The *Drosophila Hrb87F* gene encodes a new member of the A and B hnRNP protein group

Susan R.Haynes*, Diana Johnson+, Gopa Raychaudhuri[1] and Ann L.Beyer[1]
Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 and [1]Department of Microbiology, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

Received October 11, 1990; Accepted November 29, 1990 EMBL accession no. X54803

ABSTRACT

Nascent premessenger RNA transcripts are packaged into heterogeneous nuclear ribonucleoprotein (hnRNP) complexes containing specific nuclear proteins, the hnRNP proteins. The A and B group proteins constitute a major class of small basic proteins found in mammalian hnRNP complexes. We have previously characterized the *Drosophila melanogaster Hrb98DE* gene, which is alternatively spliced to encode four protein isoforms closely related to the A and B proteins. We report here that the *Drosophila* genome contains a family of genes related to the *Hrb98DE* gene. One member of the family, *Hrb87F*, is very homologous to *Hrb98DE* in both sequence and structure. The *Hrb87F* transcripts (1.7 and 2.2 kb) utilize two alternative polyadenylation sites, are abundant in ovaries and early embryos, and are present in lesser amounts throughout development. In one wildtype strain of *Drosophila* there is a naturally-occurring polymorphism in this gene due to the insertion of a 412 transposable element in the 3′ untranslated region. The larger transcript is not produced in these flies and thus is not required for viability. Sequence identities among the *Drosophila Hrb* proteins and the vertebrate A and B hnRNP proteins suggest that these proteins may form a distinct subfamily within the larger family of related RNA binding proteins.

INTRODUCTION

In eukaryotic cells, nascent RNA transcribed by RNA polymerase II is present in the nucleus as RNA-protein complexes, termed the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes [for reviews, see (1,2)]. Immunopurification of these complexes from HeLa cells indicates that they contain many protein species, most of which have not been extensively studied (3). However, six of these proteins, termed core hnRNP proteins (4), have been well characterized. These six species are the most abundant proteins in the hnRNP complex and are present in stoichiometric amounts (5). As single-stranded nucleic acid binding proteins (3,6), they may function to keep pre-mRNA from forming intramolecular hybrids so that sequences involved in specific processing steps can be recognized. Four of the core proteins (A1, A2, B1, and B2) are basic polypeptides of 30–40 kDa molecular mass and have been shown comprise a group of antigenically related proteins (7). [The other core proteins, C1 and C2, are distinct from the A and B proteins, and are probably produced from a single gene (8,9).] Although only four A and B group proteins can be distinguished on one-dimensional protein gels (on which they were originally defined), analysis of hnRNP preparations on two-dimensional immunoblots demonstrates that there are additional related A and B type hnRNP polypeptides (3,10,11). Recent experiments have begun to characterize the relationships among some of these proteins at the molecular level, and to account for the large number of related proteins. In humans, there is evidence for two transcriptionally active A1 genes, which produce variant proteins differing at two amino acids (12). The sequences of the A1 and B1 cDNAs show a high degree of homology, indicating that these proteins are encoded by related genes (9,12). Alternative splicing appears to play a major role in the production of the A and B proteins. The $A1^B$ protein is generated by an alternative splice which incorporates an optional exon in the A1 transcript (13), and a similar mechanism has been suggested for the production of two *Xenopus* A1 isoforms, A1a and A1b (14). Finally, analysis of cDNA clones for the human A2 and B1 proteins indicates that these proteins (and possibly also B2) may be produced by alternative splicing of a single primary transcript (9). Thus, transcription of closely related genes and alternative splicing of those transcripts generate several members of the A and B group of hnRNP proteins.

We have been characterizing a *Drosophila* gene, *Hrb98DE*, that encodes putative *hn*RNA *b*inding proteins that are closely related to the A and B hnRNP proteins (15,16). As in the case of the human and *Xenopus* genes, multiple proteins are generated from a single gene by the use of alternative exons. Transcripts encoding four protein isoforms are produced by use of alternative promoters and splice sites. The isoforms differ only at their N-termini, and show significant sequence and structural homology

* To whom correspondence should be addressed

+ Permanent address: Department of Biological Sciences, George Washington University, Washington, DC 20052, USA

to previously characterized A and B proteins. The N-terminal halves of these proteins consist of two copies of an ~80 amino acid sequence, variously termed the RNP consensus domain (17), RNA recognition motif (18), or RNP motif (19,20). This region is considered to be an RNA binding domain, for it has been implicated in the binding of the A1 protein to nucleic acid (21,22), and a similar domain is required for the specific association of the 70K (18) and A (19) proteins of U1 snRNPs with U1 RNA. The C-terminal halves of the proteins are glycine rich (38–44% glycine) with interspersed aromatic amino acids; while the exact sequences are poorly conserved between the different proteins, the compositions are very similar.

Given the similarity of the *Hrb98DE* proteins to the human A and B hnRNP proteins, we were interested in determining whether there is a similar genomic organization as well, i.e. whether the *Drosophila* genome contains multiple loci encoding proteins related to the *Hrb98DE* proteins. We report here the results of genomic Southern blot experiments demonstrating that the *Drosophila* genome contains one locus that is closely related to the *Hrb98DE* gene, and several others that are more distantly related. We have isolated genomic and cDNA clones for the closely related locus, which we have named *Hrb87F*. Sequence analysis confirms the close homology to the *Hrb98DE* gene at both the nucleotide and amino acid level. In addition, there are significant similarities in transcriptional regulation between the two genes. Our results indicate that the *Drosophila* genome, like the human genome, contains multiple genes for this family of hnRNA binding proteins.

MATERIALS AND METHODS

The genomic clone was isolated from a Canton S genomic library (23), the R31 and R2-1 cDNA clones from an Oregon R 0–3 hr embryonic cDNA library (24), and the ov20 cDNA clone from a Canton S ovarian cDNA library (25). Library screening and Southern hybridizations were done as previously described (26), except that for low stringency washes of filter hybridizations, 1× SET was used instead of 0.2× SET. Dideoxy sequencing, reverse transcription, S1 protection, PCR analyses and in situ hybridizations to polytene chromosomes were done as described (16), except that the probe was labeled with biotinylated UTP and detected by deposition of a colored alkaline phosphatase reaction product. Preparation of RNA, methylmercuric hydroxide gels and hybridizations followed protocols described in (26).

RESULTS

A family of genes related to Hrb98DE

To identify genes related to *Hrb98DE*, Southern blots of *Drosophila* genomic DNA were probed at reduced stringency with a coding fragment derived from a *Hrb98DE* cDNA clone. The probe contained sequences corresponding to the N-terminal half of the protein, but lacked sequences corresponding to the glycine-rich C-terminal domain [*pen* repeat sequences (15)], which hybridize to many unrelated loci in *Drosophila*. Figure 1A shows that the *Hrb98DE* probe hybridizes to a number of bands in addition to those specific for that locus (arrowheads). Because there appear to be few or no pseudogenes in the *Drosophila* genome, it is likely that many of these bands represent active genes. Note that one band in each lane hybridizes noticeably more strongly than do the other related bands, and presumably is very similar to the *Hrb98DE* coding sequences.

To isolate cDNA clones corresponding to the putative related gene, the *Hrb98DE* probe was then used in a reduced stringency screen of an early embryonic cDNA library (on the assumption that the related gene might be abundant at early stages of development, as is *Hrb98DE*). A clone was isolated and shown to be derived from the putative related gene by the fact that it hybridized to genomic DNA fragments of the appropriate sizes (Figure 1B); this new clone was then used to obtain additional clones from cDNA and genomic libraries. (See Materials and

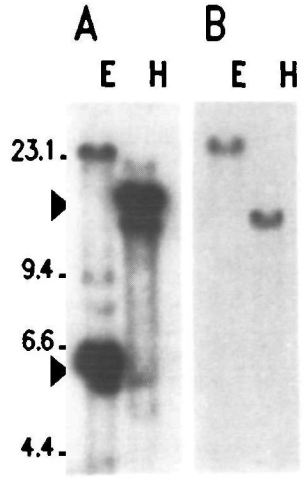

Figure 1. Identification of fragments related to *Hrb98DE*. Oregon R genomic DNA was digested with *Eco*RI (E) or *Hin*dIII (H) and hybridized (A) at low stringency with a fragment from the p9 cDNA clone encoding the two *Hrb98DE* RNP motifs or (B) at high stringency with a partial cDNA clone from the *Hrb87F* gene. The arrowheads mark the genomic fragments corresponding to the *Hrb98DE* locus; size markers are in kb.

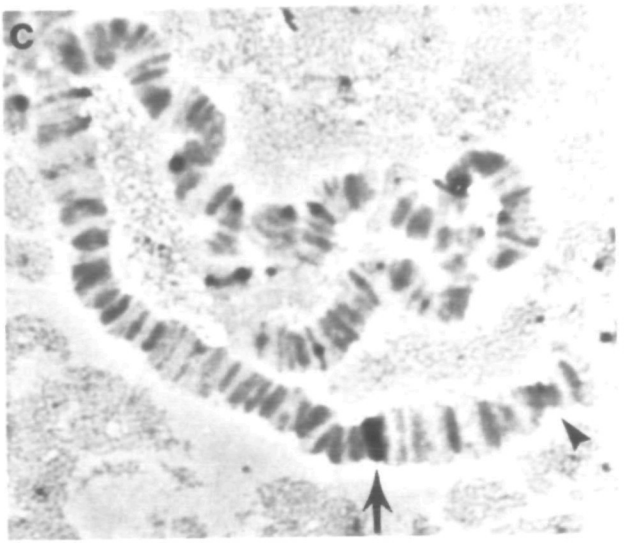

Figure 2. *In situ* hybridization of polytene chromosomes. Fragments from the genomic clone corresponding to part of the *Hrb87F* gene and several kb 5′ of it were labeled and hybridized to squashes of salivary gland chromosomes. The probe lacked *pen* (GGN) repeat and 412 element sequences. A single site of hybridization is seen on the third chromosome at 87F (arrow). The figure shows the right arm of the third chromosome from the chromocenter (C) to just beyond a constriction at 89E (arrowhead). [Note that this gene had previously been incorrectly localized to 32AB and named *Hrb32AB* (51).]

Methods for details of the libraries screened.) To determine the cytological location of the newly isolated gene and confirm that these clones represented a gene distinct from *Hrb98DE*, a probe derived from the genomic clone was hybridized to polytene chromosomes. A single band was seen on the right arm of the third chromosome at 87F (Figure 2); the *Hrb98DE* locus is found on the same chromosome arm, but closer to the telomere, at 98DE. Based on its cytological location and its sequence homology to *Hrb98DE* (see below), this gene has been named *Hrb87F*.

Characterization of the Hrb87F gene

Figure 3 shows maps of the genomic DNA covering the *Hrb87F* locus and of three representative cDNA clones: R31, ov20 and R2-1. The diagram below the genomic map shows the exon-intron structure of the transcribed region, which is ~3.2 kb long. The *Hrb87F* gene has four exons, separated by introns ranging in size from 72 to 657 nt. The sequences surrounding the exon/intron junctions agree with the *Drosophila* consensus splice junction sequences (27). In contrast to what is seen for the *Hrb98DE* gene, we can find no evidence for the use of alternative N-terminal exons or splice sites, either by analysis of multiple cDNA clones or by S1 protection analysis of RNA from early embryonic stages (data not shown). However, there are two classes of cDNA clones, which differ in length of the 3′ untranslated region (507 or 980 nt). These classes probably correspond to usage of alternative polyadenylation sites, since most of the clones ended in a short poly (A) sequence, preceded at an appropriate distance by a consensus polyadenylation signal (28), AATAAA or a close variant.

The nucleotide and inferred amino acid sequences of the *Hrb87F* cDNAs are shown in Figure 4. The locations of the exon-intron boundaries are indicated by filled triangles and the polyadenylation signals are underlined. Both polyadenylation signals are within the same exon, and the choice of polyadenylation site does not affect the protein sequence. Primer extension and S1 protection analyses (data not shown) define a single transcription start site (nucleotide 1 in the figure). Translation probably begins at the first ATG (nt 133), which is in a reasonable context for *Drosophila* translation initiation (29). The encoded protein contains 386 amino acids, and has a calculated molecular mass of 39.5 kDa and a pI of 9.96. Both

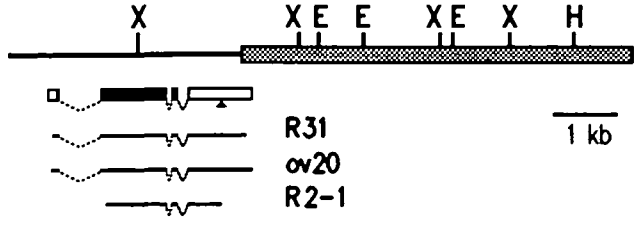

Figure 3. Map of the *Hrb87F* genomic region and cDNA clones. The top line shows a portion of the genomic clone containing the *Hrb87F* gene. The hatched box shows the location of the 412 transposable element in the Canton S genomic DNA. The exons of the *Hrb87F* transcripts are shown as boxes in the line below. Open boxes are untranslated regions, filled boxes are coding sequence, and the alternative polyadenylation site within the last exon is indicated by an arrowhead. The lines below show three of the cDNA clones that were analyzed; the R31 clone is truncated at the 3′ end. The sequences of the cDNAs diverge from that of the genomic Canton S DNA in the region of the 412 element. X = *Xho*I; E = *Eco*RI.

the nucleotide and amino acid sequences are very homologous to those of *Hrb98DE*, as expected from the Southern blot analysis presented in Figure 1 (see Figure 7 and the Discussion).

Transcripts from the *Hrb87F* gene are present throughout development, with an abundance profile similar to those of the *Hrb98DE* gene. Figure 5 shows that two transcripts, 2.2 and 1.7 kb, are found in ovaries and early embryos, consistent with the analysis of cDNA clones. The lower panel of the figure shows the hybridization of the same blot with a probe from the ribosomal protein gene *rp49*; these transcripts are present at approximately equal levels throughout development (30) and provide an estimate of the amount of RNA loaded in each lane. The levels of both *Hrb87F* transcripts decline and remain low in late embryogenesis and early larval development. In contrast to *Hrb98DE*, in which transcript levels remain low until pupation (16), the level of the 2.2 kb *Hrb87F* transcript begins to increase during the second larval instar. However, the level of the smaller *Hrb87F* transcript remains low throughout the rest of development. (The increase seen in the third larval instar lane is due to overloading; compare the relative levels of the *rp49* transcript.) Because the transcripts seen in Figure 5 represent steady-state levels of mRNA, we do not know whether the transcripts in late embryos and first instar larvae represent persistent maternal transcripts, or whether they represent newly synthesized zygotic transcripts. However, it is clear that late in development, when the *Hrb87F* locus is again being actively transcribed, the levels of the two transcripts are differentially regulated. This difference could arise by regulation of polyadenylation or by differential stability.

A transposable element insertion within the *Hrb87F* gene

Comparison of the genomic and cDNA sequences revealed that the last few hundred nucleotides of the cDNA sequence of clone R31 (and several others) did not correspond to the genomic sequence, but the sequences surrounding the point of divergence were not consistent with a splice site. Since R31 was derived from an Oregon R strain cDNA library, and the genomic clone was derived from a Canton S strain library, it was possible that the sequence differences were due to differences between the two strains. Consistent with this, genomic Southern blot experiments (data not shown) revealed restriction site variability in this region, e.g. the *Hrb87F* gene hybridizes to a 19 kb *Eco*RI fragment from Oregon R and a 5.8 kb fragment from Canton S genomic DNA. Also, the Canton S 5.8 kb *Eco*RI fragment hybridized strongly to multiple DNA bands, suggesting that it contained repetitive DNA. (Under the experimental conditions used, *pen* repeat sequences (15) would not be expected to account for the extra bands.)

The sequence organization of the genomic DNA from this region was compared with that of the cDNAs by polymerase chain reaction (PCR). Oligonucleotides designed to amplify the final 460 nt of the fourth exon were used in PCR reactions with genomic Oregon R DNA or a *Hrb87F* cDNA clone as template. The amplified fragments from the two PCR reactions were identical in size, indicating that the cDNA and the Oregon R genomic DNA are colinear in this region (data not shown). In contrast, no product was detected when Canton S genomic DNA was used as the template, although primers from outside of the region of divergence did generate the expected products. All of these observations are consistent with the idea that the sequence organization of the 3′ portion of the *Hrb87F* gene differs in Canton S and Oregon R, and that this difference could be due to the insertion of repetitive DNA. Comparison of the restriction

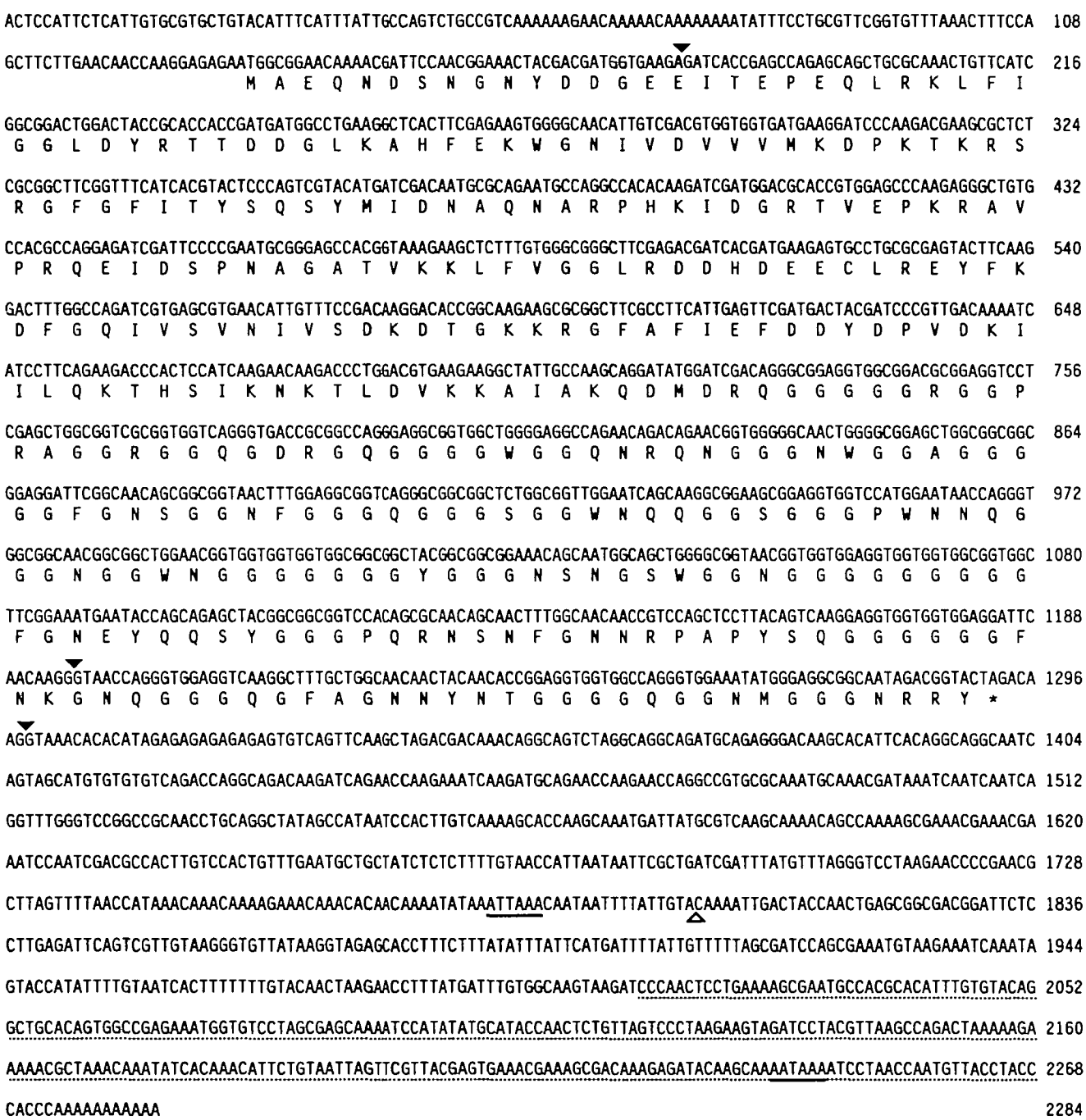

Figure 4. Sequence of the *Hrb87F* transcripts. The nucleotide sequence and inferred amino acid sequences are shown. The sequence is that of clone R31, except for the extreme 5′ and 3′ ends and regions of two short cloning artifacts; these sequences were determined from other clones. Exon/intron junctions are indicated by filled arrowheads; the junctions were determined by comparison of restriction maps of the genomic and cDNA clones and limited sequencing of the genomic DNA in regions of interest. Polyadenylation signals are indicated by a solid underline; the open triangle marks the end of the 1.7 kb transcript. The dotted underline marks sequences present in the cDNAs but missing from the Canton S genomic DNA clone.

map of this region and of the sequence around the point of divergence from the cDNA sequence (Figure 6A) with those of known *Drosophila* repetitive elements (31) revealed that a 412 transposable element (32,33) is present in the 3′ untranslated region of the *Hrb87F* gene in the Canton S strain (see Figure 3). The insertion site (nt 2011 in Figure 4) is located between the two alternative polyadenylation sites that are used in the Oregon R strain. It should be noted that not all Canton S isolates have this insertion. The isolate from which the ovarian cDNA library was made apparently does not, for clone ov20 and others obtained from this library have the Oregon R nucleotide sequence at the 3′ end.

Based on the genomic Southern experiments, it appears that the chromosome with the 412 insertion is homozygous in the Canton S flies. Although the 412 insertion is within the 3′ untranslated region and thus would not be expected to affect the protein sequence, it should alter or abolish expression of the larger *Hrb87F* transcript, potentially influencing protein levels. To

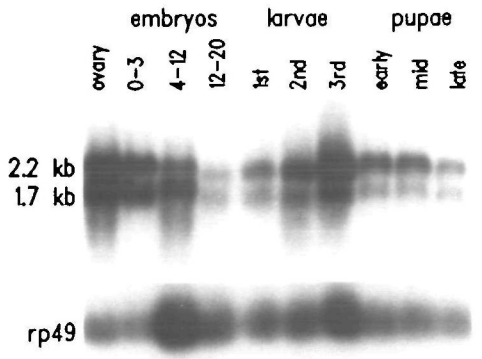

Figure 5. Developmental Northern blot of *Hrb87F* transcripts. 1 μg of poly $(A)^+$ RNA from each of the indicated stages was hybridized with a cDNA probe from the *Hrb87F* gene, then the blot was stripped and rehybridized with a probe for the *rp49* gene (30). Sizes were determined by comparison with RNA size markers.

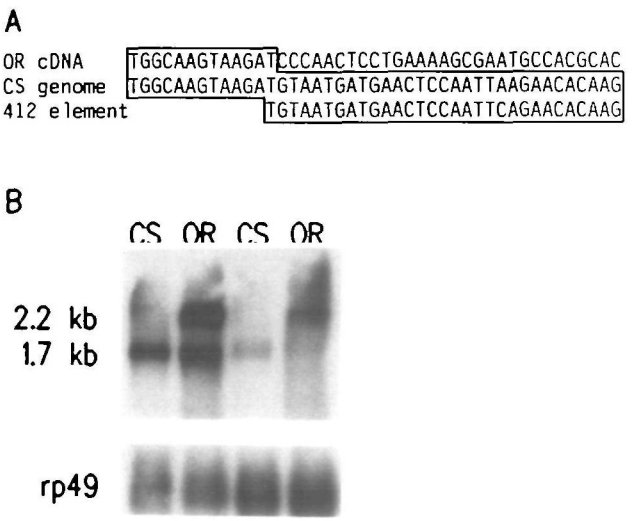

Figure 6. Identification of 412 element sequences and their effect on transcription of the *Hrb87F* gene. A. The sequences of the Oregon R (OR) *Hrb87F* cDNA (from nt 1998 of Fig. 4), the corresponding region of the Canton S (CS) genomic DNA, and the end of the LTR of the 412 element (33) are compared. B. 1 μg of poly $(A)^+$ RNA from CS females (lane 1), OR 0–3 hr embryos (lane 2) or third instar larvae (lanes 3 and 4) was hybridized with a probe from a *Hrb87F* cDNA clone. The blot was reprobed with fragment from the *rp49* gene as described in the legend to Figure 5.

examine this, poly A^+ RNA was prepared from Canton S adult female flies and mixed sex third instar larvae, and the *Hrb87F* transcripts were compared with those from Oregon R (Figure 6B). The Canton S females express predominantly one transcript (lane 1) of the same size as the smaller Oregon R transcript (lane 2). In Canton S larvae (lane 3), the level of the 1.7 kb transcript remains low, indicating that it is regulated similarly to the corresponding Oregon R transcript (lane 4). The amount of the 1.7 kb transcript in Canton S larvae is slightly elevated as compared to the amount of the 1.7 kb transcript in Oregon R larvae, but is clearly much less than the total amount of *Hrb87F* transcripts at this stage in Oregon R. This implies that the amount of protein produced from the 1.7 kb transcript alone is sufficient for normal development.

DISCUSSION

The studies reported here complement those performed in other organisms and provide a consistent view of the organization of genes encoding the A and B group of hnRNP proteins. These closely related hnRNP proteins are generated by transcription of related genes and alternative splicing of individual gene transcripts (9,12–14). In *Drosophila*, the *Hrb98DE* gene encodes four protein isoforms that are highly homologous to the A and B proteins (16). In addition, there are approximately 6 *Eco*RI fragments in the genome with detectable homology to the RNP motifs of the *Hrb98DE* gene. One of these encodes the *Hrb87F* gene, which appears to be the most closely related to the *Hrb98DE* gene. Other fragments presumably encode more distantly related genes. This has been confirmed for one of them; sequence analysis of the corresponding cDNA clone reveals a potential RNA binding protein with two copies of the RNP motif (S.R.H., unpublished data). Thus the *Drosophila* genome contains a family of genes encoding proteins related to the A and B hnRNP proteins. Whether all of them are components of *Drosophila* hnRNP complexes remains to be determined.

The *Hrb87F* transcription unit produces two major transcripts corresponding to usage of alternative polyadenylation sites. The transcripts are present at all stages of development, although usage of the polyadenylation sites may be developmentally regulated, and the absolute levels of the transcripts vary. The changes in abundance of the transcripts follow a pattern very similar to that seen for the *Hrb98DE* transcripts. Both genes are transcribed maternally, decay rapidly after early embryogenesis, and are synthesized again during late larval and pupal stages. In contrast to the approximately equal usage of the two polyadenylation sites in the maternal *Hrb87F* transcripts, the zygotic transcripts preferentially employ the 3′-most site; this site utilizes the consensus AATAAA polyadenylation signal, as opposed to the ATTAAA signal used at the first polyadenylation site. In at least one isolate of the Canton S strain, a 412 transposable element has inserted into the DNA between the two polyadenylation sites. In these flies, stable *Hrb87F* transcripts use only the first site. There is a very minor amount of a larger transcript, of approximately the size expected for transcripts extending to the second site, which can be seen on long exposure of the gel in Figure 6B. This could be due either to use of a cryptic polyadenylation site within the 412 element which fortuitously yields the 'correct' transcript size, or a low level of excision of the 412 element from the precursor by splicing, as has been shown to occur in a mutant of the *vermilion* gene (34).

The *Hrb87F* gene was isolated by cross-hybridization to a probe encoding only the RNP motifs of *Hrb98DE* but the homology between the two genes extends throughout the entire protein coding portion. There are two tandem copies of the RNP motif, followed by a glycine-rich C-terminal region. At the nucleotide sequence level, *Hrb98DE* and *Hrb87F* are 76% identical in the RNP motifs, and 67% identical in the glycine-rich regions. The similarity between the two genes extends to certain aspects of their exon/intron structure, which are also shared with the human A1 gene (35). The first exon of all these genes encodes the translation start and a few (4 to 16) additional amino acids, which are not well conserved between the various proteins. This short 'leader' peptide is followed, in the second exon, by the first RNP motif. Whether the variability in the sequences of the N-termini of the proteins indicates functional specialization is unknown. Of particular interest is the observation that in these three genes, the 3′-most exon is completely noncoding, and is preceded by

Figure 7. Comparison of amino acid sequences of A and B group hnRNP proteins. The sequences of the *Xenopus* A1 [XA1 (14)], human A1 [HA1 (41)] and B1 [HB1 (9)], and *Drosophila Hrb98DE* [D98 (16)] and *Hrb87F* (D87; this paper) proteins are shown; the two RNP motifs are aligned. Shaded residues of the RNP motifs are identical in all five proteins; between the individual sequences, a colon indicates identity and a period indicates conservative replacement. The + overlining marks positions within the RNP motif identified in (18) as being highly conserved (see text).

a short exon containing the translation stop codon. This unusual organization is found in fewer than 4% of all genes for which exon/intron boundaries are known (36), and its preservation in species as distant as humans and *Drosophila* suggests that it may be important for the proper regulation of these genes.

The RNP motif has been identified in over a dozen proteins, including snRNP proteins (37–40), components of hnRNP complexes (6,9,14,41,42), the poly(A) binding protein (17,43), two proteins in the *Drosophila* sex determination pathway (44,45), and nucleolin (46). Sequence comparisons demonstrate that the *Drosophila Hrb* proteins are most closely related to members of the A and B hnRNP protein group. The A1 protein is almost identical in the three mammalian species for which sequence data has been obtained (6,41,47). However, the *Xenopus* A1 proteins (14) have diverged from the mammalian ones, in both the RNP motifs and the glycine-rich regions. Figure 7 shows a comparison of the human and *Xenopus* A1 proteins with the human B1 sequence and the two *Drosophila* sequences. The upper two blocks contain the N-terminal 'leader' peptide and the RNP motifs, shown with the two copies aligned. The lower two blocks are the glycine-rich sequences. Within the glycine-rich regions, gaps have been introduced to maximize the alignment of conserved sequences. Overall, the C-terminal regions are poorly conserved, with multiple insertions (or deletions). However, there are small patches of homology that do not involve solely glycine residues; these tend to be most conserved between the three vertebrate sequences, or between the two *Drosophila* sequences, although there are a few regions common to all five. Within the RNP motifs, the two A1 sequences are 91% identical to each other, and 80% identical to the B1 sequence. For the *Drosophila*-vertebrate comparisons, the percentage of identical residues is much lower, ranging from 56–59%. However, it is significant that most of the conserved amino acids (46% of the total residues; shaded in Figure 7) are identical in all five proteins. Query et al. published an alignment of RNP motifs [termed the RNA recognition motif—RRM; see Fig. 7 of (18)], and a comparison of their consensus with the sequence identities shown here reveals some interesting features. The + symbols at the top of the two RRMs in Figure 7 mark the highly conserved residues identified in the RRM alignment. The sequence conservation among the A and B group proteins defines a region slightly larger than the conserved RRM, 91 amino acids vs 80 for the RRM. The residues which are highly conserved in all RRM motifs are similarly conserved here, but over half of the identities are in less conserved amino acids, and 25% are in positions that could not be assigned a consensus residue in the RRM comparison. These identities frequently are at different positions and involve different amino acids in the two RNP motifs. This suggests that this particular constellation of amino acid identities may define a subfamily within the family of proteins possessing RNP motifs. Since the RNP motif (plus a few additional residues) has been shown in several cases to be both necessary and sufficient for specific high affinity binding to RNA (18,19), some of these identical amino acids may be

important in specifying the binding interactions that distinguish the A and B proteins from other RNP motif-containing RNA binding proteins. In general, these hnRNP proteins exhibit less sequence specificity in RNA binding than do most RNP motif-containing proteins (18,19,48–50).

We have prepared antibodies to the *Drosophila Hrb* proteins and have obtained evidence that they are components of nuclear RNP complexes (G.R., S.R.H. and A.L.B., in preparation), consistent with the idea that they are hnRNP proteins. Are either of these proteins the *Drosophila* A1 or B1 protein? The overall structure of the *Drosophila* proteins (two RNP motifs followed by a C-terminal glycine-rich domain) and the sequence homologies clearly indicate that they are members of the same family as the A and B proteins. Both of them resemble the A1 protein slightly more than they do the B1 protein. There are 27 positions in which the RNP motifs of the two A1 proteins are identical but differ from the B1 protein. Some of these residues could be important for possible A1-specific functions, and thus serve to identify an A1-like protein. At these locations, the *Drosophila* proteins more frequently match the A1 rather than the B1 sequence (9–10 matches to A1 versus 4–6 matches to B1). However, in 40–50% of the positions, the *Drosophila* sequences match neither A1 nor B1, and thus such comparisons are inconclusive without further information regarding which are the critical residues. The real issue is whether the *Drosophila Hrb* proteins serve the same function(s) in hnRNP complexes as the A1 or B1 protein does. Resolving this question will involve further characterization of hnRNP complexes in both insects and mammals.

ACKNOWLEDGMENTS

We thank Paul Adler for help with the chromosome *in situ* hybridizations and Brian Kay for comments on the manuscript. This work was supported in part by Public Health Service grant GM39271 to A.L.B. A.L.B. is the recipient of an American Cancer Society Faculty Research Award. D.J. acknowledges sabbatical support from The George Washington University.

REFERENCES

1. Dreyfuss,G. (1986) *Ann. Rev. Cell Biol.*, **2**, 459–498.
2. Chung,S.Y. and Wooley,J. (1986) *Proteins: Structure, Function and Genetics*, **1**, 195–210.
3. Pinol-Roma,S., Choi,Y.D., Matunis,M. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.
4. Beyer,A.L., Christensen,M.E., Walker,B.W. and LeStourgeon,W.M. (1977) *Cell*, **11**, 127–138.
5. Lothstein,L., Arenstorf,H.P., Chung,S., Walker,B.W., Wooley,J.C. and LeStourgeon,W.M. (1985) *J. Cell Biol.*, **100**, 1570–1581.
6. Cobianchi,F., SenGupta,D.N., Zmudzka,B.Z. and Wilson,S.H. (1986) *J. Biol. Chem.*, **261**, 3536–3543.
7. Leser,G.P., Escara-Wilke,J. and Martin,T.E. (1984) *J. Biol. Chem.*, **259**, 1827–1833.
8. Merrill,B.M., Barnett,S.F., LeStourgeon,W.M. and Williams,K.R. (1989) *Nucleic Acids Res.*, **17**, 8441–8449.
9. Burd,C.G., Swanson,M.S., Gorlach,M. and Dreyfuss,G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9788–9792.
10. Wilk,H.-E., Werr,H., Friedrich,D., Kiltz,H.H. and Schafer,K.P. (1985) *Eur. J. Biochem.*, **146**, 71–81.
11. Leser,G.P. and Martin,T.E. (1987) *J. Cell Biol.*, **105**, 2083–2094.
12. Buvoli,M., Biamonti,G., Tsoulfas,P., Bassi,M.T., Ghetti,A., Riva,S. and Morandi,C. (1988) *Nucleic Acids Res.*, **16**, 3751–3770.
13. Buvoli,M., Cobianchi,F., Bestagno,M.G., Mangiarotti,A., Bassi,M.T., Biamonti,G. and Riva,S. (1990) *EMBO J.*, **9**, 1229–1235.
14. Kay,B.K., Sawhney,R.K. and Wilson,S.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1367–1371.
15. Haynes,S.R., Rebbert,M.L., Mozer,B.A., Forquignon,F. and Dawid,I.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1819–1823.
16. Haynes,S.R., Raychaudhuri,G. and Beyer,A.L. (1990) *Mol. Cell. Biol.*, **10**, 316–323.
17. Adam,S.A., Nakagawa,T., Swanson,M.S., Woodruff,T.K. and Dreyfuss,G. (1986) *Mol. Cell. Biol.*, **6**, 2932–2943.
18. Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89–101.
19. Scherly,D., Boelens,W., Van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.*, **8**, 4163–4170.
20. Scherly,D., Boelens,W., Dathan,N.A., Van Venrooij,W.J. and Mattaj,I.W. (1990) *Nature*, **345**, 502–506.
21. Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem.*, **263**, 3307–3313.
22. Cobianchi,F., Karpel,R.L., Williams,K.R., Notario,V. and Wilson,S.H. (1988) *J. Biol. Chem.*, **263**, 1063–1071.
23. Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.
24. Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
25. Steinhauer,W.R., Walsh,R.C. and Kalfayan,L.J. (1989) *Mol. Cell. Biol.*, **9**, 5726–5732.
26. Digan,M.E., Haynes,S.R., Mozer,B.A., Dawid,I.B., Forquignon,F. and Gans,M. (1986) *Dev. Biol.*, **114**, 161–169.
27. Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res.*, **15**, 7155–7174.
28. Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
29. Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
30. O'Connell,P. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495–5513.
31. Finnegan,D.J. and Fawcett,D.H. (1986) *Oxf. Surv. Eukaryot. Genes*, **3**, 1–62.
32. Rubin, G.M., Finnegan, D.J. and Hogness, D.S. (1976) in Progress in Nucleic Acid Research Vol. 19, Cohn, W. and Volkin, E. Eds. pp. 221–226, Academic Press, New York.
33. Will,B.M., Bayev,A.A. and Finnegan,D.J. (1981) *J. Mol. Biol.*, **153**, 897–915.
34. Fridell,R.A., Pret,A.-M. and Searles,L.L. (1990) *Genes Dev.*, **4**, 559–566.
35. Biamonti,G., Buvoli,M., Bassi,M.T., Morandi,C., Cobianchi,F. and Riva,S. (1989) *J. Mol. Biol.*, **207**, 491–503.
36. Hawkins,J.D. (1988) *Nucleic Acids Res.*, **16**, 9893–9908.
37. Theissen,J., Etzerodt,M., Reuter,R., Schneider,C., Lottspeich,F., Argos,P., Luhrmann,R. and Philipson,L. (1986) *EMBO J.*, **5**, 3209–3217.
38. Spritz,R.A., Strunk,K., Surowy,C.S., Hoch,S.O., Barton,D.E. and Francke,U. (1987) *Nucleic Acids Res.*, **15**, 10373–10391.
39. Sillekens,P.T.G., Habets,W.J., Beijer,R.P. and Van Venrooij,W.J. (1987) *EMBO J.*, **6**, 3841–3848.
40. Habets,W.J., Sillekens,P.T.G., Hoet,M.H., Schalken,J.A., Roebroek,A.J.M., Leunissen,J.A.M., van de Ven,W.J.M. and Van Venrooij,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2421–2425.
41. Riva,S., Morandi,C., Tsoulfas,P., Pandolfo,M., Biamonti,G., Merrill,B., Williams,K.R., Multhaup,G., Beyreuther,K., Werr,H., Henrich,B. and Schafer,K.P. (1986) *EMBO J.*, **5**, 2267–2273.
42. Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.*, **7**, 1731–1739.
43. Sachs,A.B., Bond,M.W. and Kornberg,R.D. (1986) *Cell*, **45**, 827–835.
44. Amrein,H., Gorman,M. and Nothiger,R. (1988) *Cell*, **55**, 1025–1035.
45. Bell,L.R., Maine,E.M., Schedl,P. and Cline,T.W. (1988) *Cell*, **55**, 1037–1046.
46. Lapeyre,B., Bourbon,H. and Amalric,R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1472–1476.
47. Williams,K.R., Stone,K.L., LoPresti,M.B., Merrill,B.M. and Planck,S.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5666–5670.
48. Wilk,H.-E., Angeli,G. and Schafer,K.P. (1983) *Biochemistry*, **22**, 4592–4600.
49. Conway,G., Wooley,J., Bibring,T. and LeStourgeon,W.M. (1988) *Mol. Cell. Biol.*, **8**, 2884–2895.
50. Pullman,J.M. and Martin,T.E. (1983) *J. Cell Biol.*, **97**, 99–111.
51. Haynes,S.R., Raychaudhuri,G., Johnson,D., Amero,S. and Beyer,A.L. (1990) *Mol. Biol. Reports*, **14**, 93–94.