Regulation of the Gene Encoding GPR40, a Fatty Acid Receptor Expressed Selectively in Pancreatic β Cells

Reut Bartoo-Shifman, Gabriela Ridner, Keren Bahar, Nir Rubins1, and Michael D. Walker2
From the Department of Biological Chemistry, Weizmann Institute of Science, P. O. Box 26, Rehovot 76100, Israel

GPR40 is a G protein-coupled receptor expressed preferentially in pancreatic β cells. It is activated by long-chain fatty acids and has been implicated in mediating physiological and pathological effects of long-chain fatty acids on β cells. We mapped the GPR40 transcription start site to a location 1044 bp upstream of the translation start site. This permitted definition of the GPR40 core promoter and the organization of the gene, which comprises a 24-bp non-coding exon, a 698-bp intron and a 4402-bp second exon, containing the entire protein coding sequence. Sequence analysis of the GPR40 locus revealed three 4402-bp second exons, containing the entire protein coding sequence. Sequence analysis of the GPR40 locus revealed three evolutionarily conserved regions upstream to the translation start site (HR1–HR3). DNAse I-hypersensitive sites were present in the HR2 and HR3 regions in β cells but not in non-β cells. The 5′-flanking region of the GPR40 gene was capable of directing transcriptional activity selectively in β cells. An important component of this is attributable to the HR2 region, which showed strong β cell-specific enhancer activity. Systematic mutagenesis of HR2 revealed several important sub-regions. Mutagenesis of sub-regions 4–5, and 9 reduced transcriptional activity by ~60 and 40%, respectively. These sub-regions can bind the β cell-specific transcription factors PDX1 and BET2A2, respectively, both in vitro and in vivo. Thus, cell-specific expression of the GPR40 gene involves a characteristic chromatin organization of the locus and is controlled at the transcriptional level through HR2, a potent β cell-specific enhancer.

Type 2 diabetes mellitus (T2DM) is a serious metabolic disease characterized by impaired insulin action on peripheral tissues, combined with defective insulin secretion by pancreatic β cells (1). The incidence of T2DM worldwide is increasing dramatically, driven by a sedentary lifestyle and the associated obesity in the population (2). Obesity and T2DM are associated with elevated levels of lipids such as LCFA in the serum (3, 4). LCFA have complex, divergent effects on β cells: whereas acute exposure leads to enhanced glucose-stimulated insulin secretion, prolonged exposure results in impaired β cell function (“lipotoxicity”) (5, 6). It has been proposed that this impairment may contribute to the defective β cell function associated with T2DM (7). Yet the precise mechanisms linking obesity and T2DM remain incompletely understood.

The mouse GPR40 gene is a member of a small gene family encoding G protein-coupled receptors (8). These receptors are activated by fatty acids: GPR40 responds to medium- and long-chain fatty acids (9–11), whereas the other two members of the family GPR41 and GPR43 respond to short-chain fatty acids (12, 13). The gene family is clustered in a region of ~40 kb on mouse chromosome 7, downstream to the gene encoding CD22, a receptor involved in modulation of B lymphocyte function (14).

GPR40 is expressed in highly selective manner. In rodents, GPR40 mRNA is present at high levels in pancreatic islets and β cell lines, but at very low levels in all other tissues examined (9, 15). In humans, it has been reported that GPR40 is expressed not only in pancreas, but also in brain and monocytes (11). A recent report suggests that GPR40 protein is expressed in mouse pancreatic α cells (16). Several studies have examined the role of GPR40 in β cells. In cultured β cells, down-regulation of GPR40 inhibits LCFA-dependent increases in intracellular [Ca2+] and insulin secretion (9, 17–19). Consistent with this, islets isolated from mice lacking GPR40 show impaired insulin secretion in response to short term exposure to LCFA (15). Interestingly, these mice display significant protection from obesity-induced hyperglycemia and fatty liver (15): on the other hand, overexpression of GPR40 in β cells leads to impaired β cell function and diabetes (15). Taken together, these results indicate that GPR40 plays a role in mediating both acute and chronic effects of LCFA on β cells, and may therefore help explain the link between obesity and T2DM.

The important physiological role of GPR40 prompted us to study the regulation of the gene. We identified the GPR40 transcription start site (TSS) and showed that the promoter is selectively active in β cells. The GPR40 5′-flanking region contains two highly conserved regions that display DNase I hypersensitivity in β cells: one of these, HR3, contains the GPR40 core promoter, whereas the other, HR2, is a strong transcriptional enhancer that plays an important role in directing GPR40 expression selectively to β cells.
Regulation of the GPR40 Gene

EXPERIMENTAL PROCEDURES

RNA Preparation

Total RNA was extracted from cell cultures using the TRI-reagent procedure (Molecular Research Center Inc.) according to the manufacturer’s instructions. Where indicated, RNA was treated with DNase I (RQ1, Promega, Madison, WI) to remove traces of genomic DNA.

5’-Rapid Amplification of cDNA Ends (5’-RACE)

First strand cDNA was synthesized from 5 μg of DNase I-treated βTC1 RNA, using reverse transcriptase (SuperScript II, Invitrogen) according to the manufacturer’s instructions. Primers used were complementary to a specific sequence located either within the GPR40 coding region or upstream to it. The cDNA was purified and a poly-G tail was added to the cDNA 3’-end using terminal deoxynucleotidyl transferase (Promega) by incubating 80% of the cDNA purified above with 1× terminal deoxynucleotidyl transferase buffer (Promega), 0.83 mM dGTP and 20 units of enzyme in a final volume of 60 μl at 37 °C for 1 h. The reaction was stopped by heating to 65 °C for 15 min. The cDNA was purified, and PCR was performed using the Expand high fidelity PCR system (Roche Applied Science), with 5 μl of cDNA and 30 pmol of a reverse primer complementary to GPR40 sequences nested to the primers used for reverse transcription and a forward primer: GAATTC(C)24. A fraction of the PCR product was resolved on a 1% agarose gel. When a band appeared, it was excised from the gel, purified, and sequenced. When a smear appeared, the PCR product was purified using the Roche Applied Science purification kit and subcloned into pGEM T-Easy vector (Promega). Plasmids from the resulting bacterial clones were isolated, fractionated on 1% agarose gel, and analyzed by Southern blot, using as a probe, a radioactive end-labeled primer specific for GPR40 and nested to the primer used for PCR. Three of the relevant plasmids were sequenced.

RNase Protection Assay (RPA)

A plasmid (pBS-RPA3-T7antisense) was constructed in which a genomic fragment, spanning the GPR40 transcription start site, was inserted antisense to the T7 promoter and sense to the T3 promoter. The insert contained 133 bp of the GPR40 second exon, the entire first exon (24 bp), and 52 bp upstream to the transcription start site. In vitro transcription reaction was carried out as described (20). The resulting radioactive probe was 278 bases, of which 209 bases are derived from the GPR40 genomic sequence. RPAs were performed using 50 μg of βTC1 (21) RNA, αTC1 (22) RNA, yeast RNA, or 25 pg of positive control RNA (prepared by incubating pBS-RPA3-T7antisense in the presence of T3 RNA polymerase) (20). Samples were resolved on polyacrylamide urea gels and visualized by autoradiography.

Cell Culture

The following established cell lines were used in this study: βTC1 (mouse β cells) (21), HIT M2.2.2 (hamster β cells) (23, 24), CHO (Chinese hamster ovary cells), 293T (human embryonic kidney cells), and NIH-3T3 (mouse fibroblast cells). βTC1, HIT, CHO, NIH-3T3, and 293T cells were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum (or 15% horse serum, 2.5% fetal calf serum for βTC1 cells), penicillin (200 IU/ml), and streptomycin (100 μg/ml).

Genomic Library Screening

A DNA fragment spanning the GPR40 gene was isolated from a mouse genomic library using a 32P-labeled probe containing the 900-bp GPR40 open reading frame. The clone obtained is 16,195 bp long, begins in the tenth intron of the CD22 gene, contains the entire GPR40 gene, and ends 3 kb downstream to the GPR41 open reading frame. The DNA fragment was characterized by restriction enzyme analysis and partial sequencing.

Transient Transfections

Transfections were carried out by the calcium phosphate co-precipitation technique (25). CHO and HIT transfections consisted of a mixture of 2 μg of reporter construct and 250 ng internal control plasmid. When expression plasmids were used, 1–2 μg of each was added. The total amount of expression plasmid DNA was equalized by addition of the empty expression plasmid pcDNA3. Transfection of 293T cells utilized 4 μg of expression plasmid and 0.5 μg of plasmid pCMV-GFP. Total DNA amount was adjusted to 10 μg by addition of pUC18 DNA. Glycerol shock was applied to HIT cells (20% glycerol) and CHO cells (10% glycerol) 4–7 h after transfection. Cells were harvested 40–48 h after transfection, and whole cell extracts were prepared as described below.

Cell Extracts

Whole Cell Extracts—Cells were washed with PBS, collected, and pelleted by centrifugation for 30 s in a microcentrifuge. The pellet was resuspended in KPi buffer (0.1 M potassium phosphate, pH 7.8, 1 mM dithiothreitol), and extracted by 3 cycles of freezing in liquid nitrogen and thawing at room temperature. Cell debris was removed by 5 min centrifugation in a microcentrifuge at 4 °C.

Nuclear Extracts—Cells were washed with PBS and collected by centrifugation for 30 s in a microcentrifuge. Nuclear extracts were prepared as described (26). Protein concentration of cell extracts was determined by the Bradford method (27) using bovine serum albumin as a standard.

Luciferase Assays

Firefly luciferase and Renilla luciferase assays were carried out as follows: whole cell extracts containing 5–50 μg (1–5 μl) of protein were added to 100 μl of either firefly luciferase assay buffer (20 mM Tricine, 0.1 mM EDTA, 1.07 mM (MgCO3)2Mg(OH)25H2O, 2.67 mM MgSO4, 3.3 mM dithiothreitol, 270 μM coenzyme A, 470 μM luciferin (Promega #E1602), and 530 μM ATP, pH 7.8) or Renilla luciferase assay buffer (0.1 M potassium phosphate, pH 7.4, and 0.5 μM coelenterazine (Calbiochem)). The light output was determined over a 10-s interval using a LUMAC Biocounter M2500 luminometer. Firefly luciferase activity was normalized according to the activity of Renilla luciferase.
**Mutagenesis of GPR40 Promoter**

Site-directed mutagenesis was carried out using the QuikChange system (Stratagene) according to the manufacturer’s protocol. Oligonucleotides designed for mutagenesis of the HR2 region of GPR40 are indicated in the supplemental information.

**Bioinformatics**

Sequence alignment of genomic DNA sequences was performed using NCBI-BLAST (www.ncbi.nlm.nih.gov/BLAST) and software from GCG. Identification of transcription factor binding sites was performed using vRista program and Genomatix suite programs MatInspector and FrameWorker.

**Electrophoretic Mobility Shift Assay (EMSA)**

**Probe Labeling**—The double-stranded oligonucleotides from the HR2 region of the mouse GPR40 promoter, which were used as probes are described in the supplemental materials. Radioactive probes were generated as described (28). Specific activity obtained was typically ~2,000 cpm/fmol.

**Binding Reaction**—Nuclear extracts of βTC1, NIH-3T3, 293T, and transfected 293T cells were prepared as described above. Binding reactions and electrophoresis conditions used for PDX1 were as follows: protein extracts (0.5–2.5 μg) were incubated for 10 min on ice in binding buffer (20 mM Hepes, pH 7.9, 1 mM EDTA, 70 mM KCl, 10% glycerol, 0.5 mM dithiothreitol, 5 mM MgCl₂, and 0.2% Nonidet P-40) with addition of 900 ng of poly(dI-dC) (Sigma) in a final assay volume of 14 μl, in the presence or absence of antibody, as specified in the figure legends. Competition experiments included 10-, 50-, 100-, or 200-fold molar excesses of the following three unlabeled oligonucleotides: HR2-2-6, HR2-M4-5, and the E1 element from the rat insulin gene promoter (29). Following this incubation, 1 μl of [³²P]labeled probe was added (~25 fmol), and incubation was allowed to continue on ice for an additional 20 min. Samples were subsequently resolved on 5.5% polyacrylamide gels (37.5:1 acrylamide:bisacrylamide) in 45 mM Tris borate, 1 mM EDTA) at 20 V overnight. Southern blot analysis was performed as described (20).

**DNase I Hypersensitivity Assay**

**Preparation of Nuclei**—Cells were washed with PBS, trypsinized, and resuspended in growth medium. They were pelleted at 1000 rpm for 5 min, washed twice in cold PBS, and incubated on ice for 10 min in Dounce buffer (20 mM Tris-HCl, pH 7.0, 3 mM CaCl₂, 2 mM MgCl₂, and 0.3% Nonidet P-40). Cells were then homogenized on ice using a Dounce homogenizer (20 strokes, pestle B (loose)), and nuclei were pelleted at 1000 rpm for 7 min at 4 °C. Nuclei were then resuspended in cold RSB buffer (10 mM Tris-HCl, pH 7.0, 10 mM NaCl, and 3 mM MgCl₂), counted in a hemocytometer, pelleted again, and resuspended in 1× RQ1 DNase buffer at concentration of 10⁶ nuclei/ml.

**DNase I Treatment**—A series of digestions using DNase I (RQ1) was performed in the presence of 10⁷ nuclei and increasing concentrations of DNase (0–20 units/ml) for 15 min at 37 °C. Reactions were terminated by addition of 1 volume of 2× TNESK (20 mM Tris-HCl, pH 7.0, 200 mM NaCl, 2 mM EDTA, 2% SDS, and 200 μg/ml proteinase K). Samples were incubated overnight at 37 °C and then phenol-chloroform-extracted, ethanol-precipitated, and resuspended in 100 μl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.8). The concentration of purified genomic DNA was determined using a NanoDrop spectrophotometer.

**Chromatin Immunoprecipitation (ChIP)**

Soluble chromatin was prepared as described (31) from 2× 10⁷ βTC1 cells. For each immunoprecipitation reaction, 25 μg of pre-cleared chromatin was incubated with anti-PDX1 or anti-BETA2 or control antiserum overnight at 4 °C on a rotator wheel. Following centrifugation, the extracts were incubated with 20–30 μl of 50% (w/v) protein A-Sepharose suspension at 4 °C for 2 h on a rotator wheel, and the beads were then washed. The immune complexes were then eluted by incubating the beads three times with 100 μl of elution buffer (1% SDS, 0.1 M NaHCO₃, 20 μg/ml glycogen) for 10 min each. The combined eluates were heated at 65 °C overnight to reverse the formaldehyde cross-links. The eluates were extracted once with phenol-chloroform and once with chloroform, precipitated with ethanol, and resuspended in 20 μl of TE buffer. A 2-μl sample was used for amplification by quantitative real-time PCR.

The real-time PCR reactions were performed in an ABI 7300 real-time PCR system using the “absolute quantification method” option with SYBR-Green PCR master Mix (ABI) and 333 nm primers. The primers used are described in the supplemental materials. In each experiment, a standard curve for each chromatin was plotted using serial dilutions of chromatin input. The PCR program employed 40 cycles of denaturation (95 °C for 10 s), annealing, elongation, and fluorescence detection (60 °C for 1 min). At the end of the final cycle, a dissociation curve of all samples was performed. The DNA recovered from each immunoprecipitate was calculated as a percentage of the total input chromatin.

**Statistical Analysis**

Statistical significance between two groups was assessed using unpaired two-tailed t test. The data in Fig. 5 was analyzed by two-factor analysis of variance using experiment and mutant as the factors. Means for each mutant were compared with wild-type mean using the Dunnett test (α = 0.05). The data in
Regulation of the GPR40 Gene

A. Exon 1 Exon 2

ATG

Primer E2bot4
Primer E2bot5
Primer 44413

B. βTC1 Positive cDNA RNA

400 bp

M 1 2 3 4

C. 5’ flank Exon 1 Exon 2 T7 promoter

RPA probe 278 bp
Full protection 209 bp
Product of major TSS 157 bp

D. RNA: βTC1 αTC1 in vitro-transcribed yeast

FIGURE 1. Mapping of the transcription start site of the GPR40 gene by 5’-RACE and RPA. A, primers used for 5’-RACE. Primers E2bot4 and E2bot5 match the 5’-end of the GPR40 open reading frame, and primer 44413 matches 5’-untranslated region sequences in exon 2. Orientation of primers is indicated. B, 5’-RACE results. DNase I-treated RNA from βTC1 cells was reverse transcribed using primer E2bot4. The cDNA was poly(dG)-tailed and amplified by PCR using a poly-C primer and primer E2bot5 (lane 1). As a positive control, the cDNA was amplified using primers E2bot5 and 44413 (lane 2). As a negative control PCR was carried out with RNA that was not reverse-transcribed (− RT, lane 3) or without RNA (− RNA, lane 4). C, vector used for generation of the RPA probe and its predicted protected products. The RNA probe is complementary to genomic sequence −52 to +24, containing 52 bp of the proximal 5’-flank and exon 1, and +723 to +856, containing 133 bp of the 5’-end of exon 2, according to the 5’-RACE result. D, RPA results. 50 μg of βTC1 RNA (lane 1) or αTC1 RNA (lane 2) was hybridized to the 32P-labeled RNA probe, then subjected to RNase digestion and separated on urea-polyacrylamide gel. The assay was also performed using 25 pg in vitro transcribed sense RNA as a positive control (lane 3) or 50 μg of yeast RNA as a negative control (lane 4). M indicates molecular weight marker. Full protection band (209 bases) corresponds to protection of the entire probe sequence derived from the GPR40 gene, whereas the positive control band (239 bases) corresponds to this full protection and an additional 30 bases derived from the vector used for in vitro transcription.

Fig. 6C was analyzed using the Wilcoxon two-sample test corrected for multiple comparisons by the Holms method. Values are shown as mean ± S.E.

RESULTS

Characterization of the GPR40 Transcription Start Site and Intron-Exon Organization—As a first step toward understanding the molecular mechanisms controlling cell-specific regulation of the mouse GPR40 gene, we identified the 5’-end of the GPR40 mRNA. 5’-RACE experiments using βTC1 RNA and primers complementary to the open reading frame of GPR40 (Fig. 1A) produced a major product of ~400 bp (Fig. 1B, lane 1). Comparison of its sequence with that of mouse genomic DNA revealed the presence of a 698-bp intron beginning 321 bp upstream of the ATG initiation codon, and a 24-bp first exon. Thus, the major TSS is located 1044 bp upstream of the translation initiation site. Additional weaker bands, suggested the existence of minor upstream transcription start sites (Fig. 1B, lane 1). Indeed, a 5’-RACE experiment performed with a primer starting 4 bases downstream to the major transcription start site gave rise to several products: sequencing of three of these showed distinct 5’-ends mapping 100–150 bp upstream of the major transcription start site (data not shown).

To confirm this result, RPAs were performed using a 278-base RNA probe comprising sequences corresponding to the 52 bp immediately upstream of the putative transcription start site, the 24 bp of exon 1 and the 133 bp at the 5’-end of exon 2 (Fig. 1C). A protected band of ~157 bases was observed using βTC1 RNA, but not αTC1 RNA (Fig. 1D, lanes 1 and 2), confirming the location of the start site indicated by the 5’-RACE result. An additional weak full protection band (209 bases) was also seen, consistent with the existence of the additional minor transcription start sites previously indicated by 5’-RACE.

Additional 5’-RACE experiments performed using primers distributed across the GPR40 gene (data not shown), and examination of expressed sequence tag (EST) sequences present in public databases, suggest that no additional introns are present in the gene. Taken together, these data establish the organization of the GPR40 gene, which is comprised of two exons of 24 and 4402 bp separated by a 698-bp intron. This arrangement of a short, non-coding first exon, followed by a long exon containing the complete protein coding information (Fig. 1A), is common in G protein-coupled receptors (32).

Characterization of the GPR40 Gene Promoter and Identification of Conserved Regions—Comparison of genomic DNA sequences of different species can often assist in identifying important control elements (33). We therefore compared the intergenic region between the CD22 gene and the GPR40 gene in mouse, rat, and human. The alignments revealed three conserved regions, which we named HR1–HR3 (Fig. 2A). HR1 is a part of an Alu repeat; it is 38 bp long, is 91% identical between mouse and human, and is located 2126 bp upstream from the
transcription start site. HR2 is 194 bp long; it is 82% identical between mouse and human and 93% identical between mouse and rat (Fig. 2B). HR3 is 185 bp long; it is 81% identical between mouse and human and 89% identical between mouse and rat (Fig. 2B). These sequences can serve as binding sites for basic helix-loop-helix (bHLH) transcription factors (34). Also present in the region is a sequence that is similar to the binding site for HNF4α (35) at −1017. Analysis of conserved sequences in the HR2 region of human, chimpanzee, rat, and mouse, performed using the rVista program and Genomatix suite programs MatInspector and FrameWorker, revealed in addition the binding sites for the transcription factors AP4, ETS family, RFX1, PBX1, MEIS1, Evi1, Bel1, and homeodomain proteins of the HOX family.

Transcription start sites and promoter elements are often conserved among species. Indeed, the HR3 homology region (Fig. 2C) spans the major transcription start site; hence it is expected to contain at least part of the promoter. Inspection of HR3 revealed characteristic features of a TATA-less promoter, including a GC-rich region (64% in mouse and 63% in human) and two putative Sp1 binding sites (Fig. 2C). The sequence at the GPR40 TSS matches the consensus sequence for the initiator (Int) element (PyPyA+1NT/APyPy) (36) except for the last two pyrimidines. The rest of the sequence, including the A at the +1 position, matches in all three species (Fig. 2C). In addition, the sequence between +28 to +32 matches the DPE consensus sequence (36) in all three species. Indeed, DPE-containing promoters typically possess also the Int element, and the DPE element is located exactly at positions +28 to +32 relative to the A at +1 of the Int (36). In DPE-Int promoters there is a preference for G at position +24, and in fact G is present at position +24 in the GPR40 promoter of all three species. Thus, the GPR40 core promoter displays the typical features of a TATA-less, Int-DPE core promoter.
We also identified in HR3 a conserved A-element (TAAT) at position 89 to 86, which may serve as a binding site for transcription factors of the homeodomain family (Fig. 2C). Other conserved potential transcription factor binding sites found in HR3 using the rVista program include AP2, AP4, ELK1, GATA family, IK2, MZF1, PEA3, and STAT family.

Chromatin Accessibility at the 5′-Flank of the GPR40 Gene—To compare the chromatin structure at the GPR40 locus in β versus non-β cells, we tested the DNase I sensitivity of nuclei from βTC1 and NIH-3T3 cells. We focused on an 8.8-kb KpnI genomic fragment spanning exon 9 of CD22 to the middle of the GPR40 coding sequence, including 7.5 kb upstream to the GPR40 TSS and the GPR40 intron (Fig. 3A). A radiolabeled probe of 870 bp corresponding to the 3′-end of the region (Fig. 3A) was used for identification of the genomic band and sub-bands.

The analysis showed that in βTC1 cells, which express GPR40, the accessibility of chromatin to DNase I was higher than in the non-expressing cell line NIH-3T3, as indicated by the greater sensitivity of the genomic fragment to nuclease digestion (Fig. 3, B and C). Furthermore, although in NIH-3T3 cells no sub-bands appeared following DNase I digestion, clear sub-bands of 2.5 kb and 1.5 kb were seen in βTC1 cells (Fig. 3B), indicating sites of DNase I hypersensitivity. Because the radiolabeled probe we used corresponds to the terminus of the examined genomic region, we were able to determine the location of these hypersensitive sites. Interestingly, they correlated extremely well with the HR2 and HR3 homology regions. Because the GPR40 core promoter is located in the HR3 region, it is expected that in β cells this region will be highly accessible. In addition, the hypersensitivity of the HR2 region is consistent with an important function for this region.

Transcription Regulation Conferred by Sequences from the 5′-Flank of the GPR40 Gene—To determine whether the 5′-flanking
The ability of conserved region HR2 to activate transcription from a heterologous promoter activity was also examined. For this purpose, a 258-bp fragment spanning HR2 was fused to a herpes simplex virus thymidine kinase (TK) promoter controlling the firefly luciferase reporter gene. HR2-TK-LUC had 11.4-fold higher activity than TK-LUC in the β cell line HIT, whereas in the non-β cell line CHO, the HR2 region had little or no effect on the TK promoter (Fig. 4C). The HR2 region was able to activate the TK promoter in an orientation-independent fashion (data not shown). Thus, HR2 contains a potent β-cell-specific enhancer.

**Systematic Mutagenesis of the HR2 Region**—Based on the 5’ deletion analysis, and the evolutionary conservation of the HR2 element, it is likely that the 5’ 100 bp of the HR2 element (−1191 to −1091) contains a critical control sequence. This 100-bp region was therefore systematically mutated to define key transcriptional control elements. Accordingly, a series of ten substitution mutations (M1–M10), each spanning 10 bp, was generated in 2095-GPR40P-LUC (Fig. 5A), and their transcriptional activity following transfection to HIT cells was compared with the wild-type promoter (Fig. 5B). Striking reduc-
tions in activity were observed with the adjacent mutations M4 and M5 (remaining activity, 38 and 40%, respectively). Several mutations showed smaller, yet statistically significant reduction in activity. The observed activities were: M1, 59%; M8, 74%; and M9, 59%. Thus, the analysis defines three sub-regions within HR2 that are important for transcription activation: sub-region 1, sub-region 4–5, and sub-region 8–9.

To examine the transcriptional activity of the cis-elements in the HR2 region independently of the remainder of the promoter, the same series of substitution mutations of HR2 was introduced to HR2-TK-LUC and tested in the β cell line HIT. The results obtained were similar to those of 2095-GPR40 mutants (Fig. 5C).

Binding of PDX1 to HR2 in Vitro and in Vivo—Bioinformatic analysis identified within sub-region 4–5 potential binding sites for transcription factors of the homeodomain family (Fig. 5A). We therefore tested whether PDX1, an important β cell-specific homeodomain protein can bind this sequence in vitro using EMSA. We used a radioactive probe (designated HR2-2-6) that contains the wild-type sequence of region HR2 from the middle of sub-region 2 to the middle of sub-region 6. Incubation with nuclear extracts from 293T cells transfected with a PDX1 expression vector, generated a strong band (Fig. 6A, lane 2). Incubation of the probe with βTC1 cell nuclear extracts yielded a band of similar mobility, whereas incubation with NIH-3T3 cell extracts did not show this band (Fig. 6A, lanes 3 and 4). Addition of anti-PDX1 anti-serum abolished the band (Fig. 6A, lane 6), whereas addition of PBS, pre-immune serum (anti-P33) or irrelevant anti-serum (anti-N22) had no effect (Fig. 6A, lanes 5, 7, and 8). To examine the specificity of binding, we performed quantitative competition experiments using increasing concentrations of unlabeled oligonucleotides (Fig. 6B). The oligonucleotides used were HR2-2-6, HR2-M4-5 (which contains HR2-2-6 sequence mutated at sub-regions 4 and 5), and E1 (which contains an E-box element from the rat insulin I promoter, and thus is not expected to bind PDX1). HR2-M4-5 and E1 oligonucleotides showed significantly reduced ability to compete for binding (Fig. 6B), indicating that PDX1 binds in a sequence-specific manner and with highest affinity to the wild-type sequence of sub-region 4 to 5 within HR2.

To determine whether PDX1 binds to the HR2 enhancer in vivo, we performed ChIP experiments using anti-PDX1 or control antibody. PDX1 occupancy of the HR2 region in βTC1 cells was significantly higher than its occupancy of GPR40 or CD22 coding sequences (CDS), which reside within the vicinity of HR2, but are not expected to bind PDX1 (Fig. 6C). In contrast, PDX1 occupancy on HR2 was similar to its occupancy on the promoter of the mouse insulin II gene, a bona fide PDX1 target gene (data not shown). As control, we used pre-immune serum and observed significantly lower binding to HR2 and the insulin II gene promoter as compared with anti-PDX1 antibody (Fig. 6C and data not shown).

Binding of BETA2 to HR2 in Vitro and in Vivo—Mutation M9 overlaps the 5’ E-box of the HR2 region. Because E-box ele-
Regulation of the GPR40 Gene

The pancreatic β cell plays a crucial role in maintaining metabolic homeostasis through the production and regulated secretion of the key metabolic hormone insulin in response to physiological needs (39). Inadequate production of insulin is a hallmark of both major forms of diabetes mellitus. Insulin secretion is regulated mainly according to fluctuating levels of blood glucose but also in response to nutrients, neurotransmit-

ments represent potential binding sites for transcription factors of the bHLH family, we examined whether BETA2 (NeuroD1), an important β cell-specific bHLH protein (37), can bind this sequence in vitro using EMSA. BETA2 typically binds E-box sequences as a heterodimeric complex with products of the ubiquitously expressed E2A gene (38). We used a radioactive probe (designated HR2-7-10) that contains the wild-type sequence of HR2 from sub-region 7 to the end of sub-region 10. Incubation of the probe with βTC1 cell nuclear extract yielded a slow migrating complex (Fig. 7A, lane 1). Addition of anti-BETA2 antibody to the incubation mixture resulted in a supershift of the complex, whereas addition of pre-immune serum (anti-P33) did not (Fig. 7A, lanes 2 and 3). Incubation of the probe with nuclear extracts from 293T cells did not generate this complex (Fig. 7A, lane 4), whereas 293T cells transfected with a BETA2 expression vector generated a clear band whose mobility is similar to the complex observed with βTC1 extract (Fig. 7A, lane 5). The complex was supershifted upon addition of anti-BETA2 antibody (Fig. 7A, lane 6).

To examine the specificity of binding, we performed quantitative competition experiments using increasing concentrations of unlabeled oligonucleotides (Fig. 7B). The oligonucleotides used were HR2-7-10, HR2-EM (which contains a mutated sequence of the E-box element), and P1 (which contains an A-element from the rat insulin 1 gene promoter, and thus is not expected to bind BETA2). P1 and HR2-EM oligonucleotides showed significantly reduced ability to compete for binding (Fig. 7B), indicating that BETA2 is capable of binding to sub-region 8–9 within HR2 with high affinity and in a sequence-specific manner.

To determine whether BETA2 binds to the HR2 enhancer in vivo, we performed ChIP experiments using anti-BETA2 antibody. BETA2 occupancy of the HR2 region in βTC1 cells was significantly higher than its occupancy of the CDS of GPR40, CD22, or the HR3 region, none of which were expected to bind BETA2 (Fig. 7C). The signal observed with control antibody was low for all regions tested (Fig. 7C).

DISCUSSION

The pancreatic β cell plays a crucial role in maintaining metabolic homeostasis through the production and regulated secretion of the key metabolic hormone insulin in response to physiological needs (39). Inadequate production of insulin is a hallmark of both major forms of diabetes mellitus. Insulin secretion is regulated mainly according to fluctuating levels of blood glucose but also in response to nutrients, neurotransmit-
Regulation of the GPR40 Gene

A.

BETA2 binds the HR2 region of the GPR40 promoter in vitro and in vivo. A, EMSA analysis. Nuclear extracts from βTC1 cells (lanes 1–3), untransfected 293T cells (lane 4), and 293T cells transfected with BETA2 expression vector (lanes 5 and 6) were incubated with 32P-labeled HR2-7-10 probe and resolved on a non-denaturing polyacrylamide gel. The incubation reactions were performed in the presence of anti-BETA2 antibody (lanes 2 and 6) or pre-immune (lane 3). B, competition experiments were performed by incubation of nuclear extracts from βTC1 cells with 32P-labeled HR2-7-10 probe and increasing concentrations of unlabeled P1, HR2-EM, or HR2-7-10 oligonucleotides. Binding is expressed relative to that observed in the absence of unlabeled oligonucleotides. C, ChIP. Cross-linked chromatin from βTC1 cells was immunoprecipitated using either anti-BETA2 antibodies or pre-immune serum. DNA fragments recovered from immunoprecipitations were amplified using specific sets of primers corresponding to the βTC cell-specific enhancer HR2, the coding sequence (CDS) of GPR40 and CD22 genes, and the GPR40 core promoter HR3. Binding of antisera to the tested regions was analyzed using quantitative real-time PCR and presented as percentage recovery of input chromatin. Results are expressed as the mean ± S.E. of three or more independent ChIP experiments. Asterisks indicate statistical significance (p < 0.05) compared with control antibody and compared with GPR40 CDS, CD22 CDS, and HR3.

To fulfill their complex role, β cells express a characteristic profile of genes, including many that are expressed selectively in β cells. Regulation of these genes is mediated in large part through transcriptional control mechanisms involving the combinatorial actions of ubiquitous and cell-specific transcription factors. It has recently been shown that the GPR40 gene is expressed selectively in β cells, where it appears to play an important role in mediating the effects of fatty acids on insulin secretion.

In this study, we have examined the molecular basis for the selectivity of GPR40 expression in the β cell. We have defined the organization of the gene and the location of the core promoter, which lacks a TATA element but contains an Int-like element and a DPE sequence. Reporter gene experiments demonstrated that β cell-specific expression of the GPR40 gene is mediated through selective transcription via the 5'-flanking region of the gene. Cell-specific transcription of mammalian genes generally involves evolutionarily conserved sequences located upstream of the core promoter. Indeed, we demonstrated that the conserved HR2 element, which coincides with a nuclease hypersensitive site, functions as a strong β cell-specific transcriptional enhancer.

Systematic mutagenesis of the HR2 enhancer, revealed three important domains. One of these (sub-region 8–9) contains an E-box element that can bind transcription factors of the bHLH family. Indeed, BETA2, a β cell-specific bHLH protein, implicated in regulation of several β cell genes, was shown to bind in vitro to sub-region 8–9 of HR2 in a sequence-specific manner, and was found to occupy the HR2 region in vivo. A second domain (sub-region 3–5) contains an A-element that can bind transcription factors of the homeodomain family. We were able to demonstrate that PDX1, a crucial β cell homeodomain transcription factor, binds to this region both in vitro and in vivo.

These results represent the first characterization of the GPR40 gene promoter. Like other β cell-specific genes, cell specificity is controlled at the transcriptional level through enhancer elements located in the 5' flank. This may be mediated by interactions among PDX1, BETA2, and additional transcription factors. Accumulating evidence has indicated a critical role for GPR40 in mediating the responsiveness of β cells to LCFAs; furthermore, persistent activation of GPR40 in obesity has been proposed as a mechanism to explain the well-established link between obesity and type 2 diabetes.
centrations of the protein. Thus, regulated transcription of the gene is probably essential for maintaining β cell function, and defects could lead to perturbed insulin secretion. The present study has elucidated the molecular basis for understanding β cell-specific transcription of GPR40 gene expression. Future studies will be aimed to determining the detailed mechanisms modulating GPR40 gene transcription in normal β cells and to investigate possible defects in diabetes.

Acknowledgments—We thank Dr. Dena Leshkowitz and Dr. Shifra Ben-Dor for valuable bioinformatics advice, Dr. Hillary Voet for statistical consultation, and Sara Weiss for expert technical assistance.

REFERENCES

1. Kahn, B. B. (1998) Cell 92, 593–596
2. Zimmet, P., Alberti, K. G., and Shaw, J. (2001) Nature 414, 782–787
3. Lazar, M. A. (2005) Science 307, 373–375
4. Moller, D. E., and Kaufman, K. D. (2005) Annu. Rev. Med. 56, 45–62
5. Poitout, V. (2003) Trends Endocrinol. Metab. 14, 201–203
6. Danforth, A. C., Efrat, S., Mojssov, S., Spector, D., Habener, J. F., and Hansa-

7. Poitout, V., and Robertson, R. P. (2002) Endocrinology 143, 339–342
8. Brown, A. J., Jupe, S., and Briscoe, C. P. (2005) DNA Cell Biol. 24, 54–61
9. Itoh, Y., Kawamata, Y., Harada, M., Kobayashi, M., Fujii, R., Fukusumi, S.,

10. Kotarsky, K., Nilsson, N. E., Flodgren, E., Owman, C., and Olde, B. (2003) Biochem. Biophys. Res. Commun. 301, 406–410
11. Briscoe, C. P., Tadayyon, M., Andrews, J. L., Benson, W. G., Chambers,

12. Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L.,

13. Nilsson, N. E., Kotarsky, K., Owman, C., and Olde, B. (2003) Biochem. Biophys. Res. Commun. 303, 1047–1052
14. Poe, J. C., Fujimoto, Y., Hasegawa, M., Haas, K. M., Miller, A. S., Sanford,

15. Steeneberg, P., Rubins, N., Bartoo-Shifman, R., Walker, M. D., and Edlund,

16. Flodgren, E., Olde, B., Meidute-Abaravici, S., Winzell, M. S., Ahren, B., and Salehi, A. (2007) Biochem. Biophys. Res. Commun. 354, 240–245
17. Shapiro, H., Shachar, S., Sekler, I., Hershfield, M., and Walker, M. D.

18. Salehi, A., Flodgren, E., Nilsson, N. E., Jiménez-Feltstrom, I., Miyazaki, J., Owman, C., and Olde, B. (2005) Cell Tissue Res. 322, 207–215
19. Fujiwara, K., Maekawa, F., and Yada, T. (2005) Am. J. Physiol. 289, 670–677
20. Ausubel, M. F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G.,

21. Efrat, S., Linde, S., Kofod, H., Spector, D., Delannoy, M., Grant, S., Han-

22. Powers, A. C., Efrat, S., Mojssov, S., Spector, D., Habener, J. F., and Han-

23. Santerre, R. F., Cook, R. A., Crisel, R. M. D., Sharp, J. D., Schmidt, R. J.,

24. Edlund, T., Walker, M. D., Barr, P. J., and Rutter, W. J. (1985) Science 230, 912–916
25. Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Uralab, G., and Chasin, L.

26. Aronheim, A., Ohlsson, H., Park, C. W., Edlund, T., and Walker, M. D.

27. Bradford, M. (1976) Anal. Biochem. 72, 248–254
28. Glick, E., Leshkowitz, D., and Walker, M. D. (2000) J. Biol. Chem. 275, 2199–2204
29. German, M., Ashcroft, S., Docherty, K., Edlund, T., Edlund, H., Goodison,

30. Liberonzo, A., Rodier, G., and Walker, M. D. (2004) Nucleic Acids Res. 32, 54–64
31. Aihbinder, E., Revach, M., Wolstein, O., Moshonov, S., Diamant, N., and

32. Sosinsky, A., Glusman, G., and Lancet, D. (2001) Genomics 70, 49–61
33. Pennacchio, L. A., and Rubin, E. M. (2001) Nat. Rev. Genet. 2, 100–109
34. Kadesch, T. (1992) Immunol. Today 13, 31–36
35. Deryckere, F., McNair, A., and Gannon, F. (1996) Gene (Amst.) 175, 35–41
36. Smale, S. T., and Kadonaga, J. T. (2003) Annu. Rev. Biochem. 72, 449–479
37. Naya, F. J., Huang, H. P., Mutoh, H., DeMayo, F. J., Leiter, A. B.,

38. Park, C. W., and Walker, M. D. (1992) J. Biol. Chem. 267, 15642–15649
39. Hutton, J. C., Bailyes, E. M., Rhodes, C. J., Rutherford, N. G., Arden, S. D.,

40. Henquin, J. C., Ravier, M. A., Nenquin, M., Jonas, J. C., and Gilon, P. (2003) Eur. J. Clin. Invest. 33, 742–750
41. Wilson, M. E., Scheel, D., and German, M. S. (2003) Mech. Dev 120, 65–80
42. Rutter, G. A. (2003) Curr. Biol. 13, R403–405