MOLECULAR AND CELLULAR BIOLOGY, Sept. 1988, p. 3898–3905
0270-7306/88/093898-08$02.00/0

Vol. 8, No. 9

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

CLARE HUXLEY, TREVOR WILLIAMS,† 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, England

Received 29 April 1988/Accepted 16 June 1988

The mouse surfeit locus is unusual in that it contains a number of closely clustered genes (*Surf-1*, *-2*, and *-4*) that alternate in their direction of transcription (T. Williams, J. Yon, C. Huxley, and M. Fried, Proc. Natl. Acad. Sci. USA 85:3527–3530, 1988). The heterogeneous 5′ ends of *Surf-1* and *Surf-2* are separated by 15 to 73 base pairs (bp), and the 3′ ends of *Surf-2* and *Surf-4* overlap by 133 bp (T. Williams and M. Fried, Mol. Cell. Biol. 6:4558–4569, 1986; T. Williams and M. Fried, Nature (London) 322:275–279, 1986). A fourth gene in this locus, *Surf-3*, which is a member of a multigene family, has been identified. The poly(A) addition site of *Surf-3* lies only 70 bp from the poly(A) addition site of *Surf-1*. Transcription of *Surf-3* has been studied in the absence of the other members of its multigene family after transfection of a cloned genomic mouse DNA fragment, containing the *Surf-3* gene, into heterologous monkey cells. *Surf-3* specifies a highly expressed 1.0-kilobase mRNA that contains a long open reading frame of 266 amino acids, which would encode a highly basic polypeptide (23% Arg plus Lys). The other members of the *Surf-3* multigene family are predominantly, if not entirely, intronless pseudogenes with the hallmarks of being generated by reverse transcription. The role of the very tight clustering on regulation of expression of the genes in the surfeit locus is discussed.

The MES-1 (mouse expression sequence) element was isolated by its ability to restore expression to a test gene lacking its 5′ regulatory sequences, using the expression selection technique (2). More recently, MES-1 has been found to display enhancerlike activity with heterologous genes in both long- and short-term expression assays (17). In its chromosomal location in the mouse genome, MES-1 has been found in a highly transcribed region called the surfeit locus (22). MES-1 lies between the 5′ ends of two closely spaced divergent transcription units, *Surf-1* and *Surf-2*, whose heterogenous 5′ ends are separated by only 15 to 73 base pairs (20). Another transcription unit, now referred to as the *Surf-4* transcription unit, lies downstream from *Surf-2* with the 3′ end of one of its processed transcripts overlapping the 3′ end of the converging processed *Surf-2* mRNAs by 133 bp (21). In this communication, we describe the structure and location of the *Surf-3* transcription unit, which is also located in this very tight gene cluster. Figure 1 shows the organization of the four identified genes (*Surf-1* to *-4*) in the surfeit locus.

We have studied *Surf-3* transcription independent of other members of the mouse multigene family to which it belongs after DNA transfection of a cloned mouse DNA fragment into monkey COS cells. The *Surf-3* mRNAs are highly expressed, and the putative *Surf-3* translation product is a highly basic polypeptide. The poly(A) addition site of *Surf-3* lies only 70 bp away from the poly(A) addition site of *Surf-1*. The other members of the multigene family are shown to be predominantly, if not entirely, processed pseudogenes.

* Corresponding author.

† Present address: Department of Biochemistry, University of California at Berkeley, Berkeley, CA 94720.

MATERIALS AND METHODS

Southern and Northern (RNA) blotting. Mouse BALB/c 3T3 DNA was digested with *Bam*HI or *Hin*dIII or *Eco*RI, fractionated by electrophoresis on a 0.8% agarose gel, and transferred to nitrocellulose as described previously (5). For Northern blotting, total cellular RNA was prepared from BALB/c TS-A-3T3 cells (14) by the guanidinium-CsCl method and poly(A) selected on oligo(dT)-cellulose.

The isolation of poly(A)$^+$ RNA and analysis by Northern blotting was essentially as described previously (9). Poly(A)$^+$ RNA was fractionated on a 1% agarose–formaldehyde gel and transferred to nitrocellulose after partial alkaline hydrolysis. Hybridization conditions were identical for Southern and Northern blotting, and both types of blots were washed in 0.5× SSC (1× SSC is 0.1 M NaCl plus 0.015 M sodium citrate) at 68°C. Probes were labeled by nick translation.

Sequencing. DNA was sequenced by the dideoxy method (15) after appropriate restriction fragments had been subcloned into either M13mp18 or M13mp19 (23).

cDNA and genomic libraries. The cDNA was made as described previously (4), except that after second-strand synthesis, the cDNA was repaired with T4 DNA polymerase and ligated into λGT10 (Strategene) after the addition of *Eco*RI linkers. To construct the genomic DNA library, BALB/c 3T3 DNA was digested with *Eco*RI, ligated into λGT10 (Strategene), and packaged in Gigapack Gold (Stratagene).

Transfection of COS cells. One day prior to transfection, 1.5×10^6 to 3.0×10^6 COS-1 cells were seeded onto 15-cm-diameter tissue culture dishes. These cells were transfected with 40 μg of plasmid DNA by the method described by Wigler et al. (19) with the following modifications. After addition of the DNA-$CaPO_4$ precipitate, the cells were grown overnight at 32°C and then glycerol shocked by the addition of 3 ml of 15% (vol/vol) glycerol in DMEM plus

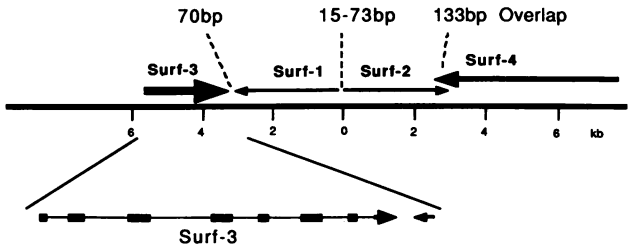

FIG. 1. Map of the mouse surfeit locus. Genomic DNA is shown as a solid bar with distances in kilobase pairs marked underneath. The direction and extent of the *Surf-1* to *-4* transcription units are shown as solid arrows above the genomic DNA, with the thickness of the arrows representing the steady-state level of the mRNAs. The 133-bp overlap of *Surf-2* and *-4* (21), the 15 to 73 bp between the 5′ ends of *Surf-1* and *-2* (20), and the 70 bp between the 3′ ends of *Surf-3* and *-1* (Fig. 6) are indicated. Underneath the genomic clone there is an expanded diagram of the *Surf-3* transcription unit, with the exons of *Surf-3* shown as solid boxes and the introns as a connecting line (Fig. 5). The 3′ exon of *Surf-1* is also shown.

0.1% (wt/vol) glutamine for 60 s. The cells were washed twice with DMEM plus 0.1% (wt/vol) glutamine, refed, grown for 48 h, and then harvested for RNA. The RNA was prepared as described above and treated with RNase-free DNase I before poly(A) selection.

RESULTS

Identification of the *Surf-3* transcription unit. A 19-kilobase (kb) *Eco*RI fragment containing MES-1 and the *Surf-1* and *Surf-2* transcription units has previously been isolated from mouse cellular DNA (20). Subsequently, sequences located directly downstream from *Surf-1* have been found to hybridize to a discrete abundant poly(A)$^+$ RNA species of about 1.0 kb. In order to study the transcription from this region in more detail, cDNA clones derived from poly(A)$^+$ RNA from mouse A31 BALB/c 3T3 cells hybridizing to the unique sequence in the 1.6-kb *Bgl*II-*Bam*HI mouse DNA fragment located directly downstream of the 3′ end of *Surf-1* were isolated. One cDNA clone (MY3), containing a 0.85-kb insert homologous to this region, was studied in greater detail. A discrete abundant mRNA of about 1.0 kb was detected in mouse poly(A)$^+$ RNA when MY3 was used as a probe (Fig. 2a). The region from which these transcripts are derived is termed the *Surf-3* transcription unit (Fig. 1).

***Surf-3* is an expressed member of a multigene family.** To ascertain whether the *Surf-3* transcription unit represented a single-copy sequence in the mouse genome, the MY3 cDNA was used to probe Southern blots of endonuclease-restricted mouse DNA. MY3 hybridized to multiple bands in mouse DNA digested separately with a number of different restriction enzymes (Fig. 2b). This indicates either that MY3 contains a repetitive element or that it is derived from a multigene family of about 15 to 30 genes. Therefore, MY3 was subdivided into three fragments of approximately 300 bp and each was used individually as a probe for Southern blotting (Fig. 3; probes A, B, and C). All three probes were found to hybridize to multiple bands in mouse cellular DNA digested separately with a number of different restriction enzymes (Fig. 3a). Thus, it appears that the *Surf-3* transcription unit present in the 19-kb clone is a member of a multigene family. Consequently, MY3 may be derived from an mRNA transcribed from any member of this family and not necessarily the copy present in the 19-kb clone containing the surfeit locus. In addition, *Surf-3* also appears to be a

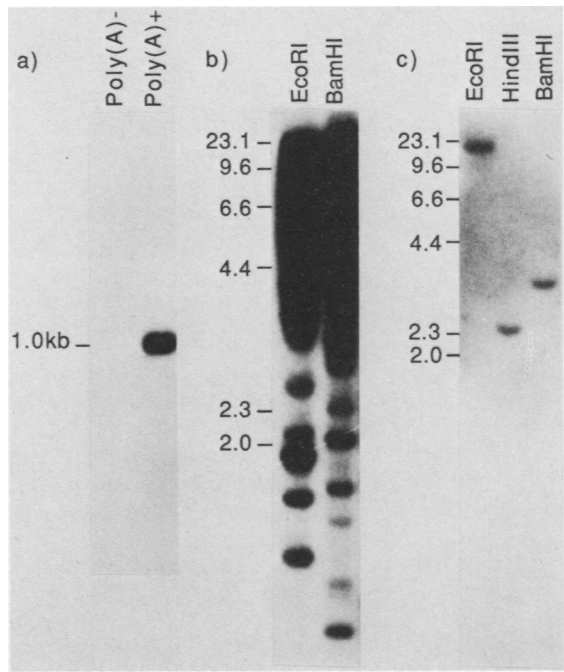

FIG. 2. (a) Northern blot showing *Surf-3* mRNA. Poly(A)$^+$ RNA (1 μg) or 20 μg of Poly(A)$^-$ RNA from mouse TS-A-3T3 cells (14) were fractionated, and the blot was probed with the cDNA MY3. The size of the mRNA detected was calculated from the positions of the 18S and 28S rRNAs. (b) Southern blot of mouse BALB/c 3T3 DNA cut with *Eco*RI or *Bam*HI and probed with *Surf-3* cDNA MY3. (c) Southern blot of mouse BALB/c 3T3 DNA cut with *Eco*RI, *Hin*dIII, or *Bam*HI and probed with the third intron of *Surf-3*. The numbers on the left indicate the size in kilobases of the six largest *Hin*dIII fragments of λ phage DNA used as size markers.

member of a multigene family in both monkey and human DNA (Fig. 3b).

In order to analyze the expression of *Surf-3* in the absence of possible related transcripts derived from elsewhere in the mouse genome, the previously cloned 19-kb *Eco*RI fragment (20) containing sequences from the four *Surf* genes was introduced into the heterologous monkey COS cells (3). COS cells contain the simian virus 40 (SV40) replication protein large T antigen and are capable of replicating DNA molecules containing the SV40 origin of replication. Two plasmids containing the 19-kb clone were constructed (Fig. 4a). Plasmid p19 contains the 19-kb EcoRI fragment cloned into pAT153 and is not capable of efficient replication in COS cells, while pSV19 also contains the SV40 origin of replication, in addition to the 19-kb fragment, and can replicate in COS cells.

Poly(A)$^+$ RNA was isolated from COS cells 48 h after transfection with either pAT153, p19, or pSV19 and analyzed by Northern blotting by using the *Surf-3*-specific 1.6-kb *Bgl*II-*Bam*HI fragment as a probe (Fig. 4b). This probe hybridizes to the abundant *Surf-3* mRNA species of 1.0 kb in TS-A-3T3 mouse fibroblast and undifferentiated F9 teratocarcinoma mouse cells but also identified similarly sized transcripts in the COS cells transfected with pAT153 alone (Fig. 4b). This indicates that mRNAs with some homology to the *Surf-3* transcript are also present in monkey cells. In cells transfected with p19 the level of the *Surf-3* transcripts was increased approximately three- to fivefold above the background level of the pAT153-transfected COS cells (Fig. 4b). In contrast, cells transfected with pSV19

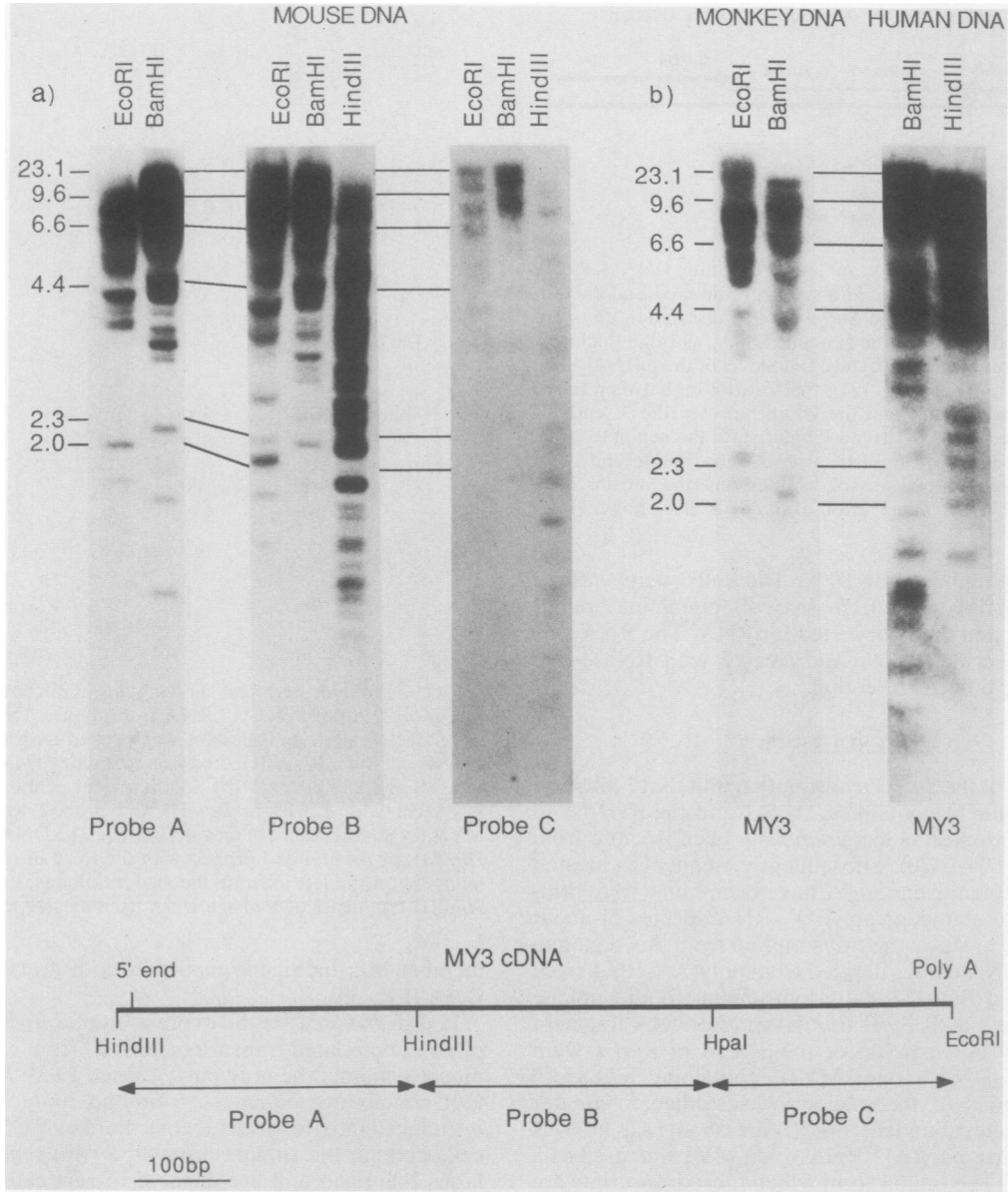

FIG. 3. Southern blots of genomic DNA probed with *Surf-3* cDNA sequences. (a) Genomic DNA isolated from mouse TS-A-3T3 cells was digested with *Eco*RI, *Hin*dIII, or *Bam*HI and probed with probe A, probe B, or probe C (see map below). (b) Monkey or human cellular DNA digested with *Eco*RI, *Hin*dIII, or *Bam*HI and probed with the complete *Surf-3* cDNA MY3. Beneath the blots is a map of the *Surf-3* cDNA MY3 showing the three probes, A, B, and C used above. The numbers on the left indicate the size in kilobases of the six largest *Hin*dIII fragments of λ phage DNA; the relative positions of these markers in the six gels are shown by the connecting lines.

produced large amounts of a 1.0-kb *Surf-3* transcript (Fig. 4b). Thus it appears that *Surf-3* is correctly transcribed and processed in monkey cells. Therefore, the *Surf-3* transcription unit is likely to be a bona fide gene, rather than a nonexpressed pseudogene. In addition, it implies that the sequences required for efficient transcription of *Surf-3* are also present within the 19-kb *Eco*RI fragment.

Analysis of *Surf-3* cDNAs. To obtain further information on the *Surf-3* transcripts produced by the 19-kb fragment, a cDNA library was constructed from poly(A)$^+$ RNA isolated from pSV19-transfected COS cells. Approximately 200,000 recombinant lambda clones were screened by using the 1.6-kb *Bgl*II-*Bam*HI fragment as a probe, and 24 mouse-specific positive clones were isolated and subcloned into plasmid vectors. The eight positives containing the largest cDNA inserts were analyzed in further detail and found to have similar restriction maps to the MY3 cDNA isolated from mouse cells (data not shown). The 5′ ends of five of these cDNAs were sequenced and found to have 100% identity with sequences at the 5′ end of MY3 and to end within a couple of nucleotides from the 5′ end of the mouse MY3 cDNA clone (Fig. 5). One of the cDNAs, C2C, derived from the transfected monkey cells was sequenced and found to have complete identity with sequences of the 19-kb clone specifying *Surf-3* exons (Fig. 5). The sequence of C2C differed from the mouse cell-derived MY3 cDNA at only one nucleotide of the 850-bp sequence (Fig. 5). This single-base difference could be due to the presence of a base change in the other mouse chromosomal allele of *Surf-3* from which MY3 is derived. Alternatively, it could be the result of a

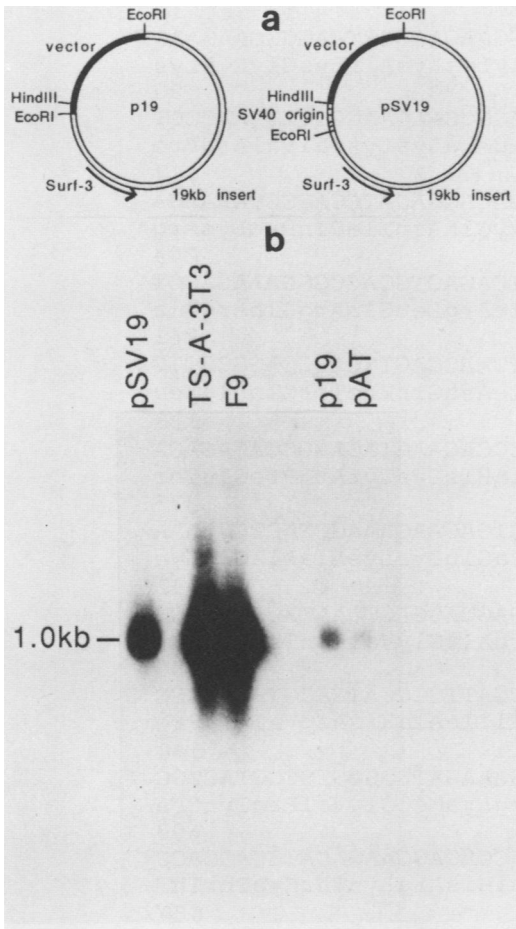

FIG. 4. Expression of *Surf-3* in COS-1 cells. (a) Plasmids used to transfect COS-1 cells. Vector sequences (▬), genomic sequences (▭), and the SV40 origin region (▥) are shown. P19 has the previously cloned 19-kb *Eco*RI mouse cellular DNA fragment (20) inserted as shown, with the arrow indicating the direction and extent of the *Surf-3* transcription unit. PSV19 has, in addition to the 19-kb *Eco*RI fragment, the SV40 origin of replication. The SV40 origin is contained in a *Hin*dIII-to-*Eco*RI fragment taken from the plasmid pSVOD (11) and corresponds to SV40 nucleotides 5171 to 160. (b) Northern blot of RNA from transfected and control cells probed with the 1.6-kb *Bam*HI-*Bgl*II fragment. Starting from the left, the lanes contain: 1.2 μg of poly(A)$^+$ RNA from COS cells transfected with pSV19, 4.5 μg of poly(A)$^+$ RNA from mouse BALB/c TS-A-3T3 cells, 4.5 μg of poly(A)$^+$ RNA from mouse F9 cells, 1.5 μg of poly(A)$^+$ RNA from COS cells transfected with p19, and 1.5 μg of poly(A)$^+$ RNA from COS cells transfected with pAT153 as indicated.

mutation generated during the in vitro synthesis of the MY3 cDNA or its growth in bacteria. In Fig. 5 is presented the DNA sequence and potential translation product of the *Surf-3* mRNA. The *Surf-3* transcript contains a large open reading frame of 266 amino acids which would code for a basic protein containing 61 arginine and lysine residues (23%) distributed over the entire polypeptide chain. A comparison of the cDNA sequence with the sequence of the 19-kb genomic clone has located the positions of eight exons in the *Surf-3* gene (Fig. 1). Each of the *Surf-3* intron-exon boundaries conforms to the consensus sequences for splice junctions (Table 1). Furthermore, when endonuclease-restricted mouse cellular DNA is hybridized with a probe from intron 3 of *Surf-3*, single-copy restriction fragments characteristic of the surfeit locus (e.g., 19-kb *Eco*RI fragment, 2.4-kb *Hin*dIII fragment, and 3.0-kb *Bam*HI fragment) are detected (Fig. 2c), showing that there is only one copy of the *Surf-3* intron in the mouse genome.

The *Surf-3* poly(A) addition site, as determined from the cDNA, is 12 bp downstream from the variant poly(A) addition signal AAUAUA (12, 16). The *Surf-3* poly(A) addition site is on the opposite strand and only 70 bp away from the poly(A) addition site of the converging *Surf-1* transcription unit (20) (Fig. 6).

The other members of the *Surf-3* multigene family are predominantly, if not entirely, intronless pseudogenes. In order to investigate the nature of the other members of the *Surf-3* multigene family, a library of *Eco*RI fragments from mouse cellular DNA was constructed by using the lambda gt10 vector. A total of 10 positives were detected from a total of 50,000 plaques screened with a probe containing the 300 nucleotides from the extreme 5′ end of the *Surf-3* MY3 cDNA. Eight of these positive clones were analyzed further after subcloning of the *Eco*RI inserts into bluescript plasmids. All eight clones were found to be positive when screened with a probe with the extreme 3′ end of the MY3 cDNA. A restriction enzyme analysis indicated that the sequence hybridizing to the 5′ and 3′ MY3 cDNA probes in each of the eight clones was contained in about 1.0 kb of DNA (data not shown) as opposed to 2.5 kb containing the *Surf-3* transcription unit (Fig. 1). This result indicates that the eight clones might be intronless pseudogenes of *Surf-3*. The presence of pseudogenes was confirmed by sequencing three of the clones (Fig. 7). All three are intronless pseudogenes and show 87 to 95% identity to the *Surf-3* cDNA as a result of additions, deletions, and base changes which alter and prematurely terminate the *Surf-3* reading frame. In one of the pseudogenes, the *Surf-3* ATG start codon is also lost by mutation (Fig. 7). Each of the pseudogenes has the hallmarks of being generated as a result of reverse transcription. In addition to the loss of introns, each pseudogene contains a different direct repeat of 13 to 15 bp at its 5′ and 3′ ends. The 3′ repeat is preceded by a run of A residues which begins at the site of polyadenylation of the *Surf-3* cDNAs (compare Fig. 6 and 7), indicating that a poly(A)-containing *Surf-3* mRNA was used as the template for the reverse transcription.

DISCUSSION

The *Surf-3* transcription unit, located in the mouse surfeit locus (22), has been identified and characterized. The analysis of *Surf-3* is complicated by its being a member of a mouse multigene family estimated to consist of 15 to 30 genes. To overcome this complication, we have studied *Surf-3* transcription in the absence of other members of the multigene family. This was accomplished by transfection of a cloned fragment of mouse DNA containing *Surf-3* into heterologous monkey cells. Transcription and cDNA analysis of the transfected monkey cells has demonstrated that a 1.0-kb mRNA is encoded by the *Surf-3* transcription unit which is identical (except for a single base) to the sequence of the mRNA produced in mouse cells.

By cloning, restriction enzyme mapping, and DNA sequence analyses, the other members of the *Surf-3* multigene family have been shown to be predominantly, if not entirely, intronless pseudogenes, whereas there is only one copy of the *Surf-3* third intron in the mouse genome. The *Surf-3* pseudogenes have the hallmarks of being created by reverse transcription. From a comparison of the sequence differ-

FIG. 5. Sequence of *Surf-3* mRNA and its potential translation product (266 amino acids). The top line indicates the DNA sequence derived from cDNA and genomic sequences, and the line underneath shows the potentially coded protein starting at the first ATG. The cDNA MY3 contains a C instead of the T at position 544. The T is found in the genomic DNA and cDNA C2C. The AATATA poly(A) addition signal is underlined, and the position of each intron is indicated (see also Fig. 1). Numbering is from the A of the initiation codon, and * indicates the 5′ end of cDNA MY3.

ences between *Surf-3* and the pseudogenes, it can be estimated that the pseudogenes were formed 10 to 25 million years ago (6). It is interesting to note that multigene families with homology to mouse *Surf-3* are also found in monkey and human DNA (Fig. 3b) and are presumably also mainly composed of pseudogenes. *Surf-3* has the properties of a housekeeping gene, being abundantly transcribed in all differentiated cell types tested (22). This is consistent with the detection of a large multigene family (15 to 30 genes), primarily composed of processed pseudogenes which were presumably derived from reverse transcription of abundant mRNAs present in the germ line and subsequent integration into the chromosomal DNA of the germ line to become incorporated into the species (for a review, see reference 18).

The putative translation product of *Surf-3* is a highly basic protein (23% Arg plus Lys) (Fig. 5), suggesting that the function of the *Surf-3* gene product might be to interact with nucleic acids. *Surf-3* does appear to be highly conserved in other species (J. Yon and M. Fried, manuscript in preparation). We cannot find any extensive homology of *Surf-3* with any other genes in the Surfeit locus or any other DNA or

TABLE 1. Splice junctions in *Surf-3*[a]

5′ junction	Size (bp) of intron	3′ junction
1[b] AAGATG | GTGAGT	174	CTCTTCCGTACAG | CCCAA
2 GCATTG | GTAAGG	311	CTTAACTTCCCAG | GACAG
3 AGACAG | GTGAGG	431	ATGTGTTTTCCAG | CTACT
4 GAGCAG | GTCAGT	199	TTGTATGTTCCAG | GAGTC
5 ATTGAG | GTAGGT	225	TGTTTGCTTCTAG | CTGGT
6 TAACTC | GTGAGT	185	ATTGCCTTTCCAG | GGAAG
7 GACGAG | GTGAGT	262	TCTGTCATTTCAG | ATCCG
Consensus		
$^{A}_{C}$AG | GT$^{G}_{A}$AGT		TTTTTTTT$_{N}^{T}$ CCCCCCCC$_{C}^{A}$AG | G

[a] See Fig. 5 for the positions of each intron in the mRNA. The splice junction consensus sequences are taken from Mount (13).
[b] Intron.

protein sequence in the EMBL, GenBank, or PIR data bases. Two candidate gene groups to which *Surf-3* might belong are the ribosomal proteins and the HMG, or high-mobility group proteins. The mRNAs specifying members of both of these groups of nucleic acid-binding proteins are highly abundant, and the genes for these proteins are members of processed pseudogene-containing multigene families (1, 7, 8). We are presently attempting to produce anti-peptide antibodies against the putative *Surf-3* translation product in order to be able to study the protein in greater detail.

The genes in the surfeit locus are very tightly clustered (Fig. 1). The heterogeneous 5′ ends of the divergent *Surf-1* and *Surf-2* transcription units are separated by only 15 to 73 bp (20), and the 3′ ends of the convergent *Surf-2* and *Surf-4* processed transcripts overlap by 133 bp (21). The location of *Surf-3* is in keeping with the clustered nature of the genes in this locus. The poly(A) addition site of *Surf-3* lies only 70 bp away from the poly(A) addition site of the converging *Surf-1* transcription unit (Fig. 1 and 6). The clustered organization of the surfeit locus, which may contain more than the four genes already identified, is conserved in other mammalian species, suggesting some sort of *cis* interaction between the *Surf* genes (22). Transcription termination has not been clearly defined in mammalian cells, but it is clear that polyadenylated mRNA is derived by processing of a much longer precursor RNA transcript. It might be expected that abundant transcription of the *Surf-3* gene would interfere with *Surf-1* gene expression. This might occur either through

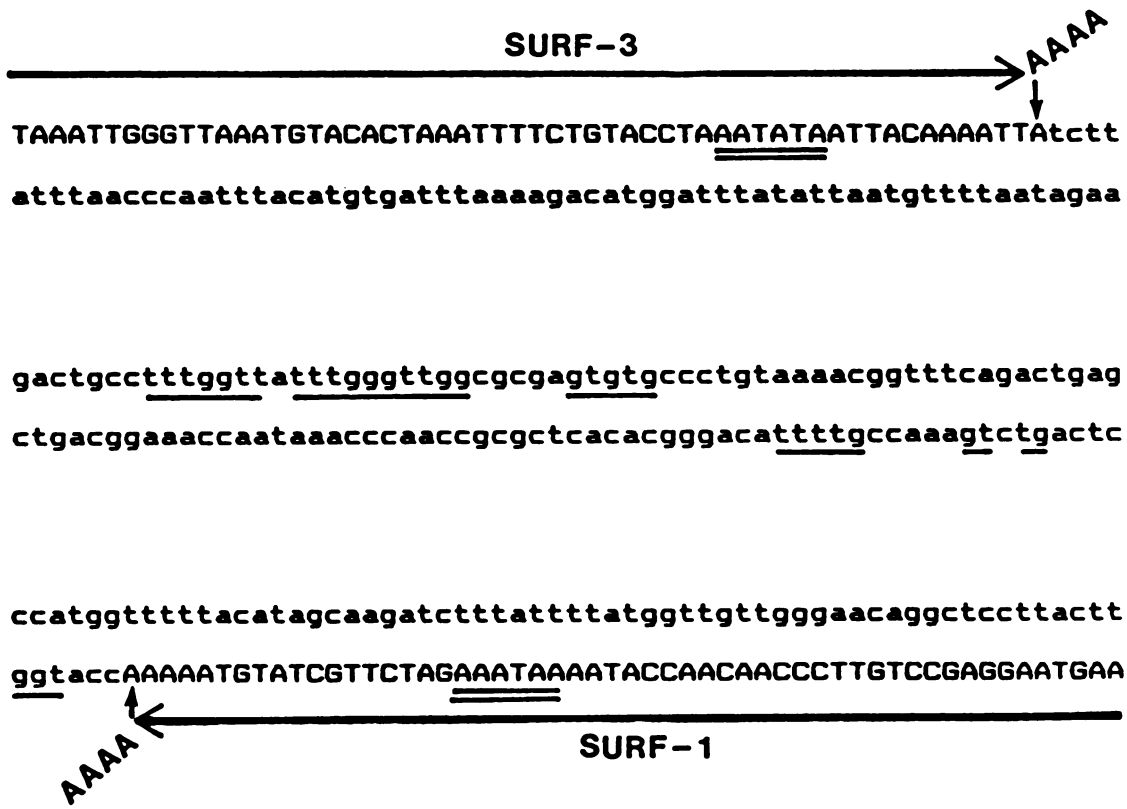

FIG. 6. DNA sequence spanning the 70 bp between the 3′ ends of *Surf-3* and *Surf-1*. The two converging genes are shown as solid arrows, and the coding strand of each gene is shown as capital letters. Each poly(A) addition signal is underlined twice, and the TG-rich sequences seen downstream of many poly(A) sites (10) are underlined on the appropriate strand for each gene.

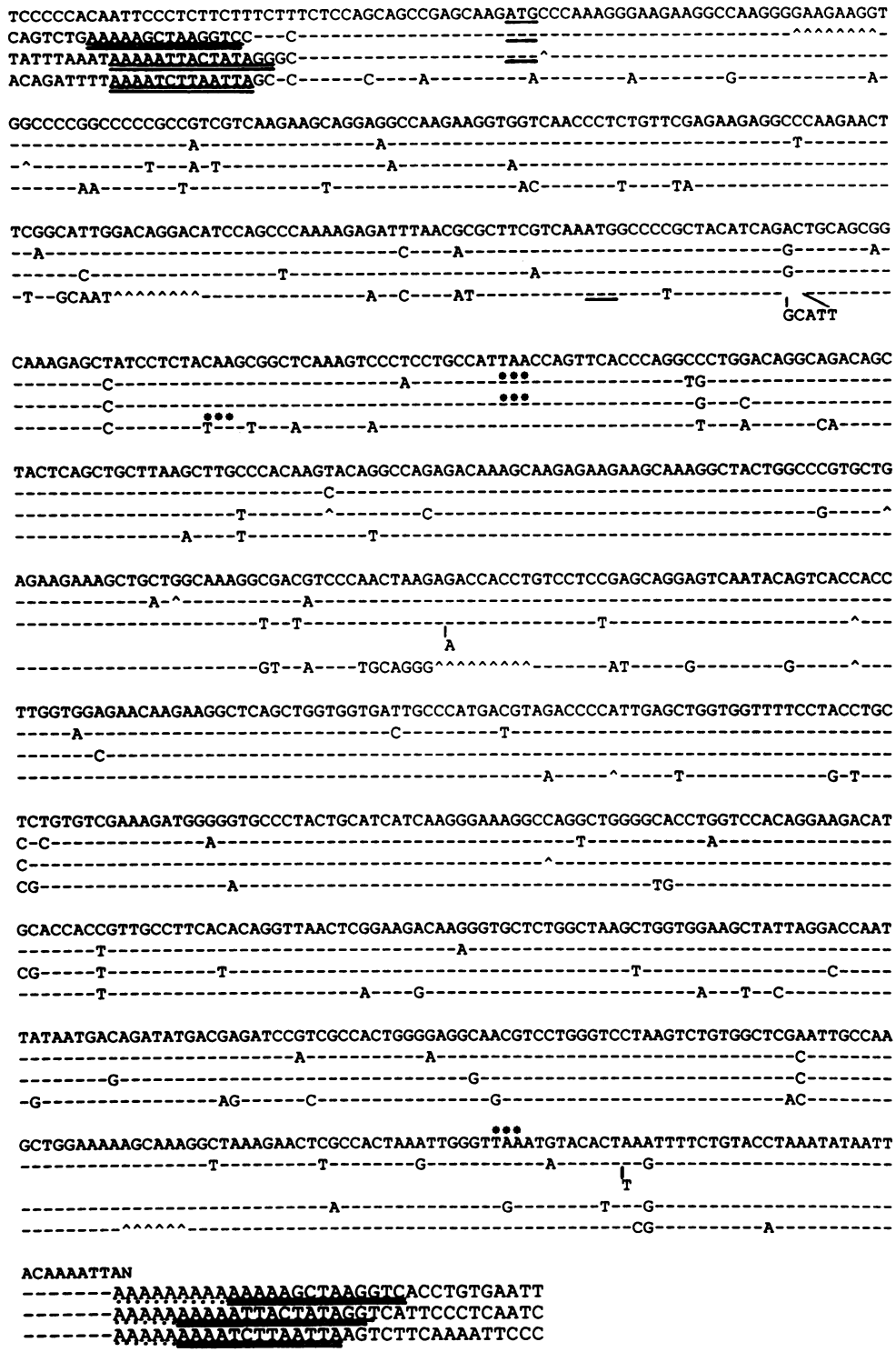

FIG. 7. Comparison of *Surf-3* and three of its pseudogenes. The top line shows the sequence of *Surf-3* (Fig. 5). Underneath are shown three different pseudogenes. Bases which are the same as those in *Surf-3* (–) and deletions (^) are indicated. Additions and changes are marked by the relevant bases. The first AUG of each gene is indicated by underlining, and the first in-frame termination codon is indicated by asterisks above. The 13- to 15-bp terminal repeats are underlined twice, and the poly(A) stretch preceding the 3′ repeat is indicated by dots beneath the sequence. The AN at the 3′ end of *Surf-3* indicates the poly(A) tract.

the production of an excess of anti-sense RNA to the less abundant *Surf-1* transcript in the form of the very abundant *Surf-3* primary transcript or via polymerase interference from the abundant *Surf-3* transcription with the *Surf-1* promoter which lies only 3.1 kb downstream from the 3′ end of the *Surf-3* gene. Alternatively, the presence of a transcription termination signal in the 70 bp that separates *Surf-1* and *Surf-3* might prevent readthrough of *Surf-3* into *Surf-1*. A further study of the transcriptional and posttranscriptional events involving the genes in the surfeit locus is required to determine the role of the very tight clustering in their expression.

ACKNOWLEDGMENTS

We are indebted to Moira Read for her excellent technical assistance. We also thank C. Dixon, S. Goodbourn, and N. Jones for their helpful comments during the preparation of this manuscript.

LITERATURE CITED

1. **Dubov, K. P., and R. P. Perry.** 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell **37**:457–468.
2. **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.
3. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell **23**:175–182.
4. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
5. **Hayday, A., H. E. Ruley, and M. Fried.** 1982. Structural and biological analysis of integrated polyoma virus DNA and its adjacent host sequences clones from transformed rat cells. J. Virol. **44**:67–77.
6. **Kimura, M.** 1983. The neutral theory of molecular evolution, p. 208–233. *In* M. Nei and R. K. Kohn (ed.), Evolution of genes and proteins. Sinauer Associates Inc., Sunderland, Mass.
7. **Landsman, D., N. Soares, F. J. Gonzalez, and M. Bustin.** 1986. Chromosomal protein HMG-17. Complete human cDNA sequence and evidence for a multigene family. J. Biol. Chem. **261**: 7479–7484.
8. **Mager, W. H.** 1988. Control of ribosomal protein gene expression. Biochim. Biophys. Acta **949**:1–15.
9. **Maniatis, R., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **McLauchlan, J., D. Gaffney, J. L. Whitton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
11. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. Cell **27**:279–288.
12. **Milner, R. J., C. Lai, K. A. Nave, D. Lenoir, J. Ogata, and J. G. Sutcliffe.** 1985. Nucleotide sequences of two mRNAs for rat brain myelin proteolipid protein. Cell **42**:931–939.
13. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
14. **Ruley, H. E., L. Lania, F. Chaudry, and M. Fried.** 1982. Use of a cellular polyadenylation signal by viral transcripts in polymoma virus transformed cells. Nucleic Acids Res. **10**:4515–4524.
15. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
16. **Tosi, M., R. A. Young, O. Hagenbuchle, and U. Schibler.** 1981. Multiple polyadenylation sites in a mouse α-amylase gene. Nucleic Acids Res. **9**:2313–2323.
17. **von Hoyningen-Huene, V., C. Norbury, M. Griffiths, and M. Fried.** 1986. Gene activation properties of a mouse DNA sequence isolated by expression selection. Nucleic Acids Res. **14**: 5615–5627.
18. **Wagner, M.** 1986. A consideration of the origin of processed pseudogenes. Trends Genet. **2**:134–137.
19. **Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin.** 1979. DNA-mediated transfer of the adenine phosphoribosyltransferase locus into mammalian cells. Proc. Natl. Acad. Sci. USA **76**:1373–1376.
20. **Williams, T., and M. Fried.** 1986. 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.
21. **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.
22. **Williams, T., J. Yon, C. Huxley, and M. Fried.** 1988. The mouse surfeit locus contains a very tight cluster of four "housekeeping" genes which is conserved through evolution. Proc. Natl. Acad. Sci. USA **85**:3527–3530.
23. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.