Molecular Characterization of Neurally Expressing Genes in the *para* Sodium Channel Gene Cluster of Drosophila

Chang-Sook Hong[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 *para*, which encodes the major class of sodium channels in the Drosophila nervous system, we have tried to locate upstream *cis*-acting regulatory elements by mapping the transcriptional start site and analyzing the region immediately upstream of *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 (*Cnx*), actin related protein 14D (*Arp14D*), calcineurin A 14D (*CnnA14D*), and chromosome associated protein (*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, *para* and *Dsc* (SALKOFF *et al.* 1987; LOUGHNEY *et al.* 1989; RAMASWAMI and TANOUYE 1989). The *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 *Dsc* is very limited until the pupal and adult stages when its expression becomes much more widespread and overlaps with *para* expression at least in the CNS (TSENG-CRANK *et al.* 1991; AMICHOT *et al.* 1993; HONG and GANETZKY 1994). The differential expression of *para* and *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 $para^+$ expression, in either direction, had profound effects on neuronal excitability and further suggested that the level of 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 *para* locus. Achieving this goal has been complicated by the size and complexity of this gene. The open reading frame of ~6 kb consists of a minimum of 35 exons distributed over ≥60 kb of genomic DNA. The mature transcript is >15 kb, indicating the presence of long untranslated segments at the 5′ and/or 3′ ends. In addition, extensive alternative splicing of the *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 *para* expression, we set out to map the transcriptional start site and to try to locate upstream *cis*-acting regulatory elements.

In the course of this analysis, we have carried out a detailed molecular analysis of the ~60 kb genomic segment immediately upstream of the *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 *para*, which are calnexin (*Cnx*), actin-

Corresponding author: Barry Ganetzky, Laboratory of Genetics, 445 Henry Mall, University of Wisconsin, Madison, WI 53706.
E-mail: ganetzky@facstaff.wisc.edu

[1] *Present address:* 114 Center for Neurologic Diseases, Harvard Medical School, Brigham and Women's Hospital, 221 Longwood Ave., Boston, MA 02115.

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; FYRBERG *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*, *Δ2-3(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 λZAP library (kindly provided by T. SCHWARZ, Stanford University) and a λEXLX library (PALAZOLLO *et al.* 1990). Genomic clones, λ12.3 and cos35.1 were isolated from the previous work (LOUGHNEY *et al.* 1989) and LC, MA and PC were provided kindly by R. MIASSOD (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 (λZAP library) in pBluscript SK-plasmids were autoexcised using the Exassist helper phage system (Stratagene) or cDNA inserts (λEXLX library) were released by *Sac*I/*Apa*I double restriction digestion and then subcloned into pBluscript 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 ^{35}S-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 cDNA 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 ^{32}P-labeled probes or nonradioactive 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 unstaged 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 DOMDEY *et al.* (1984) with some modifications. Oligonucleotide primers were ^{32}P-labeled at the 5′ end with T4 polynucleotide kinase (New England Biolabs). cDNA was synthesized in a 50-μl reaction mixture containing ~10^6 cpm of ^{32}P-labeled primer; 10 μg poly(A)$^+$ RNA or 100 μg total RNA; 500 μM 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 μl of 0.5 M NaOH and incubating at 100° for 3 min, then chilled on ice. The NaOH was neutralized by adding 12.5 μl of 0.5 M HCl and 12.5 μl 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. MERRILL, R. ORDWAY and L. PALLANCK (University of Wisconsin, Madison).

***In situ* hybridization:** Nonradioactive *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 β-galactosidase: Genomic DNA fragments were cloned into the pCaSpeR-ATG-*lacZ* transformation vector (THUMMEL *et al.* 1988). *P*-element mediated transformation was carried out as described in SANTAMARIA (1986) and SPRADLING (1986) with some modifications. Dechorionated embryos of *w*, *Δ2-3 (99B)/TM3* or *TM6* were injected with the plasmids at 1000 μg/ml in 5 mM KCl, 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 β-galactosidase was performed as described in HONG and GANETZKY (1994). Embryos of transformants were incubated for 2 hr with anti-β-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, λ12.3, 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 (*CnnA14D*) 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 intiation 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 downstream 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 (GKWXAPLIXNPNY) 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-

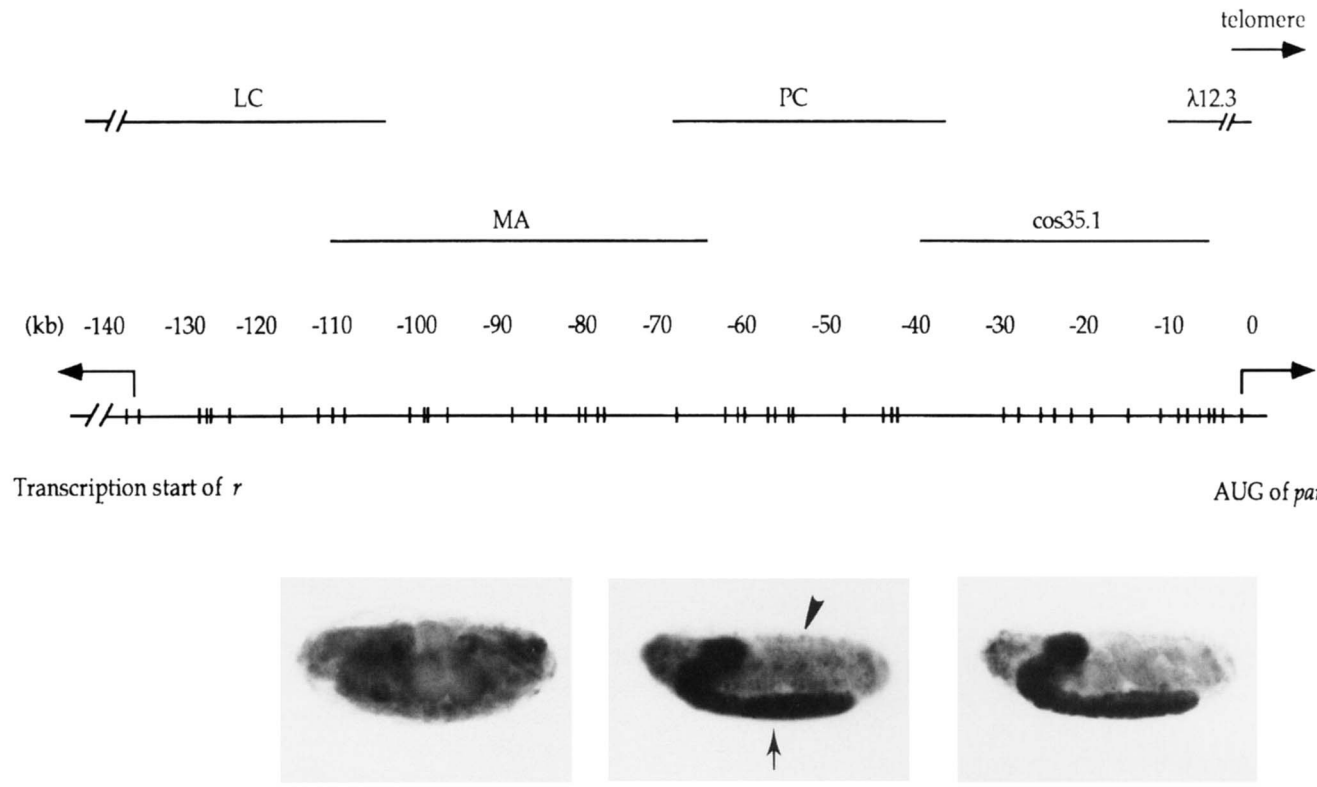

FIGURE 1.—Molecular map and embryonic *in situ* hybridization pattern of genomic clones from region 14D on the *X* chromosome. An *Eco*RI map of a genomic segment that extends from the *para* translational start site to the 5′ region of *r* is shown. The map is based on four overlapping cosmids clones (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 (SURDEJ *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.

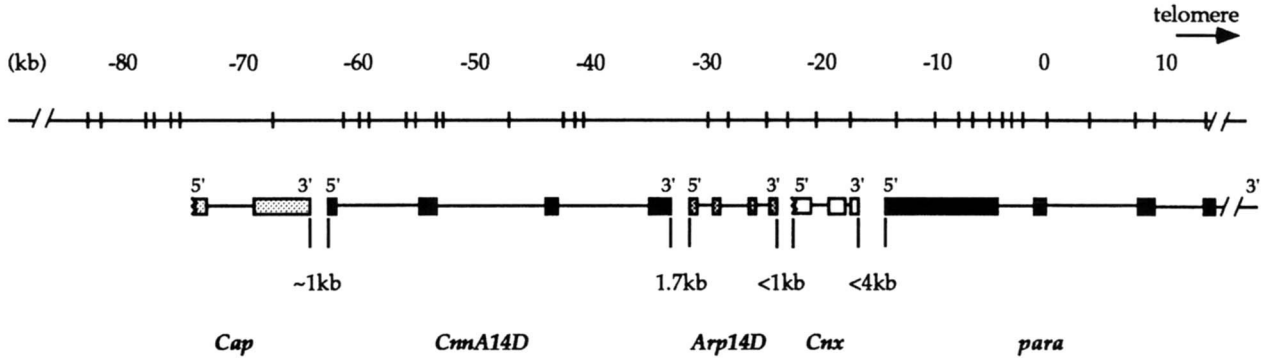

FIGURE 2.—Molecular map of region 14D on the *X* chromosome showing the alignment of cDNAs from the four genes upstream of the *para* locus. Coordinates are as in Figure 1. All five genes shown are transcribed in the same direction as indicated. Boxes represent exons deduced by comparison of restriction 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 *CnnA14D* 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 (OHSAKO *et al.* 1994; WATANABE *et al.*

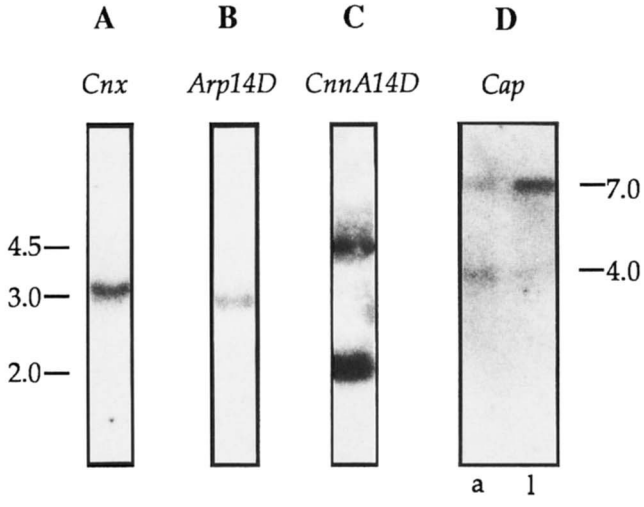

FIGURE 3.—Northern blot analysis of poly(A)$^+$ RNA isolated from *CS* adults and larvae. Northern blots of adult RNA were probed with cDNA clones from each of the genes indicated. The blot probed with the *Cap* cDNA (D) includes both adult (a) and larval (l) RNA. The size of the bands for blots A–C are marked on the left, and for blot D on the right.

1994) and some that are present in membranes other than ER including cardiac sarcoplasmic reticulum and hepatic nuclear membranes (CALA *et al.* 1993; GILCHRIST and PIERCE 1993). Thus, *Cnx* may be just one member of a gene family in Drosophila, other members of which function in cells outside the nervous system.

Characterization of *Arp14D*: Five overlapping cDNAs mapped by Southern blot hybridization to a segment of the genomic DNA between 25 and 32 kb upstream of the *para* translational start site and <1 kb upstream of the *Cnx* transcription unit. (Figure 2). The composite cDNA of 3 kb is the same size as the transcript detected by cDNA probes on Northern blots (Figure 3B), indicating that the composite cDNA is full length.

Expression of the corresponding transcript is readily detected by *in situ* hybridization throughout embryogenesis beginning at the stage of germ band elongation (data not shown). In mature embryos, the transcript is strongly expressed in the CNS, as well as in the hindgut and proventriculus (Figure 4C). Faint expression is also seen in the epidermis and antennomaxillary complex. All of the expressing tissues are ectodermal derivatives (CAMPOS-ORTEGA and HARTENSTEIN 1985). The embryonic expression pattern of this gene overlaps that of *para* and *Cnx* but is distinct from both. Expression of *Arp14D* outside the CNS was not detected in our initial *in situ* hybridization experiments using cosmid sequences as probes (Figure 1). Most likely, this difference is because of the greater sensitivity of detection when *Arp14D* expression is examined with a specific cDNA probe *vs.* probes from a large genomic segment that contains the gene. In third instar larvae, the transcript is expressed strongly in the brain hemispheres and in some cells in the ventrolateral and dorsolateral 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 Ca^{2+}-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

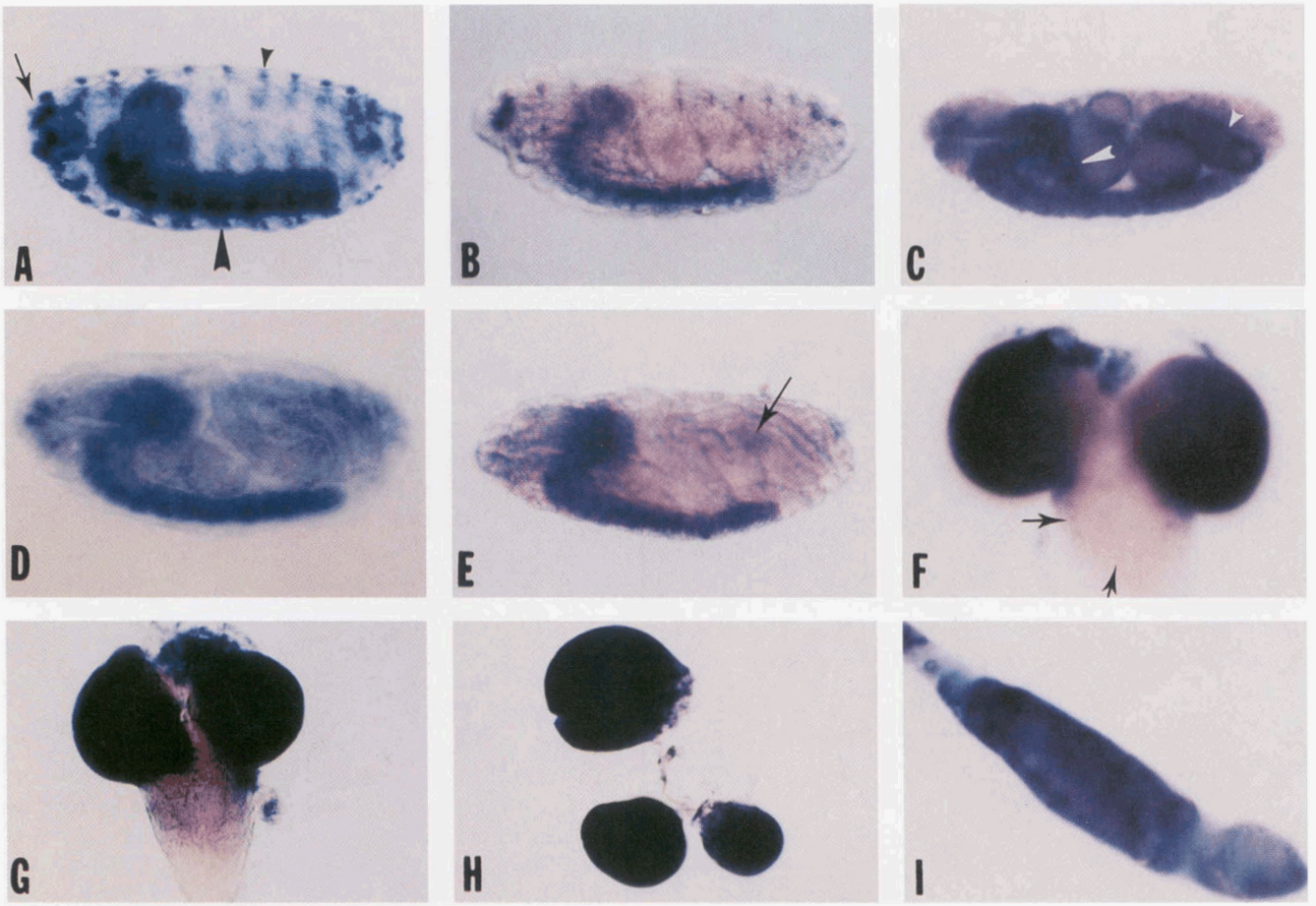

FIGURE 4.—Embryonic and larval expression patterns of the four genes upstream of *para* determined by whole-mount *in situ* hybridization. The embryos are oriented with anterior to the left and ventral down. The larval ganglia are oriented with anterior up. (A) Stage 16 embryo hybridized with *para* 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 *para*. (C) In stage 16 embryos, *Arp14D* 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 *para*. (E) *Cap* is expressed in the CNS, gonad (arrow) and antennomaxillary complex. (F) In the third instar larva, *Arp14D* expression is detected in the brain hemispheres and the ventral ganglion (arrows). (G) *Cap* 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.

carrying the smallest construct, SK, had uniform strong staining in the entire embryo. These results suggest the possibility that normal expression pattern of *Arp14D* 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 *Arp14D* in the nervous system may also require the activity of upstream enhancing elements. Database searches of the entire upstream sequence of *Arp14D* identifed only one stretch with similarity to other known transcriptional regulatory elements. This sequence between 887 and 902 bp upstream of the start sites 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 *Arp14D* 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 (MAUE *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 *Arp14D* 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 *Arp14D* (Figure 2). On Northern blots, these

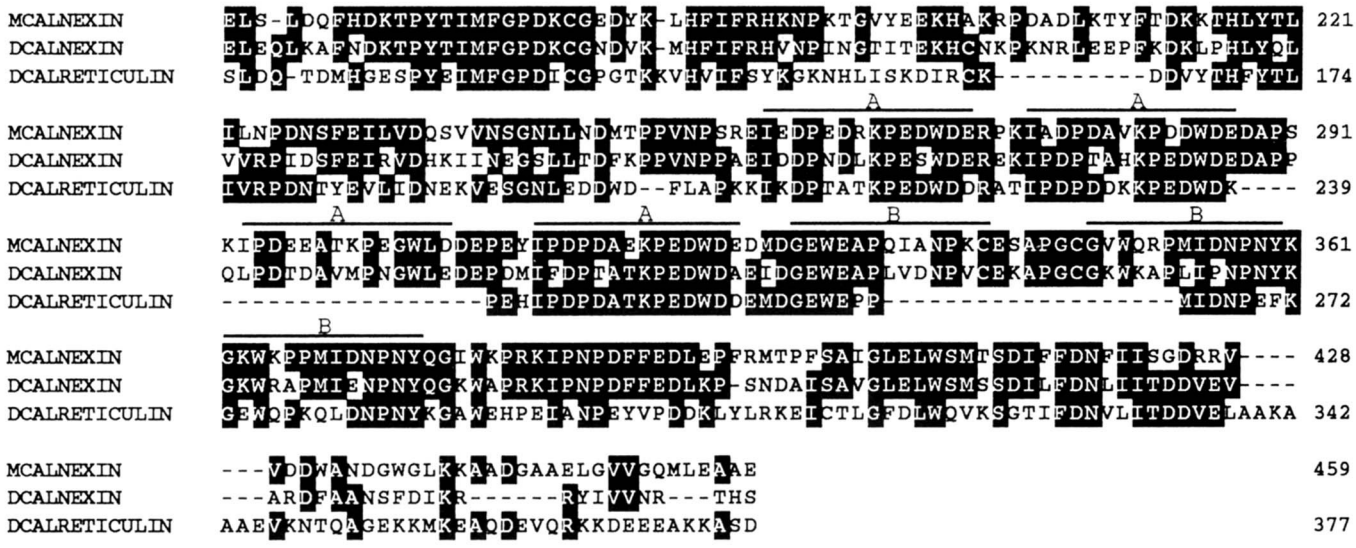

FIGURE 5.—Alignment of the deduced amino acid sequence of a partial *Cnx* cDNA sequence (DCALNEXIN) with those of a mouse calnexin (MCALNEXIN; SCHREIBER *et al.* 1994) and Drosophila calreticulin (DCALRETICULIN; SMITH 1992). Amino acid identities are shaded.

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 14D* 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

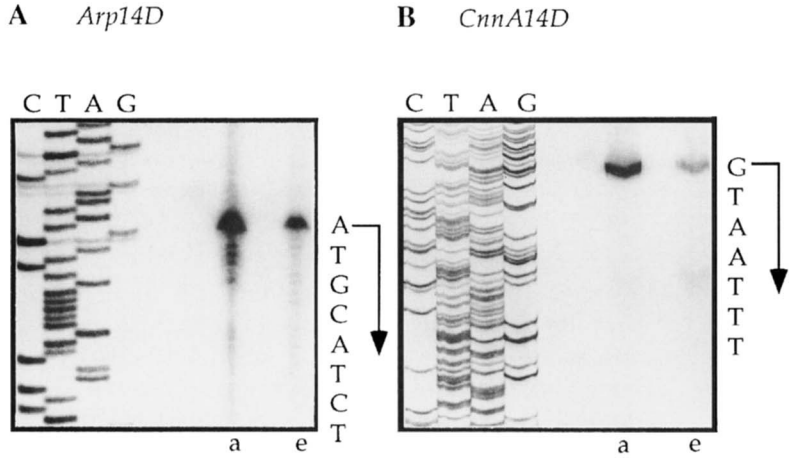

FIGURE 6.—Determination of the 5′ ends of the *Arp14D* and *CnnA14D* transcripts by primer extension analysis. Adult (a) and embryonic (e) poly(A)$^+$ (*Arp14D*) or total (*CnnA14D*) RNA were used. Sequencing reactions (C, T, A, G) of genomic clones corresponding to the cDNAs were carried out using the 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).

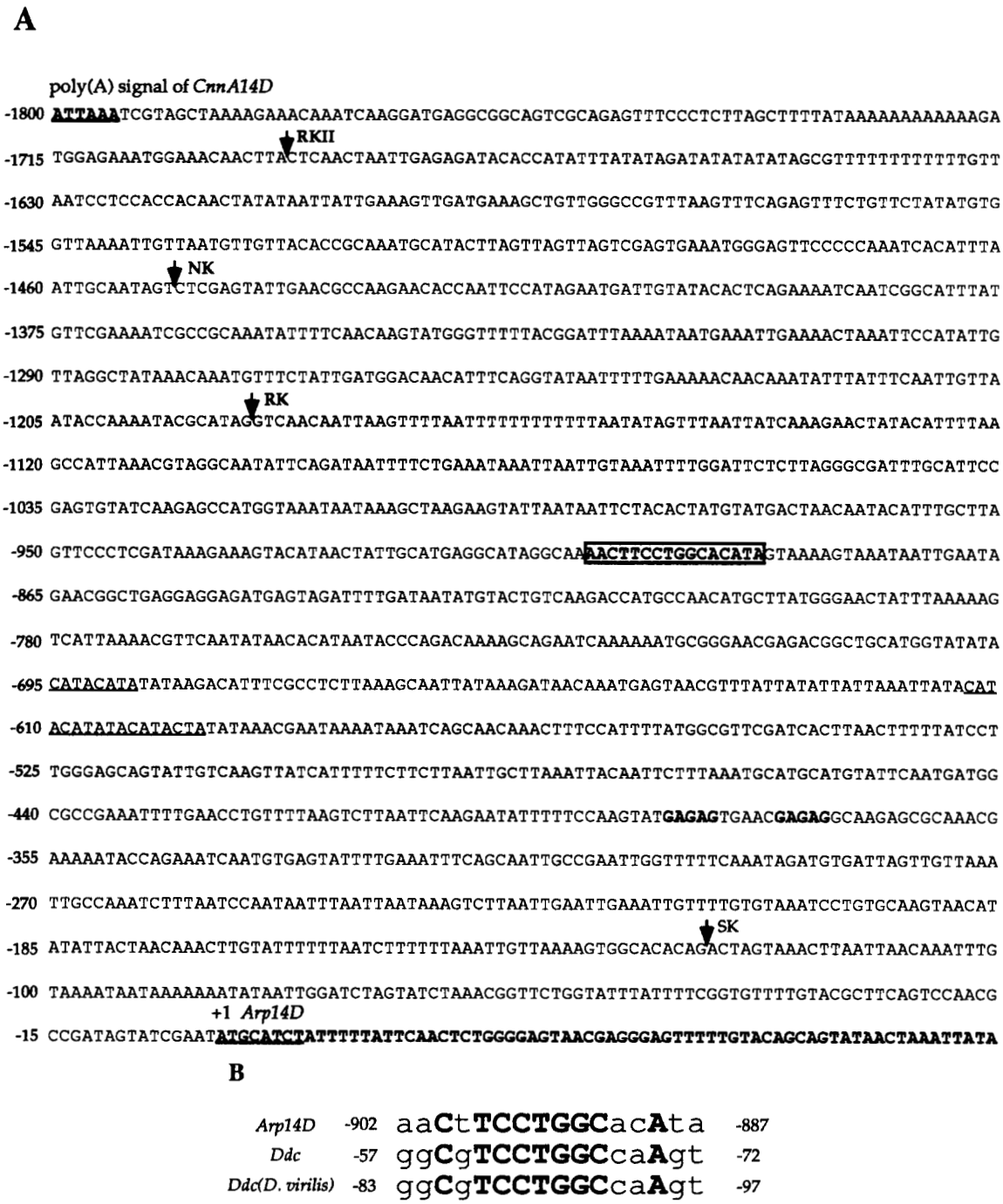

FIGURE 7.—Sequence analysis of the genomic region between the 5′ end of *Arp14D* and the poly(A) signal of *CnnA14D*. (A) The transcriptional initiation site of *Arp14D* and the putative poly(A) signal of *CnnA14D* (underlined) are shown. Arrows mark the 5′ end of the four transformation constructs (SK, 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; ROBINSON *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

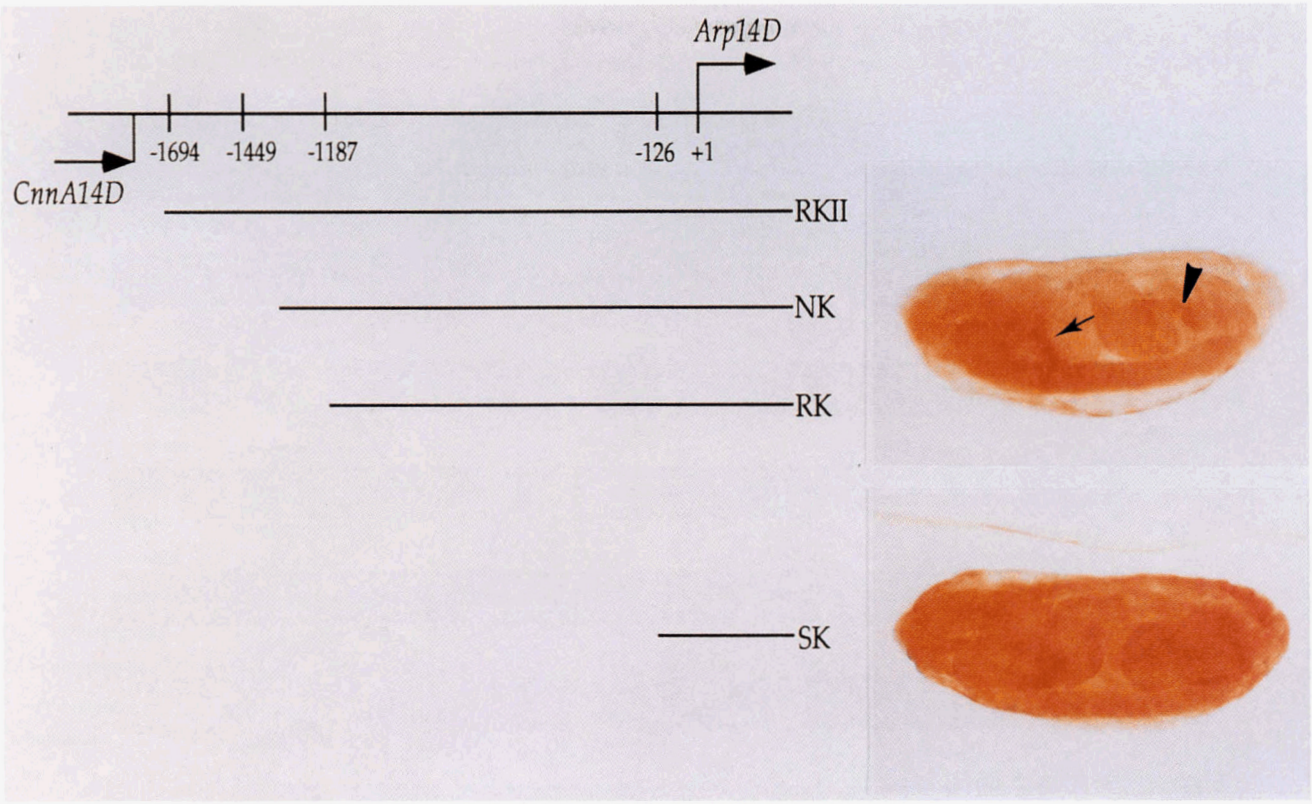

FIGURE 8.—Embryonic expression patterns of fusion constructs containing various extents of upstream DNA fused to a *lacZ* reporter. A schematic diagram of the genomic region between *Arp14D* and *CnnA14D* indicating the extents of four different fusion constructs is shown. Each fusion contains ~300 bp 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.

isoform 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 +25 downsteam and the T-rich initiation site (ARKHIPOVA 1995). Because the distance between the 5′ end of the *CnnA14D* form I transcript and the next gene

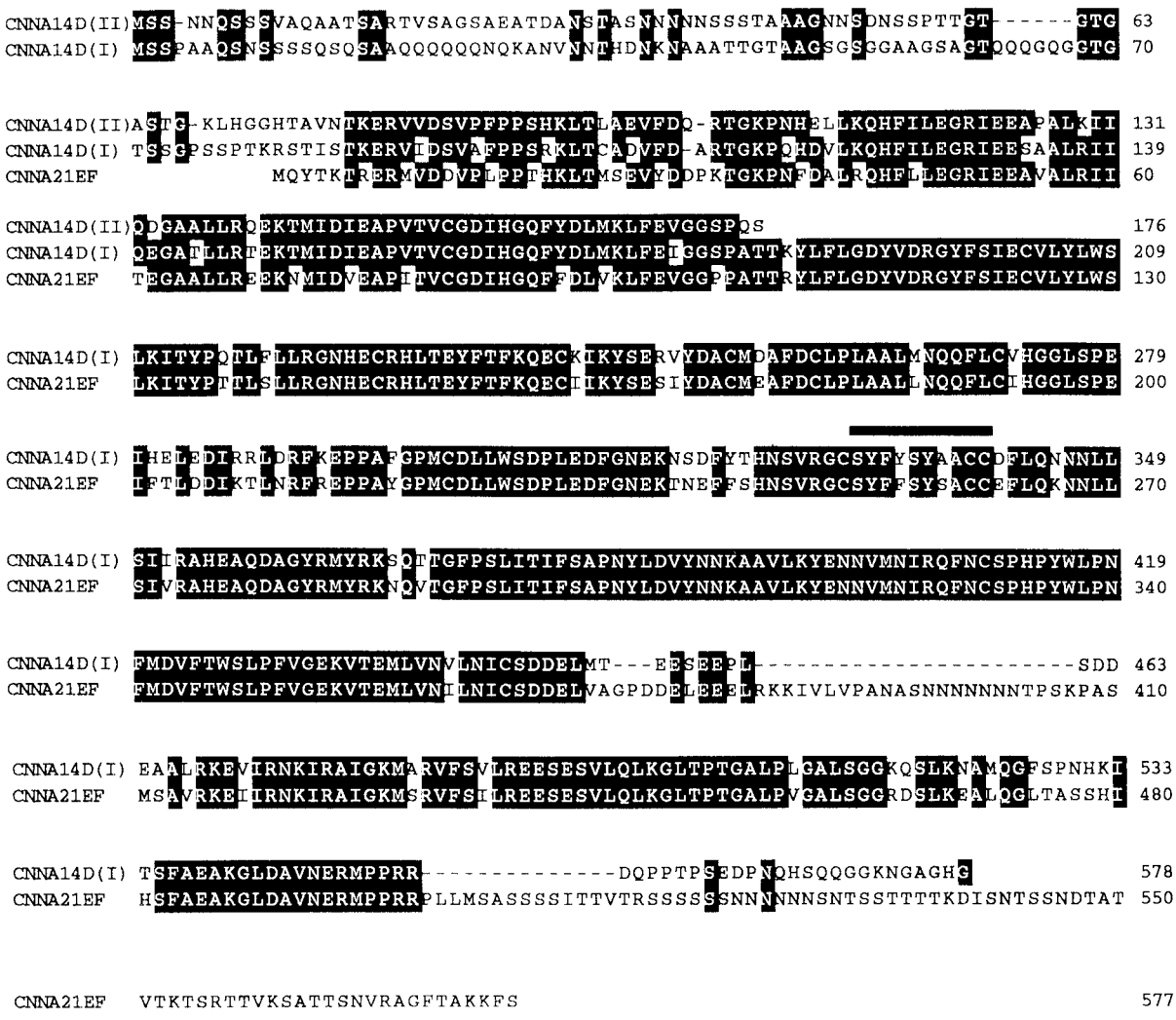

FIGURE 9.—Alignment of Drosophila calcineurin A subunits. The amino acid sequences of two alternatively spliced cDNAs, *CnnA14D(I)* and *CnnA14D(II)* are compared with another calcineurin A subunit encoded by a gene, *CnnA21EF*, mapping to polytene chromosome region 21EF. Only the first 176 amino acids of CNNA14D(II) 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-ATG-*lacZ* vector and ≥10 independent transformants were isolated. None of these transformants reproduced the endogenous expression pattern of *CnnA14D* in either young or mature embryos (data not shown). These results suggest that the regulatory elements necessary for normal expression of *CnnA14D* 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* 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 4I). These results suggest preferential expression of the gene in mitotic cells and in cells containing polytene chromosomes.

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 identifed in *C. elegans,* Xenopus, and mammals. The encoded proteins are abundant chromosome proteins required for normal chromosome condensation and chromatin 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-termini are NTP-binding sites A and B, respectively. In other proteins containing these sites such as the DEAD box 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 reasonable to anticipate that additional SMC-related genes will be identified in Drosophila including some that are more closely related to the subgroup containing the *XCap-C, Dpy-27,* and *SMC1* genes.

A

```
MVHFSGICHI MRLLATQVVA SDFVQKIAYN KMHIKQIIIQ GFKSYKDQTV   50
VEPFDKRHNV VVGRNGSGKS NFFYAIQFVL SDEFTHLRPE QRQSLLHEGT  100
GARVISAYVE IIFDNSDNRV PIDKEEIFLR RVIGAKKDQY FLNKKVVPRN  150
EVVNLLESAG FSSSNPYYIV KQGKINQMAT AADSYRLKLL REVAGTRVYD  200
ERKEESLNLL RETDSKVEKI SEYLKTIEDR LQTLEEEKEE LKEYQKWDKT  250
RRTLEYIRYE TELKDTKKAL DELQLQRKSS SDKKKIYNIE IQKAQEKIKD  300
VQKNLKEAKK KVQSTKEERS VLMTEQQQLL REKTKLDLTI VDLNDEVQGD  350
NKSKERADQE LKNLKVTIAE REKELDDVKP KYEAMKRKEE DCSRELQLKE  400
QKRKELYAKQ GRGSQFSSRE DRDKWITNEL KSISKQTRDK IAHHAKLVED  450
LKKDATSEKD LGQKIEEHSS ELEQLRLQID EHNKKYYELK KTKDQHQSMR  500
NELWRKETQM TQQLQTHKEE LSRADQALRS MAGKPILNGC DSVRKVLDSF  550
VERGGQSAEI ARAYYGPVIE NFSCDKTIYT AVEVTAANRL FHHIVESEYE  600
GTQILKEMNK LKLPGEVTFM PLNRLQVKIH DYPDDPDSIP MISKLKYDEQ  650
HDKALRYIFG KTLICRNLER ATELAKSTGL DCVTLDGDQV SSKGSLTGGY  700
FNTSRSRLEM QKKRTEYTSQ IAEFEKKLSK LRNELKSTEN NINSIVSEMQ  750
KTETKQGKSK DVFEKVQGEI RLMKEELVRI EQYRAPKERS LAQCKASLES  800
MTSTKSSLEA ELKQELMSTL SSQDQREIDQ LNDDIRRLNQ ENKEAFTQRM  850
QFEVRKNKLD NLLINNLFRR RDELIQALQE ISVEDRKRKL NNCKTELVSA  900
EKRIKKVNSD LEEIEKRVME AVQLQKELQQ ELETHVRKEK EAEENLNKDS  950
KQLEKWSTKE NMLNEKIDEC TEKIASLGAV PLVDPSYTRM SLKNIFKELE 1000
KANQHLKKYN HVNKKALDQF LSFSEQKEKL YRRKEELDIG DQKIHMLIQS 1050
LEMQKVEAIQ FTFRQVAQNF TKVFKKLVPM GAGFLILKTK DNEGDEMEKE 1100
VENSDAFTGI GIRVSFTGVE AEMREMNQLS GGQKSLVALA LIFSIQKCDP 1150
APFYLFDEID QALDAMHRKA VANMIHELSD TAQFITTTFR PELLENAHKF 1200
YGVRFRNKVS HIDCVTREEA KVFVEDDSTH A                     1231
```

B NTP-binding A site

```
XCAPE  GFKSYASRTP LNGFDPLFNAITGLNGSGKSNILDSICFLL  10-49
SCII   GFKSYASRTP IRDFDPLFNAITGLNGSGKSNILDSICFLL  10-49
SMC2   GFKSYATRTV ITDWPQQFNAITGLNGSGKSNILDAICFVL  10-49
DCAP   GFKSYKDQTV VEPFDKRHNVVVGRNGSGKSNFFYAIQFVL  41-80
XCAPC  NFKSYAGRRI LGPFHKRFSCLIGPNGSGKSNVIDSMLFVF  85-124
DPY27  NFKSYAGKHI LGPFHKNLTMLIGPNGSGKSNVIDALLFVF 100-139
SMC1   NFKSYRGVTK VGPFGESNFTSIIGPNGSGKSNMMDAISFVL  11-50
```

C Hinge domain

```
XCAPE  KAMEYVFGTTLVCDTMDNAKKVTFDKRIMTKTVTLGSDTFDPQGTLSGG 619-667
SCII   KAMEYVFGTTLVCSSMDNAKKVTFDKRIMRKTVTLQGDIFDPQGTLSGG 618-666
SMC2   KAMEPIFGNSIICSDPETAKKITFHPKIRARSIPLQGDVVDPQGTLSGG 620-668
DCAP   KALRPIFGKTLICRNLERATELA--KSTGLDCVTLDGDQVSSKGSLTGG 653-699
XCAPC  PAFYFALRDTIVANNLDDATRVAFQKDKRVRVVTLQGQIIEQSGTMTGG 700-748
DPY27  REFYFLIHDILVVDSLEEATRIDKKYPGRHRYCLNGSLLNRSGALTGG 715-763
SMC1   KAMDFVCSDSIISNTLNIAKDLKWKSGIRSKLVTIEGALLHKAGLMTGG 620-668
```

D NTP-binding B site

```
XCAPE  LSGGQRSLVALSLILAMLFKPAPLYILDEVDAALD 1086-1121
SCII   LSGGQRSLAALSLILAILLFKPAPLYILDEVDAALD 1085-1120
SMC2   LSGGQRSLTALSLIMALQQFQPAPMYILDEVDAALD 1084-1119
DCAP   LSGGQKSLVALALIFSLQKCDPAPFYLFDEIDQALD 1129-1164
XCAPC  LSGGEKTFSSLALVFALHHYKPTPLYFMDEIDAALD 1185-1220
DPY27  LSGGEKTLASLCFVFAMHHYRPTPLYVMDELDAALD 1245-1280
SMC1   LSGGEKTVAALALIFAINSYQPSPFFVLDEVDAALD 1129-1164
```

FIGURE 10.—The Drosophila *Cap* gene encodes a member of the SMC family. (A) The deduced amino acid sequence of the complete ORF is shown. Two signature sequences, the NTP-binding A site and the NTP-binding B site present in all member of the SMC protein family are underlined and in boldface. A central hinge domain (underlined) also shares significant similarity with other members of the family. (B–D) Amino acid sequences of the NTP-binding A site, the hinge domain, and the NTP-binding B sites are shown in alignment with SMC family members XCAPE (Xenopus), SCII (chicken), SMC2 (*S. cerevisiae*), XCAPC (Xenopus), DPY-27 (*C. elegans*), and SMC1 (*S. cerevisiae*). Amino acid identities are shaded. The underlined region of NTP-binding A site is a putative ATP-binding site.

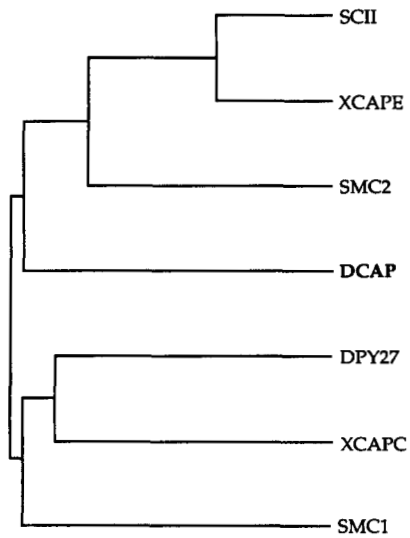

FIGURE 11.—Dendrogram showing the relationship among various proteins in the SMC family including DCAP. The horizontal branch lengths are inversely proportional to the similarity between the seequences. The distances were calculated on the basis of identities throughout the entire amino acid sequence of each protein using the DISTANCES program from the GCG software package.

Concluding remarks: Our results demonstrate the existence of a cluster of five genes, including the *para* locus, all of which are transcribed in the same direction and all of which have a neuronal expression pattern during embryonic and larval stages. In describing this collection of genes as a cluster, we are referring specifically to the tight spacing between adjacent genes in this group (*cf.* STATHAKIS *et al.* 1995). The distance between genes in this cluster ranges from <1 to 4 kb with an average of <2 kb (Figure 2). If the primary transcripts of these genes extend for hundreds of nucleotides beyond the polyadenylation signal, the gaps between genes are potentially even smaller. Consequently, ≥94% of the genomic segment of >125 kb that includes this gene cluster is transcribed. For comparison, we may consider the *Ddc* cluster, which is one of the few examples in Drosophila where a large genomic region is known to contain a very high proportion of transcribed sequences (STATHAKIS *et al.* 1995). The *Ddc* cluster includes two dense subclusters of genes, a proximal 23-kb subcluster that is 75% transcribed and a distal 27.4-kb subcluster that is 82% transcribed. However, in contrast with the *Ddc* cluster where expression of each gene appears to be regulated individually, in the *para* cluster we have been unable to identify regulatory elements in the intervals upstream of promoter regions that direct the normal expression patterns of these genes. Such regulatory elements may be contained within introns. Another possibility is that some regulatory elements may be shared by one or more genes in the cluster.

Other densely arranged gene clusters have been found in Drosophila and in most of these cases the genes in the cluster are known to be functionally and/or structurally related (*e.g.*, EVELETH and MARSH 1986; WRIGHT 1987; KNUST *et al.* 1992; FURIA *et al.* 1993; 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 answer these questions one way or the other, functional relationships between the upstream gene products and *para*-encoded sodium channels can readily be envisioned. The *para* polypeptide contains 10 putative glycosylation sites and several putative sites for phosphorylation by cAMP-dependent protein kinase or protein kinase C (LOUGHNEY *et al.* 1989; THACKERAY and GANETZKY 1994). In mammals, recent studies have shown that calcineurin modulates sodium channel activity by dephosphorylating residues that have been phosphorylated 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 important aspects of *para*-encoded sodium channel processing, 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 (KURODA *et al.* 1991). A homologue of *Cap* in nematodes is involved in dosage compensation as is the *mle* protein in Drosophila. The existence of unusual alleles of *mle*, mle^{nap}, with a temperature-sensitive paralytic phenotype associated with a block in action potential propagation (WU *et al.* 1978; GANETZKY and WU 1986) revealed that this gene somehow affected sodium channels. Subsequent studies indicated that expression of *para* is reduced in an mle^{nap} 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^{nap} 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 expression or function of sodium channels, there is still the question of why they are so tightly spaced. The polypeptides encoded by the genes in this cluster do not interact exclusively with sodium channels and must have other important biological functions. Possibly the ar-

rangement of these genes is coincidential 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 GANETZKY laboratory for their encouragement and advice, RAYMOND MIASSOD for providing cosmid clones, MICHAEL HOFFMANN, RICHARD FFRENCH-CONSTANT, SEAN CARROLL, CHING KUNG and ROBERT REENAN for comments on the manuscript. This work was supported by National Institutes of Health grant GM-43100 to B.G. This is paper 3447 from the Laboratory of Genetics, University of Wisconsin, Madison.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

AMICHOT, M., C. CASTELLA, J. B. BERG and D. PAURON, 1993 Transcriptional analysis of the *para* gene by *in situ* hybridization and immunological characterization of its expression product in wild-type and mutant strains of Drosophila. Insect Biochem. Mol. Biol. **23:** 381–390.

ARKHIPOVA, I. R., 1995 Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. Genetics **139:** 1359–1369.

BARBAS, J. A., J. GALCERAN, I. KRAH-JENTGENS, J. L. DE LA POMPA, I. CANAL *et al.*, 1991 Troponin I is encoded in the haplolethal region of the Shaker gene complex of Drosophila. Genes Dev. **5:** 132–140.

BERGERON, J. J. M., M. B. BRENNER, D. Y. THOMAS and D. B. WILLIAMS, 1994 Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. Trend Biochem. Sci. **13:** 471–474.

BROWN, L., M. X. CHEN and P. T. W. CHEN, 1994 Identification of a cDNA encoding a Drosophila calcium/calmodulin regulated protein phosphatase, which has its most abundant expression in the early embryo. FEBS Lett. **339:** 124–128.

CALA, S. E., C. ULBRIGHT, J. S. KELLY and L. R. JONES, 1993 Purification of a 90-kDa protein (Band VII) from cardiac sarcoplasmic reticulum. J. Biol. Chem. **268:** 2969–2975.

CAMPOS-ORTEGA, J. A., and V. HARTENSTEIN, 1985 *The Embryonic Development of Drosophila melanogaster.* Springer Verlag, New York.

CATTERALL, W. A., 1993 Structure and function of voltage-gated ion channels. Trends Neurosci. **13:** 201–206.

CHUANG, P.-T., D. G. ALBERTSON and B. J. MEYER, 1994 *Dpy-27*: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. Cell **79:** 459–474.

COHEN, P., 1989 The structure and regulation of protein phosphatases. Annu. Rev. Biochem. **58:** 453–508.

DE LA POMPA, J. L., 1994 Functional relationships between genes of the Shaker gene complex of Drosophila. Mol. Gen. Genet. 244: 197–204.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DOMDEY, H., B. APOSTOL, R.-J. LIN, A. NEWMAN, E. BRODY *et al.*, 1984 Lariat structures are *in vivo* intermediates in yeast pre-mRNA splicing. Cell **39:** 611–621.

ENSLEN, H., and T. R. SODERLING, 1994 Roles of calmodulin-dependent protein kinases and phosphatase in calcium-dependent transcription of immediate early genes. J. Biol. Chem. **269:** 20872–20877.

EVELETH, D. D., and J. L. MARSH, 1986 Evidence for evolutionary duplication of genes in the dopa decarboxylase region of Drosophila. Genetics **114:** 469–483.

FERRUS, A., S. LLAMAZARES, J. L. DE LA POMPA, M. A. TANOUYE and O. PONGS, 1990 Genetics analysis of the Shaker gene comlpex of *Drosophila melanogaster.* Genetics **125:** 383–398.

FRANKEL, S., M. B. HEINTZELMAN, S. ARTAVANIS-TSAKONAS and M. S. MOOSEKER, 1994 Identification of a divergent actin-related protein in Drosophila. J. Mol. Biol. **235:** 1351–1358.

FURIA, M., P. P. D'AVINO, S. CRISPI, D. ARTIAACO and L. C. POLITO, 1993 Dense cluster of genes is located at the ecdysone-regulated 3C puff of *Drosophila melanogaster.* J. Mol. Biol. **231:** 531–538.

FYRBERG, C., L. RYAN, M. KENTON and E. FYRBERG, 1994 Genes encoding actin-related proteins of *Drosophila melanogaster.* J. Mol. Biol. **241:** 498–503.

GANETZKY, B., and C.-F. WU, 1986 Neurogenetics of membrane excitability in Drosophila. Annu. Rev. Genet. **20:** 13–45.

GILCHRIST J. S. C., and G. N. PIERCE, 1993 Identification and purification of a calcium-binding protein in hepatic nuclear membranes. J. Biol. Chem. **268:** 4291–4299.

GUERINI, D., C. MONTELL and C. B. KLEE, 1992 Molecular cloning and characterization of the genes encoding the two subunits of *Drosophila melanogaster* calcineurin. J. Biol. Chem. **267:** 22542–22549.

HEIN, J., 1990 Unified approach to alignment and phylogenies. pp. 626–645 in *Method in Enzymology,* Vol. 183. Academic Press, San Diego.

HIRANO, T., and T. J. MITCHISON, 1994 A heterodimeric coiled-coil protein required for mitotic chromosome condensation *in vitro.* Cell **79:** 449–458.

HONG, C.-S., and B. GANETZKY, 1994 Spatial and temporal expression patterns of two sodium channel genes in Drosophila. J. Neurosci. **14:** 5160–5169.

HULTMARK, D., R. KLEMENZ and W. R. GEHRING, 1986 Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22.* Cell **44:** 429–438.

ISHIGURO, H., K. T. KIM, T. H. JOH and K.-S. KIM, 1993 Neuron-specific expression of the human dopamine b-hydroxylase gene requires both the cAMP-response element and a silencer region. J. Biol. Chem. **268:** 17987–17994.

JOHNSON, W. A., and J. HIRSH, 1989 Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. Nature **343:** 467–470.

KERNAN, M. J., M. I. KURODA, R. KREBER, B. S. BAKER and B. GANETZKY, 1991 *nap*^{ts}, a mutation affecting sodium channel activity in Drosophila, is an allele of *mle,* a regulator of X chromosome transcription. Cell **66:** 949–959.

KNUST, E., E. SCHRONS, F. GRAWE and J. A. CAMPOS-ORTEGA, 1992 Seven genes of the Enhancer of split complex of *Drosophila melanogaster* encode helix-loop-helix proteins. Genetics **132:** 505–518.

KURODA, M. I., M. J. KERNAN, R. KREBER, B. GANETZKY and B. S. BAKER, 1991 The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. Cell **66:** 935–947.

LIEBERMAN, D. N., and I. MODY, 1994 Regulation of NMDA channel function by endogenous Ca^{2+}-dependent phosphatase. Nature **369:** 235–239.

LIU, J.-P., A. T. R. SIM and P. J. ROBINSON, 1994 Calcineurin inhibi-

tion of dynamin I GTPase activity coupled to nerve terminal depolarization. Science **265:** 970–973.

LOUGHNEY, K., R. KREBER and B. GANETZKY, 1989 Molecular analysis of the para locus a sodium channel gene in Drosophila. Cell **58:** 1143–1154.

LUAN, S., W. LI, F. RUSNAK, S. M. ASSMANN and S. L. SCHREIBER, 1993 Immunosuppressants implicate protein phosphatase regulation of K^+ channels in guard cells. Proc. Natl. Acad. Sci. USA **90:** 2202–2206.

MANDEL, G., 1992 Tissue-specific expression of the voltage-sensitive sodium channel. J. Membr. Biol. **125:** 193–205.

MAUE, R. A., S. D. KRANER, R. H. GOODMAN and G. MANDEL, 1990 Neuron-specific expression of the rat brain type II sodium channel gene is directed by upstream regulatory elements. Neuron **4:** 223–231.

MCLEAN, J. R., R. BOSWELL and J. O'DONNELL, 1990 Cloning and molecular characterization of a metabolic gene with developmental functions in Drosophila. I. Analysis of the head function of *Punch*. Genetics **126:** 1007–1019.

MORI, N., C. SCHOENHERR, D. J. VANDENBERGH and D. J. ANDERSON, 1992 A common silencer element in the SCG10 and type II Na^+ channel genes binds a factor present in nonneuronal cells but not in neuronal cells. Neuron **9:** 45–54.

MURAMATSU, T., and R. L. KINCAID, 1993 Molecular cloning of a full-length cDNA encoding the catalytic subunit of human calmodulin-dependent protein phosphatase (calcineurin Aa). Biochim. Biophys. Acta **1178:** 117–120.

MURPHY, B. J., S. ROSSIE, K. S. D. JONGH and W. A. CATTERALL, 1993 Identification of the sites of selective phosphorylation and dephosphorylation of the rat brain Na^+ channel a subunit by cAMP-dependent protein kinase and phosphoprotein phosphatases. J. Biol. Chem. **268:** 27355–27362.

OHSAKO, S., Y. HAYASH and D. BUNICK, 1994 Molecular cloning and sequencing of calnexin-t. J. Biol. Chem. **269:** 14140–14148.

PALAZZOLO, M., B. A. HAMILTON, D. DING, C. H. MARTIN, D. A. MEAD *et al.*, 1990 Phage lamda cDNA cloning vectors for subtractive hybridization, fusion protein expression and Cre-LoxP automatic subcloning. Gene **88:** 25–36.

PAUSE, A., and N. SONENBERG, 1992 Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. **11:** 2643–2654.

PIND, S., J. R. RIORDAN and D. B. WILLIAMS, 1994 Participation of the endoplasmic reticulum charperone calnexin (p88, IP90) in the biogenesis of the cyctic fibrosis transmembrane conductance regulator. J. Biol. Chem. **269:** 12784–12788.

PONGS, O., J. LINDEMEIER, X. R. ZHU, T. THEIL, D. ENGELKAMP *et al.*, 1993 Frequnin—a novel calcium-binding protein that modulates synaptic efficacy in the Drosophila nervous system. Neuron **11:** 15–28.

PRAT, A. G., A. M. BERTORELLO, D. A. AUSIELLO and H. F. CANTIELLO, 1993 Activation of epithelial Na^+ channels by protein kinase A requires actin filaments. Am. J. Physiol. **265:** C224-C233.

RAMASWAMI, M., and M. TANOUYE, 1989 Two sodium channel genes in Drosophila: implications for channel diversity. Proc. Natl. Acad. Sci. USA **86:** 2079–2082.

ROBINSON, P. J., J.-P. LIU, K. A. POWELL, E. M. FYKSE and T. C. SÜDHOF, 1994 Phosphorylation of dynamin I and synaptic-vesicle recycling. Trends Neurosci. **17:** 348–353.

SAITOH, N., I. G. GOLDBERG, E. R. WOOD and W. C. EARNSHOW, 1994 ScII: an abundant chromosome scafold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. J. Cell Biol. **127:** 303–318.

SALKOFF, L., A. BUTLER, A. WEI, N. SCAVARDA, K. GIFFEN *et al.*, 1987 Genomic organization and deduced amino acid sequence of a putative sodium channel gene in *Drosophila*. Science **237:** 744–749.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning, A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SANTAMARIA, P., 1986 Injecting eggs, pp. 159–173 in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford, UK.

SCHOLNICK, S. B., S. J. BRAY, B. A. MORGAN, C. A. MCCORMICK and J. HIRSH, 1986 CNS and hypoderm regulatory elements of the *Drosophila melanogaster* dopa decarboxylase gene. **234:** 998–1002.

SCHREIBER, S. L., 1992 Immunophilin-sensitive protein phosphatase action in cell signaling pathways. Cell **70:** 365–368.

SCHREIBER, K. L., M. P. BELL, C. J. HUNTOON, S. RAJAGOPALAN, M. B. BRENNER *et al.*, 1994 Class II histocompatibility molecules associate with calnexin during assembly in the endoplasmic reticulum. Int. Immunol. **6:** 101–111.

SCHROER, T. A., E. FYRBERG, J. A. COOPER, R. H. WATERSON, D. HELFMAN *et al.* 1994 Actin-related protein nomenclature and classification. J. Cell Biol. **127:** 1777–1778.

SCHWOB, E., and R. P. MARTIN, 1992 New yeast actin-like gene required late in the cell cycle. Nature **355:** 179–183.

SMITH, M. J., 1992 Nucleotide sequence of a *Drosophila melanogaster* gene encoding a calreticulin homologue. DNA Seq. **3:** 247–250.

SMITH, M. J., and G. L. E. KOCH, 1989 Multiple zones in the sequence of calreticulin (CRP55, calregulin, HACBP), a major calcium binding ER/SR protein. EMBO J. **8:** 3581–3586.

SPRADLING, A. C., 1986 P element mediated transformation, pp. 175–196 in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford, UK.

SRINIVASAN, Y., M. LEWALLEN and K. J. ANGELIDES, 1992 Mapping the binding site on ankyrin for the voltage-dependent sodium channel from brain. J. Biol. Chem. **267:** 7483–7489.

STATHAKIS, D. G., E. S. PENTZ, M. E. FREEMAN, J. KULLMAN, G. R. HANKINS *et al.*, 1995 The genetic and molecular organization of the *dopa decarboxylase* gene cluster of *Drosophila melanogaster*. Genetics **141**:629–655.

STERN, M., R. KREBER and B. GANETZKY, 1990 Dosage effects of a Drosophila sodium channel gene on behavior and axonal excitability. Genetics **124:** 133–143.

STRUNNIKOV, A. V., V. L. LARIONOV and D. KOSHLAND 1993 *SMC1*: an essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family. J. Cell Biol. **123:** 1635–1648.

STRUNNIKOV, A. V., E. HOGAN and D. KOSHLAND 1995 *SMC2*, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. Genes Dev. **9:** 587–599.

SURDEJ, P., G. CLAUDE and R, MIASSOD, 1990 Developmental expression pattern of a 800 kb DNA continuum cloned from the Drosophila X chromosome 14B-15B region. Biol. Cell **68:** 105–118.

THACKERAY, J. R., and B. GANETZKY, 1994 Developmental regulated alternative splicing generates a complex array of Drosophila *para* sodium channel isoforms. J. Neurosci. **14:** 2569–2578.

THACKERAY, J. R., and B. GANETZKY, 1995 Conserved splicing patterns and splicing signals in the Drosophila sodium channel gene *para*. Genetics **142:** 203–214.

THUMMEL, C. S., A. M. BONLET and H. D. LIPSHITZ, 1988 Vectors for Drosophila P-element-mediated transformation and tissue culture transfection. Gene **74:** 445–456.

TSENG-CRANK, J., J. A. POLLOCK, I. HAYASHI and M. A. TANOUYE, 1991 Expression of ion channel genes in Drosophila. J. Neurogenet. **7:** 229–239.

WADA, I., D. RINDRESS, P. H. CAMERON, W.-J. OU, J. J. DOHERTY II *et al.* 1991 SSRa and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. J. Biol. Chem. **266:** 19599–19610.

WATANABE, D., K. YAMADA, Y. NISHINA, Y. TAJIMA, U. KOSHIMIZU *et al.*, 1994 Molecular cloning of a novel Ca^{2+}-binding protein (calmegin) specifically expressed during male meiotic germ cell development. J. Biol. Chem. **269:** 7744–7749.

WRIGHT, T. R. F., 1987 The genetic and molecular organization of the dense cluster of functionally related, vital genes in the dopa decarboxylase region of the *Drosophila melanogaster* genome, pp. 95–120 in *Eukaryotic Chromosomes: Structure and Function, Vol. 14. Results and Problems in Cell Differentiation*, edited by W. HENNIG. Springer-Verlag, Heidelberg.

WU, C.-F., B. GANETZKY, Y. N. JAN, L. Y. JAN and S. BENZER, 1978 A Drosophila mutant with a temperature-sensitive block in nerve conduction. Proc. Natl. Acad. Sci. USA **75:** 4047–4051.

Communicating editor: R. S. HAWLEY