MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4558–4569
0270-7306/86/124558-12$02.00/0

Vol. 6, No. 12

The MES-1 Murine Enhancer Element Is Closely Associated with the Heterogeneous 5′ Ends of Two Divergent Transcription Units

TREVOR J. WILLIAMS AND MIKE FRIED*

Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC1A 3PX, United Kingdom

Received 23 July 1986/Accepted 2 September 1986

The location in the mouse genome of the 149-base pair MES-1 element, previously isolated by its ability to restore expression to an enhancerless selectable gene, was analyzed. The active moiety of the single-copy MES-1 element is located between the 5′ ends of two divergent transcription units, SURF-1 and SURF-2, both of which specify more than one mRNA species by differential splicing. The heterogenous 5′ ends of the SURF transcripts are separated by only 50 to 75 base pairs, and this sequence possesses a high G+C content (65%) and contains neither the TATA and CAAT box motifs normally associated with many highly expressed genes nor the GC box motif (Sp1-binding site) associated with a number of housekeeping genes. Although MES-1 appears to have enhancerlike properties when linked to heterologous genes, its normal genomic location suggests that it functions as a bidirectional promoter. Thus, MES-1 may represent a new class of enhancer-promoter element.

Several sequence elements upstream from a large number of eucaryotic genes transcribed by RNA polymerase II have been identified as important for normal gene expression. The TATA box sequence located approximately 30 bases upstream from the cap site (−30) is necessary for precise initiation of transcription, as deletion of this sequence results in transcription starting at heterogeneous sites (1). More distal promoter elements are involved in the modulation of the frequency of transcriptional initiation. One such sequence, the CAAT box, has been identified in several genes at about −80, and at least for the globin genes is important for efficient expression (17). In addition, regulatory sequences involved in either the tissue-specific expression of certain genes or the ability of particular genes to respond to environmental stimuli are often found in the distal promoter region (for a review, see reference 14). A further class of DNA sequence elements involved in *cis* activation of transcription are the enhancers which are defined by their ability to increase expression from the normal cap site of a gene in a relatively position- and orientation-independent manner. Enhancer elements were initially identified in viruses but have subsequently been found in a number of cellular genes including the immunoglobulin gene in which the enhancer occurs in an intron downstream of the promoter after heavy-chain rearrangement (for reviews, see references 14 and 43). Recently some classically defined promoter elements have been shown to display some of the properties of enhancers, and so the distinction between promoters and enhancers has become less rigid (15, 44).

Most of our understanding of the organization of cellular promoters and enhancers has come from the study of highly expressed tissue-specific genes such as those for immunoglobulin and globin. The recent isolation of genes which are usually constitutively expressed at low levels in all cell types, housekeeping genes, has led to the discovery of a different kind of promoter structure (5, 27, 30, 34, 46, 51). Promoters for housekeeping genes are typified by their high G+C content, the absence of TATA and CAAT box motifs, and the presence of short direct repeat sequences and potential binding sites for the transcription factor Sp1 (for a review, see reference 24).

* Corresponding author.

The various sequence elements involved in the control of transcription have been identified owing to their association with cloned genes. We used an alternative approach to isolate such sequences by their ability to reactivate a test gene lacking its own expression sequences (9). By this technique, a 149-base-pair (bp) fragment of mouse DNA was isolated (originally termed H2 but now renamed MES-1, mouse expression sequence-1) which displays some of the properties of an enhancer (9, 52). Here we report a detailed transcriptional analysis of the normal genomic location of MES-1 which indicates that it lies between two divergent transcription units. Sequence analysis of this potential bidirectional promoter reveals that although it has similarities with the promoters of housekeeping genes, it may represent a new class of expression element.

MATERIALS AND METHODS

Genomic cloning. Total BALB/c 3T3 DNA (500 μg) was digested to completion with *Eco*RI and fractionated on a 10 to 40% sucrose gradient. Fractions possessing the 19-kilobase (kb) fragment containing MES-1 were identified by Southern blotting analysis with the plasmid pB1 (9) as a probe. Positive fractions were pooled, precipitated, and ligated into the *Eco*RI site of the λ vector EMBL4 (10). Approximately 5×10^4 recombinant clones were screened with pB1 as a probe, and two positive plaques were isolated which contained identical inserts by restriction enzyme analysis. Appropriate restriction fragments from this 19-kb *Eco*RI clone, including the 0.8- and 1.0-kb *Bam*HI fragments, were then further subcloned into the plasmid pxf3 (19).

Southern and Northern blotting analysis. For Southern blotting, 16 μg of BALB/c 3T3 DNA was digested with either *Eco*RI or *Hin*dIII, fractionated by electrophoresis on a 0.8% agarose gel, and transferred to nitrocellulose as described previously (20). For Northern blotting, total cellular RNA was prepared from BALB/c TS-A-3T3 cells (38) by the guanidinium-CsCl method and poly(A) selected on oligo(dT)-cellulose (25). A 5-μg sample of poly(A)$^+$ RNA was fractionated on a 1% agarose-formaldehyde gel and transferred to nitrocellulose after partial alkaline hydrolysis (25). Hybridization conditions were identical for Southern and Northern blotting, and both types of blot were washed to

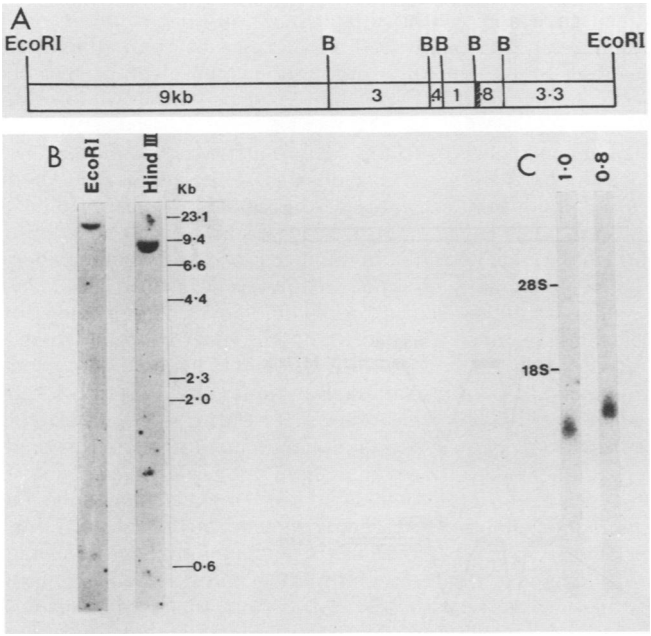

FIG. 1. Analysis of the normal genomic location of MES-1. (A) Restriction map of the 19-kb *Eco*RI fragment which contains MES-1 (shaded). The restriction sites for *Eco*RI and *Bam*HI (B) are shown. The sizes in kilobases of the fragments generated by *Bam*HI restriction of the 19-kb fragment are indicated. (B) Detection of a single copy of MES-1 in the mouse genome. A Southern blot of mouse genomic DNA digested with either *Eco*RI or *Hin*dIII was hybridized to the 149-bp MES-1 probe. The numbers on the right show the size of the markers. (C) Northern blotting analysis of mouse BALB/c TS-A-3T3 poly(A)$^+$ RNA with either the 1.0-kb *Bam*HI fragment (lane 1.0) or the 0.8-kb *Bam*HI fragment (lane 0.8) as probes (see panel A). On the left are shown the positions of the 28S and 18S rRNA markers.

a stringency of 0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 68°C essentially as described previously (20). Probes were labeled to high specific activity by nick translation (26).

S1 mapping and primer extension analysis. Probe B used for S1 mapping analysis was 5′ end labeled at the *Hin*fI site at nucleotide (nt) R482 with polynucleotide kinase and [γ-^{32}P]ATP as described previously (26), recut with *Bam*HI, and purified by fractionation on an agarose gel. Probe B was annealed with either 40 μg of BALB/c TS-A-3T3 total RNA or 40 μg of yeast RNA. Procedures for RNA-DNA hybridization and S1 nuclease digestion were essentially as described previously (8). Primers X, Y (see Fig. 4 and 7), and Z (see below) were 5′ end labeled as described above. Primers X and Z were strand separated, after denaturation, on a cracking polyacrylamide gel (26), and primer Y was strand separated on a denaturing polyacrylamide gel. After electroelution, the strand from each primer which was complementary to the SURF transcripts was identified by sequence analysis (see below). Primers were annealed to either 1 μg of BALB/c TS-A-3T3 poly(A)$^+$ RNA or 1 μg of yeast RNA as described previously (8), and first-strand cDNA synthesis was performed (18). The products of the S1 mapping and primer extension reactions were visualized by autoradiography after fractionation on denaturing polyacrylamide gels.

cDNA cloning. The isolation of cDNA T3C from a BALB/c TS-A-3T3 cDNA library has been described previously (54). In addition, cDNAs T1B, T2A, and IDE were isolated from this library owing to their homology to the 1.0-kb *Bam*HI fragment. A second cDNA library was constructed with a single-stranded primer, Z, which was complementary to SURF-2 transcripts and extended from the *Pvu*II site at nt R3433 to the *Hin*fI site at nt R3590. The primer was annealed to BALB/c TS-A-3T3 poly(A)$^+$ RNA, and the cDNA library was generated as described previously (54). cDNAs H1 and J3 were isolated from this library owing to their homology to the 0.8-kb *Bam*HI fragment. All cDNAs were then further subcloned into the *Eco*RI site of the plasmid pAT153.

Sequencing. The sequence of the cDNA clones and the genomic sequence were obtained for both strands by the dideoxy method (40) after appropriate restriction fragments had been subcloned into either M13mp18 or M13mp19 (55). End-labeled DNA was sequenced as described previously (26).

RESULTS

Isolation of mouse genomic clone containing MES-1. The 149-bp MES-1 element was initially identified owing to its ability to restore expression to a test gene lacking its 5′ regulatory sequences (9). More recently, MES-1 has been found to display enhancerlike activity in both long- and short-term expression assays (52). To characterize the normal status of this element in the mouse genome, MES-1 was used to probe Southern blots of mouse DNA and was found to occur as a single-copy (Fig. 1B). The 19-kb *Eco*RI fragment containing MES-1 (Fig. 1B) was subsequently cloned from mouse cellular DNA. The restriction map of this 19-kb clone, in which MES-1 (indicated by shading) resides in a 0.8-kb *Bam*HI fragment, is shown in Fig. 1A. The sequence of this copy of MES-1, isolated from its normal location in the mouse genome, was found to be identical to that of the copy of MES-1 isolated in the initial experiment (9). This indicated that the MES-1 element had not been fortuitously generated during the transfection process but represents a naturally occurring sequence in the mouse genome.

Transcription units are present both upstream and downstream of MES-1. The 1.0-kb *Bam*HI fragment upstream of MES-1 and the 0.8-kb *Bam*HI fragment containing MES-1 and its 3′ sequences were used to probe Southern blots of mouse DNA, and both of these fragments were found to exist as single-copy sequence (data not shown). Therefore, to assess whether MES-1 was associated with any transcription units in its normal genomic location, we used these two unique sequence fragments to probe Northern blots of mouse RNA. The 1.0-kb *Bam*HI fragment identified poly(A)$^+$ RNA species of about 1.1 kb (Fig. 1C), termed the SURF-1 transcripts, which extended into the adjacent 0.4- and 3.0-kb *Bam*HI fragments (Fig. 1A). In addition, the 0.8-kb *Bam*HI fragment hybridized to poly(A)$^+$ RNA species of about 1.2 kb (Fig. 1C), termed the SURF-2 transcripts, which extended into the adjacent 3.3-kb *Bam*HI-*Eco*RI fragment (Fig. 1A). To analyze the relationship between MES-1 and these two transcription units in more detail, we isolated cDNA clones corresponding to either SURF-1 or SURF-2 from mouse cell cDNA libraries.

Analysis of the transcription unit SURF-2. Three cDNA clones of approximately 1 kb corresponding to SURF-2 transcripts were isolated owing to their homology to the 0.8-kb *Bam*HI fragment. Subsequent sequence analysis showed that, in the region where the cDNA clones overlapped, two of them (H1 and T3C) were identical, while the

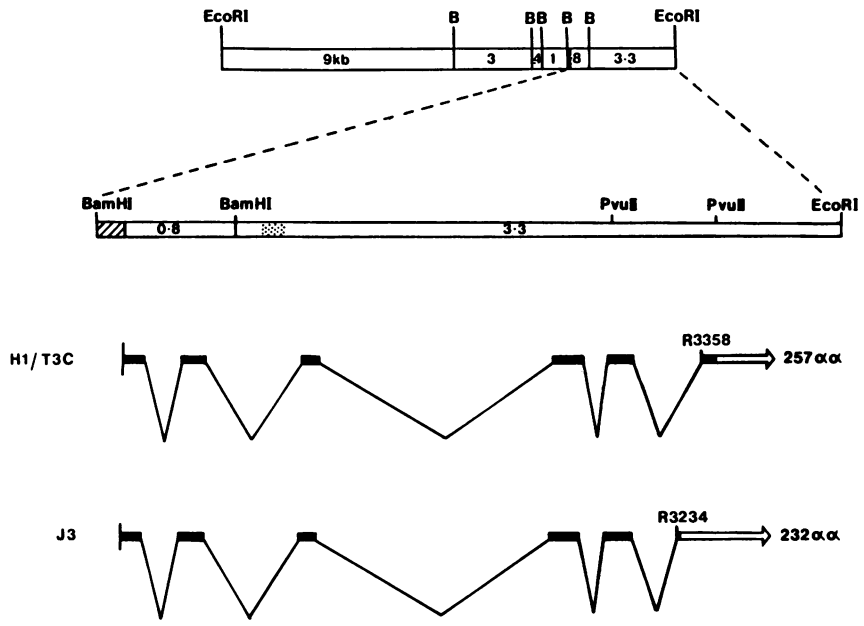

FIG. 2. Organization of the SURF-2 transcription unit. At the top is shown the structural map of the 19-kb *Eco*RI fragment which contains MES-1 (as in Fig. 1A), and beneath this is an expanded map of the 0.8- and 3.3-kb restriction fragments. The restriction sites for *Bam*HI, *Eco*RI, and *Pvu*II are shown. In addition, the positions of the MES-1 (diagonal hatching) and a mouse B1 repetitive sequence element (dotted shading) are indicated. At the bottom of the figure are shown the structures of the cDNAs H1, T3C, and J3. In both cases the vertical bar represents the 5′ end of the cDNAs and the arrow signifies both the direction of transcription and the 3′ end of the mRNAs. Boxes represent exon sequences, and the positions of the long open reading frame (closed box) and 3′-untranslated regions (open box) are indicated. Sloping lines designate the positions of intron sequences. The position of the splice acceptor site for exon 6, which differs between the two classes of mRNA, is shown by a number above the final exon (numbers refer to the distance to the right of the *Hae*II site at the 3′ end of MES-1). The size of the predicted protein products in amino acids (αα) is shown at the 3′ end of the cDNAs.

third (J3) contained an additional 124 bp of sequence located approximately in the middle of the cloned insert. To examine the relationship of these cDNAs to MES-1 and to determine whether the extra sequence present in clone J3 was due to the presence of an alternative additional exon, the sequences of the three cDNA clones were compared with the sequence obtained from this region of the 19-kb genomic clone. A diagrammatic representation of the structures of these two classes of cDNA clones (H1/T3C and J3) compared with the genomic map is presented in Fig. 2, in which the open reading frames are shown by closed boxes and the 3′-untranslated regions are shown by open boxes. Figure 3 shows the sequences of the two classes of cDNA compared with the genomic sequence to illustrate the intron-exon boundaries. Beneath the DNA sequence is shown the predicted amino acid sequence for the longest open reading frame for both classes of cDNA.

The cDNA clones corresponding to SURF-2 were found to start just downstream of MES-1 and to terminate at a poly(A) addition site at nt R3803. (Nucleotide numbers beginning with R refer to the distance to the right of the *Hae*II site at the 3′ end of MES-1, while those beginning with L refer to the distance to the left of this site.) In addition, all three cDNA clones were found to consist of six exons, sharing the same intron-exon boundaries for the first five exons. However, the cDNAs differ in the choice of splice acceptor for exon 6, and this accounts for the additional 124 bp in clone J3. This clone uses a splice acceptor site at nt R3234 and has a 570-bp final exon, while clones H1 and T3C use a splice acceptor site at nt R3358 and have a 446-bp final exon (Fig. 2 and 3). The two potential splice acceptors for exon 6 both fit the consensus splice acceptor sequence $(^{T}_{C})_{11}N^{T}_{C}AG/G$ (32). In addition, nuclease S1 mapping confirmed that both splice acceptor sites were used in vivo, although 95% of the mRNA used the site at nt R3358 (H1/T3C), while only 5% used the site at nt R3234 (J3) (data not shown). The relatively minor usage of the splice acceptor at nt R3234 may explain why no mRNA of 1.3 kb, corresponding to J3, was seen in Fig. 1C.

The first five exons, common to both classes of cDNA, contain a long open reading frame beginning at nt R51 and extending into exon 6. However, the different splice acceptor sites used for exon 6 by J3 and H1/T3C alter the potential coding capacity of their respective mRNAs. In the case of mRNA corresponding to J3 there is an in-frame stop codon shortly after the splice acceptor site at nt R3234, and so the final exon only encodes an additional three amino acids, generating a protein of 232 amino acids in total (Fig. 2 and 3). In contrast, for mRNAs corresponding to H1 and T3C, exon 6 encodes an additional 28 amino acids, generating a protein of 257 amino acids in total (Fig. 2 and 3). Thus, the potential protein products encoded by the two classes of mRNA would have an identical N-terminal 229 amino acids but would differ in length and composition at their C termini.

Determination of 5′ ends of SURF-2 transcripts. The relationship between the 5′ start sites of the SURF-2 transcripts and MES-1 was determined more precisely by the complementary techniques of primer extension and 5′ S1 nuclease analysis. To perform the 5′ S1 nuclease mapping of cap sites, it is usually necessary to use as a probe a genomic fragment which is kinased labeled in the first exon of the transcript and extends into the sequences upstream from the start sites. However, in the case of SURF-2 there were no convenient restriction sites in the first exon to enable this approach to be

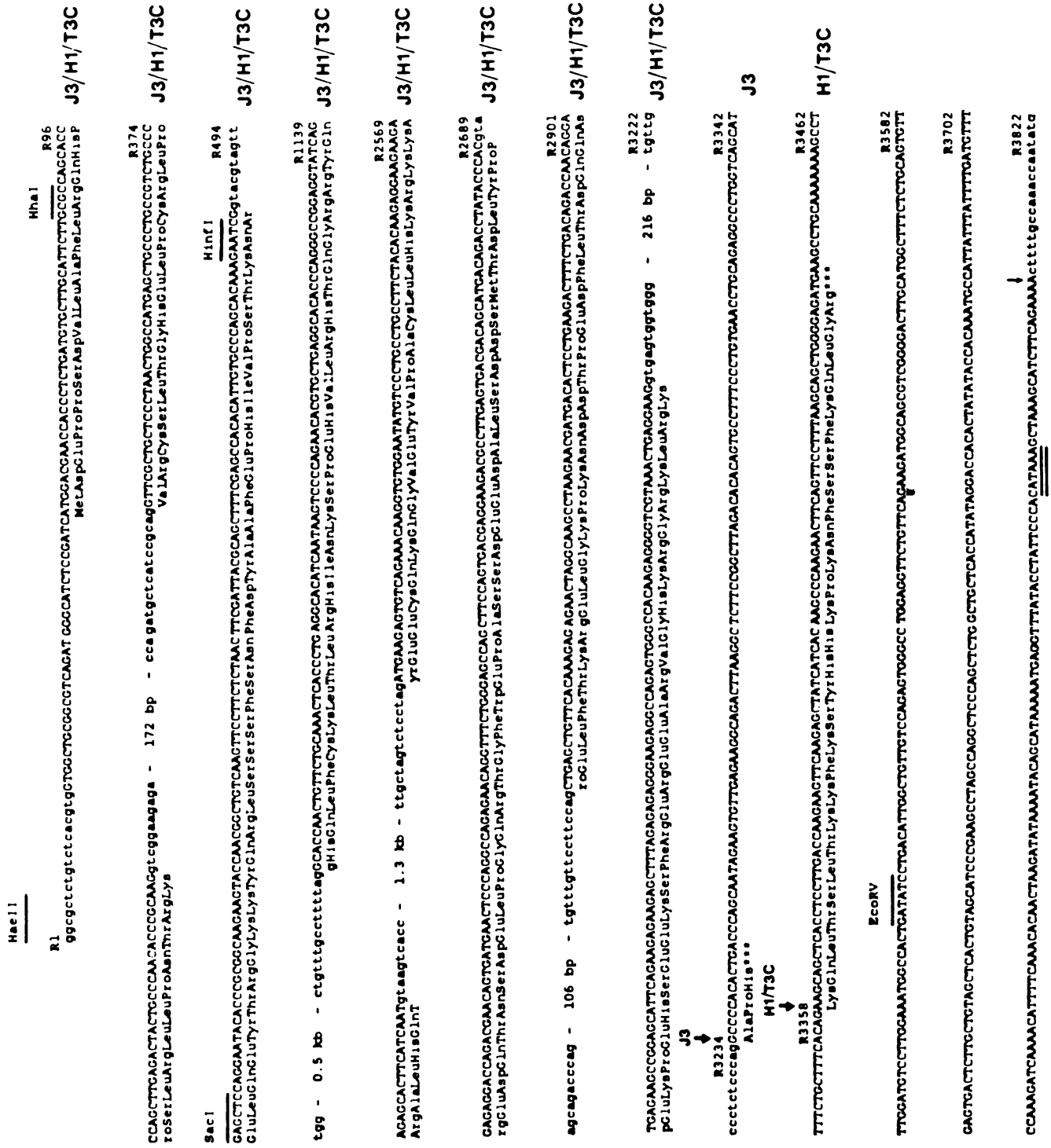

FIG. 3. Comparison of the sequence of the cDNAs H1/T3C and J3 with the genomic sequence of the SURF-2 transcription unit. The sequence is presented from the *Hae*II site at the 3′ end of MES-1 (nt R1) to a position 3′ of the poly(A) addition site of the transcription unit. Nucleotide numbers above the sequence refer to their positions in the genomic fragment. The locations of restriction sites referred to in the text are shown by a single line above the DNA sequence. Uppercase letters refer to sequences present in the cDNAs, while lowercase letters refer to 5′-flanking, 3′-flanking, or intronic sequences. The position and size of the introns are indicated. The putative poly(A) addition signal is denoted by double underlining, and the poly(A) addition site is shown by a small vertical arrow. The characterization of the 3′ end of the SURF-2 transcription unit has been reported previously in reference to an overlapping third transcription unit (54) (see Fig. 8A). The positions of the two different splice acceptors for exon 6 are indicated by the large vertical arrows (J3 and H1/T3C) above the DNA sequence. The 124 nts which represent part of an exon for cDNA J3 and part of an intron for H1/T3C (between nts R3234 and R3357) are denoted by italicized uppercase letters. The predicted amino acid sequence of the long open reading frame for both classes of cDNA is shown beneath the DNA sequence. The notation J3/H1/T3C in the right margin indicates that this amino acid sequence is common to both classes of cDNA. J3, however, refers to the amino acid sequence that would only be encoded by mRNAs which use the splice acceptor site at nt R3234, i.e., corresponding to cDNA J3. Similarly, H1/T3C refers to the amino acid sequence that would only be encoded by mRNAs which use the splice acceptor site at nt R3358, i.e., corresponding to cDNAs H1 and T3C.

used. Therefore, to facilitate the S1 nuclease mapping, we constructed the clone pHHA (Fig. 4). This clone consists of the cDNA sequence of clone T3C from its 3′ terminus, corresponding to nt R3803, to the *Hha*I site at nt R86 in exon 1 ligated to the genomic sequence of the 0.8-kb *Bam*HI fragment from this *Hha*I site to the *Bam*HI site at nt L139. This enabled a 420-nt probe, kinase labeled at the *Hin*fI site in exon 2 (corresponding to nt R482) and extending to the *Bam*HI site at nt L139 (Fig. 4, probe B), to be contiguous with SURF-2 transcripts into the first exon and then continue upstream from the presumptive cap sites into the genomic sequence. The probe was annealed to mouse RNA, and S1 nuclease-resistant products were fractionated on a polyacrylamide gel (Fig. 4, lane 2). In addition, a 195-nt single-stranded primer complementary to SURF-2 transcripts was obtained from the cDNA clone H1 for primer extension analysis (Fig. 4, primer X). The primer was kinase labeled at the same position as probe B to allow a direct comparison with the S1 nuclease mapping data (i.e., the *Hin*fI site corresponding to nt R482) and extended to the *Hha*I site at nt R86. The primer was annealed to mouse poly(A)$^+$ RNA and elongated by reverse transcriptase. The extended products were visualized by autoradiography after fractionation on a polyacrylamide gel (Fig. 4, lane 4). Both S1 mapping and primer extension analyses generated multiple bands ranging in size from 240 to 300 nts, with an accuracy of 2 to 3 nts, the most abundant products being bands of 248 nts and 267 to 273 nts. This indicated that SURF-2 transcripts were initiated from heterogeneous sites located between nts L16 and R43. However, within this range the majority of the transcripts start around nt R34 (70%) and further upstream at nts R9 to R15 (25%). An examination of the sequences upstream from these SURF-2 initiation sites revealed no obvious transcriptional control elements such as TATA or CAAT box motifs (Fig. 8).

Analysis of transcription unit SURF-1. A similar analysis was also performed on the SURF-1 transcription unit. Three cDNA clones of approximately 1.0 kb were isolated owing to their homology to the 1.0-kb *Bam*HI fragment, and all were found to be slightly different upon subsequent sequence analysis. Figure 5 shows a diagrammatic representation of the structure of these three SURF-1 cDNA clones in comparison with the genomic map. Figure 6 shows the sequence of these cDNAs compared with the genomic sequence to illustrate their intron-exon boundaries. Beneath the DNA sequence of the three cDNAs is shown the predicted amino acid sequence for their respective open reading frames. Figure 6 also shows the sequence of intron 1 which contains two copies of a 47-bp direct repeat sequence sharing 41 of 47 bp in common. In addition, a third related copy of this direct repeat sequence which is shorter and less homologous is also present at a site further upstream in this intron where it spans the *Bam*HI site.

All three SURF-1 cDNA clones begin within MES-1 and are derived from mRNAs transribed in the opposite direction to the SURF-2 mRNAs. In addition, all three cDNA clones are composed of nine exons and terminate at a poly(A) addition site at nt L3043. The intron-exon structure of the three cDNA clones is identical for exons 1, 2, and 4 to 9 and differs only in the choice of splice acceptor for exon 3 (Fig. 5 and 6). For the three cDNAs three different splice acceptor sites are used: at nt L604 for clone T2A, at nt L624 for clone T1B, and at nt L644 for clone IDE. The three potential splice acceptor sites for exon 3 all fit the consensus splice acceptor sequence $(^{T}_{C})_{11}N^{T}_{C}AG/G$ (32). In addition, nuclease S1 mapping confirmed that all three splice acceptor sites were used in vivo, although to various extents. Approximately 72% of the SURF-1 transcripts appear to use the splice acceptor site at nt L644 (IDE), less than 1% use the site at nt L624 (T1B), and the remaining 28% use the site at nt L604 (T2A) (data not shown).

The use of these three potential splice acceptors causes the open reading frame encoded in the first two exons to splice into the three different reading frames in exon 3 and so give rise to three potentially different proteins which share a common N-terminal 42 amino acids (Fig. 5 and 6). For mRNA corresponding to T2A which uses the earliest splice acceptor site at nt L604, exon 3 only encodes an additional 17 amino acids before an in-frame stop codon is reached. Similarly, for the rare mRNA equivalent to T1B, which use the middle splice acceptor at nt L624, exon 3 only encodes an additional two amino acids. Therefore, the mRNAs corresponding to T2A and T1B are only capable of generating short peptides of 59 and 44 amino acids, respectively. However, the majority of the mRNA, corresponding to IDE, uses the latter splice acceptor site at nt L644, and this generates a long open reading frame extending from exons 1 to 9 capable of encoding 306 amino acids in total.

Determination of 5′ ends of SURF-1 transcripts. The absence of suitable restriction sites in exon 1 precluded the analysis of the transcriptional start sites of SURF-1 by 5′ S1 nuclease mapping. However, when the most 5′ sequences of the longest cDNA clone T1B were used for 5′ S1 nuclease mapping, only products corresponding to full-length protection of the SURF-1 sequences by mRNA were obtained (data not shown). This indicated that SURF-1 transcripts were initiated further upstream than nt L49. Therefore the positions of these upstream start sites and their relationship to MES-1 were determined more accurately by primer extension analysis. A 52-nt single-stranded primer complementary to SURF-1 transcripts was obtained from the cDNA clone T1B (primer Y, Fig. 7). The primer was kinase labeled at the *Dde*I site (corresponding to nt L420) located just 5′ to the splice donor site in exon 2 and extended to an *Apa*I site formed by the splicing event joining exons 1 and 2 together (Fig. 7). Therefore, this primer essentially consists of the exon 2 sequence of SURF-1 transcripts which is common to all three species of cDNA. The primer was annealed to mouse poly(A)$^+$ RNA and extended with reverse transcriptase. The sizes of the resulting products were determined by polyacrylamide gel electrophoresis, followed by autoradiography (Fig. 7). This experiment generated a predominant extended product of 130 nts, although several minor bands of 133 nts and 136 to 137 nts were also obtained, which suggests that the initiation sites for SURF-1 transcripts are microheterogeneous. These results also indicate that 95% of the transcripts start at nt L41, while the remaining 5% initiate further upstream at nts L34 to L38.

Thus, the initiation sites for SURF-1 mRNAs are in close proximity to those for SURF-2 mRNAs but on the opposite strand (Fig. 8). Therefore, SURF-1 and SURF-2 mRNAs are transcribed in opposite directions away from a common upstream region centered on MES-1, which lacks obvious transcriptional control elements, such as TATA or CAAT box motifs, on either strand.

DISCUSSION

The 149-bp MES-1 element was initially isolated owing to its ability to reactivate a selectable gene devoid of its own expression sequences (9). More recently, MES-1 has been found to display enhancerlike activity, i.e., when placed in

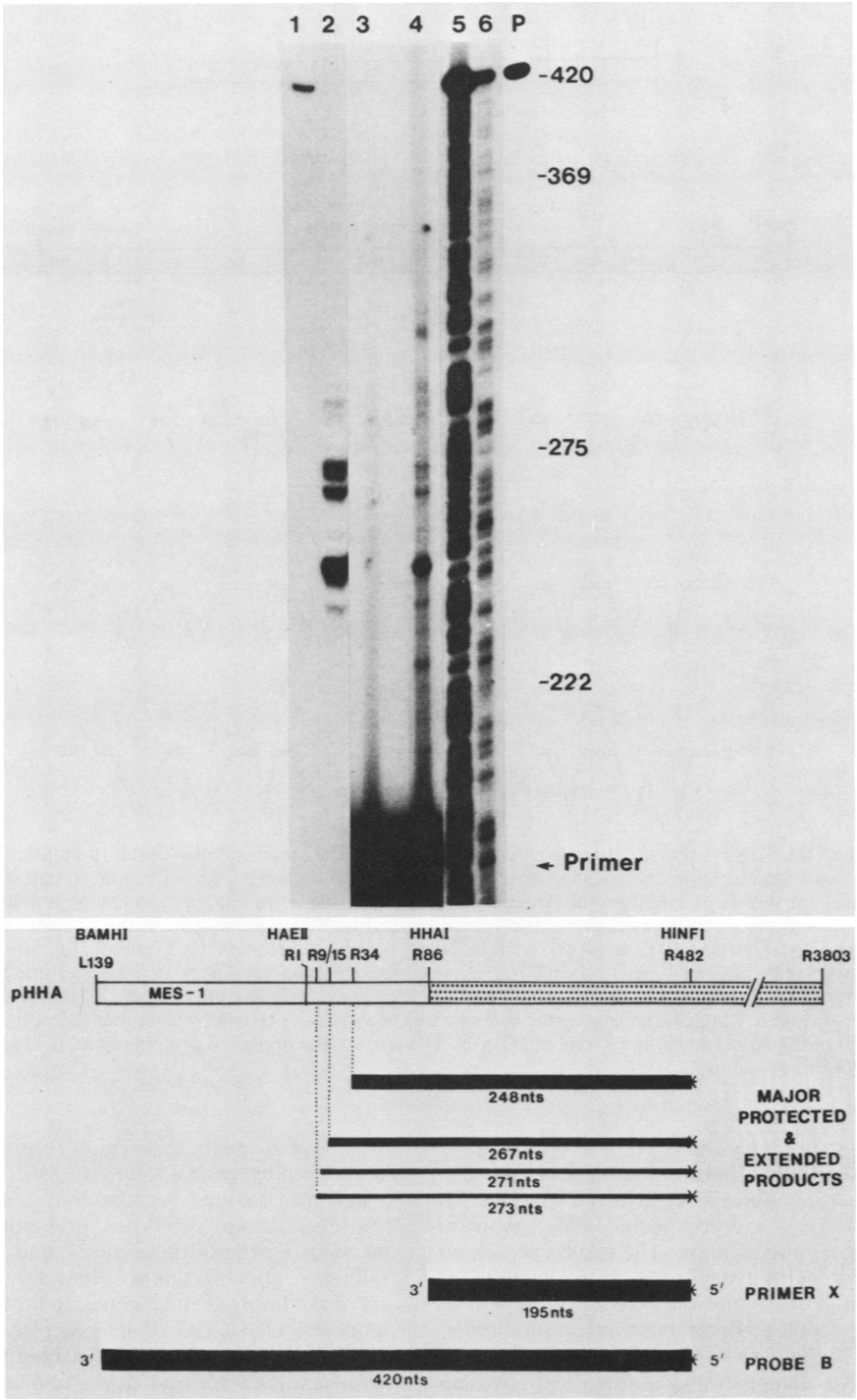

FIG. 4. Determination of the 5′ ends of the SURF-2 transcripts by 5′ S1 nuclease mapping and primer extension. On the top is presented a gel showing the results of the 5′ S1 nuclease mapping and primer extension experiments with probe B and primer X, respectively. At the bottom is shown a diagrammatic representation of probe B and primer X and the major products of the S1 nuclease mapping and primer extension experiments in terms of the structural map of the pHHA construct. The pHHA plasmid consists of the sequence of cDNA T3C (Fig. 3) from the *Hha*I site at nt R86 to the poly(A) addition site at the 3′ end of the clone at nt R3803 (stippled box), joined to the genomic sequence of the 0.8-kb *Bam*HI fragment from nt L139 to nt R86 (open box) (see text). The position of the 149-bp MES-1 element is indicated. The 420-nt probe B and 195-nt primer X are illustrated by the shaded boxes at the bottom. Both probe B and primer X are 5′ labeled at the *Hin*fI site at nt R482, shown by an *. The major protected and extended fragments produced by probe B and primer X, respectively, are denoted by thick lines between the pHHA construct and primer X. The thickness of the line represents the relative abundance of the product. Lanes: 1, probe B plus yeast RNA; 2, probe B plus BALB/c TS-A-3T3 RNA; 3, primer X plus yeast RNA; 4, primer X plus BALB/c TS-A-3T3 RNA; 5 and 6, G+A and C+T Maxam-Gilbert sequence of probe B, respectively; P, untreated probe B. On the right of the gel is shown the position of primer X and marker DNA sizes in base pairs.

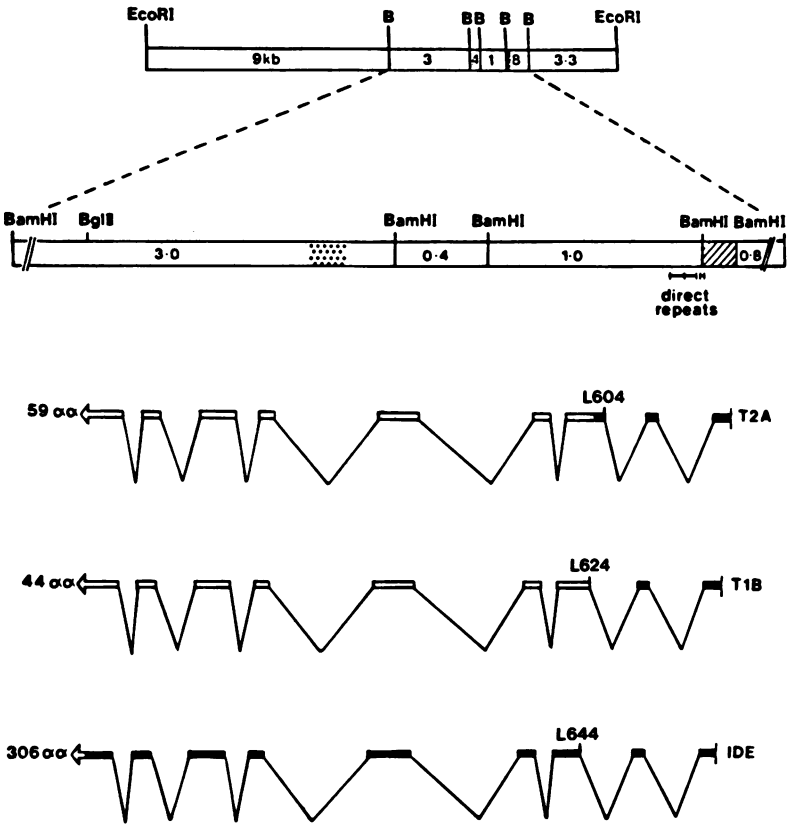

FIG. 5. Organization of the SURF-1 transcription unit. At the top is shown the structural map of the 19-kb *Eco*RI fragment which contains MES-1 (as in Fig. 1A), and beneath this is an expanded map of the 0.8-, 1.0-, 0.4-, and 3.0-kb restriction fragments. The restriction sites for *Bam*HI, *Eco*RI, and *Bgl*II are shown. In addition, the positions of the MES-1 sequence (diagonal hatching), a mouse B2 repetitive sequence element (dotted shading), and the direct repeats present in intron 1 (see Fig. 6) are indicated. At the bottom of the figure are shown the structures of the cDNAs T2A, T1B, and IDE. In all three cases the vertical bar represents the 5′ end of the cDNAs and the arrow signifies both the direction of transcription and the 3′ end of the mRNA. Boxes represent exon sequences, and the positions of the open reading frames (closed box) and 3′-untranslated regions (open box) are indicated. Sloping lines designate the positions of intron sequences. The position of the splice acceptor site for exon 3, which differs among the three classes of mRNA, is shown by a number above the final exon (numbers refer to the distance to the left of the *Hae*II site at the 3′ end of MES-1). The sizes of the predicted protein products in amino acids (αα) are shown at the 3′ end of the cDNAs.

either orientation at various positions 5′ to a heterologous gene, MES-1 activates transcription initiating from the cap site of the test gene (52). Therefore, it was of interest to determine whether MES-1 was associated with any transcripts in its normal genomic location. The results presented here demonstrate that MES-1 occurs as a single-copy sequence in the mouse genome and that the active region of MES-1 is located between two divergent transcription units, SURF-1 and SURF-2 (Fig. 8).

Recent results have demonstrated a functional overlap between enhancer and promoter elements for the control sequences present upstream of the interferon and metallothionein genes (15, 44). Similarly, although MES-1 can behave as an enhancerlike element when linked to a heterologous gene, its normal genomic position suggests that it functions as a bidirectional promoter. SURF-1 transcripts initiate at one major site and several minor sites within MES-1, while the presumptive cap sites for SURF-2 transcripts are dispersed over a 60-bp region both within and to the right of MES-1 (Fig. 8B). There are no obvious TATA or CAAT box motifs within the 5′-flanking region of either transcript; instead, the region has a high G+C content (65%) and contains a 10-bp direct repeat sequence (9) (Fig. 8B). In the above respects this region resembles the promoter of a housekeeping gene (5, 27, 30, 34, 46, 51). Housekeeping genes are so called because they tend to be expressed at a low level in many cell types. Preliminary results suggest that the same is true of the SURF-1 and SURF-2 transcripts (T. Williams, unpublished observation). However, the promoters of the housekeeping genes so far analyzed all contain the sequence GGGCGG or its complement CCGCCC, termed the GC box, which acts as a binding site for the cellular *trans*-acting factor Sp1 (for a review, see reference 24). In contrast, the region described here lacks both the GC boxes typical of the promoters of housekeeping genes and the alternative Sp1-binding sites identified in the acquired immune deficiency virus long terminal repeat (22) and so may represent a different class of promoter.

It might be expected that bidirectional promoters would need a complex organization of various sequence elements to enable RNA polymerase II to initiate transcription in either direction. However, it should be noted that many promoter and enhancer elements are capable of functioning in either orientation (for a review, see reference 43; also see references 6, 16, 23, 28). Indeed, it has been suggested that the function of the TATA box is to determine both the

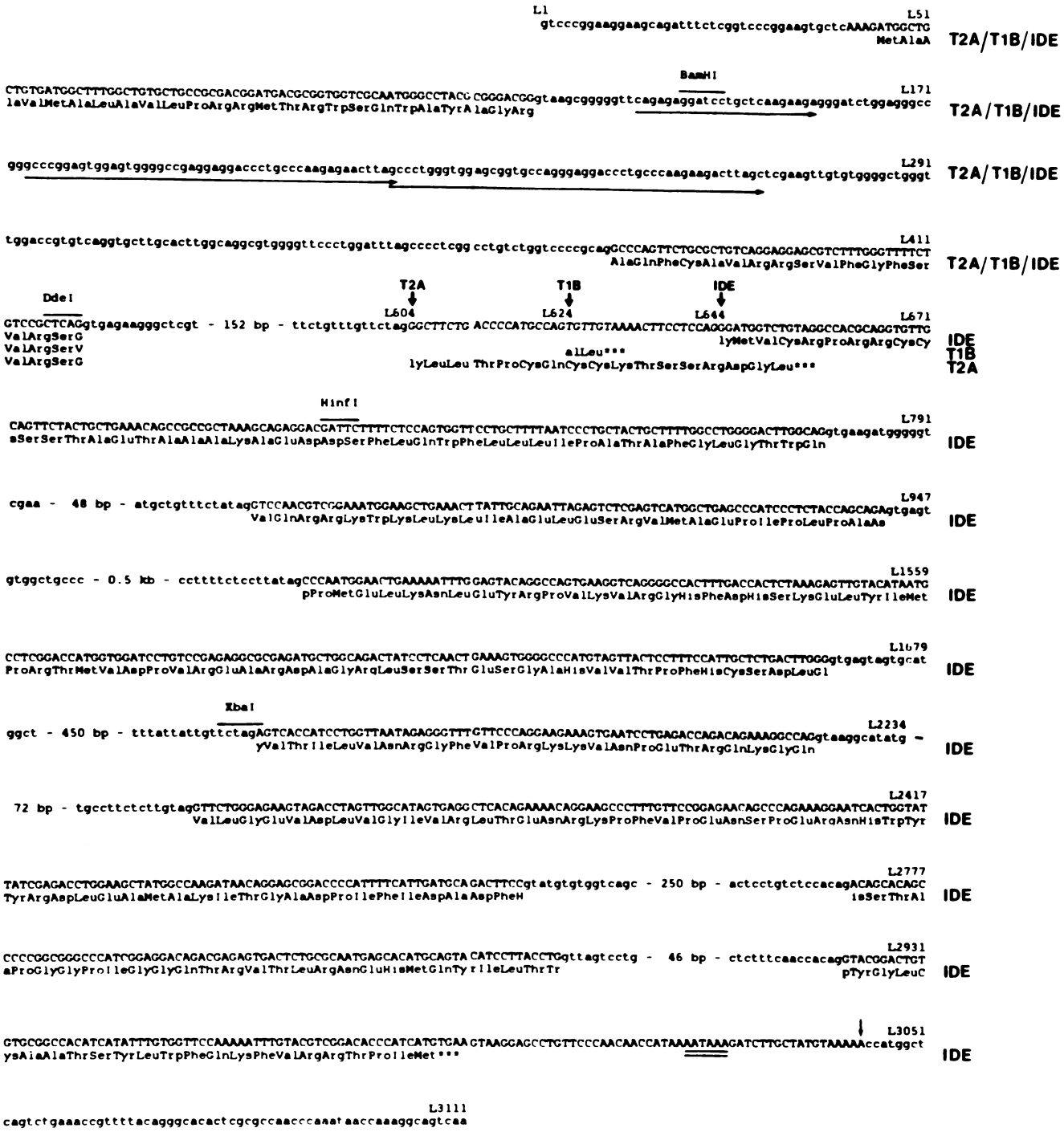

FIG. 6. Comparison of the sequences of the cDNAs T2A, T1B, and IDE with the genomic sequence of the SURF-1 transcription unit. The sequence is presented from the first nucleotide to the left of the *Hae*II site at the 3′ end of MES-1 (nt L1) to a position 3′ of the poly(A) addition site of the transcription unit. Nucleotide numbers above the sequence refer to their positions in the genomic fragment. The locations of restriction sites referred to in the text are shown by a single line above the DNA sequence. Uppercase letters refer to sequences present in the cDNAs (see below), while lowercase letters refer to 5′-flanking, 3′-flanking, or intronic sequences. The position and size of the introns are indicated, and the location of the direct repeats present within intron 1 are shown by arrows beneath the sequence. The putative poly(A) addition signal is denoted by double underlining, and the poly(A) addition site is shown by a small vertical arrow. The positions of the three different splice acceptors for exon 3 are indicated by the large vertical arrows (T2A, T1B, and IDE) above the DNA sequence. The 40 nts which represent part of an exon for cDNA T2A and part of an intron for IDE and partly intron and exon sequence for T1B (between nts L604 and L643) are denoted by italicized uppercase letters. The predicted amino acid sequence of the long open reading frame for all three classes of cDNA is shown beneath the DNA sequence and, with reference to the primer extension data presented below in Fig. 7, is presented from the ATG codon at nt L45. (The most 5′ cDNA clone only extends to nt L49.) The notation T2A/T1B/IDE in the right margin indicates that this amino acid sequence is common to all three classes of cDNA. IDE, however, refers to the amino acid sequence that would only be encoded by mRNAs which use the splice acceptor site at nt L644, i.e., corresponding to cDNA IDE. Similarly, T1B and T2A refer to the amino acid sequence that would only be encoded by mRNAs which use the splice acceptor sites at nts L624 and L604, respectively. Note that the sequence of MES-1 at nts L70 to L71 has been amended to rectify an error in the previously published sequence (9).

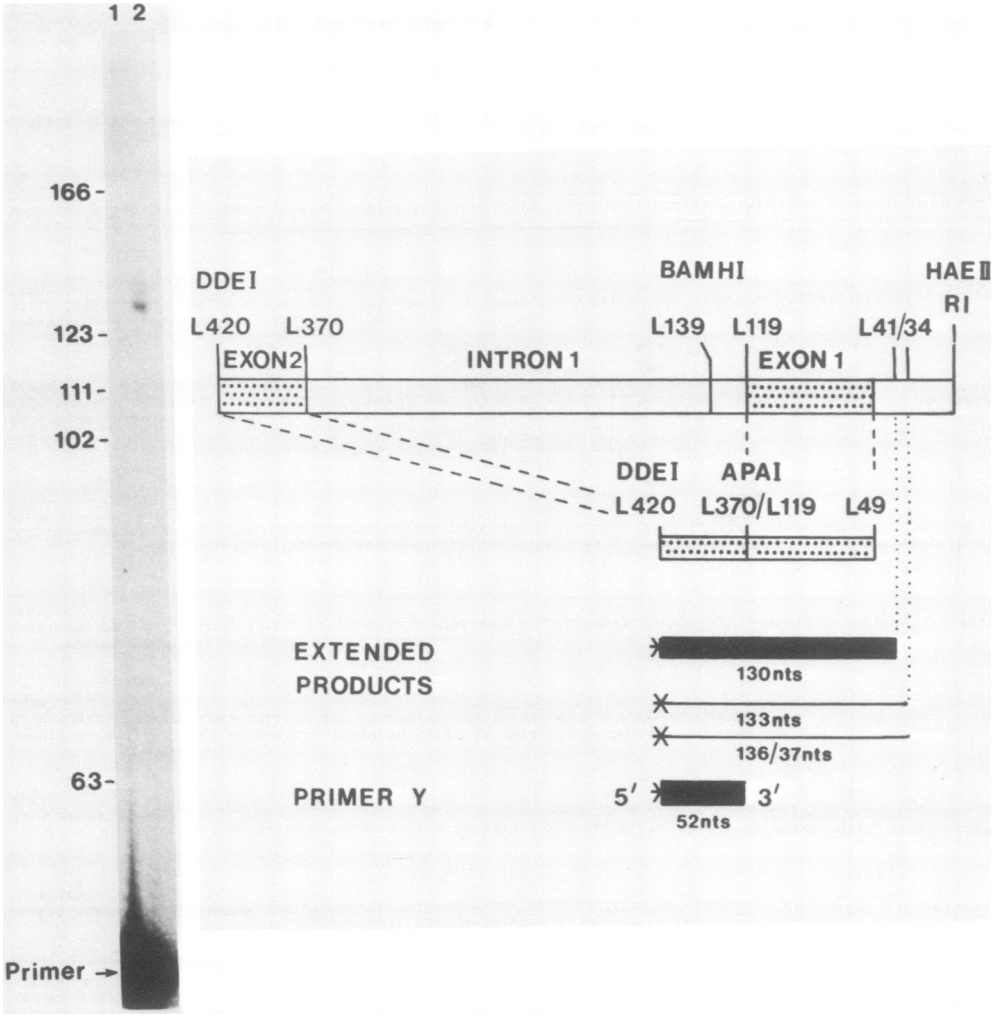

FIG. 7. Determination of the 5′ ends of the SURF-1 transcripts by primer extension analysis. On the left is presented a gel showing the results of the primer extension analysis with primer Y. At the right is shown a diagrammatic representation of primer Y, 5′ labeled at the *Dde*I site, and the major products of the primer extension analysis (as described in the legend to Fig. 4) in terms of the SURF-1 transcription unit. A structural map of the 5′ end of the SURF-1 transcription unit, indicating the 5′-flanking and intron sequences (open boxes) and exon sequences present in cDNA T1B (stippled boxes), is shown to the right of the gel (Fig. 6). Below this is shown the structural map of the 5′ end of cDNA T1B (stippled box) which was used toconstruct primer Y. Lanes: 1, primer plus yeast RNA; 2, primer plus BALB/c TS-A-3%3 RNA. On the left of the gel is shown the position of primer Y and marker DNA sizes in base pairs.

direction of transcription and the location of the cap site (7). Bidirectional promoters may also facilitate the coordinate expression of two genes by the same regulatory elements. For example, the ability of the divergently transcribed *GAL1* and *GAL10* genes of *Saccharomyces cerevisiae* to respond to galactose is controlled by a galactose-responsive upstream activator sequence located between the two genes (13). Similarly, studies on the *MATa* mating-type locus in *S. cerevisiae* have identified one sequence motif which acts as an upstream activator sequence for the divergently transcribed $\alpha 1$ and $\alpha 2$ genes and another motif which is involved in the diploid-specific repression of the transcription of both genes (45). In addition, coordinately regulated divergent transcription units have also been found in the genomes of *Drosophila* species and *Bombyx* species (11, 12, 21, 49). However, the expression of divergent transcripts which share common upstream promoter elements need not necessarily be coordinately controlled, as, for example, for the *his3-pet56* genes of *S. cerevisiae* (48). Although these genes share a common AT-rich sequence element important for their constitutive expression, only the transcription of the *his3* gene is induced in response to amino acid starvation. The early and late regions of simian virus 40 and polyomavirus, which are expressed at different stages of the lytic cycle (50), can also be viewed as divergent transcription units sharing common promoter elements which are not coordinately regulated. It will be of interest to determine whether the expression of the SURF-1 and SURF-2 transcription units is differentially or coordinately regulated.

In addition to the MES-1-associated element, other potential bidirectional promoters have also been described in the mammalian genome (2, 7, 31, 39). In particular, opposite-strand transcripts have been detected which are homologous to the 5′-flanking region of the housekeeping gene dihydrofolate reductase (*DHFR*) in both mouse and hamster DNA (2, 7, 31). In common with SURF-1 and SURF-2, both divergent transcripts associated with the *DHFR* promoter initiate at multiple sites and lack a correctly positioned

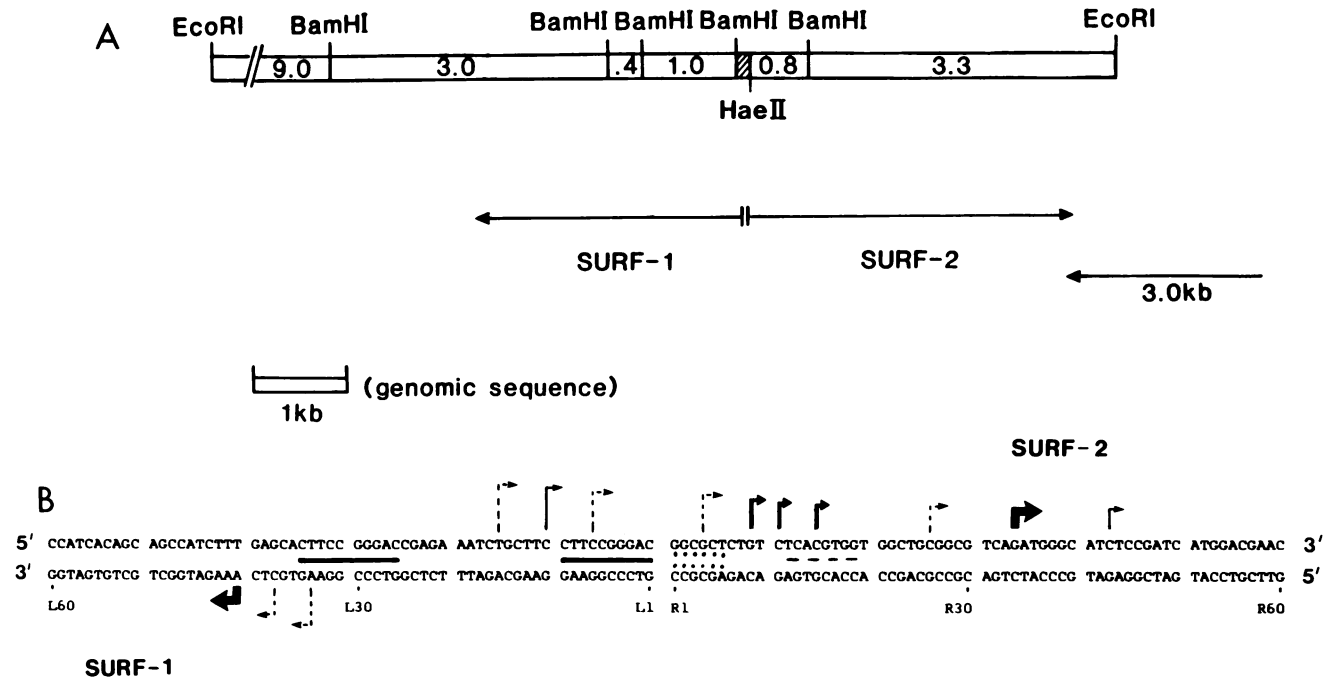

FIG. 8. Spatial arrangement of SURF-1 and SURF-2 transcription units. (A) Structural map of the 19-kb *Eco*RI fragment containing MES-1 (Fig. 1A) illustrating the location and orientation of the SURF-1 and SURF-2 transcription units with respect to this map. In addition, the position of a third transcription unit which overlaps SURF-2 at its 3' end is indicated (54). (B) Sequence of the region around the *Hae*II site containing the initiation sites of the SURF-1 and SURF-2 transcripts (see panel A). The 5' ends of the SURF-1 and SURF-2 transcripts are indicated by arrows below or above the sequence, respectively (the prominence of the initiation site is proportional to the width of the arrow; dashed arrows represent very minor start sites). The nucleotide numbers are given beneath the sequence. The locations of both the 10-bp direct repeat (solid lines) and the USF-binding site (41) (dashed line) are shown in between the two strands of the sequence. The position of the *Hae*II site (starting at nt R1) marking the right boundary of MES-1 is shown by a double row of dots in between the two stands of the sequence.

TATA or CAAT box motif in their 5'-flanking regions. However, the putative murine bidirectional promoters associated with MES-1 and *DHFR* differ in several respects. First, the distance between the major start sites of SURF-1 and SURF-2 (50 to 75 bp) is much shorter than the distance between the major start sites of the divergent transcripts associated with the *DHFR* promoter (approximately 130 bp) (2, 7). Second, as noted above, the MES-1 region lacks the GC boxes which are characteristic of the promoters of housekeeping genes such as *DHFR*. In addition, although both putative bidirectional promoters contain direct repeat sequences located between the major initiation sites of the divergent transcripts, these repeats are much shorter for MES-1. The potential importance of the 10-bp direct repeats associated with MES-1 was demonstrated in the initial deletion studies (9). A unidirectional deletion which removed the left copy of this repeat correspondingly greatly reduced the ability of MES-1 to activate the transcription of a heterologous gene. In addition, because of the short distance between the 5' ends of the two transcription units, some of the elements involved in the regulation of expression of SURF-1 and SURF-2 may actually be present outside MES-1 and perhaps within the transcription units themselves. One such potentially important motif, located outside MES-1 and bordering the SURF-2 transcription unit, is the sequence TCACGTGG (Fig. 8B) which represents the core recognition site for the *trans*-acting factor USF in the adenovirus major later promoter (41).

The SURF-1 and SURF-2 transcription units are located in unusually close proximity for mammalian genes. Furthermore, two other transcription units have been identified at this locus. One of these additional transcription units, encoding a differentially regulated 3.0-kb poly(A)$^+$ RNA species, overlaps SURF-2 at its 3' end (54) (see Fig. 8A). The fourth transcription unit, which remains to be fully characterized, ends 70 bp from the 3' end of SURF-1 (T. Williams, unpublished observation). The significance of the tight clustering of these four transcription units is presently under investigation.

Both SURF-1 and SURF-2 are capable of generating more than one species of poly(A)$^+$ RNA by differential splicing. Normally, such internal differences in mRNAs derived from the same primary transcript are the result of the presence or absence of a particular exon (3, 29, 37). However, in some instances such diversity may also be generated by the use of alternative splice donors (33, 35) or acceptors (4, 42) for the same exon. This latter mechanism is also responsible for the generation of the different SURF-1 and SURF-2 poly(A)$^+$ RNA species. Two different transcripts generated by differential splicing were identified from the SURF-2 transcription unit (Fig. 2 and 3). Both SURF-2 transcripts contain long open reading frames and could encode proteins of approximately 250 amino acids. Three different transcripts generated by the use of alternative splice acceptor sites were identified for the SURF-1 transcription unit (Fig. 5 and 6). In this instance only the major mRNA species (IDE) contains a long open reading frame which would encode a protein of approximately 300 amino acids. The other two minor mRNA species could only code for short peptides of 59 and 44 amino acids. Whether these minor RNA species are functional

mRNAs or represent nonfunctional transcripts (generated by the splicing apparatus recognizing fortuitous splice acceptor sequences) remains unclear at this time.

In summary, most of our knowledge of mammalian expression sequences has been obtained from the study of highly expressed or differentially regulated genes. More recently, the study of the 5'-flanking sequences of the small nuclear RNA genes (47), ribosomal protein genes (53), and housekeeping genes (5, 27, 30, 34, 46, 51) has revealed that various alternative promoter structures also exist. The expression selection technique used to isolate MES-1 (9) enables other such alternative enhancer-promoter structures to be isolated without relying on the necessity of initially cloning a companion gene. Indeed, it may turn out that these alternative promoter designs, including the potential bidirectional promoter described here, are a more widespread feature of the mammalian genome than they appear to be at present.

ACKNOWLEDGMENTS

We are indebted to Moira Griffiths for excellent technical assistance. We would also like to thank B. Davies, C. Norbury, V. von Hoyningen-Huene, M. van Wiles, G. Matlashewski, and R. Moore for helpful discussion during the course of this work. We are grateful to N. Jones, I. Kerr, C. Dickson, and V. Simanis for critical reading of the manuscript and to Georgina Briody and Audrey Symons for help in the preparation of this paper.

LITERATURE CITED

1. **Benoist, G., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–315.
2. **Crouse, G. F., E. J. Leys, R. N. McEwan, E. G. Frayne, and R. E. Kellems.** 1985. Analysis of the mouse *dhrf* promoter region: existence of a divergently transcribed gene. Mol. Cell. Biol. **5**:1847–1858.
3. **de Ferra, F., H. Engh, L. Hudson, J. Kamholz, C. Puckett, S. Molineaux, R. A. Lazzarini.** 1985. Alternative splicing accounts for the four forms of myelin basic protein. Cell **43**:721–727.
4. **Denoto, F. M., D. D. Moore, and H. M. Goodman.** 1981. Human growth hormone DNA sequence and mRNA structure: possible alternative splicing. Nucleic Acids Res. **9**:3719–3730.
5. **Dush, M. K., J. M. Sikela, S. A. Khan, J. A. Tischfield, and P. J. Stambrook.** 1985. Nucleotide sequence and organization of the mouse adenine phosphoribosyltransferase gene: presence of a coding region common to animal and bacterial phosphoribosyltransferases that has a variable intron/exon arrangement. Proc. Natl. Acad. Sci. USA **82**:2731–2735.
6. **Everett, R. D., D. Baty, and P. Chambon.** 1983. The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. **11**:2447–2464.
7. **Farnham, P. J., J. M. Abrams, and R. T. Schimke.** 1985. Opposite-strand RNAs from the 5' flanking region of the mouse dihydrofolate reductase gene. Proc. Natl. Acad. Sci. USA **82**:3978–3982.
8. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimentional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
9. **Fried, M., M. Griffiths, B. Davies, G. Bjursell, G. La Mantia, and L. Lania.** 1983. Isolation of cellular DNA sequences that allow expression of adjacent genes. Proc. Natl. Acad. Sci. USA **80**:2117–2121.
10. **Frischauf, A., M. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinkers. J. Mol. Biol. **170**:827–842.
11. **Garabedian, M. J., M.-C. Hung, and P. Wensink.** 1985. Independent control elements that determine yolk protein gene expression in alternative *Drosophila* tissues. Proc. Natl. Acad. Sci. USA **82**:1396–1400.
12. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**:765–789.
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. **Gluzman, Y. (ed.).** 1985. Current communication in molecular biology. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Goodbourn, S., K. Zinn, and T. Maniatis.** 1985. Human β-interferon gene expression is regulated by an inducible enhancer element. Cell **41**:509–520.
16. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell **44**:565–576.
17. **Grosveld, G. C., A. Rosenthal, and R. A. Flavell.** 1982. Sequence requirements for the transcription of the rabbit β-globin gene *in vivo*: the −80 region. Nucleic Acids Res. **10**:4951–4971.
18. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
19. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
20. **Hayday, A., H. E. Ruley, and M. Fried.** 1982. Structural and biological analysis of integrated polyoma virus DNA and its adjacent host sequences cloned from transformed rat cells. J. Virol. **44**:67–77.
21. **Jones, C. W., and F. C. Kafatos.** 1980. Structure, organization and evolution of developmentally regulated chorion genes in a silkmoth. Cell **22**:855–867.
22. **Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian.** 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science **232**:755–759.
23. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559–572.
24. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
27. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5' termini of mouse dihydrofolate reductase mRNAs. J. Biol. Chem. **260**: 2307–2314.
28. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpesvirus *tk* gene share a common hexanucleotide control sequence. Cell **37**:253–262.
29. **Medford, R. M., H. T. Nguyen, A. T. Destree, E. Summers, and B. Nadal-Ginard.** 1984. A novel mechanism of alternative RNA splicing for the developmentally regulated generation of troponin T isoforms from a single gene. Cell **38**:409–421.
30. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.
31. **Mitchell, P. J., A. M. Carothers, J. H. Han, J. D. Harding, E. Kas, L. Venolia, and L. A. Chasin.** 1986. Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5' region of the CHO *dhrf* gene. Mol. Cell. Biol. **6**:425–440.
32. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
33. **Nagata, S., M. Tsuchiya, S. Asano, O. Yamamoto, Y. Hirata, N. Kubota, M. Oheda, H. Nomura, and T. Yamazaki.** 1986. The chromosomal gene structure and two mRNAs for human granulocyte colony-stimulating factor. EMBO J. **5**:575–581.

34. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
35. **Reynolds, G. A., J. L. Goldstein, and M. S. Browne.** 1985. Multiple mRNAs for 3-hydroxy-3-methylglutaryl coenzyme A reductase determined by multiple transcription initiation sites and intron splicing sites in the 5′-untranslated region. J. Biol. Chem. **260**:10369–10377.
36. **Rigby, P. W. J., C. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labelling to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
37. **Rosenfeld, M. G., J.-J. Mermod, S. G. Amara, L. W. Swanson, P. E. Sawchenko, J. Rivier, W. W. Vale, and R. M. Evans.** 1983. Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. Nature (London) **304**:129–135.
38. **Ruley, H. E., L. Lania, F. Chaudry, and M. Fried.** 1982. Use of a cellular polyadenylation signal by viral transcripts in polyoma virus transformed cells. Nucleic Acids Res. **10**:4515–4524.
39. **Saffer, J. D., and M. F. Singer.** 1984. Transcription from SV40-like monkey DNA sequences. Nucleic Acids Res. **12**:4769–4788.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
41. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major later promoter upstream of the TATA box region. Cell **43**:165–175.
42. **Schwarzbauer, J. E., J. W. Tamkun, I. R. Lemischka, and R. O. Hynes.** 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. Cell **35**:421–431.
43. **Serfling, E., M. Jasin, and W. Schaffner.** 1985. Enhancers and eukaryotic gene transcription. Trends Genet. **1**:224–230.
44. **Serfling, E., A. Lubbe, K. Dorsch-Hasler, and W. Schaffner.** 1985. Metal-dependent SV40 viruses containing inducible enhancers from the upstream region of metallothionein genes. EMBO J. **4**:3851–3859.
45. **Siliciano, P. G., and K. Tatchell.** 1984. Transcription and regulatory signals at the mating type locus in yeast. Cell **37**:969–978.
46. **Singer-Sam, J., D. H. Keith, K. Tani, R. L. Simmer, L. Shively, S. Lindsay, A. Yoshida, and A. D. Riggs.** 1984. Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase. Gene **32**:409–417.
47. **Skuzeski, J. M., E. Lund, J. T. Murphy, T. H. Steinberg, R. R. Burgess, and J. E. Dahlberg.** 1984. Synthesis of human U1 RNA. J. Biol. Chem. **259**:8345–8352.
48. **Struhl, K.** 1985. Nucleotide sequences and transcriptional mapping of the yeast *pet56-his3-ded1* gene region. Nucleic Acids Res. **13**:8587–8601.
49. **Swaroop, A., J.-W. Sun, M. L. Paco-Larson, and A. Garen.** 1986. Molecular organization and expression of the genetic locus *glued* in *Drosophila melanogaster*. Mol. Cell. Biol. **6**:833–841.
50. **Tooze, J. (ed.).** 1981. Molecular biology of tumour viruses, part 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
51. **Valerio, D., M. G. C. Duyvesteyn, B. M. M. Dekker, G. Weeda, T. M. Berkvens, L. van der Voorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4**:437–443.
52. **von Hoyningen-Huene, V., C. Norbury, M. Griffiths, and M. Fried.** 1986. Gene activation properties of a mouse DNA sequence isolated by expression selection. Nucleic Acids Res. **14**:5615–5627.
53. **Wagner, M., and R. P. Perry.** 1985. Characterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. Mol. Cell. Biol. **5**:3560–3576.
54. **Williams, T. J., and M. Fried.** 1986. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature (London) **322**:275–279.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.