MOLECULAR AND CELLULAR BIOLOGY, Mar. 1991, p. 1281–1294
0270-7306/91/031281-14$02.00/0

Vol. 11, No. 3

The Bidirectional Promoter of the Divergently Transcribed Mouse *Surf-1* and *Surf-2* Genes

ANDREW C. LENNARD AND MIKE FRIED*

Department of Eukaryotic Gene Organization and Expression, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

Received 8 October 1990/Accepted 3 December 1990

The ubiquitously expressed mouse *Surf-1* and *Surf-2* genes are divergently transcribed, and their heterogenous start sites are separated by up to a maximum of only 73 bp. By using in vitro DNase I, dimethyl sulfate methylation, and gel retardation assays, we have identified five putative promoter control elements between and around the *Surf-1* and *Surf-2* start sites. The effects of each site on the regulation of *Surf-1* and *Surf-2* transcription have been studied in vivo, and four sites were found to be functional promoter elements. A novel binding site is required for efficient use of the intermediate but not the major start site of *Surf-1*. Three elements function in a bidirectional manner and are shared for efficient and accurate expression of both *Surf-1* and *Surf-2*. One is an UEF (USF, MLTF) binding site which had a small effect on the use of the intermediate start sites of *Surf-1* and also affected the major start sites of *Surf-2*. Another has sequence homology to the RPGα binding site associated with some ribosomal protein gene promoters and is required for efficient expression of the major but not intermediate start sites of *Surf-1* and all start sites of *Surf-2*. The third, an RPGα-like site, is used for all start sites of both *Surf-1* and *Surf-2*. Dissection of this cellular promoter region showed that different binding sites affect the use of different start sites and revealed a complex interaction between multiple elements that constitute a bona fide bidirectional promoter.

In higher eukaryotes, genes are usually separated by thousands of base pairs of DNA and close, divergent transcription units are rare. The Surfeit locus of at least six tightly clustered genes (*Surf-1* to *-6*) is therefore particularly unusual (15). The organization and sequence of four of the Surfeit genes (*Surf-1* to *-4*) have been determined; no more than 73 bp separate any two adjacent genes, and two genes overlap by 133 bp (14, 15, 32, 33). The *Surf-1* and *Surf-2* genes are divergently transcribed, and their heterogenous transcription start sites are separated by a maximum of only 73 bp (32). The general organization of the Surfeit locus is conserved in higher eukaryotes, which strongly suggests a functional basis for the maintenance of the juxtaposition of the genes in the locus (34). The Surfeit genes have been found to be transcribed to some degree in all cell types studied (34). In addition, the closely spaced divergent *Surf-1* and *Surf-2* genes have other properties typical of housekeeping genes, with their 5′ ends being located in a hypomethylated CpG-rich island and the absence of a consensus TATA box sequence 25 to 30 bp upstream from their transcriptional start sites (32, 34).

Bidirectional transcription is a potential mechanism to coordinate the expression of two genes in either a positive or negative manner. Therefore, we were interested in investigating the underlying control mechanism for the expression of the divergent *Surf-1* and *Surf-2* genes and determining whether the promoter region is bidirectional or composed of two independent promoters. In this study, we have identified and individually characterized five in vitro factor binding sites within the *Surf-1* and *Surf-2* promoter region. The role of each of these sites for the expression of *Surf-1* and *Surf-2* in vivo has been determined. Three of the factor binding sites are required for efficient expression of both genes. One site affects the location of only the *Surf-1* intermediate start sites, and another had no effect on the expression of either gene. We find that different binding sites can affect the use of different start sites and that elements of the *Surf-1* and *Surf-2* promoter region are bidirectional.

* Corresponding author.

MATERIALS AND METHODS

Preparation of nuclear extracts and DNase I footprinting. Nuclear extracts were prepared from HeLa and L cells essentially as described by Wildeman et al. (31). Protein concentration was determined by using the Bio-Rad phosphoric acid protein assay procedure. To prepare probes for DNase 1 footprinting, the 317-bp fragment containing the *Surf-1* and *Surf-2* bidirectional promoter was excised with *Xho*II from a previously subcloned 0.8-kb *Bam*HI fragment of BALB/c 3T3 DNA (32) and ligated into the *Bam*HI site of pUC18. The resultant plasmid (pS12A) was cut with *Hin*dIII for the coding strand of *Surf-2* or with *Eco*RI for the noncoding strand, end labeled with T4 polynucleotide kinase (Biolabs), and recut with *Eco*RI or *Hin*dIII, respectively. Isolated probe was incubated with 12 μg of HeLa or L-cell nuclear extract and 1 μg of poly(dI-dC) (Pharmacia) and then digested with DNase I as previously described (19), with modifications. The digestions proceeded at 20°C in a 25-μl standard reaction buffer of 3 mM $MgCl_2$, 50 mM KCl, 10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.9), 1 mM dithiothreitol, and 10% (vol/vol) glycerol unless otherwise stated. For HeLa extracts, L-cell extracts, and in the absence of extract, 200, 85, and 15 ng of DNase I (Boehringer) were added for 90 s each. Maxam-Gilbert G+A reactions were carried out on aliquots of probe for sequence markers.

Gel retardation and methylation interference assays. Single-stranded synthetic oligonucleotides (100 ng) were labeled with T4 polynucleotide kinase and hybridized to the corresponding oligonucleotide, and the free label was remove by Sephadex G-50 columns. Probe (20,000 cpm) was incubated

with 12 μg of HeLa extract and 4 μg of poly(dI-dC) or 12 μg of L-cell extract and 1 μg of poly(dI-dC) in the standard binding buffer described above for the DNase I digestions at room temperature for 20 min. Complexes were resolved on 6% (30:1) nondenaturing polyacrylamide gels run in 0.5 × TBE. For methylation interference assays, double-stranded, labeled probes (10^6 cpm) were treated with dimethyl sulfate (DMS) as described by Maniatis et al. (21) for G sequencing except that the reactions proceeded for 8 min. The methylated probes (100,000 cpm) were incubated with 60 μg of extract and a proportionately increased amount of poly(dI-dC), and the complexes were resolved on nondenaturing gels. Excised bands were electroeluted, and the DNA was purified and treated with piperidine as for Maxam-Gilbert sequencing. Aliquots (2,000 cpm) were electrophoresed on 20% 19:1 denaturing gels. Maxam-Gilbert G+A reactions were carried out on aliquots of probe as described by Maniatis et al. (21) except that the reaction time was increased to 15 min.

Oligonucleotides and cloning procedures. The following oligonucleotides were synthesized with *Bam*HI and *Eco*RI ends: Su1/Su2, 5′-GATCCAAGCCATCACAGCAGCCATC TTTGAGCACTTCCGGGACCGAGAAG-3′; Su1, 5′-AAT TCGACACAGCAGCCATCTTTGAGGTCG-3′; Su2, 5′-AA TTCGACGAGCACTTCCGGGACCGAGTCG-3′; Su3, 5′-AATTCCGAGAAATCTGCTTCCTTCCGGGACGGGCT CTG-3′; Su′x′, 5′-GATCCTTGCGACCACCGCGTCATC CGTCGCGGCAGGTCTTAA-3′; and Su′y′, 5′-GATCCG GACGGCGCTCTGTCTCACGTGGTGGCTGCGGCGTC AG-3′. The mutated oligonucleotides are described in Fig. 3. The following oligonucleotides contain known factor binding sites: rpL32 RPGα, 5′-GATCCACCCAGAGCCGGAAG TGCTTCCCTG-3′; rpL32 RPGδ, 5′-AATTCCGTAAAACA GATGGCAGCCACCTCG-3′; adenovirus E4.ATF, 5′-AAGC TTCTAAAAAATGACGTAACGG-3′; and Ad2MLP.UEF, 5′-GATCCGTAGGCCACGTGACCGG-3′. The sequence of the neomycin sequencing primer was 5′-ACCTGCGTGCA ATCCATCTTGTTCAATCATGCGAA-3′; the sequence of the *Hin*dIII-*Bam*HI adaptor was 5′-AGCTTCCTG-3′.

The wild-type *Surf-1* and *Surf-2* bidirectional promoter was synthesized by ligating 10 overlapping synthetic oligonucleotides to form the promoter sequence from −141 to +67 (see Fig. 2) into a vector that expresses the neomycin gene. The neomycin gene fused to a thymidine kinase termination sequence was excised in a *Bgl*II-to-*Pvu*II fragment from pSHL72 (30) and inserted into pBR322 between nucleotides (nt) 375 and 2066 (*Bam*HI to *Pvu*II). This recombinant was then cut with *Bam*HI and *Bgl*II and then phosphatase treated prior to insertion of the *Surf-1* and *Surf-2* promoter. *Bam*HI and *Bgl*II sites were introduced at the ends of the promoter sequence, thereby allowing the promoter to be cloned in both directions. The oligonucleotides were designed such that individual factor binding sites could be mutated simply by replacing a pair of oligonucleotides with an appropriately mutated pair of oligonucleotides. Recombinants were screened by double-stranded sequencing of minipreps (36) from the neomycin sequencing primer described above. To construct pCS 12neo, the chloramphenicol acetyltransferase (CAT) gene was isolated from pSV2CAT in an *Aat*II-*Hin*dIII (nt 2560 to 5018) fragment and inserted into pS2neo (see Fig. 5) cut at *Bam*HI and *Aat*II, using a *Bam*HI-*Hin*dIII adaptor (see above).

Expression of the *Surf-1*/*Surf-2* neomycin fusion recombinants. NIH 3T3 cells (2×10^6 cells per 14-cm petri dish) were transfected with 1.5 μg of pSPα118 (an α-globin expression vector and kind gift from S. Goodbourn), 30 μg of the *Surf-1*/*Surf-2* neomycin test plasmid, and 28.5 μg of pUC18 by the calcium phosphate technique (2, 17). The concentration of plasmids was determined by optical density at 260 nm and checked on agarose gels. Forty-eight hours after transfection, the cells were harvested and total cytoplasmic RNA was isolated (11). Aliquots of RNA (typically 30 μg) were analyzed for *Surf-1*/*Surf-2* neomycin transcripts by S1 mapping. The α-globin probe was 5′ end labeled at the *Bss*HII site and extended to the *Bam*HI site, neomycin probes A and B (see Fig. 5) were labeled at the *Nar*I site and extended to the *Hin*dIII site, and CAT probe C was labeled at the *Eco*RI site and extended to the *Bgl*II site. RNA was hybridized with 20,000 cpm each of an α-globin probe and appropriate *Surf-1*/*Surf-2* neomycin probe at 52°C overnight, digested with 100 U of S1 nuclease (Boehringer) at 25°C for 1.5 h (12), and analyzed on 6% denaturing polyacrylamide gels. Autoradiographs were quantitated on a densitometer.

RESULTS

DNase I footprinting reveals multiple factor binding sites within and around the *Surf-1* and *Surf-2* intergenic region. To determine the positions of binding sites responsible for efficient and accurate transcription of the *Surf-1* and *Surf-2* genes, we initially identified those sequences between and around the *Surf-1* and *Surf-2* transcription start sites protected from DNase I when incubated with nuclear extracts. Both mouse (L-cell) and human (HeLa) extracts were used to generate DNase I footprints on the coding and noncoding strands. The terms coding and noncoding are used with reference to the *Surf-2* gene. L-cell nuclear extracts (the homologous source) were found to be less stable than HeLa nuclear extracts; therefore, data for HeLa extracts are presented throughout unless otherwise stated.

Four regions of protection from DNase I were observed on both DNA strands, using either L-cell or HeLa nuclear extracts (Fig. 1). The extent of protection for each site and the positions of sites hypersensitive to DNase I were essentially the same for both extracts, with only minor differences. Figure 2 shows each footprinted region, subdivided into individual factor binding sites (Su1, Su2, and Su3; Su′x′, Su′y′, and Su′z′; see below) in relation to the start sites of *Surf-1* and *Surf-2*. Sites Su′x′ (−110 to −129 on the *Surf-2* coding strand, relative to the *Surf-2* major start site at +1) and Su′z′ (+50 to +60 on the *Surf-2* coding strand) lie downstream of the major start sites of *Surf-1* and *Surf-2*, respectively, whereas the adjacent sites Su1 and Su2 (−63 to −87 on the *Surf-2* coding strand) lie over the major start site of *Surf-1*, and site Su′y′ (−9 to −24 on the *Surf-2* coding strand) lies just upstream of the major start site of *Surf-2*. Factor binding site Su3, as discussed below, did not generate a DNase I footprint when crude nuclear extracts were used; it lies between sites Su2 and Su′y′. The effect of $MgCl_2$ on the footprint was investigated (Fig. 1A). Lower $MgCl_2$ concentrations (1 mM) were found to increase binding of a factor to site Su1 and to destabilize binding to site Su′y′.

Site Su′z′ formed three DNA-protein complexes in vitro, as detected by gel retardation assays. Although they were competed for by the addition of unlabeled site Su′z′ sequences, the formation of these complexes was not prevented by triple point mutation of the site (data not shown). Furthermore, the DMS methylation pattern produced by the site Su′z′ complex was weak. Therefore, these complexes appear not to be specific. However, deletion of site Su′z′ had a small qualitative effect on the start site usage for *Surf-2* transcription. There was a shift in use of the major site to the

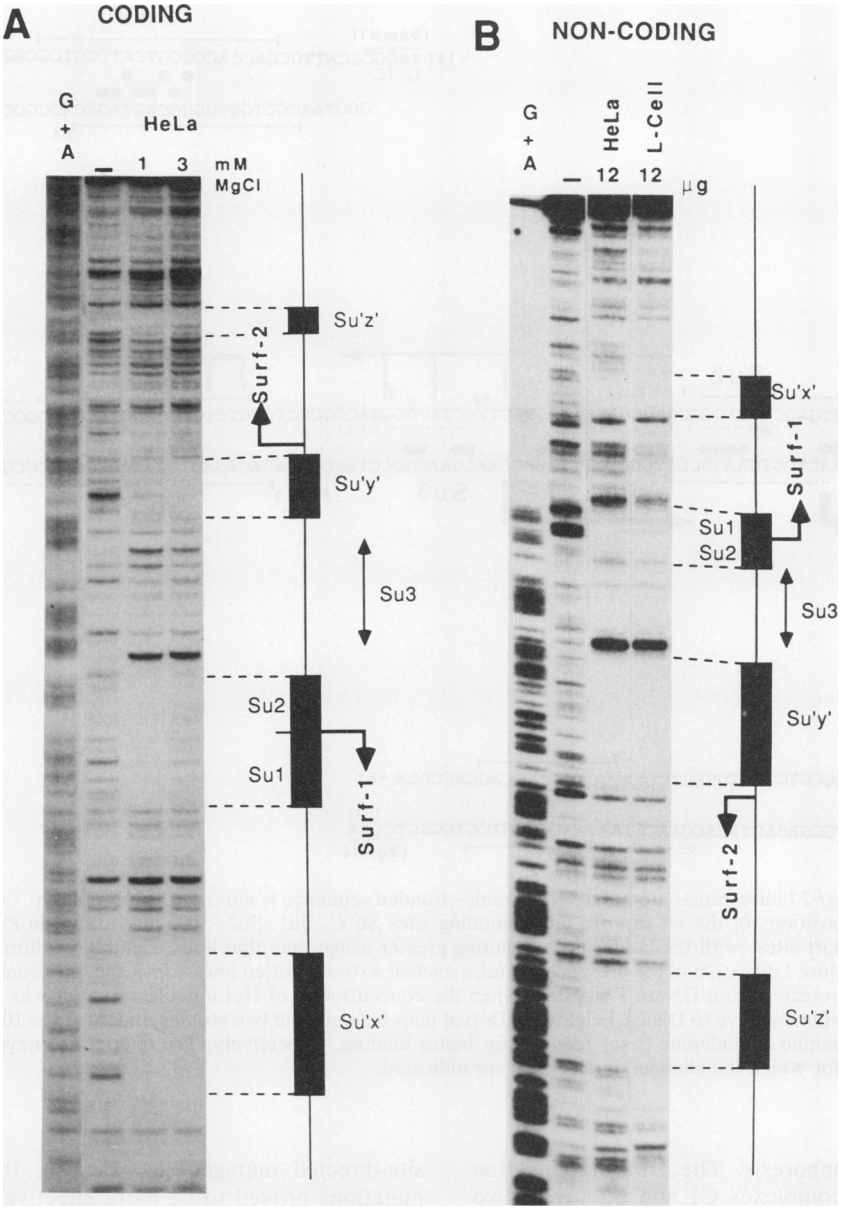


FIG. 1. DNase I footprinting of the *Surf-1* and *Surf-2* promoter region. Probes labeled at the *Hin*dIII or *Bam*HI sites of plasmid pS12A (see Materials and Methods) for the coding (A) or noncoding (B) strand of *Surf-2*, respectively, were incubated with 12 μg of HeLa of L-cell nuclear extract, briefly digested with DNase I, and analyzed on sequencing gels. A footprint on the coding strand is also shown, using HeLa extracts at 1 or 3 mM $MgCl_2$ (A). Probes digested with DNase I in the absence of nuclear extract are shown (−). Promoter schemes of the footprints are shown in relation to the *Surf-1* and *Surf-2* major start sites (large arrowheads; see also Fig. 2). Maxam-Gilbert G+A sequence reactions are shown for each probe.

intermediate start sites (data not shown). We consider this effect to be due to the sequence per se rather than to an activity of the observed DNA-protein complexes.

Sites Su1 and Su2 bind two distinct, independent complexes in vitro. The Su1 and Su2 factor binding sites and the complexes that bind to them were analyzed by gel retardation assays (Fig. 4A) and DMS methylation interference (Fig. 4B). Gel retardation assays using a wild-type synthetic oligonucleotide that spans the sequence of the adjacent sites Su1 and Su2 (−86 to −57; Fig. 2) revealed two DNA-protein complexes in vitro of different mobilities (complexes C1 and C2; Fig. 4A) upon nondenaturing gel electrophoresis with HeLa or L-cell nuclear extracts. The specificity of these complexes was demonstrated by the cocompetition of both complexes upon the addition of unlabeled wild-type Su1/Su2 oligonucleotide in the gel retardation assay, whereas site Su'y' sequences did not compete for either complex. The DNase I footprints suggest dual occupancy of factors to sites Su1 and Su2; however, double occupancy was not evident in the gel retardation assays. Absence of double occupancy is most likely due to differences in the stability of single- and double-occupancy complexes in gel retardation assays compared with DNase I footprinting combined with the removal of excess DNA, protein, poly(dI-dC), and changes in salt

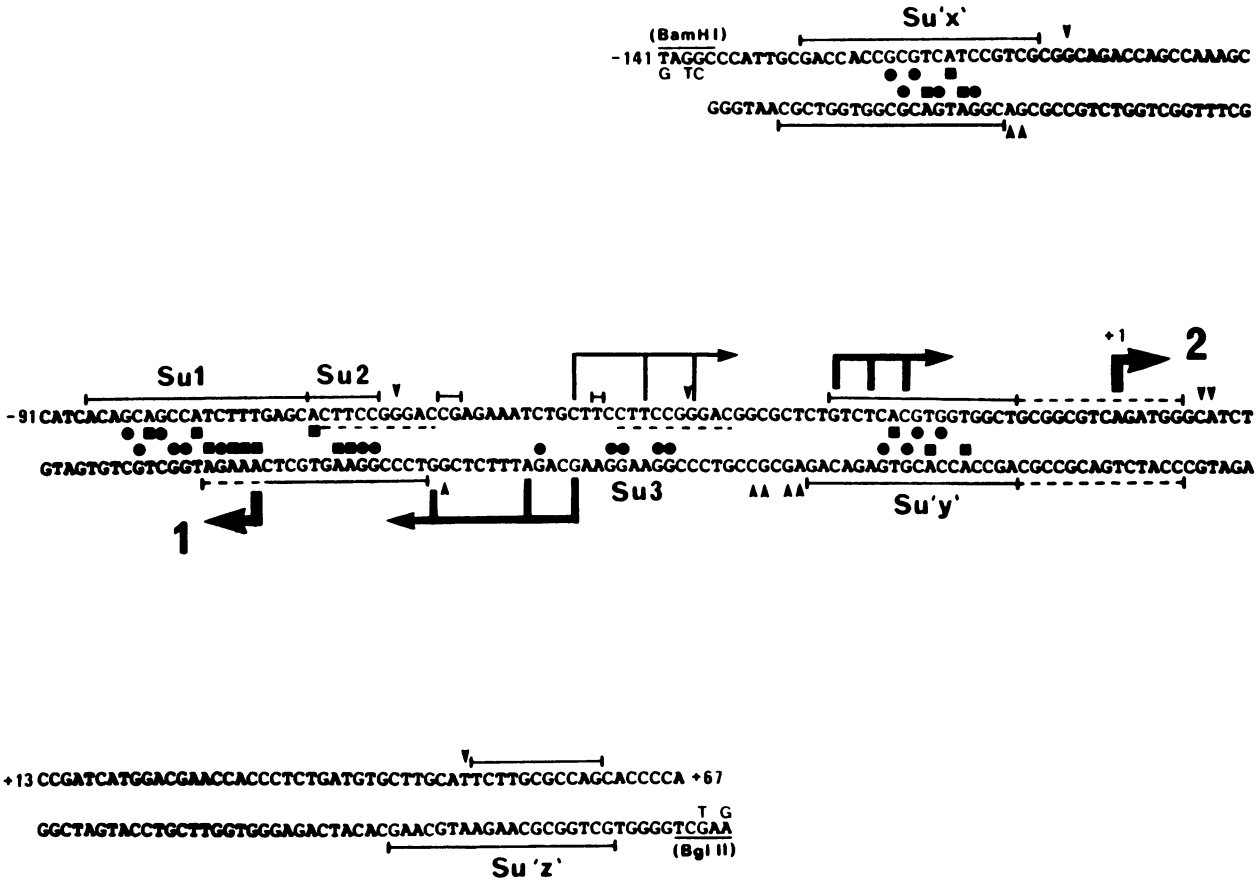


FIG. 2. The *Surf-1* and *Surf-2* bidirectional promoter. The double-stranded sequence is numbered from −141 to +67 relative to the *Surf-2* major start site at +1, and positions of the six in vitro factor binding sites Su'x', Su1, Su2, Su3, Su'y', and Su'z' are indicated. Arrows represent the transcription start sites, with the bolder lines denoting greater usage than thin lines. Continuous horizontal lines delimit the sequences protected from DNase I digestion in the presence of HeLa nuclear extract; dotted lines above the top strand and below the bottom strand illustrate sequences protected from DNase I digestion when the concentration of HeLa nuclear extract was increased. The vertical arrowheads designate sites hypersensitive to DNase I cleavage. Dotted lines between the two strands underline the 10-bp repeated sequence. Circles and squares locate guanine and adenine bases required for factor binding, respectively. The restriction enzyme sites in parentheses are introduced cloning sites for which the changed nucleotides are indicated.

conditions during electrophoresis. The DMS methylation interference patterns of complexes C1 and C2 reveal two distinctly separate clusters of essential guanine and adenine nucleotides along sites Su1 and Su2 (Fig. 4B) that share no sequence homology (Fig. 2). Therefore, complexes C1 and C2 are probably generated by different DNA sequence-specific binding factors. The two factors recognize different sequences within the Su1 and Su2 sites; the C1 factor(s) binds to site Su1 at −84 to −73 (5′-GCAGCCATCTTT-3′), and the C2 factor(s) binds to site Su2 at −68 to −63 (5′-ACTTCC-3′). Site Su1 is situated just downstream of the major start site of *Surf-1*, and site Su2 is located between the major and intermediate start sites of *Surf-1* (Fig. 2).

Further evidence for two different factor binding sites was obtained by competition gel retardation assays of complexes C1 and C2 by synthetic mutant oligonucleotides (Fig. 3). The mutant oligonucleotide **A**, in which there is a C-to-A change at nt −80 in site Su1, was shown by DMS methylation interference (Fig. 4B) to be involved in C1 formation. Oligonucleotide **A** was still able to partially compete for the C1 factor(s) against the Su1/Su2 wild-type sequence (Fig. 4A). Therefore, mutant oligonucleotides containing double or triple point mutations were constructed in subsequent site-directed mutageneses (Fig. 3). Indeed, double point mutations proved to be more effective; e.g., mutant oligonucleotide **B**, in which the CC dinucleotide at nt −79 and −80 in Su1 is changed to AA, was unable to abolish C1 while still capable of effectively binding and competing for C2 (Fig. 4A). A similar result was observed for mutant oligonucleotide **C**, in which the TT at nt −74 and −75 in site Su1 was changed to GG (Fig. 3). The mutation in the triple point mutant oligonucleotide **D**, in which the GAG at nt −70 to −72 was changed to TCT, lies between sites Su1 and Su2 as defined by the DMS methylation interference patterns (Fig. 4B). As expected, this mutation had no effect on the formation of C1 or C2 in vitro (Fig. 4A). Alternatively, mutations in the Su2 binding site prevent the formation of complex C2 and not complex C1. Oligonucleotide mutation **E**, in which the CT dinucleotide at nt −66 and −67 in site Su2 is changed to AG (Fig. 3), was unable to abolish C2, whereas it still efficiently bound and thus abolished C1 (Fig. 4A). Identical results were obtained with use of HeLa or L-cell nuclear extracts. Competition with mutated oligonucleotides that affect either complex C1 or complex C2 did not increase the binding of factors to site Su1 or site Su2, respectively.

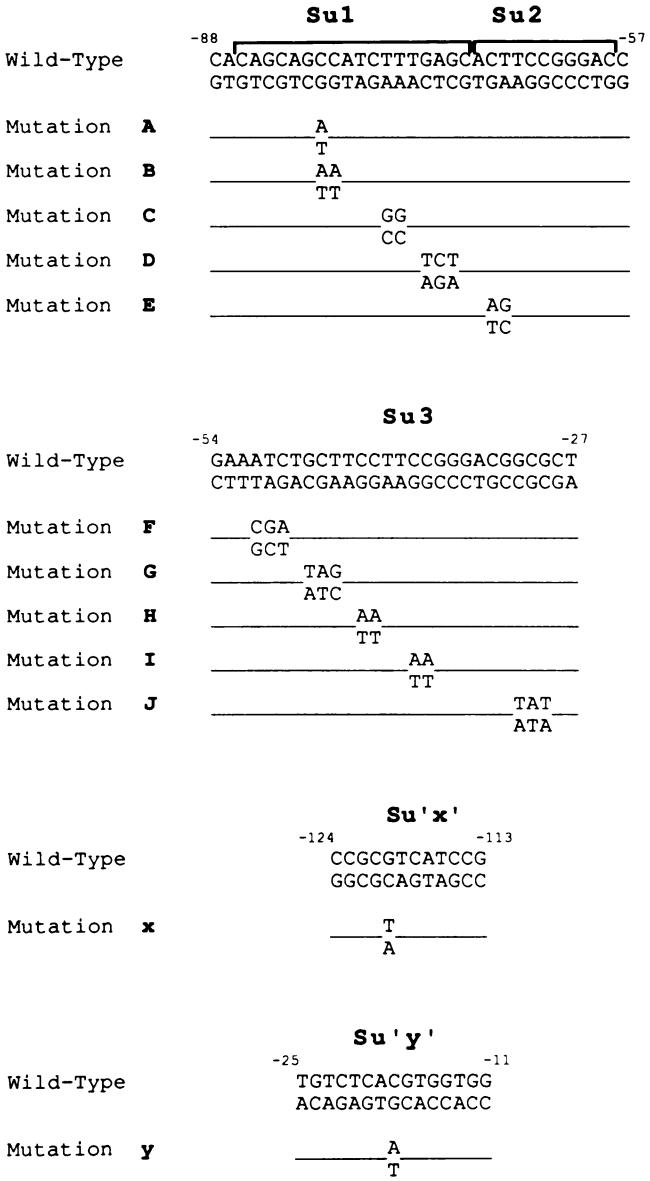


FIG. 3. Mutations of factor binding sites. Below the wild-type sequence for each of the factor binding sites Su1, Su2, Su3, Su'x', and Su'y' (Fig. 2) are shown the base changes made to generate the mutations used for the in vitro and in vivo analyses. See Materials and Methods for details of the synthesized oligonucleotides.

Therefore, two different factors appear to bind independently to two sites Su1 and Su2 in vitro.

The site Su2 sequence (5'-CACTTCCGG-3') has perfect homology with the RPGα factor binding site found in the mouse ribosomal protein gene L32 (rpL32) promoter (1) and differs in only the most 5' base from the XrpF1 binding site (5'-AACTTCCGG-3') in the *Xenopus laevis* rpL14 promoter (5). We tested the possibility of a common factor binding to the rpL32 promoter and site Su2 by competition of C2 with an RPGα site oligonucleotide (containing the RPGα core sequence and rpL32 flanking sequences). This rpL32 oligonucleotide was able to abolish complex C2 at least as effectively as the wild-type sequence (Fig. 4C). Another rpL32 promoter element, RPGδ (1), did not compete for the site Su2 factor. Conversely, when oligonucleotides containing the RPGα sequence were used as the probe for gel retardation assays, a single complex comigrating with C2 was generated which was efficiently abolished by RPGα or Su2 sequences (data not shown). Therefore, site Su2 may bind the same factor found to bind in vitro to the rpL32 promoter. We note that the DMS methylation interference pattern for the published *X. laevis* rpL14 XrpF1 binding site (5) differs slightly from that observed for the mouse *Surf-1*/*Surf-2* site Su2. For rpL14, guanine −53 is required for XrpF1 binding; the equivalent guanine −67 in site Su2 is not required for C2 formation (Fig. 4B). This may reflect a difference between homologous factors of the different amphibian and mammalian species.

Effect of the Su1 and Su2 binding sites on expression of *Surf-1* and *Surf-2* in vivo. To assess the role of sites Su1 and Su2 in vivo for the expression of *Surf-1* and *Surf-2*, the 208 bp of DNA containing sites Su'x' to Su'y' (−141 to +67; Fig. 2) were inserted upstream of the neomycin gene. The promoter sequence was inserted in either orientation to allow the expression of *Surf-1* and *Surf-2* to be studied individually (pS1neo and pS2neo; Fig. 5A). In addition, the same mutations present in the oligonucleotides in Fig. 3 were incorporated into both *Surf-1* and *Surf-2* neomycin fusion plasmids (Fig. 5B), and their role on transcription was tested in vivo by transfection (Fig. 4D). Total cytoplasmic RNA was assayed by S1 mapping after transient cotransfection of NIH 3T3 cells with the test plasmid and a reference plasmid expressing the α-globin gene. Figure 4D shows the S1 mapping data for transcription of the wild-type *Surf-1* and *Surf-2* neomycin fusion plasmids pS1neo and pS2neo. The inserted promoter sequence was sufficient for accurate transcription from the endogenous *Surf-1* and *Surf-2* cellular start sites equivalent to those previously reported (32). *Surf-2* has a major start site at +1, three intermediate start sites at −18, −21, and −24, and three minor start sites at −36, −40, and −46. The major, intermediate, and weak start sites are used in a ratio of 60:30:10. *Surf-1* has a major start site at −73 (relative to +1 of *Surf-2*) and three intermediate-strength start sites at −46, −50, and −58. Fifty percent of the *Surf-1* RNA is initiated at the major start site, and the initiation of the other 50% of the *Surf-1* RNA is spread among the three intermediate start sites. Note that the positions of the start sites have been more precisely determined in this study (see Fig. 2 for a summary).

The **B** and **E** mutations incorporated into the *Surf-1* (pS1neo.**B** and pS1neo.**E**) and *Surf-2* (pS2neo.**B** and pS2neo.**E**) neomycin fusion plasmids (Fig. 5) were used to study the in vivo roles of sites Su1 and Su2 (Fig. 4D). Site Su1 and Su2 mutations lie downstream of *Surf-1* start sites. Therefore, to obtain quantitative data from the S1 mapping analyses of RNA from cells transfected with the mutant *Surf-1* and *Surf-2* plasmids, two types of probes were used: (i) neomycin probe B (pS1neo; Fig. 5), which forms mismatches at the nucleotide differences in the heterologous RNA-DNA hybrids formed between the wild-type DNA probe and mutant RNA transcripts which can be cleaved by S1 nuclease, and (ii) probe from the mutant plasmid which is completely homologous to the mutant RNA transcripts. The former type of probe was used for each assay to eliminate any variation in the results generated by differences in the specific activities of different probe preparations. The data shown used the latter type of probe.

Figure 4D shows the S1 mapping data for steady-state RNA from cells transfected with the wild-type plasmid pS2neo and the mutant plasmids pS2neo.**B** and pS2neo.**E** for *Surf-2* and the wild-type plasmid pS1neo and the mutant

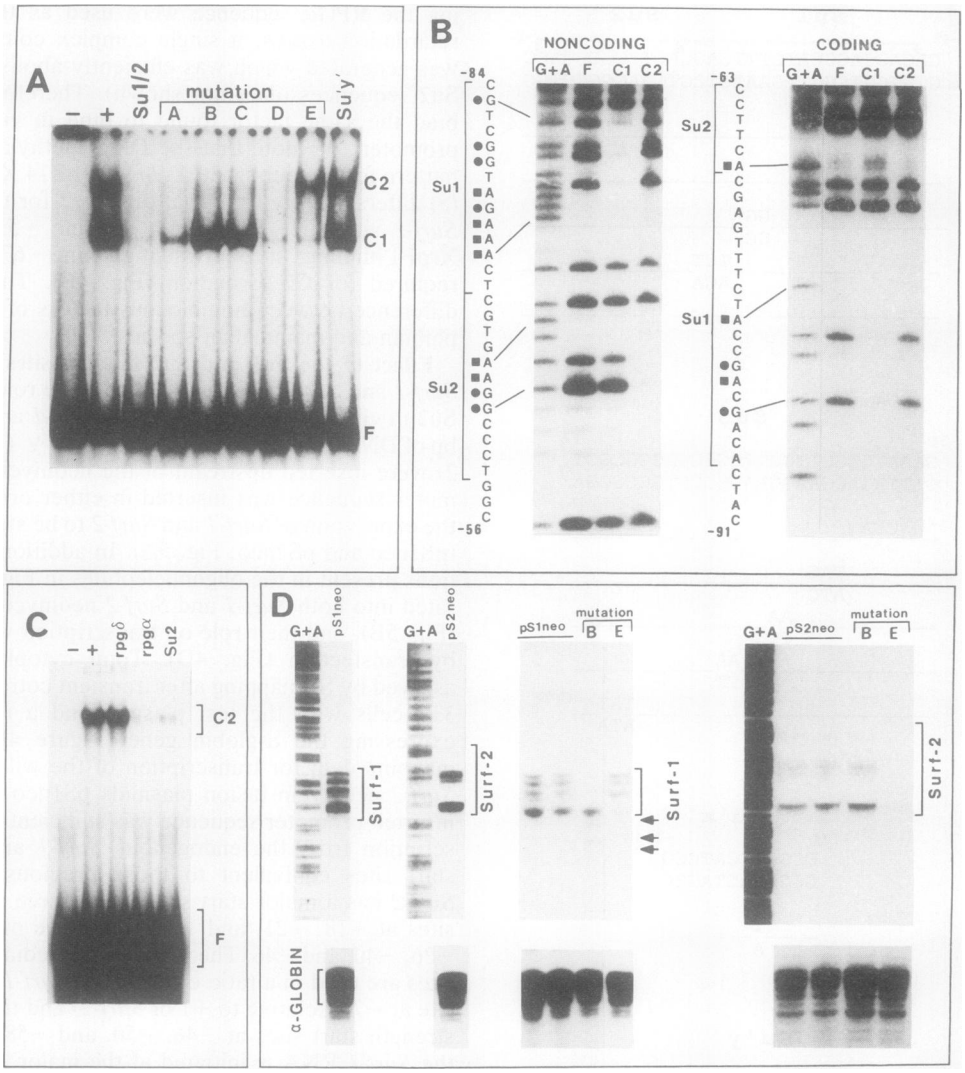
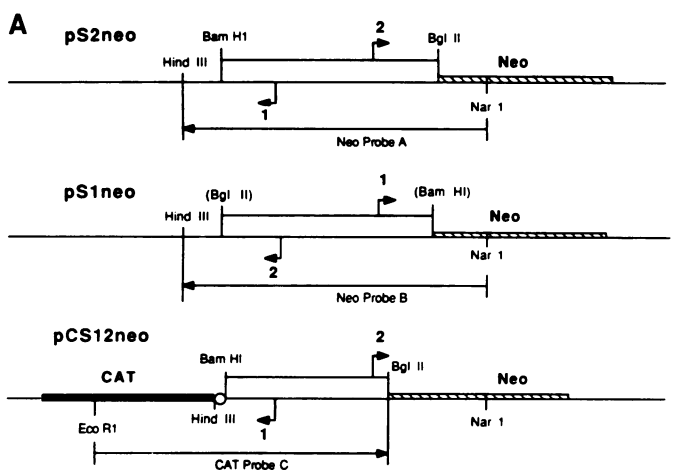


FIG. 4. Analysis of sites Su1 and Su2 in vitro and in vivo. (A) Gel retardation analyses of sites Su1 and Su2 with competition by wild-type and mutated Su1 and Su2 sequences. A labeled synthetic oligonucleotide probe from −53 to −96 encompassing the Su1 and Su2 binding sites was incubated with HeLa nuclear extract, and complexes C1 and C2 were competed for by oligonucleotides containing mutations A to E (see Fig. 3) or oligonucleotides containing site Su'y' or Su1/Su2. Free probe (F) in the absence of extract (−) complexes C1 and C2 in the absence of competitor (+) are also shown. (B) DMS methylation interference patterns for sites Su1 and Su2. Wild-type Su1/Su2 oligonucleotide was labeled on the noncoding or coding strand, partially methylated with DMS, and incubated with HeLa nuclear extract. The DMS methylation interference pattern was then determined for complexes C1 and C2 individually. The free probe methylation patterns (F) and Maxam-Gilbert G+A sequencing reactions are also shown. The results are summarized on the adjacent sequences; circles indicate guanine bases, and squares indicate adenine bases required for complex formation. (C) Competition of C2 with RPGα. Synthetic oligonucleotide containing the site Su2 sequence was incubated with HeLa nuclear extract in the absence (+) or presence of competing oligonucleotides containing the sequences of the RPGδ site, the RPGα site, and site Su2. The free probe (F) in the absence of extract (−) is shown. (D) In vivo expression of *Surf-1*/*Surf-2* neomycin fusion genes by the wild-type and Su1 or Su2 mutated promoter of *Surf-1* and *Surf-2*. NIH 3T3 cells were cotransfected with pSPα118 and pS1neo, pS2neo, pS1neo.**B**, pS1neo.**E**, pS2neo.**B**, or pS2neo.**E** as indicated. Total cytoplasmic RNA was analyzed by S1 mapping using neomycin probe A to detect transcripts from the *Surf-2* start sites, neomycin probe B to detect transcripts from the *Surf-1* start sites, and an α-globin probe (see Fig. 5 and Materials and Methods). Maxam-G+A sequence reactions are shown alongside. Arrows indicate novel transcription start sites observed when cells were transfected with pS1neo.**B**.

plasmids pS1neo.**B** and pS1neo.**E** for *Surf-1*. Clearly, mutation of site Su1 had no effect on the accuracy or efficiency of transcription from any of the *Surf-2* start sites. However, site Su1 does play a role in the positioning of the *Surf-1* transcription start sites. Although mutation of site Su1 did not alter the efficiency of transcription from the *Surf-1* major start site, transcription from the intermediate start sites was reduced to 20% of the wild-type level and three new sites were generated downstream of the major start site at −82, −91, and −96. The new sites represented 30% of total transcription; thus, there was no decrease in the overall level of steady-state RNA.

Mutation of site Su2 had a dramatic effect on the transcription of both *Surf-1* and *Surf-2*. Comparison of the transcription from pS2neo with pS2neo.2**E** (Fig. 4D) reveals a fivefold decrease in total transcription which is reflected in the



B

PLASMIDS

Surf-1	Surf-2	SITE MUTATED
pS1neo	pS2neo	----
pS1neo.**B**	pS2neo.**B**	**Su1**
pS1neo.**E**	pS2neo.**E**	**Su2**
pS1neo.**F**	pS2neo.**F**	**Su3**
pS1neo.**G**	pS2neo.**G**	**Su3**
pS1neo.**H**	pS2neo.**H**	**Su3**
pS1neo.**I**	pS2neo.**I**	**Su3**
pS1neo.**J**	pS2neo.**J**	**Su3**
pS1neo.**x**	pS2neo.**x**	**Su'x'**
pS1neo.**y**	pS2neo.**y**	**Su'y'**

FIG. 5. Schemes for the recombinant plasmids constructed to assess the function of *Surf-1* and *Surf-2* promoter elements in vivo. (A) Diagrammatic representation of the three basic constructs used for the in vivo expression assays and the probes used for the S1 mapping. The *Surf-1* and *Surf-2* bidirectional promoter from −141 to +67 (□; see Fig. 2) was inserted, in either orientation to detect transcription from the *Surf-1* start sites (pS1neo) or from the *Surf-2* start sites (pS2neo), upstream of the neomycin gene (▨). Alternatively, pCS12neo contains the promoter from −139 to +15. Plasmid pCS12neo has the CAT gene (■) downstream of the *Surf-1* start sites, oriented in the divergent direction to the neomycin gene. Thin lines represent pBR322, and the open circle represents a *Bam*HI-*Hin*dIII adaptor. Restriction enzyme sites in parentheses were destroyed by the cloning procedure. The major start sites for *Surf-1* and *Surf-2* are indicated by arrows. Neomycin probes A and B (from *Nar*I to *Hin*dIII) and CAT probe C (from *Eco*RI to *Bgl*II) were used for the S1 mapping analyses. These maps are not to scale. (B) Plasmids that contain wild-type or mutated promoter sequences that express the neomycin gene from the *Surf-1* (pS1neo series) or *Surf-2* (pS2neo series) start sites. Bold letters designate the promoter mutation present in the plasmid (see Fig. 3 for sequences of mutations). Each site mutated within the promoter is also listed.

major, intermediate, and minor start sites. Therefore, site Su2 is critical for *Surf-2* expression. Similarly, expression from the *Surf-1* major start site was decreased to 30% of the wild-type level when site Su2 was mutated. However, mutation of site Su2 had no effect on expression from the intermediate start sites of *Surf-1*. Thus, the total decrease in transcription is reduced to 70% of the wild-type level. It should be noted that mutation of site Su2 did not induce any novel start sites.

Site Su3 binds three factors in vitro and is functional in vivo. It is possible that there are functionally important transcription factor binding sites that were not identified by the DNase I footprinting technique. Gel retardation assays using a synthetic oligonucleotide comprising the sequence between sites Su2 and Su'y' (nt −24 to −63; Fig. 2) revealed three specific complexes (C3.1, C3.2, and C3.3) with use of HeLa (Fig. 6A) or L-cell (Fig. 6B) extracts. This site is referred to as Su3. The specificity of these three complexes is shown by the ability of unlabeled oligonucleotide containing the site Su3 sequence (Fig. 3) to inhibit formation of these complexes (Figs. 6A and B), whereas an oligonucleotide containing site Su1 or Su'x' did not affect formation of the complexes. No specific binding complexes were detected at other intersite sequences under a range of $MgCl_2$ or poly(dI-dC) concentrations.

The relative amounts of each Su3 complex C3.1, C3.2, and C3.3 differ strikingly between the two extracts. With HeLa extracts, C3.1 and C3.2 predominate and C3.3 is minor; with mouse extracts, C3.1 and C3.3 predominate and C3.2 is minor. We consider selective losses of factors during the preparation of extracts to be unlikely. The boundaries of site Su3 are unknown, since this site did not give rise to DNase I footprints (Fig. 2). However, two guanine nucleotides located at −38 and −39 were shown by DMS methylation interference to be involved in the formation of the HeLa complexes C3.1 and C3.2 and the L-cell complexes C3.1 and C3.3 (Fig. 6C). Further guanine bases are involved in C3.1 formation at −43 in both extracts and at −42 in L-cell extracts alone, and guanine −49 is required to generate C3.2.

Guanine residues in Su3 at −38, −39, and −42 lie within a 10-bp sequence repeated in site Su2 (10, 32). It is therefore possible that sites Su2 and Su3 bind the same or a similar factor. Indeed, we found that complexes C3.1 and C3.3 were both abolished by the sequence for site Su2 (Fig. 6A and B) or by the RPGα sequence (Fig. 6B). It is not likely that C3.3 was a proteolytic product of C3.1, since site Su2 formed only a single complex that comigrated with C3.3. Therefore, C3.3 binds the same or a similar factor as site Su2 and RPGα. C3.1 may represent a complex of the C3.3 factor and another protein. Complexes C3.2 and C3.3 appear to be formed by distinct factors that recognize overlapping DNA sequences to produce very similar methylation interference patterns, since C3.2 was not abolished by the site Su2 sequence. To further define the binding sites of the site Su3 complexes, oligonucleotides with the series of mutations **F** to **J** (Fig. 3) were synthesized for competition gel retardation assays (Fig. 6A and B). The mutant **H** and **I** oligonucleotides (Fig. 3), which contain altered cytidine nucleotides at −42 and −43 and at −38 and −39, respectively (G residues on the other DNA strand were shown to be important by DMS methylation interference [Fig. 6C]), were both unable to compete for C3.1 in HeLa or L-cell extracts. The mutant **H** and **I** oligonucleotides were also unable to compete for C3.3 in L-cell extracts. This result agrees with the DMS methylation interference data showing that these guanines were involved in C3.1 and C3.3 complex formation (Fig. 6C). Mutant oligonucleotides **F** and **I** (Fig. 3) were unable to compete for C3.2 in HeLa or L-cell extracts. Therefore, as suggested by the methylation interference pattern, C3.2 seems to have a recognition site different from but overlapping those of the other complexes. Mutant oligonucleotide **G**, which contains mutations situated between those of oligonucleotides **F** and **I**

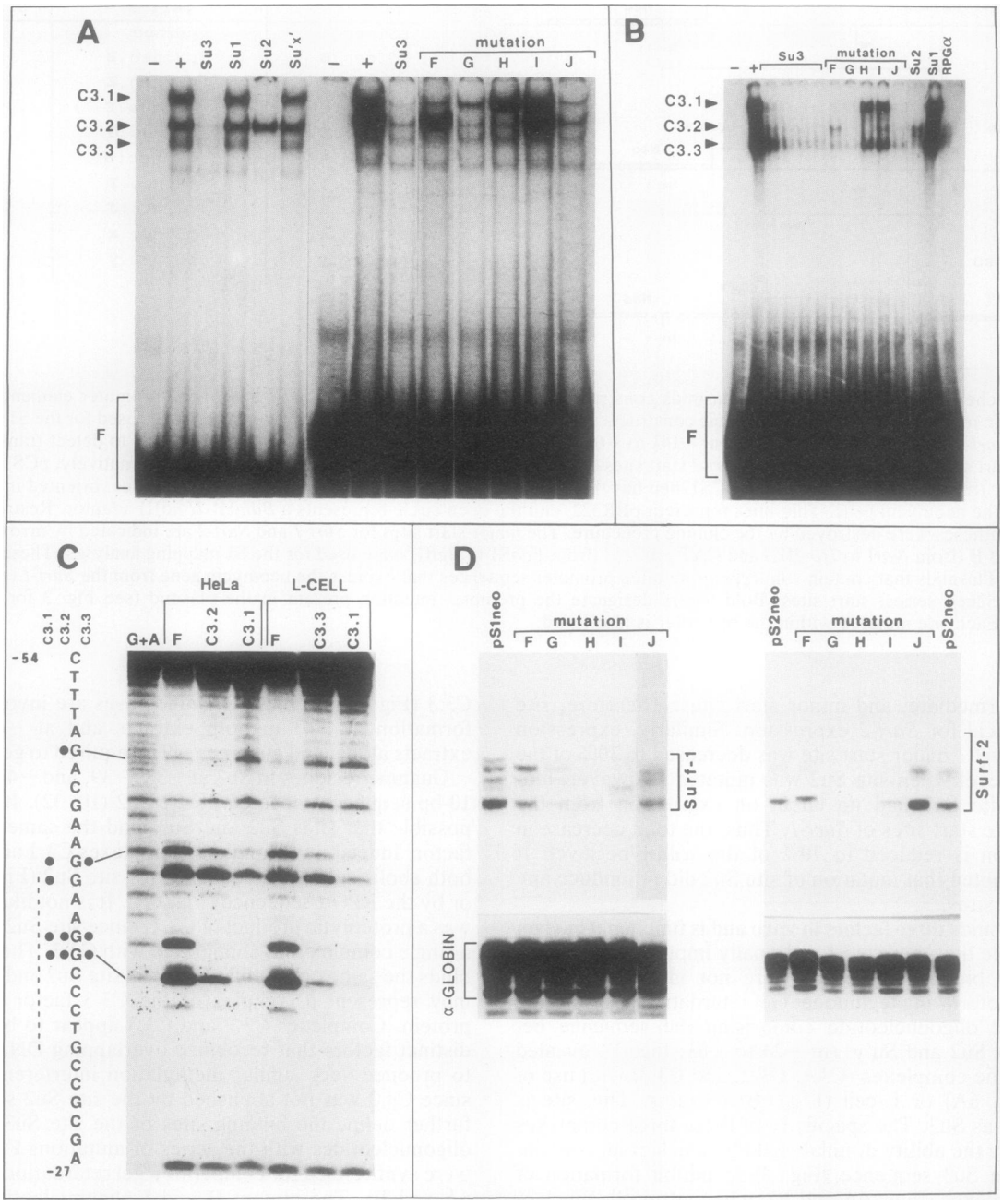


FIG. 6. Analysis of site Su3 in vitro and in vivo. (A and B) gel retardation analyses of site Su3 (−24 to −63) with competition by wild-type and mutated Su3 sequences. (A) Probe was incubated in the absence (−) or presence of HeLa nuclear extract, and complexes C3.1, C3.2, and C3.3 were competed for with oligonucleotides for sites Su3, Su1, Su2, and Su'x' and site Su3 mutations **F**, **G**, **H**, **I**, and **J** (see Fig. 3). Complexes in the absence of any competitor sequences (+) and free probe (F) are indicated. The binding conditions are as described in Materials and Methods except that 1 mM $MgCl_2$ was used. (B) Su3 probe was incubated in the absence (−) or presence (+) of L-cell nuclear extract, and the complexes were competed for with 50, 100, 250, and 500 times molar excess of oligonucleotide Su3 or 500 times excess of Su3 mutations **F**, **G**, **H**, **I**, and **J** or Su2, Su1, and RPGα. (C) DMS methylation interference patterns on the noncoding DNA strand of individual complexes C3.1, C3.2, and C3.3 generated by HeLa and L-cell nuclear extracts. Lane F contains the guanine sequence produced by oligonucleotide in the absence of extract. Lane G+A contains the guanine and adenine Maxam-Gilbert sequences. The patterns are summarized on the adjacent sequence. The dotted line indicates the 10-bp repeated sequence (32). Circles indicate guanine bases required for the formation of each complex. (D) In vivo expression of wild-type and site Su3 mutated *Surf-1*/*Surf-2* neomycin fusion recombinants. NIH 3T3 cells were cotransfected with pSPα118 and pS1neo, pS1neo.**F**, pS1neo.**G**, pS1neo.**H**, pS1neo.**I**, pS1neo.**J**, pS2neo, pS2neo.**F**, pS2neo.**G**, pS2neo.**H**, pS2neo.**I**, or pS2neo.**J** as indicated. Total cytoplasmic RNA was analyzed by S1 mapping using probe A to detect transcripts from the *Surf-2* start sites, probe B to detect transcripts from the *Surf-1* start sites, and an α-globin probe (see Fig. 5 and Materials and Methods).

(Fig. 3), did not alter the ability of site Su3 to form any of its complexes. Thus, complex C3.2 does not seem to lie along a continuous DNA sequence but may interact across a DNA helical turn.

The functional importance of site Su3 was studied by introducing the series of site Su3 mutations **F** to **J** (Fig. 3) into the *Surf-1* and *Surf-2* neomycin fusion vectors (Fig. 5, pS1neo.**F** to -**J** and pS2neo.**F** to -**J**). These plasmids were transiently cotransfected with the reference α-globin plasmid into NIH 3T3 cells. Figure 6D shows the S1 mapping data for RNA from the *Surf-1* and *Surf-2* start sites. The neomycin probes A and B (pS2neo and pS12neo; Fig. 5) caused a change in the position of the minor start sites in the S1 mapping data for several site Su3 mutations because mutations **F** and **G** lie within the *Surf-1* minor start sites and mutations **G** to **J** lie within the *Surf-2* minor start sites (compare sequences in Fig. 2 and 3). The results show that mutations **G**, **H**, and **I** had a deleterious effect on expression from all of the *Surf-1* and *Surf-2* transcription start sites (Fig. 6D). However, each of these site Su3 mutations had a significantly greater effect on *Surf-2*. Mutations **G** and **H** were the most deleterious, resulting in 72 and 90% reductions in transcription from the *Surf-2* start sites compared with the wild-type level of transcription. The *Surf-1* start sites had 40 and 58% reductions in transcription. Mutation **I** had a lesser effect on the transcription signals, with a 65% reduction for *Surf-2* and a 21% reduction for *Surf-1*. Mutations **F** and **J** gave wild-type levels of transcription from *Surf-1* and *Surf-2* start sites.

These results show that the most deleterious site Su3 mutations lie within and close to the 10-bp repeated DNA sequence that binds factors C3.1, C3.2, and C3.3 detected in vitro. Although the effects of mutations **H** and **I** correlate with the C3.1 and C3.3 binding sites, the severely deleterious mutation **G** is located at a region that was not required for the binding of any of the characterized site Su3 factors. Also, mutation **F** prevented binding of complex C3.2 yet had no effect on in vivo *Surf-1* or *Surf-2* expression. The role of site Su3 is thus complex; the results suggest that C3.1, C3.3, or both may be functional since the in vitro and in vivo results are in good agreement. C3.2, the minor complex formed with the homologous L-cell source of extract, does not appear to be functional in vivo. It remains possible that other undetected factors bind to site Su3 at the **G** mutation sequence. Alternatively, it may be the conformation of the DNA around mutation **G** that is significant.

Sites Su′x′ and Su′y′ bind specific factors in vitro, yet only Su′y′ is functional in vivo. Oligonucleotides containing the sequences of site Su′x′ and Su′y′ (Fig. 2) were used as probes for gel retardation assays. Each oligonucleotide generated single discrete complexes, Cx for site Su′x′ and Cy for site Su′y′ (Fig. 7A and B, respectively), which were efficiently abolished by the presence of excess unlabeled oligonucleotides for site Su′x′ and site Su′y′. Identically migrating complexes were obtained with use of HeLa or L-cell extracts. These were sequence-specific complexes, since they were stable in the presence of excess oligonucleotide of unrelated Su1-Su2 sequence. To determine the guanine and adenine nucleotides essential for the formation of Cx and Cy, DMS methylation interference assays were carried out (Fig. 7C) on the coding and noncoding strands. Clear methylation patterns were generated for each complex within the sequences protected from DNase I. On the basis of this information, single point mutations were introduced into the binding sites for Su′x′ and Su′y′ (Fig. 3). Mutation x (G to T at nt −121) for site Su′x′ and mutation y (C to A at nt −20) for site Su′y′ were no longer able to form their respective complexes in gel retardation assays (Fig. 7A and B).

Inspection of the sequence of site Su′x′ reveals perfect homology from −122 to −118 with the consensus sequence for an ATF/CREB binding site (5′-CGTCA-3′) (reference 13 and references therein), and the DMS methylation interference pattern agrees with the DMS protection footprint for the ATF-adenovirus type 2 (Ad2) E3.BS2 promoter complex (13). Indeed, competition gel retardation assays show that an adenovirus E4 ATF binding site effectively competed for Cx, whereas the major late promoter (MLP) upstream sequence did not (Fig. 7D). We conclude that an ATF/CREB-like factor binds to site Su′x′ in vitro.

It was previously noted (32) that a short sequence of the intergenic DNA of *Surf-1* and *Surf-2* had some homology with the binding site of the Ad2 MLP upstream factor (UEF, USF, MLTF [23, 26]). The palindromic homology (5′-CACGTG-3′ at −17 to −22) lies within site Su′y′. Furthermore, the DMS protection footprint pattern for the Ad2 MLP upstream element (19) agrees with the DMS methylation interference pattern obtained for site Su′y′. We therefore competed Cy with an oligonucleotide of the MLP upstream sequence or an oligonucleotide of an ATF/CREB site from the adenovirus E4 promoter (Fig. 7D). As expected, the UEF binding site efficiently abolished Cy, whereas the ATF binding site competed very poorly for Cy. We conclude that site Su′y′ binds to an UEF-like factor in vitro.

The in vivo functions of sites Su′x′ and Su′y′ were studied by expressing *Surf-1* and *Surf-2* neomycin fusion genes under the control of a wild-type or a mutated promoter as described above. Although there is a single nucleotide mismatch between transcripts from the site Su′x′ mutated plasmid (pS2neo.**x**) and neomycin probe B and between transcripts from the site Su′y′ mutated plasmid (pS1neo.**y**) and neomycin probe A, they did not interfere with the S1 analyses. Despite the strong in vitro binding of nuclear factors to sites Su′x′ and Su′y′, only mutation of site Su′y′ had any effect on the overall efficiency or accuracy of *Surf-1* or *Surf-2* expression (Fig. 7E; compare the wild type with the site Su′x′ mutant plasmids pS1neo.**x** and pS2neo.**x** with the site Su′y′ mutant plasmids pS1neo.**y** and pS2neo.**y**). Mutation of site Su′y′ decreased transcription from one of the *Surf-1* intermediate start sites at nt −58, with the appearance of a novel, intermediate start site at the site of mutation at nt −20. Also, the site Su′y′ mutation resulted in a 64% decrease in transcription from the *Surf-2* major start site with no effect on the intermediate or minor start sites, which resulted in an overall reduction of transcription to 56% of the wild-type level.

Bidirectional transcription by the *Surf-1* and *Surf-2* promoter. We have demonstrated bidirectional activity for a 208-bp fragment containing the sequence between and around the major start sites for *Surf-1* and *Surf-2*. These experiments detected transcription from each orientation of the promoter sequence individually. We were therefore interested in showing that this promoter could function in a truly bidirectional manner in a single plasmid. A shorter 154-bp promoter sequence containing nt −139 to +15 (Fig. 2), which did not contain site Su′z′, was cloned such that the *Surf-2* gene was fused to the neomycin gene. The CAT gene was then inserted downstream of *Surf-1* in the same plasmid, pCS12neo (Fig. 5). This plasmid was transiently transfected into NIH 3T3 cells, and the total cytoplasmic RNA was probed for *Surf-1*/CAT and *Surf-2*/neomycin fusion tran-

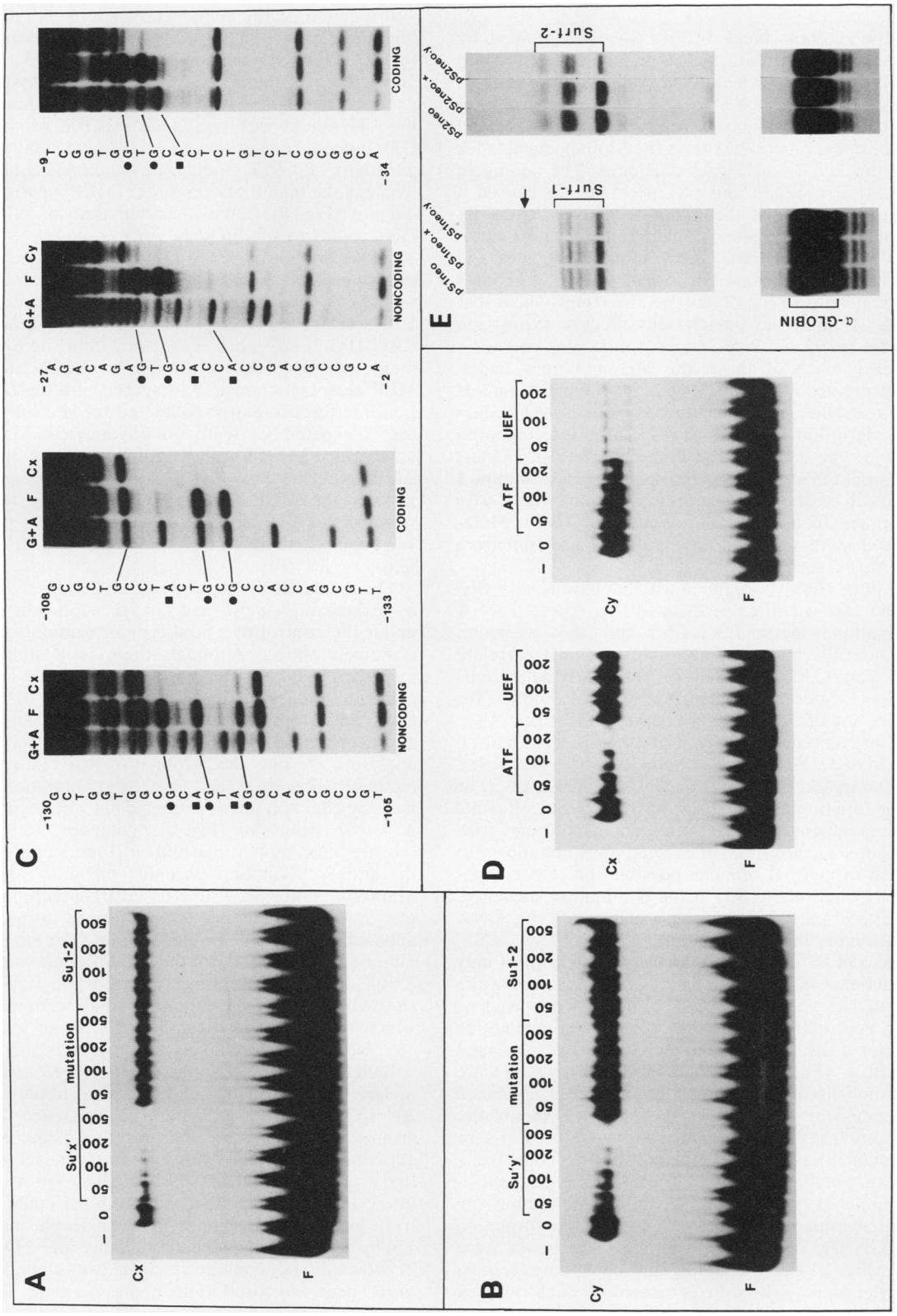

A
Su'x'
mutation
Su1-2
0
50
100
200
500
Cx
F
B
Su'y'
Cy
C
G+A
F
Cx
Cy
CODING
NONCODING
D
ATF
UEF
E
Surf-1
Surf-2
α-GLOBIN
pS1neo
pS1neo.x
pS1neo.y
pS2neo
pS2neo.x
pS2neo.y

FIG. 7. Analysis of sites Su'x' and Su'y' in vitro and in vivo. (A) Gel retardation analysis of site Su'x' with competition by wild-type and mutated Su'x' sequences. Labeled synthetic oligonucleotide probe from −104 to −133 encompassing site Su'x' was incubated with HeLa nuclear extract in the absence (0) or presence of competitor oligonucleotide (50, 100, 200, and 500 times molar excess). Complex Cx was resolved from free probe (F) by electrophoresis on nondenaturing gels. Complex Cx was competed for by the addition of wild-type Su'x', mutated Su'x' (see Fig. 3), or site Su1-2 sequences. Probe in the absence of extract is also indicated (−). (B) Gel retardation analysis of site Su'y' (+3 to −36); as for panel A with competition of Cy by wild-type site Su'y', mutated Su'y' (see Fig. 3), and site Su1-2. (C) DMS methylation interference patterns for sites Su'x' and Su'y'. Wild-type Su'x' and Su'y' probes were 5' end labeled on the coding or noncoding strands, partially methylated with DMS, and incubated with HeLa nuclear extract. The methylation interference patterns were then determined for complexes Cx, Cy, and the free probes (F). Maxam-Gilbert G+A sequencing reactions are also shown for each probe. The results are summarized on the adjacent sequences; circles indicate guanine bases and squares represent adenine bases required for complex formation. (D) Identification of Su'x' and Su'y' as ATF and UEF (USF, MLTF) binding sites, respectively. Labeled, wild-type site Su'x' or Su'y' was incubated with HeLa extract, and complexes Cx and Cy competed for by oligonucleotides (50, 100, and 200 times molar excess) to an ATF or UEF binding site. F, Free probe for Su'x' and Su'y'. Control lanes of probe in the absence of extract (−) and complexes in the absence of competitor sequences (0) are indicated. (E) In vivo expression of *Surf-1*/*Surf-2* neomycin fusion genes by wild-type and site Su'x' and Su'y' mutated *Surf-1* and *Surf-2* promoter. NIH 3T3 cells were cotransfected with pSPα118 and pS1neo, pS1neo.x, pS1neo.y, pS2neo, pS2neo.x, or pS2neo.y (see Fig 5). Total cytoplasmic RNA was analyzed by S1 mapping using neomycin probe A to detect transcripts from the *Surf-2* start sites, neomycin probe B to detect transcripts from the *Surf-1* start sites, and an α-globin probe (see Fig. 5 and Materials and Methods). The arrow indicates a novel start site observed in RNA from cells transfected with pS1neo.y.

scripts by S1 mapping. Transcription from this construct is from both *Surf-1* and *Surf-2* bona fide start sites (Fig. 8). There is little resolution of the various *Surf-1* start sites because of the length of the CAT probe, and the increased usage of the *Surf-2* intermediate sites relative to the major start site was due to the deletion of site Su'z' as discussed above. As suggested by the constructs that determine the usage of *Surf-1* and *Surf-2* start sites individually and the endogenous levels of *Surf-1* and *Surf-2* (34), *Surf-1* and *Surf-2* are expressed at essentially equal levels. Therefore, the 154-bp fragment is sufficient to support the correct in vivo ratio of *Surf-1* to *Surf-2* transcription.

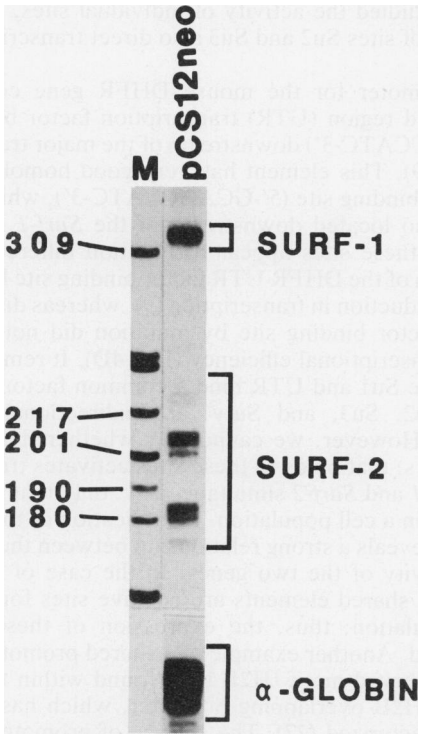


FIG. 8. Bidirectional transcription from the *Surf-1* and *Surf-2* promoter. NIH 3T3 cells were cotransfected with pSPα118 and pCS12neo (see Fig. 5). Total cytoplasmic RNA was analyzed by S1 mapping using neomycin probe C to detect transcripts from the *Surf-1* start sites, a probe derived from pCS12neo (*Nar*I to *Bam*HI) to detect transcripts from the *Surf-2* start sites, and an α-globin probe (see Materials and Methods). All probes used were double stranded. An *Msp*I cut pBR322 marker (M) is shown alongside the S1 mapping products.

DISCUSSION

Although several examples of close divergently transcribed genes have been identified in higher eukaryotes, e.g., dihydrofolate reductase (DHFR)/Rep-1 (20), collagen α1(IV) and α2(IV) (25), histone H2A and H2B (27), and HTF9 (18) genes, those that have been most fully studied are from prokaryotic, viral, and yeast systems. We were therefore interested in thoroughly investigating the promoters for the mouse *Surf-1* and *Surf-2* divergently transcribed genes (32). Five putative transcription factor binding sites have been characterized between and around the *Surf-1* and *Surf-2* major start sites (Su1, Su2, and Su3; Su'x' and Su'y'). The results of mutations of these sites on transcription in vivo are summarized in Fig. 9. Three of these promoter elements (sites Su2, Su3, and Su'y') are shared between *Surf-1* and *Surf-2* for efficient and accurate transcription of these genes, one element (site Su1) is required for accurate transcription from the *Surf-1* intermediate start sites only, and one binding site (site Su'x') was found to be nonfunctional in NIH 3T3 cells. However, site Su'x' may play a more significant role in regulating *Surf-1* and *Surf-2* expression under different con-

GENE	START SITE	Su'x'	Su1	Su2	Su3	Su'y'
SURF-1	major	—	—	↓	↓	—
	intermediate	—	↓	—	↓	⇣
SURF-2	major	—	—	↓	↓	↓
	intermediate and minor	—	—	↓	↓	—

FIG. 9. Diagrammatic summary of the in vivo expression results (Fig. 4D, 6D, and 7E). The degree of depression of transcription by mutations in the various binding sites is shown by the thickness of the arrow (the thickest arrow indicating the greatest degree of depression).

ditions or in different cell types. Thus, the *Surf-1* and *Surf-2* intergenic region can be considered as a single bidirectional promoter. We would expect that transcription factor binding sites within the *Surf-1* and *Surf-2* bidirectional promoter that are important for expression of these genes would be conserved through evolution. We are presently analyzing the homologous human promoter, and preliminary results show that the most critical promoter elements for *Surf-1* and *Surf-2* gene expression (Su1, Su2, and Su3) are conserved. Furthermore the first introns for both genes have no obvious homology, suggesting that there are no additional promoter elements proximal to the region from sites Su'x' to Su'y'.

Although site Su1 is required for accurate positioning of the *Surf-1* transcription start sites, this site does not appear to influence the total amount of steady-state *Surf-1* RNA. However, since the *Surf-1* translation start is not known, it remains possible that the first ATG (only 4 bp from the major start site) is the first codon (32); the new transcription start sites generated by mutation of site Su1 could result in truncated and potentially defective *Surf-1* polypeptides. Alternatively, if the first codon lies further downstream, truncated 5' untranslated RNA may affect translation, post-transcriptional modification, or RNA stability. Therefore, site Su1 may not have a quantitative affect on RNA levels but may quantitatively affect the amount of functional protein.

Site Su2 mutations are deleterious for transcription from the major *Surf-1* start site, while mutations in site Su1 are deleterious for transcription from the intermediate *Surf-1* start sites. Therefore, different *Surf-1* start sites are controlled by different promoter elements. A similar situation exists for *Surf-2* transcriptional start sites for which the site Su'y' mutation is deleterious for the major *Surf-2* start site and not the intermediate or minor sites. A number of genes, especially those lacking TATA box consensus sequences, contain multiple start sites. The functional importance, if any, of the use of multiple sites as opposed to a single start site is unknown, but usage of different start sites may be under the control of separate promoter elements. In the case of *Surf-1* and *Surf-2*, which are ubiquitously expressed (housekeeping) genes, it is feasible that in certain cell types, in different stages of the cell cycle, or under different conditions, certain transcription factors may vary in concentration. Thus, the differential usage of multiple start sites as a result of a change in control by using alternative sets of promoter elements would serve to maintain the expression of these housekeeping genes under a variety of conditions. The observation of control of usage of different start sites by alternative promoter elements has been previously described for the viral early gene EIIa promoter of Ad2 (35). Here we show that this mechanism is used for the major, intermediate, and minor start sites of the cellular *Surf-1* and *Surf-2* mouse genes.

Sites Su2 and Su3 were found to bind the same or similar factor(s) that was originally identified for two ribosomal protein genes, RPGα in mouse rpL32 (1) and XrpF1 in *Xenopus* rpL14 (5). *Surf-1* and *Surf-2* are unique genes (32) and are transcribed at a low level (34), and thus they most likely do not encode ribosomal proteins, which are members of multigene families containing processed pseudogenes and are transcribed at a high level (14). Therefore, it would appear that RPGα/XrpF1 binding sites are not specific to ribosomal protein genes but may occur in a number of housekeeping genes.

A sequence similar to site Su2 (5'-ACTTCCG-3') is present in the polyomavirus enhancer at the binding site for PEA3 (5'-ACTTCCT-3') (reference 29 and references therein). We find that the PEA3 sequence is able to compete for the Su2 and RPGα complexes, but a far greater excess of PEA3 oligonucleotide is required than of Su2 or RPGα oligonucleotide for effective competition (data not shown). Furthermore, the PEA3 sequence formed a comparatively weak complex that comigrated with the Su2 and RPGα complexes, which was very efficiently competed for by Su2 or RPGα sequences. Therefore, it is possible that the polyomavirus PEA3 binding site is a low-affinity site for the same or similar factor(s) that binds to Su2.

Although a few exceptions exist (e.g., triosephosphate isomerase [3]), the so-called housekeeping genes are characterized by a high G+C content, the lack of a canonical TATA box sequence, and multiple transcription start sites, and they are often expressed at a low level (22). These are also properties of the *Surf-1* and *Surf-2* genes but are not general properties of divergent genes. Although the collagen α1(IV) and α2(IV), DHFR/Rep-1, and HTF9 genes do not have consensus TATA box elements, the chicken histone H2A and H2B genes do have consensus TATA boxes. The TATA-box-binding factor(s) (TFIID [24], BTF1 [6]) directs transcription to a single start site at about 30 bp downstream of its binding site. In the absence of a TFIID binding site, other factors, e.g., Sp1 (28), GCN4 (7), and CP1 (28), that in a different context act primarily as activators of transcription can determine the start sites of transcription for housekeeping genes as well as controlling the efficiency of transcription. We have shown that site Su1 plays a role in directing the start of *Surf-1* transcription; however, since we have not studied the activity of individual sites, we cannot say which of sites Su2 and Su3 also direct transcription start sites.

The promoter for the mouse DHFR gene contains an untranslated region (UTR) transcription factor binding site (5'-GCTGCCATC-3') downstream of the major transcription start site (9). This element has very good homology to the Su1 factor binding site (5'-GCAGCCATC-3'), which intriguingly is also located downstream of the *Surf-1* start sites. However, these sites appear to function differently, since destruction of the DHFR UTR factor binding site by deletion caused a reduction in transcription (9), whereas disruption of the Su1 factor binding site by mutation did not affect the overall transcriptional efficiency (Fig. 4D). It remains possible that site Su1 and UTR bind a common factor.

Sites Su2, Su3, and Su'y' are bidirectional promoter elements. However, we cannot say whether the transcription factor(s) that binds to these sites activates transcription from *Surf-1* and *Surf-2* simultaneously, only that both genes are active in a cell population. Identification of these shared elements reveals a strong relationship between the transcriptional activity of the two genes. In the case of *Surf-1* and *Surf-2*, the shared elements are positive sites for transcriptional regulation; thus, the expression of these genes is coregulated. Another example of a shared promoter element is the conserved motif (H2B box) found within the histone H2A and H2B overlapping promoter, which has been partially characterized (27). The sharing of promoter elements for *Surf-1* and *Surf-2* gene expression strongly supports the hypothesis for a functional relationship between the *Surf-1* and *Surf-2* gene products. Functional relationships between the products of divergent genes are often apparent; they may encode enzymes involved in the same metabolic pathway (e.g., *Saccharomyces cerevisiae GAL-1* and *GAL-10* genes [16]), polypeptides that interact to form the quaternary structure of a protein or protein complex, [e.g., collagen

α1(IV) and α2(IV) genes (4, 25) and chicken histone H2A and H2B genes (27)], or one product that regulates the expression of the other product (e.g., *Escherichia coli araC* and *araBAD* genes [8]). We are presently analyzing the *Surf-1* and *Surf-2* gene products to determine their possible functional relationship.

ACKNOWLEDGMENTS

We are indebted to Iain Goldsmith for providing the numerous synthetic oligonucleotides that this work relied on and to Moira Read for her excellent technical assistance in cell culture. We also thank Steve Goodbourn and Nic Jones for providing plasmids and Steve Goodbourn, Nic Jones, and Richard Triesman for critical reading of the manuscript.

REFERENCES

1. **Atchison, M. L., O. Meyuhas, and R. P. Perry.** 1989. Localization of transcriptional regulatory elements and nuclear factor binding sites in mouse ribosomal protein gene rpL32. Mol. Cell. Biol. **9**:2067–2074.
2. **Baty, D., H. A. Barrera-Saldana, R. D. Everett, M. Vigneron, and P. Chambon.** 1984. Mutational dissection of the 21 bp repeat region of the SV40 early promoter reveals that it contains overlapping elements of the early-early and late-early promoters. Nucleic Acids Res. **12**:915–932.
3. **Boyer, T. G., J. R. Krug, and L. E. Maquat.** 1989. Transcriptional regulatory sequences of the housekeeping gene for human triosephosphate isomerase. J. Biol. Chem. **264**:5177–5187.
4. **Burbelo, P. D., G. R. Martin, and Y. Yamada.** 1988. α1(IV) and α2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. Proc. Natl. Acad. Sci. USA **85**:9679–9682.
5. **Carnevali, F., C. La Porta, V. Ilardi, and E. Beccari.** 1989. Nuclear factors specifically bind to upstream sequences of a *Xenopus laevis* ribosomal protein gene promoter. Nucleic Acids Res. **17**:8171–8184.
6. **Cavallini, B., I. Faus, H. Matthes, J. M. Chipoulet, B. Winsor, J.-M. Egly, and P. Chambon.** 1989. Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. Proc. Natl. Acad. Sci. USA **86**:9803–9807.
7. **Chen, W., and K. Struhl.** 1989. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. EMBO J. **8**:261–268.
8. **Dunn, T. M., S. Hahn, S. Ogden, and R. Schleif.** 1984. An operator at −280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. Proc. Natl. Acad. Sci. USA **81**:5017–5020.
9. **Farnham, P. J., and A. L. Means.** 1990. Sequences downstream of the transcription initiation site modulate the activity of the murine dihydrofolate reductase promoter. Mol. Cell. Biol. **10**: 1390–1398.
10. **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.
11. **Gilman, M.** 1987. Preparation of cytoplasmic RNA from tissue culture cells, p. 4.1.2–4.1.6. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, N.Y.
12. **Greene, J. M.** 1987. S1 analysis of messenger RNA using single-stranded DNA probes, p. 4.6.1–4.6.9. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York.
13. **Hurst, H. C., and N. C. Jones.** 1987. Identification of factors that interact with the E1a-inducible adenovirus E3 promoter. Genes Dev. **1**:1132–1146.
14. **Huxley, C., T. Williams, and M. Fried.** 1988. One of the tightly clustered genes of the mouse surfeit locus is a highly expressed member of a multigene family whose other members are predominantly processed pseudogenes. Mol. Cell. Biol. **8**:3898–3905.
15. **Huxley, C., and M. Fried.** 1990. The mouse surfeit locus contains a cluster of six genes associated with four CpG-rich islands in 32 kiloases of genomic DNA. Mol. Cell. Biol. **10**:605–614.
16. **Johnston, M., and R. W. Davies.** 1984. Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4**:1440–1448.
17. **Kingston, R. E.** 1987. Calcium phosphate transfection, p. 9.1.1–9.1.4. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, New York.
18. **Lavia, P., D. Macleod, and A. Bird.** 1987. Coincident start sites for divergent transcripts at a randomly selected CpG-rich island of mouse. EMBO J. **6**:2773–2779.
19. **Lennard, A. C., and J. M. Egly.** 1987. The bidirectional upstream element of the adenovirus-2 major late promoter binds a single monomeric molecule of the upstream factor. EMBO J. **6**:3027–3034.
20. **Linton, J. P., J.-Y. J. Yen, E. Selby, Z. Chen, J. M. Chinsky, K. Liu, R. E. Kellems, and G. F. Crouse.** 1989. Dual bidirectional promoters at the mouse *dhfr* locus: cloning and characterization of two mRNA classes of the divergently transcribed *Rep-1* gene. Mol. Cell. Biol. **9**:3058–3072.
21. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Melton, D. W.** 1987. Strategies and mechanisms for the control of transcriptional initiation of mammalian protein-coding genes. J. Cell Sci. **88**:267–270.
23. **Miyamoto, N. G., V. Moncollin, J. M. Egly, and P. Chambon.** 1985. Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J. **4**:3563–3570.
24. **Nakajima, N., M. Horikoshi, and R. G. Roeder.** 1988. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity and TATA box-promoter interactions of TFIID. Mol. Cell. Biol. **8**:4028–4040.
25. **Poschl, E., R. Pollner, and K. Kuhn.** 1988. The genes for the α1(IV) and α2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. EMBO J. **7**:2687–2695.
26. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
27. **Sturm, R. A., S. Dalton, and J. R. E. Wells.** 1988. Conservation of histone H2A/H2B intergene regions: a role for the H2B specific element in divergent transcription. Nucleic Acids Res. **16**:8571–8586.
28. **Taylor, I. C. A., and R. E. Kingston.** 1990. Factor substitution in a human HSP70 gene promoter: TATA-dependent and TATA-independent interactions. Mol. Cell. Biol. **10**:165–175.
29. **Wasylyk, C., P. Flores, A. Gutman, and B. Wasylyk.** 1989. PEA3 is a nuclear target for transcription activation by non-nuclear oncogenes. EMBO J. **8**:3371–3378.
30. **Weeks, D. L., and N. C. Jones.** 1983. Ela control of gene expression is mediated by sequences 5′ to the transcriptional starts of the early viral genes. Mol. Cell. Biol. **3**:1222–1234.
31. **Wildeman, A. G., P. Sassone-Corsi, T. Grundstrom, M. Zenke, and P. Chambon.** 1984. Stimulation of *in vitro* transcription from the SV40 early promoter by the enhancer involves a specific *trans*-acting factor. EMBO J. **3**:3129–3133.
32. **Williams, T. J., and M. Fried.** 1986. The MES-1 murine enhancer element is closely associated with the heterologous 5′ ends of two divergent transcription units. Mol. Cell. Biol. **6**:4558–4569.

33. **Williams, T., 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.
34. **Williams, T., J. Yon, C. Huxley, and M. Fried.** 1988. The mouse surfeit locus contains a very tight cluster of four "housekeeping" genes that is conserved through evolution. Proc. Natl. Acad. Sci. USA **85**:3527–3530.
35. **Zajchowski, D. A., H. Boeuf, and C. Kedinger.** 1985. The adenovirus-2 early EIIa transcription unit possesses two overlapping promoters with different sequence requirements for EIa-dependent stimulation. EMBO J. **4**:1293–1300.
36. **Zhang, H., R. Scholl, J. Browse, and C. Somerville.** 1988. Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res. **16**:1220.