MOLECULAR AND CELLULAR BIOLOGY, Aug. 1992, p. 3614–3627
0270-7306/92/083614-14$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 12, No. 8

Novel DNA-Binding Proteins Regulate Intestine-Specific Transcription of the Sucrase-Isomaltase Gene

PETER G. TRABER,* GARY D. WU, AND WEI WANG

Division of Gastroenterology, Department of Internal Medicine, The University of Michigan Medical School, and the Veterans Administration Medical Center, Ann Arbor, Michigan 48109

Received 17 January 1992/Accepted 18 May 1992

Sucrase-isomaltase (SI) is an enterocyte-specific gene which exhibits a complex pattern of expression during intestinal development and in the adult intestinal mucosa. In the studies described in this report, we demonstrate that enterocyte-specific transcription of the SI gene is regulated by an evolutionarily conserved promoter that extends approximately 180 bp upstream of the transcription start site. DNase I footprint analysis allowed the identification of three nuclear protein-binding sites within the SI promoter (SIF1, SIF2, and SIF3 [SI footprint]), each of which acted as a positive regulatory element for transcription in intestinal cell lines. SIF1 was shown to bind nuclear protein complexes present in primary mouse small intestinal cells and in an intestinal cell line (Caco-2). However, SIF1-binding proteins were absent in a variety of other epithelial and nonepithelial cells. In vitro mutagenesis experiments demonstrated that the SIF1 site is required for high-level promoter activity in intestinal cells. The SIF3 element formed prominent binding complexes with intestinal and liver nuclear extracts, whereas nuclear proteins from other epithelial and nonepithelial cells formed weaker complexes of different mobilities. The SIF2 element bound nuclear proteins in a pattern similar to that of SIF3, and cross-competition studies suggested that SIF2 and SIF3 may bind the same nuclear proteins. Taken together, these data have allowed the identification of novel DNA-binding proteins that play an important role in regulating intestine-specific transcription of the SI gene.

The epithelium of the mammalian intestinal tract is composed of four cell types that arise from stem cells located in crypts (16, 22, 42). Absorptive enterocytes, mucus-producing goblet cells, and enteroendocrine cells migrate from crypts onto villi and are eventually shed into the intestinal lumen (16, 22, 42). The fourth cell type, Paneth cells, are longer lived and migrate to the base of crypts (14). Cell division in crypts and migration onto the villus is a dynamic process resulting in renewal of the epithelium every 3 days in rodents and every 4 to 5 days in humans (22). The major cell type responsible for intestinal function is the absorptive enterocyte, which constitutes 85 to 90% of the cells in crypts and greater than 95% of the cells located on villi (15). The process of intestinal cell turnover and differentiation involves a complex series of changes in gene expression, reflecting both lineage determination and differentiation. Thus, for example, in crypts, enterocyte precursor cells are able to proliferate and are relatively undifferentiated. However, upon migration onto the villus, they lose their capacity to divide and express genes associated with a functional, differentiated phenotype. Relatively little is currently understood about the molecular mechanisms that regulate the processes of cell lineage determination within the epithelium and conversion of proliferating, undifferentiated crypt cells into nonproliferating, differentiated villus enterocytes. With transgenic mice, it has been shown that *cis*-acting elements present in fatty acid-binding protein genes are responsible for directing intestine-specific transcription as well as developmental and spatial patterns of expression (22, 49, 51). However, the *cis*-acting elements and their cognate transcriptional proteins have not been characterized.

Sucrase-isomaltase (SI) is an intestinal brush border disaccharidase that can serve as a model for the study of the regulation of intestinal gene expression as well as intestinal differentiation and development. SI is found only in small intestinal enterocytes (31, 35, 52, 53) and has a complex pattern of developmental (26) and spatial (31, 52, 53) expression within the intestine. Furthermore, its expression is altered in response to dietary changes (12, 13, 30) and diabetes mellitus (30). Finally, SI is expressed in the majority of colonic adenomas and adenocarcinomas, a process which may reflect a reversion to a fetal colonocyte phenotype (6, 56, 57). In previous studies, we have demonstrated that expression of SI is regulated along the crypt-villus axis at the level of mRNA accumulation in both humans (53) and rats (52). By in situ hybridization, SI mRNA was undetectable in crypt cells but appeared in cells located at the crypt-villus junction. SI mRNA was most abundant in lower and mid-villus cells, with low levels in villus tip cells. Therefore, there is a complex pattern of expression of SI mRNA as cells migrate from crypts onto villi and are finally extruded from villus tips into the lumen.

We previously reported the isolation and characterization of the human SI gene and demonstrated that the 5′-flanking region was able to direct intestinal cell line-specific transcription (59). Although 3,424 bases of the 5′-flanking region were required for high-level expression, it appeared that tissue specificity was conferred by elements within 324 bases of the start of transcription (59). In the studies described in this report, we have attempted to elucidate the molecular mechanisms underlying tissue-specific expression of the SI gene. Using transient transfections, we show that enterocyte-specific transcription of the SI gene is controlled by a 183-bp evolutionarily conserved promoter located immediately upstream of the transcriptional start site. This promoter contains at least three nuclear protein-binding sites (SIF1, SIF2, and SIF3, where SIF indicates an SI footprint), each of which acts as a positive *cis*-acting element for SI transcription. The SIF1 element binds intestine-specific nu-

* Corresponding author.

clear protein complexes, and, by using in vitro mutagenesis, we have found that SIF1 is required for transcriptional activity of the SI promoter in intestinal cells. The SIF3 element formed prominent binding complexes with intestinal and liver nuclear extracts and weaker complexes with nuclear extracts from a variety of other cell lines. The SIF2 element bound nuclear proteins in a pattern similar to that of SIF3, and cross-competition studies suggested that SIF2 and SIF3 may bind the same nuclear proteins. Taken together, these studies identify the SI promoter and its cognate nuclear proteins as a useful system for future studies of developmentally regulated and tissue-specific gene expression in the intestine.

MATERIALS AND METHODS

Screening of mouse genomic library. A mouse genomic library, constructed by Clontech Laboratories (Palo Alto, Calif.), was plated on 150-mm plates on a lawn of LE392 bacteria at a density of 10,000 to 15,000 plaques per plate. Lifts were taken with nitrocellulose disks (Stratagene, La Jolla, Calif.), and the filters were prehybridized in a solution containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 5× Denhardt's solution, 100 μg of salmon sperm DNA per ml, and 0.5% sodium dodecyl sulfate (SDS) at 65°C for 3 h and then hybridized for 16 h in the same solution containing 1×10^6 cpm of a ^{32}P-labeled mouse cDNA (described in Results) per ml. The filters were washed successively in 2× SSC–0.5% SDS, 1× SSC–0.5% SDS, and 0.5× SSC at 60°C for 30 min each. Positive plaques were rescreened under the same conditions on 100-mm plates.

Phage DNA was purified from 50 ml of liquid lysates with the Qiagen lambda preparation system (Qiagen, Studio City, Calif.). Insert DNA was subcloned into pBluescript KS− (Stratagene), and ordered deletions were prepared by exonuclease III digestion as described before (59). Oligonucleotides were synthesized to sequence regions of the gene that were not accessible in restriction-generated or deletional subclones. Double-stranded plasmids were sequenced with Sequenase 2.0 (U.S. Biochemical, Cleveland, Ohio).

Primer extension. Total RNA was incubated with a ^{32}P-labeled oligonucleotide, and the annealed primer was extended for 1 h at 37°C with Moloney murine leukemia virus reverse transcriptase as described previously (59). RNA was hydrolyzed in 0.1 M NaOH, and the DNA was precipitated with ethanol and separated in an 8% polyacrylamide–7 M urea gel. Controls included labeled primer to assess self-annealing and yeast tRNA and *Escherichia coli* total RNA to assess nonspecific binding.

Anchored PCR cloning of mouse SI cDNA. cDNA was synthesized with Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) and the reaction conditions described for primer extension. Excess nucleotides and primer were removed from the reactions with Centricon-100 microconcentrator tubes (Amicon), and guanosines were added to the 3′ end of the cDNA by using terminal deoxynucleotidyl transferase as described previously (59). The reaction mixture was diluted 1:10, and 2 ml (containing 50 μM each dATP, dCTP, dGTP, and TTP; 50 mM KCl; 0.01% gelatin; 1.5 mM $MgCl_2$; 0.2 μM gene-specific primer; 0.2 μM poly(C) adapter primer; and 2.5 U of *Taq* polymerase [Bethesda Research Laboratories]) was used in a 100-ml polymerase chain reaction (PCR). The sequence of the oligonucleotide used to prime the reverse transcription reaction was the same as that used previously to clone the rat SI cDNA (52). The gene-specific primer for the PCR was designed from the rat SI cDNA sequence: RSI-6, 5′-GTGGAATTCATAGCCATGGTTGTCTGCAAA GAAG-3′ (*Eco*RI site underlined). The poly(C) adapter primer was the same as used previously (59). The reaction mixture was denatured at 94°C for 1 min, annealed at 50°C for 2 min, and polymerized at 72°C for 3 min. Thirty-five cycles were performed, followed by a 10-min extension at 72°C. The PCR product was digested with *Bam*HI and *Eco*RI, subcloned into pBluescript KS−, and sequenced.

Construction of luciferase plasmids. Luciferase (LUC) vectors included pA3-LUC (promoterless) and pRSV-LUC (Rous sarcoma virus promoter and enhancer) (37, 58) (gifts from Dr. Wood, University of Colorado). Portions of the SI gene promoter were inserted into pA3-LUC by PCR as described previously (59). The 3′ end of the SI promoter fragments was located at base +54 as defined by the oligonucleotide HS-4X (see below). In this article, the deletional constructs are named for the 5′-most base in the SI gene, i.e., −183 to +54 human SI-LUC = Δ183 hSI-LUC.

The following primers were used for amplification (the underlined site is identified in parentheses):

−183 hSI: 5′-GTGGGTACCTGACAGTACAATTACTAA-3′ (*Kpn*I site)
−156 hSI: 5′-GTGGGTACCATTCCCAGAGAGAAACAT-3′ (*Kpn*I site)
−126 hSI: 5′-GTGGGTACCCTTTCAGGGTTAAGGCTT-3′ (*Kpn*I site)
−94 hSI: 5′-GTGGGTACCTTGAGCAGAAGATTATTA-3′ (*Kpn*I site)
−66 hSI: 5′-GTGGGTACCGGCTGGTGAGGGTGCAAT-3′ (*Kpn*I site)
−33 hSI: 5′-GTGGGTACCAGGTCAATATATACCTAA-3′ (*Kpn*I site)
HS-4X: 5′-GTGAAGCTTAGCCTGTTCTCTTTGCTA-3′ (*Hin*dIII site)

PCR products were digested with *Hin*dIII and *Kpn*I and subcloned into pA3-LUC. Before use in transfection experiments, each luciferase construct was digested with *Bgl*II and *Bam*HI to check for rearrangement of the stop signal cassette (37, 58), and the SI portion was sequenced in its entirety to check for misincorporation of nucleotides.

Site-directed mutagenesis. Mutations within the 5′-flanking sequence of the human SI gene were made by PCR as described before (28). This procedure was used to generate a two-base mutation at bases −50 and −51 of the human SI gene within a −183 to +54 SI gene fragment. The primers used included −183 hSI, HS-4X, and the following two mutant primers (mutations underlined): HSm1-5′, 5′-TGAGGGTGC TCTAAAACTTTATGAGTAGGTCAA-3′; and HSm1-3′, 5′-AGTTTTAGAGCACCCTCACCAGCCTAT-3′.

A human SI genomic clone was first amplified in two separate reaction mixes containing HSm1-5′ and HS-4X in one reaction mix and HSm1-3′ and −183 hSI in the other reaction mix. The products of these two separate reactions were annealed and reamplified with −183 hSI and the 3′ primer to yield a −183 to +54 product that contained a mutation at bases −51 and −50. The mutation was verified by sequencing.

Cell culture, transfection, and luciferase and β-galactosidase assays. Caco-2, HepG2, MDCK, and HeLa cells were grown under the conditions described previously (59). Plasmids for transfection were prepared by alkaline lysis followed by two successive bandings on cesium chloride density gradients. Test plasmids were cotransfected with CMV-βGal (59) and, 48 h later, were lysed with a buffer containing 1% Triton X-100, and the supernatant was analyzed for luciferase activity with an LKB luminometer as described previously (59). β-Galactosidase activity was measured in cell extracts by monitoring the hydrolysis of *O*-nitrophenol-β-D-galactoside to galactose and *O*-nitrophenol (59). Results are reported as light units generated per unit of β-galactosidase activity.

Preparation of nuclear extracts and DNase I footprinting. Nuclear extracts from cultured cells were prepared by a modification (4) of the method of Dignam et al. (20). Nuclear proteins from mouse liver were isolated exactly as described for rat liver by Hattori et al. (25). Nuclear proteins from small intestine were extracted from intestinal epithelial cells, which were isolated from mouse intestine by a modification of a previously validated method (8). Mice were anesthetized with an intraperitoneal injection of pentobarbital, and the intestine was rinsed in situ with CMF buffer containing 200 mM dithiothreitol, 0.1 M phenylmethylsulfonyl fluoride, and 2 μg each of aprotinin, leupeptin, and bestatin per ml (CMF contains 5 mM KCl, 0.3 mM KH_2PO_4, 138 mM NaCl, 4 mM $NaHCO_3$, and 0.3 mM Na_2HPO_4 [pH 7.4]). The mouse was then perfused for 3 min via the left ventricle with preoxygenated, prewarmed CMF buffer containing 30 mM EDTA, 200 mM dithiothreitol, 0.1 M phenylmethylsulfonyl fluoride, and 2 μg each of aprotinin, leupeptin, and bestatin per ml. The intestine was removed, the contents were collected by rinsing the lumen with additional CMF buffer containing protease inhibitors, and the cells were washed three times. Nuclear protein was then extracted by the method described for cultured cells. The protein concentration of extracts was determined (11), and aliquots were stored at −80°C.

DNase I footprinting was performed essentially as described by Gumucio et al. (24). The DNA for footprinting the human SI gene was −303 to +60 HS-CAT (59). To label the coding strand, the plasmid was cut with *Xba*I, the overlapping end was filled in with ^{32}P-labeled deoxynucleoside triphosphates (dNTPs) and Klenow enzyme, the labeled fragment was released with *Sal*I, and the product was gel purified with a DEAE membrane. The noncoding strand was labeled in the same manner on the *Sal*I end. Footprinting of the mouse SI gene was done with a subclone of λMS35 that contained bases −202 to +54. The coding and the noncoding strands were labeled at a *Hin*dIII or *Xba*I site, respectively. Nuclear protein was incubated for 10 min at room temperature in a solution containing 6% polyethylene glycol, 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid, pH 7.5), 60 mM KCl, 1 mM $MgCl_2$, 0.5 mM dithiothreitol, 12.5% glycerol, 0.1 mM EDTA, and 1 μg of poly(dI-dC). Labeled DNA (20,000 cpm) was then added and incubated at room temperature for 20 min. After binding, the solution was made 5 mM $MgCl_2$ and digested with 5 μg of DNase I per ml for 90 s at 25°C. After phenol-chloroform extraction and ethanol precipitation, the DNA was separated on a denaturing 8% polyacrylamide gel and autoradiographed. A Maxam-Gilbert A+G cleavage reaction was performed on part of the probe and run in an adjacent lane of the gel so that protected bases could be identified. Also, a dideoxy sequencing reaction was run in four adjacent lanes to further identify protected bases.

EMSA. For electrophoretic mobility shift assays (EMSAs), complementary oligonucleotides with overlapping ends were synthesized, annealed, and labeled by filling in the ends with ^{32}P-labeled dNTPs and Klenow enzyme. Binding reactions with labeled probe and nuclear extracts were performed as described for DNase I footprinting with the omission of polyethylene glycol. The binding reaction mixes were separated in a 4% polyacrylamide gel (acrylamide-bisacrylamide, 30:1) with a buffer containing 0.5× TBE (Tris-borate-EDTA) without recirculation. In some experiments, a high-ionic-strength buffer system was used (4), which gave the same results as the low-ionic-strength buffer.

The following oligonucleotides were used (uppercase letters represent the SI promoter coding strand sequences, and lowercase letters represent bases added for labeling and subcloning; the 5′ ends have a *Bam*HI site, and the 3′ ends have a *Bgl*II site; mutations are in boldface):

SIF1: 5′-gatccGGCTGGTGAGGGTGCAATAAAACTTTATGAGTA-3′
SIF1m: 5′-gatccTGAGGGTGCTCTAAAACTTTA-3′
SIF1m-ext:
5′-gatccGGCTGGTGAGGGTGCTCTAAAACTTTATGAGTA-3′
SIF1-2: 5′-gatccATAAAACTTTATa-3′
SIF2: 5′-gatccCAGAAGATTATTAAACACTGAa-3′
SIF2-ext:
5′-gatccATGTTGAGCAGAAGATTATTAAACACTGATAGGCTa-3′
SIF3: 5′-gatcCTGACAGTACAATTACTAATTAACTTAGA-3′
C/EBP: 5′-GATCAAGCTGCAGATTGCGCAATCTGCAGCTT-3′

RESULTS

Mouse and human SI gene promoters are evolutionarily conserved. Anchored PCR was used to clone a mouse SI cDNA that extended to the cap site of the mRNA (pMS-A1). The mouse SI cDNA contained 372 bases of the 5′ end of the mRNA, included an untranslated region of the same length as the human SI mRNA (59), and recognized an mRNA species of approximately 6.5 kb on Northern (RNA blot) analysis which corresponds to the size of the mRNA in rats (52), rabbits (29), and humans (53) (data not shown). A mouse genomic library was screened with this cDNA, yielding four independent clones. A map of two of these clones is shown in Fig. 1; the two remaining clones were contained entirely within λMS27. Primer extension analysis demonstrated that the major sites of transcriptional initiation corresponded to those identified previously for the human SI gene (59) (Fig. 2). The +1 nucleotide was assigned based on the most intense start site in the human gene; the −1 and +4 sites were also seen in the human gene, but the +5 site was more prominent in the mouse than in the human gene. A total of 3,347 bases of the 5′-flanking region of the mouse gene and exon 1 were sequenced and compared with the same region of the human gene with the Pustell DNA matrix algorithm (MacVector; International Biotechnologies, Inc.). The only region of significant similarity was within the first 180 bases from the start of transcription, which was 82% identical between the two species (Fig. 3). There is complete identity in the region of the TATA consensus, which begins 27 bases upstream from the start of transcription in both species.

Promoter region of the mouse and human SI genes binds both intestine-specific and more widely expressed nuclear proteins. We previously demonstrated that DNA elements that direct intestinal cell line-specific transcription of the human SI gene appeared to be located within 324 bases of the transcriptional start site (59). The marked degree of similarity between species further suggested the possibility that this region contained important regulatory elements for SI gene transcription. Rather than assessing function by random deletion of DNA, we chose to first identify elements that bind nuclear proteins and then assess the function of those elements. Nuclear proteins were isolated from Caco-2 cells, an intestinal cell line that we have previously shown supports transcription of SI (59), and HepG2 cells, a hepatoblastoma cell line that does not express SI. Caco-2 cells have been shown to attain a more differentiated phenotype after they have been maintained in a confluent state (41). Although SI mRNA is expressed in both pre- and postconfluent Caco-2 cells, the steady-state level of mRNA is greater in postconfluent cells (2). Therefore, we chose to examine nuclear extracts from Caco-2 cells that had been maintained in culture for 4 and 19 days to determine whether

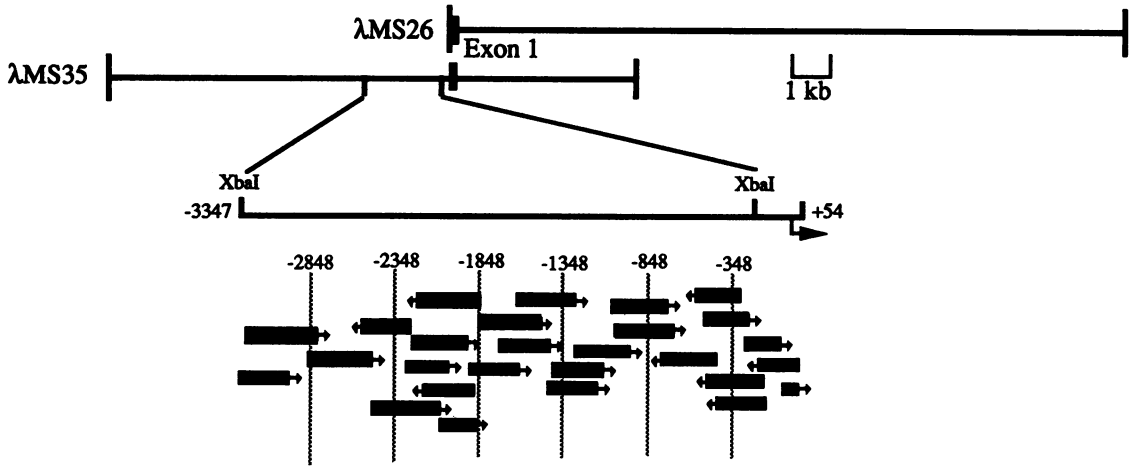

FIG. 1. Mouse SI genomic clones. Map of genomic clones obtained by screening a mouse genomic library with the pMS-A1 cDNA probe. The cloning and description of this cDNA are given in Materials and Methods and Results. The sequencing strategy for bases −3347 to +54 is shown. The transcriptional start site (arrow) was assigned according to primer extension analysis (Fig. 2).

there were qualitative differences in the pattern of nuclear proteins that bind to the promoter.

DNase I footprint analysis with nuclear proteins from Caco-2 cells demonstrated four protected regions on the coding strand of the human SI 5′-flanking region (Fig. 4). The presence of zinc ions in the binding reaction mix did not affect these binding sites or reveal new areas of binding. The region of the TATA box was protected from approximately bases −30 to −22. The next protected region, designated SIF1 (SI footprint 1), lies within a 22-bp segment that is identical between the mouse and human genes (Fig. 3 and 4). The 3′ end of SIF1 was difficult to determine because DNase I cleavage of the bare DNA was poor in this region. Analysis of the noncoding strand confirmed the 5′ boundary and did not further delineate the 3′ end of the footprint (data not shown). However, analysis of several different gels suggested that the footprint extended from bases −56 to −36. SIF1-binding protein was not present in nuclear extracts from HepG2 cells. When a labeled coding strand from the mouse gene 5′-flanking region was used, binding proteins were identified at the same sites as in the human gene (Fig. 4). Two regions of the mouse and human promoters were protected by nuclear extracts from both Caco-2 and HepG2 cells. SIF2 extends from bases −86 to −72, and SIF3 extends from bases −165 to −145, with surrounding sites of hypersensitive cleavage (Fig. 4).

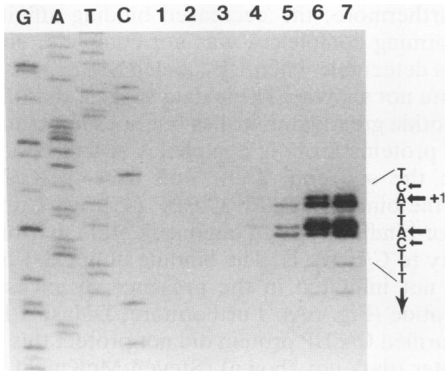

FIG. 2. Primer extension analysis of mouse SI gene. An oligonucleotide primer complementary to bases +34 to +54 was labeled with ^{32}P at the 5′ end and used to prime reverse transcriptase after annealing to RNA (see text). The products were separated in an 8% acrylamide gel. A sequencing reaction with a subclone of λMS35 and the labeled primer was run in parallel to identify the size of the extension product. The +1 site was assigned based on the major start site in the human sequence so that optimal alignment between the two genes was possible. GATC, sequencing reaction; lane 1, probe alone; lane 2, yeast tRNA, 50 μg; lane 3, *E. coli* total RNA, 50 μg; lanes 4, 5, 6, and 7, 5, 10, 25, and 50 μg of mouse jejunal total RNA, respectively.

There were no qualitative differences in the patterns of footprints obtained from day 4 or day 19 Caco-2 nuclear proteins. Therefore, if there are new transcriptional proteins that interact with the SI promoter which are expressed following confluence of Caco-2 cells, they are not detectable by this assay, or they are involved in protein-protein interactions and not protein-DNA interactions.

Protein-binding sites in the SI promoter correspond to positive *cis*-acting regulatory elements which drive intestine-specific transcription. The functional significance of the SIF1, SIF2, and SIF3 elements was assessed in transfection experiments. We have previously shown that human SI-luciferase gene (hSI-LUC) constructs initiate transcription at the same site as the endogenous SI gene (59). Δ183 hSI-LUC directed much greater cell-specific transcription in Caco-2 than in HeLa and HepG2 cells (Fig. 5). Deletion of SIF3 reduced expression by fourfold in Caco-2 cells, demonstrating that this region contains a positive regulatory element. Removal of SIF2 resulted in a loss of luciferase expression, which shows that SIF2 also contains a positive regulatory element for SI gene transcription. Finally, deletion of SIF1 led to a further fourfold reduction in luciferase expression, which was reduced to near the level of the construct containing only the TATA box region. Therefore, all three of the SI footprints correspond to positive regulatory elements for SI gene transcription in Caco-2 cells. Luciferase expression in HepG2 and HeLa cells was low for each of the constructs, and the intestine specificity of expression was maintained until SIF1 was deleted. Addition of SIF2 and SIF3 resulted in small increases in luciferase expression in HepG2 and

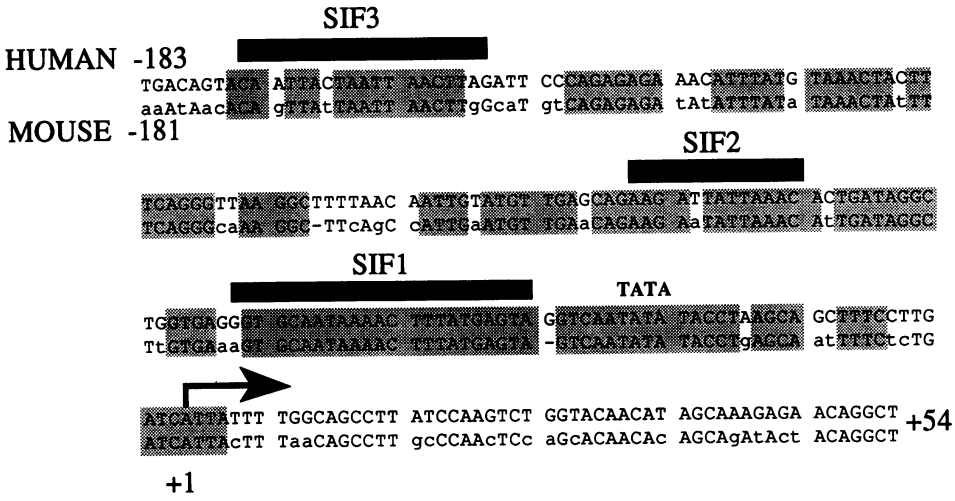

FIG. 3. Comparison of mouse and human SI genes. The immediate 5′-flanking regions of the mouse and human SI genes are shown as aligned with MacVector (International Biotechnologies, Inc.), with the human gene above the mouse gene. The transcriptional start site, +1, is indicated by an arrow. The +54 site, within the first exon, represents the 3′ end of constructs used in the studies described. The TATA consensus sequence is indicated. The bars labeled SIF1, SIF2, and SIF3 delineate the bases that are protected in DNase I footprinting analysis, shown in Fig. 4. Shaded areas represent regions of complete identity between the human and mouse genes that extend for 3 bases or more. Only the areas of identity located upstream of the transcriptional start site are shown.

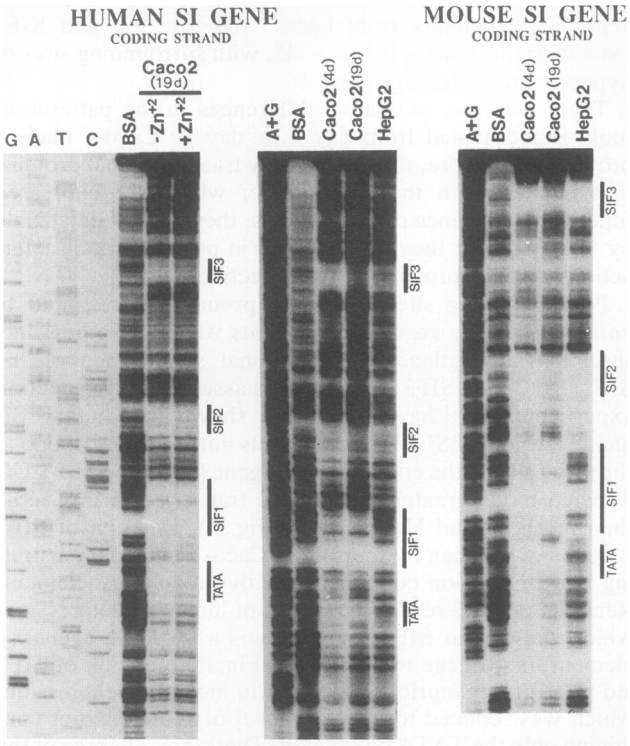

FIG. 4. DNase I footprinting analysis of mouse and human SI gene promoters. DNase I footprinting was performed with end-labeled human and mouse SI gene promoters as described in Materials and Methods. Protected bases were identified by the A+G chemical cleavage sequencing reaction as well as by dideoxynucleotide termination sequencing. BSA, bovine serum albumin. Caco-2 cells were used at day 4 or 19 of culture, as indicated.

HeLa cells over that with constructs containing the SIF1 and the TATA box alone, but the levels remained quite low.

SIF1 nuclear DNA-binding proteins are expressed in intestinal cell lines. EMSA was used to further characterize nuclear proteins that bind to the SIF1 DNA element. Two specific complexes, designated A and B, were formed with Caco-2 nuclear extracts (Fig. 6A). Complex A was detected with 5 μg of nuclear extract, whereas complex B was identified with increasing amounts of protein. An oligonucleotide containing a two-base mutation at bases −50 and −51 (SIF1m) was unable to displace proteins that bound to the wild-type sequence, forming complex A (Fig. 6A). However, SIF1m was able to compete to a limited degree with SIF1 for binding of the proteins forming complex B (Fig. 6A, lower panel). Furthermore, the decreased binding affinity of the proteins forming complex A was not complete, since complex A was detectable when ^{32}P-labeled SIF1m was used for EMSA (data not shown). These data suggest that the SIF1m oligonucleotide greatly diminishes but does not eliminate the binding of proteins forming complexes A and B.

Because the sequence from −53 to −49 (GCAAT) is similar to the binding site for C/EBP (45), we attempted to compete for binding with an oligonucleotide that binds with high affinity to C/EBP (1). The binding of nuclear protein to SIF1 was not inhibited in the presence of excess C/EBP oligonucleotide (Fig. 6A). Furthermore, DNase I footprinting with purified C/EBP protein did not protect this region of the promoter (data not shown) (Steven McKnight provided the C/EBP protein). Another potentially important structural feature of the SIF1 element is a 5-base inverted repeat separated by two bases, 5′-ATAAAACTTTAT-3′. Other DNA regulatory elements with this structural feature have been found to bind protein dimers (1). However, a double-stranded oligonucleotide (SIF1-2) containing only this inverted repeat did not bind nuclear proteins (data not shown) and did not compete for binding of complex A or B (Fig. 6A).

Nuclear extracts from other intestinal, nonintestinal epithelial, and nonepithelial cell lines were examined for the presence of SIF1-binding protein (Fig. 6B; results summa-

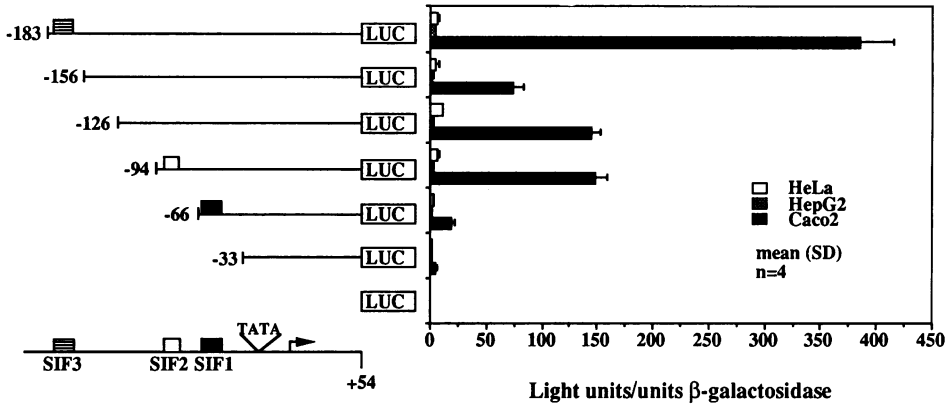

FIG. 5. Functional analysis of deletional mutants of human SI gene promoter. Mutants in which the deleted human SI gene was linked to luciferase were cotransfected with CMV-βGal into Caco-2, HepG2, or HeLa cells as described in Materials and Methods. Each bar represents four independent transfections, and the mean and standard deviation are shown.

rized in Table 1). The colon cancer cell lines HT29 and T84 had a specific complex that was slightly smaller than complex A found in Caco-2 cells. Differentiated C2T cells, a mouse skeletal muscle line, had an inconsistent band in the region of complex A that did not compete well with an excess of SIF1. Nuclear extracts from HepG2, IEC6 (a cell line derived from normal rat neonatal intestine [44]), NT2 (a human embryonal carcinoma cell line [3]), K562 (a human erythroleukemia line), MEL (a mouse erythroleukemia line), 3T3 fibroblasts, and B310 (a B-lymphocyte line) had no detectable SIF1-binding protein. A specific DNA-protein complex that migrated with complex A was detected in HeLa nuclear extracts at low concentrations. However, even when large amounts of HeLa extract were used, complex B was not detected (Fig. 6B). Each of the extracts used in this analysis contained intact nuclear proteins, as evidenced by the detection of specific complexes to a binding site which recognizes the ubiquitous transcription factor Sp1 (23) (data not shown). This analysis shows that the SIF1-binding protein(s) forming complex A is expressed in several intestinal cell lines and at low levels in HeLa cells, but is absent in other nonintestinal epithelial and nonepithelial cell lines. The SIF1-binding protein(s) that forms complex B is expressed exclusively in Caco-2 cells.

In vitro mutagenesis demonstrates importance of the SIF1 element for SI gene transcription. The SIF1-binding site in Δ183 hSI-LUC was mutated at bases −50 and −51, the same mutation that failed to compete well with wild-type SIF1 in EMSA, and was used for transfection experiments in Caco-2 and HeLa cells. The level of luciferase expression directed by the mutant in Caco-2 cells was 84% lower than that by the wild-type construct (Fig. 7). This marked decrease in expression demonstrates the importance of this element for transcription in intestinal cell lines. The mutation had little effect on transcription in HeLa cells, although the low level of transcription with the wild-type construct may make differences difficult to discern. The less than complete inhibition of transcription in Caco-2 cells resulting from the SIF1 mutation may be due to residual binding of transcription factors to the SIF1 site, since the mutation did not completely eliminate binding of either complex A or B in EMSAs. Alternatively, additional sites surrounding the TATA box region may be important for transcriptional regulation. Further fine mutational analysis will be required to evaluate these possibilities.

SIF2 and SIF3 complex with nuclear proteins expressed in a tissue-restricted pattern. Abundant and specific SIF3-binding proteins were identified in Caco-2, HepG2, HT29, and T84 nuclear extracts (Fig. 8A). The other intestinal cell line tested, IEC6, had barely detectable binding activity. Binding complexes were also identified in a number of other nuclear extracts, including HeLa, NT2, MEL, B310, and C2T cells (Fig. 8A; results summarized in Table 1). However, the patterns of binding proteins in these cell lines differed from those in intestinal and liver cells. Caco-2, HepG2, HT29, and T84 nuclear extracts demonstrated a high-mobility doublet complex, with the upper band the most intense. HeLa, NT2, MEL, B310, and C2T nuclear extracts formed a complex of much lower intensity and of different mobility from those found in intestinal cell lines (Fig. 8A). In addition to this complex, B310 and C2T cells also had specific complexes of lower mobility.

The patterns of SIF2-nuclear protein complexes were very similar to those obtained with SIF3, although the intensity of the complexes was lower (Fig. 8B; summarized in Table 1). A high-mobility doublet was identified in Caco-2, HT29, and HepG2 cells. In contrast to SIF3-binding complexes, the intensity of the two bands was similar, and T84 had no detectable complexes. HeLa, NT2, MEL, B310, and C2T cells formed a complex with the same mobility as the lower band in Caco-2 cells, which also differed from the pattern found with SIF3. It is worth noting that the same smaller complex identified with SIF3 in B310 extracts was also found with the SIF2 probe.

Because of the similarities in the size of complexes and the tissue pattern of expression, we examined the possibility that SIF2 and SIF3 bound the same nuclear protein(s). Competition with either SIF2 or SIF3 completely inhibited the binding of nuclear protein to SIF2, whereas SIF1 had no effect (Fig. 8C). SIF3 protein complexes were also partially inhibited by competition with SIF2 but not with SIF1. An extended oligonucleotide containing additional sequences flanking SIF2 did not show more avid binding than the original SIF2 oligonucleotide (data not shown; see Materials and Methods for sequence of SIF2-ext). These data suggest that SIF2 and SIF3 may bind the same nuclear protein(s). Furthermore, the relative intensities of the bands on EMSA and the less potent competition of SIF2 for the SIF3-protein complex suggests that SIF3 binds these nuclear proteins with greater affinity than SIF2. Definitive proof of whether

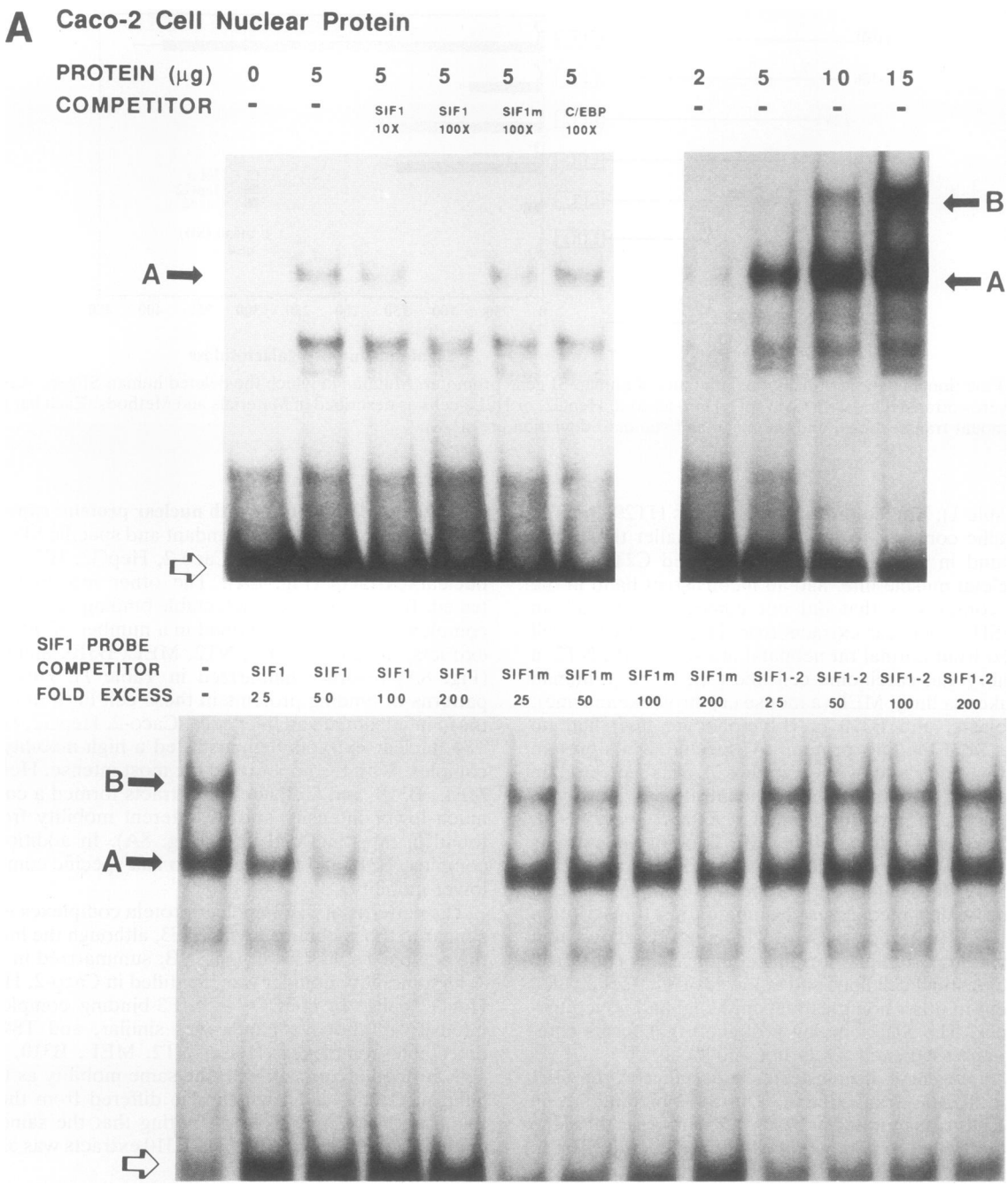

FIG. 6. EMSA of SIF1-binding proteins. (A) Analysis of Caco-2 cell nuclear extract with the SIF1 DNA element. The specific activity of the probe used in these studies was 2.47×10^8 cpm/μg of DNA. Specific complexes are marked A and B. In the upper panel, competitions were performed with either a 10- or 100-fold molar excess (10× and 100×, respectively) of the indicated probe. In the lower panel, the fold excess of the competing oligonucleotides is indicated. The sequences of oligonucleotides used (SIF1, SIF1m, C/EBP, and SIF1-2) are given in Materials and Methods. The open arrow indicates free probe. (B) Analysis of various cell line extracts for expression of SIF1-binding protein. The cell lines used in these experiments included the colon cancer cell lines HT29 and T84; differentiated C2T cells, a mouse skeletal muscle line; HepG2; IEC6, a crypt cell line derived from normal rat neonatal intestine; NT2, a human embryonal carcinoma line; K562, a human erythroleukemia line; MEL, a mouse erythroleukemia line; 3T3 fibroblasts; B310, a B-lymphocyte line; and HeLa cells. Competitor used was a 100-fold molar excess of unlabeled SIF1 double-stranded oligonucleotide. The A and B complexes are indicated with a solid arrow. Open arrow indicates free probe.

the same proteins bind SIF2 and SIF3 will require additional work.

Analysis of a nonintestinal, polarized epithelial cell line. The nonintestinal epithelial cells that we studied, HepG2 and HeLa, do not form polarized monolayers in culture as do Caco-2 cells. It is conceivable that SIF1 nuclear protein and promoter activity might simply be properties of polarized epithelial cells rather than reflecting intestine specificity.

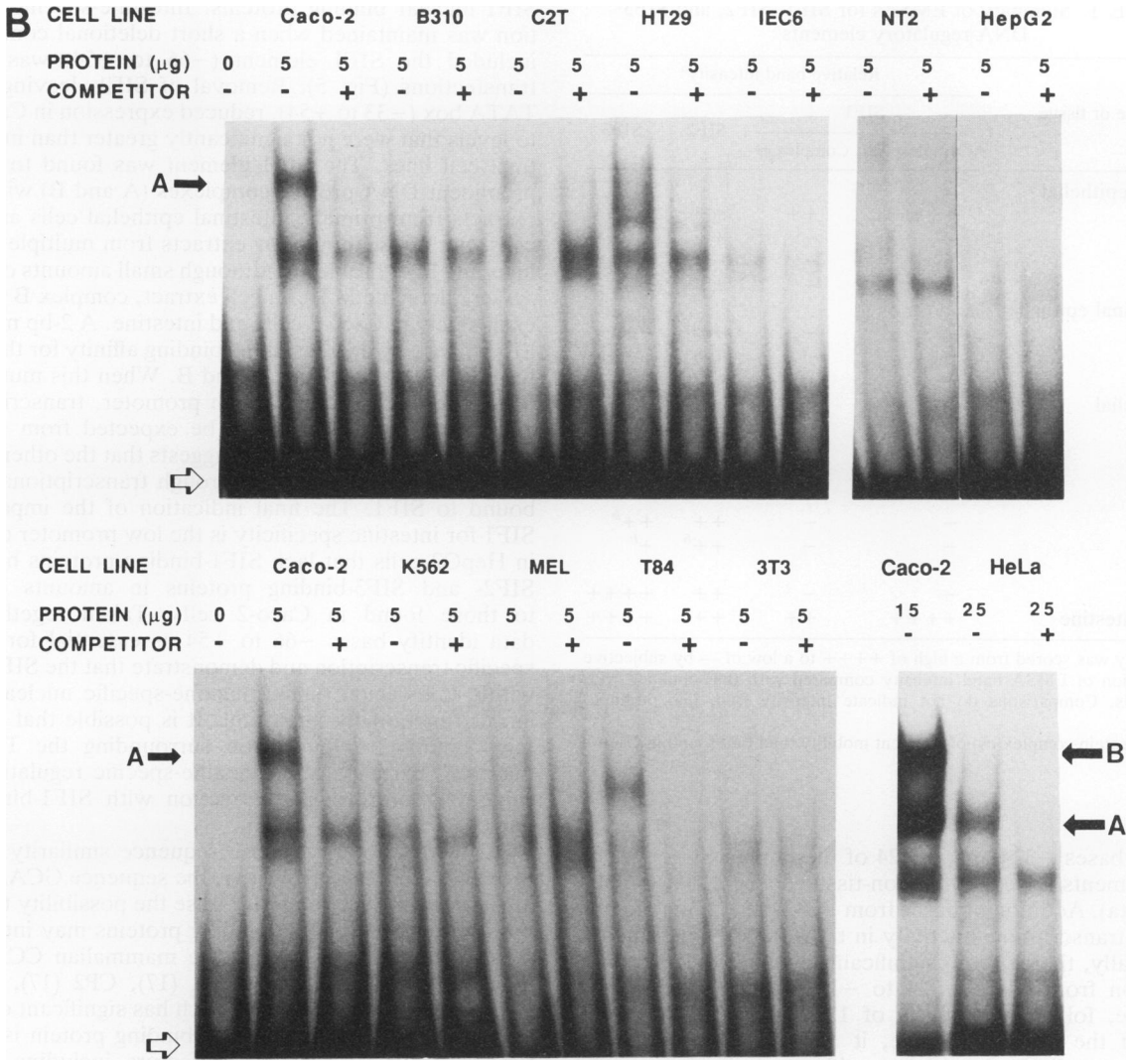

FIG. 6—*Continued.*

Therefore, we studied MDCK cells, which were derived from canine kidney, form a polarized monolayer with a well-developed brush border, and have been extensively examined as a model for vectorial transport in epithelia. Following confluence, nuclear proteins were isolated from MDCK cells and used for EMSA (Fig. 9). SIF1-binding protein was not present, whereas SIF2-, SIF3-, and Sp1-binding proteins were identified. When Δ324 hSI-LUC was transfected into preconfluent MDCK cells, there was very little expression on the order of what was seen with HeLa and HepG2 cells (data not shown). These data further substantiate that transcription from the SI promoter and the pattern of nuclear binding proteins are tissue specific.

Tissue-specific expression of SI promoter DNA-binding proteins in mouse tissues. To determine whether the pattern of expression of nuclear binding proteins in intact tissues was similar to that seen in cell lines, mouse small intestinal and liver nuclear proteins were evaluated. SIF1-binding proteins forming both complexes A and B were present in small intestinal nuclear extracts (Fig. 10A). However, there were no detectable SIF1-binding proteins in liver extracts (Fig. 10B). SIF2-binding protein was detectable in small amounts, and SIF3-binding protein was present in large amounts in both small intestine and liver. Both Sp1 and C/EBP proteins were expressed in small intestine and liver, which indicates that the nuclear extracts contained proteins known to be expressed in both tissues (7, 46).

DISCUSSION

There is currently no information on the DNA-regulatory elements or transcription factors that direct intestine-specific gene expression. The temporal, spatial, and cell lineage-dependent patterns of SI gene expression suggested that it may serve as an excellent model to identify intestine-specific DNA-regulatory elements and their cognate DNA-binding proteins. In the present studies, we show that intestine-specific transcription of the SI gene is dependent on a 183-base promoter which contains elements that are highly conserved between the mouse and human genes. In previous work, we found potential enhancers for transcription located

TABLE 1. Summary of EMSAs for SIF1, SIF2, and SIF3 DNA-regulatory elements

Cell line or tissue	Relative band intensity[a]			
	SIF1		SIF2	SIF3
	Complex A	Complex B		
Intestinal epithelial				
Caco-2	++++	++	++	++++
HT29	+[b]	−	++	++++
T84	+[b]	−	−	++
IEC6	−	−	±[b]	±[b]
Nonintestinal epithelial				
MDCK	−	−	+	++
HepG2	−	−	++	++++
HeLa	+	−	+[b]	+[b]
Nonepithelial				
NT2	−	−	++[b]	+[b]
K562	−	−	−	−
MEL	−	−	++[b]	+[b]
3T3	−	−	−	−
B310	−	−	++	++[b]
C2T	−	−	++[b]	+[b]
Tissues				
Liver	−	−	++	++++
Small intestine	++++	++	++	++++

[a] Intensity was scored from a high of ++++ to a low of — by subjective approximation of EMSA band intensity compared with that obtained with Caco-2 cells. Comparisons do not indicate intensity from one probe to another.

[b] DNA-protein complex(es) of different mobility than that found in Caco-2 extracts.

between bases −324 and −3424 of the human SI gene, but these elements appear to be non-tissue specific (59; unpublished data). Addition of bases from −3424 to −12 kb did not enhance transcriptional activity in transfection experiments (59). Finally, there are no significant regulatory elements in the region from bases −324 to −183 (unpublished data). Therefore, following analysis of 12 kb of the 5′-flanking region of the human SI gene, it appears that sequences sufficient to direct intestine-specific transcription reside within an evolutionarily conserved promoter.

DNase I footprint analysis of the mouse and human SI promoters revealed three regions that proved to be positive regulatory elements: SIF1, located from bases −56 to −36; SIF2, located from bases −86 to −73; and SIF3, located from bases −176 to −156. Of these regulatory elements, SIF1 contains a 22-base region that is perfectly conserved between mouse and human and appears to be crucial for intestine-specific transcription. This conclusion is based on transfection experiments and on the tissue-specific pattern of SIF1 nuclear binding proteins. Intestine-specific transcription was maintained when a short deletional construct that included the SIF1 element (−66 to +54) was used for transfections (Fig. 5). Removal of SIF1, leaving only the TATA box (−33 to +54), reduced expression in Caco-2 cells to levels that were not significantly greater than in nonintestinal cell lines. The SIF1 element was found to form two prominent DNA-protein complexes (A and B) with nuclear extracts from primary intestinal epithelial cells and Caco-2 cells, but not with nuclear extracts from multiple other cell lines and liver (Table 1). Although small amounts of complex A were detected in HeLa cell extract, complex B was found exclusively in Caco-2 cells and intestine. A 2-bp mutation in SIF1 markedly decreased the binding affinity for the proteins forming both complexes A and B. When this mutation was introduced into the full-length promoter, transcription was decreased more than would be expected from simple removal of SIF1 (Fig. 7). This suggests that the other promoter elements may, in part, act through transcriptional proteins bound to SIF1. The final indication of the importance of SIF1 for intestine specificity is the low promoter expression in HepG2 cells that lack SIF1-binding proteins but contain SIF2- and SIF3-binding proteins in amounts equivalent to those found in Caco-2 cells. Taken together, these data identify bases −66 to +54 as essential for intestine-specific transcription and demonstrate that the SIF1 element within this region binds intestine-specific nuclear proteins and is functionally important. It is possible that other proteins binding to the region surrounding the TATA box may also participate in intestine-specific regulation, either separately or through interaction with SIF1-binding proteins.

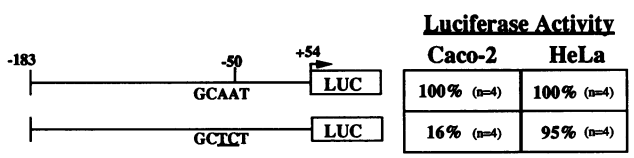

FIG. 7. Functional analysis of SIF1 mutant. Δ183 hSI-LUC and the same construct with bases −50 and −51 mutated as shown were cotransfected with CMV-βGal into Caco-2 cells and HeLa cells as described in Materials and Methods. Data were normalized to the values for the wild-type construct to facilitate comparison between the two cell lines. The results represent four independent transfections.

SIF1 does not have close sequence similarity to known promoter elements. However, the sequence GCAAT and its location within the promoter raise the possibility that one of the known CCAAT box-binding proteins may interact with SIF1. The binding sites for the mammalian CCAAT box-binding proteins including CP1 (17), CP2 (17), CTF/NF1 (17), and CBF (36), each of which has significant differences from SIF1. Furthermore, SIF1-binding protein is absent in tissues known to express these factors, including MEL cells for CP1 and CP2 (5), HepG2 cells for CBF (36), and liver for CTF/NF1 (33). GCAAT is part of the core sequence for C/EBP, which is expressed in small intestine (7) and is important for tissue-specific gene expression (18, 21, 32, 40) and control of proliferation (54). However, a high-affinity binding site for C/EBP was unable to compete for binding with SIF1, purified C/EBP did not footprint the SI promoter, and the pattern of binding with the C/EBP sequence in small intestinal nuclear extract was different than with SIF1. Therefore, it is unlikely that SIF1 binds a member of the C/EBP family of nuclear transcription factors (55).

The SIF1 sequence does not have the core bases that identify binding sites for helix-loop-helix (9) or homeodomain-containing transcription proteins (19, 47), nor is there significant homology with over 93 functional elements for vertebrate genes (34). A search of the 5′-flanking regions of genes that are expressed in intestine, including rat and human intestinal fatty acid-binding protein (50), intestinal alkaline phosphatase (27), human (48) and pig (39) aminopeptidase N, and human lactase (10), did not reveal elements that are similar to SIF1. However, it is possible that once important bases for binding are identified and/or the transcription proteins are isolated or cloned, binding sites in these or other intestinal genes will become evident. From

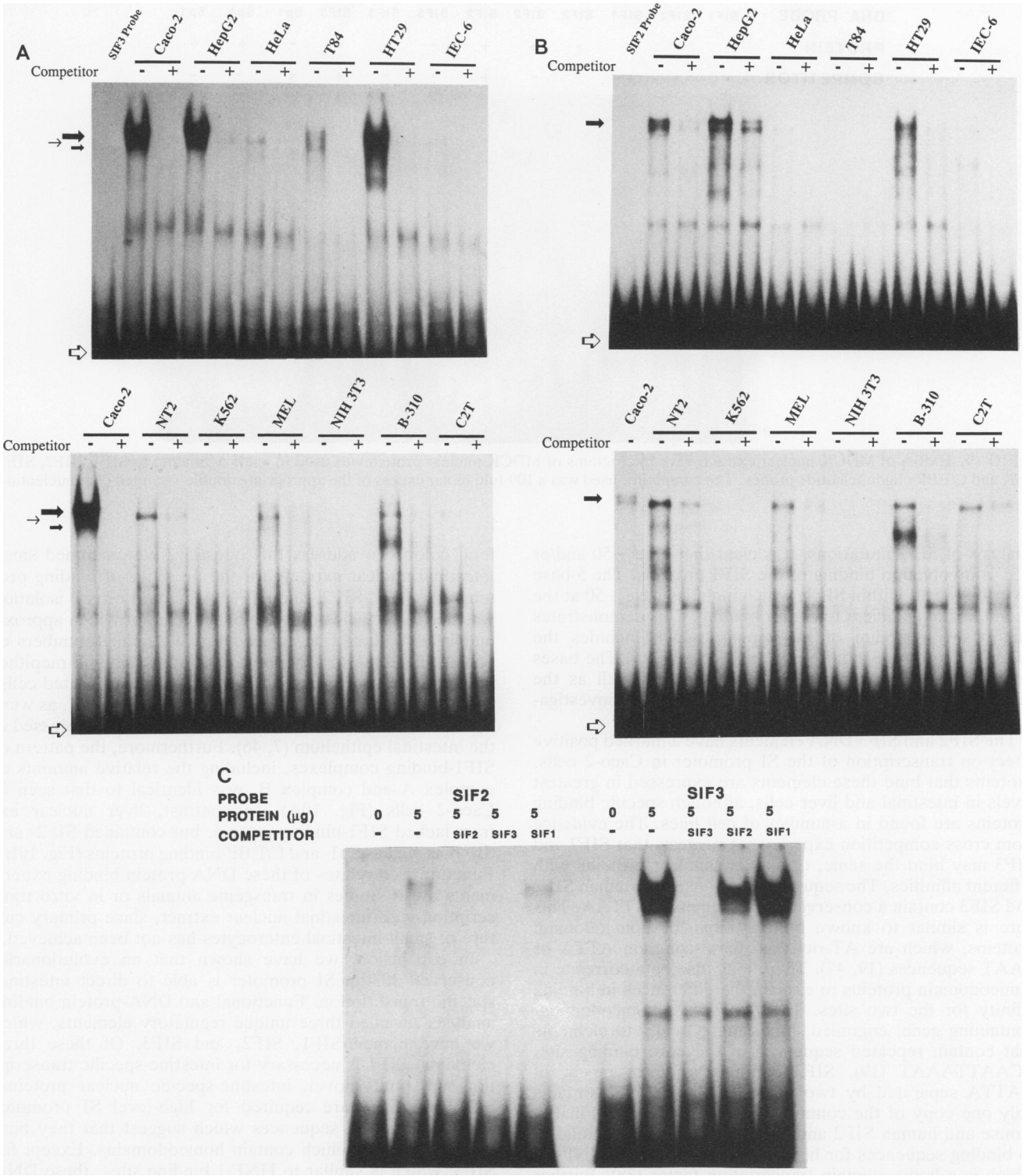

FIG. 8. EMSA of SIF2- and SIF3-binding proteins. (A) SIF3-binding proteins were assessed in multiple cell lines. Five micrograms of nuclear protein was used for each cell line. The specific activity of the probe used in these studies was 2.83×10^8 cpm/μg of DNA, and the films were exposed for 12 h. The competitor used was a 100-fold molar excess of unlabeled SIF3 double-stranded oligonucleotide. The two thick arrows indicate the positions of the major bands in Caco-2, HepG2, and HT29 cells. The thin arrow between the two thick arrows indicates the position of bands in HeLa, NT2, MEL, B310, and C2T cells. The open arrow indicates free probe. (B) SIF2-binding proteins were assessed in multiple cell lines. Five micrograms of nuclear protein was used for each cell line. The specific activity of the probe used in these studies was 3.36×10^8 cpm/μg of DNA, and the films were exposed for 16 h. The competitor used was a 100-fold molar excess of unlabeled SIF2 double-stranded oligonucleotide. The solid arrow indicates the position of the doublet band in Caco-2, HepG2, and HT29 cells. The open arrow indicates free probe. (C) Demonstration that SIF2 and SIF3 bind similar nuclear proteins in Caco-2 cells. The competitor used was a 100-fold molar excess of unlabeled SIF1, SIF2, or SIF3 double-stranded oligonucleotide.

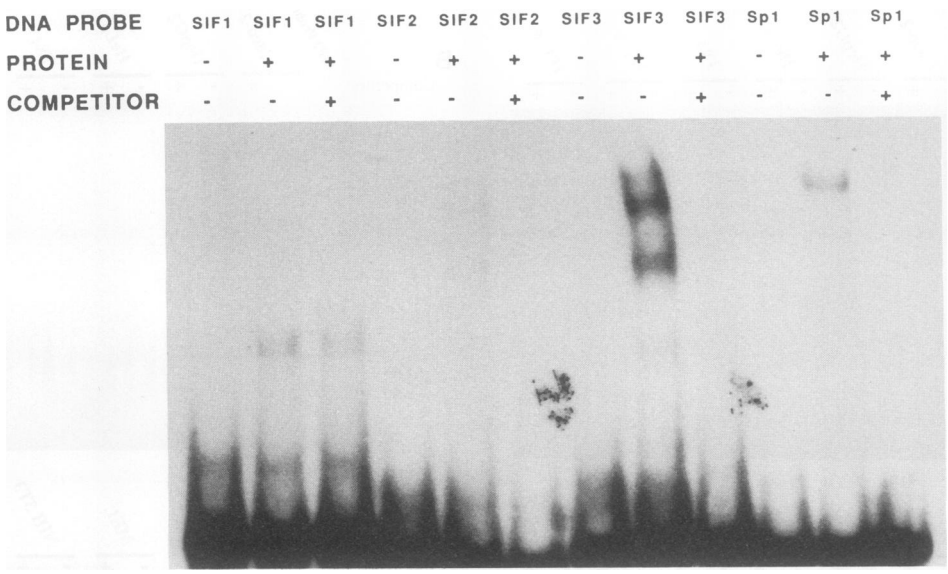

FIG. 9. EMSA of MDCK nuclear extract. Five micrograms of MDCK nuclear protein was used to analyze binding to SIF1, SIF2, SIF3, Sp1, and C/EBP oligonucleotide probes. The competitor used was a 100-fold molar excess of the appropriate double-stranded oligonucleotide.

analysis of SIF1 mutations, it is clear that base −50 and/or −51 is involved in binding of the SIF1 proteins. The 5-base inverted repeat within SIF1, which includes base −50 at the 5′ end, is not sufficient to confer binding. This demonstrates that bases upstream of the repeat, which includes the GCAAT motif, are required for binding activity. The bases required for binding of the SIF1 proteins as well as the number of binding proteins will require further investigation.

The SIF2 and SIF3 DNA elements have a marked positive effect on transcription of the SI promoter in Caco-2 cells. Proteins that bind these elements are expressed in greatest levels in intestinal and liver cells, although specific binding proteins are found in a number of cell lines. The evidence from cross-competition experiments suggests that SIF2 and SIF3 may bind the same, or similar, nuclear proteins with different affinities. The sequences of mouse and human SIF2 and SIF3 contain a conserved core sequence, ATTAA. This core is similar to known binding sites for homeodomain proteins, which are AT-rich and have common ATTA or TAAT sequences (19, 47). There may also be a correlate in homeodomain proteins to explain the differences in binding affinity for the two sites. The *Drosophila* homeodomain-containing gene, engrailed, binds more avidly to elements that contain repeated sequences of the core binding site, TCAATTAAAT (19). SIF3 contains a direct repeat of AATTA separated by two bases, whereas SIF2 contains only one copy of the common sequence ATTAA. Finally, mouse and human SIF2 and SIF3 have sequence similarity to binding sequences for hepatic nuclear factor 1 (HNF-1), which is a homeodomain transcription factor (38). Further work will be required to define the bases in SIF2 and SIF3 that are necessary for binding and to identify the proteins that bind these elements.

Caco-2 cells were derived from a human colonic adenocarcinoma, and therefore, one must question whether findings with this cell line will be relevant for expression of the SI gene in small intestine. Since SI is expressed in fetal colonocytes (43), regulation of the gene may be more directly related to the mechanisms that regulate expression in fetal colon. To address this question, we examined small intestinal nuclear extracts for the presence of binding proteins for SIF1, SIF2, and SIF3. A method of cell isolation was used that yields only epithelial cells, of which approximately 95% should be enterocytes, with small numbers of goblet cells, enteroendocrine cells, and possibly intraepithelial lymphocytes. In nuclear extracts of these isolated cells, each of the three SI-binding proteins was identified, as were Sp1 and C/EBP, both of which are known to be expressed in the intestinal epithelium (7, 46). Furthermore, the pattern of SIF1-binding complexes, including the relative amounts of complex A and complex B, was identical to that seen in Caco-2 cells (Fig. 10A). In contrast, liver nuclear extract lacked SIF1-binding protein, but contained SIF2- and SIF3- as well as Sp1- and C/EBP-binding proteins (Fig. 10B). Functional correlates of these DNA-protein binding experiments await studies in transgenic animals or in vitro transcription with intestinal nuclear extract, since primary culture of small intestinal enterocytes has not been achieved.

In conclusion, we have shown that an evolutionarily conserved 183-bp SI promoter is able to direct intestine-specific transcription. Functional and DNA-protein binding analysis revealed three unique regulatory elements, which we have named SIF1, SIF2, and SIF3. Of these three elements, SIF1 is necessary for intestine-specific transcription and binds novel, intestine-specific nuclear proteins. SIF2 and SIF3 are required for high-level SI promoter activity and have sequences which suggest that they bind nuclear proteins which contain homeodomains. Except for SIF3, which is similar to HNF-1 binding sites, these DNA elements have not, to our knowledge, been identified previously as regulatory elements in other genes. Furthermore, this represents the first report of intestine-specific transcriptional elements and their cognate DNA-binding proteins. Further characterization, including purification and/or cloning, of these proteins may provide insight into SI gene transcription as well as regulation of other intestinal genes. Finally, these studies identify the SI promoter and its cognate DNA-binding proteins as a useful system for future

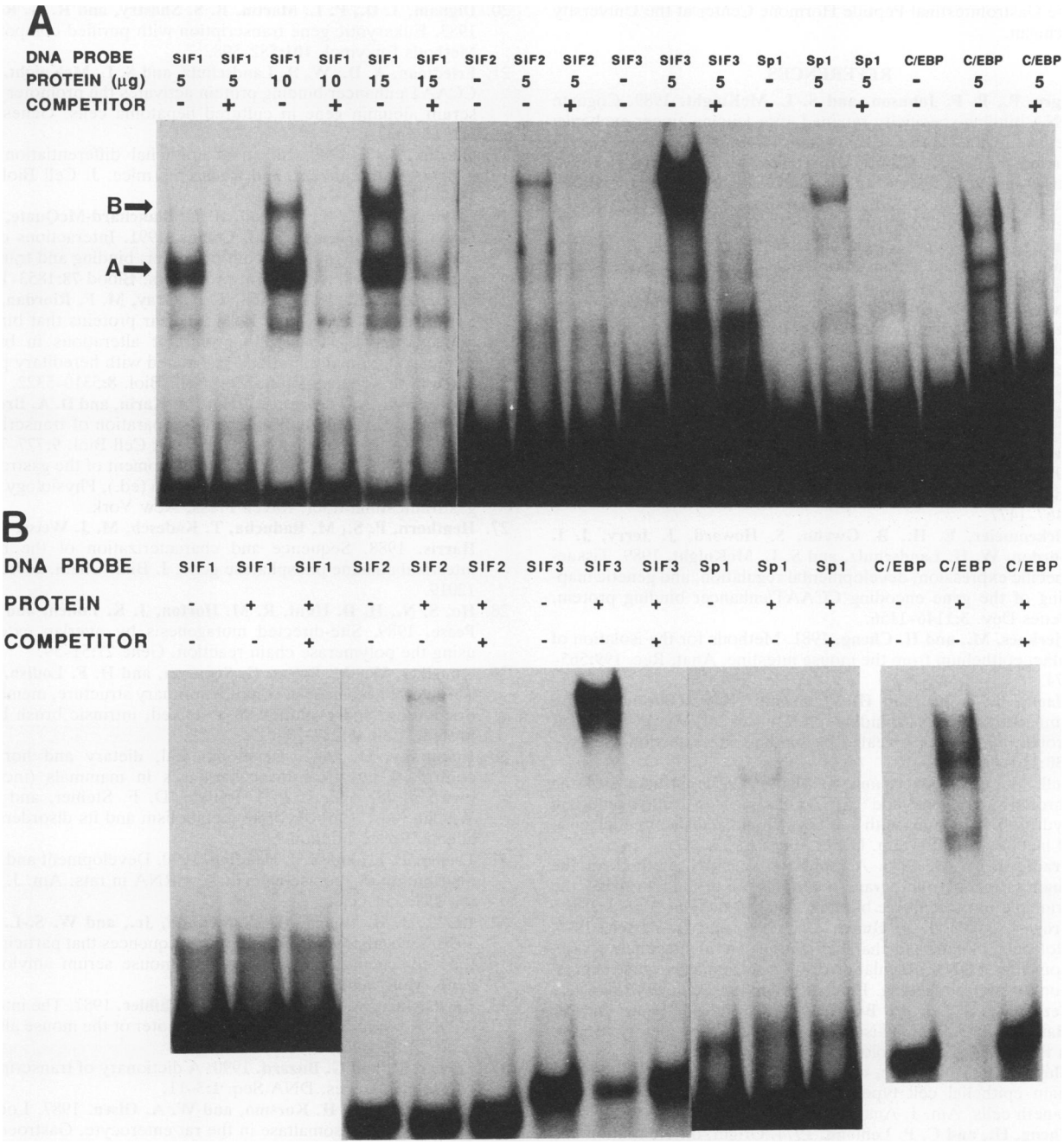

FIG. 10. EMSA of mouse small intestinal and liver nuclear extracts. (A) Mouse small intestinal nuclear protein, isolated as described in Materials and Methods, was used for EMSA with SIF1, SIF2, SIF3, Sp1, and C/EBP oligonucleotide probes. The competitor used was a 100-fold molar excess of the appropriate double-stranded oligonucleotide. The A and B complexes are indicated by arrows. (B) Five micrograms of mouse liver nuclear protein was used to analyze binding to SIF1, SIF2, SIF3, Sp1 and C/EBP oligonucleotide probes. The competitor used was a 100-fold molar excess of the appropriate double-stranded oligonucleotide.

studies of developmentally regulated and tissue-specific gene expression in the intestine.

ACKNOWLEDGMENTS

We appreciate the kind assistance of our colleagues who allowed us to use nuclear extracts isolated from cell lines. Michael Parmacek provided C2T and 3T3 extracts; Deborah Gumucio provided K562, MEL, B310, IEC6, HT29, and T84 extracts; and Juanita Merchant provided NT2 cell extracts. We also thank Steven McKnight for kindly providing purified C/EBP for use in these studies. We thank Jeffery Leiden for his careful review of the manuscript.

This work was supported by NIH grant DK41393, a VA Research Associate Award and Merit Review (P.G.T.), and pilot/feasibility grants from the Michigan Diabetes Research and Training Center

and the Gastrointestinal Peptide Hormone Center at the University of Michigan.

REFERENCES

1. **Agre, P., P. F. Johnson, and S. L. McKnight.** 1989. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. Science **246:**922–926.
2. **Anderson, J. M., C. M. Van Itallie, M. D. Peterson, B. R. Stevenson, E. A. Carew, and M. S. Mooseker.** 1989. ZO-1 mRNA and protein expression during tight junction assembly in caco-2 cells. J. Cell Biol. **109:**1047–1056.
3. **Andrews, P. W.** 1984. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. Dev. Biol. **103:**285–293.
4. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
5. **Barnhart, K. M., C. G. Kim, S. S. Banerji, and M. Sheffery.** 1988. Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine α-globin gene. Mol. Cell. Biol. **8:**3215–3226.
6. **Beaulieu, J.-F., M. M. Weiser, L. Herrera, and A. Quaroni.** 1990. Detection and characterization of sucrase-isomaltase in adult human colon and in colon polyps. Gastroenterology **98:** 1467–1477.
7. **Birkenmeier, E. H., B. Gwynn, S. Howard, J. Jerry, J. I. Gordon, W. H. Landschulz, and S. L. McKnight.** 1989. Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev. **3:**1146–1156.
8. **Bjerknes, M., and H. Cheng.** 1981. Methods for the isolation of intact epithelium from the mouse intestine. Anat. Rec. **199:**565–574.
9. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
10. **Boll, W., P. Wagner, and N. Mantei.** 1991. Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. Am. J. Hum. Genet. **48:**889–902.
11. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
12. **Broyart, J. P., J. P. Hugot, C. Perret, and A. Porteu.** 1990. Molecular cloning and characterization of a rat intestinal sucrase-isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by sucrose feeding. Biochim. Biophys. Acta **1087:**61–67.
13. **Cezard, J. P., J. P. Broyart, P. Cuisinier-Gleizes, and H. Mathieu.** 1983. Sucrase-isomaltase regulation by dietary sucrose in the rat. Gastroenterology **84:**18–25.
14. **Cheng, H.** 1974. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. IV. Paneth cells. Am. J. Anat. **141:**521–536.
15. **Cheng, H., and C. P. Leblond.** 1974. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. Am. J. Anat. **141:**461–480.
16. **Cheng, H., and C. P. Leblond.** 1974. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian theory of the origin of the four epithelial cell types. Am. J. Anat. **141:**537–562.
17. **Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp.** 1988. Human CCAAT-binding proteins have heterologous subunits. Cell **53:**11–24.
18. **Christy, R. J., V. W. Yang, J. M. Ntambi, D. E. Geiman, W. H. Landschulz, A. D. Friedman, Y. Nakabeppu, T. J. Kelly, and M. D. Lane.** 1989. Differentiation-induced gene expression in 3T3-L1 preadipocytes: CCAAT/enhancer binding protein interacts with and activates the promoters of two adipocyte-specific genes. Genes Dev. **3:**1323–1335.
19. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1988. The sequence specificity of homeodomain-DNA interaction. Cell **54:**1081–1090.
20. **Dignam, J. D., P. L. Martin, B. S. Shastry, and R. G. Roeder.** 1983. Eukaryotic gene transcription with purified components. Methods Enzymol. **101:**582–598.
21. **Friedman, A. D., W. H. Landschulz, and S. L. McKnight.** 1989. CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells. Genes Dev. **3:**1314–1322.
22. **Gordon, J. I.** 1989. Intestinal epithelial differentiation: new insights from chimeric and transgenic mice. J. Cell Biol. **108:** 1187–1194.
23. **Gumucio, D. L., K. L. Rood, K. L. Blanchard-McQuate, T. A. Gray, A. Saulino, and F. S. Collins.** 1991. Interactions of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters. Blood **78:**1853–1863.
24. **Gumucio, D. L., K. L. Rood, T. A. Gray, M. F. Riordan, C. I. Sartor, and F. S. Collins.** 1988. Nuclear proteins that bind the human gamma-globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemoglobin. Mol. Cell. Biol. **8:**5310–5322.
25. **Hattori, M., A. Tugores, L. Veloz, M. Karin, and D. A. Brenner.** 1990. A simplified method for the preparation of transcriptionally active liver nuclear extracts. DNA Cell Biol. **9:**777–781.
26. **Henning, S. J.** 1987. Functional development of the gastrointestinal tract, p. 285–300. *In* L. R. Johnson (ed.), Physiology of the gastrointestinal tract. Raven Press, New York.
27. **Henthorn, P. S., M. Raducha, T. Kadesch, M. J. Weiss, and H. Harris.** 1988. Sequence and characterization of the human intestinal alkaline phosphatase gene. J. Biol. Chem. **263:**12011–12019.
28. **Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease.** 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77:**51–59.
29. **Hunziker, W., M. Spiess, G. Semenza, and H. F. Lodish.** 1986. The sucrase-isomaltase complex: primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. Cell **46:**227–234.
30. **Kodolvsky, O.** 1981. Developmental, dietary and hormonal control of intestinal disaccharidases in mammals (including man), p. 481–522. *In* P. J. Randle, D. F. Steiner, and W. J. Whelan (ed.), Carbohydrate metabolism and its disorders. Academic Press, London.
31. **Leeper, L. L., and S. J. Henning.** 1990. Development and tissue distribution of sucrase-isomaltase mRNA in rats. Am. J. Physiol. **258:**G52–G58.
32. **Li, X., J. H. Huang, H. Y. Rienhoff, Jr., and W. S.-L. Liao.** 1990. Two adjacent C/EBP-binding sequences that participate in the cell-specific expression of the mouse serum amyloid A3 gene. Mol. Cell. Biol. **10:**6624–6631.
33. **Lichtsteiner, S., J. Wuarin, and U. Schibler.** 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell **51:**963–973.
34. **Locker, J., and G. Buzard.** 1990. A dictionary of transcriptional control sequences. DNA Seq. **1:**3–11.
35. **Lorenzsonn, V., H. Korsmo, and W. A. Olsen.** 1987. Localization of sucrase-isomaltase in the rat enterocyte. Gastroenterology **92:**98–105.
36. **Lum, L. S. Y., L. A. Sultzman, R. J. Kaufman, D. I. H. Linzer, and B. J. Wu.** 1990. A cloned human CCAAT-box-binding factor stimulates transcription from the human *hsp70* promoter. Mol. Cell. Biol. **10:**6709–6717.
37. **Maxwell, I. H., G. S. Harrison, W. M. Wood, and F. Maxwell.** 1989. A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. BioTechniques **7:**276–280.
38. **Mendel, D. B., and G. R. Crabtree.** 1991. HNF-1, a member of a novel class of dimerizing homeodomain proteins. J. Biol. Chem. **266:**677–680.
39. **Olsen, J., L. Laustsen, U. Karnstrom, H. Sjostrom, and O. Noren.** 1991. Tissue-specific interactions between nuclear proteins and the aminopeptidase N promoter. J. Biol. Chem. **266:**18089–18096.
40. **Park, E. A., W. J. Roesler, J. Liu, D. J. Klemm, A. L. Gurney, J. D. Thatcher, J. Shuman, A. Friedman, and R. W. Hanson.**

1990. The role of the CCAAT/enhancer-binding protein in the transcriptional regulation of the gene for phosphoenolpyruvate carboxykinase (GTP). Mol. Cell. Biol. **10**:6264–6272.

41. **Pinto, M., S. Robine-Leon, M.-D. Appay, M. Kedinger, N. Triadou, E. Dussaulx, B. Lacroix, P. Simon-Assmann, K. Haffen, J. Fogh, and A. Zweibaum.** 1983. Enterocytelike differentiation and polarization of the human colon carcinoma cell line caco-2 in culture. Biol. Cell **47**:323–330.
42. **Potten, C. S., and M. Loeffler.** 1990. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons from the crypt. Development **110**:1001–1020.
43. **Potter, G. D.** 1989. Development of colonic function, p. 545–559. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
44. **Quaroni, A., J. Wands, R. L. Trelstad, and K. J. Isselbacher.** 1979. Epithelioid cell cultures from rat small intestine: characterization by morphologic and immunologic criteria. J. Cell Biol. **80**:248–265.
45. **Ryden, T. A., and K. Beemon.** 1989. Avian retroviral long terminal repeats bind CCAAT/enhancer-binding protein. Mol. Cell. Biol. **9**:1155–1164.
46. **Saffer, J. D., S. P. Jackson, and M. B. Annarella.** 1991. Developmental expression of Sp1 in the mouse. Mol. Cell. Biol. **11**:2189–2199.
47. **Scott, M. P., J. W. Tamkun, and G. W. Hartzell.** 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta **989**:25–48.
48. **Shapiro, L. H., R. A. Ashmun, W. M. Roberts, and A. T. Look.** 1991. Separate promoters control transcription of the human aminopeptidase N gene in myeloid and intestinal epithelial cells. J. Biol. Chem. **266**:11999–12007.
49. **Sweetser, D. A., E. H. Birkenmeier, P. C. Hoppe, D. W. McKeel, and J. I. Gordon.** 1988. Mechanisms underlying generation of gradients in gene expression within the intestine: an analysis using transgenic mice containing fatty acid binding protein-human growth hormone fusion genes. Genes Dev. **2**:1318–1332.
50. **Sweetser, D. A., E. H. Birkenmeier, I. J. Klisak, S. Zollman, R. S. Sparkes, T. Mohandas, A. J. Lusis, and J. I. Gordon.** 1987. The human and rodent intestinal fatty acid binding protein genes. J. Biol. Chem. **262**:16060–16071.
51. **Sweetser, D. A., S. M. Hauft, P. C. Hoppe, E. H. Birkenmeier, and J. I. Gordon.** 1988. Transgenic mice containing intestinal fatty acid-binding protein-human growth hormone fusion genes exhibit correct regional and cell-specific expression of the reporter gene in their small intestine. Proc. Natl. Acad. Sci. USA **85**:9611–9615.
52. **Traber, P. G.** 1990. Regulation of sucrase-isomaltase gene expression along the crypt-villus axis of rat small intestine. Biochem. Biophys. Res. Commun. **173**:765–773.
53. **Traber, P. G., L. Yu, G. Wu, and T. Judge.** 1992. Sucrase-isomaltase gene expression along the crypt-villus axis of human small intestine is regulated at the level of mRNA abundance. Am. J. Physiol. **262**:G123–G130.
54. **Umek, R. M., A. D. Friedman, and S. L. McKnight.** 1991. CCAAT-enhancer binding protein: a component of a differentiation switch. Science **251**:288–292.
55. **Williams, S. C., C. A. Cantwell, and P. F. Johnson.** 1991. A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers in vitro. Genes Dev. **5**:1553–1567.
56. **Wiltz, O., C. J. O'Hara, G. D. Steele, and A. M. Mercurio.** 1990. Sucrase-isomaltase: a marker associated with the progression of adenomatous polyps to adenocarcinomas. Surgery **108**:269–276.
57. **Wiltz, O., C. J. O'Hara, G. D. Steele, and A. M. Mercurio.** 1991. Expression of enzymatically active sucrase-isomaltase is a ubiquitous property of colon adenocarcinomas. Gastroenterology **100**:1266–1278.
58. **Wood, W. M., M. Y. Kao, D. F. Gordon, and E. C. Ridgeway.** 1989. Thyroid hormone regulates the mouse thyrotropin beta subunit gene promoter in transfected primary thyrotropes. J. Biol. Chem. **264**:14840–14847.
59. **Wu, G. D., W. Wang, and P. G. Traber.** 1992. Isolation and characterization of the human sucrase-isomaltase gene and demonstration of intestine-specific transcriptional elements. J. Biol. Chem. **267**:7863–7870.