Gene Expression, Vol. 6, pp. 209-218, 1996
Printed in the USA.

1052-2166/96 $10.00 + .00

Tissue-Specific Processing of the Surf-5 and Surf-4 mRNAs

KEN GARSON,[1] TREVOR DUHIG, AND MIKE FRIED[2]

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

The mouse surfeit locus is an unusually tight cluster of at least six "housekeeping" genes that do not share any sequence homology and whose gene organization may play a role in gene expression. The transcription of each of the five well-characterized genes (Surf-1 to -5) alternates with respect to its neighbor(s) and no more than 159 bp separates any two adjacent genes with the Surf-4 and Surf-2 genes overlapping at their 3′ ends by 133 bp. In this work, the expression of the Surf-5 and Surf-4 genes has been examined in various mouse tissues. In addition to the ubiquitously expressed 3.5-kb Surf-5 mRNA, a second alternatively spliced Surf-5 mRNA, Surf-5b, was discovered that was highly expressed in the brain, heart, testis, and skeletal muscle. The alternative splice donor site of the Surf-5b mRNA is similar to splice donor sites found in neuron-specific mRNAs. Surf-5b encodes a unique protein, which, like the ubiquitous Surf-5 protein, has been found to be primarily located in the soluble fraction of the cytoplasm. The expression of the Surf-5b protein was also found to increase in embryonal carcinoma cells differentiated into neuronal cultures. Although the Surf-5 gene is highly conserved through evolution, the presence of the Surf-5b alternative splice may be restricted to higher vertebrates. The Surf-4 gene was ubiquitously expressed in eight different mouse tissues; however, the ratios of the three previously reported Surf-4 mRNAs (two of which are known to derive from different sites of polyadenylation) altered dramatically between tissues. The use of different forms of mRNA processing for regulation of tissue-specific expression of ubiquitously expressed genes is discussed.

Surf-4 mRNA Surf-5 mRNA Tissue-specific processing Mouse surfeit locus

THE mouse Surfeit locus (Fig. 1) is a very compact cluster of at least six "housekeeping" genes with unique properties that define a novel type of gene cluster (10). The 5′ end of each gene is associated with a CpG-rich island. The five characterized genes (Surf-1 to -5) are unrelated by sequence homology. A maximum of 159 bp separates any of the adjacent well-characterized Surfeit genes (Fig. 1). More specifically, the heterogeneous 5′ ends of the Surf-1 and Surf-2 genes are separated by only 15–73 bp (25), the heterogeneous 5′ ends of the Surf-5 and Surf-3 genes are separated by 159 bp (6), the 3′ ends of the Surf-3 and Surf-1 genes are separated by 70 bp (11), and the 3′ ends of the Surf-2 and Surf-4 genes overlap by 133 bp (23) (see Fig. 1). This extremely close juxtaposition of genes suggests that gene expression may be regulated by mechanisms involving the sharing of 5′ regulatory elements, promoter occlusion, and possible antisense regulation. One well-characterized example of shared regulatory elements within this locus is the regulation of the expression of the Surf-1 and Surf-2 genes by a bidirectional promoter (7,12,13). Both the organization of the Surfeit locus and the juxtaposition of the Surfeit genes are conserved between man, mouse, and chicken (4,24,26).

Examination by Northern blot analysis of the expression of the mouse surfeit genes in a variety of tissue culture cell lines indicated that the Surf-1

Received June 5, 1996; revision accepted December 2, 1996.
[1]Present address: Ottawa Centre for Cancer Research, Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, Canada, K1H 8M5.
[2]Address correspondence to Mike Fried. Tel: 44-171-269-3297; Fax: 44-171-269-3581; E-mail: fried@icrf.icnet.uk

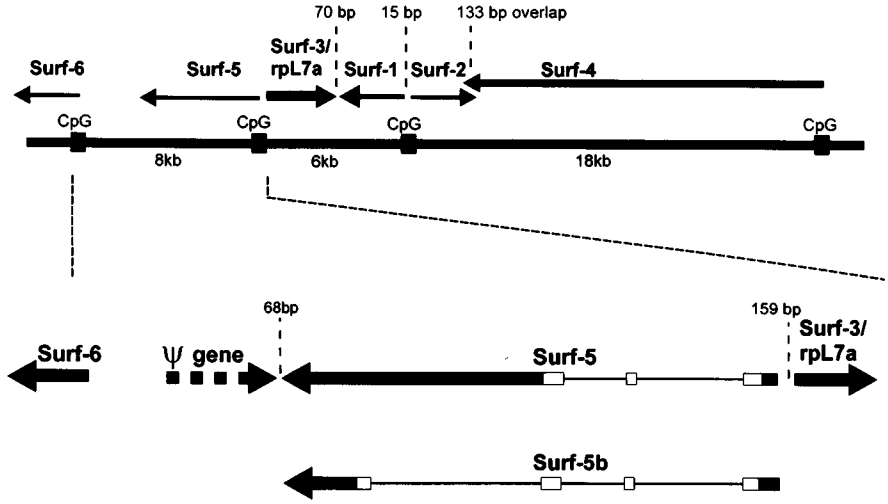

FIG. 1. Map of the mouse surfeit cluster. The genomic DNA at the top is represented as a solid bar with the distances between adjacent CpG islands (black boxes) indicated below. The directions of transcription of the surfeit genes (Surf-1 to -6) are indicated with arrows and the intergenic distances given in base pairs. The Surf-5 region has been expanded to show the intron/exon structure of the Surf-5 and Surf-5b transcripts and its relationship to the Surf-3/rpL7a gene and a potential pseudogene (Ψ) located near its 3′ end (6). Noncoding Surf-5 exon sequences are indicated by a thicker black box whereas Surf-5 coding sequences are shown as speckled boxes.

to -4 genes were ubiquitously expressed (24). The downregulation of Surf-4 expression in undifferentiated but not differentiated F9 cells (24) indicated that the expression of Surf-4 may be subject to transcriptional control. In addition, the Surf-4 gene may be subject to posttranscriptional control as three mRNAs of 2.8, 2.0, and 1.4 kb have been identified for this gene (10). The 2.8-kb mRNA has been cloned and sequenced (10) and encodes a protein shown to be an integral membrane protein associated with the endoplasmic reticulum (ER) (18). The 3′ end of the less abundant Surf-4 2.0-kb mRNA was mapped and found to use an alternative upstream polyadenylation signal (10). The Surf-4 2.0-kb mRNA retains the same open reading frame as the 2.8-kb mRNA, but has a shorter 3′ untranslated region that would not overlap the adjacent converging Surf-2 mRNA as has been observed for the 2.8 kb Surf-4 mRNA (23) (see Fig. 1). The structure of the minor 1.4-kb Surf-4 mRNA has not yet been established.

The mouse Surf-5 gene has recently been characterized (6) and found to express a single 3.5-kb mRNA in fibroblast cell lines. The open reading frame of the Surf-5 3.5-kb mRNA specifies a hydrophilic polypeptide product of 140 amino acids. This Surf-5 protein product is primarily located in the soluble fraction of the cytoplasm. The Surf-5 gene is highly conserved through evolution. In this report we have examined the expression of the Surf-5 gene in a variety of mouse tissues. We find that in addition to the ubiquitously expressed Surf-5 3.5-kb mRNA an alternatively spliced 1.5-kb Surf-5 mRNA (Surf-5b), which can specify a 200-amino acid polypeptide, is expressed to a high degree in the brain, heart, testis, and skeletal muscle. The alternative splice acceptor site of Surf-5b has the properties of a neuron-specific splice. A cDNA derived from the Surf-5b mRNA was cloned and sequenced and the subcellular location of the novel protein product expressed from this mRNA was determined. In addition we further examined the differential expression of the three Surf-4 mRNAs in a variety of mouse tissues. We found that the relative ratios of the three Surf-4 mRNAs varied dramatically between different tissues.

MATERIALS AND METHODS

Cell Culture

Mouse NIH 3T3 and Neuro 2A cells were grown in E4 medium supplemented with 10% fetal calf serum. P19 embryonal carcinoma cells were cultured as previously described (19). The differentiation of P19 cells into enriched neuronal cultures was in serum-free media as described by MacPherson et al. (14).

Isolation of Poly(A)$^+$ RNA and Northern Blotting

Poly(A)$^+$ mRNA was isolated from NIH 3T3, Neuro 2A cells, or mouse brain using the Fast Track mRNA isolation kit (In Vitrogen). Total RNA was isolated from P19 cells before and after differentiation into neurons using the TRIzol reagent (Gibco/BRL). RNA was fractionated by electrophoresis on a 1% agarose gel containing formaldehyde essentially as described (20). The RNA was transferred from the gel onto Hybond N membrane (Amersham) by capillary transfer according to the manufacturer's instructions. The resulting membranes and a multiple tissue Northern blot obtained from Clonetech were hybridized with appropriate probes in 5 × SSPE (0.9 M NaCl, 50 mM NaH_2PO_4, 5 mM EDTA, pH 7.4), 10 × Denhardts (0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2 % bovine serum albumin, Pentax Fraction V), 50% formamide, 2% SDS, and 100 μg/ml denatured herring sperm DNA at 42°C overnight. Membranes were washed to a final stringency of 0.1 × SSC, 0.1% SDS at 65°C. Signals from the hybridized membranes were quantitated on a Molecular Dynamics phosphorimager using MD Imagequant software (v3.0). Probes used for the hybridization included a near full-length probe, a coding region probe, and a 3′ untranslated region probe corresponding to nucleotide positions 154–3444, 227–649, and 2775–34444 of the Surf-5 mRNA, respectively.

Isolation and Sequencing of a Surf-5b cDNA

A mouse brain (BALB/C postnatal day 20, whole brain) oligo (dT) primed cDNA library in the Uni-ZAP XR vector (Stratagene) was plated as recommended for Lambda ZAP (Stratagene). Plaques were lifted onto Hybond N membrane (Amersham) and then denatured for 2 min in 1.5 M NaCl, 0.5 M NaOH, neutralized for 2 min in 1.5 M NaCl and 0.5 M Tris, pH 7.4, and washed for 2 min in 2 × SSC. After air drying, membranes were hybridized as described for Northern blots. Hybridizing plaques were plaque purified and the cDNAs rescued on bluescript phagemids as recommended for Lambda ZAP and sequenced directly using Sequenase (United States Biochemical) following the manufacturer's instructions. The human cDNA was isolated from a fetal brain (17-week-old embryo) oligo (dT) primed library in the pSPORT1 vector (kindly provided by Dr. Sebastian Meier-Ewert).

Analysis of DNA and protein sequences used the Genetics Computer Group software (5), and data base searches for protein homologies were done using FASTA (17), Blast (1), and searches by the method of Collins et al. (3). Prosite (2) and Blocks (9) data bases were used to search for potential motifs in the protein sequence.

Generation of Surf-5b Antibodies and Western Blotting

An *Ava*I fragment from the GST/Surf-5 expression vector, pGST-Surf-5 (6), which included all but the most immediate 5′ sequence of the Surf-5 gene, was replaced with a corresponding *Ava*I fragment from the Surf-5b cDNA. The resulting plasmid, pGST-Surf-5b, encoded the GST open reading frame in-frame with the full open reading frame of the Surf-5b gene. GST fusion proteins were purified from bacteria harboring this plasmid following induction with IPTG using glutathione agarose (Sigma) following methods previously described (21). The Surf-5b open reading frame was transferred from pGST-Surf-5b on an *Eco*R1/*Xho*1 fragment to the vector pMal-cR1 (New England BioLabs) to allow the production of a Surf-5b fusion with the maltose binding protein (MBP) as recommended by the manufacturer. Purified GST/Surf-5b fusion protein (500 μg) mixed 1:1 with complete Freund's adjuvant was injected at multiple subcutaneous sites into rabbits followed by boosts administered in a similar fashion except with incomplete Freund's adjuvant every 4 weeks. Test bleeds were taken 7 days following any given immunization. Surf-5b antibodies were purified from crude sera by affinity chromatography on MBP/Surf-5b protein sepharose columns. The column matrix was prepared by covalent linkage of purified MBP/Surf-5b fusion proteins to CNBr-activated Sepharose 4B (Pharmacia) following the manufacturer's recommendations.

Protein fractions were purified from NIH 3T3 and Neuro 2A cells as follows. Confluent 100-mm dishes of cells were washed two times in cold PBS, scraped in 1 ml of cold PBS, and pelleted at 650 × *g* for 4 min. The cell pellet was resuspended in 10 mM Tris, pH 7.5, 1 mM EDTA, 3 mM $CaCl_2$, 1 mM PMSF, 2 μg/ml leupeptin, 2 μg/ml pepstatin, 5 μg/ml aprotinin, and rested on ice for 15 min. Cells were then lysed by 40 strokes of dounce homogenization and nuclei pelleted at 650 × *g* for 4 min. The supernatant (cytoplasmic fraction) was retained for direct use in Western blot assays for further fractionation. The nuclear pellet was

resuspended in 800 μl of homogenization buffer supplemented with NP40 to a final concentration of 1% and incubated on ice for a further 5 min. Nuclei were pelleted at 650 × *g* for 4 min. The cytoplasmic fractions were further fractionated by spinning at 30,000 × *g* for 30 min in a TLA-100.2 rotor. The pellet (P30) was retained for Western blot analysis and the supernatant was centrifuged at 100,000 × *g* for 3 h. The pellet (P100) and supernatant (S100) were also analyzed by Western blot analysis. The cytoplasmic fractions as described above were similarly prepared for P19 cells and neuronal cultures of differentiated P19 cells 8 days following treatment with retinoic acid.

Protein samples were separated on 15% PAGE and then transferred onto Hybond C extra membrane for Western blot analysis. Membranes were blocked in TBS (150 mM NaCl, 20 mM Tris, pH 8.3) containing 5% skimmed milk powder for 30 min and then incubated with 1 ml of primary antibody (0.2 μg) in 5% skimmed milk powder, 0.5% Tween 20 for 2 h. Following three washes in antibody buffer, the secondary antibody, HRP-conjugated goat anti-rabbit (Tago Immunologicals) was incubated for a further 2 h. Following three washes as above and a final wash in TBS alone, the signal was observed by chemiluminescence using the Amersham ECL detection reagents.

RESULTS

The Expression of the Mouse Surf-5 Gene in Different Mouse Tissues

To analyze the expression of the Surf-5 gene in different mouse tissues, a multiple tissue Northern blot was hybridized with a near full-length Surf-5 cDNA. In addition to the 3.5-kb Surf-5 mRNA previously identified (6), a 1.5-kb mRNA, Surf-5b, was also found to be predominant in several tissues (Fig. 2A). Based on phosphoimage analysis, after correcting for mRNA sizes, the smaller Surf-5b mRNA was found to be expressed at less than 9% (spleen) to 210% (brain) relative to the level of the 3.5-kb mRNA. In fact, the Surf-5b mRNA was found to be the predominant mRNA in the heart, brain, and skeletal muscle, and almost equivalent in amount to the 3.5-kb mRNA in testis. The expression of the Surf-5b mRNA in all other tissues was extremely low in comparison to the 3.5-kb mRNA.

Isolation and Sequencing of a Surf-5b cDNA

To define the relationship of the Surf-5b transcript to the previously defined Surf-5 transcript, a mouse brain cDNA library in Lambda Zap was plated out and 200,000 plaques screened with a near full-length Surf-5 cDNA probe (see Materials and Methods). Three positive clones were plaque purified and analyzed following rescue of the bluescript plasmids. All three cDNAs were derived from the 1.5-kb Surf-5b mRNA and the longest cDNA was sequenced fully on one strand and compared to the sequence of Surf-5 cDNA previously reported (6).

The sequence presented in Fig. 3A shows the sequence of the Surf-5b cDNA with additional 5′ sequences added to include sequence up to the longest mapped 5′ end of the Surf-5 gene as previously reported (6). Comparison of the Surf-5b sequence with that previously reported for Surf-5 indicated that the Surf-5b mRNA results from the use of alternative splice donor and acceptor sites located in the last exon of the Surf-5 mRNA at bases 640 and 2736, respectively. As this alternative splice donor lies seven bases before the termination codon of the Surf-5 mRNA, the open reading frame of the Surf-5b mRNA continues into sequence following the new splice junction (Fig. 3B). As a result, the predicted protein sequence of Surf-5b has identical amino acid sequence with the first 137 amino acids of the Surf-5 protein sequence, but encodes an additional 63 C-terminal amino acids distinct from the three remaining C-terminal amino acids encoded by Surf-5 (see Fig. 1). Data base searches for related sequences, or for motifs present in the Surf-5b amino acid sequence, failed to identify any related genes.

Conservation of the Surf-5b Differential Spliced Transcript in Humans

The 140-amino acid protein encoded by the previously defined Surf-5 transcript is highly conserved (98.5%) between the mouse and humans (6). To see if the Surf-5b transcript was also highly conserved between mouse and man, 100,000 clones of a human fetal brain cDNA library were screened with a probe from the human Surf-5 coding region. Two positive clones containing an insert of approximately 1.5 kb were isolated. DNA sequence analysis revealed that both of these clones specified the Surf-5b transcript containing a splice join at the identical location of the mouse Surf-5b transcript and encoding an additional 63 amino acids, which shows high homology to the equivalent region of the mouse Surf-5b coding region (Fig. 4).

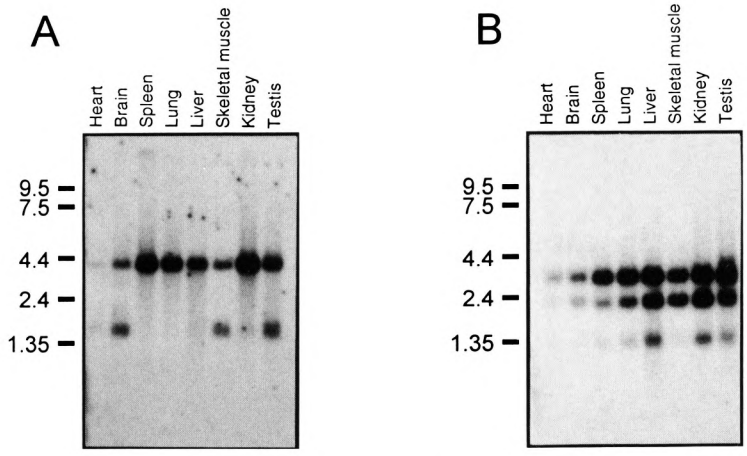

FIG. 2. Multiple-tissue Northern blot. A multiple-tissue Northern blot was hybridized with a Surf-5 near full-length cDNA probe (A) or a Surf-4 coding region probe (B). Poly(A)$^+$ mRNA (2 μg) isolated from the heart, brain, spleen, lung, liver, skeletal muscle, kidney, and testis were separated on each lane. The sizes (kilobases) and positions of molecular weight markers are shown.

Differential Expression of the Mouse Surf-5b Transcript

To confirm the differential expression of the Surf-5b transcript in the brain but not fibroblasts, mRNA was isolated from mouse brain and NIH 3T3 cells and analyzed by Northern blot analysis with a full-length mouse Surf-5 cDNA probe (Fig. 5A). Even on lengthy exposures, only the 3.5-kb mRNA was evident in NIH 3T3 cells, whereas the 1.5-kb mRNA appeared to be the predominant species expressed in mouse brain. Based on the high expression of the Surf-5b mRNA in the brain, we sought to determine if this mRNA was also expressed in the mouse neuroblastoma cell line Neuro 2A. A Northern blot of mRNA from Neuro 2A cells was hybridized with a probe derived from the coding region of the Surf-5 cDNA. As shown in Fig. 5B, the Surf-5b mRNA also appears as the predominant mRNA in this neuroblastoma cell line.

Identification of the Surf-5b Gene Product

To verify that the Surf-5b transcript expressed a stable protein, antisera were raised in rabbits to GST fusions of the Surf-5b coding region to permit the immunological detection of the Surf-5b protein product. The Surf-5b antibodies were affinity purified on columns with MBP/Surf-5b fusion proteins. The purified antibodies were used to Western blot fractionated protein from mouse cell lines. NIH 3T3 cells, which express only the 3.5-kb mRNA, and Neuro 2A cells, which express both mRNAs, were used for comparison in this assay (Fig. 6). The expression of the protein products paralleled the expression pattern of the mRNAs. Whereas the Surf-5 protein was expressed in the cytoplasm of both NIH 3T3 and Neuro 2A cells, the Surf-5b protein, which runs with an apparent molecular weight of 26 kDa, was expressed at detectable levels only in the Neuro 2A cells (Fig. 6A). Analysis of the cytoplasmic fraction with antibodies to a nuclear protein revealed no discernible cross-contamination (data not shown). Further fractionation of the cytoplasmic fraction revealed that the Surf-5b protein also resides in the soluble S100 fraction of the cytoplasm as reported for the Surf-5 protein (Fig. 6B) (6). Unfortunately, the Surf-5b antibodies were unable to immunoprecipitate the Surf-5b protein product from cell lysates and gave inconsistent results when used in indirect immunofluorescence experiments.

The Expression of the Surf-5b Protein Is Regulated by the Differentiation of Mouse P19 Teratocarcinoma Cells Into Neuronal Cells

We determined whether the expression of the two Surf-5 proteins might be regulated during differentiation. We chose to examine the Surf-5 mRNAs and proteins expressed in undifferentiated P19 teratocarcinoma cells and in neuronal cultures derived from P19 teratocarcinoma cells following differentiation induced with retinoic acid. A Northern blot of total RNA from differentiated and undifferentiated P19 cells was hybridized with a 3′ Surf-5 UTR probe, which was fully

A

```
1    AGGCGGGCGG GCCCTGGCAC CGGACCGCAG GCTCCGAGAG GCGGCGGGAG GCGCCAGGAG TGAACGGAAG CCGCGGGAAG GAGCTGCGGG TGCTTTGTTC   100
101  CCTCGCCGTA AGGGCCTGCT GCCGGGCGTG TGTTCGTGGA CGGAGCTGTT GTATCGGAAA TCCCGGTCCC CAGCCCTCAG TACTCGGTTC GGGCCGGACC   200
201  AGACTGGTTC TCGTCCAGGG CCCGCCATGG CCCAGCAGAG AGCCCTTCCC CAGAGCAAGG AGACGCTGCT GCAGTCATAC AACAAGCGGC TCAAAGACGA   300
                          M  A  Q  Q  R  A  L  P  Q  S  K  E  T  L  L  Q  S  Y  N  K  R  L  K  D  D
301  CATCAAGTCC ATCATGGACA ACTTCACCGA GATCATCAAG ACCGCCAAGA TTGAGGATGA GACGCAGGTG TCTCGGGCTA CTCAGGGCGA ACAGGACAAT   400
      I  K  S  I  M  D  N  F  T  E  I  I  K  T  A  K  I  E  D  E  T  Q  V  S  R  A  T  Q  G  E  Q  D  N
401  TATGAGATGC ACGTTCGAGC AGCCAACATC GTTCGAGCTG GTGAGTCACT GATGAAGCTG GTATCCGATC TCAAGCAGTT TCTGATCCTC AACGACTTCC   500
      Y  E  M  H  V  R  A  A  N  I  V  R  A  G  E  S  L  M  K  L  V  S  D  L  K  Q  F  L  I  L  N  D  F  P
501  CGTCTGTGAA TGAAGCCATC GACCAGCGTA ACCAGCAGCT TCGAGCCCTG CAGGAGGAGT GTGACCGCAA GCTCATCACC CTGCGGGACG AGGTCTCCAT   600
       S  V  N  E  A  I  D  Q  R  N  Q  Q  L  R  A  L  Q  E  E  C  D  R  K  L  I  T  L  R  D  E  V  S  I
601  TGACCTATAT GAGCTGGAGG AGGAGTACTA CTCGTCCAGC TCAAGTCTCT GCGAAGCTAA TGACCTGCCT CTGTGTGAAG CTTACTGGAG GCTGGACCTC   700
      D  L  Y  E  L  E  E  E  Y  Y  S  S  S  S  S  L  C  E  A  N  D  L  P  L  C  E  A  Y  W  R  L  D  L
701  GACGCAGATT CTGCTGATGG CCTCTCAGCC CCTCTTCTGG CGTCCCCGGA GACTGGTGCT GGTCCCTTAC AGTCTGCAGC CCCTGTCCAC TCCCATGGTG   800
      D  A  D  S  A  D  G  L  S  A  P  L  L  A  S  P  E  T  G  A  G  P  L  Q  S  A  A  P  V  H  S  H  G  G
801  GTGGCCCTGG CCCTACAGAA CACACCTGAG CCTCTTGGGG TACCATCGCA GGGTAGCCCT CCCTAGGGCC TAGTTTATCA GAAAGGAAGT GCTGTGGATA   900
       G  P  G  P  T  E  H  T  *
901  CCCAGCACAG CCTAGCTTTT TTCAGTTTTA GCTCCCTCTG GGGACTTTGG GATCAGGAGC CACTGTAGCC TCAGGAACCT GTCCAGGCCA GGGCAGTTCC  1000
1001 CCGTGATGGT TGGGGAATGC TGTCCAGTTA TTTACAGCAG GTGGCTCCCT TCCTGACTGC CAGTGGAAGT TTGTCCCATT GAAGACCTTG AATAGGGAGG  1100
1101 GTGGTTGGTG ATAGGAAGAC AGCTCAGTTT GTTTAGAATT GTGAGGTATT CAGTGGACCT TAGTCAAGTC CAGTTACCCA TCCTTAGGTT CAGGTTCTAT  1200
1201 AGAGAGCCCT GTCCACTGCC TGCTGCCTGG AGCATAAATG TCTCTGCCAC AGCTGACTGT CCAAGATTGT GTGAGAGGCA GCTAGCCAGG GTGTCTTCCT  1300
1301 GGATTCTCCC ACCATTTCCT TCTGCTACTT TTGAGATGAG ACTGTTGAAT AAAACATTAG CTTTTAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA  1397
```

B

```
surf-5    E  L  E  E  E  Y  Y  S  S  R  Y  K  *
          GAG CTG GAG GAG GAG TAC TAC TCG TCC AGg tac aaa tag // gtggtggcaaagC TCA AGT CTC
surf-5b   E  L  E  E  E  Y  Y  S  S  S                          S  S  L
```

FIG. 3. Sequence of the Surf-5b cDNA. (A) The sequence of the surf-5b cDNA was based on the sequence from the cDNA and sequence previously reported for the 5′ UTR of Surf-5 up to the most 5′ mapped transcriptional start site. The position of the introns are denoted by the bold v shape. (B) The region around the Surf-5b-specific splice is shown. The letters above the DNA sequence indicate the amino acid sequence of the C-terminus of the Surf-5 protein and the letters beneath the DNA sequence indicate the amino acid sequence around the third splice of the Surf-5b protein (see A). The capital and small letters of the DNA sequence indicate the exon coding and intron regions of the Surf-5b protein, respectively.

```
mouse      1 MAQQRALPQSKETLLQSYNKRLKDDIKSIMDNFTEIIKTAKIEDETQVSR 50
             |||||||||||||||||||||||||||||||||||||||||||||||||||
human      1 MAQQRALPQSKETLLQSYNKRLKDDIKSIMDNFTEIIKTAKIEDETQVSR 50

mouse     51 ATQGEQDNYEMHVRAANIVRAGESLMKLVSDLKQFLILNDFPSVNEAIDQ 100
             |||||||||||||||||||||||||||||||||||||||||||||||||||
human     51 ATQGEQDNYEMHVRAANIVRAGESLMKLVSDLKQFLILNDFPSVNEAIDQ 100

mouse    101 RNQQLRALQEECDRKLITLRDEVSIDLYELEEEYYSSSSSLCEANDLPLC 150
             |||||||.|||||||||||||||:|||||||||||||||||||||||||||
human    101 RNQQLRTLQEECDRKLITLRDEISIDLYELEEEYYSSSSSLCEANDLPLC 150

mouse    151 EAYWRLDLDADSADGLSAPLLASPETGAGPLQSAAPVHSHGGGPGPTEHT 200
             ||| |||||.|||||||||||||||.:||||| |||.|||:||||||||.
human    151 EAYGRLDLDTDSADGLSAPLLASPEPSAGPLQVAAPAHSHAGGPGPTEHA 200
```

FIG. 4. A comparison of the amino acid sequence of Surf-5b in mouse and human. The amino acid sequences of the mouse and human Surf-5b proteins are shown with homologous amino acids indicated by vertical lines, nonconservative differences by spaces, and conservative differences by one or two dots. The first 137 amino acids are contained in the Surf-5 protein.

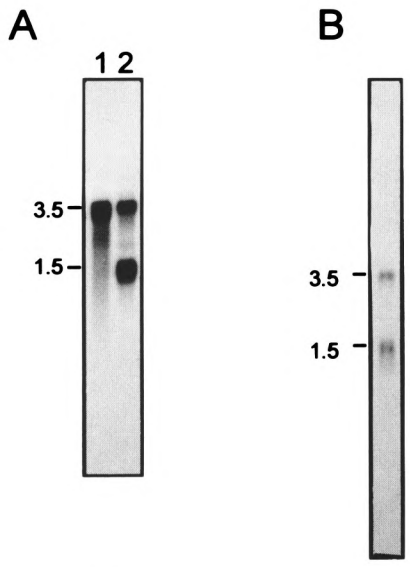

FIG. 5. Expression of the Surf-5b transcript in Neuro 2A cells. Poly(A)$^+$ RNA (2 μg) isolated from NIH 3T3 cells (A, lane 1), mouse brain (A, lane 2), or Neuro 2A cells (B) were separated on a 1% agarose gel containing formaldehyde, transferred to Hybond N, and probed with a full-length Surf-5 cDNA probe (A) or a Surf-5 coding region probe (B). The positions of the 3.5- and 1.5-kb transcripts are indicated.

represented in both the Surf-5 and Surf-5b mRNAs (Fig. 7A). Phosphoimage analysis revealed that the Surf-5 and Surf-5b mRNA were present at equal levels in the undifferentiated cells whereas the Surf-5b mRNA was expressed at four times the level of the Surf-5 mRNA in the neuronal cultures. In agreement with the Northern blot data, Western blot analysis (Fig. 7B) showed that the Surf-5b protein became the predominant species expressed in the differentiated neuronal cultures.

The Expression of the Mouse Surf-4 Gene in Different Mouse Tissues

The multiple mouse tissue Northern blot was hybridized with a probe derived from the mouse Surf-4 coding region. The results (Fig. 2B) revealed the presence of the three previously reported Surf-4 mRNAs of 2.8, 2.0, and 1.4 kb (10) in all mouse tissues, but the ratios of the three transcripts were found to vary dramatically between tissues. Based on quantitation using a phosphorimager, the 2.8-kb mRNA was the most predominant mRNA in all tissues. The 2.0-kb mRNA was expressed at levels ranging from 37% (spleen) to 93% (kidney) of the level of the 2.8-kb mRNA. Similarly, the 1.4-kb mRNA was expressed at levels ranging from less that 5% (brain) to 30% (liver) of the level of the 2.8-kb mRNA.

DISCUSSION

The mouse surfeit locus is a highly conserved gene cluster (4,24,26) containing at least six genes that do not share any sequence homology (6,10,11,15,25). At present it is unclear whether any of the genes have related functions. At least five of the surfeit genes have remained associated in the surfeit cluster throughout 600 million years of divergent evolution (4). The conservation of this cluster may be explained if the organization of the genes plays an important role in the control

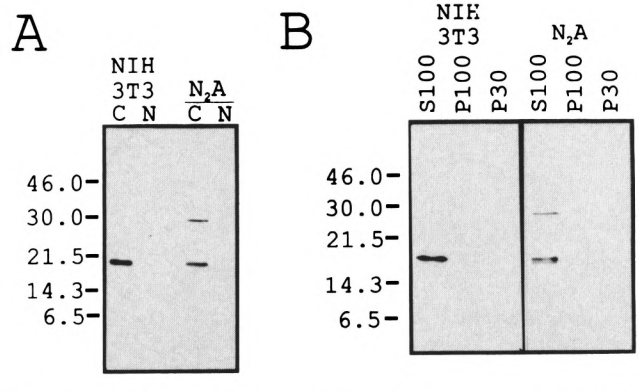

FIG. 6. Subcellular regulation of the Surf-5 proteins. (A) NIH 3T3 and Neuro 2A (N2A) cells were fractionated into nuclear and cytoplasmic fractions and Surf-5 proteins detected by Western blot analysis. (B) The cytoplasmic fractions from both cell lines were fractionated by differential centrifugation to yield the S100, P100, and P30 fractions and Surf-5 proteins detected by Western blot analysis. The positions and sizes in kilodaltons of prestained molecular weight markers are indicated (A + B).

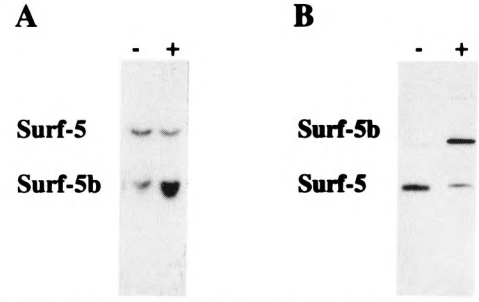

FIG. 7. Regulation of Surf-5b expression in differentiating P19 cells. (A) Total RNA from untreated (−) and day 8 retinoic acid-treated (+) P19 cells were separated on a 1% agarose gel containing formaldehyde, transferred to Hybond N, and probed with a 3′ UTR Surf-5 probe. The Surf-5 (3.5 kb mRNA) and the Surf-5b (1.5 kb) mRNAs are indicated. (B) Total cell lysates from untreated (−) and 8 day retinoic acid-treated (+) P19 cells were analyzed by Western blot analysis for the expression of Surf-5 proteins. The positions of the Surf-5 and Surf-5b proteins are indicated.

of their expression. In addition to possible forms of regulation as a result of the gene organization of the locus, other posttranscriptional mechanisms appear to exist to regulate the expression of at least Surf-5 and Surf-4.

One such posttranscriptional mechanism is the control of alternative splicing (8). Here we report that the recently characterized Surf-5 gene is also differentially spliced to generate a 1.5-kb transcript in addition to the previously reported 3.5-kb transcript (6). Although the smaller 1.5-kb transcript is undetectable in fibroblast cells and also absent or very poorly expressed in the spleen, lung, liver, and kidney, it is a major, often predominant, species in the brain, heart, skeletal muscle, and testis. The 1.5-kb transcript, Surf-5b, encodes a new Surf-5 protein that is 200 amino acids in length and contains 63 amino acid residues not found in the Surf-5 protein. Based on the subcellular fractionation of Neuro 2A cells, the Surf-5b protein resides in the cytoplasmic S100 subcellular fraction (Fig. 6) as previously shown for the Surf-5 protein (6).

In addition to being regulated in a tissue-specific manner, the expression of the two alternative Surf-5 gene products was also regulated in differentiating cells. The P19 cell line has been used frequently as a model for developmentally regulated events in neurogenesis and myogenesis (16). We examined the expression of the two Surf-5 proteins in differentiating P19 cells and found that the expression of the Surf-5b mRNA and protein increases dramatically relative to the ubiquitous Surf-5 mRNA and protein following differentiation into neurons. P19 cells might provide a useful model for the study of the regulation of alternative splicing of the Surf-5 gene.

The Surf-5b splice acceptor sequence deviates from the consensus mammalian splice acceptor sequence in containing an A, instead of a C, at the −3 position and being very purine rich, instead of pyrimidine rich, in the preceding 10 bases pairs (Fig. 3B). Based on these features the Surf-5b splice acceptor site appears to be similar to splice acceptor sites used for alternative splicing in neurons (22) and thus may be responsible for the high amount of Surf-5b mRNA in the brain. Alternative spliced mRNAs found in the brain can also be found in heart and muscle tissue (22).

The Surf-5b splice is conserved in humans and human Surf-5b cDNAs were easily isolated from a library derived from human fetal brain mRNA. Although there are only two amino acids difference between the mouse and human in the 140 amino acids that constitute the ubiquitous Surf-5 protein (6) there are six differences between mouse and human in the 63 amino acids that are specific to the Surf-5b protein (Fig. 4). The Surf-5 gene is conserved in *C. elegans* (6). A *Drosophila melaganoster* genomic P1 sequence containing sequences homologous to Surf-5 (accession number L46826) as well as a *C. elegans* cosmid containing the Surf-5 homologue (accession number Z49073) have recently been placed on the DNA data base. The *Drosophila* sequence contains 1.5 kb and the *C. elegans* cosmid at least 20 kb downstream from the Surf-5 homologous sequences. In neither of these downstream DNA regions has any sequence been detected that could specify a peptide homologous to the additional 63 amino acids found in mammalian Surf-5b. This indicates that the Surf-5b protein is specific to only certain differentiated cells of higher vertebrates. An *Arabidopsis thaliana* homologue of Surf-5 has recently been deposited on the DNA data base (accession number H76340). This plant Surf-5 homologue se-

quence is derived from a cDNA and only shows homology to the 5′ end of the Surf-5 gene so, at this time, it is not known whether a Surf-5b protein also exists in *Arabidopsis thaliana*.

A second important posttranscriptional mechanism of gene regulation is alternative polyadenylation as found with Surf-4. Although the 2.8- and 2.0-kb Surf-4 transcripts encode the same protein (10), only the former mRNA overlaps at its 3′ end with the adjacent 3′ end of the Surf-2 gene (24). The ratio of these transcripts varied almost threefold between different mouse tissues (Fig. 2B). If transcriptional interference or antisense regulation occurs between the Surf-4 and Surf-2 genes, then the choice of 3′ polyadenylation signals used to generate the 2.8- and 2.0-kb mRNAs may modulate this effect to different degrees in different cell types.

The differential processing of the Surf-5 and Surf-4 mRNAs in different cell types demonstrates another form of regulation of these genes in a cell-specific manner. In the case of the Surf-4 gene, which encodes a ubiquitously expressed integral membrane protein associated with the ER (18), the coding capacity of the major 2.8- and 2.0-kb mRNAs with alternative polyadenylation sites is the same. Differences in mRNA stability or inability to be efficiently translated due to interaction with the overlapping Surf-2 mRNA (23) (see Fig. 1) may be overcome by the use of alternatively processed mRNAs with the same coding capacity. In the case of the Surf-5 gene the ubiquitous Surf-5 mRNA is expressed in all cell types. The alternatively spliced Surf-5b mRNA, which encodes a variant of the Surf-5 protein, appears to be required in specific cell types of higher vertebrates.

ACKNOWLEDGEMENTS

We would like to thank Drs. Clive Dickson, Charles E. Owen, and Malcolm Parker for their helpful discussions in the preparation of this manuscript. K.G. was a recipient of a Medical Research of Canada Postdoctoral Fellowship.

REFERENCES

1. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. 215:403–410; 1990.
2. Bairoch, A. PROSITE: A dictionary of sites and patterns in proteins. Nucleic Acids Res. 20:2013–2018; 1992.
3. Collins, J. F.; Coulson, A. F. W. Molecular sequence comparison and alignment. In: Bishop, M. J.; Rawlings, C. J., eds. Nucleic acid and protein sequence analysis: A practical approach. Oxford: IRL Press; 1987:323–358.
4. Colombo, P.; Yon, J.; Garson, K.; Fried, M. Conservation of the organization of five tightly clustered genes over 600 million years of divergent evolution. Proc. Natl. Acad. Sci. USA 89:6358–6562; 1992.
5. Devereux, J.; Haeberli, P.; Smithies, O. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395; 1984.
6. Garson, K.; Duhig, T.; Armes, N.; Colombo, P.; Fried, M. Surf5: A gene in the tightly clustered mouse surfeit locus is highly conserved and transcribed divergently from the rpL7A (Surf3) gene. Genomics 30:163–170; 1994.
7. Gaston, K.; Fried, M. YY1 is involved in the regulation of the bi-directional promoter of the Surf-1 and Surf-2 genes. FEBS Lett. 347:289–294; 1994.
8. Green, M. R. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. 7:559–599; 1991.
9. Henikoff, S.; Henikoff, J. G. Automated assembly of protein blocks for database searching. Nucleic Acids Res. 19:6565–6572; 1991.
10. Huxley, C.; Fried, M. The mouse surfeit locus contains a cluster of six genes associated with four CpG-rich islands in 32 kilobases of genomic DNA. Mol. Cell. Biol. 10:605–614; 1990.
11. Huxley, C.; Williams, T.; Fried, M. 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; 1988.
12. Lennard, A.; Gaston, K.; Fried, M. The Surf-1 and Surf-2 genes and their essential bidirectional promoter elements are conserved between mouse and human. DNA Cell Biol. 13:1117–1126; 1994.
13. Lennard, A. C.; Fried, M. The bidirectional promoter of the divergently transcribed mouse Surf-1 and Surf-2 genes. Mol. Cell. Biol. 11:1281–1294; 1991.
14. MacPherson, P. A.; McBurney, M. W. P19 embryonal carcinoma cells: A source of cultured neurons ammenable to genetic manipulation. In: Conn, P. M.; Murphy, S., eds. Neuroprotocols, a companion to methods in neuroscience. San Diego: Academic Press; 1995
15. Magoulas, C.; Fried, M. The Surf-6 gene of the mouse surfeit locus encodes a novel nucleolar protein. DNA Cell Biol. 15:305–316; 1996.
16. McBurney, M. W. P19 embryonal carcinoma cells. Int. J. Dev. Biol. 37:135–340; 1993.
17. Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85:2444–2448; 1988.

18. Reeves, J. E.; Fried, M. The surf-4 gene encodes a novel 30 kDa integral membrane protein. Mol. Membr. Biol. 12:201–208; 1995.
19. Rudnicki, M. A.; McBurney, M. W. Cell culture methods and induction of differentiation of embryonal carcino cell lines. In: Robinson, E. J., ed. Tetratocarcinomas and embryonic stem cells, a practical approach. Oxford: IRL Press; 1987.
20. Sambrook, J.; Fritsch, E. F.; Maniatiss, T. Molecular cloning—a laboratory manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
21. Smith, D. B.; Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67:31–40; 1988.
22. Stamm, S.; Zhang, M. Q.; Marr, T. G.; Helfman, D. M. A sequence compilation and comparison of exons that are alternatively spliced in neurons. Nucleic Acids Res. 22:1515–1526; 1994.
23. Williams, T.; Fried, M. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature 322:275–279; 1986.
24. Williams, T.; Yon, J.; Huxley, C.; Fried, M. 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; 1988.
25. Williams, T. J.; Fried, M. The MES-1 murine enhancer element is closely associated with the heterogeneous 5′ ends of two divergent transcription units. Mol. Cell. Biol. 6:4558–4569; 1986.
26. Yon, J.; Jones, T.; Garson, K.; Sheer, D.; Fried, M. The organization and conservation of the human surfeit gene cluster and its localization telomeric to the c-abl and can proto-oncogenes at chromosome band 9q34.1. Hum. Mol. Genet. 2:237–240; 1993.