MOLECULAR AND CELLULAR BIOLOGY, Dec. 1985, p. 3560–3576
0270-7306/85/123560-18$02.00/0

Vol. 5, No. 12

Characterization of the Multigene Family Encoding the Mouse S16 Ribosomal Protein: Strategy for Distinguishing an Expressed Gene from Its Processed Pseudogene Counterparts by an Analysis of Total Genomic DNA

MICHAEL WAGNER† AND ROBERT P. PERRY*

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received 10 June 1985/Accepted 9 September 1985

Two genes from the family encoding mouse ribosomal protein S16 were cloned, sequenced, and analyzed. One gene was found to be a processed pseudogene, i.e., a nonfunctional gene presumably derived from an mRNA intermediate. The other S16 gene contained introns and had exonic sequences identical to those of a cloned S16 cDNA. The expression of this gene was demonstrated by Northern blot analysis of nuclear poly(A)$^+$ RNA with cDNA and unique sequence intron probes. Each S16 intron contains a well-preserved remnant of the TACTAAC motif, which is ubiquitous in yeast introns and known to play a critical role in intron splicing. A sequence comparison with two other mouse ribosomal protein genes analyzed in our laboratory, L30 and L32, revealed common structural features which might be involved in the control and coordination of ribosomal protein gene expression. These include the lack of a canonical TATA box in the −20 to −30 region and a remarkably similar 12-nucleotide pyrimidine sequence (CTTCCYTYYTC) that spans the cap site and is flanked by C+G-rich sequences. The nature of the other members of the S16 family was evaluated by three types of experiment: a DNase I sensitivity analysis to measure the extent of chromatin condensation; an analysis of the thermal stability of cDNA-gene hybrids to estimate the extent of divergence of each gene sequence from that of the expressed gene; and a restriction fragment analysis which distinguishes intron-containing genes from intronless processed genes. The results of these analyses show (i) that all genes except the expressed S16 gene are in a condensed chromatin configuration associated with transcriptional quiescence; (ii) that most of the genes within the S16 family have sequences >7% divergent from the expressed S16 gene; and (iii) that at least 7 of the 10 S16 genes lack introns. We conclude that the ribosomal protein S16 multigene family contains one expressed intron-containing gene and nine inactive pseudogenes, most or all of which are of the processed type.

The eucaryotic ribosome is a complex organelle composed of four RNA species and over 70 different proteins. The biosynthesis of these various components is coordinately regulated during embryonic development and in response to changes in cell growth rate by a variety of mechanisms operating at several levels of gene expression (21, 27, 36, 52, 59, 62). To understand how these mechanisms operate in higher eucaryotes it is important to know how the individual genes are organized and expressed. While there is considerable knowledge of the structure and expression of the rRNA genes of higher eucaryotes, much less is known about the ribosomal protein (rp) genes.

Our earlier investigations revealed that the sequences encoding individual rps are present in mammalian genomes as multigene families containing 10 to 25 members (55). Initially, this great multiplicity of rp genes seemed to present a daunting obstacle to those wishing to understand the genetic basis of how mammalian rp genes are coordinately regulated. However, the results of recent studies have indicated that this problem may not be as formidable as previously anticipated. In these studies, selected members of four mouse rp gene families were cloned, characterized, and found to consist of both intron-containing genes and intronless processed genes, many of which are clearly nonfunctional pseudogenes (19, 38, 57a, 88). Moreover, Northern blot analyses of nuclear RNA with unique sequence intron probes suggested that there might be only a single expressed intron-containing gene in the rp L30 and rp L32 families (19, 88). Yet, inasmuch as only a portion of the family membership was cloned and analyzed in these studies, the issue of whether there are any expressed processed genes or additional intron-containing genes that are not ubiquitously expressed could not be clearly resolved.

In this report, we present a detailed structural analysis of the gene that encodes the small subunit protein S16. A comparison of the rpS16 sequence with that of the expressed L32 and L30 genes has revealed some novel common features which could conceivably be involved in their coordinate regulation. To evaluate the nature of the other members of the S16 family, we devised a strategy, based on analyses of total genomic DNA, which circumvents the necessity of cloning and sequencing the entire gene family. This strategy should prove useful for characterizing other multigene families, in particular, families of housekeeping genes that normally contain a high proportion of processed members (83).

MATERIALS AND METHODS

Materials. Cytoplasmic RNA was isolated from cultured mouse L cells as previously described (61). Nuclear RNA, generously supplied by D. Kelley, was obtained from a mouse plasmacytoma by hot-phenol extraction of citric acid-purified nuclei (67, 69). Poly(A)$^+$ RNA was purified by

* Corresponding author.
† Present address: The Rockefeller University, New York, NY 10021.

deoxythymidylic acid-cellulose chromatography. BALB/c mouse DNA was extracted from the nuclei of cultured MPC 11 plasmacytoma cells by the method of Blin and Stafford (7) as modified by Maniatis et al. (46).

The S16 cDNA probe used in these studies was initially isolated by Meyuhas and Perry (53). For nick translations, a 1.2-kilobase (kb) *Hha*I fragment containing the 342-base-pair (bp) S16 cDNA insert flanked by plasmid (pMB9) DNA was used.

All restriction enzymes and T4 ligase were purchased from New England BioLabs, Inc. (Beverly, Mass.), T4 polynucleotide kinase was purchased from P-L Biochemicals, Inc. (Milwaukee, Wis.), bacterial alkaline phosphatase and DNase I were from Worthington Diagnostics (Freehold, N.J.), and S1 nuclease was from Miles Laboratories, Inc. (Elkhart, Ind.) [γ-^{32}P]ATP (3,000 Ci/mmol) was supplied by Amersham Corp. (Arlington Heights, Ill.), and [α-^{32}P]dCTP (800 Ci/mmol) was supplied by New England Nuclear Corp. (Boston, Mass.).

Isolation and analysis of genomic clones. The rp genomic clones were isolated from a recombinant Charon 4A bacteriophage genomic library constructed by M. Davis from BALB/c sperm DNA partially digested with *Alu*I and *Hae*III (14).

The library was screened with the S16 cDNA probe by standard procedures (4). S16 gene-positive bacteriophage recombinants were replated and rescreened several times to ensure single-plaque purity. Analytical amounts of recombinant Charon 4A bacteriophage DNA were prepared by a minilysate procedure (46); preparative amounts of genomic clone DNA were obtained by the method of Tiemeier et al. (80).

Restriction site mapping, subcloning, and DNA sequencing. An adaptation of the procedure described by Schibler et al. (68) was used to map restriction sites in the inserts of recombinant bacteriophage genomic clones. We used *Cla*I to generate the fragment to be mapped and an *Eco*RI-*Cla*I fragment from the left arm of the Charon 4A vector as the mapping probe. Selected restriction fragments were subcloned into the plasmid vector pBR322 (70, 82) and propagated in *Escherichia coli* HB101. The subclones were mapped with additional restriction enzymes by the method of Smith and Birnstiel (73). DNA sequencing was done by the modified Maxam and Gilbert procedure (46).

S1 nuclease mapping of the 5′ end of S16 genes. The S1 mapping procedure used here is based primarily on the procedure of Favaloro et al. (22). Briefly, a restriction fragment containing the putative 5′ end of the rp S16 gene was end labeled by T4 polynucleotide kinase and gel purified. For hybridization with mRNA, this fragment was either strand separated to provide each single strand of the duplex DNA or used directly in its duplex form. The DNA fragment and 5 to 10 μg of cytoplasmic poly(A)$^+$ RNA were mixed together in 10 μl of 80% formamide–0.4 M NaCl–40 mM PIPES (piperazine-*N*,*N*′-bis[2-ethanesulfonic acid]) buffer (pH 6.4), heated at 85°C for 15 min, and then immediately placed at 52°C for up to 16 h. Thirty volumes of ice-cold S1 digestion buffer (280 mM NaCl, 30 mM sodium acetate [pH 4.4], 4.5 mM zinc acetate, 20 μg of denatured calf thymus DNA per ml) were added, followed by 1 to 10 U of S1 nuclease per μl of reaction volume. Incubation was carried out for 30 min at 15, 20, or 30°C. The reaction was quenched by the addition of 75 μl of 2.5 M ammonium acetate–50 mM EDTA and precipitated with 1 volume of isopropanol at −20°C.

S1-digested samples were electrophoresed in parallel with the chemical cleavage sequencing products of the end-labeled fragment used in the hybridization. The samples were electrophoresed for the appropriate periods of time to ensure nucleotide resolution at the presumptive cap site of the gene.

RNA analysis. RNA was analyzed by size fractionation through 2.2 M formaldehyde denaturing agarose gels (44). Transfer of the RNA to nitrocellulose filters, hybridization to nick-translated probes, and posthybridization washing regimens were carried out by procedures detailed by Thomas (79).

DNA electrophoresis and blotting. DNA, electrophoresed through agarose gels, was transferred to nitrocellulose filters by the method of Southern (75). In those cases in which low-molecular-weight genomic DNA was to be bound to nitrocellulose filters, a procedure with 1 M ammonium acetate–0.02 M NaOH as the transfer buffer was used (72).

DNA blots were hybridized (85) for at least 16 h with DNA probes at concentrations of 50 ng/ml and specific activities of no less than 3×10^7 cpm/μg.

Rehybridization of blots was carried out after melting off the probe used in the initial hybridization. This was achieved by washing the filter at 72°C in 2.5 mM Tris (pH 8.0)–0.1 mM EDTA–0.025% PP_i–0.001% Denhardt solution for 30 to 60 min (79).

DNase I sensitivity experiments. The DNase I sensitivity protocol used in these experiments was carried out essentially as described by Mather and Perry (47), with modifications as detailed below. It consists of four procedures: (i) nuclei preparation, (ii) DNase I digestion of nuclei, (iii) preparation of nuclear DNA, and (iv) restriction digestion and DNA blot analysis.

The nuclei were prepared from MPC 11 cells grown in a 500-ml spinner culture. Log-phase cells were harvested, washed with isotonic buffer, pelleted, and then suspended in 10 ml of cold 0.32 M sucrose–3 mM $MgCl_2$–1 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) (pH 6.8) (SMH). Cold SMH (10 ml) containing 0.1% (wt/vol) Triton X-100 was added to the cell suspension, and the cells were disrupted in a Dounce homogenizer. Nuclei were purified by centrifugation through 0.28 M sucrose ($650 \times g$, 10 min, 4°C). The pelleted nuclei were suspended in 3.2 ml of cold 10 mM Tris (pH 7.4)–10 mM NaCl–3 mM $MgCl_2$ (RSB) and left on ice for 3 min to allow clumped nuclei to settle. The suspended nuclei were removed and used in experimental DNase I digests; the clumped nuclei provided control DNA (i.e., no DNase I treatment or incubation).

For DNase I digestion, the nuclear suspension was divided into seven aliquots which were treated with DNase I at concentrations of 0.10, 0.25, 0.50, 0.62, 0.75, 1.25, and 2.50 μg/ml. After the DNase I was added, the final reaction volume was brought to 400 μl with RSB, and all samples were then incubated for 2 min at 26°C. DNase I digestion was quenched by adding 680 μl of a 0.46 M EDTA–0.7% sodium dodecyl sulfate solution.

DNA was isolated from the DNase I-treated nuclei by the procedure referenced above. Each DNA sample (10 μg) was supplemented with 3 mM $MgCl_2$ and digested with *Pvu*II and *Taq*I according to the directions of the supplier. The digested samples were extracted with phenol-chloroform (1:1), ethanol precipitated, electrophoresed through a 0.8% agarose gel for 14 to 18 h at 35 V, and transferred to nitrocellulose. Blots were hybridized for 16 h with an S16 cDNA probe, washed for 10 min at room temperature in several changes of 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), and then hybridized for an addi-

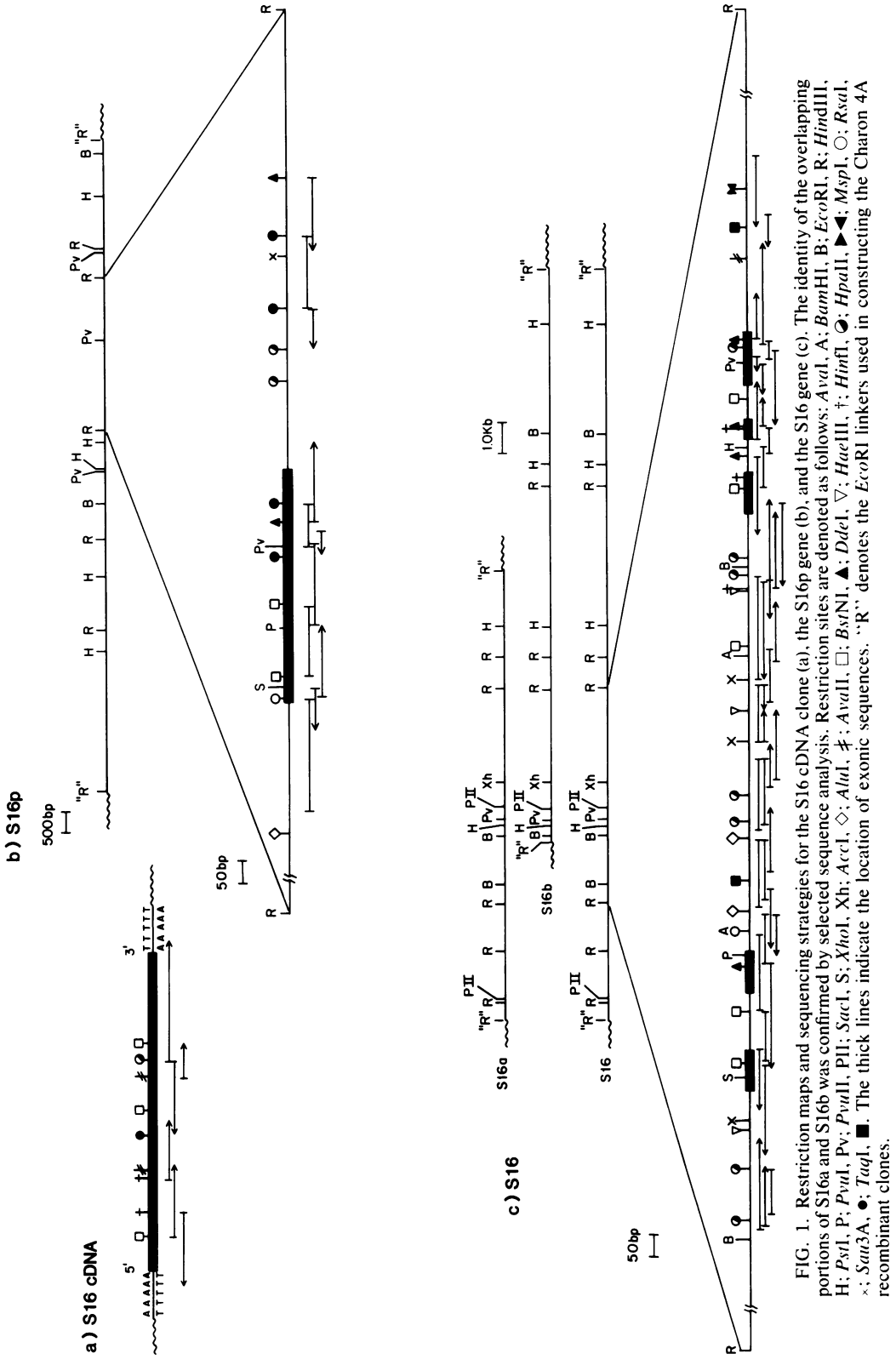


FIG. 1. Restriction maps and sequencing strategies for the S16 cDNA clone (a), the S16p gene (b), and the S16 gene (c). The identity of the overlapping portions of S16a and S16b was confirmed by selected sequence analysis. Restriction sites are denoted as follows: *Ava*I, A; *Bam*H1, B; *Eco*RI, R; *Hin*dIII, H; *Pst*I, P; *Pvu*I, Pv; *Pvu*II, PII; *Sac*I, S; *Xho*I, Xh; *Acc*I, ◇; *Alu*I, ⋕; *Bst*NI, ▲; *Dde*I, ▽; *Hae*III, †; *Hin*fI, ◕; *Hpa*II, ▶◀; *Msp*I, ○; *Rsa*I, ×; *Sau*3A, ●; *Taq*I, ■. The thick lines indicate the location of exonic sequences. "R" denotes the *Eco*RI linkers used in constructing the Charon 4A recombinant clones.

tional 16 h with an alpha-fetoprotein (AFP) probe (the appropriate restriction fragment from a mouse AFP genomic subclone [37] kindly provided by S. Tilghman).

***Sac*I-*Pvu*I fragment analysis.** Genomic DNA (200 μg) was digested with *Pvu*II and electrophoresed on a 0.7% low-melting-point agarose gel at 30 V for 18 h. After electrophoresis, the gel was cut into fractions based on the previously determined mobility of the S16 gene-containing *Pvu*II bands. The DNA from each gel fraction was isolated by adding 0.1 volume of 5 M NaCl, solubilizing the gel at 68°C, and then extracting with 0.5 M NaCl-saturated phenol. Restriction of the DNA samples with *Sac*I and *Pvu*I, gel electrophoresis, and Southern blot analysis were done as described above.

RESULTS

Gene isolation and initial characterization. Although the rp S16 family contains 10 members as indicated by Southern blot analyses (see below), only three distinctive clones (S16a, S16b, and S16p) were obtained when a BALB/c sperm DNA library was screened with an S16 cDNA probe. Restriction enzyme mapping (Fig. 1) indicated that S16a and S16b are overlapping clones of the same gene, hereafter denoted as rp S16. Our failure to pick clones representing the other eight members of the rp S16 family may be due in part to the stringent wash conditions used in our screen protocol (twice for 30 min each time at 65°C in 0.015 M NaCl); such conditions would tend to disfavor the identification of relatively divergent rp S16 pseudogenes (see references 38 and 57a). The *Eco*RI restriction fragments bearing S16 sequences were subcloned into the plasmid vector pBR322, and each subcloned as well as the insert of the S16 cDNA clone was mapped with additional restriction enzymes. The maps were used to establish sequencing strategies, and the complete sequences of the cDNA, S16, and S16p were determined.

S16p is a processed pseudogene. The sequence of S16p (Fig. 2) exhibits several features which suggest that it is a processed gene derived from an mRNA intermediate. These include colinearity with the cDNA sequence and absence of intronic sequences of the expressed S16 gene (see below), similarity in size to the S16 mRNA (53), an A-rich sequence at its 3′ end, and delimitation by a directly repeated sequence (71, 83). A comparison of the S16p sequence with the fused exonic sequences of S16 (which appears to be the only expressed S16 gene, as further evidence will show) demonstrated that insertion, deletion, and substitution of nucleotides have all contributed to the mutation of a presumed original sequence corresponding to the S16 mRNA to a nonfunctional pseudogene. The sequence divergence resulting from these mutations is about 3.7%. Given that these are neutral mutations which accumulate at a rate of approximately 0.7% per 10^6 years (58), the age of the S16p gene is estimated to be 5 million years.

Determination of the 5′ end of the S16 gene. The similarity in size between the S16p gene and the S16 mRNA suggested that the processed gene might be useful in helping to define the 5′ end of the S16 gene. This is especially advantageous because the cDNA used in these studies lacked a considerable amount of 5′-terminal sequence. To determine whether the 5′ sequences of S16p are indeed representative of the S16 mRNA sequence, an S1 nuclease protection experiment was performed. Experimental design and details are presented in Fig. 3A and its legend, respectively. The results show that the S16p (−)-strand sequences downstream of the 5′ direct repeat are protected from S1 nuclease digestion. The sequence of the S16p gene is therefore sufficiently homologous to the S16 mRNA sequence to allow us to use this informa-

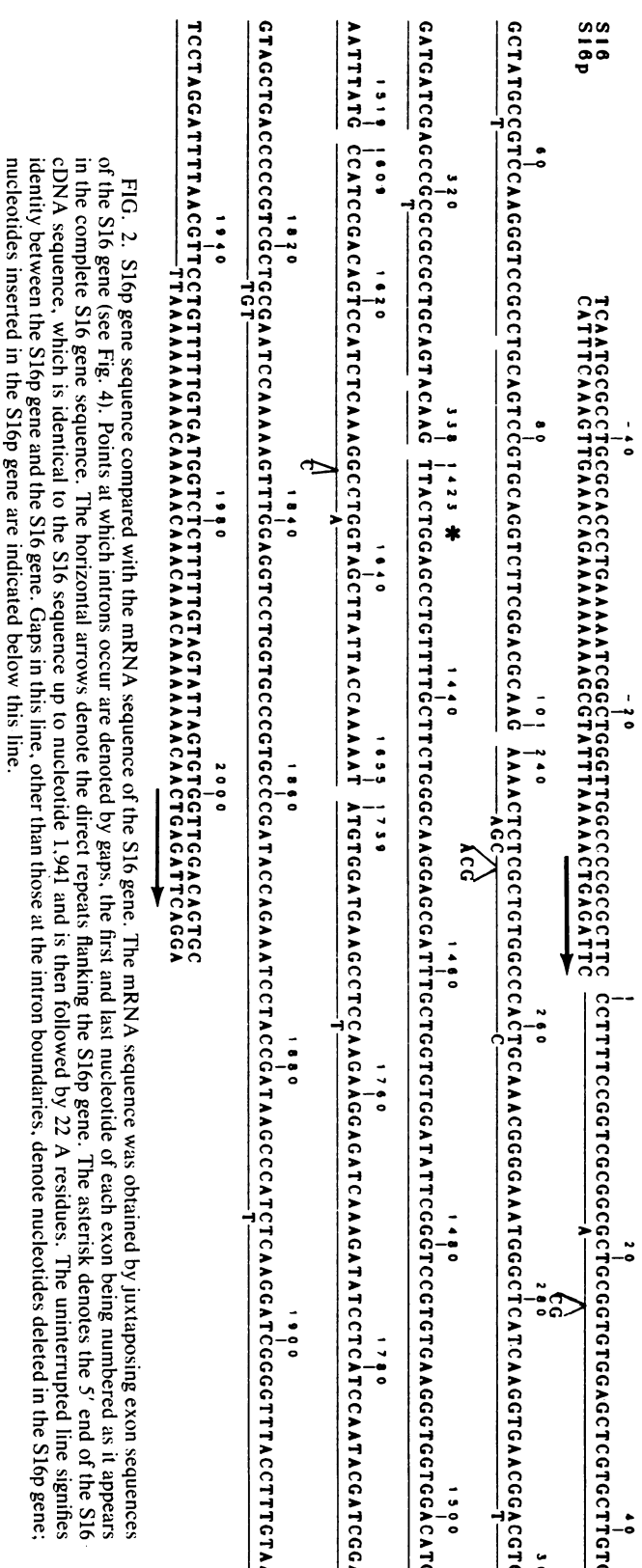


FIG. 2. S16p gene sequence compared with the mRNA sequence of the S16 gene. The mRNA sequence was obtained by juxtaposing exon sequences of the S16 gene (see Fig. 4). Points at which introns occur are denoted by gaps, the first and last nucleotide of each exon being numbered as it appears in the complete S16 gene sequence. The horizontal arrows denote the direct repeats flanking the S16p gene. The asterisk denotes the 5′ end of the S16 cDNA sequence, which is identical to the S16 sequence up to nucleotide 1,941 and is then followed by 22 A residues. The uninterrupted line signifies identity between the S16p gene and the S16 gene. Gaps in this line, other than those at the intron boundaries, denote nucleotides deleted in the S16p gene; nucleotides inserted in the S16p gene are indicated below this line.

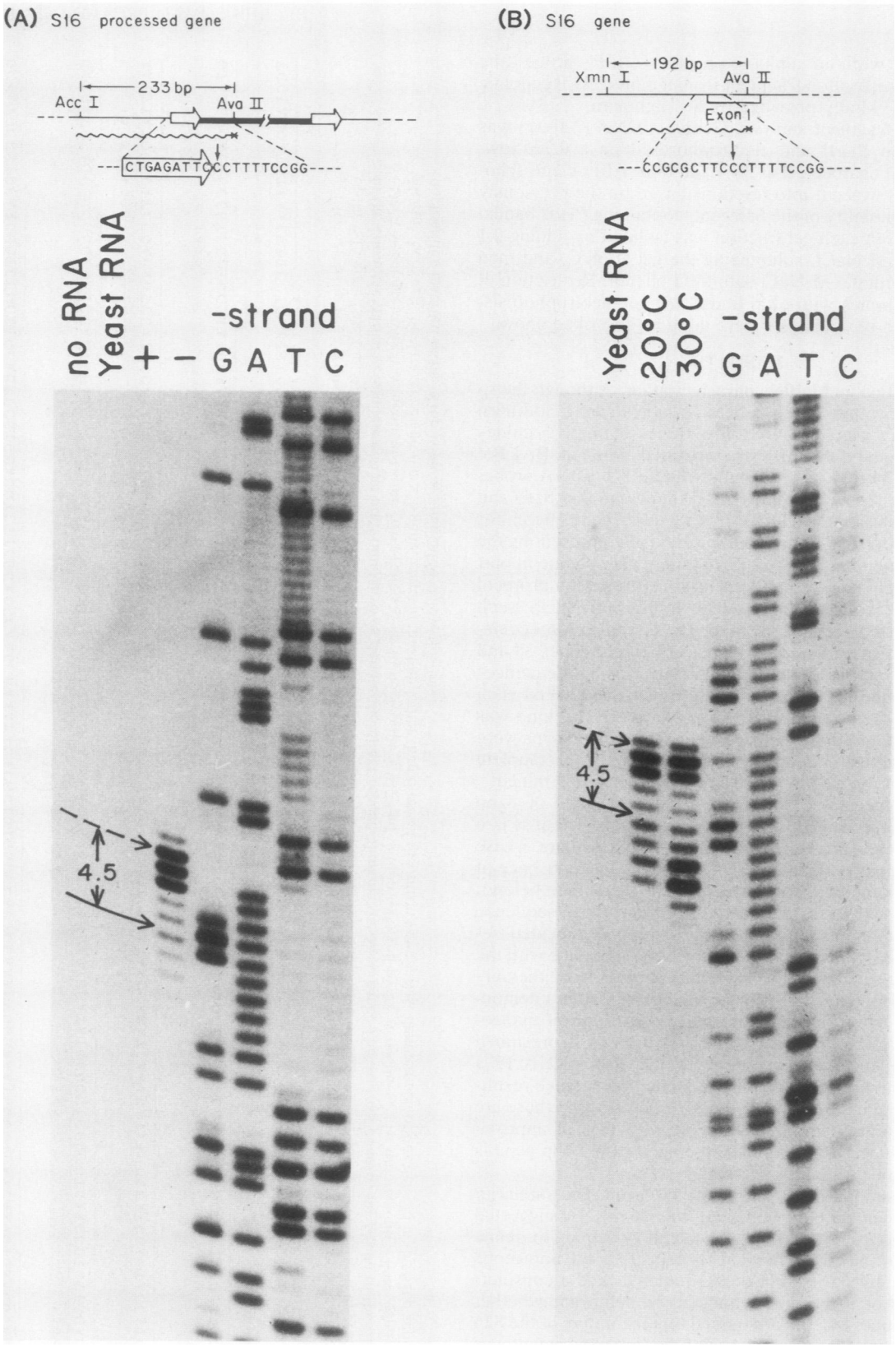


FIG. 3. Determination of the 5′ cap site by S1 nuclease protection. (A) S16p gene. (B) S16 gene. The diagrams at the top illustrate the experimental design of each S1 analysis. The fragments (wavy lines) were single stranded for panel A and double stranded for panel B. The horizontal open arrows in panel A denote the direct repeats of the processed gene. Controls included (i) no RNA input, (ii) yeast RNA input, and (iii) hybridization with the sense (+) strand. In the experiment of panel A, (+) and (−) refer to the sense and antisense strands of the restriction fragment, respectively; only the (−) strand is protected. In the experiment of panel B, two temperatures of S1 nuclease digestion were used: 20 and 30°C. Vertical arrows in the diagrams refer to the cap site nucleotide determined after subtracting a 4.5-nucleotide correction factor as indicated by horizontal dashed and unbroken arrows in the autoradiograms.

tion in identifying the 5′ exon(s) of the S16 gene and in locating its cap site.

This experiment also provides an opportunity to evaluate the precision of the S1 nuclease protection method for determining the location of the cap site. Previous studies of an immunoglobulin mRNA in which the 5′-terminal nucleotide was determined by direct chemical analysis of the cap structure as well as by S1 nuclease protection (35) indicated that the position of the capped nucleotide in an accompanying sequence ladder is 4.5 nucleotides below the position of the largest S1 nuclease-protected fragment. The 4.5-nucleotide difference is presumably due to steric hindrance of the S1 enzyme by the 5′ cap structure and to the fact that chemically sequenced DNA fragments migrate 1.5 nucleotides faster than the corresponding fragments generated by S1 nuclease digestion of RNA-DNA hybrids (74). When this correction was applied to the data of Fig. 3A, the largest S1-protected band of significant intensity corresponded to the first cytosine residue in the sequence 5′-CCTTTTCCGG-3′, which immediately follows the 5′ direct repeat of the gene. The correspondence between the 5′-terminal nucleotide predicted by the location of the direct repeat and that indicated by the S1 analysis tends to substantiate the 4.5-nucleotide subtraction for determining cap sites from S1 nuclease protection experiments with mRNA.

If S16p is indeed a full-length representation of the S16 mRNA, then the identified C residue should correspond to the cap site for the S16 mRNA and, presumably, the transcriptional start site of the gene (60). To establish this point conclusively, a similar S1 nuclease protection experiment was done with a fragment derived from the 5′ region of the S16 gene. Using the sequence information provided by the S16p gene, we examined the sequence of the S16 gene (see Fig. 4) and located the putative cap site within a 192-bp *Xmn*I-*Ava*II restriction fragment. This fragment was used in an S1 nuclease protection experiment (Fig. 3B). Applying the 4.5-nucleotide subtraction to determine the correct cap site, the 5′ end of the S16 gene is seen to begin with the sequence 5′-CCTTTTCCGG-3′, exactly as predicted from the S16p analysis.

Sequence analysis of the S16 gene. The entire sequence of the S16 gene is shown in Fig. 4. The S16 gene is 1,941 nucleotides from the cap site to the nucleotide preceding the poly(A) tail. The identification of exons and introns was accomplished by aligning the S16 cDNA and S16p sequences to appropriate portions of the S16 gene. When this was done, it was seen that the S16 gene contains five exons totalling 545 nucleotides and four introns, all of which are bounded by the requisite GT . . . AG dinucleotide splice junction sequence (9, 57). The usual pyrimidine-rich nucleotide stretch immediately upstream of the 3′ splice site (9) is also present in all introns. The sizes of the introns vary greatly; the three small introns are 82, 88, and 138 nucleotides, while the single larger intron is 1,083 nucleotides.

An open translational reading frame begins with the AUG codon at nucleotide 53 in the first exon and is terminated by a TAA stop codon at nucleotide 1,882 in the last exon. Of the 145 amino acids encoded, 25 are basic and 12 are acidic, imparting an overall basic nature to the S16 rp. The amino acid composition predicted by our sequence analysis of the S16 gene correlated well with the amino acid composition determined by biochemical analysis of the purified protein (11).

Noteworthy features at the 5′ end of the S16 gene include a 12-nucleotide pyrimidine tract spanning the cap site and the absence of a canonical TATA box in the −20 to −30 region. This region, however, contains an AAAAAT sequence flanked by stretches of high (>80%) cytosine-plus-guanine (C+G) content, and as discussed below, this sequence motif might perform the polymerase-positioning function normally attributed to the TATA box (9, 29, 50).

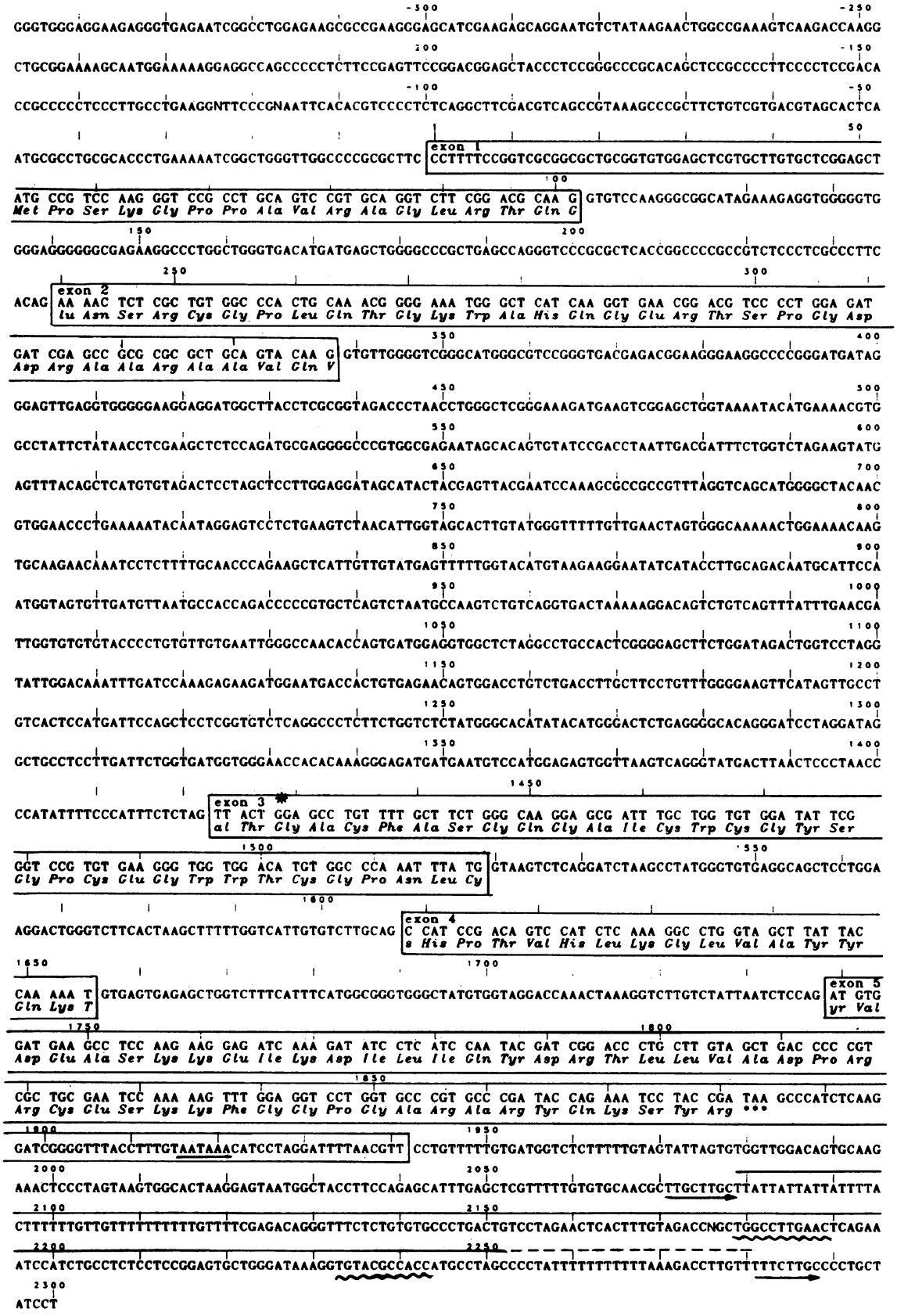
Analysis of the 3′ flanking DNA revealed a moderately repetitive B1 element 150 nucleotides downstream from the end of the S16 gene. This element had 83% homology with the published B1 sequence and, as delimited by a pair of 8-nucleotide direct repeats, was 27 nucleotides longer at the 3′ end than the consensus B1 element (32, 33). It contained a bipartite sequence, 5′-TGGCCTTGAAC--43N--TGTA CGCCACC-3′, starting at nucleotide 2,179, which resembled the consensus RNA polymerase III promoter sequence: 5′-TGGCNNAGTGG--(25–45)N-TGTACGCCACC-3′ (26). However, since the rpS16-associated sequence contains base substitutions at positions deemed essential for RNA polymerase III promoter activity (81) (underlined in consensus sequence), it may not have transcriptional capability.

Expression of rp S16 gene: nuclear poly(A)$^+$ RNA blot analysis. The expression of the rp S16 gene was investigated by Northern blot analysis with a variety of probes (Fig. 5b). Poly(A)$^+$ nuclear RNA from a mouse plasmacytoma was hybridized to the S16 cDNA probe and to probes representing the 5′ flank, intron 2, and the far 3′ flank, all shown to be unique in the mouse genome by Southern blot analysis (Fig. 5a). The intron 2 and cDNA probes both revealed a 2.1-kb primary transcript and a 1.8-kb processing intermediate. In addition, the cDNA probe detected the 0.65-kb mature S16 mRNA.

The correspondence of the components revealed by the cDNA probe and the unique sequence intron 2 probe indicates that the S16 primary transcript as well as its processed intermediates are indeed derived from the rp S16 gene. Nothing was detected with the flanking-region probes, indicating that these regions of the genome do not produce stable poly(A)$^+$ RNAs in these cells. The presence of a 1.8-kb processing intermediate containing intron 2 indicates that this intron is usually excised subsequent to the removal of the three small introns.

Strategy for characterizing the entire rp S16 family. To assess the transcriptional activity and general structural features of all members of the S16 gene family, the following series of experiments were performed: (i) DNase I sensitivity analysis to determine the extent of chromatin condensation and, by inference, the transcriptional status of each gene; (ii) thermal stability of genomic DNA-cDNA hybrids to estimate the degree of sequence homology between each gene and the S16 mRNA; and (iii) a restriction fragment analysis designed to distinguish intron-containing genes from intronless processed genes. These analyses have enabled us to examine the entire S16 family with regard to the attributes normally associated with functional genes without having to individually clone and characterize every gene. Although the interpretation of the results of each of these analyses rests on certain assumptions which cannot be rigorously verified in all cases, the combined results of all three approaches have enabled us to draw strong and definitive conclusions about the nature of the rp S16 family.

The success of such an analysis demands that all of the S16 genes be resolved on a clearly defined set of restriction fragments with each fragment containing a single S16 gene. Moreover, for an easily interpretable DNase I analysis, it is advantageous for all of the genes to reside on relatively small restriction fragments. This requirement is important for two


exon 1
exon 2
exon 3
exon 4
exon 5

reasons. First, large fragments present a relatively large target size to DNase I and will be nicked by the enzyme regardless of whether the fragment contains an actively transcribed gene or not. Second, a large fragment would have an increased probability of containing an active gene in addition to the S16 gene, and if this was the case, DNase I sensitivity of this fragment could not be unequivocally assigned to the S16 gene.

To select the appropriate restriction enzyme(s) for these studies, we digested mouse DNA with a battery of enzymes and subjected the digests to Southern blot analysis with the S16 cDNA probe. Two enzymes, *Pvu*II and *Taq*I, gave a total of 10 well-resolved fragments containing S16 sequences, the largest of the fragments not exceeding 8 kb. For the reasons given above these sets of relatively small fragments are particularly suitable for the DNase I sensitivity experiments. A subsequent restriction fragment analysis, described below, confirmed that each *Pvu*II fragment contains one S16 rp gene, thereby fulfilling the "one gene, one restriction fragment" requirement.

DNase I sensitivity experiments. The relative susceptibility of a gene to DNase I digestion is usually considered to be a reliable indicator of its transcriptional state (87): DNase I sensitivity indicating active gene transcription and DNase I resistance suggesting transcriptional quiescence. The relative susceptibility is assayed by the disappearance of the gene-bearing restriction fragment over a DNase I concentration range in which a restriction fragment bearing a transcriptionally silent gene remains intact. For our experiments the AFP gene (37) was used as the DNase I-resistant standard; this gene is not expressed in the myeloma cells used for these experiments (47), as was verified by a Northern blot analysis of nuclear RNA (data not shown).

A preparation of nuclei from rapidly growing MPC 11 cells was digested with various concentrations of DNase I, and the DNA was extracted from each sample and then digested with *Pvu*II or *Taq*I. DNase I concentrations up to 1.25 μg/ml yielded high-molecular-weight DNA from which a complete size spectrum of restriction fragments could be obtained, as indicated by ethidium bromide staining of the digested samples (Fig. 6). A Southern blot of the *Pvu*II digests analyzed with the S16 cDNA probe and an AFP probe is shown in Fig. 7a. The positions of the 10 S16 fragments (a through j) and the AFP fragment are seen in the flanking marker lanes. A comparison of the intensities of the S16 bands with that of the AFP band clearly indicates that fragment b is selectively digested at the lower DNase I concentrations. This S16 gene is thus likely to be in a chromatin configuration allowing for active transcription. All of the other S16 bands were DNase I insensitive, indicating that the remaining genes within the S16 gene family are in a chromatin structure associated with transcriptional inactivity.

To determine whether restriction fragment b contains the cloned rp S16 gene that was shown to be expressed by

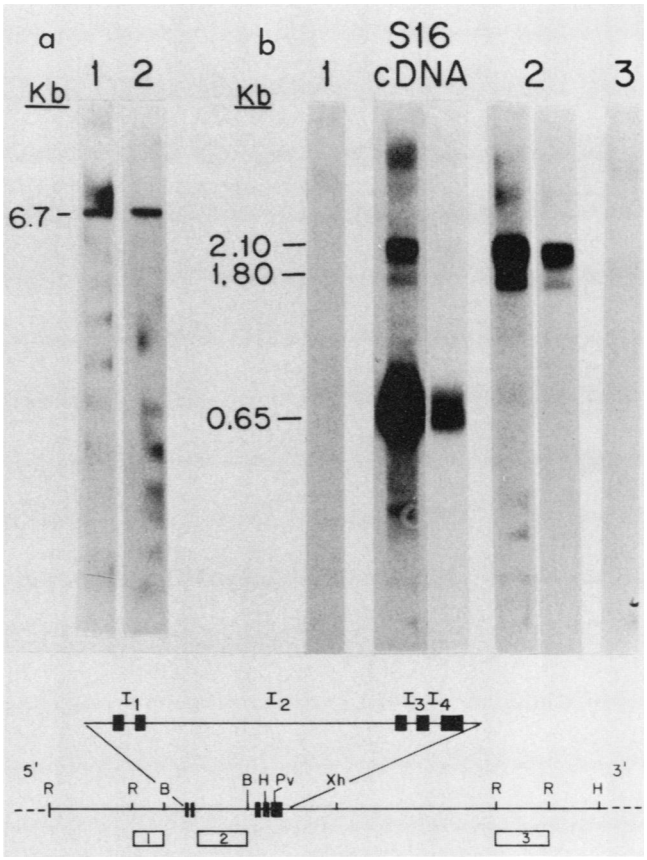


FIG. 5. Expression of the rp S16 gene as determined by Northern blot analysis. (b) Replicate samples (7.5 μg) of poly(A)$^+$ nuclear RNA from a mouse plasmacytoma were electrophoresed in a denaturing 1.5% agarose gel. The RNA was transferred to nitrocellulose and hybridized with the following nick-translated probes: 1, a unique sequence 5′ flanking *Bam*HI-*Eco*RI fragment; 2, a unique *Sma*I-*Bam*HI fragment from intron 2; 3, a unique 3′ flanking *Eco*RI fragment; and the S16 cDNA insert. Two autoradiographic exposures are shown for the blots hybridized with the cDNA probe and probe 2. Panel a shows strips of *Pvu*II-digested mouse genomic DNA hybridized with probes 1 and 2. The single 6.7-kb band demonstrates the uniqueness of these probes. A similar analysis was performed with probe 3. The sizes of the RNA components were determined with rRNA markers that were calibrated by using the exact intron/exon sizes of the sequenced gene. Each component is assumed to have a poly(A) tract of 150 nucleotides. The faint diffuse bands in the 4- to 6-kb region seen with the cDNA and intron 2 probes are not consistently observed with all poly(A)$^+$ nuclear RNA preparations and thus may be artifactual. Restriction endonuclease designations are defined in the legend to Fig. 1.

FIG. 4. Nucleotide sequence of the S16 gene. Exon sequences are boxed and arranged into triplet codons with the corresponding amino acid written beneath each codon. Numbering begins at the cap site. Negative numbers are used for 5′ flanking DNA. The polyadenylation signal AATAAA (underlined) is located 20 nucleotides from the end of the gene (nucleotide 1,941), as established by colinearity with the S16 cDNA sequence. The beginning of the cDNA sequence is denoted by an asterisk. Overlining beginning at nucleotide 2,079 and extending to nucleotide 2,252 indicates 3′ flanking sequences homologous to a B1 element. The horizontal arrows refer to direct repeat sequences associated with the B1 element (a T and G nucleotide mismatch occurs at the third nucleotide of the repeat). The broken overlining indicates that portion of the B1 element (defined here by inclusion between the two direct repeats) as sequences extending beyond the published B1 element consensus sequences (32, 33). Sequences homologous to the split RNA polymerase III promoter consensus sequence (26) are denoted by wavy underlining at nucleotides 2,179 to 2,189 and 2,233 to 2,243.

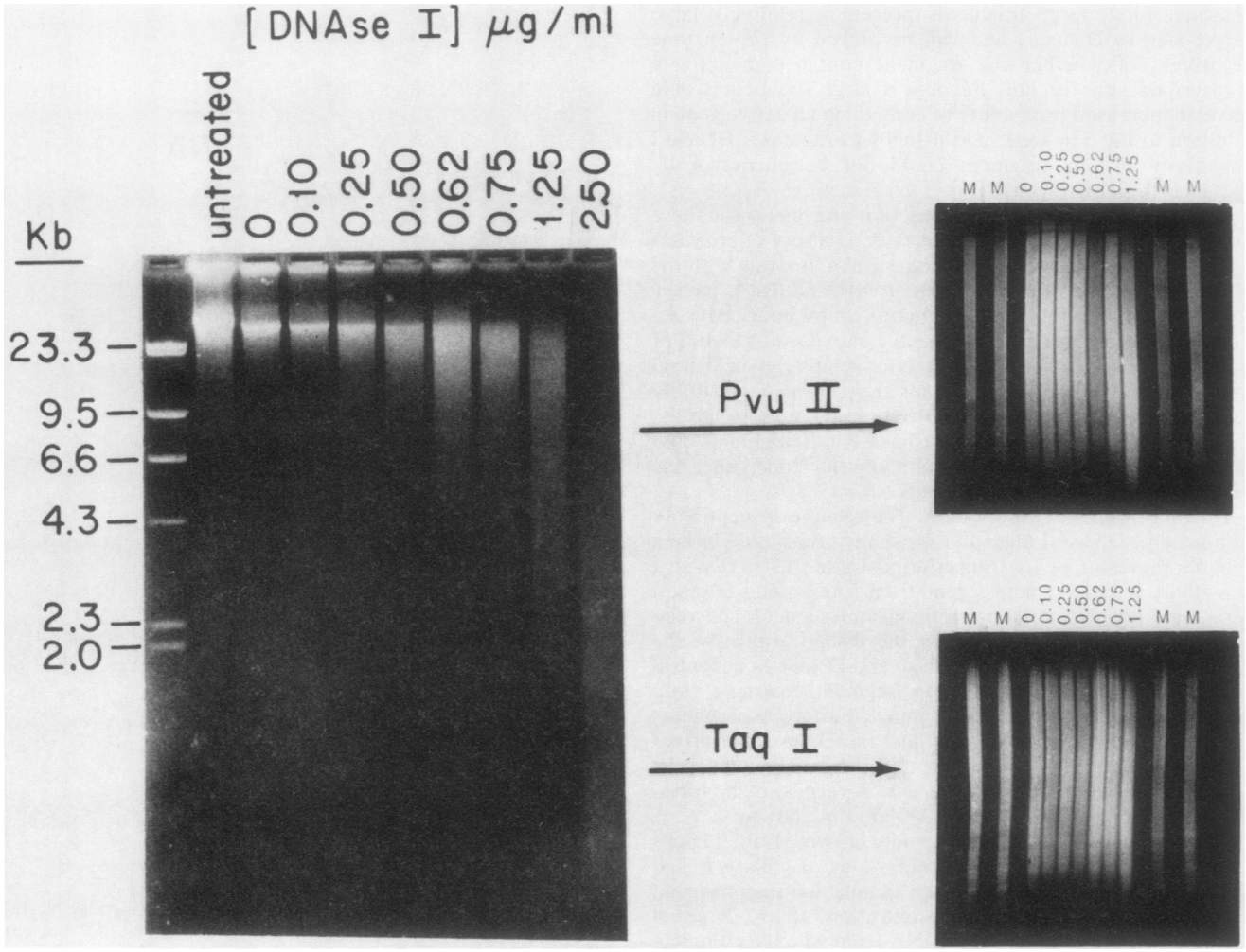


FIG. 6. Analysis by ethidium bromide staining of the DNA used to examine the DNase I sensitivity of rp S16 genes. Left panel: DNA from nuclei treated with various concentrations of DNase I compared with DNA from untreated nuclei. A set of molecular weight markers derived from *Hin*dIII-digested λ phage is shown at left. High-molecular-weight DNA was obtained at all concentrations except 2.5 μg/ml. Right panels: the *Pvu*II- and *Taq*I-digested samples used for the Southern blot analyses of Fig. 7 and 8. The tracks are marked according to the prior DNase I treatment of each sample. The M tracks contain DNA from untreated nuclei which was used to provide size markers for the AFP and S16 cDNA probes. Electrophoresis on 0.8% agarose gels.

Northern blot analysis, a comparison between cloned and genomic DNA *Pvu*II restriction fragments was carried out. Restriction fragment b aligned with the *Pvu*II fragment of the cloned rp S16 gene (Fig. 7b), indicating that it does indeed contain the expressed, intron-containing S16 gene. When similarly analyzed, S16p, the cloned S16 processed pseudogene, was found to correspond to DNase I-insensitive fragment e. Thus, as might be expected, this gene, which presumably lacks upstream transcriptional control signals, registers as transcriptionally silent by this analysis.

The *Pvu*II restriction fragment b is 6.7 kb. Since the S16 gene is 1.9 kb, this leaves 4.8 kb of flanking DNA, most of which is 5′ of the gene (Fig. 1). Such an arrangement is theoretically vulnerable to the situation mentioned above, namely, that a closely linked gene or transcription unit and not the rp gene itself might be imparting the DNase I sensitivity to fragment b. Fortunately, the *Taq*I analysis can be used to examine this possibility. From the sequence analysis, it is known that a *Taq*I digest will place the S16 gene on a 1.7-kb restriction fragment. Most of this fragment is internal to the S16 gene; it begins at a site within intron 2 and extends in the 3′ direction to a site 200 bp downstream from the 3′ end of the gene (Fig. 1). As there are no other genes on this 1.7-kb fragment, any DNase I sensitivity associated with it must be attributed solely to the S16 gene. The results of the *Taq*I analysis (Fig. 8) parallel those of the *Pvu*II analysis in that the band corresponding to the cloned S16 gene, now on a 1.7-kb fragment, was DNase I sensitive while the band containing the cloned S16p gene, band c at 4.0 kb, was not. As one might anticipate, the 1.7-kb *Taq*I fragment, which presents a smaller target for DNase I than does the larger *Pvu*II fragment, disappeared at a higher DNase I concentration (0.75 versus 0.62 μg/ml).

From the results of these experiments, it can be concluded that only one gene within the entire S16 gene family is in a chromatin configuration that suggests active transcription. All other S16 genes reside in a chromatin configuration normally associated with transcriptional inactivity. It seems, therefore, that only a single pair of alleles contributes to S16 mRNA production in the rapidly proliferating cells from which the chromatin was derived; the remaining S16 genes appear to be silent.

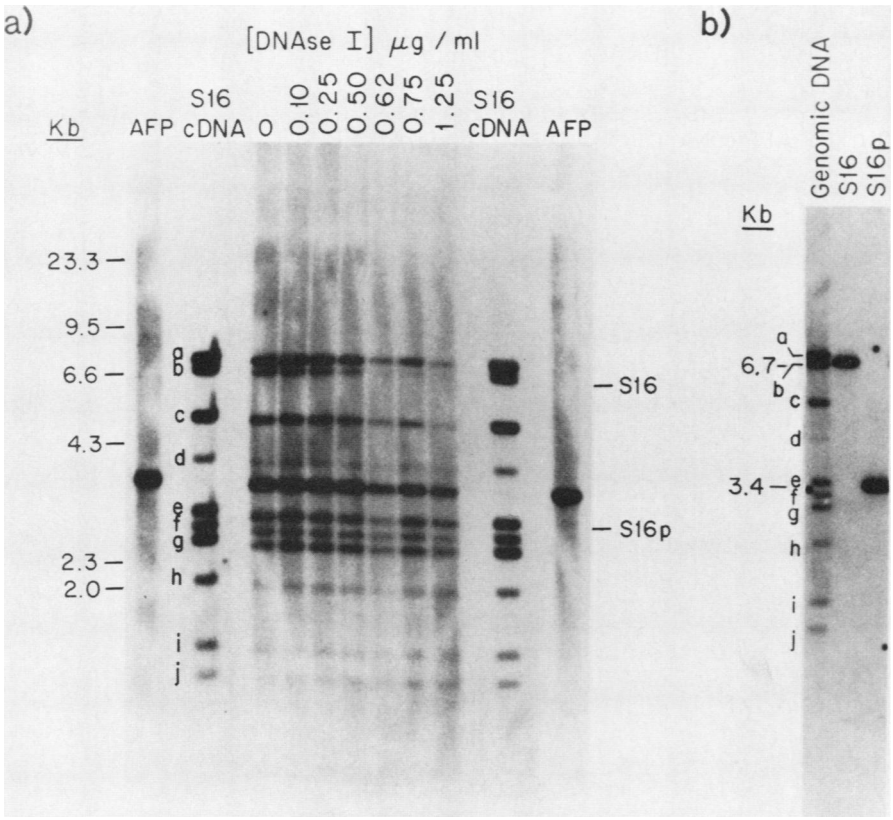


FIG. 7. DNase I sensitivity of *Pvu*II fragments containing rp S16 genes. (a) *Pvu*II fragments derived as described in Fig. 6 were transferred to nitrocellulose and hybridized sequentially (DNase I-treated series) or separately (marker DNA samples) with the S16 cDNA and mouse AFP probes. (b) Comparison of rp S16-containing *Pvu*II fragments in genomic and cloned DNA samples. BALB/c mouse DNA (10 μg) and DNA (4 ng) from clones S16a and S16p was digested with *Pvu*II, electrophoresed in parallel, and hybridized with the S16 cDNA probe. Restriction fragments b (6.7 kb) and e (3.4 kb) correspond to the expressed gene and processed pseudogene, respectively.

Thermal stability of S16 cDNA-S16 gene hybrids. Among the transcriptionally quiescent S16 genes, those exhibiting considerable sequence divergence from the expressed gene would be more likely to be nonfunctional pseudogenes. Based on this reasoning, a thermal stability experiment was performed to determine the degree of sequence divergence between each S16 gene and S16 mRNA sequences represented in the cloned cDNA. Advantage was taken of the fact that the extent of base-pair mismatching between two nonidentical hybridizing DNA strands is proportional to the difference in melting temperature (T_m) between a mismatched and a perfect hybrid (39). From this difference in denaturation temperatures, the approximate sequence divergence between expressed S16 sequences and each S16 gene may be determined.

The results of this experiment (Fig. 9) identify two genes with a melting behavior similar to that of the expressed S16 gene in fragment b. One of these is the S16p gene in fragment e; the other is the gene in fragment a. Within the limited precision allowed by the 5°C increments, there was only a marginally detectable difference in T_m between the perfectly matched hybrids formed with the relatively short S16 exons and the longer mismatched hybrid formed with the S16p gene which has 2.0% sequence divergence over the length of the S16 cDNA (Fig. 2). Although this imposes an uncertainty of about 2% on our ability to estimate the extent of sequence divergence from the decrement in T_m, useful information can be obtained from this experiment because all other genes form cDNA hybrids with significantly lower T_ms compared with those of a, b, and e. Given that a 1°C decrease in the T_m of a DNA duplex corresponds to a 1.5% sequence divergence (39), genes c and g are estimated to have a sequence divergence of about 7 to 15%, while genes d, f, h, i, and j are about 15 to 22% divergent with respect to the expressed S16 gene.

There is a formal possibility that some of the genes exhibiting sequence divergence may be expressed genes that have accrued a substantial number of nucleotide substitutions. However, such genes should be less divergent than most nonfunctional pseudogenes due to strong selective pressure to maintain the appropriate amino acid encodement and a weaker selective pressure against third-base changes in synonymous codons (54). Indeed, a sequence divergence of more than 15% would require about half of the codons to have engendered silent mutations, and there is no precedence for such large-scale neutral divergence among functional genes (54). Moreover, the existence of highly diverged functional rp genes is contradictory to the remarkable evolutionary conservation of mammalian rp mRNA sequences (17, 21, 55). Since most genes in the S16 family exhibit a large sequence divergence from the expressed S16 gene, they very likely represent pseudogenes rather than variant functional genes.

Size comparison of *Sac* I-*Pvu* I fragments. As a third means of characterizing the various members of the S16 gene family, we devised a protocol for distinguishing processed from intron-containing genes by a Southern blot analysis of genomic DNA. The rationale for this protocol (Fig. 10)

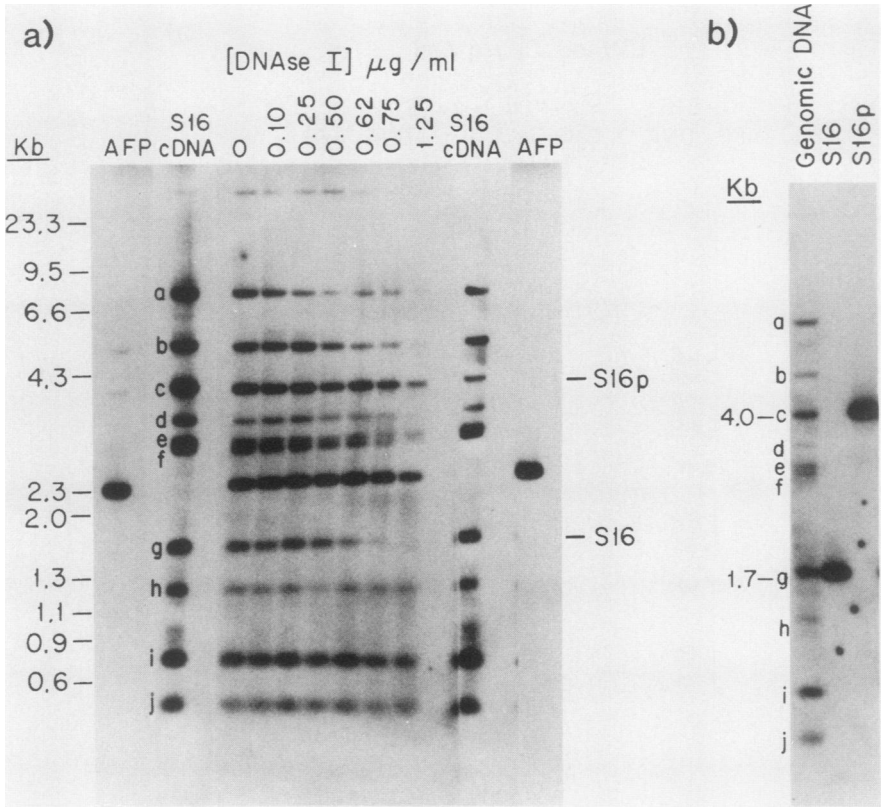


FIG. 8. DNase I sensitivity of *Taq*I fragments containing rp S16 genes. Experimental protocol is the same as that described in the legend to Fig. 7 except that *Taq*I was used instead of *Pvu*II. (a) DNase I-treated series. (b) Identification of fragments containing S16 (g, 1.7 kb) and S16p (c, 4.0 kb) genes.

makes use of two six-base restriction enzyme sites that span all of the introns of the expressed rpS16 gene: a *Sac*I site residing in the first exon and a *Pvu*I site residing in the last exon. A *Sac*I-*Pvu*I digest of the intron-containing rp S16 gene should give a 1.6-kb fragment, whereas the same digest of a processed gene such as S16p should yield a 360-bp fragment. Thus, by determining the size of the *Sac*I-*Pvu*I fragment associated with each S16 gene one may hope to assess whether it is of the processed or intron-containing type.

There are several requirements essential to this protocol. First, genomic DNA must be separated into fractions, each containing a different and complete S16 gene. The results presented earlier suggest that this could in principle be achieved by isolating *Pvu*II or *Taq*I fragments of apropriate size from a digest of mouse DNA. We chose to isolate *Pvu*II fragments for these experiments (Fig. 10B). Second, the genomic DNA should be completely digested at all *Sac*I and *Pvu*I sites to avoid having artifactual (partial digestion) products that would complicate an interpretation of the data. In our experiments, some partial digestion products were in fact encountered, but fortunately the most prevalent of these, which were due to incomplete *Pvu*I digestion, could be identified by coincident hybridization with probes for sequences on the 5′ and 3′ sides of this site (Fig. 10A). Finally, there is an inherent limitation of this analysis due to possible mutation of restriction sites. If either the *Sac*I or *Pvu*I site is mutated, as could be the case in a divergent processed gene, then the size of the fragment would be changed and yet the gene might still be of the processed type. Therefore, genes which display the diagnostic 360-bp *Sac*I-*Pvu*I fragment may be deemed to be processed genes with reasonable certainty, whereas genes which display neither the 360-bp nor the 1.6-kb fragment could be either divergent processed genes or variant intron-containing genes.

The results of this experiment are shown in Fig. 11. The purity of the fractions of isolated *Pvu*II fragments was checked by a Southern blot analysis (panel a). All fractions except those containing fragments f and g gave a single band and thus contained a single S16 gene. The analysis of the *Sac*I-*Pvu*I digests (panel b) revealed 360-bp bands for fractions a, d, f and g, h, i, and j, as well as for fraction e which contains the cloned processed gene. Thus, the S16 gene in each of these fractions is of the processed type. As expected, fraction b, which contains the expressed rp S16 gene, exhibited the 1.6-kb *Sac*I-*Pvu*I fragment.

Most of the DNA fractions including b and e, which contain the cloned, fully characterized S16 genes, displayed high-molecular-weight bands that were not expected according to the experimental rationale. This suggested possible incomplete *Pvu*I or *Sac*I digestion of the DNA samples. The *Pvu*I recognition sequence contains a CpG dinucleotide, which if methylated (5′-CGAT^{m}CG-3′) would inhibit cutting at this site (49) and thus account for the proposed incomplete digestion. This condition was verified by hybridizing the *Sac*I-*Pvu*I blot with probes representing the 5′ and 3′ portions of the S16 cDNA. Most of the high-molecular-weight bands hybridized to both probes, confirming that they are indeed partial *Pvu*I digestion products and of no serious

consequence for the present analysis. The band denoted by an asterisk in fractions a through d appeared to hybridize only with the 5′ probe and thus might represent an analogous *Sac*I partial digestion product.

Some ambiguity remains about the nature of the S16 genes of fraction c and one of the fragments of fraction f and g. Fraction c, which did not yield a 360-bp product, exhibited a 1.1-kb *Sac*I-*Pvu*I fragment. Fraction f and g displayed a 1.4-kb and a 600-bp fragment in addition to the 360-bp fragment typical of the processed gene; conceivably the 360-bp fragment could be derived from either f or g and the two larger fragments from the other S16 gene in this fraction. These larger fragments as well as the 1.1-kb fragment from fraction c could represent either partial digestion products, processed genes with mutated restriction sites, or intron-containing genes. As will be discussed below, a tentative choice among these alternatives can be made when the results of all three assays are considered together.

DISCUSSION

Similarities among mouse rp genes. Our characterization of the expressed rp S16 gene, together with similar analyses of the rp L30 and rp L32 genes (19, 88), provides three mouse rp gene sequences which can be examined for common features that might be implicated in the coordinated expression of these genes. Since these genes are apparently not evolutionarily related to each other, it seems reasonable to anticipate that highly homologous sequences occurring in similar regions could be of functional significance (13).

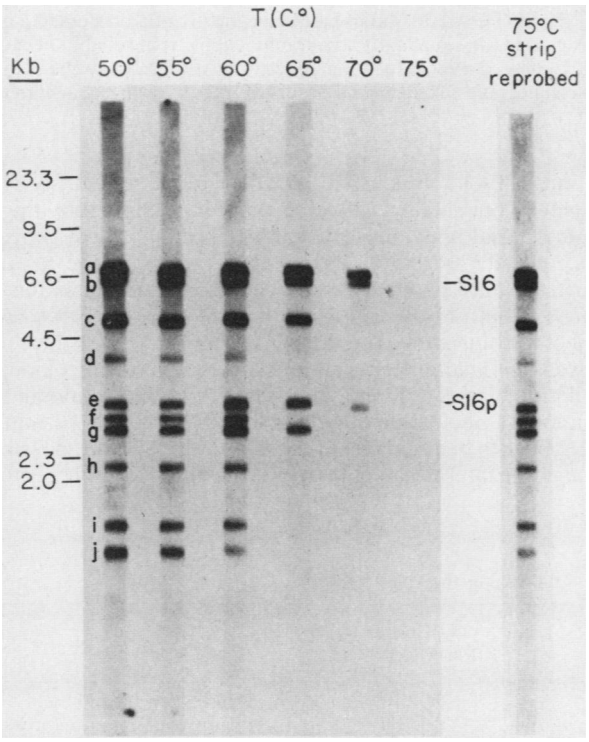


FIG. 9. Thermal stability of S16 gene-cDNA hybrids. Genomic DNA was digested with *Pvu*II, electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose, and hybridized with the S16 cDNA probe. After hybridization, the blot was cut into individual strips which were given three successive 10-min washes in 2× SSC at room temperature. Each of the strips was then washed for 30 min in 0.1× SSC at the indicated temperature. The DNA strips were dried and autoradiographed together. The 75°C strip was later reprobed with the S16 cDNA to show that no significant loss of filter-bound DNA occurs with the higher-temperature treatments.

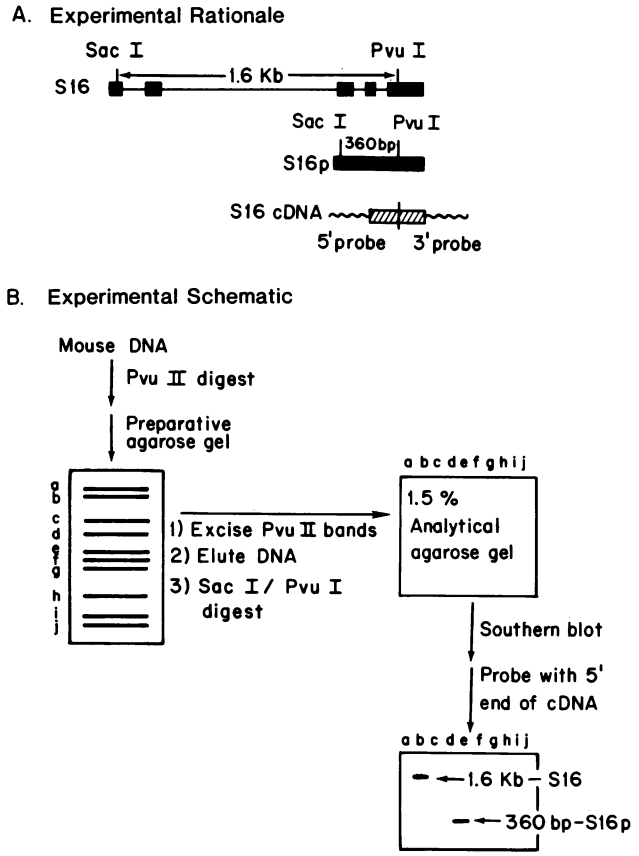


FIG. 10. Rationale and design of an experiment which distinguishes intron-containing S16 genes from intronless processed genes. (A) Diagram of the S16 and S16p genes showing the location of *Sac*I and *Pvu*I restriction sites and the sizes of the predicted restriction fragments. A probe that would specifically recognize the *Sac*I-*Pvu*I fragments (5′ probe) was derived from the S16 cDNA clone by digestion with *Pvu*I and gel purification. This procedure also yielded a 3′-specific probe which was useful for identifying *Pvu*I partial digestion products (see text). (B) A flow diagram of the various parts of this experiment.

The most striking similarity in these rp genes is at their 5′ ends. The cap site of each gene is enmeshed within a ≥12-nucleotide pure pyrimidine tract flanked by blocks of high (average, >80%) C+G content (Fig. 12a). The pyrimidine tracts all have the common motif 5′-CTTCCYTYYTC-3′. The cap sites, and by inference the sites of transcriptional initiation, are the C residues at positions 4 or 5 of this motif. Some moderate dyad symmetry in the C+G-rich blocks is also evident. The similarity in 5′ sequence organization seems especially remarkable because these particular features are not found in the majority of mammalian genes (9).

Another distinctive property of these rp genes is the lack of a canonical TATA box 20 to 30 nucleotides upstream of the transcriptional start site (Fig. 12b). In this region all three genes have a six- to seven-nucleotide box that contains five or six A·T base pairs and is generally flanked by a C+G-rich sequence. Presumably, some aspect of this sequence pattern or the novel sequence organization of the cap region or both elements acting together can perform the function normally

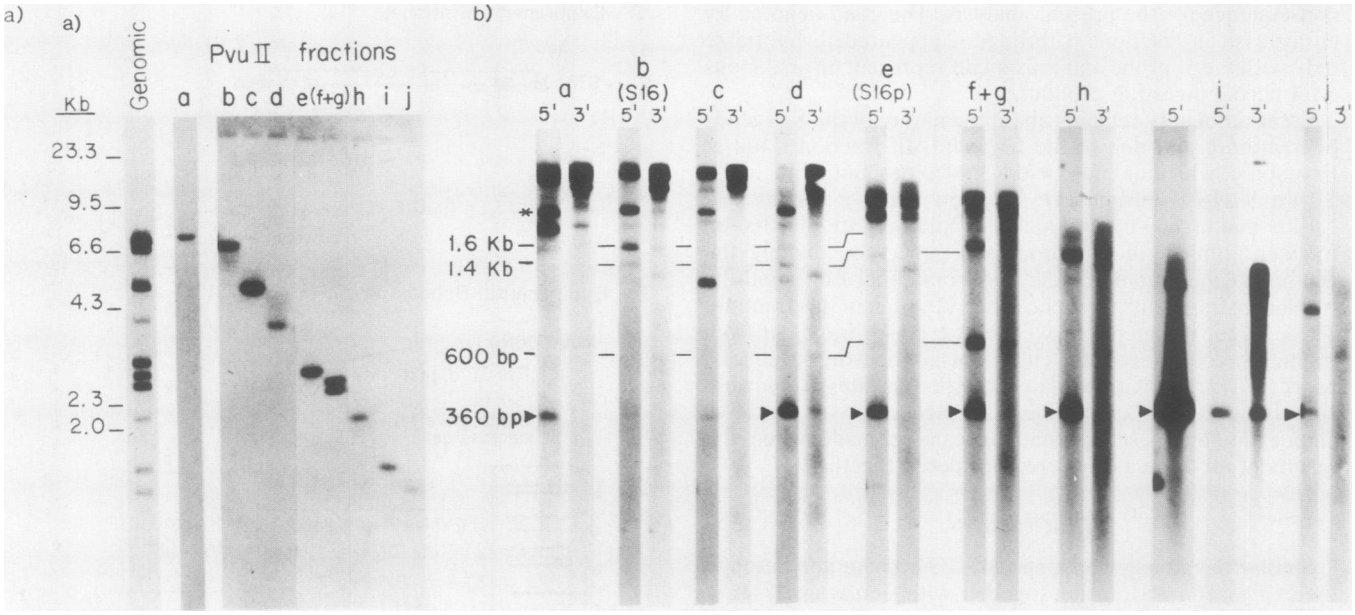


FIG. 11. Comparison of the sizes of the *Sac*I-*Pvu*I fragments derived from individual S16 genes. (a) Purity of size-fractionated *Pvu*II fragments containing S16 genes. Samples of each fraction of *Pvu*II fragments were analyzed for S16 sequences by Southern blot analysis together with a sample of *Pvu*II-digested total genomic DNA. Essentially complete separation of S16 gene-containing *Pvu*II fragments was obtained in all cases except fractions f and g; this fraction contains two S16 genes and is designated f+g. (b) Analysis of a *Sac*I-*Pvu*I digest of the fractions containing individual S16 genes. The DNA of fractions a through j was sequentially digested with *Sac*I and *Pvu*I, blotted onto nitrocellulose, and hybridized with the 5′ cDNA probe (Fig. 10A). After autoradiography the probe was melted off, and the blot was rehybridized with the 3′ cDNA probe. The arrowheads denote the position of the *Sac*I-*Pvu*I 360-bp fragment diagnostic of a processed gene. The i fraction hybridized with the 5′ cDNA probe gave an extremely intense signal at the usual autoradiography exposure, probably because of a large underestimation of the amount of DNA in this fraction; therefore, a reduced exposure is included to illustrate clearly the 360-bp fragment. Failure to completely remove all of the 5′ probe from this fragment can account for the signal observed after hybridization with the 3′ probe.

attributed to the TATA box; namely, the precise positioning of RNA polymerase II at the transcriptional start site (29, 50). Although precise transcriptional initiation in the absence of a canonical TATA box is relatively rare among genes of higher eucaryotes, it is not unique. In addition to the three rp genes described here, it has also been observed for a few viral (2) and cellular (5, 20, 23, 41, 51) genes. Interestingly, most of these cellular genes are also of the housekeeping category and share sequence features with the rp genes, e.g., the pyrimidine motif at or near to the cap site (20, 23, 41, 77) or the neighboring C+G-rich blocks (5, 20, 51, 77). Other housekeeping genes with C+G-rich 5′ regions that lack a canonical TATA box have been described previously (12, 63, 90). Conceivably, novel promoter architecture might confer special regulatory properties on this class of genes.

The possibility that the transcribed portion of the pyrimidine tract may play a role in regulating the translation of rRNAs should also be considered since these mRNAs are subject to translational control (27). The heat shock genes of *Drosophila melanogaster* may be a paradigm for this kind of regulation. Five different heat shock RNAs containing a common 13-nucleotide sequence motif at their 5′ terminus (31) appear to be regulated at the translational level, possibly through the interaction of the common motif with the protein

TABLE 1. Diagnostic features of the genes making up the rp S16 family

Gene designation	Size of *Pvu*II fragment (kb)	DNase I sensitivity	Thermal stability of gene-cDNA hybrids[a]	*Sac*I-*Pvu*I fragment size	Gene type
a	7.4	−	+ + +	360 bp	Processed, pseudo
(S16) b	6.7	+	+ + +	1.6 kb	Intron containing, expressed
c	5.0	−	+ +	1.1 kb	Processed?[b], pseudo
d	4.0	−	+	360 bp	Processed, pseudo
(S16p) e	3.0	−	+ + +	360 bp	Processed, pseudo
f	2.7	−	+	360 bp	Processed, pseudo
g	2.5	−	+ +	600 bp, 1.4 kb	Processed?[b], pseudo
h	2.1	−	+	360 bp	Processed, pseudo
i	1.5	−	+	360 bp	Processed, pseudo
j	1.2	−	+	360 bp	Processed, pseudo

[a] + + + > + + > +.
[b] A gene type cannot unequivocally be assigned for these genes. They appear to be nonexpressed pseudogenes, possibly of the processed type.

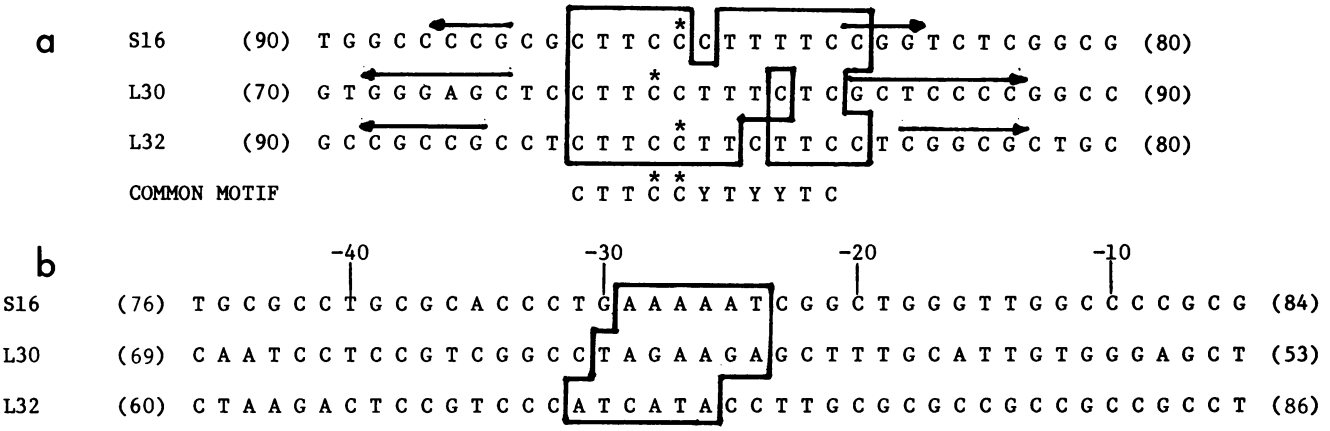


FIG. 12. Comparison of the sequences encompassing the cap site (a) and the −20 to −30 region (b) of rp genes S16, L30, and L32. Cap sites are denoted by asterisks. The common sequence motif of the pyrimidine tract is written below the gene sequences; Y indicates a pyrimidine base. Horizontal arrows indicate sequences with dyad symmetry. Numbers in parentheses are the percentage G · C base pairs in the sequences flanking the pyrimidine tract and the A+T-rich box.

translational apparatus or some heat shock-specific protein (45, 78). Conceivably, the pyrimidine tract may perform an analogous role in regulating the translation of rp mRNA.

Another interesting feature of the rp S16 gene concerns the structure of its introns and the mechanisms proposed for intron splicing. In yeasts, the intronic sequence 5′-TACTAAC-3′, located near the 3′ splice junction, is required for intron splicing (40). This sequence and a close homolog were found in comparable locations within introns of the mouse L32 rp gene (19). Recent analyses of the intermediate products of RNA splicing reactions in yeasts have revealed the existence of "lariat"-type structures with a branch point at the last A residue of the TACTAAC motif (18, 64). When the S16 introns were scrutinized, a recognizable TACTAAC remnant sequence was found at the appropriate distance from the 3′ splice junction of each intron:

CONSENSUS	T A C T A A C
I_1	/GT----101N-------C T C A C ----28N----AG/
I_2	/GT --1053N-------C T A A C ----21N----AG/
I_3	/GT-----56N----A C T A A -------23N----AG/
I_4	/GT-----52N----A C T A A -------21N----AG/

Interestingly, one of the conserved nucleotides in these remnant structures is the branch point A. Splicing of mammalian introns also involves lariat structures with branches at A residues (28, 65). However, conservation of the neighboring nucleotide sequence is, in general, less stringent than that found in the rp S16 and rp L32 remnants. Perhaps the persistence of these sequences in rp genes is a reflection of their long evolutionary history.

Characterization of the rp S16 family. The mouse rp S16 gene family consists of 10 members, 2 of which were cloned and completely sequenced. The nature of the other eight members was examined by a set of analyses that could be carried out with total genomic DNA. Our results, summarized in Table 1 and briefly reviewed below, have enabled us to give an almost complete description of the rp S16 family.

The DNase I sensitivity experiments showed that in rapidly proliferating myeloma cells the expressed rp S16 gene is the sole member of the S16 family with a chromatin configuration characteristic of an actively transcribed gene. All other S16 genes were shown to be in a chromatin structure normally associated with transcriptional quiescence. It seems, therefore, that only a single pair of rp S16 alleles contributes to S16 mRNA production in these cells. This conclusion is consistent with the observation that the intron sequence of S16 pre-mRNA is unique to this DNase I-sensitive S16 gene. These findings support theoretical arguments based on S16 mRNA lifetime, abundance, and translation rates which suggest that the expression of a single S16 gene can adequately account for the observed levels of S16 mRNA and protein (53).

To explore the structural basis for the apparent lack of transcriptional activity of the other S16 genes, two types of experiments were performed: a thermal stability measurement which assessed the extent of sequence divergence from the expressed S16 gene and a restriction fragment analysis which distinguished intron-containing genes from intronless processed genes. The thermal stability experiments demonstrated that most genes in the S16 family have sequences highly divergent from those of the expressed gene: two genes with approximately 0 to 2% divergence, two genes with 7 to 15% divergence, and five genes with 15 to 22% divergence. Extensive sequence divergence may be taken to indicate that these genes are nonexpressed pseudogenes which have evolved in the absence of any selective pressure to maintain their original protein-encoding capacity. The *Sac*I-*Pvu*I restriction fragment analysis demonstrated that at least seven S16 genes are of the processed type. The two genes that could not be unambiguously characterized by this assay are very probably pseudogenes because they are in DNase I-resistant chromatin in rapidly growing cells and have substantial sequence divergence (>7%) from the expressed S16 gene. It seems very unlikely that these divergent genes would be active in a tissue which was not examined for DNase I sensitivity.

Analyses of four other mouse rp gene families (18, 38, 57a, 88) have uncovered many processed pseudogenes but have yet to reveal more than one functional gene per family. These studies, together with the findings presented here, indicate that mouse rp gene families typically consist of one expressed gene and several inactive processed genes. Although this situation may generally apply to all mammals (17, 21, 55), it does not appear to be characteristic of lower eucaryotes or even of lower vertebrates. In *Saccharomyces cervisiae*, rp 51 (1) and presumably other rps (25, 43) are

encoded by more than one functional gene, but no processed genes have been found. In *Xenopus laevis* and other *Xenopus* species, the rp genes are either unique or of low multiplicity and so far no processed genes have been identified (8; F. Amaldi, personal communication). The apparent lack of processed rp genes in nonmammalian species may reflect the absence in germ cells of a reverse transcriptase function or an efficient illegitimate recombination mechanism, both of which are presumably needed for the proliferation of processed genes (83).

Although the predominance of processed pseudogenes within the rp gene families is remarkable, it is not without precedence. The human small nuclear RNA genes form an exceptionally large multigene family in which 500 to 1,000 pseudogenes may be represented (16). Another example is the human β-tubulin gene family. This family is composed of approximately 15 genes, of which two are expressed and the remainder are pseudogenes of either the processed or intron-containing type (41, 42, 89). This latter finding has led to the hypothesis that gene family expansion via mRNA-derived processed genes is likely to be more prevalent for those genes that are expressed in germ line cells (41). Among such genes would be constitutively expressed housekeeping genes, the products of which are required by all cells for their basic structural or metabolic needs. Both the rp genes and the tubulin genes belong to this category. Other examples of families of housekeeping genes that are known to contain processed genes include actin (56), cytochrome *c* (66), metallothionein (34, 84), argininosuccinate synthetase (24), glyceraldehyde-3-phosphate dehydrogenase (3, 30), calmodulin (76), dihydrofolate reductase (10, 48), and the small nuclear RNAs (6, 15, 16).

Since most housekeeping genes are likely to belong to multigene families heavily populated by processed members, it is clearly of considerable advantage to be able to identify all of the expressed genes in these families without the laborious task of cloning and characterizing every family member. The general strategy used here for the rp S16 family should be useful in this regard. Additional analyses such as the examination of DNA methylation may also be helpful in some cases (19), but to obtain unequivocal interpretations, such approaches usually require a fairly detailed restriction map of the gene in question.

ACKNOWLEDGMENTS

We thank Leanne Wiedemann, Kalin Dudov, and Dawn Kelley for advice and help with this project.

This research was supported by grants from the National Science Foundation and the National Institutes of Health and by an appropriation from the Commonwealth of Pennsylvania.

LITERATURE CITED

1. **Abovich, N., and M. Rosbash.** 1984. Two genes for ribosomal protein 51 of *Saccharomyces cerevisiae* complement and contribute to the ribosomes. Mol. Cell. Biol. **4**:1871–1879.
2. **Baker, C. C., and E. B. Ziff.** 1981. Promoters and heterogeneous 5′ termini of the messenger RNAs of adenovirus serotype 2. J. Mol. Biol. **149**:189–221.
3. **Benham, F. J., S. Hodgkinson, and K. E. Davies.** 1984. A glyceraldehyde-3-phosphate dehydrogenase pseudogene on the short arm of the human X chromosome defines a multigene family. EMBO J. **3**:2635–2640.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
5. **Bergsma, D. J., K. S. Chang, and R. J. Schwartz.** 1985. Novel chicken actin gene: third cytoplasmic isoform. Mol. Cell. Biol. **5**:1151–1162.
6. **Bernstein, L. B., S. M. Mount, and A. M. Weiner.** 1983. Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell **32**:461–472.
7. **Blin, N., and D. W. Stafford.** 1976. A general method for isolation of high-molecular-weight DNA from eukaryotes. Nucleic Acids Res. **3**:2303–2308.
8. **Bozzoni, I., A. Tognoni, P. Pierandrei-Amaldi, E. Beccari, M. Buongiorno-Nardelli, and F. Amaldi.** 1982. Isolation and structural analysis of ribosomal protein genes in *Xenopus laevis*. J. Mol. Biol. **161**:353–371.
9. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
10. **Chen, M., T. Shimada, A. D. Moulton, M. Harrison, and A. W. Nienhuis.** 1982. Intronless human dihydrofolate reductase genes are derived from processed RNA molecules. Proc. Natl. Acad. Sci. USA **79**:7435–7439.
11. **Collatz, E., N. Ulbrich, K. Tsurugi, H. N. Lightfoot, W. MacKinlay, A. Lin, and I. G. Wool.** 1977. Isolation of eukaryotic ribosomal proteins. J. Biol. Chem. **252**:9071–9080.
12. **Crouse, G. F., C. C. Simonsen, R. N. McEwan, and R. T. Schimke.** 1982. Structure of amplified normal and variant dihydrofolate reductase genes in mouse sarcoma S180 cells. J. Biol. Chem. **257**:7887–7897.
13. **Davidson, E. H., H. T. Jacobs, and R. J. Britten.** 1983. Very short repeats and coordinate induction of genes. Nature (London) **301**:468–470.
14. **Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. I. Joho, I. L. Weissman, and L. Hood.** 1980. An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature (London) **283**:733–739.
15. **Denison, R. A., S. W. Van Arsdell, L. B. Bernstein, and A. M. Weiner.** 1981. Abundant pseudogenes for small nuclear RNAs are dispersed in the human genome. Proc. Natl. Acad. Sci. USA **78**:810–814.
16. **Denison, R. A., and A. M. Weiner.** 1982. Human U1 RNA pseudogenes may be generated by both DNA- and RNA-mediated mechanisms. J. Mol. Cell. Biol. **2**:815–828.
17. **D'Eustachio, P., O. Meyuhas, F. Ruddle, and R. P. Perry.** 1981. Chromosomal distribution of ribosomal protein genes in the mouse. Cell **24**:307–312.
18. **Domdey, H., B. Apostol, R.-J. Lin, A. Newman, E. Brody, and J. Abelson.** 1984. Lariat structures are *in vivo* intermediates in yeast pre-mRNA splicing. Cell **39**:611–621.
19. **Dudov, K. P., and R. P. Perry.** 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell **37**:457–468.
20. **Dush, M. K., J. M. Sikela, S. A. Khan, J. A. Tischfield, and P. J. Stambrook.** 1985. Nucleotide sequence and organization of the mouse adenine phosphoribosyl transferase gene: presence of a coding region common to animal and bacterial phosphoribosyl transferases that has a variable intron/exon arrangement. Proc. Natl. Acad. Sci. USA **82**:2731–2735.
21. **Faliks, D., and O. Meyuhas.** 1982. Coordinate regulation of ribosomal protein mRNA level in regenerating rat liver. Study with the corresponding mouse cloned cDNA's. Nucleic Acids Res. **10**:789–801.
22. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
23. **Fowlkes, D. M., N. T. Mullis, C. M. Comeau, and G. R. Crabtree.** 1984. Potential basis for regulation of the coordinately expressed fibrinogen genes: homology in the 5′ flanking regions. Proc. Natl. Acad. Sci. USA **81**:2313–2316.
24. **Freytag, S. O., H.-G. O. Bock, A. L. Beaudet, and W. E. O'Brien.** 1984. Molecular structures of human argininosuccinate synthetase pseudogenes. J. Biol. Chem. **259**:3160–3166.
25. **Fried, H. M., N. J. Pearson, C. H. Kim, and J. R. Warner.** 1981.

The genes for fifteen ribosomal proteins of *Saccharomyces cerevisiae*. J. Biol. Chem. **256**:10176–10183.
26. **Galli, G., H. Hofstetter, and M. L. Birnstiel.** 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature (London) **294**:626–631.
27. **Geyer, P. K., O. Meyuhas, R. P. Perry, and L. F. Johnson.** 1982. Regulation of ribosomal protein mRNA content and translation in growth-stimulated mouse fibroblasts. Mol. Cell. Biol. **2**:685–693.
28. **Grabowski, P. J., R. A. Padgett, and P. A. Sharp.** 1984. Messenger RNA splicing *in vitro*: an excised intervening sequence and a potential intermediate. Cell **37**:415–427.
29. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. USA **77**:1432–1436.
30. **Hanauer, A., and J. L. Mandel.** 1984. The glyceraldehyde-3-phosphate dehydrogenase gene family; structure of a human cDNA and of an X chromosome linked pseudogene; amazing complexity of the gene family in mouse. EMBO J. **3**:2627–2633.
31. **Ingolia, T. D., and E. A. Craig.** 1981. Primary sequence of the 5′ flanking regions of the Drosophila heat shock genes in chromosome subdivision 67B. Nucleic Acids Res. **9**:1627–1640.
32. **Jelinek, W. R., and C. W. Schmid.** 1982. Repetitive sequences in eukaryotic DNA and their expression. Annu. Rev. Biochem. **51**:813–844.
33. **Kalb, V. F., S. Glasser, D. King, and J. B. Lingrel.** 1983. A cluster of repetitive elements within a 700 base pair region in the mouse genome. Nucleic Acids Res. **11**:2177–2184.
34. **Karin, M., and R. I. Richards.** 1982. Human metallothionein genes—primary structure of the metallothionein-II gene and a related processed gene. Nature (London) **304**:501–506.
35. **Kelley, D. E., C. Coleclough, and R. P. Perry.** 1982. Functional significance and evolutionary development of the 5′-terminal regions of immunoglobulin variable-region genes. Cell **29**:681–689.
36. **Kief, D. R., and J. R. Warner.** 1981. Coordinate control of syntheses of ribosomal ribonucleic acid and ribosomal proteins during nutritional shift-up in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **1**:1007–1015.
37. **Kioussis, D., F. Eiferman, P. van de Rijn, M. B. Gorin, R. S. Ingram, and S. M. Tilghman.** 1981. The evolution of α-fetoprotein and albumin. J. Biol. Chem. **256**:1960–1967.
38. **Klein, A., and O. Meyuhaus.** 1984. A multigene family of intron lacking and containing genes encoding for mouse ribosomal protein L7. Nucleic Acids Res. **12**:3763–3776.
39. **Laird, C. D., B. L. McConaughy, and B. J. McCarthy.** 1969. Rate of fixation of nucleotide substitutions in evolution. Nature (London) **224**:149–154.
40. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:519–527.
41. **Lee, M. G., S. A. Lewis, C. D. Wilde, and N. J. Cowan.** 1983. Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. Cell **33**:477–487.
42. **Lee, M. G., C. Loomis, and N. J. Cowan.** 1984. Sequence of an expressed human β-tubulin gene containing ten *Alu*I family members. Nucleic Acids Res. **12**:5823–5836.
43. **Leer, R. J., M. M. C. van Raamsdonk-Duin, C. M. T. Molenaar, L. H. Cohen, W. H. Mager, and R. J. Planta.** 1982. The structure of the gene coding for the phosphorylated ribosomal protein S10 in yeast. Nucleic Acids Res. **10**:5869–5878.
44. **Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker.** 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16**:4743–4749.
45. **Lindquist, S.** 1981. Regulation of protein synthesis during heat shock. Nature (London) **293**:311–314.
46. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
47. **Mather, E. L., and R. P. Perry.** 1983. Methylation status and DNase I sensitivity of immunoglobulin genes: changes associated with rearrangement. Proc. Natl. Acad. Sci. USA **80**: 4689–4693.
48. **Maurer, B. J., P. E. Barker, J. N. Masters, F. H. Ruddle, and G. Attardi.** 1984. Human dihydrofolate reductase gene is located in chromosome 5 and is unlinked to the related pseudogenes. Proc. Natl. Acad. Sci. USA **81**:1484–1488.
49. **McClelland, M.** 1983. The effect of site specific methylation on restriction endonuclease cleavage (update). Nucleic Acids Res. **11**:r169–r173.
50. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
51. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression and mutation of the hypoxanthine phosphoribosyl transferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.
52. **Meyuhas, O.** 1984. Ribosomal protein gene expression in proliferating and nonproliferating cells, p. 243–271. *In* Recombinant DNA and cell proliferation. Academic Press, Inc., New York.
53. **Meyuhas, O., and R. P. Perry.** 1980. Construction and identification of cDNA clones for mouse ribosomal proteins: application for the study of r-protein gene expression. Gene **10**:113–129.
54. **Miyata, T., and H. Hayashida.** 1981. Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. Proc. Natl. Acad. Sci. USA **78**:5739–5743.
55. **Monk, R. J., O. Meyuhas, and R. P. Perry.** 1981. Mammals have multiple genes for individual ribosomal proteins. Cell **24**:301–306.
56. **Moos, M., and D. Gallwitz.** 1982. Structure of a human β-actin-related pseudogene which lacks intervening sequences. Nucleic Acids Res. **10**:7843–7849.
57. **Mount, S. M.** 1982. A catalogue of splice sequences. Nucleic Acids Res. **10**:459–470.
57a. **Peled-Yalif, E., I. Cohen-Binder, and O. Meyuhos.** 1984. Isolation and characterization of four mouse ribosomal protein L18 genes that appear to be processed genes. Gene **29**:157–166.
58. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20**:555–566.
59. **Perry, R. P.** 1972. Regulation of ribosome content in eukaryotes. Biochem. Soc. Symp. **37**:105–116.
60. **Perry, R. P.** 1981. RNA processing comes of age. J. Cell Biol. **91**:28s–37s.
61. **Perry, R. P., and D. E. Kelley.** 1968. Messenger RNA-protein complexes and newly synthesized ribosomal subunits: analysis of free-particles and components of polyribosomes. J. Mol. Biol. **35**:37–59.
62. **Pierandrei-Amaldi, P., N. Campioni, E. Beccari, I. Bozzoni, and F. Amaldi.** 1982. Expression of ribosomal-protein genes in *Xenopus laevis* development. Cell **30**:163–171.
63. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
64. **Rodriguez, J. R., C. W. Pikielny, and M. Rosbash.** 1984. *In vivo* characterization of yeast mRNA processing intermediates. Cell **39**:603–610.
65. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. Cell **38**:317–331.
66. **Scarpulla, R. C., and R. Wu.** 1983. Nonallelic members of the cytochrome c multigene family of the rat may arise through different messenger RNAs. Cell **32**:473–482.
67. **Scherrer, K.** 1969. Isolation and sucrose gradient analysis of RNA. *In* K. Habel and N. Salzman (ed.), Fundamental techniques in virology. Academic Press, Inc., New York.
68. **Schibler, U., A. Pittet, R. A. Young, O. Hagenbuchle, M. Tosi, S. Gellman, and P. K. Wellauer.** 1982. The mouse α-amylase multigene family. J. Mol. Biol. **155**:247–266.
69. **Schibler, U., T. Wyler, and O. Hagenbuchle.** 1976. Changes in

size and secondary structure of the ribosomal transcription unit during vertebrate evolution. J. Mol. Biol. **94**:503–517.
70. **Schleif, R. F., and P. C. Wensink.** 1981. Practical methods in molecular biology, p. 134–136. Springer-Verlag, New York.
71. **Sharp, P. A.** 1983. Conversion of RNA to DNA in mammals: *Alu*-like elements and pseudogenes in the human genome. Nature (London) **301**:471–472.
72. **Smith, G. E., and M. D. Summers.** 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. Anal. Biochem. **109**:123–129.
73. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids Res. **3**:2387–2398.
74. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in *X. laevis*. Cell **18**:485–499.
75. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
76. **Stein, J. P., R. P. Munjall, L. Lagrace, E. C. Lai, B. W. O'Malley, and A. R. Means.** 1983. Tissue-specific expression of a chicken calmodulin pseudogene lacking intervening sequences. Proc. Natl. Acad. Sci. USA **80**:6485–6489.
77. **Stone, E. M., K. N. Rothblum, M. C. Alevy, T. M. Kuo, and R. J. Schwartz.** 1985. Complete sequence of the chicken glyceraldehyde-3-phosphate dehydrogenase gene. Proc. Natl. Acad. Sci. USA **82**:1628–1632.
78. **Storti, R. V., M. P. Scott, A. Rich, and M. Pardue.** 1980. Translational control of protein synthesis in response to heat shock in *D. melanogaster* cells. Cell **22**:825–834.
79. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
80. **Tiemeier, D. C., S. M. Tilghman, and P. Leder.** 1977. Purification and cloning of a mouse ribosomal gene fragment in coliphage lambda. Gene **2**:173–191.
81. **Traboni, C., G. Ciliberto, and R. Cortese.** 1982. A novel method for site-directed mutagenesis: its application to an eukaryotic $tRNA^{Pro}$ gene promoter. EMBO J. **1**:415–420.
82. **Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter, and H. M. Goodman.** 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science **196**:1313–1319.
83. **Vanin, E.** 1984. Processed pseudogenes: characteristics and evolution. Biochim. Biophys. Acta **782**:231–241.
84. **Varshney, V., and L. Gedamo.** 1984. Human metallothionein MT-I and MT-II processed genes. Gene **31**:135–145.
85. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76**:3683–3687.
86. **Wallace, J. C., and M. Edmonds.** 1983. Polyadenylated nuclear RNA contains branches. Proc. Natl. Acad. Sci. USA **80**:950–954.
87. **Weintraub, H., and M. Groudine.** 1976. Chromosomal subunits in active genes have an altered conformation. Science **193**:848–856.
88. **Wiedemann, L. M., and R. P. Perry.** 1984. Characterization of the expressed gene and several processed pseudogenes for the mouse ribosomal protein L30 gene family. Mol. Cell. Biol. **4**:2518–2528.
89. **Wilde, C. D., C. E. Crowther, and N. J. Cowan.** 1982. Diverse mechanisms in the generation of human β-tubulin pseudogenes. Science **217**:549–552.
90. **Yang, J. K., J. N. Masters, and G. Attardi.** 1984. Human dihydrofolate reductase gene organization. Extensive conservation of the C+G rich 5′ noncoding sequence and strong intron size divergence from homologous mammalian genes. J. Mol. Biol. **176**:169–187.