high affinity" and "low and high salt eluting" because under the conditions at which Sp1-327C, Sp1-236C, and Sp1-168C elute from the column, the Sp1-778C and Sp1-516C proteins remain bound to the resin. In other words, the high salt-eluting proteins have a higher affinity to DNA than the low salt-eluting proteins at salt concentrations at which the low salt forms are eluted from the resin. These experiments do not directly address, however, the relative affinity of the larger versus the smaller Sp1 species at low salt concentrations where both species bind.


19. A. D. Johnson, B. J. Meyer, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5061 (1979); M. Ptashne *et al.*, *Cell* **19**, 1 (1980).
20. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986); G. Gill and M. Ptashne, *ibid.* **51**, 121 (1987); E. Giniger and M. Ptashne, *Nature* **330**, 670 (1987); J. Ma and M. Ptashne, *Cell* **48**, 847 (1987); *ibid.* **51**, 113 (1987).


21. We thank W. Lim for construction of the Sp1 cDNA extension libraries, S. P. Jackson for the Sp1 antibodies, and J. Brosius for the plasmid vector. We are grateful to K. Perkins, K. Arndt, and S. Jackson for improving the quality of this manuscript. We also thank K. Ronan for assistance in the preparation of the figures. J.T.K. is a Lucille P. Markey Scholar in Biomedical Science. A.J.C. is supported by a postdoctoral fellowship from the American Cancer Society. Supported in part by a research grant from the NCI to R.T.

15 July 1988; accepted 5 October 1988


# The *elav* Gene Product of *Drosophila*, Required in Neurons, Has Three RNP Consensus Motifs


STEVEN ROBINOW, ANA REGINA CAMPOS, KWOK-MING YAO, KALPANA WHITE



**A sequence of developmental events transforms neurons from their immature state to their mature, terminally differentiated state. The *elav* locus is one of the first examples of a gene that is expressed in neurons early during this developmental sequence. This gene has been shown to be required for the proper development of young neurons and for the maintenance of mature neurons. DNA sequence data presented in this report suggest that the *elav* gene product is an RNA binding protein, based on the presence of RNP (ribonucleotide) consensus sequences. This leads to the proposal that this protein is involved in the RNA metabolism of neurons.**


AS A NEWLY BORN NEURON MATURES, it must alter its morphology and physiology in order to become a functional component of a nervous system. To investigate the genetic basis of these events of neuronal maturation we studied the locus *embryonic lethal, abnormal visual system (elav)* of the fruit fly *Drosophila melanogaster*. Genetic and molecular data concerning this locus suggest that the *elav*-encoded function is required in all neurons from their birth, throughout their maturation, and during their maintenance (*1–5*). Transcripts from this locus are expressed in young neurons but not in neuroblasts, the neuronal progenitor cells (*5*). Thus, transcription of *elav* is one of the first demonstrated molecular events in the process of neuronal maturation.

To gain insight into the molecular nature of the product or products of this locus, we determined the nucleotide sequence of a 2.5-kb embryonic cDNA clone (*elav* cDNA-1) (*3*) and of a 9.2-kb genomic DNA fragment. An analysis of these data reveal an open reading frame (ORF) that can code for a 50.76-kD protein. We present molecular data that show that this ORF encodes a product of the *elav* locus. An analysis of the deduced amino acid sequence of this ORF reveals three repeats of two previously defined consensus sequences; an octapeptide


Department of Biology, Brandeis University, Waltham, MA 02254-9110.


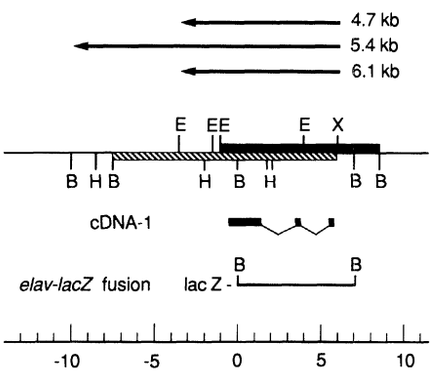


**Fig. 1.** Genomic organization of the *elav* locus. Above the genomic DNA restriction map, each embryonic transcript is drawn such that the underlying genomic region detects that transcript by RNA analysis (*3*). The maps of the 5.4-kb and 6.1-kb transcripts have been revised from the map shown in a previous paper (*3*). The black bar on the genomic map represents the 9.2-kb of genomic DNA sequenced. The hatched region of the genomic map indicates a fragment that contains sufficient information to provide the *elav* function required for viability (*3*). The splicing pattern of cDNA-1 is shown below the genomic map. Also diagrammed is an *elav-lacZ* gene fusion that has been constructed such that the *lacZ* gene is fused at the Bam HI site (genomic position −1) in frame with the 1449-nucleotide ORF (*16*). Transcription from this construct is driven from the *elav* promoter. The Elav-LacZ fusion protein is predicted to have a molecular size of 161 kD on the basis of an analysis of the DNA sequence of the fusion gene. The genomic DNA and cDNA-1 were subcloned into appropriate vectors [Bluescript (Stratagene) and pGEM1 (Promega)] and sequenced by established techniques (*23*). In all cases, the sequence was determined for both strands.

```
  1  MDFIMANTGA GGGVDTQAQL MQSAAAAAAV AATNAAAAPV QNAAAVAAAA QLQQQQVQQA
 61  ILQVQQQQTQ QAVAAAAAAV TQQLQQQQQA VVAQQAVVQQ QQQQAAAVVQ QAAVQQAVVP
121  QPQQAQPNTN GNAGSGSQNG SNGSTETRTN LIVNYLPQTM TEDEIRSLFS SVGEIESVKL
                                       ******
181  IRDKSQVYID PLNPQAPSKG QSLGYGFVNY VRPQDAEQAV NVLNGLRLQN KTIKVSFARP
                                 ••••••••
241  SSDAIKGANL YVSGLPKTMT QQELEAIFAP FGAIITSRIL QNAGNDTQTK GVGFIRFDKR
           *  *****                                          •  •••••••
301  EEATRAIIAL NGTTPSSCTD PIVVKFSNTP GSTSKIIQPQ LPAFLNPQLV RRIGGAMHTP
361  VNKGLARFSP MAGDMLDVML PNGLGAAAAA ATTLASGPGG AYPIFIYNLA PETEEAALWQ
                                                  ******
421  LFGPFGAVQS VKIVKDPTTN QCKGYGFVSM TNYDEAAMAI RALNGYTMGN RVLQVSFKTN
                                 ••••••••
481  KAK°
```

**Fig. 2.** Conceptual translation of the 1449-nucleotide ORF. A single letter representation of the 483–amino acid protein is shown. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine. This protein has a predicted molecular size of 50,760 daltons. Stretches of alanines (A) and glutamines (Q) are found between amino acids 24 and 126. Seventy-three percent of these residues are either alanine or glutamine. ····· Underlie the residues of the RNP1 motif. ****** Underlie the residues of the RNP2 motif. ° Indicates a stop codon that terminates the ORF.

consensus sequence ($^{K}_{R}$GF$^{G}_{A}$FVX$^{F}_{Y}$), termed RNP1, and a hexapeptide consensus sequence, termed RNP2 (*6, 7*). RNP2 is less conserved than RNP1 but is characterized by its aliphatic and aromatic nature and its position relative to RNP1 (*7*). These data suggest that the Elav protein is involved in the RNA metabolism of neurons.

The sequenced 9.2-kb genomic fragment and the exon map of cDNA-1 relative to the genomic map of the *elav* region (*8*) are shown in Fig. 1. The 5′ end of cDNA-1 maps to nucleotide 5292 within the 9.2-kb genomic fragment. Two intervening sequences must be spliced from a primary transcript to generate cDNA-1 from the genomic sequences. These introns are 1289 and 2208 nucleotides long, respectively. At the exon-intron boundaries, sequences are present that correspond to the consensus sequences for the splice junctions of the 5′ intervening sequence [*elav*: AAT/GTAAGA and CAA/GTGAGT; consensus: $^{C}_{A}$AG/GT$^{A}_{G}$AGT (*9*)], and the 3′ intervening sequence [*elav*: ATCAG/A and CTCAG/T; consensus: $(^{T}_{C})_{n}$N$^{C}_{T}$AG/G (*9*)].

Within the sequenced genomic fragment, multiple copies of two transcriptionally relevant sequence motifs are observed (*10, 11*).

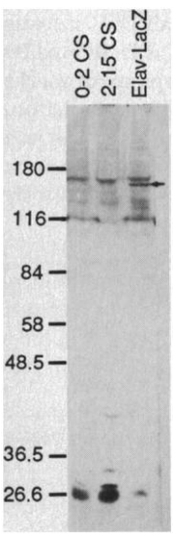


**Fig. 3.** Embryos carrying the *elav-lacZ* gene fusion produce a high molecular weight protein detectable by immunoblot analysis with the use of antibodies to β-galactosidase. Lanes: 0–2 CS, protein of 0- to 2-hour Canton-S embryos; 2–15 CS, protein of 2- to 15-hour Canton-S embryos; Elav-LacZ, protein of 0- to 21-hour embryos that carry the *elav-lacZ* gene fusion. Molecular size standards are indicated (in kilodaltons). The arrow indicates the high molecular weight protein found in the embryos carrying the fusion gene. This protein is not detected in Canton-S embryos. The other bands seen in all lanes are *Drosophila* proteins that cross-react with the primary antibody. Approximately equal amounts of protein have been loaded in each lane based on duplicate gels stained with Coomassie brilliant blue. Embryos were collected, dechorionated in 50% Clorox, and homogenized in buffer containing phosphate-buffered saline, 0.5% NP-40, 1 m*M* EGTA, and 1 m*M* phenylmethylsulfonyl fluoride. The protein of these embryos was then separated by polyacrylamide gel electrophoresis (*24*) and transferred to a nitrocellulose membrane (*25*). Using a rabbit primary antibody to β-galactosidase (Cappel; 1:1000) and an alkaline phosphatase–conjugated secondary antibody to rabbit IgG (Sigma, 1:7500), we processed this immunoblot as described by Promega.

| | CONSENSUS | $^{K}_{R}$ G F $^{G}_{A}$ F V X $^{F}_{Y}$ | |
|---|---|---|---|
| 203 | L G Y G F V N Y | V R P Q D A E Q A V N V L N G |
| 290 | K G V G F I R F | D K R E E A T R A I I A L N G |
| 443 | K G Y G F V S M | T N Y D E A A M A I R A L N G |

**Fig. 4.** Conserved residues are located near the RNP1 motifs. The numbers at the left denote the position of the first residue of each RNP1 repeat in the Elav protein (Fig. 2). The three regions of the Elav protein that contain the RNP1 motif are boxed. Above these sequences, the RNP1 consensus sequence is displayed (*6*). The underlined letters highlight residues that are outside the RNP1 and RNP2 motifs and are still conserved in all three Elav repeats.

Exact repeats of the GAGA motif are located at nucleotide 5401, which is 109 nucleotides upstream from the 5′ end of cDNA-1, and in the second intron at nucleotide 2075. Eleven copies of the sequence $^{C}_{T}$GAG$^{C}_{T}$G, which can be bound by the Zeste protein, are located throughout the 9.2-kb genomic fragment (*12*). Another feature of the fragment is a 30-nucleotide region, 1.1 kb upstream from the 5′ end of cDNA-1, which extends from nucleotide 6405 to nucleotide 6376. This sequence, ATACATACATATACACATATATGTATGTAT, has a perfect 13-bp inverted repeat around a four-nucleotide CACA core. Dyad (rotational) symmetry is a characteristic of sequences bound by certain transcription factors (*13*). The significance of these motifs for *elav* transcription is not known.

Although cDNA-1 is only a partial cDNA [the shortest transcript being approximately 4.7 kb (Fig. 1) (*3*)], it contains a complete ORF that starts 492 nucleotides from its 5′ end. The initial ATG methionine codon is generated when the second splice joins an A nucleotide from the 5′ splice junction with a TG nucleotide pair at the 3′ splice junction. This ATG is embedded in a sequence (AACAATG), which is in good agreement with the *Drosophila* consensus sequence for the site of translation initiation as defined by Cavener ($^{C}_{A}$AA$^{A}_{C}$ATG) (*14*). The region immediately downstream from the ORF is rich in stop codons in all reading frames (*8*).

This 1449-nucleotide ORF potentially encodes a 483–amino acid protein with a calculated molecular size of 50.76 kD (Elav protein) (Fig. 2) (*8*). As predicted, RNA transcribed in vitro from cDNA-1 produces a protein of approximately 50 kD when translated in a reticulocyte lysate (*15*). An independent *elav* cDNA clone from an adult head cDNA library (cDNA-2) (*16*) contains the same 1449-nucleotide ORF and produces a 50-kD protein on in vitro translation (*17*), indicating that transcripts containing this ORF are present in adults as well.

Two independent lines of evidence show that this *elav* ORF is used in vivo. First, an *elav-lacZ* gene fusion (Fig. 1) is expressed in flies transformed with a P-element vector that contains this gene fusion (*16*). As expected, β-galactosidase activity is detected in the developing embryonic nervous system (*16*). Immunoblot analysis with antibodies to β-galactosidase revealed that a protein of the predicted molecular size was present only in the transformed line (Fig. 3). We conclude that the β-galactosidase protein is being translated as an Elav–β-galactosidase fusion protein. Therefore, the predicted ORF is translated in vivo. Second, Bier *et al.* (*18*) cite data that suggest that a 50-kD neuron-specific nuclear antigen recognized by the monoclonal antibody MAb44C11 may be a product of the *elav* locus. That cDNA-1 contains an ORF that codes for a 50-kD protein is consistent with their proposal. Moreover, we have evidence that MAb44C11 recognizes an epitope contained within the Elav protein (*19*).

A search of the Protein Identification Resource (PIR) database revealed that the Elav protein is similar to rat helix destabilizing protein and to yeast polyadenylate binding protein (*20*). The region of similarity spans both the RNP1 and RNP2 consensus sequences found in these and other RNA binding proteins (*6, 7, 20, 21*). A feature shared by these RNA binding proteins is that the RNP2 motif is located approximately 30 residues toward the amino-terminus from the RNP1 motif (*7*). These motifs are part of a larger domain of 80 to 100 residues that are often repeated in any given protein (*6, 7, 20, 21*). These characteristics are also displayed by the Elav protein in which these motifs are repeated three times (Figs. 2 and 4). Other examples of proteins that contain these consensus sequences have also been observed in *Drosophila* (*22*). Because the Elav protein contains three copies each of the RNP1 and RNP2 motifs, and because these motifs are juxtaposed in a manner consistent with that of other RNA binding proteins (*7, 20, 21*), we infer that this product of the *elav* gene is an RNA binding protein also.

The repeated domains within the Elav protein show similarity to each other outside of the RNP motifs. Of particular interest are the identical amino acids in Fig. 4. Although the physiological properties of these residues are not known, the fact that they and their position with respect to the RNP1 consensus sequence are conserved in all three repeats within the Elav protein suggests that this is a functionally significant region.

Finally, we consider the role that the Elav protein may play in the development of the neuron. The *elav* locus is not expressed in neuroblasts but is expressed in young neurons (*5*), and it continues to be expressed in neurons through all developmental stages (*5*). Although the *elav* gene function is not required to generate neurons, it is essential for the formation of a structurally normal embryonic central nervous system (*4*) and adult retina and optic lobe (*1, 2*). Given the probable RNA-binding capability of the Elav protein, we propose that the product (or products) of this locus provides a function essential for the RNA metabolism of all neurons. More specifically, we propose that the Elav protein is required for the proper posttranscriptional processing of transcripts of other genes that participate in neuronal maturation and maintenance.


REFERENCES AND NOTES

1. A. R. Campos, D. Grossman, K. White, *J. Neurogenet.* **2**, 197 (1985).
2. T. Homyk, Jr., K. Isono, W. L. Pak, *ibid.*, p. 309.
3. A. R. Campos, D. R. Rosen, S. N. Robinow, K. White, *EMBO J.* **6**, 425 (1987).
4. F. Jiménez and J. A. Campos-Ortega, *J. Neurogenet.* **4**, 179 (1987).
5. S. Robinow and K. White, *Dev. Biol.* **126**, 294 (1988).
6. M. S. Swanson, T. Y. Nakagawa, K. LeVan, G. Dreyfuss, *Mol. Cell. Biol.* **7**, 1731 (1987).
7. G. Dreyfuss, M. S. Swanson, S. Piñol-Roma, *Trends Biochem. Sci.* **13**, 86 (1988).
8. The sequence of the 9.2-kb genomic DNA fragment and cDNA-1 are available through GenBank under accession nos. M21152 and M21153, respectively, or on request to the authors. The translation of the 1449-nucleotide ORF is available through the National Biomedical Research Foundation (NBRF)–PIR database under accession no. A30030.
9. R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853 (1978); S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
10. The GAGA sequence motif is bound by a factor that activates transcription of the *Drosophila engrailed* and *Ubx* genes [W. C. Soeller, S. J. Poole, T. Kornberg, *Genes Dev.* **2**, 68 (1988); M. D. Biggin and R. Tjian, *Cell* **53**, 699 (1988)].
11. The sequence $^{C}_{T}$GAG$^{C}_{T}$G is recognized by the Zeste protein, which activates transcription of the *Ubx* promoter [M. D. Biggin, S. Bickel, M. Benson, V. Pirrotta, R. Tjian, *Cell* **53**, 713 (1988)].
12. The 11 copies of the sequence $^{C}_{T}$GAG$^{C}_{T}$G, or its inverse, are distributed as follows: three are located 2.6-kb to 3.1-kb upstream from the 5′ end of cDNA-1, one begins ten nucleotides from the end of the first exon of cDNA-1, three more copies are clustered in the first intron within 165 nucleotides of each other, another three copies are dispersed in the second intron, and the last copy is in the third exon of cDNA-1 at postion −81.
13. J. Topol, D. M. Ruden, C. S. Parker, *Cell* **42**, 527 (1985); M. Ptashne, *A Genetic Switch: Gene Control and Phage Lambda* (Blackwell, Boston, 1986).
14. D. R. Cavener, *Nucleic Acids Res.* **15**, 1353 (1987).
15. S. Robinow, unpublished data.
16. A. R. Campos, thesis, Brandeis University, Waltham, MA (1988).
17. The *elav* cDNA-2 was isolated from an adult head cDNA library (provided by Y. Citri). DNA sequence analysis demonstrates that this cDNA contains the same 1449-nucleotide ORF as cDNA-1 (S. Robinow, unpublished observation). RNA transcribed from this cDNA clone was translated in vitro in a reticulocyte lysate and the protein generated comigrates with the protein product of cDNA-1 (S. Robinow, unpublished data).
18. E. Bier, L. Ackerman, S. Barbel, L. Jan, Y. N. Jan, *Science* **240**, 913 (1988).
19. By immunoblot analysis, MAb44C11 recognizes two different fusion proteins produced in *Escherichia coli*. Both of these fusion proteins express the 1449-nucleotide ORF. The only sequences common to these two proteins are those of the 1449-nucleotide ORF (S. Robinow, unpublished data).
20. S. A. Adam, T. Nakagawa, M. S. Swanson, T. K. Woodruff, G. Dreyfuss, *Mol. Cell. Biol.* **6**, 2932 (1986); A. B. Sachs, M. W. Bond, R. D. Kornberg, *Cell* **45**, 827 (1986); F. Cobianchi, D. N. SenGupta, B. Z. Zmudzka, S. H. Wilson, *J. Biol. Chem.* **261**, 3536 (1986).
21. B. Lapeyre *et al.*, *J. Biol. Chem.* **261**, 9167 (1986); D. K. Lahiri and J. O. Thomas, *Nucleic Acids Res.* **14**, 4077 (1986); T. Grange, C. Martins de Sa, J. Oddos, R. Pictet, *ibid.* **15**, 4771 (1987); A. Y.-S. Jong, M. W. Clark, M. Gilbert, A. Oehm, J. L. Campbell, *Mol. Cell. Biol.* **7**, 2947 (1987).
22. S. R. Haynes, M. L. Rebbert, B. A. Mozer, F. Forquignon, I. B. Dawid, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1819 (1987); L. R. Bell, E. M. Maine, P. Schedl, T. W. Cline, *Cell*, in press; T. Goralski and B. Baker, personal communication.
23. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); S. Henikoff, *Methods Enzymol.* **155**, 156 (1987). Sequence analysis was accomplished with the DNA Inspector II+ DNA analysis program (Textco).
24. U. K. Laemmli, *Nature* **227**, 680 (1970).
25. L. G. Davis, M. D. Dibner, J. F. Battey, *Basic Methods in Molecular Biology* (Elsevier, New York, 1986), pp. 311–314.


26. We thank E. Bier, L. Jan, and Y. N. Jan for generously supplying monoclonal antibody MAb44C11, and M. Rosbash for getting us started on the protein analysis. We are grateful to L. L. Restifo and J. C. Hall for comments and suggestions concerning this manuscript. Supported by NIH grant GM-33205.

16 August 1988; accepted 19 October 1988


# Lineage-Specific Development of Calcium Currents During Embryogenesis


LUCIANA SIMONCINI, MELODYE L. BLOCK, WILLIAM J. MOODY



**The development of electrophysiological properties of isolated, identified ascidian blastomeres was followed from the fertilized egg to the neurula, and the stage at which cells of different lineages first express different functional ion channel populations was determined. Little has been known about such events because of the difficulties of making voltage-clamp recordings from small embryonic cells and of identifying their developmental fates in dissociated preparations. The problem of small cell size was circumvented by using the whole-cell patch clamp, and identification was facilitated by the use of a species of ascidian, *Boltenia villosa*, in which endogenous pigment marks cells of specific developmental fates. Within approximately 3 hours after gastrulation, muscle-lineage blastomeres in these embryos developed a voltage-dependent calcium current while surrounding blastomeres of other lineages did not. At about the same time, all cells developed delayed outward potassium currents and lost the inwardly rectifying potassium currents present at earlier stages.**


THE MECHANISMS BY WHICH CELLS of different developmental fates acquire their characteristic electrical properties during embryogenesis are poorly understood. Later stage blastomeres are difficult to isolate and record from if conventional microelectrode techniques are used. It is also difficult in most preparations to identify the developmental fates of cells that have been separated from the embryo. A number of earlier studies have used cytochalasin B to


Department of Zoology, University of Washington, Seattle, WA 98195.


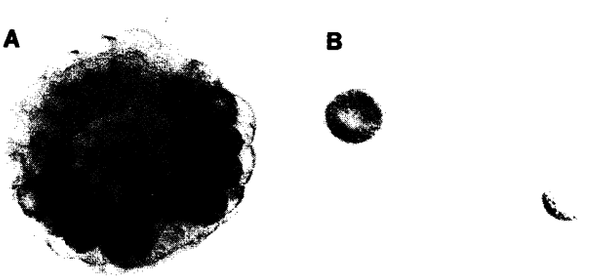


**Fig. 1.** (**A**) Photograph of a *Boltenia* embryo at the gastrula stage, approximately 13 hours after fertilization. The embryo is viewed from the dorsal aspect, showing the blastopore, with the left-right axis of symmetry running from upper left (anterior) to lower right (posterior). The diameter of the gastrula is about 150 μm. Muscle-lineage cells lie along the posterior rim of the blastopore and are easily recognized by their endogenous orange pigment. (**B**) Two cells dissociated from the gastrula and photographed under differential interference contrast optics. The muscle-lineage cell can be distinguished by its orange color but is otherwise similar in morphology to its nonmuscle neighbor. The cells are about 30 μm in diameter.