Goblet-cell-specific transcription of mouse intestinal trefoil factor gene results from collaboration of complex series of positive and negative regulatory elements

Hiroshi ITOH, Nagamu INOUE and Daniel K. PODOLSKY¹

Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, 32 Fruit Street, Boston, MA 02114, U.S.A.

Intestinal trefoil factor (ITF) is expressed selectively in intestinal goblet cells. Previous studies of the rat *ITF* gene identified one *cis*-regulatory element, designated the goblet-cell-response element (GCRC), present in the proximal region of the promoter. To identify additional *cis*-regulatory elements responsible for goblet-cell-specific expression, a DNA fragment containing 6353 bp of the 5'-flanking region of the mouse *ITF* gene was cloned and its promoter activity was examined extensively. In human and murine intestinal-derived cell lines (LS174T and CMT-93), the luciferase activities of a 6.3-kb construct were 5- and 2-fold greater than the smaller 1.8-kb construct, respectively. In contrast, the activity in non-intestinal cell lines (HepG2 and HeLa) was 2–4-fold lower than the smaller construct. In the region downstream from the 1.8-kb position, strong luciferase activities in LS174T and HepG2 cells were observed using a 201-bp construct. Interestingly, increased activity was almost completely suppressed in cells transfected with a 391-bp construct.

Detailed analyses of this region revealed the existence of a 11-bp positive regulatory element (–181 to –170; ACCTCTTCCTG) and a 9-bp negative regulatory element (–208 to –200; ATT-GACAGA) in addition to the GCRC. All three elements were well conserved among human, rat and mouse *ITF* gene promoters. In addition, a mutant 1.8-kb construct in which the negative regulatory region was deleted yielded the same approximate luciferase activity as a 6.3-kb construct, suggesting binding of a goblet-cell-specific silencer inhibitor (SI) between –6.3 and –1.8 kb. The SI present in goblet cells may block the silencers' binding to the pre-initiation complex and allow increased transcriptional activity driven by specific and non-specific enhancers. High-level expression of the mouse *ITF* gene specifically in intestinal goblet cells may be achieved through the combined effects of these regulatory elements.

Key words: colon, enhancer, epithelium, promoter, silencer.

INTRODUCTION

The trefoil peptides constitute a family of small regulatory peptides that have one or more trefoil motifs (also called a P-domain) comprised of six conserved cysteine residues (CX₂C-X₂CX₂CCX₂WCF) (reviewed in [1–4]). Three different members of the trefoil peptides are present in mammals, designated spasmolytic polypeptide (SP) [5], pS2 peptide [6] and intestinal trefoil factor (ITF) [7]. SP has two trefoil motifs and is expressed primarily in the stomach [8,9], although the porcine homologue was originally isolated from pancreas [5]. pS2 peptide, initially cloned as the product of an oestrogen-responsive gene from a breast cancer cell line MCF-7 [6], has a single trefoil motif and is also normally expressed in the stomach, although the expression pattern differs from SP [10]. In pathological conditions such as inflammatory bowel disease, these peptides are also expressed in cells adjacent to ulcers of the small and large intestine [11,12]. Tumour cells may also express these in an 'ectopic' pattern [13].

ITF was first cloned from rat intestinal villous epithelial cells and contains one trefoil motif [7]. In contrast to pS2 and SP, ITF is expressed selectively in high concentrations by mucus-producing goblet cells of the small and large intestine [7,14,15].

Studies *in vitro* demonstrated that ITF promotes epithelial migration through a transforming growth factor- β -independent pathway [16] and protects intestinal-cell monolayers against a variety of injurious agents in co-operation with mucin glycoprotein [17]. Furthermore, studies *in vivo* have demonstrated that increased ITF expression was observed in the recovery phase of experimental colitis [18,19], and oral administration of ITF was shown to protect gastric mucosa from injury [20]. Mice lacking the *ITF* gene have impaired mucosal healing after experimental induction of colitis [21]. Therefore, ITF, a dominant product of differentiated goblet cells, appears to play a critical functional role in promoting healing after mucosal damage. Recently, conserved ITF expression in the progression from normal mucosa to adenoma and carcinoma in humans has been observed [22]. Low-level expression of ITF in human hypothalamus, pituitary [23] and breast epithelial cells [24] has also been reported, although its pathophysiological role is unknown.

Despite accumulating knowledge about trefoil peptide function, mechanisms conferring cell-specific expression remain unclear. The promoter of the *pS2* gene normally expressed in gastric mucosa has been partially characterized [25,26], although these studies were performed using breast cancer cell line MCF-

Abbreviations used: ITF, intestinal trefoil factor; mITF, mouse ITF; rITF, rat ITF; SP, spasmolytic polypeptide; GCRC, goblet-cell-response element; GMSA, gel-mobility-shift assay; EGF, epidermal growth factor; LA-PCR, long-and-accurate PCR; SI, silencer inhibitor; BAC, bacterial artificial chromosome; SV-40, simian virus 40.

¹ To whom correspondence should be addressed (e-mail Podolsky.Daniel@mgh.harvard.edu).

The nucleotide sequence data reported will appear in GenBank Nucleotide Sequence Database under the accession number AF051339.

7. Although promoter sequences of several genes expressed exclusively in the intestinal columnar epithelium, such as fatty acid-binding protein [27], sucrase-isomaltase [28] and lactase [29], have been examined, the *MUC2* gene is the only other promoter of a gene expressed in goblet cells that has been reported [30]. Since goblet cells are major mucus-producing cells of the intestine and thought to play an important role in mucosal protection [30], detailed exploration of the *ITF* gene promoter may provide additional insights into the regulatory mechanism of goblet-cell-specific gene expression.

Previous studies of the rat *ITF* (*rITF*) gene promoter identified a *cis*-regulatory element, designated the goblet-cell-response element (GCRC), present in the proximal region of the promoter, which supported goblet-cell-associated expression [31,32]. However, nuclear proteins from several non-intestinal cell lines bound to this element and the luciferase activity transiently expressed in HepG2 cells was still high using the limited region of the *rITF* gene promoter studied [32], suggesting the presence of additional goblet-cell-specific regulators in the region of the promoter, far upstream. In order to identify additional *cis*-regulatory elements responsible for goblet-cell-specific expression, a longer 5'-flanking region of the mouse *ITF* (*mITF*) gene was cloned and its promoter activity was examined extensively by transient-transfection assay and gel-mobility-shift assay (GMSA).

MATERIALS AND METHODS

Cloning and sequencing of the 5'-flanking region of the *mITF* gene

A mouse (129/SvJ) genomic bacterial artificial chromosome (BAC) library (Genome Systems, St. Louis, MO, U.S.A.) was screened with an 860-bp DNA fragment corresponding to the 5'-flanking region of the *mITF* gene as a probe. The probe was made by PCR using the primers R1 (sense) 5'-TTTCTGACTC-TGCATGCTCATGCAC-3' and R1 (antisense) 5'-GGCAGC-AGGCAAGCTTCAGGACCACTGCACAGGAT-3' and the plasmid PT7, which contains the full-length *mITF* gene cloned previously in our laboratory [33], as a template. Recombinant DNAs from positive BAC clones were extracted by standard alkaline lysis and purified by phenol/chloroform and ethanol precipitation according to the manufacturer's protocol. Subsequently, the DNA was divided into pools and each was digested with different restriction enzymes and subjected to Southern-blot analysis. The DNA pool digested with *Bam*HI containing the longest 5'-flanking region of the *mITF* gene was subcloned into a pBluescript II KS+ phagemid vector (Stratagene, La Jolla, CA, U.S.A.) and screened again by colony hybridization using 5'-flanking region of the *mITF* gene as a probe. By these steps, plasmid pBS-5'ITF, containing a positive 10-kb *Bam*HI fragment, was obtained. After growing in a large-scale preparation, the pBS-5'ITF plasmid was used for sequencing with the T7 Sequenase 7-deaza-dGTP kit (Amersham Life Science, Cleveland, OH, U.S.A.). Sequencing was obtained in both directions to obtain maximum accuracy for the sequence given (sequencing strategy depicted in Figure 1B). The total 6353 bp of the 5'-flanking region of the *mITF* gene was included in the plasmid pBS-5'ITF.

Preparation of the luciferase reporter constructs

The promoterless pGL3-basic (Promega, Madison, WI, U.S.A.), which contained a luciferase structural gene immediately downstream of a polylinker, was used for reporter constructs. Since no appropriate restriction sites were present near the start codon of

the *mITF* gene, a DNA fragment spanning between -6353 and +24 of *mITF* 5'-flanking region was made by the long-and-accurate PCR (LA-PCR) technology using LA *Taq* DNA polymerase (Takara, Otsu, Japan) and a subcloned 10-kb fragment as a template. Primers used were as follows: T3(+)M2, 5'-AGCTTGATAACGCGTTCCTGCAGCCCCGGGGAATCC-3'; and 5'(-)B, 5'-AGCTTCAGGATCCCTGCACAGGATGTGCAAGGTCA-3'. The T3(+)M2 primer contained the *Mlu*I restriction site at the 5' end and disrupted the upstream *Bam*HI site of the 10-kb fragment, while the 5'(-)B primer contained a *Bam*HI site just upstream of the start codon. The resulting 6.5-kb PCR products were subcloned into *Mlu*I and *Bgl*II sites of a pGL3-basic vector, because *Bgl*II forms a compatible end with *Bam*HI. The subcloned DNA construct comprised of the -6353/+24 *mITF* gene linked to the luciferase gene (designated 6353WT) was transformed into competent *Escherichia coli* DH5 α cells (Clontech, Palo Alto, CA, U.S.A.). To confirm the correct orientation and preservation of the start codon of luciferase, subcloned DNA constructs were subjected to restriction mapping and sequencing of the insertion junctions.

Among deletion constructs of the 6353WT, a -201/+24 *mITF*-luciferase construct (201WT) was made by simply ligating after digestion of 6353WT with *Mlu*I and *Bgl*II and blunt-end formation by Klenow enzyme (Promega). This took advantage of a convenient *Bgl*II site at position -201, because a *Bgl*II site in the polylinker was disrupted by ligation with a *Bam*HI site when subcloning. In the same manner, a -1848/+24 construct (1848WT) was made by ligation after digestion of 6353WT with *Mlu*I and *Eco*RI. Further deletion plasmids, designated 1756WT, 1590WT, 1370WT, 1195WT, 708WT and 391WT, were constructed by treatment of 1848WT with exonuclease III and mung-bean nuclease after digestion with *Kpn*I and *Mlu*I using the Erase-a-Base system (Promega). Smaller constructs, 341WT, 294WT, 241WT, 224WT, 213WT, 184WT, 164WT, 150WT and 120WT, were generated by ligation of *Mlu*I- and *Bam*HI-digested PCR products into a pGL3-basic vector. The deletion constructs 6353M1 and 1841M1 were made by ligating the *Mlu*I- and *Bam*HI-digested LA-PCR products, corresponding to positions -6353/-1848 and -1848/-1370, into the *Mlu*I- and *Bgl*II-digested 6353WT plasmids, respectively, using the *Bgl*II restriction site at position -201.

Constructs driven under the simian virus 40 (SV-40) promoter were also prepared using pGL3-promoter vector (Promega). An *Mlu*I- and *Bam*HI-digested LA-PCR product corresponding to positions -6353/-1848 was ligated into *Mlu*I and *Bgl*II sites of this promoter vector in the same manner as 6353M1, and designated 4.5kP. A construct 1.9kP was generated by self-ligating after digestion of 4.5kP with *Sca*I and blunt-end ligation by Klenow enzyme. The 2.6-kb *Sca*I fragment after digestion of 4.5kP was used for generating a construct 2.6kP.

Mutant constructs, ET1-5 and ST1-6, were prepared by replacing the wild-type sequences with mutated sequences generated by the QuikChange[™] site-directed mutagenesis kit (Stratagene) using the constructs 224WT and 201WT as templates, respectively. Constructs 120M3 and SR were generated by ligation of phosphorylated double-stranded synthetic oligonucleotides (corresponding to positions -186/-164 and -224/-196) into *Mlu*I sites of 201WT and 120WT, respectively. Inverted synthetic oligonucleotides were inserted into the vector to yield the construct designated SR.

All constructs were verified by DNA sequencing. Plasmid-preparation purity was confirmed by an $A_{260/280}$ value of > 1.6, and supercoiling of DNA was established by the appearance of a quickly migrating band on agarose-gel electrophoresis prior to use in transfection experiments.

Cell lines

All cell lines used (human colon cancer cell line LS174T, mouse rectal cancer cell line CMT-93, human hepatocellular carcinoma cell line HepG2, human cervix epithelioid cancer cell line HeLa, human fibrosarcoma cell line HT1080 and mouse immortalized fibroblast cell line NIH3T3) were obtained from the American Type Culture Collection (Rockville, MD, U.S.A.). LS174T cells exhibit a goblet-cell-like phenotype and express high level of endogenous ITF [34], whereas CMT-93 cells have a rectal-epithelial-cell-like phenotype and express minimal amounts of endogenous ITF. LS174T cells were grown in Eagle's minimum essential medium and the other cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal calf serum, 4 mM L-glutamine, 50 units/ml penicillin and 50 mg/ml streptomycin in 5% CO₂ at 37 °C.

Transient transfections and promoter analyses

Transient transfection was performed by the calcium phosphate-precipitation method using the calcium phosphate mammalian cell transfection kit (5 Prime and 3 Prime, Boulder, CO, U.S.A.), or the lipofectin method using Tfx[®]-50 lipofectin reagent (Promega), according to each manufacturer's protocol. Prior to transfection (16–24 h), cells were plated out in triplicate in 35-mm wells of a six-well cell-culture plate. Complete media were refreshed 2 h prior to transfection. Efficiency of transfection was standardized by co-transfecting with pSV β -galactosidase control vector (Promega) and adjusting for the amount of β -galactosidase activity. Calcium phosphate-precipitated plasmid DNA was added to each well and incubated at 5% CO₂ for 4 h before a 2-min exposure to 15% glycerol, while serum-free medium containing Tfx[®]-50-treated DNA was added to each well and incubated for 1 h before addition of serum-containing medium. In both methods, cells were cultured for 48 h in the same conditions prior to assay for reporter-gene expression. For determination of luciferase and β -galactosidase activities, cells were lysed and assayed immediately using a commercial luciferase-assay system (Promega) and the Luminescent β -gal Genetic Reporter System II (Clontech), respectively, measured in a Monolite 2010 luminometer (Analytical Luminescence Laboratory, Ann Arbor, MI, U.S.A.). Luciferase activity was adjusted for transfection efficiency and harvesting efficiencies by dividing the value of luciferase activity by that of β -galactosidase activity. Promoter activity was expressed as a percentage of the expression of the maximal promoter construct, pGL3-control vector (Promega), consisting of the SV-40 promoter and enhancer connected to the luciferase gene. Reproducibility of the results was confirmed by at least three independent transfections and each transfection was done in triplicate. The values are expressed as the means $\pm$ S.E.M. for the nine determinations.

Extraction of nuclear proteins and GMSAs

Nuclear extracts from cultured cells were prepared by Nonidet P-40 detergent lysis and 0.5 M NaCl extraction performed as described by Schreiber et al. [35]. Protein concentration was determined by the Bradford assay [36]. The wild-type double-stranded synthetic probes used in this study were as follows: E1, ⁻²⁰¹GATCTGCCATGGCTGTCCCCACCTCTTCCTGGTCC-TTGTTTCTCCCTAG⁻¹⁵¹; E3, ⁻¹⁸⁹CTGTCCCCACCTCT-TCCCTGGTCCCTG⁻¹⁶⁴; S1, ⁻²⁴¹CAGAGTCTGCTTCTAGACTAGGTGTACACATTGATTGACA⁻²⁰²; and S2, ⁻²²⁴CTAGGTGTACACATTGATTGACAGATC⁻¹⁹⁸. They were label-

led by Klenow fill-in reaction in a buffer consisting of 10 mM Tris/HCl, 5 mM MgCl₂, 7.5 mM dithiothreitol, 33 mM dATP, dGTP and dTTP, 0.33 mM [α -³²P]dCTP and 1 unit of Klenow enzyme, as described previously [32,37]. GMSAs were carried out by incubating 10 mg of nuclear extract with 5 fmol of probe (20000 c.p.m.) in 20 ml of binding reaction containing 10 mM Tris/HCl (pH 7.5), 50 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 10% glycerol and 1 mg of poly(dA-dT). After incubation at room temperature for 30 min, samples were loaded on to 6% polyacrylamide, 0.25 $\times$ Tris borate gels and electrophoresed at 10 V/cm for 2 h. Competition experiments were carried out by preincubating the nuclear extracts with a 100-fold excess of unlabelled wild-type (E1–4, ET0, S1–6 and ST0) or mutant (Em1, Em2, ET1–5, Sm1 and ST1–6) competitor oligonucleotides prior to addition of the probe. The competitor rS6 corresponded to the rITF sequence at positions –191/–175 [31]. The positions and nucleotide sequences of these competitors are indicated in Figures 5(A) and 6 (see below). The gels were dried for 30 min and exposed to Kodak X-AR film for 6–24 h at –80 °C. Reproducibility of GMSAs was confirmed by at least three independent assays.

RESULTS

Cloning and sequencing of the 5'-flanking region of the *mITF* gene

A 10-kb *Bam*HI fragment, including approx. 6.5 kb of the 5'-flanking region of the *mITF* gene was obtained from a BAC library and then sequenced in both directions to obtain maximal accuracy. The nucleotide sequence of the total 6353 bp of the 5'-flanking region of the *mITF* gene is shown in Figure 1(A). A previous description of the *mITF* gene was limited to 1.8 kb of the *mITF* 5'-flanking region [38]. Within that 1.8 kb, divergence at 24 nucleotide positions from the published sequence was found and confirmed. Based on our previous results yielding the nucleotide sequence of products obtained by rapid amplification of 5'-cDNA ends from mouse intestinal RNA [18] and the results of primer-extension analysis by Chinery et al. [38], the transcription-start site of *mITF* was determined to be 35 bp upstream from start codon and was designated +1. A presumed RNA polymerase-II promoter site (TATA box) was present in nucleotide positions +25 to +30. As described previously [33,38], a 240-bp AT-rich region containing a Pit-1 homeobox domain (positions –667 to –424) and a 14-fold repeat of AC dinucleotide capable of forming a z-form DNA sequence (positions –781 to –752), similar to that previously observed in the *rITF* promoter, were found. Other unique sequences included a 17-fold repeat of ATCT sequence (positions –3127 to –3060) and an incomplete poly-A sequence (positions –3798 to –3745) in the region far upstream of the *mITF* promoter. A computer database (TRANSFAC v3.2) search revealed potential binding sites for known transcription factors, including activator protein 1 (AP-1) and Oct-1 and -2 in this promoter (shown in Figure 1A). No region of high sequence similarity was found when the sequence of the *mITF* promoter was compared with those of rat intestinal fatty acid-binding protein [27] and human pS2 [25]. Neither the 13-bp imperfect palindrome considered to be the oestrogen-responsive element nor the epidermal growth factor (EGF)-responsive element of the pS2 promoter [26] was found in the *mITF* promoter, but several palindromic sequences that could potentially serve as steroid-hormone-responsive elements were seen in the region between –285 and –187 (shown in Figure 1A).

Figure 2 Transient expression of *mITF* promoter constructs in intestine-derived LS174T and CMT-93 cells and non-intestinal HepG2 and HeLa cells

Deletion constructs (6353WT, 1848WT, 4.5kP, 1.9kP and 2.6kP), numbered according to the distance from the transcription-start site, and control pGL3 and pGL3P constructs are indicated on the left. Results are expressed as a percentage of luciferase activity compared with cells transfected with the maximal Rous sarcoma virus promoter driving the luciferase (Luc) gene. Luciferase activity was measured as relative light units adjusted for efficiency of transfection by co-transfection with pSV β -galactosidase control vector. All transfections were done in triplicate and repeated at least three times. Results are shown as the means $\pm$ S.E. indicated by error bars. Activity was compared between (A) the longest 6353WT and a shorter 1848WT construct or (B) three deletion constructs driven by the SV-40 promoter.

Characterization of *mITF* promoter activity by transient-transfection assay

In the absence of established non-tumorigenic goblet-cell lines, promoter analysis was undertaken using LS174T, a human colon cancer-derived cell line with goblet-like phenotype, and CMT-93, a mouse line derived from a rectal carcinoma. In preliminary analysis using LS174T (human) and CMT-93 (mouse) intestine-derived cell lines, expression of a 6.3-kb construct (6353WT) was 5- and 2-fold greater, respectively, than a smaller 1.8-kb construct (1848WT), as assessed by luciferase activity. In contrast, in non-intestinal cell lines (HepG2 and HeLa), the activity of the longer construct was 2–4-fold lower than the smaller construct (Figure 2A). Therefore, we first considered that both enhancer and silencer elements might exist in the region between –6353 and

–1848. The former may be active in intestinal cells but not in non-intestinal cells and the latter may be active in non-intestinal cells but not in intestinal cells. To evaluate these possibilities, we made constructs driven by the SV-40 promoter. Unexpectedly, the expression levels of luciferase of all constructs examined were significantly decreased in both non-intestinal HepG2 and HeLa cells and intestinal LS174T and CMT-93 cells (Figure 2B). This result suggests that the region spanning –6353 to –1848 is likely to have at least one or more general silencer elements, but not any general enhancers.

In the region downstream from –1848, strong luciferase activities in LS174T and HepG2 cells and weak activities in CMT-93 and NIH3T3 cells were observed after transfection with a 201-bp construct (201WT), which contained the GCRE identified previously in the *rITF* gene [37] (Figure 3). The increased

Figure 3 Detailed analysis of transient expression of the *mITF* promoter constructs in LS174T, HepG2, HeLa, CMT-93, NIH3T3 and HT1080 cell lines

Deletion constructs, numbered according to the length from the transcription-start site, and control pGL3 are indicated on left. Results are shown in same manner as described for Figure 2.

activity was almost completely suppressed in all cells transfected with a 391-bp construct (391WT) (Figure 3). This result suggested the presence of adjacent positive and negative regulatory elements, including GCRE, within 391 bp 5' from the transcription-initiation site. Luciferase activity was almost constant in all cells transfected, with constructs containing from -1370 to -391 of the *mITF* 5'-flanking region (391WT, 708WT, 1007WT, 1195WT and 1370WT). However, constructs ranging from -1848 to -1590 (1590WT, 1756WT and 1848WT) yielded a 2-fold increase in LS174T and HepG2 cells, suggesting the presence of another positive regulatory element in the region between -1590 and -1370, which includes potential Oct-1 and -2-binding sites (see Figure 1). Luciferase activity was low in all cells transfected with construct 120WT (less than 2% of control), indicating that a positive and negative regulatory element complex is present within the region between -391 and -120.

Thus detailed transient-transfection analysis of this region using LS174T and HepG2 cells revealed the presence of two positive regulatory elements (nucleotide positions from -184 to -164 and from -150 to -120) and one negative regulatory element (from -214 to -201) (Figure 4A). One positive regulatory element was attributable to the GCRE present in the region between -150 and -120, whereas the region between -186 and -164 contained a second 'new' positive regulatory element, different from GCRE. Indeed, these regions contain sequences that show high sequence similarity among the mouse, rat and human *ITF* 5'-flanking regions [32,39] (Figure 4B). In contrast, the single negative regulatory element spanning -213

to -201 showed only limited sequence similarity among the three species (Figure 4B). Interestingly, the mutant construct 120M3, which consisted of oligonucleotide E3 (-186/-164) ligated directly to 120WT and represented only the upstream positive element without GCRE, did not show increased luciferase activity, suggesting that the novel positive regulatory element needed to associate with the downstream 3'-GCRE.

Of most interest was that diminished luciferase activity observed after transfection with smaller constructs was significantly greater in intestine-derived LS174T cells transfected with the longest 6353WT construct. In contrast, luciferase activity was further reduced in non-intestinal HepG2 and HeLa cells transfected with the 6353WT construct (Figures 2A and 3). Transfection with a mutant construct containing the far-upstream region (-6353 to -1848; designated 6353M1) showed decreased luciferase activity in non-intestinal cells but was not altered in intestinal cells, compared with activity in cells transfected with the 201WT construct (Figure 3). Moreover, transfection with another mutant construct, 1848M1, a 679-bp construct derived from 1848WT and presumed to contain all positive regulatory elements without the negative regulatory elements identified in this study, yielded maximal promoter activity in LS174T cells equivalent to the 6353WT construct (Figure 3). Interestingly, luciferase activity in non-intestinal cells transfected with 1848M1 was at almost the same maximum level as that in cells transfected with a 201WT construct (Figure 3). These results suggest that the presence of a promoter-specific enhancer between -6353 and -1848 is unlikely. This is based on the comparison of 6353WT

Figure 4 (A) Transient expression of *mITF* promoter constructs in LS174T and HepG2 cells and (B) comparison of the nucleotide sequence of GCRC identified previously [32] and positive and negative regulatory elements identified in this study between mouse, rat and human *ITF* 5'-flanking regions

Results are shown in the same manner as described in legend to Figure 2. Two positive regulatory elements are present between the positions -184 and -164, and between -150 and -120, and one negative regulatory element is present between -213 and -201. *, Sites of bases conserved between species.

with 6353M1 and the fact that binding of a protein (a silencer inhibitor, or SI) to the intestinal-cell-specific element present between -6353 and -1848 can block the negative regulatory elements, based on the comparison of 6353WT with 1848M1.

Characterization of nuclear proteins binding to the positive and negative regulatory elements immediately 5' of GCRC in the *mITF* gene promoter

Synthetic oligonucleotides of various lengths and corresponding mutants were generated as probes or competitors for use in EMSA experiments to assess nuclear extracts from the goblet-cell-like LS174T cells for the presence of proteins interacting with the 5'-flanking region implicated in transient transfection studies (see Figure 5A). When using E1 (which corresponded to positions -201 to -150 and contains the positive regulatory element 5' from GCRC) as a probe, major bands representing probe DNA and DNA-binding-protein complexes derived from nuclear extracts of the LS174T cells were observed. Competition EMSA studies using this probe showed that the E1 (unlabelled) and E3 oligonucleotides competed specifically for binding to one of the major bands (Figure 5B). In addition, two mutant oligonucleotides, Em1 and Em2, which contained 7-bp mutations in the 5' end (Em1) and 3' end (Em2) of E3, also competed for binding of the wild-type probe (Figure 5B), indicating that a

Figure 5 Binding of LS174T nuclear extracts to the promoter sequence by EMSA

Schematic representation of GCRC followed by negative and positive regulatory elements identified. The nucleotide positions of probes and competitors are also indicated. Dashed lines indicate mutated sequences. Radiolabelled probe E1 (B) or S1 (C) was mixed with nuclear extracts prepared from LS174T cells in the absence or presence of 100-fold excess of unlabelled double-stranded E1 or S1 or in the presence of one of various competitors indicated in (A), as described in the Materials and methods section. Bands representing specific protein-DNA complexes and unbound free probes are indicated by arrows.

DNA-binding site for this nuclear protein might be present within the 10 bp of -179 to -170. These findings are consistent with the results of transient-transfection assays, indicating that the region between -184 and -164, which is highly homologous between mouse, rat and human, contains a new positive regulatory element different from GCRC.

In the upstream negative regulatory region, four major bands were observed when using a probe designated S1, which corresponds to positions -241 to -201. Competition studies showed that the S1 (unlabelled), S2, S6 and mutant Sm1 (containing an 11-bp mutation upstream of S2) oligonucleotides competed for binding to the uppermost among four major bands (Figure 5C), indicating that a DNA-binding site for this nuclear protein might be present within the 15 bp of -214 to -201. These findings are also consistent with the results of transient-transfection assays, suggesting that the region between -213 and -201 contains a negative regulatory element.

To examine the interactions between nuclear proteins from LS174T cells and positive and negative regulatory elements in more detail, EMSAs were carried out with various mutant competitors. Since the appearance of some non-specific bands in EMSAs resulted in interference with the band of interest, shorter specific E3 and S2 oligonucleotides were used as probes in the

Figure 6 Identification of nuclear protein-binding to positive and negative regulatory elements by GMSA

GMSA of LS174T nuclear extracts probed with E3 (**A**) or S2 (**B**) in the absence or presence of the unlabelled competitors was performed as described for Figure 5 and in the Materials and methods section. Nucleotide substitutions in wild-type and mutant competitors used and the corresponding nucleotide sequences of the human and rat *ITF* gene promoters are shown above the figures. The element identified for each specific binding is boxed and the corresponding human and rat elements are also indicated. Bands representing specific protein–DNA complexes and unbound probes are indicated by arrows.

subsequent detailed analysis. Nucleotide substitutions in wild-type and mutant competitors used are shown in Figure 6. In the positive regulatory region, the wild-type ET0 oligonucleotide competed for binding to the single major protein complex. Competition for binding to this band by the mutant ET5 but not ET1–4 was also observed (Figure 6A). Similarly, in the negative regulatory region, competition for binding to the band of interest was observed with unlabelled excess wild-type ST0 and mutants ST1, ST2 and ST6, but not ST3–5 (Figure 6B). These results demonstrate the presence of the nuclear proteins in LS174T cells capable of binding to a 11-bp positive regulatory element (–181 to –170; ACCTCTTCCTG) and a 9-bp negative regulatory element (–208 to –200; ATTGACAGA), in addition to the GCRE. The former is highly homologous among mouse, rat and human *ITF* gene promoters, whereas the latter is less homologous among these three species (Figure 6). Interestingly, the band present in GMSA binding to the negative regulatory region was also competed with by the competitor rS6, which represents the corresponding region of the *rITF* promoter sequence, indicating that nuclear proteins from LS174T cells could bind the *rITF* promoter sequence corresponding to the mouse negative regulatory element. Therefore, the consensus nucleotide sequence of the negative regulatory element might be A(A/T)TG(A/T)–

C(A/T)G(A/T). The same GMSA results were obtained using nuclear proteins derived from non-intestinal HepG2 cells (results not shown).

Expression of the reporter-gene luciferase driven under the mutated *mITF* promoter

In order to determine whether the presence of nuclear proteins binding to the newly defined positive and negative regulatory elements in GMSAs actually correlates with promoter activity of these elements, the effects of various mutations of the *mITF* promoter on gene expression were assessed by transient-transfection assay. Minimal DNA constructs, 201WT and 224WT, which confer maximal enhancement of and reduction of luciferase activities, respectively, were used as templates for generating various additional mutant constructs by site-directed mutagenesis. Constructs were transiently transfected into LS174T cells and gene expression was analysed by measuring relative luciferase activity. In studies of the positive regulatory element, the luciferase activity in LS174T cells and HepG2 cells transfected with ET1–4 was decreased by 2–8-fold compared with control wild-type ET0 (the same as 201WT).

Figure 7 Transient expression of mutants of positive (A) and negative (B) regulatory elements

DNA constructs containing the wild-type (ET0 and ST0) and mutant (ET1–5 and ST1–6) elements spanning –166 to –181 (A) and –196 to –214 (B) of the *mITF* promoter are indicated on the left. The nucleotides chosen for mutagenesis and the corresponding mutated nucleotides are the same as those shown in Figure 6. In (B), the luciferase activity of a mutant construct SR, in which inverted wild-type sequence was inserted, was also tested. Results are shown in the same manner as described in the legend to Figure 2.

In contrast, no significant reduction in expression was seen in the cells transfected with ET5 (Figure 7A). In studies assessing the negative element, luciferase activity in cells transfected with ST3–5 remained at the same high level as ET0. However, more than 50% reduction in expression was seen in the cells transfected with ST1 and ST2, although the level was slightly higher than the control wild-type ST0 (same as 224WT; Figure 7B) due to the faint binding to the nuclear-protein binding. These results demonstrated the correspondence between sequences attributable to both elements in their ability to specifically bind nuclear proteins and regulate promoter activity in transient-transfection assays. Transfection with a mutant construct SR, containing wild-type sequence inserted in an inverse orientation, was not active, suggesting that the negative regulatory element did not work in the reverse direction.

Characterization of nuclear proteins from various cell lines binding to the positive and negative regulatory elements

Using nuclear extracts from various intestinal and non-intestinal cell lines and E3 and S2 probes, GMSAs were performed to characterize cell specificities of nuclear proteins bound to the positive and negative regulatory elements identified. As shown in Figure 8(A), the nuclear factor bound to the positive regulatory element was abundant in intestinal goblet-cell-like LS174T cells, minimally present in non-intestinal HepG2 and NIH3T3 cells, and not detected in intestinal CMT-93 and non-intestinal HT1080 cells. The absence in CMT-93 cells could in part be responsible for the lower level of endogenous ITF and constant expression

compared with the LS174T cell line. The similar-sized band found in non-intestinal HeLa cells appeared to be distinct from the factor bound to the positive element identified, because this band was not competed with by excess unlabelled competitors. These results remain consistent with the observed increased luciferase activities in LS174T, HepG2, CMT-93 and NIH3T3 cells transfected with a 201WT construct. Weakly increased luciferase activity in CMT-93 cells by transient-transfection assay may have resulted from extremely low-level expression of the nuclear protein bound to this element or the isolated effect of downstream GCRE-binding nuclear factor. Although the expression level varied, the nuclear factor bound to the negative regulatory element was observed in all cells examined (Figure 8B), suggesting that this factor may be expressed universally.

DISCUSSION

Goblet cells are major mucus-producing cells in the small and large intestine, but characterization of the molecular basis of goblet-cell differentiation and function has been quite limited [31]. Among the growing list of cloned genes whose products are intestine-specific [27–29], only *ITF* and *MUC2* appear to be expressed specifically by goblet cells. The present studies have used a variety of model cell lines in attempts to define regulatory elements that result in goblet-cell-specific gene expression. It should be noted that none of these are ideal representatives of the murine intestinal goblet cell, which normally expresses *mITF*. Whereas LS174T does exhibit a goblet-like phenotype, it is human in origin, and CMT-93, a murine rectal epithelial line,

Figure 8 Identification of nuclear protein-binding to positive and negative regulatory elements in various types of cells by EMSA

EMSA was performed with radiolabelled E3 (A) or S2 (B) probes and crude nuclear extracts made from a variety of cell lines indicated above the figures in the absence or presence of a 100-fold excess of unlabelled double-stranded probe. Bands representing specific protein–DNA complexes and unbound free probes are indicated by arrows. (A) Nuclear factor bound to the positive regulatory element was abundant in intestinal goblet-cell-like LS174T cells, minimally present in non-intestinal HepG2 and NIH3T3 cells, but not detected in intestinal CMT-93 and non-intestinal HT1080 cells, whereas (B) nuclear factor bound to the negative regulatory element was observed in all cells examined.

lacks the pronounced goblet-like appearance of LS174T, consistent with relatively low endogenous expression of *mITF*. Both lines are tumour-derived with the possibility of altered transcriptional regulation in association with malignant transformation. Therefore, extrapolation of studies using these lines to normal goblet cells must be undertaken with caution. Nonetheless, the goblet-like phenotype of LS174T, including endogenous products of *ITF* and the apparent conservation of promoter sequences among *ITF* genes from humans, mice and rats, lends plausibility to the relevance of the present findings in this line, while the CMT-93 represents the nearest murine counterpart line.

In previous studies of the *rITF* gene promoter [32,37], we reported the identification of a goblet-cell-specific enhancer element in the *rITF* gene promoter, designated GCRE, bound by a goblet-cell nuclear protein. At least two nuclear proteins bound the GCRE (CCCCCTCCCC) to promote increased reporter-gene expression. In addition to a 'distinctive' GCRE-specific factor in the goblet-cell line LS174T, another nuclear protein was a member of the Sp-binding protein family bound to the Sp-1-like motif (CCTCCCC) in the GCRE [37], a family of zinc finger proteins binding the enhancer regions of diverse genes [39–41]. However, these mutant studies also suggested that binding of our cognate nuclear factor to the GCRE is not sufficient to determine the goblet-cell specificity of *ITF* gene expression.

In the present extensive studies of the *mITF* promoter, we demonstrate the presence of a 11-bp positive regulatory element (–181 to –170; ACCTCTTCCTG, named E1 in Figure 9) and an adjacent 9-bp negative regulatory element (–208 to –200; ATTGACAGA, named S1 in Figure 9) just 5' of the previously identified GCRE [37]. These two novel elements co-operate with

the GCRE to regulate *ITF* gene expression. The nucleotide sequences and the positions of two of these three elements, including GCRE, are highly conserved among mouse, rat and human *ITF* gene promoters. Lesser homology is apparent for the third element. Although the human element appears to be located slightly further 5' than the corresponding positions of the mouse and rat elements, the human sequence is numbered from just 5' of the initiation ATG codon because its transcriptional-start site has not yet been determined [42]. Our recent studies of the human *ITF* gene promoter suggest that the transcription-start site is actually 26 bp upstream from the ATG codon (N. Inoue, H. Itoh and D. K. Podolsky, unpublished work).

In addition to E1 and S1, our present results suggest the presence of at least one or more silencer elements in the region between –6353 and –1848 (S2 and S3 in Figure 9) and an enhancer element in the region between –1590 and –1370 (E2 in Figure 9). On the basis of transient-transfection study results, we postulate binding of a goblet-cell-specific SI to the region between –6.3 and –1.8 kb in a goblet-cell-specific manner as shown in the model illustrated in Figure 9. Binding of SI may block binding of the silencers S1–3 to the pre-initiation complex and allow increased transcriptional activity driven by specific and non-specific enhancers (E1 and E2) in intestinal goblet-like LS174T cells transfected with a 6353 WT construct. S1 suppresses the activity of E1 and E2 in the cells transfected with a 1848WT construct because the inhibitor is inactive. This model is compatible with the observed higher total promoter activity in LS174T cells transfected with a 6353WT construct compared with cells transfected with the smaller 1848WT construct.

This model would provide a basis for selective expression of *ITF* in goblet-like cells and lack of significant expression in other

Figure 9 Hypothetical model of *mITF* transcription

The putative regulatory elements including positive and negative regulatory elements identified in this study and their locations in the *mITF* 5'-flanking region are indicated within the box at the top. E1, E2, S1–3 and PIC indicate positive and negative regulatory elements and pre-initiation complex, respectively. The SI implicated in this study is also indicated.

model cells. We speculate that non-intestinal cells, including HepG2, lack SI. This is consistent with the observed lower total promoter activity in HepG2 cells transfected with the 6353WT construct compared with cells transfected with 1848WT; all silencers (S1–3) were active in the 6353WT construct, whereas only S1 was active in 1848WT. In the mutant construct 6353M1, the negative regulatory element identified (S1) and upstream putative enhancer E2 were unable to work. Total promoter activity derived entirely from E1 in LS174T cells transfected with this construct, exhibiting promoter activity comparable with the 201WT construct, whereas the activity was suppressed by S2 and S3 in HepG2 cells. In the mutant construct 1848M1, all silencers should be inactive in both LS174T and HepG2 cells, accounting for the maximum promoter activity observed in both cells transfected with this construct. After transfection with the 4.5kP construct, luciferase activity suppressed by upstream silencers (S2 and S3) was observed in both LS174T cells and HepG2 cells, suggesting that the proposed SI associates with the *mITF* promoter pre-initiation complex in a gene-specific manner, but not with the SV-40 promoter pre-initiation complex. Collectively, specific intestinal goblet-cell-specific expression of *mITF* may be achieved through the combined effects of these regulatory elements.

Of note is that similar mechanisms regulating cell-specific gene transcription were reported recently in the plasminogen activator inhibitor type-2 gene promoter [42]; total transcription of this gene is regulated by transactivator (anti-silencer) binding to a far-upstream region (–5100/–3300). In the human *MUC2* gene

promoter, which is also expressed in intestinal goblet cells, a sequence (ATTGGGGCT; –135/–127) resembling S1 is present just upstream of an Sp-1-like element (GCCACACCC; –88/–80) and approx. 150 bp downstream of a GCRC-like sequence (CCCCTCCCC; –298/–290) [43].

There has been limited past characterization of the trefoil-gene regulatory elements. As yet, the regulatory elements that are responsible for the regional selective expression of the different trefoil peptides have not been identified. The presence of an EGF-responsive element in the 5'-flanking region of the pS2 gene [26] has suggested that EGF may induce expression of trefoil proteins in response to mucosal injury [44,45]. However, scrutiny of the long 5'-flanking region of the *mITF* gene demonstrated no known EGF-response elements or previously described regulatory elements that play a role in intestine-specific expression of other genes [26–29,46,47].

In summary, we have demonstrated the presence of an adjacent 11-bp positive regulatory element and 9-bp negative regulatory element just upstream of GCRC, which co-operate with the GCRC to regulate total *ITF* gene expression. In addition, our present results suggest the presence of sequences responsible for binding a goblet-cell-specific SI far upstream, which may block silencers binding to the pre-initiation complex and allow increased transcriptional activity driven by specific and non-specific enhancers in the goblet cells. Further characterization of these enhancing elements, especially the SI and the element to which it binds, may provide insight into the molecular basis of the goblet-cell phenotype.

This study was supported by National Institutes of Health grants DK43351 and DK46906. We wish to thank Ms. K. Lynch-Devaney for expert assistance and Dr. P. L. Beck and Dr. I. M. Rosenberg for helpful discussion and critical reading of the manuscript.

REFERENCES

- Hoffmann, W. and Hauser, F. (1993) Trends Biochem. Sci. **18**, 239–243
- Poulsom, R. and Wright, N. A. (1993) Am. J. Physiol. **265**, G205–G213
- Poulsom, R. (1996) Baillière's Clin. Gastroenterol. **10**, 113–134
- Sands, B. E. and Podolsky, D. K. (1996) Annu. Rev. Physiol. **58**, 253–273
- Jørgensen, K. D., Thim, L. and Jacobsen, H. E. (1982) Regul. Pept. **3**, 207–219
- Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A. and Chambon, P. (1982) Nucleic Acids Res. **10**, 7895–7903
- Suemori, S., Lynch-Devaney, K. and Podolsky, D. K. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 11017–11021
- Tomasetto, C., Rio, M.-C., Gautier, C., Wolf, C., Hareuveni, M., Chambon, P. and Lathe, R. (1990) EMBO J. **9**, 407–414
- Jeffrey, G. P., Oates, P. S., Wang, T. C., Babyatsky, M. W. and Brand, S. J. (1994) Gastroenterology **106**, 336–345
- Lefebvre, O., Wolf, C., Kédinger, M., Chenard, M. P., Tomasetto, C., Chambon, P. and Rio, M. C. (1993) J. Cell Biol. **122**, 191–198
- Rio, M. C., Chenard, M. P., Wolf, C., Marcellin, L., Tomasetto, C., Lathe, R., Bellocq, J. P. and Chambon, P. (1991) Gastroenterology **100**, 375–379
- Wright, N. A., Poulsom, R., Stamp, G., Noorden, S. V., Sarraf, C., Elia, G., Ahnen, D., Jeffery, R., Longcroft, J., Pike, C., Rio, M. C. and Chambon, P. (1993) Gastroenterology **104**, 12–20
- Henry, J. A., Bennett, M. K., Piggott, N. H., Levett, D. L., May, F. E. B. and Westley, B. R. (1991) Br. J. Cancer **64**, 677–682
- Podolsky, D. K., Lynch-Devaney, K., Stow, J. L., Oates, P., Murgue, B., DeBeaumont, M., Sands, B. E. and Mahida, Y. R. (1993) J. Biol. Chem. **268**, 6694–6702
- Hauser, F., Poulsom, R., Chinery, R., Rogers, L. A., Hanby, A. M., Wright, N. A. and Hoffmann, W. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 6961–6965
- Dignass, A., Lynch-Devaney, K., Kindon, H., Thim, L. and Podolsky, D. K. (1994) J. Clin. Invest. **94**, 376–383
- Kindon, H., Pothoulakis, C., Thim, L., Lynch-Devaney, K. and Podolsky, D. K. (1995) Gastroenterology **109**, 516–523
- Tomita, M., Itoh, H., Ishikawa, N., Higa, A., Ide, H., Murakumo, Y., Maruyama, H., Koga, Y. and Nawa, Y. (1995) Biochem. J. **311**, 293–297

- 19 Itoh, H., Tomita, M., Uchino, H., Kobayashi, T., Kataoka, H., Sekiya, R. and Nawa, Y. (1996) *Biochem. J.* **318**, 939–944
- 20 Babyatsky, M. W., deBeaumont, M., Thim, L. and Podolsky, D. K. (1996) *Gastroenterology* **110**, 489–497
- 21 Mashimo, H., Wu, D.-C., Podolsky, D. K. and Fishman, M. C. (1996) *Science* **274**, 262–265
- 22 Taupin, D., Ooi, K., Yeomans, N. and Giraud, A. (1996) *Lab. Invest.* **75**, 25–32
- 23 Probst, J. C., Zetzsche, T., Weber, M., Theilemann, P., Skutella, T., Landgraf, R. and Jirikowski, G. F. (1996) *FASEB J.* **10**, 1518–1523
- 24 Poulson, R., Hanby, A. M., Lalani, E. N., Hauser, F., Hoffmann, W. and Stamp, G. W. (1997) *J. Pathol.* **183**, 30–38
- 25 Jeltsch, J. M., Roberts, M., Schatz, C., Garnier, J. M., Brown, A. M. C. and Chambon, P. (1987) *Nucleic Acids Res.* **15**, 1401–1414
- 26 Nunez, A. M., Berry, M., Imler, J. L. and Chambon, P. (1989) *EMBO J.* **8**, 823–829
- 27 Sweetser, D. A., Birkenmeier, E. H., Klisak, I. J., Zollman, S., Sparkes, R. S., Mohandas, T., Lusi, A. J. and Gordon, J. I. (1987) *J. Biol. Chem.* **262**, 16060–16071
- 28 Chantret, I., Lacasa, M., Chevalier, G., Ruf, J., Islam, I., Mantei, N., Edwards, Y., Swallow, D. and Rousset, M. (1992) *Biochem. J.* **285**, 915–923
- 29 Boll, W., Wagner, P. and Mantei, N. (1991) *Am. J. Hum. Genet.* **48**, 889–902
- 30 Gum, J. R., Hicks, J. W. and Kim, Y. S. (1997) *Biochem. J.* **325**, 259–267
- 31 Allen, A. (1986) *Scan. J. Gastroenterol.* **21**, 71–77
- 32 Sands, B. E., Ogata, H., Lynch-Devaney, K., DeBeaumont, M., Ezzell, R. M. and Podolsky, D. K. (1995) *J. Biol. Chem.* **270**, 9353–9361
- 33 Mashimo, H., Podolsky, D. K. and Fishman, M. C. (1995) *Biochem. Biophys. Res. Commun.* **210**, 31–37
- 34 Ogata, H. and Podolsky, D. K. (1997) *Am. J. Physiol.* **273**, G348–G354
- 35 Schreiber, E., Matthias, P., Muller, M. M. and Schaffner, W. (1989) *Nucleic Acids Res.* **17**, 6419
- 36 Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
- 37 Ogata, H., Inoue, N. and Podolsky, D. K. (1998) *J. Biol. Chem.* **273**, 3060–3067
- 38 Chinery, R., Poulson, R. and Cox, H. M. (1996) *Gene* **171**, 249–253
- 39 Kadonaga, J. T., Carner, K. R., Masiaz, F. R. and Tjian, R. (1987) *Cell* **51**, 1079–1090
- 40 Kadonaga, J. T., Courey, A. J., Ladika, J. and Tjian, R. (1988) *Science* **242**, 1566
- 41 Seib, T., Dooley, S. and Welter, C. (1995) *Biochem. Biophys. Res. Commun.* **214**, 195–199
- 42 Antalis, T. M., Costelloe, E., Muddiman, J., Ogbourne, S. and Donnan, K. (1996) *Blood* **88**, 3686–3697
- 43 Alison, M. R., Chinery, R., Poulson, R., Ashwood, P., Longcroft, J. M. and Wright, N. A. (1995) *J. Pathol.* **175**, 405–414
- 44 Wright, N. A., Poulson, R., Stamp, G. W. H., Hall, P. A., Jeffery, R. E., Longcroft, J. M., Rio, M. C., Tomasetto, C. and Chambon, P. (1990) *J. Pathol.* **162**, 279–284
- 45 Merchant, J. L., Demediuk, B. and Brand, S. J. (1991) *Mol. Cell. Biol.* **11**, 2686–2696
- 46 Treisman, R. (1986) *Cell* **46**, 567–574
- 47 McDonnell, S.E., Kerr, L. D. and Matrisian, L. M. (1990) *Mol. Cell. Biol.* **10**, 4284–4293

Received 9 October 1998/4 April 1999; accepted 7 April 1999