Isolation of a murine gene encoding a nucleic acid-binding protein with homology to hnRNP K

Kyungmin Hahm, Gregg Kim, Christoph W.Turck[1] and Stephen T.Smale*
Howard Hughes Medical Institute, Molecular Biology Institute, and Dept. of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, CA 90024-1662 and [1]Protein Structure Lab. and Dept. of Medicine, University of California School of Medicine, San Francisco, CA 94143, USA

Received June 22, 1993; Accepted June 28, 1993 GenBank accession no. L19661

Within a eukaryotic cell, mRNA transcripts are incorporated into large, multiprotein complexes called the heteronuclear ribonucleoprotein complex (hnRNP) (1). This complex may play roles in mRNA processing, stability, and/or transport (1). Several of the proteins in this complex have been isolated, some of which contain clear RNA binding domains (2). Others lack obvious binding domains, yet interact with nucleic acids with high affinity (3, 4). One of these proteins, hnRNP K, binds to single-stranded poly(rC) and poly(dC) with exceptionally high affinity. Although hnRNP K lacks a consensus RNA binding domain, Siomi *et al.* (4) have speculated that a 45 amino acid repeated sequence mediates RNA binding. Here we report the isolation of a murine cDNA encoding a 40 kD protein with repeated domains that are homologous to those found in hnRNP K. Based on this homology, a ubiquitous expression pattern, and an unusually high abundance of the transcript, we suspect that this protein is a new member of the hnRNP complex.

The 40 kD protein was isolated from cultured murine T cells during the purification of a lymphocyte-specific transcription factor (5). Crude nuclear extracts from RLm11 cells were fractionated by heparin–agarose chromatography. A protein fraction eluting between 100 and 200 mM KCl was diluted to 100 mM KCl and applied to a DNA affinity column containing multimerized double-stranded oligonucleotides with the sequence TCCGTTTTGGGAGAAAGGA. The protein fraction was incubated with a large amount of poly (dI.dC) before applying it to the column. Three major proteins eluted from the affinity column between 0.3 and 0.6 M KCl. A candidate for the gene encoding the lymphocyte-specific transcription factor has been isolated and will be reported elsewhere (K.H. *et al.* unpublished data). A 40 kD protein also bound with high affinity to the column and we isolated its gene to determine its identity. The gene was isolated by first obtaining the amino acid sequences of four tryptic peptides derived from the purified protein. Degenerate oligonucleotide primers were prepared and used for polymerase chain reaction to isolate a portion of the cDNA fragment from murine thymoma RNA. The DNA fragment was then used to screen a murine cDNA library to isolate full-length cDNAs.

Five cDNAs were characterized, all of which contain DNA sequences that encode the four tryptic peptides. Two of these cDNAs appeared to be full-length and contained an open reading frame encoding a protein of 331 aa (36 kD). One of the partial cDNAs contained a 93 bp (31 aa) insertion at amino acid 193, which presumably results from alternative splicing. Within the protein sequence, three repeats were identified that are homologous to three repeats found in the hnRNP K protein (Figure 1, ref. 4). Homologous repeats are also found in a variety of other proteins (4), including the *E.coli* ribosomal protein S3 (4, 6) and the chicken vigillin gene (7). Although hnRNP K and S3 bind to RNA, the function of vigillin, which contains 14 of these repeats, remains unknown. Northern blot analysis revealed a 1.9 kb transcript that was found in numerous tissues and cell lines. The transcript appears to be extremely abundant, since more than 40 positive plaques were found on each plate of 50,000 during the library screen. Based on this information, we suspect that this protein will be a novel member of the hnRNP complex. Most likely, it bound to single-stranded oligonucleotides that did not anneal properly during preparation of our double-stranded DNA affinity column. At this point, we have no evidence to suggest that this protein binds to a specific nucleic acid sequence, and we anticipate that the protein will bind both to single-stranded DNA and to RNA. We have called this gene and protein hnRNP X.

ACKNOWLEDGEMENTS

This work was supported by NIH Grant # DK43726 and by the Howard Hughes Medical Institute.

REFERENCES

1. Dreyfuss,G. (1986) *Annu. Rev. Cell Biol.* **2**, 459–498.
2. Franker,A.D., Mattaj,I.W. and Rio,D.C. (1991) *Cell* **67**, 1041–1046.
3. Matunis,M.J., *et al.* (1992) *Mol. Cell. Biol.* **12**, 167–171.
4. Siomi,H., *et al.* (1993) *Nucleic Acids Res.* **21**, 1193–1198.
5. Lo,K., Landau,N.R. and Smale,S.T. (1991) *Mol. Cell. Biol.* **11**, 5229–5243.
6. Brauer,D. and Röming,R. (1979) *FEBS Lett.* **106**, 352–357.
7. Schmidt,C., *et al.* (1992) *Eur. J. Biochem.* **206**, 625–634.

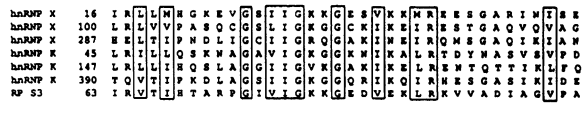

Figure 1. Comparison of the the amino acid sequence of murine hnRNP X with those of human hnRNP K and *E.coli* ribosomal protein S3. The number of the first amino acid is indicated and the most highly conserved residues are contained within boxes.

* To whom correspondence should be addressed