Vol. 195, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1993 Pages 317-326

ISOLATION AND STRUCTURAL CHARACTERIZATION OF THE RAT GENE ENCODING THE BRAIN SPECIFIC snRNP-ASSOCIATED POLYPEPTIDE "N"

Franca Esposito, Francesca Fiore, Filiberto Cimino and Tommaso Russo*

Dipartimento di Biochimica e Biotecnologie Mediche, Università degli Studi di Napoli "Federico II", via S. Pansini 5, 80131 Napoli, Italy

Received June 18, 1993

SUMMARY: We isolated the gene encoding the brain specific snRNP-associated polypeptide "N" from a rat genomic library. Some intronless genes have previously been reported for this polypeptide in the rat. In contrast, our gene consists of at least nine exons (the last exon(s) encoding the 3'-end of the mRNA has not been isolated). The second intron of the gene is probably so long as to hamper the cloning of the complete gene into a single phage vector. Therefore, we used polymerase chain reaction amplification of the rat genome with oligonucleotides designed on the basis of known cDNA sequences that allowed us to isolate and clone the first two exons at the 5'-end of the gene. Primer extension studies revealed multiple transcription start sites, all of them contained within the first exon. The intron/exon organization coincides with the alternative splicing events suggested by the sequences of the various cDNA species isolated so far. As also reported by others, genomic Southern analysis suggests the presence of other genomic regions containing sequences strongly hybridizing to the cDNA. We isolated and characterized two of these regions. They contain sequences very similar to the cDNA, not interrupted by introns and flanked by a polyA tail at their 3'-ends; hence they are considered two different pseudogenes.

The molecular basis of tissue-specific gene expression has been explored in several experimental systems. Numerous studies are leading to the identification of the promoter regions and of the trans-acting factors regulating transcription in a specific cell type (for a review see reff. 1,2). As yet, little is known about the molecular

*To whom correspondence should be addressed.

0006-291X/93 $4.00

mechanisms regulating gene expression in the mammalian nervous system. A useful approach to this problem is the identification and cloning of cDNA fragments hybridizing to mRNA esclusively present in rat brain (3-5).

Recently, we and others have cloned, from different sources, some cDNAs encoding a polypeptide, called polypeptide "N", which is associated to a small nuclear ribonucleoparticle (snRNP) (3,6-9). snRNP polypeptides, together with a number of snRNA (U1, U2, U3, U4/U6) and of non-snRNP polypeptides, interact with the pre-mRNA and are assembled in a large and structurally dynamic complex, the spliceosome, in which all the ordered steps of the splicing reactions take place (10,11). The expression of this gene, that we named "FE294", is restricted to brain. In fact, the mRNA is very abundant in rat brain, undetectable in rat liver, lung, kidney, testis (3) and it cross-hybridizes with related mRNA species present in rat heart but undetectable in rat brain (6). The isolation of mRNA species encoding an snRNP polypeptide only in the brain suggests its involvement in the regulation of tissue-specific splicings events, which are important mechanisms in the generation of protein diversity among different tissues and therefore in producing differentiated phenotypes.

The identification of the genomic sequence encoding this mRNA has been elusive. It clearly belongs to a multigene family that includes several intronless genes, and previous attempts to isolate the polypeptide "N" rat gene have led only to the identification of some intronless genes (12). This suggested that the true template for this rat mRNA species could lack introns, although the corresponding human gene, recently isolated and sequenced, shows a typical intron/exon organization (13).

Here we report the isolation and structural characterization of an intron-containing rat FE294 gene, that contains a complex intron/exon arrangement in the region encoding the 5'-untranslated region of the mRNA.

MATERIALS AND METHODS

Southern blot. Genomic DNA was prepared from rat blood cells (14); 15μg DNA were digested with EcoRI restriction endonuclease (Boheringer), size-fractionated on a 1% agarose gel, transfered to Nitran membranes (Schleicher and Schuell) and hybridized with the entire cDNA region ^{32}P-labeled by the random priming reactions (14). Hybridization was carried out at 65°C for 12hrs in Church buffer as described (15).

Screening of genomic DNA libraries. The two genomic DNA libraries (by Clontech) were constructed in the EMBL 3A and Charon 4A phage vectors, respectively. The entire cDNA region, or fragments of it, were ^{32}P-labeled with the random priming method (14) and used as probes; the nitrocellulose filters (Schleicher and Schuell) were prehybridized and hybridized at 65°C as described by Sambrook et al. (14). Washing was done at 65°C in 2X SSC, 0.1% SDS for 30 mins, 1X SSC, 0.1% SDS for 30 mins and in 0.1X SSC, 0.1% SDS for 30 mins (2 washings). Phage DNAs from the positive plaques, were purified according to the Qiagen protocol.

Nucleotide sequence analysis. Proper restriction fragments of the genomic DNA were subcloned into the pGEM 3Z vector and the nucleotide sequences of the exon and of the exon/intron junctions were determined with the dideoxy chain termination procedure (16) using the T7 DNA polymerase (Pharmacia).

PCR amplification. 1 μg rat genomic DNA was amplified with 30 cycles of PCR using the following protocol: denaturation at 94°C for 1 min, annealing to the appropriate oligonucleotide primers at 55°C for 1 min and elongation at 72°C for 3 min. The amplification products were then filled with T4 DNA polymerase (Boheringer), cloned into the SmaI site of pGEM 3Z vector and sequenced (16).

Primer extension analysis. The end-labeled oligonucleotide primer was hybridized to the rat brain and liver total RNA (16μg) and the RNAs reverse transcribed as described in ref.14. The extended products were separated on a 7M Urea /6% polyacrylamide gel, and the autoradiography was developed after a 3-day exposure at -80°C.

RESULTS AND DISCUSSION

The Southern blot analysis of the rat genomic DNA hybridized to the complete FE294 cDNA is shown in Fig. 1. The hybridization pattern of the EcoRI-digested DNA consists of numerous bands supporting the hypothesis that the FE294 gene belongs to a multigene family, and the presence of several strongly hybridizing bands confirms previous observations indicating the existence of FE294 pseudogenes (12).

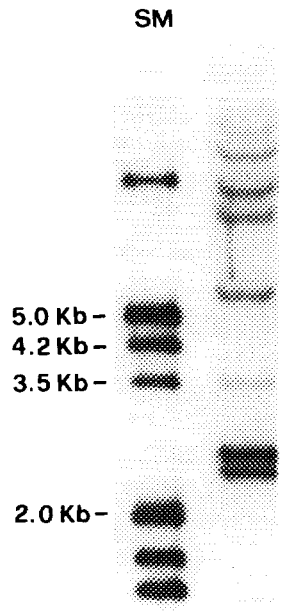

Figure 1. Southern blot analysis of rat genomic DNA hybridized to the FE294 cDNA probe. Rat genomic DNA preparation and hybridization conditions are described under Materials and Methods. Size markers (SM) are shown in the first lane and are from EcoRI-HindIII digested bacteriophage lambda DNA labeled with ^{32}P-dATP by filling-in the extremities. The second lane shows the hybridization pattern of the EcoRI digested DNA (see the text). The autoradiography was developed after a 7-day exposure at -80°C with intensifier screens.

We screened two genomic libraries in very stringent hybridization conditions using as probes the entire FE294 cDNA (about 1400 bp-long) or fragments of its 5' regions, to enrich for genomic clones containing the first exon(s) of the gene. Many positive clones were isolated and four of them, hybridizing at very stringent conditions, were characterized further. The sequence analysis of two of them demonstrated that these clones contain FE294 pseudogenes. In fact, both these genomic clones contain a nucleotide sequence almost identical to that of the cDNA, which is not interrupted by introns. Numerous point mutations are present either common to the two genes or peculiar to one of them. The two sequences show a polyA tail at the 3'-end of the cDNA homology and show different flanking sequences in which a direct repeat flanks the 5'-end and the 3'-end of the cDNA

homology. One of these two regions is identical to that of the previously published intronless gene hybridizing to the polypeptide "N" cDNA (12).

Two other genomic clones contain six exons of the gene (corresponding to the cDNA region from nt 100 to 1000), split by five introns (see Fig. 2). Unfortunately, both genomic clones, which cover a very similar genomic region, lack the first 100 bp of the untranslated region of the cDNA. In fact, as shown in Fig. 2, the first exon (Exo C) present in these clones is preceded by about 2000 nucleotides that are not homologous to the 5'-end of the cDNA. To isolate the 5' exons of the gene, we designed an oligonucleotide complementary to an intronic sequence, in which repeated sequences are absent, between exons C and D that would avoid the hybridization to pseudogenes. With this probe we screened two libraries, but failed to isolate phages containing sequences hybridizing to the leader region of the FE294 cDNA. This finding supports the hypothesis that a very long intron could have hampered the cloning of the leader exons together with exon C into a single phage vector.

To find the 5'-exon(s) missing from the genomic regions contained in the two cloned phages, we exploited PCR procedures based on the existence of two different leader sequences present in the known FE294 cDNAs. In fact, a cDNA species, named SM51, has been isolated from rat cells. This species has the same sequence and tissue distribution as FE294 mRNA, the only difference being its 5'-leader sequence upstream exon C (6). This observation supports the hypothesis that the leader sequences of clones SM51 and FE294 belong to different exons alternatively present in the mRNAs. We therefore synthesized various oligonucleotides complementary to these two different leader sequences and used them as primers to amplify rat genomic DNA with a standard PCR procedure (see Fig. 2, panel B - oligoPCR1 and oligoPCR2). We thus identified and cloned a 500-bp region containing the first two exons of the gene separated by a 200-bp intron (see Fig. 2, Panel A), named exon A and exon B, coding for the

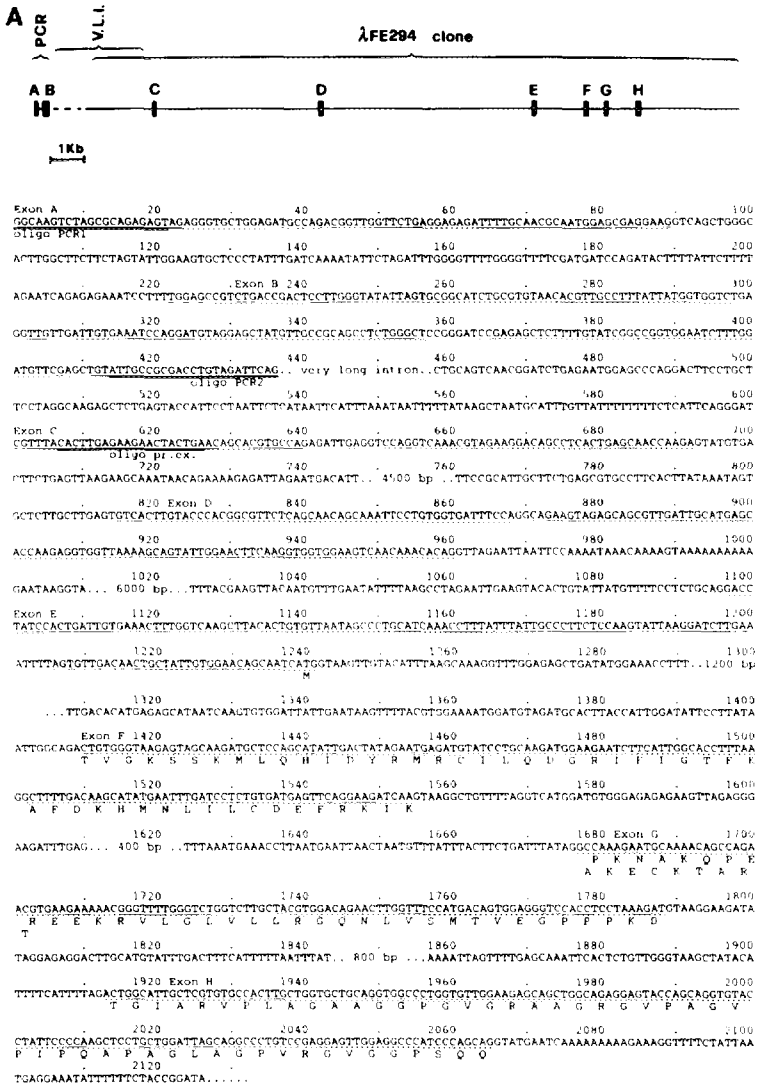

Figure 2. Structure and nucleotide sequence of the FE294 gene. Panel A: Exon-intron organization of the FE294 gene. Filled boxes labelled with letters from A to H indicate exons. Hatched line indicates the part of the second intron that was not cloned. PCR, polymerase chain reaction. V.L.I., very long intron. Panel B: Nucleotide sequence of the FE294 gene. Exonic sequences are underlined. Oligo PCR1 and oligo PCR2, the first identical and the second complementary to the thick underlined sequences, were used in the PCR amplification analyses as described under Materials and Methods. Oligo pr.ex., complementary to the thick underlined sequences, was used in the primer extension studies.

FE294 and SM51 leader sequences, respectively. We have no evidence that the SM51 exon is alternatively spliced from the pre-mRNA or that it contains a minor transcription start site, very poorly represented. In

fact, the primer extension analysis, performed on rat brain mRNA using an oligonucleotide complementary to the SM51-specific sequence (see below), was not contributory. However, the sequence of the first intron contains, just upstream from the region identical to the SM51 leader sequence, a consensus element for 3' splicing (Py_nNCAG-see Fig. 2, panel B) and this favours the hypothesis of an alternative splicing of exon B.

The unusual length of the second intron of the FE294 gene is probably why the 5' exons of the gene did not appear in our genomic clones. This is also in agreement with the results obtained by Schmauss and Lerner (12) who failed to clone the FE294 intron containing gene by PCR amplification of the rat genome. In fact, the intron/exon structure deduced from the sequence analysis of our clones indicates that two of their 5' oligonucleotide primers fall on two exon-intron junctions and therefore are not expected to prime synthesis. The third 5' oligonucleotide primer is probably very distant from the 3' oligonucleotide primer and therefore the amplification of intronless pseudogenes is favoured in such experiments.

To establish that exon A is the first exon of the gene we used a synthetic oligonucleotide, complementary to the upper strand of the exon C (see Fig. 2, panel B - oligo pr.ex.) as primer for the reverse transcription of total rat brain and liver RNAs. As shown in Fig. 3, multiple transcription start points are revealed by this analysis. From the length of these extension products we were able to locate all the transcription start sites within the first exon of the gene. The fact that we did not find extension products of higher molecular weight is evidence that the mRNAs including the exon B are minor species . This was confirmed by the primer extension analysis with an oligonucleotide complementary to the exon B (data not shown).

The intron-exon structure of the FE294 gene confirms the alternative presence of exons in various mRNA species (3, 6, 7). In fact, another cDNA species, SM21, sequenced by Li and co-workers (6), lacks 153 nucleotides that correspond to exon F (see Fig. 2, panel B).

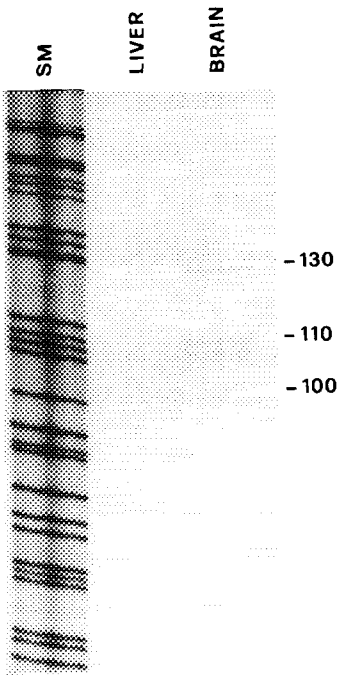

Figure 3. Primer extension analysis of FE294 mRNA. Total rat brain and liver RNAs were reverse-transcribed as described under Materials and Methods using as primer the oligo pr.ex reported in Fig. 2. The size markers are from the pUC19 sequence (ddT reaction).

The alternative splicing of this exon also warrants particular comments. The last three nucleotides of exon E are, in fact, the AUG methionine initiation codon that, as in many other eukaryotic genes, is contained in a separate exon. The link between exon E and exon G, observed in SM21 alternative splicing, results in a change of the reading frame. In this reading frame a stop codon is present 27 nucleotides downstream from AUG and therefore the mRNA would code for a 10 amino acid peptide. On the other hand, stop codons are present upstream from AUG in all the three reading frames and therefore there is no possible alternative translation start site.

A comparison between the published human gene (13) and the rat gene described here shows an evolutionary conserved intron/exon organization. The most striking difference between the two genes is that the first exon in the human gene corresponds to the fourth exon of the rat gene (exon D - see Fig. 2). However, it cannot be excluded that

also the human gene possesses further exonic sequences at the 5'-end of the gene that have not been identified. To ascertain if a further transcription start site was present in the rat gene at the level of exon D, we performed a primer extension analysis using rat brain RNA as template and an oligonucleotide complementary to the FE294 mRNA in the region encoded by exon D, as primer. This oligonucleotide primed the retrotranscription of a single-stranded DNA that terminates at the 5' end of the rat exon A, thus excluding the possibility of a significant transcription of the gene starting from the exon D.

ACKNOWLEDGMENTS

This work was supported by grants from CNR P.F."Ingegneria Genetica", P.F. "Invecchiamento", from AIRC and from MURST 40% and 60%.

REFERENCES

1. Mitchell, P. J. and Tjian, R. (1989) *Science* 243, 371-378.

2. Treacy, M. N. and Rosenfeld, M. G. (1992) *Annu. Rev. Neurosci.*15, 139-165.

3. Esposito, F., Ammendola, R., Duilio, A., Costanzo, F., Giordano, M., Zambrano, N., D'Agostino, P., Russo, T. and Cimino, F. (1990) *Dev. Neurosci.* 12, 373-381.

4. Milner, R. J. and Sutcliffe, J. G. (1983) *Nucl. Acids Res.* 11, 5497-5520.

5. Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Fields, C. and Venter, J. C. (1992) *Nature* 355, 632-633.

6. Li,S., Klein, E. S., Russo, A. F., Simmons, D. M., and Rosenfeld, M. G. (1989) *Proc. Natl. Acad. Sci.* USA 86,9778-9782.

7. Mc Allister, G., Amara, S. G. and Lerner, M. R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 5296-5300.

8. Schmauss, C., McAllister, G., Ohosone, Y., Hardin, J. A., and Lerner, M. R. (1989) *Nucl. Acids Res.* 17, 1733-1743.

9. McAllister, G., Roby-Shemkovitz, A., Amara, S.G., and Lerner, M. R. (1989) *EMBO J.* 8, 1177-1181.

10. Bindereief, A. and Green, M. R. (1990) *Genet. Eng.* 12, 201-224.

11. Guthrie, C. and Patterson, B. (1988) *Annu. Rev. Genet.* 22,387-419.

12. Schmauss, C. and Lerner, M. R. (1990) *J.Biol.Chem.* 265, 10733-10739.

13. Schmauss, C., Brines, M. L., and Lerner, M. R. (1992) *J. Biol. Chem.* 267, 8521-8529.

14. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory , Cold Spring Harbor, NY.

15.Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* 81,1991-1994.

16. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.