The Product of the *H19* Gene May Function as an RNA

CAMILYNN I. BRANNAN, ELIZABETH CLAIRE DEES, ROBERT S. INGRAM, AND SHIRLEY M. TILGHMAN*

Howard Hughes Medical Institute and Department of Biology, Princeton University, Princeton, New Jersey 08544

Received 18 July 1989/Accepted 21 September 1989

The mouse *H19* gene was identified as an abundant hepatic fetal-specific mRNA under the transcriptional control of a *trans*-acting locus termed *raf*. The protein this gene encoded was not apparent from an analysis of its nucleotide sequence, since the mRNA contained multiple translation termination signals in all three reading frames. As a means of assessing which of the 35 small open reading frames might be important to the function of the gene, the human *H19* gene was cloned and sequenced. Comparison of the two homologs revealed no conserved open reading frame. Cellular fractionation showed that *H19* RNA is cytoplasmic but not associated with the translational machinery. Instead, it is located in a particle with a sedimentation coefficient of approximately 28S. Despite the fact that it is transcribed by RNA polymerase II and is spliced and polyadenylated, we suggest that the *H19* RNA is not a classical mRNA. Instead, the product of this unusual gene may be an RNA molecule.

For the majority of cloned eucaryotic genes, their protein products were already known prior to cloning. More recently, differential hybridization screens of cDNA libraries have been used to clone genes with specific tissue and temporal patterns of expression, without regard to their products. The murine *H19* gene is such a gene. It was originally identified in a screen of a fetal liver cDNA library to find genes that were coordinately regulated with the α -fetoprotein (AFP) gene (25). The aim of the screen was to find fetal-specific mRNAs under the control of *raf*, a *trans*-acting locus that determines, at least in part, the adult basal level of AFP mRNA (2, 24). The only cDNA that fulfilled these criteria was the *H19* cDNA. This gene was expressed at high levels in fetal and neonatal liver, and its 600-fold repression in the adult was under the control of *raf*. Its coordinate expression with AFP extended to the visceral endoderm of the yolk sac and the fetal gut. However, unlike AFP, *H19* mRNA was also found in both fetal skeletal and cardiac muscle at high levels and at reduced levels in all adult muscle, although in these tissues it is not under the influence of *raf* (2).

Sequence analysis of the single-copy *H19* gene revealed the presence of multiple small open reading frames (ORFs), none of which spanned more than two of its five exons. The largest of these, called ORF5 (referred to as ORF1 in reference 26), was located entirely within the first exon and could potentially encode a 132-amino-acid protein. This ORF begins 680 bases downstream of the cap site and is preceded by four small ORFs, the largest of which could encode a protein of 27 amino acids. As there was no precedent for such a gene organization, we decided to investigate its possible significance.

As a first step, we cloned and sequenced the *H19* gene from the human genome, reasoning that a comparison between the murine and human homologs would identify regions that are conserved and therefore important to the function of the gene. The demonstration that the genes shared no ORF led us to investigate the subcellular distribution of *H19* mRNA, which indicated that the majority of the mRNA is located in a cytoplasmic particle.

MATERIALS AND METHODS

cDNA library construction. A 20- μ g sample of poly(A)⁺ RNA from Hep3B human hepatoma cells in 90 μ l of distilled H₂O was heat denatured at 65°C for 3 min and then placed at 4°C. To this was added 20 μ g of oligo(dT) (Boehringer Mannheim Biochemicals), 80 U of RNasin (Promega Biotech), 1.5 mM each dGTP, dATP, dCTP, and dTTP, 50 mM Tris hydrochloride (pH 8.3), 10 mM MgCl₂, 70 mM KCl, 10 mM dithiothreitol, and 300 U of reverse transcriptase (Life Sciences, Inc.) to a final reaction volume of 200 μ l; the mixture was incubated at 42°C for 1 h. This preparation was then mixed with 400 μ l of a solution containing 20 mM Tris hydrochloride (pH 7.5), 5 mM MgCl₂, 100 mM KCl, 10 μ g of bovine serum albumin, 4 U of RNase H (Boehringer Mannheim), 400 U of T4 DNA ligase (New England BioLabs, Inc.), and 1 μ Ci of [α -³²P]dCTP and incubated for 60 min at 12°C, 60 min at 22°C, and 10 min at 65°C. Next, 40 U of T4 polymerase (Boehringer Mannheim) was added, and the reaction mixture was incubated at 37°C for 10 min. The double-stranded cDNA was extracted with phenol and chloroform and precipitated in 0.5 volume 7.5 M ammonium acetate and 2.5 volumes of ethanol. The cDNA was subjected to *Eco*RI methylation as specified by New England BioLabs, and then 0.75 μ g of kinase-treated *Eco*RI linkers was ligated. The excess linkers were digested, and the cDNA was size fractionated on a Bio-Gel A-50m column. cDNAs of greater than 500 base pairs (bp) were cloned into λ Zap (Stratagene) to yield 1.2×10^5 recombinant bacteriophage, which were subjected to one round of amplification.

Screening of libraries and sequencing. The cDNA library was screened by the procedure of Benton and Davis (3), using a 720-bp *Eco*RI-*Bam*HI fragment from the 5' end of the murine *H19* gene as a probe. Plaque purification and growth of positively hybridizing phage were performed as previously described (34). The recombinant phage were then excised to yield Bluescript plasmids containing cDNA inserts. The largest clone was used to generate exonuclease III deletions, which were sequenced according to the Sequenase kit (U.S. Biochemical Corp.) protocol.

A human genomic DNA library, obtained from P. Leder, Harvard University, consisted of an *Mbo*I partial digestion of human placental DNA in the vector λ Charon 28. The library was screened as described above, using the human

* Corresponding author.

H19 cDNA as a probe, and positively hybridizing phage were purified. The recombinant phage was mapped with restriction enzymes to orient the gene, and fragments containing the 5' end and spanning the entire gene were subcloned into the single-stranded DNA M13 vectors (21). The nucleotide sequences of the inserts were determined by using the Sequenase kit (U.S. Biochemical Corp.).

Polysome profiles. Procedures for isolation and analysis of polysomes were modified from Schneider et al. (30), Zakeri et al. (38), and Kleene et al. (13). Two hours before the ICR:Ha (icr) mice were sacrificed, they were injected intraperitoneally with cycloheximide (25 µg per adult; 5 to 20 µg per neonate). Twenty minutes before harvesting of tissue culture cells, the cells were treated with 25 µg of cycloheximide per ml. Livers or cells were homogenized at 2 g of tissue per ml in 50 mM Tris hydrochloride (pH 7.5)–50 mM KCl–10 mM MgCl₂–6 mM 2-mercaptoethanol–10 U of RNasin–5 mM dithiothreitol (solution A), using a Teflon pestle. The homogenate was spun to generate a postmitochondrial supernatant. Half of the supernatant (0.5 ml) was layered onto an 11-ml 10 to 50% linear sucrose gradient containing 50 mM Tris hydrochloride (pH 7.5), 50 mM KCl, 10 mM MgCl₂, 10 U of RNasin per ml, and 3 mM dithiothreitol. EDTA was added to 30 mM to the other half of the supernatant, which was then placed on ice for 15 min and subsequently layered onto a gradient containing 30 mM EDTA. The gradients were centrifuged in a Beckman ultracentrifuge with an SW40Ti rotor at 38,000 rpm for 6 h at 4°C. As a control for each run, 250 µg of poly(A)⁺ RNA was loaded onto a parallel gradient and spun simultaneously. Fractions (0.5 ml) were collected from each gradient by dripping from the bottom, and A₂₆₀ nm was measured. For each fraction, RNA was isolated by adding sodium dodecyl sulfate to 1% and NaCl to 0.2 M and extracting with phenol and chloroform. The NaCl concentration was increased to 0.5 M, and RNA was precipitated with 0.6 volumes of isopropanol. The RNA was pelleted by centrifugation, washed with 90% ethanol, dried, and suspended in 7 µl of distilled H₂O. Each sample was denatured in 70% formamide–6% formaldehyde–morpholinepropanesulfonic acid (MOPS) at 65°C for 5 min and electrophoresed in a 6% formaldehyde–1% agarose gel in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA [pH 7.0]). Conditions for transfer to nitrocellulose, hybridization, and washing were as previously described (33). The DNA probe used was either a 710-bp *Sac*I or a 730-bp *Eco*RI–*Bam*HI fragment derived from the murine *H19* gene and labeled by nick translation (20). The extent of hybridization was determined by autoradiography, using Kodak XAR film. Filters were then stripped and reprobbed with a 900-bp *Eco*RI–*Hind*III fragment from the murine albumin cDNA (12).

Nuclear run-on assay. Assays were done as described by Konieczny and Emerson (14), using S2 myoblasts with the following modifications: α-amanitin was added to 5 µg/ml in one reaction, and the filters were hybridized in 4X SET (1× SET is 150 mM NaCl, 50 mM Tris hydrochloride [pH 8.0], 5 mM EDTA), 1× Denhardt solution (0.02% [wt/vol] bovine serum albumin, 0.02% [wt/vol] polyvinylpyrrolidone 40, 0.2% [vol/vol] Ficoll 400), 100 µg of sheared and denatured salmon sperm DNA per ml, 100 µg of yeast tRNA per ml, and 50% formamide at 42°C. The filters were then washed once in 1× SET–0.1% sodium dodecyl sulfate at 25°C for 30 min and twice at 60°C in 0.1X SET–0.1% sodium dodecyl sulfate.

RESULTS

Is the murine ORF5 translated? Our first approach to identifying the function of the murine *H19* gene was to raise polyclonal antibodies to ORF5, the largest of the 35 ORFs of this gene. A fusion protein consisting of the ORF5 peptide fused in frame to β-galactosidase was generated in *Escherichia coli* and used as an immunogen in rabbits. Although the resulting antibodies could recognize ORF5 protein produced in vitro by using SP6-directed in vitro transcription followed by in vitro translation, they consistently failed to detect a protein from either the endogenous gene or transfected genes in a variety of tissue culture cell lines that express the mRNA at high levels (C. I. Brannan, unpublished observation). In addition, the *E. coli lacZ* gene was inserted in frame in the C terminus of ORF5 in situ, and the resulting chimeric gene was transfected into a human hepatoma cell line, Hep3B. Neither β-galactosidase activity nor immunoprecipitable protein could be detected despite expression of the appropriate mRNA. These results suggested that either the ORF5 product is synthesized at undetectable levels or that it is not the polypeptide product of the *H19* gene.

Isolation and sequence of the human *H19* gene. In the absence of any evidence that the largest ORF encoded a protein product, the next approach we adopted was to ask which regions, if any, of the *H19* gene were conserved in evolution. To do this, the human *H19* gene was isolated from two sources: a cDNA library constructed from poly(A)⁺ Hep3B RNA and a genomic DNA library derived from human placental DNA. The cDNA library was screened by using a 5'-distal fragment from the murine *H19* structural gene, and several independent cDNA clones were isolated. Genomic clones were then isolated from a human genomic library.

DNA sequencing of both cDNA and genomic clones revealed that the structure of the human *H19* gene is identical to that of the mouse gene. It is organized into five exons 1,328, 135, 113, 123, and 633 bp in length interrupted by four very small introns 96, 90, 81, and 80 bp in length (Fig. 1). The mouse introns are likewise unusually small, ranging in size from 53 to 82 bp (26).

Comparison of the human and murine *H19* genes. The sequence of the human *H19* gene, aligned with the sequence of the mouse gene by using the BESTFIT algorithm of the Genetics Computer Group, is shown in Fig. 1. Overall, the genes share 77% sequence identity, which is greater than the degree of homology found for the AFP or albumin genes in rodents and humans, two genes that are thought to be rapidly evolving (1, 22, 36). However, the relatedness of the two *H19* genes is not evenly distributed throughout the genes. The most conserved domain begins toward the 3' end of the first exon, at nucleotide (nt) 960, and extends through exon 2. This region is 85% identical in sequence. In contrast, exons 3, 4, and 5 are less conserved, exhibiting 65, 80, and 72% homologies, respectively. The four small introns are only slightly less well conserved (~69%).

The human mRNA sequence contains 21 AUG codons (Fig. 2A). As in the case of the mouse *H19* gene (Fig. 2B), most of these initiators are followed shortly thereafter by terminators. Nine of the AUG codons are shared with the mouse gene; a comparison of the proteins sequences that could be encoded by them is illustrated in Fig. 3A. The most striking feature is the absence of any ORF that is conserved along its entire length between the two species. Instead, the ORFs that begin at common AUG codons terminate trans-

CTGACGGGCCCCAACACCTCACCAGGCGCAAGTGGTGACCGACGCCACGACGGG	-770	AGCATCCAGGGAGTCAAGGGCATCGGGCGAGACGAGCTAGCGGAGCGGGGGGGGG	+1167
GGTGGCTGGGGAGTCCAACTCCGCACTCCCACTCCCAACCGTGCTCCCCAC	-710	- TT-----CA--AT--AG-----G-----A-----A-----	
CGGGGCTGGGAGTCTGTAGGGCGCCACCGCTTGTAGTAGAGTGGCGGGGAGG	-650	AGTGAATGAG CTCTCAGGAGGAGGAT GGTCCAGGAGCGGCTGAGGAGCGGAGCGG	+1225
CGTAAGCACAGCCGGCAACATCGGGTCTTCAGACAGAAAGTGGCGGGAATGGACGG	-590	-A-----TT--AG-----T-G-----T--A- C---T--- TG---T	
GGGTGCCAGCGGCTGTGGGACTTGTCTCTGGGAAACCGGGTGACGAGCAGAACTC	-530	GGGAGC GGAAGGCACTGCGCTCCAGAGCGGCTGGGCAAGCGGGGCTCGGGGGCGGC	+1283
GGTCACTGGATGGGAATCGGCTGGGGGGTGGCAGCGGCCACGAGGGGTTTGGG	-470	--T---CA-----T-----A-----T-----	
CACCTCCCTCTGGCCCTCAGCACCCACCCCTACTCTCAGGAACCTGAGCTCTGAGCG	-410		
TGATGGTGGCAGAAAGGGGCTCTGTGGCATCCGAGTCCCGAGGACCGGAGCTGGCC	-350		
CCAGCCATGTGCAAGTATGTGACGGGGCTGGCAGGACGGAGCAGGATGGTGT	-290		
CCCCTGAGGGAGACAGTGTCTGGGAGGAGAGCTCTCGACCTGACGGAGCTGAT	-232	3' exon1	
T-TA---A---CTT---A---CTAG--G-GT-T-T-T--C-A-A-C-A--GA-AT--CT		GACGGAGCGGGATCGGTGCTC AGCGTTCGGGCTGACAGCAGG gtgagtttttcc	+1340
		-----A-T-----T---GA--A-C---A-----T---g---a---	
GGGCAATGCTCAGCCCTGTCTCGGATGCCAA GGAGGGTGCGGGGAGGGCGTC	-176	cctctgtccaccctcagccccaccgccccccccacacacacacagtttccccac	+1400
G---A-----CAG-----AA---T---GTG---TT---GT-A-C-C-T---		g---t-c-g---gtgt-ac--a-t-t-ct---g---aatta--	
TTTG GACAATTCAGGATGGGTGC TGGTGAGAGAGAGCTGTGCTGGAAGTGC	-124	5' exon 2	
---C---T-----T-----C-----GGGTCTGG--CGGA---C-----GA		acgactctctgtttctccccacag GCAGTCTCCAGCTGGGCTGACGCTGCCACAGC	+1459
TCCAGGCGGAGGTGGGCTGGGGGGCTGGGAGGGCCCTGCTCTGATTGGCGGG	-64	t-atg---tcac---t-t- -----A-A-----T---TG---T	
-A---T-----C-----GC-C-GG-----A--TA-----TT-A-		TGCGAAGCGCAAGACCGGCTCGGTGAGCTGACAGCAGGACATGATGCTCGCG	+1519
AGGGCAGGGCGGAATTCGCGGGGCCACCCCACTTAGAAAAAGCCGGGTAGGACCG	-4	CA-T-----G-G-T-A-----GT---CA-T-----A-A-----	
-A-T-----C-G-N-T-----AG--C		3' exon 2	
cap site		TGTGACGGGAGGACAGCAGGAGCGG TCGGCGCTCTGT gtgagctgtctgtccctcc	+1578
AGG AGCAGGCTGAGGAGC GCGTGGATCGGTCGGCGTAACGGCGGAACTCG	+51	----T-A-----A-G-CA-TCA--A-----T- -----a-	
--C -C-----T-----GGGT--G--TGG-----T-----		ctgctcaggagcgccctgccagaccgccccgccaccatctcactgccccgacctct	+1638
GGAATGCGGAACCGAGGGGCAACGAGGCAAGATCGGCTGCTGAGGAGAGCTTGTGG	+111	act--- --agt---ca-----a-gaa---	
----AT-----A-----C-----		5' exon 3	
AGCAAGGAGCAGCTTCGACATCTGAGTCTGCGAGGAGTATGACGGGTGACGGGCTA	+171	gtttctctacag AACACCTTAGGCTGCTGGGCTGGGCAAGCGGCTCTGTTCTTT	+1697
GT--G-----A-----A-T---A--ACT-----		---c-t-g--- -----A-G-----C-C--TGTCT-----C-----CT--C	
GCTGAGGAGGCGCTGCTGGGCGCTGAGGCAAGATGCTGAGTGTGCGGAGC	+231	ACTTCTCGACGGAGTGGCAGACTATGGCTGCGCTCTGCGGCTCGC AC	+1745
--CTC-----ACGAA-----		---TTC---AA---CTAA---TTC-CT---T-----ATC-T-CTCCCCCTACCTT-	
ATGCTGACAGACGGGAGACATCAG CTGGGAGGATGGGCTGAGCTAGGGTGGG AAG	+289	3' exon 3	
---CAA--G-T--AAA-----G-AG--AA-CT--A-----T---G-G-		AACCCACAAACATGAAAG gtgagggcttctctgccaccttgggggtgggggacgcgag	+1804
AAAGGGAGGCGGCAATTCATCCGGTCACTTTTGGTTACAGGAGCTGGCAGCTGCTGG	+349	----T--G-----ttctc-t---c-tg--t-g-a-a--gtg-gat-	
---T-----A-----T-----G-----C-----		5' exon 4	
AGG AGGGAGCTGCTGGGAGGCTTTCATCCAGGGCTGGGCAAGCGAGCTGTAGCTG	+408	aggagctgagtgaggacctcaactctctccccatccacg AAATGCTGTACCCAGCTCA	+1863
-TAA-----C-----A---C-C-CT--A-----GT-A---A-----		cca-- --ca--gt--- -t--ct-----t--	
GCAGCAGCGGAGCTGAGGACC CCAATCGCGGCGAGCTGAGTCCCTTCG CTCGCC	+465	AGCGTGGGCTTTGAATCGGAGAC AAAACCTCTACCTTGAA ATGAATAT	+1915
---T-----T-----G-G-----T---C-T-TT--T-T		T-T-----GG--TCTTT--GT--G---C-T-CT--TC-A--G--	
AGGCTCGCTTCCCGAGCTTCTGAAGAGAGGAGCTTACGGGATCGAGGCTCGCGGG	+525	3' exon 4	
T-----C-A-TGA-----A-C-----A- AG-G--C---T-A-A		GCTGCACTTACAACTGCACTACCTGACTCAGGAATCGGCTCTGAAAG gtgagcac	+1974
AGAAGCAGACACCTCCAGCAGAGGGGCA CCAATCGGGCAAGAGCTTAGCAAAAGCTG	+584	----A -C-G-----A-----C-----	
-G-CTGG---G-----AT---CAT-GAAA-----T-----A-		cagcgtctcttnggaagctccagggccccggagcaccctgccccatccccacacgtgt	+2034
ACATCTTCTCGGGGAGCGGAGACTGCGCAAGGCTCGGGCTTATCGGCCCTTCCAGC	+644	-t-g-a-a-c--t-g-nmn--gcc-a-	
-----A-AAGG-----ACC--T--GACC--G--		5' exon 5	
CAGAAAGGC AAGAGGCGAGGAGGAG	+666	cgctatctct ag GTGAAGCTAGAGGAACAGCACTCATCAGCCCAATCAAGA	+2089
TG-----TCTTAGAGAG--AA-A-A-GT---GGTGGCAAGAGTGAAGGC		-tt--t---ttgc-- -----GA-A---TGCTG--	
CACAGGGTGGCAGCGTAGGCTCCAG	+700	CACCATCGGAACAGCAGCGCGGCGGAGCAGCCACCGGCGAGCGGAGCTCATCTTCATCG	+2149
CGCTGCTCCAGGAGGTGATAGGACTCTTGAAG---T--CA-AG-C-A-G-A-----		-ATT-T-A-G-----A-T-AT-----A-C-T-A-A-----	
ACGTGGCTGCTACCCAGGCTTGGTGCAGACGGAGATGCAAGGAGCAGAGCAGAGA	+760	CCACCGGCTGGGCTGAGCGTTGACACGAGGACCATCATCCAGAGCTGAGCTCCT	+2209
-A-AACA-AA-C-TT-T---TG-----A-----CA- -AT---T---A---C-		-A-T-TGC-CTGACC---GA-----TG-----CA-----CT--	
G GGGTCCCGAGCTGCCAG CTCACCCAGCGCAATTCAATTAGTAGCAGGACAG	+813	CCAGCGGATGAGCGGCTGCCAGCAGCTCGCTCTTCTTCTTTTTCATCCTT CTGTCT	+2267
-AT-----C-----AACAT-C-----T-----C-----A-A---TT--A		A-CC-----TT-A-AT-T--TC-TG---T---C---C---TC---AA	
GGGAGCTCCGGCAGGCTTCTGAGGCTATGCGGAGGCTGAGGCGCTGAGAGCGG	+872	CTTGTCTTCTGAGCTTCTGCTCTTCTTCTTTTCTGAGAGATTCAAGCGCTGCAG GAC	+2326
-A-TG-----T---AA-- --CAAAG-T---A-----A-TAA-----AT--A		T-C-----TTC---TGGT--C-----GC-----C-----AAG--CGT--	
GAAGACAGGCACTGCTCGGGAGTTCAGCAGGAGCTCACCAGGAGGAGGAGCGGC	+930	TCTGTTTCCCGGT CCGTCTGAATTTAATTGCACTAAGTCAATTGCACTGG TT	+2381
A---AG-----T-----A-----GA-G-G---TG		-----AT-TACC-----G-G-----GA-----	
ACCGAGCGGGGCGCGGAGATTGCGGAGCAAGGAGCTGCAG GGGTGGGCTCGGC	+989	GGAGTGTGAGAGCGGCTTGA GTCTAGTACAGTGT GGTGAGTGTGAGCC	+2434
CA--T-T-TC--ACT--A-A-GG---TTA-----T---TC-A-----A-		---CCC---TA-T-----CT---C---T--A--ATACA--A-T---A-A--	
GGCGGCTCCAGGAGGAGTGGGCGGAGGCTCTGCTGGGAGAGGGCCACAGTGC	+1047	ACCTTGGCAAGTGGCTGTGAGGGGCGGGCGGCTGCATCTGGGG GGTGACTGG	+2491
ACA-A-A-T-----T-----A-----		T-T-----TGCGG---TA--AG-T-T-G-----CC-T-TGA--CA-----	
ACTTGTGAGCGTATGCGCTCAGCGCTCAGCGCTGGGCTGGCTTGGCAGAGTAC	+1107	GGCGGCTGTGCTGGGAGGCGCTCAGCGTGGCTGGCTGCTGGAAGCTGAGCGGA	+2551
-----A-----A-----T-----T-----A-A-----A-		-T-T---C-----TT-----G-T-C-----AGTTCC	
		CATCAGGAGGAGGCTTCAAGCATTCATTAGCGGCACTGCTGCTGCGGCTGCCCA	+2611
		-----A---TG-- --AC-T---T-CT--ATC--T-----T--	
		polyadenylation	
		CGAGGCTTACAGCAGGCGGCTGAGTATCATC AATAAAGCTGTTAC	+2663
		-----C---C---AG-T---T-----T-A-----	

FIG. 1. Sequence of the human *H19* gene. The human *H19* gene sequence from bp -860 to +2381 is shown on the upper line, with the cap site and the polyadenylation signal indicated. Exon sequences are in capital letters; introns are in lowercase letters. For comparison, murine *H19* sequence (26) is shown below the human sequence; a dash indicates identity, and only base changes are shown. Gaps in either sequence reflect insertions or deletions. The sequence was determined on both strands of both cDNA and genomic DNA clones, using dideoxynucleotide-chain termination procedures (29). The mouse *H19* gene sequence has been edited on the basis of corrections of the sequence in Pachnis et al. (26) by Poirier et al. (in preparation) at positions 639 (+C), 853 (T to C), 1081 (+A), 1248 (+GC), 1279 (A to T), 1584 (+GGC), 1779 (C to G), 1790 (G to C), and 1884 (+G).

FIG. 2. Positions of translation initiation and termination codons in the human (A) and mouse (B) *H19* genes. Boxes represent positions of the five exons in the *H19* RNA. Below, the relative positions of all initiation (AUG) codons (○) and UGA, UAA, and UAG termination codons (▲) in the three possible reading frames of the RNAs are depicted.

lation at different locations, and in the one instance in which the terminator codon is conserved (ORF15), the two proteins would initiate at different AUG codons. The largest of the ORFs using a common AUG, ORF13, could encode a 97-amino-acid protein in the human gene and 56-amino-acid protein in the mouse gene. However, these two proteins bear little resemblance to one another except for the amino terminus, where 9 of 11 amino acids are identical.

The longest ORF in the human gene is ORF6, located in reading frame 2, which begins at residue 629 in the first exon and terminates in the second exon (Fig. 2). It would encode a 256-amino-acid protein. However, neither its AUG nor its termination codon is conserved in the mouse gene (Fig. 3B). The mouse ORF5, the largest ORF in the murine gene, is contained entirely within human ORF6 and uses initiator and terminator codons that have no complement in the human gene. The lack of relatedness of the two hypothetical proteins is the consequence of multiple insertions and deletions as well as base substitutions in the region. This results in multiple frameshifts, which are represented in Fig. 3 by shifts between upper- and lowercase letters in the mouse sequence. The most notable feature of these two proteins is a conserved domain toward the carboxy terminus of the mouse ORF, where a stretch of 24 of 31 residues are identical. This corresponds to the most conserved region of nucleotide sequence in exon 1.

The *H19* gene is conserved in chicken. The degree of conservation between the murine and human *H19* genes (Fig. 1) ruled out the possibility that the absence of a long ORF in the mouse gene was the consequence of its becoming a pseudogene gene before mammalian speciation. The comparison did suggest that there are regions of the gene that are more highly conserved than others and might represent

functionally important domains. Additional support for this premise was obtained by examining the conservation of the gene in more distantly related organisms. Two hybridization probes were prepared from the human *H19* gene, one that spanned the least conserved region of exon 1, from nt 550 to 792, and another that encompassed the most conserved domain, from nt 891 to 1571, of the cDNA. The *H19* cognate could be detected in human, monkey, mouse, rat, and chicken DNAs with the conserved probe (Fig. 4B) but only in human and monkey DNAs with the nonconserved segment (Fig. 4A). The analogous nonconserved mouse probe detected the *H19* gene cognates only in mouse and rat DNAs (data not shown). These results, which were confirmed with other restriction digests not shown, demonstrate that the *H19* gene is conserved in both mammalian and avian species and suggest that the comparison of the human and mouse genes has indeed identified the most conserved segments of the genes.

***H19* mRNA is not associated with ribosomes.** The absence of a large ORF conserved between mouse and human *H19* genes led us to question whether *H19* encodes a protein at all. If it functions as an mRNA, it should be associated with the protein synthesis machinery in the cytoplasm. Fractionation of neonatal liver and S2 myoblasts indicated that the majority of the RNA was indeed located in the cytoplasmic fraction (data not shown). The cytoplasm was further fractionated in the presence of nonionic detergents on 10 to 50% sucrose gradients to examine the distribution of polyribosomes. Northern (RNA) blotting of the gradient fractions revealed that *H19* mRNA was present in the nonpolysomal region at the top of the gradient. Albumin mRNA, which served as a control for mRNAs undergoing active transla-

Exon		AUG	BP	Protein Sequence	
A	1	9	1063	MPSPSPWGLGRQYSIQSGQHGRPD*	
				--*	
	1	11	1173	MSSQEGGWCROG*	
				--P*	
	2-5	13	1506	MTWSGVTAETEEARPAFLNTLGMWCGKKRVCFFTSSTESAHYGCPLGSQ	NQHERSRNGATQLKPGPLNPDTPKSSLENNMLHFTTTTALPDSGIGSGR*
B				-----ME---GQsssl-ehh-l apcrrsr1--hf-qrantallls--sspyle---	--Qseplhylpqesapr*
	2-3	14	1511	MTGVVNVTERHD-----RAVIQPS*	
				MVRCDGEDRGASGLPEHLRLVGLRQEA GLFLYFLHGVGGT L WLP SGLPEPTT*	
	3-4	15	1723	MAALWAPRTHNKMWLPSSSLGL*	
				-----C-----	
	5	20	2146	MATPCGGRLLTTSHHIIPELSSSSGMPSPPPSPSSFSFCLFVSELSCLSFLLRDSKPPRLCFFPRPF*	
				--NSALT-R--T ---L-P-LYP---S-S-scifllppfl*	
	1-2	6	629	MGPFOAERARGQGGSTGVASVSGSTWGGTGLGQTGTWQGTQDRGVPSCH	LTHRNFSRRHRSSSGTAFSGLCRSLEGNRAGRQAVLGE LQODVTRRAKRP REGGP
				m---knrsil-w-----qdgv-rh-w---R-nipptvihleegs-valak -QV-Q--NN--Mekks agchqksggac-YRT-	
				GTLSRKEAA GARPAGAPTRHCGPSGAERAHSGLDGDAVCPHRSAPGAGLADSTASRGVKGMRDQTRCGRGVNELSGRMVQAGVRSAAAGGEREALASRARGOGGP	
				EGFY-----V-PACRHTI-M-----clyaltaqslgla*	
				RGRRRSRDCLSVRAGDEARSPAGVDVPTSCRRPRRQVRWT*	

FIG. 3. Comparison of ORFs in the human and mouse *H19* genes. (A) Comparison of human and mouse *H19* ORFs that have common AUG codons. The exon in which the ORF is encoded, its number, determined by ordering AUG codons from a 5'-to-3' direction in the human sequence (see Fig. 2), and the base pair (BP) at which each human ORF begins are indicated. The sequence on the top line is the human amino acid sequence. The mouse sequence is aligned below. Dashes represent identities; capital letters represent nonidentical amino acids that are in frame with respect to the human sequence and lowercase letters represent those that are out of frame; stars indicate termination codons. (B) Comparison of the largest human ORF, ORF6, with the corresponding region in the murine *H19* gene (ORF5).

FIG. 4. Conservation of the *H19* gene in evolution. Genomic DNAs (20 μ g each) from human, rhesus monkey, BALB/CJ mouse, Sprague-Dawley rat, chicken, *Xenopus laevis*, *D. melanogaster*, and *S. cerevisiae* were digested with *Eco*RI. Each digest was divided in two and loaded onto two separate 1% agarose gels. Each DNA was electrophoretically separated, transferred to nitrocellulose, and hybridized to probe A (A), derived from nt 550 to 792, or probe B (B), which encompasses nt 891 to 1571 of the cDNA, as shown in the diagram. Both blots were washed for 3 h at 52°C in 0.1% sodium dodecyl sulfate–0.5X SSC (SSC is 0.15 M NaCl plus 0.015 M sodium citrate).

tion, was associated with the polysomal region (data not shown).

Next, the nonpolysomal section of the gradient was analyzed in more detail by spinning the gradients longer in the absence of a cushion. Figure 5A shows the analysis of gradient fractions derived from livers of 5-day-old mice, an age when *H19* mRNA is at its highest level (25). *H19* mRNA was localized in fractions 8 and 9, cosedimenting with a 28S rRNA marker run in a parallel gradient. The presence of EDTA, which serves to disrupt polyribosomes, did not appreciably affect the migration of *H19* mRNA. In contrast, the majority of albumin mRNA pelleted with the heavy polysomes, with a minority sedimenting with small polysomes and monosomes located in fractions 1 and 2. As expected, EDTA treatment shifted the sedimentation of albumin mRNA to the nonpolysomal region.

Figure 5B shows the identical experiment performed with extracts from C3H10T1/2 cells, a cell line that can differentiate into myoblasts, chondrocytes, and adipocytes. These cells have previously been shown to express *H19* mRNA upon growth arrest (26). Once again, *H19* mRNA was found in the identical region of the gradient, indicating that there is no difference in intracellular localization of this RNA in endoderm- and mesoderm-derived cells.

Total cellular RNA from S2 myoblasts was run in parallel to compare the sedimentation of the *H19* RNA in the cytoplasmic extract to purified RNA. Naked *H19* RNA ran more slowly than the cytoplasmic form (Fig. 5C), consistent with the interpretation that the *H19* RNA is complexed with proteins in the cytoplasm.

FIG. 5. Sedimentation analysis of cytoplasmic RNA from day 5 neonatal mouse liver and C3H10T1/2 cells. Postnuclear fractions were sedimented through 10 to 50% sucrose gradients, and fractions were collected. RNA was extracted and analyzed by gel electrophoresis and blotting onto nitrocellulose. The filters were hybridized to a segment of *H19* cDNA (26) and then stripped and reprobbed with a segment of albumin cDNA (12), as indicated. Fraction 1 represents the bottom of the gradient, and fraction 18 represents the top. M is a marker lane with 1 μ g of newborn mouse liver poly(A)⁺ RNA. (A) Analysis of liver extracts from day 5 mice in the absence or presence of EDTA; (B) analysis of extracts from confluent C3H10T1/2 cells; (C) analysis of purified total cellular RNA.

The sedimentation properties observed for *H19* RNA in liver at 5 days after birth was reproduced in the livers of mice at other ages as well. Figure 6 shows a sedimentation profile of hepatic polysomes corresponding to embryonic day 17 and postnatal days 2, 5, 10, and 15. At each age assayed, the *H19* particle cosedimented with the 28S marker (asterisks in Fig. 6), whereas albumin mRNA pelleted or was located in the heaviest fractions (data not shown).

Extracts from the human hepatoma cell line, Hep3B, were also analyzed to ascertain whether the mRNA behaved in a similar manner in human cells. An identical sedimentation pattern was observed (data not shown). Thus, the lack of association of *H19* mRNA with ribosomes is not unique to the mouse.

The *H19* Gene is transcribed by polymerase II. *H19* mRNA exhibits properties characteristic of classical mRNAs such as polyadenylation and RNA splicing. However, the absence

FIG. 6. Sedimentation analysis of cytoplasmic *H19* RNA at different stages of mouse development. Postnuclear extracts were analyzed by sucrose gradient centrifugation and RNA analysis as described in the legend to Fig. 5. Livers were isolated from mice whose ages are indicated by the number on the left. Fraction 1 represents the bottom of the gradient, and fraction 18 represents the top. M is a marker lane with 1 µg of newborn mouse liver poly(A)⁺ RNA. The migration of 28S rRNA in each gradient is indicated by the asterisk above the appropriate lane.

of a conserved ORF and the finding that the RNA is located in a cytoplasmic particle are properties commonly associated with structural RNAs. Such RNAs may be transcribed by any one of the three RNA polymerases. To ascertain which RNA polymerase transcribes *H19* RNA, nuclear run-on experiments using nuclei isolated from the murine myoblast cell line, S2, were performed in the presence and absence of α -amanitin. In the presence of 5 µg of α -amanitin per ml, *H19* transcription was inhibited in parallel with actin gene transcription, whereas neither 18S rRNA nor 5S RNA transcription was affected (Fig. 7). Thus, the *H19* gene is transcribed by RNA polymerase II.

DISCUSSION

The mouse *H19* gene has now been isolated in two other differential cDNA library screens. Davis et al. (6) identified it in a screen for genes that activate early during the differentiation of C3H10T1/2 cells and called it MyoH. More

FIG. 7. In vitro nuclear transcription of the *H19* gene. Nuclei from S2 myoblasts were isolated and incubated in the presence of [α -³²P]UTP. The labeled RNA was isolated and hybridized to nitrocellulose filters to which plasmid DNA containing segments of the murine *H19*, α -actin (23), rRNA, and 5S RNA genes as well as pUC18 DNA had been bound.

recently, Poirier et al. (F. Poirier, C.-T. J. Chan, P. Timmons, E. Robertson, and P. J. Rigby, manuscript in preparation) isolated it as a cDNA that is activated early upon differentiation of embryonic stem cells. In each instance, it was cloned because of its regulatory pattern of expression, a pattern that reflects its high level of expression in most endodermal and mesodermal cell types in the developing mouse embryo (Poirier et al., in preparation; H.-S. Kim and S. M. Tilghman, unpublished observations).

That *H19* might have an unusual function was first suggested by the absence in the mouse gene of a significant ORF that spanned the exonic borders. Although there are examples of genes in which the 5'- and 3'-most exons are not translated (4, 7, 19), there is as yet no example of a gene in which multiple 5' or 3' exons are untranslated. We originally concentrated on ORF5, the longest of the mouse ORFs. However, we could marshal no evidence that it was ever translated either in vivo or in vitro from the full-length mRNA. Consequently, we adopted a strategy to identify the *H19* gene product that was not biased in favor of any one ORF, to clone the human *H19* gene, and compare it with the mouse gene with the aim of identifying conserved regions. To our surprise, even though the genes were well conserved, there was no common ORF.

Does *H19* mRNA encode a protein or act as an RNA in the cell? It shares features common to other mRNAs: it is transcribed by RNA polymerase II, processed by RNA splicing, and polyadenylated. Unlike mRNAs, it is not associated with ribosomes in any tissue or at any stage in development that was examined. There are three caveats to concluding from this finding that the RNA is never translated. First, if a small fraction (<5%) of *H19* mRNA was associated with ribosomes, we would have failed to detect it because of the insensitivity of the analysis. Second, it is possible that translation of *H19* mRNA occurs for a very limited time during development, as is the case for murine protamine mRNA, which is stored for up to 1 week during spermatogenesis before translation is initiated (13). Finally, *H19* mRNA may be translated only under conditions of stress, similar to the product of the yeast *GCN4* gene, which is translated only during amino acid starvation (10, 32, 35). Thus, if *H19* RNA functions as a mRNA, then it must be under stringent translational control.

If *H19* mRNA were translated, then what would the protein product be? The sequence comparison reveals that the mouse and human genes would translate different proteins, either from a conserved AUG codon (Fig. 3A) or potentially from different initiator codons. The likeliest

candidate in the human gene is ORF6, the longest ORF. It lies in the same general region of the gene as the mouse ORF5, beginning 629 bp downstream of the cap site within the first long exon. Its size, sufficient to encode a 26-kilodalton protein, is significantly longer than would be expected by chance alone. Finally, it encompasses the most conserved region in the gene, at the end of exon 1. However, the human ORF6 and the mouse ORF5 would share neither initiator nor terminator codons and very few amino acids, with the exception of a short stretch toward their carboxy termini (Fig. 3B). In addition, the human ORF6 bears the sequence GGGTTATGG surrounding its translational initiation site, which conforms to only two positions of the Kozak consensus (15–17), not a favorable configuration for efficient translation.

If the two genes encode substantially different proteins like the ORF6 and ORF5 products, it would, to our knowledge, represent only the second example of such phenomenon. The other case involves a *Drosophila melanogaster* heat shock gene, termed *hsrw*, which is a major site of transcription during heat shock. This gene is moderately conserved among *Drosophila* species and until recently was not thought to encode a protein. However, Fini et al. (8) determined that the mRNA for this gene is associated with monosomes and disomes. The translated protein is 23 to 27 amino acids in length, depending on the species, but only the first 3 amino acids are conserved. The authors suggest that the function of the gene lies in monitoring the state of translational activity of the cells rather than in the specific function of the protein product (8). Our failure to observe the association of either human or mouse *H19* RNA with ribosomes makes this a less likely possibility for the mammalian *H19* gene.

The degree of sequence homology observed between the mouse and human genes precludes their having become pseudogenes prior to mammalian radiation. However, it is conceivable that one or even both of them have become pseudogenes since the species separated 30 million years ago. If only one of them has become inactivated, it is difficult to explain why both RNAs were found in association with the same size particle in the cytoplasm. The major argument against both of them having suffered deleterious mutations is the high levels of transcription of the genes. The mRNA of the mouse gene accumulates to greater than 1% of total mRNA in fetal liver, gut, cardiac and skeletal muscle and declines in all tissues after birth (25). Likewise, the human gene is transcribed at high levels in hepatoma cells. This is reflected in significant sequence homologies in the transcriptional control regions of the two genes, especially the first 175 bp upstream of the cap site (80%) and 250 bp of the 3-distal enhancer at +8 kb (~90%; data not shown). The maintenance of transcriptional competence is not normally associated with a loss of function of genes.

There are several additional, albeit unlikely, explanations for the lack of a common ORF between the two *H19* homologs. It is possible that a common ORF is generated by a splicing pattern different from the one revealed by the mouse and human cDNAs. If this is the case, then the alternate RNA must be rare, as no other RNA species are detected by Northern blotting or S1 nuclease mapping. The recent observation that translation may initiate at CTG codons (9) does not mitigate the problem, as the density of terminators in both genes does not permit ORFs longer than the human ORF6 or the mouse ORF5. Last, it is possible that frameshifts occur during the translation of one or both genes (5, 11), or as in the case of the apoprotein B gene,

posttranscriptional modifications of the mRNA may affect its reading frame (27).

This leads us to consider the possibility that the product of the *H19* gene functions as an RNA molecule. Analysis of its subcellular distribution in both endodermal and mesodermal cell types of both humans and mice revealed that a large percentage of the RNA exists in a cytoplasmic particle, with a sedimentation coefficient of 28S. A growing number of RNAs that are required for and probably catalyze reactions involved in RNA processing, transport of proteins, and protein synthesis have been identified. Some of these, such as the U series of small nuclear RNAs, are transcribed by RNA polymerase II, although to date none have been found to be polyadenylated or spliced. It is clear that the *H19* gene is not as well conserved among mammals as are the majority of these structural RNAs (for a compilation of these sequences, see reference 28). This is partly the consequence of the fact that the *H19* gene is not uniformly conserved over its length. Rather, there is interspersed conservation and divergent domains, which range from 65% in the least conserved domains to >90% in the most conserved segment of exon 1. This pattern holds for the chicken *H19* gene as well (Fig. 5). Such a pattern is reminiscent of the pattern observed when small RNAs are compared over much longer evolutionary distances. For example, although members of the U series of RNAs from yeast are generally longer and have diverged from the mammalian counterparts, conserved domains can be found (31). Likewise, whereas the RNA components of RNase P from *Saccharomyces cerevisiae* and *Saccharomyces pombe* share only 52% homology overall, there are subregions that are 80% similar (18). Thus, if *H19* is functioning as an RNA, it is diverging far more rapidly than other genes of its kind but in a similar manner.

In conclusion, the conservation between the mammalian and avian *H19* genes has suggested that the gene has a common function in these organisms, which may not necessarily be to encode a protein. Future experiments will focus on the nature of the cytoplasmic particle containing *H19* RNA. In addition, the phenotypic effects of overproducing wild-type and mutant RNAs in vivo and of deleting the gene from the mouse genome will be investigated.

ACKNOWLEDGMENTS

We thank Louise Pape for DNA probes to the rRNA and 5S RNA genes, Philip Leder for the human genomic library, and members of the laboratory for many helpful discussions.

This work was supported by grants from the National Institutes of Health (Public Health Service grant CA44976) and the National Foundation for Birth Defects. C.I.B. is supported by a Public Health Service cancer training grant from the National Institutes of Health.

LITERATURE CITED

1. Baker, M. E. 1988. Evolution of alpha-fetoprotein: sequence comparisons among AFP species and with albumin species. *Tumour Biol.* 9:123–136.
2. Belayew, A., and S. M. Tilghman. 1982. Genetic analysis of alpha-fetoprotein synthesis in mice. *Mol. Cell. Biol.* 2:1427–1435.
3. Benton, W. D., and R. W. Davis. 1977. Screening λ gt recombinant clones by hybridization to single plaques in situ. *Science* 196:180–182.
4. Butnick, N. Z., C. Miyamoto, R. Chizzonite, B. R. Cullen, B. Ju, and A. M. Skalka. 1985. Regulation of the human c-myc gene: 5' noncoding sequences do not affect translation. *Mol. Cell. Biol.* 5:3009–16.
5. Clare, J., and P. Farabaugh. 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. USA* 82:2829–2833.

6. Davis, R. L., H. Weintraub, and A. B. Lassar. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51:987-1000.
7. Ecker, R. L., and H. Green. 1986. Structure and evolution of the human involuerin gene. *Cell* 46:583-589.
8. Fini, M. E., W. G. Bendena, and M. L. Pardue. 1989. Unusual behavior of the cytoplasmic transcript of Hsr omega: an abundant, stress-inducible RNA which is translated, but which yields no detectable protein product. *J. Cell Biol.* 6:2045-2057.
9. Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman. 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* 52:185-195.
10. Hinnebusch, A. G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci. USA* 81:6442-6446.
11. Jacks, T., and H. E. Varmus. 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science* 230:1237-1242.
12. Kioussis, D., F. Eiferman, P. van de Rijn, M. B. Gorin, R. S. Ingram, and S. M. Tilghman. 1981. The evolution of alpha-fetoprotein and albumin. II. The structure of the alpha-fetoprotein and albumin genes in the mouse. *J. Biol. Chem.* 256:1960-1967.
13. Kleene, K. C., R. J. Distel, and N. B. Hecht. 1984. Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. *Dev. Biol.* 105:71-79.
14. Konieczny, S. F., and C. P. Emerson, Jr. 1985. Differentiation, not determination, regulates muscle gene activation: transfection of troponin I genes into multipotential and muscle lineages of 10T1/2 cells. *Mol. Cell. Biol.* 5:2423-2432.
15. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233-5252.
16. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids. Res.* 12:857-872.
17. Kozak, M. 1986. Bifunctional messenger RNAs in eukaryotes. *Cell* 47:481-483.
18. Lee, J.-Y., and D. R. Engelke. 1989. Partial characterization of an RNA component that copurifies with *Saccharomyces cerevisiae* RNase P. *Mol. Cell. Biol.* 9:2536-2543.
19. Mangin, M., K. Ikeda, B. E. Dreyer, and A. E. Broadus. 1989. Isolation and characterization of the human parathyroid hormone-like peptide gene. *Proc. Natl. Acad. Sci. USA* 86:2408-12.
20. Maniatis, T., A. Jeffrey, and D. G. Kleid. 1975. Nucleotide sequence of the rightward operator of phage λ . *Proc. Natl. Acad. Sci. USA* 72:1184-1188.
21. Messing, J., B. Bronenborn, B. Muller-Hill, and P. H. Hofschneider. 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind II fragment of the *lac* regulatory region in M13 replicative form *in vitro*. *Proc. Natl. Acad. Sci. USA* 74:3642-3646.
22. Minghetti, P. P., S. W. Law, and A. Dugaiczky. 1985. The rate of molecular evolution of alpha-fetoprotein approaches that of pseudogenes. *Mol. Biol. Evol.* 2:347-58.
23. Minty, A. J., M. Caravatti, B. Robert, A. Cohen, P. Daubas, A. Weydert, F. Gros, and M. E. Buckingham. 1981. Mouse actin messenger RNAs. *J. Biol. Chem.* 256:1008-1014.
24. Olsson, M., G. Lindahl, and E. Rouslahti. 1977. Genetic control of alphafetoprotein synthesis in the mouse. *J. Exp. Med.* 145:819-827.
25. Pachnis, V., A. Belayew, and S. M. Tilghman. 1984. Locus unlinked to alpha-fetoprotein under the control of the murine raf and Rif genes. *Proc. Natl. Acad. Sci. USA* 81:5523-5527.
26. Pachnis, V., C. I. Brannan, and S. M. Tilghman. 1988. The structure and expression of a novel gene activated in early mouse embryogenesis. *EMBO J.* 7:673-681.
27. Powell, L. M., S. C. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott, and J. Scott. 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 50:831-840.
28. Reddy, R. 1988. Compilation of small RNA sequences. *Nucleic Acids Res.* 16(Suppl.):r71-r85.
29. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5468.
30. Schneider, R. J., C. Weinberger, and T. Shenk. 1984. Adenovirus VA1 RNA facilitates the initiation of translation in virus-infected cells. *Cell* 37:291-298.
31. Siliciano, P. G., M. H. Jones, and C. Guthrie. 1987. *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. *Science* 237:1484-1487.
32. Thireos, G., M. D. Penn, and H. Greer. 1984. 5' untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl. Acad. Sci. USA* 81:5096-5100.
33. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* 77:5201-5205.
34. 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.
35. Tzamarias, D., I. Roussou, and G. Thireos. 1989. Coupling of GCN4 mRNA translational activation with decreased rates of polypeptide chain initiation. *Cell* 57:947-954.
36. Young, P. R., and S. M. Tilghman. 1986. The evolution of the AFP and albumin genes, p. 55-83. *In* D. G. Hardie and J. R. Coggins (ed.), *Multidomain proteins—structure and evolution*. Elsevier Science Publishers B.V., The Netherlands.
37. Zakeri, Z. F., C. Ponzetto, and D. J. Wolgemuth. 1988. Translational regulation of the novel haploid-specific transcripts for the c-abl proto-oncogene and a member of the 70 kDa heat-shock protein gene family in the male germ line. *Dev. Biol.* 125:417-422.