Isolation and Characterization of Three Rat U3 RNA Pseudogenes Colinear with U3 RNA*

Ram Reddy, Dale Henning, S. Chirala, Lawrence Rothbium, David Wright, and Harris Busch

From the Department of Pharmacology, Baylor College of Medicine, and the Department of Genetics, University of Texas System, Cancer Center, M. D. Anderson Hospital, Houston, Texas 77030

Three different 15-kilobase rat genomic clones that contained sequences colinear with U3 RNA were isolated. These inserts hybridized only to U3 RNA in a mixture of total cellular 4-8 S RNA labeled in vivo which showed that genes or pseudogenes for most other small RNAs were absent in these U3 DNA clones.

DNA sequence analysis showed that the three subcloned genes contained full-length U3-coding sequences but each had sequence variations, insertions, and/or deletions when compared to rat U3A or U3B RNA. Two of these pseudogenes contained poly(A) sequences on the 3'-end and were flanked by 6-15-nucleotide long direct repeats. None of the three clones was transcribed when injected into Xenopus oocyte nuclei. One clone was a template for a small RNA slightly larger than U3 RNA, but this transcript was not related to the U3 RNA sequences. The structural features of two of these three U3 DNAs are supportive of the hypothesis that some pseudogenes arose from RNA-mediated DNA synthesis and insertion into the genome at random sites (Van Arsdel, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T., and Gesteland, R. F. (1981) Cell 26, 11-20). This is the first instance where full-length, colinear, U3 RNA pseudogenes have been isolated and characterized.

Studies on genes for small nuclear RNAs have yielded very interesting results. The genes of U1 and U2 RNAs appear to be clustered and/or tandemly repeated in several species (1-7). No intervening sequences have been found in any small RNA genes, and there appear to be 10 times as many U1 RNA pseudogenes in human and rodent genomes as real genes (8-10). It has been hypothesized that many of these pseudogenes arose by RNA-mediated DNA synthesis and integration into the genome at random sites (11-14).

U3 RNA is one of the six capped small nuclear RNAs found in all eukaryotic cells; this RNA is found only in the nucleolus and is in part hydrogen-bonded to preribosomal RNA (15-17). Although the detailed mechanism of action of this RNA is not known, it appears to be involved in the maturation of precursor ribosomal RNA to mature ribosomal RNA (15-17). At present, not much is known about the structure and organization of the genes coding for U3 RNA. A gene for D2 RNA isolated from Dictyostelium was shown to be 40% homologous to rat U3 RNA (18). The only studies on U3 RNA genes were by Bernstein et al. (11) who isolated and characterized four U3 pseudogenes containing the 5' 69 or 70 nucleotides of the U3 RNA sequence.

Since U3 RNA is localized to the nucleolus, it is interesting to study the structure and organization of U3 RNA genes and pseudogenes and compare their structure and frequency to those of U1 and U2 RNA which are localized to the nucleolosphere of the cell. The small nuclear RNA pseudogenes were classified into four classes, based on their structural characteristics (14). These are as follows: class I pseudogenes containing full-length U1 RNA with scattered base substitutions and showing extensive flanking homology with U1 true gene family; class II pseudogenes containing slight truncation at the 3'-end with scattered base substitutions, terminating in A-rich regions, and showing no flanking homology to U1 gene family; class III pseudogenes containing full-length U1 RNA with scattered base substitutions, terminating in A-rich or AT-rich region, flanked by short direct repeats, and showing no flanking homology to U1 true gene family; and class IV pseudogenes which are severely truncated at the 3'-end with few or no base substitutions, containing no A-rich region on the 3'-end, and flanked by short direct repeats. In this study, three clones containing sequences homologous to U3 RNA were isolated and characterized. All three clones contained DNA sequences colinear with U3 RNA; however, each of the three clones contained mismatches when compared to the two major U3 RNAs found in Novikoff hepatoma cells. Of the three pseudogenes characterized in this study, one pseudogene fits the class I type and two pseudogenes belong to class III. The results indicate that the mechanisms involved in the generation of U3 RNA pseudogenes were probably similar to those involved in the generation of U1 and U2 RNA pseudogenes.

MATERIALS AND METHODS

Enzymes—RNA ligase and polynucleotide kinase were from P-L Biochemicals. DNA ligase and restriction enzymes were obtained from New England Nuclear and Amersham Corp.

U3 RNA—Novikoff hepatoma U3 RNA containing U3A, U3B, and U3C RNAs was isolated from polyacrylamide gels (19) and was 3'-end-labeled with [32P]pCp and RNA ligase (20). The labeled RNA was again purified on an acrylamide gel and used for hybridizations. The specific activity of the RNA was 1 x 10^6 cpm/μg. The U3A RNA comprised 30% and U3B RNA 60% of the total U3 RNA, and they did not differ significantly in specific activity.

Isolation of Clones Containing U3 RNA Sequences—Screening of the rat genomic library in λ phage Charon 4A was carried out by the method of Benton and Davis (21). The partial EcoRI genomic library made from rat liver DNA in λ phage Charon 4A was kindly provided by Dr. T. Sargent and Dr. J. Bonner (Division of Biology, California Institute of Technology, Pasadena). Hybridizations with increasing stringency at 42, 52, or 58 °C were done in 50% formamide, 4 x SSPE (1 x SSPE is 0.15 M NaCl, 1 mM EDTA, 0.03 M Tris-HCl, pH 8.0, 2
x Denhardt's reagent (1 × Denhardt's reagent is 0.1% (w/v) bovine serum albumin, 0.1% (w/v) polyvinylpyrrolidone, 0.1% (w/v) Ficoll), 0.1% sodium dodecyl sulfate, and 10 µg/ml yeast or serum albumin, 0.1% (w/v) polyvinylpyrrolidone, 0.1% (w/v) Ficoll), transfer RNA. The nitrocellulose filters were hybridized for 16 h and washed two times each for 15 min with 3 x SSC, 1.5 × SSC, and 0.5 × SSC containing 0.1% sodium dodecyl sulfate at 42, 52, or 58 °C. The dried filters were autoradiographed using Kodak XR-5 film and DuPont Lightning Plus screens at -70 °C. Kinase labeling of DNA fragments and sequencing reactions were carried out according to Maxam and Gilbert (22).

Transcription of Plasmid DNAs Containing Sequences Homologous to U3 RNA—The transcription of plasmid DNAs in Xenopus oocyte nuclei was carried out as described by Kressmann et al. (23). In vitro transcription using different plasmids was carried out using whole cell extracts as described by Manley (24). Dot hybridizations using phage DNA were done as described by Kafatos et al. (25). The hybridized 32P-labeled small RNAs were eluted into water at 100 °C, precipitated with carrier tRNA, and electrophoresed on 5% polyacrylamide gels as described earlier (19). The in vivo 32P-labeled 4-8 S RNA used for dot hybridizations was prepared as described earlier (19) and had a specific activity of 1 × 10^6 cpm/µg RNA. Fingerprinting of RNAs and RNA transcripts obtained from in vitro transcription studies was carried out as described by Brownlee et al. (26).

RESULTS

Approximately 200,000 plaques of λ phage Charon 4A containing rat genomic DNA inserts were screened with 3'-end-labeled rat U3 RNA as a probe. Twenty-six positive clones from the first screening at 42 °C were rescreened at 52 and 58 °C. Three positive clones that hybridized at 52 °C were purified and assayed for their ability to hybridize to small RNAs using mega-dot hybridization (3). The whole cell 4-8 S RNA labeled in vivo used for hybridization is shown in of Fig. 1, lane 1. The three purified phages designated U3-1, U3-2, and U3-3 hybridized to U3 RNA, 4.5 S RNA and 6 S RNA, and 6 S-B RNAs hybridized to clones U3-2 and U3-3 (Fig. 1) possibly in regions homologous to type I Alu, type II Alu, or other repeated DNA sequences (27-30). Minor RNA bands noted in the size range of 100-150 nucleotides (Fig. 1) may be small RNAs, or possible degradation products of U3 RNA. These results indicated that clones U3-1, U3-2, and U3-3 contain sequences homologous to U3 RNA.

Mapping and Subcloning of U3 DNA-containing Fragments—Phage DNA from clones U3-1, U3-2, and U3-3 was digested with various restriction endonucleases, fractionated on 1% agarose gels, and transferred onto nitrocellulose filters. The DNA fragments containing U3 DNA were identified by hybridization to labeled U3 RNA (results not shown). The results showed that each phage clone was unique and contained only one U3 sequence. A 4.5-kilobase DNA segment obtained by digesting phage U3-1 with EcoRI and PstI enzymes, a 1.7-kilobase DNA segment obtained by digesting phage U3-2 with the SalI enzyme, and a 3-kilobase DNA segment obtained by digesting phage U3-3 with EcoRI and XhoI were subcloned into plasmid vector pBR322 in the appropriate restriction sites.

Sequencing of DNA Homologous to U3 RNA in the Plasmid Subclones—The plasmid subclones were further analyzed by digestion with various restriction enzymes and Southern hybridization analysis using labeled U3 RNA. The appropriate restriction fragments were sequenced using Maxam-Gilbert sequencing techniques (22). Fig. 2 shows the sequence of the three U3 DNA clones aligned with U3A RNA and U3B RNAs. The results show that the three U3 clones contain full-length U3 pseudogenes of 90-94% homology to the U3B RNA, and slightly less homology to U3A RNA (79-87%), but none of the clones contained sequences identical to U3A RNA or U3B RNA. Clone U3-3 was colinear with U3B RNA with no insertions or deletions (Fig. 2); however, there were 13 single nucleotide substitutions when compared to U3B RNA. Of the 13 substitutions, 11 were either purine → purine or pyrimidine → pyrimidine substitutions. The overall homology between clone U3-3 DNA and U3B RNA was 94%.

Clone U3-1 was colinear with U3B RNA but had a single three-nucleotide (TTC) deletion corresponding to nucleotides 94 and 95, and a dinucleotide TC, between nucleotides 164 and 165 (Fig. 2). In addition, there were 13 base substitutions in the U3-1 DNA when compared to U3B RNA. The overall homology was 94%. Clone U3-2 DNA was also colinear with U3B RNA, but two insertions were found. The insertions are the heptanucleotide AACAGC, between nucleotides 94 and 95, and a dinucleotide TC, between nucleotides 164 and 165 (Fig. 2). In addition, there were 22 base substitutions, of which 19 were either purine → purine or pyrimidine → pyrimidine transitions. The overall homology was 90%.

Sequences Flanking the U3 DNAs—The three U3 DNA clones contained poly(A) sequences on the 3'-ends; the length of this poly(A) sequence varied from 9 nucleotides in U3-1.
The +51 to +63 region that resembled AGAGCTGAGCCAT prepared from U3B RNA, and the cDNA was sequenced. The earlier studies (31, 32) contained several discrepancies with no significant homologies were found between the 5'-flanking sequences of the three U3 clones. The U3-1 DNA was found in the U3A RNA sequence in this region when 3' end-labeled U3A RNA was sequenced by the chemical degradation method of Peattie (33) (results not shown). The positions corresponding to 185 and 186 showed reactivity in G, C, and U reactions. Therefore, U or C is also indicated for these two positions of U3A RNA. The cap nucleotide is not included in the numbering, and the U3B RNA is numbered from +1 to +215 nucleotides. The 5'-flanking sequences are numbered -1 to -120, and the 3' flanking regions are numbered +1 to +70. The nucleotides identical to U3B RNA are marked by solid arrows under the sequences. The direct repeats found in U3-1, U3-2, and U3-3 DNA sequences are indicated by the horizontal arrows on top of the sequence.

### Fig. 2. Nucleotide sequence of three U3 pseudogenes and their flanking regions.

The DNA sequencing was carried out by the method of Maxam and Gilbert (22). The DNA sequences of the three pseudogenes, U3-1, U3-2, and U3-3, were aligned for maximum homology with rat U3A and U3B RNAs. The Novikoff hepatoma U3A and U3B sequences shown here contain modifications of earlier sequences (31, 32). The U3B sequence shown here was confirmed by sequencing a cDNA clone. The sequence of U3A RNA was confirmed by sequencing 3' end-labeled U3A RNA by the chemical degradation method of Peattie (33). The positions corresponding to 185 and 186 showed reactivity in G, C, and U reactions. Therefore, U or C is also indicated for these two positions of U3A RNA. The cap nucleotide is not included in the numbering, and the U3B RNA is numbered from +1 to +215 nucleotides. The 5'-flanking sequences are numbered -1 to -120, and the 3'-flanking regions are numbered +1 to +70. The nucleotides identical to U3B RNA are marked by solid arrows under the sequences. The direct repeats found in U3-1, U3-2, and U3-3 DNA sequences are indicated by the horizontal arrows on top of the sequence.

The broken lines indicate the optional extension of these direct repeats. A heptanucleotide repeat TAGGGGT found in the U3-3 pseudogene at position +24 to +37 is indicated by two horizontal arrows on top of the sequence.

### Isolation/Characterization of Three Rat U3 RNA Pseudogenes

Three clones containing U3 DNA sequences were isolated from the rat liver genomic bank and characterized. The three clones contained sequences that were 79-94% homologous to the two most abundant U3 RNA families (U3A and U3B) of Novikoff hepatoma cells. These genes could code for minor U3 RNA species or another U3 RNA expressed in different tissues or different stages of development as reported for other small nuclear RNAs (34, 35), the RNA transcript was analyzed by T1 RNase fingerprinting (results not shown). The -250-nucleotide long transcript was found to be unrelated to the U3-3 DNA sequence.

### Discussion

Three clones containing U3 DNA sequences were isolated from the rat liver genomic bank and characterized. The three clones contained sequences that were 79-94% homologous to the two most abundant U3 RNA families (U3A and U3B) of Novikoff hepatoma cells. These genes could code for minor U3 RNA species or another U3 RNA expressed in different tissues or different stages of development as reported for other small nuclear RNAs (36, 37). For example, the 5 S RNA of embryonic and adult tissues is produced from different developmental gene sets of Xenopus (37). Of the two U1 RNAs present in Ehrlich ascites cells (U1A, U1B) (38), only one U1 RNA is found in some mouse tissues (36).

However, it appears that these three U3 clones are pseudogenes for U3 RNA rather than real genes for the following reasons. The sequences of all three U3 DNAs did not match U3A or U3B RNA sequences, and U3-1 and U3-2 DNAs and U3-2 DNA to 18 nucleotides in U3-3 DNA (Fig. 2). In the U3-3 clone, the poly(A) was preceded by a CTTT sequence. The 5'-flanking sequences appear to be unique, and no significant homologies were found between the 5'-flanking sequences of the three U3 clones. The U3-1 DNA was found to be flanked by a 15-nucleotide long direct repeat with no mismatches. The U3-2 DNA was found to be flanked by a six-nucleotide long direct repeat; however, this six-nucleotide long sequence extended into the 5'-end of U3 RNA sequence. The U3-3 DNA contained AGAGCTGAGCCAT sequence at the +51 to +63 region that resembled AGAGCTGAGCCAT sequence at the -1 to -13 region (Fig. 2).

### Nucleotide Sequence of U3A and U3B RNAs

The nucleotide sequences of U3A and U3B RNAs determined in earlier studies (31, 32) contained several discrepancies with the three pseudogene sequences determined in this study and also with the four pseudogenes characterized by Berstein et al. (11). Therefore, these regions of U3 RNA were again sequenced, and the errors were corrected in the U3A and U3B sequences as published earlier (31, 32). A cDNA clone was prepared from U3B RNA, and the cDNA was sequenced. The cDNA contained sequences homologous to U3B RNA from nucleotides 109 to 215 and was colinear with U3-3 DNA. This corrected U3B sequence is shown in Fig. 2 and differs from the published U3B sequence beginning at nucleotide 140 (31). Two truncinucleotides, CAG and CUG, originally positioned at 155-160, were found instead of at 140-145. Similar errors were found in the U3A RNA sequence in this region when 3'-end-labeled U3A RNA was sequenced by the chemical degradation method of Peattie (33) (results not shown). The revised sequences are shown in Fig. 2.

### Transcription of Plasmid DNAs Containing U3 Sequences

Novikoff hepatoma cells contain at least three U3 RNAs, designated U3A, U3B, and U3C RNAs (31). Of these, U3A and U3B RNAs accounted for 90% of the total U3 RNAs (31, 32). One possibility is that U3-3 DNA may be a variant U3 gene which may code for another U3 RNA such as U3C RNA or a tissue-specific or developmentally regulated U3 RNA. To evaluate whether these U3 clones could be transcribed, the Xenopus oocytes were microinjected with plasmid DNAs, pU3-1, pU3-2, and pU3-3. No transcript corresponding to U3 RNA was detected when the transcripts were analyzed on acrylamide gels. Only U3-3 DNA coded for an RNA transcript which was slightly larger (~250 nucleotides) than U3 RNA. Since there are several reports that small nuclear RNAs such as U1 and U2 RNAs are synthesized as slightly larger precursors and processed to yield mature small nuclear RNAs (34, 35), the RNA transcript was analyzed by T1 RNase fingerprinting (results not shown). The ~250-nucleotide long transcript was found to be unrelated to the U3-3 DNA sequence.
contained insertions and/or deletions, and were not transcribed in *Xenopus* oocytes. Therefore, it is probable that all three U3 DNAs are pseudogenes.

Several types of pseudogenes for U1, U2, and U3 RNAs were characterized by Weiner and his colleagues (3, 8, 14). These were classified into four classes based on their structural characteristics. Class I U1 pseudogenes contain full-length U1 DNA and have extensive flanking homology with the bona fide U1 multigene family. The class II U1 pseudogenes are slightly truncated at the 3'-end, contain scattered base substitutions, terminate in A-rich regions, and contain no flanking homology to the U1 gene family. The class III U1 pseudogenes contain full-length U1 DNA with scattered base substitutions followed by A-rich or AT-rich regions, are flanked by short direct repeats, and contain no homology to the U1 gene family. The class IV pseudogenes are severely truncated on the 3'-end with few or no base substitutions, contain no A-rich regions, and are flanked by short direct repeats.

We attempted to fit the three U3 RNA pseudogenes characterized in this study into the above classification. The U3-1 pseudogene corresponded well to the class III pseudogenes in that it contained full-length DNA, scattered base substitutions, and an A-rich region on the 3'-end and was flanked by a 15-nucleotide long direct repeat (Fig. 2). The U3-2 DNA may also belong to class III in that it contained full-length DNA, scattered base substitutions, and an A-rich sequence on the 3'-end; however, the direct repeat was only six nucleotides long, and this AAAGAC sequence was also part of the sequence at the 3'-end was not immediately adjacent to the U3 RNA sequence, either poly(A) was added to a larger, truncated on the 3'-end with few or no base substitutions, or it belongs to class I pseudogenes and will show homology to U3B RNA.

Accordingly, U3-3 DNA may have arisen through mechanisms similar to the sequence 5718-5721 of U3B RNA is A, whereas the three pseudogenes contained G at the corresponding position, as did U3A RNA. Therefore, it is possible that these pseudogenes may not have arisen from the U3 RNAs now found in rat cells.

This study was initiated to attempt to isolate the real gene(s) coding for rat U3 RNAs. The protocol of Manser and Gesteland (10) was employed in which the stringency of hybridization was increased at each stage of plaque purification. These authors isolated six different U1 genes containing sequences identical to human U1 RNA. In this study, we screened the genomic bank equivalent to half the rat genome, and three clones that hybridized at 58°C were full-length pseudogenes. The estimates of the number of real genes for human U1 RNA are from 10 to 100 (2, 9, 40). Since the number of U3 RNA copies in both rat and human genomes is one-fifth that of U1 RNA (2 x 10^5 copies of U3 RNA versus 1 x 10^6 copies of U1 RNA), however, the probe used contained two times as many U3B RNA as U3A RNA, and this may be a factor. Perhaps more pseudogenes arose from U3B RNA, since it is more abundant in the cell. However, it is notable that position 188 in U3B RNA is A, whereas the three pseudogenes contained G at the corresponding position, as did U3A RNA. Therefore, it is possible that these pseudogenes may not have arisen from the U3 RNAs now found in rat cells.

**Acknowledgments**—We wish to thank Drs. T. Sargent and J. Bonner for providing the rat genomic library in a phage Charon 4A.

**REFERENCES**