Analysis of the Regulation of the A33 Antigen Gene Reveals Intestine-specific Mechanisms of Gene Expression*

Cameron N. Johnstone‡§, Sara J. White‡, Niall C. Tebbutt‡, Fiona J. Clay‡, Matthias Ernst‡, William H. Biggs§, Carrie S. Viars‡, Suzanne Czekay‡, Karen C. Arden‡, and Joan K. Heath‡‡

From the ‡Ludwig Institute for Cancer Research, Melbourne Branch, Post Office Royal Melbourne Hospital, Parkville, Victoria 3050, Australia and §Ludwig Institute for Cancer Research, San Diego Branch, University of California San Diego, La Jolla, California 92093-0660

The A33 antigen is a transmembrane protein expressed almost exclusively by intestinal epithelial cells. The level of its expression is robust and uniform throughout the rostrocaudal axis of the human and mouse intestines. In the colon, strong expression is found in the basolateral membranes of both the proliferating cells in the lower regions of the crypts and the differentiating cells in the upper regions of crypts. Similarly, in the small intestine, the protein is highly expressed by all the epithelial cells in the crypts and by the differentiated cells migrating over the villi. Thus, the A33 antigen has emerged as a definitive marker for all intestinal epithelial cells, irrespective of cell lineage and differentiation status. To understand the molecular mechanisms mediating this rare tissue-specific expression pattern, we undertook a comprehensive analysis of the 5′-regulatory region of the human A33 antigen gene. This allowed us to point to positive cis-regulatory elements incorporating consensus Krüppel-like factor and caudal-related homeobox (CDX)-binding sites, located just upstream from the human A33 antigen transcription start site, as being important for the intestine-specific expression pattern of this gene. Further analysis provided evidence that the A33 antigen may be one of only a few target genes to be described thus far for the intestine-specific homeobox transcription factor, CDX1. Taken together, our data lead us to propose that the activity of CDX1 is pivotal in mediating the exquisite, intestine-specific expression pattern of the A33 antigen gene.

Efforts to understand the mechanisms underlying tissue-specific patterns of gene expression have focused on the recognition of conserved regulatory sequences within tissue-specific promoters and the identification of transcription factors that are expressed in a tissue-specific manner. In some tissues, the acquisition of specific characteristics appears to be governed by a single tissue-specific transcription factor (or family of transcription factors) responsible for the activation of multiple target genes (1–6). Thus, a number of “master genes” encoding transcription factors that appear to determine the differentiation of the entire tissue have emerged. These include the runt family transcription factor, Cbfa-1, which is required for osteoblast differentiation (1, 2), and the myogenic regulatory network (comprising the helix-loop-helix transcription factors, MyoD, Myf-5, Myf-6, and myogenin) responsible for specifying muscle (3–6). Whether similar mechanisms govern the specialized patterns of gene expression seen in other tissues is currently under intense investigation.

The intestinal epithelium is a complex and dynamic tissue containing migrating cells that undergo phases of proliferation, differentiation, and apoptosis over the course of a few days (7). Such diverse cellular processes probably require transient patterns of gene expression. In many instances, cells positioned in the proliferative compartment in the lower regions of the crypts in the colon and small intestine express different genes compared with cells differentiating in the upper regions of the crypts and those in the small intestine migrating along the villi. The caudal-related family of transcription factors, which includes CDX1 and CDX2, appears to differentially regulate the expression of many genes that contribute to intestinal function (8–15).

To provide further insights into mechanisms employed to maintain intestine-specific patterns of gene expression, we investigated the regulation of expression of the human and mouse genes encoding the intestine-specific A33 antigen (16–18). The human and mouse A33 antigens are type I transmembrane glycoproteins of the immunoglobulin superfamily (17, 18) that share a strikingly similar and highly restricted pattern of expression. In both species, A33 antigen is expressed in epithelial cells throughout the rostrocaudal axis of the small and large intestines (16, 18), where the protein appears to play a role in modulating the gut immune system. In all cells the protein is specifically trafficked to basolateral membranes and is absent from apical membranes (18). An interesting feature of A33 antigen expression is its robust and uniform appearance in both the proliferating immature cells in the lower regions of crypts and the postmitotic, differentiating cells in the upper regions of crypts and throughout the small intestinal villi (16, 18). In addition to strong intestinal expression, mouse A33 antigen (mA33 antigen) expression has also been detected in...
The expression pattern of the A33 antigen is characterized with gradients of gene expression indicated for a number of intestine-specific genes. For the A33 antigen, solid bars indicate uniform expression throughout the intestine, whereas other genes, such as villin and Cdx2, are differentially expressed (8, 19–22). The gradients (Fig. 1A) indicate differential expression along the rostrocaudal axis of the intestine, whereas the vertical bars indicate differential expression depending on the position of the cell along the crypt-villus axis in the small intestine (Fig. 1B) and along the crypt axis in the colon (Fig. 1C).

The expression pattern of the A33 antigen is characterized by robust and uniform expression throughout the intestine. The intestinal (iPabp) and liver (iPabp) fatty acid-binding protein genes have a rostrocaudal gradient of expression with high expression in the small intestine compared with the colon. Both proteins are restricted to the mature epithelial cells in the upper regions of the crypts (B and C) and the vili (B). Other genes, such as villin and Cdx2, are expressed in both the small and large intestines (A), with the highest levels of expression in mature epithelial cells in the upper regions of the crypts and in the small intestinal vili (B and C). Meanwhile, Cdx1, Ca1, and dra (down-regulated in adenoma) are most strongly expressed in the colon, with Cdx1 expression restricted to the epithelial cells residing in the lower, proliferative compartment of the colon (C).

The pattern of intestinal expression exhibited by the A33 antigen appears to be unique among intestine-specific proteins, as shown schematically in Fig. 1. The drawing depicts the small intestine, cecum, and colon represented as a linear tube as shown schematically in Fig. 1. The drawing depicts the small and large intestines (A), and along the crypt axis in the colon (Fig. 1C).

This background prompted us to undertake a study of the human A33 (hA33) antigen gene with the aim of delineating localizations, exon-intron structures, and transcription start sites of both the hA33 and mA33 antigen genes. A comprehensive chromosomal localization, exon-intron structures, and transcription start sites of both the hA33 and mA33 antigen genes. A comprehensive

**EXPERIMENTAL PROCEDURES**

**Cell Lines and Tissue Culture**—All cell lines (SW1222, LS174T, LIM1215, LIM1890, LIM1863, KM12SM, SW480, Caco-2, Caco-2-C2Bbe1, LIM2094, LIM2290, HCT116, LIM2537-Colo526, and U2-OS) were cultured in RPMI 1640 medium supplemented with ADDS (0.025 unit/ml insulin, 1 μg/ml hydrocortisone, and 0.001% thyroxine) and heat-inactivated fetal calf serum (10%; CSL Ltd., Parkville, Australia) in a 5% CO2 humidified incubator at 37 °C. The LIM cell lines were derived in-house from human colorectal cancers (27). All other CRC cell lines plus U2-OS (derived from a human osteosarcoma) and Colo526 (derived from a human uterine carcinosarcoma) are available from the American Type Culture Collection (Manassas, VA).

**Cloning and Characterization of the hA33 Antigen Gene and 5′-Flanking Sequence**—The exon-intron structure of the hA33 antigen gene was assembled from genomic clones, and PCR products were amplified from human genomic DNA. Approximately 10^6 clones of a human placental genomic DNA library in Lambda FIX II (Stratagene, La Jolla, CA) were screened with a 2.6-kb hA33 antigen cDNA probe (17) labeled with [α-32P]dCTP (30 Ci/mmol; Amersham Biosciences) using the Megaprime DNA labeling system (Amersham Biosciences). The inserts from one putative clone (size, 17 kb) was excised using NolI and subcloned into pBluescript KS+ (Stratagene). In addition, PCRs of human genomic DNA with exon-specific primers designed to amplify individual introns were carried out using the EXPAND 20kb™ PCR system (Roche Molecular Biochemicals). Each exon-intron boundary and all of introns 5 and 6 were sequenced on both strands. The 5′-flanking sequences of the hA33 antigen gene, a 488-bp probe was generated by PCR using forward primer 2843 (5′-TTTCCCCAGTGAGCTCTTCTTCTT) and reverse primer 2844 (5′-TTTGATTGGTGTCGTCGTGACT) to amplify a region in intron 1 in clone SW1. One putatively positive clone (28A) was obtained from a screen of the human placental genomic library, and the insert (size, ~17 kb) was excised using NolI and subcloned into pBluescript KS+.

The transcription start site of the hA33 antigen gene was determined using a 5′-RNA ligase-mediated rapid amplification of cDNA ends (RACE) method (GeneRacer; Invitrogen). The GeneRacer strategy aims to amplify only cDNA molecules containing the ultimate 5′ terminus. Briefly, 5 μg of total RNA extracted from both LIM1215 cells and normal human colon epithelium with calf intestinal phosphatase to remove the 5′-terminal anchor sequence from all RNA molecules without a 5′-methyl guanosine cap structure. The RNA was then treated with tobacco acid pyrophosphatase to remove all the caps and leave free 5′-phosphates, which were ligated to synthetic RNA adapter molecules provided in the kit. Human A33 antigen gene-specific antisense primer, FJ65 (5′-GAGTGTAAACAACACAGGCCAAGATCGCTTCTCTC) was used with THERMOScript RT (Invitrogen) at 65 °C to generate cDNA strands complementary to full-length mRNA molecules. These were amplified in touchdown PCRs using PLATINUM Taq DNA polymerase High Fidelity (Invitrogen) to generate products containing the transcription start site. The PCR products were gel-purified and cloned into the pCR 4-TOPO cloning vector (Invitrogen), and multiple clones were sequenced on both strands.

**Cloning and Characterization of the mA33 Antigen Gene and 5′-Flanking Sequence**—A 217-bp PCR product encompassing partial gene of the V87-domain of mA33 antigen was generated using forward primer 1865 (5′-TACGAAAAGAAATACATC) and reverse primer 1866 (5′-TCTGGCTTGGAGGTTG), radiolabeled, and used to screen 12.5×10^6 clones of a 129/SvJ mouse genomic library in the Lambda FIX II vector (Stratagene). The inserts from three putatively positive clones (1A, 4A, and 7B; size range, 16–18 kb; Fig. 2) were excised using NolI and subcloned into pBluescript KS+. The distribution of exons was determined by Southern analysis using mA33 antigen exon-specific oligonucleotide probes. To confirm A33 antigen gene expression was determined using Southern analysis with mA33 antigen cDNA confirming the expression pattern. The orientation of the A33 antigen gene was determined by genomic DNA, approximately 10^6 clones of a 129/SvJ mouse genomic library in the Lambda FIX II vector.

**Hybridization Analysis**—To test for the presence of PCRs products were generated using forward primer 3479 (5′-GCAACACCGACCAGACAGCC) and reverse primer 3480 (5′-TGCACAGAGCATC-AGGCATTTACCTG) using the PCRs were gel-purified and cloned into the pCR 4-TOPO cloning vector (Invitrogen), and multiple clones were sequenced on both strands.

**Cytogenetic Localization of the Human A33 Antigen Gene**—Using fluorescence in situ hybridization, the entire hA33 antigen genomic clone, SW1, was nick-translated using biotin-14 dATP and biotin-14 dCTP (Invitrogen) and hybridized to normal human metaphase spreads in two independent experiments. Chromosome preparations were obtained from phytohemagglutinin-stimulated normal human peripheral blood lymphocytes and cultured for 72 h. To induce R-banding, some of the cultures were synchronized with thymidine for 48 h, incubated overnight at 37 °C, and treated with 5-bromodeoxyuridine the next morning, during the final S phase, and harvested 6 h later (28).

**Cytogenetic harvests and slide preparations were performed using standard methods. Fluorescence in situ hybridization to metaphase chromosomes was performed as described previously (29), with a few modifications. The biotin-labeled probe (100 ng), 0.1 μg of Cot-1 DNA, and 0.5 μg of herring sperm DNA were dissolved in DenHyb hybridization mixture (Sigma Chemical Co.) and denatured at 72 °C for 10 min to...
allow the Cot-1 DNA to anneal to repetitive sequences in the probe. The probe mixture was then applied to the slide and co-denatured with the metaphase spreads for 10 min at 82 °C on a slide warmer. Hybridization was allowed to proceed for a minimum of 24 h in a 37 °C incubator. The slides were washed at 37 °C in 50% formamide for 18 min. Biotin-labeled probe detection was accomplished by incubation with the fluorescein isothiocyanate-avidin conjugate (Intergen, Purchase, NY). Chromosome identification was performed by simultaneous hybridization with a chromosome 1-specific α-satellite repeat probe (Intergen) or by R-handing using 5-bromodeoxyuridine and mounting the slides in a modified anti-fade (p-phenylenediamine, pH 11, containing 0.01 μg/ml propidium iodide as counterstain) to produce an R-handing pattern (30).

Chromosomal Localization of the mA33 Antigen Gene—A 2.1-kb mA33 antigen cDNA probe (18) was used in Southern analyses to identify mA33 antigen restriction fragment length polymorphisms between C57BL/6J and Mus spretus mouse genomic DNAs digested with a number of restriction enzymes. Genetic mapping was achieved using a (C57BL/6J × Mus spretus)F1 × Mus spretus backcross (BSS panel) generated and distributed by the Jackson Laboratory (Bar Harbor, ME). Genomic DNA from 94 backcross animals was digested with BamHI, EcoRV, PstI, SacI, and ScaI, electrophoresed in 1% agarose gels, and transferred to Hybond N+ (Amersham Biosciences). Membranes were hybridized with the mA33 antigen cDNA probe, and the segregation of mA33 antigen alleles was analyzed. Gene linkage and order were determined using the MapManager v2.6.5 program (31).

**RNA Preparation and Northern Blot Analysis**—Total RNA was prepared from cell lines either by the method of Chomczynski and Sacchi (32) or by using Trizol reagent (Invitrogen) according to the manufacturer’s instructions. RNA pellets were dissolved in autoclaved H2O treated with 0.1% (v/v) diethylpyrocarbonate (Sigma Chemical Co.). Gels were run in 0.25% agarose mini-gels in a Hoefer miniVE vertical electrophoresis system (Amer- sham Biosciences) and the Klenow fragment of DNA polymerase (New England BioLabs) was used to generate a series of nested fragments of the 5′-flanking sequence. The inclusive positions of the oligonucleotide primers are described relative to the initiation of translation, where A in ATG is designated 105 for Colo526, U2-OS, and 293T. The next morning, cells (10^5) in 1 ml of ice-cold phosphate-buffered saline (1× phosphate-buffered saline + 1 mM EDTA), and centrifuged. The cell pellet was resuspended in 1 ml of phosphate-buffered saline + 1 mM EDTA, re-pelleted, resuspended in 400 μl of ice-cold hypotonic buffer A (10 mM Hepes, pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 μM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, and 1× complete protease inhibitor mixture (Roche Molecular Biochemicals)), and incubated on ice for 15 min. Cells were lysed by the addition of 25 μl of Nonidet P-40 detergent (final concentration, 0.6% (v/v)) and vortexed vigorously for 10 s. The nuclei were pelleted by centrifugation, and nuclear protein was extracted by resuspension in 50 μl of ice-cold buffer C (20 mM Hepes, pH 7.9, 0.4 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, and 1× complete protease inhibitor mixture) and incubated on ice for 15 min. The samples were centrifuged, and the supernatant containing nuclear proteins was collected and stored at −70 °C. Protein concentration was determined using the Bradford Assay (Bio-Rad, Hercules, CA).

**Western Blot Analysis**—20 μg of nuclear protein extract were subjected to Western blot analysis as described previously (18). Human CDX1 was detected after overnight incubation at 4 °C with a 1:200 dilution of affinity-purified rabbit anti-mouse CDx1 polyclonal antisemur (33), followed by incubation (60 min, 25 °C) in horseradish peroxidase-conjugated goat anti-rabbit immunoglobulin G at a 1:10,000 dilution (Zymed Laboratories, South San Francisco, CA). Interactions were visualized by enhanced chemiluminescence (Amersham Biosciences).

**Electrophoretic Mobility Shift Assays**—Nuclear protein extract (10 μg) was added to binding solution (10 mM Hepes, pH 7.9, 62.5 mM KCl, 4 mM MgCl2, 0.1 mM EDTA, 0.25 mM dithiothreitol, 100 ng/ml bovine serum albumin, 12.5% glycerol, and 3 μg of polydeoxyinosine-deoxyctyo- tose (Sigma Chemical Co.)), centrifuged briefly, and incubated for 15 min at room temperature. 32P-labeled double-stranded oligonucleotide probes were generated using [32P]dATP and [32P]dCTP (Amersham Biosciences) and the Klenow fragment of DNA polymerase (New England Biolabs). The wild-type probe (5′-GATCCTCCCCCTTTTTATATTTT-GGAT) encompassed the more proximal putative CDX1/2-binding site (shown in bold) in the hA33 antigen promoter (Fig. 4), whereas the mutant probe (5′-GATCCTCCCTGGGATATTTTGGAT) contained three nucleotide changes in the CDX1/2-binding site. 12,000 dpm of probe (3,000 dpm/nl) were incubated with the samples for an additional 15 min at room temperature, mixed with 4 μl of loading buffer (40% glycerol and 20 mM dithiothreitol), and loaded onto 6% polyacrylamide mini-gels in a Hoefer miniVE vertical electrophoresis system (Amersham Biosciences) that had been pre-electrophoresed for 0.5 h at 100 V. Gels were run in 0.25× Tris borate/EDTA (TBE) buffer (1 TBE = 100
Antigen Genes—Both the human and mouse A33 antigen genes comprise seven exons spanning ~38 and 35 kb of genomic DNA, respectively (Fig. 2). Both the V-type and C2-type immunoglobulin-like domains of the A33 antigen extracellular domain are encoded by two exons (half domain exons; Fig. 2), as has been shown for several other members of the immunoglobulin superfamily including MUC18 and N-CAM (34, 35). Conservation of features (exon/intron organization, phase of splicing, and presence of a novel pair of cysteine residues) among hA33/mA33 antigens, CTX (marker of cortical thymocytes in Xenopus), and CAR (receptor for Coxsackie group B viruses and adenoviruses types 2 and 5) has defined a new family of cell differentiation fragment length polymorphisms between C57BL/6J and M. spretus (data not shown) and independently confirmed and refined by genetic mapping. Restriction fragment length polymorphisms between C57BL/6J and M. spretus were identified in genomic DNA digested with various restriction enzymes. Informative PsI restriction fragment length polymorphisms were identified in 88 of 94 backcross animals. C, the haplotype for chromosome 1 in the region of the mA33 antigen gene. Black boxes indicate inherited C57BL/6J alleles, whereas open boxes indicate inherited M. spretus alleles. R, recombination distance in centimorgans (cM), SE, S.E. of the mean. D, partial map of chromosome 1 in the region of the gene encoding the mA33 antigen.

RESULTS

Cloning and Characterization of the Human and Mouse A33 Antigen Genes—Both the human and mouse A33 antigen genes contain a 3′ untranslated region, extending by 51 bp the 5′-untranslated region, respectively. To determine the transcription start sites of both genes, we carried out 5′-RNA ligase-mediated-RACE-PCR. Our results demonstrated that in human, a major transcription start site does indeed occur 345 bp upstream of the initiation of translation (Table II), corresponding exactly with the 5′ end of the longest cDNA clone (clone 18) (17). In addition, a number of other more downstream, transcription start sites were also identified by our analysis (Table II). These are likely to represent true minor transcription start sites because we conducted the reverse transcription reaction at 65 °C, thereby avoiding a technical limitation of the method employed (41). In mouse, we also obtained evidence for multiple start sites (Table II), with the major and most 5′ start site corresponding to a position 153 bp upstream of the initiation of translation, extending by 51 bp the 5′ end we had previously obtained using conventional 5′-RACE-PCR (18). These data suggest that the transcription start sites in the human and mouse genes do not correspond precisely.

Cloning and Sequencing of the 5′-Flanking Sequence of the mA33 Antigen Gene—Given the extremely similar expression to the distal region of mouse chromosome 1 (Fig. 3C). The most likely gene order positions mA33 antigen 1.10 ± 0.9 cm distal to PmX1 (37, 39) and 8.87 ± 5.57 cm proximal to Apoa2 (Fig. 3D) (37, 40). This region is syntenic with the human chromosomal region 1q23-q25, thereby providing compelling evidence that the human and mouse genes cloned in this study are orthologs.

Determination of the Transcription Start Sites of the Human and Mouse A33 Antigen Genes—Our previous experiments (17, 18) had revealed that the human and mouse genes contained at least 345 and 102 bp of 5′-untranslated region, respectively. To determine the transcription start sites of both genes, we carried out 5′-RNA ligase-mediated-RACE-PCR. Our results demonstrated that in human, a major transcription start site does indeed occur 345 bp upstream of the initiation of translation (Table II), corresponding exactly with the 5′ end of the longest cDNA clone (clone 18) (17). In addition, a number of other more downstream, transcription start sites were also indicated by our analysis (Table II). These are likely to represent true minor transcription start sites because we conducted the reverse transcription reaction at 65 °C, thereby avoiding a technical limitation of the method employed (41). In mouse, we also obtained evidence for multiple start sites (Table II), with the major and most 5′ start site corresponding to a position 153 bp upstream of the initiation of translation, extending by 51 bp the 5′ end we had previously obtained using conventional 5′-RACE-PCR (18). These data suggest that the transcription start sites in the human and mouse genes do not correspond precisely.
patterns of the hA33 and mA33 antigens \textit{in vivo}, it might be expected that the regulatory sequences critical for controlling tissue-specific expression would be conserved between the human and mouse promoters, as had previously been observed for the sucrase isomaltase (SI) promoter (42). The hA33 antigen promoter sequence was compared with the corresponding sequence in the mA33 antigen gene (Fig. 4) and found to be 65% identical, corresponding closely with the level of identity (67%) between the coding sequences in the two genes. There are two canonical TATA (TATAAA) box sequences in the human gene that lie 12 and 224 nucleotides upstream from the most 5'-transcription start site, but these are not conserved in mouse. Furthermore, because neither TATA box lies in a region 25–30 bp upstream of the start site, we infer that the initiation of transcription is not dependent on TATA sequences. However, both genes contain a conserved sequence, TCAGTTT (shown in brackets, Fig. 4), which closely corresponds to the consensus transcription initiator (Inr) sequence (Py Py A, N T/A Py Py) described by Lo and Smale (43), and a closely spaced conserved sequence GGACTTTG (Fig. 4), which corresponds to a consensus downstream promoter element (44). The configuration of these elements is consistent with them playing a role in the initiation of transcription.

The promoter region alignment revealed several conserved binding sites for transcription factors previously implicated in the regulation of gene expression in intestinal epithelial cells (Fig. 4). A potential binding site for the gut and intestine enriched Krippel-like factors (GKLF/KLF4 and IKLF/KLF5) (23, 45) and two putative CDX-binding sites (24) were perfectly conserved in the human and mouse A33 antigen genes (Fig. 4).

**Characterization of the 5'-Flanking Sequence of the hA33 Antigen Gene**—The availability of a panel of human CRC-derived cell lines either positive (+) or negative (−) for hA33 antigen expression (Ref. 17; see also Fig. 6C) provided us with a suitable model system with which to analyze the regulation of the hA33 antigen promoter in transient transfection assays. A schematic diagram depicting the various constructs and some of the consensus transcription factor-binding sites is shown in Fig. 5A. The longest fragment tested (2.3 kb) exhibited substantial promoter activity (black bars) in the two hA33 antigen (+) CRC-derived cell lines examined (SW1222 and LS174T) but was inactive in the hA33 antigen (−) CRC-derived cell line, SW480, and the three hA33 antigen (−) cell lines derived from nonintestinal tissues (U2-OS, 293T, and Colo526; Fig. 5B). All six cell lines supported the activity of the SV40 promoter and enhancer (Fig. 5B). This result suggested that all the cis-acting elements required to recapitulate endogenous hA33 antigen transcription in vitro may be contained in the −2.3-kb region.

The promoter activities of a series of nested fragments of A33-luciferase constructs (−2.3, −1.2, −0.6, and −0.44 kb) were then examined (Fig. 5C). All four constructs assayed were active in the four hA33 antigen (+) CRC cell lines (SW1222, LS174T, LIM1215, and KM12SM) and inactive in the four hA33 antigen (−) cell lines (SW480, LIM2099, Caco-2, and Colo526). Over a series of experiments \(n = 4\), there were no reproducible significant differences in activity among the four constructs within any individual cell line. Compared with the −2.3-kb construct, the −1.2-kb construct lacked a region of 5' flanking sequence containing the consensus HNF1- and e-MYB-binding sites (Fig. 5A). These data imply that the HNF1 and e-MYB sites do not greatly influence transcription of the hA33 antigen gene and do not play a significant role in the differential hA33 antigen expression in CRC-derived cell lines.

The identification of putative cis-acting elements was refined further by analyzing the activities of a final series of deletion constructs. The −0.44-, −0.42-, −0.41-, −0.38-, and −0.36-kb constructs were all active in hA33 antigen(+) cell lines (SW1222 and LS174T) and inactive in the hA33 antigen(−) cell lines (SW480 and Caco-2; Fig. 5D). Deletion of 21 bp from the −0.44-kb construct (yielding a −0.42-kb construct) removed the potential GFLK/IKLF-binding site (45) and resulted in a consistently reproducible 3-fold decrease in reporter gene activity in SW1222 cells and a consistently reproducible 2-fold decrease in reporter gene activity in LS174T cells (Fig. 5D). Deletion of an additional 17 bp from the −0.42-kb construct (yielding the −0.41-kb construct) removed the more distal of two consensus binding sites for CDX1/CDX2 (24) and resulted in a further consistent and reproducible 2-fold decrease in activity in SW1222 and LS174T cells. Removal of 21 bp from the −0.41-kb construct (yielding the −0.38-kb construct), a region that lacks potential transcription factor-binding sites, did not produce an appreciable change in reporter gene activity in either SW1222 or LS174T cells. Deletion of an additional 21 bp from the −0.38-kb construct (yielding the −0.36-kb construct), resulting in removal of the more proximal consensus CDX1/CDX2-binding site, produced an 6-fold decrease in reporter gene activity in SW1222 cells and an 4-fold decrease in reporter gene activity in LS174T cells. Taken together, these data strongly implicate the consensus GFLK/KLF4-binding site and both the distal and proximal consensus CDX-binding sites in regulating the transcriptional activity of the hA33 antigen gene in CRC-derived cells \textit{in vitro}. Finally, deletion of 141 bp from the −0.36-kb construct (producing the −0.22-kb construct) removed the major, most 5’ transcription start site of the hA33 antigen gene, resulting in negligible activity in the hA33 antigen(+) cells (Fig. 5D).

**Expression Pattern of CDX1, CDX2, GFLK/KLF4, and IKLF/KLF5 in CRC-derived Cell Lines**—The mRNA expression patterns of CDX1 and CDX2 and the two KLFs expressed in intestinal epithelium, GFLK/KLF4 and IKLF/KLF5 (46, 47), were determined in the CRC-derived cells and correlated with the distribution of hA33 antigen mRNA. In addition, the expression pattern of CDX1 protein was examined. Strikingly, five hA33 antigen(+) cell lines (SW1222, LS174T, LIM1215, LIM1899, and LIM1863) expressed CDX1 protein and mRNA, and five hA33 antigen(−) cell lines (SW480, Caco-2, LIM2405, HCT116, and LIM2537) did not (Fig. 6C–E). CDX2 mRNA was expressed by all hA33 antigen(+) cell lines examined (SW1222, LS174T, LIM1215, LIM1899, KM12SM, and LIM1863) and also by two of the hA33 antigen(−) lines (Caco-2 and SW480; Fig. 6A). GFLK/KLF4 mRNA appeared to be weakly expressed by all CRC-derived cell lines examined (Fig. 6B). As observed for GFLK/KLF4, IKLF/KLF5 mRNA was weakly expressed by all hA33 antigen(+) cell lines studied (SW1222, LS174T, LIM1215, LIM1899, and LIM1863) as well as the hA33 antigen(−) lines (SW480, Caco-2, LIM2405, and HCT116; data not shown). Thus, our analysis revealed that CDX1 was the only transcription factor of those tested whose expression pattern in CRC-derived cell lines correlated with that of the hA33 antigen.

**CDX1 Binds to a cis-Regulatory Element in the A33 Antigen Gene**—To determine whether CDX1 could interact with con-

---

**Table II**

| Position of transcription start sites in human and mouse A33 antigen genes |
|-----------------------------|-----------------------------|
| **Human**                  | **Mouse**                  |
| Most 5' start site          | Downstream sites            |
| −345                       | −326, −306, −121            |
| −153                       | −143, −113, −105            |
sensus CDX-binding sites contained within the putative cis-regulatory region of the hA33 antigen gene promoter, we performed electrophoretic mobility shift assays using nuclear extracts from transfected 293T cells. Using a probe encompassing the more proximal of the two CDX-binding sites, a prominent retarded complex was produced by nuclear extracts from 293T cells transfected with mouse Cdx1 (Fig. 7, lane 5). This band was absent when the extract was incubated with a mutated probe (Fig. 7, lane 6), and its intensity was markedly diminished when a 25-fold excess of unlabeled probe was added (Fig. 7, lane 7). When an antibody raised against mouse Cdx1 was included in the incubation, the intensity of the band corresponding to the putative CDX1-DNA complex was markedly diminished, and a proportion was supershifted (Fig. 7, lane 8). This analysis strongly suggests that CDX1 binds specifically to the regulatory region in the A33 antigen gene defined here.

**DISCUSSION**

We undertook a comprehensive characterization of the human A33 antigen gene to shed light on the mechanisms regulating its intestine-specific expression pattern. We had demonstrated previously that human CRC-derived cell lines may be either positive (+) or negative (−) for hA33 antigen expression (17) (Fig. 6C), even though normal intestinal epithelium and most primary CRC tumors are strongly positive. In our study, we assumed that the group of six cell lines that express the A33 antigen (SW1222, LS174T, LIM1215, LIM1899, KM12SM, and LIM1863) had retained expression of all the molecules that normally play a role in regulating A33 antigen expression in vivo, whereas the non-A33 antigen-expressing cell lines (Caco-2, LIM2099, LIM2405, HCT116, SW480, and LIM2537) had not. Accordingly, such cell lines provided us with a suitable model system with which to analyze the highly restricted ac-

![Alignment of sequences upstream of the initiation of translation in the human and mouse A33 antigen genes.](http://www.jbc.org/)

The sequences were aligned with the MegAlign program (DNAStar Inc.) using the Clustal algorithm with a PAM250 matrix followed by manual adjustment. The initiation of translation codons (ATG) are in **open boxes**, and the individual nucleotides corresponding to the most 5′ transcription start sites we detected for each of the two genes are in **black boxes**. Conserved nucleotides are shaded **gray**. The hA33 and mA33 antigen gene promoters share 65% nucleotide sequence identity over the region shown. The hA33 antigen gene contains a 25-bp tandem repeat, indicated by the forward and reverse **dotted arrows** between positions −198 and −149 (relative to the initiation of translation), that is not conserved in mouse. The 5′ ends of the nested fragments of the hA33 antigen gene promoter are also indicated with **forward arrows**. All terminate with the same 3′ end, indicated by the **reverse arrow** after position −36. The initiator (Inr) and downstream promoter elements that may contribute to the initiation of transcription are indicated by the **bracketed area** and the **dashed box**, respectively, between the −0.36- and −0.22-kb constructs. The conserved consensus gut/intestine-enriched Krüppel-like factor-binding site located between the beginning of the −0.44- and −0.42-kb constructs is highlighted in **black**. The two conserved consensus CDX-binding sites located between the −0.42- and −0.36-kb constructs are **boxed**.
activity of the hA33 antigen promoter, and we believe that our in vitro findings may be highly relevant to the regulation of A33 antigen expression in intestinal epithelial cells in vivo. The results of our dissection of the hA33 antigen promoter demonstrate that −400 bp of the 5′-flanking sequence (−443 to −36 by reference to the translation start site) are capable of recapitulating endogenous hA33 antigen gene expression in CRC-derived cells. Further refinement of this region points to a consensus GKLF/IKLF-binding site and two consensus CDX1/2-binding sites as important positive cis-acting elements in the hA33 antigen gene. The perfect conservation of these elements in the mA33 antigen gene adds weight to this concept, and data generated by electrophoretic mobility shift assay demonstrated that mCdx1 was capable of binding to the proximal CDX1/2-binding site in the regulatory region. Moreover, our analysis of mRNA expression of the relevant transcription factors in 10 CRC cell lines suggests that CDX1 may be absolutely required for expression of the hA33 antigen gene, at least in vitro. Furthermore, the results we obtained with SW480 cells, which express mRNAs encoding CDX2, GKLF/KLF4, and IKLF/KLF5 but not CDX1, indicated that these three transcription factors, acting alone or in combination, are insufficient to drive hA33 antigen reporter gene expression in vitro.

The clustering of critical regulatory elements within 100 bp 5′ of the major transcription start site in the hA33 antigen gene suggests that this region may comprise the core promoter of the TATA-less hA33 antigen gene. Inspection of CDX-binding sites in many other known target genes expressed by intestinal epithelium reveals very similar findings. Such genes include SI (48, 49), lactase-phlorizin hydrolase (LPH) (9–11), guanylyl cyclase C (GC-C) (12, 13), claudin-2 (26), CA1 (50, 51), and proglucagon (52). In all cases, the promoters contain CDX-binding sites within a region 25–110 bp upstream of the transcription start site. Therefore, our studies, along with those of others, reinforce the notion that CDX transcription factors play a pivotal role in determining the expression patterns of many intestine-specific genes.

Comparing the expression patterns of the mA33 antigen and Cdx1 genes in vivo reveals that mA33 antigen and Cdx1 expression are closely coordinated during intestinal development. Expression of Cdx1 first appears in the distal hindgut endoderm at E13.5, −24 h before the onset of mA33 antigen expression in exactly the same region (53, 54). Meanwhile, Cdx2 expression is found exclusively in the proximal midgut endoderm at E13.5 (53). The Cdx1 expression domain then moves proximally and is distributed universally throughout
the gut by E14.5 (53). The observation that the expression of the mA33 antigen precisely recapitulates the expansion of the Cdx1 expression domain exactly 24 h later (54) is entirely consistent with a role for Cdx1 in inducing mA33 antigen expression during intestinal development. Meanwhile, the genes encoding the pancreatitis-associated protein 1 (25) and claudin-2 (26) are the only other putative Cdx1 target genes identified to date, and their expression patterns closely resemble that of Cdx1 in intestinal epithelium in vivo (26, 55). In contrast, genes expressed specifically by the differentiated cells in the small intestine (SI, LPH, and GC-C) and colon (CA1) are targets for CDX2 in vitro, and their expression patterns are almost superimposable with that of CDX2 in vivo (Fig. 1) (8–12, 50, 51, 56).

In the adult, both the A33 antigen and CDX1 are co-expressed in the proliferative cells in intestinal crypts (16, 33); however, A33 antigen expression is also seen in the differentiated cells. This could result from induction of A33 antigen expression by CDX1 in cells at the base of the crypts, and maintenance of A33 antigen expression during cell migration and differentiation via mRNA and/or protein stabilization. This question could be further addressed by determining the localization of mA33 antigen mRNA in vivo using in situ hybridization.

The temporospatial expression pattern of GKLF/KLF4 and IKLF/KLF5 during intestinal development through stages E10.5 to E15.5 has not been studied in detail. However, expression of IKLF/KLF5 in the adult intestine is reminiscent of that of CDX1, being found at its highest level in the lower crypt region (47). Meanwhile, expression of GKLF/KLF4 more closely mimics that of CDX2, being expressed by the differentiated cells in the small intestine and colon, where it has been implicated in regulating exit from the cell cycle by direct repression of cyclin D1 promoter activity (57). It is conceivable therefore that CDX1 and IKLF/KLF5 together regulate gene expression in the lower, proliferative compartment, whereas CDX2 and GKLF/KLF4 produce augmented patterns of gene expression in

FIG. 6. Expression of CDX1, CDX2, and GKLF in confluent human CRC cell lines. Total RNA was extracted from 14 cell lines and normal colon and subjected to Northern analysis for expression of CDX1 protein. Samples of mRNA or protein extracted from ha33 antigen (+) colorectal cancer cell lines are enclosed in parentheses. The right-hand lanes in C and D contain mRNA extracted from normal human colon (positive control).

FIG. 7. Cdx1 binds to the proximal CDX element in the A33 antigen gene. Binding of nuclear proteins extracted from untransfected 293T cells and 293T cells transfected with pRCmCdx1, encoding mouse Cdx1, to a double-stranded radiolabeled DNA probe encompassing the proximal CDX1/CDX2-binding site. Lanes 1, 3, 5, 7, and 8 contain a radiolabeled wild-type (wt) probe (5’-GATCTCCCTTTTTATATGGATC) encompassing the more proximal putative ha33 antigen CDX-binding site sequence (shown in bold); lanes 2, 4, 6, and 9 contain a radiolabeled probe containing a mutant version of this sequence (5’-GATCTCCCTTGTTGATGATC). Lanes 3 and 4 show complexes formed as a result of binding of nuclear proteins from parental 293T cells to 4 ng of a radiolabeled probe (3,000 cpm/ng). Lanes 5–9 show complexes formed as a result of binding of nuclear proteins extracted from 293T cells transfected with the pRC-mCdx1 plasmid. The protein-DNA complexes seen in lane 7 were formed in the presence of a 25-fold excess of unlabeled probe (wild-type competitor). The protein-DNA complexes seen in lanes 8 and 9 were formed in the presence of an antibody to mCdx1 (α-Cdx1). The bottom black arrow indicates the position of a complex in lane 5 that is supershifted in the presence of anti-mCdx1 antibody (lane 8; top black arrow). The black arrowhead on the left of the panel indicates a band in lanes 4, 6, and 9 that represents a complex formed only with the mutant probe.
the upper, differentiated compartment.

Finally, our results suggest that a small region (400 bp) of the promoter of the A33 antigen gene is likely to be sufficient to produce intestine-specific patterns of gene expression in transgenic mice. Whereas this idea has yet to be tested, the general utility of the mA33 antigen gene locus in driving uniform expression in the intestinal epithelium was recently demonstrated in a proof of principle experiment. A “knock-in” mouse model was created whereby a cassette encoding an internal ribosome re-entry site and a truncated, oncogenic form of β-catenin (ΔNβ-cat) was inserted into the mA33 antigen gene between the stop codon and 3’-untranslated region by homologous recombination (58). Mice homozygous for the targeted transgene expressed ΔNβ-cat and developed a higher frequency of both spontaneous and chemically induced aberrant crypt foci and intestinal adenomas than their wild-type counterparts (58).

Acknowledgments—We thank Debra Silberg for the gift of anti-Cdx1 antibody, Nathan Hall for bioinformatics analysis, and Janna Stockland for expert assistance with the figures. We also thank Helen Abud and Anil Rustgi for helpful comments on the manuscript.

REFERENCES

1. Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shinagawa, R., Yamanaka, N., Izumiya, T., Koseki, N., Itoh, K., Ohtsu, Y., Ueno, Y., and Ishii, T. (1997) Cell 89, 1091–1101.
2. Nakauchi, H., Fujisawa, S., and Naka, K. (1985) J. Exp. Med. 161, 139–148.
3. Stappenbeck, T. S., Dittmer, D. S., and Prince, R. A. (2000) Dev. Biol. 223, 312–323.
4. Rudnicki, M. A., Braun, T., Hinuma, S., and Jaenisch, R. (1992) Cell 70, 109–119.
5. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
6. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
7. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
8. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
9. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
10. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
11. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
12. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
13. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
14. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
15. Rudnicki, M. A., and Jaenisch, R. (1995) Bioessays 17, 526–530.
Analysis of the Regulation of the A33 Antigen Gene Reveals Intestine-specific Mechanisms of Gene Expression
Cameron N. Johnstone, Sara J. White, Niall C. Tebbutt, Fiona J. Clay, Matthias Ernst, William H. Biggs, Carrie S. Viars, Suzanne Czekay, Karen C. Arden and Joan K. Heath

J. Biol. Chem. 2002, 277:34531-34539.
doi: 10.1074/jbc.M204865200 originally published online July 11, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M204865200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 57 references, 18 of which can be accessed free at http://www.jbc.org/content/277/37/34531.full.html#ref-list-1