et al. 1994) and Drosophila calreticulin (DCALRETICULIN; SMITH 1992). Amino acid identities are shaded.

cDNAs detect two transcripts of 4.5 and 2 kb in adult poly(A)⁺ RNA (Figure 3C). The same two transcripts were also detected in larval RNA (data not shown). The composite cDNA has a total length of 4.2 kb with a poly(A) tail at the 3' end and thus appears to represent the 4.5 kb transcript. Although the 2-kb transcript appears to be more abundant on Northern blots, none of the ten cDNAs isolated specifically correspond to the 2-kb transcript. One possible explanation is that the 2-kb transcript may be more abundant in locations other than the head because cDNAs in this study were isolated from a head library.

Transcripts from this gene detected by *in situ* hybridization with cDNA probes are expressed in several distinct patterns at different stages of embryogenesis. Before gastrulation, strong expression occurs throughout the embryo but expression is no longer detectable shortly after gastrulation (data not shown). Shortly after neurogenesis, expression is detected in the CNS and PNS in a pattern that is similar to that of *para* (Figure 4D). These results suggest that the encoded protein may have several different functions during embryogenesis—an early function required throughout the embryo and a later function that is neuronally specific.

Sequence analysis of the cDNAs revealed a complete ORF encoding a deduced polypeptide of 578 amino acids (form I; Figure 9). A second splice isoform (form II) that differs from the first in the 5' untranslated leader sequence and in the N-terminal 176 amino acids was also identified. Database searches indicate that the form II polypeptide is identical with a Drosophila calcineurin A homologue (originally termed PP2B and renamed here as CnnA14D) previously cloned via PCR (BROWN et al. 1994). The polypeptides encoded by CnnA14D are ~75% identical to mammalian calcineurin A polypeptides. Mammalian calcineurin (also termed protein phosphatase-2B) is a highly conserved, calcium-activated protein phosphatase consisting of a catalytic subunit (A) and a regulatory subunit (B) that functions in a wide variety of signaling pathways.

FIGURE 6.—Determination of the 5' ends of the *Arp14D* and *CnnA14D* transcripts by primer extension analysis. Adult (a) and embryonic (e) poly(A)⁺ (Arp14D) or total (CnnA14D) RNA were used. Sequencing reactions (C, T, A, G) of genomic clones corresponding to the cDNAs were carried out using the same primers as in the extension reactions. The complete sequence of the genomic region of *Arp14D* is shown in Figure 7. Both *Arp14D* (A) and *CnnA14D* (B) are transcribed from single start sites in adults and embryos. The sequences of their respective initiation sites (ATGCATCT) and (GTAATTT) are in good agreement with the Drosophila consensus sequence (see text).
A

poly(A) signal of CnnA14D

-1800 AAGGGATCC

-1820 CGGCAACCAAGTTTTCG

-1795 TCGG

-1755 GT

-1720 AATGGG

-1715 AACCTTTG

-1700 AG

-1690 GAATTC

-1680 CT

-1670 GA

-1655 GG

-1635 GT

-1620 A

-1610 C

-1595 G

-1585 T

-1570 C

-1560 A

-1555 G

-1545 T

-1535 A

-1520 A

[FIGURE 7.-Sequence analysis of the genomic region between the 5' end of Arp14D and the poly(A) signal of CnnA14D. (A) The transcriptional initiation site of Arp14D and the putative poly(A) signal of CnnA14D (underlined) are shown. Arrows mark the 5' end of the four transformation constructs (SK, RK, NK, and RKII) shown in Figure 8. Two putative GAGA factor binding sites (bold face) and several repeats of CATA residues (underlined) are shown. A short segment of homology with the Drosophila Ddc (dopa decarboxylase) promoter region is boxed. The alignment of this segment with the element I sequence from the D. melanogaster and D. virilis Ddc genes is shown in B. Identical nucleotides are highlighted. Note that the orientation of these elements in the region upstream of the Ddc genes is opposite from that in the Arp14D promoter region. (SCHREIBER 1992; ENSLEN and SODERLING 1994). Calcineurin is highly enriched in mammalian brain where it modulates the activities of several different ion channels (LUAN et al. 1993; MURPHY et al. 1993; LIEBERMAN and MODY 1994) and is also involved in synaptic vesicle recycling (LIU et al. 1994). Calcineurin B is the Ca\(^{2+}\)-binding subunit and is apparently encoded by a single gene in mammals, whereas at least three genes encoding calcineurin A polypeptides have been isolated. Transcripts from each of the three genes undergo alternative splicing, further increasing the diversity of calcineurin A isoforms. Each

(SCHREIBER 1992; ENSLEN and SODERLING 1994). Calcineurin is highly enriched in mammalian brain where it modulates the activities of several different ion channels (LUAN et al. 1993; MURPHY et al. 1993; LIEBERMAN and MODY 1994) and is also involved in synaptic vesicle recycling (LIU et al. 1994). Calcineurin B is the Ca\(^{2+}\)-binding subunit and is apparently encoded by a single gene in mammals, whereas at least three genes encoding calcineurin A polypeptides have been isolated. Transcripts from each of the three genes undergo alternative splicing, further increasing the diversity of calcineurin A isoforms. Each
isofrom is predominantly expressed in different tissues (GUERINI et al. 1992; MUARMATSU and KINCAID 1993). The cDNA probes used in our in situ hybridization were from segments common to forms I and II, so we do not know whether the splice isoforms of CnnA14D have differential expression.

Another calcineurin A homologue (referred to here as CnnA21EF) that maps to 21EF on the polytene chromosomes has been isolated by low stringency hybridization (GUERINI et al. 1992). The CnnA21EF polypeptide is ~71% identical with mammalian calcineurin A but only ~67% identical with the CnnA14D polypeptide suggesting that the two Drosophila genes originated after a duplication that occurred before the evolutionary separation of vertebrates and invertebrates ~600 mya. The N and C termini of the CnnA14D and CnnA21EF polypeptides are highly divergent but the catalytic and regulatory domains (Figure 9) are well conserved. The regulatory domain is composed of subdomains containing a putative calcineurin-B binding site, a calmodulin-binding site, and an autoinhibitory region (COHEN 1989). Although the respective functions of the two genes remain to be studied, CnnA21EF is expressed at low levels (GUERINI et al. 1992) compared with CnnA14D, which apparently encodes the predominant form(s) of Drosophila calcineurin.

To look for regions involved in the transcriptional control of CnnA14D, we first determined the location of the promoter region. Because the two forms of calcineurin encoded by CnnA14D have different untranslated leader sequences, CnnA14D may have multiple promoters. Consistent with this interpretation, a cDNA probe specific for the 5' end of form I mapped to coordinate −63 kb whereas a probe specific for the 5' end of form II mapped to −45 kb (Figure 2). We chose to map the distalmost promoter with greater resolution by primer extension experiments. A primer near the 5' end of form I cDNA resulted in the synthesis of a single extension product using either embryonic or adult total RNA as template (Figure 6B). The size of this product indicates that transcription of CnnA14D form I initiates at 349 bp upstream from the 5' end of our form I cDNA. The initiation site of this transcript (GTAATTT) is in good agreement with the Drosophila consensus sequence (HULTMARK et al. 1986). No recognizable TATA box was found in this region but several AT-rich elements are located nearby and overall features of the promoter region fit the characteristics of TATA-less promoters in Drosophila including a CGTG element at −1−25 downstream and the T-rich initiation site (ARKHPOVA 1995). Because the distance between the 5' end of the CnnA14D form I transcript and the next gene

---

**FIGURE 8.**—Embryonic expression patterns of fusion constructs containing various extents of upstream DNA fused to a lacZ reporter. A schematic diagram of the genomic region between Arp14D and CnnA14D indicating the extents of four different fusion constructs is shown. Each fusion contains ~300 bp from the first exon of Arp14D. Representative staining patterns with anti-β-galactosidase antibodies of embryos carrying each of the transformed constructs are shown at right. Three constructs, RKII, NK, and RK all reproduce the endogenous expression pattern of Arp14D in the CNS, hindgut (arrowhead), and proventriculus (arrow) in stage 16 embryos. In contrast, the construct SK give uniform expression over the entire embryo.
FIGURE 9.—Alignment of Drosophila calcineurin A subunits. The amino acid sequences of two alternatively spliced cDNAs, CnnA14D(I) and CnnA14D(II), are compared with another calcineurin A subunit encoded by a gene, CnnA21EF, mapping to polytene chromosome region 21EF. Only the first 176 amino acids of CnnA14D(I) are shown; it is identical with CnnA14D(I) thereafter. The overlined 11 amino acids were not in the previous sequence (Brown et al. 1994) but exist in both our cDNAs. Identical amino acids are shaded.

upstream (see below) is only ~1 kb (Figure 2), we tested this region to see if it contained the information necessary to drive expression of a reporter construct in the appropriate pattern. A 2-kb fragment containing 400 bp of the upstream leader sequence was fused upstream of the lacZ reporter in the pCaSpeR-ATGlacZ vector and 210 independent transformants were isolated. None of these transformants reproduced the endogenous expression pattern of CnnA14D(I) in either young or mature embryos (data not shown). These results suggest that the regulatory elements necessary for normal expression of CnnA14D(I) do not reside in the region between the distalmost promoter and the next gene upstream but may lie within introns upstream of the form II promoter.

Characterization of Cap: Five independent overlapping cDNAs mapped to coordinates -75 through -64 kb on the genomic map. The composite cDNA has a length of 4.7 kb and has a poly(A) tail at the 3' end. The distance separating the 3' end of this transcription unit from the 5' end of CnnA14D(I) is only ~1 kb. Northern blot analysis of adult and larval poly(A)+ RNA reveals two transcripts of 4 and 7 kb (Figure 3D). Expression in larvae appears to be higher than in adults and the 7-kb transcript predominates in larvae, whereas both are expressed about equally in adults. The composite cDNA has a total length of 4.7 kb with 406 bp of untranslated sequence at the 3' end and 548 bp of untranslated sequence at the 5' end. Thus, our cDNAs apparently represent the 7-kb transcript and are incomplete at the 5' end by ~2 kb.

Transcripts from this gene are first detected by in situ hybridization with cDNA probes during the blastoderm stage and continue to be present throughout the germ band (data not shown). After germ band retraction, expression is detected only in the CNS and gonads (Fig-
ure 4E). At later stages of development, strong expression is detected in larval brains and imaginal discs including wing and leg discs (Figure 4, G and H). Expression in imaginal discs occurs throughout the entire disc without any specific pattern. Expression was also readily detected in larval salivary glands (Figure 4).

Sequence analysis revealed a single complete ORF encoding a deduced polypeptide of 1231 amino acids (Figure 10A). Database searches demonstrated that the polypeptide shares significant similarity with the recently discovered SMC family of proteins (Figure 10, B-D) that are required for mitotic chromosome assembly and organization (Strunnikov et al. 1995). SMC1 (stability of minichromosomes) protein was identified originally from an S. cerevisiae mutant that had an increased frequency of minichromosome nondisjunction (Strunnikov et al. 1993). Subsequently, related family members have been identified in C. elegans, Xenopus, and mammals. The encoded proteins are abundant chromosome proteins required for normal chromosome condensation and chromatid organization (Hirano and Mitchison 1994; Saitoh et al. 1994; Strunnikov et al. 1995). The family member in nematodes is encoded by the dpy-27 gene, which was originally identified by a mutational defect in dosage compensation, indicating that the effect of these proteins on chromosome organization can also affect global regulation of gene expression (Chuang et al. 1994).

Of the various known family members, the Drosophila gene appears to be most closely related to the XCap-E gene (Xenopus chromosome associated protein E; Figure 10) so we have named the Drosophila gene Cap. The Cap gene product contains several structural features in common with other SMC family members including the presence of three globular domains in the amino-terminal, central, and carboxy-terminal regions connected by two coiled coil segments. There is little sequence conservation in the coiled coil regions but the globular regions are more highly conserved. Within the globular domains at the amino- and carboxy-terminal are NTP-binding sites A and B, respectively. In other RNA helicase proteins, the NTP-binding A site is known to be involved in binding ATP and the NTP-binding B site is required for ATP hydrolysis (Pause and Sonenberg 1992; Saitoh et al. 1994). Although XCap-E and Cap share only ~23% amino acid identities overall, the NTP-binding A and B sites are, respectively, 55 and 62% identical.

The dendrogram shown in Figure 11 indicates that Cap belongs to a subgroup of the SMC family containing the yeast SMC2 gene, the Xenopus XCap-E gene, and the chicken SCII gene (Hirano and Mitchison 1994; Saitoh et al. 1994; Strunnikov et al. 1995). It is reason-
we may consider the
beyond the polyadenylation signal, the gaps between
genes in this cluster ranges from <1 to 4 kb with an
cludes this gene cluster is transcribed. For comparison,
Another possibility is that some regulatory elements
sequences (STATHAKIS et al. 1995). Consequently our results raise the
questions of whether there are functional interactions
among these genes in the para cluster and whether the
expression of these genes is coordinately regulated to
some degree. Although we cannot yet definitively an-
swer these questions one way or the other, functional
relationships between the upstream gene products and
para-encoded sodium channels can readily be envi-
ioned. The para polypeptide contains 10 putative glyco-
sylation sites and several putative sites for phosphoryla-
ion by cAMP-dependent protein kinase or protein
kinase C (LOUGHNEY et al. 1989; THACKERAY and GANET-
zyk 1994). In mammals, recent studies have shown that
calcineurin modulates sodium channel activity by de-
phosphorylating residues that have been phosphory-
lated by cAMP-dependent protein kinases (MURPHY et al. 1993) and that calnexin is involved in appropriate
folding and transport into membranes of glycosylated
proteins including a chloride channel protein (PIND et al. 1994). In addition, cytoskeletal proteins such as actin
filaments and ankyrin interact with sodium channels and
regulate channel activity in mammals (SRINIVASAN et al. 1992; PRAT et al. 1993). Thus, calnexin, calcineurin
and actin-related protein could be involved in impor-
tant aspects of para-encoded sodium channel pro-
cessing, function, modulation, or distribution.

The Cap locus is the only member of the cluster
whose gene product lacks any apparent relationship
with sodium channels. In this regard, it is of interest to
note that there are some similarities between the gene
products of Cap and mle. Like the Cap polypeptide, the
mle polypeptide also contains both NTP-binding sites
A and B and is a chromosomally associated protein
(KUKUDA et al. 1991). A homologue of Cap in nema-
todes is involved in dosage compensation as is the mle
protein in Drosophila. The existence of unusual alleles
of mle, mle<sup>op</sup>, with a temperature-sensitive paralytic
phenotype associated with a block in action potential propa-
gation (WU et al. 1978; GANETZKY and WU 1986) re-
vealed that this gene somehow affected sodium channels. Subsequent studies indicated that expression of
para is reduced in an mle<sup>op</sup> background (STERN et al.
1990). This phenotypic effect is known to be distinct
from the effect of mle on dosage compensation (KERNAN
et al. 1991). Recent results have shown that processing of
the para transcript is defective in mle<sup>op</sup> mutants (R.
REENAN and B. GANETZKY, unpublished results). Thus,
it is possible that Cap is also involved in regulating some
aspect of para expression.

If other members of the para cluster do affect expres-
sion or function of sodium channels, there is still the
question of why they are so tightly spaced. The polypep-
ptides encoded by the genes in this cluster do not inter-
act exclusively with sodium channels and must have
other important biological functions. Possibly the ar-
Arrangement of these genes is coincidental but again it is tantalizing to speculate that there is some functional basis for it that provides a selective advantage. It will be of interest to carry out a detailed genetic analysis of the other genes in the *para* cluster, which should help answer some of these questions.

Finally, it is worth noting that the existence of a gene complex in the region including the Sh potassium channel gene has been proposed on the basis of genetic interactions and similar phenotypes among mutations in this region (Ferrus et al. 1990; de La Pompa 1994). Recently, a protein kinase, a troponin I homologue and a Ca$^{2+}$-binding protein (Frequenin) have been cloned from the Sh complex. Several of these proteins have been shown to affect development of the nervous system or synaptic transmission (Barbas et al. 1991; Pongs et al. 1993). The molecular organization of genes within the Sh complex has not been characterized yet so it remains to be seen whether there will be a dense clustering similar to that of the *para* region. Perhaps more detailed genetic and molecular studies of the *Sh* region as well as regions containing other ion channel structural genes will reveal that dense clustering of functionally related genes is common for reasons that remain to be determined.

We thank members of the Galantay laboratory for their encouragement and advice, Raymond Miassod for providing cosmID clones, Michael Hoffmann, Richard French-Constant, Sean Carroll, Ching Chung and Robert Reenan for comments on the manuscript. This work was supported by National Institutes of Health grant GM-43100 to B.G. This paper is 3447 from the Laboratory of Genetics, University of Wisconsin, Madison.

LITERATURE CITED


Liu, J.-P., A. T. R. Sim and P. J. Robinson, 1994 Calcineurin inhibi-


Communicating editor: R. S. Hawley