

The *Drosophila* Gene *rbp9* Encodes a Protein That Is a Member of a Conserved Group of Putative RNA Binding Proteins That Are Nervous System–Specific in Both Flies and Humans

Young-Joon Kim and Bruce S. Baker

Department of Biological Sciences, Stanford University, Stanford, California 94305

The *rbp9* gene of *Drosophila melanogaster* has been molecularly characterized and shown to be expressed solely in the CNS, where it encodes proteins with three RNA recognition motifs (RRMs). Sequencing of genomic and cDNA clones of *rbp9* revealed a complex gene with three alternative promoters as well as alternative patterns of splicing. The deduced amino acid sequence of the RBP9 proteins is highly similar to those of three other nervous system–specific genes, human HuC and HuD and *Drosophila elav*, which also encode proteins with three RRMs. Developmental Northern analysis revealed that *rbp9* is expressed from the late third instar larva through adult stages. The RBP9 protein was detected specifically in nuclei of the nervous system after morphogenesis of the adult CNS in the mid-pupal stage. Thus, the RBP9 protein does not appear until substantially later than *rbp9* transcripts are detected. The adult nervous system nuclear-limited expression pattern, the presence of RRMs, and the high similarity to a group of nervous system–specific proteins in flies and humans suggest that *rbp9* belongs to a nervous system–specific RRM protein gene subfamily that may participate in the processing of RNAs involved in the development of the CNS.

[Key words: RNA recognition motif (RRM), CNS, nuclear protein, neural RRM protein]

Substantial postembryonic neurogenesis is required for the synthesis of the adult CNS in *Drosophila*. In particular, during the transition from larva to adult, the number of neurons in the CNS increases drastically. Large numbers of new neurons are produced from a small number of larval neuronal stem cells. These neurons are growth arrested during the early stages of development only to differentiate into adult neurons at the onset of metamorphosis (Truman and Bate, 1988). As these new neurons mature, their physiological as well as morphological nature changes. These physiological and morphological changes accompany specific changes in gene expression. Therefore, it is apparent that a tightly orchestrated regulation of gene expression is required for proper CNS development.

Gene expression in the developing embryonic CNS has been shown to be regulated by a number of homeotic genes, which

encode putative transcription factors (Doe and Scott, 1988). However, the discovery of nervous system–specific genes with RNA recognition motifs (RRMs)—a region of 80–90 amino acids, across which there are 21 conserved residues and which have been shown to be the RNA binding site in several proteins with this motif (for reviews, see Dreyfuss et al., 1988; Bandziulis et al., 1989; Mattaj, 1989; Kenan et al., 1991)—such as *elav* of *Drosophila* (Robinow et al., 1988), *nrv-1* of *Xenopus* (Richter et al., 1990), and two human nervous system–specific proteins, HuD (Szabo et al., 1991) and HuC (H. M. Furneaux, personal communication), raises the possibility that the neurogenesis may also be regulated at the level of RNA metabolism.

The involvement of differential RNA processing mechanisms in the regulation of development has been shown in many cases. For example, the sex determination genes of *Drosophila* (reviewed in Baker, 1989), the rat tropomyosin gene (Breitbart and Nadal-Ginard, 1987), several viral genes (Berk and Sharp, 1978; Ziff, 1982), and the calcitonin/calcitonin gene–related peptide gene (Rosenfeld et al., 1984) make different proteins in different tissues by regulating alternative splicing. In particular, the tissue-specific production of neuropeptides in the nervous system (Kitamura et al., 1983; Rosenfeld et al., 1984) strongly indicates that nervous system development may also be controlled at the RNA processing level.

Our previous studies on the RRM-type RNA binding protein gene family in *Drosophila* (Kim and Baker, in press) identified several RRM sequences that have a high similarity to tissue-specific RRM-containing proteins. Among them, RRM2, RRM9, and RRM10 have a high similarity to a group of nervous system–specific proteins. Therefore, we characterized the gene encoding one of them, RRM9, with the hope of finding links between nervous system development and RNA processing.

Here we report a detailed molecular characterization of the structure and expression pattern of the gene *rbp9* that encodes RRM9. Our results show that the expression of *rbp9* is highly regulated both temporally and spatially. We also show that the RBP9 protein is expressed only in the nuclei of the nervous system after the mid-pupal stage and that it has a high sequence similarity to human and *Drosophila* nervous system–specific proteins that also contain RRMs. These results suggest that *rbp9* may be involved in the development of the nervous system perhaps by acting as a regulator of nervous system–specific RNA processing decisions.

Materials and Methods

Screening of cDNA and cosmid libraries. Recombinant phages (2×10^5) were screened from each of the following libraries: a λ gt10 adult fly and

Received July 23, 1992; revised Sept. 11, 1992; accepted Sept. 16, 1992.

We thank Wendy Wall, Lisa Ryner, and Claudia Zraly for comments on the manuscript. We also thank Gennet Bohm for preparing media and laboratory supplies. This work was supported by a grant from the National Institutes of Health.

Correspondence should be addressed to Bruce S. Baker at the above address.
Copyright © 1993 Society for Neuroscience 0270-6474/93/131045-12\$05.00/0

a pupa cDNA library (provided by L. Kauvar, USCF), a λ nvx third instar larva cDNA library (provided by S. Elledge, Baylor College of Medicine), and λ zap head cDNA library (provided by Y. Jan, USCF). We also screened 1×10^4 colonies from a *Drosophila* cosmid library (provided by J. W. Tamkun, NIH), and 2×10^5 recombinant phages from a Charon4 *Drosophila* genomic library (Maniatis et al., 1978). All the procedures used for the isolation of positive clones followed conventional protocols (Sambrook et al., 1989).

Sequencing of the cDNA and genomic DNA. The cDNA fragments from the nine positive cDNA clones were subcloned into pBluescript (pSK+) vectors. Unidirectional deletions were introduced into the cloned fragments using Exonuclease III and nuclease SI. Single-stranded templates were prepared by standard methods, and sequenced by the method of Sanger et al. (1977). The *rbp9* gene was localized within a 12 kilobase (kb) genomic fragment by Southern hybridizations. A contiguous set of genomic fragments encompassing the entire region homologous to the cDNAs was subcloned into pSK+ vectors and sequenced as described above for the cDNAs.

Computer programs used for sequence analysis were from the University of Wisconsin GCG package (Devereux et al., 1984). The GenBank (release 62.0) and EMBL (release 22.0) nucleotide sequence databases were scanned with the program TFASTA for the homologous proteins. The promoter *cis*-element analysis used the FPS-dependent sequence retrieval programs (Bucher and Bryan, 1984) in combination with the program FASTA to screen the Eukaryotic Promoter Database (EPD) for common promoter elements.

Primer extension and anchor PCR. Seven oligonucleotides were used to map and clone the 5' end of each transcript. The two oligonucleotides were designed from exon 1 (primer 11a: 5'-TTGGCGCTGATATTGCCTTC-3') and exon 5 (primer 55: 5'-TGAGTAATGTGGTATGTACG-3'). The 32 P-kinked primers were annealed separately with 20 μ g of polyA⁺ RNA and then extended with reverse transcriptase using the method described by Sambrook et al. (1989). The extended cDNAs were separated on the 5% acrylamide, 7 M urea gel and autoradiographed.

cDNA was also prepared from 25 μ g of polyA⁺ RNA using an oligo dT₁₄₋₁₈ primer. The 3' end of cDNA was dG tailed with dGTP and terminal transferase, and then amplified using primer 11a or 55 with an anchor PCR primer mix (9 parts oligonucleotide anchor C: 5'-GCATGCGCGCGCGCGGAGGCCCGCCCCCCCCCCCCCCC-3', 1 part oligonucleotide anchor: 5'-GCATGCGCGCGCGCGGAGGCCCG-3') by PCR for 10 cycles of 1' at 95°C, 1' at 50°C, and 1' at 74°C. The PCR products were precipitated with a half volume of 20% PEG6000, 2.5 M NaCl to remove the primers. They were then reamplified using primer 11b (5'-CCTGGGGATCTCACCCACAAAAT-3'), primer 33 (5'-CACACACTAAGGGTTTTCGAAT-3'), or primer 44 (5'-TTTGATTCTGTACGGGGAAAAT-3') with the oligonucleotide anchor for another 20 cycles under the same conditions. The amplified fragments were subcloned into pSK+ vector and sequenced.

Polymerase chain reaction for exon usage. The primer 1 (5'-ATTCTTCGAATTCATCG-3') of exon 1, the primer 2 (5'-GAAACCCCTAGTGTGT-3') of exon 3, and the primer 3 (5'-GTTGACGAACGCGTTC-3') of exon 4 are homologous to the 5' ends of the transcripts from P1, P2, and P3, respectively. The primer ex9(-) (5'-gcccttgagagacCAGTG-3') and the primer ex9(+) (5'-gcccttgagagacCCGCA-3') share the same 14 nucleotides (shown as lower case) from exon 10 at their 5' ends, but have five nucleotides (shown as upper case) at the 3' ends that are homologous to exon 8 and exon 9, respectively. One hundred nanograms of *rbp9* cDNA was used as a template for each combination of primers. PCRs were carried out for 25 cycles at 1' at 95°C, 1' at 57°C, 1' at 72°C, and the products were resolved on the 1% agarose gel electrophoresis.

Northern analysis. For developmental Northern analysis, 5 μ g of polyA⁺ RNA from each developmental stage was separated on a 1% agarose-formaldehyde gel and transferred to Nytran. The filters were hybridized with 32 P-labeled probes prepared from the type P1, type P2, and type P3 cDNAs by random hexamer labeling. Hybridizations were done using the conditions described by Sambrook et al. (1989).

Production of polyclonal antibodies to lacZ-RBP9 fusion protein. A 2.0 kb EcoRI fragment from type P1 cDNA was subcloned into the EcoRI site of plasmid pWR590-0 (Guo et al., 1984) to make pZrbp9, which contained a truncated *Escherichia coli* β -galactosidase coding region fused in frame to the carboxy-terminal 176 amino acids of *rbp9*.

One liter of dYT broth containing 100 μ g/ml carbenicillin was inoculated with JM101-pZrbp9, and the culture was induced with IPTG (isopropyl β -D-thiogalactopyranoside) (final concentration of 0.5 mM)

at the late log phase. After 2 hr at 37°C, the cells were harvested and frozen at -80°C until use. The cells were resuspended in 50 mM Tris-HCl (pH 7.5), 250 mM NaCl and lysed with lysozyme and 1% Triton X-100. After sonication, the lysate was centrifuged at $10,000 \times g$ for 30 min through a 20% sucrose cushion to pellet inclusion bodies. The pellet was resuspended in 8 M urea solubilization buffer, and the salts were removed by dialysis. After dialysis, the soluble fraction was separated on a 4% stacking, 8% polyacrylamide SDS gel. The fusion protein band was eluted from the gel, dialyzed against distilled water, and then lyophilized. About 200 μ g of fusion protein was added to one vial of MPL+TDM+CWS adjuvant (RIBI Immunochem) and injected into two New Zealand White rabbits according to the manufacturers' instructions.

Affinity purification of anti-RBP9 antibodies. The same 2.0 kb EcoRI fragment used to make pZRBp9 was subcloned into plasmid pATH3 to make pTrbp9, which has the C-terminal 176 amino acids fused in frame after an *E. coli* trpE. Five milligrams of trpE fusion protein was prepared from a liter of an M9 minimal media culture of JM101-pTRBP9 induced by indoleacetic acid using the protein purification procedure described above for the lac-RBP9 fusion protein. The purified trpE-RBP9 fusion protein was coupled to Affigel-10 (Bio-Rad) at a concentration of 2 mg/ml resin according to the manufacturer's instructions. After coupling, blocking, and washing, the resins were mixed overnight with a 5–10-fold diluted serum from immunized rabbits. The next morning, after plenty of washing, the bound antibodies were eluted with 2 or 3 ml of 0.1 M glycine, pH 2.8, and neutralized immediately with 30 ml of 1 M Tris-HCl, pH 9.0, and then stored with 0.02% Na-azide and 1% BSA.

Western blots and tissue section staining. Crude nuclear extracts from embryos, larvae, pupae, and adult flies were prepared as described in Kuroda et al. (1991). Electrophoresis of nuclear extracts, transfer of protein to nitrocellulose, and immunostaining of the blots were performed as described in Harlow and Lane (1988).

To make sections for immunostainings, γ w embryos, larvae, pupae, and adults were embedded in OCT compound and frozen with liquid CO₂. Frozen flies were cut 8 μ m thick in an SLEE cryostat, and serial sections were mounted on subbed slides. The sections were fixed in PBS containing 4% paraformaldehyde for 30 min. After blocking with 5% goat serum in 50 mM Tris-HCl (pH 8), 150 mM NaCl, 0.5% NP-40, the sections were incubated with affinity-purified anti-RBP9 antibodies at a dilution of 1:100. Fluorescein-conjugated goat anti-rabbit IgG antibody (Cappel) at a 1:100 dilution and 1 μ g/ml bis-benzimide were used, respectively, to make the RBP9 protein-antibody complex and DNA visible. Slides were mounted in 80% glycerol, 2% *n*-propyl gallate. The sections were viewed under epifluorescence optics using a Zeiss Axio-phot microscope and photographed with Kodak Ektachrome P800/P1600 film.

Results

Cloning of the *rbp9* gene. In a previous report (Kim and Baker, in press), we reported the identification of 12 new RRM sequences. Among the 12 RRM sequences, RRM9 and RRM10 showed significant similarities to the RRM of the *Drosophila* nervous system-specific gene *elav*. Since we were particularly interested in RRM9 that might be part of proteins having regulatory roles at the level of RNA processing during development, we chose RRM9 to characterize further. Later, the sequence analysis of the gene (*rbp9*) encoding RRM9 revealed that *rbp9* has three RRM9 and that both RRM9 and RRM10 were derived from different RRM9s in the same gene; RRM9 was amplified from the third RRM whereas RRM10 came from the first RRM.

In order to clone the *rbp9* gene, and determine its structure, nine independent cDNA clones containing sequences homologous to the RRM9 were isolated from four different cDNA libraries. Comparison of the restriction maps of these cDNAs suggested that there are several types of *rbp9* mRNAs (data not shown).

To determine the structures of the *rbp9* gene and its transcripts, the nucleotide sequences for each of the 9 cDNA clones

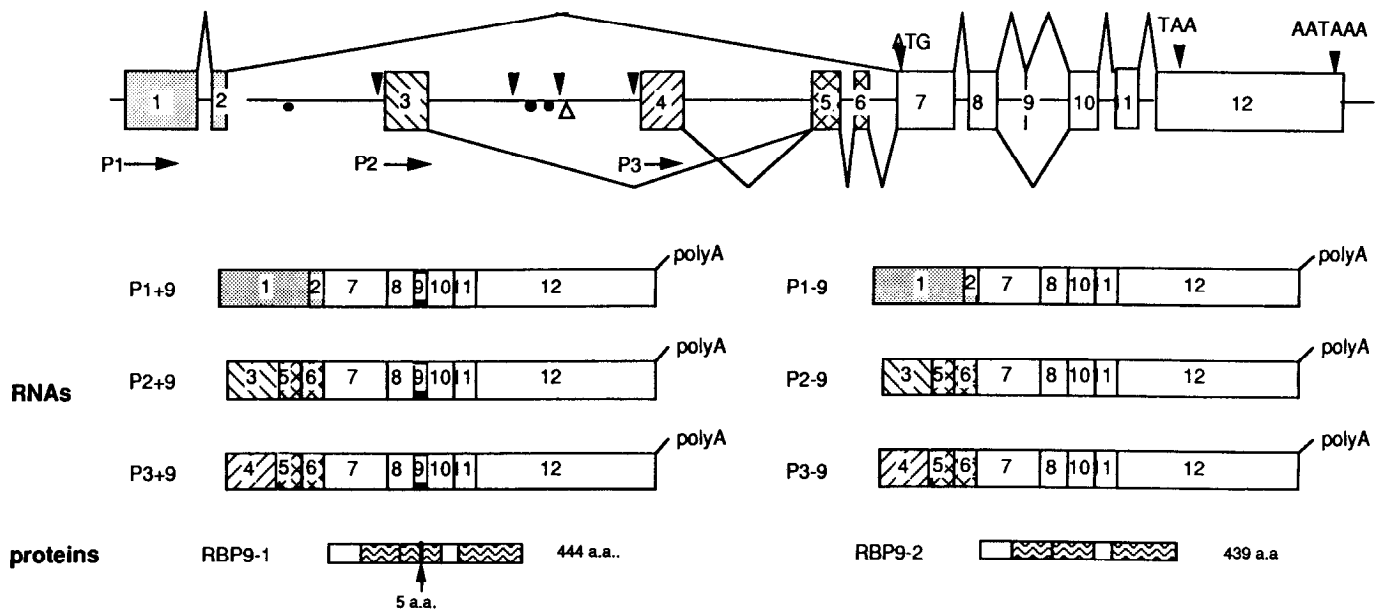


Figure 1. Alternative promoters and splicing patterns of *rbp9*. The alternatively used promoters and exons are shown together with the six types of RNAs and two types of proteins produced by *rbp9*. The arrangement of exons and splicing patterns are shown at the top. The three promoters and the translation initiation and stop codons along with the polyA signal are shown. The alternatively used exons are also marked by shading and hatching. The positions of the tentative promoter elements found in the second and third promoter region are indicated by solid circles (CT repeats), an open triangle (a putative ecdysone-responsive element), and arrowheads (sequences homologous to the ADH promoter). Two groups of RNAs with, or without, exon 9 are shown in the middle, and the exons used in each RNA are indicated with boxes. Below the RNAs, their translational products are indicated. The wavy regions within each protein indicate the RRM.

as well as 12 kb of *Drosophila* genomic DNA containing the entire *rbp9* gene and flanking sequences were determined. Comparison of the cDNA sequences to each other and to the genomic sequence revealed that all the cDNA inserts correspond to genomic sequences between positions +148 and +10241 (Figs. 1, 2). Each cDNA used one of the three different sequences at the 5' noncoding regions (denoted P1, P2, and P3), and these cDNAs with the alternatively used exon 9 (see below) are designated types P1+9, P2+9, and P3+9. In order to map the 5' ends of type P1+9, P2+9, and P3+9 mRNAs, anchor PCR (Loh et al., 1989) and primer extension assays were performed using specific primers made from the 5' end of corresponding cDNA. This revealed that type P1+9, P2+9, and P3+9 mRNAs start from genomic positions +148, +2288, and +4235, respectively. Thus, each transcript is transcribed from one of the three different promoters, P1, P2, and P3, respectively, rather than being generated by alternative processing of a precursor RNA. The structures of the *rbp9* transcripts representing each cDNA types are depicted in Figure 1.

The type P1+9 mRNA has eight exons and encodes an open reading frame (ORF) of 444 amino acids (RBP9-1) with the AUG located at +6618 and the stop codon at +8871 of the genomic sequence. This mRNA is made from P1 and uses exons 1 and 2 at the 5' noncoding region. The 3' untranslated region extends about 1.5 kb beyond the stop codon, and no ORF with the *Drosophila* codon-preference pattern is found within this region. The 3' end of the mRNA is polyadenylated at +10241 between the first AAUAAA (+10217) and T-stretch (+10252-10260). Three more AAUAAA exist within 600 base pair (bp) downstream of the first polyA signal, but there is no obvious GT-stretch after them. In addition, only the first polyadenylation site was used in the cDNAs we sequenced.

The type P2+9 mRNA is made from the P2 promoter and

has exons 3, 5, and 6 in the 5' noncoding region. This mRNA differs from the type P1+9 mRNA only in the 5' noncoding sequence. Except for this 5' end, this mRNA shares the same coding sequence, 3' noncoding sequence, and polyadenylation site with the type P1+9 mRNA.

The type P3+9 mRNA is made from the P3 promoter and has exons 4, 5, and 6 in the 5' noncoding region. Furthermore, this mRNA shares the same coding, 3' noncoding sequences, and polyadenylation site with the previous two mRNA types.

In addition to the variations at the 5' noncoding region resulting from the alternative usage of promoters, exon 9 was alternatively used in the type P2 cDNAs. The mRNA with exon 9 is designated type P2+9, whereas mRNA without exon 9 is designated type P2-9. Type P2-9 mRNA differs from type P2+9 mRNA only by the absence of exon 9. Since exon 9 is 15 nucleotides, P2-9 type mRNAs encodes a 439 amino acid ORF (RBP9-2), which is the same as the 444 amino acid ORF (RBP9-1) of type P2+9 mRNAs except for the 5 amino acid deletion. Though both RBP9-1 and RBP9-2 encoding cDNAs were identified from type P2 cDNAs, only RBP9-1-encoding cDNAs were found from the type P1 and P3 cDNAs we isolated. This observation prompted us to ask whether there were any links between the promoter used and the alternative splicing of the mini exon 9.

Alternative usage of exon 9. The pattern of alternative usage of the 15 nucleotide exon 9 was confirmed by PCR amplification. Two DNA fragments were amplified from adult fly cDNA with primers made from the neighboring exons 8 and 10 (Fig. 3A). The nucleotide sequences of the two PCR-amplified fragments showed that the small fragment was amplified from the RNA without exon 9 and the large fragment was amplified from the RNA with exon 9 (data not shown). In order to find out whether exon 9 usage was specifically coupled to the usage of

Figure 2. Complete genomic sequence of the *rbp9* gene. The sequence of *rbp9* and flanking DNA is presented. Coding sequence is translated into single-letter amino acid abbreviations below the DNA sequence. Exon sequences are underlined, and the first nucleotides underlined for each promoter are the major transcription initiation sites. The polyA signal, AAUAAA, is double underlined.

AATACCTGTGCACGTAAGTGTATGTTTTCGCAAAATCCAATTATAAATTACCTCCGCCGTGAGCATCTGAGGAATCCACCTTGCAGATATATGIGCTC 4300
 GTGCTGTAGGAAAACATATAGCCGTGTAAGTTTGATCCATTCCCATCGGGTGGCTTTCCTTCCCGAGCGGGCGAAGAGCGCTCGACAATGACGAATTC 4400
 TGTTTCTGGCCAAGAAGCGGCGCTCTCTGTAGCTCTTGGGGCTCTGCTCTCCCGATTCTTGGACATGAGGCAGCAGTACGATTGCCCTCGTCCGTGGG 4500
 CATAAAATAATTTCCACGCCGCTCGGCAAAATCAGTTTATTTCTGTTCTGTTGAGTTGCAGGAACGCGTCTTTTCTACTTTCCGTTCCGTTTATTTTC 4600
 TTATTATTATTGTTATTACGAATTAGTTTCAGCTTTCGGTTTGGCTTGGTTTTCATTTTTCAACCGGCACACACCATCAACACAATTTGATCAATAATA 4700
 TTCCAGATATTATTGCTTCGGTTATTTCTATTGTAATGAGTGAACCTCTTCGGTTTGTCTCGGCTTATGTTTGTGGAGCGCTTTCTCGTTCGGTTATT 4800
 ATTAAAAATATAAAACCAACATCGCCAGAAAATTCGTTTCTGTAATCGGTAGGAATCTACTGATTAGCAACCAGTTTCATTGATTTCCTCGTACAGAA 4900
 ATCAAGGTAAATGTCGGAATCTCTTCTCACTTTATGTAATCCGAACATATTTCCAATATTTTTCGATTTCCTGAGTTCGTTGTAACAAGCTACACTT 5000
 AACCTCAACACATAACGTCAGCATCTCTCGCATCAGCTGCGATCAGATCGGCCAGATCGTTGTAGCCTCAGAAATATTTTCGAAAATATAATCGCCGCG 5100
 GCTACGCAGCGCGCAATTTAGCCTCTCGCTCTCTCACACCGTCCGCCATCCAACCGATTGTTATTTTAAATATAACTTTTTCGCACGATTG 5200
 CGGCGTTGAATGCAACAACTGCTCAGACATGTTGTGAAAACCTCAATTGAAGTGGAACATTCTGTACCTGATTTCATAGCCGATGAGAAAGAATAATAC 5300
 TTGTGGAATAAGAAGCATAAATGGCTTGATCAATAAACGTAATCTTTTGAACTTAACCTTTTAGGTTTTCACCGTATTTTATATTATTTCATTCAATG 5400
 CTCGTCATATGCATATTTTATTTTAACTAAGAACTCTCCAGTGTATTCAATCGTAAAGTCTTACTAATTGCGTGCCACAATTATAAAGGCCAATT 5500
 TGAATGTGGGCTAAAAACTTGTCTGCTTTTGCCAAATTTATGTTTGGTTTAAAGAAGCAGATACGCTAAAAGAAGCGGCCAAACACACACATTA 5600
 AGGAAAGAAACCATACGTCAGTTGGTTATTTAAGCTATAAACAAATACACCTTTGCGAGTATGGGTTTACTTGTCTATTAGGTTTTCACAATCCCC 5700
 CAATTGAGTATGCTTTCTTCCATTAAAGTGAAGTGAAGGCCAATCAAGAATAGTTATCCGTAGGCCTTCAATTCGCTTAAGGGGGCAAAATTAAT 5800
 TGAATCGATTTTCCAGTCAATGCTGCTCGTGTATTTTGGCCCTATTGCAATTACAGCCAGTCAGTCAGCCGCTGCTGTAGCCAAGTCAGCCAGCAAT 5900
 TTGCCAACCGCAGAGGCAAGAGCCCTCGACATTTGTATGGGTTTGTGCAAAATGTTTACTCTTGTATTTTCCCAATTTCCAAGAAGTGCAGGCCCGC 6000
 ATTTCAATTAAATTAATTAACCAATTGATTTTATTTGTAACCTTTGATTTTCGATTCTCTTAATTTCTCTTTCAGTAATCTCTCTGACATACCACATTA 6100
 CTCATTCTGAGTGGATATCGTTGGTGTGGCGCTTTTAACTAGGCTATAGTAGGAGGAGCCAAAGACGGGCAGCGGACTATCTTTAATCCACTGCCGTTGC 6200
 TAGGAGTGTGCATCTGAAAGGGGCAAGTAGGGGTTGAATTTACGGGCTGAAGCTTCAACTAATCATTACTTTTATTGAGACTTTCACAGCAGTATTATA 6300
 TCTACAGCCACTTATTGATTTAGCGCCATCTCTATTCTGTCTATTGAACAAGTTTTCGAAGAAGAAAGCTGTAAGTTGATCACCTTCCAATTAACA 6400
 ATTGCGGTTAAATTAATAGCAGCAATCTCCATTAGTTAATATCTCAAGCACTCAAGCTCTTGTTCGTAGAATTTGTGGTATTAGTTTCTCCGC 6500
 ATCGATAGCATTATATCAAAATTTAAGTCTGCAAGTGATTCAATCGCTAGATCTAAGTCTGTAGAAACAGCTGATTATTAACCAGCAATCTTTCA 6600
 GCAGCAGCGTTTCGAAGAGTTCGAGGCTCAGACAGCGCTTCCAGCAGCAGCAGCAACCATCCGCGCGGGCGGTCGAGCGCGGTGGGAAGCACCAC 6700
 M V E G Q T A V Q Q Q Q Q P S G A G G A S G V G S T T
 CGGATCCGCTGGAGGACCGCTACGGCAACCAATGTGACCAACAGTCAGGCCCAACCAATGGGGGCACCAGTCCGACCAGGACTGCCGACGGGGTGGC 6800
 G S A G G P A T A N N V T N S Q A Q T N G G T T A T T T A A A G A
 GGATCCAGGACAAATCGGCGCTCGGCCAGGCCACAGCAAAATGCTGCGAGCAACAAACAAATTAATAATAATACCAACAACAATAACAATA 6900
 G S T T N A A V G Q A T A N N A A S N N N N N N N N T N N N N N N N
 ACGCCACTGCAAAACAACAACAATAATGAGCCAGACCCCAAGACCAATTTGATTGTGAACTACTTGCCACAGACGATGTCGAGGAGCAGATCCGTTTC 7000
 A T A N N N N N N N E P D P K T N L I V N Y L P Q T M S Q D E I R S
 GTTGTTCGTCAGTTTTCGGGAGGTGGAGAGCTGCAAGTTGATACGCGACAAGTGACAGTAGGATAGCTGCAAAATGATTATAAATGATAGTTGCATAT 7100
 L F V S F G E V E S C K L I R D K V T G
 ATTAATCCGATCCAATCGTTACAGGACAAAGTCTGGGCTACGGATTTCGTGAACCTACGTGAAGCAGGAGGATGCGGAAAGGCAATCAACGCATTGAACG 7200
 Q S L G Y G F V N Y V K Q E D A E K A I N A L N G
 CCCTCAGGCTGCAGAACAAACCATCAAGGCTCCATTGCACGCCCCAGCTCCGAGAGCATCAAGGGTGCCCAATCTCTATGTATCGGCTTCCAAAAA 7300
 L R L Q N K T I K V S I A R P S S E S I K G A N L Y V S G L P K N
 TATGACACAGTCCGATTTCGAATCATTTGTTCACTCCATACCGCAAGATCATAACCTCCGCTATTTCTTTGTGATAATATCACTGGTGGTGTGATCCTAA 7400
 M T Q S D L E S L F S P Y G K I I T S R I L C D N I T E
 GAGGCAAGAAATATATGGTAGAATACGATGGCTAAATACTATTTGAAGGGGCTCTCGGTTATCGGGTGTAGTCTCTTTTGAATGCATGTCCAATCTA 7500
 AAATCTATTAAAAATGTTTGTCTAATTAATTTAGTCAATGGGTTTAGCACACTTTTGTTCATATTTGCTGCGCTACACAATATATTCTCCATGA 7600
 ATTATAATTAATGCAAAATTTATTTACACTTTTACTTTTACATGTATTACCAAAATAGATGAACATGCTGCGGATATGTTTATTTTATTTTCTCTATG 7700
 E H A A G
 TTTCACTATTTTATTTGCTGTTTCATCGATCATGAAGTAACACTTATTATGTTATACTTATTTAGACCAGACGTTTATTAGAACGCTCTGGGTAATAATT 7800
 AAAATAGCATCGCCGACATTAACACACTCATCAAAAGTTAGTAACnCATGGCTTAAAAATATGTACATATCTGCGCATAACTAAGAGTTATTTTCA 7900
 CTCGCTGCACATATGGCTGCATAAAATTTACGAACCTTATTGTCATATGATGACATGCGACTTGATTTACAGTCTCTCCAAGGGCGTTGGCTTTATACG 8000
 L S K G V G F I R
 CTTTCGATCAGCGATTTCGAGCCGACCGGCCATTAAAGAGCTGAACGGCAGCACTCCGAAGAATTCACCGAABCCCATACCGTCAAGTTTCGCCAACAAT 8100
 F D Q R F E A D R A I K E L N G T T P K N S T E P I T V K F A N N

CCTAGCAGCAATAAGAACAGCATGCAGCCTTTGGCAGCCTATATAGCCCCCAAAATACCGTGGCGGACGAGCATTTCGGGCCAACGCTGCTGCCGAG 8200
 P S S N K N S M Q P L A A Y I A P Q N T R G G R A F P A N A A A G A
 CGGCAGCCGCCCGCTGCAGCTGCCATCCACCCCAATGCGGGCCGATAGAGTCAGGAATAGATTATATGTACACTAAATGAATCGAAAGATATTGCT 8300
 A A A A A A A A I H P N A G R Y S
 GTAAGACATTGTAATCGAACTAACTTCAATTTCGCATCACAGCTCTGTAATCTCGCGATATTCACCGTTGACCACTGACCTGATCACTAATGGCATGATCC 8400
 S V I S R Y S P L T S D L I T N G M I Q
 AAGGCAACACCATTTGCCAGCTCCGGCTGGTGCATTTTCGTCTACAACTGGCGCCCGATACCGAGGAGAACGTGCTGTGGCACTGTTTGGGCCCTTCGG 8500
 G N T I A S S G W C I F V Y N L A P D T E E N V L W Q L F G P F G
 AGCAGTGCBAATCTGTTAAGGTGAGTACCAATTCAATTGGAAGTTCGTGCTTTCCTAGGTCGAATTACCAGTAAATTTGAAAATTACTGAATCTTATGA 8600
 A V Q S V K
 TCTTTGACCAGGTGATTCTGTGATCTGCAGAGCAACAAGTGCAAGGGCTTTGGCTTCGTCACTATGACGAATTACGAAGAGGGTGTCTTGGCCATACAGTC 8700
 V I R D L Q S N K C K G F G F V T M T N Y E E A V L A I Q S
 GCTGAACCGCTACACACTTTGGCAACCGAGTGCCTTCAGGTCAGCTTTAAGACCAACAAGAACAAGCAACGTAATTTATTAACCAAGTTGGCTGCTGTTGTC 8800
 L N G Y T L G N R V L Q V S F K T N K N K Q T
 AATGCCAATGGCGCGCGCAGCTCTGGCGGCCAATGGTCTAGTCTCCATAATCATCACCACCAAGCAACATCGGTAGCATCAACAATAATAGTAACA 8900
 ACAGCAACCGCTGCAGTGCAGCAGCCTTCCACTTGGCAATACGTTAACCACAGCAGCAACAACAGGCAGCACAACATCAACAATAGCACTAGCAACAG 9000
 GAGCAACAGAGGCCAGCCTCCACGAGCAACAAGCAACAATCAATGGCGCCACCTAAGAGATTAGCGAGCAGCACTGCTGCAACAACGGGGAGCTTGGCC 9100
 GCAACTGCTGCCACCAACAGCAGTAGCAGCAGTCACTCGGCTCCAGCGGAGATGCTGGAGCCACCGGCACAGCCGCATCTCTCTGCTCATCGTCTGCT 9200
 CGACAAACAACAACATCTCTGAAGACATCCACGAAATTTATTTACAAACAGCCCGAGAACCAATACGAGTTGTTGCAGCAGCAGCTGCAGCTGCAGCAGGA 9300
 GTTCGCCCAGTTTGTGACCAATGTGGCCAAACAGCGCCGCACTCCATCCAGCAACGCTCGGCCAACTCGACAAACAGCGTAATGGAGGAGGAACCATC 9400
 GGCATGGGTCTGGGAATGTGGGCCAACTCAGCGCGCGGAGGTTCCGGCAACAAAACAGTAACGGCAACCAATGGCTATATGCCAACCTGGTGAATTTGGT 9500
 TCTTTTAAAGTTCACTTTTGGTTTGTGTCATGCTCCATTTTACGGATCTTTTGAATTTTAGGCAACGATTACTCAGTTTGTGTTTATGATATAAGTTTG 9600
 AATGCGCGAATAACAGAATTGTTTACGTTTGTAGTTATGAATTTGTTTTCAGTGAAATAATGTAACCATAAATTTCTTACGATGACTAAGTTCTGAA 9700
 ATGTTTTTTTTTTTTTGTCTATGCTCCCAATTCTTTTAAAGTTACATTCAAATTTAATGTAGAAGTTTACATCTTTCGTTAATGCACCTTTTAAACTGT 9800
 CTAGTCGACTCTCCCGAAGAAAAAGCTTCCATGAGGAAAAGGGAATACAGAAAAGATAATGCTATAGTTAAGTTTGTGTTGAGGTGAGGAAACGCAT 9900
 ACTCTACGGATATTTTAAACAAAGAGTAAABATACAAATTTGCAGGCATAGAAGCCCCAAACCACCAAGAAATGAAACAAACTTGACAAATATTAAAGC 10000
 TATAAATAAATGTATAAAAGCGATACGAAAACAGAACTGAATGAGTAAACAGATAGAGGCGTTTGGCTTATAGTCAAACCACTTGCAAAATAGACAATGCA 10100
 GAATGAGAAAGCAGCTAGATGCGAAAGAAAACCAACAAAGTAAACAGAAAAGCAAAATAAATAAGACATACCGCTGTTTATGTTTCATAAATCTAATG 10200
 TAAAACTTTTGTCTAAATAAATGGTGTATGCAATAATACAAAGTAAACCTTTTATTTCCATTACCATTTGTTTCGATTGCCCGAGTTAAACTTCGCCTA 10300
 GCCACTTTCATACATATGTTAACCATTACCTATCCTTCGGCACTACACTTAAATATAGACTCTCTAAGATCTTAACAGGGACATTAGCATGAGGGACAGA 10400
 AAGCCGAGAGGAAGGGCCACGAGAATTCTTCAATACAAACGAATAATTTTAAAGAAATTTGAAACAGCAAGTGGAATGCAACACCTGTCCGAGTC 10500
 AGAGAGAGAGAGAAATCAAAATACAAAGTAACAGCGATTGTAAGCTTAAACAAAAGCCAGCTAAGTACGCTAAGCAGCCCACTCCAATCGAAGTTAG 10600
 TTAATATAACTAAAGACAAAGATCAGAAACAAAGTCCCTCCCTAGACAAAGTGAAGAAATGAAATGAAACTTACATACATAGATGACAAAACGCGTT 10700
 TCAGCCTAACGAATTTGTTTAAAGCCCACTCAACAGCAATAAATAAACAACGTAACGAACAAATTAACAAATTTGCCAAACTAAATATAGCAAATTTT 10800
 AATAAACAGTAATGGACATGGAATGAGAAATGGTTGATAATTGAAGTTTTCAGAGCCTGGCAATGACAAGTTATGACAAGACAAGTTATACATATGATG 10900
 GTGCTTCACAGAATTCCTTGTAGAAGAAGTTAAATGCAAAAAATAAATAGTATTGTAAATGACAGTTTATGTTCCCTGTTTCAAGTTGATCCTTTCC 11000
 TTTAAAGTGCCGTGAACCCATTATGGAATAATCGAGATGTTACTCAACATTTACTATGCTATATCTCTCTATAAAATCTATAACTATTATGAAAATAT 11100
 TACATATGATTAAACAATATATTGATGTTGCTATTGCAAAATTTAAATGTTTATACATACATACATACCGGAAAAATTTATGAAATACAAATTCATGGC 11200
 ATGCGAATAACGGAAATCCATTTTCATGTATACGTCAAGTAAATTTATGTCATACTACTTAAACTAATTACCATGAACAAATTTATATATTCAGTATAT 11300
 ATATTGCGATTATTAGGTATGTATGTGTAGTGCATAGAGAAACCG 11347

particular promoters, we examined the presence of exon 9 in each type of transcript (P1, P2, and P3) using 5' primers specific to each transcript type. The 5' primer for each of the P1, P2, and P3 mRNAs was made from sequences within exon 1, exon 3, and exon 4, respectively, whereas the two 3' primers were made to distinguish the transcript with exon 9 from the one without the exon by making each of the 3' primers complementary to the splice junction sequences between exon 8 and 10, and between exon 9 and 10, respectively [primer ex9(−)

and ex9(+) share the same 14 nucleotides from exon 10 at their 5' ends, but have 5 nucleotides at their 3' ends, which are homologous to exon 8 or exon 9, respectively]. Since the splice junction lies very close to the 3' end of the primers, no PCR amplification was observed when P2+9 cDNA was used with ex9(−) primer, nor when P2-9 cDNA was used with primer ex9(+). Since exon 9 is only 15 bp, the PCR-amplified products from the same promoter-type mRNAs were not resolved during agarose gel electrophoresis. However, the 5' primers were made

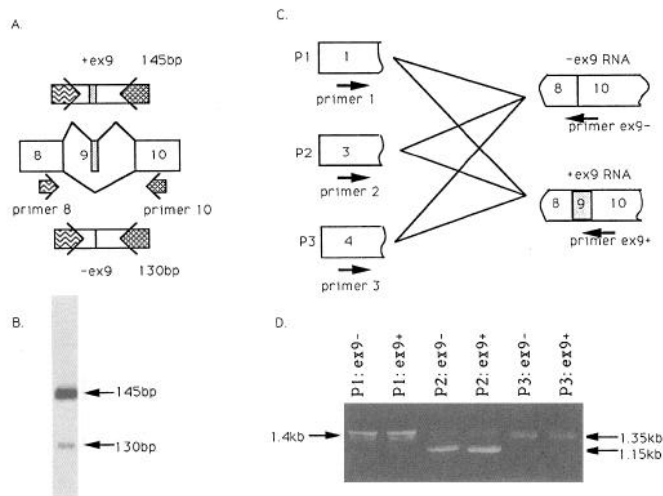


Figure 3. Alternative splicing of microexon 9. *A* and *B*, Alternative splicing of exon 9. The neighboring exons and alternative splicing pattern of exon 9 are shown. The positions of the primers used for the PCR are shown as arrows, and the structures of the PCR products are drawn. An autoradiogram of the PCR products separated by PAGE is shown. *C* and *D*, The positions of the primers used to distinguish each type of the transcript and the combination of the primers used in the PCR. The result of the PCR products separated on 0.7% agarose gel is shown. Since the expected size of the fragment amplified using each of the 5' primers and ex9(-) primer is different only 15 nucleotides from that amplified using the same 5' primer and ex9(+) primer, the two PCR products from the same type RNA appear to have the same size in the agarose gel electrophoresis; their size from each reaction is as follows: 1.40 kb from type P1 RNA, 1.15 kb from type P2 RNA, and 1.35 kb from type P3 RNA.

in such a way that PCR-amplified fragments from each of the 5' primers had a unique size: 1.40 kb from P1 RNA, 1.15 kb from P2 RNA, and 1.35 kb from P3 RNA. Fragments of the expected sizes, both with and without exon 9, were amplified from transcripts of all three promoter types (Fig. 3*B*). Thus, the alternative usage of exon 9 is independent of which promoter is used. Therefore, a total of six types of mRNAs and two types of proteins are made from the *rbp9* gene. However, the size difference between the six *rbp9* transcripts is so small, despite the utilization of three different promoters and alternative exon use, that they are not expected to be resolvable on Northern analysis. Similarly, the two types of the RBP9 proteins would have very similar size, thereby making the two proteins indistinguishable on the basis of size in Western analysis.

Promoter sequence analysis. *Drosophila* promoter sequences from -200 to +10 relative to transcriptional start sites were

retrieved from EPD (Bucher, 1988) using an FPS-dependent sequence retrieval program (Bucher and Bryan, 1984). In order to identify sequence similarities between the promoters of *rbp9* and those in the database, the 1398 nucleotides between exon 2 and 3 (Pt2), the 1878 nucleotides between exon 3 and 4 (Pt3), and the 147 nucleotides (which was the available sequence) in front of the exon 1 (Pt1) were compared with the *Drosophila* promoter sequences retrieved from the EPD. The database search revealed three sequences that were similar to sequences in the promoters of known genes; two had distinctive motifs, the functions of which have been identified, and one had a significant nucleotide sequence repeat of unknown function (Fig. 4). Their positions relative to a transcription initiation site are shown at the right side of each sequence in Figure 4, and the comparisons of these sequences to the known promoter elements are also shown.

First, a repeat of "GCCGCTG" at -83 to -55 of Pt2 is similar to the Adf-1 binding sequence. England et al. (1990) showed that "GCCGCTG" is a consensus Adf-1 binding sequence and activates transcription from the promoters of *Antp* (Schneuwly et al., 1986), *Adh* (Heberlein et al., 1985), and *Ddc* (Scholnick et al., 1986) when Adf-1 binds to them. The Adf-1 binding is sequence specific and enhances the level of transcription in conjunction with other transcription factors.

The second motif was a 35 nucleotide sequence from Pt3, which shows 72% identity to a sequence in the promoter of the ecdysone-inducible gene *Eip28/29* (Cherbas et al., 1991). In addition to the sequence similarity, these 35 nucleotides also have the sequence "TGAATC" with a partially inverted symmetry around the first "T," which has been found to be a consensus sequence of ecdysone-responsive elements of *Eip28/29* (Cherbas et al., 1991). Because the *rbp9* transcript is first detectable in the late larval period when metamorphosis is beginning under the influence of high levels of ecdysone, the presence of a potential ecdysone responsive element in the promoter of *rbp9* may indicate the involvement of ecdysone in the regulation of *rbp9* transcription.

The third motif found in the *rbp9* promoters was a long "CT" repeat that was found three times in the P2 and P3 promoters. A similar sequence was also found in the promoter region of the *as-c 14* transcription unit (Villares and Cabrera, 1987) that is involved in neurogenesis in the embryo. Even though there are no data as to the role of these repeats, the degree of conservation of the repeats may suggest that *t4* and *rbp9* have some common aspects to their regulation, perhaps due to functioning of both of these genes in the nervous system.

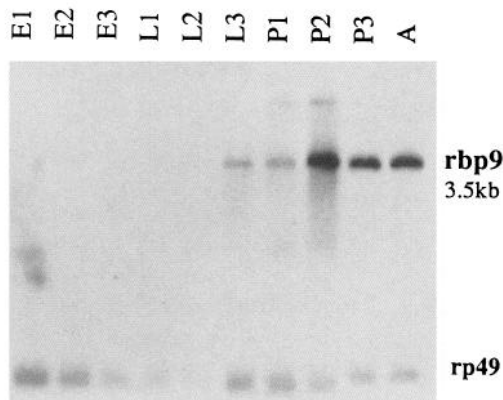
Developmental profile of the *rbp9* RNA and protein expression.

A) P2 Antp Adh-adult DDC	GgAgCTCTcGCCGCTGgGCGGcGGAAGTGC	(-83--55)
	GtGcCGCcGCCGCTGcGACTgCGCTGGA	(-97--68)
	CtGcTGCTcGCCGCTcGAGTcGACTGC	(-82--54)
	GaAgTTCaCGGCTGcgGACTgCGATTGa	(-98--69)
B) P2 P3 P3 AS-C	GCTCCTgCctTgCTCTCTCTCTCTgaTgCTCTTTctagCTtgCGCTC	(-955--906)
	cCtgCgCTCTCACTGTCTCTccCgCTCTATctctCCgcCCATA	(-1140--1098)
	ACTCCTtCgcT-CGCTCTCTCTCTgtCtCTCTTtcatCCgcTGCCCT	(-998--949)
	GCTCTTgCcgT-CACTCTCTCTCTCTctTtCTCTCCgattCTctCGCCC	(-112--63)
C) P3 Eip28/29 EcRE CONSENSUS	TTTtggggGCGT-GAATCGaaAGAGGagGAAAGAG	(-754--721)
	TTTaatcaGCGTaGAATCGgAGAGGgaGAAAGAG	(-159--126)
	Y GAACY	

Figure 4. Putative promoter elements of *rbp9*. Putative promoter elements found in *rbp9* are aligned with known promoter elements of other genes. Conserved bases are shown in *capital letters*, and identical bases are shown in *boldface*. The positions of these elements relative to their transcription initiation sites are shown to the right. *A*, Possible Adf-1 consensus binding sequence. *B*, CT repeat. *C*, Possible ecdysone responsive element. The inverted repeats are *double underlined*.

Figure 5. Developmental expression pattern of *rbp9*. *a*, Developmental Northern analysis. Two micrograms of polyA⁺ RNA from each developmental stage was analyzed with a type P1+9 cDNA probe. The blot was also hybridized with *rp49* probe to show the difference of RNA loaded in each lane. Lane 1, 0–3-hr-old embryo; lane 2, 3–12-hr-old embryo; lane 3, 12–24-hr-old embryo; lane 4, first instar larva; lane 5, second instar larva; lane 6, third instar larva; lane 7, 1-d-old pupa; lane 8, 2-d-old pupa; lane 9, 3-d-old pupa; lane 10, adult. *b*, Developmental Western analysis. Five micrograms of crude fly nuclear extracts from each developmental stage was analyzed with the anti-RBP9 antibodies. Lane 1, embryo; lane 2, larva; lane 3, 1-d-old pupa; lane 4, 3-d-old pupa; lane 5, adult.

a) Developmental Northern



b) Developmental Western

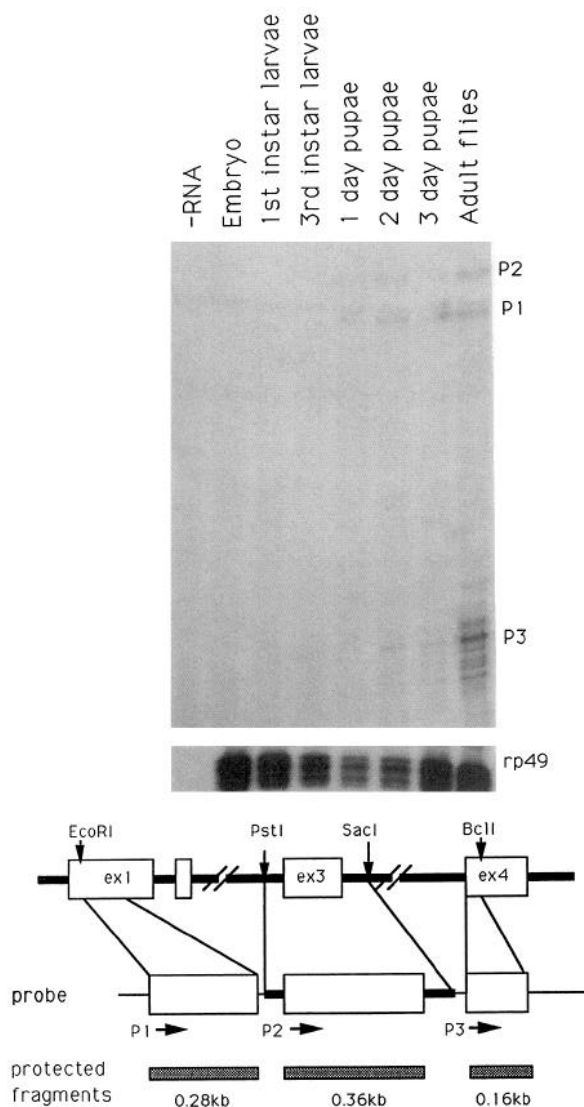
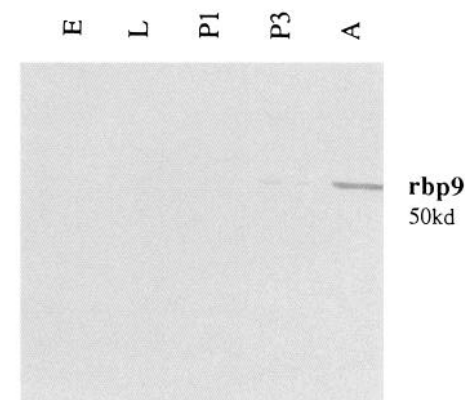


Figure 6. Developmental analysis of type P1, P2, and P3 transcripts. Ten micrograms of polyA⁺ RNAs from each developmental stage of flies was analyzed by RNase protection assay. A probe with regions homologous to each of the three promoters was made to determine directly the relative abundances of transcripts accumulating from the

The developmental Northern analysis of *rbp9* showed that a 3.5 kb RNA species was expressed from the third larval instar through the adult stages (Fig. 5*a*).

To determine the temporal profile of the expression of each promoter, we used an RNase protection assay that allows direct comparison of the amount of RNAs produced from each of the three promoters throughout development. A uniformly labeled riboprobe harboring small pieces of *rbp9* genomic DNA near the transcription initiation site of each promoter was used to detect the use of each promoter by generating RNase-protected fragments of distinct sizes (Fig. 6). The RNase protection assay showed that the type P1 and P2 transcripts are expressed from the third instar larval period through adulthood. However, the amounts of the P1 and P2 transcripts in the larval and early pupal stages were greater than in the adult. The type P3 transcript, on the other hand, was most highly expressed in the adult stage (Fig. 6). Therefore, most of the adult transcripts detected in the developmental Northern analysis are the product of the P3 promoter, while usage of the P1 and P2 promoters is mainly responsible for the transcripts of the larval and early pupal stages.

Polyclonal anti-RBP9 antibodies were raised against the carboxy-terminal 176 amino acids of the RBP9 protein using a lacZ-RBP9 fusion protein as an immunogen and affinity purified using a trpE-RBP9 fusion protein column. Western analysis of the fusion proteins using the affinity-purified anti-RBP9 antibodies showed that the anti-RBP9 antibodies reacted strongly to the trpE-RBP9 and lacZ-RBP9 fusion proteins with little or no reactivity to lacZ and trpE proteins, respectively. No specific reactivity was detected from the preimmune sera (data not shown). Nuclear extracts were prepared from each develop-

three promoters. The restriction enzyme sites used to subclone the genomic fragments are indicated, and linker sequences used to clone the genomic sequences into a vector are shown as thin lines. The exon sequences are indicated as boxes, and the intron sequences are shown as thick lines. The expected sizes of the protected fragments are shown at the bottom of the probe as shaded boxes. As a reference for the amount of RNA used in each lane, *rp49* probe was incubated together with the *rbp9* probe, and the intensity of the signal from *rp49* is also shown. Lane 1, no RNA; lane 2, embryo; lane 3, first instar larva; lane 4, third instar larva; lane 5, 1-d-old pupa; lane 6, 2-d-old pupa; lane 7, 3-d-old pupa; lane 8, adult.

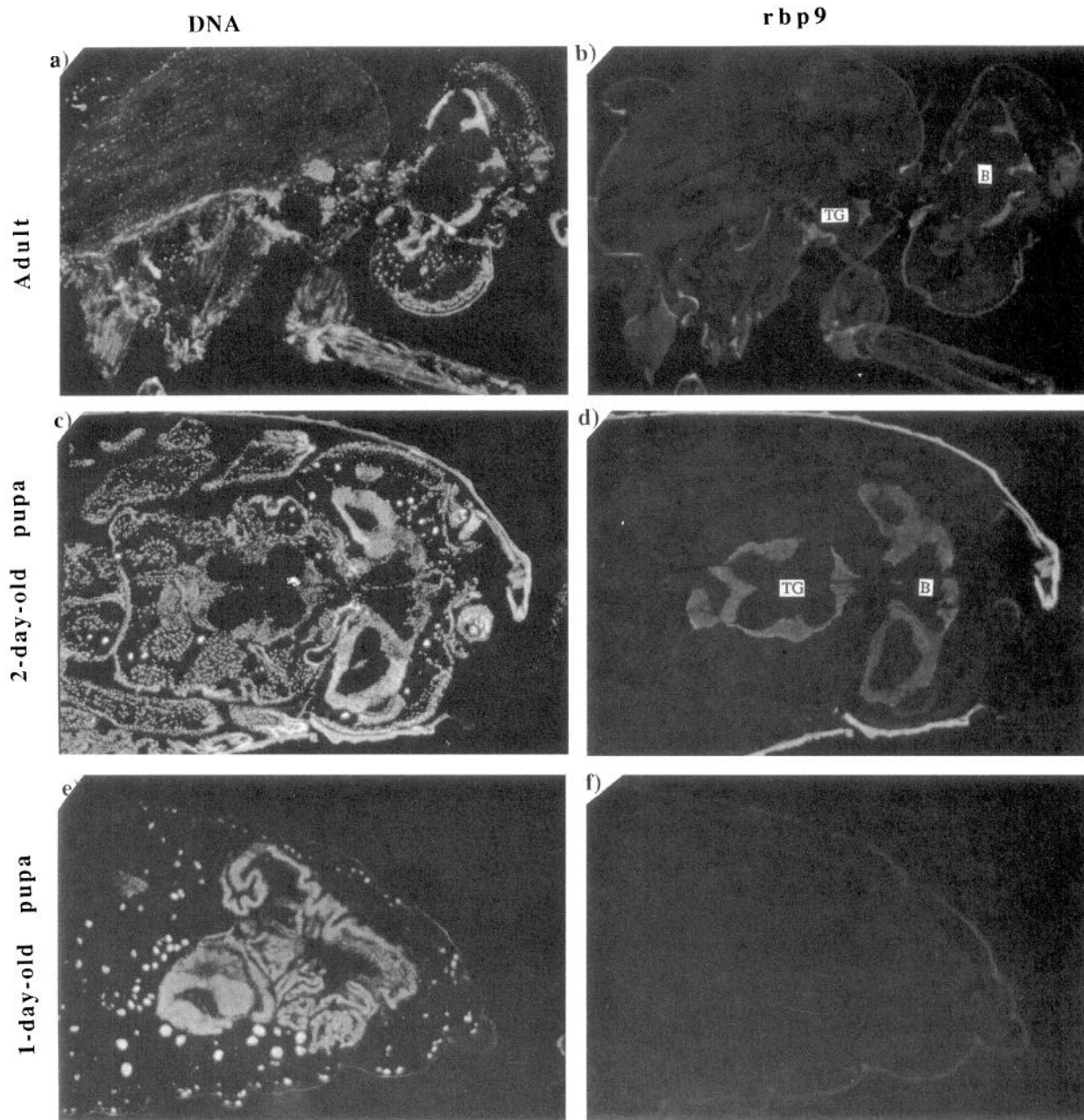


Figure 7. Tissue- and developmental-specific expression of RBP9 protein. Serial sections of flies from different developmental stages were stained with anti-RBP9 antibodies using indirect immunofluorescence (*b, d, f*) and bis-benzimide (*a, c, e*). *a* and *b*, A vertical section of an adult fly. *c* and *d*, A horizontal section of a 2-d-old pupa. *e* and *f*, A horizontal section of a 1-d-old pupa. Brain (*B*) and thoracic ganglion (*TG*) are indicated.

mental stage and identical amounts of protein were used for a developmental Western analysis. This analysis showed that, in the pupal and adult stages, the anti-RBP9 antibodies recognized only a single band of 50 kDa that was approximately the size of the RBP9 protein predicted from the cDNA sequence. However, no protein was detected by the anti-RBP9 antibodies in the nuclear extracts of embryos and larvae (Fig. 5*b*).

A comparison of the developmental Northern and Western analyses of *rbp9* shows that RBP9 protein is only detected at stages substantially later than when transcripts first appear. Several mechanisms may account for this regulation, and we are currently in the process of testing various possibilities.

Localization of RBP9 protein using indirect immunofluorescence. To examine the expression pattern of RBP9 protein, cry-

osections of *Drosophila* were reacted with anti-RBP9 antibodies and visualized by indirect immunofluorescence. As controls for the tissue staining, preimmune sera with or without a secondary antibody, or a secondary antibody alone was used to stain tissue sections. No positive staining was observed in these controls (data not shown). When sections of adult flies were stained with the anti-RBP9 antibodies, the RBP9 protein is detected only in the nerve cell bodies of the CNS (Fig. 7*b, d*). The RBP9 protein is specifically localized on the nuclei of the nerve cells. In particular, the nuclei of the brain and thoracic ganglion are the most heavily labeled, whereas low levels of the RBP9 protein are detected in the nuclei of the lamina cell body layer and antennae.

In order to examine the pattern of RBP9 protein expression

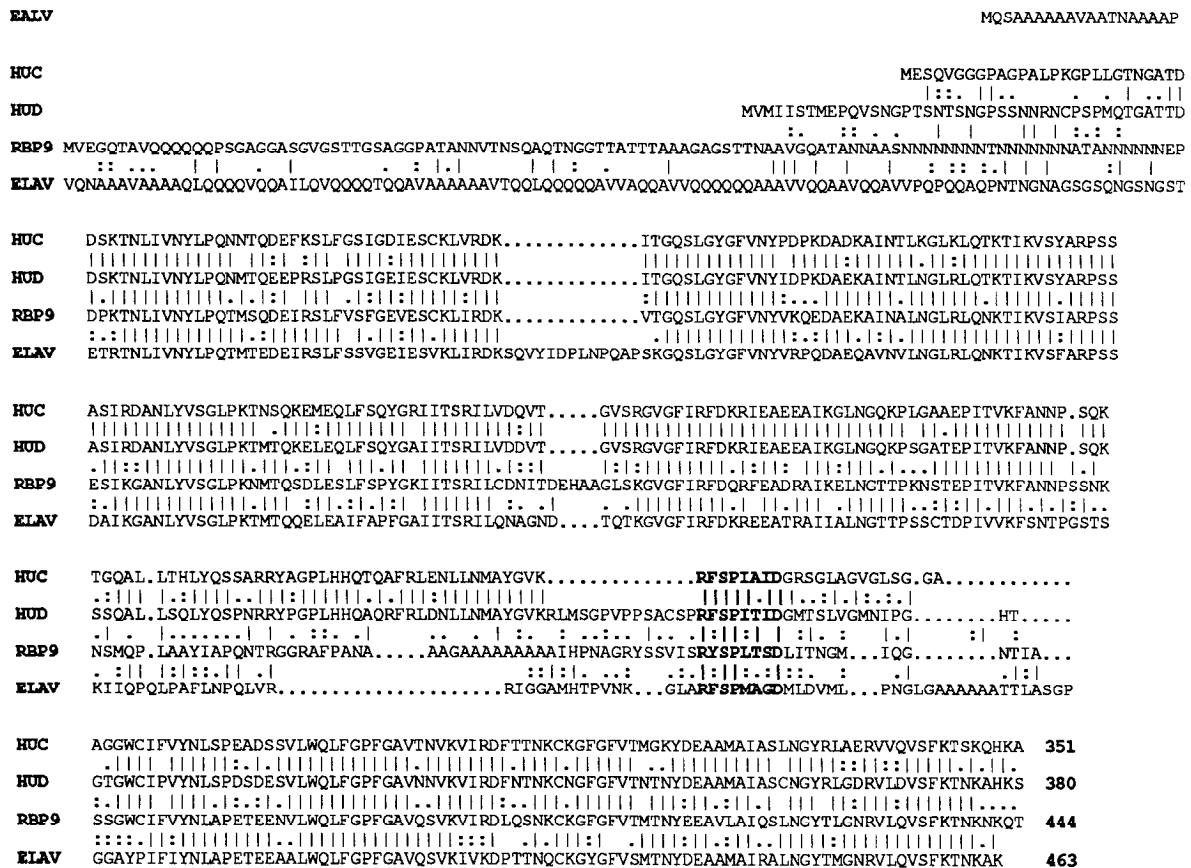


Figure 8. Sequence homology of RBP9 to HuC, HuD, and ELAV proteins. The amino acid sequences of HuC, HuD, RBP9, and ELAV proteins are aligned together using the **BESTFIT** program. Identical matches are shown with *vertical bars* and functionally related matches are marked with *single* or *double dots*. Gaps in the alignment are shown by *periods*. The first two rows of the alignment are the N-terminal sequences; the third, fourth, and sixth rows of the alignment are the three RRM_s, respectively; and the fifth row is the second and third RRM-connecting region. The conserved octapeptide of the RRM connecting region is *boldface*.

throughout development, flies from larval, pupal, and adult stages were serially sectioned (8 μ m) and the distribution of RBP9 protein was visualized by indirect immunofluorescence. These sections were also stained with bis-benzimide to visualize nuclei. The nuclear staining patterns documented the substantial morphological change that the CNS undergoes during the mid-pupal period (Fig. 7*a,c,e*; reviewed in Kankel et al., 1980). After metamorphosis, the CNS has two distinct masses, the brain and thoracic ganglion, and RBP9 protein is detected in virtually all of nuclei of the brain and thoracic ganglion. Since we detected a low level of RBP9 proteins in early pupae by the Western analysis, we carefully examined the serial sections of earlier stages for the possibility that only a small number of cells express *rbp9*. However, RBP9 protein is not detected in CNS at early stages (embryos, larvae, and 1-d-old pupae; Fig. 7*b,d*). This suggests that neither the larval nervous system, nor the precursor cells of the adult CNS express RBP9 protein.

rbp9 has a high similarity to *HuD* and *elav*. The conceptual translation product of *rbp9* can be divided into five domains: the 106 amino acid N-terminal region, first RRM, second RRM, the 71 amino acids between the second and third RRM, and the third RRM. The amino acid sequence of *rbp9* has a striking similarity to the proteins encoded by human *HuC* (Furneaux, personal communication), *HuD* (Szabo et al., 1991), and *Drosophila elav* (Robinow et al., 1988) genes (Fig. 8). These three genes, like *rbp9*, encode proteins with three RRM's and are

specifically expressed in the nervous system. The comparison of the sequences of these nerve-specific RRM proteins shows that they have a high degree of similarity: the three RRMs of *rbp9* have 68%, 62%, and 69% identity to the corresponding RRMs of *elav*, whereas the *rbp9* RRMs have 76%, 72%, and 69% identity to the corresponding RRMs of HuD. In addition, HuC and HuD showed 88% identity. The 70–90% identity shown here is outstanding considering that less than 20% identity is typically observed from the comparison of two random RRMs.

Even though the RRM sequences are highly conserved, two places, one each in the first and second RRM, show differences. In the first RRM, *elav* has 13 extra amino acids near RNP-1, whereas RBP9-1 has 5 extra amino acids near the RNP-1 of the second RRM. Since the extra 5 amino acids in RBP9-1 originate from the alternative use of exon 9, the second RRM of RBP9-2 has even higher similarity to those of *elav*, HuC, and HuD (about 1% higher than RBP9-1). A tertiary structure analysis of an RRM showed that both of these places are within the loop-3 structure that lies next to RNP-1 consensus sequence in a β -sheet structure (Nagai et al., 1990; Hoffman et al., 1991).

Essentially all the similarity between these proteins is restricted to their three RRM; other regions of these proteins show only limited similarities. In N-terminal sequences, there is no obvious similarity between the proteins, but stretches of the Asn and Gln of *rbp9*, or the Ala and Gln of *elav* are noticeable, while HuC and HuD have a short N-terminal sequence

without such sequences. The region between the second and third RRM also shows no sequence conservation, and many gaps exist in the alignment. However, an octapeptide "R(F/Y)SP(I/L)(A/T)XD" is conserved in all proteins. Besides this, stretches of Ala are also found in this region of *elav* and *rbp9*.

In situ hybridizations to larval salivary gland polytene chromosomes with an *rbp9* cDNA probe showed that *rbp9* gene is located in the cytological interval, 23C of the second chromosome. However, no genes with obvious nervous system-specific phenotypes are currently known in this region.

Discussion

There are three features of our findings that are striking. The first is that *rbp9* encodes an RRM-type RNA binding protein. Second, *rbp9* is expressed only in the adult nervous system. Third, a group of four nervous system-specific RRM genes, including *rbp9*, show high similarities to one another.

The fact that *rbp9* has RRM is very intriguing, since studies on a number of developmental processes have shown that RNA processing is a major mechanism for regulating the expression of developmentally important genes. Although more than 100 genes have been shown to be regulated at the level of RNA processing, only a few of the factors involved in regulating RNA processing choices have been isolated. However, of the few RNA processing regulators known, several have RRM, and the presence of the RRM in a protein is very informative since many of the RRM proteins identified so far have a role in aspects of RNA metabolism. Moreover, in cases where the RNA binding activity of RRM proteins has been mapped, it has been shown that the RRM is necessary for RNA binding activity in each case, and in most, but not all, of these cases the RRM has also been shown to be sufficient for RNA binding (Sachs et al., 1987; Query et al., 1989; Scherly et al., 1989). Therefore, the discovery of RRM in *rbp9* suggests that *rbp9* may be directly involved in an RNA processing event.

Second, the tissue- and stage-specific expression pattern of *rbp9* suggests that the RBP9 protein may be a regulatory factor involved in adult neurogenesis. RBP9 protein is localized to the nuclei of the CNS only after metamorphosis during the mid-pupal stage. During this metamorphosis, the neurons in the CNS change their physiology as well as their morphology drastically (Truman and Bate, 1988). Therefore, the appearance of the RBP9 protein just after this morphological change indicates that the expression of *rbp9* may be turned on in conjunction with differentiation of the adult nervous system and that *rbp9* may be involved in an adult nervous system-specific function or in the maintenance of the differentiated nervous system.

The adult- and nervous system-specific expression of *rbp9* indicates that the expression of *rbp9* is highly regulated. The sequence analysis of the *rbp9* cDNA and genomic DNA with the view to find putative regulation mechanisms revealed several putative promoter elements, and a complex gene structure including the three alternatively used promoters and alternative splicing of the small exon in the coding region. The alternatively used promoters do not affect the coding sequence. This suggests that the expression of *rbp9* transcripts from each promoter may be regulated differently to achieve a very fine control of *rbp9* expression. In addition, the existence of two RBP9 products resulting from the alternative splicing of exon 9 raises the possibility that the expression of the two RBP9 proteins is under the control of alternative splicing.

Lastly, the sequence of *rbp9* has similarity to a group of RRM

proteins, *Drosophila elav* and the human neural proteins HuC and HuD. These proteins also have nervous system-specific expression patterns. *Drosophila elav* is required for embryonic neurogenesis and the maintenance of the nervous system (Robinow and White, 1988). The functions of the human genes HuC and HuD are not known, but a nervous system disorder observed in patients having an autoimmune serum that recognizes these proteins suggests that they have roles in nervous system-specific RNA processes (Grans et al., 1985). Recently, P. King and J. D. Keene (personal communication) have isolated another *rbp9* homolog from humans. These findings suggest that the gene family containing these nervous system-specific RRM genes may be sizable, and that the members of this gene family may participate in an important nervous system-specific process.

In our initial characterization of RRM sequences from *Drosophila*, we showed that a numerical analysis of sequence similarities clustered some of the RRM we had cloned with the RRM of previously identified proteins (e.g., *rbp9* with *elav*) (Kim and Baker, in press). We speculated that these sequence similarities might reflect functional similarities. The findings of similar tissue-specific patterns of expression of *rbp9* and *elav* lends support to the speculation that they might encode similar functions. When taken together with similar findings of functional similarity between another of the RRM that we cloned (*rbp1*; Kim et al., in press) and the gene with which it clustered (ASF/SF2; Ge et al., 1991; Krainer et al., 1991), these results suggest that the numerical analysis we used may be a useful tool for gaining insight into the functions of members of gene families.

Although the sequence comparison of *rbp9* to *elav*, HuC, and HuD revealed more than 67% identities within the RRM, it shows no significant similarities between the sequences outside RRM of these proteins. The N-terminal sequences of these proteins have no detectable similarity to each other. While most of the region between the second and third RRM has also diverged between these proteins, the presence of a conserved octapeptide sequence within this region suggests that the role of this region is more than just connecting these two RRM. In addition, the differences within the RRM of these proteins are very interesting considering the high degree of conservation of the RRM sequences between these proteins. In particular, the fact that both differences are at the loop-3 region, which lies at the N-terminal end of RNP-1 β -sheet, is intriguing. Since the loop-3 and RNP-1 sequences of U1-A and U2-B' have been shown to determine the binding specificity of these proteins (Scherly et al., 1990; Bentley and Keene, 1991), the optional addition of 5 amino acids to the RBP9 protein in this region and the 13 extra amino acids of *elav* at the loop-3 region may confer on them different RNA binding specificities or affinities. In addition, the facts that two related fly proteins, *elav* and *rbp9*, have a different expression profile—*elav* is required for the early embryo neurogenesis whereas *rbp9* is expressed only at the late developmental stages—and that a *rbp9* cDNA under the *elav* promoter failed to rescue the embryonic lethal phenotype of *elav* (K. M. Yao and K. White, personal communication) are consistent with the idea that these two proteins are involved in different functions rather than having synonymous functions in the CNS development.

The nervous system-specific expression patterns of members of this RRM gene family suggest that they may participate in some aspect of RNA processing during the nervous system de-

velopment. However, we do not currently know for any of these four genes whether their putative RNA processing functions are at the level of alternative RNA splicing or some other aspect of RNA processing. The isolations of *rbp9* mutants as well as the target RNA(s) to which the RBP9 protein may bind are obviously the next steps to understand the function(s) of *rbp9*.

References

- Baker BS (1989) Sex in flies; the splice of life. *Nature* 340:521–524.
- Bandziulis RJ, Swanson MS, Dreyfuss G (1989) RNA-binding proteins as developmental regulators. *Genes Dev* 3:431–437.
- Bentley RC, Keene JD (1991) Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B' protein and through interactions with U2 snRNP-A' protein. *Mol Cell Biol* 11:1829–1839.
- Berk AJ, Sharp PA (1978) Structure of the adenovirus 2 early mRNAs. *Cell* 14:695–711.
- Breitbart RE, Nadal-Ginard B (1987) Developmentally induced muscle-specific trans-factors control the differential splicing of alternative and constitutive troponin T exons. *Cell* 49:793–803.
- Bucher P (1988) The eukaryotic promoter database of the Weizmann Institute of Science. EMBL nucleotide sequence data library, Heidelberg.
- Bucher P, Bryan B (1984) Signal search analysis: a new method to localize and characterize functionally important DNA sequences. *Nucleic Acids Res* 12:287–305.
- Cherbas L, Lee K, Cherbas P (1991) Identification of ecdysone response elements by analysis of the *Drosophila* Eip28/29 gene. *Genes Dev* 5:120–131.
- Devereux J, Haerberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Doc CQ, Scott MP (1988) Segmentation and homeotic gene function in the developing nervous system of *Drosophila*. *Trends Neurosci* 11:101–106.
- Dreyfuss G, Swanson MS, Pinol RS (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* 13:86–91.
- England BP, Heberlein U, Tjian R (1990) Purified *Drosophila* transcription factor, Adh distal factor-1 (Adf-1), binds to sites in several *Drosophila* promoters and activates transcription. *J Biol Chem* 265:5086–5094.
- Ge H, Zuo P, Manley JL (1991) Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* 66:373–382.
- Graus F, Cordon-Cardo C, Posner J (1985) Neuronal antinuclear antibody in sensory neuronopathy from lung cancer. *Neurology* 35:538–543.
- Guo L, Stepien PP, Tso JY, Brousseau R, Narang S, Thomas DY, Wu R (1984) Synthesis of human insulin gene. VIII. Construction of expression vectors for fused proinsulin production in *Escherichia coli*. *Gene* 29:251–254.
- Harlow E, Lane D (1988) Antibodies: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Heberlein U, England B, Tjian R (1985) Characterization of *Drosophila* transcription factors that activate the tandem promoters of the *Alcohol Dehydrogenase*. *Gene* 41:965–977.
- Hoffman DW, Query CC, Golden BL, White SW, Keene JD (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA* 88:2495–2499.
- Kankel DR, Ferrus A, Garen SH, Harte PJ, Lewis PE (1980) The structure and development of the nervous system. In: *The genetics and biology of Drosophila*, Vol 2d (Ashburner M, Wright TRF, eds), pp 295–368. London: Academic.
- Kenan DJ, Query CC, Keene JD (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci* 16:214–220.
- Kim Y-J, Baker BS (in press) Isolation of RRM-type RNA binding protein genes and the analysis of their relatedness using a numerical approach. *Mol Cell Biol*, in press.
- Kim Y-J, Zuo P, Manley JL, Baker BS (in press) A *Drosophila* RNA binding protein RBP1 is localized to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. *Genes Dev*, in press.
- Kitamura N, Tagagaki Y, Furuto S, Tanaka T, Nawa H, Nakanishi S (1983) A single gene for bovine high molecular weight kininogens. *Nature* 305:545–549.
- Krainer AR, Mayeda A, Kozak D, Binns G (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* 66:383–394.
- Kuroda MI, Kernan MJ, Kreber R, Ganetzky B, Baker BS (1991) The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell* 66:935–947.
- Loh EY, Elliott JF, Cwirla S, Lanier LL, Davis MM (1989) Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain. *Science* 243:217–220.
- Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D, Sim GK, Estratiadis A (1978) The isolation of structural genes from libraries of eukaryotic DNA. *Cell* 15:687–701.
- Mattaj JW (1989) A binding consensus: RNA-protein interactions in splicing, snRNPs, and sex. *Cell* 57:1–3.
- Nagai K, Oubridge C, Jessen TH, Li J, Evans PR (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A [see comments]. *Nature* 348:515–520.
- Query CC, Bentley RC, Keene JD (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* 57:89–101.
- Richter K, Good PJ, Dawid IB (1990) A developmentally regulated, nervous system-specific gene in *Xenopus* encodes a putative RNA-binding protein. *New Biol* 2:556–65.
- Robinow S, White K (1988) The locus *elav* of *Drosophila melanogaster* is expressed in neurons at all developmental stages. *Dev Biol* 126:294–303.
- Robinow S, Campos AR, Yao KM, White K (1988) The *elav* gene product of *Drosophila*, required in neurons, has three RNP consensus motifs. *Science* 242:1570–1572 [erratum, *Science* 243:12].
- Rosenfeld MG, Amara SG, Evans RM (1984) Alternative RNA processing; determining neuronal phenotype. *Science* 225:1315–1320.
- Sachs AB, Davis RW, Kornberg RD (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol Cell Biol* 7:3268–3276.
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.
- Scherly D, Boelens W, Van VW, Dathan NA, Hamm J, Mattaj JW (1989) Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J* 8:4163–4170.
- Scherly D, Boelens W, Dathan NA, Van VW, Mattaj JW (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B' and their cognate RNAs. *Nature* 345:502–506.
- Schneuwly S, Kuroiwa A, Baumgartner P, Gehring WJ (1986) Structural organization and sequence of the homeotic gene *Antennapedia* of *Drosophila melanogaster*. *EMBO J* 5:733–739.
- Scholnick SB, Bray SJ, Morgan BA, McCormick CA, Hirsh J (1986) CNS and hypoderm regulatory elements of the *Drosophila melanogaster* *Dopa decarboxylase* gene. *Science* 234:998–1002.
- Szabo A, Dalmau J, Manley G, Rosenfeld M, Wong E, Henson J, Posner JB, Furneaux HM (1991) HuD, a paraneoplastic encephalomyelitis antigen, contains RNA-binding domains and is homologous to Elav and Sex-lethal. *Cell* 67:325–333.
- Truman JW, Bate M (1988) Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila melanogaster*. *Dev Biol* 125:145–157.
- Villares R, Cabrera CV (1987) The *achaete-scute* gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* 50:415–424.
- Ziff EB (1982) Transcription and RNA processing by the DNA tumour viruses. *Nature* 287:491–499.