MOLECULAR AND CELLULAR BIOLOGY, Nov. 1994, p. 7340–7351
0270-7306/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 14, No. 11

A Homeodomain Protein Related to caudal Regulates Intestine-Specific Gene Transcription

EUNRAN SUH, LILI CHEN, JENNIFER TAYLOR, AND PETER G. TRABER*

Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Received 29 March 1994/Returned for modification 9 May 1994/Accepted 1 August 1994

The continually renewing epithelium of the intestinal tract arises from the visceral endoderm by a series of complex developmental transitions. The mechanisms that establish and maintain the processes of cellular renewal, cell lineage allocation, and tissue restriction and spatial assignment of gene expression in this epithelium are unknown. An understanding of the regulation of intestine-specific gene regulation may provide information on the molecular mechanisms that direct these processes. In this regard, we show that intestine-specific transcription of sucrase-isomaltase, a gene that is expressed exclusively in differentiated enterocytes, is dependent on binding of a tissue-specific homeodomain protein (mouse Cdx-2) to an evolutionarily conserved promoter element in the sucrase-isomaltase gene. This protein is a member of the caudal family of homeodomain genes which appear to function in early developmental events in *Drosophila melanogaster*, during gastrulation in many species, and in intestinal endoderm. Unique for this homeodomain gene family, we show that mouse Cdx-2 binds as a dimer to its regulatory element and that dimerization in vitro is dependent on redox potential. These characteristics of the interaction of Cdx-2 with its regulatory element provide for a number of potential mechanisms for transcriptional regulation. Taken together, these findings suggest that members of the Cdx gene family play a fundamental role both in the establishment of the intestinal phenotype during development and in maintenance of this phenotype via transcriptional activation of differentiated intestinal genes.

The epithelium of the intestinal tract arises from the visceral endoderm by progression through a series of developmental transitions (reviewed in references 33, 34, and 42). The early signals for intestinal development have been shown to be partially directed by the interaction of mesoderm with endoderm, although the mechanism of this effect has not been elucidated (reviewed in references 23 and 24). Later events include the emergence of different epithelial cell lineages, formation of crypts (proliferative compartment) and villi (differentiated compartment), and the coordinated expression of cell-specific genes. The resultant mature epithelia of the small intestine and colon have marked regional differences in architecture, cell populations, and gene expression which define distinct functional zones of the adult gut (reviewed in references 20, 21, and 50). These region-specific phenotypes are maintained throughout the life of the animal. The regulatory mechanisms that orchestrate these complex developmental transitions and maintain the well-ordered regional phenotypes along the cephalocaudal and crypt-villus axes have not been defined.

Experiments with transgenic mice have shown that the developmental and positional signals for gene expression in the intestinal epithelium reside in regulatory elements of several intestinal genes (20, 25 [and references therein], 41). Analysis of the regulatory regions of the intestinal and liver fatty acid binding proteins (20, 25 [and references therein]) and sucrase-isomaltase (SI) (41) has shown that there are multiple positive and negative elements that remain to be well defined which direct the correct cell lineage and spatial patterns of expression along the crypt-villus and cephalocaudal axes. It is clear, however, from analysis of these three genes that the process by which expression is restricted to the intestine is regulated at the level of transcription. The transcriptional mechanisms that regulate this intestine-specific gene expression are poorly understood.

Many transcriptional proteins that are essential for tissue-specific gene expression in other tissues have been identified and characterized. The elucidation of the mechanisms of tissue-specific gene transcription in these other tissues has provided tools to examine the regulation of developmental and differentiation processes (examples are reviewed in references 48, 49, and 52). Therefore, to advance understanding of the mechanisms that direct the development of the intestinal mucosa, information on transcription factors that are involved in intestinal gene expression is required. To investigate these issues, our laboratory has studied the mechanisms of transcriptional regulation of the SI gene. In the adult small intestinal mucosa, this gene encodes an intestinal brush border disaccharidase that is expressed exclusively in absorptive enterocytes and is absent in the other three epithelial cell types: goblet, enteroendocrine, and Paneth cells (41, 58, 61). Furthermore, SI mRNA is expressed in a complex pattern along the intestinal crypt-villus axis with absent expression in crypt cells, high-level expression in enterocytes located on the lower half of the villus, and lower levels of expression in enterocytes located at the villus tip (41, 58, 61). Finally, its expression mirrors a number of developmental transition points in the small intestine and, therefore, has served as a marker of intestinal development for many investigators (15, 35).

We previously reported that intestinal cell-specific transcription of the SI gene is dependent on an evolutionarily conserved promoter that extends approximately 180 bases upstream of the transcriptional start site (60, 66). The promoter contains a number of elements that are important for transcription which also bind specific nuclear proteins (60). The most important of these elements for intestinal cell-specific transcription is a 22-base region from bases −32 to −54 that is completely

* Corresponding author. Mailing address: 600 Clinical Research Building, 415 Curie Blvd., Philadelphia, PA 19104-6144. Phone: (215) 898-0154. Fax: (215) 573-2024.

conserved between the mouse and human genes (60). This element, designated SIF1 (SI footprint 1), was shown to bind specific nuclear proteins only from intestinal cell lines or mouse small intestinal enterocytes (60).

We now report the molecular cloning of a mouse homeodomain transcription factor related to caudal that binds with high affinity to the SIF1 element and is able to activate transcription of the minimal intestine-specific SI promoter. This predicted amino acid sequence of our cDNA is identical but for one amino acid to the predicted sequence of mouse Cdx-2 (mCdx-2) (28). In addition, this protein is highly homologous to the Cdx-3 homeodomain protein that was cloned from Syrian hamsters, suggesting that Cdx-3 is the hamster ortholog of mCdx-2. DNA-protein interaction experiments using antisera to mCdx-2 show that the major SIF1 binding protein in human intestinal cell lines is Cdx-2. An interesting aspect of the interaction of Cdx-2 with the SIF1 element was that the protein was able to bind to the element as either a monomer or a dimer. Moreover, the formation of a dimer binding complex was accelerated under reducing conditions, suggesting that dimerization of Cdx-2 on the SIF1 element is dependent on the redox state of the protein. Finally, the mCdx-2 mRNA was expressed only in intestinal tissue in mice. Taken together, our results show that Cdx-2 is important for intestine-specific transcription of the SI gene. In addition, as a member of the homeodomain family of genes, our findings identify Cdx-2 as a possible regulatory protein involved in intestinal development.

MATERIALS AND METHODS

Analysis of DNA-protein interactions. Extraction of nuclear proteins from cell lines and electrophoretic mobility shift assays (EMSA) were performed exactly as previously described (60). Copper-orthophenanthroline (Cu/OP) footprinting was performed by using a modification (22) of a previously described method (37). Briefly, EMSA was performed, and the gel was incubated in a solution of 1,10-phenanthroline–copper (37), the gel was transferred to DEAE membrane, and the bands corresponding to complexes A and B were eluted from the membrane and separated in a sequencing gel.

Vector construction and transfection analysis. For transfection studies, mutations in the SI promoter were incorporated by using a PCR strategy as previously described (60). Fidelity of the mutagenesis was verified by sequencing of all constructs. Plasmids for transfection were purified by two rounds of cesium chloride gradient centrifugation, and cells were transfected and analyzed as previously described (60).

Prokaryotic expression cloning. The strategy for cloning SIF1 binding proteins was based on interactions of the proteins with the DNA element (56, 57, 63). Polyadenylated RNA was isolated from adult mouse jejunum (Fast Track; Invitrogen Corp., San Diego, Calif.), and a cDNA library was constructed in λgt11 by using both oligo(dT) and random hexamer priming (Clontech Inc., Palo Alto, Calif.). The mouse small intestinal cDNA library was screened by using a multimerized SIF1 double-stranded oligonucleotide, as previously described (63). The library was plated on *Escherichia coli* Y1090 at a low density (10^4 PFU/150-mm-diameter culture plate), and plaques were transferred to nitrocellulose filters (Stratagene, La Jolla, Calif.). The nitrocellulose filters were subjected to a denaturation-renaturation process with 6 M guanidine-HCl as previously described (57, 63). To probe the filters, a self-annealed, double-stranded oligonucleotide with overlapping complementary ends (*Bam*HI and *Bgl*II sites on the 5′ and 3′ ends, respectively) were ligated to form multimers, and the mixture of multimers was labeled by nick translation, as previously described (31). Binding of the probe to the filters and washing of the filters were performed as previously described (57). Positive plaques were purified to homogeneity with several rounds of binding, and individual plaques were picked for analysis of DNA and expressed protein.

Protein from the purified λ phage was expressed in a lysogen preparation (53). Y1089r− clones were grown on a Luria-Bertani plate at 30°C for 18 to 24 h, and individual colonies were transferred onto Luria-Bertani plates and incubated at 30 or 42°C. Lysogenic plaques, identified by their inability to grow at 42°C, were grown at 30°C until mid-log phase, transferred to 44°C with added isopropyl-β-D-thiogalactopyranoside (IPTG) for 15 min, and then grown 37°C for 2 h with shaking. Aliquots of the culture were pelleted and then resuspended in buffer containing Tris-HCl (pH 7.5), 1 mM EDTA, 5 mM dithiothreitol, and 50 μg of phenylmethylsulfonyl fluoride per ml. The samples were quickly frozen and then thawed, lysozyme (2 mg/ml) was added, and an extraction was performed in 5 M NaCl for 30 min at 4°C. The supernatant, cleared by centrifugation, was used directly for EMSA.

DNA was purified from lysis cultures of purified λ phage by using the preparation system of Qiagen (Studio City, Calif.). Insert DNAs were subcloned as *Eco*RI fragments into pBluescript KS− (Strategene). Sequencing reactions were performed with alkaline-denatured, double-stranded plasmid, as previously described (60, 66).

In vitro expression of cloned protein. For both in vitro and in vivo expression, the coding region of the cloned mCdx-2 cDNA was cloned into pRc/CMV (Invitrogen Corp., San Diego, Calif.), which includes both a prokaryotic transcriptional promoter (T7) and the cytomegalovirus (CMV) promoter and enhancer for expression in eukaryotic cells. The complete coding region of the mCdx-2 cDNA was amplified by PCR with oligonucleotides corresponding to the translational start site and the translational termination site. The 5′ primer is 5′-GGG<u>AAGCTT</u>ACC**ATG**TACGTGAGCTACCTTCTG-3′, and the 3′ primer is 5′-CTA<u>TTCGAA</u>TCACTGGGTGA CAGT-3′. (The *Hin*dIII endonuclease site is underlined, and the translational start codon is in boldface type.)

The amplified DNA fragment was subcloned into the *Hin*dIII site of pRc/CMV, and the insert was sequenced in its entirety to ensure the fidelity of the amplification. The DNA template was sequentially transcribed and translated in the same reaction mixture containing rabbit reticulocyte lysate, amino acid mixture, and T_NT T7 RNA polymerase, by the protocol described by the manufacturer (Promega, Madison, Wis.). Translated protein was analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and directly assayed by EMSA.

GST fusion proteins and production of polyclonal antibodies. A mCdx-2–glutathione *S*-transferase (GST) fusion protein was made by using the pGex system (Pharmacia). The region of the mCdx-2 cDNA from nucleotides 404 to 619 was amplified by PCR and subcloned into the *Bam*HI and *Eco*RI restriction sites of pGex-2TK. The subcloned fragment was sequenced in its entirety to ensure fidelity of the amplification. The expressed fusion protein was purified by using glutathione-Sepharose 4B from the cleared supernatant of IPTG-amplified bacterial cultures. The purification of the 34-kDa fusion protein was confirmed by Coomassie blue staining of SDS-polyacrylamide gels and by immunoblots with an antibody to GST. Polyclonal rabbit antisera were produced by subcutaneous and intramuscular injections of two rabbits with 100 μg of GST–Cdx-2 fusion protein mixed with complete Freund's adjuvant for the initial inoculation and 50 μg mixed with incomplete Freund's adjuvant for the booster injections (Co-

calico Biologicals, Inc. Reamstown, Pa.). The serum was purified by protein A-Sepharose chromatography.

RNase protection assay. RNase protection assays were performed essentially as previously described (41). Single-stranded ^{32}P-labeled probe for mCdx-2 was synthesized from pKS-Cdx-2 linearized with *Dra*I, using T7 RNA polymerase (Promega) and [^{32}P]CTP (3,000 Ci/mmol). This yielded a probe of 260 bp in length. After synthesis, the ^{32}P-labeled RNA probes were purified by digestion with RNase-free DNase I, extraction with phenol and chloroform, and precipitation with ethanol. Solution hybridization was performed for 16 h at 45°C in a solution containing 10 μg of total RNA, 10^5 cpm of probe, 80% formamide, 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4), 0.4 M NaCl, and 1 mM EDTA. Samples were then digested with RNases A and T_1 for 60 min at 15°C, extracted with phenol-chloroform (1:1), and precipitated with ethanol. The pellets were resuspended in loading buffer (80% formamide, 0.1% bromphenol blue, 0.1% xylene cyanol, 1 mM EDTA [pH 8.0]) and separated in a 6% denaturing polyacrylamide gel.

RESULTS

Two protein complexes bind to the SIF1 element at overlapping sites. We previously showed that the SIF1 element in the SI promoter formed two specific complexes with nuclear proteins extracted from intestinal cell lines and mouse small intestinal enterocytes (60). The two complexes were defined as complex A (faster mobility in gel) and complex B (slower mobility in gel). To delineate the protected regions of the SIF1 element in these two complexes, we performed Cu/OP footprinting which provides near-single-base accuracy for defining protected regions of DNA (37). Nuclear protein from Colo-DM cells, a cell line derived from a colon cancer (51), was used for this analysis, since these cells have a large amount of SIF1 binding protein and can transactivate the SI promoter through the SIF1 element in transfection experiments (65, 67). Complex A protected a region of the SIF1 element extending from base −53 to −45, and complex B protected a region of approximately double that length, extending from −53 to −37 (Fig. 1). Therefore, the total region protected in complex B was approximately 17 bases, with the 5′ half of this region protected in complex A. The core of this protected sequence (complex B) is an inverted repeat of ATAAA, although the entire protected area is not completely symmetric (Fig. 1). Therefore, this analysis demonstrated that there were overlapping binding sites for intestinal nuclear proteins interacting with the SIF1 element. Furthermore, there is an inverted repeat at the core of the B complex binding site, and the protein(s) in the A complex binds to the 5′ half of the inverted repeat.

Characterization of the two SIF1-protein complexes. Oligonucleotides containing mutations in the SIF1 element were used in EMSA experiments to further define the characteristics of the DNA-protein interaction. Mutations throughout the binding region were able to abolish formation of complex B (Fig. 2a). However, complex A was formed regardless of the mutation. The double mutation, M3 plus M5, which introduced mutations into the two half-sites of the SIF1 inverted repeat, eliminated specific binding of both complex A and B (Fig. 2a). These data suggested that protein complexes of a similar size formed with the two half-sites of the SIF1 inverted repeat and that complex B might be a dimer of the proteins binding to the two half-sites. However, the Cu/OP footprinting suggested that the A complex only protected the 5′ half-site of the SIF1 element. The findings of the Cu/OP footprinting and

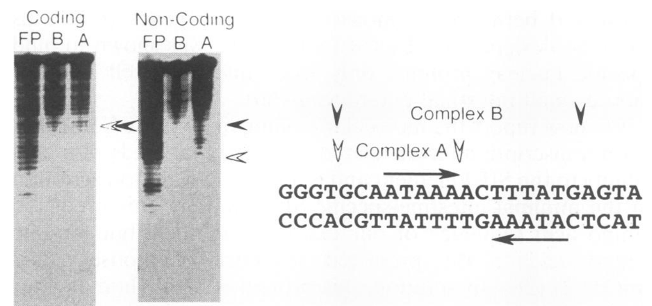

FIG. 1. Cu/OP footprinting of the SIF1 element. Cu/OP footprinting was performed with an extended double-stranded oligonucleotide to provide sufficient flanking sequence. A Maxam-Gilbert sequencing reaction (A+G) of the labeled probe alone was run in the lanes immediately adjacent to the pictured reaction lanes in order to identify the bases in the digestion ladder (not shown). Colo-DM is a cell line derived from a colon cancer (51) that has a large amount of SIF1 binding protein and expresses transfected chimeric SI-reporter gene constructs (data not shown). We previously defined the two specific DNA-protein complexes that form with SIF1 as complex A (faster mobility in gel) and complex B (slower mobility in gel). Protein in complex A protected a region of the SIF1 element extending from base −53 to −45, and proteins forming complex B protected a region of approximately double that length, extending from base −53 to −37 (numbering shown in Fig. 1b). Therefore, the total region protected in complex B is approximately 17 bases, with the 5′ half of this region protected in complex A. The core of this sequence is an inverted repeat of ATAAA (arrows). FP, free probe; A and B, A and B complex, respectively.

the EMSA with SIF1 mutants may be explained by different affinities of nuclear proteins for the two SIF1 half-sites. In support of this possibility, quantification of EMSA complexes showed that the amount of nuclear protein bound to mutants of the 3′ half-site of SIF1 was two- to fivefold greater than proteins binding to mutants of the 5′ half-site (assessed by PhosphorImager) (Fig. 2a).

To directly examine the hypothesis of differential binding to SIF1 half-sites, truncated oligonucleotides corresponding to the half-sites were used for EMSA (Fig. 2b). Both half-sites bound nuclear protein to form a complex of a similar size. Competition studies showed that the affinity of protein for the 5′ half-site (M8) was much greater than the affinity for the 3′ half-site (M9) (Fig. 2b). Therefore, these data are consistent with the results obtained with Cu/OP footprinting which showed that the 5′ half-site was occupied in the A complex with little or no binding to the 3′ half-site.

To assess the size of the proteins in the two protein-DNA complexes, EMSA was performed on size-fractionated Colo-DM nuclear protein. One fraction, corresponding to a protein with an M_r of 37,000 to 38,000 (fraction 28), yielded two specific complexes that were the same size as the A and B complex obtained when using unfractionated Colo-DM extract (Fig. 2c). Of note, there was another, slightly smaller, protein-DNA complex that formed with protein from fraction 30 (lower molecular weight). This complex likely represents a smaller SIF1 binding protein rather than degradation of the protein in the higher-molecular-weight fraction because there was no binding activity in the intervening protein fraction (fraction 29). Therefore, proteins from a narrow molecular weight range were able to form both complex A and complex B on the SIF1 element. Further evidence of the interaction of the proteins forming the two complexes was provided by EMSA experiments with Colo-DM extracts which demon-

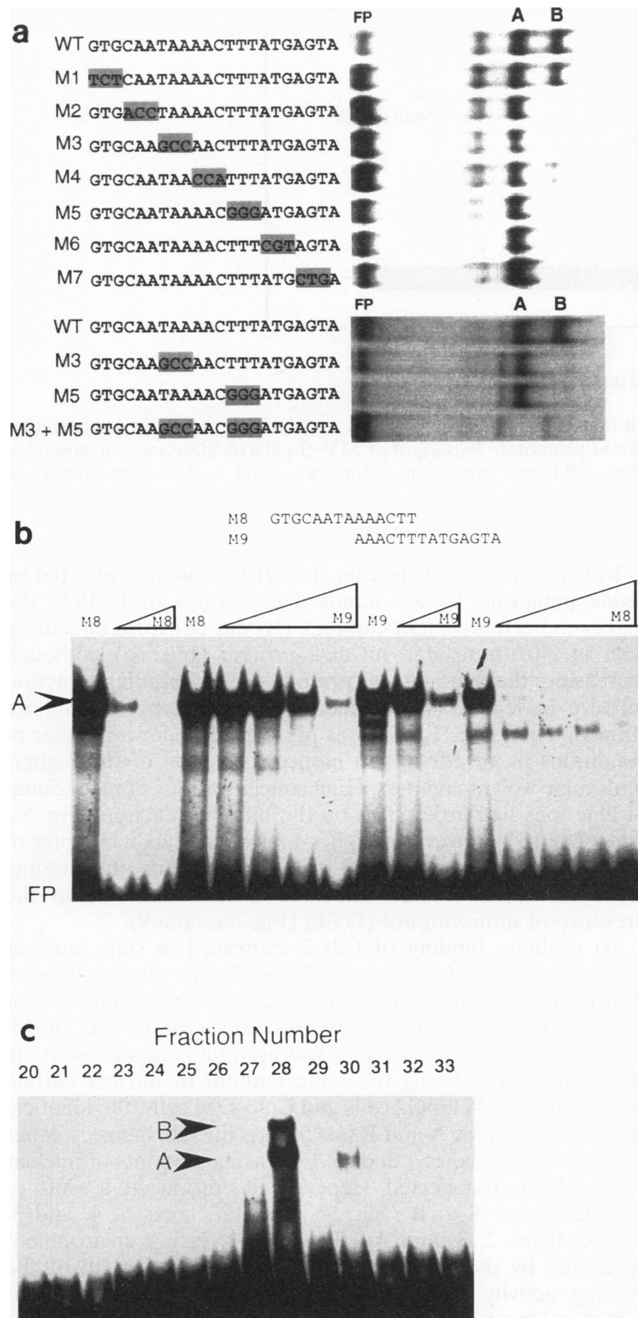

FIG. 2. Characterization of nuclear protein binding to SIF1. (a) EMSA of wild-type SIF1 and mutants. Colo-DM nuclear extracts were used in these assays to assess the effects of mutations in the SIF1 sequence on DNA-protein interaction. (b) EMSA with truncated SIF1 element oligonucleotides. Oligonucleotides that corresponded to the putative half-sites of the SIF1 element were synthesized. M8, 5′ half-site; M9, 3′ half-site. Colo-DM nuclear extract (5 μg) was used for each binding reaction mixture. Both M8 and M9 bound to protein to yield a complex of nearly the same size as complex A formed with the wild-type SIF1 element. Results of competition experiments are shown in the lanes with the triangles, with the competing oligonucleotide indicated in the triangle. The fold increased amounts of competitor were 25-, 50-, 100-, and 200-fold. The M8 oligonucleotide competed for binding to both half-sites much better than the M9 oligonucleotide. The faster-migrating complexes seen with M9 were determined to be nonspecific. The complex immediately below A competed with all added DNA. (c) EMSA of gel-purified fractions of nuclear proteins. Colo-DM nuclear extract was denatured and separated in an SDS-polyacrylamide gel, and the gel was then cut into 50 2.5-mm-width slices. The protein was eluted from each gel piece and analyzed by EMSA. Analysis of fractions 20 through 33 is shown; there was no binding with other fractions. The A and B bands shown in lane 28 were eliminated by incubation with a 100-fold molar excess of unlabeled wild-type SIF1 oligonucleotide (data not shown).

strated that complex B was increased with the temperature of incubation, concentration of nuclear protein, and time of incubation (data not shown).

Taken together, these results of the binding of crude intestinal cell nuclear extracts with the SIF1 element allows proposal of a model for DNA-protein interaction. A protein(s) with an M_r of approximately 38,000 is able to bind preferentially to the 5′ half-site of the SIF1 element. A higher-molecular-weight complex is formed from the occupation of both half-sites of the SIF1 element. This higher-molecular-weight complex is most likely a dimer which could be a homodimer of the same protein that binds preferentially to the 5′ half-site or a heterodimer of two proteins of similar molecular weight.

Functional effect of mutations in the SIF1 element. To examine the functional effect of mutations in the two half-sites of the SIF1 element, mutations corresponding to M3, M5, or M3 plus M5 were introduced into the complete evolutionarily conserved SI promoter (−183 to +54) (Materials and Methods). Transfection studies were performed with Colo-DM cells to compare the results with those of the DNA-binding studies. Results demonstrated that mutation of both half-sites (M3 plus M5) decreased transcription of the SI promoter to 4% of full activity (Fig. 3). Mutation of the 3′ half-site (M5) provides partial function of the promoter, suggesting that the binding of a protein to the 5′ half-site is sufficient for some transcriptional activation. In contrast, mutation of the 5′ half-site (M3) diminishes transcriptional activation to near the levels seen with a double mutant, suggesting that the 3′ half-site alone has little activity. This transfection analysis of mutant SI promoter constructs provides functional corroboration of the proposed model of SIF1-nuclear protein interactions.

Molecular cloning of a SIF1 binding protein. To further analyze the interaction of intestinal nuclear proteins with the SIF1 element, we cloned a SIF1 binding protein from mouse small intestine. The SIF1 sequence was used to screen a mouse small intestinal λgt11 expression library (Materials and Methods). Seven positive clones were identified after 300,000 plaques were screened with a multimerized SIF1 double-stranded oligonucleotide. EMSA using the λ lysogen of these seven clones showed that one (λgt11-3a) had the same pattern of specific binding to the SIF1 element as seen with crude intestinal nuclear extract (data not shown). The sequence of this clone revealed that the 1.37-kb insert contained an initiation codon at nucleotide position 263 resulting in an open reading frame which encoded for a protein of 311 amino acids (Fig. 4a). The sequence upstream of the initiation codon (GCCACCATG) conforms closely to the consensus sequence for translational initiation (7) as well as initiation sites known to have excellent function when tested in vitro (36).

Comparison of the predicted cDNA and protein sequence to the databases at the National Center for Biotechnology Information was performed with the BLAST algorithm. The complete amino acid sequence was 93.6% similar to the recently cloned hamster Cdx-3 (hamCdx-3) (19) (Fig. 4b), suggesting that we isolated the mouse homolog of the hamCdx-3 cDNA.

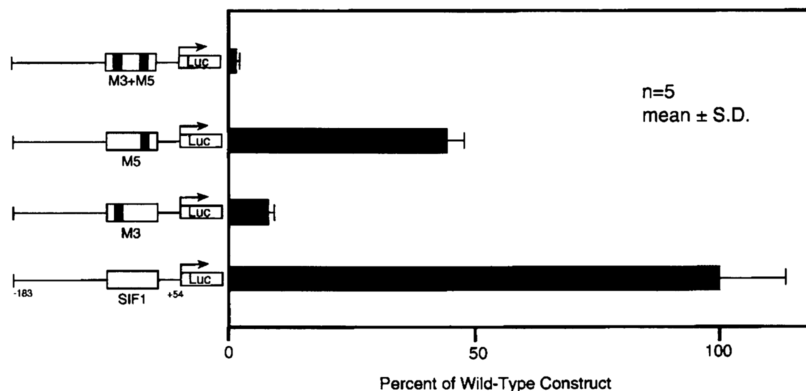

FIG. 3. Transfection analysis of SIF1 mutants. Caco-2 and Colo-DM cell lines were transfected with 8 μg of either the wild-type promoter construct linked to the luciferase reporter gene or the indicated mutants of the SI promoter. Two μg of pCMV–β-galactosidase was cotransfected to control for transfection efficiency (60). Results are expressed as percent wild-type expression, which was 4.04 × 10^5 light units/U of β-galactosidase.

However, the sequence of our clone that encodes a portion of the homeodomain is identical to the portion of a mouse cDNA, called Cdx-2, that was cloned from mouse small intestine by PCR (29). Furthermore, the complete sequence of the mCdx-2 cDNA has recently been cloned, and the predicted amino acid sequence differs from our clone only at residue 69, which is a histidine in our sequence and a tyrosine in the other published sequence (28) (Fig. 4b). Therefore, our clone is mCdx-2. There have been 12 homeobox genes related to caudal cloned from a variety of species, each of which has a high degree of similarity in the sequence of the homeodomain (Fig. 4c). In addition to the homeodomain, there are other regions of the proteins that have been identified as conserved between members of the caudal family (17). The boxed areas labeled A, B, and C in Fig. 4b have been identified as conserved regions between proteins related to caudal (17) and are nearly identical between ham Cdx3 and mCdx-2. Region C contains the hexapeptide, a conserved region which has been used to classify homeodomain proteins, including the family of genes related to caudal (17). It is of interest that this region in both mCdx-2 and hamCdx3 contains a cysteine which is unique within the caudal family. Furthermore, there is a second conserved cysteine which is not present in the other caudal genes.

Characterization of mCdx-2 interaction with the SIF1 element. The mCdx-2 cDNA was cloned into pRc/CMV (Materials and Methods), which served as a vector for in vitro transcription and expression in cell lines. In vitro transcription and translation of the mCdx-2 cDNA produced a protein with an M_r of approximately 37,000 (data not shown), which was very close to the estimated size from the fractionated nuclear extract. However, this size is greater than the predicted molecular weight of the amino acid sequence (M_r = 33,330). Other investigators have noted that homeodomain proteins migrate through SDS-polyacrylamide gels in a fashion indicating a higher molecular weight than the actual size (44). In vitro-translated mCdx-2 protein formed a protein-DNA complex with the wild-type SIF1 element of a mobility identical to that of the A complex formed with Colo-DM nuclear extract (Fig. 5a, lane 2) which was found to be specific in competition experiments with an excess of unlabeled SIF1 (Fig. 5a, lane 3). Competition with mutant oligonucleotides showed that mCdx-2 bound to the M5 mutant (Fig. 5a, lane 5) much more avidly than to M3 (Fig. 5a, lane 4). This is consistent with the finding that Colo-DM nuclear extract binds to the 5′ half-site more avidly than the 3′ half-site (Fig. 2b).

Dimerization of mCdx-2 on the SIF1 element is affected by redox potential. Under standard conditions of EMSA, the higher-molecular-weight complex (B) did not form effectively with in vitro-translated mCdx-2 protein (Fig. 5a), although increasing the amount of protein in the binding reaction mixture did lead to some formation of the B complex (data not shown). Therefore, EMSA was performed under a number of conditions in an attempt to induce formation of the higher-molecular-weight complex. High concentrations of magnesium or zinc ions had little effect on the binding reaction (Fig. 5a, lanes 7 and 8). However, a higher-molecular-weight complex of the same size as complex B was formed when the binding reaction was performed under reducing conditions in the presence of dithiothreitol (DTT) (Fig. 5a, lane 9).

To evaluate binding of Cdx-2 expressed in cells, nuclear proteins were extracted from HepG2 cells that had been transfected with an expression vector for either mCdx-2 or hamCdx-3. The hamCdx-3 expression vector (a gift of M. German, San Francisco, Calif.) was used for comparison to the mouse homolog. Using the same amount of nuclear extract from transfected HepG2 cells and Colo-DM cells, the identical pattern of complex A and B was found (Fig. 5b, hamster, lanes 5 and 2; mouse, lanes 1 and 3). Increasing amounts of nuclear extract from transfected HepG2 cells produced a shift of complex from A to B (Fig. 5b, hamster, lanes 5, 4, and 3; mouse, lanes 2, 3, and 4). Finally, a reducing environment promoted by the addition of DTT shifted nearly all of the binding activity from complex A to B (Fig. 5b, hamster, compare lane 5 with lane 6). Similar results were found for mCdx-2 protein incubated in DTT (data not shown). Taken together, these data demonstrate that mCdx-2 is able to form homodimers on the SIF1 element and that the dimerization is dependent on the redox state of the binding reaction.

Cdx-2 is the major SIF1 binding protein in Colo-DM nuclear extract. Polyclonal antibodies were raised in rabbits to a bacterially expressed fusion protein of GST linked to a portion of the mCdx-2 protein extending from residue 48 to 120 (Materials and Methods). This region of the protein is markedly different between mCdx proteins, and the comparable region was previously used to produce antibodies to mCdx-4 (17). Immunoblot analysis confirmed that the antisera recognized in vitro-translated mCdx-2 as well as a single protein with an M_r of 37,000 in Colo-DM nuclear extract (data not shown). Therefore, there is a single nuclear protein in Colo-DM cells that is the appropriate size for Cdx-2 and that

a

```
                                                50                                                100
GGAATTCCGG TCTACTTTCG TAGGAACAGT TAATGACAGC TTACTGGCAA GGAGGTGGGA GGAAAGAAGG AAGAGAGGGG AGGAGGCAGG ACGGAGGGAG
                                               150                                                200
GGACTGCCCG GGAGGCAGAA GCTCTGCAAG GAGCCGACGG AGCACCGTGG GCTGAGGTGC AGCCAGCTAC CTTTATCTCT AGCCCCCTTC TCCTCCCTCT
                                               250                                                300
GGCAGCCTTC AACGTTTGTC CCCAGACAGC ATGGTGAGGT CTGCTCTGGG TCCCTCGCCA CCATGTACGT GAGCTACCTT CTGGACAAGG ACGTGAGCAT
                                                                    M  Y  V   S  Y  L   L  D  K   D  V  S  M>
                                               350                                                400
GTATCCTAGC TCCGTGCGCC ACTCCGGCGG CCTGAACCTG GCTCCGCAGA ACTTTGTCAG TCCTCCGCAG TACCCGGACT ACGGTGGTTA CCACGTGGCG
 Y  P  S   S  V  R   H  S  G  G   L  N  L   A  P  Q   N  F  V  S   P  P  Q   Y  P  D   Y  G  G  Y   H  V  A>
                                               450                                                500
GCCGCGGCGG CTGCTACGGC GAACTTGGAC AGCGCTCAGT CCCCAGGGCC ATCCTGGCCC ACCGCGCACG GCGCCCCTCT CCGCGAGGAC TGGAATGGCT
 A  A  A   A  A  T  A   N  L  D   S  A  Q   S  P  G  P   S  W  P   T  A  H   G  A  P  L   R  E  D   W  N  G>
                                               550                                                600
ACGCACCCGG GGGCGCTGCG GCAGCCAACG CGGTAGCCCA CGGTCTCAAT GGTGGCTCCC CGGCCGCCGC TATGGGCTAC AGCAGCCCCG CCGAATACCA
Y  A  P  G   G  A  A   A  A  N   A  V  A  H   G  L  N   G  G  S   P  A  A  A   M  G  Y   S  S  P   A  E  Y  H>
                                               650                                                700
CGCGCACCAT CACCCGCATC ATCACCCGCA CCATCCGGCC GCCTCGCCGT CCTGCGCCTC CGGCTTGCTG CAGACGCTCA ACCTCGGCCC CCCGGGGCCC
 A  H  H   H  P  H   H  H  P  H   H  P  A   A  S  P   S  C  A  S   G  L  L   Q  T  L   N  L  G  P   P  G  P>
                                               750                                                800
GCAGCCACCG CCGCCGCCGA ACAGCTGTCC CCCAGCGGCC AGCGGCGAAA CCTGTGCGAG TGGATGCGGA AGCCCGCGCA GCAGTCCCTA GGAAGCCAAG
 A  A  T   A  A  A  E   Q  L  S   P  S  G   Q  R  R  N   L  C  E   W  M  R   K  P  A  Q   Q  S  L   G  S  Q>
                                               850                                                900
TGAAAACCAG GACAAAAGAC AAATACCGGG TGGTGTACAC AGACCATCAG CGGCTGGAGC TGGAGAAGGA GTTTCACTTT AGTCGATACA TCACCATCAG
V  K  T  R   T  K  D   K  Y  R   V  V  Y  T   D  H  Q   R  L  E   L  E  K  E   F  H  F   S  R  Y   I  T  I  R>
                                               950                                               1000
GAGGAAAAGT GAGCTGGCTG CCACACTTGG GCTCTCCGAG AGGCAGGTTA AAATTTGGTT TCAGAACCGC AGAGCCAAGG AGAGGAAAAT CAAGAAGAAG
 R  K  S   E  L  A   A  T  L  G   L  S  E   R  Q  V   K  I  W  F   Q  N  R   R  A  K   E  R  K  I   K  K  K>
                                              1050                                               1100
CAGCAGCAGC AACAGCAGCA GCAGCAACAA CAGCCTCCAC AGCCGCCGCC ACAACCTTCC CAGCCTCAGC CGGGTGCCCT GCGGAGCGTG CCCGAGCCCT
 Q  Q  Q   Q  Q  Q  Q   Q  Q  Q   Q  P  P   Q  P  P  P   Q  P  S   Q  P  Q   P  G  A  L   R  S  V   P  E  P>
                                              1150                                               1200
TGAGTCCTGT GACCTCCTTG CAAGGCTCAG TGCCTGGTTC TGTCCCTGGG GTTCTGGGGC CAGCTGGAGG GGTTTTAAAC TCCACTGTCA CCCAGTGACC
L  S  P  V   T  S  L   Q  G  S   V  P  G  S   V  P  G   V  L  G   P  A  G  G   V  L  N   S  T  V   T  Q>
                                              1250                                               1300
CCTCCCGTGG TCTGAAGCGG CGGCGGCACA GCAATCCCAG GCTGAGCCAT GAGGAGTATG GACGCTGCGA GAATCCTCAG AAGAGATTCC TCTCCTCCTA
                                              1350
CCCACGAACA GCATCTACTG ATGGAGATTG AGGACAGAAG ATGAGTGGAA TTATGGACCT CGGGGCCGGA ATTC
```

b

```
                 A                                                                  ▼  B
                  10         20         30         40         50         60         70         80         90        100
mCdx-2   [illegible]PSSVRH SGGLNLAPQN FVSPPQYPDY GGYHVAAAAA AtANLDSAQS PGPSWpT[illegible]NGY aPGGAAAANA VAHGLNGGSP>
                     ^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^[illegible]^^^ ^^^^^^^^^^ ^^^^^^^^^^
hCdx-3   [illegible]PSSVRH SGGLNLAPQN FVSPPQYPDY GGYHVAAAAA AAANLDSAQS PGPSWST[illegible]NGY PPGGAAAANA VAHGLNGGSP

                 110        120        130        140        150        160   C          180        190        200
         AAAMGYSSPA eYHAHHHPHH HPHHPAAsPS [illegible]ASGLLQTLN lGPPGPAATa AAEQLSPSGQ RRN[illegible]K PAQqSLGSQV KTRTKDKYRV VYTDHQRLEL>
         ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ [illegible]^^^^^^^^^ v^^^^^^^^^ ^^^^^^^^^^ ^^^[illegible]^ ^^^_^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
         AAAMGYSSPA DYHAHHHPHH HPHHPAAAPS [illegible]ASGLLQTLN PGPPGPAATG AAEQLSPSGQ RRN[illegible]K PAQPSLGSQV KTRTKDKYRV VYTDHQRLEL

                 210        220        230        240          250        260        270        280        290
         EKEFHfSRYI TIRRKsELAA TLGLSERQVK IWFQNRRAKE RKIK-KK-QQ QQQQQQQQQp pqPPPQPSQP QPGaLRSVPE PLSPVtSLQG SVPGSVPGVL>
         ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^ ^^^ ^^ ^^^^^^^^^v ^v^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^ ^^^^^^^^^^
         EKEFHYSRYI TIRRKAELAA TLGLSERQVK IWFQNRRAKE RKINKKKLQQ QQQQQQQQQL ASPPPQPSQP QPGSLRSVPE PLSPVSSLQG SVPGSVPGVL

        300        310
         GPAGGVLNsT VTQ>
         ^^^^^^^^^^ ^^^
         GPAGGVLNPT VTQ
```

(Homeodomain residues 185–244, KDKYRV … RKIK / RKIN, are printed in boldface.)

c

```
mCdx-2    KDKYRVVYTDHQRLELEKEFHFSRYITIRRKSELAATLGLSERQVKIWFQNRRAKERKIK
mCdx-1    ---S-------------------Y--------------N---T-----------------VN
mCdx-4    -E---------------------CN------------VN---------------------MI
hamCdx-3  -----------------------Y---------A----------------------------N
rCdx-1    -----------------------Y--------------N---T-----------------VN
CHox-cad  -----------------------Y---------A----A---T-----------------VN
CHox-cad2-E---------------------CN------------VN---------S------------I
Xcad1     ----------Q------------Y---------A---VN-----T----------------N
Xcad2     -----------------------Y---------A----A---T-----------------VN
Zfcad1    -E----------------------N------------VN--------------------ML-
ceh-3     A----M--S-Y------------T-PF--SD---Q-STM-S-T---I-I---------D-RD-
caudal    ----------F--------YCT----------Q---VN-S------------------TSN
```

FIG. 4. (a) mCdx-2 cDNA. Complete sequences of the mCdx-2 cDNA and translated protein which was contained in the p3a clone are shown. The cDNA insert of 1.37 kb was sequenced in both orientations. The polyadenylation signal is not included in this cDNA. (b) Amino acid sequence comparison. The homeodomain is in boldface type. The boxed regions labeled A, B, and C have been identified as regions of conservation between proteins related to caudal (17). Regions A and B show a high degree of similarity between hamCdx-3 and mCdx-2 proteins. Region C contains the hexapeptide, a conserved region which has been used to classify homeodomain proteins, including the family of genes related to caudal (17). This region in both mCdx-2 and hamCdx-3 contains a cysteine which is unique within the caudal family. Furthermore, there is a second conserved cysteine (boxed) which is not present in the other caudal genes. These conserved cysteines may be of significance in the regulation of the binding to the SIF1 element by redox potential (see text). The arrowhead indicates the one residue difference between our sequence and the other reported sequence for mCdx-2 (28). (c) Comparison of homeodomains of proteins related to caudal. Sequences of the following proteins are shown: mouse mCdx-1 (11), mCdx-2 (references 28 and 29 and this paper), and mCdx-4 (6, 17), hamster hCdx-3 (19), rat rCdx-1 (12), chicken CHox-cad (10) and CHox-cad2 (55), *X. laevis* Xcad1 and Xcad2 (3), zebra fish Zfcad1 (30), *C. elegans* ceh-3 (5), and *D. melanogaster* caudal (45).

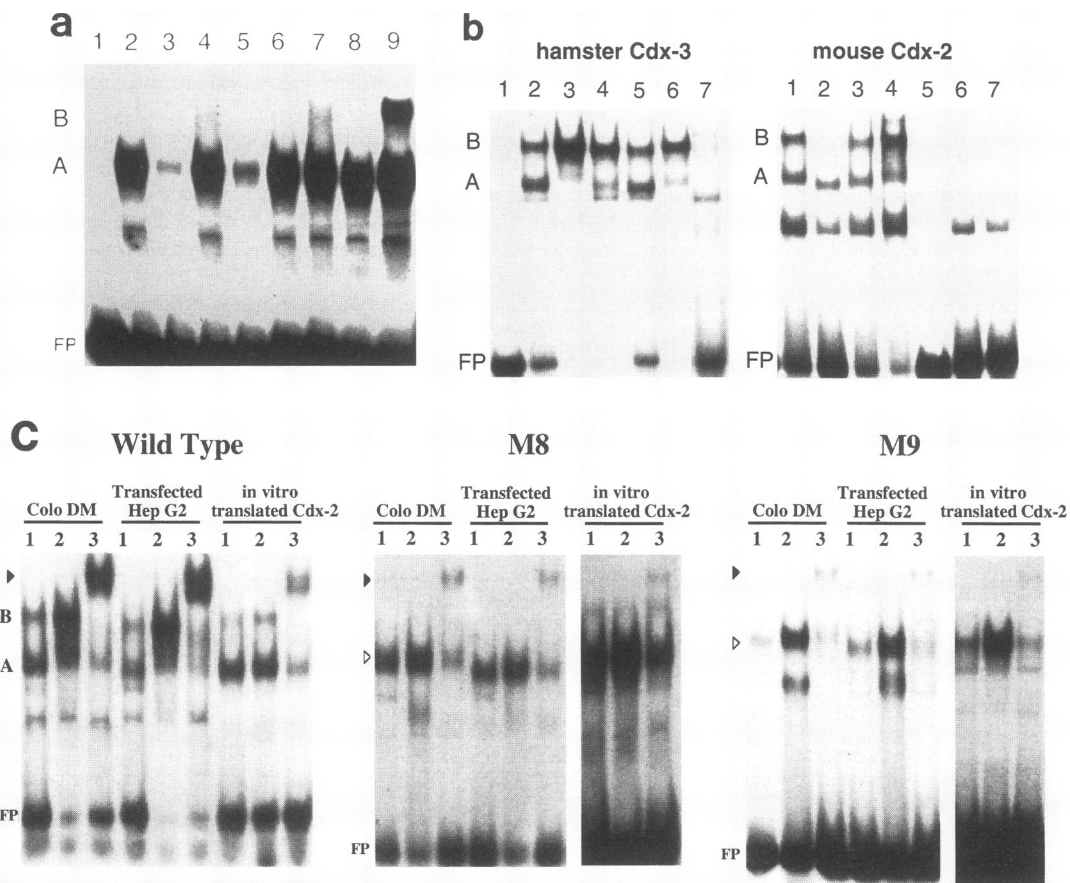

FIG. 5. (a) EMSA with in vitro-translated mCdx-2 protein. All binding reaction mixtures used 5 μg of translation product. Lane 1, wild-type SIF1 probe alone; lane 2, wild-type SIF1 probe plus protein; lane 3, wild-type SIF1 probe plus protein and a 100-fold molar excess of unlabeled wild-type SIF1; lane 4, wild-type SIF1 probe plus protein and a 100-fold molar excess of unlabeled M3; lane 5, wild-type SIF1 probe plus protein and a 100-fold molar excess of unlabeled M5; lane 6, wild-type SIF1 probe plus protein and a 100-fold molar excess of unlabeled double mutant (M3 plus M5); lane 7, wild-type SIF1 probe plus protein and Mg^{2+} to a concentration of 25 mM; lane 8, wild-type SIF1 probe plus protein and Zn^{2+} to a concentration of 50 mM; lane 9, wild-type SIF1 probe plus protein and the addition of DTT to a concentration of 60 mM. The A and B labels refer to the sizes of the complexes found with Colo-DM extract. This was determined from EMSA with both Colo-DM and mCdx-2 on the same gel (data not shown). (b) EMSA with nuclear extract from HepG2 cells transfected with expression vector for hamCdx-3 and mCdx-2. HepG2 cells were transfected with an expression vector for hamCdx-3 (phCdx-3) (a gift from M. S. German and W. J. Rutter [19]) or with mCdx-2, and nuclear protein was extracted 72 h later and used for EMSA. (Left panel) Lane 1, wild-type SIF1 probe alone; lane 2, wild-type SIF1 probe plus 5 μg of Colo-DM nuclear extract; lane 3: wild-type SIF1 probe plus 25 μg of HepG2 nuclear protein extracted following transfection with Cdx-3 expression vector; lane 4, wild-type SIF1 probe plus 10 μg of HepG2 nuclear protein extracted following transfection with Cdx-3 expression vector; lane 5, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein extracted following transfection with Cdx-3 expression vector; lane 6, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein extracted following transfection with Cdx-3 expression vector plus the addition of DTT to a concentration of 60 mM; lane 7, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein. (Right panel) Lane 1, wild-type SIF1 probe plus 5 μg of Colo-DM nuclear extract; lane 2, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein extracted following transfection with Cdx-2 expression vector; lane 3, wild-type SIF1 probe plus 10 μg of HepG2 nuclear protein extracted following transfection with Cdx-2 expression vector; lane 4, wild-type SIF1 probe plus 25 μg of HepG2 nuclear protein extracted following transfection with Cdx-2 expression vector; lane 5, wild-type SIF1 probe alone; lane 6, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein; lane 7, wild-type SIF1 probe plus 5 μg of HepG2 nuclear protein extracted following transfection with pRc/CMV. (c) EMSA with antibodies to Cdx-2. EMSA was performed with the wild-type SIF1 probe (left panel), the 5′ half-site (M8) (middle panel), and the 3′ half-site (M9) (right panel). For each panel, EMSA was performed with Colo-DM nuclear extract, nuclear extract from HepG2 cells transfected with expression vector for hamCdx-3 (similar results obtained for mCdx-3, but not shown on this gel), and in vitro-translated mCdx-2 protein, in the lanes indicated. Five micrograms of Colo-DM nuclear extract or transfected HepG2 extract was used for EMSA. One microliter of the 50-μl T_NT reaction mixture (Materials and Methods) was used for the in vitro-translated Cdx-2 protein. The lanes for each extract are as follows: lanes 1, protein alone; lanes 2, protein plus 1 μl of preimmune rabbit serum; lanes 3, protein plus 1 μl of Cdx-2 antiserum. In these experiments, the antibody was added to the binding reaction mixture before the addition of the radiolabeled oligonucleotide. In other experiments, antibody was added just before loading on the gel, which gave identical results (data not shown). Closed arrows show the positions of supershifted complex as a result of binding of antibodies. Open arrows indicate the protein binding complexes with mutant oligonucleotides (M8 and M9) which have a slightly different mobility than with wild-type SIF1, likely because of the difference in size of the oligonucleotides. The sequences of M8 and M9 are shown in Fig. 2b. FP, free probe.

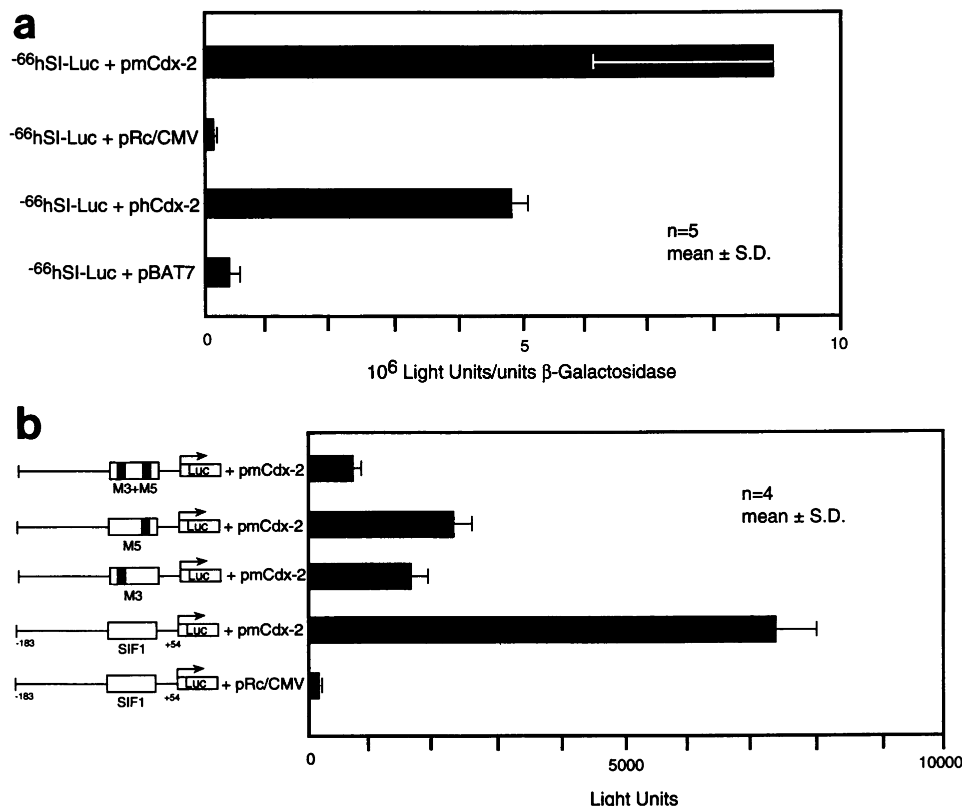

FIG. 6. Activation of SI promoter by hamCdx-3 and mCdx-2. (a) HepG2 cells were transfected with a SI gene expression vector containing bases −66 to +54 of the human SI promoter linked to the luciferase gene (−66h SI-Luc) (60) (4 μg) and with expression vectors for hamCdx-3 and mCdx-2 proteins (8 μg). The hamCdx-3 expression vector contained the hamCdx-3 cDNA in the pBAT7 vector (phCdx-3) (a gift from M. S. German and W. J. Rutter [19]). The mCdx-2 expression vector was made by amplification of the coding sequence of mCdx-2 with oligonucleotide primers containing restriction enzyme sites and a strong translational initiation site on the 5′ end and subsequent subcloning into pRc/CMV (Invitrogen Corp.). The plasmid construct was sequenced to verify the fidelity of the amplification. Two micrograms of pCMV–β-galactosidase was cotransfected to control for transfection efficiency, as previously reported (60). (b) NIH 3T3 cells were cotransfected with the same constructs shown in Fig. 3 and an expression vector for mCdx-2 or the vector alone (pRc/CMV).

reacts with the antibody to mCdx-2. The antisera were then used in EMSA to determine whether Cdx-2 was present in SIF1–Colo-DM nuclear protein complexes (Fig. 5c). By using the wild-type SIF1 element, the antisera supershifted the complexes formed with in vitro-translated mCdx-2 protein, with nuclear extracts from HepG2 cells transfected with either the mCdx-2 or hamCdx-3 expression vectors, and with Colo-DM nuclear extract (Fig. 5c, left panel). Furthermore, both the A and B complexes were shifted with this antibody. We next investigated whether mCdx-2 bound independently to both half-sites of the SIF1 element. The anti-Cdx-2 antibody supershifted in vitro-translated mCdx-2 protein, nuclear extracts from HepG2 cells transfected with either the mCdx-2 or the hamCdx-3 expression vectors, and Colo-DM nuclear extracts bound to truncated oligonucleotides corresponding to SIF1 half-sites (Fig. 5c, middle and right panels). These results demonstrate that the predominant protein binding to the SIF1 element in Colo-DM cells is Cdx-2. Furthermore, it shows that our antibodies recognize the homologous Cdx-2 proteins in mice, hamsters, and humans.

mCdx-2 functions as a transcriptional activator of the SI promoter. To examine the function of Cdx-2 on the SIF1 element in the SI promoter, we performed transient-transfection experiments with HepG2 and NIH 3T3 cells which neither express SIF1 binding protein nor support SI gene transcription (60). A minimal intestine-specific promoter containing only the SIF1 element and the TATA box (bases −66 to +54) was linked to the luciferase gene and used as a reporter construct in cotransfection experiments with an expression vector for either hamCdx-3 or mCdx-2 (Materials and Methods). Coexpression of both proteins with the chimeric SI promoter-reporter gene construct in HepG2 cells demonstrated marked stimulation of transcription of the SI promoter (Fig. 6a). NIH 3T3 cells were also cotransfected with the mutant constructs shown in Fig. 3 and the expression vector for mCdx-2 (Fig. 6b). These data show that transcription of the SI promoter is activated by mCdx-2 and that this activation is dependent on binding to the SIF1 element in the promoter.

mCdx-2 mRNA is expressed exclusively in intestinal tissue. To examine the tissue distribution of mCdx-2, both RNase protection and Northern (RNA) blot assays were performed with RNAs isolated from multiple mouse tissues. These results showed that mCdx-2 mRNA was expressed in all regions of the intestine, including duodenum, proximal and distal jejunum, ileum, cecum, and proximal and distal colon (Fig. 7). This pattern of expression was similar to that previously reported for mCdx-1 and mCdx-2 mRNAs (28, 29). Transcripts for mCdx-2 were not detectable from the other tissues, including stomach, liver, kidney, spleen, brain, heart, lung, pancreas, skeletal muscle, and testis (Fig. 7). The lack of expression in

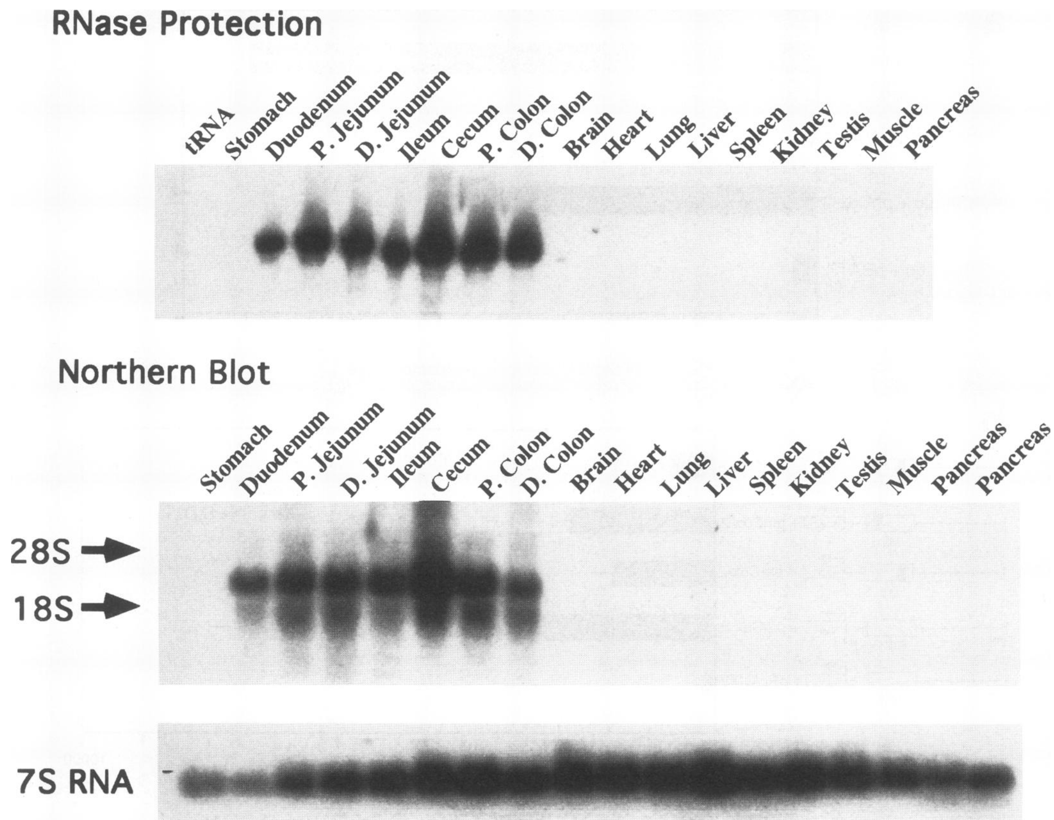

FIG. 7. Tissue expression of mCdx-2. RNase protection and Northern blot assays were performed on 10 μg of total RNA isolated from multiple mouse organs. Probes for both RNase protection and Northern blot were derived from the 3′ untranslated region of the mCdx-2 cDNA to ensure specificity of binding. The Northern blot was stripped of the mCdx-2 probe and rehybridized with a cDNA for a 7S RNA which has previously been used to normalize for expression in intestine tissue (8).

pancreas tissue is notable, since hamCdx-3 was cloned from a pancreatic islet cell line (19). hamCdx-3 mRNA also was not detected in pancreas tissue (19), but this could be due to an inability to identify a small cell population expressing this gene.

DISCUSSION

A member of the homeodomain family related to caudal regulates transcription of the SI gene. The SIF1 element, located in the SI gene promoter, was previously shown to be essential for intestinal cell line-specific transcription (60). We now demonstrate that the homeodomain protein mCdx-2 is able to bind to the SIF1 element and activate transcription of the SI promoter. We also show that hamCdx-3 is the homolog of mCdx-2, as indicated by sequence, DNA-binding characteristics, and function in transcriptional activation. Cdx-2 is a member of a growing family of homeodomain proteins related to the caudal gene (*cad*) of *Drosophila melanogaster* (45). Genes related to caudal have been found in mice (mCdx-1 [11], mCdx-2 [28, 29], and mCdx-4 [6, 17]), Syrian hamsters (ham Cdx-3 [19]), rats (rCdx-1 [12]), chickens (CHox-cad [CdxA] [10] and CHox-cad2 [55]), *Xenopus laevis* (Xcad1 and Xcad2 [3]), zebra fish (Zfcad1 [30]), and *Caenorhabditis elegans* (ceh-3 [5]). In addition to the SI gene, genes related to caudal have functional interactions with regulatory elements of the insulin (19) and fushi tarazu (*ftz*) genes (9). Syrian hamster Cdx-3 has been shown to activate the insulin gene, but its role relative to other transcription factors that also activate the same element is not clear (19). The interaction of *cad* and the segmentation gene *ftz* is important in *Drosophila* development (discussed below).

Interaction of Cdx-2 with the SIF1 promoter element. The interaction of caudal homeodomains with DNA elements has been most closely examined for CHox-cad (CdxA) (39). DNA sequences that interact with CHox-cad were selected from both random oligonucleotides and genomic DNA fragments (39). This analysis showed that the consensus for Chox-cad binding was A/C TTTAT A/G (Fig. 8). Four similar binding sites for the *cad* protein in the *ftz* gene promoter were identified by DNase I footprinting (9) (Fig. 8). Hamster Cdx-3 interacts with the Flat-E and the Flat-F sites in the insulin minienhancer (19), both of which have some similarities to the Chox-cad consensus (Fig. 8). However, the binding of these sites individually has not been carefully investigated. If one considers the binding sites in SIF1 to be seven bases in length, as suggested by the Chox-cad consensus, the two SIF1 half-sites also correspond very well to the consensus (Fig. 8).

There are novel aspects of the binding of Cdx-2 to the SIF1 element that have not previously been reported for the homeodomain proteins related to caudal. First, we found that Cdx-2 was able to bind to the SIF1 element as part of a higher-molecular-weight complex. Experiments with in vitro-translated Cdx-2 protein suggested that this larger complex is a homodimer of Cdx-2. This is consistent with the palindromic nature of the SIF1 element and with the demonstration that Cdx-2 binds independently to both SIF1 half-sites. Interestingly, the binding sites for *cad* in the *ftz* promoter and the

Chox-cad consensus		A/C	T	T	T	A	T	A/G
Ftz promoter	a.	T	T	T	T	A	G	G
	b.	G	T	T	T	A	T	G
	c.	T	T	T	T	A	T	G
	d.	C	T	T	T	A	T	G
Insulin Flat Element								
Flat-E		A	A	T	T	A	G	A
Flat-F		A	T	T	A	A	C	A
SIF1 Element								
3' half-site		C	T	T	T	A	T	G
5' half site (inverse)		T	T	T	T	A	T	T

Ftz promoter elements

```
   a.
T T T T A G G G A A C C A T A A A C
A A A A T C C C T T G G T A T T T G
                          b.

   c.           d.
T T T T A T G T C T T T A T G
A A A A T A C A G A A A T A C
```

Insulin Flat element

```
T G T T A A T A A T C T A A T T A
A C A A T T A T T A G A T T A A T
    Flat-F          Flat-E
```

SIF1 Element

```
C A A T A A A A C T T T A T G A G T A
G T T A T T T T G A A A T A C T C A T
```

FIG. 8. Comparison of the binding sites for caudal and homeodomain proteins related to caudal. The binding sites for proteins related to caudal are shown. The Chox-cad consensus, determined by both random oligonucleotide and genomic DNA selection, is used as a point of comparison for sites in the *ftz*, insulin, and SI promoters. The relationships of the binding sites in intact promoters are also shown.

binding sites in the Flat element are also grouped in closely associated pairs with few intervening bases (Fig. 8). The finding that there is an interaction between Cdx-2 molecules on the SIF1 element adds to the accumulating evidence that protein-protein interactions are involved in homeodomain protein function. Several classes of diverged homeodomain proteins are able to bind to DNA elements as dimers, including hepatocyte nuclear factor 1α and 1β, which homo- or heterodimerize in a DNA-independent fashion (43), and members of the POU-homeodomain family, which require their DNA element for dimerization (27, 52). Recently, it has been shown that proteins of the paired class of homeodomain proteins are able to form homo- and heterodimers and that these dimers bind to DNA in a cooperative fashion (64). Cooperative binding has been demonstrated for the homeodomain of Ultrabithorax to both adjacent and distant sites (2) and for HoxB5 to tandem sites (16). Finally, the interaction of the yeast homeodomain protein α2 with MCM1 leads to cooperative binding (62). Therefore, there are a number of mechanisms by which homeodomain proteins may interact on DNA regulatory elements. The functional importance of the binding of Cdx-2 to closely adjacent sites in the SIF1 element, whether there is true cooperativity of binding, and the nature of the interaction between Cdx-2 proteins will require additional investigation.

The second aspect of Cdx-2 interaction with the SIF1 element that is novel for proteins related to caudal is that the formation of the dimer of Cdx-2 on the SIF1 element is dependent on the redox state of the binding reaction. Redox-related posttranslational modifications affecting DNA-protein binding is a phenomenon that has been increasingly described for transcription factors (1), including homeodomain proteins (16, 54). The cooperative interaction of HoxB5 to tandem sites is dependent on redox potential (16), with an increase in binding upon oxidation. For HoxB5, a cysteine located in the homeodomain was shown to mediate the enhanced binding found with oxidation (16). The effect of redox state on the binding of Cdx-2 is opposite to that of HoxB5, since reduction of the binding reaction enhances Cdx-2 binding to the SIF1 element. In addition, there is no cysteine in the homeodomain of Cdx-2 as there is in HoxB5 (16). Although we have not explored the mechanism of the effect of chemical reduction on protein binding, the presence of two conserved cysteines in the amino-terminal protein domain suggests that this region of the protein plays a role in protein-protein interactions that are regulated by redox potential.

Role of Cdx-2 in the overall regulation of SI gene expression. The SI gene is expressed in complex patterns during development of the small intestine and colon, in intestinal epithelial cell lineages, and along the crypt-villus and cephalocaudal axes of the adult small intestine (reviewed in references 41 and 59). A number of lines of evidence indicate that Cdx-2 is not the sole determinant of the patterns of expression of the SI gene. First, transgenic mouse experiments suggest that sequences outside the SI promoter are important for patterns of SI gene expression. The mouse SI promoter (bases −201 to +54) is able to direct intestine-specific gene transcription in transgenic mice, but there are marked differences in the lineage-specific gene expression from the endogenous gene (40a). Sequences between bases −201 and −8500 are able to partially correct those differences in expression (reference 41 and unpublished data), but there remain important differences from the endogenous gene. Second, Cdx-2 mRNA is expressed in the adult colon, whereas SI is not expressed in adult colon. SI is expressed transiently in fetal human colon (68) and newborn mouse colon (40), suggesting that there are elements in the SI promoter that suppress expression in adult colon or that mCdx-2 is not active in the colon.

More information is required before the complete role of Cdx-2 in the transcriptional regulation of SI can be elucidated. This information includes the cellular and temporal patterns of Cdx-2 expression, the identification of posttranslational modifications that may have an effect on activity and the cells in which those modifications occur, and the identification of transcriptional proteins that may interact with Cdx-2 to affect DNA-binding or transcriptional activity.

Role of homeodomain proteins related to caudal in development. Homeodomain proteins have important roles in the regulation of developmental events in multiple organisms (4, 18, 26, 32). The fact that the Cdx-2 protein plays an important role in transcriptional activation of an intestine-specific gene raises the issue of whether Cdx proteins have a more general role in the regulation of other intestinal genes and in directing intestinal development. The *cad* gene has been shown to have fundamental effects during early developmental events in *D. melanogaster*. *cad* is expressed as an early maternal transcript

and as a zygotic transcript in the posterior cellular blastocyst and then is expressed in proctodeum, Malpighian tubules, hindgut, and posterior midgut (45, 46). Mutation of *cad* results in fusion, or elimination, of posterior body segments, including the hindgut and posterior midgut (38). Early overexpression of *cad* leads to disruption of head development and segmentation (47).

There is indirect evidence of a role for Cdx genes in vertebrate intestinal development. In mice, mCdx-1 mRNA and protein are first expressed in the ectoderm and mesoderm of the primitive streak and in derivatives of these tissues until approximately embryonic day 12 (E12), after which expression appears to be extinguished (44). At embryonic day 14, before the endoderm-intestinal epithelial transition, mCdx-1 is expressed in endoderm and later in intestinal epithelium (44). Mouse Cdx-4 mRNA and protein are expressed early in gastrulation but do not appear to be expressed later in development (17). Expression of genes related to caudal in other species also suggests a function in early gastrulation (3, 14, 17, 30, 44) and endodermal development (10–12, 14, 29). Recently, the developmental expression of CHox-cad (CdxA) protein in chickens has been examined (13). There were some similarities with the expression of Cdx-1 in mice, with expression during gastrulation and subsequent by loss of expression and reexpression in developing endoderm and adult intestinal epithelium (13). The bimodal pattern of expression noted for genes related to caudal during development is consistent with the developmental expression of other homeobox genes. These intriguing patterns of Cdx gene expression in endoderm and intestine suggest a developmental role.

Summary. We have shown that Cdx-2 is important for transcription of a gene that is expressed in terminally differentiated intestinal epithelial cells. Furthermore, the pattern of developmental expression of genes related to caudal suggests a role in intestinal development. There are similarities between these findings and the function of the POU-homeodomain protein Pit-1/GHF-1 in both pituitary development and in differentiated gene expression in mature pituitary cells (reviewed in reference 52). Taken together, these data suggest that the family of caudal homeodomain proteins act as regulators of early developmental transitions in the intestine and in the maintenance of intestinal differentiation in the adult epithelium.

ACKNOWLEDGMENTS

These investigations were supported by funding from the National Institutes of Health (grant RO1 DK46704 to P.G.T.).

We acknowledge the contribution of mouse RNA samples by Arnold Markowitz and the hamCdx-3 expression vector by Michael German. We also thank Mitchell Lazar and Gary Wu for critical review of the manuscript.

REFERENCES

1. **Allen, J. R.** 1993. Redox control of transcription: sensors, response regulators, activators and repressors. FEBS Lett. **332:**203–207.
2. **Beachy, P. A., J. Varkey, K. E. Young, D. P. vonKessler, B. I. Sun, and S. C. Ekker.** 1993. Cooperative binding of an ultrabithorax homeodomain protein to nearby and distant DNA sites. Mol. Cell. Biol. **13:**6941–6956.
3. **Blumberg, B., C. V. E. Wright, E. M. DeRobertis, and K. W. Y. Cho.** 1991. Organizer-specific homeobox genes in xenopus laevis embryos. Science **253:**194–253.
4. **Botas, J.** 1993. Control of morphogenesis and differentiation by HOM/Hox genes. Curr. Opin. Cell Biol. **5:**1015–1022.
5. **Burglin, T. R., M. Finney, A. Coulson, and G. Ruvkun.** 1989. Caenorhabditis elegans has scores of homeobox-containing genes. Nature (London) **341:**239–243.
6. **Candia, A. F., J. Hu, J. Crosby, P. A. Lalley, D. Noden, J. H. Nadeau, and C. V. E. Wright.** 1992. Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. Development **116:**1123–1136.
7. **Cavener, D. R., and S. C. Ray.** 1991. Eukaryotic start and stop translation sites. Nucleic Acids Res. **19:**3185–3192.
8. **Cohen, M. L., N. A. Santiago, J.-S. Zhu, and G. M. Gray.** 1991. Differential regulation of intestinal amino-oligopeptidase gene expression in neonatal and adult rats. Am. J. Physiol. **261:**G866–G871.
9. **Dearolf, C. R., J. Topol, and C. S. Parker.** 1989. The caudal gene product is a direct activator of fushi tarazu transcription during Drosophila embryogenesis. Nature (London) **341:**340–343.
10. **Doll, U., and J. Niessing.** 1993. Continued expression of the chicken caudal homologue in endodermally derived organs. Dev. Biol. **156:**155–163.
11. **Duprey, P., K. Chowdhury, G. R. Dressler, R. Balling, D. Simon, J.-L. Guenet, and P. Gruss.** 1988. A mouse gene homologous to the Drosophila gene caudel is expressed in epithelial cells from the embryonic intestine. Genes Dev. **2:**1647–1654.
12. **Freund, J.-N., R. Boukamel, and A. Benazzouz.** 1992. Gradient expression of cdx along the rat intestine throughout postnatal development. FEBS Lett. **314:**163–166.
13. **Frumkin, A., G. Pillemer, R. Haffner, N. Tarcic, Y. Gruenbaum, and A. Fainsod.** 1994. A role for CdxA in gut closure and intestinal epithelia differentiation. Development **120:**253–263.
14. **Frumkin, A., Z. Rangini, A. En-Yehuda, Y. Gruenbaum, and A. Fainsod.** 1991. A chicken caudal homologue, CHox-cad, is expressed in the epibleast with posterior localization and in the early endodermal lineage. Development **112:**207–219.
15. **Galand, G.** 1989. Brush-border membrane sucrase-isomaltase, maltase-glucoamylase and trehalase in mammals. Comp. Biochem. Physiol. B **94:**1–11.
16. **Galang, C. K., and C. A. Hauser.** 1993. Cooperative DNA binding of the human HoxB5 (Hox 2.1) protein is under redox regulation in vitro. Mol. Cell. Biol. **13:**4609–4617.
17. **Gamer, L. W., and C. V. E. Wright.** 1993. Murine Cdx-4 bears striking similarities to the Drosophila caudal gene in its homeodomain sequence and early expression pattern. Mech. Dev. **43:**71–81.
18. **Gehring, W. J.** 1987. Homeo boxes in the study of development. Science **236:**1245–1252.
19. **German, M. S., J. Wang, R. B. Chadwick, and W. J. Rutter.** 1992. Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes Dev. **6:**2165–2176.
20. **Gordon, J. I.** 1989. Intestinal epithelial differentiation: new insights from chimeric and transgenic mice. J. Cell Biol. **108:**1187–1194.
21. **Gordon, J. I.** 1993. Understanding gastrointestinal epithelial cell biology: lessons from mice with help from worms and flies. Gastroenterology **104:**315–324.
22. **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.
23. **Haffen, K., M. Kedinger, and P. Simon-Assmann.** 1987. Mesenchyme-dependent differentiation of epithelial progenitor cells in the gut. J. Pediatr. Gastroenterol. Nutr. **6:**14–23.
24. **Haffen, K., M. Kedinger, and P. M. Simon-Assmann.** 1989. Cell contact dependent regulation of enterocytic differentiation, p. 19–40. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
25. **Hermiston, M. L., and J. I. Gordon.** 1993. Use of transgenic mice to characterize the multipotent intestinal stem cell and to analyze regulation of gene expression in various epithelial cell lineages as a function of their position along the cephalocaudal and crypt-to-villus (or crypt-to-surface epithelial cuff) axes of the gut. Sem. Dev. Biol. **4:**275–291.
26. **Holland, P. W. H., and B. L. M. Hogan.** 1988. Expression of homeo box genes during mouse development: a review. Genes Dev. **2:**773–782.
27. **Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S.**

Kapiloff, L. Wilson, and M. G. Rosenfeld. 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. Cell **61**:1021–1033.
28. **James, R., T. Erler, and J. Kazenwadel.** 1994. Structure of the murine homeobox gene cdx-2: expression in embryonic and adult intestinal epithelium. J. Biol. Chem. **269**:15229–15237.
29. **James, R., and J. Kazenwadel.** 1991. Homeobox gene expression in intestinal epithelium of adult mice. J. Biol. Chem. **266**:3246–3251.
30. **Joly, J.-S., M. Maury, C. Joly, P. Duprey, H. Boulekbache, and H. Condamine.** 1992. Expression of a zebrafish caudal homeobox gene correlates with the establishment of posterior cell lineages at gastulation. Differentiation **50**:75–87.
31. **Karpinski, B. A., G. D. Morle, J. Huggenvik, M. D. Uhler, and J. M. Leiden.** 1992. Molecular cloning of human CREB-2: an ATF/CREB transcription factor that can negatively regulate transcription from cAMP response element. Proc. Natl. Acad. Sci. USA **89**:4820–4824.
32. **Kessel, M., and P. Gruss.** 1990. Murine developmental control genes. Science **249**:374–379.
33. **Klein, R. M.** 1989. Small intestinal proliferation during development, p. 367–392. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
34. **Klein, R. M.** 1989. Cell proliferative regulation in developing small intestine, p. 393–436. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
35. **Koldovsky, O.** 1989. Development of absorption of monosaccharides, p. 437–449. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
36. **Kozak, M.** 1987. At least six nucleotides preceding the AUG initiator colon enhance translation in mammalian cells. J. Mol. Biol. **196**:947–950.
37. **Kuwabara, M. D., and D. S. Sigman.** 1987. Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: *Escherichia coli* RNA polymerase-lac promoter complexes. Biochemistry **26**:7234–7238.
38. **Macdonald, P. M., and G. Struhl.** 1986. A molecular gradient in early Drosophila embryos and its role in specifying the body pattern. Nature (London) **324**:537–545.
39. **Margalit, Y., S. Yarus, E. Shapira, Y. Gruenbaum, and A. Fainsod.** 1993. Isolation and characterization of target sequences of the chicken CdxA homeobox gene. Nucleic Acids Res. **21**:4915–4922.
40. **Markowitz, A. J., and P. G. Traber.** 1993. Developmental expression of mouse sucrase-isomaltase and a human SI-human growth hormone fusion gene in transgenic mice. Gastroenterology **110**:A634.
40a. **Markowitz, A. J., and P. G. Traber.** Unpublished data.
41. **Markowitz, A. J., G. D. Wu, E. H. Birkenmeier, and P. G. Traber.** 1993. The human sucrase-isomaltase gene directs complex patterns of gene expression in transgenic mice. Am. J. Physiol. **265**:G526–G539.
42. **Menard, D.** 1989. Growth-promoting factors and the development of the human gut, p. 123–150. *In* E. Lebenthal (ed.), Human gastrointestinal development. Raven Press, New York.
43. **Mendel, D. B., L. P. Hansen, M. K. Graves, P. B. Conley, and G. R. Crabtree.** 1991. HNF-1a and HNF-1b (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. Genes Dev. **10**:1042–1056.
44. **Meyer, B. I., and P. Gruss.** 1993. Mouse cdx-1 expression during gastrulation. Development **117**:191–203.
45. **Mlodzik, M., A. Fjose, and W. J. Gehring.** 1985. Isolation of caudal, a Drosophila homeo box-containing gene with maternal expression, whose transcripts for a concentration gradient at the pre-blastoderm stage. EMBO J. **4**:2961–2969.
46. **Mlodzik, M., and W. J. Gehring.** 1987. Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. Cell **48**:465–478.
47. **Mlodzik, M., G. Gibson, and W. J. Gehring.** 1990. Effects of ectopic expression of caudal during Drosophila development. Development **109**:271–277.
48. **Olson, E. N., and W. H. Klein.** 1994. bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev. **8**:1–8.
49. **Pevny, L., M. C. Simon, E. Robertson, W. H. Klein, S.-F. Tsai, V. D'Agati, S. H. Orkin, and F. Costantini.** 1991. Erythroid differentiation in chimaric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature (London) **349**:257–260.
50. **Potten, C. S., and M. Loeffler.** 1990. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons from the crypt. Development **110**:1001–1020.
51. **Quinn, L. A., G. E. Moore, R. T. Morgan, and L. K. Woods.** 1979. Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogenously stained regions. Cancer Res. **39**:4914–4924.
52. **Ruvkun, G., and M. Finney.** 1991. Regulation of transcription and cell identity by POU domain proteins. Cell **64**:475–478.
53. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
54. **Sanchez-Garcia, I., and T. H. Rabbitts.** 1993. Redox regulation of in vitro DNA-binding activity by the homeodomain of the Isl-1 protein. J. Mol. Biol. **231**:945–949.
55. **Serrano, J., L. Scavo, J. Roth, E. J. de la Rosa, and F. de Pablo.** 1993. A novel chicken homeobox-containing gene expressed in neurulating embryos. Biochem. Biophys. Res. Commun. **190**:270–276.
56. **Singh, H., R. G. Clerc, and J. H. LeBowitz.** 1989. Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. BioTechniques **3**:252–261.
57. **Singh, H., J. H. LeBowitz, A. S. Baldwin, and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52**:415–423.
58. **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.
59. **Traber, P. G.** 1994. Differentiation of intestinal epithelial cells: lessons from the study of intestine-specific gene expression. J. Lab. Clin. Med. **123**:467–477.
60. **Traber, P. G., G. D. Wu, and W. Wang.** 1992. Novel DNA-binding proteins regulate intestine-specific transcription of the sucrase-isomaltase gene. Mol. Cell. Biol. **12**:3614–3627.
61. **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.
62. **Vershon, A. K., and A. D. Johnson.** 1993. A short, disordered protein region mediates interactions between the homeodomain of the years a2 protein and the MCM1 protein. Cell **72**:105–112.
63. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight.** 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2**:801–806.
64. **Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan.** 1993. Cooperative dimerization of paired class homeo domains on DNA. Genes Dev. **7**:2120–2134.
65. **Wu, G. D., A. J. Markowitz, E. Birkenmeier, and P. G. Traber.** 1992. A negative transcriptional regulatory element is involved in enterocyte and enteroendocrine cell lineage determination. Clin. Res. **40**:312A.
66. **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.
67. **Wu, G. D., W. Wang, and P. G. Traber.** 1993. Molecular cloning of a transcriptional repressor protein (SIRP-1) which binds to the intestine-specific promoter region of the sucrase-isomaltase gene. Gastroenterology **104**:A290.
68. **Zweibaum, A., N. Triadou, M. Kedinger, C. Augeron, S. Robine-Leon, M. Pinto, M. Rousset, and K. Haffen.** 1983. Sucrase-isomaltase: a marker of foetal and malignant epithelial cells of the human colon. Int. J. Cancer **32**:407–412.