MOLECULAR AND CELLULAR BIOLOGY, Nov. 1988, p. 4707–4715
0270-7306/88/114707-09$02.00/0

Vol. 8, No. 11

Two Regulatory Domains Flank the Mouse H19 Gene

HEEJA YOO-WARREN,† VASSILIS PACHNIS,‡ ROBERT S. INGRAM, AND SHIRLEY M. TILGHMAN*

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

Received 16 May 1988/Accepted 16 August 1988

The mouse H19 gene was identified by virtue of its coordinate regulation with the mouse α-fetoprotein gene. Both genes are expressed in the fetal liver, gut, and visceral endoderm of the yolk sac and are repressed shortly after birth in the liver and gut. They are both under the control of two *trans*-acting loci: *raf*, which affects the adult basal levels of the two mRNAs, and *Rif*, which affects their inducibility during liver regeneration. One crucial difference between the two genes is the activation of the H19 gene in mesoderm derivatives, skeletal and cardiac muscle. As a strategy for explaining both the similarities and differences in their modes of expression, the regulatory domains responsible for the expression of the H19 gene in liver were identified by transiently introducing the gene into a human hepatoma cell line. Two regions necessary for high-level expression of the gene could be identified, a promoter-proximal domain immediately preceding the start of transcription and an enhancer domain which lies between 5 and 6.5 kilobases 3′ of the polyadenylation site. The 3′ domain consists of two separable enhancer elements, each of which exhibits the properties of tissue-specific enhancers. Nucleotide sequence comparisons between the two H19 and three α-fetoprotein enhancers revealed limited similarities which are candidates for binding of common regulatory factors. Sequences which lie 3′ of the gene are also required for the expression of the H19 gene following differentiation of teratocarcinoma cells into visceral endoderm.

A central goal of developmental biology is to elucidate the mechanisms by which sets of genes are activated and subsequently modulated in a coordinate manner during cellular commitment and differentiation. Several years ago, investigators in our laboratory initiated a study to identify regulatory genes in the mouse which act in *trans* to affect the transcription of a set of unrelated genes and would thus qualify as genes involved in coordinate regulation. Two such loci, designated *raf* and *Rif*, were identified by screening inbred strains of mice for variants that differ in the regulation of the α-fetoprotein (AFP) gene. The *raf* locus affects the adult basal level of AFP mRNA, and the *Rif* locus affects its inducibility during liver regeneration (5, 30).

Subsequent genetic and molecular studies demonstrated that *raf* and *Rif* regulate at least one other structural gene, which we refer to as H19 (31, 32). The H19 gene, which is present as a single copy on chromosome 7, is unrelated in both structure and sequence to the AFP gene. Therefore, the two genes constitute an unrelated set of genes coordinately regulated by at least two different *trans*-acting loci. A detailed analysis of the tissue specificity of the expression of the H19 gene (31) showed that in addition to activation in the liver, the gene is activated in the visceral endoderm and fetal gut, which also express the AFP gene. However unlike AFP, the H19 gene is also expressed in skeletal and cardiac muscle. The level of H19 gene expression in adult cardiac muscle is ca. 20-fold lower than in fetal cardiac muscle, but this postnatal development regulation does not appear to be affected by the *raf* gene (31).

Our long-term goal is to investigate the molecular basis for the similarities in the regulation of the AFP and H19 genes, that is, their endoderm tissue specificity and their modulation by *raf* and *Rif*. The unique expression of the H19 gene in mesoderm raises a second question regarding the mechanism by which the regulatory domain of a single gene can specify expression in several unrelated cell types. That is, does the expression in mesoderm proceed through the endoderm *cis*-acting elements, or is it encoded in separate domains?

Toward these ends, in this communication we demonstrate that the expression of the H19 gene in liver cells is controlled by two distinct domains of regulatory elements: a tissue-specific promoter at the 5′ end of the gene and at least two enhancer elements that lie 3′ of the gene. The 3′-regulatory domain is also involved in the expression of the gene in visceral endoderm.

MATERIALS AND METHODS

Cell lines. Hep3B and HepG2 human hepatoma cell lines (25) were obtained from B. Knowles, Wistar Institute, Philadelphia, Pa., and G. Darlington, Baylor College of Medicine, Houston, Texas. They were maintained in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum.

The PC 13 teratocarcinoma cell line (1), provided by Brigid Hogan, Medical Research Council, Mill Hill, London, England, was maintained in Dulbecco modified Eagle medium with 10% fetal bovine serum. For the differentiation to visceral endoderm in monolayers, 1.5×10^5 cells were seeded in gelatin-treated tissue culture dishes in Dulbecco modified Eagle medium-F12 medium with 10% fetal bovine serum and 10^{-6} M all *trans* retinoic acid. The cells were fed after 3 days and every day thereafter as described previously.

Construction of 5′ and 3′ deletions of the H19 gene. All constructs used to generate the series of 5′ truncations (see Fig. 2) contained a 3.5-kilobase-pair (kb) *Eco*RI-*Sal*I fragment spanning the H19 structural gene, into which two *Xba*I linkers had been inserted at a *Sac*I site at +580 base pairs (bp) in exon 1 (see Fig. 7A). The minimal construct also contained 48 bases of 5′ flanking sequences and 8 kb of 3′

* Corresponding author.

† Present address: Molecular Therapeutics Inc., West Haven, CT 06516.

‡ Present address: Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, NY 10032.

flanking sequences, as well as a 500-bp *Alu*I fragment which encompasses the adenovirus VA1 gene (18), inserted into the unique *Pvu*II site of pBR322. The endpoints of the deletions in the 5′ flanking DNA were the restriction endonuclease sites *Eco*RI (−4 kb), *Xba*I (−700 bp), *Xmn*I (−240 bp), *Hin*fI (−207 bp), *Hph*I (−127 bp), and *Rsa*I (−76 bp), which were all converted to *Eco*RI sites by addition of linkers.

For the 3′ deletion constructs used (see Fig. 3 and 4), various restriction fragments were either made blunt ended or converted to an appropriate restriction site by the addition of synthetic linkers and then inserted into either the *Bgl*II site at −2 kb, the *Xba*I site at −700 bp, or the *Pvu*II site at −400 bp. In each case the H19 structural gene (+*Xba* linkers) was present on the 3.5-kb *Eco*RI-*Sal*I fragment, with approximately 400 bp of 3′ flanking DNA.

Hybrid constructs containing the herpes simplex virus (HSV) thymidine kinase (*tk*) gene utilized a *Pvu*II fragment which began at −200 bp in the 5′ flanking region and extended to 500 bp downstream of the polyadenylation site (27). This was inserted into pUC18, which allowed the addition of both AFP- and H19-derived fragments into the additional multiple cloning sites.

Transient-expression assays. Transient-expression assays were performed as described by Scott and Tilghman (38). Cells (1×10^6 to 4×10^6) were seeded in 10-cm tissue culture plates. The next day, 10 to 15 μg of plasmid, precipitated in calcium phosphate, was added and left on the cells for 12 to 16 h. The DNA precipitate was then removed, and the cells were washed with sterile phosphate-buffered saline and refed fresh medium. The cells were harvested 48 h later, and poly(A)$^+$ RNA was prepared by hot-phenol extraction (37) and oligo(dT)-cellulose chromatography (3).

RNA and DNA analysis of transfected cells. For Northern blots, 1 to 7 μg of poly(A)$^+$ RNA was denatured in 70% formamide–6% formaldehyde in MOPS buffer (20 mM morpholinepropanesulfonic acid [MOPS], 5 mM sodium acetate, 1 mM EDTA [pH 7]) at 55°C for 15 min and electrophoresed in a 6% formaldehyde–1.2% agarose gel in MOPS buffer. Conditions for transfer, hybridization, and washing were as described previously (42, 46). The specific probes, a 3′ genomic-distal DNA fragment consisting of a 2-kb *Hin*dIII fragment, a mouse ribosomal protein probe (9), an HSV *tk* gene *Pvu*II fragment (27), and a 0.5-kb *Alu* fragment containing the VA1 gene of adenovirus type 2 (18), were gel purified and labeled with [^{32}P]dCTP by nick translation.

To normalize the DNA uptake for each transfection experiment, we lysed 10% of the transfected cells by the method of Hirt (22). After digestion of the cells with proteinase K, the supernatant was extracted with phenol-chloroform (1:1) and precipitated in ethanol. Approximately 10 μg of DNA was digested with appropriate restriction enzymes, separated on a 1% agarose gel in 40 mM Tris-acetate–1 mM EDTA (pH 7.2), and transferred to nitrocellulose filters (41). Conditions used for prehybridization, hybridization, and washing were identical to those used for RNA blots.

S1 nuclease mapping. The restriction enzyme fragment used for S1 protection analysis was asymmetrically labeled at its 5′ ends by cleavage with *Bam*HI, and incubation with bacterial alkaline phosphate followed by polynucleotide kinase in the presence of [γ-^{32}P]ATP. The fragment was digested with a second enzyme and purified by gel electrophoresis.

The S1 protection assay was performed as described by Scott and Tilghman (38) by hybridizing 0.5 to 1.0 μg of labeled DNA with 1 to 5 μg of poly(A)$^+$ RNA in 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES [pH 6.4])–400 mM NaCl–1 mM EDTA–80% deionized formamide at 42 to 52°C for 12 to 14 h. The solution was adjusted to 300 mM NaCl, 30 mM sodium acetate (pH 4.5), 3 mM $ZnCl_2$, and 20 μg of denatured calf thymus DNA per ml, and digested with S1 nuclease at 37°C for 45 min. The nucleic acids were precipitated in ethanol and resolved on 5 to 8% polyacrylamide–urea gels in 90 mM Tris-borate–2.5 mM EDTA.

DNA sequencing. The 2-kb *Xba*I-*Xba*I-*Bgl*II fragment (+8 to +10 kb) spanning the H19 gene enhancers was subcloned by several strategies into the vectors M13mp10 and M13mp11 (28), and the sequence was determined on both strands by using the dideoxynucleotide chain termination method (36).

RESULTS

Two transcriptional-regulatory domains in the H19 gene. The initial strategy adopted to localize the *cis*-acting regulatory elements in the H19 gene was to introduce modified copies of the gene into liver-derived cells and to assay the level of its transcription 48 h later. To have the capability of examining possible regulatory effects in homologous cells expressing the endogenous gene (see, e.g., Fig. 7), we modified the cloned gene by inserting two tandem *Xba*I linkers into a *Sac*I site at +580 bp in exon 1. In this way, the H19 mRNA derived from the endogenous mouse gene can be distinguished from that of the reintroduced gene by an S1 nuclease assay.

The modified H19 gene, along with varying amounts of both 5′- and 3′-flanking DNA, was introduced transiently by calcium phosphate precipitation into Hep3B human hepatoma cells, which also transcribe the human H19 gene (data not shown). At 48 h, the cells were harvested and poly (A)$^+$ RNA was prepared. Transcripts were measured by gel electrophoresis, blotting onto nitrocellulose filters, and hybridization to a labeled 3′-derived genomic DNA probe. The human endogenous H19 mRNA is not detected by this assay, owing to a significant mismatch between the human and mouse sequences in the region of the probe. The same blots were also hybridized to a ribosomal protein cDNA (Fig. 1C) to ensure that all lanes contained equivalent amounts of RNA.

To control for the efficiency of transfection, we inserted the adenovirus type 2 VA1 gene into the *Pvu*II site of the pBR322 vector. The transcriptional activity of this gene, which is transcribed by RNA polymerase III and therefore is insensitive to the presence of RNA polymerase II regulatory elements (6), was determined by agarose gel electrophoresis and blotting with the isolated VA1 gene as a probe (Fig. 1B). Densitometric scanning of the blots hybridized to these two control genes were then used to normalize the densitometric scans of similar blots hybridized to a mouse H19 probe.

No transcripts were detected from the minimal construct, pH19.RS, which is composed of a 3.5-kb *Eco*RI-*Sal*I restriction fragment that begins at −48 bp relative to the start of transcription and ends approximately 400 bases downstream from the polyadenylation site (Fig. 1A, lane 1). When a 4-kb *Eco*RI fragment was included 5′ to the gene (pH19.RRS) (Fig. 1, lanes 2), a 2.5-kb mRNA was detected at low levels. No further increase was observed when an additional 8 kb of 5′-flanking DNA was included (pH19.RRRS; data not shown). However, a 50- to 70-fold further increase in the activity of pH19.RRS was obtained when an additional 8 kb of DNA 3′ to the gene was included (pH19.RRSR; Fig. 1, lanes 3). S1 nuclease mapping of the 5′ ends of the H19 gene transcripts indicated that they were being initiated at the

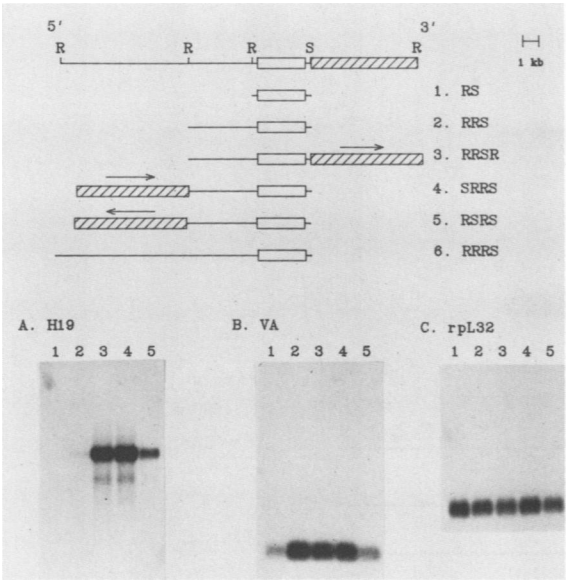

FIG. 1. A cellular enhancer in the 3'-flanking domain of the H19 gene. Symbols: ▭, H19 gene; ▨, 3'-flanking 8-kb *Sal*I (S)-EcoRI (R) fragment. Constructs which contain either 48 bp (construct 1) or 4 kb (constructs 2 and 5) or 11 kb (construct 6) of 5'-flanking sequence with or without the 3'-flanking domain were generated. Constructs 1 to 5 were transfected into Hep3B cells, and RNA was isolated 48 h later. Total RNA (B) or poly(A)$^+$ RNA (A and C) were run on 6% formaldehyde gels, transferred to nitrocellulose filters, and hybridized to an internal *Hin*dIII fragment of the H19 gene (A), an adenovirus VA1 gene probe (B), or an rpL32 pseudogene (C). The lanes are numbered according to the constructs.

authentic H19 cap site (data not shown). These experiments suggested that at least two regions were essential for full activity of the H19 gene in the transient-expression assay: one 5' and the other 3' to the gene.

It was possible that the requirement for sequences 3' to the gene involved a requirement for correct termination of transcription and not necessarily a requirement for a regulatory element that affected the rate of transcription per se. To discriminate between these two possibilities, the 8-kb 3'-distal restriction fragment was inserted 4 kb upstream of the transcription start site in both orientations. When these constructs were tested by transient assay in Hep3B cells, both were able to restore high-level transcription relative to pH19.RRS in a manner equivalent to pH19.RRSR, in which the fragment is in its normal 3' position (Fig. 1, lanes 4 and 5). The position and orientation independence of the 8-kb 3' fragment indicates that it contains a cellular enhancer.

Deletion mapping of the 5'-flanking domain of the H19 gene. We defined the minimum sequences in the 5'-flanking region that are necessary for H19 gene expression by generating a series of increasingly large deletions in that region, keeping the 3'-flanking domain intact, and these were tested for their transcriptional competence as described in the legend to Fig. 1. As shown in Fig. 2, 127 bp upstream of the transcription start site was sufficient to maintain expression of the gene at high levels. This level was reduced twofold when the gene was truncated to -76 bp, and expression was lost when a further deletion to -48 bp was tested. The decrease in VA1 transcripts in this experiment was not observed in other experiments and does not reflect a concomitant loss of H19 and VA transcription. When the identical constructs were tested in HeLa cells, which do not express the endogenous H19 gene, no transcription was detected (data not shown). This implies that the sequence information required for tissue-specific expression of the H19 gene lies downstream of −127 bp.

The DNA sequence of the 5' region has been determined previously (32). Unlike the AFP-proximal promoter (38), the H19 promoter is G+C-rich, with consensus sequences at −55 and −42 bp for the binding of the transcription factor Sp1 (10). A CCAAT motif on the noncoding strand is located farther upstream at −72 bp. It is likely that the loss of transcriptional competence upon deletion between −76 and −48 bp is the consequence of the loss of one or more of these motifs.

Localization of two cellular enhancers in the 3'-flanking domain. To localize the regulatory sequences in the 3'-flanking region, we inserted a series of overlapping restriction fragments in that region into convenient restriction sites 5' of the gene. The decision to use 5' cloning sites was based on a wish to eliminate possible effects of changing the transcriptional termination site on the accumulation of stable transcripts. The resulting constructs were assayed by transient expression in Hep3B cells. The 3'-distal enhancing activity was associated with a 2.4-kb *Xba*I-*Xba*I-*Bgl*II fragment, located approximately 8 kb 3' of the transcription start site (Fig. 3, lane 6). The 5 kb between the end of the gene and the 5'-proximal *Xba*I site retained no enhancer activity when tested alone (lane 5). When the 1-kb *Xba*I fragment between +8 and +9 kb was tested alone (lane 8), it exhibited approximately half the activity of the intact 3' domain. Likewise, when the *Xba*I fragment was deleted from the 3' domain, the remaining fragment retained approximately half the original activity. These data suggested that either a single enhancer element was interrupted by the 3' *Xba*I site or there was more than one element within the 2.4-kb fragment.

More-detailed mapping of the 2.4-kb region (Fig. 4) supported the latter interpretation, i.e., that there are two separate enhancer elements. The 5' element was localized to a 269-bp *Xba*I-*Pst*I fragment, and the 3' enhancer lies 700 bp farther downstream within a 297-bp *Hin*cII-*Pst*I fragment. Each element increased the level of H19 mRNA approximately 30-fold, which is 40% of the level achieved when the entire 2.4-kb fragment is tested. A 700-bp *Hin*cII fragment which separates the 5' and 3' elements had no biological activity (Fig. 4).

Transcriptional activity of the H19 enhancers in other cell types. The tissue specificity of the H19 enhancers could not be tested within the context of the H19 promoter, since it was inactive in heterologous cells (data not shown). Therefore, the 8-kb 3' domain was inserted upstream of the HSV *tk* gene, and its transcriptional activity in both Hep3B and HeLa cells was determined. The *tk* mRNA level was ca. fivefold higher in Hep3B cells when the H19 enhancers were present, compared with the *tk* promoter alone (Fig. 5). In addition, this enhancement was not seen in HeLa cells, in which both the AFP and simian virus 40 enhancers have been shown to be active (15, 38), demonstrating that the enhancers function in a tissue-specific manner. It is interesting that the enhancement of the H19 enhancers was significantly lower than that achieved by one of the AFP gene enhancers (Fig. 5, gel 1, lane 2). It was also lower than the enhancement obtained when the enhancers were tested within the context of the H19 promoter and structural gene (Fig. 1, 3, and 4). It is not clear whether this represents a synergy between the H19 enhancers and promoter, or whether the *tk* promoter, by virtue of its high basal activity

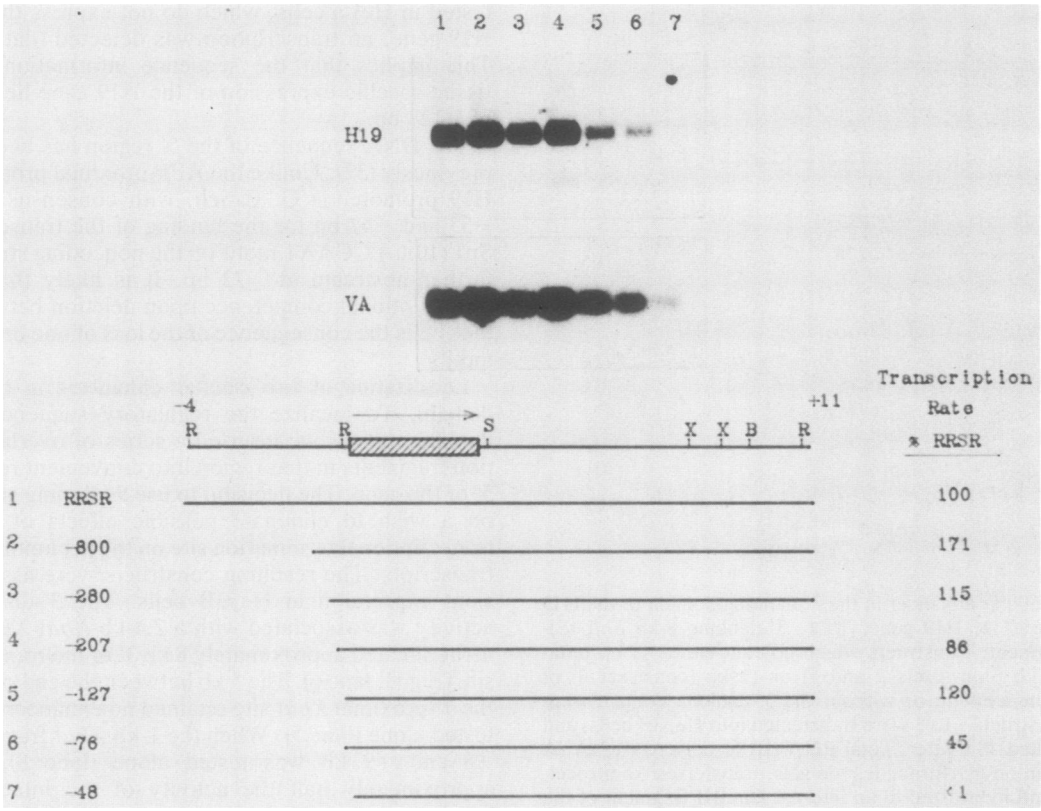

FIG. 2. Transcriptional analysis of the 5'-flanking domain of the H19 gene. The plasmid RRSR, which extends from −4 to +11 kb, and a set of 5'-flanking deletions whose endpoints are indicated on the left of the figure were transfected into Hep3B cells. RNA was prepared 48 h later, and the levels of mouse H19 mRNA and VA1 RNA were determined by agarose gel electrophoresis and hybridization to H19 and VA1 gene probes, as described in the legend to Fig. 1. The rates of transcription of each construct, normalized to the amount of VA1 RNA, are expressed as the percentage of the rate obtained with pH19 RRSR. The values are the average of three experiments.

in Hep3B cells, is less responsive to the presence of an enhancer. To control for transfection efficiencies, we prepared DNA from each transfection by the method of Hirt (22) and quantitated it by gel electrophoresis and hybridization to a *tk*-specific probe following transfer onto nitrocellulose (data not shown).

The 3' domain contains at least two separate enhancer elements (Fig. 4). Since these two elements did not exhibit any functional differences in liver cells, we sought to test their activities in the context of the heterologous *tk* promoter in expressing and nonexpressing cells. When inserted upstream of the HSV *tk* gene, both enhancers (labeled I and II in Fig. 5) increased transcription of the *tk* gene approximately fivefold over the control level in Hep3B cells, and the tissue specificity of expression was retained with both elements, in that no enhancement was observed in HeLa cells (Fig. 5B, lanes 4 to 6). These experiments demonstrate that the H19 gene contains at least two enhancers within its 3'-flanking region which behave as classic tissue-specific cellular enhancers in that they can increase the rate of transcription in an orientation- and position-independent manner and exhibit cell type specificity.

The H19 gene was originally identified by virtue of its similarities in expression to the AFP gene (31). Both genes not only respond to *raf* and *Rif*, but also activate in parallel during the differentiation of either F9 or PC 13 teratocarcinoma stem cells to visceral endoderm (31). Naïvely, one might therefore expect that they share the same *trans*-acting factors in visceral and definitive endoderm at least. However, several lines of evidence suggested a qualitative difference between the AFP and H19 genes in the manner in which they interacted with cellular factors in liver. First, repression of the H19 gene occurs 1 week after that of the AFP gene in the neonatal liver, although the decline in both cases is dependent on the *raf* gene product (31). Second, HepG2 cells, which express both endogenous and transfected AFP genes at high levels, do not express either endogenous or transfected H19 genes, suggesting that at least one crucial factor differs between them. Therefore, we decided to use these two cell lines to test whether the lack of H19 expression in HepG2 was mediated by one of the regulatory elements we had identified.

For these experiments, the 2.4-kb fragment which contains both of the H19 enhancers was linked to an AFP minigene promoter (38) and tested for its activity in both HepG2 and Hep3B cells. This fragment, which exhibits full enhancing activity in Hep3B cells, did not affect transcription of the AFP minigene in HepG2 cells (Fig. 6, lanes 3). On the other hand, when we tested the H19 promoter and gene with the AFP enhancer, H19 mRNA levels were comparable in the two cell types (Fig. 6, lanes 4). This suggested that the factors which interact with the promoter of the H19 gene are present in both HepG2 and Hep3B cells but that the H19 enhancers are not recognized in HepG2 cells.

This difference in the two liver-derived cell lines could be explained in two ways. First, H19 enhancers require at least one factor for their expression which is not required by the AFP enhancers, and this factor is not present in HepG2

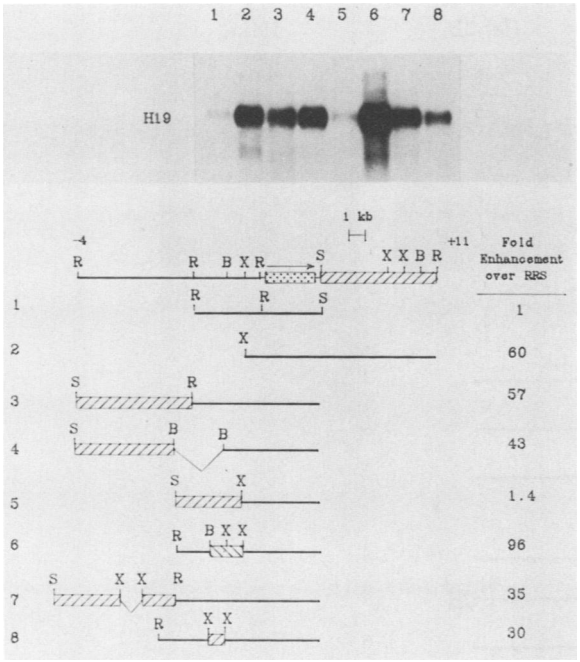

FIG. 3. Localization of a cellular enhancer in the 3'-flanking domain. The H19 gene ( ), along with 4 kb of 5'-flanking sequence and 8 kb of 3'-flanking sequence ( ), is depicted on the first line of the diagram. Below are drawn constructs in which the 3' domain was deleted (construct 1, RRS), present in its normal configuration (construct 2), or placed at sites in the 5'-flanking domain (constructs 3 to 8). These were transfected into Hep3B cells, RNA was prepared 48 h later, and the level of H19 mRNA was determined as indicated in the legend to Fig. 1, after normalization with VA1 RNA and rpL32 RNA. The level of induction of each construct over RRS is indicated (-fold). The restriction sites used to generate the constructs are as follows: R, *Eco*RI; B, *Bgl*II; X *Xba*I; S, *Sal*I. The values are the average of three experiments.

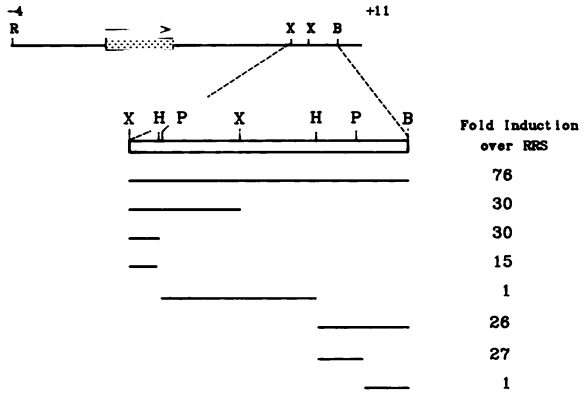

FIG. 4. The 3'-flanking domain contains two separate enhancer elements. The plasmid pH19 RRSR is shown on the top line of the figure, with the H19 gene body indicated ( ). The *Xba*I-*Bgl*II fragment from +8 kb to +10 kb was subcloned into a *Pvu*II site at −400 bp in the 5'-flanking domain of the H19 gene. This construct and a series of smaller inserts into the same site were transfected into Hep3B cells. At 48 h, RNA was prepared, and H19 RNA and VA1 RNA were quantitated by the procedure described in the legend to Fig. 1. The data are expressed as fold induction over RRS, which lacks the 3'-flanking domain. Restriction sites used in the constructs are as follows: X, *Xba*I; H, *Hin*cII; P, *Pst*I; R, *Eco*RI; B, *Bgl*II. The values are the average of four experiments.

cells. Alternatively, this difference is simply due to a lower concentration in HepG2 cells of one or more factors common to the expression of both genes. This lower concentration would then be limiting for H19 but not for AFP transcription.

Requirement for 3'-flanking sequences for expression in visceral endoderm. The demonstration that two distinct regions were required for high-level expression of the H19 gene in liver-derived cells raised the issue of whether the 3'-flanking domain was necessary for expression in the other endoderm cells which normally express the gene. To address this, we took advantage of the mouse embryonal carcinoma cell line PC 13, which differentiates into visceral endoderm in the presence of retinoic acid in monolayer cultures (1). The endogenous H19 gene is silent in the stem cells and is activated during the differentiation process (31).

PC 13 cells were transfected either before or 4 days after differentiation with pH19.RRS and pH19.RRSR, which differ only by the presence of the 3'-flanking domain (Fig. 1). To distinguish the introduced genes from the endogenous gene, we used an S1 protection assay (Fig. 7A). An *Rsa*I-*Bam*HI restriction fragment containing the first 680 bp of exon 1, including two artifical *Xba*I linkers used to mark the exogenous gene, 48 bp of 5'-flanking sequence, and 500 bp of pBR322 vector sequence was labeled at the 3' *Bam*HI site. Endogenous mouse H19 mRNA protects a 100-nucleotide (nt) fragment, as the consequence of the two linkers in the DNA probe creating an S1 nuclease-sensitive discontinuity in the hybrid formed. The H19 mRNA produced from the transfected DNA would protect all the exon 1 sequences and generate a 680-nt fragment.

When PC 13 stem cells were first treated with retinoic acid for 4 days to induce differentiation and then transfected with pH19.RRSR containing the 3' domain, a signal at 680 nt, indicative of the exogenous H19 gene transcript, was detected (Fig. 7B, lane 3). The deletion of the 3' domain from the construct (lane 4) resulted in a significantly lower signal. The strong band at 100 nt in lanes 3 and 4 reflects the level of induction of the endogenous H19 mRNA. When the construct containing the 3' domain was tested in a separate experiment in PC 13 cells both before and after retinoic acid treatment (lanes 5 and 6), low levels of transcript were obtained in the stem cells, although much higher levels were observed in the differentiated derivatives. This suggested that during differentiation the stem cells either acquire the capability of responding positively to an element 3' to the H19 gene or lose the capability of repressing its function.

Nucleotide sequences of the H19 gene enhancers. The nucleotide sequences of the two H19 gene enhancers were determined by the dideoxynucleotide chain termination method (36) following subcloning into M13 vectors (28) (Fig. 8). When they were compared with each other, there were no striking similarities. They were then screened for sequence motifs which are present in the three AFP gene enhancers and promoter, as well as in other liver- and muscle-specific genes. One of these, TGTTTGCAG, which is found in two of the AFP enhancers and the AFP promoter (15), as well as the hepatitis B virus enhancer (40), is perfectly conserved in enhancer II at position 145 and at eight of nine matches in enhancer I at position 258. The motif in enhancer II is preceded by 8 bases (dotted underline) which are found upstream of the equivalent motif in one of the AFP enhancers, although these are displaced by an additional 7 bases. A second motif present in all three AFP enhancers (GCTTTGAGCAA [underlined]) is found at position 249 in enhancer II, with 10 of 11 bases matching the

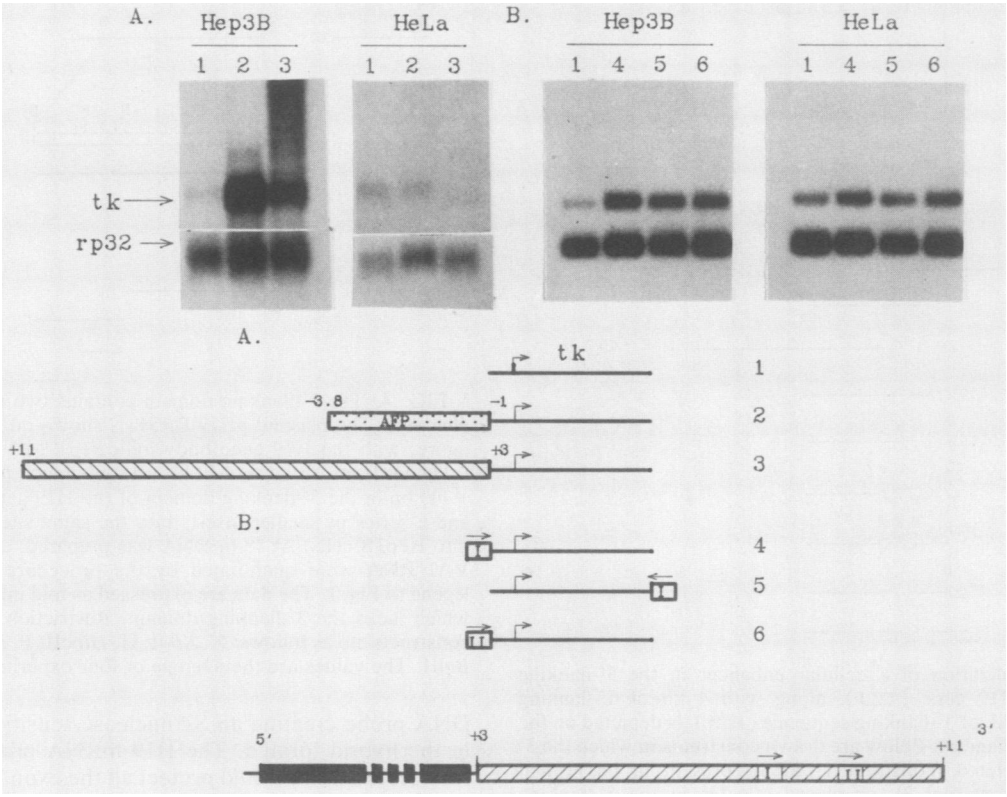

FIG. 5. Cell type specificity of the H19 gene enhancers. (A) The HSV *tk* gene in the absence (lanes 1) or presence of the AFP enhancer I from −1 to −3.8 kb (lanes 2) or the H19 3′-flanking domain (lanes 3) was transfected into either Hep3B or HeLa cells. After 48 h, poly(A)$^+$ RNA was prepared and analyzed by agarose gel electrophoresis and blotting to *tk* and rpL32 gene probes. In addition, Hirt DNAs were prepared from each transfection and used to correct the RNA signals for transfection efficiency. (B) The 5′ *Xba*I-*Pst*I fragment, labeled I, and the 3′ *Hin*cII-*Pst*I fragment, labeled II, contained within the 3′ enhancer domain of the gene (▨), were cloned either 5′ or 3′ of the HSV *tk* gene, in the orientation indicated. The transcriptional activity of these constructs in both Hep3B and HeLa cells was tested as above. The bottom of the figure contains a line drawing of the H19 gene, with the exons shown (▬).

consensus. Neither the motifs defined by Grayson et al. (16) for two liver-specific genes nor the muscle-specific motifs described by Minty and Kedes (29) and Jaynes et al. (23) were found.

DISCUSSION

The purpose of this study was to identify the regulatory domains which govern the expression of the H19 gene in endodermal cells in an effort to elucidate the basis for its coordinate regulation with the AFP gene. In a previous report we established that the AFP and H19 genes are dissimilar in both structure and sequence (32), and from this we argued that their similar modes of expression arose independently in evolution. We began this study with the assumption that these similarities in expression could ultimately be explained by the presence of common sites in their regulatory domains which were binding sites for *trans*-acting factors. This model was based on the paradigms developed for bacterial cells and more recently for *Saccharomyces cerevisiae*, for which such common denominators have been defined (2, 13, 20, 34).

To date, our studies have identified more differences than similarities between the two genes. To our knowledge, all regulation of the AFP gene proceeds via sequences which are present in its 5′-flanking domain, which spans 6.5 kb of DNA (14, 15, 19, 39). The studies in this report show that the 5′-flanking domain of the H19 gene contains a single promoter-proximal regulatory element, which is no larger than 127 bp. The only enhancers we have been able to map in the H19 gene lie 3′ to the gene itself, between 5 and 7 kb downstream of the polyadenylation site.

Enhancers were originally defined as regulatory elements capable of activity in a position- and orientation-independent manner, and thus it was to be expected that they would be found at a variety of positions surrounding the genes they affect. Indeed, two of the first cellular enhancers to be described were contained within introns of the μ and κ immunoglobulin genes (4, 7, 12, 33, 35). Subsequently, 3′ enhancers were identified several hundred bases away from the chicken (8) and human (44) β-globin genes and the histone H5 gene (43) and 18 kb downstream from the T-cell receptor β gene (26). It has not been determined, either in this or in any other case, whether there is any regulatory consequence to having enhancers 3′ to the gene. It may be, for example, that such enhancers control the expression of additional downstream genes. We are currently testing this premise within the H19 gene locus.

The 3′ location of the H19 gene enhancer is of especial interest because of the postnatal repression of H19 transcription. Experiments with the yeast mating-type locus have suggested that positioning a negative element between the promoter and upstream activating sequence results in maximal repression (24). Recent studies with transgenic mice

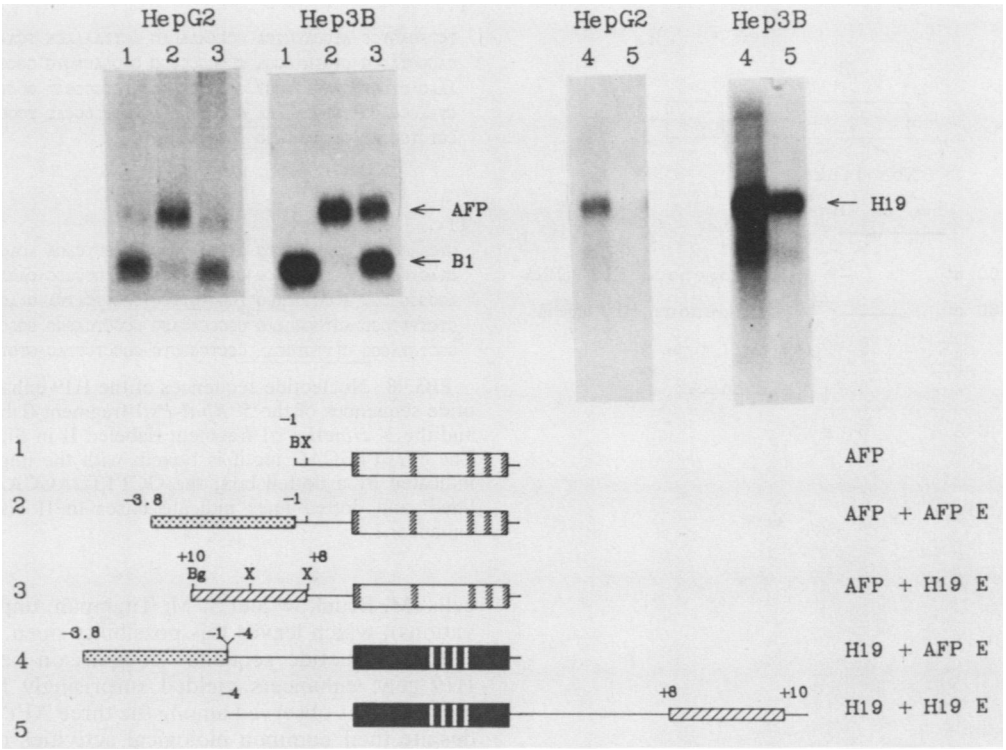

FIG. 6. The H19 gene enhancers are inactive in HepG2 cells. The AFP minigene ZE.5 (38), consisting of the first three and last two exons of the AFP gene along with 1 kb of 5'-flanking and 0.3 kb of 3' flanking DNA, is illustrated in line 1 (AFP). The AFP enhancer from -1 to -3.8 kb (▦) (line 2 [AFP + AFP E]) or the H19 enhancer from +8 to +10 kb (▨) (line 3 [AFP + H19 E]) were cloned into either the *Bam*HI site at -1 kb or the *Xba*I site at -0.8 kb of ZE.5. The AFP enhancer from -1 to -3.8 kb was subcloned upstream of the H19 gene, which spanned from the *Bam*HI site at -2 kb to the *Sal*I site at +3.5 kb (line 4 [H19 + AFP E]), and line 5 (H19 + H19 E) is the H19 gene from −4 to +11.5 kb. These constructs were transfected into both Hep3B and HepG2 cells, and the amount of AFP minigene or H19 gene transcript generated after 48 h was quantitated. To correct for transfection efficiency for the AFP gene constructs (lines 1 to 3), we also quantitated an AFP gene-specific B1 transcript whose RNA polymerase III-dependent template is contained within intron 1 of the AFP minigene (14, 47). The low level of B1 RNA in lane 2 indicates a low transfection efficiency for the construct in this experiment only. The numbers above the gel lanes refer to the numbers of the constructs below.

have indicated that the negative regulation of the AFP gene is mediated by a *cis*-acting element which is in a similar configuration, that is, between the 5' promoter and enhancers (J. Vacher and S. M. Tilghman, unpublished observations). For the H19 gene, a negative element between the promoter and enhancers would lie downstream of the transcriptional start site, rather than upstream as in the AFP gene. The fact that the H19 gene is repressed approximately 20-fold less efficiently than the AFP gene (31) suggests that the negative element may not be in an optimal position.

Our transient-expression studies have localized the liver-specific enhancer activity in the 3' domain to two separable elements, each approximately 270 to 300 bp in length. These are indistinguishable by transient-expression assay, each being capable of enhancing transcription of the H19 gene approximately 30-fold in hepatoma cells. The three AFP enhancers are also equivalent when they are assayed in a transient-expression assay (14, 15). However, when each of these was tested individually in transgenic mice, they exhibited distinct tissue preferences, with all three capable of driving high-level expression of the AFP gene in the visceral endoderm, two of them in the liver and only one of them in the gut (19). We interpreted these differences as a reflection of the diverse levels at which the endogenous AFP gene is expressed in the three tissue types. That is, the enhancers were acting additively in visceral endoderm to achieve very high levels of expression, but not in the gut, which expresses the gene at very low levels. Such an explanation cannot provide a rationale for the multiplicity of the H19 gene enhancers, since the gene is expressed at very similar levels in all fetal tissues examined (31).

The 3' domain contains at least one regulatory element which is active in visceral endoderm as well, as evidenced by its requirement for high-level expression in PC 13-derived visceral endoderm cells. It remains to be determined whether the two enhancers we have identified are also responsible for the expression of the gene in muscle cells or whether separate regulatory elements are required. In *Drosophila melanogaster*, there are several examples in which different tissue specificities are mediated by distinct sets of elements. The yolk protein gene is regulated by two separable elements in fat bodies and ovaries (11), and the *fushi tarazu* gene contains at least three kinds of upstream elements, one being responsible for the zebra stripes in embryos, one being responsible for expression during neurogenesis, and a third acting like a general enhancer (21).

At this juncture we cannot be certain that we have identified all of the regulatory elements of the H19 gene. Our survey has included only 11 kb of the 5'-flanking region and 8 kb of the 3'-flanking region. Recent data of Grosveld et al. (17) showed that widely separated *cis*-acting sequences in the human globin gene cluster (originally identified by map-

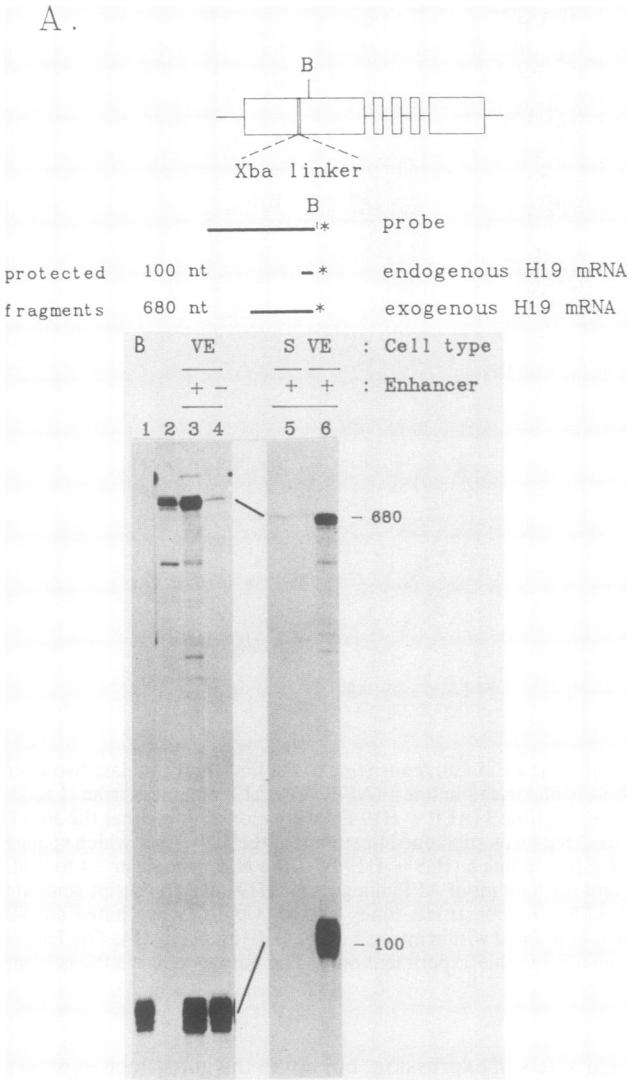

FIG. 7. The 3′ flanking domain is required for expression in visceral endoderm. (A) Two tandemly repeated *Xba*I linkers were inserted into a *Sac*I site at +580 bp in the H19 gene exon 1. When a restriction fragment derived from the marked gene extending from an *Eco*RI site at −48 bp to the *Bam*HI site at +680 bp is labeled at the 5′ phosphate of the *Bam*HI site and hybridized to authentic H19 mRNA, a 100-nt fragment is protected from S1 nuclease digestion. The mRNA derived from the gene containing the *Xba*I linkers protects a 680-nt fragment. (B) The labeled *Bam*HI-*Eco*RI restriction fragments derived from the *Xba* linker-containing marker H19 gene (lane 1) or the endogenous H19 gene (lane 2) was hybridized to 0.2 μg of poly(A)$^+$ RNA from fetal mouse liver, treated with S1 nuclease, and analyzed on an acrylamide gel. The labeled fragment from the marked gene was also hybridized to 5 μg of poly(A)$^+$ RNA from PC 13 embryonal carcinoma-derived visceral endoderm (lanes VE) cells transfected with RRSR (lane 3) or RRS (lane 4). In a separate experiment, poly(A)$^+$ RNA was isolated from PC 13 stem cells (S; lane 5) or PC 13-derived visceral endoderm cells (VE; lane 6) that had been transfected with RRSR. The fragments protected from S1 nuclease digestion were separated on acrylamide gels.

ping DNase I-hypersensitive sites in chromatin [45]) were capable of conferring position-independent expression on the β-globin gene in transgenic mice. Similar strongly DNase I-hypersensitive sites downstream of the two H19 enhancers have recently been mapped in chromatin of neonatal liver

I.

TCTAGAGTCC ATGCATCTGA GGAAAAACAG ACTTAAAACA TCAATCCGAG CCTGACCTCT
CAGGGTAGTG CTGTGAGGGA CGGGGCCTGG GGGAACAGTG GGGCCCCATG AAAGGCTGGG
AAAGACTGGT TTATTCAGGT GGCATTCAGG TTGTGCGCAT GCGTGCCAGA GCCCATGGGA
CTAAGGGAGG TGTACACAAA GCCCTCAATA AGTTCCCCAT TGGCTACTGG TGACTTCACT
CCTGTGTCAA CACCACCATG TCTGCAG

II.

GTCAACCTGG CTCTGCCCTG CATGAACCCC TCCCCTCAGA AGAGTCAGCA GTGTCCCCTG
CTAGAGGGTG AATGGCCTCC CTCTCCCTCG GGTCGTGGGG CAGAGTACAT GTGTGCCAAG
AGGGGGCCCC ATGGGCCACG TGGCTGTTTG CAGTTCATGA GAAGCAGCTG AGATAAAAAG
CTGTTTTCCC AGTGGGCAAG GGGCCAGAGG GGGCTGAGGG GGGCGTGGGA GTCAAGGCCA
GGCTGTAGGG CTTTGTGCAA CCCTGGCCTC CCCCTCCTAC GTGTTTGTAC CTGCAG

FIG. 8. Nucleotide sequences of the H19 enhancers. The nucleotide sequences of the 5′ *Xba*I-*Pst*I fragment (labeled I in Fig. 5B) and the 3′ *Hin*cII-*Pst*I fragment (labeled II in Fig. 5B) are shown. The TGTTTGCAG motif is boxed, with the imperfect copy in II indicated by a dotted box; the GCTTTGAGCAA motif is underlined; and dotted lines indicate bases in II also found in AFP enhancer I.

cells (M. Brunkow and S. M. Tilghman, unpublished observations), which leaves this possibility open.

The nucleotide sequence comparison between the two H19 gene enhancers yielded surprisingly little homology, less than was observed among the three AFP enhancers (15), despite their common biological activities in the transient-expression assay. Comparisons among the enhancers of the AFP and H19 genes revealed two regions which might serve as common regulatory motifs, the TGTTTGCAC motif (also found in the hepatitis B virus enhancer) and a GCTTTGAG CAA motif. These are both well conserved in H19 enhancer II, but only the former is present, less well conserved, in enhancer I. We had previously noted very little similarity between the promoter regions of the two genes (32).

One major difference between the AFP and H19 genes is the failure of many hepatoma cell lines to express the H19 gene. Of five lines derived from both human and rodent species tested (Hepa1-6, BW1-j, H4-II, HepG2, and Hep3B), only Hep3B cells expressed the gene (V. Pachnis, unpublished observations). The failure of HepG2 cells to transcribe the exogenous H19 gene was attributed to the inactivity of the H19 gene enhancers (Fig. 6). These data argue that the enhancers of the two genes differ in the requirement for at least one factor, in either a qualitative or a quantitative manner. Thus, perhaps the interplay of proteins required to activate the genes may be dissimilar, despite their common pattern of temporal activation.

ACKNOWLEDGMENTS

We thank B. Knowles and G. Darlington for Hep3B and HepG2 cells, B. Hogan for PC 13 cells, and R. Perry for the ribosomal protein gene probe. We also thank Noel Mann for expert assistance in the preparation of the manuscript.

This work was supported by Public Health Service grant CA44976 from the National Institutes of Health and a grant from The March of Dimes Birth Defects Foundation.

LITERATURE CITED

1. **Adamson, E. D., and B. L. M. Hogan.** 1984. Expression of EGF receptor and transferrin by F9 and PC13 teratocarcinoma cells. Differentiation **27**:152–157.
2. **Arndt, K. T., C. Styles, and G. R. Fink.** 1987. Multiple global regulators control HIS4 transcription in yeast. Science **237**:874–880.

3. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin mRNA by chromatography on oligo thymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
4. **Banerji, J., L. Olson, and W. Schaffner.** 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33**:729–740.
5. **Belayew, A., and S. M. Tilghman.** 1982. Genetic analysis of α-fetoprotein synthesis in the mouse. Mol. Cell. Biol. **2**:1427–1435.
6. **Berger, S. L., and W. R. Folk.** 1985. Differential activation of RNA polymerase III-transcribed genes by the polyomavirus enhancer and the adenovirus E1A gene products. Nucleic Acids Res. **13**:1413–1428.
7. **Bergman, Y., D. Rice, R. Grosschedl, and D. Baltimore.** 1984. Two regulatory elements for κ immunoglobulin gene expression. Proc. Natl. Acad. Sci. USA **81**:7041–7045.
8. **Choi, O., and J. D. Engel.** 1986. A 3′ enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β-globin gene. Nature (London) **323**:731–734.
9. **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.
10. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
11. **Garabedian, J. J., B. M. Shepard, and P. C. Wensink.** 1986. A tissue-specific transcription enhancer from the Drosophila yolk protein 1 gene. Cell **45**:859–867.
12. **Gillies, S. D., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33**:717–728.
13. **Giniger, E., S. M. Varnum, and M. Ptashne.** 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell **40**:767–774.
14. **Godbout, R., R. Ingram, and S. M. Tilghman.** 1986. Multiple regulatory elements in the intergenic region between the α-fetoprotein and albumin genes. Mol. Cell. Biol. **6**:477–487.
15. **Godbout, R., R. S. Ingram, and S. M. Tilghman.** 1988. Fine-structure mapping of the three mouse α-fetaprotein gene enhancers. Mol. Cell. Biol. **8**:1169–1178.
16. **Grayson, D. R., R. H. Costa, K. G. Xanthopoulos, and J. E. Darnell.** 1988. One factor recognizes the liver-specific enhancers in α_1-antitrypsin and transthyretin genes. Science **239**:786–788.
17. **Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias.** 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell **51**:975–985.
18. **Guilfoyle, R., and R. Weinmann.** 1981. Control region for adenovirus VA RNA transcription. Proc. Natl. Acad. Sci. USA **78**:3378–3382.
19. **Hammer, R. E., R. Krumlauf, S. A. Camper, R. L. Brinster, and S. M. Tilghman.** 1987. Diversity of alpha-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements. Science **235**:53–58.
20. **Hinnebusch, A. G., and G. R. Fink.** 1983. Positive regulation in the general control of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **80**:5374–5378.
21. **Hiromi, Y., and W. J. Gehring.** 1987. Regulation and function of the Drosophila segmentation gene *fushi tarazu*. Cell **50**:963–974.
22. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. **26**:365–369.
23. **Jaynes, J. B., J. E. Johnson, J. N. Buskin, C. L. Gartside, and S. D. Hauschka.** 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. Mol. Cell. Biol. **8**:62–70.
24. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MATα2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell **42**:237–247.
25. **Knowles, B., C. Howe, and D. Aden.** 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science **209**:497–499.
26. **Krimpenfort, P., R. de Jong, Y. Uematsu, Z. Dembic, S. Ryser, H. von Boehmer, M. Steinmetz, and A. Berns.** 1988. Transcription of T cell receptor β-chain genes is controlled by a downstream regulatory element. EMBO J. **7**:745–750.
27. **McKnight, S. L.** 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucleic Acids Res. **8**:5949–5964.
28. **Messing, J., B. Bronenborn, B. Muller-Hill, and P. H. Hofschneider.** 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a *HindII* fragment of the *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA **74**:3642–3646.
29. **Minty, A., and L. Kedes.** 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. **6**:2125–2136.
30. **Olsson, M., G. Lindahl, and E. Ruoslahti.** 1977. Genetic control of alphafetoprotein synthesis in the mouse. J. Exp. Med. **145**:819–827.
31. **Pachnis, V., A. Belayew, and S. M. Tilghman.** 1984. A locus unlinked to α-fetoprotein under the control of the murine *raf/Rif* genes. Proc. Natl. Acad. Sci. USA **81**:5523–5527.
32. **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.
33. **Picard, D., and W. Schaffner.** 1984. A lymphocyte-specific enhancer in the mouse immunoglobulin kappa gene. Nature (London) **307**:80–82.
34. **Pinkham, J., and L. Guarente.** 1986. Cloning and molecular analysis of the *HAP2* locus: a global regulator of respiratory genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **5**:3410–3416.
35. **Queen, S., and D. Baltimore.** 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell **33**:741–748.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
37. **Scherrer, K.** 1969. Isolation and sucrose gradient analysis of RNA, p. 413–432. *In* K. Habel and N. P. Saltzman (ed.), Fundamental techniques in virology. Academic Press, Inc., New York.
38. **Scott, R. W., and S. M. Tilghman.** 1983. Transient expression of a mouse α-fetoprotein mini-gene: deletion analyses of promoter function. Mol. Cell. Biol. **3**:1295–1309.
39. **Scott, R. W., T. F. Vogt, M. Croke, and S. M. Tilghman.** 1984. Tissue specific activation of a transfected AFP gene during F9 cell differentiation. Nature **310**:562–567.
40. **Shaul, Y., and R. Ben-Levy.** 1987. Multiple nuclear proteins in liver cells are bound to hepatitis B virus enhancer element and its upstream sequences. EMBO J. **6**:1913–1920.
41. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
42. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
43. **Trainor, C. D., S. J. Stamler, and J. D. Engel.** 1987. Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3′ enhancer. Nature (London) **328**:827–829.
44. **Trudel, M., and F. Costantini.** 1987. A 3′ enhancer contributes to the stage-specific expression of the human β-globin gene. Genes Dev. **1**:954–961.
45. **Tuan, D., W. Solomon, L. S. Qiliang, and M. L. Irving.** 1985. The "β-like-globin" gene domain in human erythroid cells. Proc. Natl. Acad. Sci. USA **32**:6384–6388.
46. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzylomethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. USA **76**:3683–3687.
47. **Young, P. R., R. W. Scott, D. H. Hamer, and S. M. Tilghman.** 1982. Construction and expression in vivo of an internally deleted mouse α-fetoprotein: presence of a transcribed *Alu*-like repeat within the first intervening sequence. Nucleic Acids Res. **10**:3099–3116.