

## Multiple Elements in the Upstream Glucokinase Promoter Contribute to Transcription in Insulinoma Cells

KATHY D. SHELTON, ALAN J. FRANKLIN, ANDRAS KHOOR,† JOSEPH BEECHEM,  
AND MARK A. MAGNUSON\*

*Departments of Molecular Physiology and Biophysics and of Medicine,  
Vanderbilt University Medical School, Nashville, Tennessee 37232*

Received 28 February 1992/Returned for modification 31 March 1992/Accepted 23 July 1992

**$\beta$ -cell type-specific expression of the upstream glucokinase promoter was studied by transfection of fusion genes and analysis of DNA-protein interactions. A construct containing 1,000 bp of 5'-flanking DNA was efficiently expressed in HIT M2.2.2 cells, a  $\beta$ -cell-derived line that makes both insulin and glucokinase, but not in NIH 3T3 cells, a heterologous cell line. In a series of 5' deletion mutations between bases -1000 and -100 (relative to a base previously designated +1), efficient expression in HIT cells was maintained until -280 bp, after which transcription decreased in a stepwise manner. The sequences between -280 and -1 bp contributing to transcriptional activity in HIT cells were identified by studying 28 block transversion mutants that spanned this region in 10-bp steps. Two mutations reduced transcription 10-fold or more, while six reduced transcription between 3- and 10-fold. Three mutationally sensitive regions of this promoter were found to bind to a factor that was expressed preferentially in pancreatic islet  $\beta$  cells. The binding sites, designated upstream promoter elements (UPEs), shared a consensus sequence of CAT(T/C)A(C/G). Methylation of adenine and guanine residues within this sequence prevented binding of the  $\beta$ -cell factor, as did mutations at positions 2, 3, and 5. Analysis of nuclear extracts from different cell lines identified UPE-binding activity in HIT M2.2.2 and  $\beta$ -TC-3 cells but not in AtT-20, NIH 3T3, or HeLa cells; the possibility of a greatly reduced amount in  $\alpha$ -TC-6 cells could not be excluded. UV laser cross-linking experiments supported the  $\beta$ -cell type expression of this factor and showed it to be ~50 kDa in size. Gel mobility shift competition experiments showed that this  $\beta$ -cell factor is the same that binds to similar elements, termed CT boxes, in the insulin promoter. Thus, a role for these elements (UPEs or CT boxes), and the  $\beta$ -cell factor that binds to them, in determining the expression of genes in the  $\beta$  cells of pancreatic islets is suggested.**

An understanding of cell type-specific gene expression requires identification and characterization of the *cis*-acting elements and *trans*-acting factors involved. For a given cell type, this can be achieved by first identifying genes expressed specifically or preferentially in the particular cell type and then characterizing the particular elements utilized and the factors that bind to them. In pancreatic  $\beta$  cells, the insulin gene has been, until recently, the only available model system for such studies. While progress has been made by several groups in determining the *cis*-acting elements and *trans*-acting factors that determine expression of the rat and human insulin genes (2, 7, 16, 17, 24, 31), studies of a single gene are likely to result in only a limited understanding of the elements and factors involved in regulating  $\beta$ -cell type-specific gene expression. A more comprehensive understanding of the elements and factors necessary for gene expression in pancreatic  $\beta$  cells can be achieved by identifying and studying additional genes that are transcribed primarily or exclusively in this cell type. Thus, we have begun an analysis of the *cis*-acting element(s) and *trans*-acting factor(s) responsible for cell type-specific expression of the upstream glucokinase promoter in  $\beta$  cells.

Glucokinase (hexokinase type IV) plays a key role in maintaining blood glucose homeostasis by catalyzing the high- $K_m$  phosphorylation of glucose, the proximal and rate-limiting step in the utilization of glucose by liver and

pancreatic  $\beta$  cells (22). Two alternative transcription control regions which lie at least 12 kb apart in a single glucokinase gene are involved in determining the cell type-specific expression of glucokinase mRNA (20). The upstream glucokinase promoter region is active in both pancreatic  $\beta$  cells and the anterior pituitary, while the downstream glucokinase promoter is active only in hepatocytes (13, 14, 19, 21). Despite the expression of glucokinase mRNA in the pituitary, there is no glucokinase activity in this tissue because of alternative RNA splicing of the glucokinase gene transcript in the pituitary that alters the reading frame of the protein (19).

In pancreatic islets, glucokinase immunoreactivity is limited to  $\beta$  cells. However, among different  $\beta$  cells, glucokinase expression is heterogeneous (15). This and other findings suggest that glucokinase gene expression in pancreatic  $\beta$  cells is under regulatory control different from that of the insulin gene. Thus, information obtained by studying  $\beta$ -cell type-specific expression of the upstream glucokinase promoter is likely to complement data obtained by studying the insulin promoter, as well as characterizing the regulation of this important metabolic enzyme. To identify the elements and factors responsible for  $\beta$ -cell-specific expression of the glucokinase gene, we undertook a detailed fusion gene analysis of the upstream glucokinase transcription control region. Both 5' deletion and block transversion mutational strategies were used to identify sequences that contribute to transcriptional activity of this promoter in insulinoma cells, and DNA-protein interactions at three sites were studied in detail. We identified an element, termed the upstream promoter element (UPE), that bound a factor of ~50 kDa which

\* Corresponding author.

† Present address: Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN 37232.

was preferentially expressed in pancreatic  $\beta$  cells. UPEs were found to be the same as CT boxes, elements that have been identified in insulin promoter sequences. Together, these findings suggest that this element and the factor that binds to it are involved in  $\beta$ -cell type-specific regulation of both the insulin and glucokinase genes.

## MATERIALS AND METHODS

**General procedures.** Oligonucleotides were synthesized by either the Vanderbilt DNA Core Laboratory or Midland Reagent Co. (Midland, Tex.). All DNA manipulations were done by standard procedures (26) unless otherwise stated.

**Plasmids.** Luciferase expression vector pSVOAPL2L (5) contains the luciferase insert L-AA5' (8) and a polylinker with recognition sites for *Kpn*I, *Sac*I, *Sma*I, *Hind*III, and *Bgl*II. Two simian virus 40-derived splice and polyadenylation sequences in pSVOAPL2L, upstream of the polylinker, were utilized to reduce readthrough transcription from cryptic initiation sites in the procaryotic DNA sequences (8). prInsII-luc was made by removing the rat insulin II promoter sequences in pPLFCAT (provided by R. Stein, Department of Molecular Physiology and Biophysics, Vanderbilt University [31]) by digestion with *Hind*III and *Bgl*II and then ligation of the fragment into the same sites of pSVOAPL2L. The resulting rat insulin II promoter fragment contained sequences from -342 to +8 and a neutral mutation at -100 that does not affect expression in cultured insulinoma cells (31). A Rous sarcoma virus (RSV) promoter-luciferase fusion gene (pRSV-luc) was obtained from S. Subramani (Department of Biology, University of California at San Diego [8]). All plasmids were propagated in *Escherichia coli* and purified by alkaline lysis, followed by two isopycnic CsCl gradient centrifugations.

**Linker-primer PCR amplifications.** Glucokinase promoter fragments with different 5' and 3' endpoints were generated by using linker-primer polymerase chain reactions (PCR). The oligonucleotide primers used contained 18 bp of a hybridizing sequence, a 6-bp enzyme recognition site, and two additional nucleotides at the 5' ends to increase restriction efficiency. A *Kpn*I site was introduced into the top-strand primers, while a *Bgl*II site was inserted into the bottom-strand primers. DNA amplification reactions (100  $\mu$ l) took place in a DNA thermocycler (Perkin Elmer Cetus) under standard conditions (25) using *Thermus aquaticus* DNA polymerase (AmpliTaq; Perkin Elmer Cetus). From 1 to 10 ng of pGK7.B2 (21) DNA was used as the template for amplification in the presence of 100 pmol of each primer. Twenty-five temperature cycles were performed by using a step cycle program of 94°C for 1 min, 42°C for 1 min, and 72°C for either 1 or 2 min, and then the amplified DNA fragment was extracted with phenol and separated from unused primers by spun-column chromatography through Sephacryl S-200. After digestion with *Kpn*I and *Bgl*II, the DNA fragment was extracted with phenol, precipitated with ethanol, and ligated into the corresponding sites of pSVOAPL2L.

**Block transversion mutagenesis.** Site-specific mutagenesis by the method of Kunkel (18) was used to alter regions of the glucokinase promoter 10 bp at a time. The mutating oligonucleotides were each 34 bases long and consisted of 12 unaltered bases on each side of the 10 bases that were mutated. The base transversions generally followed the strategy A $\leftrightarrow$ C and G $\leftrightarrow$ T, except for two stimulatory mutations (-230/-221 and -180/-171) for which additional mutations were made by using a different sequence. The -230/

-221 region of the upstream glucokinase promoter was changed to ACCTGCAGTG, creating a *Pst*I recognition site, and the -180/-171 region was changed to GCAAGCTTGA, creating a *Hind*III recognition site. All site-specific mutations were confirmed by dideoxy DNA sequence analysis.

**RNA blot transfer analysis.** Total cellular RNA was isolated by homogenization of cells in guanidinium isothiocyanate and centrifugation through a 6 M CsCl cushion (3). Poly(A)<sup>+</sup> RNA was obtained by chromatography on oligo(dT)-cellulose (Type II; Collaborative Research) and size fractionated in a 1.0% agarose gel containing 6.3% formaldehyde, 1 mM EDTA, and 20 mM MOPS (morpholinepropanesulfonic acid) at pH 7.4. The RNA was blotted to nitrocellulose, bound to the filter by UV irradiation, and prehybridized for 6 h in a buffer containing 50% formamide, 5 $\times$  SSPE (1 $\times$  SSPE is 0.18 M NaCl, 10 mM NaPO<sub>4</sub> [pH 7.4], and 1 mM EDTA), 5 $\times$  Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% albumin), 0.1% sodium dodecyl sulfate (SDS), and 100  $\mu$ g of sheared, denatured salmon testis DNA per ml. One million counts per minute of a denatured glucokinase cDNA probe per milliliter was added and allowed to hybridize for 48 h. The filter was subsequently washed three times for 20 min each time at 45°C with 0.2 $\times$  SSC (1 $\times$  SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate)-0.1% SDS, and autoradiography was performed.

**Transfection assays.** HIT M2.2.2 cells (9, 27), a simian virus 40-transformed hamster  $\beta$ -cell line, and NIH 3T3 cells, a mouse fibroblast line, were grown in Dulbecco's modified medium supplemented with 15% (vol/vol) horse serum, 2.5% (vol/vol) fetal calf serum, and 50  $\mu$ g each of streptomycin and penicillin per ml. Cells were plated onto 75-cm<sup>2</sup> flasks 1 day prior to transfection by using the calcium phosphate-DNA coprecipitation method (11). One milliliter of CaPO<sub>4</sub>-DNA precipitate was made by using 10  $\mu$ g of test plasmid DNA and 1  $\mu$ g of a plasmid containing an RSV- $\beta$ -galactosidase fusion gene. The precipitate was incubated with cells for 4 h, after which the cells were exposed to 20% (vol/vol) glycerol for 2 min and then rinsed with serum-free medium. Fresh medium containing serum was added, and the cells were incubated at 37°C for 40 h. The cells were harvested by being rinsed twice with phosphate-buffered saline and addition of 1 ml of trypsin-EDTA. After several minutes, 3 ml of medium containing serum was added and the cells were collected by centrifugation at 300  $\times$  g at 4°C for 5 min, washed with 2 ml of cold phosphate-buffered saline, and then centrifuged again. The supernatant was aspirated, and the cell pellets were suspended in 1 mM phenylmethylsulfonyl fluoride-250 mM Tris (pH 7.8) and transferred to a clean 1.5-ml tube. A cellular lysate was prepared by sonication at the lowest setting for 10 s. Debris was removed by centrifugation at 12,000  $\times$  g for 15 min at 4°C. The supernatant was gently aspirated, transferred to a clean 1.5-ml tube and kept at 4°C until assayed for luciferase and  $\beta$ -galactosidase activities. Luciferase activity was determined by the method of de Wet et al. (8). Light measurements were performed with an AMINCO photometer coupled to a chart recorder. Reactions containing 10 to 70  $\mu$ l of cellular extract were initiated by injecting 100  $\mu$ l of 1 mM luciferin in 0.4 M NaHCO<sub>3</sub> into the reaction chamber. The light output (determined by peak height) of each extract was determined at least twice, and the average was taken.  $\beta$ -Galactosidase activities were determined spectrophotometrically with *o*-nitrophenyl- $\beta$ -D-galactopyranoside as the substrate (23).

**Transcription start site analysis.** HIT M2.2.2 cells were transfected with 20  $\mu$ g of DNA per 150-cm<sup>2</sup> plate, and total

RNA was harvested 24 h later (4). Two hundred micrograms of total RNA was treated with 5 U of RNase-free DNase I (RQ DNase; Promega) for 15 min at 37°C, extracted with phenol and phenol-CHCl<sub>3</sub>, and precipitated with ethanol. The entire pellet of total RNA was used to make first-strand cDNA in a 50- $\mu$ l reaction containing 2.5  $\mu$ g of oligo(dT)<sub>12-18</sub>, a 0.5 mM concentration of each deoxynucleoside triphosphate, 500 U of Moloney murine leukemia virus reverse transcriptase, and the buffer provided by the supplier (Bethesda Research Laboratories). The reaction was incubated at 37°C for 1 h and then terminated by freezing at -20°C. PCRs (50  $\mu$ l) were performed by using 2.5  $\mu$ l of the first-strand cDNA and 25 temperature cycles with a step cycle program of 94°C for 1 min, 50°C for 1 min, and 72°C for 1.5 min. A 15- $\mu$ l aliquot of the amplified DNA was fractionated on a 1.5% agarose gel, denatured, and blot transferred to a nitrocellulose filter. After UV fixation, the blot was probed with a <sup>32</sup>P-end-labeled oligonucleotide probe with sequences complementary to the amplified DNA product. Hybridization was performed overnight at 40°C in a buffer containing 2 $\times$  SSC, 10 $\times$  Denhardt's solution, and 0.15 mg of salmon testis DNA per ml. The blot was washed first in 2 $\times$  SSC for 5 min at 40°C and then three times for 20 min each at 30°C prior to autoradiography.

**Nuclear extracts.** Nuclear extracts were prepared by the method of Shapiro et al. (29), except that protease inhibitors were added. A 1:500 dilution of an inhibitor cocktail containing aprotinin (1 mg/ml), antipain (1 mg/ml), leupeptin (1 mg/ml), chymostatin (1 mg/ml), pepstatin A (1 mg/ml), trypsin-chymotrypsin inhibitor (5 mg/ml), and benzamide (50 mM) was added to the hypotonic buffer used during extract preparation. The nuclear extracts were aliquoted after dialysis and stored at -70°C until they were used. The integrity of each extract was assessed by examining the pattern of binding to a probe for a ubiquitous CCAAT transcription factor (10).

**Gel mobility shift assays.** Oligonucleotides for mobility shift assays were synthesized with a 5' TCGA overhang on each end and purified by elution from 18% acrylamide gels. The oligonucleotide pairs were annealed by mixing equal portions (~10  $\mu$ g) in 10 mM Tris (pH 8.0)-1 mM EDTA-250 mM KCl and then heating them to 100°C for 10 min. After 10 min, the heat was turned off and the bath was allowed to cool to room temperature. Klenow fill-in reactions using [ $\alpha$ -<sup>32</sup>P]dATP (3,000 Ci/mmol) were performed by using 50 pmol of the annealed oligonucleotide pairs; probes were purified by chromatography in 2-ml Sephadex G-50 columns. Protein binding reactions contained 12.5 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.8), 2.5 mM dithiothreitol, 0.1 mM EDTA, 5% glycerol, 1  $\mu$ g of poly(dI-dC), 2  $\mu$ g of denatured salmon sperm DNA, 10 mM KCl, approximately 4  $\times$  10<sup>4</sup> cpm of the <sup>32</sup>P-labeled probe, and 20  $\mu$ g of nuclear extract. Reactions were incubated at 4°C for 20 min, and then 2  $\mu$ l of loading buffer was added and the sample was fractionated on a 4% acrylamide gel pre-cooled to 4°C in a buffer of 250 mM Tris-1.9 M glycine-10 mM EDTA.

**UV laser cross-linking.** Protein-DNA-binding reactions were performed as described for the gel mobility shift assays, except that 2  $\times$  10<sup>5</sup> cpm of the radiolabeled probe was added to 40  $\mu$ g of nuclear extract in a total volume of 40  $\mu$ l. The binding reactions were incubated at 4°C for 30 min and then irradiated with the output of a frequency-quadrupled, 266-nm, pulsed Nd:YAG laser (Coherent, Palo Alto, Calif.). The net energy output of the laser was ~3 mJ/s. After cross-linking with 600 mJ, 20  $\mu$ l of the binding reaction was

mixed in a 1:1 ratio with SDS-(polyacrylamide gel electrophoresis (PAGE) sample buffer, incubated at 100°C for 5 min, and fractionated by SDS-PAGE on a 7.5% gel. The gels were then dried, and autoradiography was performed.

**Methylation interference.** The top or bottom strand of an oligonucleotide pair that spanned from -111 to -81 (200 ng) was labeled with T4 polynucleotide kinase (New England BioLabs) and [ $\gamma$ -<sup>32</sup>P]ATP (>5,000 Ci/mmol, Amersham) for 1 h at 37°C. The kinase was inactivated by being heated to 100°C for 5 min, and the labeled oligonucleotide was annealed to a 2 M excess of the complementary strand by being heated to 100°C for 10 min and then cooled to room temperature over a 30-min period. The resulting oligonucleotide pair was methylated by addition of 50 mM sodium cacodylate and 2  $\mu$ l of dimethyl sulfate for 10 min at 22°C. Sodium acetate (pH 7.0) and 2-mercaptoethanol were added to total concentrations of 500 and 333 mM, respectively, to arrest methylation. The unincorporated nucleotides were subsequently separated by G-50 Sephadex chromatography. Binding reactions containing 2.5  $\times$  10<sup>6</sup> cpm of the annealed probe and 80  $\mu$ g of nuclear extract were performed as described for the gel mobility shift assays. After electrophoresis on a 4% polyacrylamide gel, the bound and free probes were identified by autoradiography and excised. Gel slices were cast into a 1% agarose gel, and the DNA was electrophoresed onto a DEAE membrane and eluted by incubation twice in 150  $\mu$ l of elution buffer containing 1 M sodium chloride, 20 mM Tris (pH 8.0), and 0.1 mM EDTA at 65°C. The eluted DNA was extracted once with phenol-chloroform (1:1) and precipitated with 2.5 volumes of ethanol. The precipitated DNA was suspended in 50  $\mu$ l of 1 M piperidine and incubated at 90°C for 30 min. The piperidine was vacuum evaporated and the DNA pellet was suspended twice in 10  $\mu$ l of H<sub>2</sub>O, which was followed each time by vacuum evaporation. Sequencing loading buffer was added, and an equivalent number of counts per minute of each sample was fractionated on a 20% denaturing polyacrylamide gel. The gel was covered with cellophane wrap, and autoradiography at -70°C was performed.

## RESULTS

**Cell type-specific expression of the upstream glucokinase promoter region.** HIT M2.2.2 cells, an insulin-secreting  $\beta$ -cell-derived line (9, 27), were examined for glucokinase mRNA expression by Northern (RNA) blot analysis and found to express the message (Fig. 1, lane 1). In contrast, NIH 3T3 cells, a fibroblast-derived cell line, did not express glucokinase mRNA (Fig. 1, lane 2). The mRNA detected in HIT M2.2.2 cells was approximately 200 bases longer than that detected in hepatic RNA (Fig. 1, lane 3), indicating that transcription of the glucokinase gene initiated from the upstream transcription control region. A fusion gene containing glucokinase DNA sequences from -1000 to +14, linked to the firefly luciferase gene ( $\beta$ GK-1000luc), was transiently expressed in both HIT M2.2.2 and NIH 3T3 cells, and luciferase activities, normalized to the expression of a second reporter, were determined (Table 1). In addition, luciferase activities were assayed from pInsII-luc, a fusion construct containing rat insulin II promoter sequences, pRSVluc, a construct containing the RSV long terminal repeat promoter, and pSVOAPL2L, the promoterless luciferase vector. Transcription of the  $\beta$ GK-1000luc fusion gene in HIT M2.2.2 cells was at least 25 times greater than that of the promoterless vector. However, in NIH 3T3 cells there was no difference in expression of the fusion gene

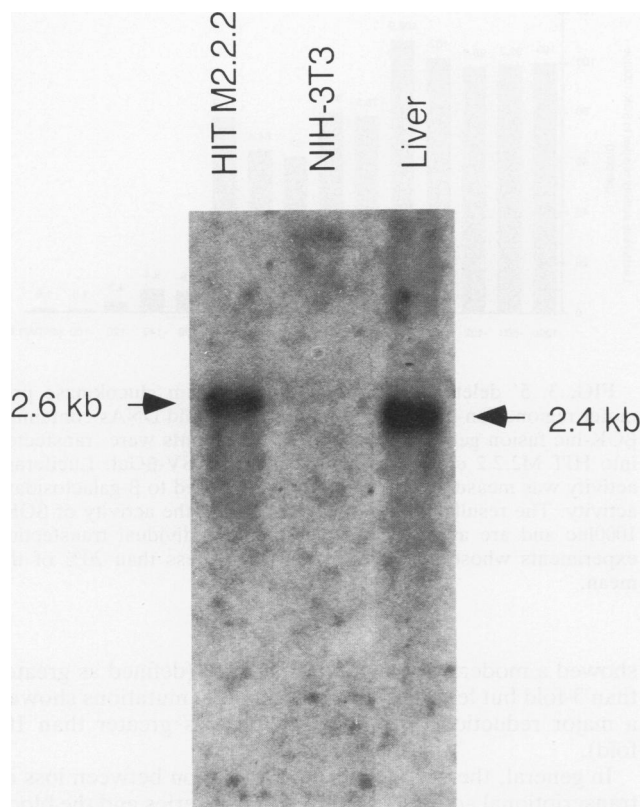


FIG. 1. RNA blot transfer analysis of glucokinase mRNA expression. Ten-microgram samples of poly(A)<sup>+</sup> RNA from HIT M2.2.2 and NIH 3T3 cells and 2  $\mu$ g of poly(A)<sup>+</sup> RNA from the liver of a diabetic rat treated with insulin for 4 h were fractionated on a 1.0% agarose-formaldehyde gel and then blot transferred to a nitrocellulose filter. The blot was hybridized with a <sup>32</sup>P-labeled glucokinase cDNA fragment and then washed three times at 45°C with 0.2 $\times$  SSC-0.1% SDS, after which the autoradiograph shown was obtained.

containing glucokinase promoter sequences and the promoterless vector. The  $\beta$ -cell-specific pattern of expression of the  $\beta$  GK-1000luc gene was similar to that seen for the rInsII-luc gene, although the luciferase activities of the latter were approximately 3.2-fold lower under the same experimental conditions. In contrast, the RSV-luc gene was expressed efficiently in both cell types. The  $\beta$  GK-1000luc gene had about 6% of the activity of the RSV promoter in HIT M2.2.2 cells.

**Analysis of transcription initiation occurring on transfected  $\beta$  GK fusion genes.** To ascertain whether the differences in luciferase activity generated by transfection of  $\beta$  GK-1000luc into the different cell lines were transcriptional in origin, the locations of the transcription initiation sites on two  $\beta$  GK-luc fusion genes were determined. Transcription initiation from the upstream glucokinase promoter region has previously been shown to span at least 60 bp (21). Initial attempts to map these sites by using an RNase protection assay were unsuccessful, presumably because the wide region of transcription initiation made it difficult to distinguish the assay signal from the background. A PCR-based assay was therefore used to determine the approximate location where transcription initiation occurred on the transfected  $\beta$  GK-luc fusion genes. cDNA templates were prepared from RNA isolated from cells transfected with either  $\beta$  GK-1000luc,

TABLE 1. Cell type-specific expression of the upstream glucokinase gene promoter region<sup>a</sup>

Plasmid	Transcriptional activity	
	HIT M2.2.2	NIH 3T3
p $\beta$ GK-1000luc	2,087	41
prInsII-luc	642	75
pRSV-luc	34,800	16,500
pSVOAPL2L	35	145

<sup>a</sup> Ten micrograms of plasmid DNA was cotransfected with 1  $\mu$ g of pRSV- $\beta$ Gal into the indicated cell lines. Luciferase activity was measured 40 h later and normalized to  $\beta$ -galactosidase activity. The results are averages of two representative experiments with each of the cell types. The units of transcriptional activity are peak light output (peak height in millimeters) per unit of  $\beta$ -galactosidase activity.

p $\beta$ GK-100luc, or pSVOALPL2L, and different primer pairs were used for amplification, as illustrated in Fig. 2B. Twenty-five amplification cycles were performed, an empirically determined number which allowed semiquantitative amplification results. Lanes 1 and 2 in Fig. 2 show a Southern blot of the PCR products obtained when cDNA made from cells transfected with  $\beta$  GK-1000luc was amplified with either primers A and C or B and C. A strong signal was obtained with primers B and C, whereas a very weak signal was obtained with primers A and C. This indicates that most transcription initiation on the transfected  $\beta$  GK-1000luc gene occurs between primers A and B, the region where transcription initiation has previously been determined. The same amplifications were performed on cDNA made from cells transfected with a plasmid containing  $\beta$  GK promoter DNA from -100 to +14 and with the promoterless vector itself. As shown and described below, the  $\beta$  GK promoter fragment from -100 to +14 produces no more luciferase than the promoterless vector. Very weak amplification signals were obtained from cells transfected with  $\beta$  GK-100luc (Fig. 2A, lanes 3 and 4), while cells transfected with pSVOAPL2L (Fig. 2A, lanes 5 and 6) showed no amplification product. The finding of either very weak or nonexistent PCR amplification signals from these two plasmids supports the conclusion that measurement of luciferase activity produced by the  $\beta$  GK-luc plasmids is an accurate reflection of  $\beta$  GK-luc mRNA transcription.

**Analysis of 5' deletion mutants.** To locate regions in the upstream glucokinase promoter necessary for the observed transcriptional activity, a series of 15 mutants with different 5' endpoints were constructed and analyzed; each of these had the same 3' endpoint at +14. The results obtained from transfection of these plasmids into HIT M2.2.2 cells are shown in Fig. 3. Relative to the expression of  $\beta$  GK-1000luc, there was no loss of luciferase activity with deletion of DNA to -603, -402, -300, or -280 bp. However, deletion of DNA inside -280 bp resulted in stepwise loss of transcriptional activity. Both the -260 and -242 deletions were about 80% as active as  $\beta$  GK-1000luc. An additional 20% decrement of luciferase activity was seen with deletion of DNA to -221 that was unchanged by the deletions to -200 and -181 bp. The single largest loss of transcriptional activity, from 71% to about 9% of the relative activity, occurred with deletion of the DNA between -181 and -159 bp. Deletion of DNA to -140 bp was without apparent consequence; however, further deletion to -120 bp eliminated half of the remaining activity, leaving 4.3% activity relative to that of  $\beta$  GK-1000luc. Finally, deletion of DNA to -100 bp eliminated all of the remaining transcriptional activity measured.

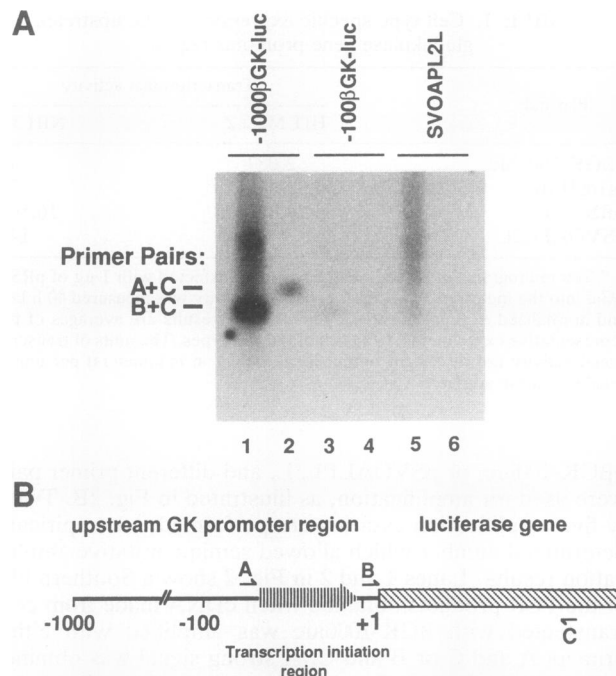


FIG. 2. Analysis of transcription initiation sites on  $\beta$ GK-luc genes by PCR. Total RNA was harvested from HIT M2.2.2 cells 24 h after transfection with either  $\beta$ GK-1000luc,  $\beta$ GK-100luc, or pSVOAPL2L. Equal amounts of total RNA were used to make first-strand cDNA, of which equal fractions were used as templates for PCR amplification. The amplified DNA fragments were size fractionated on an agarose gel, denatured, and then blot transferred to a nitrocellulose membrane. The blot was hybridized with an oligonucleotide complementary to a sequence contained within the amplified fragments. The sequence of oligonucleotide A was AGAA ACCACAGGCCCTCT (identical to positions  $-80$  to  $-62$  in the  $\beta$ GK gene), that of oligonucleotide B was CTGTGACTGAGCTGAT (identical to positions  $+2$  to  $+20$  in the  $\beta$ GK gene), and that of oligonucleotide C was AATGTCATACTGTTGAG (reverse complement to positions  $+394$  to  $+411$  in the luciferase gene [8]). Amplification products using primers A and C are shown in lanes 1, 3, and 5. The products from primers B and C are shown in lanes 2, 4, and 6. The template cDNAs used for PCR amplifications were generated from cells transfected with  $\beta$ GK-1000luc (lanes 1 and 2),  $\beta$ GK-100luc (lanes 3 and 4), and pSVOAPL2L (lanes 5 and 6).

ble, since this DNA fragment produced no more luciferase activity than did the promoterless vector (about 1.5% of that of  $\beta$ GK-1000luc).

**Block transversion mutagenesis analysis.** A higher-resolution analysis of the DNA sequences in the proximal region of the upstream glucokinase promoter that contribute to transcriptional activity in HIT M2.2.2 cells was performed by creating and analyzing a series of block transversion mutations of this region. By using a strategy of base transversions (e.g.,  $A \leftrightarrow C$  and  $G \leftrightarrow T$ ), 28 different mutations of the  $\beta$ GK promoter region were made, each of which altered sequential 10-bp regions from  $-280$  to  $-1$  bp within a promoter fragment from  $-402$  to  $+14$ . The transcriptional activity of each mutant, relative to that of the wild-type promoter fragment ( $-402$  to  $+14$  bp), was determined by transfection into HIT M2.2.2 cells. Both inhibitory and stimulatory effects of the mutations were observed (Fig. 4). Eighteen mutations showed little or no alteration of activity (defined as a less-than-3-fold effect), two mutations stimulated activity above that observed for the wild-type plasmid, six

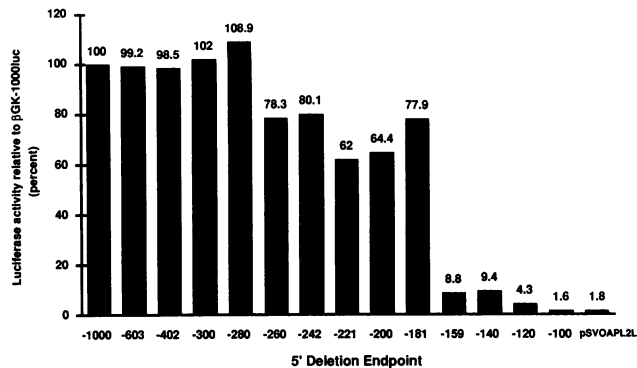


FIG. 3. 5' deletion analysis of the upstream glucokinase promoter region. Ten-microgram samples of plasmid DNAs containing  $\beta$ GK-luc fusion genes with different 5' endpoints were transfected into HIT M2.2.2 cells along with  $1 \mu\text{g}$  of pRSV- $\beta$ Gal. Luciferase activity was measured 40 h later and normalized to  $\beta$ -galactosidase activity. The results are expressed relative to the activity of  $\beta$ GK-1000luc and are averages of at least four individual transfection experiments whose standard deviation was less than 20% of the mean.

showed a moderate reduction of activity (defined as greater than 3-fold but less than 10-fold), and two mutations showed a major reduction of activity (defined as greater than 10-fold).

In general, there was a strong correlation between loss of transcriptional activity in the 5' deletion series and the block transversion mutagenesis series, with two exceptions (Fig. 5). In the first case, no loss of activity was seen for any of the block mutations between  $-280$  and  $-241$  bp, although a 5' deletion between  $-280$  and  $-260$  bp resulted in a 20% reduction in promoter activity; no explanation for this apparent discrepancy was evident. In the second case, no detectable luciferase activity remained inside  $-100$  bp in the 5' deletion series, suggesting that there are no functionally important elements between  $-100$  and  $+14$  bp. However, the block transversion mutagenesis experiments showed decreased transcription with the  $-100/-91$ ,  $-90/-81$ ,  $-80/$

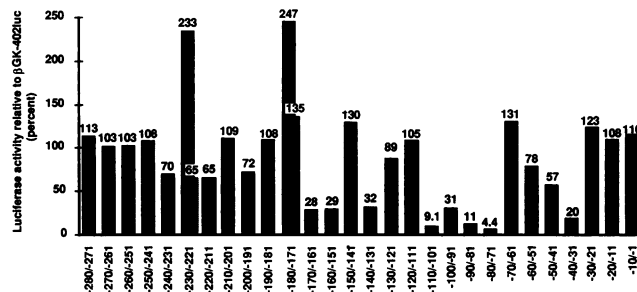


FIG. 4. Block transversion mutagenesis analysis of the upstream glucokinase promoter region. Ten-microgram samples of plasmid DNAs containing  $\beta$ GK-luc fusion genes with different block transversion mutations ( $A \leftrightarrow C$  and  $G \leftrightarrow T$ ; black bars) were cotransfected with  $1 \mu\text{g}$  of pRSV- $\beta$ Gal into HIT M2.2.2 cells. Luciferase activity was measured 40 h later and normalized to  $\beta$ -galactosidase activity. The results are expressed relative to the wild-type activity of  $\beta$ GK-402luc and are averages of at least four individual transfection experiments whose standard deviation was less than 20% of the mean. The shaded bars show the results of two mutations which altered the sequences of  $-230/-221$  and  $-180/-171$  to contain different sequences.

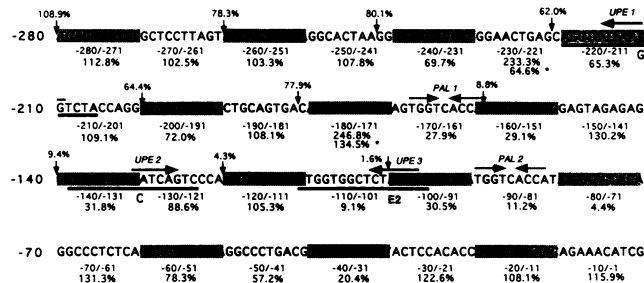


FIG. 5. Sequence of the upstream glucokinase promoter with locations of block transversion mutations. The locations of the block transversion mutations between  $-280$  and  $-1$  bp are shown by the alternating hatched and unhatched areas. The small vertical arrows indicate the locations of the last bases in the 5' deletion mutations shown in Fig. 3. Indicated below each 10-bp region is the effect of the transversion mutation relative to the expression of the unmutated promoter ( $-402/+14$ ). Shown below the  $-230/-221$  and  $-180/-171$  mutations are the effects of a second mutation described in Materials and Methods. The underlined sequences denote the three probes used in mobility shift assays containing the UPEs. The sequences designated by the arrows (UPE-1 to UPE-3, Pal1, and Pal2) are discussed in the text.

$-71$ , and  $-40/-31$  mutants. An explanation for this discrepancy is that sequences upstream of  $-100$  bp are required for transcription and that without these upstream DNA sequences, the function of regions between  $-100$  and  $-31$  is not apparent. Two block transversion mutations,  $-230/-221$  and  $-180/-171$ , stimulated transcription; however, judged by the results of different mutations of the same bases, this appears to have been due to accidental creation of an enhancer element rather than disruption of an inhibitory element.

The single most deleterious mutation identified was  $-80/-71$ , which decreased luciferase activity to 4.4% of the wild-type level. The magnitude of the effect of this mutation can be brought into better perspective by comparing the results of a similar analysis of the rat insulin I promoter. In a study by Karlsson et al., the largest effect of any single 10-bp mutation of the rat insulin I promoter region reduced transcription only 10-fold and this mutation affected the TATA box (16). Other regions in the glucokinase promoter region mutation of which resulted in a greater-than-threefold reduction in luciferase expression were  $-170/-161$ ,  $-160/-151$ ,  $-140/-131$ ,  $-110/-101$ ,  $-90/-81$ , and  $-40/-31$  bp. The noncontiguous nature of the mutations that affect transcriptional activity in HIT cells, which was apparent in both the 5' deletion and block mutagenesis series, argues for the role of multiple elements that together contribute to the total transcriptional activity of the upstream glucokinase promoter in pancreatic  $\beta$  cells.

**Binding of cell type-specific nuclear factors to an element in the upstream glucokinase promoter region.** Inspection of the mutationally sensitive sequences in the upstream glucokinase promoter region (Fig. 5) revealed two copies of the palindromic sequence TGGTCACCA located between both  $-82/-90$  and  $-159/-168$  bp and have been designated Pal1 and Pal2, respectively. Gel mobility shift experiments were performed to determine whether Pal1 and Pal2, as well as other mutationally sensitive sequences in the proximal promoter region, bind cell type-specific nuclear factors. Nuclear extracts from HIT M2.2.2 and NIH 3T3 cells were tested with probes containing sequences of Pal1 and Pal2, and a complex pattern was observed (data not shown); however,

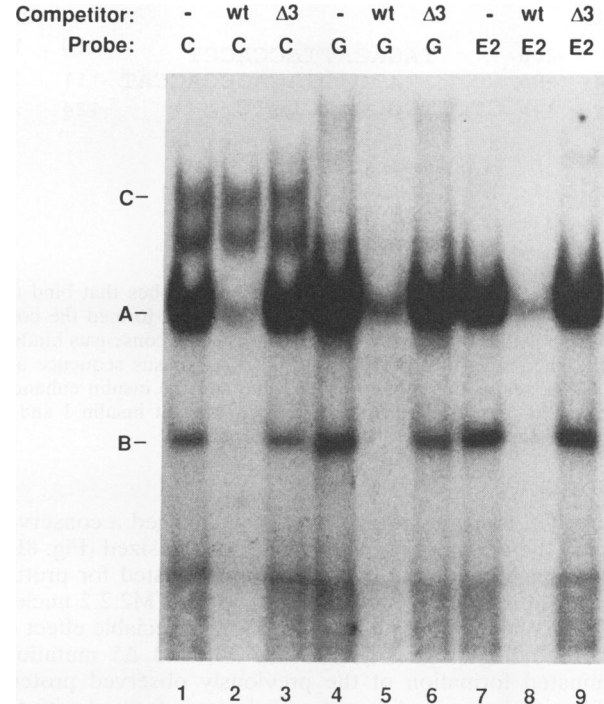


FIG. 6. Binding of factors with similar mobilities to three different regions of the upstream glucokinase promoter. Gel mobility shift analysis was used to examine protein binding to three fragments of the upstream glucokinase promoter region. Radiolabeled probes C, G, and E2, which span the regions  $-220/-206$ ,  $-139/-124$ , and  $-111/-96$ , respectively, in the upstream glucokinase promoter region were incubated with  $20 \mu\text{g}$  of nuclear extract from HIT M2.2.2 cells. The effect of adding a 1,000-fold excess of competitor DNAs is also shown; wt refers to a  $-111/-81$  oligonucleotide pair that contains UPE-3, and  $\Delta 3$  refers to an oligonucleotide pair with the same sequence except for a T $\rightarrow$ G transversion at position 3 of the UPE consensus sequence. An autoradiograph of the resulting gel is shown.

none appeared  $\beta$ -cell type specific. Next, three probes designated G, C, and E2 and containing sequences from  $-220/-206$ ,  $-139/-124$ , and  $-111/-96$ , respectively, were used to assess protein interactions with three other less mutationally sensitive regions of the upstream glucokinase promoter (Fig. 5 shows the sequences of the probes). These probes were each found to form complexes when a nuclear extract from HIT M2.2.2 cells was used (Fig. 6) that were not observed when an extract from NIH 3T3 cells was used (data not shown). Two complexes (designated A and B) with different mobilities were observed when each of the three probes was used. The mobilities of the complexes formed with the G, C, and E2 probes were similar to each other. Since the DNA probes used were also similar in length (22 or 24 bp), we considered that closely related or identical factors bound to the three different probes. Comparison of the sequences of these three probes revealed a similar 6-bp motif with a consensus of CAT(T/C)A(C/G); four of six bases of this consensus were conserved in each of the three probes. Hereafter, these sequences are referred to as UPE-1, UPE-2, and UPE-3, corresponding to the binding sites in the G, C, and E2 probes, respectively (Fig. 7).

Several experiments were performed to characterize further the factor interactions with the UPE sequences in the C, G, and E2 probes. By using the UPE-2 site as the wild type

Probe		UPE	
G	-206	<b>TAGACATTACCACCT</b>	-220 1
E2	-95	<b>ATCATTAGAGCCACCAT</b>	-111 3
C	-139	<b>CCTTTGGCCATCAGTC</b>	-124 2

Consensus	<b>CAT</b> <sup>T</sup> <b>TAG</b> <sup>C</sup>
CT I and II	<b>CATTAG</b> <sup>G</sup> <sup>G</sup>
IEB - ICE	<b>CATCTGT</b>

FIG. 7. Sequence similarities in the three probes that bind the cell type-specific factor. The three probes which formed the complexes shown in Fig. 6 were aligned to identify a consensus binding motif. The similarities between the UPE consensus sequence and the CT boxes of the human insulin gene and the insulin enhancer box-insulin control element (IEB-ICE) in the rat insulin I and II genes are shaded.

(probe C), four new probes that each changed a conserved base in the consensus sequence were synthesized (Fig. 8B). These single-point-mutated probes were tested for protein binding in a mobility shift assay using HIT M2.2.2 nuclear extract. While the  $\Delta 1$  mutation had no appreciable effect on factor binding (Fig. 8A), the  $\Delta 2$ ,  $\Delta 3$ , and  $\Delta 5$  mutations eliminated formation of the previously observed protein-DNA complexes. A different complex was formed with the  $\Delta 2$  mutation, and little or no complex formation was seen with either the  $\Delta 3$  or the  $\Delta 5$  mutation.

To further demonstrate the sequence-specific nature of the DNA-protein complex formed on the UPEs, oligonucleotide pairs containing UPE-3 and UPE-3 with the nonbinding  $\Delta 3$  point mutation were synthesized and used as unlabeled DNA competitors in a mobility shift assay. Excess cold UPE-3 competitor DNA displaced both the upper (A and B) complexes formed when the C, G, and E2 probes were used, whereas the same amount of UPE-3 competitor DNA with the  $\Delta 3$  point mutation had little or no effect on protein binding to these three probes (Fig. 6). Additional evidence that the same factor bound to the C, G, and E2 probes was provided by UV laser cross-linking experiments. Use of a pulsed UV laser with an energy output of 3 mJ/s at 266 nm allowed efficient cross-linking of unmodified thymidine bases in the DNA probes to proteins in crude nuclear extracts (12). As shown in Fig. 9A, lane 1, cross-linking of a protein of ~50 kDa to UPE-2 (probe C) was observed. This interaction was diminished more by addition of excess cold UPE-3 competitor DNA than by addition of the same amount of a  $\Delta 3$ -mutated UPE-3 competitor. The same ~50-kDa protein-DNA complex was formed with the C, G, and E2 probes when both the  $\beta$ -TC-3 and HIT M2.2.2 nuclear extracts were used, although the migration of the band in the HIT M2.2.2 extract was slightly faster than that in the  $\beta$ -TC-3 extract, as shown in Fig. 9B.

**Similar sequences in the glucokinase and insulin promoters.** The sequences of the glucokinase gene UPEs were found to be similar to those of the CT boxes of the human insulin gene and to that of an unnamed element in the rat insulin I promoters (Fig. 7). To determine whether the factor(s) that binds to the glucokinase UPEs also bind to the insulin genes, a mobility shift competition analysis was performed. Both wild-type (B) and point-mutated (Bm2) CT box oligonucleotide pairs were made to match exactly the sequence reported by Boam and Docherty (2). Nonbinding, point-mutated CT box oligonucleotide Bm2 contains the same T→G transversion ( $\Delta 3$ ) that was characterized for UPE-2 in Fig. 8. The

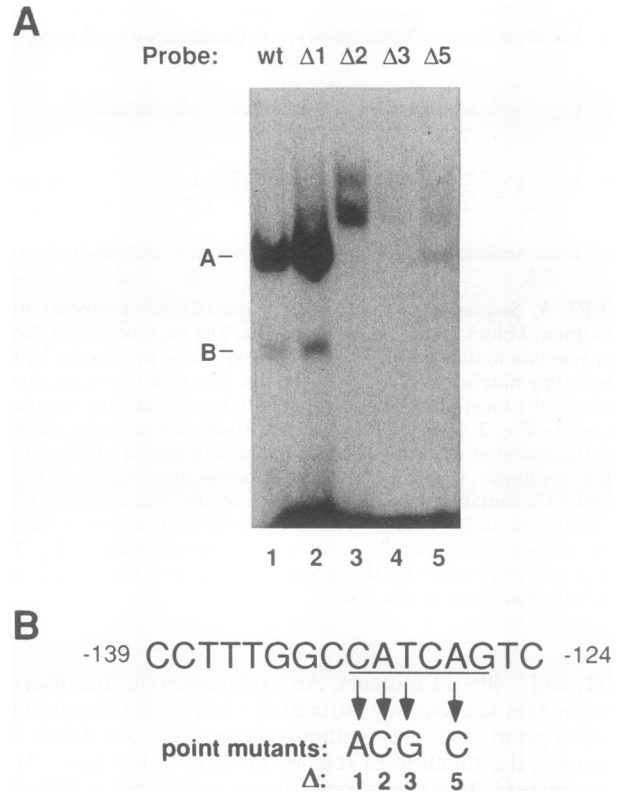
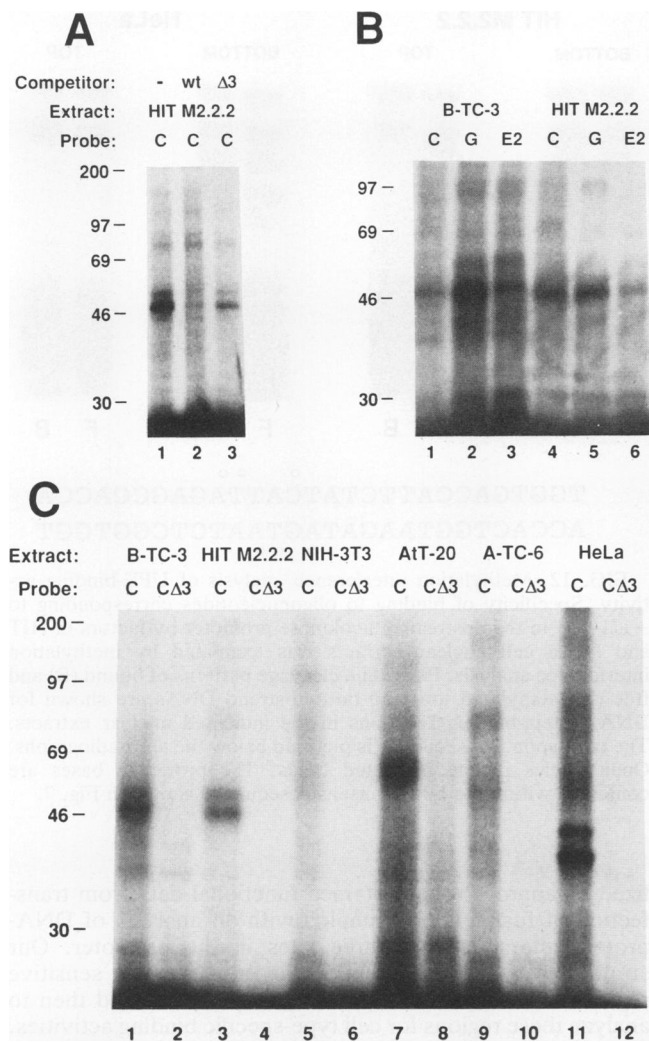


FIG. 8. Point mutational analysis of a factor that binds to UPE-2. (A) Autoradiograph of a mobility shift gel showing complex formation when either probe C or four point mutations of probe C ( $\Delta 1$ ,  $\Delta 2$ ,  $\Delta 3$ , and  $\Delta 5$ ) were used. wt, wild type. (B) Sequence of probe C and locations of the point mutations.

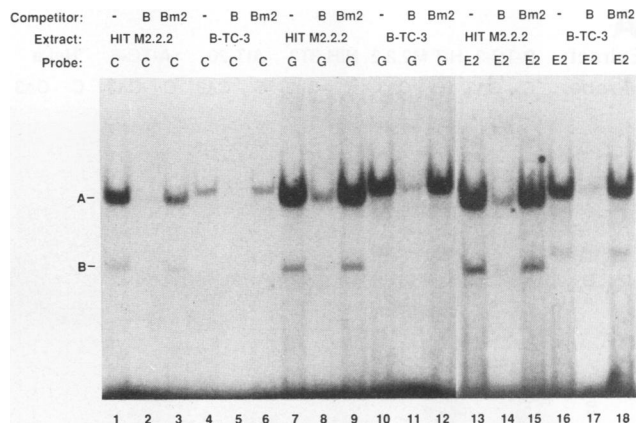
addition of excess wild-type CT box DNA (B) interfered with formation of both the A and B complexes on all three of the glucokinase UPEs, while addition of the same amount of the  $\Delta 3$  point-mutated competitor (Bm2) did not interfere (Fig. 10). The same competition result was obtained by using both HIT M2.2.2 and  $\beta$ -TC-3 nuclear extracts, as is also shown in Fig. 10. These results support the conclusion that the factor that binds to the UPEs in the glucokinase gene has the same DNA-binding specificity as the factor that binds to the human insulin gene, thereby indicating that the same or a closely related factor binds to both genes.

**Cell type-specific distribution of UPE-binding activity.** To determine whether the glucokinase UPE-binding activity was present specifically in pancreatic  $\beta$  cells, as suggested previously by Boam and Docherty, a gel mobility shift analysis of several different nuclear extracts was performed. Binding to the C and  $\Delta 3$  probes was compared for each extract to identify more clearly factors that bind to UPE-2. The two expected complexes (designated at the left of Fig. 11 as A and B) that were formed on probe C by using HIT M2.2.2 nuclear extract were not observed with the  $\Delta 3$  probe. An extract from a mouse  $\beta$ -cell-derived line,  $\beta$ -TC-3, showed bands with slightly slower mobility than the bands seen in the cells of hamster origin. Neither of these complexes (A' and B' in Fig. 11) formed on the  $\Delta 3$  probe, suggesting that they correspond to the A and B complexes from hamster-derived HIT M2.2.2 cells. A similar difference in migration rate for this factor was observed in the UV cross-linking experiment (Fig. 9B), suggesting that the factor



**FIG. 9.** UV laser cross-linking analysis of UPE-binding activity. Cross-linked protein-DNA complexes were obtained by UV laser irradiation of protein-DNA-binding assays. The sizes of the cross-linked protein-DNA complexes were determined by SDS-7.5% PAGE followed by autoradiography. Mobilities of molecular mass standards (in kilodaltons) are indicated at the left of each autoradiograph. (A) Competition analysis. HIT M2.2.2 nuclear extract was incubated with radiolabeled probe C in the presence of a 200-fold molar excess of -111/-81 DNA (wt) or -111/-81 DNA containing a T→G transversion at position 3 of the UPE consensus ( $\Delta 3$ ). (B) Mobilities of cross-linked protein-DNA complexes from either HIT M2.2.2 or  $\beta$ -TC-3 nuclear extracts incubated with radiolabeled UPE-1, UPE-2, and UPE-3 probes followed by UV laser cross-linking. (C) Complexes identified after UV laser irradiation of  $\beta$ -TC-3, HIT M2.2.2, NIH 3T3, AtT-20,  $\alpha$ TC-6, and HeLa nuclear extracts with either probe C or the same probe with a T→G transversion at position 3 of the UPE consensus sequence (CA3). Molecular masses of markers in kilodaltons are shown to the left of each panel.

that binds the UPEs is of slightly different sizes in hamsters and mice. A non-cell type-specific complex (designated D) was observed in all lanes of this experiment. Another, more slowly migrating complex (designated C) was seen in some lanes but was not cell type specific; in some cases, formation of this complex was sensitive to the  $\Delta 3$  point mutation. No bands with mobilities similar to that of the A' complex were



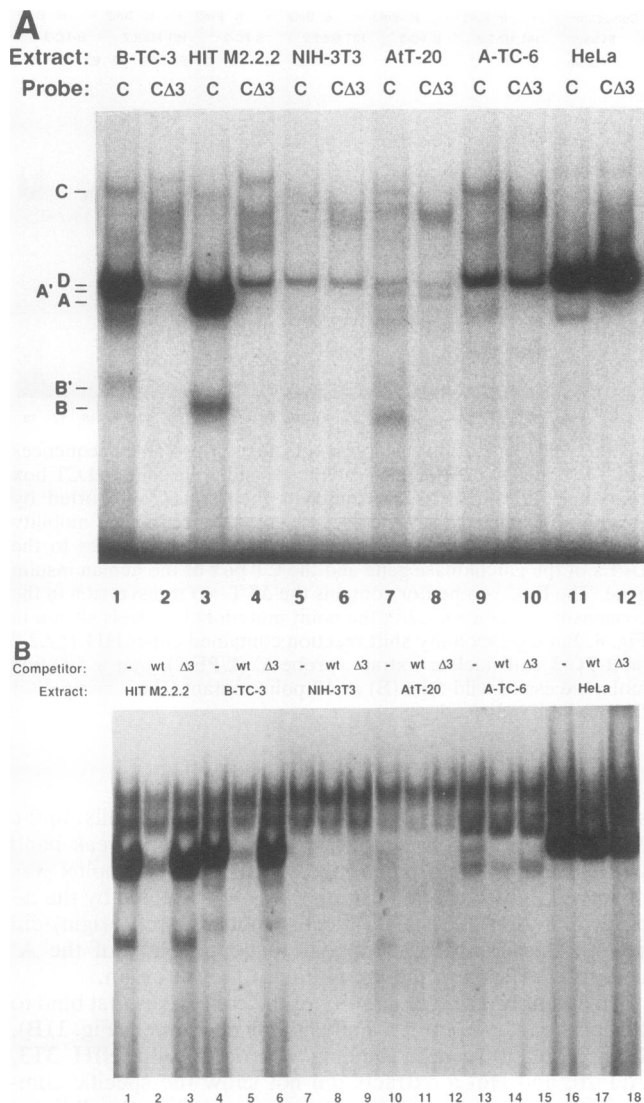
**FIG. 10.** Competition analysis with insulin promoter sequences of UPE-binding activities. Wild-type (B) and point-mutated CT box oligonucleotides (Bm2) that matched the sequence reported by Boam and Docherty (2) were used as competitors in a gel mobility shift analysis to show the relationship of binding activities to the UPEs of the glucokinase gene and the CT box of the human insulin gene. The Bm2 competitor contains the  $\Delta 3$  T→G transversion in the consensus sequence used in the point mutational analysis shown in Fig. 8. Each gel mobility shift reaction contained either HIT M2.2.2 or  $\beta$ -TC-3 cell nuclear extract, probe C (UPE-2), and a 500-fold molar excess of wild-type (B) or  $\Delta 3$  point mutant (Bm2) competitor DNA where indicated.

seen in the extract from mouse-derived NIH 3T3 cells. In the AtT-20 extract (mouse pituitary origin), a very weak band with a mobility approximating that of the A' complex was observed; however, this complex was not affected by the  $\Delta 3$  point mutation. The  $\alpha$ -TC-6 cells (mouse  $\alpha$ -cell origin) did not, in this experiment, appear to contain any of the A' complex, as seen in the  $\beta$ -cell line of mouse origin.

To assess further the identity of the complexes that bind to this probe, a competition analysis was performed (Fig. 11B). Complexes formed in binding reactions with NIH 3T3, AtT-20, and HeLa extracts did not show the specific competition pattern seen for the A complex in two  $\beta$ -cell lines. However, when  $\alpha$ -TC-6 extract was used, a small amount of a complex that migrated at a rate similar to that of the A complex in the  $\beta$ -TC-3 nuclear extract was observed. This complex was competed against more successfully by the UPE-3 competitor than by the  $\Delta 3$  point-mutated competitor (Fig. 11B). UV laser cross-linking of these other nuclear extracts was used as another means to examine the cell type-specific expression of the UPE-binding activity (Fig. 9C). SDS-PAGE analysis of cross-linked  $\beta$ -cell-derived extracts using the C probe (UPE-2) revealed an ~50-kDa doublet, whereas AtT-20 and HeLa extracts showed binding activities of ~65 kDa and a doublet of ~40 kDa, respectively. The cross-linked complex observed in the  $\alpha$ -TC-6 extract was larger than that seen for either  $\beta$ -TC-3 or HIT M2.2.2 cells. The  $\Delta 3$  mutation decreased cross-linking of complexes in all extracts, perhaps owing to the elimination of a photoreactive thymidine residue.

As a final measure of the interaction of a factor from  $\beta$ -cell lines with UPE sequences, a methylation interference assay was performed (Fig. 12). In this experiment, the A complexes from both HIT M2.2.2 and HeLa cells, as seen in Fig. 11A, were examined. Methylation of adenine and guanine residues within the UPE-3 sequence on both the top and bottom strands of the probe interfered with binding of the



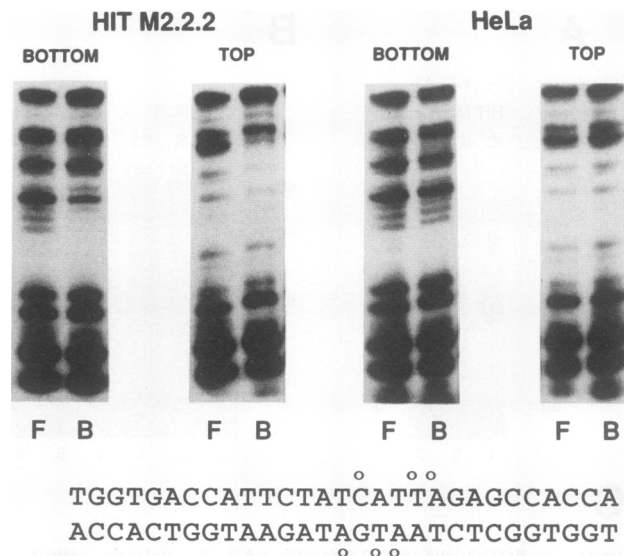


**FIG. 11.** Mobility shift analysis of UPE-2-binding activity in different cell types. (A) Effect of point mutation. Gel mobility shift analysis was used to examine the distribution of proteins that bind to the UPE-2 site. Extracts were prepared from  $\beta$ -TC-3, HIT M2.2.2, NIH 3T3, AtT-20,  $\alpha$ TC-6, and HeLa cells and incubated with either probe C or probe C $\Delta$ 3. After electrophoresis, the autoradiograph shown was obtained. (B) Competition analysis. Extracts from the cell types shown in panel A were incubated with the probe C (UPE-2). The effect of adding a 1,000-fold excess of either a specific competitor (wild type [wt], -111/-81) or a nonspecific competitor ( $\Delta$ 3, -111/-81 with a T $\rightarrow$ G transversion at position 3 of the UPE consensus) is shown.

factor from HIT M2.2.2 cells; however, no interference was seen when HeLa extract was used. The different methylation interference pattern provides convincing evidence for the expression in pancreatic  $\beta$ -cell lines of a factor that binds specifically to the UPE.

#### DISCUSSION

The aim of this study was to begin to delineate the elements and factors responsible for the  $\beta$ -cell type-specific expression of the upstream glucokinase promoter. We uti-



**FIG. 12.** Methylation interference analysis of UPE-binding activity. Specificity of binding to oligonucleotides corresponding to -111/-81 in the upstream glucokinase promoter by factors in HIT and HeLa cell nuclear extracts was examined by methylation interference analysis. Piperidine cleavage patterns of bound (B) and free (F) methylated top- and bottom-strand DNAs are shown for DNA-protein-binding reactions in the indicated nuclear extracts. The corresponding sequence is pictured below the autoradiographs. Open circles denote protected bases. The protected bases are contained within the UPE consensus sequence shown in Fig. 7.

lized an approach that obtained functional data from transfection of fusion genes coupled with an analysis of DNA-protein interactions at three sites in the promoter. Our strategy was first to identify the mutationally sensitive regions of the upstream glucokinase promoter and then to analyze these regions for cell type-specific binding activities. As a model for  $\beta$ -cell type-specific expression of glucokinase, we utilized cell lines that either expressed or did not express the upstream glucokinase promoter. HIT M2.2.2 cells, a  $\beta$ -cell line, were found to express fusion genes containing sequences from the upstream glucokinase promoter efficiently, while NIH 3T3 cells, a fibroblast cell line, did not, thereby providing a cell culture system for studying the transcriptional mechanisms involved in the cell type-specific expression of glucokinase. Similar strategies have been used for transfection on the basis of studies of both the human and rat insulin genes (1, 7, 9, 31).

**Transcription from the upstream glucokinase promoter region involves multiple elements.** Both a 5' deletion analysis and a block transversion mutation analysis were used to identify the DNA sequences important for transcriptional activity of the upstream glucokinase promoter. These different experimental approaches were complementary. In the 5' deletion analysis, maximal transcriptional activity of the upstream glucokinase promoter in HIT cells was preserved until deletion of DNA 280 bp upstream of the designated start site. Thereafter, stepwise loss of luciferase activity was observed, suggesting that several different elements contribute to the transcriptional activity observed with this promoter. These results suggest that all of the sequences necessary for full transcriptional activity of the upstream glucokinase promoter in HIT M2.2.2 cells are contained within a 294-bp DNA fragment (-280 to +14 bp); longer

DNA fragments were no more active in this system. The block mutations corroborated the results of the 5' deletion analysis and, in addition, provided evidence for a greater degree of complexity in the sequences that contribute to the total transcriptional activity in HIT M2.2.2 cells. At least eight of the 10-bp blocks between -280 and -1 bp were sensitive to mutation. Mutations in six of these reduced transcription between 3- and 10-fold, while two others reduced transcription over 10-fold, thus providing convincing evidence for a model involving multiple regulatory elements in the upstream glucokinase promoter that, together, contribute to the total transcriptional activity of this promoter in insulinoma cells.

**Identification of multiple elements in the upstream glucokinase promoter.** Inspection of the mutationally sensitive sequences between -280 and -1 bp suggests that at least two different elements are involved in regulation of expression of the upstream glucokinase promoter in  $\beta$  cells. One element is a pair of perfect palindromic repeats separated by a single base, TGGTCACCA, that have been termed Pal1 and Pal2. This element is found twice in the proximal promoter region at both -168/-160 and -90/-82 (Fig. 5). The sequences of Pal1 and Pal2 appear to be different from those of any other *cis*-regulatory elements identified in either viral or eucaryotic genes. Deletion or mutation of either Pal1 or Pal2 diminished glucokinase fusion gene activity in HIT cells. The block mutation which altered eight of nine bases of Pal1 (-170/-161) exhibited only 27.9% of the transcriptional activity of the wild type, while the block mutation which altered all nine bases of Pal2 (-90/-81) reduced transcription to 11.2% of that of the wild type. The palindromic nature of these elements suggests that they are likely sites for binding of proteins, a notion supported by a preliminary analysis of proteins that bind to this element. However, the role of Pal1 and Pal2, by themselves, in  $\beta$ -cell type-specific expression of this gene remains to be explored. It is possible that factors that bind to this site act in combination with other cell type-specific elements in this gene to enhance its cell type-specific transcription.

A second element that appears to contribute to transcriptional activity of this promoter in  $\beta$  cells is the UPE; three of these elements were identified in the proximal upstream promoter region. While mutation of the UPEs by block transversion analysis suggests that these sequences contribute to transcription of the upstream glucokinase promoter in HIT cells, the contribution of the UPEs may be less than that of Pal1 and Pal2. The block mutations that altered four or more bases of UPE-1, -2, and -3 (-220/-211, -130/-121, and -100/-91, respectively) reduced transcription to 65.3, 88.6, and 30.5% of that of the wild type, respectively. However, the -110/-101 mutation of UPE-3 was the most pronounced and reduced transcription to 9.1% of that of the wild type. Deletion of both UPE-2 and UPE-3, as occurred in the 5' deletion analysis, reduced transcriptional activity from 9.4 to 1.6% of that of the wild type. Judging from this finding UPE-3 may be functionally more important than either UPE-1 or UPE-2, perhaps because its position is closest to the transcription initiation sites. UPE-2 appears to be the least important, since deletion of it had only a minor effect (-130/-121, 88.6%) although removal of this element by 5' deletion reduced transcriptional activity by about half. The less severe effect of the -100/-91 mutation, which had two critical sites of UPE-3 mutated (positions 2 and 3), compared with the -110/-101 mutation, which only had one critical site mutated (position 5), argues against simple interpretations of this finding. Since it is likely that more

than one protein-binding site was affected in some of the mutations generated, the apparent discrepancy between the effects of the -110/-101 and -100/-91 mutations may be accounted for by the elimination of a second important protein-binding site in the -110/-101 mutation.

**The UPE-binding activity appears  $\beta$ -cell type specific.** Mobility shift analysis of proteins from the nuclei of six different cell lines representing islet  $\alpha$  and  $\beta$  cells, pituitary corticotrophs, fibroblasts, and cervical carcinoma cells suggests that the UPEs bind a factor that is expressed primarily or exclusively in a  $\beta$ -cell type-specific manner. In the  $\beta$ -cell-derived lines (HIT M2.2.2 and  $\beta$ -TC-3), two different complexes were observed when probes with UPE sequences were used. Whether these two complexes are due to related or different proteins was not determined, although their DNA-binding specificities appear to be very similar, if not identical. A consensus sequence for the UPE of CAT(T/C)A(C/G) was determined by comparing the three elements found in the glucokinase promoter. Mutation of the conserved bases in the UPE-2 motif (positions 1, 2, 3, and 5) showed that three are required for complex formation. Furthermore, methylation interference data showed protection of adenine and guanine residues within the UPE sequence by the factor in HIT M2.2.2 cells but not by a factor in HeLa cells that forms a complex of similar mobility in gel shift assays.

Transcription from the upstream glucokinase promoter region has been detected in both insulinoma cell lines and corticotroph-derived AtT-20 cells (13, 19). An important question, therefore, is whether the same elements and factors are utilized in these different cell types to activate or allow transcription of the glucokinase promoter. While functional data on the activity of these deletions in AtT-20 cells have not been obtained, the factors that bind to UPE-2 in both corticotroph- and  $\beta$ -cell-derived lines have been compared and appear to be different. In two  $\beta$ -cell-derived lines, UPE-binding activity was identified whereas this factor was not detected in AtT-20 cells. Thus, while the UPEs may be involved in  $\beta$ -cell type specific expression of the upstream glucokinase promoter, other elements and factors must be utilized in the pituitary. In cells of pancreatic islet origin, UPE-binding activity was preferentially expressed in cells of  $\beta$ -cell origin, although the possibility of a small amount of the factor in a line of  $\alpha$ -cell origin could not be excluded.

**Summary of evidence for binding of the same factor to glucokinase and insulin promoter regions.** The UPEs in the glucokinase gene are similar to an element in the human and rat I insulin gene enhancers (2, 24), and the factors which bind to the three UPEs in the upstream glucokinase promoter share many traits with a factor termed insulin upstream factor 1 (IUF-1) (2), leading to the conclusion that this is the same DNA-binding activity. IUF-1 binds between nucleotides -210 to -217 and -77 to -84 in the human insulin gene 5'-flanking DNA. The sequence C(T/C)CTAA TG, reported by Boam and Docherty (termed CT boxes I and II), is similar to the reverse complement of the consensus sequence we report (Fig. 7), although the sequence of Boam and Docherty contains an additional 2 bp. The CT boxes in the human insulin gene are conserved in the rat insulin I and II genes (2).

The following data support the common identity of the UPE- and CT box-binding factors. (i) The pattern of complexes seen in mobility shift analysis is the same. Boam and Docherty (2) observed two complexes, a more intense upper band and a less intense lower band; the same pattern is shown in Fig. 9. (ii) The same point mutation (represented by

CA3 in this report and Bm2 in reference 2) prevents formation of the complexes. (iii) The contact points determined in methylation interference experiments of Boam and Docherty coincide exactly with our core consensus sequence and our methylation interference results. (iv) The cell type-specific distributions of the binding activities in our studies are similar. Boam and Docherty identified the binding activity in  $\beta$ -cell-derived RINm-5F and HIT.T15 cells but not in HeLa, HL60, or AtT-20 extracts. We found abundant binding activity in  $\beta$ -cell-derived HIT M2.2.2 and  $\beta$ -TC-3 cells, but not in AtT-20, NIH 3T3, or HeLa cell extracts. (v) The molecular mass of the cross-linked protein, ~50 kDa, approximates that of IUF-1 (~46 kDa), which Scott et al. have characterized by Southwestern and gel mobility shift analyses (28). (vi) CT box sequences were able to compete for UPE binding, whereas a  $\Delta$ 3 point mutation of the CT box (Bm2) did not compete. Together, these findings indicate that the binding activity we associate with the UPEs in the upstream glucokinase promoter is the same as that which binds to the insulin promoters.

A role for the CT box-binding factor in transcription of the insulin gene has not been clearly defined, although large deletions containing a CT box lead to loss of transcriptional activity (2). The pairing of the CT boxes with a second element in the insulin gene termed the GC motifs has led to the suggestion that juxtaposition of the two elements may constitute a  $\beta$ -cell-specific protoenhancer (2).

Ohlsson et al. (24) have also identified a factor termed insulin promoter factor 1 (IPF-1) that binds to a region between -64 and -85 of the rat insulin I gene. The binding site contains the sequence TTAATG, which is a five-of-six-base-pair match for the reverse complement of the UPE consensus binding site. IPF-1 has been found only in  $\beta$ -cell-derived lines and is not present in an  $\alpha$ -cell-derived line. IPF-1, IUF-1, and the factors described here which bind to the UPEs in the glucokinase gene have similar cell type-specific distribution patterns and appear to bind to closely related sequences. It is likely, therefore, that all three groups have identified the same or very similar factors. Ohlsson et al. (24) have shown that IPF-1 is immunologically distinct from Isl-1, a homeodomain containing LIM protein (17) which binds to AT-rich regions in the insulin promoter. Furthermore, Isl-1 is present in both  $\alpha$  and  $\beta$  cells (24), which clearly distinguishes it from IPF-1.

It should also be noted that the UPEs in the glucokinase gene are distinct from the insulin enhancer box (16) or insulin control element (30) in the insulin genes; sites for binding of a helix-loop-helix heterodimer (6, 24, 30). While the elements are similar at four bases, they always differ at position 5 of the consensus (A in the UPE versus T in the insulin enhancer box-insulin control element; Fig. 7). DNA fragments containing a binding site for the insulin enhancer factor did not compete for binding to the glucokinase gene UPEs (data not shown).

**Concluding comments.** The identification of the same element in the glucokinase and insulin genes suggests that the  $\beta$ -cell type-specific expression of these genes involves some common transcriptional control mechanisms. Future studies pertaining to the function of UPEs and the factors that bind to them should provide insight into the mechanisms that regulate the  $\beta$ -cell type-specific expression of both the upstream glucokinase promoter and the insulin promoter.

#### ACKNOWLEDGMENTS

This work was supported by a grant from the Juvenile Diabetes Foundation and grants from the Public Health Service (DK42612 and DK42502).

We thank R. Stein, D. K. Granner, E. Leiter, D. N. Orth, S. Subramani, and J. Ozer for providing either cells, plasmids, or other reagents used in this study, T. Jetton for photographic assistance, and K. Docherty for helpful discussions.

#### REFERENCES

1. Boam, D. S. W., A. R. Clark, and K. Docherty. 1990. Positive and negative regulation of the human insulin gene by multiple trans-acting factors. *J. Biol. Chem.* **265**:8285-8296.
2. Boam, D. S. W., and K. Docherty. 1989. A tissue-specific nuclear factor binds to multiple sites in the human insulin-gene enhancer. *Biochem. J.* **264**:233-239.
3. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**:5294-5299.
4. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**:156-159.
5. Chrapkiewicz, N. B., C. M. Davis, D. T. W. Chu, C. M. Caldwell, and D. K. Granner. 1989. Rat gene 33: analysis of its structure, messenger RNA and basal promoter activity. *Nucleic Acids Res.* **17**:6651-6667.
6. Cortle, S. R., E. Henderson, H. Masuoka, P. A. Weil, and R. Stein. 1991. Pancreatic  $\beta$ -cell-type-specific transcription of the insulin gene is mediated by basic helix-loop-helix DNA-binding proteins. *Mol. Cell. Biol.* **11**:1734-1738.
7. Crowe, D. T., and M.-J. Tsai. 1989. Mutagenesis of the rat insulin II 5'-flanking region defines sequences important for expression in HIT cells. *Mol. Cell. Biol.* **9**:1784-1789.
8. de Wet, J. R., K. V. Wood, M. DeLuca, D. R. Helinski, and S. Subramani. 1987. Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.* **7**:725-737.
9. Edlund, T., M. D. Walker, P. J. Barr, and W. J. Rutter. 1985. Cell-specific expression of the rat insulin gene: evidence for role of two distinct 5' flanking elements. *Science* **230**:912-916.
10. Faber, M., and L. Sealy. 1990. Rous sarcoma virus enhancer factor I is a ubiquitous CCAAT transcription factor highly related to CBP and NF-Y. *J. Biol. Chem.* **265**:22243-22254.
11. Graham, F. L., and A. J. Van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**:456-467.
12. Hockensmith, J. W., W. L. Kubasek, W. R. Varachek, E. M. Evertsz, and P. H. von Hippel. 1991. Laser cross-linking of protein-nucleic acid complexes. *Methods Enzymol.* **208**:211-236.
13. Hughes, S. D., C. Quaade, J. L. Milburn, L. Cassidy, and C. B. Newgard. 1991. Expression of normal and novel glucokinase mRNAs in anterior pituitary and islet cells. *J. Biol. Chem.* **266**:4521-4530.
14. Iynedjian, P. B., G. Mobius, H. J. Seitz, C. B. Wollheim, and A. E. Renold. 1986. Tissue-specific expression of glucokinase: identification of the gene product in liver and pancreatic islets. *Proc. Natl. Acad. Sci. USA* **83**:1998-2001.
15. Jetton, T. L., and M. A. Magnuson. 1992. Heterogeneous expression of glucokinase among pancreatic  $\beta$  cells. *Proc. Natl. Acad. Sci. USA* **89**:2619-2623.
16. Karlsson, O., T. Edlund, J. B. Moss, W. J. Rutter, and M. D. Walker. 1987. A mutational analysis of the insulin gene transcription control region: expression in beta cells is dependent on two related sequences within the enhancer. *Proc. Natl. Acad. Sci. USA* **84**:8819-8823.
17. Karlsson, O., S. Thor, T. Norberg, H. Ohlsson, and T. Edlund. 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and Cys-His domain. *Nature (London)* **344**:879-882.
18. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488-492.

19. Liang, Y., T. Jetton, E. Zimmerman, H. Najafi, F. M. Matschinsky, and M. A. Magnuson. 1991. Effects of alternate RNA splicing on glucokinase isoform activities in the pancreatic islet, liver, and pituitary. *J. Biol. Chem.* **266**:6999-7007.
20. Magnuson, M. A. 1992. Tissue-specific regulation of glucokinase gene expression. *J. Cell. Biochem.* **48**:115-121.
21. Magnuson, M. A., and K. D. Shelton. 1989. An alternate promoter in the glucokinase gene is active in the pancreatic beta cell. *J. Biol. Chem.* **264**:15936-15942.
22. Matschinsky, F. M. 1990. Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes* **39**:647-652.
23. Miller, J. H. 1972. Assay of beta-galactosidase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. Ohlsson, H., S. Thor, and T. Edlund. 1991. Novel insulin promoter- and enhancer-binding proteins that discriminate between pancreatic alpha and beta cells. *Mol. Endocrinol.* **5**:897-904.
25. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487-491.
26. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. Santere, R. F., R. A. Cook, R. M. D. Crisel, J. D. Sharp, R. J. Schmidt, D. C. Williams, and C. P. Wilson. 1981. Insulin synthesis in a clonal cell line of simian virus 40-transformed hamster pancreatic beta cells. *Proc. Natl. Acad. Sci. USA* **78**:4339-4343.
28. Scott, V., A. R. Clark, J. C. Hutton, and K. Docherty. 1991. Two proteins act as the IUF1 insulin gene enhancer binding factor. *FEBS Lett.* **290**:27-30.
29. Shapiro, D. J., P. A. Sharp, W. W. Wahli, and M. J. Keller. 1988. A high-efficiency HeLa cell nuclear transcription extract. *DNA* **7**:47-55.
30. Whelan, J., S. R. Cordle, E. Henderson, P. A. Weil, and R. Stein. 1990. Identification of a pancreatic  $\beta$ -cell insulin gene transcription factor that binds to and appears to activate cell-type-specific expression: its possible relationship to other cellular factors that bind to a common insulin gene sequence. *Mol. Cell. Biol.* **10**:1564-1572.
31. Whelan, J., D. Poon, P. A. Weil, and R. Stein. 1989. Pancreatic  $\beta$ -cell-type-specific expression of the rat insulin II gene is controlled by positive and negative cellular transcriptional elements. *Mol. Cell. Biol.* **9**:3253-3259.