Regulatory Sequences on the Human Villin Gene Trigger the Expression of a Reporter Gene in a Differentiating HT29 Intestinal Cell Line*

Sylvie Robine, Christian Sahuquillo-Merino, Daniel Louvard†, and Eric Pringault

From the Institut Pasteur, Département de Biologie Moléculaire, URA Centre National de la Recherche Scientifique 1149, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

To develop a molecular tool for tissue-specific targeting of gene expression in immature and differentiated epithelial cells of the small and large intestinal mucosa, we have isolated the 2-kb 5′-flanking region of the human villin gene. This region contains numerous short sequences that are conserved among other tissue-specific promoters of genes expressed in differentiated enterocytes. This DNA fragment promotes the transcription and expression of the luciferase reporter gene in villin-positive intestinal enterocytes, and hepatoma cell lines but not in a villin-negative keratinocyte cell line. The pattern of expression corresponds that of the endogenous gene, indicating that this sequence can direct intestine-specific transcription. In the differentiating HT29 intestinal cell line, expression of the reporter gene is already detectable in undifferentiated cells, and dramatically increases when terminal differentiation is induced. Thus, as previously reported for the endogenous gene the isolated 5′-flanking region of the villin gene responds positively to conditions known to stimulate terminal differentiation of these cultured epithelial intestinal cells. The reported results indicate that this genomic fragment contains sufficient regulatory elements to recapitulate the expression pattern of the villin promoter during intestinal differentiation.

Most reports of gene targeting in the digestive tract have used tissue-specific promoters in cells exhibiting highly specialized metabolic activity such as hepatocytes (Sandgren et al., 1988; Sifers et al., 1988; Dalemans et al., 1990; Dubois et al., 1991; Cartier et al., 1992) or exocrine and endocrine cells of the pancreas (Swift et al., 1984; Hanahan, 1985; Efrat et al., 1988). Indeed, only a few tissue-specific promoters of genes expressed in the intestinal mucosa or in other simple columnar epithelial cells have been isolated. The 5′-flanking regions of genes thus far characterized include those of human intestinal alkaline phosphatase (Millan, 1987; Henthorn et al., 1988), human sucrase-isomaltase (Wu et al., 1992), rat liver pyruvate kinase (Cognet et al., 1987), human apolipoprotein A4 (Elshourbagy et al., 1987), pig neutral aminopeptidase (Olsen et al., 1989), and human fatty acid-binding protein (Sweetser et al., 1987).

These genes share one common functional feature; they are only activated in fully differentiated epithelial cells lining the villi of the small intestine mucosa. In contrast, no promoter able to trigger transcription in undifferentiated cells of the intestinal mucosal crypt has been characterized thus far. Furthermore, promoters active in the large intestine mucosa are also lacking. It is expected that, such promoters could be used to target immortalizing genes (such as SV40 large T antigen) and relevant oncogenes and in view of recently published data (Hautf et al., 1992) in situ transformation of immature crypt intestinal epithelial cells should be favored over transformation of mature postmitotic differentiated enterocytes. In vitro benefits could include the development of new epithelial cell lines thus far not obtained by others means. In vivo, this strategy could lead to the establishment of a murine model of colorectal adenocarcinoma, a malignant tumor that occurs in human beings.

The isolation and characterisation of regulatory elements on the villin gene, which display a tissue specificity and a pattern of early expression in immature cells, could allow targeting of heterologous genes in these cell types. Villin was first characterized as a major cytoskeletal protein of the brush border in epithelial cells of the digestive and urogenital tracts, specifically epithelial cells of small and large intestinal mucosa, gallbladder, kidney proximal tubules, and testis efferent tubules (Bretscher and Weber, 1979; Robine et al., 1985; Gröne et al., 1986; Osborn et al., 1988; Elsässer et al., 1991). Villin gene expression has been studied in detail during the differentiation of epithelial cells of the digestive and urogenital tracts, both in differentiating cell lines (Pringault et al., 1986; Dudouet et al., 1987), adult tissues (Robine et al., 1985; Boller et al., 1988), and in developing mouse embryos (Ezzell et al., 1988; Maunoury et al., 1988; Maunoury et al., 1992). In addition to expression in cells forming a brush-border, villin is expressed in few epithelia derived from the endodermic lineage (e.g. ductal cells of the exocrine pancreas and biliary tract cells of the liver). In the developing mouse embryo, villin is expressed early and thus can be used to identify initial definitive endodermal cells and their derivatives. Finally, we have previously demonstrated that villin plays a key function in the biogenesis of the brush border (Friederich et al., 1989, 1992).

Previous molecular genetic studies have shown that the villin gene maps to chromosome 2 q35–36 in humans and chromosome 1 in the mouse, and belongs to a cluster of genes conserved between the two species (Rousseau-Merck et al., 1988). Expression of villin mRNA has been studied in various...
tissues from different species (Pringault et al., 1986). In humans, there are two villin mRNA transcripts differing only in an extension of 500 bp 1 in the 3'-noncoding region (Arpin et al., 1988). Cloning and characterization of the complete human villin gene (Pringault et al., 1991) have shown that the two mRNAs are equally transcribed from one gene, and could arise from a common pre-mRNA by the alternate random choice of one of two polyadenylation signals located in the last exon.

Molecular and developmental studies have disclosed the following unique features of villin gene regulation. During development, multiple levels of villin gene regulation are implicated in the histogenesis of the digestive tract: (i) a switch on at the onset of primitive and definitive endodermal cell differentiation; (ii) a silencing in cells constituting the upper part of the primitive gut and during terminal differentiation of liver parenchymal cells and acinar cells of the exocrine pancreas; and (iii) enhanced expression in fully mature enterocytes. In adults, the villin gene is already active in immature precursor cells of the intestinal crypt. Its expression increases 10-fold as enterocytes migrate and differentiate along the crypt/villus axis. The regulation of villin synthesis is controlled at the transcriptional level in embryonic and definitive endoderm and in adult normal and neoplastic tissues. In the course of neoplastic development, villin expression is retained only in those tumors and cell lines derived from villin-positive epithelia of the digestive and excretory tracts (Moll et al., 1987; Carboni et al., 1987; West et al., 1988; Bacchi and Gown, 1991).

These observations support our hypothesis that isolation of the corresponding regulatory elements of the villin gene would provide a necessary molecular tool for oncogene targeting in proliferating small and large intestinal mucosa, for triggering hyperplasia, and for the development of corresponding solid tumors.

Here we report the cloning and characterization of a 2-kb 5'-flanking region of the human villin gene which shares sequence homologies with promoters of other genes expressed in enterocytes. This region promotes the cell-specific transcription of a reporter gene when transfected into CaCO2 and HT29 colonic intestinal cell lines, the HepG2 hepatoma cell line or the LLCPK1 renal proximal tubule cell line (all of which express the endogenous villin gene) but remains inactive when transfected into villin-negative SKP keratinocyte-derived cell line. Moreover, reporter activity is up-regulated when terminal differentiation of the HT29 intestinal cell line is induced. This regulation corresponds to that previously observed for the endogenous villin gene (Pringault et al., 1986; Dudouet et al., 1987; Huet et al., 1987).

Plasmids—The plasmids pRSVL and pSVOAL (gift of J. P. Rousset, Pasteur Institute, Paris) contain the firefly luciferase gene, respectively, under the control of the LTR promoter of the Rous sarcoma virus or without any promoter (de Wet et al., 1987). The plasmid pCH110 (gift of J. P. Rousset) contains the Lac Z fragment from the Drosophila promoter of pDrLuc with the 1.7 kb Saal-HindIII insert of BSSH1.7. A HindIII-PasI fragment from BSSH0.9, which contained 300 nucleotides of immediately proximal 5'-noncoding sequence and 77 nucleotides of the first exon, was subcloned into a new BS vector (designated BSSH0.93) and inserted into an empty BS vector fragment by digestion with these endonucleases. pPRVL was constructed by inserting this fragment into the HindIII site of pPrV1.7L, using an EcoR1/HindIII adaptor. Recombinant clones with the correct insert orientation of DNA inserts were selected by double-strand DNA sequencing. The resulting plasmid pPVRL thus contained a genomic fragment corresponding to nucleotides -1886 to +77 of the human villin gene, upstream of the firefly luciferase gene and SV40 polyadenylation elements. All junctions derived from cloning steps were checked by double-stranded DNA sequencing using synthetic oligonucleotides.

S1 Mapping—The HindIII-HindIII restriction fragment from BSSH0.9 was subcloned into a bacteriophage M13 mp18 vector. A 380-bp single-stranded DNA probe spanning the 5'-end of the villin mRNA was generated by incorporation of [32P]dCTP in the synthesis of the noncoding strand of this genomic restriction fragment from an oligonucleotide primer located in the first exon of the gene. An aliquot of the gel-purified probe (10,000 cpm) was hybridized at 48°C for 16 h in the presence of 10 µg of total RNA from either HeLa cells or HT29-differentiated cells, 80% formamide, 40 mM Pipes, 400 mM NaCl, 1 mM EDTA; pH was adjusted to 6.4. Heteroduplex hybrids were then digested for 1 h with S1-nuclease (Pharmacia). Hybridization controls were performed on the same EcoR1/HindIII DNA fragment subcloned into a bacteriophage M13 mp18 vector. Oligonucleotide primers (Pharmacia), respectively, internal synthetic oligonucleotides were used to complete the sequence on the coding strand. The sequence of the opposite strand of this 1.7-kb fragment was determined using a series of synthetic oligonucleotides. The sequence of the 0.3-kb insert BSSH0.3 (described above) was confirmed for both strands using M13 universal and reverse primers. Oligonucleotides were purchased from an LKB-Pharmacia synthesizer.

Sequence Analysis—The villin gene was analyzed for the presence of sequence homologies to known response elements for transcription factors (Locke and Buzard, 1990) using the DNAin program (Pharmacia).

Sequences of the various promoters controlling genes expressed in enterocytes were obtained from the EMBL/GenBank data base. A T-matrix program (Staden, 1977) was used to identify conserved bases among the various promoters. Optimal alignments of nucleotide sequences were generated using the NUCALN program (Wilbur and Lipman, 1983).
Cell Culture and Transfections—Human CaCO2 and HT29 adenocarcinoma-derived cell lines (Fogh and Trepke, 1975) were maintained as described (Pinto et al., 1982, 1983). Other cell lines used were the human HepG2 hepatoma cell line (Aden et al., 1979), the pig LLCPK1 renal cell line derived from proximal tubule of the kidney (Rabito et al., 1984), and the SKP keratinocyte line derived from a penile squamous carcinoma.

Cells were transfected using a lipofection reagent (Bethesda Research Laboratories) to induce the formation of lipid-DNA complexes (Felgner et al., 1987). Cells were grown until they are approximately 50% confluent. Two hours before the transfection, culture medium was changed. The procedure for transient transfection was performed as described by the supplier except that culture medium was also changed once more 6 h after transfection.

Establishment of a Stable Cell Line—For the stable transformation, CaCO2, HT29, and SKP cells were grown to approximately 50% confluency before transformation by procedure identical to that used for transient expression. Cells were cotransfected with pSV2neo and pSVOAL or pSV2neo and pPrVL before plating at low density.

Selection of stable transfectants was accomplished by the addition of 0.7 mg/ml of G418 to the culture medium. CaCO2 and SKP cell colonies formed at 3 days whereas resistant HT29 cell colonies were isolated and amplified separately.

Luciferase and β-Galactosidase Assays—Luciferase assays were performed as described by Nguyen et al. (1988) with slight modifications. Briefly, epithelial cells were scraped (48 h after transfection for transient expression experiments with luciferase vectors), into ice-cold phosphate-buffered saline. Transfected cells were pelleted by centrifugation in a microcentrifuge for 5 min at 4 °C and lysed by three repeated freezings in dry ice/ethanol and thawing at 37 °C. Pellets were resuspended in 100 μl of LB extraction buffer (25 mM Tris-phosphate, pH 7.8; 8 mM MgCl2; 1 mM dithiothreitol; 1 mM EDTA; 1% Triton X-100; 1% bovine serum albumin; 15% glycerol). An aliquot (10 μl) was analyzed using a Lumat LB 9501 luminometer (Berthold) which was set to inject 200 μl of LB buffer containing 5 mM ATP and 0.01 mM Luciferin (Sigma).

The amplification protocol employed 30-40 cycles (depending on the primer extension experiments were unsuccessful. The absence of CAAT or TATA boxes in housekeeping genes supports the contention that this cytosine residue does represent the actual transcription start site.

We found a unique putative activator protein-1 (AP-1) response element (Angel et al., 1987; Curran and Franza, 1988) at position −1105 and six putative activator protein-2 (AP-2) half-response elements (Mitchell et al., 1987; Imagawa et al., 1987) spanning positions −1702 to +3. A chicken lysozyme silencer 2 element (Banashmad et al., 1987) is located at nucleotide −894. A glucocorticoid response element (glucocorticoid response element-like sequence GGTCCACTCGTTC, Slater et al., 1985; Gloss et al., 1987) is located at nucleotide −253. No evidence of direct glucocorticoid-mediated regulation of villin gene activity have been demonstrated in cultured cells. However, as is the case with the majority of epithelial cells, in vivo terminal differentiation of the intestine mucosa might depend on glucocorticoid stimulation. These steroid hormone response elements could be activated only at specific stages of development and/or in

RESULTS

Structure of the 5'-Flanking Region of the Human Villin Gene—To isolate 5'-flanking sequences from the villin gene, we started with the previously isolated EMBL3-85e recombinant bacteriophage which has been used to elucidate the genomic structure of the human villin gene (Pringault et al., 1991). A 2-kb fragment immediately 5' of the first exon was subcloned (see “Materials and Methods”), and its nucleotide sequence was established (Fig. 1).

In order to map the initiation site for villin transcription, an antisense genomic probe encompassing 300 bp upstream from the ATG translational start site of the villin gene was hybridized with total RNA from the HT29 intestinal cell line. Resultant heteroduplexes were digested with S1 nuclease, and protected fragments were analyzed on an acrylamide gel (Fig. 2). Control for protection specificity was carried out with total RNA isolated from villin-negative HeLa cells. The length of the protected fragment was determined by comparison with a sequence ladder from the same genomic fragment and the original primer used to synthesize the probe. The single product extended 20 nucleotides upstream of the translation initiation site (to the bold underlined cytosine residue in the sequence CCGTCTCCCCAGGCCTACCATGACC). Overexposure of the gel did not reveal any additional protected fragments (data not shown). Thus, the S1 nuclease mapping method predicts that the transcription of the two human villin mRNAs initiates from a unique site (a cytosine residue subsequently designated as nucleotide +1) located 21 nucleotides upstream from the ATG codon. However, one cannot exclude the possibility that an intron longer than 300 bp, the length of the genomic probe used in this experiment, was located just upstream from this putative site and could be responsible for the length of the observed protected fragment. As previously reported in the cloning of the 5'-end of the villin cDNA (Arpin et al., 1988) and for another actin-binding protein, the gelsolin gene (Kwiatowski et al., 1988), primer extension experiments were unsuccessful. The absence of an AG splice acceptor site just upstream from this cytosine residue, a canonic feature which was shown to be also the rule in all introns on the villin gene (Pringault et al., 1991), further supports the contention that this cytosine residue does represent the actual transcription start site.

The sequence of the 5'-flanking region (Fig. 1) did not reveal the presence of identifiable promoter elements typical of RNA-polymerase II-transcribed genes (CAAT and TATA elements) at the usual positions (Bucher and Trifonov, 1986). Potential recognition sequences for SP1 transcription factor (GGGGCGG boxes), whose presence is often correlated with the absence of CAAT or TATA boxes in housekeeping genes (Kadonaga et al., 1987), were not found either.

We sequenced a 2-kb fragment immediately upstream from the first ATG codon in the HT29 cell line. The sequence of the 5' flanking region (Fig. 1) did not reveal the presence of identifiable promoter elements typical of RNA-polymerase II-transcribed genes (CAAT and TATA elements) at the usual positions (Bucher and Trifonov, 1986). Potential recognition sequences for SP1 transcription factor (GGGGCGG boxes), whose presence is often correlated with the absence of CAAT or TATA boxes in housekeeping genes (Kadonaga et al., 1987), were not found either.

We sequenced a 2-kb fragment immediately upstream from the first ATG codon in the HT29 cell line. The sequence of the 5' flanking region (Fig. 1) did not reveal the presence of identifiable promoter elements typical of RNA-polymerase II-transcribed genes (CAAT and TATA elements) at the usual positions (Bucher and Trifonov, 1986). Potential recognition sequences for SP1 transcription factor (GGGGCGG boxes), whose presence is often correlated with the absence of CAAT or TATA boxes in housekeeping genes (Kadonaga et al., 1987), were not found either.

We sequenced a 2-kb fragment immediately upstream from the first ATG codon in the HT29 cell line. The sequence of the 5' flanking region (Fig. 1) did not reveal the presence of identifiable promoter elements typical of RNA-polymerase II-transcribed genes (CAAT and TATA elements) at the usual positions (Bucher and Trifonov, 1986). Potential recognition sequences for SP1 transcription factor (GGGGCGG boxes), whose presence is often correlated with the absence of CAAT or TATA boxes in housekeeping genes (Kadonaga et al., 1987), were not found either.
cooperation with other transcription factor response elements such that a direct effect of glucocorticoids on villin expression may not be apparent in cultured cells.

A search for liver-associated elements revealed the presence of four half-HNF4/LFA-1 putative binding sites (Hardon et al., 1988) distributed between nucleotides −253 and −1867. Two of them are associated with an AP-2 element. Two hexanucleotides corresponding to half-HNF4/LFA-1 response elements (Ott et al., 1984; Hardon et al., 1988, Cerighini et al., 1988) were found at positions −123 (associated with an AP-2 element) and −1883. Although currently described as liver-associated transcription factors, HNF4/LFA-1 and HNF4/LFA-1 are also expressed in intestinal mucosa (Baumhueter et al., 1990; Sladek et al., 1990). The HNF4/LFA-1 and HNF4/LFA-1 mRNA pattern of expression in mammalian tissues (review in De Simone and Cortese, 1991) is closely related to that observed for villin. However, intestinal genes regulated by these factors have not been identified. The overall distribution of these putative response elements along the 2-kb 5′-flanking sequence of the villin gene is depicted in Fig. 4.

A computer-assisted analysis of the sequence revealed the presence of two sets of Alu sequences located between nucleotides −344/−634 and nucleotides −1474/−1777, respectively (Figs. 1–4). We also found three pairs of tandemly repeated short sequences: AGTT/(A/G)ACCCCCC, CCTGATGGGAG (A/G)AG (A/G)CCA, and AAAGCAACCATG (Fig. 4). An hexanucleotide CCCA(AG/T)G repeated five times (two tandems and a single) (Fig. 1). We calculated that the chance occurrence of such a 6-bp sequence element in a 2000-bp nucleotide fragment is close to 0.5.

The 5′-flanking sequence was then compared to several other known promoters of tissue-specific genes expressed in enterocytes: human intestinal alkaline phosphatase (Millan, 1987; Henthorn et al., 1988), human apolipoprotein A4 (Elshourbagy et al., 1987), pig neutral aminopeptidase (Olsen et al., 1989), human fatty acid-binding protein (Sweetser et al., 1987), and sucrase-isomaltase (Traber et al., 1992). Repeated sequences identified in the 5′-flanking region of the villin gene are not found in the sequences of the above promoters, except a short hexanucleotide CCCAA/TG which is present twice in the promoter of the fatty acid binding protein gene. A closely related sequence CCCAG was previously identified (a tandem repeat and three single elements) in the promoter of the intestinal alkaline phosphatase gene, in addition to the two copies present in the first intron. T-matrix plots and optimal alignments of the 5′-flanking region of the villin gene compared with the promoters of the intestinal genes listed above have allowed us to identify 11 additional short conserved sequences in the human intestinal alkaline phosphatase, human apolipoprotein A4, pig neutral aminopeptidase, and human fatty acid-binding protein promoters, as shown in Fig. 3, but not in the sucrase-isomaltase promoter. These tight homologies seem to be highly specific; except for the element 1 (see Fig. 3) located at the positions −83 and −247 (inverted repeat) in the villin gene, these sequences were not found in any other genes in the EMBL/GenBank data base. Although it was found in numerous nonintestinal genes, element 1 could be of particular interest since it corresponds to a sequence identified in the human apolipoprotein A4 as protected from DNase I digestion by nuclear extracts from the CaCO2 intestinal cell line.3

Fig. 4 shows the overall distribution in the 5′-flanking region of the villin gene of the sequences which are conserved among other active intestinal promoters. We observed a concentration of these conserved sequences in the 300 nucleotides region immediately 5′ of the villin transcription initiation site. The occurrence of conserved sequences is maximal in the promoter of the human intestinal alkaline phosphatase gene. These sequences found in promoters of both the intestinal alkaline phosphatase and the villin genes were, indeed, searched for in the promoter of the liver/kidney/bone (L/K/B) alkaline phosphatase gene (Weiss et al., 1988). The first conserved sequence (′ in Fig. 3) was also identified in the L/K/B alkaline phosphatase promoters, showing a lack of intestinal specificity. However, the other conserved sequences (2–11 in Fig. 3), the hexanucleotide CCCAG repeated in the

11429

3 S. Bovard, Pasteur Institut, France, personal communication.
Villin Promoter Characterization

**Fig. 2. Determination of the initiation of transcription by S1 nuclease mapping.** A 380-bp single-stranded DNA probe spanning the 5'-end of the villin mRNA was generated by synthesizing the noncoding strand of a genomic restriction fragment subcloned in the bacteriophage M13 mp18, starting from an oligonucleotide primer located in the first exon of the gene. This probe was labeled by incorporating [32P]dCTP during synthesis. The probe was hybridized with total RNA (10 μg) from either HeLa cells (A) or HT29 differentiated cells (B). The hetero-duplex hybrids were then digested by S1-nuclease at concentrations ranging from 0 to 1000 units/ml. Protected fragments were analyzed on 8% polyacrylamide gel containing 7 M urea, in parallel with an enzymatic sequence reaction (on the left of the panel). Nucleotide sequence of the region comigrating with the protected fragment is shown. The dashed box indicates the putative transcription initiation site. The open box points out the ATG initiation codon.

intestinal alkaline phosphatase promoter and the CCCAAG element repeated in the 5'-flanking region of the villin gene, were missing in the L/K/B alkaline phosphatase promoters. Such comparison between regulatory regions of tissue-specific isoforms of alkaline phosphatase indicates the likelihood of the intestinal specificity of these sequences.

The binding sites for proteins from nuclear extracts have been described for the sucrase-isomaltase promoter (Traber et al., 1992) and for the aminopeptidase N promoter (Olsen et al., 1991). These elements have been searched in the 5'-flanking region of the villin gene. No conserved sites corresponding to the sequences SIF1, SIF2, and SIF3, identified in the sucrase-isomaltase promoter, nor to the sequences DF or UF described in the aminopeptidase N promoter, have been found in the villin gene (above 70% homology).

Due to the degeneracy of transcription factor binding sites, other previously described sites could be present in the 5'-flanking region of the villin gene. Foot-printing experiments, already underway using an appropriate panel of epithelial nuclear extracts should allow us to formally identify such sequences and determine if the conserved elements listed in Fig. 3 are actually utilized in specific intestinal gene regulation.

**Cell-specific Transcription of Luciferase Reporter Gene Driven by the 5'-Flanking Region of the Villin Gene**—To assess whether or not the 2-kb 5'-flanking region contains cis-acting promoter elements necessary for the controlled expression of the villin gene in a panel of epithelial cell lines, we constructed a chimeric gene containing this region linked to firefly luciferase cDNA (de Wet et al., 1987, see “Materials and Methods”), subsequently referred to as the villin/luciferase hybrid gene. The resulting plasmid pPrVL was introduced by lipofection into CaCO2, HepG2, LLCPK1, and SKP cells and transient expression of luciferase measured. Plasmid pCH110 containing the β-galactosidase reporter gene was used as a control to normalize for transfection efficiency between cell lines. A representative luciferase assay is shown in Fig. 5. In the villin-expressing human CaCO2 intestinal cell line, significant luciferase activity was observed. A similar result was obtained in the human HepG2 hepatoma cell line as well as in the LLCPK1 pig kidney proximal tubule cell line. In contrast, only a basal luciferase activity could be detected in the villin-negative SKP cell line, derived from human keratinocytes. Thus, the 2-kb 5'-flanking sequence appears to be transcriptionally competent and to direct expression of the luciferase gene only in the three villin-expressing cell lines, thereby conferring cell-specific transcription. The average luciferase activity observed in CaCO2, HepG2, and LLCPK1 cell lines was about 2% of that obtained when the luciferase was under the control of a strong viral promoter such as the LTR of Rous sarcoma virus (not shown).

Permanently transfected SKP, HT29, and CaCO2 cell lines were also established. For each cell line we derived two different clones, the first permanently transfected by pSVOAL and the second permanently transfected by pPrVL. The presence of the transgene was analyzed by genomic DNA amplification (polymerase chain reaction). We used a pair of primers allowing amplification of a 400-bp fragment located within the luciferase cDNA or a 850-bp fragment spanning the 5'-flanking region of the villin gene and the contiguous luciferase cDNA. Amplification of a 400-bp fragment, significant for the presence of the luciferase gene, was obtained in samples from...
Fig. 3. Conserved sequences in the villin 5′-flanking region and promoters from genes expressed in the intestine mucosa. Eleven short nucleotide sequences, conserved between the 5′-flanking sequence of the villin gene and one or several promoters of genes active in intestine mucosa, are shown. Nucleotide sequences of these promoters were obtained from the EMBL/GenBank data base and were individually compared to the sequence of the 5′-flanking region of the villin gene by a T-MATRIX program as well as by optimal alignment (NUCALN), using a gap penalty of +7, a window size of 20, and a k-tuple value of 3.

Outlined hVZL, 5′flanking region of the human villin gene; hALP, human intestinal alkaline phosphatase promoter (Millan et al., 1987); hApo A4, human apolipoprotein A4 (Elshourbagy et al., 1987); pAPn, porcine neutral aminopeptidase (Olsen et al., 1989); hFABP, human intestinal fatty acid binding protein (Sweetser et al., 1987). The dark stippled boxes indicate sequence identity. The position of the conserved elements in the genes was numbered with respect to the initiation site of transcription of each promoter.

c, complementary strand; i, inverted; a, a sequence identified in human apolipoprotein A4 which was protected from DNAse I digestion by nuclear extracts from the CaCO2 intestinal cell line.

all permanently transfected cell lines, whereas the amplification of a 850-bp fragment, significant for the presence of the hybrid villin/luciferase transgene, was obtained only in samples from pPrVL CaCO2 and pPrVL SKP cell lines (not shown). The genomic integration of the transgenes was also confirmed by Southern blot analysis (not shown). Expression of luciferase was analyzed by measurement of enzymatic activity in cell extracts. As shown on Table I, cell-specific transcription of the reporter gene was conserved in permanently transfected cell lines. As a control, luciferase activity was assayed in cells transfected with pSVOAL which contains no promoter. The chimeric gene was only expressed in CaCO2 and HT29 cell lines but not in the SKP cell line.

The pPrVL CaCO2 cell line has maintained a high degree of differentiation potential, since we observe dome formation after cell confluency and subsequent apical localization of sucrase-isomaltase (data not shown). These parameters demonstrate that this luciferase-expressing cell line is able to establish and maintain a polarized phenotype characteristic of the original cell line.

Controlled Expression of the Reporter Gene during Terminal Differentiation of the HT29 Intestinal Cell Line—We have previously shown that villin gene activity is dramatically enhanced when intestinal cells differentiate (Robine et al., 1985; Pringault et al., 1986; Dudouet et al., 1987; Boller et al., 1988). In order to assess the presence of regulatory elements in the 5′-flanking region of the villin gene, we used the HT29 cell line which had been permanently transfected with the villin/luciferase hybrid gene. Terminal differentiation of HT29 cells can be induced by replacing glucose with galactose in the culture medium (Pinto et al., 1982). When cultured in glucose-containing medium, HT29 cells form a multilayer of
Villin Promoter Characterization

**FIG. 4.** Overall structure of the 5'-flanking region of the human villin gene. The distribution along the 5'-flanking region of the human villin gene of putative response elements for known transcription factors and of the elements conserved between several promoters active in the intestinal mucosa (described in detail in Table I and Fig. 3) is shown. The thick horizontal black line represents the 5'-flanking region of the human villin gene. *Stippled boxes* represent two repeated human Alu sequences. The black box represents the first exon of the human villin gene. *CAP*, initiation site of the transcription. *AP1*, short sequence homologous to the response element for AP1 transcription factor. *AP2*, short sequence homologous to the response element for AP2 transcription factor. *GRE*, short sequence homologous to the glucocorticoid response element. *LYS-SIL2*, short sequence homologous to the lysozyme silencer element 2. *LFA-1*, short sequence homologous to the response element for HNF4/LFA-1 liver transcription factor. *HNF-1*, short sequence homologous to the response element for HNF-1/LFB-1 liver transcription factor. *h iALP*, the position of a short sequence conserved in the human intestinal alkaline phosphatase promoter (Millan et al., 1987); *h Apo A4*, the position of a short sequence conserved in the human apolipoprotein A4 (Elshourbagy et al., 1987); *p APn*, the position of a short sequence conserved in the porcine neutral aminopeptidase (Olsen et al., 1989); *h FABP*, the position of a short sequence conserved in the human intestinal fatty acid binding protein (Sweetser et al., 1987). Numbers in *open boxes* refer to the conserved elements described in Fig. 3. The scale is indicated in base pairs (bp).

**FIG. 5.** Transient expression of luciferase, under the control of the 5'-flanking region of the human villin gene, in a panel of epithelial cell lines. Enzymatic activity of luciferase measured in various cell lines transfected by the pPrVL plasmid containing the villin/luciferase hybrid gene, is standardized to the efficiency of transfection. A typical experiment is shown. Variation of luciferase activity did not exceed 20% in repeated experiments (*n = 5*). *SKP*, human keratinocyte cell line; *CaCO2*, human intestinal cell line; *HepG2*, human hepatocytes cell line; *LLCPK1*, pig kidney cell line (derived from proximal tubule).

**TABLE I**

Luciferase activities in permanently transfected CaCO2, HT29, and SKP cell lines

Luciferase activity is reported as arbitrary units (light emission per 10⁶ cells). A typical experiment is shown. Variation of luciferase activity did not exceed 20% in repeated experiments (*n > 5*).

| Constructs | Cell lines | HT29 | SKP |
|------------|------------|------|-----|
| pSVOAL     | CaCO2      | 94   | 98  |
| pPrVL      | CaCO2      | 12840| 2850| 102 |

*HT29 clone 8, undifferentiated cells cultured in glucose-containing medium.*

**TABLE II**

Luciferase expression in three HT29 cell clones permanently transfected with the pPrVL plasmid containing the villin/luciferase hybrid gene

A typical experiment is shown. Variation of luciferase activity did not exceed 20% in repeated experiments (*n > 5*). Glu¹, glucose-containing medium, undifferentiated growth conditions; Gal², galactose-containing medium, differentiated growth conditions. Glucose and galactose were added to the growth medium at a final concentration of 10 mM.

| HT29 cell clones | Luciferase activity (arbitrary units) | Per mg of protein | Per 10⁶ cells |
|-----------------|--------------------------------------|------------------|--------------|
| 1 Glu¹          | 12.8                                 | 3200             |
| 2 Glu¹          | 84.1                                 | 2335             |
| 3 Glu¹          | 24.9                                 | 6225             |
| 4 Glu¹          | 129.7                                | 4595             |
| 5 Glu¹          | 11.4                                 | 2850             |
| 6 Glu¹          | 70.1                                 | 24538            |

*Fold stimulation of luciferase activity.

unpolaredized cells and synthesize only small amounts of villin. In contrast, when differentiation is induced by growth in galactose they form a monolayer of polarized enterocyte-like cells with tight junctions and a 5–10-fold increased expression of villin protein and mRNAs (Dudouet et al., 1987; Pringault et al., 1986). Three independently isolated HT29 cell clones (numbered 1, 3, and 8) containing the villin/luciferase hybrid gene and expressing luciferase were first grown in glucose-containing medium (undifferentiated growth conditions, see Table I), and then were grown in galactose. When cultured in galactose-containing medium these three HT29 cell clones reproduced the cytodifferentiation characteristic of the parental HT29–18 cell line. After induction of differentiation, the levels of luciferase activity were measured in each of these three clones and compared to those observed when cultured in glucose-containing medium (Table II). Upon differentiation, we noted a 5–6-fold increase in luciferase activity per mg of protein when the HT29 cells differentiate. This is
correlated to a 7-9-fold increase in luciferase activity when standardized per 10^6 cells. This difference parallels the observed increase in protein content when HT29 cells undergo differentiation (data not shown).

Thus, the isolated 5'-flanking region of the villin gene contains an enhancing activity which is able to respond positively to nutritional conditions known to stimulate terminal differentiation of these cultured epithelial intestinal cells.

**DISCUSSION**

**The Villin Promoter as a Model of Molecular Genetic Regulation during the Histogenesis of the Digestive Tract**—Attempts to characterize tissue-specific promoters of genes expressed in the digestive tract have focused on liver-specific genes. This approach has considerably improved our knowledge of the molecular regulation which leads to tissue-specific, developmental, and hormonal control of liver-specific gene expression and establishment of the hepatocyte phenotype. In contrast, the molecular genetics of intestinal differentiation are not as well documented. As a dogma, widely accepted as a model of tissue-specific regulation of this gene (Mieusset-Snyder et al., 1992).

Such complex cooperation or competition between response elements and trans-acting factors would account for the alternative regulation of the villin gene during differentiation of digestive epithelial cells. Elucidation of cis-acting regulatory elements of the villin gene, which is tightly regulated during the development of the digestive tract, would allow the characterization of transcription factors involved in the differentiation of the intestinal cell lineage. This goal should be facilitated by our documentation of the complete profile of villin expression during mouse embryogenesis (Maunoury et al., 1988; Maunoury et al., 1992).

**Gene Targeting Using the Villin Promoter**—In addition to these long-term molecular genetic goals, the villin promoter could provide a molecular tool for targeting heterologous genes into a selected population of cells of the digestive tract. Such techniques could also allow one to disrupt or create a function, or elicit the expression of a gene normally silent in the targeted cell. In particular, this would allow oncogene targeting in transgenic mice for induction of tissue-specific tumors and establishment of new cell lines (Hanahan, 1988).

Transfections of cultured intestinal cell lines have shown that the 5'-flanking region of the villin gene contains at least some of the regulatory elements responsible for villin expression in immature enterocyte precursors and for the specific up-regulation observed upon terminal differentiation. These regulatory elements might be used to drive the expression of heterologous genes in immature cells of the definitive endodermal lineage and in the proliferative precursors of the adult intestinal crypts as well as in differentiated cells from the small and large intestine mucosa (Pringault, 1990). However, we must first demonstrate that the cell-specific promoting activity of the 2-kb 5'-flanking region of the villin gene described in this study is functional in transgenic mice harboring the same construct. The luciferase reporter gene under the control of this 2-kb genomic fragment has been introduced in the germ line of mice, and the offspring of transgenic animals are currently being analyzed to clarify this point.

Murine models reproducing several steps of the progression of human colorectal cancers could then be obtained by targeting the associated oncogenes or mutated tumor suppressor genes to colonocytes. Moreover, new cell lines derived from the digestive tract could be established by targeting the SV40 T antigen to the precursors of intestinal cells as well as to biliary duct cells of the liver and ductal cells of the exocrine pancreas. **Acknowledgments**—We thank Dr. F. Breitburd and Prof. G. Orth for the gift of the SKP human keratinocyte line. We thank Prof. A. Walker, Dr. A. Israel, and Dr. G. Andreason for critically reading the manuscript, and Dr. D. Sourdie for help in computer-assisted analysis of sequences.

**REFERENCES**

Aden, D. P., Fogel, A., Plotkin, S., Dananjovan, I., and Knowles, B. B. (1979) *Nature* 282, 615-616

Angel, R., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Angel, P., Imagawa, M., Chiu, R., and Karin, M. (1988) *Cell* 47, 729-739

Arpin, M., Pringault, E., Finidori, J., Garcia, A., Jeltsch, J.-M., Vandekerckhove, J., and Loward, D. (1988) *J. Biol. Chem.* 263, 1759-1766

Bachiller, A., Muller, M., Steiner, Ch., and Renwitz, R. (1988) *EMBO J.* 7, 2297-2303

Baumhauer, S., Measurement, D. B., Conley, P. B., Kuo, C. J., Turk, C., Graves, M. L., Rohde, C. A., Courtois, G., and Crabbe, G. R. (1990) *Genes Dev.* 4, 372-379

Boller, K., Arpin, M., Pringault, E., Mangeat, P., and Reggio, H. (1988) *Nature* 339, 57-61

Bretacter, A., and Weber, K. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 2321-2325

Buchler, P., and Trifonov, E. N. (1986) *Nucleic Acids Res.* 14, 10009-10025

Carbon, J. M., Howe, C. L., West, A. B., Barbick, W. R., Mooseker, M. S., and Fish, H. I. (1987) *Am. J. Pathol.* 129, 583-600

Carette, M., Niquet, L., Tulliez, M., Lepetit, N., Levetr, F., Grimber, G., Brand, P., and Kahn, A. (1992) *Endocrinology* 14, 113-123

Coppin, J., Blumenfeld, M., and Yaniv, M. (1987) *Cell Genes & Dev.* 2, 967-974

Cogen, P., Lynch, C. Y., Vaulont, S., Kahn, A., and Marie, J. (1987) *J. Mol. Biol.* 196, 11-25

Curran, T., and Franza, B. R. Jr. (1988) *Cell* 55, 395-397

Daleman, W., Ferraud, F., LeMarce, G., Gerlinger, P., Courtcnay, M., and Pavirani, A. (1990) *Biologicals* 18, 191-198

de Simone, V., and Cortese, R. (1991) *Curr. Opin. Cell Biol.* 3, 960-965

de Wet, J. R., Wood, R. V., DeLucca, M., Helmink, D. R., Subramani, S. (1987) *Mol. Cell. Biol.* 7, 725-737

Dubois, N., Benuonn, M., Allemand, I., Molina, T., Grimber, G., Daude-Monsac, M., Abeallan, K., and Brand, P. B. (1991) *J. Hepatol.* 10, 227-239

Edelstein, C., Robin, R., Huang, M., and Yaniv, M. (1988) *Genes Dev.* 2, 101-119

Elshourbagy, N. A., Walker, D. W., Pak, Y.-K., Boguski, M. S., de Wet, D. A., Fogel, A., Plotkin, S., Subramani, S. (1987)* Proc. Natl. Acad. Sci. U. S. A.* 84, 743-747

Fog, J., and Trempe, G. (1975) in *Human Tumor Cells In Vitro* (Fogh, J., ed) 115-116, Plenum Publishing Corp., New York

Friederich, M., Huet, C., Arpin, M., and Loward, D. (1988) *Cell* 59, 461-475

Friedrich, E., Vancompernelle, R., Huet, C., Goethals, M., Finidori, J., Vandekerckhove, J., and Loward, D. (1992) *Cell* 70, 81-92

Gloss, B., Bernard, H. U., Seedorf, K., and Klock, G. (1987) *EMBO J.* 6, 3735-3743

Grone, H.-J., Weber, K., Helmchen, U., and Osborn, M. (1986) *Am. J. Pathol.* 124, 294-302

Hanahan, D. (1985) *Nature* 315, 115-122

Hanahan, D. (1988) *Annu. Rev. Genet.* 22, 479-519

Hall, C., Jacob, E., Rignold, G., and Lee, F. (1983) *J. Mol. Cell. Appl. Genet.* 1, 101-109

Hardon, E. M., Frain, M., Paonessa, G., and Cortese, R. (1988) *EMBO J.* 7, 1711-1719

Haut, S. M., Kim, S. H., Schmidt, G. H., Pease, S., Rees, S., Harris, S., Roth, R. A., Randall, Whangbough, J., Colacino, S. M., Ahnen, D. J., Wright, N. A., Gogg, B. L. A., and Gorden, M. C. (1990) *Cell* 59, 239-239

Hespert, P. S., Raducha, M., Kadesch, T., Weiss, M. J., and Harris, H. (1988) *J. Biol. Chem.* 263, 1201-1210

Herbolom, B., Bourchet, B., and Vaniy, M. (1984) *Cell* 39, 653-662

Hue, C., Sahutugqillo-Merino, C., Coudierre, D., and Loward, D. (1987) *J. Cell. Biol.* 105, 345-357

Imagawa, M., Chiu, R., and Karin, M. (1987) *Cell* 47, 251-260

Kadonas, T., Erkser, K. R., Massier, F. R., and Tjoss (1990) *Cell* 51, 445-457

Kwiatkowski, D. J., Mifl, R., and Yin, H. L. (1988) *J. Cell. Biol.* 106, 375-384

Loker, J., and Buzard, G. (1990) *J. DNA Sequencing Mapping* 1, 3-1

Maunoury, R., Robin, S., Pringault, E., Huet, C., Coudrier, J. L., Gaillard, J. A., et al.
