ELSEVIER

Gene 284 (2002) 169–178

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

The human homologue of the mouse *Surf5* gene encodes multiple alternatively spliced transcripts

Antonietta Angiolillo[a,1], Giulia Russo[a,1], Antonio Porcellini[b,2], Silvia Smaldone[a], Felicia D'Alessandro[a], Concetta Pietropaolo[a,*]

[a]*Dipartimento di Biochimica e Biotecnologie Mediche, Università 'Federico II' and CEINGE Biotecnologie Avanzate, Via Sergio Pansini 5, I-80131 Naples, Italy*

[b]*Dipartimento di Medicina Sperimentale e Patologia, Università 'La Sapienza', I-00185 Rome, Italy*

Received 18 September 2001; received in revised form 3 December 2001; accepted 21 December 2001
Received by R. Di Lauro

Abstract

Hu-Surf5 is included within the *Surfeit locus*, a cluster of six genes originally identified in mouse. In the present study, we have cloned and characterized the *Hu-Surf5* gene and its mRNA multiple transcripts. Comparison of the most abundant cDNA and genomic sequence shows that the *Hu-Surf5* is spread over a region of approximately 7.5 kb and consists of five exons separated by four introns. The nucleotide sequence of the genomic region flanking the 3′-end of the *Hu-Surf5* gene revealed the presence of a processed pseudogene of human ribosomal protein L21 followed by *Hu-Surf6* gene. Only 110 bp separate the transcription start site of *Hu-Surf5* and *Hu-Surf3/L7a* gene and the transcription direction is divergent. Earlier studies defined the 110 bp region essential for promoter activity of *Hu-Surf3/L7a*. Here, we show that this region stimulates transcription with a slightly different efficiency in both directions. The bidirectional promoter lacks an identifiable TATA box and is characterized by a CpG island that extends through the first exon into the first intron of both genes. These features are characteristic of housekeeping genes and are consistent with the wide tissue distribution observed for *Hu-Surf5* expression. *Hu-Surf5* encodes three different transcripts, Surf-5a, Surf-5b, and Surf-5c, which result from alternative splicing. Two protein products, SURF-5A and SURF-5B have been characterized. Production of chimaeras between the full-length SURF-5A or SURF-5B and the green fluorescent protein (GFP) allowed to localize both proteins in the cytoplasm.

Keywords: Human Surfeit locus; Bidirectional promoter; Post-transcriptional control; Subcellular localization

1. Introduction

The human *Surfeit* locus has been mapped on chromosome 9q34.1 (Yon et al., 1993). The locus includes six tightly clustered housekeeping genes (*Surf1-6*), and the gene organization is similar in human, mouse and chicken *Surfeit* locus (Colombo et al., 1992; Duhig et al., 1998). The *Surfeit* genes are separated by a rather low number of base pairs, the 5′-end of each gene is associated with a CpG rich island, and, at least in mouse and humans, the direction of transcription of each gene has been demonstrated to be opposite with respect to that of its neighbour (Huxley and Fried, 1990; Garson et al., 1995). There are numerous examples of clustered genes in the mammalian genome, which have been generated by duplication of an ancestral gene. However, there are very few clusters of mammalian genes with such a close spacing as in the *Surfeit* locus, which are not related either by sequence or by functions. The *Surfeit* locus is so far unique in mammalian genome because of the lack of any significant nucleotide or amino acid homology among the genes (Huxley and Fried, 1990); furthermore, *Surfeit* genes seem to be functionally unrelated. The *Surfeit* locus in mouse has been extensively characterized. *Surf3* gene encodes the ribosomal protein L7a (Giallongo et al., 1989); *Surf4* gene encodes a cytoplasmic integral membrane protein most likely associated with the endoplasmic reticulum (Reeves and Fried, 1995), *Surf6* product appears to be a nucleolar protein (Magoulas and Fried, 1996) and *Surf1* has been shown homologous to the yeast SHY1 (Mashkevich et

Abbreviations: aa, amino acid(s); bp, base pair(s); CAT, Cm acetyltransferase; cDNA, complementary DNA; CMV, cytomegalovirus; Δ, deletion; DMEM, Dulbecco's modified Eagle's medium; GFP, green fluorescent protein; kb, kilobase(s); kDa, kilodalton(s); nt, nucleotide(s); PCR, polymerase chain reaction, mRNA, messenger ribonucleic acid; M_r, relative molecular mass; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3 citrate, pH 7.6; UTR, untranslated region(s)

* Corresponding author. Tel.: +39-81-746-3065; fax: +39-81-746-3074.
E-mail address: pietropaolo@dbbm.unina.it (C. Pietropaolo).

[1] These authors contributed equally to the work.

[2] Present address: INM, Neuromed, Pozzilli, I-86077 Isernia, Italy.

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00379-7

al., 1997) encoding a factor involved in the biogenesis of cytochrome *c* oxidase. More recently, mutations in the human homologue of *Surf1* have been associated with Leigh disease (Tiranti et al., 1998; Zhu et al., 1998), and the human *Surf6* gene has been isolated and characterized (Magoulas and Fried, 2000). The expression of the *Surf5* gene in mouse has been shown post-transcriptionally regulated and producing two RNA species of 3.5 and 1.5 kb, respectively (Garson et al., 1996). The protein products show a high degree of identity in mouse and humans.

We have previously isolated, from a human genomic library, a recombinant lambda EMBL3 phage which contains 17 kb of human DNA spanning *Hu-Surf3* gene (De Falco et al., 1993), *Hu-Surf5* and the 5′-moiety of *Hu-Surf6* gene. Here, we describe: (i) the genomic organization of the *Hu-Surf5* gene, and its structural relationship with the contiguous *Surf6* and *Surf3* genes, (ii) the structural features of the bidirectional promoter shared by *Hu-Surf5* and *Hu-Surf3*, (iii) the identification of three mRNA transcripts and cloning of the respective cDNAs, and (iv) the expression of two SURF5 protein products and their subcellular localization.

2. Materials and methods

2.1. Isolation of the Hu-Surf5 gene, and cloning of cDNA transcripts

A genomic DNA clone, λ37, isolated by screening a human genomic library with a cDNA fragment coding for a human L7a (De Falco et al., 1993), contained the *Hu-Surf3/L7a* gene and the entire *Hu-Surf5* gene. Restriction fragments from the *Hu-Surf5* gene were subcloned into pGEM-4Z and the entire sequence was determined on both strands. A human cDNA library from a teratocarcinoma cell line in the λgt10 vector (D'Esposito et al., 1994) was screened to isolate clones encoding *Surf-5* cDNA. Approximately 5×10^5 recombinant plaques were screened by transferring to nitrocellulose filters and hybridizing with probe c (Fig. 1A) in stringent conditions. Positive plaques were identified by autoradiography and purified by re-screening. The largest specific clone (2568 bp in length), namely *Surf-5a*, was fully sequenced by exploiting T7 and SP6 flanking primers as well as walking primers to sequence both strands. Upon complete analysis *Surf-5a* was recognized as a partial clone; the missing 5′-end was obtained by PCR amplification using a cDNA library from human 8-week embryo (D'Esposito et al., 1994) as a template. As primers, a *Surf5*-specific oligonucleotide (Surf -5a/b in Table 1), and the forward and reverse primers specific for λgt10 were applied. The PCR conditions were as suggested for AmpliTaq DNA polymerase (Boehringer). The amplification products were electrophoresed, transferred onto nylon membrane (Amersham) and hybridized with a γ-^{32}P-labelled internal oligonucleotide. Hybridization-positive PCR products were cloned into the *Eco*RI site of a T-tailed pGEM vector. Clones containing inserts were sequenced using primers for the flanking SP6 and T7 promoters.

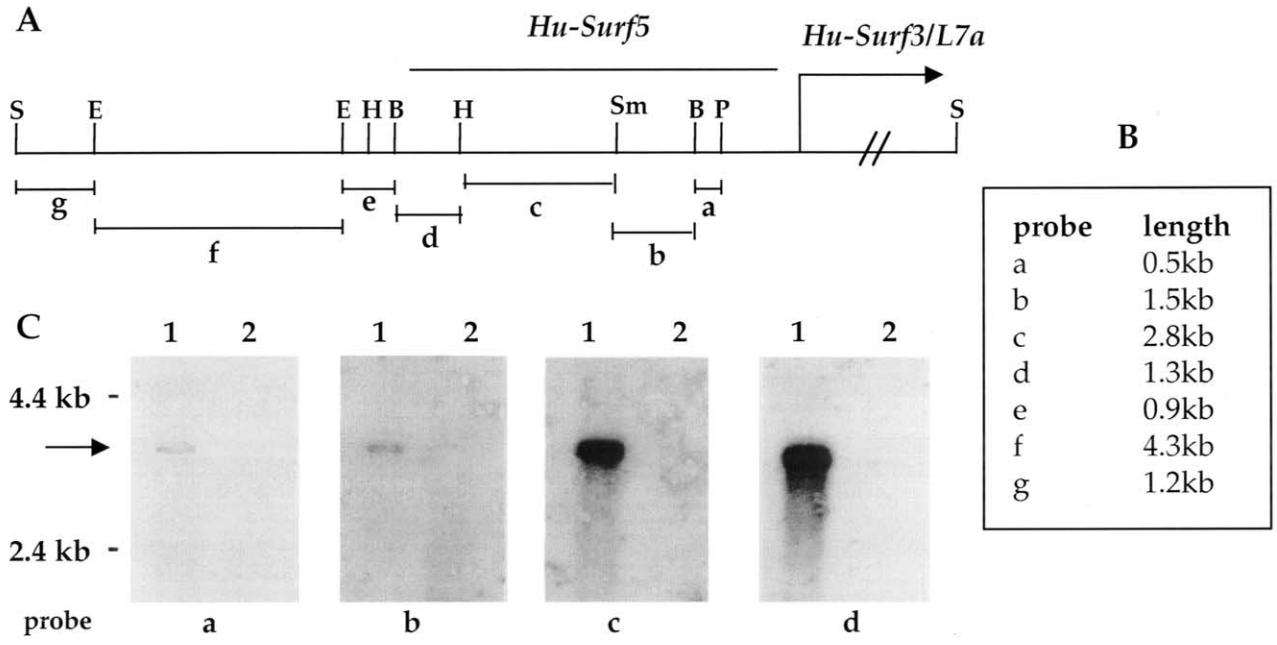


Fig. 1. Restriction and transcript analysis of the human *Surf5* genomic region. (A) Restriction map of the region (14,053 bp) of the human *Surfeit locus* upstream from the human *Surf3-L7a* gene (*Hu-Surf3/L7a*) (E, *Eco*RI; B, *Bam*HI; Sm, *Sma*I; H, *Hin*dIII; S, *Sal*I; P, *Pst*I). The direction of transcription of *Hu-Surf3/L7a* is indicated by the arrow. The restriction fragments used as probe for Northern blot analysis are indicated and marked with an alphabetical letter. (B) Nucleotide length of the indicated probes. (C) A1251 cells RNA analysis; 2 μg of poly(A)$^+$ RNA (lanes 1) and 20 μg of poly(A)$^-$ RNA (lanes 2) were fractionated on a 1.5% agarose gel, blotted into a nylon membrane and hybridized with the indicated probes.

Table 1
Sequence of synthetic oligonucleotide primers used for PCR amplification of Surf-5 cDNA species

Surf-5a/b	5′-AGCCGCTTGTTGTAGGACTGCAGCAGCGTC-3′
Surf-5c, 1	5′-GACTCGAGTCGACATTTCCAGGCCAGCCTGTGTC-3′
Surf-5c, 2	5′-GACTCGAGTCGACATGCTAGCTGAGCTCTTCACCT-3′
Surf-5c, 3	5′-GACTCGAGTCGACATCCTTTCAGAATGCTAGGGTGGC-3′
λgt10, forward	5′-CTTTTGAGCAAGTTCAGCCTGGTTAAG-3′
λgtl0, reverse	5′-GAGGTGGCTTATGAGTATTTCTTCCAGGGTA-3′

Surf-5b and *Surf-5c* cDNAs were isolated by the same PCR approach using DNA of a cDNA library from human teratocarcinoma cell line (D'Esposito et al., 1994) as a template. As primers, a panel of the *Hu-Surf5*-specific oligonucleotides and the forward and reverse primers specific for λgt10 were used (Table 1). PCR amplification was carried out as described above. The resulting clones were subcloned into a T-tailed pGEM vector. SP6 and T7 primers were used to sequence the 5′- and 3′-ends of the clones. Internal sequences of the amplified DNA clones were obtained by using oligonucleotides designed on *Surf-5a* sequence information.

2.2. DNA sequence analysis

Sequencing was performed on double-stranded DNA by dideoxy chain-termination procedure (Sanger et al., 1977) using [α-^{35}S]dATP (1000 Ci/mmol, Amersham) and Sequenase 2.0 kit (U.S. Biochemical) following the instructions provided by the supplier. Some sequences were performed using the dideoxy chain termination method with fluorescent dye terminators in an ABI 373A automated DNA sequencer (Applied Biosystems Division, Perkin Elmer, Foster City, CA).

2.3. Nucleic acid analysis

Genomic DNA extraction, and all common techniques of routine DNA manipulations, including restriction analysis, gel fractionation and blotting, were carried out according to standard procedures (Sambrook et al., 1989).

Total cellular RNA was isolated by the guanidinium-thiocyanate single-step extraction method (Chomczynski and Sacchi, 1987). Poly(A)$^+$ RNA was purified by one passage on oligo (dT)-cellulose columns as described (Sambrook et al., 1989). For Northern blots 2 μg of poly(A)$^+$ and 20 μg of total RNA were fractionated on a 1.5% agarose-formaldehyde denaturing gels. RNA transfer was achieved by blotting in 20 × SSC onto nylon membrane (Amersham) according to Sambrook et al. (1989). Hybridization and washing conditions were as described (Church and Gilbert, 1984). Lastly, autoradiography was performed at −80 °C with an intensifying screen.

Blots containing, per lane, 2 μg of poly(A)$^+$ RNA from eight different human tissues were obtained commercially (Multiple Tissue Northern, Clontech), and used in hybridization experiments following the manufacturer's instructions. For use in successive probing, membranes were stripped by incubating (2 × 15 min) at 98 °C in 0.5% SDS. The loss of bound probe was verified by exposure to a Kodak X-AR autoradiographic film.

2.4. Cell culture and transient transfections

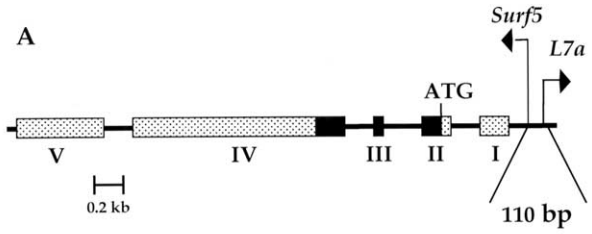
All cell lines, A1251 (Rosen et al., 1980), HepG2 (ATCC, HB-8065), HeLa (ATCC, CCL-2.2), JEG-3 (ATCC, HTB-36), N-Tera-2 (D'Esposito et al., 1994), were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco), supplemented with 10% foetal calf serum (FCS), 5 mM L-glutamine and penicillin–streptomycin 50 U/ml and 50 μg/ml, respectively, at 37 °C in a 5% CO_2 humidified incubator. For CAT assays, 5 h prior to transfection, cells were subcultured to yield 50% confluent cultures, and then transfected overnight by calcium phosphate method (Wigler et al., 1977) using 10 μg of DNA purified through a CsCl gradient centrifugation. Cells were harvested 24 h after calcium phosphate removal and assayed for CAT activity. All constructs were expressed in the pEMBL8-CAT eukaryotic transcription vector harbouring a promoterless prokaryotic *cat* gene (De Falco et al., 1993). For immunofluorescence studies, cells were seeded on coverslips in 60-mm dishes; 18 h after plating DNA transfection was carried out using 2 μg of the appropriate recombinant pEGFP-N1 (Clontech) vector and the calcium phosphate precipitation method. Fifteen hours after transfection, the cells were fixed with 3.7% formaldehyde in PBS, kept for 20 min at room temperature and then the coverslips were mounted with Moviol (Hoechst, Germany) on microscope slides. Cells were analyzed on a Zeiss laser scan microscope (LSM410) using fluorescein filters 450-490, FT510-LP520.

2.5. CAT assays

Cell extracts were prepared and CAT assays performed as described (Gorman et al., 1982). Reactions were visualized by autoradiography. All transfections were normalized for protein concentration and transfection efficiency was monitored by including 1 μg of pCMV-luciferase per transfection and assaying for luciferase activity (Sambrook et al., 1989). Each transfection was performed at least three times using different plasmid DNA preparations.

2.6. EGFP expression constructs

Constructs directing the expression of green fluorescent



B

Exon No	Exon Size (bp)	Splice donor and acceptor sites 5′ donor	3′ acceptor	Intron No	Intron Size (bp)
I	177	GGCGG/gtatg	tctag/GTCCC	1	1217
II	161	CCAAG/gtggg	cccag/ATTGA	2	1282
III	81	ACATC/gtgag	cacag/GTCCG	3	837
IV	1631	CCCCT/ggggg	cccct/ACAGG	4	1136
V	657				

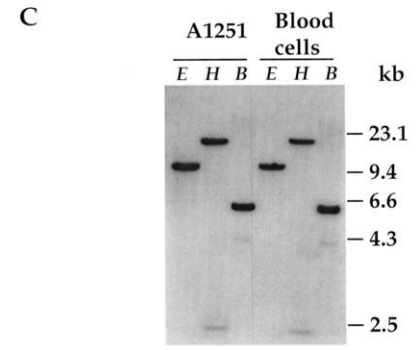


Fig. 2. Structural organization of the *Hu-Surf5* gene. (A) Boxes represent exon sequences, the open reading frame (dark box) and 5′- and 3′-untranslated regions (dotted box). Solid lines connecting exon regions designate the intron sequences. The 110 bp separating 5′-ends of *Hu-Surf3/l7a* and *Hu-Surf5* and the position of the translation initiation site for Surf-5a and Surf-5b mRNAs are indicated. The direction of transcription of *Hu-Surf3/l7a* and *Hu-Surf5* is indicated by the arrows. (B) Intron–exon organization of the *Hu-Surf5* gene. (C) Southern blot analysis on genomic DNA from A1251 cells and blood cells digested with the restriction enzymes, as indicated (E, *Eco*RI; H, *Hin*dIII; B, *Bam*HI), fractionated through a 0.8% agarose gel, blotted into a nylon membrane and hybridized with a Surf-5a cDNA probe.

(EGFP) fusion proteins containing the coding region of wild-type Surf-5a, Surf-5b and Surf-5b mutant (Surf-5bΔ) were generated by PCR. Briefly, the vectors pGEM4Z-Surf-5a and pGEM4Z-Surf-5b, containing wild-type Surf-5a and Surf -5b were used as a template with primer pairs 5′-CGGAATTCCCCGGCGGACGCGGCGTC-3′ plus 5′-CGGGATCCTTATACCTGGACGAGTAATAC-3′ and 5′-CGGAATTCCCCGGCGGACGCGGCGTC-3′ plus 5′-CGGGATCCGCGTGCTCAGTGGGGCCAGGG-3′, containing an overhanging *Eco*RI or a *Bam*HI site (underlined) to generate 400 and 600 bp products, respectively. The open reading frame between the 5′-terminus of the EGFP coding region and the Surf-5a and Surf-5b inserts was restored by the deletion of a single nucleotide flanking the 3′-end of the *Bam*HI site. The PCR products were cut with *Eco*RI and *Bam*HI and sub-cloned into *Eco*RI/*Bam*HI of pEGFP-N1 vector. To construct the Surf-5bΔ mutant harbouring a 179 bp deletion spanning from residue 268 to 446 of the coding sequence, a *Pst*I/*Pst*I fragment, 272 bp long, was removed from pEGF-N1 Surf-5b and replaced with a 93 bp *Pst*I/*Pst*I fragment obtained by PCR. To obtain the 93 bp *Pst*I/*Pst*I fragment, pGEM4Z-Surf-5b DNA was used as template, with the primers, 5′-AACTGCAGAAGCTGGTGTCCGACCTC-3′ and 5′-AACTGCAGTGTGCGCAGCTGCTGGTT-3′. All PCR products as well DNA junctions generated by sub-cloning were verified by sequence analysis.

3. Results and discussion

3.1. Structure of the Hu-Surf5 gene and cloning of the Surf-5a cDNA

In a previous study (De Falco et al., 1993), a human genomic DNA clone (λ37, 17 kb *Sal*I fragment) containing the *Hu-Surf3/l7a* gene was found extending for at least 14 kb at the 5′-end into the human *Surfeit locus*, toward the *Hu-Surf5* gene. Restriction enzyme analysis of this clone yielded the map shown in Fig. 1A. To characterize the *Hu-Surf5* gene, we analyzed poly(A)$^+$ RNA from A1251 cells by Northern blot hybridization, using the probes shown in Fig. 1A. We found that the probes a, b, c and d identified a single mRNA species of about 3.0 kb as shown in Fig. 1C. In contrast, the fragments e, f, and g revealed no transcript (data not shown), indicating that the probe d included the end of the transcript, and thus the end of the *Hu-Surf5* gene. The genomic organization of the gene was determined by subcloning and sequencing the restriction fragments of clone λ37 (Fig. 1A) and by comparing with the Surf-5a cDNA sequence (described below). The *Hu-Surf5* gene (Fig. 2A,B) contains five exons, and four introns. The splice sites conform to the AG/GT splice rule, with the exception of intron 4 where a GG/CT sequence replaces the consensus. Exon I includes part of the 5′-untranslated region, the ATG start codon is in exon II, and the entire coding region is contained within three exons (II–IV), flanked by 2068 bp of 3′-untranslated sequence in exons IV and V. Analysis of the first intron revealed the presence of an Alu repetitive DNA sequence.[3] The presence of the additional intron (number 1 in Fig. 2) in the 5′-UTR of the mRNA transcripts of Hu-Surf5 has been reported already (Duhig et al., 1998) as a major difference in the organization of the Surf5 gene between mouse and humans. Upon the

[3] The sequence data of the genome region including *Surf5* gene, the intergenic region and part of *Surf6* gene has the EMBL/GenBank accession number AJ224639. The Hu-Surf-5a, Hu-Surf-5b, and Hu-Surf-5c cDNAs were deposited with the EMBL/Genbank Data under accession numbers AJ224358, AJ224359 and AJ224360, respectively. The SURF-5A, SURF-5B protein products are in the SWISS-PROT with accession number Q15528.

complete sequence analysis of Hu-Surf5 gene, we report here the presence of an intron also in the 3′-UTR of larger Hu-Surf5 mRNA transcript. This gene feature has not been previously reported either in mouse or in humans.

To determine the copy number of *Hu-Surf5* in the human genome, a Southern blot analysis was performed on human genomic DNA digested with three different restriction enzymes. As shown in Fig. 2C, hybridization with a *Surf-5a* cDNA probe detected DNA fragments whose size was consistent with the size expected from the restriction analysis of the genomic clone (see Fig. 1A). These results strongly suggested that the *Hu-Surf5* gene is present as a single copy per haploid genome.

The analysis by comparison with sequences listed in the GenBank of the 7 kb region extending downstream the *Surf5* gene allowed us to assign a sequence of 495 bp with a small poly(dA) tail to a pseudogene of human L21 ribosomal protein. It has the characteristic of a processed pseudogene, including a full-length L21 cDNA with a poly(dA) tail, flanked by two imperfect direct repeats. Some base changes and small deletions and insertions are also present in the coding sequence of the L21 pseudogene, which is followed by the human *Surf6* gene (Magoulas and Fried, 2000).

Probe c (Fig. 1A) was used to screen a cDNA library representing poly(A)$^+$ RNA from a human teratocarcinoma cell line (N-Tera-2). We screened a total of 500,000 plaques at high stringency and the positive clones were plaque-purified and analyzed. The longest clone (2568 bp) appeared to be lacking an ATG translation initiation codon, thus did not cover the full-length mRNA. By using sequence information from the extreme 5′-end of this cDNA, we designed a *Hu-Surf5* specific oligonucleotide (see Section 2.1) that permitted the amplification of *Surf5* specific DNA fragments from a human 8-week embryo cDNA library. These fragments were sequenced and the information obtained was used to assemble the complete sequence of the cDNA clone described as *Surf-5a*, 2711 bp long.

3.2. Bidirectional promoter

Comparison of the DNA sequence of the isolated Surf-5a cDNA with that of genomic DNA showed that, as already observed in mouse (Garson et al., 1995), the *Hu-Surf5* gene is positioned in a head to head configuration with *Hu-Surf3* gene. To confirm the gene functional organization a sense RNA and an antisense RNA, transcribed using Surf-5a cDNA as a template, were used as a probe in a Northern blot experiment with total RNA from three human cell lines (Fig. 3A). A specific hybridization was obtained only with the antisense probe, thus demonstrating that the transcription direction of *Hu-Surf5* is divergent from the transcription direction of *Hu-Surf3*. *Hu-Surf5* and *Hu-Surf3* join a growing list of divergently transcribed gene pairs. In a few cases such gene pairs may have a related function and their promoters may contain common elements that facilitate coordinated expression. Examples include histone H2A and H2B genes (Hentschel and Birnstiel, 1981), the avian *GPAT* and *AIRC* genes (Gavalas and Zalkin, 1995), and the mouse and human α1 (IV) and α2 (IV) collagen genes (Burbelo et al., 1988; Schmidt et al., 1993). In the majority of cases, the gene pairs share no obvious common function, e.g. *DHFR*/*Rep*-1 (Linton et al., 1989), *ATM*/*E14* (*NPAT*) (Byrd et al., 1996; Imai et al., 1996), and the Wilms tumor *WT*-1/*Wit*-1 locus (Huang et al., 1990; Gessler and Bruns, 1993), and their expression may or may not be coordinately controlled. In some cases the core of bidirectional promoters are flanked in both directions by TATAA boxes (Hentschel

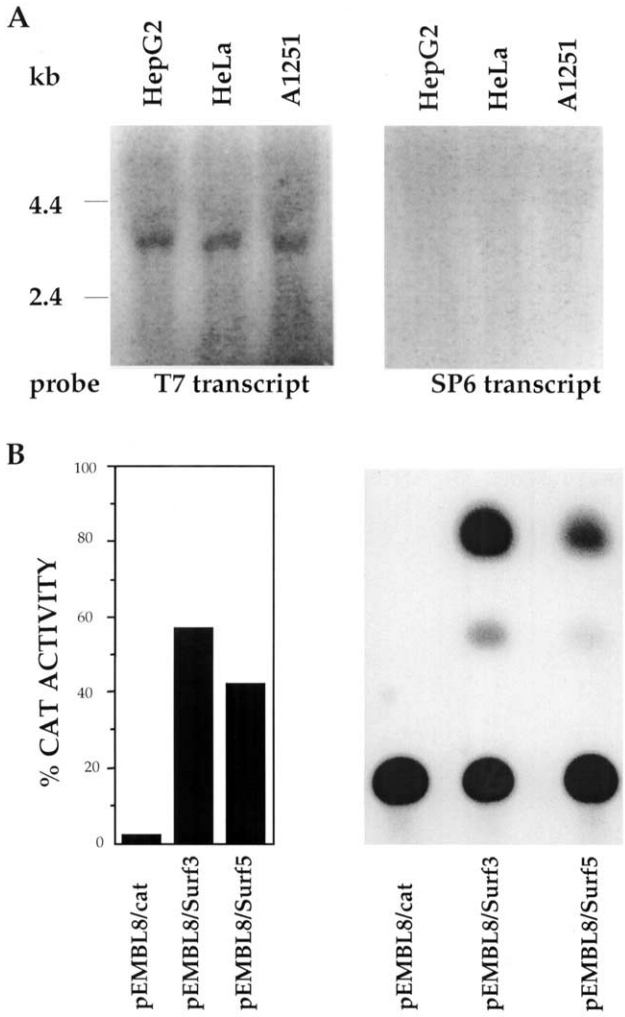


Fig. 3. Analysis of a bidirectional promoter. (A) Twenty micrograms of total RNA from different human cell lines were fractionated on a 1.5% agarose gel, blotted into a nylon membrane and hybridized to a sense (Sp6) or antisense (T7) riboprobe produced by transcription of Surf-5a cDNA. (B) Representative data from CAT assays on HeLa cell extracts, following transient transfection with 10 μg of pEMBL8, pEMBL8/Surf5 or pEMBL8/Surf3 plasmids. In all the experiments 1 μg of pCMV-luciferase was present in the precipitate and used as internal control for transfection efficiency. The panel on the right shows the autoradiographic results of a typical CAT assay analysis; the histogram is a representation of the resultant average CAT activity from three different transfections, all of which have been performed in duplicate using two different plasmid DNA preparations.

and Birnstiel, 1981). In other cases, the promoters appear to lack TATAA boxes in either orientation and may be bidirectional due to the absence of elements that specify a unique transcriptional orientation (Burbelo et al., 1988; Gavalas and Zalkin, 1995). In the present study we have demonstrated that the 110 bp segment separating *Hu-Surf3* and *Hu-Surf5* (see Fig. 2A) does not contain any TATAA or CCAAT box sequences. We have reported already that this region contains sufficient information to direct the transcription of *Hu-Surf3* gene (De Falco et al., 1993) and, thus, we tested whether it was able as well to drive the transcription of *Hu-Surf5* gene. To this aim, the intergenic DNA fragment was amplified by PCR using specific oligonucleotide primers and cloned in a pEMBL8-CAT vector harbouring a promoterless prokaryotic *cat*. The correct orientation was determined by DNA sequencing. The resulting vector (pEMBL8/Surf5) was transfected into HeLa cells for transient expression experiments along with the empty vector, as a negative control, or the same vector containing the 110 bp fragment in the opposite orientation (pEMBL8/Surf3). This last construct, in fact, has been demonstrated able to direct the transcription of *Hu-Surf3/L7a* (De Falco et al., 1993). The plasmid pEMBL8/Surf5 showed a sustained level of CAT activity, although lower than pEMBL8/Surf3 (Fig. 3B). The results of these experiments demonstrate that this region functions as a bidirectional promoter, allowing a different level of expression for *Hu-Surf5* and *Hu-Surf3* genes, as supported also by the different amounts of specific RNA detected by Northern blot analysis (data not shown). A similar mechanism of transcriptional control has been described in human and also in mouse for some *Surfeit* genes, *Surf1* and *Surf2* (Lennard and Fried, 1991), whereas, at least in mouse, other *Surfeit* genes exploit different mechanisms including transcriptional interference and antisense regulation (Williams and Fried, 1986).

3.3. Expression of Hu-Surf5 gene in human tissues

To investigate *Hu-Surf5* gene expression, Northern blot analysis was performed on total RNA from different human cell lines (Fig. 4A) and on poly(A)$^+$ RNA from human tissues (Fig. 4B), using the coding region of Surf-5a cDNA as a probe. The analysis of N-Tera-2 cell line revealed, in addition to the 2.7 kb mRNA species already identified in A1251 cells, a 1.4 kb species. The 2.7 kb species was expressed in all cell lines tested, the 1.4 kb species, on the other hand, was present only in N-Tera-2 cells. Poly(A)$^+$ RNA analysis (Fig. 4B) confirmed the

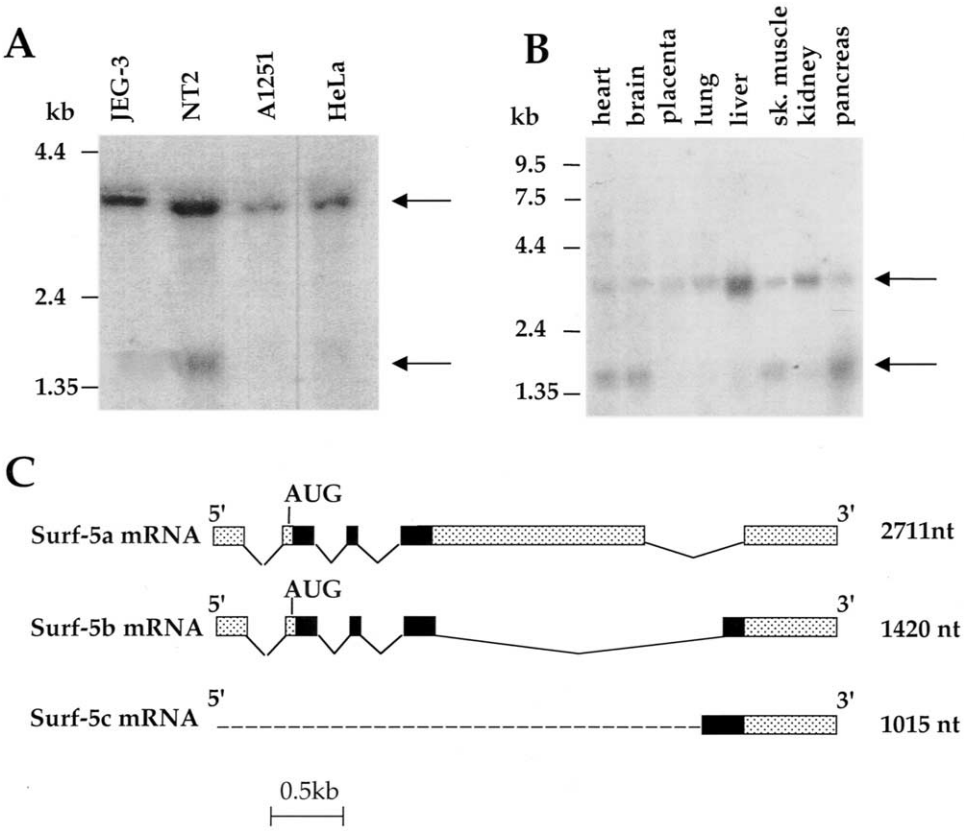
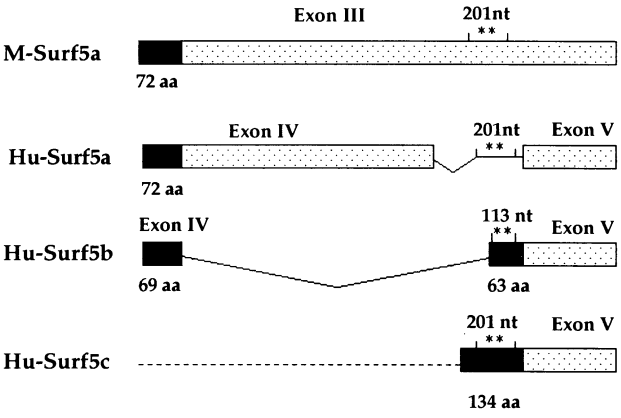


Fig. 4. Expression of human *Surf5* gene. Total RNA, 20 μg, from the indicated human cell lines (A) and 2 μg of poly(A)$^+$ RNA from the indicated tissues (B) were fractionated on 1.5% agarose gel, blotted into a nylon membrane and hybridized with a probe spanning the coding region of Surf-5a cDNA. The Surf-5a and Surf-5b transcripts are indicated by arrows. (C) Schematic representation of Surf-5 mRNA generated by alternative splicing in the 3′ region of the *Hu-Surf5* gene. Boxes represent exon sequences, coding exons (dark box) and non-coding sequences (dotted box) are indicated. Sloping lines designate intron sequences. The position of the translation start site is indicated.

ubiquitous expression of the 2.7 kb species, and a restricted expression of the 1.4 kb species. In fact, the 1.4 kb RNA transcript was clearly detected in heart, brain, skeletal muscle and pancreas. The 1.4 kb species might derive from an alternative splicing of the mRNA, from utilization of an alternative promoter, or from an alternative polyadenylation site. To investigate further the relationships between the two transcripts, a panel of primer pairs, spanning the sequence of several segments along the full-length *Surf-5a* cDNA and a segment of the fourth intron sequence, were used to screen a cDNA library from N-Tera-2 cell line by PCR technique. This analysis allowed the identification of three different *Surf-5* cDNAs (see Fig. 4C). The sequence analysis revealed also that the longest cDNA clone was identical to Surf-5a cDNA. Thus, the *Surf-5a* mRNA (Fig. 4C) contained 217 bp of 5′-untranslated sequence followed by an AUG translation initiation codon, which specified the start of a 420 nucleotide open reading frame 140 aa long (deduced $M_r = 16.680$ kDa, p$I = 4.75$), and then by, 2068 nucleotides of 3′-untranslated sequence containing one unusual polyadenylation signal (AUUAAA). The AUG in Fig. 4C has been considered as the start codon since: (i) the presence of an in-frame stop codon closely upstream, (ii) a good homology of the translational start site sequence context with the consensus sequence for eukaryotic initiation sites (Kozak, 1987), and (iii) the coding region in the mRNA shares a high degree of homology (90% identity) with the mouse homologue Surf5 mRNA (Garson et al., 1996).

Fig. 4C shows schematically the structural relationship

A



B

```
M-Surf5a: 2647  tttcttaactctgcatctgtcttgctttcccct-ccctggctttc--ctctgcctgctcc 27
                |||||||||||||| |||||||||| |||||| |||||||||||  ||||||||||||
Hu-Surf5c: 123  tttcttaactctgcttctgtcttgctctccctcccctggctttcctctctgcctgctcc 18

M-Surf5a: 2704  tgtcactcctgggggtttcttgtggtggcaaagctcaagtctctgcgaagctaatgacct 27
                ||||  ||||  ||||||||| ||||||| |||||||||| || ||||||||||||||||
Hu-Surf5c: 183  tgtctctccctggggtttctggtggtggaaaagctcaagcctttgcgaagctaatgacct 24
                                      ||||||||||||||||||||||||||||||
Hu-Surf5b: 630                        agctcaagcctttgcgaagctaatgacct 65

M-Surf5a: 2764  gcctctgtgtgaagcttactggaggctggacctcgacgcagattctgctgatggcctctc 28
                ||||||||| ||||||||| |||||||||||||||| |||| ||||||||||||||||
Hu-Surf5c: 243  gcctctgtgcgaagcttacgggaggctggacctcgacacagactctgctgatggcctctc 30
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Hu-Surf5b: 659  gcctctgtgcgaagcttacgggaggctggacctcgacacagactctgctgatggcctctc 71

M-Surf5a: 2824  agcccctcttctggcgtccccggag 2848
                 |||||||| ||||||||||||||
Hu-Surf5c: 303  ggcccctctgctggcgtccccggag 327
                |||||||||||||||||||||||||
Hu-Surf5b: 719  ggcccctctgctggcgtccccggag 743
```

Fig. 5. Relationship between mouse and human Surf5 cDNAs. (A) Diagram illustrating the organization of the 3′-region of Surf5 transcripts; boxes represent exon sequences, the coding region (dark box) and 3′-UTR (dotted box). Solid lines designate the intron sequences; ** indicate the location of shared sequence. (B) Homology between mouse *Surf-5a* (M-Surf5a), human *Surf-5b* (Hu-Surf5b), and human *Surf-5c* (Hu-Surf5c) cDNA sequences.

between the different mRNA transcripts of the *Hu-Surf5* gene. The second cDNA clone, *Surf-5b*, was 1.4 kb long consistent with the length of the shorter transcript detected in different tissues and in the teratocarcinoma cell line. Surf-5a and Surf-5b mRNA shared 100% identity in the 5′-UTR, and throughout a region coding for 137 aa. Then they diverge, since a differential splicing occurs in the 3′-UTR of Surf-5a transcripts, and in the coding region of the Surf-5b transcript. The result of the differential splicing is, in Surf-5b mRNA, a longer reading frame coding for a protein of 200 aa (21 kDa) whose first 137 aa are identical as in the SURF-5A protein. Lastly, a third cDNA (*Surf-5c*) was isolated and sequenced; it was 1.0 kb long, the 656 bp coding for the 3′-UTR were shared with *Surf-5a* and *Surf-5b*, while nucleotides 1–214 were unique for *Surf-5c* DNA and derived from the transcription of intron 4 sequence. This cDNA clone was not complete; in fact we could not find any ATG start codon; however, the isolation of this cDNA strongly suggested the existence of an additional alternatively spliced transcript. Furthermore, we found that several cDNA sequences in the human ESTs database extend through the intron four–exon five junction (e.g. accession numbers BF310449, BG683741, BG170936, etc.), some extending further than the *Surf-5c* clone toward the 5′-end, although still lacking an AUG initiation codon.

The 3′-region of the Surf5 gene appeared to be mostly affected by the alternative splicing, leading to multiple gene transcripts. Fig. 5 shows schematically the relationship between the 3′-region of mouse and human cDNAs. A 201 nt segment of the mouse larger transcript (M-Surf5a) is not present in the corresponding human Surf-5a cDNA (Hu-Surf5a), but is present in the human Surf-5c (Hu-Surf5c) and, partially, in the human Surf-5b (Hu-Surf5b) cDNAs, where becomes coding sequence, due to alternative splicing. Fig. 5B shows the high degree of identity in the shared sequences. The functional significance of these multiple transcripts is unknown although, as stated above, a tissue-specific splicing pattern has been observed. The expression pattern of human Surf-5a and Surf-5b transcripts is consistent with the expression pattern of the mouse *Surf-5* gene (Garson et al., 1996). However, there are no data on the occurrence of additional transcripts in mouse. The 2.7 kb transcript, Surf-5a, encoded a protein of 140 aa (SURF-5A, 16 kDa) that is identical to the protein encoded by the 1.4 kb transcript (Surf-5b) except for the last three amino acids that are substituted in SURF-5B by a 63 aa long peptide leading to a protein of 200 aa (21 kDa). As already reported by others (Garson et al., 1996), only two amino acid residues are different between human and mouse SURF-5A and eight amino acid residues are different between human and mouse SURF-5B.

3.4. Sub-cellular localization of SURF-5A and SURF-5B proteins

To study the intracellular localization of SURF-5 proteins we have constructed expression plasmids encoding fusion proteins between the full-length Surf-5a and Surf-5b cDNAs and the green fluorescent protein (GFP) from the jelly fish, *Aequorea victoria*. The subcellular distribution of the GFP-

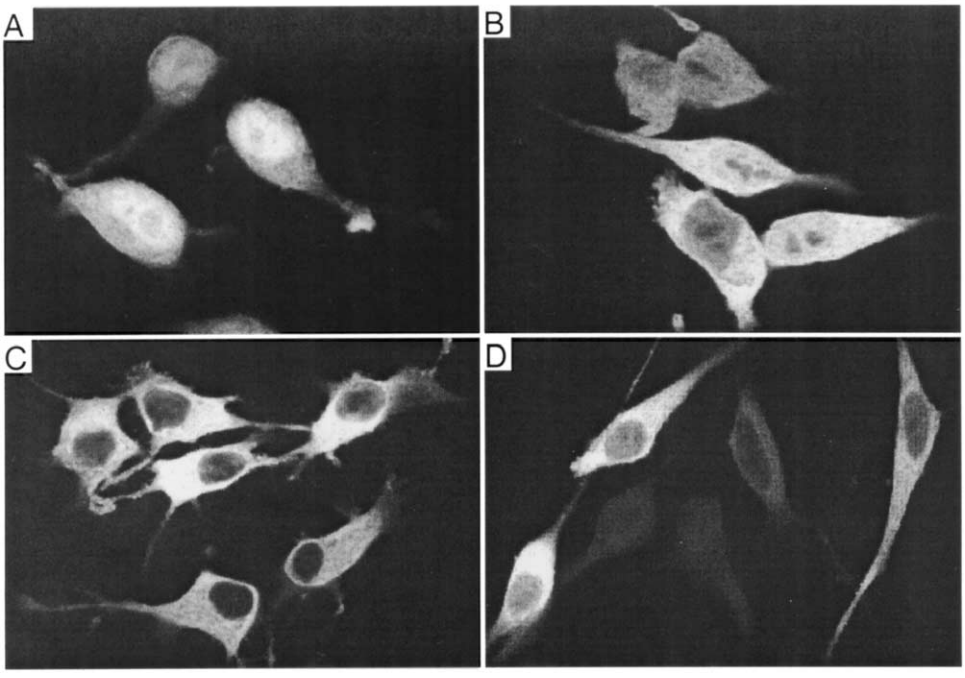


Fig. 6. Subcellular distribution of GFP-SURF-5 fusion proteins in HeLa cells. HeLa cells were transfected with DNA from expression vectors encoding GFP alone (A), a GFP-SURF-5A fusion protein (B), a GFP-SURF-5B fusion protein (C), or the mutant GFP-SURF-5BΔ (D). The GFP fluorescence was analysed by using a Zeiss laser scan microscope (LSM410).

SURF-5 fusion proteins was investigated in transient expression experiments by transfecting DNA into HeLa cells followed by confocal microscopy observation. Cells transfected with a construct encoding GFP alone displayed diffuse fluorescence throughout (Fig. 6A). When GFP was fused to the amino terminus of SURF-5A and SURF-5B proteins, an intense cytoplasmic fluorescence was observed (Fig. 6B,C), although some nuclear staining could be observed in cells transfected with Surf-5a cDNA. To examine the possibility that a difference in molecular weight could be responsible for the different behaviour of the two proteins, we designed an additional construct coding for a mutant GFP-SURF-5B protein (GFP-SURF-5bΔ), in which 60 amino acids (aa residues 17–77) shared between the two proteins were deleted, resulting in a construct encoding for a mutated SURF-5B protein with the same molecular weight as SURF-5A. Analysis of the cells transfected with this construct showed a cellular localization of the GFP-SURF-5BΔ protein similar to the SURF-5A protein (Fig. 6D). We concluded on the basis of these findings that both proteins share a cytoplasmic localization and demonstrated a crucial role for the molecular size in the localization of a protein, when a specific targeting signal is not present.

Acknowledgements

This research has been supported by Consiglio Nazionale delle Ricerche (CNR, Rome), and by MURST. The authors wish to thank Isabella Buttino for help with the confocal microscopy experiments, and Notis Argiriou for the sequence analysis.

References

Burbelo, P.D., Martin, G.R., Yamada, Y., 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.

Byrd, P.J., Cooper, P.R., Stankovic, T., Kullar, H.S., Watts, G.D.J., Robinson, P.J., Taylor, A.M.R., 1996. A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: Amino acid sequence, structure and expression studies. Hum. Mol. Genet. 5, 1785–1791.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Church, G.M., Gilbert, W., 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA 81, 1991–1995.

Colombo, P., Yon, J., Garson, K., Fried, M., 1992. Conservation of the organization of five tightly clustered genes over 600 millions years of divergent evolution. Proc. Natl. Acad. Sci. USA 89, 6358–6362.

D'Esposito, M., Mazzarella, R., Pengue, G., Jones, C., D'Urso, M., Schlessinger, D., 1994. PCR-based immortalization and screening of hierarchical pools of cDNAs. Nucleic Acids Res. 22, 4806–4808.

De Falco, S., Russo, G., Angiolillo, A., Pietropaolo, C., 1993. Human L7a ribosomal protein: sequence, structural organization, and expression of a functional gene. Gene 126, 227–235.

Duhig, T., Ruhrberg, C., Mor, O., Fried, M., 1998. The human surfeit locus. Genomics 52, 72–78.

Garson, K., Duhig, T., Armes, N., Colombo, P., Fried, M., 1995. 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.

Garson, K., Duhig, T., Fried, M., 1996. Tissue-specific processing of the Surf-5 and Surf-4 mRNAs. Gene Expr. 6, 209–218.

Gavalas, A., Zalkin, H., 1995. Analysis of the chicken GPAT/AIRC bidirectional promoter for de novo purine nucleotide synthesis. J. Biol. Chem. 270, 2403–2410.

Gessler, M., Bruns, G.A., 1993. Sequence of the WT1 upstream region including the Wit-1 gene. Genomics 17, 499–501.

Giallongo, A., Yon, J., Fried, M., 1989. Ribosomal protein L7a is encoded by a gene (Surf-3) within the tightly clustered mouse Surfeit Locus. Mol. Cell. Biol. 9, 224–231.

Gorman, C.M., Moffat, L.F., Howard, B.H., 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2, 1044–1051.

Hentschel, C.C., Birnstiel, M.L., 1981. The organization and expression of histone gene families. Cell 25, 301–313.

Huang, A., Campbell, C.E., Bonetta, L., McAndrews-Hill, M.S., Chilton-MacNeill, S., Coppes, M.J., Law, D.J., Feinberg, A.P., Yeger, H., Williams, B.R.G., 1990. Tissue, developmental, and tumor-specific expression of divergent transcripts in Wilms tumor. Science 250, 991–994.

Huxley, C., Fried, M., 1990. The mouse Surfeit Locus contains a cluster of six genes associated with four Cp-G islands in 32 kilobases of genomic DNA. Mol. Cell. Biol. 10, 605–614.

Imai, T., Yamauchi, M., Seki, N., Sugawara, T., Saito, T., Matsuda, Y., Ito, H., Nagase, T., Normura, N., Hori, T.-A., 1996. Identification and characterization of a new gene physically linked to the ATM gene. Genome Res. 6, 439–447.

Kozak, M., 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.

Lennard, A.C., Fried, M., 1991. The bidirectional promoter of the divergently transcribed mouse Surf-1 and Surf-2 genes. Mol. Cell. Biol. 11, 1281–1294.

Linton, J.P., Yen, J.-Y.J., Selby, E., Chen, Z., Chinsky, J.M., Liu, K., Kellems, R.E., Crouse, G.F., 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.

Magoulas, C., Fried, M., 1996. The Surf-6 of mouse surfeit locus encodes a novel nucleolar protein. DNA Cell Biol. 15, 305–316.

Magoulas, C., Fried, M., 2000. Isolation and genomic analysis of the human Surf-6 gene: a member of the Surfeit locus. Gene 243, 115–123.

Mashkevich, G., Repetto, B., Glerum, D.M., Jin, C., Tzagoloff, A., 1997. SHY1, the yeast homolog of the mammalian SURF-1 gene, encodes a mitochondrial protein required for respiration. J. Biol. Chem. 272, 14356–14364.

Reeves, J.E., Fried, M., 1995. The surf-4 gene encodes a novel 30 kDa integral membrane protein. Mol. Membr. Biol. 12, 201–208.

Rosen, S.W., Weintraub, B.D., Aaronson, S.A., 1980. Non random ectopic protein production by malignant cells; direct evidence in vitro. J. Clin. Endocrinol. Metab. 50, 834–841.

Sambrook, J., Fritsch, E.F., Maniatis, M., 1989. Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kuhn, K., Poschl, E., 1993. Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. Biochim. Biophys. Acta 1174, 1–10.

Tiranti, V., et al., 1998. Mutations of SURF-1 in Leigh disease associated with cytochrome c oxidase deficiency. Am. J. Hum. Genet. 63, 1609–1621.

Wigler, M., Silverstein, S., Lee, L., Pellicer, A., Cheng, T., Axel, R., 1977.

Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell 11, 223–232.

Williams, T., Fried, M., 1986. A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature 322, 275–279.

Yon, J., Jones, T., Garson, K., Sheer, D., Fried, M., 1993. 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.

Zhu, Z., Yao, J., Johns, T., Fu, K., De Bie, I., Macmillan, C., Cuthbert, A.P., Newbold, R.F., Wang, J., Chevrette, M., Brown, G.K., Brown, R.M., Shoubridge, E.A., 1998. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. Nat. Genet. 20, 337–343.