Volume 11 Number 6 1983 Nucleic Acids Research

Structural analysis of gene loci for rat U1 small nuclear RNA

N.Watanabe-Nagasu, Y.Itoh, T.Tani, K.Okano, N.Koga, N.Okada and Y.Ohshima

Institute of Biological Sciences, University of Tsukuba, Sakura-mura, Ibaraki-ken 305, Japan

Received 23 December 1982; Revised and Accepted 18 February 1983

ABSTRACT

Four phage clones which hybridize with U1 small nuclear RNA were obtained from a rat gene library. Two clones contain a presumed pseudogene. A third clone includes two gene candidates that are co-linear with the rat U1-RNA, 3.6kb apart and in the opposite orientation. The two genes are surrounded by identical sequences of 491bp upstream and 178bp downstream. The upstream sequences do not contain a TATA box, but share many block homologies with those for the human U1-RNA gene(1-3). A 101bp "identifier(ID) sequence", which was reported to be specifically expressed in rat brain(4), is inserted immediately after the shared sequence downstream of one of the genes. In the fourth clone, there are two putative pseudogenes, which have one or three nucleotide changes, 3kb apart and in the same orientation. Southern blot analysis of total rat DNA reveals about 50 U1-RNA genes/pseudogenes in the genome.

INTRODUCTION

In eukaryotic nuclei there is a group of small stable RNAs which are called small nuclear RNAs(snRNAs)(5). U1-RNA is the most abundant type in mammals. It is known to exist in the complex form with several proteins(6-8). The U1-RNA sequence(9, 10) has a region complementary to the consensus sequence at the intron-exon boundaries. This finding suggested involvement of U1-RNA in RNA splicing(11-13). Now there is evidence for this suggestion(14,15). As an approach to elucidate the nature and role of snRNAs, we have undertaken isolation and characterization of mammalian genes for these RNAs. A report on the sequence of gene loci for mouse U6-RNA has already been published (16). Here we report on cloning and sequence analysis of rat genomic loci which have a homology with U1-RNA. Although snRNA genes so far reported are dispersed, we have found two sets of duplicate gene loci.

MATERIALS AND METHODS

Isolation of recombinant phage clones

The rat gene library in Charon 4A, which had been prepared from Hae3 partial digest of DNA from Sprague-Dawley rat liver, was kindly provided by Dr. J. Bonner. The phage clones were screened by plaque hybridization(17) with U1-RNA which had been purified and labeled with ^{32}P, as described previously(16).

Structural analysis of the clones

Appropriate fragments in the phage DNA which hybridize with the U1-RNA were subcloned with pBR322 in χ1776 as described(16). Plasmid clones that include the desired fragment alone were selected by restriction enzyme analysis of plasmid preparations obtained by the small scale alkaline method(18). DNA sequencing was done by the method of Maxam and Gilbert(19). Other methods were described previously(16).

Southern blot analysis of total DNA

10μg DNA from the liver, brain or kidney of an 8 week male Wistar or Sprague-Dawley rat was digested with a restriction enzyme. Blotting of DNA from the gel was as described by Southern(20). The Sau3A-HhaI 120bp intragenic fragment of the U1-RNA gene in clone 3-1(A) was subcloned in pBR322 with EcoR1 linker after digestion with S1-nuclease. The inserted fragment was cut out and was labeled by nick-translation(21) for the probe. The Southern blot was hybridized with the probe as described by Wahl et al.(22) except for the omission of dextran sulfate.

RESULTS AND DISCUSSION

Analysis of the clones

About 100,000 phage plaques from the rat gene library were screened by hybridization with ^{32}P-labeled rat liver U1-RNA. Four positive clones were obtained. The DNA of the phages were analysed with restriction enzymes of 6bp recognition and Southern blot-hybridization with U1-RNA(Fig.1). As will be confirmed later, clone 6-6 and 3-1 include two hybridizing regions whereas clone 3-3 and 3-4 have one each. As shown in Fig.1, the restriction maps of the four clones are entirely different from one another, which indicate that they were derived from different

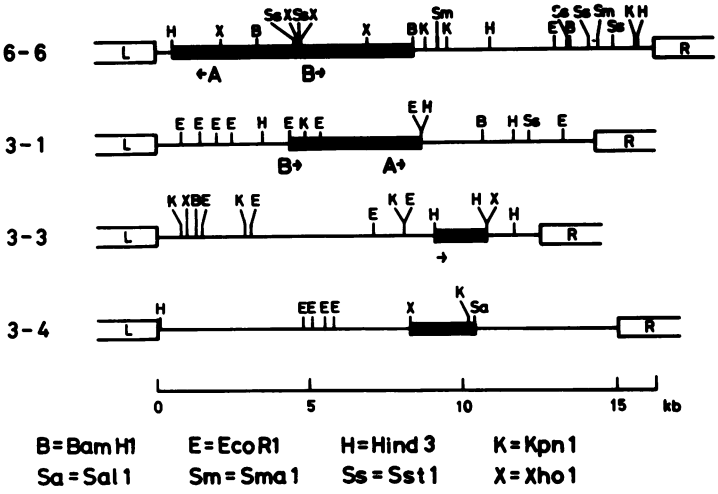

Fig.1. Restriction maps of the phage clones. The thick solid lines represent fragments which hybridize with U1-RNA. L and R denote Charon 4A arms. Arrows below the hybridizing region indicate the location and orientation of the U1-RNA gene or pseudogene.

loci in the rat genome.

The hybridizing fragments in the three clones(1.7kb Hind3 fragment in clone 3-3, 1.2kb and 3.3kb EcoR1 fragments in clone 3-1, 3.0kb Hind3-BamH1 and 5.1kb BamH1 fragments in clone 6-6) were subcloned in pBR322. Using the plasmid DNAs, detailed restsiction maps were constructed, and the gene regions were sequenced(Fig.2 and Fig.3).

Clone 6-6

Both the gene loci in clone 6-6 are completely co-linear with the rat U1-RNA(Fig.2). They are 3.6kb apart and in the opposite orientation(Fig.1). Interestingly, they are surrounded by identical sequences of 491bp upstream(up to the arrow in Fig. 3a) and 178bp downstream(down to the arrow in Fig.3b). Upstream of -491, no appreciable homology is found. Downstream of 342, there is another pair of identical sequences: 343 and down in region A, and 452 and down in region B(underlined in Fig.3b). Therefore the downstream sequence in region B is interrupted by a 109bp sequence. It has turned out to include a 101bp sequence (indicated by asterisks in Fig.3b) that is highly homologous,

```
          1       10        20        30        40        50        60        70        80        90
6-6 (A)   ATACTTACCTGGCAGGGGAGATACCATGATCACGAAGGTGGTTTTCCCAGGGCGAGGCTTATCCATTGCACTCCGGATGTGCTGACCCCT
6-6 (B)
3-1 (A)                                 T
3-1 (B)             T                   T
3-3       G                       G             -       A

                100       110       120       130       140       150       160
6-6 (A)   GCGATTTCCCCAAATGCGGGAAACTCGACTGCATAATTTGTGGTAGTGGGGGACTGCGTTCGCGCTCTCCCCTG
6-6 (B)
3-1 (A)
3-1 (B)                           T
3-3               T                 G                         T    A
```

Fig.2. The nucleotide sequences of the U1-RNA homology regions. The coding sequence in 6-6(A), which is completely co-linear with that of the U1-RNA, is shown at the top. For the other homologous loci, only nucleotide changes are shown.

although in the opposite orientation, to the "identifier(ID) sequences" described by Sutcliffe et al.(4, see the last section). The presence of an identifier sequence only downstream of gene B indicates that the two gene loci A and B have not been formed by a duplication in E. coli during the cloning procedure, but that the region cloned in 6-6 represents a rat genomic locus per se. A pair of 7bp direct repeats AGGCCAA(indicated by dots in Fig.3b) abut on the ID sequence found here. Therefore the region B seems to show an instance of transposition of an ID sequence, supporting the idea that it is movable.

In locus A upstream of the conserved region, there are direct repeats CA/CTCTC repeated 6 times(beginning at -540, shown by horizontal arrows in Fig.3a).

Formation of the two gene loci in clone 6-6 could be explained by three steps: duplication or gene conversion(23,24) of the ancestral single locus, inversion of either of the two daughter loci and transposition of an identifier sequence to locus B. The order of inversion and transposition might have been reverse.

Clone 3-1

The two loci in clone 3-1 that have homology with the U1-RNA are 3kb apart and in the same orientation(Fig.1). Locus A includes a sequence which has one nucleotide change from that of the U1-RNA coding sequence(Fig.2). The change is located at

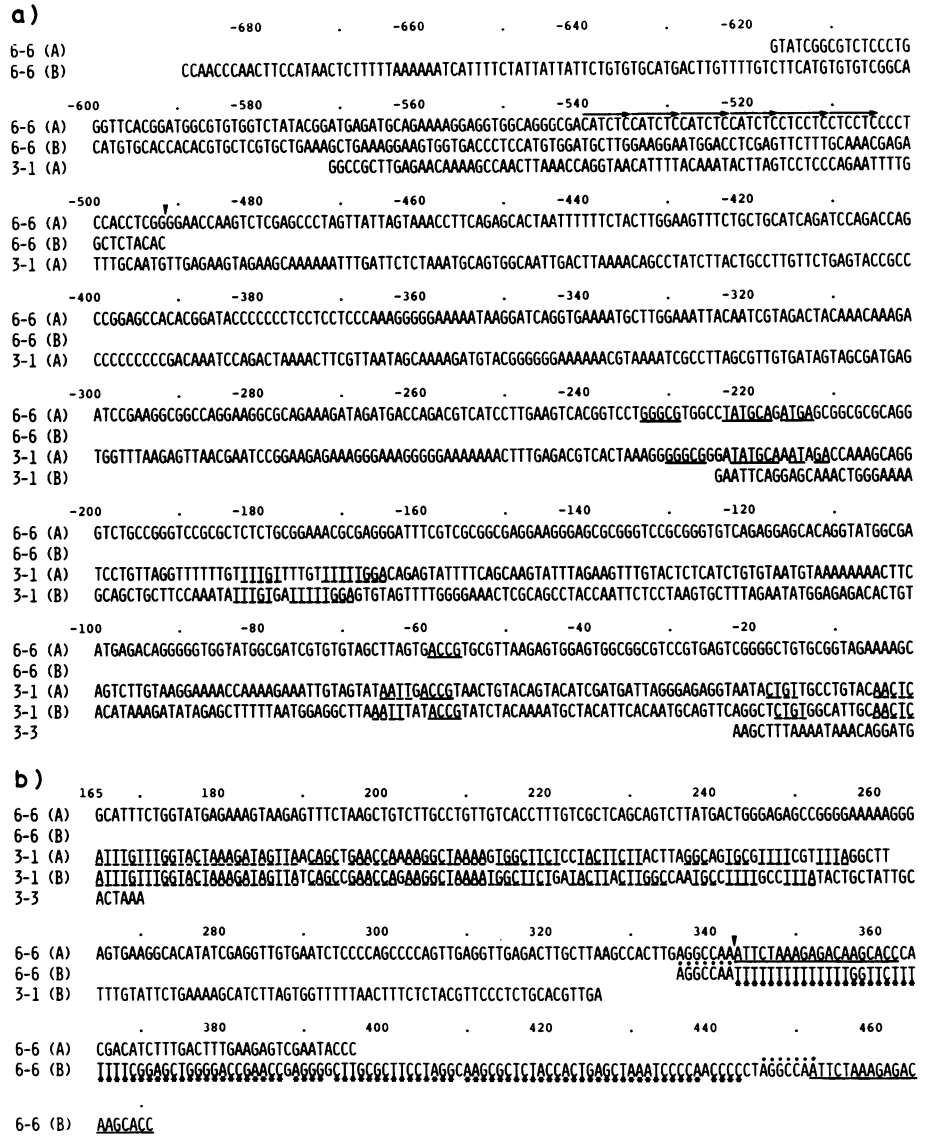

Fig.3. The nucleotide sequences of the regions upstream(a) or downstream(b) of the U1-RNA homologies. The sequences in 6-6(B) that are identical and in phase with those in 6-6(A) are left blank. A pair of sequences which are downstream of the arrow in (b) and shared by 6-6(A) and (B) are underlined. The sequence homologies between the genes in 6-6 and 3-1 are also underlined (a). The sequences shared by 3-1(A) and (B) are shown by dotted underlines. An identifier(ID) sequence inserted downstream of 6-6(B) is indicated by asterisks.

position 33, which is variable among the U1-RNAs of rat, man and chicken(10). Locus B includes a sequence which has three nucleotide changes. The 5'-flanking regions of loci A and B share only a limited homology with each other and with those flanking the genes in clone 6-6(Fig.3a). The 3'-flanking regions share a considerable homology with each other(about 80% for the regions sequenced, shown by dotted underlines in Fig.3b), but only a limited homology with those which are downstream of the genes in clone 6-6. The two loci in clone 3-1 are presumed to be pseudogenes.

The two loci in this clone may have been formed by duplication of a single gene locus by unequal crossing over(25), followed by genetic drift of the two daughter loci. Therefore clone 3-1 may be an example of pseudogene formation by duplication of an ancestral gene and following divergence. Such a mechanism has been postulated for "class I" pseudogenes of human snRNAs(26).

Duplicate gene loci found in clone 6-6 and 3-1 may suggest existence of a cluster of more than two U1-RNA gene loci, since duplicate sequences can expand by unequal crossing over(25). In fact, genes for rRNA and 5S RNA form large clusters(27).

Clones 3-3 and 3-4

In clone 3-3, there is a single hybridizing locus which has nucleotide changes at 8 positions(Fig.2). This locus seems to represent a typical pseudogene for U1-RNA. The locus is preceded by an A-rich sequence, which was also found surrounding probable pseudogenes for U6-snRNA(16).

We also tried to subclone the 1.8kb XhoI-SalI fragment of clone 3-4(Fig.1) into the SalI site of pBR322, but were unsuccesful. Therefore, the XhoI-SalI fragment was prepared from the phage DNA and was analysed with restriction enzymes of 4bp recognition. There is one hybridizing region in the fragment, but it lacks an Hpa2 site. Since the DNA which is co-linear with the rat U1-RNA should have sites for Sau3A, Hpa2, TaqI and HhaI in this order, the results suggest that there is a single pseudogene in clone 3-4.

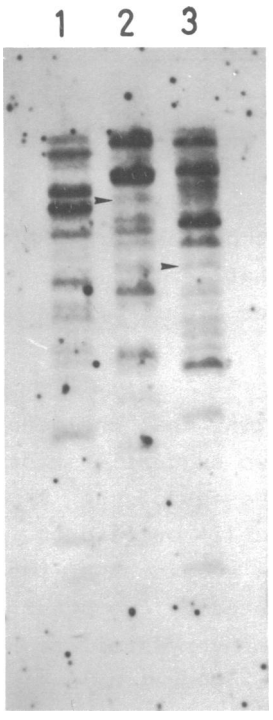

Fig.4. Southern blot analysis of rat DNA. 10μg DNA from Wistar rat liver was digested with EcoRl(lane 1), Sstl(lane 2) or BamHl (lane 3) for Southern blot-hybridization with 32P-labeled Sau3A-Hhal fragment within the Ul-RNA gene(3-1 A).

Multiple Ul-RNA gene loci in the rat genome

To obtain more information on the genomic organization of the Ul-RNA gene loci, total rat DNA was digested with restriction enzymes followed by Southern blot-hybridization(Fig.4). The probe used was the Sau3A-Hhal 120bp fragment within the Ul-RNA gene, which had been labeled with ^{32}P by nick-translation. Fig.4 shows that at least 15 positive bands are generated from the rat liver DNA with each of the three restriction enzymes (EcoRl, Sstl, BamHl). The 8.8kb Sstl fragment and 5.1kb BamHl fragment, which include the gene B in clone 6-6 seem to correspond to the minor bands shown by the arrows in Fig.4. Southern blot-hybridization with a DNA fragment within the shared region upstream of the two genes in clone 6-6 revealed two more gene loci linked with the conserved sequence(data not shown).

When we made a Southern blot such as shown in Fig.4 in parallel with a blot for copy number standard in which one or ten copy-equivalent amount of cloned DNA was used, a faintest band

exhibited approximately the same intensity as that of a single copy equivalent(data not shown). On the basis of these results, the total number of homologies to U1-RNA in the rat genome may be estimated to be about 50. We do not know whether the thick bands in Fig.4 are due to several or many fragments belonging to a conserved gene family, or a single fragment including several gene loci. Judging from the prevalence of pseudogenes for snRNA in mammalian genomes studied here or reported earlier(16, 26, 28, 29), the majority of the detected homologies are likely to be pseudogenes.

Expression of the U1-RNA gene

Although we have no direct evidence indicating that the gene A or B in clone 6-6 is functional, there are two suggestions for this idea. Fig.5 compares the 5'-flanking sequences of the genes in clone 6-6 with those flanking the human(2) or chicken (30) U1-RNA gene or mouse U6-RNA gene(16). The sequences of the 5'-flanking regions of human U1-RNA genes reported by two other groups(1, 3) are almost completely identical, or very similar, with the sequence shown in Fig.5(2). There are several block homologies between the sequences upstream of the human gene and our rat genes(underlined in Fig.5). Since the human U1-RNA gene

Fig.5. Comparison of the 5'-flanking sequences of the U1-RNA gene in clone 6-6, human U1-RNA gene(2), chicken U1-RNA gene (30), and mouse U6-RNA gene(16). Homologies shared by two or more clones are underlined. TATA box sequence before the U6-RNA gene is boxed.

has evidence indicating that it is functional(1, 2), these homologies suggest that the U1-RNA genes in clone 6-6 may be functional as well. The identity of the sequences surrounding the two genes in clone 6-6 may also suggest that these genes are functional.

The evidence has been reported indicating that U1-RNA is transcribed by RNA polymerase II(2, 30, 31). There are also suggestions for a much longer transcription unit for U1-RNA than the mature RNA(32, 33). The finding that the *in vitro* transcription for human U1-RNA genes starts 183bp upstream of the coding region reported by Murphy *et al.*(2) gives a direct support for the hypothesis of a long transcription unit. However, the suggestions cited above(32, 33), and the conservation of the sequences upstream of the human U1-RNA genes extending beyond 2kb(1) suggest that the U1-RNA transcription *in vivo* might start further upstream. The fact that we do not find any usual TATA box sequence in the upstream regions of more than 600bp in clone 6-6 may be in favor of this idea. In contrast, there is a TATA box sequence TATAAAT beginning 31bp before the mouse U6-RNA gene (boxed in Fig.5; Ref.16). We are currently studying *in vivo* transcripts of the U1-RNA genes starting upstream of the coding region.

Downstream of gene B in clone 6-6, there is a repetitive element called as "identifier(ID) sequence" in the opposite orientation(Fig.3b). Identifier sequences were found in rat brain-specific poly(A)-RNA(4), in an intron of the rat growth hormone gene(34) and in a pseudogene for rat α-tubulin(35). An identifier sequence might be regarded as a kind of enhancer sequence that enhances expression of a gene in a tissue(brain)-specific manner. An enhancer sequence identified so far is functional regardless of its orientation(36), and even when it is located downstream of a gene(37). Therefore it is possible that the expression of gene B is specifically enhanced in the brain.

ACKNOWLEDGEMENTS

We would like to thank Prof. Y. Watanabe and other members of the Institute for constant hospitality and encouragement. We

thank Mr. T. Sudoh for a part of sequencing and Mr. K. Tatei for preparation of plasmid DNA. This research was aided by Scientific Research Funds to Y. O. from the Ministry of Education of Japan.

REFERENCES

1. Manser, T. and Gesteland, R.F.(1982) Cell 29, 257-264.
2. Murphy, J.T., Burgess, R.R., Dahlberg, J.E. and Lund, E. (1982) Cell 29, 265-274.
3. Monstein, H-J., Westin, G., Philipson, L. and Pettersson,U. (1982) The EMBO J. 1, 133-137.
4. Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R.A. (1982) Proc. Natl. Acad. Sci. USA 79, 4942-4946.
5. Busch, H., Reddy, R., Rothblum, L. and Choi, Y.C.(1982) Ann. Rev. Biochem. 51, 617-653.
6. Lerner, M.R. and Steitz, J.A.(1979) Proc. Natl. Acad. Sci. USA 76, 5495-5499.
7. Lerner, M.R. and Steitz, J.A.(1981) Cell 25, 298-300.
8. Takano, M.T., Golden, S.S., Sharp, G.C. and Agris, P.F. (1981) Biochemistry 20, 5926-5936.
9. Reddy, R., Ro-Choi, T.S., Henning, D. and Busch, H.(1974) J. Biol. Chem. 249, 6486-6494.
10. Branlant, C., Krol, A., Ebel, J-P., Lazar, E., Gallinaro, H. , Jacob, M., Sri-Widada, J. and Jeanteur, P.(1980) Nucleic Acids Res.8, 4143-4154.
11. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A.(1980) Nature 283, 220-224.
12. Rogers, J. and Wall, R.(1981) Proc. Natl. Acad. Sci. USA 77, 1877-1879.
13. Ohshima, Y., Itoh, M., Okada, N. and Miyata, T.(1981) Proc. Natl. Acad. Sci. USA 78, 4471-4474.
14. Yang, V.W., Lerner, M.R., Steitz, J.A. and Flint, S.J.(1981) Proc. NAtl. Acad. Sci. USA 78, 1371-1375.
15. Calvet, J.P. and Pederson, T.(1981) Cell 26, 363-370.
16. Ohshima, Y., Okada, N., Tani, T., Itoh, Y. and Itoh, M. (1981) Nucleic Acids Res.9, 5145-5158.
17. Benton, W.D., and Davis, R.W.(1977) Science 196, 180-182.
18. Brinboim, H.C. and Doly, J.(1979) Nucleic Acids Res. 7, 1513 -1523.
19. Maxam, A. and Gilbert, W.(1980) in Methods in Enzymolozy, Grossman, L. and Moldave, K. Eds., Vol 65, p479-560, Academic Press, New York.
20. Southern, E.M.(1975) J. Mol. Biol. 98, 503-517.
21. Rigby, P.J.W., Dieckman, M., Rhodes, C. and Berg, P.(1977) J. Mol. Biol. 113, 237-251.
22. Wall, G.M., Stern, M. and Stark, G.R.(1979) Proc. Natl. Acad . Sci. USA 76, 3683-3687.
23. Slightom, J.L., Blechl, A.E. and Smithies, O.(1980) Cell 21, 627-638.
24. Baltimore, D.(1981) Cell 24, 592-594.
25. Smith, G.P.(1976) Science 191, 528-535.
26. Denison, R.A. and Weiner, A.M.(1982) Mol. Cell. Biol. 2, 815 -828.

27. Long, E.O. and Dawid, I.B.(1980) Ann. Rev. Biochem. 49, 727-764.
28. Denison, R.A., Van Arsdell, S.W., Bernstein, L.B. and Weiner, A.M.(1981) Proc. Natl. Acad. Sci. USA 78, 810-814.
29. Manser, T. and Gesteland, R.F.(1981) J. Mol. Appl. Genet. 1, 117-125.
30. Roop, D.R., Kristo, P., Stumph, W.E., Tsai, M.J. and O'Malley, B.W.(1981) Cell 23, 671-680.
31. Hellung-Larsen, P., Kuramowicz, I. and Frederiksen, S.(1980) Biochem. Biophy. Acta 609, 201-204.
32. Eliceiri, G.L.(1979) Nature 279, 80-81.
33. Tam, I., Kikuchi, T., Darnell, J.E., Jr. and Salditt-Georgieff, M.(1980) Biochemistry 19, 2743-2748.
34. Barta, A., Richards, R.I., Baxter, J.D. and Shine, J.(1981) Proc. Natl. Acad. Sci. USA 78, 4867-4871.
35. Lemischka, I. and Sharp, P.A.(1982) Nature 300, 330-335.
36. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P.(1981) Nucleic Acids Res. 9, 6047-6068.
37. Banerji, J., Rusconi, S. and Schaffner, W.(1981) Cell 27, 299-308.