Cloning of a grasshopper cDNA coding for a protein homologous to the A1, A2/B1 proteins of mammalian hnRNP

E.E.Ball[1, 2], E.J.Rehm[1] and C.S.Goodman[1]

[1]Howard Hughes Medical Institute, Department of Molecular and Cell Biology, LSA # 519, University of California, Berkeley, CA 94720, USA and [2]Molecular Neurobiology Group, Research School of Biological Sciences, Australian National University, PO Box 475, Canberra City, ACT 2601, Australia

Submitted November 28, 1990 EMBL accession no. X54670

We have cloned a 2217 bp cDNA that contains a 1026 bp open reading frame from the grasshopper, *Schistocerca americana* (Ball, E.E., Rehm, E.J. and Goodman, C.S. (1990) X54670). The putative 342 amino acid protein product of this gene is homologous to the A1, A2/B1 proteins of mammalian heterogeneous nuclear ribonucleoprotein (hnRNP) complexes (1, 2) and to the products of the *Drosophila* Hrb98DE locus (3, 4). All of these proteins contain similar, tandemly repeated RNA recognition motifs at their N terminal ends (Fig. 1). Although the degree of amino acid identity is reduced in their C-terminal portions, all of the proteins remain structurally very similar, consisting predominantly of glycine (Fig. 2). Tissue in situ hybridizations demonstrate that the grasshopper gene product is ubiquitous and abundant in embryos younger than 50% of embryonic development.

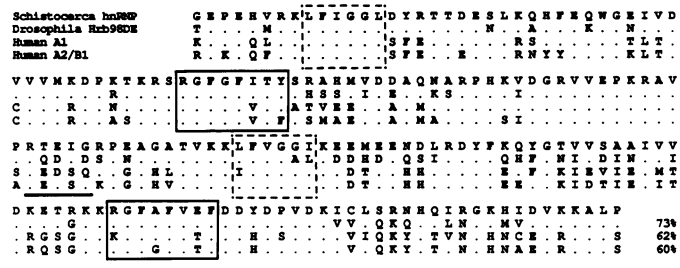

Figure 1. Comparison of the derived amino acid sequences of the RNA recognition motif of the *Schistocerca* hnRNP with the products of the *Drosophila* Hrb98DE locus (all of which are identical within these motifs) (4), with the human A1 hnRNP (1), and with the human A2 and B1 hnRNPs (2), which are identical in this portion of their RNA-recognition motifs. For clarity we have shown only the amino acids which differ from those in *Schistocerca*. The RNP 2 (dashed box) and the RNP-CS (solid box) are indicated, and the five amino acids connecting the two RNA-recognition motifs are underlined. Percent identity of the 175 amino acids shown for each molecule to the grasshopper sequence is given at the lower right.

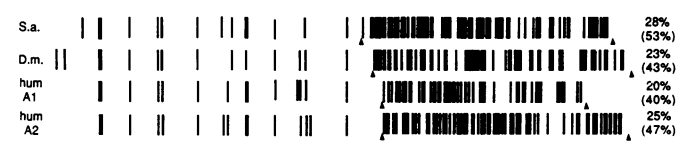

Figure 2. Occurrence of glycines in the A/B hnRNPs of *S. americana, D. melanogaster* (3), human A1 hnRNP (1) and human A2 hnRNP (2). Note the highly conserved pattern of glycine occurrence within the putative RNA-binding portions of all four molecules and the abundance of glycines at their 3′ ends. The percentage of glycine in the whole molecules and between the carats are indicated at the right, unbracketed and bracketed, respectively.

ACKNOWLEDGEMENTS

We thank Dr. Kai Zinn for the grasshopper cDNA library, Drs. Michael Levine and Tim Hoey for the probe used to clone the grasshopper hnRNP, members of the Goodman and Miklos labs for help and advice, and Ms. Sacha Holland for help with the sequencing.

REFERENCES

1. Biamonti,G., Buvoli,M., Bassi,M.T., Morandi,C., Cobianchi,F. and Riva,S. (1989) *J. Mol. Biol.* **207**, 491–503.
2. Burd,C.G., Swanson,M.S., Gorlach,M. and Dreyfuss,G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9788–9792.
3. Haynes,S.R., Rebbert,M.L., Mozer,B.A., Forquignon,F. and Dawid,I.B. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1819–1823.
4. Haynes,S.R., Raychaudhuri,G. and Beyer,A.L. (1990) *Mol. Cell. Biol.* **10**, 316–323.