A U3 RNA pseudogene in mouse: sequence and organization in genomic DNA

P. Ferrer, L.H. Qu, G. Bouche and J.P. Bachellerie*

Centre de Recherche de Biochimie et de Génétique Cellulaires du CNRS, Université Paul Sabatier, 31062 Toulouse cedex, France

A mouse U3 RNA pseudogene has been identified; it corresponds to a U3B full length coding sequence with a 3'-oligo(A) tail, precisely flanked at both ends by a pair of 15 bp direct repeats. These structural features suggest that it has arisen through an RNA-mediated mechanism involving an insertion at staggered nicks in the genome. Sequence data indicate that this mouse specimen has been generated by a different event as compared to the recently described rat pseudogenes. It represents the first reported case, for a pseudogene of this class, to be present at more than one copy per genome.

U3 RNA   Pseudogene   Retroposon   genomic DNA   (Rodent)   Evolution

1. INTRODUCTION

The capped small RNAs (termed U1–U7) found in the nucleus of eucaryotic cells have received much attention in the recent years mostly due to their potential involvement in the control of RNA processing; specific roles have been proposed for U1 [1,2] and U2 [3] in the splicing of pre mRNAs, for U4 [4] and U7 [5] in the formation of the 3'-end of mRNAs, and for U3 in the maturation of preribosomal RNA [6]. Surprisingly, in the mammalian genomes a majority of the small nuclear RNA coding sequences appear as pseudogenes [7] generated by mechanisms which still remain to be fully elucidated [8–10]. For U3 RNA, the only examples of pseudogenes were reported for human [11] and very recently for rat [12,13]. We describe here a mouse U3 RNA pseudogene which is likely to have been generated by a distinct genetic event as compared to the other mammalian specimens analyzed so far.

* To whom correspondence should be addressed

2. MATERIALS AND METHODS

2.1. Preparation of a U3 probe

Total cellular RNA was extracted from mouse liver according to [14] with some modifications [15]. U3 RNA was recovered from the 4–8 S fractions of a sucrose gradient centrifugation and further purified by 2 electrophoreses onto 7 M urea/acrylamide (first, 5%; second, 8%) gels in 50 mM Tris-boric acid, 1 mM EDTA, pH 8.3, and recovered as in [16]. U3 cDNA synthesis was carried out in 40 mM Tris-HCl, pH 8.3, 80 mM NaCl, 6 mM $MgCl_2$, 5 mM DTT in the presence of 5 μM [α-^{32}P]dATP and 500 μM dCTP, dGTP and dTTP, with 20 units AMV reverse transcriptase (Life Science). After 1 h incubation at 44°C, reaction was stopped by adding 10 mM EDTA and RNA was hydrolyzed by an overnight treatment in 0.5 M NaOH at 37°C. After neutralization by acetic acid, cDNA was analyzed on 7 M urea/8% acrylamide. Most of the radioactivity migrated as a 74 nucleotide cDNA band which was recovered from gel and purified according to [16].

2.2. Isolation and analysis clones containing U3 sequences

A recombinant library of mouse DNA randomly cleaved with AluI and HaeIII then inserted into CHARON 4A [17] was screened [18] with the ^{32}P-labelled U3 cDNA probe. Hybridization and filter washes were performed as described in the next section. Positive plaques were purified and rescreened at higher stringency (i.e. washed at 55°C in 0.1SSC, 0.1% SDS). Extracted phage DNA was then analyzed by restriction enzyme cleavage. DNA restriction fragments carrying sequences complementary to U3 RNA were subcloned into pBR322 (PP1 recombinant) or M13 vectors, and sequenced according to [16] or [19].

2.3. Genomic DNA analysis

High molecular mass genomic DNA was isolated from mouse liver according to [20]. Restricted DNA was separated by electrophoresis on 0.9% agarose gels then transferred onto nitrocellulose [21]. After a prehybridization step (42°C, 4–16 h), filters were hybridized (42°C, 16–20 h) with the nick-translated DNA probe in 50% formamide, 5 × SSC, 0.1% SDS, 50 μg/ml sonicated E. coli DNA, 5 × Denhardt, 40 mM NaP, pH 7.0. Subsequent washes at 42°C: 3 times for 10 min in 2 SSC, 0.1% SDS, then 3 times for 10 min in 0.1 SSC, 0.1% SDS.

3. RESULTS AND DISCUSSION

3.1. Isolation of a mouse genomic locus homologous to U3 RNA

Mouse U3 RNA, like its human and rat homologs [11] can act in vitro as a self-priming template for AMV reverse transcriptase, giving rise to a 74 nucleotide long cDNA product extending from the 3'-end of intact U3 RNA (not shown). This ^{32}P-labelled cDNA probe was used for screening about 10^5 plaques of recombinant phages by filter hybridization. Twenty plaques showing positive hybridization signals were rescreened at higher stringency, resulting in the final isolation of eleven recombinant clones. Preliminary restriction analyses of these 11 clones revealed that for 3 of them, the U3 sequence was arranged within a very similar environment extending over at least 8 kb of DNA. One of these 3 clones, termed λCH9, was selected for further analysis. A restriction map of

its 18 kb mouse DNA insert is shown in fig.1 (top). The sequence analysis around the U3 coding region was carried out after subcloning of the 1.3 kb BamHI fragment of λCH9 into either pBR322 (PP1 recombinant, fig.1, bottom) or M13 vectors.

3.2. Sequence of the U3 locus in PP1 recombinant

The sequence of a 630 bp region extending from the leftward BamHI site and encompassing the entire U3 coding region is shown in fig.2. This mouse U3 sequence appears exactly colinear with its rat U3A and U3B RNA homologs [12,13] over its entire length. Sequence alignments (fig.3) indicate a high level of homology with both rat U3 forms but reveals clearly that the mouse PP1 U3 sequence is much more related to rat U3B than to rat U3A (10 and 32 base changes, respectively, over the 215 bp). When the sequences flanking the mouse PP1 U3 coding region were compared to the regions flanking rat U3 bona fide genes [13] no significant homology was detected. A search for the presence of the conserved sequence elements (two upstream boxes around − 250 and − 50, and one downstream box at + 20) which could be involved in the transcriptional control of U3 genes [13] was negative. Instead examination of U3 flanking sequences reveals the classical characteristics of a retroposon sequence [7]: the U3 sequence boundaries correspond precisely to the limits of the U3 RNA species; an oligo(A) tail is precisely located at the 3'-end of the RNA sequence; and the oligo(A) tailed RNA sequence is immediately flanked by a pair of direct repeats at the 5'- and 3'-ends. By these characteristics, the
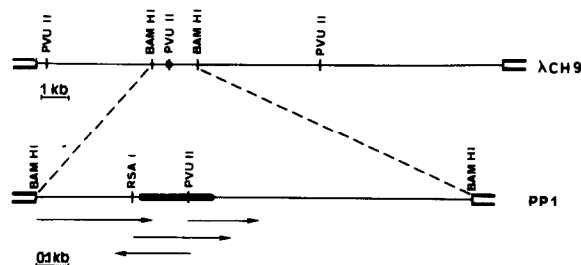

Fig.1. Structure of mouse DNA around the PP1 U3 pseudogene. The 18 kb insert in recombinant phage λCH9 is shown at the top. An expanded view of the 1.3 kb BamHI-BamHI fragment subcloned in pBR322 is shown at the bottom with indication of the sequenced segments (arrows).

mouse PP1-U3 clone clearly belongs to the class III-URNA pseudogenes [9,10]. These pseudogenes are thought to have arisen through an RNA-mediated DNA synthesis followed by insertion in the genome, with the flanking repeats likely to result from a repair synthesis at the staggered nicks of the target site. The 5'-terminal part of the direct repeat of PP1 is A-rich, a property also shared by human *Alu*I retroposons [22], suggesting that the integration process may have involved a direct hybridization between the 5'-end of the repeat and a T-rich 5'-end of a single-stranded cDNA.

Three full length U3 pseudogenes have been recently described for rat [12,13], all clearly generated from U3B RNA. One of them, termed U3.1 [12] is a typical class-III pseudogene. Sequence comparisons (fig.3) suggest that the mouse PP1 U3 pseudogene has not been generated by the same insertion event as any of the rat pseudogenes; the U3 coding region of mouse PP1 U3 is definitely much more closely related to the rat bona fide U3B gene (10 differences) than to any of the rat

pseudogenes (17, 19 and 29 differences). Moreover when mouse PP1 U3 and the rat class-III pseudogene are compared to the rat U3B RNA sequence, it is also noteworthy that among the 10 and 15 nucleotide changes, respectively, only one is common to both pseudogenes $(A \rightarrow G^{188})$. The possibility that both mouse and rat class III-U3 pseudogenes have been generated by independent genetic events is also reinforced by the examination of their U3 flanking regions, where no significant homology could be detected.

3.3. *Organization of the U3 sequences in the mouse genome*

We attempted to evaluate how representative the PP1 U3 form was among the U3 sequences present in the mouse genome, by Southern blot hybridization of restricted mouse genomic DNA.

In a first set of experiments (fig.4), we selected as hybridization probe the *Rsa*I-*Pvu*II fragment of PP1 recombinant (see fig.1 for location), extending from position -26 to position $+142$ by

```
                10         20         30         40         50         60
                 |          |          |          |          |          |
   1  GGATCCAAAC ACCCTTTTCT GGTTCCCATA GGCCGTGCAT GCACATGGAC ATACACACAG GTAAAATACT

  71  CATTGTGGTG GTTTGAATAC GCTTGGCCCA TGAGAAGAGG CGCTTTTAGG AGGTGTGACC TTATTGGAGG

 141  AAGTGTGTCA CTGAACACTA CCATCCCTCC TCCATGGCTG GGAAGAAAGA ACGAGAATAC GAGGAAACGA

 211  GGGTCCCACA GTCTCCTTCA TGACACACGT CTCTGGTAAC TTGAATGCTT GCTGTTGGCC CTTCTTCTCG

 281  AATGTACTCC GTTCCTAAAA GTGCTGCCCA CAAGACTATA CTTTCAGGGA TCATTTCTAT AGTTCGTTAC

 351  TAGAGAAGTT TCTCTGACTG TGTAGAGCAC CTGAAACCAC GAGGACGAGA CGTAGCGTCC CCTCCTGAGC

 421  GTGAAGCCGG CTCTAGGTGC TGCATGATTG CAGCTGCCTC TTGCCATTGA TGATCGTTCT TCTCTCCTTC

 491  AGGAGGGTGA GAGGGAAGGA ACGCAGTCTG AGTGGGAAAA AAAAAGTGCT GCCCGCTGGA AAGTGTCCCT

 561  TGGAACTCTG CCTTTCAACC GCTGGGCTGT TCTGAAGAAC TCGGGTGCAA GGTATTTAAC AGCCTCAAAG
```
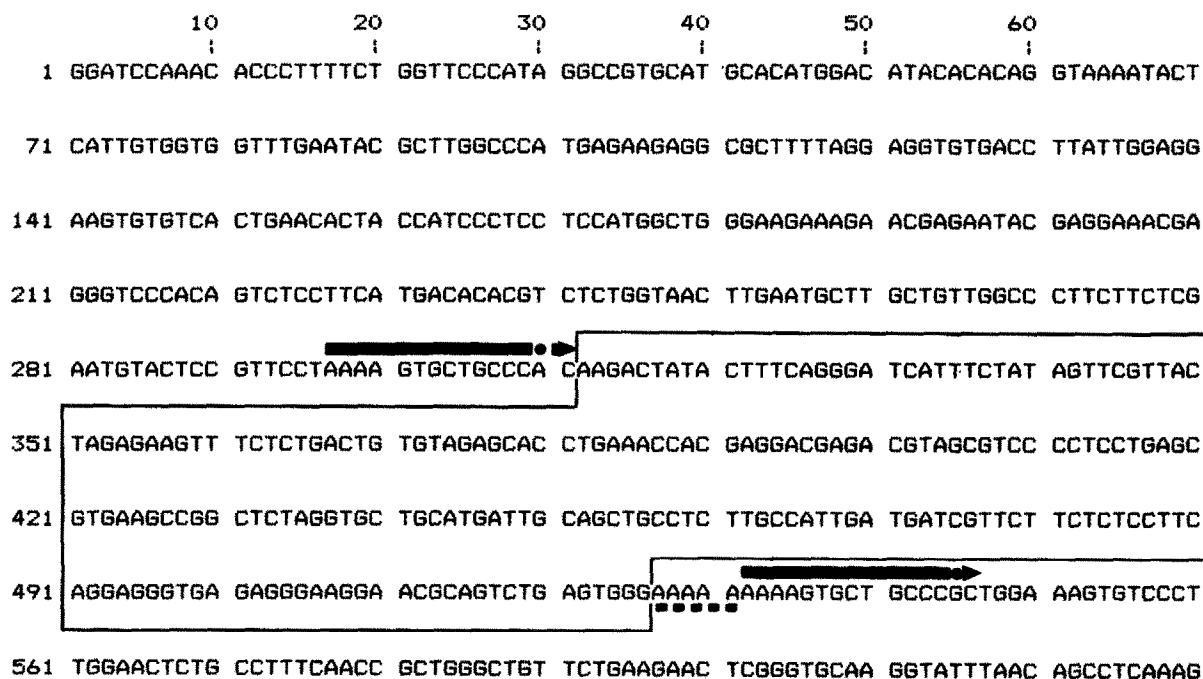
Fig.2. Sequence of the mouse U3 locus in PP1 recombinant. The sequence is numbered from the leftward *Bam*HI site. The boxed region is precisely colinear with rat U3 RNA_s [12,13] over its entire length. The broken underlining denotes the short poly(A) tract immediately downstream of the U3 coding region. The 15 bp direct repeats which flank the U3 coding region are overlined by arrows.
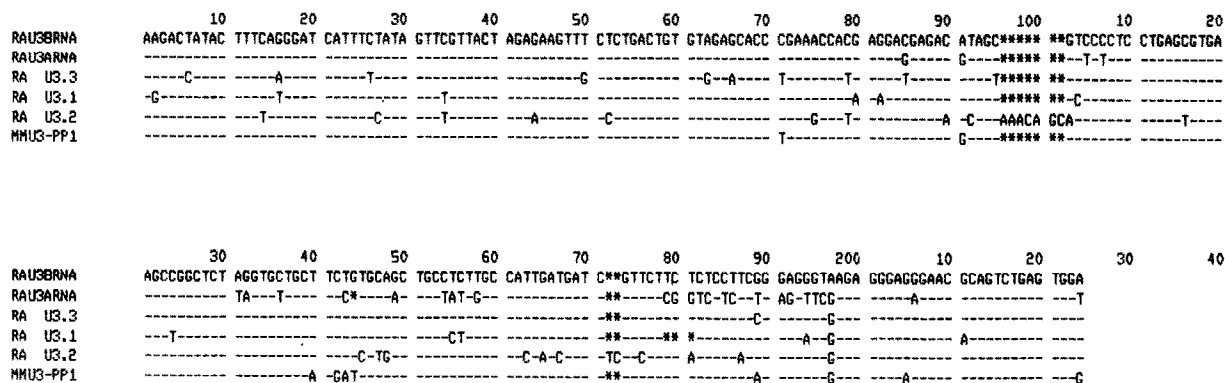
```
                   10         20         30         40         50         60         70         80         90        100         10         20
RAU3BRNA   AAGACTATAC TTTCAGGGAT CATTTCTATA GTTCGTTACT AGAGAAGTTT CTCTGACTGT GTAGAGCACC CGAAACCACG AGGACGAGAC ATAGC***** **GTCCCCTC CTGAGCGTGA
RAU3ARNA   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----C----- G----***** **--T-T--- ----------
RA  U3.3   -----C---- -----A---- -----T---- ---------- ---------G ---------- --G--A---- T-------T- ----T----- ----T***** **-------- ----------
RA  U3.1   -G-------- -----T---- ---------- ---T------ ---------- ---------- ---------A -A-------- -----***** **-C------ ----------
RA  U3.2   ---------- ---T------ ------C--- ---T------ ---A------ -C-------- ---------- ----G---T- ---------A -C--AAACA GCA------- -----T----
MHU3-PP1   ---------- ---------- ---------- ---------- ---------- ---------- T--------- ---------- G----***** **-------- ----------
```

```
                   30         40         50         60         70         80         90        200         10         20         30         40
RAU3BRNA   AGCCGGCTCT AGGTGCTGCT TCTGTGCAGC TGCCTCTTGC CATTGATGAT C**GTTCTTC TCTCCTTCGG GAGGGTAAGA GGGAGGGAAC GCAGTCTGAG TGGA
RAU3ARNA   ---------- TA---T---- --C*----A- ---TAT-G-- ---------- --**----CG GTC-TC--T- AG-TTCG--- -----A---- ---------- ---T
RA  U3.3   ---------- ---------- ---------- ---------- ---------- --**------ -------C-- ------G--- ---------- ---------- ----
RA  U3.1   ---T------ ---------- ---------- ----CT---- ---------- --**-----** *--------- ---A--G--- ---------- A--------- ----
RA  U3.2   ---------- ---------- ----C-TG-- ---------- --C-A-C--- -TC--C---- A-----A--- ------G--- ---------- ---------- ----
MHU3-PP1   ---------- --------A -GAT------ ---------- ---------- --**------ --------A- ------G--- ----A----- ---------- ---G
```

Fig.3. Alignment of mouse PP1 U3 sequence with the rat U3 genes and pseudogenes. Only differences from the rat U3B RNA sequence (top line) are shown. The rat U3A and U3B RNA sequences have been corrected as recently proposed [12]. The sequences and designation of the 3 rat U3 pseudogenes are from [12]. For rat U3A positions 185 and 186, which showed heterogeneity by RNA sequencing [12], the sequence from a cloned functional U3A gene was used [13]. The correction from G to A at position 182 of rat U3B sequence is as proposed in [13]. Note that these alignments are numbered by reference to rat pseudogene U3.2 (longer than U3B or U3A by 9 nucleotides).

reference to the 5'-end of the U3 coding sequence. Since the stretch of 5'-flanking sequences present in this fragment was very short relative to the length of the U3 coding region, this probe was meant to detect indistinctly any form of U3 sequence in the mouse genome in our hybridization conditions. For each type of digestion, a rather complex pattern of radioactive bands was ob-

tained, in addition to the band detected in λCH9 DNA (denoted by a star). These data indicate that about 20–25 copies of U3 coding sequences are present in the mouse genome, distributed over about 10 different types of DNA fragments after restriction by either BamHI, PvuII or PstI. As for the particular DNA band, also detected in the λCH9 recombinant for each type of digestion, it corresponds to about 4–5 copies of U3 genomic equivalent in each case.
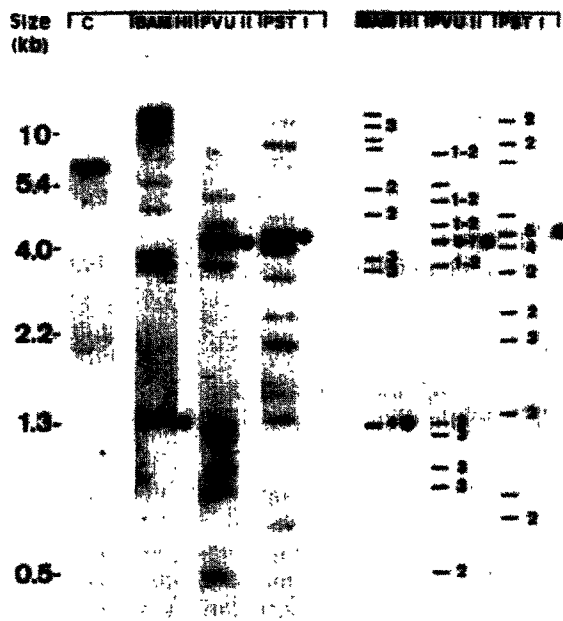

Fig.4. Southern blot analysis of mouse genomic DNA with U3 coding sequences. (A) Identical amounts (10 µg) of mouse liver DNA were restricted as indicated. The digests were analyzed by agarose gel electrophoresis, transferred onto nitrocellulose and hybridized with the nick-translated 170 bp RsaI-PvuII fragment of PP1 recombinant. In lane C, 22 pg of PP1 recombinant DNA were analyzed in parallel; for this amount of marker DNA, the intensity of the signal must correspond to what would be obtained, from 10 µg of mouse genomic DNA, for a fragment present in 5 copies per genome. (B) Schematic representation of the autoradiograph; radioactive bands are represented by bars with the approximate number of genomic equivalents (when higher than 1) indicated on the right side of each lane. Bands denoted by a thicker bar and a star (on the right side of the lane) were also detected when a similar Southern blot analysis was carried out on λCH9 recombinant DNA (not shown).
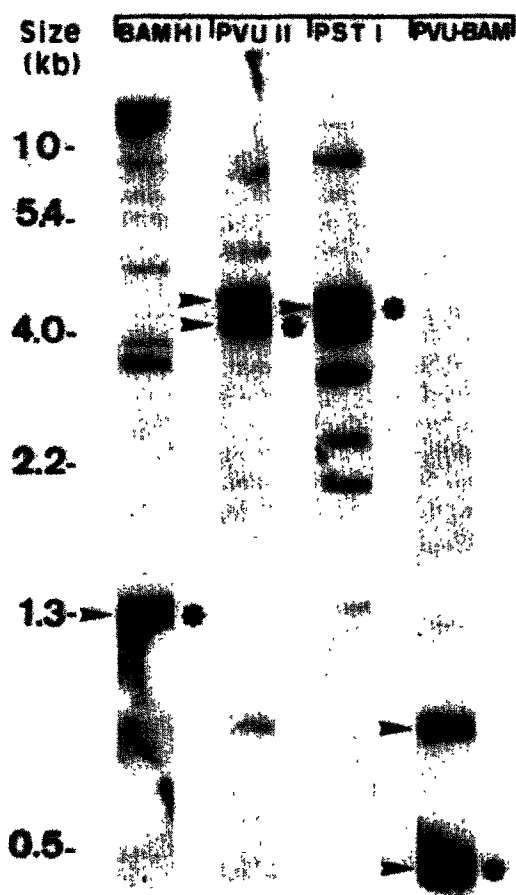
Fig.5. Southern blot analysis of mouse genomic DNA with PP1 sequences. (A) Identical amount (10 μg) of mouse liver DNA were restricted as indicated. The digests were analyzed by agarose gel electrophoresis, transferred onto nitrocellulose and hybridized with the nick-translated 1.3 kb BamHI-BamHI insert of PP1 recombinant plasmid. Fragments which were also detected in a similar Southern blot analysis of λCH9 recombinant DNA (not shown) are denoted by stars. Arrowheads point to bands which either were negative with the RsaI-PvuII probe (compare to fig.4) or give selectively enhanced hybridization signals with the 1.3 kb BamHI-BamHI (as compared to the RsaI-PvuII probe).

To confirm the identification and titration of the genomic DNA fragments represented in the λCH9 recombinant, a second set of experiments was carried out (fig.5) using total PP1 DNA insert as hybridization probe. Taking into account both the nick-translation conditions (average size of the

labelled fragments, 150–200 nu) and the relatively small (about 15%) contribution of U3 coding sequences within the entire PP1 insert, this second probe was expected to result in markedly enhanced signals for fragments present in PP1 as compared to fragments from other types of U3 genomic loci. Results shown in fig.5 are in complete agreement with this prediction, with all the intensified bands (as compared to fig.4) also present in the λCH9 insert. Moreover their intensity is again consistent with the U3 locus cloned in λCH9 being present at 4–5 copies per genome (quantitative markers not shown).

Class-III pseudogenes described so far are inserted into sequences present as single copy in the genome [9]. Accordingly the mouse PP1 U3 specimen represents a new form of class-III elements. One may assume that it was first inserted into one of the 4–5 genomic copies of a long (more than 8 kb) DNA unit, then propagated to the other copies, through recombinational events [23]. Alternatively the target site may have been a single copy DNA and the insertion followed later by a few duplications of a large domain around the pseudogene.

The sequence analysis (now in progress) of other mouse U3 recombinants corresponding to different genomic loci could provide new insights into the mechanisms involved in the appearance and dispersion of U3 pseudogenes during the evolution of rodents and their potential relevance to the proposed role of U3 RNA in pre-rRNA processing [6].

REFERENCES

[1] Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220–224.
[2] Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) Cell 33, 509–518.
[3] Krainer, A.R. and Maniatis, T. (1985) Cell 42, 725–736.

[4] Berget, S.M. (1984) Nature 309, 179–182.

[5] Strub, K., Galli, G., Busslinger, M. and Birnstiel, M.L. (1984) EMBO J. 3, 2801–2807.

[6] Bachellerie, J.P., Michot, B. and Raynal, F. (1983) Mol. Biol. Rep. 9, 79–86.

[7] Rogers, J.H. (1985) Int. Rev. Cytol. 93, 231–279.

[8] Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) Cell 26, 11–17

[9] Denison, R.A. and Weiner, A.M. (1982) Mol. Cell Biol. 2, 815–828.

[10] Weiner, A.M. and Denison, R.A. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 1141–1149.

[11] Berstein, L.B., Mount, S.M. and Weiner, A.M. (1983) Cell 32, 461–472.

[12] Reddy, R., Henning, D., Chirala, S., Rothblum, L., Wright, D. and Busch, H. (1985) J. Biol. Chem. 260, 5715–5719.

[13] Stroke, I.L. and Weiner, A.M. (1985) J. Mol. Biol. 184, 183–193.

[14] Auffray, C. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314.

[15] Le Meur, M., Glanville, N., Mandel, J.L., Gerlinger, P., Palmiter, R. and Chambon, P. (1981) Cell 23, 561–571.

[16] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.

[17] Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687–701.

[18] Benton, W.D. and Davis, R.W. (1977) Science 196, 161–169.

[19] Sanger, F., Nicklen, S. and Coulson, R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[20] Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32–38.

[21] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.

[22] Daniels, G.R. and Deininger, P.L. (1985) Nucleic Acids Res. 13, 8939–8954.

[23] Scherer, S. and Davis, R.W. (1980) Science 209, 1380–1384.